Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2024 SCPE. Volume 25, Issues 2, pp. 1107–1116, DOI 10.12694/scpe.v25i2.2604

RESEARCH ON NETWORK SECURITY SITUATION AWARENESS TECHNOLOGY BASED ON SECURITY INTELLIGENT MONITORING TECHNOLOGY

BINGYU YANG*

Abstract. This paper uses data mining technology to dynamically monitor tobacco Industrial Enterprise' information systems. This paper builds an Internet security situation awareness system under a big data environment. The weight clustering method is used to classify users' network behavior. The spacing of weights is optimized to ensure the maximum difference in classification. Then, NAWL-ILSTM technology establishes a security situational awareness model for the Internet environment. In this project, the extended and short-memory Nadam optimal algorithm (NAWL) is used to realize data deep learning. Finally, the tobacco industry network security situation assessment method is designed to complete the dynamic monitoring of tobacco industry network security based on data mining. Simulation results show that the proposed method can effectively improve the safety evaluation performance of the system and reduce evaluation errors.

Key words: Association analysis; Intelligent monitoring; Algorithm detection; Situational awareness; Network security

1. Introduction. The US Air Force developed situational awareness technology in the 1980s. Its role is to analyze and judge the situation of the war and provide relevant information for the situation of the war. It is used in high-level decisions to win wars. At the network security level, situation perception research focuses on the analysis of information about the forms and development trends of network attacks. The research of situation awareness has the characteristics of global, dynamic, complex, effective and accurate. The research on situation awareness can be divided into three levels [1]. The first level of situational awareness refers to collecting large amounts of information or data. The main content includes host, network, security, application, physical, intelligence, threat, etc. The second level is to discuss the association and data integration. Data fusion is a method to process multiple observations from a timing sequence on a computer. It includes data association, merging, and extraction. It displays assets, threats, risks, weaknesses and trends at all levels [2]. The visualization, isomerization, automation and real-time processing of multi-source information are realized through the in-depth study of situation awareness. It provides a scientific basis for risk assessment, decision making and prediction. In recent years, data mining technology has been applied more and more. Some scholars have improved the MD5 algorithm and carried on the hardware design. Although this method can solve hardware-related problems such as encryption and decryption of situational awareness, there is still a lack of targeted methods for accurate authentication in user clustering. Currently, the existing hierarchical and clustered WSN protocol only carries out the optimal verification for different types of user groups and lacks anything related to hardware. Applying the existing theory and technology to practice is a complex problem. Therefore, this paper uses data mining technology to realize the situation awareness and security monitoring of tobacco enterprise information systems.

2. Network security situation awareness system structure.

2.1. Overall system framework structure. Figure 2.1 shows the architecture of the network security situational awareness system. The architecture is distributed and open architecture. Its core concepts are "decentralized access" and "partition management". It can be divided into three levels: information acquisition, factor extraction, and scenario decision-making. This system uses a log-based sensor, SNMP sensor and NetFlow sensor to collect data [3]. The network environment's host, switching equipment, and security equipment are analyzed and studied. The element extraction layer is designed for obtaining many different types of information. The compression and extraction of information are realized by employing aggregation and fusion. The

*China Tobacco Anhui Industrial Co., LTD., Hefei, 230088, China (magicyby123@163.com)

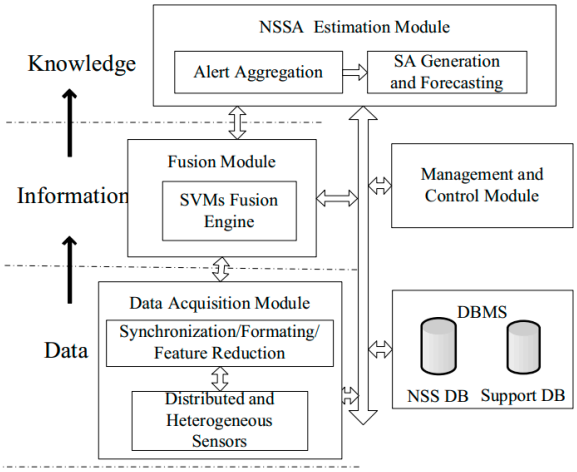

Fig. 2.1: Network security situational awareness system architecture.

