Journal of Physics: Conference Series **513** (2014) 062008 doi:10.1088/1742-6596/513/6/062008

Deployment of a WLCG network monitoring infrastructure based on the perfSONAR-PS technology

S Campana[1], A Brown[2], D Bonacorsi[3], V Capone[4], D De Girolamo[5], A F Casani[6], J Flix[7,11], A Forti[8], I Gable[9], O Gutsche[10], A Hesnaux[1], S Liu[10], F Lopez Munoz[11], N Magini[1], S McKee[12], K Mohammed[13], D Rand[14], M Reale[15], S Roiser[1], M Zielinski[16] and J Zurawski[17]

[1] CERN, Geneva, CH
[2] Internet2, Ann Arbor, MI 48104
[3] Università di Bologna, IT
[4] Università di Napoli and INFN, IT
[5] INFN CNAF, Bologna, IT
[6] Universidad de Valencia, ES
[7] Centro de Investigaciones Energeticas, Medioambientales y Tecnologicas, CIEMAT, Madrid, Spain
[8] The University of Manchester, Oxford Road, Manchester M13 9PL, UK
[9] University of Victoria, CA
[10] Fermi National Accelerator Laboratory, US
[11] Port d'Informacio Cientifica (PIC), Universitat Autonoma de Barcelona, Bellaterra (Barcelona), ES
[12] Physics Department, University of Michigan, Ann Arbor, MI 48109-1040, USA
[13] University of Oxford, Wellington Square, Oxford, OX1 2JD, UK
[14] Imperial College London, Blackett Lab, Physics Dept, Prince Consort Rd, London SW7 2BW, UK
[15] GARR, IT
[16] University of Rochester, US
[17] ESnet, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

E-mail: simone.campana@cern.ch

Abstract. The WLCG infrastructure moved from a very rigid network topology, based on the MONARC model, to a more relaxed system, where data movement between regions or countries does not necessarily need to involve T1 centres. While this evolution brought obvious advantages, especially in terms of flexibility for the LHC experiment's data management systems, it also opened the question of how to monitor the increasing number of possible network paths, in order to provide a global reliable network service. The perfSONAR network monitoring system has been evaluated and agreed as a proper solution to cover the WLCG network monitoring use cases: it allows WLCG to plan and execute latency and bandwidth tests between any instrumented endpoint through a central scheduling configuration, it allows archiving of the metrics in a local database, it provides a programmatic and a web based interface exposing the tests results; it also provides a graphical interface for remote management operations. In this contribution we will present our activity to deploy a perfSONAR based network monitoring infrastructure, in the scope of the WLCG Operations

Coordination initiative: we will motivate the main choices we agreed in terms of configuration and management, describe the additional tools we developed to complement the standard packages and present the status of the deployment, together with the possible future evolution.

1. Introduction

The WLCG infrastructure has moved from a very restrictive network topology, based on the MONARC model, to a more interconnected system, where data movement between regions or countries does not necessarily need to involve T1 centers. While this evolution brought obvious advantages, especially in terms of flexibility for the LHC experiment's data management systems, it also opened the question of how to monitor the increasing number of possible network paths, in order to provide a global, reliable network service. The perfSONAR [1] network monitoring framework has been evaluated and agreed to as a proper solution to cover the WLCG network monitoring use cases: it allows WLCG to plan and execute latency, traceroute and bandwidth tests between any instrumented endpoint through a central scheduling configuration, it allows archiving of the metrics in a local database, it provides a programmatic and a web based interface exposing the tests results; it also provides a graphical interface for remote management operations.

In this paper we will describe the relevant details of the perfSONAR implementation we choose and present our activity to deploy a perfSONAR based network monitoring infrastructure, in the scope of the WLCG Operations Coordination initiative. We will motivate the main choices we agreed in terms of configuration and management, describe the additional tools we developed to complement the standard packages and present the status of the deployment, together with the possible future evolution.

2. Network Monitoring for WLCG Using perfSONAR-PS

WLCG sites are globally distributed and intrinsically rely upon high-performance networks to function effectively. Experience has shown that end-to-end network issues can be difficult to identify, localize and debug because of insufficient tools and the typically many administrative domains involved in WLCG network paths between sites. One prominent example was the BNL-CNAF network issue documented in https://ggus.eu/ws/ticket_info.php?ticket=61440 (transfer problem between CNAF and BNL). Resolving the issue took over 7 months and logged 72 entries in the ticket. This and many other examples highlight the requirement for WLCG to have a pervasive network monitoring infrastructure deployed able to identify issues when they occur and aid in debugging and localizing problems for quick resolution.

Obviously, some Research and Education (R&E) networks had long ago recognized the difficulties in finding and diagnosing wide-area network (WAN) issues. A consortium of R&E networks and Universities in North/South America and Europe developed the perfSONAR infrastructure to help address such issues. Within WLCG the USATLAS sites began testing and deploying a perfSONAR implementation called perfSONAR-PS in 2007 [2]. By 2010, with its usefulness demonstrated, the LHCOPN [3] choose to also deploy the perfSONAR-PS Toolkit at the CERN Tier-0 and the 10 Tier-1's worldwide. With this experience in place it was logical that WLCG adopt the perfSONAR-PS Toolkit for its network monitoring needs. In fall 2012 a dedicated operations task-force was formed to coordinate and manage deploying perfSONAR-PS at all WLCG sites worldwide.

3. Description of the perfSONAR-PS Toolkit

The perfSONAR-PS Toolkit is the result of an open source development effort based upon perfSONAR [4]. It targets creating an easy-to-deploy and easy-to-use set of perfSONAR services and is available as a bootable all-in-one system (ISO on CD or USB) or can be deployed onto the target host systems local disk via a 'netinstall' process. The current release (v3.x) is based upon CentOS6 and is available in 32 and 64-bit architectures.

Installation of the toolkit is recommended to be on dedicated physical hardware to ensure other applications or services don't interfere with network metrics and diagnostics provide by perfSONAR-

PS. For ease of maintenance the recommend deployment method is "netinstall" to local disk which utilizes a YUM repository for perfSONAR-PS components. Updates and security fixes are then easily available via 'yum update'. Because the components are available via RPM it is also possible to integrate installing perfSONAR-PS with a site's local provisioning and configuration management systems, though this obviously requires significantly more work for the person deploying. It should be noted that while it is not recommended, some sites have deployed the perfSONAR-PS toolkit on virtualized machines (VMs).

3.1. Details of the perfSONAR-PS Toolkit

Once the perfSONAR-PS Toolkit is deployed on a host it provides a number of features:

- a web based GUI allowing the administrator to configure the host and services running, define and schedule tests and display measurement results.
- a set of services that provide throughput tests (BWCTL), ping tests (PingER), latency/pack-loss tests (OWAMP), traceroute and two on-demand diagnostics (NDT,NPAD).
- a measurement archive which stores all test results and makes them available through an API.
- an integrated Cacti instance that can monitor and graph host metrics and could be configured to query, monitor and graph local network equipment via SNMP.

A perfSONAR-PS toolkit instance is typically configured to be either a "bandwidth" or a "latency" instance. It is not recommended to run both types of services on the same node because the bandwidth tests interfere with the latency tests and can cause false-positives indications of network problems.

4. perfSONAR-PS deployment in WLCG

4.1. Early phase of perfSONAR deployment

The first deployment of perfSONAR in the WLCG infrastructure started in the scope of monitoring the OPN performance. All the 11 WLCG T1s and CERN have been instrumented with two distinct perfSONAR hosts (one for bandwidth and the other for latency/packet loss tests) and configured to execute periodic measurements among all sites. At this stage there was no tool for automatic configuration: each service manager had to specify the hostnames of target services and configure the test parameters. All information was documented in static and text-based web pages, and kept up-to-date by a responsible person. This deployment model worked very well for a small and rather immutable system such as the LHCOPN and successfully provided important matrices of test results that were used to improve the stability of the system. The model was originated in USATLAS, which started a campaign of deployment of perfSONAR-PS across T2 sites in 2007. As more sites and regional clouds joined in deploying perfSONAR-PS the limitations of the original model become obvious: a trivial change at some sites (installation of a new host, change of alias or hostname) would have to be reflected in a manual reconfiguration change at all other sites. The effort become almost unmanageable when it was agreed to utilize perfSONAR to monitor the performance of the LHCONE [5] infrastructure, with many new sites joining every month. At the same time, most experiments understood the importance of a network monitoring infrastructure and asked for a broad deployment at all sites. In Fall 2012 it was therefore agreed to consolidate the perfSONAR deployment effort in WLCG and a dedicated task force was initiated under the scope of the WLCG Operations Coordination activity.

4.2. Full scale WLCG perfSONAR deployment

The task force took a pragmatic approach and decided to start by focusing on the sites which were of primary interest from experiments. A list of 86 sites has been initially collected from ATLAS, CMS and LHCb. The sites have been organized in regions, mostly determined by their geographical location and the research network to which they belong, with few exceptions. All sites have been asked to deploy a first perfSONAR-PS instance to treat bandwidth tests and a second one to treat all other tests

(ping, packet loss, traceroute). Both instances had to be registered in the WLCG service registries (GOCDB or OIM for OSG sites) following given instructions.

The perfSONAR developers worked with USATLAS and later with WLCG to provide the ability to centrally define sets of related perfSONAR-PS Toolkits via the use of the so-called "mesh-configuration". The idea is shown in Figure 1 and illustrates in perfSONAR-PS Toolkit instance reading from specific URLs that provide host and test configuration details in JSON format. In addition, some services were not as robust as needed for a required WLCG service and we worked closely with the Toolkit developers to improve the overall resiliency through a combination of bug-fixes and configuration changes.

Thanks to the mesh-configuration functionality, the deployment model has then been centralized: benefitting from the improvements in the perfSONAR-PS toolkit, which from release 3.3.1 do allow to import the configuration from a web based URL, all configurations have been stored in a single location available through HTTP. The configurations would include the definition of hosts, regions, test parameters and test scenarios. The test scenarios are based on the concept of "mesh" meaning a group of sites to which a test definition can be assigned: all sites of the mesh would trigger the given test with the parameters specified in the definition. This powerful scheme allowed defining different test scenarios for sites within the same region and across different regions. In particular it was agreed to configure:

- bandwidth tests (with a length of 30 seconds) every 6 hours for sites within the same region, every 12 hours between a T2-T1 pair of different regions and every week for any other combination;
- latency tests (packet loss), sending 10Hz of packets across all WLCG sites and measuring the loss based on batches of 600 packets (every minute);
- traceroute tests between all WLCG sites once per hour;
- ping tests between all sites every 20 minutes;

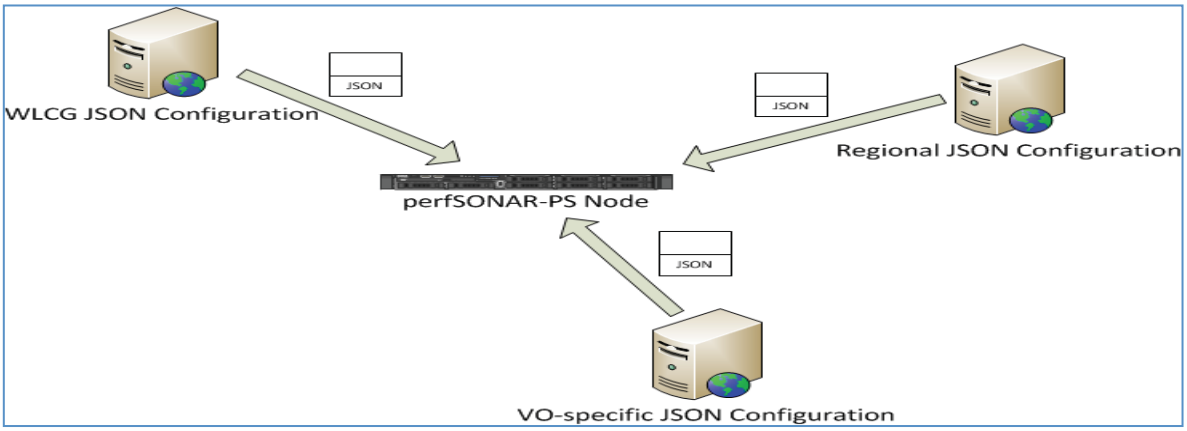

Figure 1: an illustration of the perfSONAR mesh configuration concept

Such configuration was agreed in order to achieve statistically meaningful network measurements without disrupting the overall production network activities.

