Discussion on Network Security Situational Awareness Technology Based on Big Data

Zhou Siwei

Changsha Medical College, Changsha, China
Email: 2224121008@qq.com (Z.S.W.)
Manuscript received December 20, 2024; accepted March 27, 2025; published April 25, 2025.

***Abstract*—With the rapid development of information technology, network security issues are becoming more and more prominent and have become the focus of global attention. The purpose of this paper is to discuss the design of a network security situational awareness system based on big data technology, which is able to monitor and analyze massive network data in real time in order to predict and identify potential security threats. This paper first introduces the overall framework design of the system, including the data aggregation and storage layer, the big data analysis layer, and the situational awareness and early warning business layer. Subsequently, the functional design of each layer is elaborated in detail, and it is hoped that this paper can provide some reference for the application of big data technology in the field of network security.**

***Keywords*—big data, cybersecurity posture, Hadoop, distributed storage**

I. Introduction

In the digital era, cyberspace has become an important infrastructure that people can't live without, and the importance of cybersecurity situational awareness as a key technology to maintain the security of cyberspace is self-evident. The network security situational awareness system can identify and predict potential security threats through real-time monitoring and analysis of massive data in the network, providing decision support for network security management [1]. With the development of big data technology, the traditional network security situational awareness system faces the problems of large data volume, slow processing speed, limited analyzing ability, etc., and urgently needs a new technical means to improve its performance. Based on big data technology, this paper discusses the design of a new type of network security situational awareness system, aiming at realizing rapid response and effective early warning of network security threats through efficient data processing and analyzing capabilities, as follows [2].

Overall Framework Design of Network Security Situational Awareness System

The system is constructed by Hbase+Hadoop, which provides storage and calculation processing capability for massive data. The system supports Mysql database and Oracle database. The core of this platform lies in its layered architecture and the big data analysis algorithms and models it uses. Its structure includes data collection and storage layer, big data analysis layer, and situational awareness and early warning business layer, each layer carries specific functions and tasks, which together constitute a complete network security situational awareness system, as shown in the figure below [3].

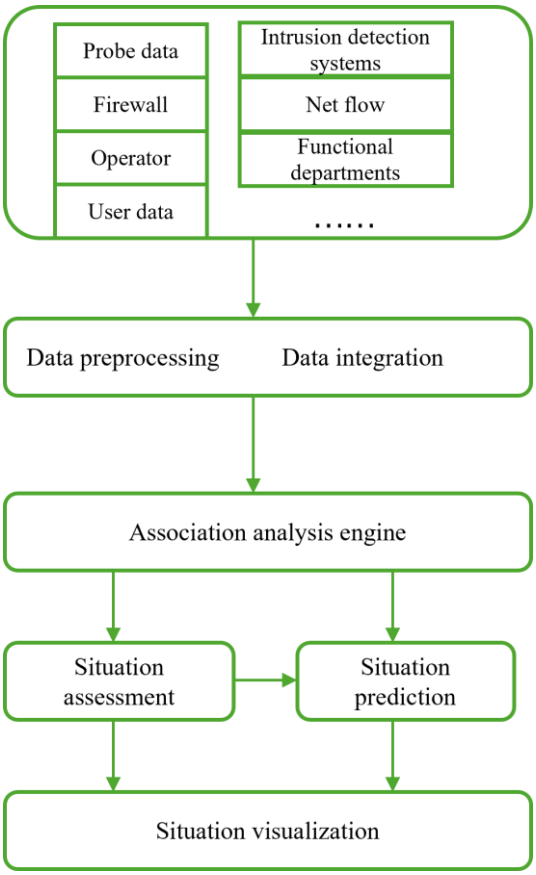

Fig. 1. Schematic diagram of the structure of the big data-based network security situational awareness platform.

A. Data Aggregation and Storage Layer

As the cornerstone of the system, the key to the design of the data collection and storage layer lies in the ability to efficiently collect and store diversified data from different network sources. The distributed storage platform constructed by using Hadoop can effectively deal with massive data and realize real-time reading and writing of large-volume data by means of distributed arrays at the same time, and the platform also supports two databases, Mysql and Oracle, so that it can adapt to a variety of data sources and storage requirements and ensure the comprehensiveness of security protection [4].