situation decision-making level realizes the comprehensive cognition of multi-source information using hierarchical evaluation and the dynamic cognition of situation awareness using nonlinear time series prediction. It gives high-level users an intuitive understanding of the system's security status. In implementing the information acquisition layer, factor extraction layer and scene decision layer, it is necessary to cooperate with the relevant database. It includes an event database, resource database, network information database, knowledge base, etc. Building such a database requires the assistance of experts, security personnel and network scans.

An information perception method based on "monitor-analysis-decision" is proposed. A variety of sensors are used for safety monitoring. Each sensor node can collect corresponding information to achieve the monitoring of the overall operation of the system. Then this paper adopts the information filtering method, verification and fusion to realize the security situation awareness of the global state.

2.2. System Physical Architecture. FIG. 2.2 is a schematic diagram of the system's physical architecture described in this paper. The system includes a sensor, analyzer, decision maker and corresponding database. In each safe zone, there are multiple sensors and multiple analysis programs [4]. In addition to the detection and analysis units, the safety area also contains a discrimination unit. The sensor monitors the central system and local area network and transmits abnormal and suspicious activity reports to the analyzer. The analysis program mainly obtains information from local sensors. And transfer that data to a global database. Decision makers obtain the overall network security degree by analyzing each safety zone and storing it in the database. The two sensors are connected circularly. The goal is to complement information across multiple sensors so that each sensor perceives the raw data. This simplifies the raw data. Similar links are also used in the various analysis programs to achieve better precision in element extraction and consistent publication of the overall strategy. In the process of data acquisition, the analysis instrument and the detection instrument adopt a two-way interactive way. On the one hand, the sensor will actively feed the collected data to the analytical machine. On the other hand, the analyzer can also issue a command to reconstruct the sensor. The analysis results of the local analyzers must not only be stored locally but also transmitted via encrypted channels to the database of the remote central monitoring system. It is submitted to the decision machine along with the results of the resulting global analysis.

Given the wide distribution of large-scale internet, many nodes, operating system and network device differences, multi-level information fusion and decision-making technology are needed. The global, real-time and dynamic security state perception of the network uses multi-source and heterogeneous information. It has the characteristics of self-adaptation and self-learning [5]. It reflects the characteristics of distribution and dynamics. It has both dynamic and static network attack detection functions. This architecture has nothing to do with network topology. This architecture can dynamically adjust the system according to different distributed

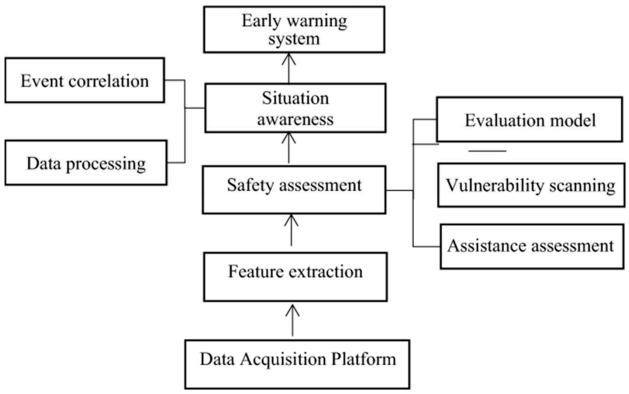

Fig. 2.2: Network security situation awareness model in multi-sensor network environment.

applications and security requirements. In addition, the architecture can effectively solve the problem of a single point of failure and improve the flexibility of the whole system [6]. At the same time, organizations within each security domain still adopt a hierarchical structure, significantly reducing the implementation complexity. This method provides a practical information acquisition method for security situation awareness in distributed large-scale networks.

