SECURITY SITUATION AWARENESS SYSTEM BASED ON ARTIFICIAL INTELLIGENCE

HAO WU

Abstract. There is rapid growth of the security threats faced by enterprises and the security attacks technology is also established at a higher level. Enterprises are facing an escalating threat landscape marked by sophisticated security attacks. To address the structural and technical challenges of information security situational awareness, a method for designing an artificial intelligence-driven system is proposed. In order to solve the system structure and key technical problems of information security situational awareness technology of artificial intelligence, a method of designing information security situational awareness system is proposed, and experiments are carried out through the method. By analyzing the data sources that the system needs to collect, including network traffic mirror data, log data, security intelligence and support data, this paper verifies the feasibility of the system method of information security situational awareness technology of artificial intelligence technology. The proposed core competence platform has the characteristics of low delay and robust real-time capabilities. By presetting the event processing topology, we can quickly build the event processing process, and build the corresponding event processing topology model according to different processing requirements to meet the business requirements. The AI-powered information security situational awareness system substantially enhances security awareness and prediction accuracy.

Key words: Network security; Situational awareness; Network defense; Real time; Event processing topology; Artificial intelligence

1. Introduction. With the continuous evolution of Internet hacker technology, the security of information network is constantly challenged, and the potential threats are becoming greater and greater. The traditional passive defense system can not meet people’s needs for network security [1]. The emergence of network security situational awareness technology opens up a new way to ensure the security of information network. With the advent of the Internet era, people’s lifestyles have undergone earth shaking changes. All walks of life are constantly upgrading and transforming under the impact of the Internet. At the same time, all kinds of network security threats in the Internet era are also increasing [2]. In recent years, cyber attacks with national and organizational backgrounds are increasing. The special roles of the government, military, finance and large enterprises often face more external attack threats. Therefore, the research on information security situational awareness system is necessary [3].

The rapid development of internet has caused a quick growth of network data. While it brings expediency to work and life of people, the large data also brings great risks of security to the network. Therefore, the security situation awareness method is developed, such as the Bayesian method based network security situation awareness model and the improved G-K algorithm based multi-node network security situation prediction awareness model. The security events in the network are detected by these two models but the security situation awareness model classification model is not good. New cascaded network security situational awareness model based on fusion decision tree algorithm is built. With rapid growth of energy internet and AI, big data are playing important roles in the management mode and value function construction. The large power data research is being paid more attention and data processing complexity is getting higher and higher, which brings challenge to the traditional security transmission management. The economical and highly scalable IT services are provided by the cloud computing for the remote computer users. The cloud-based data transmission technique is the transmission of big data from a cloud storage point to a destination storage point according to the cloud storage. This technique is advantageous as it can get rid of the hardware resource limitations. The data overflow will occur if there is limited cloud space capacity. The device itself, including trusted computing,
network equipment, security protection equipment, databases, etc. generates all the information that needs to be collected. After summary processing, it is submitted to the network security management platform on the main station and the plant side.

1.1. Contribution.
1. The article proposes a method for designing an information security situational awareness system to tackle the structural and technical hurdles associated with artificial intelligence technology.
2. Through experimentation, the method is tested, validating its efficacy in analyzing necessary data sources for the system.
3. The contribution underscores the feasibility of the proposed system in leveraging artificial intelligence to enhance information security situational awareness.

Rest of paper is organized as follows: Exhaustive literature survey is detailed in section 2 followed by the research methods in section 3. Results are discussions are presented in section 4 and the section 5 concludes the paper.

2. Literature Review. Xu, R. and others found that network security situational awareness comes from situational awareness [4]. Cohen, R. S. and others believes the recognizing and understanding environmental factors within a certain space-time range [5]. Chan, J. L. and others found that the bass functional model of network security situational awareness has been widely recognized. [6]. Starting from bass functional model, people’s research mainly focuses on the following contents: First, the related technology of data collection. Korolyov, V. and others found that due to the diversity of network sensors and the complexity of network structure, how to effectively select sensors and fuse data plays a vital role in the subsequent situation analysis [7]. Khairy, D. and others pointed out that data collection is the most important part of the whole situation analysis cycle, and divided the data sources into complete content data, session data, statistical data, packet string data and alarm data. In addition, they also proposed an application collection framework (ACF) to reduce the complexity of data collection [8].

