A feature analysis approach to network traffic in communication networks

Hong-Hao Zhao[1], Hong-Yu Dong[2], Xiao-Hui Zhang[2], Qing Ye[2], Fan-Bo Meng[1,*], Xin-Yu Sun[3] and Ying Cao[3]

1.State Grid Liaoning Electric Power Company Limited, Shenyang 110006, China
2.State Grid Fuxin Electric Power Supply Company, Fuxin 123000, China
3.Liaoning Planning and Designing Institute of Post and Telecommunication Company Limited, Shenyang 110011,China
Email:amengfb@163.com
**Corresponding author*

Network traffic holds the highlighting dynamic features. How to accurately characterize the hidden properties of network traffic has an important impact on network activities, such as anomaly detection and performance analysis. This paper proposes a feature analysis approach to describe network traffic. Firstly, the wavelet packet transformation is used to extract the multi-scale feature of network traffic. Then the principal component analysis method is exploited to refine the hidden features of network traffic in the time-frequency domain. Finally, to validate our feature analysis method, an anomaly detection test is conducted. Simulation results show that our approach is promising.

Keywords: Network Traffic; Feature Analysis; Time-Frequency Analysis; Feature Extraction.

1. Introduction

With the advance of new network technologies, the new traffic types and features have appeared in current communication networks. For some new applications, new applications cause novel traffic patterns and features. This affects network performance while traditional networks do not consider these new features [1-3]. Moreover, traffic anomalies have an important impact on users' experience and networks, such as new types of attacks, novel anomaly patterns, unknown hidden traffic nature [3-4]. Hence, how to capture network traffic features is very important for operators and users. The feature analysis of network traffic has become a hot topic in academic and industries so far [6-8].

The generalized entropy and information distance metrics were used to detect the low-rate distributed denial-of-service attack behaviors [3]. The

aggregate traffic statistics and the distributed spatial detection method were used to recognize network anomalies [6]. The feature analysis was used to diagnose anomalous network traffic [2]. Moreover, by analyzing network traffic feature, the model was built to detect network events [7]. The index of variability was used to describe network traffic [9]. The TCP traffic abnormal problem in routers was studies [10]. The wavelet transform was utilized to describe the multi-scale features of network traffic [8]. The time-frequency analysis was used to extract network traffic nature [11]. The periodicity-based anomalies in network traffic was studied and identified network anomalies [12]. Additionally, from the network-wide perspective, the anomalous network traffic was correctly detected via signal transformations [5]. The aggregate traffic statistics was employed to detect network anomaly behaviors [13]. These methods can capture the features of network traffic. However, they hold the larger errors.

This paper proposes a feature analysis approach to describe and capture network traffic in current communication networks, such as wired and wireless networks or mixed networks, which support latest network applications. Firstly, we use the signal analysis theory to extract the features of network traffic because network traffic can be taken as time signals. Due to the multi-scale and high resolution description capability of the wavelet packets, we exploit the wavelet packet transformation to extract the hidden feature of network traffic. Secondly, after performing the wavelet packet transformation for network traffic, the principal component analysis method is exploited to further refine the features of network traffic in the time-frequency domain. Finally, we propose a feature extraction algorithm to capture the hidden features in network traffic. To validate our feature analysis method, an anomaly detection test is conducted. Simulation results show that our approach is feasible and effective.

The rest of this paper is organized as follows. Our method is derived in Section 2. Section 3 presents the simulation results and analysis. We then conclude our work in Section 4.

2. Problem Statement

In the network, there exist many traffic flows from source nodes to destination nodes. The traffic of these flows exhibits some relationship such as temporal correlations. This leads to high complex of network traffic. Conversely, these features are exactly used to help capturing network traffic features. Like the general time sequence, network traffic changes over the time. Accordingly, network traffic can be regarded as the time signal to handle. In this case, the general signal processing and analysis approaches can be exploited network traffic. The wavelet packet analysis is powerful in extracting multi-scale features and selecting distinct time-frequency resolutions for time signals. Thus, the

wavelet packet is used to firstly handle network traffic. Without loss of generality, for network traffic $x_{ij} = \{x_{ij}(1), x_{ij}(2), ...\}$ from source node i to destination node j, we perform the below wavelet packet transformation:

$$d_l^{k,2n} = \sum_s a_{s-2l} d_s^{k+1,n}, \quad d_l^{k,2n+1} = \sum_s b_{s-2l} d_s^{k+1} \tag{1}$$

$$f_{ijk}^n(t) = \sum_l d_l^{k,n} x_{ij}^n(2^k t - l) = F(d_l^{k,n}, x_{ij}(t)) \tag{2}$$

where $f_{ijk}^n(t) \in U_{ijk}^n$, U_{ijk}^n denotes the subspace standing for the scale and wavelet space. According to the decomposition of the wavelet packet method, we use $\{d_l^{k,n}\}$ to attain $\{d_l^{k,2n}\}$ and $\{d_l^{k,2n+1}\}$ via the above equation (1).

