LHCOPN-LHCONE update

5th Asia Tier Centre Forum
TIFR – Mumbai, India

25th October 2019

edoardo.martelli@cern.ch

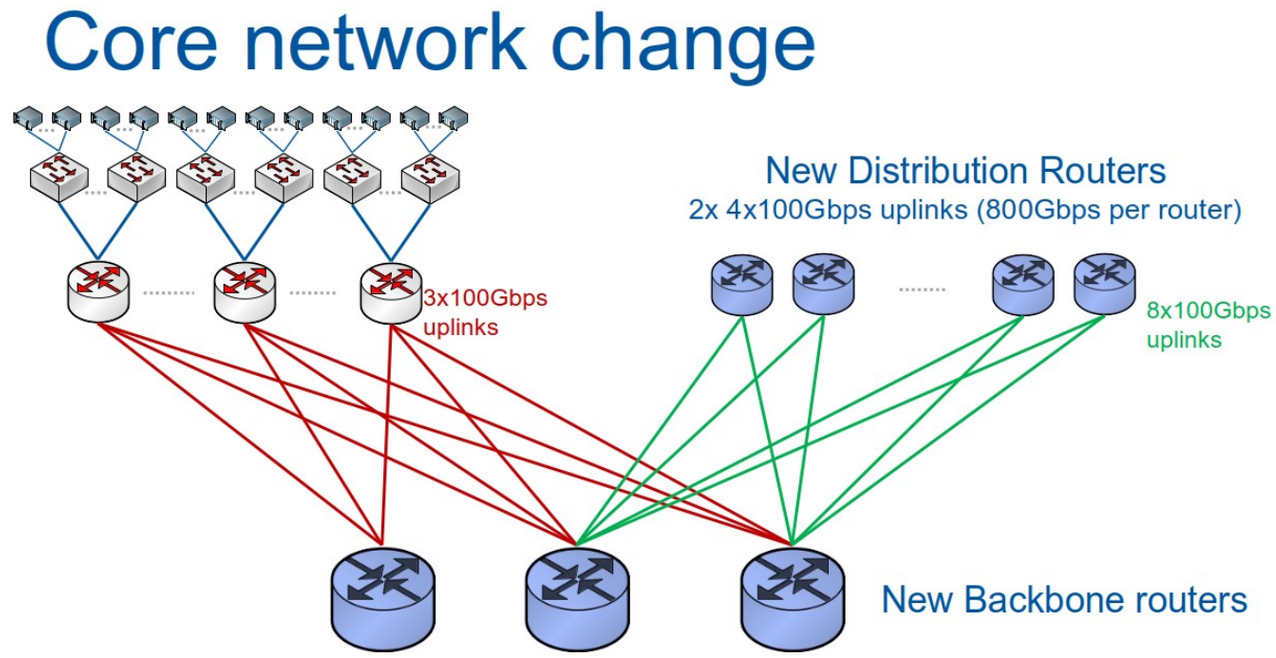
Contents

- CERN Tier-0
- LHCOPN
- LHCONE
- Networking R&D

CERN Tier-0 update

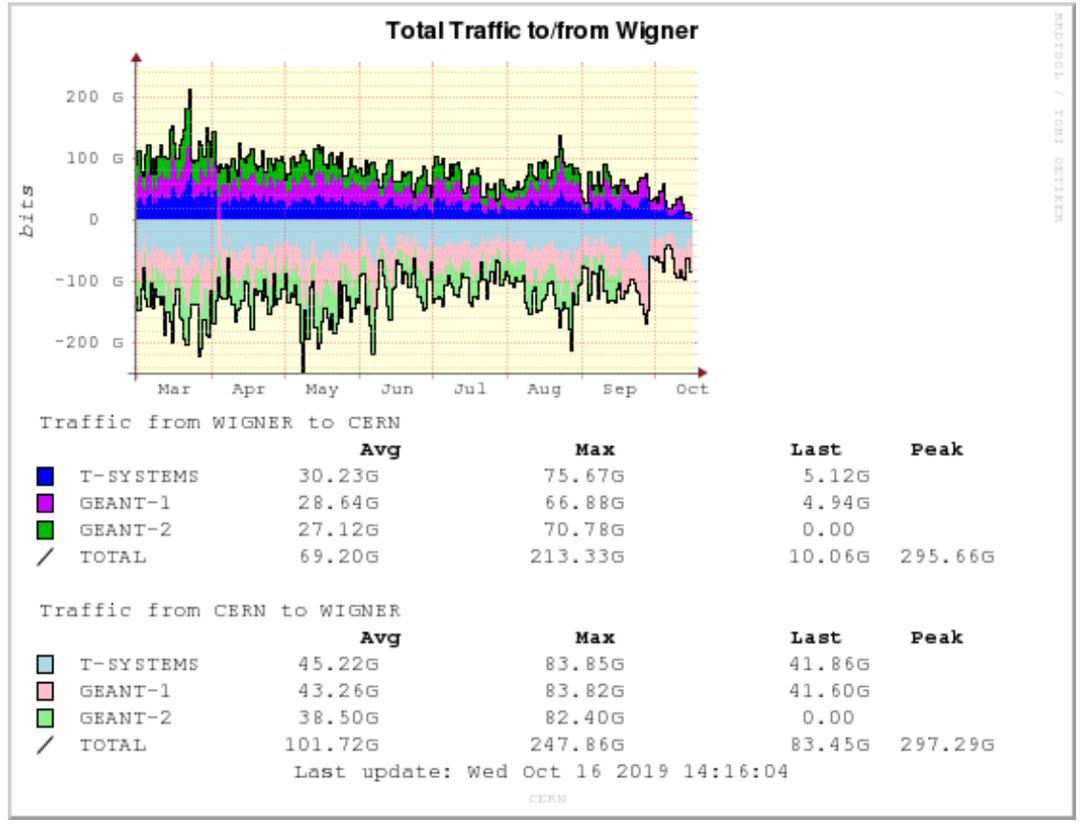
CERN data-centre B513

- Brocade MLXE being replaced by Juniper QFX10008
- Increasing new routers' interconnections to 800 Gbps with possibility to grow to 1.6Tbps
- Deploying new architecture with router redundancy using VXLAN ESI (Ethernet Segment Identifier)
- Testing Openstack integration for IP mobility



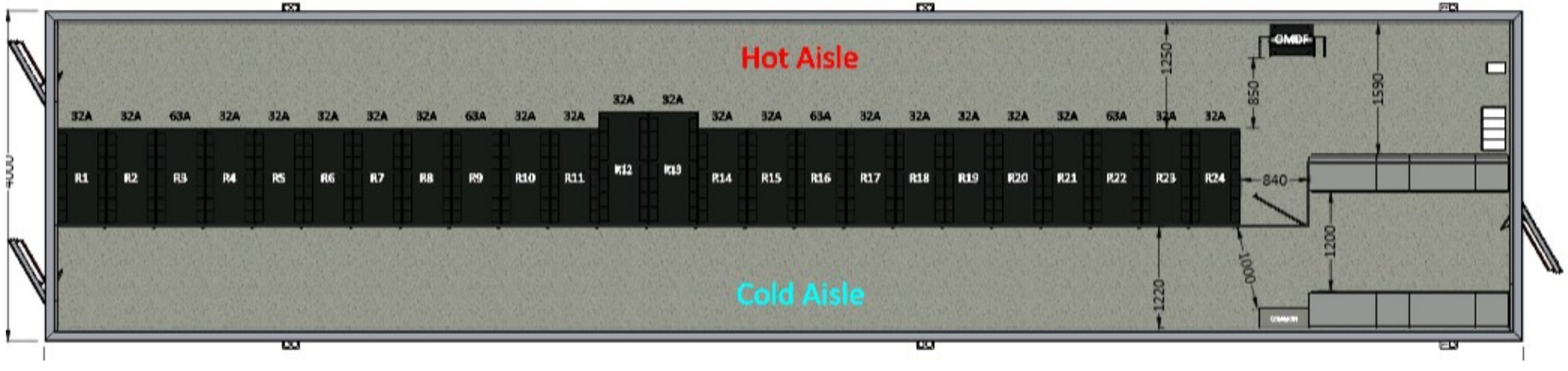
Data centre extensions

- Contract for remote data-centre in Wigner (Budapest HU) will end in December 2019
- Data draining and servers repatriation already started
- Connectivity already reduced to 200G
- Most recent servers will be re-used in Point8 DC extension (see next slide)



