Porting Commercial WhiteBox NOS to new Chipset for DC Deployment

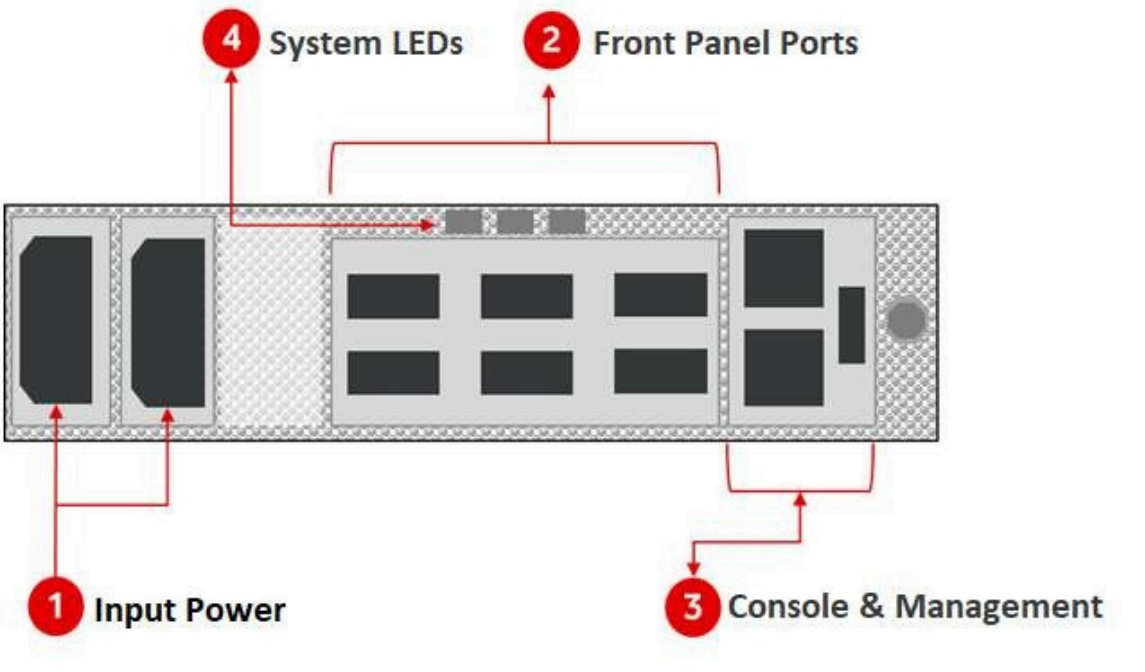
Case-study

1. Introduction

This document describes the project undertaken by PalC Networks for one of their esteemed clients to **port their commercial Whitebox Open Compute NOS (Network Operating System) to a new chipset for Data Center deployment**, which has the port density as xxx, and provide overall throughput as xxx.

1.1. Hardware Architecture

The below diagram shows the high-level mechanical design of the Open Compute switch. This includes input power with redundant power modules, front panel ports, console & management ports (RJ45), System LEDs and a USB interface as described in the figure 2. It also includes the status indicators including per port LEDs and housekeeping interfaces.



2. NOS Platform Architecture

The below diagram represents the high-level overview of the NOS architecture.

3. Network Architecture

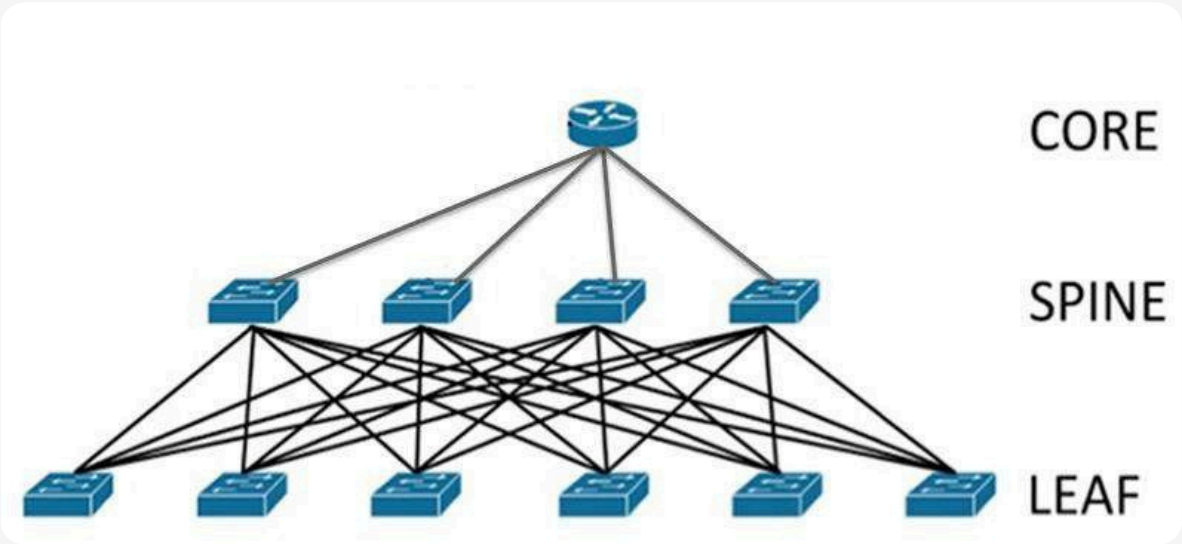
This section talks about the deployment models and network architectures where the NOS and OCP switch will be used. This includes Data Center Leaf-Spine architecture with the DUT acting as both leaf and spine switch.

3.1. Leaf-Spine Deployment

Leaf-spine is a two-layer_network topology composed of leaf switches and spine switches. Leaf-spine is a two-layer data center network topology that's useful for data centers that experience more east-west network traffic than north-south traffic. The topology is composed of leaf switches (to which servers and storage connect) and spine switches (to which leaf switches connect). Leaf switches mesh into the spine, forming the access layer that delivers network connection points for servers.

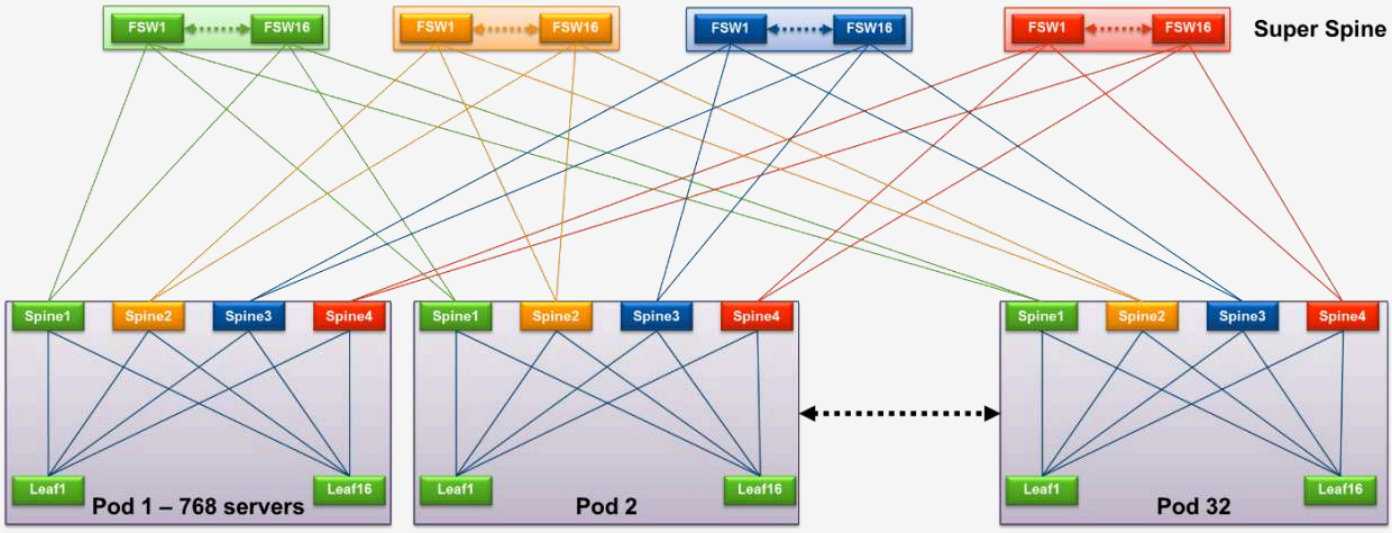
Every leaf switch in a leaf-spine architecture connects to every switch in the network fabric. No matter which leaf switch a server is connected to, it has to cross the same number of devices every time it connects to another server. (The only exception is when the other server is on the same leaf.) This minimizes latency and bottlenecks because each payload only has to travel to a spine switch and another leaf switch to reach its endpoint. Spine switches have high port density and form the core of the architecture.

A leaf-spine topology can be layer 2 or layer 3 depending upon whether the links between the leaf and spine layer will be switched or routed. In a layer 2 leaf-spine design, Transparent Interconnection of Lots of Links or shortest path bridging takes the place of spanning-tree. All hosts are linked to the fabric and offer a loop-free route to their Ethernet MAC address through a shortest-path-first computation. In a layer 3 design, each link is routed. This approach is most efficient when virtual local area networks are sequestered to individual leaf switches or when a network overlay, like VXLAN, is working.



