ESnet
ENERGY SCIENCES NETWORK

LHCONE - A global overlay network for the LHC and High Energy Physics
What it is and Why it Works

William Johnston, wej@es.net
Energy Sciences Network (ESnet)
Lawrence Berkeley National Laboratory

With input from Mike O'Connor, ESnet (moc@es.net), Edoardo Martelli, CERN (edoardo.martelli@cern.ch), and Gerben van Malenstein, SURFnet, gerben.vanmalenstein@surfnet.nl

Global Research Platform workshop

UC San Diego, 17-18 September 2019

What is the scale of data movement generated by the Large Hadron Collider at CERN?

- The LHC is the largest fully operational scientific experiment in the world.
  - It is a 27km (~17 mile) diameter (mostly) proton accelerator that has 4 detectors where colliding beams are observed
  - The ATLAS and CMS experiments (detectors) ***each generate about 50 gigabits/sec of data*** 9 months/year that needs to be stored and analyzed
  - This data is distributed to primary storage sites around the world and then further scattered to hundreds of analysis sites
    - Analysis jobs that use the data are generated by physics groups at many institutions - a process that amplifies data movement
  - ESnet serves about 25% of the LHC data storage sites and a comparable number of analysis sites (see /10/)
    - ***LHCONE traffic within ESnet is over 1 Petabyte/day***
    - GÉANT (roughly the European equivalent of Internet 2) handles about 50% more LHCONE data traffic than ESnet
- This scale is not unusual in modern science. Next generation science experiments such as the Square Kilometer Array (SKA) radio telescope, the Large Synoptic Survey Telescope (LSST) optical telescope, and the next generation Linac Coherent Light Source (LCLS), will generate comparable or greater amounts of data.

ESnet

What is LHCONE and how might its technology relate to the GRP?

- LHCONE is a ***"private" network overlay*** that connects the globally distributed data and compute facilities used by the LHC and other high energy physics (HEP) experiments
  - Architecturally, this has much in common with a distributed ScienceDMZ

How LHCONE came about

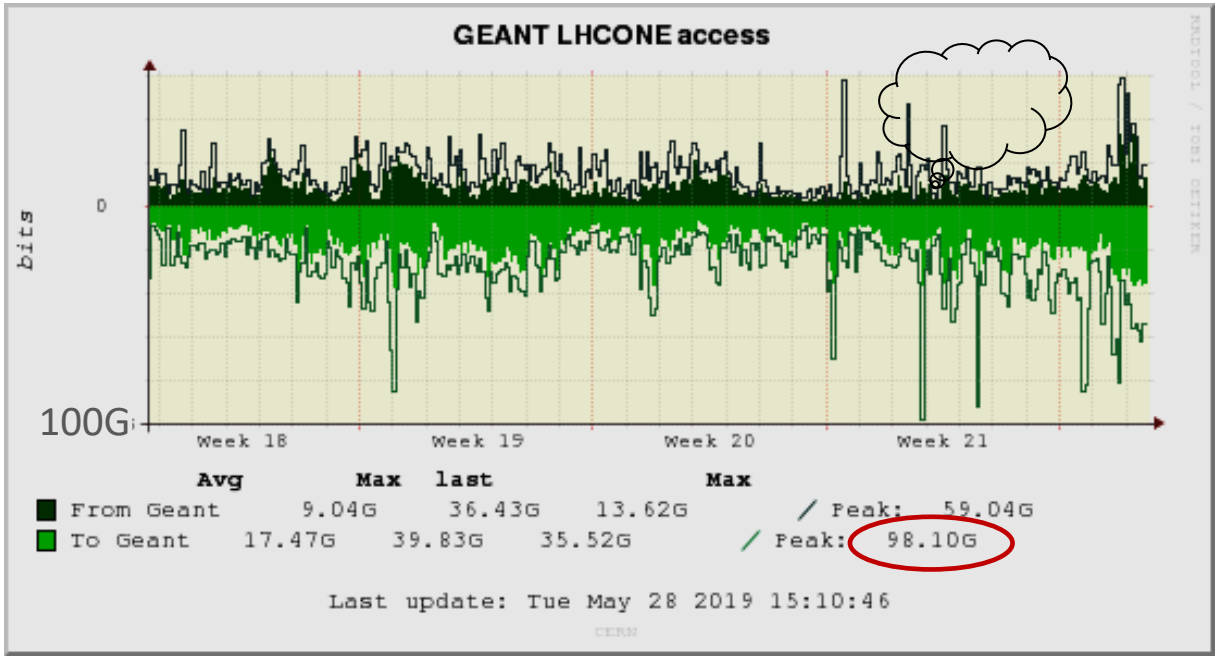
In 2010 traffic from LHC data movement between the globally distributed analysis sites was congesting the small number of transatlantic links available to the general R&E community. It was decided that a way was needed to allow network operators manage the LHC traffic in their network.

