7450 Ethernet Service Switch
7750 Service Router
Virtualized Service Router

Release 24.3.R1

Triple Play Service Delivery Architecture Guide

3HE 20117 AAAA TQZZA 01
Edition: 01
March 2024

Nokia is committed to diversity and inclusion. We are continuously reviewing our customer documentation and consulting with standards bodies to ensure that terminology is inclusive and aligned with the industry. Our future customer documentation will be updated accordingly.

Table of contents

1 Getting started

1.1 About this guide

This guide describes details pertaining to Triple Play Services Delivery Architecture (TPSDA) support provided by the operating system and presents examples to configure and implement various protocols and services.

This guide is organized into functional chapters and provides concepts and descriptions of the implementation flow, as well as Command Line Interface (CLI) syntax and command usage.

Note: Unless otherwise indicated, this guide uses classic CLI command syntax and configuration examples.

The topics and commands described in this document apply to the:

- 7450 ESS
- 7750 SR
- Virtualized Service Router

For a list of unsupported features by platform and chassis, see the *SR OS R24.x.Rx Software Release Notes*, part number 3HE 20152 000*x* TQZZA.

Command outputs shown in this guide are examples only; actual displays may differ depending on supported functionality and user configuration.

Note: The SR OS CLI trees and command descriptions can be found in the following guides:

- *7450 ESS, 7750 SR, 7950 XRS, and VSR Classic CLI Command Reference Guide*
- *7450 ESS, 7750 SR, 7950 XRS, and VSR Clear, Monitor, Show, and Tools CLI Command Reference Guide* (for both MD-CLI and Classic CLI)
- *7450 ESS, 7750 SR, 7950 XRS, and VSR MD-CLI Command Reference Guide*

Note: This guide generically covers Release 24.*x*.R*x* content and may contain some content to be released in later maintenance loads. See the *SR OS R24.x.Rx Software Release Notes*, part number 3HE 20152 000*x* TQZZA, for information about features supported in each load of the Release 24.*x*.R*x* software.

1.2 Conventions

This section describes the general conventions used in this guide.

1.2.1 Precautionary and information messages

The following information symbols are used in the documentation.

DANGER: Danger warns that the described activity or situation may result in serious personal injury or death. An electric shock hazard could exist. Before you begin work on this equipment, be aware of hazards involving electrical circuitry, be familiar with networking environments, and implement accident prevention procedures.

WARNING: Warning indicates that the described activity or situation may, or will, cause equipment damage, serious performance problems, or loss of data.

Caution: Caution indicates that the described activity or situation may reduce your component or system performance.

Note: Note provides additional operational information.

Tip: Tip provides suggestions for use or best practices.

1.2.2 Options or substeps in procedures and sequential workflows

Options in a procedure or a sequential workflow are indicated by a bulleted list. In the following example, at step 1, the user must perform the described action. At step 2, the user must perform one of the listed options to complete the step.

Example: Options in a procedure

1. User must perform this step.
2. This step offers three options. User must perform one option to complete this step.
   - This is one option.
   - This is another option.
   - This is yet another option.

Substeps in a procedure or a sequential workflow are indicated by letters. In the following example, at step 1, the user must perform the described action. At step 2, the user must perform two substeps (a. and b.) to complete the step.

Example: Substeps in a procedure

1. User must perform this step.
2. User must perform all substeps to complete this action.
   a. This is one substep.
   b. This is another substep.

2 Subscriber management

In the late 1990s, cable operators (Multiple System Operators [MSOs]) took advantage of the extra spectrum available on their shared coaxial copper cable to provide high-speed data service to their customers using the DOCSIS standard. This forced telephone companies (Telcos) to start adopting new technologies such as xDSL, which leveraged narrowband residential loops to provide more bandwidth to its subscribers than previously possible with modems. The wiring segment connecting the subscribers to the first active network element (coaxial cables in MSOs and unshielded twisted pair in Telcos) were commonly referred to as "first miles lines". Although the wiring technology has changed over time (for example, fiber at Telcos), this term remains in use today as a reference point.

The technological advancements in the first mile started the era of "broadband access", a term that was adopted to describe high-speed access to residential and small business subscribers. Broadband access enabled MSOs and Telcos, commonly referred to "Service Providers", to start offering voice, data and video service, also known as triple play services, over the existing first mile lines. Modern networks now offer a variety of high-bandwidth applications, including multimedia. These services are commonly referred to as "broadband services", distinctive from the "business services" term, which implies Layer 2 or Layer 3[1] type of connectivity, such as VLL, VPLS, or VPRN offered to the subscribers over the same broadband access.

Before broadband adaptation, Telcos focused on analog and digitalized Pulse Code Modulation (PCM) voice service over dedicated twisted coper lines, while MSOs focused on broadband broadcast traffic (video) using shared-access coaxial cable lines. This led to two different network build-out models. Although Nokia equipment can serve both types of operators, this guide is focused on the Telco architecture. In this guide, the terms "operator" and "service provider" refer to a Telco operator unless noted otherwise.

Over time, the access technology has evolved to include other physical mediums in the first mile, such as fiber (active or passive), and has even branched into wireless space.

Ethernet has become the Layer 2 technology of choice in the first mile for wireline operators because of its simplicity and low cost, effectively replacing other Layer 2 technologies such as TDM lines, ATM, frame-relay, and PPP (although for historical reasons, PPP is still in use over Ethernet – PPPoE). Ethernet in the first mile is described in IEEE 802.3ah standard (EFM), where Ethernet can run over copper, point-to-point fiber, or Passive Optical Network. With Ethernet in the first mile, it was easy to integrate subscriber access into the existing metro aggregation networks utilizing various Layer 2 or Layer 3 overlay technologies, such as VLL, VPLS, or VPRNs, without the complex conversion between the protocols in the first mile.

The resulting setup in the first mile for wireline access in modern networks include:

- IP at Layer 3
- Ethernet/PPPoE at Layer 2
- variety of physical mediums (twisted pair, fiber, coax)

Subscriber access to the network and services is described under the term "Subscriber Management". Subscriber Management is performed in network nodes called Broadband Network Gateways (BNGs) and it provides functions such as:

[1] Layer 1/Layer 2/Layer 3 refer to the physical, data link and network layer of the Open System Interconnection (OSI) model.

- security
- authentication
- billing and accounting
- Hierarchical Quality of Service
- policy management
- IP address management

Nokia BNG supports subscriber management in both Layer 2 and Layer 3 environments and can serve a multipurpose role, including providing access to business services and other value-added services such as NAT or Application-Assurance.

Subscriber management can be configured in a variety of ways, but it is critical that subscriber management integrates seamlessly with element and service management across the broadband infrastructure by, for example, the Nokia Network Services Platform (NSP). Subscriber management can also be implemented through CLI or scripted commands at the platform level, whereby a network administrator would manually configure the set of QoS, security, AAA, or anti-spoofing functions that relate to a specific subscriber.

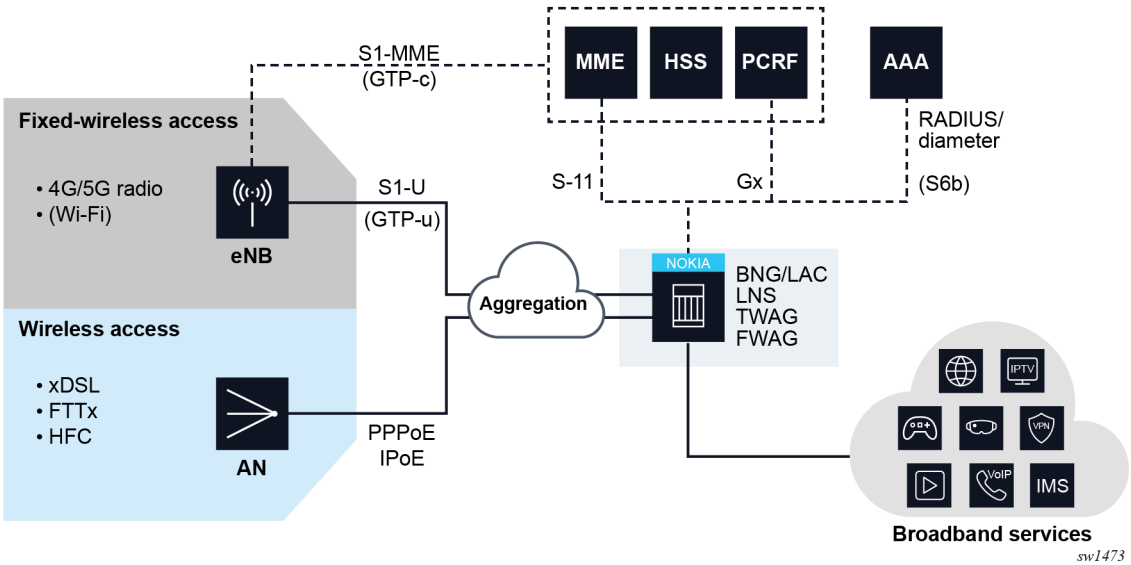
In addition to wireline subscribers, Nokia BNG also supports subscriber management for wireless users:

- subscribers connected to cellular network (4G, 5G radio) are stationary subscribers (wireless home gateways), part of Fixed Wireless Access
- wireless subscribers connected to the network via trusted wireless LAN where mobility between the Wi-Fi access points is supported.

These subscribers should not be confused with Wi-Fi devices behind the wireline subscribers, such as residential gateways.

Over the years, the BNG has evolved from copper and fiber-based wireline access to a versatile multi-access gateway (MAG) that can aggregate a wide range of wireline and wireless access technologies.

Figure 1: Nokia MAG



Market demands for scalability and cost reduction necessitated further architectural changes where the BNG control plane has been moved into the cloud. For more information about such disaggregated BNG, see "BBF TR-459: Control and User Plane Separation for a disaggregated BNG" in the https://www.broadband-forum.org/technical-reports.

2.1 Network design considerations – from customer premise to the central office

Network design, particularly from a customer premise to the central office, entails many key considerations to ensure reliable, efficient, and adaptable services. These considerations encompass network architecture, customer premises equipment, access methods, potential aggregation network, and the BNG design and location.

There are several additional considerations that could help to ensure the network is robust, efficient, and prepared for future needs, including the following:

- **network scalability**

  Design the network to allow for easy scaling. As the number of users or the demand for data grows, you should be able to add capacity without major redesigns or disruption.

- **service quality**

  Different customers and applications may have different Quality of Service (QoS) requirements. Your network design should allow for QoS differentiation and guarantees.

- **security considerations**

  Protecting the network from malicious activities is crucial. This might include strategies for preventing, detecting, and responding to attacks.

- **network automation and orchestration**

  As networks grow more complex, automation becomes more important. This can include automated provisioning, configuration, and fault management.

- **multiservice support**

  The network should be capable of supporting multiple services (voice, data, video, and so on) and types of traffic (unicast, multicast, broadcast, and so on).

- **power requirements**

  For hardware installed at customer premises or in the field, consider the power requirements and availability.

- **operational costs**

  Consider the operational costs such as maintenance, power, cooling, and network management. These factors could impact the total cost of ownership.

- **vendor interoperability**

  Make sure that the equipment from different vendors can work together without compatibility issues.

- **regulatory and legal compliance**

  Depending on the jurisdiction, there might be legal requirements to consider, such as privacy regulations.

- **emerging technologies**

Be aware of emerging technologies and industry trends, such as 5G, IoT, edge computing, and so on. A forward-thinking network design can better adapt to these technologies when they become more prevalent.

- **current and future needs**

  A network design project should start with a clear understanding of the current requirements and future needs, and then consider these and other factors to create a network that is resilient, efficient, and adaptable.

2.1.1 Centralized versus distributed architecture

A fundamental consideration is whether to implement a centralized or distributed network architecture. Centralized models pool resources into a single location, simplifying management but potentially create a single point of failure. On the other hand, distributed networks decentralize resources, enhancing resiliency, but at the cost of increased management complexity. Geographical redundancy (geo-red) is another significant consideration, allowing service continuity even if a specific geographical area experiences a failure.

2.1.2 Customer premise

The customer premise may include a variety of devices, including routed and bridged Residential Gateways (RGs), that may employ Network Address Translation (NAT). Understanding the needs of different customer types (residential or business) further impacts network design, as some types of customers might require more advanced features or higher service levels. Virtual Residential Gateway (vRGW), service per Virtual Local Area Network (VLAN), subscriber per VLAN models, or even a flat (untagged) network might be used depending on the specific requirements.

Deciding on a service per VLAN, subscriber per VLAN, or flat network model can significantly impact the network design, scalability, and management. Each model can be influenced by the following network factors:

- **service per VLAN**

  This model associates different services with distinct VLANs. For example, Internet access, VoIP, and IPTV could each be on a separate VLAN. This design allows for easier QoS management, as different services typically have different QoS requirements.

- **subscriber per VLAN**

  Each subscriber is assigned to their unique VLAN. This model can provide high granularity for managing services and security since traffic from each subscriber is isolated from the others. However, this model significantly increases the number of VLANs.

- **flat network (single VLAN or untagged)**

  In this model, all subscribers share a single VLAN or operate in an untagged network. This simplifies the network design and management because you do not have to manage multiple VLANs. However, it offers less granularity for service and security management, and it can create larger broadcast domains, which might impact performance.

When deciding between these models, consider the following factors:

- **scalability**

  How many subscribers or services does the network need to support, now and in the future?

- **service differentiation**

  Will you need to provide different QoS levels for different services or subscribers?
- **security**

  How important is it to isolate traffic from different subscribers or services?
- **management complexity**

  How equipped are you to manage a potentially large number of VLANs?
- **performance**

  How does the chosen model impact network performance?

2.1.3 Access

Network access methods can be wired, wireless, or hybrid access solutions that mix fixed and wireless access.

Physical access technologies include the following:

- **fiber optics**

  This includes Fiber to the Home (FTTH), Fiber to the Building (FTTB), Fiber to the Curb (FTTC), and Fiber to the Node (FTTN). Fiber optics offer high bandwidth, low latency, and resistance to electromagnetic interference.
- **twisted pair**

  This includes Digital Subscriber Line (DSL) technologies such as ADSL, VDSL, and G.fast. These technologies deliver broadband over the existing telephone lines.
- **coaxial cable**

  Cable broadband services, such as DOCSIS, use coaxial cables originally installed for cable TV.
- **wireless**

  Wireless broadband can include Wi-Fi, cellular (4G, 5G), satellite, and fixed wireless access.
- **satellite**

  Satellite broadband can reach areas where terrestrial broadband is not available.

Logical access protocols include:

- **Asynchronous Transfer Mode (ATM)**

  This is often used for DSL connections, but it is less common today because of its lower efficiency for packet data.
- **Ethernet**

  This is commonly used for FTTP and cable services, and for DSL services that use Ethernet over DSL (EoDSL).
- **Passive Optical Network (PON)**

  This is a type of fiber network that uses passive splitters to deliver broadband to multiple premises. Variants include GPON and EPON.
- **Data Over Cable Service Interface Specification (DOCSIS)**

  This is used for delivering broadband over coaxial cable networks.

- **G.fast**

  This is a DSL protocol designed for short distances with high speeds, often used for FTTC or FTTdp (Fiber to the Distribution Point).

- **Point-to-Point Protocol over Ethernet/ATM (PPPoE/PPPoA)**

  These are tunneling protocols often used with DSL services.

The choice of access technology and protocol depends on factors such as the existing infrastructure, required bandwidth, service area characteristics, and cost. For example, FTTH with GPON may be the best choice for a new installation in a densely populated area, while VDSL over existing telephone lines may be more cost-effective in an area with lower population density.

2.1.4 Aggregation network

The aggregation network is an optional layer that groups or aggregates multiple customer connections toward the BNG. Without an aggregation network, access nodes connect directly to the BNG. If an aggregation network is used, it could be native Ethernet or an overlay such as VPNs, Virtual Private LAN Service (VPLS), with various transport options such as Multiprotocol Label Switching (MPLS), Segment Routing over IPv6 (SRv6), or Generic Routing Encapsulation (GRE).

The decision to implement an aggregation network or directly connect access nodes to BNGs depends on several factors:

- **scale**

  In a large network with numerous access nodes, an aggregation network can simplify management by reducing the number of direct connections to the BNG. Each BNG can connect to several aggregation switches, which in turn connect to multiple access nodes. However, for a small network, direct connections from the access nodes to the BNG may be simpler and more cost-effective.

- **geographical distribution**

  If access nodes are widely dispersed, it may be more efficient to connect them to a local aggregation switch and then connect the aggregation switches to a centralized BNG. This can reduce long-haul links and save costs.

- **redundancy and resilience**

  An aggregation network can provide additional paths between the access nodes and the BNG, enhancing network resilience. If a direct link to the BNG fails, traffic can be rerouted through another aggregation switch.

- **service management**

  Aggregation networks can also offer service-specific features. For instance, they may support Multiprotocol Label Switching (MPLS) for traffic engineering or Virtual Private LAN Service (VPLS) for Layer 2 VPNs.

- **cost**

  Aggregation networks require additional equipment and potentially more complex management, which can increase costs. However, they can also reduce the number of required BNG ports, which may offset some of these costs.

- **future-proofing**

  An aggregation network can offer more flexibility to accommodate future growth or changes. For example, it can be easier to add, remove, or relocate access nodes without disturbing the BNG.

2.1.5 Subscriber termination

Subscriber terminations occur on the BNG, usually located in the central office. The design could be centralized or distributed. Centralized BNGs host a large number of subscribers, making redundancy critical to prevent service disruption. Distributed BNGs could provide better scalability and resiliency, but may require more sophisticated management and orchestration.

2.1.6 Resiliency

Resilience and redundancy are critical aspects of any network design strategy. How and where to implement redundancy depends on several factors, including the expected reliability, the cost of downtime, and the available budget. Here are some redundancy strategies:

- **redundancy at the customer premise level**

  This scenario involves setting up dual connections from the Customer Premises Equipment (CPE) to the network. These connections may be to the same network device (such as a dual-homed connection to a single access node) or to two different devices (a multihomed connection). The advantage here is that a single failed connection does not disrupt the service of the customer. However, the cost is higher as it requires more equipment.

- **redundancy at the access node level**

  In this scenario, each access node could have redundant connections to the BNG or to the aggregation network. This can provide a higher level of redundancy as it protects against both a connection failure and an access node failure. However, it requires more network resources (more connections and potentially more BNG ports) and it increases complexity, which could impact network performance or troubleshooting.

- **redundancy in the aggregation network or the BNG**

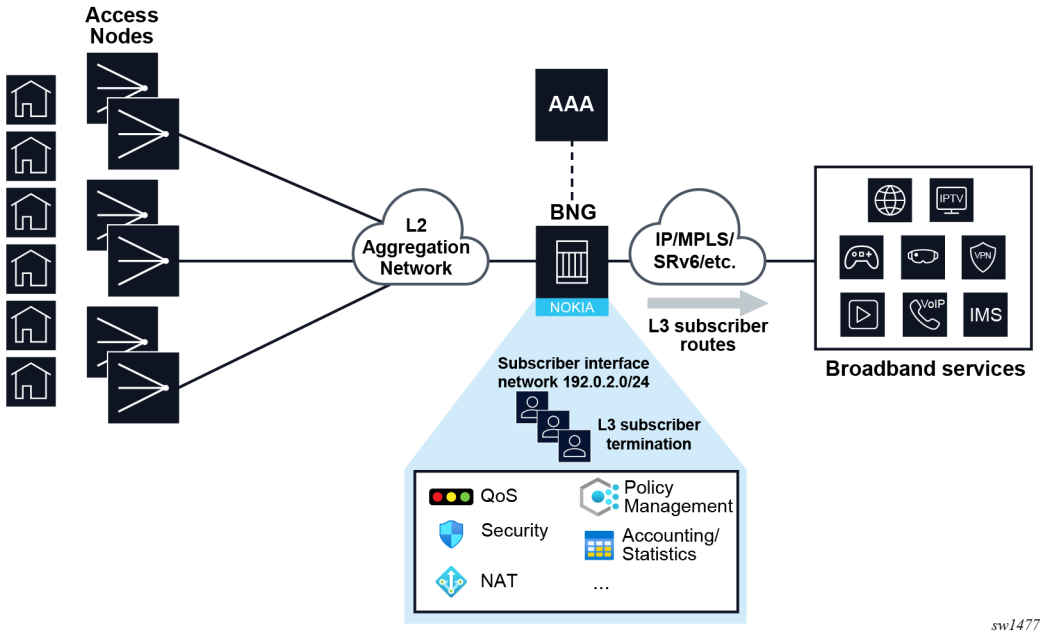
  This scenario could involve redundant connections, redundant devices, or both. For example, the aggregation network could have a redundant path to each BNG, and each BNG could have a redundant connection to the core network. Moreover, for a highly critical BNG, there could be a backup BNG that can take over if the primary BNG fails.

In all cases, network designers must balance the cost and complexity of redundancy against the cost and impact of potential downtime. Redundancy can significantly improve network reliability, but it comes with increased cost and complexity. Therefore, it is essential to understand the business needs and expectations before deciding where and how to implement redundancy.

2.1.7 Subscriber management in routed central office

Subscriber management in the routed central office (RCO) model, shown in Figure 2: RCO, represents one of the most widely adopted approaches, that offers support for various device types such as IPoEv4/v6 and PPPoEv4/v6. In this model, subscribers are typically terminated on the BNG located within a local central office. The use of an aggregation network in this setup is optional; access nodes can be directly connected to the BNG without an aggregation network or connected via such a network.

Figure 2: RCO



Note: The RCO model does not inherently dictate a centralized or distributed architecture. Subscriber management can be implemented within each local central office, aligning with a distributed architecture, or it can be centralized within a hub site that serves multiple central offices, resembling a more centralized architecture.

In RCO models, subscribers are terminated on Layer 3 interfaces. In SR OS, a conventional Layer 3 service interface is associated with a single SAP representing a single VLAN or VLAN pair. To accommodate multiple SAPs, that is multiple VLANs or VLAN pairs, under a single interface, two new interface types are introduced: subscriber and group interfaces. While subscriber interfaces define subscriber IP subnets, group interfaces play a pivotal role in aggregating SAPs. For a detailed explanation of group and subscriber interfaces and their respective roles, see Subscriber interface and Group interface.

Within each IES or VPRN service, users have the flexibility to create multiple subscriber interfaces, each of which must include at least one subnet. The distribution of these subscriber subnets into routing protocols is determined by user-defined import or export route policies.

Group interfaces are provisioned within the subscriber interface and share the associated subscriber subnets. These group interfaces can represent either a collection of subscribers on an Access Node (AN) or the AN itself.

Static SAPs are configured under the group interface. Managed (or dynamic) SAPs are automatically created under the group interfaces. For more information, see Managed SAP and Capture SAP. For example, in a VLAN-per-AN model, only one SAP per group interface is required. In contrast, in a VLAN-per-subscriber model, each subscriber on the AN necessitates its own SAP. All SAPs on a group interface must be on the same physical port or Link Aggregation Group (LAG).

For more information about static SAPs see the SAPs section in the 7450 ESS, 7750 SR, 7950 XRS, and VSR Services Overview Guide.

Various subscriber-related features, including DHCP relay, DHCP snooping, PPPoE parameters, IPv6 parameters, anti-spoofing filters, and more, are enabled at the group interface level.

Because the RCO model serves as the primary deployment model, all examples provided in this guide are contextualized within the RCO model.

2.1.8 Subscriber management in distributed network edge

For more information, see Appendix: Subscriber management in distributed access aggregation devices.

2.2 Subscriber management concepts

There are several Nokia BNG concepts that need introduction to better understand subscriber management. The following section defines those concepts and the CLI configuration details.

2.2.1 Subscriber

A broadband subscriber represents a residential or a small business user. At a nodal level, a subscriber is typically defined by a unique subscriber identifier to which an assortment of policies (or subscriber profiles) can be applied. This uniquely identifies a billable entity for the service provider.

In Nokia BNG, a subscriber is identified by a subscriber identification string (*subscriber-id*) of maximum 64 characters long. This subscriber ID can then be passed to the external systems where management, troubleshooting, and billing operations require the subscriber ID name. In Nokia BNG, the subscriber ID can be assigned by several means:

- Use RADIUS, DIAMETER, or a Local User Database (LUDB), which are commonly referred to as *authentication sources*, during the authentication phase to assign the subscriber ID.
- Use a Python script during the instantiation process of the subscriber.
- Configure explicit defaults on a SAP level, which take effect if the subscriber ID is not provided by the previous two methods. On the SAP level, *subscriber-id* can be:
  - provisioned statically as a string
  - mapped to a SAP-id
  - automatically derived as a concatenation of user-defined subscriber identification
- Implicit default is enabled on the system level and it takes effect if none of the preceding methods are enabled. The *subscriber-id* is a 10-character string consisting of the characters A-Z and 0-9 autogenerated based on the subscriber identification fields.

Although the subscriber ID can be configured in multiple places, the value that takes effect in the system is selected from a single source that is determined in the order of priority. For more information about prioritization of authentication sources, see Prioritization of authentication sources.

2.2.1.1 RADIUS returned subscriber ID

The subscriber ID can be obtained from RADIUS during the authentication phase. The name of the Nokia-specific RADIUS attribute for subscriber ID is Alc-Subsc-ID-Str.

For more information about this attribute, see the *7450 ESS, 7750 SR, and VSR RADIUS Attributes Reference Guide*.

2.2.1.2 NASREQ returned subscriber ID

The subscriber ID can be obtained from DIAMETER NASREQ in an AA-Answer message during the authentication phase. The name of the Nokia-specific AVP for subscriber ID is Alc-Subsc-ID-Str.

For more information about this AVP, see Diameter NASREQ application.

2.2.1.3 DIAMETER Gx returned subscriber ID

The subscriber ID can be obtained from DIAMETER Gx in a CC-Answer message during the authentication phase. Include the predefined string “Sub-Id:<*sub-id-string*>” in the Charging-Rule-Install / Charging-Rule-Name AVP.

For more information about this AVP, see the *7750 SR and VSR Gx AVPs Reference Guide* and Policy management via Gx interface.

2.2.1.4 LUDB returned subscriber ID

The subscriber ID can be obtained from LUDB during the authentication phase. Use the following commands to configure the subscriber ID:

- **MD-CLI**

```
configure subscriber-mgmt local-user-db ipoe host identification subscriber-id
configure subscriber-mgmt local-user-db ppp host identification subscriber-id
```

- **classic CLI**

```
configure subscriber-mgmt local-user-db ipoe host identification-strings subscriber-id
configure subscriber-mgmt local-user-db ppp host identification-strings subscriber-id
```

For more information about LUDB see LUDB only access and "Local User Database for the Enhanced Subscriber Management" section in *7450 ESS, 7750 SR, and 7950 XRS Advanced Configuration Guide Part III*

2.2.1.5 Python returned subscriber ID for DHCP based hosts

A DHCP-based host can be assigned a subscriber ID from a Python script configured in Nokia BNG. This can be triggered by the DHCP messages and using Nokia provided Python APIs.

For more information about Python, see Python script support for ESM and LUDB access by DHCPv4 server.

2.2.1.6 Defaults for subscriber ID

Explicit defaults for subscriber ID are configured under the SAP or MSAP policy and can be configured as:

- a static configured string
- the SAP identifier

- a concatenation of user-defined subscriber identification fields obtained during session initiation, such as SAP ID, MAC, circuit-ID, remote-id, and PPPoE session ID

Implicit defaults for subscriber ID are configured at the node level and generates 10-character encoded string as the subscriber ID. It takes effect if no other method to obtain the subscriber ID is available.

For more information about auto generated defaults for subscriber IDs, see Auto-sub ID.

2.2.1.6.1 Static subscriber ID

Static subscriber ID is configured as a custom string under the **sap** or **msap-policy** context and takes effect if the subscriber ID is not returned through authentication sources (RADIUS, Diameter or LUDB) or through a Python script.

Assigning a static subscriber ID makes more sense for static SAP where each SAP corresponds to a subscriber, as shown in the following example.

Example: Static subscriber ID (MD-CLI)

```
[ex:/configure service vprn "demo-vprn" subscriber-interface "demo-sub-if" group-interface
 "demo-grp-if"]
A:admin@node-2# info
    sap 1/x1/1/c1/1:1.1 {
        sub-sla-mgmt {
            defaults {
                subscriber-id {
                    string "user-1"
                }
            }
        }
    }
```

Example: Static subscriber ID (classic CLI)

```
A:node-2>config>service>vprn>sub-if>gpp-if# info
----------------------------------------------
sap 1/x1/1/c1/1:1.1 create
    sub-sla-mgmt
        def-sub-id string "user-1"
    exit
exit
----------------------------------------------
```

In a deployment model where SAP is created dynamically via the subscriber initiation control packets (for example, DHCP, PPPoE, ARP, or even a data trigger), the static subscriber ID is configured in the **msap-policy** context, as shown in the following example.

Example: Static subscriber ID for dynamically created MSAPs (MD-CLI)

```
[ex:/configure subscriber-mgmt]
A:admin@node-2# info
    msap-policy "demo-msap" {
        sub-sla-mgmt {
            defaults {
                subscriber-id {
                    string "user-1"
                }
            }
        }
```

```
    }
```

Example: Static subscriber ID for dynamically created MSAPs (classic CLI)

```
A:node-2>config>subscr-mgmt# info
----------------------------------------------
        msap-policy "demo-msap" create
            sub-sla-mgmt
                def-sub-id string "user-1"
            exit
        exit
----------------------------------------------
```

In this case, there is only one subscriber ID for all SAPs sharing this **msap-policy**.

For more information about dynamic SAP instantiation, see Managed SAP and Capture SAP.

2.2.1.6.2 SAP ID as subscriber ID

A subscriber ID can be mapped to a SAP ID. Similar to static subscriber ID assignment, this is configured under a **sap** or **msap-policy** context and it takes effect if the subscriber ID is not returned through authentication sources (RADIUS, Diameter or LUDB) or through a Python script.

In the following example a subscriber ID becomes (1/x1/1/c1/1:1.1).

Example: SAP ID as subscriber ID (MD-CLI)

```
[ex:/configure service vprn "demo-vprn" subscriber-interface "demo-sub-if" group-interface
 "demo-grp-if"]
A:admin@node-2 # info
    sap 1/x1/1/c1/1:1.1 {
        sub-sla-mgmt {
            defaults {
                subscriber-id {
                    sap-id
                }
            }
        }
    }
```

Example: SAP ID as subscriber ID (classic CLI)

```
A:node-2>config>service>vprn>sub-if>gpp-if# info
----------------------------------------------
sap 1/x1/1/c1/1:1.1 create
    sub-sla-mgmt
        def-sub-id use-sap-id
    exit
exit
----------------------------------------------
```

For dynamically created SAPs, the subscriber ID is mapped to the SAP ID at the time of the SAP creation.

Example: SAP ID as subscriber ID for dynamically created MSAPs (MD-CLI)

```
[ex:/configure subscriber-mgmt]
A:admin@node-2# info
    msap-policy "demo-msap" {
```

```
        sub-sla-mgmt {
            defaults {
                subscriber-id {
                    sap-id
                }
            }
        }
    }
```

Example: SAP ID as subscriber ID for dynamically created MSAPs (classic CLI)

```
A:node-2>config>subscr-mgmt# info
----------------------------------------------
        msap-policy "demo-msap" create
            sub-sla-mgmt
                def-sub-id use-sap-id
            exit
        exit
----------------------------------------------
```

For more information about dynamic SAP instantiation, see Managed SAP and Capture SAP.

2.2.1.6.3 Autogenerated subscriber ID

The autogenerated format of the subscriber ID name can be a user-friendly string based on the subscriber identification fields that are dependent on the session type (DHCP or PPPoE).

Examples of the host or session identification fields that are common to DHCP and PPPoE include:

- MAC address
- SAP ID
- circuit ID
- remote ID

For the complete list of command options for the **configure subscriber-mgmt auto-sub-id** command, see the following:

- *7450 ESS, 7750 SR, 7950 XRS, and VSR Classic CLI Command Reference Guide*
- *7450 ESS, 7750 SR, 7950 XRS, and VSR MD-CLI Command Reference Guide*

The automatic subscriber ID is generated at the end of the host or session initiation process (after the authentication phase is completed) and only in the case when the subscriber ID has not been already provided by any other more specific means (RADIUS, Diameter, LUDB, or Python).

Similar to the previous two cases, the explicit auto subscriber ID generation is enabled as follows:

- **MD-CLI**

```
configure service vprn subscriber-interface group-interface sap sub-sla-mgmt defaults
 subscriber-id auto-id
configure service ies subscriber-interface group-interface sap sub-sla-mgmt defaults
 subscriber-id auto-id
```

- **classic CLI**

```
configure service ies subscriber-interface group-interface sap sub-sla-mgmt def-sub-id use-
auto-id
```

```
configure service vprn subscriber-interface group-interface sap sub-sla-mgmt def-sub-id use-auto-id
```

The subscriber host or session identification fields used in automatic generation of the subscriber ID are configured on a nodal level, which means that there can be only a single set of subscriber identification fields defined per host or session type (IPoE or PPPoE) per BNG.

Auto subscriber ID identification fields can be reconfigured only if no active sessions using this method of assignment of the subscriber ID are present in the system. Auto-sub-id is not applicable to static subscribers.

2.2.2 Subscriber Session

A subscriber session in the BNG represents a user device to which the subscriber management functions are applied. This device can be a single stack device (IPv4 or IPv6) or dual-stack device (IPv4 or IPv6 with various combinations of IPv6 addresses, SLAAC, DHCP-IA-NA, DHCP-IA-PD).

A subscriber session can be either statically configured or dynamically learned in the BNG based on trigger packets. Trigger packets are typically DHCP and PPPoE control packets starting with DHCPv4 Discover, DHCPv6 Solicit, or PPPoE Active Discovery Initiation (PADI) messages. However, they can also be as simple as an ARP packet, IPv6 Router Solicitation, or even a generic IP data packet.

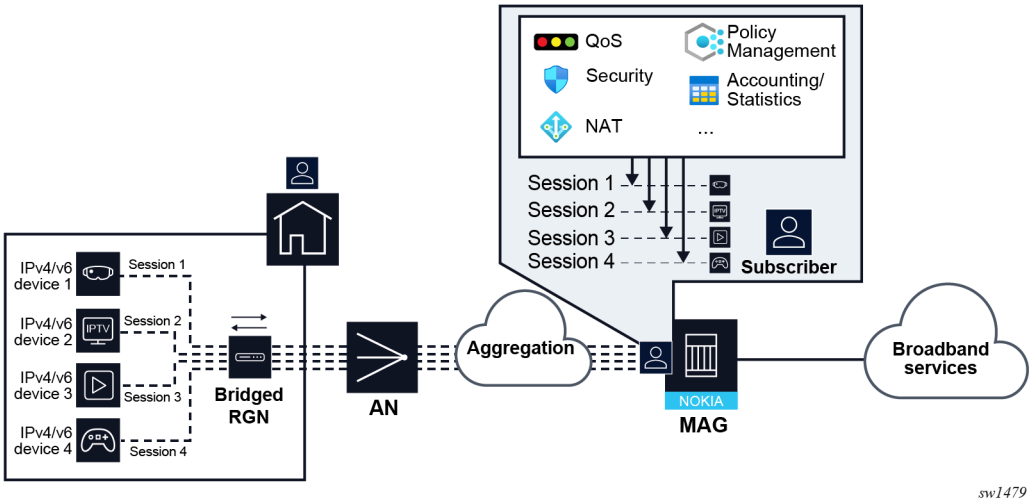
The subscriber sessions are instantiated on a SAP and are uniquely identified by various fields such as MAC address, IP address, circuit ID, remote ID, and others. Subscriber sessions of the same subscriber can be instantiated on a single SAP or spread across multiple SAPs on the same port.

During the subscriber session instantiation process, resources such as queues, filters, and counters are associated with the session. Although the queues, filters, and counters are not the only resource allocated with the session, they are the primary ones through which various policies are applied.

A subscriber can have a single or multiple sessions instantiated, depending on the configuration of the user premise. It is expected that all sessions of a subscriber originate from the same site and are reached by the same port on the BNG. The following three figures show some of the configurations:

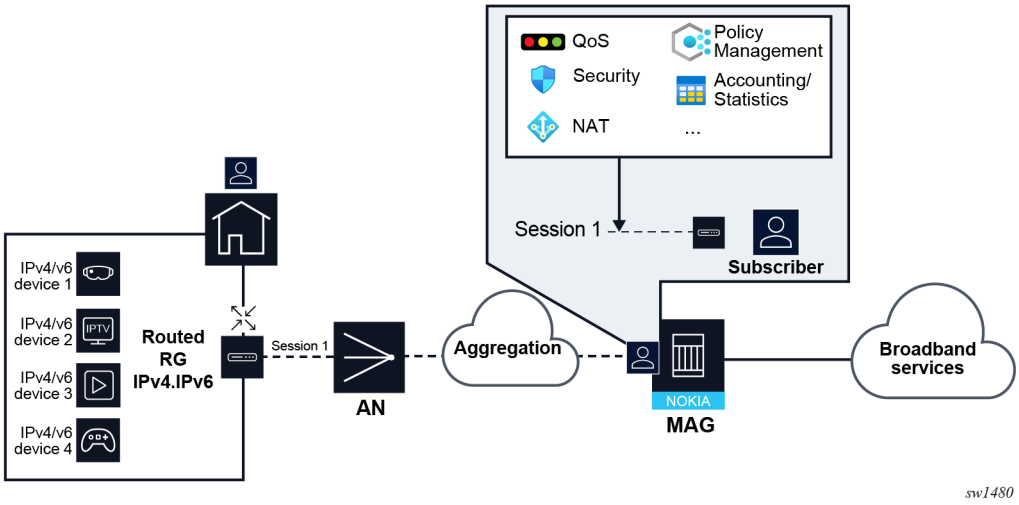
- Figure 3: Subscriber sessions in residential deployment with bridged RG shows the residential premise with a bridged residential gateway (RG), where home devices are directly exposed to the BNG. For example, the BNG is directly involved in their IP address assignment through DHCP and PPPoE control packets. In this configuration, a subscriber is associated with multiple sessions, each representing a device within a residence.

Figure 3: Subscriber sessions in residential deployment with bridged RG



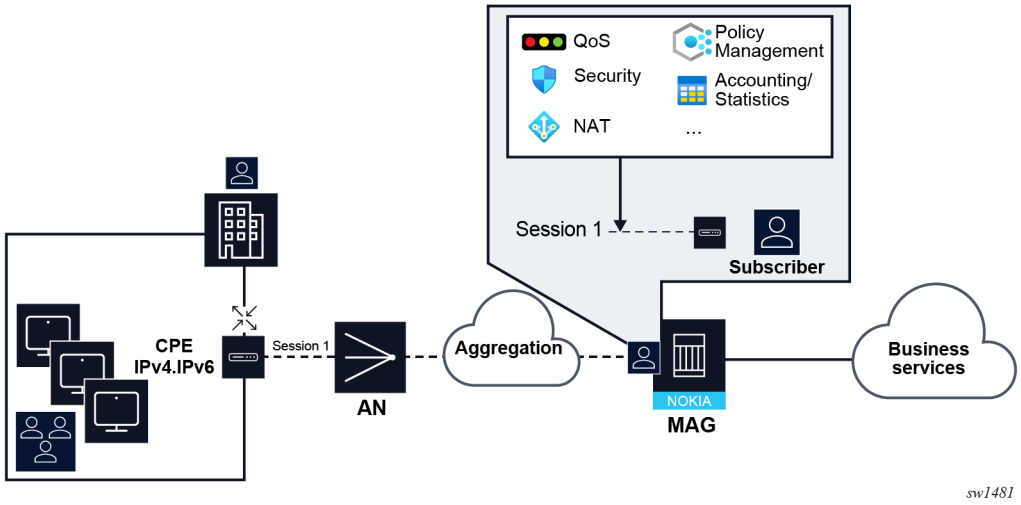
- Figure 4: Subscriber sessions in residential deployment with routed RG shows the residential premise with routed RG where the residential homes are assigned the IP addresses by the RG and in this sense their control packets never reach the BNG. The IP address of the RG is managed by the BNG and therefore, the RG represents the subscriber session. In this configuration, a subscriber has a single session that represents the RG itself. The RG may or may not run a NAT that hides the IP addresses of devices in the residence.

Figure 4: Subscriber sessions in residential deployment with routed RG



- Figure 5: Subscriber sessions in small business deployments with routed CPE shows the small business deployment with routed CPE. This scenario is similar to the previous one, where the service offered to the subscribers are different.

Figure 5: Subscriber sessions in small business deployments with routed CPE



Note: Sometimes the term "subscriber host" is used instead of "subscriber session". The reasons are historical and mainly related to dual-tack DHCP deployments where the IPv4 and IPv6 address families of a dual-tack device were authenticated and treated as two separate devices (or hosts). PPPoE natively supports the concept of a session with the session ID as the shared identifier for the IPv4 and IPv6 address families of a dual-stack device, while DHCP does not. SR OS Release 13 introduced an option to tie the IPv4 and IPv6 address families of a dual-stack DHCP device under the concept of a single IPoE session. Also, a statically configured session is commonly referred to as static host.

2.2.3 Service Level Agreement profile

The Service Level Agreement (SLA) profile contains QoS, security and other settings that are applicable to one, a subset, or all session in a home. An SLA profile is a template that can be shared by many subscriber sessions. Some of the settings in the SLA profile include:

- egress and ingress QoS configuration
- egress scheduler policy HQoS
- egress and ingress IP filters
- session limits
- idle timeout for the sessions
- statistics collection
- credit control parameters

The SLA profile becomes instantiated when the resources defined in the template are allocated, at which point the SLA profile becomes an SLA profile instance (SPI). By default, SPIs are created during the session setup process for every unique combination of:

- SLA profile name
- SAP

- Subscriber-ID

If the SLA profile does not explicitly define an ingress or egress QoS policy or filters, the default SAP ingress or default SAP egress QoS policy or filter is used.

Use the commands in the following context to configure SLA profile.

```
configure subscriber-mgmt sla-profile
```

For the complete list of command options for the **configure subscriber-mgmt sla-profile** command, see the following:

- *7450 ESS, 7750 SR, 7950 XRS, and VSR Classic CLI Command Reference Guide*
- *7450 ESS, 7750 SR, 7950 XRS, and VSR MD-CLI Command Reference Guide*

The following example shows the SLA profile called “demo”. The commands presented in this example are only a subset of a larger set of commands under the SLA profile.

Example: MD-CLI

```
[ex:/configure subscriber-mgmt sla-profile "demo"]
A:admin@node-2# info
    credit-control-policy "demo-cc-policy"
    egress {
        ip-filter "demo-filter-egress"
        qos {
            sap-egress {
                policy-name "demo-qos"
            }
        }
    }
    session-limits {
        ipoe 8
        overall 10
        pppoe {
            local 8
        }
    }
    ingress {
        ip-filter "demo-filter-ingress"
    }
```

Example: classic CLI

```
A:node-2>config>subscr-mgmt>sla-prof# info
----------------------------------------------
            credit-control-policy "demo-cc-policy"
            session-limits
                ipoe 8
                pppoe-local 8
                overall 10
            exit
            ingress
                ip-filter 11
            exit
            egress
                qos 20
                exit
                ip-filter 10
            exit
```

```
-----------------------------------------------
```

2.2.4 Subscriber profile

The subscriber profile is a template that contains the applicable settings for all sessions that belong to the same subscriber. Examples include:

- ingress and egress QoS scheduler policy and policer control policy
- egress aggregate rate limit
- RADIUS accounting policy
- IGMP policy
- MLD policy
- PIM policy
- session limits

Subscribers are either explicitly mapped to a subscriber profile template or are dynamically associated with a subscriber profile.

Attempting to delete any subscriber profile (including the profile named "default") while in use by an existing active subscriber, fails.

Use the commands in the following context to configure a subscriber profile.

```
configure subscriber-mgmt sub-profile
```

For the complete list of command options for the **configure subscriber-mgmt sub-profile** command, see the following:

- *7450 ESS, 7750 SR, 7950 XRS, and VSR Classic CLI Command Reference Guide*
- *7450 ESS, 7750 SR, 7950 XRS, and VSR MD-CLI Command Reference Guide*

The following example shows the subscriber profile called “demo”. A subscriber using this profile has an egress aggregate rate limit of 100Mbps, is using RADIUS accounting policy “acct-1”, and has a limit of 10 IPoE sessions. The commands presented in this example are only a subset of a larger set of commands under the subscriber profile.

Example: MD-CLI

```
[ex:/configure subscriber-mgmt sub-profile "demo"]
A:admin@node-2# info
    egress {
        qos {
            agg-rate {
                rate 100000
            }
        }
    }
    radius-accounting {
        policy ["acct-1"]
    }
    session-limits {
        ipoe 10
    }
```

Example: classic CLI

```
A:node-2>config>subscr-mgmt>sub-prof# info
----------------------------------------------
            session-limits
                ipoe 10
            exit
            radius-accounting
                policy "acct-1"
            exit
            egress
                agg-rate-limit 100000
            exit
----------------------------------------------
```

2.2.5 Subscriber identification policy

A subscriber identification policy is used by dynamically created sessions and it contains:

- URL pointers to the Python scripts used for dynamic session identification based on the information present in the DHCPv4 ACK message. These scripts represent a legacy method for dynamic DHCP session identification. For more information on Python scripting in ESM on the BNG , see Python script support for ESM.
- A subscriber profile map and an SLA profile map. The subscriber profile map creates a mapping between the sub and SLA profile strings returned by the Python script, LUDB, RADIUS or NASREQ and the profile names configured in the BNG. See SLA and subscriber profile mapping.
- DHCP options from which the subscriber identification strings can be derived. Some deployments require that DHCPv4 server provides subscriber strings necessary for subscriber identification.

A subscriber identification policy is defined using the following command:

- **MD-CLI**

```
configure subscriber-mgmt sub-ident-policy name "sub-ident-policy-1"
```

- **classic CLI**

```
configure subscriber-mgmt sub-ident-policy "sub-ident-policy-1"
```

2.2.5.1 SLA and subscriber profile mapping

In the BNG, the following elements must be assigned to a subscriber during the instantiation phase:

- subscriber identification string (subscriber ID). The subscriber ID can be assigned statically via configuration, or obtained via authentication sources, where it can be derived based on various other parameters obtained during the session instantiation phase. For more information, see Subscriber.
- SLA profile
- subscriber (SUB) profile
- IP address

Both, the SLA and SUB profiles are statically configured in the BNG, but their association to subscribers can be flexible.

Static sessions setup under SAPs link directly to the SLA and SUB profiles using their designated names.

In contrast, dynamic sessions use SLA and SUB profile strings retrieved from authentication sources. These strings are subsequently mapped, or translated, to their corresponding SLA and SUB profile names configured in the BNG. This mapping process is configured in the subscriber identification policy.

In many deployments, the SLA and SUB profile strings and their respective SLA and SUB profile names are identical and a default direct mapping can be configured.

Example: RADIUS returns the SLA-profile string for the session:

```
Alc-SLA-Prof-Str = "demo-1"
```

The following example shows the actual SLA profile as defined in the BNG.

Example: MD-CLI

```
[ex:/configure subscriber-mgmt]
A:admin@node-2# info
    sla-profile "sla-profile-1" {
        egress {
            qos {
                sap-egress {
                    policy-name "demo-qos-policy-1"
                }
            }
        }
        ingress {
            ip-filter "demo-filter-1"
        }
    }
```

Example: classic CLI

```
A:node-2>config>subscr-mgmt# info
----------------------------------------------
        sla-profile "sla-profile-1" create
            ingress
                ip-filter 40
            exit
            egress
                qos 10
                exit
            exit
        exit
----------------------------------------------
```

The mapping between the SLA profile string and the SLA profile name (as defined in the BNG) is defined in the subscriber identification policy.

Example: MD-CLI

```
[ex:/configure subscriber-mgmt sub-ident-policy "test"]
A:admin@node-2# info
    sla-profile-map {
        entry "demo-1" {
            sla-profile "sla-profile-1"
        }
    }
```

Example: classic CLI

```
A:node-2>config>subscr-mgmt>sub-ident-pol# info
----------------------------------------------
            sla-profile-map
                entry key "demo-1" sla-profile "sla-profile-1"
            exit
----------------------------------------------
```

The **sub-ident-policy** is referenced under the SAP (or msap-policy) on which the session is instantiated. In this case, a session for which the SLA profile string ‘demo-1’ is returned via RADIUS is associated with the SLA profile “sla-profile-1”.

In case that translation is not required, the following CLI can be used.

Example: MD-CLI

```
[ex:/configure subscriber-mgmt sub-ident-policy "test"]
A:admin@node-2# info
    sla-profile-map {
        use-direct-map-as-default true
    }
    sub-profile-map {
        use-direct-map-as-default true
    }
```

Example: classic CLI

```
A:node-2>config>subscr-mgmt>sub-ident-pol# info
----------------------------------------------
            sub-profile-map
                use-direct-map-as-default
            exit
            sla-profile-map
                use-direct-map-as-default
                entry key "demo-1" sla-profile "sla-profile-1"
            exit
----------------------------------------------
```

2.2.6 Subscriber interface

A subscriber interface is a special type of interface under which subscriber sessions are instantiated. This is an internal loopback interface that supports multiple IP subnets that can be shared across many access nodes on different ports within the BNG. Those subnets aggregating subscribers are advertised via routing protocols on the network side for downstream reachability. Optionally, this interface can be unnumbered, in which case each subscriber session is advertised individually into the network. A subscriber interface contains IP addressing configuration as well as session-specific configurations.

2.2.7 Group interface

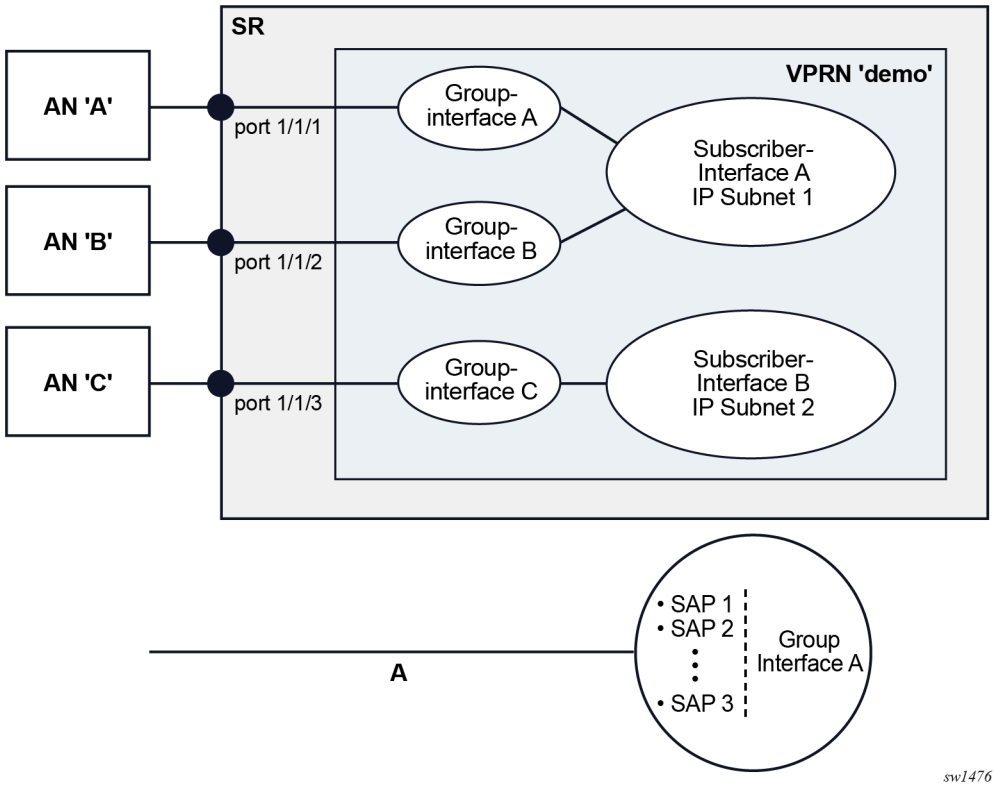
Group interfaces are configured under the subscriber interfaces and they allow multiple SAPs on the same port to be part of the same interface. This is in contrast to traditional interfaces where each interface is associated with a single SAP. This aggregation of SAPs is also relevant in multichassis redundancy where states of the SAPs can be communicated between the nodes on an aggregate level instead of on an

individual SAP level. The group interface is an unnumbered interface. The interface is operationally up if it is administratively enabled and if at least one SAP has been defined and is up and the parent subscriber interface is administratively up. The first SAP defined determines the port for the group interface. If the user attempts to define a subsequent SAP that is on a different port, it results in an error. When the subscriber interface or the group interface is administratively disabled, no packets are delivered or received to or from the subscriber session. However, the subscriber sessions, both dynamic and static, are maintained as long as the lease times are still valid.

Multiple group interfaces under the same subscriber interface share the same IP subnets. Each group interface can accommodate subscribers from a single port on the BNG, whether it is a faceplate, LAG, or a PW-port. The first SAP associated with the group interface determines the port for the group interface.

The concept of subscriber and group interface is shown in the figure that follows.

Figure 6: Subscriber and group interface concept



The following is a sample configuration of the subscriber interface and a single group interface under it (although multiple group-interfaces can be configured). No SAPs are configured because they are automatically assigned through capture SAP configuration. With the **oper-up-while-empty** configuration command, the group-interface is operationally up with no SAPs assigned. For more information about configuring a capture SAP, see Managed SAP and Capture SAP. This configuration supports IPv4 and IPv6 DHCP and PPPoE configuration with IP subnets shown in the following example.

Example: MD-CLI

```
[ex:/configure service vprn "esm" subscriber-interface "sub-int-1"]
A:admin@node-2# info
    admin-state enable
```

```
    ipv4 {
        address 10.10.0.254 {
            prefix-length 24
        }
        address 10.10.1.254 {
            prefix-length 24
        }
    }
    ipv6 {
        delegated-prefix-length variable
        prefix 2001:db8:bbbb::/56 {
            host-type wan
        }
        prefix 2001:db8:bbbb:100::/56 {
            host-type pd
        }
    }
    group-interface "group-int-1" {
        admin-state enable
        oper-up-while-empty true
        ipv4 {
            dhcp {
                admin-state enable
                server [192.168.0.1]
                trusted true
                gi-address 10.10.0.254
                match-circuit-id true
                option-82 {
                    action keep
                    vendor-specific-option {
                        pool-name true
                    }
                }
                lease-populate {
                    max-leases 100
                }
                client-applications {
                    dhcp true
                }
            }
        }
        ipv6 {
            auto-reply {
                neighbor-solicitation true
                router-solicitation true
            }
            dhcp6 {
                pd-managed-route {
                }
                relay {
                    admin-state enable
                    server ["2001:db8::1"]
                    client-applications {
                        dhcp true
                        ppp true
                    }
                }
            }
            router-advertisements {
                admin-state enable
                force-mcast ip-mac
                options {
                    managed-configuration true
                    reachable-time 10000
```

```
                    retransmit-timer 3
                }
            }
            router-solicit {
                admin-state disable
            }
        }
        ipoe-session {
            admin-state enable
            ipoe-session-policy "ipoe-session-policy-1"
            user-db "demo-1"
            sap-session-limit 100
        }
        pppoe {
            admin-state enable
            policy "ppp-policy-1"
            session-limit 100
            sap-session-limit 100
            user-db "demo-1"
        }
        local-address-assignment {
            admin-state enable
            ipv4 {
                server "dhcpv4"
                client-applications {
                    ppp true
                }
            }
        }
    }
```

Example: classic CLI

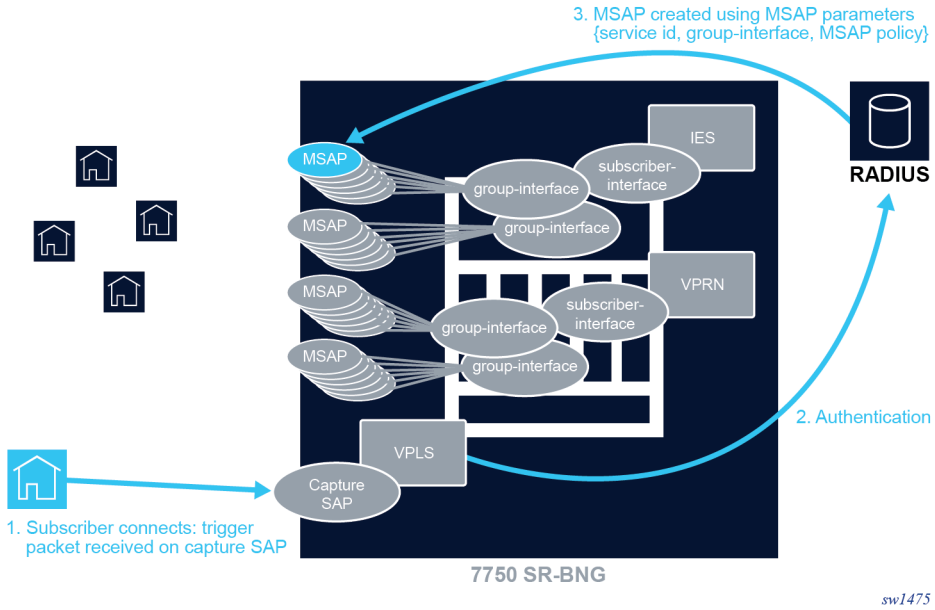
```
A:node-2# configure service vprn "esm" subscriber-interface "sub-int-1"
A:node-2>config>service>vprn>sub-if# info
----------------------------------------------
                address 10.10.0.254/24
                address 10.10.1.254/24
                ipv6
                    delegated-prefix-len variable
                    subscriber-prefixes
                        prefix 2001:db8:bbbb::/56 wan-host
                        prefix 2001:db8:bbbb:100::/56 pd
                    exit
                exit
                group-interface "group-int-1" create
                    ipv6
                        auto-reply
                            neighbor-solicitation
                            router-solicitation
                        exit
                        router-advertisements
                            force-mcast ip mac
                            managed-configuration
                            reachable-time 10000
                            retransmit-time 3
                            no shutdown
                        exit
                        dhcp6
                            pd-managed-route next-hop ipv6
                            relay
                                server 2001:db8::1
                                client-applications dhcp ppp
```

```
                            no shutdown
                        exit
                    exit
                exit
                local-address-assignment
                    server "dhcpv4"
                    client-application ppp-v4
                    no shutdown
                exit
                dhcp
                    option
                        action keep
                        circuit-id
                        no remote-id
                        vendor-specific-option
                            pool-name
                        exit
                    exit
                    server 192.168.0.1
                    trusted
                    lease-populate 100
                    gi-address 10.10.0.254
                    match-circuit-id
                    no shutdown
                exit
                ipoe-session
                    ipoe-session-policy "ipoe-session-policy-1"
                    sap-session-limit 100
                    user-db "demo-1"
                    no shutdown
                exit
                oper-up-while-empty
                pppoe
                    policy "ppp-policy-1"
                    session-limit 100
                    sap-session-limit 100
                    user-db "demo-1"
                    no shutdown
                exit
            exit
-----------------------------------------------
```

2.2.8 Managed SAP and Capture SAP

Subscriber sessions are created on a subscriber SAP. For a shared VLAN deployment model, these SAPs are usually statically configured as the limited number of VLANs are known. For a VLAN per-subscriber deployment model, it is advantageous that the subscriber SAPs are automatically created and deleted when subscriber sessions connect or disconnect. These are called Managed SAPs (MSAPs).

