DOE/MICS/SciDAC Network Research Program

Title: DOE UltraScience Net

PI: Nageswara S. V. Rao, William R. Wing
PI Institution: Oak Ridge National Laboratory

Project Website: http://www.csm.ornl.gov/ultranet

Abstract:
DOE UltraScienceNet is an experimental research network testbed with an objective to enable the development, testing and tuning of advanced networking and related technologies needed for DOE's large-scale science applications. It provides on-demand and advance-reserved, dedicated, high bandwidth channels for large data transfers, and also high resolution, high-precision channels for fine control operations. It is provisioned at the backbone using SONET switches and at the edges using Ethernet switches. The data-plane consists of multiple OC192 links from Atlanta to Chicago to Seattle to Sunnyvale, and the control-plane is implemented over an out-of-band VPN.

Introduction

National Leadership Class Facility (NLCF) supercomputers for large-scale science applications plan to provide more than 100 teraflops speeds within an year. They hold an enormous promise for meeting the demands of Department of Energy (DOE) large-scale science projects and programs, which span fields as diverse as earth science, high energy and nuclear physics, astrophysics, fusion energy science, molecular dynamics, nanoscale materials science, and genomics. These applications are expected to generate petabytes of data at the computing facilities. This data must be transferred, visualized and steered by geographically distributed teams of scientists. Similarly, in the experimental science arena, DOE currently or will soon operate several extremely valuable experimental facilities, such as the Spallation Neutron Source (SNS) and the Relativistic Heavy Ion Collider (RHIC). It also participates in the Large Hadron Collider (LHC) project. The ability to conduct experiments remotely and transfer the large measurement data sets is critical to ensuring the productivity of these facilities and the scientists using them. Indeed, the high-performance network capabilities add a whole new dimension to the access of these computing and experimental facilities, by eliminating the “single location, single time zone" bottlenecks that currently plague these valuable resources.

As has been shown in simulations and in practice, that it is very difficult, if not impossible, to sustain multiple 10 Gbps transport channels needed in these applications over shared IP networks such as current ESnet or Internet. Even for sub-lambda speeds of control channels, the requirements of both usable and stable bandwidth are extremely difficult to meet over current networks. This is primarily due to the shared nature of these TCP/IP networks, which leads to unpredictable and complex transport dynamics. By utilizing dedicated channels over switched circuits these difficulties can be overcome to a large extent. But, the existing network testbeds that provide such channels typically have an inadequate footprint or bandwidth to provide optimized solutions for high-performance deployments that connect various DOE sites.

DOE UltraScience Net

UltraScience Net (USN) is an experimental network research testbed with an objective to facilitate the development, testing and tuning of networking and related technologies for supporting distributed large-scale DOE science applications. It links Atlanta, Chicago, Seattle and Sunnyvale as shown in Figure 1, where each connection is currently supported by two 10 Gbps OC192 long-haul links. It provides on-demand and advance-reserved dedicated layer-1

and layer-2 channels at multi-, single- and sub-lambda resolutions between its core and edge switches. Its backbone is implemented using robust carrier-class SONET switches capable of setting up circuits at OC1 (50Mbps) granularity. In addition, Multi Service Provisioning platforms (MSPP) are located at the edges, which provide dedicated Ethernet channels at various resolutions. The hosts connected to edge switches provide environments to support the development and testing of protocols, middleware, and applications.

Figure 1. USN showing the span provided by dual 20 Gbs lambdas.

The backbone of dual 10Gbps lambdas currently utilizes the ORNL network infrastructure from Oak Ridge to Chicago, and National Lambda Rail from Chicago to Sunnyvale. At present, it provides connectivity to ESnet, Oak Ridge National Laboratory, Fermi National Laboratory, Pacific Northwest National Laboratory and California Institute of Technology.

User and Project Support

USN is based on the concept of giving users and applications a direct access to layer-1 (SONET) paths with zero packet re-ordering, (almost) zero jitter, and zero congestion. In addition, it also provides dedicated layer-2 (Ethernet) paths with low re-ordering, low jitter and no congestion. Users and applications can provision the dedicated channels over USN as needed or in advance for tasks such as data transfers or computations scheduled on supercomputers, or testing of new protocols, storage and file systems, or middleware. They can either utilize hosts located at USN edges or connect its core or edge switches using their own end connections. Various protocols, storage systems, middleware and application research projects in support of DOE large-scale science applications can make use of the provisioned dedicated channels. Certain production-level connectivity for research traffic can also be supported on these channels during the allocated time periods.

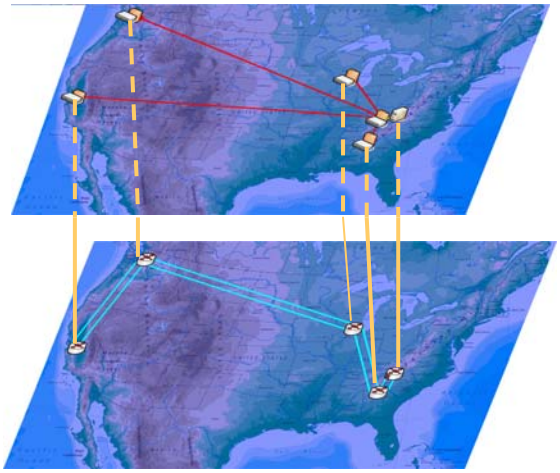

Figure 2. Out-of-band secure control-plane.

Control-Plane

The ability of the applications to actively access the control-pane potentially opens the whole infrastructure to cyber attacks that could hijack the control-plane and then prevent recovery through denial-of-service attacks. USN control-plane implemented over a Virtual Private Network (VPN) provides protection against such attacks. The control-plane operations are coordinated by a centralized system that: (a) maintains the state of bandwidth allocations on all links; (b) accepts and grants requests for current and future channels to applications; and (c) sends signaling messages to switches as per the schedule for setting up and tearing down the channels. Currently signaling is supported through TL1/CLI of the respective switches. The plans include peering with networks with Multiple Protocol Label Switching (MPLS) and Generalized GMPLS-based control planes, such as ESnet and NSF CHEETAH, respectively.