



# **OSG** **CAMPUS** INFRASTRUCTURES SERIES

## **Methods for Network Troubleshooting in Distributed High Throughput Environments --- Webinar**

**Shawn McKee**

OSG Networking / University of Michigan Physics

January 25, 2013

# Presentation Abstract

**ABSTRACT:** Often in grid and distributed campus environments, users will notice problems with "the network". This takes many forms including an inability to reach remote sites, very poor application performance and/or very slow data transfer rates. There are a number of challenges with "network" problems, not the least of which is verifying it is really a network problem. Debugging network problems and isolating the location of the problem can be very difficult and time-consuming. The perfSONAR project was designed to help aid in the diagnosis and location of network problems. OSG is working on providing a robust installation of the perfSONAR-PS toolkit and services to configure and gather network metrics from OSG sites with perfSONAR-PS. In this talk we address the details of perfSONAR-PS: how to install, configure and use it to help better understand issues in the network. The presentation will include practical information about using the perfSONAR-PS toolkit installation along with relevant URLs for further information.

# Webinar Overview

- Overview: OSG, Campus grids and “network problems”
  - Understanding the overall system
  - Categorizing problems and localizing them
- perfSONAR-PS is your friend: What it is and why it can help
- Details about perfSONAR-PS
  - Installation and configuration
  - Using perfSONAR-PS when there are problems
- Walkthru of a documenting a problem
- Summary and Resources for Going Further

# OSG Network Monitoring

- Distributed collaborations rely upon the network as a critical part of their infrastructure, yet finding and debugging network problems can be difficult and, in some cases, take months.
- There is typically no differentiation of how the network is used amongst the OSG users.  
(Quantity may vary)
- **We need a standardized way to monitor the network and locate problems quickly if they arise**
- We don't want to have a network monitoring system per VO!

# Data Movement for Science

This should not be news to anyone here ...

Flows getting larger (e.g. *Science datasets* in the R&E world)

- Special requirements (e.g. *Streaming media* is sensitive to jitter, *bulk data transfer* is sensitive to loss)
- Number of users/devices is increasing
- Locations are spread out
- ***Everything*** is cross domain



(for Spinal Tap fans...)

Slide from Jason Zurawski

# OSG, Campus grids and “network problems”

- The Open Science Grid (OSG) advances science through open distributed computing.
  - Underlying OSG is a robust network, tying together distributed compute and storage resources
- Grids in general also assume and require a ubiquitous and high-performing network interconnecting all relevant resources.
- What happens when things don't work as planned?
  - In many cases users and local administrators face a daunting and potentially LONG period of diagnosing, debugging and waiting....
  - The “network” is often blamed...but is it the issue?
  - If it IS the network...which part? Where? What?



# Network Realities

- Where are the problems?

*Network Core?* Everything is well connected, well provisioned, and flawlessly configured, **RIGHT?**

*End Systems?*

Properly tuned for optimal TCP performance (no matter the operating system), proper drivers installed and functioning optimally, **RIGHT?**

*LAN? Regional Net?*

Better to ask “Where aren’t there problems?”



Slide from Jason Zurawski

# Understanding Distributed Infrastructures

- As noted, when **problems** involving distributed resources occur the first thing to get the blame is “the Network”...sometimes that is even correct!
- **Problems** span a wide range of symptoms:
  - Loss of connectivity
  - Poor utilization of the network
  - Intermittent connectivity or widely varying performance
  - Application errors or misbehavior
- Grid systems and campus infrastructures are complex and involve many “moving pieces”
- Network diagnosis is non-trivial for many reasons, not the least of which is that networks span multiple administrative domains and no single entity usually has visibility into more than their part of the path.
- **The critical issue is to understand, even approximately, where the problem lies.**



# The Possible “Where”

- When problems arise in a distributed system there are many possible sources since they can be **anywhere, end-to-end**
- First step is to understand if it is the **network** or the **end-systems**.
  - This can be *subtle*. Local network issues may not have a significant impact on local network flows but can cripple Wide Area Network (WAN) flows.
  - Likewise poorly tuned hosts or ones with faulty hardware may perform OK for local network transfers but very badly on WAN connections

# What to do?

- Users, local administrators and network engineers are faced with the daunting issue of trying to track down a problem where they have limited information and access to the full end-to-end path.
- In general we need to divide-and-conquer
  - Examine end-hosts for problems and misconfigurations
  - Measure network path behavior

# End-Hosts and Applications

- Many end-to-end problems are actually issues at the ends.
- The good news is users and admins usually have access-to, and experience-with, end-systems.
  - Important to understand network host tuning
  - Application may not be designed properly for WAN network use (poor buffering, many small, latency-sensitive transactions, etc.)

# End-Host / Application Checklist

- This could be a whole other Webinar 😊
- Check for errors and faulty hardware
- Document/understand how the host connects to the Local Area Network(LAN)
- Use 'ifconfig' and 'ethtool' to gather network and network interface info
  - ethtool -S eth0 (or ethN) displays NIC stats
- Has the application in question worked before? Differences in use?
- See 'ESnet page <http://fasterdata.es.net/host-tuning/>

# Need for a “Finger Pointing” Tool

- Once we have checked the end-host setup and tuned appropriately, what if the problem persists?
- How can we quickly identify when problems are network problems and help isolate their locations?
- The perfSONAR project was designed to help do this

# Network Monitoring Motivation

- Finding a solution to network performance problems can be broken into two distinct steps:
  - Use of *Diagnostic Tools* to locate problems
    - Tools that actively measure performance (e.g. Latency, Available Bandwidth)
    - Tools that passively observe performance (e.g. error counters)
  - *Regular Monitoring* to establish performance baselines and alert when expectation drops.
    - Using diagnostic tools in a structured manner
    - Visualizations and alarms to analyze the collected data
- Incorporation of either of these techniques must be:
  - *ubiquitous*, e.g. the solution works best when it is available everywhere
  - *seamless* (e.g. *federated*) in presenting information from different resources and domains

Slide from Jason Zurawski



# History of perfSONAR

- perfSONAR: a joint effort of **ESnet**, **Internet2**, **GEANT** and **RNP** (and others) to standardize network monitoring protocols, schema and tools
- perfSONAR as a project provides a common way to define network metrics the makes implementations able to share data and inter-operate
  - Many Research and Education networks worldwide have adopted perfSONAR as the standard way to gather network metrics
- One example implementation from **Internet2** and **ESnet** is the perfSONAR-PS Toolkit and that is what I will discuss in this Webinar

# Example perfSONAR Use Case

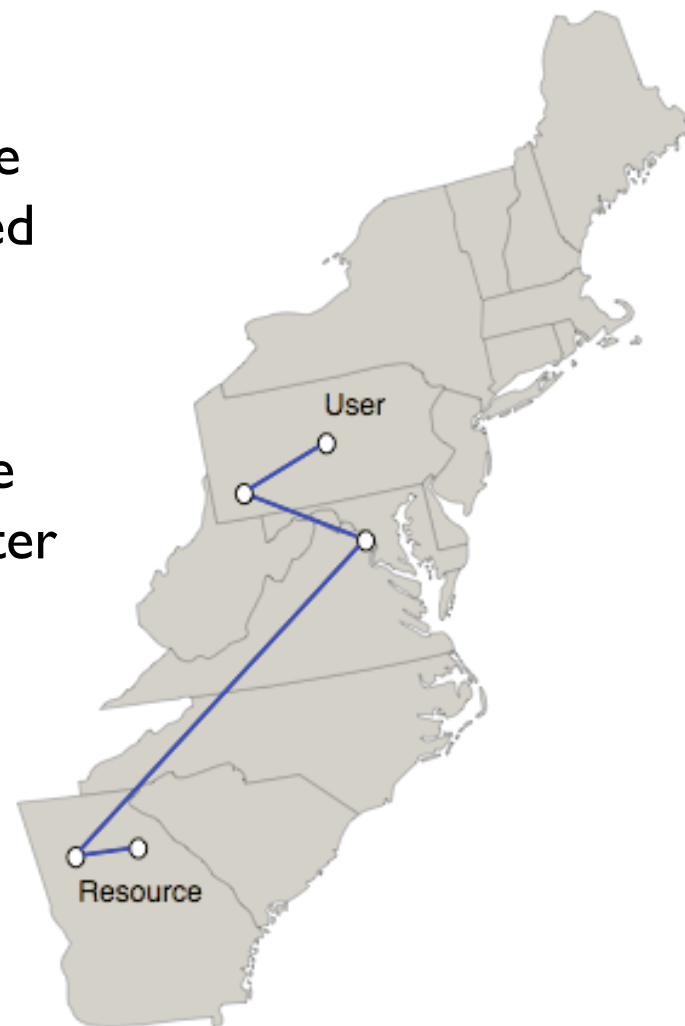
- perfSONAR should be used to diagnose an end-to-end performance problem
  - User is attempting to download from a remote resource
  - Resource and user are separated by distance
  - Both are assumed to be connected to high speed networks
- Operation does not go as planned, where to start?