Point8 (LHCb) Data-centre extension

- Two LHCb containers (out of six) will be used by CERN IT to host servers during Run3
- 24 racks per container
- Being filled with refurbished servers coming from Wigner
- 800Gbps connection to Meyrin Data-centre with DWDM PAM4 system
- To be returned to LHCb at the end of Run3



LHCb datacentre in containers

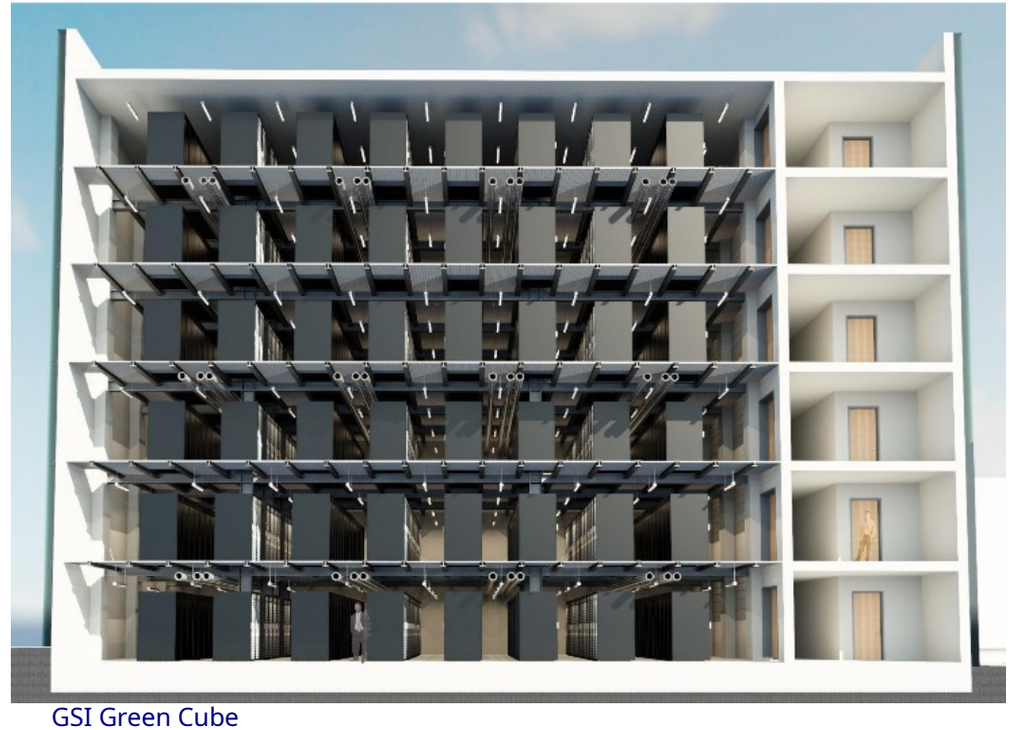
- Two containers for LHCb High Level Trigger, two lent to IT for Run3, two more to build

LHCOPN

PCC: Prevessin Computer Centre

Plan for the Construction of new Computer Centre in the CERN French site of Prevessin:

- Project fully supported by CERN management
- To be built during Run3, to be ready for Run4
- Machines only building, inspired to GSI Green Cube

