LHCOPN-LHCONE meeting #43
CERN
GDB Summary Report

15th January 2020 – v1.0
edoardo.martelli@cern.ch
smckee@umich.edu

Venue

Hosted by CERN

https://indico.cern.ch/event/828520/

Participants

- 74 Participants
- 47 Institutes
- 4 Collaborations
  (WLCG, BelleII, SKA, DUNE)
- 8 Research Networks

Day 1: Experiment's Requirements

Introduction and Welcome

We had welcome from Frederic Hemmer: "The Higgs Nobel Prize would have not been possible without the work of the LHCOPN/ONE community"

Ian Bird covered "WLCG: new challenges and collaboration with other science projects":

- HL-LHC and others facing Exabyte scale data challenges in the next 5-10 years
- Prototyping an infrastructure to enable FAIR data preservation, open access, reproducibility
- Projects such as ESCAPE should help shape the EOSC and next generation e-infrastructures for science

ALICE

https://indico.cern.ch/event/828520/contributions/3535736/attachments/1968068/3273049/ALICE_feedback_Hardi_LHC_OPN_One_Workshop.pdf

- ALICE is happy with LHC OPN/One and in general with the network performance during Run 2 *always one step ahead of/above the needs*
- Our computing model favors local data access
  - WAN access for file replication and in case of issues with local storage
  - Run 3 model will continue using the same principles
- File transfers (data recovery and storage rebalancing) use will continue at the current level
- T0 to T1s data transfer of Pb-Pb data - higher LHCOPN use for 2-3 months/year
- More data from the experiment, but no general increase on the pressure for LHC networking
- Bandwidth allocation/L2P2P service is interesting to us
  - Especially to cover the AF case

ATLAS

https://indico.cern.ch/event/828520/contributions/3534878/attachments/1968119/3273213/2020_LHCOPN_ONE_-_ATLAS__Networking.pdf

New model: Nuclei (storage) and Satellite (computing) sites. Satellites take data from the closest Nuclei

HL-LHC will produce 300PB/y of raw data

Tier 1s should foreseen 1 Tbps LHCOPN links o CERN Tier0

New analyses model will be tested and refined during Run 3

FTS will be the central service for data transfers. It's the natural candidate to interact with the network

Very supportive of efforts to better identify experiments network traffic

CMS

https://indico.cern.ch/event/828520/contributions/3535003/attachments/1968087/3273252/LHCONE-LPCON_Meeting.pdf

CMS is collaborating with the DOMA activity to reduce storage needs by leveraging network availability and performances. Caches can be one of the tools to achieve this.

HPC/Commercial clouds will not replace WLCG sites, but will be an important contribution. They will need high speed network connections though.

CMS encourage activities on Network Monitoring, traffic marking and pacing to improve operations and data management

LHCb

https://indico.cern.ch/event/828520/contributions/3538546/attachments/1968209/3273325/20200113_LHCb_LHCONE_LHCOPN.pdf

With current upgrade, WAN network needs will increase noticeably already in Run 3

Role of Tier2s will increase, thus they must consolidate their network connectivity

More structured monitoring would be useful to improve operations

DUNE

https://indico.cern.ch/event/828520/contributions/3666186/attachments/1968172/3273249/200113-LHCOPN-DUNE-PELC.pdf

DUNE will have a computing model similar to other LHC experiments.
There will be a strong overlap with WLCG sites, also for network needs

DUNE computing model similar to what DOMA is proposing

DUNE data will be stored equally between American and European sites

Data production expected to ramp up to 40PB/y in 2024

DUNE endorse the FNAL-CERN collaboration on the multiONE project

Belle II

https://indico.cern.ch/event/828520/contributions/3571695/attachments/1968262/3273536/Silvio_Pardi_LHCONE-LHCOPN-13-Jan-2020-v12.0.pdf

Belle II has started data taking and is working hard to reach the maximum luminosity with a roadmap up to 2027.

Max usage of the Network is expected from 2024

Large improvement in the global Network Connectivity happened and tested in 2019.

