# Using Network Performance Data in Facilities Operations

**Shawn McKee/University of Michigan**

**ATLAS T1/T2/T3 Jamboree / CERN**

**December 11th, 2012**

# Introduction

- This talk focuses on how the ongoing and developing activities in networking might be incorporated into our Facilities operations
  - Networking could benefit from better information about planned use
  - Facilities could benefit from better optimization and use of network resources
- I will discuss two main aspects of "Networking" broadly:
  - **Monitoring** – What information is (or should be) available and how to benefit from this information
  - **Control** – The ability to negotiate with networking services to acquire dedicated bandwidth on-demand or scheduled
- Note that while this Jamboree is ATLAS, much of the ongoing effort is LHC-wide and this is reflected in the slides
- Feel free to ask questions at anytime during the presentation…

# 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!

# Network Monitoring for LHC: 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
- **Next:** continuing measurements to track the network, alerting on problems as they develop
- **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.

# Current Network Monitoring

- We have a nice existing system (the ATLAS Dashboard) which is tracking end-to-end transfer results between sites.
  - Very good to understand how the overall system is performing
  - Difficult to understand if performance issues are from the network or the end systems or some complex interaction of the two
- We also have a broadly deployed network monitoring infrastructure based upon perfSONAR
  - Measures characteristics of the network path between sites
  - Gathers bandwidth, latency, packet loss and routing information
  - Not yet covering all our sites
  - Not yet “integrated” in our facilities operations

# WLCG perfSONAR-PS Deployments

- We want to measure (to the extent possible) the entire network path between LHC 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

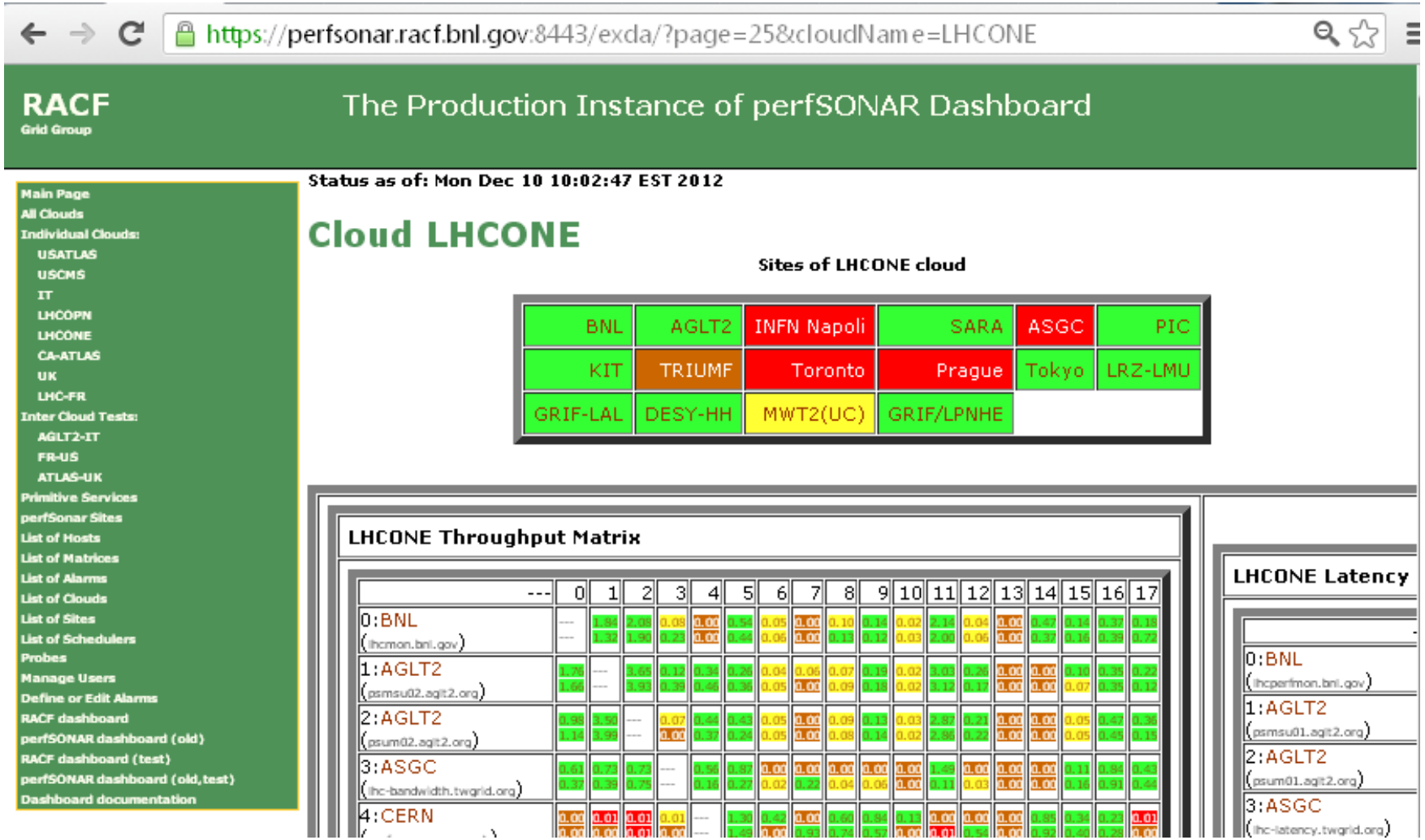
# Modular Dashboard: Centralized Info

- Having a large number of perfSONAR deployments is great for instrumenting our networks, but all these instances are not easy to track, summarize or understand.
- The current modular dashboard is being used to track a large number of LHC perfSONAR-PS installations:

https://perfsonar.racf.bnl.gov:8443/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.

# Example of Dashboard showing LHCONE




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

# perfSONAR 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
  - Effort is now being moved to GitHub under a modified BSD license
  - Everyone interested is welcome to participate
- OSG will be hosting the new dashboard “service” in the future and will also provide a packaged version of the code to allow those interested to deploy their own version
  - This service will also provide a source of network data, accessible through the developing user API

# Plans for WLCG Operations

- Simone will present the details about the near-term plans during the GDB tomorrow but summarizing here:
  - Encouraging all sites to deploy and register two instances ASAP
  - All sites to use the "mesh" configuration (central configs)
  - One set of test parameters to be used everywhere
  - GDB: http://indico.cern.ch/conferenceDisplay.py?confId=155075
- 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

# 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 were the network metrics are useful.**

# Network "Control"

- Going beyond monitoring the network, we can imagine **negotiating with the network to do what you want.**
- Is that possible? It certainly has not been possible in the past but that is changing. Many networks are now providing services for the creation of "virtual circuits"
- There is lots of ongoing work in **Software Defined Networking (SDN), OpenFlow** being a primary example.
  - Network providers may favor this mode of providing additional bandwidth in the future because of cost and management reasons
  - We need to understand what options are available and how best to take advantage of them within ATLAS
  - The future may hold new options for us…will we be ready?
- The question is how best for ATLAS to capitalize upon this capability?

# Existing Virtual Circuit Capabilities

- There are numerous projects/services related to reserving guaranteed bandwidth point-to-point: ION, OSCARS, Autobahn, DYNES, NS1 and many others.
- These services are often "production quality" in terms of the characteristics (availability and robustness) but are also "development" in the sense of the application use-cases and interfaces that may eventually be needed.
  - There is an opportunity for discussions to update/optimize APIs
- Setting up circuits requires a negotiation process and takes **~minutes** to instantiate circuits or bandwidth guarantees.
  - This is important when considering which tasks will benefit from controlling the network path
  - This setup time will presumably decrease at the underlying technologies mature.

# Using Network Control

- For our ATLAS infrastructure to effectively utilize “network control” we need the following:
  - Information about which paths support “control” interactions
  - Details of the negotiation process and API(s) involved
  - Understanding of the workflow and dataflow requirements of our planned work (on various timescales)
  - Priorities between competing tasks
  - Estimates of the network usable capacities
- My belief is we need to work incrementally and iteratively to integrate network control capability into our systems.
- **The LS1 period gives us a chance to provide the needed software changes to enable “network control” as an option in our infrastructure.**

# Ongoing LHC Networking Activities

- The WLCG Network Working Group led by Michael Ernst
- A WLCG operations sub-group (led by Simone Campana and Shawn McKee) is guiding the installation of perfSONAR at all Tier-2 sites (Tier-1's already instrumented)
- OSG has a new Networking Area (led by Shawn McKee) focused on hardening perfSONAR-PS, evolving the perfSONAR modular dashboard and providing OSG network services
- Two funded research efforts focused on the overlap between LHC software systems and networking:
  - **A**dvanced **N**etwork **S**ervices for **E**xperiments (ANSE), NSF funded (Caltech, Michigan, Vanderbilt and U Texas Arlington)
  - Next Generation Workload Management and Analysis System for Big Data, PANDA integration with networking, DOE funded (BNL, U Texas Arlington)
- These efforts need to interact with each other AND the Vos
- https://indico.cern.ch/conferenceDisplay.py?confId=215393

# Network Monitoring Challenges Ahead

- Getting hardware/software platform installed at all sites
- Dashboard development: Need additional effort to produce something suitable quickly. 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
- **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

# How to Make Progress?

- Using the LHCONE case as an example it seems possible to make significant progress in getting a standardized monitoring infrastructure in place quickly.
- Need to improve installs to be “set-it and forget-it”
- Integration with the experiments software stacks and DDM systems is now a high-priority (LS1 is an opportunity)
  - First network monitoring metrics
  - Next: SDN (Software Defined Networking)
- All VOs need to be aware of the need for network monitoring and the possibilities for sharing a common solution. Requires VO “pressure” to get sites to deploy
- Begin testing the use of metrics within the facility operations
- Plan for incorporating “network control” capabilities

# Discussion/Questions

Questions or Comments?

# 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
- This page is being shared with the perfSONAR-PS developers and we expect many of the “fixes” will be incorporated into future releases.
- Please feel free to add to the Wiki (either directly or by emailing me updates/changes/additions).