Pong: Diagnosing Spatio-Temporal Internet Congestion Properties

Leiwen Deng and Aleksandar Kuzmanovic
Department of Electrical Engineering and Computer Science, Northwestern University
Evanston, Illinois, USA
karldeng@cs.northwestern.edu, akuzma@cs.northwestern.edu

ABSTRACT

The ability to accurately detect congestion events in the Internet and reveal their spatial (*i.e.*, where they happen?) and temporal (*i.e.*, how frequently they occur and how long they last?) properties would significantly improve our understanding of how the Internet operates. In this paper we present *Pong*, a novel measurement tool capable of effectively diagnosing congestion events over short (*e.g.*, ∼ 100 ms or longer) time-scales, and simultaneously locating congested points on an end-to-end path at the granularity of a single link.

Pong (*i*) uses queuing delay as indicative of congestion, and (*ii*) strategically combines end-to-end probes with probes targeted to intermediate nodes. Moreover, it (*iii*) achieves high sampling frequency by sending probes to *all* intermediate nodes, including un-congested ones, (*iv*) dramatically improves spatial detection granularity (*i.e.*, from path segments to individual links), by using short-term congestion history, (*v*) considerably enhances the measurement quality by adjusting the probing methodology (*e.g.*, send 4-, 3-, or 2-packet probes) based on the observed path topology, and (*vi*) deterministically detects moments of its own inaccuracy. We conduct a large-scale measurement study on over 23,000 Internet paths and present their spatio-temporal properties as inferred by *Pong*.

Categories and Subject Descriptors: C.2.5 [Computer-Communications Networks]: Local and Wide-Area Networks

General Terms: Measurement, Performance

Keywords: coordinated probing, Pong.

1. INTRODUCTION

Despite existing measurements of spatial (*e.g.*, [1, 2]) and temporal (*e.g.*, [3]) Internet congestion characteristics, developing a more insightful and *joint spatio-temporal* congestion "picture" of today's Internet is valuable for several reasons. First, obtaining answers to questions such as (*i*) Where does the congestion occur (*e.g.*, at the edges, in the core, between ASes)? (*ii*) How long and how intensive is the congestion depending on its location? or (*iii*) What is the probability of observing one or more congested points on an end-to-end path? *etc.*, would significantly improve our understanding of the Internet. Moreover, the ability to accurately pinpoint congested locations in real time is useful for fault diagnosis (*e.g.*, [2]), for the design of advanced delay-based congestion control protocols (*e.g.*, [4, 5]), for building enhanced distributed monitoring systems (*e.g.*, [6, 7]) or for the design of efficient overlay systems.

Our contributions are threefold. First, we develop a novel measurement methodology capable of accurately inferring spatio-temporal congestion properties of Internet paths. We demonstrate that our methodology achieves high precision in both spatial and temporal domains. Second, we implement our methodology and build a novel measurement tool — *Pong*. Third, we conduct a large-scale measurement study (over 23,000 Internet paths), provide novel insights and enhance our understanding of spatio-temporal Internet congestion properties.

2. METHODOLOGY

Our methodology consists of two steps. First, endpoints strategically combine end-to-end probes with probes targeted to a certain intermediate node to estimate queuing delays on the two half paths between that intermediate node and both endpoints. Despite the fact that the resulting set of equations is under-constrained [8], and thus the unique solution does not exist, we manage to develop tight *lower and upper bounds* for the queuing delays on targeted path segments. Second, we sequentially repeat the above coordinated measurements to *all* nodes along the path. By correlating queuing delay estimates that are close in space and time, we manage to locate congestion at the granularity of a single link.

2.1 Coordinated Probing

The goal of the first step is to estimate queuing delays on *unidirectional* half paths between endpoints and an intermediate node. Because queuing can happen on both forward and backward directions of a path, it is impossible to directly measure queuing delays of targeted unidirectional half paths. Thus, we develop several novel methods to measure them indirectly.

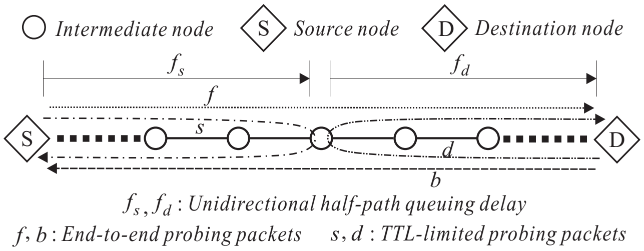

Figure 1: 4-p probing: a symmetric path scenario

SIGMETRICS'07, June 12–16, 2007, San Diego, California, USA.
ACM 978-1-59593-639-4/07/0006.

The most typical method is *coordinated 4-packet probing*, as indicated in Figure 1. For each intermediate node, we send four coordinated probing packets: two end-to-end packets (one goes along the forward path, the other goes along the backward path) and two TTL-limited packets that probe the intermediate node (one from the source, the other from the destination). By correlating measured delays of the four probing packets, we derive lower and upper queuing-delay bounds for the two unidirectional half paths. Figure 1 only shows a symmetric path scenario as the example. Indeed, *4-packet* probing also works very well in many asymmetric path scenarios. In addition, for scenarios with unfavorable path asymmetries and congestion patterns, we develop several "demoted" (*i.e.*, *3-packet*, and *2-packet*) probing methods to handle them. We adaptively choose the suitable probing method based on the asymmetry and congestion properties of related paths.