B. Big Data Analytics Layer

The big data analysis layer, on the other hand, is the core of the system, which is responsible for in-depth analysis and processing of the data in the storage layer. The data preprocessing module first cleans and transforms the raw data to improve data quality and lay a solid foundation for

subsequent analysis. The security posture assessment module based on Sim Hash algorithm quickly identifies abnormal patterns and potential threats in the network through similarity detection. The attack tree inference model, on the other hand, predicts possible attacks by constructing attack paths [2] and provides guidance for the deployment of defense measures. The online real-time mining analysis and offline mining analysis modules are responsible for real-time monitoring of network traffic and in-depth mining of historical data, respectively, to discover potential security threats and trends [5].

C. Situational Awareness and Early Warning Business Layer

The Situational Awareness and Warning business layer is then the output layer of the system, which translates the results of the big data analysis layer into specific business applications. The Network Security Threat Alert module is able to issue timely warnings to network administrators, pointing out specific security threats and possible impacts. The real-time network attack monitoring module provides a platform for real-time monitoring of network status, enabling administrators to quickly respond to various network attacks [6]. The Network Risk Awareness and Early Warning module, on the other hand, analyzes historical data to predict possible future security events, providing forward-looking guidance for network security management. Finally, the posture display module displays complex data analysis results to users in an intuitive way through visualization technology, making the network security posture clear at a glance [7].

II. Functional Design for Each Level of the System

A. Data Aggregation and Storage Layer

The core of this tier is a distributed storage system based on Hadoop, which provides solid technical support for processing and storing massive amounts of cybersecurity data with its excellent scalability and high reliability. Hadoop, through its HDFS component, realizes the distributed storage of data, allowing the system to store a large amount of data across multiple nodes, while guaranteeing the high availability and fault tolerance of the data. Due to the huge volume of cybersecurity data, the system must be able to process the data stream efficiently to ensure real-time data availability and integrity [8]. In addition, data security is also a key concern at this level; therefore, encryption and access control mechanisms during data transmission and storage are strictly implemented to prevent data leakage or unauthorized access. Finally, in order to support the scalability of the system, the data aggregation and storage tier is designed with a modularized storage architecture that allows the system to dynamically adjust the storage resources according to the amount of data and processing requirements [9].

B. Big Data Analytics Layer

1) Data pre-processing

Data preprocessing is the process of cleaning, transforming, and normalizing the raw data collected prior to data analysis, with the goal of removing noise and inconsistencies from the data and improving the quality of the data to make it more suitable for subsequent analysis and mining. Data transformation is the process of converting data into a form suitable for analysis, which may include normalization processes such as converting all text data to lowercase or standardizing date formats, as well as feature engineering, which is the process of extracting information from the raw data that is useful for machine learning models to understand [10]. Data statutes, on the other hand, aim to reduce the size of the data while preserving the useful information it contains, which is particularly important for processing large-scale cybersecurity datasets as it reduces storage requirements and increases processing speed. When implementing data preprocessing, it is important to consider data privacy and security issues to ensure that sensitive information is not compromised during processing. In addition, the preprocessing steps need to be flexible and scalable to adapt to changing cybersecurity threats and data environments.

2) Security posture assessment based on SimHash algorithm

This system adopts the security posture assessment based on SimHash algorithm, which is a key technology to realize the big data analysis in this system. SimHash algorithm, as a locally sensitive hash algorithm, is centered on converting textual data into fixed-length hash values, and evaluating the similarity of the data content by comparing the similarity between these hash values. In the field of cybersecurity, this algorithm is used to quickly identify and aggregate similar security events, so as to effectively assess the cybersecurity posture. The basic principle of the SimHash algorithm is to represent textual data as vectors in a high-dimensional space, where each dimension corresponds to a feature (e.g., a word or phrase), and each component of the vector denotes whether or not the feature occurs in the text or the number of its occurrences. Specifically, for a given text T, we first extract its feature set F, and then assign each feature a weight w_i in the interval [0, 1], which can be determined based on the frequency of the feature in the text or its inverse document frequency (IDF) in the corpus. Next, we multiply the weight corresponding to each feature by a randomly chosen sign (+1 or −1) on its corresponding dimension and sum the weighted signs of all features to obtain a high-dimensional vector V. Finally, we project this vector to a fixed-length hash H, where the value of each dimension depends on whether the vector has a sign of +1 or −1 on that dimension. its SimHash algorithm can be denoted as [11]:

$$H(T) = \sum_{i=1}^{n} w^i s^i mod n$$

where H(T) is the SimHash value of T, ωi is the weight of the ith feature, si is the symbol (+1 or −1) corresponding to the ith feature, and n is the total number of features.

