Tomography-based Overlay Network Monitoring

Yan Chen, David Bindel, Randy H. Katz (UC Berkeley)
{yanchen, dbindel, randy}@cs.berkeley.edu

With the rapid growth of the Internet, new large-scale globally distributed network services and applications have emerged, such as overlay routing and location systems, application-level multicast, and peer-to-peer file sharing. As these systems have flexibility in choosing their communication paths and targets, they can benefit significantly from dynamic network distance prediction (e.g., latency and loss rate).

Existing network distance estimation systems can be grouped into two categories: *static estimation* [1], [2] and *dynamic monitoring* [3], [4], [5]. Previous static estimation systems, such as Global Network Positioning (GNP) [1], achieve a high level of accuracy, but also incur high overhead for continuously updating the estimates.

Dynamic monitoring can detect path outages and periods of degraded performance within seconds. However, existing schemes either require pair-wise measurements for all end hosts, and thus lack scalability [5]; or they can only estimate latency, but not congestion or failures [3], [4]. Existing scalable systems, such as [3], [4], cluster end hosts based on their network proximity or latency similarity under normal conditions. However, end hosts in the same cluster may not have similar losses, especially when the losses happen in the last mile.

In this poster, we describe a scalable overlay network congestion/failure monitoring system, which is highly accurate and incrementally deployable. Consider an overlay network of n end hosts, we define a path to be a routing path between a pair of end hosts, and a link to be an IP link between routers. A path is a concatenation of links. There are $O(n^2)$ paths among the n end hosts, and we wish to select a minimal subset of paths to monitor so that the loss rates and latencies of all other paths can be inferred. The loss rates are used to estimate the congestion/failures on the overlay paths.

To this end, we propose a tomography-based overlay network monitoring system in which we selectively monitor a *basis set* of k paths (typically $k \ll n^2$). Any end-to-end path can be written as a unique linear combination of paths in the basis set. Consequently, by monitoring loss rates for the paths in the basis set, we infer loss rates for all end-to-end paths. This can also be extended to latency. The end-to-end path loss rates can be computed even when the paths contain *unidentifiable links* for which loss rates cannot be computed. We provide an intuitive picture of this characterization process in terms of *virtual links*. The monitoring system architecture is shown in Fig. 1.

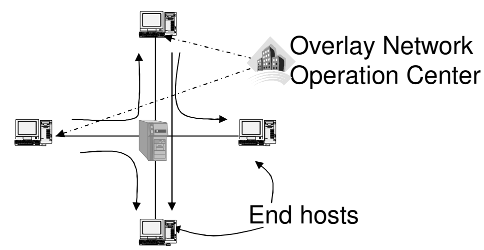

Fig. 1. Architecture of a tomography-based overlay network monitoring system. Assuming symmetric routing, we select 4 out of 6 paths for monitoring.

Our key observation is that k grows relatively slowly as a function of n. The dimension k is bounded by the number of links in the subgraph induced by the routing paths. In an Internet-like topology with a power-law degree distribution, there are $O(N)$ links, where N is the total number of end hosts in the network. This is because a small number of nodes have high degree and the links between them are heavily used [6]. Consequently, if $n = O(N)$, then $k < O(n)$. However, even when $n \ll N$, the moderately hierarchical structure of the network causes many routing paths to overlap [7], so that the number of links in the routing path subgraph grows much slower than $O(n^2)$. Our extensive study of both synthetic and real Internet topologies suggests that for a randomly selected subset of n end hosts, k grows like $O(n \log n)$ when n is sufficiently large (say 100).

Furthermore, our technique is tolerant to topology measurement inaccuracies, and is adaptive to topology changes. We also describe efficient algorithms to add/remove end hosts and to handle routing changes.

Besides simulating our system with various synthetic and real topologies, we implemented our system on the Planet Lab testbed [8]. We deployed it on 51 global hosts (each from a different organization) and ran the experiments over four weekdays with a total of 76.5M UDP packets. Both simulation and implementation results show we achieve high accuracy when estimating path loss rates with k measurements. For example, the average absolute error of loss rate estimation for the Internet experiments is only 0.0027 with average $k = 872$ out of a total of $51 \times 50 = 2550$ paths. On average, for 245 of the 2550 paths, the routing information obtained via traceroute are unavailable or incomplete, which shows that our technique is robust against topology measurement errors.

More information about this project is at: `http://www.cs.berkeley.edu/~yanchen/research/wnmms`.

References

[1] T. S. E. Ng and H. Zhang, "Predicting Internet network distance with coordinates-based approaches," in *Proc.of Infocom*, 2002.
[2] S. Ratnasamy et al., "Topologically-aware overlay construction and server selection," in *Proc. of IEEE Infocom*, 2002.
[3] P. Francis et al., "IDMaps: A global Internet host distance estimation service," *IEEE/ACM Trans. on Networking*, Oct. 2001.
[4] Y. Chen, K. Lim, C. Overton, and R. H. Katz, "On the stability of network distance estimation," in *ACM SIGMETRICS Performance Evaluation Review (PER)*, Sep. 2002.
[5] D. G. Andersen, H. Balakrishnan, M. F. Kaashoek, and R. Morris, "Resilient overlay networks," in *Proc. of ACM SOSP*, 2001.
[6] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationship of the Internet topology," in *ACM SIGCOMM*, 1999.
[7] H. Tangmunarunkit et al., "Network topology generators: Degree-based vs structural," in *ACM SIGCOMM*, 2002.
[8] Planet Lab, "http://www.planet-lab.org/," .