ORAN-WG9.XPSAAS.0-R003-v07.00

Technical Specification

O-RAN Open Xhaul Transport Working Group 9

Xhaul Packet Switched Architectures and Solutions

O-RAN ALLIANCE e.V., Buschkauler Weg 27, 53347 Alfter, Germany

1 Revision History

Date	Revision	Author	Description
2020/11/11	V01.00	All Authors	1st revision outlining a packet switched O-RAN solution based on an underlay transport infrastructure based on MPLS or IPv6 with Segment Routing (SRv6) with mobile services provided by Multi-protocol BGP based VPNs.
2021/07/01	V02.00	Simon Spraggs, Krzysztof Grzegorz Szarkowicz, Luis Miguel Contreras Murillo, Ivan Bykov, Lujing Cai	Update of references as required. Addition of annex F describing a packet switched slicing solution appropriate to slicing phase 1 described in O-RAN.WG1.Slicing-Architecture-v05.00
2022/02/27	V03.00	Simon Spraggs, Krzysztof Grzegorz Szarkowicz	Updates to annex F to support slicing phase 1 and 2 as described in O-RAN.WG1.Use Cases Detailed Specification-v06.00. Addition of a section on TNE security.
2022/10/28	V04.00	Krzysztof Grzegorz Szarkowicz, Phil Bedard, Simon Spraggs	Updates to Annex F to support slicing phase 3 as described in O-RAN.WG1.Slicing-Architecture-v07.00. Updates of section 17 about TNE security (adding physical security, user security, control plane security and IEEE 802.1X). Updated Section 18 about service slice to underlay transport plane mapping. Updates to reference list. Updated Section 11.5.1 about Seamless MPLS with transport planes. Updated Section 9.4 to clarify which IETF documents can be referenced. Annex ZZZ (O-RAN Adopter License Agreement) removed, as it is no longer needed
2023/02/24	V05.00	Krzysztof Grzegorz Szarkowicz	Updates to Section 17.4 about IEEE 802.1X
2023/06/09	V06.00	Ivan Bykov, Lujing Cai, Krzysztof Grzegorz Szarkowicz	Updates of references as required. Updates to Section 10 regarding CTI references. Updates to Annex F regarding slicing phase 4 in Midhaul.
2023/09/27	V07.00	Krzysztof Grzegorz Szarkowicz	Updates references, as required (added 3GPP references related to certificate management, added IETF reference related to 5G transport slicing). Updates sectiopn 17,4 (IEEE 802.1X authenticatioin towards Computes and O-Clouad-Gateways; additional section about certificate management)

1.1 Contributors

Editor: Krzysztof Grzegorz Szarkowicz
Contributors in alphabetical order: Jennifer Andreoli-Fang, Phil Bedard, Lujing Cai, Francois Fredricx, Ivan Bykov, Kashif Islam, Luis Miguel Contreras Murillo, Toby Rees, Simon Spraggs, Krzysztof Grzegorz Szarkowicz, Reza Vaez-Ghaemi, Nader Zein, Jeffrey Zhang

2 Contents

3 Scope

This Technical Specification has been produced by the O-RAN Alliance. The document is intended to describe best practises for O-RAN transport based on end-to-end packet switching technology. It is recognised that other solutions, not based on packet switching, could be employed or mixed with a packet switching solution. Beyond the solutions described in this document, other packet switching solutions may be adequate for Xhaul transport networks and can be considered in future versions of this document.

This specification defines an architecture for an Open Xhaul transport network based on an end-to-end packet switching architecture that utilises statistical multiplexing and a hierarchy of packet switching "Transport Node Equipment" (TNE) starting at the cell site in the access layer and going to the core layer of the transport network capable of supporting the requirements outlined in the O-RAN WG9 Transport Requirements document [19].

Within the transport core, aggregation and pre-aggregation it is assumed the L0/L1 transport technology connecting the packet switches are high capacity, low delay Ethernet point to point circuits. These circuits can be derived from dark fibre, WDM or other technologies capable of presenting Ethernet interfaces and where the delay component primarily consists of light propagation within the fibre. This technology is clearly very important but out of scope of this document.

To allow the operators to offer the most flexibility in designing their RAN infrastructure the access network should utilise the same design paradigm as the transport core, aggregation and pre-aggregation. However, in some instances this may not be an option, so this document identifies other potential access technologies, provides a description and considerations/trade-offs for their usage.

Figure 3-1 illustrates the scope of network segments covered by WG9. The area inside the dotted orange line characterizes the transport networks composed of a number of Transport Network Elements (TNE) deployed among different components defined in other O-RAN WGs. WG9 does not define the interfaces along the dotted orange line. As an example, the fronthaul interface of an O-RU or O-DU are defined by WG4.

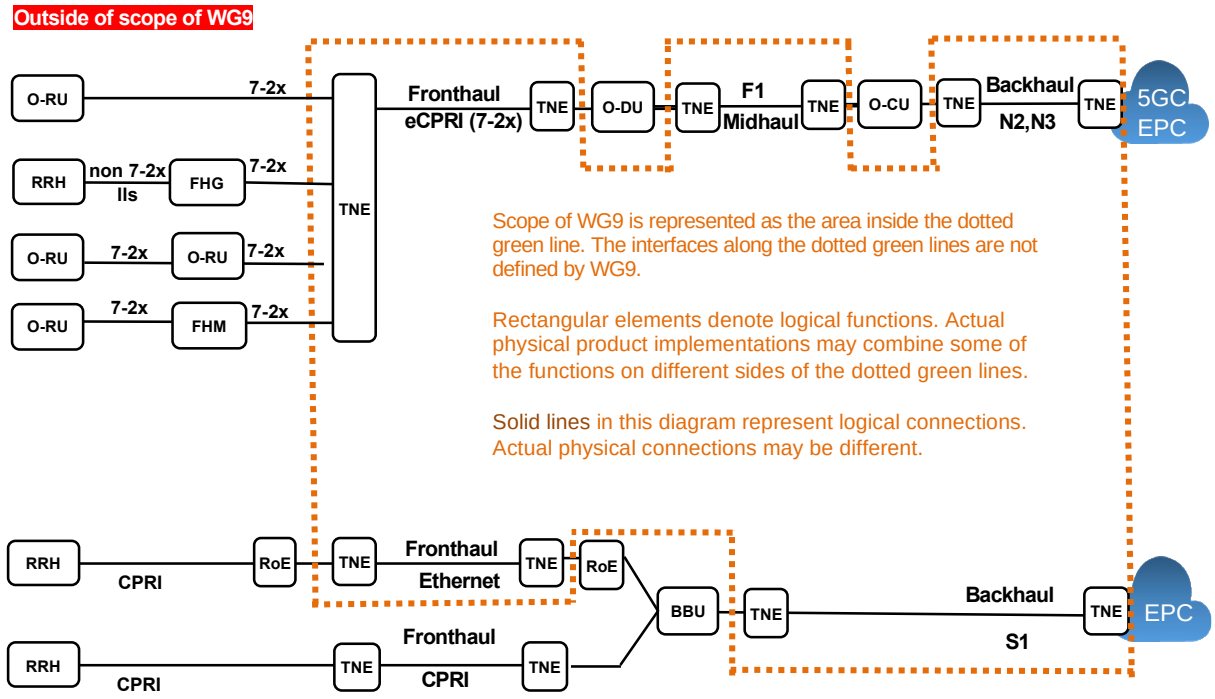


Figure 3-1 Xhaul Transport Network Overview

WG9 focuses on option 7-2x. Functional elements translating option 7-2x to non O-RAN lower layer split option (Fronthaul Gateways FHG) are considered to be part of radio network, and beyond the scope of this document. The same applies to Fronthaul Multiplexer (FHM) and cascaded radios. Radio over Ethernet mapping (RoE, IEEE 1419.3) is covered only as a service provided by the packet-based network and is not defined by WG9. The stated functions are logical ones, actual product implementation may combine several of these functions. For example, vendors may market a Fronthaul Gateway product that combines different elements such as TNE, FHG and RoE mapper in one and the same physical box.

WG9 sub-teams are working on several solution documents. This document focuses on packet-based transport technologies. It also includes sections on some of the technologies used in physical network such as PON, DOCSIS® networks and wireless Xhaul. WDM in the access network providing Fronthaul services is covered in a different document [20].

Version 1 of this document covers:

1. High level transport requirements
2. 5G Fronthaul, Midhaul, Backhaul transport requirement
3. 5G operator use cases

The document then considers packet infrastructure and how the packet infrastructure can support the identified requirements and use cases. It describes a packet architecture consisting of an underlay packet switching infrastructure which supports L2 and L3 services.

The document covers two potential underlay solutions; MPLS based or SRv6 based. In both cases the service infrastructure consists of MP-BGP VPN solutions supporting L2 and L3 services.

The last part of the document outlines, using examples, how the packet switching infrastructure can support the operator use cases outlined earlier in the document.

Version 3 builds on the MPLS and SRv6 functionality described in Version 1 and 2 and outlines a transport slicing solution capable of supporting phase 1 and 2 of WG-1 slicing architecture [24]. This is primarily covered in an informational annex (Annex F: Transport network slicing solution for WG1 Slicing (informational)).

This document does not cover:

1. Timing and synchronization –Another O-RAN WG-9 effort is underway. [21]
2. WDM in access for Fronthaul services -Another O-RAN WG-9 effort is underway[20].
3. Highly specialised URLLC, for example motion control in an industrial setting. I.e., private network, tightly constrained, requiring TSN in the Backhaul and RAN looking like a TSN bridge. This can be covered in a later edition.
4. OTN or SPN/G.mtn as a transport layer.

This document uses information and requirements published by O-RAN, 3GPP, IEEE, ITU-T, IETF, CableLabs, NGMN, MEF, BBF and many other standard bodies and industry associations.

Version 4 brings following updates/enhancements to the document:

- Section 9.4 provides more clarification, which IETF work can be referenced in this document
- in Section 11.5.1 evolution of Seamless MPLS architecture, with colored underlay transport, is added.
- major update to Section 17 provides details about desired TNE security, including physical security, user access security, transport network control plane security, and port-based access control.
- Section 18 is enhanced with additional mapping option for service slice to underlay transport plane mapping, as well has enhanced in general to provide more details about each mapping method.
- Small changes to Annex F are added to align with slicing phase 3 as described in O-RAN.WG1.Slicing-Architecture [24]
- Annex ZZZ (O-RAN Adopter License Agreement) is removed, as it is no longer needed

Version 5 brings small update to Section 17.4: Port-based network access control (IEEE 802.1X-2020), adding some discussion about the TNE-to-TNE authentication methods, as well as describing the behavior when a certificate expires.

Version 6 brings small updates to Section 10 (reference to CTI documents updated), as well as additions in Annex F related to the description of network slicing phase 4 in midhaul.

Version 7 brings update to Section 17.4, adding discussion about operator-signed certification management for TNEs providing Open Fronthaul transport connectivity. Annex F is slightly enhanced to reference relevant IETF TEAS WG 5G transport slicing draft.

4 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document in Release 17.

3GPP references

[1] 3GPP TS 23.501 V17.10.0: “System Architecture for 5G System (5GS)”

[2] 3GPP TS 23.203 V17.2.0: “Policy Control and Charging Architecture”

[3] 3GPP TS 28.316 V17.0.0: “Management and orchestration; Plug and Connect; Data formats”

[4] 3GPP TS 29.060 V17.4.0: “General Packet Radio Service (GPRS); GPRS Tunnelling Protocol (GTP) across the Gn and Gp interface”

[5] 3GPP TS 29.274 V17.9.0: “3GPP Evolved Packet System (EPS); Evolved General Packet Radio Service (GPRS) Tunnelling Protocol for Control plane (GTPv2-C); Stage 3”

[6] 3GPP TS 29.281 v17.4.0: “General Packet Radio System (GPRS) Tunnelling Protocol User Plane (GTPv1-U)

[7] 3GPP TS 32.509 V16.0.0.: “Telecommunication management; Data formats for multi-vendor plug and play eNode B connection to the network”

[8] 3GPP TS 33.310 V17.7.0: “Network Domain Security (NDS); Authentication Framework (AF)”

[9] 3GPP TS 36.422 V17.1.0: “Evolved Universal Terrestrial Radio Access Network (E-UTRAN); X2 signalling transport”

[10] 3GPP TS 36.424 v17.0.0: “Evolved Universal Terrestrial Radio Access Network (E-UTRAN); X2 data transport”

[11] 3GPP TS 38.306 V17.6.0: “NR; User equipment (UE) radio access capabilities”

[12] 3GPP TS 38.401 V17.6.0: “NG-RAN; Architecture description”

[13] 3GPP TS 38.415 V17.0.0: “NG-RAN; PDU Session User Plane Protocol”

[14] 3GPP TS 38.422 V17.1.0: “NG-RAN; Xn signalling transport”

[15] 3GPP TS 38.462 V17.0.0: “NG RAN; E1 signalling transport”

[16] 3GPP TS 38.474 V17.0.0: “NG-RAN; F1 data transport”.

O-RAN references

[17] O-RAN.WG4.CUS.0-v09.00 “Open Fronthaul Control, User and Synchronization Plane Specification Version v09.00”, August 2022

[18] O-RAN.WG4.MP.0-v09.00: “Open Fronthaul Management Plane Specification v09.00”, April 2022

[19] O-RAN.WG9.XTRP-REQ-v01.00 “Xhaul Transport Requirements v01.00”, November 2020

[20] O-RAN.WG9.WDM.0-v02.00 “WDM-based Fronthaul Transport v02.00”, November 2021

[21] O-RAN.WG9.XTRP-SYN.0-v03.00 “Synchronization Architecture and Solution Specification v03.00”, July 2022

[22] O-RAN.WG9.XTRP-MGT.0-v04.00 “Management interfaces for Transport Network Elements v04.00”, July 2022

[23] O-RAN.WG4.CTI-TCP.0-R003-v04.00, “Cooperative Transport Interface Transport Control Plane Specification v04.00”, April 2023

[24] O-RAN.WG4.CTI-TMP.0-R003-v04.00, “Cooperative Transport Interface Transport Management Procedures Specification v04.00”, April 2023

[25] O-RAN.WG1.Slicing Architecture-v07.00, April 2022

[26] O-RAN.WG11.Security-Requirements-Specifications-v04, November 2022

[27] MITRE-12-05-2022-WG1-CR-0001-UseCaseAnalysisReport-RAN-Sharing-Via-Midhaul-v09, May 2022

IEEE references

[28] IEEE 802.3-2018: “IEEE Standard for Ethernet”

[29] IEEE 802.1Q-2018: “IEEE Standard for Local and metropolitan area networks— Bridges and Bridged Networks”

[30] IEEE 802.1CM-2018: “Time-Sensitive Networking for Fronthaul”

[31] IEEE Std 1914.1TM-2019: “IEEE Standard for Packet-based Fronthaul Transport Network”

[32] IEEE 802.3av “Physical Layer Specifications and Management Parameters for 10 Gb/s Passive Optical Networks”

[33] IEEE 802.3bk “Physical Layer Specifications and Management Parameters for Extended Ethernet Passive Optical Networks”

[34] IEEE 802.3ca “Physical Layer Specifications and Management Parameters for 25 Gb/s and 50 Gb/s Passive Optical Networks”

[35] IEEE Std 1914.3-2018 – IEEE standard for radio over Ethernet Encapsulations and Mappings

[36] IEEE Std 802.1X-2020 “IEEE Standard for Local and Metropolitan Area Networks--Port-Based Network Access Control”

IETF references

[37] IETF RFC 791: “INTERNET PROTOCOL”

[38] IETF RFC 768: “User Datagram Protocol”

[39] IETF RFC 1195: “Use of OSI IS-IS for routing in TCP/IP and dual environments”

[40] IETF RFC 1771: “A Border Gateway Protocol 4”

[41] IETF RFC 2205: “Resource ReSerVation Protocol”

[42] IETF RFC 2209: “Resource ReSerVation Protocol (RSVP) --Version 1 Message Processing Rules”
[43] IETF RFC 2210: “The use of RSVP with IETF Integrated Services”
[44] IETF RFC 2328: “OSPF Version 2”
[45] IETF RFC 2283: “Multiprotocol Extensions for BGP-4”
[46] IETF RFC 2475: “An Architecture for Differentiated Services”
[47] IETF RFC 2474: “Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers”
[48] IETF RFC 2545: Use of BGP-4 Multiprotocol Extensions for IPv6 Inter-Domain Routing”
[49] IETF RFC 2597: “Assured Forwarding PHB Group”
[50] IETF RFC 2598: “An Expedited Forwarding PHB”
[51] IETF RFC 2745: “RSVP Diagnostic Messages”
[52] IETF RFC 2747: “RSVP Cryptographic Authentication”
[53] IETF RFC 2865: “Remote Authentication Dial In User Service (RADIUS)”
[54] IETF RFC 2866: “RADIUS Accounting”
[55] IETF RFC 2961: “RSVP Refresh Overhead Reduction Extensions”
[56] IETF RFC 2983: “Differentiated Services and Tunnels”
[57] IETF RFC 3031: “Multiprotocol Label Switching Architecture”
[58] IETF RFC 3032: “MPLS Label Stack Encoding”
[59] IETF RFC 3097: “RSVP Cryptographic Authentication—Updated Message Type Value”
[60] IETF RFC 3209: “Extensions to RSVP for LSP Tunnels”
[61] IETF RFC 3212: “Constraint-Based LSP Setup using LDP”
[62] IETF RFC 3215: “LDP State Machine”
[63] IETF RFC 3246: “An Expedited Forwarding PHB (Per-Hop Behavior)
[64] IETF RFC 3107: “Carry Label Information in BGP-4”
[65] IETF RFC 3443: “Time To Live (TTL) Processing in Multi-Protocol Label Switching (MPLS) Networks”
[66] IETF RFC 3477: “Unnumbered Links in Resource ReSerVation Protocol - Traffic Engineering (RSVP-TE)”
[67] IETF RFC 3478: “Graceful Restart Mechanism for Label Distribution Protocol”
[68] IETF RFC 3579: “RADIUS (Remote Authentication Dial In User Service) Support For Extensible Authentication Protocol (EAP)”
[69] IETF RFC 3630: “Traffic Engineering (TE) Extensions to OSPF Version 2”
[70] IETF RFC 3719: “Recommendations for Interoperability using ISIS”
[71] IETF RFC 4072: “Diameter Extensible Authentication Protocol (EAP) Application”
[72] IETF RFC 4090: “Fast Reroute Extensions to RSVP-TE for LSP Tunnels”
[73] IETF RFC 4115: “A Differentiated Service Two-Rate, Three-color Market with Efficient Handling of in-Profile Traffic”
[74] IETF RFC 4182: “Removing a Restriction on the use of MPLS Explicit NULL”
[75] IETF RFC 4210: “Internet X.509 Public Key Infrastructure Certificate Management Protocol (CMP)”
[76] IETF RFC 4271: “A Border Gateway Protocol 4 (BGP-4)”
[77] IETF RFC 4303: “OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)”
[78] IETF RFC 4206: Label Switched Paths (LSP) Hierarchy with Generalized Multi-Protocol Label Switching (GMPLS) Traffic Engineering (TE)”

[79] IETF RFC 4443: “ICMPv6 (ICMP for IPv6)”
[80] IETF RFC 4552: “Authentication/Confidentiality for OSPFv3”
[81] IETF RFC 4558: “Node-ID Based Resource Reservation Protocol (RSVP) Hello”
[82] IETF RFC 4561: “Record Route Object (RRO) Node-Id Sub-Object”
[83] IETF RFC 4594: “Configuration Guidelines for DiffServ Service Classes”
[84] IETF RFC 4364: “BGP/MPLS IP Virtual Private Networks (VPNs)”
[85] IETF RFC 4684: “Constrained Route Distribution for BGP/MPLS IP VPNs” Computation Element Protocol (PCEP)”
[86] IETF RFC 4760: “Multiprotocol Extensions for BGP-4”
[87] IETF RFC 4875: “Extensions to Resource Reservation Protocol – Traffic Engineering for Point to Multipoint TE Label Switched Paths”
[88] IETF RFC 5036: “LDP Specification”
[89] IETF RFC 5130: “A Policy Control Mechanism in IS-IS Using Administrative Tags”
[90] IETF RFC 5216: “The EAP-TLS Authentication Protocol”
[91] IETF RFC 5283: “LDP Extension for Inter-Area Label Switched Paths (LSPs)”
[92] IETF RFC 5291: “Outbound Route Filtering Capability for BGP-4”
[93] IETF RFC 5292 “Address-Prefix-Based Outbound Route Filter for BGP-4”
[94] IETF RFC 5302 “Domain-Wide Prefix Distribution with Two-Level IS-IS”
[95] IETF RFC 5303: “Three-Way Handshake for IS-IS Point-to-Point Adjacencies”
[96] IETF RFC 5305: “IS-IS Extensions for Traffic Engineering”
[97] IETF RFC 5308: “Routing IPv6 with ISIS”
[98] IETF RFC 5310: “IS-IS Generic Crypto Authentication”
[99] IETF RFC 5329: “Traffic Engineering Extensions to OSPF Version 3”
[100] IETF RFC 5340: “OSPF for IPv6”
[101] IETF RFC 5420: “Encoding of Attributes for MPLS LSP Establishment Using Resource Reservation Protocol Traffic Engineering (RSVP-TE)”
[102] IETF RFC 5440: “Path Computational Element (PCE) Communications Protocol (PCEP)”
[103] IETF RFC 5443: “LDP IGP Synchronization”
[104] IETF RFC 5496: “The Reverse Path Forwarding (RPF) Vector TLV”
[105] IETF RFC 5512: “The BGP Encapsulation Subsequent Address Family Identifier (SAFI) and the BGP Tunnel Encapsulation Attribute”
[106] IETF RFC 5561: “LDP capabilities”
[107] IETF RFC 5838: “Support of Address Families in OSPFv3”
[108] IETF RFC 5925: “The TCP Authentication Option”
[109] IETF RFC 5926: “Cryptographic Algorithms for the TCP Authentication Option (TCP-AO)”
[110] IETF RFC 6232: “Purge Originator Identification TLV for IS-IS”
[111] IETF RFC 6388: “Label Distribution Protocol Extensions for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths”
[112] IETF RFC 6391: “Flow-Aware Transport of pseudowires over an MPLS Packet Switched Network”
[113] IETF RFC 6512: “Using Multipoint LDP When the Backbone has No Route to the ROOT”
[114] IETF RFC 6513: “Multicast in MPLS/BGP IP VPNs”
[115] IETF RFC 6514: “BGP Encodings and procedures for multicast in MPLS/BGP IP VPNs”

[116] IETF RFC 7432: “BGP MPLS-Based Ethernet VPN”
[117] IETF RFC 7471: “OSPF Traffic Engineering (TE) Metric Extensions”
[118] IETF RFC 7524: “Inter-Area Point to Multipoint Segmented Label Switched Paths (LSPs)”
[119] IETF RFC 7570: “Label Switched Path (LSP) Attribute in the Explicit Route Object (ERO)”
[120] IETF RFC 7752: “BGP Link State (BGP-LS)”
[121] IETF RFC 7761: “Protocol Independent Multicast”
[122] IETF RFC 7810: “IS-IS Traffic Engineering (TE) Metric Extensions”
[123] IETF RFC 8029: “Detecting Multiprotocol Label Switched (MPLS) Data-plane Failures”
[124] IETF RFC 8200: “Internet Protocol, Version 6 (IPv6) Specification”
[125] IETF RFC 8214: “Virtual Private Wire Service Support in Ethernet VPN”
[126] IETF RFC 8231: “Path Computational Element Communications Protocol (PCEP) Extensions for Stateful PCE”
[127] IETF RFC 8277: “BGP and Labeled Address Prefixes”
[128] IETF RFC 8287: “Label Switched Path (LSP) Ping/Traceroute for Segment Routing (SR) IGP-Prefix and IGP-Adjacency Segment Identifiers (SIDs) with MPLS Data Planes”
[129] IETF RFC 8317: “Ethernet-Tree (E-Tree) Support in Ethernet VPN (EVPN) and Provider Backbone Bridging EVPN (PBB-EVPN)”
[130] IETF RFC 8366: “A Voucher Artifact for Bootstrapping Protocols”
[131] IETF RFC 8370: “Techniques to Improve the Scalability of RSVP-TE Deployments”
[132] IETF RFC 8395: “Extension to BGP-Signaled pseudowires to support flow-aware transport labels”
[133] IETF RFC 8402: “Segment Routing Architecture”
[134] IETF RFC 8476: “Signaling Maximum SID depth using OSPF”
[135] IETF RFC 8491: “Signaling Maximum SID depth using IS-IS”
[136] IETF RFC 8571: “BGP – Link State (BGP-LS) Advertisement of IGP Traffic Engineering Performance Metric Extensions”
[137] IETF RFC 8572: “Secure Zero Touch Provisioning (SZTP)”
[138] IETF RFC 8577: “Signaling RSVP-TE Tunnels on a Shared MPLS Forwarding Plane”
[139] IETF RFC 8660: “Segment Routing with MPLS Dataplane”
[140] IETF RFC 8661: “Segment Routing MPLS Interworking with LDP”
[141] IETF RFC 8664: “PCEP Extensions for Segment Routing”
[142] IETF RFC 8666: “OSPFv3 Extensions for Segment Routing”
[143] IETF RFC 8667: “IS-IS Extensions for Segment Routing”
[144] IETF RFC 8754: “IPv6 Segment Routing Header (SRH)”
[145] IETF RFC 8814: “Signaling Maximum SID Depth (MSD) Using the BGP – Link State”
[146] IETF RFC 8907: “The Terminal Access Controller Access-Control System Plus (TACACS+) Protocol”
[147] IETF RFC 8986: “SRv6 Network Programming”
[148] IETF RFC 9085: “Border Gateway Protocol - Link State (BGP-LS) Extensions for Segment Routing”
[149] IETF RFC 9136: “IP prefix Advertisements in Ethernet VPN (EVPN)

[150] IETF RFC 9252: “BGP Overlay Services Based on Segment Routing over IPv6 (SRv6)”
[151] IETF RFC 9256: “Segment Routing Policy Architecture”
[152] IETF RFC 9259: “Operations, Administration, and Maintenance (OAM) in Segment Routing Networks over IPv6 (SRv6)”
[153] IETF RFC 9350: “IGP Flexible Algorithm”
[154] IETF RFC 9351: “Flexible Algorithm Definition Advertisement with BGP Link-State”
[155] IETF RFC 9352: “IS-IS Extensions to support SR over IPv6 dataplane”
[156] Draft-ietf-bess-bgp-multicast-controller-11: “Controller based BGP Multicast signalling”
[157] Draft-ietf-bess-evpn-bum-procedure-updates-14: “Updates on EVPN BUM procedures” – submitted to IESG for publication
[158] Draft-ietf-bess-evpn-vpws-fxc-08: “EVPN VPWS Flexible Cross-Connect Service”
[159] Draft-ietf-bess-evpn-mh-pa-08: “EVPN multi-homing port-active load-balancing”
[160] Draft-ietf-bess-evpn-pref-df-11: “Preference-based EVPN DF Election”
[161] Draft-ietf-mpls-ri-rsvp-frr-15: “Refresh-interval independent FRR facility protection”
[162] Draft-ietf-pim-sr-p2mp-policy-06: “Segment Routing Point-to-Multipoint Policy”
[163] Draft-ietf-rtgwg-segment-routing-ti-lfa-11: “Topology Independent Fast Reroute using Segment Routing”
[164] Draft-ietf-pce-binding-label-sid-16: “Carrying Binding Label/Segment-ID in PCE-based Networks”
[165] Draft-ietf-pce-segment-routing-ipv6-18: “PCEP Extensions for Segment Routing leveraging the IPv6 data plane”
[166] Draft-ietf-pce-segment-routing-policy-cp-12: “PCEP extension to support Segment Routing Candidate Path
[167] Draft-ietf-lsr-ospfv3-srv6-extensions-15: “OSPFv3 Extensions for SRv6”
[168] Draft-ietf-idr-bgpls-srv6-ext-14: “BGP Link State extensions for IPv6 Segment Routing (SRv6)”
[169] Draft-ietf-mpls-seamless-mpls-07: “Seamless MPLS Architecture”
[170] Draft-ietf-spring-srv6-srh-compression-06: “Compressed SRv6 Segment List Encoding in SRH”
[171] Draft-ietf-teas-5g-ns-ip-mpls-00: “A Realization of IETF Network Slices for 5G Networks Using Current IP/MPLS Technologies”
[172] Draft-ietf-netconf-trust-anchors-21: “A YANG Data Model for a Truststore”

Others

[173] NGMN “5G RAN CU-DU network architecture, transport options and dimensioning, version 1.0 12 April 2019”
[174] MEF 61.1: “IP Service attributes”
[175] MEF 10.3: “Ethernet Service attributes”
[176] MEF 6.2: “EVC service definition”
[177] Broadband Forum TR-101 “Migration to Ethernet-Based Broadband Aggregation”
[178] Broadband Forum TR-156 “Using GPON Access in the context of TR-101“

[179] ITU-R M.2083: “IMT Vision – framework and overall objectives of the future development of IMT for 2020 and beyond.

[180] ITU-T GSTR-TN5G – Transport network support of IMT 2020/5G

[181] ITU-T G.Sup.66 “5G wireless Fronthaul requirements in a passive optical network context”

[182] ITU-T G.9807 series “10-Gigabit-capable symmetric passive optical network”, ITU-T G.989 series “40-Gigabit-capable passive optical networks (NG PON2)“, On-going work ITU-T G.HSP

[183] ITU-T G.8271 “Time and phase synchronization aspects of telecommunication networks”

[184] ITU-T G.8271.1 “Network limits for time synchronization in packet networks with full timing support from the network”

[185] ITU-T G.8273.2 “Timing characteristics of telecom boundary clocks and telecom time slave clocks”

[186] ITU-T G.8275.1 “Precision time protocol telecom profile for phase/time synchronization with full timing support from the network”

[187] 40m Transmission of OAM mode and Polarization Multiplexing in E-band, Globcom, Dec 2019. M.Hirabe, et. Al

[188] CableLabs “Low Latency Mobile Xhaul over DOCSIS Technology” https://www.cablelabs.com/specifications/CM-SP-LLX

[189] CableLabs “Synchronization Techniques for DOCSIS Technology Specification” https://www.cablelabs.com/specifications/CM-SP-SYNC

[190] CableLabs “Data-Over-Cable Service Interface Specifications 3.1, MAC and Upper Layer Protocols Interface” https://www.cablelabs.com/specifications/CM-SP-MULPIv3.1

[191] CableLabs “Data-Over-Cable Service Interface Specifications 4.0, MAC and Upper Layer Protocols Interface” https://www.cablelabs.com/specifications/CM-SP-MULPIv4.0

[192] CableLabs “Remote PHY Specification” https://www.cablelabs.com/specifications/CM-SP-R-PHY

[193] “Study on new radio access technology: Radio access architecture and interfaces” 3GPP TR 38.801 Table A-1.

[194] Cisco Press, MPLS and VPN Architectures, Volume 1, 420 pages, by Ivan Pepelnjak, and Jim Guichard, 2001

[195] Cisco Press, MPLS and VPN Architectures, Volume 2, 470 pages, by Ivan Pepelnjak, Jim Guichard, and Jeff Apcar, 2003

[196] O’Reilly, MPLS in the SDN Era, 890 pages, by Antonio Sánchez-Monge, and Krzysztof Grzegorz Szarkowicz, 2015

[197] BBF TR-221: Technical Specification for MPLS in Mobile Backhaul Networks, 99 pages, Oct 2011

[198] BBF TR-221, Amd.1: Technical Specifications for MPLS in Mobile Backhaul Networks, 24 pages, Nov 2013

[199] BBF TR-221, Amd.2: Technical Specifications for MPLS in Mobile Backhaul Networks, 22 pages, Sep 2017

[200] ETSI GR mWT 012 V1.1.1 (2018-11): 5G Wireless Backhaul/Xhaul

[201] Microwave and millimeter-wave technology overview and evolution, Workshop on Evolution of Fixed Service in Backhaul support of IMT 2020 / 5G, Geneva, 29 April 2019, https://www.itu.int/en/ITU-R/study-groups/workshops/fsimt2020/Pages/default.aspx

[202] ISO/IEC 10589-2002: “Intermediate System to Intermediate System intra-domain routeing information exchange protocol for use in conjunction with the protocol for providing the connectionless-mode network service”

5 Definitions and abbreviations

5.1 Definitions

The key words "**SHALL**", "**SHALL NOT**", "**SHOULD**", "**SHOULD NOT**", "**MAY**", and "**OPTIONAL**" in this document are to be interpreted as described in IETF RFC 2119 [25]. All key words must be in upper case, bold text.
Items that are **REQUIRED** (contain the words **SHALL** or **SHALL NOT**) will be labelled as **[Rx]** for required. Items that are **RECOMMENDED** (contain the words **SHOULD** or **SHOULD NOT**) will be labelled as **[Dx]** for desirable. Items that are **OPTIONAL** (contain the words **MAY** or **OPTIONAL**) will be labelled as **[Ox]** for optional**.**
Items, if supported, are not meant to be active at all times, but should be available for use. Their state (active or not active) should be based on configuration.

5.2 Abbreviations

Abbreviations defined in this document take precedence over the definition of 3GPP

Abbreviations	Meaning or explanation
3GPP	Third Generation Partnership Project – Standards Development Organization
4G	Fourth-generation mobile network
5G	Fifth-generation mobile network
ABR	Area Border Router
Amd	Ammendment (term used by BBF)
API	Application Programming Interfaces
ARPU	Average revenue per user
AS	Automated (traffic) steering
AS	Autonomous System
ASBR	Autonomous System Border Router
BBF	Broadband Forum
BBU	Baseband unit
BGP	Border Gateway Protocol
BGP-LU	BGP labelled unicast
BGP-LS	BGP link state
BIER	Bit Indexed Explicit Replication
BITS	Building Integrated Timing System
C-RAN	Centralized Radio Access Network
CA	Carrier Aggregation
CA	Certification Authority
CapEx	Capital expenditure

CBWRED	Class based Weighted Random Early Detection
CDN	Content delivery network
CDMA	Code Division Multiple-Access – a mobile radio standard
CMPv2	Certificate Management Protocol version 2
COTS	Commercial off the shelf
CPE	Customer premises equipment
CPRI	Common public radio interface
CSP	Communications service provider
CSR	Cell Site Router
CU	Centralized unit
CUPS	Control/User Plane Separation
D-RAN	Distributed Radio Access Network
DB	Dynamic Bandwidth assignment
DC	Data center
DHCP	Dynamic Host Configuration Protocol
DHCPv4	Dynamic Host Configuration Protocol for IPv4 hosts
DHCPv6	Dynamic Host Configuration Protocol for IPv6 hosts
DiffServ	Differentiated services – a quality-of-service mechanism
DSCP	DiffServ Code Point
DWDM	Dense Wavelength Division Multiplexing
EAP	Extensible Authentication Protocol
ECMP	Equal-cost multipath
eCPRI	Enhanced Common Radio Interface (CPRI)
eMBB	Enhanced mobile broadband
eNB	Enhanced Node B
EPC	Evolved packet core
ESMC	Ethernet Synchronization Message Channel
EVPN	Ethernet VPN
EXP	Experimental
FDD	Frequency division duplexing
FHG	Fronthaul Gateway
FIB	Forwarding Information Base
FIFO	First In, First Out
FMC	Fixed-mobile convergence
FMC	Fixed-mobile convergence
FQDN	Fully Qualified Domain Name
Fronthaul	Portion of the mobile network supporting O-RAN 7.2x, eCPRI, RoE or CPRI protocols
FRR	Fast Re-Route
Gbps	Gigabits per second

GNSS	Global Navigation Satellite System (example being GPS)
GPS	Global Positioning System
HLS	High-level split
HSR	Hub Site Router
IANA	Internet Assigned Numbers Authority
iBGP	internal Border Gateway Protocol
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers – Standards Development Organization
IETF	Internet Engineering Task Force – Standards Development Organization
IGP	Interior Gateway Protocol
IoT	Internet of Things (see also mMTC)
IP	Internet Protocol
IS-IS	Intermediate System to Intermediate System
ISO	International Standardization Organization
ITU-T	International Telecommunication Union-Telecommunication – Standards Development Organization
ITU-R	International Telecommunication Union-Radiocommunication – Standards Development Organization
JSON	JavaScript Object Notation
L1 / L2 / L3	Layer 1 / layer 2 / layer 3 of the network protocol stack
L3VPN	Layer 3 Virtual Private Network
LDP	Label Distribution Protocol
LLS	Low-level splits
LSDB	Link state database
LSP	Label Switched Path
LTE	Long Term Evolution (generation of Mobile networks – see 4G)
LTE-A	Long Term Evolution – Advanced
MEC	Formerly Mobile Edge Compute, now Multi-Access Edge Compute
mMTC	Massive machine type communications
MIMO	Multiple-Input Multiple-Output (number of antennas)
MNO	Mobile Network Operator
MP-BGP	Multi-protocol Border Gateway Protocol
MPLS	Multiprotocol Label Switching
MSD	Maximum Segment Depth
MVNO	Mobile virtual network operator
NAT	Network Address Translation
NETCONF	Network Configuration Protocol
NLRI	Network Layer Reachability Information
NFV	Network functions virtualization
NGFI	Next-generation Fronthaul interface
NR	New radio

NSI	Network Slice Instance
NSSI	Network Subnet Slice Instance
NTP	Network Time Protocol
O-CU	Open Central Unit
O-DU	Open Distributed Unit
O-RU	Open Radio Unit
OAM	Operations, administration, and maintenance
ODN	On-demand next hop
ODN	Optical Distribution Network
OLT	Optical Line Termination
ONU	Optical Network Unit
OpEx	Operational expenses
ORF	Outbound Route Filter
OSPF	Open Shortest Path First
OTN	Optical Transport Networking
P	Provider (router)
PCC	Path computation client
PCE	Path computation element
PCEP	Path computation element protocol
PE	Provider edge (router)
PIM	Protocol for IP Multicast
PKI	Public Key Infrastructure
PON	Passive Optical Network
PTP	Precision Time Protocol
QoS	Quality of Service
RA	Registration Authority
RAN	Radio access network
RE	Radio equipment
REC	Radio equipment controller
RoE	Radio over Ethernet
RRO	Record Route Object
RRU	Remote Radio Unit
RSVP	Resource Reservation Protocol
RSVP-TE	Resource Reservation Protocol – Traffic Engineering
RT	Route Target
RTT	Round-trip times
RU	Remote (radio) unit
SD-WAN	Software-defined wide area network
SDH	Synchronous Digital Hierarchy – a digital communications system

SDN	Software-defined networks
SID	Segment Identifier
SLA	Service level agreement
SMB	Small and medium business
SONET	Synchronous Optical Networking – a digital communications system
SPF	Shortest Path First
SR	Segment Routing
SR-DPM	Segment Routing-data plane management
SR-TE	Segment Routing – traffic engineering
SR-PCE	Segment Routing – path computation element
SRH	Segment Routing header
SRLG	Shared risk link groups
SSU	Synchronization Supply Unit
SyncE	Synchronous Ethernet
T-GM	Telecom – Grand Master – a PTP clock type
T-TSC	Telecom – Time Slave Clock – a PTP clock type
TAE	Time Alignment Error
TC	Traffic Class
TDD	Time-division duplexing – a radio communications technique
TDM	Time division multiplexing
TE	Traffic engineering
TE	Time Error
TI-LFA	Topology-independent loop-free alternative
TLS	Transport Layer Security
TLV	Type-Length-Value
TNE	Transport Network Equipment (an O-RAN term to denote a transport device)
ToD	Time of the Day
TR	Technical report
TTL	Time to live
TWDM	Time and Wavelength Domain Multiplexing
UE	User equipment
URLLC	Ultra-reliable low-latency communications
UPF	User plane functions
VIM	Virtualized infrastructure managers
VNF	Virtual network function(s)
VPN	Virtual private networks
VRF	Virtual Routing and Forwarding
WAN	Wide area network
WDM	Wavelength Division Multiplexing

Xhaul	Collective name for Fronthaul, Midhaul, and Backhaul
XML	eXtensible Markup Language
YANG	Yet another next generation – data modelling language

1

6 5G Transport network requirements

For full details of the O-RAN transport requirements, bandwidth and delay estimates of a 5G network see [19].

Requirements for the transport architecture can be characterized in following categories:

1. Latency, Frame Loss Ratio and Bandwidth requirements for Fronthaul, Midhaul, and Backhaul.
2. Operability requirements that include fault and performance management.
3. Synchronization requirements.
4. "ITU-T GSTP-TN5G: Transport support of IMT-2020/5G" [180] identifies the need for the transport to be multi-service in nature. In addition to mobile services, the infrastructure needs to support fixed line consumer and enterprise services. These services are not explicitly covered in the document, but the architecture must enable L2 or L3 services to be created between any two edge TNEs regardless of relative position to each other in the transport network.
5. End to end support of 4G/5G mobile infrastructure including Fronthaul / Midhaul / Backhaul.
6. Concurrent support for RAN deployment scenarios outlined in "ITU-T GSTP-TN5G: Transport support of IMT-2020/5G" [180] running from a single cell site location. These shown in Figure 6-1 are:

    a. Co-located O-CU and O-DU – O-RAN split 7.2x from cell site
    b. Independent O-RU, O-CU, O-DU locations – O-RAN 7.2x from cell site
    c. O-RU and O-DU integration on cell site – O-RAN split 2 from cell site
    d. O-RU, O-DU and O-CU integration on cell sites – Split 1 from cell site

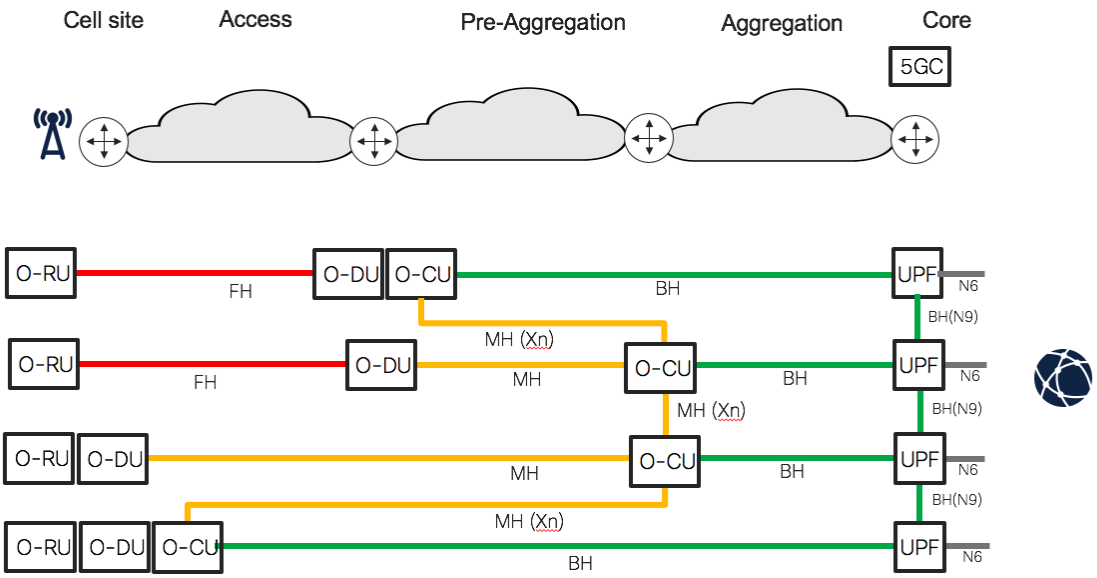


Figure 6-1: ITU-T 5G use cases

7. Central, distributed or a mix of the two for user plane termination. Central, distributed or a mix of two for 5G control plane placement. (Figure 6-2)

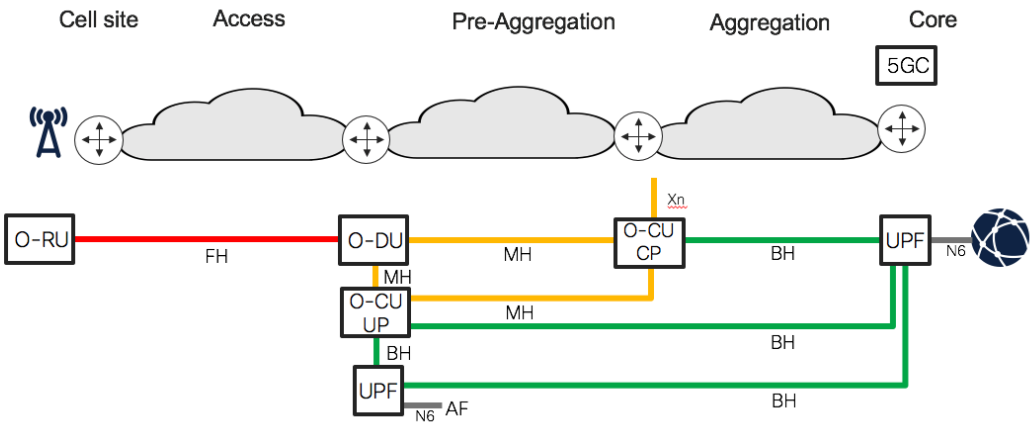


Figure 6-2: 5G central and distributed user plane termination

8. Concurrent support for 4G and 5G RAN solutions running from a single cell site. In addition to the use cases outlined in earlier figures, support for 4G radios running alongside 5G radios is required. Figure 6-3 illustrates the two 4G architectures that need to co-exist with the 5G architectures outlined above. In the upper case, the 4G RAN infrastructure utilizes a C-RAN architecture where the RRH and BBU support split 8 which is converted to 7.2x or RoE for transportation across the packet access infrastructure. In the lower case, the 4G RAN infrastructure uses D-RAN or split 1 architecture.

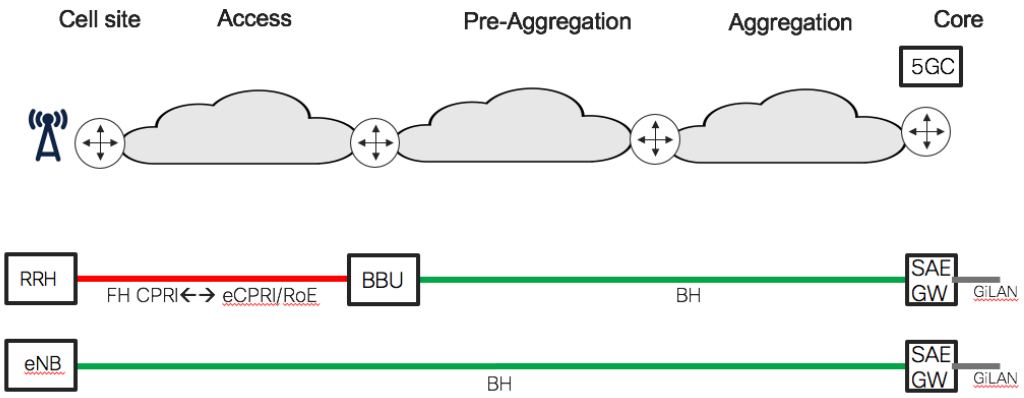


Figure 6-3: 4G use cases

9. End to end support for 5G slicing and 5G service types.

7 5G logical connectivity requirements

The transport network needs to be very flexible as depending on the use case and the RAN design each part of the physical transport network may need to support multiple slices, multiple 5G services and also different 3GPP interfaces. This section covers the logical transport connectivity requirements of the 5G Fronthaul, Midhaul and Backhaul components. Further details can be found in O-RAN WG9.Transport Requirements document [19].

7.1 Fronthaul

The Fronthaul infrastructure potentially needs to support:

- O-RAN 7.2x Fronthaul (for 5G NR)
- Non-ORAN Fronthaul

7.1.1 O-RAN 7.2x Fronthaul

The O-RAN 7.2x is a split 7 “Low Level Split” (LLS) that runs between the O-RU and the O-DU (optional more than one for network and DU based redundancy). The associated mobile interfaces for the Fronthaul are the Control, User and Synchronization and Management planes. The synchronization plane is covered in a separate WG-9 Timing and Synchronization Architecture and solutions document [21].

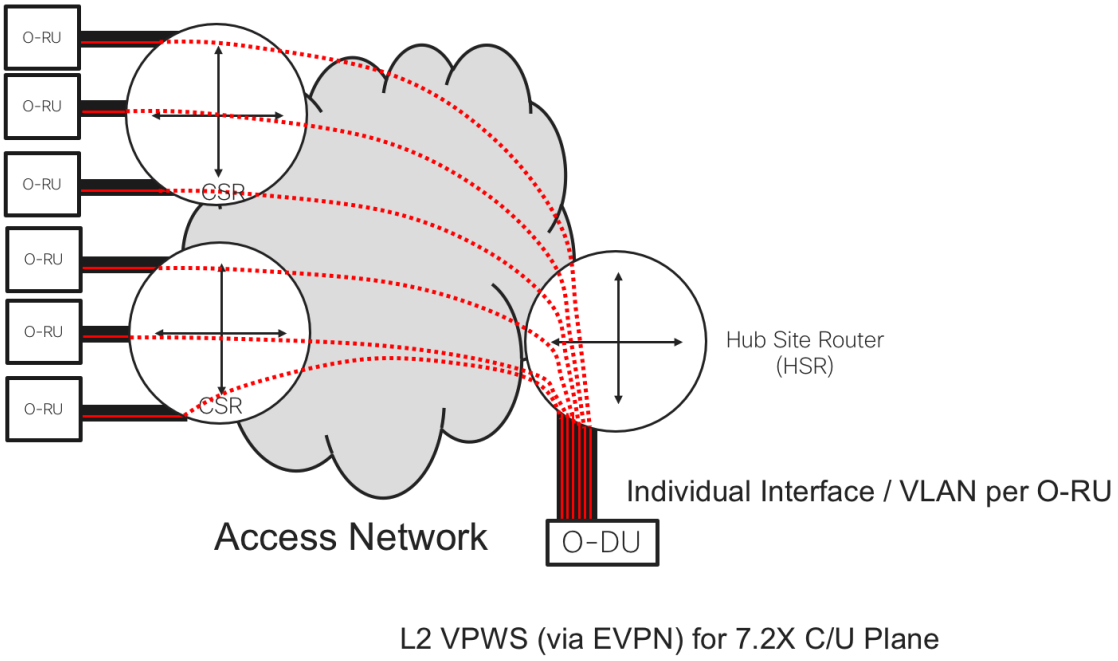


Figure 7-1: Fronthaul O-RAN 7.2x control and user plane using an Ethernet encapsulation

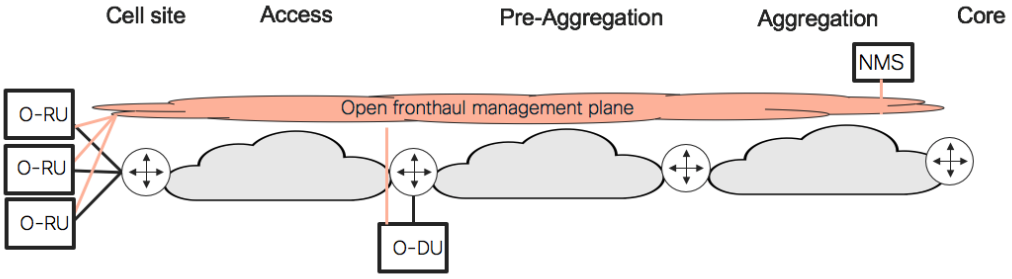


Figure 7-2: O-RAN 7.2x Hybrid Management plane

- **O-RAN 7.2x Control and User planes:** These interfaces are described in O-RAN.WG4.CUS.0-v09.00: Control, User and Synchronization Plane Specification [17]. These interfaces run between the O-RUs and their serving O-DU. Ethernet encapsulation is a mandatory requirement and IP encapsulation is optional and applies if the transmitting and receiving nodes support IP capabilities. In both cases the payload is one or more eCPRI transport headers with respective application data. The latency requirements associated with 7.2x control plane and user plane traffic are very low, and the bandwidth requirements are generally high but vary based on the level of user data traffic being transmitted.
- **O-RAN 7.2x Synchronization plane:** This interface is described in O-RAN.WG4.CUS.0-v9.00: Control, User and Synchronization Plane Specification [17]. In C-RAN architectures accurate synchronization between the O-DU and O-RUs is required to support "Time Division Duplex" (TDD), Carrier Aggregation (CA) using multiple O-RUs, MIMO and other processes. In an O-RAN Fronthaul environment using an Ethernet transport layer, protocols such as PTP and SyncE are used to achieve synchronization between the O-RUs and O-DUs. For more details refer to WG-9 Timing and Synchronization Architecture and solutions document [21].
- **O-RAN 7.2x Management plane:** This interface is described in O-RAN.WG4.MP.0-v09.00 Management plane specification[18]. Two M-Plane models are defined.
    - Hierarchical model: In this model, an O-RU is managed by one of more O-DUs. These O-DUs are entirely responsible for sub-ordinate O-RUs, which means the NMS only needs to interact with the O-DU level. In this mode the O-RAN 7.2x M-Plane interface only runs between the O-DU and sub-ordinate O-RUs.
    - Hybrid model: In this model, an O-RU is managed by one or more NMSs, in addition to the serving O-DU. In this mode the O-RAN 7.2x M-Plane interface runs between the O-RUs, O-DUs and the NMS.

  The O-RAN 7.2x M-plane uses IP/NETCONF and the basic transport requirement is end-to-end IP connectivity between the O-RU and the elements managing it. IPv4 shall be supported as a mandatory transport protocol for M-Plane and IPv6 support is optional.

7.1.2 O-RAN Fronthaul logical transport requirements

Details of the O-RAN transport requirements are illustrated in Figure 7-1.

C/U-Planes:

1. Ethernet connectivity from O-RUs to serving O-DU with potential backup to a redundant O-DU.

2. Optional IP connectivity from O-RUs to serving O-DU with potential backup to a redundant O-DU.
3. O-RUs and serving O-DU in close proximity to each other to meet delay criteria associated with Fronthaul. It is unlikely the Fronthaul components (ie O-RUs and serving O-DU) will extend beyond the access transport network.

S-Plane:

1. This is covered in the "WG-9 Timing and Synchronization Architecture and Solutions document [21].

M-Plane

Details of the O-RAN M-Plane transport running in hybrid mode are illustrated in Figure 7-2.

1. IP connectivity allowing NMS to communicate with O-DUs and O-RUs if running the hybrid model.
2. IP connectivity allowing serving O-DU to communicate with all its sibling O-RUs if running in either hybrid or hierarchical models.

Note: In addition to the 7.2x management plane other management components may need to communicate with entities at the cell site. For example, remote monitoring of sensors and actuators in the cell site. These are not explicitly covered in this document, but the specified transport architecture can cater for scenarios, where management and monitoring, are based on either Ethernet or IP connectivity.

7.2 Non-ORAN Fronthaul

Legacy Fronthaul scenarios are those C-RAN use cases where the Fronthaul traffic is transported over a packet switch network while not using the O-RAN compliant encapsulation protocol that supports 7.2x split. The two most likely legacy Fronthaul scenario in a packet switched transport network are:

- eCPRI based C-RAN solutions: using eCPRI encapsulation protocol not compliant to O-RAN WG4 CUS specifications [17]. i.e., a non O-RAN 7.2x split.
- RoE based C-RAN solutions: CPRI encapsulated by the RoE protocol.

7.2.1 eCPRI based C-RAN solutions

An operator may choose a packet-based C-RAN Fronthaul architecture that is not O-RAN 7.2x compliant. In this case the RU and DU uses eCPRI as the packet encapsulation protocol to packetize the Fronthaul data but implements a non-O-RAN compliant radio message protocol to support a different function split.

Following 3GPP recommendation [9], the possible function splits may include

- Option 6, split between MAC and PHY layers
- Option 7.1, 7.2, or 7.3, splits within PHY layer
- Option 8, split between radio and PHY layer

7.2.2 Radio over Ethernet (RoE) based C-RAN solutions

In this scenario an operator has an existing radio deployed at the cell site and the operator intends to convert these radios to support a C-RAN Fronthaul architecture whilst implementing their 5G RAN infrastructure. In this case it is likely the legacy equipment (RRH and BBU) only supports an optical CPRI interfaces. To implement a C-RAN Fronthaul architecture in a packet switched transport network the CPRI needs to be converted to an Ethernet frame at the cell site, transported over the access network as packet and then get converted back to CPRI at the hub for processing from the BBU. In this scenario a component called a RoE mapper is used to perform a CPRI to Ethernet frame conversion at the cell site and an Ethernet frame to CPRI conversion at the location where the BBU is located.

7.2.2.1 RoE Mapper transport

The design, implementation, management, and interface from the RoE mapper to the legacy equipment is not in the scope of WG-9 but the expectation is the RoE mapper will present either O-RAN 7.2x compliant Ethernet or Ethernet IP packets or IEEE 1914.3 Radio over Ethernet packets [35] to the transport network. At this time O-RAN has not defined a CPRI to O-RAN 7.2x conversion capability so implementations out in the market are based on CPRI to Radio Over Ethernet.

It is then the responsibility of the transport network to transport these packets to their destination with appropriate characteristics for the legacy connection to function.

7.2.2.2 Radio over Ethernet

Radio over Ethernet (RoE) is defined in IEEE Std 1914.3-2018 – IEEE standard for radio over Ethernet Encapsulations and Mappings [35]. Two mapping techniques for supporting CPRI to packet conversion are defined.

Structure-agnostic RoE mapper
The structure-agnostic RoE mapper captures bits from one end of a constant bit rate link, packetizes the bits into Ethernet frames, sends the frames across the network, and then recreates the bit stream at the far end of the link. While the constant bit-rate data stream is commonly encoded with the CPRI protocol, it could also be of any other protocol, provided it is within the range of data rates supported by that equipment.
The structure-agnostic RoE mapper has two main modes of operation:

- Tunnelling mode or type 0 works as a simple Ethernet tunnel. It does not remove any line coding bits and does not interpret any special characters (such as K-characters). If the source data is 8b/10b-encoded, the 10-bit symbols present on the line will be tunnelled by this RoE mapper as 10 bits of data. Similarly, 66-bit symbols will be sent for 64b/66b-encoded data as 66 bits of data. The entire stream is simply packetized.

- Line-coding-aware mode or type 1 removes the line coding bits such as for CPRI encoded with 8b/10b or 64b/66b. If the source data is 8b/10b-encoded, the 8-bit symbols present on the line will be tunnelled by this RoE mapper as 8 bits of data. Similarly, if the source data is 64b/66b encoded, the 64-bit symbols present on the line will be tunnelled by this RoE mapper as 64 bits of data. To allow the restoration of the 10-bit or 66-bit symbols at the de-mapper, the RoE mapper/de-mapper must have some awareness of the protocol it is

mapping/de-mapping; the locations of the special characters must be known a priori relative to an event that is indicated in the RoE frame.

Structure-aware RoE mapper

In this mode only the useful information in the CPRI stream is packetized into the RoE frames and different types of data (such as control words and data words) can be encapsulated into separate frames for prioritized processing. Those unused fields within the CPRI stream will be ignored thus bring the benefit of Fronthaul BW reduction. This mode requires full knowledge of the protocol layout of the CPRI and due to the proprietary nature of CPRI will require input from the radio vendor.

7.2.3 Non O-RAN Fronthaul logical transport requirements

The main use case identified by operators for non O-RAN fronthaul is 4G equipment that uses CPRI between the RRH and the BBU (see section 8.4). To support this over a packet based fronthaul network, the CPRI stream needs to be packetized as it ingresses the packet network and the CPRI stream reconstructed as it egresses the packet network. Bandwidth, delay and jitter characteristics in the packet network clearly depend on the technology used to perform this function, which is outside the scope of O-RAN. To provide some guidance on support of this traffic it has been assumed that it is presented as Ethernet and has similar delay and jitter requirements as 7.2x fronhaul traffic.

7.3 Midhaul logical transport requirements

3GPP TS 38.401 [9] defines the de-aggregated RAN, it's characteristics and outlines the F1-U, F1-C and E1 interfaces. Figure 7-3, taken from 3GPP TS 38.401 illustrates the components and interfaces. The Midhaul transport infrastructure is responsible for supporting these interfaces.

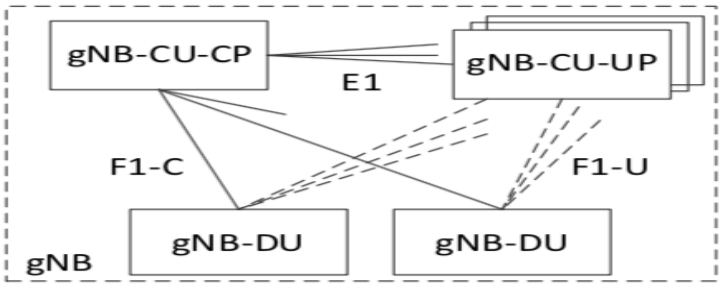


Figure 7-3: Deaggreated gNB

The characteristics of a disaggregated gNB are:

- A gNB may consist of a gNB-CU-CP, multiple gNB-CU-UPs and multiple gNB-DUs
- DUs and CU-UPs are connected to one CU-CP via the E1 interface or the F1-C interface
- DUs can connect to multiple CU-UPs
- Multiple CU-UPs can connect to one CU-CP
- For resiliency reasons, DUs and CU-UPs may connect to multiple CU-CPs

The 3GPP interface associated with O-DU and O-CU communication is the F1 interface. It has a control (F1-C) and data (F1-U) plane component.

Note: W1 interface is the 4G equivalent of the F1 interface. It will not be discussed further in this document as its characteristics are expected to be like the 5G equivalent.

The 3GPP interfaces associated with intra O-CU communications is the E1 interfaces. It runs between the gNB-CU-CP and a gNB-CU-UP. It allows these two components to run as separate entities and potentially in different locations.

7.3.1 Overall Midhaul logical transport requirements

1. Control Plane:
    a. Multi-point at the IP interfaces level (IPv4 or IPv6) between O-CU-CP and multiple O-DUs (F1-C interface).
    b. Multi-point at the IP interfaces level (IPv4 or IPv6) between O-CU-CP and multiple O-CU-UPs (E1 interface).
2. Data Plane: Multi-point at the IP interfaces level (IPv4 or IPv6) between O-CU-UP and multiple O-DUs (F1-U interface).
3. IP connectivity between O-CUs for Xn interface.
4. Some operators may wish to run the user plane interface (F1-U) separately from the control plane interfaces (E1 and F1-C).
5. Some operators may wish to treat Midhaul and Backhaul as a single logical network.
6. Some operators may wish to treat Midhaul and Backhaul as discrete logical networks.

7.4 Backhaul logical transport requirements

Figure 7-4 shows components and the 3GPP interfaces in the mobile Backhaul. It has a control plane and user plane component. It is not uncommon to see the control plane and the data plane divided into separate closed user groups (VPNs) at the transport layer to ensure a clear demarcation between customer user data and the 3GPP control plane.

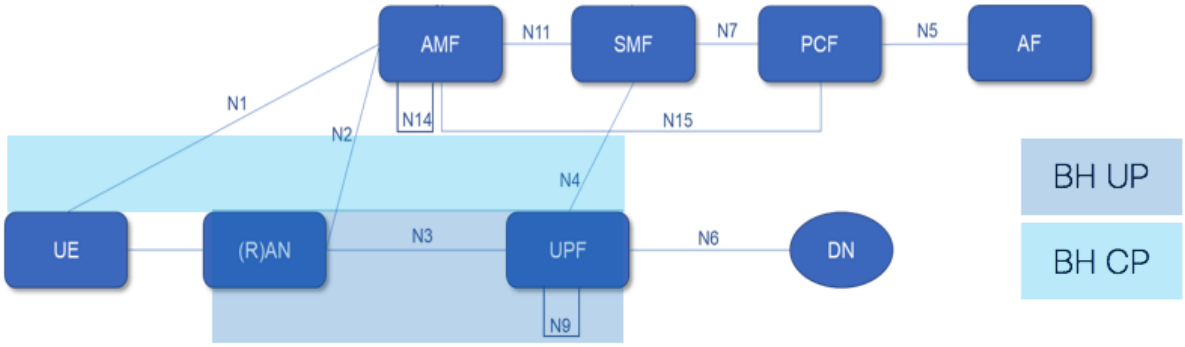


Figure 7-4: 5G Backhaul components and interfaces Source: Adapted from 3GPP TS 23.501 v6.4.0(2020-03): System Architecture for 5G [1] with control plane / user plane shading added by document authors.

The 5G 3GPP interfaces associated with Backhaul are:

- **N1 interface** is a logical control plane interface between the mobile core network and the UE. From a physical perspective it flows via the RAN through the Backhaul infrastructure to the AMF. It is a signalling interface between the UE and the AMF.

- **N2 interface** supports control plane signalling between RAN and 5G core. It is primarily concerned with connection management, UE context and PDU session management, and UE mobility management. In addition, Non-Access Spectrum (NAS) signalling between the UE and the AMF is transported over the N2 connection for that UE. This signalling includes information regarding access control, authentication and authorization, and session management procedures.

- **N4 Interface** is the bridge between the control plane and the user plane of the 5GC. It runs between the SMF and the UPF and is responsible for conveying policy rules regarding policy handling, forwarding and usage reporting to the UPF.

- **N3 interface** is the user plane interface between the O-CU component of the (gNB) and the initial UPF.

- **N9 interface** is a user plane interface than runs between two UPFs. (i.e. an intermediate UPF and the UPF session anchor).

7.4.1 Overall Backhaul logical transport requirements

1. Control Plane: Multi-point at the IP interfaces level (IPv4 or IPv6) between O-CU, UPF and 5GC components (N1, N2, N4, Xn-c).
2. User Plane: Multi-point at the IP interfaces level (IPv4 or IPv6) between O-CU←→UPF (N3), UPF←→UPF (N9) and O-CU←→O-CU (Xn-u).
3. Some operators may wish to run the Backhaul user plane (N3/N9) separate from the Backhaul control plane (N1/N2/N4).
4. Some operators may wish to treat Midhaul and Backhaul as a single logical network.
5. Some operators may wish to treat Midhaul and Backhaul as discrete logical networks.

8 Operator's use cases

Operators have indicated the following use cases are of interest. Any combination of these scenarios may apply in practical deployments though they are described individually. Unless otherwise stated, the eCPRI traffic herein is O-RAN compliant and presents the data according to O-RAN Open Fronthaul CUS plane and management plane specifications [17][18].

Multicast use cases are deferred to section 15.1, after logical network and services have been described, so that they can be provided with better background.

Before describing the individual deployment scenarios, it is worth clarifying the terminology of C-RAN vs. D-RAN. In a 5G Distributed RAN (D-RAN) architecture, the O-CU, O-DU, and O-RU all reside at the cell site (Figure 8-1).

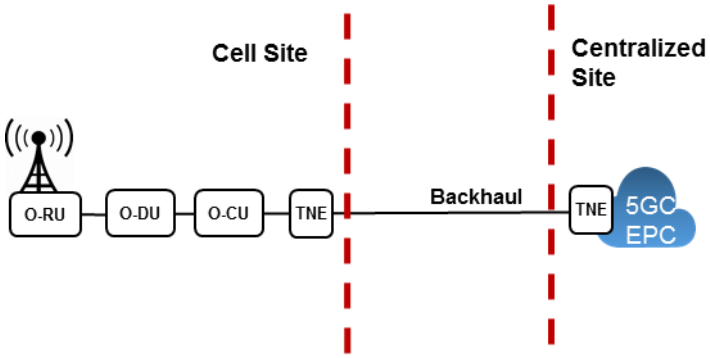


Figure 8-1 5G O-RAN D-RAN architecture

5G Centralized RAN (C-RAN) architectures splits the radio components into discrete components which can be in different locations. In 5G O-RAN model (Figure 8-2), operators may decide to only place the O-RU at the cell site, and centralize the O-DUs or the O-DUs together with O-CU in a central location. The other alternative can be represented by locating the O-RUs and O-DUs at the cell-site and centralizing the O-CUs in a location farther away.

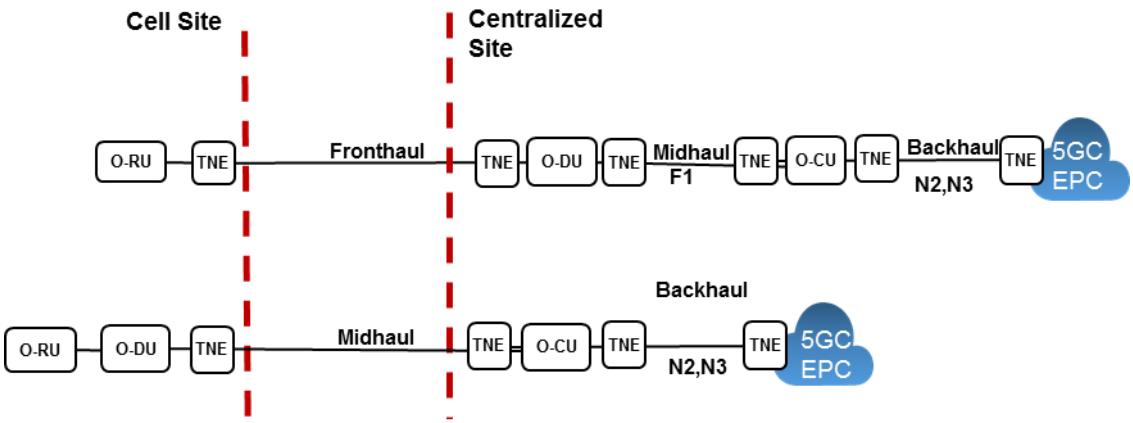


Figure 8-2 5G O-RAN C-RAN architectures

When considering a C-RAN architecture there are both positive and negative impacts. On the positive side, it can increase component efficiency by pooling RAN elements in a centralised location and improve co-ordination between the radio component. On the negative side a C-RAN architecture can significantly increase the bandwidth required if the Fronthaul protocols are traversing the transport network.

Note: The representation of the access and aggregation networks in the use case figures below are not to scale. In scenarios where a Fronthaul component exists, the access network will be geographically constrained due to delay requirements of the Fronthaul protocols. In contrast, the aggregation network in the figures can cover much greater areas due to more relaxed delay requirements of the 5G Midhaul and Backhaul protocols.

8.1 Scenario 1: C-RAN architecture with collocated O-DU and O-CU

As illustrated in Figure 8-3 for this scenario, O-DU and O-CU are collocated at a Hub site, therefore the Midhaul traffic between O-DU and O-CU is local and not going through the transport network.

The eCPRI traffic from the 7.2x split from the O-RU, is transported by a packet switched access transport network to the C-RAN Hub. At the other end of the C-RAN Hub, the aggregation transport network transports the Backhaul traffic to the mobile core.

The O-RU can be a new NR radio as well as a legacy LTE radio, as long as they are made to have the O-RAN compliant Fronthaul interface with 7.2x function split.

The Xn traffic that performs the inter-gNB coordination is considered as part of Midhaul traffic, but it will not reach to the Core. Instead, it is routed to another O-CU, either within the same Hub site or between the Hub sites, via the aggregation transport network.

There is inter-connection between the access and aggregation transport networks for the passing through services such as management traffic. In some cases, the two transport networks can share a same edge router that naturally completes the connection.

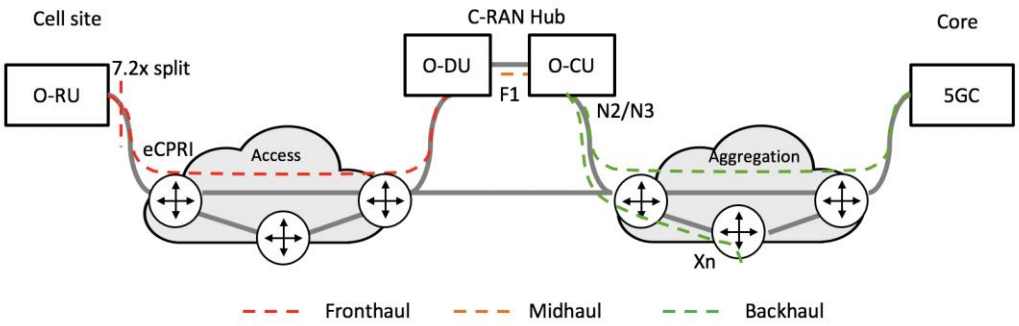


Figure 8-3: Scenario 1 – C-RAN architecture with collocated O-DU and O-CU

8.2 Scenario 2: C-RAN architecture with collocated O-RU and O-DU

In this scenario, both O-RU and O-DU are deployed at the cell site and O-CU is located at the Hub site. By 3GPP standard, the O-DU and O-CU is split by the functional split option 2 and connected with F1 interface, which is transported by the access transport network that connects the cell site and the Hub site. The aggregation transport network carries the N2/N3 Backhaul traffic to the 5G core.

O-RU and O-DU may be collocated or be in close proximity around the cell site. In both cases, O-RU and O-DU communicate to each other by the Fronthaul interfaces via the Cell Site Router (CSR), which is located at the cell site and is considered as part of access transport network, as shown in Figure 8-4

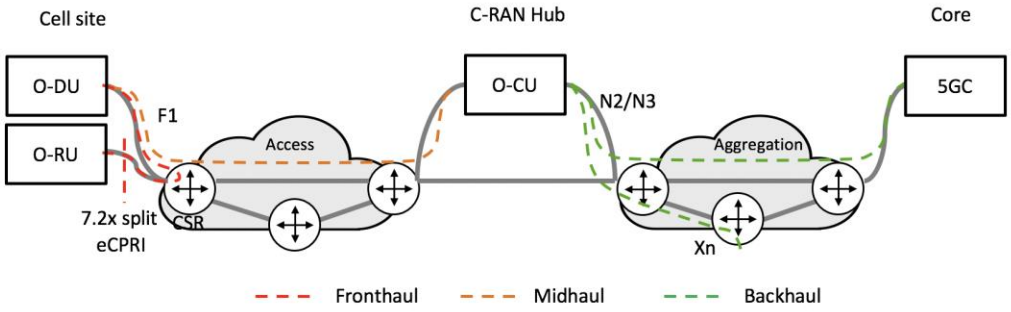


Figure 8-4: Scenario 2 – C-RAN architecture with collocated O-RU and O-DU

NOTE: In this scenario the location of the CSR, O-DU and O-RU relative to each other may be close or in some scenarios more distant. For example, within a stadium environment, the O-DU could be centralised and the O-RUs distributed around the stadium. In all cases consideration needs to be given to the low latency requirements associated with the Fronthaul interfaces.

8.3 Scenario 3: C-RAN architecture with coexistence of legacy Backhaul traffic

This scenario is an extension of Scenario 1, or Scenario 2, where 3G / 4G D-RAN deployments coexist with a 5G C-RAN deployment. The access transport network thus carries the Backhaul traffic from LTE (S1) or 3G UMTS (lu/lub) in addition to the Fronthaul traffic from Scenario 1, or Midhaul traffic from Scenario 2, as illustrated in Figure 8-5 and Figure 8-6 respectively.

The aggregation transport network is responsible for carrying the Backhaul traffic for both new and legacy mobile services.

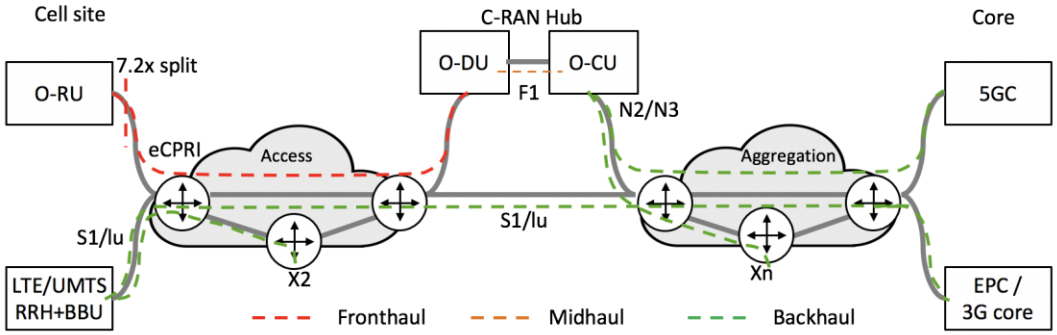


Figure 8-5 Scenario 3a – 7.2x interface combined with legacy Backhaul at access layer

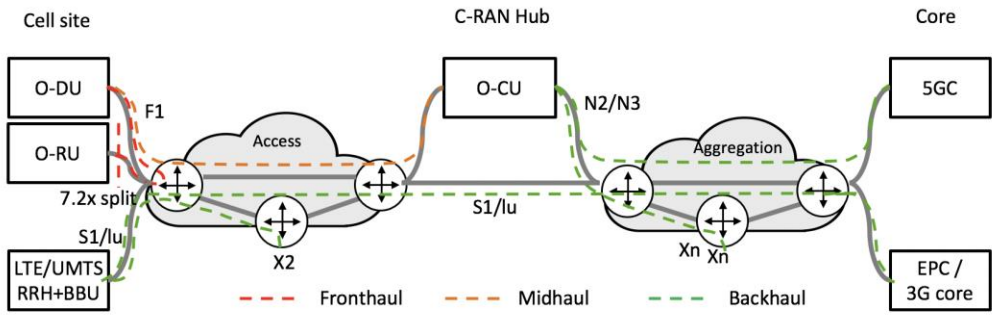


Figure 8-6: Scenario 3b – Midhaul combined legacy Backhaul at access layer

These mixed D-RAN and C-RAN scenarios are important for the brown field deployment.

8.4 Scenario 4: C-RAN architecture with coexistence of legacy Fronthaul traffic

In this scenario, shown in Figure 8-7, both the 4G and 5G networks utilise a C-RAN Fronthaul deployment in the access transport network. In this scenario it is assumed the 4G radio infrastructure consists of RRHs communicating with their serving BBU using CPRI. To migrate these services to a packet based Fronthaul, the native CPRI is converted to packets as it enters the packet infrastructure and converted from packets back to native CPRI as it egresses the packet infrastructure. This scenario assumes the CPRI to packet conversion function is performed by an RoE Mapper using the IEEE 1914.3 standard [35]. The RoE mapper is out of scope of this specification. The impact on the access transport network is it carries 7.2x and RoE fronthaul traffic.

In the Backhaul network, S1 from LTE is transported together with NR Backhaul N2/N3.

