DOE UltraScienceNet

Experimental Network Testbed for High-Performance Network technologies and Applications

Nagi Rao, Bill Wing, Steven Carter, Qishi Wu
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

OAK RIDGE NATIONAL LABORATORY
U. S. DEPARTMENT OF ENERGY

UT-BATTELLE

Contents

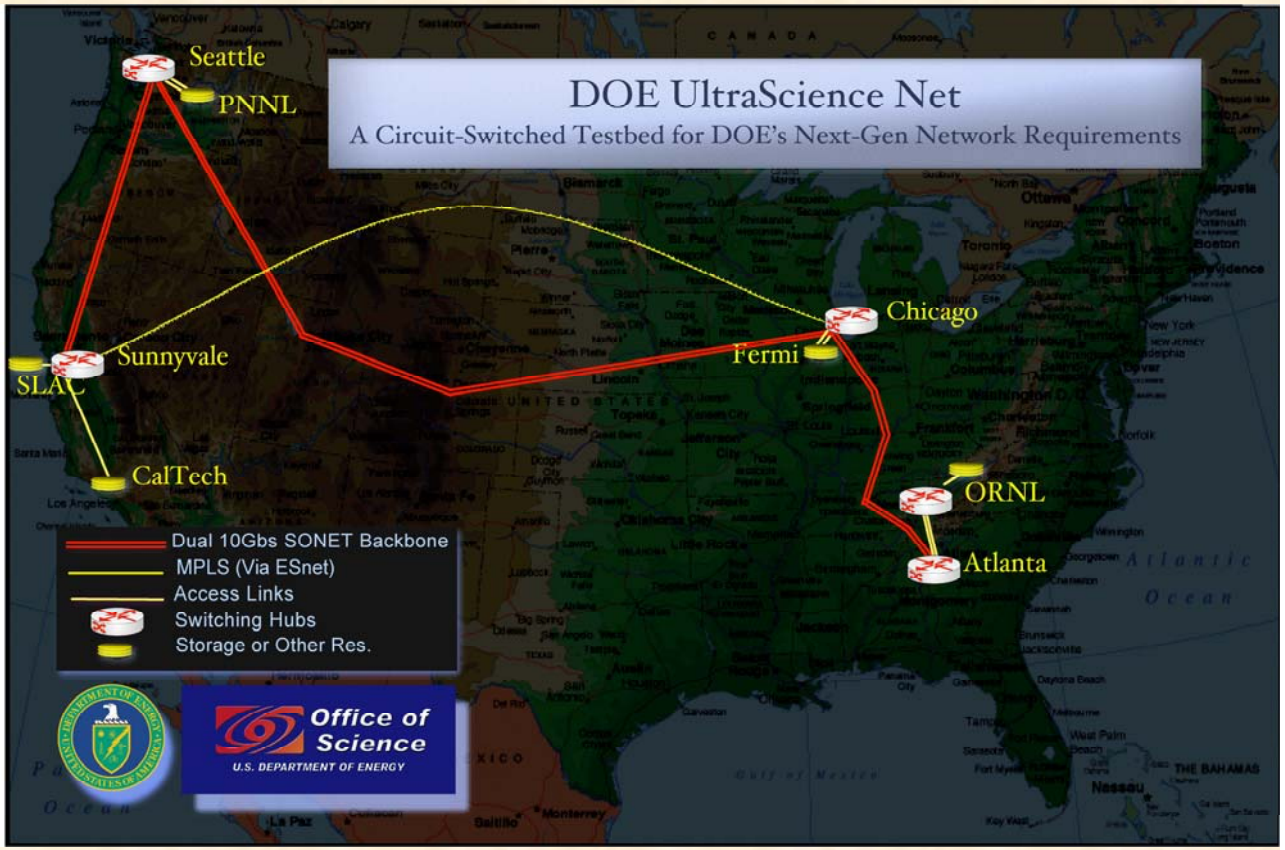
DOE UltraScience Net – In a Nutshell

Experimental Network Research Testbed:

To support advanced networking and related application technologies for DOE large-scale science projects

Features

- End-to-end guaranteed bandwidth channels
- Dynamic, in-advance, reservation and provisioning of fractional/full lambdas
- Secure control-plane for signaling
- Proximity to DOE sites: NLCF, FNL,NERSC
- Peering with ESnet, NSF CHEETAH and other networks



