Network Monitoring for LHC VOs
An Overview of perfSONAR Use

Shawn McKee/University of Michigan

WLCG Grid Deployment Board

April 18th, 2012

Introduction

- I was asked to provide an overview of the use of perfSONAR and how we might leverage its us for LHC
- Outline:
  - A brief motivation and history
  - Use to date
  - Where to go from here
- 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!

History of perfSONAR

- 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.
- Modular dashboard developed by Tom Wlodek/BNL based upon USATLAS requirements to better understand deployed infrastructure
- LHCOPN choose to adopt in June 2011…mostly deployed within 3 months (by September 2011).

Monitoring LHCONE: Goals/Purpose

- We needed to understand how a transition to LHCONE impacts our LHC infrastructure.
- **First step:** get monitoring in place to create a baseline of the current situation
- **Second step:** as sites transition to using LHCONE, characterize the impact based upon measurements
- To gather the before/after measurements we choose the **perfSONAR-PS** toolkit given its extensive use for LHCOPN and the capabilities of the modular dashboard.
- **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.

Summary for LHCONE

- Our specific goal in setting up perfSONAR-PS for LHCONE is to acquire before and after network measurements for the selected early adopter sites. This is **not** the long-term network monitoring setup for LHCONE…that is TBD
- Details of which sites and how sites should setup the perfSONAR-PS installations is documented on the Twiki at: https://twiki.cern.ch/twiki/bin/view/LHCONE/SiteList
- In the next few slides I will highlight some of the relevant details

LHCONE perfSONAR-PS

- 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
  - The **bandwidth instance** measures achievable bandwidth via a short test (20-60 seconds) per src-dst pair every 4 hour period

Network Impact of perfSONAR

- To provide an idea of the network impact of a typical deployment here are some numbers as configured in the US
  - **Latency tests** send 10Hz of small packets (20 bytes) for each testing location. USATLAS Tier-2's test to ~10 locations. Since headers account for 54 bytes each packet is 74 bytes or the rate for testing to 10 sites is **7.4 kbytes/sec**.
  - **Bandwidth tests** try to maximize the throughput. A 20 second test is run from each site in each direction once per 4 hour window. Each site runs tests in both directions. Typically the best result is around **925 Mbps on a 1Gbps link for a 20 second test**. That means we send 4x925 Mbps*20 sec every 4 hours per testing pair (src-dst) or about 5 Mbps average.
  - Tests are configurable but the above settings are working fine.

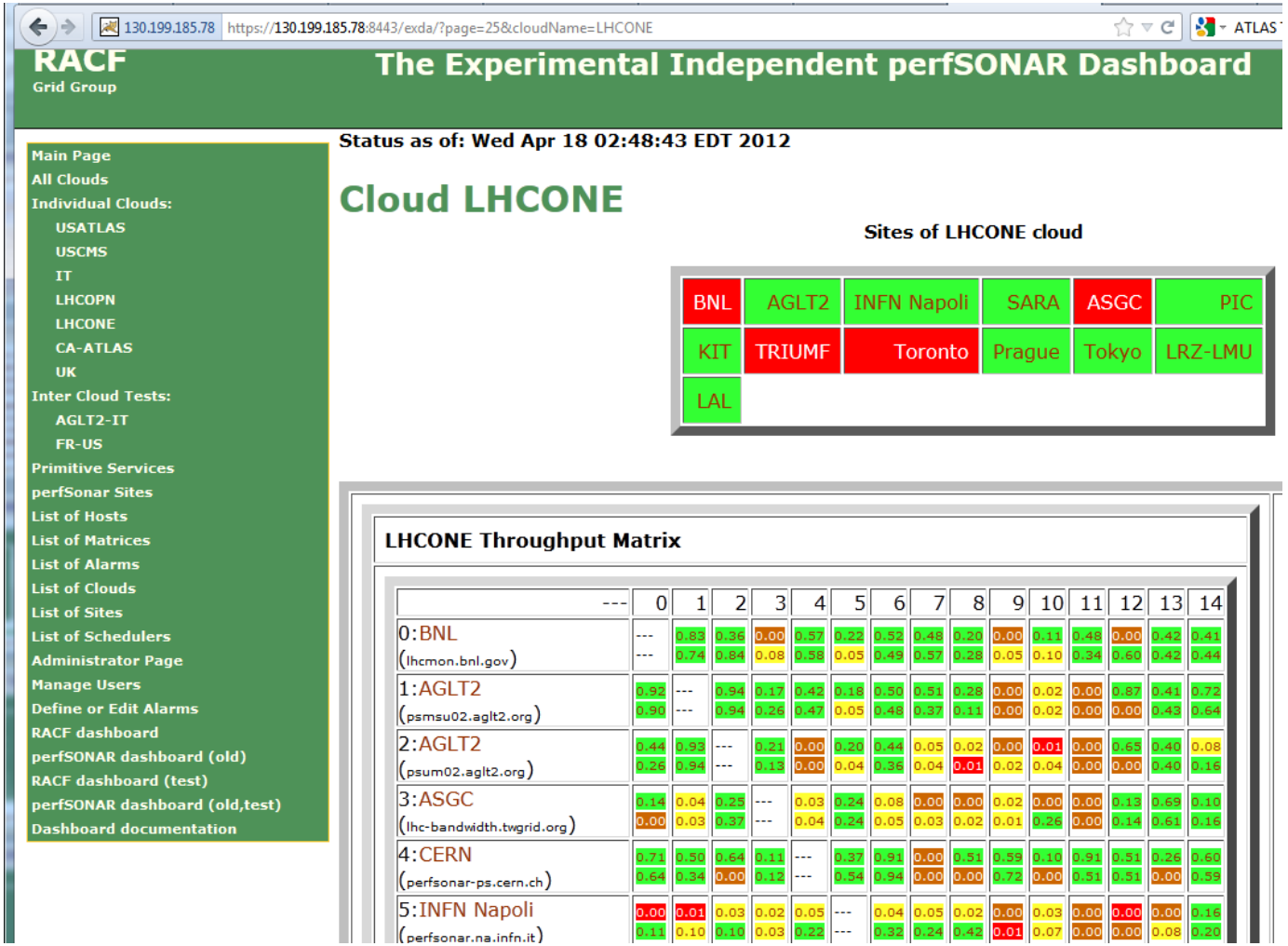
perfSONAR-PS Issues Observed

- Getting working monitoring deployed is a **first main step**
  - Focusing on a set of inter-site monitoring configuration raises awareness of the current shortcomings in our infastructure
- Two primary problems we noted:
  - Traffic between Tier-2Ds and Tier-1s is:
    - **Often routed on congested GPN** links
    - **Passing thru a firewall, limiting performance**
- Issue with MTU setting. Suggestion for LHCONE is to use jumbo frames. We need to understand the impact on our measurements.
- Test durations: 1G vs 10G. 20 seconds OK for 1G, but what about 10G? 60 seconds seems more reasonable.
- Getting alerts running: Issues with false positives.
- Higher level alarms: when, how?
- Modular dashboard: intro, use, future, issues

Modular Dashboard

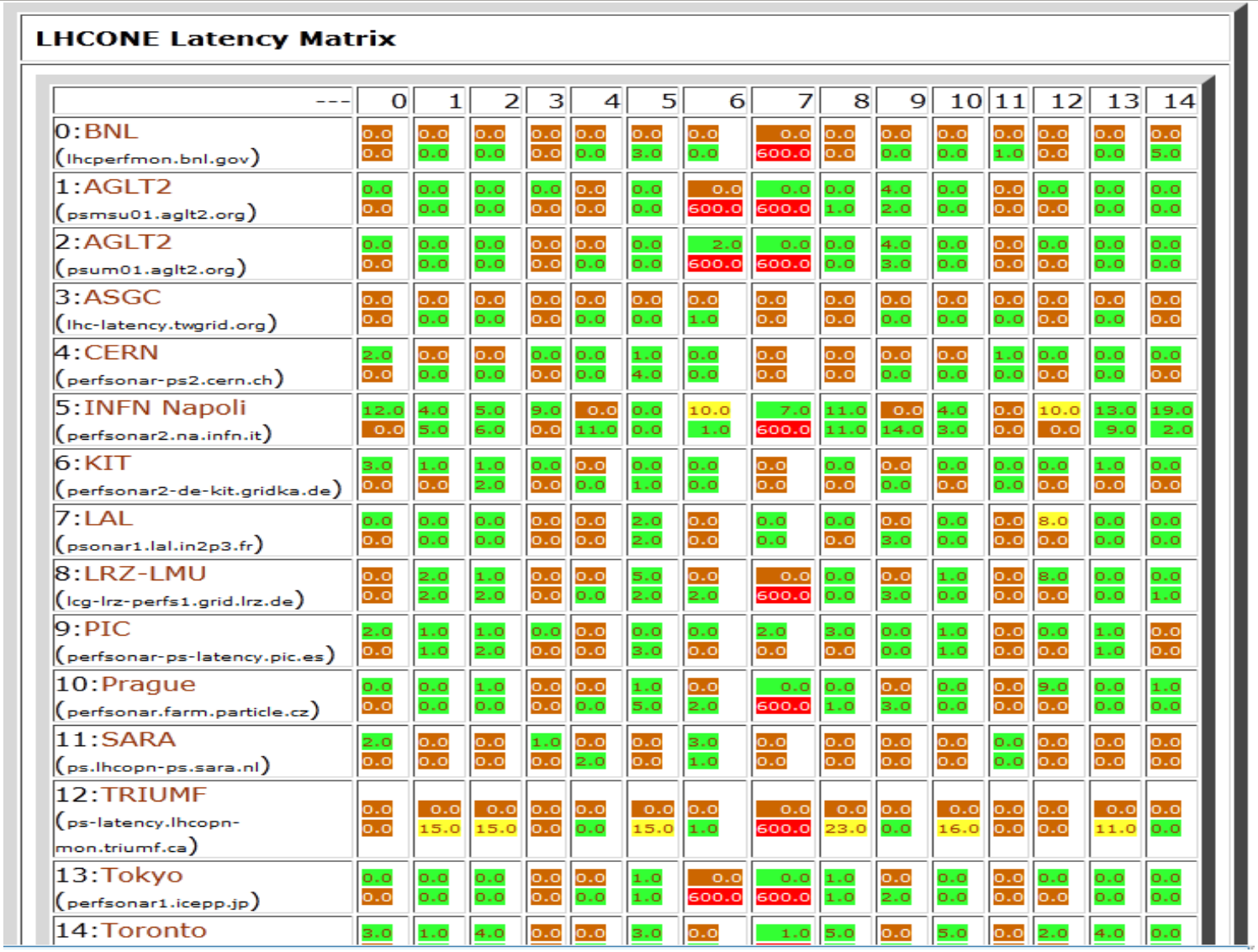
- Thanks to Tom Wlodek's work on developing a "modular dashboard" we have a very nice way to summarize the extensive information being collected for the near-term network characterization.
- The dashboard provides a highly configurable interface to monitor a set of perfSONAR-PS instances via simple plug-in 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 for LHCONE



See https://130.199.185.78:8443/exda/?page=25&cloudName=LHCONE

LHCONE Latency Matrix



LHCONE Latency Matrix

---	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0:BNL (lhcperfmon.bnl.gov)	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 3.0	0.0 0.0	0.0 600.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 1.0	0.0 0.0	0.0 0.0	0.0 5.0
1:AGLT2 (psmsu01.aglt2.org)	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 600.0	0.0 600.0	0.0 1.0	4.0 2.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0
2:AGLT2 (psum01.aglt2.org)	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	2.0 600.0	0.0 600.0	0.0 0.0	4.0 3.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0
3:ASGC (lhc-latency.twgrid.org)	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 1.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0
4:CERN (perfsonar-ps2.cern.ch)	2.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	1.0 4.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	1.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0
5:INFN Napoli (perfsonar2.na.infn.it)	12.0 0.0	4.0 5.0	5.0 6.0	9.0 0.0	0.0 11.0	0.0 0.0	10.0 1.0	7.0 600.0	11.0 11.0	0.0 14.0	4.0 3.0	0.0 0.0	10.0 0.0	13.0 9.0	19.0 2.0
6:KIT (perfsonar2-de-kit.gridka.de)	3.0 0.0	1.0 0.0	1.0 2.0	0.0 0.0	0.0 0.0	0.0 1.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	1.0 0.0	0.0 0.0
7:LAL (psonar1.lal.in2p3.fr)	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	2.0 2.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 3.0	0.0 0.0	0.0 0.0	8.0 0.0	0.0 0.0	0.0 0.0
8:LRZ-LMU (lcg-lrz-perfs1.grid.lrz.de)	0.0 0.0	2.0 2.0	1.0 2.0	0.0 0.0	0.0 0.0	5.0 2.0	0.0 2.0	0.0 600.0	0.0 0.0	0.0 3.0	1.0 0.0	0.0 0.0	8.0 0.0	0.0 0.0	0.0 1.0
9:PIC (perfsonar-ps-latency.pic.es)	2.0 0.0	1.0 1.0	1.0 2.0	0.0 0.0	0.0 0.0	0.0 3.0	0.0 0.0	2.0 0.0	3.0 0.0	0.0 0.0	1.0 1.0	0.0 0.0	0.0 0.0	1.0 1.0	0.0 0.0
10:Prague (perfsonar.farm.particle.cz)	0.0 0.0	0.0 0.0	1.0 0.0	0.0 0.0	0.0 0.0	1.0 5.0	0.0 2.0	0.0 600.0	0.0 1.0	0.0 3.0	0.0 0.0	0.0 0.0	9.0 0.0	0.0 0.0	1.0 0.0
11:SARA (ps.lhcopn-ps.sara.nl)	2.0 0.0	0.0 0.0	0.0 0.0	1.0 0.0	0.0 2.0	0.0 0.0	3.0 1.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0
12:TRIUMF (ps-latency.lhcopn-mon.triumf.ca)	0.0 0.0	0.0 15.0	0.0 15.0	0.0 0.0	0.0 0.0	0.0 15.0	0.0 1.0	0.0 600.0	0.0 23.0	0.0 0.0	0.0 16.0	0.0 0.0	0.0 0.0	0.0 11.0	0.0 0.0
13:Tokyo (perfsonar1.icepp.jp)	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	1.0 1.0	0.0 600.0	0.0 600.0	1.0 1.0	0.0 2.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 0.0
14:Toronto	3.0	1.0	4.0	0.0	0.0	3.0	0.0	1.0	5.0	0.0	5.0	0.0	2.0	4.0	0.0

LHCONE Throughput Matrix

LHCONE Throughput Matrix

---	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0:BNL (lhcmon.bnl.gov)	--- ---	0.83 0.74	0.36 0.84	0.00 0.08	0.56 0.58	0.22 0.05	0.54 0.49	0.48 0.57	0.20 0.28	0.00 0.05	0.11 0.10	0.52 0.34	0.00 0.60	0.42 0.42	0.41 0.44
1:AGLT2 (psmsu02.aglt2.org)	0.92 0.90	--- ---	0.94 0.94	0.17 0.26	0.42 0.48	0.18 0.05	0.50 0.48	0.51 0.37	0.28 0.10	0.00 0.00	0.02 0.02	0.00 0.00	0.87 0.00	0.41 0.43	0.72 0.64
2:AGLT2 (psum02.aglt2.org)	0.44 0.26	0.93 0.94	--- ---	0.21 0.13	0.00 0.00	0.20 0.04	0.44 0.36	0.05 0.04	0.02 0.01	0.00 0.02	0.01 0.04	0.00 0.00	0.65 0.00	0.40 0.40	0.08 0.16
3:ASGC (lhc-bandwidth.twgrid.org)	0.14 0.00	0.04 0.03	0.25 0.37	--- ---	0.03 0.04	0.24 0.24	0.08 0.05	0.00 0.03	0.00 0.02	0.02 0.01	0.00 0.26	0.00 0.00	0.13 0.14	0.69 0.61	0.10 0.16
4:CERN (perfsonar-ps.cern.ch)	0.71 0.65	0.50 0.34	0.64 0.00	0.11 0.12	--- ---	0.37 0.54	0.91 0.94	0.00 0.00	0.47 0.00	0.61 0.72	0.09 0.00	0.91 0.51	0.51 0.51	0.26 0.00	0.59 0.59
5:INFN Napoli (perfsonar.na.infn.it)	0.00 0.11	0.01 0.10	0.03 0.10	0.02 0.03	0.05 0.22	--- ---	0.04 0.32	0.05 0.24	0.02 0.00	0.00 0.01	0.03 0.07	0.00 0.00	0.00 0.00	0.00 0.08	0.16 0.20
6:KIT (perfsonar-de-kit.gridka.de)	0.47 0.46	0.51 0.28	0.43 0.47	0.02 0.08	0.94 0.94	0.59 0.38	--- ---	0.00 0.00	0.32 0.00	0.00 0.67	0.80 0.00	0.93 0.93	0.32 0.35	0.00 0.00	0.49 0.00
7:LAL (psonar2.lal.in2p3.fr)	0.56 0.56	0.48 0.45	0.48 0.48	0.00 0.00	0.00 0.00	0.38 0.14	0.00 0.00	--- ---	0.52 0.00	0.00 0.00	0.07 0.10	0.00 0.00	0.00 0.00	0.24 0.24	0.52 0.42
8:LRZ-LMU (lcg-lrz-perfs2.grid.lrz.de)	0.40 0.36	0.20 0.29	0.24 0.53	0.00 0.00	0.00 0.92	0.00 0.14	0.00 0.64	0.00 0.93	--- ---	0.00 0.00	0.11 0.10	0.00 0.00	0.15 0.00	0.16 0.16	0.51 0.46
9:PIC (perfsonar-ps-bandwidth.pic.es)	0.00 0.00	0.00 0.00	0.01 0.00	0.02 0.00	0.01 0.02	0.01 0.01	0.01 0.00	0.01 0.00	0.01 0.00	--- ---	0.01 0.01	0.01 0.00	0.00 0.00	0.00 0.00	0.00 0.00
10:Prague (perfsonar-bw.farm.particle.cz)	0.02 0.07	0.06 0.03	0.04 0.05	0.10 0.00	0.00 0.00	0.11 0.12	0.00 0.80	0.20 0.18	0.04 0.05	0.08 0.01	--- ---	0.00 0.00	0.04 0.00	0.03 0.01	0.04 0.05
11:SARA (ps.lhcopn-ps.sara.nl)	0.47 0.42	0.00 0.00	0.00 0.00	0.00 0.00	0.54 0.56	0.00 0.00	0.56 0.62	0.00 0.00	0.00 0.00	0.00 0.47	0.00 0.00	--- ---	0.00 0.43	0.00 0.00	0.00 0.00
12:TRIUMF (ps-bandwidth.lhcopn-mon.triumf.ca)	0.61 0.00	0.00 0.01	0.00 0.00	0.01 0.01	0.34 0.34	0.00 0.00	0.33 0.25	0.00 0.00	0.00 0.00	0.16 0.00	0.00 0.00	0.41 0.00	--- ---	0.00 0.00	0.73 0.91
13:Tokyo (perfsonar2.icepp.jp)	0.38 0.38	0.39 0.22	0.38 0.26	0.56 0.86	0.00 0.00	0.20 0.20	0.00 0.00	0.20 0.19	0.07 0.00	0.07 0.00	0.01 0.01	0.00 0.00	0.29 0.00	--- ---	0.29 0.27
14:Toronto	0.31	0.08	0.58	0.00	0.10	0.16	0.00	0.06	0.02	0.00	0.03	0.00	0.12	0.16	---

Using the Dashboard

- The dashboard is very useful for all of us to use to get a quick picture of the status for a particular grouping (cloud)
- It is also **very useful for sites** to debug their configurations!
- Note that you can quickly drill down and get error details as well as history plots or tables.
- I strongly wish to encourage anyone interested in network monitoring to use the dashboard to check the capabilities:

  https://130.199.185.78:8443/exda/?page=25&cloudName=LHCONE

- Authorization for Mgmt via X509 supported.

Challenges Ahead

- Getting hardware/software platform installed at all sites
- Dashboard development: Currently USATLAS/BNL and soon OSG, Canada (ATLAS, HEPnet) and USCMS. More ?
- Managing site and test 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)?
  - Improving the management of the configurations for VOs/Clouds
  - Tools to allow “central” configuration
- 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.

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.
- All VOs need to be aware of the need for network monitoring and the possibilities for sharing a common solution. Will require VO “pressure” to get sites to deploy
- VOs must assign effort to configure and gather VO view of network from shared perfSONAR measurement locations

Discussion/Questions

Questions or Comments?