A Novel Research on Real-Time Intrusion Detection Technology Based on Data Mining

Julan YI[1, a]

[1]XINYU UNIVERSITY, Xin Yu 338004, China

[a]julanyi@126.com

Keywords: Real-Time Intrusion Detection; Data Mining; Network Security

Abstract. with the wide application of network technology and the continuous development of network economy, network security issues are also increasingly prominent, it gives the network itself and the information system based on network has brought great threat. Adaptive detection based on data mining is the development trend of intrusion detection system, it can adapt to different network environment, and does not depend on expert training data. Based on the reference on the basis of predecessors, this paper proposes a real-time intrusion detection system based on data mining framework structure. Research based on data mining technology, with adaptability and real-time detection function and the structure of the distributed intrusion detection system, to promote the network information security technology and products to develop in the direction of the omni-directional three-dimensional testing and national information safety inspection work is meaningful.

Introduction

With the rapid development of network technology, computer network has been widely applied to various fields of human activity, the network impact on social economy and people's life more and more. The problem of network security is becoming more and more widely attention, all kinds of network security technology and product emerge. Intrusion detection technology is one of the important technology. In this paper, aiming at the shortcomings of the existing intrusion detection systems, puts forward applying data mining to intrusion detection methods to improve its performance [1-2].

Based on the reference on the basis of predecessors, this paper proposes a real-time intrusion detection system based on data mining framework structure. In view of the current intrusion detection system detection strategy of a single, cannot cope with the complex environment changes, high false alarm and non-response rates. On the basis of the distributed real-time framework, we increase the adaptive strategy manager module and adaptive model manager. The two module USES the data mining technology, reducing the dependence on experts. It can realize detection strategy and model of automatic generation and distribution. We put our focus on system architecture design and configuration, it includes sensors, detectors, data warehouse, data analysis, adaptive model management and strategy management, and other components. The system structure using the algorithm of association rules and frequent pattern detection model, greatly reduces the manual coding, further improve the ability of the system of automatic learning algorithm to construct classifier and generate strategy.

Principle of intrusion detection system

Intrusion detection, which is found for invading behavior. It through the computer network or computer system in a number of key points to collect information and carries on the analysis, found in the network or system whether there is a violation of security policy and the signs of being attacked. It does not measure from outside intrusion behavior, but also detect unauthorized activities of the internal users. Intrusion Detection System (Intrusion Detection System, IDS) is to perform Intrusion Detection work of hardware and software products. IDS, through the analysis of the real

time to check the specific attack mode, system configuration, system vulnerabilities, flawed version of program and system or user behavior patterns, monitoring and safety related activities [3]. A basic IDS need to solve two problems: one is how to fully describe the behavior characteristics of the data extracted reliably. The second is how to according to the characteristics of the data, the legitimacy of the decision behavior efficiently and accurately. Intrusion detection system consists of data collection, data analysis and processing results of three parts, in addition, also include other components, such as configuration management, interface management, communication with other systems, and so on [4-5]. Data source is diversiform, can be a host information, information on the Internet and other information systems, etc. Intrusion detection system structure as shown in figure 1.

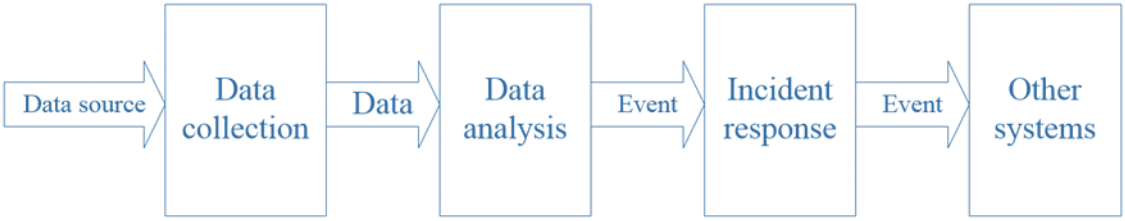

Figure 1. The structure of intrusion detection system

(1) Data collection. Intrusion detection is the first step in data (information) collection, collection content includes system, network, data and user activity state and behavior. By placing sensors in different network segments or agent to collect the information of different hosts, including log file system and network, network traffic, abnormal changes in the directories and files, abnormal program execution.

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

(3) Result processing. Incident response and console the predefined response caused by the alarm to take corresponding measures, can be reconfigured router or a firewall, to cut off the connection, change, termination process file attributes, can also be just a simple alarm

Key technology of real-time intrusion detection Based on Data Mining

Real-time intrusion detection system based on data mining is mainly composed of data collection, data mining, pattern matching, and decision making module. Data collection module from the data source, such as system logs and network data packet, to extract the raw data, will be after preprocessing the audit data is submitted to the data mining module, data mining module of the audit data sorting, analysis, found that can be used for the real-time detecting patterns and knowledge, and then submitted to the pattern matching analysis module, make a final judgment, finally by the decision module response is given. The basic model of the whole system is shown in figure 2.

