



Re-assessing the Foundations of Internet Data Transport

Department of Computer Science

University of North Carolina at Chapel Hill

August 2007

Research Theme

The Internet started as a small research network connecting a few hundred computers and users. In just over two decades, this network has grown by several orders of magnitude, now connecting millions of users and serving thousands of new applications. Unfortunately, the basic transport mechanisms used in the operation of the Internet have not changed much in this time. In particular, after observing network congestion in the late 80s, several congestion control mechanisms—instantiated mostly as different versions of the popular TCP protocol—were developed. These mechanisms are ubiquitously deployed today. Meanwhile, the Internet infrastructure has grown by several orders of magnitude, both in its structure as well as traffic-carrying capacity. It is difficult to blindly believe that the assumptions on which network designs were based a decade ago are valid even today. It is natural to ask four kinds of questions: (i) Are assumptions used in the design of legacy transport mechanisms valid any more? For instance, is there any congestion in today's Internet? (ii) How does the invalidity of assumptions impact past work? (iii) How can mechanisms and analysis techniques be redesigned after discarding invalid assumptions?

Unfortunately, we do not know the answers to most of these fundamental questions. It is the goal of our research to answer these by:

1. Developing measurement and analysis techniques that enable end-users to develop a fundamental understanding of the transport performance of networks and validate legacy assumptions.
2. Studying the impact of invalid assumptions on the design, analysis, and evaluation of existing transport mechanisms.
3. Designing new mechanisms and analysis techniques after discarding high-impact invalid assumptions.

Below we provide a general overview of a few research projects guided by the above theme.

Project: Passive TCP Performance Analysis

TCP is the dominant transport protocol used for Internet data transfers. It is well-accepted that packet losses can degrade TCP performance---what is not well-understood, however, is the extent to which they do so in the real-world. In this project we study two critical interaction between TCP and losses: (i) TCP's efficiency in detecting and recovering from packet losses, when they occur, and (ii) TCP's ability to avoid packet losses when possible. Specifically, we make the following contributions:

- *A state-of-the-art tool for analyzing TCP traces:* We develop a new tool, *TCPdebug*, for analyzing traces of real TCP

Highlights

A *state-of-the-art* tool for passive analysis of TCP connection traces – the tool has more than 99% classification accuracy, which is more than twice of previously-developed tools.

The *first* comprehensive analysis of configuration of TCP loss detection in 4 prominent OS implementations, including Windows XP, Linux, Solaris, and MacOS – our conclusions can help reduce the duration of more than 25% of Internet TCP transfers by more than 1/3rd.

The *first* comprehensive analysis of the efficacy of delay-based congestion detection – our findings show that this area holds significant promise in upcoming high-speed networks.

The *first* set of detailed evaluations of Available Bandwidth Estimation Tools – our research helps identify several crucial, but previously-unidentified, factors that affect the performance of these tools.

A *scalable* approach for monitoring the N^2 all-pairs available bandwidth in an overlay network with only $O(N)$ overhead.

A graph-theoretical framework for analyzing the placement of probing beacons in a monitoring infrastructure – this framework leads to several insights, including the reduction in number of beacons previously-required by more than 50%.

connections. *TCPdebug* incorporates several OS-specific details of TCP implementations and has been shown to be significantly more accurate than existing analysis tools.

- *Evaluation of TCP loss detection:* We use *TCPdebug* to analyze the timeliness and accuracy of loss detection mechanisms in millions of TCP connection traces. We find that by re-configuring these mechanisms in Linux and Windows XP, it is possible to reduce – to less than 1/3rd – the time spent in more than 30% of current Internet data transfers.
- *Evaluation of delay-based congestion estimators:* We use *TCPdebug* to analyze the delays and losses experienced by millions of real-world TCP connections and study the ability of delay-based congestion estimators, DBCEs, to predict (and help avoid) losses. We find that the most prominent DBCEs are quite ineffective. We also find that DBCEs are likely to significantly benefit high-throughput connections and, consequently, hold promise in future high-speed networks.

Project: Design of Tomographic Infrastructures and Techniques

Recent interest in using tomography for network monitoring has motivated the issue of whether it is possible to use only a small number of probing nodes (beacons) for monitoring all edges of a network in the presence of dynamic routing. Past work has shown that minimizing the number of beacons is NP-hard, and has provided approximate solutions that may be fairly suboptimal. In our work, we use a two-pronged approach to compute an efficient beacon set:

- i) we design algorithms for computing the set of edges that can be monitored by a beacon under all possible routing states; and
- ii) we minimize the number of beacons used to monitor all network edges.

