Making the Most out of Bandwidth: Networking in OSG March 16, 2016 Shawn McKee



0

Networks Supporting Science

- While some of us are interested in (or worried about!) networks it is fair to say most scientists would rather not have to think about them.
 - Ideally networks are "transparent" and always do the right thing, allowing data to move as fast as possible, from anywhere to anywhere at anytime ⁽³⁾
 - As Irene noted this morning: data must be able to move!
- The challenge is twofold:
 - Networks underlie all our distributed infrastructures and <u>must work well</u> for us to use our grids, clouds and HPC resources.
 - Problems in the network can be very hard to identify, isolate and fix
- OSG is working to better monitor, manage and diagnose our networks for all our benefit.



OSG Networking Area Mission

- OSG Networking was added at the beginning of OSG's second 5-year period in 2012
- The "**Mission**" is to have OSG become the network service data **source** for its constituents
 - Information about network performance, bottlenecks and problems should be easily available.
 - Should support our VOs, users and site-admins to find network problems and bottlenecks.
 - **Provide network metrics** to higher level services so they can make informed decisions about their use of the network (Which sources, destinations for jobs or data are most effective?)
- The GOAL: to make the most out of the bandwidth (network) we have! How?



Components of OSG Networking

Network Monitoring via perfSONAR

- Having perfSONAR fully deployed with a global dashboard is giving us powerful options for better management and use of our network
- A network datastore host all network metrics

• Tools to manage and maintain our infrastructure

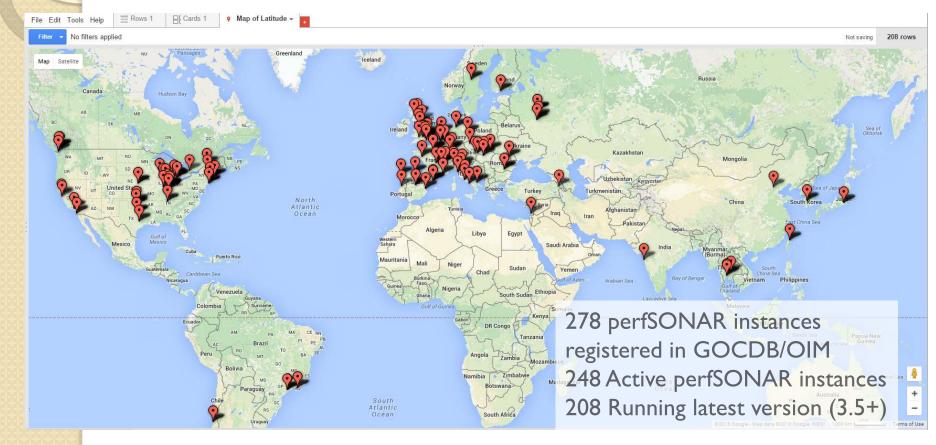
- A Modular dashboard (MaDDash); critical for "visibility" into networks. We can't manage/fix/respond-to problems if we can't "see" them.
- OMD/Check_mk (used to monitor and verify the state of many globally distributed perfSONAR services); required to maintain the overall proper functioning of the monitoring infrastructure.
- The development of the "mesh-configuration" and corresponding GUI interface; critical to creating a scalable, manageable deployment for WLCG/OSG
- Documentation --- Installation, debugging, How-tos
- Outreach and Support
 - With the network R&E community, VOs, software developers
 - OSG Support provides network ticket triage and routing



Open Science Grid

A Global Monitoring Infrastructure

• We enabled a global deployment of perfSONAR to instrument our networks (both IPv4 & IPv6)



http://grid-monitoring.cern.ch/perfsonar_report.txt for stats (updated daily)

