ISSN 1330-3651 (Print), ISSN 1848-6339 (Online) https://doi.org/10.17559/TV-20211130024245
Original scientific paper

A Cloud Based Network Intrusion Detection System

Li YOU, Zhanyong WANG*

Abstract: Along with the development of intrusion detection systems so far, remarkable results have been achieved in aspects of dynamic monitoring and network defense. However, regarding the ever-increasing volume of network detection data, the limited data processing capacity of intrusion detection systems impedes its pace of development and meanwhile restrains the effectiveness and timeliness of detection of the system. However, the application of cloud computing with its powerful computing capacity in intrusion detection systems can solve this bottleneck problem. Therefore, on the basis of cloud computing, this paper conducts design analysis of the network intrusion detection system and carries out empirical study with reference to the 10% training datasets of KDDCup99 of Lincoln Laboratory. Results show that the cloud computing based network intrusion detection system can effectively detect four types of mainstream attacks, namely Probe, DoS, U2R and R2L, with detection rates all above 94%, the highest false alarm rate being only 4.32% and the longest detection duration being only 50 s, which verifies the feasibility of system detection.

Keywords: cloud computing; feasibility; network intrusion detection system

1 INTRODUCTION

Nowadays, excessive network flow and volume of detection data impede the effectiveness and timeliness of the intrusion detection system [1-7]. And cloud computing emerges as the times require along with the rapid development of the network technology and the new-generation large-scale Internet comes into play in the fields of Internet application and service and becomes a new driving force for current economic growth [8-9]. As a popular emerging technology today, cloud computing is the outcome of the integration of network and traditional technologies (parallel computing, virtualization technology, utility computing and network computing, etc.) [10]. Cloud computing can provide a flexible, scalable, efficient and reliable IT service delivery platform for users and provide convenient service for users as required. Such platforms can preferably handle various problems in this big data era [11]. The problems of excessive data volume and insufficient detection of the intrusion detection system can be properly solved. Therefore, this paper will study the network detection intrusion system based on cloud computing and test the feasibility of the system by means of experiments.

2 SUMMARY OF SECURITY RISKS IN CLOUD COMPUTING ENVIRONMENT

Many security risks derive from the continuous development of cloud computing [12-14], mainly including the following aspects: first, in cloud computing environment, imperfect user registration and login systems lead to the abuse and malicious use of cloud computing; second, in order to reinforce the computing capacity of cloud computing, virtualization management software systems such as VMWare and vSphere will be added to the infrastructure, while the design and development of management software systems barely pay attention to encoding loopholes in the program and then lead to security risks such as virtualization level attacks to the system; attack the host machine through coding vulnerabilities, causing all virtual machines above the host machine to crash and to business interruption; the virtual machine escapes, after acquiring control of the host machine, the host machine is used to infiltrate the cloud computing platform; use host control rights to obtain sensitive data of other virtual machines under the same host. Third, since the resources of cloud computing can be shared among users, such sharing behavior restricts the independence performance of cloud systems, makes user access data vulnerable to be tampered, copied, etc. and threatens the effective guarantee of user resource security.

3 SECURITY REQUIREMENTS OF CLOUD COMPUTING BASED NETWORK INTRUSION DETECTION SYSTEM

Different from traditional computing patterns, the computing pattern of cloud computing faces new challenges and risks; therefore, the cloud computing based network intrusion detection system needs to meet numerous security requirements [15-17], which can mainly be summarized as the following: first, it needs to meet the distributivity requirement. Cloud computing is mainly composed of a large number of distributed, isomerous and parallel computers. The cloud computing based network intrusion detection system should be able to detect all physical host attack behaviors to ensure the security of cloud computing host facilities. Second, it needs to meet the effectiveness requirement. In spite of effectively detecting intrusion behaviors in normal environments, the cloud computing based network intrusion detection system should be able to detect intrusion behaviors in special environments such as virtualization level attacks including account hijacks, virus and Trojans. The system should be able to conduct timely restoration and clearance when these intrusion behaviors occur in order to avoid deterioration of events. Third, it needs to meet the scalability requirement. The number of users and user resources in cloud computing is not static, but is dynamically changing; therefore, the cloud computing based network intrusion detection model should be scalable so as to satisfy the requirements of this dynamic process. Fourth, it needs to meet the timeliness requirement. The cloud computing based network intrusion detection system can shorten the duration of detection and analysis by virtue of its computing resources so as to ensure the timeliness and accuracy of detection when the system detects mass intrusion data.

