

July 22nd 2013, XSEDE Network Performance Tutorial

Jason Zurawski – Internet2/ESnet

Kathy Benninger - Pittsburgh Supercomputing Center

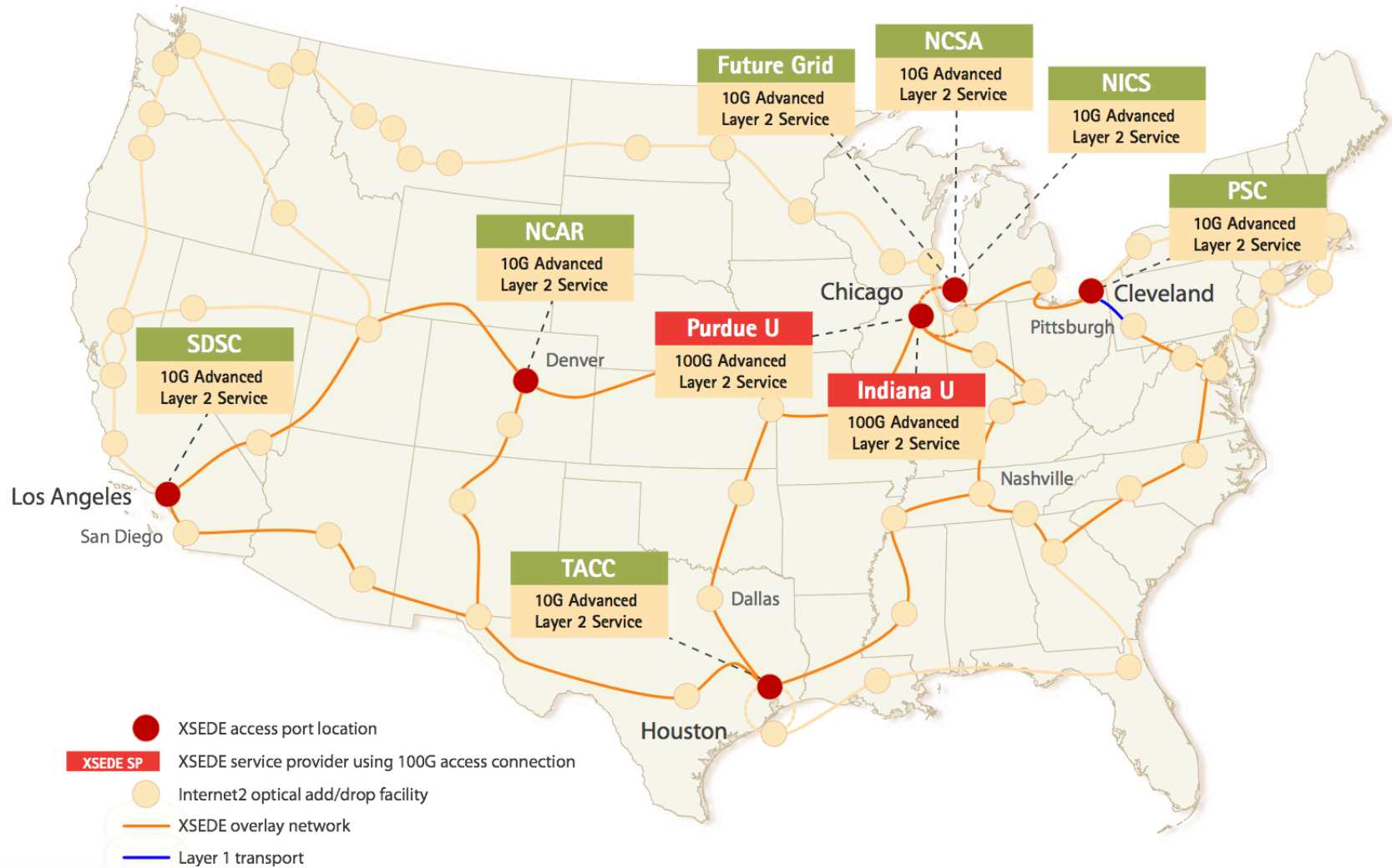
XSEDE Network Architecture Overview and Measurement Resources

Outline

- XSEDE Network Architecture
- Measurement Points
 - XSEDE
 - Internet2
 - ESnet
- XSEDE GridFTP speedpage
- ESnet IO Testers

