Network activities - update

DOMA meeting, CERN 27th December 2019

edoardo.martelli@cern.ch

Networking activities

- NOTED
- multiONE
- DTN Data Transfer Nodes and Automated Provisioning
- Low level protocol alternatives
- HEPiX NFV Working group
- LHCOPN and LHCONE

NOTED

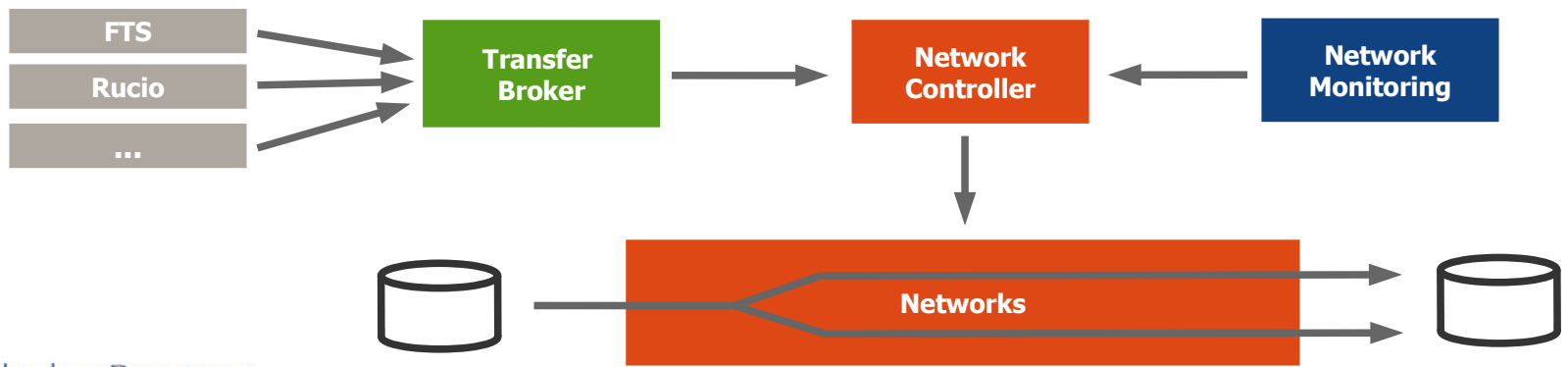
NOTED

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

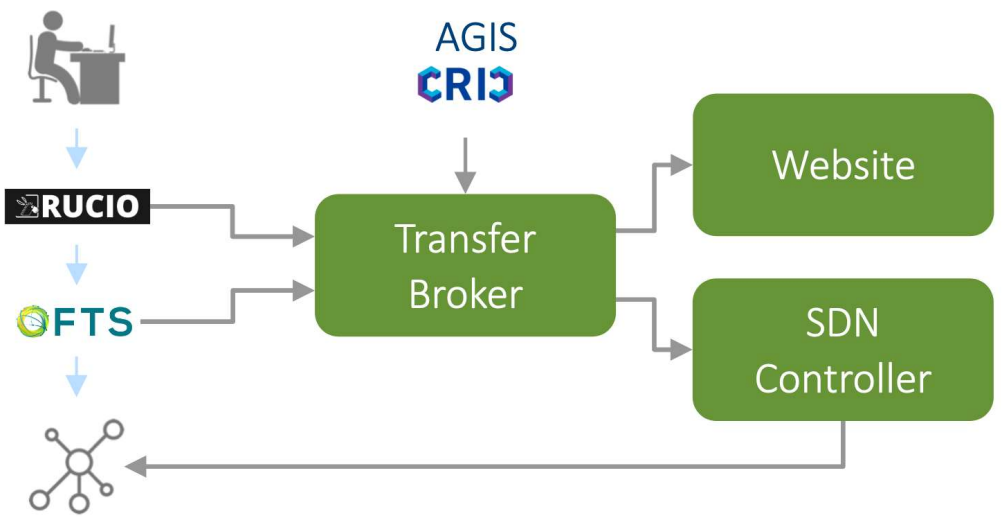
Transfer Broker

Prototype ready:

- Data provided by Rucio difficult to use to extrapolate information useful to make network optimizations
- Now using data provided by FTS via CERN Grafana: estimation of volume of upcoming data transfers and identification of source-destination storage elements.

