Detection of Fuzzy Clustering Anomaly Intrusion Behavior Based on Feature Selection in Web Big Data Environment

Hongyan Diao

Wuxi South Ocean College, Wuxi, Jiangsu, 214081, China

Keywords: Fuzzy Clustering; Abnormal Intrusion; Computer Network Security; Web Big Data

Abstract: With the continuous development of modern modern society, information technology has been more and more widely used. In the current daily life, network information has become the main channel for the public to obtain information, and in the future, the proportion of the channel It will be even more important, but it is precisely because of this that the construction of information security is becoming more and more important. In the face of malicious attacks by lawless elements, effectively improving the security defense methods of the network and establishing an efficient method for detecting abnormal behavior becomes a computer. An important project in the field of network security development, this paper discusses some of the popular fuzzy clustering anomaly intrusion detection [1].

1. Introduction

In the process of network security construction, it is like a contest between network hackers and network security experts. On the one hand, hackers invade through various illegal means to obtain damage and store various data resources in an unauthorized environment. In order to threaten the security construction of the system, on the one hand, the network security construction personnel detect and defend the illegal intrusion means through various technologies, and effectively ensure the security inside the network. In the long-term battle between the two sides, various hacker technologies emerge one after another, and network security construction experts have also developed a relatively mature and perfect defense method system[2]. This is the intrusion detection system. By monitoring the abnormality of network data in real time, the system judges whether the attack is abnormally attacked, and effectively identifies the path and mode of the attack, thus opening the relevant protection tools. In general, the current intrusion detection The system has the following basic features: first, it can monitor and analyze the activities of users and systems; second, it can check the configuration and vulnerabilities of the system; third, it can effectively identify behavioral anomalies and promptly report alarms; fourth, can Statistical analysis of abnormal attacks; Fifth, the integrity of internal data can be evaluated and protected in a timely manner; sixth, audit management of the operating system can be performed, and user activities can be safely rated. The detection of the system is mainly reflected in the direction of usability, real-time, security and scalability[3].

In the detection of intrusion behavior, two methods of misuse detection and anomaly detection have been developed[4]. The detection of misuse detection is earlier. It is a mode expression directly related to intrusion behavior, but this method Detection of unknown attacks is not ideal. The method of abnormality detection is to establish a normal operation mode, and mark behaviors other than the normal operation mode as abnormal behaviors, thereby judging whether or not the attack is performed. This method does not need to identify and classify each abnormal mode, so It is very good defense against unknown attacks, but this method needs to be built on a large amount of statistical data and the false positive rate is high. However, with the development of modern computer information technology, the data acquisition and processing speed has been significantly improved[5]. Therefore, the advantage of this method based on data statistics is increasingly obvious, and its defense efficiency is in obvious improvement. The detection of fuzzy clustering anomaly intrusion based on Web big data environment has become an important research direction of the current network security industry[6].

DOI: 10.25236/mmmce.2019.086

2. The Principle Analysis of Fuzzy Clustering Abnormal Intrusion Behavior Detection

According to the characteristics of computer data mining and machine learning, the similarity of data objects is compared and divided into classes and clusters. Clustering is the core idea of building M×N matrices. For some actual data fluctuation range, M and N have a certain relationship most of the time, but when the relationship between M and N is abnormal and is greater than a certain critical value, the data is said to have an abnormality to describe the abnormality. The quantity of the relationship usually selects one or several of the interval scale variable, the binary variable, the categorical variable and the ordinal variable, wherein the interval scale variable has several kinds of Euclidean distance, Manhattan distance, and Minkowski distance. Algorithms, algorithms for other variables are algorithms based on conventional mathematical definition methods. These data variables are the main selection points for identifying abnormal quantities in anomaly detection[7].

At present, there are several main classifications, such as hierarchical methods, partitioning methods, density-based clustering algorithms, grid-based methods and model-based methods. Different methods have different monitoring advantages[8]. When selecting a method, it is necessary to make a reasonable choice based on the actual anomaly detection requirements.

