A WAN-in-Lab for Protocol Development

George S. Lee, Lachlan L. H. Andrew, David X. Wei, Bartek P. Wydrowski, Cheng Jin,
John Doyle, Steven H. Low, Harvey Newman
California Insitute of Technology, 1200 E. California Blvd, Pasadena, CA 91125, USA
Email: {gsl,lachlan,weixl,bartek,cjin,slow}@caltech.edu, doyle@cds.caltech.edu, newman@hep.caltech.edu

***Index Terms*— Testbed, flow control**

I. Introduction

It has recently become apparant that very high speed wide area networks (WANs) are pushing the limits of existing protocols, such as the transmission control protocol (TCP). This has led to many proposed replacements for TCP [1], [2], [3], [4], [5], [6], [7], [8], which must be evaluated, optimized and eventually developed into the next generation of TCP. Although initial design and testing can be performed using mathematical modeling and software simulation, there is ultimately a need to implement the selected algorithms in real networks. This paper describes a test facility called WAN-in-Lab, consisting of a complete WAN in a laboratory environment, which is available for use by the global networking research community.

After a description of the motivation for this approach, the specific hardware will be described, followed by examples of protocols that have already been tested on the facility. Finally, the future direction of the WAN-in-Lab will be outlined.

II. Approaches to protocol testing

Developing protocols requires many forms of testing. Existing tools are listed below, in decreasing order of abstraction. Except recirculating loops, all of these play a role in developing congestion control protocols. However, there is a big gap between emulation and production networks, which WAN-in-Lab seeks to fill.

a) Mathematical modeling explores fundamental limits of algorithms. It is the only tool that can explore infinite classes of networks rather than specific instances, but analysis requires simplification of actual protocols, ignoring important implementation issues.

b) Simulation is the most inexpensive way to test new protocols, and is a useful first step. However, network simulation is hundreds of times slower than real-time, and slower for high-speed networks. Moreover, it does not allow actual prototype deployments to be tested and optimized.

c) Emulation of large delays and link impairments using Dummynet also works well for links below 1 Gbit/s. A successful example of this approach is EmuLab [9]. Unfortunately, software emulation introduces artefacts into the traffic, which become increasingly severe at higher link rates.

d) Recirculating optical loops can also emulate long physical links with high delay. This approach is suitable for studying physical-layer effects on single packets, but cannot delay a continuous stream of packets, and hence are unsuitable for flow control experiments.

e) Production networks are the ultimate testing ground for new protocols, allowing "black-box" evalution through limited end-to-end measurements. This is suitable for tests which will not disrupt other traffic, but not for testing many failure modes, such as heavily overloaded networks, or equipment failures.

III. Hybrid WAN in Lab

The aim of WAN-in-Lab is to provide a realistic but controlled environment, which enables detailed monitoring of all aspects of protocol operation. This will allow an integrated approach, where theory development, implementation, experiments, and deployment inform and influence each other intimately. WAN-in-Lab is an open resource, intended for use by the entire networking community.

WAN-in-Lab is centred around an SDH/Sonet optical backbone with four OC48 routers, with wireless and gigabit Ethernet access networks.

Delay is provided by 24 spools of 100 km of single-mode fiber, which can be seen in the lower middle of the following figure:

To increase the delay, data traverses a spool four times. A router's 2.5 Gbit/s OC48 stream is multiplexed onto a 10 Gbit/s OC192 stream, which is amplified and returned. The returning stream is multiplexed onto a second OC192 slot, and the process is repeated. A single pass through each spool gives a delay of only 0.5 ms, but with multiple passes and the associated multiplexing delays, the round trip delay approaches 5 ms per spool.

Currently, WAN-in-Lab has four Cisco routers, each with a single OC48 line card. This allows two links with an aggregate delay of 100 ms, in steps of 5 ms. More complex topologies can be created with short gigabit Ethernet (GbE) links between the routers and the 20 high-speed servers. Some of these servers are configured as Dummynets, while others are software routers to test active queue management (AQM) protocols.

IV. Preliminary Results

To date, WAN-in-Lab has been used to test two protocols: FAST [1], [2] and MaxNet [8], [10].

A. MaxNet

MaxNet is a flow control algorithm which uses explicit feedback from routers to set a sender's transmission rate. Thus, it can only be tested in networks which provide access to the bottleneck router, such as WAN-in-Lab. MaxNet encodes a congestion signal in the IP header; if the congestion level of a router is larger than the value of an incoming packet, it overwrites the field. The sender then adjusts its sending rate based on the indicated congestion in a provably stable way.

A four-node MaxNet network has been established on WAN-in-Lab. Two additional servers are configured as software routers to implement MaxNet's AQM algorithm.

These experiments have resulted in important improvements to the original MaxNet algorithm of [8]. In particular, a quick-start phase has been added to improve the performance of the algorithm when flows arrive and depart dynamically. These experiments are reported in [10].

