SC26 Network Research Exhibition: Demonstration Abstract

Prototype Multi-Tbps WAN Infrastructure: Architecture, Technology, and Control Systems

- Joe Mambretti, j-mambretti@northwestern.edu, Jim Chen, jim-chen@northwestern.edu, Fei Yeh, fyeh@northwestern.edu, International Center for Advanced Internet Research - Northwestern University, Linden Mercer (NRL), linden.mercer.ctr@nrl.navy.mil, Bill Fink (NASA), Tom Lehman (ESnet), Rod Wilson (Ciena), rwilson@ciena.com, Marc Lyonnais, (Ciena), mlyonnai@ciena.com, Scott Kohlert, (Ciena), skohlert@ciena.com, Gaurauvdeep Shami, (Ciena), gshami@ciena.com

Abstract

Large scale data production within and among science research collaborations and sites continues to increase, a long-term trend that continues to accelerate rapidly, driven especially by the deployment of new science facilities, including high luminosity research instrumentation and next generation computational science centers. Consequently, the science networking community is planning networking infrastructure that can support services beyond 100Gbps, with one thematic focus on 400 Gbps LANs and WANs (represented also by other NRE demonstrations planned by this consortium). However, this consortium is also investigating capabilities for WAN services beyond 400 Gbps, including 800 Gbps, 1.2 Tbps, 1.6 Tbps, and multi-Tbps WAN and LAN services. This NRE describes the required infrastructure for prototype multi-Tbps WANs. *NB: A companion NRE describes a demonstration of the Multi-Tbps services that will be supported by this infrastructure.*

Overview

When 400 Gbps WAN/LAN services were initially being anticipated, the science networking community began to develop techniques to effectively utilize that level of capacity. Standard implementations and configurations as commonly used were not able to rise to the level of available network capacity. Similarly, today, the requirements and implications of WAN and LAN services beyond 400 Gbps are being explored, including technologies that potentially will be able to support 800 Gbps, 1.2 Tbps, 1.6 Tbps, and multi-Tbps WAN services E2E. (NB: LHC Tier 1 site and some Tier 2 sites are planning for at a minimum 1 Tbps connections.) These SC26 demonstrations will showcase capabilities and technologies that may assist in providing such services by demonstrating prototype multi-Tbps WAN transport services from the StarLight International/National Communications Exchange Facility in Chicago to the SC26 venue, 3*1.2 Tbps and 3*800 Gbps between the JBDT Facility in McLean Va. to the SC26 venue, including channels using coherent optics, and 3*1.2 Tbps between StarLight and the JBDT Facility at McLean, Virginia, and 1.2 Tbps between the LHC testbed at the University of Victoria in Canada, over CANARIE to PacWave, to the PNWGP, to 3*400 Gbps NA REX paths to the StarLight Exchange, to the SC26 venue.

Goals

With its research partners including the SCinet WAN group, the Ciena research group, NASA, the NLR, the International Center for Advanced Internet Research (iCAIR) at Northwestern University is designing multi-Tbps WAN services among the sites noted. Multiple issues are being investigated and developed to enable the utility of multi-Tbps WAN services E2E.

1. One area of investigation is the potential to implement E2E lightpaths for data intensive science across multidomain networks (an important consideration for large scale science), avoiding as much as possible (and in some cases completely) L3 devices

2. At the edges of the WAN paths, this project is implementing multiple optimized Gen5 Data Transfer Nodes (400 Gbps DTNs, minimal 3 at each site. The potential for Gen6 DTNs is being

explored to enable 800 G on single DTNs, enabling 1.6 Tbps WAN services.).

3. These DTNs are being implemented with 400 Gbps smart NICs, high performance optical transport interfaces. Currently, 800 Gbps smart NICs are also being investigated by this consortium.

4. These DTNs are connected to deep buffer 400 Gbps L2 switches. This consortium is also researching the evolution of 800 Gbps switches.

5. In addition, these demonstrations will showcase the performance, flexibility, reliability, and programmability of WAN open line systems and open transponder platforms

6. Other technologies being explored are high performance optical transport switches, long range transceivers (including ZR/ZR+ transceivers, with various formats being investigated including those based on Indium Phosphide that do not require EDFAs or tunable optical filters and silicon photonics), and various types of breakout cables.