The leaf-spine networks scale very simply, just adding switches incrementally as growth is needed, but they do have some natural sweet spots. For example, since a 32-port spine switch can connect to 32 leaf switches, a natural pod size might be 32 racks of servers with 2 tiers of switching, serving around 1500 10/25GbE servers

If you need a larger network, you will deploy these leaf/spine switches in “Pods” that would represent a logical building block that you could easily replicate as needed. In each Pod, you would reserve half of the spine ports for connecting to a super-spine which would allow for non-blocking connectivity between Pods. A best practice in leaf/spine topologies with a lot of east/west traffic is to keep everything non-blocking above the leaf switches



The Leaf-spine architecture provides

- Better fabric performance with ECMP tuning
- Non-disruptive failover
- Hitless upgrades
- Automated provisioning
- Stretching VLANS & multi-tenant security using VxLAN
- Simple IP Mobility

4. Supported Features

Below is the list of features, which we had supported for the Leaf-spine deployment.

5. Approach

We understood the customer requirement and discussed with the customer about the list of features to be ported for the use case. Once the use case was understood we

- The NOS was running on a Debian7 based OS. Since the board CPU doesn't have support for linux kernel version 3.x. We upgraded the NOS to run on Debian 9 with linux kernel 4.x kernel
- Brought the board up with the new NOS which is packaged as a ONIE installer
- Bring all the system features which is running on the management interface
- Write the hardware hook code to interact with the chipset from control plane

1. Initialize the board, different tables, interfaces etc.
2. Wrote the code to manage all the peripheral devices in the board like fan, led, power supply, temperature sensor etc.
3. Bring the Interface up with different speed, mtu and breakout
4. Port L2 features
5. Port L3 features
6. Port ACL & QoS

- Perform test

1. Test the features individually
2. Test the use case

Write config guide

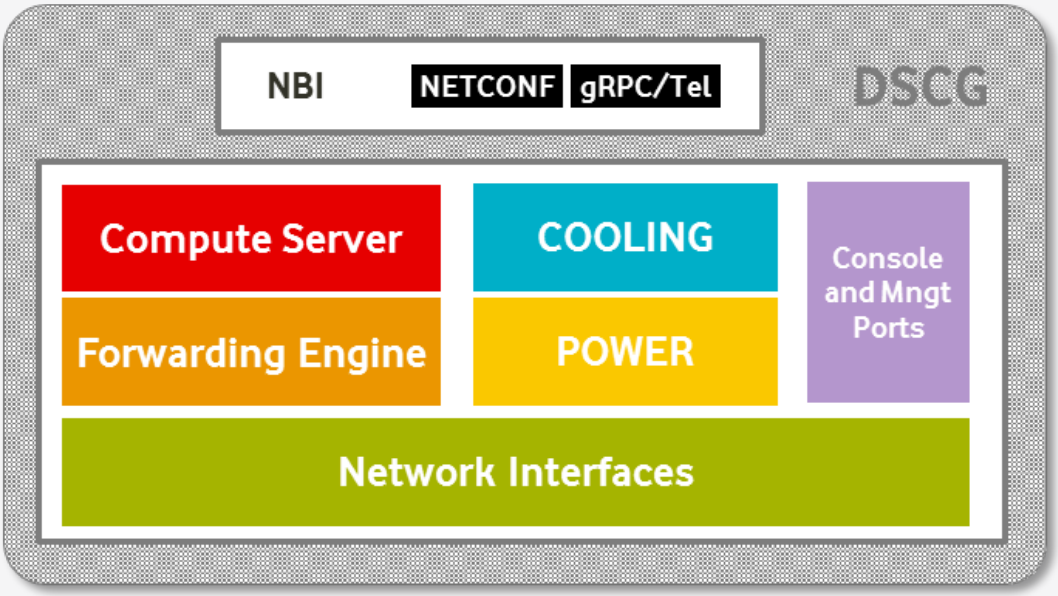


Figure 1. DCSG platform high level architecture components

5.1. Form Factor, Environmental and Power Supply Requirements

The DCSG will be used in standard cell site locations with DC/AC power supplies, 1U form factor with 300 x 600 mm for installation in standard 19” rack.
The equipment shall conform to or exceed ETSI standard ETS 300 019-1-3 Class 3.4 (-40°C to +70°C):

- -40 to 70°C, up to 1,000 feet (300m)
- -40 to 65°C, up to 6,000 feet (1800m)
- -40 to 55°C, up to 13,000 feet (4000m)

5.2. Network Interfaces and Forwarding Capacity

In terms of forwarding capacity, the DCSG shall be able to cope with current demands (for 3G/4G networks) and also with future and higher capacity demands for 5G deployments. The interfaces shall be Ethernet based and the platform shall support between 6 and 8 of them, working at different speeds from FE to 10GE initially and an additional variant shall be produced with same number of ports of which two of them at least shall be capable to work at 25GE & 50GE. All the interfaces shall be able to work either as UNI or NNI.

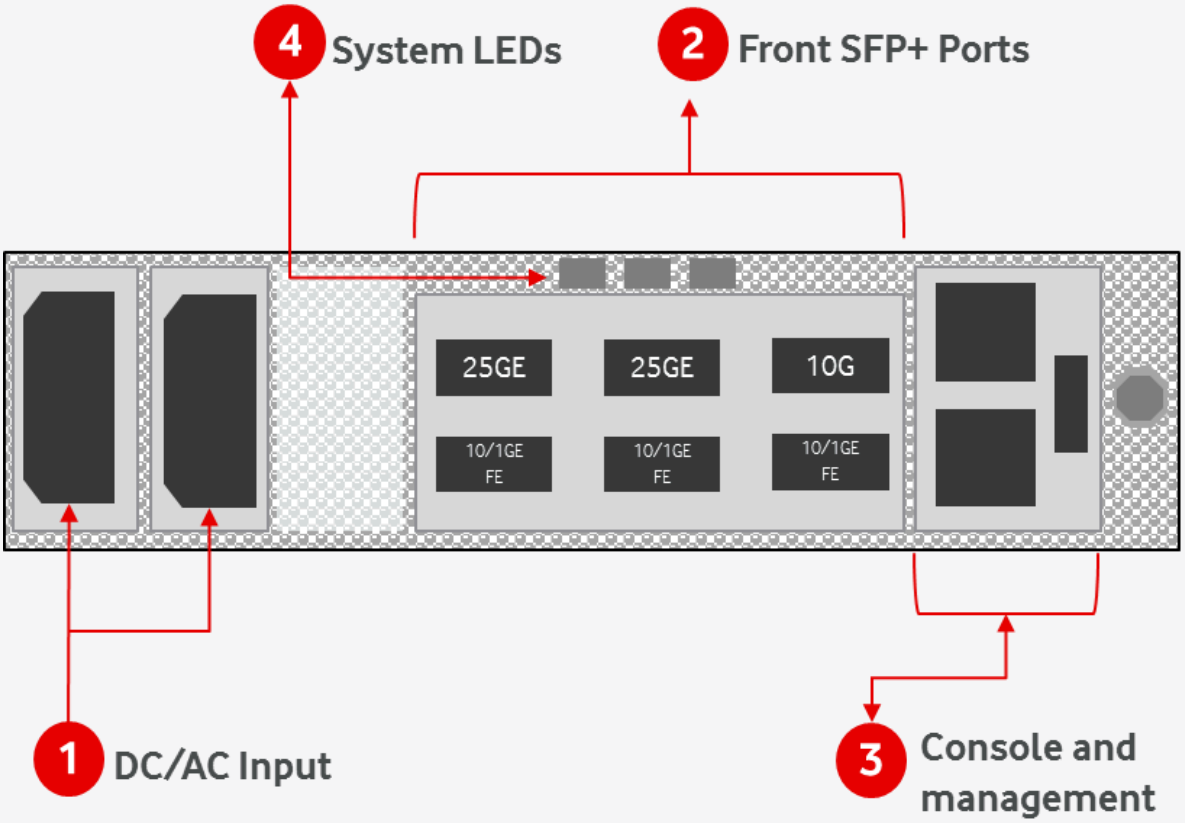


Figure 2. Example of DCSG target front view

The ports shall be able to support SFP/SFP+ modules and configurable by software. The platform shall be compatible with a pay as you grow model defined by software licences.

As mentioned before, a first implementation of the platform could include only 10GE/1GE interfaces. Future versions would be required to have the option of 25GE or faster Ethernet interfaces (at least 2 of them). The platform shall be fully interoperable with any 3rd party modules (SFP/SFP+).

5.3. Management Interfaces and Miscellaneous

The platform shall include console and management ports (RJ45) and a USB interface as described in the figure 2, for local configuration & debugging

The platform shall include status indicators including per port leds and housekeeping interfaces (see excel attached in section 5.1).

5.4. CPU and ASIC

The DCSG CPU shall be based on a X86 architecture, possibly 64bits or 32bits (TBC). The ASIC shall support capacity with full port / line rate support. The forwarding capacity of the platform shall be able to support all interfaces at full rate with no limitations. ASIC shall support all features and packet types defined in software section.

5.5. Time Sync Distribution

The DCSG shall be able to provide time sync to the nodeB which are connected to it. Additionally, as it will be explained in section 3, the DCSG will be potentially deployed in many different scenarios so it shall also be able receive the time sync signal and propagate it to other network elements close to it in the network (e.g. other DCSGs in a ring).

