Measurements On Hybrid Dedicated Bandwidth Connections

Nageswara S. V. Rao[*], William R. Wing[*], Qishi Wu[¶], Nasir Ghani[†], Tom Lehman[‡], Chin P. Guok[§] and Eli Dart[§]
[*] Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
Email: {raons, wrw}@ornl.edu
[†] Department of Electrical and Computer Engineering, Tennessee Technological University, Cookville, TN 38505, USA
Email: nghani@tntech.edu
[‡] Information Sciences Institute East, University of Southern California, Arlington, VA 22203, USA, Email: tlehman@isi.edu
[§] Network Engineering Services Group, ESnet, Berkeley, CA94720, USA, Email: {chin, dart}@es.net
[¶] Department of Computer Science, University of Memphis, Memphis, TN 38152, USA, Email: qishiwu@memphis.edu

I. Introduction

Next generation large-scale science and commercial applications are expected to generate datasets in the range of terabytes to petabytes, which have to be transported over wide-area networks. Efforts to support such applications on shared IP networks have not been very successful since the available bandwidth varies in response to "other" network traffic. The dedicated bandwidth connections are promising because they can offer: (i) large unimpeded link capacity for massive data transfer operations, and (ii) dynamically stable bandwidth for monitoring and steering operations. Several network research projects are currently underway to develop such capabilities. They include User Controlled Light Paths (UCLP) [9], UltraScience Net (USN) [8], Circuit-switched High-speed End-to-End Transport ArcHitecture (CHEETAH) [11], Enlightened [4], Dynamic Resource Allocation via GMPLS Optical Networks (DRAGON) [1], Japanese Gigabit Network II [7], Bandwidth on Demand (BoD) on Geant2 network [5], On-demand Secure Circuits and Advance Reservation System (OSCARS) [2] of ESnet, Hybrid Optical and Packet Infrastructure (HOPI) [6], Bandwidth Brokers [10], and other networks. Such deployments are expected to proliferate widely as reflected by production networks, such as Internet2 and ESnet, offering on-demand circuits, Multiple Protocol Label Switching (MPLS) tunnels and dedicated Virtual Local Area Networks (VLAN) connections.

Dedicated bandwidth connections may be provisioned at layers 1 through 3 or as combinations. They can be MPLS tunnels over routed network as in ESnet [2], or Ethernet over SONET as in CHEETAH [11], or Infiniband (IB) over SONET as in USN [3], or pure Ethernet paths [1]. An objective comparison of the characteristics of connections provisioned using these varied technologies is critical to making deployment decisions for production networks. Once deployed, the costs of replacing one by another could be extremely high, for example, replacing MPLS tunnels with SONET circuits entails replacing all routers with switches. Towards this objective, we collect measurements and compare the throughputs and message delays over OC21C SONET connections, 1Gbps MPLS tunnels, and their concatenations over USN and ESnet.

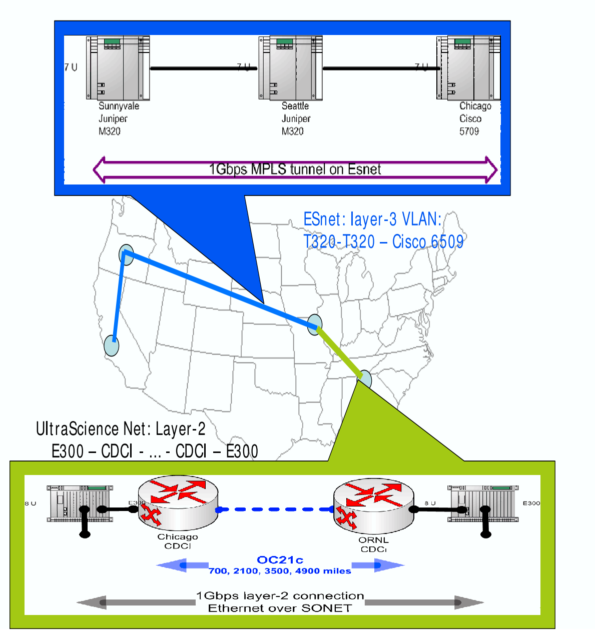

Fig. 1. **USN switched OC21C connections and MPLS tunnels implemented using ESnet routers are peered using an Ethernet switch.**

II. USN and ESnet Connections

On USN we utilize the OC192 links between Ciena SONET switches at ORNL and Chicago to realize OC21C connections of lengths 700, 1400, ..., 6300 miles by suitably switching them. At the end points we map 1Gbps Ethernet onto OC21C, thereby realizing 1GigE connections of various lengths. On ESnet, 1Gbps VLAN-tagged MPLS tunnel is setup between Chicago and Sunnyvale via Cisco and Juniper routers, which is about 3600 miles long. USN peers with ESnet in Chicago as shown in Figure 1, and 1GigE USN and ESnet connections are cross-connected using Force10 Ethernet switch. Together, this configuration provides us hybrid dedicated channels of varying lengths, namely 4300, 5700, ... ,9900 miles, composed of Ethernet mapped layer 1 and layer 3 connections.