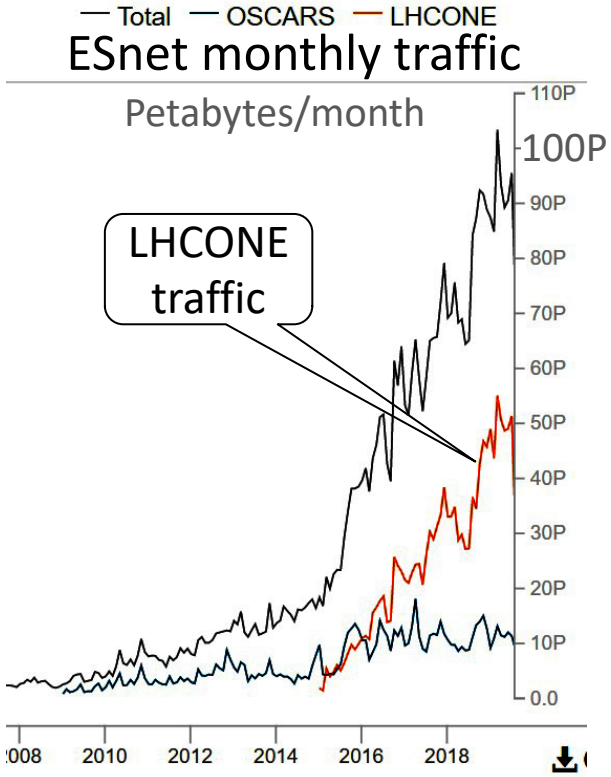
After considerable discussion, the approach of a Layer 3 VPN was adopted. This was a network engineering solution that, once implemented, was not visible to the users.

ESnet

What is the scale of LHCONE?

- IPv6: 209 destinations
- IPv4: 313 destinations
- 25 NREN providers (e.g. CANARIE (CA), ESnet, Internet2 (US), GEANT (EU), RENATER (FR), GARR (IT), KREONET (KR), SINET (JP), etc.)
- 127 sites connected (national laboratoires, institutes, universities)
- Traffic examples (see /8/ and /9/)