Second, the related technology of object extraction. Azar, R. and others believe that the original data has the characteristics of large amount of data and more redundant information. The object extraction process is the process of detecting the collected data in the network security situational awareness system. It takes the original data as the input and obtains high-level objects based on the original data, such as abnormal events and alarms [9]. Munir, A. and others proposed an alarm aggregation algorithm based on a commercial product. The algorithm aims to eliminate the interference of redundant alarms and obtain high-value aggregated alarms. The aggregated alarms here are the extracted objects [10]. Third, the related technology of situation extraction. The objects extracted from the original data are stored in the object library. The objects in the object library are the basis of situation extraction. The discovery of attack scene is a typical situation extraction process. Elia, G. studied the alarm correlation method based on alarm aggregation for the attack scenarios [11].

Fourth, the related technologies of threat assessment. In the field of network security situation awareness, threat assessment belongs to the category of situation assessment. At present, the mainstream situation assessment methods include knowledge-based reasoning method, statistical analysis method and so on. The methods of knowledge-based reasoning include Bayesian network, D-S evidence theory and so on. The methods based on statistical analysis include weight analysis method and analytic hierarchy process. In general, Lappin, Y. and others found that the current research on network security situational awareness is still in the preliminary exploratory stage, and the network security situational awareness technology is not mature and needs further research [12]. For many years, the information security of enterprises has been in the passive cycle of "defense discovery repair". The common practice is to find the loopholes or risks in the network and information system as early as possible and repair them in time through penetration test or risk assessment. At the same time, when the attack behavior is found, the attack behavior is determined by analyzing the relevant security device logs and network traffic, and the attack is blocked as soon as possible. In this passive defense infosec life cycle as shown in Figure 2.1, the vast majority of enterprises focus 95% on defense and 5% on discovering attacks. Basically, repair is based on the passive repair of patches issued by the original manufacturer of products / devices, and the second is to continuously optimize and improve the defense strategy and improve the defense ability. With the deepening of enterprise information security construction, the defense means of information security are also gradually strengthened. Most enterprises have built security systems such as terminal man-
security, access control, security audit and vulnerability discovery, which ensure the safe operation of business to a certain extent. However, each system has its own way and is independent of each other, so it is impossible to achieve unified management, unified early warning, unified tracing and traceability. On the other hand, due to the large-scale network of large enterprises, there are a large number of logs in different formats, such as Syslog log, web service log, firewall log, NetFlow log, etc. These logs come from various business system servers and many security devices and network devices, which are widely distributed and large in number. These log data are often not effectively managed and fully utilized, and cannot give full play to the analysis role of logs, especially without high-speed collection, normalized storage and correlation analysis of all logs. Shibuya, Y. and others found that in recent years, the more advanced and advanced the technology is, the more attacks the enterprise network faces. Moreover, with the continuous application of new technologies, the means and methods of attack are becoming more and more hidden and difficult to find [13].

Advanced persistent threat (APT) is a complex and covert attack means that can bypass various traditional security detection and protection measures and realize fixed-point attack through careful camouflage, long-term latency and continuous penetration. From the current research on enterprise information security situational awareness system and the current situation of information security protection in China, enterprise information security has been in a passive cycle of discovery and repair. Enterprises generally install corresponding defense systems in combination with their own work characteristics and production nature, find hidden dangers and risk problems in the network system as soon as possible through risk assessment, penetration test and other methods, and take targeted measures to solve them. After discovering the offensive behavior, the system will comprehensively investigate and analyze the whole security equipment log and network traffic, so as to determine the specific degree of behavior, and solve these problems as much as possible. In the enterprise passive circulation information security defense system, the vast majority of enterprises pay more attention to the defense process, but ignore the determination and analysis of the cause of the attack.

Verizon Data Breach Report is shown in Figure 2.2. The current attacks can lead to the leakage of enterprise data and even system paralysis in a few minutes or hours, while it takes weeks or even months for enterprises to find these attacks and effectively stop them. This makes the enterprise’s network, system and data in a dangerous state for a long time, and after the old vulnerabilities are repaired, the attacker will find and exploit new vulnerabilities, resulting in the information security personnel struggling to cope with [14].

