Jisc

Janet End-to-End Performance Initiative

Dr Tim Chown, Network Development Manager, Jisc

HEPSYSMAN – RAL – 14 June 2017

Overview

» Overall objective: help sites make optimal use of their Janet connectivity

» Seeing a growth in data-intensive science applications

- › Includes established areas, most notably GridPP (many of you here ☺)
- › As well as new areas like cryo-electron microscopy

» Seeing an increasing number of remote computation scenarios

- › e.g., scientific networked equipment, no local compute - might require 10Gbit/s to return computation results on a 100GB data set to a researcher for timely visualisation

» Starting to see more 100Gbit/s connectivity requests

- › Likely to have challenging data transfer requirements behind them, e.g. SKA

Janet End-to-End Performance Initiative

- The goals include:
  - Engaging with existing and data-intensive research communities and identifying emerging communities
  - Creating dialogue between Jisc, computing service groups, and research communities
  - Holding workshops, facilitating discussion on e-mail lists, etc.
  - Helping researchers manage expectations
  - Establishing and sharing best practices in identifying and rectifying causes of poor performance
  - Promoting good practices in campus network engineering, esp. 'Science DMZ'
  - Promoting deployment of performance measurement tools, esp. perfSONAR

- More information:
  - https://www.jisc.ac.uk/rd/projects/janet-end-to-end-performance-initiative

Understanding the factors affecting E2E

- Achieving optimal end-to-end performance is a multi-faceted problem.
- It includes:
  - Appropriate Janet provisioning between the end sites
  - Properties of the local campus network (at each end), including capacity of the Janet connectivity, internal LAN design, the performance of firewalls and the configuration of other devices on the path
  - End system configuration and tuning; network stack buffer sizes, disk I/O, memory management, etc.
  - The choice of tools used to transfer data, and the underlying network protocols, e.g. xrootd, Globus, rsync, Aspera, or…?

- To optimise end-to-end performance, you need to address each aspect
- There will inevitably be a bottleneck somewhere
  - Our initial focus has been network elements, but includes the end systems

Getting a handle on requirements

- Is there good dialogue at your site between data-intensive researchers and the IT staff?
  - If so, is this happening on a regular basis?
  - Ideally want to be able to plan ahead, rather than adapt on the fly

- Do you conduct networking 'future looks'?
  - The LHC community is quite good at this; experiments are well-defined in advance
  - Any step changes – certainly at 10Gbps - might dwarf your site's organic growth
  - The capacity issue ought to have attention at CIO or PVC Research level

- Do you know what your application elephants are?
  - What's the breakdown of your site's network traffic?
  - How are you monitoring network flows?

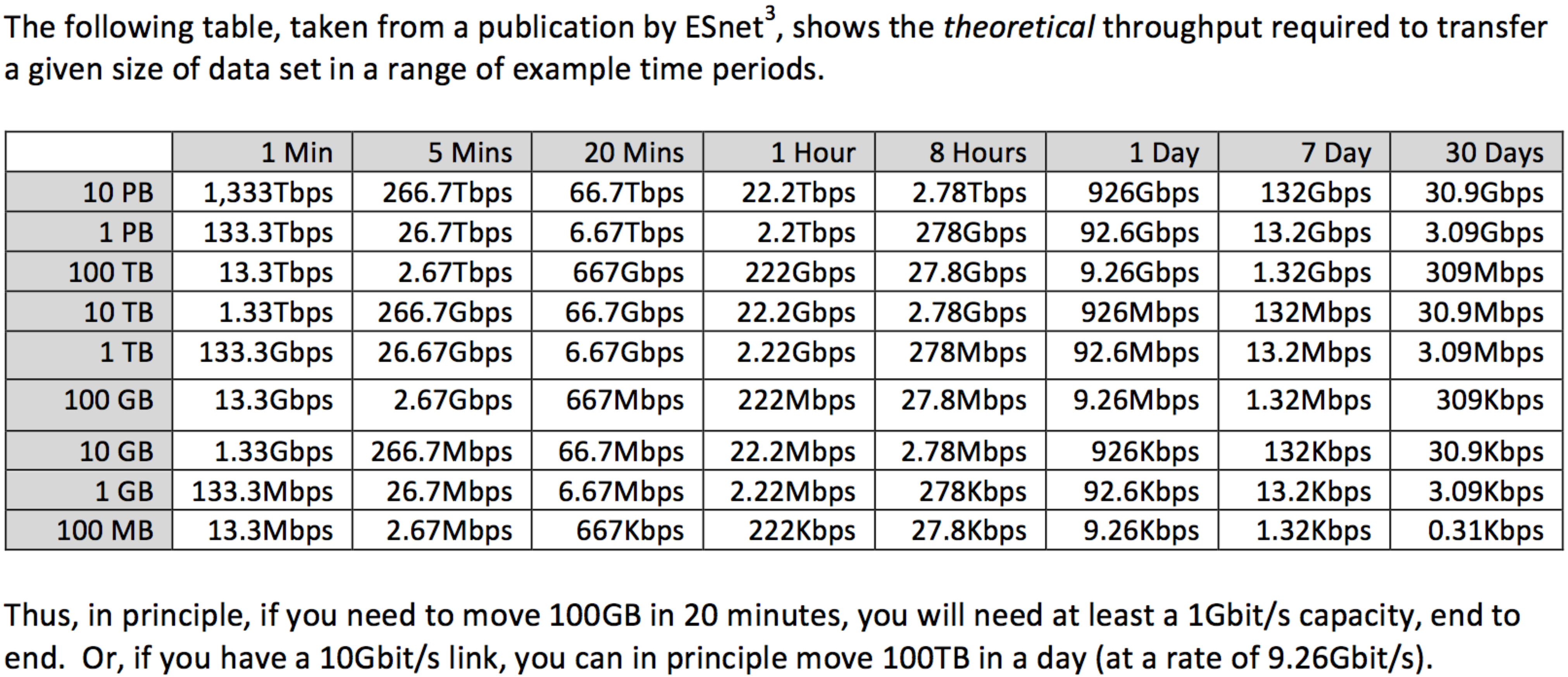
Researcher expectations

- How can we help set and manage researcher expectations?

- One aspect is helping them articulating their network requirements
  - At a simple level, volume of data in time X => data rate required
- It's also about understanding practical limitations
  - Can determine theoretical network throughput
  - e.g., in principle, you can transfer 100TB over a 10Gbit/s link in 1 day
  - But in practice, many factors may prevent this, as mentioned earlier
- We're encouraging researchers to speak to their computing service
  - Not to try, give up, and start sending their data on hard disks!

- Initial guidance document (feedback welcomed!):
  - https://community.jisc.ac.uk/system/files/1862/Network-Expectations-v1-0.pdf

Example from the doc - theoretical throughput



The following table, taken from a publication by ESnet[3], shows the *theoretical* throughput required to transfer a given size of data set in a range of example time periods.

