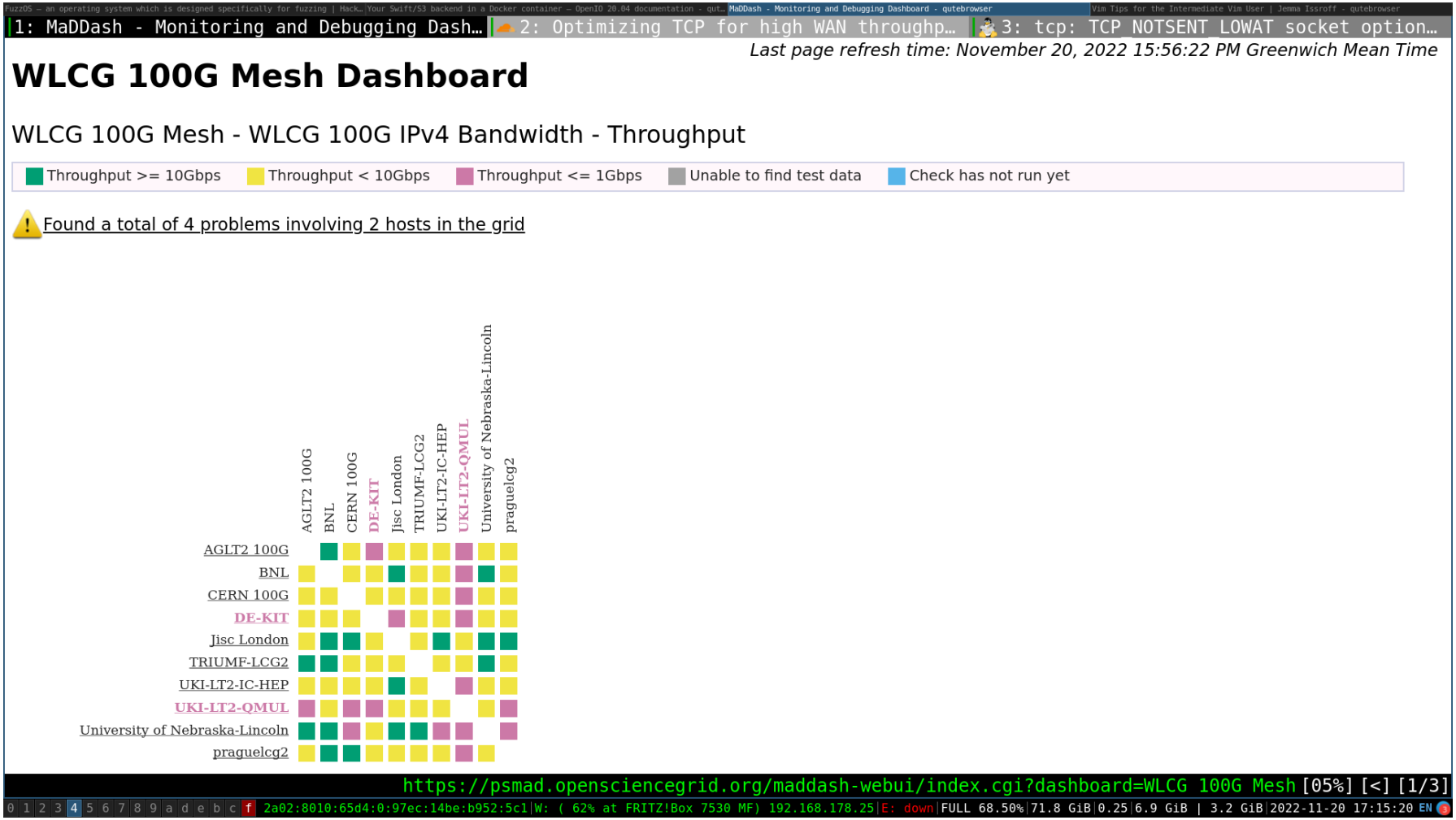
Odd and nice things in perfSONAR land

Raul H C Lopes (Jisc)
Raul.Lopes@jisc.ac.uk

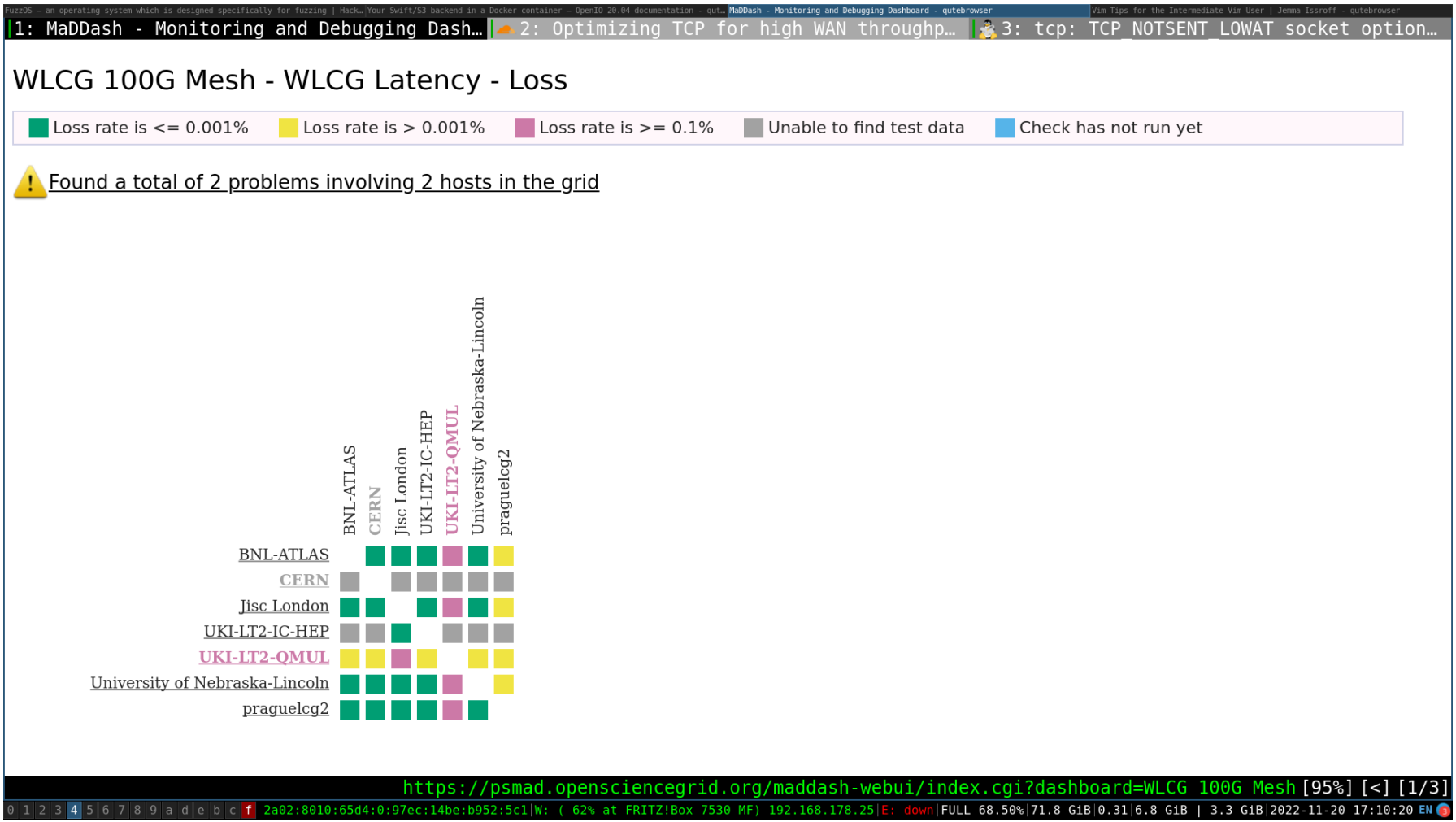
Facts

- Jisc London was added to the WLCG dashboard in July.
- It shows green with 5 out of 9 peers.
- We've seen at least one serious glitch.
- We see several permanent issues in the dashboard related to
  - latency
  - plotting
  - throughput
- I would like to talk about issues and further use of perfSONAR. I have lots of questions. Sorry!