Slides from Jason Zurawski

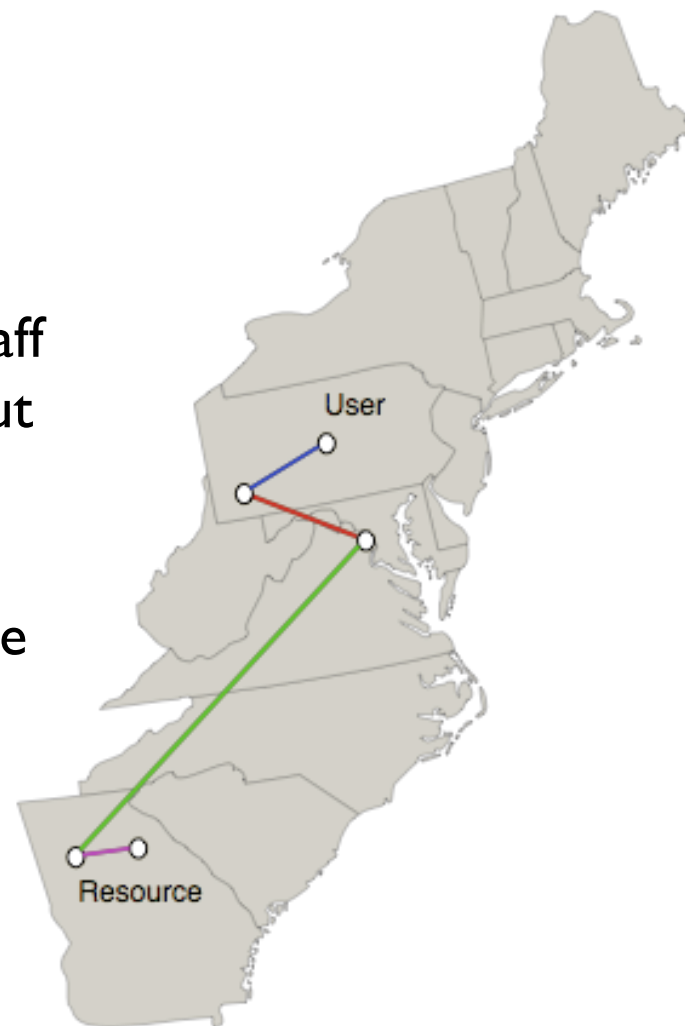
# Example perfSONAR Use Case

- Simple tools like **traceroute** can be used to determine the path traveled
- There could be a performance problem anywhere in here
- The problem may be something we could fix, but the chances are greater that it is not



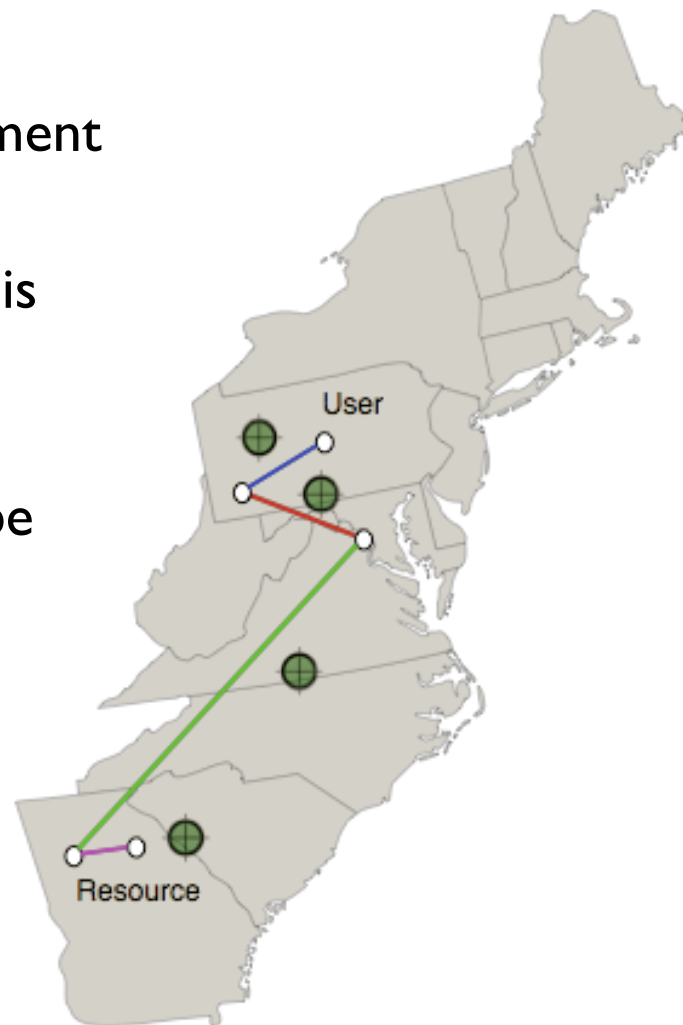
# Example perfSONAR Use Case

- Each segment of the path is controlled by a different domain.
- Each domain will have network staff that could help fix the problem, but how to contact them?
- All we really want is some information regarding performance



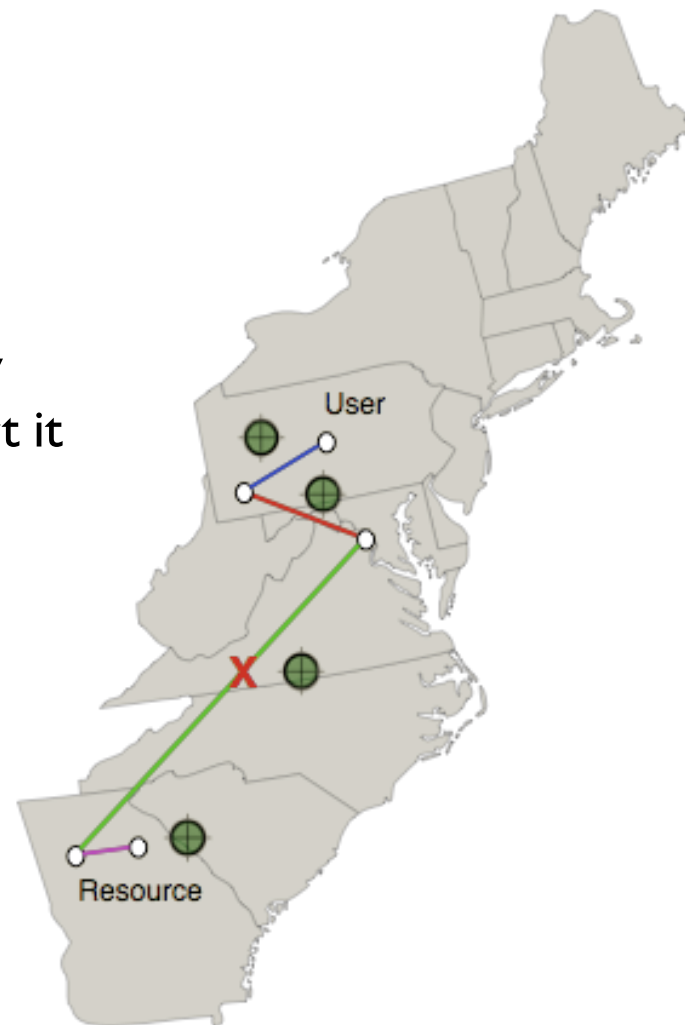
# Example perfSONAR Use Case

- Each domain has made measurement data available via perfSONAR
- The user was able to discover this automatically
- Automated tools such as visualizations and analyzers can be powered by this network data



# Example perfSONAR Use Case

- In the end, the problem is isolated based on testing.
  - May have gone unnoticed in some cases (e.g. a “soft failure”)
  - Could have been observed by many others ... that didn't think to report it
- The user (or operations staff) can contact the domain in question to inquire about this performance problem
- When fixed the transfer should progress as intended





# What Does perfSONAR-PS Provide?

- Installing perfSONAR-PS on a host provides a number of tools and options
  - Instance can be configured to perform regularly scheduled tests: Bandwidth (BWCTL), Latency/Packet-loss (OWAMP), Traceroute and PingER (Round-trip times)
  - Provides an on-demand test point for the above
  - Has built-in diagnostic tools accessible via web:
    - NDT (Runs on `http://<fqdn>:7123` by default)
    - NPAD (Runs on `http://<fqdn>:8000` by default)

# Installing perfSONAR-PS Toolkit

- Identify the hardware to use  
<http://fasterdata.es.net/performance-testing/perfsonar/ps-howto/hardware/>
  - Recommendation is two instances: one for latency/traceroute and one for bandwidth
- Get/assign DNS name and IP addresses
- Install (either on-disk via 'Netinstall' image or as an appliance via CDROM/USB image)
  - Quick start at <http://code.google.com/p/perfsonar-ps/wiki/pSPerformanceToolkit33>
- Configure it!

