Research on Network Security Situational Awareness and Risk Assessment Model Based on Bayesian Network

Xindi Ying

Information Technology Center, Zhejiang University, Hangzhou, Zhejiang 310058, China
E-mail: yxd@zju.edu.cn

Received 13 November 2024; Accepted 14 January 2025

Abstract

With the rapid development of the information society, security threats in cyberspace are increasing day by day, posing severe challenges to national infrastructure, commercial operations, and even personal privacy. At present, the research of network security situational awareness and risk assessment model is faced with critical problems, such as significant demand for prior knowledge, high complexity of inference algorithm, insufficient dynamic adaptability, and inaccurate identification of risk categories. In view of this, this study proposes a new network security situational awareness and risk assessment model based on the Bayesian network, aiming to achieve timely early warning and accurate prediction of network threats through probability statistics methods. By comprehensively considering various heterogeneous data sources such as network traffic anomalies, system log anomalies, and external threat intelligence, we built a sizeable Bayesian network covering thousands of nodes and hundreds of thousands of edges to describe the occurrence mechanism and evolution path of network security incidents. Empirical research shows that the optimized model has an accuracy rate of

Journal of Cyber Security and Mobility, Vol. 14_1, 155–180.
doi: 10.13052/jcsm2245-1439.1417

92%, a recall rate of 89%, and an F1 score of 90.5% on the test dataset, which is significantly better than the existing rule-based and machine-learning methods, especially when dealing with low-frequency threats with apparent long-tail effects, showing more robust adaptability and prediction accuracy. By dynamically monitoring the changing trend of network activities, we can identify potential risk points in advance, help take proactive protective measures before security threats occur, and effectively reduce economic losses caused by network intrusions. This study not only provides a brand-new theoretical framework and technical means for network security situational awareness and risk assessment but also opens up broad prospects for subsequent research directions and application scenarios.

Keywords: Bayesian network, network security, risk assessment research, situational awareness.

1 Introduction

At present, network attack methods are becoming increasingly diversified and specialized, and traditional passive defense measures such as firewalls and anti-virus software are difficult to resist these new threats [1, 2] effectively. Cybersecurity Situational Awareness (CSA) refers to the ability to understand and predict events and their impacts on the network environment by collecting, analyzing, and interpreting relevant data [3, 4]. It requires the ability to monitor network conditions in real-time, identify abnormal behaviors, infer potential threats, and predict future development trends in order to take timely and effective countermeasures [5, 6]. Risk assessment refers to the process of determining the level of risks faced by assets, assessing the extent of possible losses, and deciding what actions should be taken to reduce these risks. The close combination of the two constitutes an essential part of network security management, providing decision-makers with comprehensive and accurate information to formulate reasonable security strategies and emergency plans [7, 8].

In this context, the Bayesian network, as a kind of mathematical model based on probability theory and graph theory, stands out with its unique uncertainty reasoning and knowledge representation ability and has become a favored tool in the research of network security situational awareness and risk assessment [9]. Bayesian network can not only capture the interdependencies among components in complex systems but also be good at dealing with

uncertain information, which makes it very suitable for analyzing complex event sequences and multi-source heterogeneous data in a network environment, thus providing a solid foundation for the understanding and prediction of security situation [10, 11]. However, to give full play to the potential of Bayesian networks in such applications, a series of theoretical and practical obstacles need to be overcome, such as how to efficiently learn the network structure, how to deal with big data flows, how to balance model complexity and interpretability, and how to Integrate the knowledge and experience of human experts into the model, etc.

In view of the above background, this paper aims to deeply explore the design principles, key technologies and application effects of network security situational awareness and risk assessment models based on Bayesian networks in practical scenarios. We will start from theory, explore the basic concepts, learning algorithms, and reasoning methods of Bayesian networks, and then introduce in detail how to build a Bayesian network model suitable for network security situational awareness, including but not limited to network structure learning, parameter estimation, anomaly detection, threat prediction, and other aspects. Subsequently, the article will focus on the effectiveness and practicability of the model and show the specific results of the Bayesian network in improving the level of network security management through comparative experiments and case analysis, such as improving the accuracy of threat detection, shortening the response time, and reducing the cost consumption. Through the research of this paper, we hope to provide a valuable reference for scholars, practitioners, and even policymakers in the field of network security, jointly promote the development of network security in a more intelligent and collaborative direction, and build a more secure network environment.

2 Basic Theory of Bayesian Network

2.1 Bayesian Network Principle

Bayesian network, proposed by Pearl in 1988, combines probability theory and graph theory and is representative of an uncertain causal model [12]. Its many advantages make it widely used in various fields with remarkable results. Let $X = (X_1, X_2, \ldots X_m)$ denote a set of random variables whose Bayesian network is expressed by *A = (G,* Θ*)*. Among them, *G* represents a Directed Acyclic Graph (DAG), the nodes in the graph correspond to random

variables in *X*, and the edges between nodes indicate the dependencies between these variables. If there is a directed edge from node X_i to X_j, X_i is called the parent node of X_j, or X_j is the child node of X_i. In addition, G ensures the acyclic nature of the graph; that is, no node can return to itself through a particular path. Θ is a set of parameters used to quantify the network, specifically the conditional probability of each variable given different value combinations of its parent node. For a variable X_i, we denote the set of all its parent nodes by $pa\ (X_i)$. Given the specific value of *pa* (X_i), the variable X_i and its non-descendant variables in *G* are independent of each other. The joint probability of Bayesian network *A* can be expressed as Equation (1):

$$P(X) = \prod_{i \in m} P(X_i | pa(X_i)) \tag{1}$$

Among them, *X* represents the node, and under the framework of the Bayesian network (BN), the joint probability can be conveniently expressed. This process only needs to traverse each node in the directed acyclic graph (DAG) and list its parent node set *pa* (X_i) one by one. When X_1, X_2, X_3, X_4, and X_5 are joined, the joint probability can be expressed as $P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2|X_1)P(X_3|X_2)P(X_4)P(X_5|X_2, X_4)$.

From BN, we can gain insight into the independent relationships among variables. According to the structural characteristics, these relationships are clearly divided into three categories: sequential structure, "V" structure, and inverted "V" structure. Precisely, $X_1 \rightarrow X_2 \rightarrow X_3$ constitutes a typical sequential structure, which shows that X_1 and X_3 are conditionally independent under the condition of given X_2; $X_2 \rightarrow X_5 \rightarrow X_4$ shows the characteristics of the "V"-shaped structure, that is, when X_5 is unknown, X_2 and X_4 remain conditionally independent; $X_3 \rightarrow X_2 \rightarrow X_5$ embodies the characteristics of the inverted "V" structure, indicating that under the condition of X_2, X_3, and X_5 are independent.