1) The intrusion detection technology based on association rules [6]. In the network security system, can use a correlation analysis to find out the correlation between various intrusion behavior of the invaders. Association rule mining is one of the most widely used technology in data mining, is also the earliest used in intrusion detection technology. Now has a variety of association rules algorithm such as Apirori algorithms used in intrusion detection. Association rule is used in the analysis of network flow data at the earliest, then will be the result of association rule mining as input data, after mining, so as to dig up more optimal results.

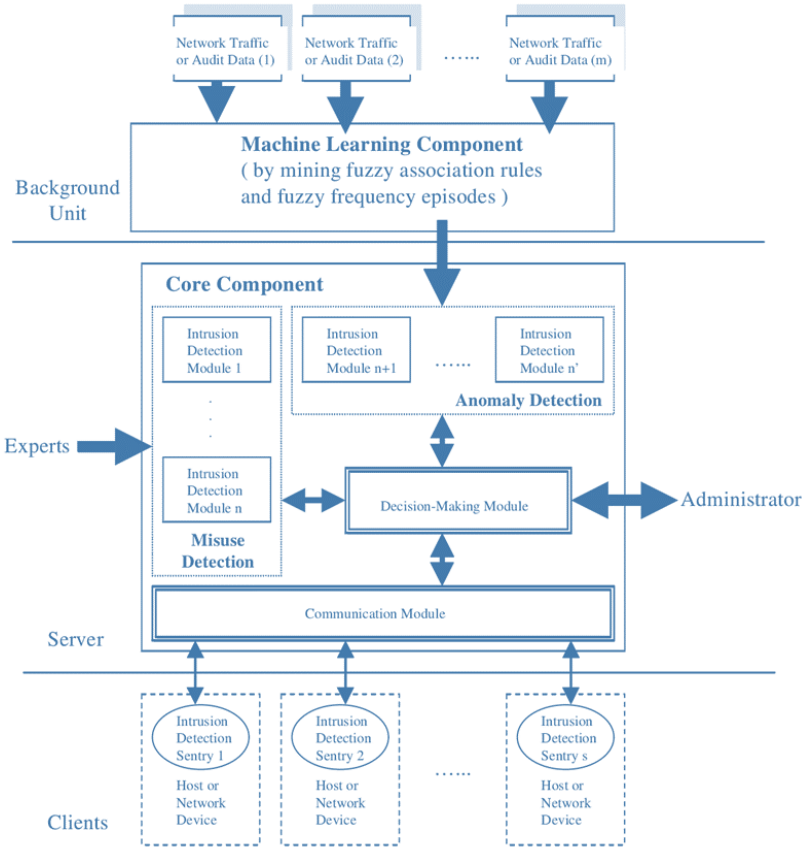

Figure 2. Real-time intrusion detection model based on data mining

2) The intrusion detection based on clustering analysis [7]. Cluster analysis is a kind of unsupervised learning methods, it is some unknown pattern can be divided into several classes, if the distance between the feature vector within a certain error range are equal, thinks that they are the same type. Intrusion detection algorithm based on clustering analysis is the basic idea is based on two assumptions, namely invasion and the number of different and normal behavior on the normal mode should be greater than the number of intrusion behavior two conditions, with this data set can be divided into different categories, which distinguish between normal and abnormal behavior to detect intrusions. Commonly used clustering algorithm is k - clustering, fuzzy clustering, self-organizing mapping M neural network clustering and genetic clustering, etc. Clustering method does not need manual or other classification, also do not need training. So can find new and unknown intrusion types.

3) The intrusion detection technology based on frequent sequence [8]. Due to network attack is closely related with time variables, so the analysis model on the basis of correlation analysis and further analysis correlation attack time. Main methods of frequent sequence mining successively relationship between security incidents, using the sequence analysis found that the sequence of intrusion behavior relationship, extract intrusion behavior between time series characteristics. Sequence pattern analysis is generally not used alone, it can be used in the intrusion detection process is one of the steps, the user sequential patterns in the data mining, to extract the knowledge can be used in intrusion detection and pattern.

4) Based on the analysis of the classification of intrusion detection technology, intrusion detection can be thought of as a data classification problem. Intrusion detection classification mining shall, first of all, choose a set of training data, the training set to mark the normal or abnormal data, using methods such as classification rules, decision tree classification rules are extracted from the data set and construct a suitable classifier. Then use constructing classifier classifying gather real-time network data flow, data can be divided into normal behavior or some kind of intrusion behavior, to determine whether there is any intrusion behavior. The classification

process should be repeated and evaluation, in the hope can get the optimization of intrusion detection classifier.