3. Design of Fuzzy Clustering Anomaly Intrusion Behavior Detection Based on Feature Selection in Web Big Data Environment

3.1. Overview of the system model

The general structural model of anomaly intrusion detection based on Web big data is shown in Figure 1:

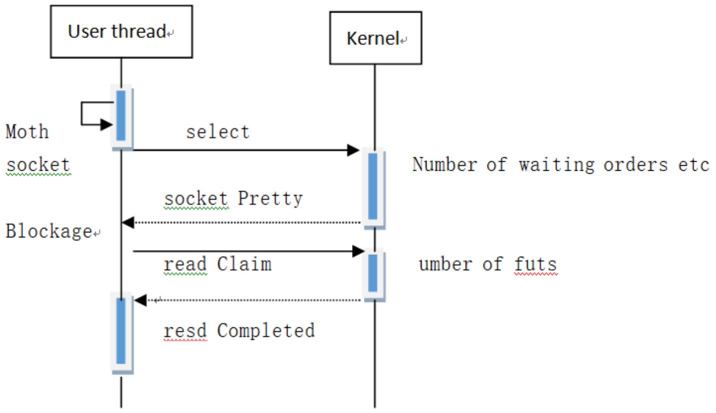

Fig.1. Intrusion detection system model based on feature selection and clustering (as shown)

When referring to the model first, it is found that in the monitoring and analysis of network traffic data, the combination of traditional misuse monitoring and abnormal clustering monitoring is adopted[9]. Although misuse detection is difficult to identify unknown attacks, it is familiar in the face. When attacking, it has the characteristics of fast matching and accurate identification. Therefore, it can still be retained in the construction of large-scale anomaly detection[10]. When cluster analysis is performed, on one hand, abnormalities can be checked and alarm responses can be timely. On the one hand, the rule extraction is integrated into In the database, this can improve the recognition ability of the computer in the face of the same attack. This clustering analysis extraction mode for the computer to join the learning method can make the system defense ability grow continuously, which is a great advantage of cluster analysis anomaly detection[11].

3.2. System model construction process

Under the guidance of the above system model, analyze the difficulties in the construction of each node, and gradually solve the main methods of establishing the monitoring system at this time. Pay attention to the following links in the construction focus:

1) In data collection, to capture and analyze the rules before the data enters the application, the current data capture tool used by Win32 is WinPcap, which can realize capture and open source analysis at speed, and provides two under WinPcap. The libraries, namely packet.dll and wpcap.dll, are responsible for capturing the underlying data interface and the functions of the upper layer to ensure that the packet capture behavior can be completed independently of the network hardware and operating system. In the process, the initial processing of filtering and saving to the heap file can be completed by the provided function.

2) After the data capture process is completed, the data is analyzed immediately. The commonly used analysis softwares are ethereal, Snort, TCPdum, etc. In the use of small TCP/IP networks, Snort is commonly used for analysis and recording, in content extraction. On the record, the main record content of the analysis layer is connected content features, network connection features and statistical features[12].

3) Data standardization processing. The main goal of this session is to standardize the data of different types of data, and pay attention to the two types of data in different processing methods when facing discrete and continuous data.

4) Cluster analysis.

The final task of clustering analysis is to identify whether it is an abnormal event by judging the clustering result[13]. In the process, the class in the clustering algorithm is first identified, which is convenient for providing a comparison template for subsequent data to be detected. In the actual identification classification process, the number of monitoring results entering normal data is much larger than the monitoring result of abnormal data. In order to reduce the probability of false positives and then set the threshold of the data, it is necessary to find the appropriate number of clusters or The soft segmentation method is used to improve the accuracy of clustering results[14]. When extracting the training set data, the proportion of normal data should be consciously increased, and the normal data should be balanced as much as possible in each type of proportion, which can greatly reduce the distance between the centroids of some rare normal data to the data clusters in the training set. Thereby reducing the chance of being falsely reported[15].

