DOE UltraScienceNet - Update

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

Brief Review of USN

Next Phase Activities

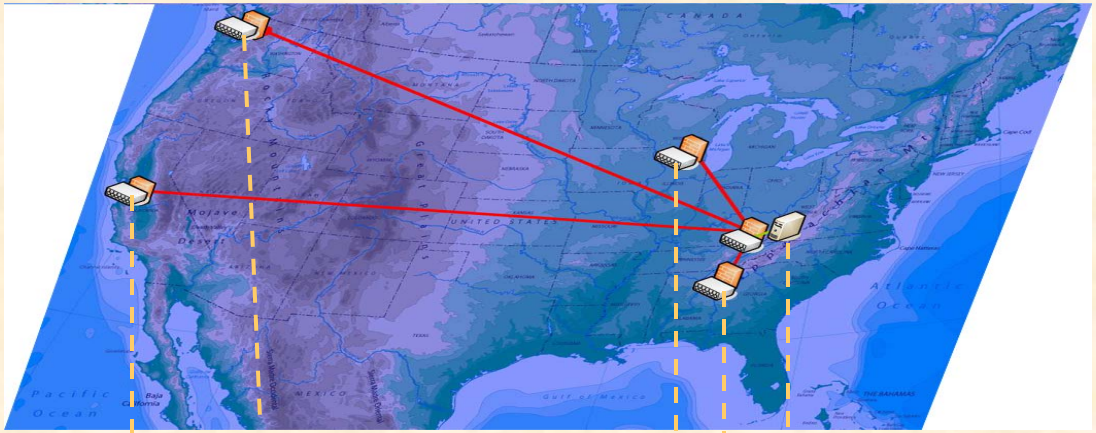
VLANS over L2 and L3 Networks

- Baseline measurements
- SONET vs MPLS VLANs
- Composite SONET-MPLS VLANS
- Summary

USN Architecture:
Separate Data-Plane and Control-Planes

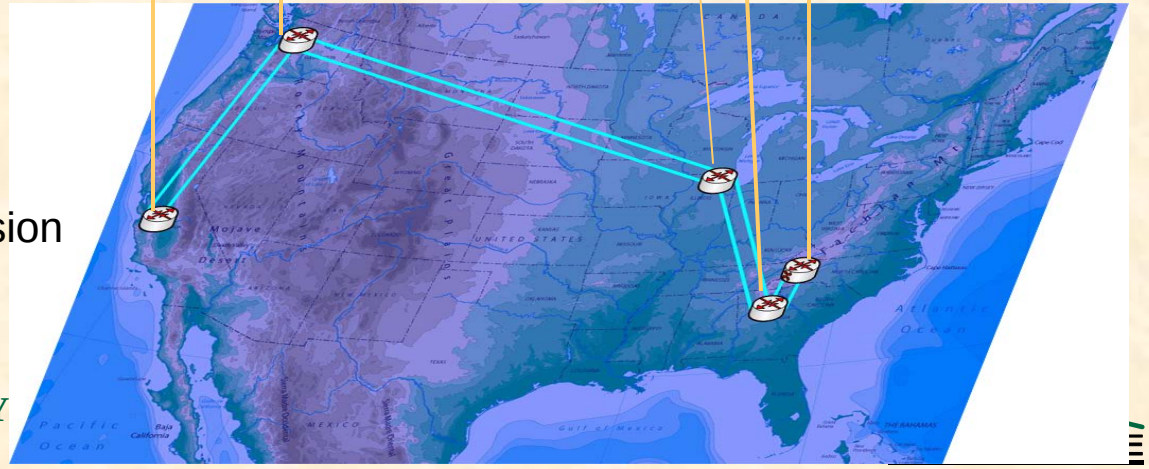
Secure control-plane with:

- Encryption, authentication and authorization
- On-demand and advanced provisioning



Dual OC192 backbone:

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



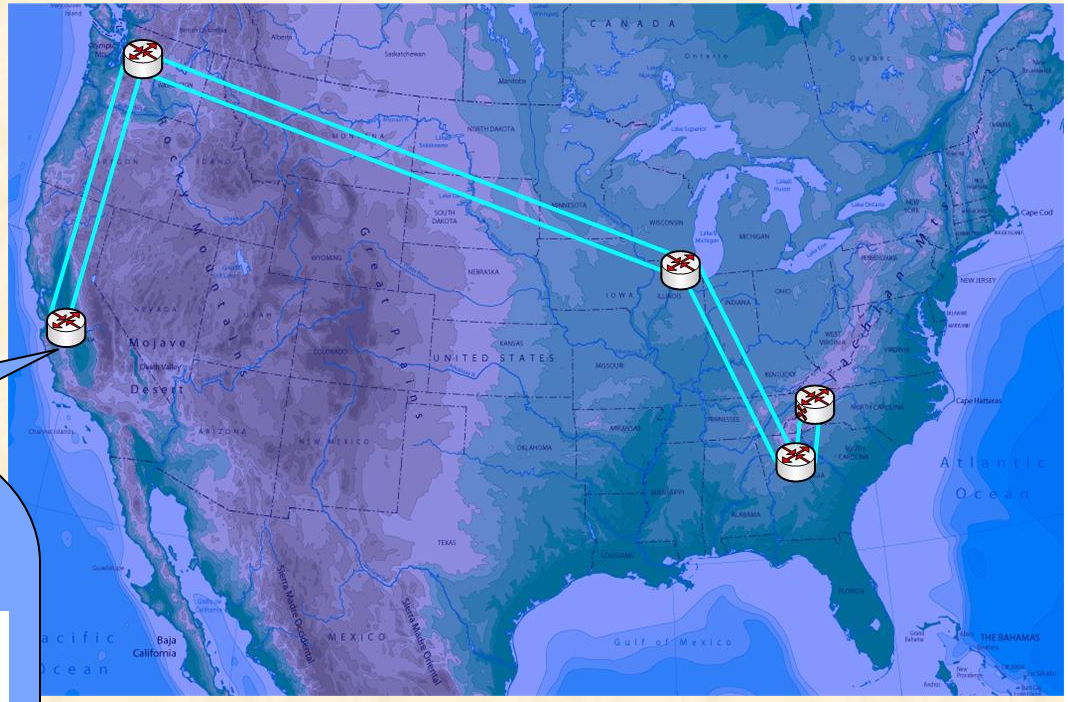
USN Data-Plane: Node Configuration

In the Core:

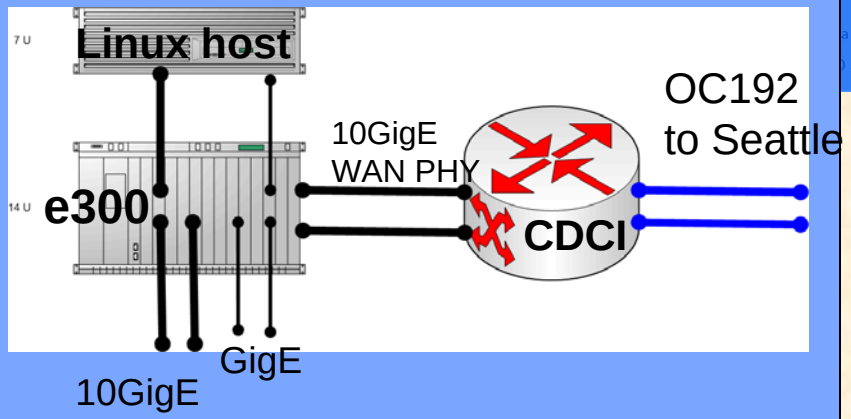
- **Two OC192 switched by Ciena CDCIs**

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

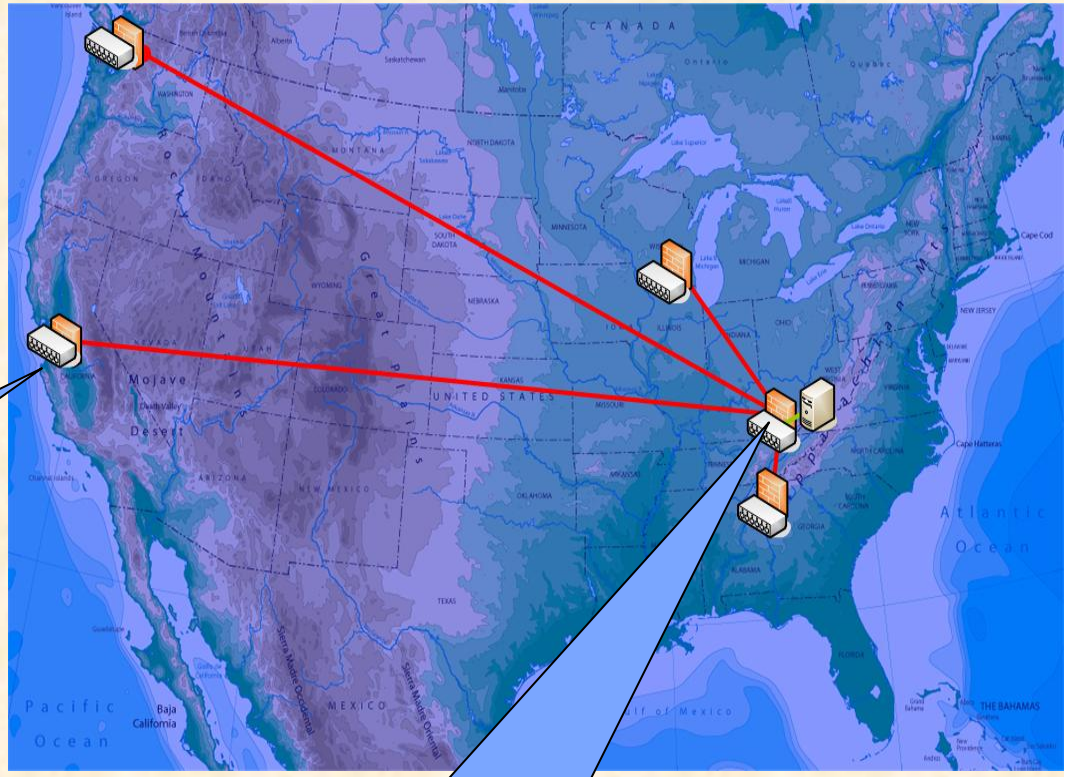
Utilize UltraScience Net hosts

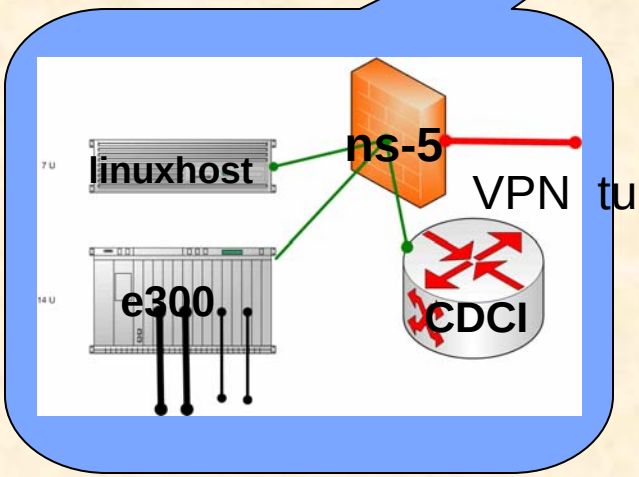
UT-BATTELLE

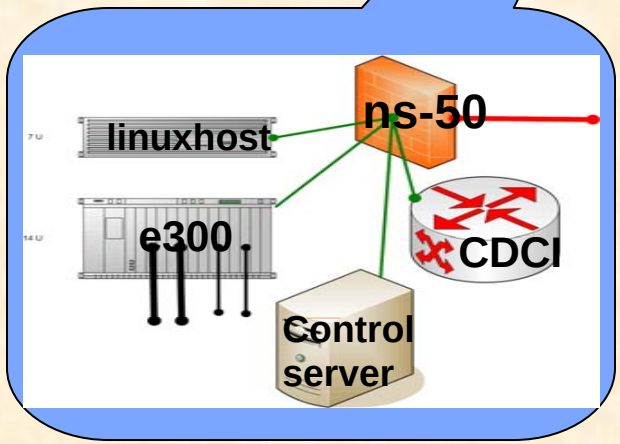
Secure Control-Plane

VPN-based authentication, encryption and firewall

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



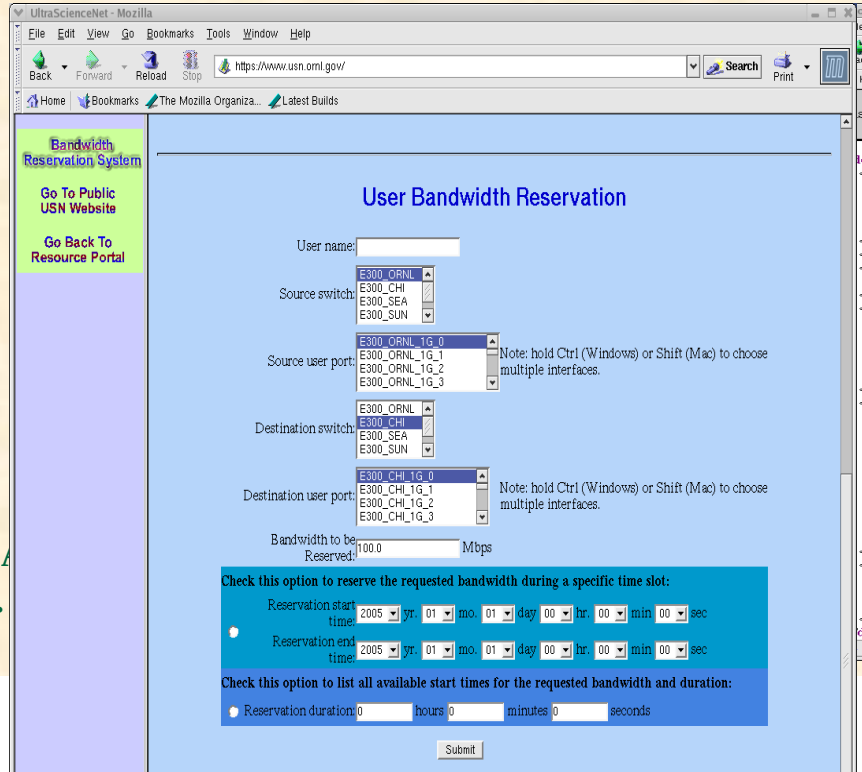


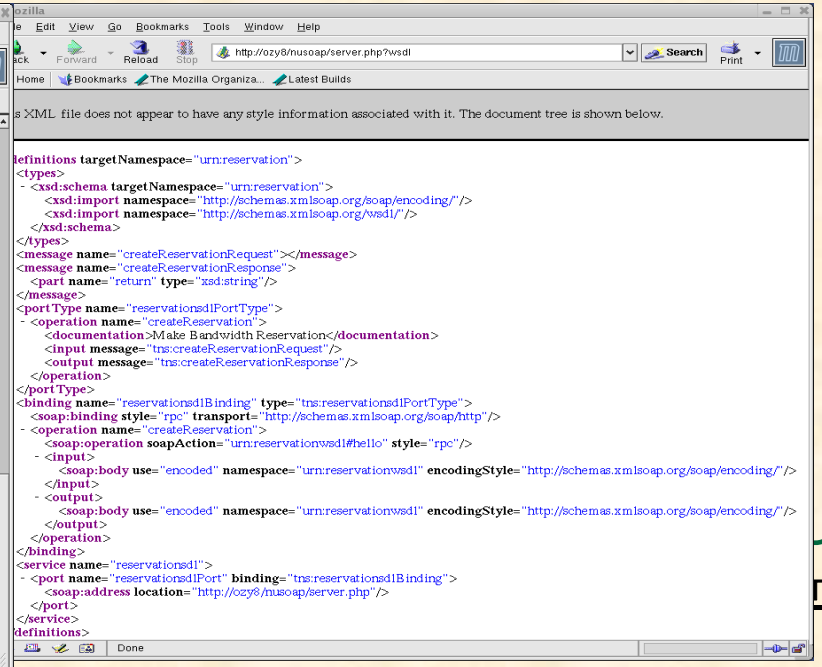


UT-BATTELLE

USN Control Plane

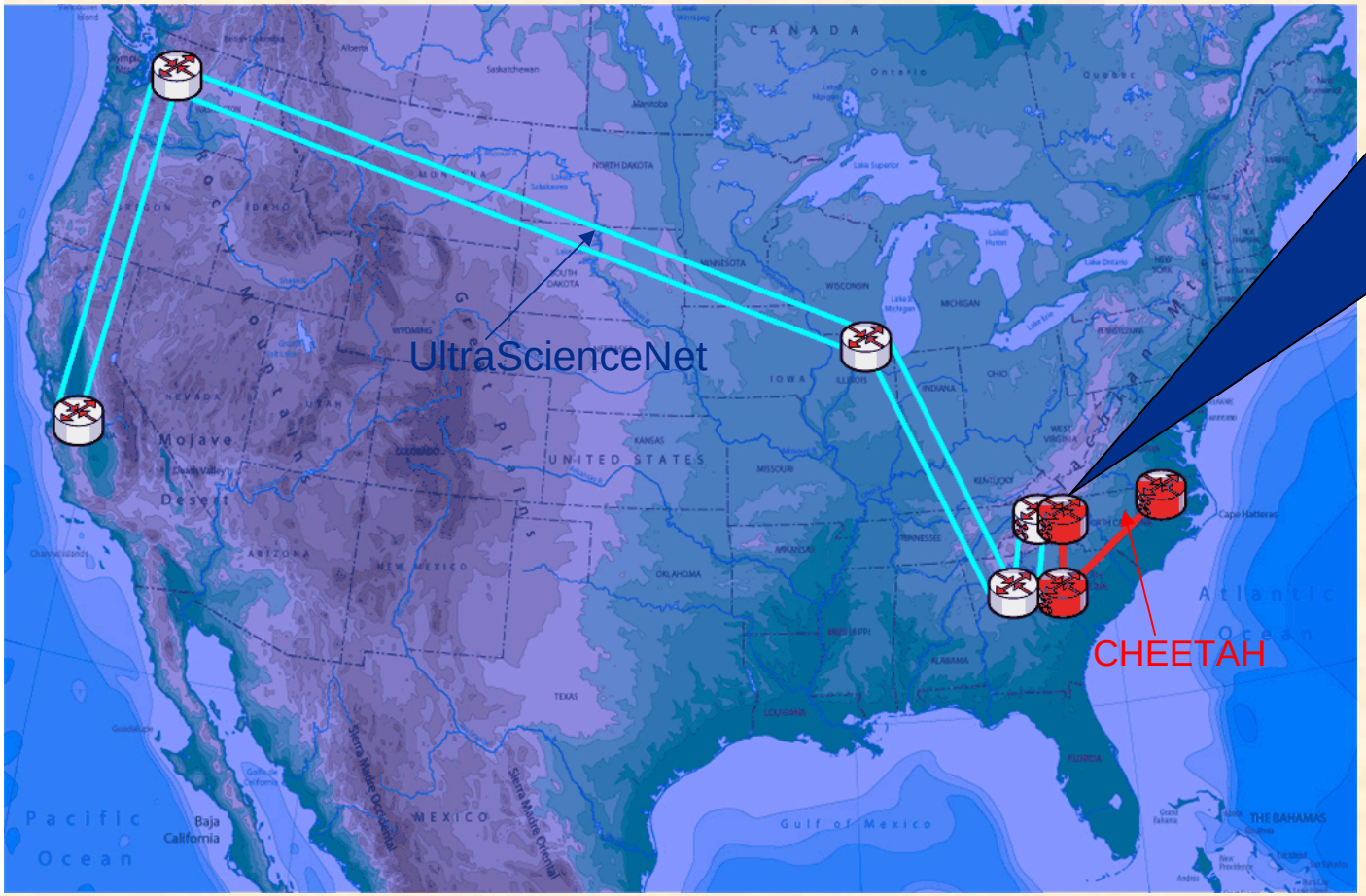
- **Phase I**
  - **Centralized path computation for bandwidth optimization**
  - **TL1/CLI-based communication with CoreDirectors and E300s**
  - **User access via centralized web-based scheduler**
- **Phase II (current)**
  - **GMPLS wrappers for TL1/CLI**
  - **Inter-domain "secured" GMPLS-based interface**
  - **Webservices interface**
  - **X509 authentication for web server and service**