XSEDE Network Architecture

- XSEDE network transition to Internet2 was completed in March 2013
- Built on Internet2's 100 Gb/s Advanced Layer2 Service (AL2S) backbone
- Indiana University and Purdue University connect at 10 Gb/s over a shared 100 GE
- TACC plans to move to 100 GE summer 2013
- Other XSEDE Tier 1 Service Provider (SP) sites connect at 10 GE with potential for upgrade to 100 GE
- NCSA Blue Waters connects at 100 GE



XSEDE Network Architecture

- XSEDE SP site connections are backhauled to nearest AL2S node
- Sites are connected via point-to-point VLANs
- Services available via AL2S are evolving
- Future:
 - Provision on multi-point VLAN when supported by the infrastructure, providing much greater flexibility for adding connections to XSEDE in the future (e.g. campus bridging)
 - SDN/OpenFlow

Outline

- XSEDE Network Architecture
- Measurement Points
 - XSEDE
 - Internet2
 - ESnet
- XSEDE GridFTP speedpage
- ESnet IO Testers

Measurement Points

- Each network provider has a litany of tools available for end-user consumption*:
 - Visualizations of collected measurements (SNMP, Netflow/sFlow, Throughput, Latency)
 - Active measurement points for specific protocols (BWCTL, Iperf, Nuttcp, OWAMP)
 - Procedures for evaluating IO performance (GridFTP)
- The following sections will outline some of these tools

* Note that not all providers make these things available. Caveat Emptor

Measurement Points – XSEDE

- One 10 GE-connected perfSONAR at each XSEDE SP site running throughput and latency tests
- Located near the network edge
- Accessible as <http://ps.<site>.xsede.org> where sites are:
 - Indiana University (iu)
 - National Center for Atmospheric Research (ncar)
 - National Center for Supercomputing Applications (nca)
 - National Institute for Computational Sciences (nics)

Measurement Points – XSEDE

- Pittsburgh Supercomputing Center (psc)
- Purdue University (purdue)
- San Diego Supercomputer Center (sdsc)
- Texas Advanced Computing Center (tacc)

Measurement Points – XSEDE

- Available perfSONAR tools:
 - bwctl
 - owamp
 - NDT
 - NPAD
 - Reverse traceroute
 - Reverse ping
 - perfSONAR-BUOY (bwctl and owamp)
 - Measurement archive
 - Maddash GUI

Measurement Points – Internet2

- Internet2
 - 4 Machines in each PoP on the current network (2 x Throughput Test Machine, 1 User Test Machine, 1 Latency Test Machine)
 - Internal Testing (<http://owamp.net.internet2.edu>), and 100s of community initiated tests per week
 - Central Netflow/SNMP Monitoring
 - Assistance available – rs@internet2.edu
 - Next generation still under discussion
 - Layer 2 (e.g. AL2S) Network could feature single server for all measurement tools
 - Layer 3 (e.g. AL3S) Network could have less machines, but still same service split and regular testing ability

Measurement Infrastructure – Internet2

- <http://noc.net.internet2.edu/i2network/advanced-layer-2-service/network-status.html>

Supported by the GlobalNOC at Indiana University

SEARCH... [+] Internet2 NOC

Internet² Network
Internet2 Network NOC

SUPPORT
LIVE: NETWORK STATUS
IP R&E NETWORK
COMMERCIAL PEERING SERVICE
ADVANCED LAYER 2 SERVICE
• USER INTERFACE
• [NETWORK STATUS](#)

GRNOC > [Internet2 NOC](#) > [Advanced Layer 2 Service](#) > [Network Status](#)

Show: [MostRecent](#)