2.3. System Conceptual Architecture. This paper presents the basic architecture of network security situation awareness. It is mainly used to reflect the flow of information at various levels of perception. From the lowest point of view to the highest point of view are the network security situation element extraction, situation assessment, and situation prediction. It is matched by data information, feature information, and situation information. The first layer is "network security scenario element extraction," which is the basis for the cognition of network security scenarios. At this level, we will focus on extracting the information that can play an essential role in future network security from a large number of multi-source security data. And convert it into a standard XML file. The research results of this project will provide necessary technical support for information security evaluation and early warning in the future network environment of our country [7]. The second layer is network security situation assessment, which is the core of network security situation awareness. It analyzes data from various scenarios to calculate possible hazards to services, hosts, and networks. Level 3 is the prediction of the network security situation. The system predicts the future network security situation according to the past and present network security situation. It allows policymakers to control potential dangers in advance to make the right decisions (Fig.2.3).

3. Design of network security situation awareness algorithm.

3.1. Adaptive weighted clustering method. A dynamic clustering model based on weights is proposed. It eliminates the distinction between data characteristics. Then, a method based on adaptive weighted clustering is proposed to fuse the features with the same network behavior [8]. The method aims to minimize the sum of squares of inter-class deviations. The guiding principle based on this is to maximize the differences between groups. Select a small number of feature sets according to the principle of random non-repetition. Finally, the solution of the increased in-class variance is obtained. The sum of squares of the errors of each cluster determines the weighting of each cluster. When a physical object is reassigned, it must be set according to the distance of the weight. The minimization algorithm with weights is adopted to ensure the maximum difference in classification [9]. The proposed method can effectively improve the system's running speed and reduce the system's calculation amount to achieve the optimal purpose. This method can not only overcome the dependence of traditional clustering methods on cluster center selection but also effectively solve the problem of large-scale feature sets. The flow of this algorithm is shown in Figure 3.1.

(1) Characteristics of standardization. Use $U = \{u_1, u_2, \cdots, u_n\}$ to represent the set of all IP pairs of communication example features. n represents the number of communication instances of IP. $S = \{s_1, s_2, \cdots, s_m\}$

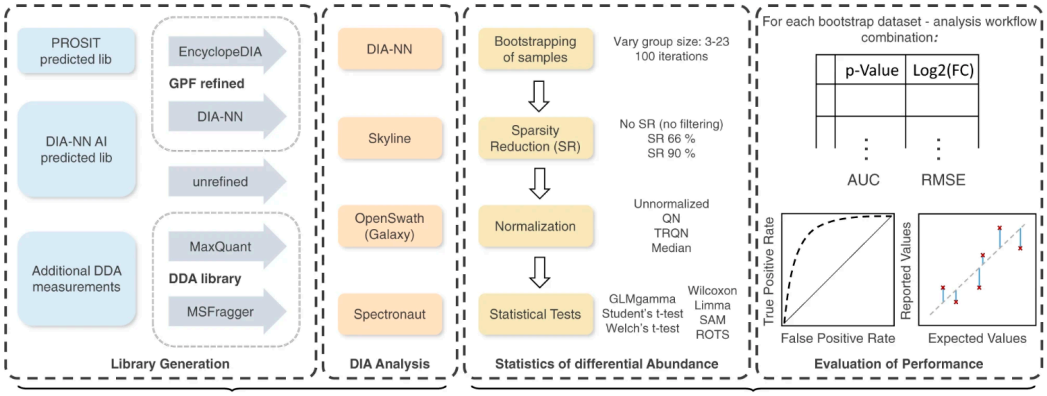

Fig. 2.3: Conceptual structure of network security situation awareness.