According to the wavelet packet method, the reconstruction and reverse transformations of the wavelet packet can be expressed as:

$$d_l^{k+1,n} = \sum_s [h_{l-2s} d_s^{k,2n} + g_{l-2s} d_s^{k,2n+1}] \tag{3}$$

$$x_{ij}(t) = F^{-1}(d_l^{k+1,n}) \tag{4}$$

By (3)-(4), we can exploit $\{d_l^{k,2n}\}$ and $\{d_l^{k,2n+1}\}$ to compute $\{d_l^{k+1,n}\}$. Thus, we can reconstruct the original network traffic signal $x_{ij}(t)$. It is clear that according to Equation (1), network traffic signal $x_{ij}(t)$ exhibits different scale features the scale space and the wavelet space. This is embodied via the wavelet packet coefficients $\{d_l^{k,n}\}$, which exhibits the distinct time-frequency feature. In general, network traffic holds different characteristics at different times and frequencies. For the wavelet packet coefficients $\{d_l^{k,n}\}$, the below low- and high-frequency components are obtained:

$$d_{low} = \{d_l^{k+1,0}, ..., d_l^{k+1,h}\} \tag{5}$$

$$d_{high} = \{d_l^{k+1,h+1}, ..., d_l^{k+1,N}\} \tag{6}$$

To get the time-domain signals in (4) via (3), (4) can be converted as:

$$\hat{d}_{low} = \{d_l^{k+1,0}, ..., d_l^{k+1,h}, \underbrace{0, ..., 0}_{N-(h+1)}\} \tag{7}$$

$$\hat{d}_{high} = \{\underbrace{0, ..., 0}_{h+1}, d_l^{k+1,h+1}, ..., d_l^{k+1,N}\} \tag{8}$$

By Equation (3), we can attain the time signals in (5) as follows:

$$x_{ij,low} = F^{-1}(\hat{d}_{low}) \tag{9}$$

$$x_{ij,high} = F^{-1}(\hat{d}_{high}) \tag{10}$$

where $x_{ij,low}$ and $x_{ij,high}$, respectively, represent the low- and high-frequency time signals for network traffic x_{ij}. Then we perform principal component analysis for Equation (9) to attain the principal and non-principal time signals of the low-frequency time signal $x_{ij,low}$, as mentioned in [8].

$$x_{ij,low} = \underbrace{R_{p,low}V_{p,low}^T}_{principle\ component} + \underbrace{R_{n,low}V_{n,low}^T}_{non-principle\ component} \tag{11}$$

$$x_{ij,low} = U_{low}D_{low}V_{low}^T \tag{12}$$

Likewise, we attain the principal and non-principal time signals of the high-frequency time signal $x_{ij,high}$ as follows:

$$x_{ij,high} = \underbrace{R_{p,high}V_{p,high}^T}_{principle\ component} + \underbrace{R_{n,high}V_{n,high}^T}_{non-principle\ component} \tag{13}$$

$$x_{ij,high} = U_{high}D_{high}V_{high}^T \tag{14}$$

Then we can get the feature model of network traffic as shown in Equations (11)-(14). The detailed steps of the feature analysis algorithm is as follow:

Step 1: Give the initial traffic matrix x_S and the number of wavelet packet transformations n_scale.

Step 2: According to Equations (1)-(2), carry out the wavelet packet transform. Then obtain the wavelet packet coefficients $\{d_l^{k,n}\}$.

Step 3: According to Equations (5)-(8), divide $\{d_l^{k,n}\}$ into low- and high-frequency components, $\hat{d}_{low}$ and $\hat{d}_{high}$.