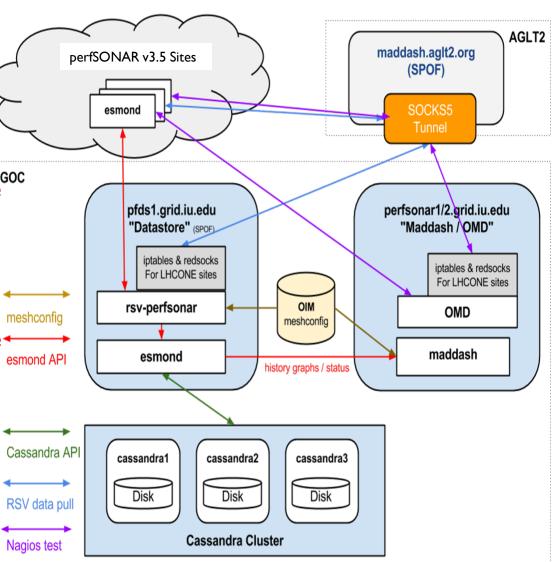


OSG Network Datastore

A <u>critical component is</u> <u>the datastore</u> to organize and store the network metrics and associated metadata

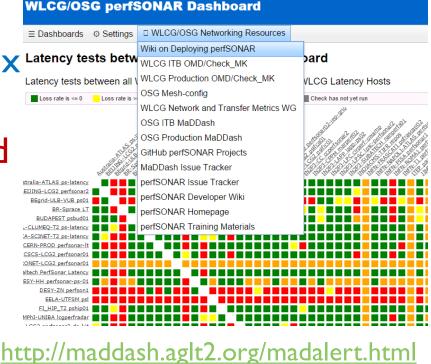
- OSG is gathering relevant metrics from the complete set of OSG and WLCG perfSONAR instances
- This data is available via an API, can be visualized and is organized to provide the "OSG Networking Service"
- In production sinceSeptember 2015





Monitoring Metrics

- Use MaDDash to view metric summaries
 - Provide quick view about how networks are working
- OSG hosts a production instance at: <u>http://psmad.grid.iu.edu/maddash-webui/</u>
- Metrics are displayed via source-destination matrix Lat
- Multiple dashboards (meshes) can be selected
- Custom menus link to relevant resources
- New release (2.0) will incorporate MadAlert





OMD/Check_MK Service Monitoring

We are using OMD & Check_MK to monitor our perfSONAR hosts and services. Provides useful overview of status/problems <u>https://psomd.grid.iu.edu/WLCGperfSONAR/check_mk/</u> [Requires x509 in your browser; update to InCommon authentication soon]