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**

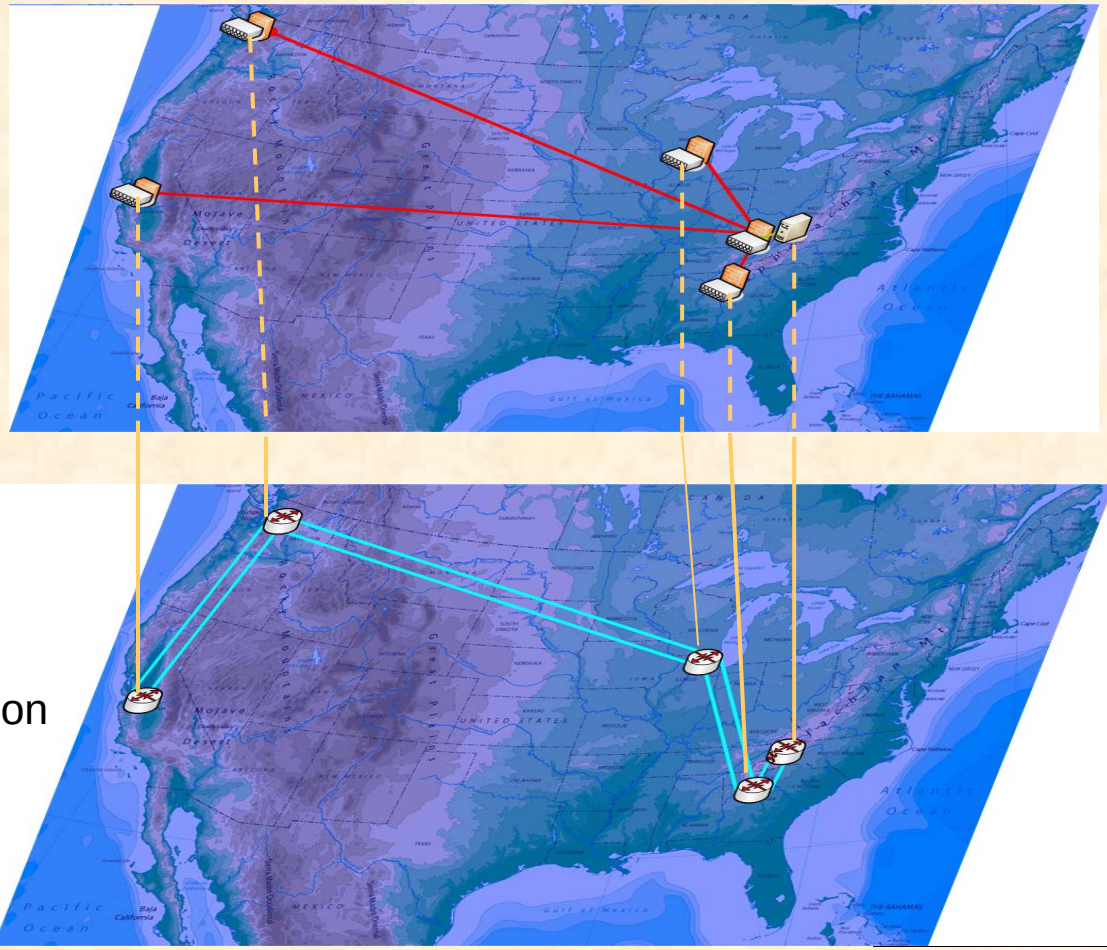
DOE-Funded Support Application Projects

- **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



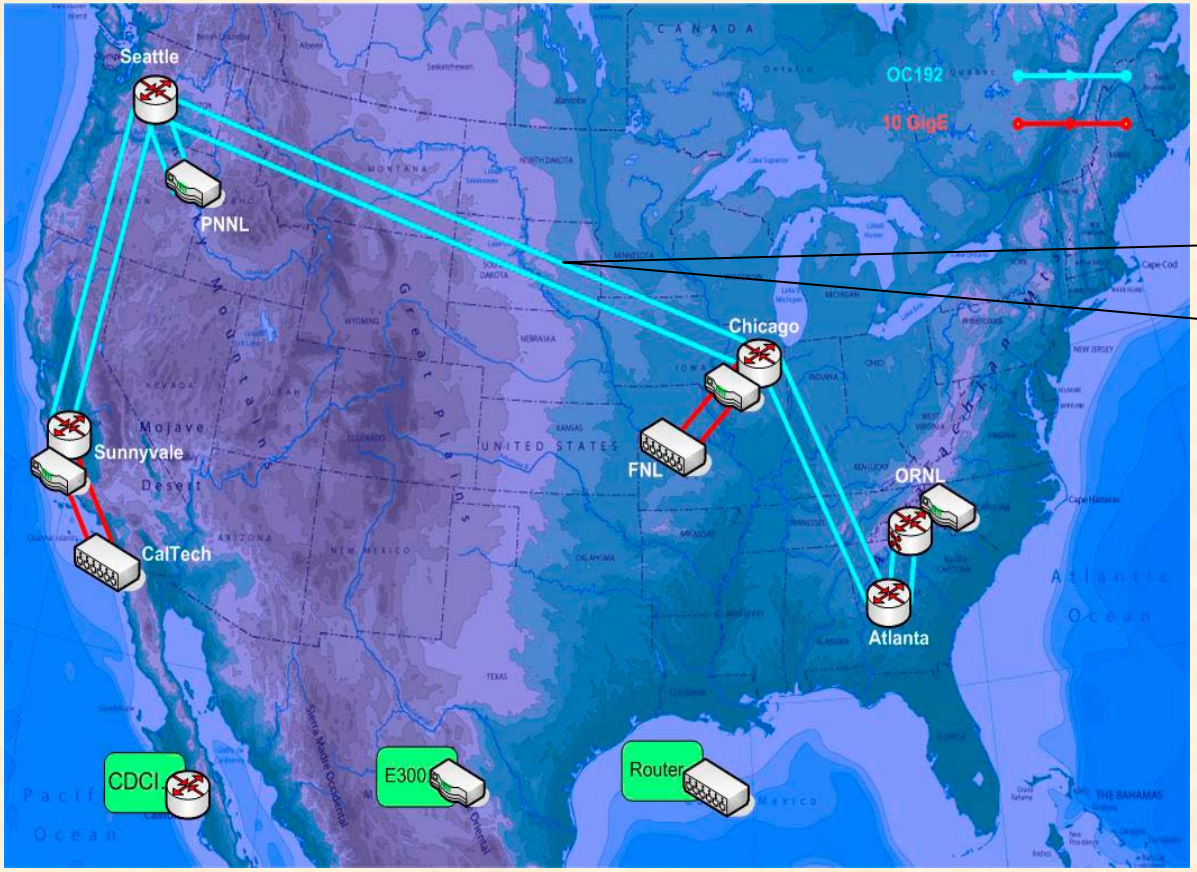
Dual OC192 backbone:

- SONET-switched in the backbone
- Ethernet-SONET conversion

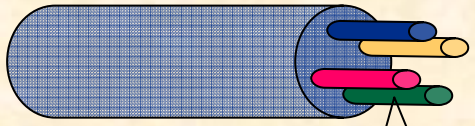
DOE UltraScience Net: Data Plane

Connects Atlanta, Chicago, Seattle and Sunnyvale:

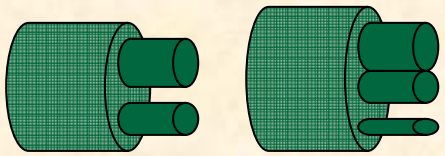
- **Dynamic and in-advance provisioned dedicated dual 10Gbps links at 50 Mbps resolution – SONET or Ethernet**



