GridPP43

28-30 August 2019

GridPP: perfSONAR refresh

Tim Chown, Jisc

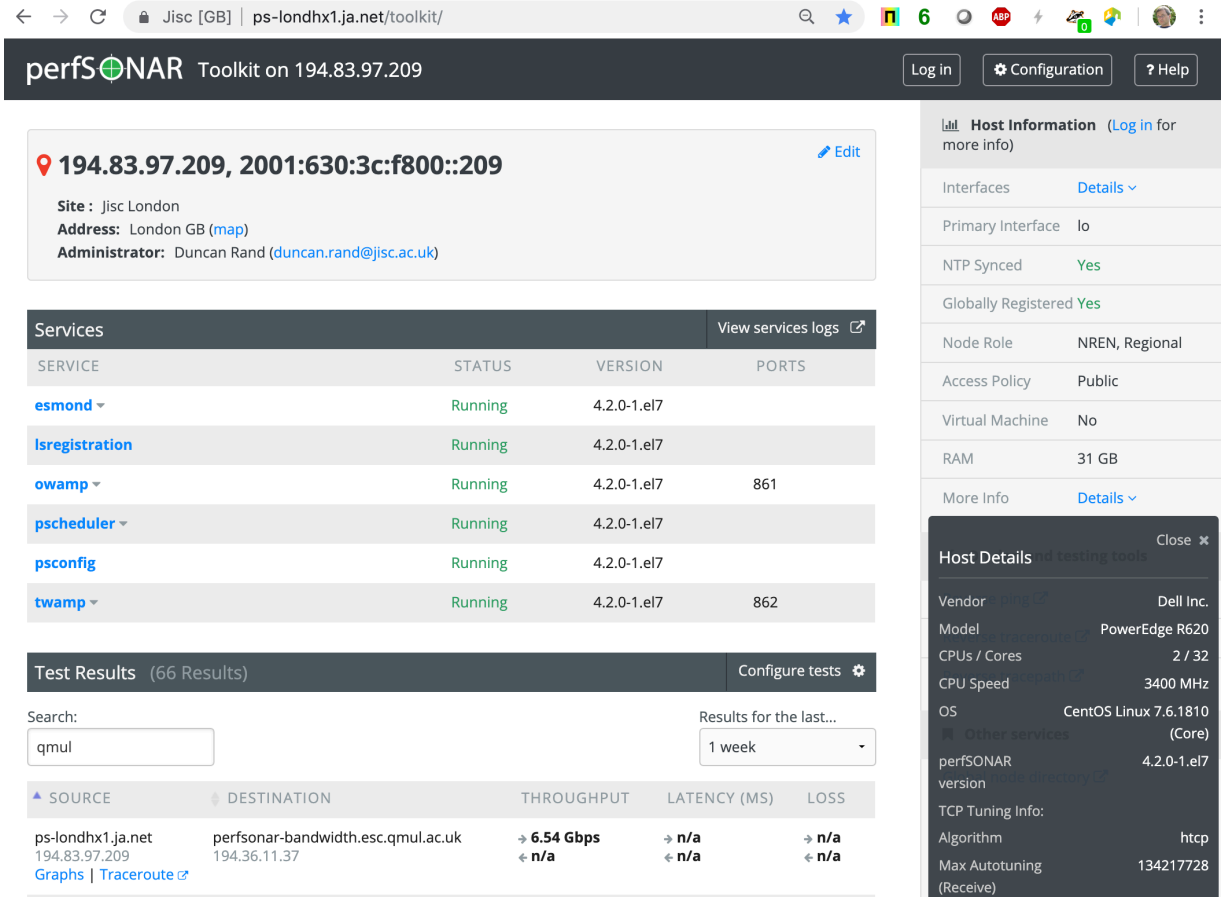
perfSONAR

Provides network characteristic measurements over time

- Open source package, available as packages or in ISO format
  - https://www.perfsonar.net/
- Measurements
  - Runs continuous tests for latency, loss, jitter and path
  - Tests throughput by default using TCP iperf every 6 hours
- It is used throughout the WLCG – for nodes see http://stats.es.net/ServicesDirectory/
  - Currently deployed at UK Grid PP sites
  - Facilitates mesh of measurements between all sites (even if not all sites exchange data)
- Ongoing perfSONAR development
  - Joint project between European and US partners; now one single package
  - Currently at v4.2
  - Many recent improvements, e.g., pScheduler, pSConfig, GridFTP throughput plugin

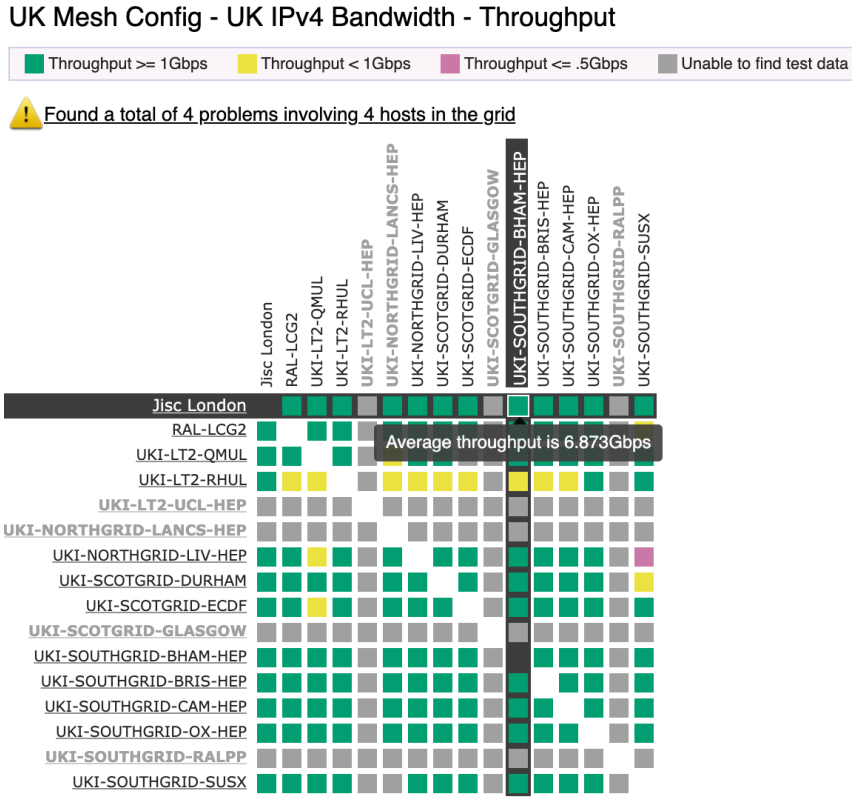
Jisc

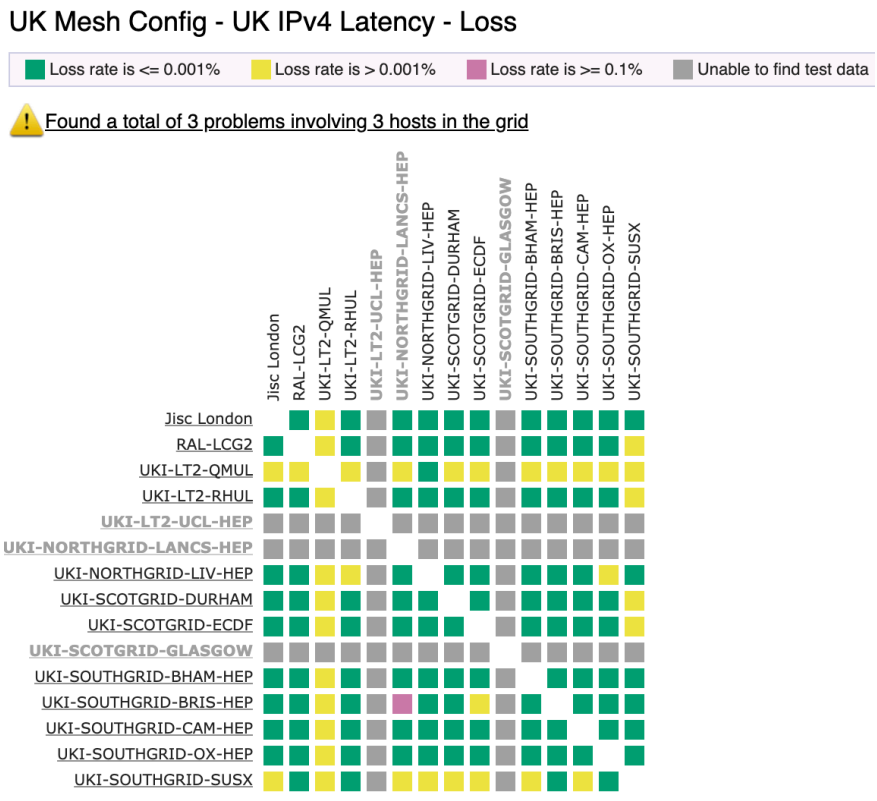
perfSONAR web UI view



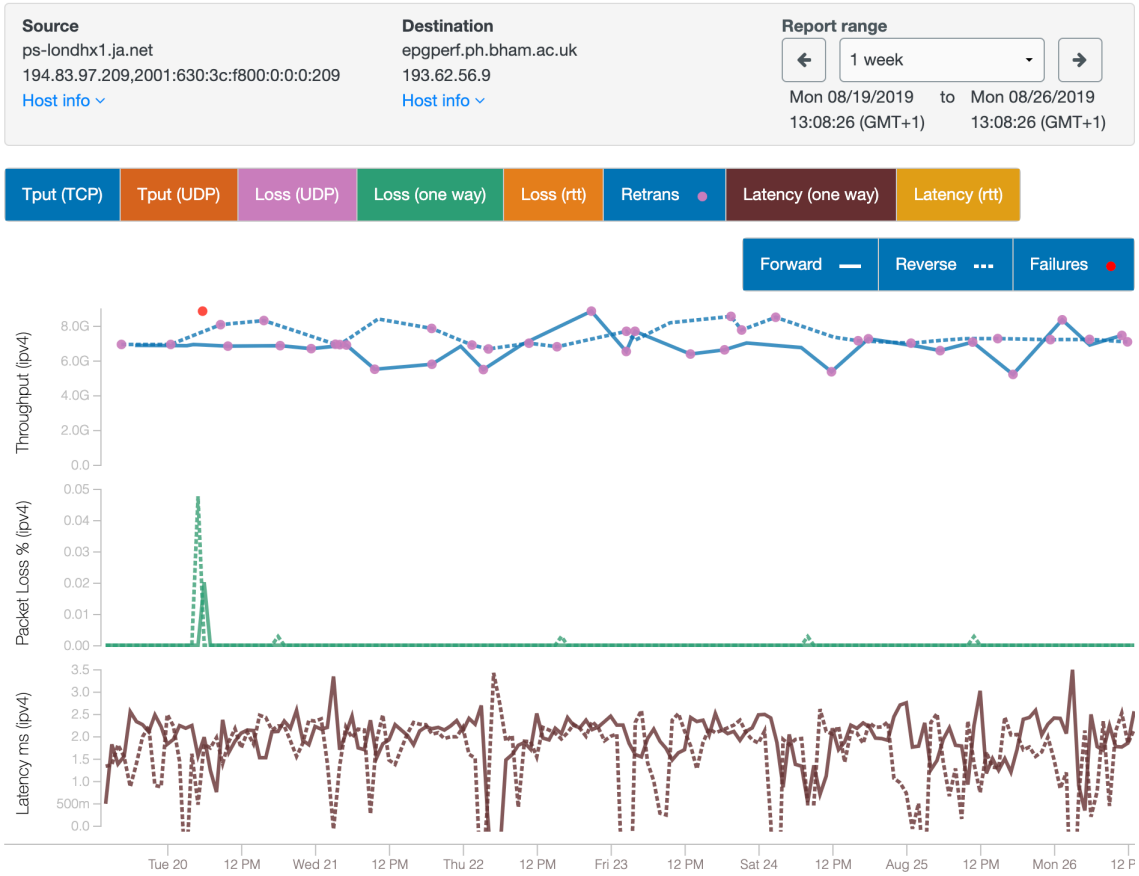
The current GridPP mesh

https://psmad.opensciencegrid.org/maddash-webui/index.cgi?dashboard=UK%20Mesh%20Config

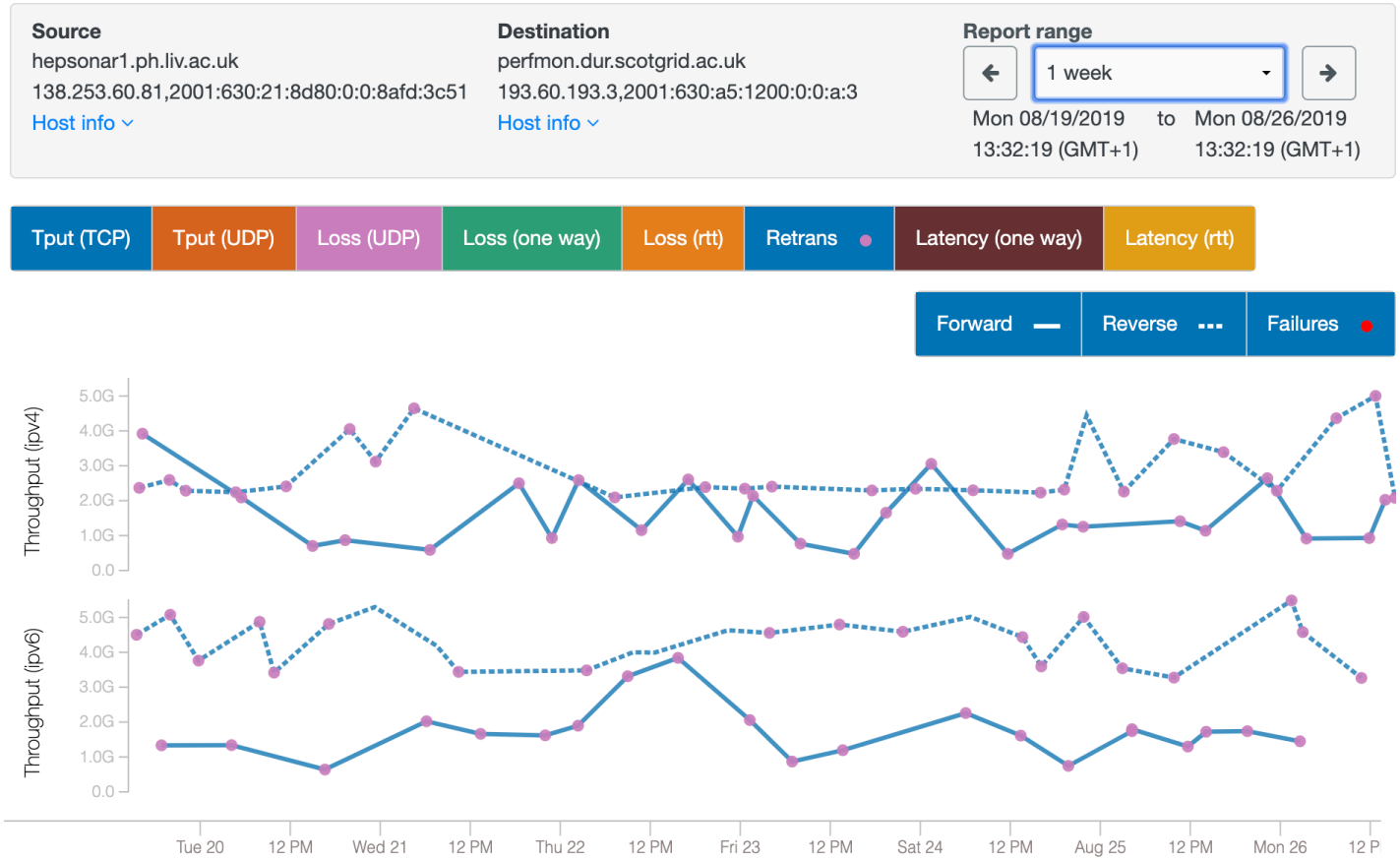




Looking at the historical data...



A more variable example



So why a refresh?

Rationale

- Large scale data transfers are of increasingly critical importance to GridPP
- Therefore we should have the means to monitor and troubleshoot network performance
- perfSONAR is the de facto WLCG measurement tool, and will be for the foreseeable future
- It is open source, and being actively developed by a large international team
- The GridPP perfSONAR systems generally have aging hardware, and do not all auto-update to new perfSONAR versions
- To maximise perfSONAR's value to GridPP, we should use well-specified hardware, and ensure the software is kept up to date
- It is important that those who maintain the GridPP perfSONAR systems have confidence in it
- A refresh is also an opportunity to improve and innovate in other aspects, such as monitoring of the perfSONAR systems themselves, and correlation of perfSONAR and measurement data from other sources (e.g., netflow data, router interface data, …)

