LHCOPN and LHCONE: preparing for HL-LHC

ISGC 2024 – Taipei TW - 28th March 2024

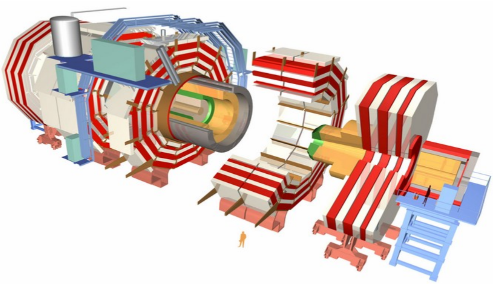
edoardo.martelli@cern.ch

Agenda

- WLCG, LHCOPN and LHCONE
- More bandwidth for HL-LHC
    - WLCG Data Challenges
    - Network R&D

WLCG and its networks: LHCOPN and LHCONE

Computing Model



Tier 0

Data source
Full data on Tape
Data reconstruction

Tier 1s

Distributed 2nd copy on Tape
Simulations, Data analyses

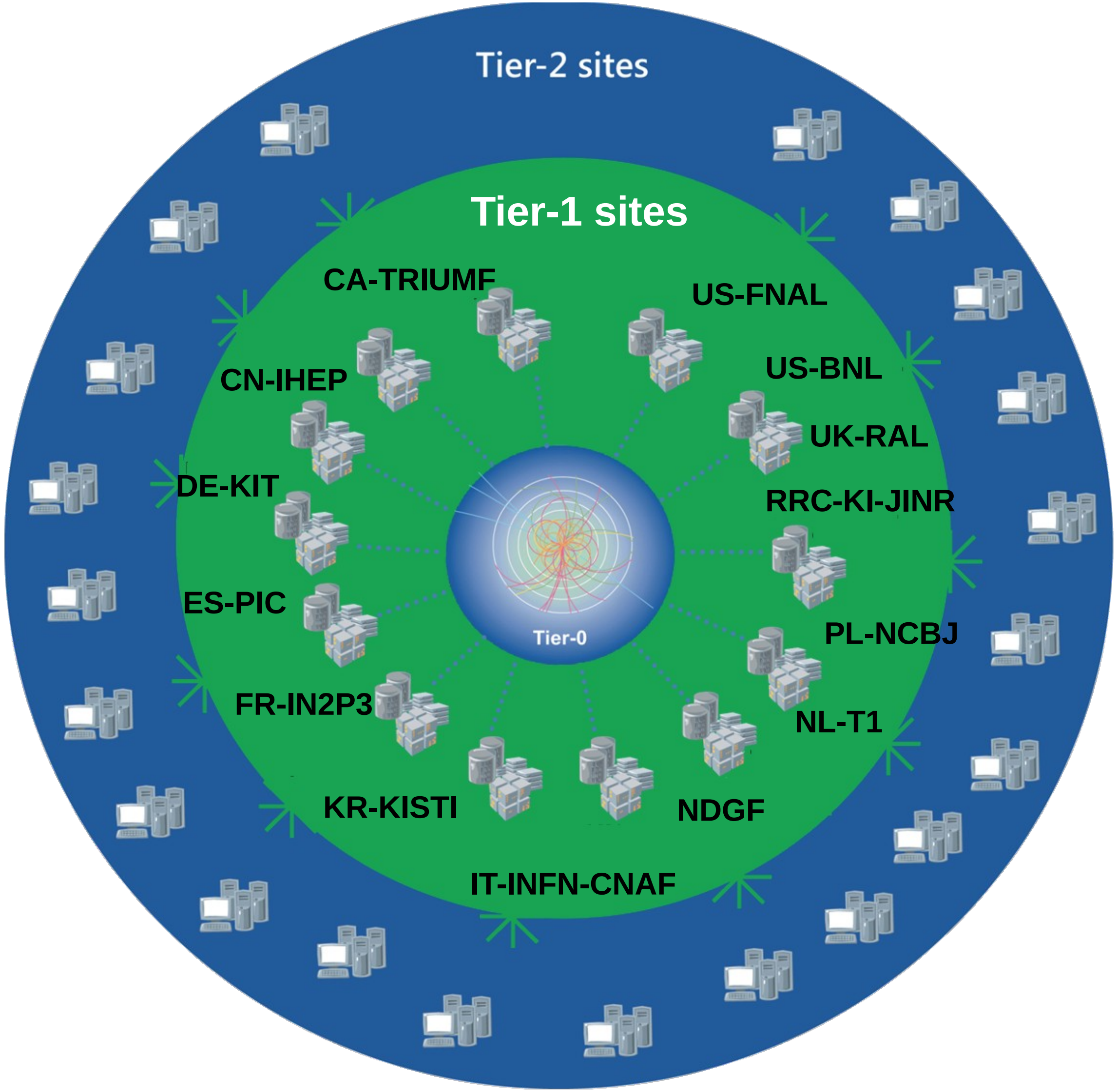
Tier 2s

Data caches
Data analyses

WLCG

The Worldwide LHC Computing Grid (WLCG) is a set of management and control applications that can use distributed computing and storage resources for the analyses of the LHC data

Computer Networks are an essential component of WLCG; they connect all the computing resources distributed in more than 150 institutes around the world



LHCOPN

Private network connecting Tier0 and Tier1s

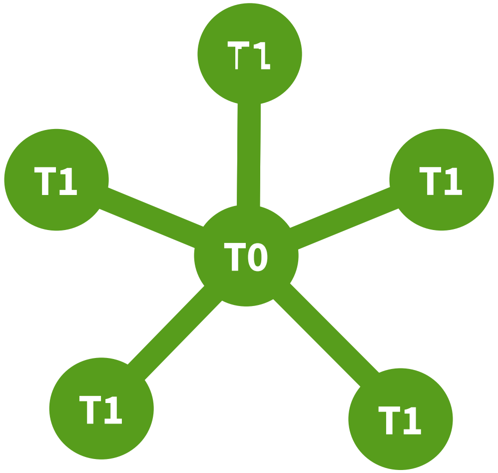
- Direct links from the Tier0 to all the Tier1s
- Dedicated to LHC data transfers

Secure:

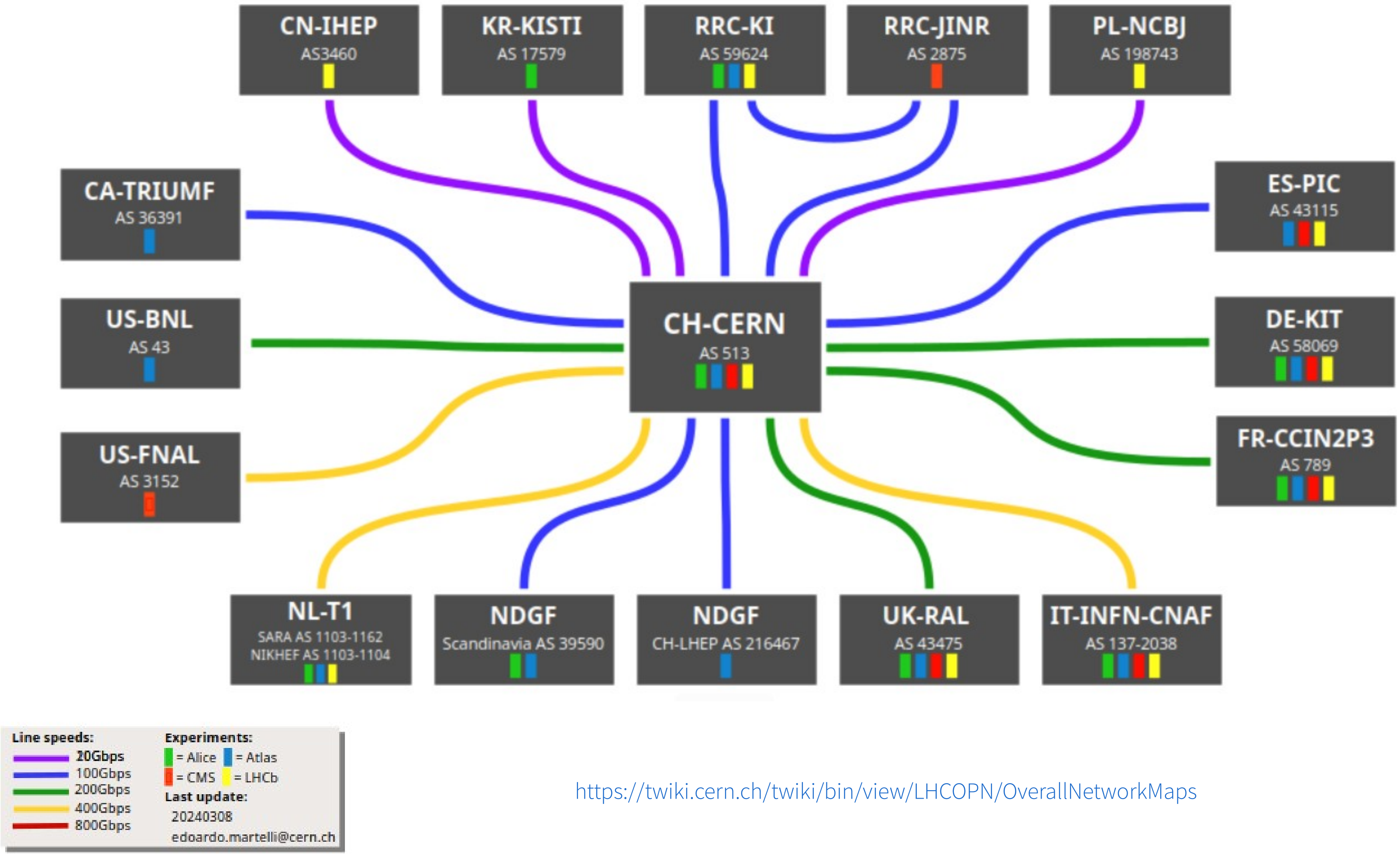
- Only declared IP prefixes can exchange traffic
- Can connect directly to Science-DMZ at sites, to bypass slow perimeter firewalls

Advanced routing:

- BGP communities for traffic engineering



LHCOPN



https://twiki.cern.ch/twiki/bin/view/LHCOPN/OverallNetworkMaps

Numbers

- 17 Tier1 sites and 1 Tier0
- 14 countries in 3 continents
- 2.66 Tbps to the Tier0

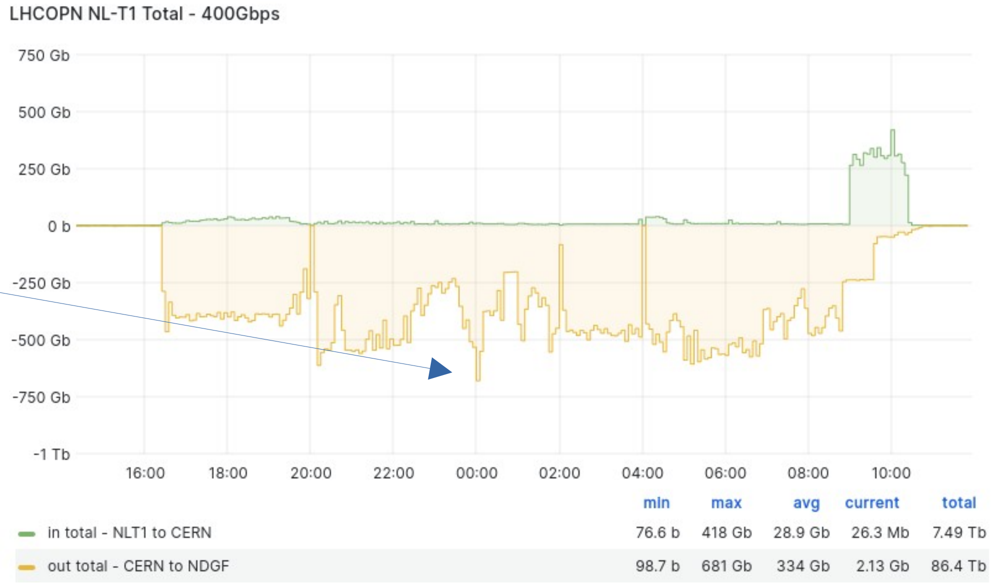
LHCOPN status

17 Tier1 sites connected

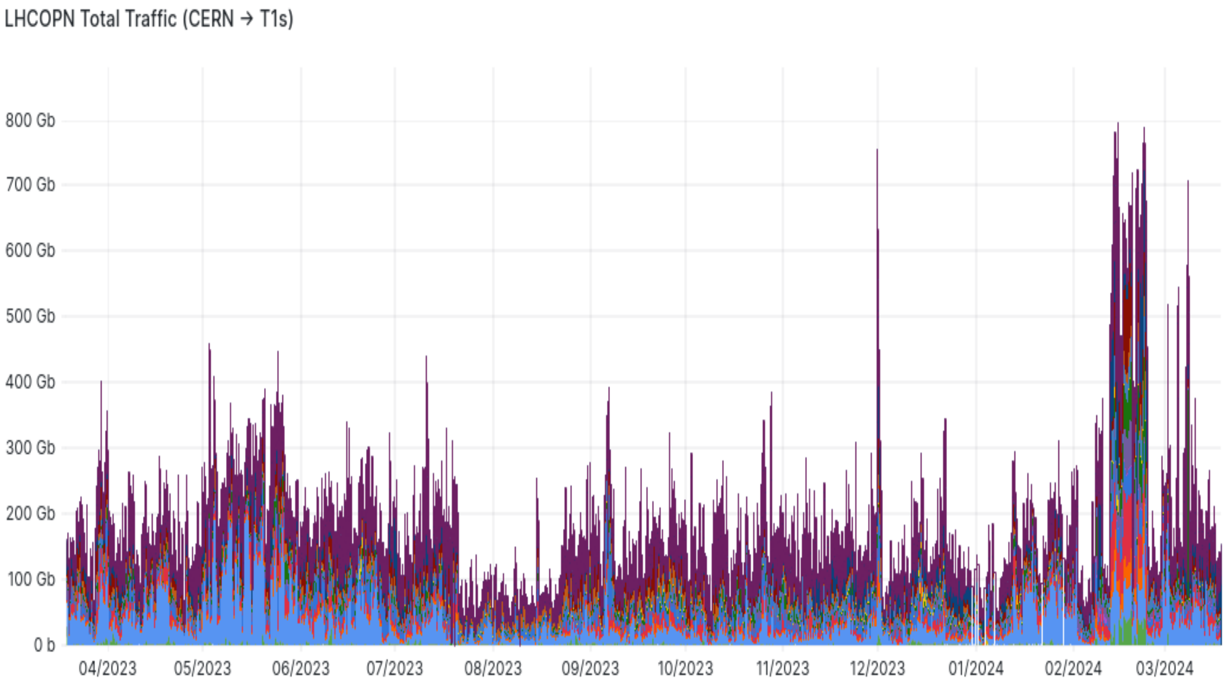
- Recently added CH-LHEP, CN-IHEP, PL-NCBJ

2.66 Tbps aggregated bandwidth

- 16x T0–T1 links
- Bandwidth from 20Gbps to 400Gbps
- Recently tested 800Gbps to NLT1 (now dismantled)



LHCOPN Traffic – last 12 months



Name	Mean	Max
Outgoing CA-TRIUMF	6.98 Gb	75.7 Gb
Outgoing CN-IHEP	15.2 Mb	21.6 Gb
Outgoing DE-KIT	51.5 Gb	203 Gb
Outgoing ES-PIC	5.96 Gb	96.6 Gb
Outgoing FR-IN2P3	14.3 Gb	169 Gb
Outgoing IT-INFN-CNAF	16.1 Gb	154 Gb
Outgoing KR-KISTI	290 Mb	18.6 Gb
Outgoing NDGF	8.19 Gb	110 Gb
Outgoing NL-T1	10.3 Gb	545 Gb
Outgoing-PL-NCBJ	690 Mb	18.9 Gb
Outgoing RU-T1	8.80 Gb	73.9 Gb
Outgoing UK-RAL	8.49 Gb	40.5 Gb
Outgoing US-BNL	14.6 Gb	148 Gb
Outgoing US-FNAL	11.2 Gb	204 Gb
Total	157 Gb	797 Gb

Numbers:

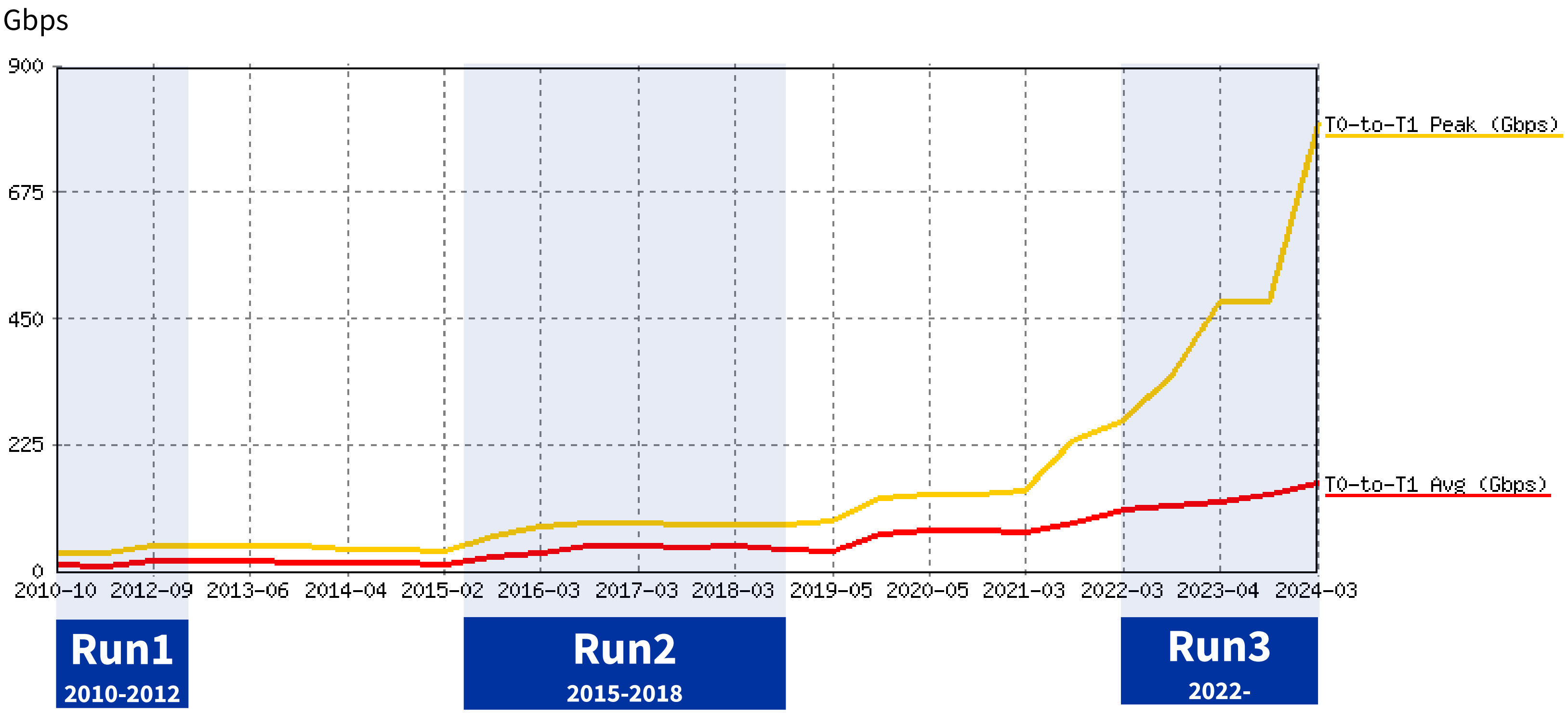
Moved ~619 PB in the last 12 months

+27% compared to previous year (488PB)

Peak at ~800Gbps (during DC24)

Ref: https://monit-grafana-open.cern.ch/d/HreVOyc7z/all-lhcopn-traffic?orgId=16

LHCOPN: traffic growth



Y-Axis: Gbps - Average bandwidth of previous 12 months

Run1: 2010-12
LS1: 2013-14
Run2: 2015-18
LS2: 2019-21
Run3: 2022-25

Ref: https://twiki.cern.ch/twiki/bin/view/LHCOPN/LhcopnStats

LHCONE L3VPN service

Private network connecting Tier1s and Tier2s

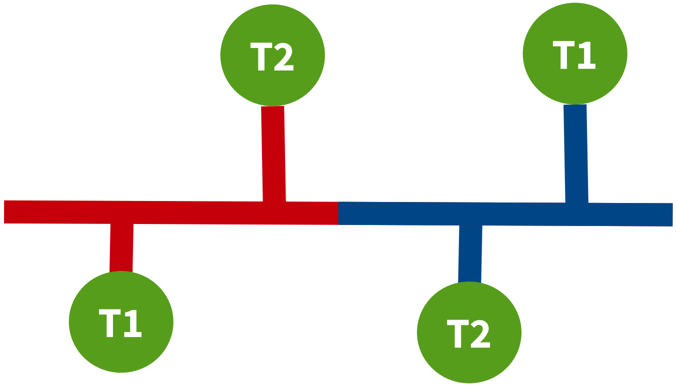
- Layer3 VPN implemented by National and International Research and Education Network operators
- Dedicated to LHC data transfers on a shared physical infrastructure

Secure:

- Only allowed sites can connect and exchange traffic
- LHCONE connections can be connect directly to Science-DMZ at sites, to bypass slow perimeter firewalls

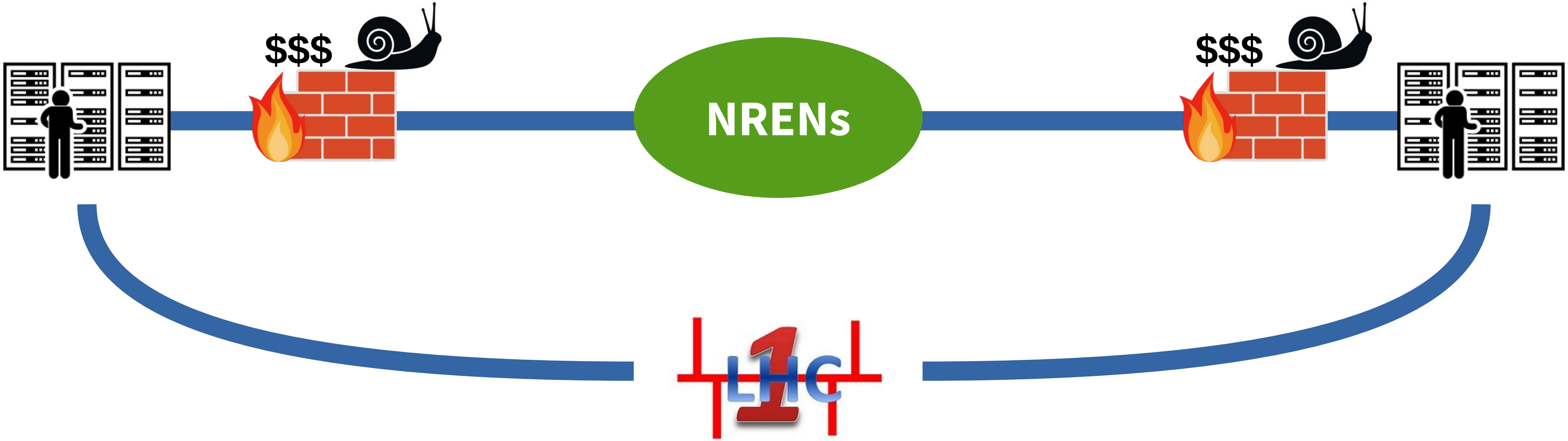
Advanced routing:

- Multi domain L3 VPN
- BGP communities for traffic engineering



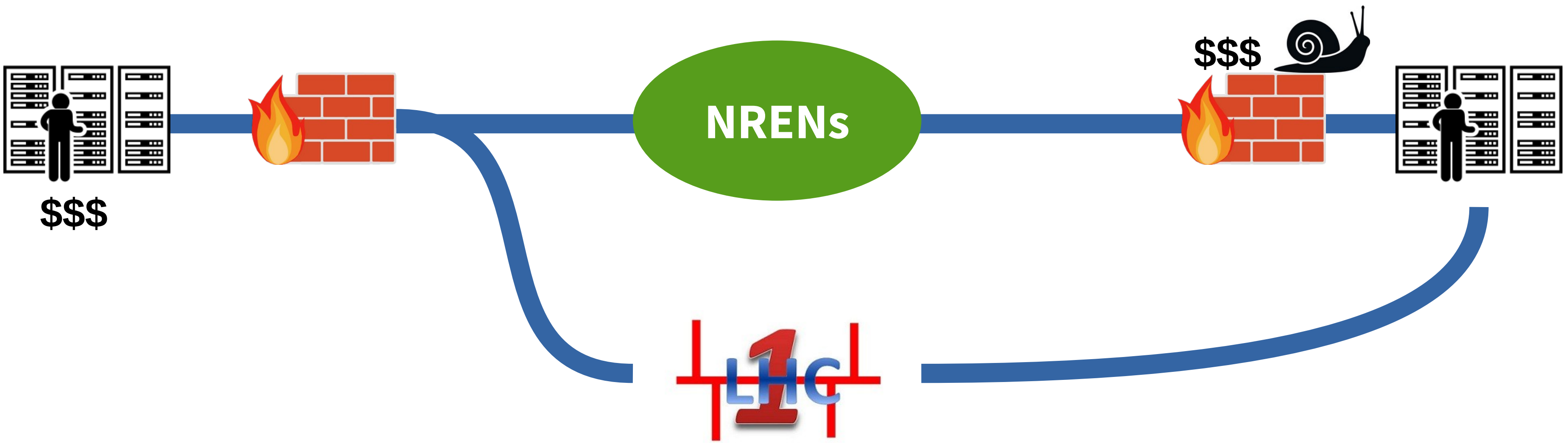
Why LHCONE is useful

LHCONE is a trusted network, more secure than a General Internet upstream. LHCONE can be connect directly to the data-centre and bypass slow perimeter firewalls (both at src and dst)

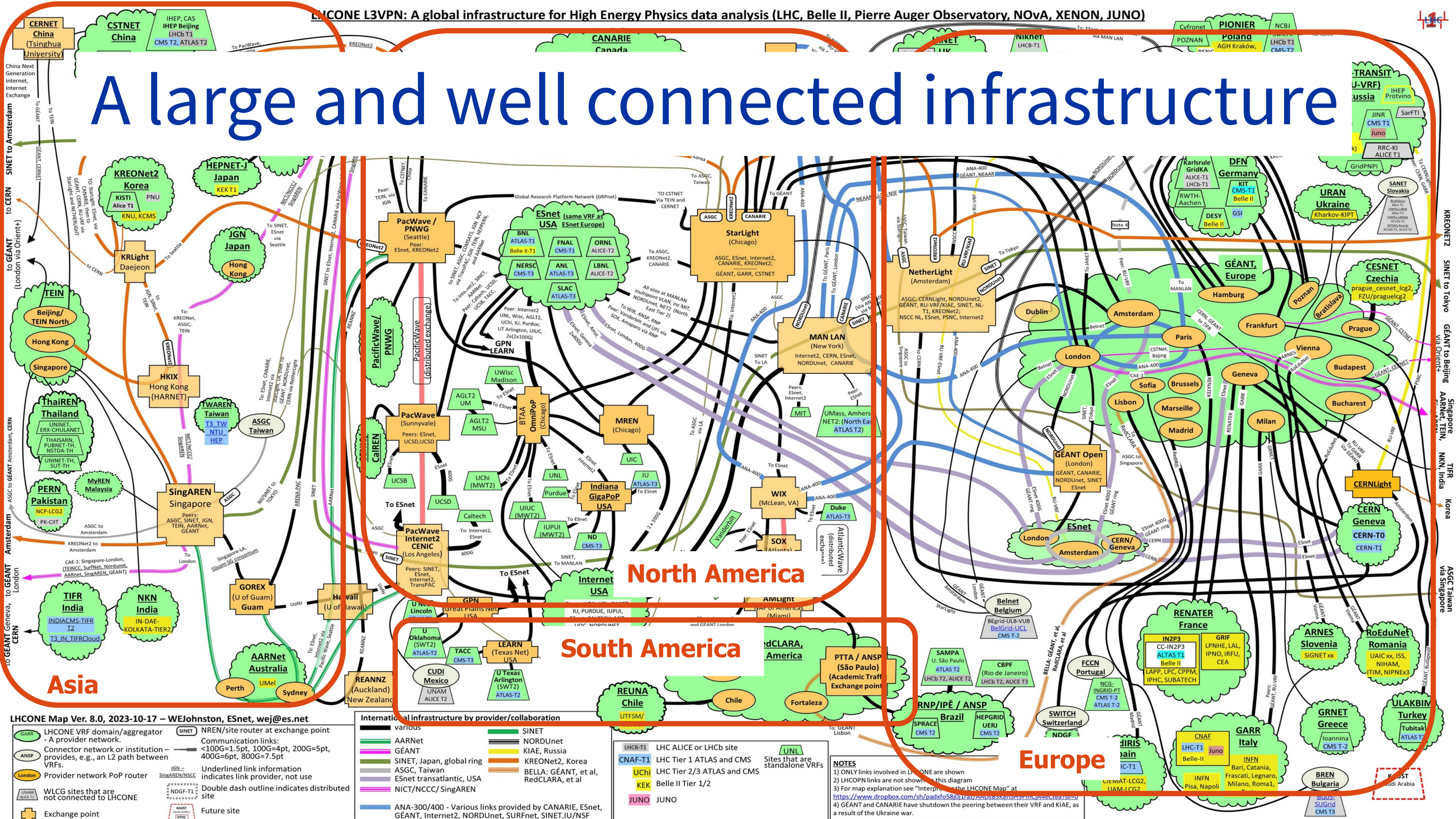


Why LHCONE is still useful

LHCONE is still useful even if don't want to bypass your Internet firewall, because remote sites may prefer to bypass their own ffirewalla



A large and well connected infrastructure



L3VPN status

- **VRFs: 30 national and international Research Networks**
- **Connected sites: ~110 in Europe, North and South America,Asia, Australia**
- Trans-Atlantic connectivity provided by ESnet, GEANT, Internet2, NORDUnet and SURF
- Trans-Pacific connectivity provided by KREOnet, SINET, TransPAC
- Interconnections at Open Exchange Points including NetherLight, StarLight, MANLAN, WIX, CERNlight, Hong Kong, Singapore and others

Open to other HEP collaborations

Belle II

Xe XENON Dark Matter Project

JUNO

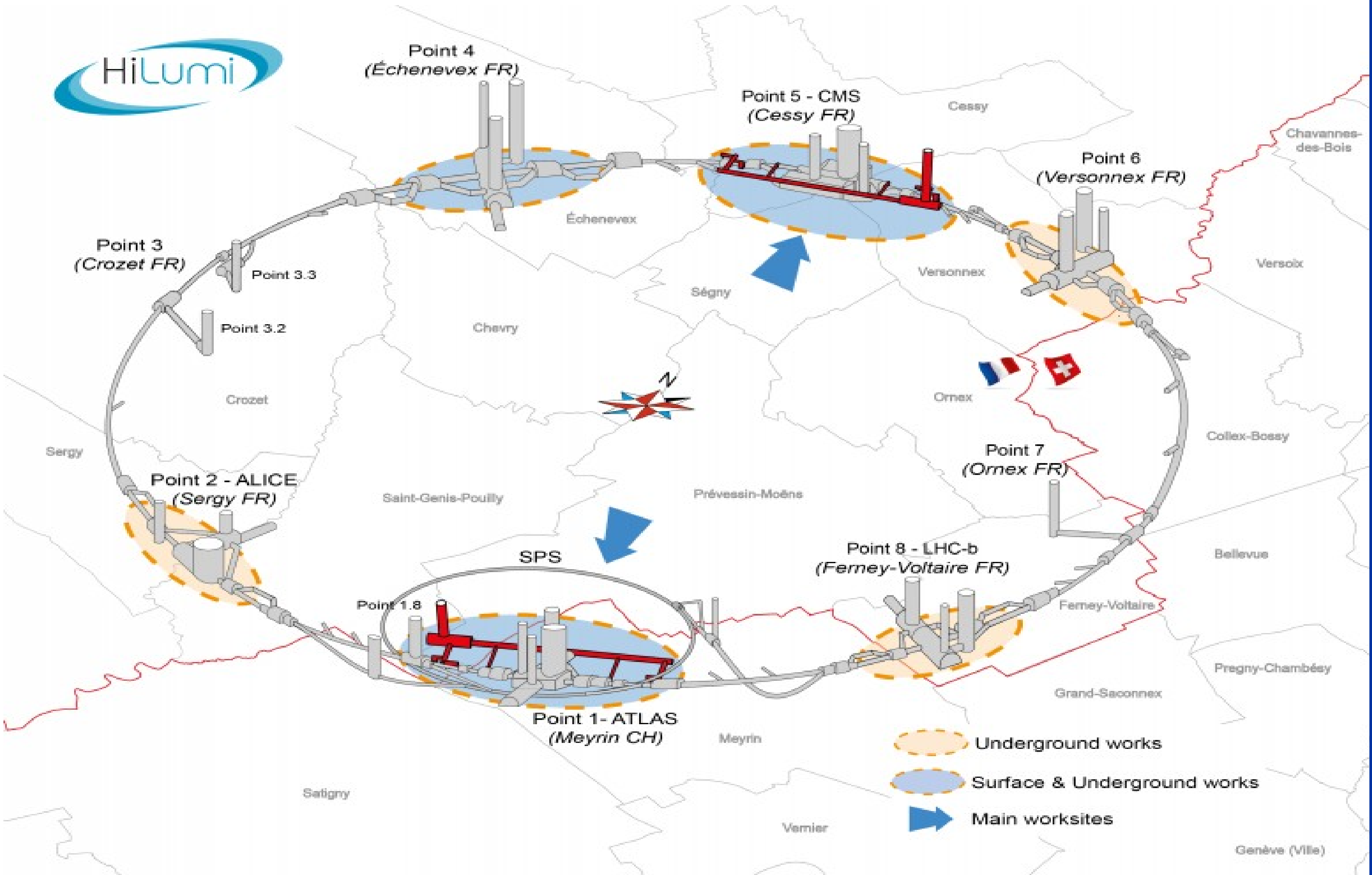
NOvA

PIERRE AUGER OBSERVATORY

DUNE DEEP UNDERGROUND NEUTRINO EXPERIMENT

HL-LHC requirements for WLCG

The HL-LHC project



The High-Luminosity Large Hadron Collider (HL-LHC) is an **upgraded version of the LHC**

It will operate at a higher luminosity or, in other words, it will be able to produce more data

The HL-LHC will enter service after 2029, **increasing the volume of data** analysed by the experiments **by a factor of 10**

HL-LHC network requirements

ATLAS & CMS T0 to T1 per experiment

- 350PB RAW, taken and distributed during typical LHC uptime of 7M seconds (3 months)
  - 50GB/s or 400Gbps
- Another 100Gbps estimated for prompt reconstruction data tiers (AOD, other derived output)
- estiimated 1Tbps for CMS and ATLAS summed

ALICE & LHCb T0 Export

- 100 Gbps per experiment estimated from Run-3 rates

Minimal Model

- Sum (ATLAS,ALICE,CMS,LHCb)*2(for bursts)*2(overprovisioning) = **4.8Tbps expected HL-LHC bandwidth**

Flexible Model

- Assumes reading of data from above for reprocessing/reconstruction in 3 month
- Means doubling the Minimal Model: **9.6Tbps expected HL-LHC bandwidth**

Overall requirements for HL-LHC

Major Tier1s:

1Tbps to the Tier0 (LHCOPN)

1 Tbps to the Tier2s (aggregated, LHCONE)

Major Tier2s:

>400 Gbps

Over provisioning main not always be an option. More efficient technology may be needed

WLCG guidelines

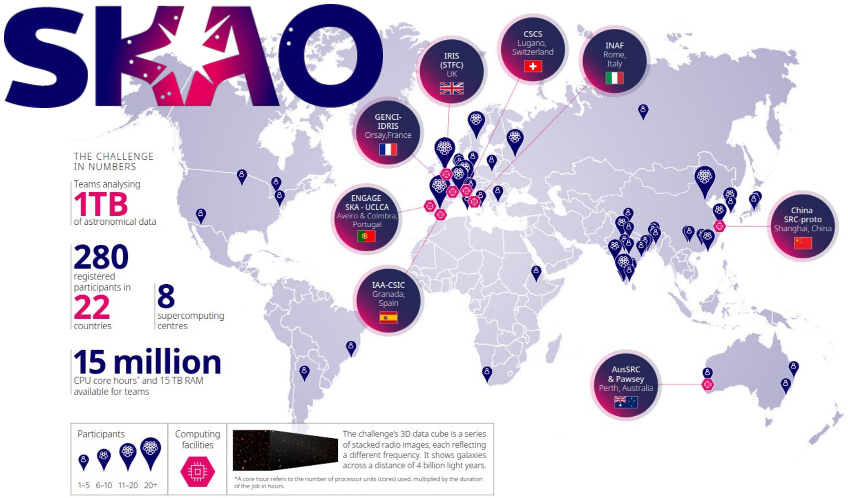
Worldwide LHC Computing Grid

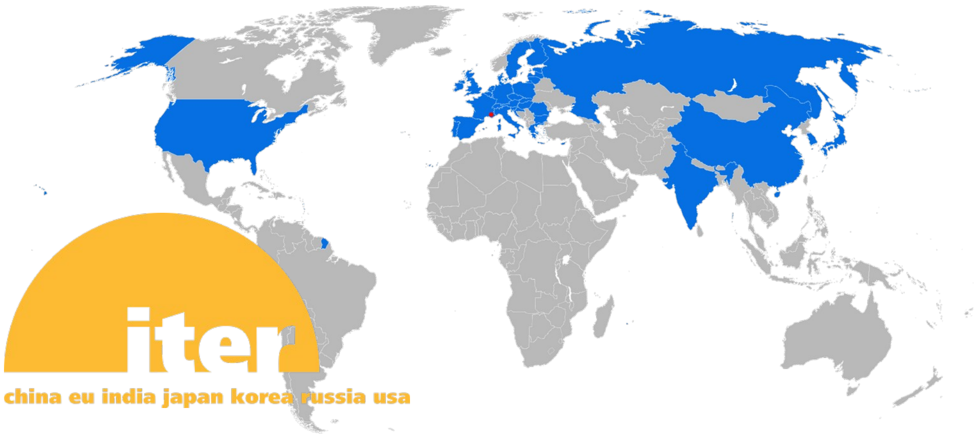
Message from Simone Campana:

In the next 10 years WLCG Networking will be faced with two major challenges:
- dealing with the HL-LHC data volumes and complexity
- the cohabitation with other experiments and sciences on the same infrastructure

