Computer Networks 67 (2014) 252–269

Contents lists available at ScienceDirect

Computer Networks

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

A hybrid procedure for efficient link dimensioning

R. de O. Schmidt [a,*], R. Sadre [b], A. Sperotto [a], H. van den Berg [a,c], A. Pras [a]

[a] University of Twente, Enschede, The Netherlands
[b] Aalborg University, Aalborg, Denmark
[c] TNO Information and Communication Technology, Delft, The Netherlands

ARTICLE INFO

Article history:
Received 15 October 2013
Received in revised form 6 March 2014
Accepted 8 April 2014
Available online 18 April 2014

Keywords:
Link dimensioning
Bandwidth estimation
Flows
NetFlow
IPFIX

ABSTRACT

An important task for network operators is to properly dimension the capacity of their links. Often, this is done by simple rules of thumb based on coarse traffic measurements provided, e.g., by SNMP. More accurate estimations of the required link capacity typically require packet-level measurements, which are hard to implement in today's high-speed networks. The challenge is, therefore, to accurately estimate the traffic statistics needed for estimating the required link capacity with minimal traffic measurement effort. This paper proposes a novel, hybrid procedure for link dimensioning that combines flow-level measurements, minimal efforts on packet captures, and an analytical traffic model. The result is an efficient and robust method to estimate required link capacities. Although the idea of estimating required capacities from flows is not new, the novelty of this paper is that it proposes a complete, efficient and deployable procedure. The proposed procedure has been extensively validated using real-world traffic captures dating from 2011 to 2012. Results show that, with minimal measurement effort, we are able to efficiently estimate the required bandwidth at timescales as low as 1 ms.

1. Introduction

An important task for network operators is to properly provision the capacity of their links. Under-provisioned links might result in immediate decrease in network performance, which can even be perceived by end users. Aiming at adequate QoS (*Quality of Service*), operators continuously monitor link usage. A commonly adopted approach is to read interface counters via SNMP (*Simple Network Management Protocol*) and use obtained values to roughly estimate required capacity for current traffic. Performing these measurements is relatively easy because such protocol is already implemented in most devices. However, the estimation of required capacity might lack accuracy since short-term traffic fluctuations are hard to capture via SNMP. Therefore, network operators tend to over-provision their links by using a *rule of thumb*: adding "large-enough" safety margins on top of the traffic averages obtained from SNMP counters. Over-provisioning, however, can lead to waste of link resources. Aiming at more efficient provisioning, in the recent past the research community has proposed several more accurate procedures for estimation of the required link capacity. Instead of relying on SNMP, these procedures often require traffic measurements solely at the packet level. However, continuous packet-level measurements in today's high-speed networks, with traffic rates of 10 Gb/s and more, are hard to deploy because they demand dedicated and mostly expensive devices.

Backbone links capacity provisioning is not the only possible application for link dimensioning approaches. These can also be used for a variety of related network management and configuration operations. Efficient

* Corresponding author. Tel.: +31 534898041.
E-mail address: r.schmidt@utwente.nl (R. de O. Schmidt).

http://dx.doi.org/10.1016/j.comnet.2014.04.008
1389-1286/

estimations of required capacity enable operators to know the residual capacity of their links (i.e., unused capacity). This information can be used, for example, to efficiently reallocate traffic in operations of load balancing and also towards energy efficiency. Furthermore, in a dynamic on-demand bandwidth service, link dimensioning can be applied on allocation of requests for resources, supporting QoS provisioning.

Contribution. This paper presents an efficient and practical link dimensioning procedure. Aiming at minimal measurement effort, this procedure uses flow-level traffic measurements (NetFlow/IPFIX-like measurements) combined with sporadic packet captures and an analytical model to efficiently describe short-term traffic fluctuations. The traffic model proposed in this paper extends the original model in [1] and allows us to predict traffic variance from flows at arbitrary timescales. This variance is then used in the dimensioning formula from [1–3]. Although the idea of using flow measurements for estimation of required link capacity is not new, the novelty of this paper is that we propose a complete and deployable procedure for link dimensioning. Our procedure has been extensively validated using real-world traffic measurements captured on universities routers and operators backbone links around the globe in 2011 and 2012. Our results show that we are able to efficiently estimate the required link capacity with minimal measurement effort at timescale as low as 1 ms.

