2012 International Conference on Innovation and Information Management (ICIIM 2012)
IPCSIT vol. 36 (2012) © (2012) IACSIT Press, Singapore

Time-Frequency Detection Algorithm of Network Traffic Anomalies

Dingde Jiang[1, 2, a+], Wenda Qin[1], Laisen Nie[1], Cheng Yao[1], and Rongfang Lin[1]

[1]College of Information Science and Engineering, Northeastern University, Shenyang 110819, China

[2]State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract. Abnormal network traffic results in a very great harm to our current communication networks, so we need to quickly detect anomalous or malicious traffic in a network. However, the existing detection methods hold a lot of computational overhead and detection errors, which will make it significantly difficult to meet the detection requirements of anomalous network traffic. This paper presents a novel detection algorithm of network traffic anomalies based on time-frequency analysis and statistic theory. We use the discrete wavelet transform to capture the time-frequency nature of network traffic. And then the nature of its low and high frequency is discussed in details. We further analyze its statistic properties in the different scales based on the sliding time window. At the same time, the abnormal nature of network traffic is extracted and the anomaly detection is performed correctly for network traffic. Simulation results indicate that our detection algorithm is feasible and effective.

Keywords: network traffic; time-frequency analysis; anomaly detection; statistic analysis

1. Introduction

With the rapid development of communication networks, network attacks, network viruses, and other abnormal network behavior often lead to network performance degradation or even make our networks impossible to operate normally [1]. Hence, how to establish an effective network anomaly detection approach to ensure the network communications efficient and reliable is currently becoming a hot research topic. Generally speaking, network anomaly behaviors often cause abnormal changes in the network traffic, so how to accurately detect abnormal traffic is an effective and feasible way for network operators to detect network anomalies, network fault, and network attacks [2, 3]. However, compared to the background traffic, the abnormal traffic is very small, holds unexpected features [4, 5], and shows short duration nature in the time [6, 7]. Therefore, nowadays researchers are interested in finding an accurate detection approach to the abnormal network traffic.

Traffic anomaly detection uses active detection methods to detect anomalies in the network traffic. Piotr et al. [3] used the combined use of wavelet transforms and change-point detection algorithms in order to detect the instants when fractality changes noticeably. Wang et al. [5] proposed a ternary content addressable memory coprocessor based solution for high speed, integrated TCP flow anomaly detection and policy filtering. Kriangkrai et al. [8] used a statistic-based anomaly detection method. Abdun et al. [9] exploited a resource conserving sampling technique to improve detection of less frequent patterns from huge network traffic under the fixed data storage capacity of the system. Marius [10] used the analytical discrete wavelet transform and high-order statistic analysis to detect network traffic anomaly. Anitha [11] used active and passive analysis mechanisms when deployed at the host and active analysis at checkpoints.

[+] Corresponding author
E-mail address: [a]jiangdingde@ise.neu.edu.cn

Different from the above methods, this paper proposes a novel detection algorithm of network traffic anomaly based on time-frequency analysis and statistic theory. Firstly, the discrete wavelet transform is exploited to capture accurately the time-frequency characteristics of network traffic. By decomposing network traffic into the different scales, we discuss and analyze the properties of its low and high frequency in details. Secondly, we further analyze its statistic properties in the different scales based on the sliding time window. And according to every sliding time window, we calculate the sample variation in the given time. Consequently, we accurately obtain the variation series in the corresponding scale. Thirdly, we computer the mean and variation of this series and decide the threshold of extracting the scale feature of anomalous network traffic. By capture the scale nature of abnormal traffic, we reconstruct its time signals. And then we can carry out the accurate anomaly detection. Finally, we use the traffic data from the real network to validate our algorithm. Simulation results indicate that our detection algorithm is feasible and effective.

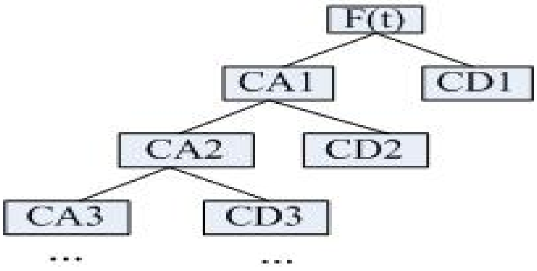


Fig.1: Discrete wavelet decomposition of signals.