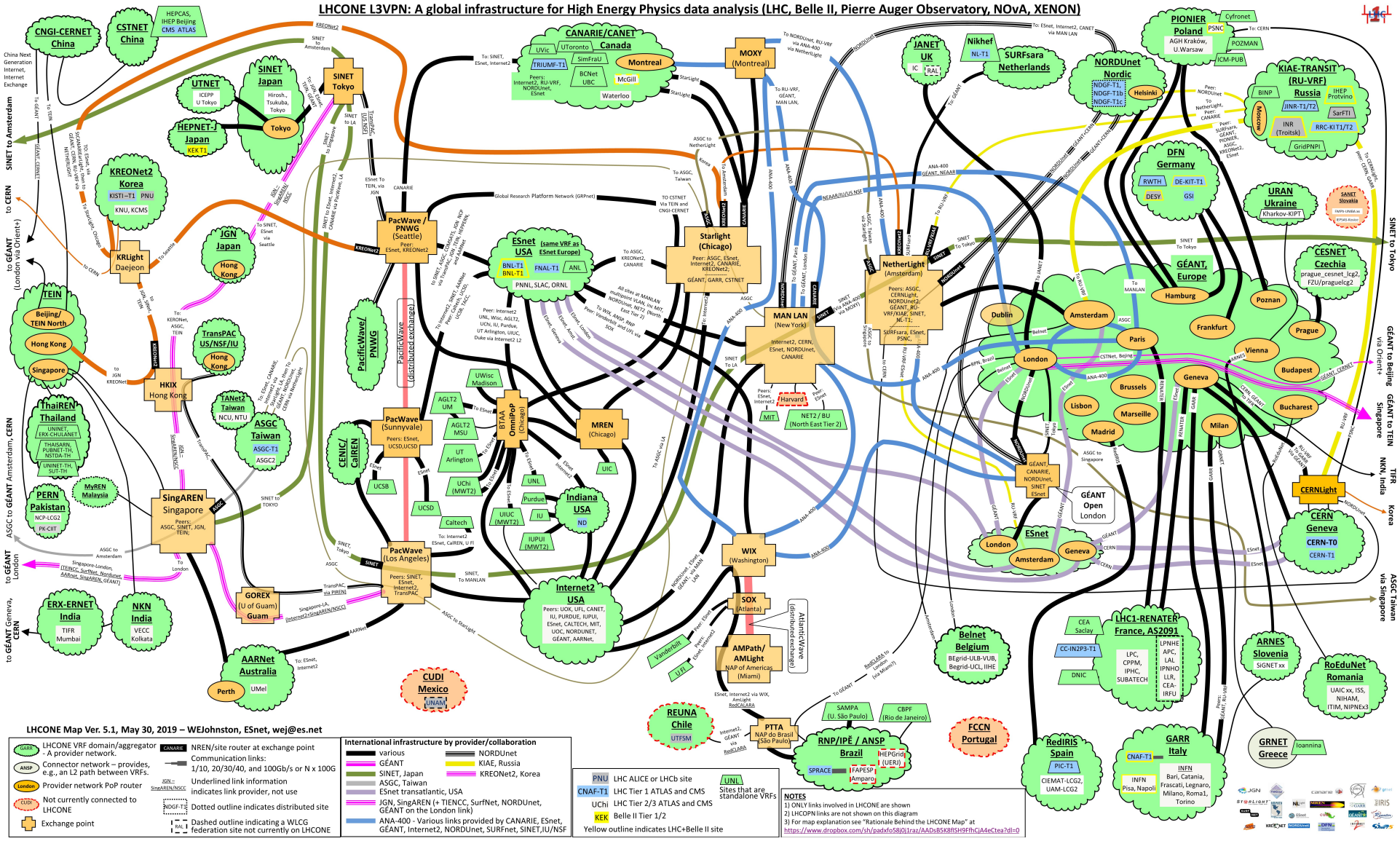
ESnet

LHCONE physical connectivity

- The physical map shows links, exchange points, provider networks, and locations of routing instances involved in LHCONE
- In addition to providing a ***quick view of who a given site or network connects*** to
  - It provides a way for small sites and countries to ***demonstrate their participation*** in the LHC community
  - It provides a way for organizations contributing resources to the LHC data infrastructure to ***showcase their contributions***
  - See “Interpreting the LHCONE Map” and “LHCONE - intermediate detail map v.5.1-2019-05-30» at /1/
- Planning for LHCONE started in mid-2010 and implementation started in mid-2011.