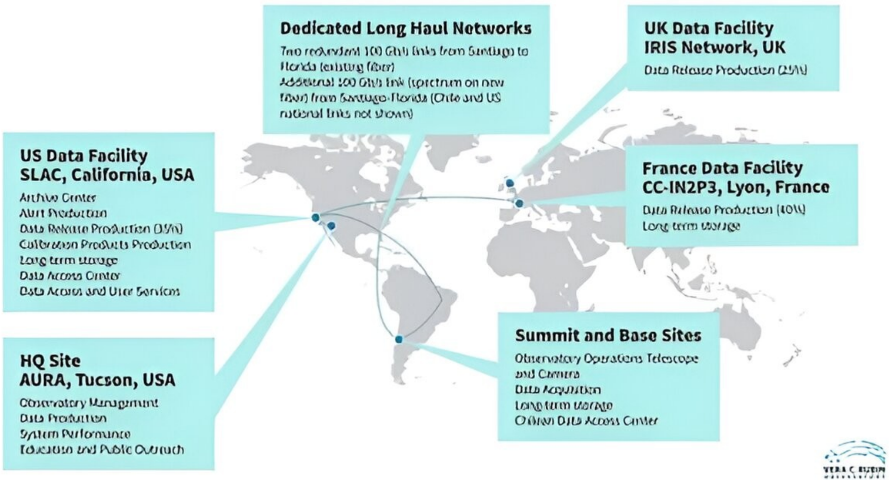
The WLCG network community can play a leading role:
- **modernize the network services**, progressing with the ongoing R&D activities and bringing early prototypes in production
- **engage with other experiments and sciences** to drive the evolution of R&E networks

More Big Data sciences coming on line





VERA C. RUBIN
OBSERVATORY



WLCG Data Challenges

Data Challenges for HL-LHC

WLCG has been mandated to execute data challenges for HL-LHC

- Demonstrate readiness for expected HL-LHC data rates
- Increasing volume/rates
- Increase complexity (e.g. additional technology)
- A data challenge roughly every two years

DOMA is the coordination and execution platform

- Agreements across the LHC experiments and beyond
    - Suited dates
    - Reasonable targets
    - Functionalities
- Help in orchestration

Dates and high level goals always approved by WLCG MB

Plan for Data Challenges

2021: 10% of HL-LHC requirements

2024: 25% of HL-LHC requirements

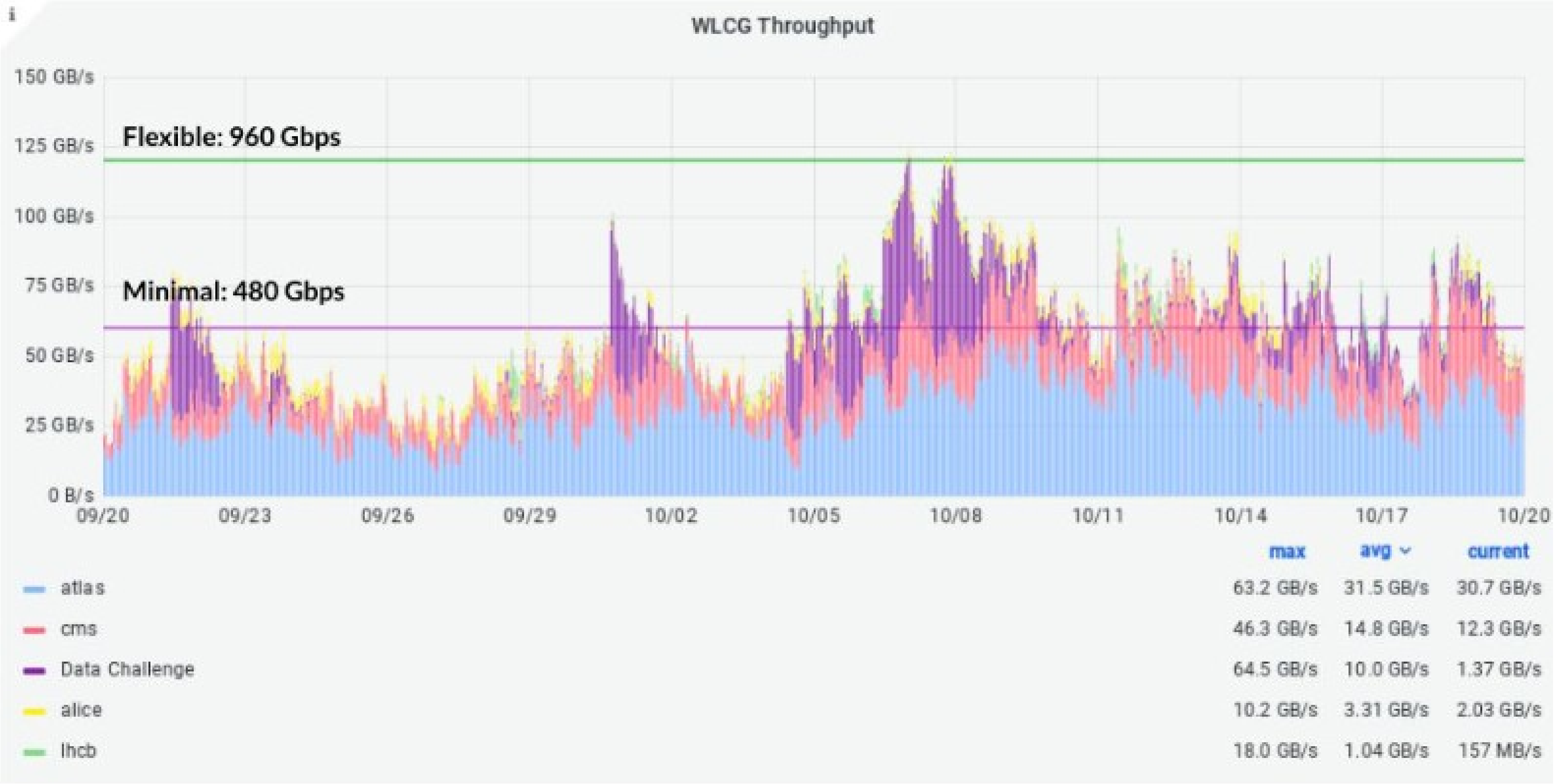
2026: 50% of HL-LHC requirements (date and % to be confirmed)

2028: 100% of HL-LHC requirements (date and % to be confirmed)

2029: start of HL-LHC (Run4)

DC21

Managed to fill 100% of the DC21 (10%) target!



DC24

DC24 just completed

- Ran 12-23 February 2024
- Target of 25% of HL-LHC requirements
- Several network projects tested during the last days

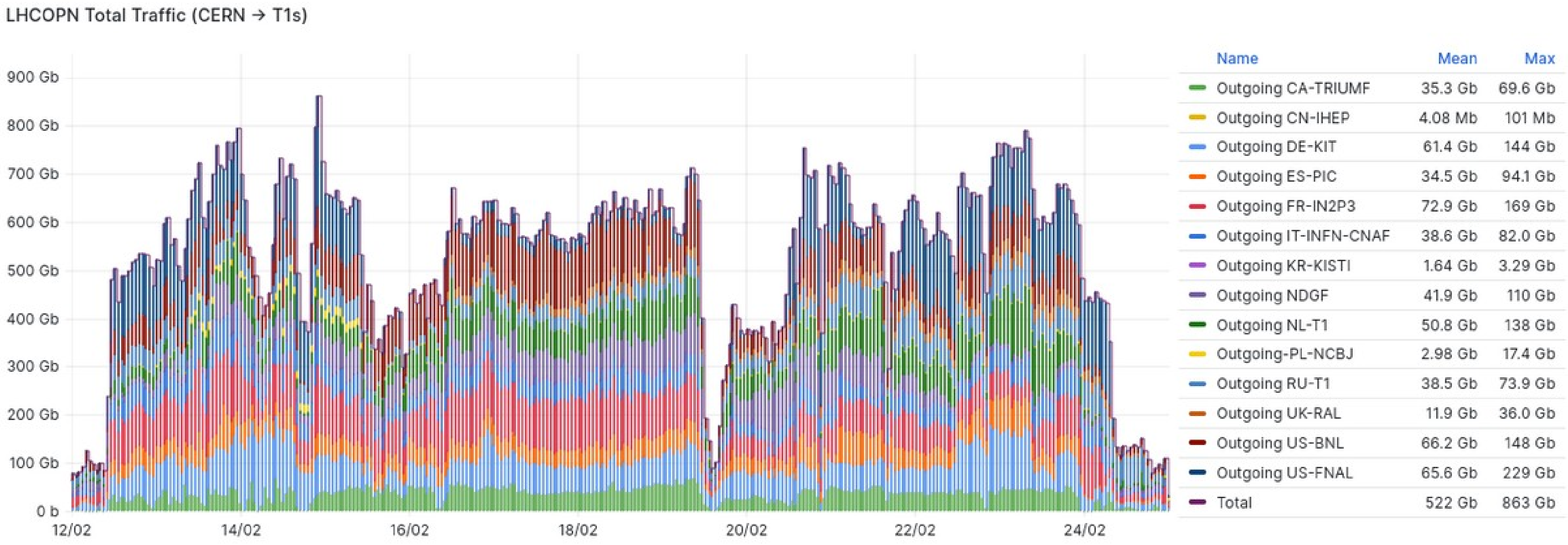
DC24 network projects

List of the projects on networking:
- Packet marking
- Packet pacing, BBR performances
- perfSONAR for network alarms and debugging
- Site Network monitoring of in/out bandwidth
- NOTED: FTS driven SDN

