ISSN 2348 - 9928
IJAICT Volume 7, Issue 9, September 2020 Doi:10.46532/ijaict-2020029 Published on 05 (09) 2020

Software Development and Design of Network Security System Under Big Data Analysis

[1] Mei Hong Chen
[1] School of Electronics and Computer Science,
[1]Peking University, Yiheyuan Road, Beijing, P.R.China 100871.
[1]hongchen@pku.edu.cn

ArticleInfo
International Journal of Advanced Information and Communication Technology
(https://www.ijaict.com/journals/ijaict/ijaict_home.html)
https://doi.org/10.46532/ijaict-2020029
Received 18 April 2020; Revised form 22 July 2020; Accepted 15 August 2020; Available online 05 September 2020.

Abstract— To explore the prediction effect of network security situational awareness on network vulnerabilities and attacks under the background of big data, this study constructs a predictive index system based on the network security situational awareness model. Based on the improved cuckoo algorithm, the cuckoo search radial basis function neural network is used to predict the situation. The weight value in the model is determined by the hierarchical analysis method, vulnerability simulation is conducted by Nessus software and network attack simulation is conducted by Snort software, and then the situation is evaluated by a fuzzy comprehensive evaluation method. Finally, Jquery and Bootstrap software is used to develop the system. The results show that the cuckoo search radial basis function model proposed in this study could predict network security situations more accurately than the radial basis function model, cuckoo search back-propagation neural network model, genetic algorithm radial basis function model and Support vector machine model based on particle swarm optimization model.

Keywords – Network security situational awareness; Radial basis function; Fuzzy comprehensive evaluation; Network vulnerabilities; Network attack

1. Introduction

With the rapid development of Internet technology, technologies such as big data, cloud computing, and artificial intelligence have been widely applied in every aspect of people's lives. Big data analysis technology refers to the process of analyzing and mining big data. To process massive data, cloud computing and big data analysis platforms have been launched successively. The emergence of new technologies provides convenience to a large extent but also increases the complexity of the network environment. At this time, network security issues become increasingly prominent [1,2]. In 2017, Russian hacker Raeputin used the flaw to gain access to the systems of more than 60 universities in different countries and U.S. government agencies and cut off a large amount of sensitive information. In 2018, a vulnerability appeared in the Intel chip. Based on this vulnerability, attackers directly accessed the Kernel memory data, including the password of the account, which affected the three major operating systems, Windows, macOS and Linux, as well as cloud servers such as Google. In 2018, CVE Detail detected more than 16,000 vulnerabilities, and Microsoft software detected more than 6,000. In the same year Aadhaar, India's national identity system, was viciously attacked by hackers, who obtained biometric data from 1.1 billion people.

These data show that network vulnerabilities and attack patterns emerge in endlessly, and the traditional network security monitoring system can't meet the current needs. The big data analysis of network security mainly focuses on the analysis of security logs and traffic, and assists in the correlation analysis of the vulnerability, user behavior, business behavior and other information [3]. The analysis based on traffic also includes detecting trojans, worms and abnormal traffic. To resist the network system is not attacked, so the emergence of firewall technology, intrusion detection and vulnerability scanning network security technology has been launched. But current technology does not help managers make the best decisions.

Traditional intrusion detection methods are misuse based detection and anomaly-based detection. These two detection methods have poor scalability and low sensitivity to network traffic anomaly monitoring. Moreover, too few feature vectors are selected in the collaborative operation of intrusion detection system, which affects the reliability of the whole detection system for abnormal state detection. Network security situation awareness is a technology that integrates network firewalls and other decentralized security technologies into one, which has gradually attracted people's attention, and network security situation awareness technology can make up for the shortcomings of traditional network security technology [4]. Moreover, studies have shown that situation prediction can analyze the information of the historical situation and predict the future situation [5].