Author details the system hardware configuration optimization and the AI synchronous operation mechanism. The information security situation inference algorithm is detailed and improved on the basis of the data support vector. The universal data security features are extracting and designed the system software structure. It is observed by the author that the information security situation awareness system based on big data and AI has improved the efficiency significantly and high accuracy as compared to the existing techniques [15]. Authors in this paper provide a comprehensive
study on the existing literature in the cyber SA for discussing the key design principles, classifications, and analysis of the techniques, and evaluation techniques [16]. Author in this paper details the security situation awareness technology which has become a new research topic in network security. A new cascaded network security situational awareness model is designed based on the fusion algorithms. An induction algorithm is also introduced for the decision tree generation on the pre-processed data for data classification. A new network security situation awareness model is shown by the results [17]. Author in this paper proposed a Power Grid Information Security Perceptual System based on AI technology. The encryption and decryption calculation method are combined and the credible risk assessment theory of dynamic cycle is established. The passive defense of power grid information security problem is solved by it and the power data risk is strengthened and the reliability of information security system of power grid is enhanced [18]. Author in this paper utilized the data mining techniques to study the power control system network security situation awareness technology. The wavelet neural network analysis method is utilized by combining the operational data collection and integrated processing. Finally, calculate the network security status through deep learning and it is concluded by the author that the AI algorithm based on wavelet NN can be utilized for power control system network security situation awareness [19]. Authors in this paper discussed the special issue of six papers on situation awareness in human machine interactive systems in teams of collaborating humans and AI [20].

The author in this paper using a big data-related technologies to analyze, filter, merge, and identify known and unknown security threats and builds a new cascaded network security situational awareness model on the basis of traditional and fusion decision tree algorithms [21]. A decision tree is generated by using the induction algorithm on the preprocessed data for the data classification according to the decision rules. A new network security situation awareness model is constructed by using decision tree calculations. Author in this paper constructed a network security situation awareness framework suitable for big data. A gate recurrent unit (GRU) model is established to extract features from the situation dataset through the deep learning algorithm [22]. This method has a good awareness effect on network threats by the experimental results and has strong representation ability. It effectively perceives the network threat situation which verifies the effectiveness of paper which improves the accuracy of security situation awareness. Author in this paper presents a NSSA that can bridge the current research status and future large-scale application and discuss the classic use cases of NSSA [23]. Finally, various challenges and potential research directions related to NSSA.
is detailed by the survey. In this paper, author summarizes the artificial intelligence and network security situational awareness classic models to provide artificial intelligence overview [24]. Starting from the machine learning, it introduces the neural-network-based network security situational awareness. Finally, summarizes the future development trends of network security situational awareness. In this paper, author presents the information security situation inference algorithm. By extracting the security features of the data source, the system software structure is designed [25]. The steps of comparison and security feature parameters are added to the information security situation awareness process. Finally, the optimal design of the information security situation awareness system is designed optimally. It is observed from the results that the information security situation awareness system on the basis of big data and artificial intelligence has improved significantly.

2.1. Research Gaps. There is rapid growth of the security threats faced by enterprises and the security attacks technology is also established at a higher level. With the advent of the Internet era, people's lifestyles have undergone earth shaking changes. All walks of life are constantly upgrading and transforming under the impact of the Internet. At the same time, all kinds of network security threats in the Internet era are also increasing. In recent years, cyber attacks with national and organizational backgrounds are increasing. The special roles of the government, military, finance and large enterprises often face more external attack threats. Therefore, the research on information security situational awareness system is necessary.

3. Methods. To combat diverse security threats, the methodology devised an information security situational awareness system after extensive research, finalizing its model. It identifies security data sources from enterprise intranet and the internet, including equipment alarms, logs, and threat intelligence. Conducting correlation analysis of internal and external data, the system detects and verifies security attacks and assesses risks using asset vulnerability dimensions. Results are displayed for monitoring. For advanced persistent threats (APTs), the system employs big data threat intelligence to search historical intranet data, visually presenting APT events for analysis and action, thus providing a comprehensive solution to security challenges.