In the cyber security posture assessment of this system, the similarity of two security events is evaluated by comparing the Hamming distance (i.e., the number of positions with different corresponding bits) between two SimHash values. The shorter the Hamming distance, the higher the similarity between the two events, which can help us to recognize the hazards caused by the same attack method or the same kind

of danger online. This method can not only effectively categorize the large number of similar security events, but also more effectively mine the security risks they face, and then accurately analyze and warn them to achieve more effective security protection [12].

3)Online real-time mining analysis

Facing the massive data in the network environment, it is of great significance to realize the timely discovery and response to the security hidden dangers in the information environment through the in-depth analysis of big data. This system can realize large amount of data collection, online processing and data analysis, and carry out rapid response to safety accidents according to the analysis results. The core of online real-time mining analysis lies in its ability to handle high-speed, large-scale data streams while maintaining the accuracy and timeliness of the analysis results [13]. To achieve this goal, the system adopts stream processing frameworks such as Apache Kafka and Apache Storm, which can support high-throughput data transmission and real-time computation, ensuring that data can be processed and analyzed quickly after being collected. Through these technologies, the system is able to monitor real-time data such as network traffic, user behavior, system logs, etc., and discover abnormal patterns and potential threats in a timely manner. During the analysis process, the system employs a variety of data mining algorithms, including cluster analysis, classification algorithms, and association rule mining, to identify abnormal behavior and potential security threats in the data. For example, through clustering analysis, the system can group similar security events together to discover the behavioral patterns of attackers; classification algorithms are used to match new security events with known attack types to determine their threat levels; and association rule mining is used to discover the correlations between different security events, providing clues to understand complex attack chains. In addition, online real-time mining analysis involves continuous learning and adaptation to real-time data, which requires the system to have the machine learning capability to learn from new data and adapt its analysis model. This adaptability is key to the system's ability to cope with new types of threats and the ever-changing cyber environment.

4)Offline mining analysis

Offline mining analytics, as a key component in the big data analytics layer, plays the role of deep insight and historical trend analysis in big data-based cybersecurity situational awareness systems. As opposed to online real-time mining analytics, offline mining analytics focuses on in-depth post-processing of historical datasets to identify long-term trends, cyclical patterns, and potentially complex attack strategies, which are often difficult to capture in real-time data streams. In offline data mining, massive data and information are analyzed in depth by means of distributed processing, and the system can also utilize techniques such as machine learning model training, deep learning, complex network analysis, and statistical modeling to achieve an in-depth understanding of the evolution law of cybersecurity posture and accurate prediction of its development trend [14]. For example, using machine learning methods, it is possible to extract the common features from the data of past security incidents and provide early warnings for similar events in the future. The main advantage of offline data mining technology is that it can provide more complete and fine-grained protection of the network environment. Through in-depth analysis of this information, the evolutionary pattern of information security can be discovered, new attack methods can be found, and the processing patterns of various methods can be rehearsed, and subtle changes and types of network attacks that are not easy to detect can be identified. In addition, in offline data mining, this system can also realize the synchronous processing of information from different sources. In order to obtain comprehensive analysis results, the system can process information from multiple sources and time points online and offline to discover the intrinsic correlation between different networks, systems and applications, and to analyze the network security posture from a holistic point of view. While conducting offline data mining, it is also necessary to take into account the privacy and security of the data to ensure that there will not be any leakage of confidential information when conducting offline data mining. At the same time, since the analysis process may involve a large amount of data processing and complex calculations, the system needs to have efficient data processing capabilities and powerful computing resources to ensure the accuracy and reliability of the analysis results.

C. Situational Awareness and Early Warning Business Applications

In the network security situational awareness system, the situational awareness and early warning business application is the key link to transform the results of data analysis into practical action, which involves four core parts, namely, network security threat alerts, real-time monitoring of network attacks, network risk perception and early warning, and situational display, which together build a comprehensive network security protection system [15].

1) Network security threat alerts

The Network Security Threat Alert module is the front line of the system's early warning mechanism. By analyzing network traffic and system logs in real time, it uses machine learning algorithms and a library of known threat intelligence to identify potential security threats and issue timely warnings to network security managers. The design of this module is based on an in-depth analysis of historical security events, and by constructing a threat model, it is able to predict and identify new attack patterns, thus sending out alerts at the early stage of an attack and buying valuable time for the deployment of defensive measures.

2) Real-time monitoring of network attacks

The real-time network attack monitoring module is an important part of the system's continuous monitoring of the network environment. This module collects all kinds of event and behavioral data in the network in real time by deploying sensors and monitoring tools, and identifies attacks in real time through data analysis techniques, such as anomaly detection and behavioral analysis. The design of this module emphasizes comprehensive monitoring of network behavior and enables timely detection and response to various attacks, including zero-day and APT attacks, thus effectively reducing the impact of attacks on the network environment.

