28/11/2017

perfSONAR – training & dev discussion summary

SIG-PMV Copenhagen, 28 Nov 2017

Tim Chown, Jisc, UK, tim.chown@jisc.ac.uk

Contents

- Brief introduction to perfSONAR

- Background: current use in Janet context
  - User communities and Janet NOC
  - Active vs passive measurements
  - perfSONAR supporting Janet E2EPI activity

- Outcomes of perfSONAR training held in Jisc's Manchester offices on 22/23 Nov
  - Attendees
  - Feature/ wish list
  - Topics arising

What is perfSONAR?

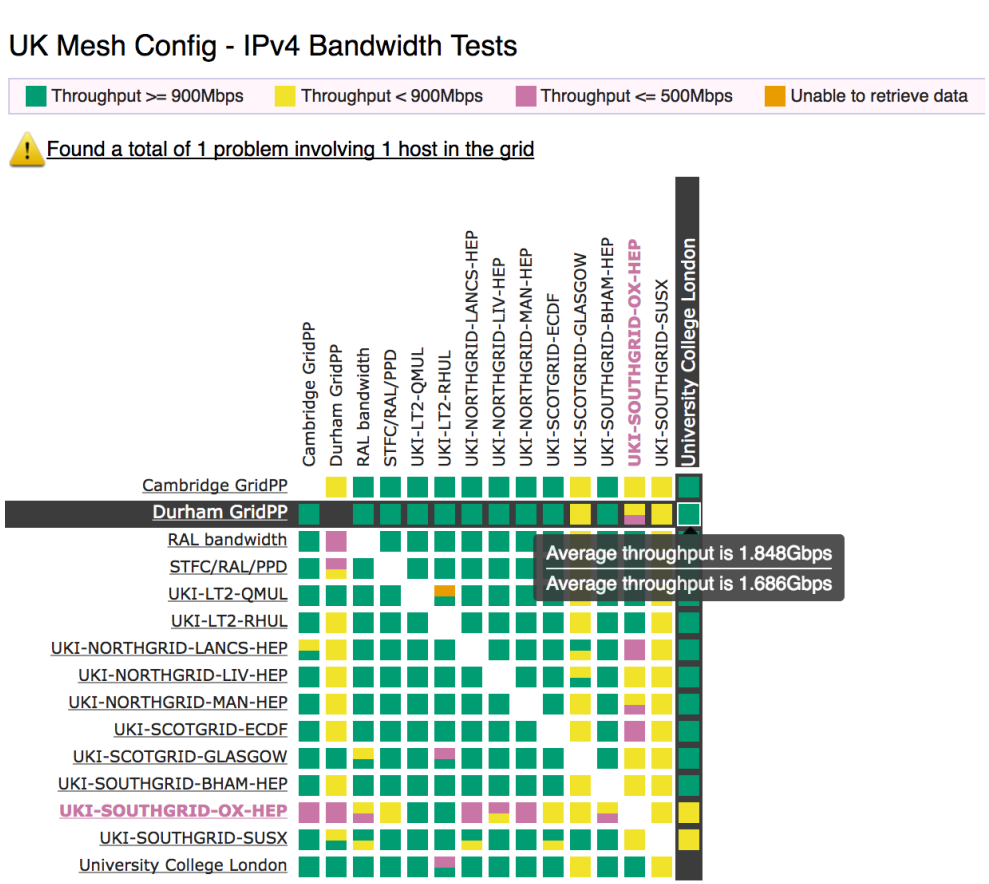
- An active measurement system:
  - Open source; Linux platforms; ISO/RPMs; current version 4.0.1
  - Measures loss, latency, throughput, path
  - Continuous loss/latency, periodic throughput (default 4x per day)
  - Scheduled or ad-hoc tests through pScheduler (introduced in 4.0)
  - Back end database to store historical data
  - Web or CLI-based management
  - Web-based visualisation tools; support for measurement meshes
- Typical deployment:
  - Alongside data transfer endpoint (DTN) and at campus edge
  - Allows performance bottlenecks to be identified
- See https://www.perfsonar.net/

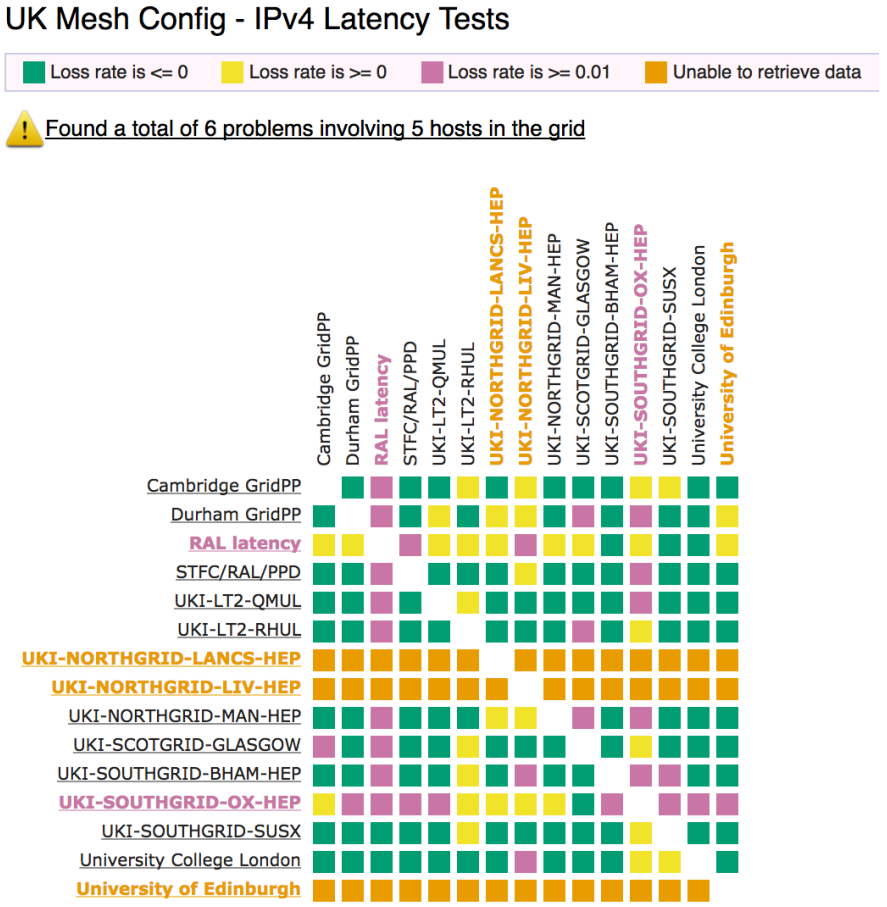
perfSONAR use in Janet context

- Used for some time by the WLCG including UK GridPP community
  - ~20 UK sites, mesh run between LHC experiment participants
  - perfSONAR nodes generally installed by campus IT staff
  - WLCG pushing IPv6, so mesh is dual-stack where possible
  - Mesh is publicaly available:
  - http://ps-dash.dev.ja.net/maddash-webui/index.cgi?dashboard=UK%20Mesh%20Config