IPv4 throughput on the WLCG 100G dashboard

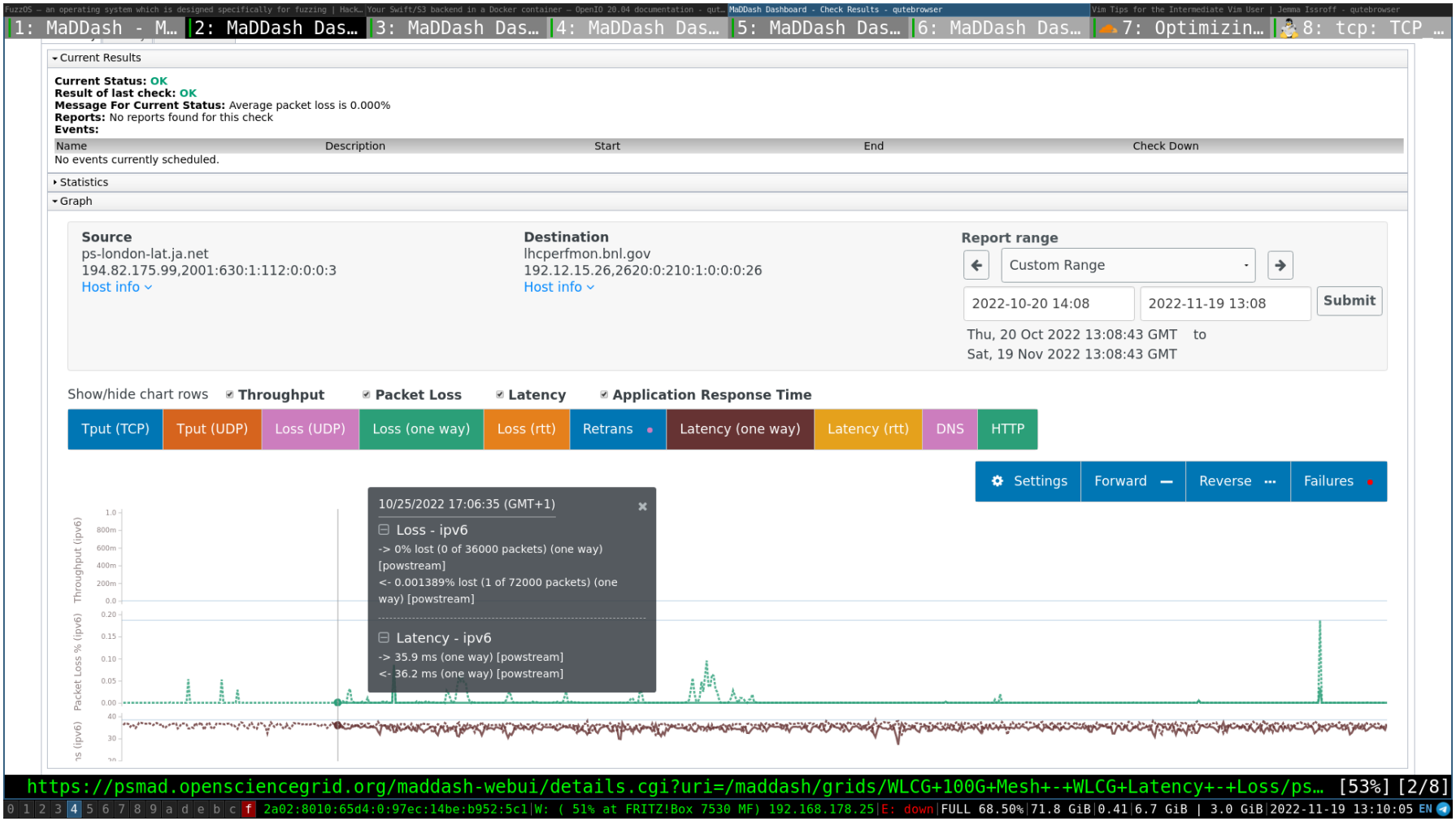


CERN and Imperial do not run latency

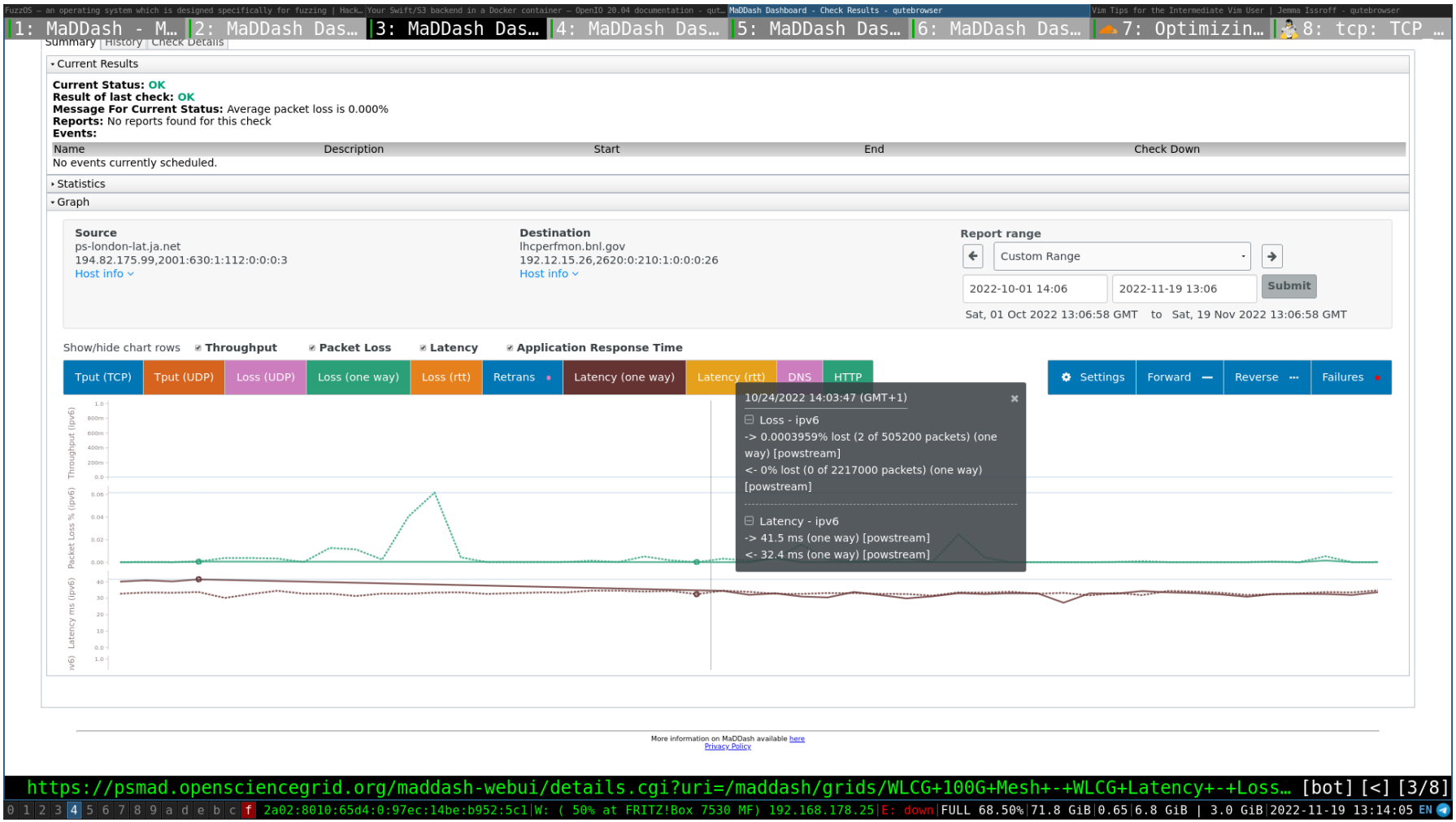


