Introduction to perfSONAR for NSF CC-NIE Awardees

Jason Zurawski

Eli Dart

Mary Hester

Lauren Rotman

Brian Tierney

ESnet Engineering & Outreach

CC-NIE Webinar

June 4th & 24th, 2013

Agenda

- Introduction and Purpose
- Science DMZ Overview
- Why Network Performance Matters
- Motivations & Technologies
- Hardware & Deployment Options
- End Results & Debugging
- Conclusions

perfSONAR powered

Introduction & Purpose

- The "*Campus Cyberinfrastructure - Network Infrastructure and Engineering (CC-NIE)*" program:
  - Invests in improvements and re-engineering at the campus level to support a range of data transfers supporting computational science and computer networks and systems research
  - Supports Network Integration activities tied to achieving higher levels of ***performance***, ***reliability*** and ***predictability*** for science applications and distributed research projects
- The bolded items can be tricky: this talk will introduce some broad concepts that will help:
  - Capable network architectures
  - Advanced data movement tools and procedures
  - ***Federated End-to-End monitoring***
- We will not be digging too deep technically – those topics will be explored at a later date (with perhaps a different crowd) if there is interest.

perfSONAR powered

NSF 13-530: http://www.nsf.gov/pubs/2013/nsf13530/nsf13530.htm

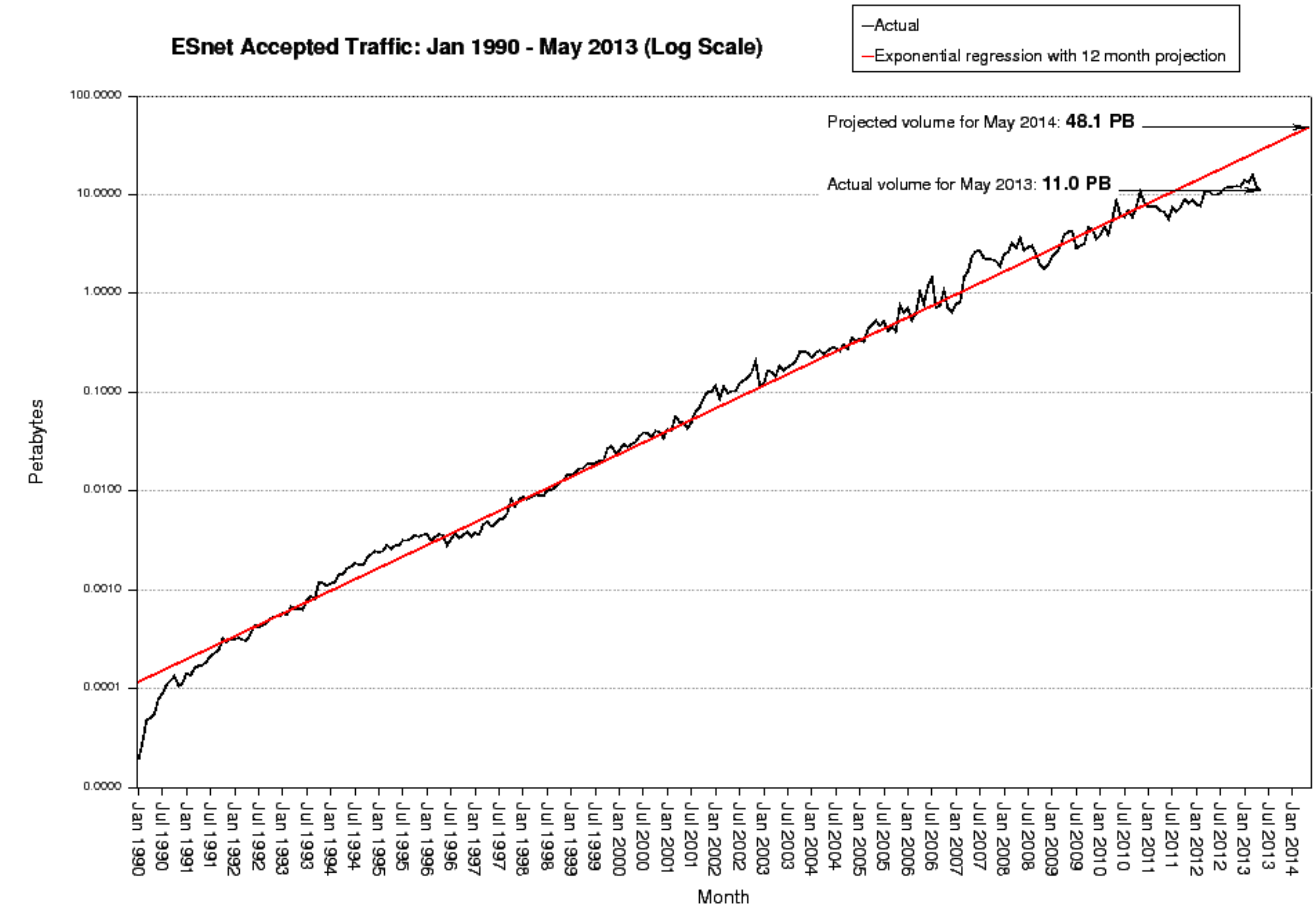
Big Data

- ***Genomics***
  - Sequencer data volume increasing 12x over the next 3 years
  - Sequencer cost decreasing by 10x over same time period
- ***High Energy Physics***
  - LHC experiments produce & distribute petabytes of data/year
  - Peak data rates increase 3-5x over 5 years
- ***Light Sources***
  - Many detectors on a Moore's Law curve
  - Data volumes rendering previous operational models obsolete
- ***Common Threads***
  - Increased capability, greater need for data mobility due to span/depth of collaboration space
  - Global is the new local. Research is no longer done within a domain. End to end involves many fiefdoms to cross – and yes this becomes **your** problem when **your** users are impacted

© Owen Humphreys/National Geographic Traveler Photo Contest 2013

perfSONAR powered

Big Data (~100 PB for 2013)



The Risks of Change

"In any large system, there is always something broken.”

Jon Postel

© Dog Shaming 2012

- Many will encounter unforeseen (and therefore challenging) situations:
  - Upgrading networks breaks them (loss of configuration, etc.)
  - Synergy between the new and the old
  - New use-cases and users
- Mitigating the risk can be done in a number of ways:
  - Analysis and alteration to architecture
  - Careful thought to security/data policies in target areas
  - Integration of software designed to exercise the network, and alert/visualize
- Proactive vs. Reactive Stance: measure twice on engineering design of the network, cut once in the implementation

perfSONAR powered

Agenda

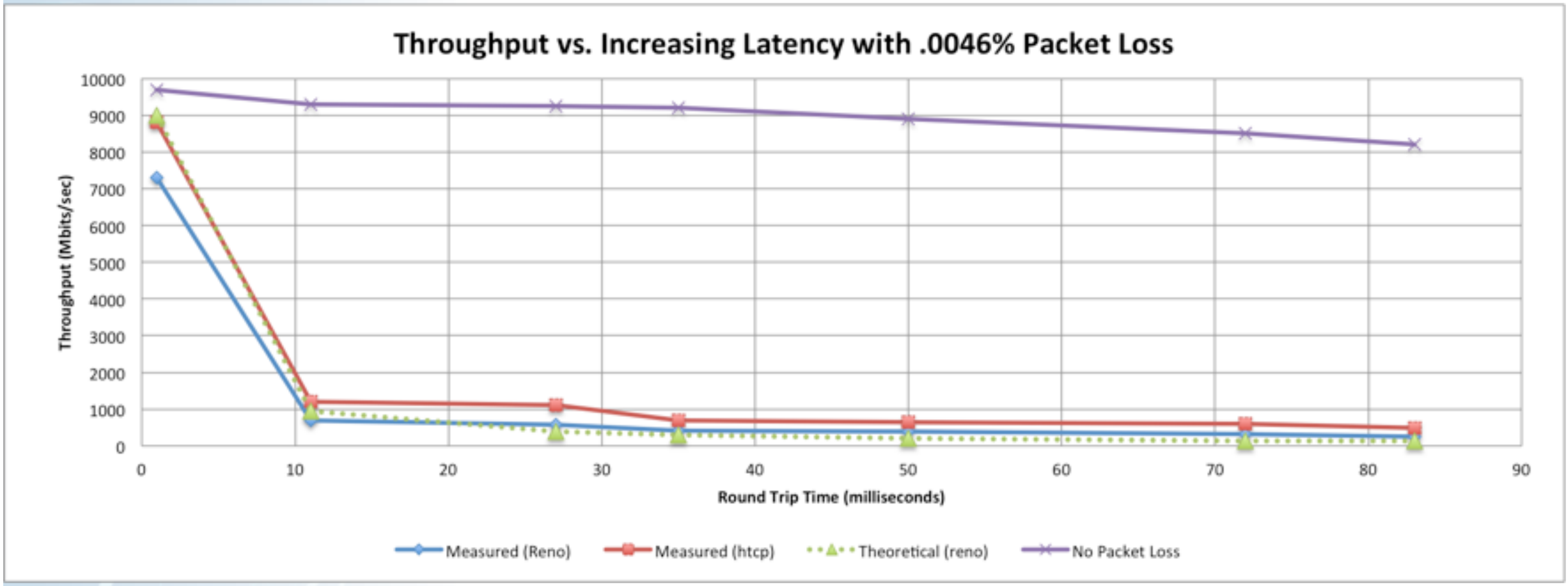
- Introduction and Purpose
- Science DMZ Overview
- Why Network Performance Matters
- Motivations & Technologies
- Hardware & Deployment Options
- End Results & Debugging
- Conclusions

Science DMZ Overview

- ESnet has a lot of experience with different scientific communities at multiple data scales
- Significant commonality in the issues encountered, and solution set
  - The causes of poor data transfer performance fit into a few categories with similar solutions
    - Un-tuned/under-powered hosts
    - Packet loss issues
    - Security devices
  - A successful model has emerged – the Science DMZ
    - This model successfully in use by CMS/ATLAS, ESG, NERSC, ORNL, ALS, and others
- The Science DMZ is a ***<u>blueprint</u>*** for network design.
  - Not all implementations look the same, but share common features
  - Some choices don't make sense for everyone, caveat emptor

perfSONAR powered

Why Network Performance Matters



perfSONAR powered

The Science DMZ in 1 Slide

Consists of **three key components**, all required:

“Friction free” network path

- Highly capable network devices (wire-speed, deep queues)
- Virtual circuit connectivity option
- Security policy and enforcement specific to science workflows
- Located at or near site perimeter if possible

Dedicated, high-performance Data Transfer Nodes (DTNs)

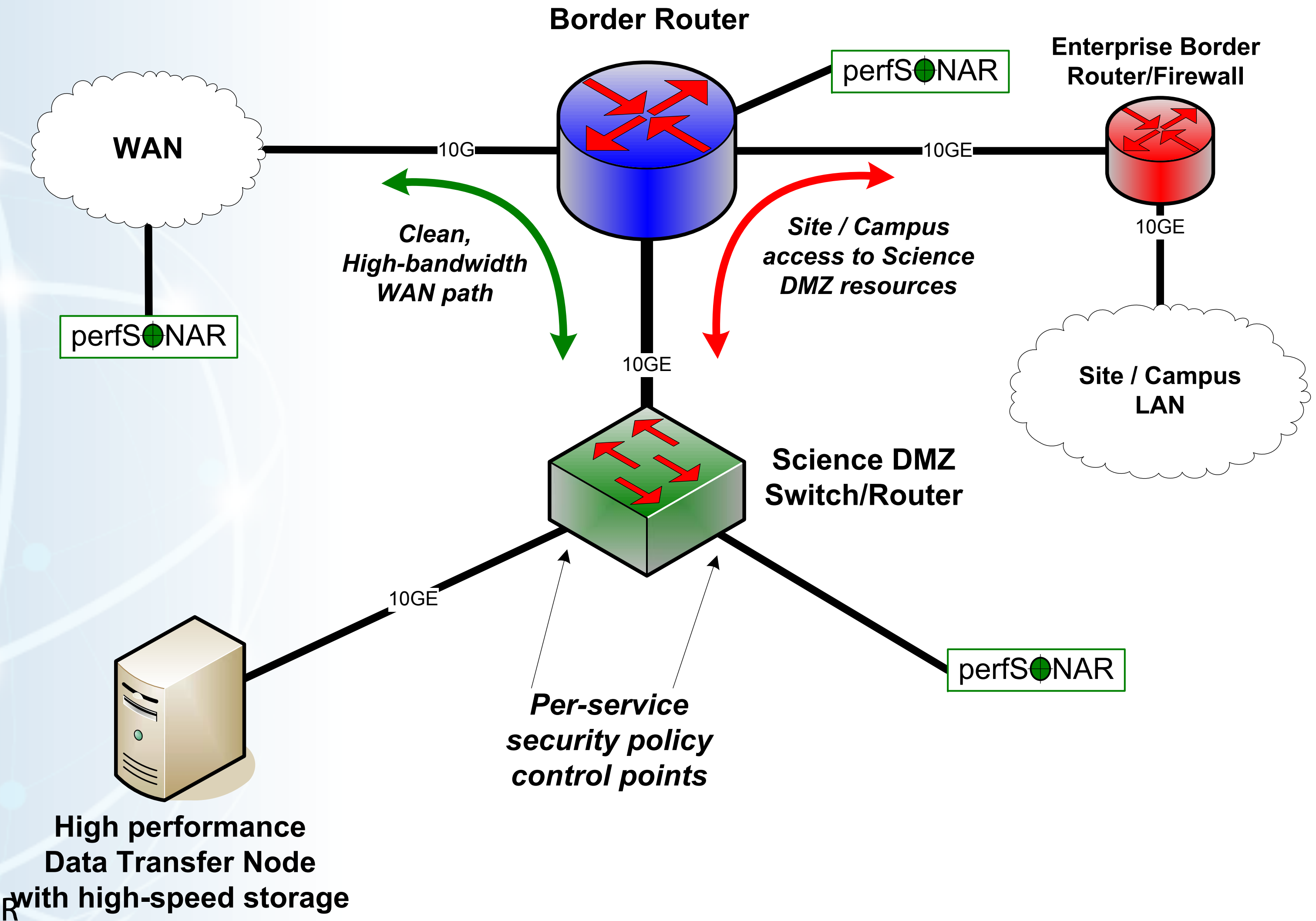
- Hardware, operating system, libraries all optimized for transfer
- Includes optimized data transfer tools such as Globus Online and GridFTP