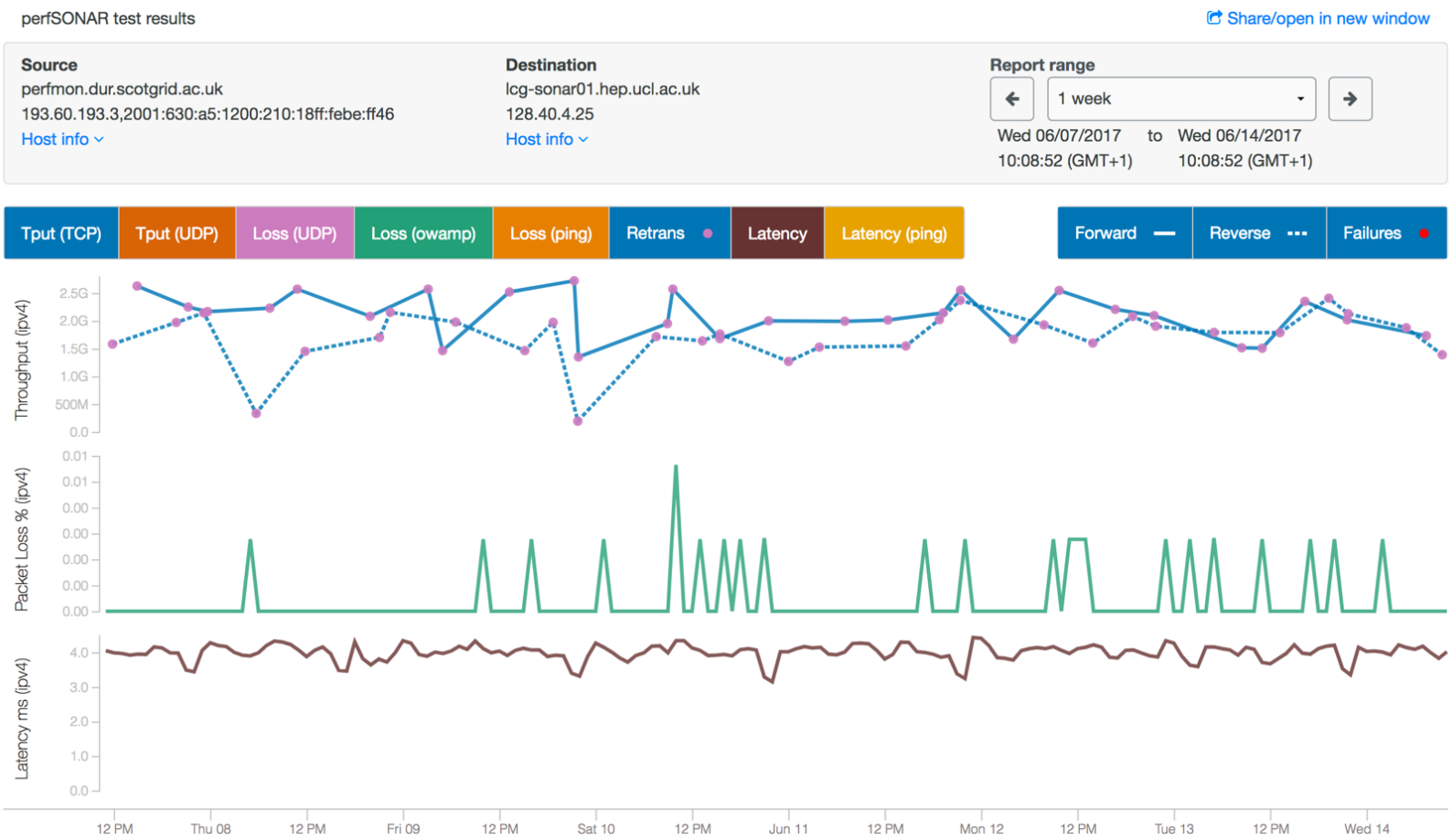
- Jisc's E2EPI is working with sites to improve application throughput
  - Identifying 'problem' cases, e.g., data transferred by hard disks / tape
  - Supporting the work through establishing baseline perfSONAR measurements
  - Encouraging wider use of perfSONAR; hence the training event