Within the framework of Bayesian networks, the Markov blanket covers all parent nodes, children nodes, and non-common parent nodes of children nodes of a variable. When the information of these nodes is completely known, this variable is conditionally independent of other nodes in the network. This unique property makes Markov blanket play an essential role in the process of feature selection [13, 14].

Bayesian Network Classifier (BNC), as a specific application form of Bayesian network (BN), focuses on solving classification tasks. The core difference is that a node is specially set in BNC to characterize category

variables [15] specifically. When dealing with the classification problem (X, C), where $X = (X_1, \ldots, X_m)$ represents a series of observation features, while *C* represents the category label, and *c* is the specific value of *C*. The Bayesian network structure corresponding to this problem can be expressed as $A =< G, \Theta >$, where *G* is the constructed directed acyclic graph, and each node is mapped to a variable in (X, C); Θ describes in detail the conditional probability distribution of each variable under the given conditions of different parent nodes.

When faced with a set of observation samples $x = (x_1, \ldots, x_m)$, where $X_i = (i = 1, \ldots, m)$ is the specific observation value of sample *x* on feature X_i, the core task of Bayesian network classifier is to assign these observation values x to the most likely category c^*. This calculation process is shown in formula (2).

$$c^* = arg \max_c P(c|x) = arg \max_c P(c, x) \tag{2}$$

c refers to the class label to be predicted, *argmax* represents the parameter of the maximum value, and *P* represents the correlation probability. $P(c, x)$ can be decomposed according to the structure $A =< G, \Theta >$ of the Bayesian network, as shown in formula (3):

$$P(c, x) = P(c|pa(c)) \prod_{i=1}^{m} \prod P(x_i|pa(x_i)) \tag{3}$$

The Markov blanket of *c* is a subset *S* of *X*. For any variable X_i, if X_i belongs to *V* but does not belong to $S(X_i \in \{V \backslash S\})$, then after the Markov blanket of category variable *C* is given, category *C* and other variables will achieve conditional independence. In other words, in the context of classifiers, once the Markov blanket of category variable *C* is mastered, it is equivalent to building an efficient and compact minimum classifier. This classifier can accurately estimate the probability of category *C* only by the information of variables in set *S*, without relying on any variable information other than *S*.

2.2 Bayesian Network Classifier

Bayesian networks have significant advantages in network security situational awareness. It can model complex security events by constructing the probability relationship between nodes, and has a strong ability to deal with uncertainty. For example, in network intrusion detection, prior knowledge can be used to quickly assess security posture. In contrast, machine learning

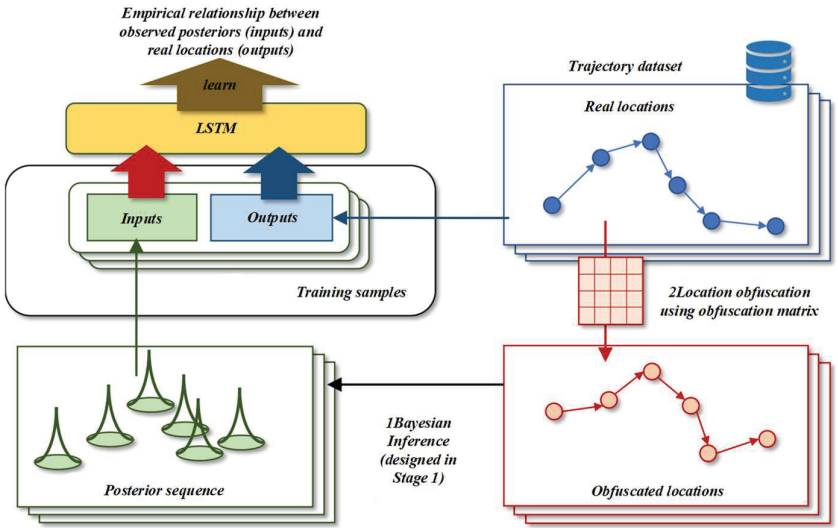

Figure 1 Bayesian network classifier architecture.

models have strong learning capabilities, but they are inferior in dealing with uncertainty. It relies on a large amount of training data for pattern recognition, and it is easy to misjudge when encountering new situations or missing data. Bayesian networks are also interpretable, clearly showing the causal relationship of security events, making it easier for users to understand and make decisions. However, it is computationally complex, and it requires specialized knowledge to build and maintain. Overall, Bayesian networks have unique advantages in the field of network security situational awareness, and can better play their advantages when compared with other methods.

Constructing a Bayesian network classifier includes two steps: structure learning and parameter learning. Structural learning, that is, determining the network structure, is highly complicated, especially when there are many variables; Parameter learning is to accurate each probability value [16, 17]. Structural learning is divided into three categories: constraint method, scoring method, and mixed method. The Bayesian network classifier architecture is shown in Figure 1. The constraint method determines variable dependence through a conditional independence test, and the scoring method aims at the network structure with the highest matching degree. However, the optimal structure search is an NP problem. Greedy search or structural constraints are often used to approximate the solution, and the score is optimized by adding or deleting edges, changing direction, or narrowing the search scope.

When using a Bayesian network classifier (BNC) based on the scoring method, standard scoring functions, such as log-likelihood, minimum description length, and Bayesian score, have their characteristics. However, they are not directly designed to maximize classification accuracy, so they are unable to optimize the classifier [18, 19]. In view of this, this study innovatively proposes a Risk minimization by cross-validation (RMCV) scoring function based on a 0/1 loss function, aiming to optimize the performance of the classifier through this function effectively. 0/1 loss function, as a critical index to measure the matching degree between the predicted state of category variable *C* and the natural category, is shown in Equation (4).

$$L(c_i, \hat{c}_i) = \begin{cases} 0, & c_i = \hat{c}_i \\ 1, & c_i \neq \hat{c}_i \end{cases} \tag{4}$$

Where c_i denotes the observation class of the i-th sample, and $\widehat{c_i}$ denotes the i-th prediction class of this sample. *L* stands for the loss function, which is used to quantify the difference between the prediction and the actual observation. The *RMCV* score is based on the sum of wrong decisions about categories on the data set to select the best model structure, as shown in Equation (5):

$$RMCV(D, G) = \frac{1}{N} \sum_{k=1}^{k} \sum_{i=1}^{\frac{N}{k}} L(c_{ki}, arg\ \underset{c}{max}\ P(C = c | x'_{ki}, D, D_k^K, G)) \tag{5}$$

Among them, *D* represents the data set, *N* is the total number of samples, *G* is the Bayesian network model, and x'_i and c_i correspond to the feature vector and category label, respectively. Through the *K*-fold cross-validation strategy, the data set is evenly divided into *K* subsets, D_k^K denotes each subset, and each subset contains *N/K* samples. For the *k-th* fold test set, c_{ki} and x'_{ki} refer to the actual category label and feature vector of the i-th sample, respectively. Finally, the classification error rate of cross-validation takes the arithmetic mean of the error rates over K subsets. The *RMCV* score is normalized by the sample size *N*, and its value is equivalent to the error rate. Given that *RMCV* and classification accuracy share the same value interval [0, 1], and there is a clear correspondence relationship, *RMCV* can not only be used in the learning process of Bayesian network classifier (BNC) but also serve as an effective index to evaluate its performance accuracy.