Jisc

Jisc's interest in perfSONAR

Various perspectives

- It is important tool for supporting our End-to-End Performance Initiative work, helping Jisc members (Janet-connected sites) make optimal use of their connectivity
  - We recommend its deployment to any site moving large volumes of data over Janet
- Jisc have deployed two 10G-connected perfSONAR servers on Janet
  - London: http://ps-londhx1.ja.net/toolkit/
  - Slough: http://ps-slough-10g.ja.net/toolkit/
- We host perfSONAR meshes in support of ongoing e2e performance investigations
- We are testing 100G perfSONAR (and DTNs) to allow us to support our member sites as they move to 100G connections to Janet
- We are also experimenting with 1G "small node" perfSONAR systems, and with perfSONAR on virtual instances, such as with AWS and on switch/router platforms such as the Cisco 9300
- In the new GÉANT GN4-3 project, Jisc (Tim) leads the work package responsible for the European part of the perfSONAR software development team

Jisc

perfSONAR refresh – issues to consider? (1)

Hardware?

- See https://www.jiscmail.ac.uk/cgi-bin/webadmin?A2=ind1201&L=TB-SUPPORT&O=D&P=5928 for the discussion in 2012 on a GridPP-funded deployment of perfSONAR at all sites, and of the desired specification and features
  - Resilience important; RAID, redundant PSUs, and a common spec if not a common vendor
- New 10G nodes? Shawn McKee's comment – "We are looking at a Dell R340, 32GB RAM, Intel Xeon E-2146G 3.5GHz, 12M cache, 6C/12T, turbo (80W), dual port SFP+ 10G, 10K SAS 1.2TB disk. I think having 32G of RAM is important, especially if you want to **use the node for BOTH latency and bandwidth (on different NICs with different FQDNs)**. The high clock on the CPU helps with 10+G transfer rates." That system would be ~£2,000.
- For 100G? – "You want at least one PCIe Gen3 x16 slot and the fastest CPU clock you can get." Shawn mentioned PowerPC, but perfSONAR is not yet ported formally to that platform
- All sites must support 10G, but **where and when might we introduce 100G tests?**
- RAL and Imperial are now connected to Janet at 100G; others will likely follow soon

Issues to consider? (2)

Other operational topics…

- Providing perfSONAR mesh views and archiving of data
  - We can continue to use the WLCG provision here. Jisc also hosts a copy of the mesh view
- Automation?
  - If sites want to fully maintain their own nodes, we should presumably let them?
  - A lot of work on perfSONAR and ansible in the GN4-3 project (PMP nodes)
- Monitoring the perfSONAR nodes
  - Important to have eyes on the status of the perfSONAR systems
- Performing richer analysis of network performance?
  - Use APIs to correlate perfSONAR data with other measurements (such as netflow data)
  - NetSage is an excellent example of this - https://portal.netsage.global/grafana/
- Commitment to act on reports
  - Important to ensure nodes are well-maintained, and issues are acted upon in reasonable time

Jisc

Also need to consider WLCG bigger picture

perfSONAR is used worldwide by the WLCG

- Excellent work being done in the WLCG team by Shawn McKee, Marian Babik and others
- UK GridPP mesh tests draw from config files on WLCG servers (*psconfig.opensciencegrid.org*)
- Important that we
  - Don't duplicate what they are doing
  - Seek collaboration to improve the utility and value of the monitoring to GridPP
- Examples of potential collaboration
  - SAND (https://sand-ci.org/); using a wider range of measurement sources to improve fault detection and analysis capabilities
  - Using tools such as *check_mk* to ensure that the GridPP perfSONAR platforms are performing as they should
- We (Jisc) will be having more detailed discussions with Shawn and Marian soon; if others are interested please let us know

Jisc

Other UK perfSONAR deployments

Most relevant examples

- STFC
  - Project starting soon with Andy Forrester
  - Jisc wrote an outline project brief to help Andy earlier in the year
  - Using perfSONAR between STFC sites, and between facilities within RAL

- DiRAC
  - Duncan talking to Mike Wilkinson
  - Early indicative mesh set up
    - https://ps-dash.dev.ja.net/maddash-webui//index.cgi?dashboard=DiRAC
    - Nodes are not (yet) all alongside DiRAC infrastructure

Jisc

Thank you – any questions?

tim.chown@jisc.ac.uk