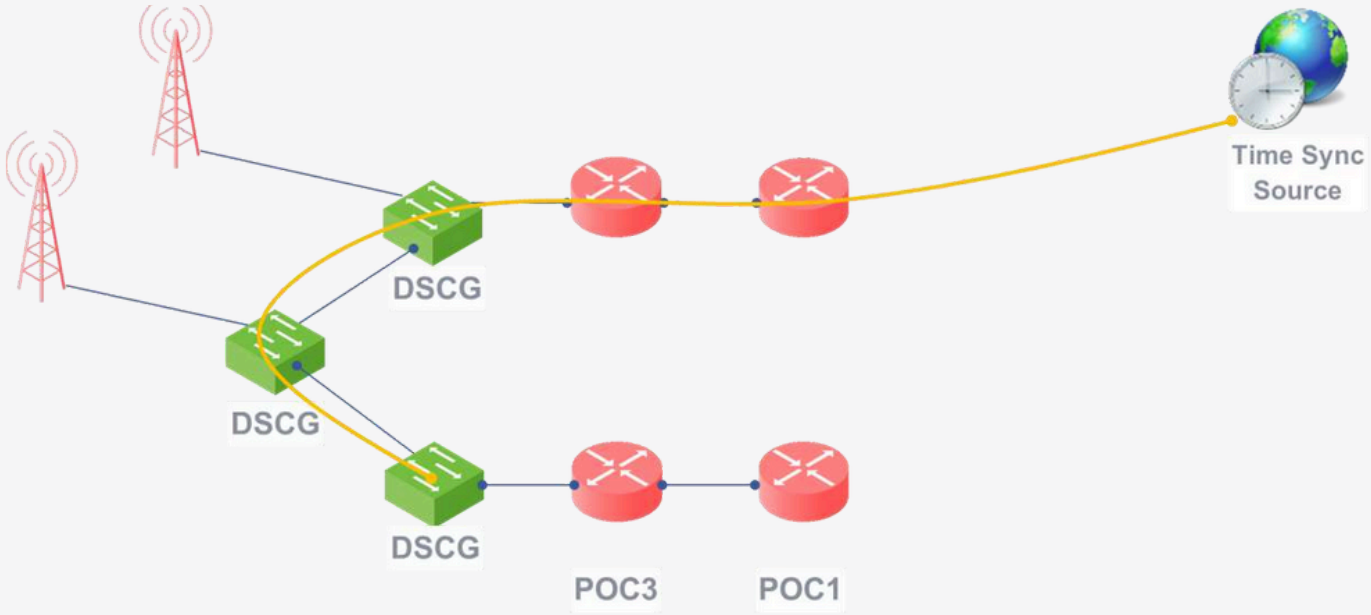


Figure 3. Time Sync distribution example with DCSG

The platform shall support the following time sync requirements:

- Network quality model (microsecond precision) according to ITU-T G.8271.1 Network conditions and reference model where the Boundary Clock is requested to work.
- Node performance (working in ideal condition) according to ITU-T G.8273.2 (7.1) Boundary clock quality objectives in holdover mode.
- Node performance (upon wander, failure, holdover) according to ITU-T G.8273.2 (7.2/7.3/annex) Boundary clock quality objectives in holdover mode
- Interoperability according ITU-T G.8275.1. Boundary clock ITU-T standard protocol features, including SyncE support for holdover purposes and Grandmaster redundant sources support

The DCSG shall also support a SFP+ input for GPS. The platform shall be interoperable with most of the SFP+ providers' solutions in the market in order to reduce interoperability issues. This GPS input shall be used only in scenarios where the time sync signal cannot be received from the backhaul network through a standard Ethernet traffic port in case time sync distribution mechanisms are used.

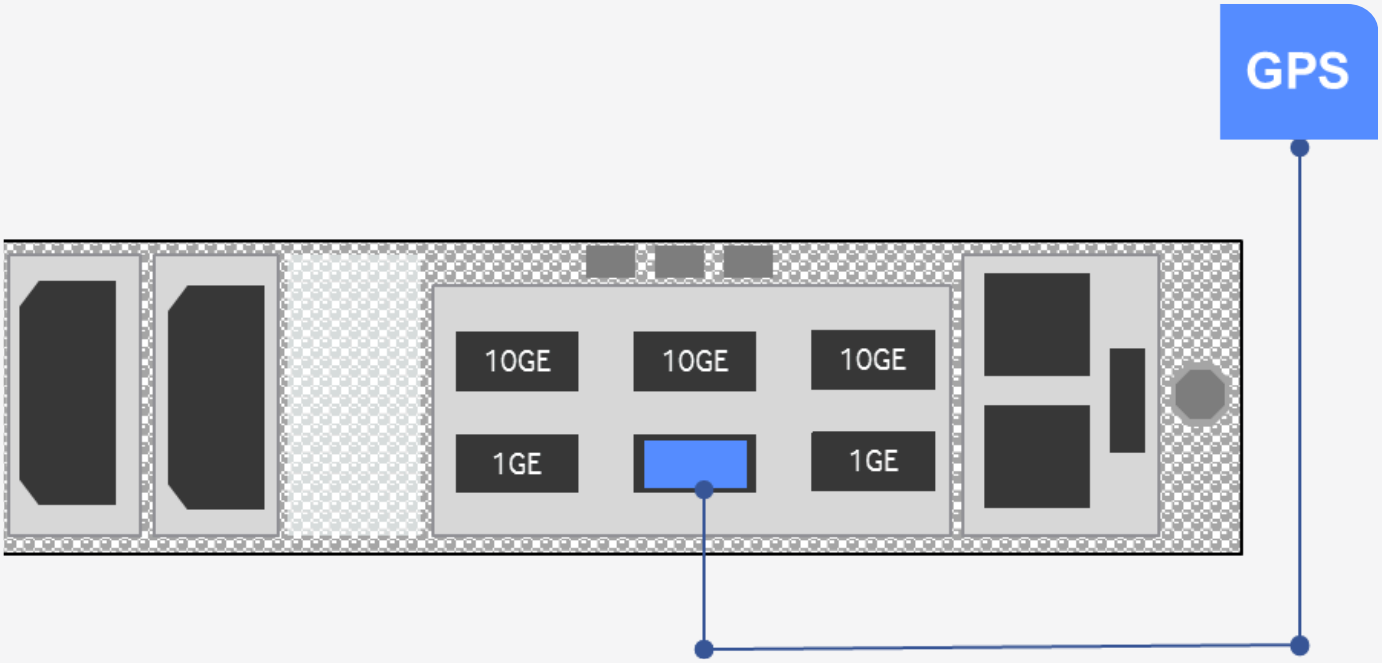


Figure 4. GPS SFP for Time Sync equipped in a traffic port

5.6. Mac Sec

Current security/encryption architectures for transport networks are evolving and the DCSG shall support the most advanced security/encryption capabilities like MACSEC.

The platform shall support the following MACSEC requirements:

- MACSEC Support
- MACSEC Security Mode: Static Connectivity Association Key Security Mode
- MACSEC Security Mode: Dynamic Secure Association Key Security Mode
- MACSEC hop by hop mode
- MACSEC multi-hop in tunnel mode

The Platform shall be at least HW ready in order to implement the abovementioned capabilities. The activation/support of MACSEC shall be available only with a SW upgrade/license activation (w/o HW changes).

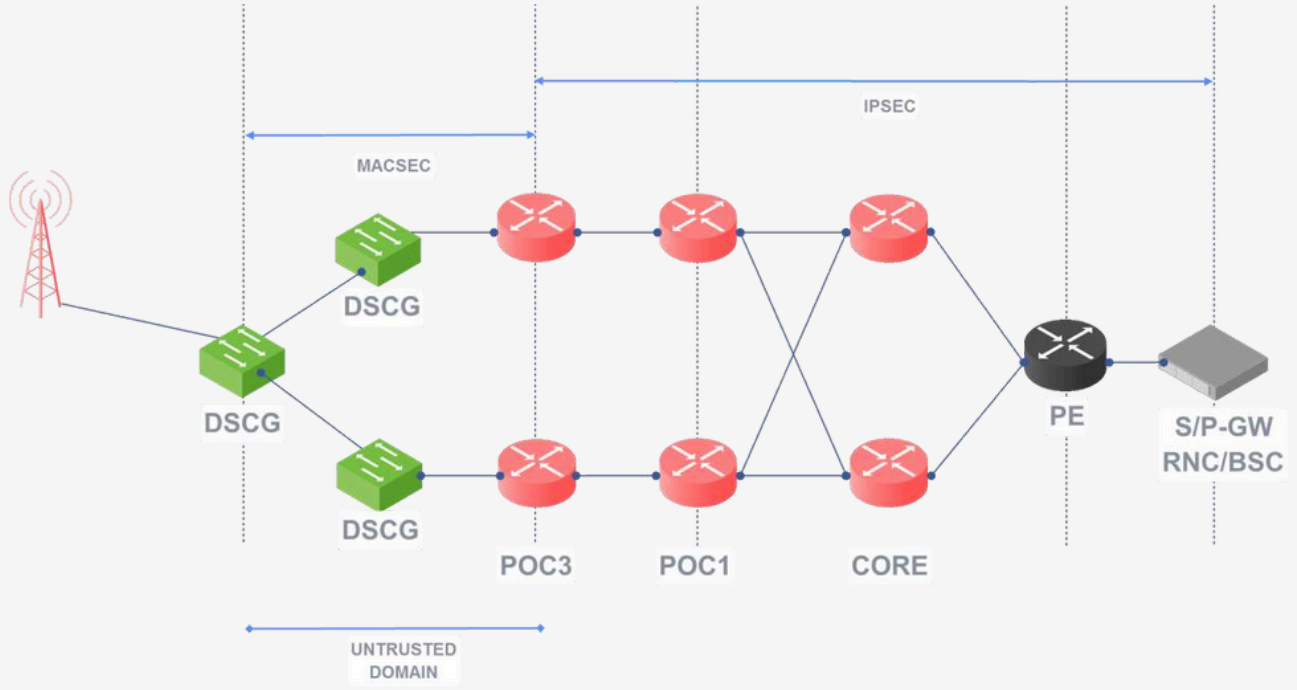


Figure 5. MACSEC application scenario

Additionally, the DCSG shall support EAP – TLS IEEE 802.1x. EAP-TLS will be used in scenarios where MACSEC is not implemented.
Finally, the DCSG shall support MD5 for authentication with other NEs/DCSGs.

