Predicting the performance of Internet-like networks using scaled-down replicas

Fragkiskos Papadopoulos, Konstantinos Psounis
University of Southern California
E-mail: fpapadop, kpsounis@usc.edu.

ABSTRACT

The Internet is a large, heterogeneous system operating at very high speeds and consisting of a large number of users. Researchers use a suite of tools and techniques in order to understand the performance of complex networks like the Internet: measurements, simulations, and deployments on small to medium-scale testbeds. This work considers a novel addition to this suite: a class of methods to *scale down* the *topology* of the Internet that enables researchers to create and observe a smaller replica, and extrapolate its performance to the expected performance of the larger Internet.

The key insight that we leverage is that only the congested links along the path of each flow introduce sizable queueing delays and dependencies among flows. Hence, one might hope that the network properties can be captured by a topology that consists of the congested links only. We have verified this in [11, 12] using extensive simulations with TCP traffic and theoretical analysis. Further, we have also shown that simulating a scaled topology can be up to two orders of magnitude faster than simulating the original topology. However, a main assumption of our approach was that uncongested links are known in advance.

We are currently working on establishing rules that can be used to efficiently identify uncongested links in large and complex networks like the Internet, when these are not known, and which can be ignored when building scaled-down network replicas.

1. INTRODUCTION AND MOTIVATION

Understanding the behavior of the Internet and predicting its performance are important research problems. These problems are made difficult because of the Internet's large size, heterogeneity and high speed of operation.

Researchers use various techniques to deal with these problems: modeling, measurement-based performance characterizations, and simulation studies. However, these techniques have their limitations.

First, the heterogeneity and complexity of the Internet makes it very difficult and time consuming to devise realistic traffic and network models. Second, due to the increasingly large bandwidths in the Internet core, it is very hard to obtain accurate and representative measurements. And further, even when such data are available it is very expensive and inefficient to run realistic simulations at meaningful scales.

To sidestep some of these problems, we have proposed in [11, 12] two methods that can be used to scale down the topology of the Internet while preserving a variety of important performance metrics, *e.g.* such as the end-to-end flow delay distributions.

In particular, by defining a link to be congested if the link imposes packet drops or significant queueing delays, we have shown that it is possible to infer the performance of the larger Internet by creating and observing a suitably scaled-down replica, consisting of the congested links only. Further, based on the observation that the majority of backbone links are uncongested, *e.g.* see [2], we have demonstrated that these methods can be used in practice, to dramatically simplify and expedite performance prediction. A main assumption of our approach was that uncongested links are known in advance, *e.g.* by utilizing a performance measurement tool.

However, while packet drops can be easily detected by a monitoring tool, accurately measuring queueing delays in high-speed backbone networks is quite difficult [13, 14, 2]. Hence, for downscaling to be practical and scalable, we need simple rules that can be used to identify links with negligible queueing delays, when these are not known. We are currently working on establishing such rules. Some of our preliminary results have been presented in [10]. Next, we give the main idea in scaling down the network topology. (For more details, the interested reader is referred to [12, 11, 10, 9].)

2. SCALING DOWN NETWORK TOPOLOGY

First, let's clearly define what we mean by an "uncongested" link in the context of downscaling. An uncongested link is a link which: (i) does not impose any packet drops, and (ii) its queueing delays are negligible compared to the total end-to-end delays of the packets that traverse it, *e.g.* one order of magnitude smaller. Most of the backbone links have both of these properties [14, 2, 13].

Now, as an illustrative example, let's consider the topology shown in Figure 1. In this topology we can see two congested links, and two groups of flows, Grp1 and Grp2. [1] Observe that Grp1 traverses only the one congested link, whereas Grp2 traverses both.

In [11, 12] we have presented two methods (DSCALEd and DSCALEs) that build a scaled replica consisting of the congested links only, along with the groups of flows that traverse them. [2] For the example shown in Figure 1, the

[1] A group of flows consists of those flows that follow the same network path.

[2] DSCALEd: Downscale using delays. DSCALEs: Down-

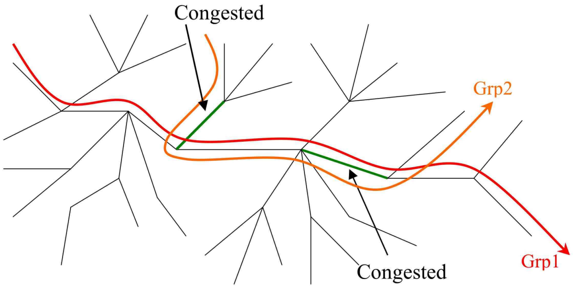

Figure 1: Original network.

resulting scaled replica is shown in Figure 2. Then, the methods adjust the round-trip times in the scaled replica appropriately, such that the performance of the replica can be extrapolated to that of the original network.

