Workgroup: TEAS Working Group
Internet-Draft: draft-srld-teas-5g-slicing-01
Published: 24 October 2022
Intended Status: Informational
Expires: 27 April 2023
Authors: K. Szarkowicz, Ed. R. Roberts
Juniper Networks Juniper Networks
J. Lucek J. Drake M. Boucadair
Juniper Networks Juniper Networks Orange
LM. Contreras I. Bykov R. Rokui
Telefonica Ribbon Communications Ciena
L. Jalil B. Setyawan
Verizon XL Axiata

A Realization of IETF Network Slices for 5G Networks Using Current IP/MPLS Technologies

Abstract

5G slicing is a new feature that was introduced by the 3rd Generation Partnership Project (3GPP) in mobile networks. It covers slicing requirements for all mobile domains, including the Radio Access Network (RAN), Core Network (CN), and Transport Network (TN).

This document describes a basic IETF Network Slice realization model in IP/MPLS networks with a focus on fulfilling 5G slicing connectivity requirements. This IETF Network Slice realization model reuses many building blocks currently commonly used in service provider networks.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on 27 April 2023.

Copyright Notice

Table of Contents

1. Introduction
  1.1. Requirements Language
2. 5G Network Slicing Integration in Transport Networks
  2.1. 5G Network Slicing versus Transport Network Slicing
  2.2. NF to NF Datapath vs Transport Network
    2.2.1. Segmentation of NF-to-NF Datapath
    2.2.2. Orchestration of Local Segment termination at ETN
  2.3. 5G slice to IETF Network Slice mapping
  2.4. First 5G slice versus subsequent slices
3. High-Level Overview of the Realization Model
  3.1. VLAN Hand-off
  3.2. IP Hand-off
  3.3. MPLS Label Hand-off
    3.3.1. Option B
4. QoS Mapping Models
  4.1. 5QI-unaware Mode
    4.1.1. Inbound Edge Resource Control
    4.1.2. Outbound Edge Resource Control
  4.2. 5QI-aware Mode
    4.2.1. Inbound Edge Resource Control
    4.2.2. Outbound Edge Resource Control
  4.3. Transit Resource Control
5. Transport Planes Mapping Models
  5.1. 5QI-unaware Mode
  5.2. 5QI-aware Mode
6. Capacity Planning/Management
  6.1. Bandwidth models
    6.1.1. Scheme 1: Shortest Path Forwarding
    6.1.2. Scheme 2: TE LSPs with Fixed Bandwidth Reservations
    6.1.3. Scheme 3: TE LSPs without Bandwidth Reservations
7. IANA Considerations
8. Security Considerations

9. References
9.1. Normative References
9.2. Informative References
Appendix A. Acronyms and Abbreviations
Appendix B. An overview of 5G Networking
B.1. Building Blocks
B.2. Core Network
B.3. RAN
B.4. Transport Network
Acknowledgements
Contributors
Authors' Addresses

1. Introduction

[I-D.ietf-teas-ietf-network-slices] introduces the framework for network slicing in the context of networks built using IETF technologies. The IETF network slicing framework introduces the concept of a Network Resource Partition (NRP), which is simply a collection of resources identified in the underlay network. There could be multiple realizations of high-level IETF Network Slice and NRP concepts, where each realization might be optimized for the different network slicing use cases that are listed in [I-D.ietf-teas-ietf-network-slices].

This document describes a basic - using only single NRP - IETF Network Slice realization model in IP/MPLS networks, with a focus on fulfilling 5G slicing connectivity requirements. This IETF Network Slice realization model reuses many building blocks currently commonly used in communication service provider (CSP) networks.

The reader may refer to [I-D.ietf-teas-ns-ip-mpls] for more advanced realization models.

Also, the reader may refer to [RFC6459] and [TS-23.501] for more details about 3GPP network architectures.

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

2. 5G Network Slicing Integration in Transport Networks

2.1. 5G Network Slicing versus Transport Network Slicing

Network Slicing has a different meaning in the mobile and transport worlds. Hence, for the sake of precision, this section provides a brief description of the objectives of 5G Network Slicing and Transport Network Slicing.

*The objective of 5G network slicing is to provide dedicated resources of the whole 5G infrastructure to certain users, application, customers or PLMN (e.g., RAN sharing). These resources are from the Transport Network, RAN and CORE Network Functions and the underlying infrastructure. [TS-28.530] defines 5G network slicing by introducing the concept of Network Slice Subnet (NSS) to represent slices within each of these domains: RAN, CORE and Transport Network (i.e., RAN NSS, CN NSS and TN NSS). As per 3GPP specifications, NSS can be shared or dedicated to a single slice.

*The objective of Transport Network slicing is to isolate, guarantee or prioritize Transport Network resources for slices such as buffers, link bandwidth or even RIB/FIB. Transport Network Slicing has two main flavors: Hard and Soft slicing. Hard slicing provides dedicated network capacity to slices. Soft Slicing provides shared network capacity with guarantees for each slice. There are different options to implement TN slices based on tooling such as VRFs for traffic separation, QoS and TE. Also, TN slice realization for 5G slices might combine elements of hard slicing in one part of the transport network, with elements of soft slicing in other parts of the transport network. An optimized 5G network slicing architecture should integrate Transport Network Slicing, however, it is possible to implement 5G network slicing without Transport Network Slicing, as explained in the next section.

Slicing in the transport network is implemented using IETF technologies, therefore, inline with [I-D.ietf-teas-ietf-network-slices], in this document the term "IETF Network Slice" (IETF NS, or INS in short) is used to describe the slice in the Transport Network domain of overall 5G architecture, composed from RAN, TN and CN domains.

2.2. NF to NF Datapath vs Transport Network

The 3GPP specifications loosely define the Transport Network and its integration in RAN and Core Network domains: the role of the Transport Network is to interconnect Network Functions. In other words, it is the end-to-end datapath between two Network Functions.

In practice, this end-to-end datapath is often a non-uniform architecture made up of several segments potentially managed by different organizations. In this document, we rather define the Transport Network with a service provider scope: the TN extends up to the PE or the CE if it is also managed by the TN Orchestration. Additionally, we assume that the Transport Network is MPLS or SRv6 capable.

2.2.1. Segmentation of NF-to-NF Datapath

This section introduces a decomposition of the datapath between two Network Functions (NFs) into two segments based on the Orchestration domains: TN segments and Local segments.

- TN Segment: the realization of this segment is driven by the IETF NSC / Transport Network Orchestrator (TNO). Generally speaking, a TN Segment provides connectivity between two sites.

- Local segment: this segments permits either to connect Network Functions within a given site or to connect a Network Function to the Transport Network. The realization of this segment is directly or indirectly driven by the 5G Orchestration without any involvement of the Transport Network Orchestration. Generally speaking, the Local Segment is a datapath local to a site. This site can be either DC, POP, CO or a virtualized infrastructure in a Public Cloud.

Note that more complex scenarios are possible, for example with extra segmentation of TN or Local Segments. Additionally, sites can be of different types such as Edge, Data Center, or Public Cloud, each with specific network design, hardware dependencies, management interface and diverse networking technologies (e.g MPLS, SRv6, VXLAN, L2VPN vs L3VPN, ...). The objective of this section is to clarify the scope of the Transport Network rather than to cover any technology or design combination.

The realization of IETF Network Slices (i.e. connectivity with performance commitments) applies therefore to the TN Segments. We consider Local Segments as an extension of the connectivity of the RAN/CORE domain without slice-specific performances requirements by assuming that the local infrastructure is overprovisioned and implements traditional QoS/Scheduling logic.

In parallel, since the TN domain can extend either to the CE or to the PE, we introduce the term Edge Transport Node (ETN) to define this boundary. The ETN is therefore a Transport Node that stitches Local segments and TN Segments. Note that depending on the design, the placement of the SDP as defined in [I-D.ietf-teas-ietf-network-slices] may or may not be enforced on

the ETN itself. The following figure is a representation of the end-to-end datapath between Network Functions including Segments and ETN (in practice PE or a managed CE), where applicable.

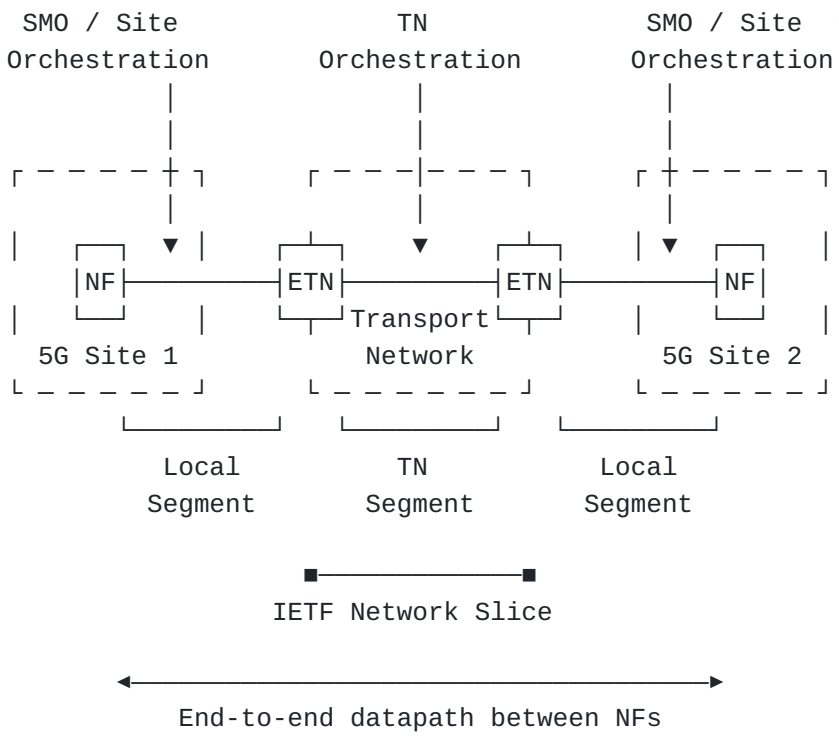


Figure 1: Segmentation of the NF-NF datapath

NFs may also be placed in the same site and interconnected via a Local Segment. In this case, there is no TN segment (i.e. no Transport Network Node is present in the datapath).

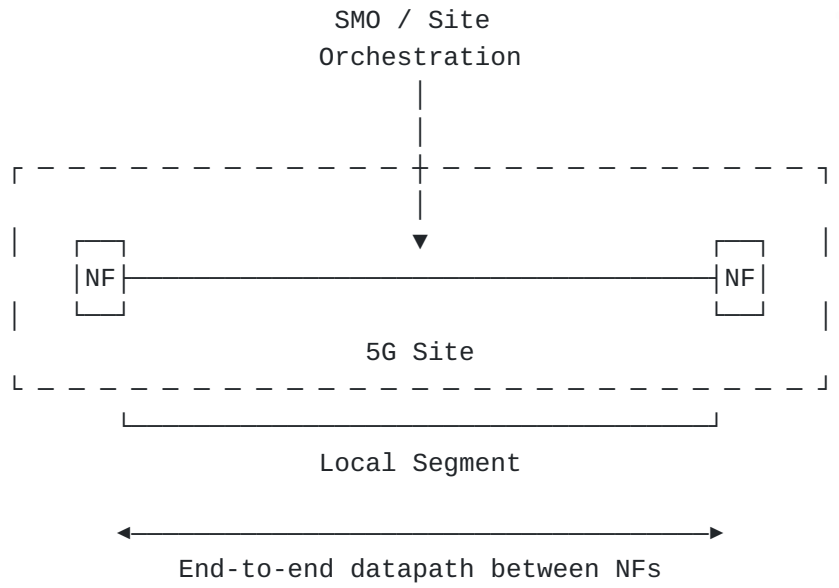


Figure 2: NF-NF datapath within single site

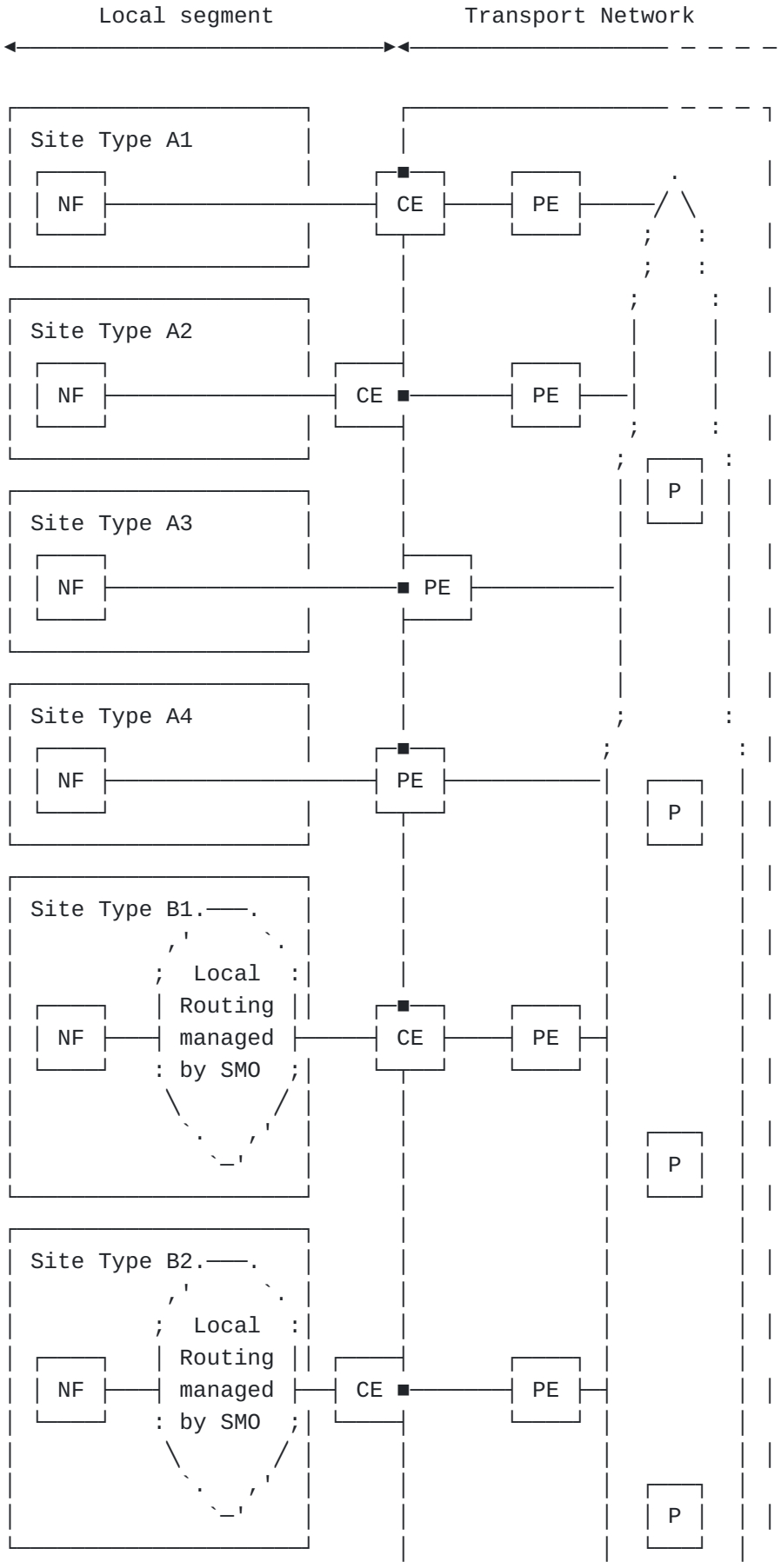
Next, the following picture provides examples of different realizations of Local and TN segments, as well the Service Demarcation Points.

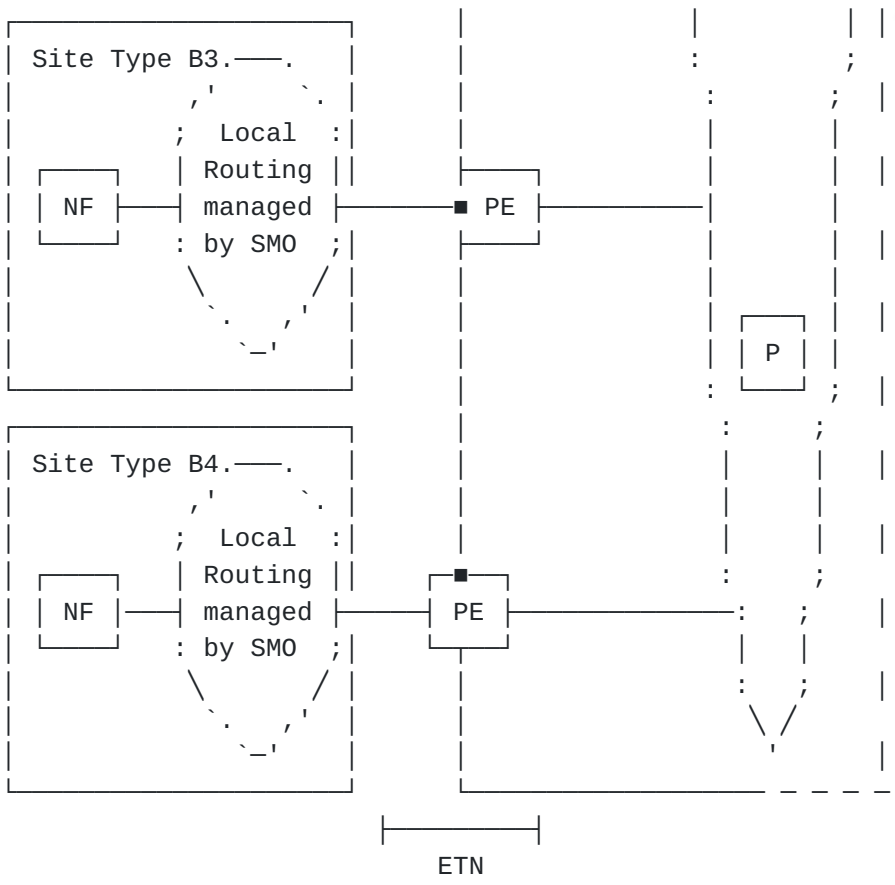
*L2 vs L3 Local Segment: the Local Segment can interconnect the NF and the ETN thanks to a unique vlan/LAN with no intermediate routing hop (the simplest example is an NF directly connected to a PE): A1, A2, A3 and A4. Alternatively, the NF interfaces may be attached in a different LAN/vlan than the ETN interface thanks to additional local routing between the ETN and the NF (e.g. CE, IP Fabric...): B1, B2, B3 and B4.