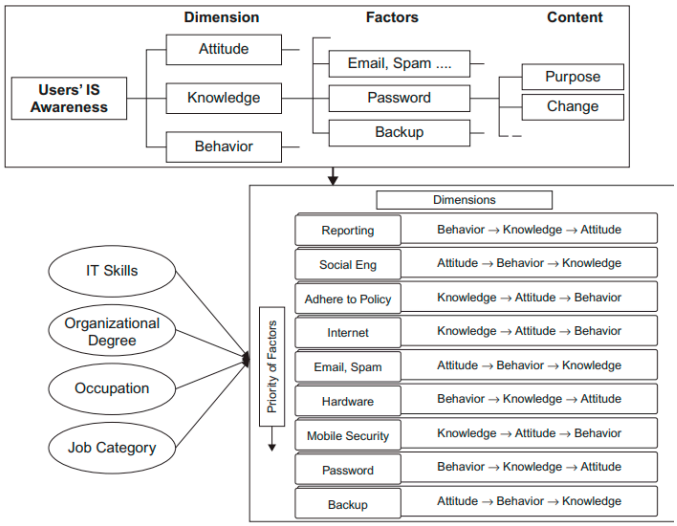

Fig. 3.1: Network in-depth behavior analysis process.

stands for a set of cluster centers. $B_i = \{u_{i,j} | 1 \leq l \leq m\}$. B_i is represented by a set of eigenvectors. The i example uses the representation of B_i. $u_{i,j}$ represents the j function under i of the communication. It normalizes the eigenvalues in the range 0 to 1.

(2) The weighted clustering method is adopted when the center of cluster m is any initial value.

(3) Z similar method is used to perform weighted clustering when determining the dimensions in the feature subset for which A is initialized.

(4) Select any one of $B\prime = \{u_i\}_{i=1}^{Z}$ number of attribute examples in the group.

(5) Calculate the distance between the cluster center and each feature example of the feature subset in each feature example. The samples are classified according to the shortest weight [10]. In this paper, an adaptive clustering method based on weighting is proposed to optimize the weighting of each index. By adaptive adjustment to a cluster, the weight of each cluster is adjusted, and the alternation of the largest and smallest

steps is carried out. The calculation formula to obtain weighting σ_j is (3.1).

$$\sigma_j = \eta\sigma_j^{t-1} + (1-\eta)\left(\frac{R_j^{\frac{1}{1-\alpha}}}{\sum\limits_{j'=1}^{M} R_j^{\frac{1}{1-\alpha}}}\right), 0 \le \eta \le 1 \tag{3.1}$$

σ_j^{t-1} stands for weighting coefficient. η represents the control factor that is weighted repeatedly. $R_j^{\frac{1}{1-\alpha}}$ stands for different weighting characteristics. The current update is determined by the previous iteration weight η. The algorithm can ensure smooth transformation of weighted values during iteration and maintain them within the maximum weighted range during iteration [11]. The premise of minimizing the sum of squared errors is to achieve higher clustering weights. The initial value α_{init} is equal to 0. In each iteration step, the α_{step} value gradually increases to α_{init} until it reaches the maximum value $\alpha_{\max}$. After the maximum value $\alpha_{\max}$, the morphology of the cluster remained at a stable level.

(6) For the cluster center, it needs to be re-calculated.

(7) Return to step (4) to perform the calculation repeatedly. Until the cluster center does not change or the maximum number of iterations is reached.

(8) Calculate the sum of squares of error of the new class M to get the difference value R_α.

(9) If $R_\alpha \ge R_{\max}$ exists, R_α and $R_{\max}$ must be upgraded. Retain the newly generated Mcluster. If there is no $R_\alpha \ge R_{\max}$, then directly cluster the newly generated M.

(10) The result of this clustering is network behavior awareness.

