

SC25 Network Research Exhibition: Demonstration Abstract

Prototype 1.2 Tbps WAN Services: Architecture, Technology and Control Systems

Joe Mambretti, Jim Chen, Fei Yeh, International Center for Advanced Internet Research - Northwestern University

j-mambretti, jim-chen, fyeh@northwestern.edu, Linden Mercer, linden.mercer.ctr@nrl.navy.mil, Bill Fink bill@wizard.sci.gsfc.nasa.gov, Tom Lehman, tlehman@es.net, Rod Wilson, Ciena, rwilson@ciena.com, Marc Lyonais, Ciena, mlyonnai@ciena.com, Scott Kohlert, Ciena, skohlert@ciena.com, Gauravdeep 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, especially driven by the deployment of new science facilities, including high luminosity research instrumentation and next generation high performance computation science centers. Consequently, the science networking community has begun to prepare infrastructure that can support high capacity services, with one key 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 those approaching 800 Gbps, 1.2 Tbps, and multi-Tbps WAN and LAN services.

NB: A companion NRE describes a demonstration of the 1.2 Tbps infrastructure required for a large scale 1.2 Tbps service. This NRE describes prototype 1.2 Tbps WAN services that will be supported by the infrastructure described in that NRE.

Overview

When 100 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 WAN services and multi-Tbps services E2E. (NB: LHC Tier 1 site and some Tier 2 sites are planning for minimal 1 Tbps connections.) These SC25

demonstrations showcase capabilities and technologies that may assist in providing such services by demonstrating prototype 1.2 Tbps WAN transport services from the StarLight International/National Communications Exchange Facility in Chicago to the SC25 venue, 1.2 Tbps from the JBDT Facility in McLean Va. to the SC25 venue, and 2*1.2 Tbps between the StarLight Exchange (including the JBDT facility at that site) and the JBDT Facility at McLean, Virginia, and 1.2 Tbps between the ESnet testbed at Berkeley and the StarLight Exchange.

Goals

With its research partners including the SCinet WAN group, the Ciena research group, the International Center for Advanced Internet Research (iCAIR) at Northwestern University is designing a 1.2 Tbps WAN service among the sites noted. Multiple issues are being investigated to enable to utility of 1.2 Tbps WAN services E2E. One area of investigation is the potential to implement E2E lightpaths across multidomain networks, avoiding as much as possible (and in some cases totally) L3 devices or services.

1. At the edges of all WAN paths, this project will implement multiple optimized Gen5 400 Gbps Data Transfer Nodes (DTNs, with a minimal of three DTNs at each site).
2. Each DTN will be implemented with 400 Gbps smart NICs configured with high performance optical transport connectors.
3. These demonstrations will leverage experimental research into the optimal design, configuration, components, and integration technologies for DTNs, including 100, 200 and

400 Gbps NICs, NVMe over Fabric, and techniques for kernel bypass using zero-copy for memory and disk copy to avoid bottlenecks in large scale data transfer over 1.2 Tbps WAN and optimal affinity bindings for NUMA architecture for higher resource utilization

4. These demonstrations will also showcase middleware required to orchestrate infrastructure resources for reliable, optimized high-speed network data transfers.
5. Technologies that will be demonstrated include Software Defined Networking (SDN), RDMA, RoCE, NVMe over fabric, GPU acceleration, and P4.
6. These demonstrations will also demonstrate measurement techniques for real-time monitoring, benchmarking and evaluation including AI toolset.
7. This WAN service investigation include examining implications for supporting 1.2 Tbps WAN services with production operation processes.

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 services (vs underlying infrastructure) that supports 1.2 Tbps services. (ref related NRE focused on 1.2 Tbps WAN infrastructure).
- 2) SCinet supports capabilities that enable innovative 1.2 Tbps WAN services and techniques to be demonstrated at a national scale.

- 3) Multiple next steps include migration to network software stacks based on Gen6 servers, 800 G smart NICs, extending capabilities to international sites, integrating new protocols, developing additional monitoring and analytics using AI techniques etc.

Resources

Required resources from SCinet WAN include 1.2 Tbps WAN transport services among the StarLight International/National Communications Exchange in Chicago, the Joint Big Data Testbed Facility in McLean Virginia, and the SC25 venue (e.g., 3*400 Gbps on FABRIC and 2*600 Gbps on the CENI testbed).

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 bill@wizard.sci.gsfc.nasa.gov,
- Tom Lehman, tlehman@es.net
- Rod Wilson, Ciena, rwilson@ciena.com
- Marc Lyonais, Ciena, mlyonnai@ciena.com
- Scott Kohlert, Ciena, skohlert@ciena.com
- Gauravdeep Shami, Ciena, gshami@ciena.com
- Ziqiang Wang, ziwang@ciena.com
- StarLight International/National Communications Exchange Facility and Consortium
- Metropolitan Research and Education Network (MREN)
- SC25 SCinet