Intrusion detection method based on cloud model and semi-supervised clustering dynamic weighting

Liping Wang
Pingxiang University, Pingxiang Jiangxi, 337055

Keywords: Intrusion detection; semi - supervised clustering; active learning; intrusion prevention

Abstract. Aiming at the problem of low detection rate and high false positive rate of intrusion detection system, a cloud model semi-supervised clustering dynamic weighting intrusion detection method is put forward. As the attribute contributes to the classification difference, the cloud was near relative degree. The method of calculating attribute weight is given. With the semi-supervised clustering algorithm as the basis, the cloud model is constructed and the cloud classifier is constructed. The dynamic weights of attributes are used to classify the cloud classifier by updating the cloud model. Finally, the simulation results show that the proposed method has better detection performance and improves the performance of intrusion detection system.

Introduction

Network information systems need to take active defense measures [1-3]. Intrusion detection technology is a kind of active protection system which has emerged in the past 20 years, and is a new type of network security technology which is free from hacker attacks. The traditional intrusion detection algorithm is based on supervised learning [4]. The detection rate is high and the false alarm rate is low, but the unknown attack can not be detected, and the data is correctly marked as normal or abnormal. There is a large amount of unlabeled data in the network environment, and it is almost infeasible to mark the data correctly. If the unsupervised learning method is applied to intrusion detection, the clustering-based intrusion detection algorithm can detect unknown attacks with high detection rate and high false alarm rate. This paper proposes an intrusion detection algorithm based on semi-supervised learning [5].

With the popularity of computers and the rapid development of the network, unauthorized access, tampering with data and denial of service attacks become more serious [6]. Increasing network connectivity not only facilitates access to large amounts of data, but also provides access to data. Network intruders based on the information provided by the network, in the understanding of how the system works, the use of system vulnerabilities to obtain permission to complete his purpose. Intruder use of invasion mode to cover up his activities track, making the system can not identify him is the intruder [7]. In order to enable network security personnel to detect and detect intrusions and intrusion attempts as much as possible, effective measures are needed to plug vulnerabilities and repair systems. Intrusion detection technology is a kind of active protection system which has emerged in the past 20 years, and is a new type of network security technology which is free from hacker attacks. Intrusion detection is considered as the second security gate after the firewall, which monitors the network without affecting the network performance, thus providing real-time protection against internal attacks, external attacks and miss operation [8].

When intrusion patterns and network behavior characteristics change, the traditional intrusion detection system can not do anything, with the clustering algorithm can detect new unknown intrusion behavior. However, the clustering algorithm processing and describing the characteristics of network behavior have limitations. Once the attack is included in the training set as normal data, it can not detect such attack and its variants. Therefore, the current intrusion detection algorithm can not accurately detect unknown attacks, false alarm rate and other issues. In order to improve the intrusion

detection algorithm, improve the detection rate and reduce false alarm rate, this paper proposes an intrusion detection algorithm based on semi-supervised clustering. The main content of this paper is to design and implement the intrusion detection algorithm based on semi-supervised clustering. It can detect variants and unknown attacks of known attacks, and has high detection rate and low false alarm rate.

1. Intrusion detection system

1.1 Intrusion detection method

Intrusion detection can be classified according to different methods, according to the analysis method, mode of operation, data sources and other aspects of classification, the common classification was shown in Figure 1.