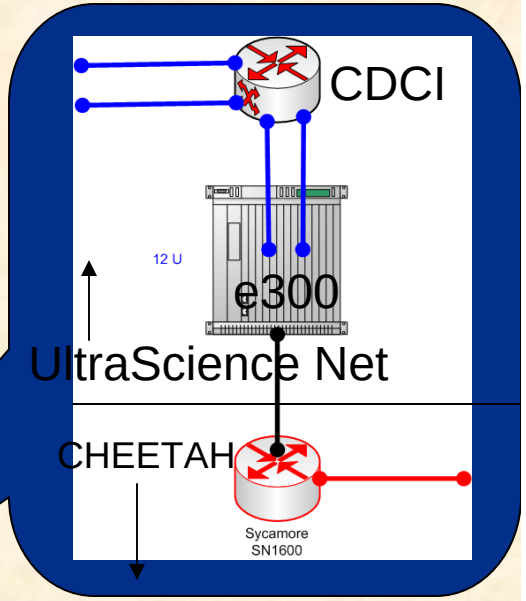




Peering: UltraScience Net – NSF CHEETAH

- **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**
- **User-connections**
  - **August 2005**
    - **PNNL, FNL, CalTech, ESnet**
  - **November 2005**
    - **SLAC**
  - **May 2006**
    - **Atlanta node installation**

Current Status: Control-Plane

- **ORNL, Chicago, Seattle, Sunnyvale, Atlanta nodes are setup**
  - **VPN, console servers are setup**
  - **Signaling modules – GMPLS extensions are being implemented**
  - **Bandwidth/channel reservation system – being tested with X.509 authentication**
  - **Webservice is being tested**
- **SC2005 node moved to Atlanta**

USN Next Steps

- Infrastructure development is close to completion
- USN will continue to play research role in advanced networking capabilities:
  - Networking Technologies for Leadership Class Facilities
    - covered last ESCC meeting results on Cray X1(E)
  - Integrated Multi-Domain Interoperation System for USN-ESnet-CHEETAH-HOPI
    - on-going efforts with OSCARS and HOPI: plans presented at JointTechs 2006 summer workshop
  - Hybrid Optical Packet and Switching Technologies
    - VLAN testing and analysis over L1-2 and MPLS connections: this presentation
    - Configuration and testing of hybrid connections

USN Next Step: Leadership Class Facilities

- Interconnects with dedicated wide-area paths.
- Alignment and impedance matching techniques for both data and execution paths.
- Testing high-performance wide-area storage and file systems over layer-2 and layer-1 standard and non-standard (infiniband) wide area transport.

USN Next Step: Multi-Domain Networking

Integrated control-plane technologies for dedicated channels spanning across USN into other networks:

1. Meta-schedulers to provide on-demand and in-advance paths over multiple domains.
2. Development, alignment and integration of provisioning and signaling technologies including MPLS of OSCARS and GMPLS of HOPI and CHEETAH.

USN Next Steps: Hybrid Optical Packet and Switching Technologies

- VLAN testing and analysis over L1-2 and MPLS connections
- Technologies for networks where shared IP connections co-exist with dedicated channels provisioned at layers 1-3.
- Technologies for provisioning dedicated channels over hybrid networks for
  - (i) point-to-point connections; and
  - (ii) forming dynamic back-plane networks of a pool of such connections between distributed NLCF systems.

Contents

Brief Review of USN

Next Phase Activities

VLANS over L2 and L3 Networks

- Baseline measurements
- SONET vs MPLS VLANs
- Composite SONET-MPLS VLANS
- Summary

Background

Relative performance of VLANs provisioned over:

- SONET layer-1 paths
- MPLS layer-3 paths

Broader Networking Question:
Layer-1 or layer-2 or layer-3 channels for dedicated bandwidth connections?

Feasibility and Performance of Composed SONET-MPLS VLANS:

Data-plane unification of dedicated paths over layer-1, layer-2 and layer-3 paths

Broader Question: Peering data-paths across networks that provide VLANs over Layer-1 or layer-2 or layer-3

Systematic analysis of application and IP level measurements:
Using USN and CHEETAH, we

- collected ping and TCP measurements
- performed comparative performance analysis
- composed and tested VLANS over SONET and IP connections

Connections and Measurements

Two types of measurements:

- Ping (standard version)
- TCP client/server: roundtrip time of 1K packets

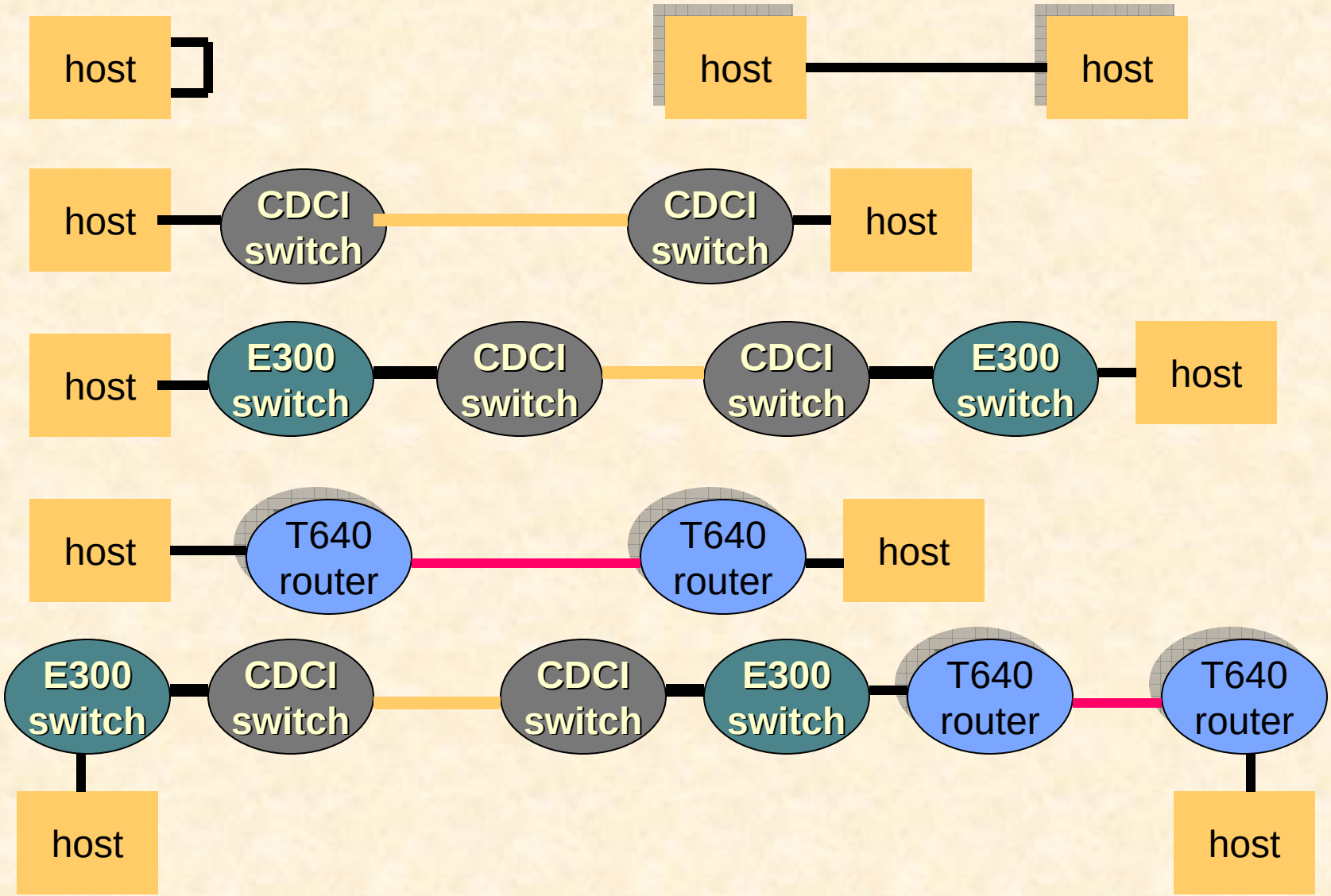
Types channels:

- host-local NIC
- host-host
- host-switch/router-host – LAN connection
- host-SONET-host – 1000 miles; OC-21
- host-MPLS-host – 300 miles; 1Gig VLAN
- host-SONET-MPLS-host; 1300 miles 1Gig VLAN

Comparison Method:

- normalization of wide-area measurements based on mean delay

Test-Configurations



Contents

Brief Review of USN

Next Phase Activities

VLANS over L2 and L3 Networks

- Baseline measurements
  - Host measurements
  - LAN first-hop measurements
- SONET vs MPLS VLANs
- SONET-MPLS VLANS
- Summary

Baseline Measurements – linux hosts

Linux host

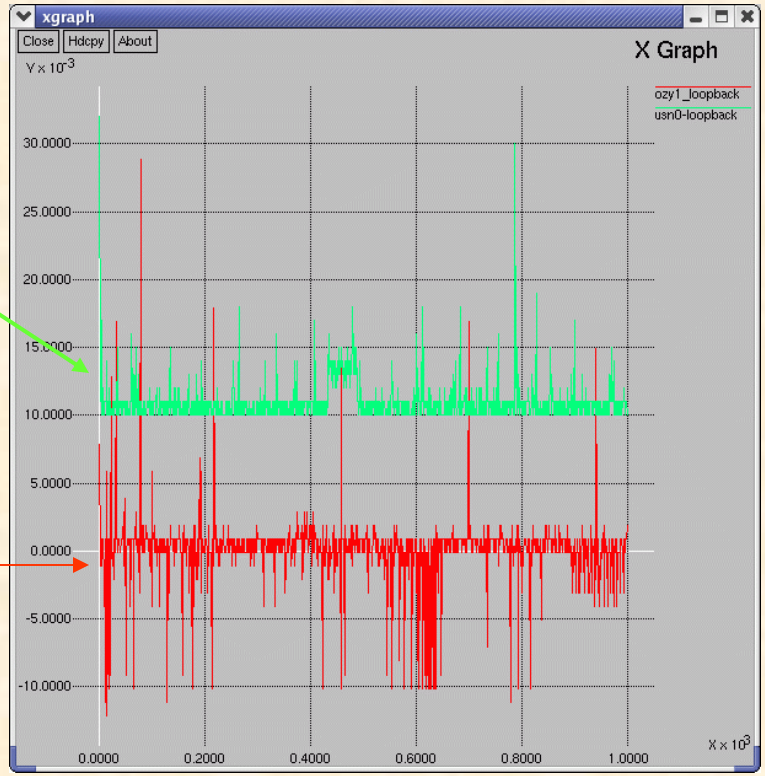
ozy1-NIC: 1GHz Pentium III

Number of ping measurements=1000
mean time=0.043117ms
percent range: [71.897528,166.987803]
:95.090275
std_deviation (percent)= 6.341750