2. Research methods

Construction of network security situational awareness model

Situational awareness is a process of extracting environmental factors, understanding the current situation and predicting the future situation. It has been widely used in aerospace, medical and military fields. However, in a complex environment, the situational awareness model needs to be judged based on the understanding of the decision-maker, the actual situation of the environment and the changing trend. At first, the situational awareness model proposed by Endsley et al mainly consists of five parts: environmental factors (system complexity, system structure and work pressure, etc.), personal factors (expectation, goal, ability, experience and training), situational awareness (factor extraction layer, understanding layer and prediction layer), decision making and execution. Personal factors are the main factors. Based on this, this study constructs a situational awareness model that can be applied to network security monitoring. Fig. 1 shows the concrete structure of the model. It is mainly divided into four levels. Firstly, according to the system log, attack vulnerability, host vulnerability, known attack and service type, the probability of occurrence, probability of successful vulnerability attack and attack weight of known attack at the level of multi-source information fusion are determined. Then the node and service weight of situation factor fusion level is determined. The node and service weight is collected into the network at the node situation fusion level. Finally, the prediction is made at the level of time series analysis.

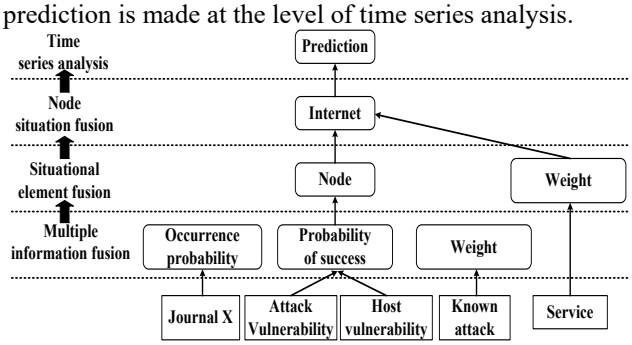

Fig 1. The network security situational awareness model

The basis of the network security situational awareness model is the selection of indicators. Different indicators have a great influence on the accuracy and comprehensiveness of the subsequent assessment and prediction. Existing literature points out that the index system of network security situation can be divided into vulnerability (open ports, safety equipment and operating system version, etc.), disaster frequency (access network, network bandwidth and network topology, etc.), stability (number of survival equipment, traffic, flow range, etc.), threatening (packet number distribution, alarm and bandwidth utilization, etc.), a total of four aspects. According to the actual situation of this study, the indicators are selected on the basis of the network security situation indicator system. In this study, three first-level indicators are selected, namely usability, vulnerability and threat. There are 10 secondary indicators, including CPU utilization, memory utilization, disk utilization and bandwidth utilization, number, score, and type of vulnerabilities; number, type and level of attack alarms. On the basis of the network security situation awareness model, a hierarchical assessment model of the network security situation is built, and Fig. 2 shows the specific structure. It indicates that data sources mainly come from CPU utilization, memory utilization, disk utilization, bandwidth utilization, vulnerability information, and attack information. All data belong to the situation elements. Big data is used to detect the network situation, analyze the security situation after obtaining the situation indicators, and finally predict the network security situation. After the visualization, the network security situation can be evaluated.

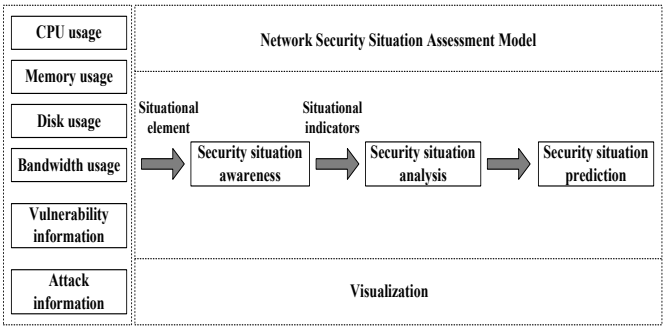

Fig 2. The hierarchical assessment model of the network security situation

Assessment method of the network security situation

According to the established network security situation indicator system and assessment model, the Fuzzy comprehensive evaluation (FCE) method is used to evaluate the network security situation. FCE is an evaluation method that combines quantitative and qualitative analysis. It mainly evaluates and calculates the fuzzy judgment matrix of the target, and normalizes the calculated results. Then the fuzzy calculation of various indicators can affect the problem and finally determine the evaluation results. In FCE, the weights need to be calculated in combination with the Analytic hierarchy process (AHP) to determine the weights.