Example: UK GridPP perfSONAR mesh





Drilling down on a specific test pair…

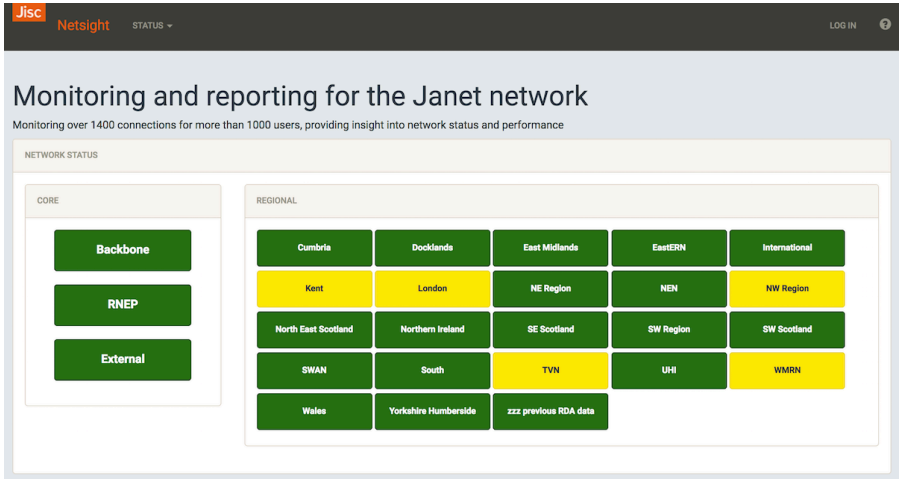


Janet network monitoring

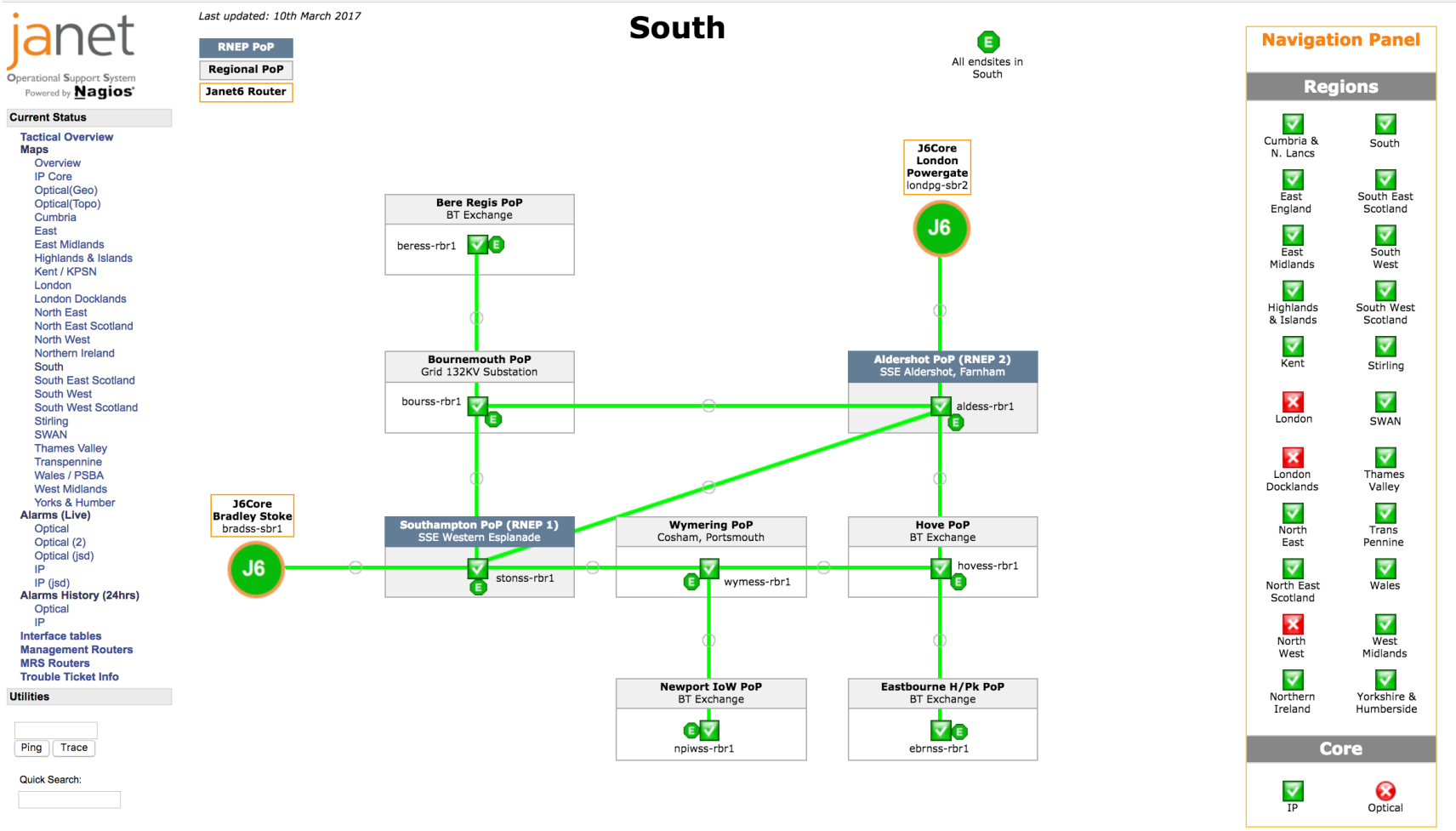
- We're undergoing a review of our network monitoring and management systems
- NOC uses a mixture of tools
- Current functions:
  - Up/down status checks on Janet network elements
  - Link utilisation data; (private) network "weather maps"
  - Site link utilisation; very helpful for site capacity planning

- The Janet NOC has an OSS (private)

- Universities have Netsight view
  - https://netsight.ja.net/
  - More detail if logged in