Jisc London went dark from 18 to 25 October

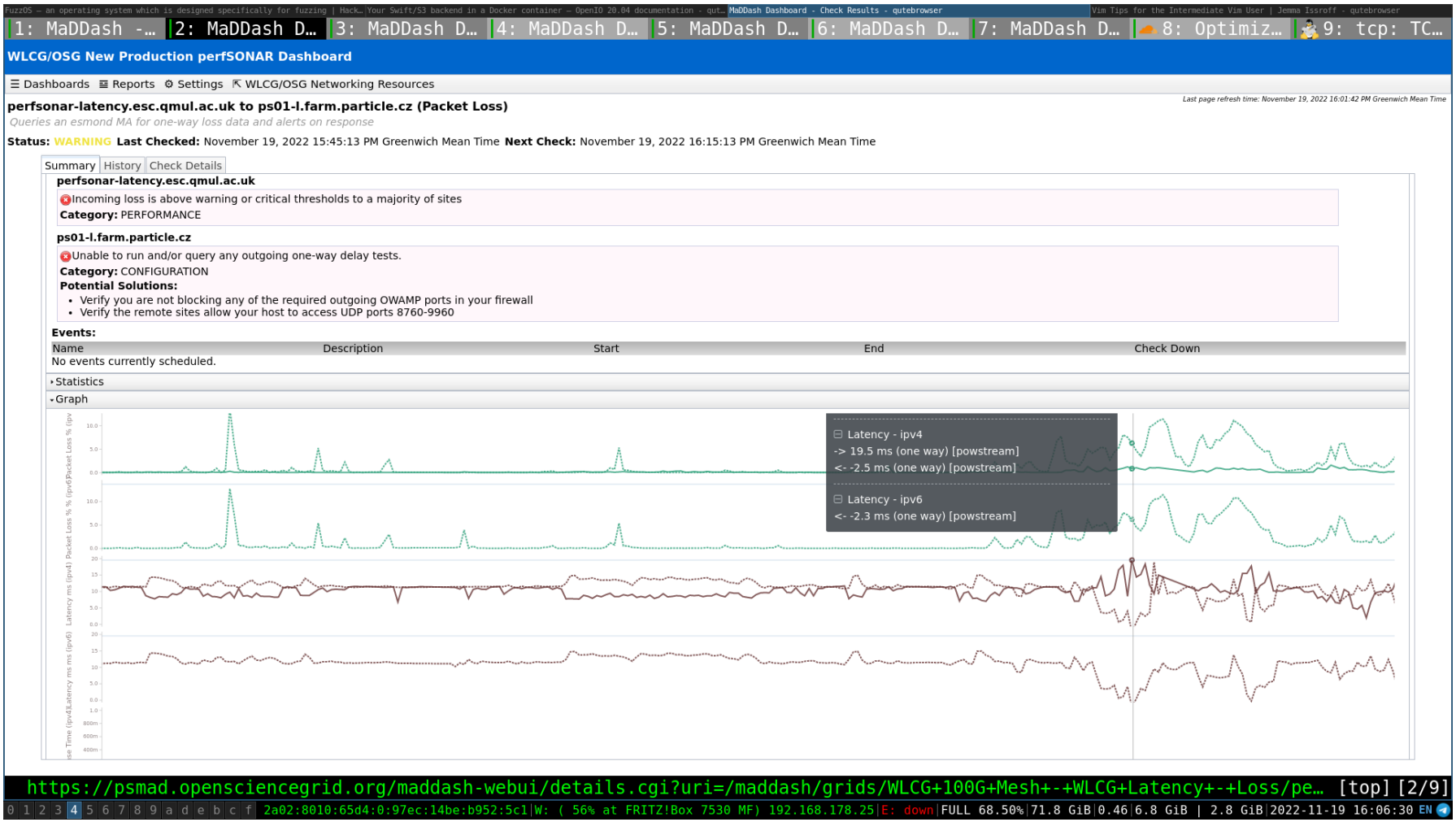
- Failure in reverse IP resolution
- **Only** latency tests were affected. Why only latency?