Linux host

usn0-NIC: Xeon 3.2 GHz

Number of measurements=1000
mean=0.010960
percent range: [91.240939,291.971025]
: 200.730086
std_deviation (percent)= 14.330526

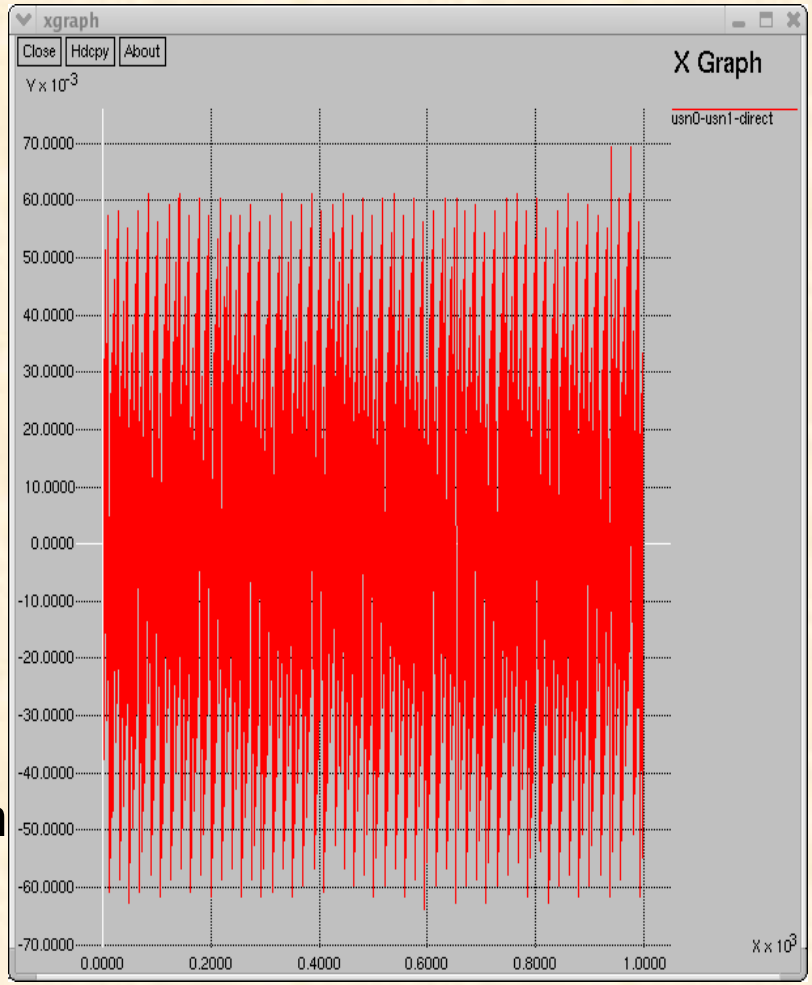


Hosts Connected Back to back

usn0-usn1 hosts: Xeon 3.2GHz

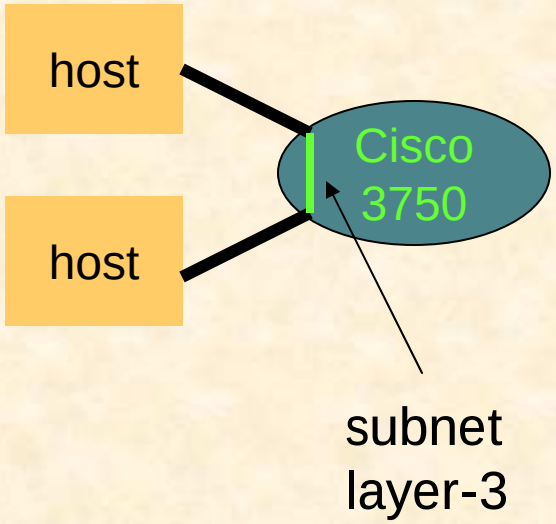
Back to back NIC connection

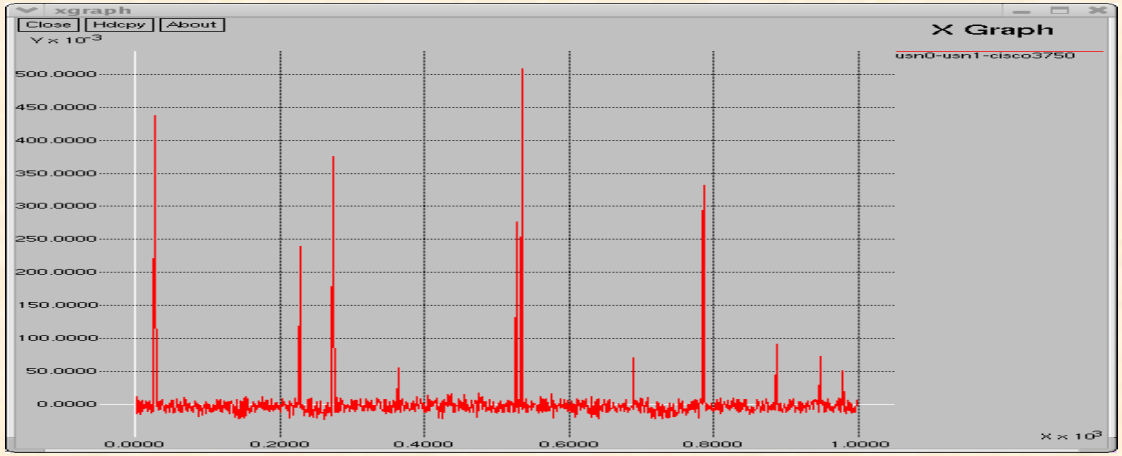
Number of ping measurements=1000
mean time =0.122728ms
percent range: [48.073878,156.443807]
: 108.369929
std_deviation (percent)= 29.437985



These dynamics need a closer inspection for possible chaotic behavior

Hosts with first hop router: Subnet Layer-3 connection





usn0-usn1-cisco3750

Number of measurements=1000

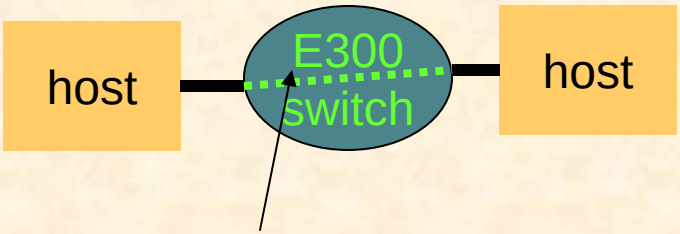
mean time=0.491746

percent range: [95.374373,203.356866]

: 107.982494

std_deviation(percent)= 6.949724

Hosts with first hop switch: Layer-2 cross-connect on Force10



Layer-2 cross-connect Via VLAN

usn0-usn1-e300:

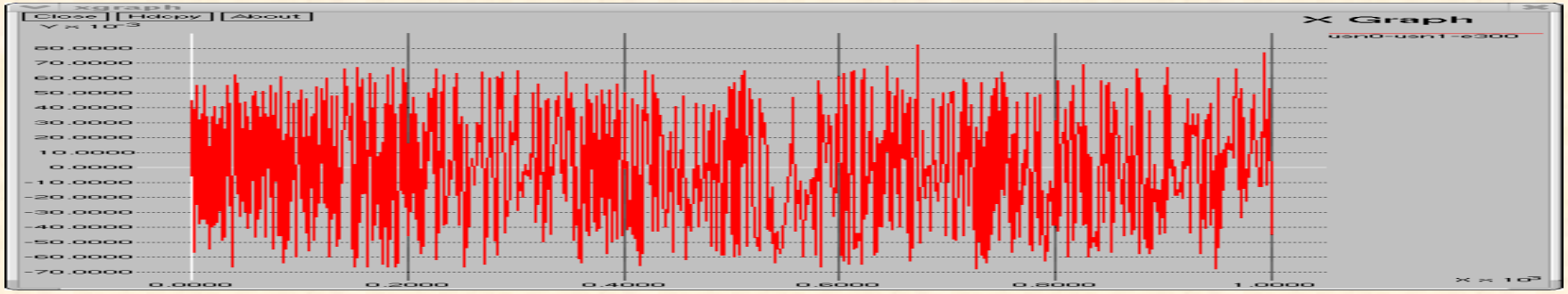
Number of **PING** measurements=1000

mean=0.211787ms

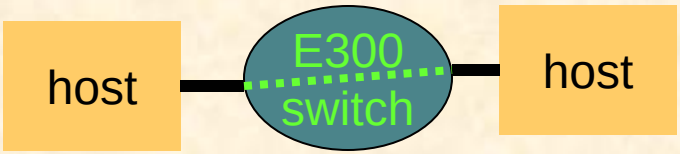
percent range: [67.992896,138.818836]

