100G tests between GEANT, SURF and RNP

Raul HC Lopes
JISC

2nd Performance Management Workshop, 8th March 2022

Public

www.geant.org

The briefing

- Test the new Bella link: 100G link from Portugal to Fortaleza (Brazil)
  - Isolate routing issues
  - Tune perfSONAR hosts
  - Find how MTU, congestion window, an other tuning can affect the transfers.
  - Pace transfers to zero retransmissions
  - Could we drive a consistent 80 Gbps TCP throughput through the link?
- The host used in the tests:
  - SURF perfSONAR host: ps2.netherlight.net
  - Fortaleza perfSONAR host: nfor.rnp.br

The path

```
1 et0-0-0.1421.jnr01.asd001a.surf.net (145.146.0.5) AS1103 0.5 ms
  SURFNET-NL SURFnet, The Netherlands, NL
2 surfnet-gw.mx1.ams.nl.geant.net (62.40.124.39) AS20965 0.8 ms
  GEANT The GEANT IP Service, NL
3 surfnet.mx1.ams.nl.geant.net (62.40.124.38) AS20965 0.5 ms
  GEANT The GEANT IP Service, NL
4 ae7.mx1.fra.de.geant.net (62.40.98.187) AS20965 7.5 ms
  GEANT The GEANT IP Service, NL
5 ae3.mx1.gen.ch.geant.net (62.40.98.181) AS20965 15.5 ms
  GEANT The GEANT IP Service, NL
6 ae2.mx1.mad.es.geant.net (62.40.98.66) AS20965 35.4 ms
  GEANT The GEANT IP Service, NL
7 ae4.mx2.lis.pt.geant.net (62.40.98.96) AS20965 44.7 ms
  GEANT The GEANT IP Service, NL
8 redclara-gw.lis.pt.geant.net (62.40.127.151) AS20965 106.9 ms
  GEANT The GEANT IP Service, NL
9 200.0.206.71 AS27750 106.2 ms
  Cooperacion Latino Americana de Redes Avanzadas, UY
```