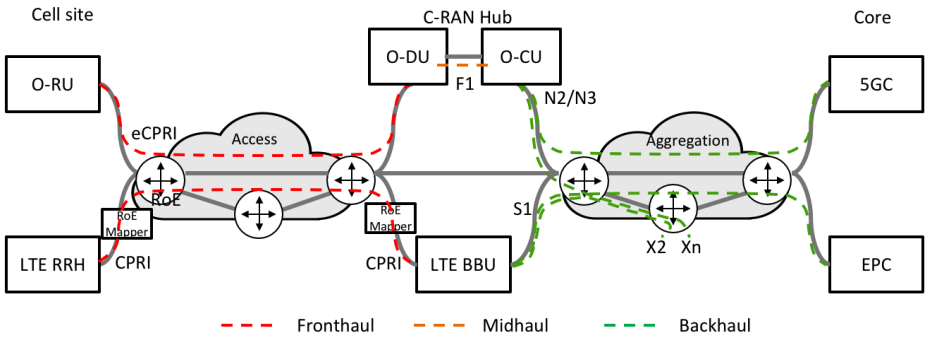


Figure 8-7: Scenario 4 – C-RAN architecture with coexistence of legacy Fronthaul traffic

This scenario is important to support migration plan from 4G to 5G, such as the NSA architecture.

8.5 Scenario 5: C-RAN architecture with further split of O-DU and O-CU

This is a C-RAN architecture with two RAN splits with the O-RUs, O-DUs, and O-CUs hosted at separate locations. The O-DU is placed at C-RAN Hub site closer to the cell sites for shorter latency and the O-CU is more centralized at a different location as shown in Figure 8-8. The Midhaul traffic, identified as F1 from the 3GPP option 2 split, is carried by an addition transport network segment (pre-aggregation transport network) that connects the two hub sites. Similar to other use cases, the Backhaul traffic is transported by the aggregation transport network.

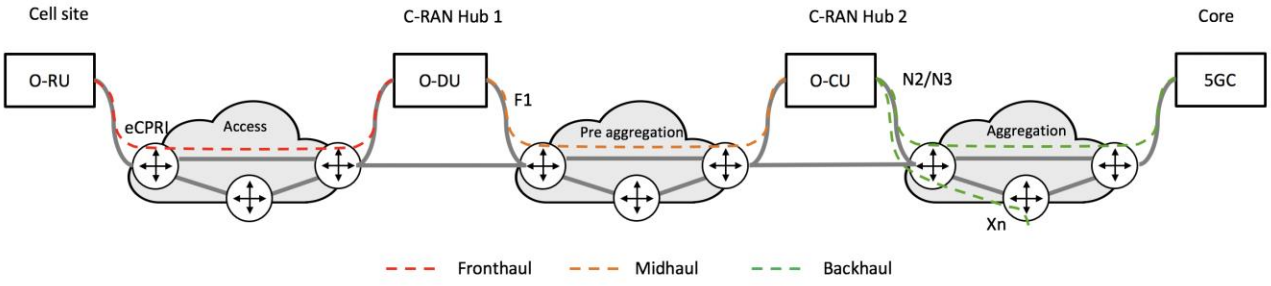


Figure 8-8: Scenario 5 – Dual split C-RAN architecture

8.6 Scenario 6: C-RAN architecture with local breakout

In supporting URLLC to reduce the data plane latency, or fixed wireless application to offload large amounts of user data locally, the User Plane Function (UPF) breakouts before the aggregation network. In this case the aggregation network only carries the control plane traffic N2 **(**Figure 8-9**).**

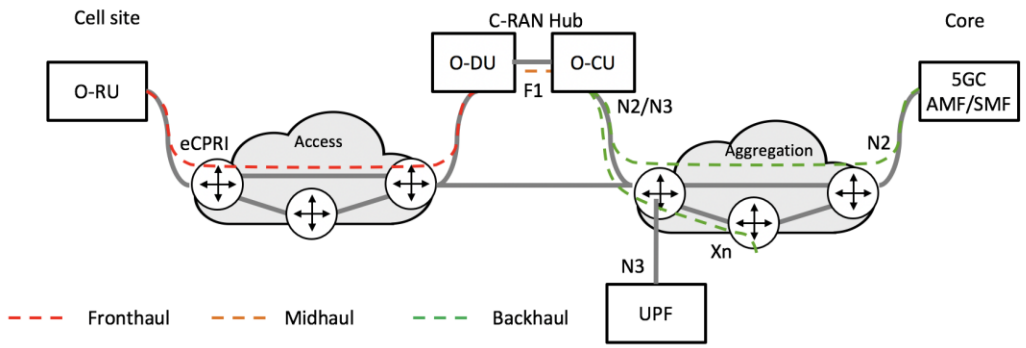


Figure 8-9: Scenario 6 – C-RAN architecture with local user plane breakout

8.7 Scenario 7: Transport slicing

An O-RAN wide initiative to develop end-to-end O-RAN slicing solutions has started, led by WG-1. The architecture and phasing are defined in O-RAN.WG1.Slicing-Architecture-v07.00 [24] and is based on operator use cases and the slicing capability of the O-RAN components. This operator scenario will now track the phasing and slicing use cases developed by WG-1.

Revision 2 of this document now includes Annex F: Transport network slicing solution for WG1 Slicing (informational), which define an example transport solution for slice phase 1 contained in O-RAN.WG1.Slicing-Architecture-v07.00 [24].

From the operator use case given in previous subsection, it is observed that the transport network, especially the access transport network, may experience all types of transport flows simultaneously when operators have the need to engage mixed use cases in their deployment. Possible transport service range widely with:

- Fronthaul, Midhaul, and Backhaul
- NR, legacy LTE and legacy UMTS
- Control plan, User plane, and Management plane
- To support transport operation for different operators
- To support different types of end-to-end services or applications (such as URLLC and eMBB)

Each of them may be multiplexed into a commonly shared transport network with largely different transport requirements, which include latency, throughput, transmission reliability.

To reduce complexity and manage network resource more efficiently, these transport flows can be classified into transport slices according to common service requirements. One example is shown in Table 1.

Transport Slices	Description	Transport flows	Transport BW	Transport timing sensitivity	Transport reliability

TS 1	Fronthaul	7.2x CUS-plane, RoE	High	High	High
TS 2_1	Data plane for Backhaul of URLLC service of Operator A	F1-U, S1-U, N3, X2/Xn-U	Medium	High	High
TS 2_2	Data plane for Midhaul, Backhaul of Operator B	F1-U, S1-U, N3, X2/Xn-U	Medium	Medium	High
TS 3	Control plane for Midhaul, Backhaul, Management plane	7.2x M-Plane, F1-C, S1-C, N2, X2/Xn-C, Management	Low	Low	Low

Table 1 Transport slicing example

The above transport slice may further split into more sub-slices if necessary, for different end-to-end user applications or different operators depending on separate needs of the priority, latency, or bandwidth.

These transport slices are designed to meet the objectives:

- Provide the transport elements to support network slicing. As part of network sub-network instances, the transport network interfaces with other network segments by the transport slices.
- Allow flexible configuration to efficiently support adding/deleting/reconfiguring slices and services
- Support protection/isolation/prioritization mechanisms to minimize inter-slice effects
- Support monitoring/reporting the slice KPIs

With reference to the network slicing use cases described in WG1 network slicing specification document [24], Figure 8-10 presents a transport realization example in supporting the phase 1 scope summarized in *Annex F: Transport network slicing solution for WG1 Slicing (informational)*. This is based on the dual split transport architecture (scenario 5) combined with scenario 3b described in previous sections. The purpose of this example is to illustrate a practical scenario of traffic distribution in a slicing transport network.

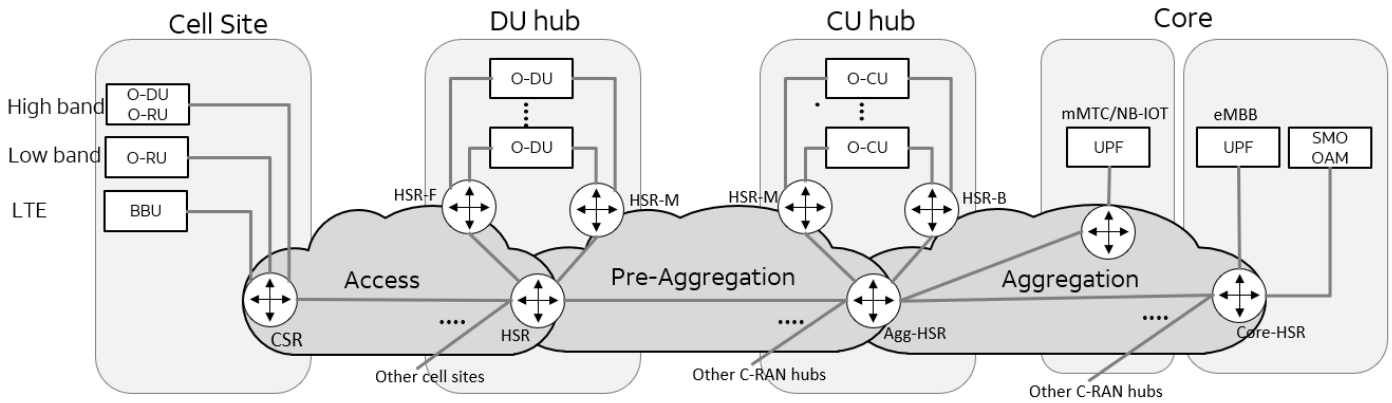


Figure 8-10: Transport based on dual split architecture

The TNEs in the figure are defined as follows:

- CSR: Cell Site Router, aggregating traffics from collocated O-RUs
- HSR: Hub Site Router, aggregating traffics from multiple cell sites
- HSR-F: Hub Site Router that distribute fronthaul traffics to O-DUs
- HSR-B/M: Hub Site Router that aggregate Backhaul or Midhaul traffics
- Agg-HSR: Hub Site Router at centralized O-CU site that aggregate the traffic from multiple O-DU hub sites
- Core-HSR: Hub Site Router at Core site that aggregate traffic from multiple O-CU hub sites

The logical link across the Access transport network between cell site and C-RAN hub is seen with complex traffic types: fronthaul, midhaul, and backhaul are mixed. Specifically, following traffic flows are jointly transported cross the transport:

UP: *7.2x C/U-P, S1-U, F1-U, X2/Xn-U*
CP: *S1-C, F1-C, X2/Xn-C*
MP: *7.2x M-P, O1, E2*

At the Pre-aggregation transport network, there is no presence of the fronthaul anymore. Those traffics are merged and carried by the network

UP: *S1-U, F1-U, X2/Xn-U*
CP: *S1-C, F1-C, X2/Xn-C*
MP: *O1, E2*

At the aggregation transport network, only backhaul traffic are distributed to UPFs that may be geographically separated for eMBB and mMTC/NB-IOT:

UP: *N3, N9, S1-U*
CP: *N2, S1-C*
MP: *O1, E2*

The transport solutions provided in this specification, as described in section 18 and *Annex F: Transport network slicing solution for WG1 Slicing (informational)* are expected to provide adequate toolbox to achieve prioritization, isolation, QoS, and configuration objectives in managing the above transport scenarios.

9 Overall packet switched Open Xhaul architecture

There are different ways packet switching could be deployed to support an Open Xhaul architecture. Factors include:

- Span of the packet switching components. Packet switching could be deployed from cell site to the transport core or mixed with other technologies, for example WDM in access, to form the end-to-end network.
- The nature of the underlying L0/L1transport used by the packet switching equipment.
- The network protocols used at the packet switching layer.
- How Xhaul services are built on the Xhaul infrastructure.

9.1 Physical layout and xHaul transport options

A simple model illustrated in Figure 9-1 relies on an end-to-end packet-based transport architecture that builds on a physical network. In this model, Fronthaul, Midhaul and Backhaul deploy packet-based transport technologies. The diagram represents a logical view of the network. The physical network implementation might be different. For example, an O-DU might be connected to a single TNE port but deliver two logical connections to a second TNE in Fronthaul network, and a third TNE in Midhaul network.

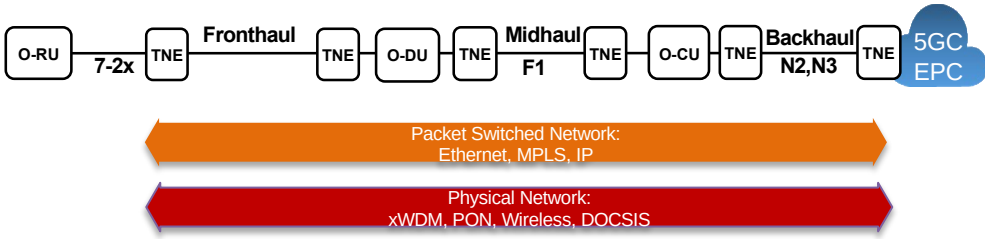


Figure 9-1 Packet based architecture in front-, mid- and Backhaul networks

Unlike the previous architecture some operators may decide to use the packet-based technology only in Mid- and Backhaul, and use simple physical networking to connect the O-DU ports with O-RUs (Figure 9-2). In this case the physical network between the O-RUs and O-DUs could, in the simplest case could be dark fiber links. The next choice in terms of simplicity can be a passive WDM system.

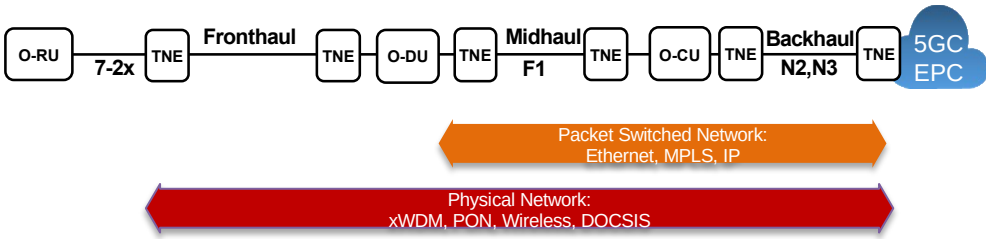


Figure 9-2: Packet based architecture in mid and Backhaul networks

Another architectural variant (Figure 9-3) can be designed by inserting TDM based layer 1 transport technologies between the packet-based and physical networks. OTN and SPN/G.mtn are two choices that allow for aggregation and transparent transport of larger traffic volumes, and design of hard slicing architectures. In this architecture, the OTN or SPN/G.mtn infrastructure must present

Ethernet clients at UNI-C/UNI-N interfaces to the packet switched network. The clients need to be transparent and operate at full Ethernet line rates, because the packet switched network elements are responsible for QoS and need to rely on client signals with proper bandwidth.

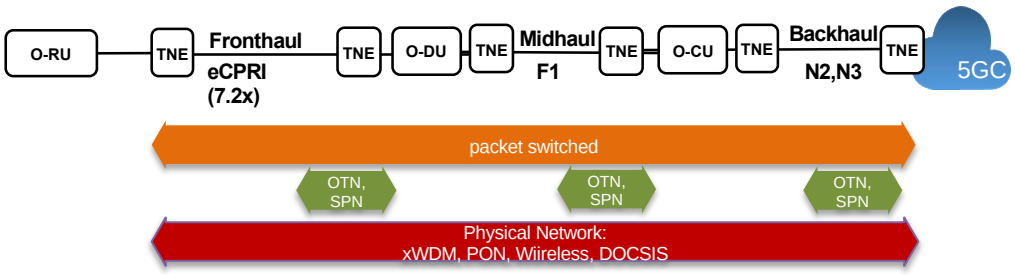


Figure 9-3 Use of TDM based technologies with Ethernet presentation in Xhaul networks

In cases where there are underlying L0/L1 transport solutions, the synchronization and timing flows must not be impaired by these layers, as any impairment will impact the synchronization function of the packet switched network, and the overall synchronization performance of the end-to-end system.

9.2 Open Xhaul architecture in revision 1 and 2

Revision 1,2 and 3 of this document describes a packet switched transport architecture illustrated in Figure 9-4.

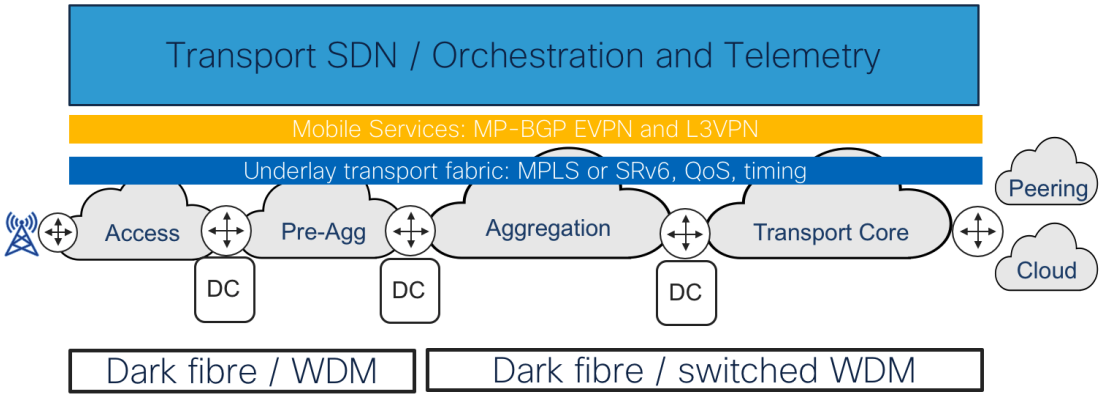


Figure 9-4 Packet switched transport for mobile Xhaul

It is a converged end to end packet switched infrastructure, beginning at the cell site, located in the edge of the access layer and stretching to the core of the transport layer. The packet switching TNEs are QoS enabled, high capacity, low latency devices interconnected by point-to-point Ethernet interfaces running at the full capacity of the Ethernet interface, typically using either point to point fibres or via a WDM infrastructure. It incorporates data centers suitably placed across the transport network infrastructure to support virtual and physical NFs associated with mobile and fixed services but also potentially the placement of “Application Functions” associated with value-add services and customer specific application.

The logical architecture is based on a common underlay packet switching infrastructure based on either MPLS or SRv6 overlaid with a L2 / L3 service infrastructure (VPNs) that uses the capabilities of the underlay packet switched network to support the mobiles interfaces.
The underlay packet switching infrastructure provides basic network services such as; any-to-any connectivity between TNEs, scaling, fast convergence, shortest path and traffic engineered forwarding, packet-based Quality of Service (QoS) and timing.

The service layer supports native Ethernet services, using EVPN technology and IP VPN services, using MP-BGP based L3VPNs. These services can utilise the facilities offered by the underlay packet switched infrastructure to support the different mobile interfaces in an appropriate fashion. Where possible transport services are built on an end-to-end basis without intermediate stitching / switching points within the transport infrastructure. This approach has been taken to minimise the transport service orchestration overhead.

9.3 Technology and architectural choices

As discussed in the introduction to this section there are other packet-switched solutions and potentially many design approaches available that may have the same capabilities of the architecture described in this revision of the document. Future versions of this document may describe alternative technologies and designs based on operator requirements and suitably mature standards.

9.4 Standardization

The underlay packet-switching technologies described in this document are based on MPLS and IPv6 with an emphasis on the Segment Routing (SR) control and data planes. The overlay service layer uses EVPN and MP-BGP L3VPNs. The IETF is the primary standardization body for these technologies. The key specification of the IETF is the “Request For Comments” (RFC). RFCs are labelled with different statuses: Internet Standard, Proposed Standard, Best Current Practice, Experimental, Informational, and Historic. RFCs usually begin as Internet-Drafts written by Individuals, or a small group and are known as individual drafts. Depending on interest an individual draft can be adopted by an IETF Work Group (WG) where it is improved and revised by a wider group and can ultimately become an RFC. This process is well defined by the IETF, with guides and considerations for the progression of a document. If the process completes, then the document is published as an RFC by the IETF RFC Editor. If a document does not follow the process or doesn’t meet the guidelines outlined, then the drafts will expire and never become an RFC. Depending on subject matter the process of moving from an WG adopted draft to an RFC can several years.

To present architectures suitable for 5G, based on modern packet-switching technologies, Standards, proposed standards, best current practise and informational RFCs and current workgroup adopted drafts with the same designations can be referenced in this document. All experimental and historic RFCs and workgroup adopted drafts and all forms of personal drafts MUST NOT be referenced. When writing requirements around these different types of documents consideration should be given to document type and their status.

Future revisions of this document MUST be updated to reflect the status of IETF drafts which are referenced. If they expire, they MUST be removed and if they complete the IETF standardization

process, references and requirements MUST be updated to remove the IETF draft and include the appropriate RFC number.

9.5 Document organization

The remainder of this document describes the physical and logical architecture of a packet switching Xhaul architecture and how it can support 5G and legacy mobile services. It is arranged as follows:

Section 10 describes the end-to-end physical network infrastructure.

Sections 11 and 12 describes two underlay packet switching architectures capable of supporting a 5G mobile environment. The first is based on MPLS, the second IPv6 with Segment Routing (SRv6). Depending on scale and preference, operators will need to select one or the other.

Section 13 describes how IP and Ethernet services suitable for 5G are delivered on a packet switched Xhaul network. The approach is common regardless of the underlay technology deployed.

Section 14 describes the QoS architecture for a packet switched Xhaul network.

Section 15 describes multicast and consideration for deployment.

Section 16 covers packet-switched orchestration and telemetry. Currently it is empty and will be completed in a later revision of this document or in a separate document.

Section 17 covers TNE and transport network security.

Section 18 and 19 describes how the packet switched Xhaul architecture supports 5G and legacy mobile use cases outlined in section 8.

10 Physical network design for packet switched Xhaul

Figure 10-1 shows two of the most common physical topologies seen in transport networks between the access and the core of the transport network. In the first, there are four layers of transport infrastructure; access, pre-aggregation, aggregation, and transport core. In the second, there are three layers of transport infrastructure; access, aggregation, and transport core. Packet switches are used within each layer and provides connectivity between the layers.

The way the mobile RAN and core infrastructure is arranged over these different segments varies based on geography and the MNO's RAN and mobile core designs and individual operator's classification of different components in the physical network. It should be noted that in most cases the transport core doesn't form part of the 5G RAN infrastructure, which is restricted to the access and aggregation infrastructure (often called the metro network), but it does play an important role in scaling and connecting metro infrastructures together to enable end to end services to be built.

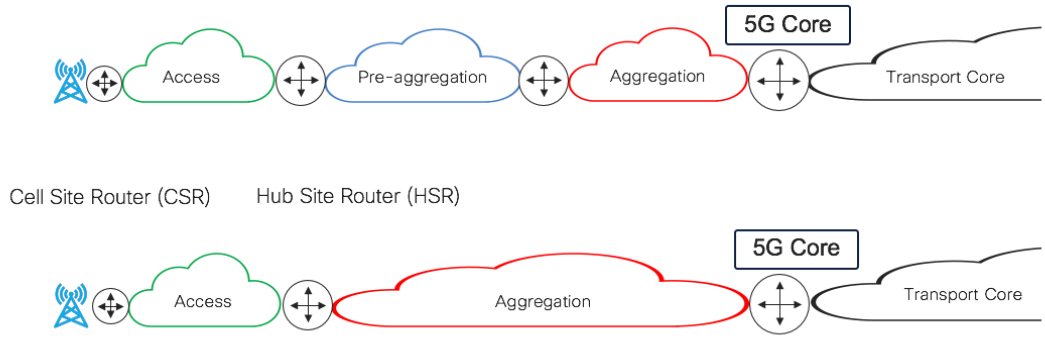


Figure 10-1: Physical transport components

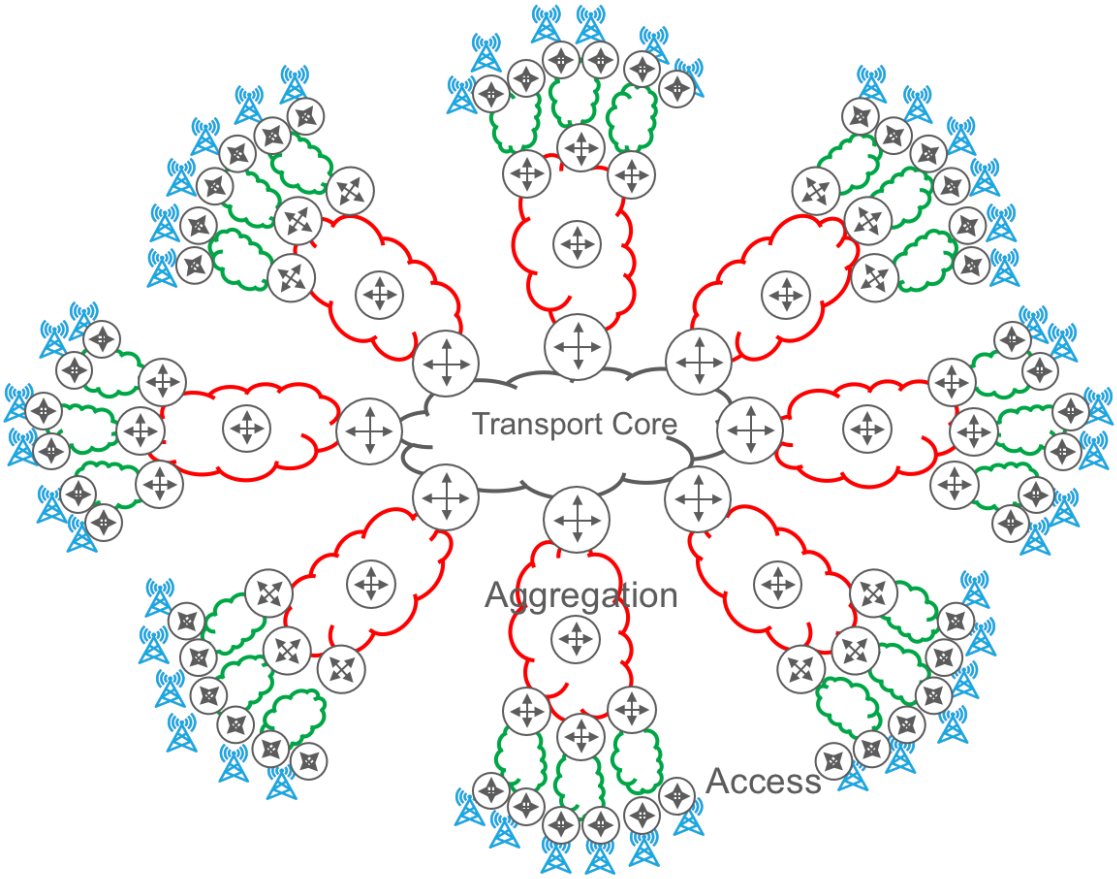


Figure 10-2 Hierarchical designed transport network consisting of access / aggregation and transport core.

Figure 10-2 shows an example of a logical transport network consisting of a number of access, aggregation domains connected hierarchically to form a single transport core domain. From a logical architecture perspective, the entire infrastructure needs to be considered as a whole, but at the physical layer there are significant differences in these discrete physical segments. These include:

1. Distances between sites
2. Layout of the physical media
3. Technology employed
4. Environmental conditions
5. Bandwidth requirements
6. Cost structures surrounding equipment and real estate.

10.1 Packet over fibre

This document concentrated primarily on a transport infrastructure built using packet over fibre solutions in all physical segments of the network.

10.1.1 Access

A transport network element (TNE) – Ethernet switch or IP router – resides in the cell site aggregating traffic received on access ports (downlinks) from end-hosts (O-RUs, or O-DUs, as outlined in Section 8, as well as other types of end-hosts, like for example 4G BBUs, business CPEs, etc. in hybrid deployment models), and statistically multiplexing this traffic onto a higher capacity Ethernet uplink, typically established over a dark fibre or WDM lambda, or any other underlying technology (for example microwave radio link). Link speed will depend on the traffic levels required by the services running in the access network. The anticipated ranges will be from 1…50 Gbps (downlink) up to bundles of 100 Gbps (uplink).

10.1.1.1 Topology

The physical topology employed in the transport access network is operator dependent and driven mainly by the three aspects:

- fibre topology, its availability
- the traffic matrix
- latency and time error requirements

The main physical architectures anticipated in the transport access are:

- ring
- chain
- hub and spoke (called as well spine and leaf), with redundant CSR connectivity
- hub and spoke (called as well spine and leaf), without redundant CSR connectivity

as shown in Figure 10-3, Figure 10-4, Figure 10-5 and Figure 10-6.

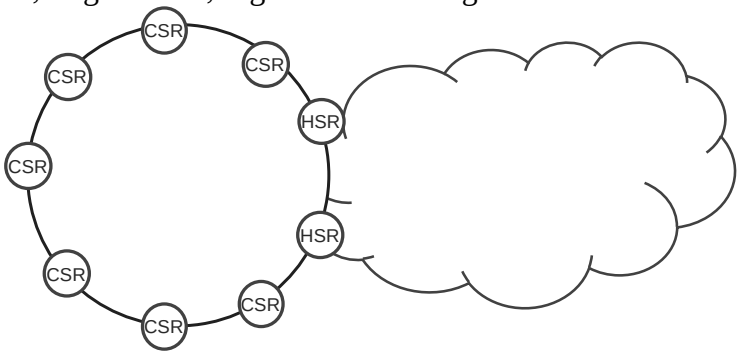


Figure 10-3: Ring based access physical topology

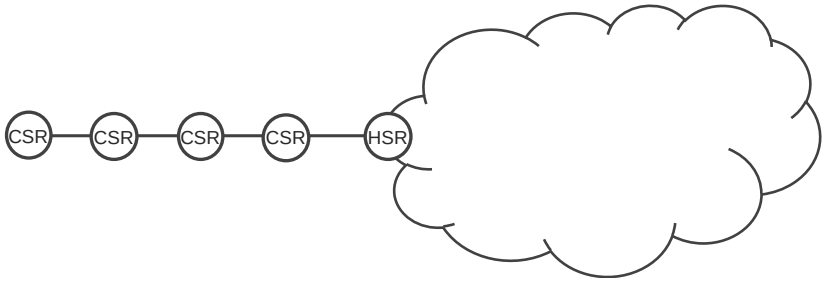


Figure 10-4 Chain based access physical topology

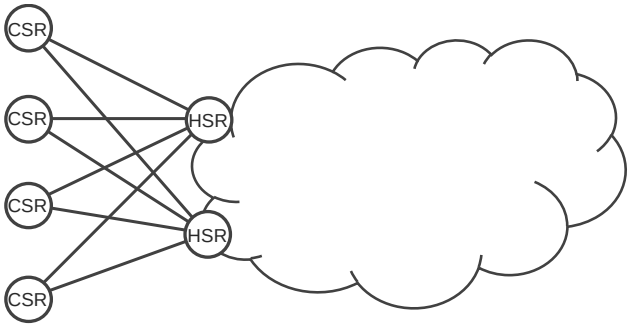


Figure 10-5 Redundant hub and spoke access physical topology (also often called spine and leaf)

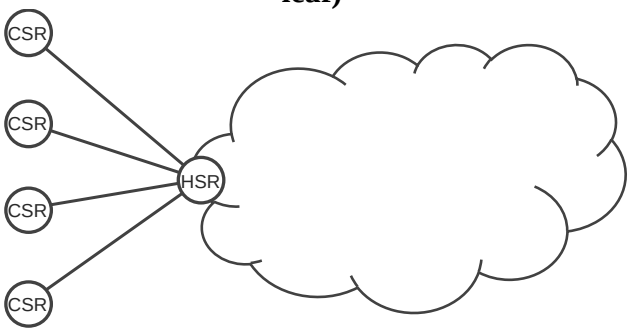


Figure 10-6 Non-redundant hub and spoke access physical topology (also often call spine and leaf)

10.1.1.2 Suitability to support O-RAN Fronthaul/Midhaul/Backhaul

As discussed in detail in Section 8, depending on the use case, O-DU function can be placed at the cell site, together with O-RU function, or placed away from the O-RU, at the hub site. The O-DU placement has big influence on the possible physical topology for the access domain of the transport network.

Open Fronthaul Interface between O-RU and O-DU has very strict latency and timing budgets (O-RAN.WG9.XTRP-REQ-v01.00 “Xhaul Transport Requirements v01.00”, November 2020 [19]). The implication of these requirements on the transport network design is, the number of Transport Network Elements (TNEs) between O-RU and O-DU should be minimized, as each TNE increases the latency and time error.

While each of the depicted access topology could be used for Fronthaul, topologies with limited number of transit transport network elements – hub-and-spoke or spine-and-leaf – leave more latency budget for delay caused by light propagation in the fiber, allowing for extending Fronthaul

over larger distances, comparing to the topologies with bigger number of transit transport network elements (ring or chain topologies).

Ring or chain topologies can be used for Fronthaul, providing that latency and timing error budgets required for particular Fronthaul deployment are maintained. For example, designing Fronthaul for standard NR performance (O-RAN.WG9.XTRP-REQ-v01.00 "Xhaul Transport Requirements v01.00", November 2020 [19] Table 3), mandates maximum 100 μs one-way latency between O-RU and O-DU. This latency budget is consumed by the fiber (~4.9 μs/km), as well as by transit transport network elements (~1-20 μs per node, depending on hardware capabilities and port speeds). Assuming 10 μs latency per transport network element, a chain or ring with 10 transport network elements would completely consume entire Fronthaul latency budget (ring might break, so any Fronthaul design must be prepared for link failures not only from traffic rerouting perspective, but as well must take into account increased number of transit network elements, leading to increased latency, under ring failure condition).

Shorter chain or ring, with for example 7 transport network elements, leaves only 30 μ (6 km) latency budget for fiber.

Note: the maximum number of TNEs can be limited due to overall synchronization requirements and the consideration of time error contributions of integrated T-BC/T-TC. Examples are given in CUS specification Annex H, and will be provided in WG9 Timing and Synchronization architecture solution document [21].

For Midhaul/Backhaul interfaces, there are not so strict requirements regarding latency budgets, so any access topology (ring, chain or hub-and-spoke/spine-and-leaf) can be used without special considerations.

Note: Hub and spoke architectures are often super-imposed over a physical ring topology using WDM technology. To minimize the Fronthaul latency, direct (dark) fibre should be used in hub and spoke (spine and leaf) access physical architecture carrying Fronthaul traffic. Otherwise (ring topology), detailed latency analysis must be performed to confirm that latency stays within mandated Fronthaul budget.

10.1.1.2.1 Time Sensitive Networking (TSN) for Fronthaul

IEEE 802.1CM [30] defines two Time Sensitive Networking (TSN) profiles, with main characteristics summarized in Table 2.

Characteristic	TSN Profile A	TSN Profile B
Max frame size for Fronthaul data	2,000 octets	2,000 octets
Max frame size for non-Fronthaul data	2,000 octets	no limit
Differentiated prioritization for Fronthaul data	yes	yes
Frame pre-emption of non-Fronthaul data	no	yes

Table 2 TSN Profile A and Profile B comparison

The major differences between TSN Profile A and TSN Profile B are therefore the maximum frame size of non-Fronthaul traffic, as well frame pre-emption, as pictured in Figure 10-7**.**

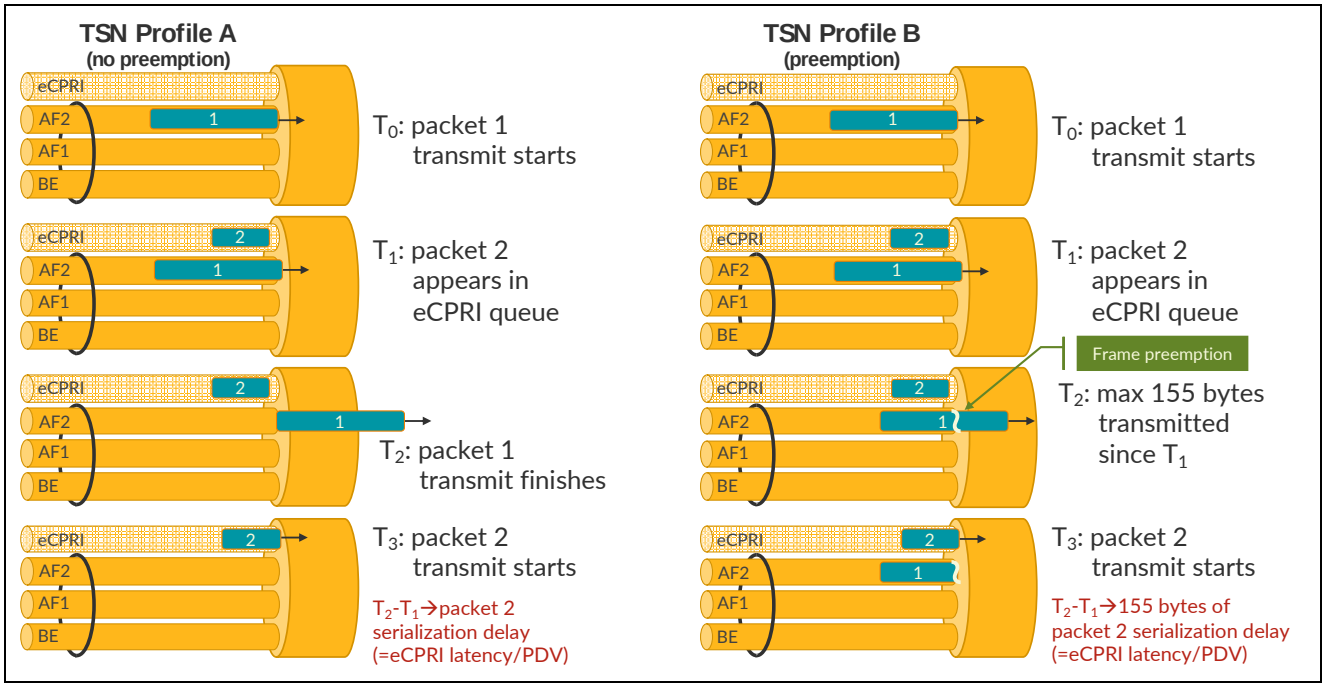


Figure 10-7 TSN Profile A and TSN Profile B operations

IEEE 802.1CM has detailed discussion about transport network element latency calculation (Section 7.2), as well as example end-to-end delay calculations (Annex B), therefore these calculations are not repeated in this document. This document, however, focuses on delay differences of Fronthaul traffic introduced by transport network element operating in accordance with TSN Profile A or TSN Profile B.

Following delay components contribute to overall transport network element delay:

- **Frame transmission delay**, which is the time taken to transmit the frame at the transmission rate of the port. Since both TSN Profile A and TSN Profile B mandate for Fronthaul data maximum frame size of 2000 bytes, this delay component is equal for Fronthaul data frames with both TSN Profile A and TSN Profile B.
- **Self-queueing delay**, which is the delay caused by other frames in the same traffic class as frame to be sent (i.e. both frames are for example Fronthaul data frames). The part of the self-queuing delay caused by frames that arrive at more or less the same time from different input ports is referred to as fan-in delay; however, it is simpler to handle fan-in delay as part of the self-queuing delay. Since both TSN Profile A and TSN Profile B mandate that Fronthaul data frames cannot be pre-empted, this delay component is equal for Fronthaul data frames with both TSN Profile A and TSN Profile B.
- **Queuing delay**, which is the delay caused by the frame of which transmission already started in an arbitrarily small time before frame X became eligible for transmission, plus the delay caused by queued-up frames from all flows with higher priority than the traffic class of frame X. Since both TSN Profile A and TSN Profile B mandate that Fronthaul data must be prioritized over non-Fronthaul data, in both profiles only single non-Fronthaul frame can contribute to this delay. In case of TSN Profile A, a non-Fronthaul frame size is limited to 2,000 bytes (2,020 bytes including preamble, start of frame delimiter and inter-frame gap). In case of TSN Profile B, a non-Fronthaul frame can be pre-empted, but at least 155 bytes of pre-empted frame are transmitted due to the characteristics of frame preemption.

- **Store-and-forward delay**, which includes all other elements of the forwarding delay that are a consequence of the internal processing of the transport network element, including the time to select the input for transmission to the egress port, assuming that the input queue under consideration and output queues are empty. This delay factor is highly hardware dependent, but is equal for both TSN Profile A and TSN Profile B on given hardware.

Summarizing, the only differentiated delay factor between TSN Profile A and TSN Profile B is queueing delay of non-Fronthaul data (queueing delay of 2,020 octets in TSN Profile A, versus queuing delay of 155 octets in TSN Profile B).

Table 3 summarizes this queuing delay for various port speeds.

Port speed	TSN Profile A Queuing delay of 2,020 bytes	TSN Profile B Queuing delay of 155 bytes	Difference	
1 Gbps	16.160 µs	1.240 µs	14.920 µs	3,045 m
10 Gbps	1.616 µs	0.124 µs	1.492 µs	304 m
25 Gbps	0.646 µs	0.050 µs	0.596 µs	122 m
50 Gbps	0.323 µs	0.025 µs	0.298 µs	61 m
100 Gbps	0.162 µs	0.012 µs	0.149 µs	30 m
200 Gbps	0.081 µs	0.006 µs	0.075 µs	15 m
400 Gbps	0.040 µs	0.003 µs	0.037 µs	8 m

Table 3 Queuing delay of non-Fronthaul data

Significant difference between TSN Profile A and TSN Profile B can be observed over port with 1 Gbps speed (14.92 µs, which is equivalent to ~3 km fiber). For port speeds 10 Gbps or above, the latency difference introduced to Fronthaul traffic (1.494 µs, which is equivalent to 305 m fiber, or less) is not so significant for designs targeting standard NR performance (with 100 µs overall latency budget).

10.1.2 Pre-aggregation / Aggregation / Core transport

Pre-aggregation and aggregation and the core transport infrastructure can support Midhaul and Backhaul requirements. Due to the delay requirements associated with Fronthaul, it is unlikely O-RAN 7.2x or RoE will be seen in these portions of the network. Additionally, in most scenarios the transport core will not be part of the Midhaul or Backhaul environment but may provide N6 connectivity to any of the following:

1. Peering points for mobile consumer services.
2. Enterprise locations for mobile enterprise solutions.
3. Inter region voice communications within the operator.
4. Other operators.

10.1.2.1 Physical media and topologies

The assumption made in this architecture is the connectivity between routers in the pre-aggregation, aggregation and transport core is provided by point-to-point Ethernet connections where the full capacity of the Ethernet interface is available to the router. Link speed will depend on the traffic levels required by the services running and the position in the network. The anticipated ranges will

be from 10Gbps up to bundles of 400Gbps with the capacity provided over dark fibre or some form of WDM solution.

The physical topology employed in the pre-aggregation, aggregation and transport core are operator dependent and driven by fibre topology, its availability and the traffic matrix. The main physical architectures seen in these segments of the network are all redundant and include ring, “hub and spoke” and mesh. These are shown in Figure 10-8.

Note: This document uses the term “hub and spoke architecture” but it is used inter-changeably with leaf / spine and fabric architectures.

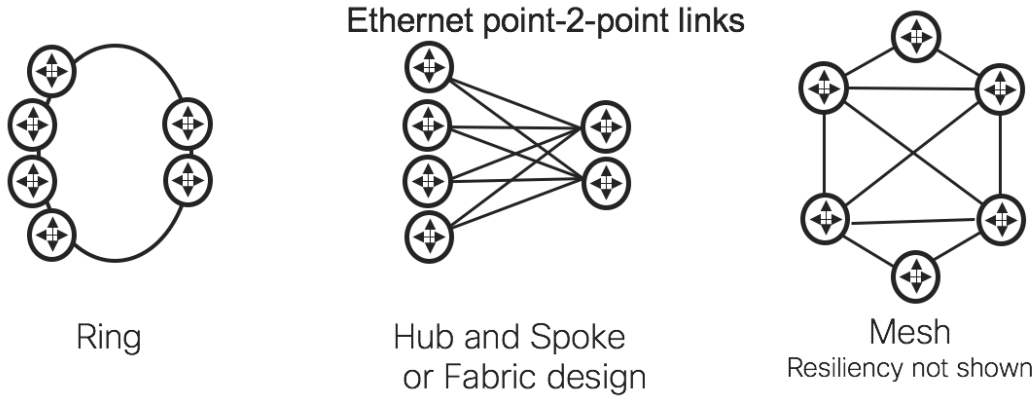


Figure 10-8 Physical topologies anticipated in Xhaul pre-aggregation, aggregation and core

1. Ring: This architecture is commonly seen in pre-aggregation and aggregation networks where traffic is being collected and aggregated from the access towards components in the aggregation or transport core. Connectivity between routers is either via direct dark fibre connections or via lambdas derived from a WDM system.

2. Dual hub and spoke: This architecture is most commonly seen in aggregation and core transport networks but also sometimes in pre-aggregation networks. It is used to collect and aggregate traffic towards common aggregation points. Connectivity between routers is typically via direct dark fibre or via lambdas derived from a WDM system.

   Note: Hub and spoke architectures are often super-imposed over a physical ring topology using WDM technology.

3. Mesh: This is architecture is normally seen in transport core networks but can be seen in aggregation networks. It is used when traffic is flowing in an any-to-any fashion. As drive distances increase and connectivity becomes meshier in nature, so there is increasing use of DWDM and ROADM technology to derive the links between routers.

10.2 Alternative physical transport solutions

This document concentrates packet switching over a fibre infrastructure. However, there are other physical solutions that can be employed in Xhaul networks. This section outlines some of those technologies and discusses their applicability in different Xhaul roles.

10.2.1 WDM in access network

The access network uses optical technology with no packet switching transport functionality. Figure 10-9 illustrates a passive WDM Fronthaul design but the same design principle applies regardless of the WDM solution. Each O-RU directly connected to the WDM multiplexer/demultiplexer by a colored optic and an optical cable. At the O-DU side in the central office, the WDM multiplexer/demultiplexer performs wavelength multiplexing/demultiplexing, which realize the one-to-one optical wavelength connection. The WDM solution can be passive, semi active or active.

Figure 10-9 WDM Fronthaul design (passive)

WG-9 has a work item underway considering WDM-based Fronthaul transport [20], consequently this document does not consider this type of access network design further.

10.2.2 Passive Optical Networks (PONs)

Passive Optical Networks (PONs) are fiber-based point-multipoint access networks, relying on TDM/TDMA for the exchange of traffic between a central Optical Line Termination (OLT) and multiple Optical Network Units (ONUs). PONs also use WDM for separation of up- and downstream traffic, and for overlaying of multiple different PON technologies over the same fiber infrastructure called Optical Distribution Network (ODN). Note that there are two basic flavours of PON;

1. Time or Time +Wavelength Division Multiplexing (TDM or TWDM) PONs apply point-multipoint connections (one up/down wavelength pair (called Channel Pair) is shared over multiple ONUs) over a point-multipoint ODN.
2. Wavelength Division Multiplexing (WDM) PONs apply point-point connections between each ONU and the OLT (one Channel Pair per ONU) over a point-multipoint ODN.

WDM PONs are like dedicated point-point links whereas TDM/TWDM PONs have some fundamental characteristics to bear in mind when considering them for mobile Xhaul, described below.

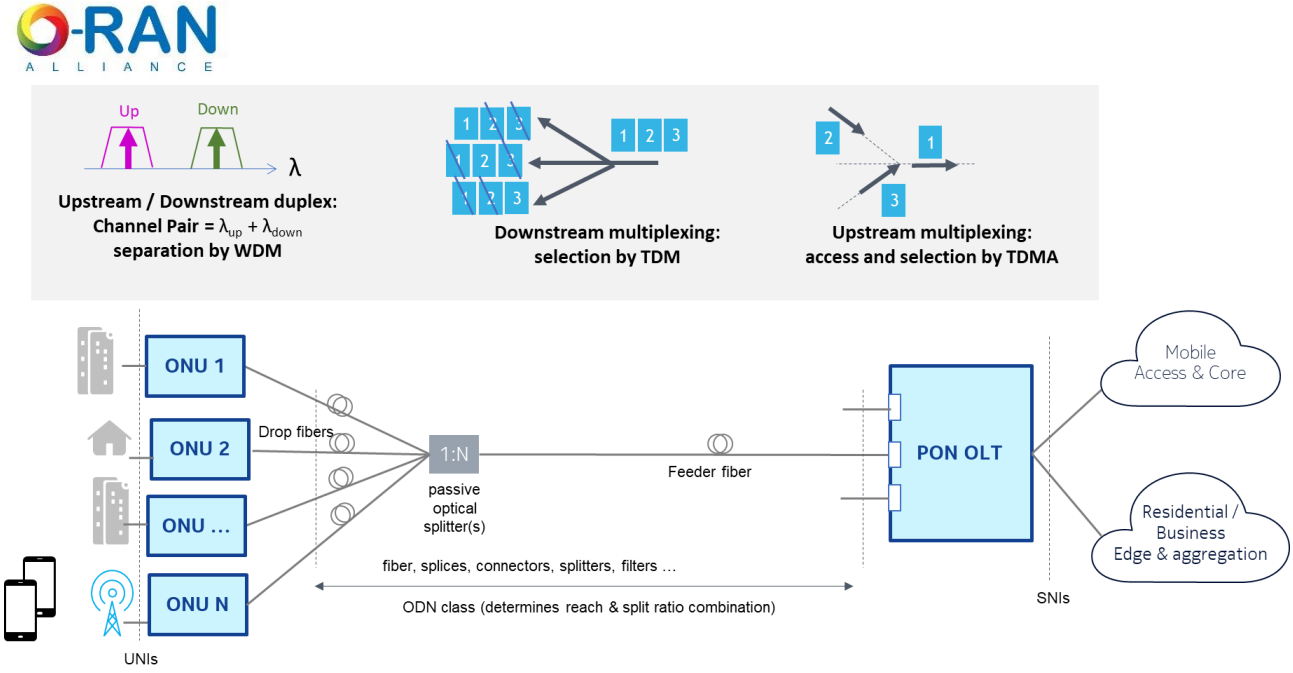


Figure 10-10: Typical TDM PON network. In TWDM PON multiple Channel Pairs are muxed per ODN

TDM/TWDM PONs are standardized technologies (see [182], [32], [33], [34]) and are characterized by generic capabilities (independently of the use case):

- PON line rates are shared over the ONUs and existing technologies reach 10 Gbit/s per channel (up to 4x 10 Gbit/s with NG-PON2), with new variants being developed at 25 Gbit/s and 50 Gbit/s. Note that some fraction is consumed by the technology-specific overheads (e.g. by parity bits for Forward Error Correction).
- Depending on the technology the maximum distance can reach 20 km or 40 km (or 60 km with a reach extender), the maximum amount of ONUs per PON port can reach 64 or 128. Note that the reach (distance) and scalability (amount of ONUs per PON ODN) are a mutual trade-off bound by the optical budget (many classes of optical budget exist, e.g. Class N1 covers optical losses from 14dB to 29dB).
- The max amount of traffic flows per PON port can reach 4k (independently of physical layer considerations).
- The upstream TDMA multiplexing is controlled by a Dynamic Bandwidth Assignment (DBA) algorithm in the OLT (which is equivalent to upstream shaping over the PON segment). The DBA allocates the bandwidth needs per transport container (each ONU can have one or multiple transport containers) and update its assigned bandwidth. The assigned bandwidth is then converted into a burstiness (rate and size of bursts). Bandwidth can also be (partially or fully) assigned as a fixed value. The upstream latency is influenced by the DBA behaviour, with a trade-off between latency and efficiency (more frequent bursts means less waiting time between bursts but more overhead).
- Protection of (parts of) the passive ODN and active OLT equipment by automated switch-over from working to protecting parts of the equipment.
- Management and configuration of the ONUs. Note that ONUs operate at Layer 2 only (e.g. for Layer 2 flow classification). An ONU can be connected to (or integrated with) a Layer 3 device like a residential gateway.

PON ports are terminated in the OLT node, which terminates the PON layers and acts as an aggregation and multiplexing node for the multiple PON ports. Contrary to the PON layer, the

OLT's features are not strictly standardized, although reference deployment models are defined in Broadband Forum documents (See [177], [178]). Typical OLT features include;

- QoS-aware packet forwarding between user side and network side (unicast, multicast, Layer 2 switching, Layer 3 routing)
- Traffic management (classification, (re)marking, queuing and scheduling, policing, shaping…)
- User-side protocol interaction (e.g., DHCP relay agent, IGMP proxy, ARP relay, …) or on the contrary protocol transparency
- Network-side protocol support (e.g., IP routing protocols, IGMP proxy, MPLS signalling, …)

The datagrams (typically IP over Ethernet) undergo several handlings when being transported from end to end over a PON system. The different steps can be summarized as follows:

Downstream (OLT SNI ingress to ONU UNI egress)

1. Traffic Management (classification, policing, queuing and scheduling) at ingress (OLT SNI),
2. Forwarding (Layer 2 switching or Layer 3 routing) towards corresponding PON MAC (PON port),
3. Encapsulation in PON MAC (according to standard, ITU-T or IEEE),
4. OLT PON port egress queuing and scheduling,
5. Fiber propagation,
6. ONU (Layer 2 based) forwarding to corresponding UNI,
7. Decapsulation of PON MAC layer,
8. ONU UNI egress queuing and scheduling (Layer 2)

Upstream (ONU UNI ingress to OLT SNI egress)

1. DBA processing in OLT for burst-based TDMA transmission (acts as policer and shaper), OLT notifies grants to ONUs by sending a Bandwidth Map to all ONUs,
2. Traffic Management (Layer 2 classification, queuing and scheduling) at ingress (ONU UNI),
3. Encapsulation in PON MAC (according to standard, ITU-T or IEEE),
4. Burst generation as per bandwidth map,
5. Fiber propagation,
6. Decapsulation of PON MAC layer,
7. Forwarding (L2 switching or L3 routing) to right OLT uplink (SNI),
8. OLT uplink (SNI) egress queuing and scheduling

PONs are widely used for all FTTx use cases (Cabinet, DPU, Office, Home, …), and are designed for multi-service support (residential triple play, business services, Wireless LAN access points). As explained in the next section and [181], PONs have been extended with additional features and are also suitable for multiple Xhaul scenario's. PONs can act as layer 2 legs in the end-end transport, or participate (from OLT upwards) to layer 3 routing.

10.2.2.1 Using PON for Xhaul: specific features

As a PON is an asymmetrical (bandwidth and latency) point-to-multipoint medium, time and frequency synchronization have to be transported using native media means using media converters (in OLT and ONU). The PON technology have in-built mechanisms for providing frequency and phase synchronization. This will be described in [21].

The PON DBA is being extended with a cooperative mode DBA (called Cooperative DBA or CO DBA in short) for interaction with O-RAN CTI ([23], [24]) making it suitable for some Fronthaul needs by improving the multiplexing while keeping transport latency low.

A specific side effect of T(W)DM PONs is the need for discovery and ranging of new ONUs by means of short interruptions (up to a couple of 100's of µs) of upstream traffic. Such interruptions are likely to impact the performance at the mobile radio layer, but several ways to mitigate the impact or reduce or eliminate the interruptions are being developed (ranging notification from OLT to O-DU by CTI, optimization of ranging timing to minimize impact, avoiding ranging on the latency-sensitive channel by means of separate dedicated activation wavelength).

10.2.2.2 Xhaul Use cases: topologies

PON for Fronthaul

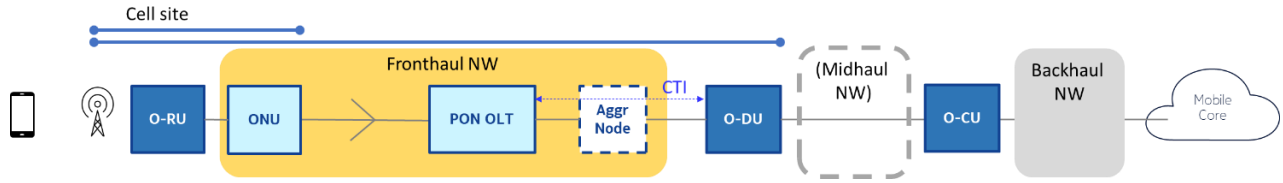


Figure 10-11 PON for transport of Fronthaul flows

Points of attention for Fronthaul deployments are latency at transport level, capacity per O-RU, and time synchronization accuracy (Time Error between the O-RUs). For Fronthaul the O-RUs are connected to ONUs, the OLT is connected to multiple O-DUs (or O-DU/O-CUs), possibly through intermediate aggregating node(s). The PON system can be on-site (e.g., cell site = event zone) or reach up to the cell site (ONU at cell site).
The appropriateness of a PON technology depends on several points:

- the considered radio bandwidth and corresponding Fronthaul throughput and the number of O-RUs per PON that can multiplexed per single PON port
- the obtained latency at packet transport level which depends on the DBA settings (and use of CTI). The requirements for Fronthaul depend on the combination of O-DU and O-RU categories, which have one-way latency budgets ranging from tens to hundreds of µs for usual Fronthaul, and multiple ms for non-ideal Fronthaul (see [17]). Plain DBA is not suited for usual fronthaul, for which Cooperative DBA with CTI is required. Plain DBA can address non-ideal fronthaul. See 10.2.2.3 for more details.
- The time accuracy between O-RUs on PON systems will be described in [21].

PON for Midhaul

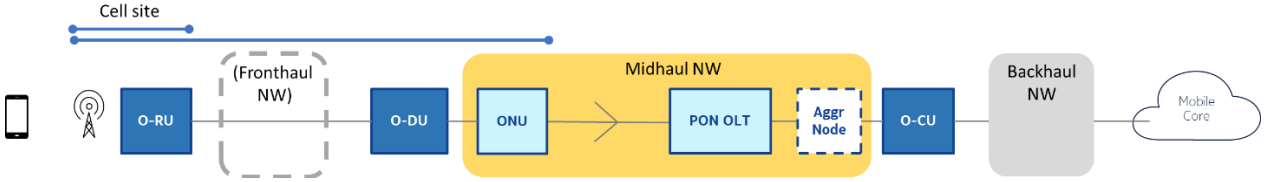


Figure 10-12: PON for transport of Midhaul flows

Points of attention for Midhaul deployments are latency at application level (e.g. for URLLC), QoS differentiation (for achieving better statistical multiplexing), capacity per O-RU. For Midhaul the

O-DUs are connected to ONUs, the OLT is connected to multiple O-CUs, possibly through intermediate aggregating node(s). The PON system (ONU) can reach up to the cell site or to some higher aggregation point.

The appropriateness of a PON technology depends on

- the considered radio bandwidth and corresponding Midhaul throughput,
- the number of O-RUs per PON that can multiplexed per single PON port and the obtained latency at application level (for higher CoS like e.g. URLLC). See 10.2.3.2 for more details.
- the requirement for time alignment accuracy of TDD transmitters between O-RU and the PRTC (Class 4A 3µs) is independent of the topology of O-RU clusters (mutual position of O-RUs on the PONs). The time accuracy with PON systems will be described in [21].

PON for Backhaul

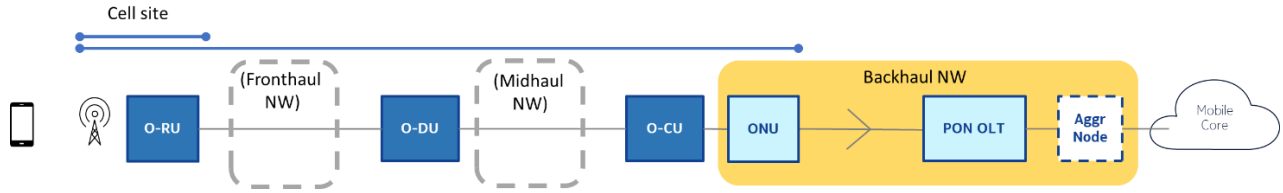


Figure 10-13 PON for transport of Backhaul flows

The points of attention are similar as for Midhaul. For Backhaul the O-CUs are connected to ONUs, the OLT is connected to the mobile core, possibly through intermediate Backhaul aggregating node(s). The PON system (ONU) can reach up to the cell site or to some higher aggregation point. The appropriateness of a PON technology is similar as for Midhaul.

PON for carrying traffic mixes

Note that PON systems are multi-service, hence mixes of different Xhaul flavours can co-exist on the same PON system (e.g. Midhaul + Backhaul) and can also co-exist with non-mobile traffic (e.g. business services). There is also the possibility of re-using the passive fiber infrastructure for different PON technologies by means of WDM overlay, e.g., one system for mobile and one system for residential services. Traffic characteristics of each PON system are then fully independent of the other system.

10.2.2.3 Using PON for Xhaul: trade-offs of the technology

PON systems can be used in many situations, but obviously not all. Each Xhaul transport use case poses specific requirements, which must be considered in light of the possible trade-offs inherent to the PON technology:

1. Latency
    a. versus bandwidth efficiency (making bursts more numerous and shorter decreases latency but increases physical layer overhead).
    b. versus statistical multiplexing (the higher the ratio of variable versus fixed BW allocations, the better the multiplexing but the higher the latency can become). Using CTI allows to reach higher statistical multiplexing while keeping latency limited.
    c. versus reach (5µs/km one-way propagation time)
    d. As guideline, using PONs with fixed bitrate allows to support very low latencies (sub-100µs) but this is not efficient for variable rate fronthaul. PONs with plain DBA efficiently follow variable rate traffic but introduce a latency in the order of one to several ms. When combining CTI with Cooperative DBA in the OLT for variable rate

fronthaul traffic (like O-RAN 7.2x), the extra latency can be reduced by an order of magnitude compared to plain DBA, and such PONs can support several use cases of usual fronthaul and all combinations of non-ideal fronthaul (see O-RU and O-DU combinations as per [17]).

2. Bandwidth per Xhauled node versus the number of nodes served per ODN (bandwidth sharing)
3. phase synchronization accuracy versus topology of O-RU clusters in PONs and path propagation difference up/down (max fiber distance and wavelength-dependent correction factors). This will be described in [21].

10.2.3 DOCSIS Networks

Today's modern cable operators deploy and manage extensive hybrid fiber coaxial (HFC) networks. The HFC plant reaches 93% of American households and connects to virtually every building across north America. The data-over-cable service interface specification (DOCSIS®) is the protocol designed to work on the HFC network. DOCSIS suite of specifications is standardized and maintained by CableLabs® [190][191], and is deployed worldwide. DOCSIS technology has evolved through six generations of progressive refinement. Originally developed for delivering high speed broadband to residential customers, the cable industry has made many advances in the HFC technology, making it a promising option for 5G transport. The HFC plants are already extensively deployed in areas where 5G will be in the most demands, especially in dense urban and suburban environments. Additionally, the HFC plant is active, providing power needed for small cell radios. The cable operators typically have the right of way on the strands where radios can be hung, eliminating the need to obtain permits from government. Leveraging the existing HFC deployments significantly reduces time-to-market and cost of deploying 5G.

10.2.3.1 DOCSIS technology overview

DOCSIS protocol is a Layer 2 transmission protocol that encapsulates the IEEE 802.3 Ethernet. The technology supports both Layer 2 and Layer 3 services and is aware of both. DOCSIS network is a point-to-multipoint access network, where the transmission of upstream and downstream is controlled by the cable modem termination system (CMTS). Figure 10-14 shows a typical HFC plant. The CMTS is typically located in either a headend or a hub, which is connected to the fiber nodes through digital or analog fiber. From there, the communication path traverses through a coax network of zero or more amplifiers to the neighborhoods. The hybrid (fiber and coax) network topology is called "N+x", where "N" is the fiber node, where the fiber ends and coax starts, and "x" is the number of amplifiers traversed. The average north American plant is between N+2 and N+5. The smaller the x, the closer the fiber node it is to the neighborhood, the larger the amount of capacity would be available to each customer.

The traditional physical CMTS controls 50k cable modems (CMs). It is expected to scale up as the CMTS moves to the cloud native architecture.

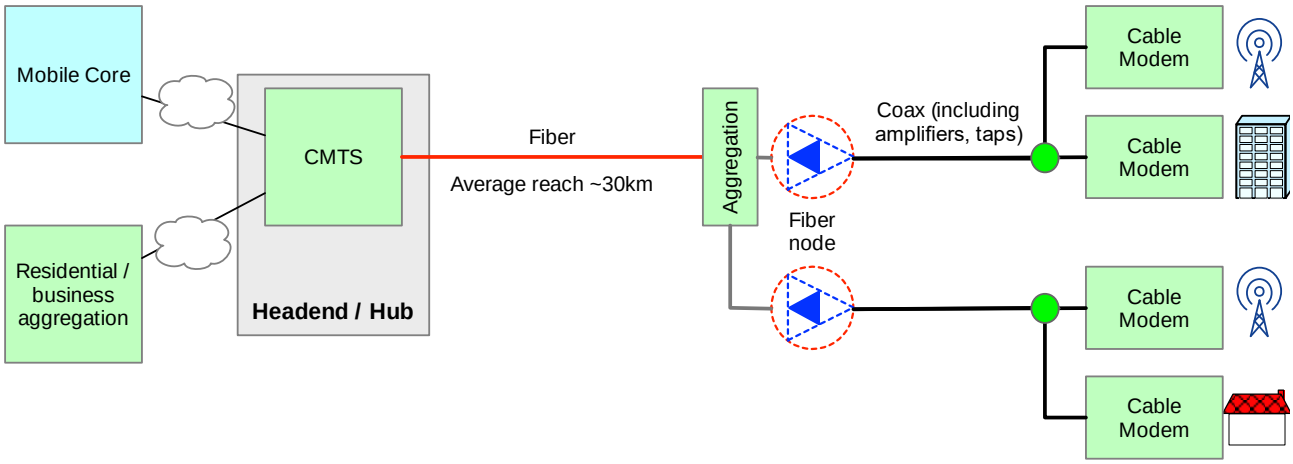


Figure 10-14: Typical HFC plant

Table 4 shows the capabilities of the most recent DOCSIS standards. While DOCSIS 4.0 will provide over 10 Gbps of downstream capacity, even higher rates can be supported. A unique aspect of the DOCSIS technology is its ability to progressively expand downstream and upstream capacities when and where it is needed. Traditionally, the same plant is used to provide video and broadband services. As the cable industry moves away from traditional video delivery, the RF spectrum that is used for video can be reclaimed for broadband. Moreover, as cable operators moves to reduce the "x" in the N+x topology, node segmentation takes place. For every node that is segmented into two, the network capacity essentially doubles.

Requirements	D3.1 today (2020)	D3.1 max (2021-22)	D4.0 (2023-24)
	Shared spectrum with video	Full spectrum through video reclamation	Extending to 1.8 GHz, possibly 3 GHz
Downstream spectrum Upstream spectrum	54 – 1002 MHz 5 – 42 MHz	258-1218 MHz 5 – 204 MHz	602 – 1794 MHz 5 – 492 MHz
DS capacity US capacity	8.5 Gbps 0.1 Gbps	8.6 Gbps 1.4 Gbps	10.8 Gbps 3.7 Gbps
Upstream latency	Best effort: 5 – 50 ms With LLX / CTI: 1 – 2 ms (can be further reduced)		
Synchronization	Frequency sync only No time sync	Frequency + time sync through DTP	

Table 4 DOCSIS Capabilities; Today and the near future

The DOCSIS specification provides a rich set of QoS mechanisms, including traffic classification, queuing, multiple scheduling services, policing, traffic shaping. Each CM supports 16-32 service flows in each direction. DOCSIS technology supports several native upstream scheduling mechanisms: best effort, unsolicited grant service (UGS), real-time polling service (RTPS), and proactive grant service (PGS). Latency on the upstream can range from minimum of 3-5 ms to average of 12 ms under best effort, to 1-5 ms under UGS or PGS. For best effort scheduling service, latency is dependent on the load, as well as frame configurations such as OFDM frame size and interleaver depth. For UGS/PGS services, there are trade-offs between incurring capacity overhead vs. lowering the latency. The recommendation is to transport mobile Xhaul traffic over the DOCSIS network via best effort service. To address the latency concern, CableLabs standardized the Low

Latency Xhaul (LLX) technology [188], which provides consistent 1-2 ms of latency on the upstream. The latency could potentially be further reduced with some configuration optimizations. The ORAN CTI ([23], [24]) specification uses a similar scheduler pipelining mechanism as LLX. Typically, the latency on the downstream is 1ms.

Cable operators are in the midst of disaggregating the traditional CMTS. The "Distributed Access Architecture" (DAA), shown in Figure 10-15, has been specified by CableLabs [192] and is being deployed. Replacing the traditional "integrated" CMTS, PHY layer component of the CMTS is being pushed to the fiber node and is called Remote PHY Device (RPD), while MAC and upper layers are implemented as DAA Core, and is located in the headend or regional data center. The DAA Core has been architected with the cloud native architecture. The links between the fiber node and the hub are being upgraded to a digital fiber network, with DWDM as a starting point, but some operators may choose to migrate to point-to-point coherent optics. A device called "CTD" is being specified to terminate the fiber from the hub, and is essentially be a Layer 2 switch that provides high-speed Ethernet to the neighborhood. A multitude of services can be provided through the CTD, including 5G Fronthaul, PON, etc, thus achieving a form of transport convergence.

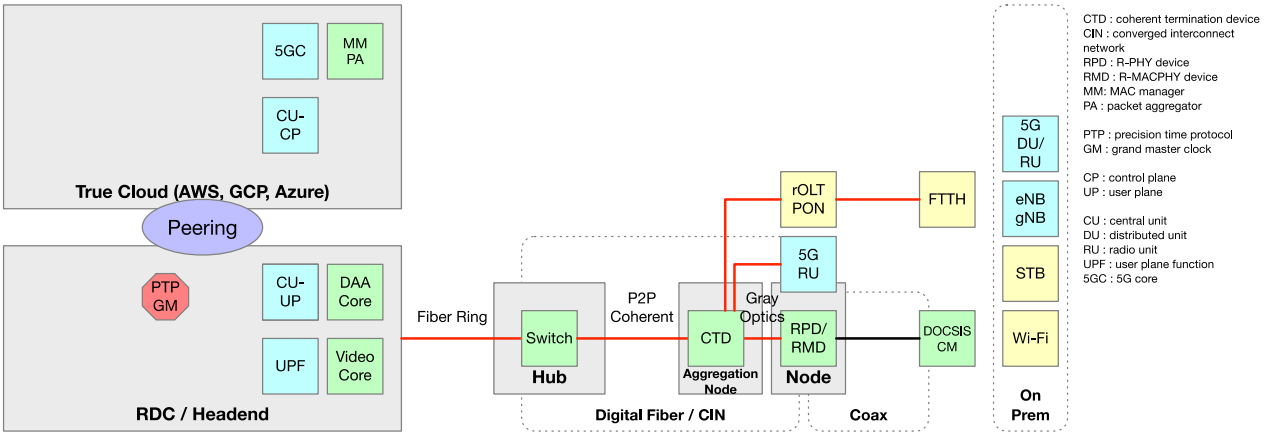


Figure 10-15 Distributed Access Architecture (DAA) for transport convergence

10.2.3.2 Xhaul transport DOCSIS network

Functionally, the DOCSIS network can interconnect the small cells or radios as shown in Figure 10-16. However, the requirement on capacity and latency vary between Backhaul, Midhaul, and Fronthaul. Backhaul and Midhaul are both based on an IP encapsulation of the original mobile transport. Thus, the bandwidth requirement on the transport network roughly matches the mobile traffic rate.

The latency requirement for Backhaul is based on the application. Additional service-level agreement (SLA) can be specified by the mobile operator. Midhaul latency is generally considered to be less than 10 ms [193]. LLX can be implemented to ensure these requirements can be met, as well as providing better latency performance, particularly for latency-sensitive flows.

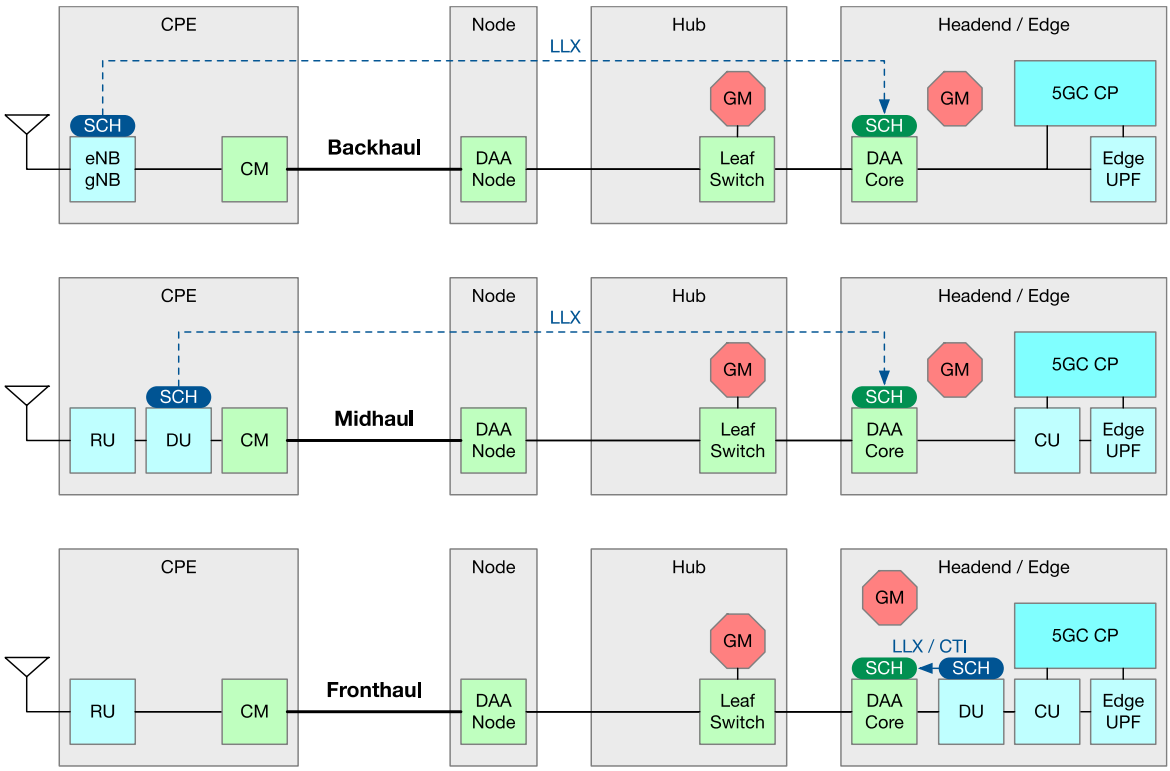


Figure 10-16 Mobile Xhaul over DOCSIS Network

Fronthaul is much more difficult to support over the coax portion of the HFC network. Studies have shown that the eCPRI-based Fronthaul transport needs significantly more bandwidth, overhead, and one-way latency in the neighborhood of 250 microseconds between the O-RU and O-DU and between TNEs 100 microseconds. Even with LLX, it will be difficult to reduce the DOCSIS latency to this level. Because of the stringent requirements, Fronthaul is better carried directly over the digital fiber network and over CTD as shown in Figure 10-16.

10.2.4 Microwave and mmwave radio transport technologies

10.2.4.1 Overview of Microwave and mmwave technology

For over 20 years, microwave has been the primary solution for the rapid and cost-effective rollout of mobile Backhaul infrastructure with over 50% of mobile sites worldwide today connected via Microwave (MW) or Millimetre Wave (mmW) radio links and up to over 90% in some networks. The evolution from 4G towards 5G presents significant challenges to all transport technologies and wireless ones make no exception.

Spectrum is a vital asset to support the mobile Backhaul requirements and certainly this topic becomes increasingly relevant as future mobile access data rates and respective Backhaul capacity requirements continue to rise. Various frequency bands are used today for mobile Backhaul depending on the requested capacity, hop length, link availability, spectrum availability and frequency re-use capability. The spectrum that is available for mobile Backhaul ranges from the traditional microwave bands up to the millimetre wave spectrum as illustrated in the Figure 10-17. These bands are allocated for fixed services by the WRC Radio regulations. Their channel plans are detailed by the ITU-R F-series recommendations as well as some regional or national regulatory regimes, such as CEPT. Fixed services spectrum uses authorisations awarded by national

administrations on the basis of link by link assignments as either coordinated or self-coordinated, or as block assignments.

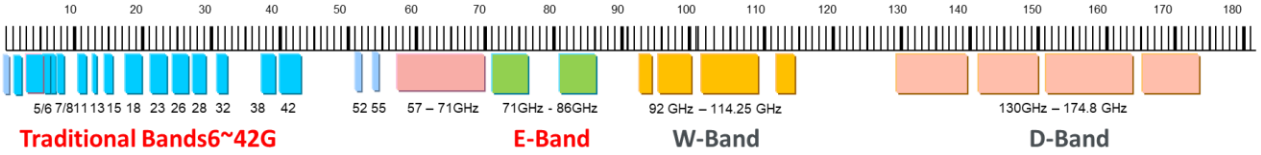


Figure 10-17 Microwave and mmwave spectrum

The engineering of a MW or mmW link involves finding the optimal combination of link length, capacity, frequency band and availability.

The physics of radio waves propagation determine the relation among capacity, availability and link length.
Since the available spectrum is proportional to the center frequency, the highest frequencies are also those that carry the most capacity, but also cover the comparatively shortest link lengths.
As a rule of thumb, frequencies below 13 GHz can be considered mostly unaffected by the intensity of rainfall and frequencies above are more and more influenced by the attenuation caused by rain, so that as a general principle higher frequencies are used for shorter links, as illustrated in Figure 10-18. It should be noted that features could be inherited for links combining different bands as indicated in the circles in this figure.

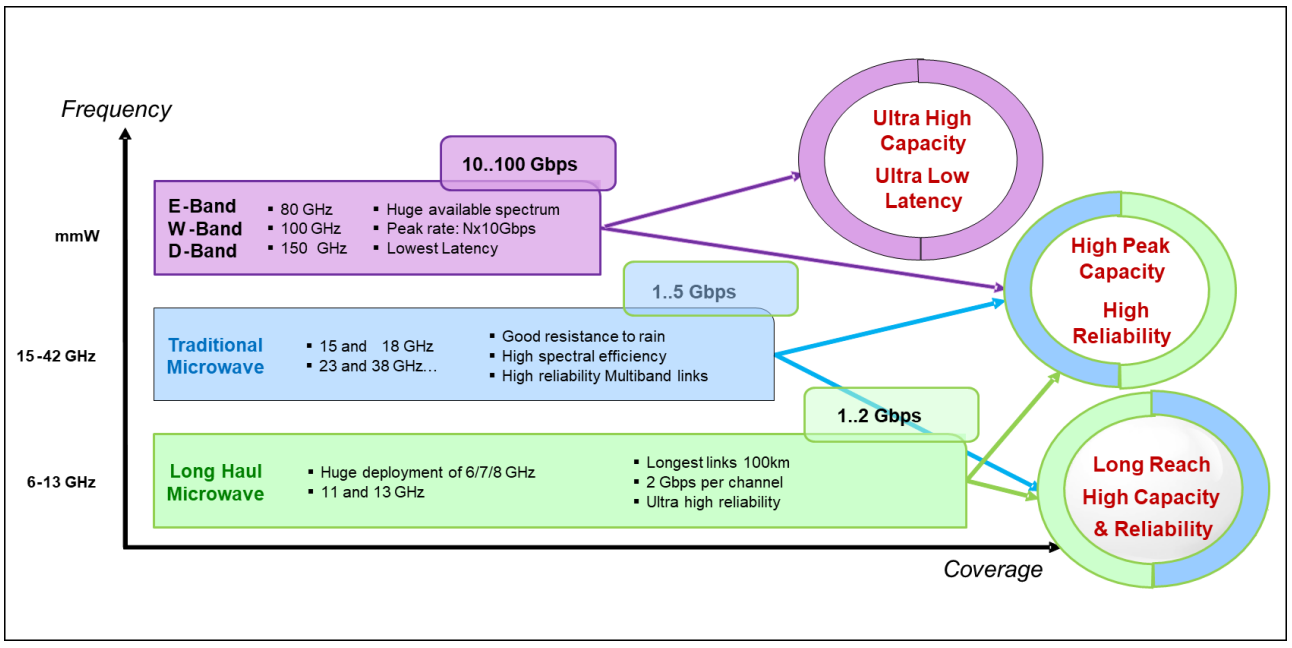


Figure 10-18 Interdependence among frequency, capacity and availability. Source ETS TS mWT

The availability of MW/mmW spectrum depends on both technological and regulatory factors. Technology is available and under development to make full use of existing (6-86 GHz) and future (90-300 GHz) spectrum: E-band (80 GHz) has been commercially deployed for several years, W-band (100 GHz) and D-band (150 GHz) are the most promising upcoming bands, with prototypes already appearing in the market.