: 70.825939

std_deviation (percent)= 17.278263



Hosts with first hop switch: TCP measurements: 1K packets roundtrip times



usn1-usn0-e300

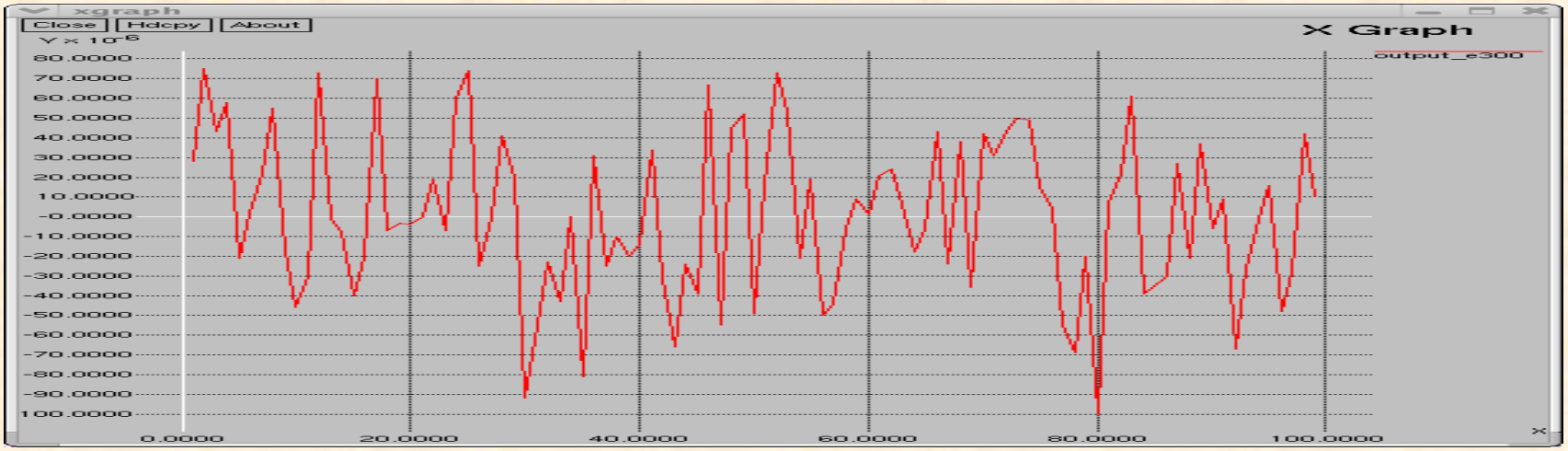
Number of measurements=100

mean roundtrip time=0.001282sec

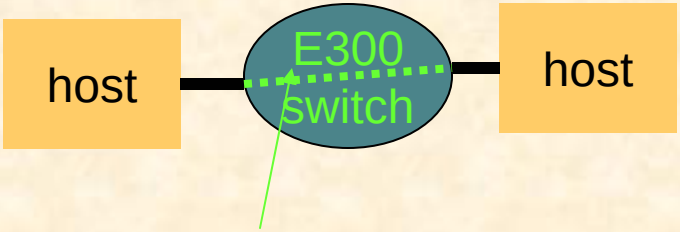
percent range: [92.177176,105.824383

actual range: [0.001182,0.001357

std_dev (%)= 3.120498



Comparison of LAN connections: Layer-2 cross-connect on Force10 subnet Layer-3 on Cisco 3750

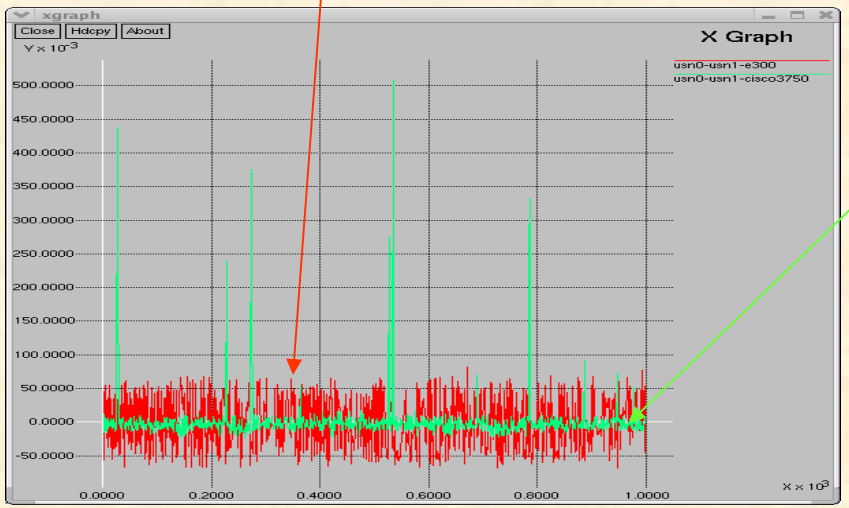


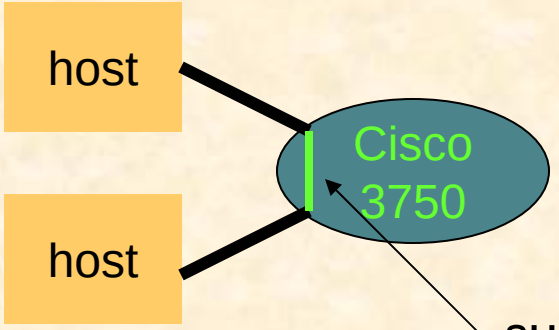
usn0-usn1-e300:

Mean time =0.211787ms

percent range: [67.992896,138.818836]

Layer-2 cross-connect via VLAN





subnet layer-3

usn0-usn1-cisco3750

mean time=0.491746ms

percent range: [95.374373,203.356866]

UT-BATTELLE

Contents

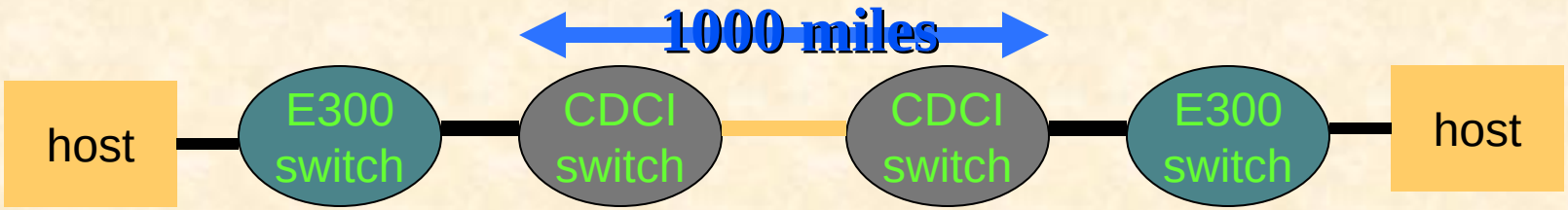
Brief Review of USN

Next Phase Activities

VLANS over L2 and L3 Networks

- Baseline measurements
- SONET vs MPLS VLANs
- Composite SONET-MPLS VLANS
- Summary

SONET Channel OC21c
1000 miles 1GigE VLAN



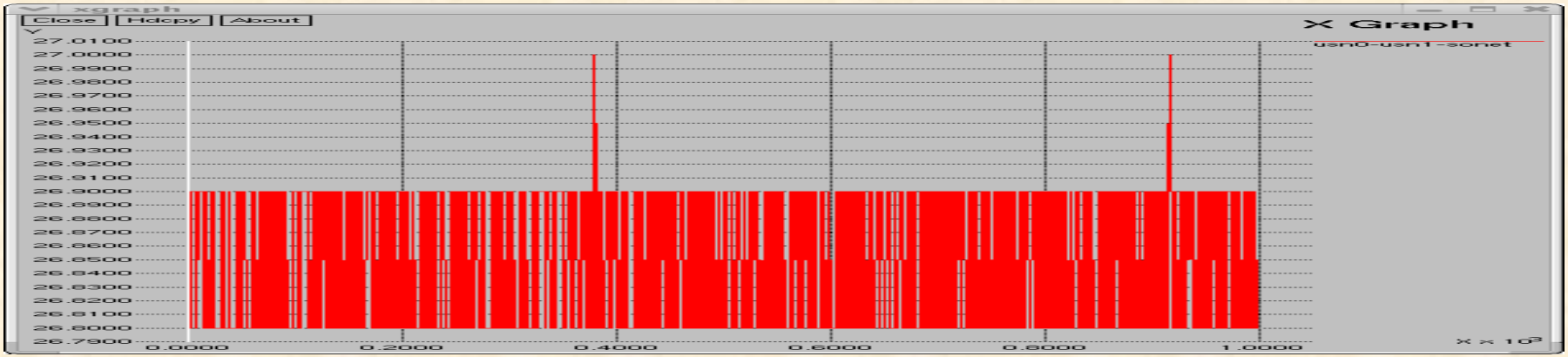
Number of PING measurements=1000

mean time=26.845877ms

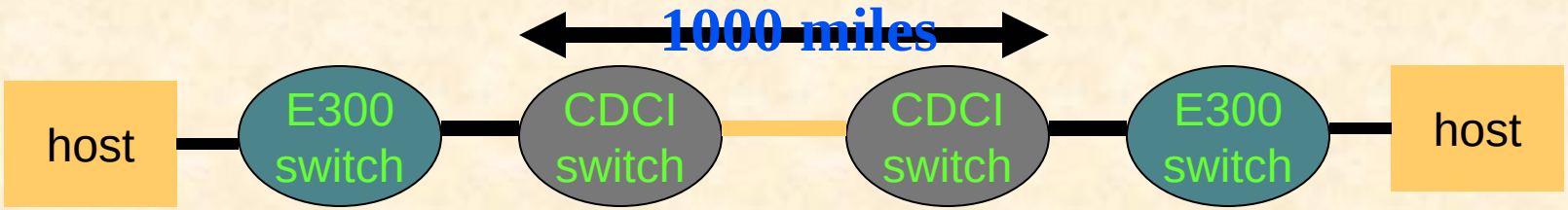
percent range: [99.829108,100.574104]: 0.744996

