Research on Network Intrusion Detection Method Based on Deep Learning Algorithm

Xinhai Liao¹, Jianguo Xie²*

¹Network and Informatization Center, Guangdong, University of Foreign Studies, Chenzhou, Hunan 510420, China
²Network and Informatization Center, Guangdong, University of Foreign Studies, Qiyang, Hunan 510420, China

*Corresponding author: caobaoyi12@126.com

Abstract. Information security is very important in enterprises. Under the tide of modern network informationization, especially intrusion detection is facing great challenges. The rise of technologies such as big data and artificial intelligence, on the one hand, helps to strengthen information security, on the other hand, it also brings great challenges to information security. Everything has two sides, and the Internet is no exception. While the network has brought us a richer and faster life, the problem of network security has become increasingly prominent, and network security risks pose new challenges to the fields of politics, economy, culture, national defense, ecology and society. The active defense strategy using intrusion detection technology can actively discover attacks when they occur or attempt to attack, realize real-time response, and provide real-time protection for network systems. This paper starts from the field of intrusion detection and combines deep learning technology to design an efficient and intelligent intrusion detection model which can be used in the field of network intrusion detection.

Keywords: Intrusion detection; network system; information security

1. Introduction
In the 21st century, with the rapid development of computer software and hardware technology, more and more individual and enterprise users upload private data to the Internet, which brings convenience to users' life and work [1]. However, due to the continuous emergence of various network attacks, there is also the risk of leakage of user sensitive privacy data [2]. With the development of the network, network attack events emerge in endlessly, network attack means are diverse, and now many attack tools are automated, which has brought great harm to our work and life [3]. With the increasing complexity and diversification of intrusion data, the model has problems of high false positive rate, high false negative rate, poor adaptability and low detection rate [4]. In today's big data era, the scale of data flow is becoming larger and larger, and the way of intrusion has become more complex. Traditional intrusion detection technologies such as artificial immune detection method and intrusion detection method based on information theory have many shortcomings [5]. Network security detection not only needs to detect the intrusion behavior that has occurred, but also needs to monitor
the unsuccessful intrusion attempts, so as to provide the necessary information to defeat the intrusion events in time [6]. With the development of artificial intelligence, the application of machine learning algorithm in the field of intrusion detection not only realizes the intelligent detection of network attacks, but also effectively improves the efficiency of intrusion detection, and reduces the rate of missing and false positives [7].

Early intrusion detection systems, such as expert system, can only carry out simple defense against some known network attacks [8]. For users using the Internet, the various services provided by networked computers are not only a high-tech tool or a quick way to obtain information, but also an indispensable part of life, and the Internet has become a new way of life [9]. Compared with the early intrusion detection system, the intrusion detection system based on traditional machine learning has better performance and performance, but with the increasingly complex network environment, the huge and high-dimensional network data makes the traditional machine learning method subject to time and space complexity constraints, resulting in the decline of detection accuracy [10]. Due to the huge data volume and high data dimension of the original data set of network intrusion detection, the traditional intrusion detection technology based on machine learning largely depends on the data dimension reduction method. A good dimension reduction method will directly affect the performance of the intrusion detection model based on machine learning [11]. At present, most of the existing machine learning and deep learning intrusion detection systems are trained by feature engineering, which makes the system heavily rely on human experience and affects the continuous improvement of system performance [12]. Starting from the field of intrusion detection, combined with deep learning technology, this paper designs an efficient and intelligent intrusion detection model which can be used in the field of network intrusion detection.

2. Requirement analysis of intrusion detection system

Intrusion Detection System (IDS) is a security management system for monitoring malicious events in network or computer. It continuously monitors network traffic to find signs of possible system attacks or abnormal behaviors violating security policies in system activities, and transmits the generated system logs to the management unit, so that the system can respond and deal with intrusion or attack behaviors in time, thus ensuring the availability, integrity and confidentiality of the system. It is the key index to judge whether an intrusion detection system is excellent or not to find the suspicious intrusion in the host or network effectively. The correct rate of malicious intrusion detection by intrusion detection system is directly related to the key technologies of intrusion detection. If the data packet cannot be captured and analyzed by the intrusion detection system in real time, the system may miss the information and cause network congestion [13]. Because the intrusion detection system can't get the traffic packets of the whole network in real time, it may lead to some malicious attacks that can't be detected effectively, resulting in false negatives. With the continuous development of intrusion detection technology, various intrusion detection systems have been developed. Because different detection systems focus on different functional points, their own system architecture has also been adjusted.