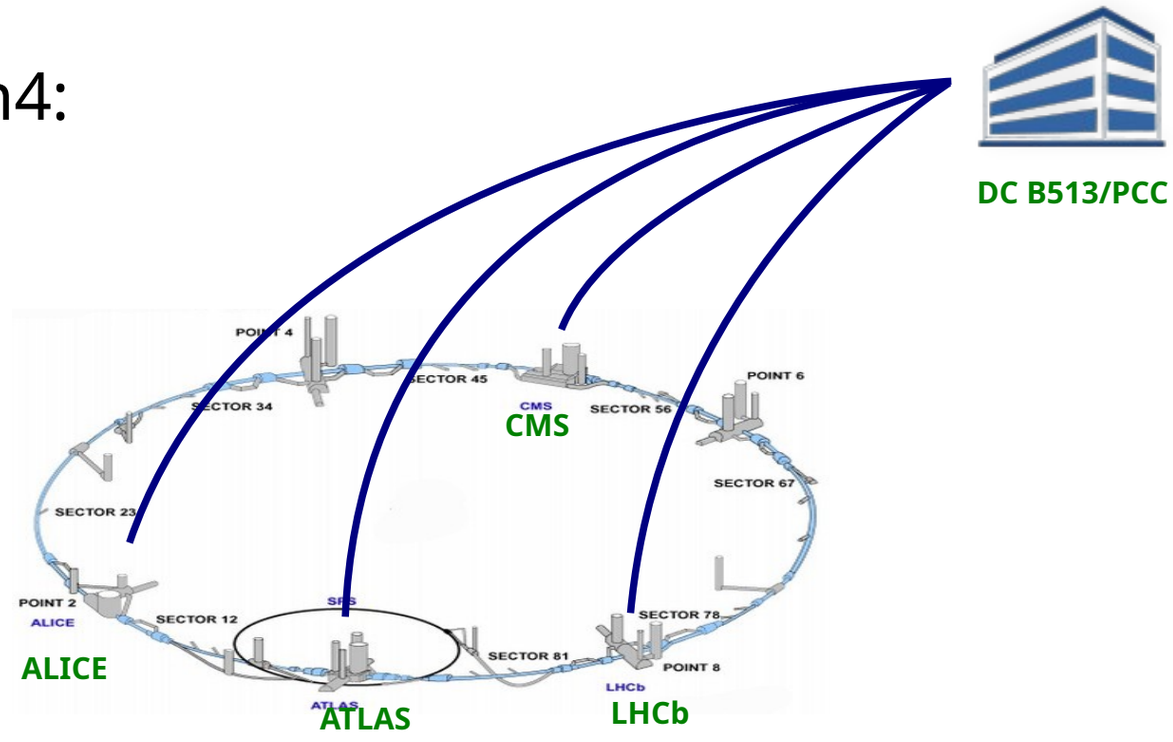


GSI Green Cube

Experiments' DAQ lines to data-centre

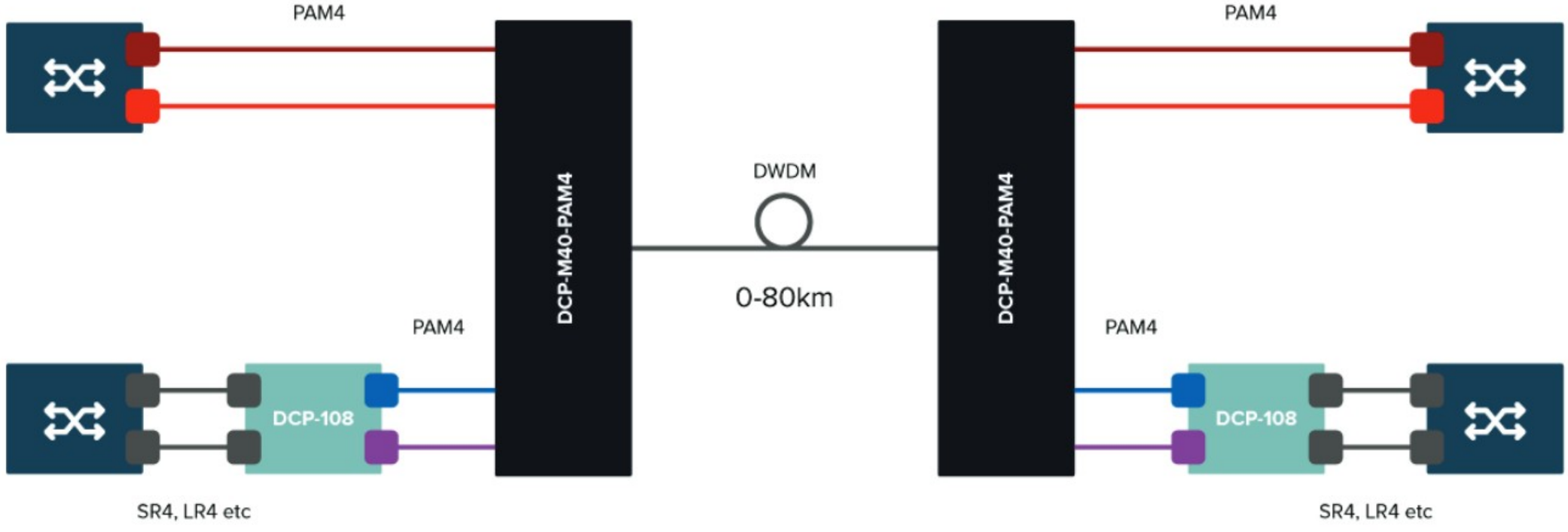
Received requirements for Run3:
- ALICE: 2Tbps
- LHCb: 1Tbps
- CMS: 400Gbps
- ATLAS: no change

Estimated requirements for Run4:
- ATLAS: 4Tbps
- CMS: 4Tbps



Experiments' DAQ lines to data-centres

- Acquired PAM4 DWDM system from Smartoptics
- To be used for LHCb and ALICE connections



CERNlight, CERN Open eXchange Point

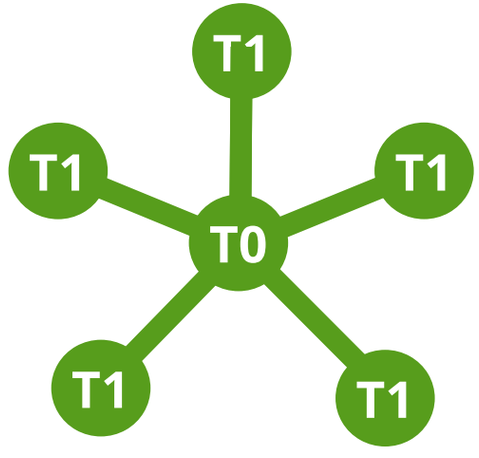
- Replaced Brocade MLXE with Juniper QFX5200
- Moved 100G Geneva-Amsterdam to new SURFnet ECI system
- SURFnet has procured a better fibre Amsterdam-Geneva to be able to light 400G lambdas
- 400G test between CERN and NL-T1 to be held in 2020

LHCOPN update

LHCOPN

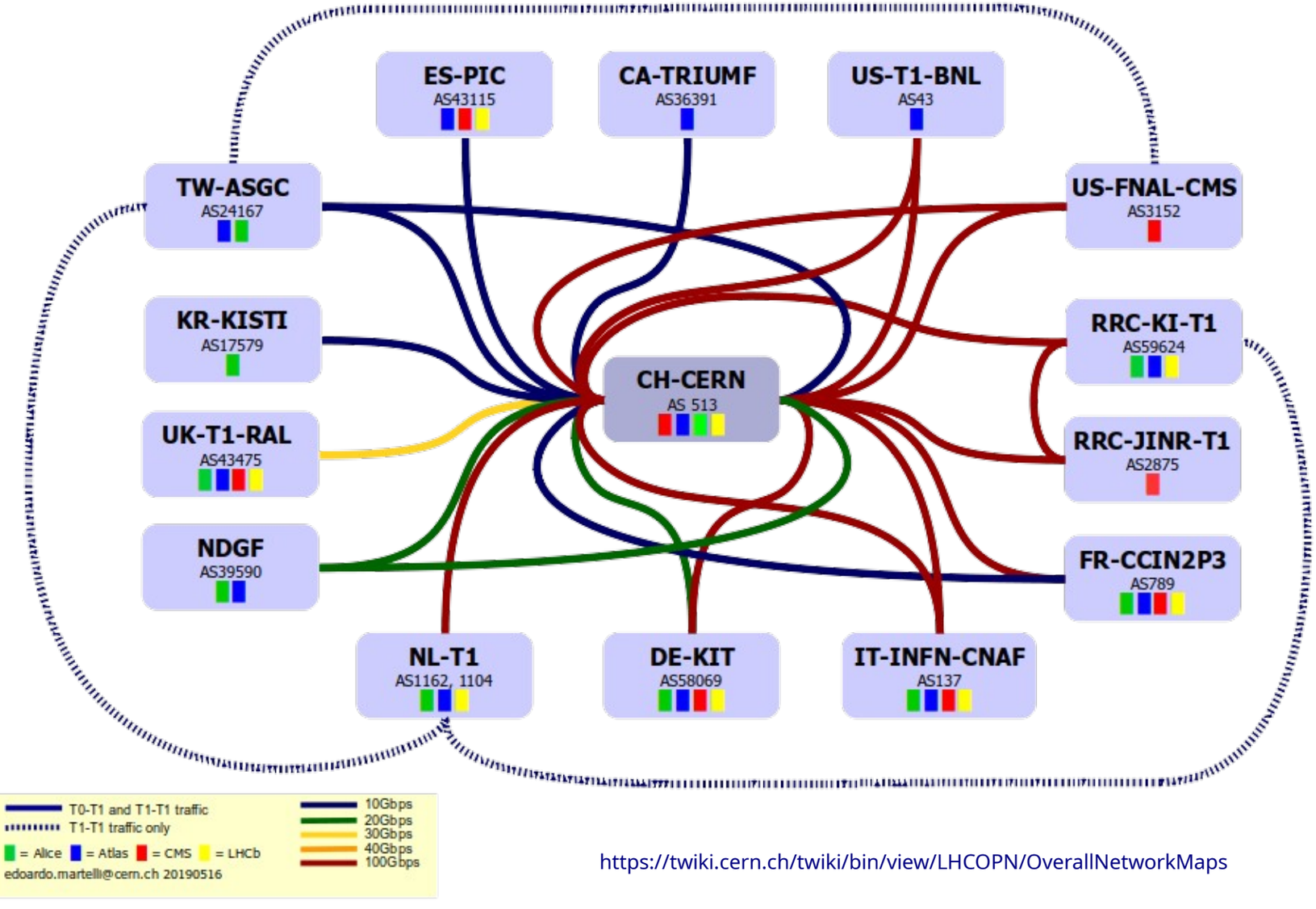
Private network connecting Tier0 and Tier1s

- Dedicated to LHC data transfers and analysis
- Secured: only declared IP prefixes can exchange traffic
- Advanced routing: communities for traffic engineering, load balancing.



