DOE UltraScienceNet - Update

Experimental Network Testbed for High-Performance Network technologies and Applications

Nagi Rao, Bill Wing, Steven Carter, Qishi Wu, Susan Hicks
Computer Science and Mathematics Division
Oak Ridge National Laboratory
{raons, wrw, scarter, wuqn}@ornl.gov
https://www.usn.ornl.gov

Sponsored by
High-Performance Networks Program
Mathematics, Information and Computational Science Division
Office of Advanced Scientific Computing Research
U.S. Department of Energy
Contents

Background and Overview
Data-Plane
Control-Plane
Experimental Results
DOE UltraScience Net: Need, Concept and Challenges

The Need
- DOE large-scale science applications on supercomputers and experimental facilities require high-performance networking
  - Moving petabyte data sets, collaborative visualization and computational steering (all in an environment requiring improved security)
- Application areas span the disciplinary spectrum: high energy physics, climate, astrophysics, fusion energy, genomics, and others

Promising Solution
- High bandwidth and agile network capable of providing on-demand dedicated channels: multiple 10s Gbps to 150 Mbps
- Protocols are simpler for high throughput and control channels

Challenges: Several technologies need to be (fully) developed
- User-/application-driven agile control plane:
  - Dynamic scheduling and provisioning
  - Security – encryption, authentication, authorization
- Protocols, middleware, and applications optimized for dedicated channels
DOE-Funded Support Application Projects

- Lambda-Station
  - FNAL-developed analysis “station” for high-energy physics
- Peering and Terascale Supernova Initiative
  - Collaborative visualization
  - Interdomain peering with NSF CHEETAH
- ESnet MPLS Tunnels
  - MPLS signaling to setup on-demand and in-advance circuits
- Remote Microscopy and Genomics Applications
  - PNNL developed remote-user control of confocal microscopy
USN Architecture:
Separate Data-Plane and Control-Planes

Secure control-plane with:
- Encryption, authentication and authorization
- On-demand and advanced provisioning

Dual OC192 backbone:
- SONET-switched in the backbone
- Ethernet-SONET conversion
DOE UltraScience Net: Data Plane

Connects Atlanta, Chicago, Seattle and Sunnyvale:
- Dynamic and in-advance provisioned dedicated dual 10Gbps links at 50 Mbps resolution – SONET or Ethernet
USN Data-Plane: Node Configuration

In the Core:
  - Two OC192 switched by Ciena CDCIs

At the Edge
  - 10/1 GigE provisioning using Force10 E300s

Data Plane User Connections:
Direct connections to:
  - core switches – SONET & 1GigE
  - MSPP – Ethernet channels
  - Utilize UltraScience Net hosts

Connections to CalTech and ESnet
USN Data-Plane: User Ports

- User connections
  - Ciena CDCI
    - SONET ports on CDCI
    - GigE ports on CDCI
  - Force10 E300
    - 10GigE ports on E300
    - GigE ports on E300

GigE ports must match at the connection end points
Secure Control-Plane

VPN-based authentication, encryption and firewall
- Netscreen ns-50 at ORNL
- NS-5 at each node
- Centralized server at ORNL
  - bandwidth scheduling
  - signalling
Need for Secure Control Plane

- Security of control plane is extremely important
  - USN switches (Ciena, Force10, Turin, Sycamore, Whiterock) do not support IPSec – do not know of any that do
  - TL1/CLI and GMPLS commands sent in the “clear”
    - Can be sniffed to profile the network
    - Can be injected to “take over” the control
  - Following cyber attacks could be easily launched
    - Hijack the dedicated circuits; sustain a DOS flood to prevent recovery
    - Takeover/flood UltraScienceNet end hosts and switching gear
- USN control-plane is out-of-band and secure
  - Uses VPN-based control channels and firewalled enclaves
Control Plane

• Phase I
  – Centralized VPN connectivity
  – TL1/CLI-based communication with CoreDirectors and E300s
  – User access via centralized web-based scheduler