2 current lambdas
2 future lambdas



provisioned at multiple 50 Mbps resolution



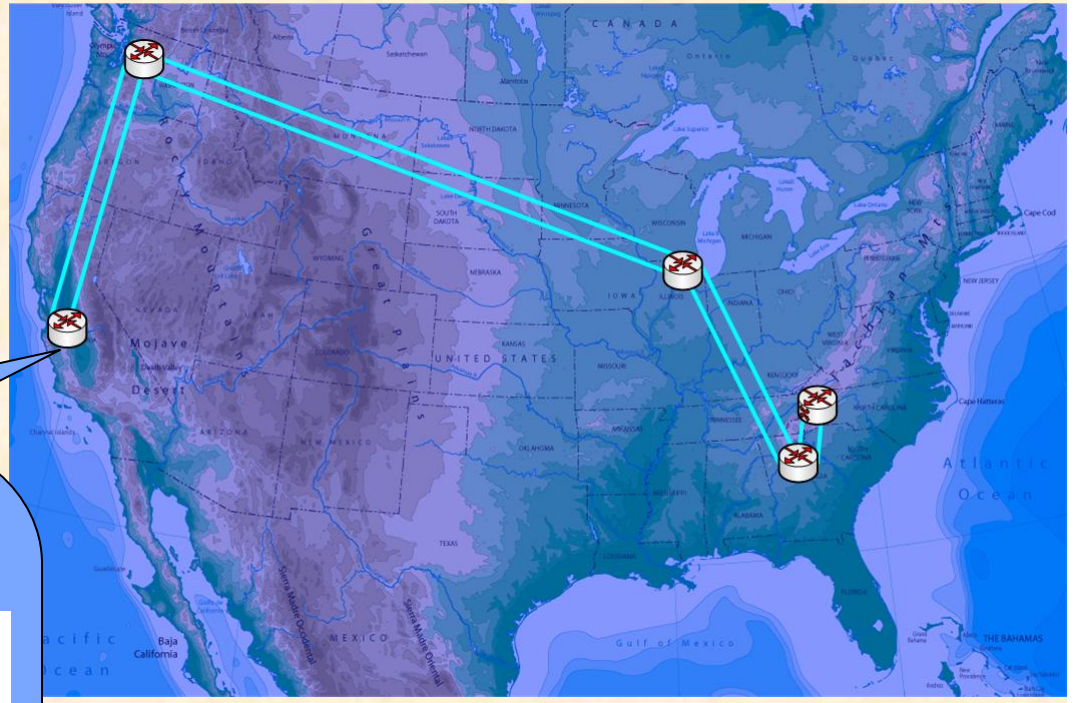
USN Data-Plane: Node Configuration

In the Core:

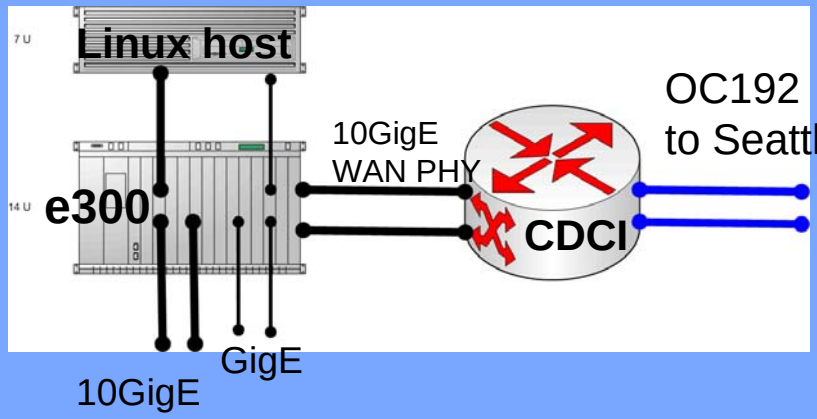
- **Two OC192 switched by Ciena CDCIs**

At the Edge

- **10/1 GigE provisioning using Force10 E300s**



Node Configuration



Connections to CalTech and ESnet

Data Plane User Connections:

Direct connections to:

core switches –SONET &1GigE

MSPP – Ethernet channels

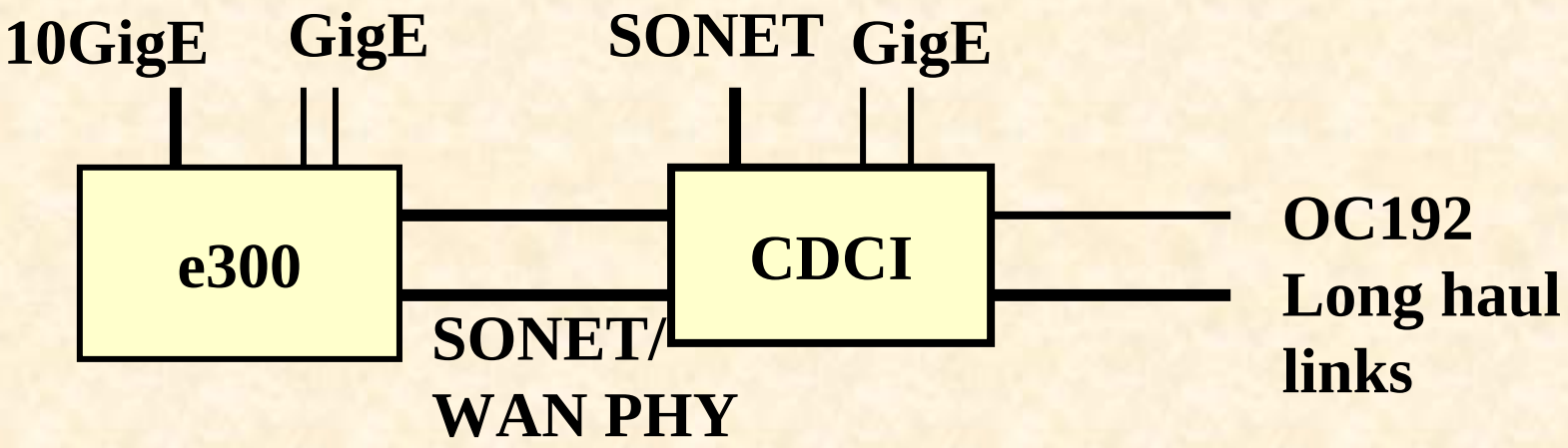
Utilize UltraScience Net hosts

UT-BATTELLE

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**

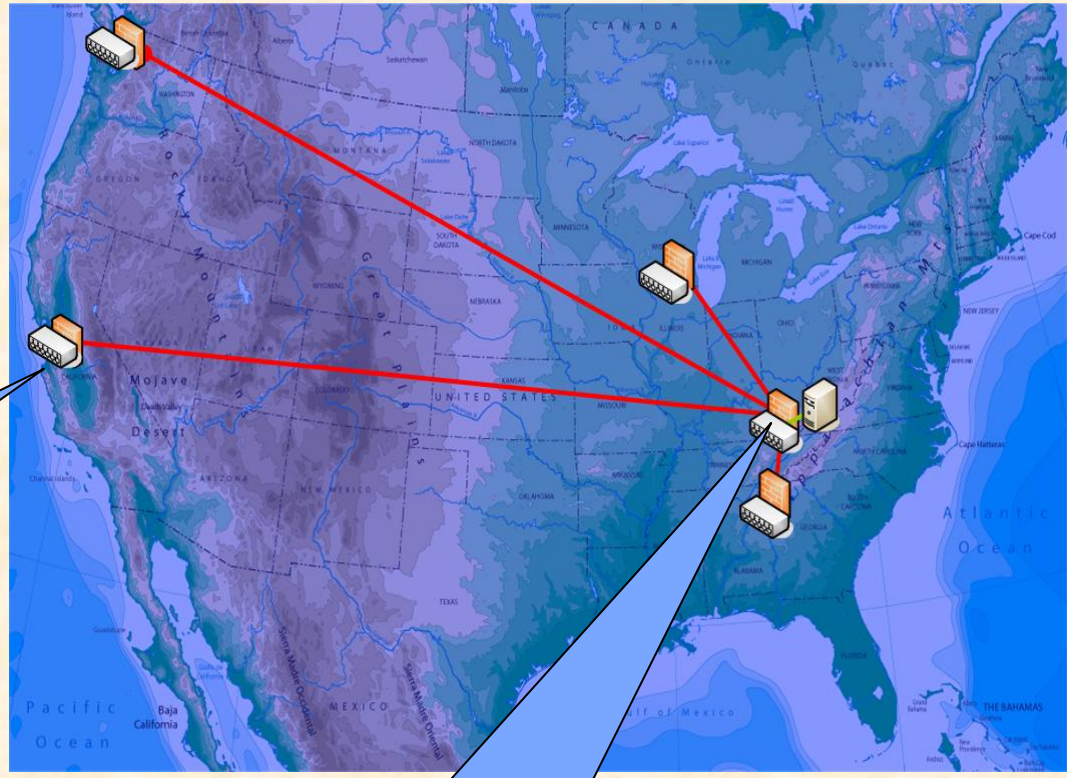
GigE ports must match at the connection end points

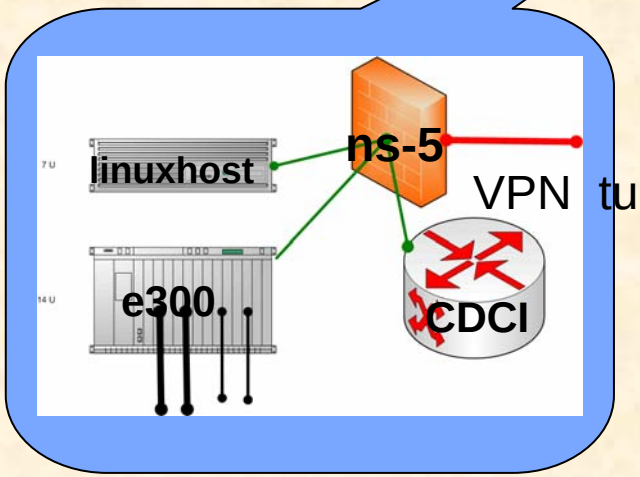


Secure Control-Plane

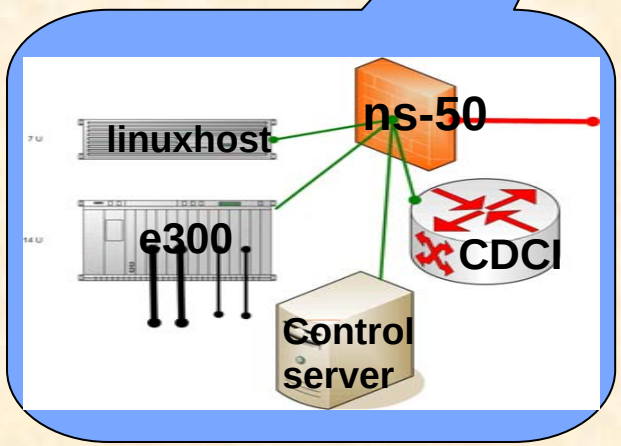
VPN-based authentication, encryption and firewall

- Netscreen ns-50 at ORNL
  NS-5 at each node
- Centralized server at ORNL
  - bandwidth scheduling
  - singnalling