2. Problem Statement

Network traffic has highly dynamic characteristics, which is difficult to be depicted. Moreover, network traffic also holds the temporal correlations. How to capture its inherent properties and extract the needed nature is a challenge. For this purpose, we employ the time-frequency analysis to analyze its joint time and frequency feature and perform the feature extraction. Here, we utilize the discrete wavelet transform to carry out the time-frequency analysis and feature extraction. Through decomposing network traffic into the different scales, its scale nature and time- frequency feature can be exhibited exactly. This is helpful to seize the abnormal properties of network traffic.

For a time series $f(t)$, its discrete wavelet transform can be denoted as follows:

$$F(a,b)=\int_{-\infty}^{+\infty} f(t)\varphi_{m,n}(t)dt, \quad and \quad \varphi_{m,n}(t)=a_0^{-m/2}(a_0^{-m}t-nb_0) \tag{1}$$

where a represents a scale and b denotes the position, and they are satisfied with:

$$a=a_0^{m}, \qquad and \qquad b=na_0^{m}b_0 \tag{2}$$

According to the wavelet analysis theory, for a discrete wavelet transform $F(a,b)$, its inverse transform can easily be expressed as:

$$f(t)=\sum_{m,n} F(a,b)\tilde{\varphi}_{m,n}(t) \tag{3}$$

Fig.1 shows the signal decomposition process of the discrete wavelet transform. We can see from Fig. 1 that the time signal $f(t)$ is decomposed for high frequency and low frequency parts. Furthermore, if these parts are to be decompose into the next level, only the low frequency part is divided into the low frequency and high frequency components. And thus Equation (1) can be denoted into:

$$G(m,j)=F(a,b)=\int_{-\infty}^{+\infty} f(t)\varphi_{m,n}(t)dt, \qquad \text{and} \qquad \varphi_{m,n}(t)=a_0^{-m/2}(a_0^{-m}t-nb_0) \tag{4}$$

where m represent the decomposition level, j=1 and j=0 respectively stand for the low frequency signal and high frequency signal of the *mth* decomposition level.

Because of the orthogonality of the discrete wavelet function, the interaction of two points caused by redundant signals is eliminated. At the same time, this orthogonality also makes the less errors. Thereby, the time-frequency analysis can be better to reflect the nature of the signal itself. Through the discrete wavelet transform showed in Fig.1 and Equation (1), the signal $f(t)$ can be transformed into the time-frequency signals of the different scales, namely,

$$f(t) \Rightarrow [G(1,0), G(1,1)], [G(2,0), G(2,1)], \dots \tag{5}$$

where $G(i,0)$ and $G(i,1)$ are the low frequency signal and high frequency signal in the *ith* decomposition level.

Generally, the high frequency parts of the discrete wavelet transform describe the rapid changes of network traffic and the low frequency parts denote the trend of network traffic. However, the abnormal traffic can not only change slowly in the time, but also can vary quickly. To accurately capture these characteristics of network traffic in the different scales, we analyze its low and high frequency characteristics concurrently.

To seize the abnormal nature of network traffic, we discuss the variance of the low and high frequency signal in the different scales. And based on sliding window, we propose a variance analysis method to extract the abnormal of network traffic. Assume that the discrete wavelet transform $G(i,j)$ (where $i = 1,2,\dots$; $j = 0$ or 1) is a series with the size k . and then it can be formulated into:

$$G(i,j) = [g(i,j,1), g(i,j,2), \dots, g(i,j,k)] \tag{6}$$

And then the variance of $G(i,j)$ at the time t can be described by the statistic variance in the sliding time window with the size h, namely

$$v(t) = \frac{1}{h-1} \sum_{z=t-\frac{h-1}{2}}^{t+\frac{h-1}{2}} \left| g(i,j,z) - \overline{g}(t) \right|^2, \quad and \quad \overline{g}(t) = \frac{1}{k} \sum_{z=t-\frac{h-1}{2}}^{t+\frac{h-1}{2}} \{ g(i,j,z) \} \tag{7}$$