Existing intrusion detection models face some problems in the detection process. In the process of network access, each network connection has multidimensional characteristics. Considering that the amount of network data and the feature dimension of each network connection are too large, in the process of intrusion detection, the intrusion detection model will select some features from the features of network connections, analyze and extract rules to realize intrusion detection. The core function of intrusion detection system is to actively find suspected intrusion behavior. Monitor the network traffic in the system in real time, find out the abnormal traffic and check it in time to find possible intrusion attempts, so as to avoid intrusion. Under the condition of complete data, the existing intrusion detection model has been able to achieve good results. However, in practical applications, the characteristics of network connections are often in an incomplete state. Before the intrusion, it actively monitors the security status of the system and uses the early warning system to stop the suspicious intrusion in real time. When invasion is happening, it takes active measures to deal with malicious
invasion and reduce the losses caused by malicious invasion. The research on the reliability of network topology needs to be analyzed from the aspects of network devices and traffic load. The algorithm flow is shown in Figure 1.

![Algorithm flow](image)

**Figure 1** Algorithm flow

With the continuous adoption of new technologies and detection methods, and paying more attention to the real-time and effectiveness of the system at the application level, the detection performance and efficiency of the intrusion detection system are getting higher and higher. Intrusion detection system is different from other network security systems in that it is an active security defense mechanism. All intrusion detection systems follow a common framework, which consists of data collection module, analysis and detection module, data configuration module, data storage module, reference data module and alarm module, etc. Because the traffic of the whole network can not be monitored in real time, the intrusion detection system has been occupying the requested resources, thus affecting the flow rate of the host or network and causing unnecessary network congestion. Therefore, the intrusion detection system needs good detection rate and real-time monitoring ability, so as to realize the monitoring and analysis of users and system activities. Intrusion detection method based on neural network mainly simulates the thinking process of human brain, constructs neural network learning model, uses normal behavior samples and illegal behavior samples to train and learn continuously, and finally realizes independent judgment, and can identify new intrusion behavior or variant forms of known intrusion behavior with a certain probability. Real-time and effective communication among modules of intrusion detection system is the basis of normal and effective work of intrusion detection system. Only through real-time and effective communication between modules can each module work together and make the whole system run safely and effectively.

3. **Intrusion detection algorithm model implementation and platform integration**

The purpose of network intrusion detection is to find suspicious intrusion behavior, and take corresponding measures to protect the network from continuous attacks and reduce the economic losses caused by it. The important task of network intrusion detection is traffic classification detection, which analyzes some key fields in traffic packets to judge whether the network is attacked or not. The model has the ability of self-evaluation of classification results, which breaks through the limitation of closed set classification and is suitable for the open set classification problem of unknown attack
detection. Be responsible for integrating and sorting out existing data sets, or collecting traffic data generated by network and host in daily activities, and providing training and testing data sets for subsequent intrusion detection models. Because of the complex network environment, the traffic data obtained by the data acquisition module are of different sizes, which is not conducive to the subsequent feature learning, so standardized measures should be taken and unified rules should be set [14]. The main difference between incomplete information and complete information lies in the reduction of data feature quantity, which will greatly affect the accuracy of intrusion detection model. Traditional intrusion detection models are difficult to adapt to the rapid changes of today's network. In the face of high-dimensional and new attack types of network traffic data, there are often problems such as high false alarm rate, low accuracy and weak adaptability. The flow information collected by the data acquisition module is preprocessed accordingly, and data cleaning and data normalization are performed according to the needs of the model to extract the original flow information containing effective features.