Figure 7: Managed SAP instantiation example



A capture SAP is a SAP on which new subscriber sessions on unconfigured VLANs are discovered. It actively monitors all session initiation traffic (trigger packet) on an underlying port, for example, PADI or DHCP traffic. If this incoming traffic does not correspond to an existing managed SAP, the capture SAP forwards this traffic to the CPM. Assuming the session is legitimate, the CPM extracts the VLANs for the ensuing session.

The reception of a valid trigger packet on a capture SAP initiates a RADIUS, DIAMETER, or local user database authentication to provide the service context where the MSAP should be created. The VLAN of the created MSAP is the same as the authenticated trigger packet. An MSAP functions like a regular SAP but its configuration is not user editable and not maintained in the configuration file. By default, an MSAP is deleted from the system when the last subscriber session active on the MSAP disconnects.

The following trigger types are supported on a capture SAP:

- **dhcp**

  DHCPv4 client messages
- **pppoe**

  PPPoE PADI messages from PPPoE clients
- **arp**

  ARP-Request from an ARP host with static configured IPv4 address
- **dhcp6**

  DHCPv6 client messages
- **rtr-solicit**

  Router Solicitation messages from a SLAAC hosts
- **data**

  An ARP-Request, IPv4 or IPv6 packet received from a data-triggered host

Multiple trigger types can be enabled on a single capture SAP. The **data** and **arp** trigger types are mutually exclusive.

A capture SAP is created in a VPLS service. A capture SAP does not forward traffic but captures received trigger packets for authentication. Similar to a default SAP, at least one of the Q-tags of a capture SAP must be a wildcard (*), meaning any tag value. See the following example configuration.

Example: MD-CLI

```
[ex:/configure service]
A:admin@node-2# info
vpls "capture-vpls" {
        admin-state enable
        service-id 10
        customer "1"
        capture-sap 1/x1/1/c1/4:*.* {
            description "capture SAP port 1/x1/1/c1/4"
            trigger-packet {
                arp true
                dhcp true
                dhcp6 true
                pppoe true
                rtr-solicit true
            }
            msap-defaults {
                policy "msap-policy-1"
                service-name "submgmt-ies"
                group-interface "group-int-1-1"
            }
            ipoe-session {
                admin-state enable
                ipoe-session-policy "ipoe-policy-1"
                user-db "ludb-1"
            }
            pppoe {
                policy "ppp-policy-1"
                user-db "ludb-1"
            }
        }
    }
```

Example: classic CLI

```
A:node-2>config>service# info
----------------------------------------------
vpls 10 name "capture-vpls" customer 1 create
    stp
        shutdown
    exit
    sap 1/x1/1/c1/4:*.* capture-sap create
        description "capture SAP port 1/x1/1/c1/4"
        trigger-packet arp dhcp dhcp6 pppoe rtr-solicit
        pppoe-policy "ppp-policy-1"
        pppoe-user-db "ludb-1"
        ipoe-session
            ipoe-session-policy "ipoe-policy-1"
            user-db "ludb-1"
            no shutdown
        exit
        msap-defaults
            group-interface "group-int-1-1"
            policy "msap-policy-1"
```

```
            service 1000
        exit
        no shutdown
    exit
    no shutdown
exit
----------------------------------------------
```

A capture SAP and default SAP cannot be configured simultaneously on a dot1q-encapsulated port. A capture SAP and default SAP cannot be configured simultaneously on a QinQ-encapsulated port when the outer tag is the same.

A SAP lookup based on the outer and inner tags is performed when a packet is received on a port. When no corresponding SAP or MSAP is found, the packet is handled by the capture SAP, meaning that the trigger packets are sent to the CPM and all other packets are dropped.

An ingress VLAN ID (VID) type **mac** filter can be configured on a capture SAP to have additional control on the VLANs that are allowed to initiate a host setup. Other filter types are not supported on a capture SAP.

For a capture SAP on a dot1q encapsulated port:

- *port-id*:*
  Matches any valid single tagged trigger packet on a *port-id* for which no more specific SAP or MSAP is found. A single Q-tag (*port-id*:tag) is available for authentication. The corresponding MSAP is created as: *port-id*:tag

For a capture SAP on a QinQ-encapsulated port:

- *port-id*:*.*
  Matches any valid double tagged trigger packet on a *port-id* for which no more specific SAP or MSAP is found.

  Both Qw-tags (*port-id*:tag1.tag2) are available for authentication.

  The corresponding MSAP is created as: *port-id*:tag1.tag2.

  The optional **allow-dot1q-msaps** command configured at the capture SAP enables additional support for single-tagged trigger packets:

  – Valid single-tagged trigger packets for which no more specific SAP or MSAP is found are matched on *port-id*.

  – A single Q-tag is available for authentication, the second tag is set to zero (*port-id*:tag.0).

  – The corresponding MSAP is created as: *port-id*:tag.0.

  – The following command should be configured where a combination of *port-id*:tag1.0 and *port-id*:tag1.tag2 MSAPs coexist. When not configured, *port-id*:tag1.0 MSAPs attract double tagged *port-id*:tag1.tag2 encapsulated traffic which is either dropped (IPoE traffic) or handled as single-tagged traffic causing PPPoE sessions to fail.

    - **MD-CLI**

      ```
      configure service system extended-default-qinq-sap-lookup
      ```

    - **classic CLI**

      ```
      configure system ethernet new-qinq-untagged-sap
      ```

Note: The *SR OS R24.x.Rx Software Release Notes* information indicates the platforms on which these commands are enabled by default.

- *port-id*:**tag1.***
  Matches any valid double-tagged trigger packet with and outer tag equaling tag1 on *port-id* and for which no more specific SAP or MSAP is found.

  Both -tags (*port-id*:tag1.tag2) are available for authentication.

  The corresponding MSAP is created as: *port-id*:tag1.tag2.

  The optional **allow-dot1q-msaps** command configured at the capture SAP enables additional support for single-tagged trigger packets:

  – Valid single-tagged trigger packets with tag equaling tag1 and for which no more specific SAP or MSAP is found are matched on *port-id*.

  – A single-tag is available for authentication; the second tag is set to zero (*port-id*:tag1.0).

  – The corresponding MSAP is created as: *port-id*:tag1.0.

  – It is a prerequisite to have the following command configured to enable both *port-id*:tag1.* capture-sap and *port-id*:tag1.0 MSAP to coexist. The *port-id*:tag1.0 capture-sap cannot be created when not configured.

    - **MD-CLI**

      ```
      configure service system extended-default-qinq-sap-lookup
      ```

    - **classic CLI**

      ```
      configure system ethernet new-qinq-untagged-sap
      ```

    Note: The *SR OS R24.x.Rx Software Release Notes* information indicates the platforms on which these commands are enabled by default.

- *port-id*:***.tag2**
  Both Q-tags (*port-id*:tag1.tag2) are available for authentication.

  The corresponding MSAP is created as: *port-id*:tag1.tag2.

  This is an inverse capture SAP that matches on a fixed inner tag with the outer tag identifying the user. The following restrictions apply when an inverse capture SAP is configured on a port:

  – The only allowed port type is Ethernet.

  – It is not possible to create y.* SAPs when there is a *.x capture SAP present on the port (y=0,1..4094,* and x=1..4094).

  – It is not possible to create a y.* network interface when there is a *.x capture SAP present on the port (y=0,1..4094,* and x=1..4094).

  – There is no support for single-tagged MSAP creation.

To enable the creation of single-tagged and double-tagged MSAPs by a QinQ-encapsulated capture SAP, enable the **allow-dot1q-msap** command in the capture SAP context:

- **MD-CLI**

```
configure service vpls capture-sap allow-dot1q-msaps
```

- **classic CLI**

```
configure service vpls sap capture-sap allow-dot1q-msaps
```

In addition, the following command should be configured for scenarios as previously described.

Note: The Release Notes indicates which platforms have these commands enabled by default.

- **MD-CLI**

```
configure service system extended-default-qinq-sap-lookup
```

- **classic CLI**

```
configure system ethernet new-qinq-untagged-sap
```

Caution: Be aware that enabling the **new-qinq-untagged-sap** or **extended-default-qinq-sap-lookup** command affects the behavior of existing *port-id*:tag1.0 SAPs.

Valid single-tagged trigger packets result in the creation of a *port-id*:tag.0 MSAP. With the **encap-tag-range** command option matching in a local user database, it is possible to specify different MSAP defaults for single or double-tagged MSAPs. s shown in the following example.

Example: defaults for dot1q MSAPs (MD-CLI)

```
[ex:/configure subscriber-mgmt local-user-db "ludb-1"]
A:admin@node-2# info
    admin-state enable
    ipoe {
        match-list [encap-tag-range]
        host "single-tagged" {
            admin-state enable
            host-identification {
                encap-tag-range {
                    from "*.0"
                    to "*.0"
                }
            }
                      msap-defaults {
            policy "msap-policy-2"
            service-name "submgmt-vprn-2"
            group-interface {
                name "group-int-2"
            }
        }
    }
}
```

Example: defaults for dot1q MSAPs (classic CLI)

```
A:node-2>config>subscr-mgmt>loc-user-db# info
----------------------------------------------
```

```
            ipoe
                match-list encap-tag-range
                host "single-tagged" create
                    host-identification
                        encap-tag-range start-tag *.0 end-tag *.0
                    exit
                    msap-defaults
                        group-interface "group-int-2"
                        policy "msap-policy-2"
                    exit
                    no shutdown
                exit
            exit
            no shutdown
----------------------------------------------
```

Example: defaults for QinQ MSAPs (MD-CLI)

```
[ex:/configure service vpls "capture-vpls" capture-sap 1/x1/1/c1/4:*.*]
A:admin@node-2# info
    allow-dot1q-msaps true
    trigger-packet {
        dhcp true
        dhcp6 true
    }
        msap-defaults {
        policy "msap-policy-1"
        service-name "submgmt-vprn-1"
        group-interface "group-int-1"
    }
    ipoe-session {
        admin-state enable
        ipoe-session-policy "ipoe-policy-1"
        user-db "ludb-1"
    }
```

Example: defaults for QinQ MSAPs (classic CLI)

```
A:node-2>config>service>vpls>sap# info
----------------------------------------------
    trigger-packet dhcp dhcp6
    allow-dot1q-msaps
    ipoe-session
        ipoe-session-policy "ipoe-policy-1"
        user-db "ludb-1"
        no shutdown
    exit
    msap-defaults
        group-interface "group-int-1"
        policy "msap-policy-1"
        service 2000
    exit
    no shutdown
----------------------------------------------
```

2.2.8.1 MSAP parameters

A set of mandatory parameters must be provisioned for MSAP creation including the following:

- **service ID**

The service context in which the MSAP is created.

- **interface name**

  The name of the group interface context in which the MSAP is created. The group interface must exist in the provided service in order for the MSAP to be installed.

- **MSAP policy**

  The name of the policy that defines the MSAP parameters. The policy must exist in the subscriber-mgmt context.

MSAP parameters can be obtained from multiple sources as described in the next sections.

2.2.8.1.1 Explicit MSAP parameters from local user database

Configure the local user database at the capture SAP and group interface using the following commands.

- **IPoE sessions**
  Use the following commands to configure a local user database for IPoE sessions:
  - **MD-CLI**

```
configure service vpls capture-sap ipoe-session user-db
configure service ies subscriber-interface group-interface ipoe-session user-db
configure service vprn subscriber-interface group-interface ipoe-session user-db
```

  - **classic CLI**

```
configure service vpls sap capture-sap ipoe-session user-db
configure service ies subscriber-interface group-interface ipoe-session user-db
configure service vprn subscriber-interface group-interface ipoe-session user-db
```

- **PPPoE sessions**
  Use the following commands to configure a local user database for PPPoE sessions:
  - **MD-CLI**

```
configure service vpls capture-sap pppoe user-db
configure service ies subscriber-interface group-interface pppoe user-db
configure service vprn subscriber-interface group-interface pppoe user-db
```

  - **classic CLI**

```
configure service vpls sap capture-sap pppoe-user-db
configure service ies subscriber-interface group-interface pppoe user-db
configure service vprn subscriber-interface group-interface pppoe user-db
```

When RADIUS or DIAMETER authentication is also required after local user database authentication, the authentication policy must be specified in the local user database. In this case, no authentication policy can be configured at the **group-interface** context.

- **IPoE sessions**
  Use the following commands to configure a authentication policy for IPoE sessions:
  - **MD-CLI**

```
configure subscriber-mgmt local-user-db ipoe host authentication nasreq-auth-policy
```

```
configure subscriber-mgmt local-user-db ipoe host authentication radius-auth-policy
```

– **classic CLI**

```
configure subscriber-mgmt local-user-db ipoe host diameter-auth-policy <name>
configure subscriber-mgmt local-user-db ipoe host auth-policy <policy-name>
```

- **PPP sessions**
  Use the following commands to configure a authentication policy for PPP sessions:

  – **MD-CLI**

```
configure subscriber-mgmt local-user-db ppp host authentication nasreq-auth-policy
configure subscriber-mgmt local-user-db ppp host authentication radius-auth-policy
```

  – **classic CLI**

```
configure subscriber-mgmt local-user-db ppp host diameter-auth-policy <name>
configure subscriber-mgmt local-user-db ppp host auth-policy <policy-name>
```

Use the following commands to configure MSAP parameters at the local user database host context.

```
configure subscriber-mgmt local-user-db ipoe host msap-defaults group-interface name
configure subscriber-mgmt local-user-db ipoe host msap-defaults policy
configure subscriber-mgmt local-user-db ipoe host msap-defaults service
```

2.2.8.1.2 Explicit MSAP parameters from RADIUS or DIAMETER authentication

When RADIUS or DIAMETER authentication is required to return the MSAP parameters without prior local user database authentication, the authentication policy should be configured at the capture SAP context. The same authentication policy must also be configured at the **group-interface** context.

Use the following commands to configure authentication:

- **MD-CLI**

```
configure service vpls capture-sap nasreq-auth-policy
configure service ies subscriber-interface group-interface nasreq-auth-policy
configure service vprn subscriber-interface group-interface nasreq-auth-policy

configure service vpls capture-sap radius-auth-policy
configure service ies subscriber-interface group-interface radius-auth-policy
configure service vprn subscriber-interface group-interface radius-auth-policy
```

- **classic CLI**

```
configure service vpls sap diameter-auth-policy <name>
configure service ies subscriber-interface group-interface diameter-auth-policy <name>
configure service vprn subscriber-interface group-interface diameter-auth-policy <name>
configure service vpls sap authentication-policy <name>
configure service ies subscriber-interface group-interface authentication-policy <name>
configure service vprn subscriber-interface group-interface authentication-policy <name>
```

The MSAP is not created if the group interface name returned from RADIUS or DIAMETER has a different authentication policy than the authentication policy configured at the capture SAP.

The following table lists the RADIUS attributes (VSAs) and DIAMETER Attribute Value Pairs (AVPs) required to obtain MSAP parameters in the authentication phase.

Table 1: RADIUS attributes/DIAMETER AVPs for MSAP parameters

Attribute Id	Attribute name	Type	Purpose and format
26.6527.31	Alc-MSAP-Serv-Id	Integer	The service ID of the service context in which the MSAP is created
26.6527.32	Alc-MSAP-Policy	String	The name of the policy that defines the MSAP parameters
26.6527.33	Alc-MSAP-Interface	String	The name of the group interface context in which the MSAP is created
241.26.6527.90	Alc-MSAP-Serv-Name	String	(RADIUS only) The service name of the service context in which the MSAP is created. Alc-MSAP-Serv-Name takes precedence over Alc-MSAP-Serv-Id if both are specified.

2.2.8.1.3 Implicit MSAP parameters specified at the capture SAP

MSAP parameters that are not obtained from a local user database lookup, and that are not returned from RADIUS or DIAMETER can be specified in the **msap-defaults** section of the capture SAP context (this is a last resort scenario).

Use the following commands to configure MSAP command options.

- **MD-CLI**

```
configure service vpls capture-sap msap-defaults group-interface
configure service vpls capture-sap msap-defaults policy
configure service vpls capture-sap msap-defaults service-name
```

- **classic CLI**

```
configure service vpls sap capture-sap msap-defaults group-interface
configure service vpls sap capture-sap msap-defaults policy
configure service vpls sap capture-sap msap-defaults service
```

2.2.8.2 MSAP creation

MSAPs can be created in the IES or VPRN group interfaces.

An MSAP is persistent when subscriber management persistence is enabled. The MSAP parameters are part of the subscriber record.

If the local user database, RADIUS, or DIAMETER authentication did not provide all the required information to create the subscriber host or session (for example, no IP address), the MSAP is created with a short timer while waiting for the host to acquire the missing information. If no host is instantiated when the timer expires, the MSAP is deleted.

Multiple subscribers, subscriber hosts or sessions can share a single MSAP. The MSAP is created with the first instantiated subscriber host or session and deleted when the last associated subscriber host or session is removed from the system.

Note: Only a single MSAP policy can be specified for a MSAP. An attempt to change the MSAP policy by a new subscriber host or session for an existing MSAP results in a host or session setup failure.

2.2.8.3 MSAP QoS configuration

MSAPs are always used in combination with subscriber management. Subscriber traffic QoS models are defined in policies associated with the SLA profile and subscriber profile and result in the instantiation of subscriber queues and policers used for subscriber traffic forwarding. The default QoS policies associated with MSAPs instantiate a single ingress and a single egress queue per MSAP for IES and VPRN services.

These MSAP queues have limited use and can be suppressed in most cases. For single-subscriber MSAPs, the MSAP queues can be suppressed with the following CLI command.

```
configure subscriber-mgmt msap-policy sub-sla-mgmt single-sub-parameters profiled-traffic-only
```

The default QoS policy associated with MSAPs may need to be changed to accommodate different scenarios, including the following:

- Saving queue resources when the **profiled-traffic-only** command cannot be used, such as when more than one subscriber is active on an MSAP.
  Mapping all forwarding classes to a policer in the QoS policy associated with an MSAP, a single policer instead of a queue is instantiated on the MSAP.
- Providing adequate QoS treatment for multicast traffic in a per MSAP replication mode
  Egress multicast traffic in per MSAP replication mode is forwarded by the MSAP queues or policers. Multicast traffic can be mapped into a dedicated queue or policer. The MSAP queue can be port parented to provide scheduling priority on the port level.

The QoS policies associated with an MSAP are configured in the MSAP policy:

- **MD-CLI**

```
configure subscriber-mgmt msap-policy ies-vprn-only-sap-parameters egress qos policy-name
configure subscriber-mgmt msap-policy ies-vprn-only-sap-parameters ingress qos policy-name
```

- **classic CLI**

```
configure subscriber-mgmt msap-policy ies-vprn-only-sap-parameters egress qos <policy-id>
configure subscriber-mgmt msap-policy ies-vprn-only-sap-parameters ingress qos <policy-id>
```

2.2.8.4 Sticky MSAP

After a subscriber session ends, the MSAP is removed from the system and the historical data of the subscriber is deleted. Sticky MSAP allows the MSAP to remain even when the subscriber session ends.

Note: This feature is only recommended for service providers who do not oversubscribe MSAPs in the network.

Sticky MSAP provides the following benefits:

- Because the sticky MSAP is never deleted, the subscriber can start a session faster; processing time is reduced because the MSAP does not have to be recreated.
- The MSAP may contain valuable historical information for the service provider. Retaining the MSAP provides a means for the service provider to look up subscriber historical data.

Only successfully created MSAPs are eligible for stickiness. The sticky MSAP introduces a new idle state. An idle MSAP indicates that the subscriber on the MSAP has disconnected and the MSAP is ready for a new subscriber connection, as shown in the output that follows.

Use the following command to display MSAP information.

```
show service sap-using msap
```

Output example

```
===============================================================================
Service Access Points
===============================================================================
PortId                          SvcId      Ing.  Ing.    Egr.  Egr.   Adm  Opr
                                           QoS   Fltr    QoS   Fltr
-------------------------------------------------------------------------------
[1/x1/1/c1/4:2211.2005](I)      1000       1     none    50    none   Up   Up
-------------------------------------------------------------------------------
Number of SAPs : 1
-------------------------------------------------------------------------------
Number of Managed SAPs : 1, indicated by [<sap-id>]
Flags : (I) = Idle MSAP
-------------------------------------------------------------------------------
===============================================================================
```

There are two ways to remove sticky MSAPs from the system:

- **Manually**

  The **clear service id msap** command removes MSAPs, including MSAPs with active subscribers. To clear only MSAPs without any active subscriber, use the **idle-only** command option.

- **Automatically**

  Sticky MSAPs can be removed if they are idle for longer than the specified time. This removal strategy can be used to keep only MSAPs that are used by regular subscribers and free the system from consuming MSAPs resources used by occasional subscribers. Use the following command to remove MSAPs:

  – **MD-CLI**

  ```
  configure subscriber-mgmt msap-policy sticky-msaps-idle-timeout
  ```

– **classic CLI**

```
configure subscriber-mgmt msap-policy sticky-msaps idle-timeout
```

Note:
Persistence restoration relies on the configured MSAP default command options under the capture SAP context. Use the following command to configure MSAP default command options. See the "Persistence" section in the *7450 ESS, 7750 SR, 7950 XRS, and VSR Basic System Configuration Guide* for more information about persistence functionality.

- **MD-CLI**

```
configure service vpls capture-sap msap-defaults
```

- **classic CLI**

```
configure service vpls sap capture-sap msap-defaults
```

With persistence enabled, Nokia recommends to avoid changing the default after the system has created hosts with these **msap-defaults** values. The hosts are not restored by the system as the **msap-defaults** values are no longer the same.

Sticky MSAP feature consumes resources because it is present in the system even after the subscriber is disconnected. These MSAPs can be cleared with the **clear service id msap** command.

2.2.9 Dynamic instantiation of subscriber sessions

In generic terms, dynamic instantiation of a subscriber session indicates that a limited configuration is required on the node. Configuring subscriber-specific parameters, including the SAP on which subscribers are instantiated, is not required. Instead, shared profiles are configured for multiple subscribers. All specifics for a subscriber, including the profile references, are assigned through authentication and address assignment sources. SAPs are then automatically determined from incoming session initiation traffic (in which a subgroup of the session control[2] traffic) arriving on the capture SAP.

This contrasts with static subscribers where everything must be explicitly configured in the node. For information about static sessions, see Static subscriber hosts.

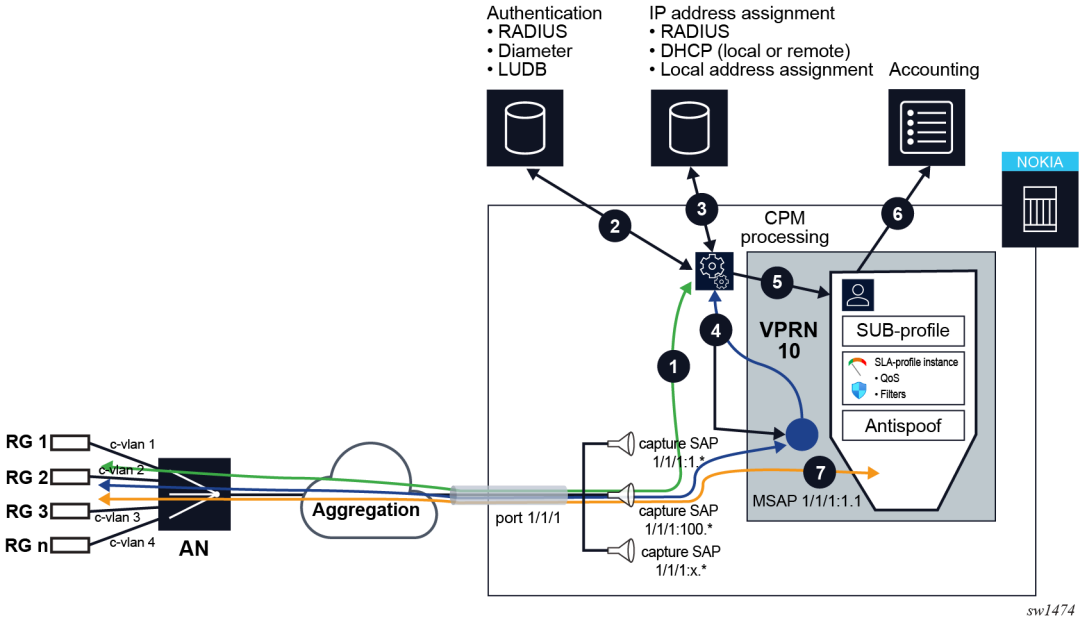
Figure 8: Dynamic instantiation of subscriber sessions shows a high-level workflow of where all parameters are assigned dynamically:

- 1. Session initiation traffic, for example DHCP Discover for IPv4 traffic, arrives on a capture SAP and sent to the control plane.
- 2. Control plane authenticates the user based on information in DHCP Discover (for example Opt82 – circuit-id).
- 3. IP address is allocated to the session.

[2] Subscriber control traffic is responsible for session setup and maintenance, and is processed in the control plane. For example, for IPoE sessions the control traffic is comprised of DHCP messages, and for PPPoE sessions the control traffic is PADx, LCP, IPCP, and PPP keepalives. This contrasts with session data traffic that flows through the node, without being processed at the control plane level.

- 4. Replies are sent back to the user, and the rest of the control traffic is exchanged (DHCP Offer, DHCP Request, DHCP Ack).
- 5. Subscriber session is setup in the system. Managed SAP is derived from the incoming control packets.
- 6. Accounting is optionally started.
- 7. Subscriber session data traffic flows

Figure 8: Dynamic instantiation of subscriber sessions



3 DHCP management

3.1 DHCP principles

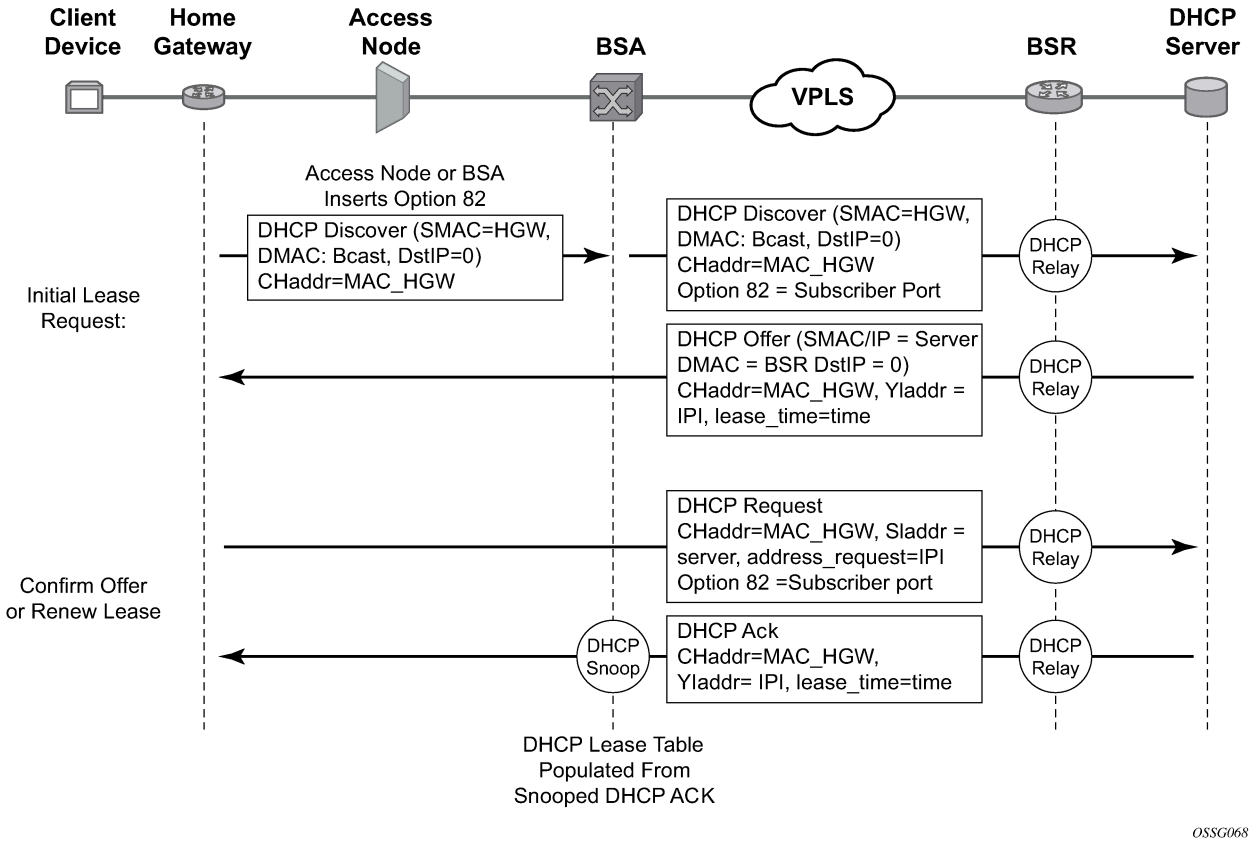
In a Triple Play network, client devices (such as a routed home gateway, a session initiation protocol (SIP) phone, or a set-top box) use the Dynamic Host Configuration Protocol (DHCP) to dynamically obtain their IP address and other network configuration information.

DHCP is defined and shaped by several RFCs and drafts in the IETF DHCP working group including the following:

- RFC 1534, *Interoperation Between DHCP and BOOTP*
- RFC 2131, *Dynamic Host Configuration Protocol*
- RFC 2132, *DHCP Options and BOOTP Vendor Extensions*
- RFC 3046, *DHCP Relay Agent Information Option*

The DHCP operation is shown in Figure 9: IP address assignment with DHCP.

Figure 9: IP address assignment with DHCP



1. During bootup, the client device sends a DHCP discover message to get an IP address from the DHCP Server. The message contains:
   - **destination MAC address**

     broadcast
   - **source MAC address**

     the MAC address of the client device
   - **client hardware address**

     the MAC address of the client device

     If this message passes through a DSLAM or other access node, typically the relay information option (Option 82) field is added, indicating shelf, slot, port, VPI, VCI, and so on, to identify the subscriber.

     DHCP relay is enabled on the first IP interface in the upstream direction. Depending on the scenario, the DSLAM, BSA, or the BSR relays the discover message as a unicast packet toward the configured DHCP server. DHCP relay is configured to insert the GIADDR (gateway IP address) to indicate to the DHCP server in which subnet an address should be allocated.
2. The DHCP server looks up the client MAC address and Option 82 information in its database. If the client is recognized and authorized to access the network, an IP address is assigned and a DHCP offer message returned. The BSA or BSR then relays this back to the client device.

3. It is possible that the discover reached more than one DHCP server and more than one offer is returned. The client selects one of the offered IP addresses and confirms that it wants to use this in a DHCP request message, sent as unicast to the DHCP server that offered it.
4. The DHCP server confirms that the IP address is still available, updates its database to indicate it is now in use, and replies with a DHCP ACK message back to the client. The ACK also contains the Lease Time of the IP address.

3.2 DHCP features

3.2.1 DHCP relay

Because DHCP requests are broadcast packets that normally do not propagate outside of their IP subnet, a DHCP relay agent intercepts such requests and forwards them as unicast messages to a configured DHCP server.

When forwarding a DHCP message, the relay agent sets the GIADDR in the packet to the IP address of its ingress interface. This allows DHCP clients to use a DHCP server on a remote network. From both a scalability and a security point of view, it is recommended that the DHCP relay agent is positioned as close as possible to the client terminals.

DHCP relay is used in a Layer 3 environment, and therefore is only supported in IES services and VPRN services. VPRN is only supported on the 7750 SR.

When DHCP clients and servers are in different VPRN routing instances of which one is the Base routing instance, route leaking (GRT-leaking) should be used to relay DHCPv4 and DHCPv6 messages between a VPRN and the Global Routing Table (GRT).

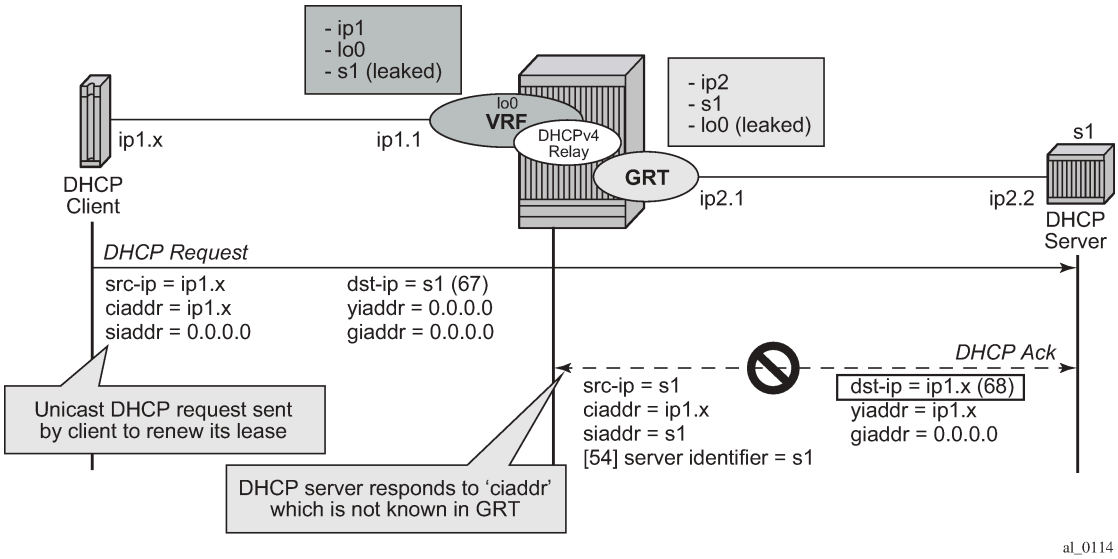
While DHCP relay is not implemented in a VPLS, it is still possible to insert or modify Option 82 information.

In a routed CO environment in the 7750 SR, the subscriber interface’s group interface DHCP relay is stateful.

3.2.2 DHCPv4 relay proxy

In network deployments where DHCPv4 client subnets cannot be leaked in the DHCPv4 server routing instance, unicast renewal messages (DHCP ACKs) cannot be routed in the DHCPv4 server routing instance, as shown in Figure 10: Unicast renewal routing problem. The DHCP server sets the destination IP address of the DHCP ACK to the client IP address (ciaddr) as received in the DHCP REQUEST message. Because there is no route available for the client subnet in the DHCP server routing instance, the DHCP ACK cannot be delivered.

Figure 10: Unicast renewal routing problem



The unicast renewal routing problem shown in Figure 10: Unicast renewal routing problem can be solved with a relay proxy function that enhances the DHCPv4 relay. With the **relay-proxy** command in the DHCPv4 relay on a regular interface or group interface, the unicast renewals are now also relayed to the DHCPv4 server, as described below and shown in Figure 11: Relay unicast messages:

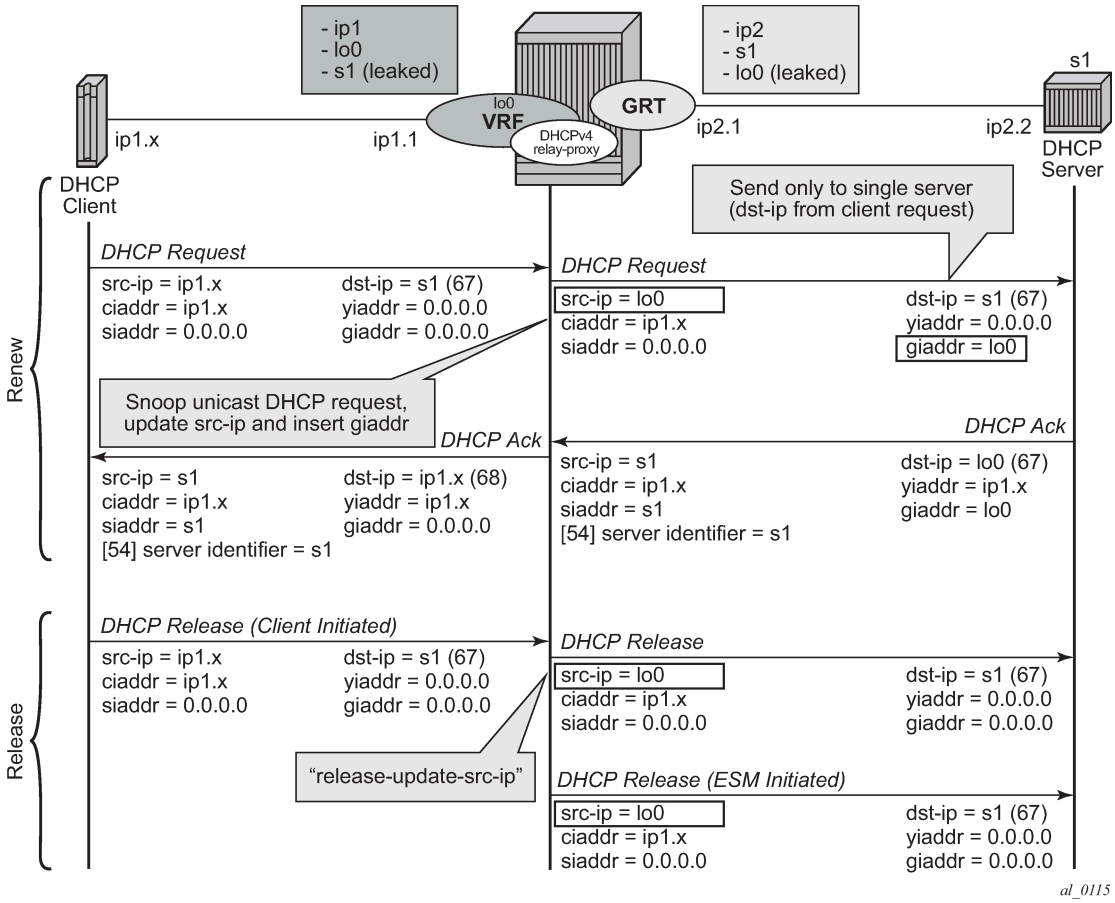
- In the client to server direction, the source IP address is updated and the gateway IP address (gi-address) field is added before sending the message to the intended DHCP server (the message is not broadcasted to all configured DHCP servers.
- In the server to client direction, the GI address field is removed and the destination IP address is updated with the value of the IP address (yiaddr) field.

When **relay-proxy** is enabled, the GI address can be configured to any local address that is configured in the same routing instance. The GI address is the only address that must be leaked in the DHCPv4 server routing instance because a DHCPv4 server always sends the response on a relayed packet to the relay agent using the gi-address as the destination IP address.

By default, unicast DHCPv4 RELEASE messages are forwarded transparently by a relay proxy function. The optional **release-update-src-ip** flag updates the source IP address with the value that is used for all relayed DHCPv4 messages, as shown in Figure 11: Relay unicast messages.

DHCPv4 FORCERENEW messages that are sent from a trusted external DHCPv4 server to a DHCPv4 relay agent configured as a relay proxy are forwarded to the DHCP client, if a corresponding DHCPv4 lease exists; otherwise, the DHCPv4 FORCERENEW messages are dropped.

Figure 11: Relay unicast messages



The **relay-proxy** command can also be used to hide the DHCPv4 server address for DHCP clients. This prevents the client from learning the DHCPv4 server infrastructure details such as the IP address and number of servers. Hiding infrastructure details helps in Denial of Service (DoS) prevention.

The optional **siaddr-override** *ip-address* parameter in relay-proxy enables DHCPv4 server IP address hiding toward the client. The client interacts with the relay proxy as if it is the DHCP server. In addition to the relay proxy functions as described earlier, the following actions are performed when DHCPv4 server IP address hiding is configured:

- In all DHCP messages to the client, the value of the following header fields and DHCP options containing the DHCP server IP address is replaced with the configured IP address:
  - the source IP address
  - the siaddr field in the DHCPv4 header if it is not equal to zero in the message received from the server
  - the Server Identification option (DHCPv4 option 54) if present in the original server message
- The DHCP OFFER selection occurs during initial binding. Only the first DHCP OFFER message is forwarded to the client. Subsequent DHCP OFFER messages from different servers are silently dropped.

The **siaddr-override** *ip-address* parameter can be any local address in the same routing instance. If DHCP relay lease split is enabled, **siaddr-override** *ip-address* has priority over the **emulated-server** *ip-address* configured in the proxy server and is used as the source IP address.

The active DHCPv4 server IP address obtained from the DHCP OFFER selection is required for the IP address hiding function and is stored in the lease state record. Therefore **lease-populate** must be enabled on the interface when **siaddr-override** *ip-address* is configured.

Figure 12: DHCP server IP address hiding/initial binding shows the initial lease binding phase of a relay proxy with DHCP server address hiding enabled. In the absence of a DHCP lease state in the initial lease binding phase, the DHCP server IP address resulting from the OFFER selection is stored in a DHCP transaction cache. After successful lease binding, the DHCP server IP address is added to the lease state record.

In a host creation failure scenario, if no transaction cache or lease state is available when a DHCP REQUEST message is received, then the DHCP REQUEST is silently dropped. The drop reason can be found by enabling DHCP debug.

Figure 12: DHCP server IP address hiding/initial binding

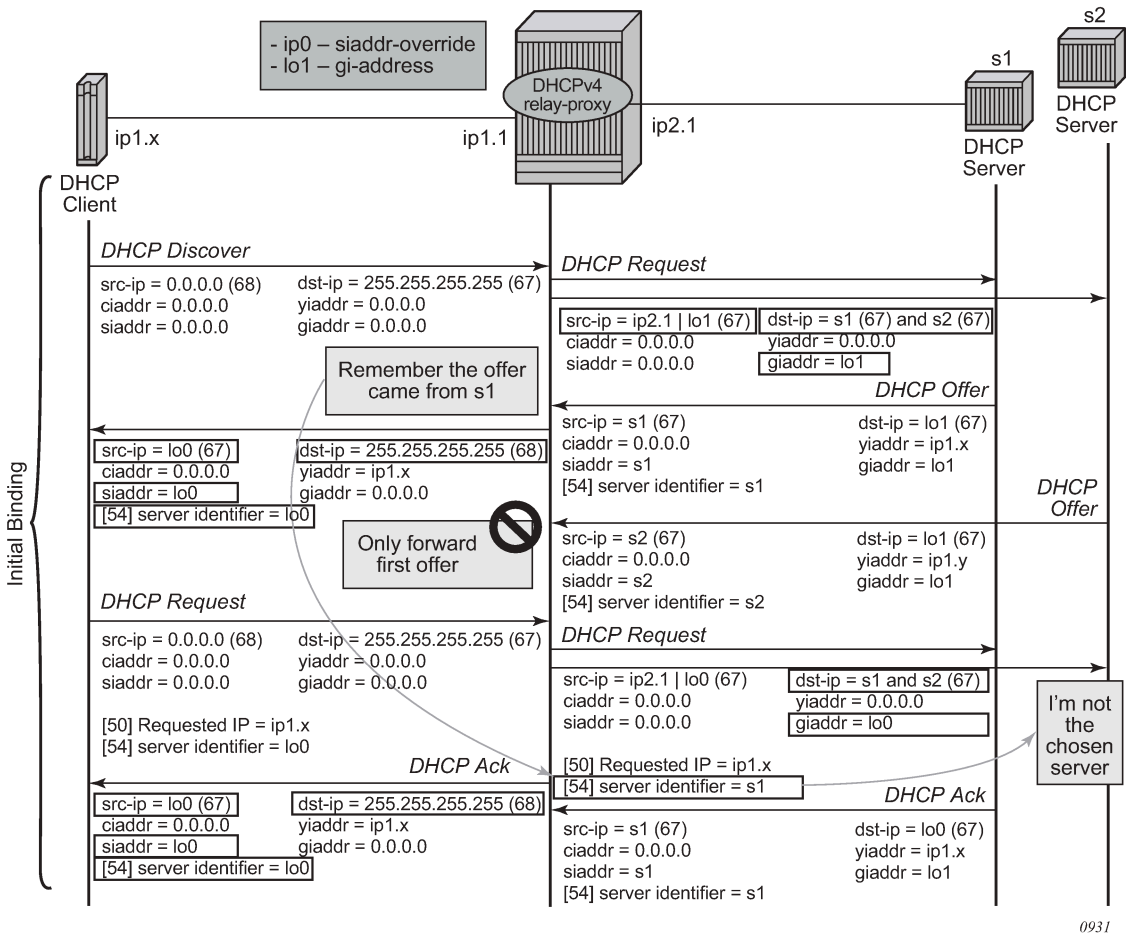


Figure 13: DHCP server IP address hiding/lease renewal shows the lease renewal phase of a relay proxy with DHCP server address hiding enabled. A unicast REQUEST (renew) is relayed only to the DHCP server owning the lease. A broadcast REQUEST (rebind) is relayed to all configured DHCP servers.

During lease renewal, the DHCP server IP address can be updated in the lease state if the DHCP ACK is received from a different server. This optimizes the DHCP proxy relay operation in a DHCP server failover scenario. This is shown in Figure 14: DHCP server IP address hiding, lease renewal with active server failure.

Figure 13: DHCP server IP address hiding/lease renewal

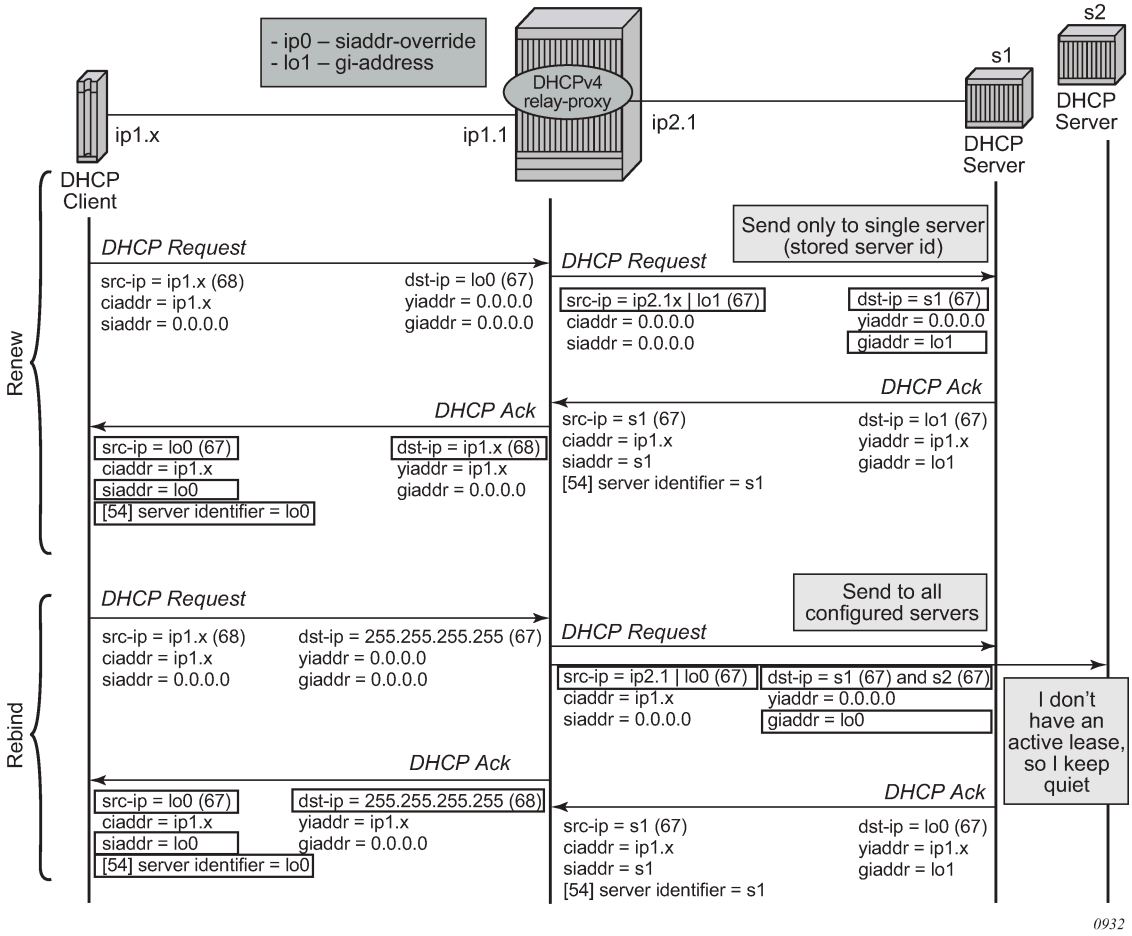


Figure 14: DHCP server IP address hiding, lease renewal with active server failure

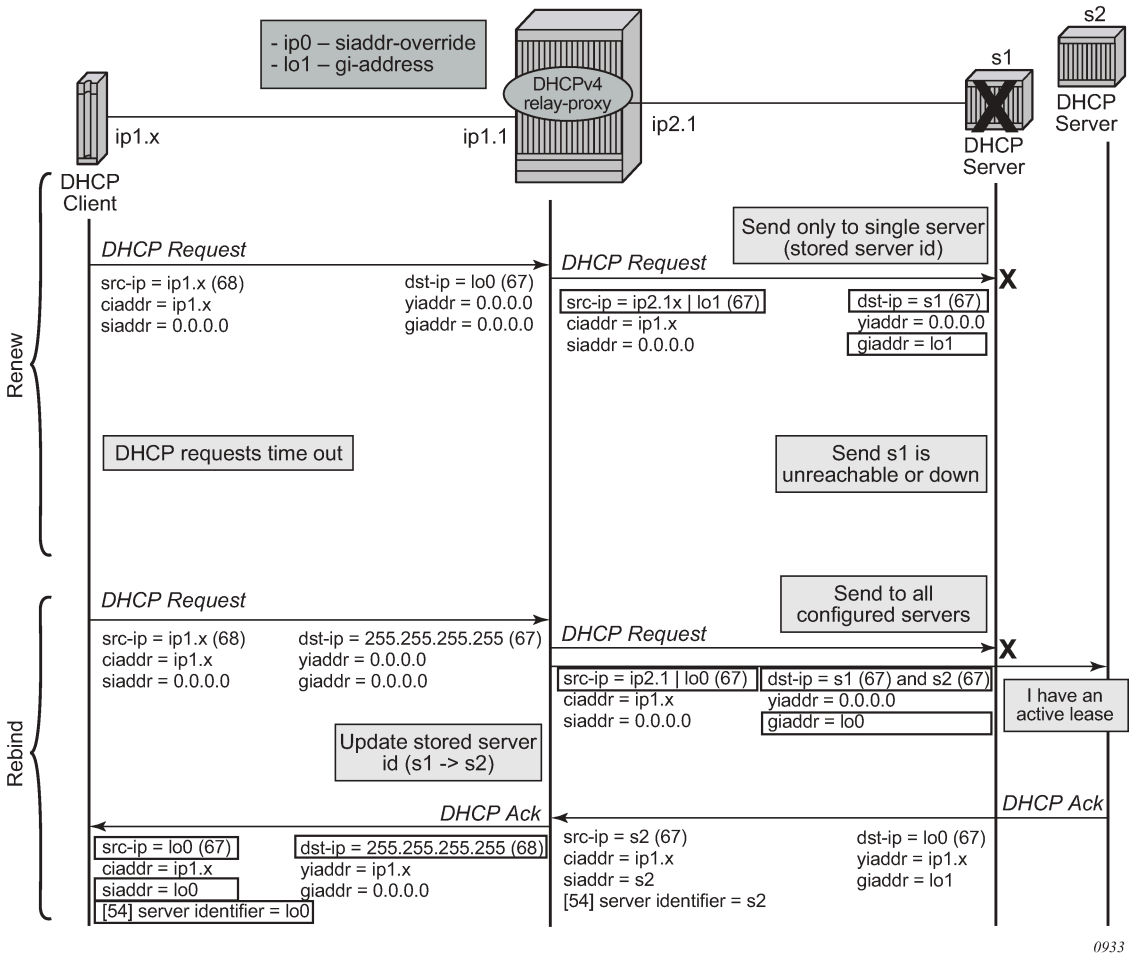
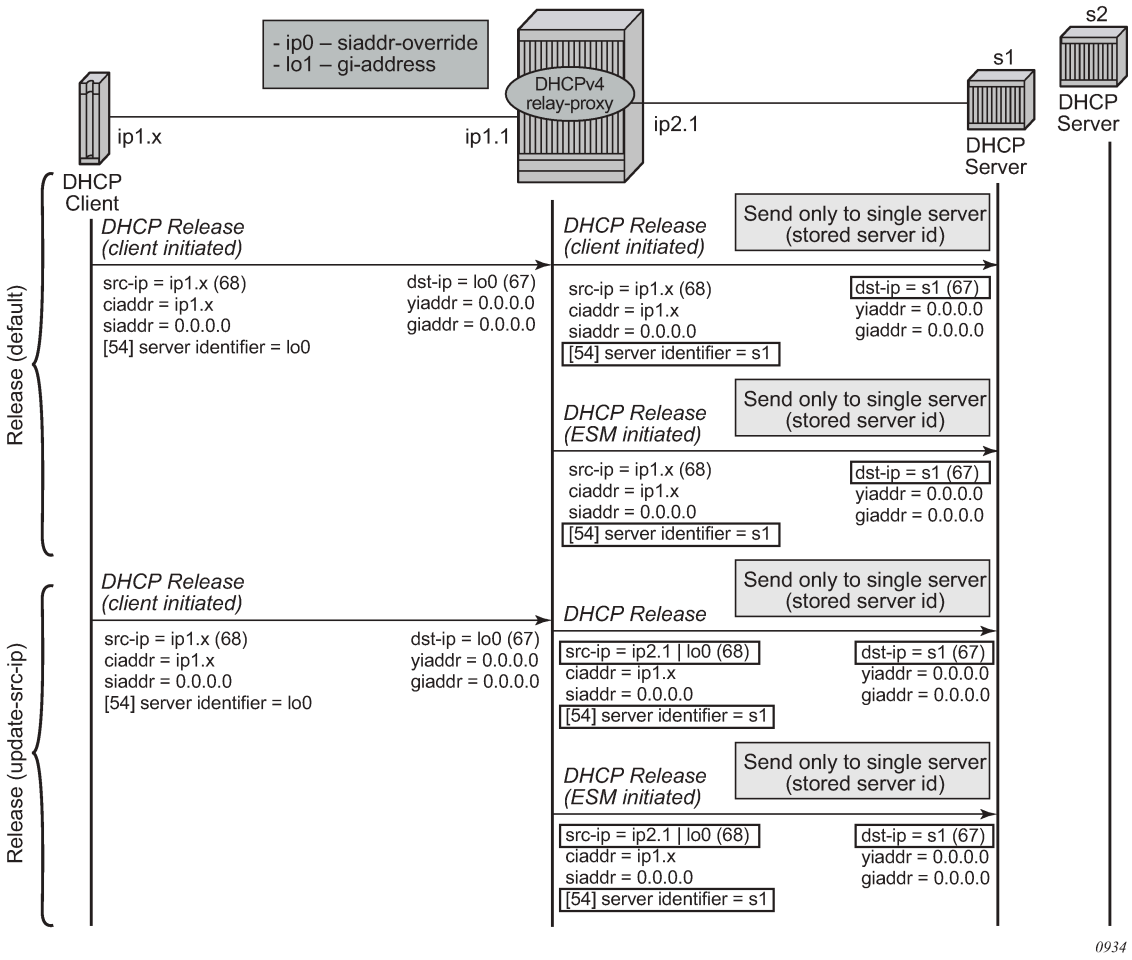


Figure 15: DHCP server IP address hiding, release shows the release in a relay proxy scenario with DHCP server address hiding enabled. The RELEASE message is sent only to the DHCP server owning the lease. Optionally, the source IP address can be updated.

Figure 15: DHCP server IP address hiding, release



Relay proxy can be enabled on subscriber group-interfaces and regular interfaces in an IES or VPRN service.

For retail subscriber interfaces, **relay-proxy** is configured at the **subscriber-interface dhcp** CLI context, as shown in the example that follows.

