

6/06/2019

100G perfSONAR and DTN hosts in Jisc

Duncan Rand and Raul Lopes

- » Jisc's End-to-End Performance Initiative (E2EPI) seeks to help our members make the most of their Janet connection
 - › Focused mainly on larger scale data transfers
 - Typically scientific data such as synchrotron and cryo-EM (DLS), particle physics (LHC), astronomy (SKA), climate (CEDA), genomics and bioinformatics etc
 - › But approaches can be applied more broadly

- » Web site for more info:
 - › <https://www.jisc.ac.uk/rd/projects/janet-end-to-end-performance-initiative>
 - › Pointers to workshops, guidance, etc.

- » E2EPI mail list:
 - › <https://www.jiscmail.ac.uk/cgi-bin/webadmin?Ao=E2EPI>

- » Providing advice to members
 - › Use Janet for data transfers, not physical media!
 - › 1TB per hour is ~2Gbit/s; 100TB per day is ~10Gbit/s
 - › Firewalls designed for thousands of small flows may not cope well with large flows
 - › Consider campus architecture – “Science DMZ”, or “Research Data Transfer Zone” (see Oct 2016 workshop slides for more details)
 - › Differentiate science and regular “business” traffic
 - › Optimise data transfer nodes (DTNs) at the campus edge

- » We have interacted with 40-50 projects or organisations

- » Two Jisc 10G-connected perfSONAR nodes available for testing:
 - › London: <https://ps-londhx1.ja.net/toolkit/>
 - › Slough: <https://ps-slough-10g.ja.net/toolkit/>

- » We also provide VM-based hosting of meshes for communities and MaDDash dashboards

- » Example:
 - › UK GridPP: <https://ps-dash.dev.ja.net/maddash-webui/index.cgi?dashboard=UK%20Mesh%20Config>
 - › Also other meshes CEDA, DIRAC, SES, STFC, SingAREN etc

- » We have deployed a reference DTN in our Slough DC
 - › Specified with NVMe SSD; can read/write at 10Gbit/s
 - › Available to member sites for disk-to-disk tests
 - › Co-located with our Slough perfSONAR system
 - › Offers a Globus Connect endpoint
 - › Have used the DTN in tests with SingAREN in Singapore (achieved ~8Gbit/s)

- » Also have a second experimental DTN in Slough
 - › Allows tests of alternative protocols and tools
 - › e.g., QUIC, TCP-BBR, WDT, ...

- » A number of Janet connected sites have or are planning to upgrade their Janet connections to 100G
 - › RAL and Imperial College London already have 100G resilient connections
 - › University of Manchester is proposed
 - › Other sites in discussion e.g. ECMWF, EBI, ACF, QMUL
- » Now we would like to help these sites make the most of the new connections
- » Janet core being prepared for such 100G connected sites
- » Janet E2EPI wishes to gain experience with 100G networking and provide a focus for the discussion and dissemination of technical aspects to member sites
- » 100G workshop last year (<https://www.jisc.ac.uk/events/100-gigabit-ethernet-networking-workshop-04-jul-2018>)

- » For sites that are upgrading their Janet connections to 100G, as well as those already connected to Janet at > 10G, we would like to provide, on the Janet network
 - › 100G connected perfSONAR host
 - › 100G DTN running Globus
- » Planning to move existing 10G perfSONAR host and 10G DTN from Slough to Northern Data Centre in Leeds
- » Move both testbed nodes from Harwell to Slough and transfer NVMe SSD from pS node to DTN to maximise available disk space

- » Set up a 100G testbed
- » Two servers in the same rack with Mellanox Connect-X4 NICs and a Mellanox MSN2100 Switch
- » Implemented fasterdata.net optimisations (TCP buffer sizes, setting CPU governor to 'performance' etc)
- » Servers running CentOS 7.6 (default kernel 3.10.0), upgraded kernel to 5.0.9
- » Using
 - › iperf3 single stream get rate of ~30 Gbps
 - › iperf2 with multiple streams get 98 Gbps

perFSONAR test results - [documentation](#)

[Share/open in new window](#)

