The Research of Network Anomaly Detection Technology Based on Data Mining

Chunhong WU[1], Wenzhong XIA[2], Fengyun LIU*[3,C]

[1,2,3]Zhangjiakou Vocational College of Technology, Zhangjiakou 075000,China

[c]fengyunliu@163.com

Keywords: Data Mining; Network Anomaly; Detection Technology

Abstract. With the rapid development of computer network, various problems also arise, especially the network security problem. How to quickly and efficiently find a variety of network intrusion behavior, to ensure the safety of the system and network resource is very important. Intrusion detection is a proactive safety protection technology, to take the initiative to monitor and track the invasion, to protect the computer system, network system, and the safety of the information infrastructure has very important practical significance. In this paper, in light of the characteristics of the network environment, this paper proposes a new network anomaly detection based on data mining technology, the method of clustering analysis in data mining to extension, raise the efficiency of intrusion detection system, guaranteeing the security of the network.

Introduction

Along with the computer network and globalization, social each domain in the Internet age has a qualitative leap, people's study, work and life are integrated into the network, people through the network to a Shared resource. After decades of development, great changes have taken place in the network environment, from simple to complex structure, all kinds of network technology in time and space is outspread, the increase of the number of users and equipment, frequent network attack behavior, such as the unstable factors in the network of difficulties make network management [1]. Only the computer network security, information to the normal development of the society, national information security, to guarantee the network life are not violated, the people as a result, network security technology research has important social significance and practical significance.

Intrusion detection is a hot spot in the study of network security technology, has achieved a certain development, have also launched many commercial products, but because of detection efficiency is not high, adaptability is not strong, and disadvantages such as lack of extensibility is difficult to fully meet the requirements of computer network security, still need to continue to study, make the intrusion detection system more perfect [2]. To this end, on the basis of the research of data mining technology, this paper designs and realizes a new data mining model, and its application to network intrusion detection, improve the deficiency of the existing detection algorithms and models, to improve the efficiency of the intrusion detection system.

The basic technology of data mining

Data mining is a long-term research and application of database technology is an inevitable result of the development of database technology but also to a more advanced stage, it can not only large amounts of historical data query, data can also be found in the history of the unknown potential link [3] . Faced with the current state of large amounts of data, the association rule is an important branch of data mining, as the study an advanced and intelligent data processing and analysis technology has become a hot spot. By association rule mining, the amount of useful information can be implied in the sea there is the potential value of the data [4]. Target association rules is an effective method to extract the most interesting patterns. So far, it has been proposed many effective association rule mining algorithm, suggesting the algorithm is proposed mining algorithm most famous Agawal algorithm is based on a priori, but it is efficient in terms of time and space scales are faced with the challenge Therefore, many researchers explore new mining method, expanding

the concept of association rules and applications.

Data mining refers to the large amount of data warehouse, reveals implicit, previously unknown, potentially important process valuable information, data mining is a decision support process, which is mainly based on artificial intelligence, machine learning, pattern recognition, statistical, database, visualization techniques, highly automated analysis of enterprise data, make inductive reasoning, dig out potential model to help the user to adjust the marketing strategy, reduce risk, make the right decisions. From the perspective of the process of data mining techniques, as shown in figure 1.

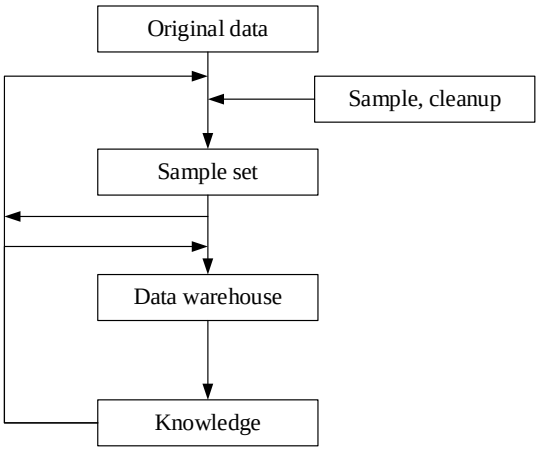

Figure 1.The basic process of data mining

After data collection, sampling and cleanup needs. Cleanup is the result of the sample data set. Data Warehouse is an effective form of data storage, data mining is very beneficial. You can use a variety of data mining algorithms. Sometimes, the need to return to the final stage of the above process.