Performance measurement/test node

- perfSONAR

Details at http://fasterdata.es.net/science-dmz/

Science DMZ – Simple Abstract Cartoon


Border Router
perfSONAR
Enterprise Border Router/Firewall
WAN
10G
10GE
10GE
Clean, High-bandwidth WAN path
Site / Campus access to Science DMZ resources
perfSONAR
10GE
Site / Campus LAN
Science DMZ Switch/Router
10GE
Per-service security policy control points
perfSONAR
High performance Data Transfer Node with high-speed storage

perfSONAR powered

Science DMZ Takes Many Forms

There are many ways to combine the Science DMZ elements – it all depends on what you need to do

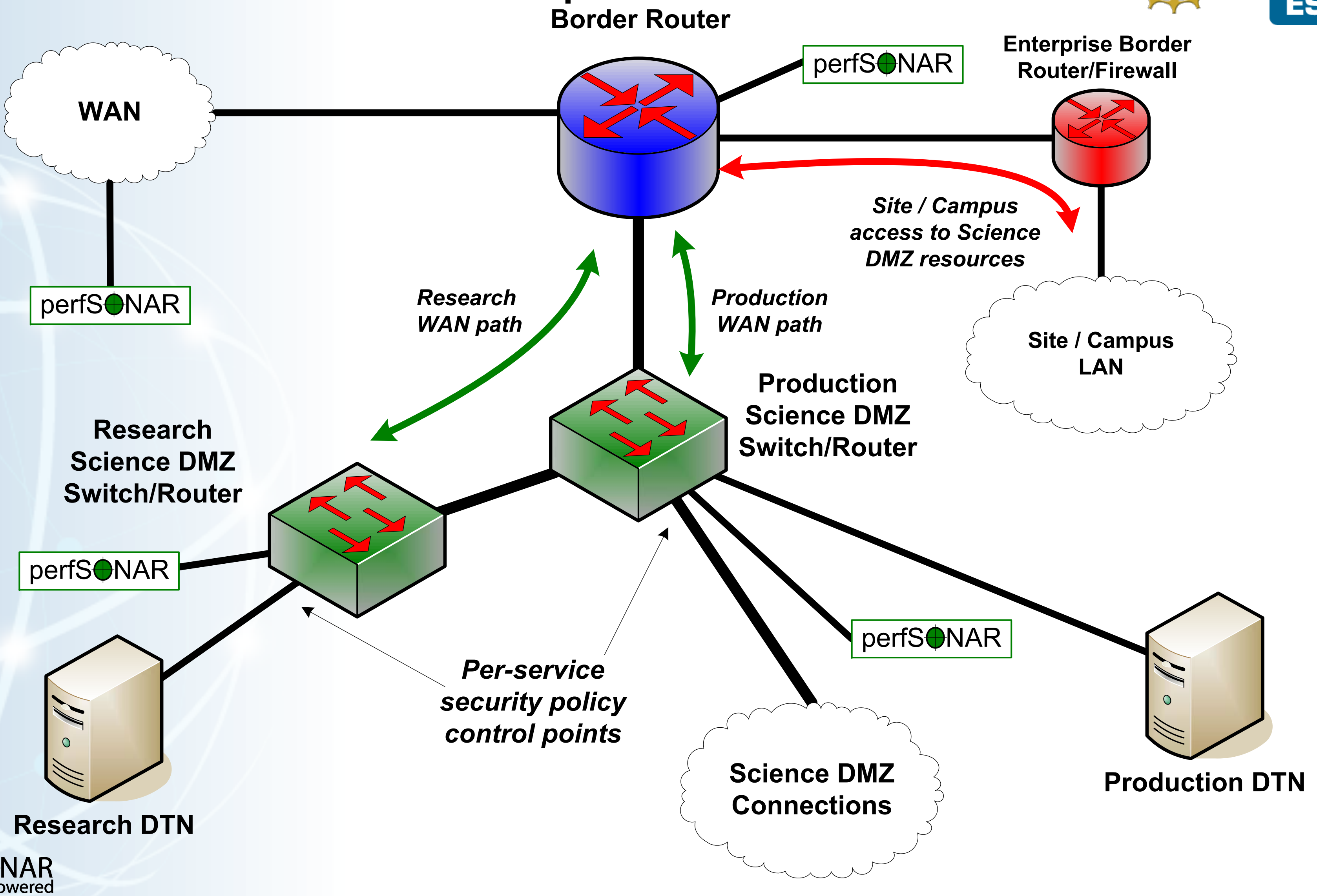
- Small installation for a project or two
- Facility inside a larger institution
- Institutional capability serving multiple departments/divisions
- Science capability that consumes a majority of the infrastructure

Some of these are straightforward, others are less obvious

- Science DMZ model used to support research
  - The network is both the environment and the subject of research
  - Science DMZ is a good fit for several reasons
    - Isolate research from production when research is in the unstable phase
    - Separation of administrative control
  - Some research projects need high-performance end to end networking, but are not network research
    - HEP/LHC, Astronomy, “Big Data,” etc.
    - The Science DMZ is production cyberinfrastructure
- Ideally, both network research and production data-intensive science could coexist

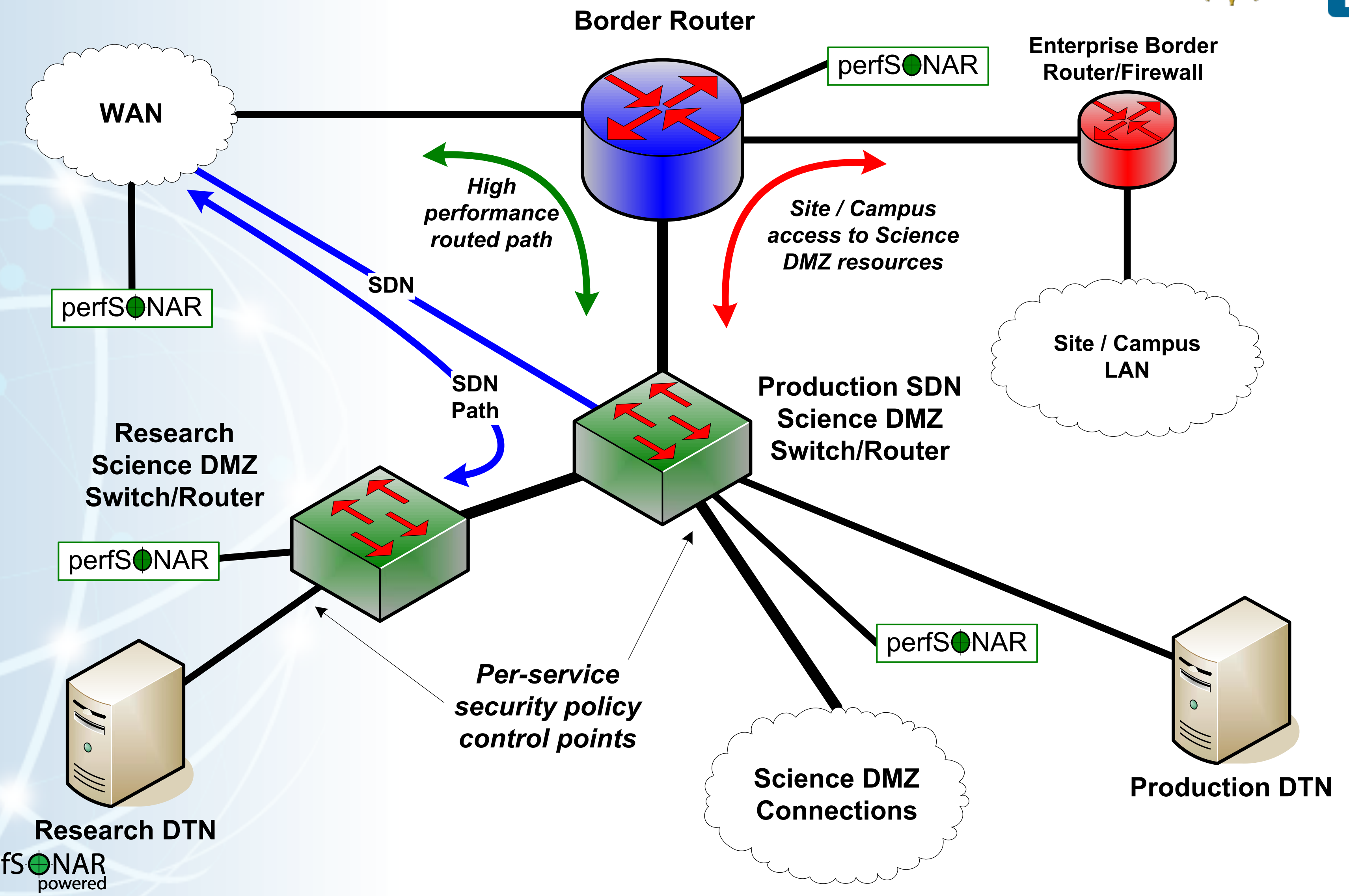
perfSONAR powered

Science DMZ With Separate Research Area


Border Router
perfSONAR
Enterprise Border Router/Firewall
WAN
Site / Campus access to Science DMZ resources
perfSONAR
Research WAN path
Production WAN path
Site / Campus LAN
Production Science DMZ Switch/Router
Research Science DMZ Switch/Router
perfSONAR
perfSONAR
Per-service security policy control points
Science DMZ Connections
Production DTN
Research DTN

perfSONAR powered

Science DMZ – Production SDN Connection


Border Router
perfSONAR
Enterprise Border Router/Firewall
WAN
High performance routed path
Site / Campus access to Science DMZ resources
SDN
perfSONAR
Site / Campus LAN
SDN Path
Production SDN Science DMZ Switch/Router
Research Science DMZ Switch/Router
perfSONAR
perfSONAR
Per-service security policy control points
Science DMZ Connections
Production DTN
Research DTN

perfSONAR powered

End Game – Enabling Data Intensive Science

Using the right tool is very important

Sample Results: Berkeley, CA to Argonne, IL (near Chicago). RTT = 53 ms, network capacity = 10Gbps.

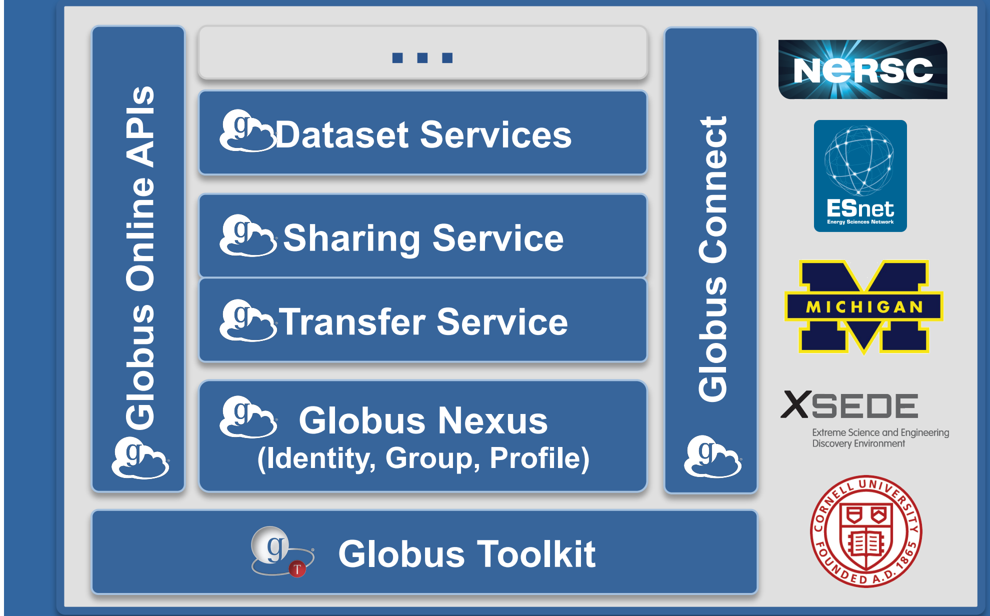
Tool	**Throughput**
scp:	140 Mbps
HPN patched scp:	1.2 Gbps
ftp	1.4 Gbps
GridFTP, 4 streams	5.4 Gbps
GridFTP, 8 streams	6.6 Gbps

Note that to get more than 1 Gbps (125 MB/s) disk to disk requires RAID.

perfSONAR powered

It should be trivial for all researchers to: **Collect, Move, Sync, Share, Analyze, Annotate, Publish, Search, Backup, & Archive** BIG DATA …but without proper kit it's very challenging

Globus Online uses SaaS approaches to address this challenge and make advanced research data management capabilities broadly accessible using just a Web browser



Globus Online APIs

…

Dataset Services

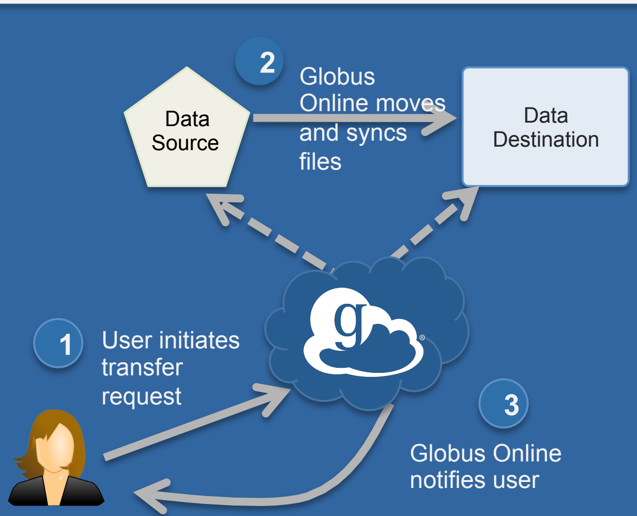
Sharing Service

Transfer Service

Globus Nexus (Identity, Group, Profile)

Globus Connect

Globus Toolkit



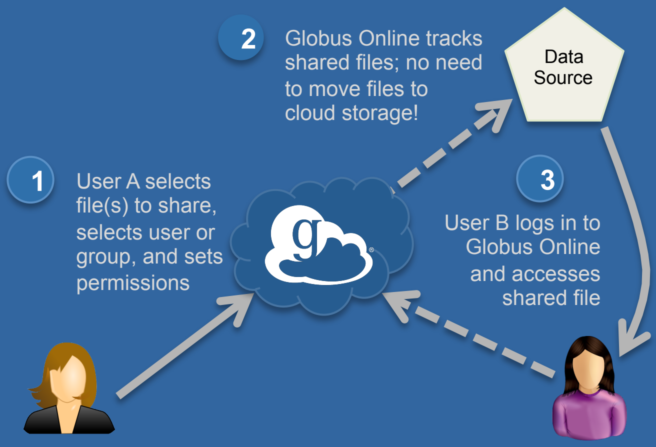
Data Source

Data Destination

1 User initiates transfer request

2 Globus Online moves and syncs files

3 Globus Online notifies user



1 User A selects file(s) to share, selects user or group, and sets permissions

2 Globus Online tracks shared files; no need to move files to cloud storage!

Data Source

3 User B logs in to Globus Online and accesses shared file

Source: R. Kettimuthu (kettimut@mcs.anl.gov)

One motivation for Science DMZ model: Soft Network Failures

Soft failures are where basic connectivity functions, but high performance is not possible.