← → C 🕑 bttps://psomd.grid.iu	edu/ WLCGperfS	ONAR/check_mk/index.p	y?start_	url=9	62FW	/LCGp	erfSC	DNA	R%2Fcheck_mk%2Fview.py%	3Fview_	name	%3E	allhos	ts%26	selection%3Db3483332 숬 🐇	¥ 📶 Ş	71	2	* (BI
_Check ´** \$	All hosts								251 rows /DC=com/DC	=DigiCe	rt-Grid	//O=C)pen Sc	ience (irid/OU=People/CN=Shawn McKe	e 83467 (a	dmin)	06	:43 【	3
Tactical Overview × Hosts Problems Unhandled		3 30s 🛛 🐣 Availabili	ty 📄																	
251 30 30	state	Host	Icons	OK W		Cr Pd	sta	ate					Cr Pd							Pd
Services Problems Unhandled 3084 725 725	UP atlas-np	t2.bu.edu	🕏 🛧	10 2	2 0	0 0	U	JP	atrogr007.nipne.ro	🕏 🛧	10	2 0	0 0	U	atrogr009.nipne.ro	🔹 🛧	10	2	0 0	0
– Quicksearch x	UP ccperfso	nar1.in2p3.fr	\$ 🕂	10 2	2 0	0 0	U	JP	ccperfsonar2.in2p3.fr	🕏 🛧	10	2 0	0 0	U	clrperf-bwctl.in2p3.fr	😵 🛧	10	2	0 0	0
	UP cirperf-o	wamp.in2p3.fr	🛸 🛧	10 2	2 0	0 0	U	JP	grid-perf1.physik.rwth-aachen.de	🛸 🛧	10	2 0	0 0	U	grid-perf2.physik.rwth-aachen.de	i 🖓 🛧	10	2	0 0	0
bnl 🔍	UP grid-per	sonar.hpc.susx.ac.uk	🕏 🛧	11 4	4 0	0 0	U	JP	gridpp-ps-band.ecdf.ed.ac.uk	😵 🛧	10	2 0	0 0	U	gridpp-ps-lat.ecdf.ed.ac.uk	ø 🛧	10	2	0 0	0
incmon.oni.gov Ihoperfmon.bnl.gov ps-development.bnl.gov	UP hcc-ps01	1.unl.edu	\$ 4	10 2	2 0	0 0	U	JP	hcc-ps02.unl.edu	24	10	2 0	0 0	U	hepinx129.pp.rl.ac.uk	24	10	2	0 0	0
– Views x	UP heplnx1	30.pp.rl.ac.uk	34	10	2 0	0 0	U	JP	ingrid-ps01.cism.ucl.ac.be	34	10	2 0	0 0	U	ingrid-ps02.cism.ucl.ac.be	24	10	2	0 0	0
	UP iut2-net1	.iu.edu	24	10 2	2 0	0 0	U	JP	iut2-net2.iu.edu	24	10	2 0	0 0	U	lapp-ps01.in2p3.fr	24	10	2	0 0	0
 ► Dashboards ▼ Hosts 	UP lapp-ps0)2.in2p3.fr	24	10 2	2 0	0 0	U	JP	lcgnetmon.phy.bris.ac.uk	24	10	2 0	0 0	U	lcgnetmon02.phy.bris.ac.uk	24	10	2	0 0	0
All hosts All hosts (Mini)	UP logperfra	adar.dnp.fmph.uniba.sk	24	10 2	2 0	0 0	U	JP	lcgperfsonar.dnp.fmph.uniba.sk	24	10	2 0	0 0	U	lcgps01.gridpp.rl.ac.uk	24	10	2	0 0	0
All hosts (tiled) Favourite hosts	UP lcqps02.	gridpp.rl.ac.uk	24	10 2	2 0	0 0	U	JP	lcgs101.jinr.ru	24	10	2 0	0 0	U	lcgs102.jinr.ru	24	10	2	0 0	0
Host search ▼ Hostgroups	UP Ihc-band	lwidth.twgrid.org	24	10	2 0	0 0	U	JP	Ihcmon.bnl.gov	24		2 0	0 0	U	Ihcperfmon.bnl.gov	24	10	2	0 0	0
Hostgroups Hostgroups (Grid)	UP Ipnhe-ps	sb.in2p3.fr	~	10 2	2 0	0 0	U	јр	Ipnhe-psl.in2p3.fr	24	10	2 0	0 0	U	lpsc-perfsonar.in2p3.fr	24	10	2	0 0	0
Hostgroups (Summary) Services	UP lpsc-per	fsonar2.in2p3.fr	4 1	10 2	2 0	0 0	U	JP	lutos.lunet.edu	24	10	2 0	0 0	UI UI	lutos1.lunet.edu	÷		2	0 0	0
✓ Servicegroups Servicegroups (Grid)		-2.lsr.nectec.or.th	~		2 0		U	ID	mwt2-	\$ 4		2 0		U	mwt2-	\$ \$		2	0 0	
Servicegroups (Summary)			4		-				ps01.campuscluster.illinois.edu						ps02.campuscluster.illinois.edu	4		-		
Services by group Business Intelligence		s01.in2p3.fr				0 0	U		nanperfs02.in2p3.fr	😵 🛧		2 0		U	-	🕏 🕁		2		0
 Problems Addons 		2.atlas-swt2.org				0 0	U		nettest.lbl.gov	8 1		2 0		U		8 1		-		0
► Other		2.zeuthen.desy.de		10 2	2 0	0 0	U	л	perfsonar-bandwidth.esc.qmul.ac.uk	-		2 0	0 0	U	perfsonar-bw.cern.ch	🤹 🛧			0 0	0
	UP perfsona bw.tier2.	ar- hep.manchester.ac.uk		10 2	2 0	0 0	U	JP	perfsonar-cms1.itns.purdue.edu	😵 🛧	10	2 0	0 0	U	perfsonar-cms2.itns.purdue.edu	😵 🛧		2	0 0	0
– Bookmarks x	UP perfsona	ar-de-kit.gridka.de	ø 🛧	10 2	2 0	0 0	U	JP	perfsonar-latency.esc.qmul.ac.uk	🕏 🛧	10	2 0	0 0	U	perfsonar-lt.cern.ch	📚 🛧	10	2	0 0	0
Add Bookmark	UP perfsona	ar-It.tier2.hep.manchester.ac.uk	i 🖓 🛧	10 2	2 0	0 0	U	JP	perfsonar-ow.cnaf.infn.it	🕏 🛧	10	2 0	0 0	U	perfsonar-ps-01.desy.de	😵 🛧	10	2	0 0	0
- WATO · Configuration ×	UP perfsona	ar-ps-02.desy.de	🖈 🍣	10 2	2 0	0 0	U	JP	perfsonar-ps-bandwidth.igfae.usc.es	🕏 🛧	10	2 0	0 0	U	perfsonar-ps-latency.igfae.usc.es	ø 🛧	10	2	0 0	0
🏠 Main Menu	UP perfsona	ar-ps.cnaf.infn.it	\$ 🕂	10 2	2 0	0 0	U	JP	perfsonar-ps.ndgf.org	🕏 🛧	10	2 0	0 0	U	perfsonar-ps2.ndgf.org	🕏 🛧	10	2	0 0	0
	UP perfsona	ar.na.infn.it	📌 🍣	10 2	2 0	0 0	U	JP	perfsonar.unl.edu	🕏 🛧	11	4 0	0 0	U	perfsonar01-iep-grid.saske.sk	😵 🛧	10	2	0 0	0
	UP perfsona	ar01.ft.uam.es	24	10 2	2 0	0 0	U	JP	perfsonar02-iep-orid.saske.sk	<i>1</i>	10	2 0	0 0	🗏 ui	perfsonar02.ft.uam.es	÷	10	2	0 0	0



