

A Bayes Theory-Based Modeling Algorithm to End-to-end Network Traffic

Hong-hao Zhao¹, Fan-bo Meng^{1,a}, Si-wen Zhao², Si-hang Zhao³, Yi Lu⁴

¹State Grid Liaoning Electric Power Company Limited, Shenyang 110006, China

²Shenyang China Resources Thermal Power Company Limited, Shenyang 110043, China

³State Grid Huludao Electric Power Supply Company, Huludao 125000, China

⁴State Grid Shenyang Electric Power Supply Company, Shenyang 110003, China

Abstract. Recently, network traffic has exponentially increasing due to all kind of applications, such as mobile Internet, smart cities, smart transportations, Internet of things, and so on. the end-to-end network traffic becomes more important for traffic engineering. Usually end-to-end traffic estimation is highly difficult. This paper proposes a Bayes theory-based method to model the end-to-end network traffic. Firstly, the end-to-end network traffic is described as a independent identically distributed normal process. Then the Bases theory is used to characterize the end-to-end network traffic. By calculating the parameters, the model is determined correctly. Simulation results show that our approach is feasible and effective.

1 Introduction

With the quick development of network technology and new applications, network traffic exhibits new features. This leads to a new challenging for network engineering [1-2]. it is important to accurately characterize and model network traffic for improving network performance. The features of network traffic, such as self-similarity nature, auto-correlations, heavy-tailed distribution, have an important impact on network optimization and routing [3-4]. The end-to-end network traffic represents the network-wide behaviors from a global point of view. Hence, modeling the end-to-end network traffic has received an extensive attention from researchers, operators, and developer all around the world [5].

The end-to-end traffic behaviors embody the path-level feature in the network. It can be used to describe network status and nature, such as path loads, throughput, network utilization, and so on. The statistical methods are employed to denote the model of network traffic from the source node to the destination node [1,3]. The gravity model [4], generic evolvement [6-7], mix method [2], and compressive sensing are utilized to capture the properties of the end-to-end network traffic. These method can attain the better prediction and estimation of the end-to-end traffic by performing a modeling process. However, these methods need the additional information from link loads or a prior information about the end-to-end network traffic. This adds the computational complexity and overhead for attaining the model parameters. The time-frequency domain analysis can be used to capture the multi-scale features and dynamic nature [1,8]. The neural network is employed to model network traffic [7,9].

^a Corresponding author: amengfb@163.com

These approaches can build the model to denote the end-to-end network traffic, while it is very difficult to exactly capture and seize their features and to build the accurate and appropriate network traffic model for traffic engineering.

This paper proposes a end-to-end network traffic modeling method to accurately characterize their features. Generally, it is significantly impossible to directly build the model about them due to their complex properties. Different form previous methods, we use the Bayes theory to establish the model about the end-to-end network traffic. Firstly, we denote the end-to-end network traffic as a independent identically distributed normal process. In the random process, there are several parameters to be estimated accurately. This is very difficult for the limited traffic information. Secondly, to this end, we use the Bases theory to characterize the end-to-end network traffic. By calculating the parameters with statistical methods, the model is determined correctly. In such a case, the model about the end-to-end network traffic is correctly built. Thirdly, we propose a new algorithm to build the model. 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

The modeling problem of the end-to-end traffic is difficult. In the network, there exists a lot of many end-

to-end network traffic. Without loss of generality, the end-to-end network traffic in the network is denoted as $x = \{x(1), x(2), \dots\}$, where $x(i)$ (where $i = 1, 2, \dots$) represents the value of traffic flow x at time slot i . Here, we assume that network traffic x follows the independent identically distributed normal process, which is denoted as:

$$X \sim N(\mu, \delta) \quad (1)$$

where μ and δ , respectively, denote the mean value and variance of the normal process X . Accordingly, the following equation can be attained:

$$P(X = x) = \frac{1}{\sqrt{2\pi\delta}} \exp\left(-\frac{(x-\mu)^2}{\delta^2}\right) \quad (2)$$

Here, parameters μ and δ describe the features of the end-to-end traffic x . However, Equations (1) and (2) only represent the statistical nature of the end-to-end traffic x . As mentioned in [10-12], the end-to-end network traffic holds the correlation property.