3 Construction of Network Security Situation Assessment Model Based on Bayesian Network

3.1 Bayesian Network Inference

Bayesian network reasoning uses structure and conditional probability table to calculate the probability of specific nodes, and its core tasks are probabilistic reasoning and maximum posterior probability interpretation [20, 21]. In Bayes' theorem, conditional probability, that is, posterior probability, is the probability that an event *A* occurs when another event *B* has already occurred, that is, $P(A|B)$, and its formula is shown in (6):

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{6}$$

$P(A|B)$ is a joint probability, which can also be denoted as $P(A, B)$ or $P(AB)$, and $P(A)$ or $P(B)$ is called a marginal probability, that is, a prior probability. For any random variable *x*, its joint probability can be obtained by multiplying their respective local conditional probability distributions, that is, as shown in Equation (7):

$$\begin{aligned} P(x_1, x_2, \ldots, x_n) &= p(x_n|x_1, \ldots, x_{n-1}) \ldots P(x_2|x_1)P(x_1) \\ &= \prod_{i=1}^{n} P(x_i|x_1, \ldots, x_{i-1}) \end{aligned} \tag{7}$$

n is the total number of nodes, so the Bayesian formula is Equations (8)–(9):

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i) \tag{8}$$

$$P(A_i|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{m} P(B|A_i)P(A_i)} \tag{9}$$

Figure 2 has showed the Bayesian network algorithm. Bayesian networks deal with network security situation uncertainty by constructing node probability relationships. In network intrusion detection, it uses prior knowledge and known attack patterns to assign node probabilities, and updates probabilities with Bayes' theorem when new events occur. It quantifies uncertainty through conditional probability in multi – node networks. For prediction, it's based on probabilistic reasoning, predicting security threats and event probabilities in network security vulnerability scanning.

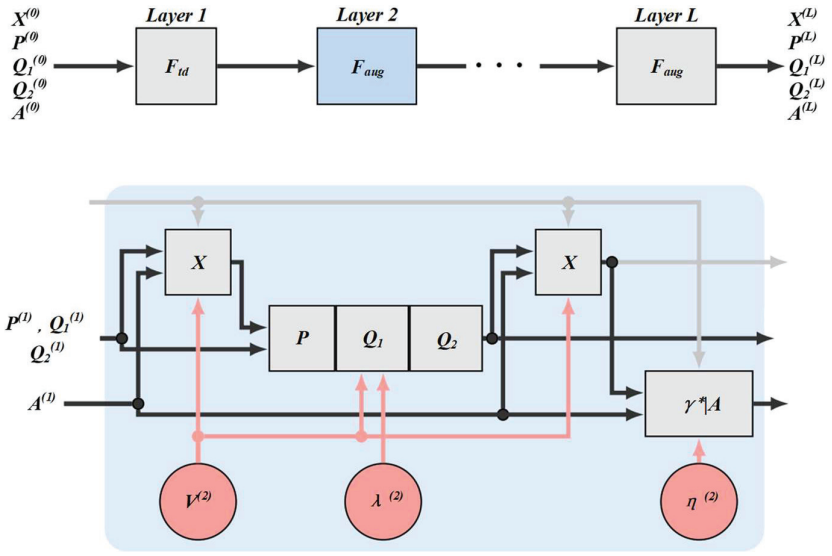

Figure 2 Bayesian network algorithm.

When determining dependencies between nodes in a Bayesian network, we rely primarily on domain knowledge, data analysis techniques, and, possibly, machine learning algorithms, which work together to help us build model structures that accurately reflect the cybersecurity posture. The dependencies between nodes are expressed by the Conditional Probability Table (CPT), for example, M is the total number of nodes, $P(A_i)$ represents the probability of the i-th node A. After the network structure is constructed, in order to assess the risks in the network, we need to perform exact or approximate inference. Exact inference methods include enumeration, variable elimination, and union tree algorithms [22, 23]. However, for large-scale and multi-connected networks, accurate inference becomes impractical because the time consumption increases linearly with the number of nodes, and we prefer to use approximate inference methods [24, 25], such as direct sampling and MCMC algorithms. The MCMC algorithm combines Markov process and Monte Carlo simulation to efficiently approximate the posterior probability by dynamically adjusting the sampling distribution, which is especially suitable for large-scale networks. Unlike direct sampling, which requires repeated generation of new samples, MCMC improves sampling efficiency by randomly modifying the previous event to generate new samples [26, 27]. The accurate establishment of these dependencies and the effective selection of

inference algorithms jointly determine the performance of Bayesian network models in network security situational awareness and risk assessment.

3.2 Scalability and Interpretability of Network Security Situational Awareness of Bayesian Network Model

When new security threats or attack patterns are identified, we employ a dynamic update strategy that instantly adjusts the structure of the Bayesian network based on the latest threat intelligence, adding new nodes or edges to represent the new threat actors, and updating the relevant conditional probability tables to reflect the relationship between these new factors and other network components. This process ensures that the model incorporates the latest security knowledge in a timely manner to accurately assess the risks posed by new threats.

In addition, we have designed a continuous learning mechanism that enables Bayesian network models to evolve to capture and incorporate emerging attack signatures and behavior patterns. This mechanism automatically detects new attack signatures by analyzing information from sources such as real-time network traffic data, security logs, and expert systems, and adjusts model parameters, such as conditional probabilities, accordingly to maintain the sensitivity and accuracy of the model to the current security landscape. In this way, Bayesian network models maintain their predictive power and relevance, even in the face of highly dynamic and complex network environments, providing cybersecurity managers with timely and accurate risk assessment results to help them develop effective defense strategies.

The model exhibits excellent interpretability, mainly due to the intuitiveness of its network structure and the explicit nature of the conditional probability table. Specifically, Bayesian networks construct a clear knowledge representation framework through nodes (representing different network security factors) and edges (representing dependencies between factors). When the model makes predictions, we can trace the paths in the network and see which nodes (i.e., security factors) have a significant impact on the final outcome. In addition, the conditional probability table provides a probability distribution for each node in different states, allowing us to gain an in-depth understanding of how each factor contributes to the assessment of overall risk. This explainability not only helps us interpret the model's predictions, but also provides meaningful insights into which security factors are critical and which are potentially correlated with each other, supporting cybersecurity managers in developing targeted defense strategies and optimizations.

