A Novel Intrusion Detection System Based on Data Mining

Xu Tao[1,a] , Zhang Wei,Li XuHong,Wang Xia,Pan Wenwen

[1] The University Of Zaozhuang, Zaozhuang 277160, ShanDong,China

[a]xutao@uzz.edu.cn

Keywords: Data Mining;Intrusion Detection; Network Security

Abstract. The rapid development of the Internet makes distributed computing has become a mainstream application, network openness, connectivity, shared characteristics, so that the risk of suffering from the growing network intrusions. How to ensure the network and information security has become very important in the field of security issues. From the initial access control mechanisms to combine packet filtering and application layer gateway firewall technology, each is not the perfect solution. Intrusion detection is a network and information security architecture an important part of the intrusion detection system put forward higher requirements. Current greatest weakness of the face of live flowers audit records cannot be quickly mass intrusion detection and high false positive rate of serious impact on system performance. This paper proposes a new intrusion detection method, and on this basis, based on data mining developed based anomaly intrusion detection system prototype network.

Introduction

With the popularity of Internet in the world and the development of computer networks have been and people learn, work closely together. However, network attacks and intrusions for national security, economic and social life caused great threat. Therefore, the information security issues become the focus of information industry, and has become an important component of national security, but also decide the country's economic sustainability of key high-speed development [1].

Intrusion detection is an important component of information security technologies, involving log analysis, vulnerability detection, attack path detection, and other technologies that rely heavily on development, including data mining techniques, including a variety of data analysis and computing technology [2-3]. Data mining techniques combined with statistical and computing, centralized access to useful patterns from large amounts of data, which can attribute to the data in the database, effectively describe the set of objects, a set of rules to produce guidance provided to the decision support system. As machine learning in a database application, data mining can provide the basis for intrusion detection, attack detection path in association rules and sequence patterns and other aspects, so as to contribute to the development of protection strategies. Based on the understanding of data mining algorithm based on the proposed method improved algorithm, and focuses on data mining technology for network security audit data, data security log correlation analysis conducted to explore the characteristics of intrusion packet and attack-series model, constructed from the test data set intrusion model.

Intrusion Detection Process Based on Data Mining

Intrusion detection is mainly for network collected a lot of data to judge, that is how it is found from normal behavior or abnormal behavior, and can be effective from a generation of intrusion detection rules, association analysis algorithm introduced earlier data mining, sequence analysis algorithm, cluster analysis algorithm can be used for intrusion detection. By applying different data mining algorithms that can extract the user or system behavior characteristic data. In the data mining algorithm for mining association analysis algorithm can be found in the relationship between the network connection data attributes, sequence analysis algorithm can be found on invasive characteristics associated with the intrusion [4]. Intruder obtained by applying the sequential pattern analysis of the relationship between action sequences, you can get the timing

characteristics of intrusion information, empathy normal behavior characteristics can be obtained, according to the time sequence characteristics of user behavior to determine the user's behavior is normal behavior or intrusion behavior. The use of correlation analysis algorithm and sequence analysis algorithm to construct normal behavior patterns applied to anomaly intrusion detection; and finally analyzed by their classification algorithm, you can get to identify normal behavior and intrusion rules from training data by tapping [5-6]. The principle mining and process shown in Figure 1.

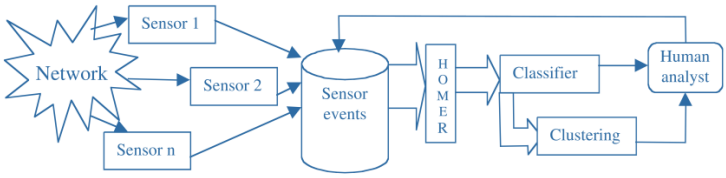

Figure 1. Based on data mining intrusion detection process

(1) Data collection. The first step in intrusion detection data (information) collection, the collection includes the status and behavior of systems, networks, data and user activity. By the sensors placed in different segments or different host agents to collect information, including system and network log files, network traffic, non-normal directory and file changes in the non-normal program execution.

