Research on Electric Power Information Systems Network Security Situation Awareness Based on Big Data Technology

Dong-Lan LIU[1]*, Dong LI[2], Lei MA[1], Xin LIU[1], Hao YU[1], Ying-Xian CHANG[2], Jian-Fei CHEN[2]

[1]State Grid Shandong Electric Power Research Institute. Jinan 250003, PR China

[2]State Grid Shandong Electric Power Company. Jinan 250021, PR China

*Email: liudonglan2006@126.com

Keywords: big data, network security, situation awareness, situation prediction, TSA.

Abstract. With the rapid development of the network scale and its applications, network security threats continue to increase, a single network security protection technology could not meet the requirement. Network security situation awareness can dynamically reflect the overall network security and predict network security development trends. Big data analytics technology provides the basis for the research of network security situation awareness. In this paper, we explore the problem of network security situation awareness for electric power information systems under big data environment. In order to monitor network security problems, a network security situation awareness technology based on multi-source logging methods by utilizing big data analysis is proposed. We apply this technique to the information system environment of a certain electric power company. We deployed network traffic security analyzer (TSA) in the export of company Internet network. It can acquire and storage the original network traffic in real time. By using the big data visualization analysis tool and rich data display component, the realization of the multidimensional graphical visualization of the analysis results is presented.

1 Introduction

With the rapid expansion of the network scale and its applications, various cyber threats are increasing. The threat of cyber viruses and DDos attacks is growing. The trend of cyber attacks is to be distributed, scale and complicate. Relying solely on firewall, intrusion detection, anti-virus, access control and other simple network security technologies could not meet the requirement of network security [1]. Therefore, we urgently need a new technology, timely find abnormal events in the network, real-time control network security situation, reduce network security risks, improve the capability of network security protection.

Network security situation awareness technology can synthesize all aspects of the safety factor, reflect on the whole dynamic network security situation, and forecast the development trend of network security. The characteristics of large data analysis technology, such as mass storage, parallel computation and efficient query, provide the basis for the research of large-scale network security situation awareness [2]. We can use the big data analysis technique to analyze the information of many network logs. The security situation of the network is analyzed and evaluated. It can realize the awareness of the abnormal events in the network and the overall security situation.

In this paper, we explore the problem of network security situation awareness for electric power information systems under big data environment. We have gathered, stored, analyzed and displayed information on all elements of data sources such as security logs, network traffic, and threat intelligence. It provides comprehensive analysis capability from attack warning, identification and analysis.

2 The Related Concepts of Network Security Situation

2.1 Cyberspace Situation Awareness

The situation awareness is the acquisition of environmental factors within a given time and space, understanding and predicting the future in the short term. Situation awareness includes situation factor acquisition, situation understanding and situation prediction [3]. Network situation refers to the current state and trend of the whole network, which is composed of all kinds of network equipment, network behavior and user behavior.

The cyberspace situation awareness (CSA) was first proposed by Tim Bass in 1999. The cyberspace situation awareness refers to obtain, understand, display and forecast of the latest development trend for security elements that can cause changes in the network dynamics in large-scale network environment.

2.2 Network Security Situation Awareness

Network security situation awareness is the use of data fusion, data mining, intelligent analysis and visualization to visualize the real-time security of the network environment. And it can provide safeguard for network security [4]. With the aid of network security situation awareness, regulators can timely understand of network state, attack, attack source and what services are susceptible to attack, and so on. Therefore, they can take action against the network that launched the attack. Network users can clearly grasp the network security status and trend, preparing to corresponding prevention, to avoid and reduce the network viruses and malicious attacks. The emergency response organization can also understand the security situation and development trend of the network from the network security situation, and provide the basis for the development of the foresight plan.

