Computer Networks 85 (2015) 36–50

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Traffic prediction for dynamic traffic engineering

Tatsuya Otoshi [a,*], Yuichi Ohsita [a], Masayuki Murata [a], Yousuke Takahashi [b], Keisuke Ishibashi [b], Kohei Shiomoto [b]

[a] Graduate School of Information Science and Technology, Osaka University, Suita-shi 565–0871, Japan
[b] NTT Network Technology Laboratories, NTT Corporation, Musashino-Shi 180–8585, Japan

ARTICLE INFO

Article history:
Received 10 September 2014
Revised 13 April 2015
Accepted 7 May 2015
Available online 18 May 2015

Keywords:
Traffic engineering
Traffic prediction
Data mining
Trend component
SARIMA model
ARIMA model

ABSTRACT

Traffic engineering with traffic prediction is a promising approach to accommodate time-varying traffic without frequent route changes. In this approach, the routes are decided so as to avoid congestion on the basis of the predicted traffic. However, if the range of variation including temporal traffic changes within the next control interval is not appropriately decided, the route cannot accommodate the shorter-term variation and congestion still occurs. To solve this problem, we propose a prediction procedure to consider the short-term and longer-term future traffic demands. Our method predicts the longer-term traffic variation from the monitored traffic data. We then take account of the short-term traffic variation in order to accommodate prediction uncertainty incurred by temporal traffic changes and prediction errors. We use the standard deviation to estimate the range of short-term fluctuation. Through the simulation using actual traffic traces on a backbone network of Internet2, we show that traffic engineering using the traffic information predicted by our method can set up routes that accommodate traffic variation for several or more hours with efficient load balancing. As a result, we can reduce the required bandwidth by 18.9% using SARIMA with trend component compared with that of the existing traffic engineering methods.

1. Introduction

In recent years, time variation of Internet traffic has increased due to wide deployments of streaming and/or cloud services. Backbone networks are expected to accommodate such time-varying traffic without congestion. So far, backbone networks have addressed this problem by preparing redundant link capacity by considering not only average traffic but also traffic surges [1,2]. However, such an approach requires overly large capacity in accordance with the level of traffic change increases and causes low bandwidth utilization. For the last dozen years, the literature has reported that average link utilization of backbone networks has been very low, such as 17–29% in Google backbone [3], less than 50% in Sprint backbone [4], and 20% utilization is targeted in Internet2 [5]. This not only causes the waste of the bandwidth due to poor utilization of the network resource but also incurs unnecessary energy consumption. Henceforth, the traffic congestion must be avoided with limited resources, which will definitely reduce the over-provisioning cost and power consumption.

Adaptive traffic engineering is a promising approach for accommodating time-varying traffic by appropriately setting up the origin–destination (OD) routes [6–10]. In such traffic engineering methods, a control server periodically measures the traffic load in the network (typically every hour) and dynamically changes the routes so as to minimize the

* Corresponding author. Tel.: +81 0668794542.
E-mail addresses: t-otoshi@ist.osaka-u.ac.jp (T. Otoshi), y-ohsita@ist.osaka-u.ac.jp (Y. Ohsita), murata@ist.osaka-u.ac.jp (M. Murata), takahashi.yousuke@lab.ntt.co.jp (Y. Takahashi), ishibashi.keisuke@lab.ntt.co.jp (K. Ishibashi), shiomoto.kohei@lab.ntt.co.jp (K. Shiomoto).

http://dx.doi.org/10.1016/j.comnet.2015.05.001
1389-1286/

network congestion. However, traffic engineering using the measured traffic only mitigates the observed congestion and never avoids the future congestion. The currently congested links are resolved by changing routes at the next control epoch. By making the control interval shorter (say, in a unit of minutes), the control server may respond quickly to such traffic changes. However, it obviously causes the heavy load at the control server and affects the performance of the upper-layer protocol TCP due to frequent route changes. Such routing oscillation degrades the throughput of TCP sessions because of packet reordering and changes of RTT [11]. Our solution here is to execute traffic engineering by predicting the future traffic changes. That is, the control server should set up routes by considering the future traffic demands, not past ones. More exactly, the control server predicts the traffic variation in the next control cycle and then determines routes that can accommodate the predicted traffic without causing congestion in the next control cycle. For deciding the traffic variation, we again have the "time-scale" problem: if we want to have stable operation, we need set up a larger control cycle, but in that case, we cannot react to the temporal changes of traffic variation within the control interval. The shorter control cycle has exactly the same problem described above.