Network information repository

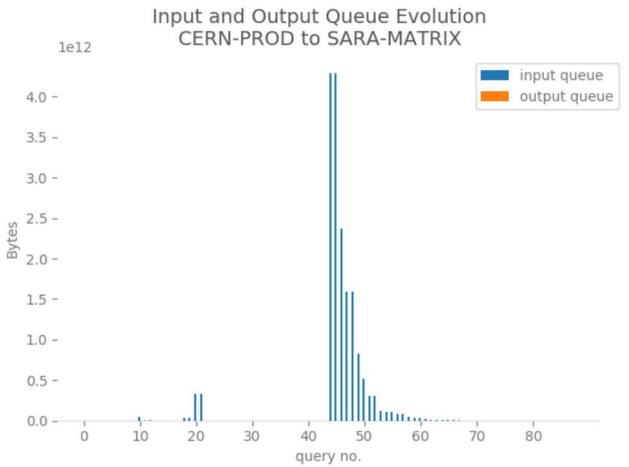
- CRIC (Computing Resource Information Catalog) being used to store IP prefixes of storage elements at sites



NOTED: simulations

- On 8th of October 2019 tested Transfer broker behavior when a large transfer between CERN and NLT1 was triggered

The Transfer Broker successfully observed how the Rucio queue fills up



- On 11th of December 2019 test will be repeated involving two Tier1s (NLT1 and DE-KIT) and **testing the full chain of components**:

FTS → Transfer Broker → SDN controller → LHCOPN/ONE load balancing

NOTED: next steps

- Coralie’s term is ending. Thank you very much for here very valuable contribution. A new Technical Student is being hired to continue the project
- Improve Transfer Broker: consolidate, consider more sources for detecting large transfers
- Evaluate other technologies for network optimizations

multiONE

Rationale

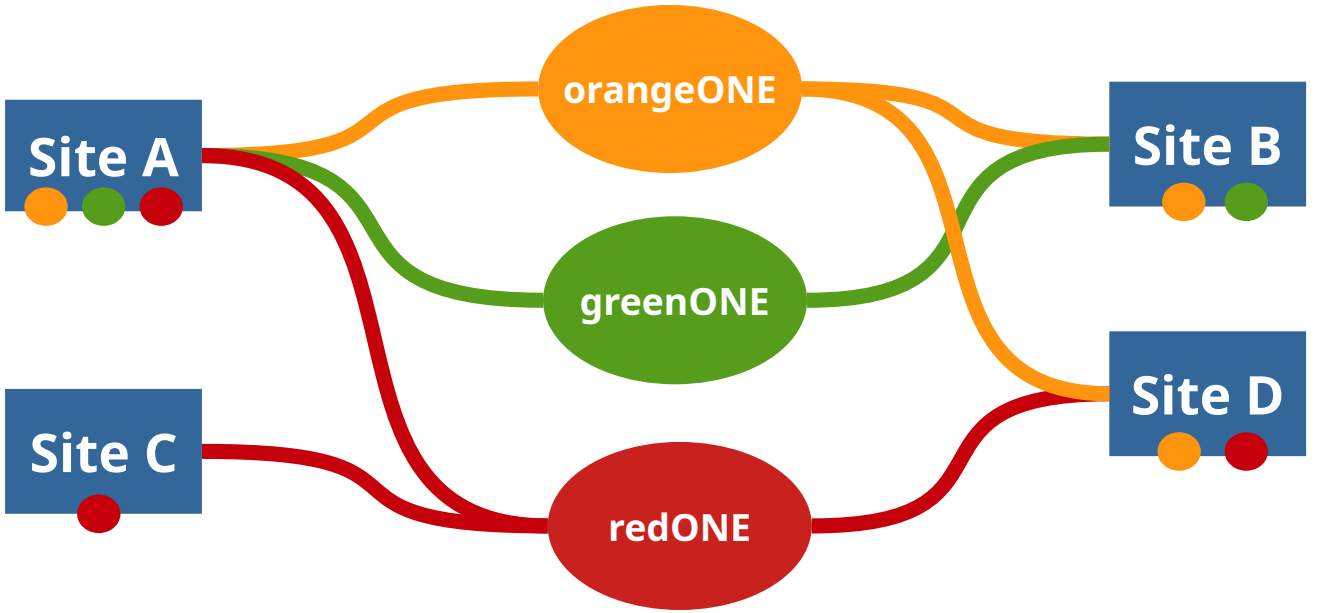
LHCONE community worried by the increasing number of collaborations using LHCONE (WLCG, BelleII, Pierre Auger oservatory, NovA, XENON): **connecting too many sites can undermine LHCONE's primary benefit** (its connection can be trusted and bypass slow firewalls)