A relay proxy function is not supported with a double DHCPv4 relay (Layer 3 DHCPv4 relay in front of a 7750 DHCPv4 relay with **relay-proxy** enabled).

Configuration example:

```
config>service>vprn
        interface "lo0" create
            address 192.0.2.10/32
            loopback
        exit
        interface "lo1" create
            address 192.0.2.11/32
            loopback
        exit
        subscriber-interface "sub-int-1" create
            address 10.1.0.254/24
            group-interface "group-int-1-1" create
```

```
                dhcp
                    server 172.16.1.1
                    lease-populate 32767
                    relay-proxy release-update-src-ip siaddr-override 192.0.2.10
                    gi-address 192.0.2.11 src-ip-addr
                    no shutdown
                exit
            exit
        exit
```

3.2.3 DHCP lease split

The DHCP lease split function enables the use of shorter lease times toward the DHCP client than the lease time committed by the DHCP server. The DHCP relay lease split function can be used for:

- liveness detection of DHCP clients, so that addresses or prefixes of inactive clients can be reclaimed sooner without generating a high volume of control traffic toward the DHCP server
- BNG reachability verification by DHCP clients without the need for an additional protocol

3.2.3.1 DHCPv4

Figure 16: DHCPv4 relay lease split illustrates the lease split function for the initial connection of a DHCPv4 subscriber host.

1. The initial Discover, Offer, and Request messages are handled by the DHCP Relay Agent as usual.
2. Before forwarding the DHCP ACK message to the DHCP client, the DHCP Relay Agent replaces the long lease time (LT) committed by the DHCP server with a shorter lease time (LT_s).

   The lease split function is active when the configured short lease time (LT_s) is less than half the lease time committed by the server (LT).
3. The DHCP client tries to extend its lease at $T1_s$, the renewal time of the short lease time (LT_s).
4. The DHCP Relay Agent extends the lease on behalf of the DHCP server by sending a DHCP ACK with the short lease time (LT_s).
5. When the next DHCP client renew request is later than T1, the renewal time of the long lease time (LT), the Relay Agent forwards the request to the DHCP server and re-authenticates when applicable.
6. The DHCP Relay Agent replaces the long lease time (LT) committed by the DHCP server with a shorter lease time (LT_s).

Figure 16: DHCPv4 relay lease split

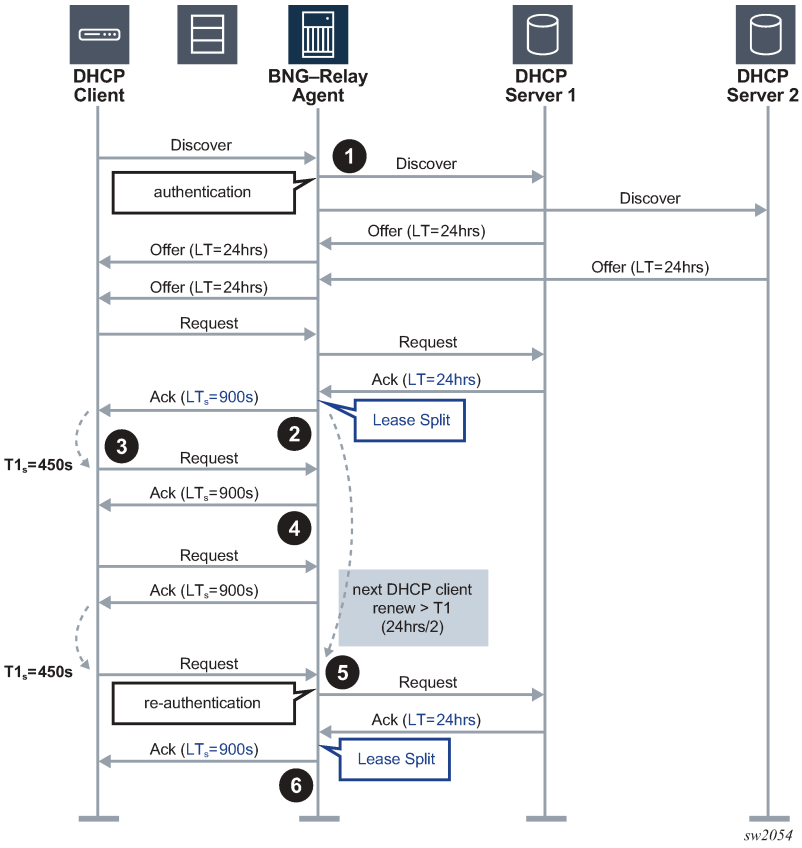


Figure 17: DHCPv4 lease split – DHCPv4 client disconnects ungracefully illustrates the lease split function when the DHCPv4 subscriber host disconnects ungracefully.

1. DHCPv4 lease split is active. When re-authenticating, the DHCP Relay Agent replaces the long lease time (LT) committed by the DHCP server with a shorter lease time (LT_s) in the DHCP ACK message.
2. The DHCP client tries to extend its lease at $T1_s$, the renewal time of the short lease time (LT_s).
3. The DHCP Relay Agent extends the lease on behalf of the DHCP server by sending a DHCP ACK with the short lease time (LT_s).
4. The DHCPv4 client disconnects ungracefully without sending a DHCP release message to the server.
5. When the short lease time (LT_s) expires, the subscriber host state is deleted from the system and the DHCP Relay Agent sends a release message to the server on behalf of the client. The maximum time before the IPv4 address becomes available for reuse after an ungraceful disconnect of the DHCP client is now determined by the short lease time (LT_s) instead of the long lease time (LT).

Figure 17: DHCPv4 lease split – DHCPv4 client disconnects ungracefully

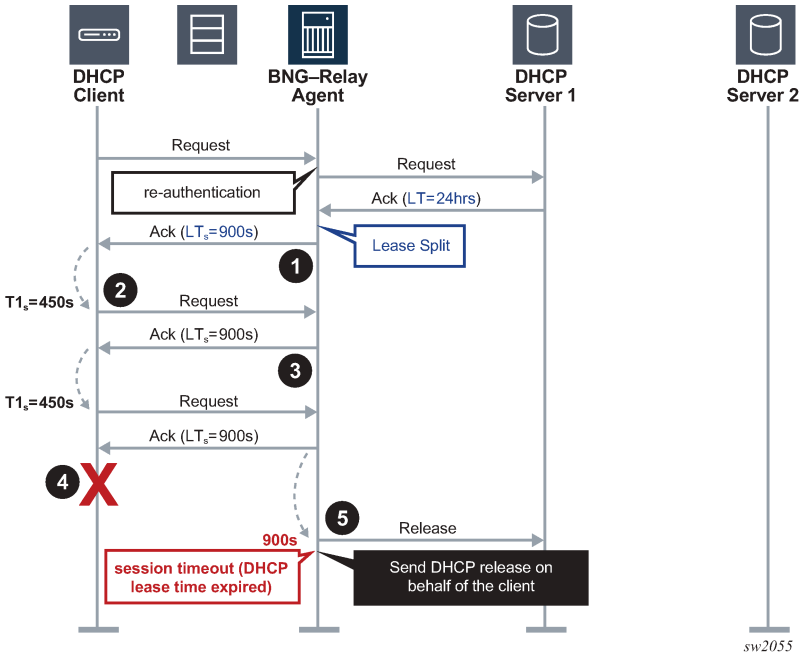


Figure 18: DHCPv4 lease split unreachable DHCPv4 servers illustrates the lease split function when the DHCPv4 servers become unreachable.

1. After the initial DHCPv4 session setup with lease split active, all DHCP servers become unreachable.

   The DHCP Relay Agent answers the DHCP client renew requests with the short lease time (LT_s) on behalf of the server until the next client renew request is later than T1, the renewal time of the long lease time (LT).

2. The DHCP Relay Agent re-authenticates (when configured) and forwards the next DHCP client renew request and all retransmissions to the DHCP server that committed the lease. No answer is received from the DHCP server.

3. Because the DHCP servers are unreachable, the client transitions to the rebinding state at $T2_s$, the rebinding time of the short lease time (LT_s). The DHCP Relay Agent re-authenticates (when configured) and broadcasts the DHCP client rebind request to all configured DHCP servers. No answer is received from any DHCP server.

4. The DHCP Relay Agent answers the first retransmission of the DHCP client rebind request with the short lease time (LT_s) on behalf of the server that committed the lease.

   The DHCP Relay Agent answers subsequent DHCP client renew requests with the short lease time (LT_s) on behalf of the server that committed the lease until either:

   - the remaining long lease time (LT_r) is less than the short lease time (LT_s) + 5 minutes. In this case, the call flow continues as described from Step 8 onward
   - the next client renew request is later than T2, which is the rebind time of the long lease time (LT)

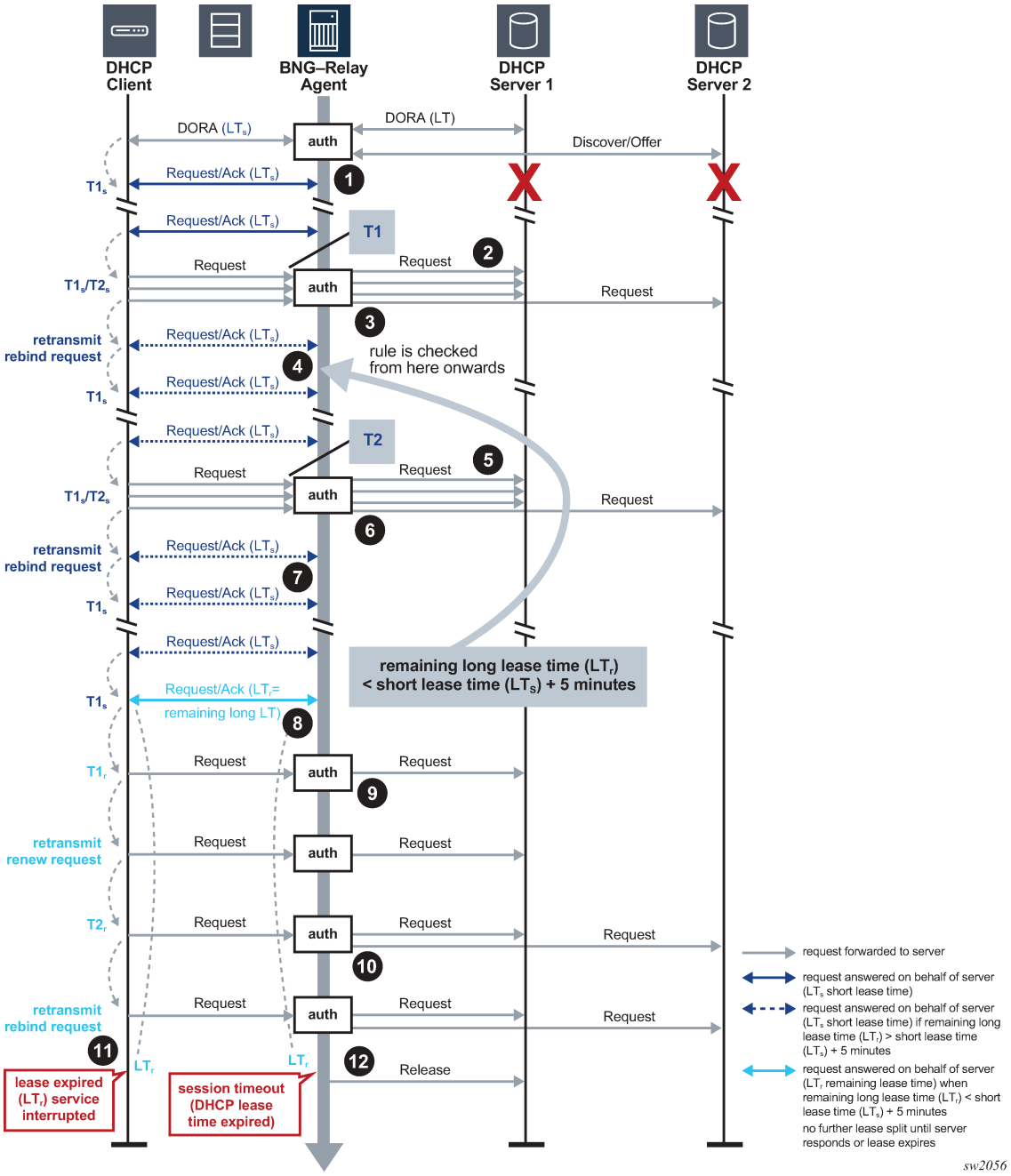
5. The DHCP Relay Agent re-authenticates (when configured) and forwards the next DHCP client renew request and all retransmissions to the DHCP server that committed the lease. No answer is received from the DHCP server.

6. Because the DHCP servers are unreachable, the client transitions to the rebinding state at $T2_s$, the rebinding time of the short lease time (LT_s). The DHCP Relay Agent re-authenticates (when configured) and broadcasts the DHCP client rebind request to all configured DHCP servers. No answer is received from any DHCP server.
7. The DHCP Relay Agent answers the first retransmission of the DHCP client rebind request with the short lease time (LT_s) on behalf of the server that committed the lease.

   The DHCP Relay Agent answers subsequent DHCP client renew requests with the short lease time (LT_s) on behalf of the server that committed the lease until the remaining long lease time (LT_r) is less than the short lease time (LT_s) + 5 minutes.
8. The DHCP Relay Agent answers the next DHCP client renew request with the remaining long lease time (LT_r) on behalf of the server that committed the lease.

   The DHCP Relay Agent no longer answers on behalf of the DHCP servers until one of the servers responds or until the lease expires.
9. The DHCP Relay Agent re-authenticates (when configured) and forwards the next DHCP client renew request (at $T1_r$, the renew time of the remaining long lease time (LT_r)) and all retransmissions to the DHCP server that committed the lease. No answer is received from the DHCP server.
10. The DHCP Relay Agent re-authenticates (when configured) and broadcasts the next DHCP client rebind request (at $T2_r$, the rebind time of the remaining long lease time (LT_r)) and all retransmissions to all configured DHCP servers. No answer is received from any DHCP server.
11. When the remaining long lease time (LT_r) expires, the DHCP client transitions to the init state and connectivity is lost.
12. When the long lease time (LT) expires, the subscriber host state is deleted from the system and the DHCP Relay Agent sends a release message to the server on behalf of the client.

Figure 18: DHCPv4 lease split unreachable DHCPv4 servers



To enable DHCPv4 lease split, configure DHCPv4 relay and administratively enable the proxy server. DHCPv4 lease split is active for a lease when the proxy server is enabled and when the configured proxy server lease-time value (the short lease time) is less than or equal to the renew time committed by the server (the long renew time) in the Option 58 Renewal (T1) Time Value or 50% of the lease time committed

by the server in the absence of DHCP Option 58 Renewal (T1) Time Value in the DHCP Ack message from the server.

```
configure service ies service-id subscriber-interface ip-int-name group-
interface ip-int-name
--- snip ---
    dhcp
        proxy-server
            lease-time min 15
            no shutdown
        exit
        server 192.0.2.3 192.0.2.4
        trusted
        gi-address 10.250.13.254 src-ip-addr
        no shutdown
    exit
```

Use the following show command to verify whether DHCPv4 lease split is active.

```
# /show service id 1000 dhcp lease-state detail
===============================================================================
DHCP lease states for service 1000
===============================================================================
--- snip ---
Remaining Lease Time : 0d 00:09:56 (Lease Split)
--- snip ---
Lease Info origin    : DHCP
--- snip ---
ServerLeaseStart     : 06/25/2020 07:35:54
ServerLastRenew      : 06/26/2020 14:42:14
ServerLeaseEnd       : 06/26/2020 15:42:14
--- snip ---
Lease-Time           : 1d 00:00:00
DHCP Server Addr     : 192.0.2.4
--- snip ---
```

Note:

- DHCPv4 lease split is supported for routed CO (IES and VPRN services, DHCP relay, and DHCP proxy) and bridged CO (VPLS service and DHCP snooping) deployment models.
- For bridged CO, the BNG does not answer on behalf of the server when the client is in rebinding state and the DHCP servers are unreachable. The DHCP client lease times out and the corresponding subscriber host state is deleted from the system when the short lease time LTs expires.

3.2.3.2 DHCPv6

The call flows for DHCPv6 lease split are similar to DHCPv4 lease split.

As shown in the following output, in a single DHCPv6 transaction, both IA_NA and IA_PD Identity Association (IA) types can be present, each with their associated timers (renew time T1, rebind time T2, preferred lifetime, and valid lifetime).

```
Option : IA_NA (3), Length : 40
        IAID : 0
        Time1: 1800 seconds
        Time2: 2880 seconds
        Option : IAADDR (5), Length : 24
```

```
          Address : 2001:db8:100::1
          Preferred Lifetime : 3600 seconds
          Valid Lifetime     : 4500 seconds
      Option : IA_PD (25), Length : 41
        IAID : 1
        Time1: 1800 seconds
        Time2: 2880 seconds
        Option : IAPREFIX (26), Length : 25
          Prefix : 2001:db8:d201::/64
          Preferred Lifetime : 3600 seconds
          Valid Lifetime     : 4500 seconds
```

DHCPv6 lease split actions are always identical for all leases in the transaction:

- DHCPv6 lease split is active for a lease when DHCPv6 relay lease split is enabled on the group interface or retail subscriber interface and when, for all IA_NA and IA_PD options in the transaction, the configured lease split valid lifetime (short lease time) is less than or equal to the following conditions:
  - the renew time T1 committed by the server (the long renew time) in the IA_NA or IA-PD Option
  - 50% of the preferred lifetime committed by the server in the IA_NA Address Option or IA_PD Prefix Option when the T1 value committed by the server equals zero
- When DHCPv6 lease split is active, the following values are updated for all IA types in the reply to the DHCPv6 client.
  - Valid lifetime (short lease time or short valid lifetime) = configured lease-split valid-lifetime
  - Preferred lifetime (short preferred lifetime) = configured lease-split valid-lifetime
  - T1 (short renew time) = 0.5 * configured lease-split valid-lifetime
  - T2 (short rebind time) = 0.8 * configured lease-split valid-lifetime

  When lease split is active, the short preferred lifetime and short valid lifetime are equal.
- Renew behavior when DHCPv6 lease split is active:
  - A client renew message is authenticated (when applicable) and relayed to the DHCP server when, for at least one of the IA options in the transaction, the next client renew message following the short renew time cycle is later than the long renew time (T1) committed by the server.
  - Otherwise, the client renew message is proxied (in other words, a renew reply is sent to the client on behalf of the server). Proxied client renew messages are not authenticated. For all IA types, the following values are included:
    - Valid lifetime (short lease time or short valid lifetime) = configured lease-split valid-lifetime
    - Preferred lifetime (short preferred lifetime) = configured lease-split valid-lifetime
    - T1 (short renew time) = 0.5 * configured lease-split valid-lifetime
    - T2 (short rebind time) = 0.8 * configured lease-split valid-lifetime

The following output shows an example of a DHCP6 relay configuration.

```
configure service ies service-id subscriber-interface ip-int-name group-
interface ip-int-name
--- snip ---
    ipv6
        dhcp6
            relay
                lease-split
                    valid-lifetime min 15
                    no shutdown
```

```
                exit
                server 2001:db8::1
                no shutdown
            exit
        exit
    exit
```

Use the following command to verify whether DHCPv6 lease split is active.

```
# /show service id 1000 dhcp6 lease-state detail
===============================================================================
DHCP lease states for service 1000
===============================================================================
--- snip ---
Remaining Lease Time : 0d 00:12:07 (Lease Split)
--- snip ---
Dhcp6 Server Addr    : 2001:db8::3
--- snip ---
Lease Info origin    : DHCP
ServerLeaseStart     : 12/15/2020 11:13:22
ServerLastRenew      : 12/15/2020 11:13:22
ServerLeaseEnd       : 12/15/2020 12:20:02
```

Note:

- DHCPv6 lease split is supported for routed CO DHCP relay deployment models.
- A Lightweight DHCPv6 Relay Agent (LDRA) in front of the DHCPv6 relay lease split is supported.

3.2.4 Subscriber identification using Option 82 field

Option 82, or the relay information option is specified in RFC 3046, *DHCP Relay Agent Information Option*, and allows the router to append some information to the DHCP request that identifies where the original DHCP request came from.

There are two sub-options under Option 82:

- *Agent Circuit ID Sub-option* (RFC 3046, section 3.1); this sub-option specifies data which must be unique to the box that is relaying the circuit
- *Remote ID Sub-option* (RFC 3046 section 3.2); this sub-option identifies the host at the other end of the circuit. This value must be globally unique

Both sub-options are supported and can be used separately or together.

Inserting Option 82 information is supported independently of DHCP relay. However, in a VPLS service (when DHCP Relay is not configured), DHCP snooping must be enabled on the SAP to be able to insert Option 82 information.

When the **circuit id sub-option** field is inserted, it can take following values:

sap-id
: the SAP index (only under an IES or VPRN service)

ifindex
: the index of the IP interface (only under an IES or VPRN service)

ascii-tuple

an ASCII-encoded concatenated tuple, consisting of [*system-name* | *service-id* | *interface-name*] (for VPRN or IES) or [*system-name* | *service-id* | *sap-id*] (for VPLS)

vlan-ascii-tuple

an ASCII-encoded concatenated tuple, consisting of the ascii-tuple followed by dot1p bits and dot1q tags

For VPRN, the ifindex is unique only within a VRF. The DHCP relay function automatically prepends the VRF ID to the ifindex before relaying a DHCP Request. VPRN is supported on the 7750 SR only.

When a DHCP packet is received with Option 82 information already present, the system can do one of three things. The available actions are:

Replace

On ingress, the existing **information-option** is replaced with the **information-option** parameter configured. On egress (toward the customer), the information option is stripped (per the RFC).

Drop

The DHCP packet is dropped and a counter is incremented.

Keep

The existing information is kept on the packet and the router does not add more information. On egress the information option is not stripped and is sent on to the downstream node.

In accordance with the RFC, the default behavior is to keep the existing information; except if the GIADDR of the packet received is identical to a local IP address on the router, then the packet is dropped and an error incremented regardless of the configured action.

The maximum packet size for a DHCP relay packet is 1500 bytes. If adding the Option 82 information would cause the packet to exceed this size, the DHCP client packet is discarded. This packet size limitation exists to ensure that there is no fragmentation on the end Ethernet segment where the DHCP server attaches.

In the downstream direction, the inserted Option 82 information should not be passed back toward the client (as per RFC 3046, *DHCP Relay Agent Information Option*). To enable downstream stripping of the Option 82 field, DHCP snooping should be enabled on the SDP or SAP connected to the DHCP server.

3.2.4.1 Trusted and untrusted

As specified in RFC 3046, *DHCP Relay Agent Information Option*, an SR OS Relay Agent discards a DHCP packet where the gi address is set to 0.0.0.0 but with a Relay Agent Information Option 82 present, unless it arrives on a "trusted" circuit.

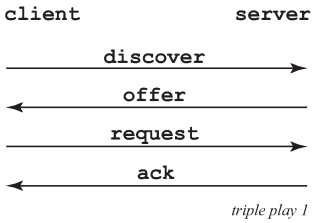
The **dhcp trusted** command enables the Relay Agent to forward such DHCP requests. The gi address is updated according the **gi-address** configuration on the interface.

3.2.5 DHCP snooping

This section discusses the Nokia routers acting as a Broadband Subscriber Aggregator (BSA) with Layer 2 aggregation toward a Broadband Subscriber Router (BSR).

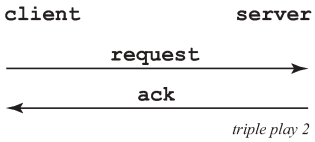
A typical initial DHCP scenario is shown in Figure 19: Initial DHCP scenario.

Figure 19: Initial DHCP scenario



But, when the client already knows its IP address, it can skip the discover, as shown in Figure 20: DHCP scenario with known IP address.

Figure 20: DHCP scenario with known IP address



The BSA can copy packets designated to the standard UDP port for DHCP (port 67) to its control plane for inspection, this process is called DHCP snooping.

DHCP snooping can be performed in two directions:

1. From the client to the DHCP server (Discover or Request messages):
   - to insert Option 82 information (when the system is not configured to do DHCP Relay), see Subscriber identification using Option 82 field.
   - to forward DHCP requests to a RADIUS server first, and not send them to the DHCP server unless the RADIUS server confirms positive identification.

   For these applications, DHCP snooping must be enabled on the SAP toward the subscriber.
2. From the DHCP server (ACK messages):
   - to remove the Option 82 field toward the client
   - to build a dynamic DHCP lease state table for security purposes, see section DHCP lease state table
   - to perform Enhanced Subscriber Management, see Triple Play Enhances Subscriber Management

   For these applications, DHCP snooping must be enabled on both the SAP and SDP toward the network and the SAP toward the subscriber.

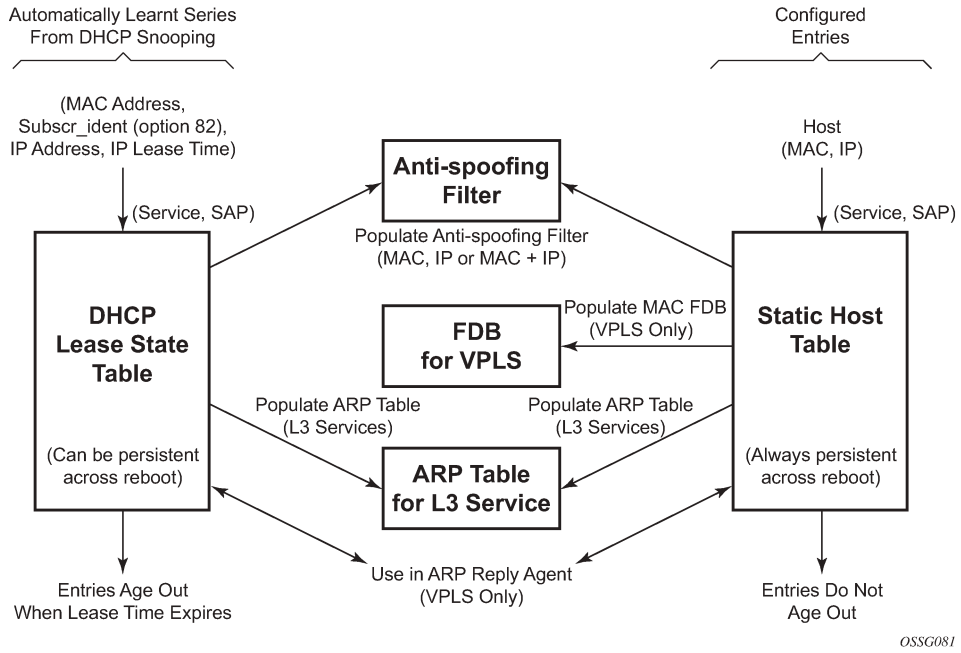
   A major application for DHCP response snooping in the context of Triple Play is security: A malicious user A could send an IP packet (for example, requesting a big video stream) with as source the IP address of user B. Any return packets would be sent to B, and therefore potentially jam the connection to B.

   As the snooped information is coming straight from the operator's DHCP server, it is considered reliable. The BSA and BSR can use the snooped information to build anti-spoofing filters, populate the ARP table, send ARP replies, and so on.

3.2.6 DHCP lease state table

The DHCP lease state table has a central role in the BSA operation, as shown in Figure 21: DHCP lease state table. For each SAP on each service it maintains the identities of the hosts that are allowed network access.

Figure 21: DHCP lease state table



When the command **lease-populate** is enabled on a SAP, the DHCP lease state table is populated by snooping DHCP ACK messages on that SAP, as described in the DHCP snooping section.

Entries in the DHCP lease state table remain valid for the duration of the IP address lease. When a lease is renewed, the expiry time is updated. If the lease expires and is not renewed, the entry is removed from the DHCP lease state table.

For VPLS, DHCP snooping must be explicitly enabled (using the **snoop** command) on the SAP or SDP where DHCP messages requiring snooping ingress the VPLS instance. For IES interfaces and VPRN IP interfaces (VPRN is supported on the 7750 SR only), using the **lease-populate** command also enables DHCP snooping for the subnets defined under the IP interface. Lease state information is extracted from snooped or relayed DHCP ACK messages to populate DHCP lease state table entries for the SAP or IP interface.

For IES and VPRN services, if ARP populate is configured, no statics ARPs are allowed. For IES and VPRN services, if ARP populate is not configured, then statics ARPs are allowed.

The retained DHCP lease state information representing dynamic hosts can be used in a variety of ways:

- To populate a SAP based anti-spoof filter table to provide dynamic anti-spoof filtering. Anti-spoof filtering is only available on VPLS SAPs, or IES IP, or VPRN IP interfaces terminated on a SAP.
- To populate a VPLS SAP-based **arp-reply-agent** table to provide dynamic ARP replies using the dynamic hosts IP assigned IP address and learned MAC address. The ARP reply agent functionality is only available for static and dynamic hosts associated with a VPLS SAP. **arp-reply-agent** is supported on the 7450 ESS only.

- To populate the system's ARP cache using the **arp-populate** feature. The **arp-populate** functionality is only available for static and dynamic hosts associated with IES and VPRN SAP IP interfaces.
- To populate managed entries into a VPLS forwarding database . When a dynamic host's MAC address is placed in the DHCP lease state table, it automatically populates into the VPLS forwarding database associated with the SAP on which the host is learned. The dynamic host MAC address overrides any static MAC entries using the same MAC and prevent learning of the MAC on another interface (implicit MAC pinning on the 7450 ESS). Existing static MAC entries with the same MAC address as the dynamic host are marked as inactive but not deleted. If all entries in the DHCP lease state associated with the MAC address are removed, the static MAC may be populated. New static MAC definitions for the VPLS instance can be created while a dynamic host exists associated with the static MAC address. To support the Routed CO model, see to Routed Central Office (CO).
- To support Enhanced Subscriber Management, see RADIUS Authentication of Subscriber Sessions.

If the system is unable to execute any of these tasks, the DHCP ACK message is discarded without adding a new lease state entry or updating an existing lease state entry; and an alarm is generated.

3.2.7 DHCP and Layer 3 aggregation

3.2.7.1 DHCPv4 snooping

The default mode of operation for DHCP snooping is that the DHCP snooping agent instantiates a DHCP lease state based on information in the DHCP packet, the client IP address and the client hardware address.

The mode of operation can be changed for DHCP snooping so the Layer 2 header MAC address is used instead of the client hardware address from the DHCP packet for the DHCP lease state instantiation. This mode is selected by enabling the l2-header in the **lease-populate** command at the DHCP level. Because SR OS routers do not have the ability to verify the DHCP information (both the **src-ip** and **src-mac** of the packet are those of the previous relay point) anti-spoofing must be performed at the access node before the SR OS routers. This mode provides compatibility with MAC concentrator devices, and cable modem termination system (CMTS) and WiMAX Access Controller (WAC).

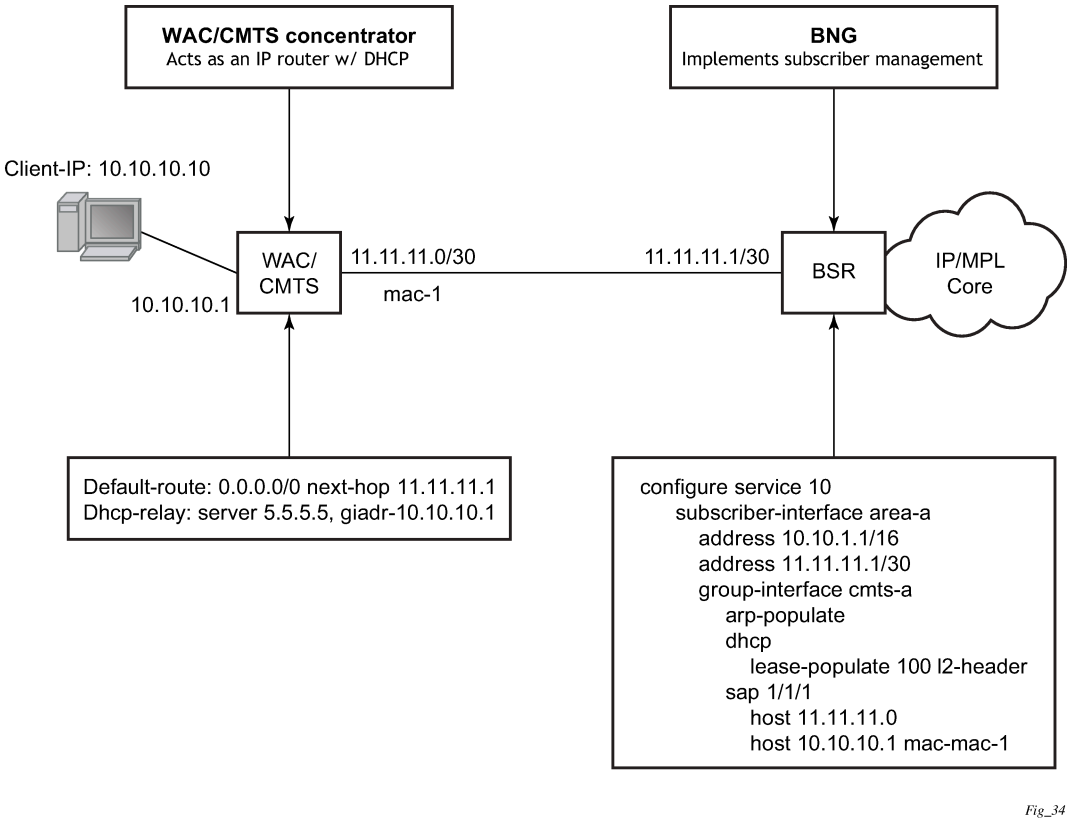
A configuration example of a cable/wireless network together with subscriber management is shown in Figure 22: CMTS/WAC network configuration example. The subnet used to connect to the CMTS/WAC must be defined as a subnet in the subscriber interface of the Layer 3 CO model under which the hosts is defined. This means that all subscriber lease states instantiated on BSR must be from a "local" subscriber-subnet, even if those are behind the router, as there is no additional Layer 3 route installed pointing to them.

The important items to notice are static hosts at the subscriber interface side:

- IP-only static host pointing to CMTS/WAC WAN link is needed to allow BSR to reply to ARP requests originated from CMTS/WAC.
- IP-MAC static host pointing to CMTS/WAC access-facing interface is required to provide BSR with an arp entry for the DHCP relay address.

When dual-homing is used the CMTS/WAC may be configured with the same MAC for both upstream interfaces. If that is not possible the BSR can be configured with an optional MAC address. The BSR then uses the configured MAC address when instantiating the DHCP lease states.

Figure 22: CMTS/WAC network configuration example



3.2.7.2 DHCPv6 snooping

Like DHCPv4, the subscriber interface SAP must be on the datapath between the subscriber host and the DHCPv6 server. The SAP snoops the DHCPv6 message exchanged between the server and the client. An ESM host is created upon snooping a "reply" message from the DHCPv6 server.

DHCPv6 messages differ from a DHCPv4 messages because it is not mandatory to have the client MAC inside the header or options. The DUID in the DHCPv6 option can be a random generated number instead of the subscriber host's MAC. The source MAC of the DHCPv6 Ethernet header cannot be used either, as a Layer 3 aggregation network replaces the client's MAC with routing. From the perspective of the BNG, all DHCPv6 message from the same downstream router has the same source MAC. By default, the BNG use the DHCPv6 Ethernet header source MAC as a host entry identifier. Therefore, it is mandatory to use the interface ID in addition to the source MAC to identify a host individually. If the interface ID is unavailable, it is possible to use python to copy another unique ID, such as DUID or remote ID, into the interface ID. The interface ID option must be on the same level as the relay forward header. Together, the interface ID and the DHCP relay MAC address are used as an identifier internally in the BNG.

If the interface ID option is on the subscriber native DHCP message (such as solicit), it is simply ignored.

The downstream router must resolve the BNG MAC before it is able to route traffic to the BNG. Traditionally, a BNG sends router advertisement to directly-connected hosts to help them resolve their default gateway and MAC address. However, routers differ from hosts and neighbor advertisements are used to resolve the neighbor's MAC instead. The downstream router has two options when programming

the BNG as the next hop. It can either use the BNG subscriber interface link-local address or the subscriber interface GUA address. If an IPv6 prefix was configured on the BNG subscriber interface, then the downstream router must use the BNG link local as the next hop. If the subscriber interface is configured as an IPv6 address, then the downstream router can configure the GUA or the link-local as the next hop. To forward traffic bidirectionally, the downstream router interface must be modeled as a static host with both IP and MAC. IP-only static host is not supported.

The DHCP relay agent use one of its interface as a IP source address in the DHCP relay-forward message. The BNG forwards a DHCP relay-reply message from the DHCP server back to the relay agent using that exact same source IP address. There are restrictions for the IP source address used by the DHCP relay agent and it depends if the relay agent is a few hops way or is directly connected to the BNG. In the case where relay agent is a few hops away, the source address used by the relay agent must not fall under the subnet or prefix range configured on the subscriber interface. For example, the loopback or the egress interface address of the DHCP relay agent can be used instead. To forward the DHCP6 relay-forward message to the relay agent, simply add a static route for the relay agent source IP address. The static route has the static IPv6 host as the next hop. In the case where the relay agent is directly connected to the BNG, there are two options. In the first option, the IPv6 static host configured on the BNG is an interface on the relay agent. If the relay agent use this as the relay-forward source address, no additional configuration is required on the BNG to forward the relay-reply to the relay-agent. The other option is to use an interface address on the relay agent which does not fall under the subnet or prefix under the subscriber interface. Like the scenario where the relay-agent is a few hops away, a static route is required to forward the DHCP relay-reply message back to the relay agent. Again, the static route must use the IPv6 static host as the next hop.

A default host is supported for IPv6 host as well. It is generally used as a failover mechanism where the host can continue to forward traffic without a host entry on the backup BNG.

For the IPv6 ESM host, it is mandatory that each host have a unique /64 prefix. Service providers who need to share the /64 prefix among multiple WAN host can use the DHCPv6 filter **bypass-host-creation na** option. All bypass hosts in general require a default host creation as well.

While ESM hosts are subject to QoS and filters rules specified in sub-profile and sla-profile, default-host follows the QoS and filters specified directly on the subscriber SAP.

DHCP6 filters perform actions based on the options inside the relay-forward DHCP message. The options must be set on the innermost level, such as, DHCP solicit. The filter ignores those options set on relay-forward levels.

ESMv6 host created by DHCP snooping is not supported with the following:

- WLAN-GW
- DHCPv6 Proxy
- Wholesale or Retail
- SHCV
- Layer 2–aware NAT
- GRT leaking or Extranet

3.2.8 Local DHCP servers

3.2.8.1 Overview

A local DHCP server functions only if there is a relay agent (gateway) in front of it. Either a GI address is needed to find a subnet or Option 82, which is inserted by the relay, to perform authentication in the local-user-db.

The local DHCP server must be configured to assign addresses in one of the following ways:

1. Use a local user database authentication (**user-db local-user-db-name**)

   The host is matched against the specified local user database. A successful user lookup should return information about one of the following valid addresses:

   - **fixed IP address**

     The IP address should not overlap with the address ranges configured in the local DHCP server.

   - **pool name**

     A free address of any subnet in that pool is offered.

   - **use-gi-address** [**scope subnet** | **pool**]

     The GI address is used to find a matching subnet. When **scope subnet** is configured, an address is allocated in the same subnet as the GI address only. When scope is pool, an address is allocated from any subnet within a local pool when that pool has been selected based on matching the "giaddr" field in the DHCP message with any of the configured subnets in the pool.

   - **use-pool-from-client**

     The pool name specified in the DHCP client message options and added by the DHCP relay agent is used. A free address of any subnet in that pool is offered.

   When no valid address information is returned from the local user database lookup, no IP address is offered to the client.

2. Without local user database authentication (**no user-db**).

   One or both address assignment options must be configured:

   - Use a pool name (**use-pool-from-client**)

     The pool name specified in the DHCP client message options and added by the DHCP relay agent is used. A free address of any subnet in that pool is offered.

   - Use the gi address (**use-gi-address** [**scope subnet** | **pool**])

     The gi address is used to find a matching subnet. When the scope is **subnet**, an address is allocated in the same subnet as the gi address only. When scope is pool, an address is allocated from any subnet within a local pool when that pool has been selected based on matching the "giaddr" field in the DHCP message with any of the configured subnets in the pool.

When both options are configured and a pool name is specified in the DHCP client message options, then the **use-pool-from-client** option has precedence over the **use-gi-address** option.

Note: The local DHCP server does not allocate any address if none of the above options are configured (**no user-db**, **no use-gi-address**, **no use-pool-from-client**).

Options and identification strings can be defined on several levels. In principle, these options are copied into the DHCP reply, but if the same option is defined several times, the following precedence is taken:

1. user-db host options
2. subnet options

3. pool options
4. from the client DHCP request

A local DHCP server must be bound to a specified interface by referencing the server from that interface. The DHCP server is then addressable by the IP address of that interface. A normal interface or a loopback interface can be used.

A DHCP client is defined by the MAC address and the circuit ID. This implies that for a specified combination of MAC and circuit ID, only one IP address can be returned. The same address is returned if a re-request is made.

Typically, the DHCP server can be configured to perform as follows:

- When a **user-db** is specified, a host lookup is performed in the local-user-db and the local user-db host defines how to get the IP address or pool to get the IP address from:
  - a fixed IP address
  - a pool

    free address of any subnet in that pool is offered
  - **use-pool-from-client**

    The pool name is taken from Option 82 vendor-specific sub-option. (If not present, proceed to the next bullet.)
  - **use-gi-address**

    The gi-address is used to find a matching subnet.
- If no IP address, subnet, or pool is found (see the points above), no IP address is offered to the client. If a subnet is found, an available address from the subnet is offered to the client. If a pool is found, an available address from any subnet of the pool is offered to the client.

Note: The local DHCP server does not allocate any address if none of the above options are configured (**no user-db**, **no use-gi-address**, **no use-pool-from-client**).

3.2.8.2 Local DHCP server support

Local DHCP servers provide a standards-based full DHCP server implementation which allows a service provider the option of decentralizing the IP address management into the network. Local DHCP servers are supported on 7750 SR-7, and 7750 SR-12 models. The 7450 ESS routers use DHCP relay and proxy DHCP server functionality.

Three applications are targeted for the local DHCP server:

- subscriber aggregation in either a single node (routed CO) or TPSDA
- business services

  A server can be defined in a VPRN service and associated with different interfaces. Locally attached hosts can get an address directly from the servers. Routed hosts receives addresses through a relay point in the customer's network. This option is supported for the 7750 SR only.
- PPP clients

  Either in a single node or a separate PPPoE server node and a second DHCP server node. The PPPoE server node may be configured to query the DHCP server node for an address and options to provide to the PPPoE client. The PPPoE server provides the information in PPP format to the client.

DHCP server scenarios include:

- DHCP servers can be integrated with Enhanced Subscriber Management (ESM) for DHCP clients, DHCP relays and PPPoE clients.
- A stand-alone DHCP server can support DHCP clients and DHCP relays.
- IPv4 is supported. DHCP servers provide increased management over the IPv4 address space across its subscriber base, with support for public and private addressing in the same router, including overlapped private addressing in the form of VPRNs in the same SR-series router.

DHCP servers are configurable and can be used in the bridged CO, routed CO, VPRN wholesaler, and dual-homing models.

3.2.9 DHCPv6

In the stateful auto-configuration model, hosts obtain interface addresses and configuration information and parameters from a server. Servers maintain a database that keeps track of which addresses have been assigned to which hosts. The stateful auto-configuration protocol allows hosts to obtain addresses, other configuration information or both from a server.

3.2.9.1 DHCPv6 relay agent

When the server unicast option is present in a DHCP message from the server, the option is stripped from the message before sending to the clients.

3.2.9.2 DHCPv6 prefix options

The prefix delegation options described in RFC 3633, *IPv6 Prefix Options for Dynamic Host Configuration Protocol (DHCP) version 6*, provide a mechanism for automated delegation of IPv6 prefixes using DHCP. This mechanism is intended for delegating a long-lived prefix from a delegating router to a requesting router, across an administrative boundary, where the delegating router does not require knowledge about the topology of the links in the network to which the prefixes are assigned. For example, the delegating router can get a /48 prefix by DHCPv6 and assign /64 prefixes to multiple requesting routers. Prefix delegation is supported for the delegating router (not the requesting router).

3.2.9.3 Neighbor resolution with DHCPv6 relay

This is like ARP populate for IPv4. When it is turned on, an SR OS needs to populate the link-layer address to IPv6 address mapping into the neighbor database based on the DHCPv6 lease information received.

If the IPv6 address of the host does not belong to the subnets of the interface, such a neighbor record should not be created. This could happen when there is a downstream DHCPv6 relay router or prefix delegation requesting router.

3.2.9.4 DHCPv6 lease persistency

DHCPv6 lease persistency is supported.

The following features are enabled:

- DHCPv6 lease information is reconciled to the standby.
- DHCPv6 lease information can be stored on a flash card.
- When rebooted, DHCPv6 lease information stored on a flash card can be used to re-populate the DHCPv6 table as well as the neighbor database if neighbor-resolution is enabled.

3.2.9.5 Local proxy neighbor discovery

Local proxy neighbor discovery is like local proxy ARP. It is useful in the residential bridging environment where end users are not allowed to talk to each other directly.

When local proxy ND is turned on for an interface, the router:

- responds to all neighbor solicitation messages received on the interface for IPv6 addresses in the subnets unless disallowed by policy
- forwards traffic between hosts in the subnets of the interface
- drops traffic between hosts if the link-layer address information for the IPv6 destination has not been learned

3.2.9.6 IPv6oE hosts behind bridged CPEs

This feature adds support for dual-stack IPoE hosts behind bridged clients, receiving globally-routable address using SLAAC or DHCPv6. The feature also provides configurable support for handing out /128 addresses to bridged hosts from same /64 prefix or a unique /64 prefix per host. Bridged hosts that share the same /64 prefix are required to be all SLAAC hosts or DHCPv6 hosts and are required to be associated with the same SAP. For SLAAC based assignment, downstream neighbor-discovery is automatically enabled to resolve the assigned address.

3.2.9.7 IPv6 link-address based pool selection

This feature provides the capability to select prefix pools for WAN or PD allocations based on configured link addresses. The scope of selection is the pool or a prefix range within the pool.

3.2.9.8 IPv6 address and prefix stickiness

This feature extends lease identification criterion beyond DUID (default) for DHCPv6 leases held in the lease database for a configured period after the lease times out. DHCPv6 leases can be held in the lease database for a configurable period, after the lease time has expired. A large configured timeout value allows for address and prefix "stickiness". When a subscriber requests a lease through DHCPv6 (IA_NA or PD), existing lease is looked-up and handed out. The lease identification match criterion has been extended beyond DUID to also include interface-id, or interface-id and link-local address.

3.2.9.9 IPv4/IPv6 linkage for dual-stack hosts or Layer 3 RGs

In case of dual-stack Layer 3 RGs or dual-stack hosts behind Layer 2 CPEs, it is beneficial to have the ability to optionally link Ipv6oE host management to DHCP state for v4. This feature provides configurable support to use circuit-id in DHCPv4 option-82 to authenticate DHCPv6. Also, a SLAAC host is created

based on DHCPv4 authentication if RADIUS returns IPv6 framed-prefix. IPv6oE host is deleted when the linked IPv4oE host is deleted. In addition, with v4 and v6 linkage configured, sending of unsolicited unicast RA toward the client can be configured when v4 host state is created and IPv6 is configured for the client. The linkage between IPv4 and IPv6 is based on SAP and MAC address. The sharing of the circuit-id from DHCPv4 for authentication of DHCPv6 (or SLAAC) allows the SR series router to work around lack of support for LDRA on Access-nodes.

3.2.9.10 Host connectivity checks for IPv6

This feature provides support to perform SHCV checks on the global unicast address (assigned by SLAAC or DHCPv6 IA_NA) and link-local address of a Layer 3 RG or a bridged host. SHCV uses IPv6 NS and NA messages. The configuration is like IPv4 support in SHCV. The **host-connectivity-check** command is extended to be configured for IPv6 or both IPv4 and IPv6.

3.2.10 Lease query

A lease query by client-id as defined in RFC 5007, *DHCPv6 Lease Query*, is supported for Local DHCPv6 servers. For security reasons this must be explicitly enabled using the CLI command **allow-lease-query**. The user identification option must be set to DUID (default value) for lease-query to work. Lease query by address is not supported. It is not possible to filter out leases with the link address, the server always returns all addresses for a client. The Relay Data and Client Link options are not supported and are not returned.

3.2.11 DHCPv6 to server option

A DHCPv6-relay **user-db** can be configured on an IES or VPRN IP interface. This allows the insertion of a **to-client** and **to-server** option on the client's DHCPv6 message. The VPRN or IES IP interface must be the first relay agent for the subscriber. The interface must also have **lease-populate** enabled. The interface can store up to eight leases per DHCPv6 (relay) reply. If the interface contains both RADIUS and user-db, RADIUS always takes precedence. The **to-client** and **to-server** option is inserted in the inner most level of the DHCPv6 packet. ESM IPoE subscriber can also use the **user-db** to insert **to-server** option. IPoE ESM subscribers are not limited to only **to-client** and **to-server** option and can use all other parameters configurable under **config**>**subscr-mgmt**>**loc-user-db**>**ipoe**>host> in LUDB.

3.2.12 Allow client ID change for DHCPv6

The client IDs of incoming DHCPv6 packets are strictly checked. Packets with different client IDs from an existing lease state are treated as suspicious and discarded.

Execute the following commands on a group interface for IES/VPRN or subscriber interface for retailer IES/VPRN, to configure client ID changes for interoperability:

- **MD-CLI**

```
configure service ies string subscriber-interface string group-interface string ipv6 dhcp6
 allow-client-id-change
```

```
configure service vprn string subscriber-interface string group-interface string ipv6 dhcp6
 allow-client-id-change
```

```
configure service ies string subscriber-interface string ipv6 dhcp6 allow-client-id-change
```

```
configure service vprn string subscriber-interface string ipv6 dhcp6 allow-client-id-change
```

- **Classic CLI**

```
configure service ies subscriber-interface group-interface ipv6 dhcp6 allow-client-id-change
```

```
configure service vprn subscriber-interface group-interface ipv6 dhcp6 allow-client-id-
change
```

```
configure service ies subscriber-interface ipv6 dhcp6 allow-client-id-change
```

```
configure service vprn subscriber-interface ipv6 dhcp6 allow-client-id-change
```

3.2.13 Flexible host identification in LUDB based on DHCPv4/v6 options

Host identification plays a critical role during the assignment of the parameters to the host through LUDB. The parameters that can be assigned to the subscriber host can range from the IP addressing parameters and the subscriber identification string all the way to the parameters that define the service to which the subscriber is entitled.

LUDB access in the context of IPoE hosts is triggered by DHCP messages passing through the interface on which the LUDB access is configured. This is true regardless of the direction of the DHCP message flow (ingress/egress).

The parameters that define the characteristics of the host are represented by an LUDB host *entry*. The parameters in the LUDB entry can be unique for each individual host, or they can be shared for a group of hosts. In the former case, the identification field for the LUDB host entry must be host specific while in the latter case the identification field for LUDB host entry could be derived from DHCP options that are common to a set of host.

The host identification in the LUDB can be based on a fixed set of predefined fields within the 7750 SR and 7450 ESS. If this predefined set of fields is not flexible enough, a custom identification field can be constructed from the DHCP options that are processed by the Python script. When this custom identifier is constructed, its value can be preserved for the duration of the DHCP transaction and it is used by the LUDB for the host identification.

An example of how this can be used is the following:

A Python script is installed in 7750 SR and 7450 ESS. This Python script intercepts incoming DHCP messages on the access side (Discover/Solicit/Request/Renew/Rebind) and consequently creates a host identification string based on DHCP options in the packet. This string then is cached and used for host identification in LUDB in both directions (access ingress and network ingress).

This functionality is supported for DHCPv4/DHCPv6 hosts.

3.2.14 DHCP caching

Subscriber host identification through LUDB is performed upon the arrival of the incoming DHCP messages on both, the access and the network side, while the host instantiation and ESM string assignment is performed only during the processing of the DHCP ACK/Reply messages. In other words, if Python without the caching is used for subscriber host identification and classification (into the correct service class by means of deriving ESM strings), the DHCP options required for host identification must be present in all DHCP messages, even the ones sent by the DHCP servers. However, DHCP servers are not required to echo DHCP options sent by the clients and relay-agents. Consequently, the missing options from the server side would cause the subscriber host instantiation to fail.

To remedy this situation and cover all deployments models (even the ones where the DHCP options are not echoed back by the DHCP servers), a caching mechanism is introduced whereby the results of the Python processing on ingress access are locally stored in 7750 SR and 7450 ESS. This ensures that the information about the subscriber host is readily available when the DHCP packet from the DHCP server arrives. Furthermore, because we already have the cached information, no additional Python processing on the network ingress is needed.

The caching is performed in a DHCP Transaction Cache (DTC), which is accessible to Python and to the ESM module. Python writes the result of its processing to it and the Enhanced Subscriber Management (ESM) module within 7750 SR and 7450 ESS can access those results.

The cache entries are relatively short lived, with the lifetime of a DHCP transaction. DHCP transaction is defined as a pair of DHCP messages that have the same DHCP transaction ID number (<Discover, Offer>, <Request, Ack>, <Solicit, Advertise>, <Request, Reply>, <Renew, Ack>, and so on).

3.2.15 Flexible creation of DHCPv4/6 host parameters

One of the facilities for flexible creation and assignment of subscriber host parameters is through Python scripting.

There are two models that allow assignment of the subscriber hosts parameters based on the Python processing, one without the utilizing the internal cache (DTC) and the other with the internal cache (DTC).

- Without utilizing the DTC, Python can process options in DHCP ACK message, derive the subscriber host parameters based on those options and consequently insert those parameters in a pre-configured DHCP option (defined in *sub-ident-policy*). The ESM module can be then instructed to extract those parameters and consequently instantiate the host with correct service levels. The drawback of this solution is that the DHCP server may not return all DHCP (v4 and v6) options that clients and relay-agents originally transmitted. Because those options are needed for subscriber parameter determination (but may be absent in the DHCP ACK/REPLY messages when the Python script is run), this solution falls short of covering all deployment cases. In addition, the range of parameters that can be assigned to a subscriber host in this fashion is smaller than the set of parameters utilizing the DTC.
- The internal cache (DTC) allows us to store the result of Python processing. The result is stored during the lifetime of the DHCP transaction. This method of string assignment does not rely on the DHCP server ability to return client's options, DHCPv4 and DHCPv6.

Parameters (ESM strings, IP addresses, and so on) present in the DTC have priority over any other source that is providing overlapping parameters when it comes to ESM processing. In other words, if the same

parameter is provided by DTC (Python), LUDB and RADIUS, the one provided by DTC is in effect. This prioritization occur automatically without any additional CLI.

For example, if the IPv4 address is provided by DTC during DHCP Discovery processing, then the mode of operation for this host is proxy-to-dhcp (ESM terminates DHCP, without going to the server), regardless of whether the IP address is also provided by LUDB or RADIUS.

This functionality is supported for DHCPv4/DHCPv6 hosts.

3.2.16 Python DTC variables and API

The following are the Python variables and APIs related to DTC:

Subscriber Host Identification

alc.dtc.derivedId

A read or write (from the Python perspective) string to store the LUDB lookup key for subscriber host identification. This key is derived from the contents of the packet. This string is used as a **match** criteria in LUDB. The derived-id can only be used when the lookup is performed in ESM. If the LUDB is attached to the local DHCP server, then the lookup based on the derived-id cannot be performed as the DHCP server has no means to derive such an ID from the DHCP message.

Caching Any Data During the Lifetime of a Transaction

alc.dtc.store(key,value)

The operator can store any data needed in one or more entries. The key can be any arbitrary string (printable ASCII characters), up to 32 bytes. The value part is 'unlimited' (memory permitting) in size.

alc.dtc.retrieve(key)

Retrieve data from the DTC. The key must be an existing key, which is a string consisting of printable ASCII characters, up to 32 bytes.

For example, this can be used to cache the DHCP options that the client inserts but the server does not echo back. Those options can still be retrieved in 7750 SR and 7450 ESS by cache in case that their presence is needed for any reason.

The lifespan of the cached data is tied to a DHCP transaction (a pair or corresponding DHCP messages flowing in opposite direction).

ESM Related Parameters (ESM strings, routing context)

DTC provides an API to supply a subset of configuration parameters that can otherwise come from RADIUS and LUDB and are used by the ESM code to setup the subscriber host.

DTC parameters as defined below should not be considered as DHCP options that can be blindly returned to the DHCP client, but instead they should be considered as real configuration settings. For example, the lease-time option is used in LUDB to enforce the lease time for the client. As such, the ESM keeps state of the lease-time. The following parameters can be used to setup a subscriber host:

alc.dtc.setESM (key-from-below, value)

Store data that is used by ESM. This data is write-only.

The keys are predefined (only these can be used) and are described in Table 2: ESM-related Python variables . These keys are read-only static variables.

The LUDB column indicates the configuration option under the **config**>**subscr-mgmt**>**loc-user-db**>**ipoe**>**host** context in LUDB.

