High-speed Network Anomaly Information Detection Method with Multiple Time Scale Synchronization

Yu Song
Department of Network Information Management Center, Sichuan University of Science and Engineering, Zigong, 643000, China

Abstract: Aiming at the problem of inaccurate detection of abnormal network information detection methods, a high-speed network anomaly information detection method with multiple time scale synchronization is proposed. The method firstly forms a network traffic time series by collecting the number of IP packets per unit time of the line. Then the Haar wavelet transform is used to decompose the sequence and remove the redundancy to obtain the normal wavelet sequence under the guidance of the "3c" rule of the normal distribution, and it is approximated as Gaussian white noise. Finally, the statistical characteristics of the normal distribution sequence are used to judge the network abnormal information. The results show that compared with the high-speed network anomaly information detection method based on data mining, the anomaly detection accuracy of this method is improved by 9.5%.

Keywords: Multiple time scale; Anomaly information detection; Time series; Haar wavelet transform

1. Introduction

With the development of global information technology, computer and network technologies are becoming more and more popular, which brings great convenience to people's work, study and life. However, the opposite is that there are threats in terms of network security, which brings great trouble to people. Network security aims to maintain the security of data and information in the Internet. To achieve this, we must first protect the security of various hardware resources and software resources[1]. Anomaly information detection is an important part of high-speed network intrusion detection systems. In recent years, researchers have proposed a number of methods through in-depth research on the detection of abnormal network flows. They are classified into statistical methods, data mining methods, and machine learning based methods according to the technologies used. Based on these methods, a large number of applied real-time network anomaly detection systems have been implemented, including NIDES from SRl International, champion E MERALD from KDD CUP 1999, and PHAD based on the packet header. These systems are applied to the actual network, and have achieved good results to some extent, but there are still some shortcomings[2]. First, the detection rate and false positive rate are still relatively high. Using the KDD CUP 1999 data set test, the detection rate of PHAD is only 27%. E MERALD is also only 50%; Second, it cannot effectively distinguish between normal burst traffic and DoS attacks. Normal flow has fractal characteristics, which change after being attacked. This change cannot be reflected at a single scale; Finally, although most systems are multivariate (for example, NIDES), they do not fuse together for multiple variables, or simply combine them, and cannot abstract the intrinsic properties of different features. Aiming at the above problems, a multiple time scale synchronization high-speed network anomaly information detection method is proposed. First, the network traffic time series is formed by the number of collected IP units per unit time. Then the Haar wavelet transform is used to decompose the sequence and use the "3c" rule of the normal distribution to remove the redundancy to obtain the normal wavelet sequence. Finally, based on the statistical characteristics of the obtained normal distribution sequence, the network anomaly information is judged[3]. It is proved by experiments that the method can accurately detect the flow anomaly compared with the traditional anomaly detection method.

2. Network Abnormal Information Detection Methods

Based on the demand of high-speed network anomaly detection and the shortcomings of existing detection methods, a multiple time scale synchronization network traffic anomaly detection method is proposed[4]. Figure 1 is a flow chart and structure diagram of a network traffic anomaly detection algorithm with multiple time scale synchronization. As an example, a two-layer wavelet decomposition is shown in the figure. The algorithm is divided into the following steps:

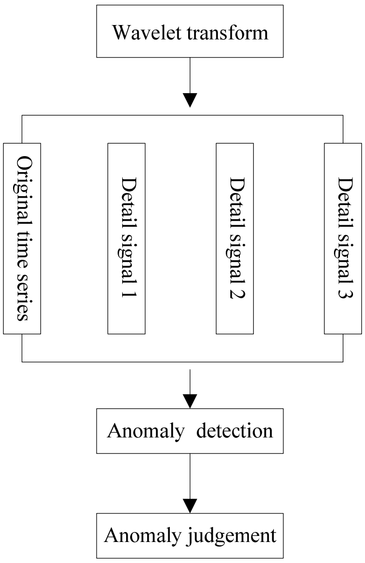

Figure 1. Two-layer wavelet decomposition process

Collecting the number of IP packets per unit time of the line to form a time series of network traffic; Decomposing the flow time series into signals on different time scales using a non-decimated Haar wavelet transform; The abnormality is judged by the statistical properties of the normal distribution sequence [5].