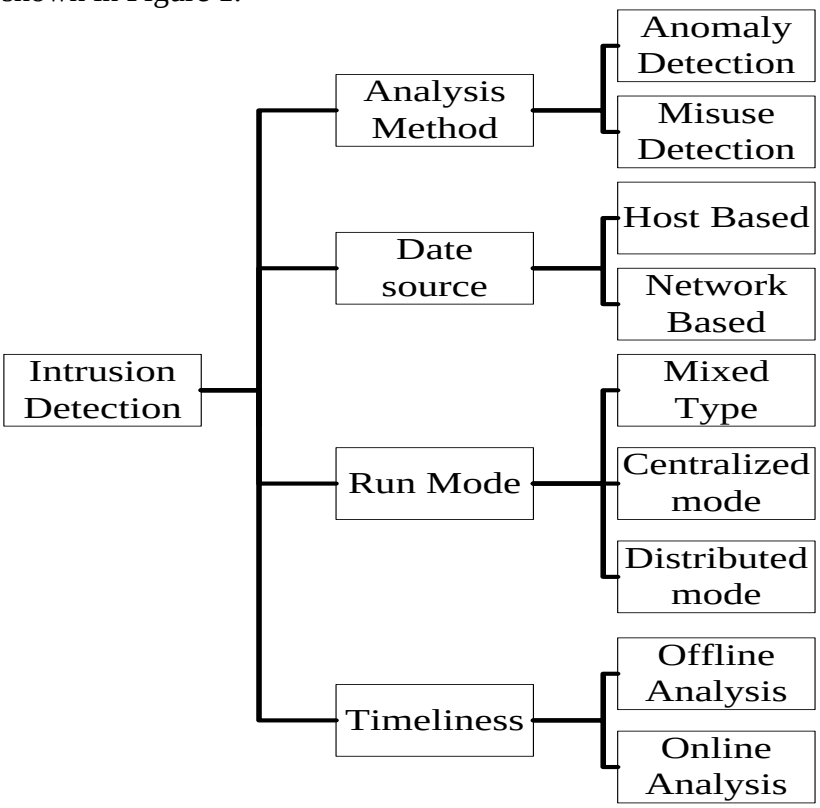

Figure 1. Classification of intrusion detection system

Different detection methods constitute the invasion model is not the same, abnormal detection is the key to the creation of abnormal models and how to use the model to detect other abnormal behavior. Abnormality detection technology first obtains the prior probability of intrusion because anomaly detection predicts changes in user behavior by observing prior probabilities, so it is important to obtain these prior probabilities of intrusion. A single anomaly should be squared and weighted:

$$a_1 S_1^2 + a_2 S_2^2 + \cdots + a_n S_n^2, a_i > 0 \tag{1}$$

1.2 Intrusion detection algorithm based on clustering

Distance can be used to measure the similarity between data. Multivariate observation properties with n data:

$$x_i = \left(x_{i1}, x_{i2}, \cdots, x_{ip} \right)^T, i = 1, 2, \cdots, n \tag{2}$$

At this time, each data can be seen as a point of p-dimensional space, n data p-element space n points. Let d (xi, yi) be the distance between data xi and xj and generally satisfy the following requirements:

$$d\left(x_i, x_j\right) = \left[\sum_{k=1}^{p}\left(x_{ik} - x_{jk}\right)^2\right]^{\frac{1}{2}} \tag{3}$$

Form a distance matrix:

$$\begin{pmatrix} 0 & d_{12} & \cdots & d_{1n} \\ d_{12} & 0 & \cdots & \\ \vdots & \vdots & & \vdots \\ d_{n1} & d_{n2} & \cdots & 0 \end{pmatrix} \tag{4}$$

The above distance is related to the dimension of each attribute index. To eliminate the influence of the dimension, we should standardize the data first, and then calculate the distance with standardized data. Standardized data:

$$x_{ik}^{*} = \frac{x_{ik} - \bar{x}_k}{s_k} \tag{5}$$

$$s_k^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_{ik} - \bar{x}_k\right) \tag{6}$$

$$\bar{x}_k = \frac{1}{n}\sum_{i=1}^{n} x_{ik}, i = 1, 2, \cdots, n; k = 1, 2, \cdots, p \tag{7}$$

2. Experiments and results

2.1 The function and basic structure of intrusion detection

In order to ensure the security of computer network and improve the resistance to external intrusion, it is necessary to establish a complete security protection system to carry out multi-level and multi-level detection and protection of network environment. Traditional security technologies such as cryptography, firewall technology, access control technology are mostly passive defense technology, can not take the initiative to find intrusion. The intrusion detection system can monitor the network, state, behavior and usage of the system in real time to detect whether the system or network contains abnormal behavior. If any abnormality is detected, the intrusion detection system will take corresponding measures such as alarming, logging events and cutting off the network. The role of intrusion detection was shown in Figure 2 below.