Detailed Service Checks

Services of Host ps-latency.clumeq.mcgill.ca 12 rows /DC=com/DC=DigiCert-Grid/O=Open Science Grid/OU=People/CN=Shawn McF 🔍 📜 🔨 🔲 2 30s 🕙 Availability ps-latency.clumeq.mcgill.ca Status detail Perf-O-Meter State Service R perfSONAR 3 4+ Toolkit Version OK tookit version found 3.4.2 2015-03-20 17:25:40 OK 15 sec 24 OK - Administrator is Simon Nderitu, email simon.nderitu@clumeq.ca (cached:0) 2014-12-11 19:57:00 OK perfSONAR Administrator Details 2 hrs OK perfSONAR esmond Freshness Latency Direct Service perfSONAR esmond Freshness Bandwidth Direct, psu... 1 row /DC=com/DC=DigiCert-Grid/0= OK perfSONAR esmond Freshness Latency Reverse perfSONAR esmond Measurment Archive OK 🔍 🔎 🔨 🔽 🚺 30s 🕙 Availability perfSONAR Homepage OK OK perfSONAR Latitude/Longitude Configured Site alias psum02.aglt2.org OK perfSONAR Mesh Configuration Hostname Service description perfSONAR esmond Freshness Bandwidth Direct OK perfSONAR NTP Service Service icons Z OK perfSONAR OWAMP One-Way Ping Service WARN Service state perfSONAR Regular Testing Service OK Servicegroups the service is member of sg esmond bw direct, sg esmond psum02.aglt2.org State Service service level OK perfSONAR 3 4+ Toolkit Version Service contact groups all Service contacts OK perfSONAR Administrator Details WARNING Found stale hosts for certain events, time-range: 3700 Output of check plugin perfSONAR BWCTL Bandwidth Test Controller OK Long output of check plugin (multiline) Time-range: 3700 WARN perfSONAR esmond Freshness Bandwidth Direct Even-types checked: packet-trace Mesh (Event-type): GOC Traceroute Tests (packet-trace) WARN perfSONAR esmond Freshness Bandwidth Reve Destinations count: 1 perfSONAR esmond Measurment Archive Missing destinations: perfsonar-cms2.itns.purdue.edu OK Mesh (Event-type): Traceroute Test Between WLCG Bandwidth Hosts (packet-trace) OK perfSONAR Homepage Destinations count: 10 Missing destinations: marperf01.in2p3.fr, perfsonar.pleiades.uni-wuppertal.de, ps02.cat.cbpf.br, perfs OK perfSONAR Latitude/Longitude Configured ps02.cism.ucl.ac.be, clrperf-bwctl.in2p3.fr, perfsonar-cms2.itns.purdue.edu, perfsoar2.hep.kbfi.ee OK perfSONAR Mesh Configuration Documentation for this check can be found at https://twiki.opensciencegrid.org/bin/view/Documentati OK perfSONAR NTP Service OK N I P synchronized 2015-01-29 20:17:01 52 min 🐲 perfSONAR Regular Testing Service OK Regular Testing enabled and running 2015-02-19 10:51:38 51 min OK OK perfSONAR Toolkit Version OK - Version 3.4.2 OK (cached:1) 2014-12-11 19:56:41 2 hrs



Open Science Grid

Existing Tools

- We have a number of tools deployed and available to help debug and understand network problems.
- There are very good presentations on these tools in the training materials provided by perfSONAR: <u>http://www.perfsonar.net/about/training-materials/</u>
- While I don't have time to cover all the details (see http://www.perfsonar.net/about/training-materials/201601-ps-training/ and especially the Measurement Tools, Use Cases and Debugging presentations from Jason Zurawski) I do want to note that command line tools exist to allow you to create on-demand 3rd party tests (between two remote instances) for bandwidth, latency and traceroute.
 - Follow the debugging strategy as a guide to finding and fixing OSG network issues using perfSONAR capabilities
- As for new tools....



Managing perfSONAR Deployments

- OSG originally developed a "mesh-config" GUI built within the OIM/MyOSG framework
 - We provided a GUI to define and organize the regularly scheduled tests between specific sets of perfSONAR instances.
 - The mesh-config was a huge benefit; no longer need to use email to hundreds of system admins to make changes to network tests and their organization. The GUI made changes easy and consistent.
- **Problem**: not able to be made easily available to others within or outside OSG.
 - Campuses deploying many perfSONARS
 - Science VOs wanting to organize/customize their perfSONARs
- Soichi Hayashi/OSG hasproduced a new standalone package which provides an even more feature-rich mesh-configuration GUI
 - Will be integrated with perfSONAR v3.6
 - See beta release info at <u>http://soichi7.ppa.iu.edu/pdoc/mca.html#</u>
 - Eventually will replace current GUI



MadAlert: A new project to analyze meshes

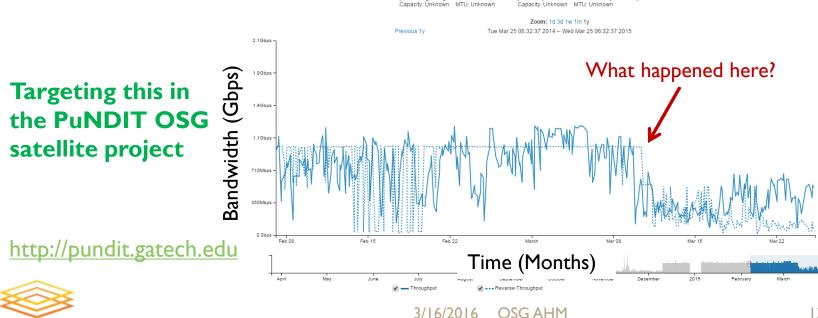
- Gabriele Carcassi/UM has been working with me on creating a new utility to analyze meshes: MadAlert
- See details at <u>http://madalert.aglt2.org/madalert/index.html</u>
 - You can see meshes and reports from the page
 - Reports find both infrastructure and network problems
- We are now working with Andy Lake/ESnet to incorporate this into the next major release of MaDDash (v2.0) due out later this year.
- Now testing a "diff" to allow us to compare meshes; e.g., IPv4 vs IPv6, testing vs production, mesh(t1) vs mesh(t2)
 - <u>http://madalert.aglt2.org/madalert/testDiff.html</u>
 - Could be really helpful for understanding new software versions or changes in time. Time based comparison will require some modifications to MaDDash to allow specifying point-in-time meshes.



OSG Network Alerting

- What kinds of capabilities can we enable given a rich datastore of historical and current network metrics?
 - Users want "someone" to tell them when there is a network problem involving their site or their workflow.
 - Can we create a framework to identify when network problems occur and locate them? (**Must** minimize the false-positives).

psum02.aglt2.org - 192.41.230.60



Open Science Grid

Link to this chart

ps2.ochep.ou.edu - 129 15 40 232 [traceroute

OSG Networking and End-to-end

- Most scientists just care about the **end-to-end** results:
 - How well does their infrastructure support them in doing their science?
- Network metrics allow OSG to differentiate end-site issues from network issues.
- There is an opportunity to do this better by having access to end-to-end metrics to compare & contrast with network-specific metrics.
 - What end-to-end data can OSG regularly collect for such a purpose?
 - Is there some kind of common instrumentation that can be added to some data-transfer tools? (NetLogger in GridFTP, having transfers "report" results to the nearest perfSONAR-PS instance?, etc)
 - Let's try to put all the information we have together...

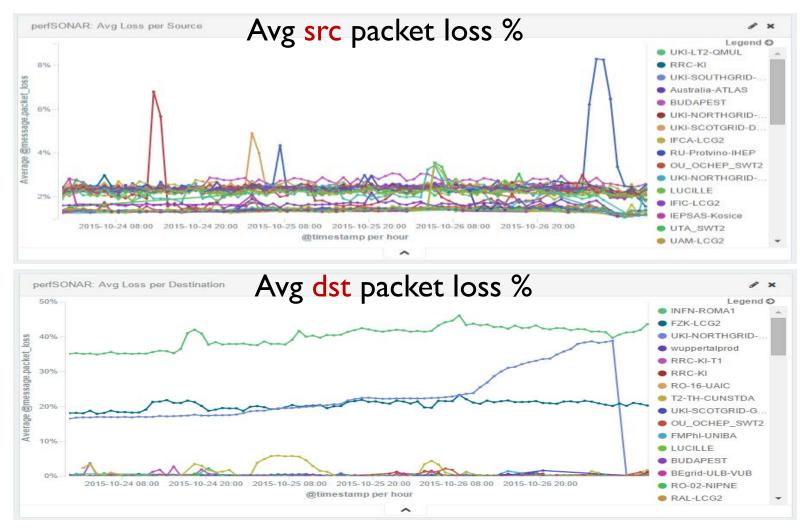


ATLAS Network Metrics Pipeline

- Ilija Vukotic, Kaushik De, Rob Gardner and Jorge Batista are working with the Network and Transfer Metrics WG to make OSG perfSONAR metrics available to PANDA
 - See Ilija's presentation at <u>http://tinyurl.com/gt92zwb</u>
- **Pipeline**: OSG Network Datastore to CERN Active MQ to Flume to ElasticSearch to PANDA
- Pipeline is operating and analysis has been performed in ElasticSearch to validate our data.
- Working on a network source-destination cost-matrix PANDA can use to evaluate options
 - Actual interface details being discussed/developed with the PANDA team
- Potentially a very powerful, customizable analysis framework to use all available metrics from the network



perfSONAR Data in ElasticSearch



http://tinyurl.com/OSGNetES for other examples using OSG data



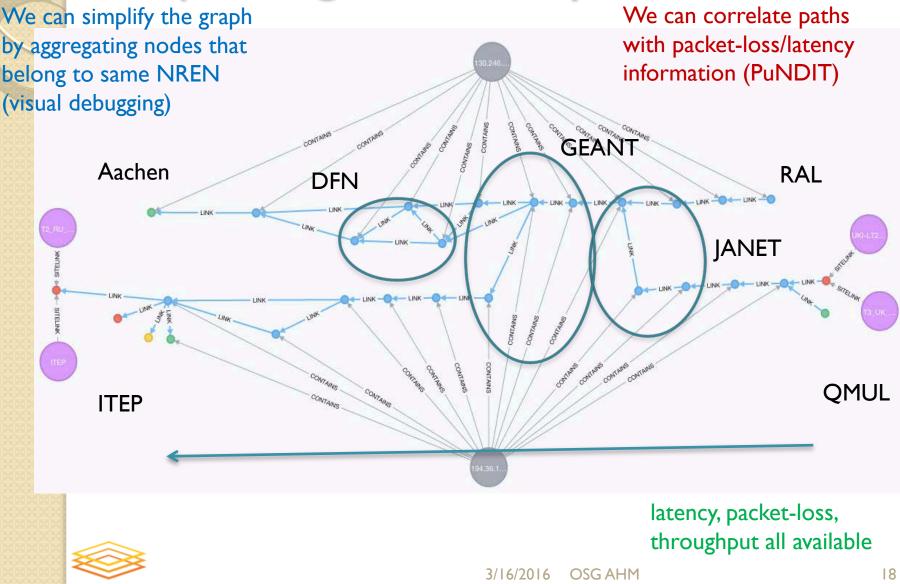
Understanding Network Topology

- One thing that may not be obvious is that all the network measurements are only really useful if we know the path being measured...topology is very important to find/fix problems.
 - OSG measures topology via traceroute and tracepath
- Can we create tools to manipulate, visualize, compare and analyze <u>network topologies</u> from the OSG network datastore contents?
- Can we build upon these tools to create a set of nextgeneration network diagnostic tools to make debugging network problems easier, quicker and more accurate?
- Even without requiring the ability to perform complicated data analysis and correlation, basic tools developed in the area of network topology-based metric visualization would be very helpful in letting users and network engineers better understand what is happening in our networks.
- This area is under active investigation in various projects. Lots of work to do here.



Exploring Path Analysis

Open Science Grid



Software Defined Networks and OSG

- Within the next few years evolving technology in the area of **Software Defined Networking(SDN)** may be able to provide OSG researchers with the ability to construct their own Wide-Area networks with specified characteristics.
- What will OSG be able to do to integrate this type of capability with the rest of the OSG infrastructure?
- We are planning for how best to enable evolving capabilities in the network for OSG users and admins.We need to address:
 - What is the impact on the OSG software stack?
 - What strategic modifications/additions/tools are useful?



Summary

- We have made significant strides in making the network "visible"
 - ~250 OSG/WLCG perfSONAR deployments globally
 - All monitored, managed and orchestrated by OSG
 - Tools to manage and maintain our infrastructure are in place
 - We have production datastore providing long term access to all metrics
- New opportunities to improve our networks and their use are possible because of the unique set of data we have
 - Exploiting the rich dataset we have is underway.
 - To identify network problems and do "targeted alerting" on them is a high priority.
 - To inform and enable higher level services, researchers and users



Questions or Comments?

Thanks!





Extra Slides

Network Problem Debugging(1/4)

- Problem with transfers from SARA to AGLT2 noted February 20th.
 - FTS transfers failing. ATLAS asked about network. Dataset had large files; transfers failed because of timeout (I MB/sec=3.6GB/hour; I hour timeout)
 - Setup 'Debug' mesh using OSG tools to track SARA, CERN to AGLT2, MWT2
 Debug Mesh (temp) - Debug LT
 - Jason Zurawski ran tests using perfSONAR. Bad throughput ~few hundred kbits/sec SARA->AGLT2 (real network issue!)