Improvement on traffic flow monitoring would be beneficial

Day 2: WLCG Requirements, Network Developments and LHCONE/LHCOPN Updates

The DOMA project: requirements

https://indico.cern.ch/event/828520/contributions/3570904/attachments/1968554/3274036/LHCONE-DOMA-01-2020.pdf

The DOMA project foresees requiring 1 Tbps links by HL-LHC (ballpark) to support WLCG needs. This is for the network backbones and larger sites.

User analysis use-cases should have less impact on the network because of a strong focus on trimming the data format for users (nanoAODs).

Caching/latency hiding will be important. DOMA exploring XCache as a mechanism. Provides latency hiding and support diskless sites (with regional data lakes).

Production of AODs (using RAW) will be a network driver, especially regionally. Effectively the “site” is expanded to encompass a “region”

How best to provide networking ala LHCONE for the many new science domain users arising? LHCONE, DUNEONE, SKAONE… ALLINONE ? Many issues and we need to be careful to ensure the experiments, NRENs and sites are collaborate on the (hopefully simple) solution!

NFV WG - update and next steps for the experiments

https://indico.cern.ch/event/828520/contributions/3699233/attachments/1968544/3274004/Network-Discussion-LHCOPN_LHCONE-2020.pdf

Shawn McKee presented the summary of the NFV phase I work and discussed the possible future work areas viewed as potentially useful for the NRENs/experiments.

There was consensus that packet marking was important from all the experiments. Some additional had interest in shaping/packing/orchestration

Areas discussed in the document (pages 53-56):

1. Making our network use visible (marking)
2. Shaping WAN data flows (pacing)
3. Orchestrating the network to enable multi-site infrastructures (orchestrating)

Next steps are for Shawn and Marian to get draft a document on how best to proceed. Need volunteers and methods to keep work connected to constituents

Cost model study on cache size/network trade off

https://indico.cern.ch/event/828520/contributions/3680450/attachments/1968433/3273903/Cost_model_study_on_cache_size_-_preGDB.pdf

Analyzing network and cache use and effectiveness for HEP workflows vs cache size

Allows estimated cost-optimization by estimating both disk and network costs.

Conclusion: simple exercise to evaluate optimal cache size. Varies strongly with use patterns and the scale of the dominant workloads.

Need to make this estimation generically usable to support changing use-cases/costs.

multiONE

https://indico.cern.ch/event/828520/contributions/3570916/attachments/1968565/3274060/LHCOPNE-20200114-Geneva-multiONE.pdf

Edoardo Martelli presented on the need for and limitations of LHCONE.

- Original AUP for LHCONE limited use to sites providing WLCG resources
- How do we handle other science domains, especially with they use the same sites?

Primary advantages of LHCONE

- Trust: allows bypassing firewalls
- Relatively small community that knows and trusts each other and understands needs.

MultiONE is a solution that would allow multi-VO sites to better manage and orchestrate their various VO specific traffic.

- DUNEONE was proposed to test the waters…

Lots of discussion about trust, scope, previous issues and how best to move forward.

DUNEONE prototyping should continue but how best to proceed longer term is still an open question.

Options for resources separations

https://indico.cern.ch/event/828520/contributions/3642464/attachments/1968664/3274267/multione_batch.pdf

Ben Jones presented from the perspective of the batch systems at CERN

Context: 500K-2M jobs/day, ~200k cores, 17k virtual machines, HTCondor, Openstack and Kubernetes

Most pools are shared but there are VO dedicated pools

Options:

1) Partition resources but this is very inefficient/wasteful!
2) DSCP bit field setting to allow network to handle specific traffic and certain ways
3) Assign IP to job (putting jobs into VLANs use VXLAN or similar)
4) Explore SDN/Kubernetes/Openstack features on **tungstenfabric**, select right network dynamically from the set of provisioned networks

Challenge: ~2 million IP addresses created/destroyed per day

Option 4 being explored using Docker containers (Docker Universe in HTCondor)

Concerns on LHCONE Security

https://indico.cern.ch/event/828520/contributions/3642464/attachments/1968664/3274267/multione_batch.pdf

Open discussion on security concerns initiated by Ian Collier

AUP/Security monitoring on LHCONE is on “unroutable packets” being injected into LHCONE

