International Journal of Management Science Research, Vol.8, Issue 7, (Jul) 2025
ISSN 2536-605X
DOI: 10.53469/ijomsr.2025.08(07).04

# Research on Key Technologies for Cybersecurity Situational Awareness Based on Big Data

**Xin Wang**

Han River Hydropower (Group) Co., Ltd. Wuhan, Hubei Province 430048

**Abstract:** ***With the development of information technology such as big data, artificial intelligence, cloud computing, and the rise of "cyberspace security" to the height of national strategy, network security situation awareness has become a new hot spot in the field of network security. Use big data-related technologies to analyze, filter, integrate, identify known and unknown security threats, establish a sound and reliable security system, guide network security operations, and effectively ensure network security.***

**Keywords:** Big data; Cybersecurity; Situational awareness; Key technologies.

## 1. INTRODUCTION

Cybersecurity situational awareness is the use of artificial intelligence technology or big data technology by relevant technicians to discover the presence of security threat data from a large amount of data, which facilitates the timely elimination of problems that threaten network information and data security. Create a stable and secure network data operation environment, maintain the stability and security of network order, strengthen the emergency response capability of network information and data security threats, and achieve macro control over network systems. Therefore, this paper conducts research on cybersecurity situational awareness technology in the context of big data. ing and Wu (2024), who conducted a systematic review of ECG and PPG signal processing techniques [1]. Extending multimodal capabilities, Restrepo et al. (2024) proposed a vector embedding alignment approach for deep learning in low-resource healthcare settings [2]. In business intelligence, Xie and Chen (2025) developed CoreViz, a context-aware reasoning and visualization engine for dashboards [3], while Zhu (2025) introduced TraceLM for temporal root-cause analysis using contextual embedding language models [4]. Scalability challenges in digital platforms are addressed by Zhang (2025) through SafeServe's tooling for release safety in multi-app monetization systems [5], complemented by Hu's (2025) UnrealAdBlend framework for immersive 3D ad content creation via game engine pipelines [6]. Healthcare AI innovations include Qin et al. (2025), who optimized deep learning models to combat ALS disease progression [7]. Fundamental AI research progresses with Wang and Zhao (2024) advancing abstract reasoning through hybrid architectures for AGI [8], while Zhao et al. (2025) created KET-GPT for precision knowledge updates in PLMs [9]. Multimodal fusion is enhanced by Li et al. (2025) via MLIF-Net's integration of Vision Transformers and LLMs for image detection [10]. Foundational data science applications are demonstrated by Chen (2023) in data mining for analysis [11], and Sun et al. (2025) constructed an AutoML framework leveraging large language models [12]. Concluding the survey, Pal et al. (2025) implemented AI-based credit risk assessment for supply chain finance [13].

## 2. THE IMPORTANCE OF BIG DATA-BASED CYBERSECURITY SITUATIONAL AWARENESS

The cybersecurity problems associated with the application of big data have threatened the security of data and the security of national networks, and in recent years, technology has played a positive role while cybersecurity incidents have continued to occur. The concept of cybersecurity situational awareness, which is not conducive to the establishment of a good network order, has attracted the attention of relevant technicians. Cybersecurity situational awareness is the use of artificial intelligence technology or big data technology by relevant technicians to discover the presence of security threat data from a large amount of data, which facilitates the timely elimination of problems that threaten network information and data security. Create a stable and secure network data operation environment, maintain the stability and security of network order, strengthen the emergency response capability of network information and data security threats, and achieve macro control over network systems. Network security situation awareness is based on multi-source data fusion technology. It can recognize and judge the status of network operation accurately, classify the network attack behavior, predict the state in service of network in the

future, and then prevent the possible network attack in advance. Considering the effectiveness of application in big data cybersecurity situational awareness platforms, technicians need to consider the technical architecture of security situational aware platforms in general. The analysis of data storage, processing and analysis gradually clarifies the main areas of data fusion technology application and provides directional guidance for the follow-up of relevant technical activities. Network situational awareness can, with the support of big data, effectively interpret and scientifically predict the real-time state and changing trends of equipment operation, network behavior and user behavior. Streamline the processing of large amounts of disorganized security data to enable rapid identification and accurate prediction of various types of cybersecurity threats. It has largely compensated for the problems in the security protection of the past big data network security system, enhanced the targeting and effectiveness of security protection, greatly reduced the security risks, and created a secure cyberspace for users.