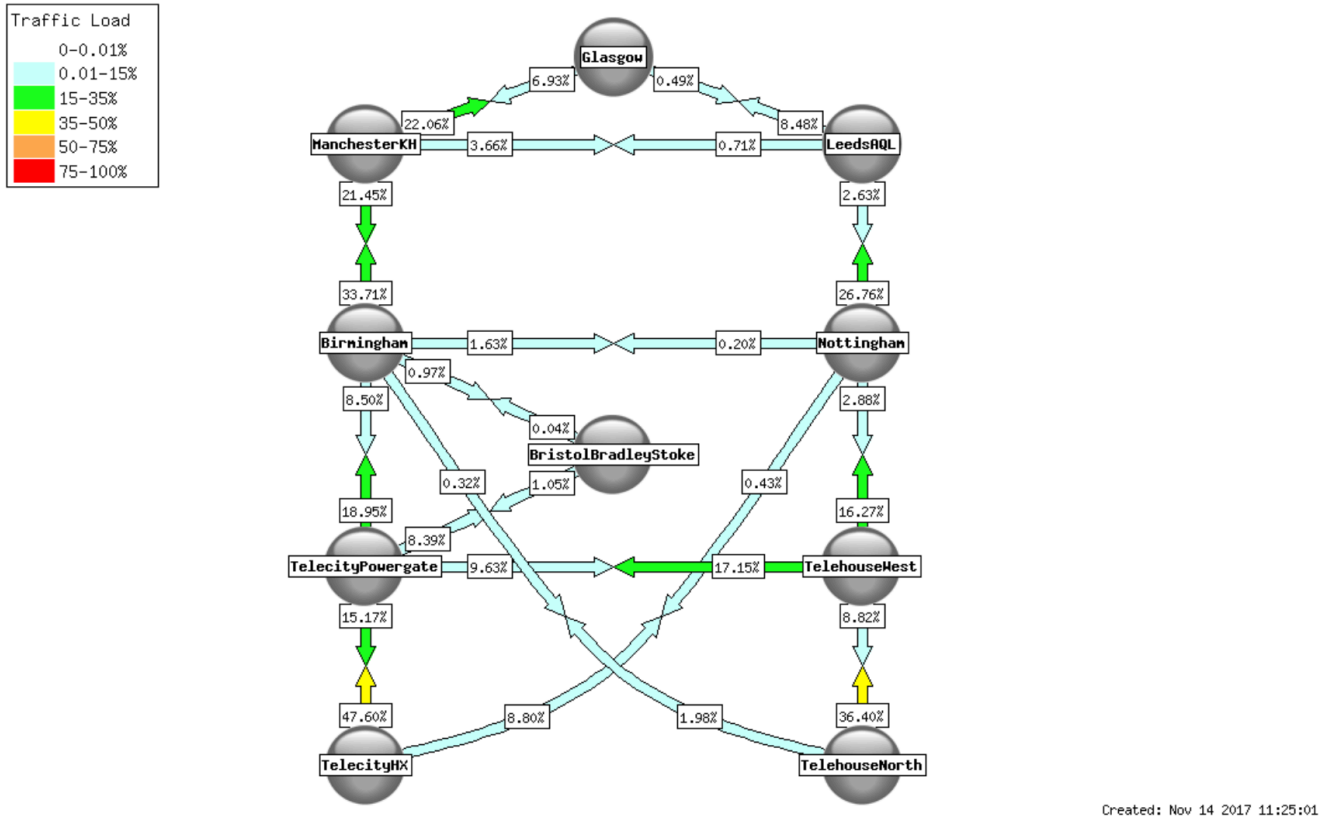


Janet OSS

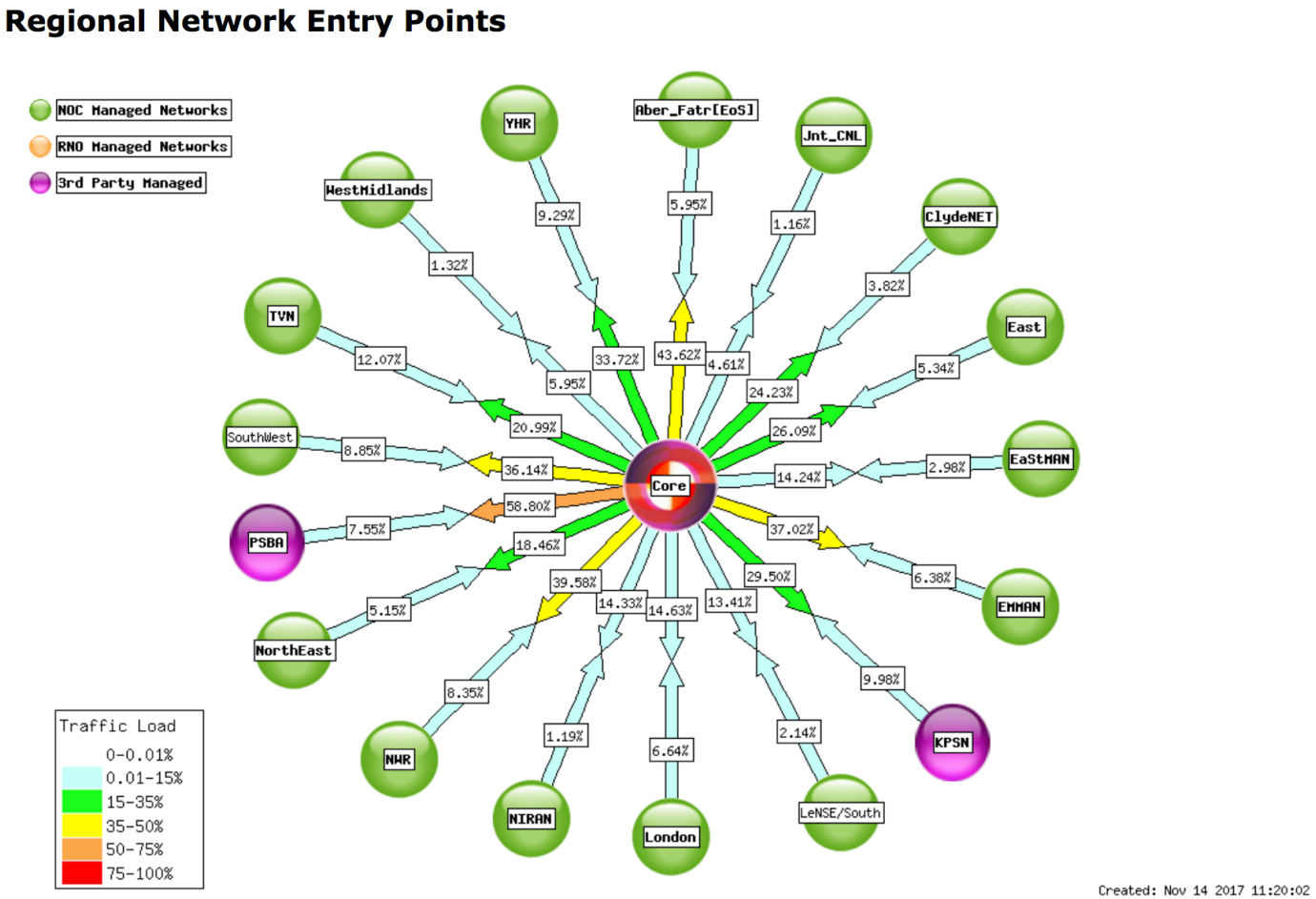

janet
Operational Support System
Powered by Nagios
Current Status
Tactical Overview
Maps
Overview
IP Core
Optical(Geo)
Optical(Topo)
Cumbria
East
East Midlands
Highlands & Islands
Kent / KPSN
London
London Docklands
North East
North East Scotland
North West
Northern Ireland
South
South East Scotland
South West
South West Scotland
Stirling
SWAN
Thames Valley
Transpennine
Wales / PSBA
West Midlands
Yorks & Humber
Alarms (Live)
Optical
Optical (2)
Optical (jsd)
IP
IP (jsd)
Alarms History (24hrs)
Optical
IP
Interface tables
Management Routers
MRS Routers
Trouble Ticket Info
Utilities
Ping
Trace
Quick Search:
Last updated: 10th March 2017
RNEP PoP
Regional PoP
Janet6 Router
South
All endsites in South
J6Core London Powergate londpg-sbr2
J6
Bere Regis PoP
BT Exchange
beress-rbr1
Bournemouth PoP
Grid 132KV Substation
bourss-rbr1
Aldershot PoP (RNEP 2)
SSE Aldershot, Farnham
aldess-rbr1
J6Core Bradley Stoke bradss-sbr1
J6
Southampton PoP (RNEP 1)
SSE Western Esplanade
stonss-rbr1
Wymering PoP
Cosham, Portsmouth
wymess-rbr1
Hove PoP
BT Exchange
hovess-rbr1
Newport IoW PoP
BT Exchange
npiwss-rbr1
Eastbourne H/Pk PoP
BT Exchange
ebrnss-rbr1
Navigation Panel
Regions
Cumbria & N. Lancs
South
East England
South East Scotland
East Midlands
South West
Highlands & Islands
South West Scotland
Kent
Stirling
London
SWAN
London Docklands
Thames Valley
North East
Trans Pennine
North East Scotland
Wales
North West
West Midlands
Northern Ireland
Yorkshire & Humberside
Core
IP
Optical

Janet backbone

Core Network J6

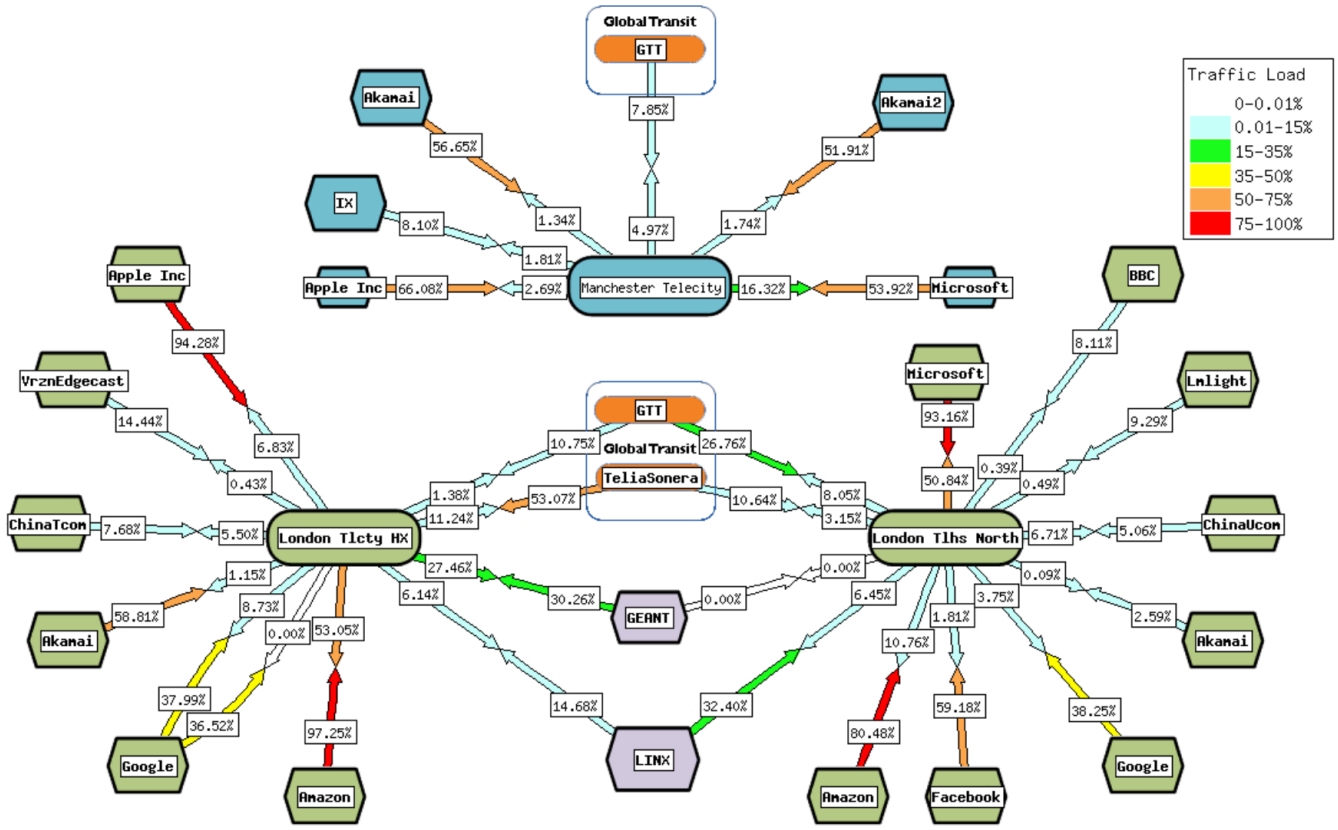


Regional network aggregation



External connectivity

Major External Links



Jisc perfSONAR deployment

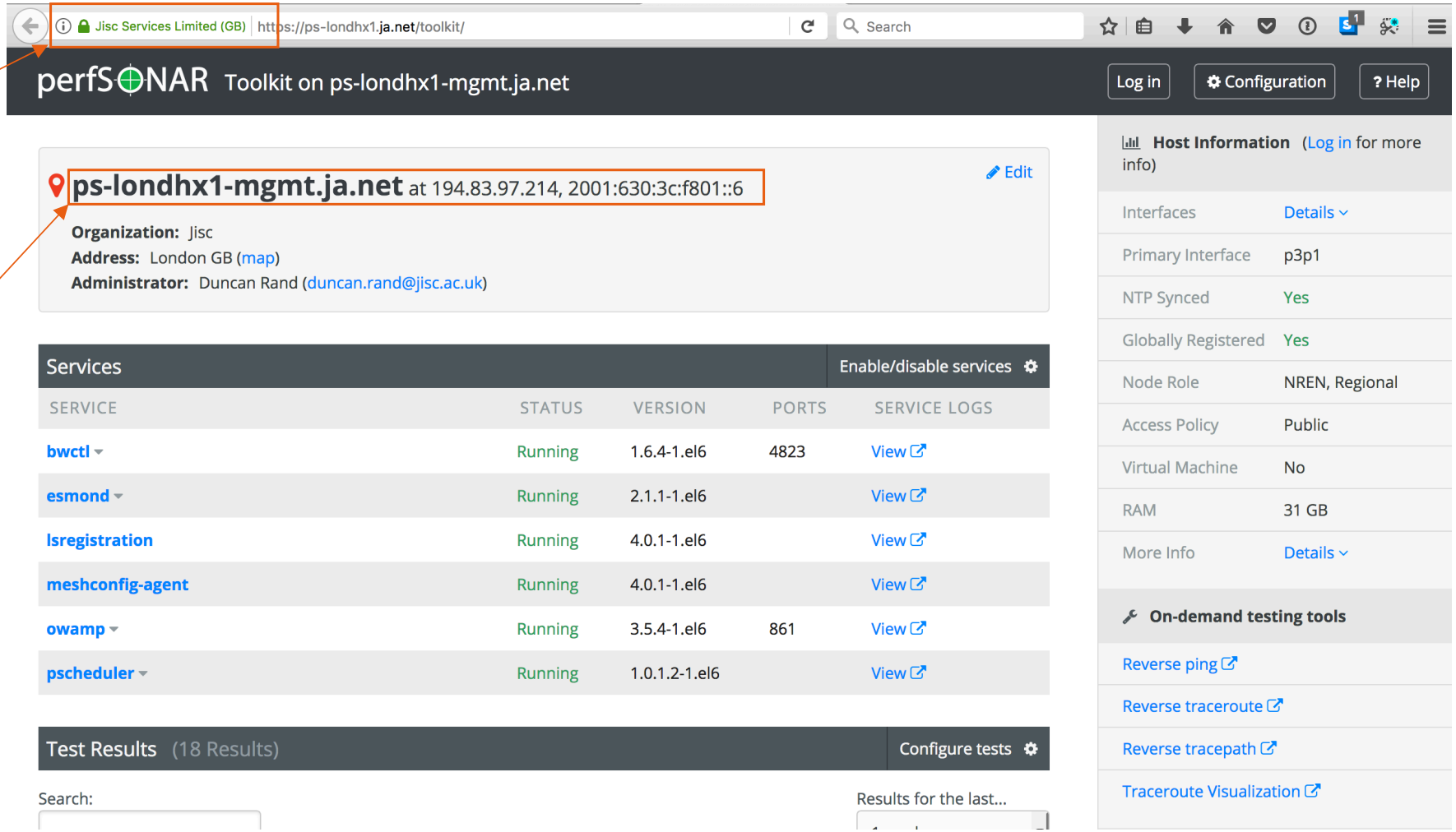
- Piloting perfSONAR nodes on Janet network
  - With a view to a wider rollout; details TBD, hence dev discussion
- Have deployed 10G perfSONAR node at London in Harbour Exchange
  - Dell R620; 2 Intel 2.6 GHz processors (32 cores) and 32 GB memory
  - Intel X520 DP 10Gb DA/SFP+ Server Adapter with 10G and 1 G single-mode SFPs
- Second 10G perfSONAR node coming soon in our Slough data centre
  - Alongside a 10G DTN for disk to disk & transfer tool tests
- Jisc is running a VM that acts as a mesh server
  - Results currently pulled from measurement points
  - In future, likely to also store data centrally

perfSONAR toolkit home page

Added a Jisc certificate

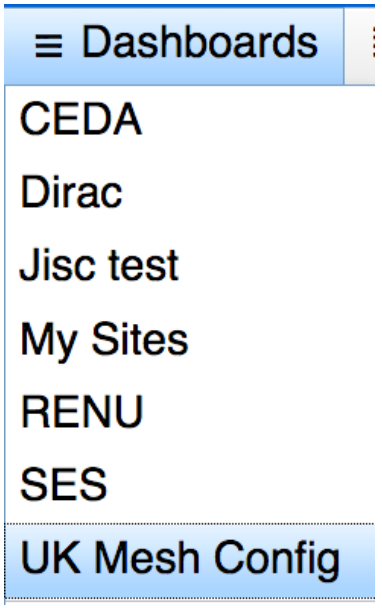
Dual-stack

Possible to set up tests manually, but better to set up a mesh…



MaDDash Dashboard

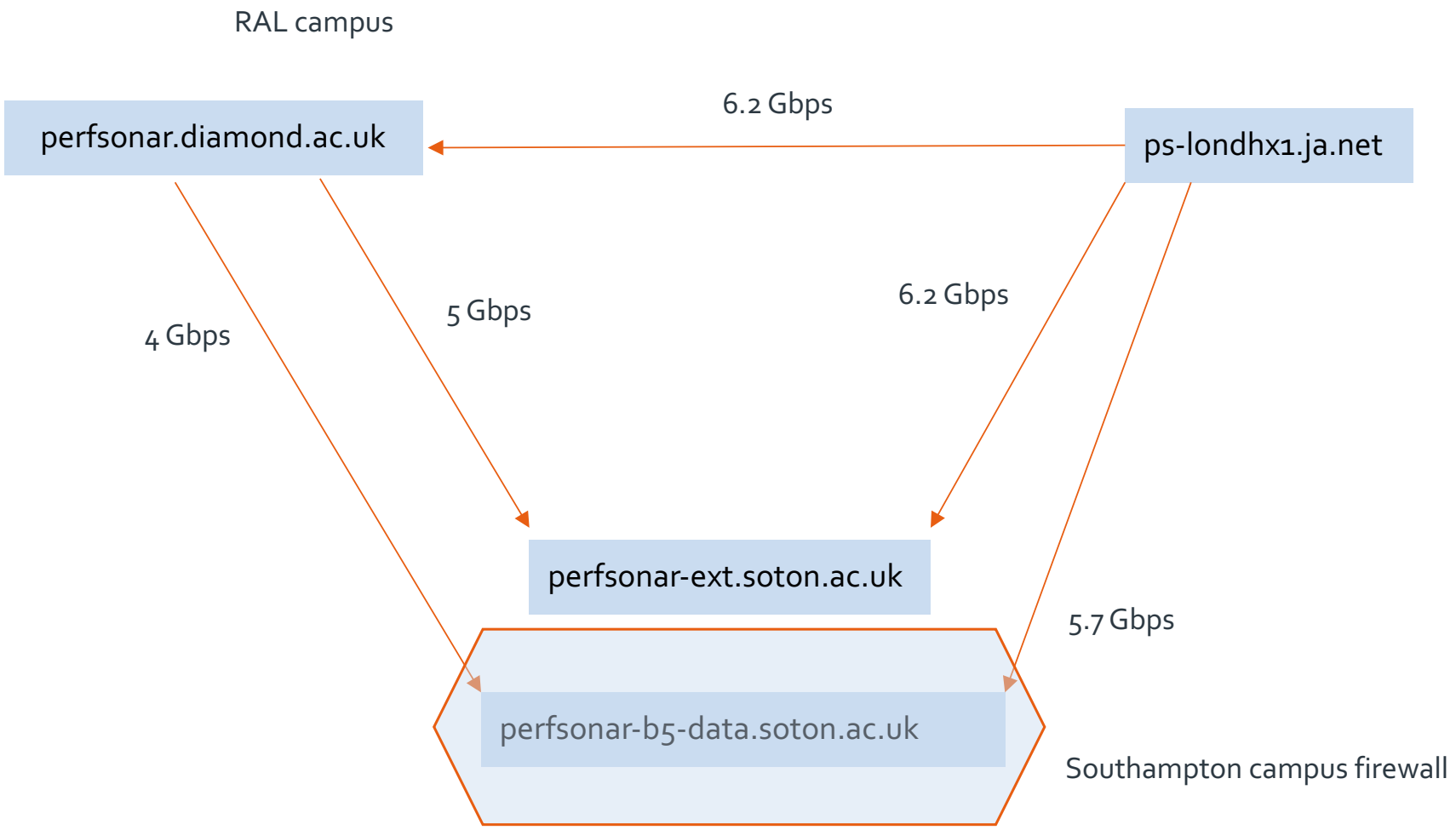
- Supporting several communities that we have been working with in the E2EPI
- *UK Mesh Config*: GridPP mesh, part of the WLCG
- *SES* - Science and Engineering South
  - Pilot with Southampton University and Diamond Light Source
- https://ps-dash.dev.ja.net/maddash-webui/



SES- Science and Engineering South

- Materials science researcher at Southampton, getting a few 10's of Mbit/s for data transfers from Diamond to local lab filestore
  - Moving 10-40TB, six times a year, on physical disks
- Initial work involved adopting Globus Connect to transfer files from DLS
  - Achieved a significant improvement; able to fill 1 Gbit/s local link
- Also installed a perfSONAR host (*perfsonar-b5-data.soton.ac.uk*) on campus next to data storage
- Network to storage upgraded to 10 Gbps
  - Then achieving a few Gbit/s
- Later a perfSONAR host (*perfsonar-ext*) was installed at the Soton border, outside the firewall
  - perfSONAR very useful for understanding effects of changes