The main tasks of network security situation awareness include both risk perception and event perception [5]. Risk perceptions include network asset awareness and network vulnerability awareness. Network asset awareness refers to the automatic, rapid discovery and collection of large-scale network assets distribution, updates, properties and other information. Network vulnerability awareness is the analysis of the vulnerability of the network and the identity and management of vulnerability. Network vulnerability includes invisible vulnerability and visible vulnerability. Event perception mainly includes security event awareness and abnormal behavioral perception. Security event awareness refers to the time, place, cause, result, and outcome that can determine the occurrence of a security event. Abnormal behavior perception refers to use the abnormal behavior to determine risk perception, to make up for the vulnerability of invisible, unknown security events found insufficient, and it is mainly oriented the perception of the unknown attacks.

3 the Deployment Architecture

In this section, we give an overview of the deployment architecture based on big data technology. This architecture is depicted in Fig. 1.

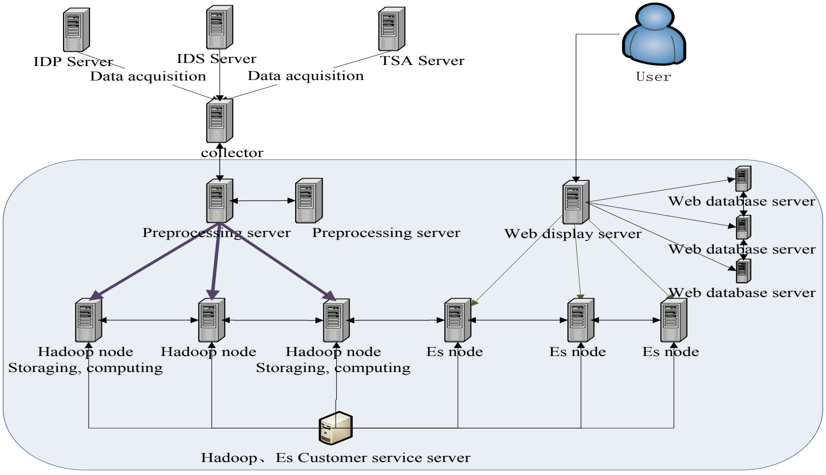

Fig. 1. the deployment architecture based on big data technology

The deployment architecture of situation awareness platform based on big data analysis technology includes front-end servers, storage servers, database servers, and Web servers. The front-end server is primarily divided into TSA servers, IDS servers, firewalls, and other servers. Each type of front-end server provides data sources for the situation awareness platform based on big data technology and provides data analysis and retrieval for the system.

The collector server is responsible for collecting data from front-end servers such as TSA, IDS, APT, and IPS. The collection server is also responsible for filtering, caching, and simply formalizing the data, and so on.

The preprocessor server is responsible for summarizing all data collected by the collector server. And it is in charge of data collection and data cleansing, and storing data to different storage systems based on different businesses.

The pre-processed data is stored on the hadoop server. The platform leverages hadoop's storage and analysis capabilities to correlate statistics and data mining. The resulting data is generated and imported into the retrieval engine for the web server to query the data.

Es nodes (Elasticsearch) server is stored the results that it formed by hadoop servers. The Es node server also makes simple secondary statistics and provides an interface for retrieving data to the web server.

The client server provides automated operation and maintenance and monitoring services for the entire network security situation awareness platform. The operations personnel are able to configure and manage the task scheduling and monitoring of the system platform through the interface provided by the client server.

The Web server is divided into Web database server and Web display server. The Web database server is primarily store a business function data for the network security situation awareness platform. The Web display server is a graphical representation of the data.

4 The Related Technologies of Network Security Situation Awareness

In the case of large-scale networks, there are many heterogeneous network environments and application platforms, which have a large number of network nodes, complex branches and data traffic. On the other hand, the network technology and attack methods showed a trend of the development of the platform, integration and automation, network attack has better concealment and longer incubation time. As a result, the threat and damage of cyber attacks is continuously increasing [6]. In order to display the entire network security situation in real time and detect the potential malicious attacks accurately, which that network security situation awareness is on the basis of the elements of network resources in the collection, through the data preprocessing, feature extraction, situation assessment, situation prediction and situation display to complete. It involves many related technical issues, mainly includes the data fusion technology, data mining technology, feature extraction technology, the trend prediction technology and visualization technology, etc [7].

