Analysis and Application of intelligent Power Control System Cyber Security Situation Awareness Based on Wavelet Neural Network

Jingcheng Zhao1*, Xiaomeng Li1, Yaofu Cao1, Junwen Liu1, Junlu Yan1, Chengwei Li1

1 State Grid Information & Telecommunication Branch., Beijing, China

*Corresponding author’s e-mail: jingchengzhao@sgcc.com.cn

Abstract. In recent years, international industrial control network security incidents have occurred frequently. As a core component of the industrial control field, intelligent power control systems are increasingly threatened by external network attacks. Based on the current research status of power industrial control network security, closely combining the development of active monitoring and defense technology in the public network field and the problems encountered by network security operators in actual work, this paper uses data mining methods to study the power control system network security situation awareness technology. Combining operational data collection and integrated processing, situation index screening and extraction, we use wavelet neural network analysis method to train the sampled data set, and finally calculate the true value of the network security status through deep intelligent learning. Finally, we conclude that the artificial intelligence algorithm based on wavelet neural network can be used for power control system network security situation awareness. In actual work, it can predict the situation value for a period of time in the future and assist network security personnel in judgment and decision-making.

1. Introduction

With the rapid development of the power grid and the continuous improvement of the level of dispatch automation, the dispatch data network is rapidly extended to the power plants, substations and new energy stations of the power grid at all levels. The space of the dispatch data network is expanding rapidly, and the number of responsible parties involved is increasing and the composition is increasing. Diversified, the network security management and control of the power system is becoming more and more complex, and new threats and loopholes are constantly emerging. Traditional network security protection methods can no longer cope with the new situation. The power system has become an important target of international cyber warfare. Among more than 200 industrial control system security incidents in America, more than half of the incidents occurred in the energy sector such as electricity. The security protection of intelligent power control systems is facing unprecedented pressure. However, there are very few researches and applications on the existing network security situation assessment and prediction technology in the power control network. Therefore, the research on the network security situation awareness of the power control system has important theoretical value and practical significance. At present, research in this field is still in its infancy. How to analyze and predict the network security situation based on the data collected from the monitoring of the power control system is a very important main research direction. This paper refers to the existing research foundation of public network situational awareness, starting from the data collection and processing, the quantification
and application of situational awareness, and conducts research on power control system network security situational awareness [1].

2. Related theories and technologies
This section introduces the characteristics of power control system, network security management platform, experimental data collection, cleaning and other theoretical basic knowledge related to situational awareness analysis.

2.1. Electric power industry control network structure
The State Grid Corporation of China has established a comprehensive power monitoring security protection system in accordance with the overall security protection strategy of "security zone, network dedicated, horizontal isolation, and vertical certification". The safety zone is the structural foundation of the safety protection system of the power monitoring system. The internal business systems of power generation companies and power grid companies based on computer and network technology are in principle divided into production control areas and management information areas. The production control area can be divided into a control area (also known as the safety zone I, with the highest safety level). The non-control area (also known as the safety zone II, with the safety level second). The horizontal isolation device is used in combination with the vertical encryption authentication device to ensure the confidentiality of the communication channel.

Vertical encryption authentication device: A vertical encryption authentication device is used between the dispatch automation system and the factory station automation system for data encryption transmission.

Horizontal security isolation device: It is used for one-way data transmission in non-network mode from the production control area to the management information area.

The schematic diagram of a typical plant topology structure is shown in the “Fig.1.”

2.2. Data collection and analysis
The power monitoring system network security management platform independently developed and fully deployed by the State Grid Corporation of China. The network security monitoring device is the
source direction of the network security management platform data and the sender of the order. Responsible for collecting safety information of monitored objects, controlling the monitoring objects to execute designated commands, providing security event data to the platform, and supporting related service calls. In accordance with the principles of equipment perception, monitoring device distributed collection, and unified management and control of the management platform, a three-tier logical structure of perception, collection, and control of network security management is constructed. The currently developed network security management platform mainly has five major application functions, including network security monitoring, warning, analysis, auditing, and verification. Security verification is for pre-event; security monitoring and security alarms are for during the event; security analysis and security auditing are for post-event.

Network security monitoring devices are divided into type I and type II. Type I network security monitoring device mainly realizes the monitoring of host equipment, network equipment (intranet switch), security equipment, and database in the master station dispatching agency and national level standby internal security zone I / II / III; Type II network security monitoring device mainly realizes the monitoring of the substation station control layer, the network-related parts of the power plant, the host equipment, network equipment, and safety equipment of the county and local dispatching, and uploading alarm information to the main station platform. The network security management platform data collection method is shown in “Fig.2.”