4. Conclusion

Through the above cluster analysis, it is found that in the actual system construction, in order to improve the accuracy of monitoring, it is necessary to effectively improve the data types in the monitoring training set. In order to achieve more comprehensive monitoring, the workload of the comparison data will be Different algorithms increase the amount of calculation greatly, which affects the efficiency of monitoring. In a system with slow data calculation, it seems that the accuracy and response speed cannot be both, but from the perspective of the actual level of technology and software development, The computing speed of the current computer is getting higher and higher. On the other hand, through the update of the software calculation method by scholars, the calculation method of the whole anomaly detection will become more and more scientific, and the response speed will be faster and faster, so the field The development prospects are enormous. In the future, through the continuous improvement of technologies such as tracking and anti-intrusion, the security construction of the entire network system will be more secure. However, as the means of the attackers are constantly changing, the tasks of the relevant network security construction scholars are still very arduous. Therefore, in the construction of the network, we continue to improve and repair, and constantly respond to the attackers. Overall, the network is now step by step.

References

[1] Li Chao, Tian Xinguang, Xiao Xi, et al. User behavior anomaly detection method based on Shell command and symbiotic matrix. Journal of Computer Research and Development, 2015, (9): 1982-1990.

[2] Rose K., Gurewitz E. Constrained clustering as an optimization method. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 8(8): 785-794.

[3] Zhang Xinbo. Two-stage fuzzy C-means clustering algorithm. Journal of Circuits and Systems, 2015, (2): 117-120.

[4] Ingunn Berget, Bjorn-Helge Mevik, Tormod Naes. New modifications and applications of fuzzy C-means methodology. Computational statistics & data analysis, 2018, 5(5): 2403-2418.

[5] Wang Zhanhu, Liu Zhijing, Chen Donghui. Research on Fuzzy C-Means Clustering Algorithm Based on Particle Swarm Optimization. Computer Science, 2015, (9): 166-169.

[6] Chen You, Cheng Xueqi, Li Yang, et al. Lightweight Intrusion Detection System Based on Feature Selection. Journal of Software, 2017, (7): 1639-1651.

[7] Mehdi Karimi-Nasab, Ioannis Konstantaras. A random search heuristic for a multi-objective production planning. Computers & Industrial Engineering, 2012, 2(2): 479-490.

[8] Chen You, Shen Huawei, Li Yang, et al. An efficient feature selection algorithm for lightweight intrusion detection systems. Chinese Journal of Computers, 2017, (8): 1398-1408.

[9] Kai-Quan Shen, Chong-Jin Ong, Xiao-Ping Li, et al. Feature selection via sensitivity analysis of SVM probabilistic outputs. Machine learning, 2018, 1(1):1-20.

[10] JUNGSUK SONG, KENJI OHIRA, HIROKI TAKAKURA, et al. A Clustering Method for Improving Performance of Anomaly-Based Intrusion Detection System. IEICE transactions on information and systems, 2018, 5(5): 1282-1291.

[11] Jengnan Tzeng, Wen-Liang Hwang, I-Liang Chern. Enhancing image watermarking methods with/without reference images by optimization on second-order statistics. IEEE Transactions on Image Processing, 2015, 7(7).

[12] Liu Zhu, Li Zhonghai, Wang Dingwei. Adaptive adjustment of image uneven gray scale. Control Engineering, 2015, (3). doi: 10.3969/j.issn.1671-7848.2003.03.018.

[13] Rushing J.A., Ranganath H., Hinke T.H., Et al. Image segmentation using association rule features. IEEE Transactions on Image Processing, 2015, 5(5).

[14] Scheunders P.. A multivalued image wavelet representation based on multiscale fundamental forms. IEEE Transactions on Image Processing, 2016, 5(5).

[15] Shyh-Shiaw Kuo, Johnston J.D. Spatial noise shaping based on human visual sensitivity and its application to image coding. IEEE Transactions on Image Processing, 2018, 5(5).