TCP was intentionally designed to hide all transmission errors from the user:

- "As long as the TCPs continue to function properly and the internet system does not become completely partitioned, no transmission errors will affect the users." (From IEN 129, RFC 716)

Some soft failures only affect high bandwidth long RTT flows.

Hard failures are easy to detect & fix

- soft failures can lie hidden for years!

One network problem can often mask others

perfSONAR powered

Agenda

- Introduction and Purpose
- Science DMZ Overview
- Why Network Performance Matters
- Motivations & Technologies
- Hardware & Deployment Options
- End Results & Debugging
- Conclusions

perfSONAR powered

Time to Copy 1 Terabyte

10 Mbps network : 300 hrs (12.5 days)

100 Mbps network : 30 hrs

1 Gbps network : 3 hrs (are your disks fast enough?)

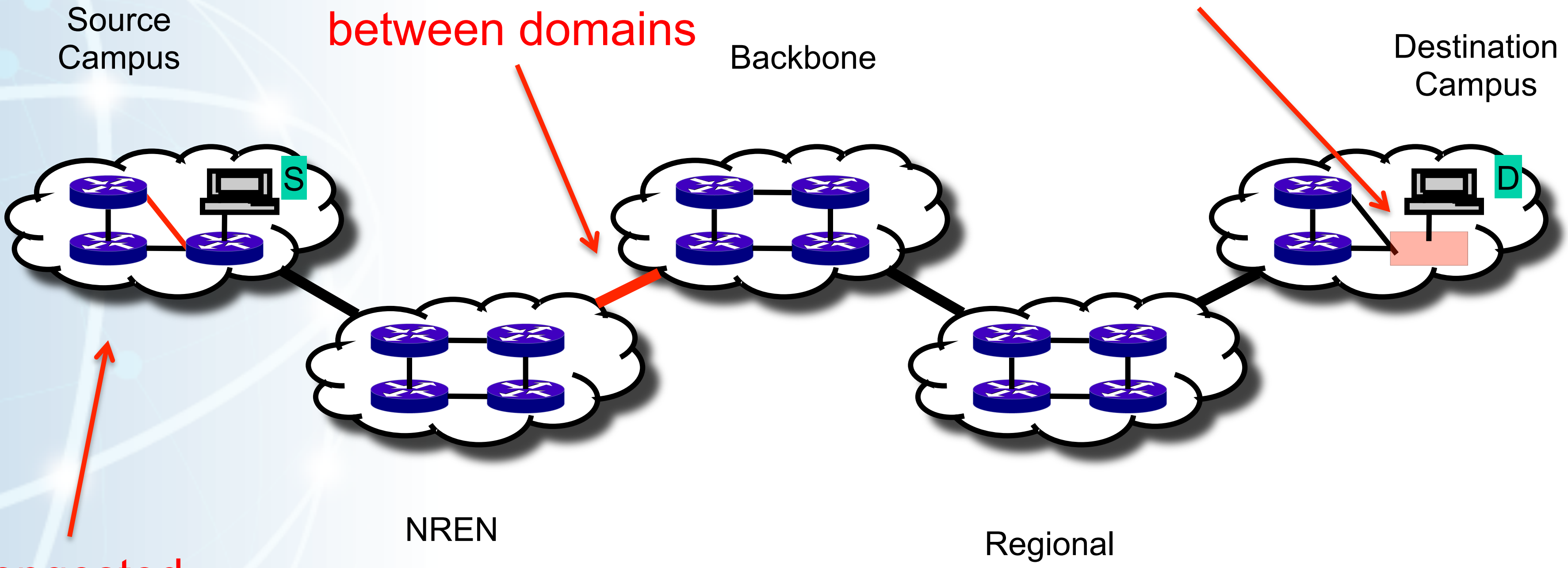
10 Gbps network : 20 minutes (need really fast disks and filesystem)

These figures assume some headroom left for other users

Compare these speeds to:

- USB 2.0 portable disk
  - 60 MB/sec (480 Mbps) peak
  - 20 MB/sec (160 Mbps) reported on line
  - 5-10 MB/sec reported by colleagues
  - 15-40 hours to load 1 Terabyte

Where Are The Problems?



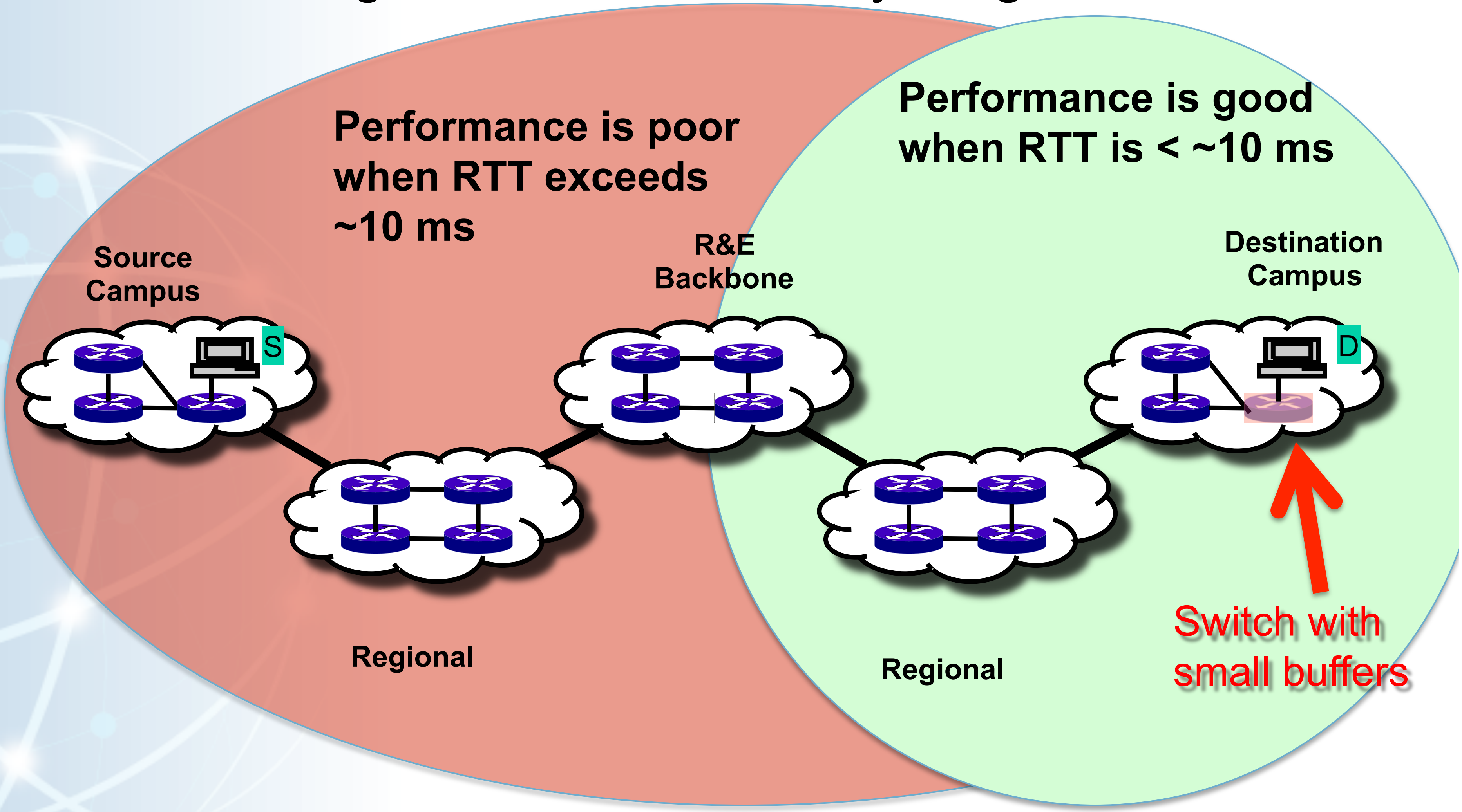
Congested or faulty links between domains

Latency dependant problems inside domains with small RTT

Congested intra- campus links

perfSONAR powered

Local Testing Will Not Find Everything

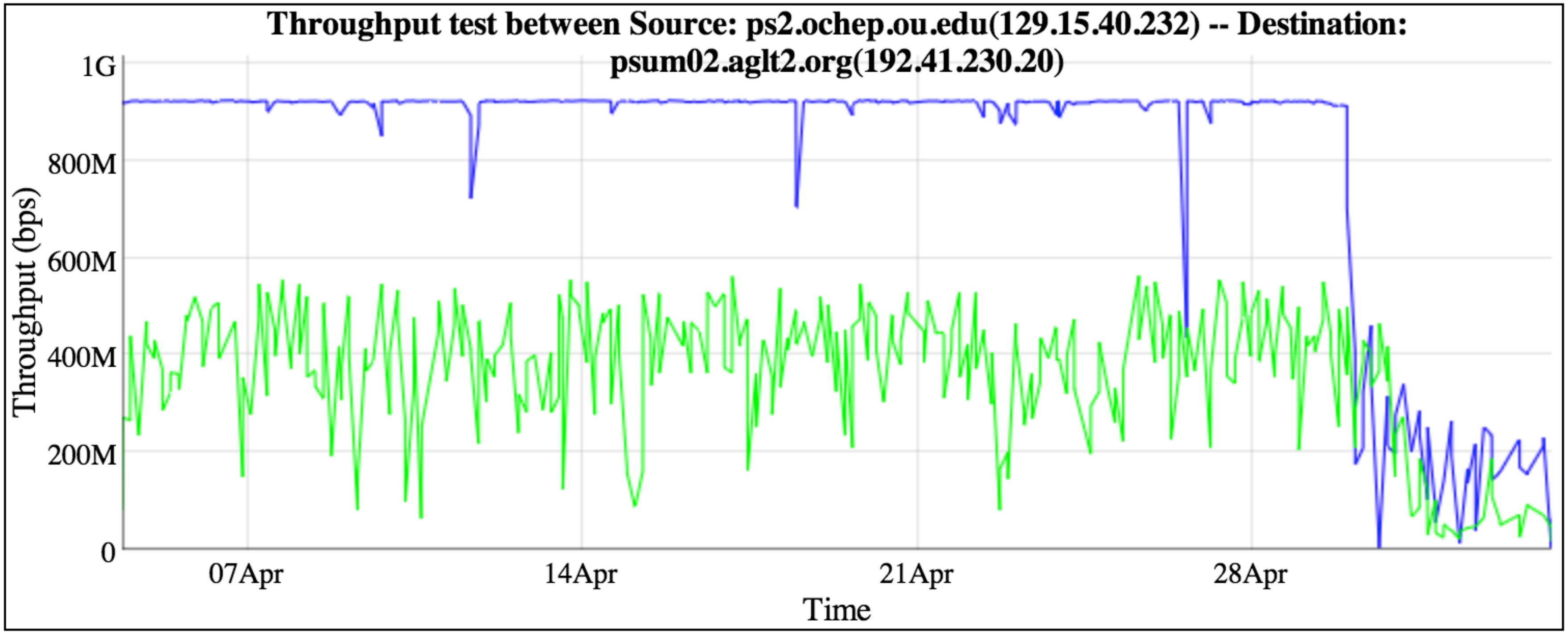


Performance is poor when RTT exceeds ~10 ms

Performance is good when RTT is < ~10 ms

Switch with small buffers

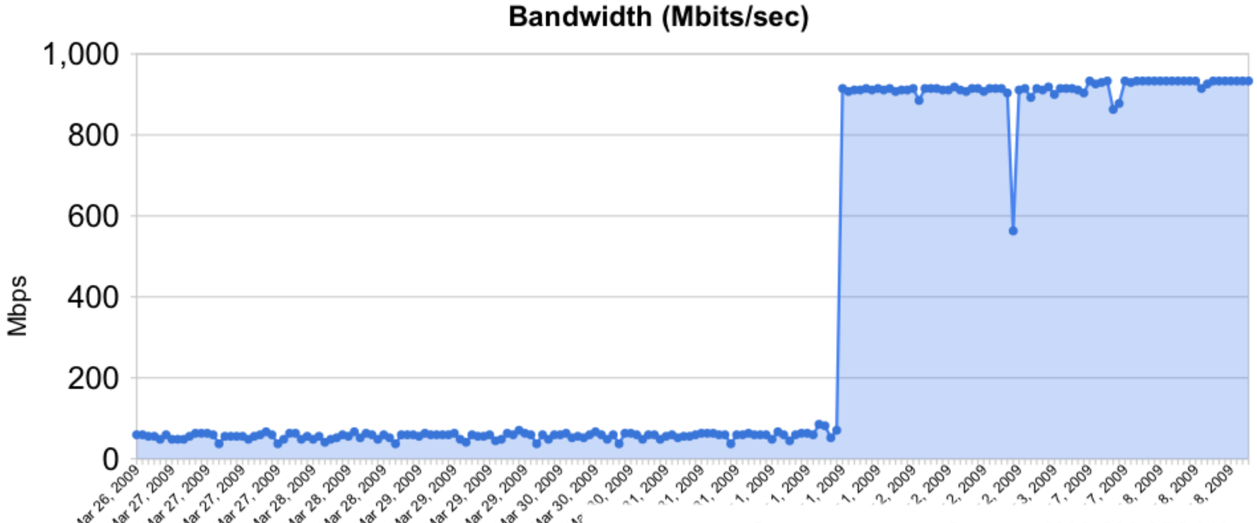
What Monitoring Can (and Cannot) Tell You



