Measurements of an XCP Implementation at Gigabit/Second Rates

Bob Braden, Aaron Falk, Ted Faber, Eric Coe

University of Southern California/Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292
Phone: 310-822-1511
`{braden,falk,faber,ecoe}@isi.edu`

Introduction

This presentation will describe preliminary results from implementation and testing of the eXplicit Control Protocol (XCP) [1,2] for TCP in a high bandwidth-delay-product network. XCP is a congestion control system that can be used by any transport protocol to establish endpoint transmission rates using available capacity feedback from a flow's bottleneck router. Simulations indicate that in high bandwidth-delay-product networks, XCP uses network capacity more efficiently and allocates that capacity more quickly and fairly than competing protocols [2]. This work is aimed at verifying those claims using an implementation of XCP under FreeBSD. This is the first implementation of XCP, and much of this initial work will be in validating and extending the early results on XCP [2].

More specifically, we wish to determine how well XCP uses the capacity of the network, how fairly it allocates capacity between multiple connections, how well XCP-based TCP connections can coexist with conventional Van Jacobson congestion control [3], and the effects of varying loss rates and dynamic bandwidth variation. We plan to use a FreeBSD-based implementation to investigate these issues.

The implementation study described in this abstract is part of a larger project looking at the feasibility of moving XCP into the Internet today. The larger project will include implementation studies on a variety of platforms, further simulation studies, studying mixed deployments of XCP-enabled routers with conventional routers, and efforts in the Internet standards bodies and implementation communities. The work is currently funded by the NSF OptIPuter Information Technology Research and Strategic Technologies for the Internet XCP Development projects.

XCP

XCP is a feedback-based congestion control system that uses direct, explicit, router feedback to avoid congestion in the network. It is designed for both scalability and generality. It was developed by Dina Katabi, starting from a suggestion by Mark Handley [1].

XCP's scalability results from the fact that it requires no per-flow state in the router to calculate the feedback. Most router-assisted congestion control systems maintain per-flow information used to allocate the resources. XCP keeps very little information in the router, and this information is chosen to minimize both the amount of router state and the per-packet operations needed to update that state.

For generality, XCP divides the resource allocation function between two controllers: a congestion controller that ensures that flows use all available capacity, and a fairness controller that ensures that flows are allocated the capacity fairly. Most congestion control systems fail to make this division, much less to implement as two conceptually distinct systems. This division allows a clear exposition and implementation of two basic resource allocation functions in XCP.

Beyond the intellectual merits of the XCP design, its direct router feedback promises improved performance. Unlike conventional feedback systems that infer the network capacity from the end and pick a sending rate by experiment [3,4], XCP is able to converge more quickly to a stable, efficient rate. Other router-assisted protocols are tied to a specific network technology [5] or only use router feedback to modulate standard feedback systems [6]. XCP can be implemented at the transport layer, and therefore in general IP networks. It should be applicable to any transport protocol that has congestion feedback.

Furthermore, because the routers are directly providing congestion data, XCP is less likely to be confused by the transient errors in high-bandwidth or lossy environments than systems that infer congestion from losses.

Experiments with XCP

ISI has prototyped XCP endpoints and routers using personal computer hardware. Initially, we have explored XCP's startup behavior compared with the existing FreeBSD implementation of TCP, which uses Van Jacobson congestion control. The FreeBSD TCP implementation is a well-tuned production system. As expected, we have found that the XCP congestion window ramps up more efficiently than conventional TCP and remains higher than that system in the absence of congestive loss. The following plot shows the evolution of the congestion window of a FreeBSD TCP and an XCP/TCP session on a gigabit network. These are two representative single connection runs. The network is shown in Figure 1.

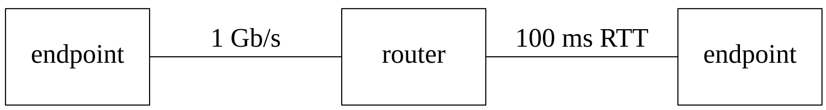

Figure 1: Network Configuration

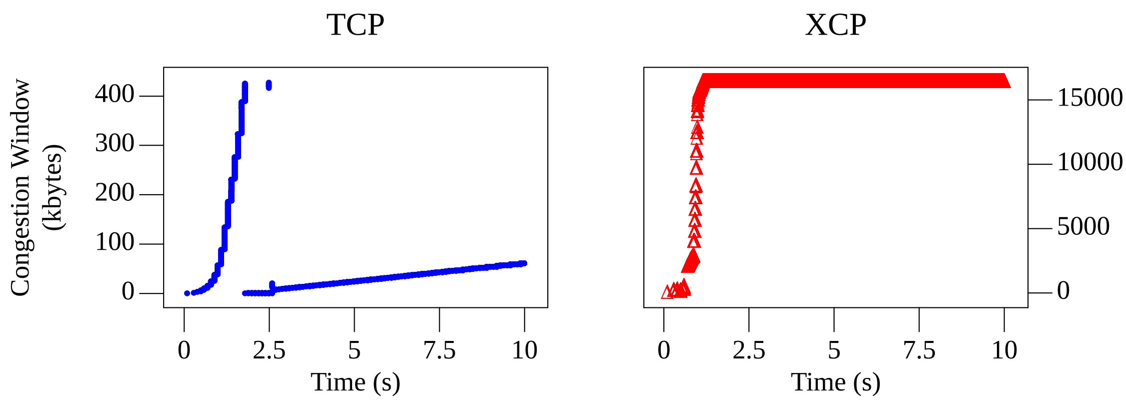

Figure 2

Figure 2 shows that XCP does achieve and hold a high congestion window, while conventional TCP gets locked into a slow probing cycle. Note the different y-axis ranges on the two plots.

The presentation for PFLD 2004 will describe other experimental results related to XCP performance across a high bandwidth-delay-product network using real components.

References

1. Dina Katabi, Mark Handley, and Charlie Rohrs, "Congestion Control for High Bandwidth-Delay Product Networks," *Proceedings of ACM SIGCOMM Symposium on Communications Architectures and Protocols*, pp. 89-102, ACM, Pittsburgh, PA (August 2002).

2. Dina Katabi, "Decoupling Congestion Control from the Bandwidth Allocation Policy and its Application to High Bandwidth-Delay Product Networks," *Ph.D Thesis*, MIT (March 2003).

3. Van Jacobson, "Congestion Avoidance and Control," *Proc. SIGCOMM Symposium on Communications Architectures and Protocols*, pp. 314-329, ACM SIGCOMM, Stanford, CA (Aug 16-19 1988).

4. Lawrence S. Brakmo, Sean W. O'Malley, and L. Peterson, "TCP Vegas: New techniques for congestion detection and avoidance," *Proceedings of ACM SIGCOMM Symposium on Communications Architectures and Protocols*, pp. 24-35, ACM, London, UK (August 1994).

5. Raj Jain, "Congestion Control and Traffic Management in ATM Networks: Recent Advances and a Survey," *Computer Networks and ISDN Systems*, vol. 28, pp. 1723-1738, Elsevier Science B. V. (1996).

6. K. Ramakrishnan and R. Jain, "A Binary Feedback Scheme for Congestion Avoidance in Computer Networks with a Connectionless Network Layer," *Proc. SIGCOMM Symposium on Communications Architectures and Protocols*, pp. 303-313, ACM SIGCOMM, Stamford, CA (Aug 16-19 1988).