To capture the feature, we build the below model

$$x(k) = a_1x(k-1) + \dots + a_nx(k-n) + \varepsilon \quad (3)$$

where a_i (where $i = 1, 2, \dots, n$) denotes parameters; ε represents the normal process $N(\alpha, \beta)$, which stands for the bias of the model in Equation (3); n indicates the number of time slots before the current time slot.

According to Equations (1) and (2), $x(i)$ (where $i = 1, 2, \dots$) is an instance of the normal process X . From Equations (1)-(3), it is very clear that we have successfully characterize the statistical and correlation nature of the end-to-end network traffic. Equation (3) shows that given the end-to-end network traffic $x(k-1)$, $x(k-2)$, ..., $x(k-n)$ at time slots $k-1$, $k-2$, ..., $k-n$ and the bias distribution ε , we can get the end-to-end traffic $x(k)$ at time slot k . At each time slot, the end-to-end network traffic can be regarded as a random process. Thereby, according to the Bayes theory, Equation (3) can be denoted as:

$$\begin{aligned} &P(X(k) | X(k-1), X(k-2), \dots, X(k-n)) \\ &= \frac{P(X(k), X(k-1), X(k-2), \dots, X(k-n))}{P(X(k-1), X(k-2), \dots, X(k-n))} \end{aligned} \quad (4)$$

Our experiments show that when $n = 3$, we can accurately capture the features of the end-to-end network traffic. In such case, Equations (3) and (4) can be simply denoted as:

$$x(k) = a_1x(k-1) + a_2x(k-2) + a_3x(k-3) + \varepsilon \quad (5)$$

$$\begin{aligned} &P(X(k) | X(k-1), X(k-2), X(k-3)) \quad (6) \\ &= \frac{P(X(k), X(k-1), X(k-2), X(k-3))}{P(X(k-1), X(k-2), X(k-3))} \end{aligned}$$

Compared with Equations (3) and (4), Equations (5) and (6) use much simpler form to describe the end-to-end network traffic. Consequently, we can easily attain the model parameters. Due to the dynamic and random nature of the end-to-end network traffic, it is significantly difficult to directly determine the model in Equation (5).

To attain the parameters in Equation (5), we define the following loss function:

$$\begin{aligned} \Delta x(k) &= x(k) - \tilde{x}(k) \quad (7) \\ &= x(k) - (a_1x(k-1) + a_2x(k-2) + \\ &\quad a_3x(k-3) + \varepsilon) \end{aligned}$$

Equation (7) show the model bias, which is a random process related with the parameters $a_1, a_2, a_3, \alpha, \beta, \mu$ and δ . Accordingly, the below equation is attained:

$$\begin{aligned} &E(\Delta x(k) | a_1, a_2, a_3, \alpha, \beta, \mu, \delta) \quad (8) \\ &= \iint \Delta x(k) p(x, \varepsilon) dx d\varepsilon \end{aligned}$$

where $p(x, \varepsilon)$ denotes the joint probability density function of the normal random processes X and ε . Generally, we wish Equation (8) to be equal to zero. Because of the time-varying and correlated nature, there always exists some bias in Equation (8). Only when the bias is minimum, can we attain the optimal model for the end-to-end network traffic.

Accordingly, to solve Equation (3), we build the below optimal equation:

$$\left\{ \begin{aligned} &\min E(\Delta x(k) | a_1, a_2, a_3, \alpha, \beta, \mu, \delta) \quad (9) \\ &s.t. \tilde{x}(k) = a_1x(k-1) + \dots + a_3x(k-3) + \varepsilon \\ &P(X = x) = \frac{1}{\sqrt{2\pi\delta}} \exp\left(-\frac{(x-\mu)^2}{\delta^2}\right) \\ &\Delta x(k) = x(k) - \tilde{x}(k) \end{aligned} \right.$$

Equation (9) indicates a multi-constraint single object optimization problem. The first equation in Equation (9) denotes the object function, which minimizes the bias in the model in Equation (3); the second one represents the estimation of the end-to-end network traffic at time slot k according to Equation (3); the three one stands for the distribution of network traffic x ; the four one indicates the model bias. Using the sample data to train and solve the model in Equation (9), we can correctly attain the model parameters. In such a case, we can perform the correct description for the end-to-end network traffic.