3.1. Overall Design. In order to deal with various information security threats faced by enterprises, Power China launched the research on information security situational awareness system. After a lot of research and demonstration, the overall model of the system is finally determined. The security data sources of enterprise intranet mainly include security equipment alarm, equipment log (network equipment, server, application, etc.), intranet security evaluation data, network traffic data at the boundary of important areas of the network, etc. Internet security data sources mainly include commercial and open source threat intelligence data from the Internet, Internet security public opinion and vulnerability monitoring data. The security situational awareness system conducts correlation analysis of internal and external security data, determines security attacks and verifies them. At the same time, combined with the dimensions of asset vulnerability, it uses the risk assessment model for comprehensive risk assessment, and finally sends the risk assessment results to the threat situation display module for display. For advanced persistent attack (APT), it mainly relies on the threat intelligence of big data, searches the historical data stored in the enterprise intranet, finds the possible unknown threats in the intranet, and visually displays the found apt attack events in the situation display module.

3.2. Proposed System platform design. The working principle of information security situation awareness system is to analyze and perceive the relevant information security situation as shown in Figure 3.1. To this end, we have built a core platform for distributed computing based on distributed storage and big data processing technology. The core technologies of the platform mainly include:

(1) **Hadoop distributed file system (HDFS).** We use Hadoop distributed file system as the file system of the system. HDFS file system has the characteristics of high fault tolerance and can be deployed on low-cost PC servers. HDFS relaxed the requirements for POSIX, so that we can access the data in the file system in the form of stream. On the other hand, HDFS also supports large-scale cluster deployment, so that the demand for high-throughput concurrent data access can be solved through unlimited expansion of nodes.

(2) **HBase - Hadoop database.** We use NoSQL database running on HDFS - HBase as the database of the system. HBase has the characteristics of high reliability, high performance, column oriented and scalability [30]. Row key: each row has a unique row key. The row key has no data type. The row key is a byte array. Column cluster: data is organized into column clusters in rows. Each row has the same column cluster, but between rows, the same column cluster does not need to have the same column modifier. In the
HBase stores column clusters in its own data files, which are defined in advance. Column modifier: a column cluster defines a real column, which is called a column modifier. The column modifier is the column itself. Version: each column can have multiple configurable versions. HBase obtains data through the version specified by the column modifier.

HBase, a data definition, storage and use mode based on columns rather than rows, is very suitable for dynamically adding data attributes. Through HBase, a large table can be created, and the attributes of this table can be dynamically added according to needs, especially suitable for the processing of unstructured data.

4. Results and Analysis. The data sources that the system needs to collect include: network traffic image data, log data, security intelligence and support data. Among them, the log data is relatively standard. You can export syslog log, web service log, firewall log, Net-Flow log, etc. by configuring the logs of relevant devices and servers [33-35]. At present, there is no unified standard for security intelligence and support data. We normalize the security intelligence data into the intelligence data that the system can identify and use. At the same time, we regularly update the intelligence base by synchronizing the cloud server or upgrade package, and store all kinds of intelligence and support data in the system for system processing and analysis.

The large-scale data acquisition and processing platform must have the ability of multi-point data acquisition and fault tolerance, especially for the large-scale data acquisition and processing center. The system pre-processes the original image traffic, uses multi-core parallel processing means to analyze, restore and analyze the original network data with large traffic, and then forms a unified traffic log format and uploads it to the big data platform for storage. The architecture of flow acquisition probe is shown in Figure 4.1.

One of the characteristics of the data layer of the attack characteristic event map is that a single attack characteristic event is a weakly connected branch of the whole attack characteristic event map. If \( E \) represents the attack characteristic event map and \( G \) represents a single attack characteristic event. From the perspective of set, \( E \) represents the complete set, and \( G \) is a division of the complete set \( E \). This design can facilitate the system to traverse each weakly connected branch of the attack characteristic event map when discovering the attack behavior in the later stage. The traffic collection probe is mainly divided into two modules. The basic traffic processing module is responsible for preprocessing the original traffic, including basic packet reorganization and traffic reorganization, and can analyze the information of traffic transmission layer and network layer; The high-level protocol processing module is also divided into abnormal behavior discovery,
FIG. 4.1: TRAFFIC COLLECTION ARCHITECTURE DIAGRAM

Protocol resolution and message transmission modules. The protocol resolution module is responsible for in-depth resolution of application layer protocols, analyzing the information of application layer protocols such as HTTP, DNS and SMTP, and extracting key information to the message transmission module. At the same time, restore the files contained in HTTP, SMTP and other protocols, and send the restored information to the big data platform for saving.