![Network security management platform data collection method](image)

**Figure 2. Network security management platform data collection method**

All the information that needs to be collected is generated by the device itself, including trusted computing, operating systems, network equipment, security protection equipment, monitoring software, databases, etc. It is submitted to the network security management platform after summary processing by the network security monitoring device on the main station and the plant side. Realize the network security data collection of monitoring system related equipment such as regulatory agencies, plants, stations, power distribution, and negative control, as well as communication and interaction with the management platform. Realize the perception and reporting of network security data of servers, workstations, switches, vertical encryption, forward and reverse isolation and other equipment.
2.3. Acquisition data processing

2.3.1. Data cleaning
Data cleaning is to process the power industrial control network data obtained in different ways, and finally get standard and clean data for subsequent situation assessment and analysis. Generally speaking, data cleaning mainly represents three methods, namely binning method, clustering method and regression method. Each of the three methods has its own advantages, and can perform a full range of data cleaning on missing, abnormal and other noisy data [2].

The binning method is that we put the data to be processed into a box according to certain rules, then test the data in each box, and take different methods to process the data differently according to the actual situation of the box.

The regression method is to use the data in the function to draw an image, and further smooth the drawn image. At present, the commonly used regression methods include single linear regression and multiple linear regression. Single linear regression is to find the best straight line between two attributes in the data, and predict the other attribute based on one attribute. The multi-linear regression is to find multiple attributes from the data, and then fit the data to a multi-dimensional surface, which can also eliminate noise.

The clustering method is to divide abstract objects into different sets, and then find the isolated points outside the set is the noise value, so that the noise data in the data can be found directly, and finally the appropriate method is selected to clear the noise value.

2.3.2. Data integration
Data integration is the physical integration of related data from different sources, different formats, etc., so that they have a unified format and interface. This can reduce data inconsistency and redundancy.

The typical data integration methods are federated database and middleware integration method. Detailed discussion in the paper [3].

The federated database is a mode integration method adopted by early people, and its data sources can share data with each other.

The middleware mode is to hide the underlying data details by providing a unified logical view of the data in the middle layer, so that users can regard the data source as a unified whole to realize the global data mode.

2.3.3. Data transformation
Data transformation is also an important method of data management. Data transformation is to transform the representation of data, such as discretizing continuous data. It mainly expands through data smoothing, data aggregation, and data standardization [4].

- Data smoothing: The purpose of smoothing is to remove the noise value in the data and discretize continuous data. Such as the binning method, regression method and clustering method mentioned above.
- Data generalization: Use higher concepts to replace lower concepts to reduce the complexity of data.
- Data aggregation: functionalize the data and summarize the processed data sets.
- Data normalization: The data is scaled according to different proportions, so that the original value can be mapped to a new specific area.

In this paper, the network security management platform has assumed the role of data processing, integrating and transforming various collected data. In the situational awareness analysis and application in Chapter 4, Only a small amount of noise data needs to be manually removed, and the basic experimental data required for situational awareness analysis can be obtained.
3. Network security situation indicators

3.1. Selection of network security situation indicators

Network security situation indicators generally refer to elements that can describe network security attributes. To establish an indicator system that can comprehensively and objectively reflect the network security situation, only two or three indicators are not enough, because the selected indicators need to cover the network security situation system. However, it is not feasible to select too many indicators, because as the number of indicators increases, it becomes more difficult to determine the priority of the indicators. At the same time, the established indicator system will be too complicated and the amount of calculation will increase significantly. Therefore, the selection of network security situation indicators needs to follow the following principles [5]:

- The principle of independence: Due to the complexity of cyberspace, the cybersecurity attributes referred to by an indicator may often include multiple situational factors. If there are too many situational factors overlap between the selected indicators, it may affect the situational assessment. Therefore, selecting relatively independent indicators and reducing the correlation between indicators can better reflect the network security situation.
- The principle of operability: The selected indicators should be easy to obtain, and the selection of data indicators that are difficult to collect and count in cyberspace should be avoided, and the data of the indicators should be reliable and accurate.
- Principle of principal components: When the data set is too large and features too many, the characteristics that better reflect the changes in the network security status should be selected as indicators, and some indicators that have little impact on the changes in the network security status can be discarded.

The extraction of industrial control security situational awareness indicators is extracted from the five modules of the network security management platform: security monitoring, security alarms, security analysis, security audits, and security verification upload information. The five modules can be set as five first-level indicators, and the detailed data extracted under each module can be set as second-level indicators. The power control system situation assessment index is shown in “Fig.3.”