Surprise when zooming in



Negative latency for QMUL-Prague

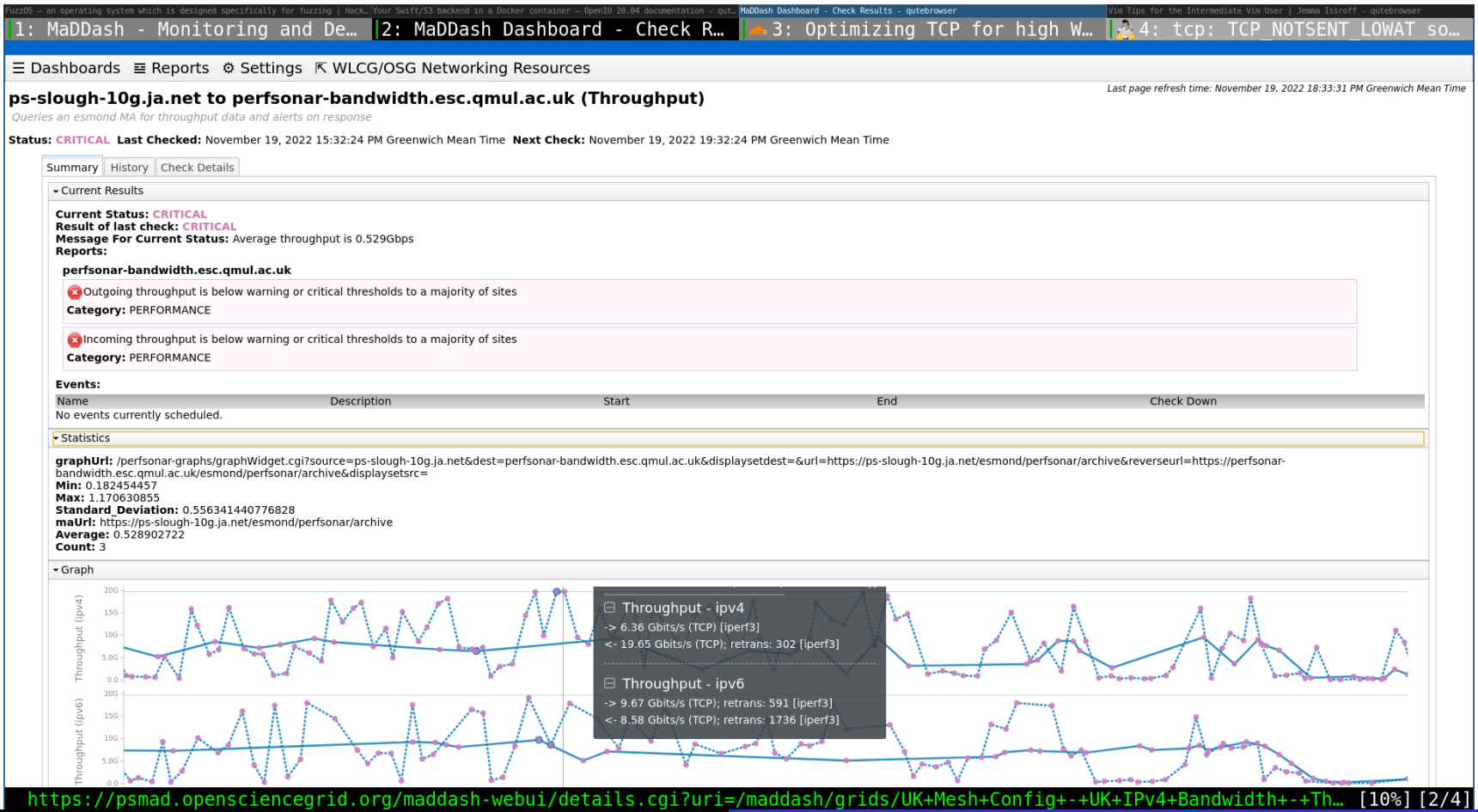


Raul H C Lopes (Jisc)Raul.Lopes@jisc.ac.uk

Odd and nice things in perfSONAR land

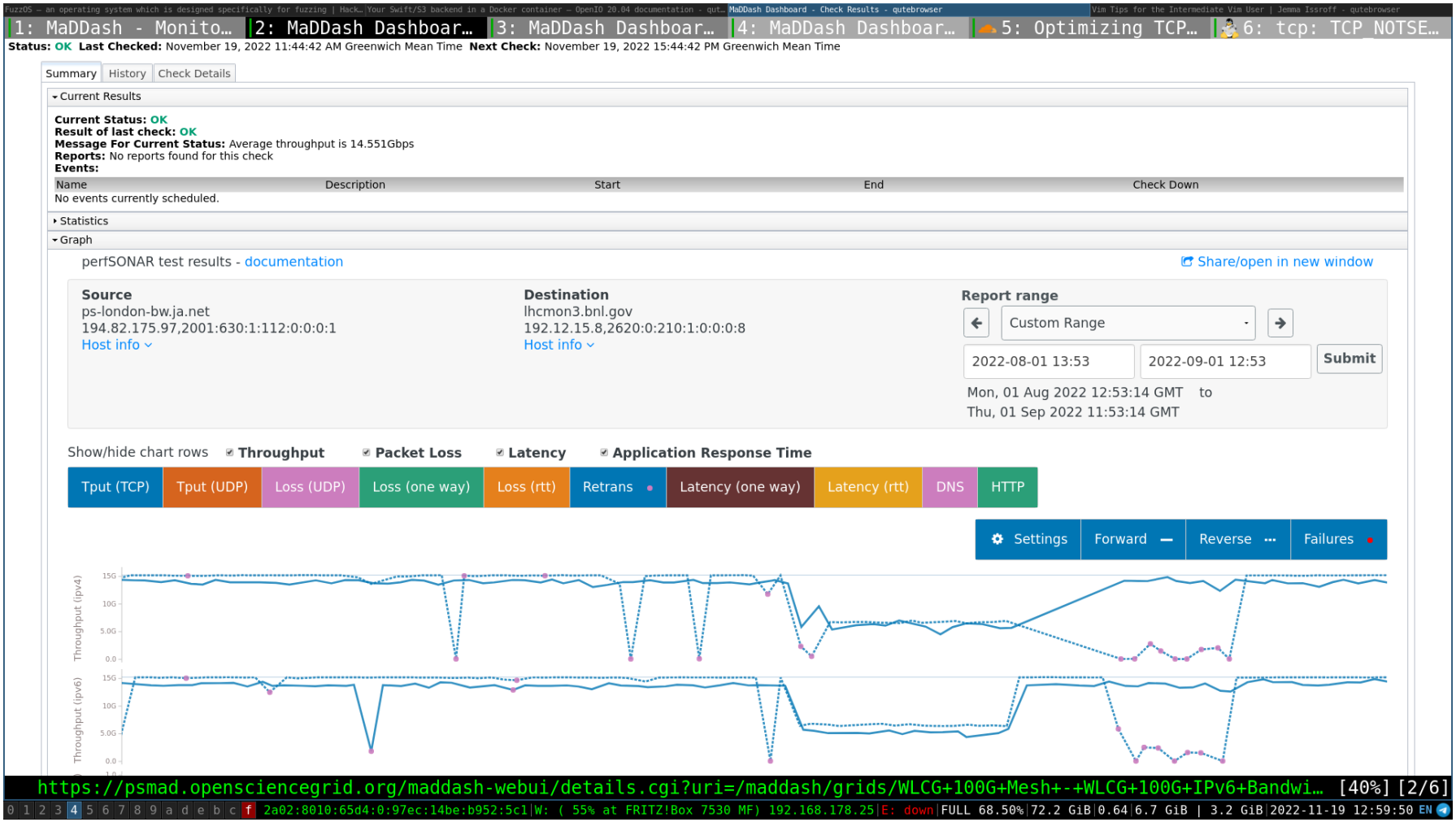
QMUL to Jisc Slough: 20G on a 10G link

