ITM Web of Conferences **7**, 09025 (2016) DOI: 10.1051/itmconf/20160709025
ITA 2016

A Wavelet Analysis-Based Dynamic Prediction Algorithm to Network Traffic

Fan-Bo Meng[1,a], Hong-Hao Zhao[1], Si-Hang Zhao[2], Si-Wen Zhao[3], Zhong-Qiu Lin[4]

[1] *State Grid Liaoning Electric Power Company Limited, Shenyang 110006, China*
[2] *State Grid Huludao Electric Power Supply Company, Huludao 125000, China*
[3] *Shenyang China Resources Thermal Power Company Limited, Shenyang 110043, China*
[4] *State Grid Dandong Electric Power Supply Company, Dandong 118000, China*

Abstract. Network traffic is a significantly important parameter for network traffic engineering, while it holds highly dynamic nature in the network. Accordingly, it is difficult and impossible to directly predict traffic amount of end-to-end flows. This paper proposes a new prediction algorithm to network traffic using the wavelet analysis. Firstly, network traffic is converted into the time-frequency domain to capture time-frequency feature of network traffic. Secondly, in different frequency components, we model network traffic in the time-frequency domain. Finally, we build the prediction model about network traffic. At the same time, the corresponding prediction algorithm is presented to attain network traffic prediction. Simulation results indicates that our approach is promising.

1 Introduction

Network traffic is very significant for network management, traffic engineering, network monitoring, routing optimization, and network measurement activities [1-2]. Network traffic, special for end-to-end network traffic, represents the behavior features of network users' and network devices' activities. Network traffic denotes the volume of flows between nodes or between Origin-Destination (OD) pairs. Particularly, OD pairs can more effectively describe the behaviors of network activities from a network-wide or global perspective [3-4]. Hence, network traffic holds an important role in the normal network activities. However, network traffic, special for OD traffic pairs, is more difficult to be estimated and predicted due to their highly time-varying nature [5-7]. In this paper, we study the network-wide traffic prediction problem, that is, we model and predict the end-to-end network traffic or the traffic amount of OD node pairs. Despite the difficulty in predicting network traffic, it has been the hot topic and received the more attentions.

Tebaldi et al. proposed a new statistical approach which was based on the statistical model of the inversion method to build network source-destination traffic flow model and to predict network traffic [4]. Tune et al. used the information theoretic to describe network traffic [8]. Takeda et al used the partial measurement information to estimate the end-to-end traffic matrix and proposed the corresponding prediction algorithm [6]. Zhang et al. [9] managed to acquire prior information of traffic matrix based on gravity model and then proposed traffic matrix estimation method. Jiang et al used neural network to model the end-to-end network traffic and attain the accurate estimation [10]. Chen et al. studied the short-time network traffic prediction [11]. The expectation maximization algorithm was used to estimate traffic matrix via the assumption that traffic demands between OD pairs were independent of each other and obeyed Gaussian distribution [12]. Jiang et al. also exploited the compressive sensing theory to propose a joint time-frequency estimation of network traffic [13]. Some of these methods had relatively large estimation errors, while other were very sensitive to prior information [9,14]. Hence, the above models and methods are difficult to accurately capture network flow traffic, so it is still significantly necessary to find more accurate model to depict network flow traffic, to lower the complexity of algorithms, and to improve the prediction accuracy.

Different from the above algorithms, this paper proposes a new Wavelet Analysis-based Prediction Algorithm (WAPA) to predict network traffic in the network. Firstly, the signal analysis theory is exploited to transform network traffic from time domain to time-frequency domain. And then the time-frequency analysis approach is used to characterize the inherent property of network traffic. Secondly, in the time-frequency domain, we decompose network traffic into different frequency components, such as low-, medium-, and high-frequency components. The low-frequency part reflects the change trend of network traffic, medium-frequency one describes their slow change, but the high-frequency one reflects their quick change. Hence, we use the autoregressive model, period signal model, and Gaussian noise model to model the low-, medium-, and high-frequency

[a] Corresponding author: amengfb@163.com

DOI: 10.1051/itmconf/20160709025

components, respectively. Thirdly, the corresponding algorithm is proposed to predict network traffic. Simulation results show that our algorithm holds much lower prediction error than previous methods and thus it is promising.