range: [26.799999,27.000000]: 0.200001

std_deviation (percent)= 0.187035



SONET Channel OC21c: TCP measurements 1000 miles 1GigE VLAN



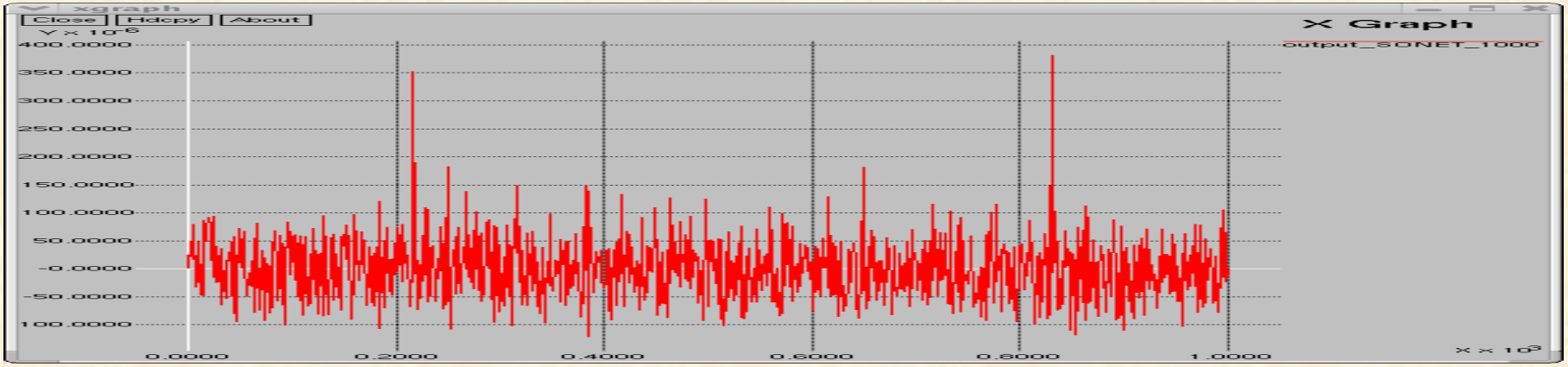
Number of TCP measurements=1000

mean=0.028064sec

percent range: [99.569830,101.358615]: 1.788785

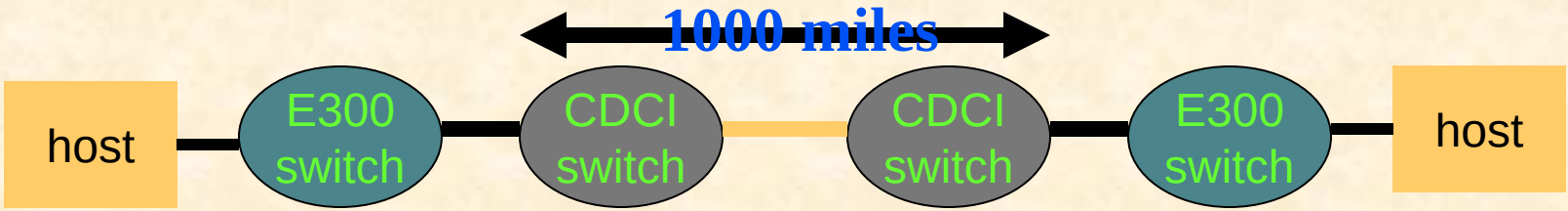
actual range: [0.027943,0.028445]: 0.000502

std_deviation_percent= 0.184920



SONET Channel OC21c: TCP measurements

1000 miles 1GigE VLAN normalized with LAN



usn1-usn0-SONET

mean=0.028037

percent range:

[99.598089,101.417140]

std_dev (%)= 0.214886

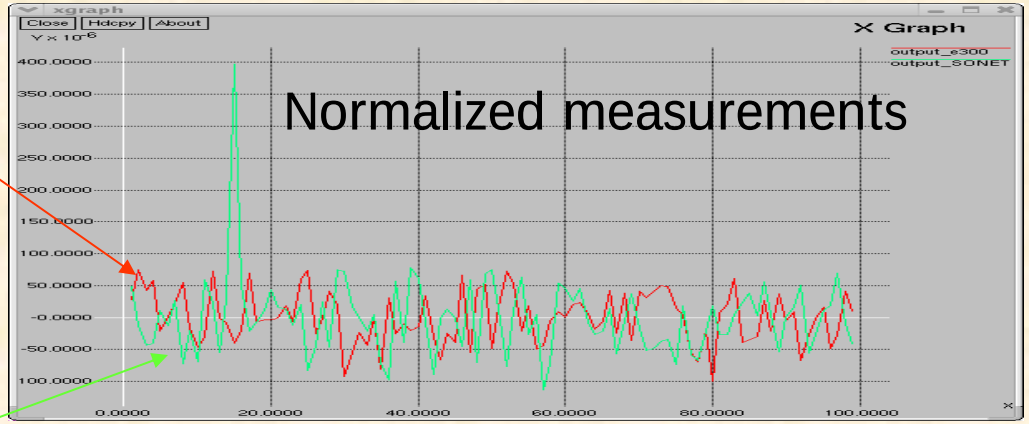
usn1-usn0-e300

mean=0.001282

percent range:

[92.177176,105.824383]

STD_dev (%)= 3.120498



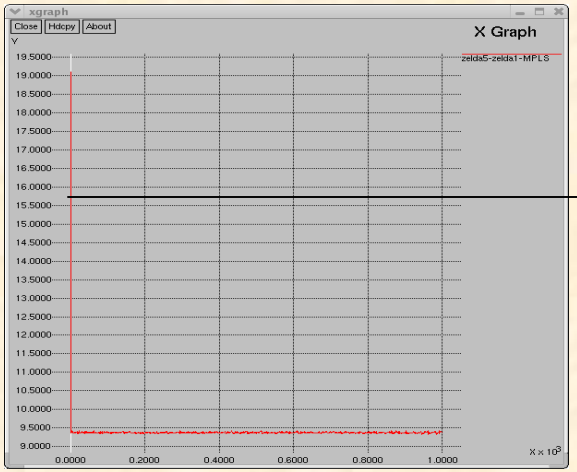
<u>Conclusion:</u>

Jitter levels of 1000 miles of SONET connection is of same order as LAN

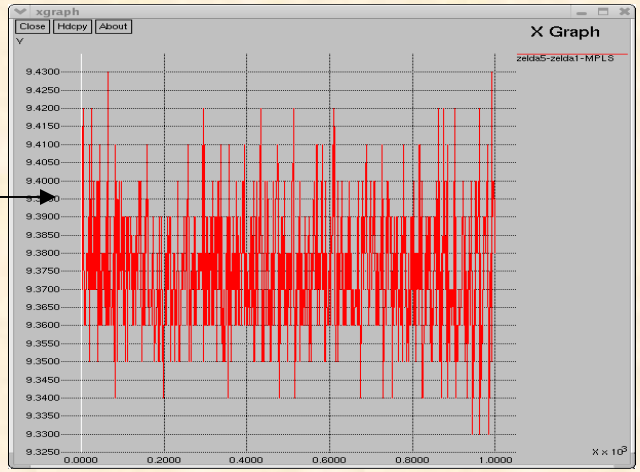
MPLS Tunnel with VLAN Tagging
1Gig Tunnel ~300 miles

1Gig MPLS tunnel

Number of ping measurements=999
mean time=9.384557ms
percent range: [99.418653,203.525865]: 104.107212
variance= 0.094847
std_deviation (%)= 3.281692

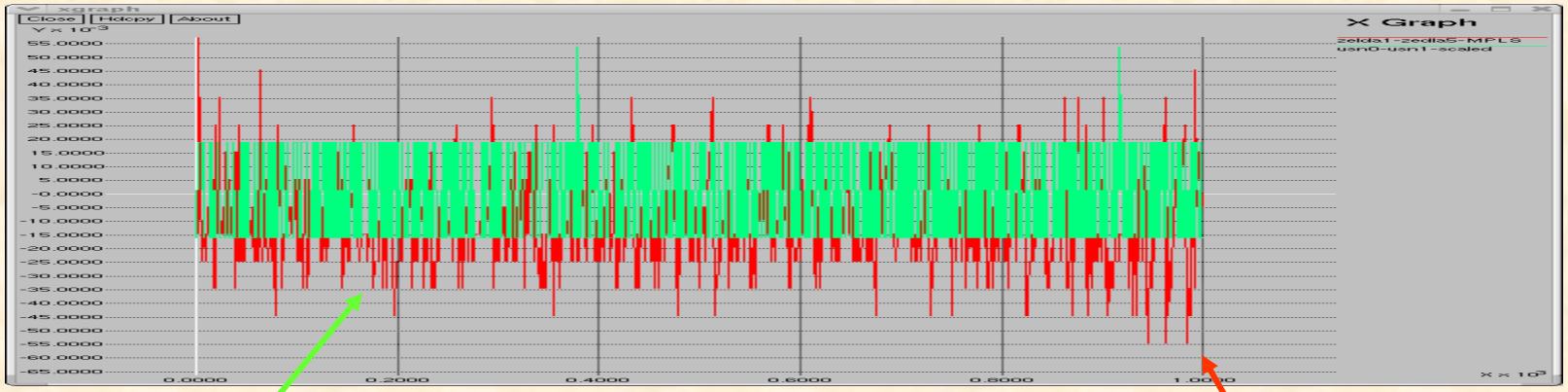


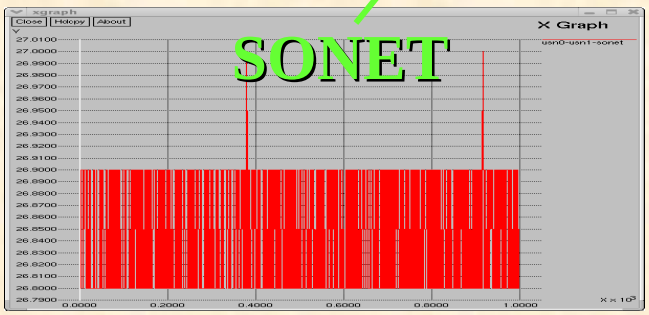
outlier removed

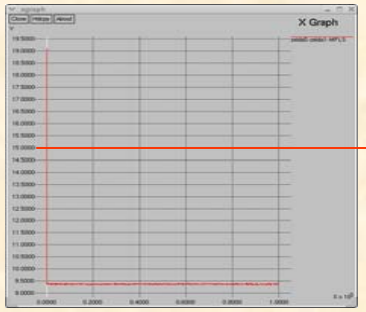


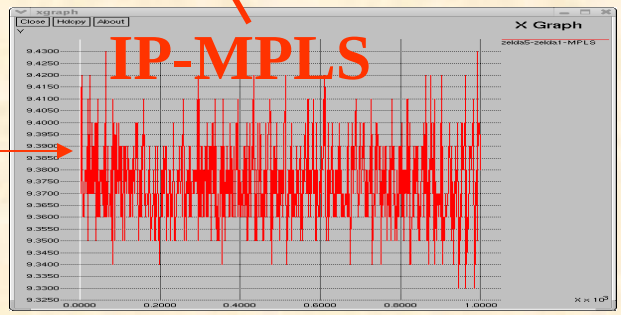
Comparison of VLANs: SONET vs. MPLS tunnels

