multiONE

GDB at FNAL - 11th of September 2019

Edoardo Martelli , Tony Cass – CERN IT-CS

LHCONE limitations and the need for multiple “ONEs” (multiONE)

LHCONE

LHCONE is a worldwide Virtual Private Network (VPN) implemented by Research and Education Network providers (RENs)

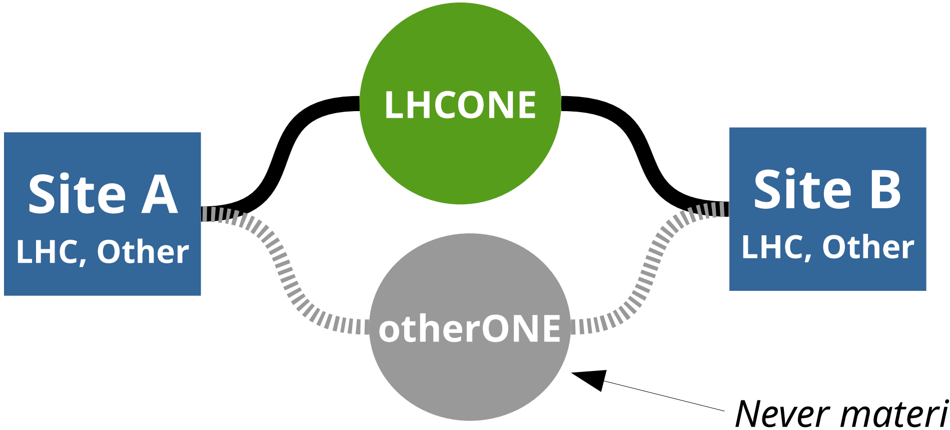
Original AUP:
- **Only WLCG** Tier1/2/3 sites can connect to LHCONE
- At sites, only resources dedicated to WLCG can use LHCONE

Main advantage: **Sites can trust LHCONE to be safe and plug it directly into their datacentre**, bypassing bottlenecks (e.g. expensive security equipment)

Adding more Collaborations

Other Collaborations would benefit of having their own “ONE”
- in fact, defining a new VPN is relatively simple for RENs;
- but it’s difficult for Sites participating in multiple collaborations to put the traffic in the corresponding VPN

Thus, over the years, few HEP collaborations (Pierre Auger, XENON, BelleII, NOVA…) have simply joined LHCONE instead of building their own VPN



Never materialized

Problems with just adding Collaborations

- The more sites join LHCONE, the **less trustable** it becomes

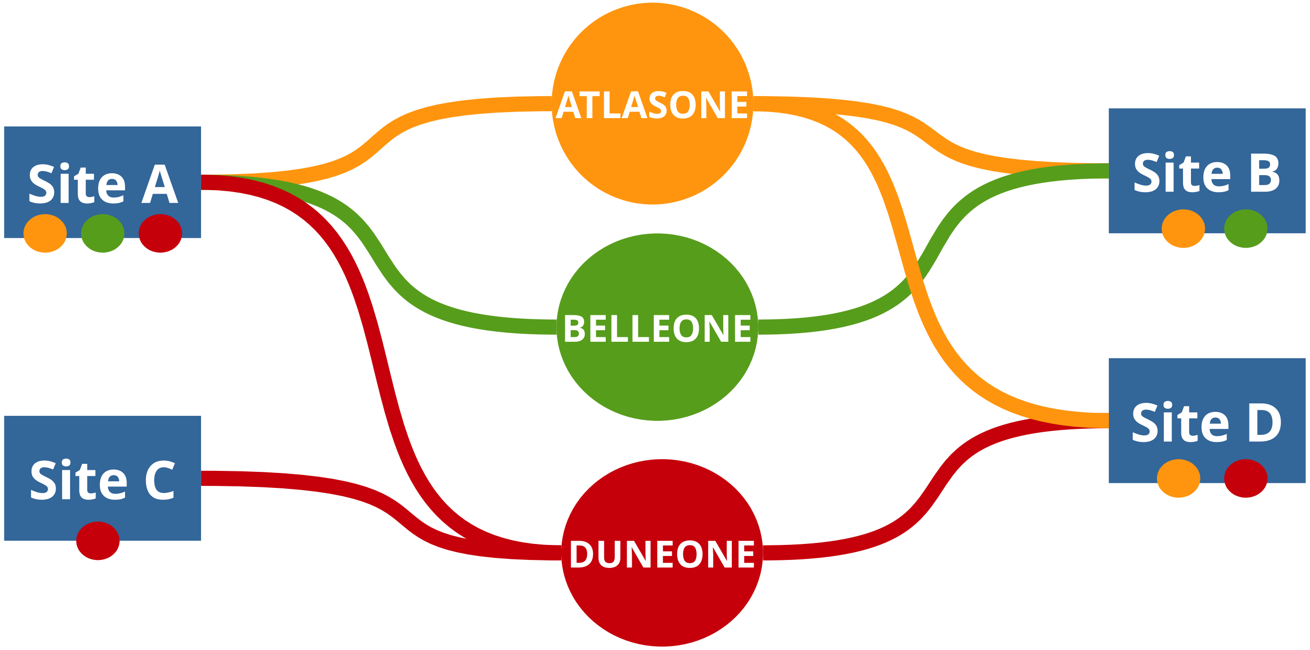
- The more the traffic volume grows in a single domain, the more difficult for RENs to shape the load in their networks