Most Popular

1. ATLAS Traffic Map
2. AL2S SNAPP
3. AL2S Router Proxy
4. AL2S SNAPP
5. AL2S SNAPP


Recent Changes

There are no recent changes in this branch

NETWORK STATUS

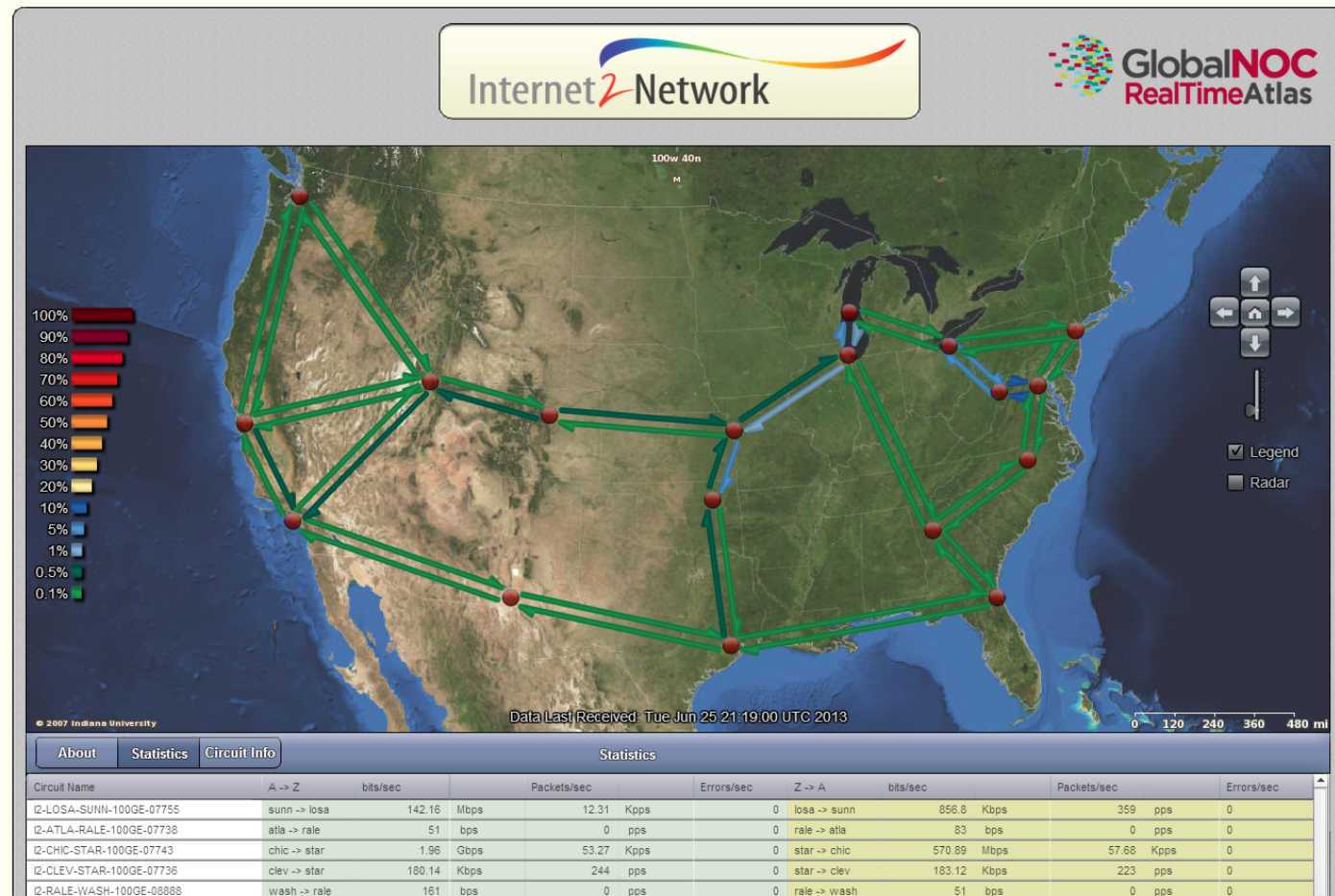
ATLAS - Advanced Layer 2 Service Traffic Map
SNAPP - Advanced Layer 2 Service Traffic Graphs
Router Proxy - The Advanced Layer 2 Service Looking Glass.