- The negative latency (previous slide) coincides with the 15-20G spikes.
- Google on TrueTime https://cloud.google.com/spanner/docs/true-time-external-consistency



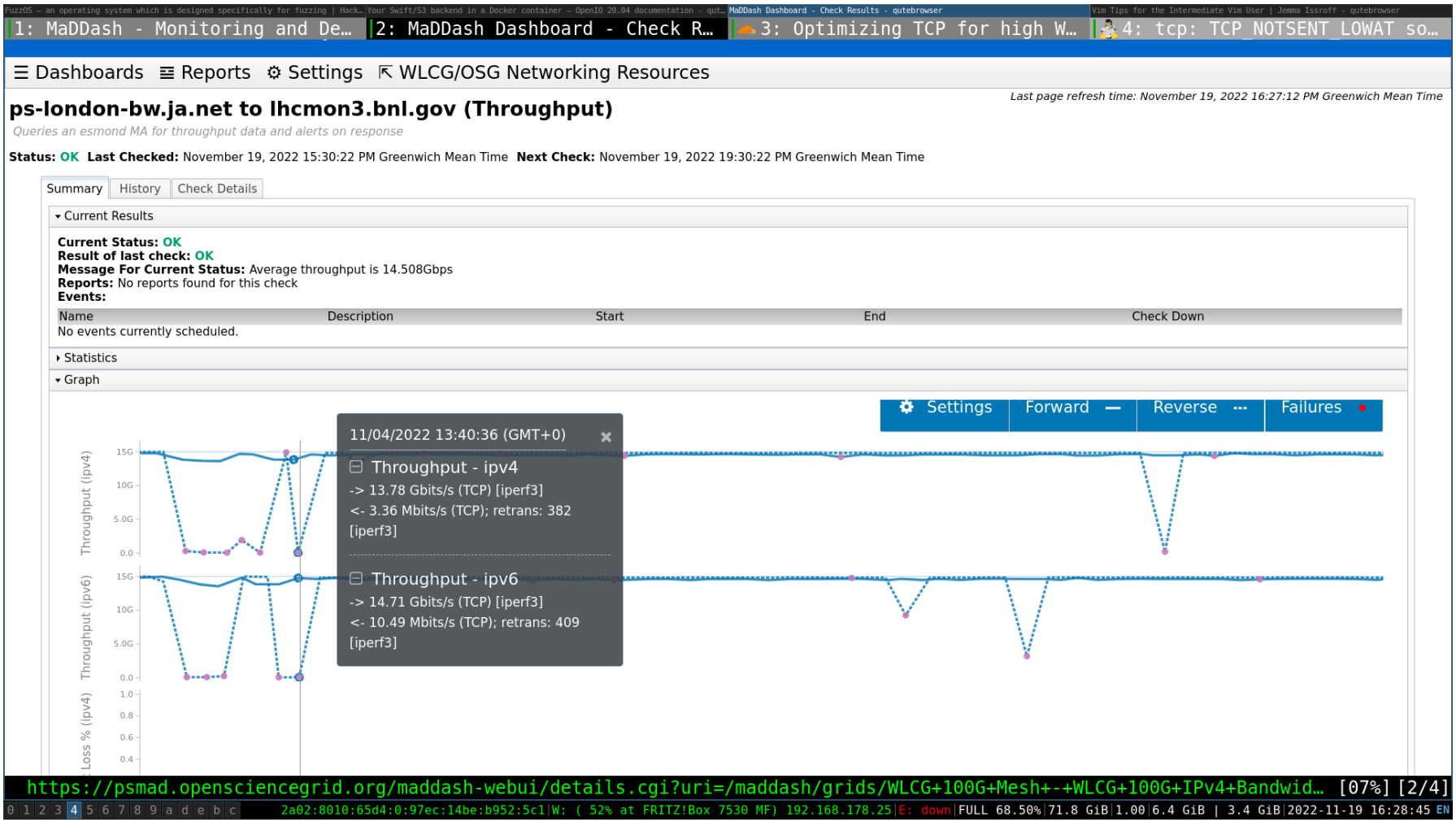
Statistics associated with plots are frequently incorrect

MTU: Tim's request for a comparison of MTU at 9000 vs 1500

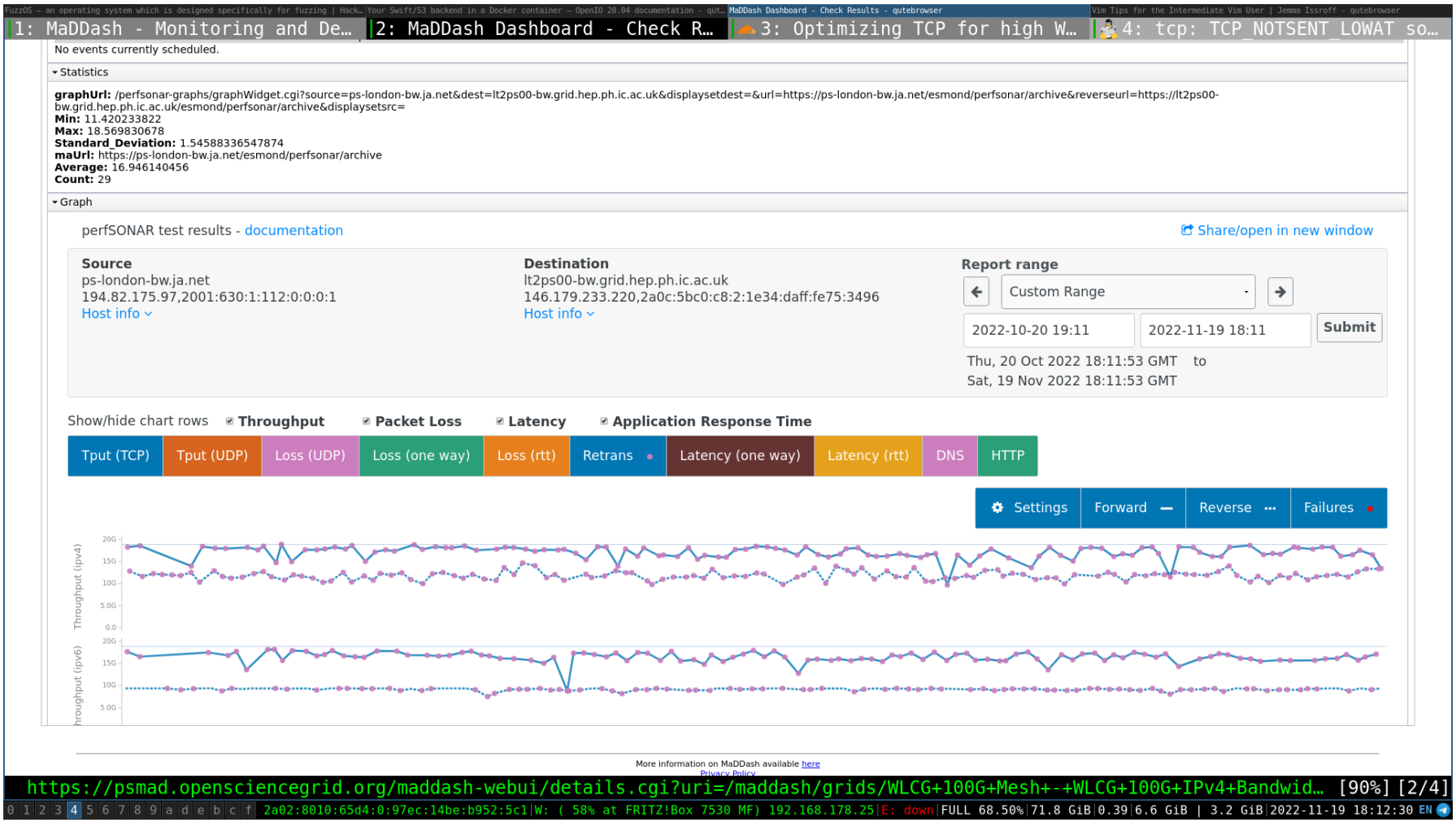


Is Brian showing a link possibly maxed out?