	1 Min	5 Mins	20 Mins	1 Hour	8 Hours	1 Day	7 Day	30 Days
10 PB	1,333Tbps	266.7Tbps	66.7Tbps	22.2Tbps	2.78Tbps	926Gbps	132Gbps	30.9Gbps
1 PB	133.3Tbps	26.7Tbps	6.67Tbps	2.2Tbps	278Gbps	92.6Gbps	13.2Gbps	3.09Gbps
100 TB	13.3Tbps	2.67Tbps	667Gbps	222Gbps	27.8Gbps	9.26Gbps	1.32Gbps	309Mbps
10 TB	1.33Tbps	266.7Gbps	66.7Gbps	22.2Gbps	2.78Gbps	926Mbps	132Mbps	30.9Mbps
1 TB	133.3Gbps	26.67Gbps	6.67Gbps	2.22Gbps	278Mbps	92.6Mbps	13.2Mbps	3.09Mbps
100 GB	13.3Gbps	2.67Gbps	667Mbps	222Mbps	27.8Mbps	9.26Mbps	1.32Mbps	309Kbps
10 GB	1.33Gbps	266.7Mbps	66.7Mbps	22.2Mbps	2.78Mbps	926Kbps	132Kbps	30.9Kbps
1 GB	133.3Mbps	26.7Mbps	6.67Mbps	2.22Mbps	278Kbps	92.6Kbps	13.2Kbps	3.09Kbps
100 MB	13.3Mbps	2.67Mbps	667Kbps	222Kbps	27.8Kbps	9.26Kbps	1.32Kbps	0.31Kbps

Thus, in principle, if you need to move 100GB in 20 minutes, you will need at least a 1Gbit/s capacity, end to end. Or, if you have a 10Gbit/s link, you can in principle move 100TB in a day (at a rate of 9.26Gbit/s).

Janet network engineering – capacity planning

» From Jisc's perspective, it's important to ensure there is sufficient capacity in the network for its connected member sites

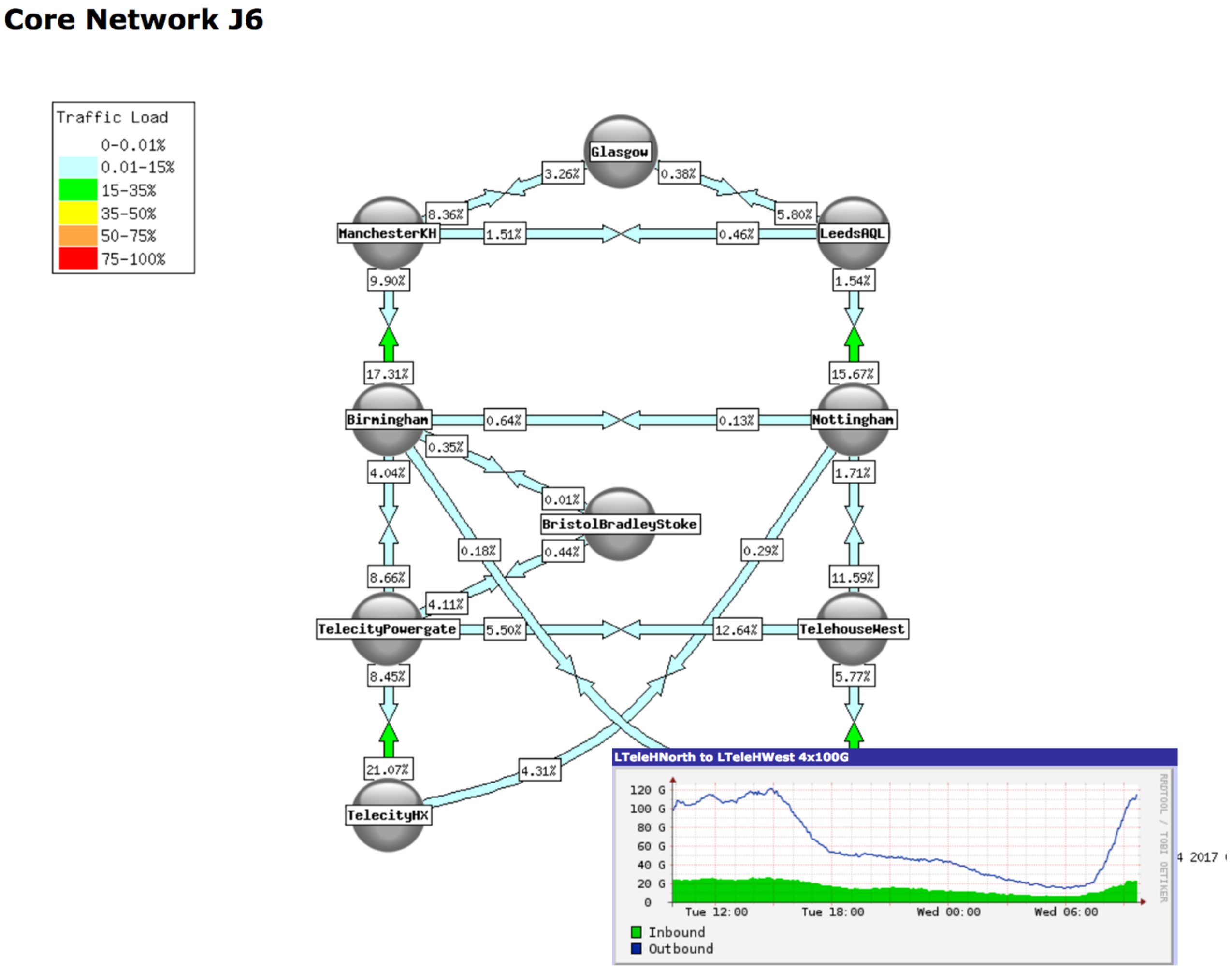
» We perform an ongoing review of network utilisation

- › Provision the backbone network and regional links
- › Provision external connectivity, to other NRENs and networks
- › Observe the utilisation, model growth, predict ahead
- › Step changes have bigger impact at the local rather than backbone scale