Future major Collaborations should get their own "ONE" (VPN), but works need to be done to correctly route traffic from shared resources at sites participating to multiple Collaborations

multiONE

- Each site joins only the VPNs it is collaborating with, to reduce the exposure of their storage and computing resources

- Each Collaboration funds its own VPN



Issues with multiple VPNs

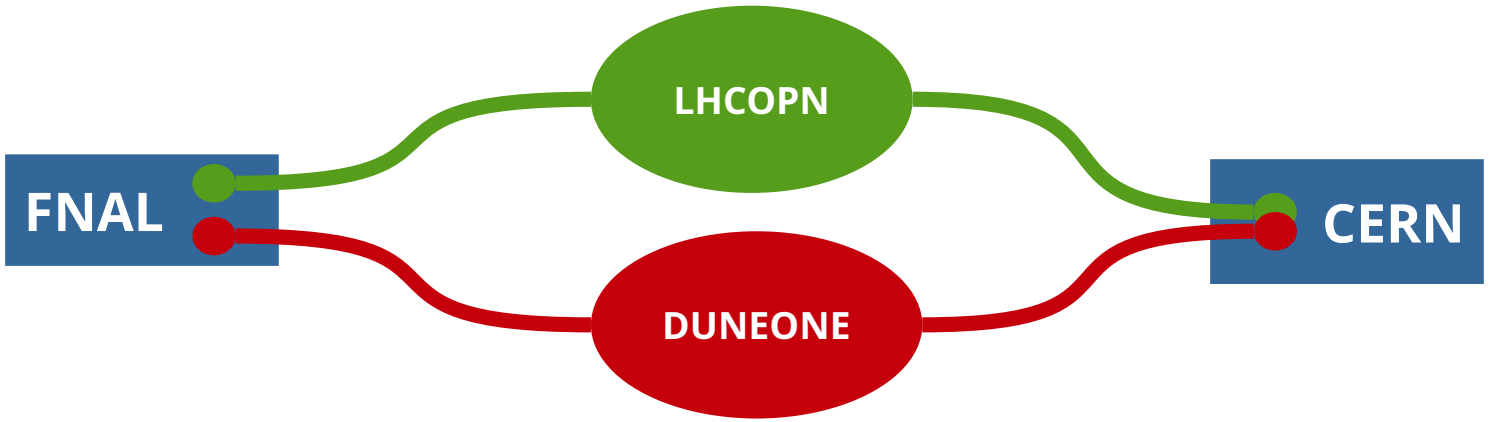
- Difficult to select what VPN to use for a Site that serves multiple Collaborations
- Even more difficult if the different Collaborations share the same servers and applications
- The simpler solution (static segregation of resources) is rather inefficient

multiONE will explore multiple solutions to efficiently separate data traffic for the different collaborations served by a site