*ETN: the ETN can be either the PE or the CE if it is managed by the TN Orchestration (A1, A2, B1, B2).

*SDP: the SDP can be located in multiple places as per [I-D.ietf-teas-ietf-network-slices] section 4.2: A1 + B1 for case i), A2 + B2 for case i), B3 + A3 for case iii) and B4 + A4 for case iv)

*Redundancy/Scale-out: no example of redundancy/multihoming/scale-out is provided for the sake of simplification. Nonetheless, each Node/NF can be multiple.


Local segment
Transport Network
Site Type A1
NF
CE
PE
Site Type A2
NF
CE
PE
P
Site Type A3
NF
PE
Site Type A4
NF
PE
P
Site Type B1
Local Routing managed by SMO
NF
CE
PE
P
Site Type B2
Local Routing managed by SMO
NF
CE
PE
P



■ - Service Demarcation Point

Figure 3: Examples of various combinations of Local Segments, ETN and SDP

2.2.2. Orchestration of Local Segment termination at ETN

The interconnection between the 5G site and the Transport Network is made up of shared networking resources. More precisely, the Local Segment terminates to an interface of the ETN, which must be configured with consistent dataplane network information (e.g. vlan-id and IP addresses/subnets). Hence, the realization of this interconnection requires a coordination between the SMO and the Transport Orchestration (NSC). In this document, we assume that this coordination is based on IETF YANG data models and APIs (more details in further sections). The following diagram is a basic example of a L3CE-PE realization with shared network resource such as vlan-ID and IP prefixes, which must be passed between Orchestrators via the Network Slice Interface.

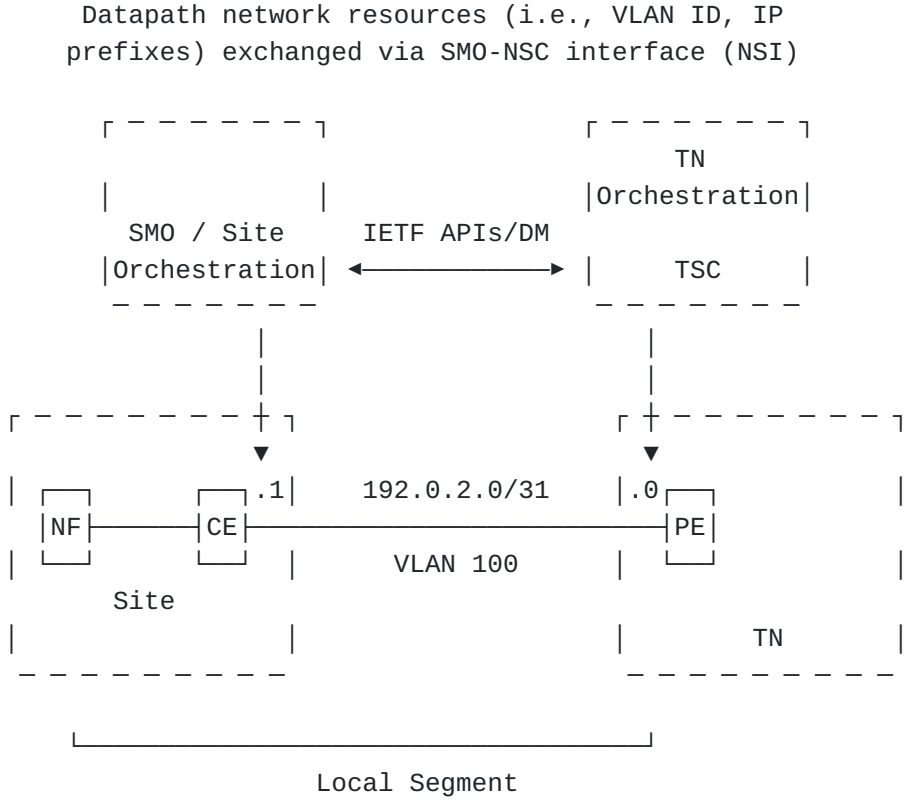


Figure 4: example

Note that the allocation of these resources (e.g. vlan or IPAM) can be either managed by the SMO or the Transport Network. In other words, the initial SMO request for the creation of a new IETF Network Slice on a given 5G site may or may not include all network

resources. In the latter case, this information is exchanged in a second step.

2.3. 5G slice to IETF Network Slice mapping

There are multiple options to map 5G network slices to IETF Network Slices:

*1 to N: A single 5G Network Slice can map to multiple IETF Network Slices (1 to N). One example of such a case is the separation of the 5G Control Plane and User Plane: this use case is represented in Figure 5 where a slice (EMBB) is deployed with a separation of User Plane and Control Plane at Transport Network level.

*N to 1: Multiple 5G Network Slices may rely on a same IETF Network Slice (i.e., in [TS-28.530] semantic, two RAN/CORE NSS rely on a shared TN NSS). In this case, the SLA differentiation of slices would be entirely controlled at 5G Control Plane, for example with appropriate placement strategies: this use case is represented in Figure 6, where a UPF network function for the URLLC slice is instantiated at the Edge Cloud close the gNB CU-UP User Plane for better latency/jitter control, while 5G Control Plane and the UPF for slice EMBB are instantiated in the Regional Cloud.

*N to M: the 5G to IETF Network Slice mapping combines both approaches with a mix of shared and dedicated associations.

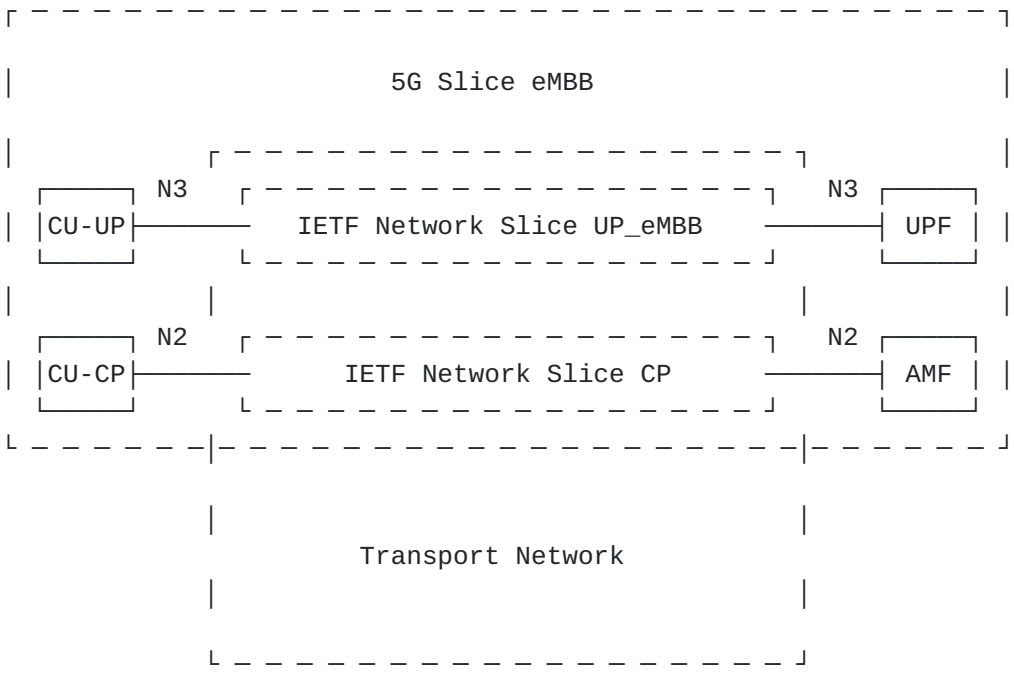


Figure 5: 1 (5G slice) to N (IETF Network Slice) mapping

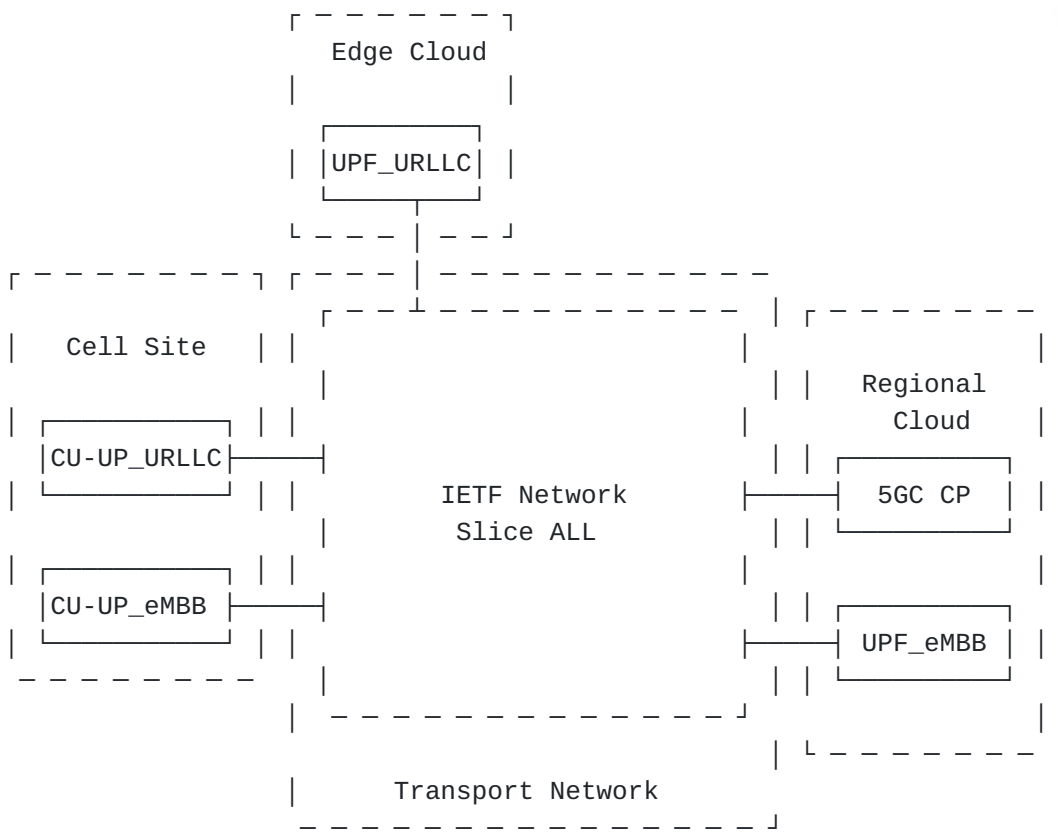


Figure 6: N (5G slice) to 1 (IETF Network Slice) mapping

Note that the actual realization of the mapping depends on several factors such as the actual business cases, the VNF vendor capabilities, the VNF vendor reference designs, as well as service provider or even legal requirements.

2.4. First 5G slice versus subsequent slices

A 5G Network Slice is fully functional with both 5G Control Plane and User Plane capabilities (i.e., dedicated NF functions or contexts). In this regard, the creation of the "first slice" is subject to a specific logic since it must deploy both CP and UP. This is not the case for the deployment of subsequent slices because they can share the CP of the First Slice, while instantiating dedicated UP. An example of an incremental deployment is depicted in Figure 7

At the time of writing, [NG.113], Section 6.2, specifies that the eMBB slice (SST=1 and no SD) should be supported globally. This 5G slice would be the first slice in any 5G deployment.

Note that the actual realization of the mapping depends on several factors such as the actual business cases, the VNF vendor capabilities, the VNF vendor reference designs, as well as service providers or even legal requirements.

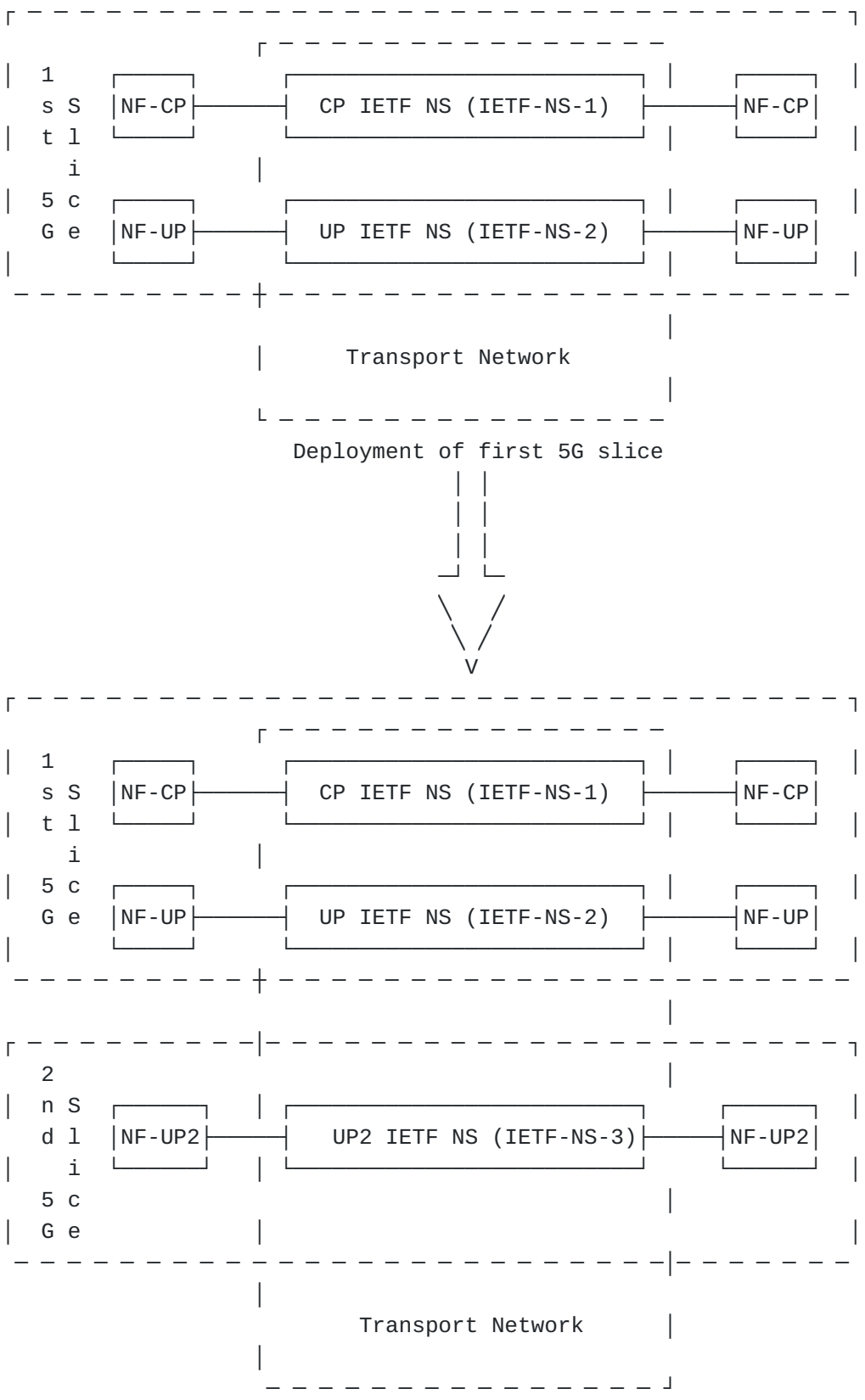


Figure 7: First and Subsequent Slice Deployment

3. High-Level Overview of the Realization Model

[I-D.ietf-teas-ietf-network-slices] introduces the concept of a Network Resource Partition (NRP), which is defined as a collection of resources identified in the underlay network. In the basic realization model described in this document, a single NRP is used with following characteristics

*L2VPN/L3VPN service instances for logical separation:

This realization model of transport for 5G slices assumes Layer-3 delivery for midhaul and backhaul transport connections, and a Layer 2 or Layer 3 (eCPRI supports both) delivery model for fronthaul connections. L2VPN/L3VPN service instances might be used as basic form of logical slice separation. Further, using service instances results in additional outer header (as packets are encapsulated/decapsulated at the nodes performing PE functions) providing clean discrimantion between 5G QoS and TN QoS, as explained in Section 4

*Fine-Grained resource control at the edge links of TN domain (attachment circuits):

This is sometimes called 'admission control' or 'traffic conditioning'. The main purpose is the enforcement of the bandwidth contract for the slice right at the edge of the transport domain where the traffic is handed-off between the transport domain and the 5G domains (i.e., RAN/Core). The toolset used here is granular ingress policing (rate limiting) to enforce contracted bandwidths per slice and, potentially, per traffic class within the slice. Out-of-contract traffic might be immediately dropped, or marked as high drop probability traffic, which is more likely to be dropped somewhere at the transit if congestion occurs. In the egress direction at the edge of the transport domain, hierarchical schedulers/shapers can be deployed, providing guaranteed rates per slice, as well as guarantees per traffic class within the slice.

*Coarse resource control at the TN transit (non-attachment circuits) links of the transport domain, using a single Network Resource Partition (NRP), spanning the entire TN domain

Transit nodes do not maintain any state of individual slices. Instead, only a flat (non-hierarchical) QoS model is used on transit links with up to 8 traffic classes. At the transport domain edge, traffic-flows from multiple slice services are

mapped to the limited number of traffic classes used on transit links.

*Capacity planning/management for efficient usage of TN edge and TN transit resources:

The role of capacity management is to ensure the transport capacity can be utilized without causing any bottlenecks. The toolset used here can range from careful network planning, to ensure more less equal traffic distribution (i.e., equal cost load balancing), to advanced traffic engineering techniques, with or without bandwidth reservations, to force more consistent load distribution even in non-ECMP friendly network topologies.

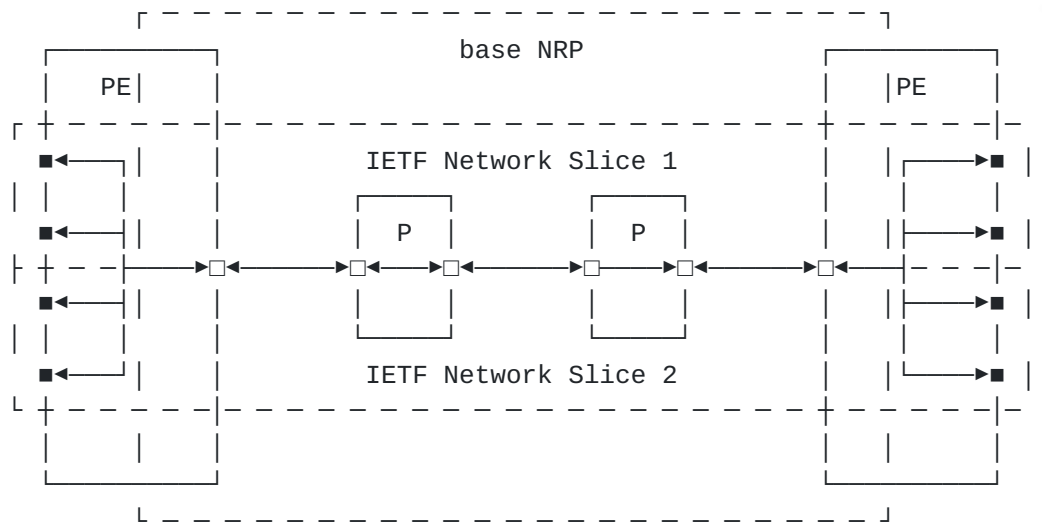


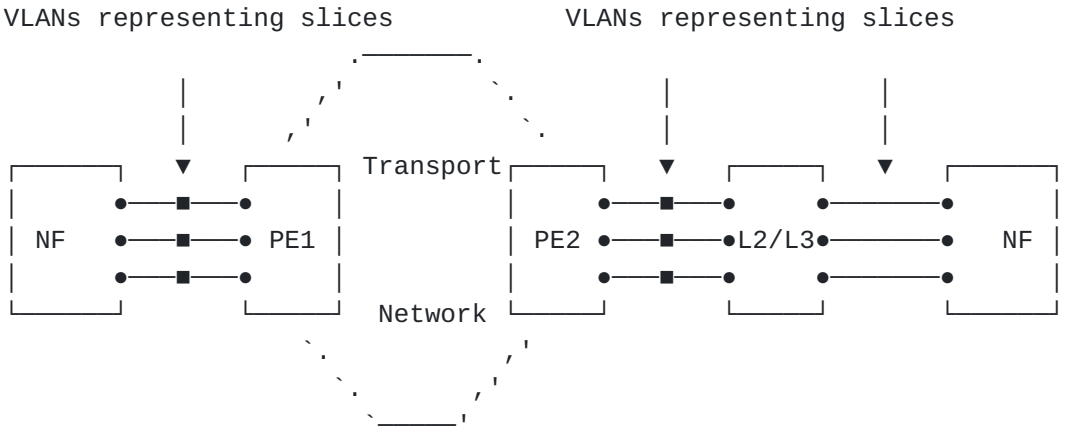
Figure 8: Resource allocation in with single NRP slicing model

The 5G control plane relies on S-NSSAI (Single Network Slice Selection Assistance Information: 32-bit slice identifier) for slice identification. The S-NSSAI is not visible to the transport domain, so instead 5G functions can expose the 5G slices to the transport domain by mapping to explicit L2/L3 identifiers such as VLAN, IP addresses or DSCP, as documented in [I-D.geng-teas-network-slice-mapping].

3.1. VLAN Hand-off

In this option, the IETF Network Slice, fulfilling connectivity requirements between NFs of some 5G slice, is represented at the SDP by a VLAN, or double VLANs (commonly known as QinQ). Each VLAN can represent a distinct logical interface on the attachment circuits, hence it provides the possibility to place these logical interfaces

in distinct L2 or L3 service instances and implement separation between slices via service instances. Since the 5G interfaces are IP based interfaces (the only exception could be the F2 fronthaul-interface, where eCPRI with Ethernet encapsulation is used), this VLAN is typically not transported across the TN domain. Typically, it has only local significance at a particular SDP. For simplification it is recommended to rely on a same VLAN identifier for all ACs, when possible. However, SDPs for a same slice at different locations may also use different VLAN values. Therefore, a VLAN to IETF Network Slice mapping table MUST be maintained for each AC, and the VLAN allocation MUST be coordinated between TN domain and extended RAN/Core domains. Thus, while VLAN hand-off is simple from the NF point of view, it adds complexity due to the requirement of maintaining mapping tables for each SDP.

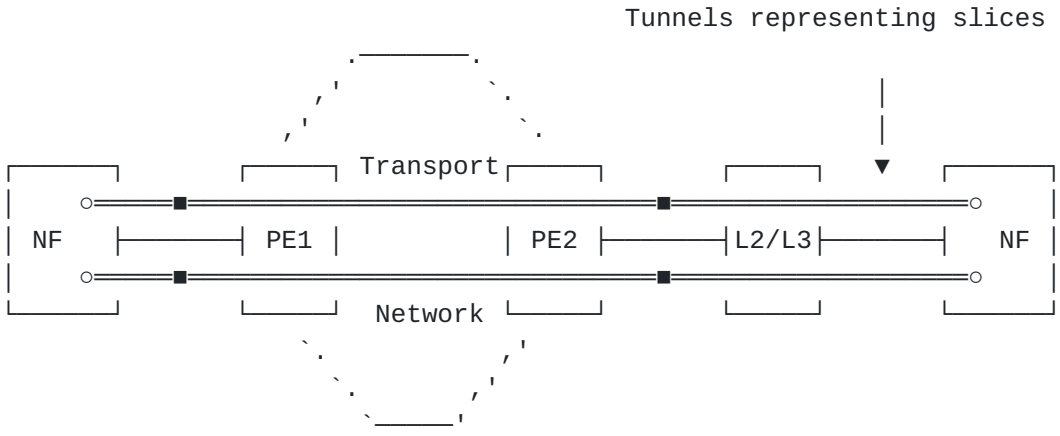


- ● – logical interface represented by VLAN on physical interface
- ■ - Service Demarcation Point

Figure 9: 5G slice with VLAN hand-off

3.2. IP Hand-off

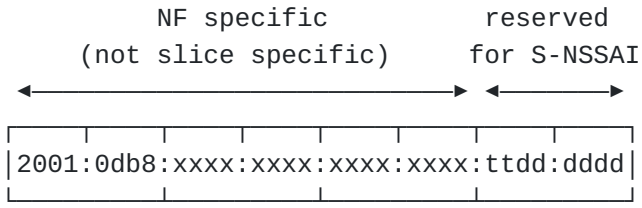
In this option, the slices in the transport domain are instantiated by IP tunnels (for example, IPsec, GTP-U tunnel) established between NFs. The transport for a single 5G slice is constructed with multiple such tunnels, since a typical 5G slice contains many NFs - especially DUs and CUs. If a shared NF (i.e., an NF that serves multiple slices, for example a shared DU) is deployed, multiple tunnels from shared NF are established, each tunnel representing a single slice. As opposed to the VLAN hand-off case, there is no logical interface representing slice on the PE, hence all slices are handled within single service instance. On the other hand, similarly to the VLAN hand-off case, a mapping table tracking IP to IETF Network Slice mapping is required.



○ – virtual interface à la loopback (not associated with a VLAN)
for tunnel (IPsec, GTP-U, ...) termination
■ - Service Demarcation Point

Figure 10: 5G slice with IP hand-off

The mapping table can be simplified if, e.g., IPv6 addressing is used to address NFs. An IPv6 address is a 128-bit long field, while the S-NSSAI is a 32-bit field: Slice/Service Type (SST): 8 bits, Slice Differentiator (SD): 24 bits. 32 bits, out of 128 bits of the IPv6 address, MAY be used to encode the S-NSSAI, which makes an IP to Slice mapping table unnecessary. This is simply an allocation method to allocate IPv6 addresses to NF loopbacks, without redefining IPv6 semantic. Different IPv6 address allocation schemes following this concept MAY be used, with one example allocation showed in Figure 11. This addressing scheme is local to a node; intermediary nodes are not required to associate any additional semantic with IPv6 address.



tt - SST (8 bits)
dddddd - SD (24 bits)

Figure 11: An Example of S-NSSAI embedded into IPv6

In the example, the most significant 96 bits of the IPv6 address are unique to NF, but do not carry any slice specific information, while the least significant 32 bits are used to embed S-NSSAI information. The 96-bit part of the address could be further divided, based for example on geographical location, or DC identification. 128 bits is

wide enough to design an IPv6 addressing scheme, which is most suitable for particular 5G deployment.

Figure 12 shows an example slicing deployment, where S-NSSAI is embedded into IPv6 addresses used by NFs. NF-A has a loopback interface, used to terminate tunnels, with unique per slice IP addresses allocated from 2001:db8::a:0:0/96 subnet, while NF-B uses loopback interface with per slice IP addresses allocated from 2001:db8::b:0:0/96. This example shows two slices: eMBB (SST=1) and MIoT (SST=3). Therefore, for eMBB the tunnel IP addresses are auto-derived (without the need for a mapping table) as {2001:db8::a:100:0, 2001:db8::b:100:0}, while for MIoT (SST=3) tunnel uses {2001:db8::a:300:0, 2001:db8::b:300:0}.

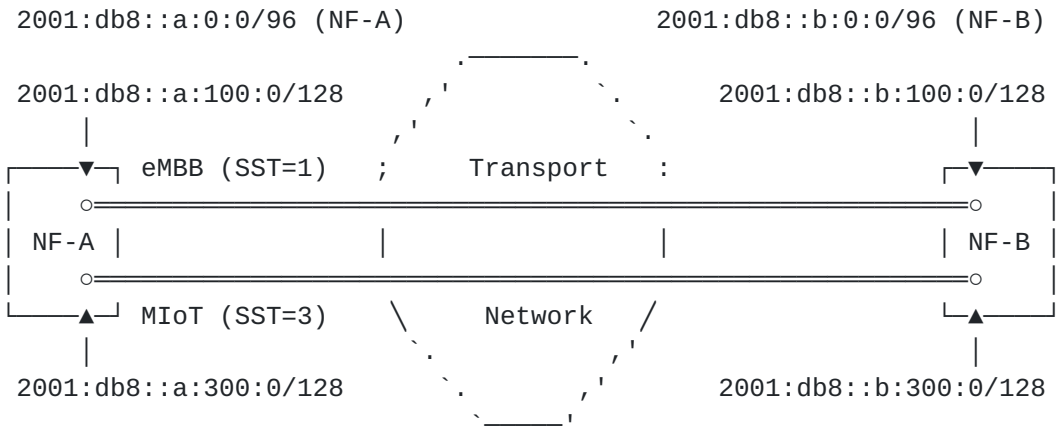


Figure 12: Deployment example with S-NSSAI embedded into IPv6

3.3. MPLS Label Hand-off

In this option, the service instances representing different slices are created directly on the NF, or within the cloud infrastructure hosting the NF, and attached to the TN domain. Therefore, the packet is MPLS encapsulated outside the TN domain with native MPLS encapsulation, or MPLSoUDP encapsulation, depending on the capability of the NF or cloud infrastructure, with the service label depicting the slice. There are three major methods (based on [RFC4364], Section 10) for interconnecting multiple service domains:

*Option A: VRF-to-VRF connections

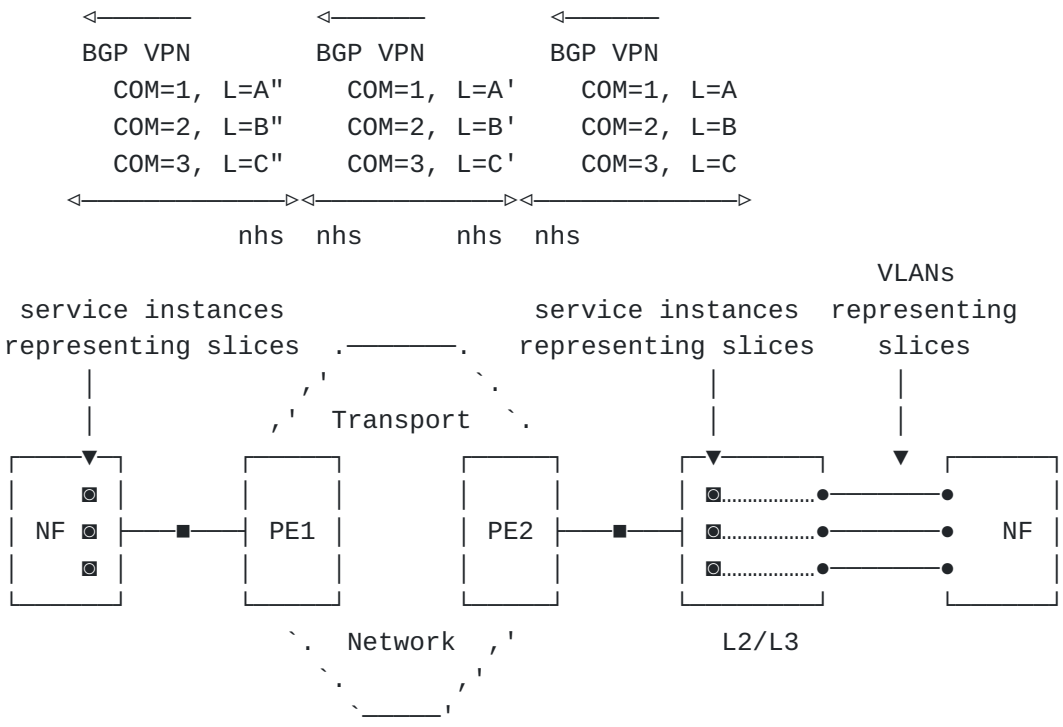
*Option B: redistribution of labeled VPN routes with next-hop change at domain boundaries

*Option C: redistribution of labeled VPN routes without next-hop change + redistribution of labeled transport routes with next-hop change at domain boundaries

MPLS is not used in VRF-to-VRF hand-offs, since services are terminated at the boundary of each domain, and VLAN hand-off is in place between the domains. Thus, it is the same as VLAN hand-off, described in Section 3.1.

3.3.1. Option B

In the Option B scenario, service instances for different IETF Network Slice services are instantiated outside the TN domain. They could be instantiated either on the compute, hosting mobile network functions (Figure 13, left hand side), or within the cloud infrastructure itself, e.g. on the top of the rack or leaf switch within cloud IP fabric (Figure 13, right hand side). Between TN domain and the (extended) RAN/CN domain, packets are MPLS encapsulated (or MPLSoUDP encapsulated, if cloud or compute infrastructure doesn't support native MPLS encapsulation), therefore the PE uses neither a VLAN, nor an IP address for slice identification at SDP, but instead uses the MPLS label.



● – logical interface represented by VLAN on physical interface
◙ - service instances (with unique MPLS label)
■ - Service Demarcation Point

Figure 13: MPLS Hand-off: Option B

MPLS labels are allocated dynamically, especially in Option B deployments, where at the domain boundaries service prefixes are reflected with next-hop self, and new label is dynamically

allocated, as visible in Figure 13. Therefore, for any slice-specific per hop behavior at the TN domain edge, the PE must be able to determine which label represents which slice. In the BGP control plane, when exchanging service prefixes between (extended) RAN/CN domains and TN domain, each slice might be represented by a unique BGP community, so tracking label assignment to the slice is possible. For example, in Figure 13, for the slice identified with COM=1, PE1 advertises a dynamically allocated label A". Since, based on the community, the label to slice association is known, PE1 can use this dynamically allocated label A" to identify incoming packets as belonging to slice 1, and execute appropriate edge per hop behavior.

It is worth noting that slice identification in the BGP control plane is at the prefix granularity. In extreme case, each prefix can have different community representing a different slice. Depending on the business requirements, each slice could be represented by a different service instance, as outlined in Figure 13. In that case, the route target extended community might be used as slice differentiator. In another deployment, all prefixes (representing different slices) might be handled by single 'mobile' service instance, and some other BGP attribute (e.g., a standard community) might be used for slice differentiation. Or there could be a deployment that groups multiple slices together into a single service instance, resulting in a handful of service instances. In any case, fine-grained per-hop behavior at the edge of TN domain is possible.

4. QoS Mapping Models

The resources are managed via various QoS policies deployed in the network. QoS mapping models to support 5G slicing connectivity implemented over packet switched transport uses two layers of QoS

*5G QoS

At this layer QoS treatment is indicated by the 5QI (5G QoS indicator), as defined in [TS-23.501]. A 5QI is an ID that is used as a reference to 5G QoS characteristics (e.g., scheduling weights, admission thresholds, queue management thresholds, link layer protocol configuration, etc.) in the RAN domain. Given the fact that 5QI applies to the RAN domain, it is not visible to the TN domain. Therefore, if 5QI-aware treatment is desired in the TN domain as well, 5G components might set DSCP with a value representing 5QI, to allow differentiated treatment in TN domain as well. Based on these DSCP values, at SDP of each TN segment used to construct transport for given 5G slice, very granular QoS enforcement might be implemented. The mapping between 5QI and DSCP is out of scope for this document. Mapping recommendations

are documented in [I-D.henry-tsvwg-diffserv-to-qci]. Each slice might have flows with multiple 5QIs, thus there could be many different 5QIs being deployed. 5QIs (or, more precisely, corresponding DSCP values) are visible to the TN domain at SDP (i.e., at the edge of the TN domain).

In this document, this layer of QoS will be referred as '5G QoS Class' ('5G QoS' in short), or '5G DSCP'.

*TN QoS

Control of the TN resources on transit links, as well as traffic scheduling/prioritization on transit links, is based on a flat (non-hierarchical) QoS model in the IETF Network Slice realization. That is, IETF Network Slices are assigned dedicated resources (e.g., QoS queues) at the edge of the TN domain (at SDP), while all IETF Network Slices are sharing resources (sharing QoS queues) on the transit links of the TN domain. Typical router hardware can support up to 8 traffic queues per port, therefore the architecture assumes 8 traffic queues per port support in general.

At this layer, QoS treatment is indicated by QoS indicator specific to the encapsulation used in the TN domain, and it could be DSCP or MPLS TC. This layer of QoS will be referred as 'TN QoS Class', or 'TN QoS' for short, in this document.

While 5QI might be exposed to the TN domain, via the DSCP value (corresponding to specific 5QI value) set in the IP packet generated by NFs, some 5G deployments might use 5QI in the RAN domain only, without requesting per 5QI differentiated treatment from the TN domain. This can be due to an NF limitation (no capability to set DSCP), or it might simply depend on the overall slicing deployment model. The O-RAN Alliance, for example, defines a phased approach to the slicing, with initial phases utilizing only per slice, but not per 5QI, differentiated treatment in the TN domain ([O-RAN.WG9.XPSAAS], Annex F).

Therefore, from QoS perspective, the 5G slicing connectivity realization architecture defines two high-level realization models for slicing in the transport domain: a 5QI-unaware model and a 5QI-aware model. Both slicing models in the transport domain could be used concurrently within the same 5G slice. For example, the TN segment for 5G midhaul (F2-U interface) might be 5QI-unaware, while at the same time the TN segment for 5G backhaul (N3 interface) might follow the 5QI-aware model.

4.1. 5QI-unaware Mode

In 5QI-unaware mode, the DSCP values in the packets received from NF at SDP are ignored. In the TN domain, there is no QoS differentiation at the 5G QoS Class level. The entire IETF Network Slice is mapped to single TN QoS Class, and, therefore, to a single QoS queue on the routers in the TN domain. With a small number of deployed 5G slices (for example only two 5G slices: eMBB and MIoT), it is possible to dedicate a separate QoS queue for each slice on transit routers. However, with introduction of private/enterprises slices, as the number of 5G slices (and thus corresponding IETF Network Slices) increases, a single QoS queue on transit links serves multiple slices with similar characteristics. QoS enforcement on transit links is fully coarse (single NRP, sharing resources among all IETF Network Slices), as displayed in Figure 14.

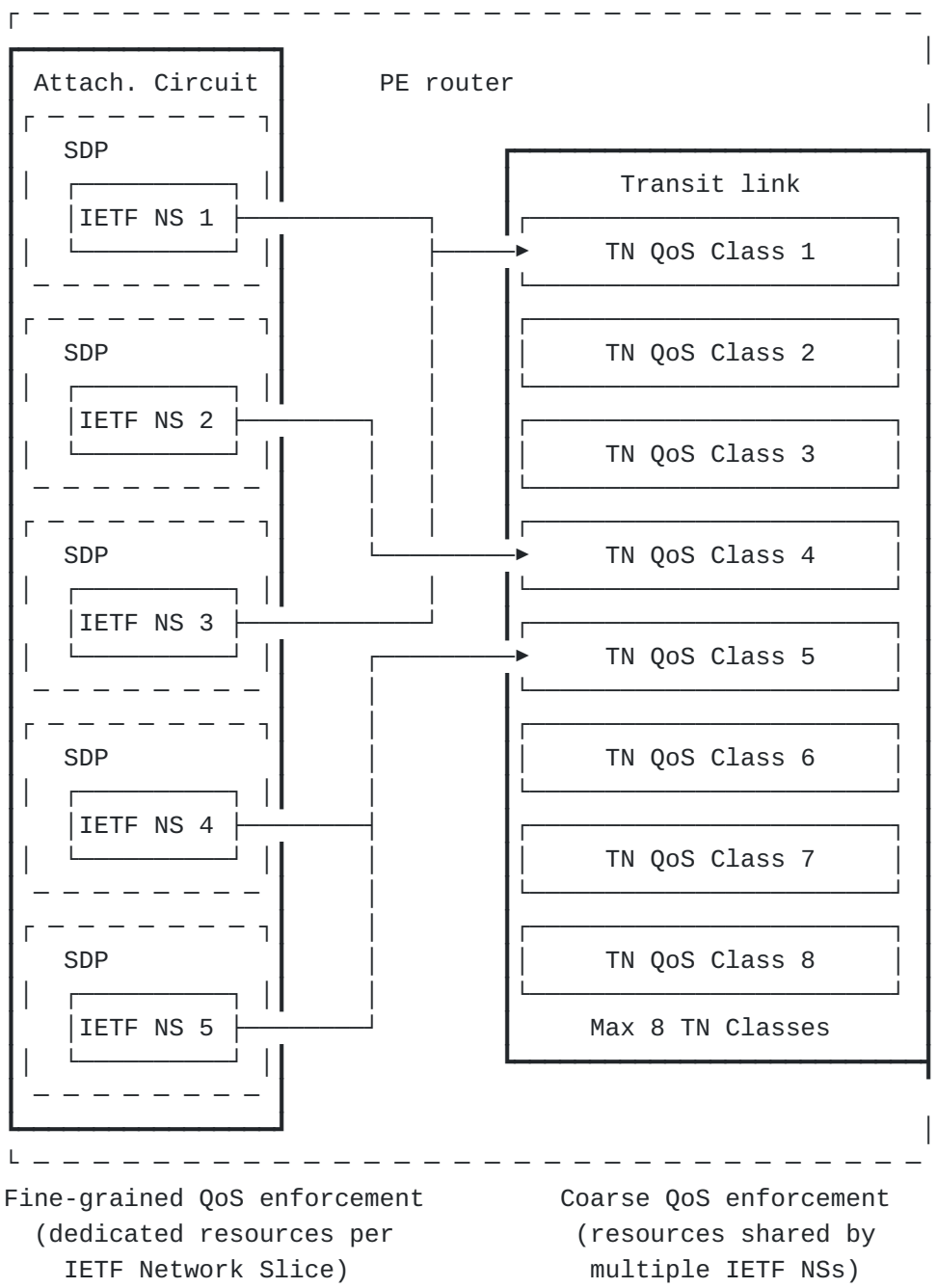


Figure 14: Slice to TN QoS mapping (5QI-unaware model)

When the IP traffic is handed over at the SDP from the extended RAN or extended Core domains to the TN domain, the PE encapsulates the traffic into MPLS (if MPLS transport is used in the TN domain), or IPv6 - optionally with some additional headers (if SRv6 transport is used in the TN domain), and sends out the packets on the TN transit link.

The original IP header retains the DCSP marking (which is ignored in 5QI-unaware mode), while the new header (MPLS or IPv6) carries QoS marking (MPLS Traffic Class bits for MPLS encapsulation, or DSCP for

SRv6/IPv6 encapsulation) related to TN CoS. Based on TN QoS Class marking, per hop behavior for all IETF Network Slices is executed on TN links. TN domain transit routers do not evaluate the original IP header for QoS-related decisions. This model is outlined in Figure 15 for MPLS encapsulation, and in Figure 16 for SRv6 encapsulation.

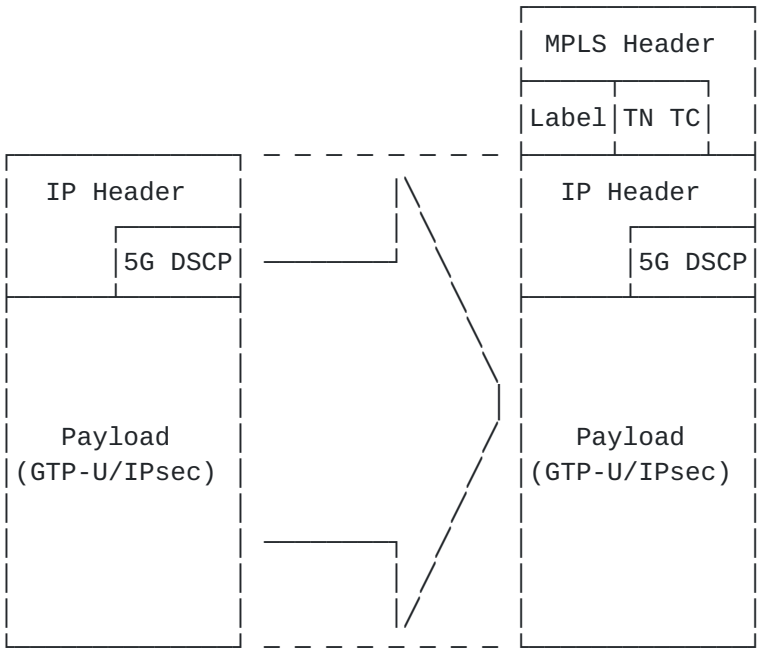


Figure 15: QoS with MPLS encapsulation

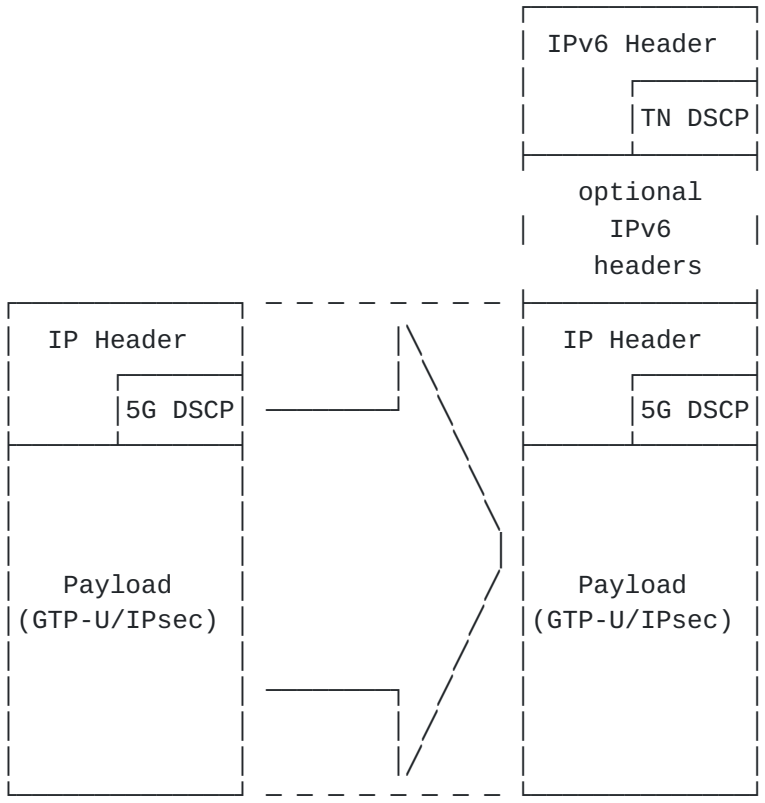


Figure 16: QoS with IPv6 encapsulation

From the QoS perspective, both options are similar. However, there is one difference between the two options. The MPLS TC is only 3 bits (8 possible combinations), while DSCP is 6 bits (64 possible combinations). Hence, SRv6 provides more flexibility for TN CoS design, especially in combination with soft policing with in-profile/out-profile, as discussed in Section 4.1.1.

Edge resources are controlled in a very granular, fine-grained manner, with dedicated resource allocation for each IETF Network Slice. The resource control/enforcement happens at each SDP in two directions: inbound and outbound.

4.1.1. Inbound Edge Resource Control

The main aspect of inbound edge resource control is per-slice traffic capacity enforcement. This kind of enforcement is often called 'admission control' or 'traffic conditioning'. The goal of this inbound enforcement is to ensure that the traffic above the contracted rate is dropped or deprioritized, depending on the business rules, right at the edge of TN domain. This, combined with appropriate network capacity planning/management, as described in Section 6, is required to ensure proper isolation between slices in scalable manner. As a result, traffic of one slice has no influence on the traffic of other slices, even if the slice is misbehaving

(i.e., DDoS attack, equipment failure, etc.) and generates traffic volumes above the contracted rates.

The slice rates can be characterized with following parameters [I-D.ietf-teas-ietf-network-slice-nbi-yang]:

*CIR: Committed Information Rate (i.e., guaranteed bandwidth)

*PIR: Peak Information Rate (i.e., maximum bandwidth)

These parameters define the traffic characteristics of the slice and are part of SLO parameter set provided by the SMO to IETF NSC. Based on these parameters the inbound policy can be implemented using one of following options:

*1r2c (single-rate two-color) rate limiter

This is the most basic rate limiter, which meters at the SDP a traffic stream of given slice and marks its packets as in-contract (below contracted CIR) or out-of-contract (above contracted CIR). In-contract packets are accepted and forwarded. Out-of contract packets are either dropped right at the SDP (hard rate limiting), or remarked (with different MPLS TC or DSCP TN markings) to signify 'this packet should be dropped in the first place, if there is a congestion' (soft rate limiting), depending on the business policy of the operator. In the second case, while packets above CIR are forwarded at the SDP, they are subject to be dropped during any congestion event at any place in the TN domain.

*2r3c (two-rate three-color) rate limiter

This was initially defined in [RFC2698], and its improved version in [RFC4115]. In essence, the traffic is assigned to one of the 3 categories:

-green, for traffic under CIR

-yellow, for traffic between CIR and PIR

-red, for traffic above PIR

An inbound 2c3r meter implemented with [RFC4115], compared to [RFC2698], is more 'customer friendly' as it doesn't impose outbound peak-rate shaping requirements on customer edge (CE) devices. 2r3c meters in general give greater flexibility for edge enforcement regarding accepting the traffic (green), de-prioritizing and potentially dropping the traffic during congestion (yellow), or hard dropping the traffic (red).

Inbound edge enforcement mode for 5QI-unaware mode, where all packets belonging to the slice are treated the same way in the TN domain (no 5Q QoS Class differentiation in the TN domain) is outlined in Figure 17.

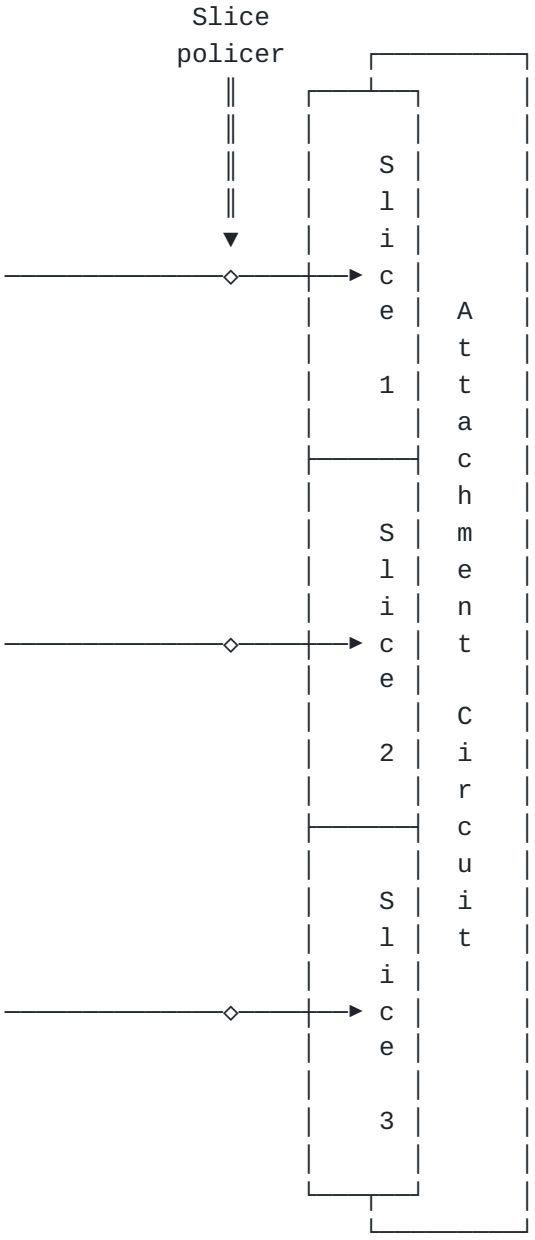


Figure 17: Ingress slice admission control (5QI-unware model)

4.1.2. Outbound Edge Resource Control

While inbound slice admission control at the transport edge is mandatory in the model, outbound edge resource control might not be

required in all use cases. Use cases that specifically call for outbound edge resource control are:

- Slices use both CIR and PIR parameters, and transport edge links (attachment circuits) are dimensioned to fulfil the aggregate of slice CIRs. If at any given time, some slices send the traffic above CIR, congestion in outbound direction on the transport edge link might happen. Therefore, fine-grained resource control to guarantee at least CIR for each slice is required.

- Any-to-Any (A2A) connectivity constructs are deployed, again resulting in potential congestion in outbound direction on the transport edge links, even if only slice CIR parameters are used. This again requires fine-grained resource control per slice in outbound direction at transport edge links.

As opposed to inbound edge resource control, typically implemented with rate-limiters/policers, outbound resource control is typically implemented with a weighted/priority queuing, potentially combined with optional shapers (per slice). A detailed analysis of different queuing mechanisms is out of scope for this document, but is provided in [RFC7806].

Figure 18 outlines the outbound edge resource control model at the transport network layer for 5QI-unaware slices. Each slice is assigned a single egress queue. The sum of slice CIRs, used as the weight in weighted queueing model, MUST not exceed the physical capacity of the attachment circuit. Slice requests above this limit MUST be rejected by the NSC, unless an already established slice with lower priority, if such exists, is preempted.

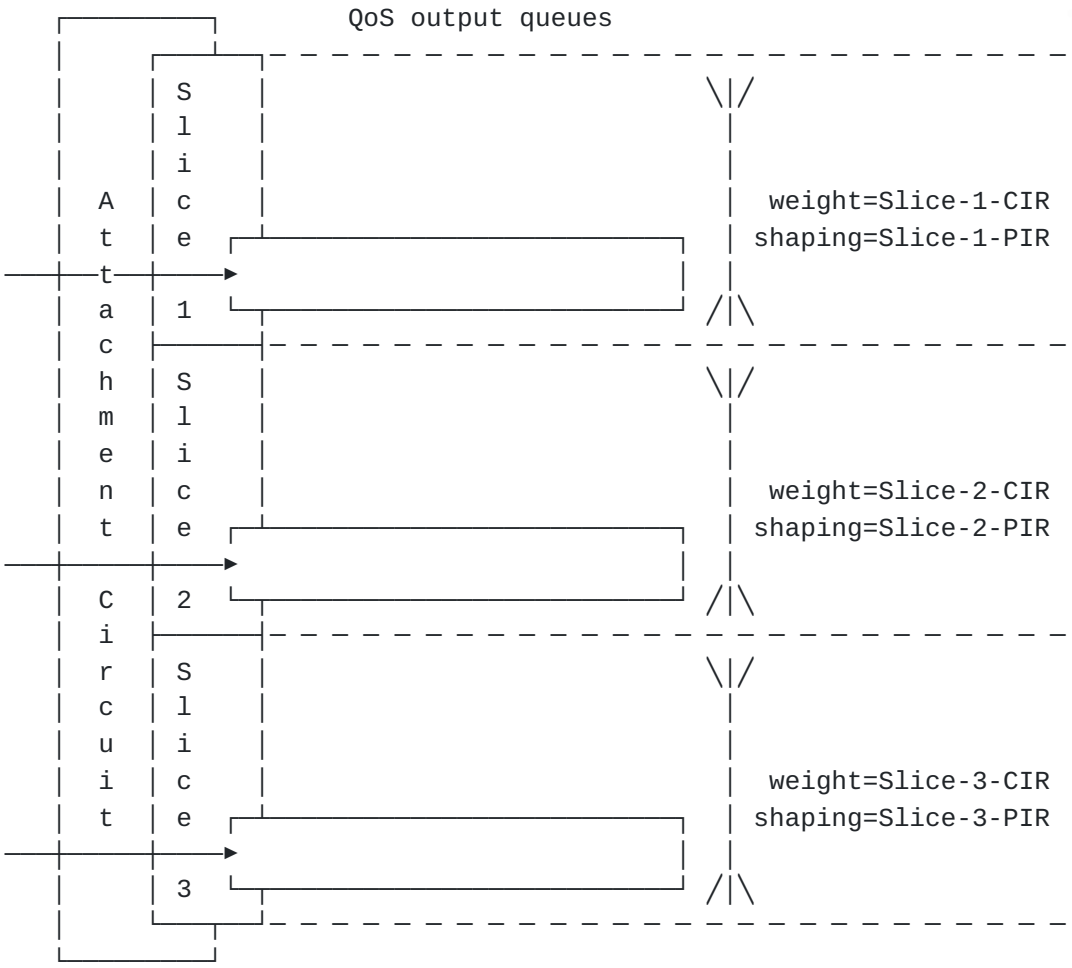


Figure 18: Egress slice admission control (5QI-unaware model)

4.2. 5QI-aware Mode

In the 5QI-aware model, potentially a large number of 5Q QoS Classes (the architecture scales to thousands of 5Q slices) is mapped (multiplexed) to up to 8 TN QoS Classes used in transport transit equipment, as outlined in Figure 19.

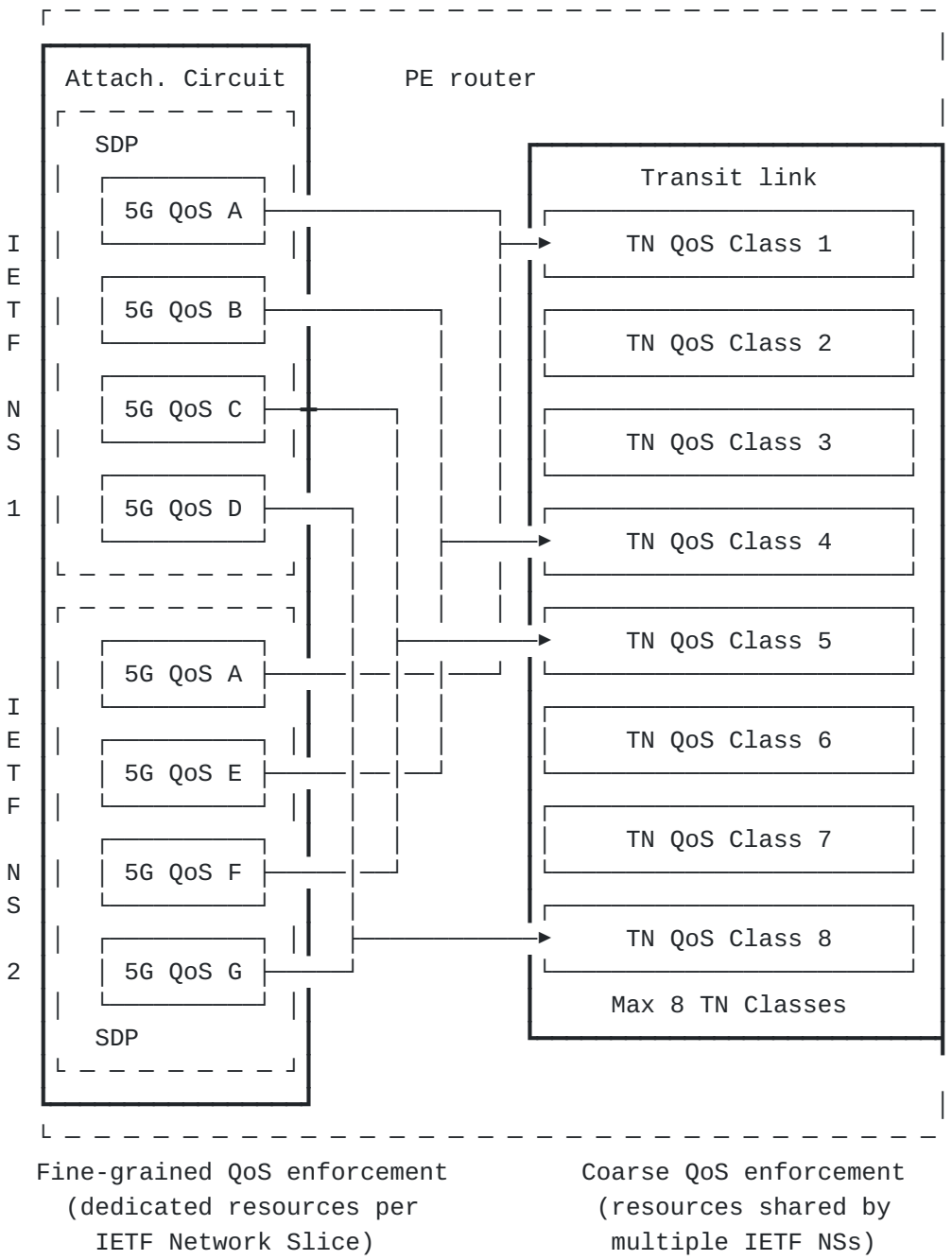


Figure 19: Slice 5Q QoS to TN QoS mapping (5QI-aware model)

Given the fact that in large scale deployments (large number of 5G slices), the number of potential 5G QoS Classes is much higher than the number of TN QoS Classes, multiple 5G QoS Classes with similar characteristics - potentially from different IETF Network Slices - can be mapped to a same TN QoS Class when transported in the TN domain. That is, common per hop behavior (PHB) is executed on transit TN routers for all packets grouped together.

Like in 5QI-unaware model, the original IP header retains the DCSP marking corresponding to 5QI (5G QoS Class), while the new header

(MPLS or IPv6) carries QoS marking related to TN QoS Class. Based on TN QoS Class marking, per hop behavior for all aggregated 5G QoS Classes from all IETF Network Slices is executed on TN links. TN domain transit routers do not evaluate original IP header for QoS related decisions. The original DSCP marking retained in the original IP header is used at the PE for fine-grained per slice and per 5G QoS Class inbound/outbound enforcement on AC link.

In 5QI-aware model edge resources are controlled in an even more granular, fine-grained manner, with dedicated resource allocation for each IETF Network Slice and dedicated resource allocation for number of traffic classes (most commonly up 4 or 8 traffic classes, depending on the HW capability of the equipment) within each IETF Network Slice.

4.2.1. Inbound Edge Resource Control

Compared to the 5QI-unware model, admission control (traffic conditioning) in the 5QI-aware model is more granular, as it enforces not only per slice capacity constraints, but may as well enforce the constraints per 5G QoS Class within each slice.

5G slice using multiple 5QIs can potentially specify rates in one of the following ways

* rates per traffic class (CIR, or CIR+PIR), no rate per slice (sum of rates per class gives the rate per slice)

* rate per slice (CIR, or CIR+PIR), and rates per prioritized (premium) traffic classes (CIR only). Best effort traffic class uses the bandwidth (within slice CIR/PIR) not consumed by prioritized classes

In the first option, the slice admission control is executed with traffic class granularity, as outlined in Figure 20. In this model, if a premium class doesn't consume all available class capacity, it cannot be reused by non-premium (i.e., Best Effort) class.

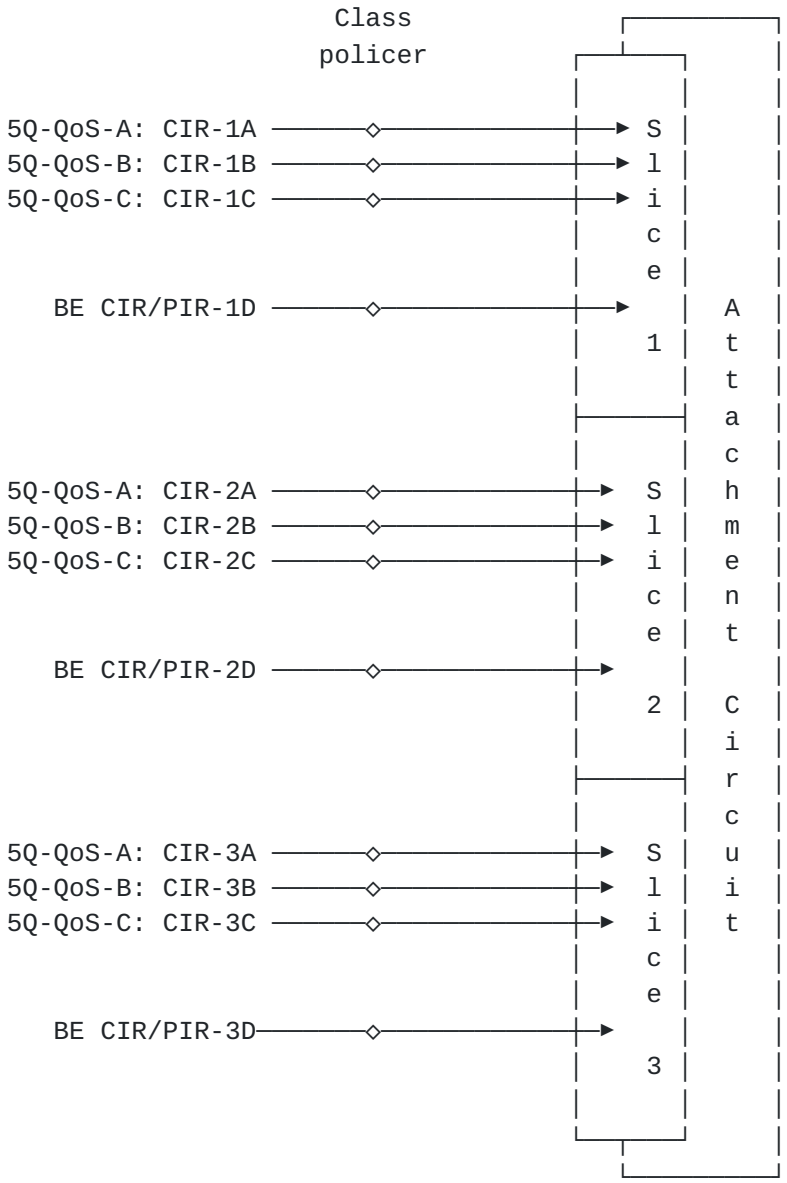


Figure 20: Ingress slice admission control (5QI-aware model)

The second model combines the advantages of 5QI-unaware model (per slice admission control) with the per traffic class admission control, as outlined in Figure 20. Ingress admission control is at class granularity for premium classes (CIR only). Non-premium class (i.e. Best Effort) has no separate class admission control policy, but is allowed to use entire slice capacity, which is available at any given moment. I.e., slice capacity, which is not consumed by premium classes. It is a hierarchical model, as depicted in Figure 21.

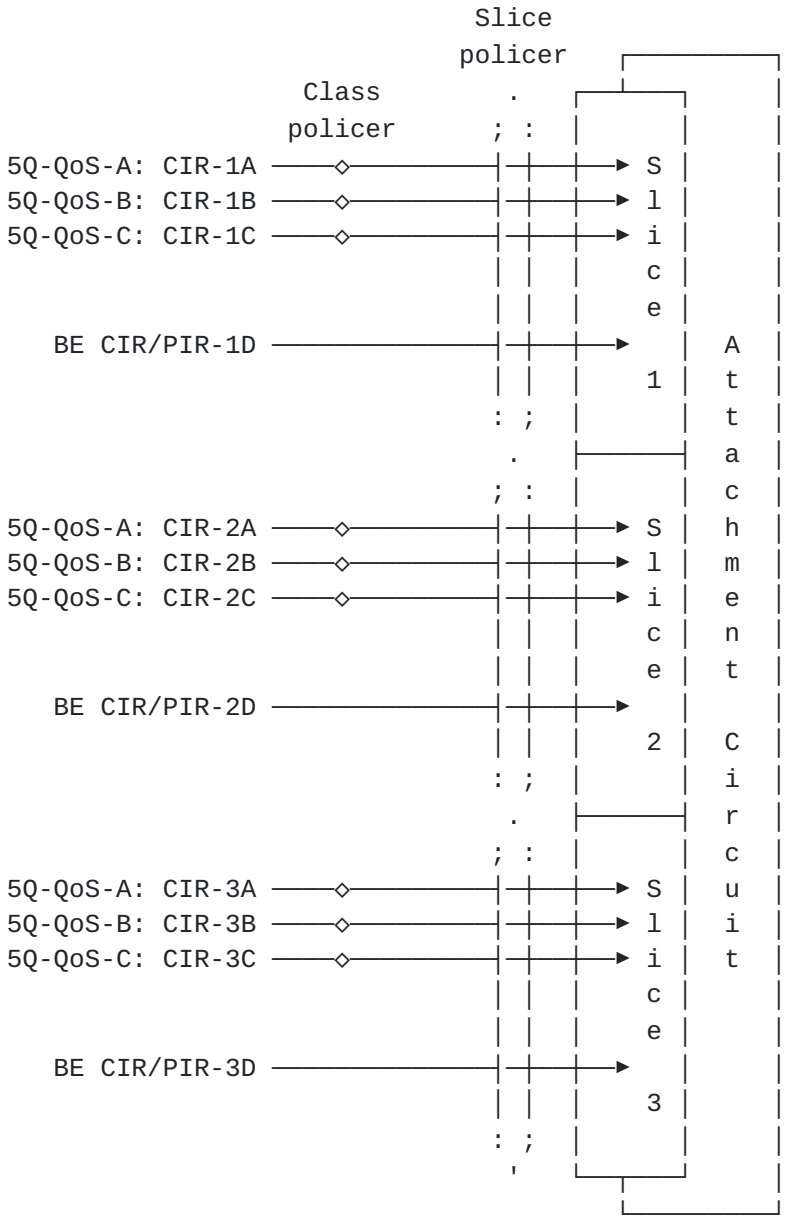


Figure 21: Ingress slice admission control (5QI-aware) - hierarchical

4.2.2. Outbound Edge Resource Control

Figure 22 outlines the outbound edge resource control model at the transport network layer for 5QI-aware slices. Each slice is assigned multiple egress queues. The sum of queue weights (equal to 5Q QoS CIRs within the slice) CIRs MUST not exceed the CIR of the slice itself. And, similarly to the 5QI-aware model, the sum of slice CIRs MUST not exceed the physical capacity of the attachment circuit.

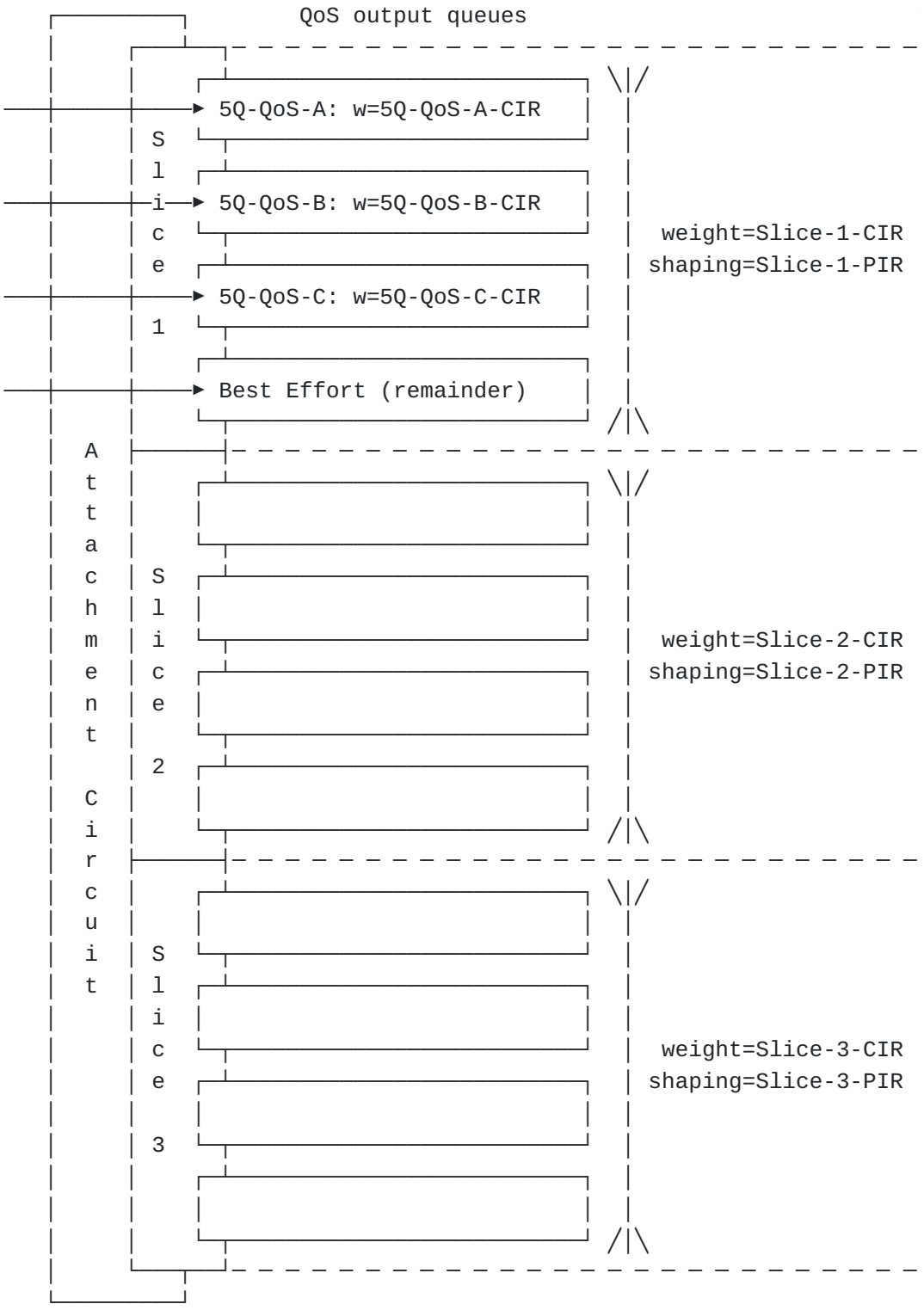


Figure 22: Egress slice admission control (5QI-aware)

4.3. Transit Resource Control

Transit resource control is much simpler than Edge resource control. As outlined in Figure 19, at the edge, 5Q QoS Class marking (represented by DSCP related to 5QI set by mobile components in the

packets handed off to the TN) is mapped to the TN QoS Class. Based in TN QoS Class, when the packet is encapsulated with outer header (MPLS or IPv6), TN QoS Class marking (MPLS TC or IPv6 DHCP in outer header, as depicted in Figure 15 and Figure 16) is set in the outer header. PHB on transit is based exclusively on that TN QoS Class marking, i.e., original 5G QoS Class DSCP is not taken into consideration on transit.