5.7. Routing/Switching capabilities (Layer 2 & Layer 3)

The platform shall support (as it will be explained in sections 3) features in order to be able to work both at Layer 2 and layer 3. In order to simplify the architecture and the feature set that the DCSG will use, the L3 capabilities (e.g. OSPF) will be used only in the required scenarios.

In terms of L3 features support, the DCSG shall, at least, support:

- OSPFv2 (RFC 2328)
- OSPFv3 for scenarios where IPv6 is used (RFC 5340)
- eBGP for CE to PE communication
- OSPFv2/v3 for CE to PE communication

The DCSG shall also support dual Stack IPv6/IPv4 (in case the NodeBs are using IPv6 addressing but the transmission network is not).

5.8. L2 Requirements (MEF)

In case the DCSG is used for providing L2 services to enterprise customers (out of mobile access/backhauling context) the DCSG shall be MEF 2.0/3.0 Compliant.
The DCSG shall support:

- G.8032 Ethernet ring protection
- QinQ IEEE 802.1ad
- Ethernet services must allow customers to use full range for CE-VLANs (4095)
- Y.1731 (OAM functions and mechanisms for Ethernet-based networks) and Y.1564 (Ethernet service activation test methodology) ITU-UT standard compliancy.
- H-QoS shaping
- Set P-bits on the network port depending on the Color of the service
- ETH CFM according to Y.1731/G.8013. Ethernet OAM IEE 802.1ag. MEF SOAM.
- ETH OAM MEG LEVEL 3 and Level 4

The DCSG shall allow (acting as a NID for L2 services) to do reporting of: Latency, Frame Loss Ratio, Frame Delay variation/Jitter, Availability.
There is also an ask on the DCSG (NID) to be able to perform bearer and service testing RFC2544 & Y.1564 respectively in order to provide real service certification results for the customer, but this will be further detailed in the NID section.

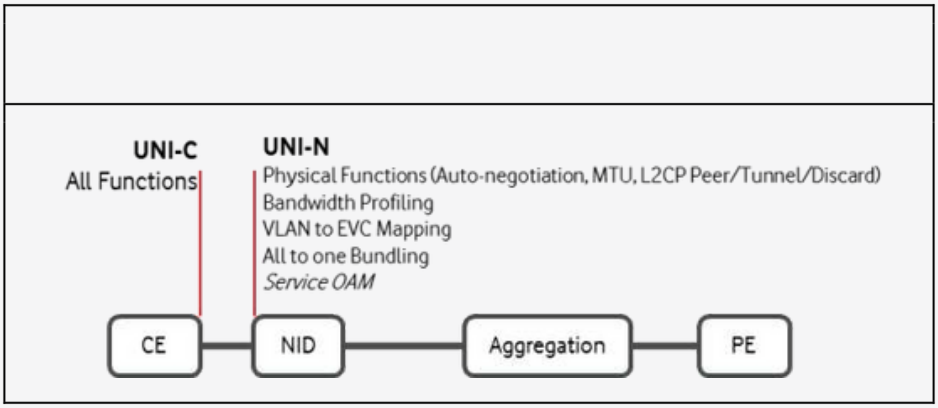


Figure 6. MEF UNI functional delivery model with DCSG working as Enterprise NID

The Metro Ethernet forum specifies multiple different types of UNI described in MEF13 and MEF20. Vodafone will initially support UNI2.1 including the 1522-byte MTU without any further optional attributes. This capability will be backward compatible with UNI type 1.0 and 1.1.

5.9. Additional Features

The DCSG shall support:

- BFD is to be evaluated, based on the resources needed and the HW cost to implement it.
- Multi-card LAG with Bidirectional Forwarding Detection (BFD) on Link Aggregation Group (LAG) Interfaces according to RFC7130

5.10. Quality of Service

The DCSG shall be able to map the mobile traffic classes defined by the GSMA into the right DSCP values, using standard compliant values for DSCP with compatible mapping with IP Precedence field if any old non DiffServ capable routers are present in the network.
As a reference, the following picture shows a potential mapping of mobile traffic into the DSCP classes.

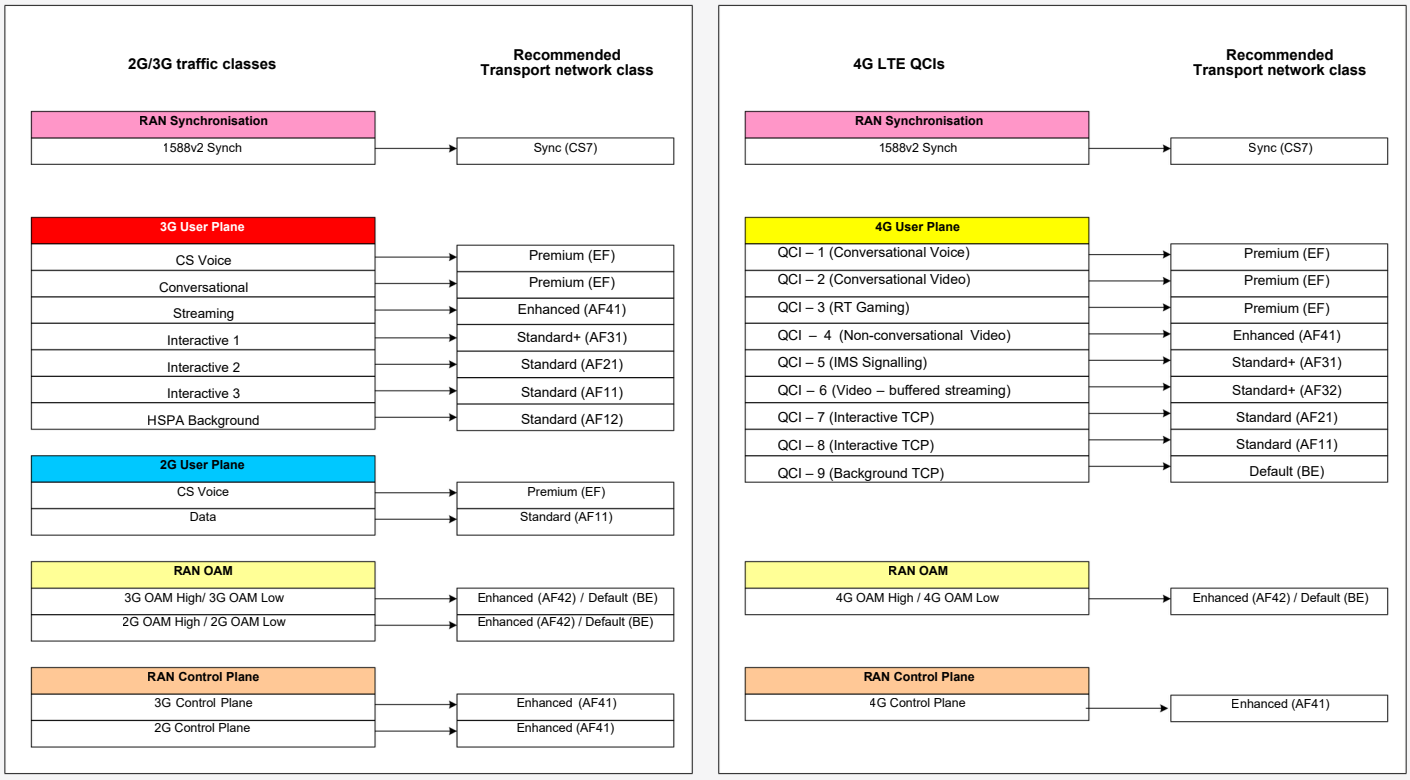


Figure 7. QoS mapping into DSCP - example

The DCSG shall also be able to map the client traffic classes into the VLAN (IEEE802.1q) tagging values for the scenarios where DCSG is acting at layer 2 (Enterprise scenario). An example (for enterprise traffic) is shown in the picture below:

Class Name	802.p
Premium	5
Enhanced	3
Standard	1
Default	0

Figure 8. Traffic classes

5.11. Performance monitoring and telemetry

The Platform shall support the following performance monitoring requirements:

- Y.1564 Tech Specifications
- Y.1564 IOT with Third Party
- ITU-T Y.1564 IOT with JDSU
- Y.1564 Layer 2
- Y.1564 Layer 3
- Y.1564 CoS Tests
- Y.1564 Time Stamping
- Y.1564 Loopback tests
- Y.1564 Full line rate

5.12. Software Scalability figures

As a reference, the following scalability figures shall be supported by the platform SW:

IP prefixes VRFs VSIs	20K
EVIs	128
EVCs	4K
Y.1731 S-OAMflows	2K
ScalingE-LANhub	256
	1024
	256 sites
Scaling EVPL hub	256 sites

Figure 9. SW platform scalability figures

6. Network Architecture

We introduce here the DCSG deployment models and network architectures where it will be used. These include IP/MPLS, Microwave and Optical network scenarios.

6.1. DCSG in IP/MPLS networks

During the last 10 years, network operators have been using different technologies to aggregate mobile traffic, one of the most implemented (probably the most common one) is IP/MPLS.