4 DESIGN OF THE CLOUD COMPUTING BASED NETWORK INTRUSION DETECTION SYSTEM

4.1 Model Design of the Cloud Computing Based Network Intrusion Detection System

At present, the network intrusion detection system is relatively weak in timeliness when processing quantified detection data and the application of cloud computing can effectively solve this problem due to its powerful computing capacity. Meanwhile, cloud computing can strengthen the security defense performance of the intrusion detection system to enable the system to effectively respond to outside invasions and malicious attacks [18-22]. Hence, the design of the cloud computing based network intrusion detection system is of great practical significance. Here, this paper designs the model of the cloud computing based network intrusion detection system which can not only detect intrusion threats quickly but also conduct real-time data processing. The model is mainly composed of the user interaction module, the cloud service catalog module, the intrusion tolerance module, the system management module, etc. as specifically shown in Fig. 1.

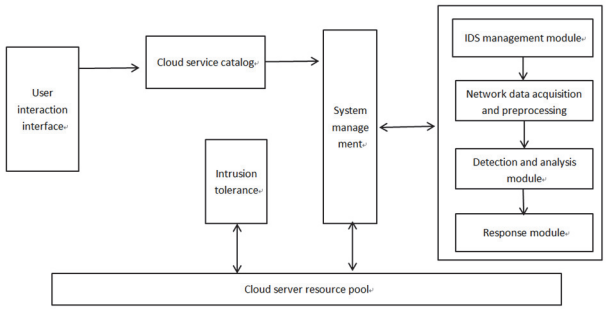

Figure 1 Model of the cloud computing based network intrusion detection system

The user interaction module can provide cloud environment interaction interfaces for users and enterprises as convenient channels for user login and cloud computing resource utilization. The cloud service catalog module enables users to choose services as demanded rapidly. The system management module employs the load balancing technology to realize information intercommunication among the modules and to carry out task management and resource allocation as needed. The intrusion tolerance module, closely connected with the system modules, is mainly responsible for the detection of computing nodes and will automatically transmit the information of node failures to the system module. The system module, when it receives such information, will deploy another node to continue the task so as to guarantee the robustness of the system. The intrusion detection module is one of the important modules in the cloud computing based intrusion detection system as it can provide various functional modules for the system. The IDS management module is mainly in charge of the dispatching of the data acquisition and preprocessing module and the detection and analysis module, and is closely connected with the system management module. It takes advantage of the system management module to improve the rationality, comprehensiveness and security of data acquisition and analysis, and thus to not only give full play of cloud computing in completing the tasks of intrusion detection and analysis but also enhance the self-defense detection capacity of the system. It can be seen that all the modules of the cloud computing based intrusion detection system model are mutually reinforcing and the model can improve the timeliness and accuracy of the system in quantified data detection. The following is detailed analysis of the kernel modules of the model.

4.2 Design Analysis of the IDS Management Module and the System Management Module

In addition to the characteristics stated above, the IDS management module will also periodically transmit the request for intrusion detection to the system management module and the management module will give feedbacks in a timely manner after receiving such request information. After receiving the feedback information from the management module, the IDS management module will immediately give work order to the data acquisition and preprocessing module to conduct distributed acquisition and preprocessing of data before it stores such data in the acquisition submodule in the form of files. The system management module mainly communicates with the IDS management module, responds to its detection requests, carries out storage and computation of processed data file resources and then allocates tasks in accordance with the comprehensive utilization condition of resources of modules in all computing nodes. Because resources are dynamically changing, resource allocation should be reasonable so as to improve the overall detection speed. This paper employs scoring mechanism in resource dispatching. To be more specific, the system management module will first score every node according to the load usage situation of processors and storages and the score of the node is positively correlated to resource allocation ratio, i.e., the higher the score, the more possible the allocation of new resources. Then, calculate the resource characteristic total score (TSC) of every node according to Eq. (1).

$$TSC_i = \frac{\sum_{j=1}^{n}\left(SC_{ij} * W_j\right)}{\sum_{j=1}^{n} W_j} \tag{1}$$

In Eq. (1), n represents the number of characteristic information of the node, and here $n = 2$; SC_{ij} represents the current score of the j-th characteristic; W_j represents the characteristic weight. Herein, the characteristic weight of the processor is set as 1 and the characteristic weight of the storage is set as 0.5. The system management module will record the scores of all the nodes in the scorebook and order the nodes according to the total scores, which not only facilitates the real-time consultation and invocation of back-stage management personnel or users but also helps with the rational allocation of tasks according to the scores. When a node receives a task, SSA algorithm is applied to lower the score of the node as shown in Eq. (2) and the score is regained by using SAA algorithm after the task is completed as shown in Eq. (3), in order to avoid repeated allocation of tasks.

