10/09/2015

Janet end-to-end performance initiative

Dr Tim Chown, Jisc

What's Jisc's interest in GridPP in the e2epi context?

- With the recent merger, Jisc now operates the Janet network. Janet(UK) is no more.
- Jisc is therefore very interested in the applications that its users/customers wish to run
- Increasingly these applications are putting greater demands on the network
  - Not least from the GridPP community ☺
- So we're very keen to hear what your applications require of the network, and what you'd like to do
  - There's clearly already good work being done
  - But there's also more we could do together
  - (while also not exposing scientists to unnecessary detail…)

What trends are placing demands on our networks?

- General growth of HE/FE user traffic
- Use of off-site data centres, including new Jisc shared DC
- Increased outsourcing of systems/services to the cloud
- Communities (such as those here today) wanting to run increasingly demanding network applications
- New communities wanting to do the same (as we heard yesterday afternoon)
- New requirements on universities, e.g. archival of research data as required by funding bodies
- A rise in trans-national education
- …

Aims of the Janet e2epi project

- Help sites get the best from their Janet IP connection
- Identify existing and emerging user communities seeking to run high performance networked applications (be that high throughput, low latency, or…)
- Determine and share best practices
- Foster discussion between research communities, network operators (inc. Janet & the campus end sites) and network researchers
- Help set expectations, and raise awareness of issues
- Raise 'high water marks'
- Offer specialist support & tools (2 FTEs being recruited)

Who are the communities?

- These include:
  - Astrophysics, cosmology
  - Particle Physics
  - Human Genome Project
  - Environmental Science
  - Oceanography
- Often, participant sites are quite diversely spread
- Applications from these communities typically may have data rates through Janet of around 1-6Gbit/s
  - We heard yesterday Lydia Heck was getting 3-4Gbit/s

Articulating current/future requirements?

- Understanding user community requirements will help us get results today, and inform future Janet planning
- Some communities better at articulating this than others
  - Familiar with their data, not with the network
    - "I need to copy this 100TB data set to RAL within 7 days"
  - Many examples given at this meeting of sizes of data sets, but not of specific network requirements
- A good example: LHC Network Future Look, 2014
  - Tier-1 likely to rise to ~20Gbit/s by 2016
  - Tier-2 likely to need at least 10Gbit/s
- Does your site have an infrastructure 'future look' plan?

What types of performance bottlenecks might we see?

- There will inevitably be a limiting factor somewhere in any given application's ability to transfer data, e.g.:
- Application protocols
  - TCP vs UDP, support for parallelisation, …
- End systems
  - TCP buffers, disk I/O, etc
- Local network architectures
  - Local LANs, Ethernet capacities, firewalls, …
- Wide-area links and connectivity
  - Links to and across Janet, and potentially beyond

The potential bottlenecks imply areas for action, e.g.:

- Tuning end systems
  - Classic TCP tuning, disk system tuning, …
  - Some very good guidance at https://fasterdata.es.net/
- Reviewing the tools being used for transfers
  - GridFTP, BBCP, FDT, HPN ssh/scp, etc
- Reviewing internal site network architectures
  - What links/devices might hinder network flows?
- Janet connectivity
  - Site access links – typically up to 10Gbit/s for HE
  - Point to point 'Lightpath' links, where used
  - Janet's peerings to other networks