Wider channels (112MHz, even 224MHz where possible) in traditional frequency bands and raw availability of spectrum (10GHz in E-band, 18GHz in W-band and 30GHz in D-band) provide the main resources to expand the capacity of MW and mmW radio systems. Overview of available spectrum for wireless transport network with corresponding capability related to link capacity, latency and hop length is illustrated in Figure 10-19.

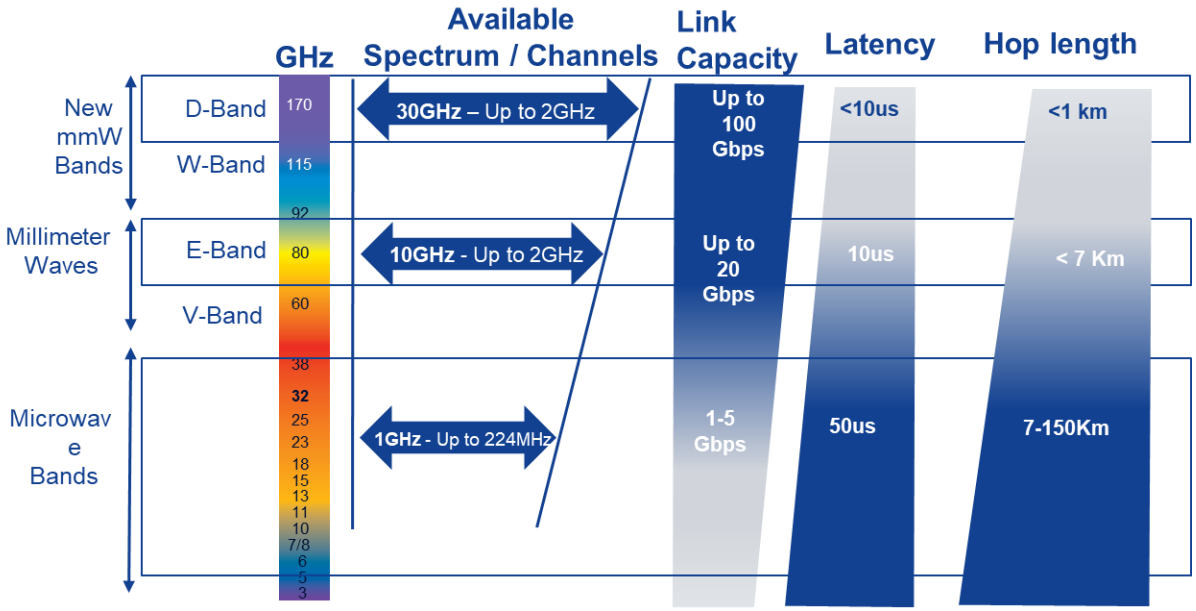


Figure 10-19 Bands characteristics and capabilities. Source ETSI TSG mWT

10.2.4.2 Evolving technologies

There are a number of existing and evolving technologies which are applied in microwave and mmwave radio to enhance its spectrum efficiency, achievable capacity and reducing latency. An overview of these technologies is presented in[201]. As follow is a summary of these applicable technologies in the microwave and mwave radio transport systems.

Capacity and spectrum efficiency enhancement

Larger channels are no longer a technology limit. In MW bands recent regulatory limit shifted up to Channel Spacing CS=224MHz, but not everywhere. Up to CS=2000MHz is standardised in E-Band and above 100GHz.

Larger CS are needed where Carrier Aggregation, in same band or adjacent band is employed.

Higher Modulation schemes reached the reasonable top at 4096QAM (and more). After 1024QAM spectral efficiency gain is less than 10% for every step. Higher order result in reduction in system gain (impacting availability), however, with adaptive modulation availability can be maintained with lower modulation order (i.e. scarifying capacity).

Frequency reuse with cross polar interference cancelation (XPIC)

This is a well-known technique for doubling the spectral efficiency by utilising the cross polarisation of the channel bandwidth occupied.

LoS-MIMO Line of Sight Multi-Input Multi-Output

Exploiting link geometry deployment, two different signals in the same channel can be transmitted.
4x4 LoS-MIMO is obtained with LoS-MIMO 2x2 plus XPIC
LoS MIMO needs optimal antennas separation.
Under optimal conditions, spectral efficiency close to four times capacity improvement can be achieved, while lower performance in case of suboptimal conditions should be expected.

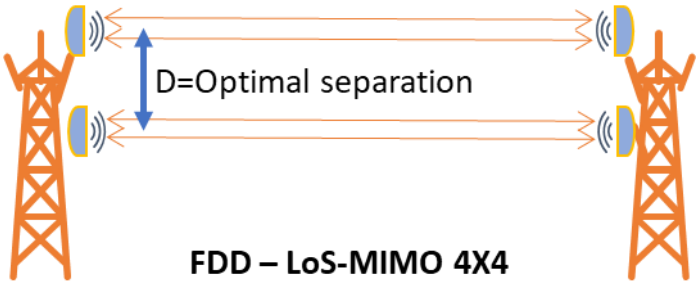


Figure 10-20 4x4 Multiplexing with LoS MIMO and XPIC

Orbital Angular Momentum

OAM is a new transmission mechanism allowing multiplexing multiple streams simultaneously over the same frequency channel without the limitation imposed by the optimum separation distance of antennas in the LoS MIMO. Using different antennas, multiple OAM signals with different spiral phase front (mode) can be transmitted. OAM modes are orthogonal of each other. A pragmatic OAM system, uses uniform circular antenna array while the OAM signal is generated at the base band as illustrated in the Figure 10-21**.**

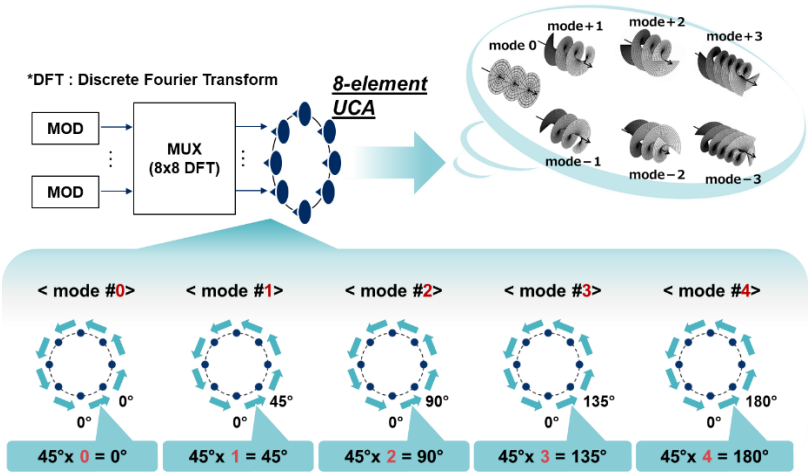


Figure 10-21 Orbital Angular momentum transmission

An OAM system with 8 Uniform Circular Array UCA antenna elements allows the transmission of 16 orthogonal streams resulting in capacity of around 105Gbps. [187]

Bands & Carriers Aggregation (BCA)

BCA joins different channels that may be even in different bands, providing a single big capacity pipe. Lower band will provide capacity pipe's segment with high availability, while higher band the best effort capacity pipe segment. Packets may be adaptively re-routed among different channels according to their priority and channels condition.

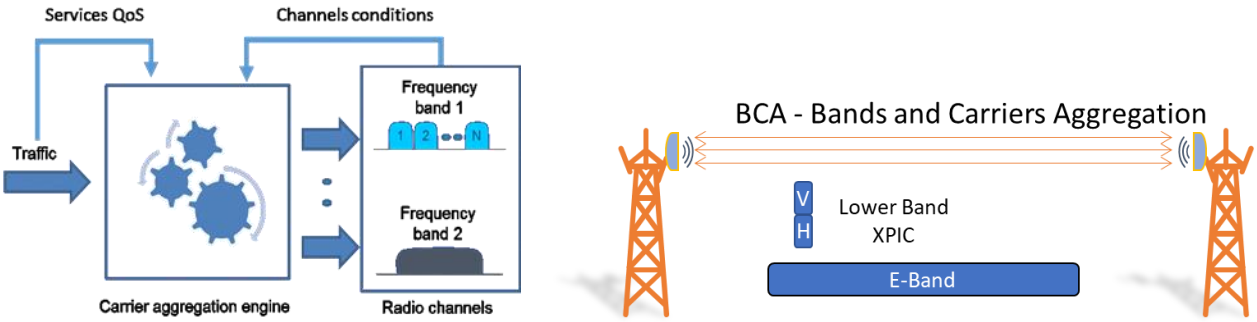


Figure 10-22 Bands and carrier aggregation concept and benefits

One of the most valuable approach is 15/18/23 GHz with E-Band where dual band antennas are available:

- Links up to 7-10Km are feasible. Capacity may even exceed 10Gbps
- High spectral efficiency obtained because E-Band can reach longer links than in traditional approach.

BCA among two MW bands is another variant when distance becomes more challenging i.e.: rural application.

Geographical spectral efficiency: Dense reuse of channels

To better exploit the scarce resource (spectrum) it is advisable to increase not only the single channel spectral efficiency but also the channel reusability in a given area, guaranteeing the "interference free operation".

Nodal configuration is the key point to understand the concept of geographical spectral efficiency. Better antenna class are introduced (e.g. ETSI Class 4), reducing significantly the minimum angle between two links using the same/adjacent channels (angle discrimination)
Cross polar (XPIC) can also be used in reducing angle discrimination.
Co-Channel Interference Canceller (CCIC) further improve the re-use of channels with very narrow angle discrimination.

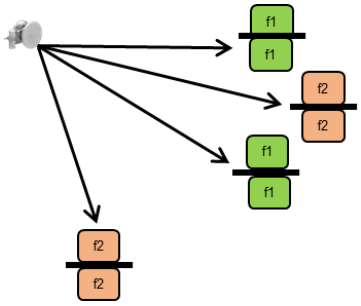


Figure 10-23 Increase nodal capacity is now easy with no additional spectrum with XPIC

When additional capacity is needed and then additional channels shall be used, CCIC permits an optimal re-use of channels with very narrow angle discrimination

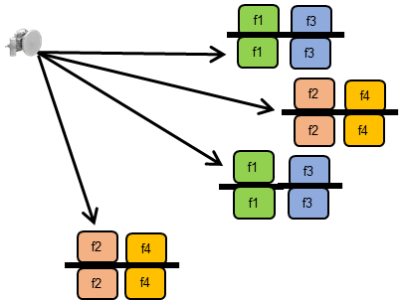


Figure 10-24 Further increase nodal capacity with increased spectral utilization by applying CCIC technique

Combination of the above technologies gives enhancements to microwave/mmwave capabilities as indicated below [200].

MW Backhaul Technology	56 MHz BW	112 MH BW	224 MHz BW	+XPIC	+ Los 2x2 MIMO	+ BCA (with higher MW Band)	+BCA (with mmW Band)
6-15 GHz	0.5 Gbps	1 Gbps		2 Gbps		3-4Gbps	
18-42 GHz	0.5 Gbps	1 Gbps	2 Gbps	2-4 Gbps	4-8 Gbps		4-10 Gbps

mmW Backhaul Technology	500 MHz BW	2 GHz BW	4 GHz BW	+XPIC	+LOS 2x2 MIMO/OAM
V-band (60GHz)		>4 Gbps			
E-band (70/80GHz)	3.2 Gbps	12.8 Gbps		25.6 Gbps	51.2 Gbps
W-band (100GHz)	3.2 Gbps	12.8 Gbps	25.2 Gbps	51.2 Gbps	102.4 Gbps
D-band (150GHz)	3.2 Gbps	12.8 Gbps	25.6 Gbps	51.2 Gbps	102.4 Gbps

Table 5 Microwave/mmwave enhancements

10.2.4.3 Support for packet

MW or mmW links are characterised by their physical capabilities related to the combination of link length, capacity, frequency band and availability, which is independent of their network layer functionalities and use cases. These radio links are usually capable of multiplexing various types of traffic over single radio channel and can be equipped with multiple network interfaces including FE/GbE (RJ45, SFP) and CWDM filter support. Various packet functions are also supported as optional features and can be configured by network operators, such as L2 pass through, feature rich L2 switch with ERPS, H-QoS among others enabling carrier grade Ethernet services. MPLS-TP is also supported for layer 2 services, and for timing and synchronisation support includes 1588v2 (TC, BC) and SyncE.

These network features make MW and mmwave radio popular in Backhaul and Xhaul network infrastructure by themselves. They are also used as a complementing technology to other physical technologies where the MW and mmwave radio links can be used as part of network segments to close links where underlying technology infrastructure is missing or absent or as a back-up in mission critical part of the network. The section below illustrates some use case examples where MW and mmwave radio links are utilised.

10.2.4.4 Topology example use cases

Dense urban/urban (C-RAN plus D-RAN support) case

The figure below illustrates an example of transport network use in an end-to-end network with C-RAN infrastructure used to provide network coverage in a Dense Urban/Urban environment with support for some D-RAN sites.

In this network, the core is connected to the O-CU via a Backhaul aggregate and pre-aggregate transport part of the network. At the aggregation/pre-aggregation transport network, a bandwidth of 10G or more will be required, mainly 25G or more. MW and mmW will be used as alternative (ring closer) or redundant circuit to fibre, and the application of wide-bandwidth products in the E/D/W bands will be the main focus.

At the Midhaul transport network, MW and mmW radio equipment can be used in almost the same band as conventional Backhaul. From the transport perspective, there is little different between 5G and 4G transport networks, with additional synchronisation and delay requirements.

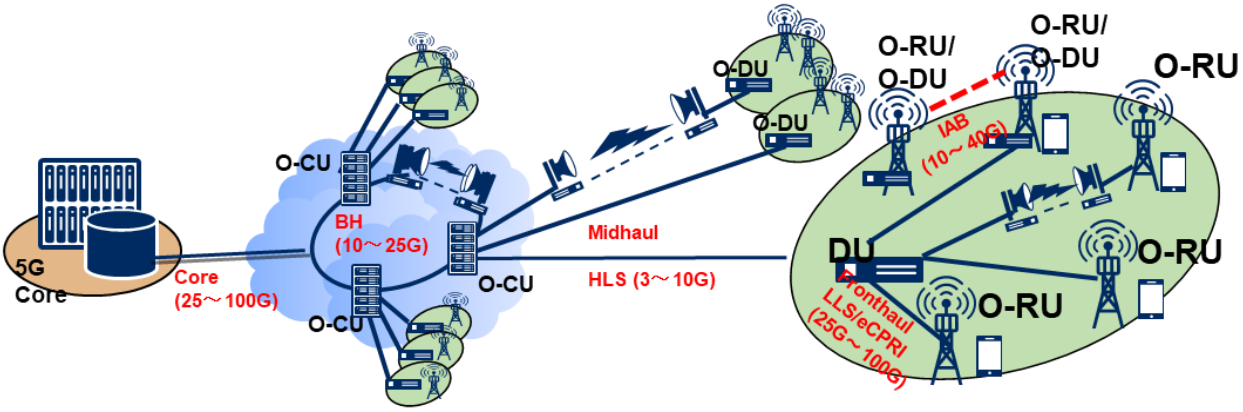


Figure 10-25 Network topology example for dense urban/urban environment with microware and mmwave links

In urban centers, high bandwidth services are required, hence Fronthaul LLS needs to be more than 25G (25-100G). In many cases, LLS is used when DU/RU are collocated in the same building. It may be in short distances. In this case, MW/mmW radio equipment is of limited application.

Within the urban environment, some D-RAN sites may also be connected to edge site. In this case IAB (integrated Access Backhaul) base station to base station communication, which cover about 100 metres with LoS, can be used.

Rural (C-RAN) case

In this topology scenario, coverage in rural areas is migrated using existing fibres in a C-RAN configuration for 4G. Generalisation of equipment is made possible by upgrading CPRI links to Ethernet utilising eCPRI or IEEE1914.3 mechanism. The link capacity requirements in this case is in the range 10 to 25G. The application of MW/mmW radio equipment is used to extend coverage and other uses due to its ease of installation.

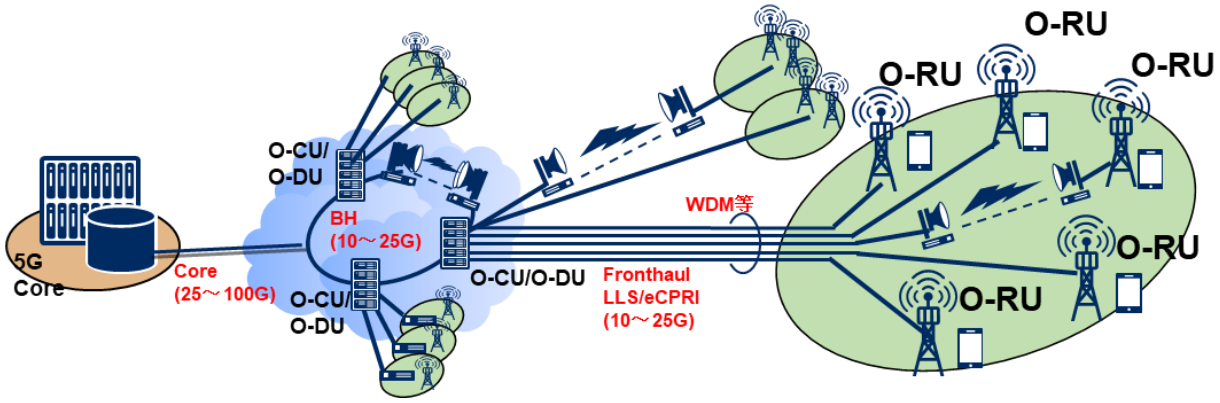


Figure 10-26 Network topology example for rural environment with microware and mmwave links.

10.3 Data Centers

Data center solutions, connectivity to the Xhaul transport infrastructure and orchestration are not in scope of this version of document. However, it is important to note the critical role Data centers play in 5G, with key components of the 5G architecture running on servers either virtualized, containerized or as bare metal located in data centers. Consequently, operators need to consider how data centers are connected to the Xhaul transport network and how end to end transport services are built that span the DC and the WAN. At a high-level two basic approaches to interconnecting DC to the Xhaul transport network are available.

10.3.1 Complete separation between DC and WAN infrastructure

This refers to data centers that are discrete from the WAN infrastructure with each domain potentially running different underlay and overlay transport network technologies. The inter-connection between the two domains is via “Data Center Interconnect” (DCIs) routers. These routers reside in both the DC and WAN domains and have an awareness of the DC environment on one side and an awareness of the WAN on the other. To provision an end-to-end network service, the data center and the WAN are provisioned separately, using a cross-domain orchestrator to stitch the two separate service environments together.

10.3.2 DC integrated into WAN infrastructure.

This refers to data centers that run the same underlay and service infrastructures as the WAN. In this case, although there maybe DCIs to create underlay separation between the WAN and the DC, it is possible to treat the network service infrastructure as a single orchestration domain with WAN transport services directly configured on the Top of Rack (ToRs) devices or on soft switches / routers residing on the server themselves.

11 Packet-switched underlay network – MPLS based

This document presents two packet switched underlay technologies, the first is based on MPLS contained in this section (section 11) and the second based on SRv6 contained in section 12. An operator wishing to implement the transport architecture outlined in this document will need to select one of the two and implement the requirements outlined in the associated section. It should be

noted that other transport architectures are potentially available but not covered in this revision of the document.

This section outlines a packet switched underlay model based on “Multi-Protocol Label Switching” (MPLS). MPLS data plane is based on label switching but has more than one control plane technology. The first, classic approach is based on the control plane developed when MPLS was first conceived in the mid 1990s and utilizes IGPs, BGP-LU, LDP and RSVP. An emerging approach, very appropriate to mobile transport networks, is based on Segment Routing (SR) extensions to IGP and BGP, which have been developed since around 2010 and aim to simplify and minimize the number of protocols running in network, as well as reduce the state held on the TNEs. This document describes scalable, multi-domain Xhaul transport architecture that can involve domains with different MPLS control plane technologies (classic: LDP, RSVP, BGP-LU, as well as emerging: SR, SR with Flex-Algo, or SR-TE), in order to support both brownfield (extending existing mobile transport network using classic control plane with additional domains using new, emerging SR control plane) and greenfield (building new Xhaul transport network) deployment use cases. However, in order to avoid duplications with BBF technical reports documenting classic MPLS control plane in mobile transport network designs, and to keep the size of this document within reasonable limits, this document puts more focus on the new, emerging control plane based on SR.

Regardless which MPLS control plane – classic or SR – is used to distribute underlay transport labels, the services are overlaid on-top. The service layer is independent from the underlaying MPLS transport layer and supports native Ethernet services and layer 3 services.

11.1MPLS data plane

MPLS architecture relies on an MPLS data plane using MPLS encapsulated IP (for L3VPN) or Ethernet (for L2VPN/EVPN) data packets. MPLS label stack, containing potentially multiple labels, pushed on the data packet might encode various information, like for example:

- path through the transport network (transport labels)
- entropy for load balancing on transit routers (flow or entropy labels)
- function/service on egress router (L3VPN/EVPN labels)

Figure 11-1 shows examples for Ethernet based eCPRI and IP based eCPRI encapsulated in MPLS.

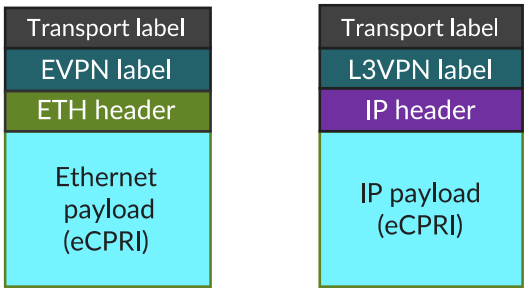


Figure 11-1 MPLS encapsulation of Ethernet frame or IP packet

To support packet switching transport with MPLS data plane, the transport device will need to support:

[R1]: MUST support MPLS architecture as defined by "Multiprotocol Label Switching Architecture", RFC 3031 [57].

[R2]: MUST support MPLS label stack encoding defined by "MPLS Label Stack Encoding" RFC 3032 [58].

[R3]: MUST support TTL processing defined by "Time To Live (TTL) Processing in Multi-Protocol Label Switching (MPLS) Networks", RFC 3443 [65].

[D1]: SHOULD support MPLS Explicit Null operation defined by "Removing a Restriction on the use of MPLS Explicit NULL", RFC 4182 [74].

11.2 MPLS control plane

The MPLS control plane is used to distribute information required to deliver the packets through the transport network. It typically includes following information:

- Topological information (nodes and links in transport network, including attributes, like metrics, IP prefixes administrative groups, SRLGs, etc., associated with these topological elements)
- IP prefixes
- MPLS transport label information

Based on the collected information, control plane responsibility is to

- calculate paths (shortest path to the destination, or path fulfilling certain administrative constraints, which are not necessarily the shortest paths – for example low latency path)
- program data plane with information required to forward the MPLS packets through the network (next-hops, MPLS labels stack manipulation rules – pop, push, swap)
- calculate and program in the data plane backup paths required for rapid protection mechanisms (fast re-route – FRR)

In a packet switched transport network consisting of a large number of routers, a mix of WAN and data centers components, the need to support traffic engineering, the underlay control plane is a typically a collection of independent routing domains that interact in a collaborate fashion.

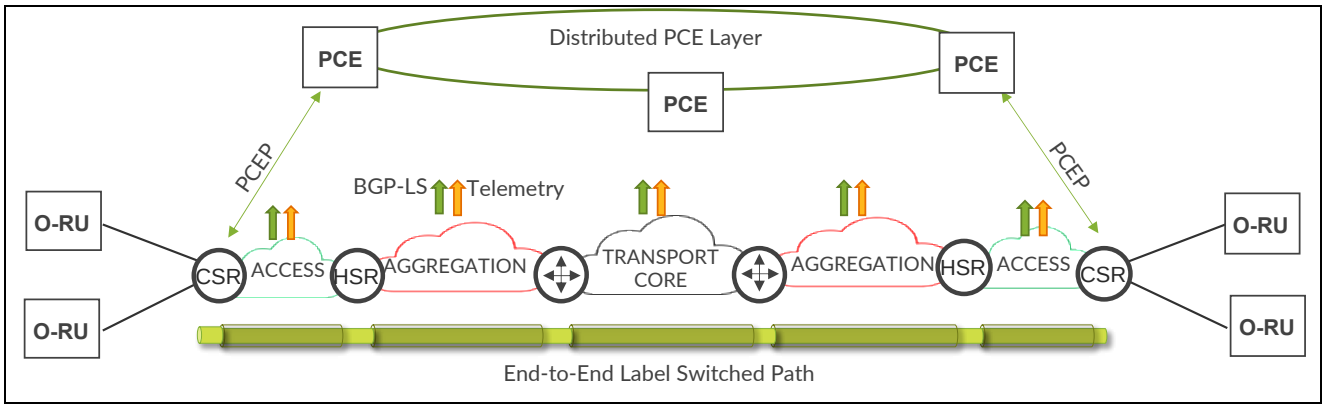


Figure 11-2 MPLS packet switched underlay architecture

Figure 11-2 illustrates how a large packet switched MPLS transport infrastructure might be designed and how different components interact. It should be noted that not all these components are required, and some serve the same purpose but in different ways.

1. Each routing domain has an IGP for internal connectivity.
2. Each routing domain has a protocol for distributing label information. It could be SR extension to an IGP, or could be as well legacy protocol like RSVP or LDP. Architecture supporting legacy MPLS transports (LDP and/or RSVP) is especially important in brown-field deployments, with legacy MPLS transport already in place.
3. Mechanisms to calculate Traffic Engineered paths within a routing domain. Depending on scale, TE paths can be calculated on the transport devices themselves (distributed constrained shortest path first – D-CSPF – calculation), calculation can be off-loaded from less powerful transport devices to more powerful transport devices (on-box PCE), or even dedicated servers could be deployed (off-box PCE). The diagram shows a distributed "Path Computation Element" (PCE) layer with "Path Computation Element Protocol" PCEP running between the edge nodes and the PCE
4. Mechanism to establish end-to-end MPLS LSPs between IGP domains. To achieve highly scalable architecture, two mechanism could be considered: Seamless MPLS Architecture, or a controller-based Architecture. Both architectures are described in more details in Section 11.5.
5. Mechanism to convey topology and network state information from the network to the PCE and other management elements. BGP Link State (BGP-LS) and Telemetry feeds are recommended tools to fulfil this requirement.

11.3Classic MPLS control plane

The classic control plane is widely implemented, and designs and requirements are well documented in:

- Cisco Press, MPLS and VPN Architectures, Volume 1, 420 pages, by Ivan Pepelnjak, and Jim Guichard, 2001 [194]
- Cisco Press, MPLS and VPN Architectures, Volume 2, 470 pages, by Ivan Pepelnjak, Jim Guichard, and Jeff Apcar, 2003 [195]
- O'Reilly, MPLS in the SDN Era, 890 pages, by Antonio Sánchez-Monge, and Krzysztof Grzegorz Szarkowicz, 2015 [196]

The application of classic MPLS control plane protocols (IGP, LDP, RSVP, BGP) to the mobile transport networks is defined by Broadband Forum (BBF) in following technical reports:

- BBF TR-221: Technical Specification for MPLS in Mobile Backhaul Networks, 99 pages, Oct 2011u [197]
- BBF TR-221, Amd.1: Technical Specifications for MPLS in Mobile Backhaul Networks, 24 pages, Nov 2013 [198]
- BBF TR-221, Amd.2: Technical Specifications for MPLS in Mobile Backhaul Networks, 22 pages, Sep 2017 [199]

11.3.1.1 LDP base requirements

If an operator wishes to use LDP MPLS control plane in some routing domain, then the routing equipment in that routing domain will require:

[D2]: SHOULD support constraint-based LSP setup using LDP defined by “Constraint-Based LSP Setup using LDP”, RFC 3212 [61]

[R4]: MUST support LDP state machine defined by “LDP State Machine”, RFC 3215 [62]

[D3]: SHOULD support graceful restart mechanism for LDP defined by “Graceful Restart Mechanism for Label Distribution Protocol”, RFC 3478 [67]

[R5]: MUST support LDP defined by “LDP Specification”, RFC 5036 [88]

[D4]: SHOULD support LDP extension for inter-area label switched paths defined by “LDP Extension for Inter-Area Label Switched Paths (LSPs)”, RFC 5283 [91]

[D5]: SHOULD support LDP IGP synchronization defined by “LDP IGP Synchronization”, RFC 5443 [103]

[D6]: SHOULD support LDP capabilities defined by “LDP capabilities”, RFC 5561 [106]

11.3.1.2 RSVP base requirements

If an operator wishes to use RSVP MPLS control plane in some routing domain, then the routing equipment in that routing domain will require:

[R6]: MUST support “Resource ReSerVation Protocol (RSVP)”, RFC 2205 [41], RFC 2209 [42] , RFC 2210 [43]

[R7]: MUST support “RSVP Diagnostic Messages”, RFC 2745[51]

[R8]: MUST support “RSVP Refresh Overhead Reduction Extensions”, RFC 2961[55]

[D7]: SHOULD support “RSVP Cryptographic Authentication—Updated Message Type Value”, RFC 3097 [59]

[R9]: MUST support “RSVP-TE: Extensions to RSVP for LSP Tunnels”, RFC 3209 [60]

[D8]: SHOULD support “Signalling Unnumbered Links in Resource ReSerVation Protocol - Traffic Engineering (RSVP-TE)”, RFC 3477 [66]

[R10]: MUST support “Traffic Engineering (TE) Extensions of OSPF Version 2”, RFC 3630 [69]

[D9]: SHOULD support “Fast Reroute Extensions to RSVP-TE for LSP Tunnels”, RFC 4090 [72]

[D10]: SHOULD support “OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)”, RFC 4203 [78]

[D11]: SHOULD support “Node-ID Based Resource Reservation Protocol (RSVP) Hello”, RFC 4558 [80]

[R11]: MUST support “Record Route Object (RRO) Node-Id Sub-Object”, RFC 4561[82]

[R12]: MUST support “IS-IS Extensions for Traffic Engineering”, RFC 5305 [96]

[D12]: SHOULD support “Traffic Engineering Extensions to OSPF Version 3”, RFC 5329 [99]

[R13]: MUST support Encoding of Attributes for MPLS LSP Establishment Using Resource Reservation Protocol Traffic Engineering (RSVP-TE), RFC 5420 [101]

[R14]: MUST support “Label Switched Path (LSP) Attribute in the Explicit Route Object (ERO)”, RFC 7570 [119]

[D13]: SHOULD support “Techniques to Improve the Scalability of RSVP-TE Deployments”, RFC 8370 [131]

[D14]: SHOULD support “Signaling RSVP-TE Tunnels on a Shared MPLS Forwarding Plane”, RFC 8577 [138]

[D15]: SHOULD support “Refresh-interval Independent FRR Facility Protection”, draft-ietf-mpls-ri-rsvp-frr [161]

11.4 SR/MPLS control plane

The following section outlines the control plane requirements for SR/MPLS environment.

11.4.1 Interior Gateway Protocol (IGP) for SR/MPLS

The distribution of labels, rapid convergence and distribution of traffic engineering information and optional calculation of TE optimized transport planes in a single SR domain is done by a link-state IGP, in the form of either IS-IS or ISPF, with suitable SR enhancements. This document separately outlines the requirements for IS-IS and OSPF with support for TiLFA for fast convergence and optional support for flex-algorithm for the creation of multiple transport planes optimised for different criteria and potentially using different topologies.

SR/MPLS requires at least one IGP with SR awareness per routing domain or autonomous system. There are two IGPs that support SR/MPLS; ISIS or OSPF. They are both link-state protocols and have similar capabilities but there are operational differences which is beyond the scope of this document.

The IGP provides internal connectivity within a routing domain. The size of a routing domain is determined by technical, operational and organizational considerations, such as protocol scalability and spans of control. In large networks, such as a 5G infrastructure, extending from the access to the transport core, it is very common to see multiple autonomous systems, IGPs and segmentation within the IGPs.

In an SR/MPLS environment the IGP is responsible for distributing node, prefix and label information to all nodes within the routing domain and for calculating the “Routing Information Base” (RIB) and the Label Forwarding Information Base (LFIB) on each TNE. Optionally it can provide fast convergence mechanisms and distribute traffic engineering attributes within the domain and build multiple transport planes. Both IS-IS and OSPF have facilities to create a hierarchical routing structure within a routing domain using levels and areas respectively. They also have the facility to bring external routing information into a routing domain and push its routing information into other routing domains using route redistribution.

Note: There are other techniques to achieve route re-distribution between routing domains, such as BGP and centralised SDN controllers.

Note: When looking at SR/MPLS IGP requirements there are sets of requirements for ISIS and OSPF. Operators need to follow the requirements associated with IGPs they are using within their network infrastructure.

11.4.1.1 Topology Independent Loop Free Alternative (TiLFA)

In the SR/MPLS underlay design outlined in this document, TiLFA can be used to achieve fast IGP convergence in the event of a network failure. Topology Independent Loop Free Alternative as defined “Topology Independent Fast Reroute” using Segment Routing (draft ietf-rtgwg-segment-routing-ti-lfa [163]) is an IGP capability supported by Segment Routing that provides a local repair

mechanism that achieves 100% coverage within a routing domain against links, nodes and SRLGs failures with convergence times, once failure detection has occurred, H/W dependent but typically less than 50ms.

For each destination in the network, TI-LFA pre-installs a backup forwarding entry for each protected destination ready to be activated upon detection of the failure of a link used to reach the destination. TI-LFA provides protection in the event of any one of the following: single link failure, single node failure, or single SRLG failure. In link failure mode, the destination is protected assuming the failure of the link. In node protection mode, the destination is protected assuming that the neighbor connected to the primary link has failed. In SRLG protecting mode, the destination is protected assuming that a configured set of links sharing fate with the primary link has failed (e.g. a linecard or a set of links sharing a common transmission pipe). The mechanics of the TiLFA in an SR/MPLS environment at failure is shown in Figure 11-3 and upon convergence in Figure 12-3Figure 11-4. One of the key advantages of TiLFA over some other fast convergence mechanisms, not well illustrated in the example below, is that the TiLFA backup path is the same as the post convergence path, thus minimising micro-loops that can occur as IGPs converge.

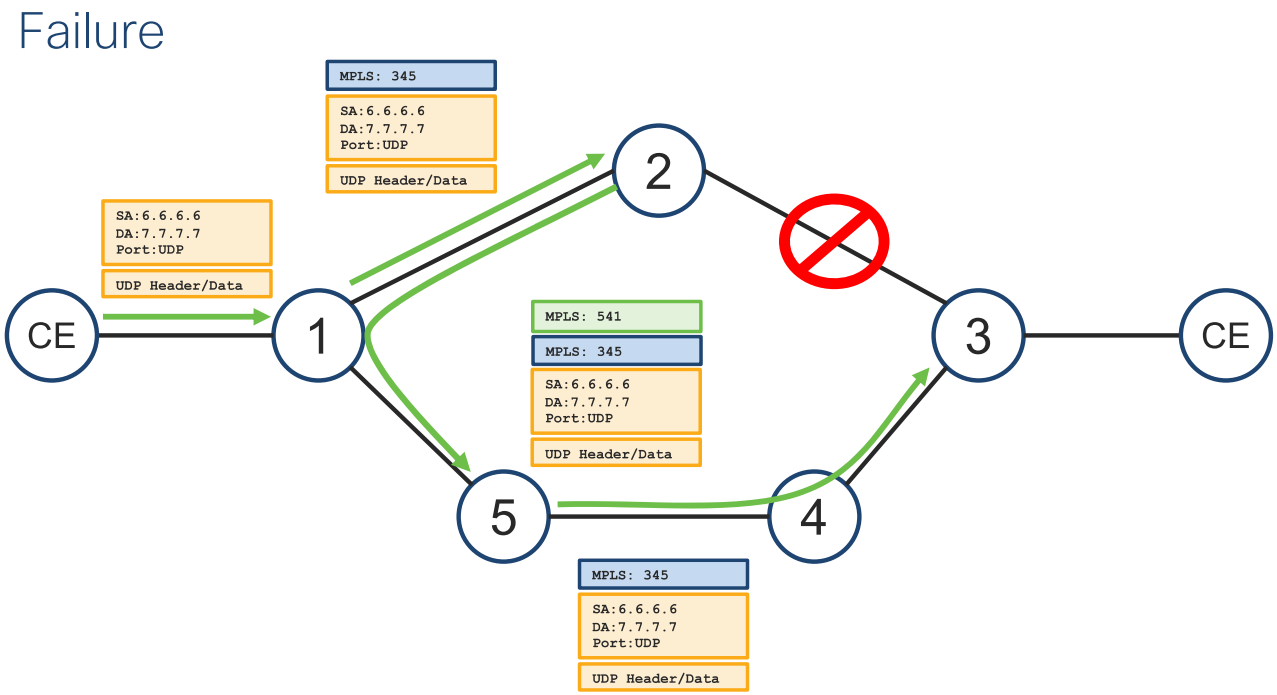


Figure 11-3 TiLFA at failure with SR/MPLS

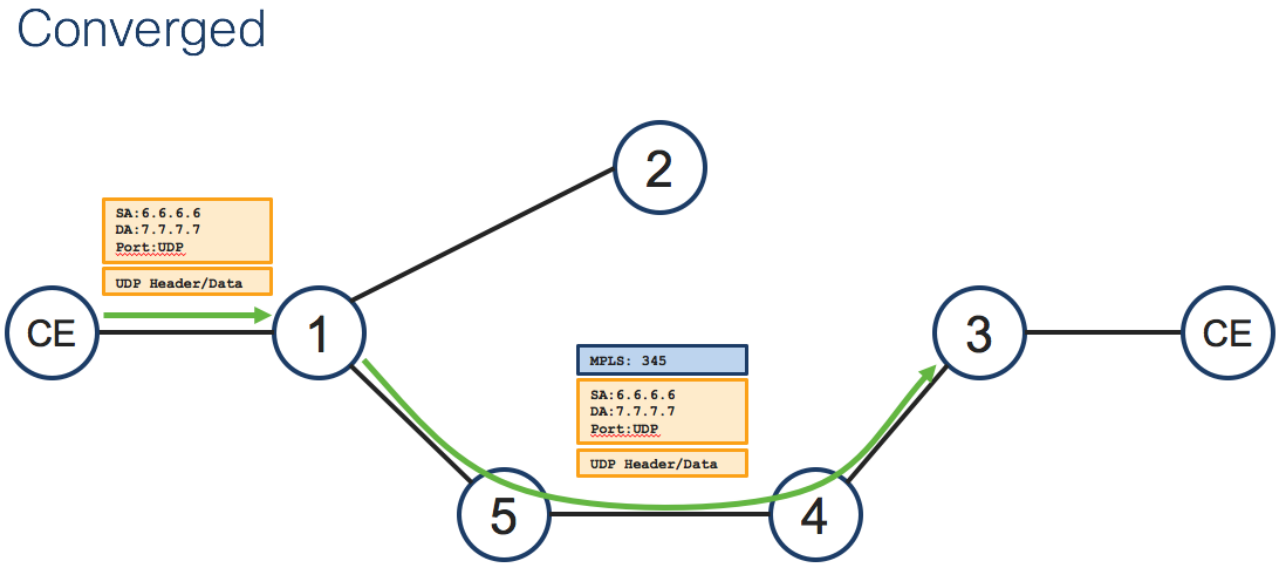


Figure 11-4 Post convergence traffic flow with SR/MPLS

11.4.1.2 IGP Flexible-Algorithm

In the SR/MPLS underlay design outlined in this document Flexible Algorithm can be used to within an IGP domain to provide an IGP based traffic engineering facility. Flexible-Algorithm is an IGP based traffic engineering technique that permits multiple forwarding tables to be created in the IGP based on operator programmed criteria. It is a simple, automatic way, in which a packet switched transport network can be traffic engineered, using only the IGP, to create different transport planes designed to meet specific operator defined criteria.

Standard ISIS and OSPF use the IGP metric on links and the “Shortest Path First” (SPF) algorithm to calculate the “best” path between an ingress and egress point in the network. For many services this approach is sufficient, however this approach cannot, for example, take into account real-time link latency, utilization, packet loss and whether an ECMP path shares links using the same underlying fibre ducts.

IGP Flex-Algo, defined in RFC 9350 [153] is an IGP based Segment Routing traffic engineering capability, that enables an operator to define their own custom algorithm, based on a wide range of variables including: link latency, packet loss, bandwidth, affinity and shared risk link groups (SRLG), to achieve a specific forwarding aim. This is illustrated in Figure 11-5.

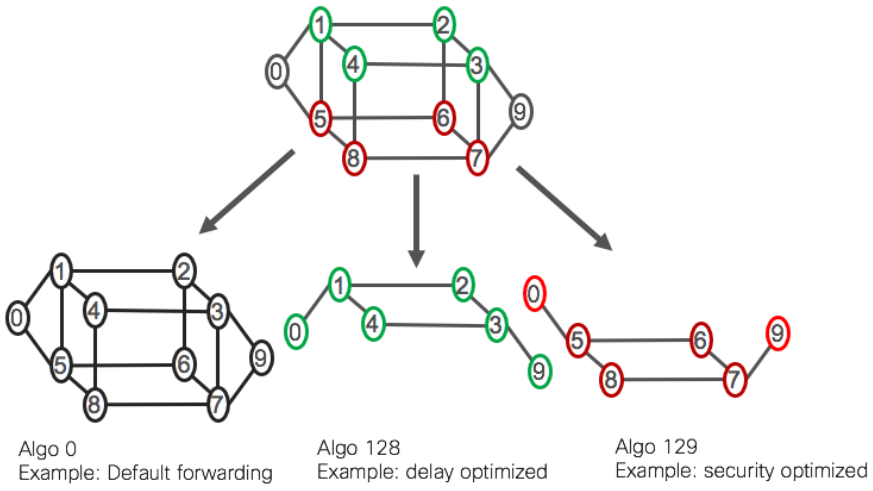


Figure 11-5 Flex-algo example with three topologies

This enables an operator, using a single IGP instance, to build multiple forwarding tables within an SR/MPLS underlay network based on different optimization criteria. The base forwarding instance is still based on IGP metrics, but an operator can build additional IGP Traffic engineered forwarding tables based, for example on lowest latency or avoiding certain links or a combination of the two.

The requirements to support flex-algorithm are all associated with the IGP requirements and contained in sections 11.4.1.3 and 11.4.1.4

11.4.1.3 IS-IS SR/MPLS base requirements

If an operator wishes to use IS-IS to support SR/MPLS with TiLFA and optionally flexible Algorithm (flex-algo) then a TNE providing the packet switching function will require:

[R15]: MUST support Routing IPv4 with ISIS, as defined in ISO/IEC 10589, RFC 1195 [39], RFC 3719, and/or Routing IPv6 with ISIS, RFC5308 [97]

[R16]: MUST support “IS-IS Extensions for Segment Routing”, RFC 8667 [143]

[D16]: SHOULD support “Signaling Maximum SID Depth (MSD) Using IS-IS”, RFC 8491 [135]

[D17]: SHOULD support “IGP Flexible Algorithm”, RFC 9350 [153]

[R17]: MUST support IS-IS TE Extensions, RFC5305 [96] and RFC7810 [122]

[D18]: SHOULD support “Topology Independent Fast Reroute using Segment Routing”, draft-ietf-rtgwg-segment-routing-ti-lfa [163]

[R18]: MUST support ISIS cryptographic extensions as defined in RFC5310 [98]

[R19]: MUST support Domain-Wide Prefix Distribution with Two-Level IS-IS as defined in RFC5302 [94].

[R20]: MUST support Three-Way Handshake for IS-IS Point-to-Point Adjacencies as defined in RFC5303 [95]

[D19]: SHOULD support purge originator identification TLV for IS-IS as defined in RFC6232 [110]

11.4.1.4 OSPF SR/MPLS basic requirements

If an operator wishes to use OSPF (v2 or v3) to support SR/MPLS with TiLFA and optionally flexible Algorithm (flex-algo) then a TNE providing the packet switching function will require:

[R21]: MUST support OSPFv2 (support for IPv4 only) as defined in RFC2328 [44], or OSPFv3 (IPv4/IPv6 support) as defined in RFC5340 [100] and RFC 5838 [107].

[R22]: MUST support "OSPFv3 Extensions for Segment Routing", RFC 8666 [142]

[D20]: SHOULD support "Signalling Maximum SID Depth (MSD) Using OSPF," RFC 8476 [134]

[D21]: SHOULD support "IGP Flexible Algorithm", RFC 9350 [153]

[R23]: MUST support "OSPF TE Extensions", as defined in RFC3630 [69] and RFC 7471 [117]

[D22]: SHOULD support Topology Independent Fast Reroute using Segment Routing "draft-ietf-rtgwg-segment-routing-ti-lfa [163]

[D23]: SHOULD support authentication/confidentiality for OSPFv3, as defined in RFC 4552 [80]

11.4.2 SR/MPLS Traffic Engineering

SR/MPLS supports two forms of SR policies or Traffic Engineering. Both solutions rely on the IGP to gather and convey topological and resource information around the network and optionally to an SR Path Computation Element (SR-PCE) via BGP-LS.

11.4.2.1 Segment Routing Traffic Engineering (SR-TE)

The SR policy consists of a list of SIDs the packet needs to traverse and is programmed into the packet on the source node. The SID list can consist of a mix of prefix and adjacency SIDs. In SR this form of TE allows a path to be:

1. Loosely source routed where some intermediate points, but not all, are specified between the ingress and egress node. Paths between these intermediate points typically use the shortest path, ECMP based routing determined by the IGP.
2. Explicitly source routed where all intermediate points and even links are specified between the ingress and egress nodes.

In all cases, topology information and live network status within a routing domain is distributed within a routing domain by the IGP with suitable extensions. In single routing domain environments path computation can be performed either by individual head-end routers or via a centralised "SR Path Computational Element" (SR-PCE). In multi-domain routing environments, then path computation generally occurs on a "centralised SR-PCE" component that has topology information

for all domains being traversed. This function can be a standalone entity or be integrated into strategically placed TNEs. This form of traffic engineering can support loose and explicitly routed paths, low-latency paths, bandwidth-guaranteed paths, and disjoint paths. Precise TE capability depends on the capabilities of the SR-PCE component.

11.4.2.2 IGP Flexible Algorithm

IGP Flexible-Algorithm is described in section 11.4.1.2. It is an IGP based traffic engineering technique that permits multiple logical topologies to be created in an IGP domain based on operator programmed criteria. When used in a single IGP routing domain the IGP calculates and maintains the path. The ingress node simply needs to address the packet to the MPLS label associated with the flexible algorithm. This removes the need for a head-end or centralised PCE path computation element and also keeps SID list contained in the label stack to a minimum.
This works in a single IGP domain but depending on area border router and autonomous system border router capabilities can be supported in multi-domain environments based on redistribution.

For flex-algo requirements for SR/MPLS, see section 11.4.1.3 and 11.4.1.4.

11.5 Scaling the MPLS infrastructure

In order to provide communication between two MPLS TNEs, end-to-end MPLS LSPs must be available between the TNEs. This implies, that unique MPLS label towards the transport network element, which possibly resides in different routing domain, must be available. In highly scaled environment the number of potential next-hops can be high, therefore care must be taken to ensure that resources, especially forwarding plane resources, are utilized in an efficient way, while providing end-to-end MPLS connectivity.

While the requirement for having unique label per remote transport network element increases the pressure on scaling, the benefit of unique label is faster detection of remote transport network element failure, resulting from unique underlay prefix/label withdrawal. For architectures, which can use prefix summarization for remote transport network element reachability, like for example SRv6 architecture discussed in Section 12, such capability is missing, since remote transport network element failure does not influence underlay summary prefix state.

11.5.1 Seamless MPLS architecture

Seamless MPLS (draft-ietf-mpls-seamless-mpls [169]) is applicable to for both LDP/RSVP and SR based control planes. Seamless MPLS is based on following architectural aspects:

- Transport network is divided into multiple smaller routing domains. Routing domains can be represented by separate BGP autonomous systems, or separate IGP domains (areas). Or, combination of BGP autonomous systems and IGP areas, where BGP autonomous systems are further divided with IGP areas.
- The size of each routing domain might vary, but must be chosen in such a way, that even the weakest transport device in given routing domain can participate in intra-domain routing and intra-domain MPLS path (SR, SR-TE, Flex-Algo, RSVP, LDP) establishment without any scaling concerns.
- Inter-domain routing and MPLS path establishment is achieved by BGP labelled unicast (BGP LU), running on top of intra-domain protocols. BGP-LU is used for both prefix distribution (PE loopbacks), as well as label distribution (labels allocated for PE loopbacks)

This concept is outlined in Figure 11-6.

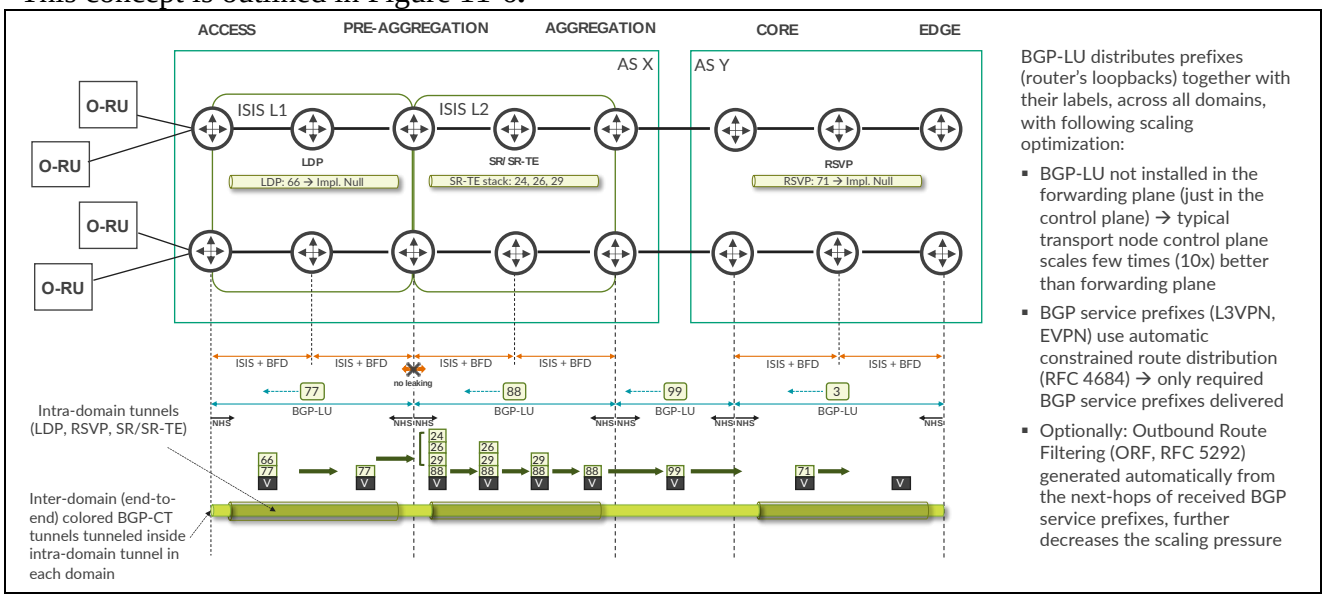


Figure 11-6 Seamless MPLS architecture

BGP-LU operation can be further optimized for enhanced scaling:

- BGP-LU prefixes (loopbacks of remote TNE) are not installed in the forwarding table (FIB) – they are just installed in the routing table (RIB), which consumes only control plane resources. On typical transport device, control plane resources scale better than forwarding plane resources
- Service NLRIs (L3VPN, EVPN) are automatically filtered on route reflectors (not shown in Figure 11-6, for diagram simplicity), and distributed only to the TNEs requiring specific service prefixes. This is achieved with constrained route distribution.
- Only these automatically filtered service NLRIs are installed in the forwarding plane, using the BGP-LU transport label of the required BGP-LU prefix (remote TNE loopback). Therefore, usage of forwarding plane resources is highly minimized.
- Further optimization of control plane resources is possible, where TNE (PE) requests from route reflector only limited set of BGP-LU prefixes. This can be achieved with outbound route filtering (ORF) of BGP-LU prefixes with ORF filter automatically generated to allow only BGP-LU prefixes required to resolve protocol next-hops of accepted (filtered with constrained route distribution mechanism) service NLRIs.

With Seamless MPLS architecture, ingress PE uses recursive next-hop resolution, where service NLRI received from remote PE is resolved over BGP protocol next-hop (BGP-LU loopback of remote PE), which in turn is resolved over local intra-domain MPLS tunnel (SR/SR-TE or legacy RSVP/LDP).

If an operator wishes to use Seamless MPLS architecture for enhanced scaling of an MPLS underlay, then the transport device will require:

[R24]: MUST support "Seamless MPLS architecture", draft-ietf-mpls-seamless-mpls [169] (note: this is an expired IETF draft but is a widely referenced document that describes the architecture)

[R25]: MUST support BGP Labelled Unicast (BGP-LU), as defined in RFC3107/RFC8277 [64][127]

[D24]: SHOULD support Constrained Route Distribution for BGP/MPLS virtual private networks, RFC4684 [85]

[R26]: MAY support Outbound Route Filter (ORF) for BGP, RFC5291/RFC5292 [92][93]

In scenarios where seamless MPLS is operating in heterogeneous environments where some domains are LDP based and some are SR based, then a Segment Routing Mapping Server can be used to provide an interworking function between the two environments.

[D25]: SHOULD support Segment Routing MPLS Interworking with LDP, RFC8661 [140]

IETF conducts work on enhancing the Seamless MPLS architecture to support inter-domain, multi-plane (multi-color) underlay transport infrastructure, as depicted in Figure 11-7. For example, one colored underlay transport plane (represented by color 'A') might be optimized for low latency, while another colored underlay transport plane (represented by color 'B') might be optimized for high-capacity, but not necessarily low-latency. More detailed discussion about colored underlay transport planes, and mapping of slice services to these colored underlay transport planes can be found in Section 18.1.

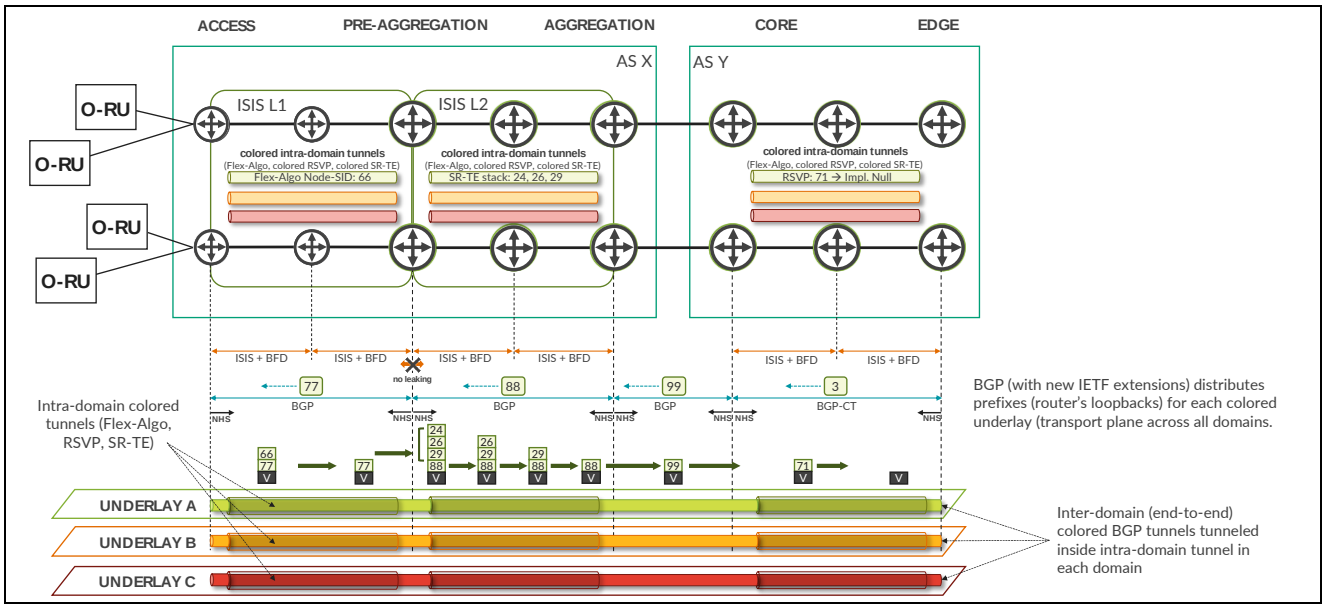


Figure 11-7 Seamless MPLS architecture with colored underlay transport planes

To this end, within the IETF Inter-Domain Routing WG there are two active experimental IETF drafts on BGP enhancements to support this enhanced architecture. As these IETF standards mature and are found to benefit packet switched xHaul, future versions of this document may describe in detail the BGP mechanisms chosen by the IETF to convey the inter-domain multi-plane/multi-color transport information.

11.5.2 Controller based network scaling architectures

In addition to the "Seamless MPLS" approach which is an infrastructure-based scaling technique, an external controller in the form of a distributed "Path Computational Element" (PCE) could be used for scaling across multiple MPLS domains. The approach aims to simplify the underlying

infrastructure by reducing interactions between domains at the infrastructure level and delegating cross domain path determination to a distributed PCE component.

Figure 11-8 shows the general architecture with a distributed PCE layer that gathers IGP topology information from their respective domain. PCE's can exchange topology information with each other using BGP-LS. In scenarios where the headend is not able to resolve the next-hop locally, it may rely on PCE to provide an End-to-End Path when requested. The headend communicates with the PCE using Path Computation Element Protocol (PCEP).-This mechanism can helps save resources on CSR's contributing to overall Xhaul Network scale.

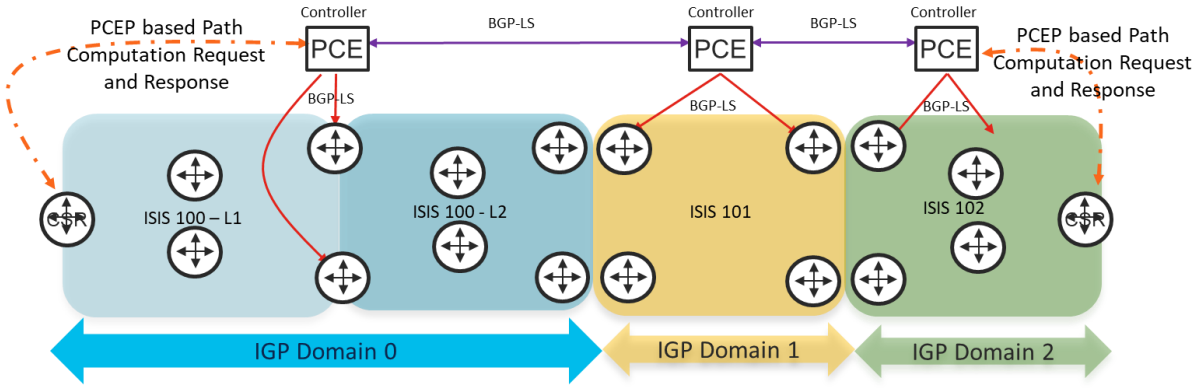


Figure 11-8 General concept of a controller-based architecture

The figure above shows network simplification and scalability by:

- Creating a multi domain topology, each with its own PCE Controller
- Removing extensive Route leaking and filtering requirements between IGP domains.
- Removing the need for the headend to populate its routing table with topology and routing information.
- Use Passive Stateful PCE Path Computation Request/Response as per RFC 8231 to create an end-to-end path.

Each domain may have its own path consideration and constraints that can independently be implemented and communicated to the PCE/PCEs. PCEs can calculate end to end path based on these constraints and communicated to the headend TNE via PCEP as shown in Figure 11-9.

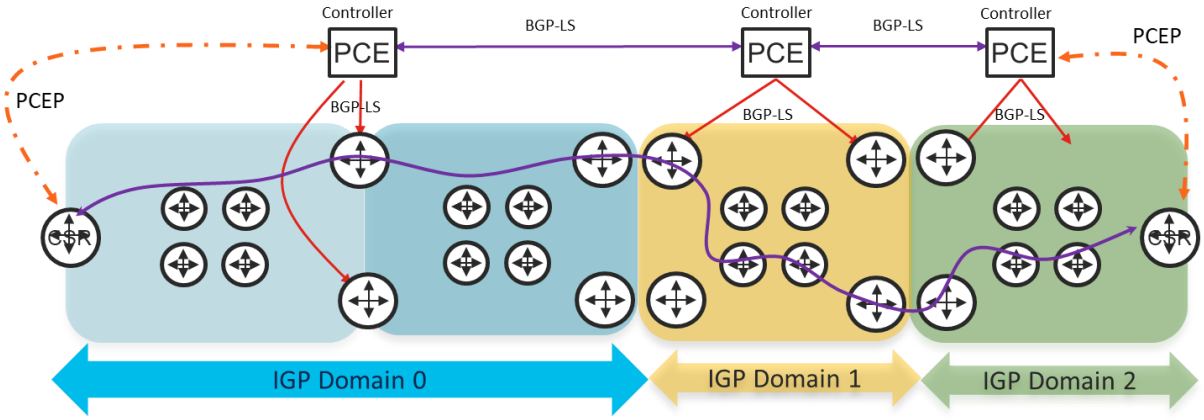


Figure 11-9 Path with Domain specific constraints using PCE based architecture

The SID list provided by the PCE to the headend is in actuality a label stack. Depending on the network size and the use-case, the SID list may exceed the platform's label stack capacity. In those cases, the SID list across domains could be further optimized by using techniques offered by one of the following two architectures.

- Using Forwarding Adjacency (FA) between domain boundaries, with associated Adjacency SID, abstracting the path between domain boundaries
- Using SR-TE policies between domain boundaries, with associated Binding SID, abstracting the path between domain boundaries

Both these approaches are optional architectures that could be used to further optimize the size of the label stack in controller-based solutions.

While the PCE is shown as a separate entity in the figure above, an operator may choose to either have PCE functionality integrated into the infrastructure router(s) or use a dedicated device or devices for PCE. As long as the PCE's in the distributed PCE layer can exchange topology information and communicate with their respective PCEP clients, the positioning of PCE functionality can be a choice an operator makes based on their respective network environment.

Following are the requirements that should be met for implementing a controller-based architecture. These additions are in addition to baseline Segment Routing and Segment Routing Traffic Engineering requirements already mentioned:

[R27]: MUST support BGP Link State (BGP-LS), RFC7752 [120]

[R28]: MUST Support BGP-LS Extensions for SR as per RFC9085 [148]

[D26]: SHOULD support "Signaling Maximum SID Depth using Border Gateway Protocol Link-State", RFC8814 [145]

[R29]: MUST support "Segment Routing Policy Architecture", RFC9256 [151]
[R30]: MUST support "Path Computation Element Protocol (PCEP)", RFC5440 [102]

[R31]: MUST support PCEP for communication between CSR and PCE as per "Path Computation Element Communication Protocol (PCEP) Extensions for Segment Routing", RFC-8664 [141]
[D27]: SHOULD support PCEP extension to support Segment Routing Policy Candidate Paths as per draft-ietf-pce-segment-routing-policy-cp [166]
[D28]: SHOULD support "PCEP Extensions for Stateful PCE", RFC8231

[D29]: SHOULD support "BGP – Link State (BGP-LS) Advertisement of IGP Traffic Engineering Performance Metric Extensions", RFC8571 [136]

11.5.2.1 Forwarding Adjacency architecture for cross domain scale

Forwarding Adjacency abstracting the path between domain boundary routers is one of the ways to optimize and scale a controller based Xhaul architecture. Forwarding Adjacency architecture is based on following architectural concepts:

- Transport network is divided into multiple smaller routing domains. Routing domains can be represented by separate BGP autonomous systems, or separate IGP domains (areas). Or, combination of BGP autonomous systems and IGP areas, where BGP autonomous systems are further divided with IGP areas.
- The size of each routing domain might vary, but must be chosen in such a way, that even the weakest transport device in given routing domain can participate in intra-domain routing and intra-domain MPLS path (SR, SR-TE, Flex-Algo, RSVP, LDP) establishment without any scaling concerns
- Head-end router (border router) of intra-domain LSP advertises the intra-domain LSP (of any kind, i.e. SR, SR-TE, Flex-Algo, RSVP, LDP) as standard link in BGP Link State (LS) topology database, with attributes similar to standard link, like for example SR Adjacency-SID, administrative group (link color, link affinity), SRLG, etc. IGP forwarding adjacency is established (RFC 4206) between domain border routers creating abstracted topology of the routing domain.
- BGP-LS (RFC 8571) is used to distribute abstracted topology. Given the fact that only border routers, and only abstracted forwarding adjacency links between these border routers are exported, the size of the topology database exported via BGP-LS is considerably smaller. BGP-LS distribution can happen in two way, depending on the overall network scale
  - BGP-LS topology is exchanged between routers in different routing domains
  - BGP-LS topology is exchanged between PCEs in distributed PCE layer

The first approach is more suitable for small to medium scale deployments. In this approach, any transport network element has the detailed topology visibility of the local routing domain, and abstracted (with Forwarding Adjacencies advertised as links in BGP-LS) visibility of the remote routing domains. This significantly reduces the topology database stored on the transport network elements, therefore, in small/medium deployments PCE might not be necessary.

The headend PE, based on the full topology database (detailed topology from local routing domain + topology abstracted with links representing forwarding adjacencies from remote routing domains), performs calculation of end-to-end SR-TE paths (local constrained shortest path first – CSPF – calculation). Each forwarding adjacency is represented in the resulting SR-TE label stack as single Adjacency-SID associated with given forwarding adjacency, regardless of the underlying MPLS transport used for given forwarding adjacency. At each border node, this Adjacency SID is mapped to the actual transport label stack (LDP, RSVP, SR, SR-TE) used for the given forwarding adjacency.

This concept is outlined in Figure 11-10**.**

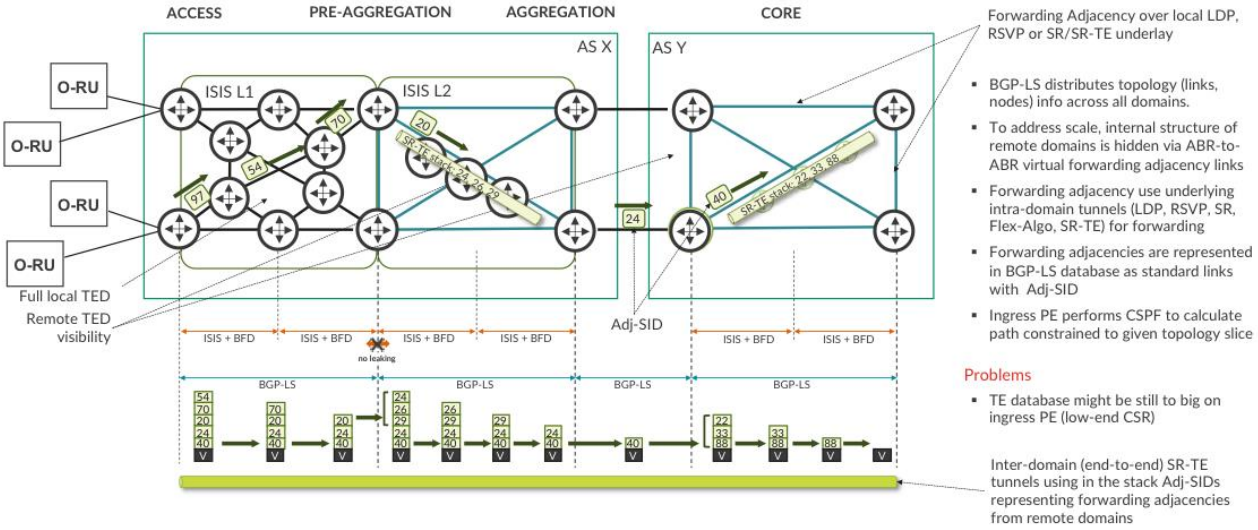


Figure 11-10 Forwarding Adjacency architecture

As further scaling optimization of this architecture, ingress PEs (typically cell site routers, with limited control plane resources required for efficient end-to-end path computation) can be off-loaded from end-to-end path computation. Distributed PCEs (Path Computation Elements) can be placed on more powerful transport devices (for example pre-aggregation or aggregation routers), or containerized/virtualized PCE function can be placed in the distributed telco cloud environment to execute end-to-end path computation duties over simplified (via forwarding adjacencies) full network topology. In this scenario, ingress PE requests from PCE on demand path computation for unresolved next-hops, as outlined in Figure 11-11**.**

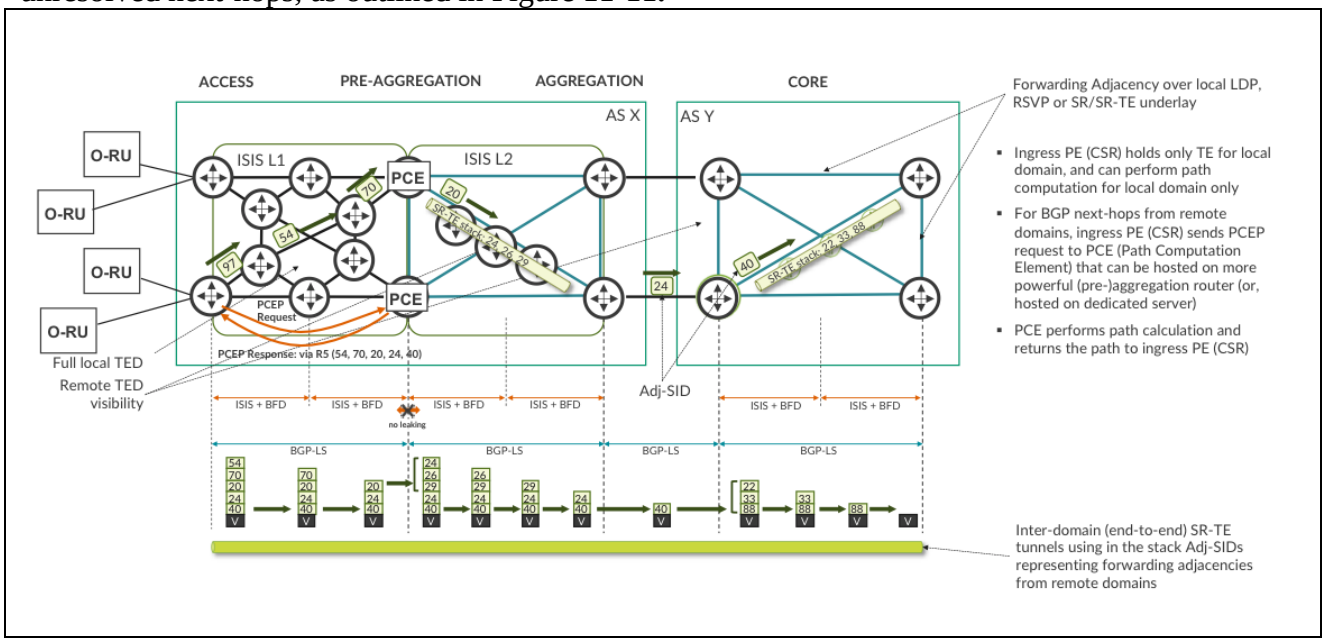


Figure 11-11 Forwarding Adjacency architecture with distributed PCE layer

If an operator wishes to use a forwarding adjacency based architecture for enhanced scaling of MPLS underlay, then the transport device will require:

[R32]: MUST support “forwarding adjacency links”, RFC4206 [128]

11.5.2.2 Using Binding SID (BSID) for cross domain scale

As mentioned earlier, a controller-based architecture make use of the PCE to provides the SID list to the headend for an end-to-end path across multiple domains. If no constraints are imposed on the traffic, the SID list would consist of Area Border Routers Prefix SIDs as next hop. Traffic within the domain will use IGP to get to the next hop while utilizing well established IGP routing mechanism such as Equal Cost Multi Path (ECMP), metric based best path etc. At the ABR, the top most SID is popped and the new next hop SID/Label is exposed, starting the routing within the next IGP are or domain. The process is repeated until traffic reaches its destination. Figure 11-12 shows such behaviour between IGP Domain 0,1 and 2.

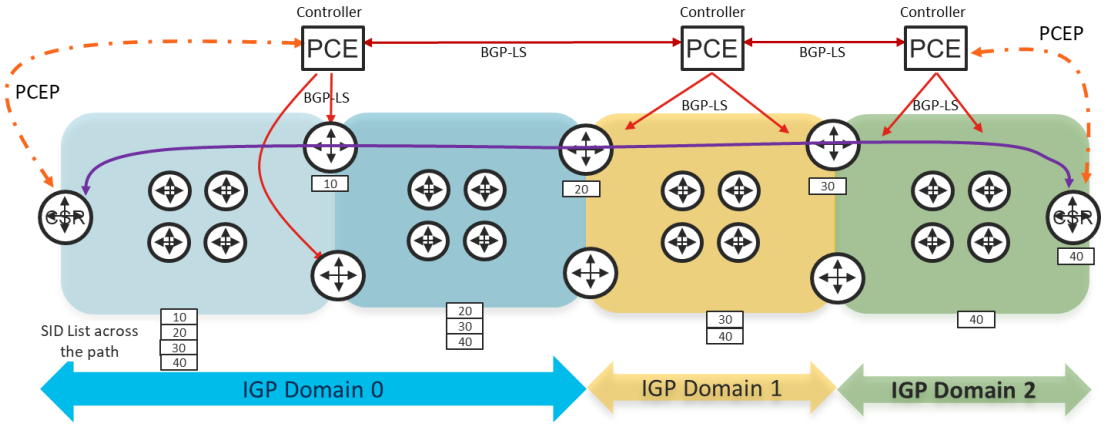


Figure 11-12 Basic SID list with no constraints

There may be scenarios where more explicit path considerations and constraints could be required, such as using low latency path or inclusion/exclusion of certain links or node. In those scenarios, rather than provide a SID list that encompasses every link/node through the path, mechanisms such as Flex Algo could be used to optimize the SID list on the head end. Figure 11-13 shows a scenario with latency and link/node constraints across the path. In this case operator may use Flex-Algo within a domain and use the PCE controller to provide SID list to the headend accordingly. Notice while the SID depth is no different from the earlier scenario Figure 11-13) and while it still only contain prefix-SIDs of ABR as next hop label, the SID's now refer to the node's Flex-Algo Prefix SID, instead of default algo SID. Due to this, the traffic behavior is also different and the next Hop SID automatically uses the desired constraints-based path.

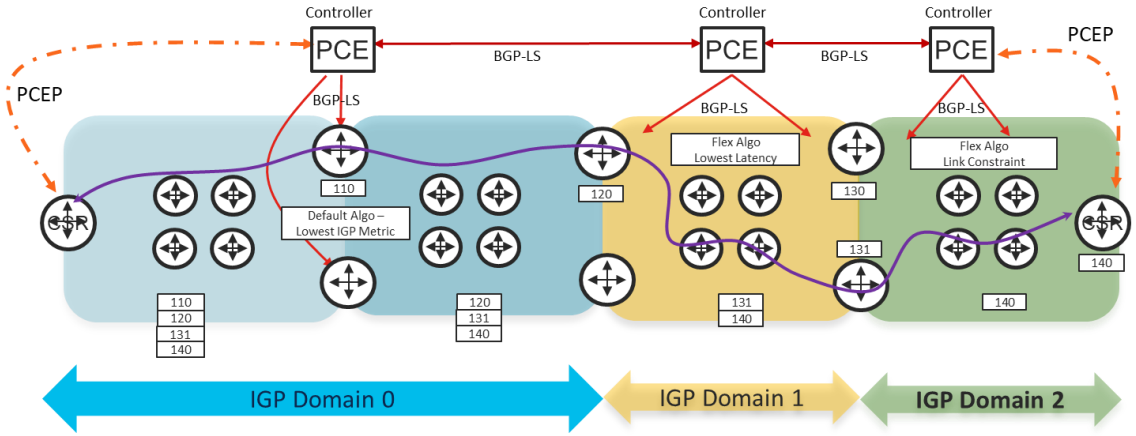


Figure 11-13 Constraint based SID list with Flex-Algorithm

In scenarios where an operator is unable or unwilling to use Flex-Algo for SID list optimization, an SRTE Binding SID (BSID) could be used. A Binding SID is a fundamental component of Segment Routing network that may be used to provide SID and network scaling. A BSID is a SID that is associated with an SRTE policy's active path and is a representation of an End-to-End candidate path. BSIDs could be used to provide additional scale by using nested SRTE policies across domains. The PCE instead of providing a full path through multiple domains, may provide the headend with a path within its own domain and a BSID corresponding to the next domain. When, exposed at the domain boundaries, the BSID gets translated into another set of SID List, providing extensibility and scalability throughout the topology. This process is shown in Figure 11-14.