Table 2: ESM-related Python variables

DTC variable	Type	LUDB	RADIUS attribute	Comment
alc.dtc.subIdent	string	identification-strings >subscriber-id	Alc-Subsc-ID-Str	—
alc.dtc.subProfileString	string	identification-strings >sub-profile-string	Alc-Subsc-Prof-Str	—
alc.dtc.slaProfileString	string	identification-strings >sla-profile-string	Alc-SLA-Prof-Str	—
alc.dtc.spiSharingGroupId	integer	identification-strings >spi-sharing-group-id	Alc-SPI-Sharing-Id	—
alc.dtc.ancpString	string	identification-strings >ancp-string	Alc-ANCP-Str	—
alc.dtc.appProfileString	string	identification-strings >app-profile-string	Alc-App-Prof-Str	—
alc.dtc.intDestId	string	identification-strings >inter-dest-id	Alc-Int-Dest-Id-Str	—
alc.dtc.catMapString	string	identification-strings >category-map-name	Alc-Credit-Control-CategoryMap	—
alc.dtc.ipAddress	string	address	Framed-IPAddress	—
alc.dtc.dhcp4DefaultGateway	string	options>default-router	Alc-Default-Router	—
alc.dtc.subnetMask	string	address	Framed-IPNetmask	—
alc.dtc.ipv4LeaseTime	integer	options>lease-time	Alc-Lease-Time	—
alc.dtc.ipv4PrimDns	string	options>dns-server	Alc-Primary-Dns Client-DNS-Pri	—
alc.dtc.ipv4SecDns	string	—	Alc-Secondary-Dns Client-DNS-Sec	—
alc.dtc.primNbns	string	options>netbios-name-server	Alc-Primary-Nbns RB-Client-NBNSPri	
alc.dtc.secNbns	string	—	Alc-Secondary-Nbns RB-Client-NBNSSec	—
alc.dtc.msapGroupInterface	string	msap-defaults>group-interface	Alc-MSAP-Interface	—

DTC variable	Type	LUDB	RADIUS attribute	Comment
alc.dtc.msapPolicy	string, integer	msap-defaults>policy	Alc-MSAP-Policy	—
alc.dtc.msapServiceId	string, integer	msap-defaults>service	Alc-MSAP-Serv-Id	—
alc.dtc.retailServiceId	string	Retail-service-id	Alc-Retail-Serv-Id	—
alc.dtc.ipv6Address	string	ipv6-address	Alc-Ipv6-Address	—
alc.dtc.ipv6DelegatedPrefix	string	ipv6-delegated-prefix	Delegated-IPv6-Prefix	—
alc.dtc.ipv6SlaacPrefix	string	ipv6-slaac-prefix	Framed-IPv6-Prefix	—
alc.dtc.ipv6WanPool	string	ipv6-wan-address-pool	Framed-IPv6-Pool	—
alc.dtc.ipv6PrefixPool	string	ipv6-delegated-prefix-pool	Alc-Delegated-IPv6-Pool	—
alc.dtc.ipv6DelegatedPrefix Length	integer	ipv6-delegated-prefix-len	Alc-Delegated-IPv6-Prefix-Length	—
alc.dtc.accountingPolicy	string	acct-policy	—	—
alc.dtc.dhcpv4GIAddr	string	gi-address	—	—
alc.dtc.dhcv4ServerAddress	string	server	—	—
alc.dtc.dhcp4SrcAddr	string	—	—	—
alc.dtc.dhcp4Pool	string	address>pool	Framed-Pool Ip-Address-Pool-Name	prim | sec ("|" – delimiter)
alc.dtc.linkAddress	string	link-address	—	—
alc.dtc.dhcp6SrcAddr	string	—	—	—
alc.dtc.dhcv6ServerAddr	string	server6	—	—
alc.dtc.setDhcpv6LinkAddr	string	link-address	—	This API applies only to regular numbered or unnumbered IPv6 interfaces (no ESM)
alc.dtc.setDhcpv6ServerAddr	string	server6	—	This API applies only to regular

DTC variable	Type	LUDB	RADIUS attribute	Comment
				numbered or unnumbered IPv6 interfaces (no ESM)
alc.dtc.ipv6PrimDns	string	options6>dns-server	Alc-Ipv6-Primary-Dns	—
alc.dtc.ipv6SecDns	string	—	Alc-Ipv6-Secondary-Dns	—
alc.dtc.dhcpv6PreferredLifetime	integer	ipv6-lease-times>preferred-lifetime	Alc-v6-Preferred-Lifetime	—
alc.dtc.dhcpv6RebindTimer	integer	ipv6-lease-times>rebind-timer	Alc-Dhcp6-Rebind-Time	—
alc.dtc.dhcpv6RenewTimer	integer	ipv6-lease-times>renew-timer	Alc-Dhcp6-Renew-Time	—
alc.dtc.dhcpv6ValidLifetime	integer	ipv6-lease-times>valid-lifetime	Alc-v6-Valid-Lifetime	—

For example, an IP address is assigned to a DTC variable as a string:

alc.dtc.ipAddress = 192.168.0.10

This is performed through the following ALU API: alc.dtc.setESM(alc.dtc.ipAddress, 192.168.0.10). The DTC logic then parses this variable and converts it into appropriate format for consumption by ESM code.

The values defined above are the ones that are mostly defined in the LUDB. Main use, however, is assigning ESM strings for the subscriber host instantiation phase during the processing of DHCP ACK/Reply messages. Consequently, the Python script needs to be run only on DHCP Request messages (no need to run it on Discoveries for ESM string assignment, unless the LUDB derived ID is also needed).

DHCP options that are blindly returned to the DHCP client without the ESM code being aware of them cannot be configured with DTC. These options should be configured with RADIUS (Alc-ToCLient-Dhcp-Options IPv4 only) or they can be inserted directly into DHCP messages with Python (bypassing DTC).

Other possible uses for DTC variables are:

- Assigning routing context information with Python (service-id, msap, msap-policy, retail service-id, and so on). For example, AN can insert specific hints in various DHCP options that would suggest (by Python) the service context to place the subscriber host.
- IP address assignment by Python (DTC). This would address the DTC-to-DHCP-Proxy case where Python script is invoked on DHCP Discovery/Solicit. For example:
  1. Discover arrives.
  2. Python generates an IP address, for example based on some DHCP options.
  3. The script stores the IP address by using alc.dtc.setEsm(alc.dtc.ipAddress, 10.0.1.2).
  4. After the script is finished, ESM starts processing the packet (no LUDB/RADIUS authentication configured).

5. ESM finds the IP address already in DTC and decides to handle all DHCP and execute proxy function instead of relay.
6. ESM sends an offer with the address that Python generated.
7. DHCP options should be provided as well in this case (lease-times, and so on).
8. The same applies to the DHCP Request.

3.2.16.1 DTC debugging facility

DTC debugging is part of the generic DHCP debugging facility that is enabled by the flowing commands:

debug router *router-id* **ip dhcp** enables DHCP debug on Layer 3 interfaces, including subscriber-interfaces.

debug service id *service-id* **dhcp** enables DHCP debugging on capture SAP

If the DTC cache is populated with Python, the corresponding DTC entries are shown as part of the matching DHCP message debug.

3.2.17 Virtual subnet for DHCPv4 hosts

The **virtual-subnet** command in the **sub-if>dhcp** context allows the system to snoop and record the default router address in DHCP ACK messages for a DHCPv4 ESM host. The system can answer ping or traceroute requests even if the default router address is not configured on the subscriber-interface.

This feature eliminates the need to configure every default-gw address on subscriber interface. Beside default router address, the system also calculates host's subnet by using an assigned address and the subnet mask option in ACK. Both recorded default router address and the subnet can be displayed with the **show service** *id* **virtual-subnet** command.

Every ESM subscriber only has one set of default router address and subnet.

3.2.18 Address reservation for sticky leases

Address reservation for sticky leases adds support in local DHCP servers to provide IP address reservation for the assignment of sticky leases. These leases are pre-provisioned in the server (by SNMP or CLI) with a specific set of user-identification parameters. This set of parameters must be unique for each pool to avoid duplicate leases. For management purposes, these leases also have an additional *hostname* parameter to easily retrieve them by SNMP. When a sticky lease is created, the corresponding requested IP address is allocated from the specified pool and this address afterwards can only be used by DHCP if they match the specified user identification parameters. It is not possible to make an existing DHCP lease sticky.

Sticky leases are persistent but not synchronized in multichassis synchronization. To support multichassis redundancy, a management system can allocate the lease on one 7750 SR, immediately retrieve the lease and populate it again on the redundant router. For this to work, sticky leases should not be combined with any other allocation method (for example, regular DHCP leases, or Local Address Assignment because these methods can already allocate the address on the standby node).

This feature targets two scenarios:

- **simplified IP address reservation**

It is not necessary to provision LUDB entries and exclude ranges for sticky leases. After the lease is reserved in the local DHCP server (triggered by an external system by SNMP or CLI), the external system can then subsequently assign the corresponding IP address to a DHCP client that matches the configured lease. A use case of this related to virtual residential gateway (vRGW) application and is described in the vRGW section.

- **pool management without DHCP**

  In pure management cases, this provides an easy method to perform pool management without implementing DHCP-specific configurations. For example, a management platform can allocate an IP address from a pool using the sticky lease mechanism and then assign this to a static host without risking overlap.

3.2.19 DHCP message processing overload protection

A DHCP message processing overload condition occurs when the arrival rate of DHCP packets is higher than what the applications can process. For example, when inadequate lease times are used in a scaled BNG setup. The SR OS measures and reports DHCP message processing overload and acts upon it by selectively dropping DHCP messages for new connections before DHCP messages for ongoing sessions. When in overload, DHCP messages are dropped in following order (similar for DHCPv6):

- Discover
- Offer
- Other DHCP messages
- Renew
- Ack

When DHCP message drops occurred within the last 5 minute interval, then a "*DHCP message processing overload detected: true*" trap is generated. When no DHCP message drops occurred in the last 5 minutes interval, a "*DHCP message processing overload detected: false*" trap is generated as shown in the following example:

```
A:pe1# show log event-control "svcmgr" 2572
===============================================================================
Log Events
===============================================================================
Application
ID#    Event Name                       P   g/s     Logged     Dropped
-------------------------------------------------------------------------------
   2572 tmnxSubDhcpOverloadDetected      WA  thr          0           0
===============================================================================
```

The DHCP message processing overload state can also be checked with the **show**>**subscriber-mgmt**>**status system** command. The following output displays an example.

```
A:pe1# show subscriber-mgmt status system
===============================================================================
Subscriber Management System Status
===============================================================================
Chassis 1
-------------------------------------------------------------------------------
Memory usage high                                    : No
DHCP message processing overload                     : No
Statistics usage high                                : No
```

```
Number of subscribers using statistics                  : 0

Data-trigger statistics
-------------------------------------------------------------------------------
Packets received                                        : 0
Packets dropped                                         : 0
Packets in queue (actual)                               : 0
Packets in queue (peak)                                 : 0

Bridged Residential Gateway statistics
-------------------------------------------------------------------------------
BRG initialized                                         : 0
BRG operational                                         : 0
BRG in connectivity verification                        : 0
BRG on hold                                             : 0
BRG authenticated by proxy                              : 0

Subscriber VLAN statistics resources
-------------------------------------------------------------------------------
Administrative state                                    : out-of-service
Number of entries                                       : 0
===============================================================================
```

Statistics for dropped DHCP packets can be displayed and cleared with the **tools**>**dump**>**dhcp-rx-stats** command. The following output displays an example.

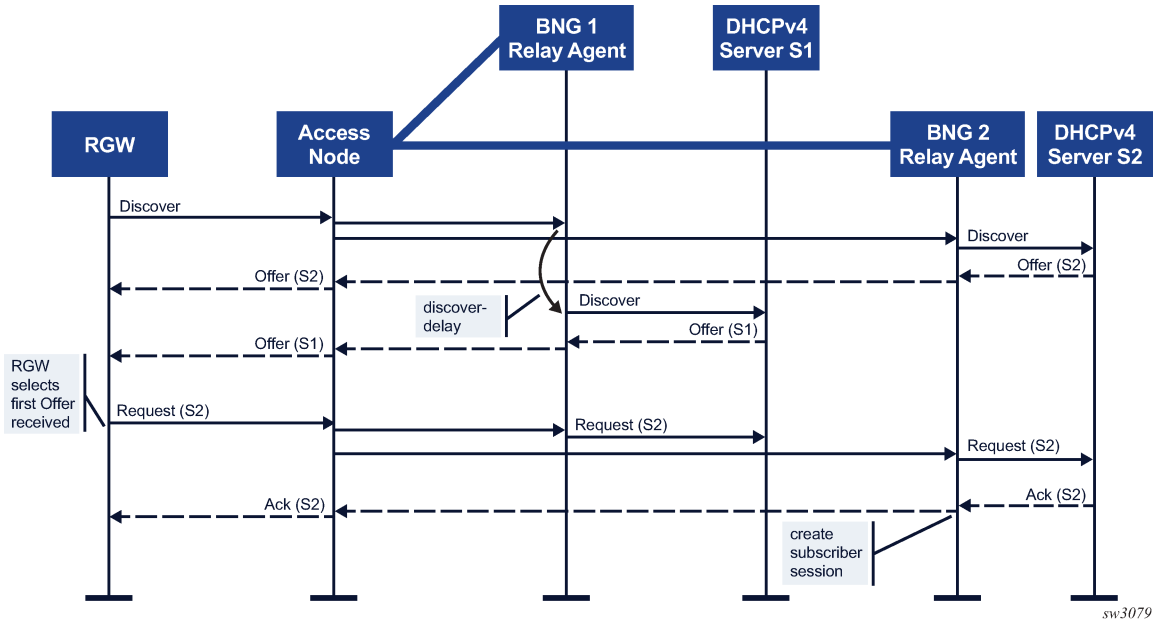
```
A:pe1# tools dump dhcp-rx-stats
===============================================================================
DHCP Received Packet Statistics
===============================================================================
     Type                  Received     Forwarded       Dropped  Dropped(ESM)
-------------------------------------------------------------------------------
IPv4 DISCOVER                      0             0             0             0
     OFFER                         0             0             0             0
     REQUEST                       0             0             0             0
     DECLINE                       0             0             0             0
     ACK                           0             0             0             0
     NAK                           0             0             0             0
     RELEASE                       0             0             0             0
     INFORM                        0             0             0             0
     FORCERENEW                    0             0             0             0
     LEASEQUERY                    0             0             0             0
     LEASEUNASSIGNED               0             0             0             0
     LEASEUNKNOWN                  0             0             0             0
     LEASEACTIVE                   0             0             0             0
     RENEW                         0             0             0             0
-------------------------------------------------------------------------------
IPv6 SOLICIT                       0             0             0             0
     ADVERTISE                     0             0             0             0
     REQUEST                       0             0             0             0
     CONFIRM                       0             0             0             0
     RENEW                         0             0             0             0
     REBIND                        0             0             0             0
     REPLY                         0             0             0             0
     RELEASE                       0             0             0             0
     DECLINE                       0             0             0             0
     RECONFIGURE                   0             0             0             0
     INFO_REQUEST                  0             0             0             0
     RELAY_FORW                    0             0             0             0
     RELAY_REPLY                   0             0             0             0
     LEASEQUERY                    0             0             0             0
     LEASEQUERY_REPLY              0             0             0             0
-------------------------------------------------------------------------------
```

```
     Total                          0            0            0            0
-------------------------------------------------------------------------------
Maximum queue length          : 0
Maximum outst pbufs total     : 0
Maximum outst pbufs to client : 0
===============================================================================
```

3.2.20 DHCPv4 offer and DHCPv6 advertise selection parameters for DHCP relay

When multiple DHCP servers in the network offer an IP address or prefix, the DCHP client must choose a server from which to request configuration parameters. In some network designs, such as a stateless multichassis redundant BNG deployment, the load balancing of IPoE subscriber sessions can be optimized by influencing the server selection at the DHCP relay. This is similar to using a **pado-delay** for PPPoE subscriber sessions. See Figure 23: DHCPv4 relay offer selection parameters.

Figure 23: DHCPv4 relay offer selection parameters



In Figure 23: DHCPv4 relay offer selection parameters, the DHCPv4 client on the RGW connects and broadcasts a DHCPv4 Discover message to obtain an IP configuration. Both the BNG 1 and BNG 2 relay agents receive the Discover message with the Layer 2 aggregation network between the Access Node and the BNGs. The DHCP Relay Agent function on BNG 1 is configured to delay the Discover message before sending it to the DHCPv4 Server S1, while the DHCP Relay Agent function on BNG 2 immediately forwards the Discover message to DHCPv4 server S2. As a result, the DHCP Offer from DHCPv4 Server S2 reaches the DHCP client first. The client selects the Offer from DHCP Server S2 and broadcasts a DHCP Request with the server identifier option set to Server S2. The BNG1 and BNG2 relay agents both forward the Request to their DHCP servers. DHCPv4 Server S1 ignores the Request targeted to server S2. DHCPv4 Server S2 Acknowledges the lease. The subscriber session is created on BNG 2 and the DHCPv4 Ack is sent to the RGW.

Similar results can be achieved for DHCPv6 clients by delaying the DHCPv6 Solicit message or by inserting a Preference Option in the DHCPv6 Advertise message to the client.

Note: The DHCPv4 Offer Selection and DHCPv6 Advertise Selection is a function of the DHCP client. This feature provides a non-standardized mechanism to influence the client's decision at the DHCP Relay, but ultimately the client decides which server is selected.

3.2.20.1 DHCPv4 offer selection parameters on a subscriber group-interface DHCP relay

With the configuration of a **discover-delay**, the forwarding of a DHCP Discover Message to the DHCP server is delayed, which results in a delayed DHCPv4 Offer to the DHCP client. A **discover-delay** (in deciseconds) can be configured for DHCP Discover messages, as shown in the following examples:

- originated by DHCP clients with odd or even MAC addresses

```
dhcp
    server 192.0.2.1
    offer-selection
        client-mac odd
            discover-delay 5
        exit
    exit
    no shutdown
exit
```

- sent to a specific DHCP server (a delay for up to eight servers can be configured)

```
dhcp
    server 192.0.2.1 192.0.2.2
    offer-selection
        server 192.0.2.2
            discover-delay 5
        exit
    exit
    no shutdown
exit
```

- for which no per client MAC or per DHCP server **discover-delay** is configured (for example, a default **discover-delay**)

```
dhcp
    server 192.0.2.1
    offer-selection
        discover-delay 5
    exit
    no shutdown
exit
```

Additional considerations:

- Configuring a per DHCP server **discover-delay** and a per DHCP client MAC address **discover-delay** is mutually exclusive.
- A default **discover-delay** can be combined with either a per DHCP server or a per DHCP client MAC delay.
- When a new Discover message, such as a retransmitted message is received from the same client while a **discover-delay** timer is running, the **discover-delay** timer is stopped, the queued Discover message is discarded, and the new Discover message is immediately forwarded without delay.

3.2.20.2 DHCPv6 advertise selection parameters on a subscriber group-interface DHCP6 relay

With the configuration of a **solicit-delay**, the forwarding of a DHCP Solicit Message to the DHCP server is delayed resulting in a delayed DHCPv6 Advertise to the DHCP client. A **solicit-delay** (in deciseconds) can be configured for DHCP Solicit messages, as shown in the following examples:

- originated by DHCP clients with odd or even MAC addresses

```
dhcp6
    relay
        advertise-selection
            client-mac odd
                solicit-delay 5
            exit
        exit
        server 2001:db8::1
        no shutdown
    exit
exit
```

- sent to a specific DHCP server (a delay for up to eight servers can be configured)

```
dhcp6
    relay
        advertise-selection
            server 2001:db8::2
                solicit-delay 5
            exit
        exit
        server 2001:db8::1
        server 2001:db8::2
        no shutdown
    exit
exit
```

- for which no per client MAC or per DHCP server **solicit-delay** is configured (for example, a default **solicit-delay**)

```
dhcp6
    relay
        advertise-selection
            solicit-delay 5
        exit
        server 2001:db8::1
        no shutdown
    exit
exit
```

With the configuration of a **preference-option value**, a DHCPv6 Preference Option (7) with the configured value is inserted in the Advertise Message to the DHCP client. A DHCPv6 client can use the preference value to select one out of multiple received Advertise messages. A **preference-option value** (0 to 255) can be configured for DHCP Advertise Messages, as shown in the following examples:

- sent to DHCP clients with odd or even MAC addresses

```
dhcp6
    relay
        advertise-selection
            client-mac odd
```

```
                preference-option
                    value 200
                exit
            exit
        exit
        server 2001:db8::1
        no shutdown
    exit
exit
```

- sent to a specific DHCP server (a **preference-option value** for up to eight servers can be configured)

```
dhcp6
    relay
        advertise-selection
            server 2001:db8::2
                preference-option
                    value 200
                exit
            exit
        exit
        server 2001:db8::1
        server 2001:db8::2
        no shutdown
    exit
exit
```

- for which no per client MAC or per DHCP server **preference-option value** is configured (for example, a default **preference-option value**)

```
dhcp6
    relay
        advertise-selection
            preference-option
                value 200
            exit
        exit
        server 2001:db8::1
        no shutdown
    exit
```

Additional considerations:

- A **solicit-delay** and **preference-option value** can be configured simultaneously.
- Configuring a per DHCP server **solicit-delay** and **preference-option value** and a per DHCP client MAC address **solicit-delay** or **preference-option value** is mutually exclusive.
- A default **solicit-delay** and **preference-option value** can be combined with either a per DHCP server or a per DHCP client MAC configuration.
- When a new solicit message is received from the same client, such as a retransmitted message while a **solicit-delay** timer is running, the **solicit-delay** timer is stopped, the queued solicit message is discarded, and the new Solicit message is immediately forwarded without delay.

3.2.21 Lightweight DHCPv6 Relay Agent

This section describes Lightweight DHCPv6 Relay Agent (LDRA) functionality.

3.2.21.1 LDRA in a VPLS service

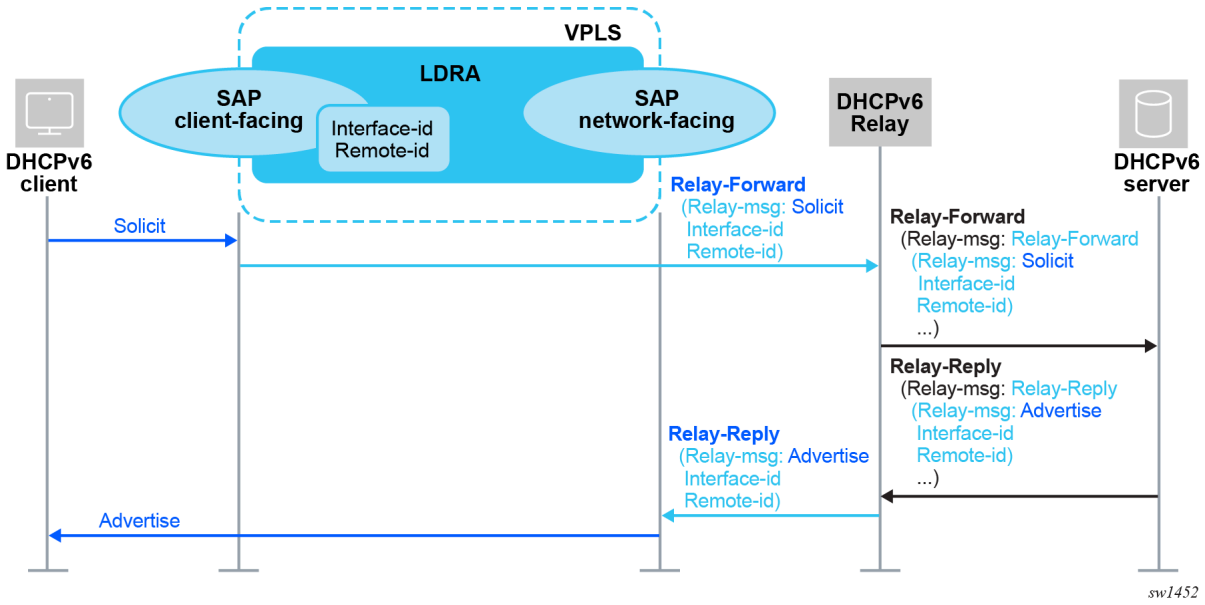
Note: LDRA is only supported on VPLS SAPs.

LDRA, as described in RFC 6221, *Lightweight DHCPv6 Relay Agent*, resides on the same IPv6 link as the DHCPv6 client and the DHCPv6 relay agent or server and inserts relay agent information options such as the interface ID and remote ID that identify the client-facing interface. A DHCPv6 server can use these options to:

- identify the client
- know where to client is attached to the network
- assign appropriate parameters

The following figure displays an initial DHCPv6 Solicit/Advertise message flow with LDRA enabled in a VPLS service.

Figure 24: DHCPv6 Solicit/Advertise message flow with LDRA in a VPLS service



An LDRA distinguishes between client-facing interfaces and network-facing interfaces to process DHCPv6 messages.

Use the following command to configure a VPLS SAP as a client-facing interface.

```
configure service vpls sap dhcp6 ldra interface-type client-facing
```

A client-facing SAP only accepts DHCPv6 client messages. The client messages are encapsulated in a Relay-Forward message, including the mandatory **interface-id** option 18 and an optional relay agent **remote-id** option 37 that can be configured using the following commands.

```
configure service vpls sap dhcp6 ldra options interface-id
configure service vpls sap dhcp6 ldra options remote-id
```

The Resulting Relay-Forward message is flooded in the VPLS. For more information about these options, see the following user guides:

- *7450 ESS, 7750 SR, 7950 XRS, and VSR MD-CLI Command Reference Guide*
- *7450 ESS, 7750 SR, 7950 XRS, and VSR Classic CLI Command Reference Guide*

A client-facing SAP is always untrusted: the DHCPv6 client must be directly connected on the same link layer without a relay agent or another LDRA in between. DHCPv6 Relay-Forward, Relay-Reply, Advertise, Reply, and Reconfigure messages received on a client-facing SAP are silently dropped.

Use following command to configure a VPLS SAP as network-facing interface.

```
configure service vpls sap dhcp6 ldra interface-type network-facing
```

A network-facing SAP only accepts DHCPv6 Relay-Reply messages. All other DHCPv6 messages are silently dropped on a network-facing SAP. The DHCPv6 Relay-Reply message is decapsulated and the extracted DHCPv6 server message forwarded in the VPLS.

LDRA DHCPv6 messages are forwarded in the VPLS following the VPLS forwarding rules:

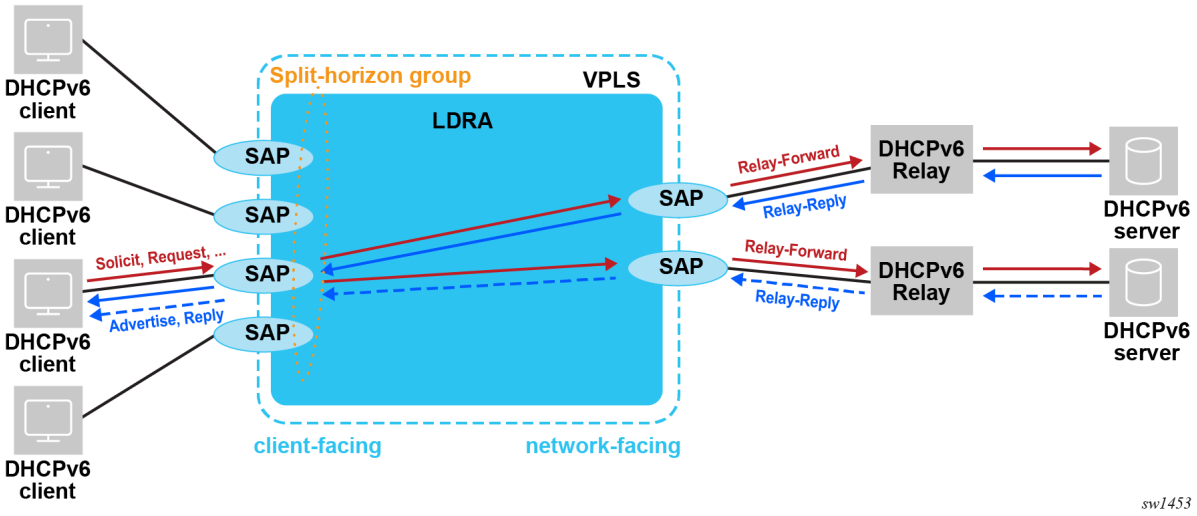
- DHCPv6 client messages received on client-facing SAPs and the resulting Relay-Forward message have the multicast destination IP address All_DHCP_Relay_Agents_and_Servers (ff02::1:2) and a multicast destination MAC address (33:33:00:01:00:02). The Relay-Forward message is flooded in the VPLS regardless of the LDRA configuration to all SDPs, all SAPs with LDRA disabled, and all SAPs with LDRA enabled as client-facing or network-facing interfaces.
- DHCPv6 Relay-Reply messages received on network-facing SAPs and the extracted server messages have the link layer address of the DHCPv6 client as a unicast destination MAC address. The extracted server message is forwarded to the client-facing SAP based on a VPLS FDB lookup of the destination MAC address. This requires that MAC learning is enabled in the VPLS, otherwise the server messages are flooded as unknown unicast to all SDPs, all SAPs with LDRA disabled, and all SAPs with LDRA enabled as client-facing or network-facing interface.

The VPLS forwarding of LDRA DHCPv6 messages as described above results in the following requirements for a correct operation when enabling LDRA on a VPLS SAP:

- Only include SAPs in the VPLS (that is, no spoke or mesh SDP)
- Enable MAC learning in the VPLS service
- Enable LDRA on all SAPs in the VPLS
- Configure at least one SAP in the VPLS as a client-facing interface and at least one SAP in the VPLS as a network-facing interface
- Include all client-facing SAPs in a split-horizon group

The following figure displays the DHCPv6 message flow for a VPLS with LDRA enabled.

Figure 25: DHCPv6 message forwarding for a VPLS with LDRA enabled



Use DHCPv6 debug in the VPLS service to troubleshoot LDRA related problems.

Use the **show service dhcp6 statistics sap** command to display a per SAP statistics of DHCPv6 snooped packets.

3.2.22 DHCP release messages

A DHCP client can send a release message to the server to indicate that the address assigned in the lease is not used. The BNG configured as DHCP relay or relay proxy can also send a release message to the server on behalf of the client in the following cases:

- the drop of a DHCPv4 ACK or DHCPv6 reply message caused by a host creation failure, for example, when there are not enough resources or there is a duplicate host
- sla-profile or sub-profile host and session limit enforcement
- lease time expiration when lease split is active
- Subscriber Host Connectivity Verification (SHCV) connectivity loss with action lease removal
- an IPoE session timeout
- manual clearing of an IPoE session or lease state; a DHCP release message is not sent when the optional **no-dhcp-release** parameter is specified in the **clear service id dhcp lease-state** or **clear service id dhcp6 lease-state** commands.

The BNG does not send a DHCP release message on behalf of the client when the remaining lease time is less than 5 minutes.

The **release-include-gi-address** command configured in the **dhcp** context of an interface sets the gateway IP Address (giaddr) field in the DHCPv4 release messages to the configured GI address. By default, the giaddr field in a DHCPv4 release message is transparently forwarded as received from the client or set to 0.0.0.0 when the DHCPv4 release message is sent by the BNG on behalf of the client.

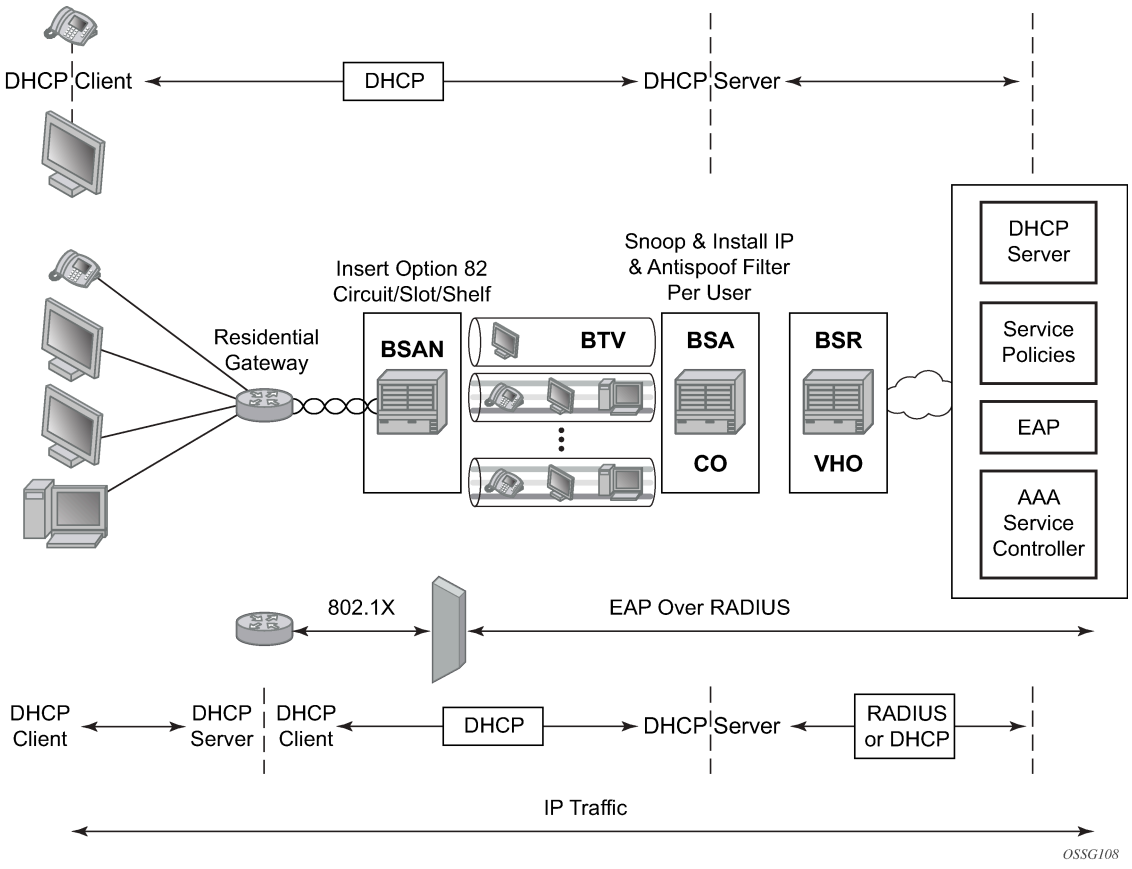
3.3 Proxy DHCP server

This section describes the implementation of proxy DHCP server capability to provide a standards-based DHCP server which front-ends to a downstream DHCP client, DHCP relay enabled devices, and interfaces with RADIUS to authenticate the IP host and subscriber and obtains the IP configuration information for DHCP client devices.

The proxy DHCP server is located between an upstream DHCP server and downstream DHCP clients and relay agents when RADIUS is not used to provide client IP information.

Service providers can introduce DHCP into their networks without the need to change back-end subscriber management systems that are typically based around RADIUS (AAA). Service providers can support the use of DHCP servers and RADIUS AAA servers concurrently to provide IP information for subscriber IP devices (Figure 26: Typical DHCP deployment scenarios).

Figure 26: Typical DHCP deployment scenarios



DHCP is the predominant client-to-server based protocol used to request IP addressing and necessary information to allow an IP host device to connect to the network.

By implementing DHCP, the complexity of manually configuring every IP device that requires connectivity to the network is avoided. IP devices with DHCP can dynamically request the appropriate IP information to enable network access.

DHCP defines three components that are implemented in a variety of device types:

- The DHCP client allows an IP device (host) to request IP addressing information from a DHCP server to enable access to IP based networks. This is typically found in:
  - end user notebooks, desktops, and servers
  - residential gateways and CPE routers
  - IP phones
  - set-top boxes
  - wireless access points
- The DHCP relay agent passes (relays) DHCP client messages to pre-configured DHCP servers where a DHCP server is not on the same subnet as the IP host. This feature optionally adds information into DHCP messages (Option 82) which is typically used for identifying attaching IP devices and their location as part of subscriber management. This is typically found in:
  - residential gateways and CPE routers
  - DSLAMs
  - edge aggregation routers
- The DHCP server receives DHCP client messages and is responsible for inspecting the information within the messages and determining what IP information if any is to be provided to a DHCP client to allow network access. This is typically found in:
  - dedicated standalone servers
  - residential gateways and CPE routers
  - edge aggregation routers
  - centralized management systems

DHCP is the predominant address management protocol in the enterprise community, however in the provider market PPP has traditionally been how individual subscribers are identified, authenticated, and provided IP addressing information.

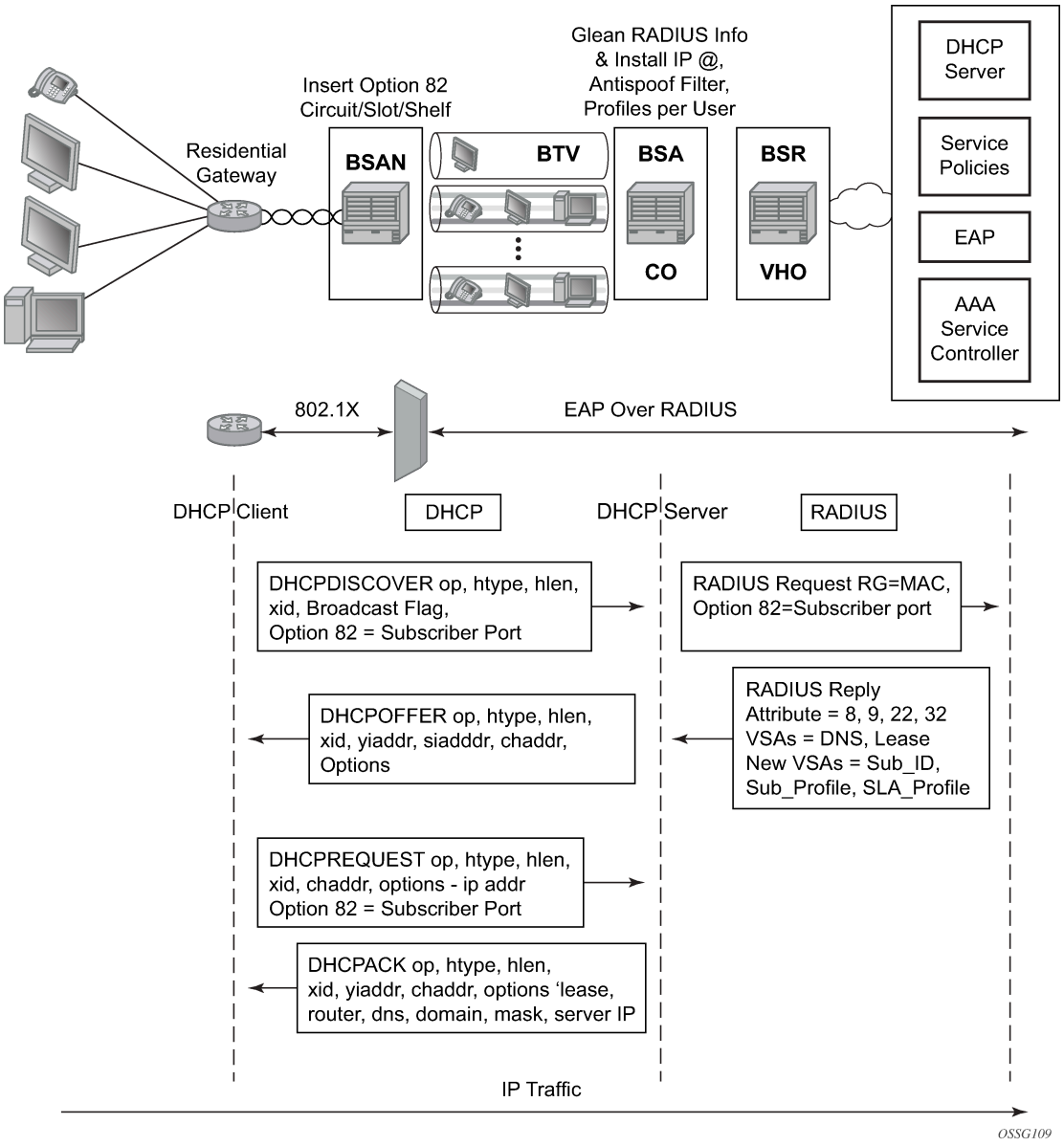
The use of DHCP in the provider market is a growing trend for managing subscriber IP addressing, as well as supporting newer devices such as IP-enabled IP phones and set-top boxes. Most subscriber management systems rely heavily on RADIUS (RFC 2865, *Remote Authentication Dial In User Service (RADIUS)*) as the means for identifying and authorizing individual subscribers (and devices), deciding whether they are allowed access to the network, and which policies should be put in place to control what the subscriber can do within network.

The proxy DHCP server capability enables the deployment of DHCP into a provider network, by acting as a proxy between the downstream DHCP devices and the upstream RADIUS based subscriber management system.

- Interact with downstream DHCP client devices and DHCP relay agents in the path.
- Interface with RADIUS to authenticate DHCP requests.
- Receive all the necessary IP information to properly respond to a DHCP client.
- Override the allocated IP address lease time, if necessary, for improved IP address management.

Figure 27: Aggregation network with DHCP to RADIUS authentication shows a typical DHCP initial bootup sequence with the addition of RADIUS authentication. The proxy DHCP server interfaces with downstream DHCP client devices and then authenticate upstream using RADIUS to a provider's subscriber management system.

Figure 27: Aggregation network with DHCP to RADIUS authentication



In addition to granting the authentication of DHCP hosts, the RADIUS server can include RADIUS attributes (standard and vendor-specific attributes (VSAs)) which are then used by the edge router to:

- Provision objects related to a specific DHCP host such as a subscriber and SLA policy.
- Provide IP addressing information to a DHCP client.
- Support the features that leverage DHCP lease state.
  - dynamic ARP population
  - ARP reply agent
  - anti-spoofing filters

  - MAC pinning
- Leverage host-connectivity-verify to determine the state of a downstream IP host.

This feature offers the ability for a customer to integrate DHCP to the subscriber while maintaining their existing subscriber management system typically based around RADIUS. This provides the opportunity to control shifts to an all DHCP environment or to deploy a mixed DHCP and RADIUS subscriber management system.

To maximize its applicability VSAs of legacy BRAS vendors can be accepted so that a network provider is not forced to reconfigure its RADIUS databases (or at least with minimal changes).

To receive data from the RADIUS server the following are supported:

- Juniper (vendor-id 4874) attributes 4 (Primary DNS server) and 5 (Secondary DNS server).
- Redback (vendor-id 2352) attributes 1 (Primary DNS) and 2 (Secondary DNS).
- Juniper attributes 6 and 7 (Primary and Secondary NetBIOS nameserver).
- Redback attributes 99 and 100 (Primary and Secondary NetBIOS nameserver).

The following attributes can be sent to RADIUS:

- Sending authentication requests: (from the DSL Forum) (vendor-id 3561), attributes 1 (Circuit ID) and 2 (Remote ID).
- DSL Forum attributes 129 and 130 (Actual Data Rate Upstream and Downstream), 131 and 132 (Minimum Data Rate Upstream and Downstream) and 144 (Access Loop Encapsulation).

The complete list of Nokia VSAs is available on a file included on the compact flash shipped with the image.

3.3.1 Local DHCP servers

3.3.1.1 Terminology

- **local 7750 SR and 7450 ESS DHCP server**

  The DHCP server instantiated on the local 7750 SR and 7450 ESS node.
- **remote 7750 SR and 7450 ESS DHCP server**

  The DHCP server instantiated on the remote 7750 SR and 7450 ESS node (external to the local 7750 SR and 7450 ESS node).
- **3rd party DHCP server**

  The DHCP server external to any 7750 SR and 7450 ESS node and implemented outside of 7750 SR and 7450 ESS.
- **intercommunication link**

  The logical link between dual-homed 7750 SR and 7450 ESS DHCP servers used for synchronizing DHCP lease states. Multi-chassis Synchronization (MCS) protocol runs over this link. When this link is interrupted, synchronization of the leases between redundant DHCP servers is impaired. This link should be well protected with multiple underlying physical paths.
- **local IP address-range and prefix**

The local failover mode in which the IP address-range and prefix is configured in dual-homed DHCP environment. The local keyword does not refer to the locality (local versus remote) of the server on which the IP address-range and prefix in configured, but rather refers to the ownership of the IP address-range and prefix. The DHCP server on which the local IP address-range and prefix is configured, owns this IP address-range and prefix and consequently can delegate the IP addresses and prefixes from it at any time, regardless of the state of the intercommunication link.

- **remote IP address-range and prefix**

  The remote failover mode in which the IP address-range and prefix is configured in dual-homed DHCP environment. The remote keyword does not refer to the locality of the server on which the IP address-range and prefix is configured but rather refers to the ownership of the IP address-range and prefix. The DHCP server on which the remote IP address-range and prefix is configured, but does not own this IP address-range and prefix during normal operation and consequently is not allowed to delegate the IP addresses and prefixes from it. Only when the intercommunication link between the two nodes transition into particular (failed) state, the DHCP server can start delegating new IP addresses from the remote IP address-range and prefix.

- **IP address-range and prefix ownership**

  The 7750 SR and 7450 ESS DHCP servers that can delegate new leases from an IP address-range and prefix that it owns. For example, an IP address-range and prefix designated as remote is not owned by the DHCP server on which it is configured unless specific conditions are met. Those conditions are governed by the state of the intercommunication link.

- **IP address-range and prefix takeover**

  The 7750 SR and 7450 ESS DHCP servers that do not own an IP address-range and prefix can take over the ownership of this IP address-range and prefix under specific conditions. When the ownership is taken, the new IP addresses can start being delegated from this IP address-range and prefix. Only the remote IP address-range and prefix can be taken over. Note that the takeover of an IP address-range and prefix has only local significance, in other words, the ownership is not taken away from some other DHCP server that has the same IP address-range and prefix designated as local. It only means that IP address-range and prefix that is configured as remote is available to takeover for new IP address delegation.

- **local PPPoX address pools**

  The method of accessing an IPv4/v6 address pool in 7750 SR and 7450 ESS DHCP4/6 server. For PPPoX clients, the IPv4/v6 addresses are allocated from those pools without the need for an intermediate DHCP relay-agent (7750 SR and 7450 ESS internal DHCP relay-agent). Although those pools are part of the local DHCP server in 7750 SR and 7450 ESS, the method of accessing them is substantially different than accessing local DHCP address pools for IPoE (DHCP) clients. IPoE (DHCP) and PPPoX hosts can share the same pool and yet each client type can access them in their own unique way:

  – IPoE client by DHCP messaging

  – PPPoE by internal API calls

- **local PPPoX pool management**

  The IPv4 address allocation/management for PPPoX clients independent of DHCP process (DHCP lease state). An IPv4 address allocated by local PPPoX Pool Management is tied to the PPPoX session. It is without the need for an internal DHCP relay-agent.

3.3.1.2 Overview

7750 SR and 7450 ESS DHCP server multihoming ensures continuity of the IP address and prefix assignment and renewal processes when an entire 7750 SR and 7450 ESS DHCP server fails or in case of a failure of the active link that connects clients to one of the 7750 SR and 7450 ESS DHCP servers in the access part of the network. DHCP server multihoming is an integral part of the overall subscriber management multichassis protection scheme.

DHCP server multihoming can be implemented outside of the BNG, without subscriber management enabled. However, in the following text, it is assumed that the subscriber management multihoming (SRRP, MC-LAG, subscriber synchronization) is deployed along with DHCP server multihoming.

Although the subscriber synchronization process and the DHCP lease states synchronization process use the same synchronization infrastructure within 7750 SR and 7450 ESS (Multi Chassis Redundancy protocol), they are two separate processes that are not aware of each other. As such, the mechanisms that drive their switchover are different. For example, the mechanism that drives subscriber switchover from one node to the other is driven by the access protection mechanism (SRRP/MC-LAG) while the switchover (or takeover) of the IP address-range and prefixes in a DHCP pool is driven by the state of the intercommunication link over which the leases are synchronized. The failure of an entire node makes those differences irrelevant because the access-link failure coincides with the intercommunication link failure and the other way around. However, link-only failures become critical when it comes to their interpretation by the protection mechanisms (SRRP, MC-LAG, DHCP server multihoming). Regardless of nature of the failure, an overall DHCP server multichassis protection scheme must be devised in so that the two 7750 SR and 7450 ESS DHCP servers never allocate the same IP address and prefix to two different clients. Otherwise, IP address or prefix duplication ensues. Unique IP address and prefix allocation is achieved by making only one 7750 SR and 7450 ESS DHCP server responsible for IP address prefix delegation out of the shared IP address-range or prefix.

There are two basic models for DHCP server dual-homing:

- Shared IP address-ranges and prefixes are designated as local on one 7750 SR and 7450 ESS DHCP server and as remote on the other.

  In this case, the DHCP relays must point to both DHCP servers; the one configured with the local IP address-range and prefix as well as the one with the remote IP address-range and prefix.

  Under normal circumstances, the new IP addresses and prefixes can be only allocated from the DHCP server configured with the local IP address-range and prefix.

  The DHCP server configured with the remote IP address-range and prefix starts delegating new lease from it only when it declares that the redundant peer with the local IP address-range and prefix becomes unavailable.

  Detection of the peer unavailability is triggered by the failure of the intercommunication link which can be caused either by the nodal failure or simply by the loss of connectivity between the two nodes protecting each other. Thus, the loss of intercommunication link does not necessarily mean that the peering node is truly gone. It can simply mean that the two nodes became isolated and unable to synchronize their DHCP leases between each other. In such environment, both nodes can potentially allocate the same IP address at the same time. To prevent this, additional intercommunication link states and associated timers are introduced to give the operator ample time to fix the problem.

  For example, the DHCP server takes over the remote IP address-range and prefix after the MCLT period expires while the intercommunication link is in PARTNER-DOWN state. The PARTNER-DOWN state is entered after a preconfigured timer (partner-down-delay) expires. The consequence of these two additional timers (partner-down-delay and MCLT) is that the new IP address delegation from the remote (shared) IP address-range and prefix is not possible until the preset timers expire. This is needed and justified if the intercommunication link is interrupted, the nodes become isolated, and

consequently, the DHCP lease state synchronization becomes impaired. On the other hand, if the DHCP server with local IP address-range and prefix becomes truly unavailable, those additional restoration times causes interruption in service because the new IP addresses from the remote IP address-range and prefix is not immediately available for delegation.

Only new IP address delegation from the remote IP address-range and prefix is affected by this behavior. The existing IP leases can be extended on both nodes at any time irrespective of whether the configured address-range and prefix is designated as local or remote.

To ensure uninterrupted service even for new lease delegation in this model (local-remote), two approaches can be adopted:

– Segment the IP address and prefix space so that each node has an IP address-range and prefix designated as local. For example, instead of designating IP address-range 10.10.10.0/24 as local on DHCP server A and as remote on DHCP server B, the 10.10.10./24 IP address is split into two: 10.10.10.0/25 and 10.10.10.128/25. The 10.10.10.0/25 would be designated as local on the DHCP server A and as remote on DHCP server B. The 10.10.10.128/25 would be designated as remote on the DHCP server A and as local on DHCP server B. In this fashion, one node is always available to assign new leases without any overlap.

– If only one shared IP address-rage/prefix is deployed, the operator can bypass the timers (partner-down-delay and MCLT) that are put in place in case that DHCP server nodes become isolated. This bypass of the timers can be achieved with configuration. In this case, a safe operation is warranted only if the operator is confident that the intercommunication link failure is caused by the nodal failure, and not the physical link failure between the two nodes.

- Shared IP address-range and prefix is designated as access-driven on both 7750 SR and 7450 ESS DHCP servers.

  In this scenario, the shared IP address-range and prefix is owned by both nodes and the ownership is not driven by the state of the intercommunication link.

  To avoid IP address duplication, only one DHCP server at any time must be responsible for IP address assignment from this shared IP address-range and prefix.

  This is ensured by the access protection mechanism (SRRP/MC-LAG) that provides a single active path from the clients to the one of the DHCP servers.

  In case that clients have access to the same IP address-range and prefix on both DHCP servers at the same time, the IP address duplication may occur.

  Consider the following case:

  Two DHCP clients send DHCP Discovers in the following fashion:

  – DHCP client A sends DHCP Discover to the DHCP server A

  – DHCP client B sends DHCP Discover to the DHCP server B

  – DHCP server A assigns IP address 10.10.10.10 to the DHCP client A

  – DHCP server B assigns IP address 10.10.10.10 to the DHCP client B

  – This is a legitimate scenario because the DHCP lease states are not synchronized until the DHCP lease assignment is completed.

  – Just before the DHCP ACK is sent to the respective clients from both nodes, the DHCP lease sync messages are exchanged between the peers.

  – DHCP servers do not wait for the reply to the sync message before they send the DHCP Ack to the client.

– After the DHCP lease syn message is received from the peer, the DHCP server realizes that the IP lease already exist. In this case, the newer IP lease overrides the older.
– The result is that clients A and B use the same IP address and consequently the forwarding of the traffic is impaired.

In access-driven model, the ESM subscriber host must be colocated with the DHCP server. In other words, the DHCP server must be instantiated in redundant BNGs. The dhcp-relays must point to the respective local DHCP servers. There must be no cross-referencing of DHCP servers in this model. In addition, the IP address that the DHCP servers are associated with, must be the same on both DHCP servers. This is necessary to ensure uninterrupted service levels when the switchover in the access occurs.

For example:

– DHCP server A on BNG A is associated with the IP address 1.1.1.1 (for example loopback interface A on BNG A)
– DHCP server B on the peering BNG B is associated with the IP address 1.1.1.1 (configured in BNG B under loopback interface B)
– DHCP relay on BNG A points to the IP address 1.1.1.1 (DHCP server in BNG A)
– DHCP relay on BNG B points to the IP address 1.1.1.1 (DHCP server in BNG B)

Consider the following when contemplating deployment of the two described models:

- Local-remote model is agnostic of the access protection mechanism. In fact, the access protection mechanism is not needed at all for safe operation.

  Fast takeover of the single shared (remote) IP address-range and prefix can be provided only in cases where the operator can guarantee that the intercommunication link failure is caused by the nodal failure (entire DHCP server node becomes unavailable). Fast takeover is provided by bypassing the partner-down-timer and MCLT.

  If multiple IP address-ranges and prefixes are deployed, bypass of the timers is not needed because the local IP address-ranges and prefixes are available on both nodes.
- Access-driven model allows a single IP address-range and prefix to be shared across the redundant DHCP server nodes. The access to the single DHCP server node from the client side is ensured by the protection mechanism deployed in the access part of the network (SRRP or MC-LAG).

3.3.1.3 DHCP lease synchronization

DHCP server leases are synchronized over Multi-Chassis Synchronization (MCS) protocol. A DHCP lease synchronization message is sent to the peering node just before the DHCP ACK/Reply is sent to the client.

For example, the message flow for DHCPv4 lease establishment is the following:

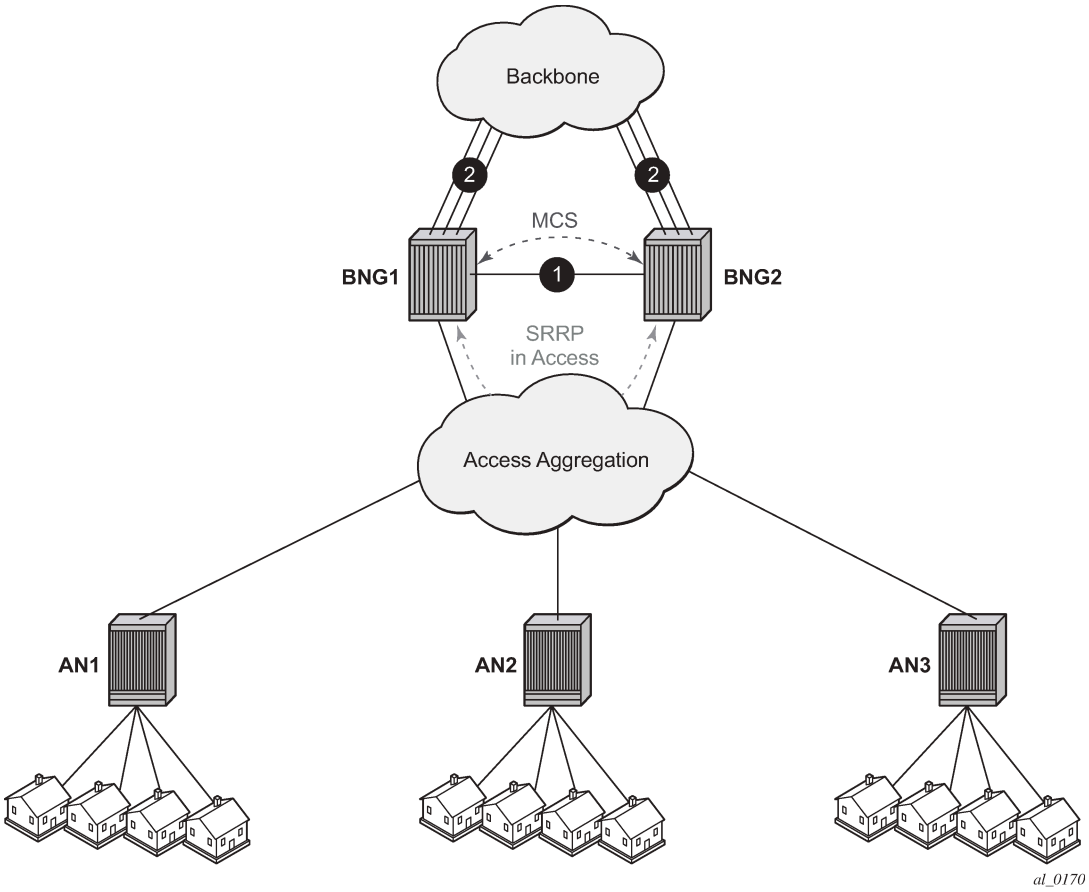
Table 3: DHCPv4 message flow

DHCP discover	DHCP client — DHCP server
DHCP Offer	DHCP server — DHCP client
DHCP Request	DHCP client — DHCP server
DHCP Sync Message	DHCP server — by MCS to the peering DHCP server

DHCP Ack	DHCP server— DHCP client

DHCP server failover mechanism in the local-remote IP address-range and prefix model relies on the detection of the failure of the link over which DHCP states are synchronized (through the MCS protocol). This link is normally disjointed from the access links toward the clients. MCS protocol normally runs over a direct link between the two redundant nodes (1) or over backbone links (2) if the direct link is not present. This is shown in Figure 28: Redundancy model.