Impact of MTU 9000

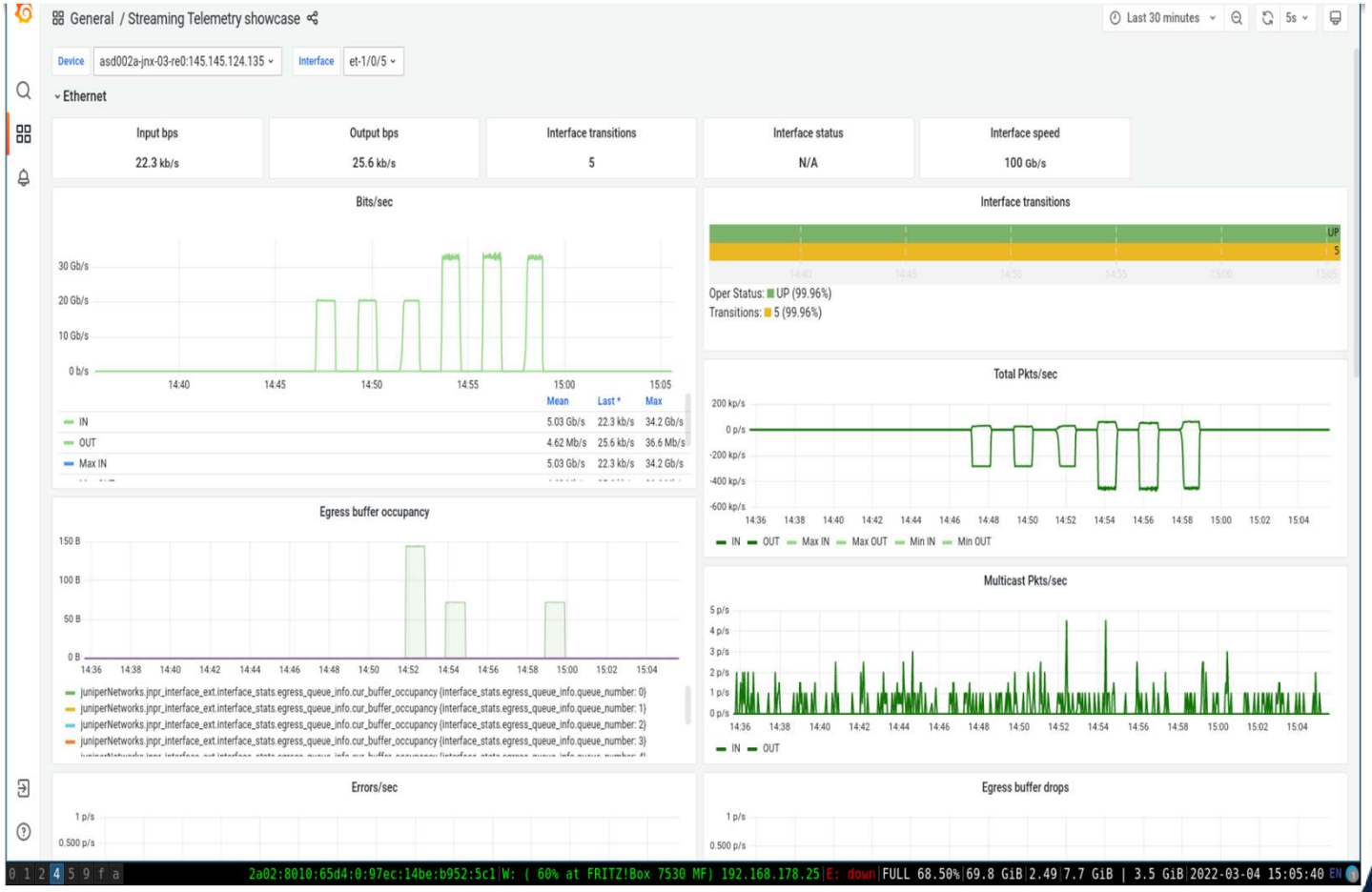
- SURF and the hosts in the RNP all have MTU equal to 9000.
- ICMPv6 is correctly configured and PMTUD was never an issue. SURF is IPv4 only.
- The question: can increasing the MTU from 1500 to 9000 improve a host's performance?
- Simulated the limitation on MTU using pscheduler option **–mss=1500**
- Disclaimer: I run a WLCG site using MTU at 1500. It works fine if you have a twenty plus storage system.
- Important notice: all plots shown come from the SURF streaming telemetry system.

MTU: iperf3 and iperf2: SURF to Fortaleza

- iperf3: two left columns: 9000 and 1500
- iperf2: two right: 9000 and 1500 (iperf2 problems with mss setting?)