Step 4: By (3)-(4), perform the converse transformation of wavelet packets and the time-domain signals $x_{ij,low}$ and $x_{ij,high}$ corresponding to $\hat{d}_{low}$ and $\hat{d}_{high}$.

Step 5: By the principal component analysis, attain eigenvector matrix, the diagonal matrix describing the energy spectrum, and the eigenflow matrix, U_{low}, D_{low}, V_{low}, and U_{high}, D_{high}, V_{high} for $x_{ij,low}$ and $x_{ij,high}$, respectively.

Step 6: According to principal component analysis, extract the k top principal components, and then obtain the parameters of the model about network traffic, $V_{low}^{'}$, $D_{low}^{'}$, $V_{high}^{'}$, $D_{high}^{'}$.

Step 7: Extract principal components $x_{ij,low,p}$ and $x_{ij,high,p}$ from $x_{ij,low}$ and $x_{ij,high}$ via the models.

Step 8: According to $x_{ij,low,p}$ and $x_{ij,high,p}$, attain new principal component $x_p = x_{ij,low,p} + x_{ij,high,p}$.

Step 9: Save the results to file and exit.

3. Simulation Result and Analysis

Now we carry out some numerical experiments to verify our feature analysis approach to network traffic in communication networks. From Fig. 1, we can easily see that there exists nearly no difference between normal and abnormal network traffic. In our experiments, abnormal network traffic in Fig. 1(b) is attained via adding an abnormal network traffic to normal network traffic in Fig. 1(a). It is very clear that we cannot nearly directly detect and diagnose the anomaly components of network traffic in Fig. 1(b).

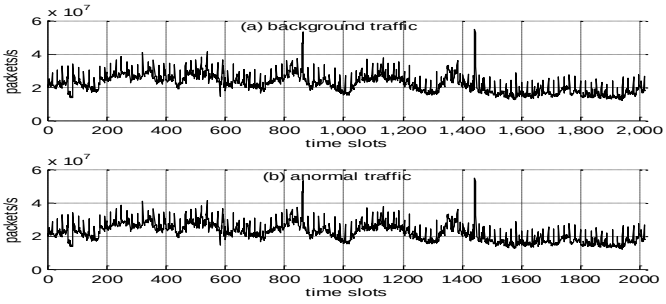

Fig. 1 Network traffic without and with abnormal properties

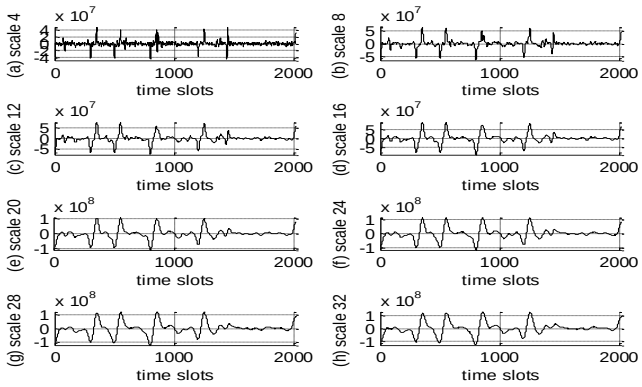

Fig. 2 Wavelet packet transformations with different scales

Fig. 2 indicates that for different transformation scales, network traffic exhibits different time-frequency features. This demonstrates that our approach can use the wavelet packet analysis to extract the features of network traffic in different scales. It is more interesting that Fig. 2(a) shows the high-frequency property in scale 4. For scales 8, 12, and 16 in Fig. 2(b)-(d), we can effectively

capture the medium-frequency nature of network traffic. However, for other scales in Fig. 2(e)-(h), the low-frequency features of network traffic can be exactly extracted. Hence, this indicates that our approach can effectively capture the features of network traffic in the time-frequency domain.