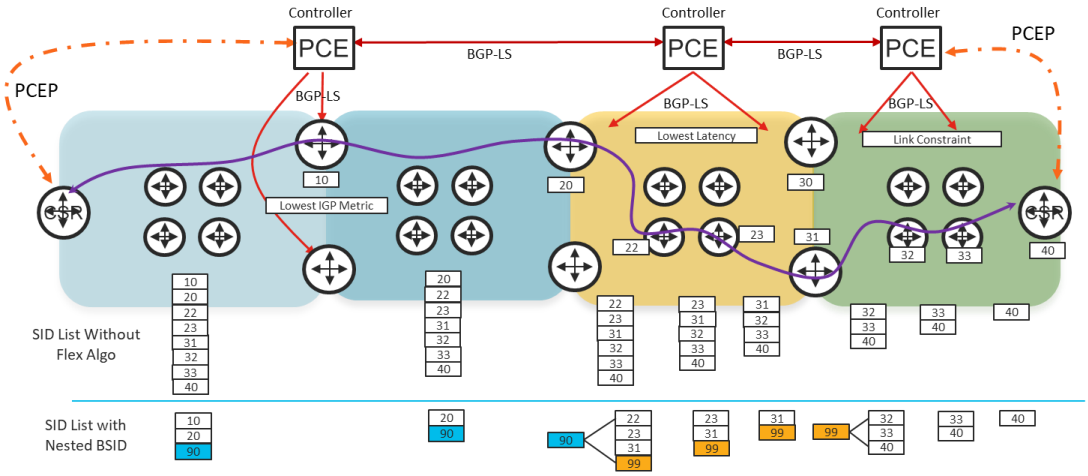


Figure 11-14 SID scaling using binding SID

Figure 11-14 compares the two scenarios where an End-to-End path is required with latency and link/nodes constraints along the path. The first scenario is where the PCE provides a full SID list to the headend for the entire path, with based path is required end to end. The second scenario uses a Binding, exposed at domain boundaries and then use the appropriate constrains based SID within the domain in question.

If an operator wishes to implement controller based architectural scaling, in addition to underlying segment-routing based transport with controllers, a device will need to support:

[R33]: SHOULD support Binding SID as per draft-ietf-pce-binding-label-sid [164]

11.6 MPLS Quality of Service

The MPLS data plane includes a 3-bit field in the MPLS header, the TC – Traffic Class (previously EXP - Experimental) bits. These bits are assignable as a marker for QoS mechanisms in transport nodes to use as a classification and marking tool. The per-hop behavior for MPLS forwarding elements can be defined based on these markings. Because of the restricted size of the TC field, the specific markings are not standardized, but are open for individual operators to define. For a full discussion of a proposed marking scheme and QoS architecture for TNEs in Xhaul, see Chapter 14.

11.7 MPLS OAM

Basic OAM tools are Ping and Traceroute. Ping is required to test liveliness to a remote MPLS underlay destination and traceroute is required to trace paths and perform hop-by-hop fault isolation to remote MPLS underlay destination. The following capabilities are required on all TNEs.

[R34]: MUST support “Detection of MPLS Data Plane Failures”, RFC8029 [123]

[R35]: MUST support LSP Ping/Traceroute for SR IGP Prefix-SID and IGP Adjacency-SID with MPLS Data Plane, as defined in RFC8287 [128]

11.8 IP/MPLS service infrastructure

For IPv4, IPv6 and Ethernet services an overlay solution is used. Please refer to section 13 for a description of overlay service recommendations for a 5G Xhaul infrastructure.

12 Packet-switched underlay network – SRv6 based

This document presents two packet switched underlay technologies, the first based on MPLS contained in section 11 and the second based on SRv6 contained in this section (section 12). An operator wishing to implement the transport architecture outlined in this document will need to select one of the two and implement the requirements outlined in the associated section. It should be noted that other transport architectures are potentially available but not covered in this revision of the document. This section outlines a packet switched underlay model based on SRv6.

SRv6 is based on the segment routing architecture as defined in RFC8402 [133]. For more information on Segment Routing see annex A. In an SRv6 infrastructure, like an MPLS environment, it is important to consider the underlay / fabric of the transport infrastructure somewhat separately from the services that run on-top of the infrastructure. The emphasis with the underlay / fabric is to provide an environment that will scale and support the services required by a 5G infrastructure. In contrast, the services infrastructure runs on-top of the underlay/fabric of the transport network and supports the different components of the 5G infrastructure (Fronthaul, Midhaul, Backhaul).

This chapter is specific to an SRv6 underlay/fabric, its data plane, control plane and how to scale it. Although there are some similarities with an SR-MPLS environment, in terms of architecture, there

are some key differences in requirements and particularly on how to scale the underlay infrastructure, which is critical in a 5G environment.

The service infrastructure and how to support 5G service requirements for both SR-MPLS and SRv6 are covered in a common services section (section 13) as the technologies used and the designs have many similarities.

12.1 SRv6 data plane

SRv6 relies on an IPv6 data plane with segments defined using SIDs contained in the IPv6 header. SRv6 SIDs are specified in "Segment Routing Header" (SRH) and standardised in RFC8754 [144]. Packet switching TNEs supporting SRv6 will need to:

[R36]: MUST support "IPv6 Segment Routing Header (SRH)" RFC8754 [144]

[R37]: MUST support "SRv6 Network Programming", RFC8986 [147]

[D30]: SHOULD support "Compressed SRv6 Segment List Encoding in SRH", draft-ietf-spring-srv6-srh-compression [170]

12.2 SRv6 control plane

The SRv6 control plane refers to routing and path control within the SRv6 underlay/fabric. This infrastructure can extend to customer devices or stop at a Provider Edge (PE) function. Its role can be summarised as learning the topology, calculation, and implementation of dynamic and explicit routes across the SRv6 infrastructure and providing rapid protection and repair mechanisms.
In a packet switched transport network consisting of a large number of routers, a mix of WAN and data centers components, the need to support traffic engineering, the underlay control plane is a typically a collection of independent routing domains that interact in a collaborate fashion.

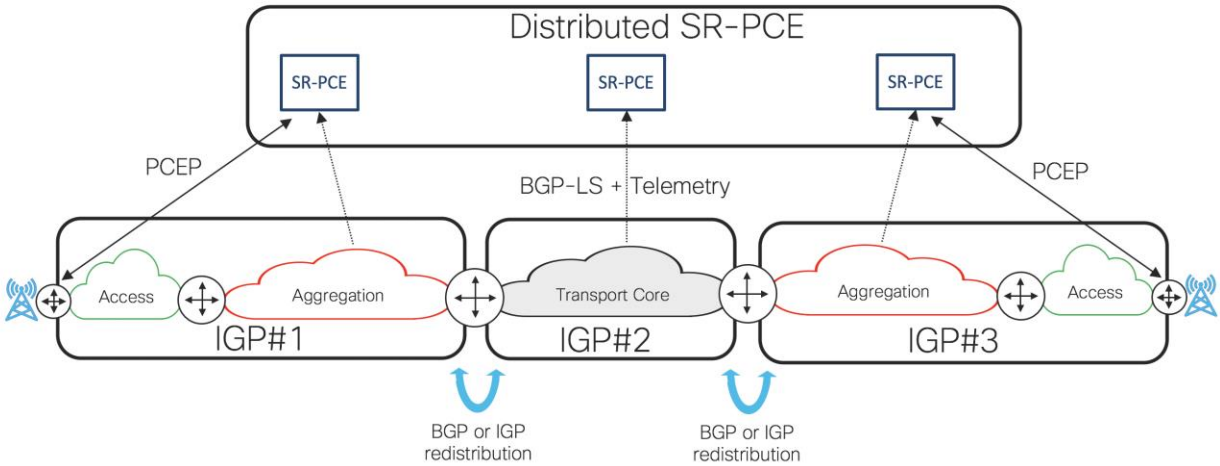


Figure 12-1 SRv6 underlay architecture

Figure 12-1 illustrates how a large packet-switched transport infrastructure might be designed and how different components interact. It should be noted that not all these components are required, and some serve the same purpose but in different ways.

1. Each routing domain has an IGP for internal connectivity.
2. Mechanisms to calculate SR policies or Traffic Engineer paths and convey them to the source nodes. Figure 12-1 shows a distributed “Path Computation Element” (PCE), with “Path Computation Element Protocol” (PCEP) running between the edge nodes and the PCE but there are other network-based mechanisms to calculate SR policies such as Flex-algo running in the IGP or head-end calculated paths.
3. Mechanisms to convey routing between IGP domains. This could be provided by summarization and redistribution, E-BGP between AS boundaries or via a multi-domain “Path Computation Element” (PCE). One of the key differences between an MPLS and IPv6 based underlay is that IPv6 enables address summarization within routing domains and also between routing domains. This makes an IPv6 based underlay network more scalable and less complex than a corresponding MPLS solution.
4. Mechanisms to convey topology and network state information from the network to the PCE and other central management elements. BGP Link State (BGP-LS) and Telemetry feeds are recommended tools to fulfil this requirement.

12.2.1 Interior Gateway Protocol (IGP) for SRv6

SRv6 requires at least one IPv6 IGP per routing domain or autonomous system. There are two IGPs that support SRv6; ISIS for IPv6 or OSPFv3. They are both link-state protocols and have similar capabilities but there are operational differences which is beyond the scope of this document.

The IGP provides internal connectivity within a routing domain. The size of a routing domain is determined by technical, operational and organizational considerations, such as protocol scalability and spans of control. In large networks, such as a 5G infrastructure, extending from the access to the transport core, it is very common to see multiple autonomous systems, IGPs and segmentation within the IGPs.

In an SRv6 environment the IGP is responsible for distributing node, prefix and SID information to all nodes within the routing domain and for calculating the forward tables between these nodes. Optionally it can provide fast convergence mechanisms and distribute traffic engineering attributes within the domain and build multiple transport planes. Both IS-IS for IPv6 and OSPFv3 have facilities to create a hierarchical routing structure within a routing domain using levels and areas respectively. They also have the facility to bring external routing information into a routing domain and push routing information into other routing domains using route redistribution.

Note: There are other techniques to achieve route re-distribution between routing domains, such as BGP and centralised SDN controllers.

Note: When looking at SRv6 IGP requirements, there are sets of requirements for ISIS for IPv6 and OSPFv3. Operators need to follow the requirements associated with IGPs they are using within their network infrastructure.

12.2.1.1 Topology Independent Loop Free Alternative (TiLFA)

In the SRv6 underlay design outlined in this document TiLFA can be used to achieve fast IGP convergence in the event of a network failure. Topology Independent Loop Free Alternative as defined “Topology Independent Fast Reroute” using Segment Routing [163] is an IGP capability supported by Segment Routing that provides a local repair mechanism that achieves 100% coverage within a routing domain against links, nodes and SRLGs failures with convergence times, once failure detection has occurred, H/W dependent but typically less than 50ms.

For each destination in the network, TI-LFA pre-installs a backup forwarding entry for each protected destination ready to be activated upon detection of the failure of a link used to reach the destination. TI-LFA provides protection in the event of any one of the following: single link failure, single node failure, or single SRLG failure. In link failure mode, the destination is protected assuming the failure of the link. In node protection mode, the destination is protected assuming that the neighbor connected to the primary link has failed. In SRLG protecting mode, the destination is protected assuming that a configured set of links sharing fate with the primary link has failed (e.g. a linecard or a set of links sharing a common transmission pipe). The mechanics of the TiLFA for SRv6 at failure is shown in Figure 12-2 and upon convergence in Figure 12-3. One of the key advantages of TiLFA over some other fast convergence mechanisms, not well illustrated in the example below, is that the TiLFA backup path is the same as the post convergence path, thus minimising micro-loops that can occur as IGPs converge.

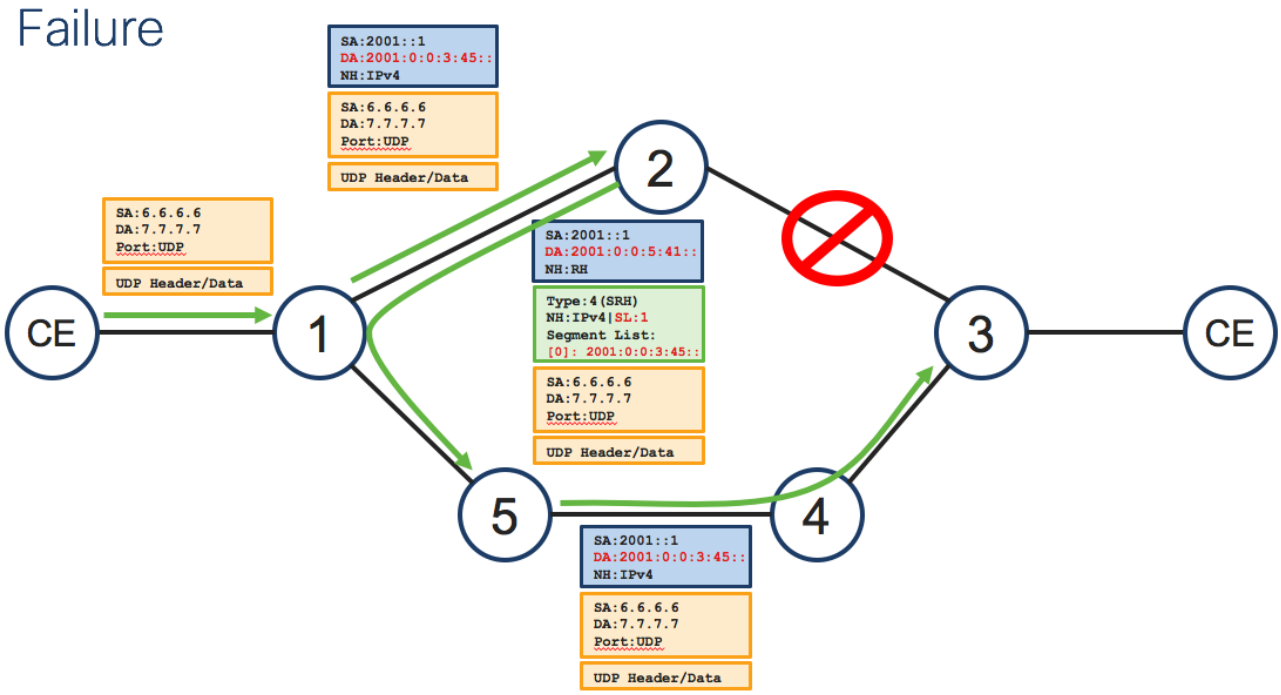


Figure 12-2 TiLFA at failure with SRv6

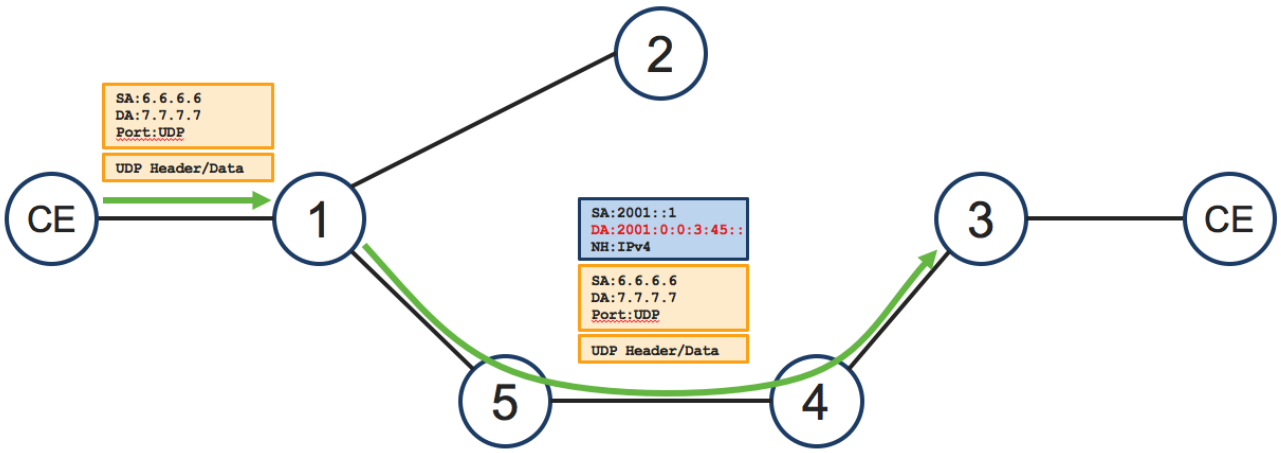


Figure 12-3 Post convergence traffic flow with SRv6

12.2.1.2 IGP Flexible-Algorithm

IGP Flexible-Algorithm is an IGP based traffic engineering technique that permits multiple forwarding tables to be created in the IGP based on operator programmed criteria. It is a simple, automatic way, in which a packet switched transport network can be traffic engineered, using only the IGP, to create different transport planes designed to meet specific operator defined criteria. Standard ISIS for IPv6 and OSPFv3 use the IGP metric on links and the "Shortest Path First" (SPF) algorithm to calculate the "best" path between an ingress and egress point in the network. For many services this approach is sufficient, however this approach cannot, for example, take into account real-time link latency, utilization, packet loss and whether an ECMP path shares links using the same underlying fibre ducts.

IGP Flex-Algo, defined in RFC 9350 [153] is an IGP based Segment Routing traffic engineering capability, that enables an operator to define their own custom algorithm, based on a wide range of variables including: link latency, packet loss, bandwidth, affinity and shared risk link groups (SRLG), to achieve a specific forwarding aim. This is illustrated in Figure 12-4.

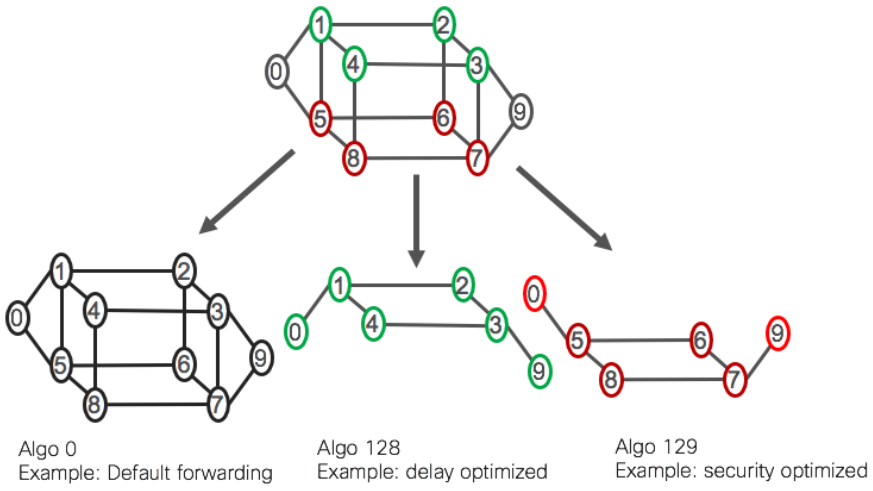


Figure 12-4 Flex-algo example with three topologies

This enables an operator, using a single IGP instance, to build multiple forwarding tables within an SRv6 underlay network based on different optimization criteria. The default forwarding instance uses shortest path forwarding based on IGP metrics, but an operator can build additional IGP Traffic Engineered forwarding tables based, for example on lowest latency or avoiding certain links or a combination of the two.

The requirements to support flex-algorithm are all associated with the IGP requirements and contained in sections 12.2.1.3 and 12.2.1.4.

12.2.1.3 ISIS for IPv6 base requirements

If an operator wishes to use ISIS for IPv6 with TiLFA and optionally Flexible Algorithm (flex-algo) with SRv6 then the TNE will require:

[R38]: MUST support routing IPv6 with ISIS as defined in RFC5308 [97].

[R39]: MUST support RFC 9352: “IS-IS Extensions to Support Routing over IPv6 Dataplane” [155]

[R40]: MUST support “Topology Independent Fast Reroute using Segment Routing”, draft-ietf-rtgwg-segment-routing-ti-lfa [163]

[R41]: MUST support IS-IS TE Metric Extensions, as defined in RFC5305 [96] and RFC7810 [122].

[D31]: SHOULD support RFC 9350: “IGP Flexible Algorithm”, [153]

12.2.1.4 OSPFv3 basic requirements

If an operator wishes to use OSPFv3 with TiLFA and optionally flexible Algorithm (flex-algo) with SRv6 then the TNE will require:

[R42]: MUST support OSPFv3, RFC5340 [100].

[R43]: MUST support “OSPFv3 Extensions for SRv6”, draft-ietf-lsr-ospfv3-srv6-extensions[167]
[R44]: MUST support “Topology Independent Fast Reroute using Segment Routing”, draft-ietf-rtgwg-segment-routing-ti-lfa [163].
[R45]: MUST support OSPFv3 TE Metric Extensions, as defined in RFC5329 [99].

[D32]: SHOULD support RFC 9350: “IGP Flexible Algorithm” [153]

12.2.2 SRv6 Traffic Engineering

SRv6 supports two forms of SR policies or SR Traffic Engineering. Both solutions rely on the IGP to gather and convey topological and resource information around the network and optionally to a Path Computation Element (PCE) optimized for SR via BGP-LS.

12.2.2.1 Segment Routing Traffic Engineering (SR-TE)

The SR policy consists of a list of SIDs the packet needs to traverse and is programmed into the packet on the source node. The SID list can consist of a mix of prefix and adjacency SIDs. In SR this form of TE allows a path to be:

1. Loosely source routed where some intermediate points, but not all, are specified between the ingress and egress node. Paths between these intermediate points typically use the shortest path, ECMP based routing determined by the IGP.
2. Explicitly source routed where all intermediate points and even links are specified between the ingress and egress nodes.

In all cases, topology information and live network status within a routing domain is distributed within a routing domain by the IGP with suitable extensions. In single routing domain environments path computation can be performed either by individual head-end routers or via a centralised “SR Path Computational Element” (SR-PCE). In multi-domain routing environments, then path computation generally occurs on a “centralised SR-PCE” component that has topology information for all domains being traversed. This function can be a standalone entity or be integrated into strategically placed TNEs. This form of traffic engineering can support loose and explicitly routed paths, low-latency paths, bandwidth-guaranteed paths, and disjoint paths. Precise TE capability depends on the capabilities of the SR-PCE component.

12.2.2.2 IGP Flexible Algorithm

IGP Flexible-Algorithm is described in section 12.2.1.2. It is an IGP based traffic engineering technique that permits multiple logical topologies to be created in an IGP domain based on operator programmed criteria. When used in a single IGP routing domain the IGP calculates and maintains the path. The ingress node simply needs to address the packet to the “locator-block” associated with the flexible algorithm. This removes the need for a head-end or centralised PCE path computation element and also keeps SID list to a minimum.

This works in a single IGP domain but depending on area border router and autonomous system border router capabilities can be supported in multi-domain environments based on redistribution. For flex-algo requirements in an SRv6 environment see section 12.2.1.3 and 12.2.1.4.

12.2.3 Inter-domain connectivity

In a large packet-switched transport network the underlay infrastructure will need to be sub-divided into multiple routing domains. These routing domains can be within an autonomous system, at the IGP level, or between autonomous systems. One motivation is IGP scalability, but other reasons include fault isolation between domains and organizational, for example different groups run the DC and WAN infrastructures. However, even though the infrastructure is divided, connectivity between routers in different routing domains is still required to build end to end services, therefore the underlay infrastructure needs mechanisms to enable inter-domain connectivity and services to be built inter-domain. In an SRv6 environment, potential mechanisms for interchanging routing between domains include:

1. IGP based hierarchy
2. Inter-domain IGP redistribution
3. Inter AS BGP
4. Controller based

12.2.3.1 IGP based hierarchy

Both ISIS for IPv6 and OSPFv3 have the concept of routing hierarchy in their basic designs. In an SRv6 environment, route summarization or injection of a default route between IS-IS levels or OSPFv3 areas is sufficient to enable EVPN or L3VPNs services to be built over an SRv6 hierarchical IGP infrastructure.

12.2.3.2 Inter-domain IGP redistribution

One or more routers are located on the boundary between two or more IGP routing domains. These routers run multiple IGP instances, so have routing awareness of all the IGPs they participate in. Routing information is mutually re-distributed between the IGP protocols allowing reachability between the domains. This redistribution can be in the form of full routing, summarised routes or even default routes. The choice is dependent on the level of awareness required between domains.

See Figure 12-5 for more detail. IGP route re-distribution in both IS-IS for IPv6 and OSPFv3 is a common capability.

12.2.3.3 Inter-AS BGP based routing

A router in each autonomous system exchanges routing information using BGP-4 supporting the IPv6 address family. It should be noted that this scheme provides better separation than inter-domain IGP redistribution but typically requires separate routers in each autonomous system. See Figure 12-6 for more detail.

12.2.3.4 Controller based path calculation

In this scenario no routing information is exchanged between routing domains at the device level. Instead, each routing domain provides its topology to an SR Path Computation Element (SR-PCE), which is a central SDN control element responsible for determining and finding routes to a destination node on behalf of a source node. Although the SR-PCE is logically a central component, for scale and resiliency reasons it can consist of multiple individual entities distributed around and

serving different parts of the network. In this case the source TNE requires very limited visibility of the overall network, as end to end cross-domain path computation is delegated to the SR-PCE.

A SR-PCE function in an SRv6 underlay transport network has three main functions.

1. Gather data about the topology of the overall SRv6 network underlay. Several options exist:
    a. The SR-PCE participates in the IGP domains and directly collects each domains IGP link-state database.
    b. Use BGP Link-State (BGP-LS) with appropriate extensions for SRv6 to collect the IGP's link-state database from each routing domain. This is done through one or more BGP session between the SR-PCE and a TNE in each routing domain participating in the domain's IGP. The TNEs retrieve information from the IGP LSDB and distribute it to the controller using the BGP link-state address family.
2. Communications between the PCE and PCC (path computation client) which is the headend TNE. In this document it is assumed to PCEP (path computation Element Protocol) is used as the communication mechanism between the source TNE and the SR-PCE.
3. Path computation. The PCE uses information gathered from the various domains to compute an SR policy consisting of a SID list which packets using the SR policy will traverse. This SID list can be a loosely sourced routed or explicitly source routed.

12.3 Scaling an SRv6 underlay infrastructure

Scaling is one of the key challenges in building a packet switched infrastructure supporting 5G services. Communications and support of L2 and L3 VPN services between two SRv6 TNEs relies entirely on IP routing mechanisms on the underlay transport fabric. In SRv6 this is based on longest IP prefix matching and forwarding, so TNEs can support routing and forwarding based on full IPv6 host routes, summarized IPv6 routes or default routing or a combination, derived from dynamic routing protocols or statically defined. Further, different TNEs in the end to end path can use different levels of route summarization appropriate to their position in the network and their computation, memory and NPU/ASIC resources.

12.3.1 Route summarization and redistribution

Summarization within ISIS for IPv6 and OSPFv3 is part of the base protocol capability and can occur between levels in IS-IS for IPv6 and between areas in OSPFv3 using "Area Border Routers" (ABRs). Summarization is also a common capability when redistributing routes between different routing domains and is also supported between Autonomous Systems using BGP-4 between "Autonomous System Border Routers" (ASBRs). This is illustrated in Figure 12-5 and Figure 12-6 and shows how PE1-4's routing tables can be controlled through route summarization.

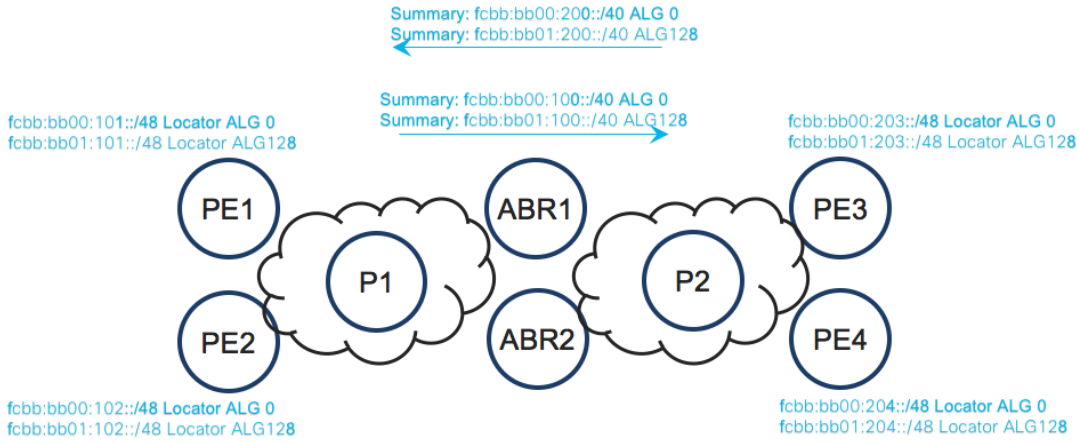


Figure 12-5 Summarization / redistribution using IGPs

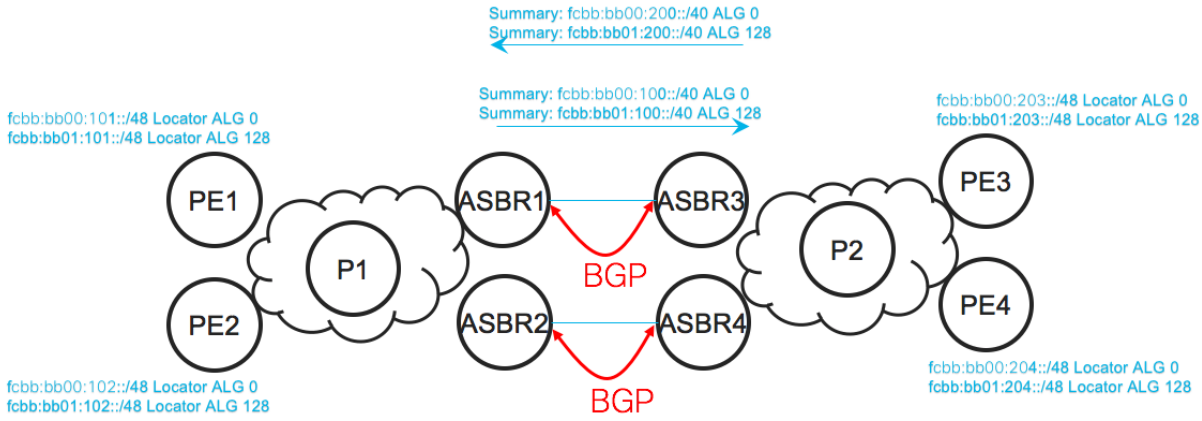


Figure 12-6 Summarization / redistribution using BGP-4

This approach offers considerable flexibility, as operators can choose the level of summarization, ranging from no summarization to default routing between domains depending on the size of the network and level of awareness required between domains. It can also offer inter-domain traffic engineering support based on flexible algorithm within an IGP domain. However, there are some considerations and limitations:

1. Address planning is critical and needs to occur upfront.
2. Route summarization might result in sub-optimal routing in certain designs.
3. Some traffic engineering scenarios cannot be achieved with route summarization alone.
4. A source TNE cannot rely on a host route withdrawal to detect a remote TNE failure covered by a summary route. In this situation over forms of device or service verification are required.

[D33]: ABRs SHOULD have flex-algorithm aware summarization/redistribution function. This is a local behavior.

To use BGP-4 to transmit underlay SRv6 routing between Autonomous Systems requires:

[R46]: BGP with IPv6 multi-protocol extensions (RFC1771, RFC2283, RFC2545) [40][45][48]

12.3.2 Controller based scaling

Route summarization and redistribution is simple and allows massive scale and allows end to end traffic engineering based on flex-algo. However, some traffic engineering scenarios cannot be achieved, such as path diversity and bandwidth optimization, because summarization hides topological details.

To achieve more complex TE solutions, end-to-end visibility of the network is required, and a controller-based solution based on an SR-PCE is used. The SR-PCE receives detailed topology information via BGP-LS protocol from each domain. The SR-PCE then has a full picture of end to end reachability. When an ingress PE need to establish an end to end path it requests information from SR-PCE via PCEP protocol. SR-PCE responds with stack of SIDs to establish path to egress PE.
Figure 12-7 illustrates an SR-PCE based solution where PE1 needs to establish shortest path to PE3.

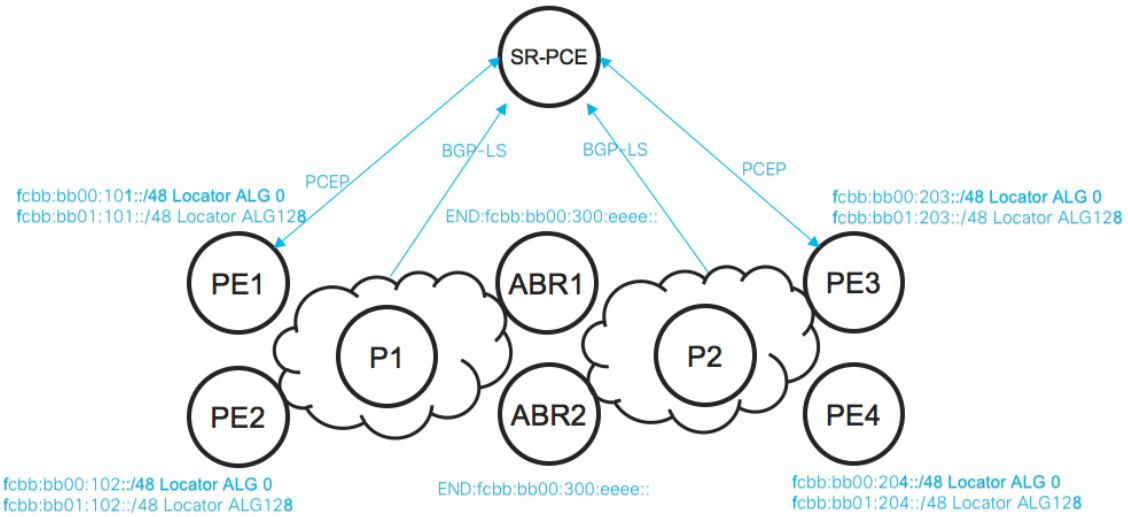


Figure 12-7 Path computation based on SR-PCE

1. SR-PCE gathers topology from each domain via BGP-LS.
2. PE1 sends a PCEP request to SR-PCE
3. PCE will respond with list of SIDs fcbb:bb00:300, fcbb:bb00:203:ffff::.

For more complex path establishment SR-PCE will provide more complex SID list.

[R47]: PCE and at least one TNE in each domain MUST support BGP Link State (BGP-LS), RFC7752 [120]

[R48]: SR-PCE and at least one TNE in each domain MUST support "BGP Link State extensions for IPv6 Segment Routing (SRv6)", draft-ietf-idr-bgpls-srv6-ext [168]

[D34]: SR-PCE and at least one TNE in each domain SHOULD support extensions to advertise Flexible Algorithm Definition as part of the topology, RFC 9351 [154]

[R49]: SR-PCE and TNEs providing PE functionality MUST support Path Computation Element Protocol (PCEP), as described in RFC5440 [102]

[R50]: SR-PCE and TNEs providing PE functionality MUST support PCEP protocol client functionality with SRv6 extensions, draft-ietf-pce-segment-routing-IPv6 [165]

[D35]: SR-PCE SHOULD support "IGP Flexible Algorithm be able to compute paths based on different flexible algorithms

12.3.3 SRv6 scaling conclusion

To build an underlay transport network that is highly scalable and supports traffic engineering the simplest way is to uses a multi-domain IGP that supports flex-algo and use address summarization. This design might not be sufficient for all application requirements. In these instances, a PCE based solution could be deployed in tandem for TE use cases that require additional network visibility.

12.4 IPv6 Quality of Service

The IPv6 packet includes a 6-bit field in the IPv6 header, the DiffServ Code Point (DSCP). These bits are assignable as a marker for QoS mechanisms in transport nodes to use as a classification and marking tool. The per-hop behavior for IPv6 forwarding elements can be defined based on these markings. For a full discussion of a proposed marking scheme and QoS architecture for TNEs in Xhaul, section 14.

12.5 SRv6 OAM

SRv6 OAM functionality is required to understand the status of the network. SRv6 OAM is described in "Operations, Administration, and Maintenance (OAM) in Segment Routing over IPv6 (SRv6)" RFC 9259 [152]

12.5.1 Ping / Traceroute to a remote IPv6 network address

Ping is required to test liveliness to a remote SRv6 underlay address and traceroute is required to trace paths and perform hop-by-hop fault isolation to remote SRv6 underlay address. The following capabilities are required on all TNEs.

[R51]: TNEs MUST support ICMPv6 (RFC4443) [79]

[R52]: TNEs MUST support IPv6 ping to query liveliness of a remote IPv6 address along the shortest path for default SPF algorithm.

[R53]: TNEs MUST support IPv6 ping to query liveliness of a remote IPv6 address along a path calculated by a flex-algorithm.

[R54]: TNEs MUST support IPv6 ping to query liveliness of a remote IPv6 address along a path designated by a list of SIDs.

[R55]: TNEs MUST support IPv6 traceroute to trace the path to a remote IPv6 address along the shortest path for default SPF algorithm

[R56]: TNEs MUST support IPv6 traceroute to trace the path to a remote IPv6 address along a path calculated by a flexible algorithm.

[R57]: TNEs MUST support IPv6 traceroute to trace the path to a remote IPv6 address along a path designated by a list of SIDs.

12.5.2 Ping / Traceroute to remote SID functions

Ping is required to test liveliness to a remote SID and traceroute is required to trace paths and perform hop-by-hop fault isolation to a remote SRv6 SID. The following capabilities are required on all Transport Nodes.

[R58]: TNEs MUST support IPv6 ping to query liveliness of a remote EVPN and L3VPN services along the shortest path for default SPF algorithm.

[R59]: TNEs MUST support IPv6 ping to query liveliness of a remote EVPN and L3VPN services along a path calculated by a flex-algorithm.

[R60]: TNEs MUST support IPv6 ping to query liveliness of a remote EVPN and L3VPN service along a path designated by a list of SIDs.

[R61]: TNEs MUST support IPv6 traceroute to trace the path to a remote EVPN and L3VPN service along the shortest path for default algorithm

[R62]: TNEs MUST support IPv6 traceroute to trace the path to a remote EVPN and L3VPN service along a path calculated by a flexible algorithm.

[R63]: TNEs MUST support IPv6 traceroute to trace the path to a remote EVPN and L3VPN service along a path designated by a list of SIDs.

12.6 SRv6 Service infrastructure

An SRv6 transport underlay supports basic IPv6 services natively at the control and data plane. For Ethernet, IPv4 and IPv6 VPN services an overlay control plane infrastructure is used to convey VPN connectivity information. At the data plane the P devices can utilise a standard IPv6 data plane with the PE devices requiring an IPv6 data plane with SRv6 VPN network programming awareness [147]. Please refer to section 13 for a description of overlay service recommendations for a 5G Xhaul infrastructure.

13 Packet-switched Xhaul services Infrastructure

To support an Xhaul environment requires the packet switched network to support L2 and L3 services. Both an MPLS and IPv6 packet switched underlays use EVPN to support L2 and MP-BGP L3VPNs.

13.1 MP-BGP design

Both EVPN and MP-BGP L3VPNs use MP-BGP, with appropriate address-family support for EVPN and L3VPN, to convey service connectivity information between Provider Edge (PE) equipment.

Typically, the MP-BGP infrastructure for L2 and L3 services uses a Route Reflector design rather than an I-BGP mesh between all PEs. In large networks, operators will normally implement a hierarchical Route Reflector (RR) design for Multi-Protocol BGP (MP–BGP) peering. A pair of Route Reflectors (RR) could be used at every domain to provide scalability and extensibility.

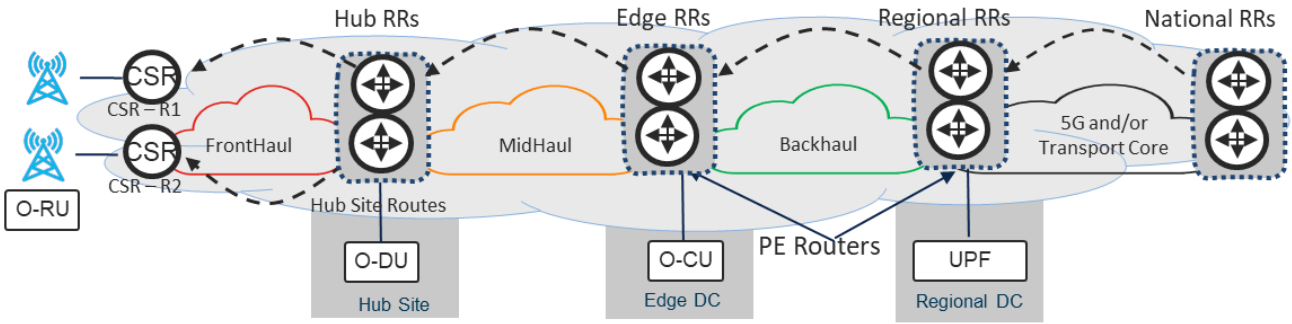


Figure 13-1 Hierarchical Route Reflector Design for Multiple Protocol BGP

Figure 13-1 shows the Hierarchical Route Reflector Design utilizing RR pairs in each network domain, all peering with National Route Reflector core.

13.2 Ethernet services

Ethernet services will be provided by EVPN over either MPLS or IPv6 depending on the underlay network employed by the operator. More details on EVPN are provided in annex B.

EVPN VPWS can be used as transport service for Open Fronthaul (eCPRI), as well a transport service for Radio over Ethernet (RoE) as outlined in following Figure 13-2 and Figure 13-3.

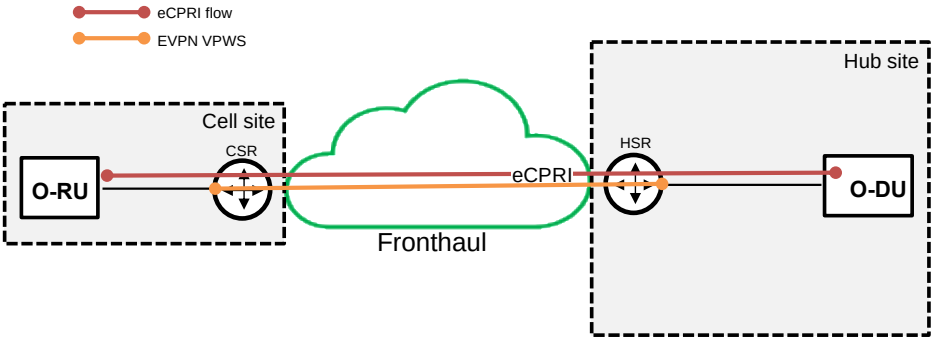


Figure 13-2 EVPN VPWS for eCPRI

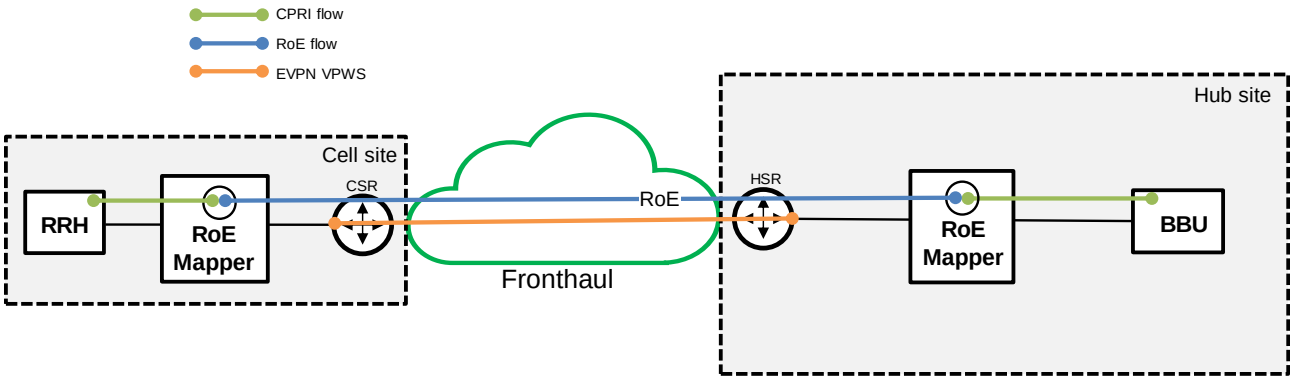


Figure 13-3 EVPN VPWS for RoE

To support EVPN VPWS for eCPRI or RoE, the transport device will need to support:

[D36]: SHOULD support "Flow-Aware Transport of Pseudowires over an MPLS Packet Switched Network", RFC 6391 [112]

[R64]: MUST support "Virtual Private Wire Service Support in Ethernet VPN", RFC 8214 [125]

[D37]: SHOULD support "Extensions to BGP-Signaled Pseudowires to Support Flow-Aware Transport Labels", Internet Engineering Task Force, RFC 8395 [132]

[R65]: If using SRv6 as the underlay technology TNEs providing PE functionality MUST support "BGP Overlay Services Based on Segment Routing over IPv6 (SRv6)", Internet Engineering Task Force, RFC 9252 [150]

13.2.1 Ethernet services redundancy

EVPN VPWS can provide redundancy for O-RU to O-DU (Open Fronthaul) interface to react to the transport network errors that might occur.

13.2.1.1 Ethernet services redundancy – Option 1

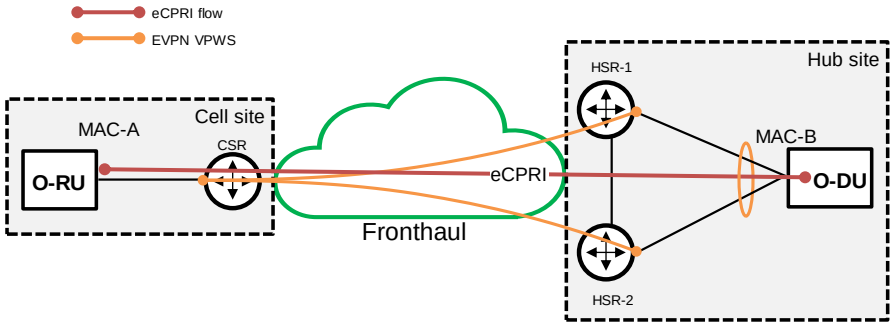


Figure 13-4 EVPN VPWS redundancy option 1

This option assumes that O-DU supports Link Aggregation Group (LAG, also called Ethernet bundling) and terminates the eCPRI stream on a MAC address associated with the Ethernet bundle (e.g. MAC-B in the diagram). Therefore, eCPRI stream can arrive on any physical interface (via HSR-1 or via HSR-2) to O-DU, and O-DU might sent eCPRI stream, again, on any interface (via HSR-1 or via HSR-2) (see Figure 13-4).

To facilitate fast failover in case of HSR⇔O-DU link failure, following architectural design choices are recommended:

- O-DU uplinks towards HSR pair should be bundled on O-DU to create Link Aggregation Group (LAG), often called an “Ethernet bundle”.
- EVPN VPWS service should be terminated on HSR pair as ‘multi-homed all-active’ service

When O-RU generates eCPRI stream towards O-DU, this stream can be sent from CSR via EVPN VPWS leg towards HSR-1, or via EVPN VPWS leg towards HSR-2. It is CSR implementation choice, towards which HSR eCPRI stream will be sent. In case multiple O-RUs are connected to single CSR, and multiple eCPRI streams are forwarded from CSR towards O-DU via HSR pair, each eCPRI stream might take differen path (i.e. via HSR-1 or HSR-2).

When for example HSR-1⇔O-DU link failure happens, following EVPN failure detection machinery, the EVPN VPWS leg between CSR and HSR-1 is disabled, and eCPRI stream flows now over HSR-2. Since EVPN VPWS termination on HSR-2 was originally ‘active’ (multi-homed all-active service) no special forwarding plane reprogramming is needed on HSR-2, allowing quick delivery of rerouted eCPRI streams towards O-DU, as outlined in Figure 13-5.

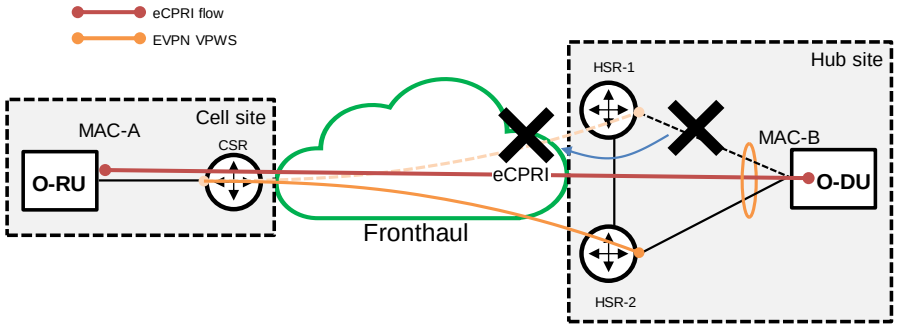


Figure 13-5 EVPN VPWS Redundancy option 1 – failure event

13.2.1.2 Ethernet services redundancy – Option 2

Option 2 is similar to option 1, with the difference that O-DU expects to receive eCPRI stream on a particular interface (for example, on the interface from HSR-1), as outlined in Figure 13-6

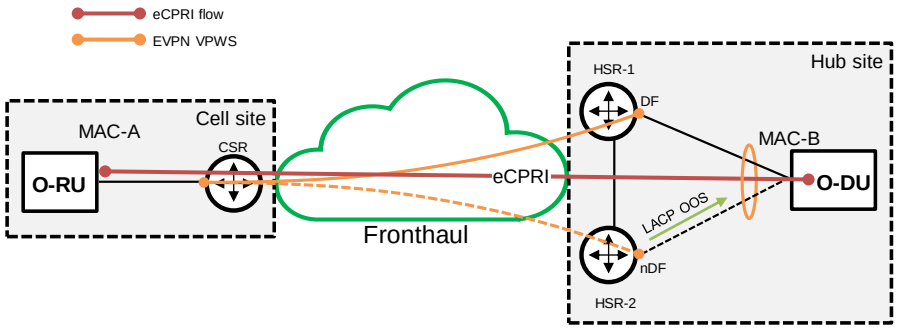


Figure 13-6 EVPN VPWS Redundancy Option 2

For this option, following architectural design choices are recommended:

- O-DU uplinks towards HSR pair should be bundled on O-DU to create Link Aggregation Group (LAG), often called an “Ethernet bundle”.
- EVPN VPWS service should be terminated on HSR pair as ‘multi-homed single-active’ service
- Deterministic Designated Forwarder (DF) election should be used on HSR pair, to ensure that one of the HSR is deterministically elected as ‘active’ forwarder (DF), and second HSR is deterministically elected as ‘standby’/’passive’ forwarder (non-DF)
- Per physical port (rather than default per-VLAN) DF election should be used on HSR pair
- HSR should signal the non-DF status to O-DU via OAM (for example LACP – Link Aggregation Control Protocol – Out of Sync signalling)

In addition to requirements already mentioned in the main section, to support this option following requirements must be supported on HSR:

[R66]: MUST support “Preference-based EVPN DF Election", draft-ietf-bess-evpn-pref-df [160]

[R67]: MUST support "EVPN multi-homing port-active load-balancing", draft-ietf-bess-evpn-mh-pa [159]

As outlined in Figure 13-7, when HSR-1⇔O-DU link failure happens, following EVPN failure detection machinery, the EVPN VPWS leg between CSR and HSR-1 is disabled, HSR-2 becomes DF and EVPN VPWS leg between CSR and HSR-2 is enabled. Additionally, LACP Out-of-Sync state from HSR-2 to O-DU is cleared, allowing forwarding over the HSR-2⇔O-DU link. eCPRI stream flows now over HSR-2. As opposed to Option 1, since EVPN VPWS termination on HSR-2 was originally ‘standby’ (multi-homed single-active service), DF election, LACP OOS clearance and forwarding plane reprogramming is needed on HSR-2, thus failover is longer than in case of Option 1.

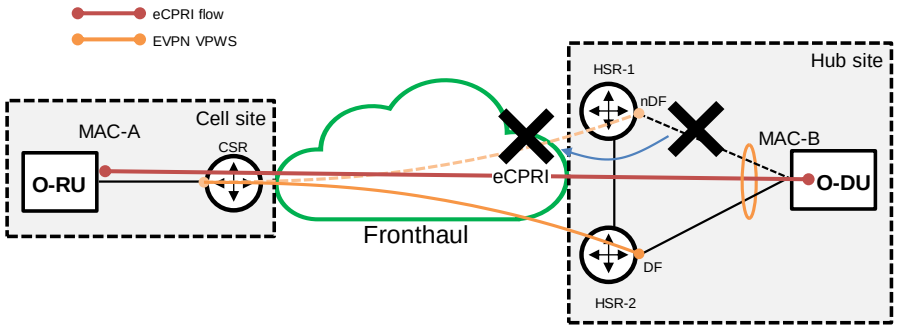


Figure 13-7 EVPN VPWS Redundancy Option 2 -failure event

13.2.1.3 Ethernet services redundancy – Option 3

Option 3 is further modification of option 2. It is suitable for O-DUs that do no support LAG bundling, but have the capability to terminate eCPRI stream on some internal virtual MAC, allowing reception of eCPRI stream over any uplink.

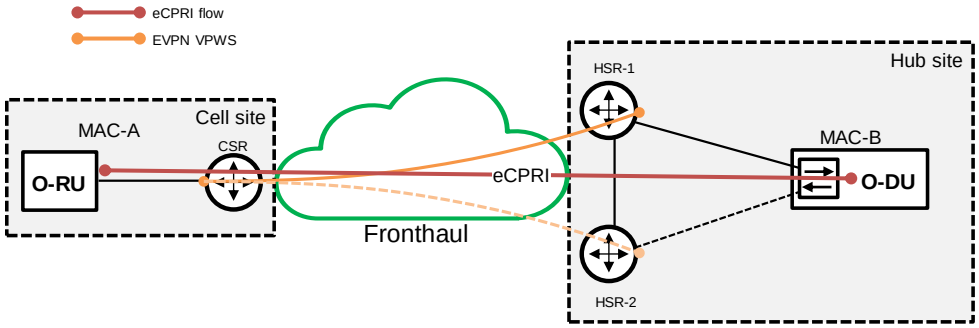


Figure 13-8 EVPN VPWS Redundancy Option 3

For this option, following architectural design choices are recommended:

- Two O-DU uplinks towards HSR pair are not bundled, but are standalone links placed inside internal bridge on O-DU
- EVPN VPWS service should be terminated on HSR pair as 'multi-homed single-active' service

In this option, one of the HSR routers (for example HSR-1) is automatically elected by EVPN control plane as 'active' (Designated Forwarder) router, and eCPRI stream is delivered over EVPN VPWS service from CSR to active HSR only, where it is forwarded to O-DU in Ethernet frames. Internal bridge at O-DU performs MAC learning (to learn O-RU MAC: MAC-A), therefore the eCPRI stream generated from O-DU side and destined to O-RU (MAC-A) follows the same path via HSR-1.

As outlined in Figure 13-9, when HSR-1⇔O-DU link failure happens, following EVPN failure detection machinery, the EVPN VPWS leg between CSR and HSR-1 is disabled, HSR-2 becomes DF and EVPN VPWS leg between CSR and HSR-2 is enabled. eCPRI stream flows now over HSR-2, and internal bridge in O-DU relearns the MAC-A via HSR-2. As opposed to Option 1, since EVPN VPWS termination on HSR-2 was originally 'standby' (multi-homed single-active service), DF election and forwarding plane reprogramming is needed on HSR-2, thus failover is longer than in case of Option 1.

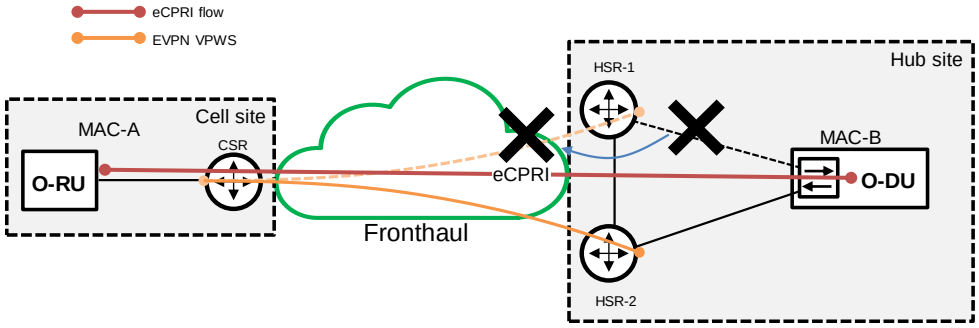


Figure 13-9 EVPN VPWS Redundancy Option 3 – failure event

13.2.1.4 Ethernet services redundancy – Option 4

Option 4 is suitable for O-DUs with only basic Ethernet support. i.e. O-DUs not supporting LAG, internal bridge, or virtual MAC for eCPRI stream termination. Therefore, O-DU is represented by two MAC addresses: MAC-B, on the link towards HSR-1, and MAC-C, on the link towards HSR-2.

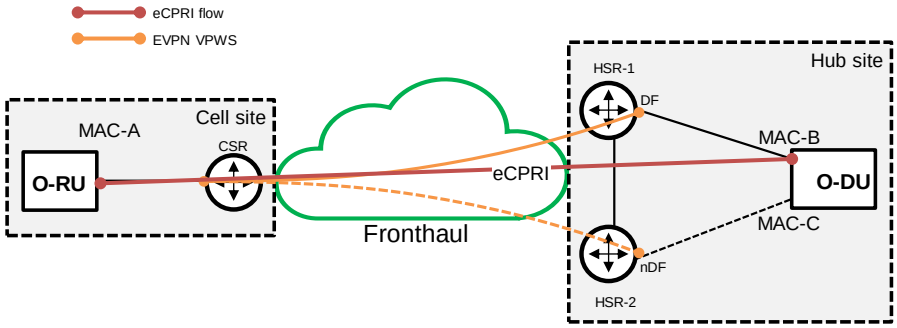


Figure 13-10 EVPN VPWS Redundancy Option 4

For this option, following architectural design choices are recommended:

- Two O-DU uplinks towards HSR pair are not bundled, but are standalone links placed inside internal bridge on O-DU
- EVPN VPWS service should be terminated on HSR pair as 'multi-homed single-active' service

In this option, eCPRI stream generated at O-RU uses of the O-DUs MAC addresses, for example MAC-B, as the destination MAC in eCPRI Ethernet frames. Therefore, EVPN DF election must be deterministic, similar to the DF election on Option 2.

In addition to requirements already mentioned in the main section, to support this option following requirements must be supported on HSR:

[R68]: MUST support "Preference-based EVPN DF Election", draft-ietf-bess-evpn-pref-df [160]

As outlined in Figure 13-11, when HSR-1⇔O-DU link failure happens, following EVPN failure detection machinery, the EVPN VPWS leg between CSR and HSR-1 is disabled, HSR-2 becomes DF and EVPN VPWS leg between CSR and HSR-2 is enabled. However, as opposed to all options discussed previously, in this option, when failure happens, O-RU configuration must be changed by the orchestrator, so that new MAC address (MAC-C) is used as the destination MAC in eCPRI Ethernet frames. This contributes to the highest failover time among all options discussed so far.

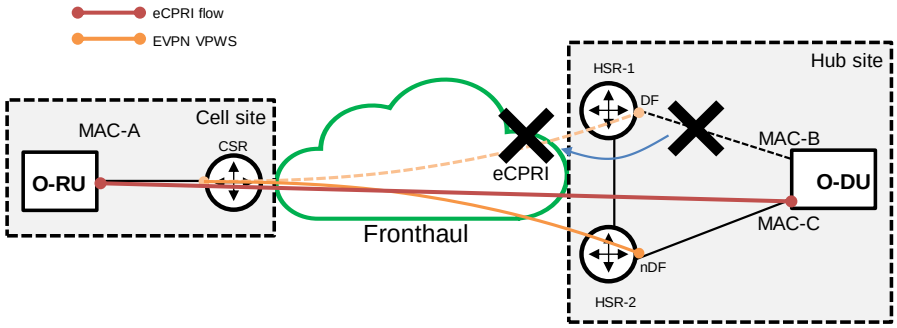


Figure 13-11 EVPN VPWS Redundancy Option 4 – failure event

13.3 IP Services

Mobile IP services will be provided by MP-BGP based L3 VPNs (RFC4364) [84] over MPLS or SRv6 depending on the underlay network technology employed by the operator. More details on L3VPN are provided in annex C.

Operators that choose to implement IP services using the BGP Based L3VPN services may use the following architectural design options:

13.3.1 Building flexible L3VPN service topologies

L3VPN services can be used to establish L3 connectivity between various mobile components. BGP L3 VPN support both IPv4 and IPv6 VPNs and offer flexible connectivity models including:

- IP N:N multipoint services
- IP 1:N multipoint services
- IP 1:1 connectivity services

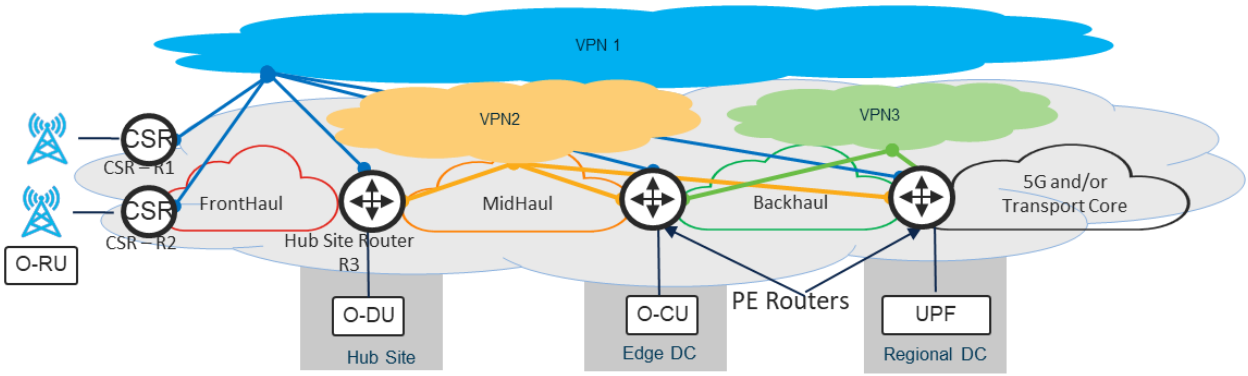


Figure 13-12 Flexible L3 VPN service topologies

13.3.2 Constraints based Traffic Steering in L3VPNs

By default, BGP based L3VPN use shortest path routing across the transport underlay. However, BGP based VPN traffic can also get automatically steered into SR policy, such as one created by a flexible algorithm or a point-to-point TE tunnel. This is achieved by coloring VPN routes within MP-BGP with extended community attributes. As illustrated in Figure 13-13 this can be done in two ways

1. While defining a VRF, define a color for all Prefixes associated with that VRF, or
2. When sending or receiving BGP routes, using a policy to “color” the route.

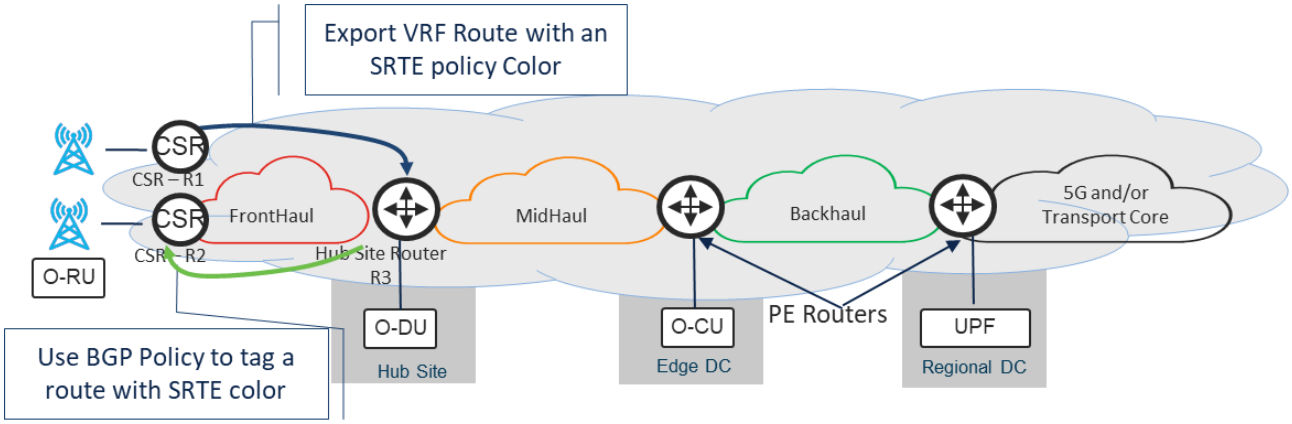


Figure 13-13 Import / Export of route colors for SR policy selection

In either case, the VPN route will now be tagged with a color and traffic for the destination will automatically get steering into an SR policy based on the color associated with the VPN route.

An operator using MP-BGP L3VPN services for Xhaul, TNEs need to support:

[R69]: MUST support Multiprotocol Extensions for BGP-4, RFC4760 [86]

[R70]: MUST Support BGP/MPLS IP Virtual Private Networks (VPNs), RFC4364 [84]

[D38]: SHOULD support SRTE Policy configuration as RFC 9256 [151], if SRTE policies are used with L3VPN services

[D39]: SHOULD support EVPN and L3VPN traffic steering based on “color extended community BGP attributes” defined in RFC5512 [105]

[D40]: SHOULD support to color extended community attribute for VPNv4, VPNv6 and EVPN BGP address families defined in RFC5512 [105]

[R71]: If using SRv6 as the underlay technology TNEs providing PE functionality MUST support “BGP Overlay Services Based on Segment Routing over IPv6 (SRv6)”, Internet Engineering Task Force, RFC 9252 [150]

14 Quality of Service in packet-switched networks

This chapter discusses the Quality of Service (QoS) capabilities that must be deployed to support the delivery of Xhaul traffic and non-Xhaul traffic in the RAN network.

14.1 Xhaul transport core interface QoS

For the purposes of this document, the transport network domain will be considered to be constructed of two interface types. Core interfaces and edge interfaces. An interface is considered to be transport network core if it is interconnecting two devices inside the transport domain. An interface is considered transport network edge if it is connecting to an element outside the transport domain.

14.1.1 Transport network core interface classification

Interfaces in the transport network core domain should perform classification on the outer transport header only. The header in use will depend on the transport model selected by the operator. An example classification scheme for MPLS TC (EXP) to queue is shown below in Figure 14-1

[R72]: TNEs MUST support classification based on MPLS TC in a MPLS underlay transport network

[R73]: TNEs MUST support classification based on IPv6 DSCP in an SRv6 underlay transport network

14.1.2 Core interface queue structure

The Xhaul domain must support a differentiated QoS architecture supporting priority queueing and scheduling with weighted fair queue scheduler for non-priority packets.

[R74]: MUST support EF forwarding described in IETF RFC 3246 [63] mapped to a strict priority scheduler with shaping and policing to prevent bandwidth starvation of other classes.

[R75]: MUST support AF forwarding model described in IETF RFC 2597 [49] mapped to a WFQ, WRR or MDRR scheduler for non-priority classes.

[R76]: MUST support the ability to manipulate queue depths

14.1.2.1 Flat queue model

In the flat queue model, the TNE should support a single level of queue on each physical interface. This model is used for a single topology infrastructure, or a “soft slicing” where the QoS model for each slice is common and bandwidth is dynamically shared between slices.

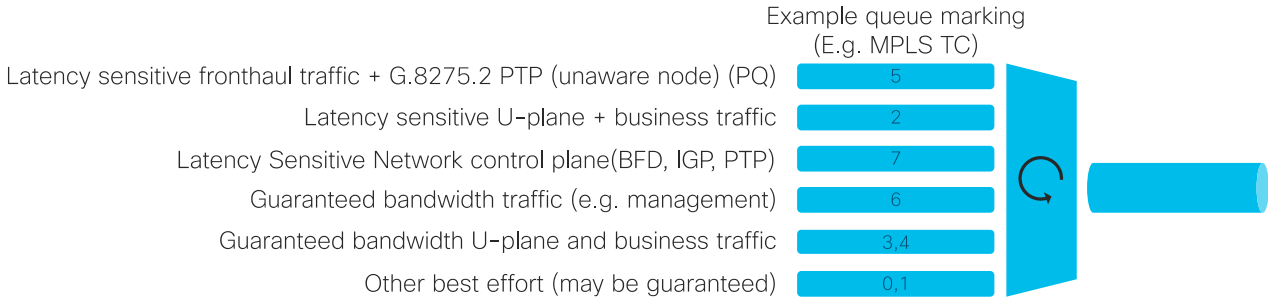


Figure 14-1 Example Xhaul queue model

The system should be capable of supporting a minimum of six queues within the core domain. An example queuing model is provided in Figure 14-1which should be adapted to fit the available queue and scheduling model available in the deployed hardware.

All deployments must support a strict priority queue dedicated to latency sensitive front haul traffic (shown as queue marking 5). This queue must be serviced in strict priority over all other queues up

to its shaped or policed limit. If the node is providing transit for G.8275.2 PTPoIP as a PTP unaware element, then PTP packets MUST also be scheduled in this queue. If the hardware supports multiple levels of strict priority queueing, and is a PTP unaware node, then PTP MAY also be scheduled in a dedicated highest priority queue to minimise packet delay variation (PDV).

Two further latency sensitive queues are defined, one for network control (shown as queue marking 7) and one for other latency sensitive transit traffic (e.g. latency sensitive Backhaul or Midhaul U-plane or latency sensitive business or consumer traffic) (shown as queue marking 2). These queues should be configured to be serviced with a latency bound suitable for the traffic associated with them. The scheduling model used will be dependent on the specific HW choice. For example, this could be a guaranteed bandwidth queue with suitable scheduler weight and queue depth, or a high priority queue with lower priority scheduling than that for the latency sensitive Fronthaul traffic.

Additional queues are defined for guaranteed bandwidth traffic (Management (shown as queue marking 6), guaranteed bandwidth U-plane (shown as queue marking 3,4) and best effort traffic (shown as queue marking 0,1)).

Additional queues may be supported if the carrier deems necessary.

The example queue model is show in Figure 14-1 provides example MPLS “Traffic Class” (TC) or often called EXP, based classification for the queues. It should be noted that a transport based on SRv6 allows the use of IP DSCP as the QoS marker, allowing for more flexibility in the queue model, and that more queues may be used if the marking scheme and hardware are appropriate.

[R77]: MUST support a minimum of 6 HW queues per physical interface

14.1.2.2 Hierarchical queue model

The hierarchical queue model provides a capability to support a segmented infrastructure “Hard slicing” where the QoS model for each slice is different, or where each slice requires dedicated bandwidth. (see Figure 14-2 for details)

Here the model the transport device must support a minimum of two levels of queue hierarchy on each physical interface.

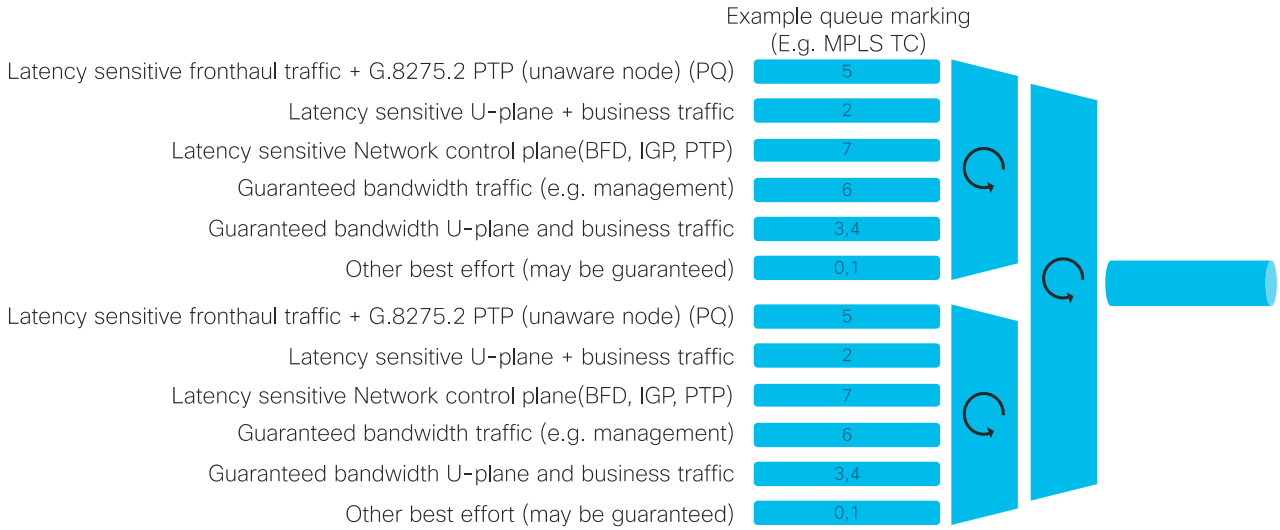


Figure 14-2 Sample hierarchical queue model

In this model the first, or child level defines the queues needed to support the capabilities associated with the slice. These might be the same as in the flat queue model or may be different. In the second or parent level, a scheduler is defined to allocate the maximum bandwidth that will be available to that hard slice. It should be noted that the sum of the guaranteed bandwidth to each shaper should not exceed the interface bandwidth. As with the flat queue model, example is provided which should be adapted for the specific HW in use.

[R78]: MUST support a minimum of 6 HW queues per logical interface

[R79]: MUST support a minimum of two tiers of configurable scheduler per physical interface

14.1.3 Transport network core interface marking structure

Transport nodes should preserve the QoS marking of transported frames and packets across the transport network. This should be accomplished via the use of the pipe or short pipe model described in IETF RFC2983[56].

The specific marking structure used to define the transport QoS behaviours is left to the individual providers and is dependent on the transport encapsulation implemented. An example marking structure for MPLS TC values is shown in Figure 14-1**.**

14.1.4Core interface scheduling model

Latency sensitive and network control plane traffic should be forwarded using an expedited forwarding model as described in IETF RFC 3246[63]. In order to preserve minimum latency, these queues should be scheduled with a maximum of one packet in each queue and be allocated a priority forwarding schedule in the NPU. To prevent bandwidth starvation of non-priority scheduled traffic, each priority queue should be shaped or policed at a combined capacity less than the physical rate of each egress interface in the core, and with sufficient bandwidth capacity available

for the latency sensitive Fronthaul services they are supporting. Where needed a TSN based scheduler may also be implemented if the interface bandwidth demands such optimisation. Details on TSN may be found in section 10.1.1.2.1.

Guaranteed bandwidth queues (management and best effort) should be scheduled using the assured forwarding model described in IETF RFC 2597[49]. These queues should be scheduled using a Weighted Round Robin (WRR), Minimum deficit round robin (MDRR) or weighted fair queue (WFQ) scheduler, depending on the capability available in the NPU. Where network control traffic is assigned to a guaranteed bandwidth queue, the scheduler for that queue must be configured to support minimum latency to ensure time sensitive network control traffic (eg BFD) is not delayed. The scheduler MUST have sufficient buffer capacity to avoid loss in these queues due to the effects of micro-burst caused by the aggregation of traffic from different ingress interfaces, or speed mismatch (higher to lower).

14.2 Xhaul transport network edge interface QoS

PE elements in the transport network domain will be responsible for providing logical separation and encapsulation for transport. It can be assumed that there are two types of traffic that will be presented for transport.

1. Ethernet frames. For example, RoE or eCPRI frames not using an IP header.

2. IP packets. For example, eCPRI packets or Backhaul traffic whose encapsulation includes an IP header.

14.2.1 Transport network domain PE ingress classification of Ethernet frames.

For the purposes of this section, it is assumed that there will be a physical or logical interface attaching the Ethernet frame device to the transport network edge PE. As an example, this could be a Mobile Client, such as a Front Haul Gateway or O-DU supporting an RU using RoE. PE nodes in the transport network will perform classification on traffic ingressing the domain using different models.

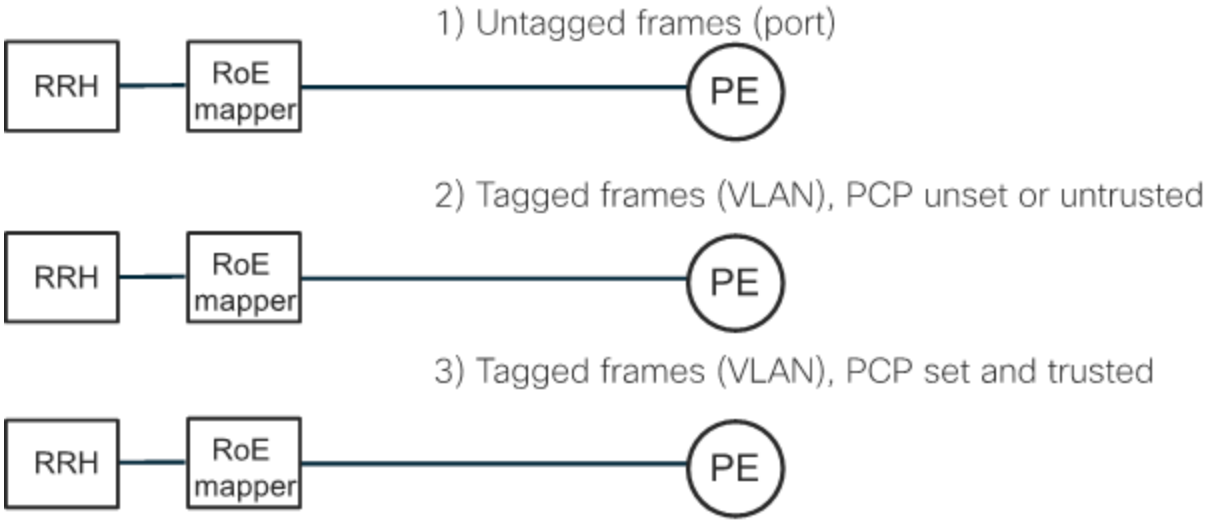


Figure 14-3 Ethernet ingress classification models

Figure 14-3 identifies three models by which transport domain PE devices will classify traffic originating from a ROE mapper. All three models must be supported.

14.2.1.1 Untagged frames

As per Figure 14-3(1) the Mobile Client (MC) presents a flow of untagged frames to the PE. The PE uses the local port as the context for the MC. This context must be preserved in the transport network and presented at the port level (untagged) at the remote PE. Because the PE has no knowledge of the MC encapsulation model, or priority needed for each frame, all frames must be given the same appropriate treatment (and transport marking) to support the encapsulation mode. It can be assumed that the operator has this knowledge at the port level and can apply the appropriate local policy for both transport marking and PHB for all frames. These frames should be encapsulated with the appropriate transport and marked with QoS marking for carried in the class defined in section 14.1.2 for Latency sensitive Fronthaul traffic.

14.2.1.2 Tagged frames (VLAN), PCP unset or untrusted

As per Figure 14-3(2) The RoE mapper presents a flow of VLAN tagged frames to the PE. The PE uses the VLAN and local port as the context for the RoE mapper. This context must be preserved in the transport network and presented at the remote PE with the same VLAN marking. Because the PE has no knowledge of the RoE mapper encapsulation model, or priority needed for each frame, all frames must be given the same appropriate treatment and transport marking, to support the encapsulation mode. It can be assumed that the operator has this knowledge at the either the VLAN or port level and can apply the appropriate local policy for both transport marking and PHB for all frames. These frames should be encapsulated with the appropriate transport and marked with QoS marking for carried in the class defined in 14.1.2 for Latency sensitive Fronthaul traffic.