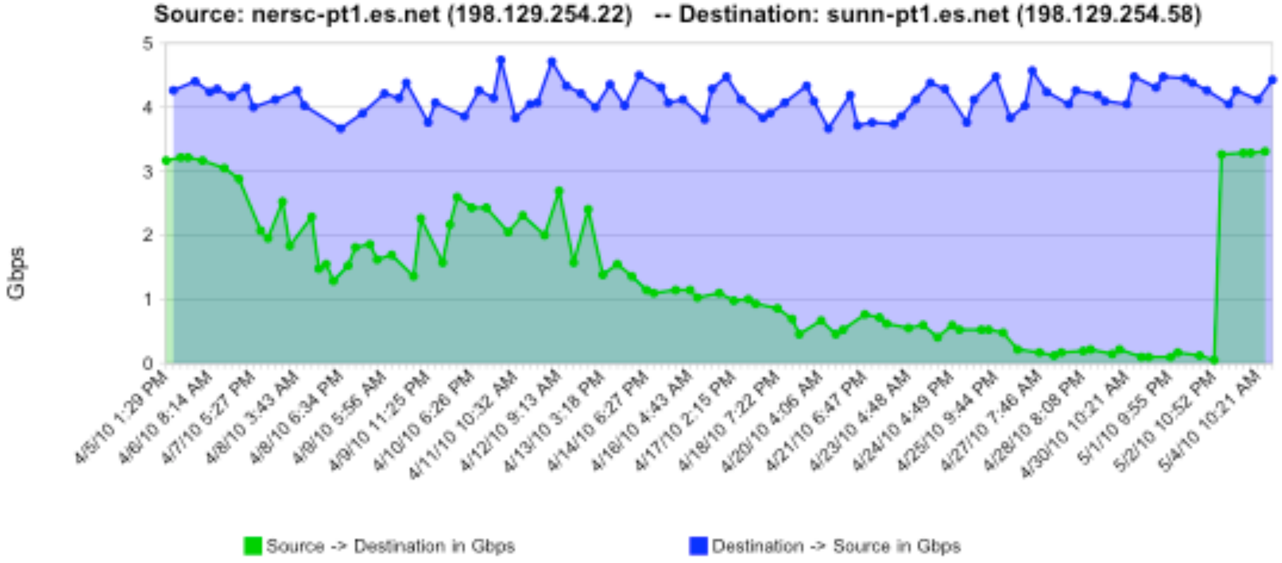
Can you tell, ***by looking,*** what is going on here?

perfSONAR powered

Sample Soft Failures

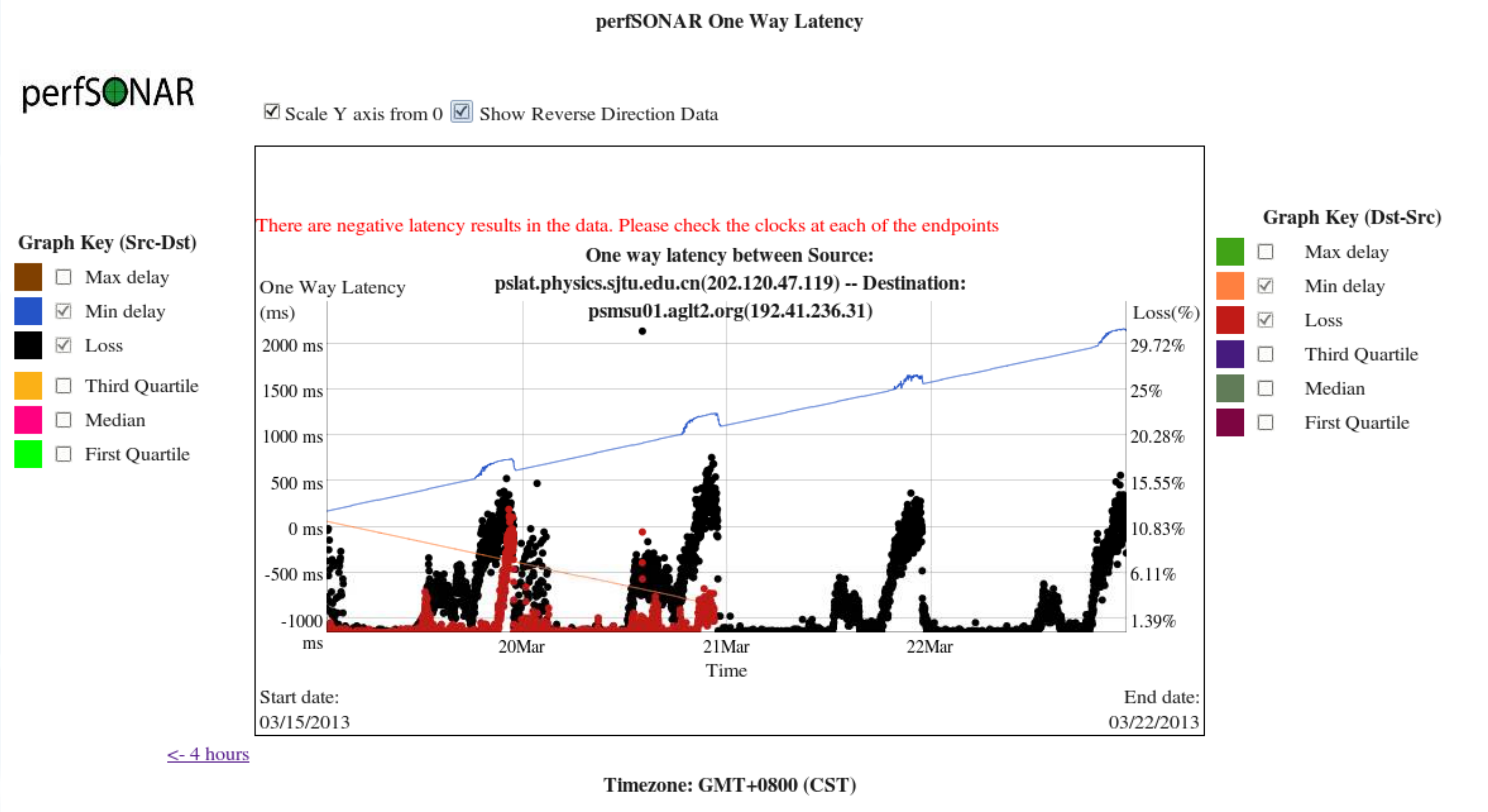


Rebooted router with full route table



Gradual failure of optical line card

Congestion on Link + Drifting Clock



perfSONAR One Way Latency

☑ Scale Y axis from 0 ☑ Show Reverse Direction Data

There are negative latency results in the data. Please check the clocks at each of the endpoints

One way latency between Source: pslat.physics.sjtu.edu.cn(202.120.47.119) -- Destination: psmsu01.aglt2.org(192.41.236.31)

Graph Key (Src-Dst)

- 🟫 ☐ Max delay
- 🟦 ☑ Min delay
- ⬛ ☑ Loss
- 🟧 ☐ Third Quartile
- 🟥 ☐ Median
- 🟩 ☐ First Quartile

Graph Key (Dst-Src)

- 🟩 ☐ Max delay
- 🟧 ☑ Min delay
- 🟥 ☑ Loss
- 🟪 ☐ Third Quartile
- 🟩 ☐ Median
- 🟥 ☐ First Quartile

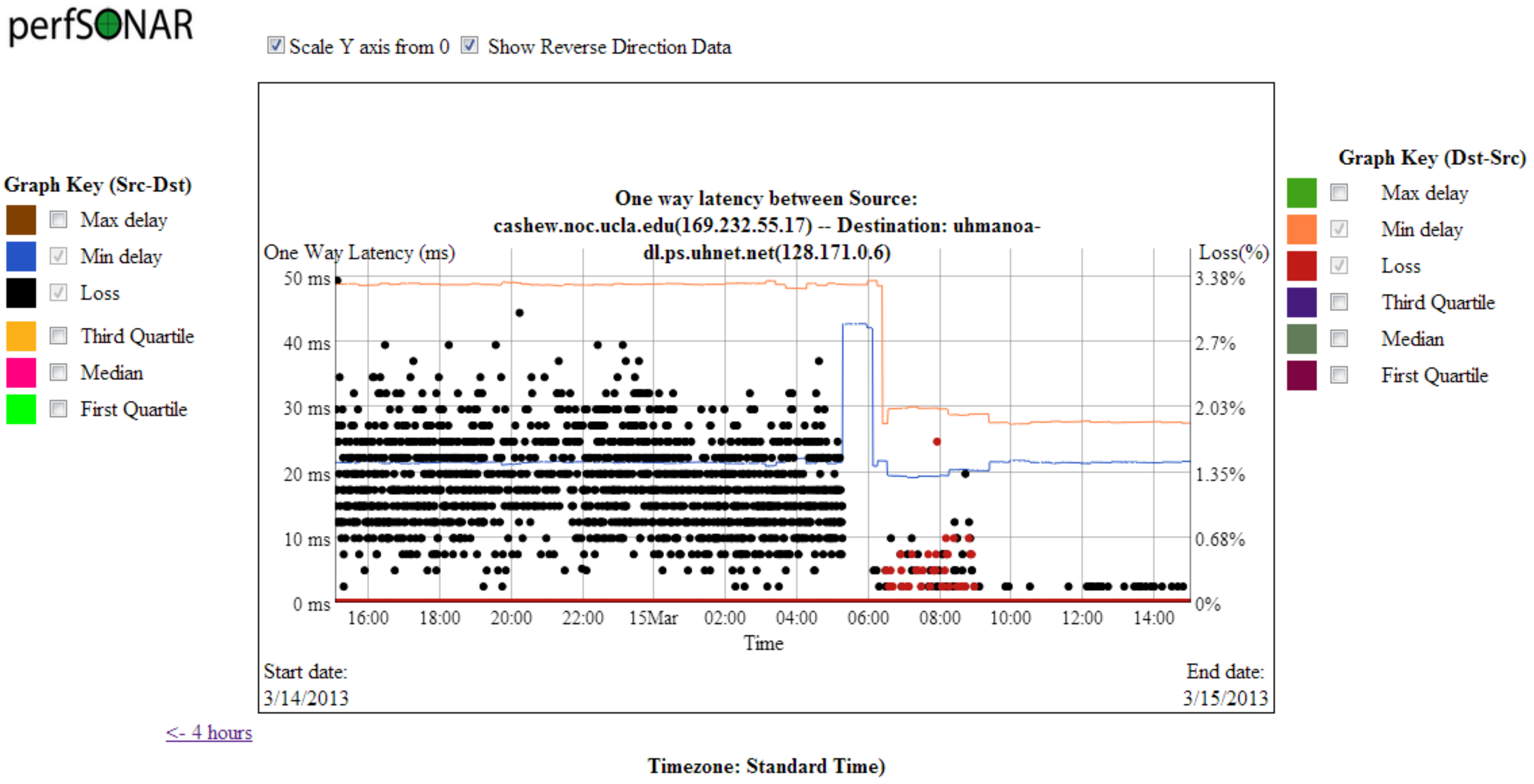
Start date: 03/15/2013

End date: 03/22/2013

<- 4 hours

Timezone: GMT+0800 (CST)

Adding Attenuator to Noisy Link



☑ Scale Y axis from 0 ☑ Show Reverse Direction Data

Graph Key (Src-Dst)

🟫	☐	Max delay
🟦	☑	Min delay
⬛	☑	Loss
🟧	☐	Third Quartile
🟪	☐	Median
🟩	☐	First Quartile

Graph Key (Dst-Src)

🟩	☐	Max delay
🟧	☑	Min delay
🟥	☑	Loss
🟪	☐	Third Quartile
🟩	☐	Median
🟪	☐	First Quartile

Start date: 3/14/2013

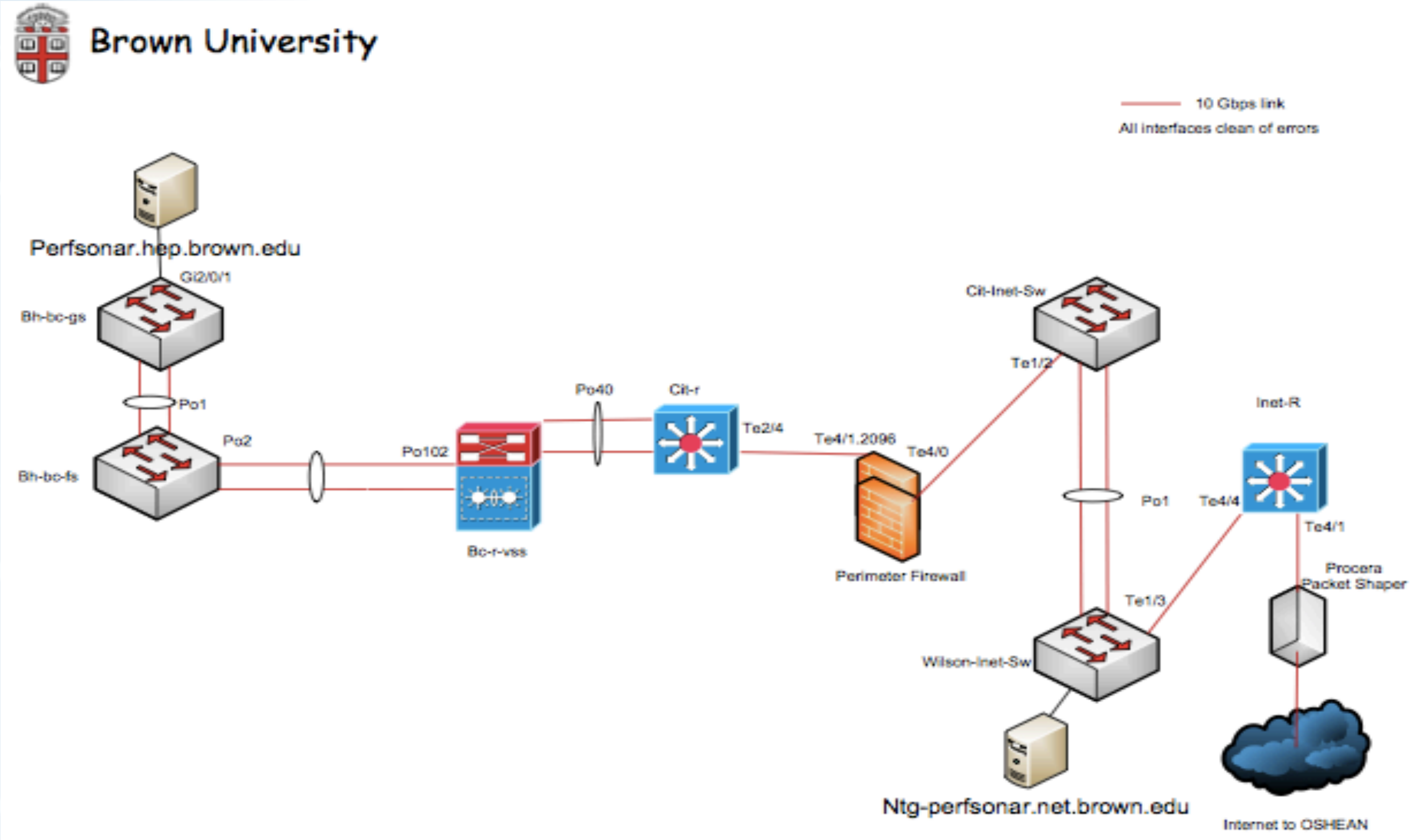
End date: 3/15/2013

<- 4 hours

Timezone: Standard Time)