Similar to human neural network, the basic component of artificial neural network is neuron, and a simple neuron structure is shown in Figure 2.

![Figure 2 Simple neuron](image)

The neuron has three inputs $x_1, x_2, x_3$, where $+1$ represents the bias term, input to the arithmetic unit $f$, and the entire neuron output is:

$$ h_{w,b}(x) = f(W^T x) f\left(\sum_{i=1}^{3} W_i x_i + b\right) $$

The processed gray image is input into the model training module, and the appropriate network is selected for relevant model training to extract the network flow characteristics, which makes the subsequent detection more accurate. The analysis of packets shows that the payload length of each packet is not equal, so in order to effectively use the original flow data to train the model. These data include signals with local correlation, signals expressed in one-dimensional or multi-dimensional arrays, targets that do not change due to slight deformation, and features that can appear at any position. On the training data set, correctly classified data and incorrectly classified data can be regarded as known types and unknown attacks respectively. The purpose of data cleaning is to clear the redundant and illegal data in the original flow data, and to effectively clean the redundant information and reduce the correlation between the same batch of data, which can facilitate the subsequent feature learning and extract the features in the original flow more effectively. MLP can be simulated as any complex function, and its complexity depends on the number of neurons in the input layer and the number of hidden layers.

After the intrusion, it collects the intrusion activity data generated by the intrusion and updates it to its own database. Misuse detection analyzes the known attack behavior and extracts the corresponding attack characteristics, establishes the corresponding attack signature database, and determines whether the suspected behavior is an attack behavior by matching the attack pattern database. Intrusion detection system monitors network traffic and computer log information, analyzes relevant data, and finds out whether there is malicious attack. Fig. 3 is the flow chart of intelligent video anomaly detection system.
The signal processing of intrusion characteristic data uses cloud computing network foundation to describe communication intrusion data through $\omega$ and $v$. Among them, $\omega$ represents the horizontal domain vector of the communication network intrusion data, and $v$ represents the vertical domain vector of the communication network intrusion data. $\alpha$ represents the initial filtering result of the intrusion feature data, then $\alpha$ is expressed as:

$$\alpha = \sqrt{W \cdot s(\omega)} + s(v) + m \quad (2)$$

Among them, $W$ represents the norm vector of the intrusion signal, $s(\omega)$ represents the norm coefficient of the horizontal domain vector, $s(v)$ represents the norm coefficient of the vertical domain vector, and $m$ represents the initial filtering constant. The signal processing result of the intrusion feature data can be expressed as:

$$R = W \cdot 2\zeta(n) \cdot \omega \quad (3)$$

In the formula, $\zeta(n)$ represents the superimposed signal processing result of the intrusion communication data.

The improved algorithm uses two fully connected layers, and when the first fully connected operation is performed, the partial deletion operation is performed to randomly deactivate some neurons in the fully connected layer. By doing so, we can ensure that the feature over-fitting phenomenon will not be caused when extracting spatial features, and at the same time, the computational overhead of neural network will be reduced. Each layer protocol obtains the required data according to the header information in the frame, and finally delivers the processed frame to the target. The reality is that the attack process is detected. If the attack is audited after the attack, the subsequent timeliness must be guaranteed, because the system is likely to face similar attacks in the future [15]. Timeliness avoids the inconvenience and technical limitations of administrators when auditing system logs to find intruders or intrusion prompts. Anomaly detection can judge whether the suspected attack behavior deviates from the normal behavior pattern by constructing the normal behavior pattern generated during the user's use of the host and the network. The application scenario of big data is very different from that of relational database. Usually, after the data source of big data is generated and saved, the data set needs to be classified and queried, but not deleted and modified. This data access scenario is a typical one-time write. Check the system log files and monitor the operation of the system, and check whether the key files of the system are missing, so as to judge whether there is abnormal network activity. If there is abnormality, record the relevant data generated by the attack behavior and save it in the knowledge base to avoid similar attacks again.