# Configuring perfSONAR-PS

- Details in the quick start URL above
- Should supply “Administrative Information” information
  - On the <http://<fqdn>/toolkit> page there is a “Toolkit Administration” box:
- You will need credentials
  - (Created during install)
- Click on ‘Administrative Information’
- Put in address/location of box suitable for finding in Google Maps for example
- Decide on relevant “Communities” to participate in

Toolkit Administration	
Administrative Information	
External BWCTL Limits	
External OWAMP Limits	
Enabled Services	
NTP	
Scheduled Tests	
Cacti SNMP Monitoring	<a href="#">↗</a>
perfSONAR Logs	<a href="#">↗</a>

# Configuring perfSONAR-PS (2)

- Define “Enabled Services”
  - Is this a latency or bandwidth node?



## User Tools

Local Performance Services  
Global Performance Services  
Java OWAMP Client [↗](#)  
Reverse Traceroute [↗](#)  
Reverse Ping [↗](#)

## Service Graphs

Throughput  
One-Way Latency  
Ping Latency  
SNMP Utilization  
Cacti Graphs [↗](#)

## Toolkit Administration

Administrative Information  
External BWCTL Limits  
External OWAMP Limits  
Enabled Services  
NTP

## Enabled Services Configuration Tool

Configuration Saved And Services Restarted

Save Reset

Services	Description
<input checked="" type="checkbox"/> PingER	Enables this host to perform scheduled ping tests. These tests will periodically ping configured host latency from their site over time.
<input type="checkbox"/> perfSONAR-BUOY Throughput Testing	Enables this host to perform scheduled throughput tests. These tests will run periodically giving adr from their site over time.
<input checked="" type="checkbox"/> perfSONAR-BUOY Latency Testing	Enables this host to perform scheduled one-way latency tests. These tests will run periodically giving their site over time.
<input checked="" type="checkbox"/> perfSONAR-BUOY Measurement Archive	Makes available the data collected by the perfSONAR-BUOY Latency and Throughput tests.
<input checked="" type="checkbox"/> NDT	Allows clients at other sites to run NDT tests to this host.
<input checked="" type="checkbox"/> NPAD	Allows clients at other sites to run NPAD tests to this host.
<input type="checkbox"/> BWCTL	Allows clients at other sites to run Throughput tests to this host
<input checked="" type="checkbox"/> OWAMP	Allows clients at other sites to run One-Way Latency tests to this host
<input checked="" type="checkbox"/> SSH	Allows administrators to remotely connect to this host using SSH
<input checked="" type="checkbox"/> SNMP MA	Makes available SNMP statistics collected by Cacti (Note: you must configure cacti for this to work)
<input checked="" type="checkbox"/> Traceroute MA	Makes available results of data collected by scheduled traceroute tests
<input checked="" type="checkbox"/> Traceroute Scheduler	Enables this host to run scheduled traceroute tests.
<input checked="" type="checkbox"/> Lookup Service	Registers your services into the global set of perfSONAR services so that they can be discovered

Only Enable Bandwidth Services

Only Enable Latency Services

Save Reset

# Example perfSONAR-PS WebGUI



## User Tools

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

## Service Graphs

- Throughput
- One-Way Latency
- Ping Latency
- SNMP Utilization
- Cacti Graphs

## Toolkit Administration

- Administrative Information
- External BWCTL Limits
- External OWAMP Limits
- Enabled Services
- NTP
- Scheduled Tests
- Cacti SNMP Monitoring
- perfSONAR Logs

## Performance Toolkit

- Configuration Help
- Frequently Asked Questions
- About
- Credits

perfSONAR

pS-Performance Node For AGLT2 In 0156 LSA, Univ of Michigan, Ann Arbor, MI USA

### Host Information

Organization Name	AGLT2
Host Location	0156 LSA, Univ of Michigan, Ann Arbor, MI USA
Administrator Name	Shawn McKee
Administrator Email	<a href="mailto:smckee@umich.edu">smckee@umich.edu</a>

### Communities This Host Participates In

USATLAS AGLT2 LHCONE pS-NPToolkit-3.2.2

### Host Status

Primary Address	psum01.aglt2.org
MTU	1500
NTP Status	Synced
Globally registered	Yes

### Services Offered

Bandwidth Test Controller (BWCTL)	Disabled
<ul style="list-style-type: none"> <li><a href="http://psum01.aglt2.org:4823">tcp://psum01.aglt2.org:4823</a></li> </ul>	
Lookup Service	Running
<ul style="list-style-type: none"> <li><a href="http://psum01.aglt2.org:9995/perfSONAR_PS/services/hLS">http://psum01.aglt2.org:9995/perfSONAR_PS/services/hLS</a></li> </ul>	
Network Diagnostic Tester (NDT)	Running
<ul style="list-style-type: none"> <li><a href="http://psum01.aglt2.org:3001">tcp://psum01.aglt2.org:3001</a></li> <li><a href="http://psum01.aglt2.org:7123">http://psum01.aglt2.org:7123</a></li> </ul>	
Network Path and Application Diagnosis (NPAD)	Disabled
<ul style="list-style-type: none"> <li><a href="http://psum01.aglt2.org:8001">tcp://psum01.aglt2.org:8001</a></li> <li><a href="http://psum01.aglt2.org:8000">http://psum01.aglt2.org:8000</a></li> </ul>	
One-Way Ping Service (OWAMP)	Running
<ul style="list-style-type: none"> <li><a href="http://psum01.aglt2.org:861">tcp://psum01.aglt2.org:861</a></li> </ul>	
perfSONAR-BUOY Regular Testing (Throughput)	Disabled
perfSONAR-BUOY Measurement Archive	Running

# Using perfSONAR-PS Toolkit On-Demand



- The toolkit provides the ability to do on-demand tests
- **NDT** (Network Diagnostics Toolkit) can be invoked from any web browser which supports Java-script. Connect to a perfSONAR-PS **host** <http://<host>:7123>
  - Click “Start” and a 20 second test is run
- **NPAD** (Network Path and Application Diagnostics) can also be run
  - Connect to <http://<host>:8000>
- Both **NDT** and **NPAD** analyze the network by using a WebI00 instrumented TCP stack
- OSG distributes “CLI client” versions of the tools to allow running from the command line

<https://twiki.grid.iu.edu/bin/view/Documentation/Release3/NetworkPerformanceToolkit>



# Using Scheduled Testing

- Most sites installing perfSONAR-PS will want to setup some regularly scheduled tests to other relevant sites across the network
  - This creates a baseline to compare to
  - It can be useful for identifying when and where problems occur
- Click “Scheduled Tests”
- Setup tests

Toolkit Administration	
Administrative Information	
External BWCTL Limits	
External OWAMP Limits	
Enabled Services	
NTP	
Scheduled Tests	
Cacti SNMP Monitoring	
perfSONAR Logs	

# Example of OWAMP Tests



## User Tools

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

## Service Graphs

- Throughput
- One-Way Latency
- Ping Latency
- SNMP Utilization
- Cacti Graphs

## Toolkit Administration

- Administrative Information
- External BWCTL Limits
- External OWAMP Limits

## perfSONAR-PS Tests

### Service type

One Way Latency

### Active Tests:

▲ - Sorted(asc) by that column. Click on column headings to sort

Source ▲	Destination	Bidirectional	Forward Direction Loss (Past 30 minutes)	Reverse Direction Loss (Past 30 minutes)	Graph
PERFSONAR01.CMSAF.MIT.EDU (18.12.1.171)	psum01.aglt2.org (192.41.230.19)	Yes	0.00%	0.00%	Select ▼
atlas-npt1.bu.edu (192.5.207.251)	psum01.aglt2.org (192.41.230.19)	Yes	0.00%	0.00%	Select ▼
iut2-net1.iu.edu (149.165.225.223)	psum01.aglt2.org (192.41.230.19)	Yes	0.00%	0.00%	Select ▼
lcg-lrz-perfs1.grid.lrz.de (129.187.131.211)	psum01.aglt2.org (192.41.230.19)	No	0.00%	*	Select ▼
lhc-latency.twgrid.org (117.103.105.191)	psum01.aglt2.org (192.41.230.19)	Yes	0.00%	0.00%	Select ▼