14.2.1.3 Tagged frames (VLAN), PCP set and trusted

As per Figure 14-3(3). The RoE mapper presents a flow of VLAN tagged frames to the PE. The PE uses the VLAN and local port as the context for the RoE mapper. This context must be preserved in the transport network and presented at the remote PE with the same VLAN marking. In this model it is assumed that the RoE mapper supports a differential marking scheme for flows requiring different treatment. The marking is placed in the PCP field of the VLAN header (described by IEEE 802.1p). The PE should be capable of classifying flows based on the PCP marking and implementing an appropriate PHB. The PE should also apply an appropriate marking to the frames as supported by the transport model in use. It can be assumed that the operator has this knowledge of the PCP markings from the RoE mapper.

Each class presented by the RoE mapper must be mapped to one of the class markings defined in section 14.1.2. It is assumed that the operator has appropriate knowledge to determine this mapping.

[R80]: MUST support classification of frames based on 802.1p PCP bit field

[R81]: MUST support classification of frames based on incoming logical interface.

[D41]: SHOULD support ingress policing using a single rate 2 color policer described in RFC2698 [50] or RFC4115 [73]

[D42]: SHOULD support ingress policing using a dual rate 3 color policer described in RFC2698 [50] or RFC4115 [73]

[R82]: MUST support ingress marking of frames based on result of soft policing using policers in [D41]: and [D42]:.

14.2.2 Transport domain PE ingress classification IP packets.

Unlike Ethernet frames, IP packets will always have a QoS marking field available to them. As a result, we do not have to be concerned with IP packets arriving unmarked from a QoS perspective and must simply determine if we trust the marking.

14.2.2.1 Marked packets DSCP untrusted

Similar to Figure 14-3(2). The client element presents a flow of DSCP marked packets to the PE. The PE uses local port as the context for the flow. This context must be preserved in the transport network and presented at the remote PE with the same DSCP marking. Because the PE has no knowledge of the client element encapsulation model, or priority needed for each packet, all packets must be given the same appropriate treatment and transport marking, to support the encapsulation mode. It can be assumed that the operator has this knowledge at the port level and can apply the appropriate local policy for both transport marking and PHB for all frames. These frames should be encapsulated with the appropriate transport and marked with QoS marking for carried in the class defined in section 14.1.2 for Latency sensitive Fronthaul traffic.

14.2.2.2 Marked packets DSCP trusted

Similar to Figure 14-3. The client element presents a flow of DSCP marked packets to the PE. The PE uses the local port as the context for the flow. This context must be preserved in the transport network and presented at the remote PE with the same DSCP marking. In this model it is assumed that the client element or IP source supports a differential marking scheme for flows requiring different treatment. The marking is placed in the DSCP field of the IP header. The PE should be capable of classifying flows based on the DSCP marking and implementing an appropriate PHB. The PE should also apply an appropriate marking to the frames as supported by the transport model in use. It can be assumed that the operator has this knowledge of the DSCP markings from the client element or IP source.

Each class presented by the client element or IP source must be mapped to one of the class markings defined in section 14.1.2. It is assumed that the operator has appropriate knowledge to determine this mapping.

[R83]: MUST support classification of frames based on IPv4 DSCP field

[R84]: MUST support classification of frames based on IPv6 DSCP field

14.2.3 Admission control

Admission control via policing should be applied to all sources of traffic at the PE. It is important to recognise that suitable capacity MUST be made available for all traffic (especially where loss is problematic) in each element of the infrastructure. Where statistical multiplexing is expected to be used to support additional service capacity in the network, excess traffic, beyond that which is committed, MUST be marked as excess and be eligible for drop to protect other committed traffic under heavy usage periods.

14.2.4 PE Egress scheduling

Traffic arriving at a PE from the infrastructure, to be transmitted to a CE device should be done so based on the model used for its classification on ingress.

14.2.4.1 Unclassified traffic

Where an interface is using unclassified traffic, that is;

all traffic is always marked as default by the CE or

traffic with no marking available e.g. ethernet without a VLAN header,

traffic should be forwarded on a FIFO basis. In this model it must be ensured that the flow of traffic into the PE toward the CE MUST be at a lower rate than the interface attaching the PE to the CE to avoid congestion resulting in latency or loss.

14.2.4.2 Classified traffic

Where an interface is using classified traffic, that is traffic has a marking scheme available, and traffic is marked using this scheme, traffic should be forwarded to the CE with a queueing scheme and scheduling model that is appropriate to the traffic for that CE. The specific marking and queueing scheme are beyond the scope of this document but MUST be understood by the operator. In this model the PE MUST support the classification and queueing of the CE destined traffic based on the marking on the CE packet or frame. This is understood as “pipe model – after” as defined in IETF RFC 2983, section 5 – 6.

[R85]: MUST support classification of encapsulated frames or packets based on 802.1p PCP bit field or IP DSCP field as described in IETF RFC 2983 section 5-6

Latency sensitive and network control plane traffic should be forwarded using an expedited forwarding model as described in IETF RFC 3246 [63]. In order to preserve minimum latency, these queues should be scheduled with a maximum of one packet in each queue and be allocated a priority forwarding schedule in the NPU. To prevent bandwidth starvation of non-priority scheduled traffic, each priority queue should be shaped or policed at a combined capacity less than the physical rate of each egress interface in the core, and with sufficient bandwidth capacity available for the latency sensitive Fronthaul services they are supporting. Where needed a TSN based scheduler may also be implemented if the interface bandwidth demands such optimisation. Details on TSN may be found in section10.1.1.2.1.

Guaranteed bandwidth queues (management and best effort) should be scheduled using the assured forwarding model described in IETF RFC 2597 [49]. These queues should be scheduled using a Weighted Round Robin (WRR), Minimum deficit round robin (MDRR) or weighted fair queue (WFQ) scheduler, depending on the capability available in the NPU. The scheduler MUST have sufficient buffer capacity to avoid loss in these queues due to the effects of micro-burst caused by the aggregation of traffic from different ingress interfaces, or speed mismatch (higher to lower).

15 Multicast

15.1 Multicast use cases

Multicast use cases in an Xhaul transport network come from two categories as described below.

15.1.1 Multicast transport for fixed line services

As mentioned in the requirements section:

"ITU-T GSTP-TN5G: Transport support of IMT-2020/5G" identifies the need for the transport to be multi-service in nature. In addition to mobile services, the infrastructure needs to support fixed line consumer and enterprise services.

Fixed line consumer/enterprise services like IPTV and VPN all require multicast support in the transport network.

15.1.2 MBMS/5MBS transport

3GPP TS 23.246 *MBMS Architecture and Functional Specification* (R15) and TR 23.757 *Study on architectural enhancements for 5G multicast-broadcast services* (5MBS, R17) all support "shared delivery method" with which multicast/broadcast traffic is transported via multicast to RAN nodes (N3/N9), who will then transmit over the air. This means multicast transport is needed to CU-UP.

Another "individual delivery method" may also be used – UPFs send individual copies of multicast traffic over PDU sessions to UEs, transparent to RAN nodes. With this method, if UPFs are distributed, then multicast should be used in DNN to the UPFs (N6). Note that, a UPF may be connected to multiple DNNs as the anchored PDU sessions may belong to multiple DNNs.

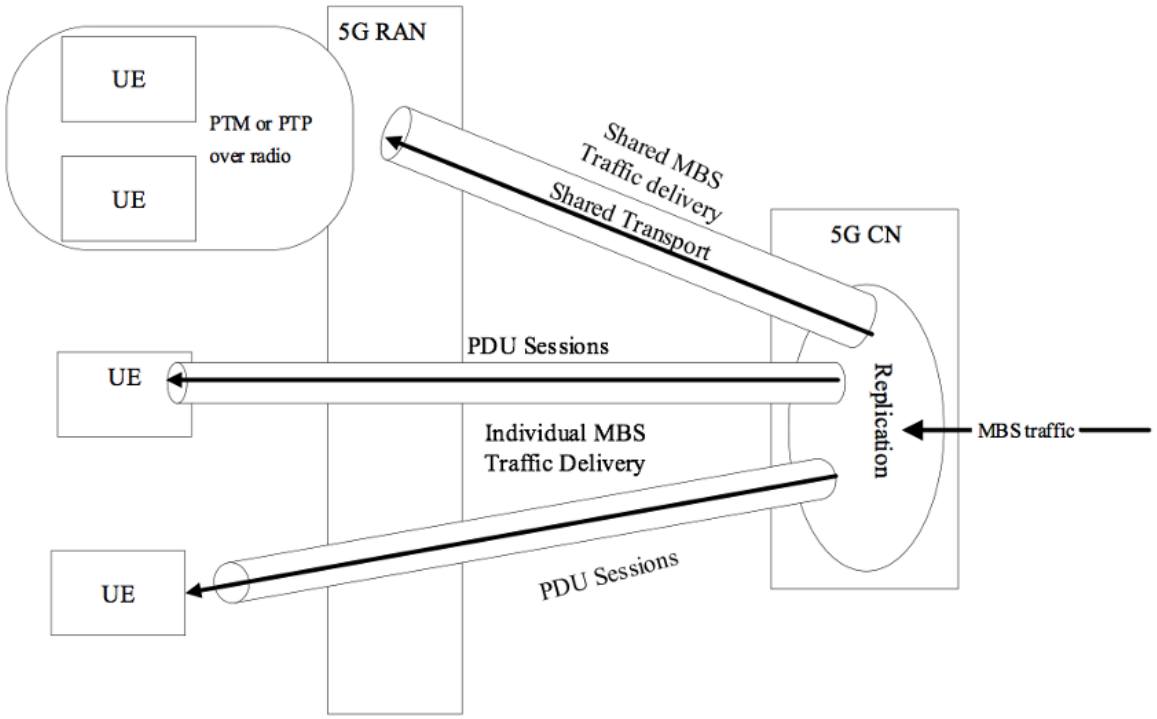


Figure 15-1 Mobile multicast use cases

15.2 Overlay and underlay multicast

The multi-service nature of 5G transport network is not only that the transport network is also used for fixed line service transport. Even for 5G itself, the same transport network is used for multiple purposes, e.g., N3/N9 transport, N6 transport and Xhaul transport.

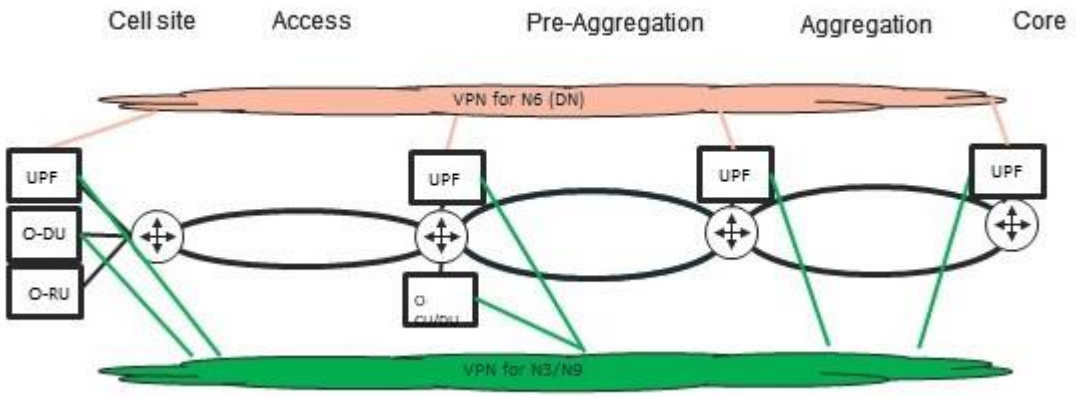


Figure 15-2 VPN infrastructure for mobile

As described in the service capability section, different transports are rendered as different IPVPN/EVPN overlay services over a common transport underlay. When it comes to multicast, it is just an aspect of the IPVPN/EVPN, as further described in section 13.

15.3 Recommendation/considerations for multicast solutions

With the background discussion provided in Annex E: Multicast Technologies background, the following multicast solutions are recommended:

1. BGP-MVPN and EVPN-BUM for overlay, with tunnel segmentation if different tunnel types/instances are necessary/desired for different ASes/areas.
2. Currently deployed underlay tunnel solutions can still be used even with Segment Routing (mLDP/RSVP protocol would only be used for multicast not unicast purposes). If end-to-end tunnels are used without tunnel segmentation, PIM RPF Vector or mLDP Recursive FEC or controller signalling need to be used.
3. Controller-signaled multicast can be used if tree calculation and signalling by controllers are desired to satisfy TE constraints, or to remove legacy LDP/RSVP protocols from the network. This includes SR-P2MP an BGP-signalled mLDP tunnels, with the latter offering the best flexibility – easy transition from existing mLDP deployment and flexible ways of tunnel identification via mLDP FEC.
4. BIER can be deployed in (part of) the network when enough routers support BIER.

Depending on the selected solution, some of the following standards may need to be supported.

[O1]: BGP Encoding and Procedures for Multicast in MPLS/BGP IP VPNs, RFC6514 [115]

[O2]: BGP MPLS-Based Ethernet VPN, RFC7432 [116]

[O3]: Update on EVPN BUM Procedures, draft-ietf-bess-evpn-bum-procedure-updates [157]

[O4]: Protocol Independent Multicast - Sparse Mode (PIM-SM) Protocol Specification, RFC 7761 [121]

[O5]: The Reverse Path Forwarding (RPF) Vector TLV, RFC 5496 [104]

[O6]: Label Distribution Protocol Extensions for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths, RFC6388 [111]

[O7]: Using Multipoint LDP When the Backbone Has No Route to the Root, RFC 6512 [113]

[O8]: Extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs), RFC4875 [87]

[O9]: Controller Based BGP Multicast Signaling, draft-ietf-bess-bgp-multicast-controller [153]

[O10]: Segment Routing Point-to-Multipoint Policy, draft-ietf-pim-sr-p2mp-policy [162]

16 Packet-switched orchestration and telemetry

Packet switched orchestration and telemetry covers how the packet switching transport network is programmed at a device and service level and how telemetry data is retrieved from TNEs. This will be covered in a separate document[22].

17 Packet Switched TNE Security

Packet Switched Transport Network Equipment (TNE) plays a vital role in building the transport infrastructure needed to support a 5G infrastructure. As such the packet switched TNEs needs to support a set of security features associated with:

1) The security of the TNEs (both physical security nad user access security)
2) The security of the network control, data and management plane protocols the TNEs run
3) Security functionality required by the TNEs to support security functionality required by the 5G mobile components.

This section, which deals with transport specific requirements, in association with O-RAN.WG11.Security-Requirements-Specifications [26] provides guidelines and requirements for security features required for an O-RAN compliant packet switched transport infrastructure.

17.1 TNE physical security

This section describes the measures to improve the security of the TNE, when the TNE is placed in unsecured, publicly accessible location. For example, such as remote unattended site, or a shared rack with a partner or customer organization

17.1.1 Console Port

Typical TNE has a console port, which is typically used for last resort, emergency access to the TNE. Console port on TNE MUST be properly secured to prevent unauthorized access over the console port.

One type of such unauthorized access could happen, when an authorized user is logged into the console port and inadvertently disconnects the cable without properly logging out. The next time someone connects to the console, they could have access to the TNE without authenticating.

[R86]: TNE MUST automatically log out a user connected over console when console cable is disconnected

The other type of unattended session occurs when a connection is left idle, which is rather common situation when a terminal server is used, as in that case the console cable is always connected.

[R87]: TNE MUST support automatic log out of users connected over console, when the session is idle for a (configurable) period

The Root or “Super user” account is the most important and most powerful user account on a TNE. Very often, TNEs do not provide idle timers for the root account. Therefore, to increase the console security, the root user MUST be prevented from directly login using the console. This essentially forces console users to initially authenticate using a regular user account, and subsequently, from within the session of the regular user, authenticate as the root user. If the root session, running within regular user session, is idle for a (configurable) period, regular user session is idle as well. Based on [R87]:, the regular user session MUST be automatically logged out, which will automatically break all other user sessions (i.e., root user) within that regular user session, too.

[R88]: TNE MUST prevent root user login over the console

17.1.2 Auxiliary Port

Many TNEs have as well an auxiliary port, which is intended to be used with a modem to provide dial-in access to a TNE in remote locations. In most cases, the auxiliary port can also be used as a secondary console port, in which case it inherits some console port security concerns. In most cases, the auxiliary port is not used in typical network deployments. Leaving that port enabled creates security concern.

[R89]: TNE MUST provide capability to disable auxiliary port, if the port is not used

[R90]: TNE MUST provide capability to secure auxiliary port in a similar manner as console port, if the auxiliary port is used

17.1.3 USB Port

Many TNEs today have external user accessible USB ports used for booting the device off a USB stored image or for loading software onto the device's internal storage. The USB port can be exploited to load malicious device software or used to easily transfer provider information in the case of a device with a compromised management plane. In many cases the device when booting will give the user an option to boot off the USB if a capable USB storage device is connected. Leaving the USB port active creates a security concern.

[R91]: TNE MUST provide capability to disable USB port, if the port is not in use

17.1.4 Front Panel functions

Many TNEs have front panel with physical buttons, LCD displays or menu options to perform system control and maintenance functions, like bringing a line-card offline or online, restoring the factory default configuration, acknowledging alarms, etc. There may be scenarios, where TNEs are placed in an unsecure location, such as remote unattended site, or a shared rack with a partner or customer organization. Such unsecure locations require enhanced security for front panel functions (buttons, LCD displays, menus, etc.).

[R92]: TNE MUST provide capability to disable/enable front panel functions (buttons, menus, LCD displays, etc.)

17.1.5 TNE software security

Due to the security risks posed by malicious software, the TNE must validate software loaded onto the device before execution. The method for implementation is outside the scope of this document, however the signing of all software images by the TNE vendor's private key is the common industry practice. Image signing applies to boot software as well as incremental updates applied to a running device.

[D43]: TNE SHOULD validate the device boot loader before execution, halting boot on revocation

[D44]: TNE SHOULD support signed boot images validated by the device boot loader before execution, halting boot on revocation

[D45]: TNE SHOULD validate all software images and updates prior to execution

17.1.6 Zero Touch Provisioning

Due to the scale of today's mobile packet switched networks, Zero Touch Provisioning (ZTP) is often used either in the field to deploy new devices or replace existing devices with hardware impairments. While ZTP operations are typically internal to a provider network, it's important to secure the communication and components used with ZTP. RFC 8572 [137] defines Secure ZTP, a framework for ensuring the ZTP process is secure end to end. The primary points of security for ZTP are:

1) ZTP client validation, to ensure the client is trusted by the server before offering ZTP artifacts or configuration information
2) ZTP server validation, to ensure the server is trusted by the client before accepting any ZTP configuration or artifacts
3) Artifact validation, to ensure the ZTP components downloaded from the ZTP server are trusted before being loaded or executed

[R93]: TNE MUST provide follow the SZTP framework defined in RFC 8572 [137] if the device supports ZTP

[R94]: TNE MUST follow methods and formats defined in RFC 8366 [130] to ensure the validity of the device from the manufacturer using an Ownership Voucher

17.1.7 PXE boot over management or data ports

Many TNEs today will attempt to boot via PXE as part of their Zero Touch Provisioning process. The ZTP process can run on the management Ethernet interface as well as on standard data interfaces. This process can be a security risk if the physical device location is compromised, allowing an attacker to load malicious software onto the device or compromise its boot integrity protections.

[R95]: TNE MUST provide capability to disable ZTP, if ZTP is not in use

17.2 TNE user access security

This section describes the measures to improve the security of the TNE, preventing unauthorized user to access the device.

17.2.1 TACACS+ Authentication

Typically, the default authentication used to authenticate user on the TNE is password authentication using locally stored user data (username, password). In large-scale deployments, like 5G Xhaul deployment, local authentication database on each TNE is very impractical. Therefore, the recommended solution is to use TACACS+ server, which should always be considered the first and principal authentication method.

After integrating the TNE into TACACS+ authentication infrastructure, the TACACS+ authentication method should be preferred over password authentication using local authentication

database. When user authentication order is set on TNE (with TACACS+ authentication as first, and authentication using local authentication database as second) the user is first authenticated with TACACS+ server. If TACACAS+ server responds (either granting the access or refusing the access), this response is used by the TNE to either allow or deny the user access to the device. However, if TACACS+ server doesn't respond (is unreachable), local password authentication is used to grant or deny access to the device. Therefore, it is always recommended to have limited number of local users defined on the TNE as a last resort for authentication. These users can be used for authentication purpose in failure cases, when TACACS+ server is not available.

TACACS+ provides not only user authentication, i.e., to determine who a user is, but as well authorization, i.e., to grant access rights to certain portions of the TNE, or to certain configuration or operational commands on the TNE. Additionally, TACACS+ performs accounting, which is the action of recording what a user is doing and/or has done. This is commonly known as AAA (Authentication, Authorization, Accounting).

[R96]: TNE MUST support TACACS+ for user authentication, in accordance with Section 5 of RFC 8907: "The Terminal Access Controller Access-Control System Plus (TACACS+) Protocol" [146]

[R97]: TNE MUST support TACACS+ for user authorization, in accordance with Section 6 of RFC 8907: "The Terminal Access Controller Access-Control System Plus (TACACS+) Protocol" [146]

[D46]: TNE SHOULD support TACACS+ for user accounting, in accordance with Section 7 of RFC 8907: "The Terminal Access Controller Access-Control System Plus (TACACS+) Protocol" [146]

[R98]: TNE MUST support specification of authentication order, preferring TACACS+ authentication (RFC 8907 [146]) over authentication using local user database

17.2.2 Login Message

Typically, by default, no message is displayed by the TNE before or after login. It is strongly advised that a system login message is present before the user logs in, as well as another message after user successfully logged in. A strongly worded login banner SHOULD include statements regarding the following:

- Clear prohibition of unauthorized access.
- Clear prohibition of unauthorized use. This warns authorized users against unauthorized use of the system.
- Warning that the system may be monitored. Note that the wording might have important ramifications for successful prosecutions. Do not say the system will be monitored.
- Statement that there is no expectation of privacy. This statement insures that no legal defense can be mounted claiming that the company should not have monitored an unauthorized user.
- Statement that results of monitoring may be provided to the appropriate authorities.

- Statement that continued use of the system implies consent to the stated terms and conditions.
- Warning to log off immediately if consent is not given.

Login banners SHOULD NOT include any of the following information:

- Information about the device type or software
- Information about the device location
- Contact information
- Administrator information

There could be two level of messages displayed by on TNE's terminal: one message appearing before the user logs in, and 2nd message after the user logs in.

[R99]: TNE MUST support modifiable messages displayed on the user session before and after user login

[D47]: Login banners SHOULD NOT include details about the device, location or contact/administrator information

17.2.3 Password requirements

Variety of requirements can be used to strengthen passwords for greater security. TNE should provide a framework allowing to enforce password rules defined by the operator to conform to a particular set of requirements defined in operator's security policy. These password rules may include such things as length, number of changes, type of characters, numbers, or letters, case, etc.

[D48]: TNE SHOULD support definition and enforcement of user password rules like length, number of changes, type of characters, numbers, or letters, case, etc.

17.2.4 Device lockout

Malicious users sometimes might try to log in to a secure TNE by guessing an authorized user account's password. Locking out a user account after several failed authentication attempts helps protect the device from malicious users.

[D49]: TNE SHOULD have the possibility to lockout the user after a number of failed attempts, and to set the amount of time before the user can attempt to log in to the device again.

17.3 Transport Network control plane security

17.3.1 Control plane DoS/DDoS protection

A denial-of-service (DoS) attack against an TNE is any attempt to disturb legitimate traffic or deny valid users access the TNE by using up all the resources of the TNE. Distributed denial-of-service (DDoS) attacks involve an attack from multiple sources, enabling a much greater amount of traffic to attack the TNE. The attacks typically use network protocol control packets to trigger a large

number of exceptions, which are packets handed over from the data plane to the control plane of the TNE. This results in an excessive processing load that disrupts normal network operations.

TNE MUST have DDoS protection capability, which enables the TNE to continue functioning while under an attack. It identifies and suppresses malicious control packets while enabling legitimate control traffic to be processed. TNE SHOULD have possibility to customize DDoS protection profiles allowing to optimize DDoS protection of network control traffic in a particular deployment.

To protect against DDoS attacks, typically filters and policers for host-bound exception traffic are used. The filters completely block specific types of control traffic (for example, if OSPF is not used in particular deployment, there is no need to send OSPF packets to the TNE's control plane), while policers specify rate limits for individual types of protocol control packets or for all control packet types for a protocol. Control traffic is dropped when it exceeds policer values. Each violation SHOULD immediately generates a notification to alert operators about a possible attack. The violation SHOULD be counted, the time that the violation starts noted, and the time of the last observed violation noted.

[R100]: TNE MUST support protection mechanisms (filters, policer) to protect own control plane from DoS/DDoS attacks

[D50]: TNE SHOULD have the possibility to customize DoS/DDoS protection parameters to protect own control plane

[D51]: TNE SHOULD have the possibility to raise alarms, when DoS/DDoS attacks are detected

17.3.2 Control plane protocols protection

Transport control plane uses various protocols to exchange information between TNEs. This information is required to build paths across transport network, exchange label information (in MPLS deployments), exchange service information (L3VPN, L2VPN), monitor the network state (periodic OAM messages), and more (for example, determine the virtual gateway state in multi-homed environment). These protocols could be divided into following groups:

- IGP (e.g., OSPFv2, OSPFv3, IS-IS)
- Multicast (e.g., PIM, MSDP)
- Protocols to distribute transport labels (e.g. LDP, RSVP)
- Other protocols (e.g. BFD, VRRP, …)

This document doesn't mandate any specific protocol to be used. For example, one deployment could use OSPFv2 as the IGP, while other deployment could use IS-IS as the IGP. However, if certain protocol is used, this document mandates minimum security requirements for each protocol that must be fulfilled.

17.3.2.1 IS-IS

All IS-IS packets (Hello, Link State, Sequence Numbers) shall be authenticated with the Hashed Message Authentication Code with the Secure Hash Algorithm (HMAC-SHA-1) cryptographic checksum in accordance with the recommendations outlined in RFC 5310: “IS-IS Generic Cryptographic Authentication” [98]. Cryptographic checksum is calculated over additional fields from IS-IS packets (LSP ID, Sequence Number) not covered by basic IS-IS checksum defined in ISO/IEC 10589:2002: “Intermediate System to Intermediate System intra-domain routeing information exchange protocol for use in conjunction with the protocol for providing the connectionless-mode network service” [202], Section 7.3.11. With basic IS-IS checksum, undetected bit errors (caused for example due to some hardware failure) in LSP ID or Sequence Number fields may lead to complete network outage in the worst case.

[R101]: In deployments using IS-IS, TNE MUST support HMAC-SHA-1 authentication for all IS-IS packets (Hello, Link State, Sequence Numbers) in accordance with RFC 5310: “IS-IS Generic Cryptographic Authentication” [98]

17.3.2.1 OSPFv2

OSPFv2 has built-in authentication mechanism defined in RFC 2328: “OSPF Version 2” [44]. It defines three authentication types:

- 0: Null authentication
- 1: Simple password
- 2: Cryptographic authentication

Null (no authentication) and simple (clear text) authentication methods are not sufficient in today’s network.

[R102]: In deployments using OSPFv2, TNE MUST support cryptographic authentication in accordance with RFC 2328: “OSPF Version 2” [44].

17.3.2.2 OSPFv3

OSPF version 3 (OSPFv3), unlike OSPF version 2, does not have a built-in authentication method and relies on IPsec to provide this functionality. This authentication functionality for OSPFv3 is defined in RFC 4552: “Authentication/Confidentiality for OSPFv3” [80].

[R103]: In deployments using OSPFv3, TNE MUST support IPsec based authentication for OSPFv3 packets in accordance with RFC 4552: “Authentication/Confidentiality for OSPFv3” [80]

17.3.2.3 BGP

BGP is TCP based protocol, therefore it benefits from RFC 5925: “The TCP Authentication Option” (TCP-AO) [108] to protect against the malicious or accidental insertion of unauthorized BGP data into the network by preventing the establishment of BGP adjacencies with any unauthenticated router. As per [108] TCP-AO specifies the use of stronger Message Authentication Codes (MACs), protects against replays even for long-lived TCP connections, and provides more details on the association of security with TCP connections than TCP MD5 introduced in the BGP specification [76]. Further, TCP-AO uses cryptographic algorithms to compute the MAC that is

used to authenticate a segment and its headers. MAC algorithms are specified in a separate RFC 5926: “Cryptographic Algorithms for the TCP Authentication Option (TCP-AO)” [109]. This allows updating the algorithm requirements independently from the protocol.

[R104]: TNE MUST support the TCP Authentication Option Authentication to secure BGP sessions in accordance with RFC 5925: “The TCP Authentication Option Authentication” [108]

[R105]: TNE MUST support HMAC-SHA-1-96 MAC algorithm for TCP-AO in accordance with RFC 5926: “Cryptographic Algorithms for the TCP Authentication Option (TCP-AO)” [109]

[D52]: TNE SHOULD support AES-128-CMAC-96 MAC algorithm for TCP-AO in accordance with RFC 5926: “Cryptographic Algorithms for the TCP Authentication Option (TCP-AO)” [109]

17.3.2.4 LDP

LDP, similarly to BGP, uses TCP for underlay transport. Therefore, TCP-AO [108] applies to LDP as well.

[R106]: In deployments using LDP, TNE MUST support the TCP Authentication Option Authentication to secure LDP sessions in accordance with RFC 5925: “The TCP Authentication Option Authentication” [108]

17.3.2.1 RSVP

IETF defined authentication for RSVP using MD5 one-way hashes in RFC 2747: “RSVP Cryptographic Authentication” [52]. RSVP protocol exchanges can be authenticated to guarantee that only trusted neighbors participate in setting up reservations. RSVP authentication uses an HMAC MD5 message-based digest. This scheme produces a message digest based on a secret authentication key and the message contents.

Small error correction to RFC 2547 [52] was provided via RFC 3097 “RSVP Cryptographic Authentication—Updated Message Type Value” [59]. IETF didn’t specified, so far, any more modern authentication method for RSVP.

[R107]: In deployments using RSVP, TNE MUST support MD5-based cryptographic authentication to secure RSVP sessions in accordance with RFC 2747: “RSVP Cryptographic Authentication” [52]

[R108]: In deployments using RSVP, TNE MUST support amendment to the cryptographic authentication in accordance with RFC 3097 “RSVP Cryptographic Authentication—Updated Message Type Value” [59].

17.4Port-based network access control (IEEE 802.1X-2020)

WG11 has standardized (O-RAN.WG11.Security-Requirements-Specifications [26]) on IEEE 802.1X-2020 [36] as the port security mechanism in O-RAN compliant network infrastructure.

Based on [26], from the transport network domain perspective, following aspects of IEEE 802.1X-2020 [36] SHALL be supported on the TNE deployed in Open Fronthaul

[R109]: Port-based network access control for point-to-point LAN segments for both port types: with Layer-2 and Layer-3 encapsulation. Note: VLAN-based access control or access control for non-point-to-point LAN segments is optional.

[R110]: IEEE 802.1X-2020 [36] authenticator functionality for each port connection towards the O-RAN compliant network elements.

IEEE 802.1X-2020 [36] can be used with various authentication methods, and with IETF RADIUS or IETF Diameter to interact with an authentication server. The exact requirements mandatory and optional methods are specified in O-RAN.WG11.Security-Requirements-Specifications [26].

Figure 17-1 outlines the scope of IEEE 802.1X in Open Fronthaul showing all features required on TNE deployed in Open Fronthaul point-to-point LAN segment links for support of TNE-to-O-RU, TNE-to-FHG, TNE-to-FHM, TNE-to-O-DU TNE-to-Compute or TNE-to-O-Cloud-Gateway. O-DU can be physical O-DU (O-DU on a bare-metal) directly connected to TNE (1), virtual O-DU deployed on a compute node directly connected to the TNE (2), or an O-DU deployed in the O-Cloud, which connects to TNE via O-Cloud gateway (3). In use cases (2) and (3) TNE uses IEEE 802.1X [36] authentication not towards O-DU, but towards physical device directly connected to TNE (compute node or O-Cloud gateway). IEEE 802.1X [36] authentication beyond the physical device (compute node or O-Cloud gateway) directly connected to TNE is out-of-scope for this document.

Intermediate relay TNEs without Layer-2 or Layer-3 switching/routing functionality enabled are out of scope for IEEE 802.1X-2020 [36] port-based network access control because there are no point-to-point LAN segment links. Furthermore, in packet switched transport network architecture using MPLS or SRv6 underlay, as described in this document (Sections 11 and 12), transport network control plane protocols (e.g., IS-IS, OSPF, BGP, LDP, RSVP) have built-in authentication mechanisms (Section 17.3.2) for authentication and authorization of the Open Fronthaul between TNEs.

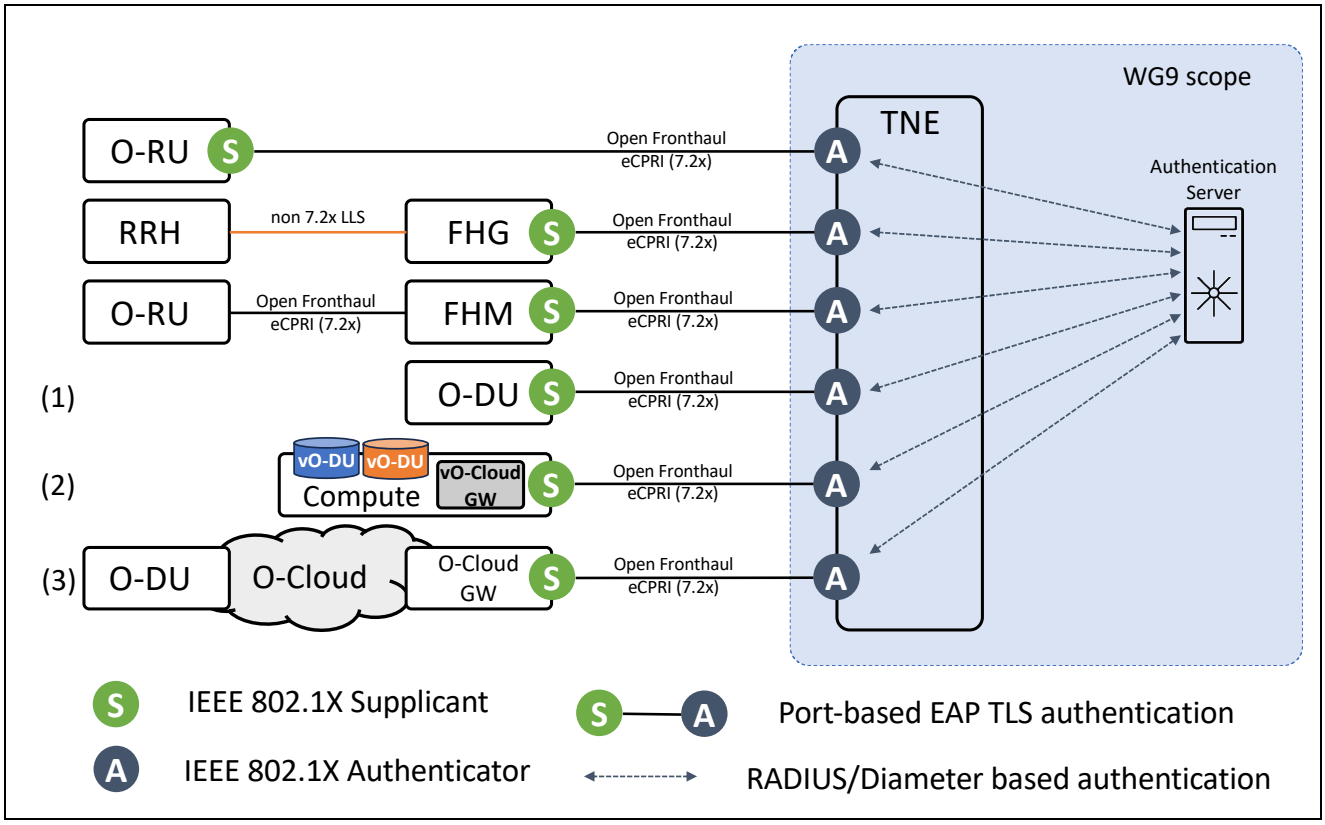


Figure 17-1: IEEE 802.1X scope in Open Fronthaul (Example)

At the high-level, the overall authentication process via EAP-TLS can be divided into three phases:

1. Limited authentication using manufacturer pre-installed certificate
2. Certificate enrollment into operator's Public Key Infrastructure (PKI)
3. Full authentication using operator installed certificate

The details of this authentication process are described and standardized in O-RAN.WG11.Security-Requirements-Specifications [26], Section 5.2.5.5.2. All three phases SHALL be supported on the TNE deployed in Open Fronthaul.

Note: IEEE 802.1X-2020 [36] support on TNEs deployed in other parts of the network (not Open Fronthaul) is optional.

17.4.1 Certificate time validation

Operation of the EAP-TLS method requires the TNE verify the certificate chain presented by the authenticator PAE, including confirming the certificate's validity periods. TNEs using IEEE 802.1X port-based access control shall be able to verify the certificate validity periods during the EAP-TLS exchange, e.g., by employing Global Navigation Satellite System (GNSS), PTP, or NTP-based ToD (Time of the day) synchronization, or employing a persistent clock previously synchronized to GNSS, PTP or NTP.

If the TNE cannot verify the certificates validity, e.g., because there is not available GNSS/PTP/NTP time source, and a persistent clock has failed or was never synchronized to

GNSS/PTP/NTP, the EAP authentication procedure will fail. The behavior to handle such situation is not specified in this document but is up to the operator's policy.

17.4.2 Certificate management for TNEs providing Open Fronthaul transport

17.4.2.1 Trust anchor provisioning

Before a TNE can establish a mutual EAP-TLS connection with a signaling peer, e.g., with an IEEE 802.1X supplicant, a TNE needs to be able to trace the peer's certificate path to a valid trust anchor. To validate against a trust anchor, the TNE shall be able to be provisioned with one or more trust anchor certificates. The TNE shall ensure that all trust anchor certificates are stored in reset persistent memory and protected from external modification.

The TNE shall be able to be provisioned with new trust anchors. The TNE shall be able to have an existing trust anchor replaced, e.g., because it has expired. The TNE should support ietf-truststore YANG model ([172]) to enable to recover the list of provisioned trust anchors and associated public keys.

The present document defines the use of Certification Management Protocol version 2 (CMPv2, [75]) Initialization Response message to allow the discovered Certification Authority (CA) and Registration Authority (RA) server to provision a trust anchor.

NOTE: The CA/RA Server identity is configured manually or by the DHCP server using techniques described in section 17.4.2.2. The Dynamic Host Configuration Protocol (DHCP) server is not considered by the TNE as a trusted source of security bootstrapping data.

When shipped, a TNE only trusts information that is signed or encrypted using a certificate chain leading to a pre-loaded trust anchor.

17.4.2.2 Operator-signed certificate enrolment

A TNE shall support certificate enrolment using CMPv2 ([75]). Reachability (IP address, Fully Qualified Domain Name – FQDN) to the CA/RA servers can be configured on TNE during initial TNE provisioning. Alternatively, 3GPP TS 32.509 ([7]), clause 4.2.2, specifies how the TNE supporting certificate enrolment over IPv4 can be configured with the IP address or FQDN of one or more CA/RA servers using DHCP Option 43.

The DHCP Options specified in 3GPP TS 32.509 ([7]), do not specify how to signal a CA/RA server identity using DHCP for IPv6 hosts (DHCPv6). Hence, TNE certificate enrolment using CMPv2 over IPv6 shall support the signaling of vendor specific options using DHCPv6 option 17. The format of the DHCPv6 option 17 follows the format of the DHCP for IPv4 hosts (DHCPv4) encoding as specified in clause 4.2.2 of 3GPP TS 32.509 ([7]), with the additional inclusion of an Enterprise Number prior to the Type-Length-Value (TLV) option data. The Internet Assigned Numbers Authority (IANA) allocated private enterprise number to be used with DHCPv6 option 17 shall be 53148 (as allocated by IANA to O-RAN Alliance).

3GPP has since published 3GPP TS 28.316 ([3]), which now specifies how to signal a CA/RA server identity with DHCPv6 option 17 messages using the 3GPP registered IANA Enterprise Number (10415). The DHCP Options defined in [7] are a subset of those defined in [3]. An operator and vendor can agree to use the 3GPP registered IANA Enterprise Number in DHCPv6 messages

that signal the CA/RA server information instead of the O-RAN registered IANA Enterprise Number.

If a TNE provides Layer-3 (IP) transport connectivity between IEEE 802.1X supplicants, the TNE shall operate as the default gateway to the IEEE 802.1X supplicant to provide connectivity between the IEEE 802.1X supplicant and a remote CA/RA server, when the IEEE 802.1X supplicant is required to perform certificate enrolment using CMPv2, the TNE shall provide means for the IEEE 802.1X supplicant to access the operator CA/RA for the IEEE 802.1X supplicant certificate enrolment at the IP address conveyed as described above. If the FQDN option is used instead, the TNE shall provide means for the IEEE 802.1X supplicant to access an operator DNS server in addition. Examples of providing such access by a TNE include:

- An IP forwarding function within the TNE providing access to the operators PKI.
- A Network Address Translation (NAT) function within the TNE providing access to the operators PKI.
- A Registration Authority function accessible by the TNE (in the same domain), which is part of the operator PKI.

A TNE shall report any discovered multi-vendor plug-and-play servers using the o-ran-dhcp YANG model (reference needed).

17.4.2.3 CMPv2 based certificate enrolment

This section covers the use case where a TNE is configured with the identity of a CMPv2 capable CA/RA server that the operator uses for certificate enrolment. The TNE shall attempt to enroll in the operator-PKI and may be issued an operator-signed certificate.

Clause 9 of 3GPP TS 33.310 ([8]) specifies the use of CMPv2 used by base stations to obtain an operator-signed certificate using a secured communication based on the vendor-signed certificate in the base station and a vendor root certificate pre-installed in the CA/RA server. While the approach has been defined for provisioning certificates on base stations for use in either IPSec or TLS, the same techniques defined for provisioning TLS certificates on base stations can be used for provisioning TLS certificates on TNEs.

The handling of certificates, including certificate profiles, shall follow the rules defined in 3GPP TS 33.310 ([8]) for TLS CA certificates. In addition:

- when a TNE generates a certificate signing request it shall populate the Subject Distinguished Name field with a string that includes the TNE manufacturer's name, model and serial number. The exact Subject DN sub-field used is defined in the operator of the CA/RA server's certificate policy.

NOTE 1: In future, an O-RAN defined certificate policy may be defined to normalize the sub-field definition across the O-RAN ecosystem.

NOTE 2: There are various characters that may not be permissible in the Subject Distinguished Name Field, e.g., ":" (colon, hexadecimal character 0x34), "." (period, hexadecimal character 0x2E), "_" (underscore, hexadecimal character 0x5F), "#" (hash, hexadecimal 0x23), "£" (pound, hexadecimal 0xa3), "*" (asterisk, hexadecimal 0x2a) or "”" (double quote, hexadecimal 0x22). Manufacturers that include such characters in their name, model and/or serial number should ensure such characters are removed before including in the Subject Distinguished Name Field.

- when transferring messages to the CA/RA server, the TNE shall use the "port number of the CA/RA server" and the “path to the CA/RA directory” as signalled using the DHCP options as specified in clause 4.2.2 of 3GPP TS 32.509 ([7]). If no DHCP based configuration is received by a TNE, the TNE shall use the default port 443 and default directory "/pkix/".

- The CA/RA server shall include the trust anchor for the operator issued certificate and the appropriate certificate chains in the initialization response message.

- The TNE shall store the operator issued certificate and corresponding certificate chain in reset persistent memory.

When configured to operate with EAP-TLS, a TNE that has a valid certificate issued by an operator PKI shall use this certificate to establish a mutually authenticated EAP-TLS connection for IEEE 802.1X-2020 [36] authentication.

17.4.2.4 Operation with vendor-signed certificates

All TNEs should be pre-provisioned at a factory with a unique device vendor-signed certificate and its entire certificate chain up to the root. If a TNE fails to enroll in an operator-PKI, as specified in section 17.4.2.2, a TNE configured to operate with TLS may use its vendor-signed certificate initially. The vendor’s public trusted certificate and certificate management services should be accessible within the operator network to allow management of vendor-signed certificates. The extended use of unmanaged vendor-signed certificates, such as outside of enrolment, is not recommended.

18 5G Slicing in a packet switched Xhaul network

Network Slicing is end-to-end partitioning of the network resources and network functions so that selected applications/services/connections may run in isolation from each other for a specific business purpose. The overall slicing architecture is shown in Figure 18-1 and covers the orchestration infrastructure and at the physical layer can cover the radio access network, the mobile core, including data centers and virtualization aspects, and the X-haul transport network.

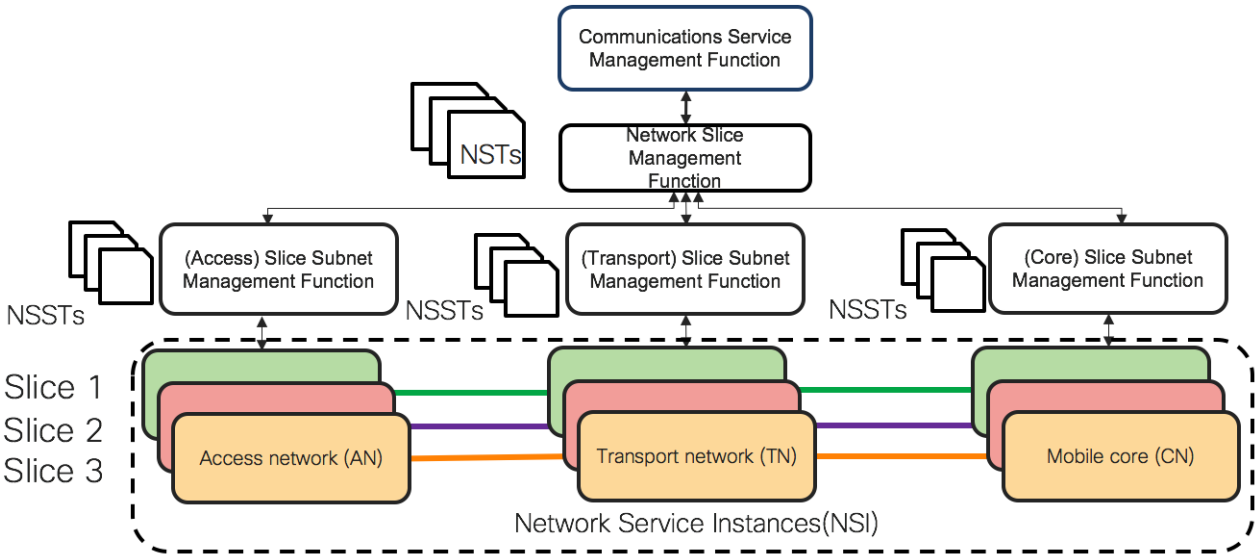


Figure 18-1: Overall 5G slicing architecture

Although slicing is a key capability of 5G, there is debate as to how it exactly translates into the transport network, which mobile interfaces (Fronthaul, Midhaul, Backhaul, N6) need to be sliced, what form they will take, and the number of slices required at the transport layer. For the purposes of this document the Xhaul transport infrastructure needs the capability to support transport requirements of the:

1. O-RAN and 3GPP control, management, and user plane interfaces.
2. Wireline consumer, enterprise, and wholesale services.

This includes 5G transport slicing, which from a mobile perspective may include:

1. Transport separation between Fronthaul, Midhaul, Backhaul interfaces.
2. Transport separation between Control, management, and user plane interfaces of each domain.
3. Flexible mapping of Network Slice Instances (NSIs) to physical or logical transport network instances.

Figure 18-2 is an example of how three NSIs could be mapped to logical networks within an Xhaul transport network. In this example there is a:

1. Common logical transport infrastructure supporting the Fronthaul and Midhaul management planes for all three NSIs
2. Common logical transport infrastructure supporting Fronthaul control and user planes for all three NSIs.
3. Common logical transport network supporting Midhaul user plane for all three NSIs.
4. Common logical transport network supporting 3GPP control plane (Midhaul and Backhaul) for all three NSIs.
5. Dedicated logical transport networks for backhaul and N6 networks for each NSI.

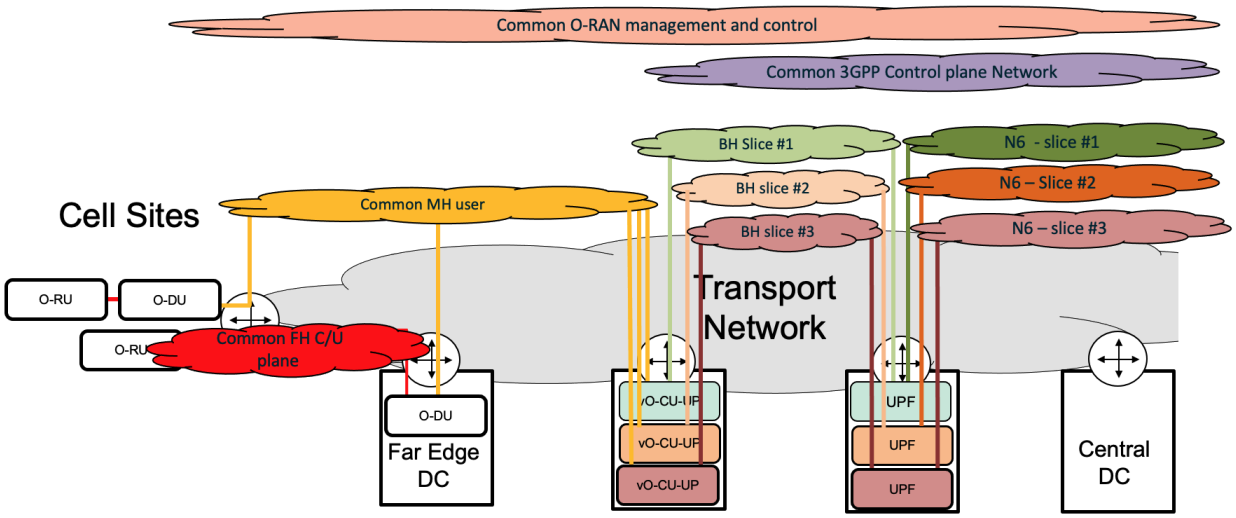


Figure 18-2: Overall 5G slicing architecture

The characteristics of a transport slice are defined in clause 5.2.3 of 3GPP Technical Report 22.891. Although not referenceable, further useful information can be found in various expired personal informational IETF drafts which provide thinking on the characteristics of a sliced transport network. Some of the key points are:

- Management and lifecycle of the network
  - Definition
  - Creation / deletion
  - Modification
- Per slice OAM
- Resource Reservation
- Slice isolation
  - Performance
  - Operational
  - Security
  - Reliability
- Abstraction
  - Virtualization of network functions (where appropriate)
  - Use of shared compute resources

In the transport space, the informal terms, hard and soft slicing has emerged. This refers to the level of isolation between different slices. In both cases they need to support the functions outlined above but the way the slice is built and managed differs considerably.

- Hard slicing: Transport resources are dedicated to a specific "Network Slice Instance" (NSI). In this case a resource is dedicated to a particular slice and not available to other slices.

- Soft slicing: The transport slice has the characteristics outlined above but a resource is shared and can be re-used by other slice instances.

A packet switched infrastructure, as described in this document, has an extensive toolset, consisting of underlay forwarding solutions, Quality of Service (QoS) and VPNs that allows an operator to scalable partition the transport network to cater for both hard and soft slices use cases. Figure 18-3 outlines some of the packet switched features that can be flexibly combined to create transport slice with different level of resource sharing.

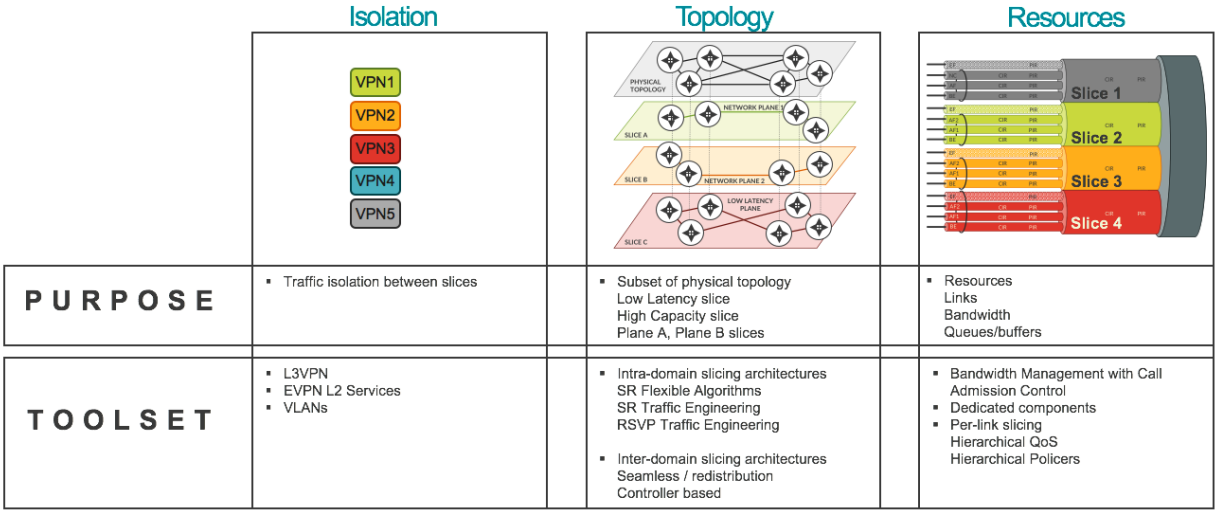


Figure 18-3: Packet switched toolset for transport level slicing

When considering hard and soft slices it is important to bear in mind they should be thought of as a spectrum of capability, rather than one or the other.

18.1 Packet-switched underlay network

The packet-switched underlay network can use MPLS, SRv6 or a combination.

18.1.1 Underlay transport plane

The transport plane determines how traffic is sent over the packet-switched underlay infrastructure. Four approaches are outlined for constructing the underlay transport plane/planes for a 5G transport infrastructure. In each of the first three cases the level of resource sharing reduces, hence the slice solution becomes harder in nature but may introduces scalability and operational challenges. Fourth mapping approach is based on 5QI/DSCP, and scaling doesn't depend on the number of slices.

18.1.2 Single transport plane for all slices

The underlay network relies solely on the routing protocols (IGPs and EGPs) to calculate a single forwarding table based on shortest path routing. The routing protocol sees all links and there is a single forwarding table based on IGP and BGP metrics. All traffic takes the shortest path between two points within the network and utilises ECMP.

In the context of slicing, all slices will see the same set of paths between end points within the network. It can be considered the softest slicing solution in terms of the underlay transport plane. However, although the infrastructure is shared between slices, it is highly scalable, widely implemented in 4G and critical enterprise environments, and is capable of delivering high quality SLAs.

18.1.3 Transport plane per 5G service type

Within the underlay network multiple transport planes are built to meet the specific forwarding behaviors associated with the different 5G services types. One or more customers can then use these transport planes. This is achieved using VPNs and traffic steering techniques. The transport planes can utilise different topologies and be optimized based on different criteria. For example, a URLLC service types could be optimised to use only the most reliable links in the network and select the best paths between endpoints based on link delay metrics. In contrast, the eMBB service types could be designed to use cheap, high bandwidth links, based on paths calculated with IGP metrics correlated with link capacity (high link capacity → low IGP metric). And, NB-IoT service types, which don't require neither low latency communication, nor high capacity, could use a 3rd transport plane with paths established based on TE metrics, favoring links for this type of services. This concept is outlined in Figure 18-4.

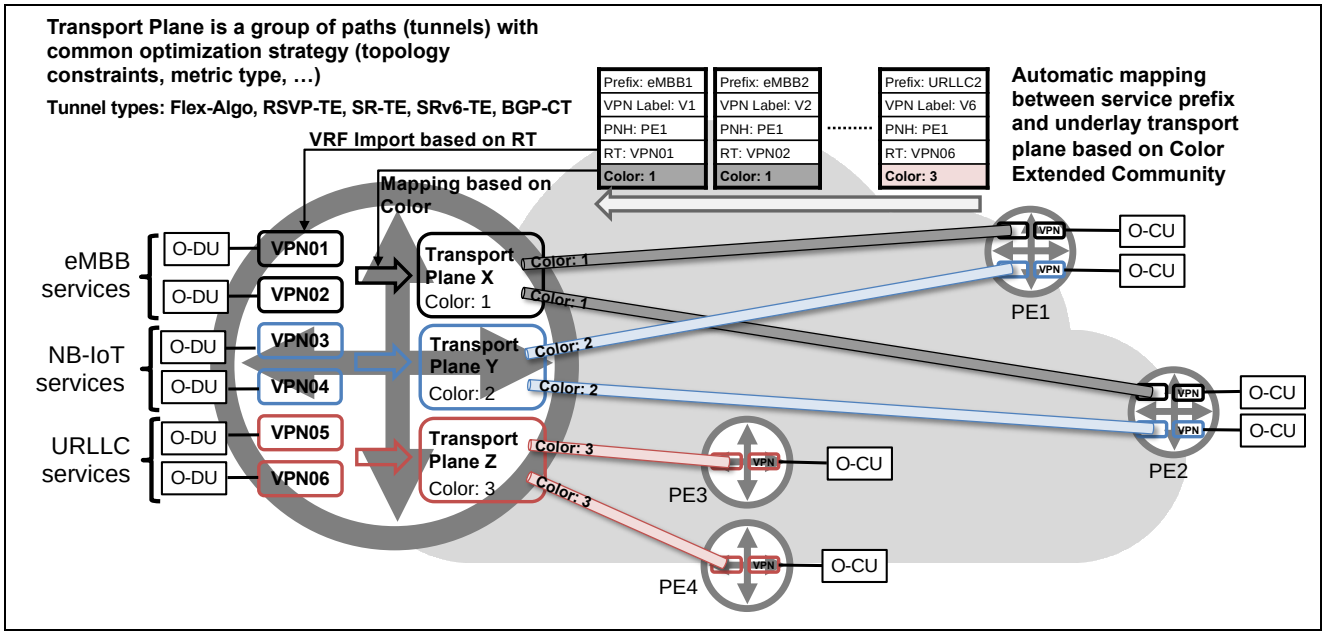


Figure 18-4: Transport Plane per 5G service type – midhaul example

To implement such an approach, a mix of shortest path routing and traffic engineering solutions could be used. Traffic engineering options include flex-algo, SR-TE or traditional MPLS TE. As with any traffic engineering approach, consideration needs to be made of scale and state held within the network. Perhaps the simplest approach would be to use the default Flex-algorithm for eMBB and the mMTC services and a delay optimized flexible algorithm for the URLLC services. Backbone links and even nodes could be considered for inclusion or exclusion into the delay optimized flex-algorithm transport planes based on whether the algorithm is enabled on a TNE and TE affinities associated with the links.

The mapping between services and transport planes can be simplified with the usage of Color Extended Community defined in RFC 5512 [100]. Each transport plane (each tunnel being part of specific transport plane) is associated with some color value. For example, latency optimized tunnels use color 1 (Transport Plane A), capacity optimized tunnels use color 2 (Transport Plane B), and remaining tunnels use color 3 (Transport Plane C). At the same time, service prefixes are advertised via BGP with appropriate Color Extended Community. URLLC service prefixes use community value 1, eMBB service prefixes community value 2, and NB-IoT service prefixes community 3. On the PE router receiving such service prefixes, selection of appropriate transport plane is automatic based on matching colors advertised via Color Extended Community with the service prefix and the color associated with the transport plane.

18.1.4 Transport plane per slice customer

This is a variation on the previous scheme. In this case rather than a transport plane per 5G service type, a transport plane for individual customers is defined, as depicted in Figure 18-5. The same techniques outlined above would be used but very careful consideration needs to be given to scale and operational complexity of such an approach. In this case, scalability is a function of the number of customers using the network rather than the 5G service types.

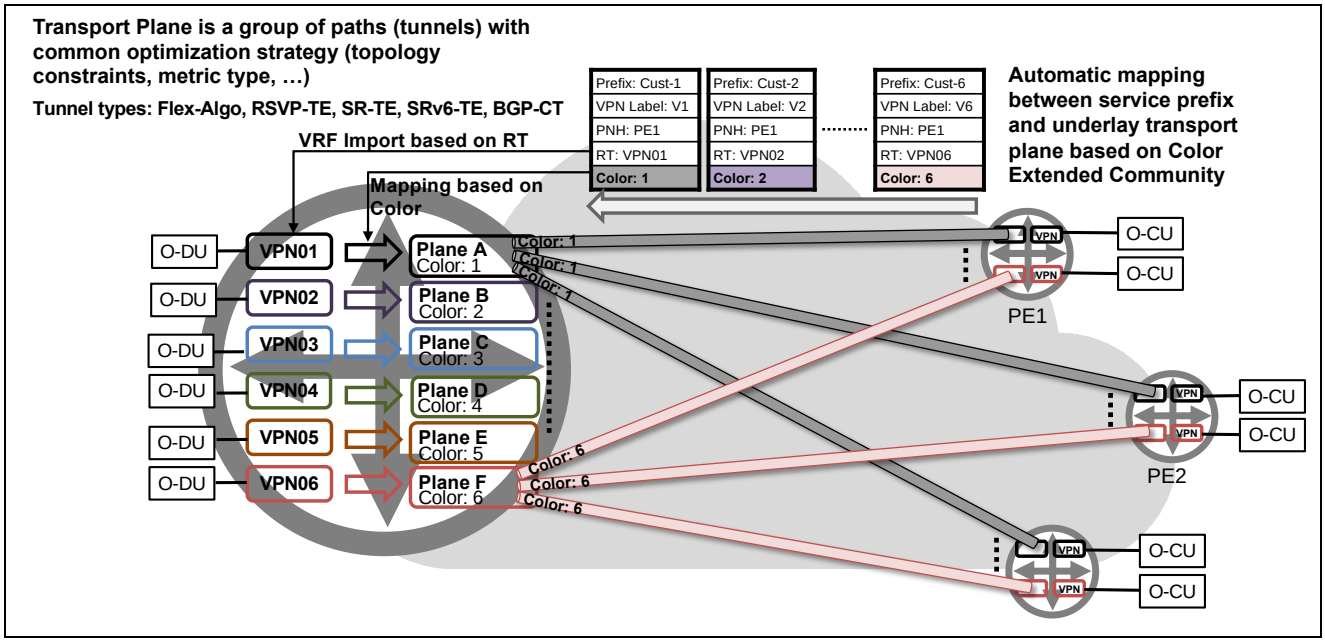


Figure 18-5: Transport Plane per slice customer – midhaul example

Mapping per customer and mapping per 5G service type might be used together to address scaling challenges. E.g., as the default approach mapping per 5G service type might be used, and only small number of selected premium customers might have mapping to individual transport planes.

18.1.5 Transport plane per 5QI group (5QI-aware mapping)

5G slices might carry traffic belonging to multiple 5QI flows. Annex F: Transport network slicing solution for WG1 Slicing (informational) proposes an approach for mapping between 5QI and DSCP, and for grouping multiple 5QI/DSCP flows to limited number of 5QI/DSCP groups, where each group contains 5QI/DSCP flows (from multiple slices) with similar characteristics. This grouping is introduced to allow easy mapping of large number of 5QIs to limited set of transport resources, like for example queues in TNEs. 5QI/DSCP grouping allows as well for intelligent transport plane mapping, as outlined in Figure 18-6.

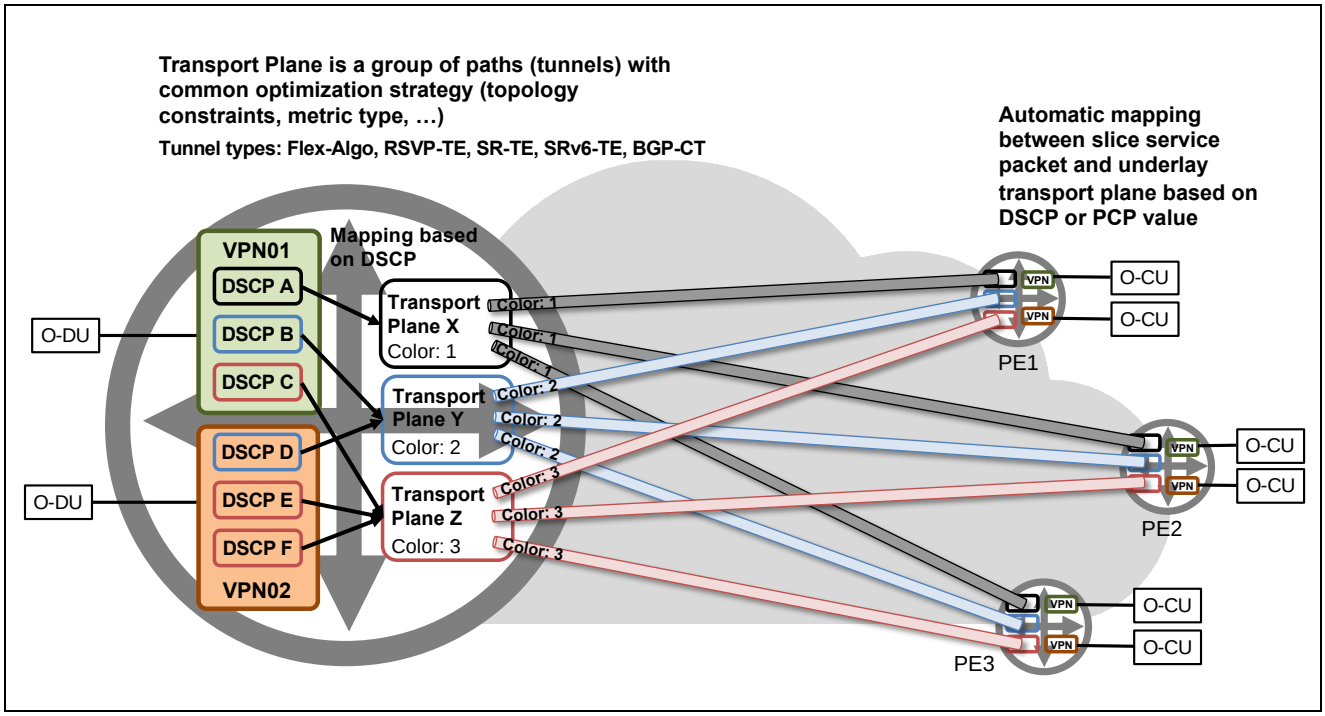


Figure 18-6: Transport Plane per 5QI group (5QI-aware mapping) – midhaul example

In this example, three 5QI groups (group X: DCSP A; group Y: DSCP B, D; group Z: DSCP C, E, F) are uniquely mapped to three transport planes. With this approach, Color Extended Community is no longer used, as the mapping is based on DSCP (which corresponds to some 5QI). 5QI-aware mapping is most advanced mapping option between the service flows and underlay transport planes.

18.2 Quality of Service

Quality of Service is an important component in slicing a transport infrastructure. As with the transport plane, different approaches could be taken depending on the level of isolation required between slices. In considering QoS it is necessary to look at the edge QoS solution and the core QoS solutions. For more info on QoS refer to section 14 and 23.

18.2.1 Edge QoS

Edge interfaces to packet transport networks are generally the slowest and also where congestion and packet drop most often occurs. Regardless of the QoS structures implemented in the transport core, it is important with slicing to support edge conditioning of traffic on ingress and scheduling on egress to the transport network. This ensures that each slice gets its contracted overall bandwidth and class bandwidth as it enters and leaves the transport network. When slices are presented via VLANs from the mobile clients, the PE router needs a hierarchical QoS capability that can account for the overall contracted bandwidth at the VLAN level and also the overall contracted class bandwidth within the VLAN.

18.2.2 Core QoS

Within the core network different QoS strategies can be applied that determine the level of bandwidth and queue sharing that occurs between different slices.

18.2.2.1 Shared queue architecture

This approach follows the classic Diffserv QoS architecture. Slice traffic would be conditioned, and optionally marked or remarked as it enters the Diffserv domain, in this case the transport network. These markings determine the behaviour or Per Hop Behaviour (PHB) the traffic receives in the transport network. The number of PHBs in the core is determined by a number of factors; the number of PHBs that can be marked, the number of queues supported on the TNEs and operational complexity of the solution. Typically, in an MPLS network the traffic-class field is used to designate the PHB in the core. This is 3 bits, giving a maximum of 8 markings. In an IPv6 network the DSCP field is used to designate the PHB in the core. This is 6 bits giving a maximum of 64 markings. Even with more marking options, the number of PHBs in the core is generally kept small (<8) to simplify configuration, management and is generally done as a one-time set-up when a link is first installed. Core PHBs are enabled using a combination of queues, congestion management and scheduling techniques and are normally set-up in a work preserving fashion. That is, if one PHB is not fully utilised, the unused capacity can be used by traffic with other PHBs.
Applying this to slicing. Each slice has an SLA based on overall and class bandwidth it receives. This is enforced on ingress at the PEs. In this QoS model, core queues are shared between slices and can be considered to be a soft form of slicing, but there is protection between slices through ingress edge conditioning of traffic. This architecture it is well tried, and tested, is simple and scales and able to support very tight SLAs.

18.2.2.2 Dedicated queue architecture

This approach uses the same Diffserv QoS architecture, except one or more PHBs is exclusively dedicated to a particular slice or slice type. An example could be a high priority PHB, which is dedicated exclusively to the URLLC slice. At the same time other slices can share the remaining PHBs. Given the small numbers of queues normally provisioned in the core (and available traffic classes in an MPLS environment), this type of solution needs to be used with extreme care, however it is a model used today successfully in production networks to support private line services which run in a dedicated high priority queue, while simultaneously supporting other services using a shared queue architecture.

18.2.2.3 Dedicated links and queues per slice

This QoS approach achieves hard slicing by either dedicating core links or parts of a core link to a specific slice. This can be achieved by one of the following approaches:

- Using dedicated links per slice in the transport network. This could be done using full physical Ethernet interfaces or using Ethernet TDM technology, such as flexE, to create dedicated Ethernet channels for different slices. Within these links, class-based queuing would be used to support different traffic types associated with the slice. This approach offers no ability to share unused bandwidth between slices as the links are hard partitioned.
- Using hierarchically scheduled bandwidth on core links. In this case core links would be split using a shaper to govern the level of bandwidth available to each slice, with class-based queuing within the shaped bandwidth. This approach does offer the ability to share unused bandwidth between slices.

In addition to dividing the core bandwidth up, mechanisms are required to guide slice traffic into the links or shaped bandwidth associated with each slice. Slice traffic could be guided into the appropriate links using a TE solution, such as flex-algo or SR-TE or alternatively the core routers

could be enhanced to schedule on a slice identifier within the packet. At this time there is discussion on this subject but as yet nothing solid. Like with the previous example, this could support a small number of slices but scaling up would be highly problematic with core provisioning, as well as edge provisioning required when a slice is defined, changed or deleted. It is also a large departure in the way QoS in core of packet networks has been managed in the past.

18.35G Services and slices

MP-BGP based L2 EVPN and L3VPNs are used to create slice instances. VPNs are very scalable in terms of numbers and endpoints and allow different flexible connectivity models. This means slices could be based on 5G service, customer or a combination of the two.
Each slice would have a VPN associated with it. By default, VPNs use the default IGP based forwarding tables to forward traffic. However, VPN traffic can be directed, either by configuration or using automated traffic steering techniques based on BGP "color extended community attribute" to use traffic engineered paths. In this way, a URLLC slice could be directed to use a delay optimised TE solution. Further control of the transport level connectivity within the VPN can be achieved using Route Target filtering. For example, "Route Target" filtering could be used to stop transport level connectivity between UPFs of different customers but allow O-CUs to communicate with any UPF.

19 Supporting mobile scenarios on a packet switched Xhaul network

What is clear in discussions with operators is that although section 8 outlines individual use cases, operator's expect multiple use cases to be present in their transport network based on the need to support 4G and 5G, maturity and timelines of the O-RAN specifications, vendor product readiness, use cases and latency requirements. Figure 19-1, an adapted operator's view, illustrates this well with multiple 4G/5G architectures present along with Enterprise and consumer services.

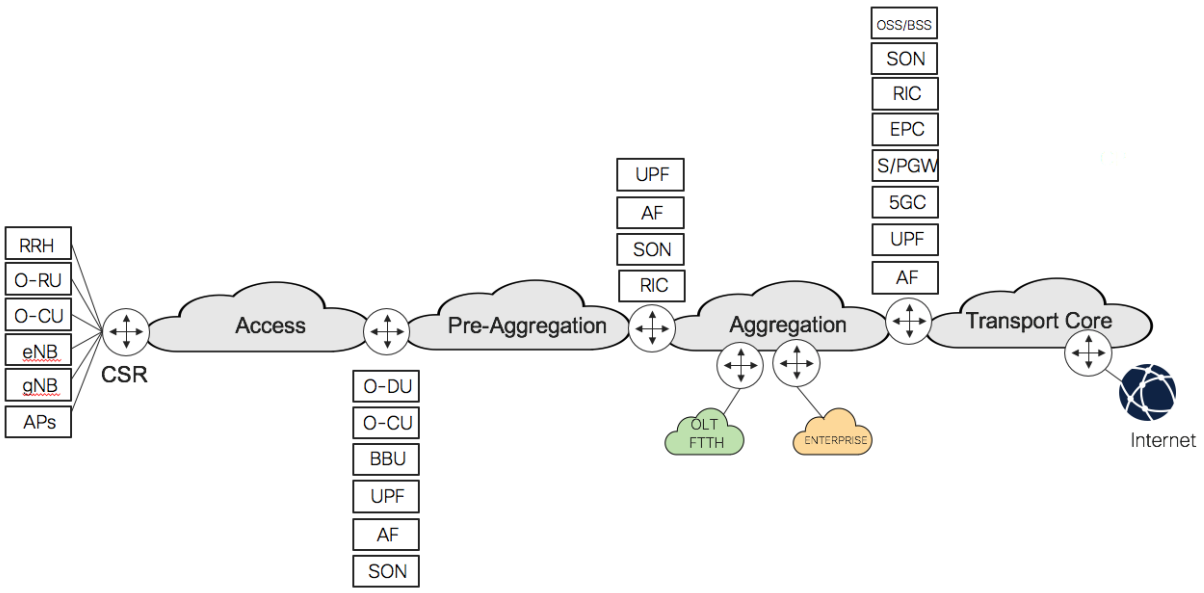


Figure 19-1: An adapted O-RAN operator's view of mobile component placement

In order to make this document generic to operators, simplification is required. For this reason, the following physical topology and underlay transport solution is used as a baseline. Operators can then use the information and customize to their specific requirements.

19.1 Physical network

Figure 19-2 illustrates physical network layout for a metro network and its connectivity to the transport core. It consists of:

1. A collection of metro networks surrounding a transport core. Each metro network consists of a number of access networks that are consolidated by an aggregation network towards to the transport core.
2. All sites which interface to customer or mobile components supports Provider Edge functionality for L2 and L3 overlay services.
3. Data centers or location for positioning mobile components potentially exist at the hub sites, edge sites and in centralised locations.
4. Connectivity between routers is provided by line-rate point-to-point Ethernet connections derived from dark fibre or WDM.
5. Access networks could be based on either small rings, star, chained or hub and spoke topologies. To simplify the following discussion, it is assumed each cell site has a CSR connected via a single high-capacity dark fibre to the HSR in a star topology. Refer to section 10 for more details on potential fiber access topologies.
6. The Aggregation network originates from the HSR and is based on either a ring or a hub and spoke topology. For some scenarios the aggregation network may be split into a pre-aggregation and aggregation component. In this case, there is an intermediate layer of packet switches and an additional layer of statistical multiplexing. (See greyed out router / switch in Figure 19-2)

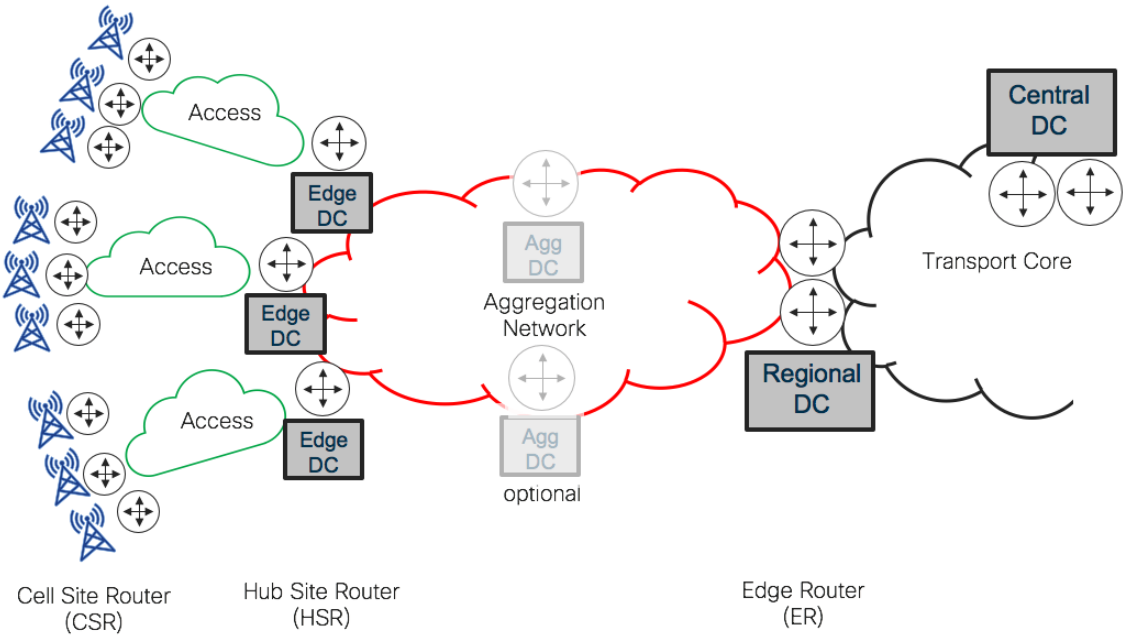


Figure 19-2: An example physical architecture

19.2 Logical underlay architecture

The logical underlay is illustrated in Figure 19-3 and consists of QoS enabled infrastructure that either uses an IP/MPLS or an SRv6 underlay data and control plane. It provides any-to-any connectivity between TNEs in the network, regardless of whether they reside in the same metro network, a different metro network or the transport core and supports both shortest path routing and traffic engineering. The underlay data planes, control plane and scaling mechanisms to enable this for IP/MPLS and SRv6 are different and discussed in their respective chapters. Both solutions have the capabilities to support best effort and traffic engineered forwarding and the capability to create discrete constrained based topologies using either traditional traffic engineering or flexible algorithm which can be optimized on criteria such as shortest path, delay and bandwidth.