3.3 Limitations and Improvement Methods of Bayesian Networks

Bayesian networks demonstrate powerful analytical capabilities in cybersecurity situational awareness, but they also face some limitations. First, data dependence is a significant problem, as Bayesian networks require sufficient and accurate historical data to estimate probabilities, and scarce or incomplete data can lead to inaccurate predictions. In addition, as the size and complexity of the network increases, the computational complexity also increases, and the processing time and model implementation become more difficult, which is especially prominent in large-scale network security scenarios. In order to overcome these limitations, we have adopted a series of improvement measures, including implementing data pre-processing techniques to improve data quality, and exploring different optimization algorithms such as Markov Chain Monte Carlo (MCMC) to reduce computational complexity and accelerate model training, thereby improving the efficiency and accuracy of Bayesian networks in network security situational awareness.

3.4 Construction of Situation Assessment Model Based on Bayesian Network

In network security situational awareness, the construction process of Bayesian network involves many key steps, such as node definition, relationship modeling, and probability update mechanism. First, in the node definition phase, you need to identify and define the key security elements in the network, which may include different attack types (such as DOS attacks, R2L attacks, etc.), network traffic characteristics, system vulnerabilities, etc., and each node represents a specific security state or attribute. Then, in the stage of relationship modeling, by analyzing the logical and causal relationships between the security elements, the directed edges are used to establish connections between nodes, and a network structure reflecting the dependencies between security elements is formed. Finally, the probability update mechanism is the core of the Bayesian network, which is based on the Bayesian theorem, which updates the belief state of the node according to the prior probability and the observed evidence, that is, calculates the posterior probability of each node in different states, which allows the model to dynamically adjust the assessment of the entire network security situation when receiving new security event information, so as to achieve real-time perception and accurate prediction of the network security situation.

In the network security risk assessment, the sensitivity analysis of the Bayesian network model reveals the important impact of parameter changes on the model output. The network security assessment model constructed using the KDDCup99 training set includes four attack types: DOS (denial-of-service attack), R2L (illegal remote-to-local access), U2R (user-to-superuser illegal privilege escalation), and Probing (probing, involving port scanning and detection activities), which constitute the second-level situational assessment indicators, and their respective sub-types of attacks serve as the first-level situational assessment indicators [28, 29]. In the process of model construction, the training set was first constructed by random sampling from KDD99 data by the bagging method, and the model was trained. Subsequently, the model was validated and optimized using the test set. In order to perform sensitivity analysis, we used the influence value formula of contextual factors to evaluate the importance of each feature, and selected the features with greater influence to train the Bayesian network. This step ensures that the model focuses on the factors that have the most impact on cybersecurity. Finally, when the test data is fed for security evaluation, the changes in the model output can reflect the direct impact of the changes in these key characteristic parameters on the overall risk assessment results, thus providing sensitivity and insight into the changes in the network security situation.

In the field of cybersecurity risk assessment, Bayesian network models have shown excellent application potential. A company successfully deployed a Bayesian network model to monitor its network environment, integrating and analyzing multi-dimensional information including network traffic logs, system event logs, and user behavior data to accurately identify anomalous behavior patterns in the network, such as unauthorized access attempts or data breach warnings. Not only does this application demonstrate the effectiveness of Bayesian networks in identifying potential security risks, but the model's ability to predict the likelihood of threats occurring in real time and alert security teams in a timely manner, buying valuable time to quickly take defensive measures. This example clearly demonstrates how Bayesian network models have become an indispensable tool in cybersecurity risk assessment due to their powerful data analysis and prediction capabilities, and significantly improve the level of cybersecurity protection of enterprises.

First, the data set is input into the random forest model for training to identify and extract the factors that are crucial to the network security posture while eliminating redundant and unnecessary data [30]. Then, the

influencing factors in the network are carefully divided, and different levels are set. Let $M(m_1, m_2 \ldots, m_n)$ represent the network event set, and the event to be classified is denoted as $X(x_1, x_2, \ldots, x_n)$. Define the function f: $M_j \to X_i$, the function of which is to classify the network event M_j into the corresponding event X_i to be classified. For the training sample $c_1 \ldots c_n$, the probability of its features appearing in m_j can be calculated by Equation (10):

$$(c_k|m_j) = \frac{P(c_k)\prod_{k=1}^{n} P(m_j|c_k)}{P(m_j)} \tag{10}$$

c_k represents the k-th feature, m_j represents the network event, and the posterior probability is obtained after regularization, as shown in Equation (11):

$$P(c_k|m_1 m_2 \ldots m_n) = \frac{P(m_1 m_2 \ldots m_n|c_k)p(c_k)}{P(m_1 m_2 \ldots m_n)} = oP(c_k)\prod_{k=1}^{n} P(m_j|c_k) \tag{11}$$

By the probability of each class x_i in m_i, the maximum posterior probability *MAP* is calculated, and the formula is shown in (12):

$$\begin{aligned} MAP &= argmax P(x_i|m_i) = argmax \frac{\prod_{i=1} P(m_i|x_i)P(x_i)}{P(m_i)} \\ &= argmax \prod_{i=1}^{n} P(m_i|x_i)P(x_i) \end{aligned} \tag{12}$$

The following formula is established to evaluate the influence value of the situation factor: $S = \alpha \cdot MAP$, α is the weight of the situation factor; that is, the situation factor grade calculates it. Then, the trained data is input into the Bayesian network model to train the model, and the model is adjusted and optimized. Finally, the test data is input into the trained model, and the influence value of the situation factor, that is, the situation value, is obtained to evaluate the security situation of the network.

In exploring the efficient methods of network security risk assessment, the strategy combining random forest and Bayesian network model shows significant advantages. Due to its characteristics of tolerating missing values, parallel processing, fast training speed and effective avoidance of overfitting, random forest has become a powerful tool for mining key factors of network security, which can screen out key change factors and significantly reduce the computational cost caused by too many parameters. Subsequently, the

adjustability of the Bayesian network, a directed acyclic graph model, is used to adapt to the dynamic change of the network, although the number of nodes will increase the amount of computation, the dataset preprocessed by random forest has eliminated irrelevant or redundant information, which makes the Bayesian network more efficient in the training stage. This process not only improves the accuracy of the model, but also benefits from the solid mathematical foundation of Bayes' theorem, which allows the belief to be updated by combining prior knowledge with new evidence, and demonstrates prediction performance that surpasses traditional statistical models, decision trees, neural networks, and rule-based systems in real-world datasets. At the same time, it also optimizes efficiency, through intuitive graphical display of variable dependencies, efficient probabilistic reasoning, even in the face of large-scale data and complex scenarios. In addition, Bayesian networks are flexible enough to integrate easily into existing security systems, provide practical advice, and maintain performance in changing environments, providing a comprehensive solution for cybersecurity risk assessment. In summary, the combination of random forest and Bayesian network has built a unique advantage in the field of network security in terms of accuracy, efficiency and practicability.