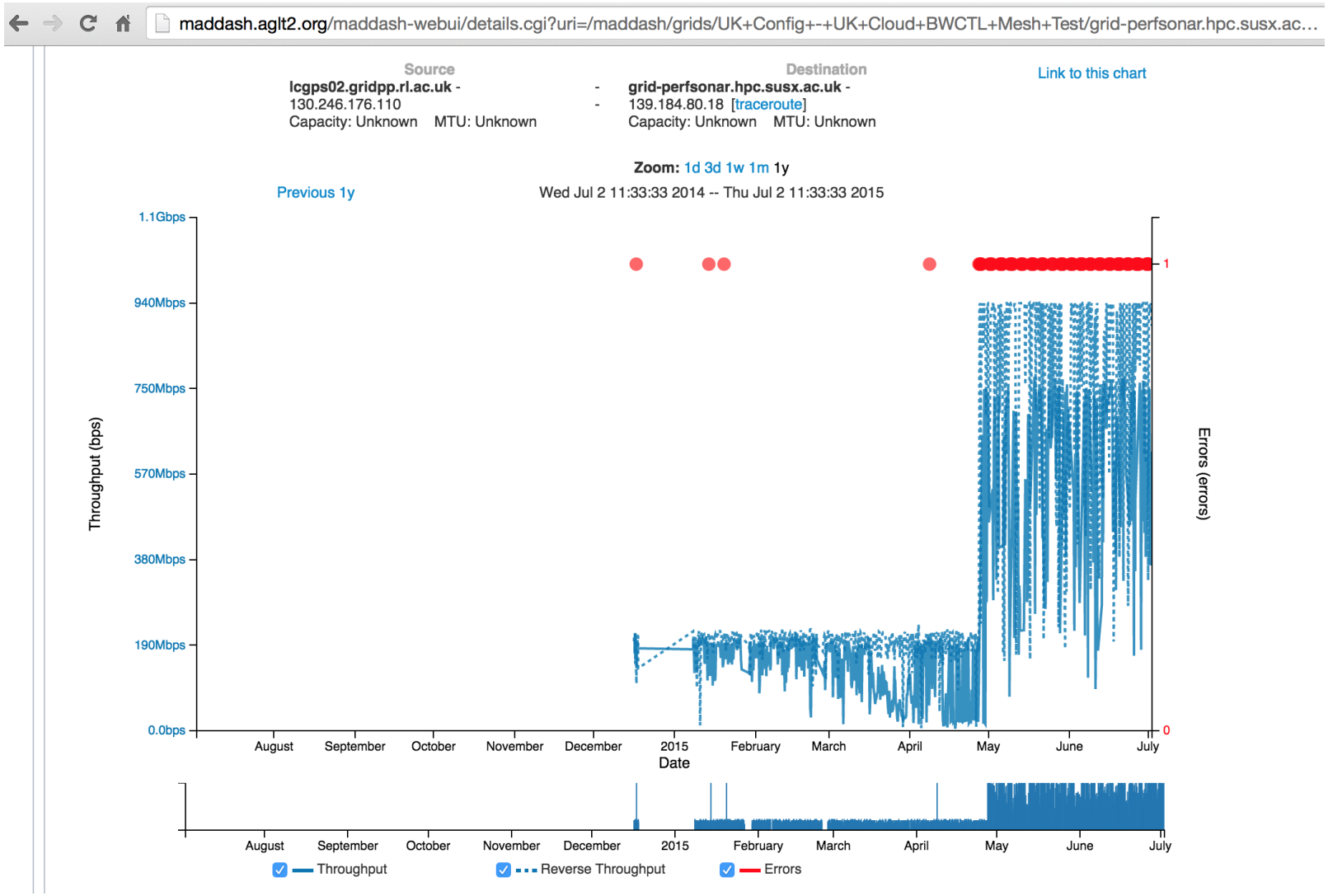
Example site issue - Firewall impact

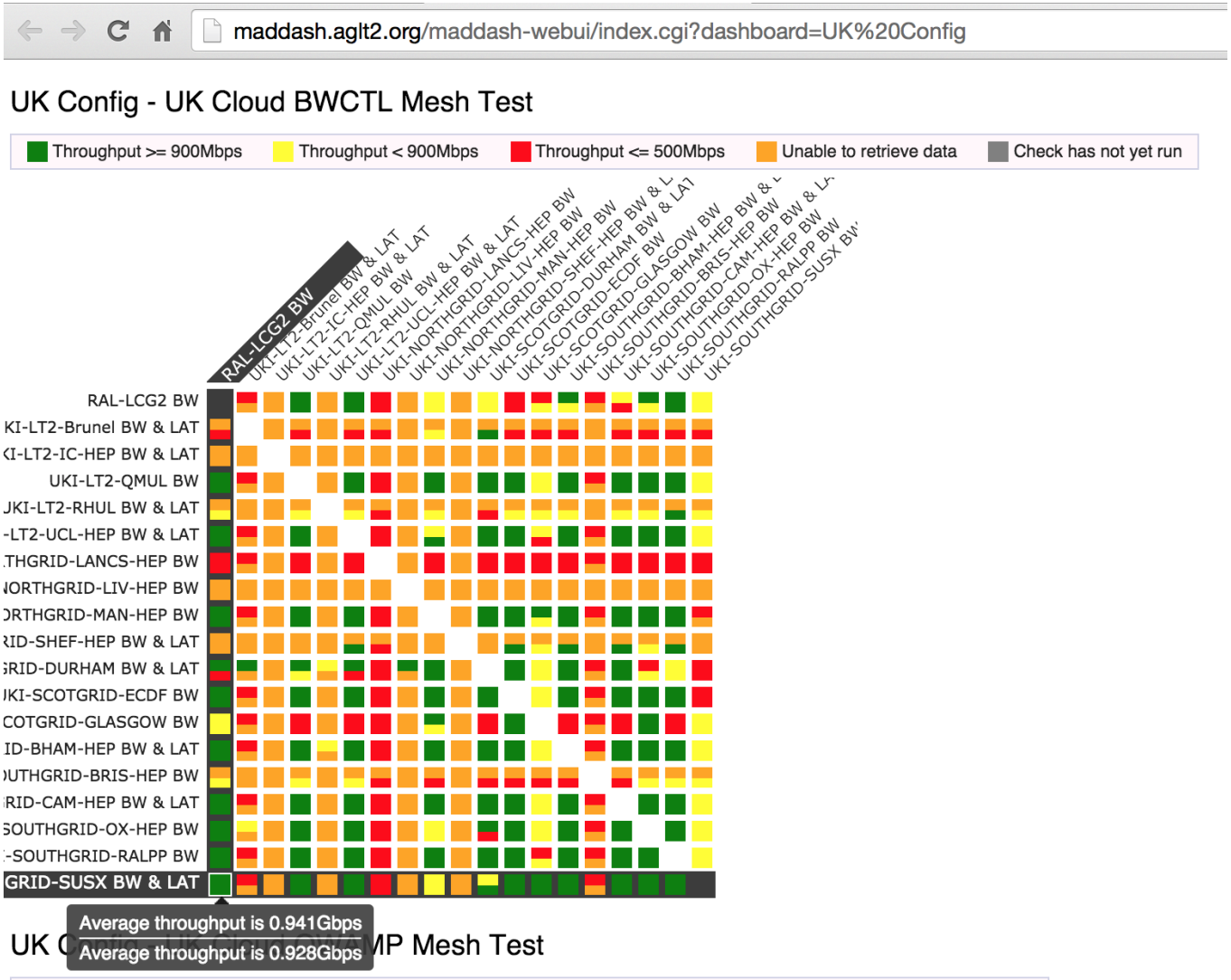
- Some firewall architectures are not well-suited to high throughput flows, or may be applying IDS (intrusion detection) to those flows
- This can often place a cap on performance, and limit an application's performance significantly
- The result is that some sites are now bypassing their firewalls, or requiring significant upgrades to them
- Recent examples:
  - Sussex (more in a moment)
  - Durham (as we heard yesterday from Lydia Heck)

A quick, specific firewall performance example

- Jeremy Maris showed me this example at a recent HPC-SIG meeting
- The first slide shows the impact of removing the firewall from the path
  - In this case it was a BSD system running pf, but similar issues can and do happen with commercial systems
- The next slide shows a perfSONAR matrix which shows the improved performance for Sussex
  - perfSONAR is a commonly used tool in a number of communities, for measuring either loss & latency, or measuring throughput