This technology mainly adopts the optical splitter image or network port image technology to export the traffic in the network, and then input it to the analysis platform for correlation analysis. Traffic restoration and data analysis can perform high-performance analysis on mainstream protocols such as HTTP and SMTP/POP3 in IPv4/IPv6 network environment, and restore the files transmitted by mainstream P2SP software through fragment file detection and P2SP reorganization.

(1) Port matching. In the process of network protocol development, a series of standard protocol specifications have been formed, which stipulate the ports used by different protocols. Although some other widely used applications do not have standardized ports, they have formed de facto standard ports. Port matching is to use TCP/UDP ports to identify behaviors according to the corresponding relationship between standards or factual standards. This method has the advantages of high detection efficiency, but it is easy to be forged. Therefore, on the basis of port detection, it is necessary to add the judgment and analysis of feature detection to further analyze the data.

(2) Traffic feature detection. There are two kinds of traffic feature detection. One is the identification of standard protocol traffic. The standard protocol stipulates a unique message, command and state migration mechanism. These traffic can be accurately and reliably identified by analyzing the proprietary fields and states of the application layer in the traffic packet; The other is the identification of undisclosed protocol traffic. Generally, it is necessary to analyze the protocol mechanism through reverse engineering and identify the communication traffic directly or through the characteristic field of message flow after decryption.

(3) Automatic connection and association. With the development of Internet applications, more and more data are transmitted on the Internet.
(4) Behavior characteristic analysis. For some data flows that are not easy to restore, we use the method of behavior characteristics for analysis, that is, the system does not try to analyze the data on the link, but uses the statistical characteristics of the link, such as the number of connections, the connection mode of a single IP, the proportion of upstream and downstream traffic, packet transmission frequency and other indicators to distinguish the data flow. Because our core platform adopts the stream framework based on big data technology, we can stream all kinds of data according to the predetermined process to ensure the accuracy of all kinds of data processing. Stream framework is a distributed structure that supports horizontal expansion. By adding cluster nodes, the concurrent processing ability of the cluster can be improved. Stream framework also has automatic fault tolerance mechanism, which can automatically handle process, machine and network exceptions to ensure the stable operation of event processing process. When processing data, the data is not written to the disk and cached in the memory of each node. Our core competence platform has the characteristics of low delay and strong real-time. By presetting the event processing topology, we can quickly build the event processing process, and build the corresponding event processing topology model according to different processing requirements to meet the business requirements. Efficiency comparison result of system operation is shown in Figure 4.2. The data security processing accuracy was tested and compared with the security situation awareness system and the result was as follows:

The big data and AI technology information based security situation awareness system has improved the security awareness and prediction accuracy significantly. The System operation accuracy contrast detection is shown in Figure 4.3.

The information security situation awareness based on the big data and AI has improved effectively. The prediction accuracy of the massive data is also improved.

5. Conclusion. In the realm of networking, the volume of data and the accuracy of awareness have emerged as pivotal concerns across various sectors. Traditional security perception systems often suffer from inadequate perception, defense accuracy, and operational efficiency. To address these issues, optimization of information security situational awareness systems leveraging big data backgrounds has been undertaken, with AI technologies ensuring enhanced accuracy and system efficiency, thereby bolstering network information environment security. Presently, the structure of information network security situational awareness systems grounded in artificial intelligence primarily encompasses stages such as information extraction, pre-processing, fusion, situational awareness, and assessment. Key performance indicators during system operation include basic operation, network vulnerability, and threat indicators, which underpin the functioning of situational
awareness systems. This technology advancement not only fortifies information network security but also integrates various technologies like data mining, fusion, and pattern recognition, effectively addressing early-stage security issues. Its significance extends to fostering the safe and reliable operation of power systems, thereby supporting societal production and daily life activities. Given the indispensable role of network information technology across diverse sectors, enterprises must prioritize innovation in information security technology and adeptly apply information management skills to ensure orderly progress and enterprise development. Looking ahead, optimizing decision tree algorithms will be considered to bolster models, overcoming local optimization limitations and enhancing efficiency.

REFERENCES

Edited by: Pradeep Kumar Singh

Special issue on: Intelligent Cloud Technologies Enabled Solutions for Next Generation Smart Cities

Received: Nov 12, 2022
Accepted: Mar 1, 2024