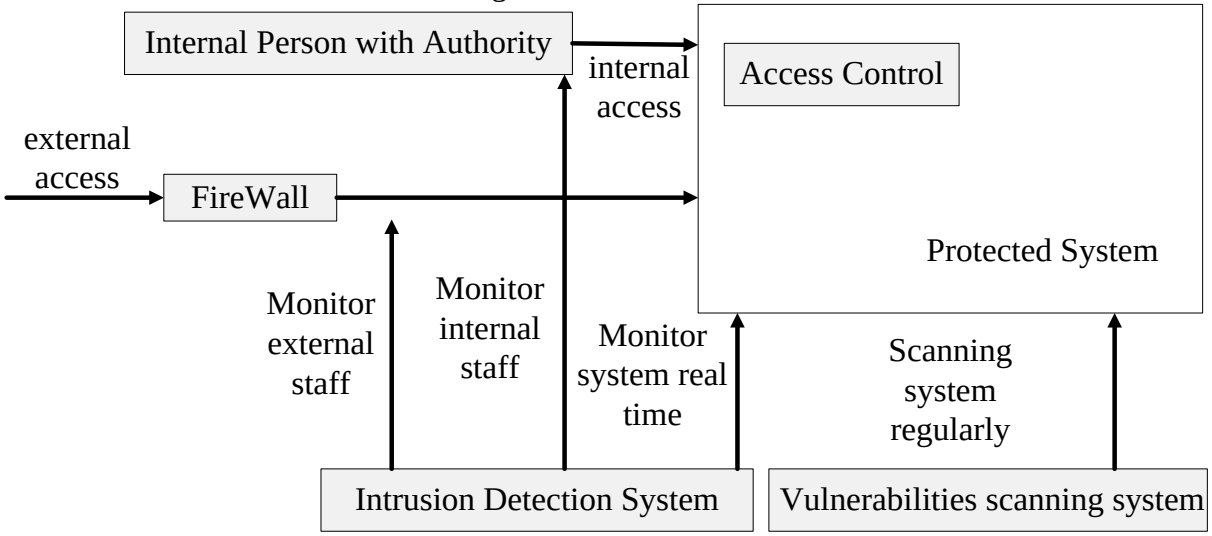

Figure 2. The function of intrusion detection

Data acquisition and preprocessing: the first step of intrusion detection is data acquisition, first from the network or system environment, select the data, and then do a simple pretreatment, it is easy

to detect the module analysis. If the data selection is appropriate, the model will be relatively perfect, the detection of intrusion behavior is relatively accurate. Therefore, the reliability and correctness of the information collected have a great impact on the performance of the intrusion detection system.

2.2 Experimental process and result analysis

Experimental results show that compared with SK-Means algorithm and SFCA algorithm, SML-KNN algorithm has higher detection rate and lower false alarm rate, which indicates that under the guidance of marker data, the algorithm establishes a better anomaly Detection model. At the same time, observing Figure 3, with the increase in the proportion of marker data, the detection rate of the algorithm is gradually increased, false alarm rate decreased significantly, reflecting the semi-supervised learning model, but also shows that more markers, the result will be better.