DC24 results

Tier0-Tier1 traffic on LHCOPN: peak at 800Gbps

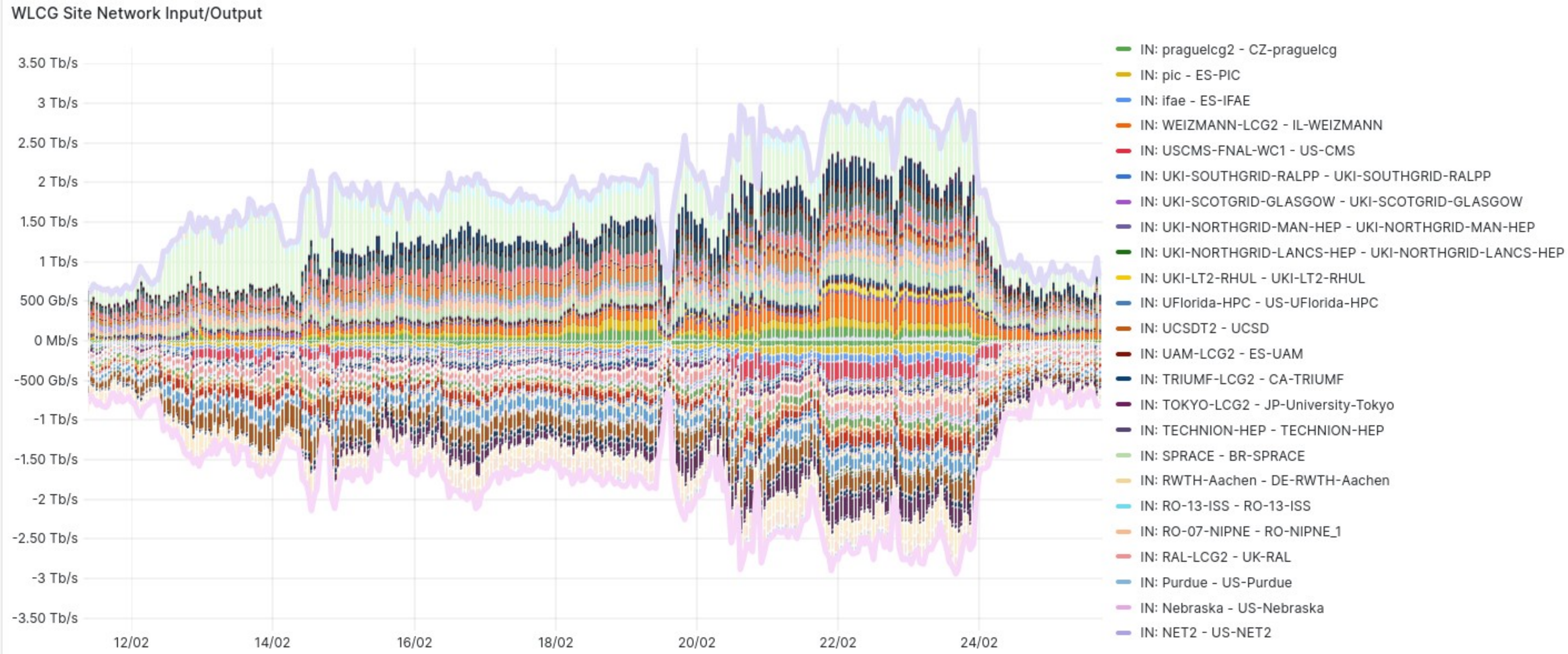
- 36% of existing bandwidth, but just ~15% of the estimated HL-LHC bandwidth



https://monit-grafana-open.cern.ch/d/HreVOyc7z/all-lhcopn-traffic?orgId=16&var-source=long_term&var-bin=1h&from=1707650640915&to=1708873382950

DC24 results

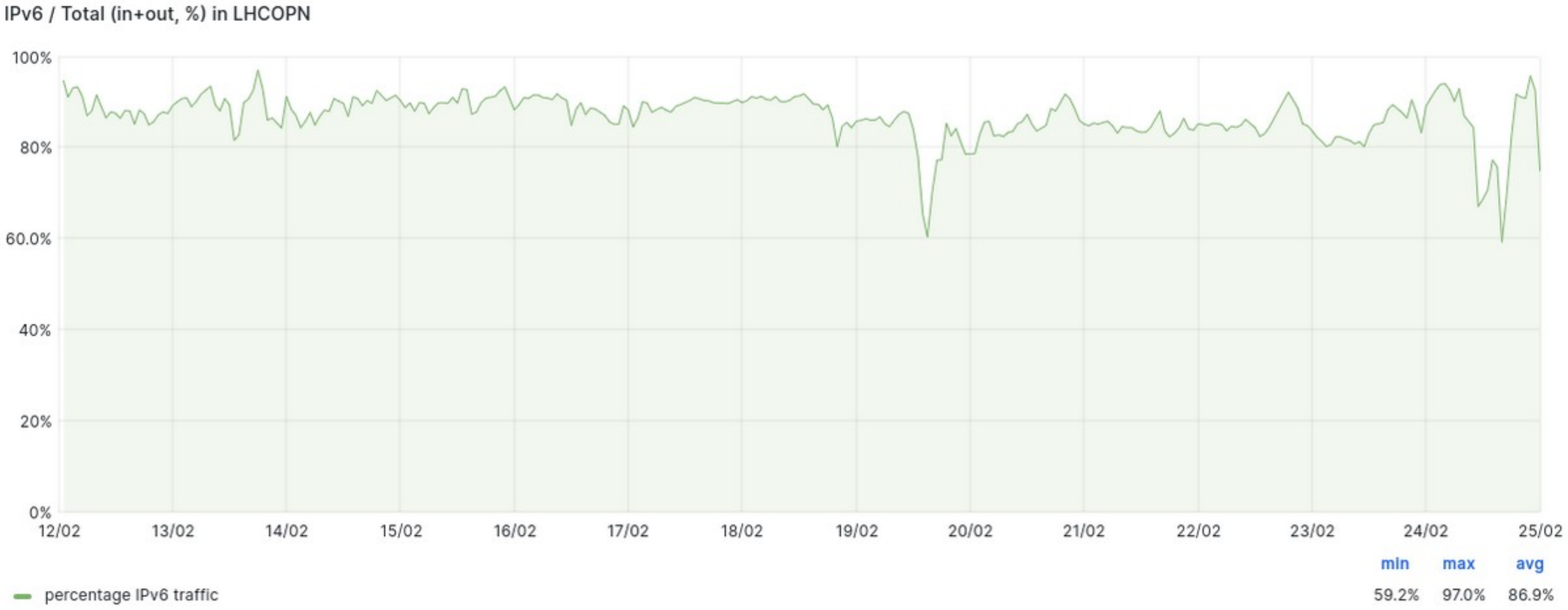
WLCG aggregated traffic exceeded 3Tbps



https://monit-grafana-open.cern.ch/d/MwuxgogIk/wlcg-site-network?from=1707638857216&orgId=16&to=1708880381142

