CERN Campus Network upgrade: requirements and plans
Authors: Carles Kishimoto (CERN IT/CS), Tony Cass (CERN IT/CS)

Abstract
The Campus Network is CERN's general purpose networking workhorse. It supports CERN's office networking requirements, the Wi-Fi infrastructure and a rapidly growing number of connected devices, connects dedicated experiment networks to other CERN services and even carries physics data from smaller experiments to the data centre. The Campus Network provides full dual-stack IPv4 and IPv6 connectivity and enables users to protect devices from unwanted connections from the general Internet where required.

The Campus Network has evolved organically since the last major upgrade in 2006 and a review was organised in 2018 to prepare for a long-overdue refresh to ensure the network can continue to meet evolving demands in the coming decade.

This document summarises the current state of the network, documents the limitations and problems we experience, documents the future requirements of specialist users and sets out the options we foresee for the network upgrade.

Today's Campus Network
The campus network, as shown in Figure 1, follows a 3-tier hierarchical model with well-defined access, distribution and backbone layers. Given the scale of CERN's campus, a fourth aggregation layer was added to optimise optical fibre usage by concentrating the optical fibre infrastructure at six geographically convenient points. Today, more than one hundred routers in the distribution layer provide connectivity to over 2,000 switches in the access layer. Figure 2 shows the relationship between the campus network and other network domains at CERN.

The last major upgrade of the campus network was in 2006 when HP Procurve routers and switches were deployed to replace the previous hardware. A minor, but still significant, upgrade took place in 2014 when the aggregation and backbone routers were replaced by more powerful Brocade MLXe routers. This upgrade was necessary both to address performance degradation due to the size of the network and growth in use and to better support IPv6, a protocol that HP routers do not support in hardware.

For network professionals, the CERN campus network is a dual-stack routed network using only OSPFv2/v3 as IGP. No vlans are used in the access layer and the uplinks are deployed at 1G with all switch ports configured untagged in the default vlan. In the distribution layer, all routers are connected via two uplinks to the aggregation router via 1G links. The interconnection of the aggregation to the backbone layer is via 10G links. The campus network is a single OSPF area with the two Brocade MLXe routers deployed in the backbone acting as ABR routers and summarizing all GPN prefixes towards the core of the network.

The routers currently used at the aggregation layer have not been available for purchase since 2015 and no hardware or software support will be provided by the manufacturer after 2020. Although we have sufficient spares to continue operation for some while, continuing with the current hardware much beyond 2020 would not be sensible even if the routers could in principle meet the service requirements.
Campus Network services

The key services provided today in the campus are the following:

- Ethernet connections at up to 1Gbps\(^1\);
- high-density Wi-Fi coverage for office buildings and key outdoor areas using the latest standard (IEEE 802.11ac) supporting roaming and a guest Wi-Fi service with a self-registration portal;

---

\(^1\) Connection speeds are limited by the cable infrastructure. Most buildings have CAT5e and CAT6 cabling although newer buildings are equipped with CAT6A allowing speeds of up to 10Gbps. A dedicated CAT6A infrastructure has recently been installed to support the Wi-Fi infrastructure.
a choice between public and private IP addresses. Devices with private IP addresses cannot access, and are not accessible from, the internet and are therefore protected from direct external attack;
- the possibility, for more fine-grained security, for users to define groups of devices (“control sets”). These control sets are used to authorise/deny traffic flowing between domains using user-configurable access lists on the router interfaces (the “Gates” shown in Figure 2);
- anti-spoofing access-lists to prevent malicious users or misconfigured devices using random source IP addresses.

In terms of networking:
- all routers in the campus are dual stack IPv4 / IPv6;
- loop protection is enabled in both access switches and distribution routers to prevent a loop accidentally created by a user of cutting the connectivity of the whole service. Only the problematic port will be disabled automatically by the network device.
- Protocol Independent Multicast (PIM) is deployed to provide routing of multicast traffic.