Experimental Analysis

Our experiment against explosion, detection rate anomaly detection procedure. We use the standard index, considering the outbreak of the detected if the corresponding detection rate is more than 50%.Because we have a total of 19 sudden attack, use the law to calculate the total detection rate. Results shown in figure 3, two of the most successful anomaly detection scheme of nearest neighbor (NN) and a LOF, neural network method to detect 14 after the outbreak of attacks, can detect LOF 13 attacks. Although the unsupervised hierarchical support vector machine (SVM) were used to detect looks good when comparison is not fair, because the rate of false positives in this case is 4%.Fixed to the rate of false positives can't maintain at that speed test data of the training data 2% false positive rate, and up to 4%.

Figure 3 shows the ROC curve of all of the proposed algorithm and the detection rate and false alarm rate showed different USES of different threshold, the most consistent anomaly detection program is a LOF of method, because it is better than low false alarm rate is just slightly smaller and 1% (1%), neural network Markov method is always inferior to the neural network method can detect just over 11 connects to the attacker. Based on Markov, the poor performance can be a solution, can have several types of normal behavior, and can't have a single profile.

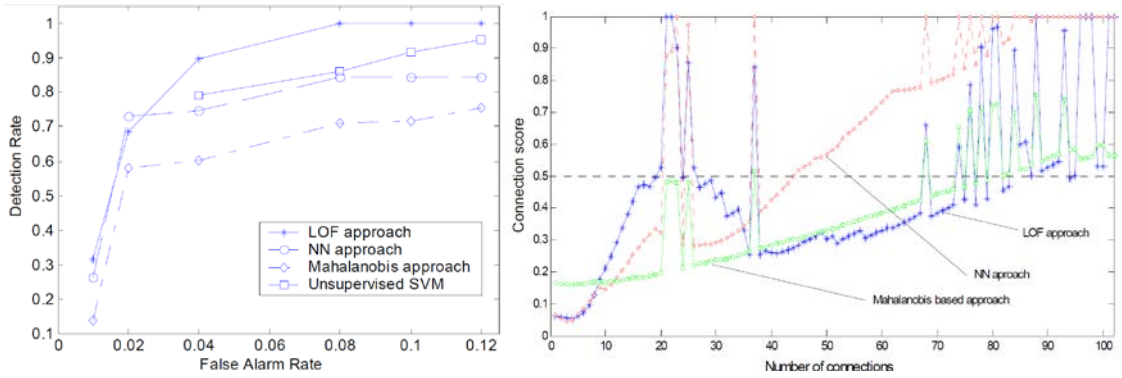

Figure 3. Real-time intrusion detection test using different algorithm of data mining

Conclusion

With the development of computer network especially the Internet technology, the network in our daily life and plays a more and more important role in our work. And growing network at the same time, more and more sensitive information is online storage and management, make the network more vulnerable, more vulnerable to all sorts of malicious or illegal users on the network attack, so network security is becoming more and more important. Based on the reference on the basis of predecessors, this paper proposes a real-time intrusion detection system based on data mining framework structure. Single, in view of the current intrusion detection system detection strategy can't cope with the complex environment changes, false alarm and high non-response rates, based on data mining technology are studied, with adaptability and real-time detection function and the structure of the distributed intrusion detection system, to promote the network information security technology and products to develop in the direction of a full range of three-dimensional detection and national information safety inspection work is meaningful.

Reference

[1] 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.

[2] 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.

[3] Phua C, Lee V, Smith K, et al. A comprehensive survey of data mining-based fraud detection research[J]. arXiv preprint arXiv:1009.6119, 2010.

[4] Garcia-Teodoro P, Diaz-Verdejo J, Maciá-Fernández G, et al. Anomaly-based network intrusion detection: Techniques, systems and challenges[J]. computers & security, 2009, 28(1): 18-28.

[5] Su M Y, Yu G J, Lin C Y. A real-time network intrusion detection system for large-scale attacks based on an incremental mining approach[J]. Computers & security, 2009, 28(5): 301-309.

[6] Sangkatsanee P, Wattanapongsakorn N, Charnsripinyo C. Practical real-time intrusion detection using machine learning approaches[J]. Computer Communications, 2011, 34(18): 2227-2235.

[7] Vaarandi R. Real-time classification of IDS alerts with data mining techniques[C]//Military Communications Conference, 2009. MILCOM 2009. IEEE. IEEE, 2009: 1-7.

[8] Corchado E, Herrero Á. Neural visualization of network traffic data for intrusion detection[J]. Applied Soft Computing, 2011, 11(2): 2042-2056.