Copyright © 2013 The Trustees of Indiana University | Copyright Complaints

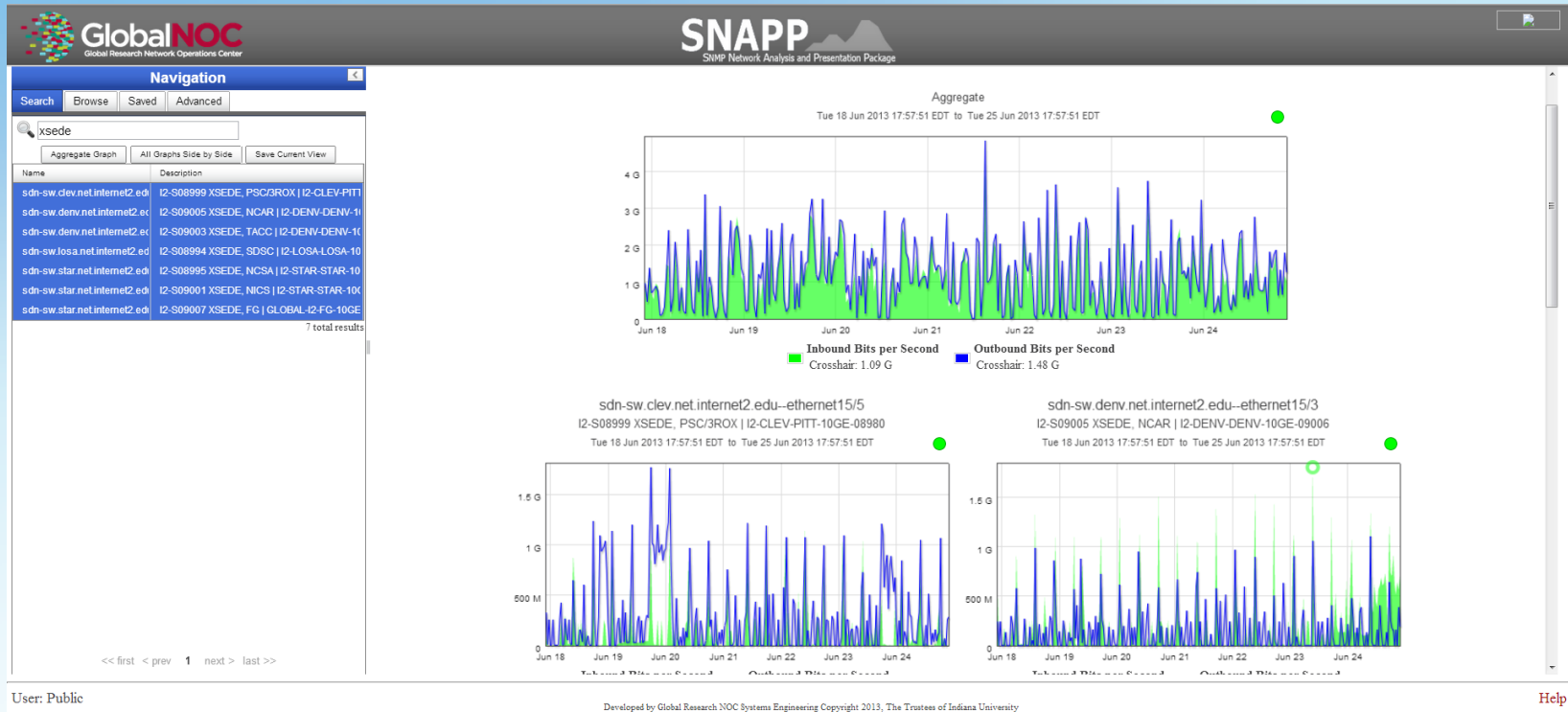
SUPPORTED BY  **GlobalNOC**
Global Research Network Operations Center
LOGIN