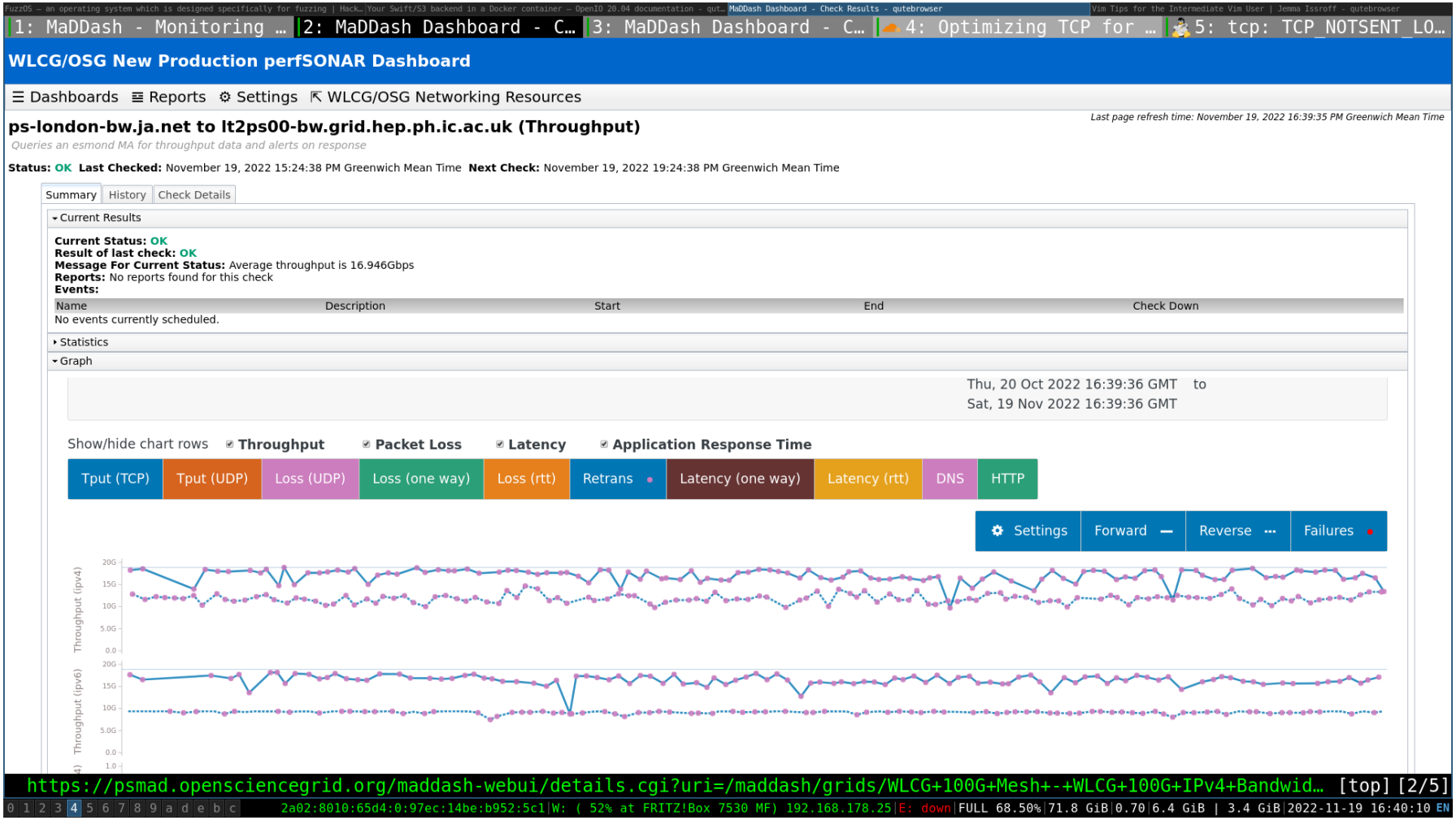
Jisc London view — BNL path



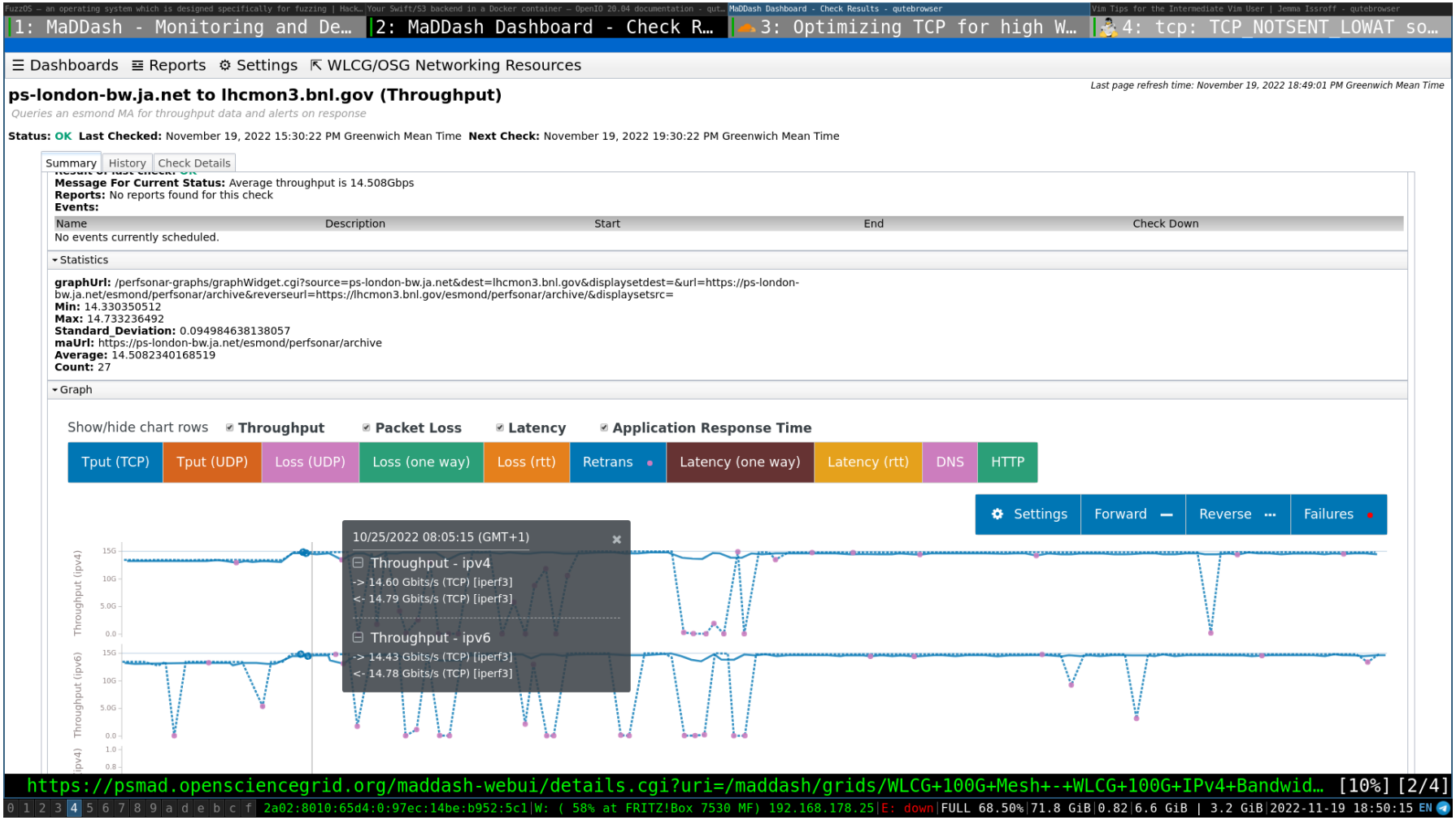
Test in the UK went unaffected



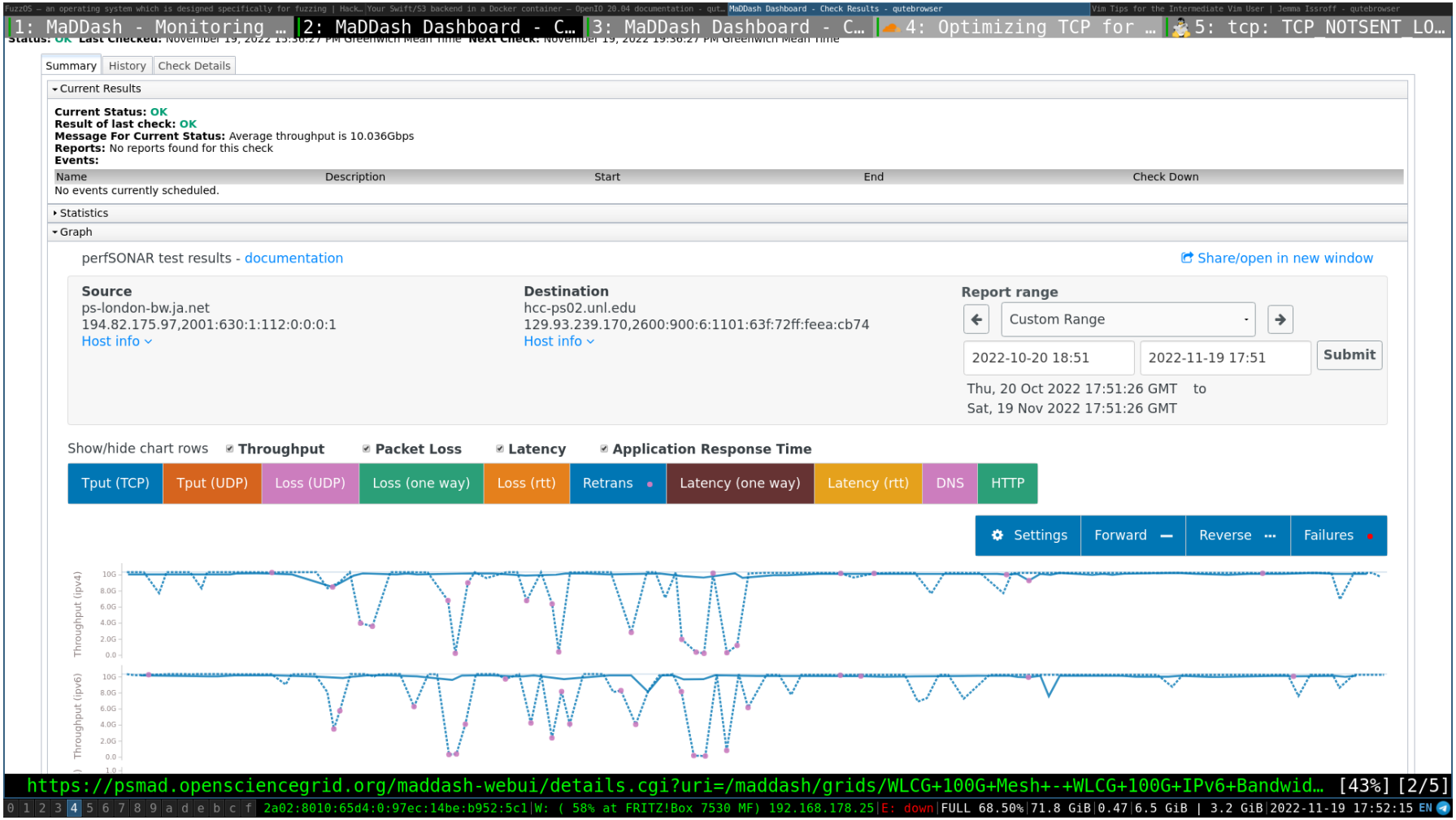
Tim: "Is Imperial capping outbound IPv6 at 10G?"