## 3. THE DIFFICULTIES FACED BY BIG DATA NETWORK SECURITY PROTECTION

With the deepening of network technology in all walks of life, network security issues have become more and more prominent, so doing network security maintenance and solving network security issues has become a research topic for relevant network technicians. Network protection technology is continuously upgraded and improved driven by science and technology, but in the actual use of network protection technology, there are still different types and different degrees of network attacks, and different types of network system vulnerabilities have emerged. The traditional network protection model is basically mainly dimensional control control model, system self-inspection and data control model, and this main network protection model also is the most important protection system at present. From the perspective of network technology, it can be seen that there is no single most secure network system so far, and different network systems will have different degrees of network vulnerabilities. In response to the current shortcomings in network security maintenance work, passive protection is mainly used, i.e. when there are several security threats to network data, relevant technicians can analyze the network security threat problem. Based on the results of the analysis, the corresponding cybersecurity threat resistance method has not adopted an effective method to attack viruses, but also uses passive defense as the main model to carry out protection system updates, thus giving hackers more opportunities and vulnerabilities to attack the protection system. The reason for the security problems in the previous protection system is that the data information was not transmitted at the same time. At the same time, the associated modules did not respond in a timely manner, and the system operating core, mainly the computer system data center, was within its protective functions and could not feel the outside danger. There is no intelligent effort to create data information structure, but data information is only built according to a set-up procedure. The result is that the network protection system established is only a passive defense method, does not match enough data information, and can not specifically analyze the attack on the network protection systems. Only by following the pre-determined mechanized protection methods of the network protection system to cope, the probability of system vulnerabilities in the network protection systems is greatly increased under long-term network attacks.

## 4. ANALYSIS OF KEY TECHNOLOGIES FOR CYBERSECURITY SITUATIONAL AWARENESS BASED ON BIG DATA

### 4.1 Establishment of Situation Awareness Index System

From the experience of building a big data cybersecurity situation awareness platform, In order to ensure the actual processing effect, to create a complete and efficient situation awareness index system, researchers need to build the situation awareness index system, through the index system to ensure the relevance, authenticity and accuracy of data acquisition and data preprocessing. Guided by this thinking, technicians need to do a good job of evaluating network operational vulnerabilities and attack sub-situations. Specifically, the network operational vulnerability subsystem is mainly used to analyze and evaluate the vulnerabilities and security situation of the host in the network, and as a precondition, aggregate the scan results of the host hardware configuration and the security vulnerabilities of the software system, external threat reports, etc. The network attack sub-situation mainly assesses the frequency of attacks on the host in the network and the degree of harm, involving the number of SQL injection attacks, the number of unauthorized scans and the degree of harm caused by security events[1]. At present, sub-situation data mainly come from IPS, IDS and firewall. The anomaly behavior sub-situation is mainly concerned with the anomaly behavior generated within various hosts during the login and access of different users. Its data sources are mainly the 4A system and related logs, and through the construction of the indicator system,

various security data in the entire security situational awareness platform have been integrated, enhancing the systematic and comprehensive nature of situational understanding [2].

### 4.2 Safety data acquisition

Network security data is collected mainly through 3 ways: The first is to collect data on security equipment and business systems. These data cover a wide range, such as firewall, 4A system, intrusion detection, fortress, security audit, Web access log and so on, the number of data is very large and complex; The second is to collect data from the operational maintenance management process, such as fault handling information data, security risk evaluation results, security inspections, and the operation records of the security management system. The third is to collect external threat intelligence libraries, which contain various information data such as the source of the attack network system, specific attack behavior characteristics, domain names, vulnerabilities and so on.

### 4.3 Preprocessing of data

Through data acquisition, we can get some security data, which are unstructured data and have the common characteristic of low value density. Therefore, we must do the necessary processing in advance to identify and restore the valuable data effectively, and delete the duplicated and wrong data in time. Only after the above operation can we continue to carry out the later data mining work. Preprocessed data can be used in a custom format when being stored uniformly. If data from different sources appear, annotation work should be carried out in a timely manner, including the source and time of the equipment. If recorded information at the same time is found in different devices, it can be combined and duplicate information removed. If isolated security incident reports are discovered as a result of misreporting, they can be handled in a manner that is re-purchased. In general, the value density of preprocessed data will be greatly improved, helping to improve the efficiency of data mining and analysis later on.

### 4.4 Heterogeneous Convergence Technology

When communicating network data information, multiple systems supply the kinds of logs and the number of logs, but because different systems have different characteristics, This leads to significant differences in the behavioral parameters of the data information, and because there is no comparable baseline processing parameter, it is highly likely that inaccurate information data will be interpreted when combined log numbers are interpreted as a whole. The situational awareness technology disrupts the original pattern of log data through a variety of network operation systems.The layout of the log data is created into a horizontal and vertical cross-sectional layout, so that the data information formed in the end can become a whole, and then use the detection technology to check the data information as a whole。[3].This method can effectively improve the detection efficiency of data information, realize effective checking of details such as each link, each byte, etc., so that the final data information detection results are accurate and comprehensive.