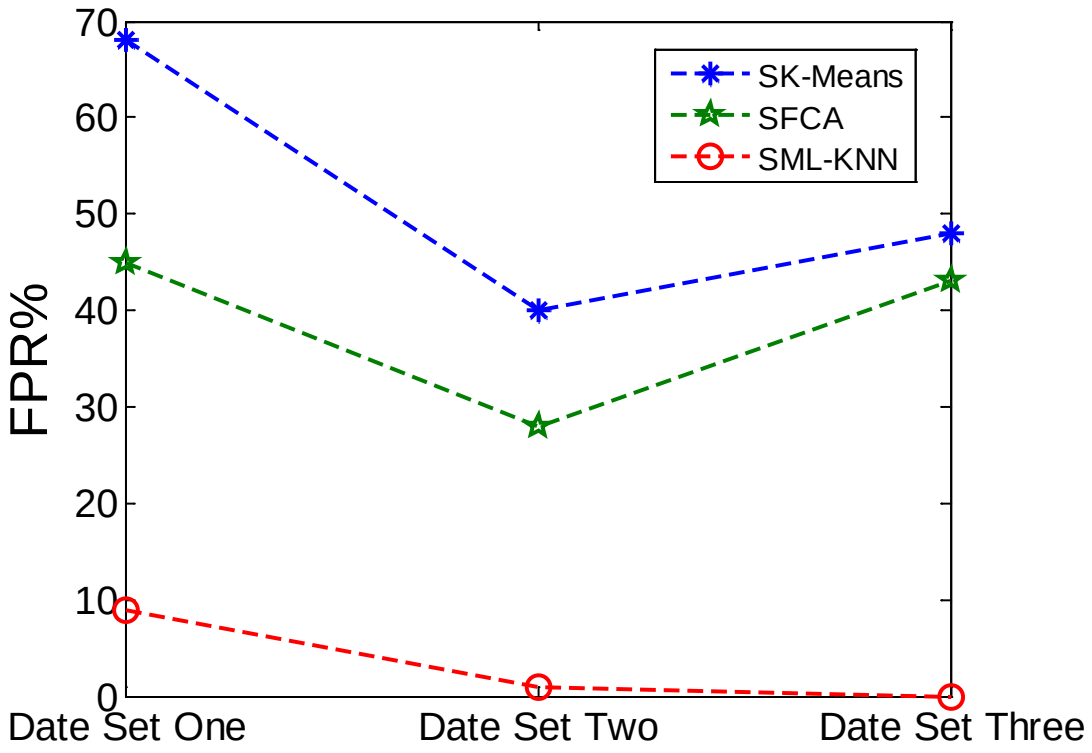

Figure 3. The comparison of FPR

Summary

Intrusion prevention system by the firewall technology and intrusion detection technology development, combined, mainly pre-attack activity and intrusion network traffic to intercept, in order to avoid losses. This paper first introduces the current common network security technology, focusing on the anomaly detection and intrusion prevention system functions and principles. In this paper, the intrusion detection technology is introduced in this paper. According to the latest development of intrusion prevention technology research at home and abroad, the multi-label learning theory, semi-supervised learning and clustering analysis are deeply researched. Intrusion detection is a network security technology using active strategy. Although the traditional intrusion detection algorithm based on supervised learning has high detection precision, it can not detect unknown and intrusion detection algorithm based on unsupervised learning can detect unknown intrusion behavior, but the false positive rate is high.

Acknowledgement

This work was supported by Science and Technology Research Project in Department of Education, Jiangxi Province, 2015 (GJJ151274), Jiangxi Provincial Intellectual Property Soft Science

Research Project (ZR201610), Humanities and Social Sciences Research Project of Jiangxi Province, 2015 (TQ1516) ,Art Science Planning Project of Jiangxi Province (YG2014255, YG2015189) and Pingxiang Science and Technology Support Program, 2015: Research and Development of Animation Rendering Service Platform in the Central Region Based on Cloud Computing.

.

References

[1] Jie, Zhang, and Li Yongzhong. "INTRUSION DETECTION METHOD BASED ON DYNAMIC WEIGHTED SEMI-SUPERVISED CLUSTERING CLOUD MODEL." Computer Applications and Software 3 (2014): 086.
[2] ZHANG, Jie, and Yong-zhong LI. "Dynamic Weighted Intrusion Detection Method Based on Cloud Model and Semi-Supervised Clustering." (2013): 011.
[3] Woźniak, Michał, Manuel Graña, and Emilio Corchado. "A survey of multiple classifier systems as hybrid systems." Information Fusion 16 (2014): 3-17.
[4] A. Tarighat, M. Sadek, A. H. Sayed, “A multi User Beamforming Scheme for Downlink MIMO Channels based on Maximizing Signal-to-Leakage Ratios”, IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 1129-1132, 2005.
[5] Bostani, Hamid, and Mansour Sheikhan. "Modification of supervised OPF-based intrusion detection systems using unsupervised learning and social network concept." Pattern Recognition 62 (2017): 56-72.
[6] K.Wong, R. Cheng, K. B. Letaeif, R. D. Murch, “Adaptive antennas at the mobile and base stations in an OFDM/TDMA system” , IEEE Transactions on Communications, vol. 49, no.1, pp. 195-206, 2001.
[7] Yongzhong, Li, and Zhang Jie. "New intrusion detection algorithm based on cluster and cloud model." Journal of Electronic Measurement and Instrumentation 12 (2014): 016.
[8] Palmieri, Francesco, Ugo Fiore, and Aniello Castiglione. "A distributed approach to network anomaly detection based on independent component analysis." Concurrency and Computation: Practice and Experience 26.5 (2014): 1113-1129.