pre-GDB on data-centre network architectures

Summary report

GDB 14th July 2021

Stefano Zani (INFN CNAF) and Edoardo Martelli (CERN)

Workshop on Data-Centre network architecture

An initiative proposed at the last LHCOPN-LHCONE meeting

183 registered people, peaked at ~100 connected people

2 sessions of 3 hours on Zoom

11 speakers

Agenda and presentations: https://indico.cern.ch/event/1028690/

The Nightmare of Securing a Multi-Purpose Data Centre

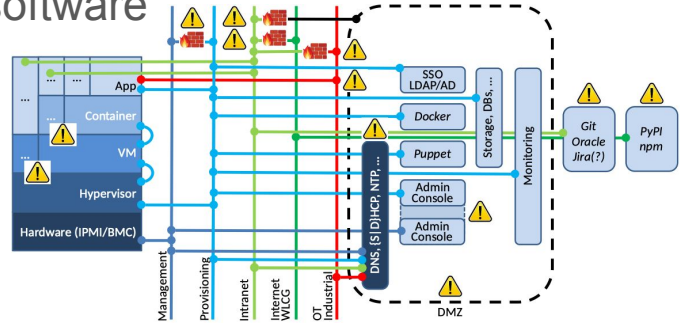
Presented by Stefan Lueders, Head of Security at CERN (CH)

Some considerations on the increasing difficulty of securing modern data centre environments, where abstraction layers and complexity are ever increasing.

Multi-purpose applications inside the DC and dependencies from common services makes extremely difficult the network segregation.

Recommandations:

- Keep it Simple and avoid over complication
- Defense in depth and at every level of hardware and software
- Network segregation and compartmentalization



The RARE routing platform and its use in data-centres

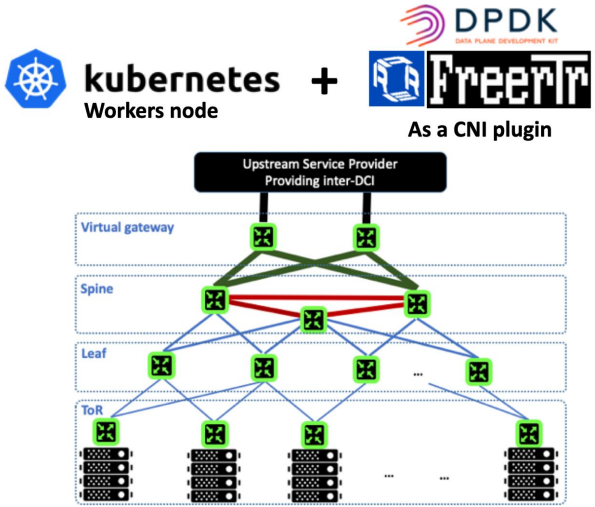
Presented by Frederic Loui, network expert at RENATER (FR)

The RARE project (GEANT) has developed the FreeRTR routing software platform which ties control plane and data plane programming

It can be used on switches with P4 programmable data plane, like the Intel Tofino ones

The project provides a P4 Lab where users can test their apps

RARE has been initially developed for Service Providers, but it can now be used also on data-centre networks, on ToR switches and Leaf-Spine routers



The CERN data-centre network

Presented by Vincent Ducret, network expert at CERN (CH)

The CERN data-centre network connects servers in 3 large rooms, with a Leaf-Spine IP fabric.