Suppose the importance of A to B needs to be evaluated, then the judgment matrix of importance can be written as:

$$S = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \quad (1)$$

In Eq. (1), a_{ii} is equal to 1; a_{ij} is the differential coefficient of a_{ji}; the maximum eigenvalue λ is calculated and judged according to the matrix in Eq. (1); after the eigenvectors and eigenvalues are further calculated, the consistency value CI of different indicators is calculated as follows:

$$CI = \frac{\lambda - n}{n - 1} \quad (2)$$

The average random consistency index RI can be obtained by looking up the table. RI values of different orders of n (n=1,2,3, ⋯,9) are -, 0, 0.58, 0.90, 1.12, 1.24, 1.32, 1.41, and 1.45, respectively. If the ratio of CI to RI is less than 0.1, it indicates that the feature vector can meet the requirement of consistency and can be used as the weight vector of the index. If the ratio of CI to RI is greater than and equal to 0.1, it indicates that the feature vector cannot meet the requirement of consistency, and the importance judgment matrix needs to be constructed again. The maximum eigenvalue λ is normalized to obtain the weight vector for evaluating the target element.

After determining the weight value, FCE is used to evaluate the network security situation. Taking the utilization rate of CPU as an example, it can be divided into three different levels: level 1 (<70%), level 2 (70%-90%) and level 3 (>90%). Therefore, the membership function of the FCE algorithm can be determined by using the following equation:

$$F(x) = \begin{cases} 1 & , \quad x < 70\% \\ \dfrac{90\% - x}{90\% - 70\%} & , \quad 70\% < x < 90\% \\ 0 & , \quad x > 90\% \end{cases} \quad (3)$$

When the CPU utilization reaches level 1, it needs to decrease from 1 to 0 at a uniform speed within a range of 70% to 90%. If the CPU utilization reaches level 2, it needs a constant increase in the range of 70% to 90%. If the CPU utilization reaches level 3, the CPU utilization has exceeded the constant value and is 1. Then, the fuzzy matrix is constructed, and the matrix $R=(d_{ij})_{nm}$ is initialized, where n is the number of indicators, m is the rating number, and d_{ij} is the membership degree of the j rank of the ith indicator. The weight vector $W=(w_1,w_2,\cdots, w_i)$ of each index is calculated and obtained, where w_i is the weight value of the ith index. The suitable fuzzy operators are selected for index W, R and synthesis operations, and the calculation equation is as follows:

$$b_j = \left(1, \sum_{i=1}^{n} w_i * d_{ij}\right) \min \quad (4)$$

In Eq. (4), b_j is the jth member of vector B. Then, the membership function is scored subjectively, and the situation calculation equation of the usability index of a single dimension is obtained:

$$S = K \cdot M(W, R)^T \quad (5)$$

In Eq. (5), K is the subjective quantitative score vector; M is the weighted average operation in Eq. (4); W is the weight vector; R is the fuzzy evaluation matrix; T is the transpose of the result.

Network situation prediction method based on the improved radial basis function neural network model

The Radical basis function (RBF) neural network is a network consisting of three feedforward layers (input layer, hidden layer, and output layer). Its main disadvantages are that the initial center selection of the clustering algorithm is difficult, the network structure of the orthogonal least square method is large, and it is easy to fall into local optimization when gradient descent. Therefore, the Cuckoo search (CS) algorithm is introduced to calculate the optimal solution of each parameter in the RBF neural network. RBF neural network requires that the number of nodes in the hidden layer should be larger than the input layer, so the hidden layer belongs to the transformation from a low dimension to a high latitude space. Therefore, this study selects 10, 15 and 20 nodes to construct the hidden layer, and finally selects the node number with the earliest prediction error for subsequent experiments. Then the output calculation equation of the hidden layer of RBF neural network is as follows:

$$Y(X) = \sum_{i=1}^{n} w_i e^{-\frac{\|X - c_i\|^2}{2\sigma_i^2}}, \quad i = 1, 2, \cdots, n \quad (6)$$