ESnet

Asia and Australia Americas Europe



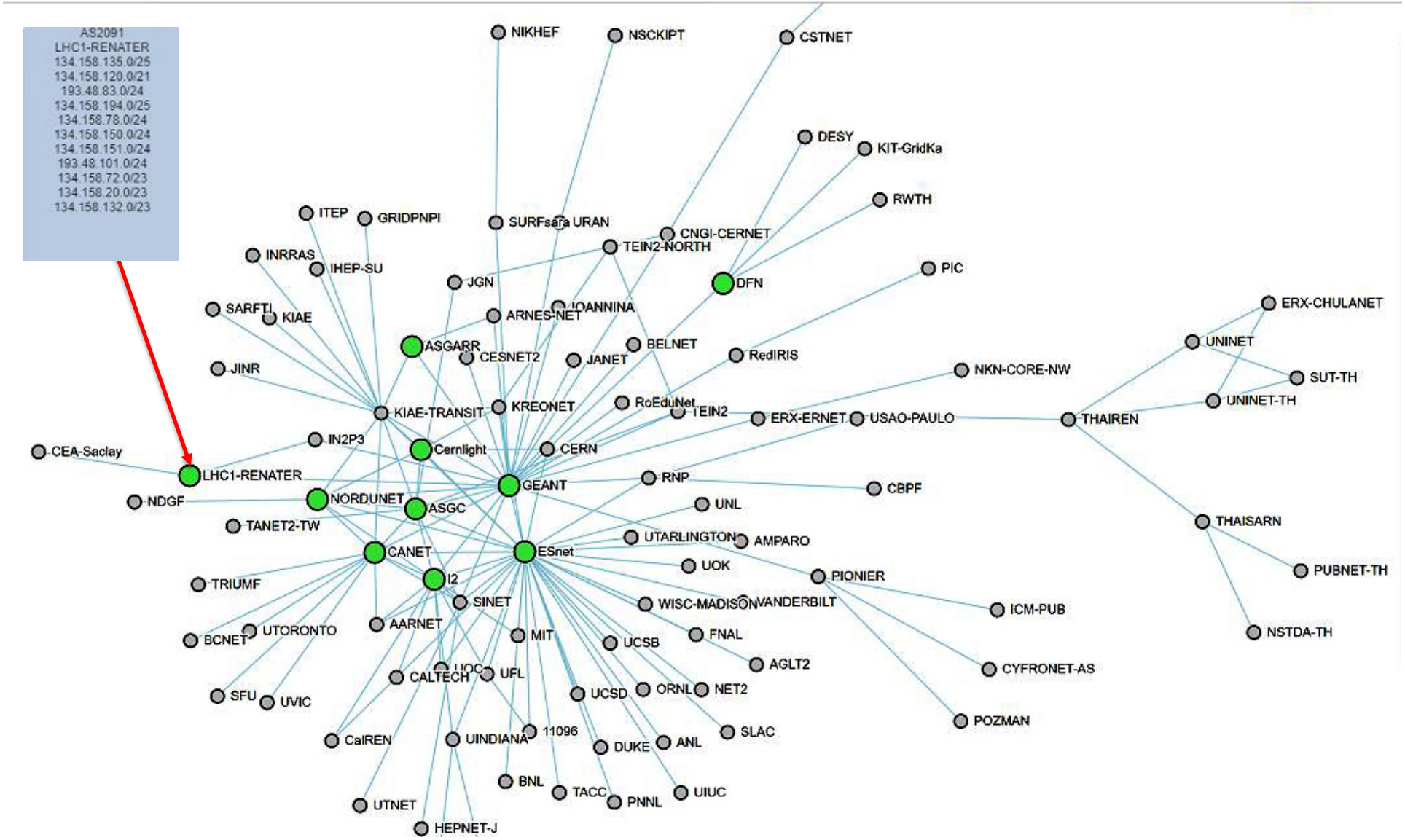
Overview of the global infrastructure of LHCONE

LHCONE logical connectivity

- The LHCONE routes that describe, i.e. how to reach the subnet within a site that has HEP resources connected to it, are published by CERN and several big NRENs
- When a site wants to connect to other LHCONE sites it will gets the needed routes and load them into its VRF (virtual routing instance for LHCONE)
  - Though increasingly most the LHCONE routes are carried in most LHCONE routers
  - Not all LHCONE networks are policy free, and their policies are represented by BGP "Community Values" that control routing policy in upstream provider network
    - This is commonly used, e.g., to "tell upstream networks not to use certain networks to reach me" (e.g. because certain networks may charge to carry my traffic)
    - These routing conditions are available from CERN and LHCONE networks are supposed to honor them
- At present there are about 150 networks/ASNs involved in LHCONE

ESnet

LHCONE logical connectivity



http://stats.nordu.net/pt/lhcone_v4.html Magnus Bergroth (NORDUNet): LHCONE peering relationships from the routing tables (see /11/)

What is LHCONE and how might its technology relate to the GRP?

- LHCONE has several defining characteristics:
  - It uses Layer 3 VPN technology that tags the traffic on shared links and allows for identifying and managing LHCONE traffic as it enters a network
    - This allows network operators to direct LHCONE traffic into parts of their network infrastructure that are not in their core network (mostly trans-oceanic links) that, due to funders wishes/policy, may not be used for general Internet traffic
      - Current technology provides some fairly static mechanisms to manage traffic within the core (e.g. MPLS paths), but in next generation networks more dynamic solutions will be provided by functionality such as MPLS RSVP-TE external path computation. See /4/
  - It has a centrally managed record of signed Appropriate Use Policy statements (at CERN) that effectively defines who participates
    - The AUP defines policy, that, among other things, limits the attached systems to only those that participate in the HEP experiments of LHCONE. See /5/

What is LHCONE and how might its technology relate to the GRP?

- In the early days of LHCONE, the combination of
  - traffic isolation
  - a restricted (and relatively small) user community
  - dedicated subnets at each site that are restricted to LHC/HEP related resources

  provided for a ***level of security somewhat better*** than the general Internet which let sites experiment with resource nodes, such as ScienceDMZs and DTNs, that were outside the site firewall
  - This is less true now because most sites have the resource nodes connected to the general Internet as well as LHCONE
  - However, the “small”, well defined LHCONE community allows for the use of Access Control Lists, which are an effective security tool, e.g. for a ScienceDMZ

ESnet

Why does LHCONE work?

- Because the community wants it to work
  - From ***network operators' point of view*** it effectively provides tagged traffic that can be traffic engineered to manage its use of special resources in the network such as trans-oceanic links.
  - From ***network politics POV***, it allows the LHC traffic to use resources (mostly trans-oceanic links) that, are not available for general Internet traffic.
  - From the ***LHC community POV***, it provides a way to create and manage a well defined and relatively trustworthy environment.
- Because there is (loose) ***central management*** at CERN
- Because LHCONE is ***operated using the same protocols***, routing, processes and procedures that are used in the general Internet.
  - This has allowed it to scale across network domains world-wide.
- Because the LHC is an important project in many R&E network environments that the engineers of the provider networks pay attention to LHCONE
- Because it is closely monitored, both by NOCs of the participating networks and by perfSONAR
  - perfSONAR monitoring is essential if high throughput is to be maintained end-to-end