Firewall Example

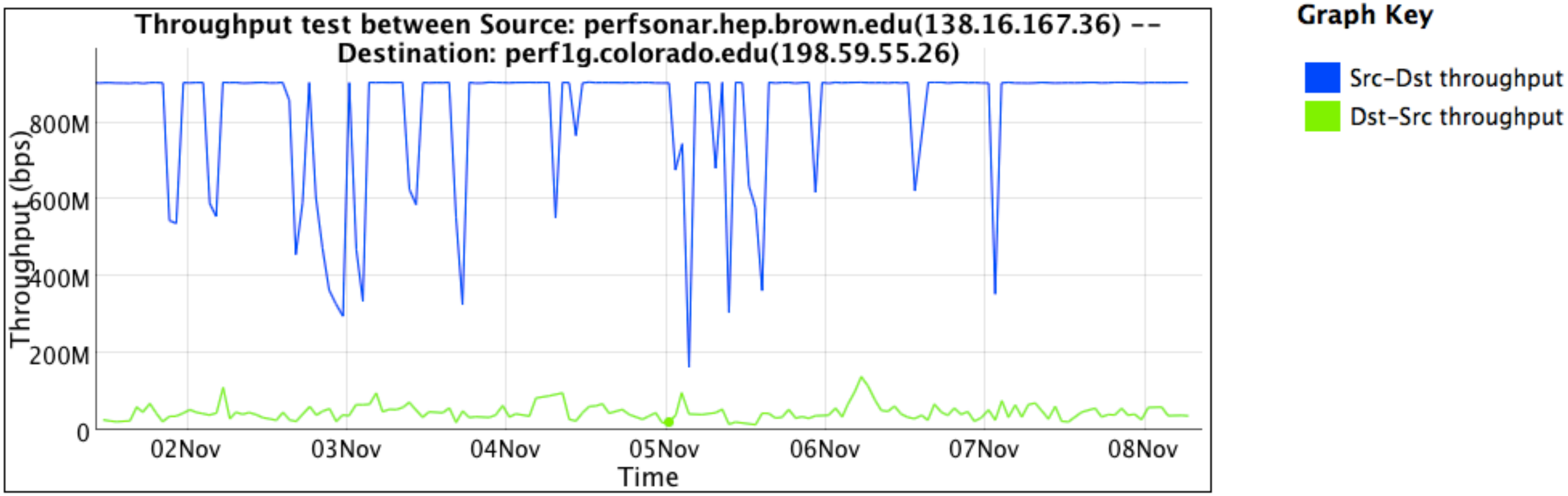
Totally protected campus, with a border firewall



Performance Behind the Firewall

Blue = "Outbound", e.g. campus to remote location upload

Green = "Inbound", e.g. download from remote location



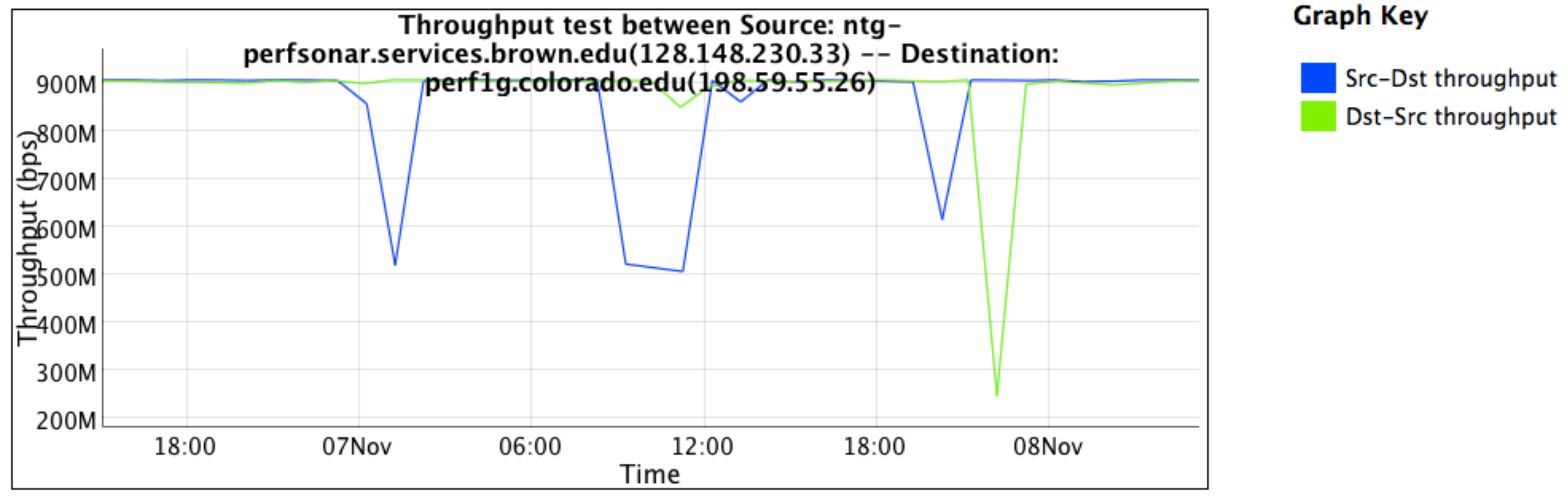
perfSONAR powered

Performance In Front of the Firewall

Blue = "Outbound", e.g. campus to remote location upload

Green = "Inbound", e.g. download from remote location

Note – This machine is in the *SAME RACK*, it just bypasses the firewall vs. that of the previous



Agenda

- Introduction and Purpose
- Science DMZ Overview
- Why Network Performance Matters
- Motivations & Technologies
- Hardware & Deployment Options
- End Results & Debugging
- Conclusions

perfSONAR powered

Motivations & Technologies

All the network diagrams have little perfSONAR boxes everywhere

- The reason for this is that consistent behavior requires correctness
- Correctness requires the ability to find and fix problems
  - ***You can't fix what you can't find***
  - ***You can't find what you can't see***
  - ***perfSONAR lets you see***

Especially important when deploying new technologies like SDN

- If there is a problem with the SDN infrastructure, need to fix it
- If the problem is not with SDN, need to prove it
  - New technology is often assumed to be the source of problems
  - The only way to correctly attribute is to find the problem

perfSONAR powered

Perf-what?

Network Monitoring

- E.g. everyone has some form on their network (e.g. SNMP, NAGIOS, etc.). Addresses the needs of local staff for knowing what is going on
  - Would this information be useful to external users?
  - Are tools such as CACTI really able to function on a multi-domain basis?
- Beyond passive methods, there are active tools.
  - E.g. Iperf can be run to get a 'throughput' number. Do we store these anywhere?
  - Wouldn't it be nice to get some sort of plot of performance over the course of a day? Week? Year? Multiple endpoints?

perfSONAR = Measurement Middleware

perfSONAR powered

What is perfSONAR?

perfSONAR is a tool to:

- Set network performance expectations
- Find network problems ("soft failures")
- Help fix these problems

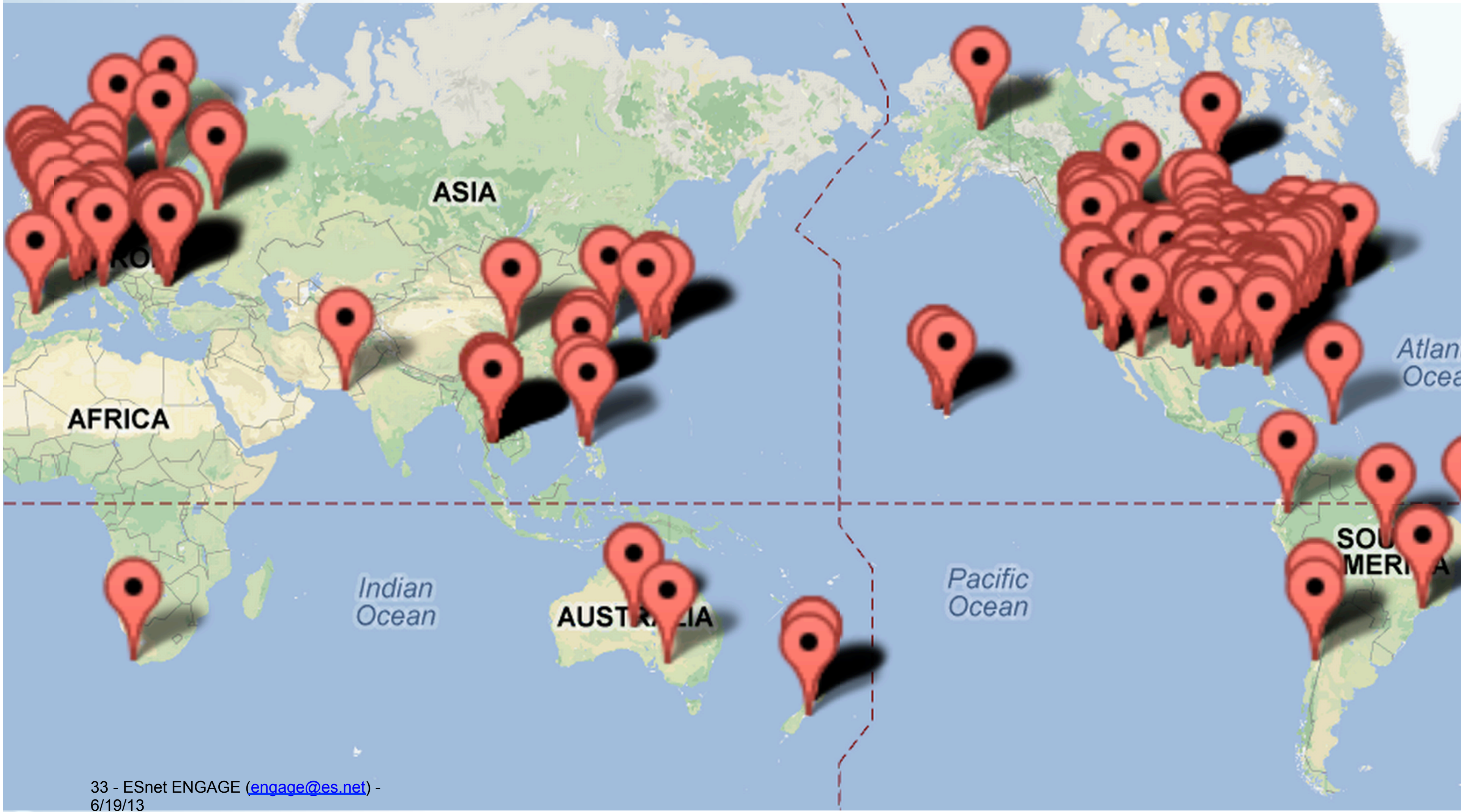
All in multi-domain environments

- These problems are all harder when multiple networks are involved

perfSONAR is provides a standard way to publish active and passive monitoring data

- This data is interesting to network researchers as well as network operators

World-Wide perfSONAR-PS Deployments:
572 bwctl nodes, 552 owamp nodes as of Jun ‘13



perfSONAR-PS Software

perfSONAR-PS is an open source implementation of the perfSONAR measurement infrastructure and protocols

- written in the perl programming language

http://psps.perfsonar.net

All products are available as RPMs.