# The 24 hour OWAMP Result BU (Boston) – AGLT2 (Michigan)

perfSONAR One Way Latency

perfSONAR

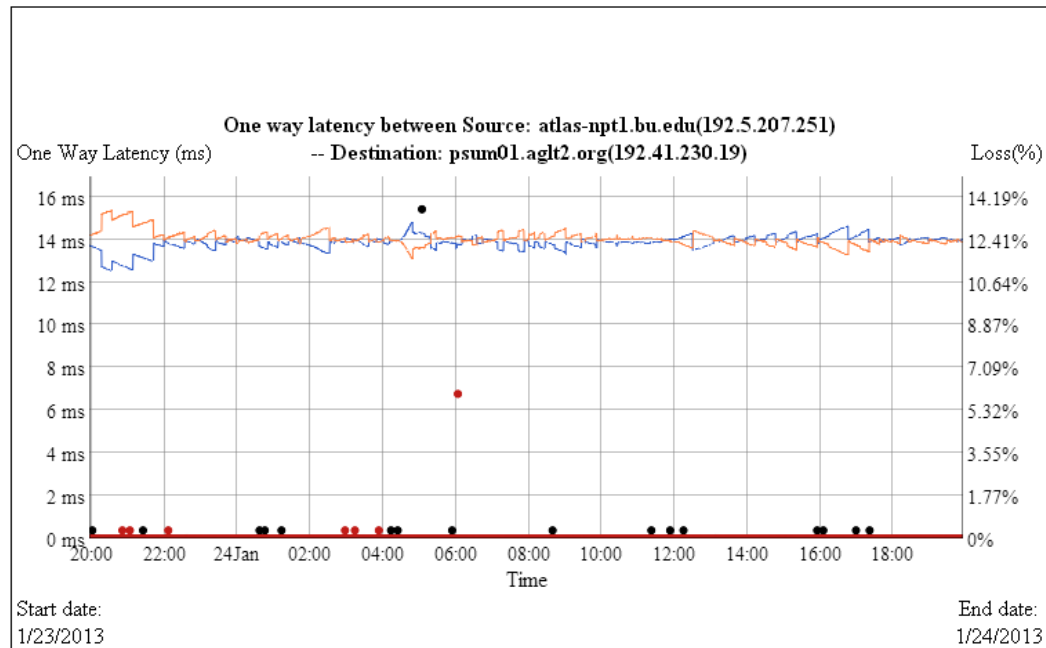
☒ Scale Y axis from 0 ☒ Show Reverse Direction Data

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



<- 4 hours

This plot shows results in both directions (tests are unidirectional). While the one-way latency can be useful, more important is the loss indication. This shows a path that while acceptable, is not as clean as we would like.

# Another OWAMP Example: SWT2 (Texas) – AGLT2 (Michigan)

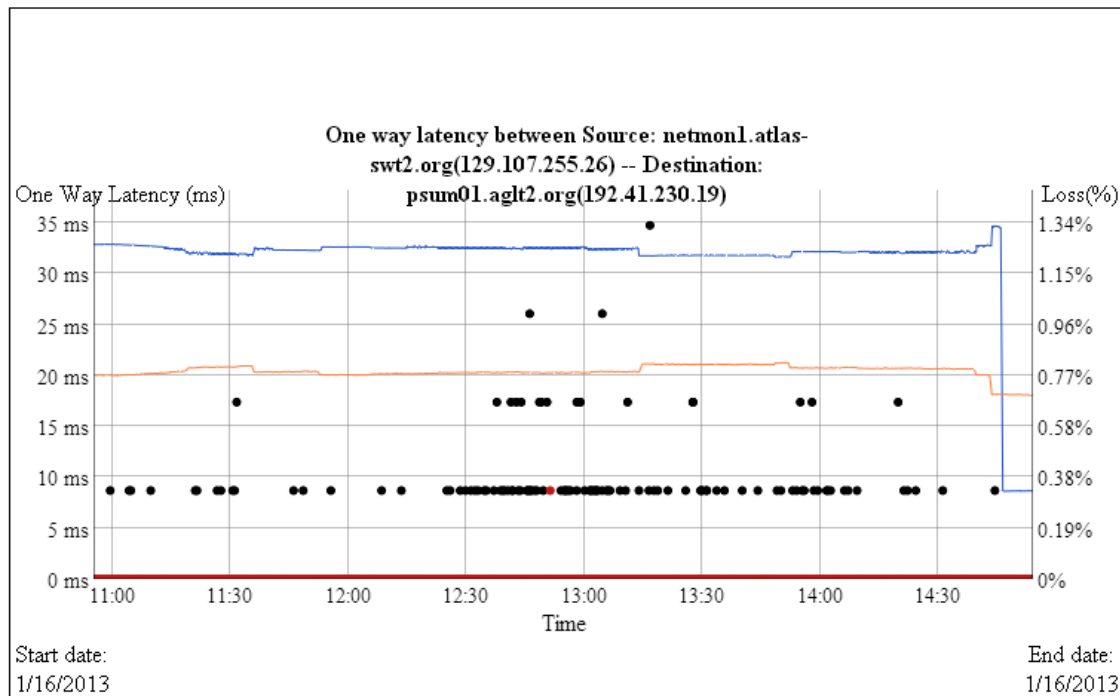
perfSONAR One Way Latency

perfSONAR

☒ 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

<- 4 hours

Timezone: Standard Time)

This plot shows results in both directions (tests are unidirectional). This shows an asymmetric route (different latency in each direction). Loss is more frequent than the last example but not as large. This may have a larger impact on throughput.

# Example of BWCTL Tests



## User Tools

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

## Service Graphs

Throughput  
One-Way Latency  
Ping Latency  
SNMP Utilization  
Cacti Graphs

## Toolkit Administration

Administrative Information  
External BWCTL Limits  
External OWAMP Limits  
Enabled Services  
NTP  
Scheduled Tests  
Cacti SNMP Monitoring  
perfSONAR Logs

## perfSONAR-PS Tests

### Service type

bandwidth(bwctl)

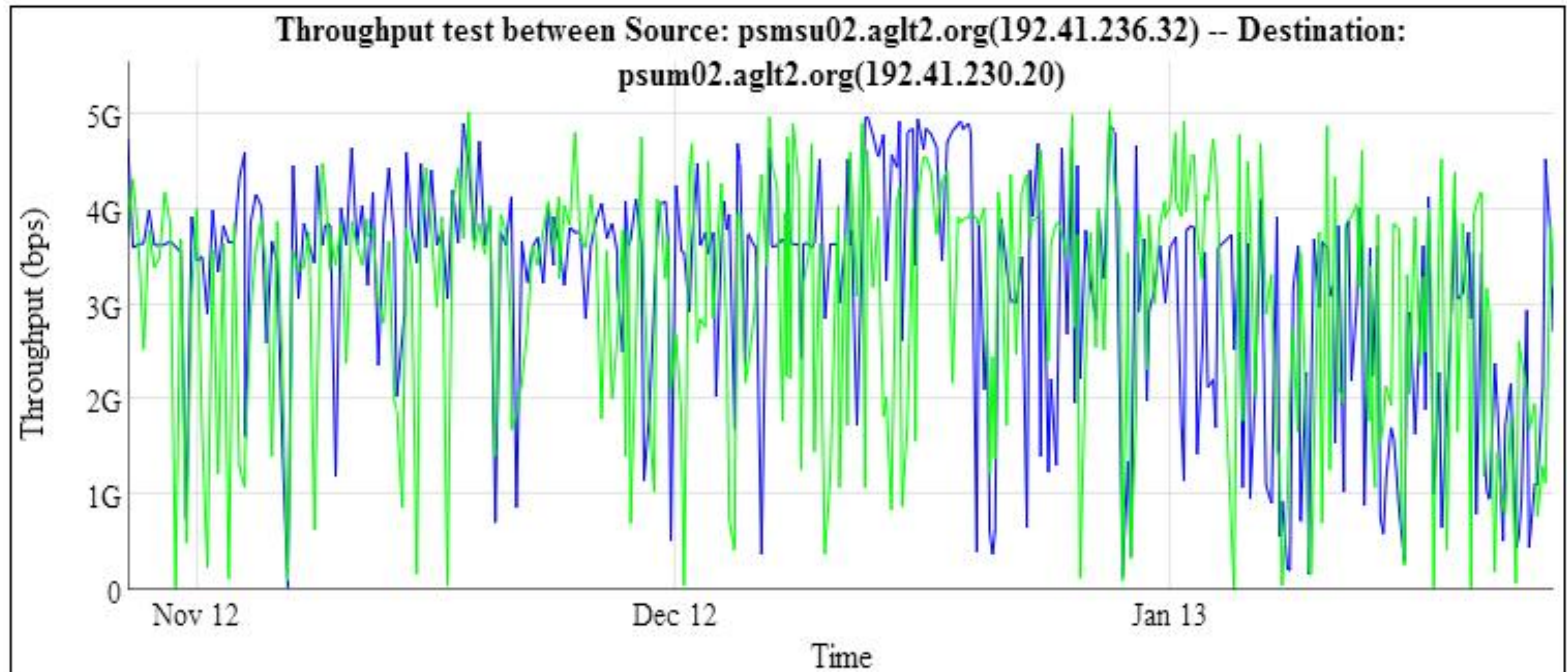
### Active Tests:

▲ - Sorted(asc) by that column. Click on column headings to sort

Source ▲	Destination	Bidirectional	Protocol	Duration	1 week Avg Throughput Src-Dst (Gbps)	1 week Avg Throughput Dst-Src (Gbps)	Graph
iperf.its.vanderbilt.edu (192.111.110.34)	psum02.aglt2.org (192.41.230.20)	No	TCP	30	0.877	No data	Select ▼
iut2-net2.iu.edu (149.165.225.224)	psum02.aglt2.org (192.41.230.20)	Yes	TCP	20	0.199	0.231	Select ▼
lhcmn.bnl.gov (192.12.15.23)	psum02.aglt2.org (192.41.230.20)	Yes	TCP	30	1.748	0.668	1 month ▼
lhcmn.bnl.gov (192.12.15.23)	psum02.aglt2.org (192.41.230.20)	Yes	TCP	20	1.911	0.596	Select ▼
lpnhe-gs9086.in2p3.fr (134.158.159.86)	psum02.aglt2.org (192.41.230.20)	Yes	TCP	30	0.263	0.21	Select ▼
perfsonar-bw.farm.particle.cz (147.231.19.52)	psum02.aglt2.org (192.41.230.20)	No	TCP	30	0.411	No data	Select ▼
perfsonar-ps-02.desy.de (131.169.98.29)	psum02.aglt2.org (192.41.230.20)	Yes	TCP	30	0.377	0.238	Select ▼
perfsonar-ps-bandwidth.pic.es (193.109.172.190)	psum02.aglt2.org (192.41.230.20)	Yes	TCP	30	0.4	0.226	Select ▼

# The 3-Month BWCTL Result

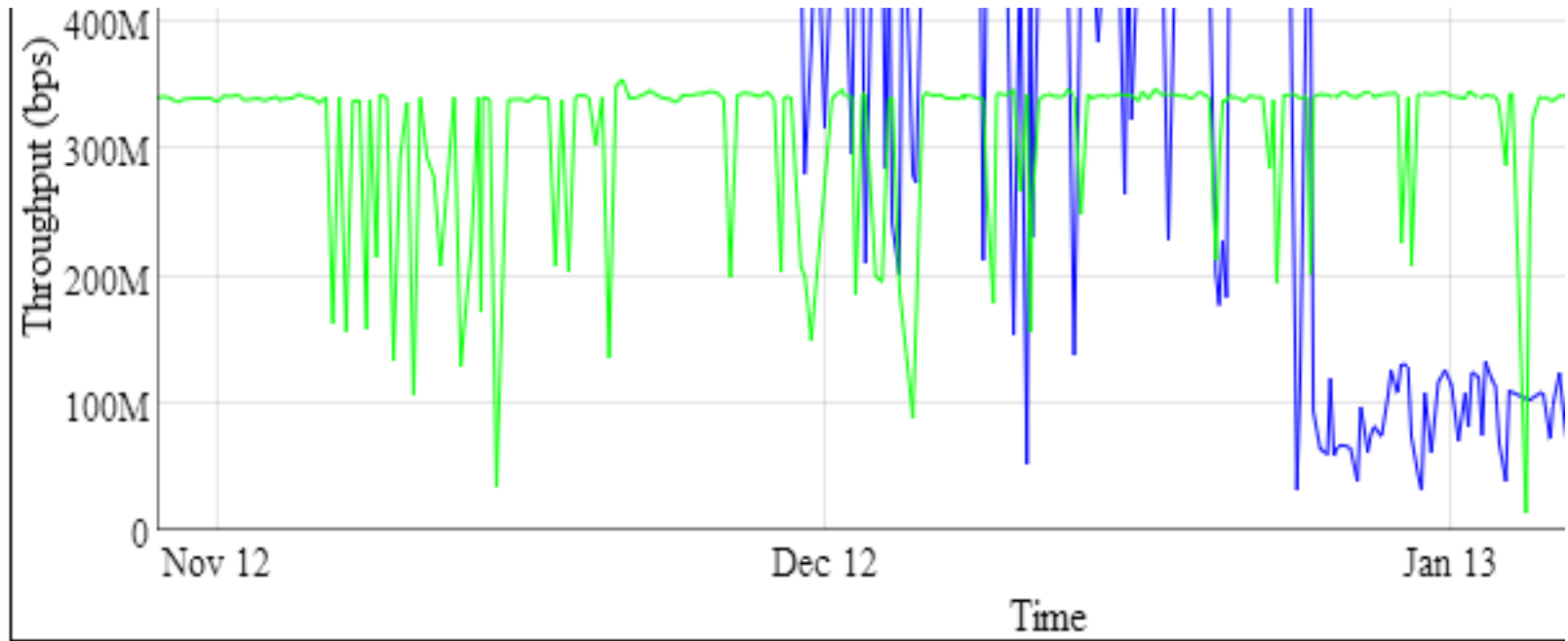
## AGLT2:Ann Arbor – East Lansing



This plot shows results in both directions Src->Dst is BLUE while Dst->Src is GREEN. These testers are obviously using 10 GE NICs. Measurements vary quite a bit, perhaps due to real loads (congestion).



# The 3-Month BWCTL Result Michigan - Japan



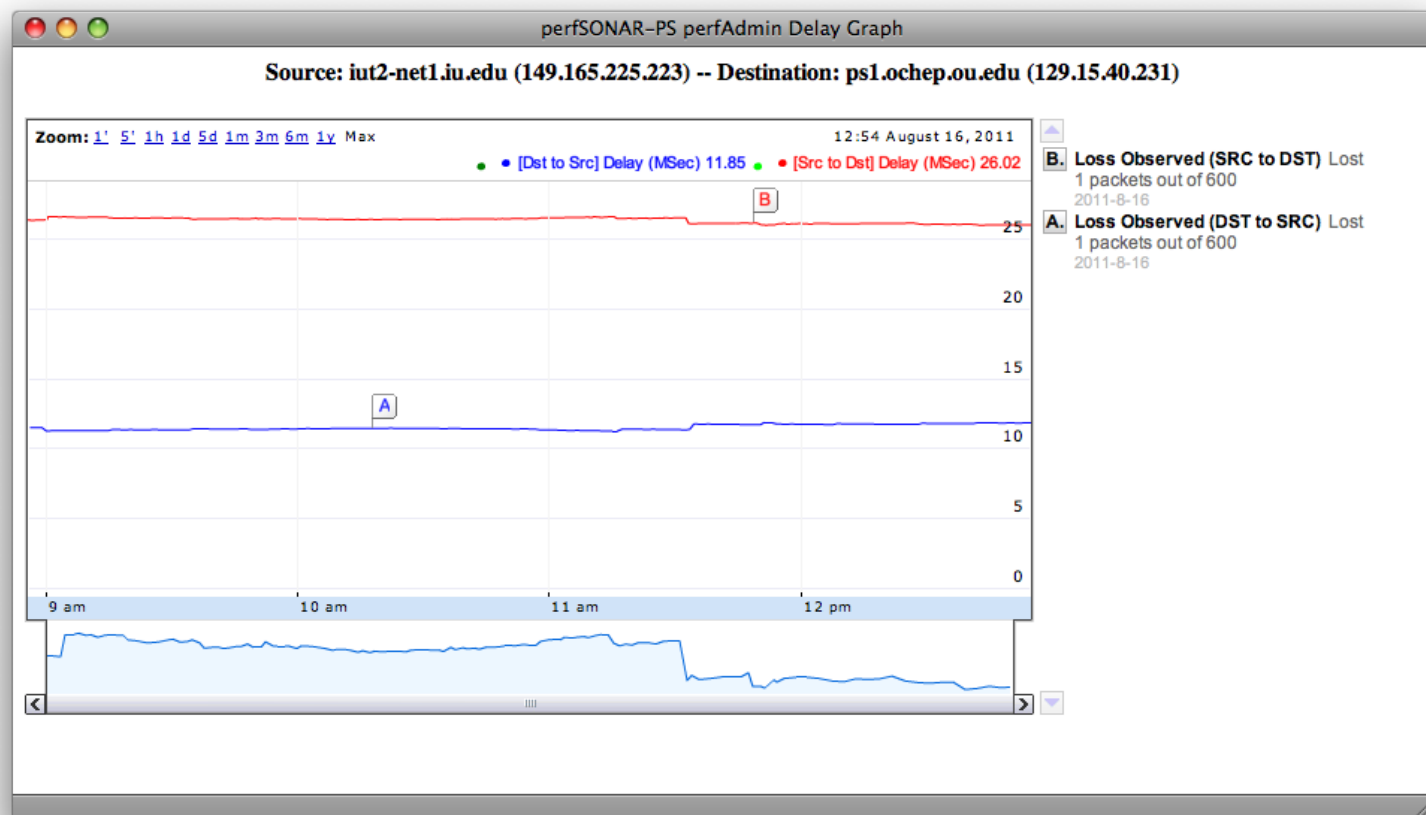
This plot shows results in both directions between Michigan and Japan. These show some much more interesting behavior than the last test. Src->Dst is BLUE while Dst->Src is GREEN. Tests from Japan->Michigan started around the end of November. Something changed around Christmas. Recent change in Dst->Src?