4.1 The Data Fusion Technology

The data on the situation awareness of cyberspace comes from a multitude of network devices. These data formats, data content, and data quality vary widely. There are different forms of storage and the semantics of expression vary. We can preprocess the data from different network locations and different formats, and then integrate them on this basis. It can provide a more comprehensive and accurate data source for network security situation awareness, thus obtaining more accurate network situation. Data fusion technology is a multi-level, multi-faceted data processing process. It mainly completes complementary integration of multiple sources of information from the network with similar or different feature patterns. It mainly deals with the automatic monitoring, correlation, estimation and combination of the data, thus obtaining more accurate and reliable conclusions [8]. Data fusion can be divided into three levels from low to high, including data level fusion, feature level fusion and decision level fusion. And feature level fusion and decision level fusion are widely used in situation awareness.

4.2 The Data Mining Technology

The network security situation awareness takes the data of a large number of network security devices that are collected and processed into a unified data unit. These data units are large and carry a lot of information, which is difficult to identify with useful information and useless information. In order to grasp the relative accuracy and real-time network security situation, we must eliminate the jamming information. Data mining is to excavate useful information from a large amount of data. That is to say, it is to discover regular and potentially useful knowledge from a large, incomplete, noisy, fuzzy data [9]. Data mining can be divided into descriptive mining and predictive mining. Descriptive mining is used to characterize the general characteristics of data in a database. Predictive mining is extrapolated and predicted on current data. The data mining methods mainly include correlation analysis, sequential pattern analysis, classification analysis and cluster analysis. Correlation analysis is used to explore the connection between data. Sequential pattern analysis focuses on the analysis of the causal relationship between data. Classification analysis is used to establish an analysis model for pre-defined classes. The common models include decision tree model, bayesian classification model and neural network model, etc. Cluster analysis does not rely on predefined classes, and its division is unknown. The common methods include fuzzy clustering, dynamic clustering, and density based methods, etc.

4.3 The Feature Extraction Technology

The feature extraction technology is to use a series of mathematical methods for the large-scale network security information fusion merge into a group or several groups within a certain range of values. These values have a series of characteristics of the real-time performance of the network to reflect the situation of network security and the degree of threat. The network security situation feature extraction is the foundation of network security situation assessment and prediction. There are important implications for the overall situation assessment and forecast. The method of feature extraction mainly includes analytic hierarchy process, fuzzy level analysis, Delphi method and comprehensive analysis [10].

4.4 The Situation Prediction Technology

Network security situation prediction is based on the network running status development and changes of the actual data and historical data, using the scientific theory, method, and a variety of experience, judgment, knowledge to speculate, estimate and analyze its possible changes in certain period of time in the future. It is an important part of network security situation awareness. The security situation of the network at different times is related to each other. The change of security situation has certain internal rule, which can predict the security situation of the network in the future. Thus, the security strategy can be carried out in a predictable way, and the dynamic network security management is implemented to prevent the occurrence of large-scale cyber security incidents. The network security situation prediction method mainly includes the neural network prediction method, the time series prediction method and the grey based prediction method.

4.5 The Visualization Technology

The network security situation generation is based on the analysis of a large number of data to show the current state and future trends. It is difficult to find useful, critical information through traditional text or simple graphical representations. The visualization technology is the use of computer graphics and image processing technology to convert the data to graphics or image, and display it on the screen [11]. It is involved in computer graphics, image processing, computer vision, computer aided design, etc. There have been a number of studies that have applied visualization techniques and visualization tools to situation awareness. Each stage of network security situation awareness takes full advantage of the visualization method to consolidate network security situation into a coherent network security situation diagram. In this way, we can quickly find out the threat of cyber security and grasp the situation of network security.

