Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Computer Networks 51 (2007) 1745–1747

Computer Networks

www.elsevier.com/locate/comnet

Guest Editorial

Protocols for fast, long-distance networks

Networks operating at 1 Gbit/s, 10 Gbit/s or even 100 Gbit/s and spanning several countries or states are now becoming commonplace. More and more users have to transfer routinely between multi-GB and multi-TB datasets over these gigabit networks. There is a growing range of application domains for such massive transfers including data-intensive Grids (e.g., in Particle Physics, Earth Observation, Bio informatics, and Radio Astronomy), database mirroring for Web sites (e.g., in e-commerce), and push-based Web cache updates. Although this high speed network infrastructure is already emerging, available transport and application protocols perform poorly over such networks. Standard TCP (TCP Reno or NewReno) is a reliable transport protocol that is designed to perform well in traditional networks. However, several experiments and analyses have shown that this protocol incurs substantial penalties when used for bulk data transfer in fast, long-distance networks.

Starting at CERN (Geneva) during the International DataTAG project in 2003, the International Workshop on Protocols for Fast Long-Distance Networks (PFLDnet) has brought together researchers from the US, Asia, and Europe working on these problems. This issue of Computer Networks focuses on the challenges of protocols for fast, long-distance networks, building upon the community of PFLDnet, the IEEE Gigabit/High-Speed Networks workshops, and other workshops held over the past decade. For this issue, twenty four original papers were received from the combination of an open call and solicited as extended versions of the best papers of the past two PFLDnet workshops. Seven were accepted and appear here, including three from PFLDnet. They include two studies of the relationship between protocols and fast, long-distance network environments, two new transport protocols designed specifically for high performance in this environment, and three new mechanisms that help transport and application protocols adjust appropriately to this environment.

During the last five years a number of TCP variants have been proposed as alternatives to standard TCP to improve the transport protocol in fast, long-distance networks, which have a characteristically large bandwidth-delay product. These variants modify the congestion control algorithm by adapting the increase and decrease factors of the AIMD (Additive-Increase Multiplicative-Decrease) algorithms, and some modify the congestion signal as well. The first paper in this special issue compares these high speed TCP variants. Ha, Le, Rhee and Xu examine the performance of these protocols in "Impact of Background Traffic on Performance of High-Speed TCP Variant Protocols" and they demonstrate how it is important to consider the background traffic in protocol evaluation methodology. The problem of multiple congested links (MCLs) and RTT (Round-Trip Time) bias appears to be more serious with these high speed TCP variants than with standard TCP. The paper "Can High-Speed Networks Survive with DropTail Queues Management?" by Chen and Bensaou shows that the drop tail algorithm increases the unfairness across multiple congested links in high speed networks. They show that this result is mainly due to the synchronized losses and that active queue management schemes can mitigate this unfairness.

New transport protocols are required to support high performance in these new environments.

1389-1286/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.comnet.2006.12.005

Emerging high speed wide area optical networks are being deployed for scientific data distribution, transferring large volumetric datasets. Gu and Grossman, in their paper "UDT: UDP-based Data Transfer for High-Speed Wide Area Networks", present a new application-level protocol with user-configurable congestion control and a more expressive API (Application Programming Interface). Their protocol trades packet-based feedback with timer-based feedback, emulating how polling often replaces interrupts for high performance network interfaces. Again considering the MCL problem, Huang, Lin and Ren propose a new transport protocol leveraging population ecology theory in "A Novel High Speed Transport Protocol Based on Explicit Virtual Load Feedback". They treat network flows as species, sending rates of the flows as population numbers and bottleneck bandwidth as the food resources, and relying on explicit router feedback to develop high performance independent of the flow RTT.