LHC/protoDUNE use case

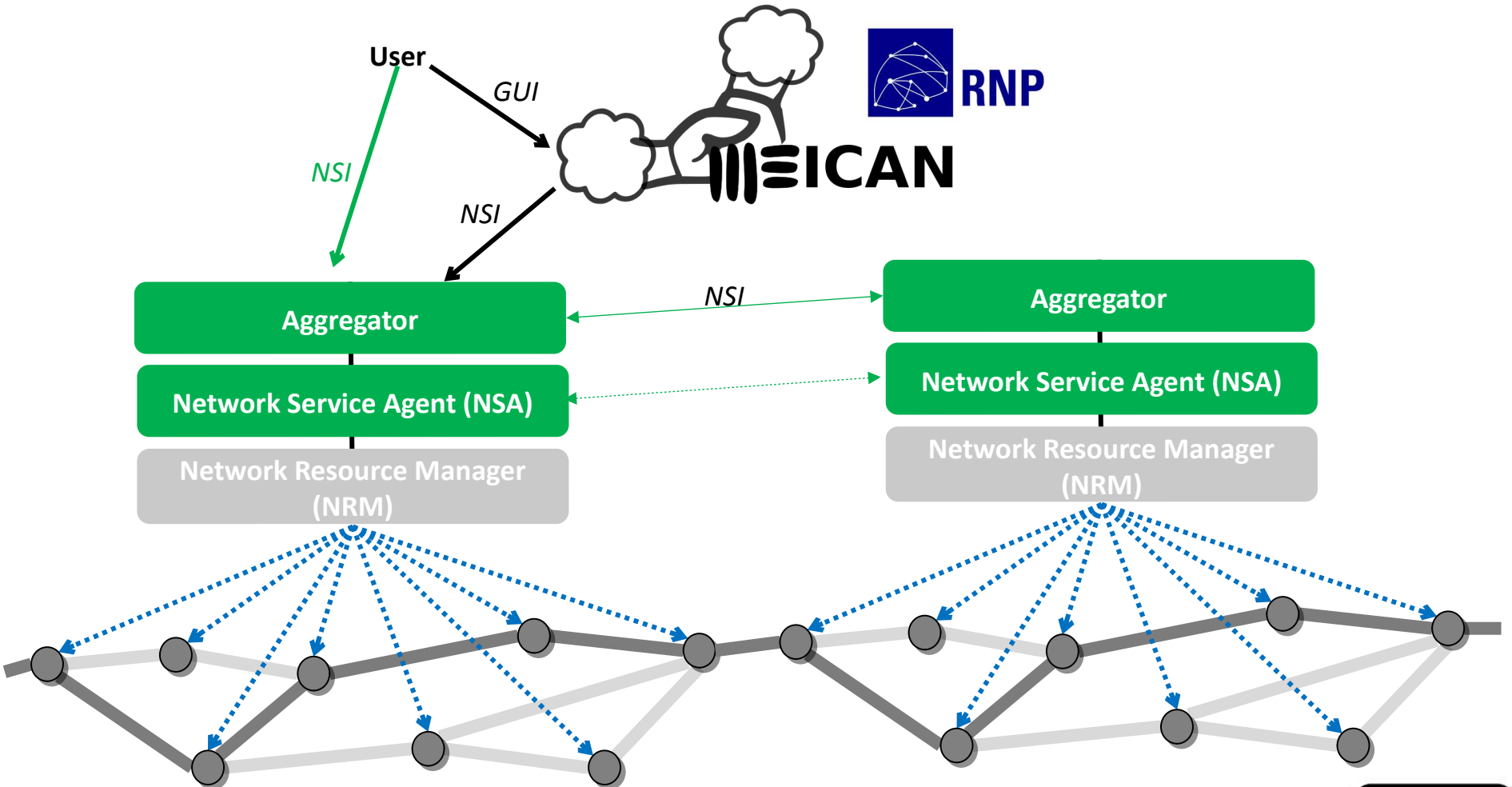
Agreed with FNAL to prototype the solution with protoDUNE between CERN and FNAL (protoDUNE is currently using the LHCOPN link of FNAL)

- 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



DTNs and Automated Provisioning Systems

AutoGOLE: multi-domain network services on-demand



AutoGOLE activities 2019

- Connected more networks around the globe using NSI *(the framework for inter-domain provisioning of connection-oriented services)*
- Started by segment CHIGAGO-MONTREAL-AMSTERDAM
- MEICAN software development by RNP
- Expanding AutoGOLE with connectivity to DTNs through SENSE

AutoGOLE activities 2020

- **Production**

  dynamic provisioning on ANA (Advanced North Atlantic)

- **Innovation**

  towards multi-resource provisioning

SENSE: SDN for E2E Networked Science at the Exascale

End-to-End (network point of view)

–DTN NIC to DTN NIC, across Science DMZ, WAN(s), Open exchange points (ideally)

Multi-domain

–Multiple administrative domains, independent policies and AUP

Provisioning automation

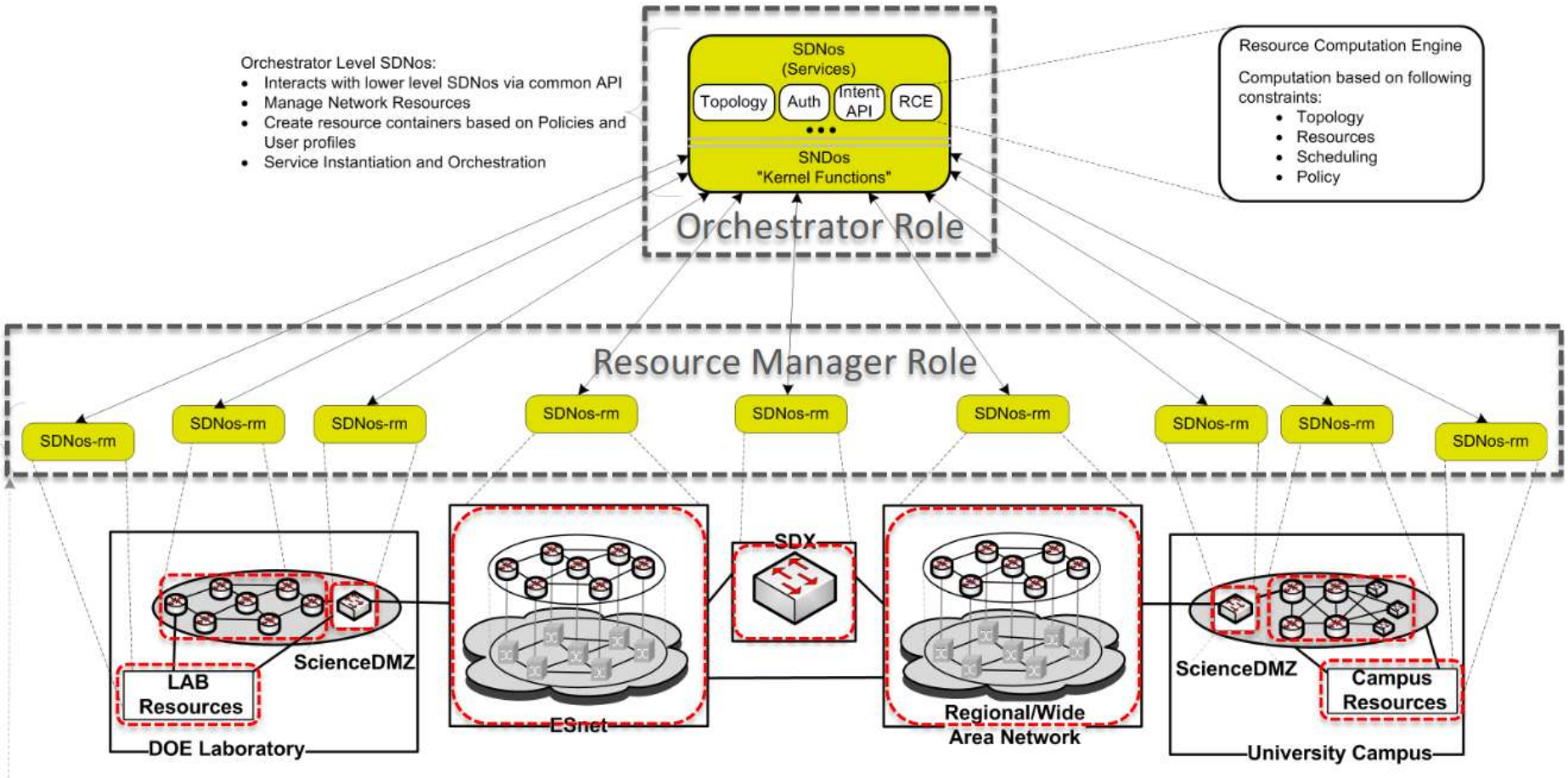
–Bring-up and management of services without interrupt-driven human involvement

Resource orchestration

–Allocation and reservation of resources including compute, storage and network

ESnet
ENERGY SCIENCES NETWORK

SENSE architecture and approach



Resource or Facility Specific SDNos
- Responsible for local resource of facility
- Implemenation system and technolgy a local decision
- Southbound APIs vary depending on resources/facility type
- Commmon Northbound API to be defined
- Resource descriptions based on extensions to NML

Defines Service Perimeter/Boundary

SDNos: SDN Operating System
SDNos-rm: SDN Operating System - Resource Manager

ESnet
ENERGY SCIENCES NETWORK

SANDIE: SDN Assisted Named Data Networking (NDN) for Data Intensive Experiments

NSF CC* Program: Integration and Innovation

CHALLENGES

- LHC program in HEP is world's largest data intensive application: handling One Exabyte by ~2018 at hundreds of sites
- Global data distribution, processing, access, analysis; large but limited computing, storage, network resources

APPROACH

- Use Named Data Networking (NDN) to redesign LHC HEP network; optimize workflow

SOLUTIONS + Deliverables

- Deploy NDN edge caches with SSDs & 40G/100G network interfaces at 7 sites; combine with larger core caches
- *Simultaneously optimize caching ("hot" datasets), forwarding, and congestion control* in both the network core and site edges
- Development of naming scheme and attributes for *fast access and efficient communication in HEP and other fields*

SCIENTIFIC and BROADER IMPACT

- Lay groundwork for an NDN-based data distribution and access system for data-intensive science fields
- Benefit user community through lowered costs, faster data access and standardized naming structures
- Engage next generation of scientists in emerging concepts of future Internet architectures for data intensive applications
- Advance, extend and test the NDN paradigm to encompass the most data intensive research applications of global extent

TEAM

- Northeastern (PI: Edmund Yeh), Caltech (PI: Harvey Newman) Colorado State (PI: Craig Partridge)
- In partnership with other LHC sites and the NDN project team

Northeastern

Demonstrations at Caltech Booth 543

NRE-019 – Global Petascale to Exascale Workflows for Data Intensive Science Accelerated by Next Generation Programmable SDN Architectures and Machine Learning Applications

NRE-019b FPGA-Accelerated Machine Learning [Caltech and 2CRSI] Inference for Trigger and Computing at LHC

NRE-013 – SENSE: Intelligent Network Services for Science Workflows
Layer2/3 Services, Full Lifecycle, Multi-Domain, Multi-Resource, Interactive, End-to-End

NRE-020 – LHC Multi-Resource, Multi-Domain Orchestration via AutoGOLE and SENSE

NRE-022 – Toward Unified Resource Discovery and Programming in Multi-Domain Networks

NRE-023 – International Data Transfer over AmLight Express and Protect (ExP) [Supporting LSST]

NRE-024 – 3 X 400GE Ring: Caltech-SCinet-Starlight/NRL with WAN Extensions to Starlight/iCAIR

NRE-035 – SANDIE: SDN-Assisted NDN for Data Intensive Experiments

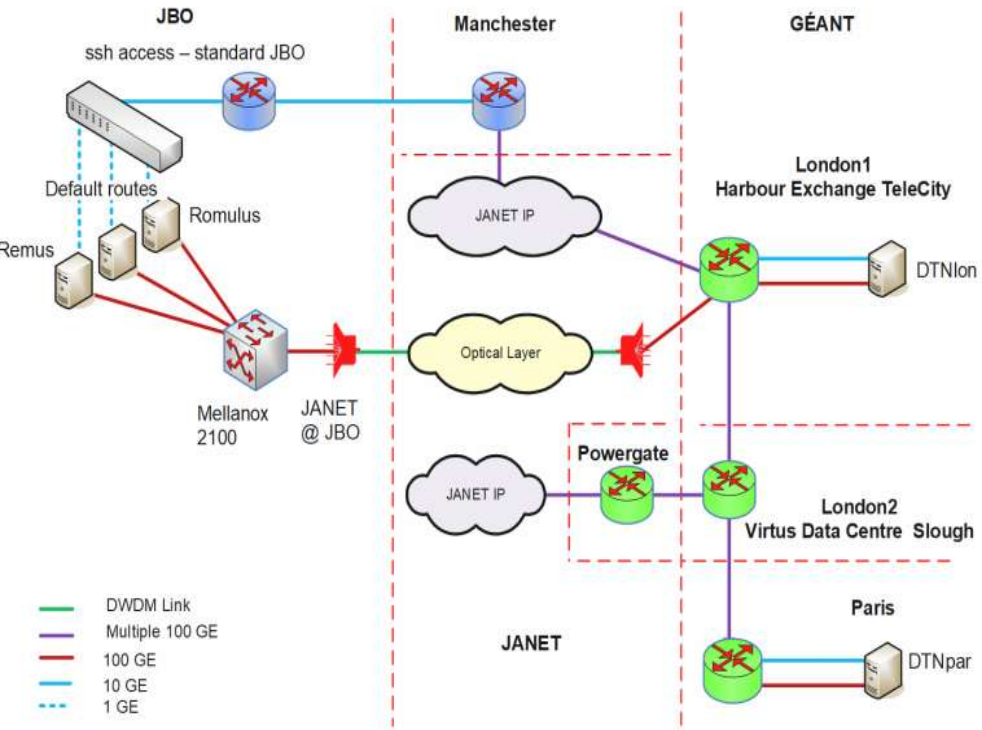
Low layers protocols alternatives

Low layers protocols alternatives

Activity in the AENEAS SKA projects

Exploring alternative transport protocols for very long RTTs data transfers

AENEAS Network Topology



What have we learnt?