The rest of this paper is organized as follows. Section II is problem statement. Section III is to derive our prediction approach. Section IV is simulation results and analysis. And finally our work in this paper is concluded in Section IV.

2 Problem Statement

The end-to-end traffic in the communication network can be expressed as $x=(x(1),x(2),...,x(m))$ for m observation time slots, where $x(t)$ (where $t=1,2,...,m$) represents the volume of network traffic x at moment t. To facilitate the processing, network traffic x is normalized as

$$z(t)=\frac{x(t)-\frac{1}{m}\sum_{i=1}^{m}x(i)}{h} \quad (1)$$

where

$$h=\max\{|x(1)-\frac{1}{m}\sum_{i=1}^{m}x(i)|,...,|x(m)-\frac{1}{m}\sum_{i=1}^{m}x(i)|\}$$

Accordingly, $z(t)$ satisfies the following constraint:

$$-1\le z(t)\le 1 \quad (2)$$

In such a case, the computational complex is reduced effectively. According the wavelet analysis theory, the new network traffic $z=\{z(1),z(2),...,z(m)\}$ is converted into:

$$\begin{cases} a_{i(j+1)}=Ha_{ij}, \quad j=0,1,2,...,I-1 \\ d_{i(j+1)}=Ga_{ij}, \quad j=0,1,2,...,I-1 \\ a_{i0}=z(t) \end{cases} \quad (3)$$

where H is the high-pass filter, G denotes the low-pass filter, $a_{i(j+1)}$ and $d_{i(j+1)}$ are, respectively, the time-frequency coefficients of z, and I represents the number of decomposition levels.

According to Equation (3), divide the time-domain signals of network traffic z into low-, medium-, and high-frequency components as follows:

$$\begin{cases} W(f,t)=Y(z_i(t)=W_l(f,t)+W_m(f,t)+W_h(f,t) \\ W_l(f,t)=a_I \\ W_m(f,t)=(d_1,d_2,...,d_k) \\ W_h(f,t)=(d_{k+1},d_{k+2},...,d_I) \end{cases} \quad (4)$$

where $W_l(f,t)$, $W_m(f,t)$, and $W_h(f,t)$, respectively, denote the low-, medium-, and high-frequency components in the time-frequency domain, and $W(f,t)$ represents the time-frequency signals of network traffic z at time t.

According to the inversion transformation of the wavelet, for the low-, medium-, and high-frequency components of z, the loss part is filled with zeros. Then the low-frequency coefficient in the time-frequency domain can be attained:

$$W_l(f,t)=(a_I,0,0,...,0) \quad (5)$$

The medium-frequency coefficient in the time-frequency domain can be expressed as:

$$W_m(f,t)=(0,d_1,d_2,...,d_k,0,...,0) \quad (6)$$

Similarly, the high-frequency coefficient in the time-frequency domain can be expressed as:

$$W_h(f,t)=(0,0,...,0,d_{k+1},d_{k+2},...,d_I) \quad (7)$$

Then we obtain the below equation:

$$\begin{cases} z_l(t)=Y^{-1}(W_l(f,t)) \\ z_m(t)=Y^{-1}(W_m(f,t)) \\ z_h(t)=Y^{-1}(W_h(f,t)) \end{cases} \quad (8)$$

where $z_l(t)$, $z_m(t)$, and $z_h(t)$, respectively, denote the reconstruction signals of the low-, medium-, and high-frequency signals $W_l(f,t)$, $W_m(f,t)$, and $W_h(f,t)$ of network traffic $z(t)$ at time t.

According to Equation (8), the low-frequency time-domain signal $z_l=\{z_l(1),z_l(2),...,z_l(m)\}$ is modeled as:

$$z_l(t)=b_1z_l(t-1)+b_2z_l(t-2)+...+b_pz_l(t-p)+\varepsilon \quad (9)$$

where p is the order number of the autoregressive model, ε represents the error, and $b=\{b_1,b_2,...,b_p\}$ stands for the parameters of the model. $b=\{b_1,b_2,...,b_p\}$ can be determined via the signal analysis. Accordingly, Equation (9) is decided correctly.