In Eq. (6), X is the input vector; Y(X) is the predicted output of X; W_i is the connection weight between the node of the ith hidden layer and the node of the output layer; C_i is the center vector of the ith node; σ_i is the variance of the ith node.

Then, parameters are coded by real coding, and the fitness function in CS is determined:

$$Fit = \frac{1}{1/n\left[\sum_{m=1}^{n}\left(y_m - \hat{y}_m\right)^2\right]} \quad m = 1, 2, \cdots, n \quad (7)$$

In Eq. (7), n is the total sample size of the test; ym is the predicted value. $\hat{y}_m$ is the actual observed value. The discovery probability (P), population size and the maximum number of iterations in the CS algorithm are adjusted, among which the population size is 40 and the maximum number of iterations is 450. Fig. 3 shows the specific steps of network security situation prediction using the Cuckoo search - Radical basis function (CS-RBF) neural network model.

Corresponding author at: Mei Hong Chen, Peking University, P.R.China.

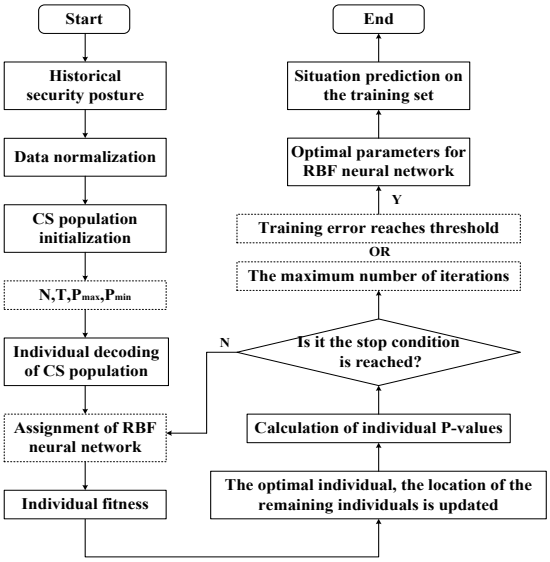

Fig 3 A network security situation prediction method based on CS-RBF neural network model

The design of the network security system

A system that can be used for network security situation awareness is constructed based on the network security situation indicator system, assessment model and prediction model constructed in the early stage. It mainly consists of six modules, which are central database (monitoring data and perception data), data acquisition module (Nessus vulnerability information collection, Snort attack information collection), central control module (task, data and tool management), situation prediction, situation assessment and user interface (assessment of prediction results, attack and vulnerability information statements). Fig. 4 shows the interaction structure between the modules. The central management module is responsible for collecting network security situation indicators and dispatching Nessus vulnerability scanning. After the central management module timer is triggered, it broadcasts the request for information collection to all host nodes. At this point, after the data acquisition module in each node receives the information acquisition request, it starts the information collection and puts the information into storage. Finally, it feedbacks collection status to the central management module. After collecting feedback information from all nodes, the central management module makes an invocation request to the situation assessment module. The situation assessment module queries the weight values of each index and node, determines the movement state of each node, and simultaneously obtains the attack and vulnerability information between two evaluation intervals. The situation assessment module carries out the situation assessment of each node, system and situation, and records the assessment results into the database. The assessment status is feedbacked to the central management module. After receiving all the assessments, the central management module completes the response and sends the prediction request to the prediction module of the situation. The prediction module of the situation receives the request and then sends a request to the database for historical situation information. CS-RBF neural network model is used to predict the situation, and then the predicted results are stored in the database for data training. Finally, the information is updated to the network.