Implementing full IP/MPLS architectures including the cell sites has been a complex task for network operators, mainly because of the lack of support of some specific features in the smaller IP/MPLS boxes. The objective in this project is to keep the architecture as simple as possible, reducing the complexity of the DCSG and the aggregation network while leveraging on advanced control capabilities provided by SDN. In this case the DCSG shall work as a simple CE router, running an OSPF instance together with the other DCSGs that might be part of the same ring and the POC3s (the interfaces that belong to the ring will be included in a IPVPN) where this ring is closed. The DCSG will be treated in this case as a customer of the IP/MPLS network. The traffic received from the DCSG will be transported by the IP/MPLS network (as an example) in a dedicated H&S L3VPN till the Core network, as described in the picture below.

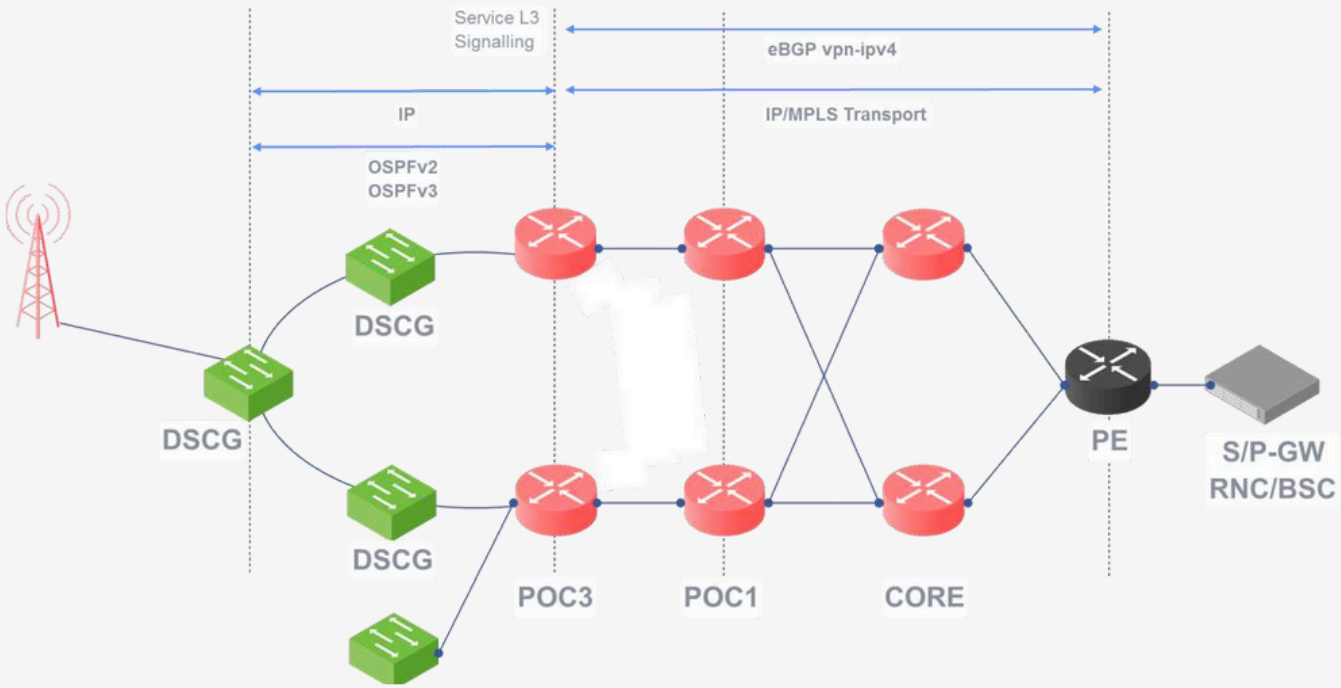


Figure 10. Scenario 1: Dual-home IP/MPLS Ring

The picture describes a network where IP/MPLS is used for performing the E2E transport of mobile traffic (including 2G, 3G, 4G and 5G) from the Backhaul to the Core network.

OSPF will also be used as a protection mechanism combined with Bidirectional Forwarding Detection (BFD) for faster detection in order to achieve better protection times. BFD intervals shall be adjusted to reduce as much as possible the detection time but also keeping in mind that BFD is quite intensive and will consume a IOT resources of the system. The BFD interval shall be adjusted based on the network scale/size. The DCSG shall support also OSPFv3 (RFC 5340) to beIPV6 capable. The use of OSPFv3 shall be decided based on the operator network addressing strategy.

6.2. DCSG as Microwave IDU

There are some scenarios in which the DCSG could be considered to be used in microwave networks acting as IDU (Indoor Unit). Those cases are limited to the ones where the microwave modem is implemented as part of the ODU (Outdoor Unit)1, so the IDU needs to implement only basic Ethernet functions (VLAN switching, QinQ, etc.) and support optimal inter-operability with microwave equipment and link peculiarities.

DCSG is connected to the ODU using standard GE and 10GE Ethernet interface. In this case the DCSG receives different VLANs from the ODU and implements QinQ to transport those VLANs towards the POC3. The Backhaul network will be in charge of transporting those VLANs up to the Core.

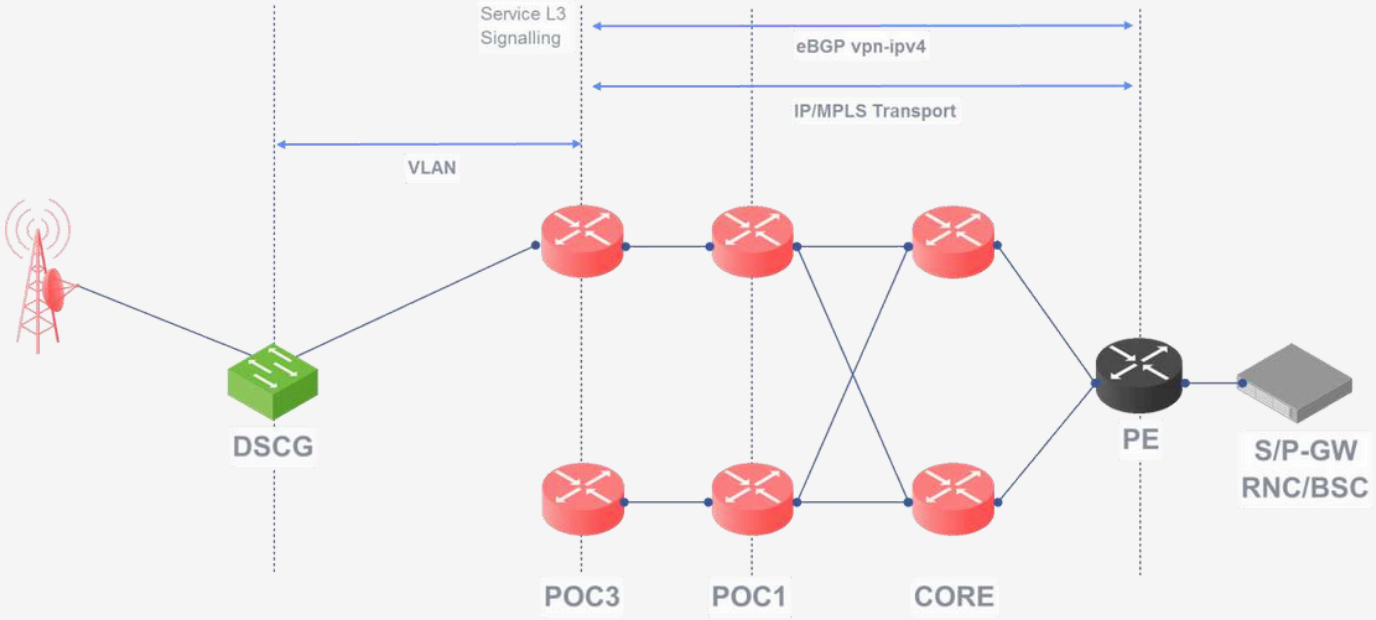


Figure 11. DCSG working as MW IDU

DCSG shall support ITU-T Y.1731 Bandwidth notification in order to be informed real time on the actual microwave link capacity to take appropriate actions (e.g. real time traffic shaping); this will ensure DCSG is fully aware of microwave link peculiarity (variable capacity due to Adaptive Code & Modulation, ACM) DCSG will also provide (using Routing/Switching capabilities (Layer 2 & Layer 3) – see section 2.7)
management connectivity for microwave ODU:

- towards management platforms (EMS/NMS/SDN controller)
- for local management on Console & Management port of DCSG (see section 2.1.3).

The case of Split Mount microwave (where modem is split from the ODU radio) is not considered for integration in the DCSG. This is because modem is proprietary technology (different across manufacturer and across products of same manufacturer) with proprietary interface towards the ODU.

6.3. DCSG in Packet Optical Networks

An alternative backhaul solution could be in place in some cases. This solution, based on packet optical (L2 switching/MPLS-TP) technologies, can also be used in order to aggregate mobile/enterprise traffic from the DCSGs up to the Core.

The different deployment scenarios will be detailed in the following sub-sections.

Alternatively, pure Ethernet aggregation into OTN can be used, but only tactically, as it is in general quite intensive in terms of bandwidth consumption on the optical network and not optimal - traffic aggregation prior to the optical network is recommended to optimise optical bandwidth usage.

6.3.1. SCENARIO 1: Dual-home/Single Home Ring

In case a ring of DCSGs is closed in 2 or 1 packet optical network elements (dual-home ring), G.8032 can be used as protection mechanism. Different transport technologies/solutions can be implemented from the POC3 up to the core (e.g. MPLS-TP, G.8032). More specific designs are out of the scope of this document. As an additional option in cases of single DCSG with dual homing to POC3 multi-chassis LAG could be considered instead of G.8032.

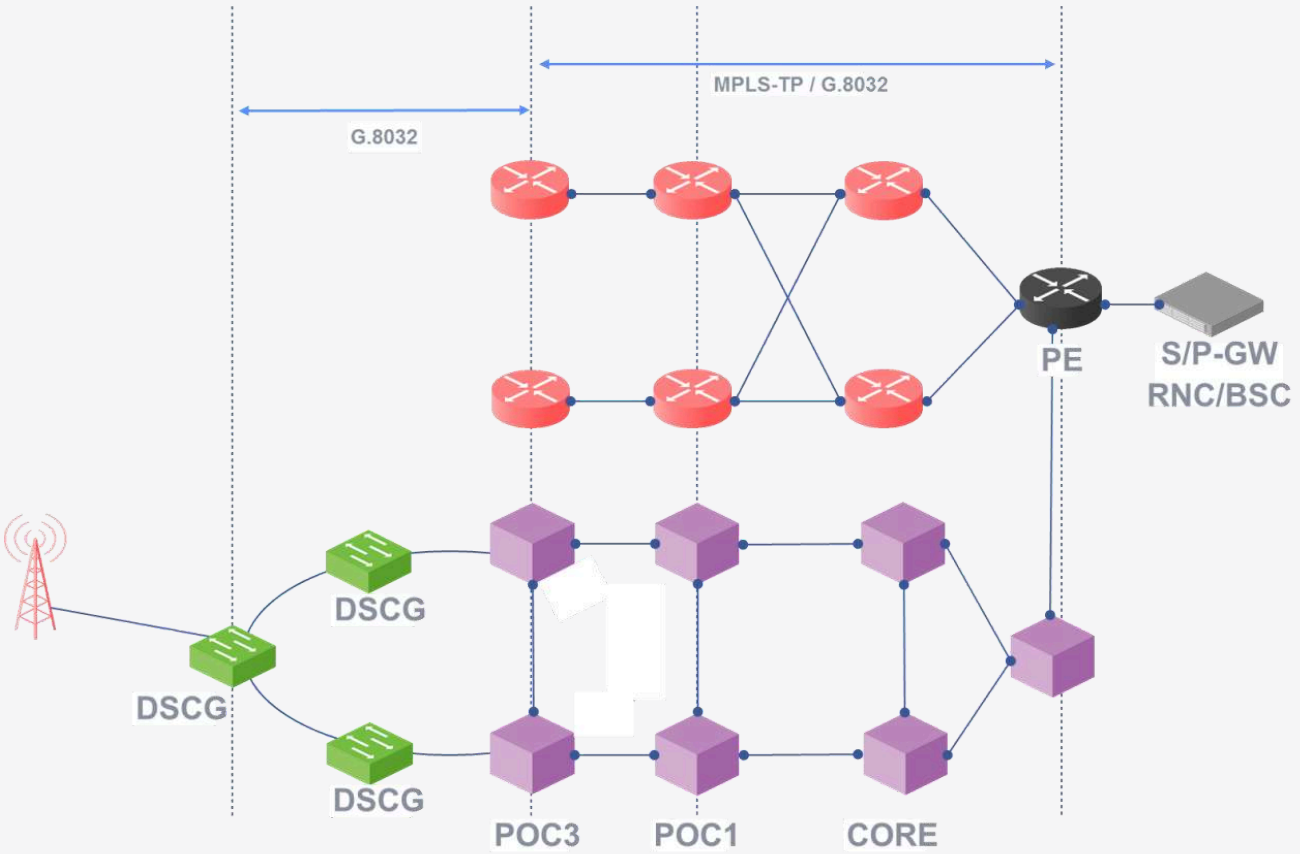


Figure 12. Scenario 1: Dual-home ring with packet optical backhaul

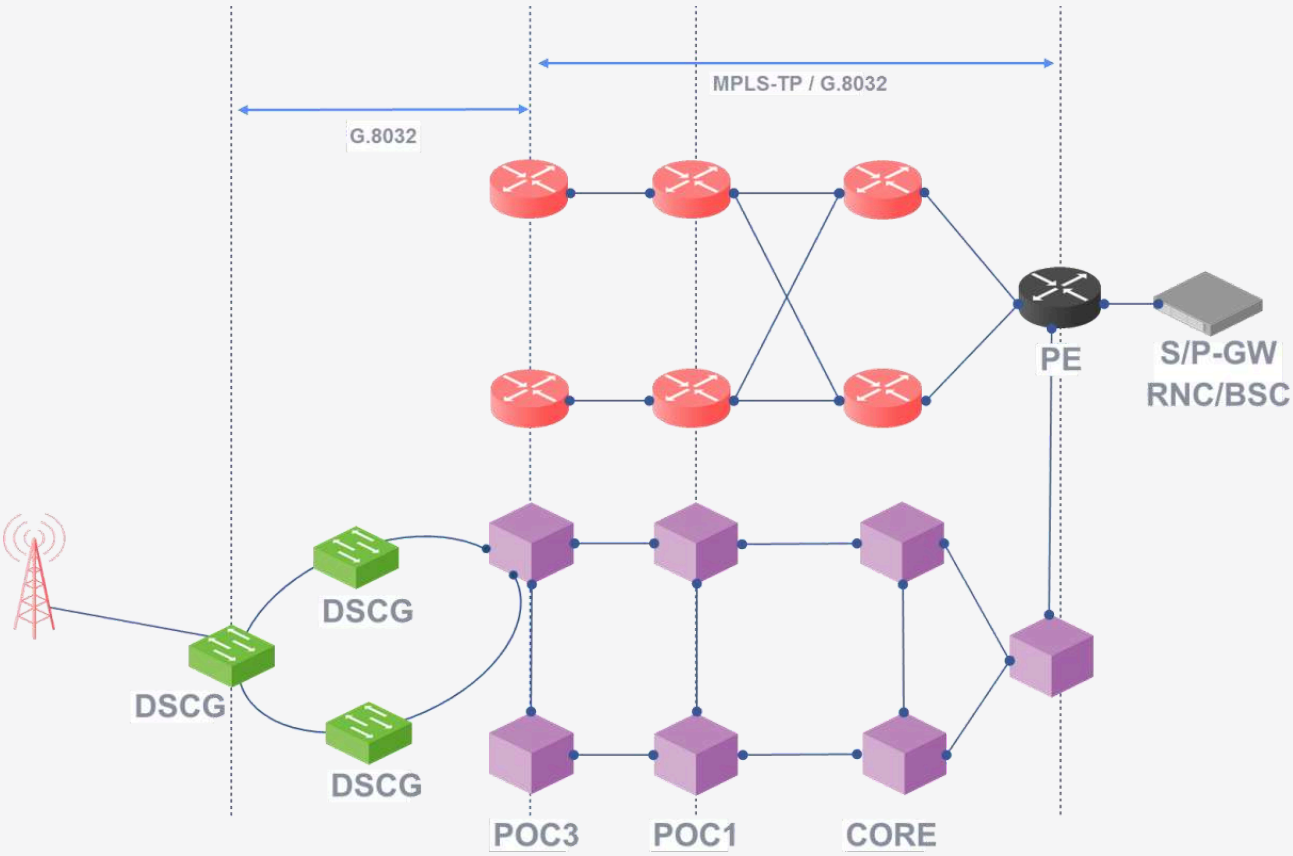


Figure 13. Scenario 2: Single-home ring with packet optical backhaul

6.3.2. SCENARIO 2: Directly Connected

In case the DCSG is directly connected to a packet optical network element (see picture below) the DCSG will only perform a simple switching functionality (QinQ). Then different VLANs will be transported up to the core based on the existing transport technology in the packet optical network.