• Phase II
  – GMPLS direct enhancements and wrappers for TL1/CLI
  – Inter-domain “secured” GMPLS-based interface
  – Webservices interface for OSCARS
Web Interface

- Allows users to logon to website
- Request dedicated circuits
- Based on cgi scripts written in c and c++
Bandwidth Scheduler

- Computes path with target bandwidth
  - Is currently available?
    - Extension of Dijkstra’s algorithm using interval sequences
  - Provide all available slots
    - Extension of closed semi group structure to sequences of reals
      - Both are solvable by polynomial-time algorithms
      - Implementation – first part almost complete; needs interface

- Notes:
  - GMPLS does not have this capability
  - Control-plane engineering taskforce interested in using it.
  - Not an NP-Complete problem
Peering: UltraScience Net – NSF CHEETAH

- Peering: data and control planes
  - Coast-to-coast dedicated channels
  - Access to ORNL supercomputers

Peering at ORNL:
Data plane:
10GigE between SN16000 and e300
Control-Plane:
VPN tunnel
Current Status: Data-Plane

- Data-Plane Connections:
  - Chicago-Sunnyvale
    - May 2005: 10GigE WAN-PHY between E300
    - August 2005: 2 x OC192 links between CDCIs
  - ORNL-Chicago
    - August 2005: 2 x OX192 links between CDCIs
    - Atlanta will be connected after SC2005
- User-connections
  - August 2005
    - PNNL, FNL, CalTech, ESnet
  - November 2005
    - SLAC
  - February 2006
    - Atlanta node installation
Current Status: Control-Plane

- ORNL, Chicago, Seattle, Sunnyvale nodes are setup
  - VPN, console servers are setup
  - signaling modules – being integrated
  - Bandwidth/channel reservation system – being integrated with signaling system
- SC2005 node will be moved to Atlanta
USN at Supercomputing2005

- Extended USN to exhibit floor: eight dynamic 10Gbps long-haul connections over time
- Moved and re-created USN-Seattle node on
- PNNL, FNL, ORNL, Caltech, SLAC at various booths to support:
  - applications and bandwidth challenge

Helped Caltech team win Bandwidth Challenge:
- 40Gbps aggregate Bw
- 164 terabytes transported in day
ESnet Related Issues

- Port Assignments:
  - 10GigE port each on E300 in Sunnyvale and Chicago
  - multiple 1GigE ports assigned on E300 in Sunnyvale and Chicago

- Cross-connects
  - 1 SM and 4 MM cross-connects in Level(3) POP in Sunnyvale and in Starlight in Chicago

- Control-Plane Issues are being addressed
OSCARS and USN Control-Plane Integration I

- Composition Front End: User
  - Back-end interaction with OSCARS and USNCP
  - Website and webservice: authentication + encryption
  - User request:
    - Scheduling
      - decomposed into OSCARS and USNCP requests
      - combine the responses and compose the path
    - Signaling
      - Pieces of paths are signaled separately

Drawbacks: VLANs need to be supported separately
OSCARS and USN Control-Plane Integration II

- VLAN transitioning
  - Scheduler explicitly allows for VLAN setup requests
    - Front end sends separate requests and handles boundaries
  - Signaling
    - Wrappers to OSCARS and USNCP to accept VLAN signaling
    - Uniform wrapper formats needed – WDSL+SOAP (?)
    - Authentication and Encryption
MPLS-GMPLS Integration

- Advanced Reservation:
  - Open issue within MPLS and GMPLS
  - Reservation front-end:
    - Scheduling a priori
    - Send MPLS-GMPLS messages for immediate setup/tear down

- Signaling
  - GMPLS wrapper for USNCP

- Scheduling Extensions of GMPLS and MPLS
  - Need to work with standards
Conclusions

USN Deployment
Data-Plane – Complete
Control-Plane – almost Complete

Request for USN Collaborations
USN channels/circuits
USN hosts – transport, middleware
Locate your hardware at USN nodes
Thank you
https://www.usn.ornl.gov