5 Network Security Situation Awareness Technology based on Multi-source Logs

With the expanding of network size and the increasing of network attack complexity, intrusion detection, firewall, anti-virus, security audit, and many other security equipment widely used in the network. While these safety devices have played a role in cyber security, there are significant limitations. Firstly, there are lots of alarms and logs of safety devices is low level of semantics, high redundancy and large storage space. And there are a lot of false positives. Secondly, most of the safety equipment is single, and the alarm message format is different. It is difficult to comprehensively analyze and organize without the realization of information sharing and data interaction, so that the overall protective effectiveness of the security equipment cannot be fully utilized. Thirdly, the result of the security equipment can only show the operation of the network in a single way, and it is difficult to provide comprehensive and intuitive information about the overall security situation and trend of the network. In order to effectively overcome the limitations of these network security management, we propose a network security situation awareness technology based on multi-source logs.

5.1 Network Security Situation Awareness Factor Acquisition Based on Multi-source Logs

Network security situation awareness technology based on multi-source logs is to extract, analyze and process a variety of security logs in network, and to achieve real-time monitoring of the network situation. It can also identify and alert potential and malicious cyber attacks. In this way, we will make full use of the overall effectiveness of all safety equipment and improve the ability of network security management.

The network security situation awareness technology based on multi-source logs mainly collects security logs at the entrance of the network. It includes firewall logs, intrusion detection logs, critical host logs in the network, and host vulnerability information. By analyzing the log information from different devices, we can fully and thoroughly explore the relevant information about the real and effective network security situation. Compared with the network security situation that is based on single log source analysis, it can improve the overall and accuracy of network security situation.

5.2 Multi-source Log Analysis with Big Data Technology

The network security situation awareness technology based on multi-source logs captures a large amount of security devices data with a variety of detection mode and event reporting mechanisms. And these original logs information exists defects such as mass, redundancy and errors, not as a direct source of situation awareness, correlation analysis and data fusion processing must be conducted. The emergence of big data extends computing and storage resources. The big data has three characteristics: Variety, Volume, and Velocity, which happens to be the need for network security situation awareness analysis based on multi-source logs. The variety of big data can make the network security situation awareness for more types of log data, including safety log, network operation, business information and application log, etc. The volume of big data is required for large volumes of log storage and processing. The velocity of big data provides technical support for deep security analysis of high-speed network traffic, and provides computing resources for high-intelligent model algorithms. Therefore, we take advantage of the basic platform and the technical support provided by big data to handle the analysis of network security situation. In this paper, we mainly explore the following three analytical models.

Correlation analysis. Security log on the network is a depiction of traffic to the security event that enters the network. A large number of logs and associated alarm records are generated for a possible attack event. These records have a lot of redundancy and relevance. Therefore, we need to convert the original log after monophyletic correlation analysis to intuitive or may cause harm to the network security events. Network security situation awareness based on multi-source log use the similarity of alarm correlation, and it can better control the alarm number after associated. And it is helpful to reduce the complexity. The process is described as follows. Firstly, we extract the main properties in the alarm log and form the original alarm. Secondly, the polymerization alarm is generated by repeated alarm polymerization. The method of calculating the similarity of each attribute of the polymerization alarm and the weight is assigned. The similarity of the two polymerization alarms is calculated by comparing the value of the similar degree to determine whether the polymerization alarm is carried out hyper-alert. Finally, we output an address range and alarm information that belongs to the same class of alarm and generate security events.