2.1. Composition of time series

Through statistical analysis of traffic, it is found that TCP traffic accounts for the majority of the overall traffic, and many attacks are also directed to TCP, so this paper only considers the feature extraction of TCP traffic.

Currently, the characteristics of network anomaly detection are mainly based on packet level and session flow level. The characteristics of the packet level are mainly the packet length, the number of packets, and so on. The TCP session flow refers to the process of sending a SYN from the host, establishing a connection to the 3rd handshake, and ending the 4th wave connection. In the actual situation, there is also a connection timeout, etc., and the timeout period is set to T. Once there are no packets in the session stream for the time T, the session stream ends. This paper will select features from these two types of features [6].

Definition 1 (active connection) According to the source IP, destination IP, source port, destination port in the TCP connection, the 4-tuple flag indicates a TCP connection. When the host is the connection initiator (send the first SYN request), it is the active connection.

Definition 2 (passive connection) According to the source IP, destination IP, source port, destination port in the TCP connection, the 4-tuple flag indicates a TCP connection. When the host is not the connection initiator (send the first SYN request), it is a passive connection.

For the connection initiated by the host, the frequency of communication between the host and the outside world can be depicted. The active connection generally responds to the characteristics of the client. If there are frequent active connections in the network, it may be a zombie host, which is launching a DDoS attack to the outside world. For the monitoring of the server on the well-known port (with 80, 25 and other ports), it can effectively describe the problem of server network service frequency and service quality [7].

In each statistical window t, the feature quantity F is counted, and in successive N time windows, time series F1, F2, F3, ..., Fn are obtained.

Figure 2~5 are graphs showing the change of uplink traffic, downstream traffic, number of communication IP, and number of TCP connections over time in a passive connection of a server.