Current limitations
The ageing campus network routers are no longer able to deliver the performance necessary to support the expectations of our users, notably in terms of capacity and performance. Areas where we regularly experience issues include the following.
- Bandwidth: The increasing data rates from some experiments connected to the campus requires increasing the uplink bandwidth from the access and distribution layers to avoid adversely affecting other users. The limited number of 10GbE ports on the current routers and the lack of support for higher bandwidth technologies such as 40GbE or 100GbE limits our ability to provide the necessary uplink bandwidth.
- ACLs: The size of the TCAM\textsuperscript{2} in the HP routers is too small to support the number of ACL entries required for some experiments.
- Router Performance: The router hardware has problems coping even with the number of access list entries that can be accommodated in the TCAM. Problems are often caused by poor performance caused by overload during routine operations such as ACL updates.
- Device isolation: Although private addresses and the use of access control lists and gates provide a reasonable level of protection, devices are still vulnerable to attack. This is of particularly concern for the growing number of connected or “Internet of Things” devices that often have poor security features and limited upgradability. Providing a virtual “IoT” domain on top of the campus network would enable a valuable separation between IoT and end-user devices. Unfortunately, the current routers do not support the necessary network protocols—nor will they ever do so as the manufacturer now provides only software fixes, not upgrades.

\textsuperscript{2} Ternary Content Addressable Memory: specialised memory for storing access control lists to ensure fast and efficient lookups.
Towards a future Campus Network

To gather user requirements for the next 10-15 years a Campus Network Working Group was organised with members from the EN, BE and SMB departments representing key network users and service providers. Several meetings were held allowing each of the groups to present a list of requirements in terms of new networking needs and the limitations in the current architecture that they would like to be lifted.

User requirements

To a large extent, the input from the Campus Network Working group reflected our internal assessment of the points to address, notably in terms of the ability to support connection speeds of at least 10GbE and to support larger ACLs. The group also confirmed our assessment of the importance of support for connected devices and was particularly helpful in clarifying the requirement for device isolation and the need for PoE support.

- Device isolation: The working group strongly supported the requirement to isolate IoT devices from end-user devices. Representatives, however, did not consider that there is a need to create multiple “IoT” networks to separate the connected devices of different communities; a single such logical network was considered sufficient.

- PoE: users would like to have support for Power over Ethernet, not just in terms of the ability to connect devices that are powered in this way, but also to have the ability to control the connection, i.e. the ability to power on or power off devices remotely. Representatives on the working group did not consider that there will be a need to provide a large-scale PoE infrastructure, however.

One other requirement expressed by the working group was the ability to provide Technical Network connections on arbitrary network sockets around the campus. As the technical network infrastructure is entirely separate at present, the cost to deliver a technical network connection in a building with no Technical Network coverage is extremely high as an additional router and switch must be installed.

A key input from the group was the clear statement that there is no requirement for a significant increase in the number of devices to be connected to the wired network in any move towards “smart buildings”. Instead, smart sensors and actuators are expected to be connected to, and accessed via, dedicated controllers.

Architecture

One option to support the requested larger ACLs is to move away from the current fully-routed 3-tier model and span VLANs from the distribution layer up to the aggregation layer. In such a configuration, switches could replace the routers in the distribution layer and the money saved used to purchase more expensive routers with extensive TCAM for the aggregation layer.

This design, however, has three drawbacks:

- increased criticality of the devices deployed in the aggregation layer. All Layer 3 features such as ACL protection, DHCP relay, InterVLAN-routing and routing updates will happen in these routers. A simple configuration error or bug in the software would affect a large number of users at CERN, something we would like to avoid.

- decreased flexibility: in a routed network, capacity can always be added by adding more links, something that cannot be done in a non-routed (Layer2) network; and

- inefficiency: traffic for data transfers between servers in the same building is routed up to the aggregation layer.
Fundamentally, however, the limitations we see come not from the routed architecture, but from the limited capacity of the routers deployed. As the routers anyway need replacing as they are reaching the end of their operational lifetime, the problems we face today are clearly better addressed by deploying up-to-date hardware rather than by changing the network architecture.

