The case for perfSONAR deployment
in support of networking for data-intensive research

This document summarises the case for university IT departments to deploy perfSONAR network monitoring systems, such that they are more readily able to identify network performance issues, particularly with respect to supporting the growing volume of data-intensive applications that their researchers may run between Janet-connected sites and beyond.

What is perfSONAR?

perfSONAR[1] is an open source toolkit that provides active measurement of network characteristics between two or more systems that are running the software. The toolkit is being developed and supported by research networking teams within the GEANT and Internet2 communities, and is built upon a suite of tried and tested open source tools, such as the *iperf* throughput measurement tool.

While perfSONAR can be used to test network characteristics over time between just two sites, it can also be configured as a "mesh" that can monitor and summarise performance between multiple sites in a research community, through an intuitive web interface. The UK GridPP[2] particle physics community is one such example, but others are now beginning to emerge, such as the community of DiRAC HPC facilities.

Why deploy perfSONAR?

With the use of data-intensive applications becoming more widespread between Janet-connected organisations, it is important that those providing networking support to researchers have access to the most appropriate tools to allow them to identify and diagnose network performance issues. As a tool that is being used widely, in particular by the worldwide particle physics community, perfSONAR has a strong pedigree, measuring multi-Gigabit throughput between participating endpoints.

There are many varied and nuanced causes of network performance issues. TCP-based applications are particularly susceptible to packet loss; just a fraction of one percent loss can cause significant throughput problems for conventional TCP protocols. It is therefore important to engineer end-to-end paths to be as free of packet loss as possible. Further, security appliances such as firewalls and intrusion detection systems can also restrict the performance of high packet-per-second network flows.

The perfSONAR package allows network operators, and also network-savvy researchers, to measure network characteristics over time, including throughput, latency and loss, and to then visualise which sites are experiencing problems. This can be useful when a new research system is first being commissioned, but also to gauge the impact of improvements to the local network architecture, or to identify issues that arise during day-to-day operation. perfSONAR can also record changes in the network path between systems over time.

Figure 1 shows an example of a development mesh showing UK GridPP sites. The perfSONAR *MaDDash* dashboard shows rows and columns for each site, with colour-coded boxes in the grid indicating the throughput being achieved, green being throughout above 900Mbps, and purple being throughput below 500Mbps (these thresholds can be changed if desired). Note that throughput tests are run in both directions.

The measurements displayed in *Figure 1* are a snapshot of the most recent throughput tests run by each perfSONAR system in the mesh. perfSONAR can archive all measurements, so these can also be viewed over time, as the example in *Figure 2* shows.

---

[1] http://www.perfsonar.net/
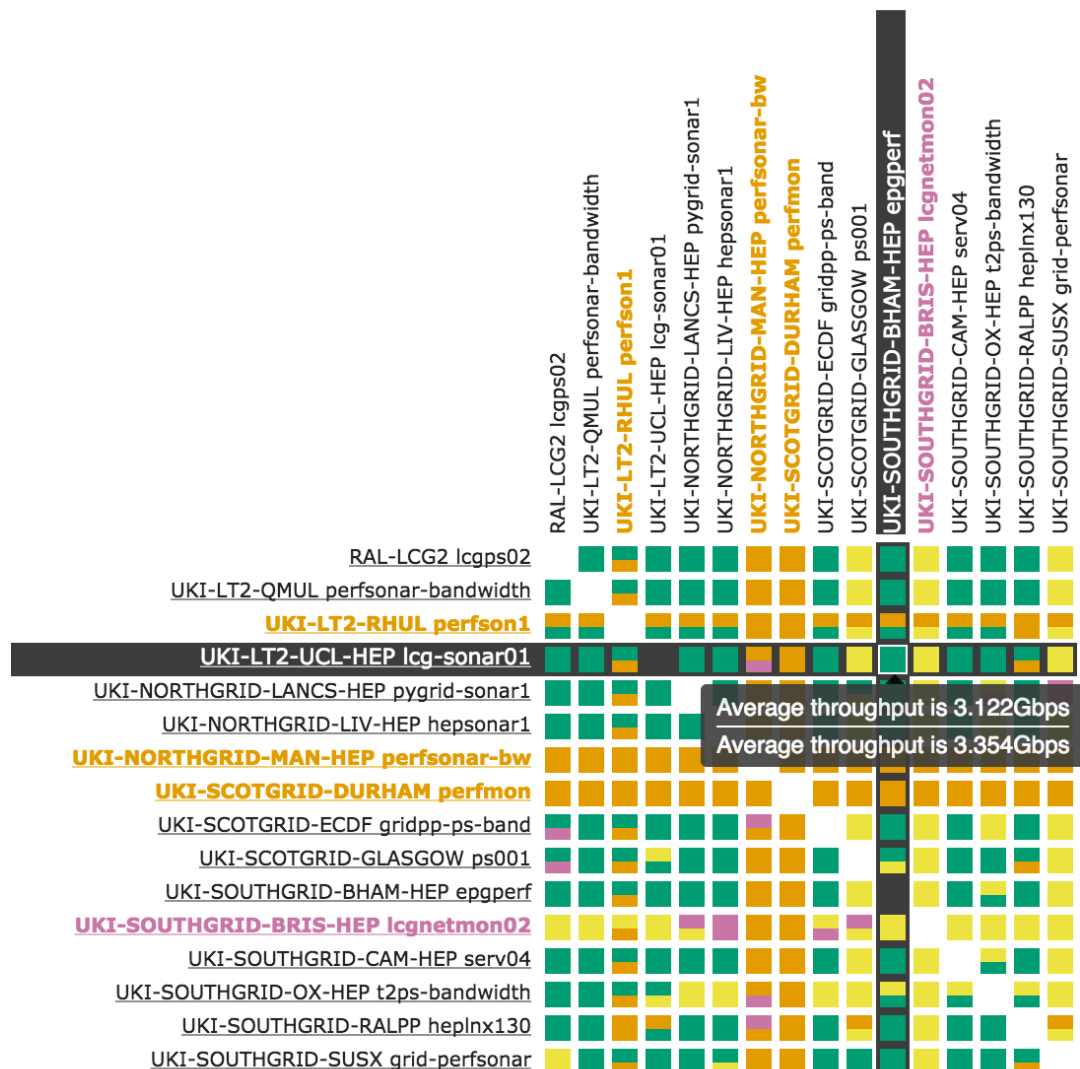[2] https://www.gridpp.ac.uk/