LHCOPN

LHCOPN



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

Numbers

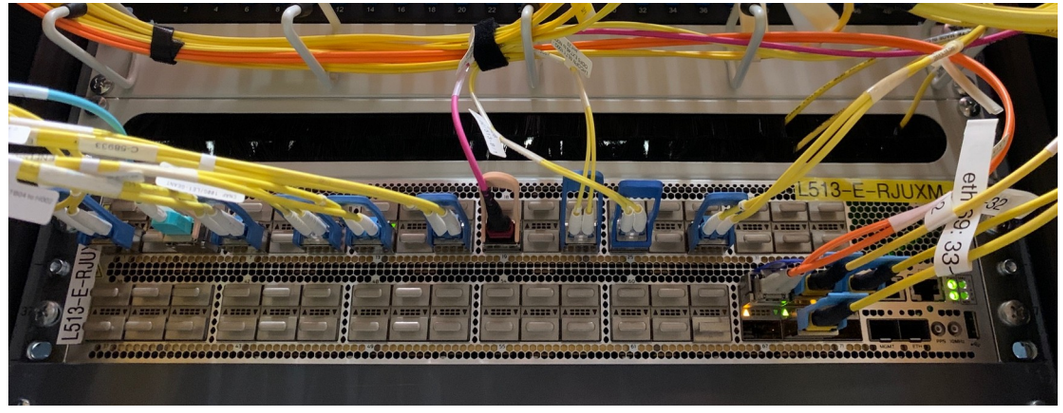
- 14 Tier1s + 1 Tier0
- 12 countries in 3 continents
- Dual stack IPv4-IPv6
- 1Tbps to the Tier0
- Moved ~224 PB in the last year (+40%)

LHCOPN latest deployments

- **NL-T1:** primary link upgraded to 100Gbps
- **RRC-T1s**: primary and secondary links upgrade to100Gbps
- **IT-INFN-CNAF**: primary and backup links upgraded to 100Gbps
- **DE-KIT**: new 100G link deployed, plan to deploy second 100G for backup
- **FR-CCIN2P3**: new 100G link deployed
- **NDGF** will upgrade to 2x100G as soon as network hardware available in Geneva (currently 4x10G)
- **ES-PIC** and **UK-RAL**: will deploy 100G link for Run3
- **CH-CERN:** legacy Brocade MLXE border routers retired. All LHCOPN and LHCONE links now connected to two Juniper QFX10002

CERN Juniper upgrade

Legacy Brocade MLXE16



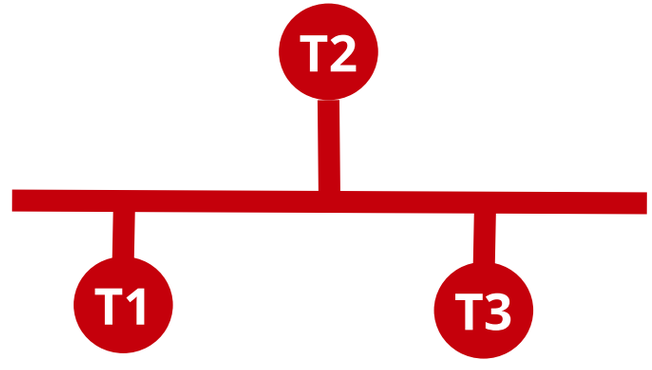
New Juniper QFX10002-72C: 72x40G or 24x100G in 2RUs

LHCONE update

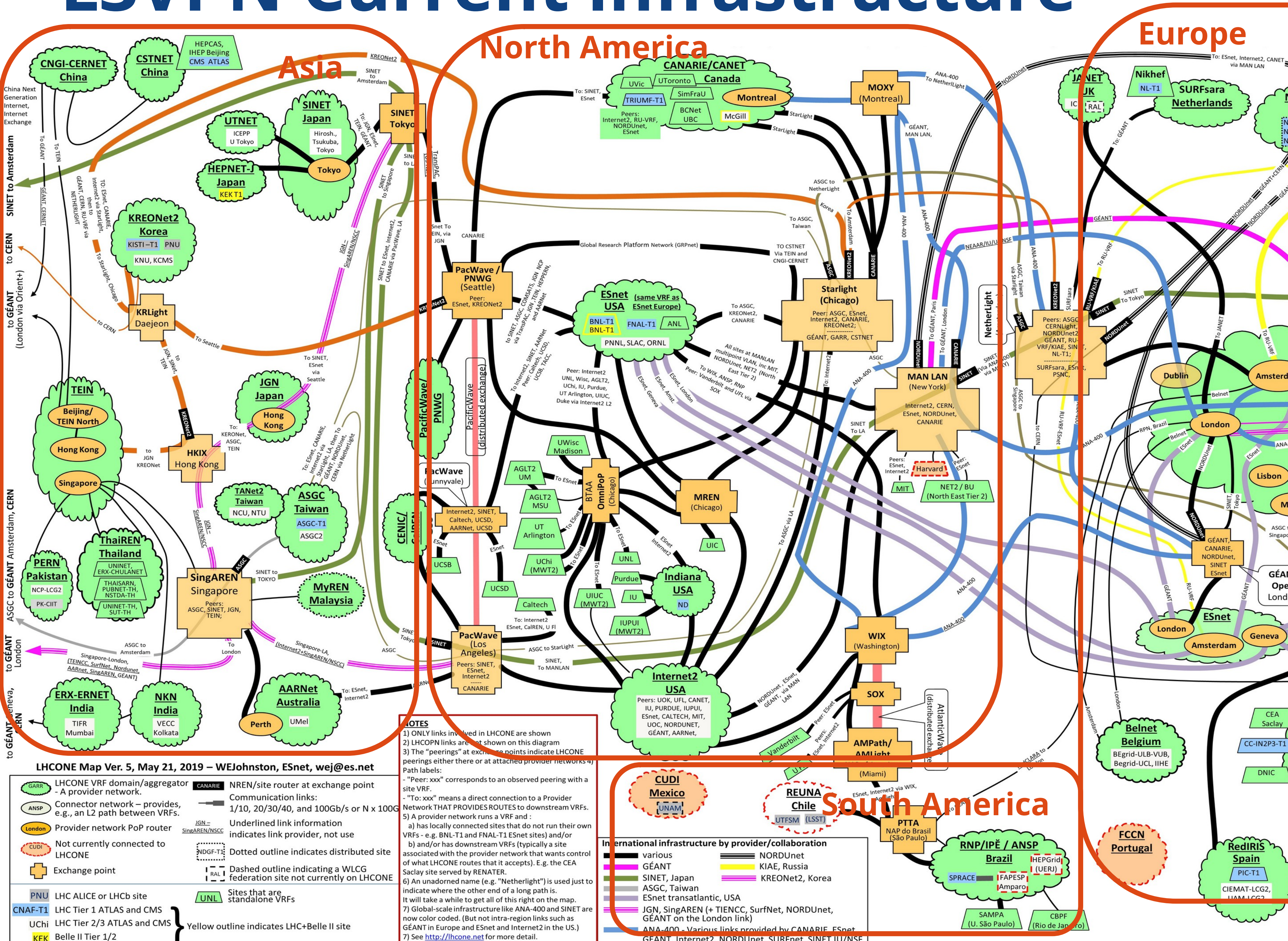
LHCONE L3VPN service

Layer3 (routed) Virtual Private Network:

- Worldwide network backbone **connecting Tier1s, T2s and T3s** at high bandwidth
- Bandwidth **dedicated to HEP** data transfers
- Trusted traffic that **can be allowed to bypass slow perimeter firewalls**