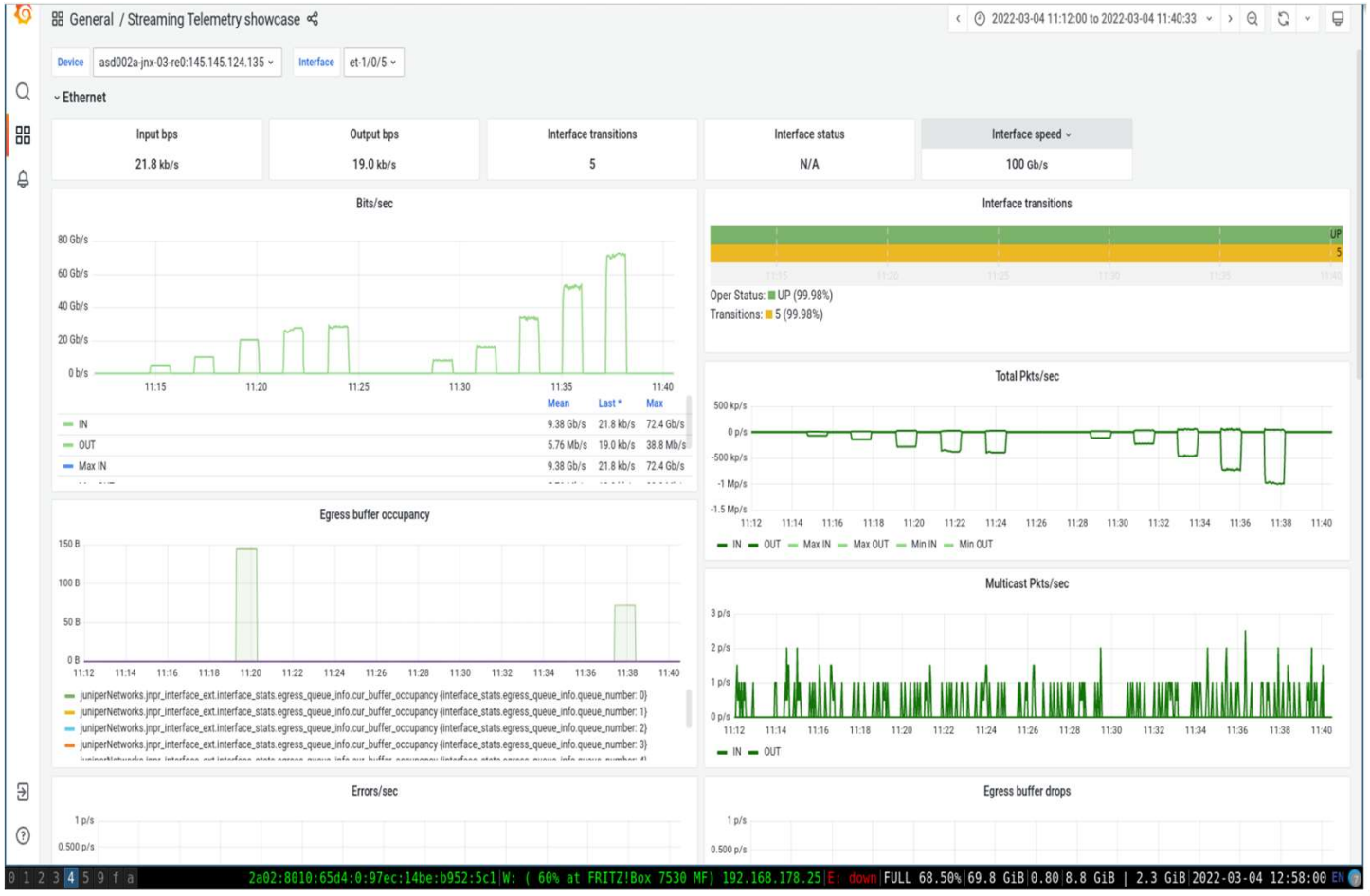
Effect of the Congestion Window Algorithm (CWA)

- CWA: reno, htcp, cubic
- iperf3 (first 3 curves) and iperf2
- 4 streams and 64M socket length



