Networking News and Developments

GridPP53, 10th April 2025
Matt Doidge, Lancaster University
on behalf of many, but special thanks to Dave, Tim and Duncan

Overview

- Landscape
- Highlights from the LHCONE/LHCOPN meeting
- Plans and Challenges in the UK.

There’s going to be a bit of a GridPP/WLCG slant.

Disclaimer - I’m a networking novice, and I’m hoping someone will chime in if I accidentally start spreading misinformation.

The Networking View, according to a Grid Site Admin

- The importance of good networking is recognised across the board
- In the UK, 25Gb/s connections in the rack, 100 Gb/s connections to Janet are becoming more common (X x 10Gb/s otherwise).
  - 40Gb/s in rack is here, 100Gb/s (kinda) affordable.
  - Rates are such that careful planning is needed to optimise intra-machine room traffic
  - 400Gb/s on its way too - but likely still a bit too pricey
- Very much all ethernet everywhere
  - No infiniband here
- The “IPv6-enable battle is almost won
  - Most services dual stacked, by hook or by crook. 90% WLCG traffic v6
  - The next step is to start removing IPv4 where we can
- Monitoring and Tuning are important to make the most out of the network
  - Perfsonar, but how to make the most out of it?
  - Jumbo frames
  - The fantastic fasterdata esnet recommendations

News from the 54th LHCONE/LHCOPN meeting

- Marked 20 years of the LHCOPN
  - There was a “birthday” cake
- Special hosts SKA, with a trip to Jodrell Bank for the second day
  - The second day was held in the SKA Council Chamber - can we get one?
  - Discussion of SKA’s networking needs and challenges
- A look at the state of Global Networking
  - A view that research network infrastructure provisioning at the national/international level was generally in a good place.
- A lot of news on the networking efforts from across the pond
  - Tb networking
  - Jumbo Frames
- “Mini” Data Challenges, with an eye on the next big DC (pencilled in for 2027)
  - Mini-DCs held in both the UK and US
- Updates on work with monitoring; perfsonar, packet tagging
- IPv6 news

And more! A more in depth summary was given at a recent WLCG-DOMA meeting.

LHCOPN/LHCONE 101

A quick reminder about what these are:

- **LHCOPN** - LHC Optical Private Network, is a set of dedicated network light paths used for the transfer of LHC data from CERN and between member sites, notably the Tier-1s.
- **LHCONE** - LHC Open Network Environment, is a (*checks notes*) Layer 3 Virtual Private Network connecting member sites. Member sites have their endpoints (prefixes) registered (in CRIC) and sign AUPs, creating a “walled garden” of increased trust.
    - BGP (Border Gateway Protocol) tagging and VRF (Virtual Routing and Forwarding) are used to create the ONE, but I’m not confident I could explain how.

LHCOPN updates

As mentioned, the OPN turned 20 this year.

- A lot of retrospective
- Not many recent updates - the OPN is in good health, and working well
  - 2.28 Tb/s traffic
- One presentation contained a sobering view of the GREN undersea cable networks, and the places that they “clump” (like the Red Sea)

LHCONE updates

- Expected uptick in traffic - pushing 1Tb/s
  - 90% of the traffic is v6, efforts into identifying the last sources of ipv4
    - and eliminating them
- Exceptionally good resilience
  - Failovers tested by fibre cuts (real and simulated)
- A reminder to make sure information is in CRIC
- Discussion, but no conclusions, were had regarding the AUP
  - WG being formed to discuss this

“Mini” Data Challenges

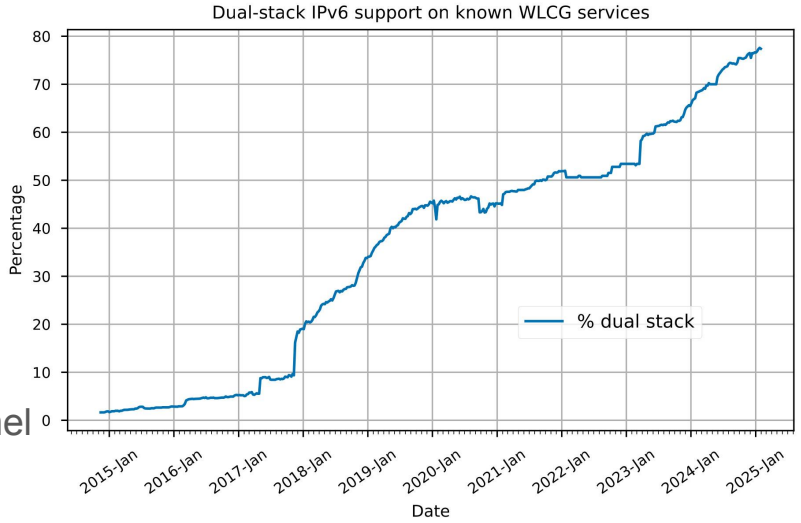
- After DC24 both US and UK institutions separately decided on the need for “mini-DCs”.
  - The UK mini-DCs are discussed tomorrow
- Similar findings and challenges faced across both sets of mDCs
  - Scope creep was an issue - difficult to keep things “mini”.
  - The network was very rarely the bottleneck - issues in both throughput and reliability tend to originate in the storage at one end or the other.
    - Where network was the issue it was often internal, not on the NREN (Janet/esnet) side
  - The mini-DCs proved very useful on both sides of the Atlantic, with an aim to repeat them.
- Naturally discussion on DC27, some regarding the size - consensus from many was that the currently projected target rates are aggressive and there’s a high level of uncertainty. Meeting them would require significant “premature investment” to meet them; spending just to meet the needs of tests is explicitly discouraged.

Monitoring

- By coincidence Perfsonar has also reached the 20 year milestone
- Perfsonar Developments and Plans
  - MTU reporting - important as more sites enable Jumbo Frames
  - Switch off v4 tests, “releasing” testing slots
    - Already active for new meshes, using v6 as default.
  - Other general improvements
- Packet Marking; Scitags and Fireflies
  - Supported by several storage solutions
  - Enables real time monitoring of flows
  - Allows to tag traffic with activity, easier accounting
  - Collectors need to be deployed by NRENs
    - Main global collector at ESNET, proposal for GEANT to host their own.

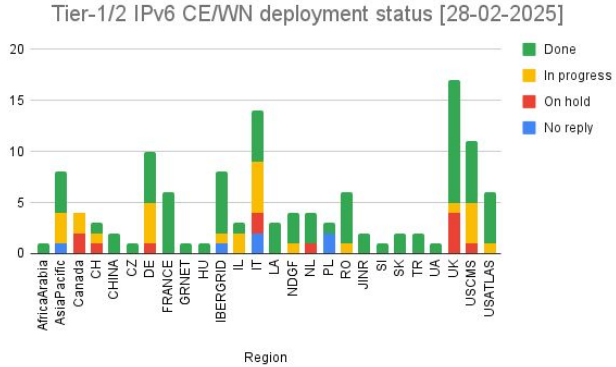
Miscellanea

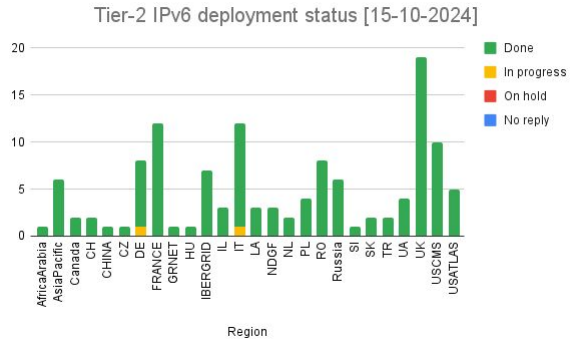
- Jumbo Frame testing in the States had promising results
  - Improvements seen, especially noticeable over large distances
  - Even when found to not provide any significant advantage there was no sign of Jumbo Frames causing issues.
- Tb-scale transfer tests
  - Utilising 3x400Gb, done over distances.
  - 400Gb seems quite common in the U.S.
- Deployment of Dual Stack IPv6 progressing well
  - “v6-only” is another matter.
  - Current WLCG campaign is IPv6 on the WNs
  - Work on v6 only WNs, test cluster for CMS at Brunel



GridPP Networking Status and Plans

- The Dual Stacking campaigns have done well.
  - Most of us are mostly dual stacked across our resources.
- Good working relationship with Jisc
- Number of storage sites on 100Gb
  - RAL on 400Gb + an OPN link!
  - Need to be able to have the storage infrastructure and tuning in place to take advantage of this.
  - Network tuning, Sufficient Gateways.
  - Transfer tuning to match the storage? How many files/streams?
- Networking Monitoring
  - Perfsonar, Packet Marking, XRootD Shoveller
  - Having network monitoring paint an accurate picture of data transfer potential is a challenge
  - Need to increase adoption of the DC24 network monitoring tool, expanded to be IP version specific
- LHCONE is considered a future requirement for the UK Storage Tier-2s.
  - Work on our DMZs/DTNs (another talk tomorrow)
- Shortage of networking effort at institutes
  - Many can count the number of network engineers on one hand.





Closing Thoughts

- Like everything else with computing, networking is throwing about numbers that a few decades ago we'd never have dreamed of - X Tb/s
- Networking improvements take time, effort and investment to put in place
  - Need to be able to justify increases
- Increasing bandwidth does not magically improve throughput
  - Need to make sure storage and transfer infrastructure can keep up.
- Monitoring is important
  - But as a corollary to the previous point, good networking metrics do not always mean good transfer metrics. Need to match the two.
- Start with careful planning and finish with thorough tuning
  - Very easy to block our own flows within our racks/data-centres and create our own sources of congestion
  - Tuning at all levels can provide significant improvements
    - But not miracles...