Organization. The remainder of this paper is structured as follows. Related work on link dimensioning is described in Section 2. Flow-level network traffic monitoring is introduced in Section 3. In Sections 4 and 5 we detail the background on which we base our contributions, and also present the proposed flow-based link dimensioning procedure. Then, a complete overview of the proposed procedure is given in Section 6. The measurements dataset used in this paper is presented in Section 7. The validation of the proposed procedure and results discussion are done in Section 8. In Section 9 we provide a discussion on the parameters of the proposed solution and provide directions on how to set them in real deployments. Finally, in Section 10, we draw our conclusions.

2. Related work

The problem of bandwidth provisioning has been extensively studied. Several of the proposed solutions are technology-specific. For example, recently, Anjum et al. [4] proposed a bandwidth allocation procedure for delay sensitive applications along a path of point-to-point MPLS (*Multiprotocol Label Switching*) connection. More general solutions, such as [5,3,6], have also been proposed, in which intelligent over-provisioning of backbone links is presented as an attractive alternative for QoS achievement; Fraleigh et al. [5] focuses on packet delay, while [3,6] on link rate exceedance. However, because they require traffic measurements at the packet level, such solutions are hard to deploy since packet monitoring in high-speed networks requires powerful and expensive technologies. Pras et al. [3] and Mandjes and van de Meent [6] also propose an indirect method towards link dimensioning, in which traffic statistics are computed from samples of the router's buffer content. Although this approach does not need on-link traffic measurements, it requires additional complexity to be implemented in the routers.

In [9] the authors propose a bandwidth estimator based on a $M/G/\infty$ model. The main limitation of this work is, however, that it requires continuous packet-level measurements to observe packet arrivals and sizes. In addition, the model is further divided in four different sets of equations, and the selection on which one to use will depend on the timescale the operator wishes to dimension a given link. Our proposed solution differs by the fact that the timescale is already modeled within the adopted dimensioning formula (originally from [2,3]). This allows flexibility to the operator without the need to readapt the dimensioning procedure if timescale is changed.

In [7], the authors propose a traffic model based on Poisson flow arrivals and i.i.d. flow rates able to predict bandwidth consumption for non-congested backbone links. Our contribution differs in the proposed way of computing the traffic variance, since in our case no assumption on the evolution over time of the traffic in a single flow is needed. In [8], the authors provide dimensioning formulas for IP access networks, and the QoS is measured by useful per-flow throughput. In such work only elastic data traffic (TCP connections) was considered, while we do not put any constraint on the nature of the traffic.

The work in [1] proposes a provisioning procedure requiring minimal measurement effort, using minimal model assumptions, and with QoS constraints expressed in link rate exceedance. However, this work focuses on traffic variation that are solely due to fluctuations at the flow level, and the proposed bandwidth provisioning method is only valid for relatively large timescales, e.g., 1 s. We build upon this modeling approach by proposing an extended version of the model. In short, we propose a flow-based formula and additional packet-based correction factors that together enable better estimations of required capacity at smaller timescales. Finally, in [10] we propose a purely flow-based approach to estimate traffic variance from flow-based time series, which proved to work at timescales as low as 1 s. The modeling approach in the present paper, however, lowers this boundary to 1 ms.

3. Flow-based traffic monitoring

In this section we provide a brief introduction to the concept of flow-based monitoring. In [11], a *flow* is defined as a set of packets that share common properties passing an observation point in the network. A commonly-used flow definition is based on a 5-tuple key consisting of source and destination IP addresses, source and destination ports, and transport layer protocol.

A flow-monitoring probe exports information on the observed flows by means of *flow records*. Flow records are usually generated on the basis of timers with configurable timeouts, namely *active* and *inactive* timeout. These are defined as: