

WLCG Network Monitoring using perfSONAR-PS

Shawn McKee/University of Michigan
For the WLCG perfSONAR-PS Task-Force
HEPiX Spring 2013 / INFN Bologna, Italy
April 17, 2013

Motivations for Common LHC Network Monitoring

- ❄ LHC 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 no differentiation of how the network is used amongst the LHC experiments. (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!**
- ❄ We can all benefit from all working with a “standard” in terms of development, maintenance and support.
- ❄ Much easier to self-support with a common toolkit

Network Monitoring for WLCG: Goals/Purpose

❄ Goals:

- ❄ Find and isolate “network” problems; alerting in a timely way
 - ❄ Characterize network use (base-lining)
 - ❄ Provide a source of network metrics for higher level services
- ❄ **First step:** get monitoring in place to create a baseline of the current situation between sites (see details later)
- ❄ **Next:** continuing measurements to track the network, alerting on problems as they develop
- ❄ Choice of a standard “tool/framework”: **perfSONAR**
- ❄ **perfSONAR’s main purpose is to aid in network diagnosis** by quickly allowing users to isolate the location of problems. **In addition it can provide a standard measurement of various network performance related metrics over time as well as “on-demand” tests.**

History of perfSONAR for WLCG

- ❄ **perfSONAR**: a joint effort of ESnet, Internet2, GEANT and RNP to standardize network monitoring protocols, schema and tools
- ❄ USATLAS adopted **perfSONAR-PS** toolkit starting in 2008. All Tier-2s and the Tier-1 instrumented by 2010.
- ❄ LHCOPN choose to adopt in June 2011; deployed within 3 months (by September 2011).
- ❄ Used to monitor initial transition of selected LHCONE sites
- ❄ **Now targeting full deployment at all WLCG Tier-2 sites worldwide (~150 locations)**

WLCG perfSONAR-PS Deployments

- ❄ We want to measure (to the extent possible) the entire network path between WLCG resources. This means:
 - ❑ We want to locate perfSONAR-PS instances as close as possible to the storage resources associated with a site. The goal is to ensure we are measuring the same network path to/from the storage.
- ❄ There are two separate instances that should be deployed: **latency and bandwidth**
 - ❑ The **latency instance** measures one-way delay by using an NTP synchronized clock and send 10 packets per second to target destinations. We also **traceroute** using this instance.
 - ❑ The **bandwidth instance** measures achievable bandwidth via a short test (30 seconds) per src-dst pair every ~n hour period
- ❄ OK, once these are in place we are set...right?
- ❄ Not so fast 😊 Need to configure scheduled tests and make sure we can use the metrics being collected....

Configuration of Network Testing

- ❄ One of the lessons learned from LHC use of perfSONAR-PS was that setting up and maintaining scheduled tests for the perfSONAR-PS toolkit instances was a challenge.
 - ❄ As sites changed, joined or left, every other site needed to update their configuration to change, add or remove tests.
 - ❑ Labor intensive, slow to get all updates in place and gets worse as we increase the size of the deployments!
 - ❄ Aaron Brown/Internet2 provided a solution: the “mesh” configuration which allows sites to track a central configuration and update themselves when it changes.
- http://www.usatlas.bnl.gov/twiki/bin/view/Projects/PerfSONAR_PS_Mesh
- ❄ For current perfSONAR-PS deployments a set of patches and new RPMS needs to be installed to enable the mesh-config
 - ❄ V3.3 (out very soon) will have all functionality for the mesh built-in

Overview of Mesh-Configuration

- ❄ perfSONAR-PS toolkit instances can get their configuration information from a URL hosting an suitable JSON file
 - A simple `agent_configuration.conf` file defines one or more URLs
- ❄ The perfSONAR-PS instances can participate in more than one configuration (WLCG, Tier-1 cloud, VO-based, etc.)
- ❄ The “Mesh-configuration” JSON file specifies which hosts, tests and test parameters are used.
- ❄ Changes to the tests or test-members can be made in one location and quickly reflected in all test sites.
- ❄ The WLCG mesh configurations are hosted in AFS/HTTP at
 - `/afs/cern.ch/project/gd/wlcg-ops/perfsonar/conf/`
 - `https://grid-deployment.web.cern.ch/grid-deployment/wlcg-ops/perfsonar/conf/`.

Modular Dashboard: Centralized Info

- ❄ Having a large number of perfSONAR deployments is great for instrumenting our networks, but all these instances are not easy to track, summarize or use to better understand the overall network performance and status.
- ❄ The current modular dashboard (developed by Tom Wlodek at BNL) is being used to track a large number of LHC perfSONAR-PS installations:

<http://perfsonar.racf.bnl.gov:8080/exda/>

The dashboard provides a highly configurable interface to monitor a set of perfSONAR-PS instances via simple plug-in test modules. Users can be authorized based upon their grid credentials. Sites, clouds, services, tests, alarms and hosts can be quickly added and controlled.

- ❄ Very good for the user community but needed recoding...

Example of Dashboard showing LHCONE

← → ↻ <https://perfsonar.racf.bnl.gov:8443/exda/?page=25&cloudName=LHCONE> 🔍 ☆ ☰

RACF
Grid Group

The Production Instance of perfSONAR Dashboard

Status as of: Mon Dec 10 10:02:47 EST 2012

Cloud LHCONE

Sites of LHCONE cloud

| | | | | | |
|----------|---------|-------------|------------|-------|---------|
| BNL | AGLT2 | INFN Napoli | SARA | ASGC | PIC |
| KIT | TRIUMF | Toronto | Prague | Tokyo | LRZ-LMU |
| GRIF-LAL | DESY-HH | MWT2(UC) | GRIF/LPNHE | | |

Main Page

All Clouds

Individual Clouds:

- USATLAS
- USCMS
- IT
- LHCOPN
- LHCONE
- CA-ATLAS
- UK
- LHC-FR

Inter Cloud Tests:

- AGLT2-IT
- FR-US
- ATLAS-UK

Primitive Services

- perfSonar Sites
- List of Hosts
- List of Matrices
- List of Alarms
- List of Clouds
- List of Sites
- List of Schedulers
- Probes
- Manage Users
- Define or Edit Alarms

RACF dashboard

perfSONAR dashboard (old)

RACF dashboard (test)

perfSONAR dashboard (old, test)

Dashboard documentation

LHCONE Throughput Matrix

| | --- | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|--------------------------------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|----|
| 0:BNL (lhcmon.bnl.gov) | --- | 1.34 | 0.28 | 0.08 | 0.00 | 0.33 | 0.05 | 0.00 | 0.10 | 0.13 | 0.02 | 0.14 | 0.04 | 0.01 | 0.02 | 0.11 | 0.37 | 0.11 | |
| 1:AGLT2 (psmsu02.agit2.org) | 1.78 | --- | 0.05 | 0.11 | 0.34 | 0.28 | 0.04 | 0.05 | 0.07 | 0.19 | 0.02 | 0.03 | 0.28 | 0.01 | 0.00 | 0.10 | 0.38 | 0.22 | |
| 2:AGLT2 (psum02.agit2.org) | 0.98 | 0.98 | --- | 0.07 | 0.24 | 0.24 | 0.05 | 0.00 | 0.08 | 0.13 | 0.03 | 0.07 | 0.27 | 0.01 | 0.00 | 0.05 | 0.27 | 0.28 | |
| 3:ASGC (lhc-bandwidth.twgrid.org) | 0.24 | 0.24 | 0.24 | --- | 0.16 | 0.16 | 0.02 | 0.02 | 0.04 | 0.06 | 0.01 | 0.11 | 0.03 | 0.01 | 0.00 | 0.11 | 0.24 | 0.24 | |
| 4:CERN | 0.00 | 0.01 | 0.01 | 0.01 | --- | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.01 | |

LHCONE Latency

| |
|------------------------------------|
| 0:BNL (lhcperfmon.bnl.gov) |
| 1:AGLT2 (psmsu01.agit2.org) |
| 2:AGLT2 (psum01.agit2.org) |
| 3:ASGC (lhc-latency.twgrid.org) |

See <https://perfsonar.racf.bnl.gov:8443/exda/?page=25&cloudName=LHCONE>

perfSONAR Modular Dashboard Efforts

- ❄ There is an ongoing effort to produce the next generation of the dashboard which is scalable and preserves the existing functionality of the current dashboard
 - ❑ We have setup a new organization in GitHub: 'PerfModDash'
 - ❑ Code is now in GitHub under a modified BSD license:
<https://github.com/PerfModDash/PerfModDash>
 - ❑ **Everyone interested is encouraged to participate**
- ❄ We need to discuss how to leverage this project for WLCG. OSG (as a subset of WLCG) plans to deploy this as a service to gather and provide network metrics for OSG sites
 - ❑ Should WLCG also bring up such a service?

Example of New Dashboard

<http://perfsonar.racf.bnl.gov:8080/PsDisplay-1.0-SNAPSHOT/matrices.jsp?id=1>



Popular Clouds
USATLAS

Matrix:USATLAS Packet Loss

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|----------------------------------------|---|---|---|---|---|---|---|---|---|---|
| 0:lhcpfmon.bnl.gov | | | | | | | | | | |
| 1:psmsu01.aglt2.org | | | | | | | | | | |
| 2:psum01.aglt2.org | | | | | | | | | | |
| 3:iut2-net1.iu.edu | | | | | | | | | | |
| 4:uct2-net1.uchicago.edu | | | | | | | | | | |
| 5:mwt2-ps01.campuscluster.illinois.edu | | | | | | | | | | |
| 6:atlas-npt1.bu.edu | | | | | | | | | | |
| 7:ps1.ochep.ou.edu | | | | | | | | | | |
| 8:netmon1.atlas-swt2.org | | | | | | | | | | |
| 9:psnr-lat01.slac.stanford.edu | | | | | | | | | | |

Plans for WLCG Operations

- ❄ Simone Campana and I are leading a WLCG (Worldwide LHC Computing Grid) operations task-force for perfSONAR:
 - ❑ Encouraging all sites to deploy and register two instances
 - ❑ All sites to use the “mesh” configuration (central configs)
 - ❑ One set of test parameters to be used everywhere
 - ❑ Site instructions (draft) at <https://twiki.cern.ch/twiki/bin/view/LCG/PerfsonarDeployment>
- ❄ The current dashboard is being used as a central source for network information. This will continue but we need to make sure we are gathering the right metrics and making them easily accessible for our applications and infrastructure
 - ❑ We need to encourage discussion about the types of metrics our frameworks and applications would like concerning the network

Network Monitoring Deliverables

- ❄️ What does a perfSONAR deployment provide for us?
 - ❑ We get measurements of the **network** behavior along relevant paths
 - ❑ The system schedules non-conflicting tests between sites
- ❄️ The **latency measurements** provide one-way latency via NTP-synced clocks at each end.
 - ❑ More interesting is the measurement of packet-loss it provides. With 600 packets/minute we can see marginal paths via their loss metric
- ❄️ The **bandwidth measurements** deliver an estimate of achievable bandwidth along paths
 - ❑ Useful to set expectations and indicate problematic paths
 - ❑ Tracking versus time provides a way to identify when problems start
 - ❑ Comparison with ATLAS Dashboard can differentiate networks vs end-site problems
- ❄️ The **traceroute measurements** track routing changes

Network Monitoring Challenges Ahead

- ❄ Getting hardware/software platform installed at all sites
- ❄ **Dashboard development:** Need additional effort to produce something suitable quickly and ensure it meets our needs...
- ❄ Managing site and test configurations
 - ❑ Testing and improving “centralized” (VO-based?) configurations
 - ❑ Determining the right level of scheduled tests for a site, e.g., Tier-2s test to other same-cloud Tier-2s (and Tier-1)?
 - ❑ Address 10G vs 1G tests that give misleading results
- ❄ Improve path monitoring (traceroute) within the tool
- ❄ **Alerting:** A high-priority need but complicated:
 - ❑ Alert who? Network issues could arise in any part of end-to-end path
 - ❑ Alert when? Defining criteria for alert threshold. Primitive services are easier. Network test results more complicated to decide
- ❄ Integration with VO infrastructures and applications

Improving perfSONAR-PS Deployments

- ❄ Based upon the issues we have encountered we setup a Wiki to gather best practices and solutions to issues we have identified:
<http://www.usatlas.bnl.gov/twiki/bin/view/Projects/LHCperfSONAR>
- ❄ Improving resiliency (set-it-and-forget-it) a high priority. Instances should self-maintain and the infrastructure should be able to alert when services fail
- ❄ This page is being shared with the perfSONAR-PS developers and we expect many of the “fixes” will be incorporated into future releases (most are in v3.3 already)
- ❄ Please feel free to add to the Wiki (either directly or by emailing me updates/changes/additions).

Facility Use of Network Metrics

- ❄️ Once we have a source of network metrics being acquired we need to understand how best to incorporate those metrics into our facility operations.
- ❄️ Some possibilities:
 - ❑ Characterizing paths with “costs” to better optimize decisions in workflow and data management
 - ❑ Noting when paths change and providing appropriate notification
 - ❑ Optimizing data-access (FAX) or data-distribution (DDM) based upon a better understanding of the network between sites
 - ❑ Identifying structural bottlenecks in need of remediation
 - ❑ Aiding network problem diagnosis and speeding repairs
 - ❑ **In general, incorporating knowledge of the network into our processes**
- ❄️ **We will require testing and iteration to better understand when and where the network metrics are useful.**

Summary

- ❄ During 2012 we converged on getting a ‘standard’ network monitoring framework in place globally for WLCG sites
- ❄ Both WLCG and OSG are focusing on getting robust installations of perfSONAR-PS in place to provide a single source of network metrics for their use.
- ❄ There have been significant improvements in the perfSONAR-PS toolkit in response to issues noted during ATLAS and CMS use. Centralized management is now possible. Further improvements coming in v3.3
- ❄ **Modular dashboard critical for “visibility” into networks**
- ❄ **Interesting times ahead as we complete deployments:**
How best to incorporate the network metrics we gather?
How best to debug network issues with this new framework in place?

Discussion/Questions

Questions or Comments?

Relevant URLs

- * perfSONAR-PS site <http://psps.perfsonar.net/>
- * Install/configuration guide: <http://code.google.com/p/perfsonar-ps/wiki/pSPerformanceToolkit33>
- * Modular Dashboard: <https://perfsonar.racf.bnl.gov:8443/exda/> or <http://perfsonar.racf.bnl.gov:8080/exda/>
- * Tools, tips and maintenance: <http://www.usatlas.bnl.gov/twiki/bin/view/Projects/LHCperfSONAR>
- * OSG networking pages <https://www.opensciencegrid.org/bin/view/Documentation/NetworkingInOSG>
- * GitHub perfSONAR Modular Dashboard: <https://github.com/PerfModDash/PerfModDash>
- * WLCG Installs <http://twiki.cern.ch/twiki/bin/view/LCG/PerfsonarDeployment>