We show that the latter problem is NP-complete and use several approximate placement algorithms that yield beacon sets of sizes within $1 + \ln(|E|)$ of the optimal solution, where E is the set of edges to be monitored. Beacon set computations for several publicly-available ISP topologies indicate that our algorithms may reduce the number of beacons yielded by past solutions by more than 50% and are, in most cases, close to optimal.

Project: How to make the art of AB estimation useful?

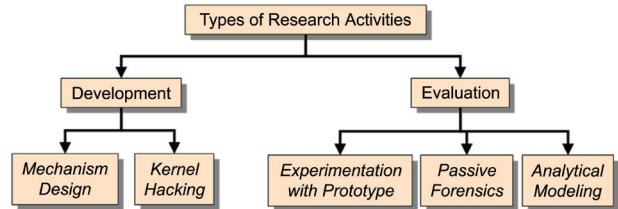
As the Internet evolves, several new applications that transfer large volumes of data are being designed – examples include scientific computing, streaming video, 3-D network games, and file-sharing applications. Such applications can benefit from the knowledge of the spare transfer capacity (*Available Bandwidth* or *AB*) of an Internet path. To support these, a myriad of tools have been designed over recent years for measuring the AB of a path. Unfortunately, none of these tools are being used by either network operators or application designers. We believe that two main limitations of existing tool designs are responsible for this. First, there are no comprehensive evaluations of existing tools—consequently, it is not clear which tool(s) perform the best and should be used. Second, most tool designs have not considered several fundamental requirements of targeted applications. In this project, we address these limitations in two ways:

- *Comprehensive evaluations of AB estimation tools:* We conduct the *first* set of evaluations that systematically study the impact of algorithmic design, systemic and implementation issues, as well as temporal aspects of measuring AB on the performance of existing tools.
- *Design of a scalable AB monitoring infrastructure:* We design techniques for inferring the all-pairs N^2 AB matrix in an overlay network of size N , using only $O(N)$ actual AB measurements. The resulting infrastructure significantly outperforms previous approaches in accuracy.

Types of Research Activities Conducted

The research projects described above rely on a good mix of several types of skills:

- *Design:* Development of new network architectures and protocol mechanisms.
- *Kernel Hacking:* Prototype implementation of new mechanisms.
- *Experimentation:* Evaluation of prototypes via large-scale emulations of network traffic and topologies.
- *Forensics:* Passive analysis of traces of real-world data transfers.
- *Modeling:* Analytical modeling of network dynamics as well as behavior of protocol mechanisms.



Researchers

Jasleen Kaur, assistant professor (jasleen@cs.unc.edu)

Don Smith, research professor (smithfd@cs.unc.edu)

Ritesh Kumar, graduate student (ritesb@cs.unc.edu)

Sushant Rewaskar, graduate student (rewaskar@cs.unc.edu)

Alok Shriram, graduate student (alok@cs.unc.edu)

Publications

S. Rewaskar, J. Kaur, and F.D. Smith. “A Performance Study of Loss Detection/Recovery in Real-world TCP Implementations,” in *Proc. of the IEEE International Conference on Network Protocols (ICNP '07)*, Beijing, China, Oct 2007.

A. Shriram and J. Kaur. “Empirical Evaluation of Techniques for Measuring Available Bandwidth,” in *Proc. of IEEE INFOCOM*, Anchorage, AK, May 2007.

R. Kumar and J. Kaur. “Practical Beacon Placement for Link Monitoring Using Network Tomography,” in *IEEE Journal on Selected Areas in Communication (JSAC)*, special issue on “Sampling the Internet: Techniques and Applications,” Dec 2006.

S. Rewaskar, J. Kaur, and F.D. Smith. “A Passive State-Machine Approach for Accurate Analysis of TCP Out-of-Sequence Segments,” in *ACM Computer Communications Review (CCR)*, July 2006.

A. Shriram and J. Kaur. “Empirical Study of the Impact of Sampling Timescales and Strategies on Measurement of Available Bandwidth,” in *Proc. of the Seventh Passive and Active Measurements Conference (PAM'06)*, Adelaide, Australia, March 2006.