Measurement Infrastructure – Internet2

- http://atlas.grnoc.iu.edu/atlas.cgi?map_name=Internet2%20AL2S



Measurement Infrastructure – Internet2 AL2S SNAPP



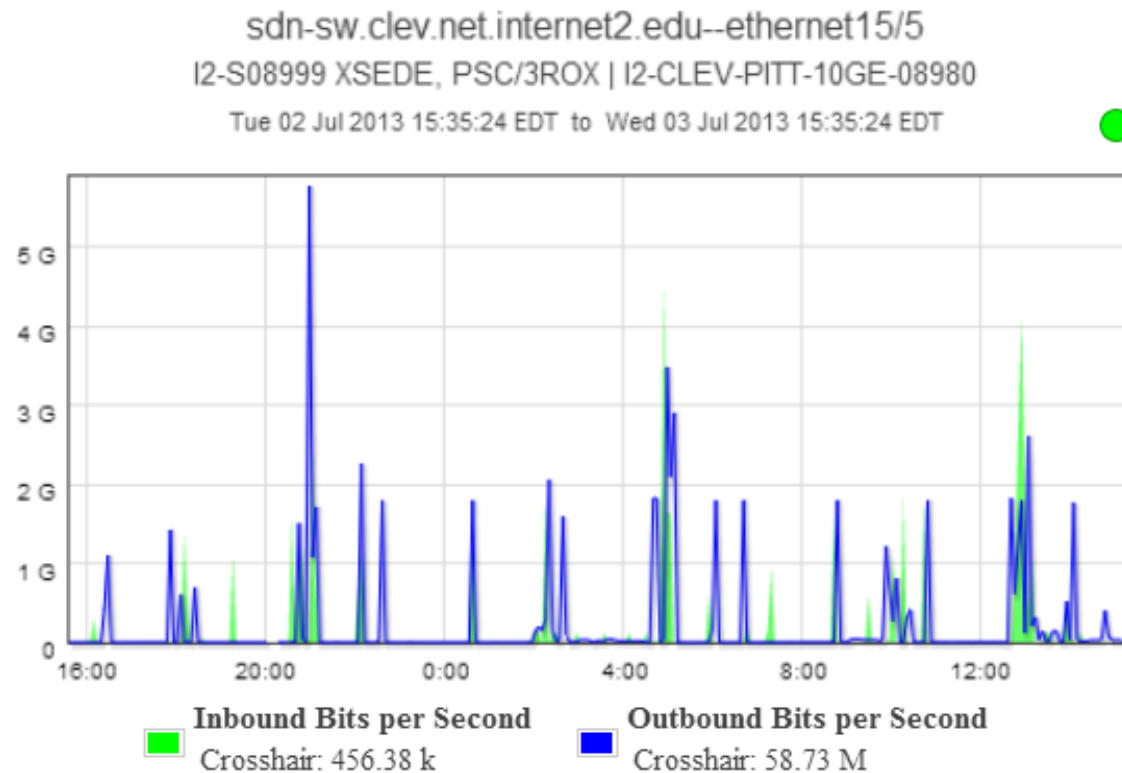
Measurement Infrastructure – Internet2 AL2S SNAPP

The screenshot shows the GNOC interface with a search bar containing 'xsede'. Below the search bar are buttons for 'Aggregate Graph', 'All Graphs Side by Side', and 'Save Current View'. A table displays 7 results, each with a Name and a Description.

Name	Description
sdn-sw.clev.net.internet2.edu-ethernet15/5	I2-S08999 XSEDE, PSC/3ROX I2-CLEV-PITT-10GE-08980
sdn-sw.denv.net.internet2.edu-ethernet15/3	I2-S09005 XSEDE, NCAR I2-DENV-DENV-10GE-09006
sdn-sw.denv.net.internet2.edu-ethernet15/4	I2-S09003 XSEDE, TACC I2-DENV-DENV-10GE-09004
sdn-sw.losa.net.internet2.edu-ethernet15/6	I2-S08994 XSEDE, SDSC I2-LOSA-LOSA-10GE-08981
sdn-sw.star.net.internet2.edu-ethernet15/1	I2-S08995 XSEDE, NCSA I2-STAR-STAR-10GE-08996
sdn-sw.star.net.internet2.edu-ethernet15/4	I2-S09001 XSEDE, NICS I2-STAR-STAR-10GE-09002
sdn-sw.star.net.internet2.edu-ethernet15/5	I2-S09007 XSEDE, FG GLOBAL-I2-FG-10GE-09008