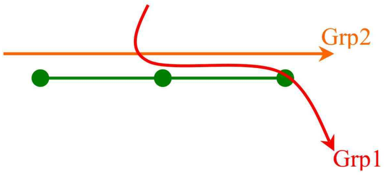

Figure 2: Scaled replica.

As mentioned earlier, our main assumption was that we know in advance which links of the original network can be considered as uncongested. However, as we have already said, while links that cause packet drops can be easily identified by a monitoring tool, measuring the queueing delays on every other link to determine whether these are negligible, is clearly a not scalable procedure. Further, it becomes critical in high-speed backbone routers, *e.g.* see [13, 14, 2]. Hence, we need simple rules that can be used to identify which of the links that do not impose drops do not impose significant queueing delays either, without having to explicitly measure their delays.

In [10] we have argued that the only case where one can decide by just inspecting the network topology, that a link imposes negligible queueing, is the case where the link carries traffic from/to links for which the sum of their capacities is smaller than the capacity of the link. For the rest of the cases, we use a model from the theory of large deviations to approximate the queue distribution. Next, we describe this model.

3. APPROXIMATING THE QUEUE LENGTH DISTRIBUTION TO IDENTIFY UNCONGESTED LINKS

Consider a link/queue. Let C be its capacity, B the buffer size, λ the average packet arrival rate and σ^2 the variance of the packet arrival process. Also, let $I(H) = \frac{(C-\lambda)^{2H}(\delta B)^{2-2H}}{2\sigma^2 K^2(H)}$, where $K(H) = H^H(1-H)^{1-H}$ ($H \in [0.5, 1)$ is called the Hurst parameter and is an index of long-range dependency in the packet arrival process), and $0 < \delta \leq 1$. Finally, let Q be the random variable representing the steady-state queue occupancy. Then, a widely accepted upper bound for Q, is [6]:

$$P(Q > \delta B) \leq \exp\left(-I(H)\right). \tag{1}$$

The above relation is known in the literature as the *large-buffer asymptotic* upper bound and the function $I(H)$ is called the *large deviations rate function*. The model assumes that the packet arrival process is well approximated by a Gaussian distribution. This is a realistic assumption for high-speed backbone links [1]. Further, if B is sufficiently large and the link's utilization is not very small (*e.g.* above $60 - 70\%$), Equation (1) can be used to approximate the queue distribution, *e.g.* see [4, 16]. [3]

When $\delta \geq \frac{1}{B}$ a better bound/approximation is [7]:

$$P(Q > \delta B) \leq \frac{1}{(\delta B)^\gamma} \exp\left(-I(H)\right), \tag{2}$$

where $\gamma = \frac{(1-H)(2H-1)}{H}$. Hence, for better approximating the queue distribution for any δ, one can take the minimum of Equations (1) and (2).

As we observe, using Equation (1) (or Equation (2)) requires knowledge of the packet level statistics λ, and σ^2. Further, it also requires knowledge of the parameter H. But, as with the queueing delays, it is difficult and not scalable to estimate these parameters by monitoring packets on every link of interest. However, it is much easier to monitor flows on a router, instead of packets [2, 3]. This argument is further strengthened by the fact that information on flows can be either collected on the link we want to study or at the edges of a backbone network. Collecting flow information at the edge routers and combined with their routing information, will give us information on each link of the network. This alleviates the burden of having to monitor many links and makes the measuring procedure scalable.

Since we are interested in links with no drops, an intuitive and well-known expression for λ (*e.g.* see [2]) is:

$$\lambda = rE[S], \tag{3}$$

where r is the flow arrival rate at the link of interest and $E[S]$ the expected flow size. [4]

The long-range dependence of Internet traffic has been shown to be the result of a heavy-tailed flow size distribution [15]. A heavy-tailed distribution is one in which $P(S > s) \sim s^{-\alpha}, 1 < \alpha < 2,$ as $s \to \infty$. At large time-scales, *e.g.* greater than the round-trip time, the parameter H is directly related

scale using sampling.

[3] Internet routers today are sized according to the rule-of-thumb, where the buffer size equals the bandwidth-delay product. Since capacities in backbone links are quite large, so that they can support a large number of flows, the buffer size B is also large. Further, we can safely consider backbone links at lower utilizations as uncongested, *e.g.* [14], without the need of using the model to approximate their queue distribution.

[4] Notice that it is not unrealistic to expect that only a small proportion of flows on a backbone link will experience drops elsewhere along their path, given that backbone links carry thousands of flows and that the number of concurrent congested Internet links is usually small. Therefore, for simplicity we can make the assumption that no flow that passes through the link under study experiences drops elsewhere along its path.

to the parameter α (called the shape parameter) of the size distribution. [5] According to [15]:

$$H = \frac{3-\alpha}{2}. \tag{4}$$

Hence, if we know the shape of the flow-size distribution (or the flow-size distribution itself) we can also compute H.