![Figure 3. Power control system situation assessment index](image)

3.2. Quantitative methods of situational awareness

Due to the complexity of the power control system, some of the data indicators collected and sent through the platform are still nominal data, which cannot be directly input into the evaluation model as
the numerical feature of the sample for training. Therefore, specific mathematical formulas need to be used to quantify different situation indicators. Based on the evaluation method of the Common Vulnerability Scoring System (Common Vulnerability Scoring System), the quantitative formulas for some secondary indicators in 3.1 of this chapter are as follows [6]:

3.2.1. Network security vulnerability level
In most cases, network attacks are initiated by exploiting vulnerabilities in the system, so different vulnerability levels will affect the value of the network security situation. The level of network vulnerabilities is mainly affected by the number of vulnerabilities in the network environment and the types of vulnerabilities. The quantitative formula is as follows:

\[ g = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} W_{ij} D_{ij}}{N} \]  

(1)

In the above formula (1), \( g \) represents the vulnerability level, \( n \) is the number of hosts, \( m \) is the number of vulnerability categories, \( N \) is the total number of vulnerabilities, and \( W_{ij} \) represents the level factor of the \( j \)-th vulnerability category in the \( i \)-th host, which can be passed through CVSS. The scoring standard is obtained. \( D_{ij} \) represents the number of the \( j \)-th vulnerability category in the \( i \)-th host, and \( Q_{ij} \) represents the importance index of the \( i \)-th device. The formula is as follows:

\[ Q_{i} = \frac{t_{i}}{\sum_{i=1}^{n} t_{i}} \]  

(2)

In the above formula (2), \( t_{i} \) represents the importance score of the information stored by the host, and its measurement standards are as follows:

\[ l_{i} = \begin{cases} 
1.0, & \text{Very important} \\
0.6, & \text{important} \\
0.3, & \text{general} 
\end{cases} \]  

(3)

3.2.2. Host device operating status
The security level of the operating state of the host device is not only affected by its hardware level and the type of installed software, but also by the version of the system kernel installed on the device. The indicator quantification formula for the kernel state of the device system is as follows:

\[ E_{y} = \sum_{i}^{n} O_{i} H_{i} \]  

(4)

In the above formula (4), \( E_{y} \) represents the device system kernel status score, \( O_{i} \) represents the power on/off status of the \( i \)-th device, power-on is 0 when it is turned off. \( H_{i} \) represents the operating system kernel version score of the \( i \)-th device.

3.2.3. Network security threat level
The severity of a network security attack will affect the value of the security situation [7]. The attack severity score is mainly determined by the main indicators of the device, the number of attacks, and the type of attack. The formula is as follows:

\[ Y = \frac{\sum_{i=1}^{n} \sum_{j=1}^{k} p_{ij} C_{ij}}{K} \]  

(5)

In the above formula (5), \( Y \) represents the score of attack severity, \( k \) represents the number of attack categories, \( K \) represents the total number of attacks in a certain period of time, \( p_{ij} \) represents the scoring factor of the \( j \)-th type of attack on the \( i \)-th device, \( C_{ij} \) represents the number of times that the \( i \)-th device has suffered the \( j \)-th type of attack, and \( Q_{ij} \) is shown in the above formula (2). Other quantitative indicators are similar to the above-mentioned indicator quantification methods and will not be listed here in detail.
3.3. Classification of network security situation

This article refers to the basic network security situation information released by the National Internet Emergency Response Center and divides the network security assessment levels into five levels: safety, slightly danger, general danger, moderate danger, and high danger [8]. In order to have a more intuitive analysis of the results of the situation assessment, the five security levels are numerically characterized as situational indexes and shown in the following “table 1.”

| Situation level       | Situation value range |
|-----------------------|-----------------------|
| Safety                | [0.00~0.20]           |
| Slightly dangerous    | [0.21~0.40]           |
| General dangerous     | [0.41~0.60]           |
| Moderately dangerous  | [0.61~0.80]           |
| Highly dangerous      | [0.81~1.00]           |

4. Situation Awareness Analysis of Power Control System

The experimental data required in this chapter is obtained in real time from the network security management platform. The data collection time starts at 8:00 on Monday and ends at 16:00 on Tuesday the next week, for a total of 200 hours. After the collection is over, export the data collected by the five major modules of the platform, such as: network security vulnerability level, host device operating status, network security threat level, opening high-risk illegal ports, inserting illegal peripherals, abnormal logins, suspicious system files in the system directory operation, security equipment configuration changes, etc. After removing a small amount of noise data in various data by manual screening, a data set that can be used for experimental analysis is obtained.