$$DC_{ij} = X_{ij}\left[1-\mathrm{e}^{-\left(\frac{S_1+S}{1000}\right)}\right] - X_{ij}\left[1-\mathrm{e}^{-\left(\frac{S_1}{1000}\right)}\right] \quad (2)$$

In Eq. (2), X_{ij} represents the linear decrease weight of the *j*-th characteristic; S_1 represents the size of the file conveyed to the *i*-th node last time; S represents the size of the file to be conveyed. The present score is calculated according to $SC_{ij} - DC_{ij}$ and the score obtained is recorded in the scorebook in real time.

$$AC_{ij} = X_{ij}\left[1-\mathrm{e}^{-\left(\frac{S_2}{1000}\right)}\right] - X_{ij}\left[1-\mathrm{e}^{-\left(\frac{S_2-S}{1000}\right)}\right] \quad (3)$$

In Eq. (3), S_2 represents the present file size and S represents the file size after the last task is completed. The score is calculated according to $SC_{ij} + AC_{ij}$ and the new score obtained should be recorded in the scorebook in a timely manner. The system management module then functions according to the scores to realize the rational dispatching and allocation of resources.

4.3 Design Analysis of the Data Acquisition and Preprocessing Module

In the link of intrusion detection, the first thing to do is to collect the network data in cloud computing by virtue of the data acquisition and preprocessing module, during which the network data packet capture tool, namely, Winpcap is utilized which can be placed in all router switch systems. After the data acquisition and processing module is initiated, the acquisition submodule will begin to collect data (network data packets cut out or monitored by all nodes). After data collection is completed, relevant information (size, location, etc.) of such data should be transmitted to the system management module to facilitate the storage, calculation and allocation of detection data by the system management module. The procedure of data acquisition and preprocessing is as shown in Fig. 2.

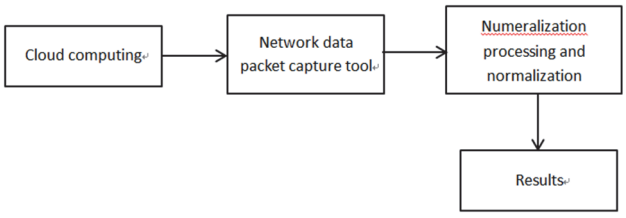

Figure 2 The procedure of data acquisition and preprocessing

In Fig. 2, Winpcap is a network access system under the Windows platform constituted of two dynamic linkbases and NPF filter. Compatible with Libpcap without rejection reactions, it is usually applied for monitoring and network scanning. Winpcap realizes network data packet capture in six steps: first, acquire relevant information of drivers and adapters of the local matcher by virtue of the Find function, open the adapter by virtue of the Open function and start the promiscuous mode to satisfy various data packets to achieve the goal of packet capture; second, after the adapter receives the data packets transmitted in the link layer, the adapter will send interrupt request order to the kernel in order to better process the data packets. During the interrupt processing, the interrupt program will conduct corresponding preprocessing operations; third, conduct filtration processing to the data packets during which Compile and Setfilter will be invoked to compile the filter equipment on the one hand and start the filter on the other hand; fourth, terminate the capture of data packets after the number of data packets captured reach the fixed set value; fifth, switch the format of the data packets to the format of files by virtue of Dump function and store them in the local file system. The data collected by Winpcap is somewhat different in terms of attributive characters, e.g. discrete data, continuous data, character string type data, etc. Different attributive characters need different processing methods and the difference restrains the detection processing of the detection module; therefore, normalization processing and numeralization processing are required for such data. As for discrete data and continuous data, HVDM is selected to conduct datamation processing, the relevant definition of which is as follows: setting that in network dataset S, every data X has n attributes, and $X_i\left(i=1,\cdots,n\right)$ is the value of the *i*-th attribute of data X, wherein $X_i\left(i=1,\cdots,m\right)$ is a continuous value and $X_i\left(i=m+1,\cdots,n\right)$ is a discrete value, then the distance between data X and data Y can be calculated by virtue of the formula below:

$$H\left(X,Y\right) = \sqrt{\sum_{i=1}^{n} d_i\left(X_i,Y_i\right)^2} \quad (4)$$

$$d_i\left(X_i,X_i\right) = \begin{cases} 1 & X_i或Y_i未知 \\ \dfrac{\left|X_i-Y_i\right|}{4\sigma_i} & if\ 第i个属性为连续值 \\ \sqrt{\sum_{c=1}^{C}\left|\dfrac{N_{iXc}}{N_{iX}}-\dfrac{N_{iYc}}{N_{iY}}\right|^2} & if\ 第i个属性为离散值 \end{cases} \quad (5)$$