Finally, σ^2 is given in the following Theorem:

THEOREM 1.

$$\sigma^2 = \frac{E[A]\,Var(W) + (E[W])^2\,Var(A)}{(E[RTT])^{2H}}, \tag{5}$$

where $E[W]$ is the average congestion window size of a flow and $Var(W)$ its variance, $E[RTT]$ is the average round-trip time of a flow, and $E[A]$, $Var(A)$ the average and variance respectively of the steady-state number of active flows at the link. [6]

PROOF. See [9]. □

In [9] we also show how $E[W]$ and $\mathrm{Var}(W)$ can be inferred from the flow-size distribution, and how $E[RTT]$ can be computed from the flow-size distribution and the average number of active flows $E[A]$ at the link of interest.

Therefore, given the flow-size distribution, the flow arrival rate, and the first two moments of the steady-state number of active flows on the link of interest, we can compute all of the packet-level statistics that we need in order to use Equations (1) and (2). (Note that tools, *e.g.* such as NetFlow, can be easily utilized to provide such flow-level information [8].) We demonstrate in [9] the accuracy of our parameter estimation and of the model using extensive simulations with TCP traffic.

4. FUTURE WORK

Our future work consists of verifying the model and our parameter estimation using different network topologies, under various loads, as well as extending our analysis for links where packet drops occur. Further, we would also like to analytically quantify the relationship between the number of uncongested links that are ignored by topological downscaling and the achieved accuracy in performance prediction. Relevant to this, we also want to theoretically establish the queueing delay threshold below which, the queueing dynamics of a link can be completely ignored when evaluating the network's performance.

5. REFERENCES

[1] G. Appenzeller, I. Keslassy, and N. McKeown. Sizing router buffers. In *Proc of ACM SIGCOMM*, August 2004.

[2] C. Barakat, P. Thiran, G. Iannaccone, C. Diot, and P. Owezarski. A flow-based model for internet backbone traffic. In *Proc. of the 2nd ACM SIGCOMM Workshop on Internet measurment*, 2002.

[3] C. Barakat, P. Thiran, G. Iannaccone, C. Diot, and P. Owezarski. Modeling Internet backbone traffic at the flow level. *IEEE Transactions on Signal Processing, Special Issue on Networking*, 51(8), August 2003.

[4] A. Erramilli, O. Narayan, and W. Willinger. Experimental queueing analysis with long-range dependent packet traffic. *IEEE/ACM Trans. Netw.*, 4(2):209–223, 1996.

[5] C. Fraleigh, F. Tobagi, and C. Diot. Provisioning IP backbone networks to support latency sensitive traffic. In *Proc. of IEEE INFOCOM*, March 2003.

[6] A. Ganesh, N. O. Connell, and D. Wischik. Big queues. *Springer-Verlang*, Berlin, 2004.

[7] O. Narayan. Exact asymptotic queue length distribution for fractional brownian traffic. *Advances in Performance Analysis*, 1(1), 1998.

[8] Cisco IOS netflow. `http://www.cisco.com/en/US/products/ps6601/products_ios_protocol_group_home.html`.

[9] F. Papadopoulos and K. Psounis. Efficient identification of uncongested links for topological downscaling of Internet-like networks. *Technical report CENG-2007-7, University of Southern California*, 2007.

[10] F. Papadopoulos and K. Psounis. Application of the many-sources asymptotic in downscaling internet-like networks. In *Proc. of the Information Theory and Applications Workshop*, January 2007 (Invited Paper).

[11] F. Papadopoulos, K. Psounis, and R. Govindan. Performance preserving network downscaling. In *Proc. of the 38th Annual Simulation Symposium*, 2005.

[12] F. Papadopoulos, K. Psounis, and R. Govindan. Performance preserving topological downscaling of Internet-like networks. *IEEE Journal on Selected Areas in Communications, Issue on Sampling the Internet: Techniques and Applications*, 24(12), 2006.

[13] K. Papagianaki, S. Moon, C. Fraleigh, P. Thiran, F. Tobagi, and C.Diot. Analysis of measured single-hop delay from an operational backbone network. In *Proc. of IEEE INFOCOM*, June 2002.

[14] K. Papagiannaki. *Provisioning IP Backbone Networks Based on Measurements.* PhD thesis, University College London, March 2003.

[15] W. Willinger, M. S. Taqqu, R. Sherman, and D. V. Wilson. Self-similarity through high-variability: Statistical analysis of Ethernet LAN traffic at the source level. *IEEE/ACM Transactions on Networking*, 5(1):71–86, February 1997.

[16] D. Wischik. Buffer requirements for high-speed routers. In *Proc of ECOC*, 2005.

[5] Note that for links that are not working at small utilizations, it is the traffic dynamics at large time-scales (larger than the round-trip time of flows) that dominate their queueing behavior [5].

[6] As usual we say that a flow is active on a link if the link belongs to the path of the flow, and the flow has more data packets to send.