Application and Implementation of Big Data Visualization Technology in Network Security System

Chao Tang, Jie Zhang*, Sheng Wang, Linghao Zhang, Hai Wang and Huihui Liang

Electric Power Research Institute of State Grid Sichuan Electric Power Company
ChenDu, Sichuan, 610000, China

*Corresponding author’s e-mail: Zhangj6459@sc.sgcc.com.cn

Abstract. In order to improve the safety and reliability of the information system of the State Grid, platforms that apply technologies such as big data and artificial intelligence to realize the safety monitoring and management of information systems have been widely used. At the same time, the application of visualization technology can fully display the current security status and situation of the network, as well as the intrusion and attack process of various attacks and threats. This paper proposes to apply big data visualization technology to realize asset visualization, behavior visualization, and attack visualization, and to show the risk status, intrusion view and process of the entire network.

1. Introduction
Since the 21st century, under the background of steady social and economic development, network security has shown a rapid development trend. At the same time, with the rise of network security, network security systems have attracted much attention. Taking network security log data as an example, it shows an explosive growth trend, but the current visualization technology is difficult to fully visualize and analyze high-latitude, multi-granular data. With the advent of the era of big data, data visualization has become a hot topic and has attracted great attention. Therefore, according to the actual situation, the big data visualization design is applied to the network security system to ensure the security of network information.

Network security data is mainly composed of two parts, including data that threatens network data security and non-monitored network data collected from monitoring equipment at different physical locations and logical levels. The technical personnel responsible for network data security mainly discover and deal with problems based on these monitoring equipment and non-monitored network data to ensure the security of network data. Network security data visualization is mainly for the actual needs of whether workers can directly and efficiently analyze whether the network data is safe or not based on the data obtained from monitoring. Network security data visualization advocates the use of advanced technology and computer equipment to help staff responsible for network data security find problems and harmful information in network data, and help staff improve their ability to identify network data security, thereby ensuring network data security. The purpose, to put it simply, is to use advanced technology to break through the human body's own limitations in the face of many potential information.

Big data visualization refers to the use of images and charts to show specific data information. With the advent of the big data era, we are faced with a lot of data every day, and information
processing has become more onerous. The visualization of data makes complex data information simple and easy to understand, and people can also reduce time consumption as much as possible when working on information analysis and data processing. Moreover, with the continuous improvement of computer technology, the computer's ability to visualize data and information has also been further improved. Therefore, the popularization of big data visualization processing technology can bring higher protection to network security systems. At the same time, it will also make it more convenient for the staff to process information.

2. Related research
For example, Nunnally T [6] extended the traditional parallel coordinate axis, and proposed the 3D Parallel Coordinate method to visualize network security data, and achieved good results. However, when the amount of data is too large, the method is between the lines. There will be overlaps, and it is difficult to find useful information. Flow-In-Spector [7] uses multi-graph integrated visualization technology to display network flow data in real time, and realizes the rendering of histograms, force guide graphs, and radiation graphs. In order to enhance the interactivity of the analysis system, a visualization model of a three-dimensional multi-layer spherical space was designed by Wu Yadong [8]. Chen Peng [9] used information entropy-based traffic anomaly data mining algorithm to improve the success rate of traffic anomaly detection, and realized A three-dimensional visualized flow monitoring system was developed. Zhang Sheng [10] used a combination of tree diagram and time series diagram to achieve a visualization system for Netflow logs. However, when the amount of data is large, the two-dimensional time series diagram will cause graphic coverage, and the tree diagram will take up a lot. The screen space is very unfriendly to the display of large amounts and multiple dimensions of data.

In the direction of visual analysis, multi-view correlation analysis and interactive query are essential to help network security managers observe the current network status from multiple dimensions. For example, Zhao Ying use time-series parallel coordinate views and multi-agent views. Matrix view, multi-agent timing view, and similarity extended tree view analyze network traffic log data. ENTVis uses radar chart, matrix chart and other visualization methods to support network attack traffic characteristics analysis based on information entropy, Shi use improve the radar chart for network event correlation analysis. Although the above techniques reduce data items and data dimensions through feature extraction, dimensionality reduction, sampling or aggregation methods, the interaction is not good.

In summary, the research on the visualization analysis technology of Netflow logs has made great breakthroughs, but there is still a high room for improvement in multi-view integrated interaction and multi-dimensional data visualization. Therefore, this article focuses on the use of three-dimensional histogram instead of Two-dimensional coordinate graph display data, it can provide multiple interactive methods such as rotation, zooming, filtering, etc. to solve the problem of dense data concealment, and the introduction of information entropy algorithm for parallel coordinate axes to achieve unified processing of multi-dimensional features, which can better display more Dimensional data. In the power grid system analysis, the visual analysis system of this article is used to realize the visual analysis of assets, attacks, and behaviors.

3. Visual analysis system

3.1. Introduction to Big Data Visualization Technology
(1) Basic concepts of data visualization
The data space is a multi-dimensional information space composed of a data set composed of n-dimensional attributes and m elements. Data development refers to the use of certain tools and algorithms to quantitatively deduct and calculate data. Data analysis is to analyze the data by slicing, block, and rotating the multi-dimensional data, so that the data can be observed from multiple angles and sides. Data visualization is to represent the data in a large data set through graphics and images, and use data analysis and development tools to discover unknown information.
(2) Standards for data visualization
In order to realize the effective communication of information, data visualization should take into account both aesthetics and function, and intuitively convey key features to facilitate the mining of the hidden value behind the data. Visualization technology application standards should include the following 4 aspects:
1) Visualization
Present the data intuitively and vividly.
2) Association
Highlight the correlation between the data.
3) Artistic
Make the presentation of data more artistic and more in line with aesthetic rules.
4) Interactivity
Realize the interaction between the user and the data to facilitate the user to control the data.

3.2. Visual analysis of assets
Asset visualization refers to the automatic discovery of network assets, identification of high-value assets, and visual display of asset status, type, location, and user information. The following figure shows the visualization of asset information scanned by the system in combination with big data visualization technology.

![Asset visualization diagram](image)

3.3. Visual analysis of attacks
From the perspective of previous designs, most of the traffic display methods adopt two-dimensional line graphs and histograms, and it is difficult to analyze the current traffic distribution on a multi-level basis. In the course of this research, the three-dimensional histogram is based on time series to count the number of bytes under different protocols over a period of time. For example, by observing the height and color depth of the histogram, the user can learn more about the abnormal traffic conditions based on the TCP protocol in the darker color and the relatively high time period of the histogram. If there is a small amount of traffic under the UDP protocol and other protocol conditions during a certain period of time, the user can effectively filter the target data by clicking the mouse, or view the data information of the blocked area in time through functions such as translation, rotation, and zoom. The port number on the network represents a network service provided by the computer. By scanning a certain port or designated port on the computer, the attacker can detect the type of network service provided by the host, and use the known vulnerabilities of these services to attack. You can start to prepare attack methods. In order to find more vulnerability, some attackers will scan more than 60,000
ports at a time. This behaviour will cause a significant increase in the number of destination ports and an increase in destination port entropy.

For network security analysts, it is possible to observe the size of the traffic in the three-dimensional histogram and locate abnormal data accurately and quickly. Considering that the staff can effectively judge the type of abnormal behaviour during a certain period of time, the multi-dimensional characteristics of Netflow can be observed to ensure that the decision-making is more accurate and efficient. Generally speaking, information entropy can effectively measure the random characteristics of network activities and reflect the uncertainty and disorder of data distribution. If the data distribution is more regular, the entropy value will be lower; if the data distribution is more chaotic, the entropy value will be greater. The following figure shows the use of big data technology to visualize cyber-attacks.