# Network Performance – Example of Use

The following example of using perfSONAR-PS is from Jason Zurawski/Internet2 and demonstrates using perfSONAR-PS measurements to identify a problem and verify the solution.

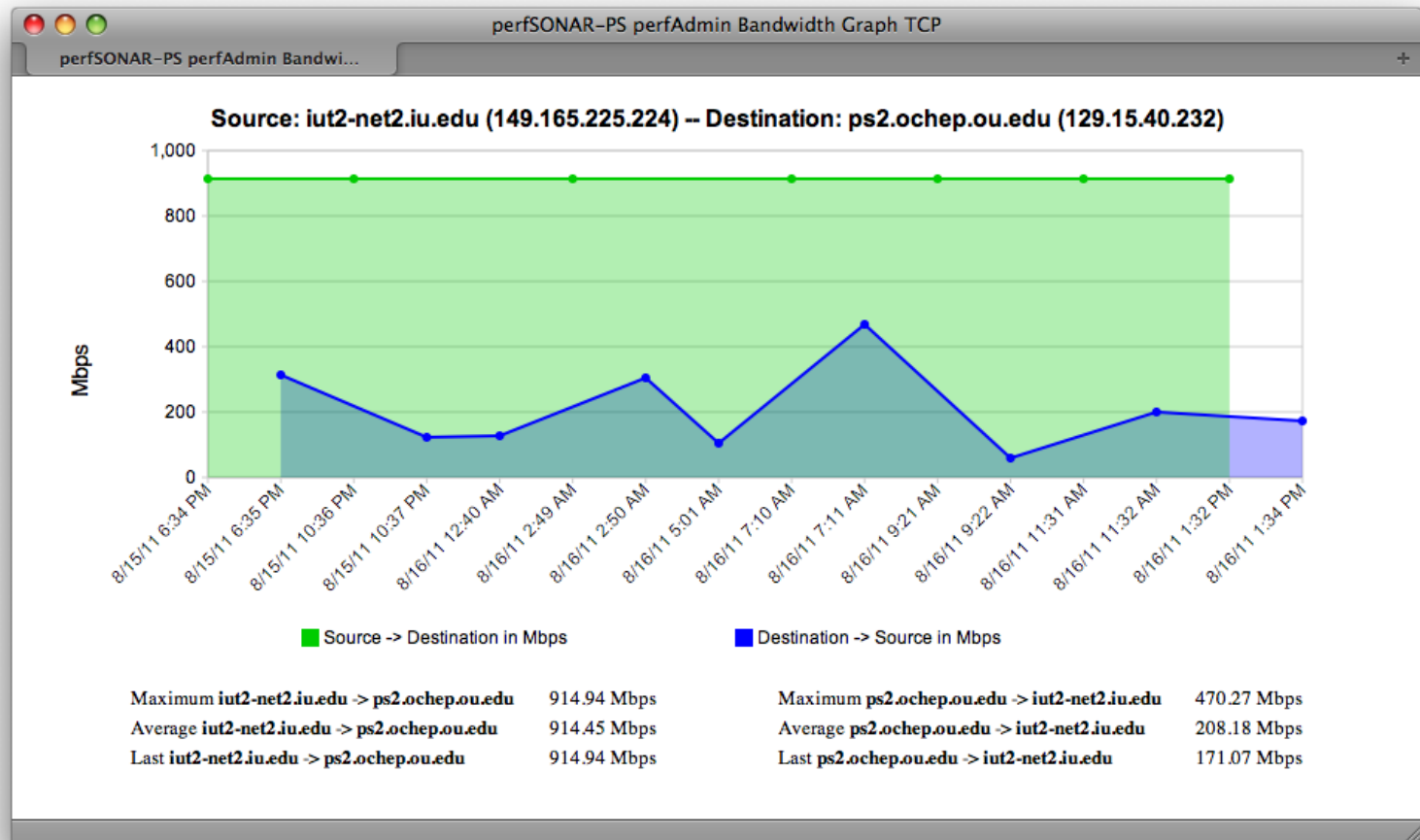
This incident happened in USATLAS where we noticed poor performance between Indiana and Oklahoma.

# Asymmetric Latency



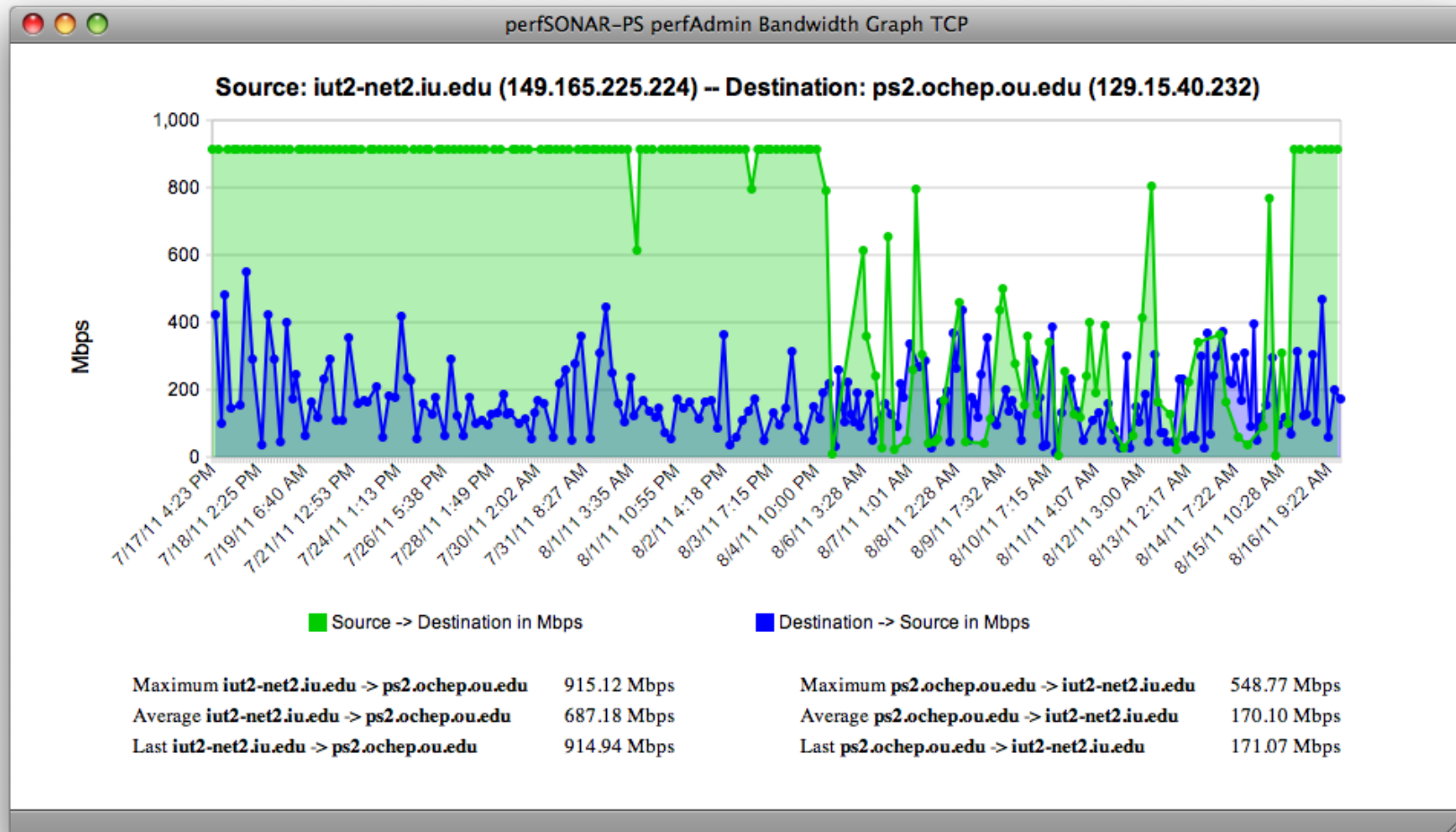
- This is not a problem by itself
  - Only a problem if there is a real problem on either side
  - If one side has a problem – think who complains (uploaders vs. downloaders)