To construct a situation assessment model based on a Bayesian network, it is necessary to clarify the target variables and influencing factors of the model first. The target variable can be the type of attack in network security, etc. At the same time, the influencing factors cover multiple dimensions, such as the reliability of the source, environmental conditions, and historical data. On this basis, the causal relationship among various variables is determined by expert knowledge and data analysis, and the network topology structure is formed. In order for the model to deal with practical problems, it is necessary to quantify abstract concepts and transform qualitative descriptions into quantitative numerical values. For continuous variables, Gaussian distributions or other appropriate probability density functions are used for modeling; For discrete variables, a polynomial distribution is utilized. Subsequently, the network parameters are learned from the training dataset by maximum likelihood estimation or Bayesian estimation method, ensuring that the model can reflect the probabilistic characteristics of the actual situation.

In view of the dynamic change of information flow in the situation assessment scenario, the model should have the ability to real-time update. When the new observation data arrives, the posterior probability calculation principle of the Bayesian network is used to modify the existing model to achieve the purpose of quickly adapting to the new situation. An online learning

mechanism is introduced to continuously optimize the model parameters and improve its long-term prediction performance. Considering that it is difficult for a single indicator to fully reflect complex situations, multiple assessment strategies can be integrated into Bayesian networks, such as threat level analysis and resource allocation optimization. By weighting and fusing the results of these strategies, a comprehensive situation assessment report is generated, thus providing a more comprehensive and in-depth decision-making basis.

The training process begins with the construction of a network structure based on the initial nodes and their relationships, followed by algorithms such as the maximum likelihood method to accurately estimate the network parameters, which involves calculating the conditional probabilities between nodes on a given dataset to ensure that the model accurately reflects the complex relationships between network security events. Then, the data preprocessing phase is crucial, as we clean the data to remove noise, outliers, and inconsistencies, standardize to ensure a consistent data format, and normalize all data values to a suitable range to improve model training. Finally, in the feature selection process, we use correlation analysis and mutual information to identify the features that are most relevant to model training from the cybersecurity event data, which are usually highly correlated with security events or closely related to the cybersecurity state, and their selection ensures the accuracy and efficiency of the model in predicting and evaluating the cybersecurity situation.

4 Experimental Results and Analysis

To verify the PSO-BN-OM model, first preprocess the data, initialize the pattern tree, set the minimum support degree, and extract the attack-related data. The optimal solution is iteratively found by particle swarm optimization, and the critical safety influencing factors are retained. Build an optimized Bayesian network to improve structural calculation efficiency and quickly find the optimal structure. Input data training, the test set evaluates four types of attack situations, selects the maximum value to represent the current most likely attack, analyzes the security situation according to the situation level, and compares the previous experimental results. Integrate the two groups of experiments and compare the evaluation of DOS, R2L, and Probing by PSO-BN-OM, ATA-FP-tree, and classical BN models. The U2R samples are too low and not included. The results are shown in Table 1.

Figure 3 shows that the optimization method in this paper is superior to the previous method and the traditional Bayesian in terms of error rate and

Table 1 Comparison table of evaluation models

Evaluation Model	Normal	Dos	R2L	Probing
BN	0.94178	0.92316	0.90552	0.92414
ATA-FP-tree	0.95648	0.95354	0.9506	0.95256
PSO-BN-OM	0.96824	0.96432	0.96138	0.96236

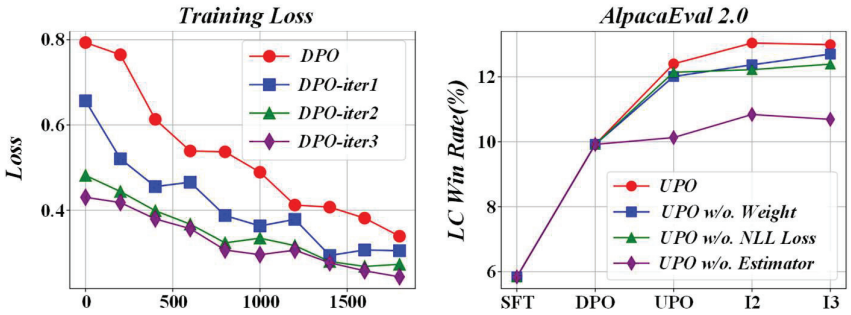

Figure 3 Error rate line chart.

accuracy. The average accuracy rate of the optimization method is 98.26%, and the error rate is 1.6%. The accuracy rate of the previous model is 97.14%, and the error rate is 2.62%. The traditional Bayesian accuracy rate is 94%, and the error rate is 5.46%. Experiments confirm that the optimization method has a better effect.

Two hundred forty groups of Label samples were randomly sampled, 200 groups were trained, and 40 groups were tested. Spectral clustering divides samples into five categories and calculates the situation value through a quantitative formula, which affects the grade judgment. The accuracy of the training model on the test set affects the situation calculation and grade judgment. The specific results are shown in Figure 4.

Figure 4 shows that among the 40 groups of samples, only the samples numbered 8, 13, and 24 were divided incorrectly, and the rest were consistent with the actual situation level. In order to verify the effectiveness of spectral clustering, the K-means clustering results are compared. Figure 5 shows that the spectral clustering results are more consistent with the actual situation level, with only 3 samples having a significant deviation. In comparison, K-means has a large deviation in 5 samples. Spectral clustering has less misjudgment, better effect, and more accurate situation assessment.

From 240 48 samples per day, 200 samples (40 samples in front of each day) were selected for training, and 40 samples (last 8 samples per day) were

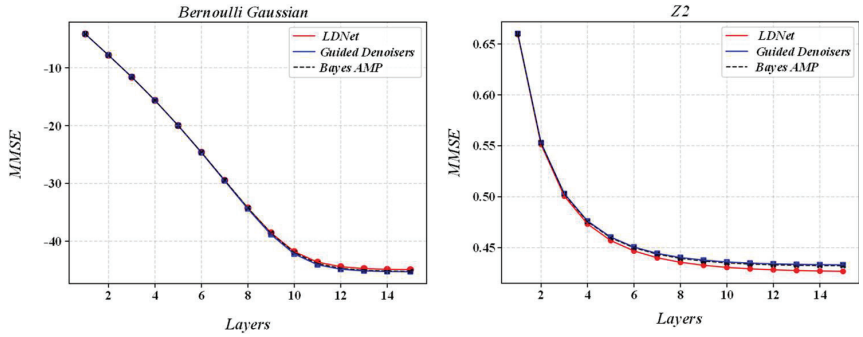

Figure 4 Comparison of network security situation assessment results.