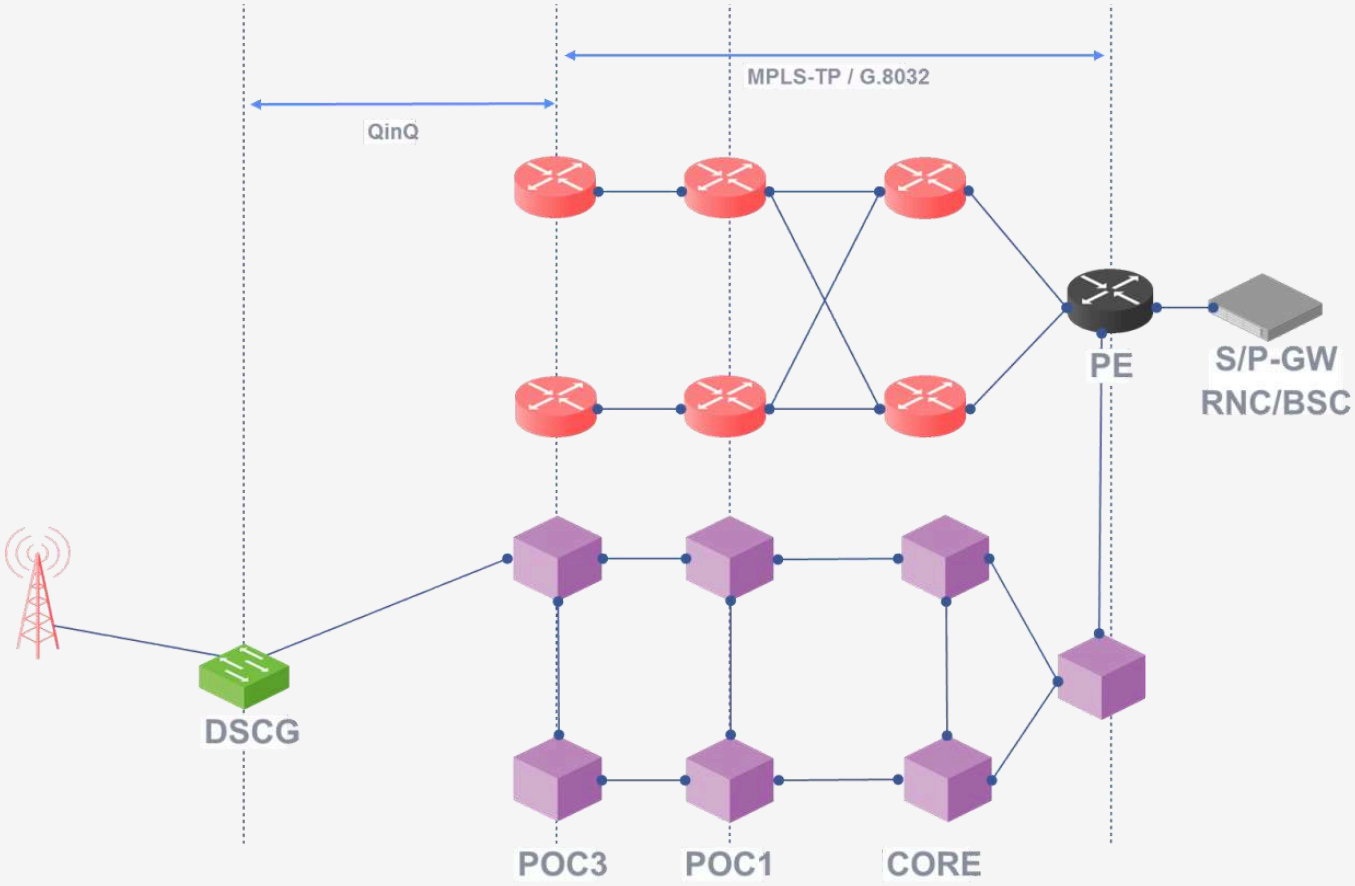


Figure 14. Scenario 2: Point to Point with packet optical backhaul

6.4. Auto Configuration

The DCSG shall be designed to support auto-configuration mechanisms. The intention is to simplify and automate as much as possible the deployment process. Any additional configuration will be performed by the SDN controller.

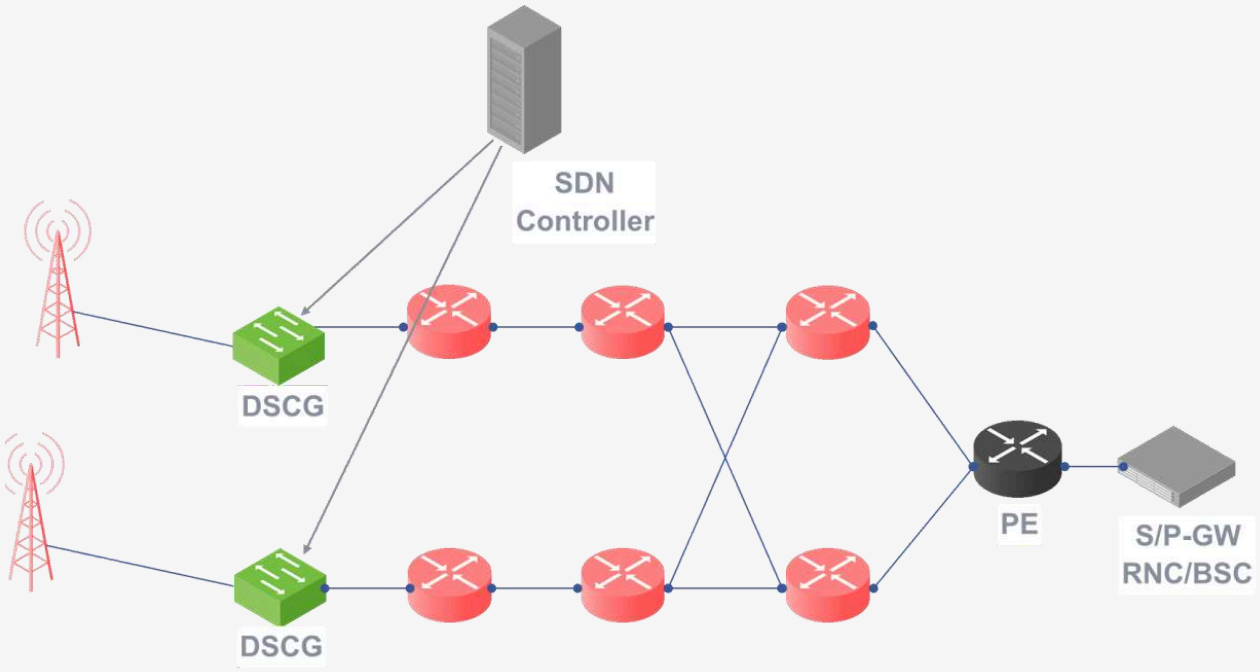


Figure 15. Auto Configuration

The different in band management models/architectures, depending on the network architecture/scenario, including auto-configuration support, will be described in the following sub- sections.

6.4.1. DCSG Auto Configuration in IP/MPLS Networks

In case DCSG is deployed in a IP/MPLS network (see section 3.1), the SDN controller needs to have connectivity to those DCSGs. For achieving this connectivity, a management and H&S IP VPN will be created in the POC3 in the network where there are DCSG connected. This L3 VPN will allow the remote (in band) management of the DCSGs from the location where the SDN controller has been deployed (typically a POC1 as per the diagrams shown before).

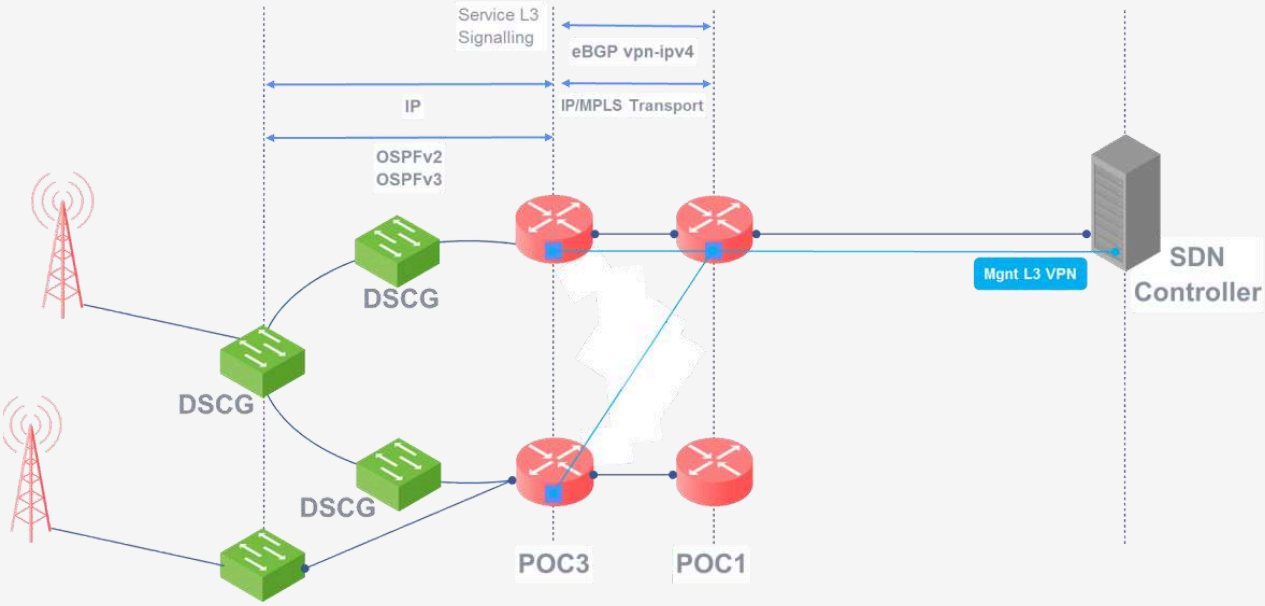


Figure 16. Auto Configuration in IP/MPLS Networks

In this case, the minimum configuration that the DCSG shall contain includes:

- IP Addresses of the line interfaces (connected to the ring) according to the addressing design of that network
- OSPF enabled in the abovementioned interfaces.
- Security certificates

6.4.2. DCSG Auto Configuration in Microwave Networks

In case DCSG is deployed Microwave IDU a VLAN will be used for management. This VLAN will terminate into a hub-and-spoke VPN within the POC3 allowing communication between the SDN controller and the DCSG.