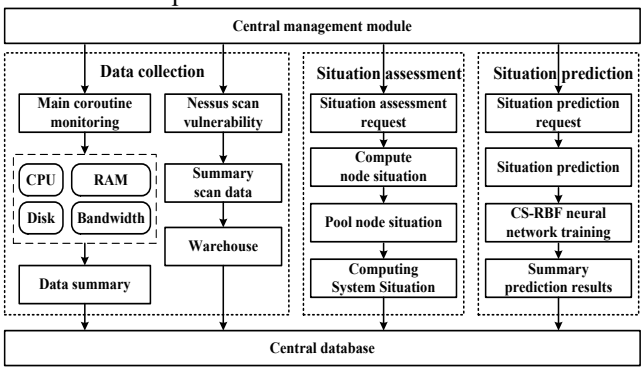

Fig 4. The module interaction structure of the network security system

The user interface module of the network security system is mainly used to show the final situation results to the user. In this study, Jquery Bootstrap software is used to develop the system, and then the situation indicators, security situation assessment and prediction results of each node, and the information of attacks and vulnerabilities of each detection point are mainly presented to users.

3. Results

The validation of radial basis function neural network model improvement

According to the network security situation indicator system and assessment model built in the early stage, this study uses Nessus software to scan vulnerabilities. The Blade IDS Informer software is used to simulate network attacks. Then CPU, memory, loan utilization, and vulnerability and attack information are collected, then AHP and FCE are used to evaluate network security situations. The obtained situation sequence data [0,100] are normalized to accelerate the convergence of the network. Finally, the total sample size is determined to be 250, the driving vector dimension is 5, the training sample size is 200, and the test sample size is 50. Then, based on the actual measured values, the predictive performance indexes of mean absolute error (MAD), mean relative error (MAPE) and mean square root error (RMSE) are calculated. Fig. 5 shows the comparison of predictions. It can be concluded from Fig. 5A that the variation trend of CS-RBF and RBF model in predicting network security situation value is basically consistent with the actual test value of 50 sample sizes. However, it can be concluded from Fig. 5B that the RBF model's MAD, MAPE and RMSE error values are all larger than the CS-RBF model, indicating that the CS-RBF model is more accurate in predicting network security situation.

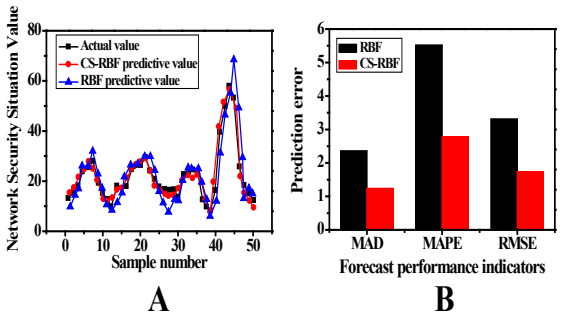

Fig 5. The results comparison of predicting network security situations of models CS-RBF and RBF

Note: Fig. A is the comparison of the prediction effects of different samples; Fig. B is CS-RBF and the prediction error ratio of the RBF model

In order to prove the excellence of the model constructed in this study, the back-propagation neural network model (CS-BPNN) optimized by CS, the RBF neural network model (GA-RBF) optimized by genetic algorithm and the support vector machine model (PSO-SVM) based on particle swarm optimization were selected to predict the network security situation. Fig. 6 shows the results compared with the CS-RBF model proposed in this study. It can be concluded from Fig. 6A that, except the PSO-SVM model, the overall trend between the rest models and the actual values is relatively consistent. As shown in Fig. 6B, the prediction error of the CS-RBF model proposed in this study is the smallest, indicating that the CS-RBF model has stronger adaptability.