DC24 results

IPv6 Traffic in LHCOPN: 86.9% of the Total



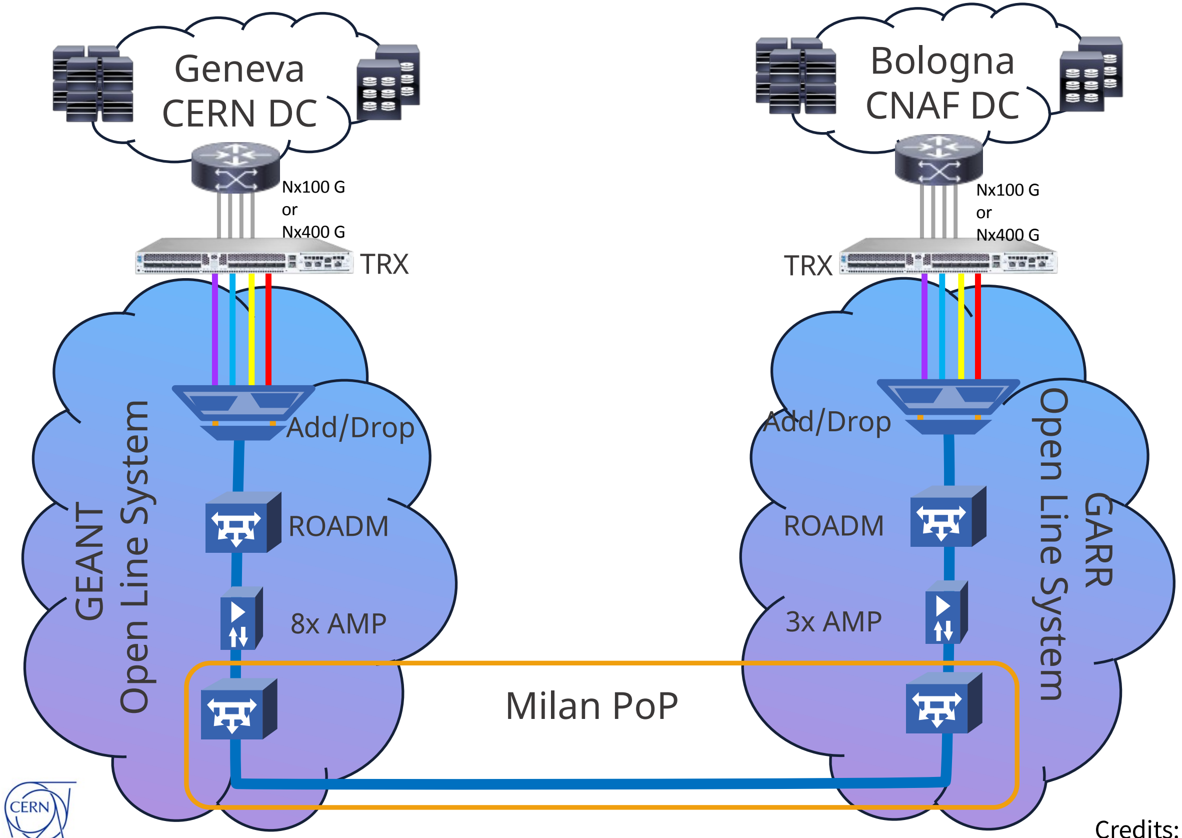
More on DC24

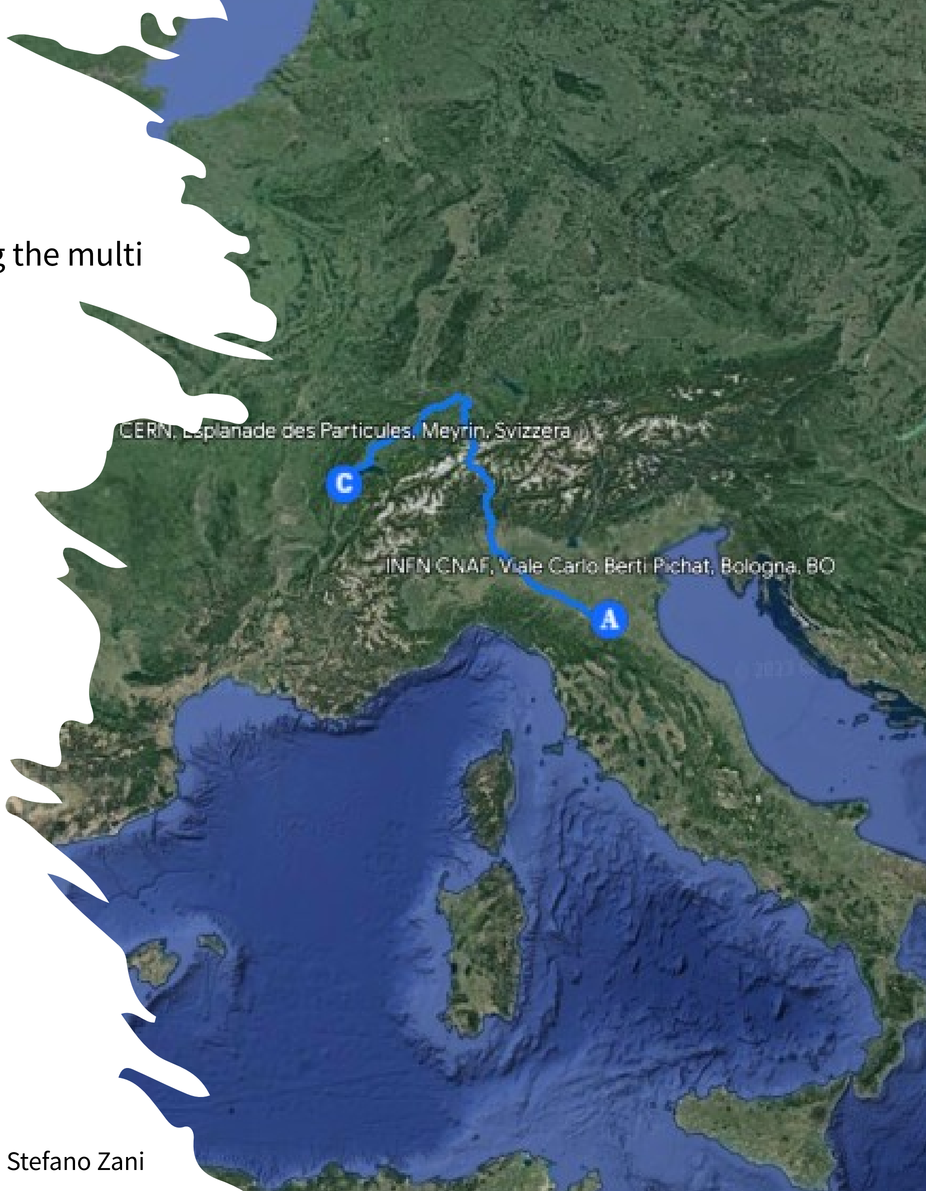
More details about DC 24 network projects in Carmen Misa’s talk

DC24 R&D network projects

CNAF-CERN DCI

Proposed in GEANT GN4-3 (WP7-T2) as a possible use case for experimenting the multi domain **S**pectrum **C**onnection **S**ervice at about 1000 km of distance.

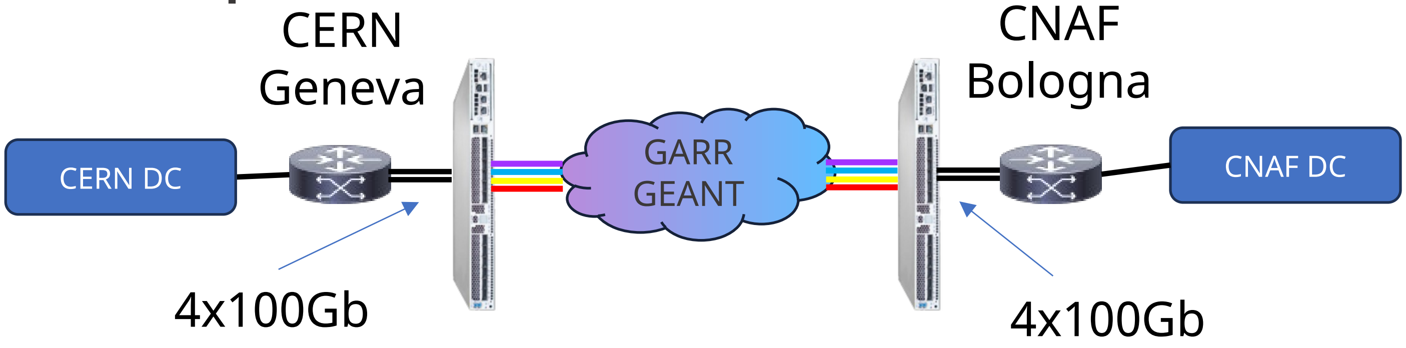




Credits: Stefano Zani

Current state

On 1000 km distance using **69 Gbaud DP-16 QAM** modulation, it is possible to reach **400Gbps per carrier**
Having 4 line ports on transpoders, it is possible to reach 1.6 Tbps on this «Circuit» that could be used as up to 4x400Gb Ethernet or 16x100Gb Ethernet .
The spectrum occupation for each carrier is 100Ghz, so 400Ghz (10% of the C band) are suffcient to transport **1.6 Tbps with 9.5 ms of Round Trip Time (Standard routed path is about 13.5ms)**



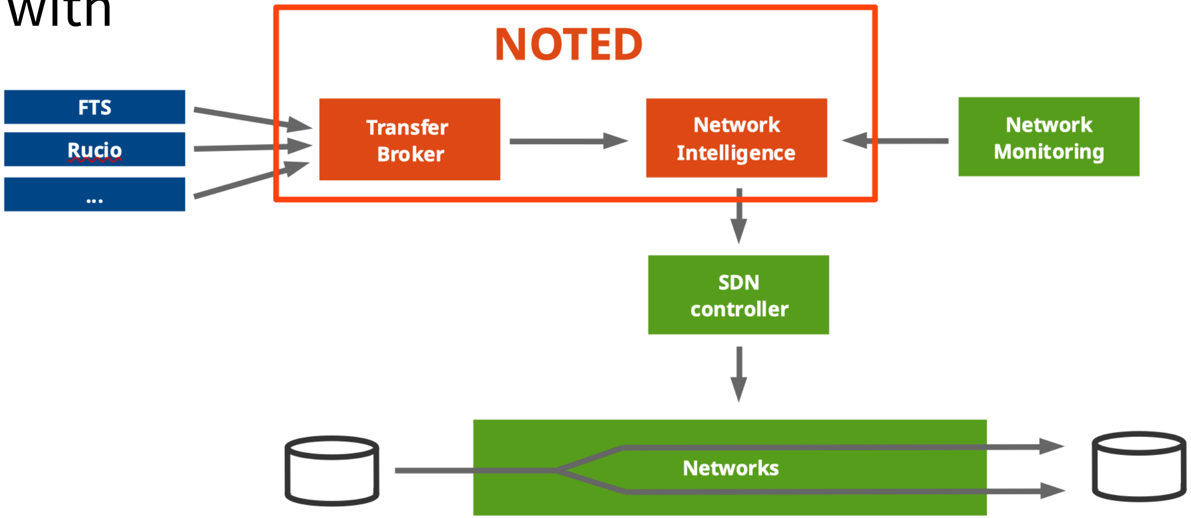
In production has main LHCOPN link for IT-INFN-CNAF. Used during DC24

NOTED SDN

NOTED is a framework that can detect large FTS data transfers and trigger network optimization actions to speed up the execution of the transfers

Already tested with production transfers:

- CERN-PIC with LHCOPN-LHCONE load balancing
- CERN-TRIUMF and KIT-TRIUMF with the activation of dynamic circuits
- New version with triggers from Network Monitoring tested during DC24



More information

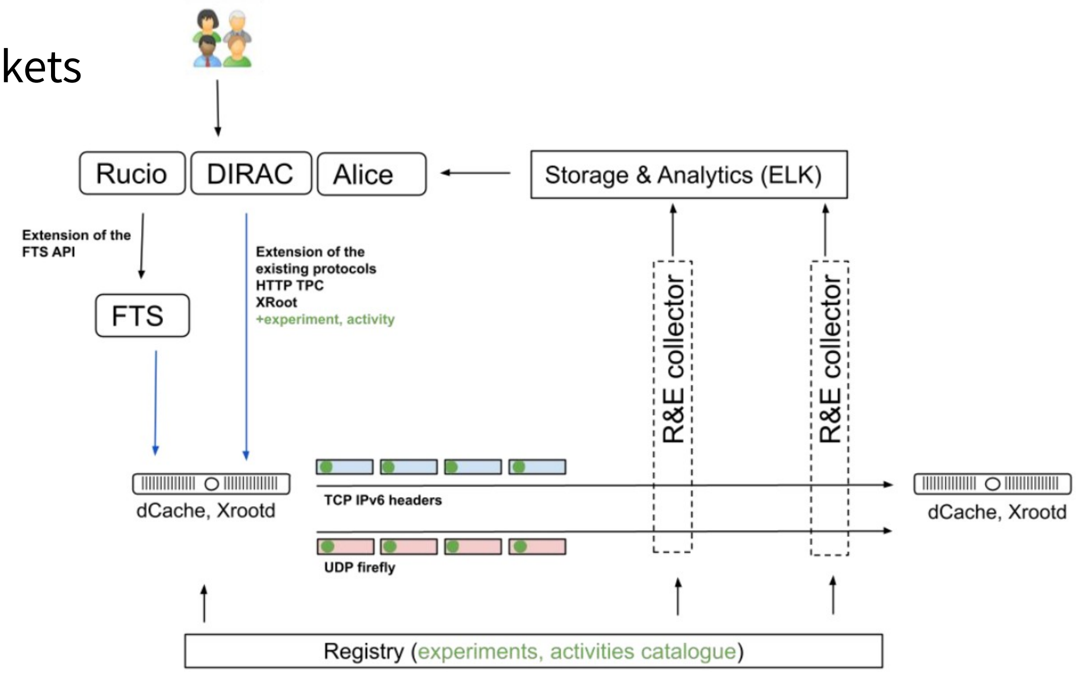
Follow Carmen Misa's talks for more information on NOTED

