LHCOPN-LHCONE meeting #52 summary notes

GDB at CERN, 10th July 2024
edoardo.martelli@cern.ch

Venue

- 10-11 of April 2024
- Hosted by INFN Catania (IT)
- Sponsored by Internet2
- co-located with
    - GEANT Next Generation Networks workshop
    - SKAO-LHCONE joint meeting
- 45 people in presence and ~20 connected remotely

- Agenda and presentations at
   https://indico.cern.ch/e/LHCOPNE52

LHCOPN - update

- 2.66Tbps of aggregated bandwidth to the Tier0

- Traffic stats: moved 619PB in the last 12 months. +27% compared to previous year

- CN-IHEP, new LHCb Tier1: connected at 20Gbps via CSTnet and GEANT

- NDGF-LHEP, new NDGF distributed Tier1 site at UniBern (CH): connected at 100Gbps

- TW-ASGC terminated Tier1 activities. All links decommissioned

- NLT1 Testing 800Gbps connection
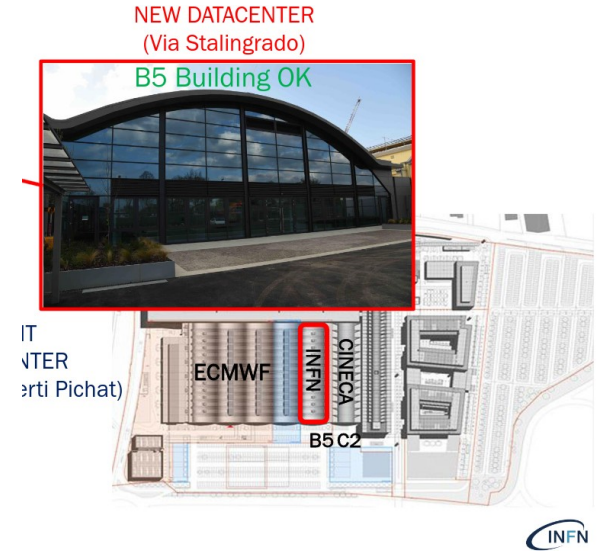
- IN2P3 upgraded to 2x100G, FNAL to 400Gbps
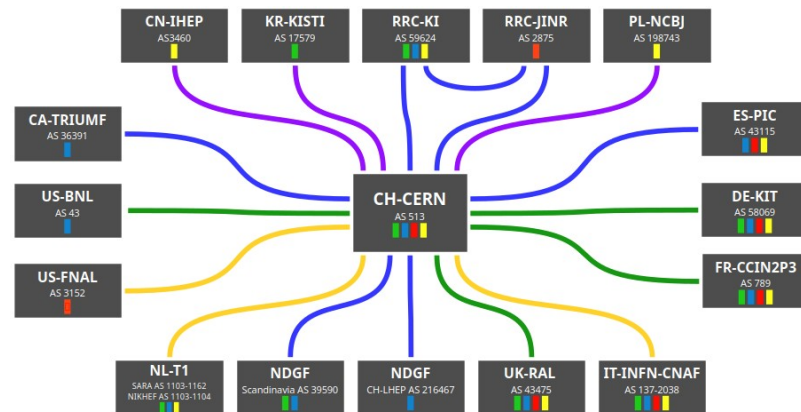
LHCOPN Total Traffic (CERN → T1s)
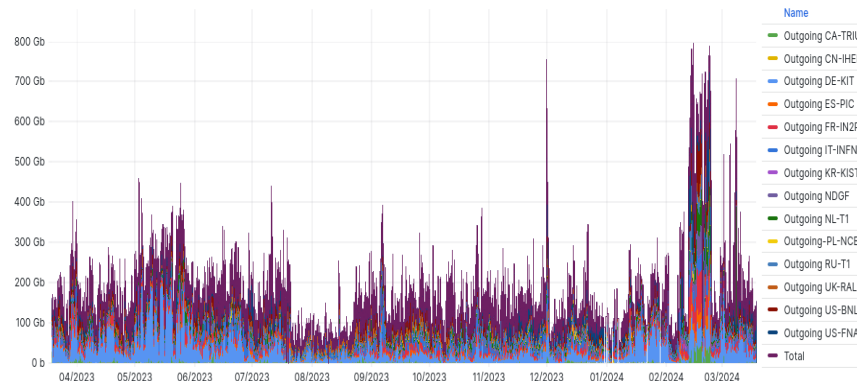
IT-INFN-CNAF update

New CNAF datacentre building is ready

- 150 racks, being equipped with network and servers
- It will be connected to CERN (LHCOPN) and to the old CNAF data-centre to migrate the data

CNAF-CERN DCI in production

- 4x100Gbps links used for DC24 and now in production
- Previous 2x100Gbps links kept for backup
- Next step:
    - connection to new datacentre
    - use of 400Gbps
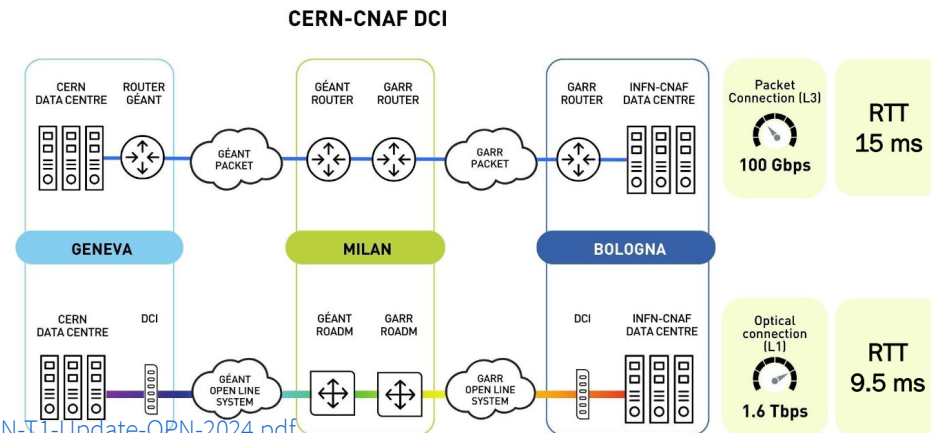
NLT1 and SURF update

800Gbps trial

- SURF and Nokia implemented a 800Gbps connection on a single wavelength over the SURF dark-fibre CERN-Amsterdam

- Link tested with 660Gbps of ATLAS transfers from CERN to NIKHEF

AMS — BRU — PAR — GEN

- ROADM site
- Amplifier site
- FOADM site
- ↔ Raman span
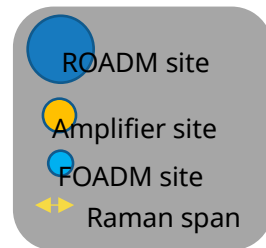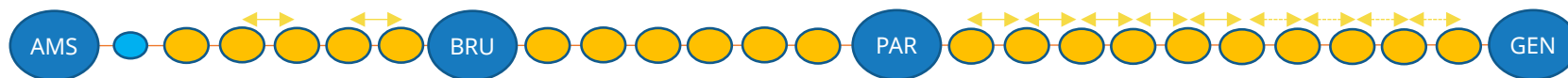
IHEP update

LHCb Tier1 at IHEP is ready to run

- 3216 CPU cores and 3.2PB disk storage

LHCOPN connectivity

- Primary 20Gbps link via Singapore Marseille

- Backup via Singapore London

- 20Gbps guaranteed on the direction IHEP→CERN. Not possible to guarantee the bandwidth in the direction CERN→ IHEP, but the links are guaranteed to not be over-provisioned

LHCONE L3VPN - update

News

- DC24 impact well visible by all LHCONE providers

- New LHCONE network providers:
    - SWITCH (Switzerland) for University of Bern-LHEP
    - FCCN (Portugal) for NCG-INGRID-PT

- GARR (Italy) access upgraded to 2 x 300G (MIL, MAR)
- RENATER (France) access upgraded to 400G in GVA, 300G in PAR (about to become 400G)

https://indico.cern.ch/event/1349135/contributions/5765308/attachments/2834555/4953142/2024-04-10_ECapone_LHCONE_L3VPN.pdf

Network bandwidth values in CRIC

Added two help pages to help user correctly fill the bandwidth values in CRIC
- https://twiki.cern.ch/twiki/bin/view/LHCONE/CRICguideNetSite
- https://twiki.cern.ch/twiki/bin/view/LHCONE/CRICguideNetworkRoute

https://indico.cern.ch/event/1349135/contributions/5765343/attachments/2834290/4952640/LHCOPNE-20240410-52-CRIC-database.pdf

ESnet Transatlantic Engineering

Trans-Atlantic upgrades

- Now In Production:
    - – 400G New York - London
    - – 400G Boston - CERN
- Currently underway:
    - – 400G Boston - London [in production at the date of GDB]
- Trans-Atlantic capacity targets:
    - – 3.2T in 2027, in advance of Run 4
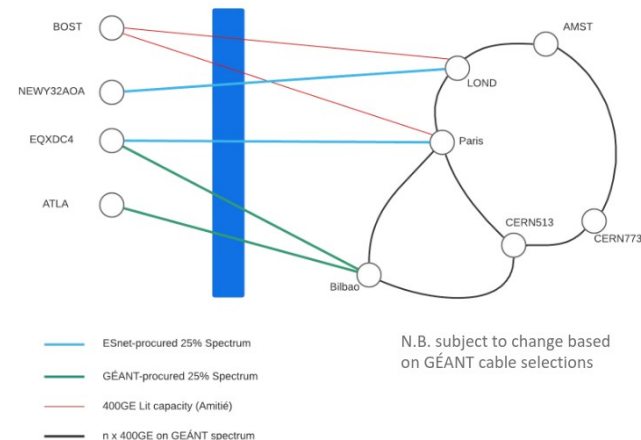
US - Europe Connectivity Plans

- Collaborating with GEANT to share spectrum on subsea
- Two additional EU PoPs: Paris (firm), Bilbao? (tbd)
- n x 400G EU rings in partnership with GÉANT

US-EU traffic engineering

- Most of the US-EU LHCONE traffic use ESnet transatlantic links

perfSONAR monitoring update

Updates to perfSONAR and OSG/WLCG network measurement platform

- perfSONAR 5.1 is coming with new features and will require all sites to update OS.

- Plan to adapt the network measurement platform to benefit from changes in 5.1

Ongoing efforts in network analytics and ML methods for our data

- Focus on pre-processing (gaps, predictive models) and anomaly detection

- Opportunity to collaborate on models and data sets

Preparing monthly meetings with site network teams:

- Discuss how sites are deploying, managing and planning for WLCG networking requirements

We have to continue to watch our network monitoring infrastructure as it is a complex system with lots of areas for issues to develop

perfSO+NAR

BelleII update

- Belle II resumed data taking in early 2024.

- The Data Challenge 2024 has been successfully completed, achieving the maximum target of 18.5Gbps outbound from KEK to RAW DCs.

- Although IPv6 traffic is already significant, the entire computing infrastructure is currently not optimized for IPv6-only resources.

- Data transfer activities are functioning correctly over the research network and are monitored via Rucio

https://indico.cern.ch/event/1349135/contributions/5826769/attachments/2834693/4953420/Belle2_LHCONE-LHCOPN-April-2024.pdf

WLCG Data Challenge 2024

Very comprehensive report from Katy Ellis

- DC24 was a success - we learnt a lot
- Testing in advance was very helpful
- Post-DC24 analysis goes on…
- There are other bottlenecks than network bandwidth
- Maintenance of DC injections was challenging
- FTS instances got pushed to their limits, particular the ATLAS one
- Overloading FTS was a concern

https://indico.cern.ch/event/1349135/contributions/5765317/attachments/2834780/4953601/DC24%20-%20LHCOPN_LHCONE%20workshop.pdf

DC24 from ESnet perspective

ESnet very committed to DC24, preparing upgrades well in advance for all the Tier1 and Tier2 sites in the US and to CERN. Unfortunately two transatlantic 400Gbps links were not delivered in time.

In addition, right before DC24 the only 400Gbps link in production was cut. ESnet managed to get some additional bandwidth from NEA3R and GEANT. Luckily the link was repaired in time.

Now preparing the High Energy Physics Requirements Review in 2025

Starting US CMS/ATLAS Tier 1 & 2 outreach to prepare for DC26

Measurements was very useful in spotting and addressing issues.
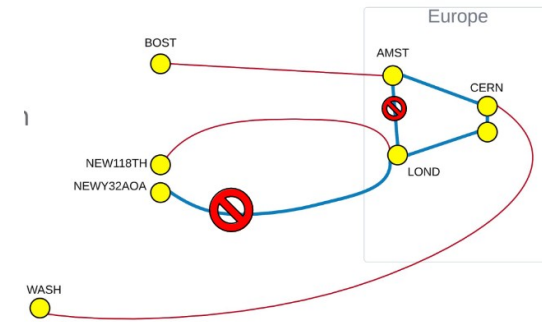
NOTED and BBR at DC24

CERN tested NOTED and BBRv1 during DC24

NOTED:
- In the first part it was dry-ran for all the LHCOPN and LHCONE connections at CERN
- During the last 3 days it ran with real actions on the production LHCOPN links for DE-KIT, CA-TRIUMF and ES-PIC
- It worked well, but there was no improvements on transfers because links were not congested

BBR:
BBRv1 was periodically configured and removed on 40 EOS servers at CERN. No benefits nor penalties identified, but again the links were not congested
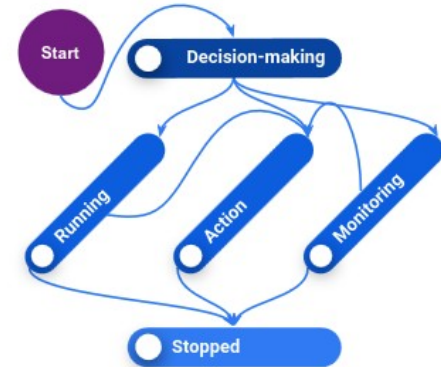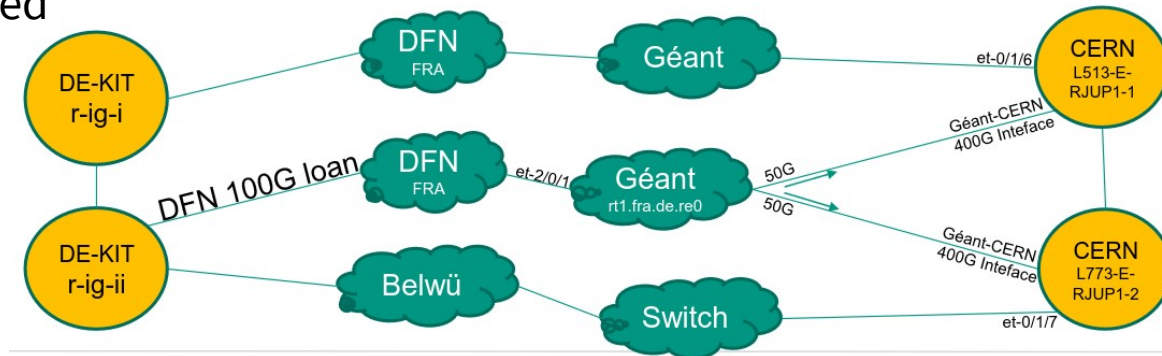
DE-KIT 3x100G links for DC24