Figure 28: Redundancy model



In the access-driven IP address-range and prefix model, the DHCP server address-ranges and prefixes are not tied to the state of the intercommunication link. Instead, the DHCP server selection for IP address assignment is only governed by the path selected by the path protection mechanism (SRRP/MC-LAG) deployed in the access part of the network.

3.3.1.4 Intercommunication link failure detection

7750 SR and 7450 ESS DHCP server is a client of Multi-chassis Synchronization (MCS) application with 7750 SR and 7450 ESS. After MCS transitions into an out-of-sync state, the 7750 DHCP server redundancy assumes that there is a failure in the network. The DHCP server failure in dual-chassis configuration relies on the failure detection mechanism of MCS.

MCS runs over TCP, port 45067 and it is using either data traffic or keepalives to detect failure on the communication link between the two nodes. In the absence of any MCS data traffic for more than 0.5 seconds, MCS sends its own keepalive to the peer. If a reply is not received within 3 seconds, MCS declares its operational state as down and the database sync state as out-of-sync. MCS consequently notifies its clients (DHCP Server being one of them) of this condition.

It can take up to 3 seconds before the DHCP client realizes that the interclass communication link failed.

MCS clients (applications) can optimally send their own proprietary keepalive messages to its partner over MCS to detect failure. DHCP Server does not use this method and it strictly relies on the failure notifications by MCS.

Note that the intercommunication link failure does not necessarily assume the same failed fate for the access links. In other words, it is perfectly possible (although unlikely) that both access links are operational while the inter-chassis communication link is broken.

The failure detection of the intercommunication link leads to specific failover state transitions on the DHCP server. The DHCP lease handling in the local-remote model depends on the failover state on the DHCP server and the duration of each failover state is determined by preconfigured timers.

3.3.1.5 DHCP server failover states

The DHCP server, when paired in redundant fashion, can transition through several states:

- TRANSITION
- SHUTDOWN
- INIT
- STARTUP

  In this state, the DHCP server recovers leases from the MCS database and does not respond to any unicast and broadcast messages.
- NORMAL

  In this state, the 7750 SR and 7450 ESS DHCP server is serving all IP leases from the local and access-driven IP addresses-ranges and prefixes (assigning new leases and extending existing ones). Remote IP addresses-ranges and prefixes are not served (new or existing ones).
- P RENORMAL

  In this state, the DHCP recovers leases from the MCS database after a resolved communication failure. The DHCP server responds to unicast and broadcast messages for addresses in the local ranges, and to unicast messages for addresses in the remote ranges.
- COMMUNICATION INTERUPTED

  IP addresses and prefixes under the local and access-driven IP address-range and prefix are served in the same fashion as in the NORMAL state. The IP addresses and prefixes under the remote IP address-range and prefix are renewed. However, the new address and prefix from the remote address-range and prefix are not allocated until the partner-down timer expires, the failover state consequently transitions into PARTNER DOWN and the MCLT timer while in the PARTNER-DOWN state expires. This is necessary in case that the failure occurred only on the intercommunication link while the access link is still operational (DHCP server nodes become isolated).

  COMMUNICATION INTERUPTED state indicates that there is a failure of some kind, but it cannot be determined whether an entire node failed or only the inter-chassis link. The access layer may still be

fully operational, but the new leases (including RENEWS/REBINDS) cannot be synchronized between the two peers.

- PARTNER DOWN

  After the DHCP Server reaches this state, the remote IP address-range and prefix is taken over and after an additional period of MCLT (Maximum Client Lead Time), the new IP addresses and prefixes from it can be delegated to clients. PARTNER-DOWN state is an indication (and assumption at the same time) that the remote node is truly down.

  Otherwise, the IP address duplication may occur when all the following conditions are met:

  – The DHCP server nodes become isolated.

  – The failover state is PARTNER-DOWN.

  – DHCP/PPPoX clients have simultaneous access to the same IP address-range and prefix on both DHCP servers.

3.3.1.6 Lease time synchronization

7750 SR and 7450 ESS DHCP server state synchronization is different from the lease state synchronization of the subscriber host itself.

- Each subscriber host lease state is synchronized by sub-mgmt client application. The host lease state can be seen by the output of following CLI command:

  show service id *id* **dhcp lease-state**

  The output of this command represents the state of our internal DHCP relay (client).

- The DHCP server lease states can be observed by the following CLI command:

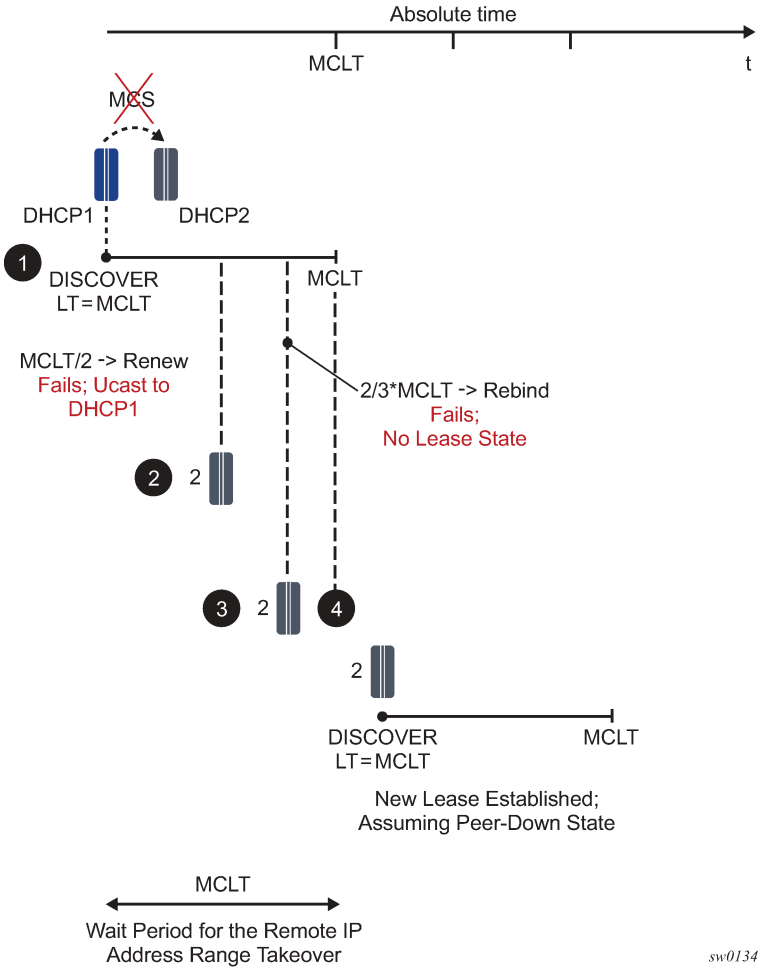
  show routed *id* **dhcp local-dhcp-server** *name* **lease**

The concern is with the latter, the 7750 SR and 7450 ESS DHCP server lease state synchronization. To ensure un-interrupted IP lease renewal process after a failure in the network, the DHCP server lease time that is synchronized between the 7750 SR and 7450 ESS DHCP servers must always lead the currently assigned lease time for the period of the anticipated lease time in the next period.

For comparison purposes the two flow diagrams are juxtaposed in Figure 29: Potential expiration time:

- The right side does not include any lead time during synchronization.
- The left side includes the lead time during synchronization.

Figure 29: Potential expiration time



On the left side of the graph, the lease time is synchronized to the same value to which it was renewed.

In point 3, the primary DHCP server renews the lease time but fails to synchronize it because of its own failure (crash). The secondary server (2) has the old lease time A' in its database. The next time the client tries to REBIND its address and prefix, the lease in the secondary server expires (4). As a result, the IP address and prefix are not renewed.

To resolve this, the primary server must synchronize the lease time as the current RENEW time plus the next lease time. This way, as depicted in point 3, when the REBIND reaches server 2, the lease in its database is active (4) and the server 2 can extend the lease for the client.

3.3.1.7 Maximum Client Lead Time

Maximum Client Lead Time (MCLT) is the maximum time that the 7750 SR and 7450 ESS DHCP server can extend the lease time to its clients beyond the lease time currently know by the 7750 SR and 7450 ESS DHCP partner node. By default, this time is relatively short (10 minutes).

The purpose of the MCLT is described in the following scenario:

The local DHCP server assigns a new IP lease to the client but it crashes before it sends a sync update to the partner server. Because of the local DHCP server failure, the remote DHCP server is not aware of the IP address and prefix that was allocated on the local DHCP server. This condition creates the possibility that the remote DHCP server allocates the same address and prefix to another client. This would cause IP address and prefix duplication. MCLT is put in place to prevent this scenario.

MCLT based solution is shown in Figure 29: Potential expiration time.

The sequence of events is the following:

1. DHCP server 1 is the local DHCP server (with the local address-range and prefix) that creates the IP lease state for a new client. The initial lease-time assigned to the client is MCLT which is normally shorter than the requested lease time.

   This DHCP server fails before it gets a chance to synchronize the lease state with the DHCP Server 2 (remote DHCP server with the remote address-range and prefix).
2. The remote DHCP server transitions into the PARTNER-DOWN state (if the partner-down timer is 0). In this state the remote DHCP server can extend the lease time to the existing clients but it cannot assign a new lease for a period of MCLT. In MCLT/2 a new RENEW request is sent directly to the local DHCP server. This server is DOWN and therefore it cannot reply.
3. The client broadcasts a REBIND request that reaches the remote DHCP server. The remote DHCP server has no knowledge of the requested lease and therefore it does not reply.
4. The lease for the client expires and the client must reinitiate the IP address and prefix assignment process.

Because the remote DHCP server is not aware of the lease state that was assigned by the local DHCP server, there is a chance that the remote DHCP server assigns to the new client the same IP address and prefix already allocated by the local DHCP server just before it crashed. Therefore the remote DHCP server needs to wait for the MCLT time to expire so that the IP addresses and prefixes allocated (but never synchronized) by the local DHCP server can time out.

When the communication channel between the chassis is interrupted, two scenarios are possible:

- The entire node becomes unavailable. In this case the redundant node takes over and it starts reducing the lease time until the lease time reaches MCLT.
  - COMMUNICATION-INTERUPTED

    The remote DHCP server only renews the leases but does not delegate new ones (primary is DOWN).

    The local DHCP server renews the leases (which eventually trickle down to MCLT) and delegates the new ones with the lease time of MCLT (secondary is down).
  - PARTNER-DOWN

    The remote DHCP server starts delegating new IP addresses and prefixes from the remote address-range and prefix after MCLT (primary is down). The lease time of the new clients is MCLT. A lease cannot be assigned for a period longer than what is agreed with the peer incremented with the MCLT. As for a "new" lease nothing is agreed yet so the sum falls back to the MCLT itself.
- The communication channel is down but the remote DHCP server is not (meaning that the clients have still access to both servers). The behavior in this case is following:
  - COMMUNNICATION-INTERUPTED

    The remote DHCP server keeps renewing existing leases but it does not delegate the new leases. There is little chance this could happen as the clients continue to send RENEWS by unicast to the

local DHCP server which is still active. The non-synched leases in the remote DHCP server times out. The local DHCP server starts ticking down the lease time to MCLT.

The local DHCP server continues to delegate the new leases, although with the MCLT lease time.

– PARTNER-DOWN State

The local DHCP server continues to extend the existing leases but also starts delegating new IP leases after the initial MCLT elapses.

Both local and remote DHCP server delegates new leases in the PARTNER-DOWN state (although the remote DHCP server in PARTNER-DOWN state must wait an additional MCLT period before it can start delegating new leases).

3.3.1.8 Sharing IPv4 address range or IPv6 prefixes

Access-driven DHCP server redundancy model ensures uninterrupted IP address assignment service from a single IP address-range and prefix if an access link forwards BNG fails. To avoid duplicate address allocation, there must be a single active path available from the clients to only one of the SR OS-based DHCP servers in redundant configuration. This single active path is ensured by a protection mechanism in dual-homed environment in the access side of the network. The supported access redundancy mechanisms are SRRP or MC-LAG.

In the access-driven DHCP redundancy model, the DHCP relay in each router of the redundant pair must point only to the IP address of its local DHCP server. In other words, the DHCP messages received on one DHCP server should never be relayed to the other. Because the IP address-ranges and prefixes are shared between the DHCP servers, accessing both DHCP servers with the same DHCP request can cause DHCP lease duplication. Moreover, the IP addresses of both DHCP servers must be the same in both nodes. Otherwise the DHCP renew process would fail.

Granting new leases out of the shared IP address-ranges or prefixes that are configured as access-driven is not dependent on the state of the inter-chassis communication link (MCS). Instead, the new leases can be granted from both nodes simultaneously and it is the role of the protection mechanism in the access to ensure that a single path to either server is always active.

This model allows the newly active node, after a SRRP/MC-LAG switchover, to be able to serve new clients immediately from the same (shared) IP range or prefix. At the same time, upon the switchover, the corresponding subscriber-interface route is re-evaluated for advertisement with a higher routing metric to the network side by SRRP awareness. The result is that by aligning the subscriber-interface routes or prefixes with access-driven DHCP address ranges or prefixes in DHCP server, the IP address ranges or prefixes are advertised to the network side only from the actively serving node (SRRP master state or active MC-LAG node). This is performed indirectly with the corresponding subscriber-interface route that is aligned (by configuration) with the DHCP address range or prefix in access-driven mode.

If SRRP or MC-LAG is not deployed in conjunction with the access-driven configuration option, the IP address duplication could occur.

Multi-chassis redundancy relies on MCS to synchronize various client applications. (subscriber states, DHCP states, IGMP states, and so on) between the two chassis. Therefore, the links over which MCS peering session is established must be highly redundant. Failed MCS peering session renders dual-chassis redundancy non-operational. Considering this fact, the DHCP failover scenario with SRRP/MC-LAG and shared IP address range should be evaluated considering the following cases described in Table 4: DHCP failover scenarios .

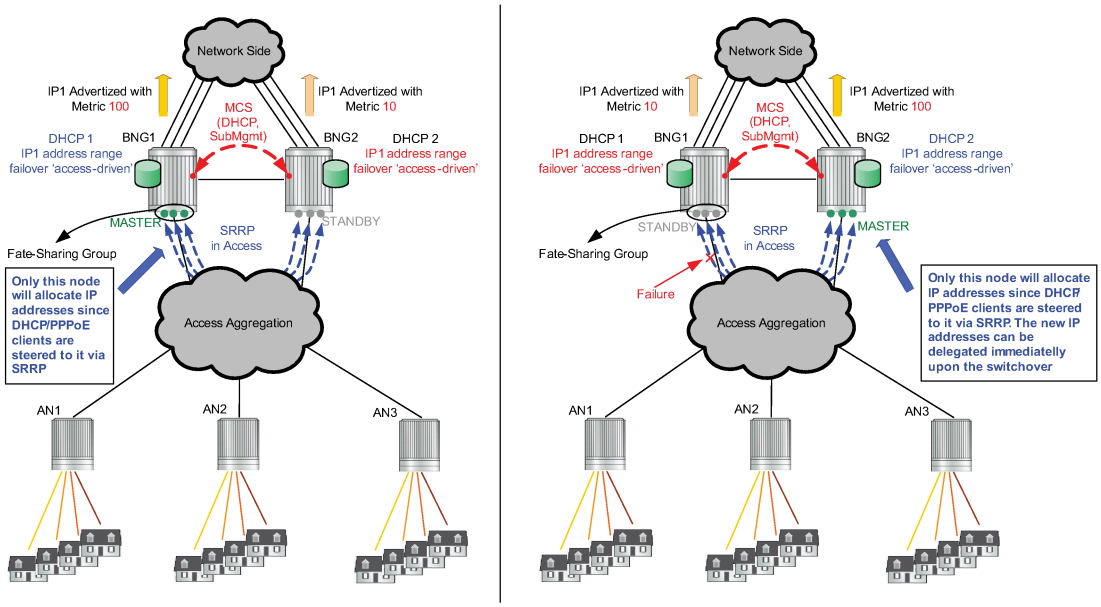
Table 4: DHCP failover scenarios

Access related failure	Inter-chassis link state	Number of failures	Action
None	None	None	Only the multichassis active BNG (SRRP master state) grants IP leases. The subnet tied to the SRRP instance is advertised to the network side on the multichassis active BNG.
None	COMM-INT	Possibly multiple	Only the (SRRP master state) BNG (SRRP master state) grants IP leases. Communication link between the two chassis has failed and the DHCP states cannot be synchronized. Operator is required to restore the links between the chassis.
None	PARTNER-DOWN	Possibly multiple	Same as above. The premise of this deployment model in general (SRRP/MC-LAG and access-driven) is that there is only one path leading to the DHCP server active. This path is governed by SRRP. Therefore, there is no change in behavior between the PARTNER-DOWN and COMM-INT states in this scenario.
Access Link Towards the multichassis active BNG (SRRP master state)	NORMAL	Single	SRRP switches over. The new IP lease grants continues from new multichassis active BNG (SRRP master state) using the same IP address range. IP leases are synched OK. Subnets are advertised from new multichassis active BNG by SRRP state awareness.
Access Link Towards the multichassis active BNG (SRRP master state)	COMM-INT	Multiple	SRRP switch over. The new leases can be handed from the DHCP server on the newly multichassis active BNG (SRRP master state). However, this DHCP server may not have its lease state table up to date because the inter-chassis communication link is non-operational. Consequently, the new multichassis active BNG may start handing out leases that are already allocated on the node with the failed link. In this case, IP address duplication would ensue. Therefore it of the utmost importance that the intercommunication chassis link is well protected and the only event that causes it to

Access related failure	Inter-chassis link state	Number of failures	Action
			go down is when the entire node goes down. Otherwise the nodes becomes isolated from each other and synchronization becomes non-operational.
Access Link Towards the multichassis active BNG (SRRP master state)	PARTNER-DOWN	Multiple	Same as above.
Access Link Towards the multichassis standby BNG (SRRP standby state)	NORMAL	Single	No effect on the operation because everything is active on the multichassis active BNG (SRRP master state) anyway.
Access Link Towards the multichassis standby BNG (SRRP standby state)	COMM-INT	Multiple	Intercommunication link is broken. The multichassis active BNG (SRRP master state) continues handing out the new leases and renewing the old ones. However, they are not synchronized to the peering node.
Access Link Towards the multichassis standby BNG (SRRP standby state)	PARTNER-DOWN	Multiple	Same as above.
Entire multichassis active BNG (SRRP master state)	COMM-INT	Single	SRRP switches over. However, lease duplication may occur on the new multichassis active BNG because the intercommunication link is broken and this new multichassis active BNG is not aware of DHCP leases that the peer (failed node) may have allocated to the clients while the intercommunication-link was broken.
Entire multichassis active BNG (SRRP master state)	PARTNER-DOWN	Single	Same as above.
Entire multichassis standby BNG (SRRP standby state)	COMM-INT	Single	Operation continues, but multichassis redundancy is lost.
Entire multichassis standby BNG (SRRP standby state)	PARTNER-DOWN	Single	Same as above.

A possible deployment scenario is shown in Figure 30: Failover scenario with SRRP and DHCP in access-driven mode.

Figure 30: Failover scenario with SRRP and DHCP in access-driven mode



3.3.1.9 Fast-switchover of IP address and prefix delegation

Some deployments require that the remote IP address and prefix range starts delegating new IP addresses and prefixes upon the failure of the intercommunication link, without waiting for the intercommunication link to transition from the COMM-INT state into the PARTNER-DOWN state and the MCLT to expire while in PARTNER-DOWN state.

In other words, the takeover of the remote IP address-range and prefix should follow the failure of the intercommunication link, without any significant delays.

This can be achieved by configuring both of the following two items under the dhcp failover CLI hierarchy:

- The partner-down-delay must be set to 0. This causes the intercommunication link to bypass the COMM-INT state upon the failure and transition straight into the PARTNER-DOWN state. The remote IP address-range and prefix can be taken over only in PARTNER-DOWN state, when the MCLT expires.
- The ignore-mclt-on-takeover flag must be enabled. With this flag enabled, the remote IP address and prefix can be taken over immediately upon entering the PARTNER-DOWN state of the intercommunication link, without having to wait for the MCLT to expire. By setting this flag, the lease times of the existing DHCP clients, while the intercommunication link is in the PARTNER-DOWN state, are reduced to the MCLT over time and all new lease times are set to MCLT. This behavior remain the same as originally intended for MCLT. this functionality must be exercised with caution. Be mindful that the partner-down-delay and MCLT timers were originally introduced to prevent IP address duplication in cases where DHCP redundant nodes transition out-of-sync because of the failure of intercommunication link. These timers (partner-down-delay and MCLT) would ensure that during their duration, the new IP addresses and prefixes are delegated only from one node, the one with local IP address-range and prefix. The drawback is that the new IP address delegation is delayed and service is impacted.

If the intercommunication link could be guaranteed to always available, then the DHCP nodes would stay in sync and the two timers would not be needed. Therefore it is important that in this mode of operation, the intercommunication link is well protected by providing multiple paths between the two DHCP nodes.

The only event that should cause intercommunication link to fail is the entire nodal failure. This failure is acceptable because in this case only one DHCP node is available to provide new IP addresses and prefixes.

3.3.1.9.1 DHCP server synchronization and local PPPoX pools

Because there is no classical DHCP lease state maintained for local PPPoX pools, the IP addresses are not synchronized by the DHCP server. Instead they are synchronized through PPPoX clients. After the PPPoX subscriber is synchronized, the respective IP address lease is updated in the respective local pool.

For example:

- A PPPoE client is created on 7750 SR and 7450 ESS 'A'.
- The IP address is assigned from the local poll on 7750 SR and 7450 ESS 'A'.
- The PPPoE client is synchronized to the peering node 7750 SR and 7450 ESS 'B'.
- After the client is synchronized in the 7750 SR and 7450 ESS 'B', the IP address assignment is synchronized by the internal PPPoE process on the 7750 SR and 7450 ESS 'B' with the local pool.

One artifact of this behavior (IP address assignment in local DHCP pools is synchronized through PPPoX clients and not by DHCP server synchronization mechanism) is that during the node boot, the DHCP server must wait for the completion of PPPoX subscriber synchronization by MCS so that it learns which addresses and prefixes are already allocated on the peering node. Because the DHCP server can theoretically start assigning IP addresses before the PPPoX sync is completed, a duplicate address assignment may occur. For example, an IP address lease can be granted by DHCP local pools while PPPoX sync is still in progress. After the PPPoX sync is completed, the DHCP server may discover that the granted IP lease has already been allocated by the peering node. The most recent lease is kept and the other is removed from both systems. To prevent this scenario, a configurable timer can be set to an arbitrary value that renders sub-if non-operational until the timer expires. The purpose of this timer is to allow the PPPoX sync to complete before subscribers under the sub-intf can be served.

3.4 Local address assignment

3.4.1 Stateless address autoconfiguration

In the stateless autoconfiguration model, hosts can be assigned address statically or dynamically. For static prefix assignment, LUDB and RADIUS can be used. For dynamic assignment, a pool name returned from LUDB or RADIUS and the local DHCPv6 server is used for address management. Although, the DHCPv6 server is used there are no lease time associated with the SLAAC prefix assigned to hosts. To use the local pool for SLAAC prefix assignment, the command **local-address-assignment** is used under group-interface. The client-application type **ppp-slaac** or **ipoe-slaac** must first be specified. Afterwards, the server name of the local-address-server must also be provisioned.

3.4.2 Local address assignment and multichassis redundancy

The internal leases for local address assignment are not synchronized through the local DHCP server multichassis synchronization (MCS) application. Instead, the SLAAC prefix is synchronized to the standby BNG through the subscriber management PPPoE or IPoE application. The standby BNG then creates

an internal lease in the local DHCP server to be used for the local address assignment. Local address assignments fail when local DHCP server failover is configured at the server or pool level.

3.5 Configuring DHCP with CLI

This section provides information to configure DHCP using the command line interface.

3.5.1 Enabling DHCP snooping

DHCP snooping is the process of copying DHCP packets and using the contained information for internal purposes. The BSA and BSR can use the snooped DHCP information to build anti-spoofing filters, populate the ARP table, send ARP replies, and so on.

For VPLS, DHCP snooping must be explicitly enabled (using the **snoop** command) on the SAP or SDP where DHCP messages ingress the VPLS instance. It is recommended to enable snooping on both the interface to the DHCP server (to snoop ACK messages) and the interface to the subscriber (to snoop RELEASE messages).

For IES and VPRN IP interfaces (VPRN is supported on the 7750 SR only), lease populate enables DHCP snooping for the subnets defined under the IP interface. The number of allowed simultaneous DHCP sessions on a SAP or interface can be limited using the **lease-populate** command with the parameter number-of-entries specified. Enabling **lease-populate** and **snoop** commands is effectively enabling "standard subscriber management".

The following output displays an example of a partial BSA configuration with DHCP snooping enabled in a service:

```
*A:ALA-48>config>service# info
----------------------------------------------
...
        vpls 600 customer 701 create
            sap 1/1/4:100 split-horizon-group "DSL-group2" create
                description "SAP towards subscriber"
                dhcp
                    lease-populate 1
                    option
                        action replace
                        circuit-id
                        no remote-id
                    exit
                    no shutdown
                exit
            exit
            mesh-sdp 2:800 create
                dhcp
                    snoop
                exit
            exit
            no shutdown
        exit
...
----------------------------------------------
*A:ALA-48>config>service#
```

3.5.2 Configuring local user database parameters

A local user data base defines a collection of host entries. There are two types of hosts: PPP and IPoE. A local user database can be used to:

- Authenticate PPP clients. For this only the host entries configured in the **ppp** CLI are matched.
- Authenticate IPoE hosts (DHCPv4, DHCPv6 IA-NA/IA-PD, SLAAC). The host entries configured in the **ipoe** CLI context are matched.
- Perform authentication and address management for the local DHCPv4 server. For this, both PPP and IPoE sections can be used depending on the client type indicated by a vendor-specific sub-option inside Option 82 of the DHCPv4 message.

Each host can be identified by a set of values. However, at any point in time only four of these values are considered for IPoE as defined by the **ipoe match-list** option and only three are considered for PPP as defined in the **ppp match-list** option.

When trying to find a matching host entry, attempts are made to match as many items as possible. If several hosts match an incoming IPoE packet, the one with most match criteria is taken.

One host entry can map on several physical clients. For instance, when using a circuit ID, by masking when the interface ID is used, the host entry is used for all the clients on that same interface.

IPoE host identification includes:

- circuit ID

  This field also matches the DHCPv6 interface-id field
- MAC address
- remote ID

  Matches on the remote-id sub-option in option 82 for DHCPv4 clients and on the remote-id option (including enterprise-id field) for DHCPv6 clients
- Option 60 from DHCPv4 message

  Only first 32 bytes are looked at
- SAP ID
- service ID
- string from vendor-specific sub-option of Option 82
- system ID
- derived-id

  a string provided via a DHCP Python script
- dual-stack-remote-id

  matches on the remote-id sub-option in option 82 for DHCPv4 clients and on the remote-id field in the remote-id option (without enterprise-id) for DHCPv6 clients
- encap-tag-range

  matches on VLAN tag ranges
- IP

  matches on the source IPv4/IPv6 address of a data-trigger packet

PPP host identification includes:

- circuit ID
- MAC address
- remote ID
- username, either complete username, domain part only, or host part only
- derived ID

  a string provided by Python script

When a host cannot be inserted in the lookup database, it is placed in an unmatched-hosts list. This can occur because

- Another host with the same host-identification exists. Only the host-identification that is specified in the match-list is considered.
- A host has no host-identification specified in the match-list.

When used for PPPoE-authentication, the fields are used as follows:

- **password**

  Verifies the PPPoE user password. This is mandatory. If no password is required then it must be explicitly set to **ignore**.
- **address**
  - **no address**

    No address information. The address must be obtained by other means, either RADIUS or DHCP server.
  - **gi-address**

    No meaning in this context. The address must be obtained by other means, either RADIUS or DHCP server.
  - **use-pool-from-client**

    No meaning in this context. The address must be obtained by other means, either RADIUS or DHCP server.
  - **pool-name**

    The address must be obtained by other means, either RADIUS or a DHCP server. When a DHCP server is used, this pool name is included in Option 82 vendor-specific sub-option.
  - **ip-address**

    This IP address is offered to the client.
- **identification-strings**

  Returns the strings used for enhanced subscriber management (ESM).
- **options**

  Only DNS servers and NBNS server are used, others are ignored.

When used from the DHCP server, the following applies:

- **password**

  Not used.
- **address**

Defines how the address must be allocated for this host.

- **no address**

  The host is not allowed. The clients mapping to this host do not get an IP address.

- **gi-address**

  Finds the matching subnet and an IP address is taken from that subnet.

- **pool-name**

  A free IP address is taken from that pool.

- **ip-address**

  This address is offered to the client.

- **use-pool-from-client**

  Use the **poolname** in the Option 82 vendor-specific sub-option. If no **poolname** is provided there, falls back to the DHCP server default (**none** or **use-gi-address**).

- **identification-strings**

  The operator can specify subscriber management strings and in which option the strings are sent back in dhcp-offer and dhcp-ack messages.

- **options**

  The operator defines which options specific to this host should be sent back in the dhcp-offer and dhcp-ack messages. The options defined here override options defined on the pool-level and subnet-level inside the local DHCP server.

The circuit ID from PPPoE or from Option 82 in IPoE messages can be masked in following ways:

- **prefix-length**

  Drop a fixed number of bytes at the beginning of the circuit-id.

- **suffix-length**

  Drop a fixed number of bytes at the end of the circuit-id.

- **prefix-string**

  The matching string is dropped from the beginning of the circuit-id. The matching string can contain wildcards (*). For example: incoming circuit-id mybox|3|my_interface|1/1/1:22 masked with *|*| leaves my_interface|1/1/1:22.

- **suffix-string**

  The matching string is dropped at the end of the circuit-id. For example: incoming circuit-id mybox|3|my_interface|1/1/1:22 masked with |* results in mybox|3|my_interface.

The following is an example of a local user database used for PPPoE authentication:

```
*A:ALA-48>config>subscr-mgmt# info
----------------------------------------------
...
        local-user-db "pppoe user db"
            description "pppoe authentication data base"
            ppp
                match-list username circuit-id
                mask prefix-string "*|*|" suffix-string "|*"
                host "john" create
                    host-identification
```

```
                        username "john" no-domain
                    exit
                    password pap "23T8yPoe0w1R.BPGHB98i0qhJf7ZlZGCtXBKGnjrIrA" hash2
                    no shutdown
                exit
                host "test.com" create
                    host-identification
                        username "test.com" domain-only
                    exit
                    password ignore
                    no shutdown
                exit
                host "john@test.com" create
                    host-identification
                        username "john@test.com"
                    exit
                    password pap "23T8yPoe0w0Tlf1yCb4hskknvTYLqA2avvBB567g3eQ" hash2
                    identification-strings 122 create
                        subscriber-id "john@test.com"
                        sla-profile-string "sla prof1"
                        sub-profile-string "subscr profile 1"
                        ancp-string "ancp string"
                        inter-dest-id "inter dest"
                    exit
                    no shutdown
                exit
                host "john@test.com on interface group-if"
                    host-identification
                        circuit-id string "group-if"
                        username "john@test.com"
                    exit
                    password pap "23T8yPoe0w1R.BPGHB98i0qhJf7ZlZGCtXBKGnjrIrA" hash2
                    address 10.1.2.3
                    no shutdown
                 exit
            exit
            no shutdown
        exit
...
-----------------------------------------------
*A:ALA-48>config>subscr-mgmt#
```

The following are some examples when a user tries to set up PPPoE:

- john@test.com tries to set up PPPoE with circuit-id pe_23|3|group-if|1/1/1: host john@test.com on interface group-if match, the PAP password is checked and the IP address 10.1.2.3 is assigned to the PPPoE to use for this host.
- john@test.com (on another interface): host john@test.com matches, the PAP password is checked, and identification strings are returned to PPPoE.
- nokie@test.com: host test.com matches, no password check, the user is allowed.
- john@nokia.com: host john matches and the password is checked.
- anybody@anydomain: does not match and is not allowed.

The following is an example of a local user database used for DHCP server for IPoE clients:

```
*A:ALA-50>config>subscr-mgmt# info
-----------------------------------------------
...
        local-user-db "dhcp server user db"
```

```
            description "dhcp server user data base"
            ipoe
                match-list circuit-id mac
                mask prefix-string "*|*|" suffix-string "|*"
                host "mac 3 on interface" create
                    host-identification
                        circuit-id string "group-if"
                        mac 00:00:00:00:00:03
                    exit
                    address 10.0.0.1
                    no shutdown
                exit
                host "maskedCircId" create
                    host-identification
                        circuit-id string "group-if"
                    exit
                    address pool "pool 1"
                    identification-strings 122 create
                        subscriber-id "subscriber 1234"
                        sla-profile-string "sla prof 1"
                        sub-profile-string "sub prof 1"
                        ancp-string "ancpstring"
                        inter-dest-id "inter dest id 123"
                    exit
                    options
                        netbios-name-server 1.2.3.4
                        lease-time min 2
                    exit
                    no shutdown
                exit
            exit
            no shutdown
        exit
...
-----------------------------------------------
*A:ALA-50>config>subscr-mgmt#
```

The following is an access example:

- MAC 00:00:00:00:00:03 on circuit-id pe5|3|group-if|1/1/1: host mac 3 on interface is matched and address 10.0.0.1 is offered to the IPoE client.
- Another MAC on circuit-id pe5|3|group-if|2/2/2: host maskedCircId is matched and an address is taken from pool1 (defined in the DHCP server). The identification-strings are copied to Option 122 in the dhcp-offer and dhcp-ack messages. The options defined here are also copied into dhcp-offer and dhcp-ack messages.
- The circuit-id pe5|3|other_group_if|1/1/3: no host is matched. The client only gets an IP address if on DHCP server level defined the use-gi-address parameter and the gi-address matches a subnet.

The following is an example of a local user database used for a DHCP server, only for PPPoE clients:

If PPPoE does not get an IP address from RADIUS or the local-user-db used for authentication, the internal dhcp-client is used to access a DHCP server which can be in the same node or in another node. These request are identified by inserting Option 82 sub-option client-id in the dhcp-discover and dhcp-request messages. When the DHCP server receives this request and has a user-db connected to it, then the PPPoE section of that user-db is accessed.

```
*A:ALA-60>config>subscr-mgmt# info
-----------------------------------------------
...
```

```
        local-user-db "pppoe user db"
            description "pppoe authentication data base"
            ppp
                match-list username
                host "internet.be" create
                    host-identification
                        username "internet.com" domain-only
                    exit
                    address "pool_1"
                    no shutdown
                exit
                host "john@internet.com" create
                    host-identification
                        username "john@internet.com"
                    exit
                    identification-strings 122 create
                        subscriber-id "john@test.com"
                        sla-profile-string "sla prof1"
                        sub-profile-string "subscr profile 1"
                        ancp-string "ancp string"
                        inter-dest-id "inter dest"
                    exit
                    address use-gi
                    no shutdown
                exit
                host "malicious@internet.com"
                    host-identification
                        circuit-id string "group-if"
                        username "internet@test.com"
                    exit
                    no shutdown
                 exit
            exit
            no shutdown
        exit
...
-----------------------------------------------
*A:ALA-60>config>subscr-mgmt#
```

The following is an access example:

- john@internet.com: GI is used to find a subnet and a free address is allocated form that subnet. Identification strings are returned in Option 122.
- anybody@internet.com: pool_1 is used to find a free IP address.
- malicious@internet.com: no address is defined. This user does not get an IP address.

The following is an example of associating a local user database to PPPoE for authentication for the 7750 SR.

```
A:pe5>config>service>vprn#
-----------------------------------------------
            subscriber-interface "tomylinux" create
               address 10.2.2.2/16
               group-interface "grp_pppoe3" create
                   pppoe
                       e "pppoe"
                   exit
               exit
-----------------------------------------------
A:pe5>config>service>vprn#
```

The following is an example of associating a local user database to a local DHCP server.

```
A:pe7>config>router>dhcp#
------------------------------------------
                local-dhcp-server my_server
                    description "my dhcp server"
                    user-db "data base 1"
                        ...
                exit
------------------------------------------
A:pe7>config>router>dhcp#
```

In PPPoE access scenarios without access node or with access nodes that do not insert PPPoE vendor specific tags Circuit-ID or Remote-ID, it may be required to configure this information in the local user database so that they can be picked up in pre-authentication phase and used for RADIUS authentication and reporting in RADIUS accounting messages. For example:

```
config>subscr-mgmt

        local-user-db "ludb-1" create
            ppp
                match-list username
                host "host-1" create
                    access-loop-information
                        circuit-id string "LUDB inserted circuit-id"
                        remote-id string "LUDB inserted remote-id"
                    exit
                    host-identification
                        username "cpe-1@domain1.com"
                    exit
                    auth-policy "auth-policy-1"
                    password ignore
                    no shutdown
                exit
            exit
```

With PPPoE, when the system accesses a LUDB during a discovery phase, a matched host could return a second LUDB via a **user-db** configuration under the LUDB host context. This second database is accessed again during the PAP/CHAP phase. The following is an example:

```
 local-user-db "padi-db" create
            ppp
                match-list derived-id
                host "testuser" create
                    host-identification
                        derived-id "testuser"
                    exit
                    msap-defaults
                        group-interface "g1"
                        service 500
                    exit
                    user-db "chap-db"
                    no shutdown
                exit
            exit
            no shutdown
        exit
        local-user-db "chap-db" create
            ppp
```

```
                match-list derived-id username
                host "testuser" create
                    host-identification
                        derived-id "testuser"
                        username "testuser"
                    exit
                    password chap "cYhRmQYWOkLW3s0LrtEnBjWlAwFa/1Kx" hash2
                    identification-strings 254 create
                        sla-profile-string "sla-2"
                    exit
                    no shutdown
                exit
            exit
            no shutdown
        exit
```

3.5.3 Configuring Option 82 handling

Option 82, or the Relay Information Option is a field in DHCP messages used to identify the subscriber. The Option 82 field can already be filled in when a DHCP message is received at the router, or it can be empty. If the field is empty, the router should add identifying information (circuit ID, remote ID or both). If the field is not empty, the router can decide to replace it.

The following output displays an example of a partial BSA configuration with Option 82 adding on a VPLS service. Snooping must be enabled explicitly on a SAP.

```
A:ALA-1>config>service>vpls#
-----------------------------------------------
            no shutdown
            description "Default tls description for service id 1"
            sap 1/1/11 split-horizon-group "2dslam" create
                dhcp
                    no description
                    snoop
                    no lease-populate
                    option
                        action replace
                        circuit-id ascii-tuple
                        no remote-id
                    exit
                    no shutdown
                exit
            exit
-----------------------------------------------
A:ALA-1>config>service>vpls#
```

3.5.4 Enabling DHCP relay

Lease populate and DHCP relay are different features in which are not both required to be enabled at the same time. DHCP relay can be performed without populating lease tables.

The following example displays DHCP relay configured on an IES interface:

```
A:ALA-48>config>service>ies>if# info
-----------------------------------------------
                address 10.10.42.41/24
                local-proxy-arp
```

```
                proxy-arp
                    policy-statement "ProxyARP"
                exit
                sap 1/1/7:0 create
                    anti-spoof ip
                exit
                arp-populate
                dhcp
                    description "relay_ISP1"
                    server 10.200.10.10 10.200.10.20
                    lease-populate 1
                    no shutdown
                exit
-------------------------------------------
A:ALA-48>config>service>ies>if#
```

4 Stateless Address Autoconfiguration (SLAAC)

Note: The information in this section applies only to the 7750 SR.

4.1 SLAAC management principles

In a Triple Play network, client devices can use SLAAC to dynamically obtain their IP address and other network configuration information.

1. During bootup, the client sends a Router Solicit (RS) message to get an IP prefix.
2. The BNG address server can assign a prefix statically to the subscriber through RADIUS or LUDB, or dynamically using the local address server.
3. The BNG address server replies to the client with a Router Advertisement which contains a /64 prefix.

4.2 Configuration overview

The ICMP6 Router Solicit is the primary trigger for SLAAC host creation. It is also possible to use the DHCPv4 message to trigger a SLAAC host creation using the "IPoE-linking" feature. The SLAAC host can use RADIUS or LUDB authentication, as well as bypass authentication. Address assignment can be assigned statically or dynamically. Static prefix assignment is accomplished through RADIUS or LUDB. Dynamic prefix assignment requires the use of the local-address-server (reusing the local DHCPv6 server), and a pool name returned from RADIUS or LUDB. The DHCPv6 server for SLAAC is used for address management only, there are no lease state associated with SLAAC users. The DHCPv6 server can be shared with regular DHCPv6 users as well.

4.3 Router-solicit trigger

The following example shows a router-solicit triggered configuration.

```
*A:eng-BNG-2>config>service>vprn>sub-if>grp-if>ipv6# info
----------------------------------------------
                        router-solicit
                            no shutdown
                        exit
```

To add authentication to the above configuration, there are two options.

- For RADIUS authentication, like DHCP and PPP authentication, add a RADIUS policy under the group interface.
- For LUDB, add the following to the **router-solicit** configuration.

```
*A:eng-BNG-2>config>service>vprn>sub-if>grp-if>ipv6# info
```

```
-----------------------------------------
                        router-solicit
                            user-db "slaac-users"
                            no shutdown
                        exit
```

4.4 SLAAC address assignment

After an RS is received to trigger the creation of a SLAAC host, address assignment can be provided statically or dynamically.

4.5 Static SLAAC prefix assignment

If using RADIUS, the attribute "framed-ipv6-prefix" VSA is used. The attribute must use a /64 prefix.

```
*A:eng-BNG-2>config>subscr-mgmt>loc-user-db>ipoe>host# info
-----------------------------------------------
                    ipv6-slaac-prefix 2001::/64
```

4.6 Dynamic SLAAC prefix assignment

SLAAC prefix can be dynamically assigned to a user at real time. Prefixes are assignment through the local DHCPv6 pool. Therefore, a DHCPv6 pool must be defined first. The following displays an example configuration.

```
*A:eng-BNG-2>config>service>vprn>dhcp6# info
-----------------------------------------------
       local-dhcp-server "dhcp6-server" create
              use-pool-from-client
              pool "pool-01" create
                     prefix 2001::/32 wan-host create
                     exit
              exit
        exit
```

To associate the DHCPv6 server for SLAAC address assignment, the following configuration is used. The server name configured under **local-address-assignment dhcp6-server** matches the name configured under the DHCPv6 pool.

```
*A:eng-BNG-2>config>service>vprn>sub-if>grp-if# info
-----------------------------------------------
                    local-address-assignment
                        ipv6
                            client-application ppp-slaac ipoe-slaac
                            server "dhcp6-server"
                        exit
                        no shutdown
                    exit
```

To specify the pool to be used for SLAAC prefix assignment, the pool name can either be returned from LUDB or RADIUS.

If using RADIUS, the attribute "Alc-slaac-ipv6-pool" is used.

If using LUDB, the following configuration is used.

```
*A:eng-BNG-2>config>subscr-mgmt>loc-user-db>ipoe>host# info
----------------------------------------------
                    ipv6-slaac-prefix-pool "pool-01"
```

In this example, the pool named "pool-01" provisioned in the LUDB or returned from RADIUS matches the pool name configured in the DHCP6 server. A prefix from the 2001::/32 pool is assigned to the SLAAC subscribers.

4.7 SLAAC prefix replacement

An SLAAC host prefix can be replaced with the VSA Alc-Ipv6-Slaac-Replacement-Prefix. This VSA is only supported through CoA or through the **tools>subscr-mgmt>coa** command. When a CoA is triggered, the original SLAAC host session terminates from the BNG. Depending on the accounting mode, an accounting stop message may be sent. Immediately following the termination of the original SLAAC host session, an SLAAC host with the new replacement prefix is created on the system. The SLAAC host inherits all the original host attributes, such as the subscriber profile and SLA profile. Nokia recommends against combining the VSA Alc-Ipv6-Slaac-Replacement-Prefix with other VSAs. An error in any of the VSAs can cause the SLAAC host recreation to fail. Service can only be restored after the subscriber performs an address request and authenticates.

The RA prefix replacement generates a single router advertisement containing both the old and new prefixes. The old prefix had both the valid and preferred lifetime parameters set to 0, informing the subscriber to deprecate the prefix as soon as possible. The new prefix has the valid and preferred lifetime parameters set as per the operator configuration. The subscriber can continue to use the old prefix for up to two hours after the RA. During the two hours, the 7750 SR drops subscriber traffic that does not match the anti-spoof criteria.

This feature is supported on an MCS setup. For persistence, the replacement SLAAC prefix is stored as the subscriber new prefix.

SLAAC prefixes that were assigned through local address assignment cannot be replaced. This feature ensures that when SLAAC replacement is performed, the address origin is not changed. This feature only replaces SLAAC host prefixes and cannot, for example, replace a DHCPv6 host with an SLAAC prefix. This feature is not supported for PPPoE sessions.

5 Point-to-Point Protocol over Ethernet management

5.1 PPPoE

Note: The information in this section applies only to the 7750 SR.

A Broadband Remote Access Server (BRAS) is a device that terminates PPPoE sessions. The Point-to-Point Protocol (PPP) is used for communications between a client and a server. Point-to-Point Protocol over Ethernet (PPPoE) is a network protocol used to encapsulate PPP frames inside Ethernet frames.

Ethernet networks are packet-based, unaware of connections or circuits. Using PPPoE, Nokia users can dial from one router to another over an Ethernet network, then establish a point-to-point connection and transport data packets over the connection. In this application subscriber hosts can connect to the router using a PPPoE tunnel. There are two command available under PPPoE to limit the number of PPPoE hosts, one to set a limit that is applied on each SAP of the group-interface and one to set the limit per group-interface.

PPPoE is commonly used in subscriber DSL networks to provide point-to-point connectivity to subscriber clients running the PPP protocol encapsulated in Ethernet. IP packets are tunneled over PPP using Ethernet ports to provide the client's software or RG the ability to dial into the provider network. Most DSL networks were built with the use of PPPoE clients as a natural upgrade path from using PPP over dial-up connections. Because the PPP packets were used, many of the client software was reusable while enhancements were made such that the client could use an Ethernet port in a similar manner as it did a serial port. The protocol is defined by RFC 2516, *A Method for Transmitting PPP Over Ethernet (PPPoE)*.

PPPoE has two phases, the discovery phase and the session phase.

- Discovery: The client identifies the available servers. To complete the phase the client and server must communicate a session-id. During the discovery phase all packets are delivered to the PPPoE control plane (CPM or MDA). The IOM identifies these packets by their Ethertype (0x8863).
  - **PPPoE Active Discovery Initiation (PADI)**

    This broadcast packet is used by the client to search for an active server (Access Concentrator) providing access to a service.
  - **PPPoE Active Discovery Offer (PADO)**

    If the access server can provide the service it should respond with a unicast PADO to signal the client it may request connectivity. Multiple servers may respond and the client may choose a server to connect to.
  - **PPPoE Active Discovery Request (PADR)**

    After the client receives a PADO it uses this unicast packet to connect to a server and request service.
  - **PPPoE Active Discovery Session-confirmation (PADS)**

    A server may respond to the client with this unicast packet to establish the session and provide the session-id. After the PADS was provided the PPP phase begins.
- **Session**

After the session ID is established connectivity is available for the duration of the session, using Ethertype 0x8864. Either client or server can terminate a session.

During the life of the session the packets may be uniquely identified by the client's MAC address and session-id. The session can terminate either by PADT sent by the client or server or by an LCP Terminate-Request packet.

During session creation, the following occurs:

- **PADI (control packet upstream)**

  This packet is delivered to the control plane. The control plane checks the service tag for service name. In the case multiple nodes are in the same broadcast domain the service tag can be used to decide whether to respond to the client. A relay tag can also be present.

- **PADO (control packet downstream)**

  The packet is generated by the control plane as response to the PADI message. The packet is forwarded to the client using the unicast packet directed at the client's MAC address. The node populates the AC-name tag and service tag. The packet sources the forwarding Ethernet MAC address of the node. If SRRP is used on the interface, it uses the gateway address as the source MAC. When in a backup state, the packet is not generated.

- **PADR (control packet upstream)**

  This packet is delivered to the control plane. The packet is destined for the node's MAC address. The control plane then generates the PADS to create the session for this request.

- **PADS (control packet downstream)**

  The control plane prepares for the session creation and sends it to the client using the client's MAC address. The session-id (16-bit value) is unique per client. The session-id is populated in the response. After a session-id is generated, the client uses it in all packets. When the server does not agree with the client's populated service tags, the PADS can be used to send a service error tag with a zero session-id to indicate the failure.

- **PADT (control packet upstream/downstream)**

  The packet is used to terminate a session. It can be generated by either the control plane or the client. The session-id must be populated. The packet is a unicast packet.

- PPP session creation supports the LCP authentication phase.

During a session, the following forwarding actions occur:

- Upstream, in the PPPoE before PPP phase, there is no anti-spoofing. All packets are sent to the CPM. During anti-spoof lookup with IP and MAC addressing, regular filtering, QoS and routing in context continue. All unicast packets are destined for the node's MAC address. Only control packets (broadcast) are sent to the control plane. Keep-alive packets are handled by the CPM.
- Downstream, packets are matched in the subscriber lookup table. The subscriber information provides queue and filter resources. The subscriber information also provides PPPoE information, such as the dest-mac-address and session-id, to build the packet sent to the client.

PPPoE-capable interfaces can be created in a subscriber interface in both IES and VPRN services (VPRN is supported on the 7750 SR only). Each SAP can support one or more PPPoE sessions depending on the configuration. A SAP can simultaneously have static hosts, DHCP leases and PPPoE sessions. See Limiting subscribers, hosts, and sessions for a detailed description of the configuration options to limit the number of PPPoE sessions per SAP, per group-interface, per SLA profile instance, or per subscriber.

RADIUS can be used for authentication. IP addresses can be provided by both RADIUS and the local IP pool, with the possibility of choosing the IP pool through RADIUS.

DHCP clients and PPPoE clients are allowed on a single SAP or group interface. If DHCP clients are not allowed, the operator should not enable lease-populate and similarly if PPPoE clients are not allowed, the operator should not enable the PPPoE node.

Note: The DHCP node can be enabled when only PPPoE clients are allowed because the DHCP relay function can be used for IP retrieval.

The DHCP lease-populate is for DHCP leases only. A similar command host-limit is made available under PPPoE for limits on the number of PPPoE hosts. The existing per sla-profile instance host limit is for combined DHCP and PPPoE hosts for that instance.

- For authentication, local and RADIUS are supported.
  - RADIUS is supported through an existing policy. A username attribute has been added.
  - For PAP/CHAP, a local user database is supported and must be referenced from the interface configuration.
- The host configuration can come directly from the local user database or from the RADIUS or DHCP server. A local host configuration is allowed through a local DHCP server with a local user database.
- IP information can be obtained from the local user database, RADIUS, a DHCP server, or a local DHCP server.

If IP information is returned from a DHCP server. PPPoE options such as the DNS name are retrieved from the DHCP ACK and provided to the PPPoE client. An open authentication option is maintained for compatibility with existing DHCP-based infrastructure.

The DHCP server can be configured to run on a loopback address with a relay defined in the subscriber or group interfaces. The DHCP proxy functionality that is provided by the DHCP relay (getting information from RADIUS, lease-split, option 82 rewriting) cannot be used for requests for PPPoE clients.

5.1.1 PPPoE authentication and authorization

5.1.1.1 General flow

When a new PPPoE session is setup, the authentication policy assigned to the group interface is examined to determine how the session should be authenticated.

If no authentication policy is assigned to the group interface or the **pppoe-access-method** is set to **none**, the local user database assigned to the PPPoE node under the group interface is queried either during the PADI phase or during the LCP authentication phase, depending on whether the match-list of the local user database contains the requirement to match on username. If the match-list does not contain the username option, PADI authentication is performed and can specify an authentication policy in the local user database host for an extra RADIUS PAP-CHAP authentication point.

If an authentication policy is assigned and the pppoe-access-method is set to PADI, the RADIUS server is queried for authenticating the session based on the information available when the PADI packet is received (any PPP username and password are not known here). When it is set to PAP-CHAP, the RADIUS server is queried during the LCP authentication phase and the PPP username and password is used for authentication instead of the username and password configured in the authentication policy.

If this authentication is successful, the data returned by RADIUS or the local user database is examined. If no IP address was returned, the DHCP server is now queried for an IP address and possibly other information, such as other DHCP options and ESM strings.

The final step consists of complementing the available information with configured default values (ESM data), after which the host is created if sufficient information is available to instantiate it in subscriber management (at least subscriber ID, subscriber profile, SLA profile, and IP address).

The information that needs to be gathered is divided in three groups, subscriber ID, ESM strings, and IP data. When one of the data sources has offered data for one of these groups, the other sources are no longer allowed to overwrite this data (except for the default ESM data). For example, if RADIUS provides an SLA profile but no subscriber ID and IP address, the data coming from the DHCP server (either through Python or directly from the DHCP option) can no longer overwrite any ESM string, only the subscriber ID and IP data. However, after the DHCP data is processed, a configured default subscriber profile is added to the data before instantiating the host.

5.1.1.2 RADIUS

Refer the 7450 ESS, 7750 SR, and VSR RADIUS Attributes Reference Guide for attributes that are applicable in RADIUS authentication of a PPPoE session.

5.1.1.3 Local user database directly assigned to PPPoE node

The following are relevant settings for a local user database directly assigned to PPPoE node:

- Host identification parameters (user name only)
- Username
- Password
- Address
- DNS servers (under DHCP options)
- NBNS servers (under DHCP options)
- Identification (ESM) strings

Incoming PPPoE connections are always authenticated through the PPPoE tree in the local user database.

The match list for a local user database that is assigned directly to the PPPoE node under the group interface is always **user-name**, independent of the match list setting in the database.

For username matching, the incoming username (user[@domain]) is always first converted to a user and a domain entity by splitting it on the first @-sign. If the no-domain parameter to the username is specified, the user component should be equal to the specific username, if the domain-only portion of the username is specified, the domain entity should be equal to the specified username and if no extra parameters are provided, the user and domain components are concatenated again and compared to the specific username.

The option number for the identification strings is not used if the local user database is assigned directly to the PPPoE node (it is only necessary if it is connected to a local DHCP server). Any valid value may be chosen in this case (if omitted, the default value chosen is 254).

If a pool name is specified for the address, this pool name is sent to the DHCP server in a vendor-specific sub-option of Option 82 to indicate from which pool the server should take the address. If the **gi-address** option is specified for the address, this is interpreted as if no address was provided.

5.1.1.4 PPP policy override parameters

The group interface PPP policy parameters apply when a PPPoE session is created in the system. At PPPoE session authentication, it is possible to override the following PPP policy parameters such that they can have different values per session or group of sessions.

5.1.1.4.1 pado-delay

An override of the **pado-delay** command applies to PADI authentication only in the following circumstances:

- via local user database authentication
  - **MD-CLI**

```
configure subscriber-mgmt local-user-db "ludb-1" ppp
    host "user-1@csp.net" {
        pado-delay 2
    }
```

  - **classic CLI**

```
A:node-2>config>subscr-mgmt# info
        local-user-db "ludb-1" create
            ppp
                host "user-1@csp.net" create
                    pado-delay 2
                exit
            exit
        exit
```

  The system ignores the **pado-delay** command in the local user database when not applicable, such as during PAP or CHAP authentication.
- via RADIUS authentication, using the 26.6527.34 Alc-PPPoE-PADO-Delay VSA in an Access-Accept

5.1.1.4.2 max-sessions-per-mac

An override of the **max-sessions-per-mac** command applies to PADI authentication only in the following circumstance:

- via local user database authentication
  - **MD-CLI**

```
configure subscriber-mgmt local-user-db "ludb-1" ppp
    host "user-1@csp.net" {
        ppp-policy-parameters {
            max-sessions-per-mac 1
        }
    }
```

  - **classic CLI**

```
A:node-2>config>subscr-mgmt# info
        local-user-db "ludb-1" create
```

```
            ppp
                host "user-1@csp.net" create
                    ppp-policy-parameters
                        max-sessions-per-mac 1
                    exit
                exit
            exit
        exit
```

The system ignores the **max-sessions-per-mac** command in the local user database when not applicable, such as during PAP or CHAP authentication.

5.1.1.4.3 LCP keepalive interval and hold-up-multiplier

An override of the LCP keepalive **interval** and **hold-up-multiplier** commands applies to both PADI and PAP or CHAP authentication in the following circumstances:

- via local user database authentication
  - **MD-CLI**

```
configure subscriber-mgmt local-user-db "ludb-1" ppp
    host "user-1@csp.net" {
        ppp-policy-parameters {
            keepalive {
                hold-up-multiplier 2
                interval 15
            }
        }
```

  - **classic CLI**

```
host "user-1@csp.net" create
     ppp-policy-parameters
          keepalive 15 hold-up-multiplier 2
          exit
     exit
```

- via RADIUS authentication, by including the 241.26.6527.92 Alc-PPPoE-LCP-Keepalive-Interval and 241.26.6527.93 Alc-PPPoE-LCP-Keepalive-Multiplier VSAs in an Access-Accept
- via Diameter Gx policy management, by including the 92 Alc-PPPoE-LCP-Keepalive-Interval and 93 Alc-PPPoE-LCP-Keepalive-Multiplier Nokia specific AVPs in the Charging-Rule-Definition AVP of a CCA message at initial PPPoE session setup

```
Charging-Rule-Install ::= <AVP Header: 1001>
+-- Charging-Rule-Definition <AVP Header: 1003>
    +-- Charging-Rule-Name <AVP Header: 1005> = “LCP keepalive”
    +-- Alc-PPPoE-LCP-Keepalive-Interval <AVP Header; vendor 6527; 92>
    +-- Alc-PPPoE-LCP-Keepalive-Multiplier <AVP Header; vendor 6527; 93>
```

Both LCP keepalive interval and hold-up multiplier override parameters must be specified when overriding.