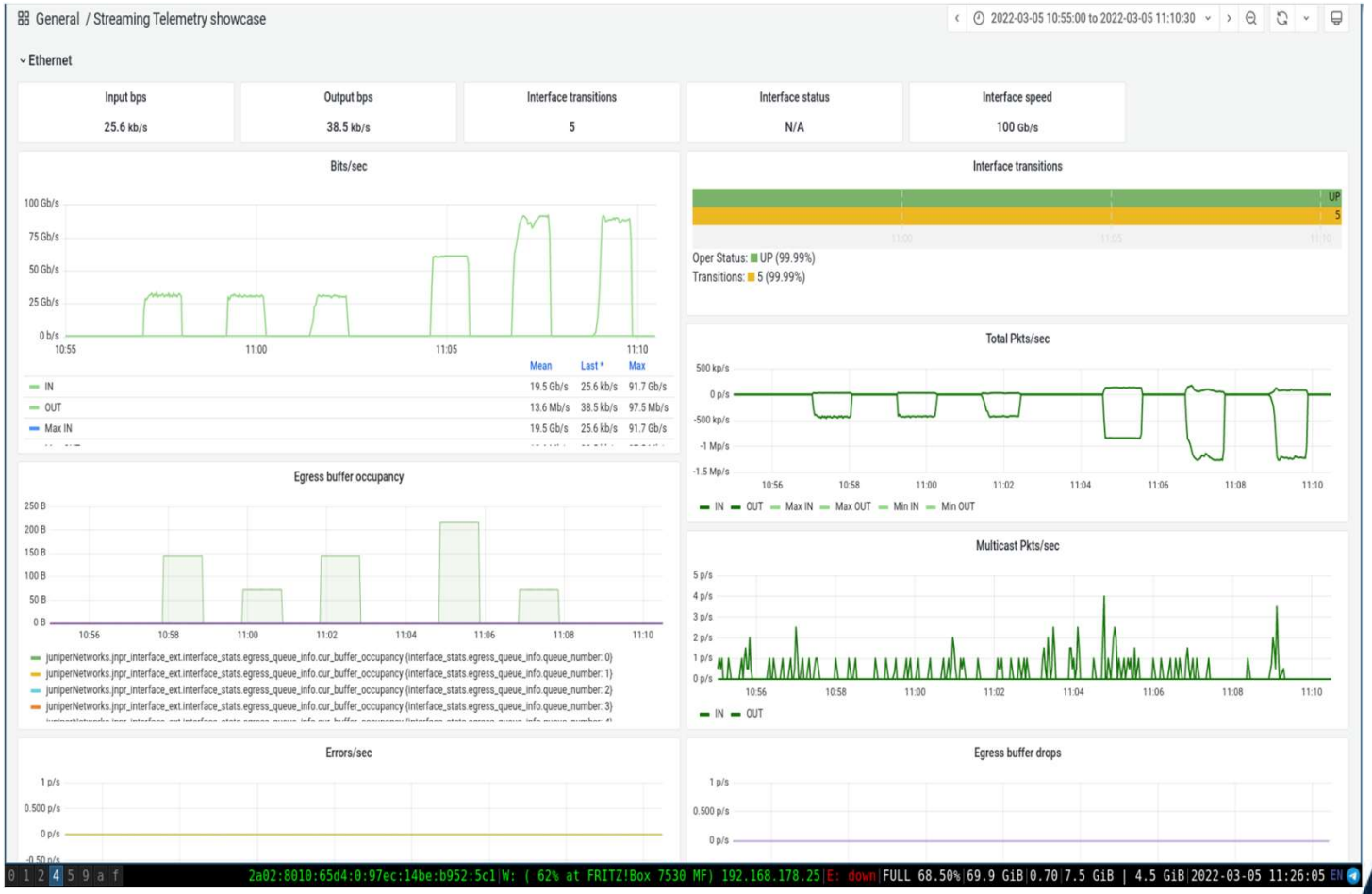
Parallel Streams

- number of streams: 1, 2, 4, 6 8
- iperf3 (first 5 curves) and iperf2
- htcp and socket at 64Mb.



Six Parallel Streams, but…

- number of streams: 6
- CWA: reno, htcp, cubic
- iperf3 (first 3 curves) and iperf2
- socket at 128Mb