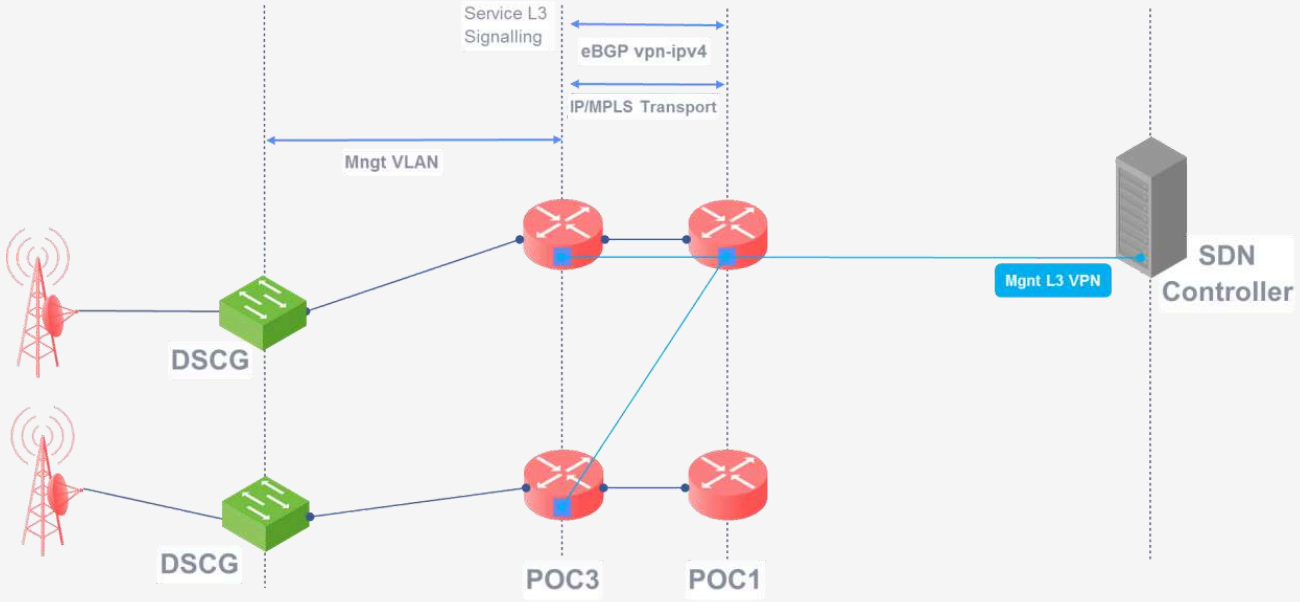


Figure 17. Auto Configuration in MW networks

In this case, the minimum configuration that the DCSG shall contain includes:

- A management VLAN that will be used for the DCSG configuration
- Security certificates

6.4.3. DCSG Auto Configuration in Packet Optical Networks

In the case of packet optical the management VLAN will be transported using an EVP-Tree or EVP-LAN (hub & spoke) to the central node where it is delivered to the SDN Controller.

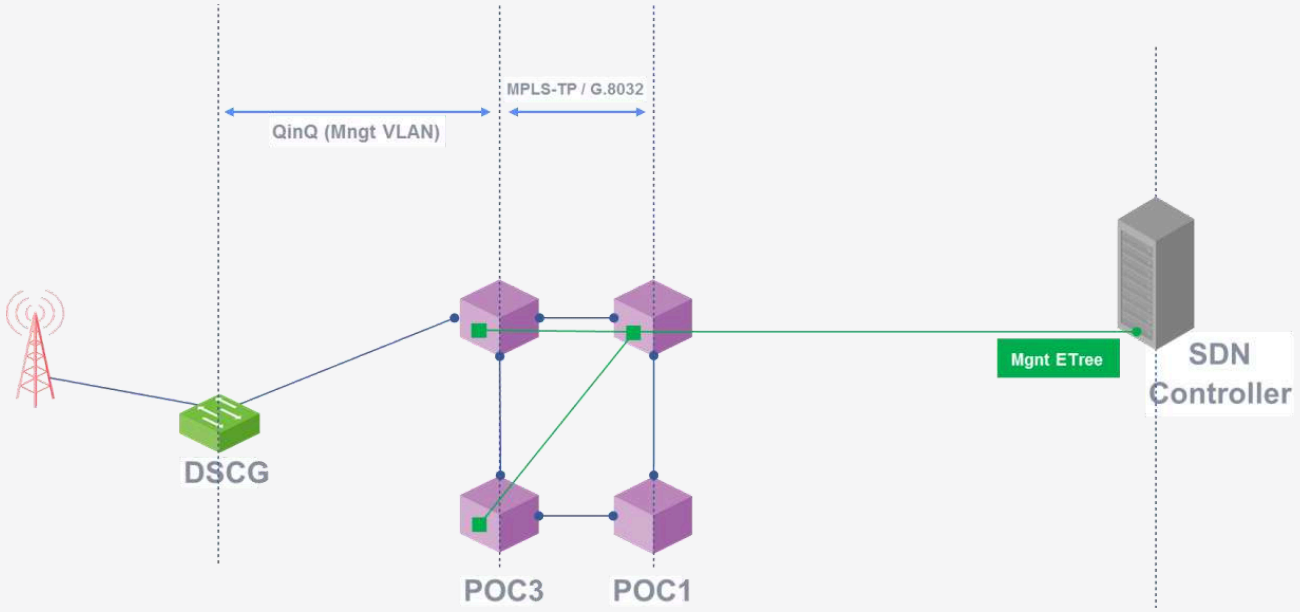


Figure 18. Auto Configuration in Packet Optical/Optical networks

In this case, the minimum configuration that the DCSG shall contain includes:

- A management VLAN that will be used for the DCSG configuration
- Security certificates

7. DCSG and Software Defined Networks

As explained in previous sections, the intention is to have a centralised control entity (an SDN Controller) managing the DCSG in a smart and automatic way.

The SDN controller shall manage and optimize the DCSG domains in order to perform SLA fulfilment and service provisioning.

All the configuration and management of the DCSG shall be done using Netconf (RFC 7803). Additionally, the controller will also use BGP-LS to collect all the OSPF-TE/L2 topology information in the DCSG domains.

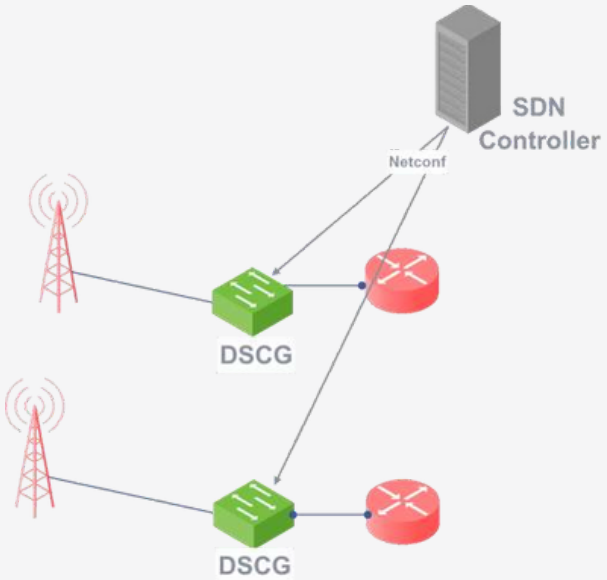


Figure 19. SDN controller and Management

Therefore, the interface needed at both the DCSG and the SDN controller is:

- Netconf (RFC 7803); connections could be encrypted using Transport Layer Security (TLS) [RFC5246] or Secure Shell (SSH) [RFC4251], being TLS the preferred option.

Standard data models shall be used as much as possible. The definition/selection of the target/needed models will be done in future releases of this specification.

The controller shall support also BGP-LS in order to collect the topology information from the DCSG domains, but the POC3 will act as a gateway for BGP-LS.

7.1. DCSG Telemetry and SDN

As explained in previous sections, the DCSG shall support advanced monitoring and telemetry features. Those features will be used by the SDN controller in order to monitor the status of the platform and the different services instantiated in the DCSG.

Therefore, the features needed at both the DCSG and the SDN controller is:

- RPC Network Management Interface (gNMI), gPB (Google Protocol Buffers) proto3 for encoding, and data exported (modelling) based on YANG models.

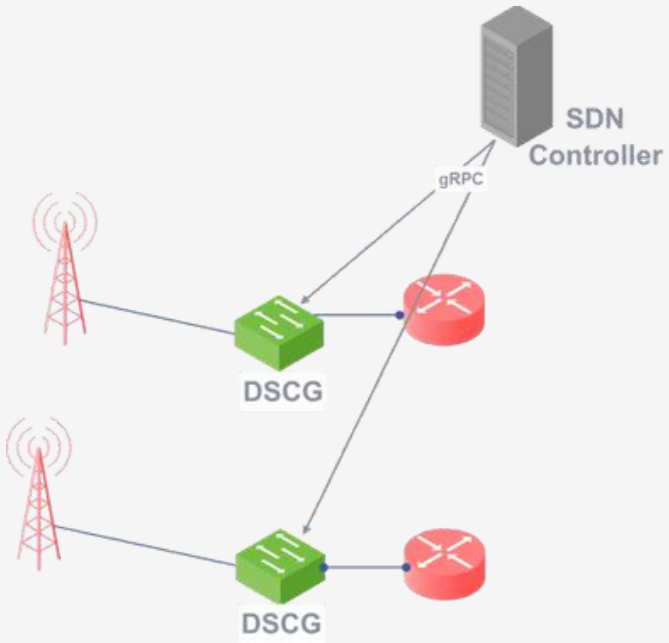


Figure 20. SDN controller and Management

8. GLOSSARY

BFD: Bidirectional Forwarding Detection
CE: Customer Edge
H&S: Hub & Spoke
IDU: Indoor Unit
ODU: Outdoor Unit
PE: Provider Edge
PoC1: Point of Concentration 1st level (aggregation level next to the core network)
PoC2: Point of Concentration 2nd level (intermediate aggregation level)
PoC3: Point of Concentration 3rd level (first aggregation point after last mile/access)
SDN: Software Defined Networks
DCSG: Disaggregated Cell Site Gateways