Research on Intrusion Detection System based on Data Mining

Jing Liang

Chongqing Vocational Institute of Safety & Technology, Chongqing 404020, China.

513001084@qq.com

Keywords: Data mining, intrusion detection system, network security.

Abstract: At present, Internet technology is in a period of rapid development, and people begin to use intelligent devices based on Internet technology frequently in their daily life. More and more people begin to realize the importance of information security. At the same time, the issue of information security has become one of the major research scholars continue to explore and think about. With the progress of science and technology, people gradually found that the traditional security protection technology and firewall technology have been outdated, it has been unable to meet the needs of the current people. Therefore, at present, there is an urgent need for a system that can find security problems in the shortest possible time and report them automatically, which is the intrusion detection system to be discussed in this paper, which can actively detect malicious attacks on the system by a variety of unauthorized objects. Moreover, it can also monitor whether the authorized object has illegal operation behavior, and can organize the intrusion behavior in time. In this paper, the definition and characteristics of data mining and intrusion detection system are described in detail. After that, this paper analyzes the problems existing in the current intrusion detection system, and points out the function and importance of intrusion detection system based on data mining. The author hopes that this paper can give some reference to those scholars who study intrusion detection system, and also hope to contribute to the development of intrusion detection system in the future.

1. Introduction

At present, with the development of Internet technology, the computer penetration rate is also greatly improved, a variety of new online services are also deep into people's daily life, therefore, how to ensure people's information security has become one of the hot topics of social discussion. In the past, most of the security protection technologies used authentication mechanism, authorization mechanism, access control mechanism, encryption mechanism and so on. However, these cannot prevent some unauthorized objects from taking advantage of the defects of the computer software and hardware system to invade the computer system, and it is difficult to prevent a variety of illegal operations carried out by authorized objects. With the development of society, more and more people begin to use intelligent devices based on Internet technology frequently. This also shows that if there is no scientific and efficient system to ensure the security of people's daily information data, then there is a great probability that it will lead to a variety of serious problems. Therefore, intrusion detection system is born, it is a kind of system which can detect unauthorized object intrusion system and authorized object to carry on abnormal behavior in time. It can not only detect the malicious intrusion of unauthorized objects to the system, but also detect the illegal operation of authorized objects to the system. However, only relying on intrusion detection system is far from enough, the current era is the era of big data. In this era, there is a huge amount of data information, so if the intrusion detection system is only to detect all the objects, then it is a very difficult and time-consuming task. With the emergence of big data technology, people began to find that in order to find meaningful information from a large number of data, we must use big data technology in data mining technology. The so-called data mining technology is to find useful data in a large amount of data.

However, although the intrusion detection system is very powerful, it is not enough to use intrusion detection system to prevent computers from being illegally invaded. In other words, people must combine traditional technical means with intrusion detection systems, so that computers can not only have firewalls, secure routers and other defense means. It can also have the real-time monitoring of intrusion detection system and the means of attack and anti-attack. Therefore, at present, the use of intrusion detection system is necessary and important, but also the needs of the current Internet era. On the basis of intrusion detection system, the fusion of data mining technology can greatly improve the efficiency of intrusion detection system, so as to have better detection and protection ability.

2. Theoretical Basis

2.1 Data Mining

Data mining is an extension of artificial intelligence technology, which can extract all kinds of useful information from massive data. It can not only accurately find the past data, but also find out the relationship between the current data and the past data, so as to provide people with all kinds of useful information. If you divide it from a functional point of view, there are four main analysis methods of data mining:

First, correlation analysis. Association analysis, according to the principle of association, to mine the relationship between those hidden in the data. Among them, the principle of relevance can scientifically and efficiently reflect the interdependence and relevance between things and other things. If there is a certain connection between this thing and other things, then people can understand and even predict other things through one of them [1].

Second, sequence pattern analysis. Sequential pattern analysis refers to the connection of the time sequence between the data, so as to mine the time association between the data.