3.2. A situation awareness method in a network environment based on NAWL-ILSTM. An information fusion method based on NAWL-ILSTM is proposed. The extended and short-memory neural networks based on Look-ahead are optimized using the Look-ahead and Naddam algorithms to perceive the network environment [12] effectively. Detailed calculation steps are as follows:

(1) Use $W = (w_1, w_2, \cdots, w_n)$ to represent the accurate timing. The extended Wsequence is a matrix.

$$\begin{bmatrix} w_1 & w_2 & \cdots & w_{x-t+1} \\ w_2 & w_3 & \cdots & w_{x-t+2} \\ & & \vdots & \\ w_t & w_{t+1} & \cdots & w_x \end{bmatrix} \tag{3.2}$$

x is the length of the sequence, and t is the sample size. In formula (3.3), $g = (w_t, w_{t+1}, \cdots, w_n)$ represents the normalized time series W of a training sample.

$$W = \frac{w_i}{\sqrt{w_i^2 + w_{i+1}^2 + \cdots + w_{i-t+1}^2}}, i = 1, 2, \cdots, x-t+1 \tag{3.3}$$

(2) Set the network parameters according to formula (3.4), and set the initial value of the network parameters.

$$\begin{cases} V_g = rand(L, F) \\ p_g = rand(1, F) \\ \vdots \\ Max\_iter = Y_1 \\ Error\_Cost = Y_2 \end{cases} \tag{3.4}$$

V_g indicates the forgotten gate weight. p_g is for forgotten door position deviation. F is the number of ganglion segments and L is the number of $LSTM$ brain segments. Y_1 stands for the maximum number of iterations Max_iter. Y_2 stands for error threshold $Error\_Cost$.

(3) The state information $\widehat{c}_t$ of the unit that needs to be ignored is calculated by formula (3.5).

$$\widehat{c}_t = \mu(V_g * [\zeta_{t-1}, w] + p_g) * \Lambda_{t-1} \tag{3.5}$$

$\mu(V_g * [\zeta_{t-1}, w] + p_g)$ is the output value of the Forget gate. Λ_{t-1} represents the condition of the unit at the last time.

(4) Calculate the amount of information stored in the battery state at A certain time using formula (3.6):

$$\hat{i}_t = \mu(V_i * [\zeta_{t-1}, w_t] + p_i) * \tan\zeta(V_\Lambda * [\zeta_{t-1}, w_t] + p_\Lambda) \tag{3.6}$$

$\mu(V_i * [\zeta_{t-1}, w_t] + p_i)$ indicates the result of the input port i. This output determines the value the battery unit needs to be modified. $\tan\zeta(V_\Lambda * [\zeta_{t-1}, w_t]$ stands for the alternate vector Λ_t, and this alternate vector is created using $\tan\zeta$.

(5) The state Λ_t of the battery is obtained from formula (3.7).

$$\Lambda_t = \hat{i}_t + \hat{c}_t \tag{3.7}$$

From (7) it is possible to see the state of the cell unit obtained by combining the states of the input grid and the forgotten grid.

(6) The output of the network is calculated using formula (3.8) at t time point:

$$\zeta_t = \mu(V_u * [\zeta_{t-1}, w_t] + p_u) * \tan\zeta(\Lambda_t) \tag{3.8}$$

$\mu(V_u * [\zeta_{t-1}, w_t] + p_u)$ represents the result of the output gate u_t. ζ_t is the confidence of the current point in time. $\tan\zeta(\Lambda_t)$ represents the condition of the unit. The reliability of the whole training sample is obtained through repeated calculations of (3.3) to (3.6).

(7) The deviation between the total perceived and actual value is obtained from formula (3.9).

$$\Gamma_{(\beta)}(g, \zeta; V, p) = \frac{1}{2}||g - \zeta||^2 \tag{3.9}$$

The network is modified by the method of back-to-back propagation [13]. Let's go back to step 3. Up to the maximum number of duplicates or error threshold. If the current number of repetitions *iter* is more than Max_iter, or *error* is more than $Error\_Cost$, the training cycle is terminated.