» Janet has no differential queueing for regular IP traffic

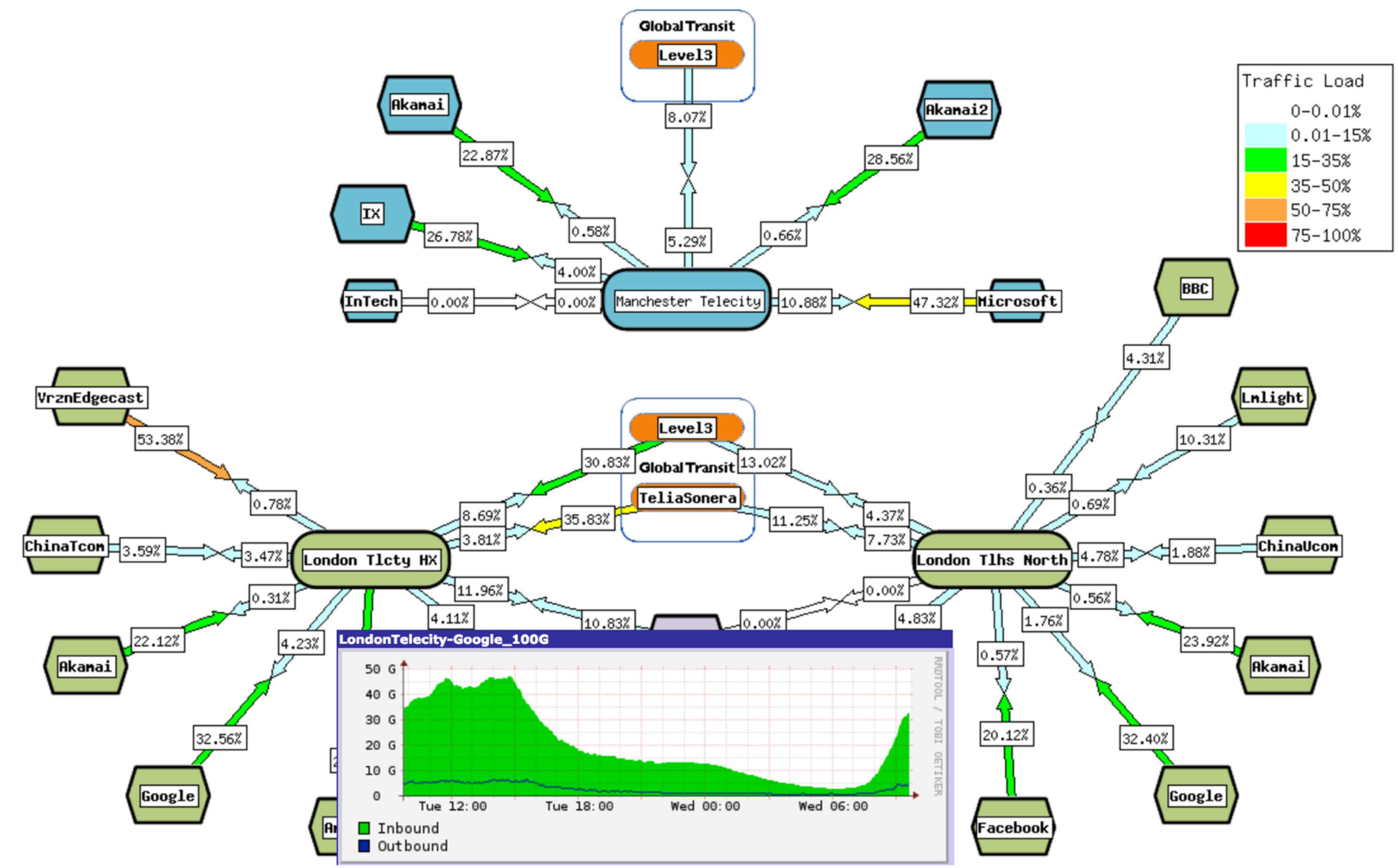
- › The Janet Netpath service exists for dedicated / overlay links
- › In general, Jisc plans regular network upgrades with a view to ensuring that there is sufficient latent capacity in the network

Janet backbone, Jun 2017 (exam period!)



Major external links, Jun 2017

Major External Links



Created: Jun 14 2017 09:40:02

Site network engineering – the Science DMZ

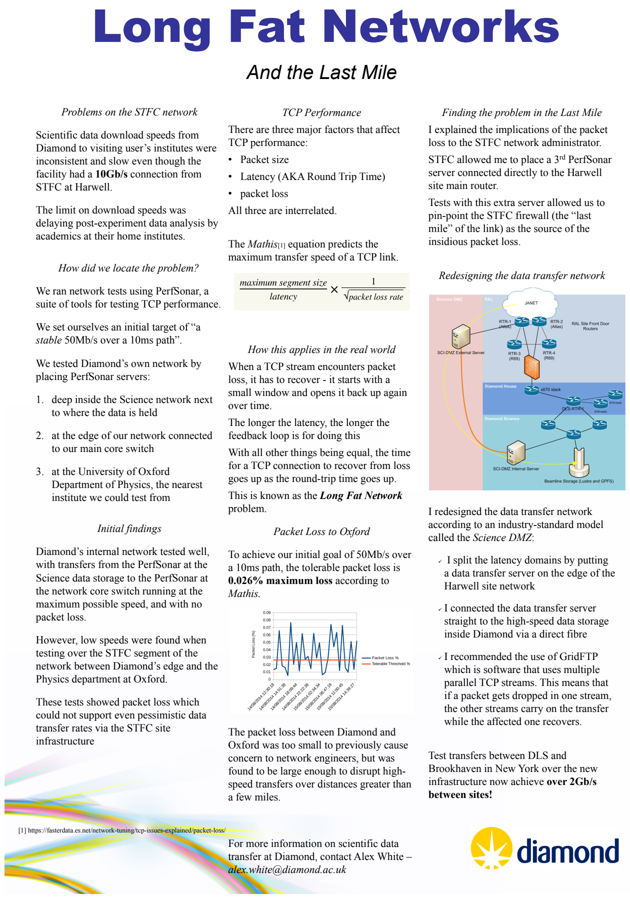
- » ESnet published the Science DMZ 'design pattern' in 2012/13
  - › https://www.es.net/assets/pubs_presos/sc13sciDMZ-final.pdf
- » Three key elements:
  - › Network architecture; avoiding local bottlenecks
  - › Network performance measurement
  - › Data transfer node (DTN) design and configuration
- » Important to apply your security policy without impacting performance
  - › Note some of Eric's comments yesterday on applying stateless filters / ACLs

- » The NSF Cyberinfrastructure (CC*) Program funded this model in over 100 US universities:
  - › See https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504748
- » No current funding equivalent in the UK
  - › But this can and should be part of your network architecture evolution

Good news – Science DMZ happening already

- There are several examples of sites in the UK that have a form of Science DMZ deployment
- In many cases the deployments were made without knowledge of the Science DMZ model!
- Science DMZ is just a set of good principles to follow, so it's not surprising that some Janet sites are already doing it, especially GridPP sites ☺

- Examples in the UK:
  - Diamond Light Source
  - JASMIN/CEDA Data Transfer Zone
  - Imperial College GridPP; supports up to 40Gbit/s of IPv4/IPv6
  - To realise the end-to-end benefit, both ends need to apply the principles


Science DMZ
RAL
JANET
RTR-1 (Atlas)
RTR-2 (Atlas)
RAL Site Front Door Routers
RTR-3 (R89)
RTR-4 (R89)
SCI-DMZ External Server
Diamond House
x670 stack
DLS-RTR-1
G19 stack
G18 stack
Diamond Science
SCI-DMZ Internal Server
Beamline Storage (Lustre and GPFS)
Packet Loss (%)
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
14/08/2014 12:30:18
14/08/2014 14:51:38
14/08/2014 18:09:44
14/08/2014 22:22:38
15/08/2014 02:34:34
15/08/2014 06:47:18
15/08/2014 11:00:45
15/08/2014 14:39:27
Packet Loss %
Tolerable Threshold %
diamond

Long Fat Networks

And the Last Mile

Problems on the STFC network

Scientific data download speeds from Diamond to visiting user's institutes were inconsistent and slow even though the facility had a **10Gb/s** connection from STFC at Harwell.

The limit on download speeds was delaying post-experiment data analysis by academics at their home institutes.

How did we locate the problem?

We ran network tests using PerfSonar, a suite of tools for testing TCP performance.

We set ourselves an initial target of "a *stable* 50Mb/s over a 10ms path".

We tested Diamond's own network by placing PerfSonar servers:

1. deep inside the Science network next to where the data is held
2. at the edge of our network connected to our main core switch
3. at the University of Oxford Department of Physics, the nearest institute we could test from

Initial findings

Diamond's internal network tested well, with transfers from the PerfSonar at the Science data storage to the PerfSonar at the network core switch running at the maximum possible speed, and with no packet loss.