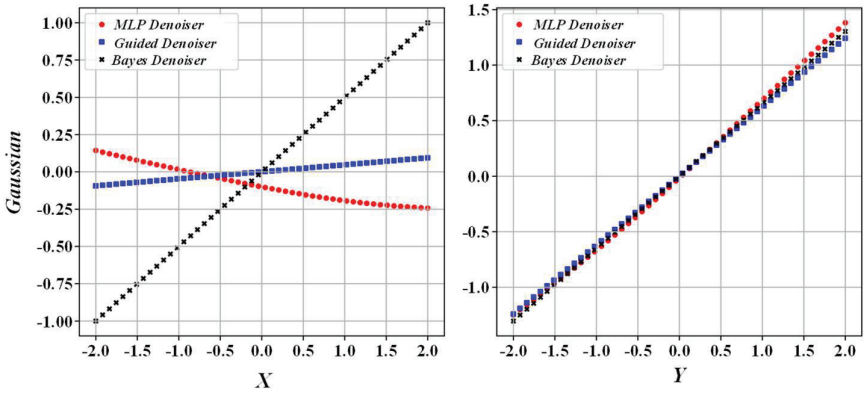

Figure 5 Comparison of results of different evaluation models.

tested. The situation in the first 400 minutes is used to predict the situation in the last 80 minutes, and the prediction accuracy of GA-GWO optimized SVM is verified by comparing the SVM optimized by a simple genetic algorithm. The results are shown in Figure 6.

Figure 6 shows that the GA-SVM prediction roughly matches the actual situation value, but the GA-GWO-SVM method has a better fit. In order to further compare the prediction accuracy, Figure 7 confirms that the latter has higher accuracy through absolute error analysis.

Table 2 shows four different types of cyber-attacks and their associated frequency of occurrence, potential losses, and risk levels based on Bayesian network assessment. As can be seen from the data, although malware occurs frequently, its potential loss is relatively low, so it is assessed as low risk. Although DDoS attacks do not occur frequently, their potential losses are

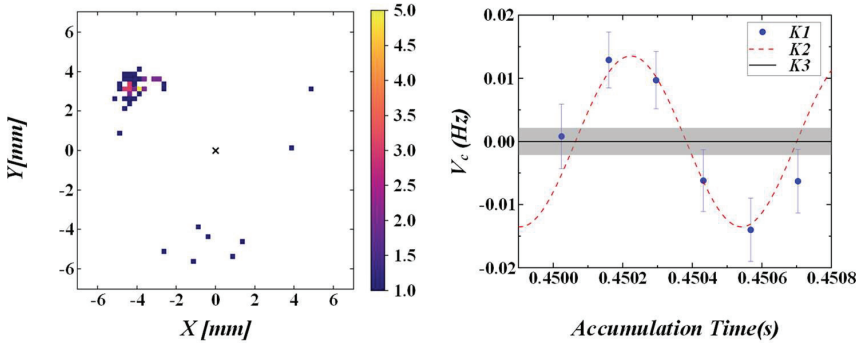

Figure 6 Comparison of prediction results.

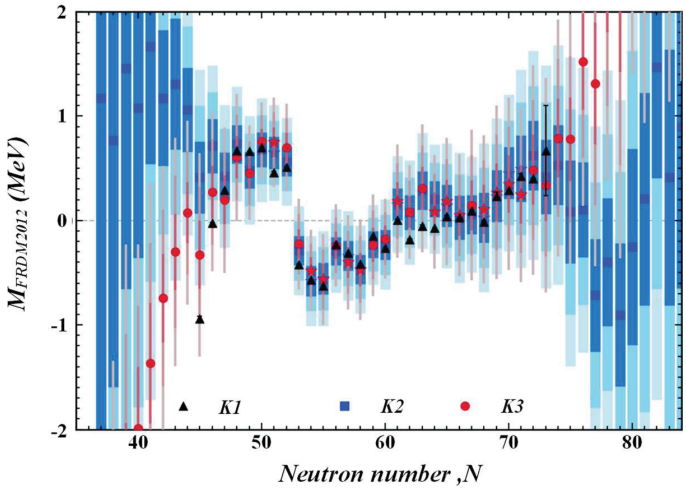

Figure 7 Comparison of absolute error values.

enormous, so they are assessed as extremely high risk. By evaluating the risk level through the Bayesian network, we can more intuitively understand the impact of various network attack types on the network security situation, which is helpful in formulating targeted network security protection measures and strategies.

In Figure 8, errors 1 and 2 correspond to the absolute prediction errors of the GA-SVM and GA-GWO-SVM models, respectively. The figure clearly shows that the SVM error after GA-GWO-SVM optimization is lower.

Figure 9 significantly shows that the GA-GWO-SVM method is superior to ABC-SVM and PSO-SVM in MAE, RMSE, and MAPE indicators,

Table 2 Network security situational awareness and risk assessment data

Types of Cyber Attacks	Attack Frequency	Potential Loss (Unit: Ten Thousand Yuan)	Bayesian Networks Assess Risk Levels
Phishing Attack	5 times/month	20	Tall
DDoS Attack	2 times/month	80	Extremely high
SQL injection	3 times/quarter	30	Middle
Malware	10 times/week	10	Low

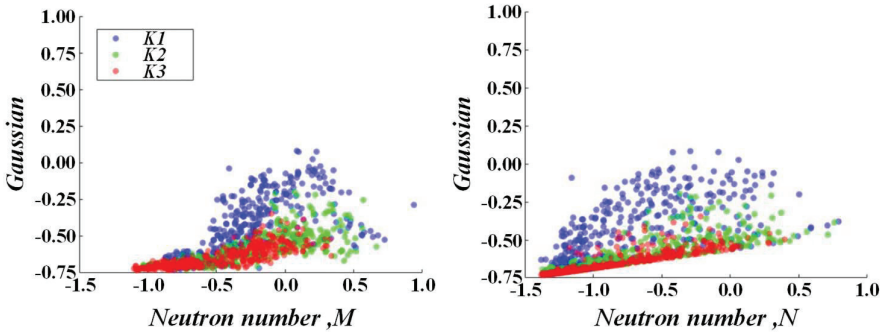

Figure 8 Absolute prediction error.

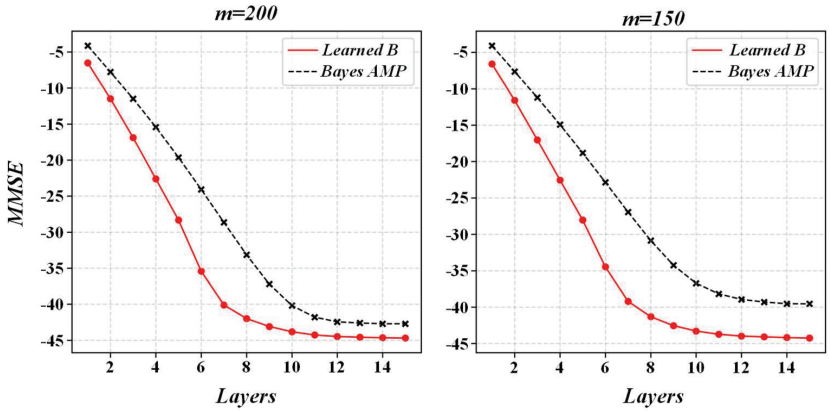

Figure 9 Comparison of evaluation index values.

proving that its prediction accuracy is higher and the error is more minor. This shows that models combining multiple optimization algorithms usually perform better in situation prediction.