Figure 1: An example perfSONAR mesh – the UK GridPP community

Janet perfSONAR test node

Jisc has recently deployed a 10Gbit/s perfSONAR node near our GEANT PoP in London. This node is available for throughput tests; please contact us via the email addresses at the end of the document if interested. We are also in the process of deploying a node in our Slough data centre.

How do I deploy perfSONAR?

The general recommendation for perfSONAR deployment is to place the system alongside the data transfer facility for which network performance is to be measured, thus in the same subnet, on the same switch equipment, as the facility. Additionally, it is desirable to also deploy a perfSONAR system at the campus edge, typically off an interface of the campus edge router. This allows measurement both between sites, and deeper into sites. Data from such measurement points allows more specific identification of the location of network throughput issues – i.e., is it within one site, or between the sites?
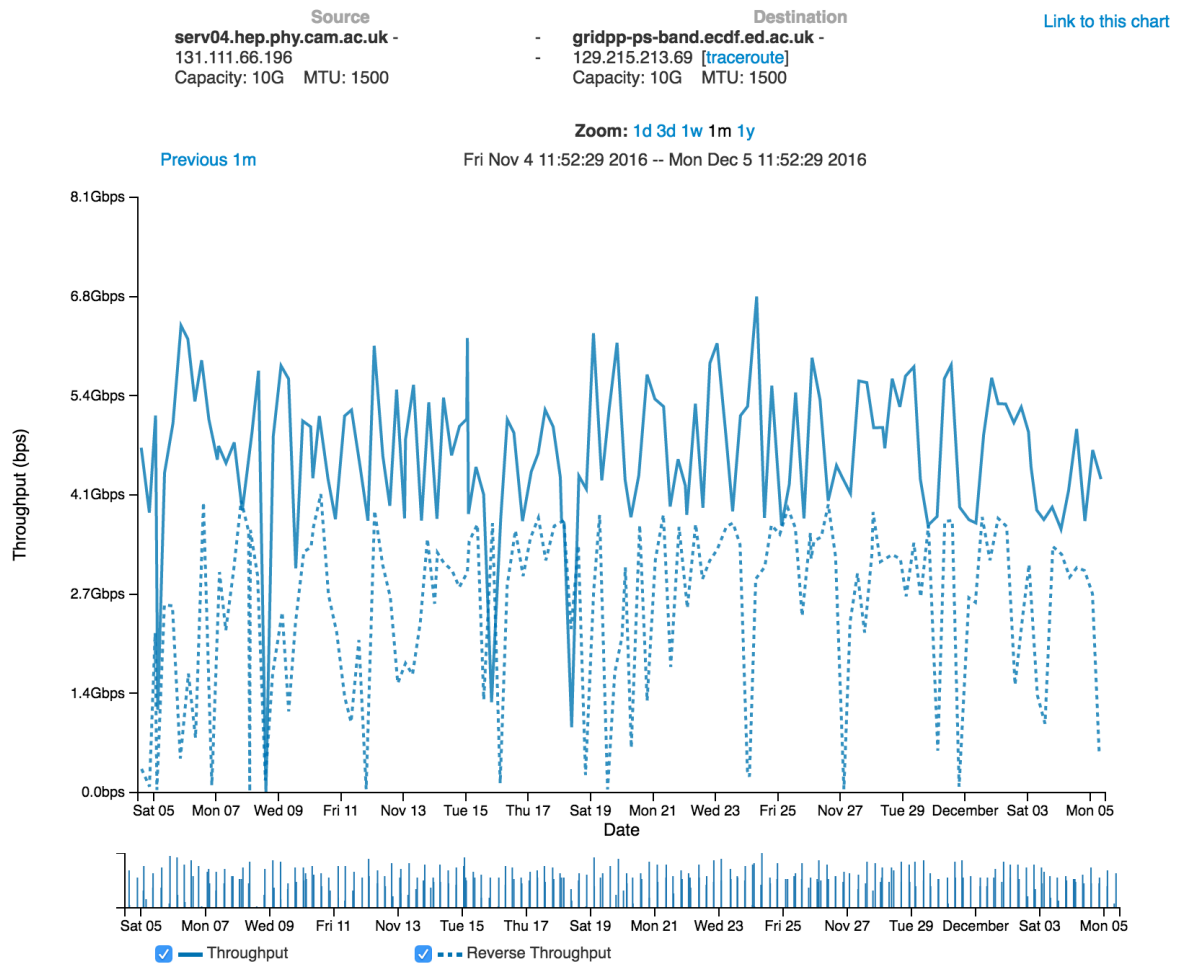
Figure 2: Example of throughput over time between two UK GridPP sites