Measurements are normalized for comparison:









mean time=26.845877ms
percent range: [99.8,100.6]
std_dev (%)= 0.187035

mean time=9.384557ms
percent range:[99.4,203.5]
std_dev (%)= 3.281692

<u>Conclusion</u>
VLANs over SONET have smaller jitter levels

Contents

Brief Review of USN
Next Phase Activities
VLANS over L2 and L3 Networks
- Baseline measurements
- SONET vs MPLS VLANs
- Composite SONET-MPLS VLANS

Summary

Composition of VLANs Over Fundamentally different networks

Feasibility and Performance of Composed SONET-MPLS VLANS:

Data-plane unification of dedicated paths over paths provisioned over layers 1 through 3-1

VLANs are typically native to layer-2: other layers need to be moved up/down to implement VLANs:

SONET connections (layer1): VLANs are provisioned using edge switches (E300 in our case)

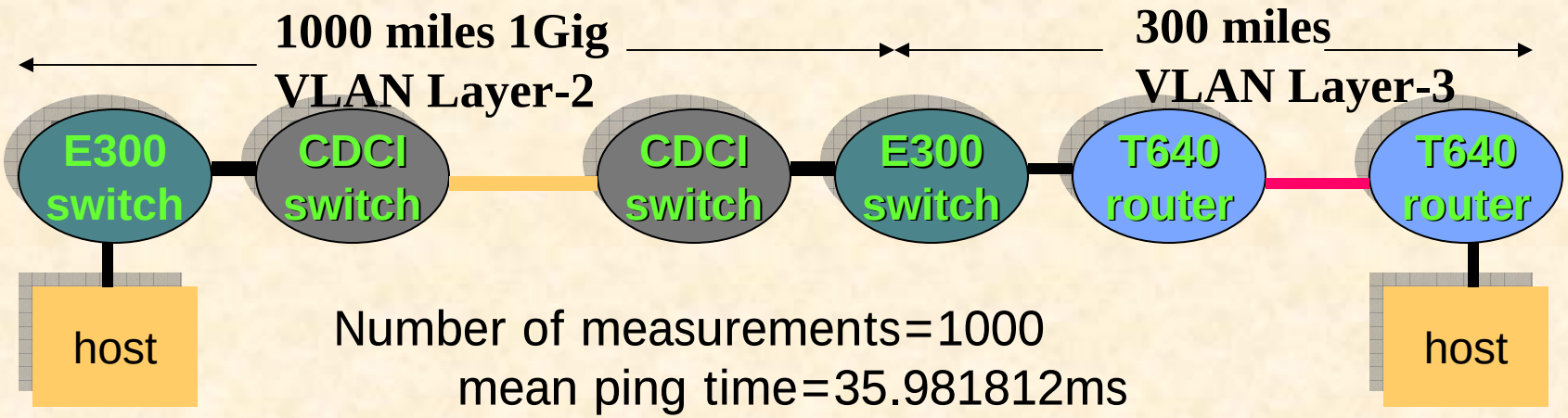
Layer-2 connections – GigE WAN/LAN PHY: VLANs are provisioned natively

IP networks (layer 3) – VLANs are provisioned over MPLS tunnels using IEEE 802.1q – routers may differ as to how they implement this feature

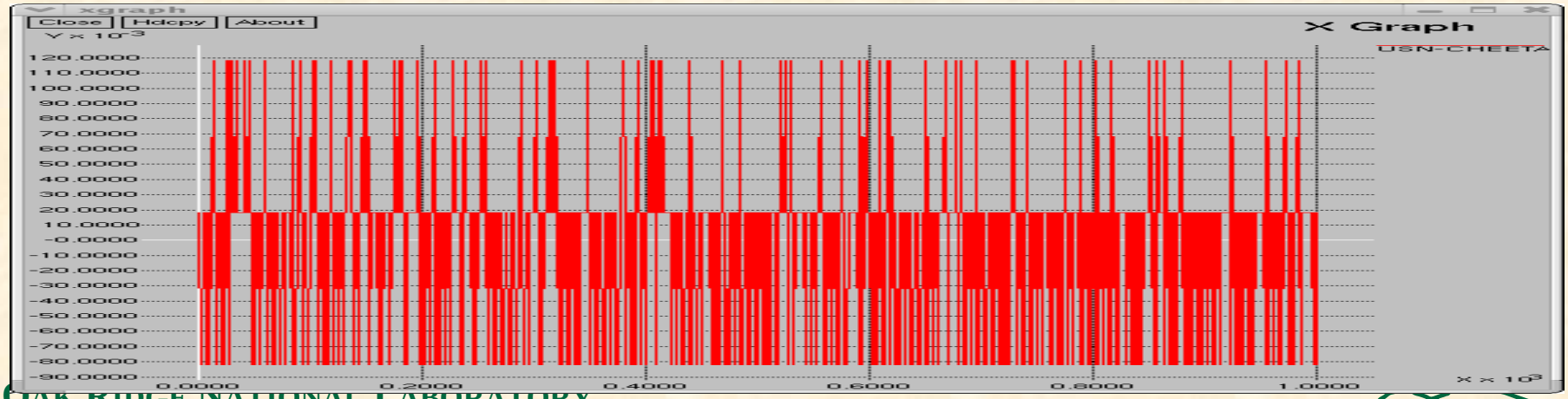
Broader Question: Peering data-paths across networks that provide VLANs over Layer-1 or layer-2 or layer-3

Composed VLAN:

SONET and Layer-3 Channels - Gig 1300 miles

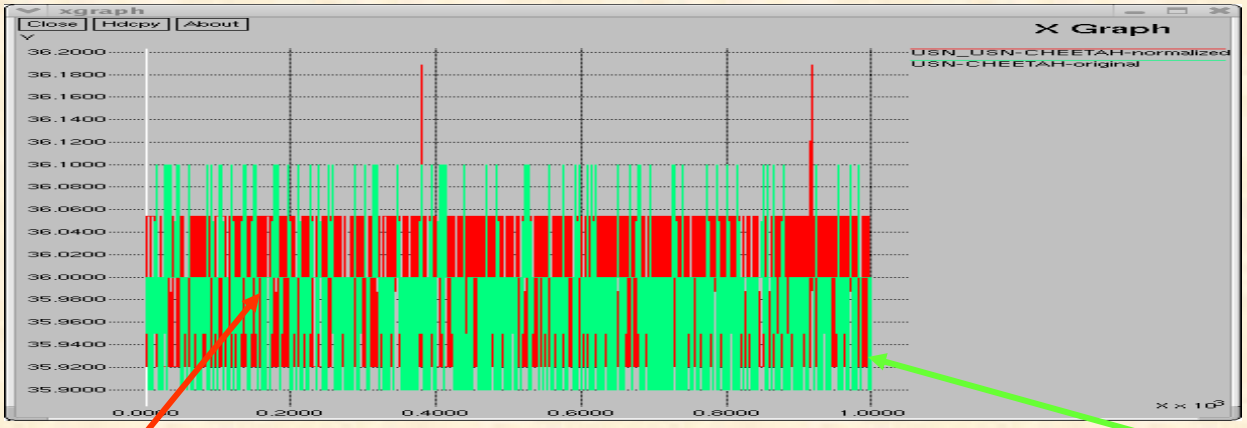


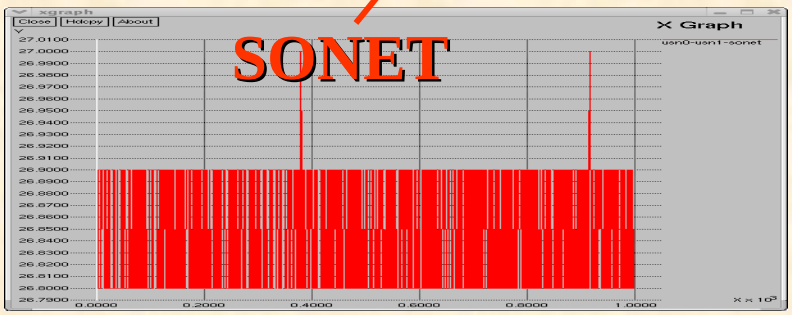
Number of measurements=1000
mean ping time=35.981812ms
percent range: [99.772635,100.328463]
range: [35.900002,36.099998]: 0.199997
std_deviation (percent)= 0.151493

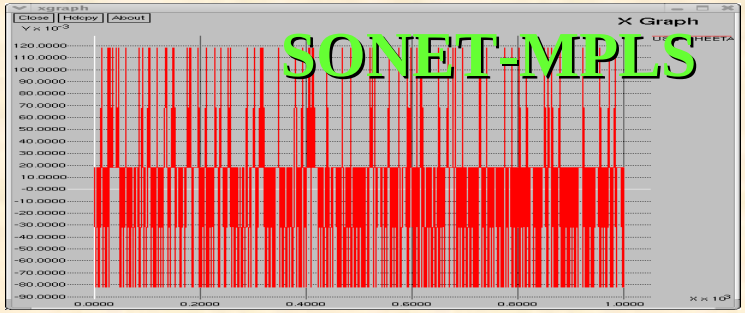


Comparison of VLANs: SONET vs. SONET-MPLS composed tunnels

Measurements are normalized for comparison:







mean time=26.845877ms
percent range: [99.8,100.6]
std_dev (%)= 0.187035

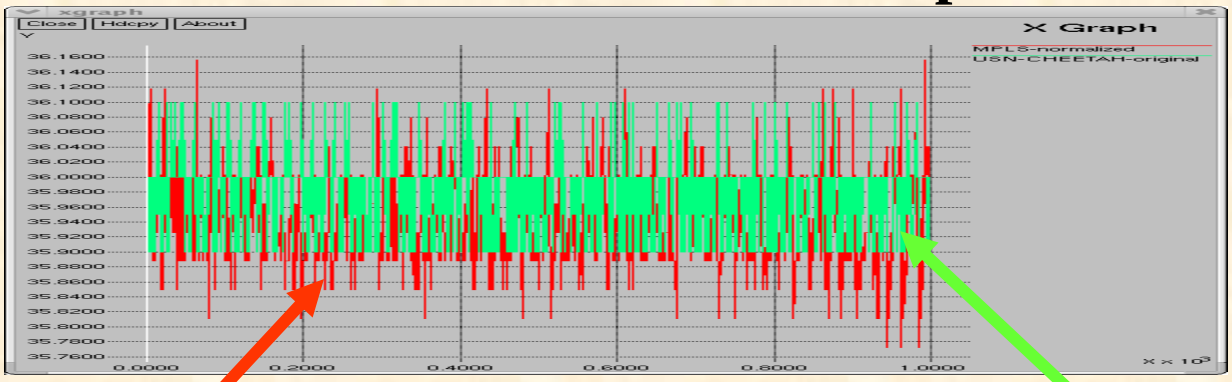
mean ping time=35.981812ms
percent range: [99.8,100.3]
std_dev (%)= 0.151493