- Hard to locate the source
- Against the AUP
- Contact to sites or networks has not resulted in responses in many cases

Romain Wartel and Dave Kelsey discussed re-using policies and procedures from similar activities to apply to LHCONE

- View is the the activity of Bruno Hoeft and Mike O’Connor to identify and quantify the unroutable packets should continue

No conclusions but good discussions on policy and possible technical options

GEANT will work on an AUP annex to clarify escalation procedures

The NOTED project

https://indico.cern.ch/event/828520/contributions/3570905/attachments/1968456/3273836/presentation_noted.pdf

Coralie Busse-Grawitz presented on the NOTED project

Goal is to publish network aware information on on-going massive data transfers; information that can be used by network operators to provide additional capacity by orchestrating the network behavior (e.g. more effective use of existing network paths; load balancing).

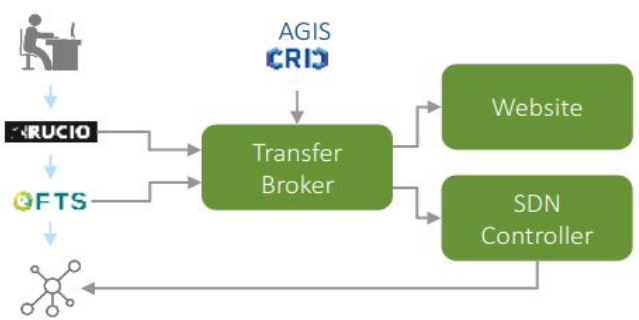
The NOTED Transfer Broker can interpret Rucio and FTS queues and translate them into network aware information with the help of the CRIC database

Network controllers can use the Transfer Broker to decide when and how apply optimizations in their domains

Currently still prototyping. Demonstrated full chain with transfer between CERN and NLT1 and DE-KIT

Exploring other load-balancing options include segment based routing.

Other file transfer services should be considered in the future



The SENSE Project

https://indico.cern.ch/event/828520/contributions/3701881/attachments/1968917/3274702/SENSE_LHCONE_LHCOPN_20200114.pdf

Chin Guok presented the SENSE project: SDN for end-to-end Networked Science at the Exascale.

Trying to address how you would construct network topologies for specific use cases

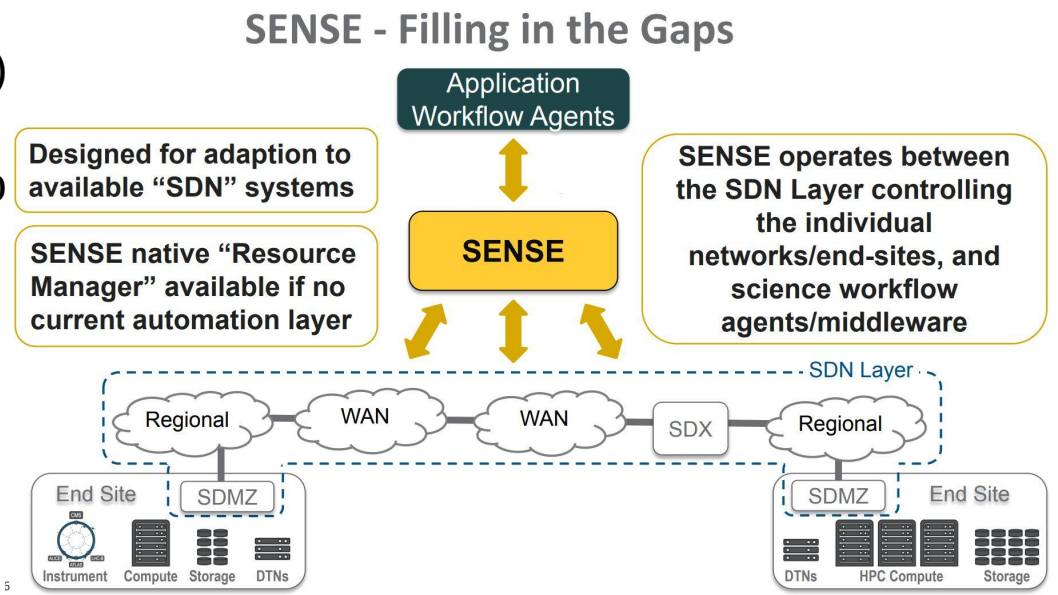
Manual implementations are not feasible...we need information and automation