However, low speeds were found when testing over the STFC segment of the network between Diamond's edge and the Physics department at Oxford.

These tests showed packet loss which could not support even pessimistic data transfer rates via the STFC site infrastructure

TCP Performance

There are three major factors that affect TCP performance:

- Packet size
- Latency (AKA Round Trip Time)
- packet loss

All three are interrelated.

The *Mathis*[1] equation predicts the maximum transfer speed of a TCP link.

$$\frac{\textit{maximum segment size}}{\textit{latency}} \times \frac{1}{\sqrt{\textit{packet loss rate}}}$$

How this applies in the real world

When a TCP stream encounters packet loss, it has to recover - it starts with a small window and opens it back up again over time.

The longer the latency, the longer the feedback loop is for doing this

With all other things being equal, the time for a TCP connection to recover from loss goes up as the round-trip time goes up.

This is known as the ***Long Fat Network*** problem.

Packet Loss to Oxford

To achieve our initial goal of 50Mb/s over a 10ms path, the tolerable packet loss is **0.026% maximum loss** according to *Mathis.*

The packet loss between Diamond and Oxford was too small to previously cause concern to network engineers, but was found to be large enough to disrupt high-speed transfers over distances greater than a few miles.

Finding the problem in the Last Mile

I explained the implications of the packet loss to the STFC network administrator.

STFC allowed me to place a 3rd PerfSonar server connected directly to the Harwell site main router.

Tests with this extra server allowed us to pin-point the STFC firewall (the "last mile" of the link) as the source of the insidious packet loss.

Redesigning the data transfer network

I redesigned the data transfer network according to an industry-standard model called the *Science DMZ*:

- ✓ I split the latency domains by putting a data transfer server on the edge of the Harwell site network
- ✓ I connected the data transfer server straight to the high-speed data storage inside Diamond via a direct fibre
- ✓ I recommended the use of GridFTP which is software that uses multiple parallel TCP streams. This means that if a packet gets dropped in one stream, the other streams carry on the transfer while the affected one recovers.

Test transfers between DLS and Brookhaven in New York over the new infrastructure now achieve **over 2Gb/s between sites!**

For more information on scientific data transfer at Diamond, contact Alex White – *alex.white@diamond.ac.uk*

[1] https://fasterdata.es.net/network-tuning/tcp-issues-explained/packet-loss/

Other examples of campus network engineering

- In principle you can just use your Janet IP service for data-intensive science
- Many sites split their connectivity, e.g. 10G campus, 10G GridPP, 20G resilience
  - And then apply Science DMZ principles to the GridPP path
- Where specific point-to-point guarantees are required, there's Janet Netpath Plus
  - See https://www.jisc.ac.uk/netpath for details
  - But then you'll not be able to exceed / burst beyond that capacity

- The WLCG has used physical / virtual overlays
  - LHCOPN (private optical network) / LHCONE (virtual network)
- But how should campuses cater for multiple data-intensive science disciplines?
  - Would one new overlay network per research community scale?

- At least one site is exploring SDN, using Cisco Application Centric Infrastructure (ACI)

Choice of data transfer tool?

- Researchers will find a wide range of data transfer tools available
  - Choice may be dictated by the facility they're accessing, or made by word of mouth
- There's the simpler old friends
  - FTP, scp
  - But these are likely to give a bad initial impression of what your network can do
- There's TCP-based tools designed to mitigate the impact of packet loss
  - GridFTP typically uses four parallel TCP streams
  - Globus Connect has become popular - https://www.globus.org/globus-connect
- There's specific tools to support management of transfers
  - FTS – see http://astro.dur.ac.uk/~dphoelh/documentation/transfer-data-to-ral-v1.4.pdf
- There's also a commercial UDP-based tool, Aspera
  - See http://www.asperasoft.com/
- It would be good to establish more benchmarking of these tools at Janet campuses

E2EPI site visits

- » We (Duncan Rand and I) have visited a dozen or so sites so far
  - › Met with a variety of networking staff and researchers
  - › And spoken to many others via email
- » Really interesting to hear what sites are doing
  - › Some good practice evident, especially in local network engineering
  - › e.g. routing those elephants around main campus firewalls
  - › Campus firewalls typically not designed for single high throughput flows

- » Seeing varying use of site links
  - › As mentioned before, GridPP sites tend to split out their GridPP traffic
  - › Some sites are using their 'resilient' link for bulk data transfers (be careful here!)
- » Some rate limiting of researcher traffic
  - › We'd encourage sites to talk to us (Jisc) about capacity rather than rate limit

Example case: Diamond – University X

- Spoke with Diamond Light Source staff
  - Become evident that some of their university customers are using hard disk transfers
  - Science DMZ deployed at Diamond, but not at many of the customer sites

- Example: University X
  - Running ~6 experiments at DLS per year, producing 10-40TB of data per run
  - Takes 3 weeks from start to finish to get data locally for further processing
  - Remember that moving 100TB requires approx 10Gbps throughput for 24 hours
- Arranged a staged plan with X:
  - Benchmark current performance, with existing 1Gbps connectivity to data store
  - Repeat using Globus tools
  - Upgrade data store node to 10Gbps <- *at this stage now; encouraging results*
  - Then repeat with 10G DTN at campus edge (Science DMZ)
- Has shown the value of getting researchers and network admins sat together ☺

Measuring network characteristics

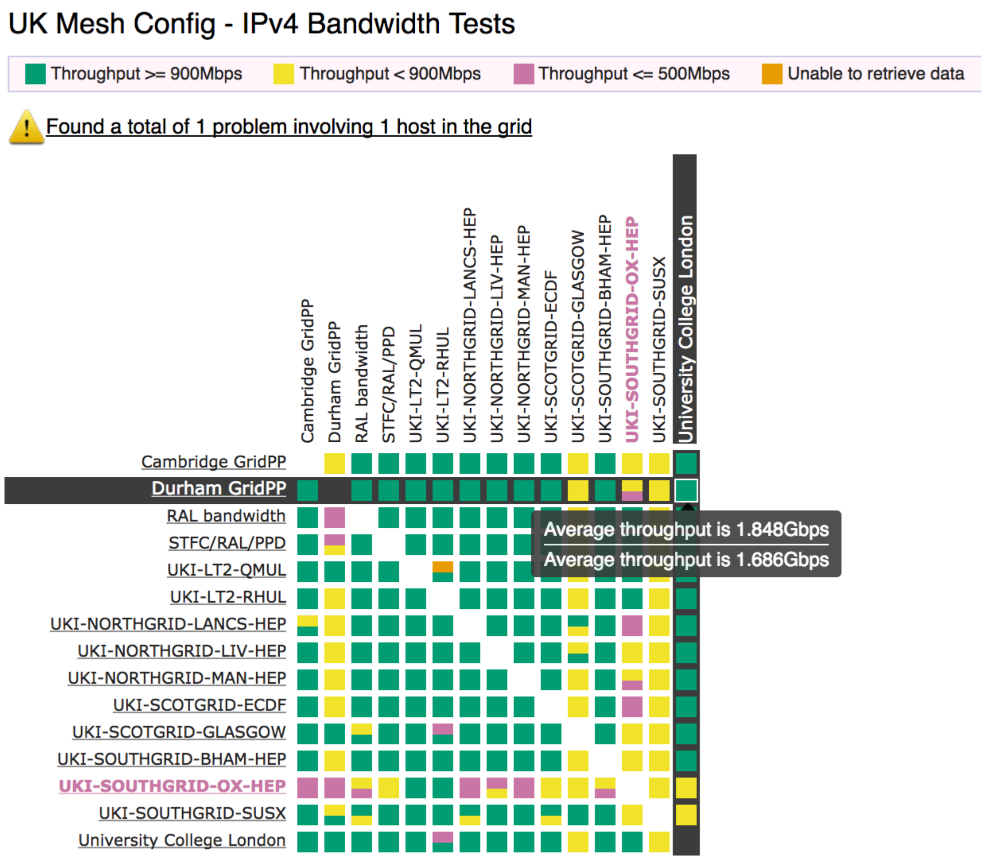
- Important to have telemetry on your network