So far, various prediction methods have been studied on the basis of traffic predictive models such as ARMA, ARIMA [12,13], ARCH [14], GARCH [15], and Neural Network [16–18]. However, to the best of our knowledge, existing prediction methods do not solve the above problem because they can predict the traffic variation accurately only for its target time scale. For example, the method proposed by Guang et al. [17] targets prediction in time scale of several hours. Therefore, it cannot obtain information about shorter-term variations because they are removed as noise before the prediction. On the other hand, a prediction method targeting a small time scale such as milliseconds or minutes [12,14,15,19] is only effective for very near future prediction because of the significant degradation of prediction accuracy in the far future.

In this paper, we propose a traffic prediction procedure intended for application to traffic engineering by separating the short-term (non-periodical or temporal) and longer-term (hour or day) variations. We directly predict the longer-term variation as existing methods and estimate the short-term variation instead of predicting it. We then obtain the range of traffic variation including short-term variation during the next several hours, which is used as a basis for calculating the necessary capacities of each route in the next control interval. That is, our key contribution here is that we investigate how to handle the prediction uncertainty in order to apply our method to traffic engineering. As described before, the prediction uncertainty stems from two factors (prediction error for periodical pattern and noisy short-term variation), and we take account of such prediction uncertainty in determining the necessary resources for each route. In this paper, we focus on the results of traffic engineering instead of the accuracy of prediction, because prediction methods with small error are not always suited to traffic engineering. Even when mean prediction error is low, congestion cannot be avoided by traffic engineering using the predicted traffic if the temporal increase of the traffic causing congestion is not predicted. On the other hand, a prediction method responsive to the traffic increase that may cause the congestion can avoid congestion even if the method's mean prediction error is large. Therefore, we evaluate our prediction procedure by investigating the influence of prediction method on traffic engineering performance.

In our earlier work [20], we only compared the effectiveness of traffic engineering using predicted traffic with observation-based traffic engineering. This paper also investigates details of the impact of traffic prediction on traffic engineering. We first investigate the impact of two parameters in our prediction procedure having a large impact on traffic engineering, the confidence level of prediction errors and the confidence level of short-term variation. We find that the confidence level of the short-term variation should be set to a large value, while a small confidence level for prediction errors is generally sufficient.

We then investigate the impact of considering periodicity, and find that even prediction without considering periodicity is sufficient if the control period is a few hours, while traffic prediction considering periodicity improves the worst-link utilization achieved by traffic engineering if the control period is larger than 24 h.

The rest of this paper is organized as follows. Section 2 surveys related work of traffic prediction and traffic engineering. Section 3 introduces the traffic engineering method using the predicted traffic. Section 4 describes the prediction procedure. Section 5 presents an evaluation of our prediction procedure. Section 6 mentions the conclusion and future work.

2. Related work

2.1. Traffic engineering

There is a large body of literature regarding TE [6–10]. The most of existing traffic engineering methods are observation-based approach in which the control server collects the current traffic information and then sets the routes so as to accommodate the observed traffic. However, such observation-based method may not be able to accommodate the future traffic because the traffic pattern will change from the observed pattern.

One approach to handling such uncertainty of the future traffic is to allocate sufficient resources to accommodate worst-case traffic patterns. For example, a static routing method called *oblivious routing* [21–23] sets a fixed route to accommodate worst-case traffic. Instead of observing current traffic, this method tries to accommodate all possible traffic patterns by minimizing the maximum link load. Wang et al. proposed a robust traffic engineering method by introducing the oblivious routing concept [6]. Their method considers the *convex hull* of a set of historical traffic patterns, namely the set of arbitrary weighted average of observed traffic. It handles uncertain future traffic dynamics by optimizing routes for this convex hull under constraints where the worst-case performance is not degraded. However, the approach requires large resources to accommodate worst-case traffic.

To accommodate the future traffic variation with a small resources, it is important to know the future traffic. Thus, our