L3VPN Current infrastructure



Numbers

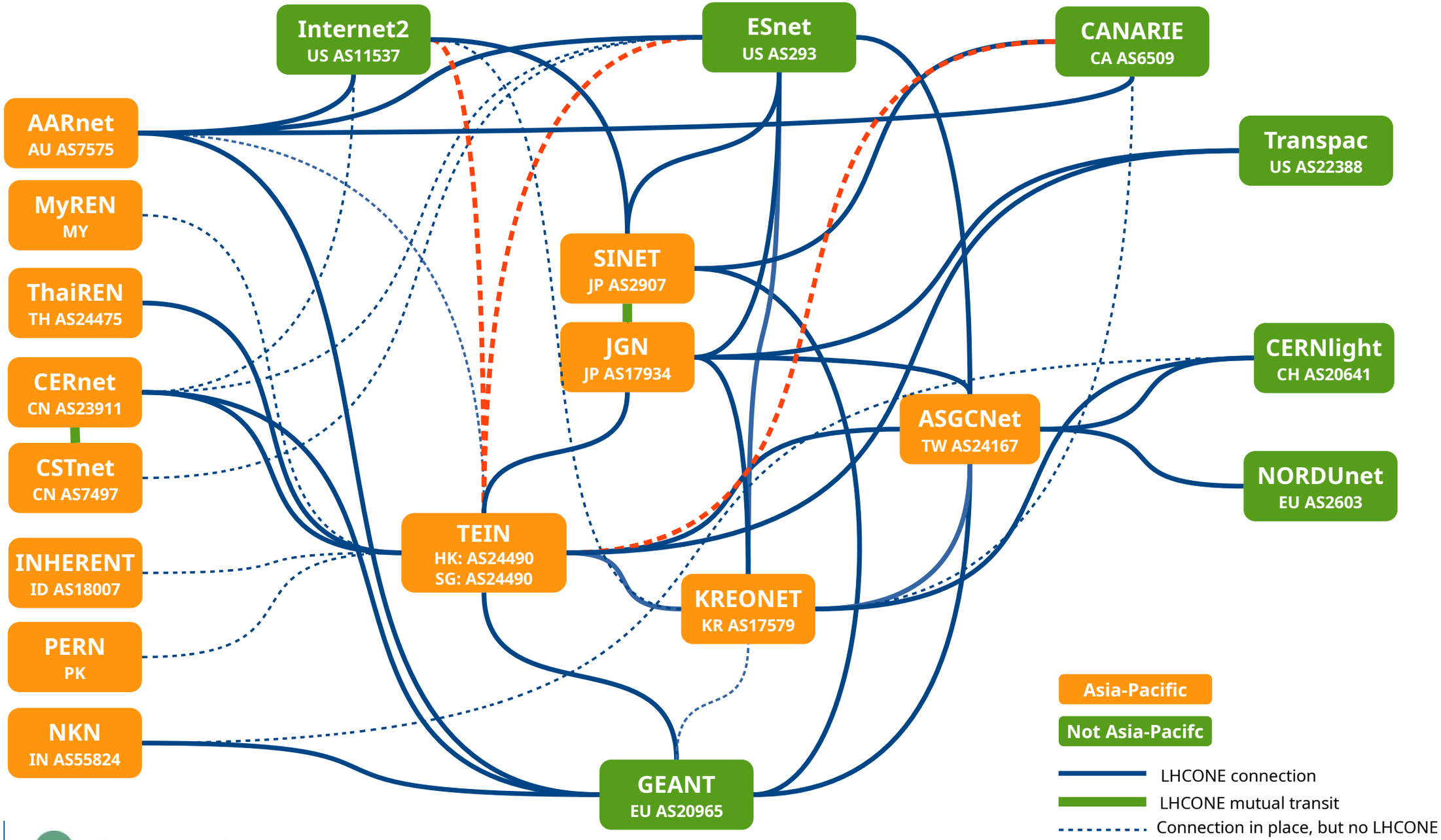
- 28 R&E networks
- 14 Tier1s and ~90 Tier2s in 5 continents
- ~250 perfSONAR instances
- 6 collaborations

LHCONE last year deployments

L3VPN latest changes

- TIFR connection moved to NKN international network. NKN has 20Gbps to CERN which will be increased to 40Gbps for Run3
- Transpac has connected to JGN and TEIN giving transit to US destinations
- UK-T1-RAL working on its connection (Tier2 connected, Tier1 following soon)
- Chile just joined. Sites connected by REUNA (Chilean NREN) via RedCLARA and GEANT
- Estonia T2 will soon connect via NORDUnet

Asia-Pacific VRFs – Current Status

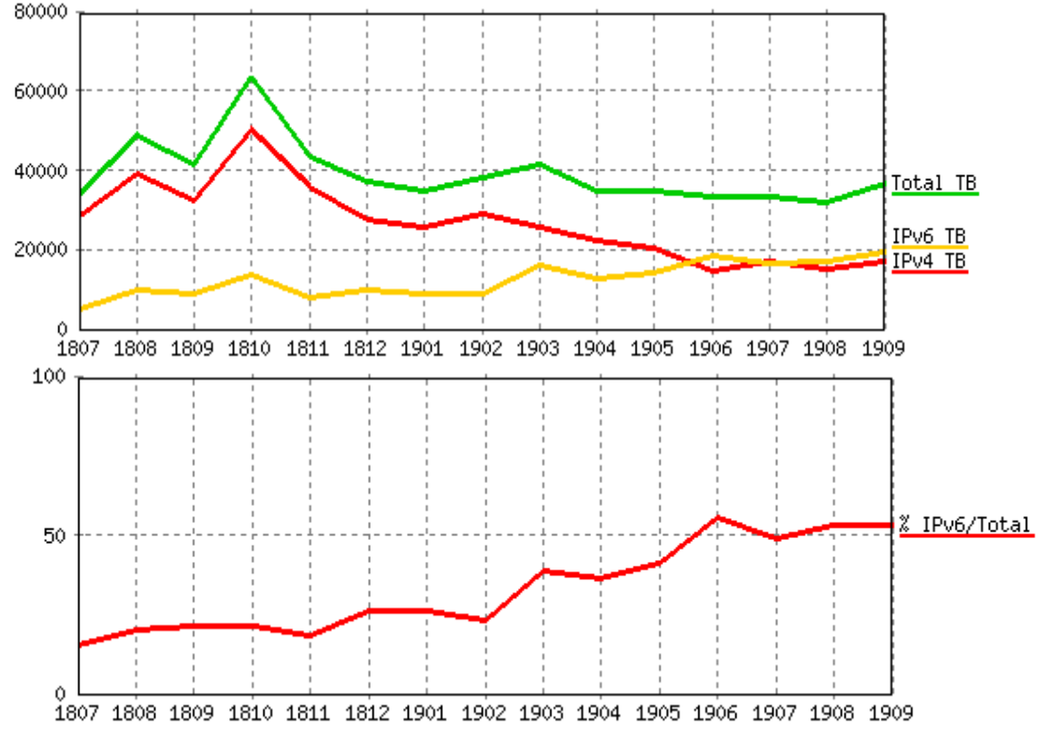


IPv6 adoption in LHCONE and LHCOPN

LHCOPN: all Tier1s connected with IPv6. IPv6 transfers happening among all site except 2 sites

LHCOPN+LHCONE traffic seen at CERN: now more than 50% over IPv6

Looking for areas where to deploy IPv6 only services



LHCOPN+LHCONE traffic seen on CERN border routers

MTU size recommendation

The working group has produced a twiki page to document the MTU issue:
https://twiki.cern.ch/twiki/bin/view/LHCONE/LhcOneMTU

MTU size recommendations for LHCONE networks
- In order to avoid issues with Jumbo frames, these recommendations on MTU size are given:
- The link layer MTU must be set to the maximum supported
- IPv4 and IPv6 MTU must be set to 9000 Bytes

The Path MTU Discovery protocols need these ICMP packets to be allowed:
- IPv4: ICMP Fragmentation Needed - Type 3, Code 4
- IPv6: ICMPv6 Packet Too Big - Type 2 (Value 0)

It's important that these packets have a routable IP address as sources, because unroutable addresses may be dropped by antispoofing filters.