Transit resource control does not use any inbound interface policy, but only outbound interface policy, which is based on priority queue combined with weighted or deficit queuing model, without any shaper. The main purpose of transit resource control is to ensure that during network congestion events, for example caused by network failures and temporary rerouting, premium classes are prioritized, and any drops only occur in non-premium (best-effort) classes. Capacity planning and management, as described in Section 6, ensures that enough capacity is available to fulfill all approved slice requests.

5. Transport Planes Mapping Models

A network operator might define various groups of tunnels, where each tunnel group is created with specific optimization criteria and constraints. This document refers to such tunnel groups as 'transport planes'. For example, transport plane A might represent tunnels optimized for latency, transport plane B for high capacity, transport plane C might represent tunnels using only the "upper half" of the transport network, and transport plane D might represent tunnels using only the "lower half" of the transport network. Figure 23 depicts an example of a simple network with two transport planes. These transport planes might be realized via various IP/MPLS techniques, for example Flex-Algo or RSVP/SR traffic engineering tunnels with or without PCE, and with or without bandwidth reservations. Section 6 discusses in detail different bandwidth models that can be deployed in the transport network. However, discussion about how to realize or orchestrate transport planes is out of scope for this document.

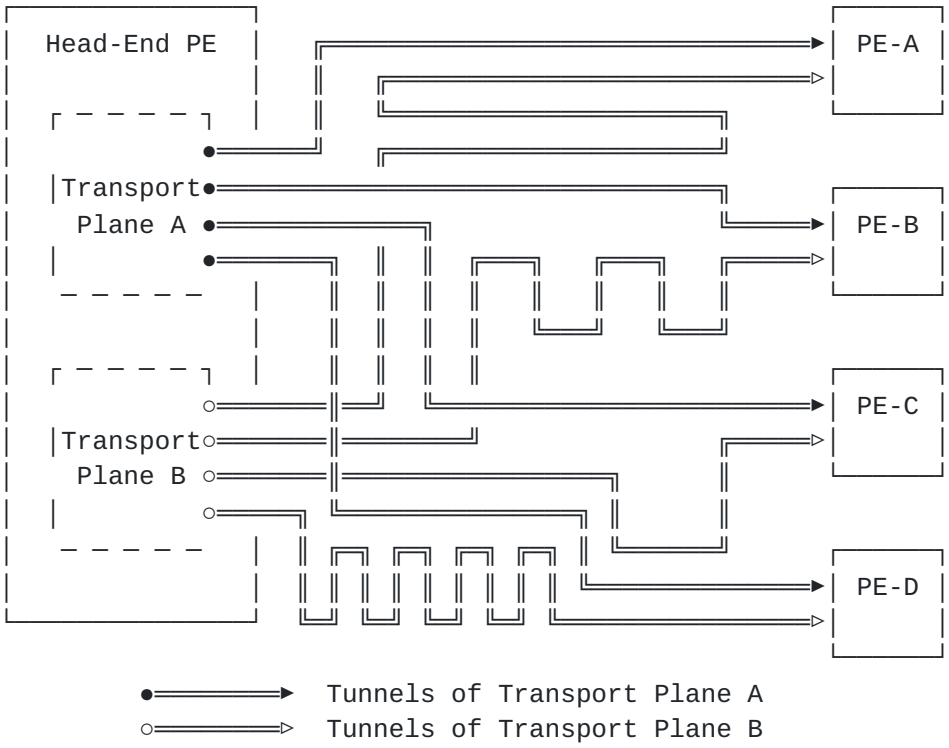


Figure 23: Transport Planes

Similar to the QoS mapping models discussed in Section 4, for mapping to transport planes at the ingress PE, both 5QI-unaware and 5QI-aware modes are defined. In essence, entire slices can be mapped to transport planes without 5G QoS consideration (5QI-unaware mode), or flows with different 5G QoS Classes, even if they are from the same slice, might be mapped to different transport planes (5QI-aware mode).

5.1. 5QI-unaware Mode

As discussed in Section 4.1, in the 5QI-unware model, the TN domain doesn't take into account 5G QoS during execution of per-hop behavior. The entire slice is mapped to single TN QoS Class, therefore the entire slice is subject to the same per-hop behavior. Similarly, in 5QI-unaware transport plane mapping mode, the entire slice is mapped to a single transport plane, as depicted in Figure 24.

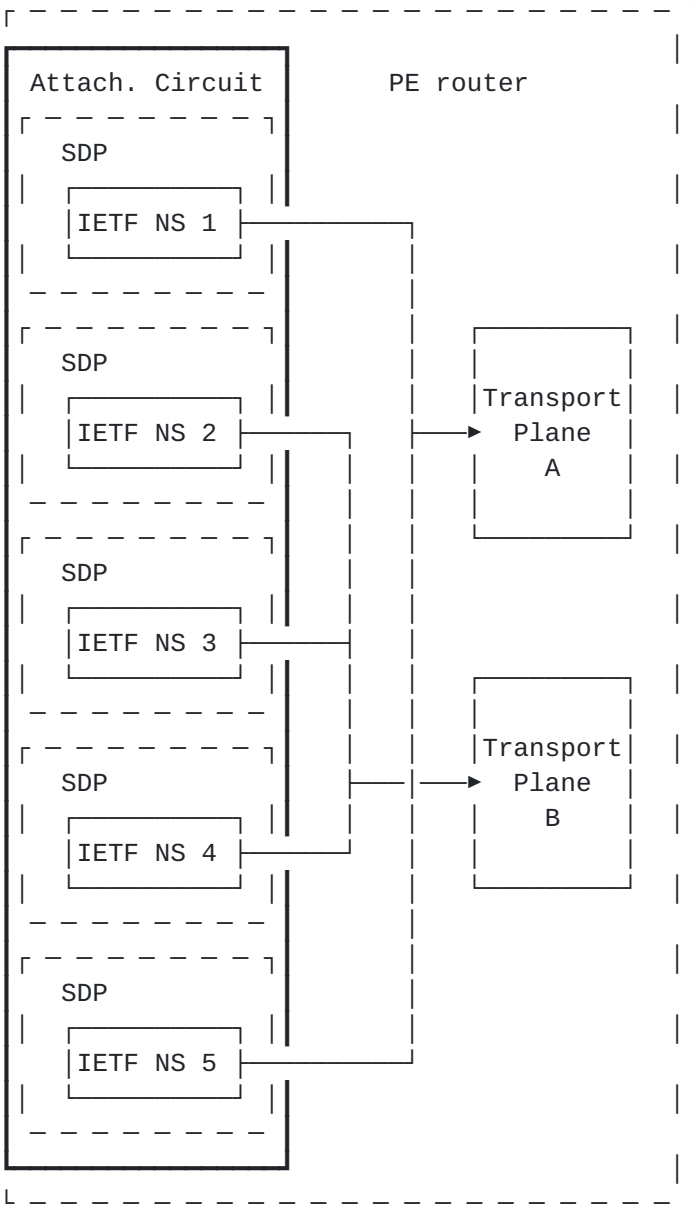


Figure 24: Slice to Transport Plane mapping (5QI-unaware model)

It is worth noting that there is no strict correlation between TN QoS Classes and Transport Planes. The TN domain can be operated with e.g., 8 TN QoS Classes (representing 8 hardware queues in the routers), and 2 Transport Classes (e.g., latency optimized transport plane using link latency metrics for path calculation, and transport plane following IGP metrics). TN QoS Class determines the per-hop behavior when the packets are transiting through the TN domain, while Transport Plane determines the path, optimized or constrained based on operator's business criteria, that the packets use to transit through the TN domain.

5.2. 5QI-aware Mode

In 5QI-aware mode, the traffic can be mapped to transport planes at the granularity of 5G QoS Class. Given that the potential number of transport planes is limited, packets from multiple 5G QoS Classes with similar characteristics are mapped to a common transport class, as depicted in Figure 25.

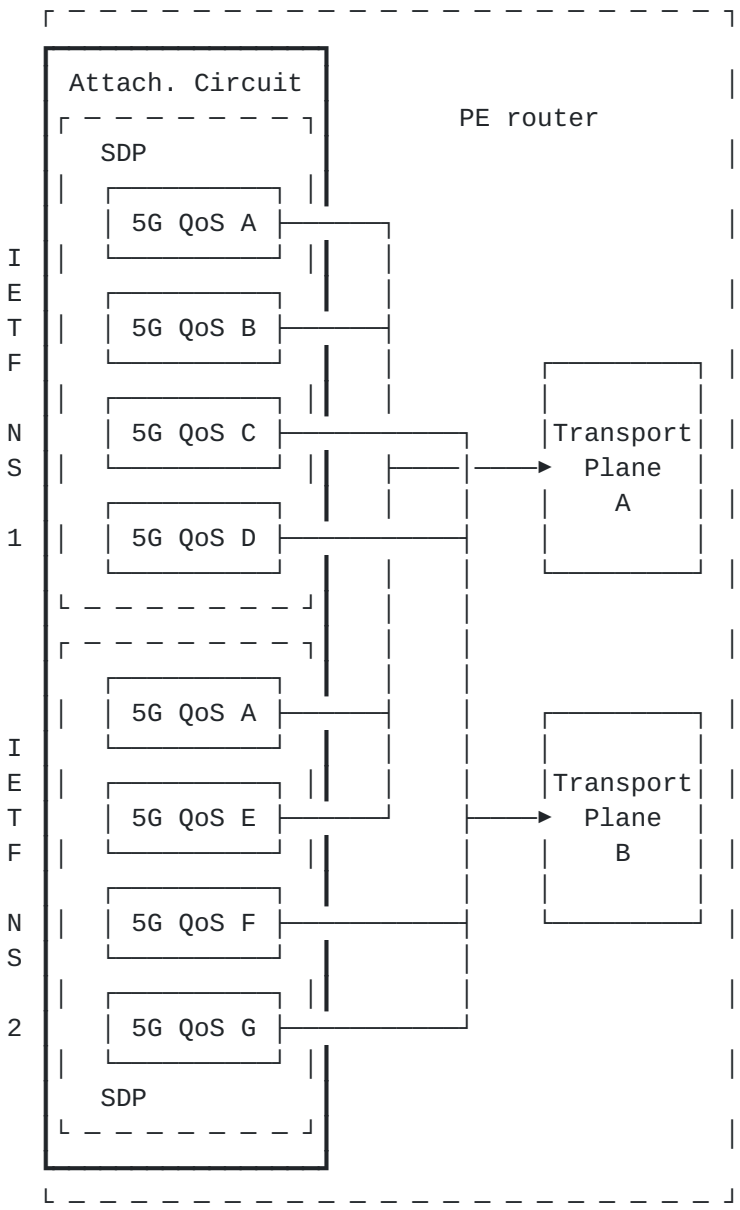


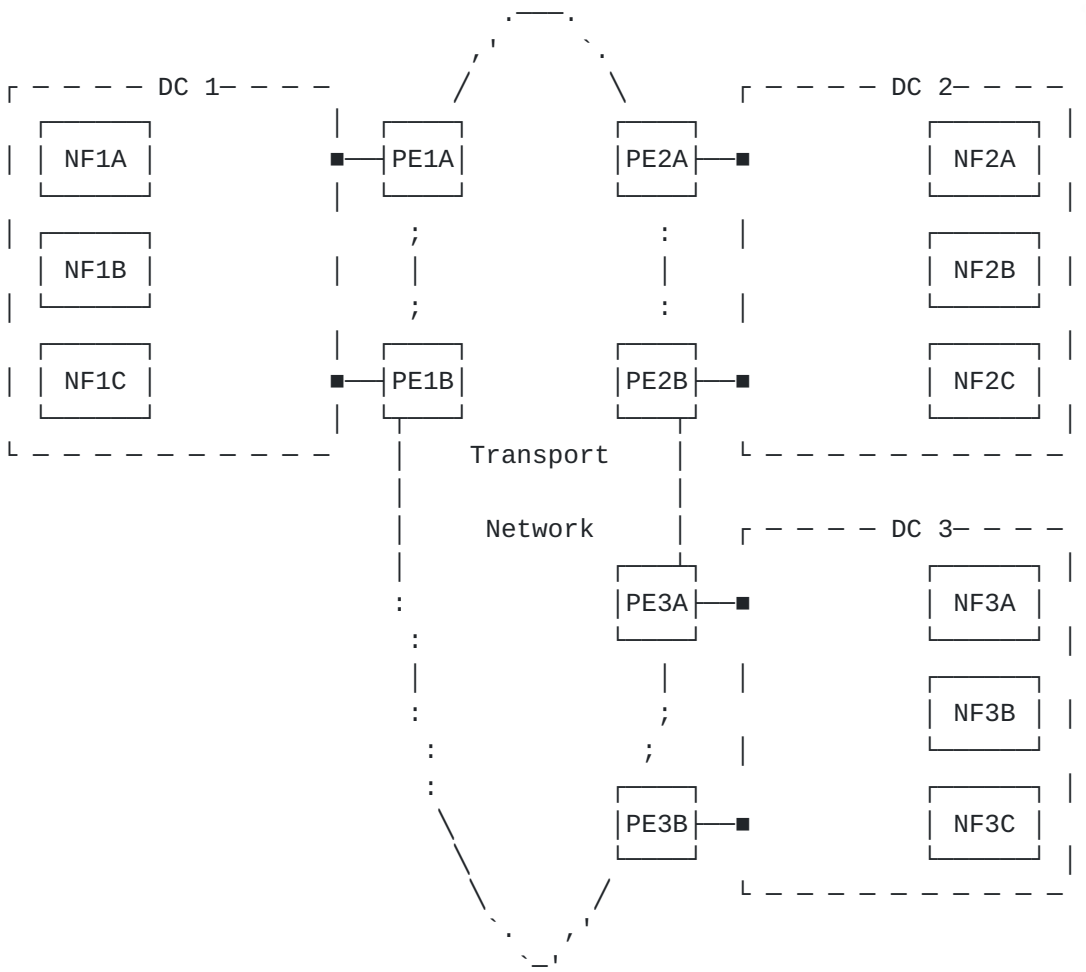
Figure 25: Slice to Transport Plane mapping (5QI-aware model)

6. Capacity Planning/Management

This section describes the information conveyed by the SMO to the transport controller with respect to slice bandwidth requirements.

Figure 26 shows three DCs that contain instances of network functions. Also shown are PEs that have links to the DCs. The PEs belong to the transport network. Other details of the transport network, such as P-routers and transit links are not shown. Also details of the DC infrastructure such as switches and routers are not shown.

The SMO is aware of the existence of the network functions and their locations. However, it is not aware of the details of the transport network. The transport controller has the opposite view - it is aware of the transport infrastructure and the links between the PEs and the DCs, but is not aware of the individual network functions.



■ - SDP, with fine-grained QoS (dedicated resources per IETF NS)

Figure 26: Multi-DC architecture

Let us consider 5G Slice X that uses some of the network functions in the three DCs. If the slice has latency requirements, the SMO will have taken those into account when deciding which network functions in which DC would participate in the slice. As a result of

that placement decision, the three DCs shown are involved in 5G Slice X, rather than other DCs. In order to make this determination, the SMO needs information from the NSC about the latency between DCs. Preferably, the NSC would present the topology in an abstracted form, consisting of point-to-point abstracted links between pairs of DCs and associated latency and optionally delay variation and link loss values. It would be valuable to have a mechanism for the SMO to inform the NSC which DC-pairs are of interest for these metrics - there may be of order thousands of DCs, but the SMO will only be interested in these metrics for a small fraction of all the possible DC-pairs, i.e. those in the same region of the network. The mechanism for conveying the information will be discussed in a future version of this document.

Figure 27 shows the matrix of bandwidth demands for 5G slice X. Within the slice, multiple network function instances might be sending traffic from DCi to DCj. However, the SMO sums the associated demands into one value. For example, NF1A and NF1B in DC1 might be sending traffic to multiple NFs in DC2, but this is expressed as one value in the traffic matrix: the total bandwidth required for 5G Slice X from DC1 to DC2 (8 units). Each row in the right-most column in the traffic matrix shows the total amount of traffic going from a given DC into the transport network, regardless of the destination DC. Note that this number can be less than the sum of DC-to-DC demands in the same row, on the basis that not all the network functions are likely to be sending at their maximum rate simultaneously. For example, the total traffic from DC1 for Slice X is 11 units, which is less than the sum of the DC-to-DC demands in the same row (13 units). Note, as described in Section 4, a slice may have per-QoS class bandwidth requirements, and may have CIR and PIR limits. This is not included in the example, but the same principles apply in such cases.

From \ To	DC 1	DC 2	DC 3	Total from DC
DC 1	n/a	8	5	11.0
DC 2	1	n/a	2	2.5
DC 3	4	7	n/a	10.0

Slice X

From \ To	DC 1	DC 2	DC 3	Total from DC
DC 1	n/a	4	2.5	6.0
DC 2	0.5	n/a	0.8	1.0
DC 3	2.6	3	n/a	5.1

Slice Y

Figure 27: Inter-DC traffic demand matrix

[I-D.ietf-teas-ietf-network-slice-nbi-yang] can be used to convey all of the information in the traffic matrix to the IETF NSC. The IETF NSC applies policers corresponding to the last column in the traffic matrix to the appropriate PE routers, in order to enforce the bandwidth contract. For example, it applies a policer of 11 units to PE1A and PE1B that face DC1, as this is the total bandwidth that DC1 sends into the transport network corresponding to Slice X. Also, the controller may apply shapers in the direction from the TN to the DC, if otherwise there is the possibility of a link in the DC being oversubscribed. Note that a peer NF endpoint of an AC can be identified using 'peer-sap-id' as defined in [I-D.ietf-opsawg-sap].