The future Campus Network: current plans

The current plan is thus to simply replace the backbone and aggregation routers with newer devices which can provide higher speeds and which have more TCAM to accommodate larger ACLs. These routers can then be configured to implement multiple virtual domains. Based on the input received from the campus network planning group, only two or three virtual domains are foreseen at present—the default domain for most devices and one or two domains for devices that need an extra level of protection. Fortunately, we can easily implement such a small number of virtual domains using VRF-lite as we already demonstrated in the data centre. There is thus no need for any developments or upgrades to our tools and procedures. Should more domains be needed later we could deploy EVPN/VXLAN. These protocols could be deployed on top of VRF-lite and the existing domains be migrated whenever our configuration tools are ready. Work is already underway to prepare for the deployment of these protocols in the data centre so we will be ready for the use of EVPN/VXLAN in the campus network should this ever be necessary.

Router options

The choice of router for the backbone and aggregation layers is fairly easy. The key requirements for these routers, in addition to sufficient TCAM, are support for high bandwidth ports, mainly 40GbE and 100GbE, and a high port density. Both requirements are met by the routers that are being deployed in the Data Centre; deploying these in the campus allows us to reuse our existing operational procedures. Juniper QFX10002-36 routers will be used for the backbone layer with QFX10002-72 routers in the aggregation layer. Both support 10GbE, 40GbE and 100GbE interfaces; we will use 40GbE links initially but will easily be able to scale out by replacing these with 100GbE links in the future if necessary.

The choice of router for the distribution layer is more problematic. As there are many more of these devices cost becomes a significant factor in the choice. Additionally, requirements are not uniform across the site as some buildings, notably those with users from the EP department, have users with high bandwidth requirements. Whilst we would prefer, for operational reasons, to deploy a single type of router everywhere in the campus, that might not be the most cost effective solution and it might be necessary to select multiple potential distribution layer routers.

The following potentially suitable routers from manufacturers with whom CERN has an ongoing supply contract have been considered.

- Ruckus ICX 7750: This seems a bit expensive due to the maintenance cost. Support for VXLAN was lacking in the past but it seems support has been added in the latest release. Several switches of this type are already deployed in the data centre.
- Juniper QFX5100: The main problem of this router is that it has no support for IPv6 egress ACLs so we favoured the next version based on a different Broadcom chipset (Trident3) with improved the ACL capabilities. This is the …
- … Juniper QFX5120, announced in July 2018 and which seems a good candidate. It supports tri-speed interfaces (1/10/25GbE) in all 48 ports and provides 40/100GbE interfaces on the uplink ports. Support for VXLAN/EVPN is available although requires purchasing an additional software license.
- Juniper EX4650: Announced a little later than the QFX5120, the EX4650 is on paper identical and has the advantage of a perpetual software licence (that for the QFX5120
must be renewed every year). However, as the EX family is targeted at the campus rather than the data centre market, the type of features supported in the future may differ. In any case this should not be a problem as the core set of features will be kept in both platforms, including support for VXLAN/EVPN as these protocols are being adopted in both data centre and campus domains.

- Ruckus ICX 7850: This is the new Ruckus router based on a Trident3 chipset with larger TCAMs so competing directly with the Juniper QFX 5120 and expected to be available in the first quarter of 2019.
- Juniper EX9208: This is the router that will be deployed in the Technical network. However, the high price per port precludes use in the campus network
- Juniper QFX10002: The problem with this platform is the lack of support for 1GbE ports. Deploying this device would thus force us to replace existing switches that only have 1GbE uplinks with ones that have 10GbE interfaces

At present we are waiting for the Ruckus ICX 7850 to become available so we can conduct our tests and compare it, both technically and economically, to the Juniper EX4650.

Emerging technologies

One of the limitations mentioned in the Campus Network Working Group is the inability today of deploying a few random connections to the Technical Network. Today, even a single connection in a building requires the deployment of a full starpoint, i.e. a router and a switch, which leads to a significant cost for the first such connection, even if subsequent connections have a low marginal cost. Fibre-to-the-Office, or FTTO, is a technology that perhaps has the potential to address this issue.

FTTO uses time division multiplexing to send multiple simultaneous data streams along an optical fibre to terminal devices—Optical Network Terminals or ONTs—that extract the relevant data stream; end-user devices are connected to the ONTs with standard UTP Ethernet cables. As data transmission is over fibre, the ONTs can be many kilometres from the central Optical Line Terminal (OLT). At CERN, therefore, a single OLT in the data centre could feed Technical Network data streams to ONTs in many buildings using the existing fibre optic infrastructure. Of course, if the FTTO infrastructure is required only for a small number of technical connections then the cost per connection is still likely to be prohibitive. It is likely, however, that a single OLT infrastructure could serve a variety of purposes and thus meet the requirement for isolated technical network connections at a reasonable cost.