- The Science DMZ model recommends perfSONAR for this
  - Current version now 4.0; would encourage upgrades; but do check h/w requirements first

- Collects telemetry over time; throughput, loss, latency, path, allows retrospective viewing of data
  - Uses proven tools under the hood such as iperf
- Can run between two perfSONAR systems or build a mesh

- Helps you assess the impact of changes to your network or systems
- It can highlight poor performance, but doesn't troubleshoot per se
- See our 'case for perfSONAR deployment' document (again, feedback welcomed!)
- https://community.jisc.ac.uk/system/files/1862/The%20case%20for%20perfSONAR%20deployment-v1-0.pdf

Example: UK GridPP perfSONAR mesh



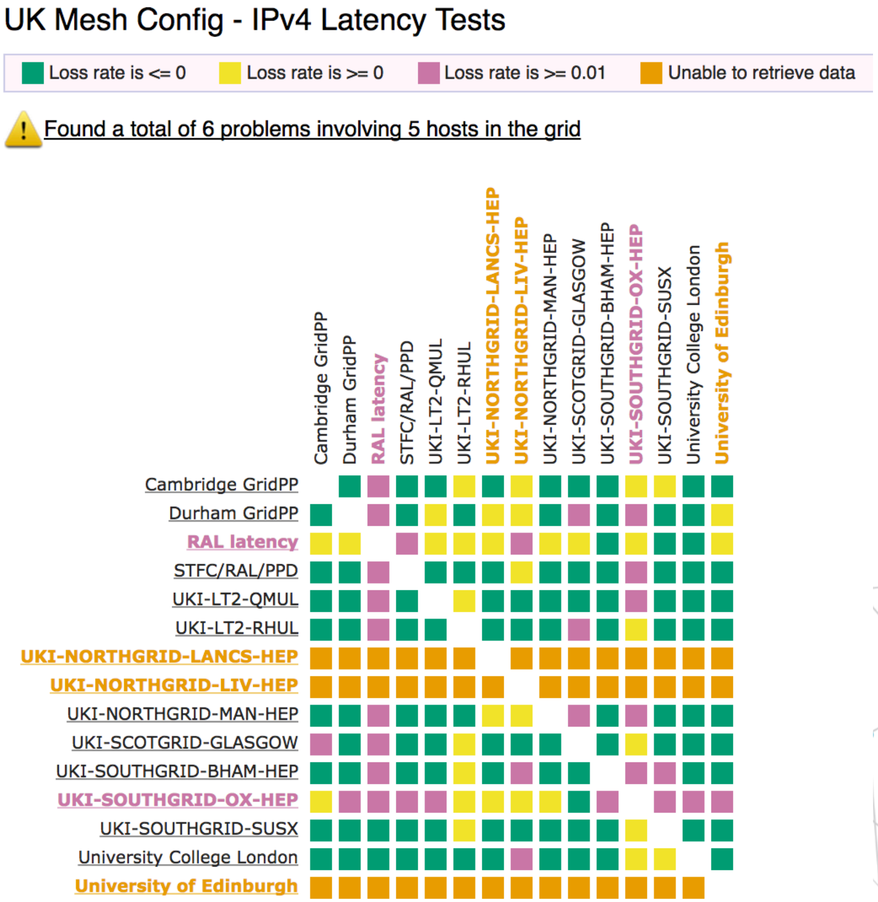
UK Mesh Config - IPv4 Bandwidth Tests

🟩 Throughput >= 900Mbps 🟨 Throughput < 900Mbps 🟪 Throughput <= 500Mbps 🟧 Unable to retrieve data

⚠ Found a total of 1 problem involving 1 host in the grid

	Cambridge GridPP	Durham GridPP	RAL bandwidth	STFC/RAL/PPD	UKI-LT2-QMUL	UKI-LT2-RHUL	UKI-NORTHGRID-LANCS-HEP	UKI-NORTHGRID-LIV-HEP	UKI-NORTHGRID-MAN-HEP	UKI-SCOTGRID-ECDF	UKI-SCOTGRID-GLASGOW	UKI-SOUTHGRID-BHAM-HEP	UKI-SOUTHGRID-OX-HEP	UKI-SOUTHGRID-SUSX	University College London
Cambridge GridPP		🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟨	🟩	🟨	🟨	🟩
Durham GridPP	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟨	🟩	🟨🟪	🟨	🟩
RAL bandwidth	🟩	🟪		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟨	🟩	🟩	🟨	🟩
STFC/RAL/PPD	🟩	🟪🟨	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟨	🟩	🟩	🟨	🟩
UKI-LT2-QMUL	🟩	🟩	🟩	🟩		🟧🟩	🟩	🟩	🟩	🟩	🟨	🟩	🟩	🟨	🟩
UKI-LT2-RHUL	🟩	🟨	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟨	🟩	🟩	🟨	🟩
UKI-NORTHGRID-LANCS-HEP	🟨🟩	🟨	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩🟨	🟩	🟪	🟨	🟩
UKI-NORTHGRID-LIV-HEP	🟩	🟨	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟨🟩	🟩	🟨	🟨	🟩
UKI-NORTHGRID-MAN-HEP	🟩	🟨	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟨	🟩	🟨🟪	🟨	🟩
UKI-SCOTGRID-ECDF	🟩	🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟨	🟩	🟪	🟨	🟩
UKI-SCOTGRID-GLASGOW	🟩	🟩	🟨🟩	🟩	🟩	🟪🟩	🟩	🟩	🟩	🟩		🟩	🟨	🟨	🟩
UKI-SOUTHGRID-BHAM-HEP	🟩	🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟨		🟨	🟨	🟩
UKI-SOUTHGRID-OX-HEP	🟪	🟪	🟨🟪	🟨	🟩	🟩	🟪	🟪🟨	🟪	🟨	🟨	🟨🟪		🟨	🟨
UKI-SOUTHGRID-SUSX	🟩	🟨🟩	🟩🟨	🟩	🟩	🟩	🟩🟨	🟩	🟩	🟩🟨	🟩	🟩	🟨		🟨
University College London	🟩	🟩	🟩	🟩	🟩	🟪🟩	🟩	🟩	🟩	🟩	🟨	🟩	🟩	🟨	