Depending on the bandwidth model used in the network (Section 6.1), the other values in the matrix, i.e., the DC-to-DC demands, may not be directly applied to the transport network. Even so, the information may be useful to the IETF NSC for capacity planning and failure simulation purposes. If, on the other hand, the DC-to-DC demand information is not used by the IETF NSC, the IETF YANG Data Model for L3VPN Service Delivery [RFC8299] or the IETF YANG Data Model for L2VPN Service Delivery [RFC8466] could be used instead of [I-D.ietf-teas-ietf-network-slice-nbi-yang], as they support conveying the bandwidth information in the right-most column of the traffic matrix.

The transport network may be implemented in such a way that it has various types of paths, for example low-latency traffic might be mapped onto a different transport path to other traffic (for example a particular flex-algo or a particular set of TE LSPs), as discussed in Section 5. The SMO can use [I-D.ietf-teas-ietf-network-slice-nbi-yang] to request low-latency transport for a given slice if required. However, [RFC8299] or [RFC8466] do not support requesting a particular transport-type, e.g., low-latency. One option is to augment these models to convey this information. This can be achieved by reusing the underlay-transport construct used in [RFC9182] and [RFC9291].

6.1. Bandwidth models

This section describes three bandwidth management schemes that could be employed in the transport network. Many variations are possible, but each example describes the salient points of the corresponding scheme. Schemes 2 and 3 use TE, other variations on TE are possible as described in [I-D.ietf-teas-rfc3272bis].

6.1.1. Scheme 1: Shortest Path Forwarding

Shortest path forwarding is used according to the IGP metric. Given that some slices are likely to have latency SLOs, the IGP metric on each link can be set to be in proportion to the latency of the link. In this way, all traffic follows the minimum latency path between endpoints.

In Scheme 1, although the operator provides bandwidth guarantees to the slice customers, there is no explicit end-to-end underpinning of the bandwidth SLO, in the form of bandwidth reservations across the transport network. Rather, the expected performance is achieved via capacity planning, based on traffic growth trends and anticipated future demands, in order to ensure that network links are not over-subscribed. This scheme is analogous to that used in many existing business VPN deployments, in that bandwidth guarantees are provided to the customers but are not explicitly underpinned end to end across the transport network.

A variation on the scheme is that Flex-Algo, defined in [I-D.ietf-lsr-flex-algo], is used, for example one Flex-Algo could use latency-based metrics and another Flex-Algo could use the IGP metric. There would be a many-to-one mapping of slices to Flex-Algos.

While Scheme 1 is technically feasible, it is vulnerable to unexpected changes in traffic patterns and/or network element failures resulting in congestion. This is because, unlike Schemes 2

and 3 that employ TE, traffic cannot be diverted from the shortest path.

6.1.2. Scheme 2: TE LSPs with Fixed Bandwidth Reservations

Scheme 2 uses RSVP-TE or SR-TE LSPs with fixed bandwidth reservations. By "fixed", we mean a value that stays constant over time, unless the SMO communicates a change in slice bandwidth requirements, due to the creation or modification of a slice. Note that the "reservations" would be in the mind of the transport controller - it is not necessary (or indeed possible for SR-TE) to reserve bandwidth at the network layer. The bandwidth requirement acts as a constraint whenever the controller (re)computes the path of an LSP. There could be a single mesh of LSPs between endpoints that carry all of the traffic types, or there could be a small handful of meshes, for example one mesh for low-latency traffic that follows the minimum latency path and another mesh for the other traffic that follows the minimum IGP metric path, as described in Section 5. There would be a many-to-one mapping of slices to LSPs.

The bandwidth requirement from DCi to DCj is the sum of the DCi-DCj demands of the individual slices. For example, if only Slice X and Slice Y are present, then the bandwidth requirement from DC1 to DC2 is 12 units (8 units for Slice X and 4 units for Slice Y). When the SMO requests a new slice, the transport controller, in its mind, increments the bandwidth requirement according to the requirements of the new slice. For example, in Figure 26, suppose a new slice is instantiated that needs 0.8 Gbps from DC1 to DC2. The transport controller would increase its notion of the bandwidth requirement from DC1 to DC2 from 12 Gbps to 12.8 Gbps to accommodate the additional expected traffic.

In the example, each DC has two PEs facing it for reasons of resilience. The transport controller needs to determine how to map the DC1 to DC2 bandwidth requirement to bandwidth reservations of TE LSPs from DC1 to DC2. For example, if the routing configuration is arranged such that in the absence of any network failure, traffic from DC1 to DC2 always enters PE1A and goes to PE2A, the controller reserves 12.8 Gbps of bandwidth on the LSP from PE1A to PE2A. If, on the other hand, the routing configuration is arranged such that in the absence of any network failure, traffic from DC1 to DC2 always enters PE1A and is load-balanced across PE2A and PE2B, the controller reserves 6.4 Gbps of bandwidth on the LSP from PE1A to PE2A and 6.4 Gbps of bandwidth on the LSP from PE1A to PE2B. It might be tricky for the transport controller to be aware of all conditions that change the way traffic lands on the various PEs, and therefore know that it needs to change bandwidth reservations of LSPs accordingly. For example, there might be an internal failure within DC1 that causes traffic from DC1 to land on PE1B, rather than

PE1A. The transport controller may not be aware of the failure and therefore may not know that it now needs to apply bandwidth reservations to LSPs from PE1B to PE2A/PE2B.

6.1.3. Scheme 3: TE LSPs without Bandwidth Reservations

Like Scheme 2, Scheme 3 uses RSVP-TE or SR-TE LSPs. There could be a single mesh of LSPs between endpoints that carry all of the traffic types, or there could be a small handful of meshes, for example one mesh for low-latency traffic that follows the minimum latency path and another mesh for the other traffic that follows the minimum IGP metric path, as described in Section 5. There would be a many-to-one mapping of slices to LSPs.

The difference between Scheme 2 and Scheme 3 is that Scheme 3 does not have fixed bandwidth reservations for the LSPs. Instead, actual measured data-plane traffic volumes are used to influence the placement of TE LSPs. One way of achieving this is to use distributed RSVP-TE with auto-bandwidth. Alternatively, the transport controller can use telemetry-driven automatic congestion avoidance. In this approach, when the actual traffic volume in the data plane on given link exceeds a threshold, the controller, knowing how much actual data plane traffic is currently travelling along each RSVP or SR-TE LSP, can tune the paths of one or more LSPs using the link such that they avoid that link.

It would be undesirable to move a minimum-latency LSP rather than another type of LSP in order to ease the congestion, as the new path will typically have a higher latency, if the minimum-latency LSP is currently following the minimum-latency path. This can be avoided by designing the algorithms described in the previous paragraph such that they avoid moving minimum-latency LSPs unless there is no alternative.

7. IANA Considerations

This memo includes no request to IANA.

8. Security Considerations

IETF Network Slices considerations are discussed in Section 6 of [I-D.ietf-teas-ietf-network-slices]. TBC.

9. References

9.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/

RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[RFC8299] Wu, Q., Ed., Litkowski, S., Tomotaki, L., and K. Ogaki, "YANG Data Model for L3VPN Service Delivery", RFC 8299, DOI 10.17487/RFC8299, January 2018, <https://www.rfc-editor.org/info/rfc8299>.

[RFC8466] Wen, B., Fioccola, G., Ed., Xie, C., and L. Jalil, "A YANG Data Model for Layer 2 Virtual Private Network (L2VPN) Service Delivery", RFC 8466, DOI 10.17487/RFC8466, October 2018, <https://www.rfc-editor.org/info/rfc8466>.

9.2. Informative References

[RFC2698] Heinanen, J. and R. Guerin, "A Two Rate Three Color Marker", RFC 2698, DOI 10.17487/RFC2698, September 1999, <https://www.rfc-editor.org/info/rfc2698>.

[RFC4115] Aboul-Magd, O. and S. Rabie, "A Differentiated Service Two-Rate, Three-Color Marker with Efficient Handling of in-Profile Traffic", RFC 4115, DOI 10.17487/RFC4115, July 2005, <https://www.rfc-editor.org/info/rfc4115>.

[RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, DOI 10.17487/RFC4364, February 2006, <https://www.rfc-editor.org/info/rfc4364>.

[RFC6459] Korhonen, J., Ed., Soininen, J., Patil, B., Savolainen, T., Bajko, G., and K. Iisakkila, "IPv6 in 3rd Generation Partnership Project (3GPP) Evolved Packet System (EPS)", RFC 6459, DOI 10.17487/RFC6459, January 2012, <https://www.rfc-editor.org/info/rfc6459>.

[RFC7806] Baker, F. and R. Pan, "On Queuing, Marking, and Dropping", RFC 7806, DOI 10.17487/RFC7806, April 2016, <https://www.rfc-editor.org/info/rfc7806>.

[RFC9182] Barguil, S., Gonzalez de Dios, O., Ed., Boucadair, M., Ed., Munoz, L., and A. Aguado, "A YANG Network Data Model for Layer 3 VPNs", RFC 9182, DOI 10.17487/RFC9182, February 2022, <https://www.rfc-editor.org/info/rfc9182>.

[RFC9291] Boucadair, M., Ed., Gonzalez de Dios, O., Ed., Barguil, S., and L. Munoz, "A YANG Network Data Model for Layer 2

VPNs", RFC 9291, DOI 10.17487/RFC9291, September 2022, <https://www.rfc-editor.org/info/rfc9291>.

[I-D.ietf-lsr-flex-algo] Psenak, P., Hegde, S., Filsfils, C., Talaulikar, K., and A. Gulko, "IGP Flexible Algorithm", Work in Progress, Internet-Draft, draft-ietf-lsr-flex-algo-26, 17 October 2022, <https://www.ietf.org/archive/id/draft-ietf-lsr-flex-algo-26.txt>.

[I-D.ietf-opsawg-sap] Boucadair, M., de Dios, O. G., Barguil, S., Wu, Q., and V. Lopez, "A YANG Network Model for Service Attachment Points (SAPs)", Work in Progress, Internet-Draft, draft-ietf-opsawg-sap-10, 4 October 2022, <https://www.ietf.org/archive/id/draft-ietf-opsawg-sap-10.txt>.

[I-D.ietf-teas-ietf-network-slices]
Farrel, A., Drake, J., Rokui, R., Homma, S., Makhijani, K., Contreras, L. M., and J. Tantsura, "Framework for IETF Network Slices", Work in Progress, Internet-Draft, draft-ietf-teas-ietf-network-slices-15, 21 October 2022, <https://www.ietf.org/archive/id/draft-ietf-teas-ietf-network-slices-15.txt>.

[I-D.ietf-teas-ietf-network-slice-nbi-yang]
Wu, B., Dhody, D., Rokui, R., Saad, T., Han, L., and J. Mullooly, "IETF Network Slice Service YANG Model", Work in Progress, Internet-Draft, draft-ietf-teas-ietf-network-slice-nbi-yang-03, 24 October 2022, <https://datatracker.ietf.org/api/v1/doc/document/draft-ietf-teas-ietf-network-slice-nbi-yang/>.

[I-D.ietf-teas-ns-ip-mpls]
Saad, T., Beeram, V. P., Dong, J., Wen, B., Ceccarelli, D., Halpern, J., Peng, S., Chen, R., Liu, X., Luis Contreras, M., Rokui, R., and L. Jalil, "Realizing Network Slices in IP/MPLS Networks", Work in Progress, Internet-Draft, draft-ietf-teas-ns-ip-mpls-00, 16 June 2022, <https://www.ietf.org/archive/id/draft-ietf-teas-ns-ip-mpls-00.txt>.

[I-D.ietf-teas-rfc3272bis]
Farrel, A., "Overview and Principles of Internet Traffic Engineering", Work in Progress, Internet-Draft, draft-ietf-teas-rfc3272bis-21, 11 September 2022, <https://

www.ietf.org/archive/id/draft-ietf-teas-rfc3272bis-21.txt>.

[I-D.geng-teas-network-slice-mapping] Geng, X., Dong, J., Pang, R., Han, L., Rokui, R., Jin, J., and J. Tantsura, "5G End-to-end Network Slice Mapping from the view of Transport Network", Work in Progress, Internet-Draft, draft-geng-teas-network-slice-mapping-05, 7 March 2022, <https://www.ietf.org/archive/id/draft-geng-teas-network-slice-mapping-05.txt>.

[I-D.henry-tsvwg-diffserv-to-qci] Henry, J., Szigeti, T., and L. M. C. Murillo, "Diffserv to QCI Mapping", Work in Progress, Internet-Draft, draft-henry-tsvwg-diffserv-to-qci-04, 13 April 2020, <https://www.ietf.org/archive/id/draft-henry-tsvwg-diffserv-to-qci-04.txt>.

[TR-GSTR-TN5G] ITU-T, "Technical Report GSTR-TN5G", 9 February 2018, <https://www.itu.int/dms_pub/itu-t/opb/tut/T-TUT-HOME-2018-PDF-E.pdf>.

[TS-23.501] 3GPP, "TS 23.501: System architecture for the 5G System (5GS)", 2021, <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3144>.

[TS-28.530] 3GPP, "TS 23.530: Management and orchestration; Concepts, use cases and requirements)", 2021, <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3273>.

[O-RAN.WG9.XPSAAS] O-RAN Alliance, "O-RAN.WG9.XPSAAS: O-RAN WG9 Xhaul Packet Switched Architectures and Solutions Version 02.00", 1 July 2021, <https://www.o-ran.org/specifications>.

[NG.113] GSMA, "NG.113: 5GS Roaming Guidelines Version 4.0", 28 May 2021, <https://www.gsma.com/newsroom/wp-content/uploads//NG.113-v4.0.pdf>.

Appendix A. Acronyms and Abbreviations

3GPP: 3rd Generation Partnership Project

5GC: 5G Core

5QI: 5G QoS Indicator

A2A: Any-to-Any

AC: Attachment Circuit

AMF: Access and Mobility Management Function

AUSF: Authentication Server Function

BBU: Baseband Unit

BH: Backhaul

BS: Base Station

CE: Customer Edge

CIR: Committed Information Rate

CN: Core Network

CoS: Class of Service

CP: Control Plane

CSP: Communication Service Provider

CU: Centralized Unit

CU-CP: Centralized Unit Control Plane

CU-UP: Centralized Unit User Plane

DC: Data Center

DDoS: Distributed Denial of Services

DN: Data Network

DSCP: Differentiated Services Code Point

DU: Distributed Unit

eCPRI: enhanced Common Public Radio Interface

FH: Fronthaul

FIB: Forwarding Information Base

GPRS: Generic Packet Radio Service

gNB: gNodeB

GTP: GPRS Tunneling Protocol

GTP-U: GPRS Tunneling Protocol User plane

HW: Hardware

ID: Identifier

IGP: Interior Gateway Protocol

IP: Internet Protocol

L2VPN: Layer 2 Virtual Private Network

L3VPN: Layer 3 Virtual Private Network

LSP: Label Switched Path

MH: Midhaul

MIoT: Massive Internet of Things

MPLS: Multiprotocol Label Switching

NF: Network Function

NR: New Radio

NRF: Network Function Repository

NRP: Network Resource Partition

NSC: Network Slice Controller

NSS: Network Slice Subnet

PE: Provider Edge

PIR: Peak Information Rate

PLMN: Public Land Mobile Network

PSTN: Public Switched Telephony Network

QoS: Quality of Service

RAN: Radio Access Network

RF: Radio Frequency

RIB: Routing Information Base

RSVP: Resource Reservation Protocol

RU: Radio Unit

SD: Slice Differentiator

SDP: Service Demarcation Point

SLA: Service Level Agreement

SLO: Service Level Objective

SMF: Session Management Function

SMO: Service Management and Orchestration

S-NSSAI: Single Network Slice Selection Assistance Information

SST: Slice/Service Type

SR: Segment Routing

SRv6: Segment Routing version 6

TC: Traffic Class

TE: Traffic Engineering

TN: Transport Network

TS: Technical Specification

UDM: Unified Data Management

UE: User Equipment

UP: User Plane

UPF: User Plane Function

URLLC: Ultra Reliable Low Latency Communication

VLAN: Virtual Local Area Network

VNF: Virtual Network Function

VPN: Virtual Private Network

VRF: Virtual Routing and Forwarding

VXLAN: Virtual Extensible Local Area Network

Appendix B. An overview of 5G Networking

This section provides a brief introduction to 5G mobile networking with a perspective on the Transport Network. This section does not intend to replace or define 3GPP architecure, it just provides a brief overview for readers that do not have a mobile background. For more comprehensive information, refer to [TS-23.501].

B.1. Building Blocks

[TS-23.501] defines the Network Functions (UPF, AMF, etc.) that compose the 5G System (5GS) Architecture together with related interfaces (e.g., N1, N2...). This architecture has native Control and User Plane separation, and the Control Plane leverages a service-based architecture. Figure 28 outlines an example 5GS architecture with a subset of possible network functions and network interfaces.

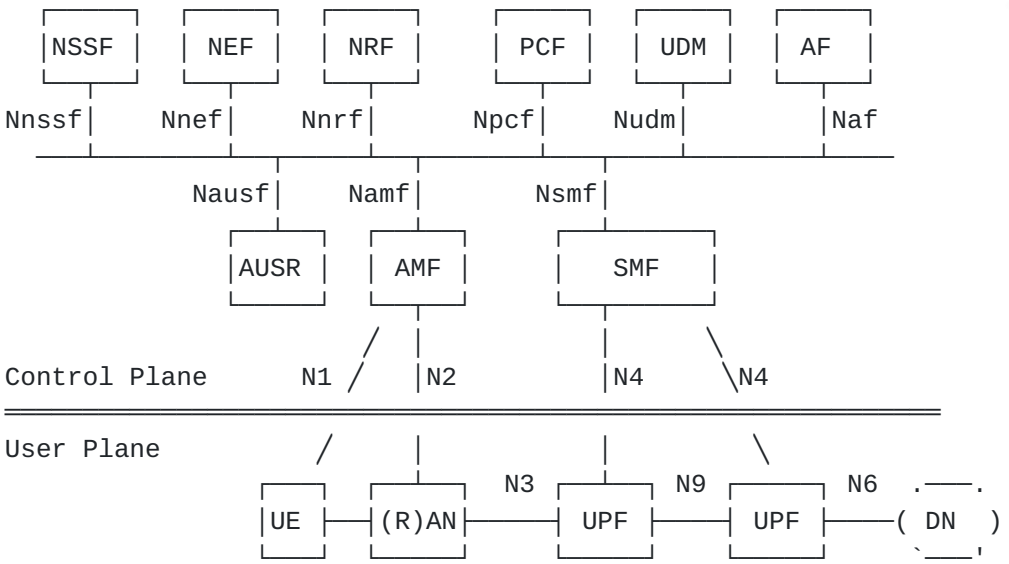


Figure 28: 5GS Architecture and Service-based Interfaces

Similar to previous versions [RFC6459], a 5G mobile network is split into 4 major domains:

*UE, MS, MN, and Mobile

The terms UE (User Equipment), MS (Mobile Station), MN (Mobile Node), and mobile refer to the devices that are hosts with the ability to obtain Internet connectivity via a 3GPP network. An MS is comprised of the Terminal Equipment (TE) and a Mobile Terminal (MT). The terms UE, MS, MN, and mobile are used interchangeably within this document.