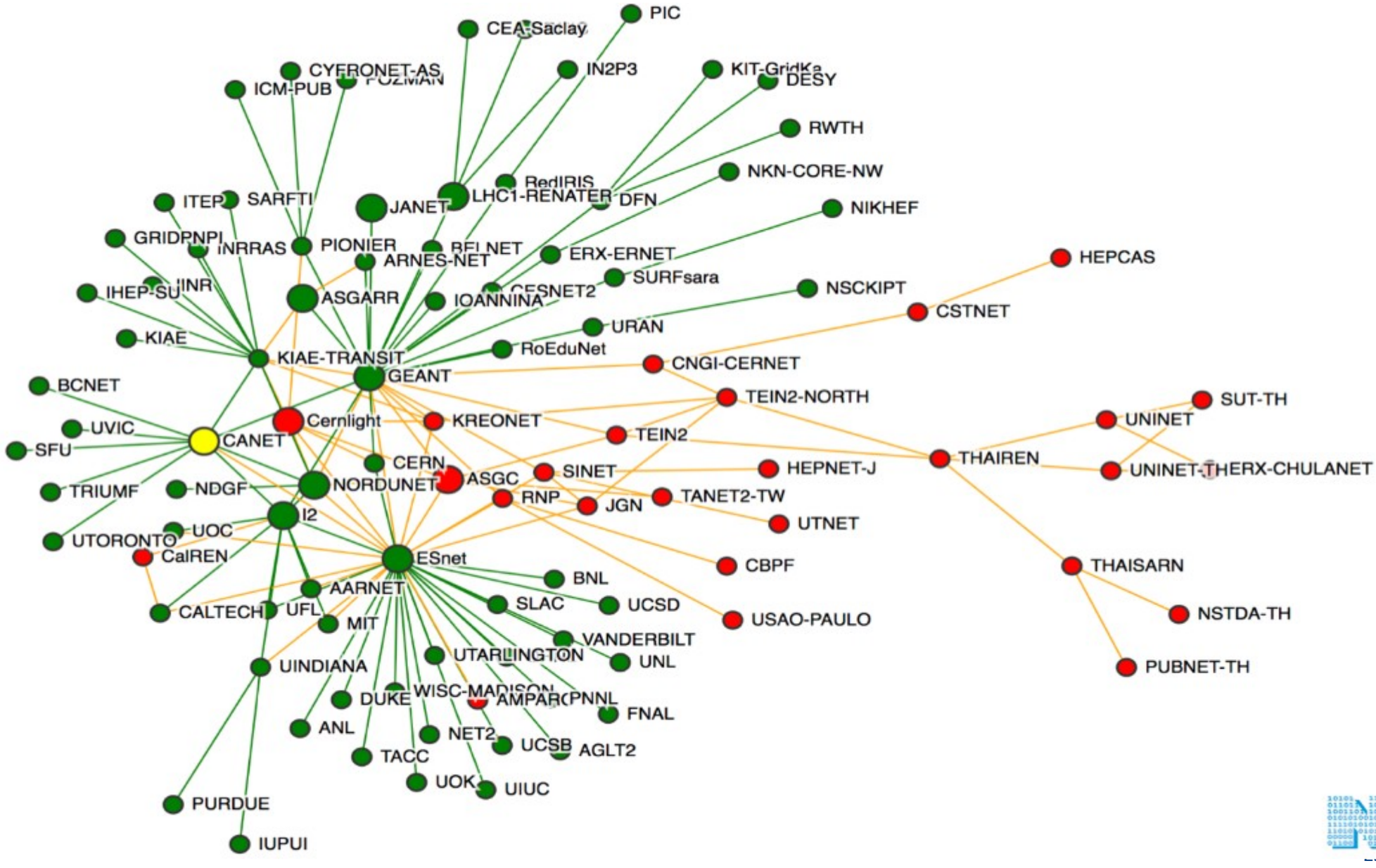
Digging in the LHCONE routing table

NORDUnet has run an analyses of the routing tables of most of the LHCONE VRFs.

It resulted that **reachability is fragmented**, especially on IPv6:
- Only GEANT has a full view of all LHCONE destinations
- Especially the sites behind TEIN cannot reach a fraction of LHCONE

The community will work on improving this situation, which will be followed-up in the future meetings

Example: CANARIE IPv4 reachability



LHCONE Looking Glass

Running looking-glass to analyses the routing tables of the VRFs

Implemented on a CERN router. Now peering with these VRFs:

- ASGC AS24167
- CANARIE AS6509
- CERNlight AS20641
- ESnet AS293
- KREOnet AS17579
- GEANT AS20965 (Geneva and Frankfurt routers)
- NORDUnet AS2603
- RU-VRF AS57484

Internet2, SINET, TEIN should follow

The looking glass is accessible at http://lhcone-lg.cern.ch/

LHCONE perfSONAR: status

~288 perfSONAR instances registered in GOCDB/OIM

~207 production perfSONAR instances

https://www.google.com/fusiontables/DataSource?docid=1QT4r17HEufkvnqhJu24nIptZ66XauYEIBWWh5Kpa#map:id=3

- Initial deployment coordinated by WLCG perfSONAR TF
- Commissioning of the network followed by WLCG Network and Transfer Metrics WG

Slide credit: Shawn McKee, University of Michigan

perfSONAR and monitoring update

- 4.2.1 is the latest release. It adds preemptive scheduling & gridftp testing
- All meshes now test throughput and traces via both IPv4 and IPv6
- OCRE Cloud testing (https://github.com/cern-it-efp/OCRE-Testsuite/)
- Testing of 100G perfSONAR servers on going. Reached ~80Gbps between CERN and NL-T1 via LHCOPN

Some dashboards:
- WLCG MadDASH
- Latency per area
- Throughput
- end2end performance
- IPv4 vs IPv6

LHCONE for HPC

Growing interest of LHC Experiments in HPC resources. Many presentations on the subject at the last HSF/OSG/WLCG workshop

Those infrastructure could be connected to LHCONE, possibly with DTNs, but only when dedicated to WLCG.

The community will investigate possible solutions in the coming meetings

Networking R&D

Networking activities around WLCG

Survey of networking research activities happening around WLCG:

- DTN Nodes (a-la ESnet) and Test Nodes (a-la GEANT)
- High level protocol alternatives (DOMA TPC)
- Low level protocol (TCP) alternatives (AENEAS SKA)
- Efficient use of WAN connections (NOTED)
- Adding additional bandwidth with Bandwidth on Demand and P2P (NOTED, LHCONE-P2P)
- Network Function Virtualization (HEPiX NFV Working group)
- Connectivity for commercial service providers (LHCONE)
- Dedicated VPNs (multiONE)

noted

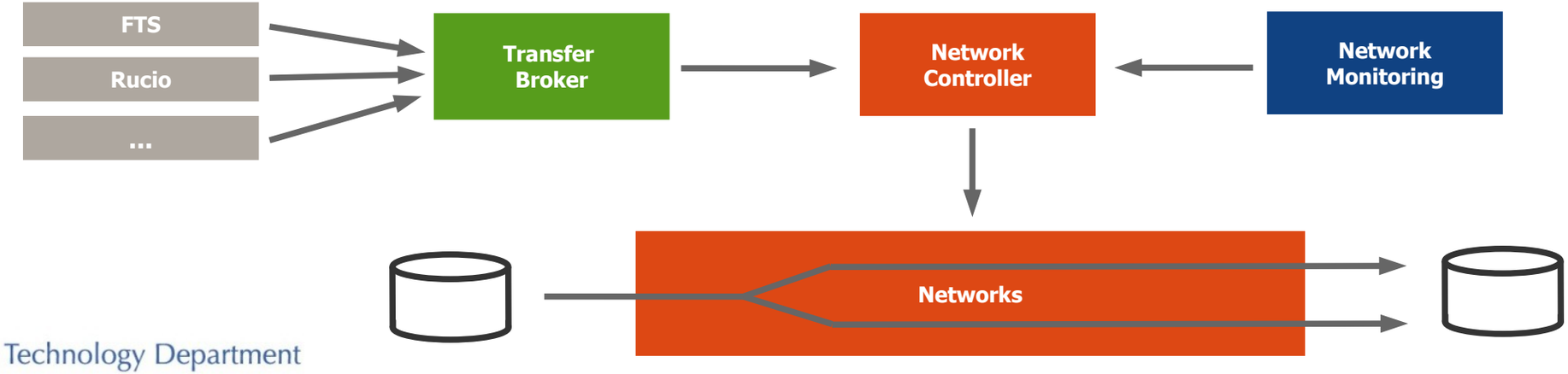
Network activity in the WLCG DOMA contest