Uses MRML (multi-resource modeling lang)

Prototypes running for 3 years; SC19 demo

Seeking users/collaborators

See https://sense.es.net (DOE funded)



Network design ideas and proposals for SKA

https://indico.cern.ch/event/828520/contributions/3677232/attachments/1968653/3274246/CERN_LHCONE_SKA-AENEAS_v1.pdf

Richard Hughes Jones covered the planning, design and prototyping of the SKA network

Lots of science topics within SKA and different types of data be utilized; different from LHC

Telescopes produce ~ 1 PB/day

Typical site traffic: 20 Gb, 40 Gb peak per telescope => 100 Gbps per site

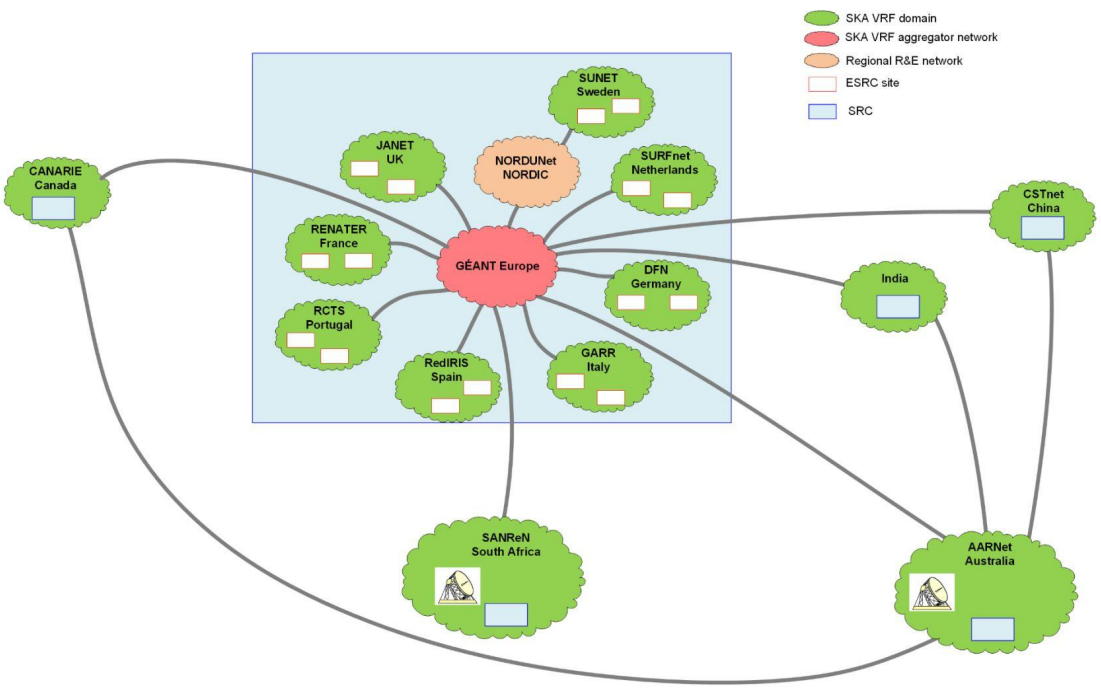
Planning for 2 large sites/country (~T1)

Setup global VRF for SKA (SKAONE?)