The active LCP keepalive interval and hold-up multiplier values for a PPPoE session is available in the output of the following CLI commands.

```
show service id ppoe session detail
show service id ppp session detail
```

LCP keepalive parameter overrides apply to PPPoE PTA sessions and L2TP LNS sessions.

5.1.1.5 Subscriber per PPPoE session index

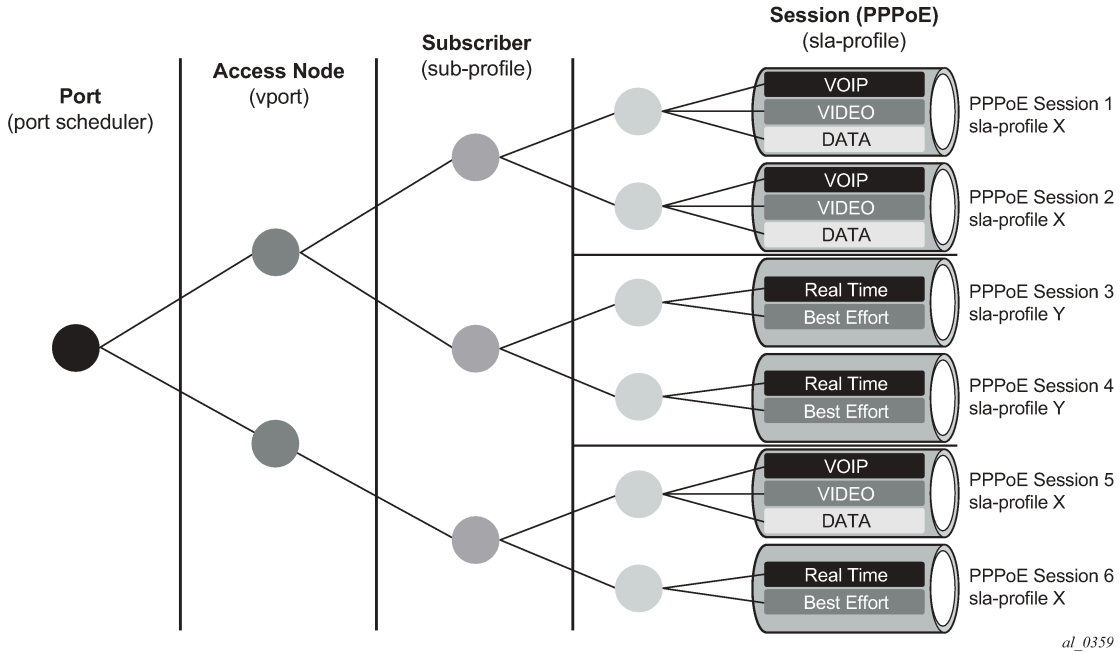
The system keeps track of the number of PPPoE sessions active on a specific SAP and assign a per SAP session index to each such that always the lowest free index is assigned to the next active PPPoE session. When PAP/CHAP RADIUS authentication is used, the PPPoE SAP session index can be sent to, and received from, the RADIUS server using the following VSA:

```
ATTRIBUTE Alc-SAP-Session-Index           180      integer
```

This is supported for all PPPoE sessions, including those using LAC and LNS, but is not supported in a dual-homing topology. It should only be used in a subscriber per VLAN model as the session index is per SAP.

The SAP session index allows PPPoE sessions to have their own set of queues for QoS and accounting purposes when using the same SLA profile name as that received from a RADIUS server. An example of this with multiple levels of HQoS egress scheduling is shown in Figure 31: Egress QoS per PPPoE session. Alternatively, this can be achieved by configuring per-session SPI sharing in the SLA profile as described in SLA Profile Instance Sharing.

Figure 31: Egress QoS per PPPoE session



This requires a set of identical SLA profiles to be configured which only differ by an index being, for example, appended to their name. The SAP session index must be sent to RADIUS in the Access-Request message, which is achieved by configuring the RADIUS authentication policy to include it as follows:

Example:

```
    configure subscriber-mgmt authentication-policy name
        include-radius-attribute
```

```
            [no] sap-session-index
```

The RADIUS server must then reflect the SAP session index back to the system in the RADIUS Access-Accept message together with the SLA profile name.

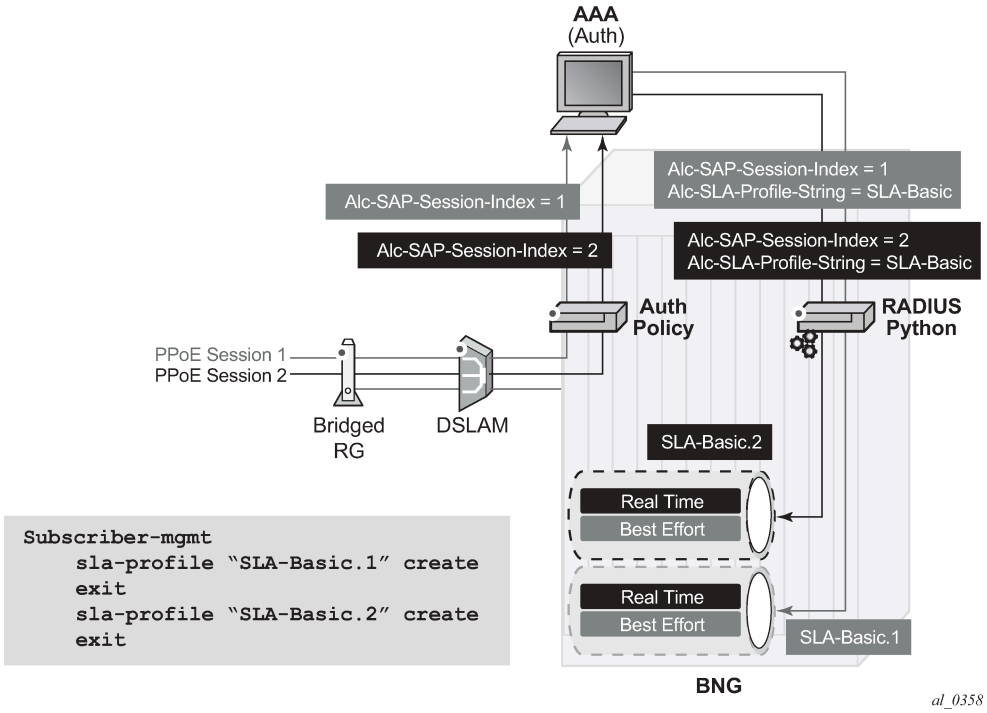
A Python script processes the RADIUS Access-Accept message to append the SAP session index to the SLA profile name to create the unique SLA profile name, in this example with the format:

sla-profile *sla-profile-name*.**suffix**

The exact format (for example, the separator used) is not fixed and merely needs to match the pre-provisioned SLA profiles, while not exceeding 16 characters. This ensures that each PPPoE session is provided its own SLA profile and consequently its own set of queues.

This processing is shown in Figure 32: Per PPPoE session SLA profile selection.

Figure 32: Per PPPoE session SLA profile selection



Below is an example Python script for this purpose:

```
import alc
import struct
from alc import radius
from alc import sub_svc

PROXY_STATE = 33
ALU = 6527
SLA_PROF_STR = 13
SAP_SESSION_INDEX = 180


################################
## QoS for Multiple PPPoE Sessions
#   This script checks if a sap-session-
```

```
index (sid) is included in the authentication
#   accept. If present, the sla-profile-string (sla) is adapted to "sla.sid"

if alc.radius.attributes.isVSASet(ALU,SLA_PROF_STR):
    sla = alc.radius.attributes.getVSA(ALU,SLA_PROF_STR)
    if alc.radius.attributes.isVSASet(ALU,SAP_SESSION_INDEX):
        ssi = alc.radius.attributes.getVSA(ALU,SAP_SESSION_INDEX)
        suffix = "" .join(["%x" % ord(x) for x in ssi])
        alc.radius.attributes.setVSA(ALU,SLA_PROF_STR,sla + '.' + "%d" %
int(suffix,16))
```

To use a CoA to change the SLA profile used, the new SLA profile name must be constructed with the same suffix (in this example) as that used for the current SLA profile. This is necessary to ensure unique use of a specifically provisioned SLA profile. This mandates that the SAP session index is included in the CoA information. Two options are proposed to achieve this:

- The CoA can specify a new SLA profile name and include the SAP session index. A Python script would then process the CoA and construct the new SLA profile name to be used by appending the suffix in the same way as was done with the RADIUS Access-Accept.
- The CoA could be using a RADIUS proxy which may make the first option unattractive. An alternative solution would be to use a Python script to append the suffix to the acct-session-id in all messages sent so that the suffix can be identified when a CoA is received that uses the acct-session-id(+suffix) for session identification. This would need to be performed for all messages sent that include the acct-session-id. CoAs would reference the session using the acct-session-id+suffix. A Python script would be required to remove the suffix and append it to the new SLA profile name. All messages received with the acct-session-id+suffix would be processed by the Python script to remove the suffix before sending the acct-session-id to the system.

  To ensure that the acct-session-id sent in RADIUS accounting messages is updated with the suffix, the user must configure **include-radius-attribute sla-profile** in the RADIUS accounting policy to be applied. The Python script needs to remove the suffix from the SLA profile and add it to the **acct-session-id** for all messages sent. Clearly, the **acct-session-id** used by any external server would then be different to that seen on the system.

5.1.1.6 Local DHCP server with local user database

If a DHCP server is queried for IP or ESM information, the following information is sent in the DHCP request:

- **Option 82 sub-options 1 and 2 (Circuit-ID and Remote-ID)**

  These options contain the circuit-ID and remote-ID that were inserted as tags in the PPPoE packets).
- **Option 82 sub-option 9 vendor-id 6527 VSO 6 (client type)**

  This value is set to 1 to indicate that this information is requested for a PPPoE client. The local DHCP server uses this information to match a host in the list of PPPoE users and not in the DHCP list.
- **Option 82 sub-option 6 (Subscriber-ID)**

  This option contains the username that was used if PAP/CHAP authentication was performed on the PPPoE connection.
- **Option 82 sub-option 13 (DHCP pool)**

This option indicates to the DHCP server that the address from the client should be taken from the specified pool. The local DHCP server only honors this request if the **use-pool-from-client** option is configured in the server configuration.

- **Option 82 sub-option 14 (PPPoE Service-Name)**

  This option contains the service name that was used in the PADI packet during PPPoE setup.

- **Option 60 (Vendor class ID)**

  This option contains the string "ALU7XXXSBM" to identify the DHCP client vendor.

- **Option 61 (Client Identifier) (optional)**

  When **client-id mac-pppoe-session-id** is configured in the **config**>**service**>**vprn**>**sub-if**>**grp-if**>**pppoe**>**dhcp-client** or **config**>**service**>**ies**>**sub-if**>**grp-if**>**pppoe**>**dhcp-client** context, the client identifier option is included and contains a type-value with type set to zero and value set to a concatenation of the PPPoE client MAC address and the PPPoE session ID.

- The WT-101 access loop options are not sent for PPPoE clients

Local user database settings relevant to PPPoE hosts when their information is retrieved from the local DHCP server using this database:

- Host identification parameters (including username)
- address
- DNS servers (under DHCP options)
- NBNS servers (under DHCP options)
- identification (ESM) strings

For username matching, the incoming username (user[@domain]) is always first converted to a user and a domain entity by splitting it on the first @-sign. If the no-domain parameter to the username is provided, the user component should be equal to the specified username, if the domain-only portion of the username is provided, the domain entity should be equal to the specified username and if no extra parameters are provided, the user and domain components are concatenated again and compared to the specified username.

To prevent load problems, if DHCP lease times of less than 10 minutes are returned, these are not accepted by the PPPoE server.

5.1.2 PPPoE session ID allocation

The following two parameters in the PPP policy control the PPPoE session ID allocation method in the discovery phase.

- **config**>**subscr-mgmt**>**ppp-policy**>**sid-allocation** {**sequential** | **random**}
- **config**>**subscr-mgmt**>**ppp-policy**>**unique-sid-per-sap** [**per-msap**]

The following list describes the various combinations of these two parameters:

The session ID range is 1 to 8191.

- **sid-allocation** sequential and **no unique-sid-per-sap** (the default allocation method)

  Each first PPPoE session with a specific client MAC address and active on a specific SAP or MSAP is configured with a session ID value of 1. The session ID for subsequent PPPoE sessions with the same client MAC address and active on the same SAP or MSAP is allocated in sequentially-increasing order.

The PPPoE session ID is unique per (client MAC address, SAP) and per (client MAC address, MSAP).

- **sid-allocation sequential** and **unique-sid-per-sap**

  Each PPPoE session that is active on a specific SAP is assigned a unique, sequentially-increasing session ID starting with a session ID value of 1.

  A unique, sequentially-increasing session ID starting with a session ID value of 1 is assigned per capture SAP: PPPoE sessions with the same or different client MAC address and that are active on the same or different MSAP have a unique session ID per capture SAP.

  The PPPoE session ID is unique per SAP and per capture SAP.

  For a **unique-sid-per-sap**, the maximum number of PPPoE sessions per SAP or per capture SAP is 8191. This limit is enforced across all derived MSAPs.

- **sid-allocation sequential** and **unique-sid-per-sap per-msap**

  Each PPPoE session that is active on a specific SAP is assigned a unique, sequentially-increasing sessions ID starting with a session ID value of 1.

  A unique, sequentially-increasing session ID starting with a session ID value of 1 is assigned per capture SAP. PPPoE sessions with the same or different client MAC address and that are active on the same MSAP have a unique session ID.

  The PPPoE session ID is unique per SAP and per MSAP.

  For a **unique-sid-per-sap per-msap**, the maximum number of PPPoE sessions per SAP or per MSAP is 8191.

- **sid-allocation random** and **no unique-sid-per-sap**

  Each PPPoE session with the same client MAC address and active on the same SAP or MSAP is assigned a unique, randomly-assigned session ID.

  The PPPoE session ID is unique per client MAC address and SAP and per client MAC address and MSAP.

- **sid-allocation random** and **unique-sid-per-sap**

  Each PPPoE session that is active on a specific SAP is assigned a unique, randomly-assigned session ID.

  A unique session ID is randomly-assigned per capture SAP: PPPoE sessions with the same or different client MAC address and that are active on the same or different MSAP have a unique session ID per capture SAP.

  The PPPoE session ID is unique per SAP and per capture SAP.

  For a **unique-sid-per-sap** configured, the maximum number of PPPoE sessions per SAP or per capture SAP is 8191. This limit is enforced across all derived MSAPs.

- **sid-allocation random** and **unique-sid-per-sap per-msap**

  Each PPPoE session that is active on a specific SAP or MSAP is assigned a unique, randomly-assigned session ID.

  The PPPoE session ID is unique per SAP and per MSAP.

  With **unique-sid-per-sap per-msap** configured, the maximum number of PPPoE sessions per SAP or per MSAP is 8191.

5.1.3 Multiple Sessions per MAC Address

To support MAC-concentrating network equipment, which translates the original MAC address of a number of subscribers to a single MAC address toward the PPPoE service, the SR OS supports up to 8191 PPPoE sessions per MAC address. Each of these sessions are identified by a unique combination of MAC address and PPPoE session ID.

To set up multiple sessions per MAC, the following limits should be set sufficiently high:

- Maximum sessions per MAC in the PPPoE policy
- The PPPoE interface session limit in the PPPoE node of the group interface
- The PPPoE SAP session limit in the PPPoE node of the group interface
- The multiple-subscriber-sap limit in the subscriber management node of the SAP

If host information is retrieved from a local DHCP server, care must be taken that, although a host can be identified by MAC address, circuit ID, remote ID or username, a lease in the DHCP server is, by default, only indexed by MAC address and circuit ID. For example, multiple sessions per MAC address are only supported in this scenario if every host with the same MAC address has a unique Circuit-ID value.

To enable IPv4 address allocation using the internal DCHCPv4 client for multiple PPPoE sessions on a single SAP and having the same MAC address and circuit-ID, the optional CLI parameter **allow-same-circuit-id-for-dhcp** should be added to the **max-sessions-per-mac** configuration in the PPP policy. The SR OS local DHCP server detects the additional vendor-specific options inserted by the internal DCHCPv4 client and uses an extended unique key for lease allocation.

5.1.4 Session limit per circuit ID

The **config**>**subscr-mgmt**>**ppp-policy max-sessions-per-cid** command limits the number of PPPoE sessions with the same Agent Circuit ID that can be active on the same SAP or MSAP.

The limit is enforced in the discovery phase, before PAP or CHAP authentication based on the Agent Circuit ID sub-option that is present in the vendor-specific PPPoE access loop identification tag added in PADI and PADR messages by a PPPoE intermediate agent.

By default, PPPoE PADI messages without the Agent Circuit ID sub-option are dropped when a **max-sessions-per-cid** limit is configured.

```
3502 2020/06
16 11:02:57.949 UTC MINOR: SVCMGR #2214 Base Managed SAP  creation failure

"The system could not create Managed SAP:1/1/
4:2111.100, MAC:00:00:64:6b:01:07, Capturing SAP:1/1/
4:*.*, Service:10. Description: Agent Circuit ID sub-
option missing in PADI and per circuit id limit configured"
```

The default behavior can be overruled with the optional allow-sessions-without-cid keyword. PPPoE sessions without an Agent Circuit ID can be established on a SAP with a max-sessions-per-cid limit configured. The **max-sessions-per-cid max-sessions-per-cid** limit is not applied to these sessions.

When the **max-sessions-per-cid** limit is not configured, no limit is placed on the number of PPPoE sessions with the same Agent Circuit ID on the same SAP or MSAP and PPPoE sessions with or without an Agent Circuit ID can be established.

5.1.5 PPP session re-establishment

The **re-establish-session** command allows a host to re-establish a PPP session if the previous PPP session has yet to be terminated. The only way for a host to reconnect is to wait for the health check to fail or a manual termination by the operator. To allow a faster reconnect, the feature allows a PADI request to terminate the host previous session and allow the host to re-establish a new PPP session. As the old PPP session terminates, the accounting record also stops. The new PPP session starts a new accounting record; this is to ensure that the subscriber is not charged for the unused time in the previous PPP session.

Note:

- Subscribers that use credit-control-policy along with PPP re-establish-session can experience a longer attempt at re-establishing a PPP connection.
- When PPP subscribers closes an L2TP connection, the terminate cause on the accounting record would show "user-request". For cases, where PPP re-establishment is enabled, stale connections are closed without requests. The terminate cause on the accounting record would show "loss of service".

5.1.6 Private retail subnets

IPoE and PPPoE are commonly used in residential networks and have expanded into business applications.

Both IPoE and PPPoE subscriber hosts terminate in a retail VPRN. It is possible for the subscriber to connect to one or more 7750 SRs for dual homing purposes. When the subscriber is dual-homed, routing between the two BNGs is required and can be performed with either a direct spoke SDP between the two nodes or by MP-BGP route learning.

For PPPoE, both PADI and PAP/CHAP authentication are supported. The PPPoE session is negotiated with the parameters defined by the retail VPRN interface. Because the IP address space of the sub-mgmt host may overlap between VPRN services, the node must anti-spoof the packets at access ingress with the session ID.

When the **config**>**service**>**vprn**>**sub-if**>**private-retail-subnets** command is enabled on the subscriber interface, the retail IP subnets are not pushed into the wholesale context. This allows IP addresses to overlap between different retail VPRNs (those VPRNs can have IPoE or PPPoE sessions). If an operator requires both residential and business services, two VPRNs connected to the same wholesaler can be created and use the flag in only one of them.

To update a subscriber attribute (such as an SLA-profile or subscriber-profile), either perform a RADIUS CoA or enter the **tools**>**perform**>**subscriber-mgmt**>**coa** command. To identify a subscriber for a CoA, the subscriber IP address can be used as part of the subscriber's key. Because it is possible for different subscribers to use the same IP address in multiple private retail VPRNs, additional parameters are required in addition to the subscriber IP address. When performing a CoA on a retail subscriber for PPPoE hosts using a private retail subnet, the following conditions apply:

- If NAS-port-ID is used, the VSA Alc-Retail-Serv-Id or Alc-Retail-Serv-Name must be included. The subscriber IP address and prefix must also be included.
- If Alc-serv-ID or Alc-Serv-Name is used, the subscriber retail service ID or name must be referenced. The subscriber IP address and prefix must also be included.

When performing a CoA on a retail subscriber for IPoE hosts using a private retail subnet, the following conditions apply:

- If NAS-port-ID is used, the VSA Alc-Client-Hardware-Addr referencing the subscriber MAC and Alc-Retail-Serv-Name must be included. The VSA Alc-Retail-Serv-Id must not be included. The subscriber IP address and prefix must also be included.
- If Alc-serv-ID or Alc-Serv-Name is used, it must reference the subscriber wholesale service ID or name. The VSA Alc-Client-Hardware-Address and the subscriber IP address and prefix must also be included.

To perform a DHCP force renew on IPoE hosts, enter the **tools>perform>subscriber-mgmt>forcerenew** command. To identify between subscribers within a private retail VPRN where overlapping IP addresses are possible, the MAC address must be used instead of the IP address in the **tools** command.

Private retail subnet are supported on both numbered and unnumbered subscriber interfaces. RADIUS host creation is possible for IPoEv4 hosts.

When IPoE and PPPoE session terminates in the retail VPRN, the node must learn the retail VPRN service ID. This can be provided by the LUDB or RADIUS. If the local user database is used, the host configuration provides a reference to the VPRN service ID. If RADIUS is used, RADIUS can return the service ID or name VSA. The retail subscriber interface must also reference the subscriber interface on the wholesale subscriber interfaces.

The following are not supported on private retail subnets:

- ARP hosts
- static hosts

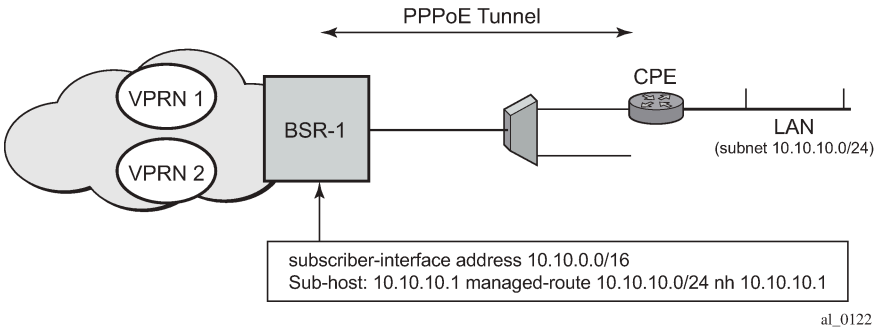
ESM multicast is not supported for IPoE hosts that use overlapping IP address between private retail VPRNs. ESM multicast is only supported for IPoE hosts that have unique IP addresses in the system. It is possible for a IPoE subscriber to contain both type of hosts, only the hosts with overlapping private IP address do not support ESM multicast.

5.1.7 IPCP subnet negotiation

This feature enables negotiation between Broadband Network Gateway (BNG) and customer premises equipment (CPE) so that CPE is allocated both ip-address and associated subnet.

Some CPEs use the network up-link in PPPoE mode and perform dhcp-server function for all ports on the LAN side. Instead of wasting one subnet for P2P uplink, CPEs use allocated subnet for LAN portion as shown in Figure 33: CPEs network up-link mode.

Figure 33: CPEs network up-link mode



From a BNG perspective, the specified PPPoE host is allocated a subnet (instead of /32) by RADIUS, external dhcp-server, or local-user-db. And locally, the host is associated with managed-route. This managed-route is a subset of the subscriber-interface subnet, and also, subscriber-host ip-address is from

managed-route range. The negotiation between BNG and CPE allows CPE to be allocated both ip-address and associated subnet.

5.1.7.1 Numbered WAN support for Layer 3 RGs

Numbered WAN interfaces on RGs is useful to manage v6-capable RGs. Dual-stack RGs can be managed through IPv4. However, with v6-only RGs or with dual-stack RGs with private only v4 address, RGs require a globally routable v6 WAN prefix (or address) for management. This feature provides support to assign WAN prefix to PPP based Layer 3 RG using SLAAC. The feature also adds a new RADIUS VSA (Alc-PPP-Force-IPv6CP, Boolean) to control triggering of IPv6CP on completion of PPP LCP. RA messages are sent as soon as IPv6CP goes into open state, and the restriction to hold off on sending RAs until DHCP6-PD is complete if a dual-stack PPP is no longer applicable.

5.1.8 IES as retail service for PPPoE host

In this application, the PPPoE subscriber host terminates in a retail IES service. The IES service ID or name can be obtained by the Alc-Retail-Serv-Id or Alc-Retail-Serv-Name attribute in the RADIUS Access-Accept packet.

Note:
Be aware that Alc-Retail-Serv-Name takes precedence over Alc-Retail-Serv-Id if both are specified. The Access-Accept packet for initial authentication can contain both the Alc-Retail-Serv-ID and Alc-Retail-Serv-Name, but the Access-Accept packet for re-authentication and CoA can have only one AVP that the subscriber session is using.

If MSAP is used, the SAP is created in the wholesale VPRN.

The PPPoE session is negotiated with the command options defined by the wholesale VPRN group interface. The connectivity to the retailer is performed using the linkage between the two interfaces.

Because of the nature of IES service, there is no IP address overlap between different IES retail services; therefore, the private-retail-subnets flag is not needed in this case.

5.1.9 Unnumbered PPPoE

Unlike regular IP routes which are mainly concerned with next-hop information, subscriber-hosts are associated with an extensive set of parameters related to filtering, qos, stateful state (PPPoE/DHCP), antispoofing, and so on. Forwarding Database (FDB) is not suitable to maintain all this information. Instead, each subscriber host record is maintained in separate set of subscriber-host tables.

By pre-provisioning the IP prefix (IPv4 and IPv6) under the **subscriber-interface** and **sub-if**>**ipv6** CLI hierarchy, only a single prefix aggregating the subscriber host entries is installed in the FDB. This FDB entry points to the corresponding subscriber-host tables that contain subscriber-host records.

When a IPv4/IPv6 prefix is not pre-provisioned, or the subscriber-hosts falls out of pre-provisioned prefix, each subscriber-host is installed in the FDB. The result of the subscriber-host FDB lookup points to the corresponding subscriber-host record in the subscriber-host table. This scenario is referred to as unnumbered subscriber-interfaces.

Unnumbered does not mean that the subscriber hosts do not have an IP address or prefix assigned. It only means that the IP address range out of which the address or prefix is assigned to the host does not have to be known in advance through configuration under the **subscriber-interface** or **sub-if**>**ipv6** node.

An IPv6 example would be:

```
configure
    router/service
        subscriber-interface <name>
            ipv6
                [no] allow-unmatching-prefixes
                delegated-prefix-length <bits>
                subscriber-prefixes
```

This CLI indicates the following:

- There is no need for any indication of anticipated IPv6 prefixes in CLI.
- However, the **delegated-prefix-length** (DPL) command is required. The DPL (or the length of the prefix assigned to each Residential Gateway) must be known in advance. All Residential Gateways (or subscribers) under the same **subscriber-interface** share this pre-configured DPL.
- The DPL range is 48 to 64.
- If the prefix length in the received PD (through the DHCP server, RADIUS or LUDB) and the DPL do not match, the host creation fails.
- If the assigned IP prefix/address (DHCP server, RADIUS, LUDB) for the host falls outside of the CLI defined prefixes and the **allow-unmatching-prefixes** command is configured, then the new address and prefix automatically installs in the FDB.

5.1.10 Selective backhaul of PPPoE traffic using an Epipe service

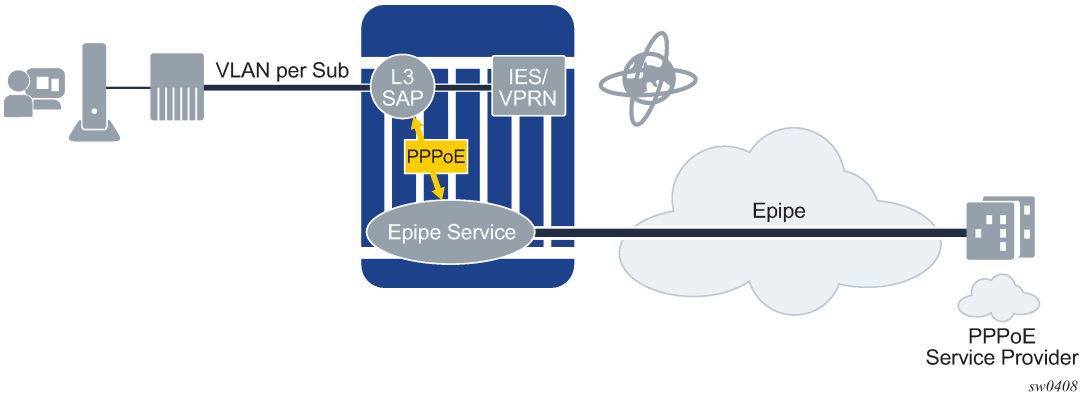
The router supports the redirection of PPPoE packets on ingress to a Layer 3 static subscriber SAP and PW-SAPs (for example, bound to an IES or VPRN service) in the upstream direction toward an Epipe service. The Epipe, which is bound to a specific SAP on a group interface of a Layer 3 service, is used to backhaul the PPPoE traffic toward a remote destination, such as, a wholesale service provider. Other non-PPPoE packets are still forwarded to the group interface on the IES or VPRN.

There is a one-to-one mapping from backhaul Epipe to the subscriber. A single backhaul Epipe service per subscriber SAP is supported. Only static configuration of subscriber SAPs is supported.

In the downstream direction, all traffic arriving on the backhaul Epipe is forwarded to the subscriber SAP and merged with IPoE traffic.

This architecture is shown in Figure 34: Redirection of traffic to Epipe for backhaul.

Figure 34: Redirection of traffic to Epipe for backhaul



If the SAP on the Layer 3 service is configured to indicate PPPoE redirection, in addition to ESM anti-spoofing, then the following processing occurs in the upstream direction:

- All PPPoE traffic, such as traffic with Ethertype 0x8863 and 0x8864, skips the anti-spoof lookup and is then forwarded to a configured Epipe service.
- All non-PPPoE is subject to anti-spoofing lookup. If the anti-spoofing fails, then the packet is dropped. Otherwise, processing continues using the host configuration.

The **fwd-wholesale** context in the Layer 3 SAP is used to configure the forwarding of PPPoE packets toward a specified Epipe service:

```
config
   service
      ies | vprn
         subscriber-interface
            group-interface
               sap 1/1/1:10.20
                  anti-spoof ...
                  fwd-wholesale
                     pppoe 10
                     exit
```

The PPPoE option specifies that only packets matching Ethertype 0x8863 and 0x8864 are redirected to the Epipe of service ID '10'. This includes all PPPoE control plane packets. The service IDs specified under **fwd-wholesale** cannot refer to a vc-switching Epipe service.

When **fwd-wholesale** is configured to an Epipe with a specified service ID, the system ensures that the Epipe exists and meets all the requirements to participate in the PPPoE redirect. There is no need for additional configuration under the Epipe. For the previous example, only the following configuration is required for the Epipe:

```
epipe 10 name "10" customer 1 create
   service-mtu 1400
   spoke-sdp 10:9 create
      no shutdown
      exit
   no shutdown
   exit
```

The system generates a CLI error if a user tries to configure additional SAPs on an Epipe that is already referenced from a **fwd-wholesale** context.

In the upstream direction, subscriber queues are used for IPoE packets destined for a local host and PPPoE backhaul traffic. However, CPM traffic is not consuming subscriber queue resources; QoS profiles are instantiated while **single-sub-parameters profiled-traffic-only** is enabled.

In the downstream direction, PPPoE traffic arriving on the Epipe is merged back into the subscriber SAP referencing the Epipe service. ESM traffic from the host uses the subscriber queues that the host is configured to use, and the backhaul traffic uses the SAP queues. If **profile-traffic-only** is configured, then all traffic uses the SAP queues.

The operational status of the Epipe with a matching service ID can only be up if the corresponding Layer 3 SAP's operational status is up. The operational or administrative status of the Epipe does not affect the status of the Layer 3 service SAP. If the Layer 3 service SAP is up, but the backhaul Epipe is down, then the system continues to redirect packets to the Epipe, but they are dropped by the Epipe service. The status of the Epipe service is indicated in the output of the **show**>**service**>**service-using** command.

5.1.11 PPPoE interoperability enhancements

Interoperability with non-conforming PPPoE client implementations is supported with the following behavior:

- The PPPoE client continues to send IPCP renegotiation messages (such as ConfReq) after negotiation is complete. BNG terminates the PPP session by default, however, the following command allows BNG to ignore the renegotiation messages:

  Classic CLI:

  config>**subscr-mgmt**>**ppp-policy**>**ncp-renegotiation ignore**

  MD-CLI:

  configure subscriber-mgmt ppp-policy ncp-renegotiation
- The PPPoE client sets incorrect identifier values in the Echo-Reply messages that are sent to BNG. BNG discards such messages and the PPP session is terminated because of the Echo timeout. However, the following command instructs BNG to ignore the Identifier field of the Echo-Reply message to keep the PPP session up:

  Classic CLI:

  config>**subscr-mgmt**>**ppp-policy**>**lcp-ignore-identifier**

  MD-CLI:

  configure subscriber-mgmt ppp-policy lcp-ignore-identifier

5.2 MLPPPoE with LFI on LNS

MLPPPoX is generally used to address bandwidth constraints in the last mile. The following are other uses for MLPPPoX:

- To increase bandwidth in the access network by bundling multiple links and VCs together. For example it is less expensive for a customer with an E1 access to add another E1 link to increase the access bandwidth, instead of to upgrade to the next circuit speed (E3).
- LFI on a single link to prioritize small packet size traffic over traffic with large size packets. This is needed in the upstream and downstream direction.

PPPoE and PPPoEoA/PPPoA v4/v6 host types are supported.

5.2.1 Terminology

The term MLPPPoX is used to reference MLPPP sessions over ATM transport (oA), Ethernet over ATM transport (oEoA) or Ethernet transport (oE). Although MLPPP in subscriber management context is not supported natively over PPP/HDLC links, the terms MLPPP and MLPPPoX terms can be used interchangeably. The reason for this is that link bundling, MLPPP encapsulation, fragmentation and interleaving can be in a broader scope observed independently of the transport in the first mile. However, MLPPPoX terminology prevails in this section in an effort to distinguish MLPPP functionality on an ASAP MDA (outside of ESM) and MLPPPoX in LNS (inside of ESM).

The terms speed and rate are interchangeably used throughout this section. Usually, speed refers to the speed of the link in general context (high or low) while rate quantitatively describes the link speed and associates it with the specific value in b/s.

5.2.2 LNS MLPPPoX

This functionality is supported through LNS on BB-ISA. LNS MLPPPoX can be used then as a workaround for PTA deployments, whereby LAC and LNS can be run back-to-back in the same system (connected by an external loop or a VSM2 module), and therefore locally terminate PPP sessions.

MLPPPoX can:

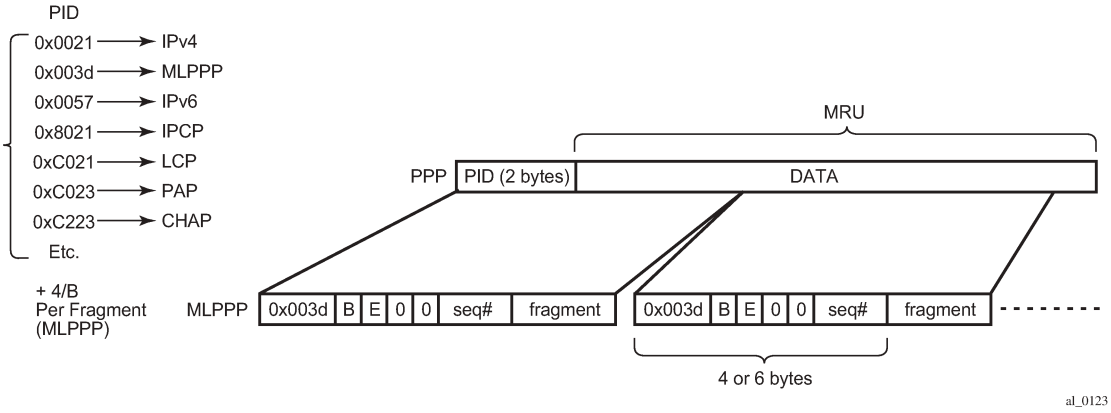
- Increase bandwidth in the last mile by bundling multiple links together.
- LFI/reassembly over a single MLPPPoX capable link (plain PPP does not support LFI).

5.2.3 MLPPP encapsulation

After the MLPPP bundle is created in the 7750 SR, traffic can be transmitted by using MLPPP encapsulation. However, MLPPP encapsulation is not mandatory over an MLPPP bundle.

MLPPP header is primarily required for sequencing the fragments. If a packet is not fragmented, it can be transmitted over the MLPPP bundle using either plain PPP encapsulation or MLPPP encapsulation. MLPPP encapsulation for fragmented traffic is shown in Figure 35: MLPPP encapsulation.

Figure 35: MLPPP encapsulation



5.2.4 MLPPPoX negotiation

MLPPPoX is negotiated during the LCP session negotiation phase by the presence of the Max-Received-Reconstructed Unit (MRRU) field in the LCP ConfReq. MRRU option is a mandatory field required in MLPPPoX negotiation. It represents the maximum number of octets in the Information field (Data part in Figure 35: MLPPP encapsulation) of a reassembled packet. The MRRU value negotiated in the LCP phase must be the same on all member links and it can be greater or lesser than the PPP negotiated MRU value of each member link. This means that the reassembled payload of the PPP packet can be greater than the transmission size limit imposed by individual member links within the MLPPPoX bundle. Packets are always be fragmented so that the fragments are within the MRU size of each member link.

Another field that could be optionally present in an MLPPPoX LCP Conf Req is an Endpoint Discriminator (ED). Along with the authentication information, this field can be used to associate the link with the bundle.

The last MLPPPoX negotiated option is the Short Sequence Number Header Format Option which allows the sequence numbers in MLPPPoX encapsulated frames/fragments to be 12-bit long (instead 24-bit long, by default).

After the multilink capability is successfully negotiated by LCP, PPP sessions can be bundled together over MLPPPoX capable links.

The basic operational principles are:

- LCP session is negotiated on each physical link with MLPPPoX capabilities between the two nodes.
- Based on the ED and the authentication outcome, a bundle is created. A subsequent IPCP negotiation is conveyed over this bundle. User traffic is sent over the bundle.
- If a new link tries to join the bundle by sending a new MLPPPoX LCP Conf Request, the LCP session is negotiated, authentication performed and the link is placed under the bundle containing the links with the same ED and authentication outcome.
- IPCP and IPv6CP is in the whole process negotiated only once over the bundle. This negotiation occurs at the beginning, when the first link is established and MLPPPoX bundle created. IPCP and IPc6CP messages are transmitted from the 7750 SR LNS without MLPPPoX encapsulation, while they can be received as MLPPPoX encapsulated or non-MLPPPoX encapsulated.

5.2.5 Enabling MLPPPoX

The lowest granularity at which MLPPPoX can be enabled is an L2TP tunnel. An MLPPPoX enabled tunnel is not limited to carrying only MLPPPoX sessions but can carry normal PPP(oE) sessions as well.

In addition to enabling MLPPPoX on the session terminating node LNS, MLPPPoX can also be enabled on the LAC by a PPP policy. The purpose of enabling MLPPPoX on the LAC is to negotiate MLPPPoX LCP parameters with the client. After the LAC receives the MRRU option from the client in the initial LCP ConfReq, it changes its tunnel selection algorithm so that all sessions of an MLPPPoX bundle are mapped into the same tunnel.

The LAC negotiates MLPPPoX LCP parameters regardless of the transport technology connected to it (ATM or Ethernet). LCP negotiated parameters are passed by the LAC to the LNS by Proxy LCP in a ICCN message. This way, the LNS has an option to accept the LCP parameters negotiated by the LAC or to reject them and restart the negotiation directly with the client.

The LAC transparently passes session traffic handed to it by the LNS in the downstream direction and the MLPPPoX client in the upstream direction. The LNS and the MLPPPoX client performs all data processing functions related to MLPPPoX such as fragmentation and interleaving.

After the LCP negotiation is completed and the LCP transition into an open state (configuration ACKs are sent and received), the Authentication phase on the LAC begins. During the Authentication phase the L2TP parameters become known (l2tp group, tunnel, and so on), the session is extended by the LAC to the LNS by L2TP. If the Authentication phase does not return L2TP parameters, the session is terminated because the 7750 SR does not support directly terminated MLPPPoX sessions.

In the case that MLPPPoX is not enabled on the LAC, the LAC negotiates plain PPP session with the client. If the client accepts plain PPP instead of MLPPPoX as offered by the LAC, when the session is extended to the LNS, the LNS re-negotiates MLPPPoX LCP with the client on a MLPPPoX enabled tunnel. The LNS learns about the MLPPPoX capability of the client by a Proxy LCP message in ICCN (first Conf Req received from the client is also send in a Proxy LCP). If the there is no indication of the MLPPPoX capability of the client, the LNS establishes a plain PPP(oE) session with the client.

Note: There is no dependency between ATM autosensing on LAC and MLPPPoX because autosensing operates on a lower layer than PPP (LCP).

5.2.6 Link Fragmentation and Interleaving

The purpose of Link Fragmentation and Interleaving (LFI) is to ensure that short high priority packets are not delayed by the transmission delay of large low priority packets on slow links.

For example it takes ~150ms to transmit a 5000B packet over a 256 kb/s link, while the same packet is transmitted in only 40us over a 1G link (~4000 times faster transmission). To avoid the delay of a high priority packet by waiting in the queue while the large packet is being transmitted, the large packet can be segmented into smaller chunks. The high priority packet can be then interleaved with the smaller fragments. This approach can significantly reduce the delay of high priority packets.

The interleaving functionality is only supported on MLPPPoX bundles with a single link. If more than one link is added into an interleaving capable MLPPPoX bundle, then interleaving is internally disabled and the tmnxMlpppBundleIndicatorsChange trap generated.

With interleaving enabled on an MLPPPoX enabled tunnel, the following session types are supported:

- Multiple LCP sessions tied into a single MLPPPoX bundle. This scenario assumes multiple physical links on the client side. Theoretically it would be possible to have multiple sessions running over the same physical link in the last mile. For example, two PPPoE sessions going over the same Ethernet

link in the last mile, or two ATM VCs o2q1231qa23n the same last mile link. Whichever the case may be, the LAC/LNS is unaware of the physical topology in the last mile (single or multiple physical links). Interleaving functionality is internally disabled on such MLPPPoX bundle.

- A single LCP session (including dual stack) over the MLPPPoX bundle. This scenario assumes a single physical link on the client side. Interleaving is supported on such single session MLPPPoX bundle as long as the conditions for interleaving are met. Those conditions are governed by max-fragment-delay parameter and calculation of the fragment size as described in subsequent sections.
- An LCP session (including dual stack) over a plain PPP/PPPoE session. This type of session is a regular PPP(oE) session outside of any MLPPPoX bundle and therefore its traffic is not MLPPPoX encapsulated.

Packets on an MLPPPoX bundle are MLPPPoX encapsulated unless they are classified as high priority packets when interleaving is enabled.

5.2.6.1 MLPPPoX fragmentation, MRRU and MRU considerations

MLPPPoX in the 7750 SR is concerned with two MTUs:

- **bundle-mtu** determines the maximum length of the original IP packet that can be transmitted over the entire bundle (collection of links) before any MLPPPoX processing takes place on the transmitting side. This is also the maximum size of the IP packet that the receiving node can accept after it de-encapsulates and assembles received MLPPPoX fragments of the same packet. Bundle-mtu is relevant in the context of the collection of links.
- **link-mtu** determines the maximum length of the payload before it is PPP encapsulated and transmitted over an individual link within the bundle. Link-mtu is relevant in the context of the single link within the bundle.

Assuming that the CPE advertised MRRU and MRU values are smaller than any configurable mtu on MLPPPoX processing modules in 7750 SR (carrier IOM and BB-ISA), the bundle-mtu and the link-mtu are based on the received MRRU and MRU values, respectively. For example, the bundle-mtu is set to the received MRRU value while link-bundle is set to the MRU value minus the MLPPPoX encapsulation overhead (4 or 6 bytes).

In addition to mtu values, fragmentation requires a fragment length value for each MLPPP bundle on LNS. This fragment length value is internally calculated according to the following parameters:

- Minimum transmission delay in the last mile.
- Fragment "payload to encapsulation overhead" efficiency ratio.
- Various MTU sizes in the 7750 SR dictated mainly by received MRU, received MRRU and configured PPP MTU under the following hierarchy:
  - configure subscriber-mgmt ppp-policy ppp-mtu (ignored on LNS)
  - configure service vprn l2tp group ppp mtu
  - configure service vprn l2tp group tunnel ppp mtu
  - configure router l2tp group ppp mtu
  - configure router l2tp group tunnel ppp mtu

The decision whether to fragment and encapsulate a packet in MLPPPoX depends on the mode of operation, the packet length and the packet priority as follows:

LFI

When Interleave is enabled in a bundle, low priority packets are always MLPPPoX encapsulated. If a low-priority packet's length exceeds the internally calculated Fragment Length, the packet is MLPPPoX fragmented and encapsulated. High priority packets whose length is smaller than the link-mtu is PPP encapsulated and transmitted without MLPPP encapsulation.

Non-LFI

When Interleave is disabled in a bundle, all packets are MLPPPoX encapsulated. If a packet's length exceeds the internally calculated fragment length, the packet is MLPPPoX fragmented and encapsulated.

A packet of the size greater than the link-mtu cannot be natively transmitted over an MLPPPoX bundle. This packet is MLPPPoX encapsulated and consequently fragmented. This is regardless of the priority of the packet in interleaving case or whether the fragmentation is enabled or disabled.

When MLPPPoX fragmentation is disabled with the **no max-fragment-delay** command, it is expected that packets are not MLPPPoX fragmented but rather only MLPPPoX encapsulated to be load balanced over multiple physical links in the last mile. However, even if MLPPPoX fragmentation is disabled, it is possible that fragmentation occurs under specific circumstances. This behavior is related to the calculation of the MTU values on an MLPPPoX bundle.

Consider an example where received MRRU value sent by CPE is 1500B while received MRU is 1492B. In this case, the bundle-mtu is set to 1500B and the link-mtu is set to 1488B (or 1486B) to allow for the additional 4/6B of MLPPPoX encapsulation overhead. Consequently, IP payload of 1500B can be transmitted over the bundle but only 1488B can be transmitted over any individual link. If an IP packet with the size between 1489B and 1500B needs to be transmitted from 7750 SR toward the CPE, this packet would be MLPPPoX fragmented in 7750 SR as dictated by the link-mtu. This is irrespective of whether MLPPPoX fragmentation is enabled or disabled (as set by **no max-fragment-delay flag**).

To entirely avoid MLPPPoX fragmentation in this case, the received MRRU sent by CPE should be lower than the received MRU for the length of the MLPPPoX header (4 or 6 bytes). In this case, for IP packets larger than 1488B, IP fragmentation would occur (assuming that DF flag in the IP header allows it) and MLPPPoX fragmentation would be avoided.

On the 7750 SR side, it is not possible to set different advertised MRRU and MRU values with the ppp-mtu command. Both MRRU and MRU advertised values adhere to the same configured ppp mtu value.

5.2.7 LFI functionality implemented in LNS

As mentioned in the previous section, LFI on LNS is implemented only on MLPPPoX bundles with a single LCP session.

There are two major tasks (Most of this is also applicable to non-lfi case. The only difference between lfi and non-lfi is that there is no artificial delay performed in non-lfi case) associated with LFI on the LNS:

- Executing subscriber QoS in the carrier IOM based on the last mile conditions. The subscriber QoS rates are the last mile on-the-wire rates. After traffic is QoS conditioned, it is sent to the BB-ISA for further processing.
- Fragmentation and artificial delay (queuing) of the fragments so that high priority packets can be injected in-between low priority fragments (interleaved). This operation is performed by the BB-ISA.

Examine an example to further clarify functionality of LFI. The parameters, conditions and requirements that are used in the example to describe the wanted behavior are the following:

- High priority packets must not be delayed for more than 50ms in the last mile because of the transmission delay of the large low priority packets. Considering that tolerated end-to-end VoIP delay

must be under 150ms, limiting the transmission delay to 50ms on the last mile link is a reasonable option.

- The link between the LNS and LAC is 1Gb/s Ethernet.
- The last mile link rate is 256 kb/s.
- Three packets arrive back-to-back on the network side of the LNS (in the downstream direction). The large 5000B low priority packet P1 arrives first, followed by two smaller high priority packets P2 and P3, each 100B.

Note:
Packets P1, P2 and P3 can be originated by independent sources (PCs, servers, and so on) and therefore can theoretically arrive in the LNS from the network side back-to-back at the full network link rate (10Gb/s or 100Gb/s).

- The transmission time on the internal 10G link between the BB-ISA and the carrier IOM for the large packet (5000B) is 4us while the transmission time for the small packet (100B) is 80ns.
- The transmission time on the 1G link (LNS->LAC) for the large packet (5000B) is 40us while the transmission time for the small packet (100B) is 0.8us.
- The transmission time in the last mile (256 kb/s) for the large packet is ~150ms while the transmission time for the small packet on the same link is ~3ms.
- Last mile transport is ATM.

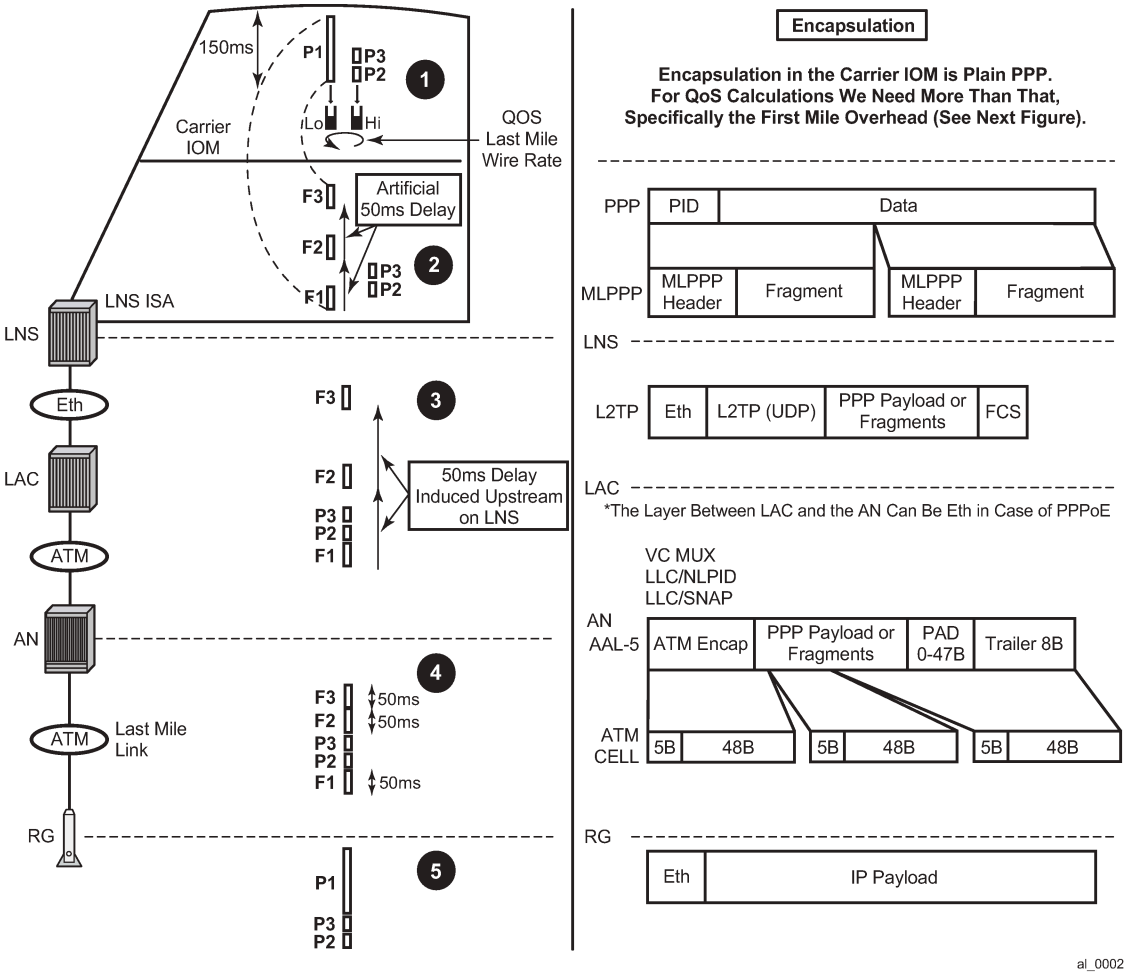
To satisfy the delay requirement for the high priority packets, the large packets are fragmented into three smaller fragments. The fragments are carefully sized so that their individual transmission time in the last mile does not exceed 50ms. After the first 50ms interval, there is an opportunity to interleave the two smaller high priority packets.

This entire process is further clarified by the five points (1-5) in the packet route from the LNS to the Residential Gateway (RG) as depicted in Figure 36: Packet route from the LNS to the RG.

The five points are:

1. Last mile QoS awareness in the LNS
2. BB-ISA processing
3. LNS-LAC link
4. AN-RG link
5. Home link

Figure 36: Packet route from the LNS to the RG



5.2.7.1 Last mile QoS awareness in the LNS

By implementing MLPPPoX in LNS, the traffic treatment functions (QoS/LFI) of the last mile to the node (LNS) that is multiple hops away is transferred.

The success of this operation depends on the accuracy at which the last mile conditions in the LNS can be simulated. The assumption is that the LNS is aware of the two most important parameters of the last mile:

- The last mile encapsulation — This is needed for the accurate calculation of the overhead associated of the transport medium in the last mile for traffic shaping and interleaving.
- The last mile link rate — This is crucial for the creation of artificial congestion and packet delay in the LNS.

The subscriber QoS in the LNS is implemented in the carrier IOM and is performed on a per packets basis before the packet is handed over to the BB-ISA. Per packet, instead of per fragment QoS processing ensures a more efficient utilization of network resources in the downstream direction. Discarding fragments

in the LNS would have detrimental effects in the RG as the RG would be unable to reconstruct a packet without all of its fragments.

High priority traffic within the bundle is classified into the high priority queue. This type of traffic is not MLPPPoX encapsulated unless its packet size exceeds the link MTU as described in MLPPPoX fragmentation, MRRU and MRU considerations. Low priority traffic is classified into a low priority queue and is always MLPPPoX encapsulated. If the high priority traffic becomes MLPPPoX encapsulated or fragmented, the MLPPPoX processing module (BB-ISA) considers it as low-priority. The assumption is that the high priority traffic is small in size and consequently MLPPPoX encapsulation or fragmentation and degradation in priority can be avoided. The aggregate rate of the MLPPPoX bundle is on-the-wire rate of the last mile as shown in Figure 37: Last mile encapsulation.

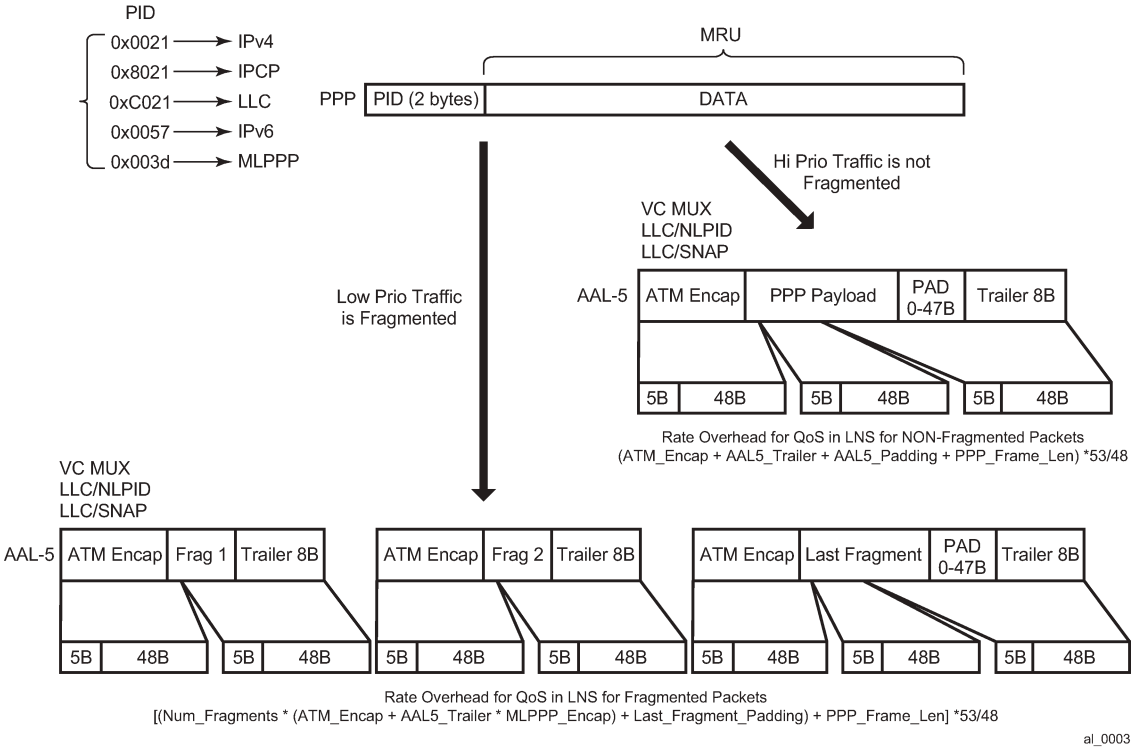
ATM on-the-wire overhead for non-MLPPPoX encapsulated high priority traffic includes:

- ATM encapsulation (VC-MUX, LLC/NLPID, LLC/SNAP).
- AAL5 trailer (8B).
- AAL5 padding to 48B cell boundary (this makes the overhead dependent on the packet size).
- Multiplication by 53/48 to account for the ATM cell headers.