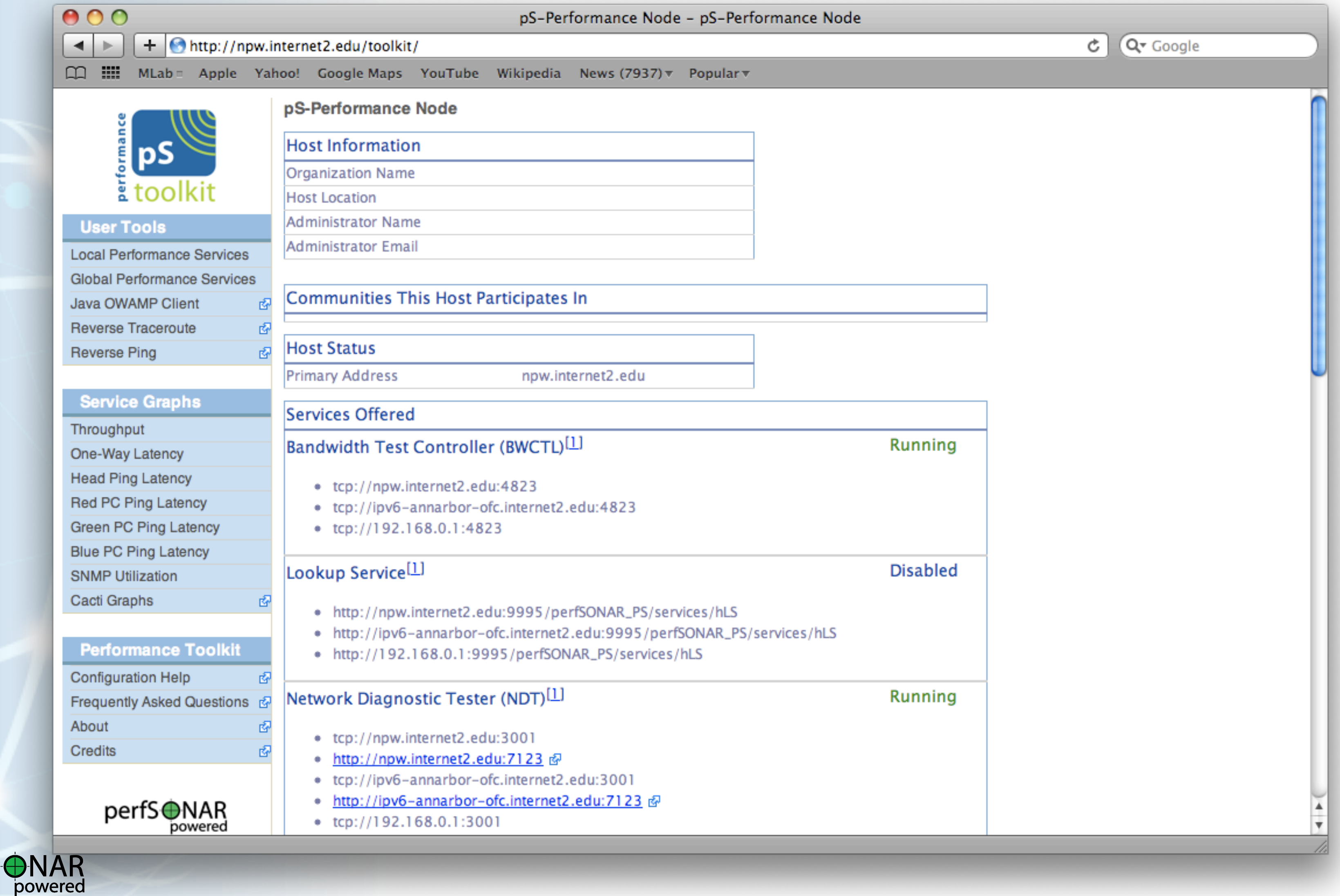
The perfSONAR-PS consortium supports CentOS (versions 5 and 6).

RPMs are compiled for i386 and x86 64

Functionality on other platforms and architectures is possible, but not supported.

- Should work: Red Hat Enterprise Linux and Scientific Linux (v5)
- Harder, but possible:
  - Fedora Linux, SuSE, Debian Variants

Toolkit Display



User Tools

- Local Performance Services
- Global Performance Services
- Java OWAMP Client
- Reverse Traceroute
- Reverse Ping

Service Graphs

- Throughput
- One-Way Latency
- Head Ping Latency
- Red PC Ping Latency
- Green PC Ping Latency
- Blue PC Ping Latency
- SNMP Utilization
- Cacti Graphs

Performance Toolkit

- Configuration Help
- Frequently Asked Questions
- About
- Credits

perfSONAR powered

pS-Performance Node

Host Information	
Organization Name	
Host Location	
Administrator Name	
Administrator Email	

Communities This Host Participates In

Host Status	
Primary Address	npw.internet2.edu

Services Offered

Bandwidth Test Controller (BWCTL)[1] Running

- tcp://npw.internet2.edu:4823
- tcp://ipv6-annarbor-ofc.internet2.edu:4823
- tcp://192.168.0.1:4823

Lookup Service[1] Disabled

- http://npw.internet2.edu:9995/perfSONAR_PS/services/hLS
- http://ipv6-annarbor-ofc.internet2.edu:9995/perfSONAR_PS/services/hLS
- http://192.168.0.1:9995/perfSONAR_PS/services/hLS

Network Diagnostic Tester (NDT)[1] Running

- tcp://npw.internet2.edu:3001
- http://npw.internet2.edu:7123
- tcp://ipv6-annarbor-ofc.internet2.edu:3001
- http://ipv6-annarbor-ofc.internet2.edu:7123
- tcp://192.168.0.1:3001

The Metrics

Use the correct tool for the Job

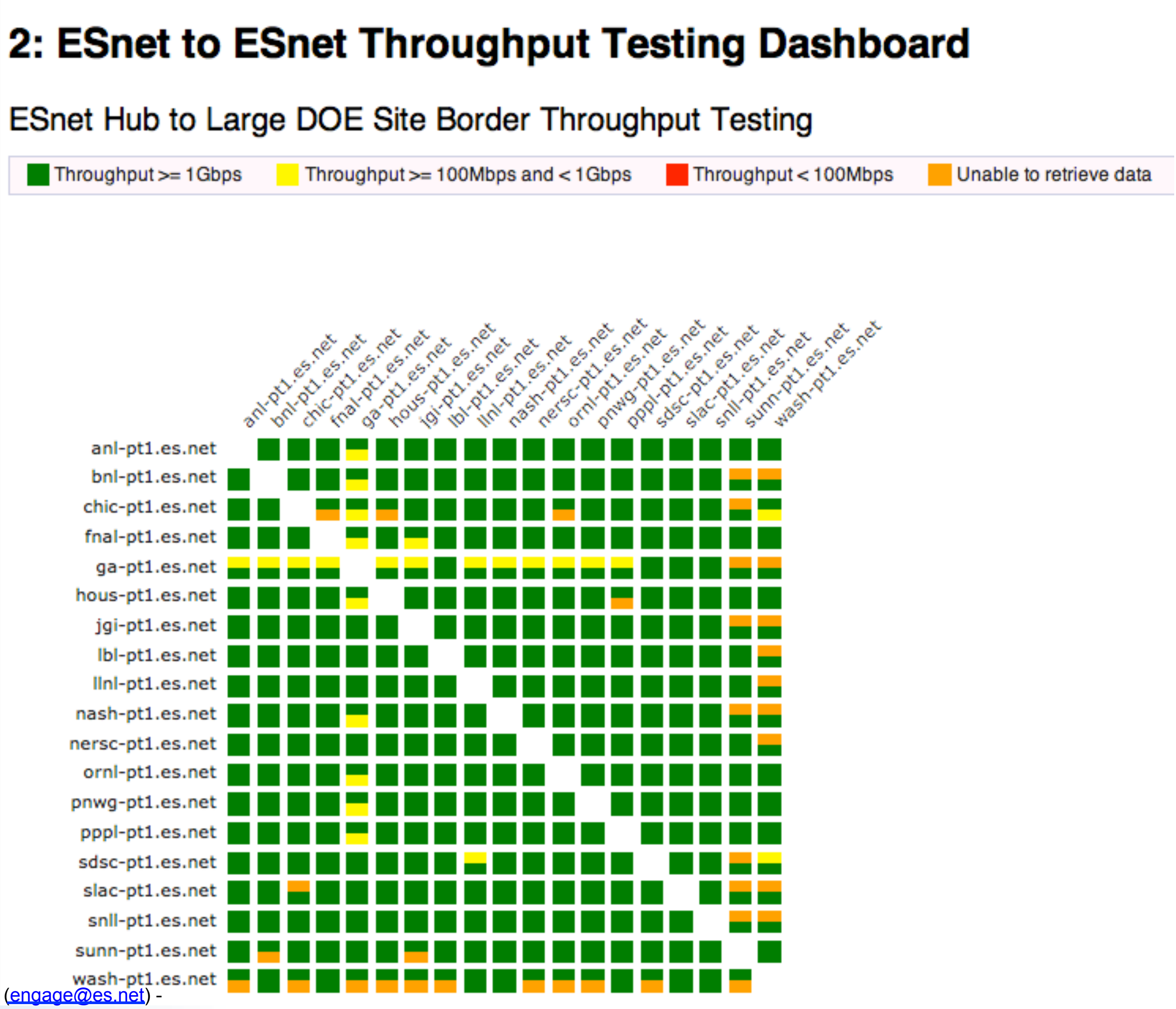
- To determine the correct tool, maybe we need to start with what we want to accomplish …

What do we care about measuring?

- Packet Loss, Duplication, out-of-orderness (transport layer)
- Achievable Bandwidth (e.g. "Throughput")
- Latency (Round Trip and One Way)
- Jitter (Delay variation)
- Interface Utilization/Discards/Errors (network layer)
- Traveled Route
- MTU Feedback

perfSONAR powered

perfSONAR Dashboard (http://ps-dashboard.es.net)



2: ESnet to ESnet Throughput Testing Dashboard

ESnet Hub to Large DOE Site Border Throughput Testing

🟩 Throughput >= 1Gbps 🟨 Throughput >= 100Mbps and < 1Gbps 🟥 Throughput < 100Mbps 🟧 Unable to retrieve data

	anl-pt1.es.net	bnl-pt1.es.net	chic-pt1.es.net	fnal-pt1.es.net	ga-pt1.es.net	hous-pt1.es.net	jgi-pt1.es.net	lbl-pt1.es.net	llnl-pt1.es.net	nash-pt1.es.net	nersc-pt1.es.net	ornl-pt1.es.net	pnwg-pt1.es.net	pppl-pt1.es.net	sdsc-pt1.es.net	slac-pt1.es.net	snll-pt1.es.net	sunn-pt1.es.net	wash-pt1.es.net
anl-pt1.es.net		🟩	🟩	🟩	🟩🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
bnl-pt1.es.net	🟩		🟩	🟩	🟩🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟧🟩	🟧🟩
chic-pt1.es.net	🟩	🟩		🟩🟧	🟩🟨	🟩🟧	🟩	🟩	🟩	🟩	🟩	🟩🟧	🟩	🟩	🟩	🟩	🟩	🟧🟩	🟩🟨
fnal-pt1.es.net	🟩	🟩	🟩		🟩🟨	🟩	🟩🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
ga-pt1.es.net	🟨🟩	🟨🟩	🟨🟩	🟨🟩		🟨🟩	🟨🟩	🟩	🟨🟩	🟨🟩	🟨🟩	🟨🟩	🟨🟩	🟨🟩	🟩	🟩	🟩	🟧🟩	🟧🟩
hous-pt1.es.net	🟩	🟩	🟩	🟩	🟩🟨		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩🟧	🟩	🟩	🟩	🟩	🟩
jgi-pt1.es.net	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟧🟩	🟧🟩
lbl-pt1.es.net	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟧🟩
llnl-pt1.es.net	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟧🟩
nash-pt1.es.net	🟩	🟩	🟩	🟩	🟩🟨	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟧🟩	🟧🟩
nersc-pt1.es.net	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟧🟩
ornl-pt1.es.net	🟩	🟩	🟩	🟩	🟩🟨	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩
pnwg-pt1.es.net	🟩	🟩	🟩	🟩	🟩🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩
pppl-pt1.es.net	🟩	🟩	🟩	🟩	🟩🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩
sdsc-pt1.es.net	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟨🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟧🟩	🟨🟩
slac-pt1.es.net	🟩	🟩	🟧🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟧🟩	🟧🟩
snll-pt1.es.net	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟧🟩	🟧🟩
sunn-pt1.es.net	🟩	🟩🟧	🟩	🟩	🟩	🟩	🟩🟧	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩
wash-pt1.es.net	🟩🟧	🟩	🟩🟧	🟩	🟩🟧	🟩🟧	🟩🟧	🟩🟧	🟩	🟩	🟩🟧	🟩🟧	🟩🟧	🟩	🟩🟧	🟩	🟩	🟩🟧	

perfSONAR powered

US Deployment

Internet2

- 4 Machines in each PoP on the current network (2 x Throughput Test Machine, 1 User Test Machine, 1 Latency Test Machine)
- Plans for single server in all PoPs on new network
- Internal Testing (http://owamp.net.internet2.edu), and 100s of community initiated tests per week
- Central Netflow/SNMP Monitoring
- Assistance available – rs@internet2.edu

ESnet

- 2 Machines in each PoP (Latency and Bandwidth Testing)
- Machines at Customer sites (e.g. federal labs and other scientific points of interest)
- Full mesh of testing (http://stats.es.net)
- Assistance available – trouble@es.net

perfSONAR powered

US Regional Networks

- ***Regionals with, or acquiring, perfSONAR:***
  - 3ROX
  - ARE-ON
  - CEN
  - CENIC
  - CIC
  - FLR
  - SOX
  - FRGP
  - GPN
  - KanREN
  - LEARN
  - LONI
  - MAGPI
  - MARIA
  - MAX
  - MCNC
  - MERIT
  - MissiON
  - MOREnet
  - MREN
- ***Regionals with, or acquiring, perfSONAR (cont):***
  - WVNET
  - NJEDGE
  - NOX
  - NYSERNET
  - OneNet
  - Oregon GigaPoP
  - PNWGP
  - PeachNet
  - UEN
  - WiscNet
- ***Regionals with unsure status:***
  - ABQG
  - C-Light
  - Indiana GigaPoP
  - IRON
  - KyRON
  - MDREN
  - Northern Lights
  - OARnet
  - OSHEAN