(8) The NAWL algorithm trains the neural network model of ILSTM. Please enter the weighting matrix to be updated $\delta_0 = [V_c, V_i, V_v, V_u]$. The initial values are optimized overall when new sampling points are added. This method only needs a simple iteration to get a new optimization result.

(9) Algorithm for dynamic parameter correction based on online observation results [14]. Add new samples δ_0 and $W_{n+1}(w_{n-t+2}, \cdots, w_{n+1})$. The perceived value ζ_{n+1} of the obtained new sample is propagated forward according to (3.3) (3.6).

$$error = error + \frac{1}{2}(\zeta_{n+1} - w_{n+2})^2 \tag{3.10}$$

(10) Assume that the observed data at the next sampling time reaches where the network was attacked. Relevant network security personnel can detect the alarm the first time the network is attacked and respond quickly [15]. Otherwise, there will be a more profound attack on the network.

4. Experimental simulation. The model's empirical analysis verifies the proposed model's effectiveness in practice. Firstly, the network operation data of tobacco Industrial Enterprise in a certain period is collected to form a sample data set for selection. The collected information includes network access information, network speed, and device storage information [16]. The fuzzy degree of the obtained data is analyzed by the fuzzy correlation degree after the data is preprocessed. The results are analyzed by game theory, and the corresponding learning model is established. The topology of the network is shown in Figure 4.1.

The data from the experiment is imported into the learning model, and the data association cluster composed of 30 central nodes is obtained. The test parameters are listed in Table 4.1.

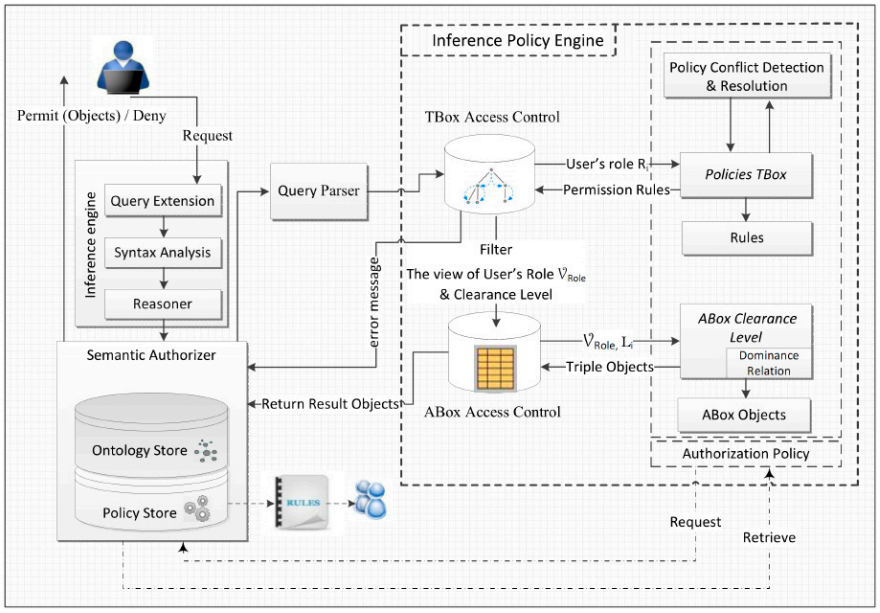

Fig. 4.1: Mesh topology.

Table 4.1: Experimental parameters.

Item	Argument
Computer	Equipped with 2.5GHzInteli5CPU
Running memory	8GB
Storage hard disk	256GB
Operating system	Windows10

The test data in Table 4.1 can meet the requirements of effective operation. The computer data operation program is used to learn and model network security information to obtain abnormal data information [17]. After eliminating the difference in time, the data's similarity is analyzed by the game method. The original data are compared with the results of relevant analysis. Correct the wrong data operation results, and finally obtain the data information that can reflect the network security situation. The results of the cumulative value at risk are shown in Figure 4.2.