During DC24, DE-KIT deployed a third 100Gbps link to avoid congestion on the 2x100Gbps link

The link was provided over the DFN, GEANT and CERN networks

It proved to be useful because in several occasions during DC24 the bandwidth between CERN and DE-KIT exceeded 200Gbps

The third link has now been removed
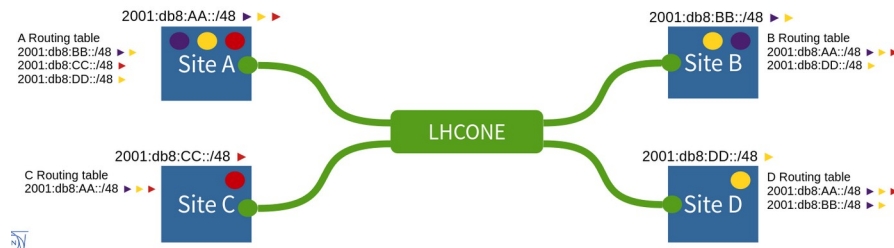
MultiONE with BGP communities

New proposal for MultiONE implementation
- Keep a single VPN (LHCONE)
- Tag each prefix announced to LHCONE with BGP communities that identify the collaborations served by the site
- Eventually accept only interesting prefixes

This proposal is less operationally complex then the previous one, since it use a common technique already used by RENs and LHCONE itself.

Mandate to move to implementation. Aiming to complete the tagging in one year from now. Filtering may start during LS3

[Activity started]

https://indico.cern.ch/event/1349135/contributions/5765335/attachments/2834291/4952642/LHCOPNE-20240410-52-multiONE-with-BGP-communities.pdf

Jumbo frames survey

Presented results of Jumbo Frame survey

53 answers received during ~1 month
- 26 sites have Jumbo on the storage (50%)
- 19 sites of the 26 have Jumbo also on worker nodes (40%)
- 4 sites of the 26 didn't have any problem with implementing Jumbo (20%)
- 10 sites of the 20 without Jumbo don't want to implement Jumbo (~20% of total)

https://indico.cern.ch/event/1349135/contributions/5765326/attachments/2834292/4952645/LHCOPNE-20240410-52-Jumbo-Frame-survey.pdf

BBRv3 preliminary results

Results of BBRv3 testing run by ESnet

BBRv3 is still not part of the mainline Linux Kernel

BBR helps on some paths
– Only minor improvements over CUBIC on clean paths with large buffered devices
– Need to find paths will small buffered devices to really see improvements

Detailed analysis of pacing behaviour is difficult: more work needed

https://indico.cern.ch/event/1349135/contributions/5765324/attachments/2838863/4961604/BBRv3%20preliminary%20testing%20results%20-%20LHCONE%20April%202024.pdf

High Touch new features

ESnet discussed some findings from the packet by packet analyses of the DC24 traffic, using the data produce by the ESnet High Touch service.

- Almost no sign of retransmission. That indicates no congestion nor problematic links. And makes the need for BBR and other congestion avoidance measures less stringent

- The average flow travelled at less than 1Gbps. There is room for improvement

- CERN will use this data to perform some analyses to understand if there's anything that could be improved by using BBR and Jumbo frames [work in progress]

Jumbo and BBR: next steps

Agreed to continue the work on BBR and Jumbo frames.

BBRv3 requires more testing in congested situation

Jumbo frames may improve the efficiency of the file transfer by relieving the storage servers' CPU and allow higher throughput for FTS.