- WebDAV/http(s) and xrd protocols both work well for moving bulk data.
- A simple client loop of “Get Chunk” - “Put Chunk” clearly reduces disk-to-disk throughput.
- Use of zero-copy e.g. sendfile() on the server gives a big improvement.
- Use of multiple parallel disk accesses is a help
- TCP will send what the app gives it – keep the socket full.
- TCP auto-tuning works well at medium RTT but may be slower at large RTT.

Next steps

- Check out v 4.11.0
- Look at multiple TCP flows

HEPiX NFV working group

NFV WG: update

- **Working on interim report (link) - need feedback on potential areas for future work with the experiments**

- Report focuses on highlighting important trends in networking - potentially critical to both data lakes and container/vm-based compute
  - Network disaggregation - open network environments becoming mainstream
  - **Cloud native networking** - rethinking network design of the DCs
  - **Network virtualisation** - report surveys a number of solutions that offer ways to build scalable, robust and cost-effective DC networks
  - DC edge services and hyper converged infrastructures and their potential impact on HEP
  - **Programmable WAN** projects focusing on the future of network provisioning and operations

- **Future work focuses on identifying potential areas that sites, experiments and R&Es could work on together**

- Dedicated session at the LHCONE/LHCOPN workshop in January 2020
  - Expecting feedback from the experiments on potential common projects in this area
  - We need to outline a plan for future networking deliverables and an associated timeline