N.B. These represent people I have talked with, if there are errors just let me know

perfSONAR powered

Agenda

- Introduction and Purpose
- Science DMZ Overview
- Why Network Performance Matters
- Motivations & Technologies
- Hardware & Deployment Options
- End Results & Debugging
- Conclusions

perfSONAR powered

perfSONAR Deployment Locations

Critical to deploy such that you can test with useful semantics

perfSONAR hosts allow parts of the path to be tested separately

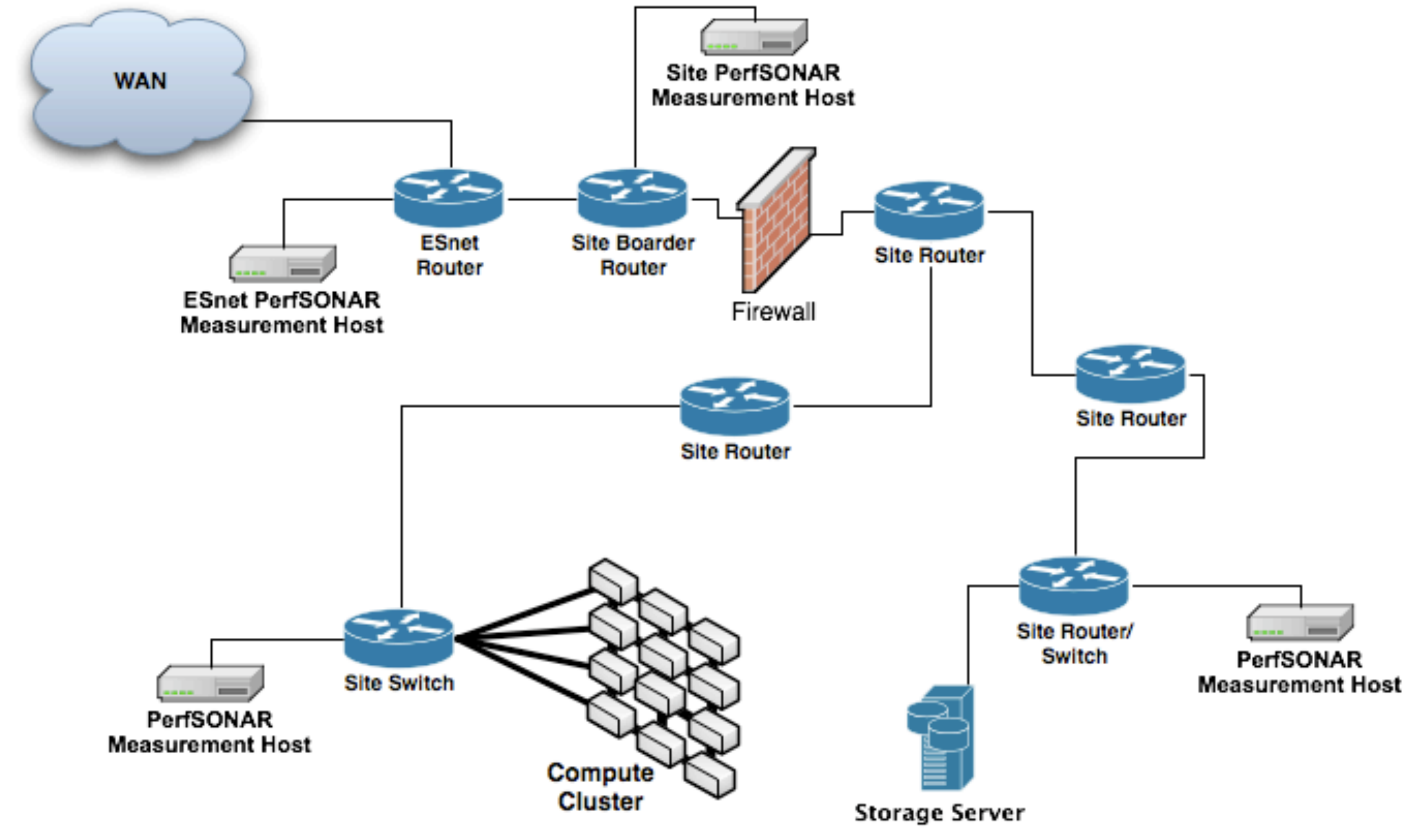
- Reduced visibility for devices between perfSONAR hosts
- Rely on counters or other means where perfSONAR can't go

Effective test methodology derived from protocol behavior

- TCP suffers much more from packet loss as latency increases
- TCP is more likely to cause loss as latency increases
- Testing should leverage this in two ways
  - Design tests so that they are likely to fail if there is a problem
  - Mimic the behavior of production traffic as much as possible
- Note: don't design your tests to succeed – it is ***not*** helpful

perfSONAR powered

Sample Site Deployment



perfSONAR powered

Importance of Regular Testing

You can't wait for users to report problems and then fix them (soft failures can go unreported for years!)

Things just break sometimes

- Failing optics
- Somebody messed around in a patch panel and kinked a fiber
- Hardware goes bad

Problems that get fixed have a way of coming back

- System defaults come back after hardware/software upgrades
- New employees may not know why the previous employee set things up a certain way and back out fixes

Important to continually collect, archive, and alert on active throughput test results

perfSONAR powered

Develop a Plan

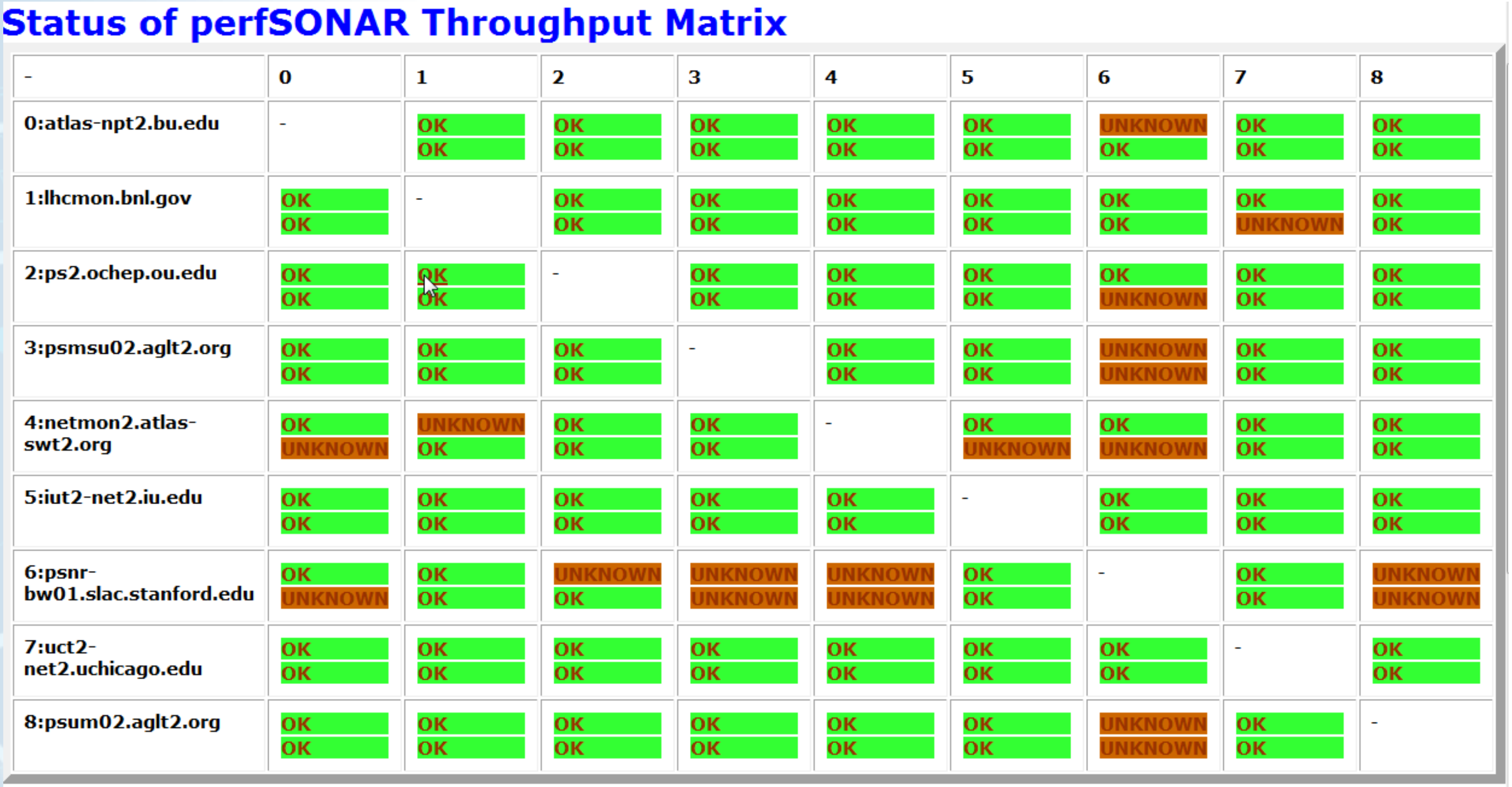
What are you going to measure?

- Achievable bandwidth
  - 2-3 regional destinations
  - 4-8 important collaborators
  - 4-10 times per day to each destination
  - 20 second tests within a region, longer across oceans and continents
- Loss/Availability/Latency
  - OWAMP: ~10 collaborators over diverse paths
  - PingER: use to monitor paths to collaborators who don't support owamp
- Interface Utilization & Errors

What are you going to do with the results?

- NAGIOS Alerts
- Reports to user community
- Post to Website

ATLAS Dashboard



Status of perfSONAR Throughput Matrix

-	0	1	2	3	4	5	6	7	8
0:atlas-npt2.bu.edu	-	OK OK	OK OK	OK OK	OK OK	OK OK	UNKNOWN OK	OK OK	OK OK
1:lhcmon.bnl.gov	OK OK	-	OK OK	OK OK	OK OK	OK OK	OK OK	OK UNKNOWN	OK OK
2:ps2.ochep.ou.edu	OK OK	OK OK	-	OK OK	OK OK	OK OK	OK UNKNOWN	OK OK	OK OK
3:psmsu02.aglt2.org	OK OK	OK OK	OK OK	-	OK OK	OK OK	UNKNOWN UNKNOWN	OK OK	OK OK
4:netmon2.atlas-swt2.org	OK UNKNOWN	UNKNOWN OK	OK OK	OK OK	-	OK UNKNOWN	OK UNKNOWN	OK OK	OK OK
5:iut2-net2.iu.edu	OK OK	OK OK	OK OK	OK OK	OK OK	-	OK OK	OK OK	OK OK
6:psnr-bw01.slac.stanford.edu	OK UNKNOWN	OK OK	UNKNOWN OK	UNKNOWN UNKNOWN	UNKNOWN UNKNOWN	OK OK	-	OK OK	UNKNOWN UNKNOWN
7:uct2-net2.uchicago.edu	OK OK	OK OK	OK OK	OK OK	OK OK	OK OK	OK OK	-	OK OK
8:psum02.aglt2.org	OK OK	OK OK	OK OK	OK OK	OK OK	OK OK	UNKNOWN UNKNOWN	OK OK	-

Host Considerations

http://psps.perfsonar.net/toolkit/hardware.html

Dedicated perfSONAR hardware is best

- Server class is a good choice
- Desktop/Laptop/Mini (Mac, Shuttle) can be problematic, but work in a diagnostic capacity

Other applications will perturb results

Separate hosts for throughput tests and latency/loss tests is preferred

- Throughput tests can cause increased latency and loss
- Latency tests on a throughput host are still useful however

1Gbps vs 10Gbps testers

- There are a number of problem that only show up at speeds above 1Gbps

Virtual Machines do not always work well as perfSONAR hosts (use specific)

- Clock sync issues are a bit of a factor
- throughput is reduced significantly for 10G hosts
- VM technology and motherboard technology has come a long way, YMMV
- NDT/NAGIOS/SNMP/1G BWCTL are good choices for a VM, OWAMP/10G BWCTL are not

perfSONAR powered

Agenda

- Introduction and Purpose
- Science DMZ Overview
- Why Network Performance Matters
- Motivations & Technologies
- Hardware & Deployment Options
- End Results & Debugging
- Conclusions

Common Use Case

Trouble ticket comes in:

> "I'm getting terrible performance from site A to site B"

If there is a perfSONAR node at each site border:

- Run tests between perfSONAR nodes
  - performance is often clean
- Run tests from end hosts to perfSONAR host at site border
  - Often find packet loss (using owamp tool)
  - If not, problem is often the host tuning or the disk

If there is not a perfSONAR node at each site border

- Try to get one deployed
- Run tests to other nearby perfSONAR nodes

perfSONAR powered

WAN Test Methodology – Problem Isolation

Segment-to-segment testing is unlikely to be helpful

- TCP dynamics will be different
- Problem links can test clean over short distances
- This also goes for testing from a Science DMZ to the border router or first provider perfSONAR host

Run long-distance tests

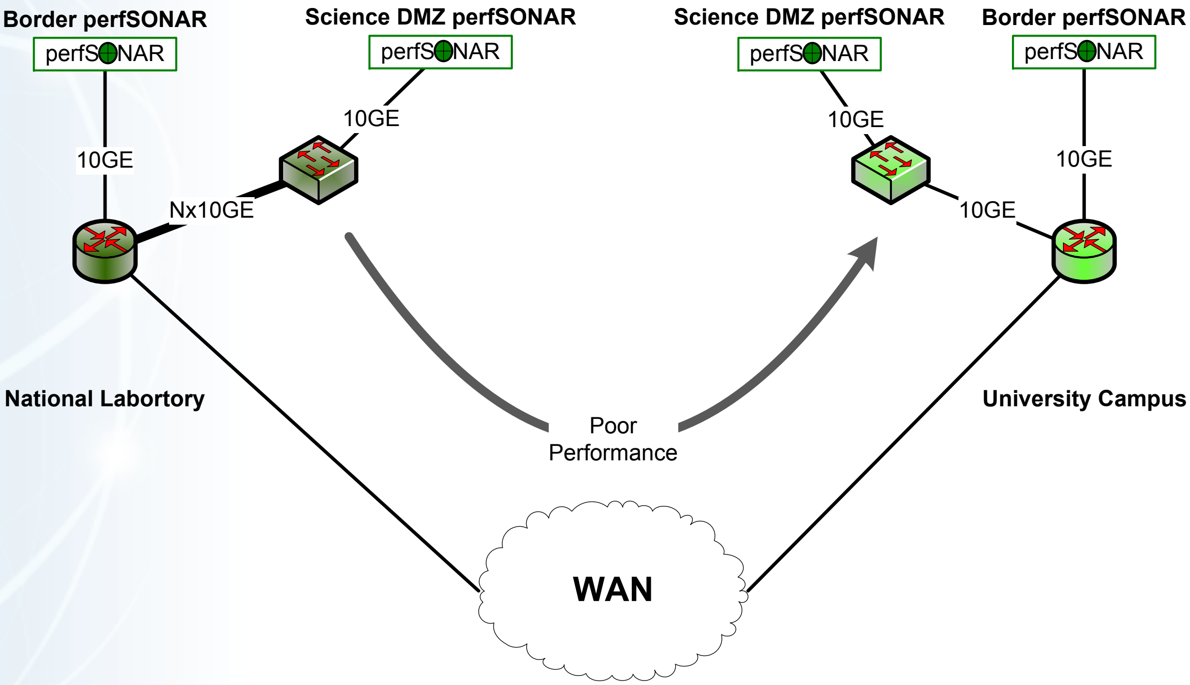
- Run the longest clean test you can, then look for the shortest dirty test that includes the path of the clean test
- In many cases, there is a problem between the two remote test locations

