The detection method of Network attacks in Campus Network

Wang Hongxu

SiChuan Information Technology College , 608040

Keywords: attack detection; Hidden Markov Model; entropy

Abstract. In the process of the detection method of network attacks in campus network, with current algorithm, there are defects such as high demand for data and low detection rate for model. Therefore, a network attack detection method based on improved Hidden Markov Model algorithm is proposed. Structure and characteristics of the campus network is firstly analyzed. Based on the hidden Markov model, the learning of the normal school network behavior sample is then established. On this basis, the information entropy of information theory is introduced to measure the degree of dispersion of source IP addresses. According to the change of the initial stage of Hurst index and entropy threshold is adaptive set to detect the attacks so as to complete attack behavior detection. The experimental results show that the detection method based on the improved hidden Markov model has high accuracy and robustness.

1 Introduction

With network scale and complexity, the university campus network as a special LAN, has now become essential part of the school teaching, scientific research, information exchange, resource sharing, document retrieval and etc. [1。2.3] Accompany great convenience of the campus networks, weaker safety awareness coupled with rougher management system and more open network environment comparing to enterprise network make the campus network face a series of safety issues, such as the external and internal attacks, virus problems, non-authorized access, etc. [4.5.6].

The detection method of network attack behavior is the effective way to solve this problem, which has caused the attention of many experts and scholars [7,8]. Due to the campus network in Colleges and universities in the network attack behavior detection method has profound significance for development, which has become the focus study of the personage and has received wide attention. There are a lot of good methods correspondingly [9.10].

At present, network attack behavior detection is mainly based on distributed algorithms, SVM algorithm and particle algorithm detection method. Among them, particle algorithm is commonly used. However, the current algorithm needs high data and obtains low detection rate.

In view of the above problems, a new method of network attack detection based on the improved hidden Markov model is proposed. The experimental results show that the detection method based on the improved hidden Markov model has high accuracy and robustness.

2 detection principle of network attack

In the detection of network attacks, real-time tracking of network data is needed, and the performance of the network is analyzed. A characteristic profile is extracted from the large amount of network data, the contour is overview for normal data on the network. Once the network data is detected and the profile of dissimilarity exceeds a certain value, it means that the network is in intrusion. Specific steps are as follows:

In the detection process, the sample set X is one that contains a classification of random n variables and the sample set can be defined as information entropy:

$$E(X) = -\sum_{i=1}^{n} p_i \log_2 p_i \qquad (1)$$

Where, p_i presents probability of ith element in X. When n is 1, E takes the minimum value. When the probability of each classification is $\frac{1}{n}$, E takes the maximum value $\log_2 n$.

3. Optimization detection principle of network attack behavior

3.1 Construction of Hidden Markov Model

In optimization process of network attack, normal network data characteristic is obtain from：

$$f = \{ path(a_1, v_1), (a_2, v_2), \ldots (a_N, v_N) \} \quad (2)$$

where $N = 0,1,2\ldots...$

X means training data set， X_P and X_N respectively presents as normal and abnormal training set in X.

In the process of optimizing the network attack behavior in the campus network, the hidden Markov model is expressed as a three tuple:

$$M = (Q, T, E) \quad (3)$$

In the above equation, Q is the state set, T is the network state transfer matrix，from the above equation.

$$p(q_i -\geq q_j) = T(q_i, q_j) \quad (4)$$

The above equation can be defined as transfer probability from the network state q_i to the state q_j. The formula can be derived as

$$p(q_i \uparrow s_i) = E(q_i, s_i) \quad (5)$$

The equation means probability of network state q_i with output s_i.

Two special states I and F of the network are defined as the initial and final state of the network. It is said that the network structure is required to start at I and end at F state. The probability $P(s_T | M)$ is defined with generated s_T by M, a model based on the hidden Markov model is built:

$$P(s_T | M) = \sum_{q_T \in QS} P(I -\geq q_1) p(q_1 \uparrow s_1) \quad (6)$$

$$P(q_1 -\geq q_2) \ldots p(q_M \uparrow s_M) P(q_M -\geq F) \quad (7)$$

Where $s_T = s_1 s_2 \ldots s_M$， $q_T = q_1 q_2 \ldots q_M$ presents network state sequence of s_T， Q_S means possible state sequence set of s_T.

3.2 Implementation of detection design

To optimize the detection process in the network attack behavior, contradiction is that as the source IP addresses in network traffic disperse, concentration degree affected by the attack behavior is very serious. Therefore the information entropy of source IP addresses is employed to measure source IP address distribution. When the entropy is larger, the distribution is more dispersed. Conversely, the source IP address distribution is more concentrated that it is effectively distinguish network attack behavior in campus network. The process can be described by using the following formula.

The information entropy of a sample set X containing a random n variable is defined as:

$$E(X) = -\sum_{i=1}^{n} p_i 1b_{pi} \quad (8)$$

Where, p_i presents probability of ith element in X. When n is 1, E takes the minimum

value 0. When the probability of each classification is $\frac{1}{n}$, E takes the maximum value $\log_2 n$.