Source

192.168.220.2

[Host info](#) ▾

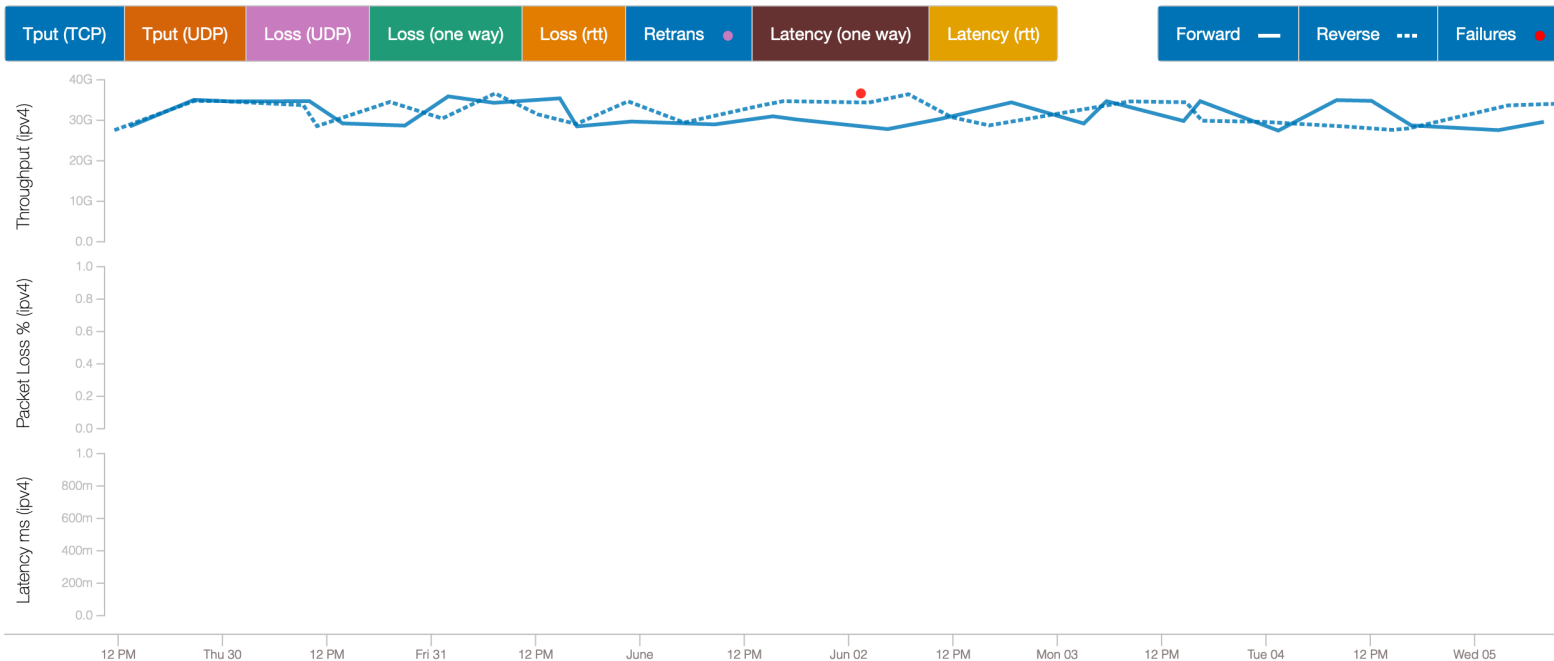
Destination

192.168.220.1

[Host info](#) ▾

Report range

Wed 05/29/2019 10:26:13 (GMT+1) to Wed 06/05/2019 10:26:13 (GMT+1)

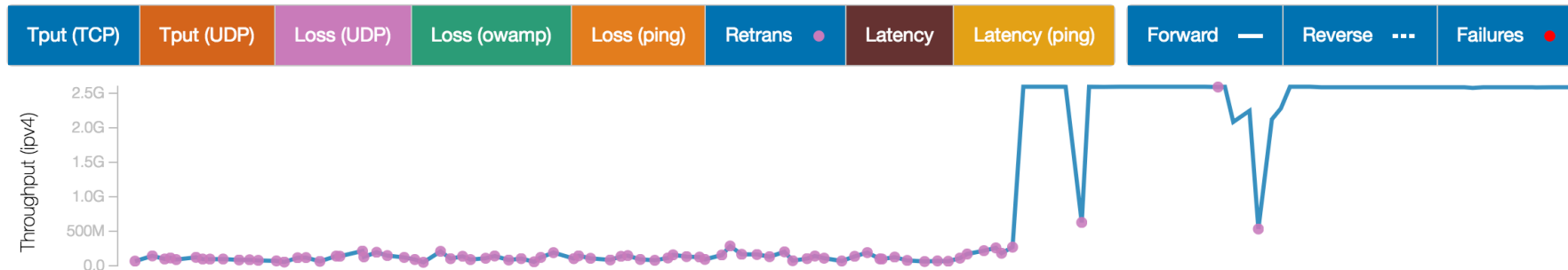


-
- » Do we want to use one (iperf3) or multiple streams (iperf2)?
 - › One stream to test performance of host, 'quality' of transfer
 - › Multiple transfers to see what an end-to-end link is actually capable of

- » Sites may be connected to their NREN at say 100G, but not necessary for hosts within that site to have 100G hosts
- » A large WLCG site might have 50-100 hosts at 10G – clearly capable of more throughput than a 100G WAN connection to NREN
- » DOE sites in the US typically have tens of hosts ([ESnet Petascale project](#))
 - › “OLCF now has 28 transfer nodes in production on 40-Gigabit Ethernet”
 - › NERSC has around 12 DTNs
- » Perhaps stating the obvious: easier to gain throughput with a large number of DTNs connected at 10G than a small number connected at 100G

- » We have been working with Imperial College and SingAREN
- » New genomics project, needs to send/receive up to 200 TB of data between sites
- » perfSONAR highlighted one-way issue on Singapore to Janet path; faulty hardware
- » Resolved with TEINCC/CERNET, now get 2.5 Gbit/s single stream, both ways

Source bwctl-10g-ps.singaren.net.sg 203.30.39.13,2001:df0:21a:0:f6e9:d4ff:fea4:6432 Host info ▾	Destination ps-londhx1.ja.net 194.83.97.209,2001:630:3c:f800:0:0:0:209 Host info ▾	Report range ← 1 month → Sat 03/10/2018 to Tue 04/10/2018 11:25:07 (GMT+0) 12:25:07 (GMT+1)
---	--	---



» We have been working with Imperial College and SingAREN

» New News: UK-Singapore connection being upgraded from 10G to 100G

» perfSONAR Both Imperial College and NSCS connected at 100G so in principle a 100G connection all the way

» Resol Imperial College Research Computing Service has a 100G perfSONAR host and is installing a 100G connected DTN running Globus

Source

bwctl-10g
203.30.39
[Host info](#)

Tput (TCP)

Throughput (ipv4)

2.5G
2.0G
1.5G
1.0G
500M
0.0

- What data transfer rates might we hope to achieve in future?

Thank you

Duncan Rand

duncan.rand@jisc.ac.uk

jisc.ac.uk