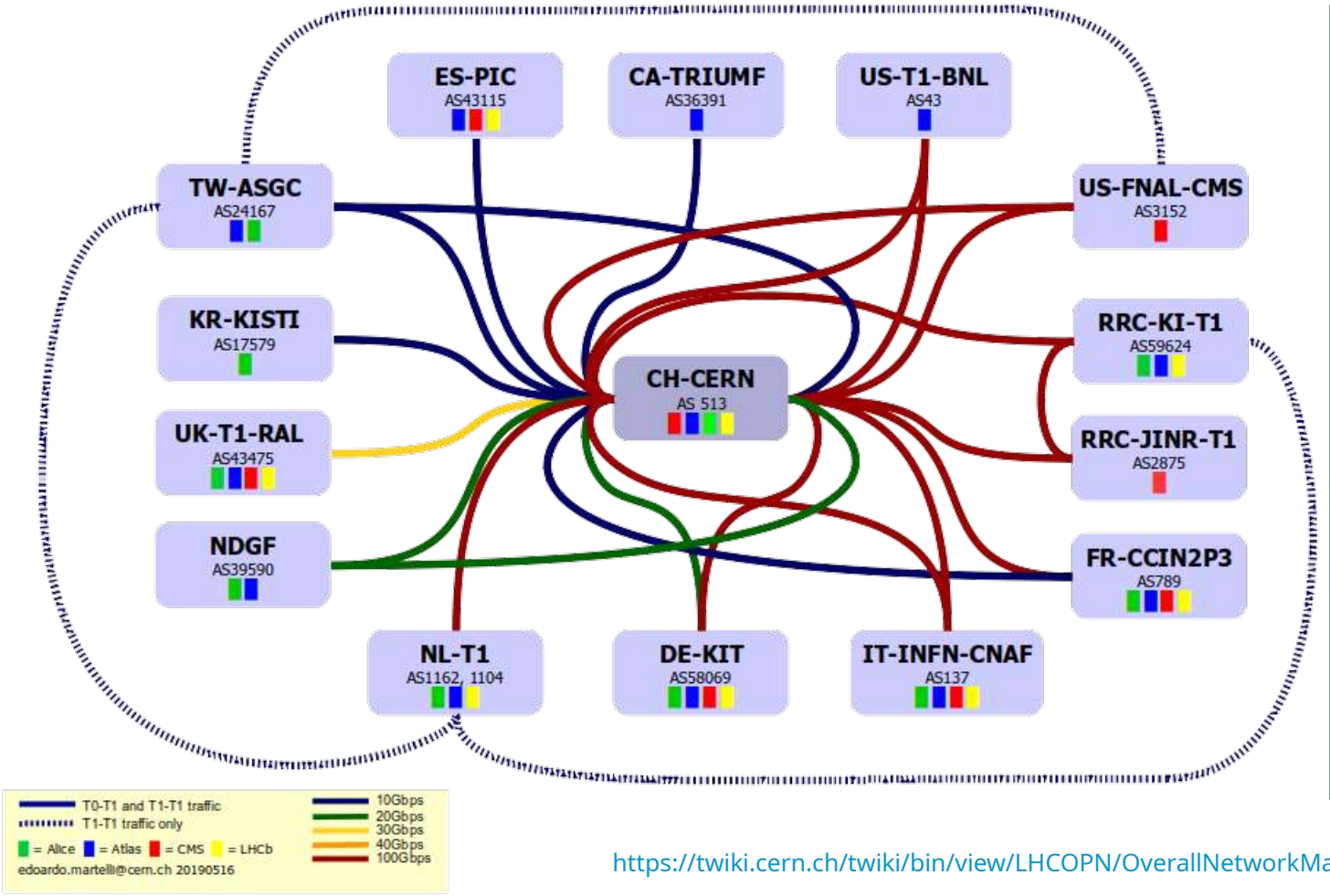
LHCOPN and LHCONE

LHCOPN latest deployments

Tier1s upgrading links to CERN Tier-0 to 100Gbps

- **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

LHCOPN



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%)

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

LHCONE last year deployments

LHCONE L3VPN: bigger sites upgrading connections to 100Gbps.

Few new sites joining in East Europe, Asia and South America

- 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

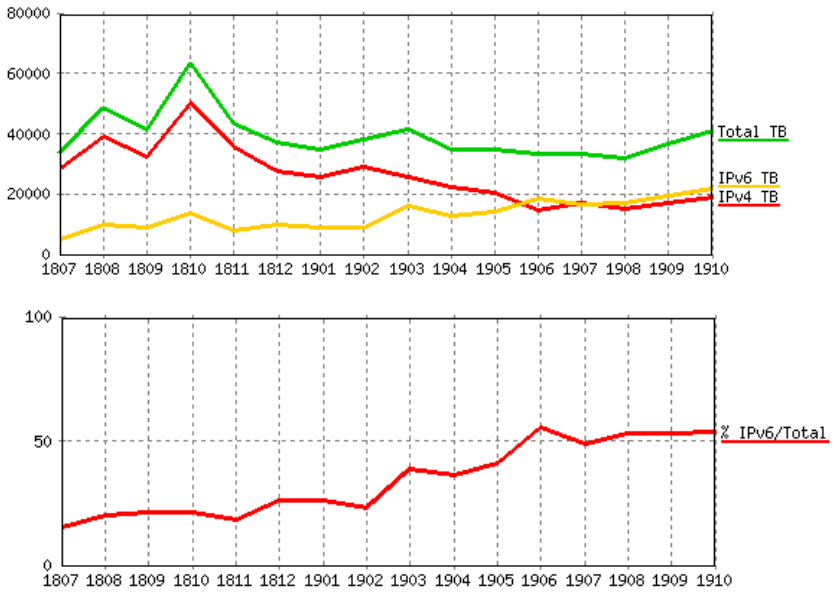
IPv6 adoption in LHCONE and LHCOPN

RIPE just completely run out of IPv4 addresses

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

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

Conclusions

Upcoming Meetings

Next LHCONE/OPN meeting with Experiments: 13-14 of January 2020 at CERN
https://indico.cern.ch/event/828520/

SIG-NGN meeting, Next Generation Networks for Science: 15-16 of January at CERN
https://wiki.geant.org/display/SIGNGN/4th+SIG-NGN+Meeting

HEPiX IPv6 Working group meeting: 16-17 of January at CERN
https://indico.cern.ch/event/855123/

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

References

NOTED

Presentation at HEPiX Fall 2019:
https://indico.cern.ch/event/810635/contributions/3592922/attachments/1926417/3188957/presentation_hepix.pdf
Presentation at CHEP 2019:
https://indico.cern.ch/event/773049/contributions/3473789/attachments/1932599/3211841/noted_CHEP.pdf

multiONE:

https://indico.cern.ch/event/739882/contributions/3520004/attachments/1906199/3148167/EM-multiONE-GDB.pdf

SENSE

http://sense.es.net/

NFV working group

WG meetings and notes: https://indico.cern.ch/category/10031/
F2F meeting: https://indico.cern.ch/event/725706/
White Paper: https://docs.google.com/document/d/1w7XUPxE23DJXn--j-M3KvXlfXHUnYgsVUhBpKFjyjUQ/edit?usp=sharing

Summary of Caltech activities:

https://www.dropbox.com/s/78w7rsjz5gyzezh/SC19NextGenCyberwithAiV3_hbn111919.pptx?dl=0

Questions?

edoardo.martelli@cern.ch