(2) Data analysis. The second step is data (information) analysis, the collected information about the system, network, data and user activity in the state and behavior information, to the detection engine, the detection engine resides in the sensor, usually by means of three techniques Analysis: pattern matching, statistical analysis and integrity analysis.

(3) The results of the processing. Incident Response and consoles produced in accordance with pre-defined alarm response to take corresponding measures, can be re-configure the router or firewall, to terminate the process, cutting off connections, change file attributes, you can simply alert.

Intrusion Detection Model Based on Data Mining

In order to properly filter network data, you must be able to generate accurate data to identify patterns of behavior normal, cluster analysis is an effective way to build a network of normal behavior patterns. Packets will not meet the normal behavior patterns considered abnormal packets, they will be in the system detector for further testing, and abnormal packet detector is not likely to be found in the new data packets generated intrusions, generate new behavior patterns after the invasion of these packages use features abnormal data extraction module for analysis, and then the invasion of these new patterns of behavior change into a new intrusion detection rule to rule libraries, so the detector can detect new unknown invasion behavior.System model framework consists of the following components: data collection preprocessing module, association rule mining module sequence rule, misuse detection rule mining module, intrusion detection engine module, shown in Figure 2.

Preprocessing of data acquisition module is mainly responsible for network connection record collection, pretreatment, and the formation of a training data set; association rule mining sequence rule mining module is responsible for generating association rules and sequence rules, found in normal mode, for anomaly detection; misuse detection Rules Mining module classification rule mining algorithm for training to learn the updated data extracted classification rules for misuse detection; intrusion detection engine module on the one hand the use of classification rules as misuse detection criterion of real-time detection of network data On the other hand using the associated error detection and sequential pattern as the basis to determine the network data is normal or invasion, and the test results and then back to the training data collection module.

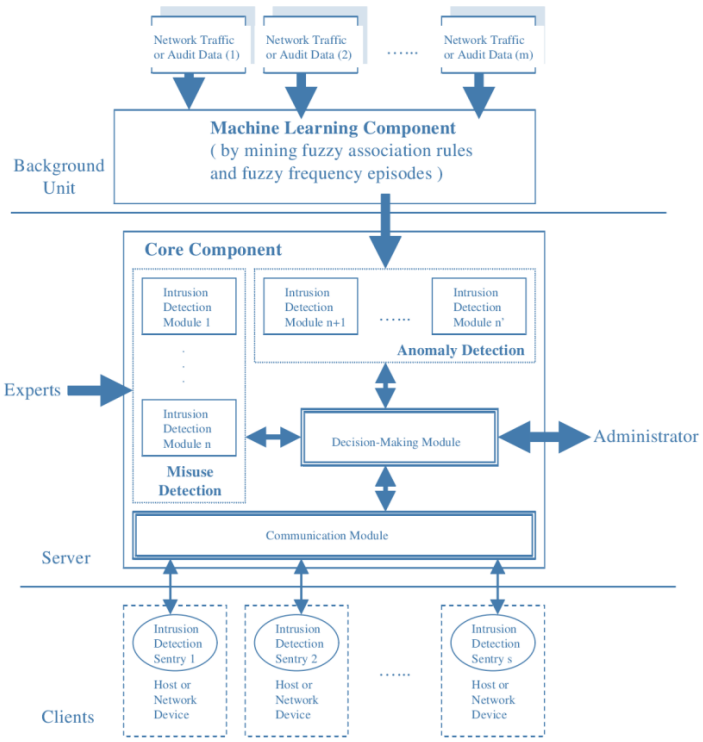

Figure 2.Intrusion Detection Model Based on Data Mining

Experimental Analysis

Our experiments attack broke out, and the detection rate for all anomaly detection program acquisition. We use the standard indicators, taking into account the outbreak detected if the corresponding burst detection rate greater than 50%. Because we are a total of 19 sudden attack, the overall detection rate was calculated using this rule. The results in Figure 3 show that the two most successful outlier detection scheme Nearest Neighbor (NN) and LOF, neural network method is able to detect 14 attacks broke out, able to detect the outbreak of LOF 13 attack methods. Although detection using unsupervised hierarchical support vector machine looks good when comparisons are unfair because the false alarm rate in this case is 4%. False alarm rate is fixed to the training data 2% false positive rate of the test data cannot be maintained at this speed, and rose to 4%.