Average throughput is 1.848Gbps
Average throughput is 1.686Gbps



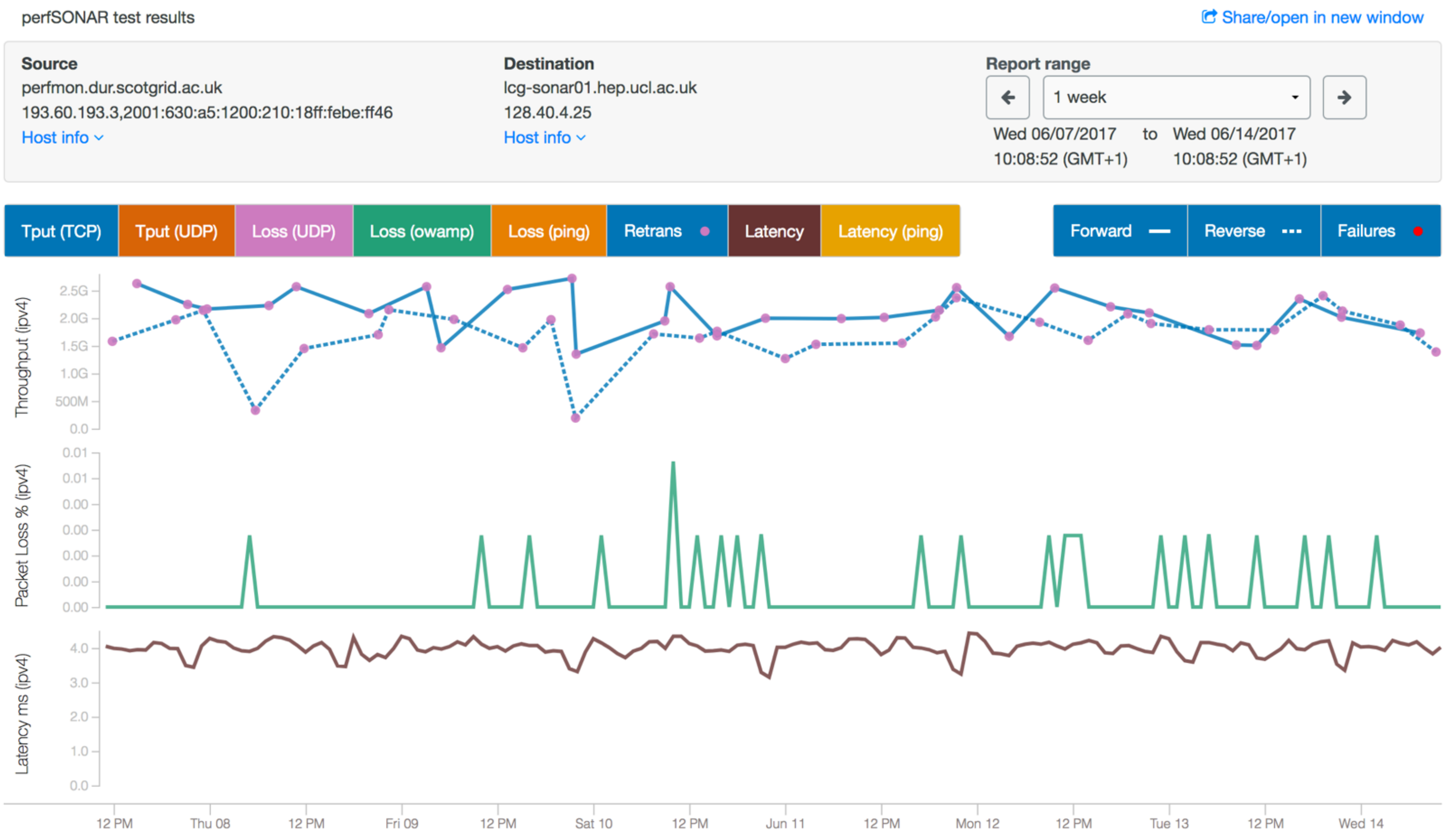
UK Mesh Config - IPv4 Latency Tests

🟩 Loss rate is <= 0 🟨 Loss rate is >= 0 🟪 Loss rate is >= 0.01 🟧 Unable to retrieve data

⚠ Found a total of 6 problems involving 5 hosts in the grid

	Cambridge GridPP	Durham GridPP	RAL latency	STFC/RAL/PPD	UKI-LT2-QMUL	UKI-LT2-RHUL	UKI-NORTHGRID-LANCS-HEP	UKI-NORTHGRID-LIV-HEP	UKI-NORTHGRID-MAN-HEP	UKI-SCOTGRID-GLASGOW	UKI-SOUTHGRID-BHAM-HEP	UKI-SOUTHGRID-OX-HEP	UKI-SOUTHGRID-SUSX	University College London	University of Edinburgh
Cambridge GridPP		🟩	🟪	🟩	🟩	🟨	🟩	🟨	🟩	🟩	🟩	🟨	🟨	🟩	🟩
Durham GridPP	🟩		🟪	🟩	🟨	🟩	🟨	🟨	🟩	🟪	🟩	🟪	🟩	🟩	🟨
RAL latency	🟨	🟨		🟪	🟨	🟨	🟨	🟪	🟨	🟨	🟩	🟨	🟩	🟩	🟨
STFC/RAL/PPD	🟩	🟩	🟪		🟩	🟩	🟩	🟨	🟩	🟩	🟩	🟪	🟩	🟩	🟩
UKI-LT2-QMUL	🟩	🟩	🟪	🟩		🟨	🟩	🟩	🟩	🟩	🟩	🟪	🟩	🟩	🟩
UKI-LT2-RHUL	🟩	🟩	🟪	🟩	🟩		🟩	🟩	🟩	🟪	🟩	🟨	🟩	🟩	🟩
UKI-NORTHGRID-LANCS-HEP	🟧	🟧	🟧	🟧	🟧	🟧		🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧
UKI-NORTHGRID-LIV-HEP	🟧	🟧	🟧	🟧	🟧	🟧	🟧		🟧	🟧	🟧	🟧	🟧	🟧	🟧
UKI-NORTHGRID-MAN-HEP	🟩	🟩	🟪	🟩	🟩	🟩	🟨	🟨		🟪	🟩	🟪	🟩	🟩	🟩
UKI-SCOTGRID-GLASGOW	🟪	🟩	🟪	🟩	🟩	🟨	🟩	🟩	🟩		🟩	🟨	🟩	🟩	🟩
UKI-SOUTHGRID-BHAM-HEP	🟩	🟩	🟪	🟩	🟩	🟨	🟩	🟪	🟩	🟩		🟪	🟪	🟩	🟩
UKI-SOUTHGRID-OX-HEP	🟨	🟪	🟪	🟪	🟪	🟨	🟨	🟨	🟨	🟩	🟪		🟪	🟪	🟪
UKI-SOUTHGRID-SUSX	🟩	🟩	🟩	🟩	🟩	🟨	🟩	🟩	🟩	🟩	🟩	🟨		🟩	🟩
University College London	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟪	🟩	🟩	🟩	🟨	🟨		🟩
University of Edinburgh	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	

New perfSONAR 4.0 visualisation



perfSONAR test results

Share/open in new window

Source
perfmon.dur.scotgrid.ac.uk
193.60.193.3,2001:630:a5:1200:210:18ff:febe:ff46
Host info

Destination
lcg-sonar01.hep.ucl.ac.uk
128.40.4.25
Host info

Report range
1 week
Wed 06/07/2017 10:08:52 (GMT+1) to Wed 06/14/2017 10:08:52 (GMT+1)