Network Debugging (2/4)

- Ticket opened by AGLT2 with Internet2 NOC on February 20th
 - Poor iperf results also between SARA and MWT2
 - Routes provided both ways by perfSONAR

Topology beginning at Tue Mar 10 00:01:17 2015 (UTC -4)

Hop	Router	IP	Delay	MTU
1	ge-0-2-0-1020.grid-r1.grid.sara.nl	145.100.17.1	0.16ms	
2	geant-lhcone-gw.mx1.ams.nl.geant.net	62.40.126.161	0.279ms	
3	et-10-0-0.3019.rtr.newy32aoa.net.internet2.edu	64.57.30.225	74.946ms	
4	192.17.10.74	192.17.10.74	105.154ms	
5	64.57.30.154	64.57.30.154	120.308ms	
6	psum02.aglt2.org	192.41.230.60	107.806ms	

Topology beginning at Tue Mar 10 13:46:17 2015 (UTC -4)

Hop	Router	IP	Delay	MTU
1	ge-0-2-0-1020.grid-r1.grid.sara.nl	145.100.17.1	0.158ms	
2	geant-lhcone-gw.mx1.ams.nl.geant.net	62.40.126.161	0.268ms	
3	et-10-0-0.3019.rtr.newy32aoa.net.internet2.edu	64.57.30.225	87.094ms	
4	192.17.10.74	192.17.10.74	102.803ms	
5	requestTimedOut	requestTimedOut		
6	psum02.aglt2.org	192.41.230.60	108.808ms	



Topology beginning at Tue Mar 10 00:40:32 2015 (UTC -4)

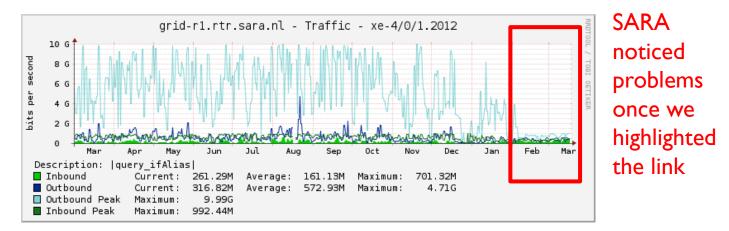
Hop	Router	IP	Delay	MTU
1	192.41.230.1	192.41.230.1	0.266ms	
2	et-8-0-0.3151.rtr.chic.net.internet2.edu	64.57.30.155	5.796ms	
3	et-10-0-0.402.rtr.newy32aoa.net.internet2.edu	64.57.30.142	33.651ms	
4	64.57.30.228	64.57.30.228	119.977ms	
5	surfnet-lhcone-gw.ams.nl.geant.net	62.40.126.160	119.973ms	
6	ps.lhcopn-ps.sara.nl	145.100.17.9	119.803ms	

Topology beginning at Tue Mar 10 13:43:23 2015 (UTC -4)

Hop	Router	IP	Delay	MTU
1	192.41.230.1	192.41.230.1	0.235ms	
2	esnet-lhc1-a-aglt2.es.net	198.124.80.53	6.266ms	
3	62.40.126.149	62.40.126.149	109.678ms	
4	62.40.126.148	62.40.126.148	108.904ms	
5	surfnet-lhcone-gw.ams.nl.geant.net	62.40.126.160	119.086ms	
6	ps.lhcopn-ps.sara.nl	145.100.17.9	108.313ms	

Network Debugging (3/4)

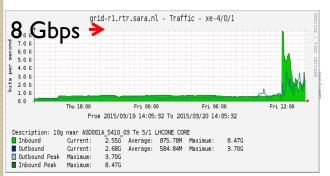
- Internet2 initially pursuing asymmetric routes and link congestion in US
- I2 opened ticket with GEANT Feb 27
- GEANT brought up LHCONE perfSONAR in Frankfurt
 - Tests to SARA(close) and AGLT2(far) showed 3x better bandwidth to AGLT2. Problem close to SARA
 - Isolated link between SARA and GEANT March 4





Problem Found/Fixed – Mar 20

- 2 weeks of link-debug
- Problem identified and fixed Mar 20
 - Bandwidth "policing" for 0 LHCONE. More than a year ago, LHCONE setup I Gbps. Never enforced.
 - Turned on policing start of 0 February.
 - Changed BW 10Gbps 0 March 20...fixed





perfSONAR Throughput SARA to AGLT2 (I month)