The intrusion detection technology

Intrusion detection system intrusion detection framework based on common framework (CIDF) proposed intrusion detection system can be simplified into three parts: data extraction module, the results of the data analysis module, a processing module [5]. The role of the data extraction module is to provide the system data, system status data source, network, data, and user activity and behavior, all of the test data sources. One of the most important data analysis module, intrusion detection system. Its role is more in-depth analysis of the data, through a process of gradual accumulation of early detection of the data model and protocol analysis, and to determine whether there is any data in violation of the policy, in accordance with normal policy directly filtered to keep its records, contrary finally passed result of the processing module.

Result processing unit module, in response, the role of the reaction unit receives the event data as a result of the analysis module, which can be connected to the cutting, change the file property, even a strong response to counter attacks such as those, but can also make a simple alarm handling. Host-based intrusion detection is shown in Figure 3.

Host-based intrusion detection system protection agent running on each host. Cost performance in a few cases, this approach may be more cost-effective, while this method can be easily monitored in some activities, such as access to sensitive files, directories, programs, or ports, based on these activities is difficult clues agreement. Once the intruder to obtain a user name and password, and according to the host agent is the easiest to distinguish between normal and illegal activities. Host-based methods are sometimes need to add specific hardware platform, generally will not be lost because of increased network traffic monitoring network behavior.

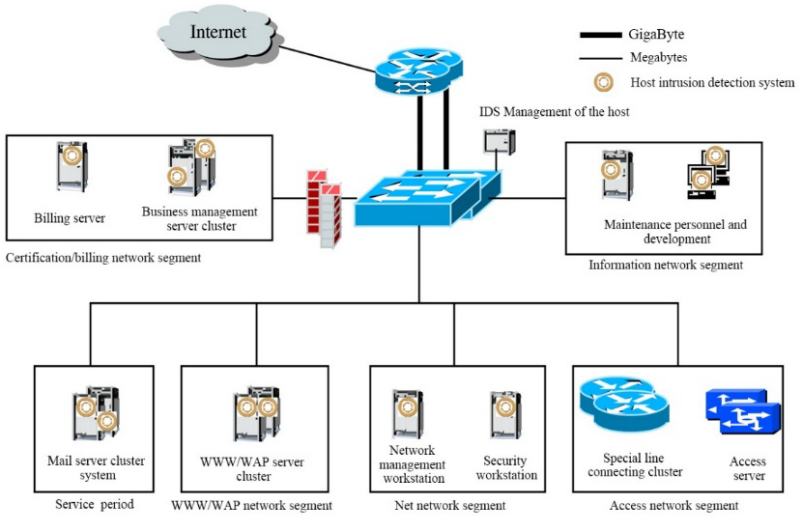

Figure 2. Host-based intrusion detection

The design of intrusion detection system based on data mining

In the algorithm of data mining algorithm for mining association analysis can be found that the network connection data attribute, the relationship between sequence analyses algorithms can be found about invasion associated characteristics of intrusion attack. Through sequence pattern analysis method is applied to obtain the intruder behavior sequence relation, be able to get intrusion behavior characteristics of the temporal information, similarly also can get the characteristics of the normal behavior, according to the time sequence characteristics of user behavior to determine the user's behavior is normal behavior or intrusion behavior. Using correlation analysis algorithm and sequence analysis algorithm to construct the normal patterns of behavior and applied to anomaly intrusion detection; finally, the algorithm carries on the classified analysis, can from the training data obtained from mining to identify normal behavior and intrusion behavior rules. The principle and the mining process is shown in figure 3.

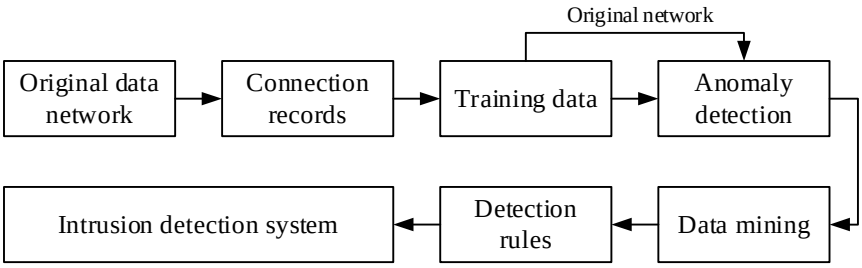

Figure 3. The mining process of intrusion detection system

Intrusion detection system model based on data mining framework basically has the following several parts: data preprocessing module, association rules mining module, misuse detection rules mining module, as shown in figure 4.