Fusion analysis. Multi-source log is characterized by redundancy, complementarity, etc. The Situation awareness with the aid of data fusion technology, it can make between multiple data sources complement each other, so as to provide protection for cognitive process. And it can generate security situation more accurately. After the single source log alarm correlation process, we can get separate security events respectively. In the case of multiple source security events from firewalls and intrusion detection logs, the d-s evidence method is used to identify the multiple source safety events. We can assess the credibility of the security event, further improve the accuracy and reduce false positives. The basic idea of the d-s evidence theory applied to the fusion of security events is depicted as follows. First, a practical method of initial trust allocation is studied, and the information degree function of firewall and intrusion detection is allocated. Then, through the synthetic rule of d-s, the reliability of the security event after the fusion is obtained.

Situational factor analysis. By mean of security analysis of security device logs at the portal, it just gets access to the target network's possible attack information. A security event that is truly critical to network security needs to be finalized by a comprehensive analysis of the knowledge base and the specific network environment. There are three main steps. Firstly, by studying a large number of network attack instances, the available attack knowledge base is obtained. It mainly covers the principles and characteristics of various cyber attacks. Secondly, by analyzing the potential vulnerabilities of system vulnerabilities and hosted services vulnerabilities, we can establish a vulnerability knowledge base for the current network environment. The network environment knowledge base is obtained by analyzing the topology and performance indexes of the current network environment. Thirdly, the vulnerability knowledge base is used to identify the effectiveness of security events, which is a network attack event that affects the current network. In the process of verification of network security events and attacks, we can extract the situation factor that are used to assess the overall security situation of the network. They include security threats to the entire network, security threats to branch networks, security threats to hosts, and the extent of these threats.

6 Experimental Evaluation

At present, the certain electric power company has deployed a network security situation awareness platform based on big data analysis technology. The platform has been connected to the Internet traffic and 12 important asset servers, 2 IPS, 2 firewalls, and the information network APT system. There are 7 high performance hardware servers, including 3 for ES, 3 for hadoop and 1 for display server. So far, the platform has been acquired threat data of 6TB, the number of threatened data of 1,180 megatons. And the platform has added 230,000 Internet threat intelligences. We have monitored the front-end device network traffic by using the platform as shown in figure 2 and the important assets under attack as shown in figure 3.

Let's take the data has monitored by the platform in November 2016. The platform has found 674,311 external threats and 1870 important asset threat events. In this case, the trigger Web attack class is 2650 times, DDoS attacks 2300 times, browser scan class 275,500 times. From the hit ratio of threat intelligence, threat IP accounts for 53%, threat domain names 32% and threat URL 12%. From the attack mode, active attacks accounts for 63% and passive attacks 37%. From the source of the attack, China accounts for 66.84%, while the U.S. accounts for 33.3%. The security situation results are shown in figure 4 and figure 5. The month-on-month threat in October and November 2016 is shown in figure 6. Figure 7 is shown the number of attack threats varied Top5. From the results of situation awareness analysis, we can clearly catch sight of the company information system facing security threats, and formulate corresponding reinforcement measures in time. In this way, we can guarantee the safe and stable operation of information system of the company.