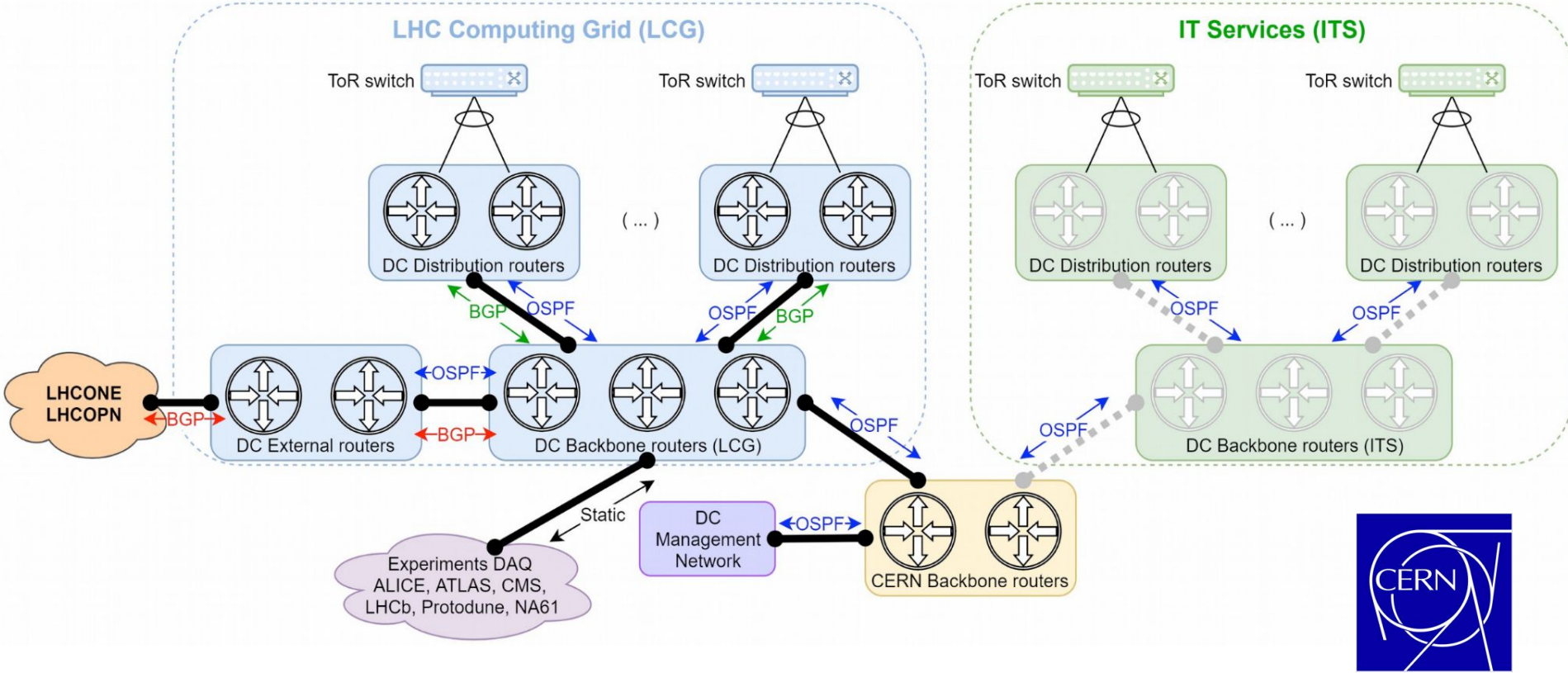
The network is divided in two virtual domains implemented using VRF-lite: ITS for general services; LCG for WLCG services with access to LHCOPN/ONE

ToR switches are connected with a LAG to two Leaf routers configured with VXLAN ESI. OSPF is used for the underlay network, BGP EVPN for the overlay

A new datacentre will be built soon. Extensive use of 400G links is expected there

DC Network design – Routing protocols

- **BGP for LHCONE/OPN + BGP for EVPN/VxLAN "overlay"**



CERN data-centre virtualization services

Presented by Ricardo Rocha, virtualization technology expert at CERN (CH)

SDN project to allow IP mobility and floating virtual IPs

Use cases: VM migrations, Load Balancer aaS, Isolation, Firewall aaS

Using Tungsten Fabric as controller, integrated with Openstack

Upcoming: VXLAN overlay networks with VTEPs on hypervisors and L3 gateways on Juniper routers, to ensure line rate performances