For low priority traffic, which is always MLPPPoX encapsulated, an additional overhead related to MLPPPoX encapsulation and possibly fragmentation must be added. In other words, each fragment carries ATM+MLPPPoX overhead.

Note: Avoid the 48B boundary padding for all fragments except the last one. This can be done by choosing the fragment length so that it is aligned on the 48B boundary (rounded down if based on max-fragment-delay or rounded up if based on the encapsulation/utilization.

Figure 37: Last mile encapsulation



For Ethernet in the last mile, the implementation always assures that the fragment size plus the encapsulation overhead is always larger or equal to the minimum Ethernet packet length (64B).

5.2.7.2 BB-ISA processing

MLPPPoX encapsulation, fragmentation and interleaving are performed by the LNS in BB-ISA. According to the example, a large low priority packet (P1) is received by the BB-ISA, immediately followed by the two small high priority packets (P2 and P3). Because the requirement stipulates that there is no more than 50ms of transmission delay in the last mile (including on-the-wire overhead), the large packet must be fragmented into three smaller fragments each of which do not cause more than 50ms of transmission delay.

The BB-ISA would normally send packets or fragments to the carrier IOM at the rate of 10Gb/s. In other words, by default the three fragments of the low priority packet would be sent out of the BB-ISA back-to-back at the very high rate before the high priority packets even arrive in the BB-ISA. To interleave, the BB-ISA must simulate the last mile conditions by delaying the transmission of the fragments. The fragments are be paced out of the BB-ISA (and out of the box) at the rate of the last mile. High priority packets can be injected in front of the fragments while the fragments are being delayed.

In Figure 36: Packet route from the LNS to the RG (point 2) the first fragment F1 is sent out immediately (transmission delay at 10G is in the 1us range). The transmission of the next fragment F2 is delayed by 50ms. While the transmission of the second fragment F2 is being delayed, the two high priority packets (P1 and P2 in red) are received by the BB-ISA and are immediately transmitted ahead of fragments F2 and F3. This approach relies on the imperfection of the IOM shaper which is releasing traffic in bursts (P2 and P3 right after P1). The burst size is dependent on the depth of the rate token bucket associated with the IOM shaper.

Note: By the time the second fragment F2 is transmitted, the first fragment F1 has traveled a long way (50ms) on high rate links toward the Access Node (assuming that there is no queuing delay along the way), and its transmission on the last mile link has already begun (if not already completed).

This is not applicable for this discussion, but worth noticing is that the LNS BB-ISA also adds the L2TP encapsulation to each packet or fragment. The L2TP encapsulation is removed in the LAC before the packet or fragment is transmitted toward the AN.

5.2.7.3 LNS-LAC link

This is the high rate link (1Gb/s) on which the first fragment F1 and the two consecutive high priority packets, P2 and P3, are sent back-to-back by the BB-ISA.

(BB-ISA->carrier IOM->egress IOM-> out-of-the-LNS).

The remaining fragments (F2 and F3) are still waiting in the BB-ISA to be transmitted. They are artificially delayed by 50ms each.

Additional QoS based on the L2TP header can be performed on the egress port in the LNS toward the LAC. This QoS is based on the classification fields inside of the packet or fragment headers (DSCP, dot1.p, EXP).

Note: The LAC-AN link is not really relevant for the operation of LFI on the LNS. This link can be either Ethernet (in case of PPPoE) or ATM (PPPoE or PPP). The rate of the link between the LAC and the AN is still considered a high speed link compared to the slow last mile link.

5.2.7.4 AN-RG link

Finally, this is the slow link of the last mile, the reason why LFI is performed in the first place. Assuming that LFI played its role in the network as designed, by the time the transmission of one fragment on this link is completed, the next fragment arrives just in time for unblocked transmission. In between the two fragments, there can be one or more small high priority packets waiting in the queue for the transmission to complete.

Note:

- On the AN-RG link in Figure 36: Packet route from the LNS to the RG that packets P2 and P3 are ahead of fragments F2 and F3. Therefore the delay incurred on this link by the low priority packets is never greater than the transmission delay of the first fragment (50ms). The remaining two fragments, F2 and F3, can be queued and further delayed by the transmission time of packets P2 and P3 (which is normally small, in the example, 3ms for each).
- If many low priority packets are waiting in the queue, then they would have caused delay and would have further delayed the fragments that are in transit from the LNS to the LAC. This condition is normally caused by bursts and it should clear itself out over time.

5.2.7.5 Home link

High priority packets P2 and P3 are transmitted by the RG into the home network ahead of the packet P1 although the fragment F1 has arrived in the RG first. The reason for this is that the RG must wait for the fragments F2 and F3 before it can re-assemble packet P1.

5.2.7.6 Optimum fragment size calculation by LNS

Fragmentation in LFI is based on the optimal fragment size. LNS implementation calculates the two optimal fragment sizes, based on two different criteria:

- Optimal fragment size based on the payload efficiency of the fragment considering the fragmentation and transportation header overhead associated with the fragment encapsulation based fragment size.
- Optimal fragment size based on the maximum transmission delay of the fragment set by configuration delay based fragment size.

At the end, only one optimal fragment size is selected. The actual fragment's length is the optimal fragment size.

- The parameters required to calculate the optimal fragment sizes are known to the LNS either through configuration or signaling. These, in-advance known parameters are:
- Last mile maximum transmission delay (max-fragment-delay obtained by CLI)
- Last mile ATM Encapsulation (in the example the last mile is ATM but in general it can be Ethernet for MLPPPoE)
- MLPPP encapsulation length (depending on the fragment sequence number format)
- The last mile on-the-wire rate for the MLPPPoX bundle

Examine closer each of the two optimal fragment sizes.

5.2.7.6.1 Encapsulation-based fragment size

Be mindful that fragmentation may cause low link utilization. In other words, during fragmentation a node may end up transporting mainly overhead bytes in the fragment as opposed to payload bytes. This would only intensify the problem that fragmentation is intended to solve, especially on an ATM access link that tend to carry larger encapsulation overhead.

To reduce the overhead associated with fragmentation, the following is enforced in the 7750 SR:

The minimum fragment payload size is at least 10times greater than the overhead (MLPPP header, ATM Encapsulation and AAL5 trailer) associated with the fragment.

The optimal fragment length (including the MLPPP header, the ATM Encapsulation and the AAL5 trailer) is a multiple of 48B. Otherwise, the AAL5 layer would add an additional 48B boundary padding to each fragment which would unnecessarily expand the overhead associated with fragmentation. By aligning all-but-last fragments to a 48B boundary, only the last fragment potentially contains the AAL5 48B boundary padding which is no different from a non-fragmented packet. All fragments, except for the last fragment, are referred to as non-padded fragments. The last fragment is padded if it is not already natively aligned to a 48B boundary.

As an example, calculate the optimal fragment size based on the encapsulation criteria with the maximum fragment overhead of 22B. To achieve >10x transmission efficiency the fragment payload size must be 220B (10*22B). To avoid the AAL5 padding, the entire fragment (overhead + payload) is rounded UP on a 48B boundary. The final fragment size is 288B [22B + 22B*10 + 48B_allignment].

In conclusion, an optimal fragment size was selected that carries the payload with at least 90% efficiency. The last fragment of the packet cannot be artificially aligned on a 48B boundary (it is a natural reminder), so it is be padded by the AAL5 layer. Therefore, the efficiency of the last fragment is less than 90% in the example. In the extreme case, the efficiency of this last fragment may be only 2%.

Note: The fragment size chosen in this manner is purely chosen based on the overhead length. The maximum transmission delay did not play any role in the calculations.

For the Ethernet-based last mile, the CPM always makes sure that the fragment size plus encapsulation overhead is larger or equal to the minimum Ethernet packet length of 64B.

5.2.7.6.2 Fragment size based on the maximum transmission delay

The first criterion in selecting the optimal fragment size based on the maximum transmission delay mandates that the transmission time for the fragment, including all overheads (MLPPP header, ATM encapsulation header, AAL5 overhead and ATM cell overhead) must be less than the configured max-fragment-delay time.

The second criterion mandates that each fragment, including the MLPPP header, the ATM encapsulation header, the AAL5 trailer and the ATM cellification overhead be a multiple of 48B. The fragment size is rounded down to the nearest 48B boundary during the calculations to minimize the transmission delay. Aligning the fragment on the 48B boundary eliminates the AAL5 padding and therefore reduces the overhead associated with the fragment. The overhead reduction improves the transmission time and also increases the efficiency of the fragment.

These two criteria along with the configuration parameters (ATM Encapsulation, MLPPP header length, max-fragment-delay time, rate in the last mile), the implementation calculates the optimal non-padded fragment length as well as the transmission time for this optimal fragment length.

5.2.7.6.3 Selection of the optimum fragment length

So far, the implementation has calculated the two optimum fragment lengths, one based on the length of the MLPPP/transport encapsulation overhead of the fragment, the other one based on the maximum transmission delay of the fragment. Both of them are aligned on a 48B boundary. The larger of the two is chosen and the BB-ISA performs LFI based on this selected optimal fragment length.

5.2.8 Upstream traffic considerations

Fragmentation and interleaving is implemented on the originating end of the traffic. In other words, in the upstream direction the CPE (or RG) is fragmenting and interleaving traffic. There is no interleaving or fragmentation processing in the upstream direction in the 7750 SR. The 7750 SR are on the receiving end and is only concerned with the reassembly of the fragments arriving from the CPE. Fragments are buffered until the packet can be reconstructed. If all fragments of a packet are not received within a preconfigured time frame, the received fragments of the partial packet are discarded (a packet cannot be reconstructed without all of its fragments). This time-out and discard is necessary to prevent buffer starvation in the BB-ISA. Two values for the time-out can be configured: 100ms and 1s.

5.2.9 Multiple links MLPPPoX with no interleaving

Interleaving over MLPPPoX bundles with multiple links are not supported. However, fragmentation is supported.

To preserve packet order, all packets on an MLPPPoX bundle with multiple links are MLPPPoX encapsulated (monotonically increased sequence numbers).

Multiclass MLPPP (RFC 2686, *The Multi-Class Extension to Multi-Link PPP*) is not supported.

5.2.10 MLPPPoX session support

The following session types in the last mile are supported:

- **MLPPPoE**

  Single physical link or multilink. The last mile encapsulation is Ethernet over copper (This could be Ethernet over VDSL or HSDSL). The access rates (especially upstream) are still limited by the xDSL distance limitation and therefore interleaving is required on a slow speed single link in the last mile. It is possible that the last mile encapsulation is Ethernet over fiber (FTTH) but in this case, users would not be concerned with the link speed to the point where interleaving and link aggregation is required.

Finally, this is the slow link of the last mile, the reason why LFI is performed in the first place. Assuming that LFI played its role in the network as designed, by the time the transmission of one fragment on this link is completed, the next fragment arrives just in time for unblocked transmission. In between the two fragments are one or more small high priority packets waiting in the queue for the transmission to complete.

As shown in Figure 38: MLPPPoE — multiple physical links, the AN-RG link in that packets P2 and P3 are ahead of fragments F2 and F3. Therefore the delay incurred on this link by the low priority packets is never greater than the transmission delay of the first fragment (50ms). The remaining two fragments, F2 and F3, can be queued and further delayed by the transmission time of packets P2 and P3 (which is normally small, in the example 3ms for each).

Note: If many low priority packets were waiting in the queue, then they would have caused delay for each other and would have further delayed the fragments in transit from the LNS to the LAC. This condition is normally caused by bursts and it should clear itself out over time.

Figure 38: MLPPPoE — multiple physical links

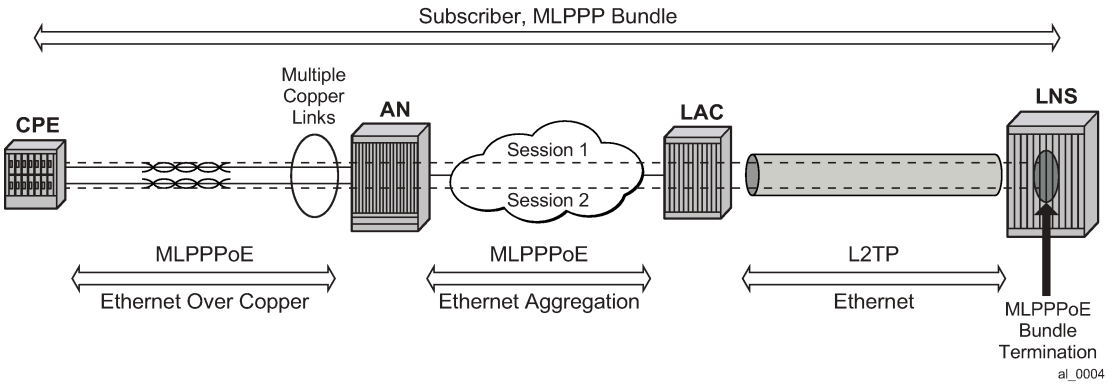
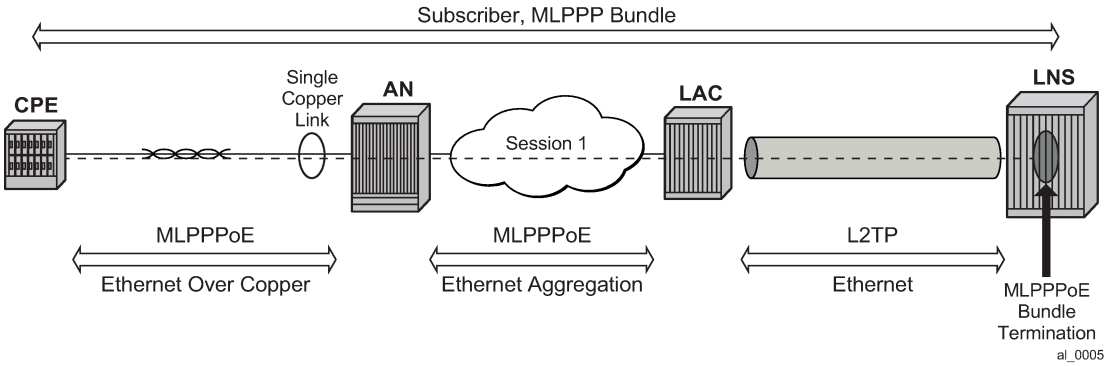


Figure 39: MLPPPoE — single physical link



- MLPPP(oEo)A — A single physical link or multilink. The last mile encapsulation is ATM over xDSL.

Figure 40: MLPPP(oE)oA — multiple physical links

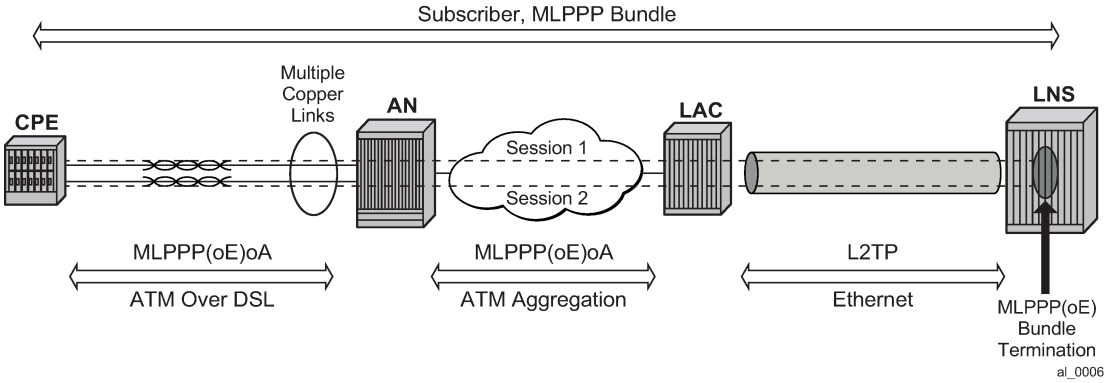
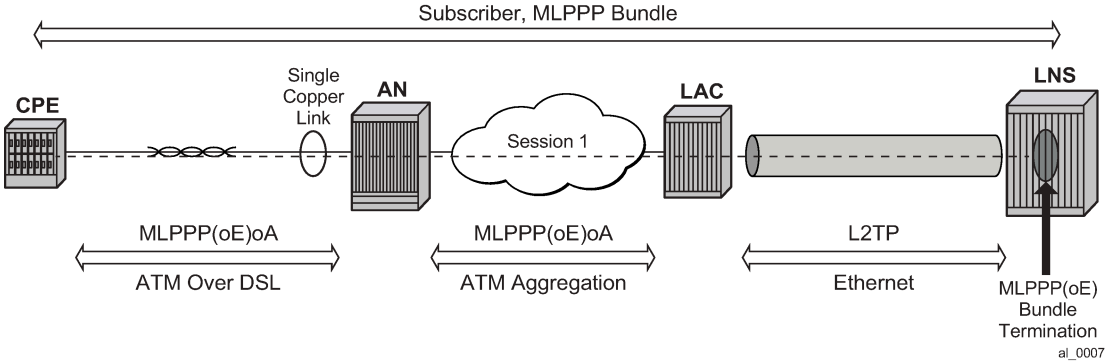


Figure 41: MLPPP(oE)oA — single physical link



Some other combinations are also possible (ATM in the last mile, Ethernet in the aggregation) but they all come down to one of the above models that are characterized by:

- Ethernet or ATM in the last mile.
- Ethernet or ATM access on the LAC.
- MLPPP/PPPoE termination on the LNS.

5.2.11 Session load balancing across multiple BB-ISAs

PPP/PPPoE sessions are by default load balanced across multiple BB-ISAs (max 6) in the same group. The load balancing algorithm considers the number of active session on each BB-ISA in the same group. The load balancing algorithm does not consider the number of queues consumed on the carrier IOM. Therefore, a session can be refused if queues are depleted on the carrier IOM even though the BB-ISA may be lightly loaded in terms of the number of sessions that is hosting.

With MLPPPoX, it is important that multiple sessions per bundle be terminated on the same LNS BB-ISA. This can be achieved by per tunnel load balancing mode where all sessions of a tunnel are terminated in the same BB-ISA. Per tunnel load balancing mode is mandatory on LNS BB-ISAs that are in the group that supports MLPPPoX.

On the LAC side, all sessions in an MLPPPoX bundle are automatically assigned to the same tunnel. In other words an MLPPPoX bundle is assigned to the tunnel. There can be multiple tunnels created between the same pair of LAC/LNS nodes.

5.2.12 BB-ISA hashing considerations

All downstream traffic on an MLPPPoX bundle with multiple links is always MLPPPoX encapsulated. Some traffic is fragmented and served in a octet oriented round robin fashion over multiple member links. However, fragments are never delayed when the bundle contains multiple links.

In a per fragment/packet load sharing algorithm, there is always the possibility that there is uneven load utilization between the member links. A single link overload can go unnoticed in the network all the way to the Access Node. The access node is the only node in the network that actually has multiple physical links connected to it. All other session-aware nodes (LAC and LNS) only see MLPPPoX as a bundle with multiple sessions without any mechanism to shape traffic per physical link. Other nodes in this case being 7750 SRs. Other vendors may have the ability to condition (shape) traffic per session.

If one of the member sessions is perpetually overloaded by the LNS, traffic is dropped in the last mile because the corresponding physical link cannot absorb traffic beyond its physical capabilities. This would have detrimental effects on the whole operation of the MLPPPoX bundle. To prevent this perpetual overloading of the member links that can be caused by per packet/fragment load balancing scheme, the load balancing scheme that considers the number of octets transmitted over each member link. The octet counter of a new link is initialized to the lowest value of any existing link counter. Otherwise the load balancing mechanism would show significant bias toward the new link until the byte counter catches up with the rest of the links.

5.2.13 Last mile rate and encapsulation parameters

The last mile rate information along with the encapsulation information is used for fragmentation (to determine the maximum fragment length) and interleaving (delaying fragments in the BB-ISA). In addition, the aggregate subscriber rate (aggregate-rate-limit) on the LNS is automatically adjusted based on the last mile link rate and the number of links in the MLPPPoX bundle.

Downstream Data Rate in the Last Mile

The subscriber aggregate rates (**agg-rate-limit**) used in (H)QoS on the carrier IOM and in the BB-ISA (for interleaving) must be wire based in the last mile. This rule applies equally to both, the LAC and LNS.

The last mile on-the-wire rates of the subscriber can be submitted to the LAC and the LNS by various means. The following discusses the break down on how the last mile wire rates are passed to each entity:

LAC

The last mile link rate is taken by the following methods in the order of listed priority:

- LUDB — rate-down command under the host hierarchy in LUDB.
- RADIUS Alc-Access-Loop-Rate-Down VSA. Although this VSA is stored in the state of plain PPP(oE) sessions (MLPPPoX bundled or not), it is applicable only to MLPPPoX bundles.
- PPPoE tags — Vendor Specific Tags (RFC 2516, A Method for Transmitting PPP Over Ethernet (PPPoE); tag type 0x0105; tag value is Enterprise Number 3561 followed by the TLV sub-options as specified in TR-101 -> Actual Data Rate Downstream 0x82)

As long as the link rate information is available in the LAC, it is always passed to the LNS in the ICRQ message using the standard L2TP encoding. This cannot be disabled.

In addition, an option is available to control the source of the rate information can be conveyed to the LNS by TX Connect Speed AVP in the ICCN message. This can be used for compatibility reasons with other vendors that can only use TX Connect Speed to pass the link rate information to the LNS. By default, the maximum port speed (or the sum of the maximum speeds of all member ports in the LAG) is reported in TX Connect Speed. Unlike the rate conveyed in ICRQ message, The TX Connect Speed content is configurable with the following command:

```
config>subscr-mgmt
    sla-profile <name>
        egress
            report-rate agg-rate-limit | scheduler <scheduler-name> | pppoe-actual-
rate | rfc5515-actual-rate
```

The report-rate configuration option dictates which rate is reported in the TX Connect Speed as follows:

- **agg-rate-limit**

  Statically configured **agg-rate-limit** value or RADIUS QoS override is reported

- **scheduler** *scheduler-name*

  Virtual schedulers are not supported in MLPPPoX

- **pppoe-actual-rate**

  The rate taken from PPPoE tags is reported.

  Note: The rate reported according to RFC 5515 can still be different if the source for both methods is not the same.

- **rfc5515-actual-speed**

  The rate is taken from RFC5515.

The RFC 5515 relies on the same encoding as PPPoE tags (vendor ID is ADSL Forum and the type for Actual Data Rate Downstream is 0x82).

Note: The two methods of passing the line rate to the LNS are using different message types (ICRQ and ICCN).

The LAC on the 7750 SR is not aware of MLPPPoX bundles. As such, the aggregate subscriber bandwidth on the LAC is configured statically by usual means (sub-profile, scheduler-policy) or dynamically modified through RADIUS. The aggregate subscriber (or MLPPPoX bundle) bandwidth on the LAC is not automatically adjusted according to the rates of the individual links in the bundle and the number of the links in the bundle. As such, an operator must ensure that the statically provided rate value for aggregate-rate-limit is the sum of the bandwidth of each member link in the MLPPPoX bundle. The number of member links and their bandwidth must be therefore known in advance. The alternative is to have the aggregate rate of the MLPPPoX bundle set to a high value and rely on the QoS treatment performed on the LNS.

LNS

The sources of information for the last mile link rate on the LNS is taken in the following order:

- LUDB (during user authentication phase, same as in LAC)
- RADIUS (same as in LAC)
- ICRQ message — Actual Data Downstream Rate (RFC 5515)
- ICCN message — TX Connect Speed

There is no configuration option to determine the priority of the source of information for the last mile link rate. TX Connect Speed in ICCN message is only be taken into consideration as a last resort in absence of any other source of last mile rate information.

After the last mile rate information is obtained, the subscriber aggregate rate **aggregate-rate-limit** is automatically adjusted to the minimum value of:

- The smallest link speed in the MLPPPoX bundle multiplied by the number of links in the bundle.
- Statically configured aggregate-rate-limit

The link speed of each link in the bundle must be the same, meaning, different link speeds within the bundle are not supported. When different link are received, speed values for last mile links within the bundle, the minimum received speed is adopted and apply it to all links.

When the obtained rate information from the last mile for a session within the MLPPP bundle is out of bounds (1 kb/s to 100 Mb/s), the session within the bundle is terminated.

Encapsulation

Wire-rates are dependent on the encapsulation of the link to which they apply. The last mile encapsulation information can be extracted by various means.

LAC

- Static configuration by LUDB.
- RADIUS — Alc-Access_Loop-Encap-Offset VSA.
- PPPoE tags — Vendor Specific Tags (RFC 2516; tag type 0x0105; tag value is Enterprise Number 3561 followed by the TLV sub-options as specified in TR-101 -> Actual Data Rate Downstream 0x82).

The LAC passes the line encapsulation information to the LNS by an ICRQ message using the encoding defined in the RFC 5515.

LNS

The LNS extracts the encapsulation information in the following order:

- Static configuration by LUDB.
- RADIUS — Alc-Access-Loop-Encap-Offset VSA.
- ICRQ message (RFC 5515)

When the encapsulation information is not provided by any of the existing means (LUDB, RADIUS, AVP signaling, PPPoE Tags), then by default **pppoa-null** encapsulation is in effect. This applies to LAC and LNS.

5.2.14 Link failure detection

The link failure in the last mile is detected by the expiration of session keepalives (LCP). The LNS tears down the session over the failed link and notify the LAC by a CDN message.

5.2.15 CoA support

CoA request for the subscriber aggregate-rate-limit change is honored on the LAC and the LNS.

CoA for the rate change of an individual link within the bundle is supported through the same VSA that can be used to initially assign the rate parameter to each member link. This is supported only on LNS. The rate override with CoA is applied to all active link members within the bundle.

Change of the access link parameters with CoA is supported in the following fashion:

- Change of access loop encap: refused (NAK)
- Change of access loop rate down:
- On L2TP LAC session: refused (NAK). On LAC the access loop rate down is not locally used for any rate limiting function but instead it is just passed to the LNS at the beginning when the session is first established. Mid-session changes on LAC by CoA are not propagated to the LNS.
- On the L2TP LNS session, the plain session is ignored. The rate is stored in the MIB table but no rate limiting action is taken. In other words, this parameter is internally excluded from rate calculations and advertisements. However, it is shown in the output of the relevant show commands.
- Bundle session: applied on all link sessions. The aggregate rate limit of the bundle is set to the minimum of the:
- CoA obtained local loop down rate multiplied by the number of links in the bundle
- The aggregate rate limit configured statically or obtained by CoA.
- Fragment length is affected by this change. In case that interleaving is enabled on a single link bundle, the interleave interval is affected.
- Non-L2TP: ignored. The rate is stored in the MIB table but no rate limiting action is taken. In other words, this parameter is internally excluded from rate calculations and advertisements. However, it is shown in the output of the relevant show commands.

Similar behavior is exhibited if at mid-session, the parameters are changed through LUDB with the exception of the rate-down parameter in LAC. If this parameter is changed on the LAC, all sessions are disconnected.

5.2.16 Accounting

Accounting counters on the LNS include all packet overhead (wire overhead from the last mile). There is only one accounting session per bundle.

On the LAC, there is one accounting session per pppoe session (link).

In tunnel-accounting mode there is one accounting session per link.

On LNS only the stop-link of the last link of the bundle carries all accounting data for the bundle.

5.2.17 Filters and mirroring

Filters and mirrors (LI) are not supported on an MLPPPoX bundle on LAC. However, filters and ip-only mirror type are supported on the LNS.

5.2.18 PTA considerations

Locally terminated MLPPPoX (PTA) solution is offered based on the LAC and the LNS hosted in the same system. An external loop (or VSM2) is used to connect the LAC to the LNS within the same box. The subscribers are terminated on the LNS.

5.2.19 QoS considerations

5.2.19.1 Dual-pass

HQoS and LFI are performed in two stages that involve double traversal (dual-pass) of traffic through the carrier IOM and the BB-ISA. The following are the functions performed in each pass:

- In the first pass through the carrier IOM, traffic is marked (dot1p bits) as high or low priority. This plays a crucial role in the execution of LFI in the BB-ISA.
- In the first pass through the BB-ISA this prioritization from the first step, is an indication (along with the internally calculated fragment size) of whether the traffic is interleaved (non MLPPP encapsulated) or not (MLPPP encapsulated). Consequently, the BB-ISA adds the necessary padding related to last mile wire overhead to each packet. This padding is in the second pass on the carrier IOM and performs last mile wire based QoS functions.
- In the second pass through the carrier IOM, the last-mile wire-based HQoS is performed based on the padding added in the first pass through the BB-ISA.
- In the second pass through the BB-ISA, the previously added overhead is stripped off and LFI/MLPPP encapsulation functions are performed.

5.2.19.2 Traffic prioritization in LFI

The delivery of high priority traffic within predefined delay bounds on a slow speed last mile link is ensured by the correct QoS classification and prioritization. High priority traffic is interleaved with low priority fragments on a single link MLPPPoX bundle with LFI enabled. The classification of traffic into the correct (high or low priority) forwarding class is performed on the downstream ingress interface. However, traffic can be re-classified (re-mapped into another forwarding class) on the egress access interface of the carrier IOM, just before packets are transmitted to the BB-ISA for MLPPPoX processing. This can be achieved with a QoS **sap-egress** policy referenced in the LNS sla-profile.

The priority of the forwarding class in regular QoS (on IOM) is determined by the properties of the queue to which the forwarding class is mapped. Expedited, non-expedited queue type, CIR and PIR rates. In contracts, traffic prioritization in LFI domain (in BB-ISA) is determined by the outer dot1p bits that are set by the carrier IOM while transmitting packets toward the BB-ISA. The outer dot1p bits are marked based on the forwarding class information determined by classification/re-classification on ingress/carrier IOM. This marking of outer dot1p bits in the Ethernet header between the carrier IOM and the BB-ISA is fixed and defined in the default sap-egress LNS ESM policy 65537. The marking definition is as follows:

```
FC be -> dot1p 0
FC l2 -> dot1p 1
FC af -> dot1p 2
FC l1 -> dot1p 3
FC h2 -> dot1p 4
FC ef -> dot1p 5
FC h1 -> dot1p 6
```

```
FC nc -> dot1p 7
```

In LFI (on BB-ISA), dot1p bits [0,1,2 and 3] are considered low priority while dot1p bits (4,5,6 and 7) are considered high priority. Consequently, forwarding classes BE, L2, AF and L1 are considered low priority while forwarding classes H2, EF, H1 and NC are considered high priority. High priority traffic is interleaved with low priority traffic. Assuming that the packet size does not exceed maximum fragment size.

The following describes the reference points in traffic prioritization for the purpose of LFI in the 7750 SR:

- Classification on downstream ingress interface (entrance point into the 7750 SR) - packets can be classified into one of the following eight forwarding classes: be, l2, af, l1, h2, ef, h1, and nc. Depending on the type of the ingress interface (access or network), traffic can be classified based on dot1p, exp, DSCP, ToS bits or ip-match criteria (dscp, dst-ip, dst-port, fragment, src-ip, src-port and protocol-id).
- Re-classification on downstream access egress interface between the carrier IOM and the BB-ISA - in the carrier IOM, downstream traffic can be re-classified into another forwarding class, just before it is forwarded to the BB-ISA. Re-classification on access egress is based on the same fields as on ingress except for the dot1p and exp bits because Ethernet or MPLS headers from ingress are not carried from ingress to egress.
- Marking on downstream access egress interface between the carrier IOM and the BB-ISA, after the forwarding class is available on the carrier IOM in the egress direction (toward BB-ISA), it is used to mark outer dot1p bits in the new Ethernet header that are used to transport the frame from the carrier IOM to the BB-ISA. The marking of the dot1p bits on the egress SAP between the carrier IOM and the BB-ISA cannot be changed for MLPPPoX even if the **no qos-marking-from-sap** command is configured under the **sla-profile** on egress.

5.2.19.3 Shaping based on the last mile wire rates

Accurate QoS require that the subscriber rates in the first mile on an MLPPPoX bundle be properly represented in the LNS. In other words, the rate limiting functions in the LNS must account for the last mile on-the-wire encapsulation overhead. The last mile encapsulation can be Ethernet or ATM.

For ATM in the last mile, the LNS accounts for the following per fragment overhead:

- PID
- MLPPP encapsulation header
- ATM Fixed overhead (ATM encap + fixed AAL5 trailer)
- 48B boundary padding as part of AAL5 trailer
- 5B per each 48B of data in ATM cell

In case of Ethernet encapsulation in the last mile, the overhead is:

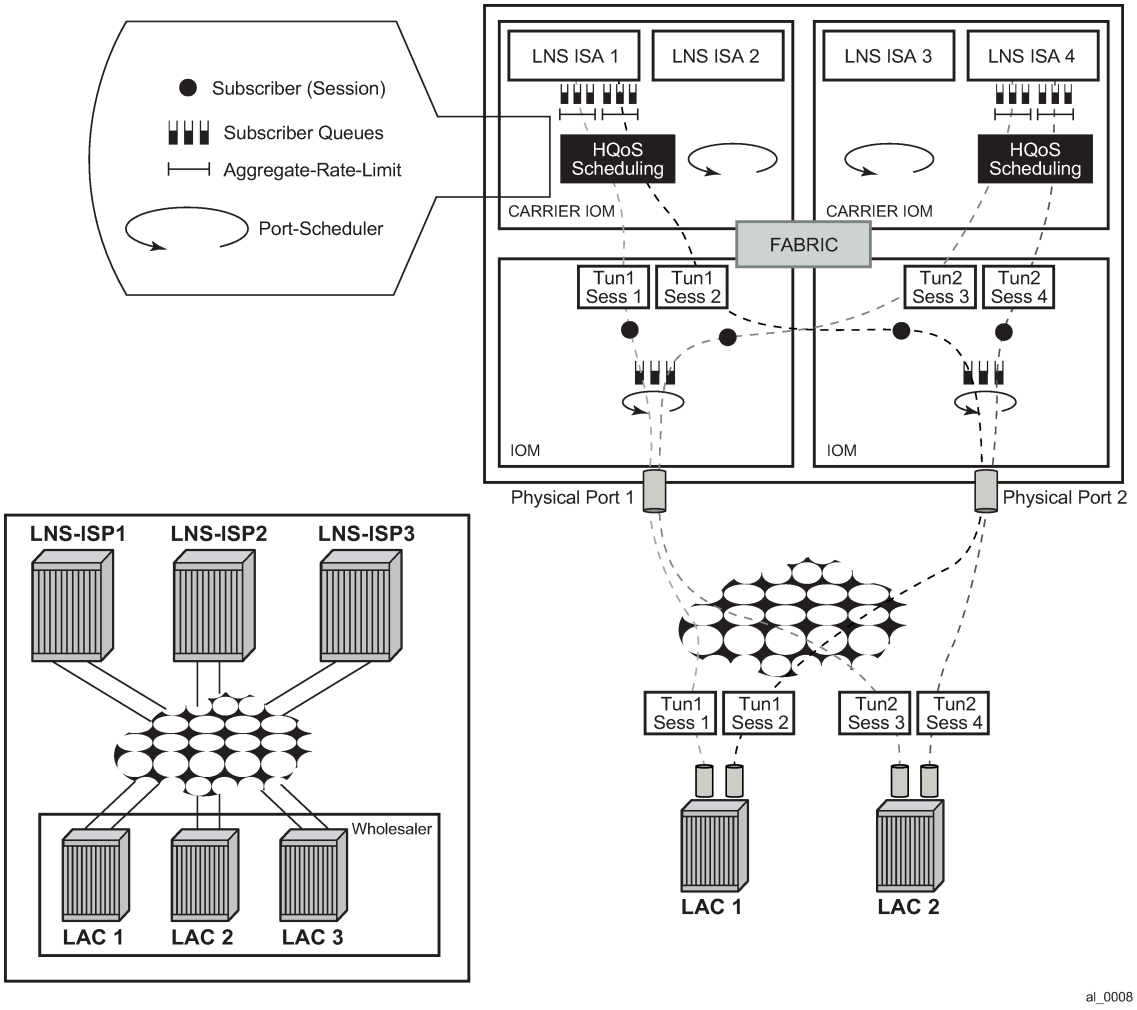
- PID
- MLPPP header per fragment
- Ethernet Header + FCS per fragment
- Preamble + IPG overhead per fragment

The **encap-offset** command in the **sub-profile**>**egress** CLI context is ignored in case of MLPPPoX. MLPPPoX rate calculation is, by default, always based on the last-mile wire overhead.

The HQoS rates (port-scheduler, aggregate-rate-limit, and scheduler) on LNS are based on the wire overhead of the entity to which the HQoS is applied. For example, if the port-scheduler is managing

bandwidth on the link between the BB-ISA and the carrier IOM, then the rate of such scheduler accounts for the q-in-q Ethernet encapsulation on that link along with the preamble and inter packet gap (20B).

Figure 42: QoS enforcement points in the LNS



While virtual schedulers (attached by sub-profile) are supported on LNS for plain PPPoE sessions, they are not supported for MLPPPoX bundles. Only aggregate- rate-limit along with the port-scheduler can be used in MLPPPoX deployments.

5.2.19.4 Downstream bandwidth management on egress port

Bandwidth management on the egress physical ports (Physical Port 1 and Physical Port 2) is performed at the egress port on the egress IOM instead on the carrier IOM. By default, the forwarding class (FC) information is preserved from network ingress to network egress. However, this can be changed with the QoS configuration applied to the egress SAP of the carrier IOM toward the BB-ISA.

L2TP traffic originated locally in LNS can be marked with the router or service vprn sgt-qos hierarchy.

5.2.20 Sub and sla profile considerations

In the MLPPPoX case on LNS, multiple sessions are tied into the same subscriber aggregate-rate-limit using a sub-profile. The consequence is that the aggregate rate of the subscriber can be adjusted dynamically depending on the advertised link speed in the last mile and the number of links in the bundle.

Note: Shaping in the LNS is performed per the entire MLPPPoX bundle (subscriber) rather than per individual member links within the bundle. The exception is a MLPPPoX bundle with the single member link (interleaving case) where the relationship between the session and the MLPPPoX bundle is 1:1.

In the LAC, the subscriber aggregate rate cannot be dynamically changed based on the number of links in the bundle and their rate. The LAC has no notion of MLPPPoX bundles. However, multiple sessions that in reality belong to an MLPPPoX bundle under the subscriber are shaped as an aggregate (agg-rate-limit under the sub-profile). This in essence yields the same shaping behavior as on LNS.

Sla-profile

Sessions within the MLPPPoX bundle in LNS share a single sla-profile instances (queues).

In the LAC, as long as the sessions within the subscriber6 are on the same SAP, they can also share the same sla-profile. This is be the case in MLPPPoX.

The manner in which sub/sla-profile are applied to MLPPPoX bundles and the individual sessions within results in aggregate shaping per MLPPPoX bundle as well as allocation of unique set of queues per MLPPPoX bundle. This is valid irrespective of the location where shaping is executed (LAC or LNS). Other vendors may have implemented shaping per session within the bundle and this is something that needs to be taken into consideration during the migration process.

5.2.21 MLPPPoX session setup flow example

LAC behavior

- A new PPP(oEoA) session request arrives to the LAC (PADI or LCP Conf Req).
- The LAC negotiates PADx session if applicable.
- The LAC may negotiate MLPPPoX LCP phase with its own endpoint discriminator, or it may reject MLPPPoX specific options in LCP if MLPPPoX on the LAC is disabled (such as, **no accept-mrru** in the LAC's ppp-policy). If MLPPPoX options (seq num header format, ED, MRRU) are rejected, the assumption is that the client renegotiates plain PPP(oEoA) session with the LAC.
- After LCP (MLPPPoX capable or not) is negotiated, the session is authenticated (PAP/CHAP).
- Upon successful authentication, an L2TP tunnel is identified to which the session belongs.
- If the session is a non-L2TP session (PTA MLPPPoX capable session for which the tunnel cannot be determined), the session is terminated.
- Otherwise, the QoS constructs are created for the subscriber hosts: the session is assigned to a sub/sla-profiles.
- The session LCP parameters are sent to the LNS by call management messages.

Note: If another LCP session is requested on the same bundle, the LAC creates a new LCP session and join this session to the existing subscriber as another host. In other words, the LAC is bundle agnostic and the two sessions appears as two hosts under the same subscriber.

The following assumes that MLPPPoX is configured on the LNS under the L2TP group or the tunnel hierarchy.

LNS behavior

- The LNS have the option to accept the LCP parameters or to reject them and start renegotiating LCP parameters directly with the client.
- If the LNS choose to renegotiate LCP parameters with the client directly, this renegotiation is completely transparent to the LAC by the means of a T-bit (control vs. data) in the L2TP header. LCP is renegotiated on the LNS with all the options necessary to support MLPPPoX.

  Note: Endpoint Discriminator is not mandatory in the MLPPPoX negotiation. If the client rejects it, the LNS must still be able to negotiate MLPPPoX capable session (same is valid for the LAC). If the client's endpoint discriminator is invalid (bad format, invalid class, and so on), the 7750 SR is not negotiated MLPPPoX and instead a plain PPP session is created.

- If the LNS is configured to accept the LCP Proxy parameters, the LNS determines the capability of the client.

If there is no indication of MLPPPoX capability in the Proxy LCP (not even in the original ConfReq), the LNS may accept plain (non MLPPPoX capable) LCP session or renegotiate from scratch the non MLPPPoX capable session.

If there is an indication of MLPPPoX capability in the Proxy LCP (either completely negotiated on the LAC or at least attempted from the client), the LNS tries to accept the MLPPPoX negotiated session by the LAC or renegotiate the MLPPPoX capable session directly with the client.

If the LCP Proxy parameters with MLPPPoX capability are accepted by the LNS then the endpoint as negotiated on the LAC is also accepted.

- After the MLPPPoX capable LCP session is negotiated or accepted, authentication can be performed on the LNS. Authentication on the LNS can be restarted (CHAP challenge/response with the client), or accepted (chap challenge/response accepted and verified by the LNS through RADIUS).

  Note: chap-challenge length is configurable in LNS.

- If the authentication is successful, depending on the evaluation of the parameters negotiated up to this point a new MLPPPoX bundle is created or an existing MLPPPoX bundle is joined. In case that a new bundle is established, the QoS constructs for the subscriber(-host) is created (sub/sla-profile). Session negotiation advances to IPCP phase.
- The decision whether a new session should join an existing MLPPPoX bundle, or trigger creation of a new one is governed by RFC 1990, *The PPP Multilink Protocol (MP)*, section 5.1.3, page 16, cases 1,2,3, and 4.

Note: Interleaving is supported only on MLPPPoX bundles with single session in them.

5.2.22 Other considerations

- IPv6 is supported.
- AA is supported at LNS where full IP packets can be redirected with AA policies.

- Intra-chassis redundancy is supported:
  - CPM stateful failover
  - BB-ISA — non-stateful failover

5.3 Configuration notes

MLPPP in subscriber management context is supported only over Ethernet transport (MLPPPoX). Native MLPPP over PPP/HDLC links is supported outside of the subscriber management context on the ASAP MDA.

MLPPPoX is supported only on LNS.

Interleaving is supported only on MLPPPoX bundles with a single member link. If more than one link is present in an MLPPPoX bundle, the interleaving is automatically disabled and a SNMP trap is generated. The MIB for this even is defined as tmnxMlpppBundleIndicatorsChange.

If MLPPPoX is enabled on LNS, the load balancing mode between the BB-ISAs within the group should be set to per tunnel. This ensures that all sessions of the same MLPPPoX bundle are terminated on the same BB-ISA. On the LAC, sessions of the same bundle are setup in the same tunnel.

Virtual schedulers are not supported on MLPPPoX tunnels on LNS. However, aggregate-rate-limit is supported.

The aggregate-rate-limit on LNS is automatically adjusted to the minimum value of:

- configured aggregate-rate-limit
- minimum last mile rate (obtained by LUDB, RADIUS, or PPPoE tags) multiplied by the number of links in the bundle.

The aggregate-rate-limit on the LAC is not adjusted automatically. Therefore, if configured it should be set to a high value and therefore the traffic treatment should rely on QoS performed on the LNS.

The rate (rate-down information) of the member links within the bundle must be the same. Otherwise the lowest rate is selected and applied to all member links.

A single CoA for a rate change (Alc-Access-Loop-Rate-Down) of an individual link in an MLPPPoX bundle modifies rates of all links in the bundle. This is applicable on LNS only.

The range of supported last mile rate (rate-down information) for the member links on an MLPPPoX session is 1 kb/s to 100 Mb/s. On the LNS, obtain the last mile rate:

- From the LAC by Tx-Connect-Speed AVP or by standard L2TP encoding as described in the RFC 5515, *Layer 2 Tunneling Protocol (L2TP) Access Line Information Attribute Value Pair (AVP) Extensions*.
- From the LAC by LUDB or RADIUS
- Directly on the LNS by LUDB or RADIUS.

The session fails to come up if the obtained rate-down information is outside of the allowable range (1 kb/s to 100 Mb/s).

A session within the MLPPPoX bundle is terminated if the rate-down information for the session is out of bounds (1 kb/s to 100 Mb/s).

If a member link in the last mile fails, traffic is blackholed until the LNS is notified of this failure. The failure detection in the LNS relies on PPP keepalives.

Shaping is performed per MLPPPoX bundle and not individually per member links.

If encapsulation overhead associated with fragmentation is too large in comparison to payload, the fragments are sized based on the encapsulation overhead (to increase link efficiency) instead of on maximum transmission delay.

There can be only a single MLPPPoX bundle per subscriber.

MLPPPoX bundles and non-MLPPPoX (plain L2TP PPPoE) sessions cannot coexist under the same subscriber.

Filters and mirrors (LI) are not supported on MLPPPoX bundles on LAC.

ip-only type mirrors are supported on MLPPPoX bundles.

In MLPPP scenario, downstream traffic is traversing Carrier IOM and BB-ISA twice. This is referred to as dual-pass and effectively cuts the throughput for MLPPP in half (for example, 5Gb/s of MLPPP traffic on a 10Gb/s capable BB-ISA).

6 Layer 2 Tunneling Protocol (L2TP)

This chapter provides information about using L2TP, including theory, supported features and configuration process overview.

Note: The information in this section applies only to the 7750 SR.

6.1 Terminology

- **Tunnel spec**

  Describes the requirements for a tunnel and is defined as a set of parameters that are used in tunnel setup and selection process. The tunnel-spec is defined in the CLI or can be supplied through RADIUS.

- **Tunnel** (instance)

  A run-time object with a unique ID terminating at a specific peer. Any change in the tunnel spec after the tunnel has been created has no bearing on the tunnel itself. The list of tunnels can be obtained using the **show router l2tp tunnel** command.

- Peer

  A run-time object that is defined by an ip-address/port combination. Multiple tunnels can be terminated on the same peer. The list of peers can be obtained using the **show router l2tp peer** command.

6.2 L2TP overview

6.2.1 LAC DF bit

The Layer 2 access concentrator (LAC) DF bit is configurable, but by default, it sends all L2TP packets with the DF bit set to 1. Clearing the DF bit allows downstream routers to fragment the L2TP packets. The LAC itself does not fragment L2TP packets. L2TP packets that have a larger MTU size than what the LAC egress ports allows are dropped. The DF bit can also be configured through RADIUS attribute **Alc-Tunnel-DF-bit**.

6.2.2 Handling L2TP tunnel/session initialization failures

6.2.2.1 L2TP tunnel/session initialization failover mechanisms on LAC

In deployment scenarios with multiple LNS nodes, a list of those LNS nodes can be presented to the LAC during the L2TP session instantiation process (either through CLI or RADIUS). An example of this would be a RADIUS Accept message with a list of tunnel peers:

```
tunnel.com Auth-Type := Local, Password == "tunnel1"
Tunnel-Type:1 += L2TP,
       Tunnel-Medium-Type:1 += IP,
       Tunnel-Client-Auth-Id:1 += lns_tun,
       Tunnel-Assignment-Id:1 += 1,
     Tunnel-Client-Endpoint:1 += 10.0.0.1,
     Tunnel-Server-Endpoint:1 += 10.0.0.2,
       Tunnel-Password:1 += TUNNELPASS,

          Tunnel-Type:2 += L2TP,
     Tunnel-Medium-Type:2 += IP,
         Tunnel-Client-Auth-Id:2 += lns_tun,
     Tunnel-Assignment-Id:2 += 2,
        Tunnel-Client-Endpoint:2 += 10.0.0.1,
        Tunnel-Server-Endpoint:2 += 10.0.0.3,
        Tunnel-Password:2 += TUNNELPASS,

               Tunnel-Type:3 += L2TP,
        Tunnel-Medium-Type:3 += IP,
        Tunnel-Client-Auth-Id:3 += lns_tun,
        Tunnel-Assignment-Id:3 += 3,
        Tunnel-Client-Endpoint:3 += 10.0.0.1,
        Tunnel-Server-Endpoint:3 += 10.0.0.4,
        Tunnel-Password:3 += TUNNELPASS,
               Tunnel-Type:4 += L2TP,

        Tunnel-Medium-Type:4 += IP,
        Tunnel-Client-Auth-Id:4 += lns_tun,
        Tunnel-Assignment-Id:4 += 4,
        Tunnel-Client-Endpoint:4 += 10.0.0.1,
        Tunnel-Server-Endpoint:4 += 10.0.0.5,
        Tunnel-Password:4 += TUNNELPASS
```

If the tunnel or the session establishment attempt fails for any reason, a search for additional operational facilities (tunnels or peers) is made to complete the establishment of the tunnel or session that failed in the previous attempt. Sometimes it is required to go beyond this automatic search for the new facilities and place the tunnel or peer in question into a denylist. A tunnel timeout always forces the corresponding peer and the tunnel into the denylist. In addition, a tunnel can be forced into the denylist by specific explicit error codes (CDN, and Stop-CCN) during the tunnel or session initialization phase. A peer is never forced on a denylist because of explicit Result-Code sent by LNS.

Denylisted peers and tunnels are not eligible to serve new incoming L2TP session until they are removed from the denylist. The exception is when all tunnel specs evaluate into a denylisted item. Then, a denylist item (tunnel) is tried.

6.2.2.2 Peer denylist

A peer is always placed into the denylist if:

- An attempt to establish a new tunnel fails because of a time out (SCCRQ and SCCCN timeouts)

- The timeout occurs on any control packet within an already established tunnel. All sessions on such tunnel are terminated (PADT is sent toward the clients, StopCCN is sent toward the LNS). Other tunnels that are terminated on the same peer times out on their own (if the peer is indeed non-operational), for example, the 7750 SR not explicitly tear them down based on the timeout of a single tunnel. The timeout of an existing tunnel is caused by the lack of acknowledgments to be transmitted control packets (ICRQ, ICCN, CDN, Hello).

A tunnel timeout occurs if an acknowledgment is not received after max-retries-established (on an established tunnel) or max-retries-not-established (for the tunnel in the process of being established) retries.

Although there is no configuration option that would control whether a peer can or cannot be denylisted (it is always denylisted on tunnel timeout), the amount of time that a peer remains in the denylist is configurable within the **tunnel-selection-blacklist** CLI node.

6.2.2.3 Tunnel denylists

A tunnel spec (that evaluates into a tunnel) is temporary unusable if that corresponding peer or the tunnel is denylisted. The following events trigger placement of the tunnel into the denylist:

1. Explicit termination of the L2TP session that is in the process of being established within this tunnel. The following CDN Result Codes places a tunnel to a denylist (text in parenthesis are CLI keywords that enable specific Result Codes as triggers and [rx,tx] is direction of the messages from the LAC perspective):
   - 02 DisconnectedSeeErrorCode, rx (**cdn-err-code**)
   - 04 TempMissingFacilities, rx (**cdn-tmp-no-facilities**)

     Transmit CDN when no session can be allocated

     Audit not yet complete
   - 05 PermanentMissingFacilities, rx (**cdn-perm-no-facilities**)

     No result code available
   - 06 InvalidDestination, rx (**cdn-inv-dest**)

     Tunnel is not usable (for example, **lns-group** is not configured on LNS)
   - 10 NotEstablishedInAllotedTime, tx (tx-cdn-not-established-in-time)
2. Explicit termination of the L2TP tunnel in the process of establishment by Stop-CCN Result-Codes:
   - (1) General request to clear control connection, rx (**stop-ccn-other**)
   - (2) General error, rx (**stop-ccn-err-code**)
   - (4) Requester is not authorized to establish a control channel, rx, tx (**stop-ccn-other**)
   - (5) Protocol version not supported, rx, tx (**stop-ccn-other**)
   - (6) Requester is being shut down, rx (**stop-ccn-other**)

Error messages identified by the received Result-Codes can be interpreted as the inability of the LNS to accept additional L2TP sessions within the tunnel (for example because of resource depletion) or to accept additional new tunnels.

The following statements further describe behavior related to the placement of tunnels into the denylist:

- A new L2TP session establishment attempt does not trigger on the tunnel that is in the denylist. Instead, another tunnel is searched according to the configured preference model.

- The tunnel or session initialization failure always triggers the selection mechanism for another tunnel. However, it is possible to control through configuration whether to denylist or not the tunnel for which the L2TP initialization process failed because of specific Result Codes in CDN or Stop CCN messages.
- After the L2TP tunnel or session is established, no events other than the timeout can force the tunnel (and the peer) into the denylist. In other words, a tunnel Stop or Call disconnect message for a stable tunnel or session does not force the tunnel into the denylist.
- Existing sessions within the L2TP tunnel are not purposefully terminated if that the tunnel is forced into the denylist because of an explicit reply from LNS indicating the tunnel or session initialization failure. In other words, although the L2TP tunnel may be denylisted and therefore prevented from serving new L2TP sessions, the existing L2TP session over this tunnel is not affected.
- A peer is not forced into the denylist if the explicit failure response from that peer. Only tunnels are denylisted in that case, if the configuration trigger is enabled. Peers are denylisted only based on timeouts and not explicit responses.

When the end-point is not in the routing table (unreachable through routing), the end-point is marked as permanently unavailable (removed from the L2TP process). Such end-point is never denylisted.

6.2.2.3.1 Tunnel timeout because of peer IP address change

When the peer address is changed mid-session (for example, from configured IP@ 1.1.1.1 to the new IP@ 2.2.2.2), and then subsequently the tunnel times-out, the new peer 2.2.2.2 would be placed in the denylist by default. The tunnel itself would not be placed in the denylist because it is originally tied to a different peer address that it is not in the denylist. As such it would be eligible for selection the next time a new session request for it arrives. To block selection of this failed tunnel, optionally (by configuration) force it into the denylist.

This behavior can be enabled with the following CLI:

```
configure router l2tp
configure service vprn <id> l2tp
    tunnel-selection-blacklist
        add-tunnel on <reason> [<reason>...(up to 7 max)]
<reason> : cdn-err-code|cdn-inv-dest|cdn-tmp-no-facilities|cdn-perm-no-
facilities|tx-cdn-not-established-in-time|stop-ccn-err-code|stop-ccn-other|addr-
change-timeout
```

6.2.2.4 Tunnel selection mechanism

After the L2TP tunnel failover is triggered (timeout or specific L2TP session or tunnel setup error message), a new tunnel spec in the list of available tunnel specs are selected. This tunnel selection mechanism can be controlled with CLI so that the new tunnel-spec is selected from the next preference level. Alternatively, the tunnel selection mechanism can be set to a mode where all the possibilities within the same preference are exhausted, tunnel specs on a higher preference level are tried.

```
configure router l2tp
configure service vprn <id> l2tp
    next-attempt same-preference-level | next-preference-level
```

When all tunnels on a specific preference levels are denylisted, then the behavior depends on the configuration option as per the following:

- **next-attempt = next-preference**

  Only one tunnel spec from the current preference level is tried before switching to the next preference level.

- **next-attempt = same-preference**

  All tunnel specs are tried before switching to the next preference level.

6.2.2.5 Tunnel probing

Tunnel probing refers to the mechanism where the denylisted tunnel or an end-point can be selected to serve only a single L2TP session initialization request. Only if this single L2TP session is successfully established over the selected tunnel, the tunnel can be removed from the denylist and consequently can serve new L2TP sessions. The tunnel is eligible for probing after its preconfigured time in the denylist has expired.

This behavior ensures that the new session initialization requests are not buffered while waiting for the tunnel to transition into operational state. Buffering would incur session setup delay and in the worst case it would cause session timeout if the L2TP tunnel cannot be established.

Without tunnel probing enabled, tunnels are automatically removed from the denylist upon the expiry of the preconfigured timer. New consecutive L2TP session initialization requests for such tunnels are always buffered.