In the following, we propose our modeling algorithm, called Bayes theory Traffic Modeling Algorithm (BTMA), to the end-to-end network traffic according to the above

analysis and derivation, namely:

Step 1: Give the h initial value of the end-to-end network traffic $x_{rr} = \{x(1), x(2), \dots, x(h)\}$ in the network.

Step 2: According to statistical theory, use initial value x_{rr} to attain the experience distribution parameters $\tilde{\mu}$ and $\tilde{\delta}$ of the end-to-end network traffic in Equations (1)-(2). Accordingly, obtain the experience distribution probability $N(\tilde{\mu}, \tilde{\delta})$.

Step 3: According to mathematical theory and signal processing theory, substitute initial value x_{rr} into Equation (5). Then determine the parameters $a_1, a_2, a_3, \alpha, \beta$. Accordingly, the experience model in Equation (3) is built.

Step 4: According to Equation (7), calculate the bias $\Delta x(k)$. Use Equation (8) to compute the statistical bias $E(\Delta x(k) | a_1, a_2, a_3, \alpha, \beta, \mu, \delta)$.

Step 5: According to Equation (9), minimize the statistical bias $E(\Delta x(k) | a_1, a_2, a_3, \alpha, \beta, \mu, \delta)$.

Step 6: Get the optimal parameters $a_1, a_2, a_3, \alpha, \beta, \mu, \delta$ in terms of Equation (9).

Step 7: If initial value x_{rr} has not been completely handled, go back to Step 4.

Step 8: The model of the end-to-end network traffic is correctly built. Then exit the modeling process.

3 Simulation Result and analysis

In this suction, we conduct some tests to demonstrate our algorithm BTMA. In order to verify the accuracy of our algorithm, we need to use real network data. The real data needed in the simulation experiment is collected by the network nodes; we use the real data from the real Abilene backbone network in the United States to validate BTMA. Matlab2010 is exploited performed the detailed simulation experiments. PCA [3], WABR [7], and HMPA [2] algorithms for the end-to-end network traffic modeling have been reported as the better performance. In this paper, we compare BTMA with them in detail. In the following, the prediction results of the end-to-end network traffic are analyzed for BTMA algorithm. The average relative errors for the end-to-end network traffic are indicated for four algorithms. Finally, we also evaluate the performance improvement of BTMA against PCA, WABR, and HMPA. In our simulation, the data of the first 500 time slots are used to train the models of four approaches, while other data are exploited to validate the performance of all algorithms.

Fig.1 shows the prediction results of end-to-end traffic flows 78 and 118, where end-to-end traffic flows 78 and 118 are selected randomly from the 144 end-to-end traffic pairs (or flows) in the Abilene backbone network. As our simulation experiments, other end-to-end traffic pairs holds similar results. Without loss of generality, we only discuss the end-to-end traffic flows 78 and 118 in this paper. Additionally, here we consider the end-to-end traffic flows equal to the Origin Destination (OD) pair.

Fig. 1(a) indicates that BTMA can effectively capture the dynamic changes of the end-to-end traffic flow 78. For different time slots, the real end-to-end network traffic

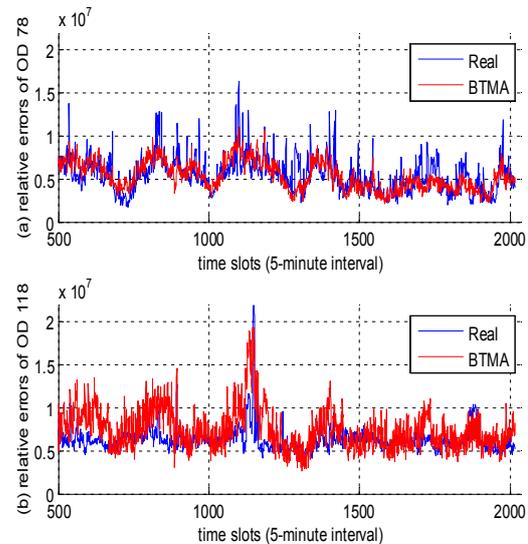


Figure 1. Prediction Results of End-to-End Traffic Flows 78 and 118.