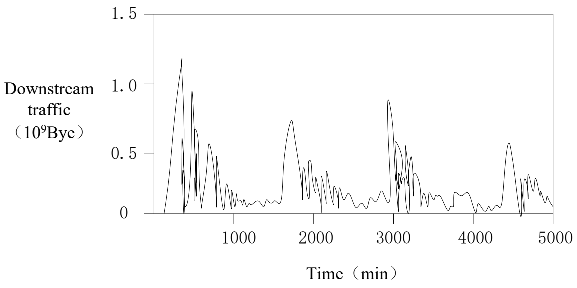

Figure 2. Upstream traffic time series

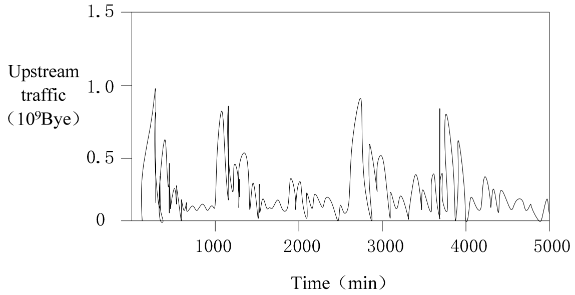

Figure 3. Downstream traffic time series

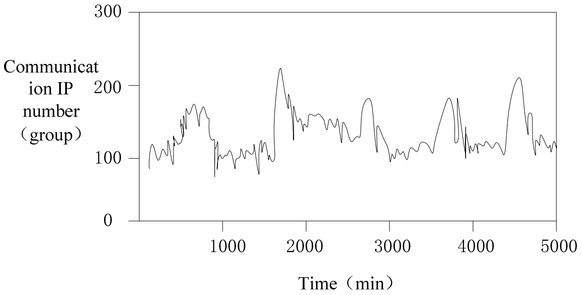

Figure 4. Communication IP quantity time series

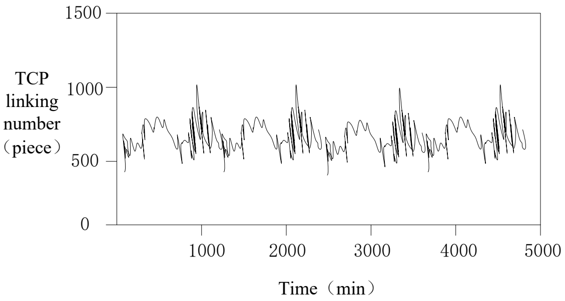

Figure 5. TCP connection number time series

As can be seen from Figures 2 to 5, the number of communication IP has obvious periodicity, and other traffic characteristics also have periodicity. However, because the anomaly is worthwhile, the periodicity is not obvious. After eliminating these abnormal values, the traffic of the entire server will be very stable and periodic[8]. When an abnormal behavior such as attack, scan, or probe occurs on the network, some traffic attributes in the network change. In this paper, 17 features in Table 1 are selected. If not specified, the features in the table need to distinguish between active connection and passive connection.

Table 1. Feature List

Number	Features	Descriptions
1	TCP Count	TCP session establishment number
2	SendPkt Count	Number of upstream packets
3	ReceivedPkt Count	Number of downstream packets
4	Send Len	Upstream traffic
5	Received Len	Downstream traffic
6	IP Count	Number of communication IP
7	Conn per IP	Average number of connections per IP

As shown in Table 1, the traffic characteristics distinguish the active connection and the passive connection according to the TCP connection direction. The TCP session establishment failure is the number of unsuccessful TCP 3 handshake.

2.2. Time series decomposition

Wavelet decomposition and reconstruction typically use a tower algorithm based on multiresolution analysis. In the decomposition process of the original signal, the number of points of the new signal sequence is reduced by half compared with the previous level. Therefore, the tower algorithm cannot establish an intuitive relationship at each time point at each scale. More importantly, the algorithm does not have time-shift invariance. If the initial values of the time series are deleted, the coefficients of the resulting wavelet transform are different from the original ones and need to be recalculated. Therefore, the non-decimated Haar wavelet transform algorithm is used [9].

Wavelet transform is a mathematical tool that has been rapidly developed in recent years. Its applications cover many research fields such as feature selection, data compression and signal processing. Compared with other orthogonal functions, the Haar wavelet function is simple in structure and convenient in calculation. Therefore, the Haar wavelet function is the simplest of all orthogonal functions, and its orthogonal set is a pulse waveform with amplitudes of +1 and -1. In addition, the Haar wavelet function has values in one interval, and the other intervals have zero values. This feature makes the Haar wavelet transform (HWT) faster than other wavelet transforms [10]. Generally used to represent the Haar wavelet function, which is defined as follows:

$$\zeta(t)=\begin{cases}-1, & \frac{1}{2}<t<1\\ 0, & \text{others}\\ 1, & 0<t<\frac{1}{2}\end{cases} \qquad (1)$$

A Haar wavelet basis function is a set of functions consisting of a set of piecewise constant functions, defined as:

$$\sigma(t)=\begin{cases}0, & \text{others}\\ 1, & 0<t<1\end{cases} \qquad (2)$$

Haar orthogonal wavelet transform can be compared to a set of image filtering, the working process is: The signals are respectively input into a decomposition high-pass filter and a decomposition low-pass filter, and a high-frequency component portion (detail information) of the signal is output from the high-pass filter, and a low-frequency component portion (approximation information) of the signal is output from the low-pass filter. The block diagram of wavelet decomposition is shown in Figure 6. The filtering decomposition algorithm utilizes the downsampling method to take only one data point in the two points of the output, and two sequences of half the length of the original signal data are generated, which are recorded as CA and CD [11].