It can be seen from Figure 6 that the deviation of the cumulative risk amount described by the method described in this paper is slight and has little impact on the regular operation of the tobacco Industrial Enterprise. The whole system works pretty steadily, and the network speed is fast. Reliable and complete data were obtained after preliminary preprocessing and preliminary screening [18]. At the same time, the time required is significantly shortened. A risk assessment of the operation of the tobacco Industrial Enterprise during the disposal period was conducted, and the assessment was "low." Experimental results show that the proposed algorithm is efficient and stable. The analysis of this data has little impact on the regular operation of the entire network. In addition, the risk to the network environment is also low. The results of the error rate cognition test are given in Table 4.2.

Through the test of the actual system, it is concluded that the algorithm's error rate proposed in this paper is minimal in practical application. This project will use the NAWL-ILSTM algorithm, game theory, machine learning and other technologies to analyze network security situations. In this system, the data information and similarity factors are operated on many levels, the relationship is established, and the time error problem is reduced to obtain a meager error rate. The results show that the calculation accuracy of the data with multi-

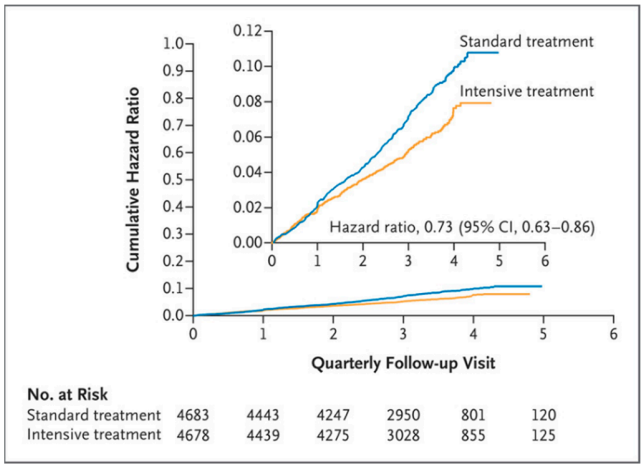

Fig. 4.2: Cumulative var results.

Table 4.2: Error rate perception experiment results.

Number of tests /n	Traditional algorithm	Text algorithm
1	11.71	1.30
2	10.56	1.13
3	12.84	1.27
4	15.88	2.02
5	11.75	1.11
6	13.22	1.70
7	13.79	1.06
8	15.13	1.27
9	14.26	1.42
10	15.50	1.48

layer operation is higher. It can reflect the current network security situation more intuitively and accurately. This project can use various calculation methods to obtain accurate information about network security issues in the environment to ensure the security and stability of the entire network. This method has high efficiency, low risk, and a meager error rate.

5. Conclusion. Combining multi-source and heterogeneous data, a network security situation perception system for tobacco Industrial Enterprise in complex environments was established. The system ring's physical structure and the system layer's concept model are presented. The real-time monitoring and data mining of tobacco enterprise information systems are realized. The application of data mining technology in the tobacco industry can analyze the security events that may occur in the industry communication and network security platform and the association rules and statistics to realize the comprehensive dynamic monitoring of network security. The method proposed in this project can effectively reduce the system's running time and improve the system's reliability for information. The experimental results show that the network security situation awareness technology studied in this project has high practical value in practice.

REFERENCES

[1] Tan, L., Yu, K., Ming, F., Cheng, X., & Srivastava, G. (2021). Secure and resilient artificial intelligence of things: a HoneyNet approach for threat detection and situational awareness. IEEE Consumer Electronics Magazine, 11(3), 69-78.