![Attack visualization diagram](image)

**Figure 2. Attack visualization diagram**

### 3.4. Behavioral visualization analysis

The rapid development of my country's economy has accelerated the pace of computer network technology development. At present, it has entered the red-hot stage of network technology development. As some loopholes in the network are gradually exposed, it is particularly easy to cause network security problems. Among the most important security issues are four major aspects: the expansion of network technology scale, the stability of data information, the transmission of network data and the development of related services. These all have certain obstacles to network security management. The stability of network information and data means that it can be kept confidential and not leaked during network data transmission. There are many insecure factors from the outside world, such as hacker intrusion and virus infection. These dangerous factors from the outside threaten the privacy of users, and these problems affect the security of our country's network data. Therefore, this article uses visualization technology based on big data to visually analyse user behaviour, automatically identify equipment and user behaviour, and visually display access relationships and usage habits. As shown below.
4. Conclusions
Visualization technology based on big data is an important development direction and research hotspot for future network security monitoring. The visualization system based on big data constructed in this paper adopts the form of multi-graph linkage, optimizes the visualization of high-dimensional data, and provides users with a wealth of possibilities. Interactive operation is helpful for users to explore the association between data and improve user experience. Improve the work efficiency of the network security staff, and also improve the guarantee for the safe operation of the network system.

Acknowledgments
This work was supported by the State Grid SiChuan Electric Power Company of China (Project Name: Research and platform development of full flow Advanced Sustainable threat intelligent analysis technology based on deep learning).

References
[1] Chen Chongtao, Wang Weiping, Meng Dan, et al. Research on storage and query processing methods for massive NetFlow data [J]. High Technology Communications, 2016, 26(6): 534-541.
[2] Wang Wenjie. Discussion on network information security and protection in the era of big data [J]. China Management Information, 2019(12): 125-126.
[3] Guo Jun. A Preliminary Study on Network Information Security Assurance Strategies in the Context of Big Data [J]. Science and Technology Communication, 2019(11): 89-90.
[4] Pang Shulan, Wang Xiaohui. Application of Big Data Technology in Network Security Analysis. Technology Wind, 2019(16): 91.
[5] Deng Shouxun, Zhang Chengtian. Research on Intranet Security Detection Based on NetFlow[J]. Computer Applications and Software, 2018,35(10):329-333.
[6] Nunnally T, Chi P, Abdullah K, et al. P3D: A parallel 3D coordinate visualization for advanced network scans.Proceedings of 2013 IEEE International Conference on Communications (ICC).Budapest, Hungary. 2013. 2052–2057.
[7] Braun L, Volke M, Schlamp J, et al. Flow-inspector: A framework for visualizing network flow data using current web technologies. Computing, 2014, 96(1): 15–26. [doi:10.1007/s00607-013-0286-4]
[8] Wu Yadong, Jiang Hongyu, Zhao Silui, et al. 3D visualization method of network security data. Journal of University of Electronic Science and Technology of China, 2015, 44(4): 594–598, 604. [doi:10.3969/j.issn.1001-0548.2015.04.020 ]
[9] Chen Peng, Si Jian, Yu Zihuan, et al. Network flow anomaly monitoring and three-dimensional visualization method based on information entropy. Computer Engineering and Applications, 2015, 51(12): 88–93. [doi:10.3778/j.issn.1002-8331.1408-0111]

[10] Zhang Sheng, Shi Ronghua, Zhao Ying. Visual fusion analysis method based on multiple heterogeneous network security data. Computer Applications, 2015, 35(5): 1379–1384, 1416. [doi: 10.3969/j.issn.1001-3695.2015.05.025]

[11] Zhao Ying, Wang Quan, Huang Ye Zi, et al. Visual analysis of network traffic time series data based on multi-view cooperation. Journal of Software, 2016, 27(5): 1188–1198.

[12] Zhou FF, Huang W, Zhao Y, et al. ENTVis: A visual analytic tool for entropy-based network traffic anomaly detection. IEEE Computer Graphics and Applications, 2015, 35(6):42–50. [doi: 10.1109/MCG.2015.97]

[13] Shi L, Wang C, Wen Z, et al. 1.5D Egocentric dynamic network visualization. IEEE Transactions on Visualization and Computer Graphics, 2015, 21(5): 624–637. [doi: 10.1109/TVCG.2014.2383380]

[14] Wang Xun. Design and implementation of anomaly detection system for LAN traffic based on Netflow [J]. Information and Computer (Theoretical Edition), 2016(21):186-188.