*Radio Access Network (RAN)

Provides wireless connectivity to the UE devices via radio. It is made up of the Antenna that transmits and receives signals to the UE and the Base Station that digitizes the signal and converts the RF data stream to IP packets.

*Core Network (CN)

Controls the CP of the RAN and provides connectivity to the Data Network (e.g., the Internet or a private VPN). The Core Network hosts dozens of services such as authentication, phone registry, charging, access to PSTN and handover.

*Transport Network (TN)

Provides connectivity between sites where 5G Network Functions are located. The TN may connect sites from the RAN to the Core Network, as well as sites within the RAN or within the CN. This connectivity is achieved by IP Networking.

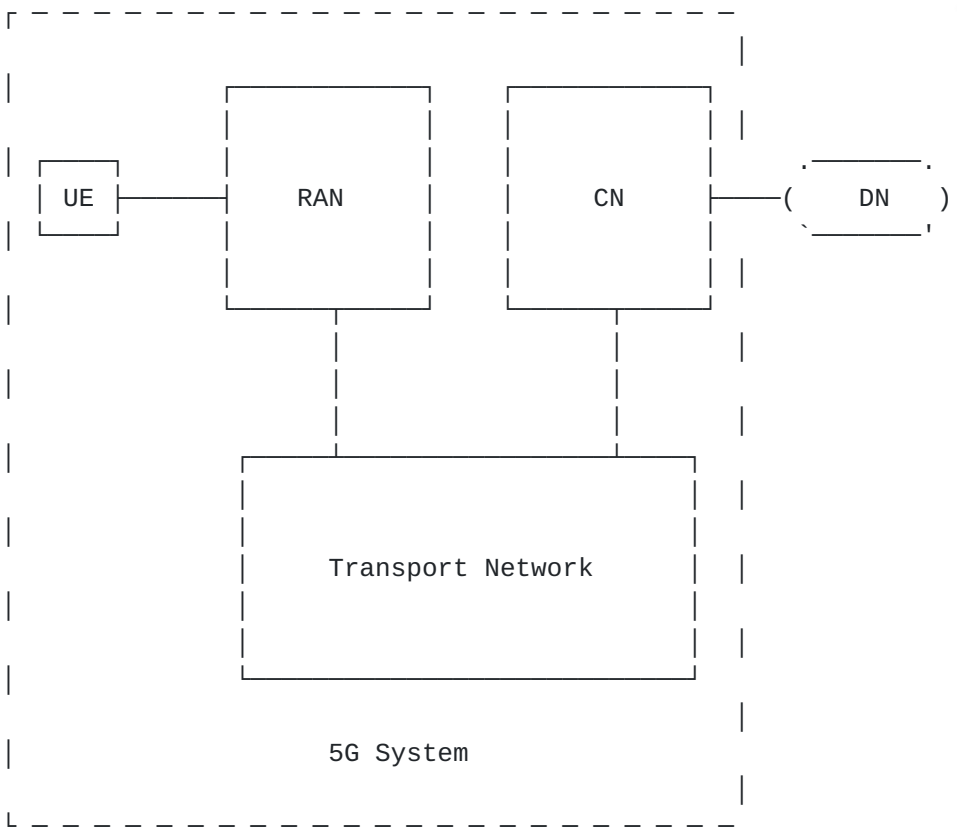


Figure 29: Building blocks of 5G architecture (high-level representation)

B.2. Core Network

The 5G Core Network (5GC) is made up of a set of Network Functions (NFs) which fall into two main categories:

- 5GC User Plane: the User Plane Function (UPF) is the interconnect point between the mobile infrastructure and the Data Network (DN). It interfaces with the RAN via the N3 interface by encapsulating/decapsulating the User Plane Traffic in GTP Tunnels (aka GTP-U or Mobile User Plane).

- 5GC Control Plane: the 5G Control Plane is made up of a comprehensive set of Network Functions. An exhaustive list and description of these entities is out of the scope of this document. The following NFs and interfaces are worth mentioning, since their connectivity may rely on the Transport Network:

  - the AMF (Access and Mobility Function) connects with the RAN control plane over the N2 interface

  - the SMF controls the 5GC UPF via the N4 interface

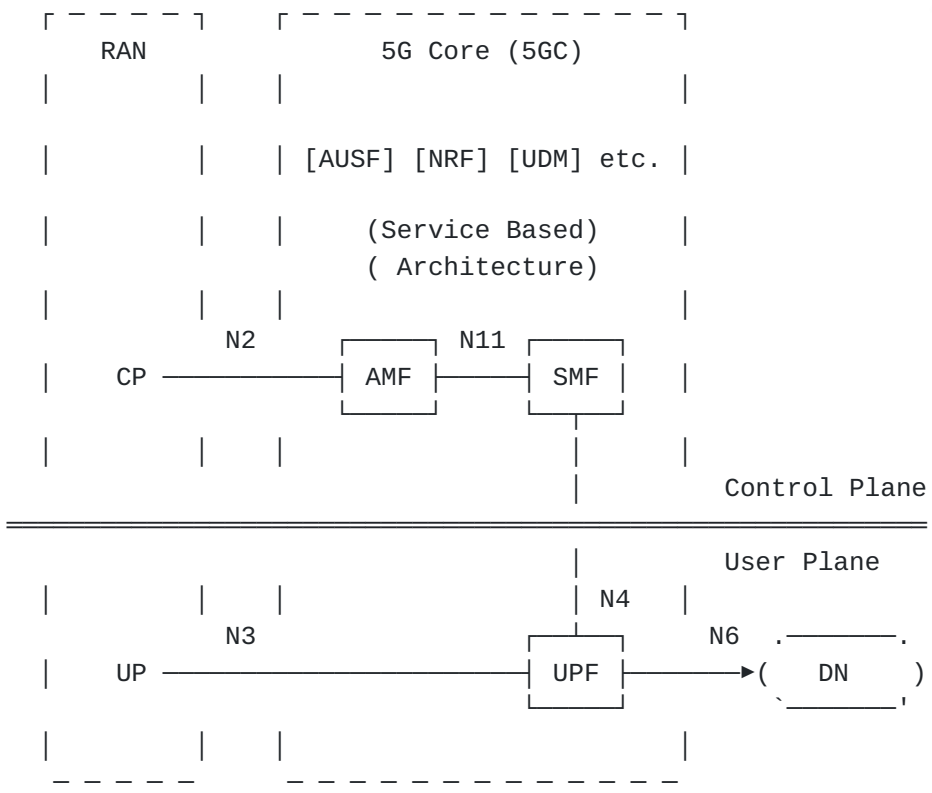


Figure 30: 5G Core Network (CN)

B.3. RAN

The radio access network (RAN) connects cellular wireless devices to a mobile Core Network. The RAN network is made up of 3 components, which form the Radio Base Station:

*The Baseband Unit (BBU) provides the interface between the Core Network and the Radio Network. It connects to the Radio Unit and is responsible for the baseband signal processing to packet.

*The Radio Unit (RU) is located close to the Antenna and controlled by the BBU. It converts the Baseband signal received from the BBU to a Radio frequency signal.

*The Antenna converts the electric signal received from the RU to radio waves

The 5G RAN Base Station is called a gNodeB (gNB). It connects to the Core Network via the N3 (user plane) and N2 (control plane) interfaces.

The 5G RAN architecture supports RAN disaggregation in various ways. Notably, the BBU can be split into a DU (Distributed Unit) for digital signal processing and a CU (Centralized Unit) for RAN Layer 3 processing. Furthermore, the CU can be itself split into Control Plane (CU-CP) and User Plane (CU-UP).

Figure 31 depicts a disaggregated RAN with NFs and interfaces.

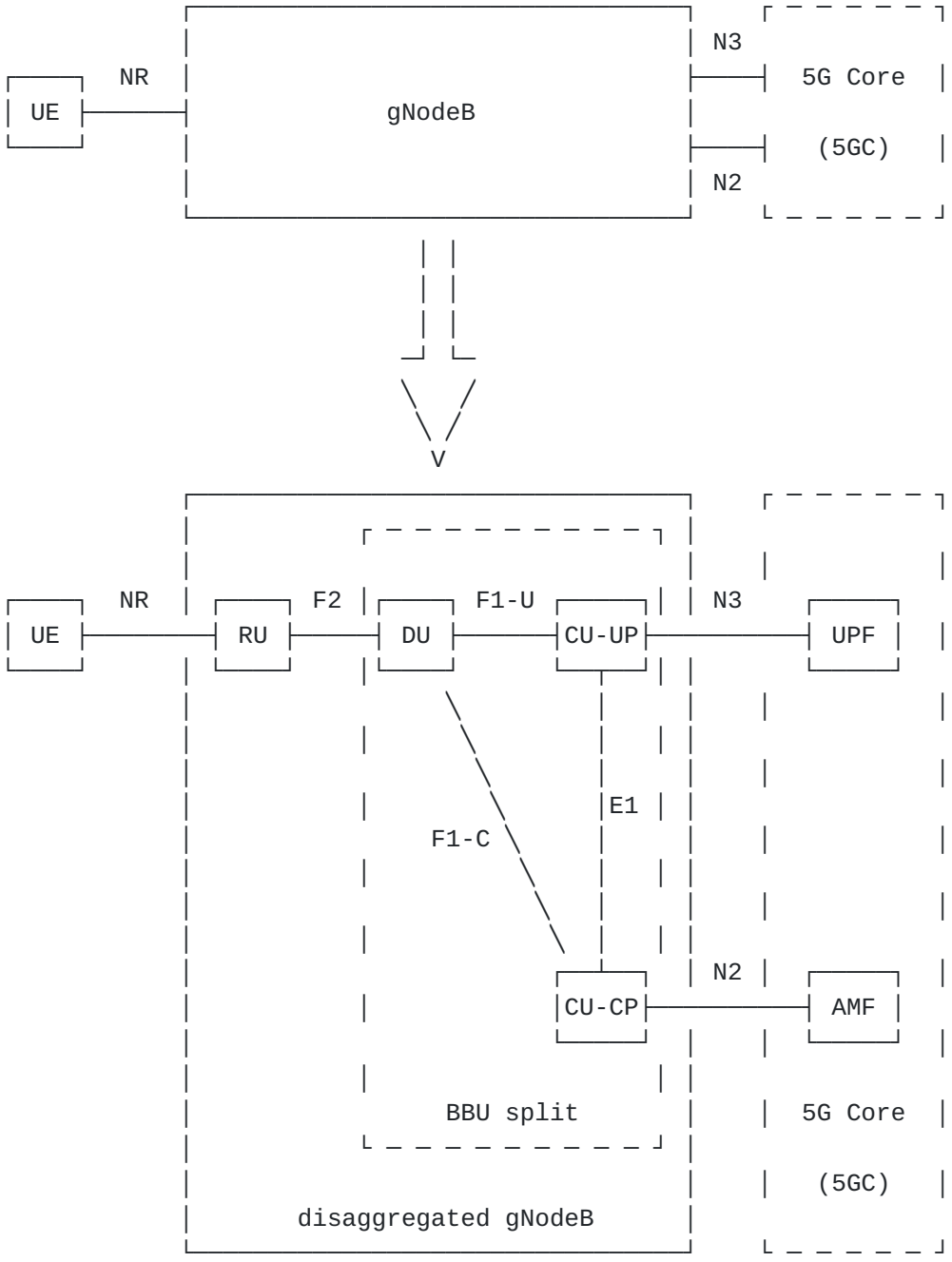


Figure 31: RAN Disaggregation

B.4. Transport Network

5G TN segments

The 5G transport architecture defines three main segments for the Transport Network, which are commonly referred to as Fronthaul (FH), Midhaul (MH), and Backhaul (BH) ([TR-GSTR-TN5G]).

*Fronthaul happens before the BBU processing. In 5G, this interface is based on eCPRI (Enhanced CPRI) with native Ethernet or IP encapsulation.

*Midhaul is optional: this segment is introduced in the BBU split presented in Appendix B.3, where Midhaul network refers to the DU-CU interconnection (i.e., F1 interface). At this level, all traffic is encapsulated in IP (signaling and user plane).

*Backhaul happens after BBU processing. Therefore, it maps to the interconnection between the RAN and the Core Network. All traffic is also encapsulated in IP.

Figure 32 illustrates the different segments of the Transport Network with the relevant Network Functions.

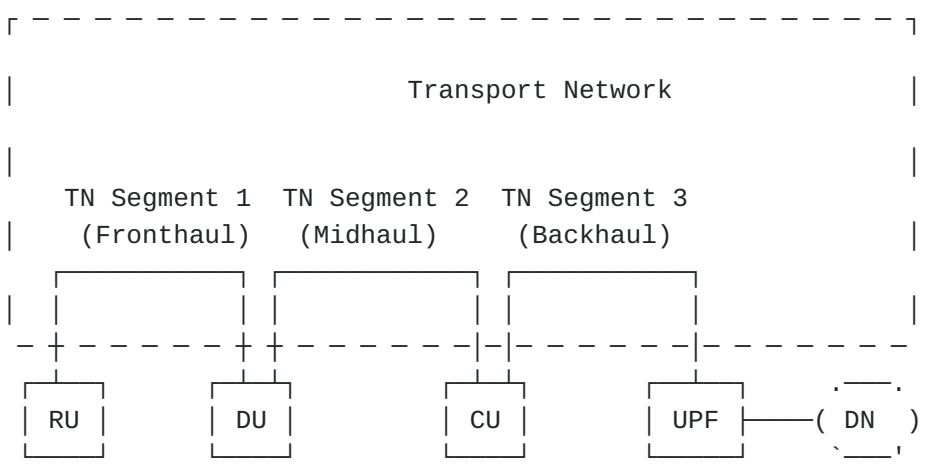


Figure 32: 5G Transport Segments

It is worth mentioning that a given part of the transport network can carry several 5G transport segments concurrently, as outlined in Figure 33. This is because different types of 5G network functions might be placed in the same location (e.g., the UPF from one slice might be placed in the same location as the CU-UP from another slice).

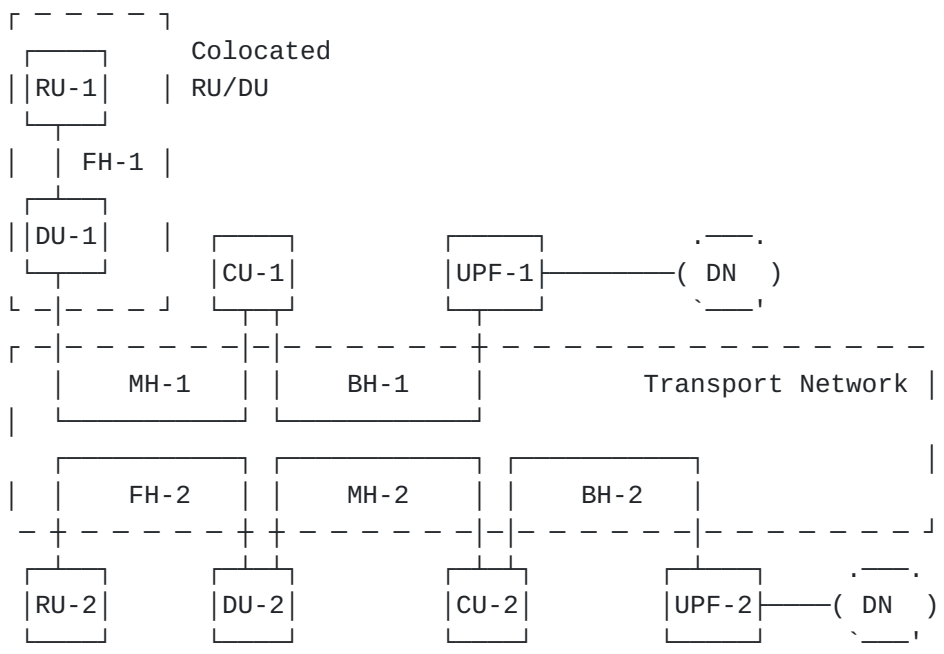


Figure 33: Concurrent 5G Transport Segments

Acknowledgements

The authors would like to thank Adrian Farrel, Joel Halpern and Tarek Saad for their reviews of this document and for providing valuable feedback on it.

Contributors

To be added later

Authors' Addresses

Krzysztof G. Szarkowicz (editor)
Juniper Networks
Wien
Austria

Email: kszarkowicz@juniper.net

Richard Roberts
Juniper Networks
Rennes
France

Email: rroberts@juniper.net

Julian Lucek
Juniper Networks
London

United Kingdom

Email: jlucek@juniper.net

John E. Drake
Juniper Networks
Sunnyvale, CA
United States of America

Email: jdrake@juniper.net

Mohamed Boucadair
Orange
Rennes
France

Email: mohamed.boucadair@orange.com

Luis M. Contreras
Telefonica
Ronda de la Comunicacion, s/n
28050 Madrid
Spain

Email: luismiguel.contrerasmurillo@telefonica.com
URI: http://lmcontreras.com/

Ivan Bykov
Ribbon Communications
Tel Aviv
Israel

Email: ivan.bykov@rbbn.com

Reza Rokui
Ciena
Ottawa
Canada

Email: rrokui@ciena.com

Luay Jalil
Verizon
Dallas, TX
United States of America

Email: luay.jalil@verizon.com

Beny Dwi Setyawan
XL Axiata

Jakarta
Indonesia

Email: benyds@xl.co.id