VPN tunnel



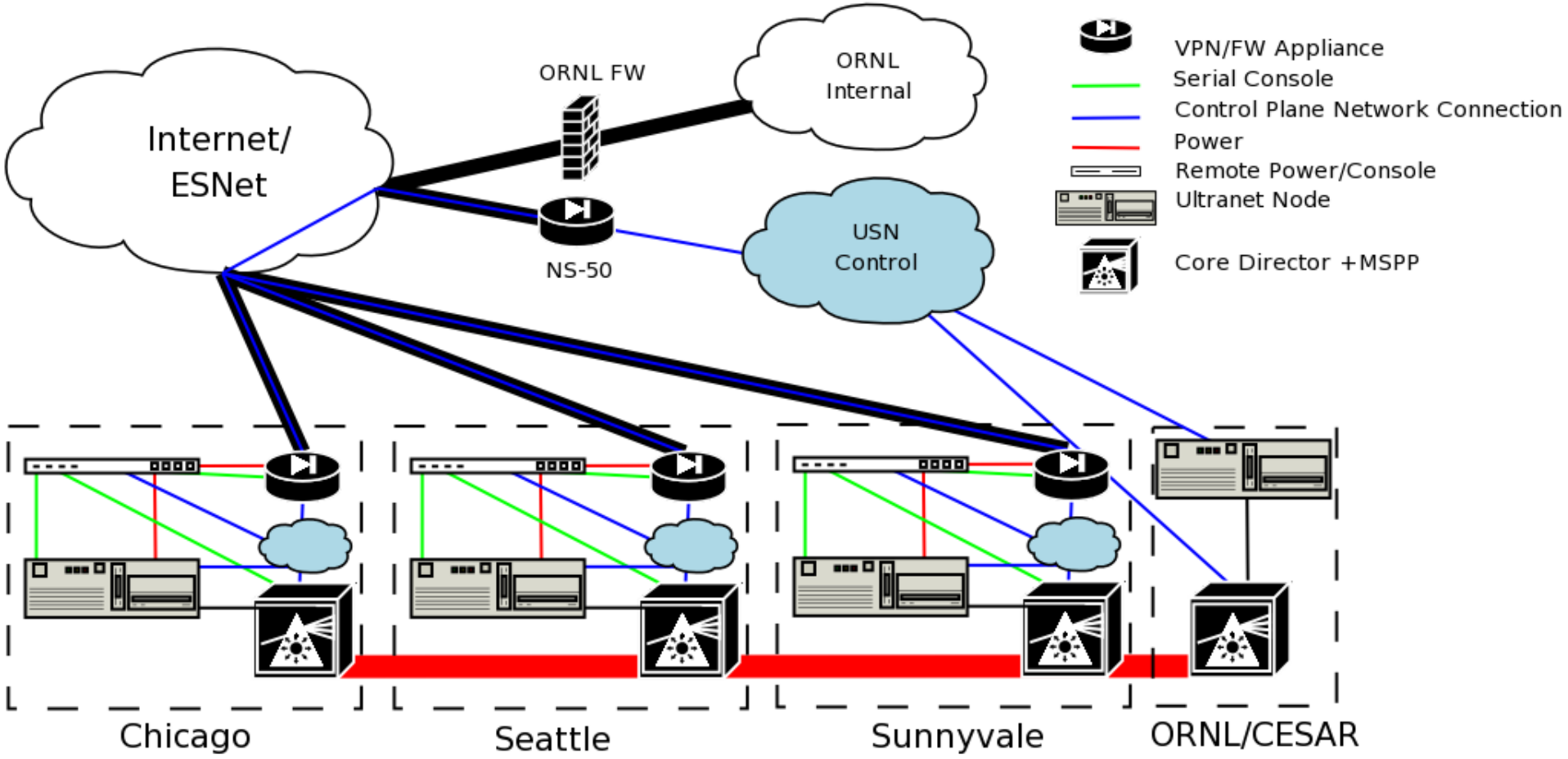
Need for Secure Control Plane

- **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

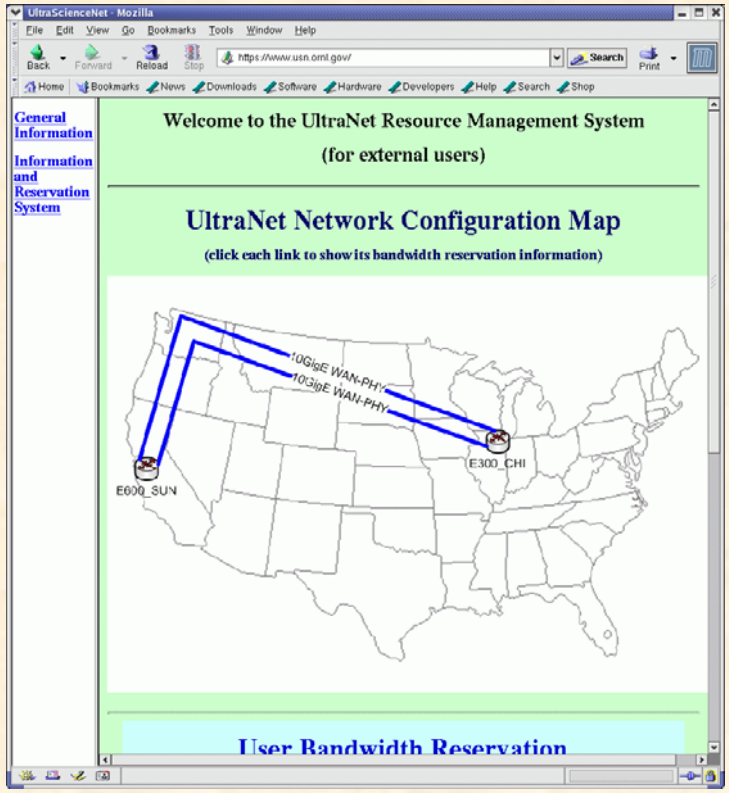
- **Phase I**
  - **Centralized VPN connectivity**
  - **TL1/CLI-based communication with CoreDirectors and E300s**
  - **User access via centralized web-based scheduler**
- **Phase II**
  - **GMPLS direct enhancements and wrappers for TL1/CLI**
  - **Inter-domain “secured” GMPLS-based interface**