First, divide the data set with 60 minutes as a unit, and divide all the data sets into 200 time periods of data. Then, according to the relevant important indicators in the situation indicator system, the effective indicator information is extracted from the data of the 200 time periods, and some indicators are quantitatively converted according to the quantitative formula of the indicator. Finally, we use wavelet neural network analysis method to obtain the true value of the network security situation [9]. Then calculate the network security situation level according to the network security situation level classification standard, and complete the situation awareness analysis.

During the experiment, the data needs to be processed in the same dimension:

$$x_i = \frac{x_i - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}$$  \hspace{1cm} (6)

The $x_{\text{min}}$ and $x_{\text{max}}$ are the minimum and maximum values in the same feature data, respectively.

In the experiment, the first 180 samples in the data set are used as the training set, and the remaining 20 samples are used as the prediction samples. Put the training set samples into the model to train the model, and then use the prediction samples to test the effect of the model. The situation level safe is marked as 1, the situation level slightly dangerous is marked as 2, the situation level general danger is marked as 3, the situation level moderately dangerous is marked as 4, and the situation level highly dangerous is marked as 5. The experimental results are shown in “Fig.4.”
It can be seen from Figure 5 that the situation assessment model based on wavelet neural network incorrectly classified category 4 into category 3 at sample number 11, and the remaining network security situation prediction value are consistent with the real value. The rate can be calculated by the following formula:

\[ S_c = \frac{N}{T_N} \times 100\% \]  

(7)

\[ A_c = \frac{\sum_{i=1}^{n} \frac{N}{T_N} \times 100\%}{n} \]  

(8)

The \( S_c \) is Situation classification accuracy, \( A_c \) is Average classification accuracy, \( N \) is Number of positive classification samples, \( T_N \) is Total number of samples, and \( n \) is the number of experiments. Calculate the average classification accuracy according to formula (7):

\[ S_c = \frac{19}{20} \times 100\% = 95\% \]

The model was tested for 10 times, and the average classification accuracy was 95.08\%, indicating that the situational awareness model is suitable for power control system network security assessment and can predict network security risks accurately.

5. Summary and outlook

This paper conducts situational awareness analysis and research in the field of power industrial control systems, based on the research results of situational awareness of public network security at home and abroad. Using the data collected by the State Grid Corporation's network security management platform as the analysis sample, a data set composed of five modules, including security monitoring, security alarms, and security audits, was used to conduct a situational awareness analysis experiment on the power control system. Using the artificial intelligence algorithm based on wavelet neural network, we obtain the true value of the network security situation. The final experimental results prove the feasibility of applying public network security situational awareness methods to intelligent power control systems, and a reliable situational awareness prediction model has been established, which provides strong support for the network security monitoring work of the State Grid Corporation of China, and provide a research foundation for the subsequent establishment of more intelligent situational awareness models.

This article only predicts and analyzes the network security situation awareness of the power industrial control system, but does not propose corresponding emergency plans for possible risks. The network security risk prediction and emergency response methods after the occurrence of the risk need to be further explored and studied.
Acknowledgments
This work is supported by The Science & Technology Project of State Grid Information & Telecommunication Branch.

References
[1] Lenders V, Tanner A, Blarer A. Gaining an edge in cyberspace with advanced situational awareness[J]. Security & Privacy IEEE, 2015, 13(02): 65–74.
[2] Endsley M R. Design and Evaluation for Situation Awareness Enhancement[J]. Human Factors & Ergonomics Society Annual Meeting Proceedings,1988,32:97-101.
[3] Bass T. Intrusion Detection Systems & Multisensor Data Fusion[J]. 2000,42(4):99-105.
[4] Hu Liu, Liqian Zhou, Jie Deng, Rui Li, Zhengwei Zhao. Network security based on support vector machines and adaptive weights situation assessment model[J]. Computer System Applications,2018,27(07):188-192.
[5] Yaodong Tao, Xintong Jia. Research on Network Security Situational Awareness Framework of Industrial Control System[J]. Information Technology and Network Security, 2018, 37(05): 3-6.
[6] Evesti A, Frantti T. ICUFN 2015 - 7th International Conference on Ubiquitous and Future Networks[C]. Sapporo, Japan: IEEE Computer Society, 2015:1-6.
[7] Yongjie Shi, Huichao Yu, Feng Lu, Chang Zhang, Yaping Wu. Research on Active Defence Technology of Industrial Control System Network Security Research and Practice[J]. Information Technology and Network Security, 2020, 39(04): 13-18.
[8] Xiaoyan Li. Research on network security situation prediction method based on wavelet neural network[D]. Hunan University, 2016.
[9] Ho W, Ma X. The state-of-the-art integrations and applications of the analytic hierarchy process[J]. European Journal of Operational Research, 2018, 267(2): 399-414.