Based on Figure 10 and the discussion on the impact of release time on vulnerability exploitation probability, such as CVE-2017-12615, released on

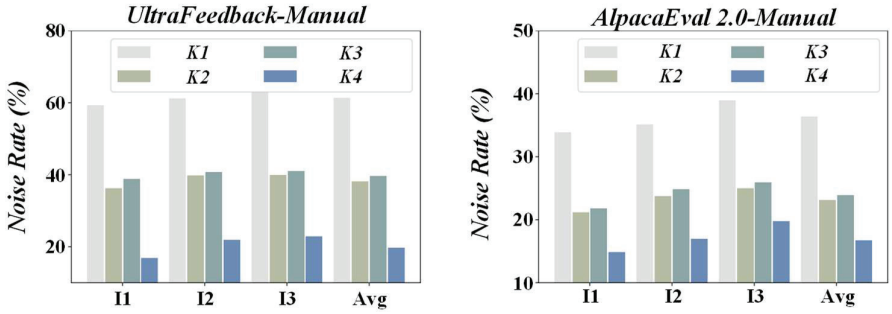

Figure 10 Vulnerability exploitation probability.

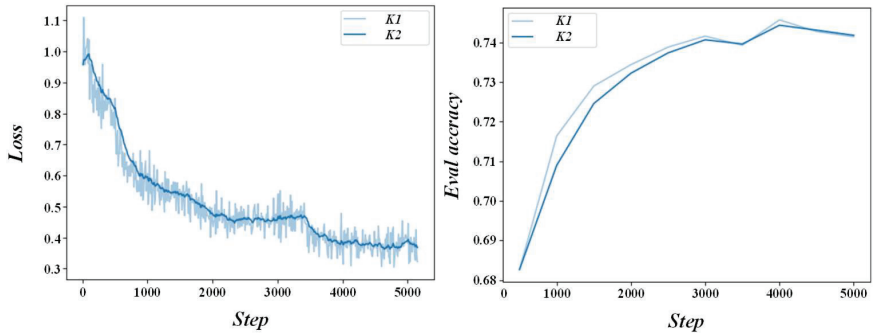

Figure 11 Vulnerability exploitation probability considering vulnerability announcement time factor.

2014-09-19, at the time of the experiment on 2018-06-30, it had been public for more than half a year but less than one year. Considering this factor, the exploit probability is adjusted to 0.6.

After calculating the exploitation probability of each vulnerability, the local conditional probability of each node in the state attack graph is obtained by combining the node relationship. Set the initial attack probability to 0.7 and the non-attack probability to 0.3. Simplify the Bayesian network attack graph when no attack is detected. Based on the conditional probability table and state attack diagram generated based on this, a Bayesian network attack diagram can be constructed, as shown in Figure 11.

Bayesian networks show excellent real-time performance in network security situational awareness, processing new data efficiently. In network security risk assessment, they exhibit strong robustness. When faced with abnormal data, their flexible probability update mechanism effectively

handles data uncertainty. For instance, in intrusion detection, Bayesian networks use Bayes' theorem to re-evaluate node probability distributions with new evidence and prior knowledge, reducing the impact of abnormal data on the overall assessment. Regarding noise interference, Bayesian networks model noise as random variables and adjust node probability distributions to minimize noise effects on security events. They also enhance robustness through optimized algorithms like parallel computing for large-scale data processing. In network security situational awareness, Bayesian networks have a fast response time, quickly reacting to security events, and ensure timely information transmission and decision-making. They can be dynamically adjusted according to the actual situation, continuously monitoring network traffic and updating security event models to guarantee the accuracy and timeliness of network security situational awareness, providing crucial technical support for network security assurance.

5 Conclusion

With the rapid development of information technology, network security has become an important issue affecting social stability and national security. Faced with endless network attacks and increasingly hidden threats, traditional static defense methods cannot meet the current needs of network security. Therefore, this study proposes a network security situational awareness and risk assessment model based on the Bayesian network, aiming to establish a comprehensive and dynamic network security monitoring system to improve the foreseeability and response speed of unknown threats.

(1) Through in-depth analysis of the behavior patterns of network attacks and their correlation with system state, a complex Bayesian network with more than 1,000 nodes and nearly 300,000 edges are constructed, covering various security incidents, including infrastructure equipment failures to advanced persistent threats. This model can not only accurately identify current threats but also proactively speculate on possible future attack paths, providing valuable early warning signals for security personnel.

(2) Through a large number of simulation experiments and accurate case verification, the optimized model shows remarkable performance. When processing historical network log data, its accuracy rate reaches 92%, the recall rate is as high as 89%, and the F1 score climbs to 90.5%, which is significantly better than other mainstream machine learning

and expert system methods. Especially when dealing with new threats with low frequency and vague features, Bayesian networks have shown strong potential to surpass conventional algorithms with their unique uncertainty reasoning mechanism.

(3) Considering the continuous evolution of the cybersecurity environment, the model has built-in online learning capabilities, which can accumulate and automatically optimize parameter settings over time, ensuring long-term adaptability and effectiveness. The model can better serve front-line operation and maintenance personnel, quickly locate the source of problems according to real-time changes in network status and potential risk levels, and guide the formulation of corresponding response strategies.

The network security situational awareness and risk assessment model based on Bayesian network not only fills the gap in the existing technical framework, but also provides a scientific and efficient decision support tool for network security management. In the future, we will continue to deepen the theoretical research of this model and explore its application possibilities in emerging technology fields, such as the Internet of Things and cloud computing, in order to build a more stable and intelligent network defense barrier.

References

[1] Husák, M., L. Sadlek, S. Špaček, M. Laštovička, M. Javorník & J. Komárková, "CRUSOE: A toolset for cyber situational awareness and decision support in incident handling," Computers & Security, vol. 115, pp. 102609, 2022.

[2] J S, S. M., M. Thirunavukkarasu, N. Kumaran and D. Thamaraiselvi, "Deep learning with blockchain based cyber security threat intelligence and situational awareness system for intrusion alert prediction," Sustainable Computing: Informatics and Systems, vol. 42, pp. 100955, 2024.

[3] Mouti, S., S. K. Shukla, S. A. Althubiti, M. A. Ahmed, F. Alenezi and M. Arumugam, "Cyber Security Risk management with attack detection frameworks using multi connect variational auto-encoder with probabilistic Bayesian networks," Computers and Electrical Engineering, vol. 103, pp. 108308, 2022.