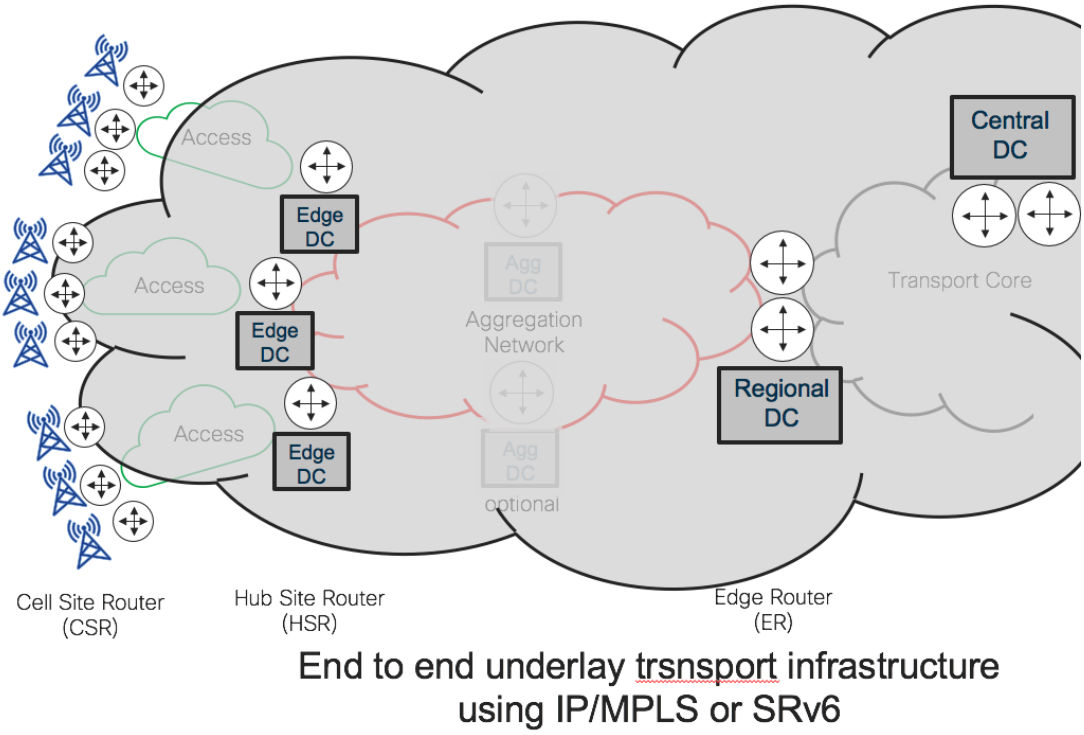


Figure 19-3: An example logical architecture

19.2.1 Underlay Quality of Service (QoS)

Refer to section 14 for fuller discussion on QoS. Underlay QoS refers to the scheduling used on the core or backbone links. All backbone links are enabled with a combination of priority and class-based queueing. Assignment of traffic to different queues within the core is based on marking within the packet header. In an MPLS backbone this is based on the EXP bits or traffic class field, a three-bit field, giving a maximum number of core behaviours of eight. In an IPv6 backbone this is based on the DSCP bits, a six-bit field, giving a maximum number of 64 markings and potential behaviours.

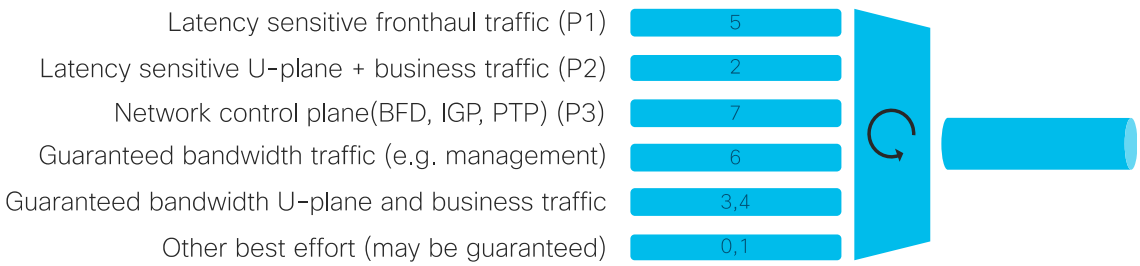


Figure 19-4: Example Xhaul queue structure

19.3 Service architecture

For mobile services it is proposed to use BGP based VPNs to overlay service functionality onto the underlay transport network as described in section 13.

NOTE: There is no reason this same transport infrastructure cannot support other VPN solutions such as VXLAN or SD-WAN type solutions.

All PEs and DCIs (which are also PEs) should be capable of supporting point to point Ethernet services using EVPN VPWS services and L3 BGP VPN services described in section 13. With these two basic service overlays all the scenarios outlined in section 8 can be accommodated.

19.3.1 Automated VPN Traffic Steering

If required traffic from different VPNs or specific flows within an VPN can be steered into different underlay transport planes based on coloring the MP-BGP VPN routes associated with L2 and L3VPNs. This capability is described in *Annex C: MP-BGP based L3VPNs* .

19.4 Mobile services

There are many ways Fronthaul, Midhaul and Backhaul services could be provisioned onto the transport network. Below are a set of assumptions and mechanisms used to illustrate how a packet switched architecture can accommodate the scenarios outlined. These can easily be adapted based on operator's individual requirements.

19.4.1 Open Fronthaul

Figure 19-5 illustrates a physical and logical design that supports Open Fronthaul traffic. It is fairly simple in nature, however, can easily be adapted based on operator designs and requirements. It uses the underlay building blocks discussed earlier chapters.

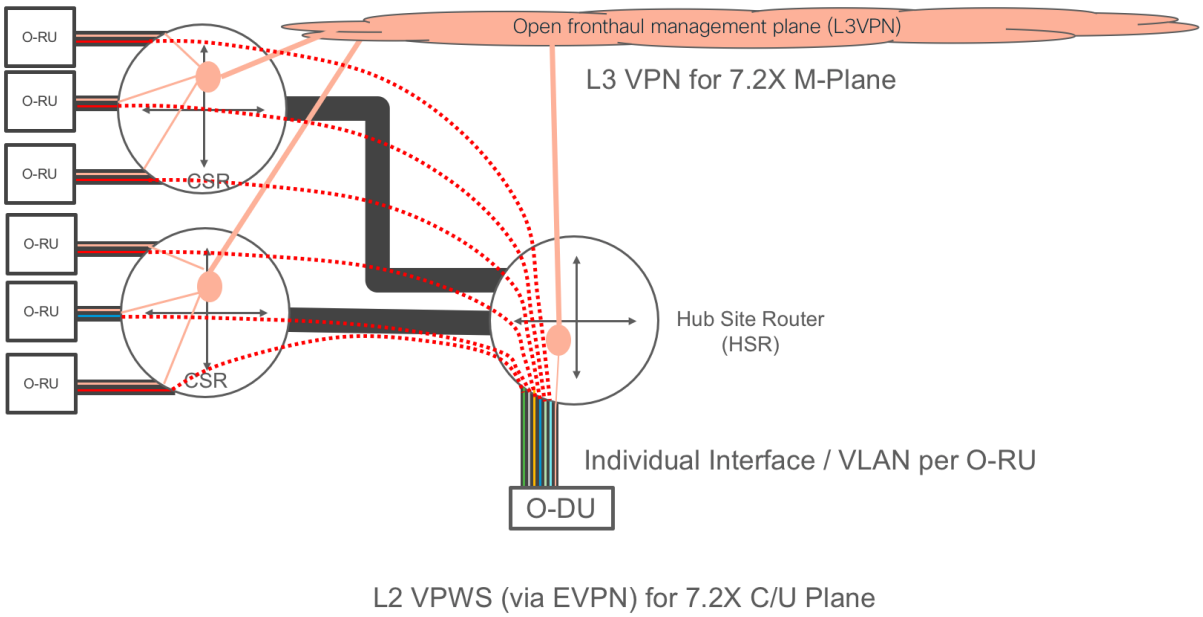


Figure 19-5: Fronthaul physical and logical topology

19.4.1.1 Assumptions

1. 7.2x C/U planes connects to the transport network via a VLAN on a physical Ethernet interfaces on O-RUs and O-DUs.

2. 7.2x C/U planes traffic uses Ethernet / VLAN encapsulation option and not the optional Ethernet / VLAN / IP encapsulation.

3. 7.2x C/U planes uses an Ethernet service across the transport network to connect O-RUs to O-DUs. This is via an Ethernet VPWS service using EVPN in the transport network.

4. If the 7.2x C/U planes traffic used an Ethernet / VLAN / IP encapsulation, then the transport network could use a L3 VPN service to transport the traffic between the O-RU and the O-DU, but this is not considered in the following sections.

5. 7.2x M-plane connects to the transport network via a VLAN on a physical Ethernet.

6. 7.2x M-plane uses an Ethernet / VLAN / IP encapsulation.

7. 7.2x M-plane is operating in hybrid mode and uses an L3 VPN service across the transport network to connect O-RUs, O-DUs and O-RAN NMS together. This is via an overlay MP-BGP based L3VPN.

8. For the following discussion, the 7.2x C/U planes and M plane share a single physical Ethernet interface on the O-RU towards the CSR with the C/U plane and M-Plane running in different VLANs.

9. Each O-RU has its own dedicated Ethernet port on the CSR. I.E There is no lower level aggregation of statistical multiplexing component between an O-RU and the CSR.

10. O-RU and O-DU set appropriate QoS markings for C/U and M-plane traffic that the transport network can trust. In other words, the TNE does not need to classify different components of Fronthaul for transport.

19.4.1.2 Open Fronthaul physical design

1. 3 O-RUs per cell site.

2. In the above example, each O-RU is connected via a dedicated Ethernet interface to a co-located CSR. This single Ethernet supports both the C/U planes and the M-plane presented through two VLANs. The required capacity of the Ethernet link between the O-RU and the CSR is primarily dependent on the O-RU’s radio capabilities. The requirements document [19] provides some bandwidth examples for Open Fronthaul bandwidth. The expectation is that the C/U plane traffic will be significantly larger than the M-Plane traffic, therefore these Ethernet interfaces should be provisioned to deal with the peak 7.2x C/U-plane traffic of an O-RU. It should be noted that unlike CPRI, the 7.2x C/U plane traffic loads are dependent on user activity, so peak loading will only occur when the O-RU is running at full theoretical capacity.

3. In the above example, each CSR is connected using a single Ethernet link across dark fibre to the HSR. Other underlay transport technologies and topologies are available, such as redundant hub and spoke, chain or ring topologies in the access network. Regardless of the WAN technology and topology employed, careful consideration needs to be given the latency and jitter requirements of C/U plane traffic between the O-RU and the O-DU. Transport considerations includes fibre distances between the O-RU and the O-DU, the number of routers/switches traversed, link speeds and switching times of the intermediate transport network equipment. (see section 10.1.1.2 for more details). As O-RAN 7.2x traffic loads are dependent on user activity and in the example there are multiple O-RUs in the cell site, the link capacity between the CSR and HSR does not need to be provisioned at the theoretical sum of the O-RU's peak rates because a statistical gain can be realised. The bandwidth provisioning should be based on usage and the real achievable peak rates associated with the O-RUs. The real achievable peak rates are dependent on factors such as number and proximity of radios to each other and environmental factors. Precise provisioning rules for 5G Fronthaul is subject to further investigation.

19.4.1.3 QoS

1. The QoS scheme outlined in section 19.2.1, Figure 19-4 will be applied across the transport network and is capable of supporting Fronthaul, Midhaul and Backhaul traffic running on the same link. If an SRv6 underlay is to be used, then at a minimum the same queue structure supported by appropriate operator selected DSCP marking should be used.
2. O-RU and O-DU set appropriate QoS markings for C/U and M-plane traffic that transport network can trust.
3. In the above example, the C/U and M-plane traffic share the same physical interface, separated using VLANs between the O-RU and CSR and between HSR and O-DU. In this situation the mobile components (O-RUs and O-DUs) will need to support:
    - An internal scheduling mechanism to prioritise C/U plane traffic over M-Plane traffic onto the Ethernet interface
    - Depending on O-RU←→ CSR and HSR←→ O-DU link speeds, the O-RU, O-DU, CSR and HSR may need to support TSN frame pre-emption to control serialization delay. (refer to section 10.1.1.2.1 for more details). Alternatively, if the O-RU or CSR do not support TSN frame pre-emption or the O-DU or HSR do not support TSN frame pre-emption, then separate physical interfaces could be used to separate the C/U plane and M-plane traffic.
5. C/U and M-plane traffic from all O-RUs in the cell site share the same transport link between the CSR and the HSR. In most scenarios where a CSR supports multiple radios running fronthaul, this link will be 25Gbps or above and therefore does not really benefit from TSN frame pre-emption. However, in some instances, the link between the CSR and HSR may be lower speed and would benefit from TSN frame pre-emption. (see section 10.1.1.2.1 for more details)
6. C/U plane traffic must be prioritized over M-plane traffic on the WAN links. CSR and HSR must be provisioned at both the link level and within the link level queuing infrastructure in a manner such that C/U plane traffic is not dropped or experiences queuing delay in either the CSR or HSR (or any additional Transport Network Equipment in the path between the O-RU and the O-DU). M-plane traffic can experience a level of delay, so be configured to use a lower priority queue or a queuing scheme such as Weighted Round Robin that reserves a small amount of bandwidth for management traffic (see section 14 for an example)

19.4.1.4 Services

1. C/U plane traffic uses an EVPN VPWS service to transport traffic between the O-RU and O-DU over the access network. To simplify cell site VLAN provisioning, common VLANs could be used between the O-RUs and the CSR and VLAN tag translation could occur on the HSR to provide a unique interface / VLAN id per O-RU towards the O-DU. Depending on the capabilities of the O-DU, EVPN has mechanisms that can provide service level resiliency but this is dependent on the O-DU network implementation.
2. The M-plane traffic on the O-RU and O-DU connects, via a VLAN or physical interface to a MP-BGP based L3 VPN service. This provides layer 3 connectivity between the management components on the O-RU, O-DU and the O-RAN "Network Management System" (NMS). "Route Target" filtering could be used to restrict transport level connectivity between management entities. For example, the transport network could be configured such that IP connectivity is only permitted between the NMS and the O-DUs and O-RUs, while no IP connectivity is permitted between O-RUs.

19.4.2 Non O-RAN Fronthaul

The assumptions, design approach and considerations discussed for the C/U planes in Open Fronthaul are used for Non O-RAN Fronthaul.

19.4.3 Midhaul and Backhaul

Figure 19-6 illustrates a physical and logical design that could be employed to support Midhaul and Backhaul traffic. It is a simple example but can easily be adapted based on operator designs and requirements. It uses the underlay building blocks discussed earlier in the chapter.

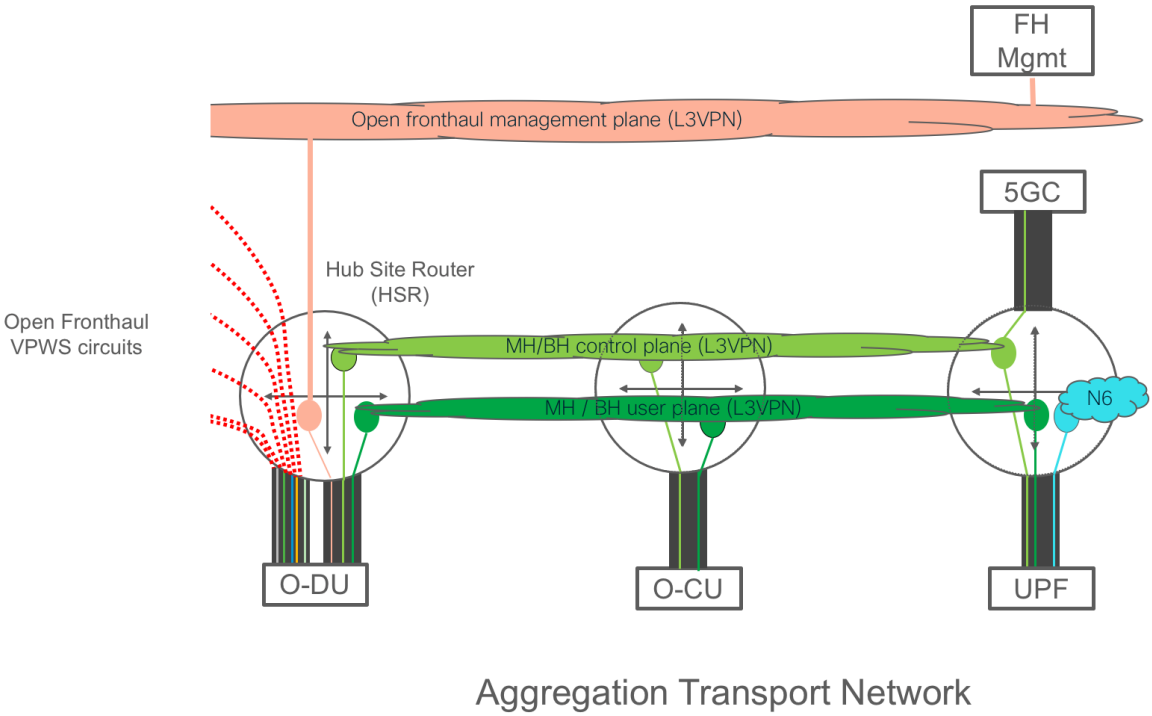


Figure 19-6: Midhaul and Backhaul logical topology

The following design assumptions have been made about the Midhaul / Backhaul components and how traffic is presented and moved across the transport network.

19.4.3.1 Assumptions

1. Midhaul and Backhaul traffic share the same L3 VPNs.
2. There are dedicated and separate L3 VPNs for control plane traffic (N1, N2) and user plane traffic (N3 and N9).
3. If legacy Backhaul is required, it shares the same L3 VPN transport services used by the 5G EMBB Midhaul and Backhaul user and control plane infrastructure.
4. If there are multiple Midhaul and Backhaul transport slices, they could run in their own independent L3 VPNs (see Annex F: Transport network slicing solution for WG1 Slicing (informational))
5. Midhaul and Backhaul mobile components will either use a discrete independent VLAN or a dedicated physical Ethernet to separate control and user plane traffic. For the following discussion it is assumed they are presented via VLANs on a single Ethernet interface.
6. Midhaul and Backhaul mobile components which require connectivity to different transport slices will either use a discrete VLAN or a dedicated physical interfaces per slice. For the following discussion it is assumed they will be presented via VLANs.
7. The transport and RAN teams have collaborated on Midhaul / Backhaul QoS marking and packets emerging from O-DUs, O-CUs and UPFs have a DSCP marking appropriate to their 5G forwarding requirements, and the TNEs can trust these markings.
8. N6 networks associated with different customers, applications or use cases will use L3 VPNs.

Note: Other service designs could be considered, for example a design where only Ethernet services are used from the cell site and the hub sites and L3 VPN services are built at the hub sites. The primary reason for the choice outlined above is to make the service orchestration as simple as possible, so services only need to be configured where the mobile clients connect to the transport network and no intermediate stitching points within the transport network are needed.

19.4.3.2 Midhaul and Backhaul physical topology

1. O-DUs connect, via Ethernet interface/interfaces, to a co-located HSR or CSR. The Ethernet interface/interfaces support Midhaul control and user plane traffic via two discrete VLANs.
2. O-CUs connect, via Ethernet interfaces to the transport network. The O-CU Ethernet interface/interfaces support Midhaul and Backhaul control and user plane traffic via two discrete VLANs.
3. The UPFs connect, via Ethernet interfaces to the transport network. The UPF connects to the common control and user plane Midhaul/Backhaul VPNs.
4. The location of the O-CUs and UPFs is service dependent and the choice of the operator. Additionally, these components may be virtualized or containerized and reside in a data center. In these scenarios, orchestration techniques maybe be required to connect the virtualized O-DU, O-CU and UPF components in the data centers to the correct WAN L3VPNs.
5. 5G Core components connect via Ethernet interfaces to the transport network in their respective locations. These components may be virtualized or containerized and reside in a data center. In these scenarios, orchestration techniques are required to connect the 5G control plane entities in the data centers to the Midhaul and Backhaul control plane WAN L3VPNs.

6. O-CU can reside in the same location as the O-DUs it controls, therefore Midhaul traffic never leaves the site and is switched between the two entities by the either the CSR or HSR. Alternatively, the O-CU can reside in a different location to the O-DUs it controls. In this case the Midhaul traffic traverses the WAN.
7. HSRs provides the boundary between the access and aggregation networks. Connectivity to the aggregation network will typically use high-capacity Ethernet interfaces or bundles of high capacity Ethernet interface. The topology of the aggregation network is normally resilient, but topologies vary based on operator choices and constraints, with hub and spoke and rings very common.
8. Midhaul and Backhaul user plane traffic levels are dependent on UE usage, the characteristics of the radio and the number the proximity of radios to each other. The O-RAN Transport Requirements document [19] outlines some of the provisioning rules widely used in planning 4G Backhaul networks. Although not directly applicable to the 5G, they provide a good starting point in terms. Longer term network capacity planning should be done based on monitoring and modelling the live network.

19.4.3.3 QoS

1. O-CU and UPFs mark the packets they produce based on the behaviour required for the mobile data the carry.
2. Most Midhaul and Backhaul traffic is far more tolerant to delay, and jitter compared with Fronthaul traffic. The exception could be some URLLC based use-cases. However, for most 5G use-cases fibre propagation delay, number of hops, transport equipment switching times and serialization delays are not of a huge issue when designing Midhaul and Backhaul networks.
3. In some scenarios, particularly when some forms of slicing are considered then edge QoS in the form of ingress and egress H-QoS will be required.

19.4.3.4 Services

1. Two L3 VPNs are created; the first is Midhaul/Backhaul control plane traffic and the second is Midhaul/Backhaul user plane traffic. O-DUs, O-CUs, UPFs connects to both L3VPNs. The 5GC connect only to the control plane L3VPN. If required “Route Target” filtering could be used to restrict transport level connectivity between mobile entities.

19.5 Scenario 1 and 5

Both scenario 1 and 5 represent a full two split C-RAN architecture. The main difference is the location and proximity of the O-DUs and the O-CUs to each other. In scenario 1 they are contained in a single hub site location, in scenario 5 they are split across different locations.

At the high level the basic service layout for scenario 1 and 5 is shown in Figure 19-7 and Figure 19-8 with a more detailed layout shown in Figure 19-9.

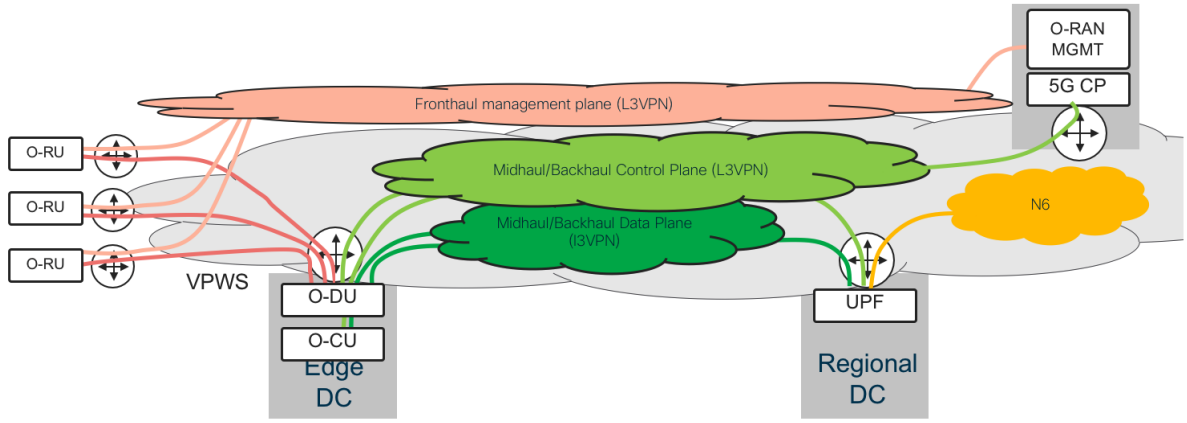


Figure 19-7: Scenario 1 layout

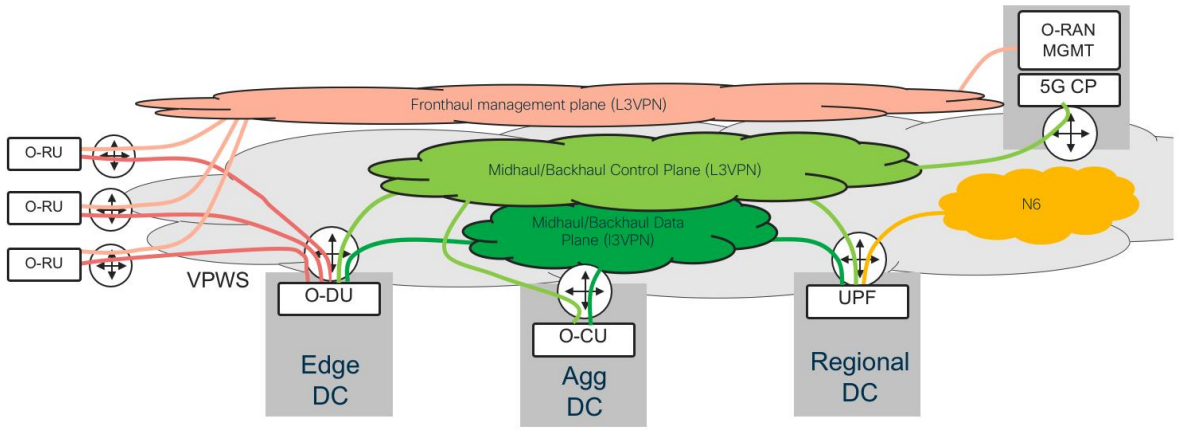


Figure 19-8: Scenario 5 layout

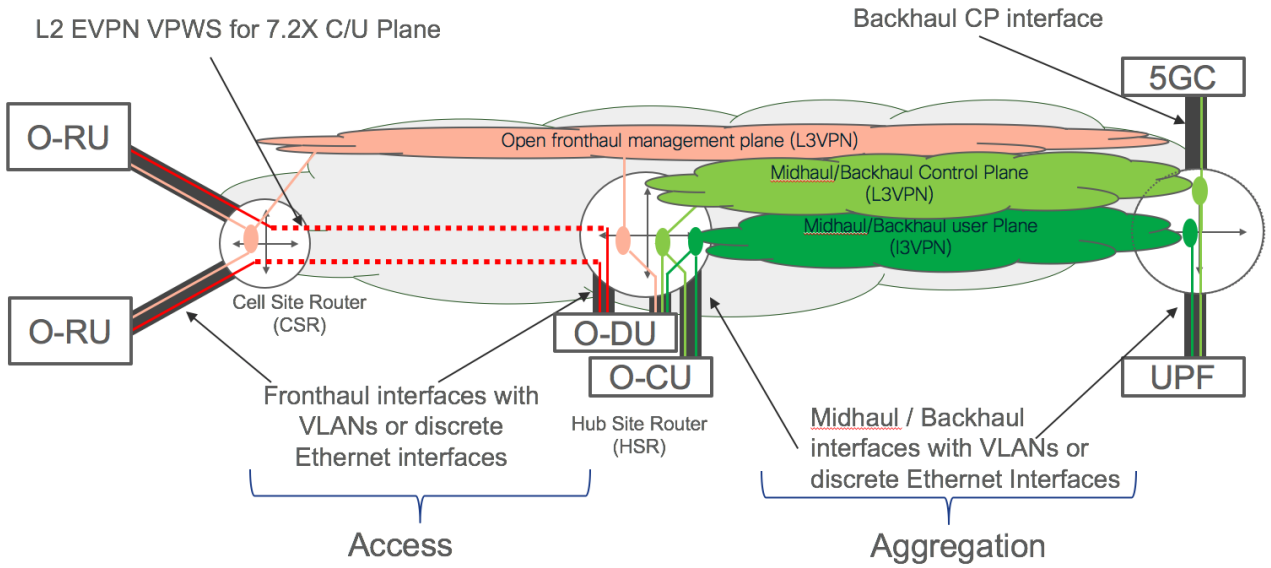


Figure 19-9: L2 / L3 service layout for scenario 1 and 5

19.6 Scenario 2

Scenario 2 represent a full two split C-RAN architecture with the O-RU and O-DU at the cell site and the O-CU at the hub site and is shown in Figure 19-10.

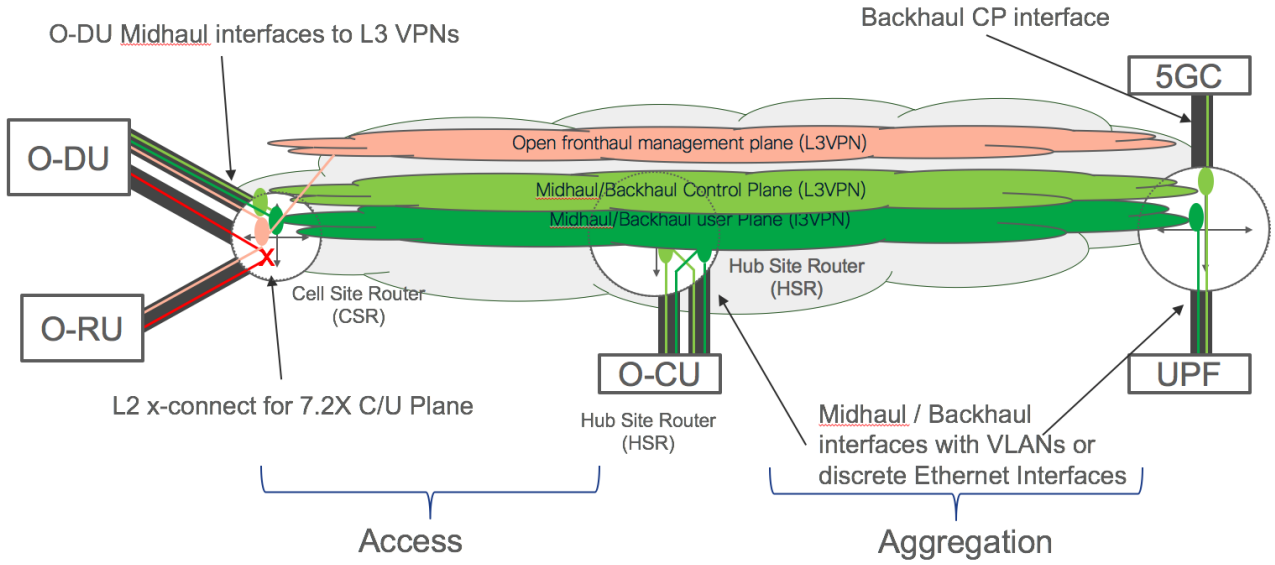


Figure 19-10: L2 / L3 service layout for scenario 2

Notes on the design:

1. Fronthaul traffic is restricted to the cell site with the CSR providing a local L2 x-connect function to get C/U plane traffic between the O-RUs and the O-DU.
2. The Midhaul/Backhaul control and user plane L3 VPNs need to be extended to the cell site router.

3. In above diagram the O-DU has two physical interfaces with one supporting Fronthaul traffic and the other supporting, via VLANs, Midhaul and Backhaul control plane, user plane and Open Fronthaul M-plane traffic. In this instance TSN frame pre-emption is not really a consideration. If the Midhaul, Backhaul and Fronthaul traffic shared the same physical interface on the O-DU and the CSR, then depending on interface speeds, TSN frame pre-emption may need to run between the O-DU and the CSR.

19.7 Scenario 3 5G C-RAN with legacy D-RAN

19.7.1 Scenario 3a

Scenario 3a is a full two split C-RAN architecture with the O-RU at the cell site and O-DU at the hub site sitting alongside an existing 4G D-RAN architecture and is shown in Figure 19-11. O-CU, UPF and 5GC location are largely immaterial but they do need connectivity to the appropriate L3VPNs.

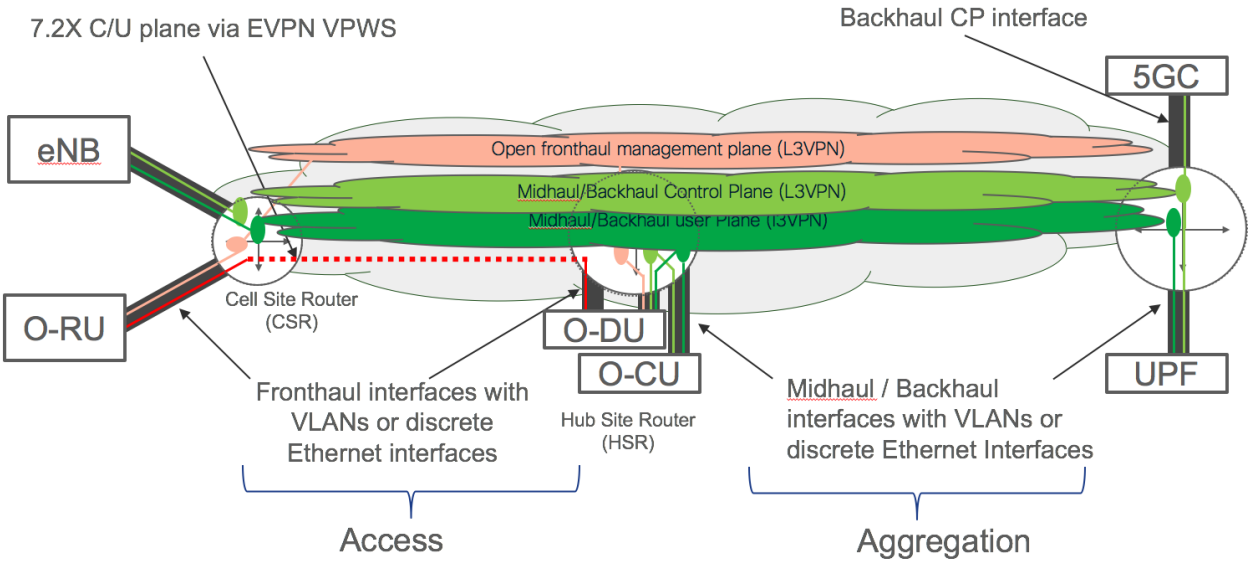


Figure 19-11: L2 / L3 service layout for scenario 3a

Notes on the design:

1. Open Fronthaul traffic runs over the access network to an O-DU located in the Hub site using a VPWS EVPN service.
2. The Midhaul/Backhaul control and user plane VPNs extend to the cell site to support the existing 4G D-RAN solution.
3. The QoS design in the access network needs to be able to cater for the requirements of the Open Fronthaul when D-RAN Backhaul traffic is sharing the same physical link.

19.7.2 Scenario 3b

Scenario 3b is a full two split C-RAN architecture with the O-RU and O-DU at the cell site and the O-CU located at the hub site sitting alongside an existing D-RAN architecture and is shown in Figure 19-12

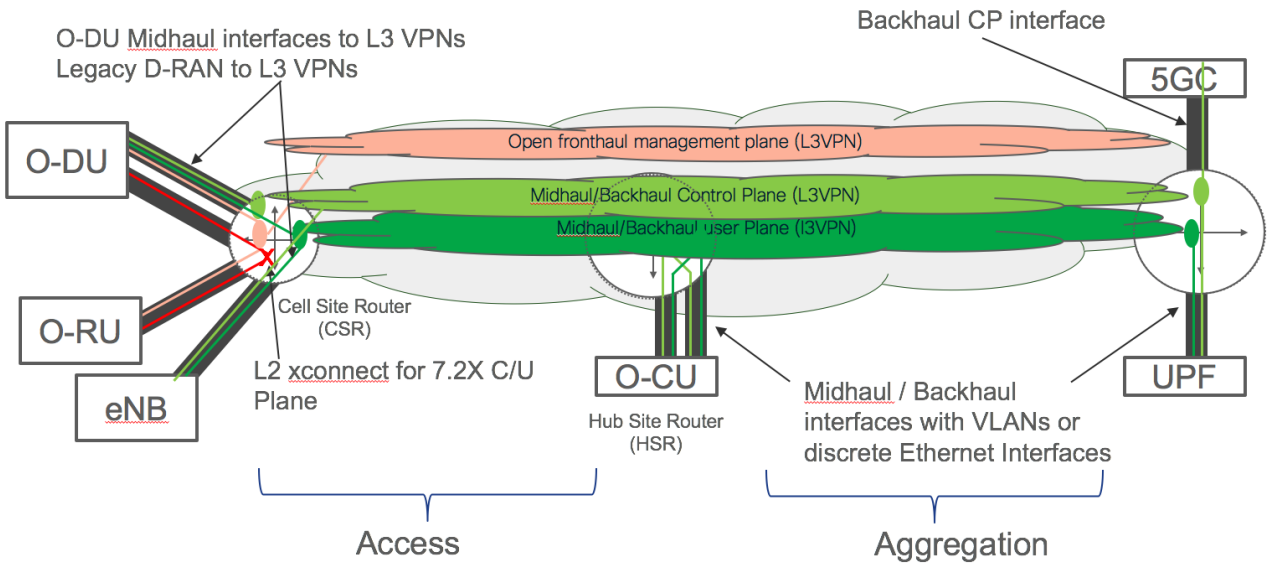


Figure 19-12: L2 / L3 service layout for scenario 3b

Notes on the design:

1. The only difference between scenario 2 is that the legacy D-RAN equipment also connect to the Midhaul/Backhaul control and user plane VPNs at the cell site.

19.8 Scenario 4 5G C-RAN with RoE mappers

Scenario 4 is a full two split C-RAN architecture with O-RAN 7.2x and non O-RAN Fronthaul protocols running across the access network and is shown in Figure 19-13**.**

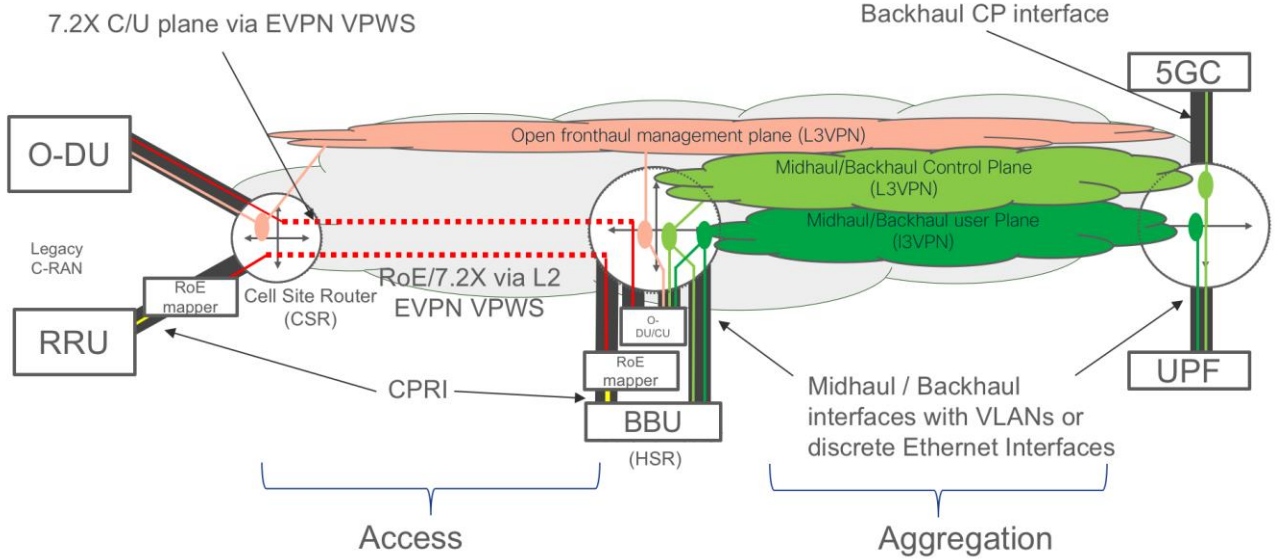


Figure 19-13: L2 / L3 service layout for scenario 4

Notes on the design:

1. O-RAN 7.2x C/U plane and non O-RAN Fronthaul packet traffic (between RoE mappers) use VPWS EVPN services.

2. The expectation is both the O-RAN 7.2x C/U plane traffic and the non O-RAN Fronthaul traffic will have similar delay and jitter characteristics so will run in the same traffic class which is priority queued.
3. The vendor of the RoE mapper function will need to provide details of traffic rates and whether they are fixed or vary depending on user data. The network planner will need to take this into account when capacity planning the access links between the CSR and the HSR.

19.9 Scenario 6 5G C-RAN with distributed UPF

Scenario 6 includes the insertion of distributed UPFs in the architecture. The solution is illustrated in Figure 19-14 and is easily to achieve by simply connecting the distributed UPF to the Midhaul/Backhaul control and user plane L3VPNs. All necessary transport routing will occur automatically through the L3VPN control plane.

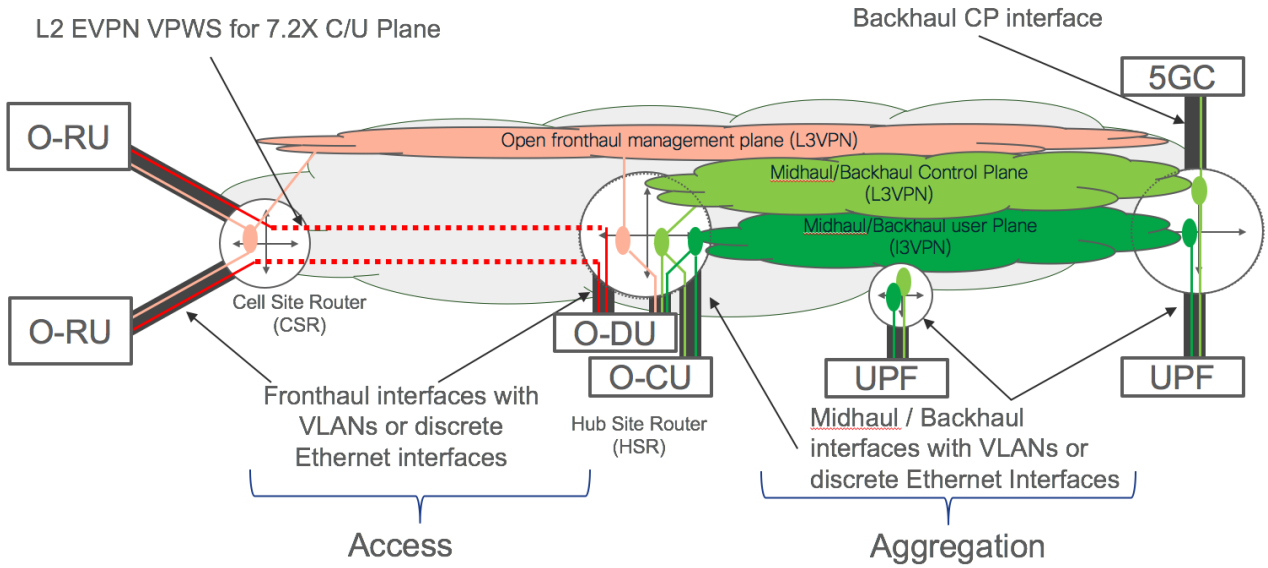


Figure 19-14: L2 / L3 service layout for scenario 6

19.10 Scenario 7 Slicing

This section has been removed because WG-1 is now developing the overall slice architecture and use cases [24] for the whole of the O-RAN organization. The WG-1 slicing task force is developing the slicing solution in phases over multiple O-RAN releases. The packet switched architecture has an extensive set of capabilities to support transport level slicing which are outlined in section 17. As the slice use cases are going to rapidly evolve based on WG-1 slice use cases, slicing scenario have been moved to Annex F where solution examples are presented.

20 Annex A: Overview of “Segment Routing” (SR)

This annex gives an overview of segment routing based on the Segment Routing Architecture defined in RFC8402 [133] and related drafts. It has been included as an annex because the main document describes two architecture/designs that utilise Segment Routing: SR-MPLS and SR for

IPv6 or SRv6. There are differences between the two implementations and capabilities of SR when using an MPLS and IPv6 data plane but there are also many similarities in the basic architectural concepts and the motivation for using the technology. Consequently, this annex describes the basic SR architecture and the "logical architecture sections" in the main document goes into more detail on the technologies and outlines implementation and design considerations when using the respective technologies.

This annex utilises many different information sources and, in some instances, takes content very literally from these sources. Information sources include IETF documents, research and vendor documents.

Background

The Internet and role of packet switching has changed over the last couple of decades and continues to evolve rapidly. This is no better illustrated than by the 5G transport requirements which anticipates billions of endpoints, massive numbers of antenna, huge bandwidths and the ability to concurrently support different 5G services. This evolution is challenging existing Internet protocols, equipment and increasing the operational complexity of the network. Segment Routing and related technology evolutions aims to address these challenges through protocol simplification, removal of state from the network and embracing technology advances such as Software Defined Networking.

Segment Routing

Segment Routing (SR) is based on the loose Source Routing concept. A node can include an ordered list of instructions in the packet headers. These instructions steer the forwarding and the processing of the packet along its path in the network. Single instructions are called segments, a sequence of instructions is called a segment list or an SR Policy. Each segment can enforce a topological requirement (e.g. pass through a node or an interface) or a service requirement (e.g. execute an operation on the packet). The term segment refers to the fact that a network path towards a destination can be split in segments by adding intermediate waypoints. The segment list can be included by the original source of the packet or by an intermediate node. When the segment list is inserted by an intermediate node, it can be removed by another node along the path of the packet, supporting the concept of *tunneling* through an *SR domain* from an *ingress* node to an *egress* node. The implementation of the Segment Routing Architecture requires a data plane which is able to carry the segment lists in the packet headers and to properly process them. Control plane operations complement the data plane functionality, allowing segment allocation (i.e. associate a segment identifier to a specific instruction in a node) and distribution of segment identifiers within an SR domain.

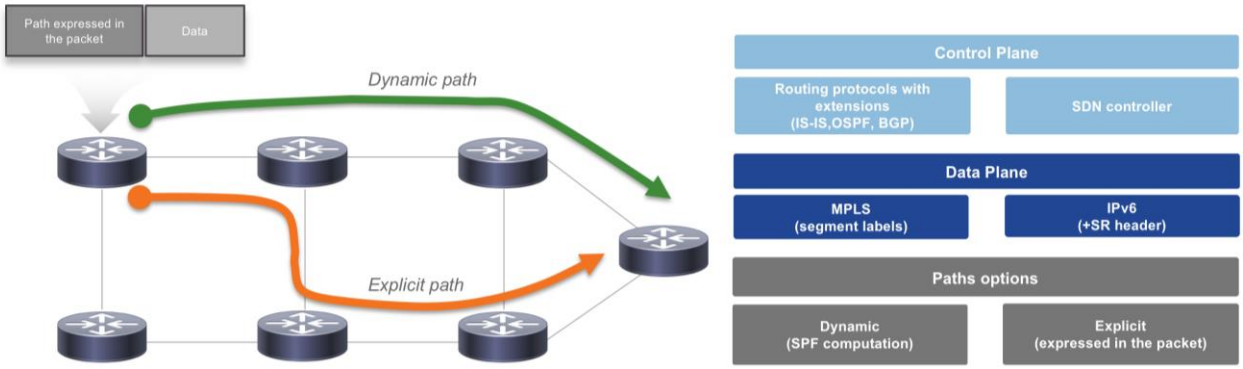


Figure 20-1: Segment Routing protocol stack and path capabilities

Segment Routing architectural principle

There is a single SR architecture which is defined in RFC8402 [133]. This defines the general concepts of SR and is independent from the specific data plane. Currently, there are two instantiations of the SR architecture designed and implemented in production code, SR over MPLS (SR-MPLS) and SR over IPv6 (SRv6).

The architecture relies on segments identifiers (SIDs) and SID lists. The SID represents a single instruction which can be related to how the packet is forwarded or a function that needs to be performed on a packet. A SID lists or the SR policy is a collection of SIDs or instructions to be performed on the packet as it traverses the SR domain. Typically, a SID list is imposed on the packets as it enters the SR domain and is either empty or completed by the time it leaves an SR domain.
There are two basic types of segment, global segments which correspond to instructions that are globally valid in an SR domain and local segments which correspond to instructions that are only valid within a single node. Some of the key segment types are:

1. IGP Prefix and IGP node segments: These are global SIDs and are instructions on how to forward a packet towards a destination IP network or destination IP node. As the name suggests these are conveyed by IGP and can be used by any node in the SR domain to send traffic to a node or IP prefix.
2. Adjacency segments: These are local SIDs which identify the links available on that node.
3. Anycast segments: These are a special IGP-prefix segment that corresponds to an anycast prefix. ie a prefix advertised by more than one router. Anycast and anycast segments are commonly used to provide high availability or load balancing solutions.
4. Binding Segments: These are a single SID that is associated with a segment list or SR policy. This means the node that processes the Binding SID replaces the single segment with a segment list.

Using Figure 20-1 as an example, the basic SR architecture can support:

1. A dynamically locally computed forwarding path between an ingress and egress point within the SR domain by the ingress node either imposing the node SID of the egress node or the prefix SID of the destination prefix. Typically, a link-state IGP or an EGP routing protocol, running locally on the routers calculate these types of paths and they exhibit the behaviour of normal routing protocols such “Shortest Path Forwarding” and “Equal Cost Multi Pathing (ECMP)
2. A fully deterministic path between the ingress and egress points. In this case the SR policy or SID list includes all nodes on the path and if required adjacency SIDs to select links between nodes.
3. A combination of the two. The SR architecture can build a path between an ingress and egress node that combines dynamically computed forwarding and explicitly routed forwarding.

Segment Routing data plane

Two data plane instantiations are designed and implement: SR over MPLS (SR-MPLS) and SR over IPv6 (SRv6).

SR-MPLS

The MPLS data plane (SR-MPLS) is specified in RFC 8660 “Segment Routing with MPLS data plane”[139], In the case of SR-MPLS, Segment Routing does not require any change to the MPLS forwarding plane. An SR Policy is instantiated through the MPLS Label Stack and the Segment IDs (SIDs) of a Segment List are inserted as MPLS Labels. The classical forwarding functions available for MPLS networks allow implementing the SR operations.

IPv6 data plane (SRv6)

For the IPv6 data plane (SRv6), a new type of IPv6 Routing Extension Header, called Segment Routing Header (SRH) has been defined in RFC 8754 IPv6 Segment Routing Header (SRH) [144]. The SRH contains the Segment List (SR Policy) as an ordered list of IPv6 addresses: each address in the list is a SID. A dedicated field, referred to as *Segments Left*, is used to maintain the pointer to the active SID of the Segment List. This SRH format and basic operation of the SRH is illustrated in Figure 20-2.

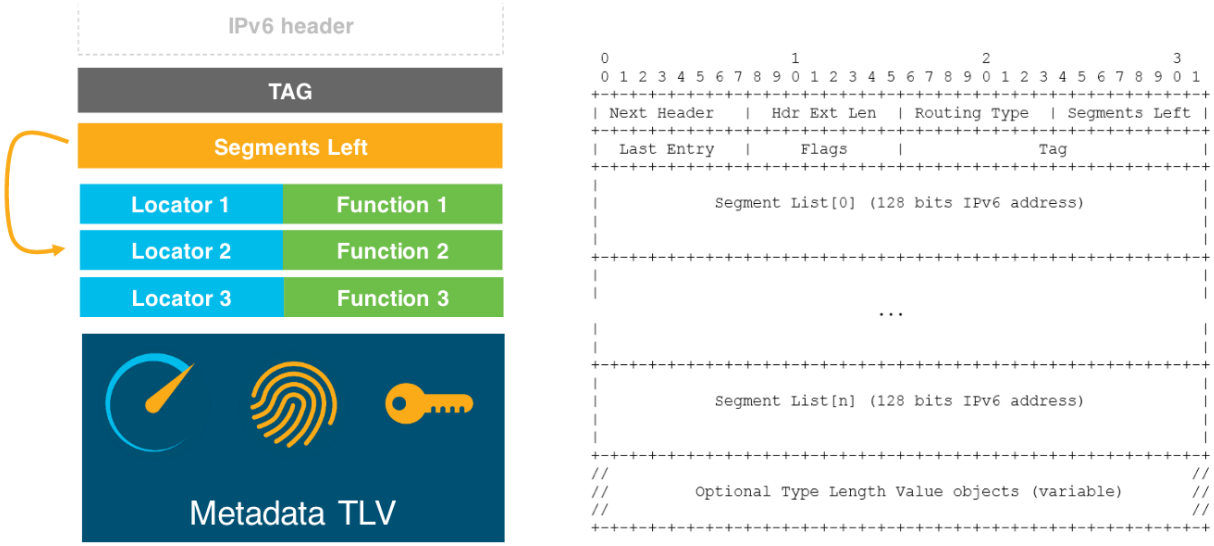


Figure 20-2 SRH header and interaction between segments left and SID list

To explain the SRv6 data plane, there are three categories of nodes: Source SR nodes, Transit nodes and SR Segment Endpoint nodes. A Source SR node corresponds to the *headend* node, it can be a host originating an IPv6 packet, or an SR domain ingress router encapsulating a received packet in an outer IPv6 header.

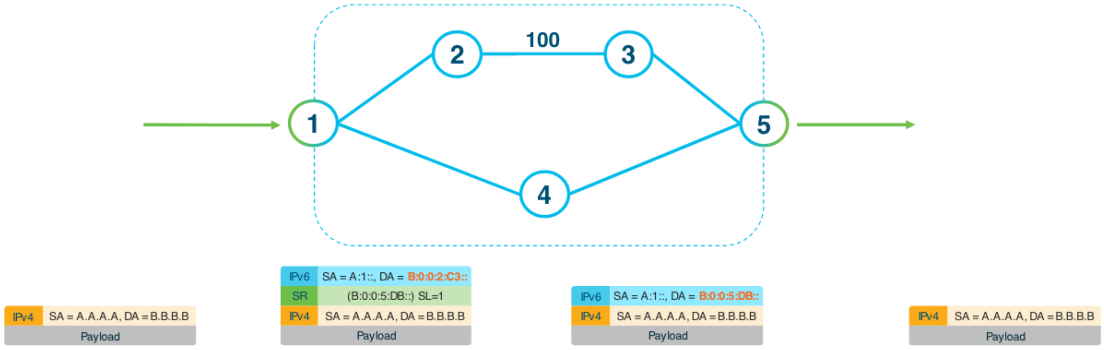


Figure 20-3 Packet and SRH construct for IPv4 payload for traffic following nodes 1,2,5

In Figure 20-3, which considers the latter case the source SR node (S1) is an edge router that encapsulates a packet (which can be IPv6, IPv4 or an Layer 2 frame) into an outer IPv6 packet and inserts the SR Header (SRH) as a Routing Extension Header in the outer IPv6 header. The traffic is destined for S5 and for policy reasons the operator wishes to send the traffic via S2 and S3 which is not the shortest IGP path. To implement this policy, a segment list is composed of S2 and S5. S3 does not need to be specified because the shortest path from S2 to S5 is via S3. To implement this SR policy and assuming the use of a reduced SRH (for details see RFC8754[139]) the packet details are shown in Figure 20-3.

In addition to the basic operations, the *SRv6 Network Programming* model [147] describes a set of functions that can be associated to segments and executed in a given SRv6 node. Examples of such functions are: different types of packet encapsulation (e.g. IPv6 in IPv6, IPv4 in IPv6, Ethernet in IPv6), corresponding decapsulation, lookup operation on a specific routing table (e.g. to support VPNs). The list of functions described in [147] is not meant to be exhaustive, as any function can be associated to a segment identifier in a node. Obviously, the definition of a standardized set of segment routing functions facilitates the deployment of SR domains with interoperable equipment from multiple vendors.

Segment Routing control plane

The control plane is responsible for calculating, building the connectivity graphs and implementing the SR policies within the network. Segment Routing can support distributed, centralised or hybrid control plane architectures. See Figure 20-4 for more details.

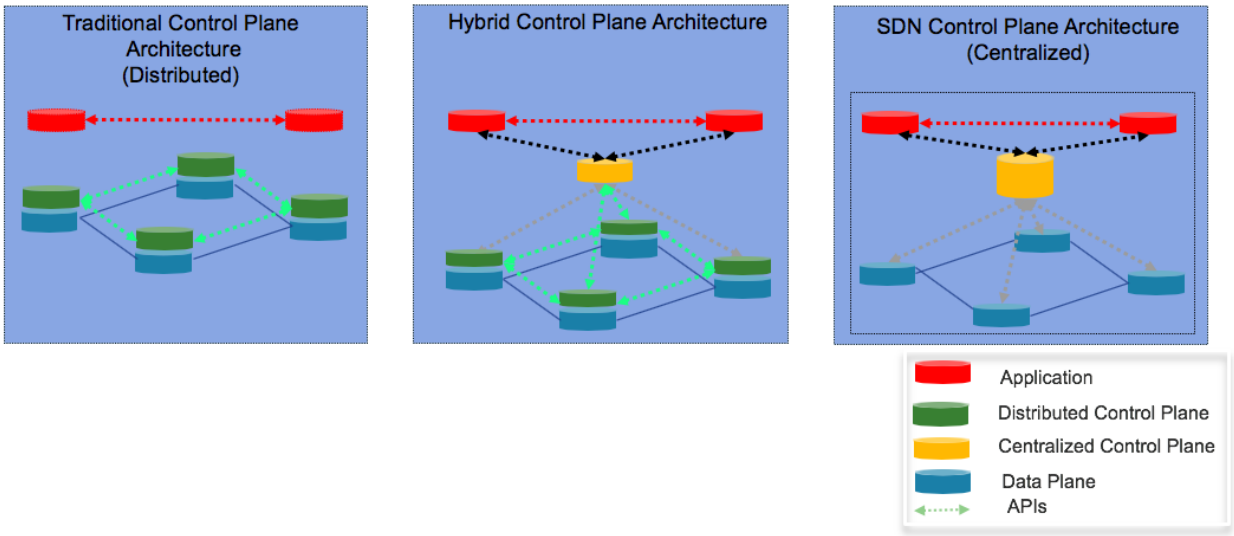


Figure 20-4 Distributed, centralised and hybrid control planes

In a distributed control plane environment, the transport nodes independently interact with each other to convey routing information and independently make decisions to set-up and enforce SR policies. This is achieved through a combination of IGP and EGP protocols running between the transport nodes.

In a centralised control plane environment, central SDN controllers have full knowledge of the network topology and calculates SR policies centrally. These are then programmed onto the TNEs in the network.

The hybrid approach splits the responsibilities, the TNEs in the network support some functions, such as building the link state database, calculating the intra-domain routing tables, monitoring the links to attached nodes, and rapidly recovering in the event of failure. At the same time the central SDN controllers has a feed from the network and is used for functions that require a network wide view such as transport optimization and inter-domain routing.

MPLS-SR distributed control plane

The MPLS-SR distributed control plane utilises ISIS or OSPF routing protocols with extensions to advertise the different types of IGP segments (prefix, node, adjacency, any cast) between TNEs.

IPv6 distributed control plane

For the IPv6 data plane, the process of advertising the IGP-prefix, IGP-node and IGP-anycast segments is simplified due to the use of IPv6 addresses as SIDs. In particular, there is no need to extend the IGP routing protocols to distribute these segment types, because they are IPv6 prefixes natively distributed by the routing protocols. This means that the Control Plane for SRv6 can use the regular IPv6 link state IGP routing protocols (OSPFv3, ISIS) to support the basic operations, while extensions are still needed ([142][143]) to distribute IGP-Adjacency segments and other SR configuration information.

Centralised control plane

Although much emphasis has been made of fully centralised control plane models in academia for path computation, they have not made their ways into production large scale WAN designs due to the distributed nature of the transport nodes and the need for the transport nodes to communicate at

all times with the central controller. As a consequence, most WAN SDN designs today are hybrid in nature, where intelligence is split between the TNEs and central SDN controllers.

Note: This discussion above only considers path computation element of a central controller. There are many aspects of a central SDN controller including orchestration, data analytics etc.

Hybrid control plane

The hybrid control plane is combination of distributed control plane and centralised SDN control plane functionality. Like the central control plane model there is no universal architecture for a hybrid control plane design, other than some functions are provided centrally and others occur locally on the TNEs. One of the most common hybrid control plane designs is where the central SDN controllers provides a “Path Computation Element” capability. In this model the PCE is responsible for path computation and inter-domain routing with the distributed control planes running on the TNEs, responsible for building the link state database, calculating the network graph and building the shortest path forwarding tables. In this scenario there is generally a tight relationship between the two components.

A. The TNEs exchange topology information, creates the link-state database and computes the shortest path forwarding tables.
B. The centralised control plane component needs to get visibility of the distributed protocol’s link-state database to understand the current topology. Some of the common approaches are:
    a. The SDN controller participates in the IGP and as a consequence builds its own link-state database and network graph.
    b. The SDN controller extracts IGP databases from the network using proprietary mechanisms.
    c. BGP-LS (BGP link state) runs between one or more TNEs and centralised controllers. In this way the centralised controller gets visibility of the network through BGP-LS.
C. In many instances the centralised control plane needs to gather other information about the network, for example link loading. Currently, there is no single standard for collecting this information. Common mechanisms include SNMP, telemetry feeds and operational data from Yang models residing on the devices.

D. The central controller and the TNEs also need to communicate when, either a TNE requests a service from the central SDN controller, or the central SDN controller wants to push an explicit policy to a TNE. An example would be if a TNE needs build an SR policy and wants to use the services of a PCE. The transport node and the PCE need to communicate, some of the common protocols used for this type of communications include:
    a. Network Configuration Protocol
    b. PCE (PCEP)
    c. BGP
    d. Openflow (SR-MPLS)

21 Annex B: IETF Ethernet Virtual Private Networks

IETF Ethernet Virtual Private Network (EVPN) is a service supporting transportation of Ethernet frames across a transport network encapsulated in MPLS or IP. There are multiple flavors of IETF EVPN service as outlined Table 6.

Service Type			Standardization status
Layer 2 VPN	E-LAN	VLAN-based	**RFC 7432 [116]**: *BGP MPLS-Based Ethernet VPN*
		VLAN-bundle	
		VLAN-aware bundle	
	E-Tree	VLAN-based	**RFC 8317[129]** : *Ethernet-Tree (E-Tree) Support in Ethernet VPN (EVPN) and Provider Backbone Bridging EVPN (PBB-EVPN)*
		VLAN-bundle	
		VLAN-aware bundle	
	E-Line (VPWS)	VLAN-based	**RFC 8214 [125]**: *Virtual Private Wire Service Support in Ethernet VPN* **draft-ietf-bess-evpn-vpws-fxc [158]**: *EVPN VPWS Flexible Cross-Connect Service*
		VLAN-bundle	
Layer 3 VPN			**RFC 9136 [149]**: *IP Prefix Advertisement in Ethernet VPN (EVPN)*

Table 6: EVPN service classification

The two major building blocks of EVPN are:

- Control plane, based on Border Gateway Protocol (BGP), to distribute all necessary information required for proper EVPN operation
- Data plane, with different underlay transport options, to suit various requirements:
  - MPLS (LDP, RSVP, BGP-LU, MPLSoUDP, SR, SR-TE)
  - IP (VxLAN, SRv6)

EVPN E-LAN service

EVPN E-LAN service is a Layer 2 multipoint-to-multipoint service. In essence, EVPN E-LAN emulates Layer 2 switch behavior, with attachment circuits (ACs) of this single emulated Layer 2 switch present on multiple Provider Edge (PE) routers, which are placed in different network locations, as outlined in Figure 21-1

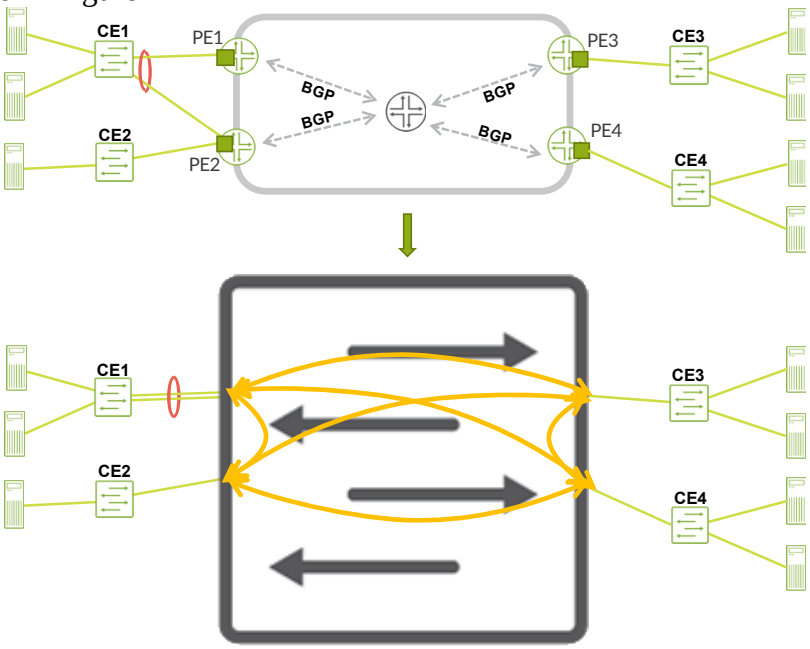


Figure 21-1: EVPN E-LAN service

Similar to ordinary Layer-2/Layer-3 switch, EVPN E-LAN service provides

- MAC learning capability (globally or per VLAN)
- capability to carry Ethernet frames without VLAN tag, as well as with single or double VLAN tags, including VLAN translation
- single-homed, as well as multi-homed (both single-active and all-active) attachment circuits capability
- BUM optimization mechanisms (like for example APR proxy/suppression, IGMP proxy, etc.)
- handover to Layer-3 domain via integrated routing and bridging (IRB) operation and more.

EVPN E-Tree service

EVPN E-Tree service is an enhancement of EVPN E-LAN service, which allows to restrict certain communication patterns, creating in essence a point-to-multipoint service model. As outlined in Figure 21-2, each attachment circuit is designated as *root* (R) or *leaf* (L). As per EVPN E-Tree definition, only *root*⇔*root* and *root*⇔*leaf* Layer 2 communication is permitted, while *leaf*⇔*leaf* Layer 2 communication is blocked.

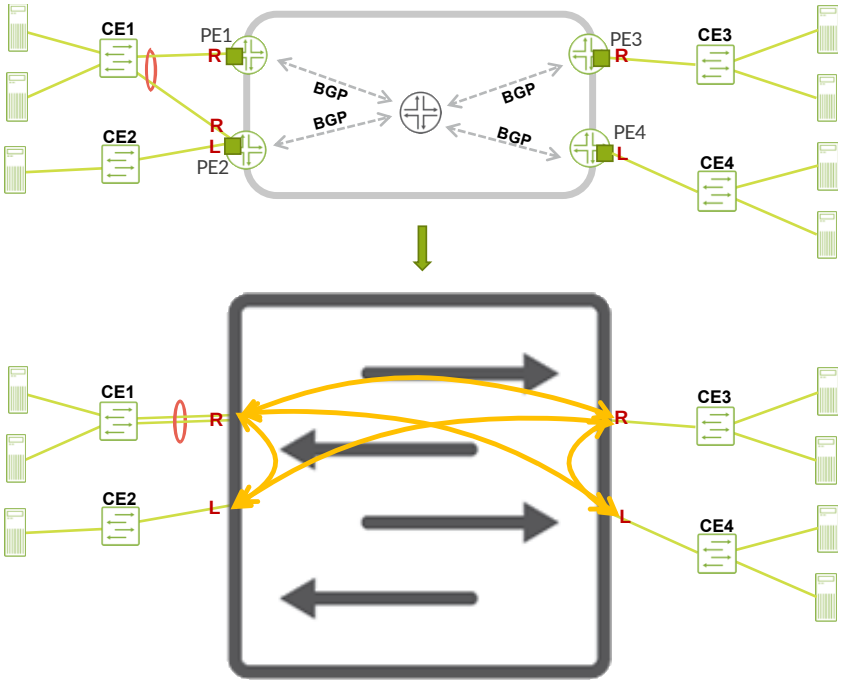


Figure 21-2 EVPN E-Tree services

At a high level, EVPN E-LAN service can be considered as a special case of EVPN E-Tree service, with all attachment circuits designated as *root*. In fact, many vendors use the same configuration constructs for EVPN E-LAN and EVPN E-Tree, defaulting the attachment circuit to a *root*, with some configuration knobs turning it to a *leaf*.

EVPN E-Line service

EVN E-Line, called as well EVPN Virtual Private Wire Service (VPWS), is point-to-point Layer 2 service, without MAC learning. Ethernet frame received from an end device connected via single-homed or multi-homed attachment circuit, is transported over this point-to-point service to the remote end device, connected again via single-homed or multi-homed attachment circuit. Since it is pure point-to-point communication, no MAC learning is required, and thus MAC learning is not part of EVPN E-Line (VPWS) architecture.

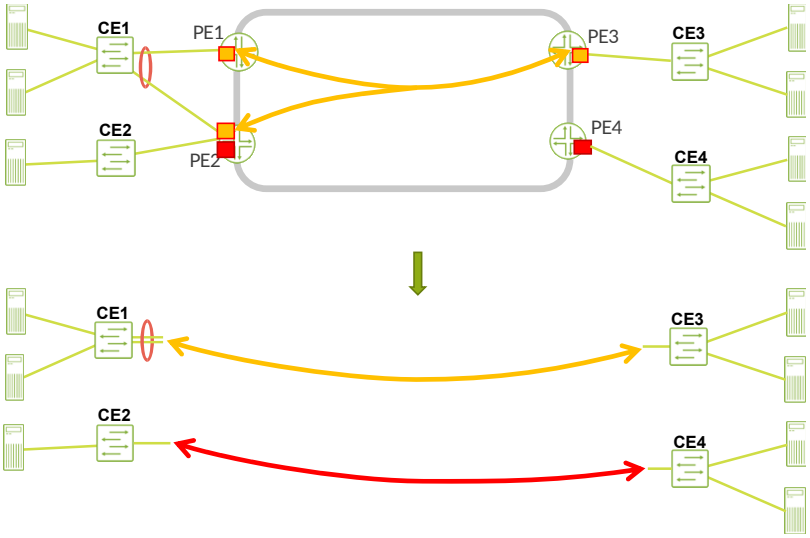


Figure 21-3 EVPN E-Line (VPWS) services

EVPN VLAN-based service

EVPN architecture uses three implementation constructs defined on the Provider Edge (PE) devices:

- EVPN Instance (EVI), called as well MAC-VRF. Both EVI and MAC-VRF will be used in this document interchangeably.
- Bridge (MAC learning) domain/table
- Broadcast (flooding) domain (i.e. VLAN)

Depending how these building blocks are used, EVPN architecture defines 3 options for constructing EVPN services:

- EVPN VLAN-based service (applicable to EVPN E-LAN, EVPN E-Tree and EVPN E-Line services)
- EVPN VLAN-bundle service (applicable to EVPN E-LAN, EVPN E-Tree and EVPN E-Line services)
- EVPN VLAN-aware bundle (applicable to EVPN E-LAN and EVPN E-Tree services only)

In EVPN VLAN-based service, each EVI contains single bridge (MAC learning) domain/table, which in turn contains only single VLAN (single broadcast/flooding domain). This is the simplest EVPN service with 1:1 mapping between VLAN and EVI as outlined in 0

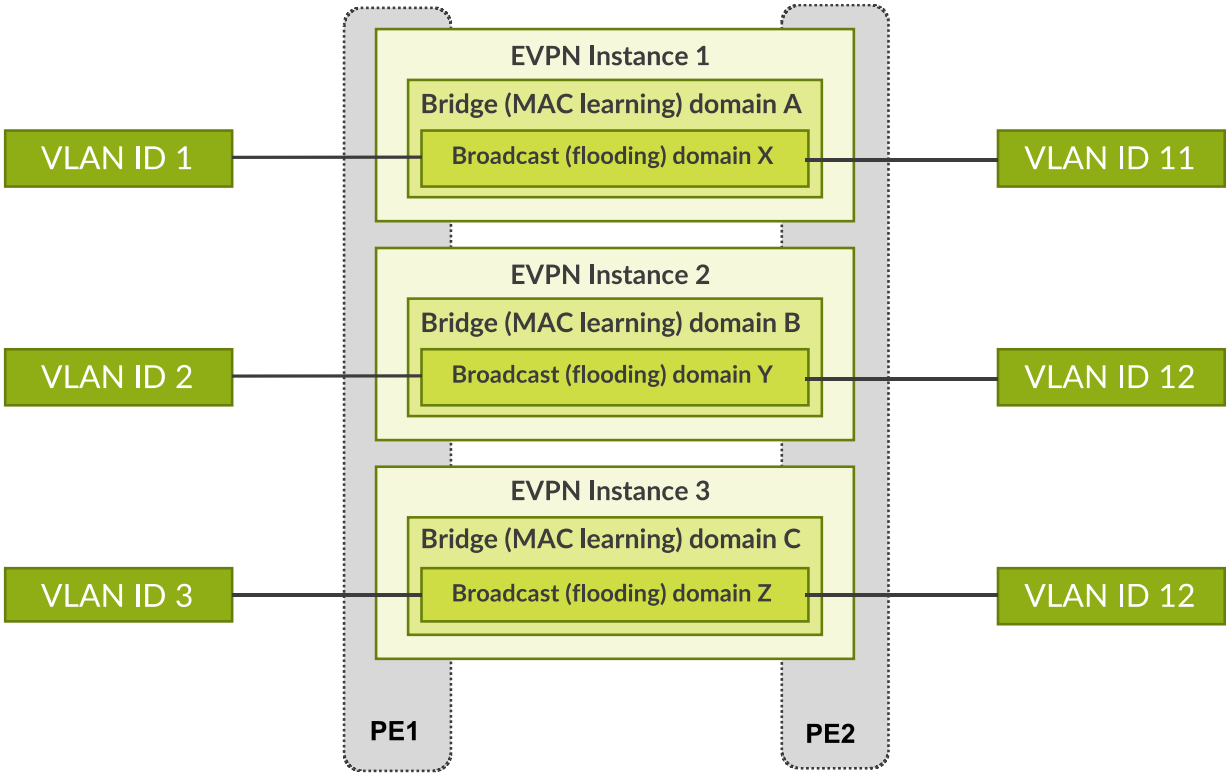


Figure 21-4 EVPN VLAN-based services (E-LAN, E-Tree, E-Line)

EVPN VLAN-based service allows for VLAN tag translation, where incoming VLAN tag (for example VLAN 2) is translated to VLAN 12. Additionally, EVPN VLAN-based service provides the option to carry the Ethernet frame with the VLAN tag across EVPN network, or to strip the VLAN tag before carrying the Ethernet frame across EVPN network.

EVPN VLAN-bundle service

The biggest challenge with basic EVPN VLAN-based service is scaling. In EVPN VLAN-based service each VLAN consumes dedicated EVI, thus the maximum numbers of VLANs that can be served is limited by the supported EVIs on given hardware platform. EVPN VLAN-bundle service addresses this scaling limitation by bundling multiple VLANs into single bridge (MAC learning) domain/table and single EVI (MAC-VRF), as outlined in Figure 21-5

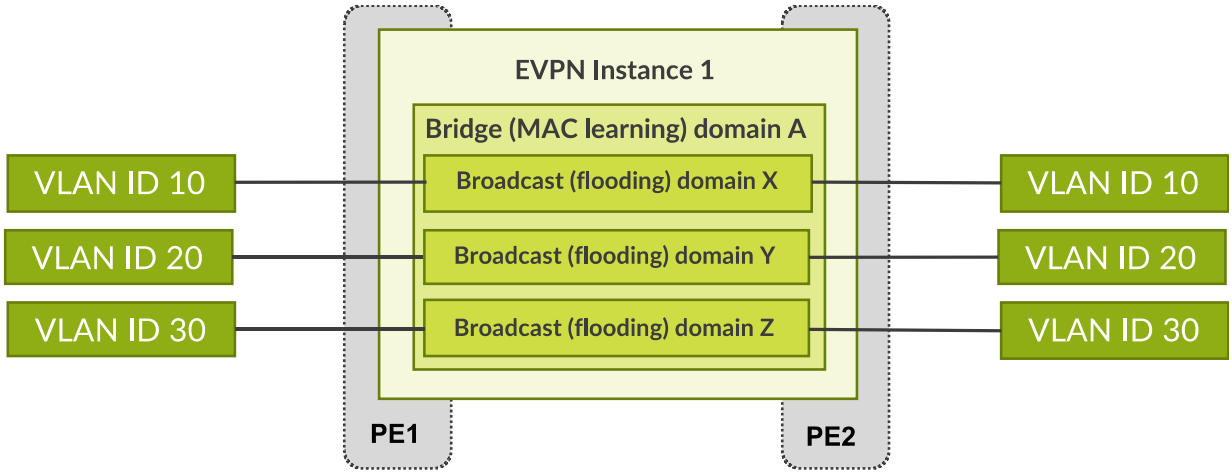


Figure 21-5 EVPN VLAN-bundle service (E-LAN, E-Tree, E-Line)

VLAN bundling allows dramatic reduction of MAC-VRFs that otherwise would be needed on the router. Therefore, EVPN VLAN-bundle service is typically used in high-scale environment, with large number of VLANs or VLAN combination (QinQ) that must be transported with EVPN service.

This scaling improvements brings improvements, as well some restrictions for service deployment, namely:

- the original VLAN tag must be carried with frame across EVPN network and VLAN translation is not supported
- MACs cannot be re-used across different VLANs

The first restriction is related to the fact, that underlying transport identification (i.e. MPLS label with MPLS underlay transport, Function:Argument with SRv6 underlay transport) identifies the bridge domain. Therefore, the underlying transport identification cannot be used to distinguish Ethernet frames - received over EVPN network - as belonging to different VLANs, if VLAN tag is stripped before sending Ethernet frames across EVPN. And this differentiation is required to maintain per-VLAN broadcast/flooding restrictions.

Second restriction is the straight result of bundling multiple VLANs in single bridge (MAC learning) domain/table. All these bundled VLANs use the same MAC table, therefore all MACs must be unique across all bundle VLANs.