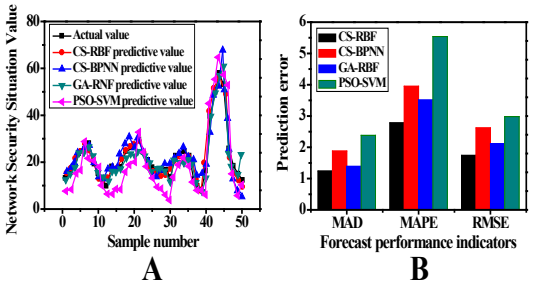

Fig 6 The results comparison of network security situation prediction by different models

Note: Fig. A is the comparison of the prediction effects of different samples; Fig. B shows the comparison of prediction errors between different models

Demonstration of network security system

The home page includes the current state of the system, forecast state, node state information, and attack information. The left side of Fig. 7 shows the information of the current situation, including the overall situation, availability, threat and vulnerability. The right side of Fig. 7 shows the situation information of nodes in the system, including three results: availability, threat and vulnerability, which indicates that the system shows the overall situation of the system through the bar chart, and shows the situation information collected by each node in the form of a table.

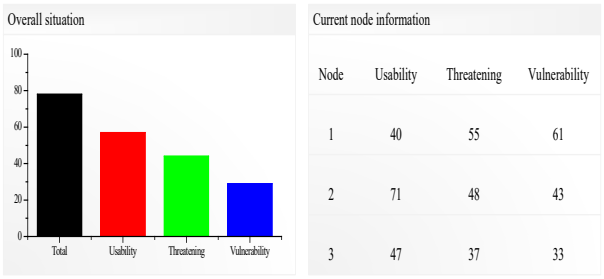

Fig 7 Network security system and node situation information

At the bottom of the homepage, the network situation changes and attack information of recent systems are displayed. It can be concluded from Fig. 8 that the left side shows the change trend of availability, threat, and vulnerability of the system in the recent period. When the user selects a specific time point, the system situation information of that time appears, which shows that the overall situation of the system is in a relatively stable state, but the situation shows a state of constant fluctuation with the increase of time. On the right is the attack information, which mainly lists the top 3 attack types of recent attacks and the types of attacks. Therefore, it can be concluded that the network security situation awareness system designed in this study can effectively complete the monitoring and prediction of network security vulnerabilities and attack information, and can visually and clearly present the predicted results to users, which can provide effective decision-making information for network personnel.

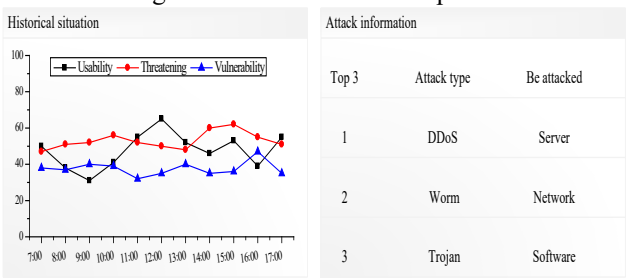

Fig 8. Network security system history and attack information

4. Discussion

With the emergence of mass data, a growing number of information security attacks also appear, the traditional network security technology has been difficult to solve the current security attacks, so the new information security attacks and big data analysis technology has become a hot research issue. Big data analysis is the process of data analysis and mining. It can reach a scale never reached before produce some operational and commercial knowledge specifically. Security analysis technology based on big data can not only solve the problems of big data collection, storage, and mining but also deal with various unknown risks in the network more actively [6]. With the increasing probability of network security time, people gradually realize that network attacks and hacker attacks have brought great threats to people's daily life. Therefore,