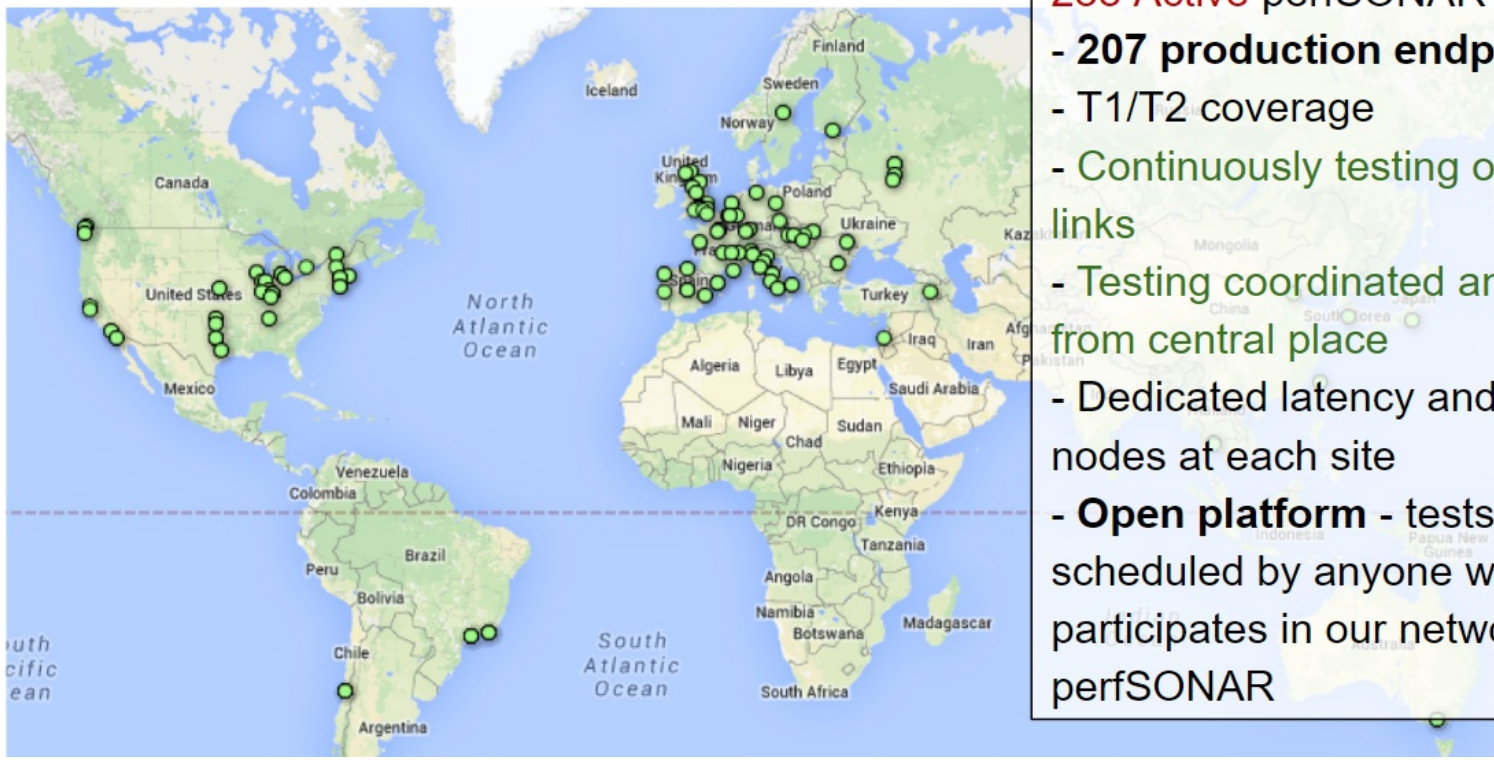
ESnet

Why does LHCONE work?

- "The main purpose for LHCONE [perfSONAR] mesh is to check connectivity/performance (possibly reachability) within LHCONE
  - "Important to have endpoints to test within R&Es (on LHCONE)

(From "monitoring and perfSONAR update," Marian Babik (CERN), Shawn Mc Kee (University of Michigan (US) /12/)

perfSONAR deployment



288 Active perfSONAR instances
- **207 production endpoints**
- T1/T2 coverage
- Continuously testing over 5000 links
- Testing coordinated and managed from central place
- Dedicated latency and bandwidth nodes at each site
- **Open platform** - tests can be scheduled by anyone who participates in our network and runs perfSONAR

Open Science Grid

WLCG Worldwide LHC Computing Grid

The road to overlay networks for other science communities

- One of the important reasons that LHCONE works is that it is a "relatively small" ***community of people who have something in common*** – in this case High Energy Physics research.
  - Initially, only the LHC community were allowed to use LHCONE.
  - Then the Belle-II experiment at KEK in Japan pointed out that they used almost all the same resources as the LHC experiments – physics department compute and storage servers, etc. – so it was agreed that they could join LHCONE
    - Similarly with the other experiments that use LHCONE
  - The opinion of the network engineers that work on LHCONE is that ***its current size is at the upper end of what this approach will support***
  - Adding an unrelated community like the LSST or SKA would add a whole new set of people and resources
  - This argues for building ***new, LHCONE-like infrastructure for other large science experiments*** like the SKA
    - (N.B., however, that this depends on the data analysis and management model for the experiment. Six years ago an analysis of the SKA data model seemed to indicate that it had many similarities to the LHC, and so an LHCONE infrastructure might be useful – that may or may not still be true. [WE Johnston comment] See /6/)
  - For some information on what is involved in setting up an LHCONE-like environment see "How to connect to LHCONE L3VPN" /2/