exhibits the significant time-varying nature. From Fig. 1(a), we have seen that BTMA can seek the trend of the end-to-end traffic flow. Likewise, the end-to-end traffic flow 118 shows the irregular and dynamic changes over the time as indicated in Fig. 1(b). From Fig. 1(b), it is very clear that although BTMA holds the larger prediction errors for the end-to-end traffic flow 118, it can still capture its change trend. This further demonstrates that BTMA can effectively predict the change of the end-to-end network traffic over the time.

Next, we discuss the predict errors of four algorithms. Generally, the time-varying nature of the end-to-end network traffic over the time is difficult to be captured only via the model. To further validate our algorithm, we compare the relative prediction errors over the time for all algorithms. To avoid the randomness in the simulation, we perform 500 runs to calculate the average relative prediction errors.

The average relative prediction errors over the time for the end-to-end network traffic are defined as:

$$e(t) = \frac{1}{N} \sum_{i=1}^N \frac{\|\hat{x}_i(t) - x_i(t)\|_2}{\|x_i(t)\|_2} \quad (10)$$

where $i = 1, 2, \dots, N$, N is the number of runs in the simulation process, $\|\cdot\|_2$ is the norm of L_2 , and $\hat{x}_i(t)$ indicates the end-to-end traffic prediction value of run i at time slot t .

Fig. 2 illustrates the average relative prediction errors of four algorithm over the time for end-to-end traffic flows 78 and 118. It is very interesting that for end-to-end

traffic flows 78 and 118, WABR, HMPA, and BTMA exhibit the lower relative errors while PCA hold the larger prediction deviation. At the same time, Fig. 2 also indicates that BTMA holds the lowest relative errors. This tells us that in contrast to PCA, WABR, HMPA, and BTMA hold the better prediction ability for the end-to-

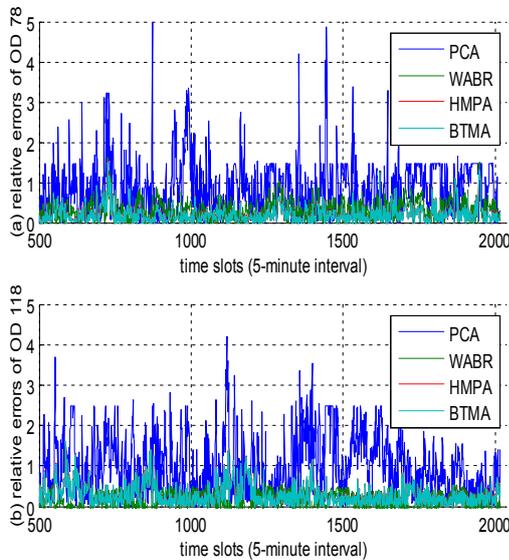


Figure 2. Average Relative Errors for End-to-End Traffic Flows 78 and 118.

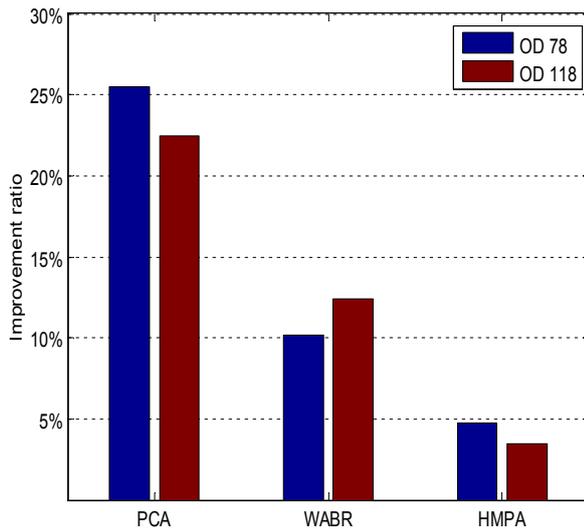


Figure 3. Improvement Ratio of End-to-End Traffic Flows 78 and 118.