7. This infrastructure will incorporate the SCinet HPC cluster and a mobile FABRIC node implemented at the SC26 venue.

8. INT components supporting monitoring and analytics include Tofino2 programmable switches used for line-rate packet inspection, e.g., using P4 programmable logic to edit payloads, extracting metadata-rich headers for analysis from general scientific traffic for analysis, b) FPGA-Based SmartNICs (e.g., Xilinx Alveo U55C), processing header-only data streams, using P4 and c) NVIDIA BlueField-2 DPUs.

Impacts

1) This series of research projects is directed at developing resources to meet the demands of increasingly large capacity WAN transport for large scale data intensive science. This NRE is focused on WAN infrastructure that can support mult-Tbps services. (ref related NRE on multi-Tbps WAN services)
2) SCinet supports capabilities that enable innovative multi-Tbps WAN techniques to be demonstrated at a national scale.
3) Multiple next steps include migration to Gen6 DTNs, 800 Gbps smart NICs, extending capabilities to international sites, integrating new protocols, integrating ZR/ZR+ coherent optics, developing additional monitoring and analytics, including those using AI techniques etc.

Resources

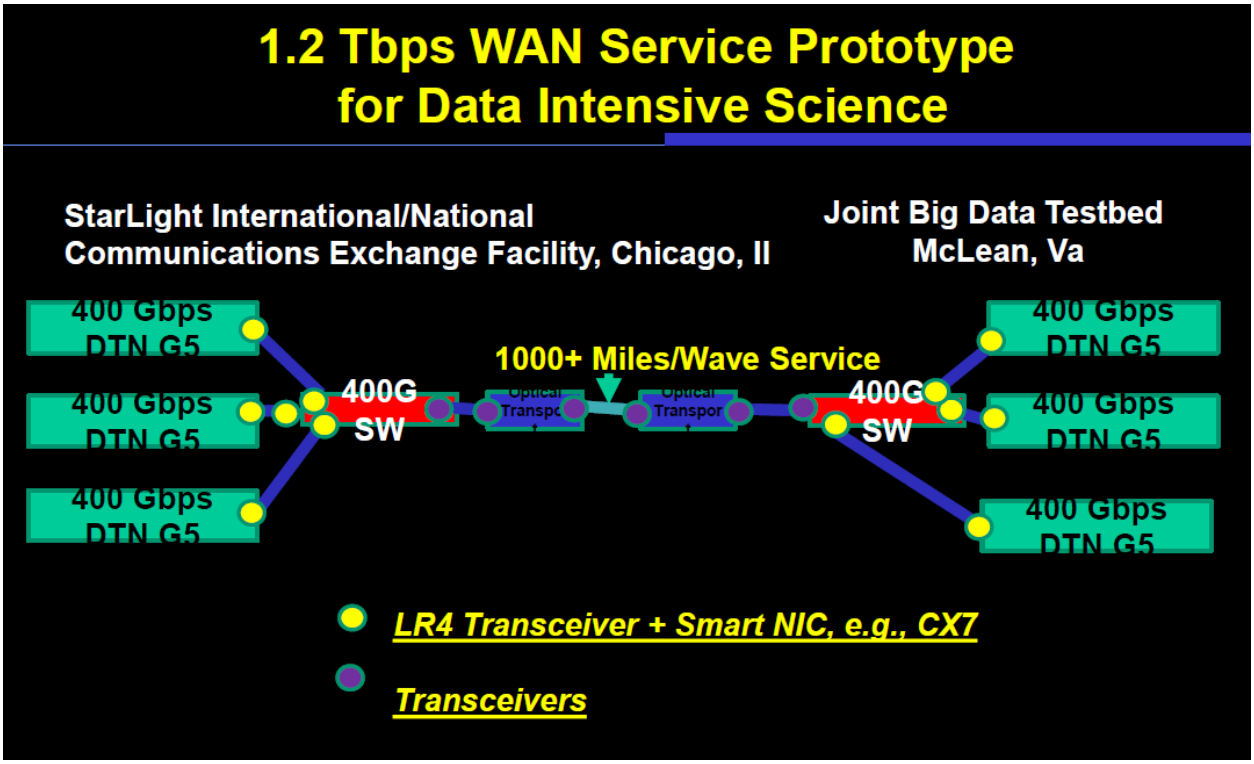
Required resources from the SCinet WAN are 3.6 Tbps and 3*800 Gbps WAN transport services from the StarLight International/National Communications Exchange Facility in Chicago to the SC26 venue, 3.6 Tbps and 3*800 Gbps from the JBDT Facility in McLean Va. to the SC26 venue (e.g., 3*400 Gbps on FABRIC and 2*1.2 Gbps on the CENI testbed), and 1.2 Tbps between the University of Victoria in Canada across CANARIE, PacWave, the PNWGP, NA REX and StarLight and the JBDT Facility at Mclean. NB: Many of these connections have been implemented. Ref NRI103.