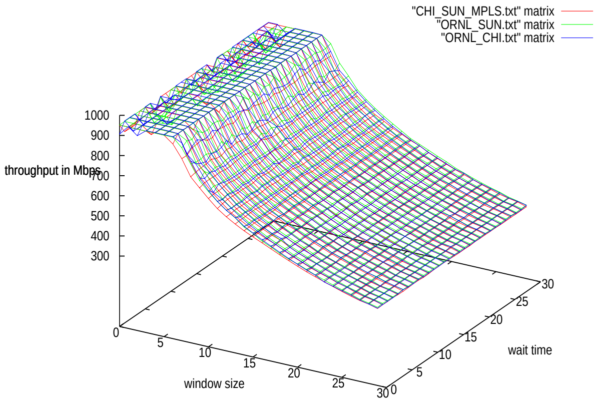

Fig. 2. **Transport profiles of 6300 mile OC21C, 3600 mile 1Gbps MPLS and their concatenated connections are very similar.**

III. Throughput Measurements

We collected throughput measurements using iperf and Peak Link Utilization Protocol (PLUT) over SONET, MPLS and concatenated connections. For iperf TCP, the number of streams *n* is varied between 1 and 10, and for iperf UDP the target rate is varied as 100, 200, ..., 1000, 1100 Mbps; each set of measurements is repeated 100 times. First, we compare USN and ESnet connections of lengths 3500 and 3600 miles respectively and their concatenation. TCP throughput is maximized when *n* is around 7 or 8 and remained constant around 900, 840 and 840 Mbps for SONET, MPLS and hybrid connections, respectively. For UDP, the peak throughput is 957, 953 and 953 Mbps for SONET, MPLS and hybrid connections, respectively. Thus there is difference of 60Mbps and 4Mbps between the TCP and UDP peak throughputs, respectively, over SONET and MPLS connections. There is a difference in peak throughput achieved by TCP and UDP in all cases, in particular, 57 and 93 Mbps for SONET and MPLS connections, respectively. This difference is in part due to the congestion control of TCP, and the high UDP bandwidth makes it a viable candidate for transport since there is no "congestion" on dedicated channels. We measured file transfer rates over these connections using UDP-based PLUT, which achieved 955, 952 and 952 over SONET, MPLS and hybrid connections, respectively. Thus the iperf UDP bandwidth estimate is indeed is achievable in actual file transfers.

We compute the connection *throughput profile* by sending UDP datagrams at varying rates and measuring PLUT goodput at the destination. The sending rate is controlled by transmitting a number of datagrams, denoted by the *window size* $W_c(t)$, in a single burst and then waiting for a time period called the *wait time* $T_s(t)$. Thus the sending rate is specified by a point in the horizontal plane, given by $(W_c(t), T_s(t))$, and its corresponding goodput is shown in Figure 2 for 6300 mile OC21C, 3600 mile 1GigE MPLS and the concatenated connection. All three transport profiles are very similar, which explains the PLUT throughput results described above.

IV. Message Delay Measurements

To estimate the jitter properties, we collected three types of measurements: (a) Ping measurements correspond to round trip time estimates based on ICMP measurements. (b) For tcpmon measurements, a fixed-size message is sent from a client to server via TCP specifying the size of a message to be sent back to client via TCP. The time duration between starting of first message and receipt of second message is estimated at the client. (b) For TCP client-server measurements, a message is sent from the client, which is read by server and sent back to client. The round trip time is estimated at the client. To estimate the trends, we compare mean delay (ms) and its range (% of mean) over 3600 mile MPLS connection with OC21C connections of lengths 2800 and 4200 miles as follows.

connection	ping	tcpmon	cli-ser
2800m OC21C	53.4; 0.1%	53.5; 0.2%	54.5; 0.4%
3600m MPLS	67.5; 0.1%	67.6; 0.1%	68.7; 0.3%
4200m OC21C	79.9; 0.2%	79.9; 0.03%	81.0; 0.3%

V. Conclusions

We demonstrated that connections provisioned at layers 1-3 can be peered and carried across networks using VLAN technologies. Throughput and message delay measurements (IP level) collected over USN and ESnet indicate a comparable performance of layer 1, layer 3 and hybrid connections. Due to the page limit, we only outlined a small sample of our measurements, and the entire set will be included in a complete version of this paper. While being instructive, these results are only anecdotal, and a careful design of experiments and a finer analysis of measurements using methods such as regression interpolation would be needed to gain a deeper understating of the relative performance of these connections.

Acknowledgments

This research is sponsored by the High Performance Networking Program of the Office of Science, U.S. Department of Energy, under Contract No. DE-AC05-00OR22725 with UT-Battelle, LLC, and by National Science Foundation Under Grant No. ANI-0335185.

References

[1] Dynamic resource allocation via GMPLS optical networks. http://dragon.maxgigapop.net.

[2] On-demand secure circuits and advance reservation system. http://www.es.net/oscars.

[3] S. M. Carter, M. Minich, and N. S. V. Rao. Experimental evaluation of high-performance file systems over local and wide-area networks. In *Proceedings of High Performance Computing Conference*. 2007.

[4] Enlightened Computing, http://www.enlightenedcomputing.org/.

[5] Geant2, http://www.geant2.net.

[6] Hybrid Optical and Packet Infrastructure, http://networks.internet2.edu/hopi.

[7] JGN II: Advanced Network Testbed for Research and Development, http://www.jgn.nict.go.jp.

[8] N. S. V. Rao, W. R. Wing, , S. M. Carter, and Q. Wu. Ultrascience net: Network testbed for large-scale science applications. *IEEE Communications Magazine*, 2005.

[9] User Controlled LightPath Provisioning, http://phi.badlab.crc.ca/uclp.

[10] Z. L. Zhang, Z. Duan, and Y. T. Hou. Decoupling QoS control from core routers: A novel bandwidth broker architecture for scalable support of guaranteed services. In *Proc. ACM SIGCOMM*. 2000.

[11] X. Zheng, M. Veeraraghavan, N. S. V. Rao, Q. Wu, and M. Zhu. CHEETAH: Circuit-switched high-speed end-to-end transport architecture testbed. *IEEE Communications Magazine*, 2005.