The AGLT2 data-centre network

Presented by Shawn McKee, network and IT expert at Univ. Michigan (USA)

The AGLT2 network is undergoing a major hardware upgrade

It will support higher speeds and network function virtualization, ready for Run3 requirements and allowing prototyping for Run4

Configurations will be managed with Ansible and Github

Developing an enhanced monitoring infrastructure to allow users to understand network performances

This new network will allow prototyping of the development promoted by the Research Networking Technical WG

ATLAS Great Lakes Tier 2
UNIVERSITY OF MICHIGAN AGLT2 MICHIGAN STATE UNIVERSITY

SURF NLT1 data-centre network

Presented by Sander Boele, network and IT expert at SURF NLT1 (NL)

Using NVIDIA/Mellanox switches with Cumulus Linux

Implemented with BGP-EVPN over VXLAN

Using automation to manage configuration of network devices: CMT, Gitlab, Patchmanager, Ansible

Testing 400Gbps internally between SURF and NIKHEF, and on the 400Gbps link to CERN

KISTI GSDC data-centre network architecture

Presented by Jin Kin, network expert at KISTI (KR)

Data-centre supporting ALICE, CMS, BelleII, LIGO and other national experiments

IP Fabric architecture, especially designed to better support Docker

Using eBGP for underlay routing, one ASn per device. Mapping rack location into IP addresses used to interconnect devices

Management tools: Puppet, Foreman, Ansible, Calico

The UK RAL data-centre network

Presented by Jonathan Churchill, network expert at STFC RAL (UK)

Data-centre network supporting several experiments and services, with low latency and high bandwidth.

Experiments have assigned resources in dedicated network PODs.

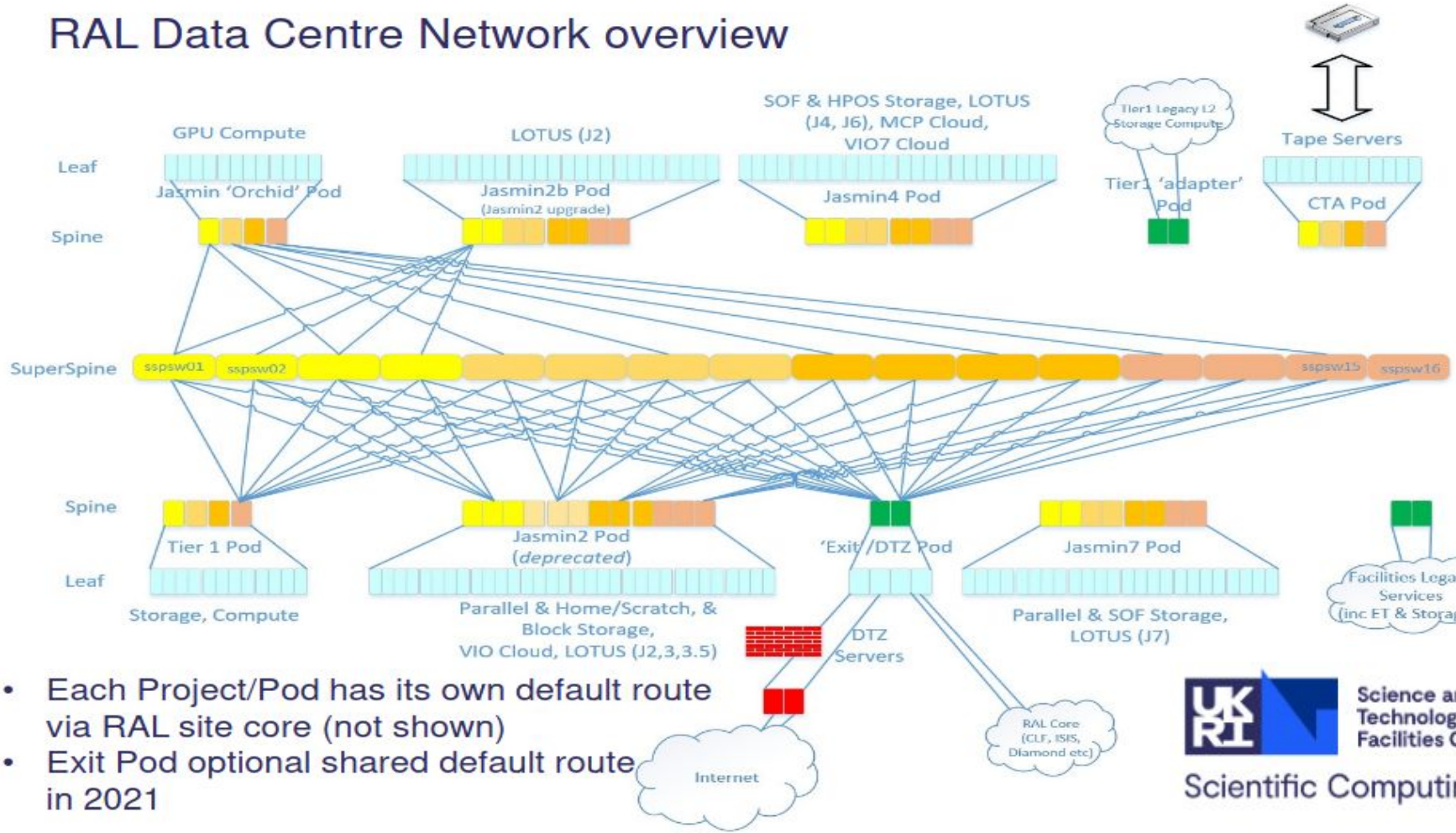
PODs Architecture: IP Fabric, spine-leaf architecture built using cheaper fix-form-factor switches running Cumulus Linux and Onyx.