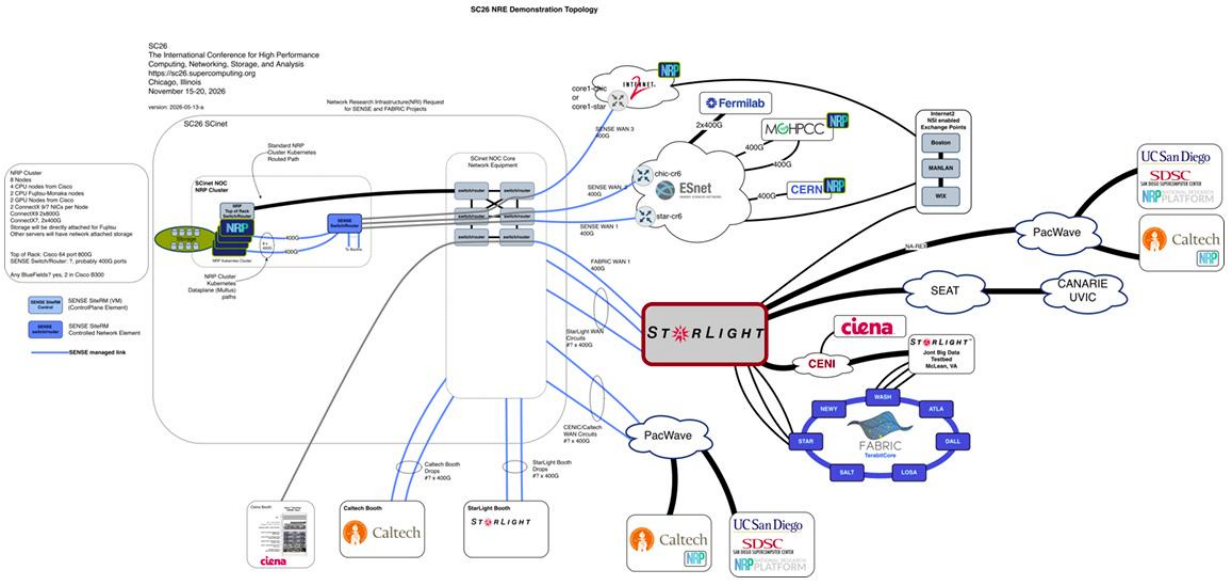
Involved Parties

- Joe Mambretti, iCAIR, jmambretti@northwestern.edu
- Jim Chen, iCAIR, jim-chen@northwestern.edu
- Fei Yeh, iCAIR,fyeh@northwestern.edu
- Linden Mercer, linden.mercer.ctr@nrl.navy.mil
- Bill Fink billfink@mindstream.com
- Tom Lehman, tlehman@es.net
- Rod Wilson, Ciena, rwilson@ciena.com
- Marc Lyonnais, Ciena, mlyonnai@ciena.com
- Scott Kohlert, Ciena, skohlert@ciena.com

- Gaurauvdeep Shami, Ciena, gshami@ciena.com
- StarLight International/National Communications Exchange Facility and Consortium
- Metropolitan Research and Education Network (MREN)
- SC26 SCinet

Below is an abstract of a prototype 1.2 Tbps infrastructure architecture





Initial SCinet topology – Tom Lehman – This topology does not include all paths, e.g., the complete NA REX infrastructure and all of the paths between the StarLight Facility and the SC26 venue.