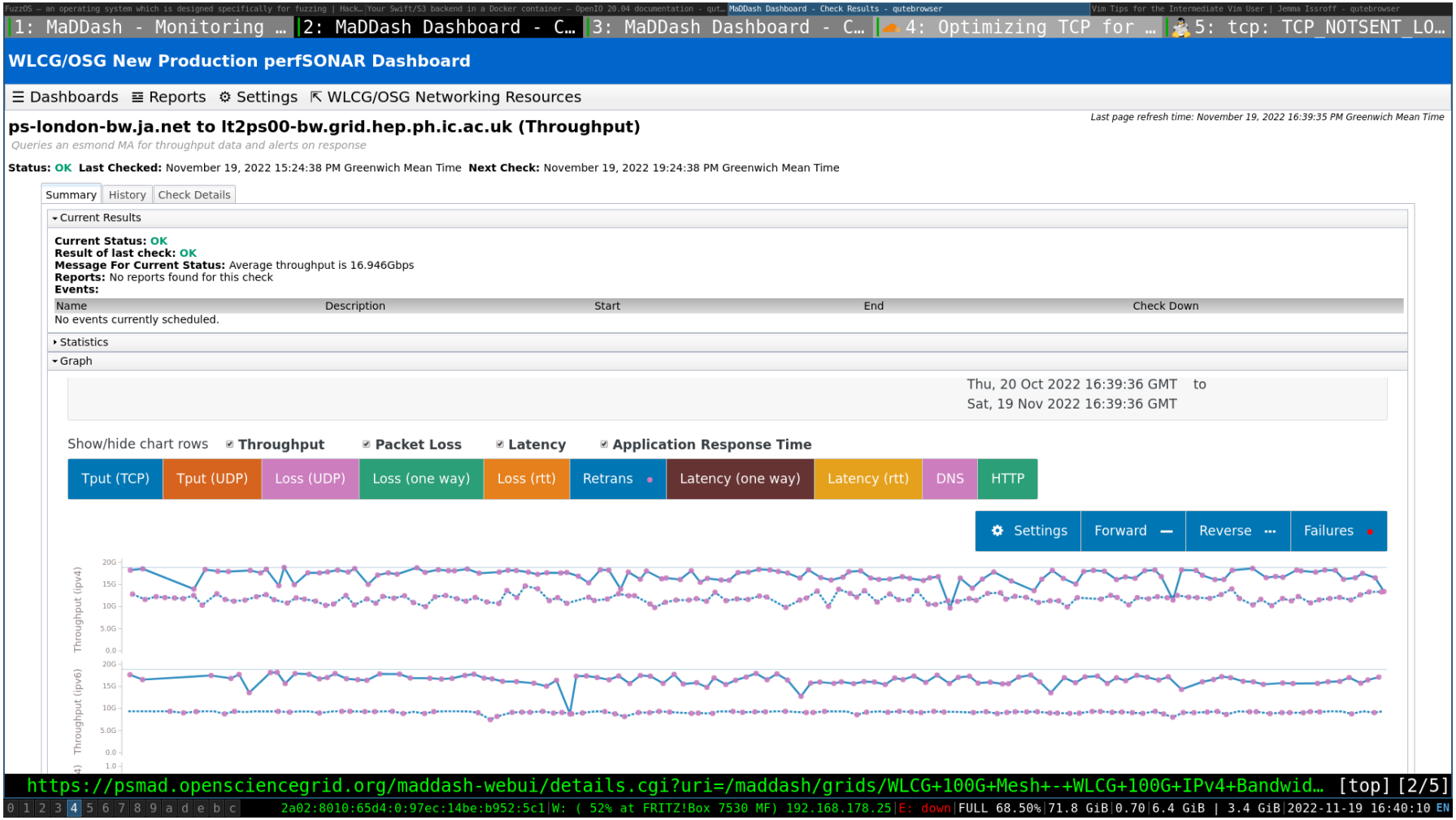
Jisc London to BNL: feel good of MTU=9000



Jisc London to UNL: feel good of MTU=9000



Jisc London to Imperial: MTU=9000 vs MTU=1500



Jisc London configuration

- Kernel: tested $5.14, 5.17, 6.0$
- CWA: tested reno, htcp, cubic, bbrv1; bbrv1 being used.
- tcp_rmem = max test 256MB and 512MB; 512MB in use.
  - Google and Cloudflare have reported that small tcp_rmem on high latency limits throughput.
- tcp_adv_win_scale $= -2$; it limits the TCP window to 3/4 of available receiving memory.
- tcp_wmem = max test 256MB and 512MB; 512MB in use.
- net.ipv4.tcp_notsent_lowat = 131072
  - limit number of unsent bytes in the writing queue
  - reduce buffer bloat https://lwn.net/Articles/560082/
- cpu power set to performance.
  - Important for smooth receiving.
  - Evidence when testing with RNP: a 4 cores server was capable of sending at 97 Gbps.
  - Same tests, same server: it never achieved more than 70Gbps while receiving.

Jisc Netperf team

- Tim Chown (task leader)
- Duncan Rand (Speaker of House and WLCG Ambassador)
- Chris Walker (University Ambassador)
- David Richardson (Network Engineer)
- Raul Lopes (some semblance of a sysadmin)
- Supermicro servers behind a Freertr.