The final three papers address new mechanisms that allow existing protocols to adapt more effectively to the fast, long distance network environment. Router feedback is used in a different way to determine an appropriate sending rate to avoid an amplified slow-start penalty for small transfers in Sarolahti, Allman, and Floyd's paper "Determining an Appropriate Sending Rate over[1] an Underutilized Network Path". They explore the use of their Quick-Start mechanism to enable a flow to advertise a desired sending rate and the network to adjust or confirm this rate explicitly. Another use of router feedback presents fine-grain information on link capacities, available bandwidth, queue length, queue size, and loss rate. In their paper "An Explicit Router Feedback Framework for High Bandwidth-Delay Product Networks," Nakauchi and Kobayashi present a framework for capture and dissemination of this data, and consider the overhead involved in supporting this capability. Using this kind of information in transport protocols presumes TCP (or its variants) or an equivalent 'TCP-friendly' congestion control (TFRC) is available. TFRC is an equation-based system that emulates TCP behavior, but experiences similar challenges in fast, long networks. Xu describes this issue in "Extending Equation-based Congestion Control to High-Speed and Long-Distance Networks", and proposes an alternative equation for this environment.

All seven papers address the challenges that fast, long distance networks present to providing high performance, efficient, and fair data transport. They explore a variety of approaches, trading positive and negative feedback, event and timer-based response, and implicit and explicit router participation. Some create entirely new protocols whereas others augment existing ones. They represent the breadth of approaches now being considered so that the future of high-speed networking will be productive for us all.

Katsushi Kobayashi received his B.E., M.E., and Ph.D. degrees in Engineering, at the University of Electro-Communications (UEC), Tokyo, Japan in 1987, 1989, and 1993, respectively. He was a research associate with Computer Center, UEC during 1993–1998. He joined the National Institute of Information and Communications Technology (NICT formerly named CRL), Japan in 1998-2006. He is now at the Grid Technology Research Center in Advanced Industrial Science and Technology (AIST), Japan. He is a member of the IEEE and the ACM.

Pascale Vicat-Blanc Primet received an M.S. in CS from INSA de Lyon, (National Institute of Applied Science) in 1984, a Ph.D. in CS in 1988 and an HDR (Habilitation à diriger des Recherches) in 2002 from the University of Lyon. From 1989 to 2004, she was Associate Professor in Computer Science and Mathematics Department at the Ecole Centrale de Lyon. She joined INRIA as permanent researcher in 2005 where she is now Research Director and leading the INRIA RESO team within the LIP laboratory of the Ecole Normale Superieure de Lyon. She is expert in the scientific "Networks and Telecoms" committees of CNRS (National Research and Science Center) and of the French National Research Agency. Her interests include network and Internet protocols, network architecture, quality of service, network measurement, high-speed and low-latency nets, programmable networks, Grid computing and Grid networking. She has co-chaired the GGF Data Transport Research Group, and participates in a variety of European Grid projects including DataGRID, DataTAG, eToile, GRID5000, and GdX. She is a member of the PFLDnet and Gridnets conference steering committees and of numerous international program committees, and is active in the OGF (Open Grid Forum).

[1] Paper uses upper-case "Over"; match use in published version (either way).

Joseph D. Touch received a B.S. (Hons.) in biophysics and CS from the University of Scranton in 1985, an M.S. in CS from Cornell University in 1987, and a Ph.D. in CS from the University of Pennsylvania in 1992. He joined the University of Southern California/Information Sciences Institute (ISI), Marina del Rey, California, in 1992, where he is Director of the Postel Center and a Research Associate Professor in USC's CS and EE/Systems Departments. He is currently supporting the US Air Force's Transformational Communications Satellite (TSAT) program as Senior Network Engineer of its Space Segment. His interests include overlay networks, Internet protocols, network architecture, high-speed and low-latency nets, and optical network device design. He is a member of Sigma Xi, ACM Sigcomm Conference Coordinator, a senior member of the IEEE, and was Infocom 2006 Program Chair. He is a member of numerous conference steering and program committees, is active in the IETF, and serves on the editorial board of IEEE Network.

Katsushi Kobayashi
National Institute of Information and
Communications Technology,
Communication Research Laboratory,
Japan
E-mail address: ikob@koganei.wide.ad.jp

Pascale Vicat-Blanc Primet
Laboratoire d'Informatique du Parallélisme:LIP,
Ecole Normale Superieure 46, Allee d'Italie 69364,
LYON, Cedex 07, France
E-mail address: pascale.primet@inria.fr

Joe Touch
USC/ISI, 4676 Admiralty Way, Marina del Rey,
CA 90292-6695, United States
E-mail address: touch@isi.edu

Available online 3 January 2007