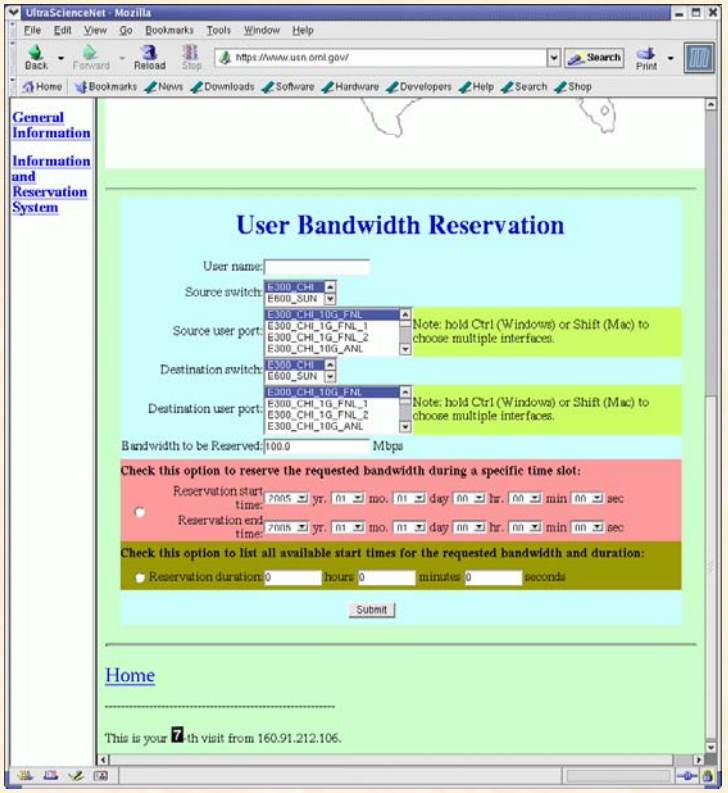
Complete Control Plane and Management Plane