SIP is a sample set that contains different n SIP addresses of the source IP address, which can be defined the distribution entropy of the source IP address of the sample set in the following form:

$$E(SIP) = -\sum_{i=1}^{n}\left(\frac{n_i}{s}\right)1b\left(\frac{n_i}{s}\right) \quad (9)$$

Where, $\left(\frac{n_i}{s}\right)$- SIP represents probability of IP address of ith source.

The H value is obtained by the initial N time windows by using the reduction dimension of the network data acquisition:

$$H_{\max} + \frac{1}{N}\left(H_{\max} - \bar{H}\right) \quad (10)$$

where, $H_{\max}$ means maximum of H values, $\bar{H}$ is average value.

After data reduction dimension, $E(SIP)$'s n values are extracted and set as:

$$E_{\max} + \left(E_{\max} - E_{\min}\right) \quad (11)$$

Where, $E_{\max}$ means maximum of $E(SIP)$, $E_{\min}$ means maximum of $E(SIP)$.

4 Simulation experiment

In order to prove the validity of the proposed method based on the improved hidden Markov model algorithm, an experiment is carried out. Experimental data are obtained from the site "oldjun.com" with all the access logs from September 28, 2013 to January 11, 2014. Experimental data set contains 400M text data. Through examination of the artificial and auxiliary program, there are a large number of malicious access attacks, including SQL injection and cross site scripting attacks. The improved algorithm and particle algorithm, as well as the SVM algorithm and distributed algorithm on different data sets to do the detection performance comparison, so as to measure the effectiveness of different algorithms.

Table 1 results of different algorithms Network attack detection

improved algorithm		particle algorithm		SVM algorithm		Distributed algorithm	
dataset	detection rate（%）	False alarm rate（%）	detection rate（%）	False alarm rate（%）	detection rate（%）	False alarm rate（%）	detection rate（%）	False alarm rate（%）
A	99.2	0.02	86	0.5	83	0.65	78	8.9
B	99.3	0.01	84	0.4	83	0.58	79	7.3
C	99.3	0.01	86	0.6	82	0.6 7	78	8.9
D	99.6	0	83	0.4	82	0.69	77	11.2

It is seen from table 1, the detection rate and the false positive rate of the improved algorithm is better than the other algorithms. It is mainly because that the improved algorithm firstly analysis on the structure and properties of campus network. Through learning on campus network samples of normal behavior hidden Markov model is established and induce information entropy in information theory to measure dispersion of the source IP address, according to initial stage of Hurst index and the entropy change adaptively setting threshold to detect the attacks and effectively completed the campus network, which effectively guarantee the accuracy of the detection algorithm.

5 conclusions

In the detection of network attacks in campus network, there are defects such as high demand for data and low detection rate for model. Therefore, a network attack detection method based on improved Hidden Markov Model algorithm is proposed. Structure and characteristics of the campus network is firstly analyzed. Based on the hidden Markov model, the learning of the normal school network behavior sample is then established. On this basis, the information entropy of information theory is introduced to measure the degree of dispersion of source IP addresses. According to the change of the initial stage of Hurst index and entropy threshold is adaptive set to detect the attacks so as to complete attack behavior detection. The experimental results show that the detection method based on the improved hidden Markov model has high accuracy and robustness.

Reference

[1]Wang Yu,Zhang Kun,Liu Jian. Research of wormhole attacks in Ad Hoc networks and the detection method [J]. Computer age, 2014, (2):20-24.

[2]Chang Shuai,Sun Yipin,Wang Yongjun. Research of network attack threatening behavior evaluation Method[J]. Mini-micro computer systems, 2015, 36(1):121-125.

[3]Zhao Na. Many means converged network attacks Detection Technology [J]. Digital Technology and Application , 2014, (9):194-194.

[4]ShiMei. Research and Simulation of Weak Network Attack Signal Effective Detection Method [J]. Computer simulation, 2014, (5):316-318.

[5]Dai Kunyu, HuBin, LeiHao. Research of Application Layer DDoS Detection Method Based on Network Traffic [J]. Microcomputer Applications, 2014, 30(9):17-19.

[6]Li qiang,Li Zhoujun,Zhou Changbin. Sybil Group Attack Detection in Kad Network [J]. Computer Research and Development, 2014, 51(7):1614-1623.

[7]Zhang Ping, Zheng Jin, Chen Xiaojun. Flooding-Attack Defense based on Campus Network Flow Detection [J]. Information Security And Communications Privacy, 2013, (8):80-81.

[8]Teng Liping. Reserch on Sinkhole attacks-based intrusion Detection system in WSN [J]. Computer applications and software2013, 30(6):312-315.

[9]Zhao Pan,Jiang Yubo,Qiu Ling. A New Detection Method of Network Attacks [J]. Sichuan University of Science & Engineering:Natural Sicence Edition, 2014, 27(4):21-23.

[10]Yang Xiaofeng,Li Wei,Sun Mingming. Web attack detection method on the basis of text clustering [J]. CAAI Transactions on Intelligent Systems, 2014, (1):40-46.