Next-Generation Transport Protocols for Ultra-High Speed Networks

Department of Computer Science **University of North Carolina at Chapel Hill** **August 2010**

Motivation End-to-end data transfer rate requirements in the physics and astronomy scientific computation communities are soon to approach the terabit-per-second regime. Even for regular Internet, end-to-end transfer rate requirements of emerging digital media applications are likely to rise to at least the multi-gigabit regime. Unfortunately, even when sufficient raw transmission capacity is available at individual links and routers traversed on an Internet path, such capacity cannot be made available to applications if the underlying transport protocols do not scale correspondingly. In this project, we consider a novel paradigm for end-to-end congestion control that can help scale transport protocols to such ultra-high network speeds.

State of the Art The design of high-speed congestion control protocols has been a fairly vibrant area of research over the last decade. While most of current designs have been shown to be much more scalable than traditional TCP, even the best-performing designs scale to at most a few gigabits-per-second of steady-state single-stream throughput. The state of the art in transport protocols for TCP/IP networks is, consequently, not prepared to serve the needs of upcoming ultra-high speed networks.

New Paradigm: *Packet-scale Congestion Control* We argue that the state of the art offers only limited scalability because all of current high-speed designs retain a legacy design framework of *RTT-scale* protocol operations—this framework fundamentally limits the ability of a protocol to operate at ultra-high speeds without nearly causing congestion collapse. Instead, we show that if this legacy mindset is discarded, it is possible to design a novel paradigm of *packet-scale congestion-control*, in which the protocol operates at a frequency close to the frequency of packet transmissions. This paradigm allows the congestion-control timescale to be shrunk by several orders of magnitude over current protocols, especially in high-speed networks. This reduced timescale can then be exploited to probe for a wide range of rates within an RTT without overloading the network. This is the most distinguishing feature of the paradigm—existing "high-speed" protocols take orders of magnitude longer to probe for a similar range; and no existing protocol can do even that without overloading the network once it gets close to the available-bandwidth.

Promised Impact We have developed a congestion-control protocol, RAPID, that is a proof-of-concept based on this paradigm. Our design effort as well as simulation-based evaluations of RAPID suggest that the impact of the paradigm is likely to be quite significant along several dimensions that have remained elusive to previous high-speed transport protocols:

Speed/Overhead: While most RTT-scale "high-speed" protocols struggle with the speed-overhead tradeoff, the packet-scale paradigm could allow a protocol to detect end-to-end bandwidth of up to multi-terabits within a few RTTs, while causing negligible router queuing footprints!

Adaptability: The fine timescale at which the paradigm operates allows it to exploit efficiently short-timescale changes in end-to-end available bandwidth—this is true while most existing protocols either do not achieve high utilization or are able to do so only by maintaining very large packet queues at the bottleneck router.

Incremental Deployability/TCP Co-existence: Due to its low-queuing footprint, the packet-scale paradigm could allow an ultra-high speed transfer to share a network with highly-multiplexed aggregates of conventional low-speed TCP transfers without affecting the performance of the latter. None of existing "high-speed" protocols have been able to achieve this property without sacrificing on their efficiency.

RTT-fairness: By shedding the legacy framework of RTT-scale operation, the paradigm allows the design of end-to-end congestion-control that is truly RTT-fair and does not favor short-RTT transfers (again, unlike any RTT-scale high-speed protocol).

Highlights

- **This project represents a *significant performance leap*—while the best of current protocols are struggling to achieve 10Gbps transfer speeds, the paradigm enables comfortable operation at terabit-and-higher speeds. This is the *first* end-to-end congestion-control protocol for TCP/IP networks to achieve this scale.**
- **This research introduces a *fairly innovative paradigm shift* in the design of transport protocols. This fresh approach enables it to tackle issues that have remained stubbornly elusive to previous protocols.**
- **The innovativeness and nature of this research requires a research methodology that adopts both *theoretical analysis and formal modeling,* as well as *practical system design, implementation, and experimentation* on wide-area high-speed networks.**

Ongoing Research The packet-scale paradigm does need to address new challenges that are non-issues for RTT-scale protocols. These include: (i) the stringent support needed from end-systems for controlling and observing fine inter-packet gaps; (ii) the impact of transient "buffering" at non-bottleneck system resources; and (iii) the sensitivity and stability of the paradigm in of dynamic traffic conditions. Our current research is focused on addressing these challenges. The very diverse nature of these challenges requires us to rely on a mix of both *theory* (formal analysis, modeling, and design) and *practice* (prototype development and evaluation in wide-area networks).

Project Members

Jasleen Kaur, Associate Professor (jasleen@cs.unc.edu)
Simona Bacanu, Graduate Student
Debapriya Basu, Graduate Student
Eric Gavaletz, Graduate Student
Vishnu Konda, Alumnus
Alok Shriram, Alumnus

Publications

E. Gavaletz and J. Kaur, "Decomposing RTT-Unfairness in Transport Protocols," in *Proceedings of the 17th IEEE Workshop on Local and Metropolitan Area Networks (LANMAN'10)*, Long Branch, NJ, May 2010.

V. Konda and J. Kaur, "RAPID: Shrinking the Congestion-control Timescale," *Proceedings of IEEE INFOCOM*, Rio de Janeiro, Brazil, April 2009.

V. Konda and J. Kaur, "Rethinking the Timescales at Which Congestion-control Operates," (invited short paper) in *Proceedings of the 16th IEEE Workshop on Local and Metropolitan Area Networks (LANMAN'08)*, Cluj-Napoca, Romania, Sept 2008.

A. Shriram and J. Kaur, "Empirical Evaluation of Techniques for Measuring Available Bandwidth," *Proceedings of IEEE INFOCOM*, Anchorage, AK, May 2007.

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

http://rapid.web.unc.edu