3) Cyber risk perception and early warning

It realizes the long-term monitoring of network security through the risk situation sensing and early warning of the Internet. And this module can also analyze the historical information and current development trend, can be possible in the future security incidents for early warning, to determine the network security events will likely occur, the system will put forward the appropriate warning. This system through the information security situation analysis and early warning function, to provide security management and protection of computer decision-making, to achieve long-term information security and security protection effectiveness.

4) Posture display

This system can also realize the user-oriented posture analysis results display function, can be analyzed using graphics, tables and other visual means, the results of the analysis of data into intuitive and easy-to-understand visual information presented in front of the operator's screen, to help the system Kusatsuzawa quickly grasp the network security situation and make effective decisions. Visualized information in the design should be intuitive, easy to understand, in order to enhance the practicality of the system.

III. Conclusion

In summary, in today's rapid development of big data technology, its application to the field of network security technology can make it realize a substantial increase in processing capacity, analysis efficiency and response time, etc., to better meet the needs of network security protection in the context of big data. This paper proposes a network security situation system based on big data, supported by Hadoop distributed structure, through data preprocessing, SimHash algorithm and other technologies, to achieve synchronized in-depth analysis of network data from various sources and intelligent early warning, aiming at achieving network security protection with higher efficiency, effectively avoiding all kinds of network anomalous attacks from different sources, and improving the security of the network.

Conflict of Interest

The author declares no conflict of interest.

References

[1] X. J. Bai, "Analysis of network security situational awareness and key technologies based on big data," *Network Security and Informatization*, vol. 11, pp. 122–124, 2024.

[2] Y. Zhu, "Research on network security situational awareness-based on big data background," *Network Security Technology and Application*, vol. 11, pp. 20–22, 2024.

[3] B. Chen, "Research and application of network security situational awareness technology based on artificial intelligence," *Communication World*, vol. 31, no. 10, pp. 31–33, 2024.

[4] S. W. Guo, S. F. Liu, Z. M. Li *et al.*, "A network security situational awareness approach based on fusion model," *Computer Engineering*, vol. 50, no. 11, pp. 1–9, 2024. DOI:10.19678/j.issn.1000-3428.0069758.

[5] M. L. Shao, Y. L. Zong, and Y. L. Zhou, "Research on network security situational awareness technology: Data integration and visualization system design," *Information Technology and Informatization*, vol. 10, pp. 166–169, 2024.

[6] L. Ying, C. Lan, and Z. Qi, "Threat intelligence-based network security posture assessment technology," *China New Communications*, vol. 26 no. 20, pp. 26–28+55, 2024.

[7] Y. Chen, Y. Zhu, Y. L. Liu *et al.*, "Design of network security situation awareness standard architecture," *Journal of Information Security Research*, vol. 9, pp. 2096–1057, 2021.

[8] W. R. Chen and C. W. Zhu, "Research status of network security situation awareness technology," *Science and Technology & Innovation*, vol. 14, pp. 1–5, 2020.

[9] X. L. Han, Y. Liu, Z. J. Zhang *et al.*, "Overview of theory and technology of network security situation awareness and research on difficult issues," *Information Security and Communications Privacy*, vol. 7, pp. 1009–8054, 2019.

[10] H. X. Li, H. Qi, L. Zhao *et al.*, "Network security situation awareness based on quantum neural networks," *Journal of Shenyang Aerospace Ace University*, vol. 40, no. 1, pp. 78-85, 2023.

[11] Y. Q. Tang, C. H. Li, J. Wang *et al.*, "IGAPSO-ELM: A network security situation prediction model," *Electronics Optics & Control*, vol. 29, no. 2, pp. 30–35, 2022.

[12] Z. J. Xue, "Planning and design of financial security situation awareness platform," *Information Technology & Informatization*, vol. 9, 2020.

[13] Y. Y. He, "Network information security situation prediction method based on big data analysis," *Computer and Information Technology*, vol. 32, no. 3, 2024.

[14] Y. D. Wang, "Research on optimization and application of network information security situation awareness in smart city construction," Wuxian Hulian Keji, vol. 21, no. 17, 2024.

[15] Y. M. Gao, H. Wei, J. P. Guo *et al.*, "Big data network security situation prediction based on improved wavelet neural network," *China Computer & Communication*, vol. 35, no. 22, 2023.