X_i or Y_i is unknown
if the *i*-th attribute is a continuous value
if the -*i*th attribute is a discrete value

In the formula above, σ_i represents the variance of the value of the ith attribute, N_{iX} represents the number of the value X_i of the ith attribute of all datasets, C represents the number of all categories, N_{iXc} represents the number of output category c of the value X_i of the *i*-th attribute of all datasets. The formula of the normalization processing is as shown in Eq. (6).

$$X_i^{'} = \frac{d_i\left(X_i,X_i,\min\right)}{d_i\left(X_i,\max,X_i,\min\right)} \quad (6)$$

In Eq. (6), $X_i,\max$ and $X_i,\min$ represent the maximum value and the minimum value of the *i*-th

attribute of all datasets respectively. Character string type data can be processed by virtue of attribute mapping method, i.e. first to assume that the Protocol type has three different values, namely ICMP, TCP and UDP in character string type attribute characters; second to use mapping method to convert such data into numeric data and order the three types of data according to alphabetical order. If the Protocol type of the data is UDP, then the corresponding attribute value should be 3. Then conduct numeralization processing and normalization processing by virtue of the above formula to acquire the final detection data before the data is stored in accordance with the multi-dimensional vector method.

4.4 Design of the Detecion and Analysis Module

To ensure the safety of the cloud environment, the acquisition and preprocessing module of the intrusion detection system will conduct large-scale acquisition of network flow data packets for the detection module to detect and analyze. The analysis module should be efficient, prompt and precise in dealing with big data; otherwise the timeliness requirement will not be met. Cloud computing, due to its super-strong computing capacity, can effectively solve the computing problem and enhance the timeliness of detection of the intrusion detection system. Here, the detection and analysis module can provide corresponding detection method for the detection system, realize the analysis and judgment of sample data and lay the basis for the safety assessment of the cloud environment. The detection effect of the intrusion detection system is necessarily related to the analysis method, and here the PCA based clustering algorithm K-means (PCA-KM) is chosen as the detection and analysis method of the system after meticulous consideration because of its preferable clustering effect and easy implementation. It can effectively identify attack behaviors without machine learning and can be conveniently and easily applied.

4.5 Design Analysis of the Intrusion Tolerance Module

Usually, the network intrusion detection system will stop working in times of node failure, which will greatly restrain the work efficiency. To avoid the occurrence of such problems, the invasion tolerance module is introduced into the network intrusion detection system in cloud environment. In cases of physical aging problems or attacks in the cloud environment, the system can operate normally in the effect of the intrusion tolerance module. This module mainly utilizes two kinds of technologies in resisting attacks, one being the system redundancy technology and the other being the resource reallocation technology, which can ensure the normal service and operation of the system when the system is not subjected to catastrophic damages. The intrusion tolerance module is mainly composed of two parts, namely the monitoring module and the backup module. Here, the monitoring module monitors and tracks the state of cloud computing nodes in real time. When the performance of a certain computing node reaches the given lower limit value or when the node is about to collapse, then such node can be deemed as a failure node and the monitoring information at that moment will be automatically transmitted to the system management module. After receiving the information, the system management module will reallocate the tasks and renew the scorebook. The backup module is mainly used to backup data information and it has a backup sheet. When storing detection data, the system module will also transmit data copies to the backup module for storage to avoid data loss. Users can define the number of copies by themselves and there are three copies under normal conditions. Users can browse the backup sheet and know clearly the specific storage location of data and the number of copies. The specific workflow of the intrusion tolerance module is as follows: first, backup the data transmitted by users by virtue of the backup module. Second, monitor and track computing nodes by virtue of the monitoring module and transmit failure information to the resource dispatching module for it to reallocate tasks in cases of node performance collapse. Third, find the location of the failure node by virtue of the backup sheet, read the data information in this location and transmit such information to the system management module for it to reallocate tasks. Fourth, renew the backup sheet in real time.