CERN will put more effort in setting up a pilot to have a group of EOS servers running with Jumbo frames and run tests with other sites

RNT-WG update

SciTags was largely tested during DC24:
- 80% of EOS CMS (production), UNL production storage
- Flow labeling functionality (fireflies)

Results:
- Confirmed the capability to propagate Scitags all the way to the storages (for both ATLAS and CMS)
- Sending fireflies (from XRootd, EOS storages)
- Collection and visualisation at ESnet collector

Implementation status

Propagation:
- Rucio supports Scitags from 32.4.0
- FTS/gfal2 support Scitags from 3.2.10/2.21.0

Storages:
- XRootD provides Scitags implementation (from 5.0+)
- EOS provides Scitags support from 5.2.19+
- Working on a project for production rollout at CERN (for WLCG)
- dCache prototype exists, roadmap for release pending
- Also working with StoRM and Pelican

Plans for LHC Networking NREs for SC24

Super Computing 2024 will be held in Atlanta, Georgia

NOTED-SENSE, SciTags, SENSE-Rucio demonstrations at SC23 went well

Aiming to a 1.2Tbps network with 400Gbps DTNs to be available for demos during SC24

Demo proposals should be submitted by June 2024 [SciTags demo accepted]

https://indico.cern.ch/event/1349135/contributions/5815690/attachments/2834452/4952942/Potential%20LHC%20Networking%20SC24%20NRE%20Demonstrations.pdf

LHC Workflow R&D/Testing using AutoGOLE/SENSE and FABRIC

FABRIC is a distributed testbed to prototype services which integrates network and computing

SENSE is a service that can provide end-to-end services which includes the network and the attached End Systems

One application is the SENSE and Rucio/FTS/XRootD project, which aims to develop an improved way to manage CMS transfers
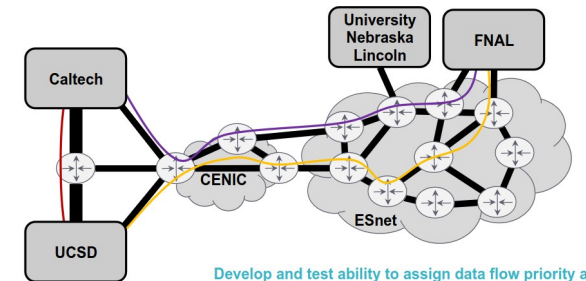
Develop and test ability to assign data flow priority and traffic engineer different end-to-end paths

The Global Research Platform

Presented the Global Research Platform

The GRP facilitates High Performance Data Gathering, Analytics, Transport (100 Gbps-Tbps E2E), Computing and Storage

Selected GRP themes:
- Orchestration Among Multiple Domains
- Large-Scale High Capacity Data WAN Transport (Highlighted at SC23: 400 Gbps, 800 Gbps, 1.2 Tbps WAN Services For Data Intensive Science)
- High-Fidelity Data Flow Monitoring, Visualization, Analytics, Diagnostic Algorithms, Event Correlation AI/ML/DL
- International testbeds for Data-Intensive Science

Conclusions

Summary

- LHCOPN: new Tier1s fully operational, 2.66 Tbps total bandwidth

- LHCONE: two new NRENs, more bandwidth over the Atlantic

- ESnet: deployment of 400Gbps in Europe and over the Atlantic progressing

- DC24: good performance of the network

- MultiONE: agreed solution with BGP communities.

- Data Transfer performances: Jumbo frames needs more testing

Actions

- Collaborate closer with SKAO

- Finalize MultiONE with BGP communities proposal and start tagging

- Increase effort in Jumbo frames testing

- Continue BBRv3 exploration

- Continue SciTags activities

Next meeting

Location: Beijing - CN
Host: IHEP
Dates: 9-10 October 2024

Agenda will be published here
https://indico.cern.ch/e/LHCOPNE53

References

Meeting agenda and presentations:
https://indico.cern.ch/e/lhcopne52

Questions?

edoardo.martelli@cern.ch