Janet e2epi project




maddash.aglt2.org/maddash-webui/index.cgi?dashboard=UK%20Config
UK Config - UK Cloud BWCTL Mesh Test
Throughput >= 900Mbps
Throughput < 900Mbps
Throughput <= 500Mbps
Unable to retrieve data
Check has not yet run
RAL-LCG2 BW
UKI-LT2-Brunel BW & LAT
UKI-LT2-IC-HEP BW & LAT
UKI-LT2-QMUL BW
UKI-LT2-RHUL BW & LAT
UKI-LT2-UCL-HEP BW & LAT
UKI-NORTHGRID-LANCS-HEP BW
UKI-NORTHGRID-LIV-HEP BW
UKI-NORTHGRID-MAN-HEP BW
UKI-NORTHGRID-SHEF-HEP BW & L
UKI-SCOTGRID-DURHAM BW & LAT
UKI-SCOTGRID-ECDF BW
UKI-SCOTGRID-GLASGOW BW
UKI-SOUTHGRID-BHAM-HEP BW & L
UKI-SOUTHGRID-BRIS-HEP BW
UKI-SOUTHGRID-CAM-HEP BW & L
UKI-SOUTHGRID-OX-HEP BW
UKI-SOUTHGRID-RALPP BW
UKI-SOUTHGRID-SUSX BW
RAL-LCG2 BW
KI-LT2-Brunel BW & LAT
KI-LT2-IC-HEP BW & LAT
UKI-LT2-QMUL BW
JKI-LT2-RHUL BW & LAT
-LT2-UCL-HEP BW & LAT
THGRID-LANCS-HEP BW
NORTHGRID-LIV-HEP BW
ORTHGRID-MAN-HEP BW
RID-SHEF-HEP BW & LAT
GRID-DURHAM BW & LAT
KI-SCOTGRID-ECDF BW
COTGRID-GLASGOW BW
ID-BHAM-HEP BW & LAT
UTHGRID-BRIS-HEP BW
RID-CAM-HEP BW & LAT
SOUTHGRID-OX-HEP BW
-SOUTHGRID-RALPP BW
GRID-SUSX BW & LAT
Average throughput is 0.941Gbps
Average throughput is 0.928Gbps
MP Mesh Test

Engineering your local site network?

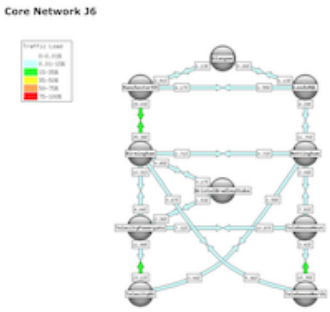
- The 'firewall bypass' is something of a quick and dirty workaround. A more strategic solution is desirable.
- The ESnet approach in the US is to promote a 'Science DMZ', where sites alter their architectures to provide a high performance 'onramp'
  - See https://fasterdata.es.net/science-dmz/
  - Architecture designed for high performance
  - Distinct science network, don't upgrade whole site
  - Appropriate security mechanisms
- Is this something we should promote for Janet sites?

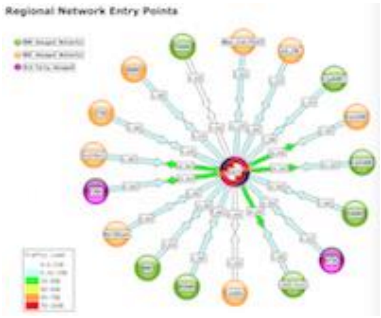
Network monitoring

- How do we see how well applications are using the network?
- What are appropriate tools to measure performance?
  - Preferably as close to the application end points as possible
  - (or can be built in to the protocol, e.g. RCTP for RTP)
- As mentioned earlier, perfSONAR is widely used
  - Includes some 'standard' tools, like owamp, bwctl, iperf
- Per-flow stats are not currently available on the Janet core
- But what other tools might we use? What would help you?
  - Some interesting examples
  - e.g. RIPE Atlas Project - https://atlas.ripe.net/

Janet network usage visualisation?

- Janet has Netsight - http://netsight.ja.net/
  - So has data on link capacities and link usage
- Question - how to make such views available to users?
  - Some 'weather maps' could help in expectation management, and also understanding ongoing issues
  - Possible concerns in making full weather maps public
  - Could we create tailored, per-community views?





How can you get involved?

- Share your experiences – what works, what doesn't – your open challenges – case studies are very welcome
- Join the e2epi Jiscmail list:
  - www.jiscmail.ac.uk/cgi-bin/webadmin?A0=E2EPI
- Come to the free Janet e2epi workshop
  - London, Oct 19th, registrations are open
  - https://www.jisc.ac.uk/events/janet-end-to-end-performance-initiative-workshop-19-oct-2015
  - Help us shape our priorities & strategy for e2epi
- We're also very keen to visit and talk to you!
- Email me – tim.chown@jisc.ac.uk

Find out more…

Contact:

Dr Tim Chown
Senior Network Services Developer

Tim.Chown@jisc.ac.uk
jisc.ac.uk