7 total results

Measurement Infrastructure – Internet2 AL2S SNAPP

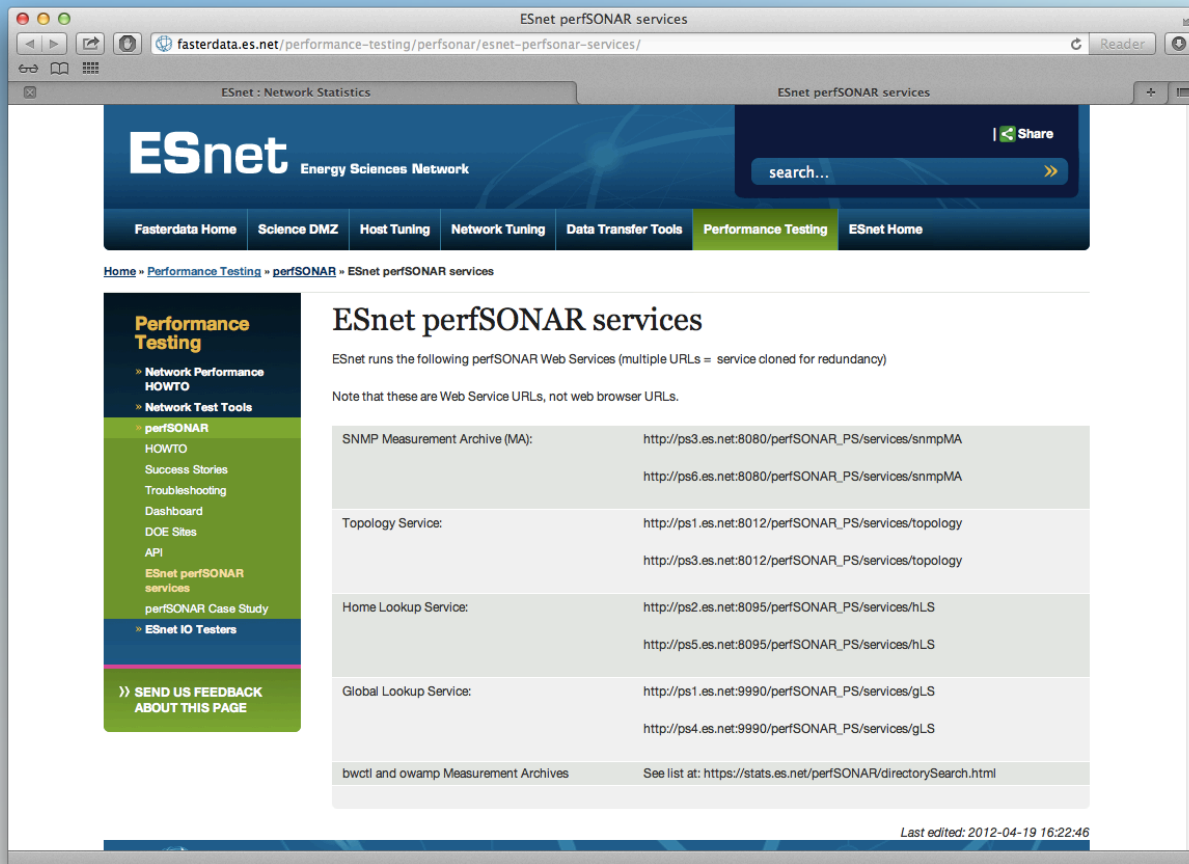


Measurement Points - ESnet

- 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

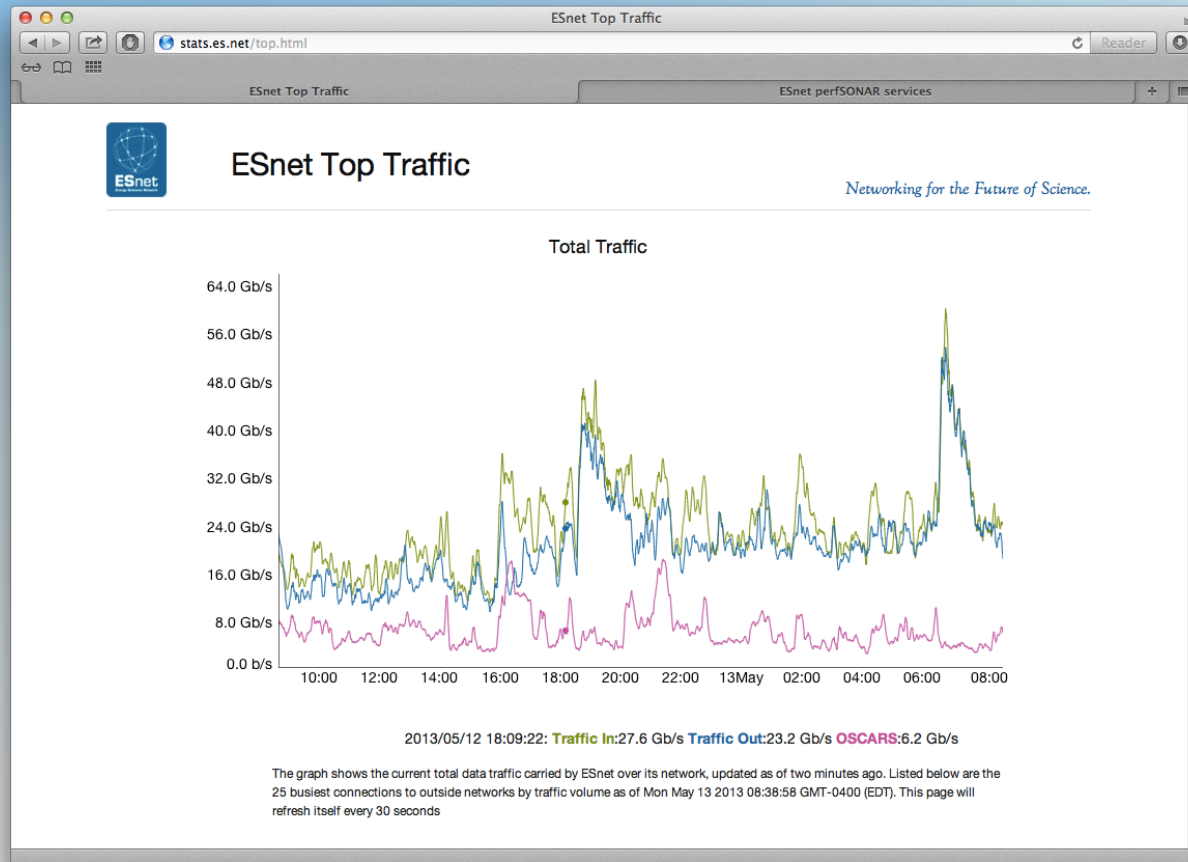
Measurement Points – ESnet FasterData

- <http://fasterdata.es.net>