Web Interface

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



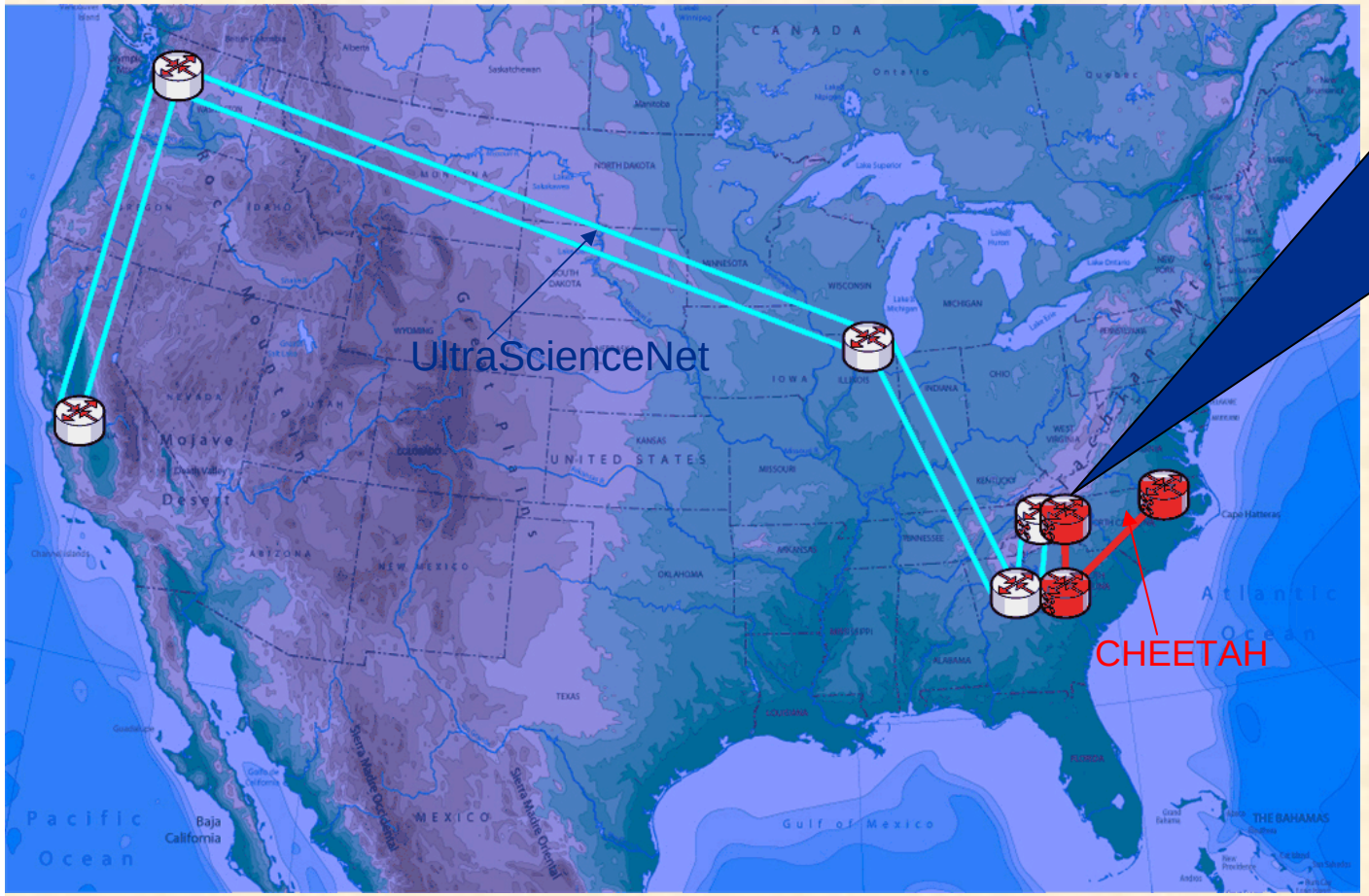


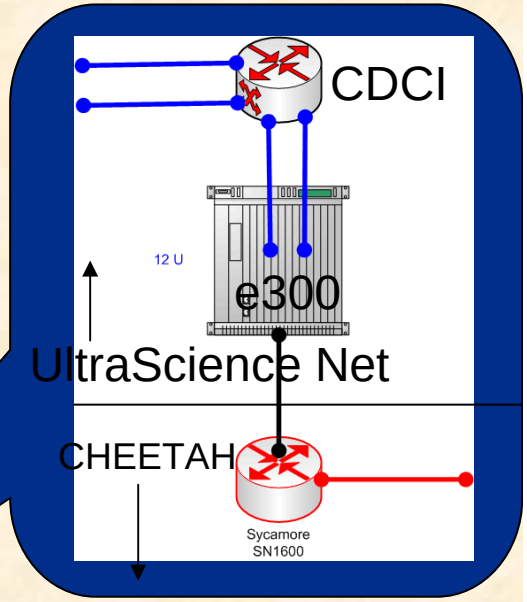
Bandwidth Scheduler

- **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

UT-BATTELLE

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**
  - **May 2005**
    - **FNL and CalTech**
  - **August 2005**
    - **PNNL, ESnet**

Current Status: Control-Plane

- **ORNL node is setup**
  - **VPN, console servers are setup**
  - **signaling modules – being tested ~ 1 month**
  - **Bandwidth/channel reservation system ~ 1 month**
- **Chicago, Sunnyvale, Seattle nodes are setup**
- **SC2005 node will be moved to Atlanta**

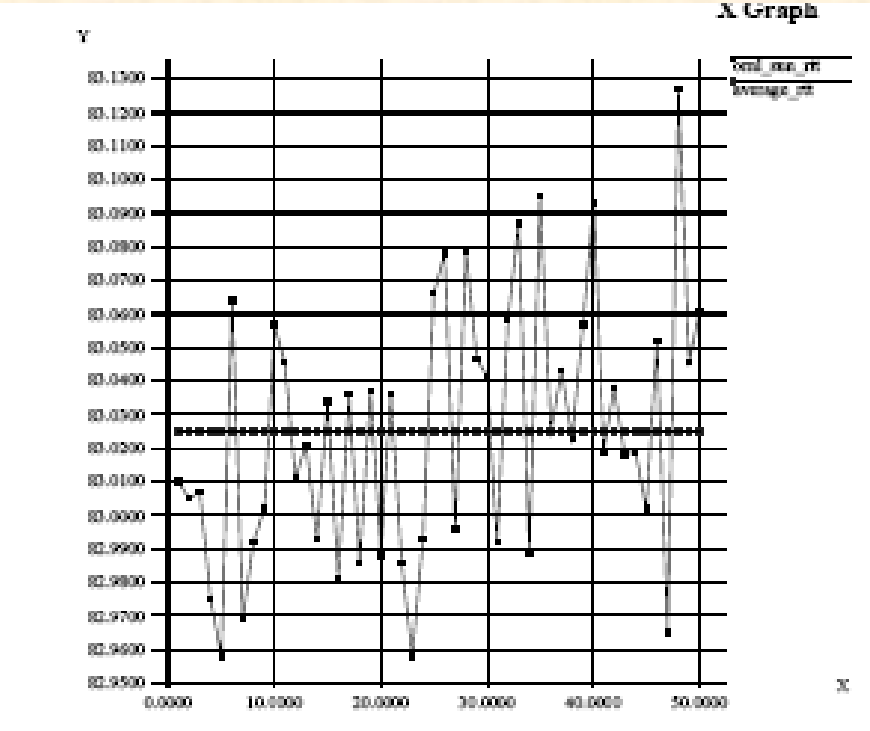
ESnet Related Issues

- **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 ordered in Level(3) POP in Sunnyvale and in Starlight in Chicago**
- **Control-Plane Issues are being addressed**

Some Experimental Results

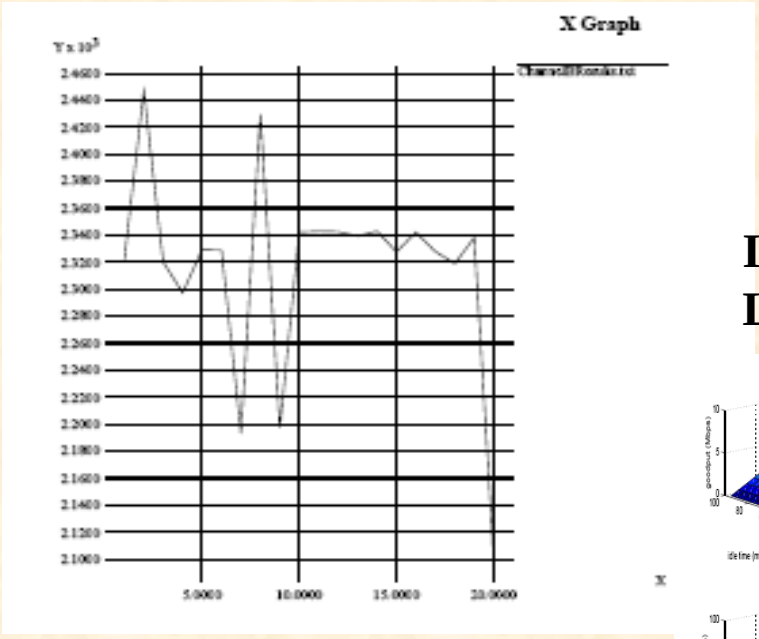
- Layer-2 double-loopback test:
  - Entire USN SONET backbone connected in 16000 mile single connection
  - 16 hours continuous zero errors

- **Jitter measurements**
  - **ORNL-SUNNYVALE host-to-host 1K packets**
  - **round-trip time:**
    - mean: 82ms
    - jitter: 0.2%