# And What Does Bandwidth Say?



- One of the paths is doing poorly

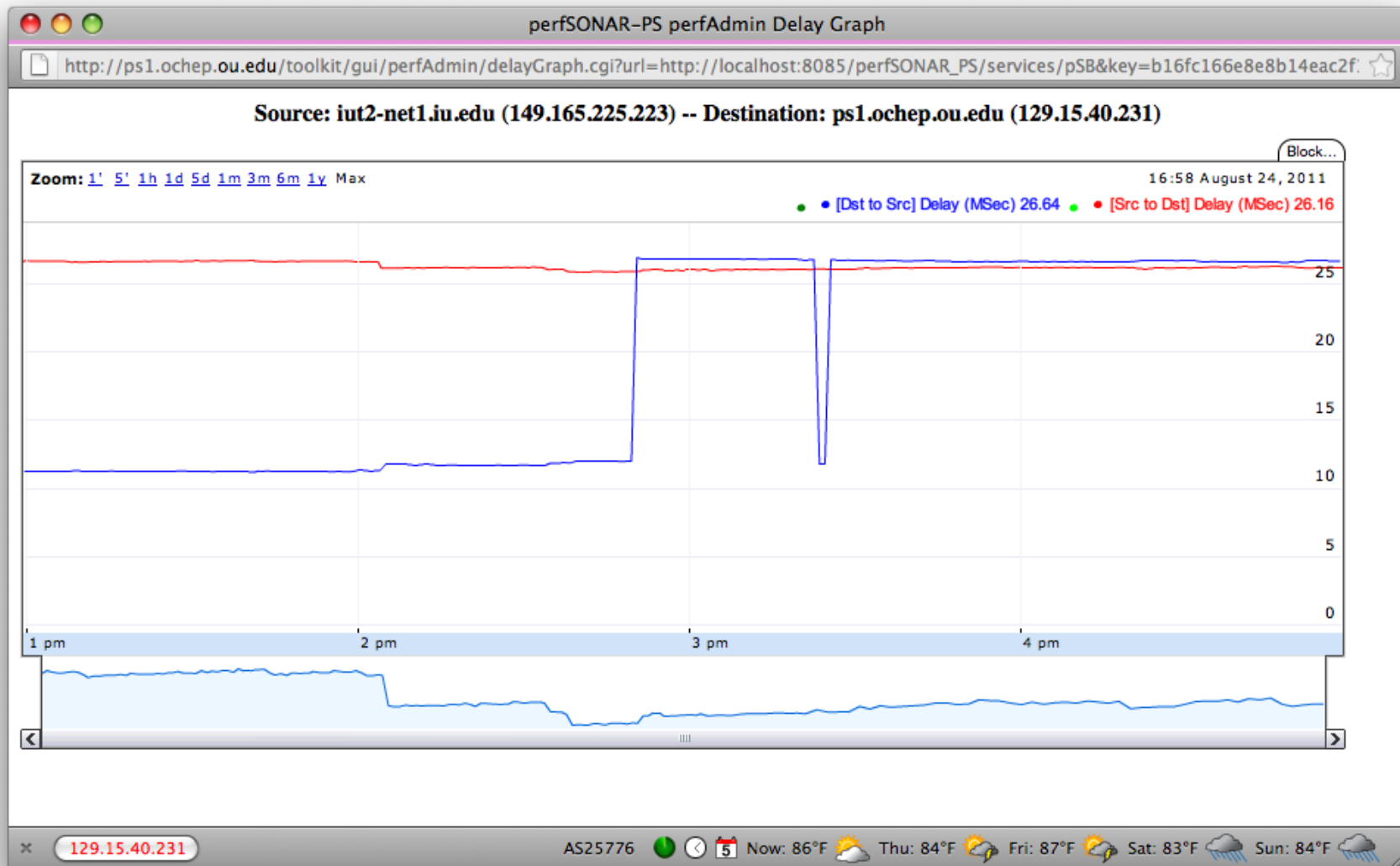
# Longer View (Problem Existed for a While)



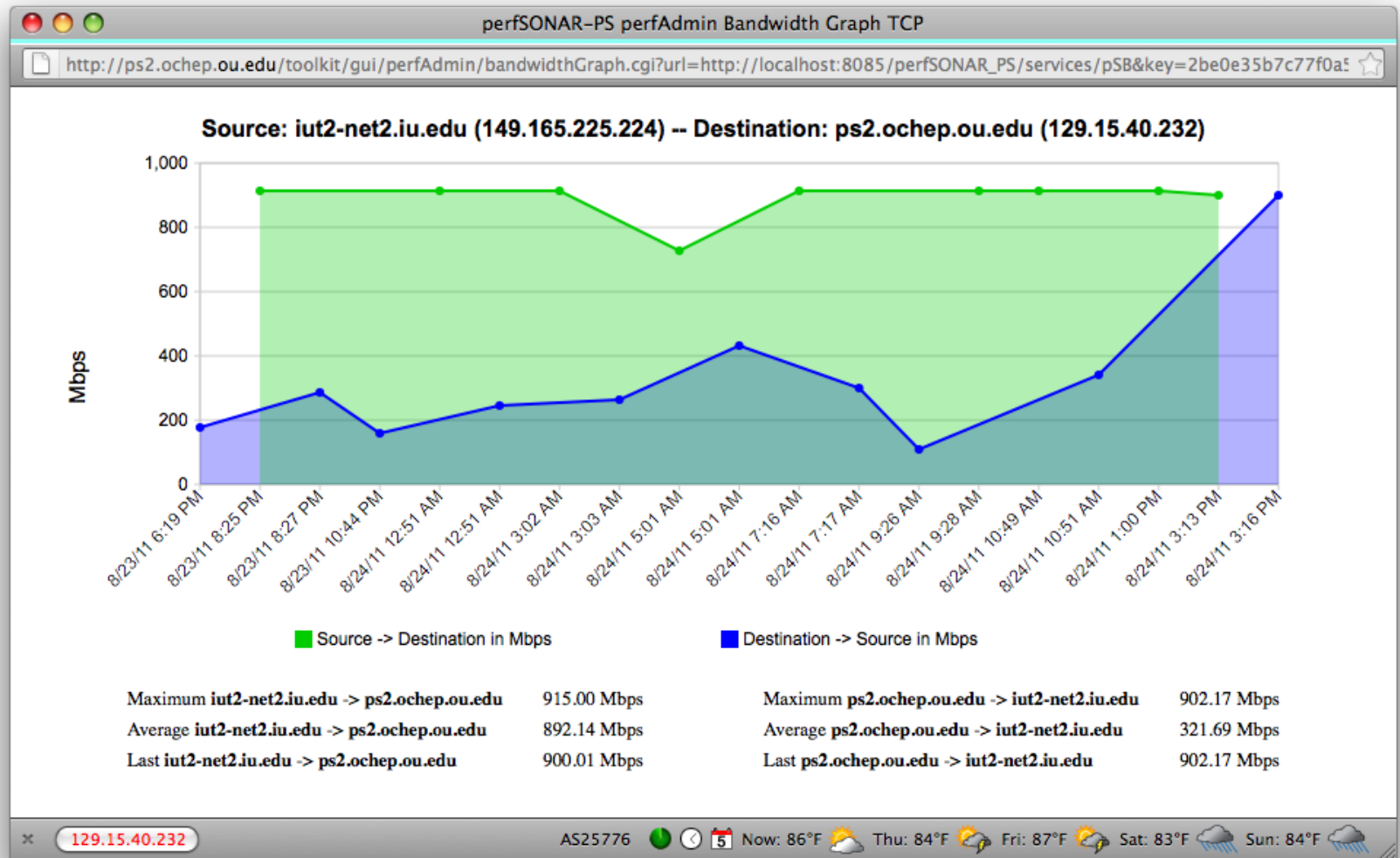
# Possible Solution Space

- Alter routing to use the same path
  - E.g. the ‘good’ path hopefully ☺
- Fix the “bad” path
- Note these two are often done together
  - Please refrain from just changing routing to avoid a problem and not telling anyone