5 SYSTEM TESTING

Testing research is carried out to examine the feasibility of the cloud computing based network intrusion detection system. The sample data used during the testing period comes from the KDDCup99 10% training dataset of Lincoln Laboratory. The dataset is stored in the tcpdump file format and mainly involves four types of attacks; first, local illegal permission elevation attack (U2R); second, remote attack (R2L); third, probing attack (Probe); and fourth, denial of service attack (DoS). To examine the feasibility of the system, the detection rates, the false alarm rates and the detection durations of the abovementioned four types of attacks are tested and analyzed and the specific sample numbers of every test subset are collected as shown in Tab. 1.

Table 1 Sample numbers of the test subsets

Test subset	Total sample number	Normal	U2R	R2L	Probe	DoS
U2R	10000	9948	52	0	0	0
R2L	20000	18886	0	1114		0
Probe	30000	26098	0	0	3902	0
Dos	40000	29910	0	0	0	10090

The test results of the four groups of test subsets are as shown in Tab. 2.

Table 2 Test results

Cluster ID	Total sample number	Cluster sample number	Normal	Attack	Detection rate / %	False alarm rate / %	Detection duration / s
Normal	10000	9520	9518	2	96.1	4.32	13
U2R		480	430	50			
Normal	20000	18214	18154	60	94.6	3.87	18
R2L		1786	732	1054			
Normal	30000	25432	25296	136	96.5	3.07	32
Probe		4568	802	3766			
Normal	40000	29368	29143	225	97.8	2.56	50
DoS		10632	767	9865			

It can be observed that the cloud computing based network intrusion detection system can effectively detect the four types of mainstream attacks, namely Probe, DoS, U2R and R2L, with detection rates all above 94%, the highest false alarm rate being only 4.32% and the longest detection duration being only 50 s. The preferable detection effect verifies the feasibility of system detection.

6 CONCLUSION

In general, this paper fist summarizes the security risks and system security requirements in cloud computing environment and points out the four major security requirements, namely the distributivity requirement, the effectiveness requirement, the scalability requirement and the timeliness requirement. Second, design analysis of the cloud computing based network intrusion detection system is conducted based on these requirements, including the design analyses of the IDS management module, the system management module, the data acquisition and preprocessing module, the detection and analysis module and the intrusion tolerance module. Last, the system is tested through empirical method and it is discovered that the system still exhibits sound detection effect when facing different types of attacks. For example, in its detection of U2R attack, the detection rate is 96.1%, the false alarm rate is 4.32% and the detection duration is 13 s, which testifies the feasibility of the cloud computing based network intrusion detection system.

7 REFERENCES

[1] Maheswari, M. & Karthika. R. A. (2021). A Novel QoS Based Secure Unequal Clustering Protocol with Intrusion Detection System in Wireless Sensor Networks. *Wireless Personal Communications*. https://doi.org/10.1007/s11277-021-08101-2

[2] Sari, T., Gules, H. K., & Yigitol, B. (2020). Awareness and readiness of Industry 4.0: The case of Turkish manufacturing industry. *Advances in Production Engineering & Management*, *15*(1), 57-68. https://doi.org/10.14743/apem2020.1.349

[3] Mangayarkarasi, R., Vanmathi, C., Vinayakumar, R., & Neeraj, K. (2021). An intrusion detection system using optimized deep neural network architecture. *Transactions on Emerging Telecommunications Technologies*, *32*(4). https://doi.org/10.1002/ett.4221

[4] Meng, J. L. (2021). Demand Prediction and Allocation Optimization of Manufacturing Resources. *International Journal of Simulation Modelling*, *20*(4), 790-801. https://doi.org/10.2507/IJSIMM20-4-CO20

[5] Zhong, M., Zhou, Y., & Chenx, G. (2021). Sequential Model Based Intrusion Detection System for IoT Servers Using Deep Learning Methods. *Sensors*, *21*(4). https://doi.org/10.3390/s21041113

[6] Balasundaram, J. & Pushpalatha, M. (2021). A novel optimized Bat Extreme Learning intrusion detection system for smart Internet of Things networks. *International Journal of Communication Systems*, *34*(7). https://doi.org/10.1002/dac.4729

[7] Support Vector Machines (2020). Researchers from University of Sharjah Describe Findings in Support Vector Machines (A Hybrid Anomaly-based Intrusion Detection System to Improve Time Complexity in the Internet of Energy Environment). Journal of Robotics & Machine Learning.