EVPN VLAN-aware bundle Service

Restrictions of EVPN VLAN-bundle services are removed by EVPN VLAN-aware bundle service. EVPN VLAN-aware bundle service is a trade-off between EVPN VLAN-based service (poor scaling, but simple service without restrictions related to VLAN translation or MAC learning) and EVPN VLAN-bundle service (good scaling, but restricting VLAN translation support and requiring MAC uniqueness).
EVPN VLAN-aware bundle brings back separate bridge (MAC learning) domain/table per VLAN, still keeping the option to bundle multiple VLANs with single EVI (MAC-VRF), as outlined in Figure 21-6**.**

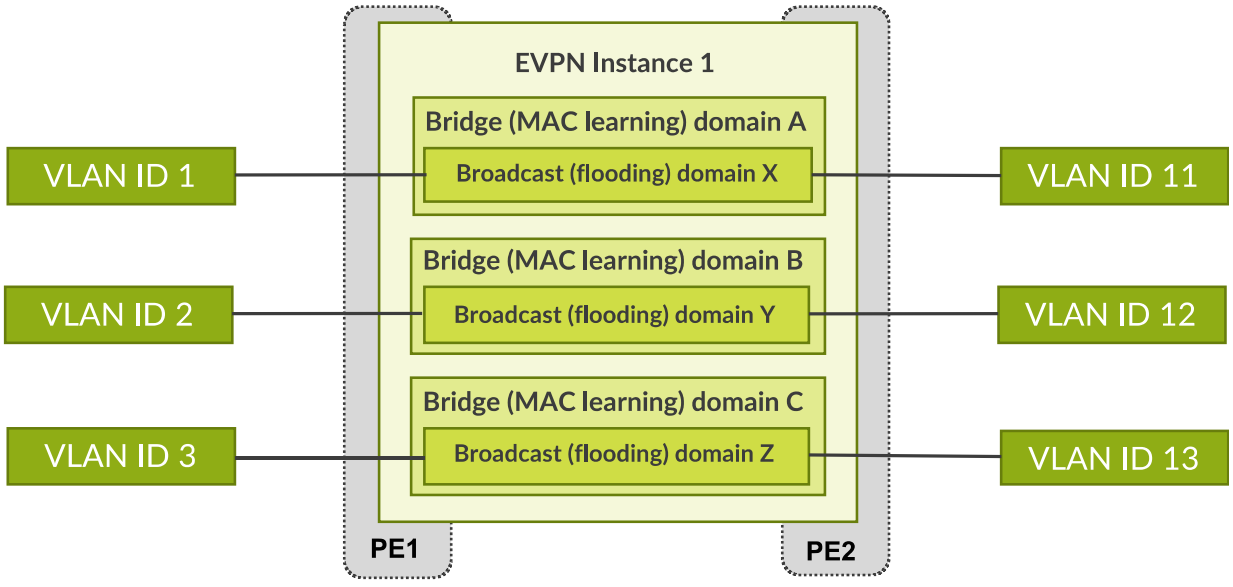


Figure 21-6 EVPN VLAN-aware bundle service (E-LAN, E-Tree)

Scaling wise EVPN VLAN-aware bundle service sits between EVPN VLAN-based service and EVPN VLAN-bundle service, optimizing number of required EVIs (MAC-VRFs), while keeping the same number of bridge (MAC learning) domain as EVPN VLAN-based service.

Table 7 summarizes different EVPN service models.

Characteristic	VLAN-based	VLAN-bundle	VLAN-aware bundle
Broadcast domains (VLANs) per EVI	=1	≥1	≥1
Bridge (MAC learning) domains per EVI	=1	=1	≥1
MACs must be unique across VLANs	no	yes	no
VLAN translation allowed	yes	no	yes
VLAN tag carried over	yes/no	yes	yes
Ethernet Tag (BD) ID on BD scoped routes	0	0	≠0
Port based service support	no	yes	yes

Table 7: EVPN VLAN service types comparison

MEF/EVPN service mapping

MEF 6.3 (Subscriber Ethernet Service Definitions) defines Ethernet services commonly used in Ethernet metro aggregation architectures. Table 8 provides the mapping between MEF service definitions and IETF EVPN service implementations.

MEF 6.3 Ethernet Service Definitions	**IETF EVPN Service Definition**
EPL: Ethernet Private Line	EVPN E-Line VLAN-based
EVPL: Virtual Ethernet Private Line	EVPN E-Line VLAN-bundle (port-based)
EP-LAN: Ethernet Private LAN	EVPN E-LAN VLAN-based
EVP-LAN: Ethernet Virtual Private LAN	EVPN E-LAN VLAN-bundle (port-based) EVPN E-LAN VLAN-aware bundle (port based)
EP-Tree: Ethernet Private Tree	EVPN E-Tree VLAN-based
EVP-Tree: Ethernet Virtual Private Tree	EVPN E-Tree VLAN-bundle (port-based) EVPN E-Tree VLAN-aware bundle (port based)

Table 8: MEF Ethernet Services to IETF EVPN Services Mapping

22 Annex C: MP-BGP based L3VPNs

BGP based layer 3 VPNs are a widely deployed connectivity application in Service Provider networks and work over an MPLS and SRv6 underlay infrastructure. BGP based layer 3 VPNs are based on RFC4364 [84] and use a peer-to-peer model that uses Border Gateway Protocol (BGP) to distribute VPN-related information. It is a highly scalable, peer-to-peer model that through route filtering allows flexible connectivity models within a VPN. VPN topologies include any-to-any L3 multi-point solutions as well as L3 tree solutions where connectivity constraints are built into the VPN infrastructure through controlled exportations and importation of routing information. This makes it an effective and efficient technology for L3 services associated with Xhaul transport.

Building blocks of a L3 VPN service

A typical L3 VPN services consists of the following:

- A Provider Edge (PE) Routers
- Virtual Routing Forwarding
- Route Distinguisher (RD)
- Route Target (RT)
- Multi-Protocol Border Gateway Protocol (MP-BGP)

Provider Edge (PE) Router

A Provider Edge router is a device which originate or terminate a L3VPN service. In other words, a PE router is a VPN Endpoint, providing connectivity to the device(s) connected to it.

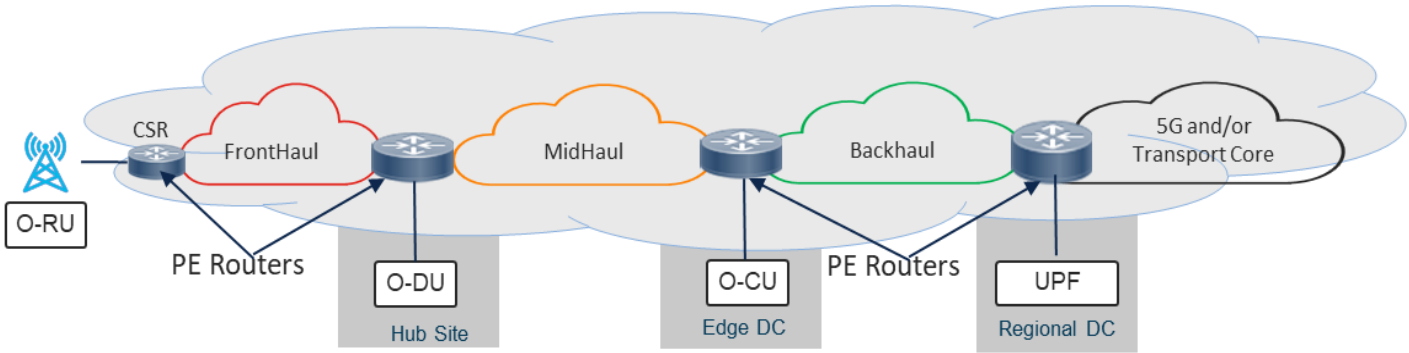


Figure 22-1 PE routers in an Xhaul infrastructure

In the context of an Xhaul packet switched transport network, the first L3 router enabling an L3VPN services to provide connectivity between the O-RAN/3GPP components (O-RU, O-DU, O-CU, UPF, 5GC) is considered a Provider Edge (PE) router. This could be a CSR or HSR in Fronthaul, Midhaul or Backhaul network.

VPN Routing and Forwarding Tables (VRF) and Route Distinguishers (RD)

The VPN routing and forwarding table (VRF) is a key element in the BGP based L3 VPN technology. A VRF essentially define a VPN instance on PE and associates a routing table instance with the VRF. A VRF exist on PEs and more than one VRF can exist on a single PE. The VRF contains routes that should be available to a particular set of sites participating in the VPN.

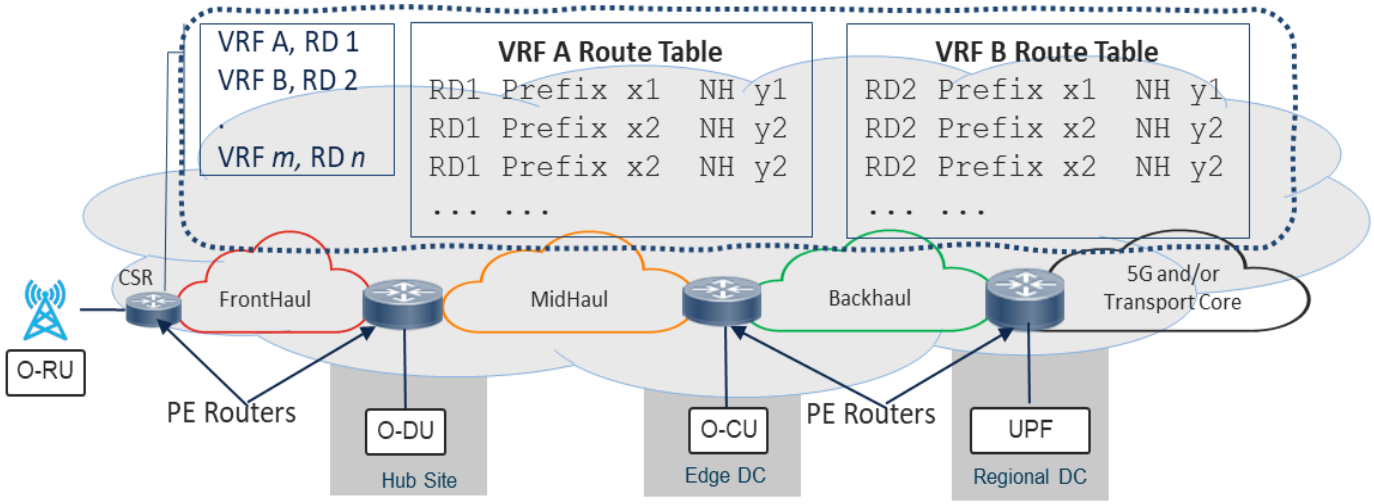


Figure 22-2: VRF routing tables

A route distinguisher (RD) is a unique value assigned to a VRF on the PE Router. A RD allows to associate and identify routes in the routing table with a particular VRF and allows for overlapping IP address space across VRFs.
An RD is a local value on the PE in the Packet Switched RAN network.

Route Targets (RT)

Route Targets (RT) are the mechanism by routing information distribution can be controlled within an BGP based VPN throughout the packet switched network. Route Target are propagated through the BGP based VPN network using route-target extended MP-BGP communities. Every PE within the MPLS VPN network define a set of RT values for import or export of VPN route information. BGP based VPN uses route-target communities as follows:

- When a VPN route is injected into MP-BGP, the route is associated with a list of VPN route-target communities.
- An import list of route-target communities is associated with each VRF. This list defines the values that should be matched against to decide whether a route is eligible to be imported into this VRF.

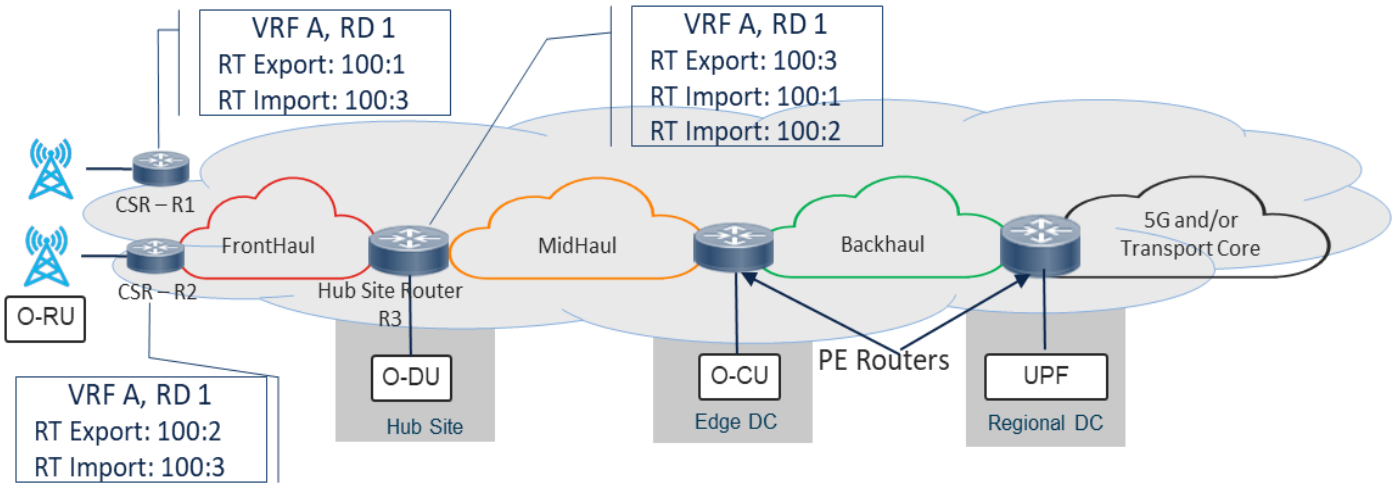


Figure 22-3: Use of route targets to control IP connectivity within a VRF

Figure 22-3 above shows a simple mechanism of import and export Route Communities within an BGP based VPN network. Where required CSR route-targets may be configured to be different than the Hub Site Router. For scalability and efficient resource usage on the CSR, the operator may choose to not import all the routes, rather importing the routes from Hub Site Route and only certain other CSR's.

When used properly, Route Targets can be used to limit the routing table information within a BGP based VPN Network and create different IP topologies at the transport layer. These include:

- Any to any topologies.
- Hub and spoke topologies.

Multi-Protocol Border Gateway Protocol (MP-BGP)

MPLS VPN information is propagated through the network using Multi-Protocol BGP (MP-BGP) [86]. All PE routers within the VPN must configure MP-BGP peering and enable VPN address-family and extended communities to exchange VPN related information including routing table information and Route Targets.

Traffic Steering into an BGP VPN

SR Policy operations is defined in [151]. It is a mechanism that can be to create an end-to-end Label Switch Path (LSP) using required metrics and constraints. While using Segment Routing Traffic Engineering (SRTE) for traffic path programmability, it is possible to automatically steer BGP based VPN traffic into an SRTE policy.
Typically, an SRTE Policy is uniquely identified using the following 3-tuple consisting of:

- **Head End:** Starting point of the LSP created by SRTE Policy
- **End Point:** Destination for the traffic using the SRTE LSP
- **Color:** A numeric value associated with the SRTE policy.

The SRTE policy color is a numeric value that is configurable in the SRTE policy. Once a VPN endpoint route is "tagged" with this color, traffic can be automatically steered into the SRTE policy.

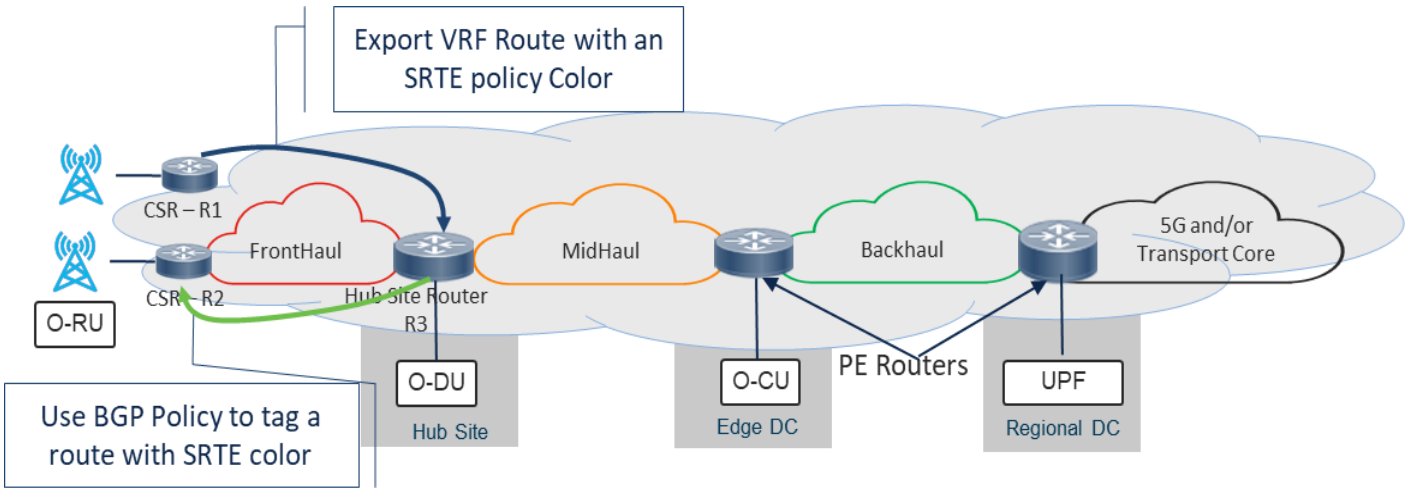


Figure 22-4: Traffic Steering with BGP based VPNs

As shown above, the VPN routes can be colored in 2 possible ways:

1. While defining a VRF, define a color for all Prefixes associated with that VRF, or
2. When sending or receiving BGP routes, using a policy to “color” the route.

In either case, the VPN route will now be tagged with a color and traffic for the destination or a flow will automatically be steered into the SRTE policy that may provide traffic path programmability.

23 Annex D: Quality of Service

This annex gives an overview of packet base Quality of Service (QoS).

What is Quality of Service?

Quality of Service (QoS) is an umbrella term that broadly covers the concept of the management of traffic flowing in an infrastructure. In general, QoS can provide two capabilities.

1. A mechanism to actively manage the transmission of data on a network link out of a node when the link is approaching or at saturation. This active management consists of determining which data to forward next, which data to store for forwarding later, and which data to drop.
2. The enforcement of service level agreements through the selective acceptance of traffic into a node (or network) based on pre-determined criteria. Again, this acceptance is based on determining which data to accept now, which data to store for acceptance later and which data to deny (drop).

Fundamentally, QoS only operates when traffic is flowing at the limits of capacity (either physical or logical). When flowing traffic rates are well below the limits, QoS is essentially benign allowing all traffic to be forwarded as it is received.

Why do we need QoS?

Applications using the network transport will often require certain capabilities from the network in order for the application to function correctly.

For example, real-time voice services (telephone calls) can only sustain a maximum latency between the end points before interactive speech becomes difficult. This maximum latency between endpoints translates to two demands from the network;

1. Minimising hop by hop latency as a voice packet flows across the network infrastructure.
2. Minimising jitter in the packets arriving in order to minimise the de-jitter buffer needed at the receiver (which adds latency).

It is useful then to be able to identify voice traffic in the network and prioritise if for forwarding such that it is not held up behind other less time sensitive traffic.

A second example might be an Electronic Point of Sale (EPoS) application. This type of application does not need to be particularly latency bound (a few 10s or 100s of milliseconds in packet arrival does not impact the service), but it does need to have some guaranteed capacity in order to function. I.e. we do not want the EPoS application struggling for bandwidth because a large data transfer is taking place. This guaranteed bandwidth requirement also translates into network demands.

1. Minimising packet loss for the traffic, if necessary, storing it locally until the local egress interface has capacity to send it on.
2. Reserving a minimum amount of transmission time (or bandwidth) on the interface for packets of the EPoS application.

QoS then is the mechanism by which operators can identify applications in an infrastructure and apply policies to the infrastructure to maintain the applications transport requirements, even when capacity is limited.

QoS functional elements

As a suite of capabilities QoS can be broken out into three distinct sets of functions;

1. Traffic Classification and marking
2. Congestion avoidance
3. Congestion management

These three function sets work together to identify (classify) different traffic types, determine how much of each traffic type should be allowed into and through the network (congestion avoidance) and how traffic should be scheduled when interfaces become congested (congestion management).

The specific implementation of these capabilities is vendor specific, but in general, the behaviours required by the different functions are defined as policies, either individually or in groups. These policies are applied to interfaces on the switching elements in a network defining the local behaviours within the device.

Network level behaviour

QoS is implicitly a mechanism that is node based; the management of traffic in one node has no direct impact on the management of traffic in surrounding nodes. However, it is important to design the QoS behaviour of a network as a whole. Just as the operator has the ability to define the routing behaviour of a network as a system of single hop by hop elements, so she has the ability to define the QoS behaviour of the network as a system.

The interaction between elements is defined by the policies applied to each device and the use of "class of service" markings applied to the packets or frames flowing between the devices. [1]
The use of markings allows an individual network element to pass some information about a given packet to a downstream element. How this downstream element interprets the information and then

[1] The term "class of service" is used here to generically to describe the specific markings in a packet or frame. The author recognises Class of Service (COS) identifies such markings in an ethernet header as defined by IEEE 802.1p. Here it is generically used to refer to Ethernet COS, IP TOS, IP PREC, IP DSCP. MPLS EXP etc.

acts on it, is out of the upstream elements control, but a network of devices under single administrative control may choose to use such information in network wide policy.

Node level behaviour

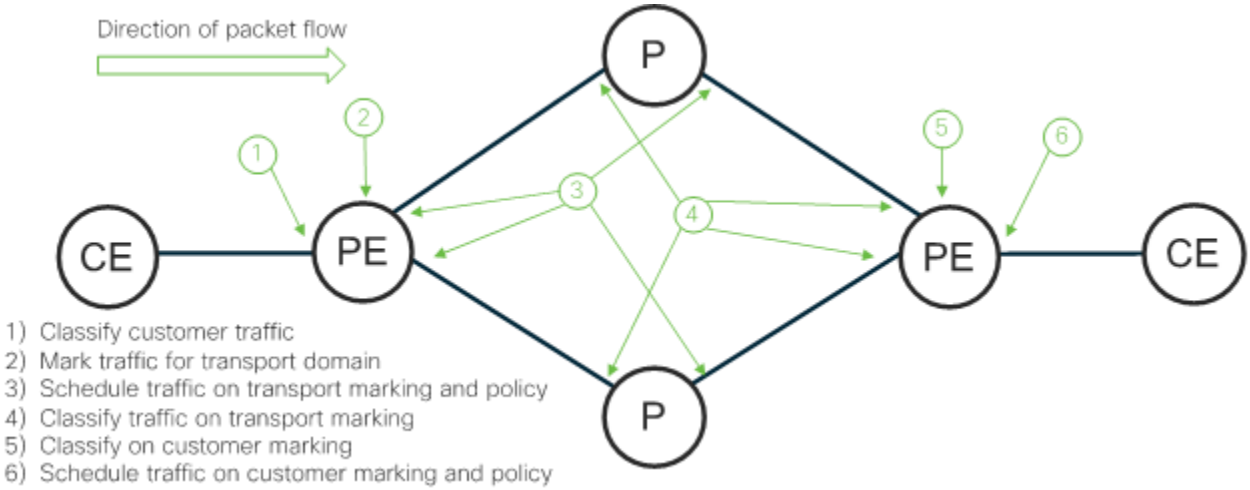


Figure 23-1: Example network QoS structure

The per-hop-behaviours (PHBs) needed in the network can generally be broken down into two sets; Those needed at the edge of the network (Edge QoS), and those needed in the core of the network (Core QoS). We can see these behaviours in Figure 23-1**.**

1. Edge QoS. Here we are classifying traffic entering into the network on ingress (1) and scheduling traffic leaving the network on egress (6). Here the ingress classification (from the client) and egress classification (5) and scheduling (6) (towards the client) will (often) be determined by the markings of the client traffic being transported across the network. In addition, ingress traffic (CE → PE), once classified, will be marked (2) to conform to the transport marking scheme.

2. Core QoS. Here traffic will be classified by the transport marking scheme (4) (mentioned above), scheduled based on those markings (3). This PHB is designed to manage traffic at an aggregate level, treating services that have common requirements (defined by marking) with a common behaviour.

These two sets of PHBs present different requirements to the underlying hardware and care should be taken when selecting a platform to ensure the operators desired PHBs can be met.

Traffic classification and marking

As previously mentioned, traffic classification is a term used to describe the identification and categorisation of traffic arriving at or inside network element. Classification broadly falls into two distinct categories.

Context based classification

In this model, traffic arriving at a node is classified, (allocated to some internal category) based on the context associated with its arrival. For example, it might be assumed that all traffic arriving from a specific logical or physical interface will be categorized as “best effort”. This specific example actually represents the default behaviour for almost all network elements, and is useful to recognise, as it forms the basis for all other classifications. I.e. that without further modification, traffic arriving at a node will be treated as best effort (without any specific explicit treatment).

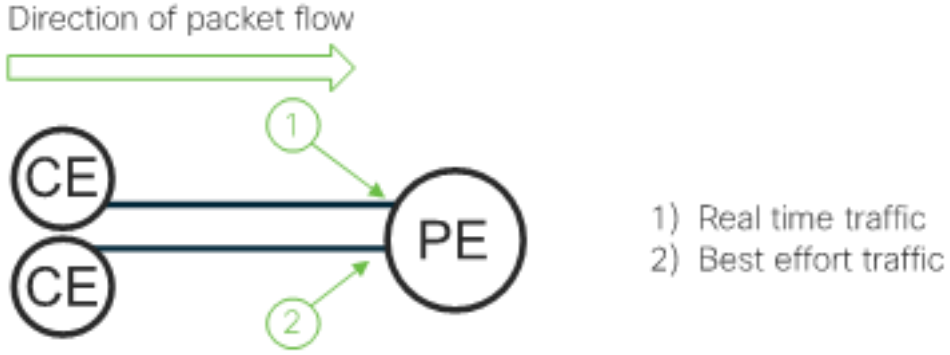


Figure 23-2: Context based classification

In a second example, all traffic arriving on a specific interface might be classified (categorised) as “real time” or “high priority” (see Figure 23-2).

Whilst use of context-based classification is important, it is a somewhat blunt tool, and assumes that all traffic arriving on an interface is of a single type and requires common treatment.

Packet based classification

Packet based classification gives us a more selective tool with which to categorize traffic arriving at a node. In this mode the node examines one or more fields in the arriving packets (or frames) to make a categorisation decision.

One approach is to classify traffic into categories based on some flow or application information. E.g. looking at source or destination, ether type, IP protocol, UDP or TCP port number etc. In this way the node can infer the required QoS capabilities (e.g. low latency, minimum bandwidth, best effort etc) from the type of traffic embedded in the packet.

However, often the most pragmatic approach is to require that packets are pre-marked with an explicit class of service marking. The specific marking model varies dependent on the transport being used, but the primary marking schemes to support are Ethernet Priority Code Point (PCP) and Drop Eligible Indicator (DEI) fields in Ethernet IEEE 802.1p (part of the IEEE 802.1q VLAN header specification), IP TOS, PREC or DSCP ECN fields (defined in IETF RFC 2474 [47] and RFC 3168) and MPLS TC (EXP) field (defined in IEFT RFC 3032 [58] and RFC 5462).

The use of a class of service (COS) marking allows a packet source or network element to encode the required QOS capabilities for a packet directly into the packet header. By predetermining the marking scheme to be used for the COS field, classification using the COS field alone is often sufficient to determine the PHB needed in the network node.

Traffic marking

As described above, a key attribute of a QoS architecture is the ability to pass information about the QoS status of a packet between nodes, and indeed within a node (e.g. from ingress to egress). This is achieved through the use of QoS markings, setting bit fields within a packet header that convey some information about the QoS related context in which this packet should be considered.

Again, as described above there are numerous traffic marking schemes that are in use, each associated with a particular header type in the packet (or frame) world. There are also bit fields defined in meta-data associated with internal forwarding in many router and switch silicon architectures. Two examples of this are “Discard Class or Loss Priority” often associated with the output of a policing function, but also explicitly settable, and “qos-group” or “forwarding-class”, an opaque field provided in many router implementations

The use and meaning associated with the values represented by these bits is entirely arbitrary, as defined and used by users and operators. However, the value in their use lies in having agreed meaning that can be passed between network elements. Some fields, for example IP DSCP, have recommend meanings that can be used between providers. Others, such as MPLS TC (EXP) are not formally defined so as to remain flexible.

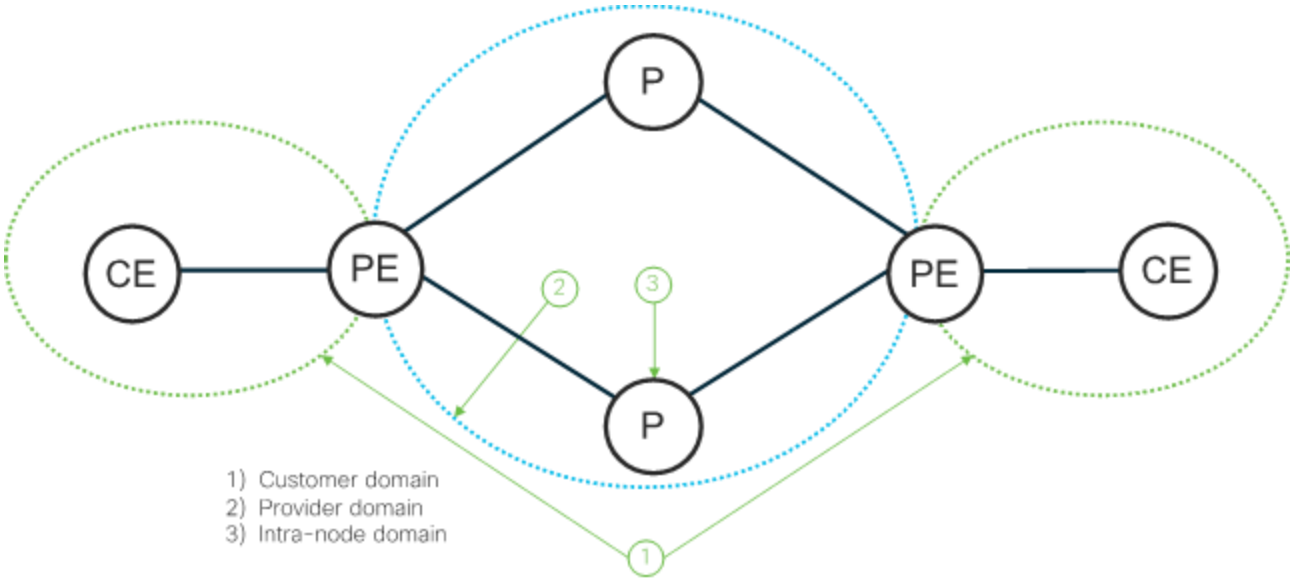


Figure 23-3: QoS marking domains

A useful concept is to consider the use of a given marking scheme within a QoS marking domain. Nodes within a domain use a common scheme, allowing identification of traffic with common properties (a traffic class). Policies are then implemented in each node to maintain the appropriate network behaviour for each identified class, when under congested conditions.

At the border between domains, the border element should be capable of classifying traffic in the first domain and remarking it for transport in the second. Usually the policy for forwarding will be implemented to honour the class structure for the domain it is in. In Figure 23-3, three domains are identified, 1) the customer domain, 2) the provider domain 3) an intra-node domain. The PE

elements on the left and right of the provider domain border the provider and customer domains and will be responsible for remarking traffic as appropriate.

Congestion management

Thus far we have discussed the options available for classifying traffic. But once classified, we must consider what behaviour to apply. As mentioned in the introduction to this section, a QoS node fundamentally only has three options available for managing traffic; forward, store for forwarding later or drop, but these options need only be applied when a bandwidth resource is congested.

A key function of QoS is that of managing traffic during congestion of an interface. Congestion occurs when the quantity of traffic to be forwarded on an interface exceeds its capacity. In its simplest form, QoS alleviates this issue by buffering the excess traffic, i.e. storing it in a local memory until such time that the interface is available to send the stored traffic.

FIFO Queueing

By storing packets that need to be sent on to this interface in the order they arrive, and then sending them in that order when the interface is free, we form a “first in - first out” (FIFO) queue of packets.

This mechanism, queueing, is a useful tool for managing temporary congestion, smoothing out the bursts of traffic flowing into an interface, and forms the basis for more complex congestion management tools.

There are four aspects of this capability that cause challenges;

1. By storing traffic for later sending, we intrinsically delay that traffic that is being stored, adding latency to the transmission path for that traffic. Further, because the delay may be variable, (different amounts of traffic may be enqueued prior to this packet arriving) we introduce jitter.
2. As the speeds of the interfaces we are managing increases, we need more memory capacity to store enough packets to substantively impact the traffic we are buffering.
3. The process is of queuing traffic is indiscriminate, in that it impacts all traffic flowing through an interface. This presents the same latency implication to all applications using this interface, some of which will be impacted by this delay, and some will not.
4. As the amount of memory available to for the queue is finite, so the queue is finite. As a result, if the rate of arrival of traffic to be forwarded is significantly higher than the interface rate or the burst is long enough, the available buffer will be filled, and no further traffic can be enqueued. At this point, the node has no option but to drop traffic. When a queue is full, and a node can no longer add packets to the end (tail) of the queue, traffic otherwise destined for that queue is indiscriminately dropped from the tail (tail drop), until such time that the queue depth reduces and space is made available again for new traffic.

It is worth remembering at this point that the application of QoS to an infrastructure is only a solution for the management of how traffic is dropped under congestion conditions. It is not a replacement for capacity planning and management for traffic that MUST be delivered. I.e. if a network has less capacity available than the amount of traffic that MUST be delivered, applying QoS is not an alternative to augmenting the capacity.

However, it is possible to significantly improve on this single FIFO queue model by combining the use of classification with a programable scheduler to determine how we enqueue traffic, how much we allow to pass, and what to do with the excess.

Class based queueing

By combining the concept of classifying traffic with the concept of a queue it is possible to allow a node to manage the traffic flowing on an interface with a behaviour that is appropriate to the applications flowing in each class.

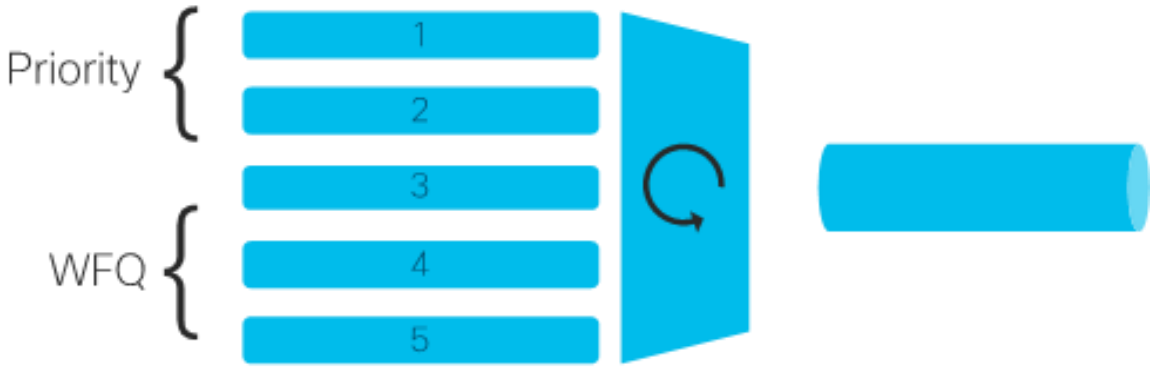


Figure 23-4: Class based queue

In this model, shown in Figure 23-4, the system identifies separate logical queues for traffic destined to flow out of a given interface. Traffic can be placed into a queue based on the classification information gained on arrival (or if the system is capable, after the packet has been forwarded – determined that it should flow out of this specific interface).

For example, traffic identified as needing low latency, and low jitter (real time – our voice calls from earlier) may be placed into queue 1. Traffic requiring some minimum bandwidth (business critical, our EPOS traffic) may be placed into queue 3. Traffic that has no specific requirements (best effort) may be placed into queue 5.

The scheduler can now be programmed to dequeue traffic and place it onto the interface following a pre-determined policy – for example;

1. Serve all high-priority traffic first (E.g. Queue 1 and 2), up to a given capacity of the interface for each queue (for example 50% and 5%). Any traffic arriving at this interface for those queues that cannot be served at that rate should be placed into the appropriate queue.
2. Serve traffic from low priority queues (queue 3, 4 and 5) with equal priority (a packet from each) until the minimum guaranteed amount of bandwidth allocated to that queue is met.
3. Continue to serve traffic from low priority queues until either the queue is empty, or the interface bandwidth is consumed.

This type of policy allows maximum use of the available bandwidth whilst guaranteeing each application class will have access to at least the capacity it needs to support the user base.

Hierarchical class-based queueing

A further extension of this model uses a tiered set of schedulers to manage sets of queues that are “grouped” to manage traffic associated with some logical entity on the physical interface.

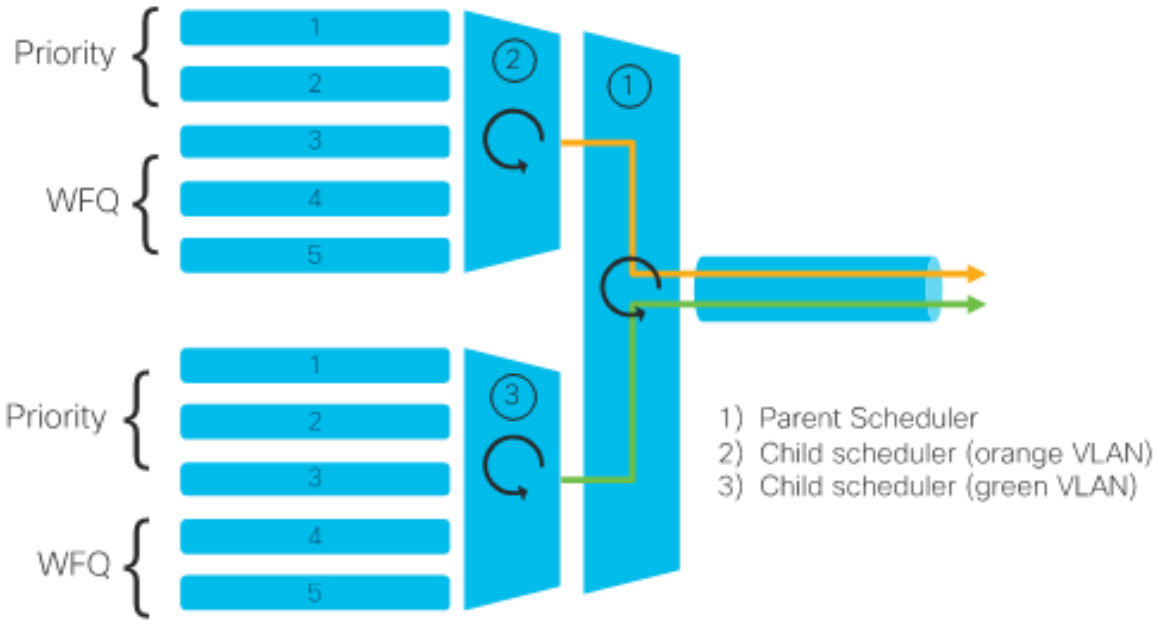


Figure 23-5: Hierarchical class-based queue model

In Figure 23-5, we have a physical interface supporting two VLANs (orange and green). Each VLAN has its own set of queues and its own scheduling policy for those queues. In this example the orange VLAN has two priority queues and three guaranteed bandwidth queues. The top scheduler (2) is responsible for managing the flow of traffic from the queues associated with that VLAN. The bottom scheduler (3) manages this flow of traffic from the queue associated with the green VLAN.
Schedulers (2) and (3) are regarded as child schedulers, and their packet flow is in turn managed by scheduler (1), the parent scheduler for the interface. The parent scheduler will be programmed with a policy to manage the traffic flowing from the child schedulers toward the interface.

This model can allow for a “hard” distribution of the bandwidth on an interface amongst the logic entities sharing that capacity, protecting the use of the bandwidth for a given VLAN (orange) from the others on the same interface (green).

Congestion avoidance

Thus far we have discussed the options available for classifying traffic, and managing traffic when interfaces are under congestion. We have examined how, when an interface is congested, we can consider which packets to forward and when. However, one important aspect of QoS is the prevention of congestion in the first place.

Congestion avoidance predominantly takes two forms;

1. Admission control
2. Selective queue management through random early discard (RED) or weighted random early discard (WRED)

Admission control

This is the process of managing the amount of traffic that we accept into the network. This can be thought of as the enforcement of the contract between the applications requirements', and the networks transport capability. Returning to the example of the voice call, let us imagine that the transport network has enough capacity to sustain 100 concurrent voice calls (Figure 23-6). If more than 100 calls are introduced there is insufficient bandwidth to sustain all the calls and some traffic will inevitably be dropped. Because the dropping of traffic will be random, all calls will be impacted by the packet loss equally, resulting in reduced voice quality or even call failure.

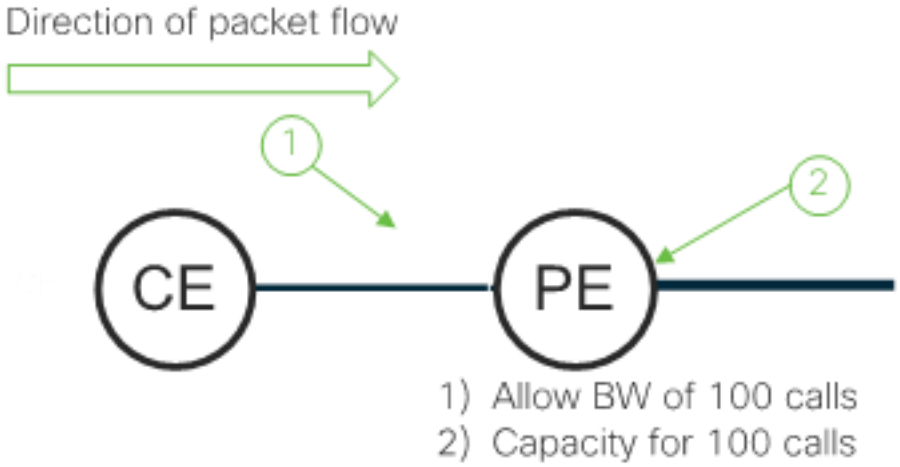


Figure 23-6: Admission control policing

By pre-agreeing that the network can only sustain a certain number of calls, we can police the traffic arriving from the application, to the agreed level. In the case of the packet network, this means agreeing a sustainable bandwidth level and honouring that level at the point that we accept traffic from the application.

Once we have classified traffic to determine that it is indeed voice traffic, we can apply a policer to that traffic class to limit the amount of traffic that is accepted. If the application violates the agreed rate, the policer will drop some packets, forwarding only the rate that was previously agreed. It should be noted that for this particular application, again voice quality will suffer, however, the onus is now on the application to stay within the terms of the agreed contract for the benefit of its own service.

So, if by dropping traffic at admission to the network we will see the application service degrade, just as if we drop traffic in the middle of the network, why bother? Should we not simply let the traffic have access and hope there is sufficient capacity to support the application demand?

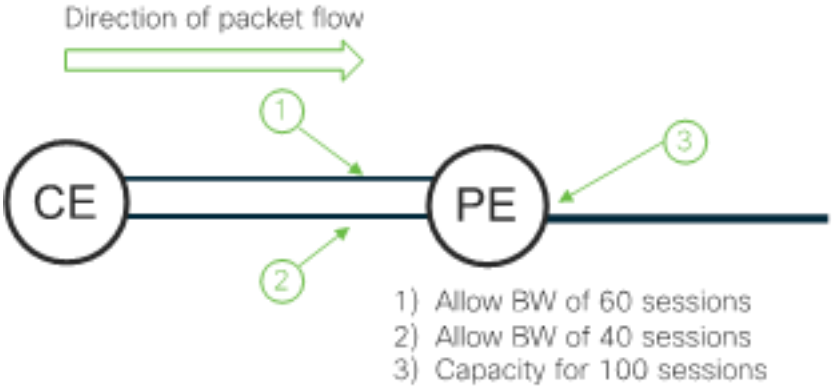


Figure 23-7: Service protection through admission control

This is where the network level administration of QoS becomes relevant. Consider the same core network resource now being shared by two customers (Figure 23-7). If we recognise that the network can only sustain 100 calls, we can ask each customer to restrict their usage, customer 1 to 60 calls and customer 2 to 40 calls. By applying a policy that is specific to each customer context we can ensure that the available network resources are not exhausted, resulting in poor performance for all.
If customer 1 exceeds their allocated capacity, some traffic will be dropped at point 1, resulting in poor quality for customer 1. However, if customer 2 continues to maintain their agreement not to exceed 40 calls, there will always be sufficient capacity to support their service without quality impairment.

Weighted Random Early Discard (WRED)

When looking at congestion management, the idea of a queue was introduced in order to store traffic that is awaiting transmission on an outgoing interface. In this previous discussion we saw that in a simple queue, when the queue is full, any additional traffic is “tail dropped” indiscriminately.

For applications that are using a TCP connection over IP, the loss of one (or more) packets from the flow will cause the TCP algorithm to time out and request re-delivery of these missing packets. This also triggers the shortening of the TCP “ack window” resulting in lower throughput for that TCP session. As many TCP session may be impacted at once in a tail drop scenario, this can cause a number of sessions to “back off” reducing the total load on the interface, and the queue to be drained.

Over time, the TCP sessions will re-open their ACK window, increasing bandwidth until the interface is once again full and the queue overwhelmed. The result is a sawtooth type bandwidth profile that oscillates around the maximum interface speed. This oscillation reduces the efficiency of the interface usage as there are periods where the interface is not full.

Weighted Random early discard (WRED) is a modification to the queue drop behaviour that can help to eliminate this effect. The premise is to manage the acceptance of traffic into the queue in a more intelligent manner.

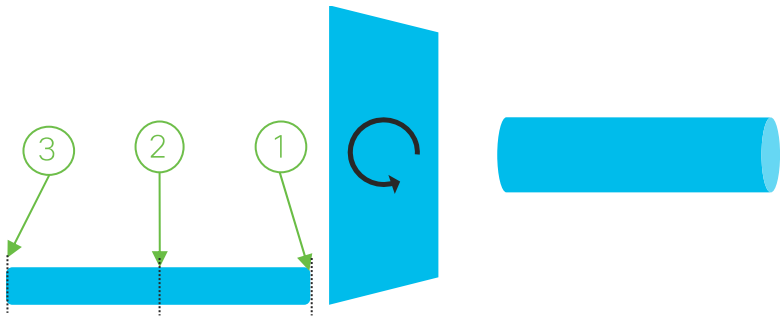


Figure 23-8: RED queue threshold

Consider the queue in Figure 23-8, with three depths at 0% (1), 50% (2) and 100 % (3).
As traffic arrives for the interface and the scheduler has no capacity available to forward them, the waiting packets are placed into the queue. Between queue depth 1 (0% full) and depth 2 (50% full), packets are added as normal to the queue.

When the queue depth reaches the lower RED threshold (2) (50% in this example), rather than adding all new packets to the queue, some will be randomly dropped. The probability of dropping any given packet changes dependent on the current (and possibly historical) queue depth. As the depth of the queue increases, so does the probability of drop until at depth 3 (100% full) the probability of drop is 1 (always drop).

The impact of this weighted behaviour is to slowly encourage individual TCP sessions flowing in the queue to find an appropriate TCP window (and hence bandwidth) such that the queue never saturates, and the link can continue flowing at 100% capacity, eliminating the sawtooth behaviour.

Class based Weighted Random Early Discard

We have seen it is possible to add traffic to a specific queue based on a classification, in order to differentiate between traffic types flowing in an interface. This allows us the flexibility to manage bandwidth to support different applications in an infrastructure. We have also seen that under some circumstances it might be possible to “borrow” bandwidth from one class to use in another class, in the event that the bandwidth in an interface is not fully used.

It is possible to extend the idea differentiated behaviour for traffic from between queues, to within a single queue using the Weighted RED scheme.

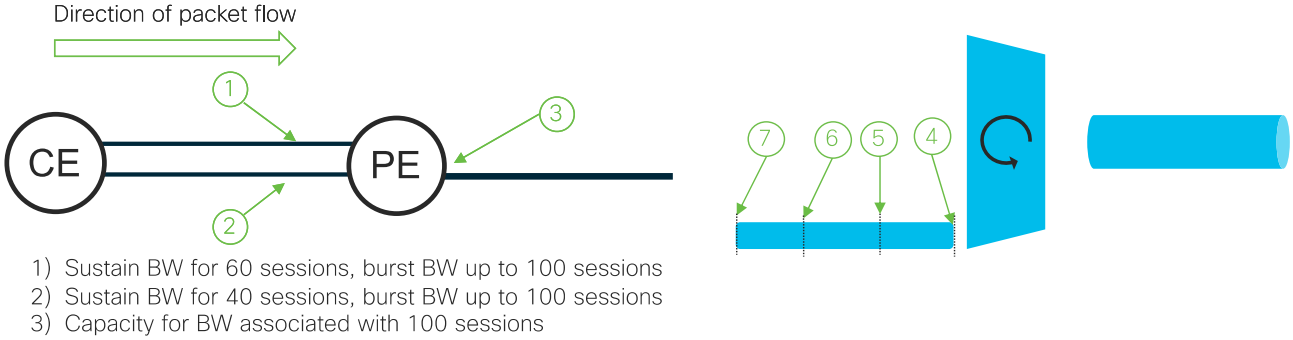


Figure 23-9: Class based WRED function and behaviour

Consider the PE node on the left of Figure 23-9, two CEs are sending EPOS traffic. The PE has capacity at (3) for 100 concurrent sessions. The number of sessions sent by each CE varies. CE 1

has a minimum sustained rate of 60 sessions, and CE 2 has a minimum sustained rate of 40. Each has a possible peak rate of 100.

As traffic arrives from CE 1 at point (1) the PE node classifies traffic as part of the EPOS application. It also measures the amount arriving and allows up to 100 sessions worth into the network. For traffic that exceeds the rate of 60 sessions, the node marks this traffic as having exceeded its contracted rate or “out of contract”. The same classification, policing and marking takes place at point (2) for CE 2 with a rate of 40 sessions. This extra marking is termed the discard class or DC.

When the incoming packets arrive at the queue depicted on the right, again a decision is made prior to enqueuing the traffic. This time the decision is based on two criteria, the current queue depth AND the discard class of the packet as determined in the ingress policing and marking stages.

A sample policy may behave as follows. For queue depth between 0% (4) and 33% (5) all traffic is enqueued.

For traffic that is marked with the out-of-contract discard class marking, between queue depth 33% (5) and 66% (6) the drop probability varies from 0 to 1; that is at depth 66% (6) all traffic that is marked out of contract will be dropped. Traffic that is marked “in contract” will continue to be queued for transmission.

For queue depths between 66% (6) and 100% (7) no out of contract traffic will be enqueued, and the drop probability for in contract traffic will increase from 0 to 1.

It should be noted that the PE is not policing “sessions”, but simply the traffic associated with those sessions. The PE has no knowledge of the nature of the traffic sent by the CE (for example the membership of a specific packet to a specific session), other than the markings being examined. It should also be noted that traffic dropped at point (3) will contain packets associated with many or all of the sessions from the CE whose traffic is being marked “out-of-contract”. The impact of this is that all TCP sessions that lose traffic will again contract their TCP windows, with the effect of lowering the BW used by each session and so reducing the load on the network.

This behaviour has the following benefits;

1. both CE users 1 and 2 potentially have access to all the bandwidth at point 3 in the diagram when the other user is not consuming it. The network resources at 3 can be optimally used when demand from one user is high and the other is low. Compare this to the example in Service protection through admission control where the network resource at point 3 may be under used.
2. If a user constrains their traffic to within the contracted limit, they are guaranteed delivery, even under congestion caused by out-of-contract traffic being generated by other users.

24 Annex E: Multicast Technologies background

Overlay multicast

Overlay multicast refers to CE to CE multicast over the transport provider/underlay network. In the following example, CE1 is a First Hop Router (FHR) connecting to the source and CE2/3 are Last Hop Routers (LHRs) connecting to receivers. Receivers send IGMP/MLD joins towards the LHRs, who then send PIM joins towards their upstream routers PE2/3. If the PEs were connected by a LAN, PE2/3 would simply send PIM joins onto the LAN towards PE1. Since PEs are actually Provider Edge routers that are not directly connected and typically serve many VPNs, Overlay Multicast signalling is used to signal the overlay multicast state over, not through, the provider/underlay network.

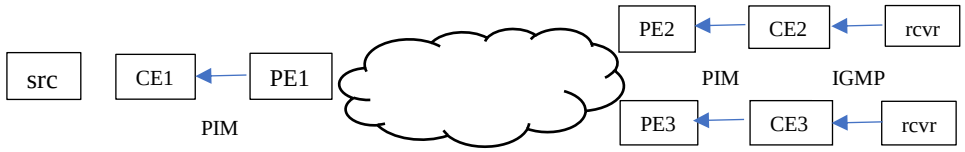


Figure 24-1: Multicast procedures

PIM-based overlay signalling for IPVPN

The most straightforward way is to emulate a per-VPN pseudo LAN over the underlay network. The pseudo LAN is instantiated by a multicast GRE tunnel for the VPN. Every PE for the VPN joins that tunnel and can send both control plane (PIM messages) and data plane traffic over the tunnel. For example, multicast group 225.1.1.1 used for VPN1, 225.1.1.2 for VPN2, etc.

This is referred to as PIM-MVPN and also commonly as Rosen-MVPN. It is documented in the historic RFC6037 and further in RFC6513 [114] (which documents both PIM- and BGP-based overlay signaling).

Essentially, a PE runs PIM sessions over each pseudo LAN for each VPN. This is different from unicast, with which no per-VPN session is used.

When there are a large number of VPNs, many PIM sessions and signalling (including message refreshes) are incurred so it does not scale well. Most implementations and deployments of PIM/Rosen-MVPN only use IP multicast tunnel in the underlay to emulate the pseudo LANs.

BGP-based overlay signalling for IPVPN and EVPN

For BGP-based signalling, the overlay signalling is done by BGP just like unicast case. It is referred to as C-multicast signalling (C for Customer) in RFC6513 [114]. BGP-based signaling use unified methods for both unicast and multicast, and inherits all BGP scaling mechanisms and properties.

Multicast (IP) VPN with BGP based overlay signalling is referred to BGP-MVPN. While RFC6513[114] covers both PIM-MVPN and BGP-MVPN concepts, BGP-MVPN specific encoding and procedures are specified in RFC6514 [115] .

BGP-MVPN also defines mechanisms to signal what type and instance of provider/underlay tunnels are used to transport overlay multicast traffic over the underlay, via Inclusive or Selective PMSI Auto-Discovery routes (often referred to as I/S-PMSI or x-PMSI routes). IP multicast, RSVP-TE P2MP, mLDP P2MP/MP2MP, Ingress Replication, BIER and future additions could all be used as underlay tunnels.

EVPN BUM (Broadcast, Unknown unicast and Multicast) only uses BGP overlay signalling and is modelled after BGP-MVPN. It is specified in RFC7432 [116] and further in draft-ietf-bess-evpn-bum-procedures-update [157].

BGP-MVPN and EVPN BUM have been widely deployed and are often referred to as Next Generation MVPN (NG-MVPN).

Underlay multicast

As mentioned above, overlay multicast traffic is transported via underlay multicast tunnels. Many tunnel types have been mature and widely deployed, but with Segment Routing and central control being widely adopted, it is necessary to review the pros and cons of various multicast technologies and see what is the best solution in the SR era.

Traditionally, multicast requires per-tree state on all routers of a tree for efficient replication. PIM, RSVP-TE P2MP, mLDP P2MP/MP2MP are all signaling protocols to instantiate tree state on those routers, either from the tree leaves towards the tree root or vice versa. The new SR-P2MP (aka tree-sid) tunnels also requires per-tree state on all root/leaf and replication nodes, except that the signalling is directly from a central controller who also calculates the tree.

Obviously, this does not go well with one principal of Segment Routing – no per-flow state inside the network. An alternative is Ingress Replication – the root tunnels individual copies directly to leaves so no per-tree state is needed inside the network. Since replication is done at the root, it is not efficient and does not work well for high fan-out hight data rate cases.

BIER is a new technology that achieves efficient replication w/o incurring per-tree state, so it is the ideal multicast solution for an SR network. However, it uses a new encapsulation and forwarding algorithm, so it requires new hardware. While this makes it difficult to deploy in non-greenfield networks, there are very well-designed brownfield deployment methods, and all major vendors have BIER implementation & hardware available/upcoming.

For an SR network, an operator may prefer to remove all legacy multicast signaling – PIM/RSVP/mLDP and switch to controller-based tree calculation and signaling via either PCEP or BGP (i.e., SR-P2MP [draft-ietf-pim-sr-p2mp-policy[162]] [draft-ietf-spring-sr-replication-segment] or BGP-signaled IP/MPLS multicast [draft-ietf-bess-bgp-multicast-controller[153]]). Notice that this still incurs per-tree state inside the network and the only real benefit is the central calculation of the tree based on many TE constraints/considerations.

In summary, depending on many factors, multicast technologies can be considered in the following preference order:

1. PIM or mLDP/RSVP-TE P2MP that are already deployed in existing network
2. Controller calculated/signalled IP multicast or mLDP/RSVP/SR-P2MP tunnel
3. BIER if enough routers in the network support it

MVPN/EVPN and Seamless MPLS/SR

Given the vast scale of the transport network, a seamless architecture is used as described previously. In this architecture, two aspects have impacts to multicast:

a) Only border routes have routes to edges nodes (that connect to 5G NFs like RAN/UPF), while internal routers only have routes to the border routes.
b) Different ASes/areas may support/prefer different types of underlay tunnels

For a), to establish end-to-end underlay tunnels across those different ASes/areas, PIM RPF vector [RFC5496][104] or mLDP Recursive FEC [RFC6512][113] can be used – a leaf router looks up the route to the tree root, with the BGP protocol nexthop being a border router in the local area/AS. It then encodes that border router's address in PIM/mLDP signaling so that internal routers will signal towards that border router instead of the original root. When that border router gets the signaling, it encodes the next border router's address so that the internal routers in the next AS/area will signal towards the next border router, and so on so forth.

An alternative solution for a) is use controller signaled multicast, since the routers become dumb forwarding devices programed with forwarding state for the tunnels by the controllers.

For b), MVPN/EVPN-BUM tunnel segmentation RFC7524 [118] and draft-ietf-bess-bgp-multicast-controller [153] can be used – MVPN/EVPN-BUM procedures will run on the border routers and the signaled tunnel type/identification are changed as MVPN/EVPN routes are re-advertised by the border routers, so that different tunnel types/instances are used in different ASes/areas.

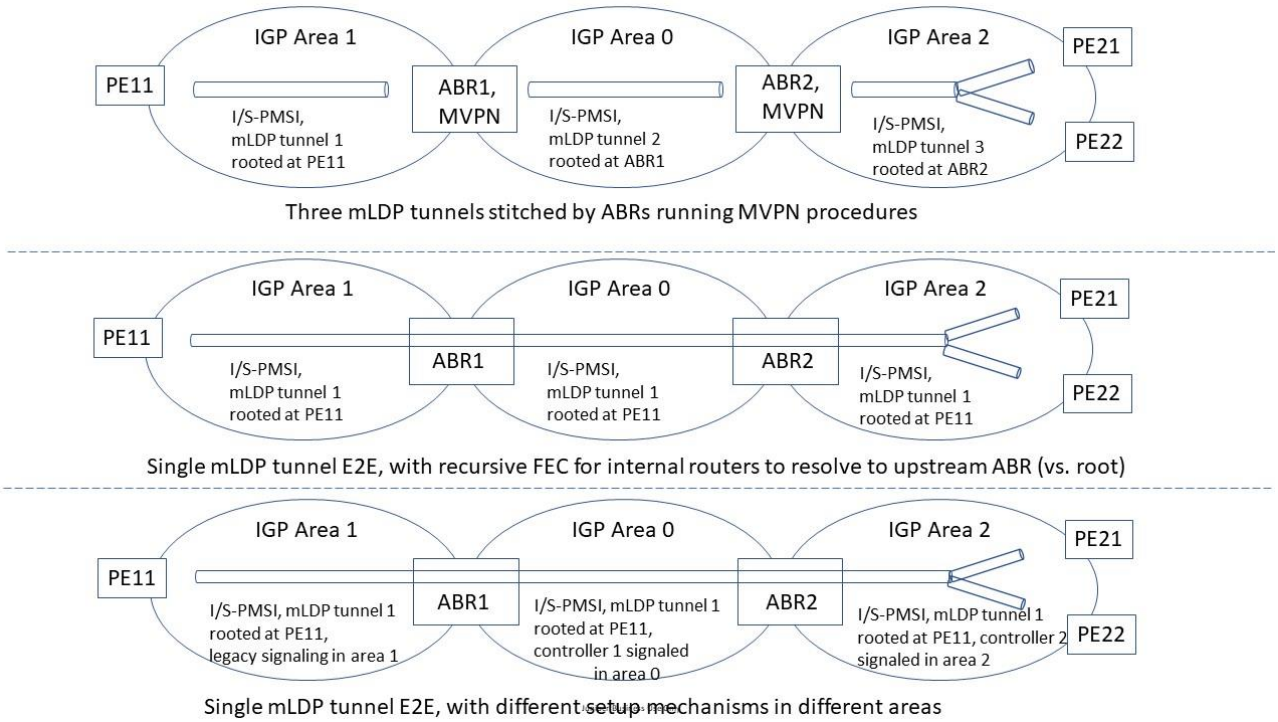


Figure 24-2: Multicast diagram

25 Annex F: Transport network slicing solution for WG1 Slicing (informational)

WG-1 is developing a slicing architecture contained in O-RAN.WG1.Slicing-Architecture-v07.00 [24]. It is a multi-phased effort based on the roadmaps of the different O-RAN working groups combined with use cases specified by O-RAN operators. The use cases are additive, so a slicing phase 2 environment must be able to concurrently support phase 1 and phase 2 use cases.

This annex currently addresses the transport network solution needed to support the use cases in phase 1 and 2 of WG-1's slicing architecture [24]. There are potentially many ways a packet switched network can address the requirements outlined by WG-1, so this section has been presented as an example solution that uses capabilities defined in the main body of this document. This annex is inline with work done by IETF TEAS WG in draft-ietf-teas-5g-ns-ip-mpls dealing with a realization of IETF network slices for 5G networks using current IP/MPLS technologies ([171]).

WG-1 Phase 1 scope

The role of the transport network in a 5G slicing scenario is to provide the necessary connectivity, isolation, and Quality of Service (QoS) so management, control and user plane traffic can flow between the mobile components, making up a slice, in an appropriate fashion.

Figure 25-1 shows the use cases for slicing phase 1 in O-RAN.WG1.Slicing-Architecture-v07.00 [24].

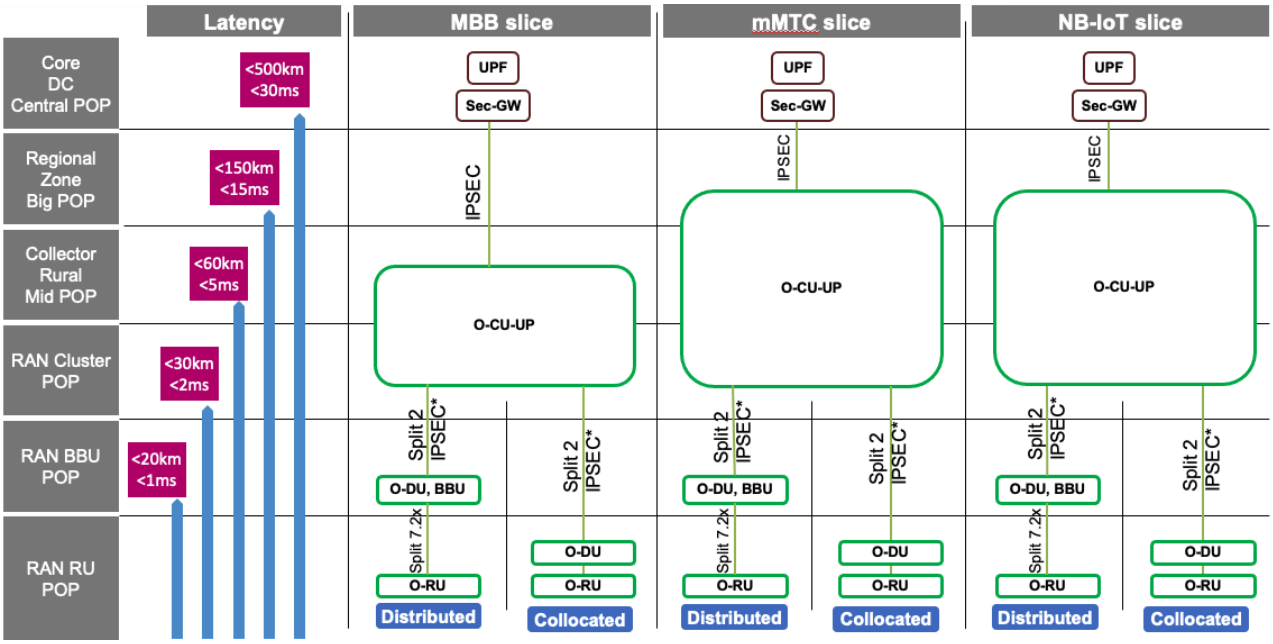


Figure 25-1: WG-1 Phase 1 use cases

In slicing phase 1 the scope is:

- Single Mobile Network Operator (MNO) and Transport Network Operator (TNO) who can be separate organizational entities or the same organizational entity.

- Single management and 3GPP and O-RAN control plane infrastructure for all slices.
- Three public geographically dispersed slices covering mobile broadband, mMTC and NB-IoT.
- Each cell site supports two radio deployments: 1) Distributed: The O-RU is in the cell site and the O-DU is in a data center, located across the transport access network. In this instance, the O-RAN 7.2x C/U protocols and procedures run across the transport access infrastructure. 2) Collocated: The O-RU and O-DU reside in the cell site. In this instance, O-RAN 7.2x protocols and procedures run locally in the cell site and Midhaul interfaces run across the transport access infrastructure.
- All three slices are present in all cell sites and all slices use both radio deployment options.

Not illustrated in Figure 25-1, the following has been assumed or are limitations in slicing phase 1.

- The transport network is multi-service and in addition to 5G needs to support:
  - 3G and 4G mobile services
  - Wireline services (consumer, enterprise, and wholesale)
- Mobile components connect directly to the transport network or reside in a data centre that connects to the transport network.
- O-RAN and 3GPP components present slices to the transport network or Data Center infrastructure via a discrete physical or logical Ethernet interface (VLANs).
- Data Centers, connected to the transport network, present slices to the transport network via logical Ethernet interfaces (VLANs).
- There is no slicing in phase 1 in Fronthaul and Midhaul due to lack of slicing support in these mobile components.
- For the transport network to support QoS appropriate to the QoS requirements of the mobile infrastructure, it is necessary for the O-RAN and 3GPP mobile components, in the Midhaul and Backhaul, to set the DSCP bits in the IP header of the GTP-U packets based on the importance of the user's traffic in the radio domain. In slicing phase 1 this is possible for Uplink and Downlink traffic in Backhaul but only for Downlink traffic in Midhaul.
- No multicast capability in required in slicing phase 1.
- Management architectures and interfaces to packet switched infrastructure for slicing are dealt with in O-RAN.WG9.XTRP-MGT.0-v04.00 "Management interfaces for Transport Network Elements v04.00", July 2022[22]

NOTE: This annex only covers 5G slice requirements, however transport network architects also need to consider the other network services outlined above when designing the transport network.

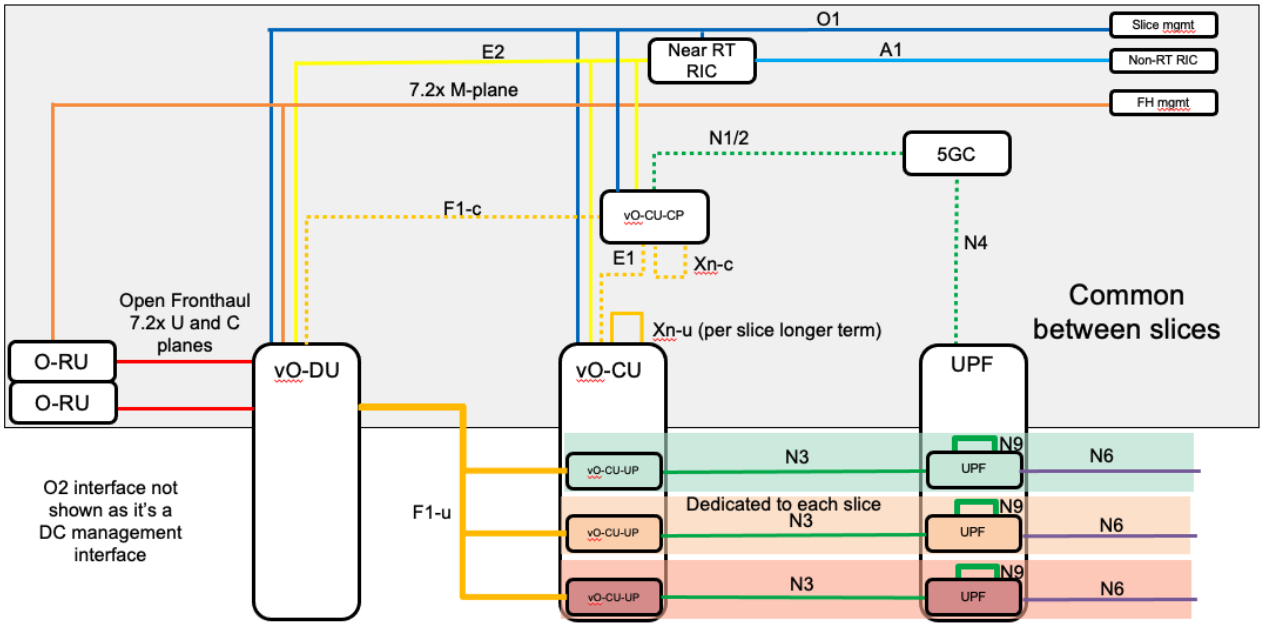


Figure 25-2: 3GPP and O-RAN connectivity requirements for phase 1 slicing architecture

Figure 25-2 developed jointly between WG-1 and WG-9 shows the capabilities and transport connectivity requirements in slicing phase 1. Components and interfaces in the grey shaded box at the top of the figure are shared entities and interfaces, while entities and interfaces shaded green, orange and pink at the bottom of the figure are dedicated to a slice. At the transport level the design in slicing phase 1 provides:

1. A single O-RAN / 3GPP control plane for all slices.
2. Fronthaul service overlay is common for all slices.
3. Midhaul service overlay is common for all slices.
4. Sliced Backhaul user planes are presented to the transport network by mobile components via a discrete VLAN per slice. User plane traffic from each slice is carried in a discrete L3 VPN in the transport network.

WG-1 Phase 2 scope

Figure 25-3 shows the new use cases introduced in slicing phase 2 in O-RAN.WG1.Slicing-Architecture-v07.00 [24].

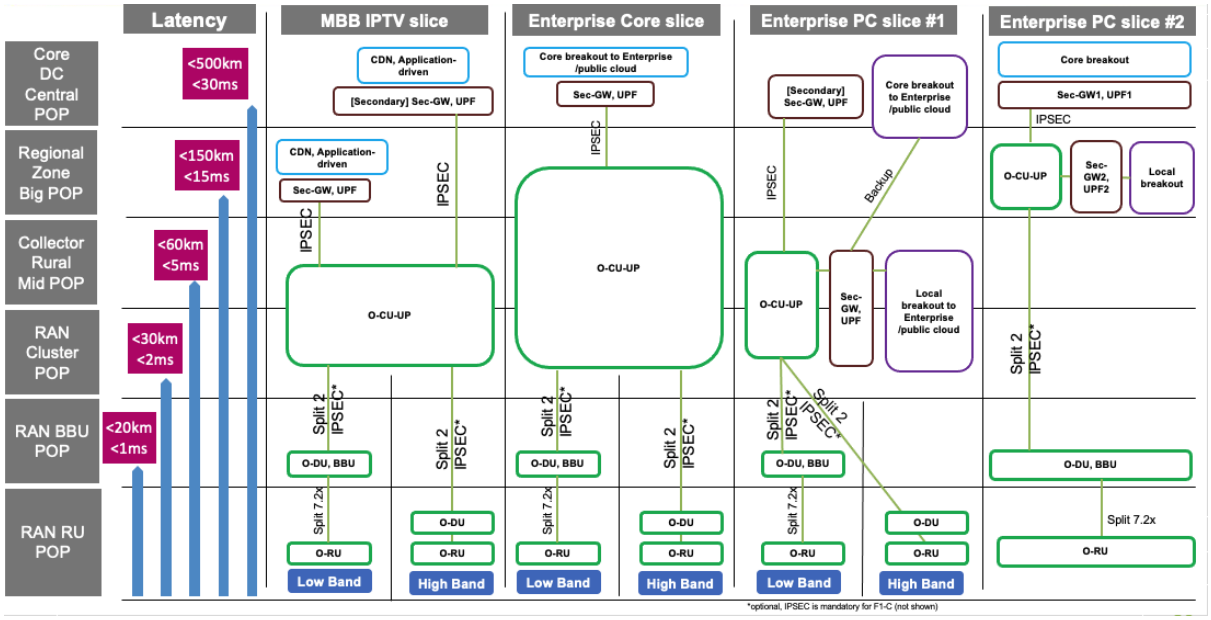


Figure 25-3: WG-1 Phase 2 use cases

In slicing phase 2 the transport infrastructure needs to concurrently support the phase 1 and phase 2 use cases. Phase 2 use cases are:

- MBB IPTV: Slice for IPTV with UPFs and CDNs located at regional and core DC levels
- Enterprise Core Slice: Slices for enterprise customers where the enterprise's UPF is located centrally.
- Enterprise Private Cloud Slice #1 and #2: Slices for enterprise customers where the enterprise's UPFs are in central and distributed locations. Two examples are shown; 1) UPFs are sited in either the RAN cluster PoP or the Collector Rural Mid PoP and the Core DC Central PoP. In this instance all UPFs share common backhaul and N6 networks. 2) UPFs are sited in Regional Zone Big PoP and the Core DC Central PoP. In this instance all UPFs share a common backhaul network but Distributed UPFs are used only for local breakout and the central UPFs are used only for centralised breakout. In this case their N6 networks are discrete logical networks.
- As per phase 1, each cell site may need to support a distributed and/or a co-located radio infrastructure.
- All phase 1 and phase 2 scenarios may be present in all cell sites and all slice types can potentially use both radio deployment options.
- Up to 1000 enterprise slices in total.

Not illustrated in Figure 25-3, the following assumptions and limitations have changed moving from slicing phase 1 to slicing phase 2.

- Slicing of the Midhaul user plane (F1-U interface) is supported in phase 2. This is achieved by presenting either a dedicated Ethernet interface or a separate VLAN interface per slice on O-DUs and the O-CU-UPs. Slicing phase 2 does not support fronthaul slicing.
- Midhaul mobile components associated with user plane traffic (O-DUs and O-CU-UPs) can set the DSCP bits in the IP header of the GTP-U packet for Uplink and Downlink traffic based on the importance of the user's traffic in the radio domain.
- Selection of UPFs for different traffic streams within a slice is the sole responsibility of the 3GPP mobile core.

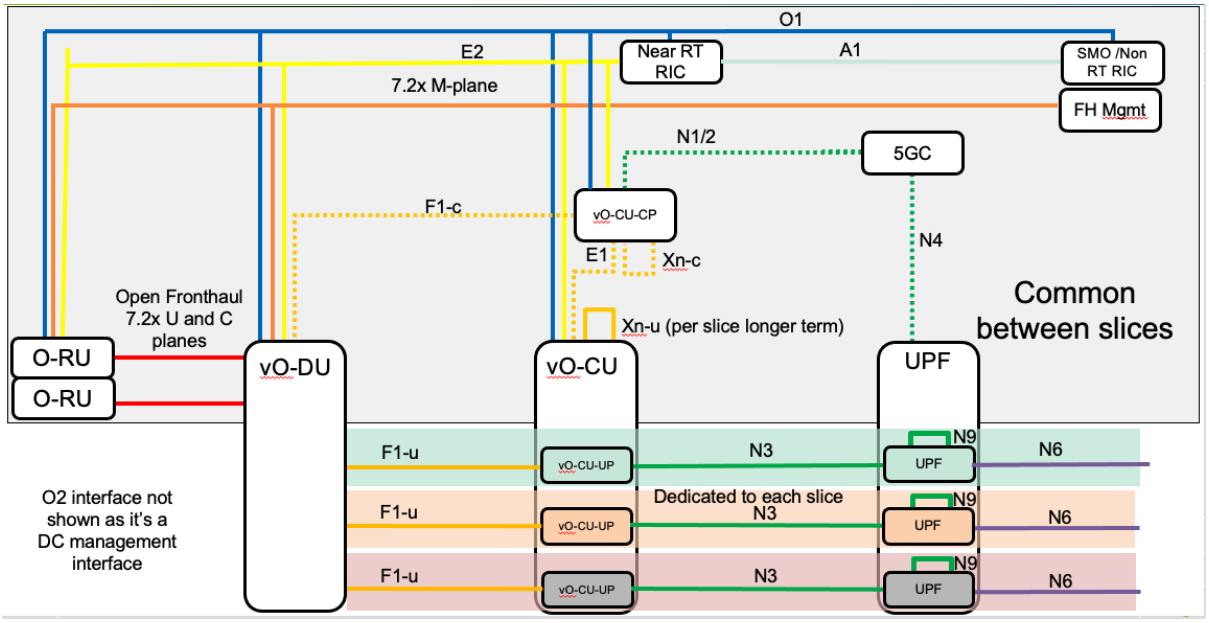


Figure 25-4: 3GPP and O-RAN connectivity requirements for phase 2 slicing architecture

Figure 25-4 developed jointly between WG-1 and WG-9 shows the capabilities and transport connectivity requirements in slicing phase 2. Components and interfaces in the grey shaded box at the top of the figure are shared entities and interfaces, while entities and interfaces shaded green, orange and pink at the bottom of the figure are dedicated to a slice. At the transport level the design in slicing phase 2 provides:

1. A single O-RAN / 3GPP control plane for all slices.
2. Fronthaul service overlay is common for all slices.
3. Sliced Midhaul and Backhaul user planes are presented to the transport network by mobile components via a discrete VLAN per slice. User plane traffic from each slice is carried in discrete L3VPNs in the transport network.

WG-1 Phase 3 scope

Figure 25-5 shows the new use cases introduced in slicing phase 2 in O-RAN.WG1.Slicing-Architecture [24].