Third, classified analysis. The main purpose of classification analysis is to construct the classification model and explore the classification rules according to some specific data.

Fourth, cluster analysis. Cluster analysis inputs a group of unclassified data, and people do not know that these data can be divided into several types before outputting the results. Through cluster analysis, people also need to follow certain classification rules, and these classification rules are determined by cluster analysis tools. Finally, people need to classify these data reasonably, determine the category of the data, and make the difference between different categories of data as large as possible, while the difference of the same category of data is small [2].

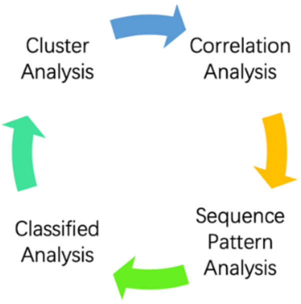

Fig. 1 Four Main Analysis Methods of Data Mining

2.2 Intrusion Detection System

Intrusion detection system, abbreviated as IDS, which is mainly to monitor the running status of the computer system. When various attempts to invade a computer system or attack a computer system are found, the system will report the computer system in time and ensure the security of the internal data of the system. In the past, the main means of defense for computers was firewalls [3]. However, with the development of the times, the ability of intrusion objects is getting stronger and stronger, and the defense function of firewall has been greatly weakened. Therefore, only relying on the firewall to protect the computer system is far from enough. At the same time, IDS can be divided into two categories, namely, network-based intrusion detection system and host-based intrusion detection system. The former is mainly to listen to all the data packets in the network, if we find those packets with attack characteristics, and then take precautions against them. The latter is mainly to detect the internal data information of the system, if those data with attack characteristics are found, then take precautions against them. At present, most of the intrusion detection systems have the functions of these two kinds of intrusion systems, which can prevent the illegal intrusion in the network and within the system at the same time.

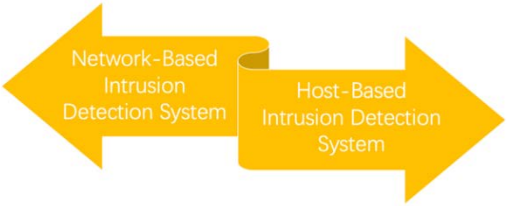
Network-Based Intrusion Detection System
Host-Based Intrusion Detection System

Fig. 2 Two Categories of Instrusion Detection System

At present, there are mainly two main intrusion detection methods, one is misuse detection, the other is anomaly detection. These two techniques greatly identify objects whose attack behavior is quite different from that of normal behavior. At present, the main method of misuse detection is pattern matching, which mainly compares the collected data with the various patterns in the predetermined feature knowledge base [4]. If it finds that there is a great difference between the behavior and the normal behavior, then it is judged that the object has the characteristic of attack. However, anomaly detection is not so, it is mainly judged from the behavior, it constructs the model of the normal behavior, and its hypothetical attack behavior is different from the normal behavior. Therefore, if it finds that the behavior is different from the normal behavior, it is judged that there is an attack, even if the difference between the two is very small. With the development of the times, the current anomaly detection method is mainly to set the limit threshold on the model, and compare the detected data with the normal behavior. If the normal limit threshold is exceeded, then the behavior is judged to be an intrusion behavior [5].

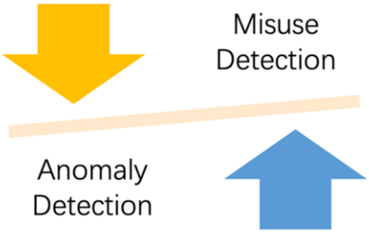

Fig. 3 Two Intrusion Detection Methods

The advantage of misuse detection is that it can accurately identify the type of attack, can build a perfect intrusion detection system scientifically, and its accuracy is incomparable to other detection methods. Its disadvantage is that it is difficult to detect new types of attack behavior, it is not sensitive to all kinds of deformation of attack behavior, and its database update lags behind the generation of attack type. The advantage of anomaly detection is that it does not need to input the attack features into the database, nor does it need to have perfect system knowledge [6]. It can detect not only known attacks, but also unknown attacks. Its adaptability is incomparable to other tests. However, there are still shortcomings, that is, it is difficult for people to set the limit threshold well, therefore, in the process of using anomaly detection, the error rate is also high.