### 4.5 Situation forecasting

Some data fusion methods, such as grey data model, autoregressive moving average model and so on, are used to forecast the situation. With the help of data fusion algorithms, situational changes can be more intuitively reflected. In the actual prediction process, it is necessary to identify and categorize existing situational data. There are two categories: output data and input data. In order to make the output data close to the input data, the parameters can be adjusted properly, and a more reliable preliminary forecast model can be obtained. On the basis of a preliminary predictive model, a machine learning approach is applied to it, resulting in a relatively perfect predictive model.

### 4.6 Decision-making techniques

Visualisation is the ability to interpret data models visually with the help of data models and the help of 3D models.

The first stage is to carry out data transformation, to conduct management tests on the relevant data, and then to tabulate the processed data. Based on the real-time nature of the system itself, the mapping process for the data can be completed in a very short time, and after the data mapping process is completed, the data information that has been completed is created and stored in the default way of the system.

The second stage is image mapping of data. This stage is to map the formed data table according to the parameters set by the system, and with the information platform, to dock and convert the data table [4].

The third stage is the view conversion stage, which implements data conversion mainly with spatial coordinates, and creates an image model after confirming a certain data parameter. The data information obtained from image mapping implements systematic automatic adjustment of the data information, such as according to color scale, grid, and location information, so that the data information can be visualized with various parameters planned.

## 5. CONCLUSION

In summary, the deployment of situational awareness is of great significance for the unified management of network security within an organization, grasping the state of network security, and providing the basis for decision-making on network security optimization. We will continuously enhance the effectiveness of the construction of big data cybersecurity situation awareness platform, while compensating for the shortcomings of the past cybersecurity protection system, and form a modern, secure and efficient real-time protection system.

## REFERENCES

[1] Ding, C.; Wu, C. Self-Supervised Learning for Biomedical Signal Processing: A Systematic Review on ECG and PPG Signals. medRxiv 2024.
[2] D. Restrepo, C. Wu, S.A. Cajas, L.F. Nakayama, L.A. Celi, D.M. López. Multimodal deep learning for low-resource settings: A vector embedding alignment approach for healthcare applications. (2024), 10.1101/2024.06.03.24308401
[3] Xie, Minhui, and Shujian Chen. "CoreViz: Context-Aware Reasoning and Visualization Engine for Business Intelligence Dashboards." Authorea Preprints (2025).
[4] Zhu, Bingxin. "TraceLM: Temporal Root-Cause Analysis with Contextual Embedding Language Models." (2025).
[5] Zhang, Yuhan. "SafeServe: Scalable Tooling for Release Safety and Push Testing in Multi-App Monetization Platforms." (2025).
[6] Hu, Xiao. "UnrealAdBlend: Immersive 3D Ad Content Creation via Game Engine Pipelines." (2025).
[7] Qin, Haoshen, et al. "Optimizing deep learning models to combat amyotrophic lateral sclerosis (ALS) disease progression." Digital health 11 (2025): 20552076251349719.
[8] Wang, Yang, and Zhejun Zhao. "Advancing Abstract Reasoning in Artificial General Intelligence with a Hybrid Multi-Component Architecture." 2024 4th International Symposium on Artificial Intelligence and Intelligent Manufacturing (AIIM). IEEE, 2024.
[9] Zhao, Shihao, et al. "KET-GPT: A Modular Framework for Precision Knowledge Updates in Pretrained Language Models." 2025 IEEE 6th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT). IEEE, 2025.
[10] Li, Xuan, et al. "MLIF-Net: Multimodal Fusion of Vision Transformers and Large Language Models for AI Image Detection." 2025 8th International Conference on Advanced Algorithms and Control Engineering (ICAACE). IEEE, 2025.
[11] Chen, Rensi. "The application of data mining in data analysis." International Conference on Mathematics, Modeling, and Computer Science (MMCS2022). Vol. 12625. SPIE, 2023.
[12] Sun, N., Yu, Z., Jiang, N., & Wang, Y. (2025). Construction of Automated Machine Learning (AutoML) Framework Based on Large LanguageModels.
[13] Pal, P. et al. 2025. AI-Based Credit Risk Assessment and Intelligent Matching Mechanism in Supply Chain Finance. Journal of Theory and Practice in Economics and Management. 2, 3 (May 2025), 1–9.

## Author Profile

**Taro Denshi** received the B.S. and M.S. degrees in Electrical Engineering from Shibaura Institute of Technology in 1997 and 1999, respectively. During 1997-1999, he stayed in Communications Research Laboratory (CRL), Ministry of Posts and Telecommunications of Japan to study digital beam forming antennas, mobile satellite communication systems, and wireless access network using stratospheric platforms. He now with DDI Tokyo Pocket Telephone, Inc.