2.2 Locating Congestion Points

To detect congested locations at the granularity of individual links or path segments *between intermediate nodes*, we adopt a *switch-point* approach, which correlates half-path congestion estimates of neighboring nodes that are measured through coordinated probing. The *switch-point* approach can determine congestion locations with high resolution and very low false positives.

2.3 Tracing Congestion Status

Once congested links are located, we trace instantaneous congestion status of these links by *reusing* probes to all intermediate nodes along the paths.

3. IMPLEMENTATION

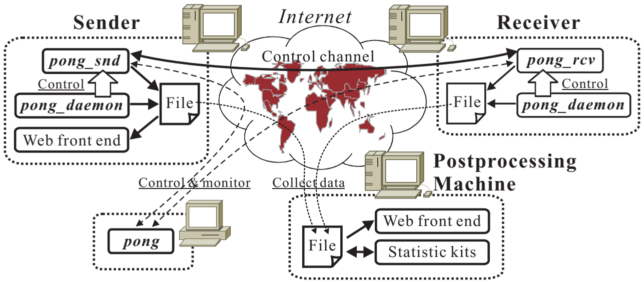

Figure 2: Components of Pong

Our implementation — a practical measurement tool *Pong* is depicted in Figure 2. *Pong* can not only perform measurements on individual paths, but also reliably automate large scale Internet measurements on a large number of diversified paths. *Pong* performs real time data processing during the measurement and generates results of path congestion status online. Therefore, it is also eligible to be used as a network monitoring tool.

4. EVALUATION

We implement our inference methodology in the *Emulab* testbed and evaluate its performance. The key idea is to compare the "ground truth," *i.e.*, various spatio-temporal congestion events that we create, with inferences made by the methodology. Numerous experiments including scenarios with UDP, short- and long-lived TCP cross traffic, and a number of congested links at both forward and backward paths demonstrate our ability to accurately pinpoint multiple congested locations in real time.

We also evaluate our measurement methodology via Internet experiments. However, in absence of the "ground truth" measurements from within the network, we apply a self-consistency validation method as a way to indirectly evaluate our methodology. The key idea is to correlate observations from different endpoints that measure the shared path segments concurrently, and to determine if the measurements are consistent in both spatial and temporal domains. We collect data useful for self-consistency validation from our large scale Internet measurement experiments. The results show both high spatial consistency and high temporal consistency in measurements. Consistency over a large number (*i.e.*, over 600 in our case) of independent measurements is a strong indicator of the validity of the approach.

5. MEASUREMENTS

We deployed *Pong* in PlanetLab, and measured spatio-temporal characteristics of over 23,000 Internet paths. Our findings are as follows. (*i*) The network edge is more frequently congested than the core; *14 times* on average. (*ii*) However, once the congestion happens, the congestion intensity at the edge is on average only 1.5 larger than that in the core. (*iii*) Overall, intra-AS links are more congested than inter-AS links. The reason is congestion occurring at *edge* ASes. (*iv*) At the Internet edges, intra-AS links are more congested than inter-AS ones. (*v*) In the core, the relationship reverses: inter-AS links are slightly more congested than intra-AS. (*vi*) Intra-AS links at the edge are significantly more congested than intra-AS links in the core. (*vii*) For inter-AS links, there is little to no difference in the congestion at the edge and in the core. (*viii*) Congestion events at edges are relatively clustered in time, while dispersed in the core. However, the congestion spike mode (peak-to-mean ratio) is uniform and location-independent. (*ix*) Approximately 12% of segments we measured experience congestion; only 3% of segments experience congestion more than 10% of time. Almost 20% of end-to-end paths experience considerable congestion. (*x*) The probability to observe multiple congested points on an end-to-end path over a given time interval grows as a power function of the interval length, and decays exponentially with the number of congested points.

6. REFERENCES

[1] N. Hu, L. Li, Z. Mao and P. Steenkiste. Locating Internet bottlenecks: algorithms, measurements, and implications. In *ACM SIGCOMM*, 2004.
[2] R. Mahajan, N. Spring, D. Wetherall and T. Anderson. User-level Internet path diagnostics. In *ACM SOSP*, 2003.
[3] Y. Zhang, N. Duffield, V. Paxson and S. Shenker. On the consistency of Internet path properties. In *ACM SIGCOMM IMW*, 2001.
[4] A. Kuzmanovic and E. Knightly. TCP-LP: Low-priority service via end-Point congestion Control. In *IEEE/ACM ToN*, 2006.
[5] D. Wei, C. Jin, S. Low and S. Hegde, FAST TCP: motivation, architecture, algorithms, performance, In *IEEE/ACM ToN*, 2006.
[6] M. Zhang, C. Zhang, V. Pai, L. Peterson and R. Wang. PlanetSeer: Internet path failure monitoring and characterization in wide-area services. In *OSDI*, 2004.
[7] A. Jain, J. Hellerstein, S. Ratnasamy and D. Wetherall. A wakeup call for Internet monitoring systems: The case for distributed triggers. In *HotNets*, 2004.
[8] Y. Zhao, Y. Chen and D. Bindel. Towards unbiased end-to-end network diagnosis. In *ACM SIGCOMM*, 2006.