The medium-frequency signal $z_m=\{z_m(1),...,z_m(m)\}$ can be denoted as:

$$z_m(t)=u\sin(\omega t+\alpha)+v\sin(\omega t+\beta)+\eta \quad (10)$$

where u and v are the coefficients of the model; ω, α, and β denote the frequency and initial phase of the period signal, and η represents the error of the model. The model in Equation (10) is decided via the signal analysis theory.

DOI: 10.1051/itmconf/20160709025

The high-frequency signal $z_h = \{z_h(1),\ldots,z_h(m)\}$ is modeled as:

$$z_h(t) = \frac{1}{\sqrt{2\pi}\delta}\exp\left(-\frac{(z_h(t)-\mu)}{2\delta^2}\right)+\lambda \qquad (11)$$

where δ and μ denote the variable and mean value of the Gaussian noise, and λ represents the error. Equation (11) is decided via the statistics theory.

According to the above discussion, the low-, medium-, and high-frequency signals corresponding to the time-frequency domain, respectively, are modeled as the autoregressive model, period signal, and Gaussian noise signal, as shown in Equations (9)-(11).

We can attain the time-domain signal as follows:

$$\hat{z}(t) = z_l(t) + z_m(t) + z_h(t) \qquad (12)$$

According to Equation (1), the resulting time-domain signal is obtained as follows:

$$\hat{x}(t) = h \times \hat{z}(t) + \frac{1}{m}\sum_{i=1}^{m} x(i) \qquad (13)$$

Accordingly, we can obtain the estimations of the end-to-end traffic by equation (13). The algorithm WAPA proposed can be described as follow:

Step 1: Give the initial data of the end-to-end network traffic data $x_0 = \{x_0(1), x_0(2),\ldots,x_0(r)\}$, where r denotes the length of initial data.

Step 2: According to Equation (1), perform the transformation to attain a new sequence series about the end-to-end network traffic $z_0 = \{z_0(1), z_0(2),\ldots,z_0(r)\}$.

Step 3: According to Equations (3), carry out the wavelet transformation, and then attain the time-frequency coefficients of $c_0 = \{a_l, d_1, d_2,\ldots,d_l\}$ for time signals.

Step 4: According to Equations (4), divide the time-frequency signals into the low-, medium-, and high-frequency components of $W_l(f,t)$, $W_m(f,t)$, and $W_h(f,t)$ in the time-frequency domain, respectively.

Step 5: Fill the loss entries in the low-, medium-, and high-frequency components of $W_l(f,t)$, $W_m(f,t)$, and $W_h(f,t)$ with zeros according to Equations (5)-(7).

Step 6: According to Equation (8), perform the inversion transformation of the wavelet, and then attain the low-, medium-, and high-frequency signals of $z_l(t)$, $z_m(t)$, and $z_h(t)$ in the time domain.

Step 7: According to Equations (9)-(11), build the models corresponding to the low-, medium-, and high-frequency signals of $z_l(t)$, $z_m(t)$, and $z_h(t)$, respectively, using the signal analysis and statistics theories via the initial data.

Step 8: According Equations (9)-(11), attain the prediction value of the low-, medium-, and high-frequency signals of $z_l(t)$, $z_m(t)$, and $z_h(t)$.

Step 9 According to Equations (12), calculate the time-domain signal $\hat{z}(t)$.

Step 10: According to Equation (13), attain the prediction signal $\hat{x}(t)$ of the end-to-end network traffic at time t.

Step 11: If the process is over, then save the results to file and exit, or go back to Step 8.