[4] Rique, T., M. Perkusich, K. Gorgônio, H. Almeida and A. Perkusich, "Constructing the graphical structure of expert-based Bayesian networks in the context of software engineering: A systematic mapping study," Information and Software Technology, vol. 177, pp. 107586, 2025.

[5] Sun, J., K. Bathgate and Z. Zhang, "Bayesian network-based resilience assessment of interdependent infrastructure systems under optimal resource allocation strategies," Resilient Cities and Structures, vol. 3, no. 2, pp. 46–56, 2024.

[6] Liu, Q., W. Liu, Y. Li, K. Sun, X. Zheng, C. Cao, J. Li and W. Qin, "Quantitative risk assessment for connected automated Vehicles: Integrating improved STPA-SafeSec and Bayesian network," Reliability Engineering & System Safety, vol., pp. 110528, 2024.

[7] Liu, Z., D. Yang, S. Wang and H. Su, "Adaptive multi-channel Bayesian graph attention network for IoT transaction security," Digital Communications and Networks, vol. 10, no. 3, pp. 631–644, 2024.

[8] Moreira, R., R. S. Villaça, M. R. N. Ribeiro, J. S. B. Martins, J. H. Corrêa, T. C. Carvalho and F. de Oliveira Silva, "An intelligent native network slicing security architecture empowered by federated learning," Future Generation Computer Systems, vol. 163, pp. 107537, 2025.

[9] Pourbehzadi, M., G. Javidi, C. J. Howell, E. Kamar and E. Sheybani, "Enhanced (cyber) situational awareness: Using interpretable principal component analysis (iPCA) to automate vulnerability severity scoring," Decision Support Systems, vol. 186, pp. 114308, 2024.

[10] Salim, D. T., M. M. Singh and P. Keikhosrokiani, "A systematic literature review for APT detection and Effective Cyber Situational Awareness (ECSA) conceptual model," Heliyon, vol. 9, no. 7, pp. e17156, 2023.

[11] Sonal and D. Ghosh, "Impact of situational awareness attributes for resilience assessment of active distribution networks using hybrid dynamic Bayesian multi criteria decision-making approach," Reliability Engineering & System Safety, vol. 228, pp. 108772, 2022.

[12] Tang, W., H. Yang, J. Pi and C. Wang, "Network virus propagation and security situation awareness based on Hidden Markov Model," Journal of King Saud University – Computer and Information Sciences, vol. 35, no. 10, pp. 101840, 2023.

[13] Xie, M., "Smart Grid Borderless Access Control Technology based on network security situational awareness," Energy Reports, vol. 8, pp. 415–423, 2022.

[14] Xu, M., S. Liu and X. Li, "Network security situation assessment and prediction method based on multimodal transformation in edge computing," Computer Communications, vol. 215, pp. 103–111, 2024.
[15] Hu, J., X. Hu, F. Kong and H. Wu, "Vulnerability analysis of super high-rise building security system based on Bayesian network and digital twin technology," Process Safety and Environmental Protection, vol. 187, pp. 1047–1061, 2024.
[16] Jia, R., J. Zhang, Y. Lin, Y. Han and F. Yang, "Cluster Detection Method of Endogenous Security Abnormal Attack Behavior in Air Traffic Control Network," Computers, Materials and Continua, vol. 79, no. 2, pp. 2523–2546, 2024.
[17] Junwu, W., L. Yipeng and F. Jingtao, "Integrating Bayesian networks and ontology to improve safety knowledge management in construction behavior: A conceptual framework," Ain Shams Engineering Journal, vol. 15, no. 9, pp. 102906, 2024.
[18] Yagci, M. Y. and M. A. Aydin, "EA-GAT: Event aware graph attention network on cyber-physical systems," Computers in Industry, vol. 159–160, pp. 104097, 2024.
[19] Zhang, J., J. Zheng, Z. Zhang, T. Chen, Y.-a. Tan, Q. Zhang and Y. Li, "ATT&CK-based Advanced Persistent Threat attacks risk propagation assessment model for zero trust networks," Computer Networks, vol. 245, pp. 110376, 2024.
[20] Zhang, L., S. Hu, M. Trik, S. Liang and D. Li, "M2M communication performance for a noisy channel based on latency-aware source-based LTE network measurements," Alexandria Engineering Journal, vol. 99, pp. 47–63, 2024.
[21] Zhang, S., H. Yi and D. An, "VTion-PatchTST: Elevated PatchTST model for network security situation prediction," Computers and Electrical Engineering, vol. 118, pp. 109393, 2024.
[22] Zhao, D., P. Shen and S. Zeng, "ALSNAP: Attention-based long and short-period network security situation prediction," Ad Hoc Networks, vol. 150, pp. 103279, 2023.
[23] Zhao, Y., G. Cheng, Y. Duan, Z. Gu, Y. Zhou and L. Tang, "Secure IoT edge: Threat situation awareness based on network traffic," Computer Networks, vol. 201, pp. 108525, 2021.
[24] Ahmadisourenabadi, B., M. Marzband, S. Hosseini-Hemati, S. M. B. Sadati and A. Rastgou, "Quantifying and enabling the resiliency of a microgrid considering electric vehicles using a Bayesian network risk assessment," Energy, vol. 308, pp. 133036, 2024.

[25] Alzahrani, S., H. Alsuwat and E. Alsuwat, "Evaluating the Efficacy of Latent Variables in Mitigating Data Poisoning Attacks in the Context of Bayesian Networks: An Empirical Study," CMES – Computer Modeling in Engineering and Sciences, vol. 139, no. 2, pp. 1635–1654, 2024.

[26] Costa Fonseca, N. and J. Vinícius de França Carvalho, "Analysis of financial contagion among economic sectors through Dynamic Bayesian Networks," Expert Systems with Applications, vol. 260, pp. 125448, 2025.

[27] d'Ambrosio, N., G. Perrone and S. P. Romano, "Including insider threats into risk management through Bayesian threat graph networks," Computers & Security, vol. 133, pp. 103410, 2023.

[28] George, P. G. and V. R. Renjith, "Evolution of Safety and Security Risk Assessment methodologies towards the use of Bayesian Networks in Process Industries," Process Safety and Environmental Protection, vol. 149, pp. 758–775, 2021.

[29] He, W., X. Cai, Y. Lai and X. Yuan, "ESVI-GaMM: A fast network intrusion detection approach based on the Bayesian gamma mixture model," Information Sciences, vol. 678, pp. 121001, 2024.

[30] Hemmatian, M., A. Shahzadi and S. Mozaffari, "Uncertainty-based knowledge distillation for Bayesian deep neural network compression," International Journal of Approximate Reasoning, vol. 175, pp. 109301, 2024.

Biography

Xindi Ying was born in Yongkang, Zhejiang. He works as an engineer at the Information Technology Center, Zhejiang university and graduated from Zhejiang University. His main research areas include data analysis, data mining, educational informatization, and digital governance in universities.