Packet Marking

Marking of data packets/flows with Experiments and Applications IDs for better accounting

Two options being investigated:

- Tag in the IPv6 flowlabel field
- Tag (and more) in UDP fireflies (UDP packets sent in parallel to each flow)



Packet Pacing

A small amount of packet loss makes a huge difference in TCP performance, especially on long distance flows

TCP can send packets in burst. These burst can be a problem in case of:

- Shallow switch buffers
- Slower receivers
- Speed mismatch on the path

Goal of pacing is to limit the burst rate of a TCP flow

BBR TCP congestion protocol has built-in pacing (transmit based on a clock, not ACKs)

More on DC24 network projects

More details about DC 24 network projects in Carmen Misa’s talk

R&D: MultiONE

Other collaborations and sciences

DUNE (Neutrino Experiment) has officially joined LHCONE

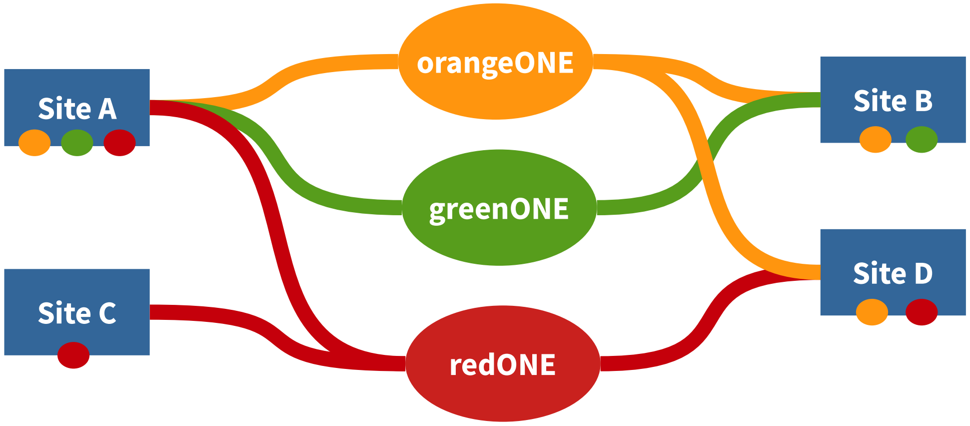
Contact established with other big science projects which in the future may compete with WLCG on network utilization

There's a growing need for Big Science collaborations to coordinate their requirements to allow an organic grow of the R&E networks

multiONE

LHCONE already very large, it could become risky to include other large science projects
Better to implement multiple VPNs, one for each collaboration:
- Each site joins only the VPNs it is collaborating with, to reduce the exposure of their data-centre

But it’s difficult to separate the traffic for sites member of multiple collaborations.



New proposal using BGP communities

Don’t add any additional VPN (or maybe just one for Other Big Sciences)

Each prefix announced to LHCONE is tagged with BGP communities that identify the collaborations served by the site

The tagging is done by the sites, or by the connecting REN if they can’t do it

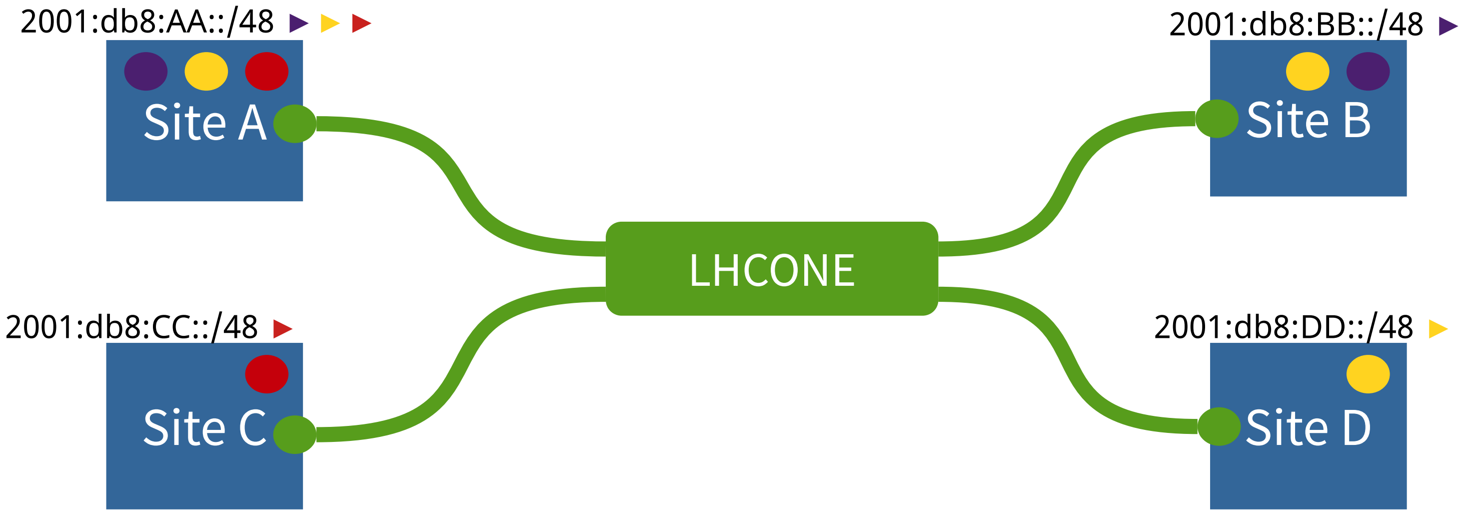
Sites can/should then decide to accept only the prefixes of the collaboration they are working with

In addition/alternative, RENs could announce to a given site only the prefixes of the collaborations related to the site

Practical example

Each site tags its prefixes announced to LHCONE with the BGP communities that identify the collaborations the site is participating in

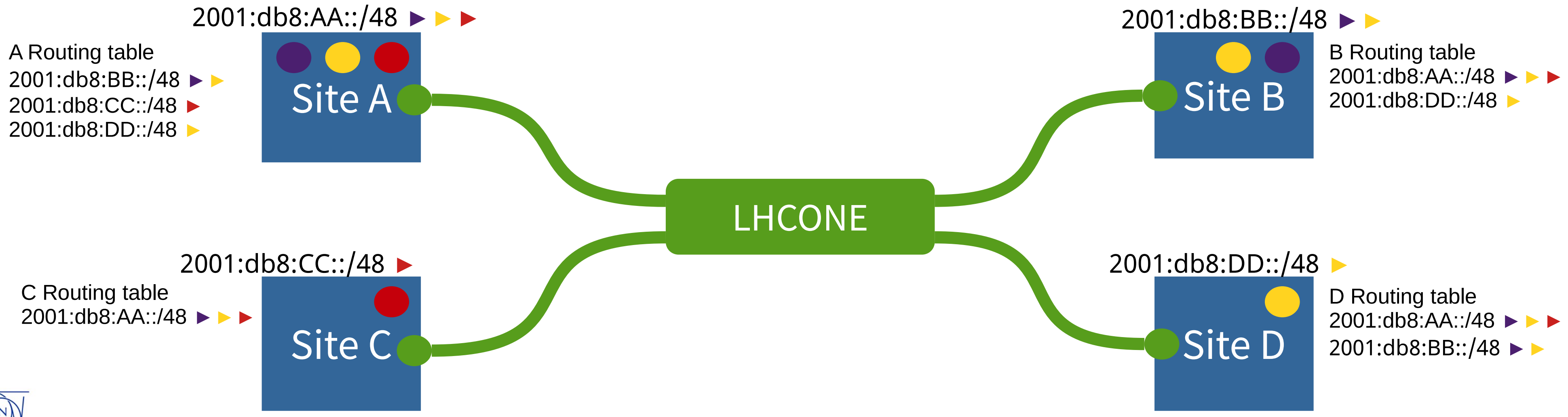
The tagging can be done by the sites or the connecting NREN



Practical example

Each site accepts only the prefixes tagged with the BGP communities of its own collaborations

The filtering can be done by the sites or the connecting NREN



Benefits

- Reduced exposures of sites
- No additional VPNs to configure
- No changes at sites when a new site connects to LHCONE
    - only when a new collaboration joins, if they are interested in it
- Communication Errors will be an incentive to adhere
- Communication Errors will highlight already existing implementation errors and weaknesses

Implementation proposal

Plan to be discussed at the next LHCONE meeting

- Define BGP communities for the different LHCONE collaborations

- Implement prefix tagging at sites
  - if sites can’t do it, RENs will do for them

- When all prefixes are tagged, gradually implement filtering at sites

Conclusions

Summary

- HL-LHC will increase data production of a factor of 10. Networks will have to increase capacity, but also make a more efficient use of available bandwidth

- LHCOPN and LHCONE are continuously evolving to be ready to support HL-LHC

- WLCG Data Challenges are helping software and networks to reach the HL-LHC requirements

- On-going Network R&D activities in preparation to HL-LHC:
    - NOTED: SDN to improved FTS transfers
    - Scitags: packet marking for better accounting
    - Packet pacing: better transfer performances
    - MultiONE: keeps LHCONE secure

Questions?

edoardo.martelli@cern.ch