The sliding time window at the time t can be expressed as:

$$h: \quad \left[t - \frac{h-1}{2}, \dots, t, \dots, t + \frac{h-1}{2} \right] \tag{8}$$

Equation (7) means that *v(t)* denotes a time series related with the time t. According to Equation (7), as mentioned in [6], we use the statistic analysis to decide the detection threshold of the series sequence $v(t)$ And then we extract the low and high frequency characteristics of network traffic in the different scales. Based on Equation (3), we get a new time series $\zeta(t)$ by the time reconstruction. Similarly, we determine a detection threshold of $\zeta(t)$ and identify when abnormal traffic takes place.

The below algorithm steps indicate the detailed detecting process of network traffic.

Step 1..Give the network traffic $f(t)$ and the decomposition threshold β of the discrete wavelet transform, and set the decomposition level i = 1.

Step 2. According to Equations (1) and (4), perform discrete wavelet transform for $f(t)$. And obtain low frequency signal series $G(i,0)$ and high frequency signal series $G(i,1)$ showed in Equation (5).

Step 3. Calculate the statistic variance of sand $G(i,1)$ by Equation (7), respectively. And then obtain a new time series.

Step 4: Discuss the statistic nature of $v(t)$ extract the specious abnormal parts of network traffic, and reconstruct it into a new time signal $\zeta(t)$.

Step 5. Analyze the statistic nature of $\zeta(t)$ and decide the detection threshold α.

Step 6. If $\zeta(t) > \alpha$, then diagnose $f(t)$ into the abnormal traffic at the time t.

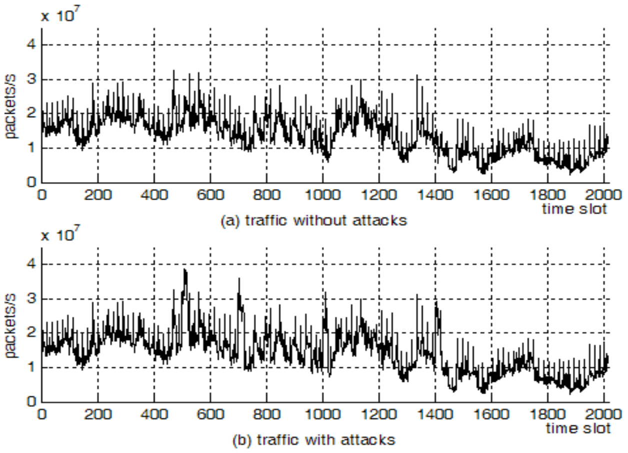
Step 7. If the current decomposition does not arrive at the decomposition threshold β, then set $f(t) = G(i,0)$ and $i = i+1$, and go back to Step 2.

Step 8. Save the detection results to files and exit the detection process.

3. Simulation Result and Analysis

To verify the effectiveness of our algorithm, we exploit the real traffic data from the Abilene backbone network and the background traffic without attacks. And we utilize the DDoS tool in the local area network to emulate the attack traffic. By injecting the emulated attack traffic into the real background traffic, we synthesize the abnormal network traffic. To validate the performance, we inject the attack traffic in the irregular interval, namely [500,519], [700,719], [1000,1019], [1400,1419].

Fig. 2 plots out the network traffic without attacks and network traffic with attacks. From Fig.2, we can not find the difference between them in these time zones. Hence, it is very difficult to detect anoma- lous components in network traffic only in the time domain.



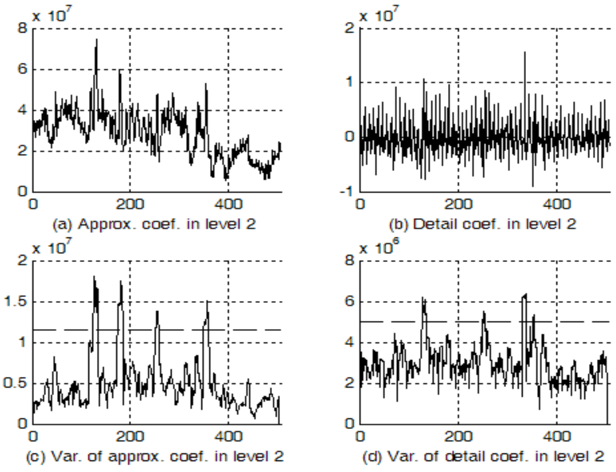


Fig.2 Network traffic without and with attacks.