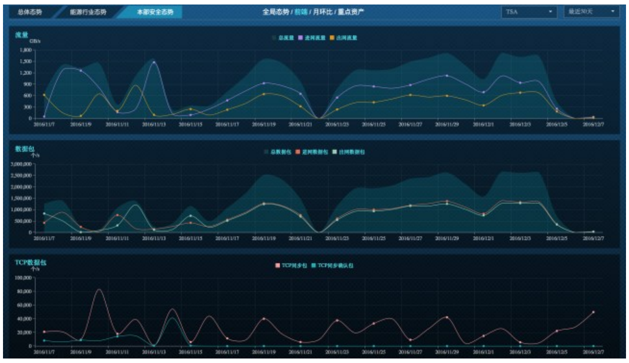

Fig.2. Monitor the network traffic of the front-end device

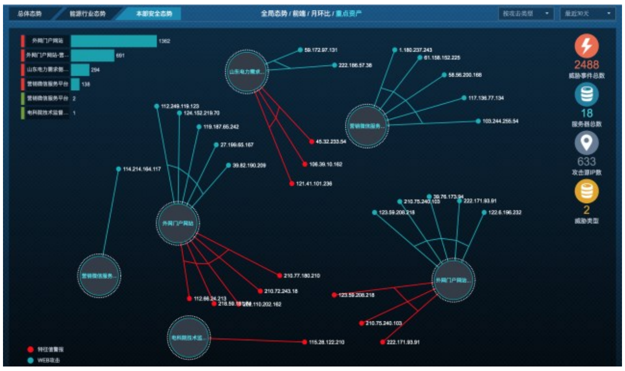

Fig.3. Monitor important assets under attack

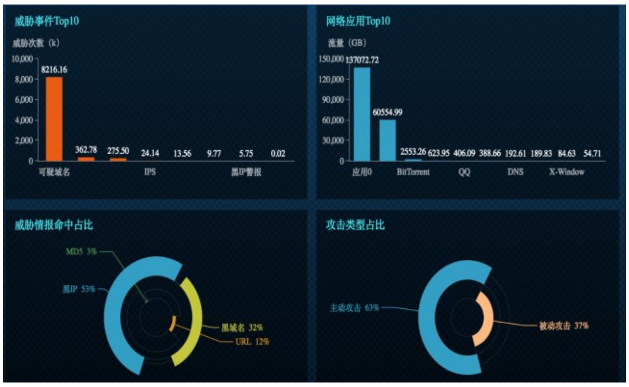

Fig.4. The proportion of threat events and attack mode

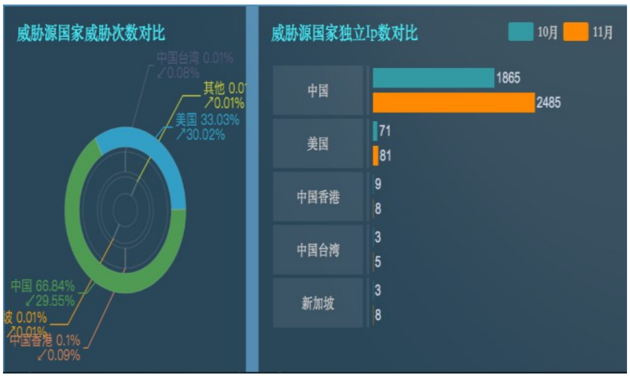

Fig.5. The proportion of threat source countries

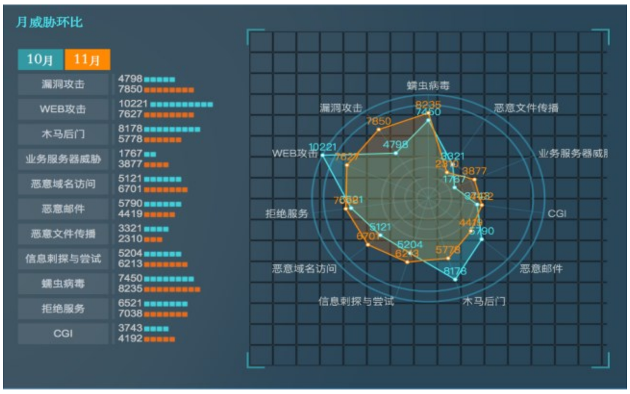

Fig.6. The month-on-month threat in October and November 2016

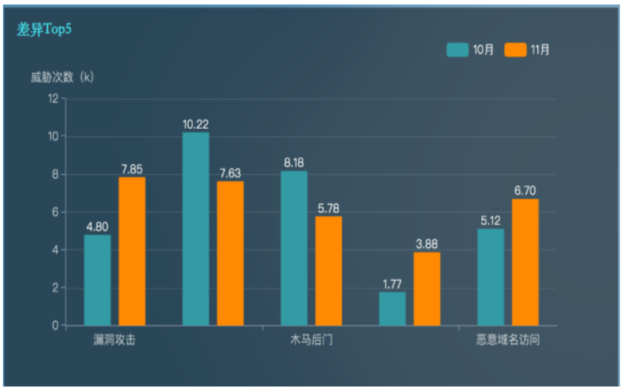

Fig.7. The number of attack threats varied Top5

7 Conclusion

In order to solve the increasing threat of cyber security, we apply situation awareness techniques to network security. In this way, we can not only fully grasp the current network security situation, but also predict the future trend of cyber security. In this paper, we have introduced the concept and technology of cyber security situation. Then, we explore the network security situation awareness technology based on multi-source logs. We focus on studying the network security situation awareness factor acquisition based on multi-source logs, and the use of big data correlation analysis, and fusion analysis and situational factor analysis. We apply this technique to the information system environment of a certain electric power company. We are able to collect and store the original network traffic in real time by deploying the network traffic security analyzer TSA in the export of the company's external network. We use the big data analysis technique to analyze the security threats in real time and backtrack the network attack process. By using big data display components, the multidimensional graphical visualization of the analysis results is intuitively presented. By building a network security situation awareness platform based on big data technology in electric power information systems, we can monitor the security threats of information systems in real time. We are able to formulate the appropriate reinforcement measures for security threats in time. In this way, we can guarantee the safe and stable operation of information system of the company.

Acknowledgments