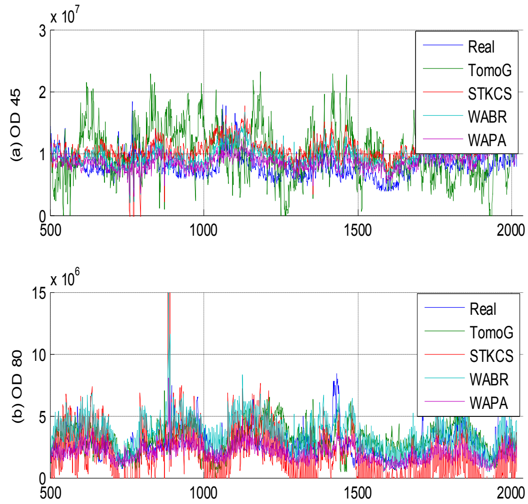

Figure 1. Prediction Results of Four Algorithms for Ods 45 and 80.

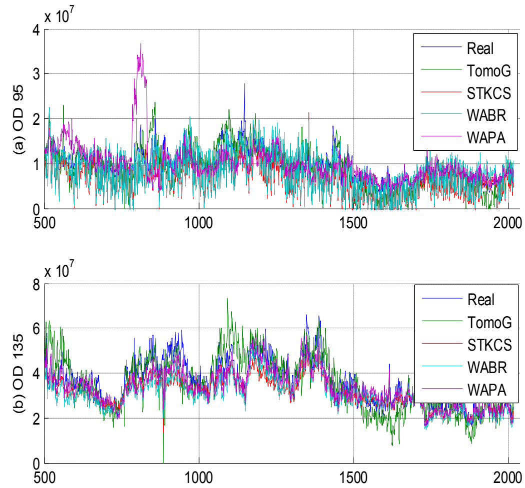

Figure 2. Prediction Results of Four Algorithms for Ods 95 and 135.

3 Simulation Result and analysis

Now we validate our algorithm WAPA. The real traffic data from the real Abilene backbone network in the United States are used to perform the simulation. STKCS [2], WABR[7], PCA [5], and TomoG [9] are reported as good prediction approaches for network traffic. In this paper, we will compare WAPA with them. at the same time, their prediction performance is analyzed in detail.

Figs. 1 and 2 plot the prediction results of four algorithms for ODs 45, 80, 95, and 135, respectively, where Real denotes the real network traffic. From Fig. 1, we can find that four algorithm can attain the good prediction results. In contrast to other three algorithms, WAPA holds the best prediction value

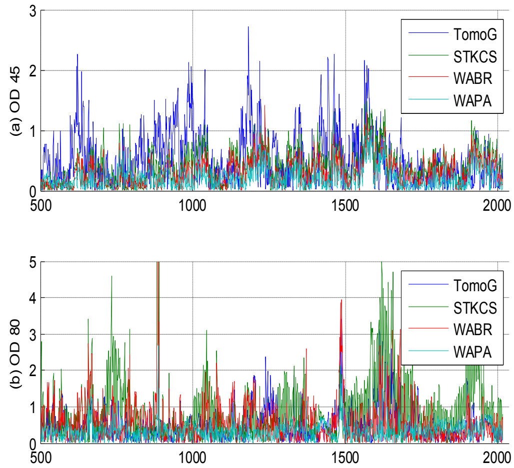

Figure 3. Relative Prediction Errors of Four Algorithms for Ods 45 and 80.

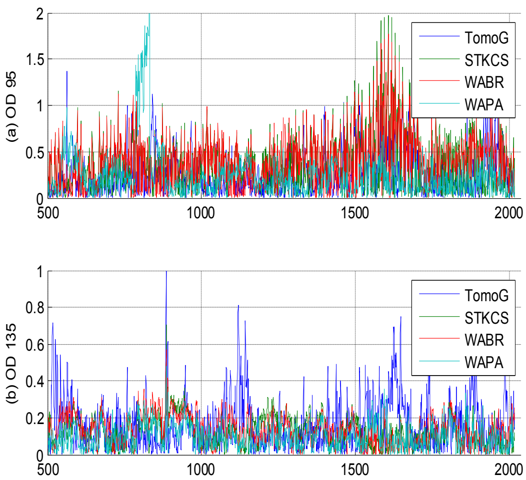

Figure 4. Relative Prediction Errors of Four Algorithms for Ods 95 And 135.

for ODs 45 and 80. TomoG has the larger prediction errors than other algorithm. Fig. 2 shows that four algorithms can also better predict the traffic of ODs 95 and 135. More importantly, when network traffic dynamically changes over time, four algorithms can capture the change trend of network traffic. Moreover, WAPA always can more accurately predict network traffic. This shows that WAPA holds the better prediction performance for network traffic.