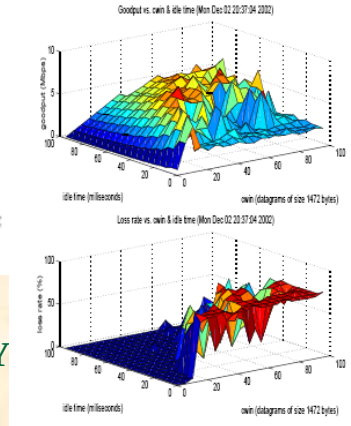


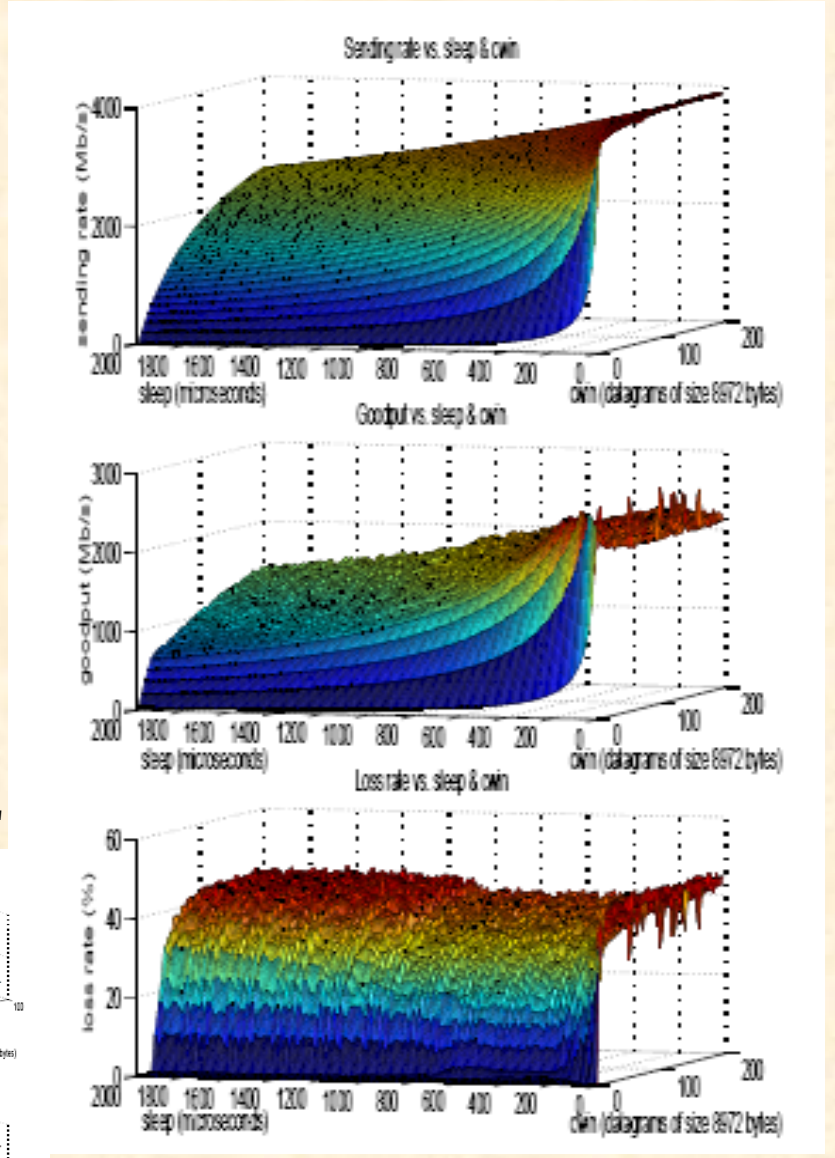
Throughput profile

- **Transport measurement**
  - **ORNL-SUN host-to-host file transfers 4000mile, 10G connection**
  - **Limited by host - Hurricane**
  - **Average throughput 2.3Gbps**
  - **Loss rate < 0.1%**



Internet
LSU-ORNL





Publications

1. N. S. V. Rao, W. R. Wing, S. M. Carter, Q. Wu, UltraScience Net: Network testbed for large-scale science applications, IEEE Communications Magazine, 2005, in press.
2. Q. Wu, N. S. V. Rao, Protocols for high-speed data transport over dedicated channels, Third International Workshop on Protocols for Fast Long-Distance Networks, 2005.
3. Q. Wu, N. S. V. Rao, A class of reliable UDP-based transport protocols based on stochastic approximation, Proceedings of IEEE INFOCOM, 2005.
4. N. S. V. Rao, Q. Wu, S. M. Carter, W. R. Wing, Experimental results on data transfers over dedicated channel, First International Workshop on Provisioning and Transport for Hybrid Networks: PATHNETS, 2004.

Related Dynamics Project

- M. Yang, J.-F. Ru, H. Chen, A. Bashi, X. R. Li and N. S. V. Rao, Predicting Internet end-to-end delay: A statistical study, to appear in Annual Review of Communications, Vol. 58, 2005.
- N. S. V. Rao, J. Gao, L. O. Chua, On dynamics of transport protocols in wide-area Internet connections, in Complex Dynamics in Communication Networks, L. Kocarev and G. Vattay (editors), 2005.
- J. Gao, N. S. V. Rao, J. Hu, J. Ai, Quasi-periodic route to chaos in the dynamics of Internet transport protocols, Physical Review Letters, 2005.
- J. Gao, N. S. V. Rao, TCP AIMD dynamics over Internet connections, IEEE Communications Letters, vol. 9, no. 1, 2005, pp. 4-6.
- Y. Guo, Z. Qu, N. S. V. Rao, A new TCP end-to-end congestion avoidance algorithm through output feedback, Asian Control Conference, 2004.
- M. Yang, X. R. Li, H. Chen, and N. S. V. Rao. Predicting Internet end-to-end delay: An overview, Proc. of 36th Southeastern Symposium on Systems Theory, 2004.
- Q. Wu, N. S. V. Rao, S. S. Iyengar, On transport daemons for small collaborative over wide-area networks, Proceedings of International Performance Computing and Communications Conference, 2005.

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