Most of this work was done during my working period at State Grid Shandong Electric Power Research Institute. The authors wish to acknowledge all colleagues and the anonymous reviewers for their helpful comments.

References

[1] Cao Rongrong. Research of Network Security Situation Awareness under Big Data Environment. Digital Library Forum[J]. 2014, 117(02), pp.11-14.

[2] Vincent Lenders, Axel Tanner, Albert Blarer. Gaining an edge in cyberspace with advanced situational awareness. Security & Privacy IEEE[J], 2015,13(2), pp. 65-74.

[3] Gong Jian, Zang Xiaodong, Su Qi, Hu Xiaoyan, Xu Jie. Survey of Network Security Situation Awareness. Journal of Software[J], 2017, 28(4), pp.1010-1026.

[4] Wei Yong, Lian Yifeng, Feng Dengguo. A Network Security Situational Awareness Model Based on Information Fusion. Journal of Computer Research and Development[J]. 2009, 46(3), pp.353-362.

[5] Liu Peng, Meng Yan, Wu Yanyan. Large-Scale Network Security Situation Awareness and Forecast. Computer Security[J]. 2013, 12(03) pp.28-35.

[6] Xi Rongrong, Yun Xiaochun, Zhang Yongzheng. An improved quantitative evaluation method for network security. Chinese Journal of Computers[J]. 2015,38(4), pp.749-758.

[7] Xin Dan, Gai Weilin, Wang Lu, Liu Xin, Hu Jianbin. Survey of cyberspace situation awareness model. Journal of Computer Applications[J]. 2013, 33(S2), pp.245-250.

[8] Liu Xiaowu, Wang Huiqiang, Lü Hongwu, Yu Jiguo, Zhang shuwen. Fusion-Based cognitive awareness-control model for network security situation. Journal of Software[J], 2016, 27(8), pp.2099-2114.

[9] Liu Zhijun, Pu Xiaowei. The Analysis of Application of Data Mining Technology in the System of Intrusion Detection. Humanities, Social Sciences and Global Business Management[M]. ISSGBM 2014, pp.75-78.

[10] Zhang Mingjun, Li Shupeng, Li Xuan. Research on technologies of underwater feature extraction and target location based on binocular vision. 2015 27th Chinese Control and Decision Conference (CCDC)[M], 2015, pp. 5798-5804.

[11] Hadi Shiravi, Ali Shiravi, Ali A. Ghorbani. A survey of visualization systems for network security. IEEE Transactions on Visualization and Computer Graphics[J], 2012,18(8), pp.1313-1329.