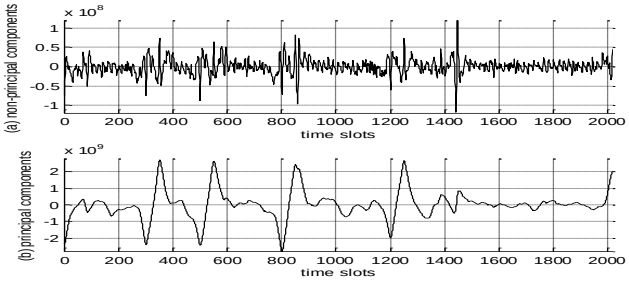

Fig. 3 Traffic feature extraction via our approach

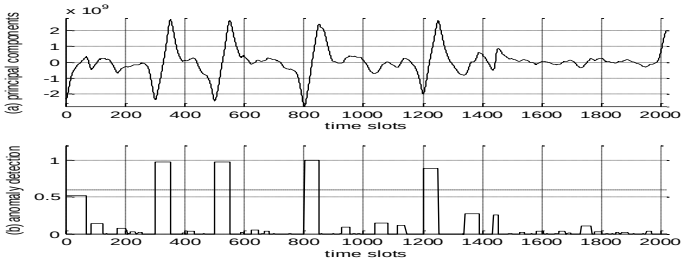

Fig. 4 Anomaly detection results

Fig. 3 illustrates the principal component feature of network traffic via our approach based on the principal component analysis. From Fig. 3, we can see that the principal component of network traffic is correctly extracted. More importantly, the dramatic changes in principal components in network traffic denotes the possible anomalies. This will help us perform the effective detection of network traffic. This also indicates that our approach can effectively capture and characterize network traffic.

Besides, we discuss the anomaly detection ability to further justify our approach. Fig. 4 plots the anomaly detection results of our approach. In our simulation, we inject the abnormal traffic in four times with the duration of 50 unit time slots, namely at times 300, 500, 800, and 1200, respectively. Fig. 4 shows that our approach can exactly detect the abnormal components of network traffic in different time slots. This further states that our approach can effectively extract anomalous components in network traffic and carry out the accurate network traffic detection.

4. Conclusion

Network traffic holds the dynamic features. How to accurately characterize the hidden properties of network traffic has an important impact on network activities, such as network failure positioning, anomaly detection, and performance analysis. To this end, this paper propose a feature analysis approach to describe network traffic. Firstly, the wavelet packets transformation is used to extract the multi-scale feature of network traffic. Then the principal component analysis method is exploited to refine the hidden features of network traffic in the time-frequency domain. Finally, to validate our feature analysis method, an anomaly detection test is conducted. Simulation results show that our approach is promising.

References

1. I. C. Paschalidis et al, Spatio-temporal network anomaly detection by assessing deviations, IEEE Trans. Netw., 2009, 17(3): 685-697.
2. D. Jiang, et al., A traffic anomaly detection approach in communication networks for applications, Multimed. Tools Appl., 2016, online available.
3. Y. Xiang, et al., Low-rate DDoS attacks detection and trace back by using information metrics, IEEE Trans. Inf. Forensic Secur., 2011, 6(2): 426-437.
4. W. Xiong, H. Hu, N. Xiong, et al., Anomaly secure detection methods by analyzing dynamic characteristics of the network traffic in cloud communications, Inf. Sci., 2014, 258(2014): 403-415.
5. D. Jiang, et al., A transform domain-based anomaly detection approach to network-wide traffic, J. Netw. Comput. Appl., 2014, 40(2): 292-306.
6. G. Thatte, U. Mitra, and J. Heidemann, Parametric methods for anomaly detection in aggregate traffic, IEEE Trans. Netw., 2011, 19(2): 512-525.
7. B. Eriksson, P. Barford, R. Bowden, et al., Basisdetect: A model-based network event detection framework, in Proc. IMC'10, pp. 451-464, 2010.
8. D. Jiang, C. Yao, Z. Xu, et al., Multi-scale anomaly detection for high-speed network traffic, Trans. Emerg. Telecommun., 2015, 26(3): 308-317.
9. G. Y. Lazarou, J. Baca, V. S. Frost, et al., Describing network traffic using the index of variability, IEEE Trans. Netw., 2009, 17(5): 1672-1683.
10. A. Vishwanath, et al., Anomalous loss performance for mixed real-time and TCP traffic in routers, IEEE Trans. Netw., 2011, 19(4): 933-946.
11. D. Jiang, et al., How to reconstruct end-to-end traffic based on time-frequency analysis, AEU-Int. J. Electron. Commun., 2014, 68(10): 915-925.
12. T. Akgül, et al., Periodicity-based anomalies in self-similar network traffic flow measurements, IEEE Trans. Instrum. Meas., 2011, 60(4): 1358-1366.
13. G. Thatte, U. Mitra, J. Heidemann, Parametric methods for anomaly detection in aggregate traffic, IEEE Trans. Netw., 2011, 19(2): 512-525.