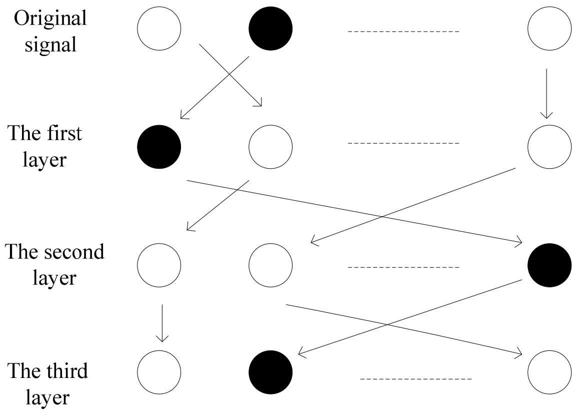

Figure 6. Block diagram of wavelet decomposition

The orthogonal filter in Figure 6 uses a Haar small filter bank. It can be seen that the relationship between the two-channel analysis and the input and output of the integrated filter bank can be expressed as follows:

$$\begin{bmatrix}\alpha_0(z)\\ \beta(z)_1\end{bmatrix}=\frac{1}{2}\sqrt{\zeta(t)\cdot}\sum\sigma(t) \qquad (3)$$

In the formula, $\alpha_0(z)$ is a low-pass filter and $\beta(z)_1$ is a high-pass filter.

2.3. Time series feature analysis

Through the wavelet transform, the original network traffic time series is decomposed into detailed signals and approximate signals on multiple time scales. The network traffic time series has a self-similarity feature, and the wavelet transform coefficient of the self-similar process under the binary grid is a stationary random sequence. Moreover, when the wavelet function is a wavelet basis with integer displacement orthogonally normalized, the wavelet transform has a strong de-correlation effect on the self-similar process, and the wavelet detail signal can be approximated as a white noise sequence[12]. Since the un-decimated Haar wavelet transform used is a redundant wavelet transform, to have the above-mentioned wavelet coefficient feature, the redundancy is first eliminated, so the i_{-th} detail signal is divided into 2i sequences, wherein the k_{-th} sequence is

$$\mathrm{A}_{i,2_i k},(k=0,1,2....) \qquad (4)$$

In the formula, Ai is the detail signal at each scale.

When new data arrives, the wavelet coefficients obtained by the decomposition belong to only one of the sequences, and have no effect on the computational complexity of the algorithm. In Figure 2, taking the first layer of detail signal sequence as an example, the black filled circle and the unfilled circle represent the sequence of two wavelet coefficients formed after the redundancy is removed [13].

The second layer of detail signals decomposed by wavelet is taken as an example to analyze the independence and normality of the wavelet-free detail signal without redundancy. Figure 7 is an autocorrelation analysis diagram of the no-redundant wavelet detail signal. The 95% confidence interval is given in the autocorrelation analysis chart. The autocorrelation coefficient falls into this interval and there is no significant difference from 0. It can be seen that the correlation of the sequence is very weak and can be approximated as an independent distribution.

Figure 8 is a diagram showing the normality of a non-redundant detail signal using a normal probability plot. The curve in Figure 8 can be better fitted with a straight line, so the sequence is substantially consistent with the normal distribution. At the same time, the detailed signal was tested by the Jarque-Bera test, and the normal distribution was observed at a significance level of 0.05. Based on the above analysis, the sequence of wavelet coefficients after redundancy elimination is approximated as Gaussian white noise [14].