B. FAST

Research into FAST TCP was the original motivation for WAN-in-Lab. FAST determines a sender's rate based on the difference between the round trip time (RTT) when queues are full and when they are empty. Thus it has been important to investigate the impact of real-world timing jitter, due to such things as operating system task scheduling, on measuring the small queueing delay of a high-speed network.

WAN-in-Lab has been used [11] to investigate the automated selection of the number of packets FAST should buffer in the network, based on the interaction between loss and delay. This issue is one of the important remaining research issues for FAST.

V. Future Directions

WAN-in-Lab is still evolving. In the near future, we expect it to have:

- A connection to the Ultralight 10 Gbit/s experimental/ production network [12]. This will allow the monitoring facilities of WAN-in-Lab to study real-world traffic, allow studies of incremental deployment of protocols, and increase the range of topologies available to WAN-in-Lab.
- A richer set of topologies, through the acquisition of additional line cards.
- Remote reconfigurability using a 144×144 port optical switch. This will allow topology changes between experiments or within an experiment. Rerouting is important for delay-based protocols such as FAST.
- Detailed monitoring of link traffic. This will allow the queue sizes at routers to be tracked, and the burstiness of data to be monitored. Both of these factors play an important role in the design of flow control protocols.
- Management tools to allow researchers to configure the network easily, both rearranging the network topology and installing custom kernels on the servers.

As well as developing flow control protocols, WAN-in-Lab will be useful for benchmarking proposals. There have been proposals to establish a general benchmark for various TCP algorithms. However, the difference in hardware setup affects the benchmark results of a TCP algorithm. WAN-in-Lab is an ideal standard hardware platform for comparing different TCP algorithms with a benchmark suite. We are currently porting a general suite of tests from an emulated Dummynet framework to use WAN-in-Lab.

Acknowledgements

WAN-in-Lab is funded by NSF (EIA-0303620), Cisco ARTI, ARO (W911NF-04-1-0095), and Corning. We also thank Julian Bunn, Raj Jayaman, Xun Su, Yang Xia of Caltech and Bob Aiken, Vijay Doraiswami, Russ Esmacher, Kevin McGrattan, Chris McGugan, Charles Smith, Doug Waltsen, and Steven Yip of Cisco.

References

[1] C. Jin, D. X. Wei, S. H. Low, G. Buhrmaster, J. Bunn, D. H. Choe, R. L. A. Cottrell, J. C. Doyle, W. Feng, O. Martin, H. Newman, F. Paganini, S. Ravot, and S. Singh. FAST TCP: From theory to experiments. *IEEE Network*, 19(1):4-11, 2005.

[2] C. Jin, D. X. Wei, and S. H. Low. TCP FAST: Motivation, architecture, algorithms, performance. In *Proceedings of IEEE Infocom*, March 2004. `http://netlab.caltech.edu`.

[3] S. Floyd. HighSpeed TCP for large congestion windows. Internet draft draft-floyd-tcp-highspeed-02.txt, work in progress, `http://www.icir.org/floyd/hstcp.html`, February 2003.

[4] R. N. Shorten and D. J. Leith. H-TCP: TCP for high-speed and long-distance networks. in *Proc. PFLDnet*, Argonne, 2004. `http://hamilton.ie/net/htcp3.pdf`.

[5] T. Kelly. Scalable TCP: Improving performance in highspeed wide area networks. *Computer Communication Review*, 32(2), April 2003. `http://www-lce.eng.cam.ac.uk/~ctk21/scalable/`.

[6] L. Xu, K. Harfoush, and I. Rhee. Binary increase congestion control for fast long distance networks. In *Proc. INFOCOM*, March 2004.

[7] D. Katabi, M. Handley, and C. Rohrs. Congestion control for high-bandwidth delay product networks. In *Proc. Sigcomm*, Aug. 2002.

[8] B. Wydrowski, L. L. H. Andrew, and M. Zukerman, "MaxNet: A congestion control architecture for scalable networks," *IEEE Commun. Lett.*, 7:511–513, Oct. 2003.

[9] B. White, J. Lepreau, L. Stoller, R. Ricci, S. Guruprasad, M. Newbold, M. Hibler, C. Barb and A. Joglekar, "An Integrated Experimental Environment for Distributed Systems and Networks", In *Proc. Fifth Symp. Operating Systems Design and Implementation* Dec. 2002, pp. 255–270. `http://www.emulab.net/`.

[10] M. Suchara, R. Witt and B. Wydrowski. TCP MaxNet - Implementation and Experiments on the WAN in Lab. In *Proc. Int. Conf. Networks* Kuala Lumpur, Malaysia, pp. 901–906, November 2005.

[11] A. Tang, C. Jin and S. H. Low. Alpha tuning for FAST TCP. Private correspondence, 2005.

[12] Ultralight: An ultrascale information system for data intensive research. `http://www.ultralight.org`.