SES- Science and Engineering South

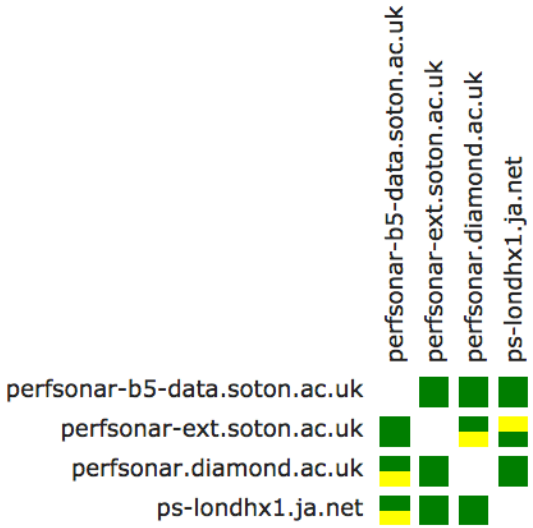


SES- Science and Engineering South

SES - Traceroute

Number of Paths is <= 1 | Number of Paths is >=

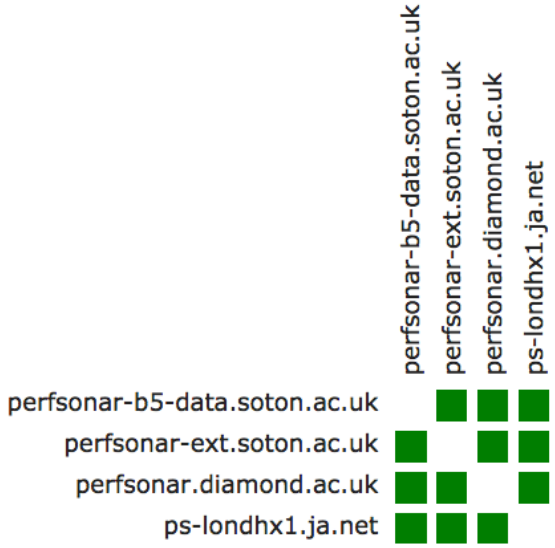
No problems found in grid



SES - Throughput Testing

Throughput >= 900Mbps | Throughput < 900Mb

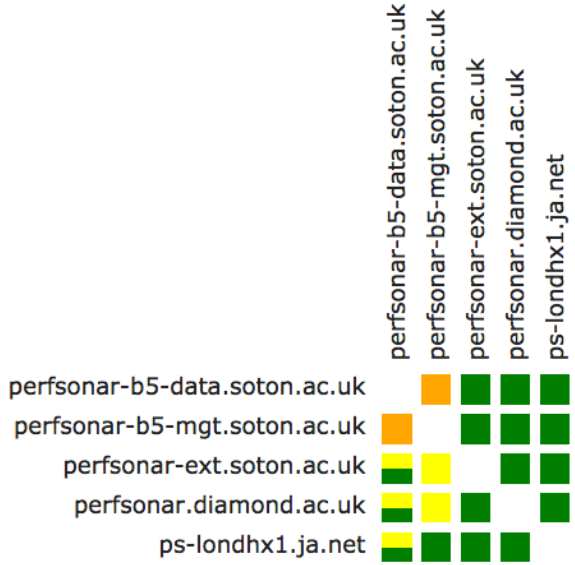
No problems found in grid



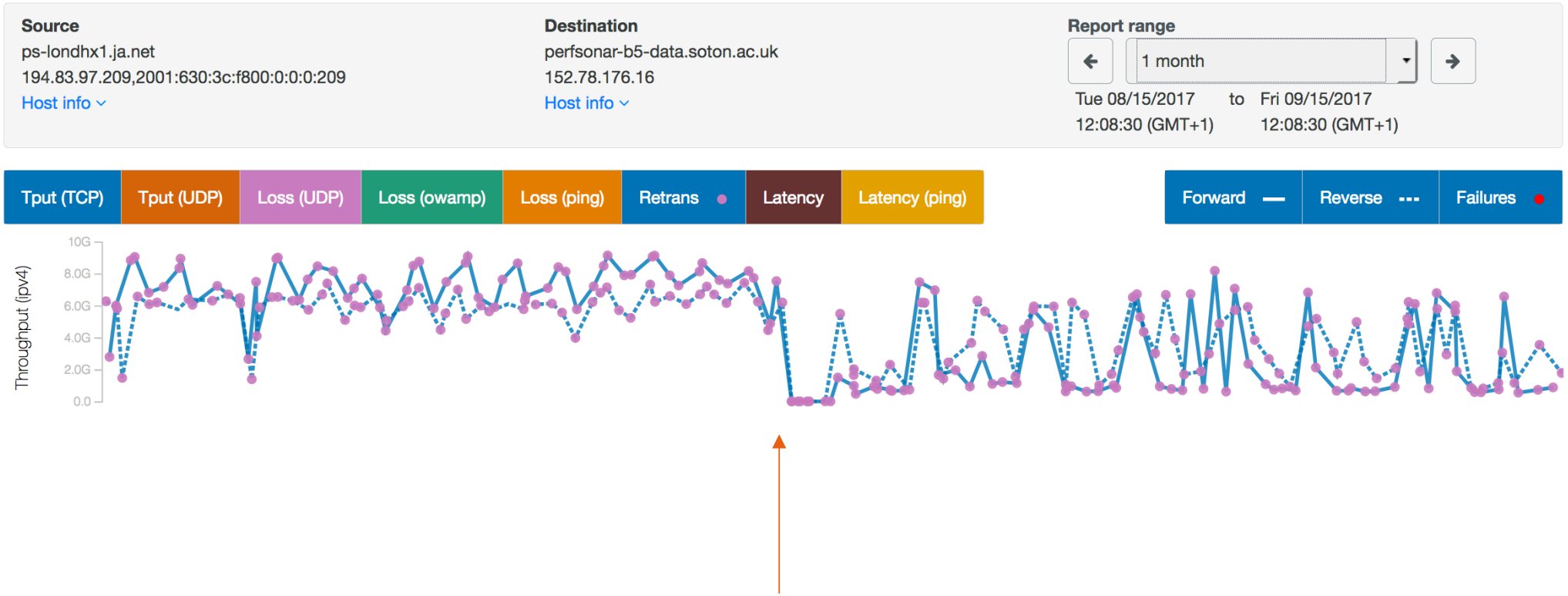
SES - Latency Testing

Loss rate is <= 0 | Loss rate is >= 0 | Loss rate is >= (

No problems found in grid



Inbound throughput to Soton



Firewall update
(30th Aug)

Inbound throughput to Soton

Source
ps-londhx1.ja.net
194.83.97.209,2001:630:3c:f800:0:0:0:209
Host info

Destination
perfsonar-b5-data.soton.ac.uk
152.78.176.16
Host info

Report range
1 year
Wed 11/16/2016 11:12:04 (GMT+0) to Thu 11/16/2017 11:12:04 (GMT+0)

Tput (TCP) | Tput (UDP) | Loss (UDP) | Loss (owamp) | Loss (ping) | Retrans | Latency | Latency (ping)

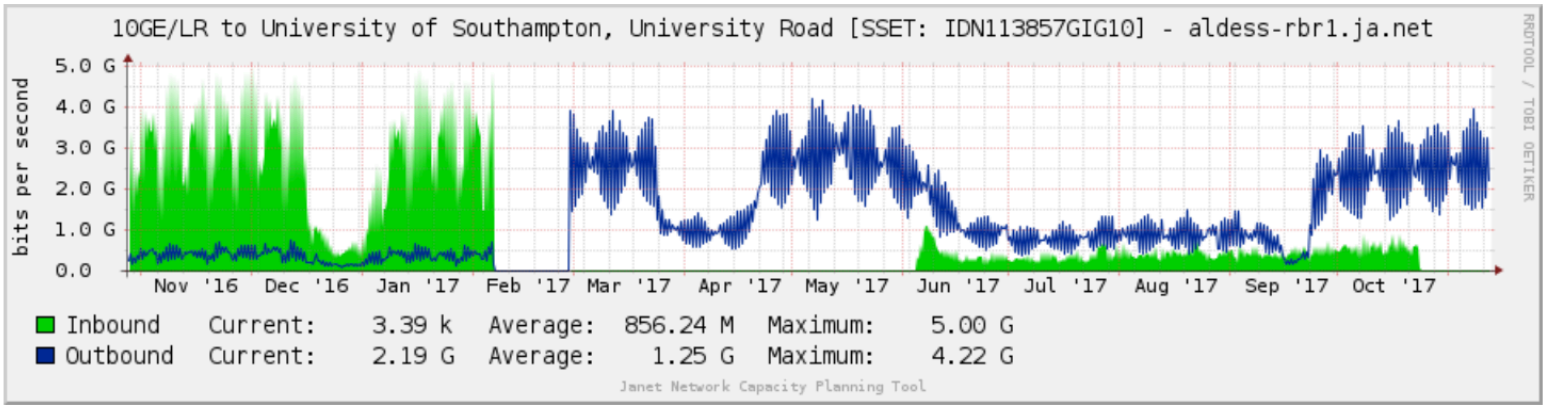
Forward | Reverse | Failures

Firewall update (30th Aug)