Fig.3 Discrete wavelet decomposition in level 2.

Figs. 3 and 4 shows the discrete wavelet decomposition in level 2 and the corresponding time signal reconstructed, respectively. Form Fig. 3(a) and (b), we easily find that although the network traffic signal is decomposed in level 2, its low and high frequency characteristics are not obvious. That is to say, we can not diagnose the anomalous parts of network traffic only by the discrete wavelet transform in level 2. Fig. 3(c) and (d) illustrate the statistic variance of the low and high frequency parts corresponding to Fig. 3(a) and (b), respectively. Moreover, Fig. 3(c) and (d) also indicate that we can recognize exactly the abnormal parts of the low and high frequency traffic signals in level 2. Obviously, this shows that the anomalous nature of network traffic has been exhibited by its scale and time-frequency feature.

Fig.4 illustrates the reconstructed time signal by the extracted nature of network traffic corresponding to Fig. 3(c) and (d). From Fig. 4, we can see that the reconstructed signal can embody when the traffic anomalies appear. As a result, we can accurately detect and find the abnormal components of network traffic. More importantly, we can find out all the abnormal parts only in level 2. This shows that our method is effective.

Likewise, we continue to decompose the network traffic in level 3 to validate the detection performance of our algorithm. Figs. (5) and (6) show the decomposition in level 3 and the corresponding reconstructed, respectively signal. We find that although all the abnormal properties can not be extracted, all the anomalous parts can still be detected accurately. This further indicate that our algorithm is feasible and promising.

4. Summary

This paper presents a new anomaly detection algorithm of network traffic based on time-frequency analysis and statistic theory. The discrete wavelet transform is exploited to capture the time-frequency nature and scale feature of network traffic. And then we can easily decompose it into the low and high frequency signals. By analyzing its statistic properties in the different scales based on the sliding time window, we can easily capture and extract the abnormal nature of network traffic. Consequently, we can accurately detect when the abnormal traffic takes place. Simulation results illuminate that our algorithm is promising.

5. Acknowledgement

This work was supported in part by the National Natural Science Foundation of China (Nos. 61071124, 61172051 and 60802023), the Fok Ying Tung Education Foundation (No. 121065), the Specialized Research

Fund for the Doctoral Program of Higher Education (No. 20100042120035), and the Program for New Century Excellent Talents in University (No. 08-0095)), the Fundamental Research Funds for the Central Universities (No. N090404014), the Open Project of State key Laboratory of Networking and Switching Technology (No. SKLNST-2009-1-04), the China Postdoctoral Science Foundation (No. 20110491516), and the Northeastern University Postdoctoral Science Foundation (No. 20110622).The authors wish to thank the reviewers for their helpful comments.

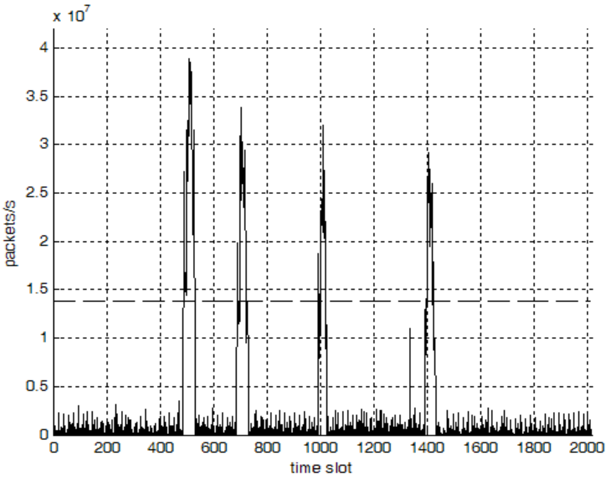


Fig.4 Reconstructed time signal from level 2.