In order for this to work, the testers need to be already deployed when you start troubleshooting

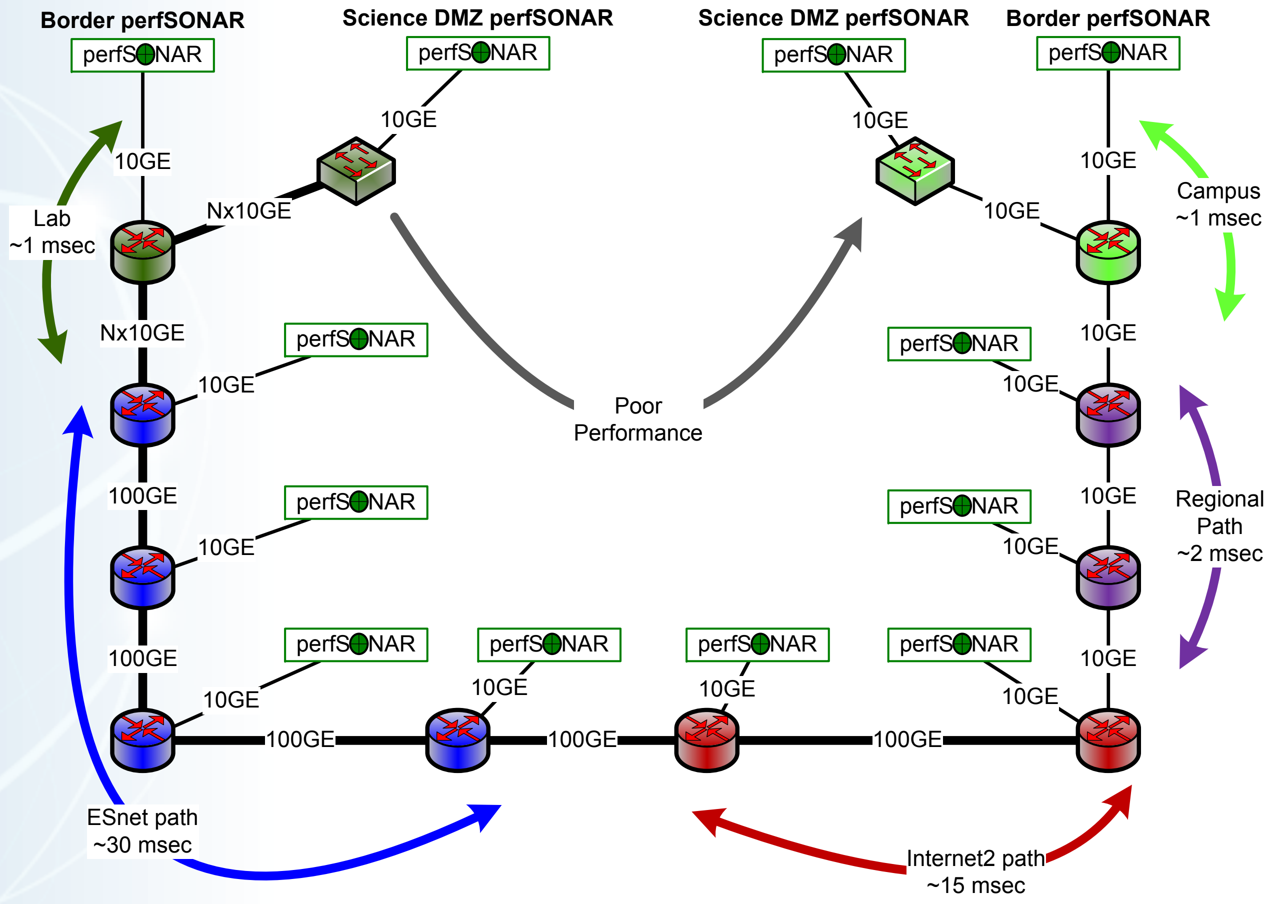
- ESnet has at least one perfSONAR host at each hub location. So does Internet2. So do many regionals.
- If your provider does not have perfSONAR deployed ask them why, and then ask when they will have it done

perfSONAR powered

Wide Area Testing – Problem Statement

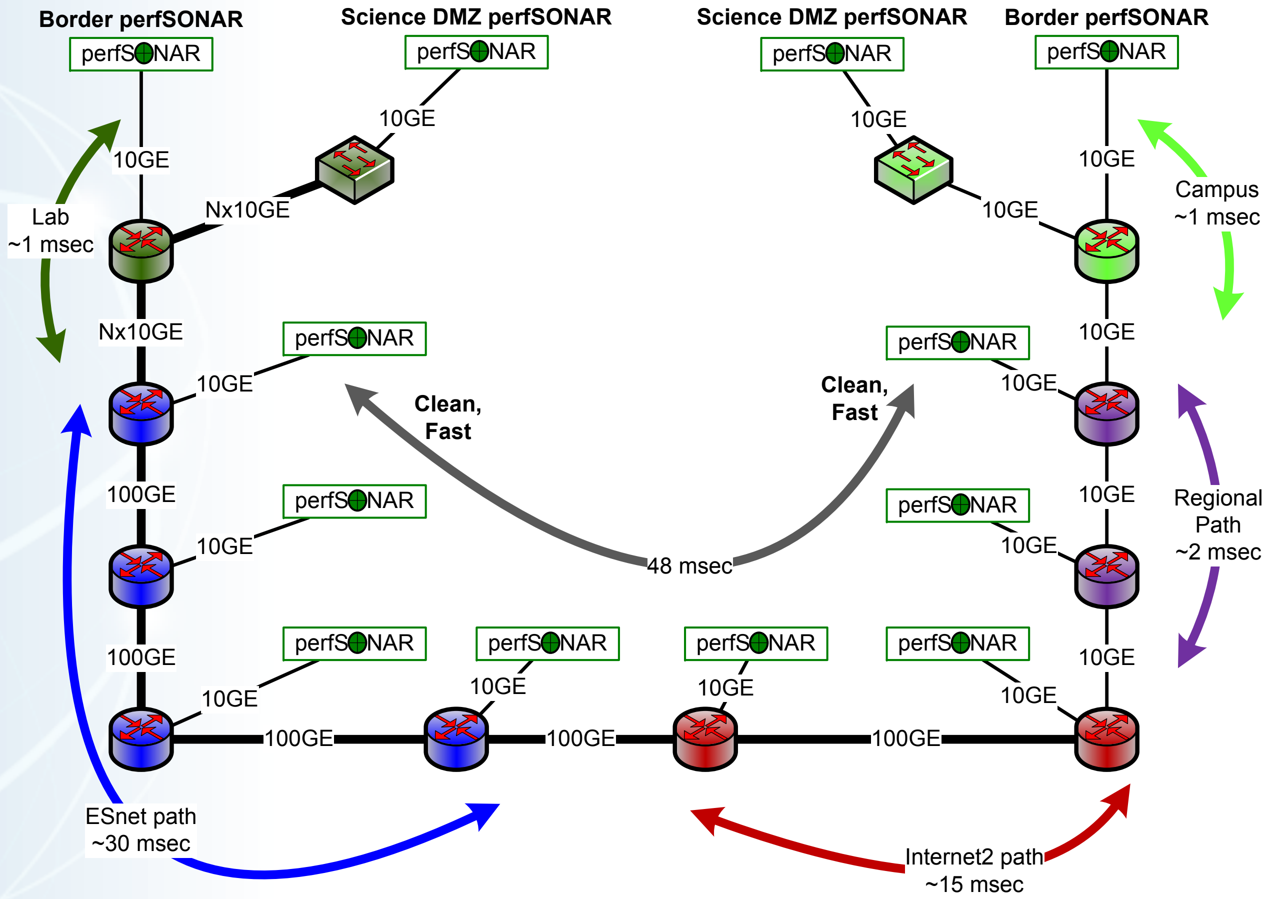


Wide Area Testing – Full Context



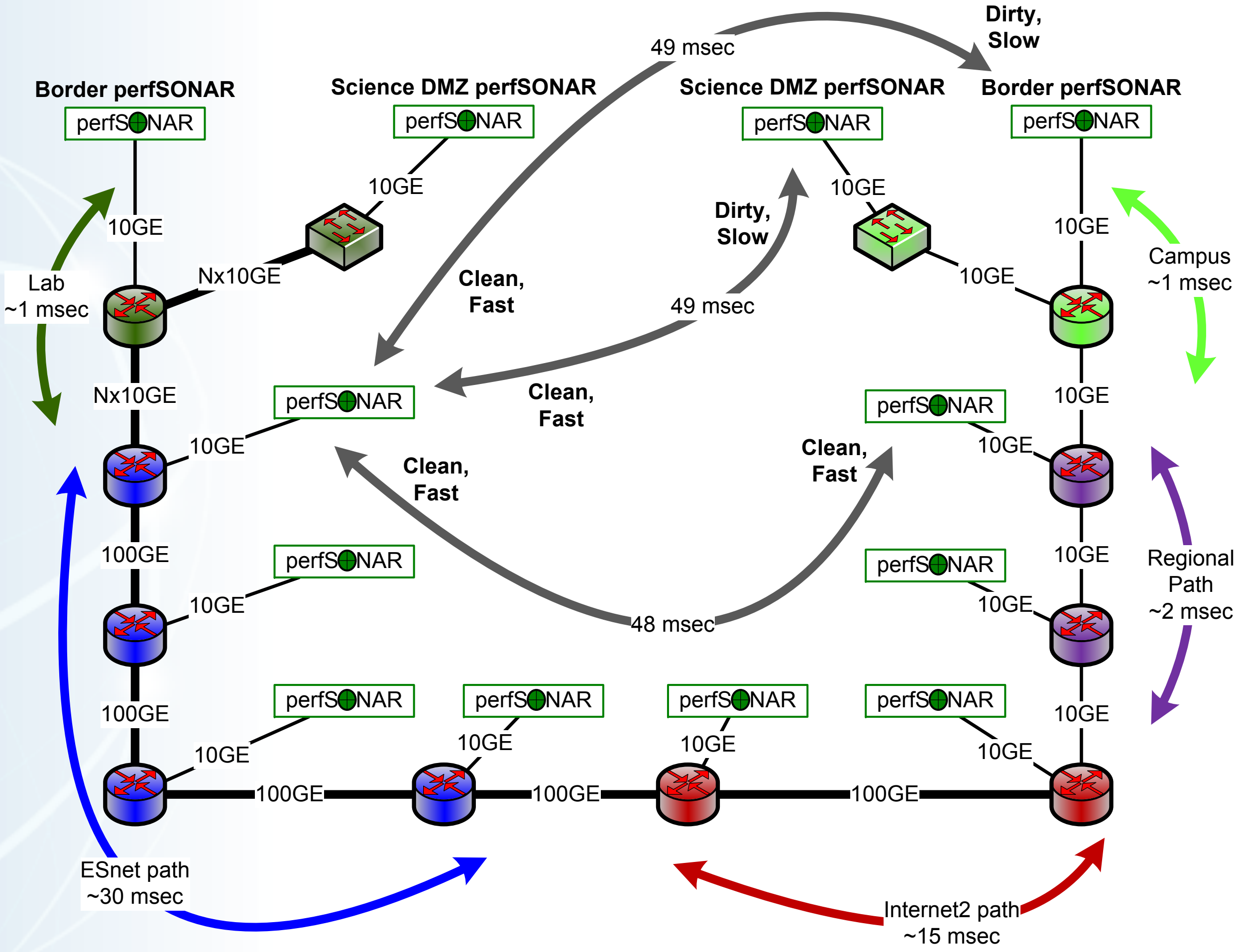
perfSONAR powered

Wide Area Testing – Long Clean Test



perfSONAR powered

Wide Area Testing – Poorly Performing Tests Illustrate Likely Problem Areas



perfSONAR powered

Lessons From The Test Case

This testing can be done quickly if perfSONAR is already deployed

Huge productivity

- Reasonable hypothesis developed quickly
- Probable administrative domain identified
- Testing time can be short – an hour or so at most

Without perfSONAR cases like this are very challenging

Time to resolution measured in months

In order to be useful for data-intensive science, the network must be fixable quickly, because it ***will*** break

The Science DMZ model allows high-performance use of the network, but perfSONAR is necessary to ensure the whole kit functions well

perfSONAR powered

Agenda

- Introduction and Purpose
- Science DMZ Overview
- Why Network Performance Matters
- Motivations & Technologies
- Hardware & Deployment Options
- End Results & Debugging
- Conclusions

Conclusions

- Lots of information today:
  - Why performance matters
  - How to implement the network to ensure success
  - How to use software to guarantee it
- Next steps:
  - Growing community – please join it!
  - Will other webinars make sense? Do your local tech folks want:
    - A more in depth (1 hr plus) tutorial on how to configure and interpret perfSONAR?
    - Installing and using Globus Online?
    - Others?

perfSONAR powered

perfSONAR Community

perfSONAR-PS is working to build a strong user community to support the use and development of the software.

perfSONAR-PS Mailing Lists

- Announcement List:
  - https://mail.internet2.edu/wws/subrequest/perfsonar-ps-announce
  - https://mail.internet2.edu/wws/subrequest/performance-node-announce
- Users List:
  - https://mail.internet2.edu/wws/subrequest/performance-node-users

Science DMZ Community

In addition to perfSONAR, the Science DMZ community is growing as well. We would encourage everyone to join the conversation as you implement your networks:

- General Info:
  - http://fasterdata.es.net/science-dmz/
- Mailing List
  - https://listserv.es.net/mailman/listinfo/sciencedmz
- Forums:
  - http://fasterdata.es.net/forums/

perfSONAR powered

http://fasterdata.es.net

ESnet maintains a “knowledge base” of tips and tricks for obtaining maximum WAN throughput

Lots of useful stuff there, including:

- TCP tuning information (in cut and paste friendly form)
- Data Transfer Node (DTN) tuning information
  - Also in cut and paste friendly form
- DTN reference designs
- Science DMZ information
- perfSONAR information

perfSONAR powered

Introduction to perfSONAR for NSF CC-NIE Awardees

Thanks!

Jason Zurawski - zurawski@es.net

Eli Dart - dart@es.net

Mary Hester – mhester@es.net

Lauren Rotman - lauren@es.net

Brian Tierneney - bltierney@es.net

http://psps.perfsonar.net

http://fasterdata.es.net