PODs are interconnected through transit super-spine. PODs may have different default routes

UKRI Science and Technology Facilities Council

Scientific Computing

RAL Data Centre Network overview



- Each Project/Pod has its own default route via RAL site core (not shown)
- Exit Pod optional shared default route in 2021

Science and Technology Facilities Council

Scientific Computing

BNL SDCC data-centre networks

Presented by Alexandr Zaytsev, network expert at BNL (US)

Large data-centre hosting HTC and HPC, serving multiple collaborations

Building a new data-centre to host 478 racks with 9.6MW of power. Will use Arista large chassis for interconnections

Architecture: Leaf-spine IP-Fabric with eBGP for the underlay

Brookhaven National Laboratory

FNAL Network Architecture, implementation experience

Presented by Andrey Bobyshev, network expert at FNAL (USA)

Computing resources distributed in two data-centres and other smaller computing facilities

Leaf Spine topology, using classic VLANs and STP.

Considering to use IP fabric and VXLAN to extend VLANs to remote locations

The network is divided in Network Modules, each assigned to a specific experiment or location.
Modules are interconnected via a transit network where security policies are applied

Fermilab

Requirements for DC supporting multi-petabyte transactions

Presented by Harvey Newman, professor at Caltech (USA)

Overview of network R&D projects aiming to fulfill the demanding requirements of LHC Run4:

- Storage caches (SoCal)
- Bandwidth on Demand (AutoGOLE, SENSE, NOTED)
- 800Gbps Ethernet
- Intelligent Data Plane opportunity using P4 (RARE)

Conclusions

Datacenter Networks have to conjugate:

- **Performance:** to match the challenging I/O experiments requirements
- **Flexibility:** to make the datacenter more dynamically reconfigurable
- **Security:** to guarantee data safety and service availability

400 Gb Ethernet technology for brief distance (<100m) is not yet “Commodity” and has to be carefully “tracked” (Cabling and optics costs are relevant).

VM and container networking have to be designed consistently with the physical network infrastructure and this is one of the main topic to work on in next years

Actions

- Propose follow-up workshop to share knowledge and experiences
- Continue the activity in the context of the LHCONE community

Thank you!

Any question?