network security monitoring and system software development are of great significance. The application of big data analysis technology to the analysis of information security state can effectively integrate the scattered security data, thus digging out the potential security problems and improving the work efficiency. Situational awareness is a method of quantifying and evaluating information from different sources to predict future trends, which was initially widely used in the prediction of information such as space environment [7]. However, network security situation awareness is a comprehensive network security assessment technology combining intrusion detection, vulnerability scanning and firewall technology, and it can also quantitatively predict the future network security situation through the assessment results, which can provide network security managers with the most direct and clear decision-making basis [8]. Chun et al. (2017) showed that the network information security early warning platform based on the security situational awareness model can realize all-sided monitoring of network and information security [9]. Studies have shown that the improved radial basis function neural network model based on the cuckoo algorithm can effectively improve the accuracy of search and classification [10]. Therefore, this study builds a situational awareness model based on network security issues to predict network vulnerabilities and attacks. The improved radial basis function neural network model based on cuckoo algorithm is applied to the prediction of network security situation. Compared with the radial basis function neural network model, it can obviously improve the accuracy of network vulnerability and attack prediction. Zhang et al. (2017) found that network situational awareness can effectively predict network intrusion [11], which is basically consistent with the results of this study that CS-RBF based network security situational awareness model can effectively predict attack and vulnerability information. DDos attack refers to that it can reasonably request resources from the server, thus occupying a variety of services so that users who normally request access can't get a response. DDoS attacks focus on exploiting specific vulnerabilities in the host [12]. There are many common types of DDoS attacks, and the system based on network security situational awareness model designed and developed in this study can effectively predict DDoS attacks.

5. Conclusions

Under the background of big data, to develop and design a network security system that can prevent and predict network vulnerabilities and attacks, a network security situational awareness model is constructed that can be used to predict network vulnerabilities and attacks. After evaluation, it is found that the network security situational awareness prediction model based on the CS-RBF algorithm can predict security issues more accurately. Jquery + Bootstrap software is used for the development of the system, and finally, the system could show users historical situation information, network system security situation information and prediction information. However, the feasibility of the network security situation prediction model proposed in this study has been verified only through simulation experiments, and further studies are needed through practical applications. In conclusion, CS-RBF based network security situational awareness model can effectively predict network security problems. The results of this study can lay a foundation for solving network security problems in the context of big data.

References

[1]. Sabar, N. R., et al., A bi-objective hyper-heuristic support vector machines for big data cyber-security, IEEE Access, 6 (2018), pp. 10421-10431
[2]. Stergiou, C., et al., Security, privacy & efficiency of sustainable cloud computing for big data & IoT, Sustainable Computing: Informatics and Systems, 19 (2018), pp. 174-184
[3]. Hyun, S., et al., Interface to network security functions for cloud-based security services, IEEE Communications Magazine, 56 (2018), 1, pp. 171-178
[4]. Hayes, J., et al., Assessing risks to paediatric patients: conversation analysis of situation awareness in huddle meetings in England, BMJ Open, 9 (2019), 5, pp. e023437
[5]. Zhao, D. M., Liu, J. X., Study on network security situation awareness based on particle swarm optimization algorithm, Computers & Industrial Engineering, 125 (2018), pp. 764-775
[6]. Rehman, M. H. U., et al., Big data analytics in industrial iot using a concentric computing model, IEEE Communications Magazine, 56 (2018), 2, pp. 37-43
[7]. Y. Gao, et al., Uav distributed swarm situation awareness model, Journal of Electronics & Information Technology, 40 (2018), 6, pp. 1271-1278
[8]. Song, Z., et al., Survey of network security situation awareness, Electronic Test, 269 (2017), pp. 3281-3286
[9]. Chun-Lin, C., et al., Practice of network and information security situation awareness in SGCC, Electric Power Information and Communication Technology. 15 (2017), 6, pp. 3-8
[10]. Ayala, H. V. H., Coelho, L. S., Multiobjective cuckoo search applied to radial basis function neural networks training for system identification, IFAC Proceedings Volumes, 47 (2014), 3, pp. 2539-2544
[11]. Zhang, B. C, et al., Network intrusion detection based on directed acyclic graph and belief rule base. ETRI Journal, 39 (2017), 4, pp. 592-604
[12]. Kesavamoorthy, R., Soundar, K. R., Swarm intelligence based autonomous ddos attack detection and defense using multi agent system. Cluster Computing, 22 (2018), Supplement 4, pp. 9469-9476