3. Problems Existing in Intrusion Detection System at Present

Most of the current intrusion detection systems, whether host-based intrusion detection system or network-based intrusion detection system, mainly rely on manual operation. At the same time, because the network attack behavior has the high complexity, the expert knowledge also often does not have the accuracy, therefore, this also causes the current intrusion detection system to be lack of science. At present, most of the intrusion detection systems are still integrated, and their expansibility is low, so it is difficult to combine them with the new intrusion detection system model [7].

With the progress of the times and the continuous development of science and technology, data mining technology is gradually used in people's production and life. If we combine data mining with intrusion detection system, it can solve the main problems of intrusion detection system scientifically and efficiently. The process of intrusion detection can be regarded as a process of data analysis.

4. Intrusion Detection System based on Data Mining

4.1 Characteristics of Intrusion Detection System based on Data Mining

Intrusion detection system based on data mining is quite different from other analysis systems. It is mainly based on data and regards the process of intrusion detection as a process of data analysis and processing. Therefore, combining data mining with intrusion detection system, even if there is an unknown attack behavior, we can find the association between this behavior and other behaviors from the database, so as to detect abnormal behavior. Therefore, adding data mining technology to intrusion detection system can improve the self-learning ability and adaptive ability of intrusion detection system, and improve its expansibility [8]. The core of fundamentally solving the problem of false positives and false positives in intrusion detection is to extract accurate patterns to describe the characteristics of behavior from a large amount of data. Intrusion detection system based on data mining mainly has the following four advantages.

4.1.1 High Intelligence and High Degree of Automation

Intrusion detection system based on data mining operates a variety of disciplines, such as statistics, decision science and neural network science, which can automatically extract network behavior patterns that experts are difficult to find from the data. This can fundamentally improve the intelligence and automation of the intrusion detection system, so as to reduce the work pressure of the relevant personnel.

4.1.2 High Detection Efficiency and Low False Alarm Rate

Data mining technology can extract valuable data from massive data and fundamentally reduce the data processing capacity of intrusion detection system in the later stage. Therefore, data mining technology can fundamentally improve the detection efficiency of intrusion detection system. At present, most intrusion detection systems only rely on simple information matching, so their alerts may be higher than the actual situation, resulting in a higher false alarm rate. With the integration of data mining technology, intrusion detection system can efficiently eliminate a variety of repeated attack data, so as to reduce the false alarm rate [9].

4.1.3 Strong Adaptability and Low Missing Report Rate

The intrusion detection system based on data mining technology is not based on the predefined detection model, so it has high adaptive ability. Specifically, when there is an unknown attack mode, the traditional intrusion detection system has a high probability that it cannot be identified, but the intrusion detection system based on data mining can quickly detect and identify the attack behavior.

4.1.4 Reducing the Amount of Overloaded Data

Data mining technology can not only analyze and deal with the relationship between all kinds of data, but also provide the characteristics of all kinds of data for the system from many aspects. It can combine previous results with the latest data, thereby reducing unnecessary data and reducing data overload.

High Intelligence and High Degree of Automation
High Detection Efficiency and Low False Alarm Rate
Strong Adaptability and Low Missing Report Rate
Reducing the Amount of Overloaded Data

Fig. 4 Four Advantages of Intrusion Detection System based on Data Mining

4.2 Construction of Intrusion Detection Model using Data Mining Technology

The architecture of adaptive intrusion detection system based on data mining includes sensor, data preprocessor, database, data mining, rule ontology description, rule base, detection engine and decision center. In addition, specifically, the working principle of intrusion detection system based on data mining has the following five points.