Janet perfSONAR / DTN test node(s)

- We've installed a 10G perfSONAR node at a Janet PoP in London
  - Lets you test your site's throughput to/from Janet backbone
  - Useful if you want to run data-intensive applications in the near future, but don't have perfSONAR at the far end, or to benchmark your site's connectivity
  - Ask us if you're interested in using it

- We're planning to add a second perfSONAR test node in our Slough DC
  - Also planning to install a 10G reference DTN there
  - Will be a FIONA appliance - see https://fasterdata.es.net/science-dmz/DTN/
  - Will allow disk-to-disk tests, using a variety of transfer tools

- We can also run a perfSONAR mesh for you, using MaDDash on a VM
- We may also deploy an experimental perfSONAR node for TCP-BBR tests

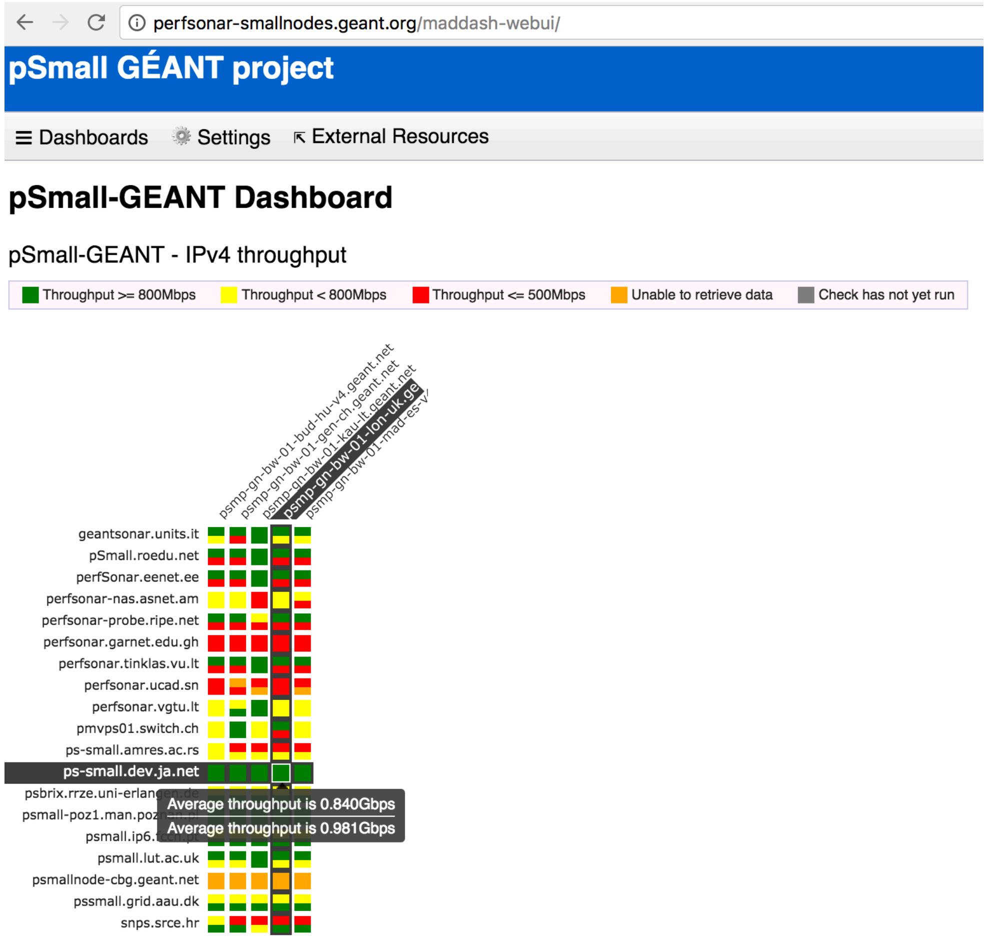
Small node perfSONAR

- In some cases, just an indicative perfSONAR test is useful
- i.e., run loss/latency tests as normal, but limit throughout tests to 1Gbit/s
- For this scenario, you can build a small node perfSONAR system for under £250
- Jisc took part in the GEANT small node pilot project, using Gigabyte Brix:
  - IPv4 and IPv6 test mesh at http://perfsonar-smallnodes.geant.org/maddash-webui/

- We now have a device build that we can offer to communities for testing
  - Aim to make them as 'plug and play' as possible
  - And a stepping stone to a full perfSONAR node

- Further information and TNC2016 meeting slide deck:
  - https://lists.geant.org/sympa/d_read/perfsonar-smallnodes/

perfSONAR small node test mesh



perfsonar-smallnodes.geant.org/maddash-webui/

pSmall GÉANT project

☰ Dashboards   Settings   ⇱ External Resources

pSmall-GEANT Dashboard

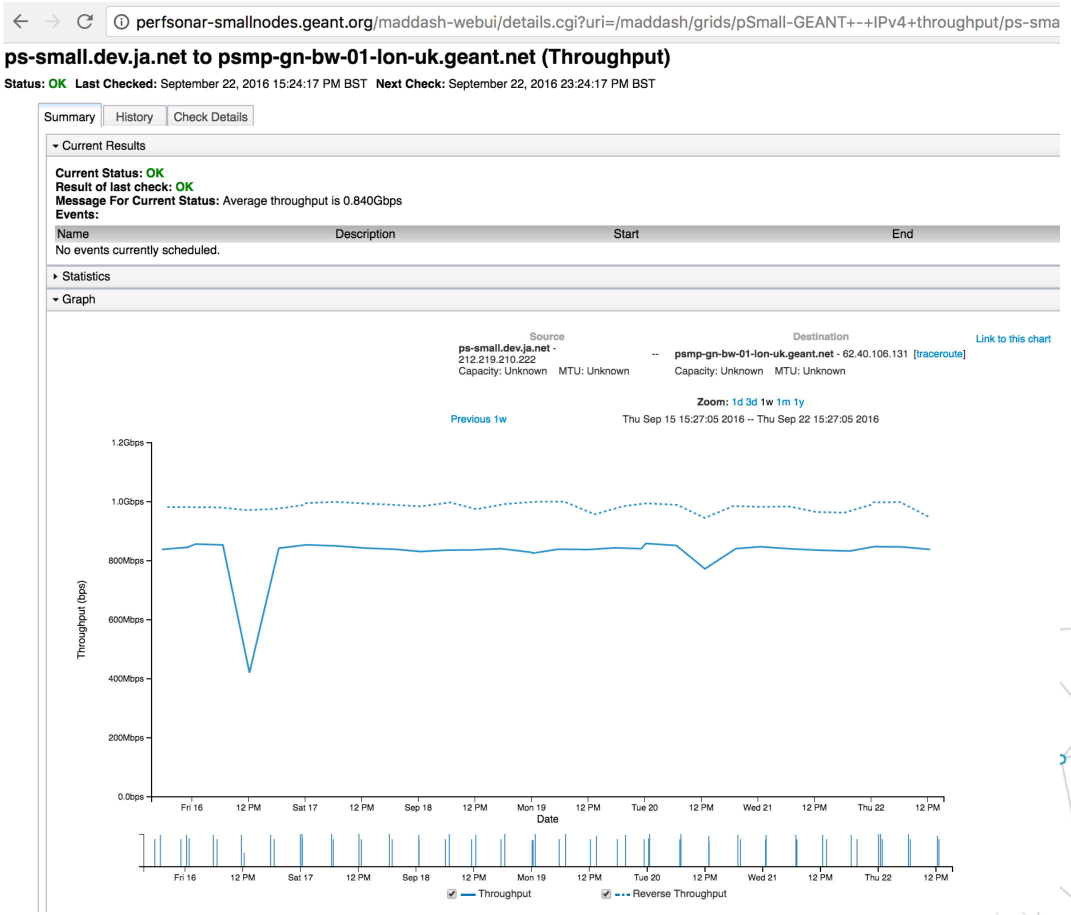
pSmall-GEANT - IPv4 throughput

🟩 Throughput >= 800Mbps   🟨 Throughput < 800Mbps   🟥 Throughput <= 500Mbps   🟧 Unable to retrieve data   ⬛ Check has not yet run