Implement a **Transfer broker:**

- Identify upcoming and on-going substantial data transfers
- get information from transfer services (FTS, Rucio …)
- map transfers to network endpoints
- make transfers info available to network providers

Demonstrate a **Network Controller:**

- takes input from Transfer Broker
- modify network behavior to increase transfer efficiency
- take into account real-time network status information

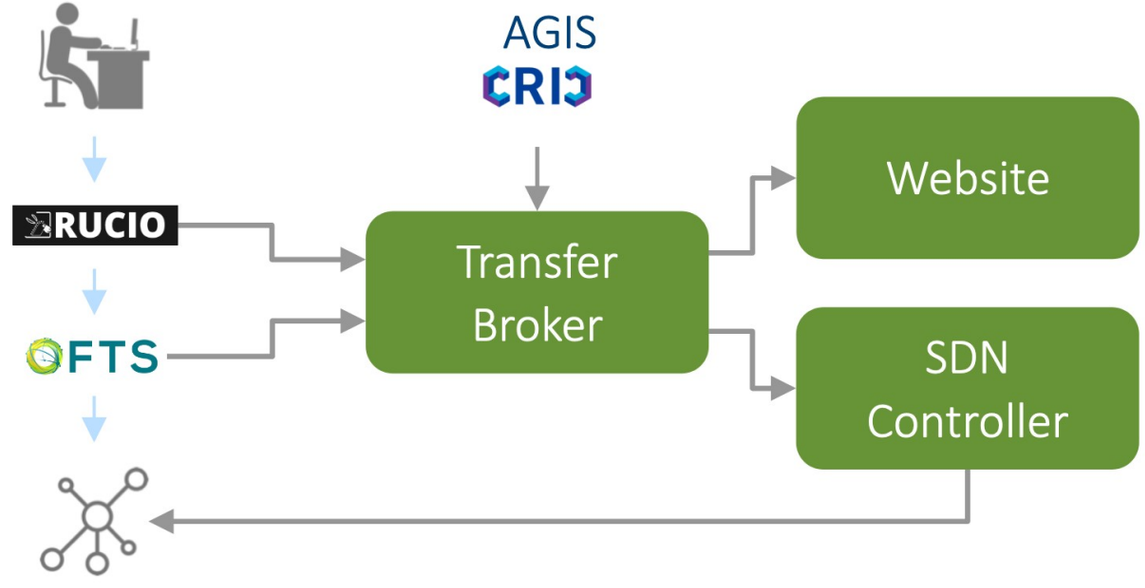


noted: status update

Transfer Broker: interpreting information provided by Rucio to estimate volume of upcoming data transfers and identify source-destination storage elements

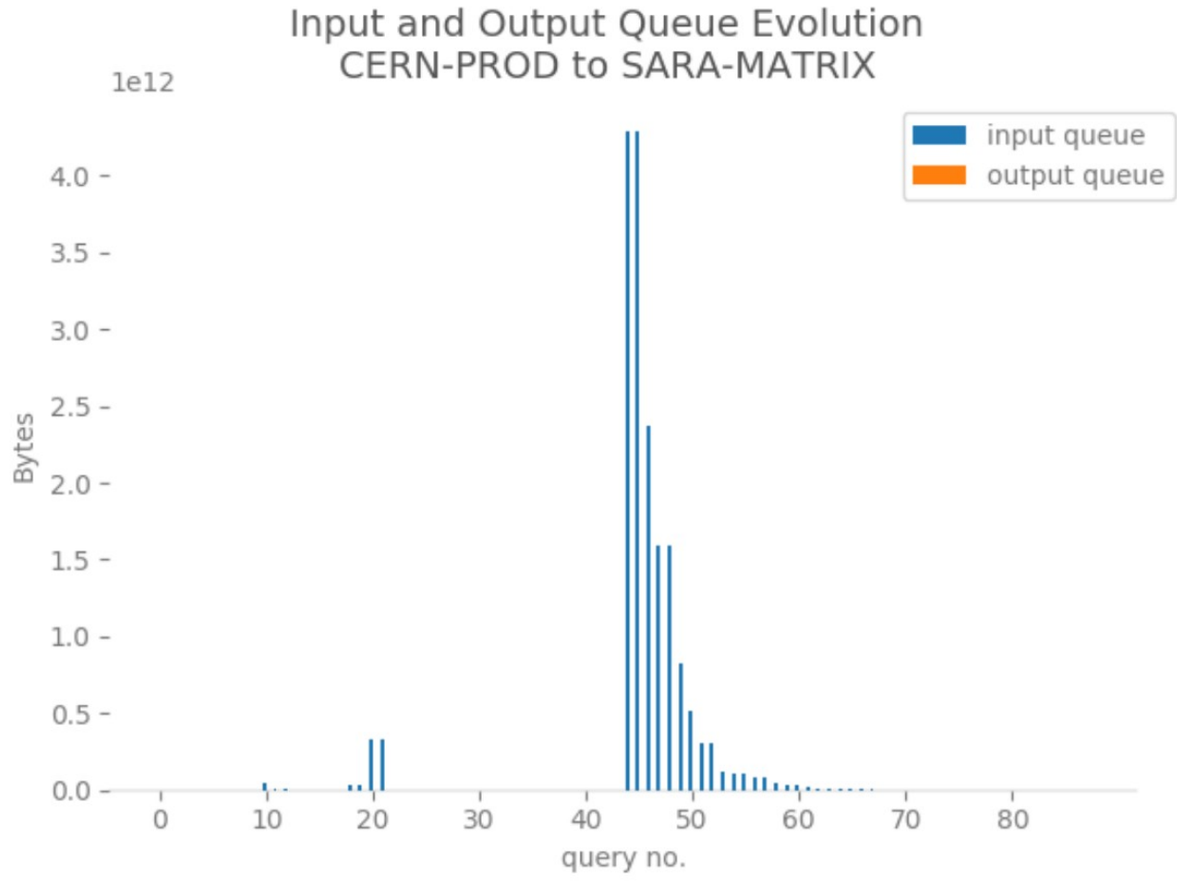
Enhanced CRIC (Computing Resource Information Catalog) to store IP prefixes of storage elements at sites

Evaluating
- Stackstorm for network controller
- Segment Routing for traffic engineering



noted: CERN-NLT1 last simulation

The Transfer Broker successfully observed how the Rucio queue fills up



https://indico.cern.ch/event/810635/contributions/3592922/attachments/1926417/3188957/presentation_hepix.pdf