- Funding agencies would prefer to have a clear distinction of **who is using the resources they fund** (in fact it was not always straightforward to accept new collaborations in LHCONE)

multiple “ONEs”

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



But there are 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

About multiONE

Issue discussed several time at LHCONE meeting

Agreed to start a project to verify if it is possible to use multiple VPNs for sites that participate to several science Collaborations

Discussion on going to identify a Collaboration to prototype a working solution. Just been agreed with FNAL to start with protoDUNE (currently using LHCOPN)

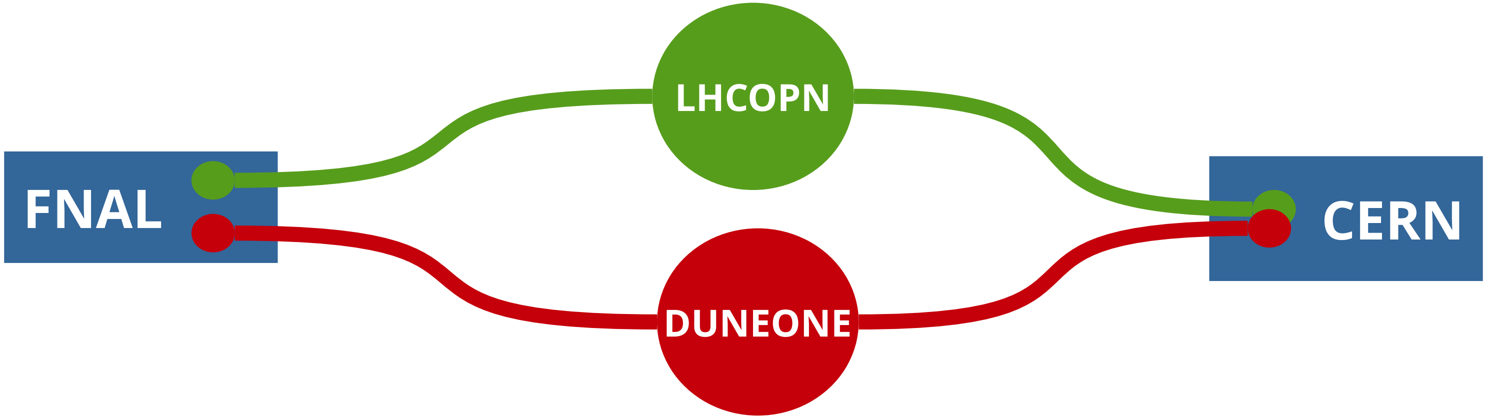
LHC-protoDUNE use case

Just 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



Possible solutions and considerations

Some possible solutions

Agile segregation of resources

- virtual domains by using eVPN/VLAN/VXLAN…
- software defined resource allocation, to avoid static segregation

Policy based routing

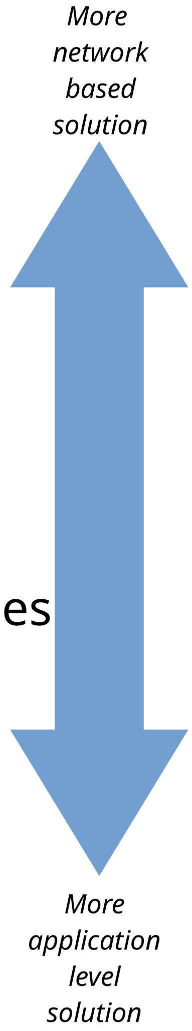
- Policy route the traffic to put it in the right VPN

Correspondent source-destination addresses

- clients and servers must use corresponding source and destination addresses
- normal routing will handle the separation

Packet tagging:

- applications tag the packets (DSCP?) with owner information
- the network policy routes the tagged packets in the corresponding VPN


More network based solution
More application level solution

Storage considerations

EOS at CERN

- already separated clusters and hardware for four major LHC experiments (but mixed together in the same datacentre network)
- shared EOSPUBLIC cluster for smaller experiments

In general

- General trend: drop gateways, drop GridFTP for xrootd
- Other different storage systems are used by other sites

More storage considerations

On using tagging or src-dst IPs:

- XRootD is heavily based on redirections
- Changing redirection logic (tagging, multipleIP,...) needs changes in both XRootD and Storage backend layers. Also needs overall support from other storage implementations
- Needs proper client support and configuration

==> technically complex and uncertain if possible at all

CPU considerations

Properties to be preserved:

- use of a batch system (i.e. HTCondor) that assign capacity to different Experiments, but allow to share spare resources.

- possibility to run different jobs from different Experiments on the same physical server

Avoid loss of efficiency:

- it must not be necessary to drain the resources that run the jobs

More CPU considerations

Setting of DSCP field: doable, but requires to control all the clients and the applications

Linux Network Namespace (aka VRFs on Linux) can help for the traffic separation.

- It could means multiple vNICs, one per served Experiment.
- Or controller managed VXLANs.

Docker (and probably Kubernetes) can make a Network Namespace for the job it starts. HTCondor could be changed to do it (development required)

Conclusions

Why do we want to do it?

- No urgent need nor specific request right now
- However size of LHCONE is already at its limits
- Most of all we need to be prepared when the next major collaboration (SKA?) will need its own ONE

Next steps

- Explore more technical implementation
- Prototype different solutions
- Test solutions with a site that serves multiple experiments
    - Implementation of DUNEONE between CERN and FNAL

Questions?

edoardo.martelli@cern.ch