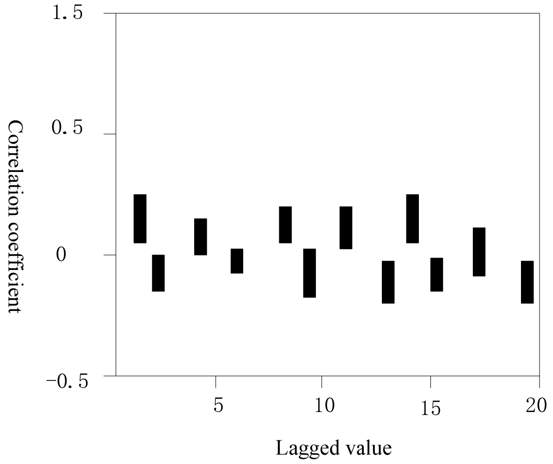

Figure 7. Autocorrelation analysis of the no-redundant wavelet detail signal

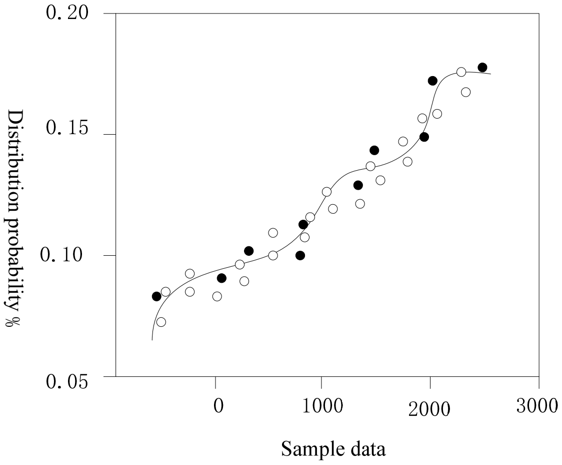

Figure 8. Non-redundant detail signal normal probability map

2.4. Abnormal information judgment

According to the analysis of the distribution characteristics of the detail signal, in the normal network traffic state, the non-redundant detail signal is a stationary sequence with a mean of 0, which can be approximated as Gaussian white noise, and its probability density function obeys a normal distribution. When the time series is abrupt, the detailed signal of the corresponding time scale must deviate significantly from the normal distribution law. Assuming that the non-redundant detail signal sequence A(t) has a mean value of B and the variance is c, then according to the normal distribution law,

$$E\left(A(t)\right) = F\left(\frac{B}{c}\right) \qquad (5)$$

In the formula, F is a distribution function of the standard normal distribution, from which the distribution characteristics of the sequence can be calculated:

$$E\left(-c \le A(t) \le c\right) = 99.56\% \qquad (6)$$

This is the "3c" rule of normal distribution. That is to say, when the wavelet coefficient value deviates from the mean value of 3 times the mean square error, the occurrence of an abnormal event can be judged. The value of the ICJ value a can be adjusted according to requirements, generally taking 3~4. Each Gaussian white noise sequence is provided with a sliding window for storing historical data of the most recent period of time as a basis for judging anomalies. The step size of the window sliding is 1 time point. Considering the timeliness of the data and the saving of the calculation space, the window length is 80, and the long historical data has little influence on the detection accuracy. When an abnormality occurs, the abnormal data will inevitably affect the abnormality detection in the subsequent stage. We use the method of exception elimination, that is, the abnormal wavelet coefficient does not participate in the mean square error calculation of the sequence to ensure that the normal data is used as the criterion for judging the abnormality [15].

The processing of this method at each point in time is mainly divided into two steps. The time complexity of the first step of wavelet decomposition is G(m), and m is the number of layers of wavelet decomposition; The second step is to calculate the mean and standard deviation of the historical data in the sliding window, and the time complexity is still G(m). So the time complexity of the algorithm is G(m). In this method, the space occupied is mainly the data in the history window. Suppose we do m-layer decomposition, we need n sliding windows. The number of historical data saved in each window is constant, so the space complexity of the algorithm is G(2m+N). Since the value of m is small, the actual occupied space is also small.

3. Contrast Experiment

In order to verify the effectiveness of the method, an abnormality detection is performed on a certain server traffic of a university. The original traffic of the server area is obtained by means of switch port mirroring. The data collection time is from April 25, 2016 to April 29, 2016 for a total of 5 days, with a statistical interval of 1 min for a total of 7 200 time periods. The total traffic is about 3.2 TB, and the TCP traffic is about 2.6 TB. The number of TCP session streams is approximately 84.46 million.

Now this method and high-speed network anomaly information detection method based on data mining are used to detect anomaly. The detection results are shown in Table 2.

Table 2. Abnormal Detection Results of the two Methods

Exception	Method in this paper%	Average accuracy%	Method based on data mining%	Average accuracy%
Port scan	98	96	86	86.5
Brute force attack	95		87	
Frequent connection	96		85	
Traffic abnormality	95		88	

It can be seen from Table 2 that the average correct rate of the high-speed network anomaly information detection method with multiple time scale synchronization is 96%; The average accuracy of high-speed network anomaly information detection based on data mining is 86.5%. Compared with the two methods, the accuracy of the former's anomaly detection is improved by 9.5%, which indicates that the method is superior to the high-speed network anomaly information detection method based on data mining.