psmp-gn-bw-01-bud-hu-v4.geant.net
psmp-gn-bw-01-gen-ch.geant.net
psmp-gn-bw-01-kau-lt.geant.net
psmp-gn-bw-01-lon-uk.ge
psmp-gn-bw-01-mad-es-v

geantsonar.units.it
pSmall.roedu.net
perfSonar.eenet.ee
perfsonar-nas.asnet.am
perfsonar-probe.ripe.net
perfsonar.garnet.edu.gh
perfsonar.tinklas.vu.lt
perfsonar.ucad.sn
perfsonar.vgtu.lt
pmvps01.switch.ch
ps-small.amres.ac.rs
ps-small.dev.ja.net
psbrix.rrze.uni-erlan
psmall-poz1.man.poz
psmall.ip6.
psmall.lut.ac.uk
psmallnode-cbg.geant.net
pssmall.grid.aau.dk
snps.srce.hr

Average throughput is 0.840Gbps
Average throughput is 0.981Gbps



perfsonar-smallnodes.geant.org/maddash-webui/details.cgi?uri=/maddash/grids/pSmall-GEANT+-+IPv4+throughput/ps-sma

ps-small.dev.ja.net to psmp-gn-bw-01-lon-uk.geant.net (Throughput)

Status: OK **Last Checked:** September 22, 2016 15:24:17 PM BST **Next Check:** September 22, 2016 23:24:17 PM BST

Summary | History | Check Details

▾ Current Results

Current Status: OK
Result of last check: OK
Message For Current Status: Average throughput is 0.840Gbps
Events:

Name	Description	Start	End
No events currently scheduled.			

▸ Statistics

▾ Graph

Source: **ps-small.dev.ja.net** - 212.219.210.222 Capacity: Unknown MTU: Unknown -- Destination: **psmp-gn-bw-01-lon-uk.geant.net** - 62.40.106.131 [traceroute] Capacity: Unknown MTU: Unknown

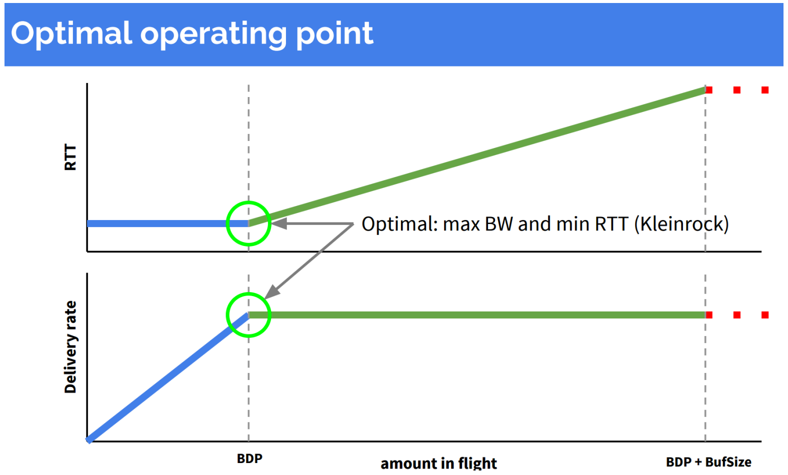
Link to this chart

Zoom: 1d 3d 1w 1m 1y

Previous 1w   Thu Sep 15 15:27:05 2016 -- Thu Sep 22 15:27:05 2016

Aside: Google's TCP-BBR

- » Traditional TCP performs poorly even with just a fraction of 1% loss rate
- » Google have been developing a new version of TCP
  - › TCP-BBR was open-sourced last year
  - › Requires just sender-side deployment
  - › Seeks high throughput with a small queue
  - › Good performance at up to 15% loss
  - › Google using it in production today
- » Would be good to explore this further
  - › Understand impact on other TCP variants
  - › And when used for parallelised TCP applications like GridFTP
- » See the presentation from the March 2017 IETF meeting:
  - › https://www.ietf.org/proceedings/98/slides/slides-98-iccrg-an-update-on-bbr-congestion-control-00.pdf


Optimal operating point

E2E performance to cloud compute

- We're seeing interest in the use of commercial cloud compute
  - e.g. to provide remote CPU for scientific equipment, or for cloudbursting
- Complements compute available at ESPRC Tier-2 HPC facilities
  - https://www.epsrc.ac.uk/research/facilities/hpc/tier2/
- Anecdotal reports of 2-3Gbit/s into AWS
  - e.g. by researchers at the Institute of Cancer Research
  - See presentations at RCUK Cloud Workshop - https://cloud.ac.uk/
- Bandwidth fo AWS depends on the VM size
  - See https://aws.amazon.com/ec2/instance-types/

- We're exploring measurement of cloud compute connectivity performance further
  - Includes AWS, MS ExpressRoute, GCP, …
  - And scaling Janet connectivity to these services as appropriate

Building on Science DMZ?

- We should seek to establish good principles and practices at all campuses
  - And the research organisations they work with
  - There's already a good foundation at many GridPP sites
- The Janet backbone is heading towards 1Tbps capacity and beyond
- Jisc can help to seed further communities of good practice on this foundation
  - e.g. DiRAC HPC community, SES consortium, …
- And grow a Research Data Transfer Zone (RDTZ) within and between campuses
  - Towards a UK RDTZ
  - Inspired by the US Pacific Research Platform (PRP) model of multi-site, multi-discipline research-driven collaboration built on the NSF's Science DMZ investment
- Many potential benefits, such as enabling new types of workflow
  - e.g. streaming data to CPUs without the need to store locally
  - And hopefully no more hard disk transfers!

Useful links

- Janet E2EPI project page
  - https://www.jisc.ac.uk/rd/projects/janet-end-to-end-performance-initiative
- E2EPI Jisc community page
  - https://community.jisc.ac.uk/groups/janet-end-end-performance-initiative
- JiscMail E2EPI list (approx 100 subscribers)
  - https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=E2EPI
- Camus Network Engineering for Data-Intensive Science workshop slides
  - https://www.jisc.ac.uk/events/campus-network-engineering-for-data-intensive-science-workshop-19-oct-2016
- Fasterdata knowledge base
  - http://fasterdata.es.net/
- eduPERT knowledge base
  - http://kb.pert.geant.net/PERTKB/WebHome

Please feel free to get in touch!

Dr Tim Chown

Network Development Manager

tim.chown@jisc.ac.uk

jisc.ac.uk