Six Parallel Streams, but…

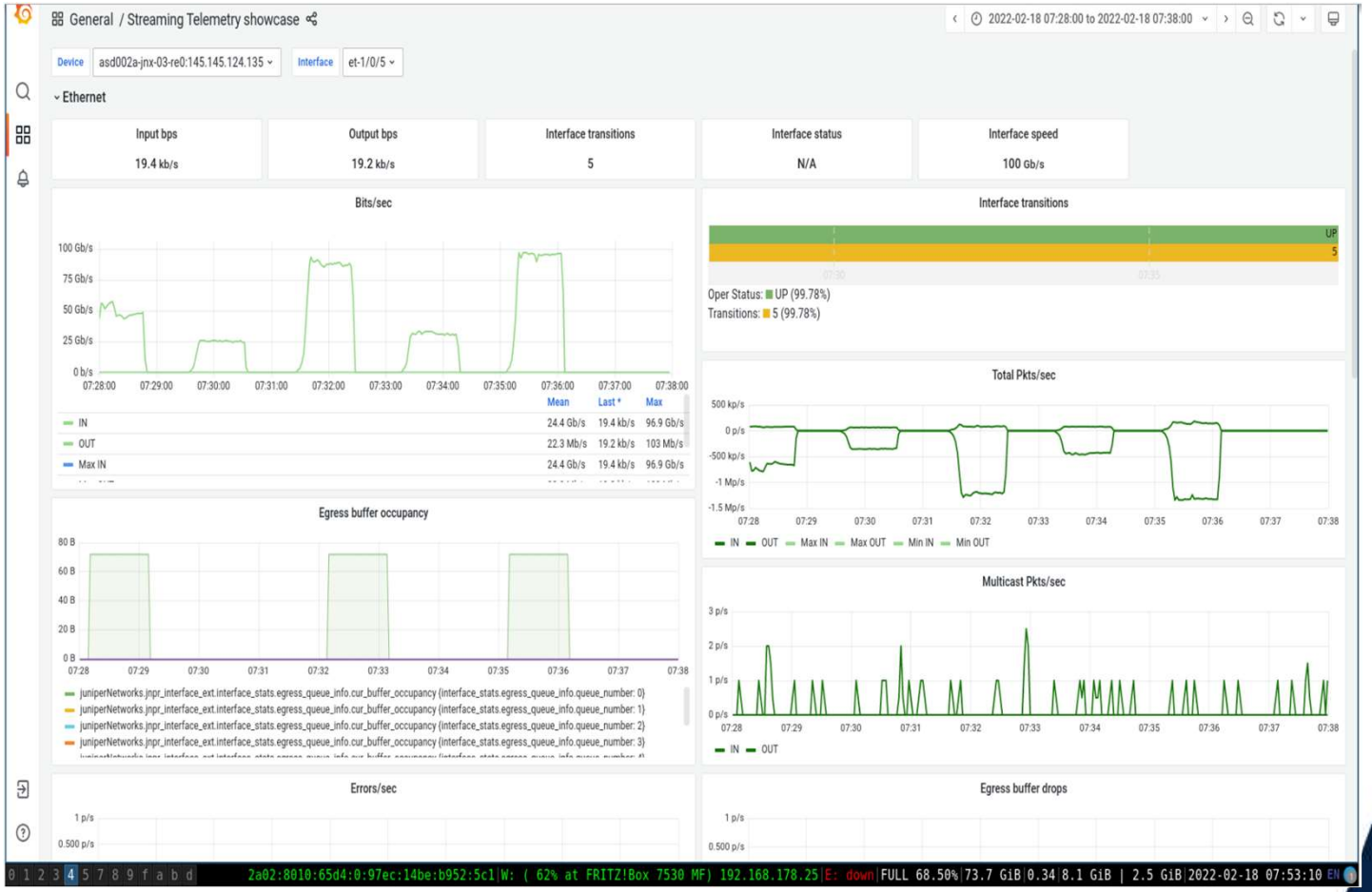
- number of streams: 6, 8
- CWA: reno, htcp, cubic
- iperf3 (first 3 curves) and iperf2
- socket at 128Mb

The issues

- Routing from to RNP was flipping between Bella link and USA
  - Corrected by Max Mudde and GÉANT colleagues in December
- BGP issues affecting transfers to São Paulo and Salvador
  - Corrected by RNP in January
- Present state
  - Transfers to Fortaleza on 2 streams with zero retransmits any time.
  - pscheduler tests using iperf3 deliver consistent 30 Gbps
  - Asymmetry to be investigated: transfer from SURF to Fortaleza show much less retransmits than the other way round.

SURF to Fortaleza: 96 Gbps on SURF telemetry system

- iperf2 (07:36): cubic
  6 streams (no gain with 8) and 256M socket
- htcp (07:32) slightly lower



Team involved

- RNP: Marcus Schwartz, Jeferson Souza
- SURF: Max Mudde
- Jisc Netperf: David Richardson, Tim Chown, Duncan Rand

- To come: Report covering experiments with RNP, USA, Europe

GÉANT
Networks • Services • People

Thank you

Any questions?

www.geant.org

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 856726 (GN4-3).