The road to other overlay networks – a dilemma

On the other hand ……

- Many of the national laboratories, high performance computer centers, etc., that have major resource pools, serve many diverse communities
  - For such institutions to carve out a resource pool so that it can isolated for LHCONE, or other similar overlay, is likely to be non-trivial to say the least
- There are numerous, potentially hard, issues here, including how the network infrastructure interacts with the data management infrastructure
  - However, there is active work in this area between Fermilab and CERN on a proto-DUNEONE (an LHCONE-like overlay) for the Deep Underground Neutrino Experiment (DUNE). See /7/
  - There has also been a lot of work on how to get HEP workflows into the HPC environment. See /3/

Primary Sources

1. "Interpreting the LHCONE Map" and "LHCONE - intermediate detail map v.5.1-2019-05-30," W.E. Johnston, ESnet, https://www.dropbox.com/sh/padxfo58j0j1raz/AADsB5K8fISH9FfhCjA4eCtea?dl=0
2. "How to connect to LHCONE L3VPN," Michael OConnor, ESnet, https://twiki.cern.ch/twiki/bin/view/LHCONE/LhcOneHowToConnect
3. "HPC resource integration into CMS Computing via HEPCloud," Dirk Hufnagel (Fermilab) for the CMS Collaboration, http://cds.cern.ch/record/2647109/files/CR2018_283.pdf?version=1;
   "Enabling production HEP workflows on Supercomputers at NERSC," Gerhardt, Mustafa, Lee, Canon, and Bhimji, Lawrence Berkeley National Lab and National Energy REsearch Scientific Computing Center (NERSC), https://indico.cern.ch/event/587955/contributions/2937411/;
   "Review on current [HEP] workflows and production on HPC centers," Dirk Hufnagel (Fermilab), https://indico.cern.ch/event/759388/contributions/3311664/attachments/1814435/2964911/hpc_production.pdf
4. See "Support of the Path Computation Element Protocol for RSVP-TE Overview," Juniper Networks, https://www.juniper.net/documentation/en_US/junos/topics/concept/pcep-for-rsvp-te-overview.html
5. "LHCONE Acceptable Use Policy (AUP)," Edoardo Martelli, CERN, https://twiki.cern.ch/twiki/bin/view/LHCONE/LhcOneAup
6. "The Square Kilometer Array: A next generation scientific instrument and its implications for networks (and possible lessons from the LHC experience),"William Johnston and Eli Dart, Energy Science Network (ESnet) and Roshene McCool, SKA Program Office, Jodrell Bank, Center for Astrophysics, https://www.dropbox.com/s/55u84jygzghp9l3/The%20Square%20Kilometer%20Array%20%E2%80%93%20A%20next%20generation%20scientific%20instrument%20and%20its%20implications%20for%20networks.v7.pptx?dl=0
7. "multiONE," Edoardo Martelli and Tony Cass – CERN IT-CS, https://indico.cern.ch/event/739882/contributions/3520004/attachments/1906199/3148167/EM-multiONE-GDB.pdf
8. ESnet monitoring portal, https://my.es.net/traffic-volume?s=linear
9. "GEANT network update," Vincenzo Capone (GÉANT), https://indico.cern.ch/event/772031/
10. "WLHC REBUS: Federation Resources," Worldwide LHC Computing Grid (WLCG),
    - Lists all T0, T1, and T2 sites and the resources that they contribute to the LHC data analaysis
    - https://wlcg-rebus.cern.ch/apps/pledges/resources/
11. "LHCONE reachability," Magnus Bergroth (NORDUnet), https://indico.cern.ch/event/772031/
12. "monitoring and perfSONAR update," Marian Babik (CERN), Shawn Mc Kee (University of Michigan (US), https://indico.cern.ch/event/772031/
13. LHCONE meeting presentations
    - E.g. https://indico.cern.ch/event/725706/
    - For past meetings: indico.cern.ch search for LHCONE meeting, refine by "Indico Event"

ESnet