4. Conclusion

In summary, with the development of society and the advancement of science and technology, Internet technology has been widely used in various fields, which has promoted the development of various industries. However, many problems have arisen, especially the intrusion of abnormal network information is happening frequently. Therefore, in view of the shortcomings of traditional network anomaly information detection methods, a new multiple time scale synchronization high-speed network anomaly information detection method is proposed. It has been verified that this method not only makes up for the shortcomings of the traditional methods, but also on the basis of innovation, improves the ability of the intrusion detection method to respond to security incidents, and guarantees the overall security of the network from many aspects.

5. Acknowledgment

This work is supported in part by 2016 Key base of tourism and scientific research of Sichuan Provincial Tourism Administration (No. ZHZ16-02), and 2017, 2018 Artificial Intelligence Key Laboratory of Sichuan Province (No.2017RYY05, No. 2018RYJ03), and 2017 Horizontal Project (No.HX2017134), and 2015 Teaching Reform Project (No. B11605035).

This work is also supported by Postgraduate Innovation Fund Project by Southwest University of Science and Technology (No. 18ycx115).

References

[1] Wang Bingyu, Sun Qiuye, Ma Dazhong, et al. Multi time scale information physics fusion model of energy Internet [J]. automation of electric power systems, 2016, 40 (17): 13-21.

[2] Xia Shu, Ji Haihua, Zhang Yang. A multi time scale analysis method of power consumption based on electricity information collection system [J]. Shanghai electric power, 2017 (3): 14-18.

[3] Xue Xin. Simulation Research on abnormal information detection in large database [J]. computer simulation, 2017, 34 (8): 399-402.

[4] Fei Jinlong, Wang Yu, Wang Tianpeng, et al. Network anomaly traffic detection based on cloud model [J]. Computer Engineering, 2017, 43 (1): 178-182.

[5] Wu Xiaoping, Zhou Zhou, Li Hongcheng. Research and implementation of network traffic anomaly detection based on unsupervised learning environment under Spark framework. [J]. information network security, 2016 (6): 1-7.

[6] Shi Zhenhua, Liu Waixi, Yang Jia Ye. Network anomaly detection method based on ICMP traffic under SDN architecture, [J]. computer system application, 2016, 25 (4): 135-142.

[7] Hu Ping, ye Kun, Liu Ruiqin. A network traffic anomaly detection method based on Chebyshev [J]. computer application and software, 2016, 33 (5): 127-131.

[8] Jiang Honghong, Zhang Tao, Zhao Xinjian, et al. Large data based anomaly detection mechanism for power information network traffic [J].Telecommunications Science, 2017, 33 (3): 134-141.

[9] Song Xianqiang, Gao Zhong He, Liu Long, et al. Research on network anomaly detection method based on data mining [J]. electronic technology, 2016, 45 (11).

[10] Li Rui, Zhang Luqiao, Li Haifeng, et al. Summary of network anomaly traffic detection based on entropy [J]. application of computer systems, 2017, 26 (6): 36-39.

[11] Yuan Jing, Zhang Yu Jin. Application of sparse noise cancellation self coding network with gradient difference information in abnormal behavior detection [J]. automation journal, 2017, 43 (4): 604-610.

[12] Xu Gang, Wang Zhan, Zang Dawei, et al. Data center network anomaly detection algorithm based on link state database [J]. computer research and development, 2018, 55 (4): 815-830.

[13] Jin Renjie, Wang Yu, Han Weijie. Abnormal traffic [J]. software based on flow template detection, 2017, 38 (4): 121-126.

[14] Mi Hong, Yang Xi Bei. Network anomaly data detection model based on intrusion feature selection [J]. modern electronic technology, 2017, 40 (12): 69-71.

[15] Wang Ying. Wireless network traffic anomaly data detection simulation [J]. computer simulation, 2017, 34 (9): 408-411.