Data collection and pretreatment: Its main is to collect all the network behavior records, preprocessing, and finally generate the training data set; Association rules or rules: excavated from a training set of data, generate and sequence rules, association rules that generate normal behavior patterns used in anomaly detection of intrusion behavior.

Misuse detection model: with the classification rule mining algorithm to deal with the updated training data and extract classification rules and misuse detection.

Anomaly detection model: complete intrusion detection of abnormal behavior, there are two main functions: one is to use the classification rule as a judge on the basis of real-time detection of

network data, another function is used as an anomaly detection based on correlation and sequence mode, network data is normal or invasion.

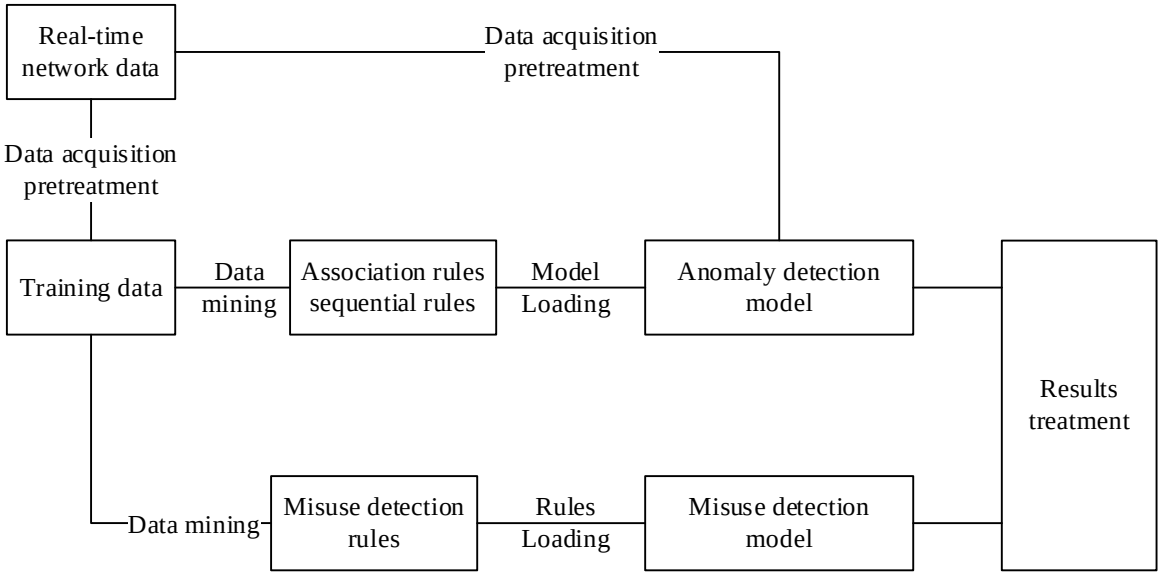

Figure 4. Intrusion detection framework based on data mining

Conclusion

At present there are two directions in the field of intrusion detection technology research: misuse detection and anomaly detection methods. A variety of methods has its own advantages and disadvantages, misuse detection can accurately detect the invasion of the known behavior, but not to the new intrusion behavior play a role of detection; Anomaly detection has better ability in intrusion detection to the unknown intrusion behavior, but the completeness is not easy to implement on the normal behavior model, and in detection rate and the rate of false positives than misuse detection performance is poor. In view of the above situation, based on the analysis of two kinds of technology methods on the basis of fully combined the advantages of the two kinds of methods, build the model of intrusion detection system based on data mining.

Reference

[1] P. Garcia-Teodoro, J. Diaz-Verdejo, and G. Maciá-Fernández: Computers & security, Vol. 28(2012) No.1, p. 18.

[2] M.A. Aydın, A.H. Zaim, and K.G. Ceylan: Computers & Electrical Engineering, Vol. 35(2010) No.3, p. 517.

[3] W. Hu, and S. Maybank: Systems, Man, and Cybernetics, Part B: Cybernetics, Vol. 38(2009) No.2, p. 577.

[4] A. Boukerche, R.B. Machado, and K.R.L. Jucá: Computer Communications, Vol. 30(20011) No.13, p. 2649.

[5] J. Kim, P.J. Bentley, U. Aickelin: Natural computing, Vol. 6(2008) No.4, p. 413.