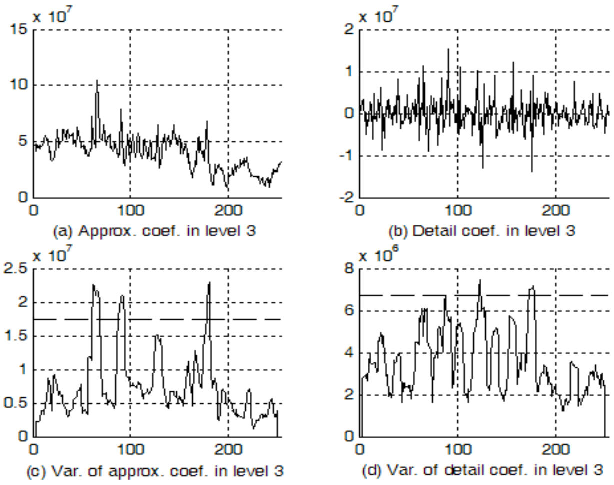


Fig5 Discrete wavelet decomposition in level 3.

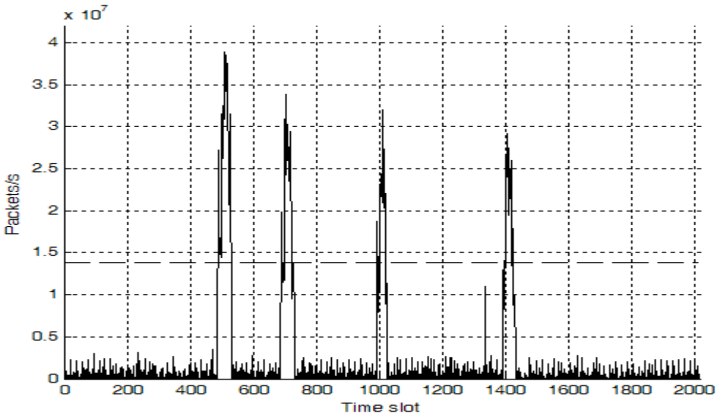


Fig.6: Reconstructed time signal from level 3.

6. References

[1] A. Lakhina and M. Crovella. Mining anomalies using traffic feature distributions. Computer Communication Review, vol. 35, no. 4, pp. 217-228, Oct. 2005.

[2] D. Jiang, J. Chen, and L. He. An accurate approach of large-scale IP traffic matrix estimation. IEICE Transactions on Communications, vol. E90-B, no. 12, pp. 3673-3676, 2007.

[3] P. Zuraniewsk and D. Rincon. Wavelet transforms and change-point detection algorithms for tracking network traffic fractality. Proc. of NGIDE'06, 2006, pp. 216-223.

[4] L. Guo, J. Cao, H. Yu, and L. Li, “Path-based routing provisioning with mixed shared protection in WDM mesh networks,” J. Lightwave Technol., 2006, 24: 1129-1141.

[5] Z. Wang, H. Che, J. Cao, et al. A TCAM-based solution for integrated traffic anomaly detection and policy filtering. Compuer Commucations, vol. 32, pp. 1893-1901, Nov. 2009.

[6] D. Jiang, X. Wang, and L.Guo. An optimization method of large-scale IP traffic matrix estimation. AEU-International Journal of Electronics and Communications, vol. 64, no. 7, pp. 685-689, 2010.

[7] L. Guo, “LSSP: A novel local segment shared protection for multi-domain optical mesh networks,” Computer Commun., 2007, 30: 1794-1801.

[8] K. Limthong, P. Watanapongse. A wavelet-based anomaly detection for outband network traffic. Proc. of APSITT'10, 2010, pp. 1-6.

[9] A. N. Mahmmood, J. Hu, Z. Tari, et al. Critical infrastructure protection: Resource efficient sampling to improve detection of less frequent patterns in network traffic. Journal of Network and Computer Applications, vol. 33, pp. 491-502, Jul 2010.

[10] M. Salagean. Real network traffic anomaly detection based on analytical discrete wavelet transform. Proc. of OPTIM'10, 2010, pp. 926-931.

[11] R. Anitha. Detecting keyloggers based on traffic analysis with periodic behaviour. Network Security, pp. 14-19, Jul 2011.