end network traffic, while BTMA indeed has the best prediction ability. More importantly, WABR, HMPA, and BTMA exhibit the lower fluctuation over the time in terms of relative errors than PCA. This shows that compared with other three algorithms, BTMA can more effectively model the end-to-end network traffic with dynamic features.

Now, we analyze the improvement of BTMA to other

three algorithms for the end-to-end network traffic. Fig. 3 plots the improvement ration of end-to-end traffic flow 78 and 118. For end-to-end traffic flow 78, BTMA attains the performance improvement of about 25.5%, 10.0%, and 4.0% against PCA, WABR, and HMPA, respectively. Similarly, For end-to-end traffic flow 118, BTMA obtains the performance improvement of about 22.8%, 12.5%, and 3.3% against PCA, WABR, and HMPA, respectively. This definitely demonstrates that in contrast to PCA, WABR, and HMPA, our algorithm BTMA can indeed model the end-to-end network traffic more effectively. From Fig. 3, we also see that relative to PCA, BTMA can reach the largest performance improvement. For WABR, BTMA only achieve the smaller improvement. However, BTMA holds the lowest improvement against HMPA, namely less than 5%. As mentioned in Fig. 2, this further shows that WABR, HMPA, and BTMA hold the better modeling capability for the end-to-end network traffic. Moreover, BTMA and HMPA hold the similar performance. Therefore, BTMA can correctly model the end-to-end traffic.

4. Conclusions

This paper proposes a Bayes theory-based method to model the end-to-end network traffic. Different from previous methods, the Bayes model is used to infer and establish the model parameters effectively. Firstly, the end-to-end network traffic is described as a independent identically distributed normal process. Secondly, the Bases method is exploited to capture the end-to-end network traffic. By calculating the parameters, we construct the corresponding network traffic model. Simulation results show that our approach is feasible and effective.

References

1. 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.
2. D. Jiang, Z. Xu, H. Xu. A novel hybrid prediction algorithm to network traffic. *Annals of Telecommunications*, 2015, 70(9): 427-439.
3. A. Soule, A. Lakhina, N. Taft, et al. Traffic matrices: balancing measurements, inference and modeling, In *Proceedings of SIGMETRICS'05 2005*, 33(1): 362-373.
4. Y. Zhang, M. Roughan, N. Duffield, et al. Fast accurate computation of large-scale IP traffic matrices from link loads, In *Proceedings of SIGMETRICS'03, 2003*, 31(3): 206-217.
5. T. Takeda, K. Shionoto. Traffic matrix estimation in large-scale IP networks, In *Proceedings of LANMAN'10, 2010*, pp: 1-6.
6. F Yingxun. *The research and improvement of the genetic algorithm*, Beijing, Beijing University of Posts and Telecommunications, 2010.
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. D. Jiang, Z. Yuan, P. Zhang, et al. A traffic anomaly detection approach in communication networks for applications of multimedia medical devices. *Multimedia Tools and Applications*, 2016, online available.
 9. D. Jiang, Z. Xu, L. Nie, et al. An approximate approach to end-to-end traffic in communication networks. *Chinese Journal of Electronics*, 2012, 21(4): 705-710.
 10. S. Vaton, J. Bedo. Network traffic matrix: How can one learn the prior distributions from the link counts only, In *Proceedings of ICC'04*, 2004, pp: 2138-2142.
 11. M. Lad, R. Oliveira, D. Massey, et al. Inferring the origin of routing changes using link weights, In *Proceedings of ICNP*, 2007, pp: 93-102.
 12. P. Tune, D. Veitch. Sampling vs sketching: An information theoretic comparison, In *Proceedings of INFOCOM*, 2011, pp: 2105-2113.