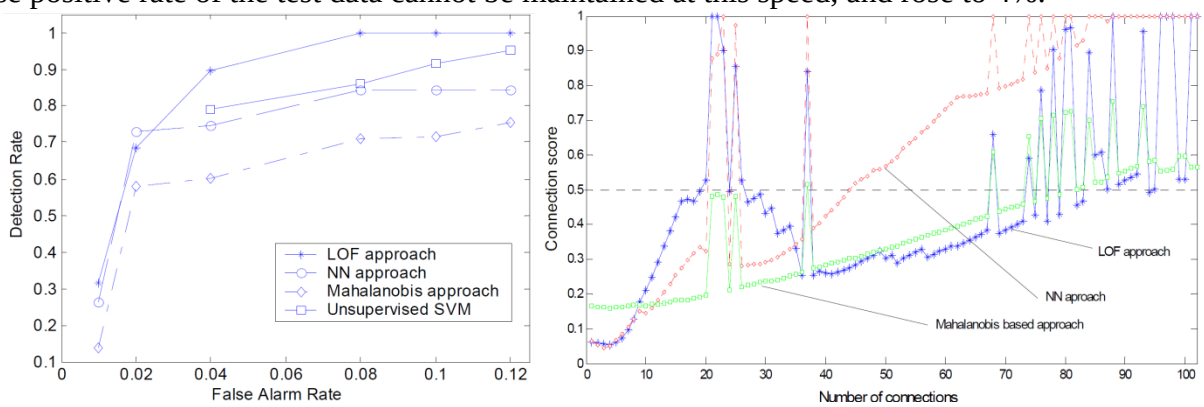

Figure 3. Intrusion detection test using different algorithm of data mining

Figure 3 illustrates the ROC curves for all the proposed algorithms and detection rate and false alarm rate show different uses of different thresholds, the most consistent abnormality detection program is LOF approach because it is only slightly smaller than the low false alarm rate (1 % and 2%), neural networks Markov pedagogy always inferior to the neural network method is able to detect just over 11 connect to the attacker. Based on the Markov,this poor performance may be solutions, normal behavior may have several types, and cannot have a single profile.

Conclusion

The main function of intrusion detection system is to detect intrusions and identify the invasion behavior, essentially data network behavior classification. The network data into normal and abnormal data types, determination of abnormal behavior is intrusion detection. Intrusion detection system can host system, run applications and their overall status, real time monitoring, and also within the system of intrusion attacks and external attacks in real-time detection, while proactively identify intrusions and alarm and so on. The above features make it play an important role in network security. In this paper, data mining technology into intrusion detection technology can only enhance the ability of intrusion detection, intrusion detection while intelligent. But the detection accuracy and detection rate is difficult taking into account. So, how can we make do with high accuracy and high rate of detection of network attacks, will be a problem long-term research needs.

Reference:

[1] Han J, Kamber M, Pei J. Data mining: concepts and techniques: concepts and techniques[M]. Elsevier, 2011.

[2] So-In C, Mongkonchai N, Aimtongkham P, et al. An evaluation of data mining classification models for network intrusion detection[C]//Digital Information and Communication Technology and it's Applications (DICTAP), 2014 Fourth International Conference on. IEEE, 2014: 90-94.

[3] Tsai C F, Hsu Y F, Lin C Y, et al. Intrusion detection by machine learning: A review[J]. Expert Systems with Applications, 2009, 36(10): 11994-12000.

[4] Mohammad M N, Sulaiman N, Muhsin O A. A novel intrusion detection system by using intelligent data mining in weka environment[J]. Procedia Computer Science, 2011, 3: 1237-1242.

[5] Zuech R, Khoshgoftaar T M, Wald R. Intrusion detection and Big Heterogeneous Data: a Survey[J]. Journal of Big Data, 2015, 2(1): 1-41.

[6] Hajian S, Domingo-Ferrer J, Martinez-Balleste A. Discrimination prevention in data mining for intrusion and crime detection[C]//Computational Intelligence in Cyber Security (CICS), 2011 IEEE Symposium on. IEEE, 2011: 47-54.