Initial investigations during 2018 strongly suggest that FTTO is a mature technology, even to the point that FTTO should be considered as the baseline option for providing office connectivity in new buildings, replacing copper-based UTP cabling. A small OLT has recently been purchased, together with various different ONTs, for tests during 2019. Should the results of these tests be positive, it may be that use of FTTO technology will not only enable us to meet the requirement for technical connections but also that we will be able to deliver an improved campus network at a lower cost than with solely traditional technologies.

Summary

A review of future networking needs has shown that the architecture remains valid. All weaknesses of the current campus network can be met simply by upgrading the aging routers with modern devices that are more performant and support newer networking protocols, notably VRF needed to implement multiple virtual campus network domains. Following extensive testing, we have identified a router, the Juniper EX4650, that meets all our requirements. We also have a potential alternative, the Ruckus ICX 7850, that might meet our requirements at a reduced cost.
At the time the review started, in early 2018, we imagined that we would follow past practice and gain experience with new routers in the campus network before deploying them in the Technical Network. This implied deployment of new routers in the campus network from early in 2019 in order to be able to complete the Technical Network upgrade during LS2. As it is not cost effective to deploy the chosen TN router, the Juniper EX9208, widely in the campus network we have decided to delay any upgrade of the campus network routers until after the TN upgrade is complete. Starting the TN upgrade as early as possible in 2019 will also greatly ease the scheduling of this upgrade during LS2. The drawback is that we will not be able to test new features or firmware upgrades in a production environment on the campus network routers before deployment on the more critical TN. To ensure that the ability to test new features in a production environment is maintained, EX9208 routers will be deployed to support the Campus network for Buildings 31 and 28.

Although the delay to the upgrade of the campus network routers will delay deployment of new features, notably the multiple virtual domains, we do not believe that any users will suffer significantly as a result. The delay will also be beneficial as it will allow us to more thoroughly investigate FTTO technology to understand whether this will help us to deliver the services our users expect but with significantly reduced need for active equipment and so at a reduced cost.
### Terminology and acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABR</td>
<td>Area Border Router</td>
</tr>
<tr>
<td>ACL</td>
<td>Access Control List</td>
</tr>
<tr>
<td>DNS</td>
<td>Domain Name System</td>
</tr>
<tr>
<td>EOS</td>
<td>End of Support</td>
</tr>
<tr>
<td>EVPN</td>
<td>Ethernet Virtual Private Network</td>
</tr>
<tr>
<td>FTTO</td>
<td>Fibre to the office</td>
</tr>
<tr>
<td>GbE</td>
<td>Gigabit Ethernet</td>
</tr>
<tr>
<td>GPN</td>
<td>General Purpose Network</td>
</tr>
<tr>
<td>IEEE</td>
<td>Institute of Electrical and Electronics Engineers</td>
</tr>
<tr>
<td>IGP</td>
<td>Interior Gateway Protocol</td>
</tr>
<tr>
<td>IoT</td>
<td>Internet of Things</td>
</tr>
<tr>
<td>L3VPN</td>
<td>Layer 3 Virtual Private Network</td>
</tr>
<tr>
<td>MPLS</td>
<td>Multiprotocol Label Switching</td>
</tr>
<tr>
<td>OSPF</td>
<td>Open Shortest Path First</td>
</tr>
<tr>
<td>PoE</td>
<td>Power over Ethernet</td>
</tr>
<tr>
<td>PIM</td>
<td>Protocol Independent Multicast</td>
</tr>
<tr>
<td>TCAM</td>
<td>Ternary Content Addressable Memory</td>
</tr>
<tr>
<td>TN</td>
<td>Technical Network</td>
</tr>
<tr>
<td>VLAN</td>
<td>Virtual Local Area Network</td>
</tr>
<tr>
<td>VRF</td>
<td>Virtual Routing and Forwarding</td>
</tr>
<tr>
<td>VXLAN</td>
<td>Virtual eXtensible Local Area Network</td>
</tr>
</tbody>
</table>