Working to understand operational modes and cost implications

Global Network Architecture for SKA

Global VRF based overlay with peering linked over the shared academic network



CERN Network and LHCOPN Update

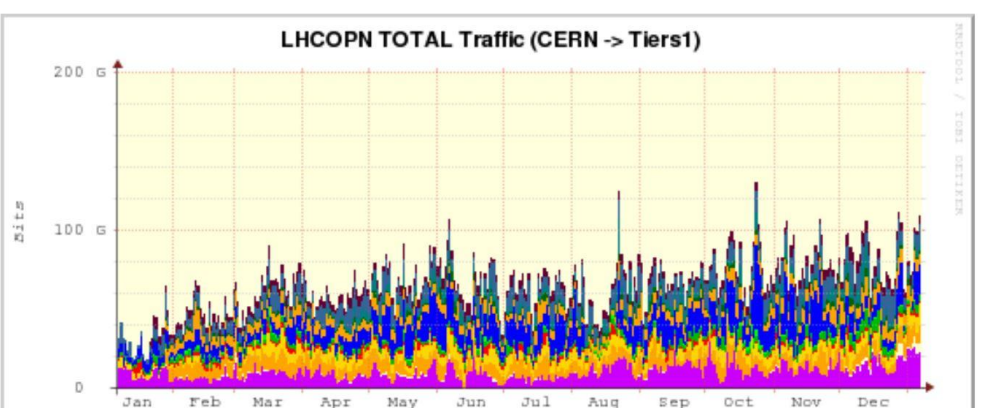
https://indico.cern.ch/event/828520/contributions/3534905/attachments/1968570/3274066/LHCOPNE-20200114-Geneva-LHCOPN-update.pdf

Updates on CERN:

- Data-centre network being renewed to increase overall switching capacity and reliability
- Interim use of containers at LHCb to host computing/storage resources
- CERN had a new Prevessin Computer Centre approved (using GSI Green Cube model). Will be ready for Run4
- CERNlight testing 400G with NL-T1 this year

LHCOPN:

- 14 T1s and the T0
- 12 countries, 3 continents
- moved 256PB in 2019
- T1s are mostly 100G (or multiples)



LHCONE Update

https://indico.cern.ch/event/828520/contributions/3534907/attachments/1968601/3274129/LHCONE-Update-ESnet.pdf
https://indico.cern.ch/event/828520/contributions/3534907/attachments/1968601/3274204/LHCONE-L3VPN_ECapone.pdf

Two presentations:

Mike O'Conner / ESnet presented an LHCONE update
- LHCONE architecture detailed for those not as familiar
- Robust, scales, cost effective, science only network with a distributed mgmt model

Presented work on unroutable packet tracing and network prefix controls. Discussed connectivity challenges for sites with multiple path options

Enzo Capone / GEANT presented a second LHCONE update
- Focused on the L3VPN service and compared 2014 with today; significant changes!
- Much greater transAtlantic capacity and coordination, much improved BW and connectivity (VRFs) in Asia

LHCONE in South America

https://indico.cern.ch/event/828520/contributions/3534910/attachments/1968802/3274549/WLCG_and_LHCONE_in_Latin_America.pdf

RedCLARA connects the South American countries and provides LHCONE to Brazilian and Chilean WLCG sites

New undersea cables will allow RedCLARA to connect directly to Europe and Africa, reducing their dependencies on US networks

CBPF (WLCG site in Rio de Janeiro) still experiencing some efficiency issues with some EU sites. To be investigated with the help of WLCG

Conclusions

Summary

Experiments:

- will improve their computing model to reduce Storage costs by increasing network utilization
- expect a ~10x increase of network traffic for Run4
- desire more complete network monitoring information to improve their operations

Data marking would help the understanding of traffic flows and network utilization

NOTED activity relies on abundant and reliable network connectivity to implement cost savings in storage

multiONE will try to achieve traffic separation to improve security and traceability

NOTED and Sense among the R&D project to improve network efficiency and automation

And many thanks to **Mike O'Connor** in advance of his retirement in June. Mike has been a major contributor-to, manager-of and participant-in LHCONE/LHCOPN for many years! We will miss him.

Actions for next meeting

Define policies and procedures to improve LHCONE Security model

Define roles and activities of a Network R&D working group

Define action plan to implement Experiments' requirements

Future Meetings

Next meeting co-located with ISGC: 8-9 of March 2020
https://indico.cern.ch/event/845506/

Meeting in Fall 2020 could be colocated with NORDUnet conference. Meeting in Spring 2021 with HEPiX in US

References

Meeting agenda and presentations:

https://indico.cern.ch/event/828520/