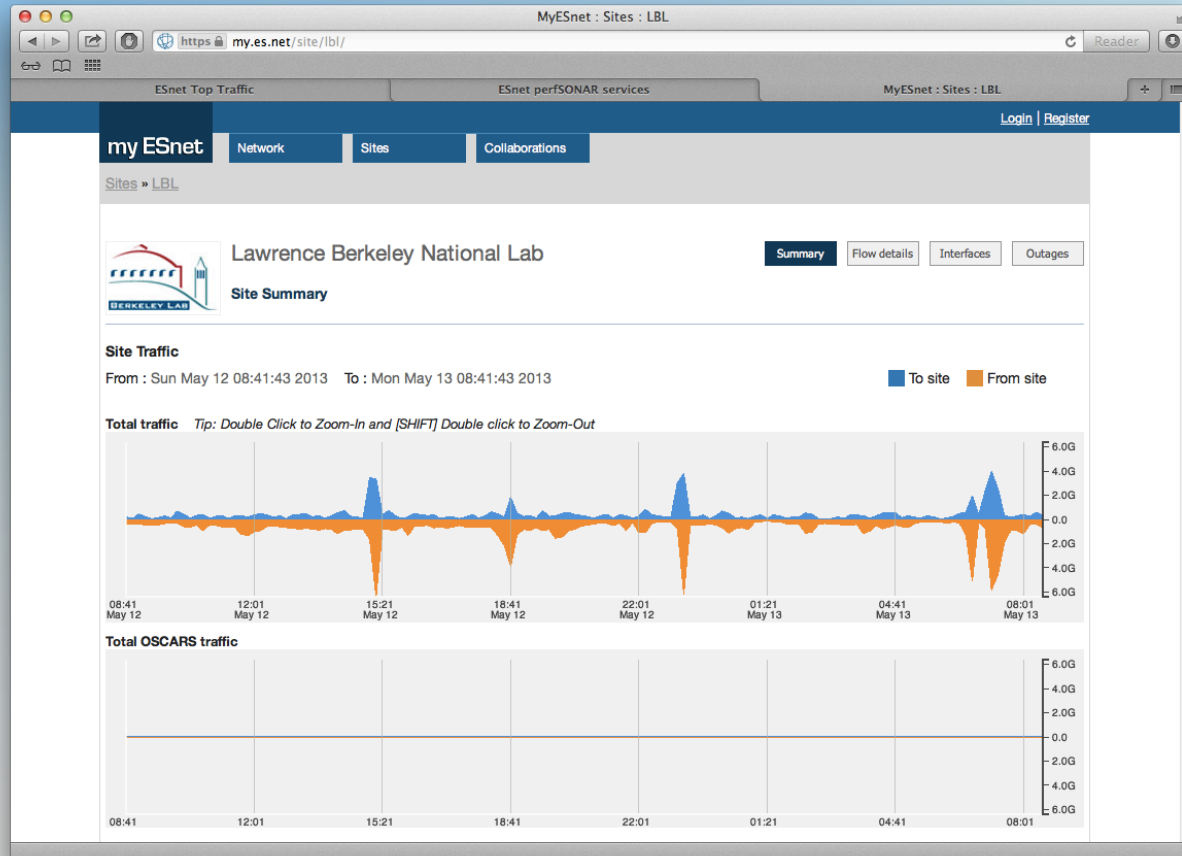
Measurement Points – ESnet Stats

- <http://stats.es.net>



Measurement Points – ESnet Portal

- <http://my.es.net>



Outline

- XSEDE Network Architecture
- Measurement Points
 - XSEDE
 - Internet2
 - ESnet
- XSEDE GridFTP speedpage
- ESnet IO Testers