4. Conclusions
With the development of the Internet, there are more and more network security incidents, and intrusion detection technology with active security defense can effectively resist network attacks. Using machine learning method to detect network traffic anomaly is the mainstream intrusion detection method at present, but the design of feature set is highly dependent on experience and the increase of space-time complexity caused by massive high-dimensional data, which seriously limits the performance improvement of intrusion detection system. Applying deep learning technology to the field of network intrusion detection not only realizes the intelligent detection of the network, but also improves the detection efficiency. Deep learning has made a breakthrough in the field of image processing, and image classification is very similar to abnormal traffic classification. At the same time, the deep learning method has the ability of representation learning, and can learn the original traffic end-to-end, thus avoiding the loss of original feature information caused by artificial feature sets.
When the event reaches the alert value, identify the threat, make a decision or send out an alarm message. The learning of neural network has obvious advantages in identifying threats, which makes the application of neural network technology in intrusion detection system possible.

References
[1] Yang Jie, Tang Yachun, Tan Daojun, Liu Xiaobing. Multi-channel autoencoder deep learning intrusion detection method. Computer Science and Exploration, vol. 147, no. 12, pp. 74-84, 2020.
[2] Gao Ni, He Yiyue, Gao Ling. A deep learning method for intrusion detection in a massive data environment. Computer Application Research, vol. 35, no. 4, pp. 1197-1200, 2018.
[3] Wang Rui, Li Xiaofeng, Shi Tianyun, et al. Research on Railway Perimeter Intrusion Detection Algorithm Based on Video Deep Learning. Transportation System Engineering and Information, vol. 20, no. 2, pp. 65-72, 2020.
[4] Di Chong, Li Tong. Deep learning method for unknown network attack detection. Computer Engineering and Application, vol. 965, no. 22, pp. 115-122, 2020.
[5] Liu Haiyan, Cong Fei. Overview of network intrusion anomaly detection based on deep learning. Information System Engineering, vol. 321, no. 9, pp. 52-53, 2020.
[6] Yu Lin. Research on Intrusion Detection Based on Deep Belief Network. Computer Science and Application, vol. 8, no. 5, pp. 687-701, 2018.
[7] Zhou Luming, Zheng Mingcai. Research on Network Intrusion Prevention Technology Based on Deep Learning. Microcomputer Application, vol. 331, no. 11, pp. 97-101, 2020.
[8] Ye Song. Research and implementation of deep learning application protocol recognition technology based on modern networks. Software Guide, vol. 17, no. 10, pp. 194-199, 2018.
[9] He Wenyu, Yang Jie, Zhang Tianlu. Orbital foreign body intrusion detection algorithm based on deep learning. Computer Engineering and Design, vol. 408, no. 12, pp. 84-91, 2020.
[10] Wei Mingjun, Peng Ning. Research on intrusion detection methods based on DL and TSVM. Computer Applications and Software, vol. 37, no. 12, pp. 334-339, 2020.
[11] Chen Guang, Han Weizhan, Zhang Wenzhi. Encrypted traffic classification and intrusion detection based on deep learning. Computer measurement and control, vol. 256, no. 1, pp. 59-65, 2020.
[12] Sun Huili, Chen Weihua, Liu Dongchao. Improved Bayesian network intrusion detection algorithm based on deep learning. Software Engineering, vol. 22, no. 4, pp. 17-20, 2019.
[13] Zhu Jun, Wu Peng. Research on Intrusion Detection of Shipbuilding Industry Network Based on Deep Learning. Ship Science and Technology, vol. 42, no. 9, pp. 186-188, 2020.
[14] Yan Chunjiang, Wang Chuang, Fang Hualin, et al. Intrusion detection of transmission line engineering vehicles based on deep learning. Information Technology, vol. 320, no. 7, pp. 36-41+46, 2018.
[15] Liu Jinghao, Sun Xiaowei, Jin Jie. Intrusion detection model based on principal component analysis and recurrent neural network. Journal of Chinese Information Processing, vol. 34, no. 10, pp. 109-116, 2020.