Figs. 3 and 4 indicate the relative prediction errors of four algorithms for ODs 45, 80, 95, and 135, relative to the real network traffic. Fig. 3 shows that WAPA exhibits the lowest relative prediction errors in four algorithms for network traffic. In contrast to WAPA, TomoG holds the largest prediction errors, while STKCS and WABR illustrate the lower prediction. This demonstrates that WAPA holds the best prediction capacity for network traffic. From Figs. 3 and 4, it is clear that WAPA exhibits the more stationary prediction errors over the time than other three algorithm. This further indicates that WAPA holds the better prediction ability for network traffic.

4. Conclusions

This paper uses the wavelet analysis theory to model network traffic. By the wavelet transformation, we can correctly capture the dynamic and time-varying features in the time-frequency domain. Network traffic is converted into the time-frequency domain to capture time-frequency feature of network traffic. For different frequency components, we model network traffic in the time-frequency domain. Then network traffic signals are divided into the low-, medium-, and high-frequency signal. Accordingly we build the prediction model about network traffic. Finally, we propose the corresponding prediction algorithm to attain network traffic prediction. Simulation results indicates that our approach is promising and feasible.

References

1. D. Jiang, Z. Xu, J. Liu, et al. An optimization-based robust routing algorithm to energy-efficient networks for cloud computing. Telecommunication Systems, Sept., 2016, 63(1): 89-98.
2. T. Akgul, S. Baykut, M. Kantarci, S. Oktug. Periodicity-based anomalies in self-similar network traffic flow measurements, In Proceedings of TIM'11, 2011, 60(4): 1358-1366.
3. D. Jiang, L. Nie, Z. Lv, et al. Spatio-temporal Kronecker compressive sensing for traffic matrix recovery. IEEE Access, 2016, online available.
4. C. Tebaldi, M. West. Bayesian inference on network traffic using link count data, Journal of American Statistics Association, 1998, 93(442): 557-576.
5. A. Soule, A. Lakhina, N. Taft, et al. Traffic matrices: Balancing measurements, inference and modeling, ACM SIGMETRICS Performance Evaluation Review, 2005, 33(3): 362-373.
6. T. Takeda, K. Shionoto. Traffic matrix estimation in large-scale IP networks, In Proceedings of LANMAN'10, 2010, pp: 1-6.
7. D. Jiang, Z. Zhao, Z. Xu, et al. How to reconstruct end-to-end traffic based on time-frequency analysis and artificial neural network, AEU-International Journal of Electronics and Communications, 2014, 68(10): 915-925.
8. P. Tune, D. Veitch. Sampling vs sketching: An information theoretic comparison[C], In Proceedings of INFOCOM, 2011, pp: 2105-2113.
9. Y. Zhang, M Roughan, N. Duffield, et al. Fast accurate computation of large-scale IP traffic

DOI: 10.1051/itmconf/20160709025

matrices from link loads, ACM SIGMETRICS Performance Evaluation Review, 2003, 31(3): 206-217.

10. D. Jiang, Z. Xu, H. Xu, et al. An approximation method of origin-destination flow traffic from link load counts, Computers and Electrical Engineering, 2011, 37(6): 1106-1121.
11. C. Chen, J. Hu, et al. Short-time traffic flow prediction with ARIMA-GARCH model[C], In Proceedings of IV, 2011, pp: 607-612.
12. C. Hu. Online updating with a probability-based prediction model using expectation maximization algorithm for reliability forecasting. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, 2011, 41(6): 1268-1277.
13. D. Jiang, Z. Xu, Z. Chen, et al. Joint time-frequency sparse estimation of large-scale network traffic, Computer Networks, 2011, 55(10): 3533-3547.
14. H. Singhal and G. Michailidis. Structural models for dual modality data with application to network tomography, IEEE Transactions on Information Theory, 2011, 57(8): 5054-5071.