LHCONE AUP review

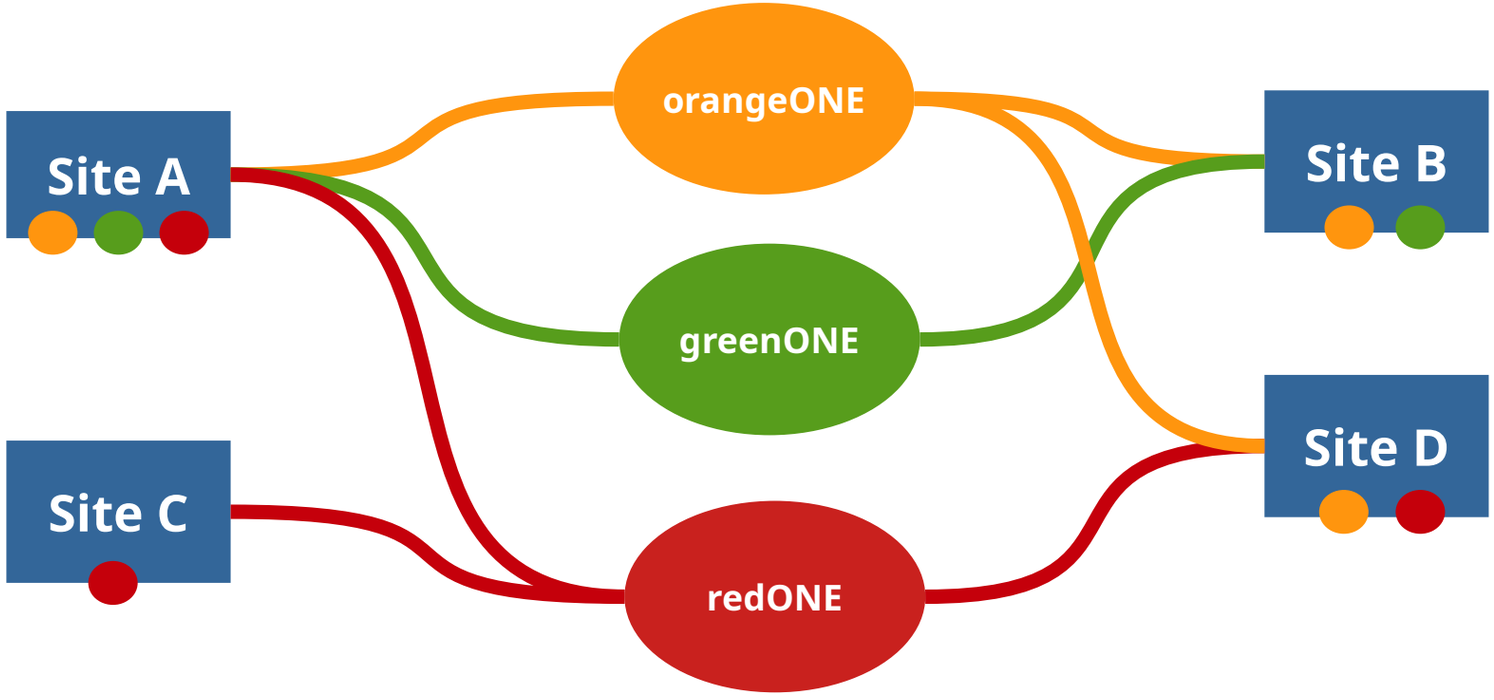
Discussed whether is necessary to review the LHCONE AUP

Still concern on upcoming big data science project: add them to LHCONE or push them to create their own VPN?
- Network providers can easily implement any VPN
- Sites may have problems in selecting the right VPN to use

Proposed to test a separated VPN with an existing or upcoming collaboration

multiONE

A solution would be to implement a VPN for each Collaboration:
- Each site joins only the VPNs it is collaborating with, to reduce the exposure of their data-centre
- Each Collaboration funds its own VPN



Issues with multiple VPNs

- Difficult to select what VPN to use for a Site that serves multiple Collaborations
- Even more difficult if the different Collaborations share the same servers and applications
- The simpler solution (static segregation of resources) is rather inefficient

multiONE will focus on finding multiple solutions that can allow sites to easily and efficiently separate traffic for the different collaborations they serve

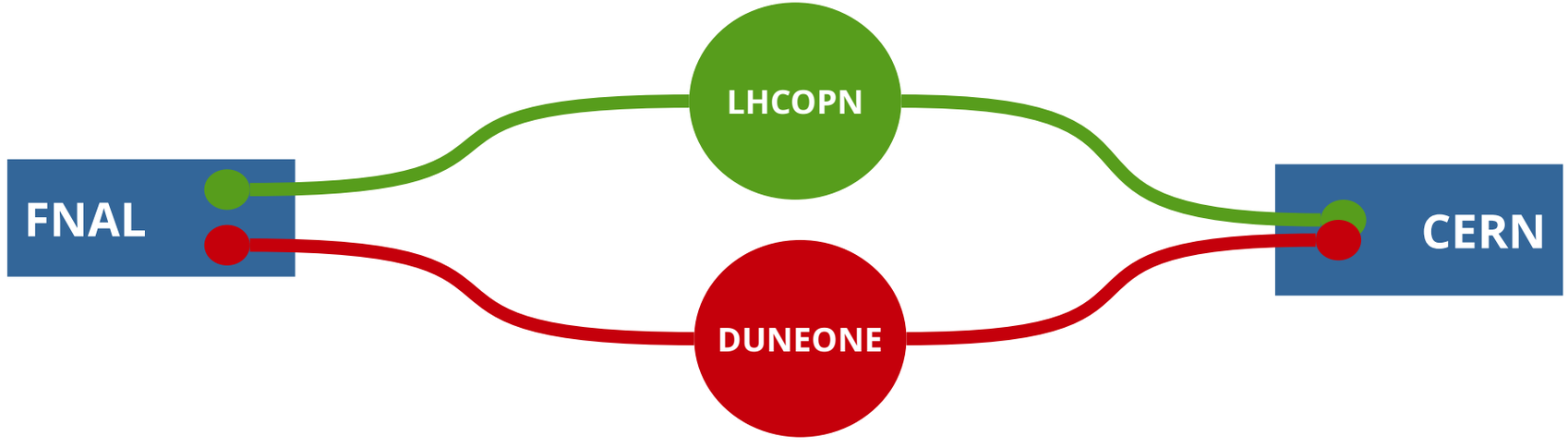
LHC/protoDUNE use case

Agreed with FNAL to prototype the solution with protoDUNE between CERN and FNAL (protoDUNE is currently using the LHCOPN link of FNAL)

New VPN DUNEONE to be agreed with ESnet

No impact on existing protoDUNE traffic and other sites

Resources already distinct at FNAL. Mixed up at CERN



Conclusions

Summary

LHCOPN:

- most Tier1s connected at 100Gbps to CERN. The others will upgrade for Run3

LHCONE:

- Traffic grown ~50% in the last year. IPv6 traffic more than 50% of the total
- The sites upgrading to 100Gbps see immediate traffic grows
- Global reachability improved for IPv4. IPv6 still lacking behind

Monitoring:

- perfSONAR moving to 100Gbps probes
- Looking glass for LHCONE has 8 peerings. More will follow

R&D:

- NOTED will implement a Transfer Broker for WLCG
- multiONE will look for solutions to use multiple VRFs

Following Meetings

Next meeting: 13th and 14th of January 2020 at CERN
https://indico.cern.ch/event/828520/

Following meeting co-located with ISGC in Taipei on 8th- 9th of March 2020
https://indico.cern.ch/event/845506/

Meeting in Fall 2020 could be co-located with NORDUnet conference (September) or HEPiX (October). Or Meeting in Spring 2021 with HEPiX in US. TBC

Questions?

edoardo.martelli@cern.ch