4.3 Data aggregation and data mining

The raw measurements provided by the various tests are stored locally at each perfSONAR-PS

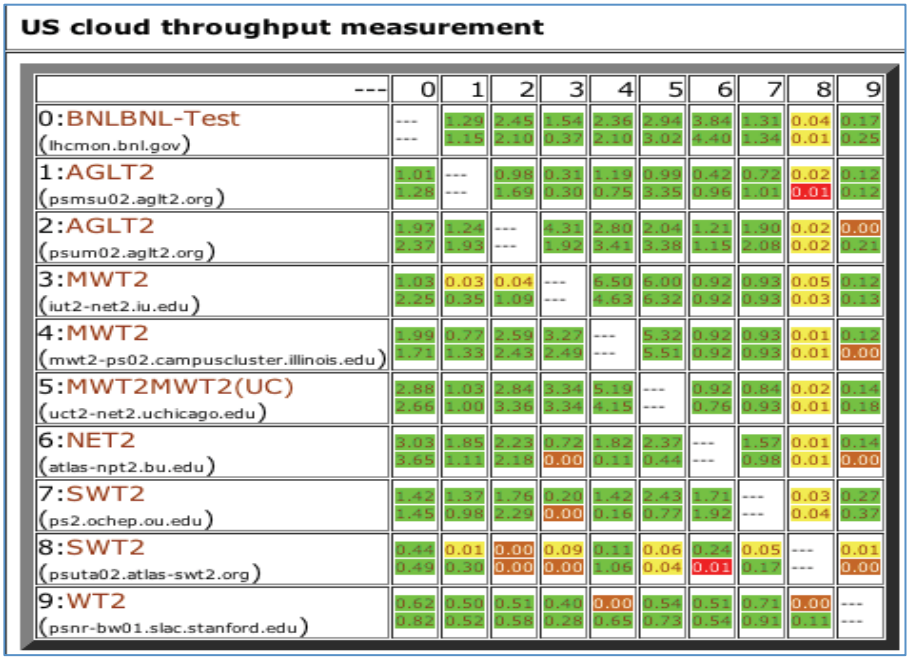

Figure 2: The perfSONAR Modular Dashboard throughput measurements for the USATLAS region

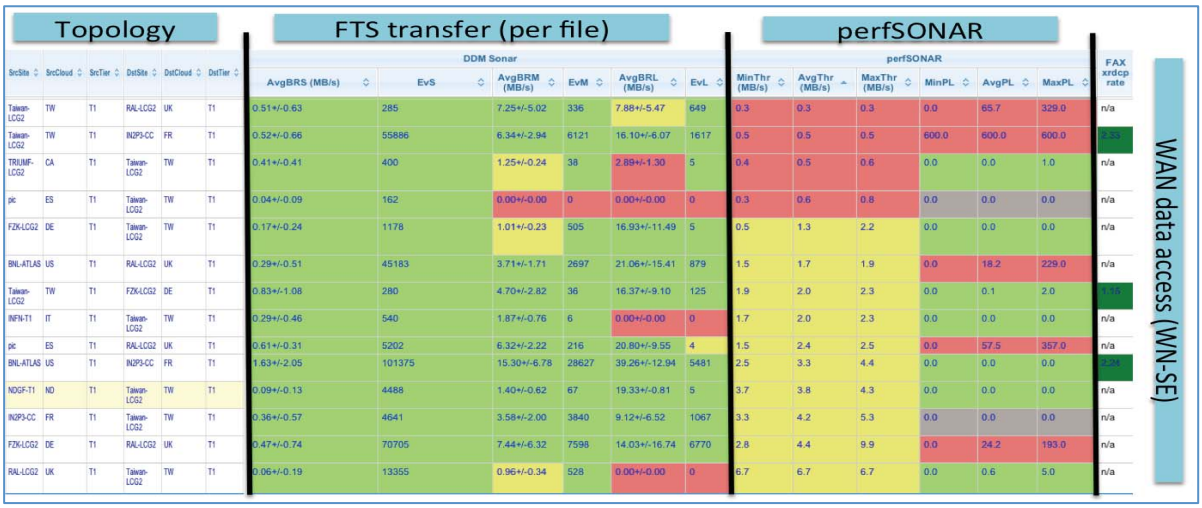

Figure 3: The ATLAS Site Status Board network measurements view

instance in a MySQL database. The information are then collected, aggregated and stored in the perfSONAR Modular Dashboard [6]. This dashboard provides a GUI for quick visualization of the test results, with aggregation by site and/or region. For example Figure 2 shows the matrix of perfSONAR bandwidth tests between USATLAS sites, as displayed in the perfSONAR dashboard. Additionally, the dashboard exposes the results programmatically via HTTP. Various applications collect and consume the JSON output and consume the numbers for different use cases. For example Figure 3 shows the ATLAS Site Status Board [7], where perfSONAR measurements of throughput and packet

loss are utilized to complement the network measurement information coming from file transfer statistics between two storages and between storage and Worker Node.