It is recommended that perfSONAR systems are deployed bare metal, i.e., not virtualised.

Historically, sites would deploy two perfSONAR hosts at each measurement point, one to measure throughout, and one to measure loss and latency. More recently, it has become common practice to deploy systems with two network interfaces, one for measuring throughput (preferably with a 10G interface), and one for measuring loss and latency (typically 1G, which would also allow network management). Such dual-interface deployment is discussed at http://docs.perfsonar.net/manage_dual_xface.html.

General guidance for deploying perfSONAR can be found at:
http://www.perfsonar.net/deploy/

Guidance on hardware specifications is available at:
http://www.perfsonar.net/deploy/hardware-selection/
https://www.perfsonar.net/deploy/hardware-selection/hardware-advice/
https://fasterdata.es.net/performance-testing/perfsonar/ps-howto/hardware/

The most recent version of perfSONAR is 4.0 RC3; with the final version 4.0 expected in April. The toolkit is available either as an image, or as a suite of rpm packages. While perfSONAR can run on a variety of platforms, CentOS7 is commonly used.

perfSONAR measurements can be configured directly from the local system to any remote perfSONAR node. To facilitate a measurement mesh, a central *MaDDash* dashboard server is required. The tests to be run are configured on the *MaDDash* server, and pulled to each measurement system via a configured URL. Typically,

each community of interest would run its own dashboard, e.g. GridPP, DiRAC, etc. It is possible for one perfSONAR node to appear in multiple dashboards. The dashboard server can be run on a VM.

Through its Janet end-to-end performance initiative[3], Jisc is able to operate a dashboard service on request for communities, at least initially, and offer general guidance on the deployment of perfSONAR nodes.

The perfSONAR mesh offers an indication of network performance issues. Once identified, further troubleshooting is required; perfSONAR does not in itself detect where such issues are occurring, though by judicious placement of measurement nodes, i.e. at the campus edge and at the internal data transfer facility at each site, more directed clues can be derived.

More information about perfSONAR, and its use for identifying and troubleshooting network performance issues, can also be found by looking through the presentations at the Jisc Campus Network Engineering for Data-Intensive Science workshop page[4].

What is the network impact of the tests?

Typically, perfSONAR is configured to only run the throughput tests relatively sparsely, e.g., every six hours for a period of around 30 seconds (long enough for TCP to ramp up to give a reasonable indication of available bandwidth). This can be changed if necessary. If running at full capacity, such tests may cause temporary contention on a site's router. However, the throughput tests are TCP, so *iperf* should back off its sending rate when any congestion occurs.

The loss and latency tests are run continuously, and are not disruptive to normal network operation.

What does it cost?

The perfSONAR software is open source, and free to download. It might be expected that a reasonably specified hardware platform, including one 10G interface, would be of the order of £3000-£4000. There are also implicit costs through network cabling/connections and the availability of a free network port into which to plug the system.

Deployment and support effort

The initial configuration of a perfSONAR system can be time consuming and should not be under estimated; a site installing two perfSONAR nodes should budget for up to two weeks of effort from a network engineer or sysadmin to move from bare metal through back-to-back testing and finally into deployment in the network. Support is however available either through Jisc or the Internet2 *perfsonar-users* mail list[5].

Monitoring of the perfSONAR nodes should not be time consuming, but shouldn't be ignored; initially daily monitoring would be recommended whilst the engineers become familiar with "normal" readings. However, as familiarity increases review of the historical data is more likely to only be needed if issues arise. The latest version of the *MaDDash* dashboard is moving to support integration with monitoring tools such as Nagios.

Tim Chown (tim.chown@jisc.ac.uk)
Duncan Rand (duncan.rand@jisc.ac.uk)
Jisc
4th April 2017
Document version 1.0

---

[3] https://www.jisc.ac.uk/rd/projects/janet-end-to-end-performance-initiative
[4] https://www.jisc.ac.uk/events/campus-network-engineering-for-data-intensive-science-workshop-19-oct-2016
[5] https://lists.internet2.edu/sympa/subscribe/perfsonar-user