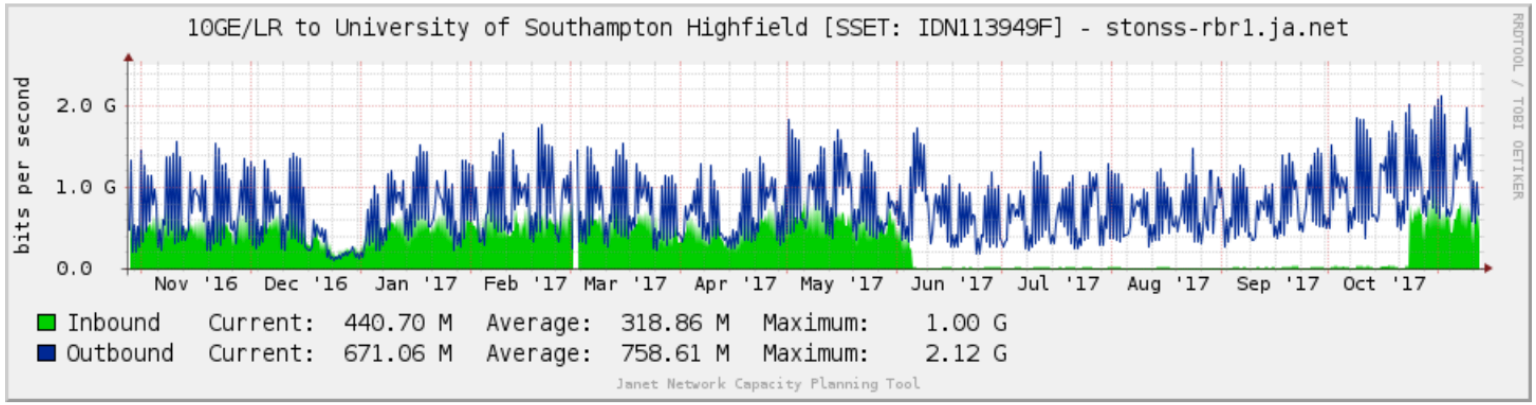
Issue fixed (24th Oct)

Univ. Southampton – passive link measurement

Last 12 Months



Last 12 Months



Active vs passive measurements

- Janet NOC uses a mixture of tools
  - Active status checks (e.g., ping)
  - Router interface utilisation is passive SNMP; e.g., campus uplinks to Janet
  - But as the Southampton example shows, such passive measurement is not so useful for detecting performance issues
  - Note: if users have never experienced good throughput, they're unlikely to complain if they get a bad experience

- Hence the desire/interest in also using active performance measurements
  - Juniper RPM probes, and maybe TWAMP, as presented at a previous SIG-PMV, for L2/L3/L7 measurements
  - perfSONAR; whether deployed by campuses or the Janet NOC
  - Ideally we would maximise the value of campus and NOC deployments

perfSONAR workshop

- Organised by Jisc, and held in Jisc's offices in Manchester on 22/23 Nov
- Trainers were developers on the GEANT project
  - Ivan Granizov and Antoine Delvaux
- Plus training VMs provided by Kurt Baumann at SWITCH
- 1.5 days training, 0.5 days developer discussion
- 20 attendees; free attendance
  - From Jisc, STFC (science facilities), ECMWF, universities
- Logistics and attendees:
  - https://eventr.geant.org/events/2785
- Agenda and materials:
  - https://wiki.geant.org/display/gn42na1/perfSONAR+training+@+Jisc+2017
- Many feature requests raised, and ideas discussed in the day 2 discussion

Training

Developer discussion

Feature / wish list - examples

- The training event raised a few "why can't I?" type questions, e.g.:
  - In the web visualisation, can we have specific time periods entered rather than fixed durations?
  - Displaying results in two directions in the mesh is confusing; is that really needed?
  - The traceroute 'errors' shown in the mesh may be due to local ECMP/load balancing; can we instruct the visualisation tool to ignore certain hops in the path?
  - Can I cancel a mesh and all future tests associated with it?
  - How can I easily archive an ad-hoc test result that I run?
  - Is there an open API to get the historic data out?

Questions…

- Could we use perfSONAR as a harness to measure the performance of different transfer tools, e.g., GridFTP, WDT, Aspera, … perhaps between DTNs?

- pScheduler has already been extended for DNS tests, tcpdump, …
- There are now multiple plugins and categories of plugins, including measurement archive category
- Can extend pScheduler functionality to support additional archivers, e.g. Elastic, or OpenTSDB
- In principle can schedule tests and post results to a database
- Requires development effort on a transfer/GridFTP plugin
- Recorded webinar on the topic on the perfSONAR YouTube channel

Questions…

- Can we use perfSONAR to run TWAMP tests against our router infrastructure (largely Cisco and Juniper)?

- Coming in 4.0.2/4.1
- Available for testing soon
- Developers would welcome people offering to test
- Will be a ‘standard’ measurement

Questions…

- How might we integrate active perfSONAR measurements with other passive measurements, e.g. Netsight (SNMP) data on link utilisation?
- Better analysis of variation in observed pS measurements
- Possible work area for GN4-3?
- Technically possible
- Need to identify resource
- Interesting questions over user interface; overlay of utilisation with perfSONAR view?
- Overlay a trend / baseline?
- Backend Esmond / Cassandra has a REST API
  - http://software.es.net/esmond/

Questions…

- What are the recommendations for automated management of an organisation's perfSONAR infrastructure?

- ESnet are using Ansible
- GEANT operations are using puppet in general, and for perfSONAR
- GEANT small node service will need a solution
- The perfSONAR packages do not hinder any particular approach

Questions…

- Is it possible to use network topology information, and known network locations of perfSONAR nodes, to arrange that tests minimise potential duplication of measurements?

- Not clear
- There is an issue if a mesh is 'full' of tests, particularly where the same pairs of hosts appear
- How to identify duplication?
- Some smarts in principle possible via meshconfig UI

Questions…

- How do we best design perfSONAR tests to effectively show up soft problems in the network?
- What general principles should we follow?

- Comes with experience
- General recommendation is for a perfSONAR node alongside your DTN, and at the campus edge, and build from there
- Look at the data flows, and place measurement points along the path

Questions…

- When setting up pScheduler tests on demand, how should we best ensure appropriate authentication / authorisation for the test to be run?
- Concern expressed on possible remote 'abuse' of the test infrastructure
- The system is inherently open
- Authorisation is IP-based through the limits file; can also limit on duration, throughput, …
- Perhaps pScheduler can alert on test requests above a certain threshold?
- You can monitor traffic volume to/from a perfSONAR node via SNMP
- Perhaps nodes on a common mesh can be trusted more?

Questions…

- What tools should we use to measure the load on a perfSONAR node? At least CPU and memory, perhaps more?
  - SNMP for network link utilisation
  - CPU/memory use
  - Disk space for archives
  - Free / available time in the test schedule
  - Ganglia?

Questions…

- What considerations are there for running a perfSONAR instance in a cloud / virtual environment, e.g., to measure performance to/from a commercial cloud provider, or a private OpenStack instance?
- Docker work is ongoing
- There is interest


Tim Chown
Network Development Manager
Jisc, UK
tim.chown@jisc.ac.uk
jisc.ac.uk