MIT/NTT Proposal
Statement of Work
Monitoring Network Routing Traffic with Low Space
Erik D. Demaine

An accurate model and general understanding of existing Internet traffic patterns and behaviors is critical for efficient network routing, caching, prefetching, information delivery, and network upgrades. Design of a network-traffic model consists of two interacting components: gathering relevant data from existing network traffic, and abstracting that data into a model with appropriate parameters. Both parts are essential. A model cannot be designed in isolation from actual traffic patterns, while on the other hand, raw data is too voluminous to suffice as a model. The proposed research focuses primarily on the first component, but there is inherent interplay with the second component, resulting in a complete approach to the problem of understanding network traffic.

A fundamental difficulty with measuring traffic behavior on the Internet is that there is simply too much data to be recorded for later analysis, on the order of gigabytes a second. As a result, network routers can collect only relatively few statistics about the data, and cannot spend much time per packet to collect these statistics. While hardware-specific solutions exist, they are expensive and cannot be widely deployed. The central problem addressed here is to use the limited memory of routers, on the order of a gigabyte total, to determine essential features of the network traffic stream. A particularly representative subproblem is to determine the top k sites to which the most packets are delivered, for a desired value of k.

These problems are made particularly difficult by the unpredictable distributions of network traffic. It is unreasonable to assume a priori that network traffic follows a standard distribution such as Poisson or Zipf. While such properties might hold on the aggregate, they cannot generally be expected to hold over a specific link of interest in the network. For example, consider two network pipes, one with a financial web site as the dominant destination and another with a local ISP as the dominant destination. In this case, one can expect a peak on traffic to the financial site at market open and close, while the ISP, used mostly from home, will follow a Gaussian distribution centered on 7:00pm.

We propose to design and analyze efficient algorithms and data structures that have provable guarantees on the quality of gathered statistics based on weak assumptions on the traffic distribution. One particularly attractive model that we have developed is that packet destinations follow an arbitrary distribution, but that their order is permuted uniformly at random. We hope that even under such weak assumptions it is possible to achieve low-space algorithms with guaranteed high probability of success. If this turns out not to be the case, we will prove lower bounds on the possible success of all algorithms in this broad model. We will consider the range from deterministic, fully guaranteed algorithms (which are extremely difficult) to randomized, probabilistically guaranteed algorithms (which are more powerful).

Solutions to these problems will impact the frontiers of Internet research. Relevant data and models of network traffic are essential to nearly all areas of network design. Our techniques and potential results will enable these areas to excel beyond current technology.

This research is already an ongoing effort with Alejandro López-Ortiz, Assistant Professor of Computer Science at University of Waterloo in Canada and former Director of

Core Research at Internap Network Services in Seattle; and with Ian Munro, Professor of Computer Science at University of Waterloo.

We have several initial ideas and approaches, and will continue to investigate the wide range of practically motivated algorithmic design problems that arise in this context. We welcome further collaboration with interested researchers at NTT.

References

[1] K. Claffy, G. Miller, K. Thompson. The nature of the beast: recent traffic measurements from an Internet backbone. *Proceedings of the 8th Annual Internet Society Conference*, ISOC, 1998.

[2] K. Claffy, T. E. Monk, and D. McRobb. Internet Tomography. *Nature*, January 7, 1999.

[3] I. D. Graham, S. F. Donelly, S. Martin, J. Martens, and J. G. Cleary. Nonintrusive and accurate measurements of unidirectional delay and delay variation in the Internet. *Proceedings of the 8th Annual Internet Society Conference*, ISOC, 1998.

[4] S. Kalidindi and M. J. Zekauskas. Surveyor: An infrastructure for Internet performance measurements. *Proceedings of the 9th Annual Internet Society Conference*, ISOC, 1999.

[5] A. Odlyzko. The current state and likely evolution of the Internet. *Proceedings of Globecom'99*, IEEE, pp. 1869–1875, 1999.

[6] X. Xiao and L. M. Ni. Reducing routing table computation cost in OSPF. *Proceedings of the 9th Annual Internet Society Conference*, ISOC, 1999.