First, the function of the data sensor is to collect all kinds of data from the external environment, and then transmit the data to the data preprocessor. Because no matter what kind of data, there is inevitably incomplete data. Therefore, in order to make the mining data both authentic and effective, it is necessary to preprocess the data source. There are four steps in the processing, namely, book cleaning, data integration and exchange, data specification, discretization and data hierarchical generation. Through these four steps, you can make the data into a unified format, and then store the data in the database [10].

Second, the use of data mining technology to learn database data, can extract the relevant behavior characteristics and rules, to understand abnormal patterns and normal patterns, and then build a perfect detection model.

Third, the state obtained from data mining and its corresponding behavior rules are described by ontology, which is conducive to the sharing and complete expression of rules and stored in the rule base.

Fourth, the model constructed by data mining technology and the rules from the rule base are analyzed, and the analysis results are transmitted to the decision center.

Fifth, let the decision center to judge the network behavior, and implement the corresponding processing scheme.

5. Conclusion

With the continuous progress of science and technology, data mining technology has also begun to enter people's production and life, in which intrusion detection system is also combined with data mining technology, so as to promote its development in the direction of intelligence. However, with the combination of the two, people gradually found that the real-time reflection ability of data mining is poor, and the mining time is also very long, and even the intrusion detection system based on data mining still has the problem of high false alarm rate. In addition, it is also found that the intrusion detection system based on data mining has weak ability to prevent unknown attacks. Thus, it can be seen that at present, the intrusion detection system of technical data mining still has a lot of room for improvement, and it still needs to be combined with other security defense means, such as firewall, anti-virus software and so on. The author believes that in the future, the combination of data mining technology, protocol analysis technology and correlation analysis technology can improve the detection efficiency of intrusion detection system to a certain extent. However, we should also realize that the current intrusion detection system based on data mining has also been a big step in the field of network security. Therefore, the author believes that in the future, with the progress of science and technology and the efforts of scientists, intrusion detection system based on data mining will be widely used.

References

[1] Lu Yong, Cao Yang, Ling Jun, et al. Intrusion detection system framework based on data mining [J]. Journal of Wuhan University (Science Edition), 2002, 48 (1): 63-66.

[2] Wang Shenghe. Design of Intrusion Detection System Based on Data Mining [J]. Computer Engineering and Design, 2004, 25 (2): 243-245.

[3] Jiang Yunyan, Changsheng. Intrusion Detection Based on Data Mining [J]. Computer Application and Software, 2006, 23 (11): 124-126.

[4] Zou Hong, Chen Hai, Wei Qinyu. Research on Intrusion Detection Technology Based on Data Mining [J]. Computer and Modernization, 2005 (04): 41-43.

[5] Cui Ting, Li Yulong. Application of Data Mining in Intrusion Detection System [J]. Computer Engineering and Design, 2008 (23): 33-34 + 79.

[6] Cheng Yuqing, Mei Denghua, Chen Longfei, et al. Intrusion Detection System Model Based on Data Mining [J]. Computer Technology and Development, 2009, 19 (12): 123-126.

[7] Shi Shaomin. Hybrid Intrusion Detection Model and Analysis Based on Data Mining [J]. Communication Technology, 2009, 42 (8).

[8] Wang Zhongcai, Li yongbi. Research on Intrusion Detection System Based on Data Mining [J]. Science and Technology Bulletin, 2012 (08): 158-160.

[9] Luo Yun, Lu Lan. Intrusion Detection System Model Based on Data Mining [J]. Computer Knowledge and Technology, 2012, 08 (3): 560-561.

[10] Tian Chunping, Liu Yun. Network Intrusion Analysis and Detection Based on Data Mining Technology [J]. Digital Communication World, 2019 (5).