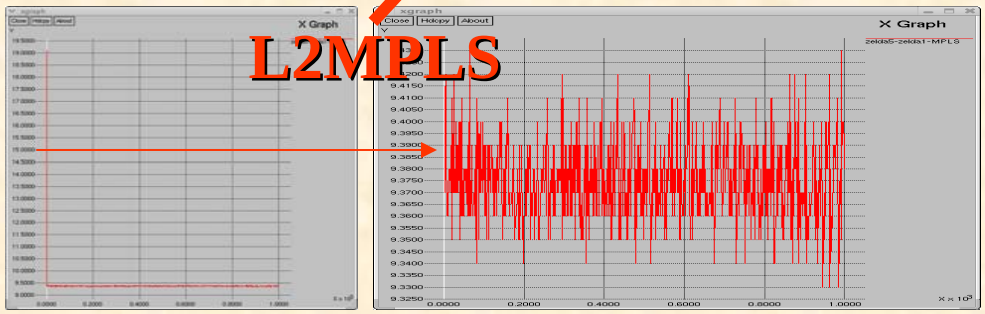
<u>Conclusion</u>
VLANs over SONET have smaller jitter levels

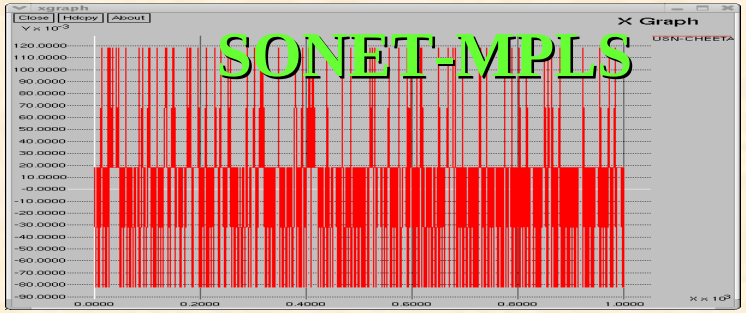
UT-BATTELLE

Comparison of VLANs: L2MPLS vs. SONET-MPLS composed tunnels

Measurements are normalized for comparison:







mean time=9.384557ms
percent range:[99.4,203.5]
std_dev (%)= 3.281692

mean ping time=35.981812ms
percent range: [99.8,100.3]
std_dev (%)= 0.151493

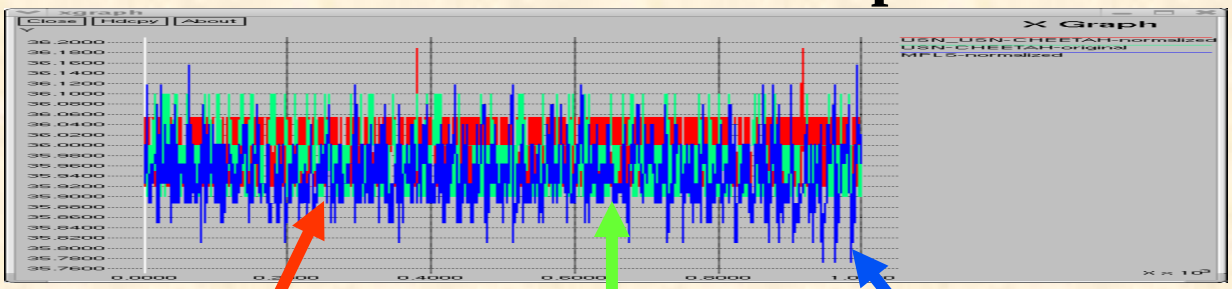
Conclusion

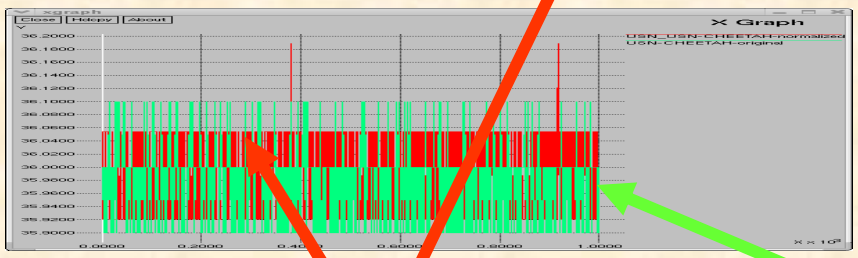
SONET-MPLS composed have smaller jitter levels

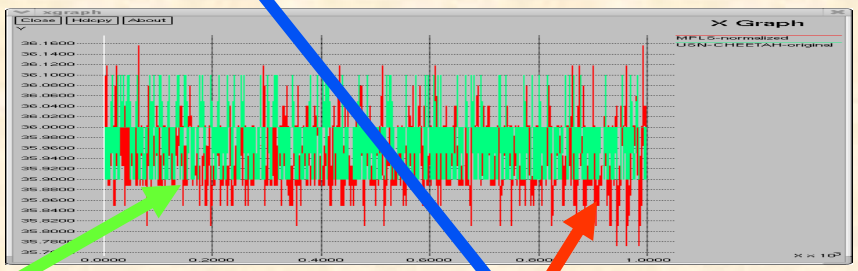
UT-BATTELLE

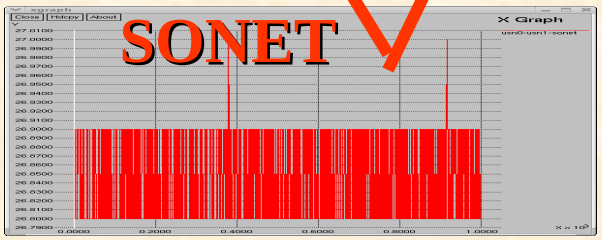
Normalized Comparison of VLANs:
SONET - SONET-MPLS composed – L2MPLS

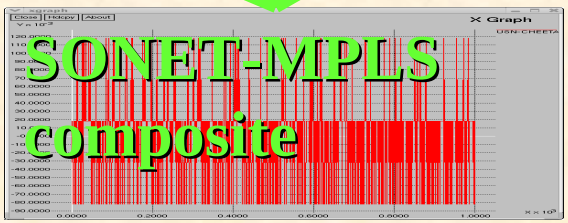
Measurements are normalized for comparison:

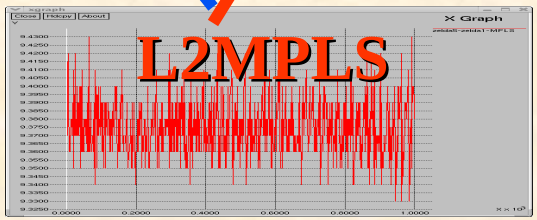












mean time=26.845877ms
std_dev (%)= 0.187035

mean time=35.981812ms
std_dev (%)= 0.151493

mean time=9.384557ms
std_dev (%)= 3.281692

SONET channels have smaller jitter levels

Normalization Framework

Basic Question: We collected measurements on two connections of different lengths and types, how do we objectively compare parameters?

Example: Ping measurements on 1000 mile SONET-VLAN and 300 mile MPLS-VLANS, can we objectively conclude about jitte on such VLANs?

$M_T(d)$ Measurements on path of type *T* of distance *d*

$\hat{M}_T(d)$ Estimates of measurements on path of type T of distance d

$P \odot \hat{M}_T(d)$ Parameters computed using measurements

$$\begin{tikzcd}
M_{T_1}(d_1) \arrow[dr, "\Theta_{T_1}"] & & M_{T_2}(d_2) \arrow[dl, "\Theta_{T_2}"'] \\
& \hat{M}_{T_1}(d) \arrow[d] & \hat{M}_{T_2}(d) \arrow[d] \\
& P \odot \hat{M}_{T_1}(d) & P \odot \hat{M}_{T_2}(d) \arrow[l, "\aleph_P"']
\end{tikzcd}$$

Jitter Comparison on SONET-MPLS VLANs

Basic Question: We collected measurements on two connections of different lengths and types, how do we objectively compare parameters?

$$M_{SONET}(1000) \xrightarrow[\text{scale}]{\Theta_{T_1}} \hat{M}_{SONET}(300) \rightarrow P \odot \hat{M}_{SONET}(300)$$

$$M_{MPLS}(300) \xrightarrow[\text{identity}]{\Theta_{T_2}} \hat{M}_{MPLS}(d) \rightarrow P \odot \hat{M}_{MPLS}(d)$$

$$P \odot \hat{M}_{MPLS}(d) \xrightarrow{\aleph_P} P \odot \hat{M}_{SONET}(300)$$

scale with respect to mean

Another Method

P **- FFT**

$\aleph_P$ **- Identity**

Conclusions:

Status Update of USN and Future Plans

Our Experiments and Analysis of VLANs show:

SONET-VLANs provide:

- Lower levels of jitter both at ping and TCP level

VLANs unify data plane connections over layers 1-3

- data paths of layers 1-3 can be aligned at VLAN level
- composed paths inherit low jitter characteristics of layer 1 components

Future Work:

- Detailed application-level and protocol-level performance evaluation
- Automated methods for VLAN alignment at disparate networks
- Security aspects of VLAN peering