[8] Iranpak, S., Shahbahrami, A., & Shakeri, H. (2021). Remote patient monitoring and classifying using the internet of things platform combined with cloud computing. *Journal of Big Data*, *8*(1). https://doi.org/10.1186/s40537-021-00507-w

[9] Baral, M. & Mohan, V. A. (2021). Cloud Computing Adoption for Healthcare: An Empirical Study Using SEM Approach. *FIIB Business Review*, *10*(3). https://doi.org/10.1177/23197145211012505

[10] Mangalampalli, S., Swain, S. K., & Mangalampalli, V. K. (2021). Multi Objective Task Scheduling in Cloud Computing Using Cat Swarm Optimization Algorithm. *Arabian Journal for Science and Engineering*. https://doi.org/10.1007/s13369-021-06076-7

[11] Liu, C. H., Chen, T. L., Chang, C. Y., & Wu, Z. Y. (2021). A reliable authentication scheme of personal health records in cloud computing. *Wireless Networks*. https://doi.org/10.1007/s11276-021-02743-7

[12] Avinash, K., Pooja, G., & Manpreet, S. (2019). A Data Placement Strategy Based on Crow Search Algorithm in Cloud Computing. *Recent Advances in Computer Science and Communications*, *13*(1). https://doi.org/10.2174/2213275912666181127123431

[13] Mateen, A., Zhu, Q., Afsar, S., Rehan, A., Mumtaz, I., & Ahmad, W. (2019). Access Control Model for Data Stored on Cloud Computing. *International Journal of Advanced Culture Technology*, *7*(4).

[14] Tarawneh, S. A. & Abdulrahman, A. A. (2019). Analysis and Management of Risk Related to Using Cloud Computing in Business: Employee's Perception. *International Journal of Management (IJM)*, *10*(6). https://doi.org/10.34218/IJM.10.6.2019.003

[15] Kanimozhi, V. & Jacob, T. P. (2021). Artificial Intelligence outflanks all other machine learning classifiers in Network Intrusion Detection System on the realistic cyber dataset CSE-CIC-IDS2018 using cloud computing. *ICT Express*, *7*(3). https://doi.org/10.1016/j.icte.2020.12.004

[16] Krishnaveni, S., Sivamohan, S., Sridhar, S. S., & Prabakaran, S. (2021). Efficient feature selection and classification through ensemble method for network intrusion detection on cloud computing. *Cluster Computing*. https://doi.org/10.1007/s10586-020-03222-y

[17] Elmasry, W., Akbulut, A., & Zaim, A. H. (2021). A Design of an Integrated Cloud-based Intrusion Detection System with Third Party Cloud Service. *Open Computer Science*, *11*(1). https://doi.org/10.1515/comp-2020-0214

[18] Rahayuningsih, P. A., Rahayuningsih, P. A., Maulana, R., Irmayani, W., Saputra, D., & Purwaningtias, D. (2020). Feature Dependent Naïve Bayes for Network Intrusion Detection System. Journal of Physics: Conference Series, *1641*(1). https://doi.org/10.1088/1742-6596/1641/1/012023

[19] Ashfaq, K. M. & Juntae, K. (2020). Toward Developing Efficient Conv-AE-Based Intrusion Detection System Using Heterogeneous Dataset. *Electronic*s, *9*(11). https://doi.org/10.3390/electronics9111771

[20] Mahfouz, A., Abuhussein, A., Venugopal, D., & Shiva, S. (2020). Ensemble Classifiers for Network Intrusion Detection Using a Novel Network Attack Dataset. *Future Internet*, *12*(11). https://doi.org/10.3390/fi12110180

[21] Latah, M. & Toker, L. (2020). An efficient flow-based multi-level hybrid intrusion detection system for software-defined networks. *CCF Transactions on Networking*, *3*(3-4). https://doi.org/10.1007/s42045-020-00040-z

[22] Zeeshan, A., Adnan, S. K., Cheah, W. S., Johari, A., & Farhan, A. (2020). Network intrusion detection system: A systematic study of machine learning and deep learning approaches. *Transactions on Emerging Telecommunications Technologies*, *32*(1). https://doi.org/10.1002/ett.4150

Contact information:

Li YOU
The information center,
Hebei Vocational University of Industry and Technology,
Shijiazhuang, China
E-mail: yolier@126.com

Zhanyong WANG
(Corresponding author)
Admissions and Career Guidance Center,
Hebei Vocational University of Industry and Technology,
Shijiazhuang, China