# Routing Change Results (Latency)

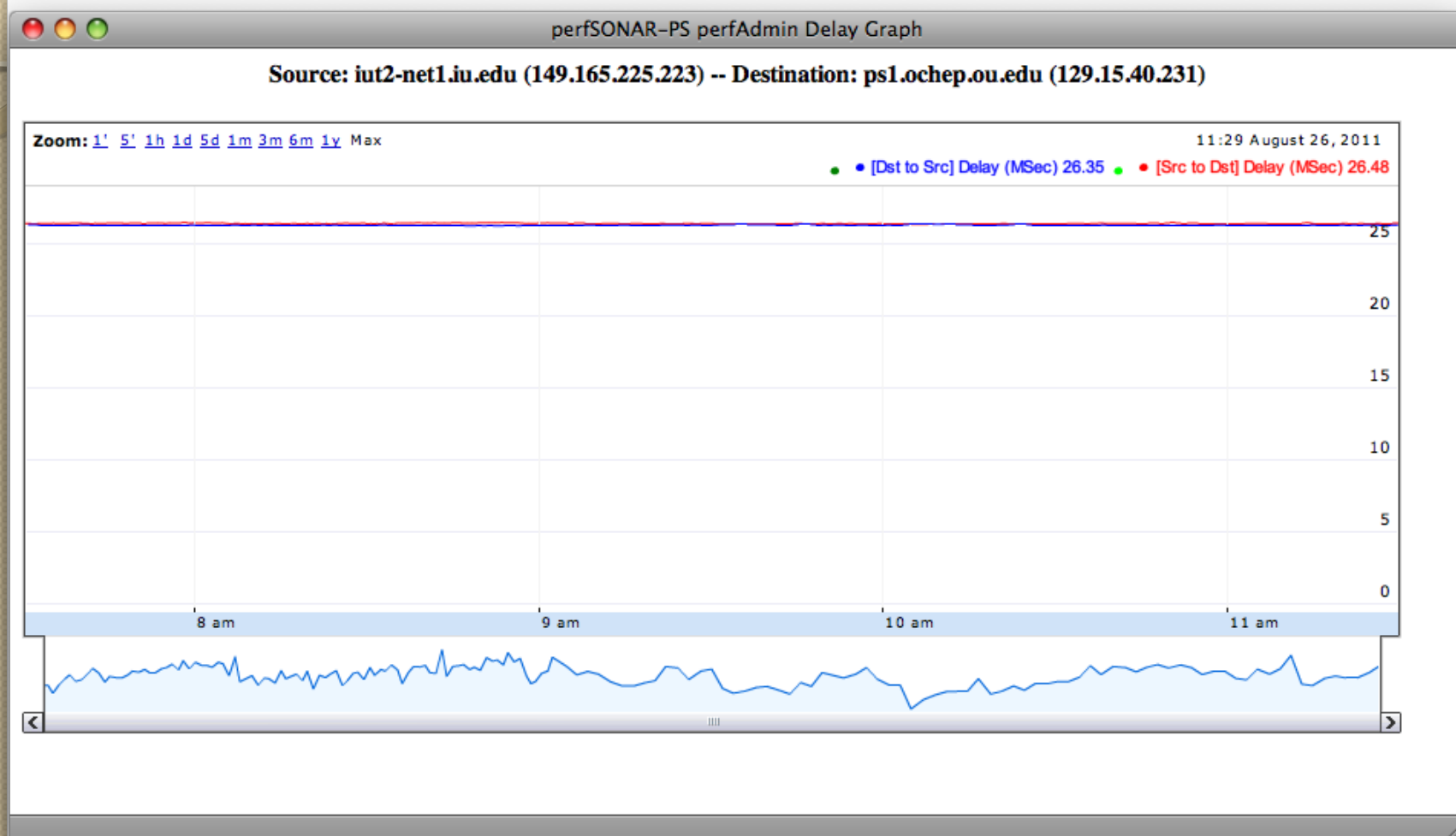


# Routing Change Results (Bandwidth)

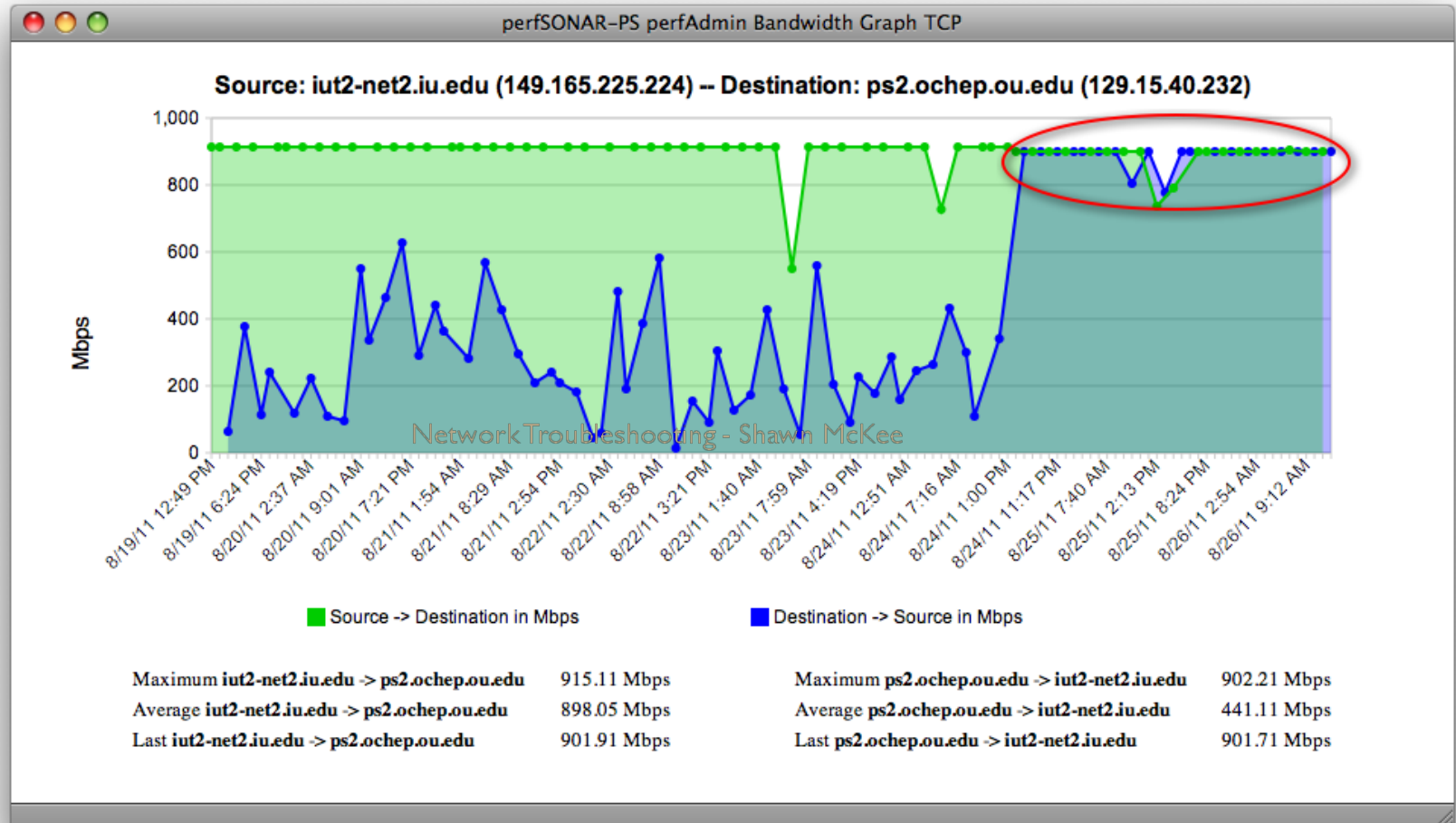




# Routing Change 1 Week Later (Latency)



# Routing Change 1 Week Later (BW)



Things are MUCH better.

# Checklist for Network Issues

- Check end-hosts for issues
- Use on-demand tests to verify a problem
- Examine any scheduled/existing network metrics for problems
- Gather the relevant data (application, host info source, destination, network metrics, problem)
- Contact your local network people with the information you collected. Also available:
  - - Research Support @ Internet2: rs@internet2.edu
  - - Trouble Reporting @ ESnet: trouble@es.net
- See <http://fasterdata.es.net/performance-testing/perfsonar/troubleshooting/overview/> for a good summary of the troubleshooting process

# The Future

- Open Science Grid has a new initiative in networking
  - Goal is to provide tools and service to support OSG sites regarding networking
  - New troubleshooting document being written
- There is a new project in GitHub focused on creating a new “Modular Dashboard” to gather, analyze and display perfSONAR metrics (see URLs at end of talk)
- perfSONAR-PS will continue to develop to meet the needs of its users

# Summary

- Network issues can be difficult to understand and localize
  - End-users shouldn't be expected to fix network problems BUT they should report them with enough detail to allow experts to do so!
- perfSONAR-PS is a central component for finding network problems. Deploy it! It can help you (and others) find/fix network problems
- Make sure you know who to contact about network problems.

# Questions or Comments?

Thanks!

# Useful References

- OSG Networking:  
<https://www.opensciencegrid.org/bin/view/Documentation/NetworkingInOSG>
- perfSONAR-PS site <http://psps.perfsonar.net/>
- perfSONAR-PS Install/configuration guide:  
<http://code.google.com/p/perfsonar-ps/wiki/pSPerformanceToolkit33>
- ESnet Fasterdata website: <http://fasterdata.es.net/>
- Modular Dashboard:  
<https://perfsonar.racf.bnl.gov:8443/exda/> or  
<http://perfsonar.racf.bnl.gov:8080/exda/>
- GitHub PerfModDash Organization:  
<https://github.com/PerfModDash>
- perfSONAR-PS Toolkit tips and maintenance:  
<http://www.usatlas.bnl.gov/twiki/bin/view/Projects/LHCperfSONAR>