[2] Zwilling, M., Klien, G., Lesjak, D., Wiechetek, Ł., Cetin, F., & Basim, H. N. (2022). Cyber security awareness, knowledge and behavior: A comparative study. Journal of Computer Information Systems, 62(1), 82-97.
[3] Guo, H., Li, J., Liu, J., Tian, N., & Kato, N. (2021). A survey on space-air-ground-sea integrated network security in 6G. IEEE Communications Surveys & Tutorials, 24(1), 53-87.
[4] Zhang, D., Feng, G., Shi, Y., & Srinivasan, D. (2021). Physical safety and cyber security analysis of multi-agent systems: A survey of recent advances. IEEE/CAA Journal of Automatica Sinica, 8(2), 319-333.
[5] Lv, Z., Qiao, L., Kumar Singh, A., & Wang, Q. (2021). AI-empowered IoT security for smart cities. ACM Transactions on Internet Technology, 21(4), 1-21.
[6] Rosenberg, I., Shabtai, A., Elovici, Y., & Rokach, L. (2021). Adversarial machine learning attacks and defense methods in the cyber security domain. ACM Computing Surveys (CSUR), 54(5), 1-36.
[7] Bécue, A., Praça, I., & Gama, J. (2021). Artificial intelligence, cyber-threats and Industry 4.0: Challenges and opportunities. Artificial Intelligence Review, 54(5), 3849-3886.
[8] Logeshwaran, J., Ramkumar, M., Kiruthiga, T., & Sharanpravin, R. (2022). The role of integrated structured cabling system (ISCS) for reliable bandwidth optimization in high-speed communication network. ICTACT Journal on Communication Technology, 13(01), 2635-2639.
[9] Zhang, J., Pan, L., Han, Q. L., Chen, C., Wen, S., & Xiang, Y. (2021). Deep learning based attack detection for cyber-physical system cybersecurity: A survey. IEEE/CAA Journal of Automatica Sinica, 9(3), 377-391.
[10] Yang, X., Shu, L., Chen, J., Ferrag, M. A., Wu, J., Nurellari, E., & Huang, K. (2021). A survey on smart agriculture: Development modes, technologies, and security and privacy challenges. IEEE/CAA Journal of Automatica Sinica, 8(2), 273-302.
[11] Butt, O. M., Zulqarnain, M., & Butt, T. M. (2021). Recent advancement in smart grid technology: Future prospects in the electrical power network. Ain Shams Engineering Journal, 12(1), 687-695.
[12] Alizadeh, D., Alesheikh, A. A., & Sharif, M. (2021). Vessel trajectory prediction using historical automatic identification system data. The Journal of Navigation, 74(1), 156-174.
[13] Sadeeq, M. M., Abdulkareem, N. M., Zeebaree, S. R., Ahmed, D. M., Sami, A. S., & Zebari, R. R. (2021). IoT and Cloud computing issues, challenges and opportunities: A review. Qubahan Academic Journal, 1(2), 1-7.
[14] Rawal, B. S., Manogaran, G., & Hamdi, M. (2021). Multi-tier stack of block chain with proxy re-encryption method scheme on the internet of things platform. ACM Transactions on Internet Technology (TOIT), 22(2), 1-20.
[15] Endsley, M. R. (2021). A systematic review and meta-analysis of direct objective measures of situation awareness: a comparison of SAGAT and SPAM. Human factors, 63(1), 124-150.
[16] Sabireen, H., & Neelanarayanan, V. J. I. E. (2021). A review on fog computing: Architecture, fog with IoT, algorithms and research challenges. Ict Express, 7(2), 162-176.
[17] Georgiadou, A., Mouzakitis, S., & Askounis, D. (2022). Working from home during COVID-19 crisis: a cyber security culture assessment survey. Security Journal, 35(2), 486-505.
[18] Zhou, X., Liang, W., Li, W., Yan, K., Shimizu, S., Kevin, I., & Wang, K. (2021). Hierarchical adversarial attacks against graph-neural-network-based IoT network intrusion detection system. IEEE Internet of Things Journal, 9(12), 9310-9319.

Edited by: Zhigao Zheng
*Special issue on:*Graph Powered Big Aerospace Data Processing
Received: Sep 12, 2023
Accepted: Oct 18, 2023