5. Conclusions

The status of the perfSONAR-PS deployment in WLCG is graphically summarized in Figure 4: 81 sites have deployed perfSONAR-PS instances and published in GOCDB/OIM (3 of them need attention concerning the publication), while 31 site are still missing. The deployment issues are clustered in regions where the WLCG perfSONAR deployment misses a strong contact person. In the near term we plan to work in parallel on three different areas:

- The perfSONAR deployment will be completed at the missing sites. Additionally, we need to verify that existing instances are running the most recent version and import correctly the central configuration.
- We will instrument service monitoring and, in future, alarming in case an instance fails to schedule tests or in general starts malfunctioning
- We will work with the perfSONAR Modular Dashboard developers to rationalize and improve the various views and configuration of the tool.

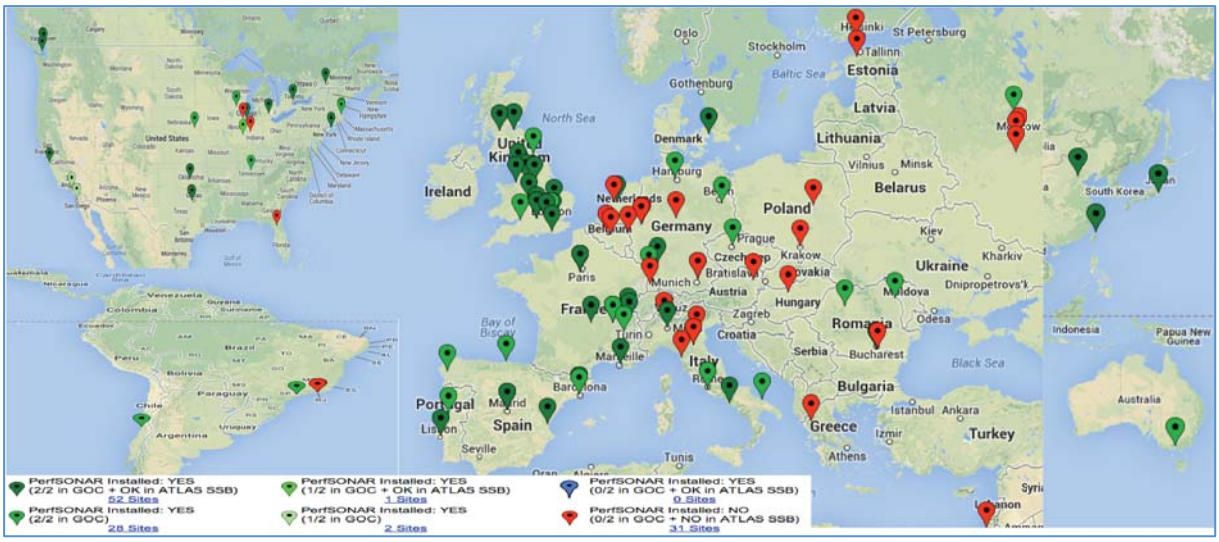

Figure 4: The perfSONAR deployment status in WLCG

References

[1] Hanemann A., Boote J. , Boyd E., Durand J., Kudarimoti L., Lapacz R., Swany M., Trocha S., and Zurawski J., "PerfSONAR: A Service-Oriented Architecture for Multi-Domain Network Monitoring", International Conference on Service Oriented Computing (ICSOC 2005), Amsterdam, The Netherlands, 2005

[2] Tierney B., Metzger J., Boote J., Brown A., Zekauskas M., Zurawski J., Swany M., Grigoriev M., "perfSONAR: Instantiating a Global Network Measurement Framework", 4th Workshop on Real Overlays and Distributed Systems (ROADS'09) Co-located with the 22nd ACM Symposium on Operating Systems Principles (SOSP), January 1, 2009

[3] http://lhcopn.web.cern.ch/lhcopn/

[4] Zurawski J., Balasubramanian S., Brown A., Kissel E., Lake A., Swany M., Tierney B., Zekauskas M.,"perfSONAR: On-board Diagnostics for Big Data", 1st Workshop on Big Data and Science: Infrastructure and Services Co-located with IEEE International Conference on Big Data 2013 (IEEE BigData 2013), October 6, 2013

[5] http://lhcone.net/

[6] The perfSONAR Modular Dashboard projec: https://github.com/PerfModDash/PerfModDash
[7] The ATLAS SSB at http://dashb-atlas-ssb.cern.ch/dashboard/request.py/siteview