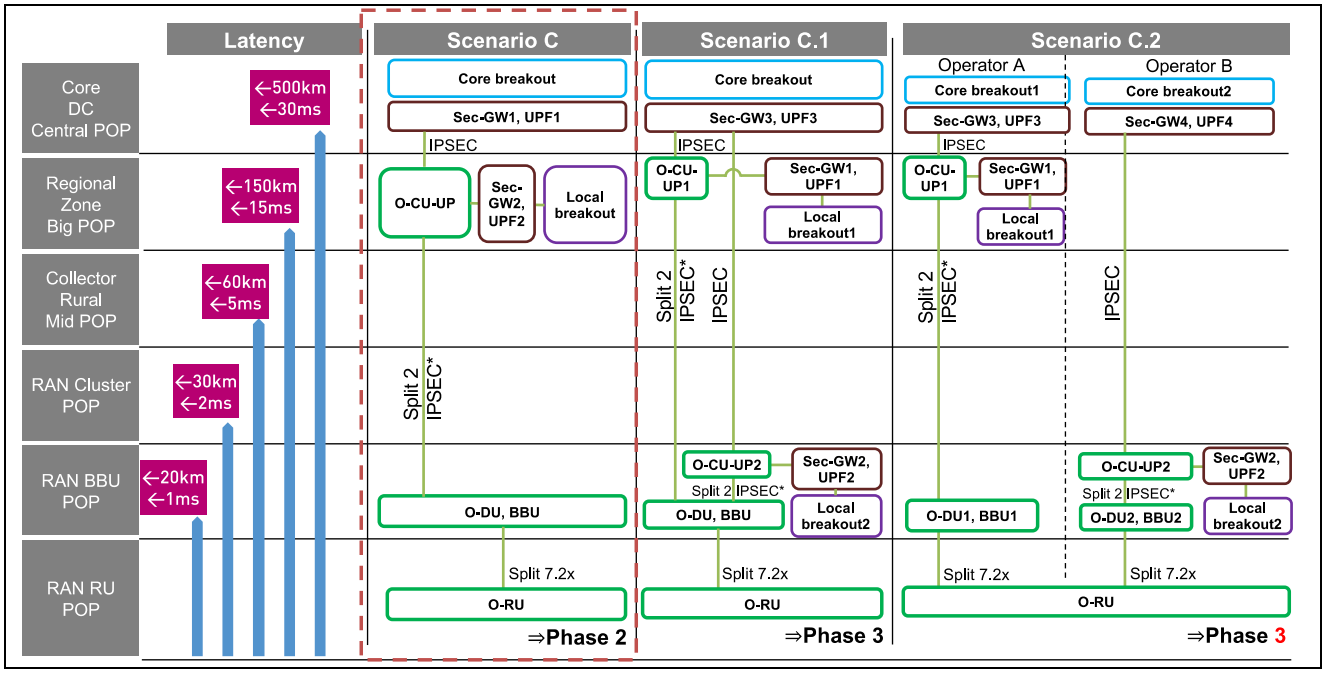


Figure 25-5: WG-1 Phase 3 use cases

WG1 phase 3 slicing doesn't introduce any new requirements on transport regarding DSCP handling. Both, WG1 phase 2 and WG1 phase 3 are based on the DSCP mechanism, in both midhaul and backhaul, as described with Phase 2.

WG1 phase 3 slicing introduces more complex scenarios regarding placement of O-RAN components (Scenario C.1 in Figure 25-5), which doesn't change how the transport network is orchestrated, how the logical isolation between slices is realized with VPNs, or how the slices are handed-off to the transport network (VLAN hand-off).

Overall Packet Switched Transport Architecture

In reading this annex the reader is expected to be fully familiar with the main body and other annexes in this document. The solution is based on a packet switched transport architecture illustrated in Figure 25-6.3

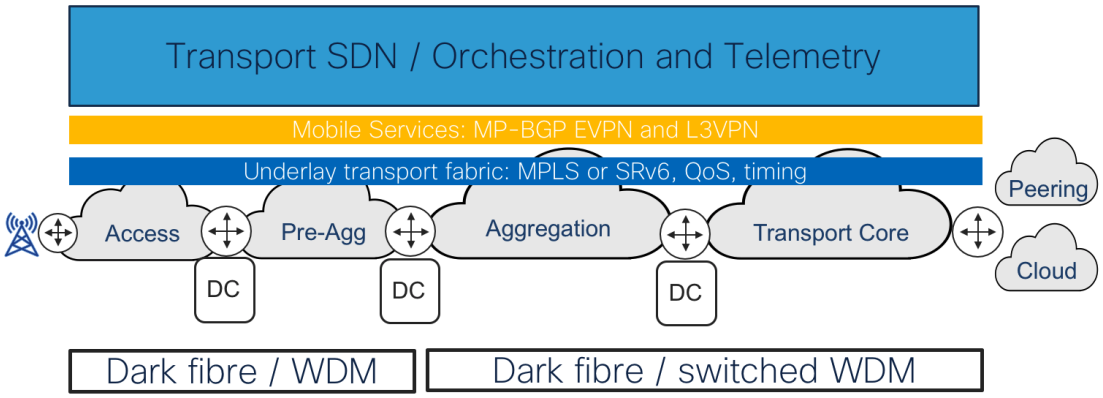


Figure 25-6 Packet switched transport for mobile Xhaul

It is a converged end to end packet switched infrastructure, beginning at the cell site, located in the edge of the access layer, and stretching to the core of the transport layer. The packet switching TNEs are IP/MPLS or SRv6, QoS enabled, high capacity, low latency devices interconnected by point-to-point Ethernet interfaces running at the full capacity of the Ethernet interface, typically using either point to point fibres or via a WDM infrastructure. It incorporates data centers suitably placed across the transport network infrastructure to support virtual and physical NFs associated with mobile and fixed services but also potentially the placement of "Application Functions" associated with value-add services and customer specific application.

The logical architecture is based on a common underlay packet switching infrastructure based on either MPLS or SRv6 overlaid with an L2 / L3 service infrastructure (VPNs) that uses the capabilities of the underlay packet switched network to support the mobiles interfaces.
The underlay packet switching infrastructure provides basic network services such as any-to-any connectivity between TNEs, scaling, fast convergence, shortest path and traffic engineered forwarding, packet-based Quality of Service (QoS) and timing.

The service layer supports native Ethernet services, using EVPN technology and IP VPN services, using MP-BGP based L3VPNs. These services can utilise the facilities offered by the underlay packet switched infrastructure to support the different mobile interfaces in an appropriate fashion. Where possible transport services are built on an end-to-end basis without intermediate stitching / switching points within the transport infrastructure. This approach has been taken to minimise the transport service orchestration overhead.
NOTE: The term MP-BGP L3VPNs is used throughout this annex. It refers to L3VPNs built using MP-BGP based on either RFC4364 [84] or RFC9252 [150] depending on whether the underlay technology is MPLS or SRv6. More details on the technology can be found in section 13 and Annex C: MP-BGP based L3VPNs of this document.

Underlay network for WG-1 slicing phase 1, 2 and 3

A packet switched network has extensive capabilities to support slicing. These are outlined in section 18. Slicing phase 1 is basic in nature with limited slicing capabilities in the O-RAN mobile components and the actual slices (mobile BB, MMTc, NB-IOT) are low priority and delay tolerant slices. In phase 2, some additional slice types are added (MBB IPTV, Enterprise Core, Enterprise Private Cloud), with more distributed 3GPP, O-RAN component and Application Functions (UPFs and CDNs can be located at regional and central DC levels), and more – but still not very critical – delay sensitivity (for example IPTV). Phase 3 adds even distribution possibilities. With this in mind, and more onerous slicing use cases in later phases, it is proposed to support slicing phase 1 and 2 in the transport network using:

- Default IGP instances running in each underlay routing domain calculating shortest path routing based on IGP metrics.
- IGP supporting fast and loop free convergence.
- Inter-domain connectivity performed using either BGP, IGP redistributing or a PCE infrastructure.
- Management, control and user planes in phase 1 and 2 use shortest path routing between TNEs. I.e., the Service VPNs associated with each slice are not mapped to TE tunnels or Flex-algos in phase 1 or 2.

- QoS based on Diffserv model defined in RFC2475 [46] and outlined in section 14 and *Annex D: Quality of Service.*

As the WG-1 slicing use cases and slicing capabilities within O-RAN equipment matures so more sophisticated forwarding techniques, such as SR-TE and Flex-algo may need to be utilised. Therefore, phase 3 introduces mapping between Service VPNs and transport planes, where each transport plane is realized via set of tunnels optimized for some common characteristics. For example, one transport plane could reflect set of tunnels optimized for latency (using latency link metrics for optimization), where another transport plane could group tunnels optimized for high capacity (but not necessarily low latency), using IGP metrics. Tunnels could be TE tunnels (RSVP, SR-TE, SRv6-TE), or different Flex-Algos. This approach might result in different path being taken by different slices. Please see Section 18.1 “Packet-switched underlay network” for discussion on how they can be applied and used to support slicing.

NOTE: When designing and implementing the transport network Service Providers may want to anticipate more onerous slicing use cases and implement facilities such as Traffic Engineering, Flex-Algo from the outset.

Service Models for WG-1 slicing phase 1, 2 and 3

Management and control plane networks are needed to support the management, orchestration and monitoring of the transport network and the data center infrastructure. In addition to these “infrastructure components”, various management and control plane networks are needed for the management and control of the 5G 3GPP and O-RAN mobile environments. In the transport slicing architecture outlined, some management networks are discrete VPN entities, while others combine multiple management functions together in a single VPN entity. The reasoning behind these choices is outlined, but individual Service Provider’s need to assess their organizational needs and adjust accordingly.

Transport and DC management networks

Both the transport network and the data center infrastructure are common capabilities used by a range of different services and in many organizations managed by different organizational groups. In this example the management networks for these two entities are separate, as shown in Figure 25-7. Both these management networks need to be set-up in advance on the transport network and are pre-requisites to building a slice across the transport network or instantiating VMs or containers within data centers.

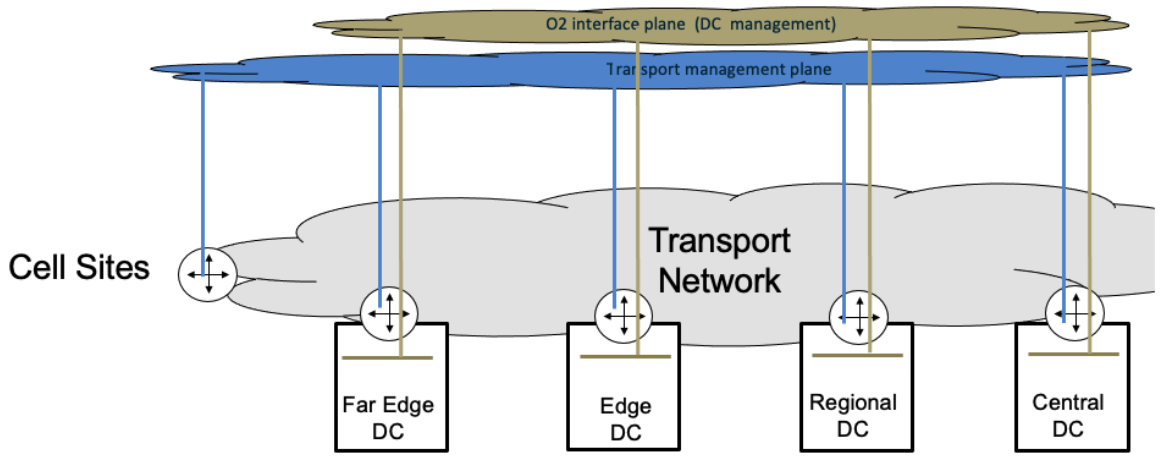


Figure 25-7: Transport network and Data centre management networks

Transport network management network

A network wide in-band L3VPN is used for day-to-day management of the TNEs. Every TNE has a local network management interface with an associated network management IP address configured at installation. This is typically a loopback interface. Management protocols associated with TNE's configuration, monitoring and service orchestration are configured to use this interface as their source interface. The TNE's management interface is placed into a transport management network built using a MP-BGP L3VPN mapped to the underlay's default routing algorithm. The transport network management network can be designed as an any-to-any network, meaning that every entity within the L3VPN can connect to every other entity in the transport network management network but are more commonly built so that the central network management functions can communicate with the TNEs and vice versa but the TNEs cannot communicate with each other.
In addition to the in-band management network Service Providers may also choose to have an out-of-band network for scenarios when in-band network connectivity has been lost. Various solutions exist, often based on access to the console port, through mobile, xDSL or PON technology.

Data Centre (DC) management network

In the packet switched transport architecture data centers are distributed around the transport network and although distributed they need to be managed centrally. The DC infrastructure supports "Network Functions" and "Applications Functions". Some functions are associated with the provision of 5G and O-RAN capabilities and services, but DCs should be considered common infrastructure facilities that support services beyond just 5G. The DC management network needs to support the O2 interface defined by O-RAN but also other data center management, monitoring and orchestration facilities. In the slicing phase 1, 2 and 3 the handoff between the DC and the transport network is via a logical Ethernet attachment circuit. The design of the management infrastructure within the DC is handled by WG-6, with connectivity between DCs and the central DC management infrastructure provided by the transport network. Each DC needs to hand-off the DC management network to a TNE via a logical Ethernet interface. On the TNE the logical Ethernet attachment circuit is placed into a MP-BGP L3VPN mapped to the underlay's default routing algorithm. This VPN connects all the DCs to the central DC management, monitoring and

orchestration functions. The DC management network can be designed as an any-to-any network, meaning that every entity within the L3VPN can connect to any other entity in the DC management network but are more commonly built so that the central network management functions can communicate with the DC components (servers and DC switches) and vice versa, but the DC components cannot communicate with each other.

O-RAN control and management networks

The O-RAN control and management networks are illustrated in Figure 25-8.

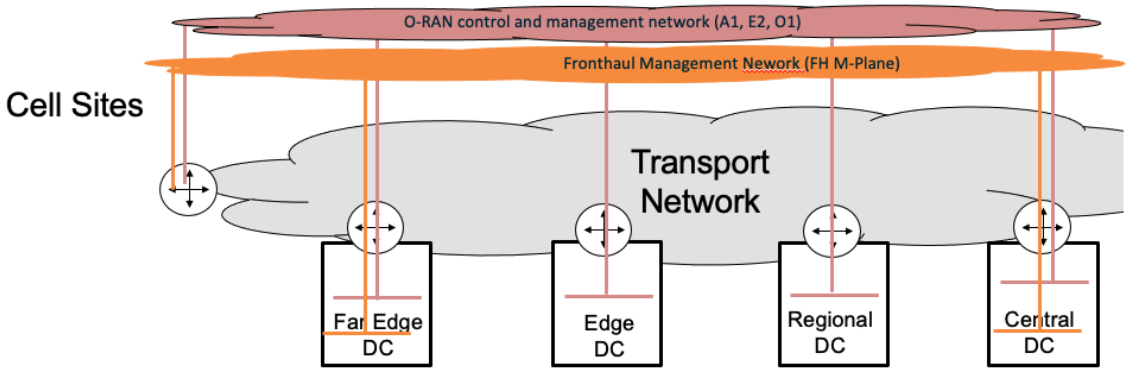


Figure 25-8: Fronthaul Management and O-RAN management and control networks

O-RAN Fronthaul Management network (M-Plane)

Based on guidance from WG-1, the O-RAN Fronthaul M-Plane interface is supported across the transport network as a dedicated VPN. This is discrete from the O-RAN Management and Control VPN that supports the O-RAN A1, E2 and O1 interfaces. The reasoning is Fronthaul management and O-RAN control and management may fall under different organizations within the Service Provider.

Full details of Fronthaul M-plane can be found in section 7.1. The design outlined assumes the Fronthaul M-plane is operating in hybrid mode, where a central Fronthaul NMS needs IP connectivity to the O-DUs and the O-RUs. The O-RAN Fronthaul Management network in the transport network uses a MP-BGP L3VPN mapped to the underlay's default routing algorithm. The VPN needs to be present on all TNEs, through VLAN attachment circuits, that support O-RAN mobile components managed using the Fronthaul M-Plane. O-RAN mobile components that use the M-Plane need to connect to the O-RAN Fronthaul Management VPN, via a logical Ethernet interface, when they are installed or instantiated.

The O-RAN Fronthaul Management VPN can be designed as an any-to-any network, meaning that every O-RAN component in the L3VPN can connect to any other entity in the O-RAN Fronthaul Management VPN, or more likely built so the M-Plane management components can communicate with the O-RUs and O-DUs but the O-RUs and O-DUs are restricted in their connectivity.

O-RAN Control and Management network (A1, E2, O1 interfaces)

Based on guidance from WG-1, the O-RAN A1, E2, O1 interfaces are supported across the transport network as single VPN, discrete from the Fronthaul management VPN. The A1 and E2 interfaces are associated with the RAN Intelligent Controllers (RICs). The O-RAN architecture has two RIC entities:

- Near-Real-Time RIC (Near-RT RIC). Near-RT RICs use the E2 interface to communicate with the O-CUs, and O-DUs they control. The interface controls the O-CUs and O-DUs with execution actions coming from the Near-RT RIC and feedback in the opposite direction.

- Non-Real-Time RIC (Non-RT RIC). Non-RT RICs use the A1 interface to communicate with the Near-RT RICs they control. The interface provides the Near-RT RICs with policies, enrichment info, and ML model updates. In the other direction the Near-RT RICs provides policy feedback to the Non-RT RIC.

The O1 interface is associated with “Service and Management Orchestration”. This interface is an FCAPs interface with configuration, registration, security, performance, and monitoring responsibilities.

The O-RAN Control and Management network in the transport network uses MP-BGP L3VPN technology, mapped to the underlay’s default routing algorithm. It needs to be present on all TNEs through VLAN attachment circuits, that support O-RAN components that use either the A1, E2 or O1 interfaces. All O-RAN components using these interfaces need to connect, via the TNEs, to the O-RAN Management and Control VPN when they are installed or instantiated.

The O-RAN control and management VPN can be designed as an any-to-any network, meaning that every O-RAN component in the L3VPN can connect to any other entity in the O-RAN control and management VPN, or more likely built as a tiered VPN using a combination of Virtual Routing and Forwarding (VRFs) instances and Route Targets (RTs) so that:

- SMO can communicate Near-RT RICs, O-CU-CPs, O-CU-UPs and O-DUs.
- Non-RT RIC can communicate with the Near-RT RICs.
- Near-RT RICs can communicate with the O-CUs, O-DUs and O-RUs.

3GPP Control Plane network

In slicing phase 1, 2 and 3 there is a single 3GPP core infrastructure running in a 3GPP control plane VPN across the transport network. See Figure 25-9 for details. It carries all the 3GPP control plane traffic associated with Midhaul and Backhaul including F1-C, E1, Xn-C, N1, N2, N4 for all slices. In the longer term it is expected that some 5G slices may have a dedicated 3GPP control plane, in these instances they could share the same transport VPN or utilise separate transport VPNs for each slice’s mobile core.

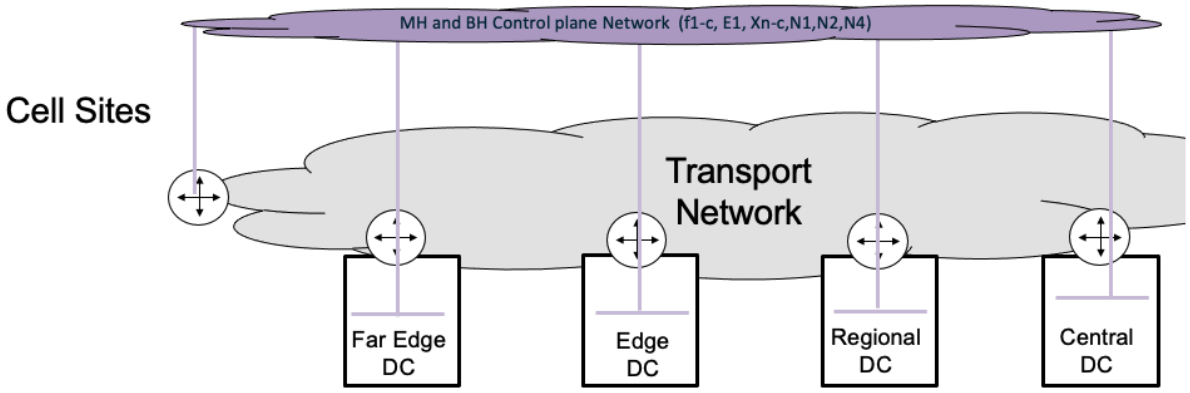


Figure 25-9: Midhaul and Backhaul control plane network

In slicing phase 1, 2 and 3 connectivity for the 3GPP control plane network across the transport network uses a MP-BGP L3VPN, mapped to the underlay's default routing algorithms. The 3GPP control plane VPN needs to be present on all TNEs via VLAN attachment circuits, where mobile components using the 3GPP control plane reside. All 3GPP mobile components using these control plane interfaces need to connect, via VLANs to the TNEs, to the 3GPP control plane VPN when they are installed or instantiated.

The 3GPP control and management VPN can be designed as an any-to-any network, meaning that every 3GPP component in the L3VPN can connect to any other entity in the 3GPP control plane VPN, or more likely built with limited connectivity based on the mobile core connectivity requirements.

O-RAN and 3GPP user planes networks

Slicing Phase 1

Slicing phase 1 transport connectivity architecture for the O-RAN and 3GPP user planes is shown in Figure 25-10. The transport connectivity architecture is driven by two factors.

- Fronthaul and midhaul do not support slicing
- UPFs for all slice types are positioned centrally

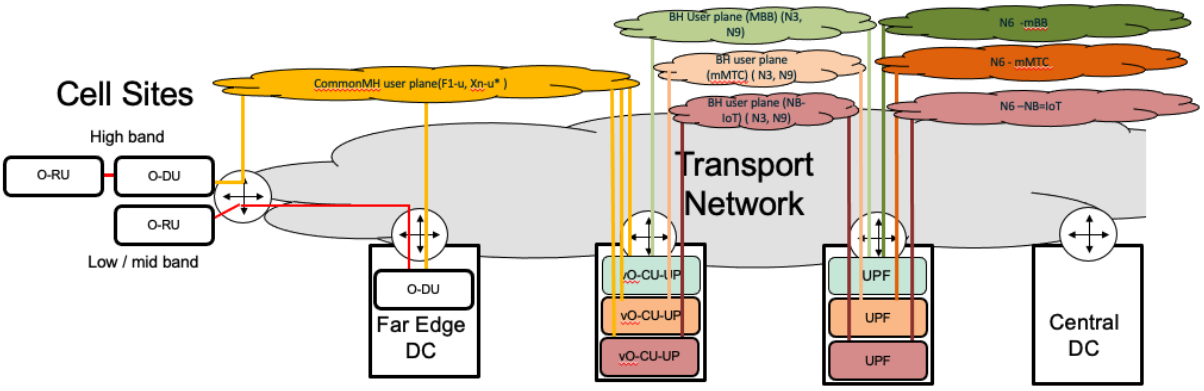


Figure 25-10: Slicing Phase 1 O-RAN and 3GPP user plane networks

Slicing phase 2 and 3

Slicing phase 2 and 3 transport connectivity architecture for the O-RAN and 3GPP user planes is shown in Figure 25-11. The transport connectivity architecture changes to accommodate slicing in midhaul and distributed UPFs and the distribution of the N6 network closer to the radio sites.

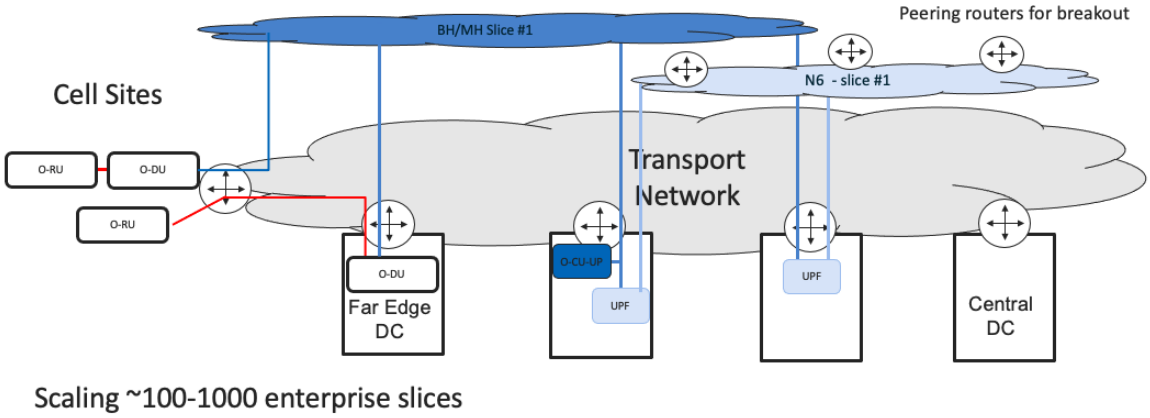


Figure 25-11: Slicing phase 2 and 3 O-RAN and 3GPP user plane networks

This basic set-up can be used for all the phase 1 use cases and those introduced into phase 2 and 3, if midhaul slicing is supported. Each slice now has a single L3 VPN that concurrently supports midhaul and backhaul user plane traffic. Figure 25-11 also shows the N6 network extending deeper into the transport network. This is required, if distributed and centralised UPFs need to be able to access common applications or take advantage of local breakout.

Fronthaul C/U plane network

In slicing phase 1, 2 and 3, slicing is not supported on O-RAN fronthaul components. In practical terms this means O-RUs and O-DUs are common across all slices and there is a single 7.2X eCPRI stream between the O-RU and O-DU which carries all the control and user plane traffic associated

with all slices. The 7.2X C/U interface must support Ethernet encapsulation or optionally support Ethernet / IP encapsulation. Slicing phase 1, 2 and 3 use cases assume the Ethernet variant of the 7.2X protocol, meaning Ethernet connectivity needs to exist between O-RUs and the parent O-DU. Slicing phase 1, 2 and 3 use cases also assumes two radio deployments options:

Co-located O-RU and O-DU at cell site.

Ethernet connectivity needs to exist between the O-DU and the O-RUs and they both reside in the cell site. There are two options:

- Direct point to point connectivity between the O-RU and the O-DU. In this case the transport network plays no part in the connectivity or the provision of an Ethernet service between the O-RU and the O-DU.
- Connectivity between the O-DU and O-RUs is via the "Cell Site Router" (CSR). In this case the Cell Site Router provides layer 2 bridging, or an L2 Ethernet cross connect function between the O-DUs port and the O-RUs port. See section 19.6 for more details.

O-RU and O-DU separated by access network

For distributed solutions, the O-RUs and the O-DUs are separated across the access portion of the transport network. In this instance, an Ethernet service needs to be created across the MPLS or SRv6 transport underlay. This Ethernet service can use either an E-Line or E-Tree service configured using EVPN over MPLS or SRv6. See sections 13.2 for more details.

Midhaul user plane network (F1-U and Xn-U)

Slicing Phase 1

In slicing phase 1 slicing is not supported on Midhaul O-RAN mobile components. O-DUs and O-CU-UPs do not have a concept of slicing with the F1-U traffic from all slices originating and terminating on the same physical/logical interface and IP address on the O-DU. In scenarios, where the O-CU-UP is shared between two or more slices there is also no concept of slicing on the O-CU-UP. In situations, as shown in Figure 25-2, where the O-CU-UP is dedicated to a slice, then the O-CU-UP function will only receive and send traffic associated with that slice, however the O-CU-UP slice assignment and selection process is completely transparent to the transport network and under the control of the O-RAN / 3GPP control planes. For this reason, in slicing phase 1 a single MP-BGP L3VPN mapped to the underlay's default routing algorithm is used across the transport network to support all Midhaul user plane traffic regardless of slice it belongs too. The L3VPN needs to be present on all TNEs and DC-TNEs that either connect directly or indirectly (through the DC infrastructure) to mobile components that support Midhaul user plane interfaces. Access to the Midhaul user plane VPN, on the TNEs, is via VLAN interfaces.

A further challenge with Midhaul user plane interfaces in slicing phase 1 is the lack of a QoS mapping function on the O-DUs to map radio orientated QoS parameters, controlled by the 3GPP / ORAN control planes, to transport orientated DSCP QoS markings. The means the transport network QoS mechanisms must treat all uplink traffic the same, regardless of the importance of the data contained within Midhaul GTP-U packets.

Slicing Phase 2 and 3

Midhaul slicing and QoS is supported in slicing phase 2 and 3. F1-U traffic segregation is supported using a VLAN per user plane slice on the O-DUs and O-CU-UPs and QoS is supported by the O-DUs and O-CU-UPs setting a DSCP value appropriate to the importance of the user payload in the GTP-U packet.

For slicing phase 2 and 3 this annex provides an example where midhaul and backhaul user plane traffic associated with a slice shares a single MP-BGP L3VPN that provides any to any connectivity. If an operator wishes to keep the midhaul and backhaul traffic separate then they could run each slice's midhaul and backhaul traffic in separate MP-BGP L3VPNs or alternatively use "Route Target" (RT) and BGP community filtering within a single L3VPN, so only backhaul mobile components can communicate with each other and only midhaul mobile components can communicate with each other. Examples of how this can be done are in the following section.

Slicing Phase 4

Since Slicing and Sharing utilizes similar toolbox in Transport Network domain, the RAN-Sharing via Midhaul" use-case analysis ([27]) captured by ORAN ATG and ORAN UCTG is included in Slicing Phase 4 of the current document scope. Transport Network isolation may be utilized for NPN 5G RAN-sharing architecture to support multiple operator's midhaul traffic. As shown in NPN RAN-sharing diagram in Figure 25-12, NPN RAN is shared and connected to multiple partner operator's O-CU via shared midhaul in Transport Network domain.

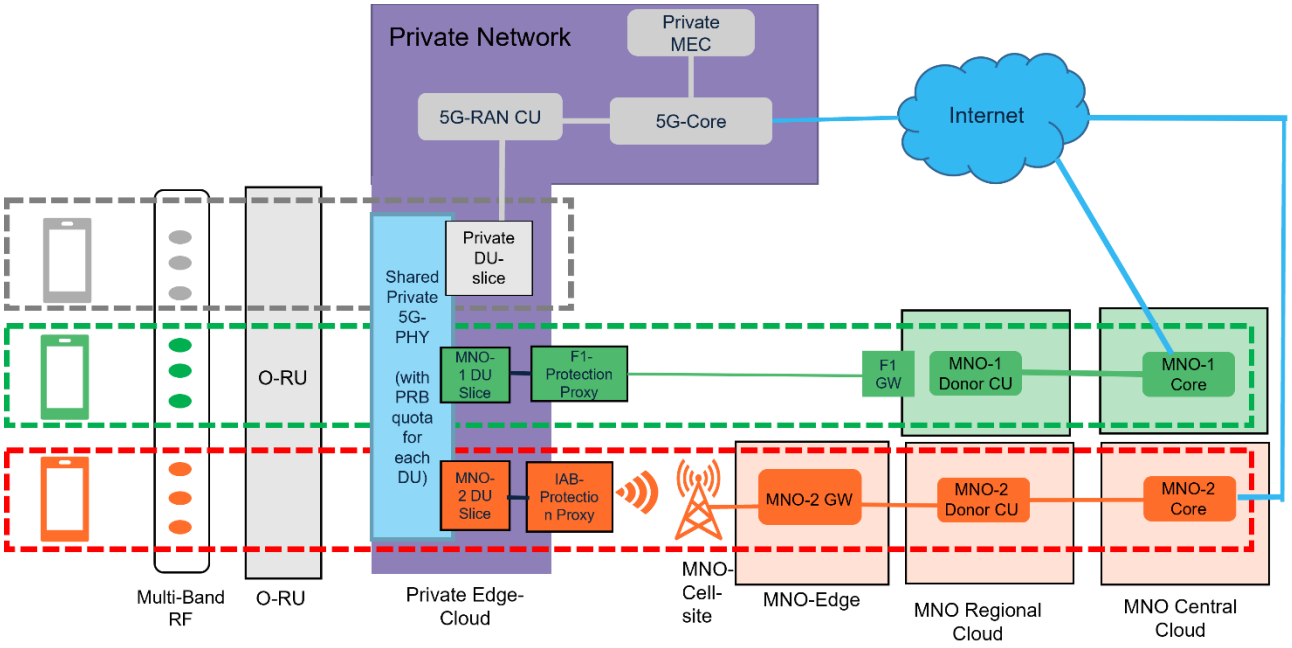


Figure 25-12: NPN RAN-sharing via midhaul

Trusted and untrusted midhaul connection in which MNO's O-CU connects to shared NPN RAN-partition may be isolated using Transport Network L2/L3VPNs and other techniques from Network Slicing Framework of this document.

Backhaul user plane network (N3 and N9)

WG-1 slicing phase 1, 2 and 3 includes user plane slicing in the backhaul network. In phase 1, 2 and 3 it is assumed traffic associated with different slices originates from individual logical Ethernet interfaces on the O-CU-UPs backhaul interfaces and UPFs. These same slice mappings can be maintained across the transport network by mapping VLAN attachment circuits to different L3VPNs on the TNEs or DC-TNEs. Within the transport network MP-BGP L3VPNs can be built per slice or shared between slices depending how the mobile component present the slice traffic to

the transport network. For slicing phase 1 and 2 these L3VPNs are mapped to the underlay's default routing algorithm. In slicing phase 3, flex-algo or Traffic Engineering can be used to map L3VPN to different underlay transport planes.

Backhaul L3VPNs (phase 1) or backhaul/midhaul L3VPNs (phase 2 and 3) associated with a slice needs to be present on all TNEs and DC-TNEs that either connect directly or indirectly (through the DC infrastructure) to mobile components supporting that slice. Access to the slice's L3VPN, on the TNEs, is via an VLAN attachment circuit.

Backhaul Transport Slicing

A full description of MP-BGP based L3VPNs is contained in sections 13.3 and *Annex C: MP-BGP based L3VPNs*. MP-BGP based L3VPNs support flexible connectivity models and are highly scalable in terms of sites, VPNs and the number of VPN IP routes so able to deal with transport connectivity requirements associated with WG-1s slicing phases and the anticipated size of 5G environments. MP-BGP L3VPN's have two basic mechanisms for controlling connectivity within a L3VPN.

- Virtual Routing and Forwarding (VRF) instances: These are local routing table associated with L3VPNs on Provider Edge (PE) TNEs. The VRF holds local and remote VPN routing information. Local routes are those learnt by the PE and normally consist of attachment circuits, routes learnt via routing protocols running over attachment circuits or via local static routing. Remote VPN routes are those learnt from remote PEs. They are conveyed between PEs using MP-BGP and held in the PE's BGP table. Installation of remote routes into the VRF routing table is determined by local policies. One of the key policy attributes used to determine what routes are installed into a VRF are Route Targets (RTs).
- Route Targets (RTs): RTs are a BGP attribute attached to VPN routes when they are distributed by MP-BGP from the PE where they were learnt. At the destination PE, filtering of routes based on RTs is one of the key mechanisms for determining what remote VPN routes are imported into a VRF.

Slicing Phase 1

Slicing phase 1 supports three general purpose 5G backhaul service slices which are network wide. For these simple slice use cases the design shown in Figure 25-13 could be used.

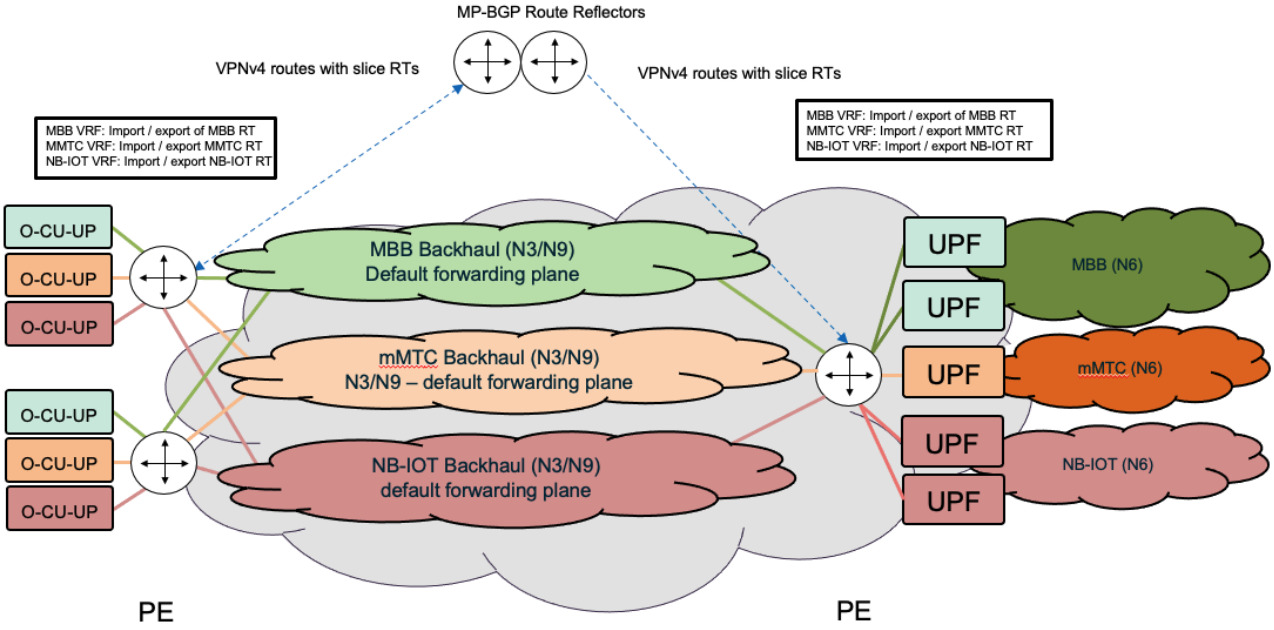


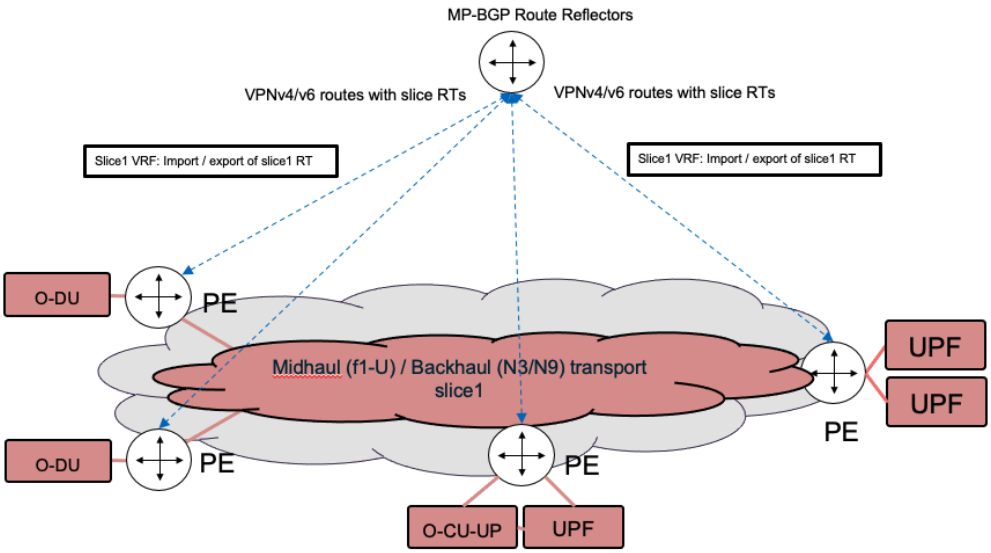
Figure 25-13: Backhaul phase 1 slice design

- Three discrete VRFs associated with the phase 1 slices are created on PEs supporting slices. Alphabetic VRF names have been used in the illustration. In reality, VRF names could also be numeric values.
  - VRF1: MBB
  - VRF2: MMTC
  - VRF3: NB-IOT
- Mobile components associated with each slice connects to the PE via a VLAN attachment circuit which is placed into the appropriate slice VRF.
- Each VRF on the PE learns local IP subnets containing the mobile slice components, either as connected routes, via a routing protocol or via static routes.
- A single RT is associated with each slice type. Alphabetic RT names have been used in the illustration. In reality, RT values are numeric values.
  - VRF1: RT = MBB
  - VRF2: RT = MMTC
  - VRF3: RT = NB-IOT
- Routes from each slice's VRF are exported where they are learnt with their associated RT.
- On remote PEs each VRF imports the RT associated with their slice.
- This creates an any-to-any transport network allowing connectivity between all mobile backhaul components in the slice.

NOTE: In many cases, Service Providers already use L3VPNs based on RFC4364 [84] and have their own methodologies for creating L3VPNs. In these instances, Service Providers could use their own VPN design methodology to create an any-to-any VPN for Backhaul slices.
NOTE: This is a very simple transport slice architecture and suitable for general-purpose slices. More complex slice use cases may need different L3VPN topologies. The combination of VRFs, RTs and BGP communities and import and export filtering provides a powerful toolset to build these connectivity models.

Slicing phase 2 and 3

The transport technology and mechanisms outlined above for slicing phase 1 are largely applicable in slicing phase 2 and 3.



-

Figure 25-14: Backhaul phase 2 slice design

The main differences are:

- A common midhaul/ backhaul MP-BGP L3VPN which extended to all TNEs and DC-TNEs containing midhaul and backhaul mobile components.
- Distributed UPFs also means the data network (N6) L3VPN may need to extend to distributed data centers.
- Up to 1000 slices, in addition to the three supported in slicing phase 1.

Data Network (N6)

The data network or N6 is not part of the RAN and is not in scope of the WG-1 slicing work item. However, transport slicing needs to extend to the N6 data network, particularly as the Data Network (N6) network extends closer to the radio infrastructure with multiple breakout points and although not discussed in this annex, the techniques outlined for the backhaul and midhaul user plane slicing are largely applicable when building the data networks across a common packet switched transport infrastructure. An important consideration when utilising multiple breakout points in the data network (N6) is to ensure traffic exits the network in the desired place but also returning traffic enters the data network (N6) in desired place. This requires UPF address block planning within the data network, how these routes are advertised externally by the routers at the peering points and how external routes are advertised within the Data network (N6).

Transport Network Quality of Service architecture

Quality of Service (QoS) in the transport network in O-RAN phase 1, 2 and 3 is based on the Diffserv architecture defined in RFC2475 [46] and further outlined in section 14 and *Annex D:*

Quality of Service of this document. The QoS architecture is widely implemented by Service Providers but there are significant differences between Service Providers in their QoS implementations. These include DSCP bit usage, edge policies, marking strategies, numbers of queues on the edge and within the core and how different Per Hop Behaviours (PHBs) are implemented. As such, this section aims to outline some of the key considerations associated with 5G traffic and provide example QoS implementations. Service Providers can use this information to adjust their existing QoS architectures to support 5G.

Transport QoS considerations in a 5G environment

Section 7 outlines the 5G interfaces the transport network needs to support. In many areas the transport QoS requirements for 5G are very similar to existing 4G and general multi-service environments. However, there are some notable considerations and exceptions.

- In the context of packet-based switching technologies for O-RAN and 3GPP, the interplay of QoS and transport slicing has two dimensions: 1) The QoS requirements of O-RAN fronthaul traffic 2) The QoS requirements of the mobile services, which depends on the kind of application being delivered to the end users.
  The transport slicing and QoS architectures needs to consider both dimensions to properly allocate QoS resources capable of satisfying the expected requirements for each kind of traffic in terms of bandwidth, latency, etc.
- O-RAN 7.2X Fronthaul Traffic: The user and control traffic stream is eCPRI based and runs between the O-RU and the O-DU. It is a new traffic type associated with 5G. The transport requirements of the Fronthaul user and control plane interfaces are contained in the O-RAN.WG9.CTRP-REQ-V01.00 [19] but can be summarised as high bandwidth and more importantly from a QoS perspective, extremely low delay (micro-seconds) traffic. Due to the low delay requirements this traffic type will only ever be seen in the access network and requires packet switching equipment to treat this traffic with the highest priority through the device and when scheduling onto attached interfaces. Today, many Service Providers use the highest priority queue for voice traffic, consequently, changes to the existing QoS schemes may be required.
- 5G introduces Midhaul: Midhaul interfaces run between the O-DUs and the O-CUs and between O-CUs. The F1 interfaces are a new set of interfaces associated with 5G, but use GTP-U and GTP-C which is the same packet types used in Backhaul. It is expected that Midhaul and Backhaul interfaces will have similar QoS requirements.
- 5G introduces new service types, each with their own QoS requirements. GTPv1-U [5] is used in the Midhaul and Backhaul for F1-U, Xn-U, N3 and N9 interfaces. The importance of user data packet within the mobile environment is indicated in the QoS Field Identifier (QFI)**Error! Reference source not found.** in the GTP-U extension header[5]. High performance TNEs do not have the capability to examine the QFI field and instead use the DSCP field in the IP header or the MPLS Traffic Class to understand the QoS requirements of packets in the transport network. For the transport network to be able to treat GTP-U packets in an appropriate fashion for the mobile service, the Midhaul (starting from O-RAN slicing phase 2) and Backhaul (starting from O-RAN slicing phase 1) mobile components (O-DUs, O-CU-UP and UPFs) must have mechanisms to set a DSCP value in the GTP IP header appropriate to the QoS requirements of the mobile service carried in the GTP-U packet. This could be achieved using a mapping mechanism on the Midhaul and Backhaul mobile components that map QFI values to DSCP values.

NOTE: In O-RAN slicing phase 1, QFI to DSCP mapping facilities is a capability seen on Backhaul mobile component's (O-CU-UP (on backhaul side), UPF) user plane interfaces. In phase 2 and 3,

QFI to DSCP mapping capability and associated management and control is enhanced to Midhaul mobile components (O-DU, O-CU-UP) user plane interfaces.

3GPP QoS flows and Transport QoS

The concept of QoS in 4G LTE is based on bearers, while the concept of QoS in 5G mobile systems is based on the QoS Flow concept [1] which provides finer granularity. Flows in 5G and 4G bearers can be divided into guaranteed bit rate flows (GBR QoS flows) and non-guaranteed bit rate flows (non-GBR QoS flows). 5G also has a new delay critical GBR flow category which is not present in LTE. QoS flows also have a service expectation which is expressed in terms of packet delay budget and packet error rate. The mobile QoS requirements are designated through the QoS Class Identifier (QCI) in LTE [2] and the 5G QoS Identifier (5QI) in 5G [2]. These identifiers can be operator defined but there is also a standardized list contained in [1]. Within the mobile domain traffic (of flows in 5G or bearers in LTE) with a given QCI or 5QI should receive the same forwarding treatment (e.g, in terms of scheduling, admission threshold, etc)

NOTE: The majority of the QCI and 5QI are the same, with very few exceptions of identifiers being defined in one case but not in the other (e.g., QCI 75 or 5QI 10).

TNEs in the transport network use DSCP or MPLS traffic class markings in the IP/MPLS header to determine QoS packet handling. To keep consistency of QoS treatment between the overlay mobile service and the underlay packet-based switching networks it is interesting to look at the different standard QCI and 5QI in qualitative terms to understand how distinct kind of flows (or bearers) at the mobile overlay level could be grouped together and treated on the underlay packet transport network. This equates to how different QCIs and QFIs could be mapped to DSCP markings on the 5G mobile components (RUs, DUs, CUs) and how these DSCPs are mapped to Diffserv Per Hop Behaviour (PHB) in the transport network.
Figure 25-15 plots the standardized QCIs and 5QIs in relation to the defined packet loss and delay characteristics (Note, the figure is not to scale). Each box in the plot firstly contains the 5QI or QCI (top part) complemented with the values for delay / packet loss (at the bottom). To distinguish profiles, that are GBR, non-GBR or delay-critical GBR, a color code is used which is indicated in the legend of the figure (note that delay-critical GBR 5QI in 5G are simply GBR QCI in LTE).

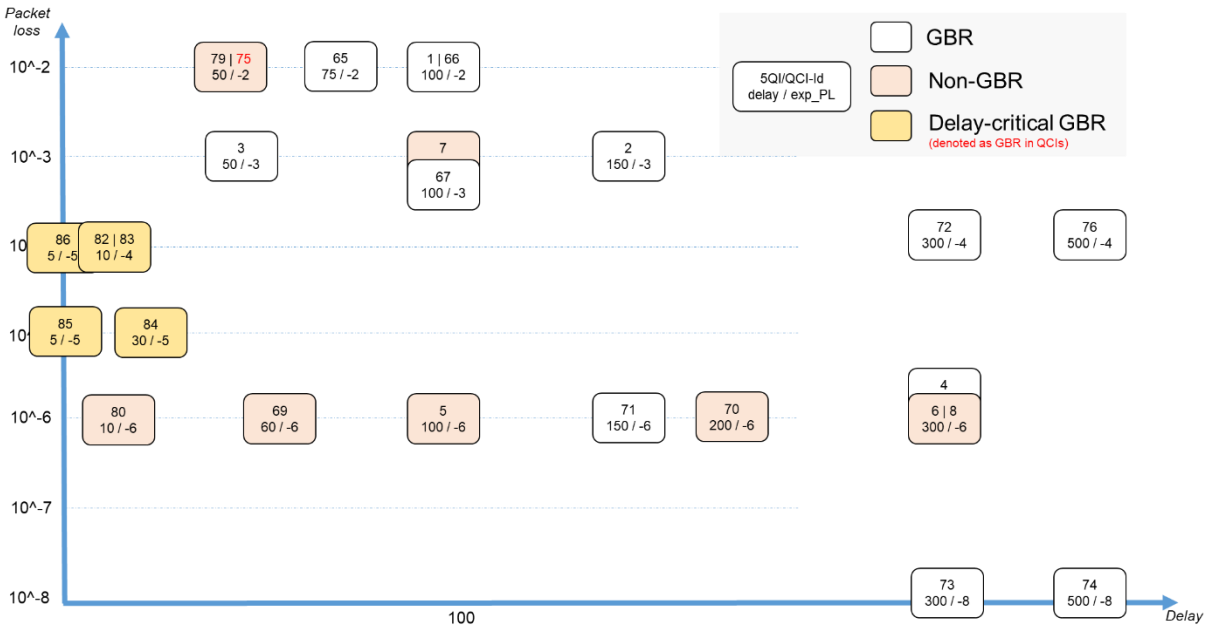


Figure 25-15: 5QI and QCI plot in relation to packet loss and delay

The plot identifies 5QI or QCI values with common QoS flow characteristics. It is now possible to group 5QI and QCI which require similar forwarding treatment in the transport network. Figure 25-16 presents a set of exemplary groupings, in which the 5QI/QCI are categorized into four distinct groups: group 1 consists of QoS flows requiring low delay, e.g., < 30 ms; group 2 consists QoS flows which are in the delay range between 50 and 75 ms; group 3 consists of the rest of GBR 5QI/QCI; while finally group 4, as default group, aggregates the rest 5QI/QCI of non-GBR profile.

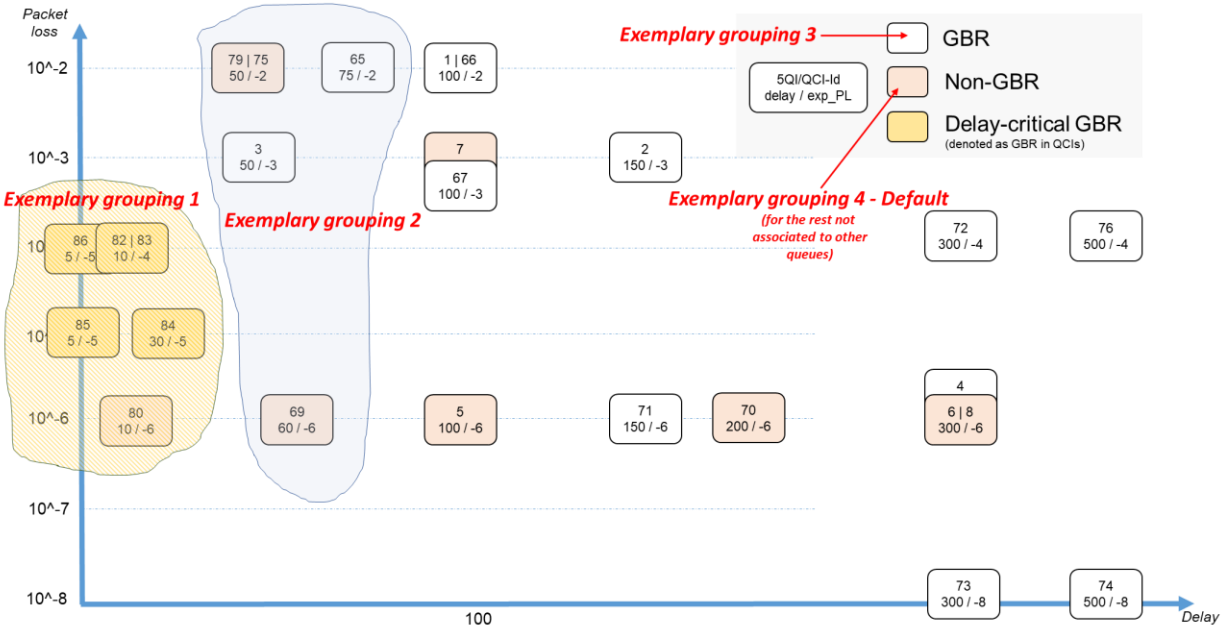


Figure 25-16: Exemplary 5QI and QCI grouping

Using the above example, a transport QoS architecture is illustrated in the next section that supports a set of transport core PHBs capable of supporting existing services, new 5G services, 5G fronthaul and protecting the infrastructure needed to deliver timing, OAM and the transport routing protocols.

This annex makes no recommendations on QFI to DSCP mapping that needs to occur on the Midhaul (starting from O-RAN slicing phase 2) and Backhaul (starting from O-RAN slicing phase 1) mobile components, as there are huge variations between operators on their usage of DSCP markings. Viable approaches include:

- A dedicated DSCP marking for some 5QI / QFI value.
- Group 5QI and QFI values, for example as shown above, and allocate all 5QIs and QFIs in the group the same DSCP value.
- A combination of the two.

Regardless of the edge marking scheme chosen, the most important action occurs on the ingress TNE which needs to map the edge DSCP marking to the appropriate transport core marking (MPLS TC for MPLS underlays and DSCP for SRv6 underlays) to ensure the appropriate PHB in the core of the transport network.

Transport QoS Architecture for O-RAN slicing phase 1, 2 and 3

To ensure that Service Level Objectives (SLOs) requested for a particular slice are maintained, appropriate QoS scheme must be deployed in the transport network.

The details of QoS functionality are provided in *Annex D: Quality of Service*. This section provides guidelines about particular use of the QoS concepts discussed in that annex to deploy the QoS architecture suitable for providing SLO guarantees for transport network slices.

As outlined in Figure 25-17, the QoS model for transport network slicing has following major building blocks:

- Ingress (input) policing (rate limiting) at the edge of the transport domain
- Mapping function at the edge of the transport domain to map potentially large number of edge traffic classes, as described earlier, into finite (typically 4 to 8) traffic classes used in the transport core
- Flat egress (output) scheduling (queuing) on the transport network transit links in the transport core with typically 4 to 8 traffic classes
- Egress (output) scheduling (queuing) at the edge – optionally hierarchical

NOTE: Figure 25-17 is an example figure and for simplicity shows the traffic in one direction only (from UPF on the left-hand side to O-CU-UP on the right-hand side). However, traffic is typically bidirectional (for example UPF → O-CU-UP and O-CU-UP → UPF), therefore the building blocks mentioned earlier are deployed in opposite direction, too. Thus, all transport transit interfaces (transport core) should have flat egress (output) scheduling (queuing) deployed, whereas all transport edge interfaces would have input policing and potentially hierarchical output scheduling (queuing) deployed. Also, the figure shows only one type of interface (backhaul), in accordance with O-RAN slicing phase 1 requirements. Similar QoS concepts apply to other sliced interfaces, for example midhaul, starting from O-RAN slicing phase 2.

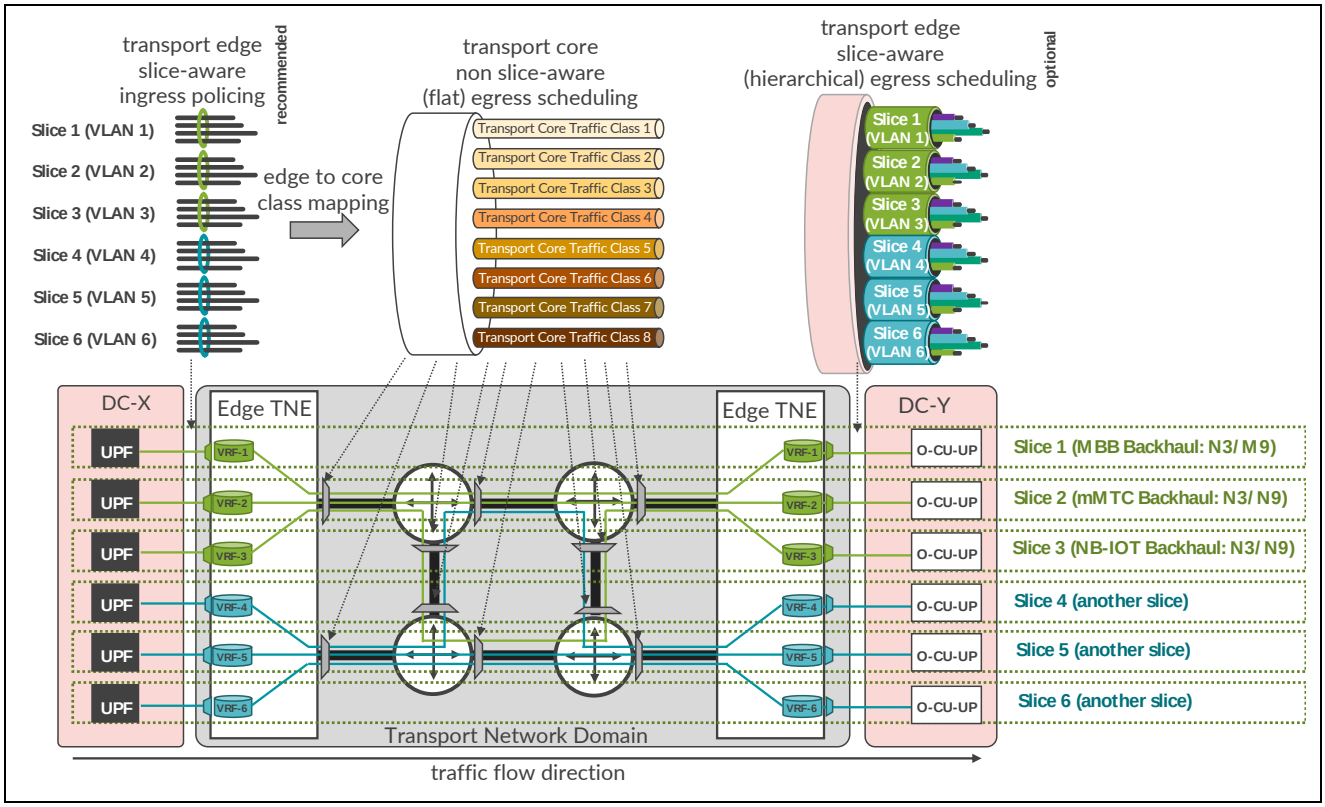


Figure 25-17: High-level QoS model for transport slices

Subsequent subsections provide detailed discussion for each element in the QoS architecture for transport network slicing.

Edge QoS

When the slice is created, mechanisms required to implement the slice in the transport network are deployed on the transport network elements (TNEs). An important building block is edge QoS to condition slice traffic at the transport edge and ensure that slice traffic does not exceed the previously requested and agreed bandwidth values.

Ingress policing

Ingress policing (rate limiting) is used to police (rate limit) the traffic of a particular slice when the traffic enters the transport network domain. This slice admission control function is required to ensure, that only traffic contracted by given slice at the time of slice creation (or at the time of last slice modification) is accepted with SLO guarantees.

NOTE: In the deployments where there is no expectation from the transport network to provide SLO guarantees (delay, jitter, bandwidth, no traffic interference between different slices, etc.) ingress policing might be omitted.

Depending on the actual slice QoS profile requested at the slice creation/modification, out-of-contract traffic is either completely dropped at the transport network edge (if requested slice guaranteed rate and maximum rate are equal, or maximum guaranteed rate is missing in the request), or, if slice guaranteed rate is not equal to slice maximum rate, slice traffic is handled at the edge in following way:

- Slice traffic under guaranteed rate: unconditionally accepted at the transport edge

- Slice traffic above guaranteed rate and below maximum rate: conditionally accepted at the transport edge. Such traffic is marked at the transport edge with QoS marking (DSCP or MPLS Traffic Class bits) representing high drop probability. Using Class based Weighted Random Early Discard (CBWRED) functionality, described in *Annex D: Quality of Service* , and implemented on transit transport network elements, when the transit links are nearing the congestion (transmit queues on transit links are building up) CBWRED kicks in, and starts to drop packets marked with QoS marking representing high drop probability. Thus, when no congestion is observed in the transport network, slice traffic can go up to the contracted value (no drop at the transport edge for traffic volumes below slice maximum rate), but when congestion is close, high drop probability packets (above guaranteed rate but below maximum rate) are being dropped at transit. This ingress marking behaviour at the transport edge can be implemented with 2-rate 3-color policer.

- Slice traffic above maximum rate (or above guaranteed rate, if maximum rate is not specified during slice creation/modification): unconditionally dropped at the transport edge. This behaviour can be implemented with 1-rate policer (when slice guaranteed rate equals to maximum rate), or 2-rate 3-color policer (when different guaranteed and maximum rates are specified for the slice)

There could be many reasons for misbehaving (send traffic above the contracted rates) slices, like for example:

- Underestimate for slice traffic volume during slice creation (or slice modification) events
- Some failure within the slice, causing extra traffic generation
- Denial of Service (DoS) attack started from within the slice

With slice admission control, implemented via ingress policing (rate limiting), it can be ensured that only legitimate in-contract slice traffic is allowed at the transport edge, thus slices that are sending traffic above the agreed rate do not negatively impact other slices. This provides traffic isolation guarantees between the different slices.

NOTE: In O-RAN slicing phase 1, only per slice SLO parameters might be specified. There is no granular per class (per 5QI) SLO parametrization within each slice in phase 1. Therefore, in phase 1 of network slicing single policer (1-rate or 2-rate-3-color) per slice might be sufficient for slice admission control implementation.

O-RAN slicing phase 2 introduces slices with multiple traffic classes (multiple 5QIs) within a single slice, like for example MBB with IPTV. Therefore, starting from slicing phase 2, more individual policers, or even hierarchical policers, might be introduced to implement slice admission control.

Ingress Edge to core traffic class mapping

As described earlier multiple 5QIs, resulting in multiple DSCP markings, are potentially being used at the edge of the transport domain in the deployments.

Transport core typically uses flat QoS model with up to 8 traffic classes. Therefore, mapping function to map multiple transport edge classes (represented by multiple 5QI/DSCP values) to limited set of transport core classes is implemented on the edge transport network element as outlined in Figure 25-18.

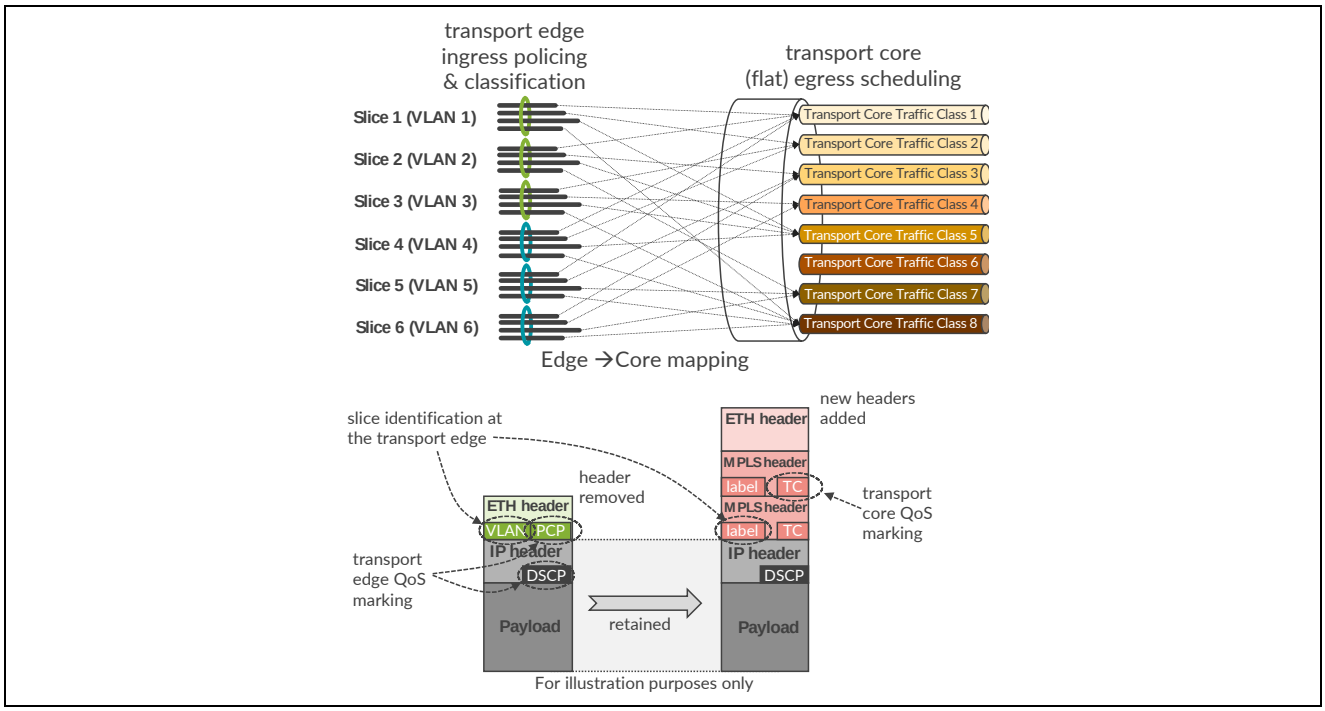


Figure 25-18: Transport edge to transport core QoS mapping

In case of MPLS encapsulation in transport core, the result of this transport edge to transport core class mapping is setting the Traffic Class (TC) field in outer MPLS label. If SRv6 is used as the encapsulation mechanism, DSCP field in the outer IPv6 header carries the transport core class marking, while DSCP field in the inner IPv4/IPv6 header, similarly to MPLS encapsulation, carried transport edge class marking.

Transport network elements (TNEs) associate transport core class markings, carried either in outer MPLS header TC field, or in outer IPv6 header DSCP field, with QoS output queues and execute appropriate scheduling mechanism to implement desired traffic prioritization during link congestion events.

Egress (hierarchical) scheduling

On the egress side of transport edge, egress scheduling is implemented to provide desired traffic prioritization among the slices, as well as among traffic classes within each slice. Depending on the hardware capabilities of the edge transport network element, as well as the intended granularity (more “soft” or more “hard”) of the enforcement, one of the following option can be used:

- flat, non-hierarchical QoS model with up to 8 queues (similar to transport core QoS model)
- hierarchical, slice-aware QoS model with separate set of queues per each slice

As described in the *Annex D: Quality of Service* , the hierarchical QoS scheduler model allows for a “hard” distribution of the bandwidth on an interface amongst the slices sharing that interface capacity, protecting the use of the bandwidth for a given slice from the other slices on the same interface.

Core QoS

Transport core QoS is based on a flat model with 4 to 8 queues. Transport core QoS settings are rather static in the nature, provisioned at the time the transport network infrastructure (interface settings, IGP settings, BGP settings, etc.) is provisioned. Especially, there is no direct correlation between transport core QoS and 5G slices being created/modified/deleted. Transport core QoS doesn't maintain any state related to individual 5G slice, neither adjust its QoS parameters based on 5G slices creation/modification/deletion events.

The goal of transport core QoS is to ensure right prioritization of transport core traffic classes during congestion events caused by network failures and subsequent traffic rerouting, so that during such congestion events traffic classes requiring low latency are using priority forwarding, and traffic classes requiring low drop are not dropped. The goal of transport core QoS is not end-to-end bandwidth management or providing granular guarantees per each 5G slice.

Table 9 contains a list of typical flows that can be observed in the multiclass transport network used to transport 5G flows as well, taking into considerations exemplary 5QI/QCI grouping described earlier in this annex. When recommending appropriate transport core QoS model, overall QoS policies for all flows must be taken into consideration.

Traffic type	Packet size (order of magnitude)	Per-hop latency (order of magnitude)[1)]	Per-hop PDV (order of magnitude)[1)]
PTP (unaware mode)[2)]	~100 bytes	constant average	~0.5 µs[3)]
CPRI (RoE)	~1500 bytes	~1-20 µs	~1-20 µs
eCPRI CU-plane	~1500 bytes	~1-20 µs	~1-20 µs
OAM with aggressive timers	~100 bytes	~1 ms	~1 ms
5QI/QCI Group 1 (low latency U-plane)	variable	~1 ms	~1 ms
Low latency business traffic	variable	~1 ms	~1 ms
Network Control: OAM with relaxed timers, IGP, BGP, LDP, RSVP, PTP aware mode (T-TC/T-BC)[4)]	variable	~5 ms	~1-3 ms
O-RAN/3GPP C-plane and M-plane[5)]	variable	~5 ms	~1-3 ms
5QI/QCI Group 2 (medium latency U-plane)	variable	~5 ms	~1-3 ms
5QI/QCI Group 3 (remaining GBR U-plane)	variable	~10 ms	~5 ms
Guaranteed business traffic	variable	~10 ms	~5 ms
5QI/QCI Group 4 (remaining non-GBR U-plane) Other traffic types (best effort)	variable	~10-50 ms	~5-25 ms

Table 9: Different flows per-hop latency/PDV (order of magnitude)

Note 1: Per-hop latency includes all latency contributors, like frame transmission delay, self-queueing delay, queuing delay, store-and-forward delay, as described in section 10.1.1.2.1. Exact per-hop requirements depend on the overall network budget, number of hops, budget allocated to fibers, etc. The table intends to emphasize only relative order of magnitude for per-hop latency/PDV to illustrate the process of assigning traffic to QoS queues.

Note 2: PTP unaware mode (i.e., router transiting G.8275.2 PTP stream, without T-TC function on the router that could add timestamps to transit PTP packets) → strict-priority queue is required to

minimize jitter (actual latency value is not relevant, but its average should be constant). Minimizing the latency via strict-priority queue minimizes jitter as well.

Note 3: Max|TE| accumulated across the network must be ≤1.1 μs.

Note 4: T-BC/T-TC (both G.8275.1 and G.8275.2) with physical layer time stamping → guaranteed bandwidth queue is OK, strict-priority queue is not required, since jitter/PDV will be accounted by physical layer timestamps in PTP packet. Also, latency value is not relevant, but average latency should be constant. QoS should ensure that PTP packets are not dropped during congestion, and guaranteed bandwidth queue is sufficient for that.

Note 5: O-RAN/3GPP C-plane and M-plane kept in separate queue to ensure minimum bandwidth guarantees during congestion events preventing failures of these planes, as well as to increase security and separation of these planes.

There are variety of QoS models, depending on the hardware support available on transport network element platform. It is out of scope for this document to discuss all the various QoS models supported by different hardware platforms of transport network elements. However, for illustration purposes, two major, most common QoS models are worth to mention: with single priority queue, and with multiple priority queues.

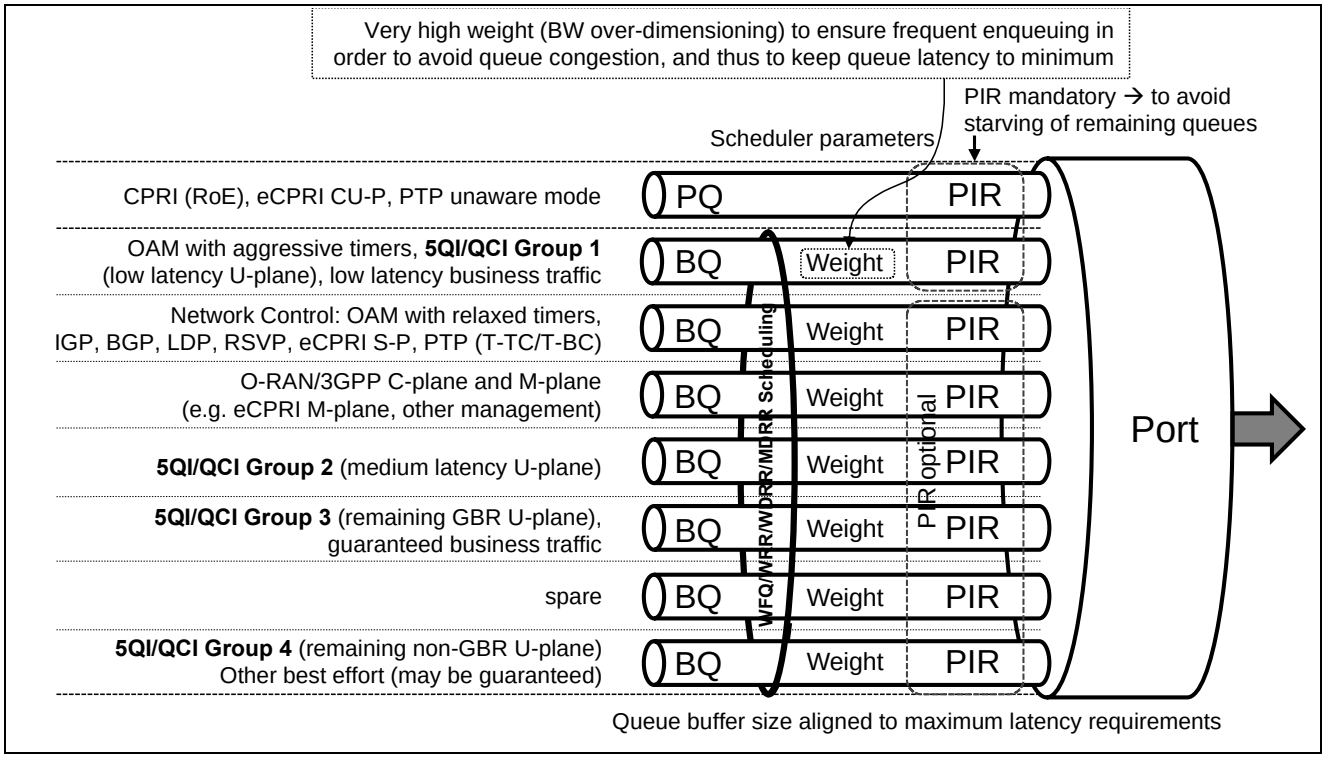


Figure 25-19: Example of QoS model with single priority queue

Figure 25-19 outlines an example QoS model with a single priority queue. In this hardware model, all flows with ultra-low latency/PDV sensitivity (PTP unaware mode, CPRI/RoE, eCPRI CU-plane) must be placed in this priority queue, while other flows should be distributed among remaining bandwidth queues. Bandwidth queue used for flows with low (but not ultra-low) latency/PDV sensitivity (OAM with aggressive timers, latency sensitive U-plane and business traffic) should be parametrized with relatively high weight used in WFQ/WRR/WDRR/MDRR (Weighted Fair

Queueing, Weighted Round Robin, Weighted Deficit Round Robin. Modified Deficit Round Robin) scheduling algorithms, so that this queue is serviced very frequently, to avoid queue congestion and to minimize latency/PDV. With high weight parameterization this queue behaves almost as priority queue.

To avoid starvation of the remaining queues during congestion events, priority queue, as well as highly over dimensioned bandwidth queue must have a policer or shaper to limit the rate (PIR – Peak Information Rate) of the queue. Queue buffer sizes in both cases must be aligned to maximum latency requirements of the traffic assigned to the queue.

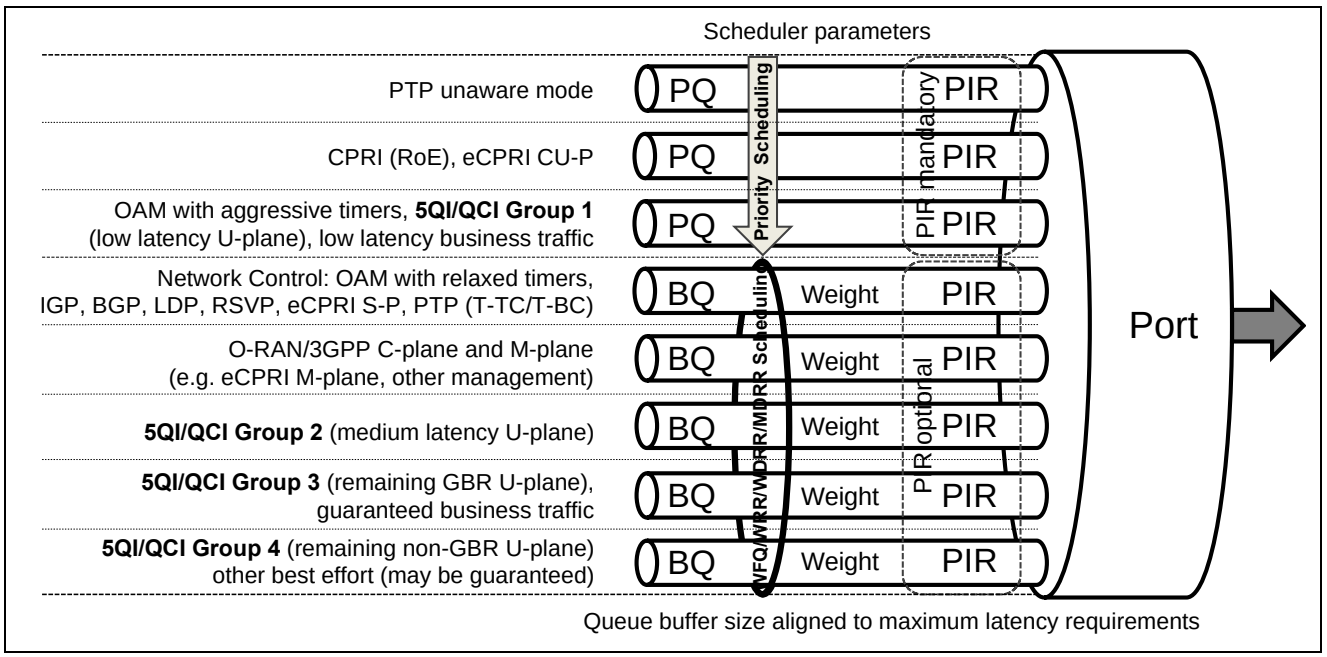


Figure 25-20: Example of QoS model with multiple priority queues

Figure 25-20 outlines an example of queue assignment on hardware platforms supporting multiple priority queues, dequeued in strict priority order. With multiple priority queues available it is recommended to place PTP packets in unaware mode in the highest priority queue, to minimize the PDV of these packets to the highest possible degree. Putting these packets above CPRI(RoE) or eCPRI has only minimal influence on CPRI/eCPRI packets PDV, since PTP packets are very small (~100 bytes). For example, serialization delay of such small packet on 10 GE interface is only 80 ns, so PDV factor contributing to CPRI/eCPRI PDV is very small as well and can be easily handled by the CPRI/eCPRI reassembly functions.