XSEDE GridFTP “speedpage”

- Carried over from TeraGrid era
- Mesh of GridFTP tests that run twice daily between XSEDE SP GridFTP servers
- Test characteristics:
 - Automated
 - Run at ~2amET and ~2pm ET
 - /dev/zero -> /dev/null
 - Local tests with no file system or network overhead
 - file -> /dev/null and /dev/zero -> file
 - Local tests, taking into account file system read and write
 - Multi-stream transfer 4GB
- <http://speedpage.psc.edu>

Outline

- XSEDE Network Architecture
- Measurement Points
 - XSEDE
 - Internet2
 - ESnet
- XSEDE GridFTP speedpage
- ESnet IO Testers

ESnet IO Testers

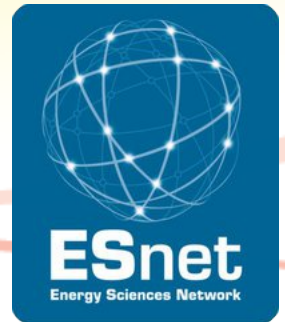
- <http://fasterdata.es.net/performance-testing/esnet-io-testers/>
- Test hosts for high-speed disk-to-disk testing
- Capable of 35 Gbps network throughput
 - 4 10GE NICS, only 1 of which is currently connected to ESnet
 - 16 Gbps disk read, and 12 Gbps disk write
- BWCTL for network testing
- Anonymous read-only GridFTP server for disk-to-disk
- These services are available to anyone on an R&E network anywhere in the world.
 - lbl-diskpt1.es.net / lbl-diskpt1-v6.es.net: Berkeley, CA
 - anl-diskpt1.es.net / anl-diskpt1-v6.es.net: Near Chicago, IL
 - bnl-diskpt1.es.net / bnl-diskpt1-v6.es.net: Near NY, NY

ESnet IO Testers

- Installing globus-url-copy:
<http://fasterdata.es.net/data-transfer-tools/gridftp/>
- Sample commands:
 - #make sure you can connect to server
 - globus-url-copy -list ftp://lbl-diskpt1.es.net:2811/data1/
 - # copy 1G file
 - globus-url-copy -vb -fast ftp://lbl-diskpt1.es.net:2811/data1/1G.dat file:///tmp/test.out
 - # copy 1G file using 4 parallel streams
 - globus-url-copy -vb -fast -p 4 ftp://lbl-diskpt1.es.net:2811/data1/1G.dat file:///tmp/test.out
 - # write to /dev/null
 - globus-url-copy -vb -fast -p 4 ftp://lbl-diskpt1.es.net:2811/data1/1G.dat file:///dev/null
 - # read from /dev/zero
 - globus-url-copy -vb -fast -p 4 -len 1G ftp://lbl-diskpt1.es.net:2811/dev/zero file:///tmp/t.out
- Each host has three 1.5 TB disk arrays, mounted as /data1, /data2, and /data3. /data1 and /data2 host test data sets (e.g. you will find test data there), while /data3 is typically used for ad-hoc testing by ESnet staff.

ESnet IO Testers

- Running disk to disk tests between the Berkeley host and the Chicago host using jumbo frames (MTU=9KB) gives these results:
 - 1 stream 175 MB/sec (1.4 Gbps)
 - 4 streams, 680 MB/sec (5.4 Gbps)
 - 8 streams, 820 MB/sec (6.6 Gbps)
- **Note:** Sites with firewalls will need to open the ports used by GridFTP (2811 and 50000-51000) and bwctl/iperf (4823 and 5001-5050). If you have a firewall that blocks outgoing connections, you may need to set the environment variable GLOBUS_TCP_SOURCE_RANGE as well. For more information see the **Globus GridFTP client firewall information** guide.



XSEDE Network Architecture Overview and Measurement Resources

July 22nd 2013, XSEDE Network Performance Tutorial

Jason Zurawski – Internet2/ESnet

Kathy Benninger - Pittsburgh Supercomputing Center

For more information, visit <http://www.internet2.edu/workshops/npw>