6.2.2.6 Controlling the size of the denylist

The size of the denylist and the time that an item remains ineligible for selection within the denylist, is configurable.

```
configure router l2tp
configure service vprn <id> l2tp tunnel-selection-blacklist
        max-time    1..60 (minutes)
        max-list-length unlimited | 1..65535
```

6.2.2.7 Displaying the content of a denylist

The content of a denylist along with the remaining time that each entity is confined to the denylist can be displayed with the following command:

```
show router <id> l2tp peer blacklisted|not-blacklisted|selectable
```

The following displays denylist information.

```
show router l2tp peer 10.100.0.2
===============================================================================
Peer IP: 10.100.0.2
===============================================================================
Roles capab/actual: LAC LNS /LAC  -     Draining          : false
Tunnels           : 1                   Tunnels Active    : 0
Sessions          : 1                   Sessions Active   : 0
Reachability      : blacklisted         Time Unreachable  : 01/31/2013 08:55:06
Time Blacklisted  : 01/31/2013 08:55:06 Remaining (s)     : 34
```

```
===============================================================================
Conn ID                     Loc-Tu-ID Rem-Tu-ID State             Ses Active
  Group                                                          Ses Total
    Assignment
-------------------------------------------------------------------------------
977207296                   14911     0         closed            0
  base_lac_base_lns                                              1
    t1
-------------------------------------------------------------------------------
No. of tunnels: 1
===============================================================================


show router l2tp tunnel detail
===============================================================================
L2TP Tunnel Status
===============================================================================
Connection ID: 831782912
State        : closedByPeer
IP           : 10.0.0.1
Peer IP      : 10.100.0.2
Tx dst-IP    : 10.100.0.2
Rx src-IP    : 10.100.0.2
Name         : lac
Remote Name  :
Assignment ID: t1
Group Name   : base_lac_base_lns
Acct. Policy : l2tp-base
Error Message: N/A
                                    Remote Conn ID    : 4294901760
Tunnel ID         : 12692           Remote Tunnel ID  : 65535
UDP Port          : 1701            Remote UDP Port   : 1701
Preference        : 50              Receive Window    : 64
Hello Interval (s): 300
Idle TO (s)       : 5               Destruct TO (s)   : 60
Max Retr Estab    : 5               Max Retr Not Estab: 5
Session Limit     : 32767           AVP Hiding        : sensitive
Transport Type    : udpIp           Challenge         : never
Time Started      : 01/31/2013 08:56:58 Time Idle       : 01/31/2013 08:56:58
Time Established  : N/A             Time Closed       : 01/31/2013 08:56:58
Stop CCN Result   : reqShutDown     General Error     : noError
Blacklist-state   : blacklisted
Blacklist Time    : 01/31/2013 08:56:58 Remaining (s)     : 49
-------------------------------------------------------------------------------
No. of tunnels: 1
===============================================================================
```

6.2.2.8 Generating a trap when the denylist is full

A log is generated when the denylist reaches its max limit of items. The log event is tmnxL2tpTunnelSelectionBlacklistFull.

6.2.2.9 Premature removal of denylisted entries

When the total number of supported tunnels and peers in a denylist and when the LAC, in general, has reached its maximum, on the new session initialization request, the oldest tunnel entry in the denylist is removed from the denylist regardless if the denylist max-time has expired.

6.2.2.10 Manual purging of entities within the denylist

The items can be manually purged from the denylist using the following commands.

```
clear router <id> l2tp tunnel-selection-blacklist
clear router <id> l2tp peer <ip-address> [udp-port <port>] tunnel-selection-
blacklist
clear router <id> l2tp group <tunnel-group-name> [tunnel <tunnel-name>] tunnel-
selection-blacklist
clear router <id> l2tp tunnel <connection-id> tunnel-selection-blacklist
```

6.2.3 CDN result code overwrite

When the number of L2TP sessions reaches the configured maximum value, the LNS sends an out-of-resource Result Code (4 or 5) in a CDN (Call-Disconnect-Notify) message to the LAC. This would trigger the LAC to fail over to another LNS that has the resources available. Similarly, when the tunnel is not usable because of the invalid destination CDN error, the Result-Code 6 is sent from the LNS.

Certain third-party LAC implementations trigger tunnel failover only when they receive Result Code 2 in CDN messages (and not 4,5 or 6). To support those scenarios, the LNS in the 7750 SR can overwrite result Codes 4, 5 and 6 with result Code 2 just before they are sent to the LAC. Result Codes can be overwritten only during the L2TP session initialization phase. These codes have the following meanings and are described in RFC 2661, 4.4.2:

- **2**

  Call disconnected for the reason indicated in error code
- **4**

  Call failed because of the lack of appropriate facilities being available (temporary condition)
- **5**

  Call failed because of the lack of appropriate facilities being available (permanent condition)
- **6**

  Invalid Destination

This functionality is enabled on LNS with the following CLI hierarchy:

```
configure router l2tp
configure service vprn <id> l2tp
   replace-result-code {cdn-tmp-no-facilities | cdn-prem-no-facilities | cdn-inv-
dest}
    no replace-result-code
```

6.2.4 LNS proxy

LNS offers a proxy LCP (with the **proxy-lcp** command) function where LCP-related information is cached temporarily in the LNS during the ICCN phase where L2TP control messages are exchanged between the LAC and LNS. The LNS can then use the cached information to bypass the LCP negotiation and immediately start the authentication state with the client. Furthermore, proxy authentication (using the **proxy-authentication** command) can also be enabled on the LNS to bypass authentication and the client

can immediately start the IPCP negotiation phase. If the proxy LCP information conflicts from the LNS configuration, then the LNS forces the client to re-start LCP negotiation. LCP negotiation is not restarted in the proxy LCP mode when:

- MRU is missing in the LastTxLcpConfReq AVP
- The magic number is missing in LastTxLcpConfReq AVP
- Async-Control-Character-Map (ACCM) with value = 0x00000000 is present in LastTxLcpConfReq and LastRxLcpConfReq AVP's
- Address-and-Control-Field-Compression (ACFC) is present in LastTxLcpConfReq and LastRxLcpConfReq. Note that PPP frames with and without address and control field (0xFF03) in the PPP header are accepted. LCP frames without 0xFF03 are also accepted as valid frames.

Also, proxy-authentication that fails then forces the client to re-start the LCP negotiation again.

6.3 L2TP LAC VPRN

Layer 2 Tunneling Protocol (L2TP) allows for PPP sessions to be carried over an IP network.

Each L2TP session transports PPP frames, irrespective of link-layer encapsulation, allows the LNS to terminate PPP sessions that were PPPoE. L2TP is carried over IPv4 packets in UDP datagrams (default port 1701).

If session data is not reliably delivered, that is, if there is a packet loss, there is no retransmission, a sequence numbers is used within each L2TP session to identify packet loss and re-ordering.

L2TP consists of the following concepts:

- L2TP tunnels — L2TP tunnel is a connection between one LAC (L2TP Access Concentrator) and one LNS (L2TP network server) that share a common control channel.
- L2TP sessions — Within each L2TP tunnel, there exists one or more L2TP sessions (one PPP session corresponds to exactly one L2TP session)

L2TP tunnels provide an IP transport for PPP frames between LAC and LNS. In some existing networks, BGP/MPLS VPNs (VPRN in SR OS) are used to contain the L2TP traffic (and the routes associated with the LAC and LNS) into a dedicated routing instance.

Like the LNS implementation, L2TP LAC in a VPRN allows L2TP control and data traffic to be sourced from and received by any valid IP interface within the VPRN (including loopback and interface addresses). L2TP frames may ingress a network port (with up to five MPLS tags) or access ports with SAPs associated with the VPRN IP interfaces.

Non-hitless multichassis LAC resiliency

In dual-homed PPPoEv4/v6 wholesale/retail environment over L2TP, the subscriber-hosts are synchronized by the Multi-Chassis Synchronization (MCS) protocol. The failover detection mechanism can be implemented by SRRP or Layer 3 MC-LAG with SRRP. When an interface or an entire node fails, the new multichassis active BNG (SRRP master state) sends PADT to all sessions that were moved over from the failed node.

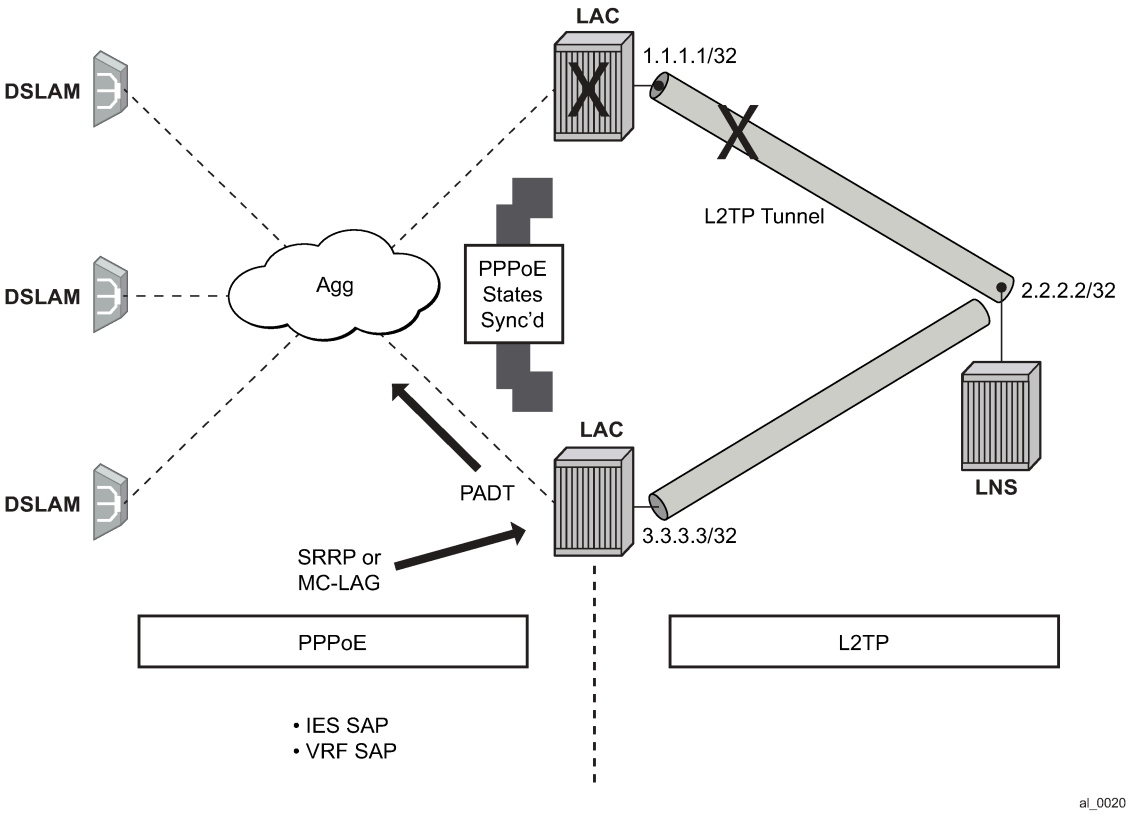
In the event of an interface-only failure, CDN is sent toward the LNS to terminate sessions on the LNS.

The PPPoE sessions are reestablished on the new multichassis active BNG, but because PADT was sent to clients the recovery time is faster (no need to wait for PPPoE session timeout). On the network side (toward the LNS) an existing tunnel toward the LNS can be used to re-establish the sessions or if none exists, a new tunnel is established. Then there is no need for a redundant interface.

Note:
The L2TP tunnel carrying the sessions must always be terminated on the multichassis active LAC (SRRP master state).

In the event of nodal failure, the sessions within the old tunnel on the LAC times out (CDN cannot be sent from the new multichassis active LAC because there is no tunnel state preserved across redundant LAC nodes). During the time-out period, the LNS must maintain double the amount of failed sessions (stale ones plus the new ones). This model is shown in Figure 43: Non-hitless interface/node protection on the LAC .

Figure 43: Non-hitless interface/node protection on the LAC

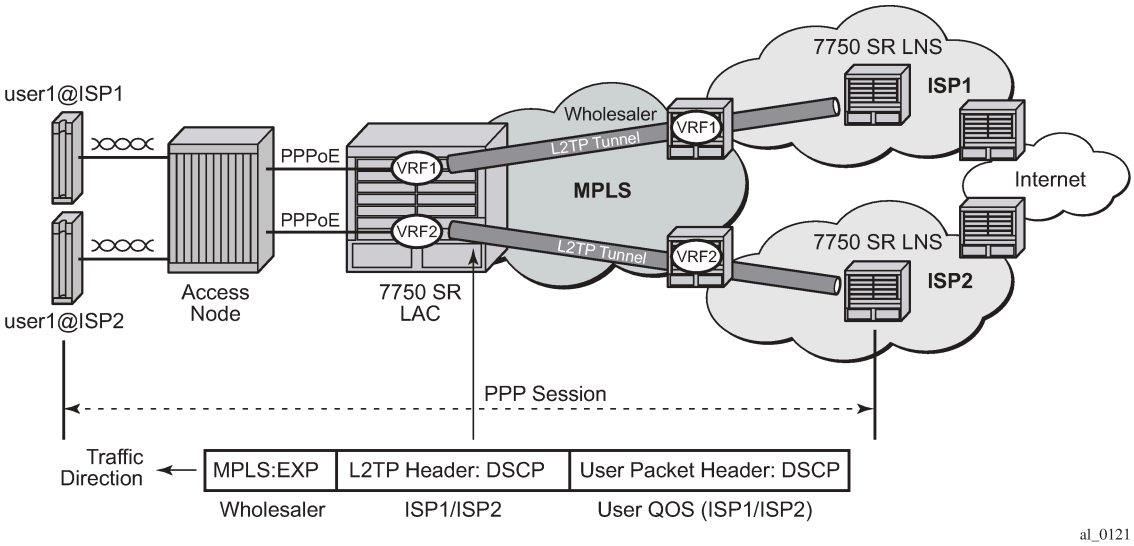


6.3.1 Per-ISP egress L2TP DSCP reclassification

Wholesale providers can deliver Internet access to directly connected PPP users through third party ISPs. This involves the users connecting to an L2TP Access Concentrator (LAC) with their traffic being tunneled to and from an L2TP Network Server (LNS) in their ISP.

If there is a requirement to support per-ISP (and per-subscriber host) QOS control for downstream traffic on the LAC toward the users based on the DSCP marking in the L2TP header, the **use-ingress-l2tp-dscp** command must be configured within the sla-profile selected for the users. An example topology is shown in Figure 44: ISP Internet access through wholesale provider.

Figure 44: ISP Internet access through wholesale provider



The downstream traffic arrives at the LAC with:

- An MPLS header (because of the VRF encapsulation). This contains EXP bits which are set based on the wholesale provider's QOS scheme.
- An L2TP header (because of the L2TP tunnel to the ISP). This contains DSCP bits in its IP header which are set by the originating ISP.
- A user IP packet header. This contains DSCP bits which could be set by the ISP or by the originating Internet application.

The network ingress on the LAC would normally use the MPLS EXP bits for traffic QoS classification, however, this matches the wholesale provider's QoS scheme.

It is possible to apply the **ler-use-dscp** parameter at the LAC network ingress to classify based on the L2TP header DSCP, but this would require the QoS schemes used by all ISPs, and the wholesale provider, to have a consistent interpretation of the DSCP bits.

If the standard egress IP reclassification is used, the QOS would be dependent on the DSCP in the user packet.

Configuring the **use-ingress-l2tp-dscp** parameter in the sla-profile of the ISP1 and ISP2 users forces the egress QoS control to be based on the DSCP from the L2TP header received on the LAC (which is set by ISP1/ISP2). This provides per-ISP (and per-subscriber host) QoS control for downstream traffic on the LAC toward the users.

6.4 Traffic steering on L2TP LAC

Traffic steering on L2TP LAC allows wholesale providers to forward L2TP-encapsulated packets going to and coming from LNS to Value-Added Services (VAS).

Traffic steering on L2TP LAC consists of the following components:

- A steering profile contains steering configuration (Access/VAS routers and next hop) information that is applied to the subscriber host for the PPPoE/L2TP session.

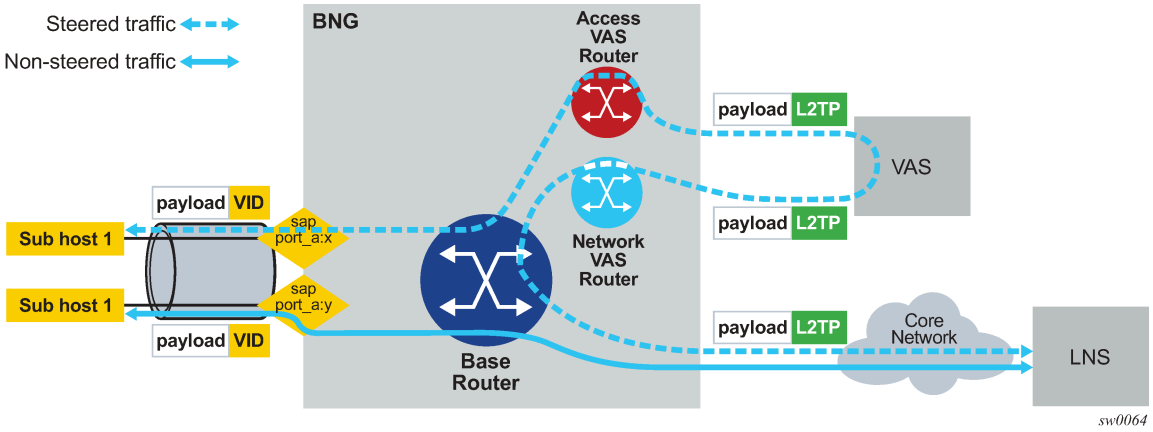
- For steered traffic all PPP packets to and from the PPPoE host that have a steering profile attached are forwarded through VAS. PPP packets include LCP/NCP control packets, LCP echo and echo reply, and user data packets.
- Non-steered traffic consists of packets for the L2TP control channel and PPP packets of the subscriber host that do not have a steering profile
- For the base router (can also be a VPRN), the routing instance terminates subscriber host and L2TP tunnels or sessions to LNS
- For an access VAS router, the routing instance forwards upstream (to LNS) steered traffic to VAS and receives downstream (to subscriber) steered traffic from VAS.

  An access VAS router must be a VPRN. A base routing instance cannot be specified as an access VAS router.
- For a network VAS router (optional), the routing instance receives upstream (to LNS) steered traffic from VAS and forwards downstream (to subscriber) steered traffic to VAS.

  The creation of a network VAS router is not mandatory and a base router can also perform the same function.
- In a network VAS next hop, the next-hop IP address to reach VAS from the network VAS Router or Base Router for downstream traffic. This address must be specified in the steering profile.

Figure 45: Traffic steering on L2TP LAC shows traffic steering on L2TP LAC.

Figure 45: Traffic steering on L2TP LAC



6.4.1 Steering activation and deactivation

Traffic steering can be activated by the LUDB and RADIUS Access-Accept/CoA messages that include the Alc-Steering-Profile VSA.

Steering can be deactivated by a RADIUS CoA that includes an Alc-Remove-Override VSA to remove Alc-Steering-Profile generated by an AAA server or BNG node itself using a CLI command.

The following CLI example shows the activation of a steering profile:

```
tools perform subscriber-mgmt coa alc-subscr-id subscriber-
1 attr evs,241,6527,25=”steering-profile-1”
```

The following CLI example shows the deactivation of a steering profile:

```
tools perform subscriber-mgmt coa alc-subscr-id subscriber-
1 attr 6527,238="deactivate 241.26.6527.25"
```

6.4.2 Steering states

Each PPPoE/L2TP session has an operational state of traffic steering that prevents a traffic black hole caused by problems such as an incorrectly configured network or a temporary VAS outage.

The steering states are:

- **Non-steered**

  a steering profile is not applied to the session and traffic steering is not performed

- **Steered**

  a steering profile is applied to the session and traffic steering is performed

- Steering-failure a steering profile is applied to the session but traffic steering is not performed, possibly because of a misconfiguration of the steering profile, L2TP, or IP routing

As the conditions to perform traffic steering are met or lost, the steering state transitions in and out of steering failure.

6.4.3 Configuring traffic steering on L2TP LAC

The following steps show the commands used to configure traffic steering on L2TP LAC.

1. Enable L2TP on a base router.

```
router
    l2tp
         no shutdown
         exit
    exit
```

2. Create VAS routers and interfaces.

```
vprn 100 customer 1 create
    description “VAS router for access”
    route-distinguisher 65001:100
    vrf-target target:65001:100
    l2tp
        no shutdown
        exit
    interface “L2TP LAC endpoint”
        address 10.0.0.1
        loopback
        exit
    interface “Access interface to VAS”
        address 10.0.100.100/24
        vas-if-type to-from-access
        exit
    static-route-entry 20.0.0.1/32 next-hop 10.0.100.200
vprn 200 customer 1 create
    description “VAS router for network”
    route-distinguisher 65001:200
```

```
        vrf-target target:65001:200
        interface "Network interface to VAS"
            address 10.0.100.200/24
            exit
        grt-lookup
            enable-grt
            exit
        static-route-entry 10.0.0.1/32 next-hop 10.0.100.100
```

The L2TP endpoint IP addresses used by the base router and the access router must be the same.

The VAS-facing interface on the access router must be configured as **vas-if-type to-from-access** to avoid a traffic loop between BNG and VAS.

3. Configure the steering profile.

```
    subscriber-mgmt
        steering-profile "SP1" create
            access router 100
                exit
            network next-hop 10.0.100.200 router 200
                exit
        exit
```

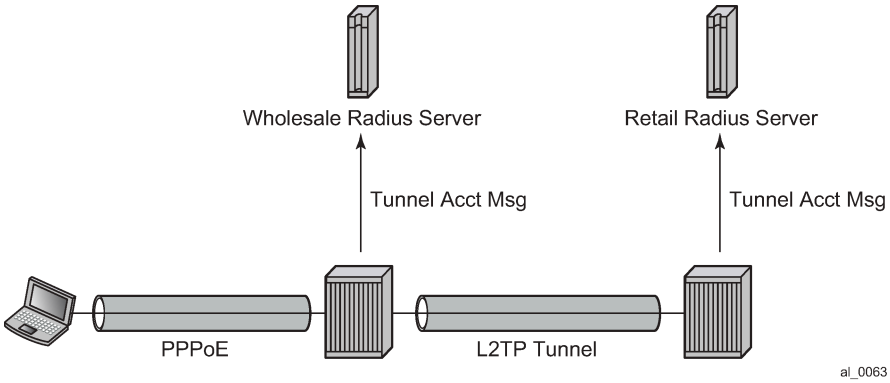
4. (Optional) Configure the LUDB host entry for the steered L2TP/PPPoE session.

```
    subscriber-mgmt
        local-user-db "LAC-steering"
            ppp
                match-list sap-id
                host "lag-1:1" create
                    host-identification
                        sap-id "lag-1:1"
                    exit
                    auth-policy "AUTH1"
                    pre-auth-policy "PRE1"
                    steering-profile "SP1"
                    identification-strings 254 create
                        sla-profile-string "SLA1"
                        sub-profile-string "SUB1"
                    exit
                    no shutdown
                exit
```

6.5 L2TP tunnel RADIUS accounting

Figure 46: L2TP tunnel accounting shows an L2TP tunnel RADIUS accounting configuration.

Figure 46: L2TP tunnel accounting



When L2TP tunnel accounting is enabled, except for **host** or **sla-profile-based** accounting packets and attributes, the following are additional accounting packets and attributes:

- Accounting packets:
  - tunnel-start/stop/reject
  - tunnel-link-start/stop/reject

  There are no interim updates for L2TP tunnel/session accounting.
- RADIUS accounting attributes:
  - Tunnel-Assignment-Id (LAC only)
  - Acct-Tunnel-Connection
  - Acct-Tunnel-Packets-Lost

These attributes were added into current account-start/stop/interim-update packets (host accounting/sla-profile accounting)

Tunnel level accounting and session level accounting can be enabled or disabled independently.

New accounting packets and related RADIUS attribute list are described in Table 5: L2TP tunnel accounting behavior .

Some considerations of RADIUS attributes are described in RADIUS attributes value considerations.

6.5.1 Accounting packets list

Table 5: L2TP tunnel accounting behavior describes L2TP tunnel accounting behavior along with some key RADIUS attributes (apply for both LAC and LNS):

Table 5: L2TP tunnel accounting behavior

Act-packet	When	Key attributes
Tunnel-Start	A new L2TP tunnel is created	Acct-Session-ID Event-Timestamp Tunnel-Type:0

Act-packet	When	Key attributes
		Tunnel-Medium-Type:0 Tunnel-Assignment-Id:0 Tunnel-Client-Endpoint:0 Tunnel-Client-Auth-Id:0 Tunnel-Server-Endpoint:0 Tunnel-Server-Auth-Id:0
Tunnel-Reject	A new L2TP tunnel creation failed	Acct-Session-Id Event-Timestamp Tunnel-Type:0 Tunnel-Medium-Type:0 Tunnel-Assignment-Id:0 Tunnel-Client-Endpoint:0 Tunnel-Client-Auth-Id:0 Tunnel-Server-Endpoint:0 Acct-Terminate-Cause
Tunnel-Stop	An established L2TP tunnel is removed	Acct-Session-Id Event-Timestamp Tunnel-Type:0 Tunnel-Type:0 Tunnel-Medium-Type:0 Tunnel-Assignment-Id:0 Tunnel-Client-Endpoint:0 Tunnel-Client-Auth-Id:0 Tunnel-Server-Endpoint:0 Tunnel-Server-Auth-Id:0 Tunnel-Server-Auth-Id:0 Acct-Session-Time Acct-Input-Gigawords Acct-Input-Octets Acct-Output-Gigawords Acct-Output-Octets Acct-Input-Packets Acct-Output-Packets Acct-Terminate-Cause

Act-packet	When	Key attributes
Tunnel-Link-Start	An L2TP session is created	User-Name Acct-Session-Id — This is the same as Acct-Session-id in access-request of host auth Event-Timestamp Service-Type — Framed Class Tunnel-Type:0 Tunnel-Medium-Type:0 Tunnel-Assignment-Id:0 Tunnel-Client-Endpoint:0 Tunnel-Client-Auth-Id:0 Tunnel-Server-Endpoint:0 Tunnel-Server-Auth-Id:0 Acct-Tunnel-Connection — See RADIUS attributes value considerations
Tunnel-Link-Reject	A new L2TP session creation is failed	Acct-Session-Id — Should be as same as Acct-Session-id in access-request of host auth Event-Timestamp Tunnel-Type:0 Tunnel-Medium-Type:0 Tunnel-Assignment-Id:0 Tunnel-Client-Endpoint:0 Tunnel-Client-Auth-Id:0 Tunnel-Server-Endpoint:0 Acct-Terminate-Cause Acct-Tunnel-Connection
Tunnel-Link-Stop	An established L2TP session is removed	User-Name Acct-Session-Id — Should be as same as Acct-Session-id in access-request of host auth Event-Timestamp Service-Type — Framed Class Tunnel-Type:0

Act-packet	When	Key attributes
		Tunnel-Medium-Type:0 Tunnel-Assignment-Id:0 Tunnel-Client-Endpoint:0 Tunnel-Client-Auth-Id:0 Tunnel-Server-Endpoint:0 Tunnel-Server-Auth-Id:0 Acct-Tunnel-Connection Acct-Session-Time Acct-Input-Gigawords Acct-Input-Octets Acct-Output-Gigawords Acct-Output-Octets Acct-Input-Packets Acct-Output-Packets Acct-Tunnel-Packets-Lost Acct-Terminate-Cause

Notes:

- Errors occur if there are multiple hosts sharing the same sla-profile instance and then these hosts go to different tunnel.
- 7750 SRs have an internal limitation of 500 pps for accounting messages. This feature shares the same limitation.

6.5.2 RADIUS attributes value considerations

- The value of Acct-Tunnel-Connection uniquely identifies an L2TP session. To match LAC and LNS accounting records, the value of Acct-Tunnel-Connection is determined by a method shared by LAC and LNS. This means for a specific L2TP session, Acct-Tunnel-Connection from the LAC and LNS are the same.
- Current ESM statistics are used in tunnel link and tunnel level accounting. This applies to the standard attribute and the 7750 SR's own VSA.
- Tunnel level accounting statistics must aggregate all session statistics that belong to the tunnel.

  Note:
  There could be sessions come and go before tunnel is down, so system need to remember the stats of every session that has been created within the tunnel.

  This applies for both standard attribute and 7750 SR's own VSA.
- The value of Acct-Tunnel-Packets-Lost is the aggregation of all discarded packets on both ingress and egress.

- L2TP tunnel accounting on LAC can be enabled from RADIUS using the Alc-Tunnel-Acct-Policy VSA. This attribute overrides the locally configured L2TP RADIUS accounting policy.
- See *7450 ESS, 7750 SR, and VSR RADIUS Attributes Reference Guide* for attributes that are applicable in RADIUS L2TP tunnel accounting.

6.6 MLPPP on the LNS side

With MLPPP, the counter on LNS side is only available for the bundle, not for each link, so the SR OS's behavior is:

- For each new link session system sends a tunnel-link-start.
- For each link session that is deleted system sends a tunnel-link-stop.
- For all link sessions except the last one system reports 0 for all counters.
- For the last link session, system reports the actual counters for the bundle.

6.7 LNS reassembly

LNS reassembly is supported in the BB-ISA. Fragments are collected and reassembled. After the entire L2TP packet is reassembled, the packet is either decapsulated or sent to the CPM without change.

The delivery of the L2TP packets to the BB-ISA depends on the specific fields in the L2TP header. The forwarding decision on the ingress LNS side in the upstream direction (LAC->LNS) is based on the tunnel-id or session-id combination and the T-bit (message type bit – control or data) in L2TP header.

Control type messages are delivered directly to the CPM. CPM performs L2TP decapsulation and processes the message (tunnel or session setup/teardown related messages or tunnel hellos). The CPM provides forwarding information to the forwarding plane (ingress/egress IOM and the carrier IOM) and to the BB-ISA (tunnel-src plus tunnel-id/session-id plus generated-mac-addr and SAP).

Data type messages are delivered directly to the BB-ISA. The BB-ISA decapsulates the L2TP packets and forwards them to the carrier IOM as a quasi-PPPoE frame (ESM forwarding module).

Because the LAC fragments the packets in the upstream direction, the L2TP header is preserved only in the first fragment. Therefore, the crucial forwarding information needed by LNS is lost in all consecutive fragments. If a fragments ends up in the wrong BB-ISA with no reassembly context for the fragment, the fragment is dropped.

Similarly, the information whether to forward the fragment to the BB-ISA (data packet) or the CPM (control packet) is lost.

To support LSN reassemble, the following configuration limitations are imposed:

- Only one pair of active/standby BB-ISAs are supported. This way all fragments are forwarded to the same active BB-ISA that maintains all reassembly contexts for all fragments.
- All fragments, regardless of the packet type, are forwarded to the active BB-ISA. After the L2TP packet is reassembled, it is determined whether the packet is:
  - A data packet — The packet is decapsulated and a quasi PPPoE packet is forwarded to the carrier IOM (ESM function).

– A control packet — The packet is not decapsulated but instead it is forwarded as L2TP packet to the CPM.

The **lns-reassembly** commands that inform the ingress forwarding plane that all L2TP packets should be sent to the BB-ISA are configured in the **config>router>l2tp** and **config**>**service**>**vprn**>**l2tp** contexts.

6.8 LNS subscriber policers

6.8.1 Policer support

LNS subscribers' SLA profiles support policers. In the egress direction, HQoS manageable policers are also supported.

The following QoS features are supported for LNS subscribers:

- ingress policer with h-pol
- egress policer with h-pol
- policer-hqos-manageable
- policer-output-queue
- egress queuing, policer to local queue
- egress queuing, bypass policers
- egress queuing, flow based

7 Triple Play security

7.1 Triple Play security features

7.1.1 Anti-spoofing filters

7.1.1.1 Anti-spoofing filter types

A SAP or interface that supports anti-spoof filtering can be configured to use one of three types of anti-spoof tables. The type of table used by the SAP is dependent on the type of anti-spoof filtering needed, only one anti-spoofing table type is supported per SAP:

- When only the incoming source MAC address is to be verified, the source MAC table must be defined (**anti-spoof type mac**).
- When only the incoming source IP address is to be verified, the source IP table must be defined (**anti-spoof type ip**).
- When both the incoming source MAC and source IP addresses are to be verified, the combination source IP and source MAC table must be defined (**anti-spoof type ip-mac**).

Note: Setting the anti-spoof filter type for the SAP is dependent on pre-existing static host definitions, for example, attempting to set the SAP anti-spoof filtering to **mac** fails if any static hosts exist that do not have a defined MAC address.

The anti-spoof table of a SAP or interface is populated from the DHCP lease state table and from any statically defined hosts on the SAP or interface.

7.1.1.2 Filtering packets

Packets from a client that match an anti-spoof filter entry when anti-spoof filtering is enabled can be further processed by the system. The matching packet is still subject to other forwarding criteria including potentially ACL filtering.

All packets that are not exempt from anti-spoofing and do not match an entry in the anti-spoof table are discarded. Every discard event increments the SAP discard packet counter. The discard event is not logged or alarmed, but a threshold alarm could be configured for the counter (see the Configuring System Monitoring Thresholds section in the *7450 ESS, 7750 SR, 7950 XRS, and VSR Basic System Configuration Guide*).

Not all ingress packets are subject to the anti-spoof filtering when enabled. Non-IP packets are exempt for anti-spoof filter lookups and can be further processed by the system. This includes ARP requests and replies, as well as PPPoE packets. The only IP packets exempt from anti-spoof filtering are DHCP packets destined for the server UDP port 67. DHCP packets destined for the client UDP port number (port 68) are not exempt.

7.1.2 Layer 2 Triple Play security features

7.1.2.1 MAC pinning

This section describes the 7450 ESS and the 7750 SR acting as a Broadband Subscriber Aggregator (BSA).

DHCP snooping and IP and MAC auto-filters can be used to prevent Theft of Service (by a malicious user spoofing another user's address). However, these auto-filters do not discard non-IP packets such as PPPoE packets, potentially allowing a MAC address to be relearned on another SAP. MAC pinning closes this loophole, by not allowing a MAC address to be relearned on another SAP.

When MAC pinning is enabled, a MAC address learned on one SAP or SDP cannot be relearned on another SAP or SDP in the same VPLS, until the FDB entry for the MAC address times out. (If MAC aging is disabled, MAC entries on a SAP or SDP with MAC pinning enabled effectively becomes permanent.)

MAC pinning is implicitly enabled when DHCP auto-filters are enabled, and cannot be disabled. For MAC addressing learned during DHCP address assignment (when DHCP snooping function is active at least on one port of the VPLS), the MAC address is tied to a specific SAP for the duration of the DHCP lease.

7.1.2.2 MAC protection

In a Layer 2 environment, a malicious subscriber could create a denial-of-service (DoS) attack by sending Ethernet frames, with as source MAC address the address of a gateway (for example, the IP next hop upstream). As MAC learning is typically enabled, this would move the learned gateway MAC from the uplink SAP or SDP to the subscriber’s SAP, causing all communication to the gateway to be disrupted. If a local content server is attached to the same VPLS, a similar attack could be launched against it.

Communication between subscribers can be disallowed using Split Horizon Groups, but this by itself is not enough to prevent such an attack. The solution is to create a mechanism to explicitly protect some MAC addresses against being relearned on other SAPs.

The **mac-protect** feature on the 7450 ESS and 7750 SR allows a list of special MAC addresses to be configured in a VPLS. Two checks can then be made on incoming packets against these protected MAC addresses:

- [**no**] **auto-learn-mac-protect**

  Used to enable the automatic protection of source MAC addresses learned on the associated object. MAC protection is used in conjunction with **restrict-protected-src**, **restrict-unprotected-dst** and **mac-protect**. When this command is applied or removed, the MAC addresses are cleared from the related object.

- restrict-protected-src

  Used to prevent DoS attacks. If the source MAC address of a packet from a subscriber matches a protected entry, it is probable that this subscriber tried to impersonate the gateway or server. If no parameter is specified, such packets are discarded, a trap is generated, and the SAP on which it arrived is placed operationally down. If the **alarm-only** parameter is specified, the packet is forwarded, and an alarm is generated but the source MAC is not learned. If the **discard-frame** parameter is specified, the packet is discarded and an alarm generated.

- **restrict-unprotected-dst**

Used to force traffic from subscribers to only go toward a few defined destinations (the gateways or servers). Any packet from a subscriber whose destination MAC address does not match a protected entry is discarded.

7.1.2.3 DoS protection

This section describes the mechanisms and limitations of DoS protection related to subscriber management snooping functions. This feature is only supported on 7750 SR-Series and 7450 ESS-Series redundant chassis models. In subscriber aggregation networks, these routers play an active role in several protocols. Subscribers either intentionally or unknowingly interfere with the operation of the node's processing capacity (for example, excessive ARP handling) or another user traffic.

Routing protocols such as OSPF and ISIS could also be a threat as packets can be injected by customers (erroneously or maliciously) which could cause high CPM overload. Service providers are concerned about DoS protection including DoS attacks when acting as a subscriber aggregation device and guarding against DoS attacks using unprovisioned protocols.

7.1.2.3.1 Subscriber aggregation network

In a subscriber aggregation network, multiple devices such as the 7750 SR or 7450 ESS routers provide access to a DHCP or a RADIUS server. These servers usually do not scale high enough to provide the means to control access to snooping functions through a controlled queue. It is possible, under severe conditions, that the network could become unavailable if the node cannot handle requests from subscribers.

Because the IOMs cannot be scaled to provide a per-subscriber queue to control traffic, a monitoring function, handled by the CPM, is provided. With this monitoring system, the CPM tracks the number of control plane messages set per subscriber and limits the rate to a specified level and provides feedback using event generation to alert a centralized system of a possible DoS attack.

The CPM provides a prioritized access to the CPU. Because the number of control packets expected from a subscriber should have a low rate, and under normal conditions, the system provides a rate limit on a per subscriber or MAC basis and drops a subscriber control packet before it is queued or processed by the CPU. The system is configured with expected arrival rate of per MAC or subscriber control packet rates and optionally total rate per interface or SAP.

The system maintains a per-second running rate monitor per SAP and per MAC. If an entry is using more than the configured rate, the system does not forward that packet to be queued. Every existing subscriber host is monitored. A subscriber host is flagged and the system observed with an excessive rate of control packets. With PPPoE, the CPM monitors subscriber hosts before the IP address is provided by the SAP, MAC, or session-id combination.

The control protocols affected by this mechanism include:

- ARP (in arp-reply-agent)
- DHCP (for discover and renew)
- ICMP
- PPPoE
- IGMP

7.1.2.3.2 Network control filtering

The 7750 SR or 7450 ESS can block network control traffic for unconfigured protocols. For example, if OSPF is not configured on an IP interface, all OSPF- related traffic should be dropped before the traffic reaches the CPU.

Protocols are blocked based on whether that protocol is configured to run on the specific IP interface. It is not required to re-configure the permitted protocols.

Protocol traffic control by this mechanism includes:

- OSPFv2
- OSPFv3
- IS-IS
- RSVP-TE
- LDP
- RIP
- PIM
- MLD
- IGMP
- BGP
- BFD
- L2PT
- PPP
- DHCP

7.1.2.4 VPLS redirect policy

This section describes the 7450 ESS or 7750 SR acting as a Broadband Subscriber Aggregator (BSA) with Layer 2 aggregation toward a Broadband Subscriber Router (BSR).

In a Triple Play network it may be necessary to route some traffic from or to subscribers through a Deep Packet Inspection (DPI) device, for example, to limit peer-to-peer traffic. However such a DPI device typically has limited bandwidth available, so only those packets that need inspection should be sent to it.

In a Layer 3 network, such policy-based redirection can be achieved using "next-hop redirect" ACL entries. In a layer 2 (VPLS) aggregation network, the same result can be achieved using "redirect to SAP" or "redirect to SDP" policy.

See the ACL Next-Hop for VPLS section in the 7450 ESS, 7750 SR, 7950 XRS, and VSR Services Overview Guide.

7.1.3 ARP handling

7.1.3.1 ARP reply agent

This section describes the 7450 ESS or 7750 SR acting as a Broadband Subscriber Aggregator (BSA).

In Triple Play networks, typically downstream broadcast is not allowed on subscriber SAPs. As a result, subscribers cannot receive ARP requests from the network. Instead, the 7450 ESS or 7750 SR responds to ARP requests from the network, with information from the DHCP lease state table.

In the upstream direction (toward the network), the ARP reply agent intercepts ARP Requests on subscriber SAPs, and checks them against the DHCP lease state table. The purpose is to prevent a malicious subscriber spoofing ARP request or ARP reply messages and therefore populating the upstream router's ARP table with incorrect entries.

When the keyword **sub-ident** is added in the ARP reply agent configuration, also the subscriber identity is checked. If an upstream ARP request is targeted to the same subscriber, it is dropped. Otherwise, it is flooded to all VPLS interfaces outside the received Split Horizon Group (SHG).

Static hosts can be defined on the SAP using the **host** command. Dynamic hosts are enabled on the system by enabling the **lease-populate** command in the SAP's **dhcp** context. If both a static host and a dynamic host share the same IP and MAC address, the VPLS ARP reply agent retains the host information until both the static and dynamic information are removed. If both a static and dynamic host share the same IP address, but different MAC addresses, the VPLS ARP reply agent is populated with the static host information.

In brief, the ARP Replay Agent operation is as follows:

- For ARP request received from a customer SAP:
  - first check in DHCP lease state table — if no match: discard
  - if (sub-ident enabled) and (destination equals the same subscriber): discard
  - otherwise: flood to all SAPs/SDPs outside this SHG
- For ARP request received from the network:
  - lookup IP address in DHCP lease state table — if no match: discard
  - otherwise: respond with MAC address from the DHCP lease state table

7.1.3.2 Dynamic ARP table population

This section describes the 7450 ESS or 7750 SR acting as a Broadband Subscriber Aggregator (BSA) with Layer 3 forwarding toward the network.

In an IES service, the system's ARP table can be populated dynamically using entries in the DHCP lease state table (in turn, populated from snooping DHCP ACK messages (see DHCP Snooping)), and from static hosts defined on the SAP. In the router ARP table these are indicated with state managed.

If both a static host is created with the same IP and MAC address as an existing managed entry, creation fails and a trap is generated.

If a DHCP Lease needs to be populated with the same IP and MAC address as an existing static host entry, creation fails and a trap is generated.

No **static-arp** creation is possible when combined with **arp-populate**.

7.1.3.3 Local proxy ARP

This section describes the 7450 ESS or 7750 SR acting as a Broadband Subscriber Aggregator (BSA) with Layer 3 forwarding toward the network.

Local proxy ARP allows the 7450 ESS or 7750 SR to respond to ARP requests received on an interface, for an IP address which is part of a subnet assigned to the interface. When the local proxy ARP feature is enabled, the switch responds to all ARP requests for IP addresses belonging to the subnet with the MAC address of the interface, and forwards all traffic between hosts in the subnet.

This feature is intended to be used in situations (such as DSL aggregation networks) when hosts belonging to the same subnet are prevented from directly communicating with each other over the subnet by the configuration of the switch (or DSLAM) to which they are connected.

Note:
When **local-proxy-arp** is enabled under a IES service, all ICMP redirects on the ports associated with the service are automatically blocked. This prevents users from learning each other's MAC address (from ICMP redirects).

The implementation of proxy ARP with support for local proxy ARP allows the router to respond to ARP requests in the subnet assigned to an IES or VPRN interface, therefore allowing multiple customers to share the same IP subnet.

7.1.4 Web portal redirect

This section describes the 7450 ESS or 7750 SR acting as a Broadband Subscriber Aggregator (BSA).

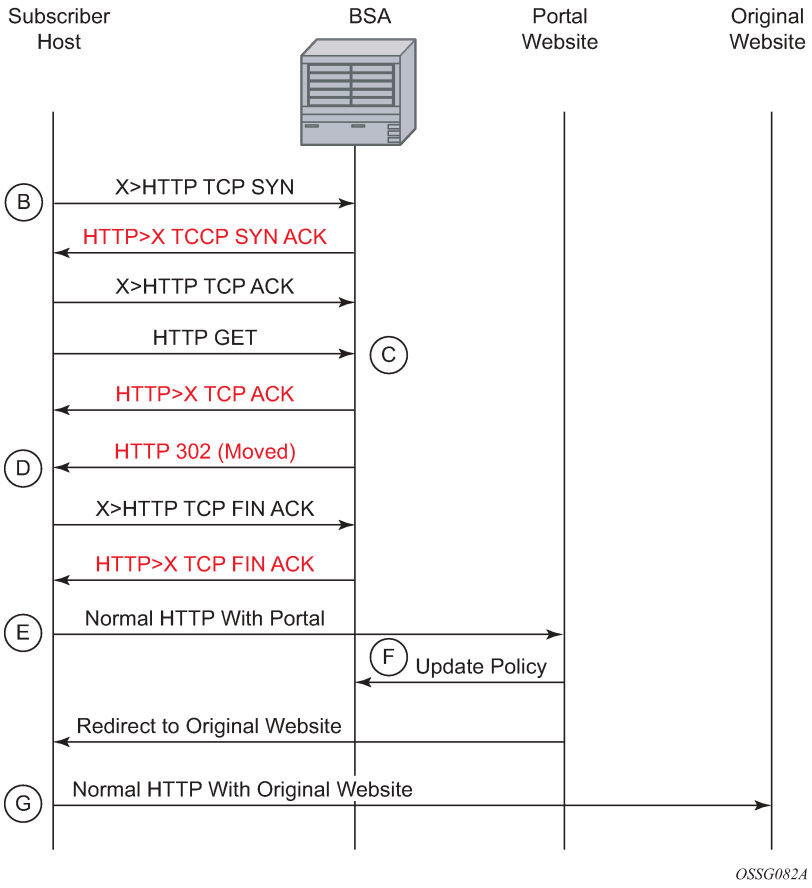
In a Triple Play network it may not be wanted (or feasible) to perform manual provisioning of new services and service changes. The ideal way of working is automatic provisioning, with the end-user supplying his details at a retailer, or (if physical connectivity is already present through an on-line customer portal).

The 7450 ESS and 7750 SR support a special ACL that automatically redirects subscribers to a predefined URL. This is done by sending a HTTP 302 (moved) message to the subscriber, regardless of the requested URL.

The message flow is as follows (see Figure 47: IP illustration of message flow in web portal redirect below):

1. The subscriber gets an IP address using DHCP (if the customer is trying to use a static IP he is blocked by the anti-spoofing filter).
2. The subscriber tries to connect to a website (TCP SYN, TCP ACK, HTTP GET).
3. The 7450 ESS or 7750 SR intercepts the HTTP GET request and discards it.
4. The 7450 ESS or 7750 SR then responds to the subscriber with a HTTP 302 message (service temporarily unavailable or moved), containing a new target URL (that of the portal) configured by the operator. This target URL can include the subscriber’s IP and MAC addresses as part of the portal’s URL string.
5. The subscriber’s web browser closes the original TCP connection and opens a new connection to the web portal where the subscriber can sign up or change his/her service Profile.
6. After approving the changes, the web portal updates the ACL (directly or through another system such as the Nokia 5750 SSC) to remove the redirection policy.
7. The subscriber can now connect to the original site.

Figure 47: IP illustration of message flow in web portal redirect



The items in red text in Figure 47: IP illustration of message flow in web portal redirect are messages the 7450 ESS or 7750 SR sends (masquerading as the destination), regardless of the destination IP address or type of service.

The following displays information that can optionally be added as variables in the portal URL (http-redirect url):

- $IP

  Customer's IP address
- $MAC –

  Customer's MAC address
- $URL

  Original requested URL
- $SAP

  Customer's SAP
- $SUB

  Customer's subscriber identification string
- $CID

The string that represents the circuit-id or interface-id of the subscriber host (hexadecimal format)

- $RID

  The string that represents the remote-id of the subscriber host (hexadecimal format)

Note:
The subscriber's IP and MAC address variables are populated from the anti-spoofing list, and therefore anti-spoofing must be enabled (see section Anti-spoofing filters).

Because most websites are accessed using the domain name, the 7450 ESS or 7750 SR needs to allow DNS queries, and an ACL entry to this effect should be included in the filter (see Configuring Web Portal Redirect).

7.2 Configuring Triple Play security with CLI

7.2.1 Common configuration tasks

7.2.1.1 Configuring anti-spoofing filters

Anti-spoofing filters are used to prevent malicious subscribers from sending IP packets with a forged IP or MAC address, and therefore mis-directing traffic. The anti-spoofing filter is populated from the DHCP lease state table, and DHCP snooping must be enabled on the SAP.

There are three types of filters (MAC, IP, and IP+MAC). One type is allowed per SAP.

The following displays an IES service interface configuration with anti-spoofing.

```
A:ALA-48>config>service>ies# info
----------------------------------------------
            interface "test123" create
                address 10.10.42.41/24
                local-proxy-arp
                proxy-arp
                    policy-statement "ProxyARP"
                exit
                sap 1/1/7:0 create
                    anti-spoof ip
                exit
                arp-populate
                dhcp
                    lease-populate 1
                    no shutdown
                exit
            exit
            no shutdown
----------------------------------------------
A:ALA-48>config>service>ies#
```

7.2.1.2 Configuring Triple Play security features

7.2.1.2.1 Configuring MAC pinning

The following example displays a partial BSA configuration with MAC pinning enabled on a SAP.

```
A:ALA-48>config>service# info
----------------------------------------------
vpls 800 customer 6001 create
    description "VPLS with residential split horizon for DSL"
    stp
        shutdown
    exit
    sap 2/1/4:100 split-horizon-group "DSL-group2" create
        description "SAP for RSHG"
               mac-pinning
    exit
    no shutdown
----------------------------------------------
A:ALA-48>config>service#
```

7.2.1.2.2 Configuring MAC protection

7.2.1.2.2.1 Preventing access by residential subscribers using protected (gateway) MAC addresses

The first step is to create a list of MAC addresses to be protected. The second step is to prevent access using these source addresses inside an SHG or a SAP.

The following example displays a partial BSA configuration with some protected MAC addresses on any SAP created inside the SHG.

```
A:ALA-48>config>service# info
----------------------------------------------
vpls 800 customer 6001 create
    no shutdown
    split-horizon-group "mygroup" create
        restrict-protected-src
    exit
    description "VPLS with residential split horizon for DSL"
    mac-protect
        mac 00:00:17:FE:82:D8
        mac 93:33:00:00:BF:92
    exit
----------------------------------------------
A:ALA-48>config>service#
```

7.2.1.2.2.2 Restricting access by residential subscribers to a small list of upstream MAC addresses

The first step is to create a list of MAC addresses to be protected. The second step is to restrict access to these addresses only from an SHG or a SAP (if the MAC address of an upstream server is not known, it can be discovered using, for example, the CPE ping OAM tool).

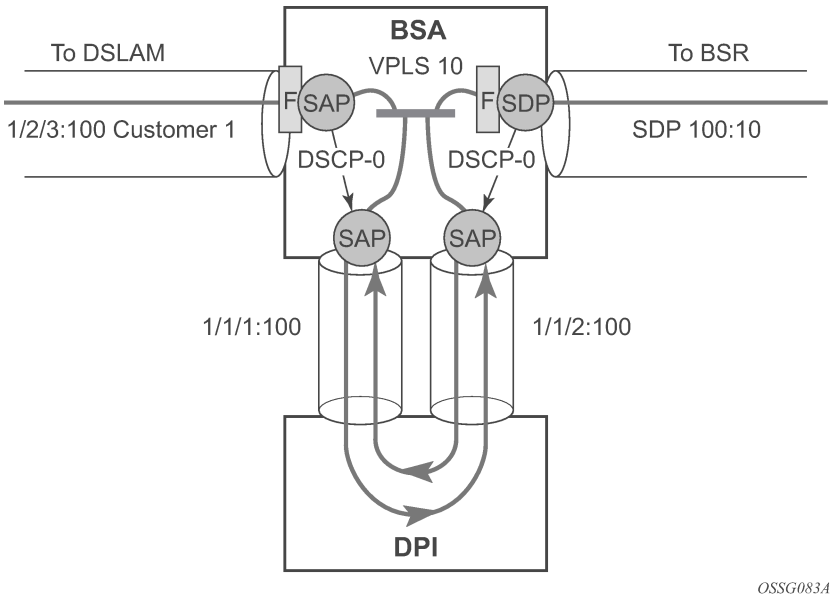
The following example displays a partial BSA configuration with restricted access to some MAC addresses from a specified SAP (an unrestricted access from any other SAP within the VPLS).

```
A:ALA-48>config>service# info
----------------------------------------------
vpls 800 customer 6001 create
    no shutdown
    description "VPLS with restricted access on a SAP"
    mac-protect
        mac 00:00:17:FE:82:D8
        mac 93:33:00:00:BF:92
    exit
    sap 1/1/4:30 create
        restrict-unprotected-dst
    exit
----------------------------------------------
A:ALA-48>config>service#
```

7.2.1.2.3 Configuring a VPLS redirect policy

Figure 48: VPLS redirect policy example displays an IP filter entry configuration for VPLS redirect policy.

Figure 48: VPLS redirect policy example



Information about defining and applying IP and MAC filters is described in the *7450 ESS, 7750 SR, 7950 XRS, and VSR Router Configuration Guide*.

7.2.1.2.3.1 Creating the filter

The following displays a redirect filter entry:

```
A:ALA-A>config>filter# info
----------------------------------------------
    ip-filter 10
        default-action forward
        entry 10
            match
                dscp be
            exit
            action forward next-hop sap 1/1/1:100
            exit
        exit
    exit
    ip-filter 11
        default-action forward
        entry 10
            match
                dscp be
            exit
                dscp be
            exit
            action forward next-hop sap 1/1/2:100
        exit
    exit
----------------------------------------------
A:ALA-A>config>filter#
```

7.2.1.2.3.2 Applying the filter to a VPLS service

The following displays how the redirection filter configured above is assigned to the ingress SAP from the DSLAM, and the ingress SDP from the BSR:

```
A:ALA-A>config>service>vpls# info
----------------------------------------------
    vpls 10 customer 1 create
        description "vpls10"
            sap 1/2/3:100 create
                ingress ip filter 10
            exit
            sap 1/1/1:100 create
            exit
            sap 1/1/2:100 create
            exit
            mesh-sdp 100:10 create
                ingress ip filter 11
            exit
        exit
    exit
----------------------------------------------
A:ALA-A>config>service>vpls#
```

7.2.1.3 Configuring ARP handling

7.2.1.3.1 Configuring proxy ARP

The implementation of proxy ARP with support for local proxy ARP allows the 7450 ESS or 7750 SR to respond to ARP requests in the subnet assigned to an IES or VPRN interface.

Configuring this command allows multiple customers to share the same IP subnet.

The following example displays an IES proxy ARP configuration:

```
A:ALA-48>config>service>ies# info
----------------------------------------------
            interface "test123" create
                address 10.10.42.41/24
                local-proxy-arp
                proxy-arp-policy "ProxyARP"
                exit
            exit
            no shutdown
----------------------------------------------
A:ALA-48>config>service>ies#
```

7.2.1.3.2 Configuring local proxy ARP

When local proxy ARP is enabled on an IP interface, the 7450 ESS or 7750 SR responds to all ARP requests for IP addresses belonging to the subnet with its own MAC address, and forwards all traffic between hosts in that subnet. Local proxy ARP is disabled by default.

Note: When **local-proxy-arp** is enabled under a IES or VPRN service, all ICMP redirects on the ports associated with the service are automatically blocked. This prevents users from learning each other's MAC address (from ICMP redirects).

The following example displays a local proxy ARP IES configuration:

```
A:ALA-A>config>service>ies# info
----------------------------------------------
        interface "test" create
            shutdown
            address 10.10.36.2/24
            local-proxy-arp
        exit
----------------------------------------------
A:ALA-A>config>service>ies#
```

7.2.1.3.3 Configuring ARP reply agent in a VPLS service

When ARP reply agent is enabled, the 7450 ESS or 7750 SR responds to ARP requests from the network, with information from the DHCP lease state table.

In the upstream direction (toward the network), the ARP reply agent intercepts ARP requests on subscriber SAPs, and checks them against the DHCP lease state table. The purpose is to prevent a malicious

subscriber spoofing ARP request or ARP reply messages and therefore populating the upstream router's ARP table with incorrect entries.

The following example displays a partial BSA configuration with ARP Reply Agent enabled on a SAP:

```
A:ALA-48>config>service# info
----------------------------------------------
...
    vpls 800 customer 6001 create
        description "VPLS with ARP Reply Agent active"
        sap 2/1/4:100 split-horizon-group "DSL-group2" create
    arp-reply-agent sub-ident
        exit
        sap 3/1/4:200 split-horizon-group "DSL-group2" create
    arp-reply-agent sub-ident
        exit
        no shutdown
...
----------------------------------------------
A:ALA-48>config>service#
```

7.2.1.3.4 Configuring remote proxy ARP

The following example displays the IES configuration to enable remote proxy ARP:

```
A:ALA-49>config>service>ies# info
----------------------------------------------
            interface "test-1A" create
                address 10.10.26.3/24
                remote-proxy-arp
            exit
            no shutdown
----------------------------------------------
A:ALA-49>config>service>ies#
```

7.2.1.3.5 Configuring automatic ARP table population in an IES or VPRN interface

The following example displays the IES DHCP configuration to enable automatic population of the ARP table using snooped DHCP information about an IES or VPRN (VPRN is supported on the 7750 SR only) interface:

```
A:ALA-1>config>service>ies>if# info
----------------------------------------------
                arp-populate
                dhcp
                    description "snooping_only"
                    lease-populate 1
                    no shutdown
                exit
----------------------------------------------
A:ALA-1>config>service>ies>if#


A:ALA-1>config>service>vprn>if# info
----------------------------------------------
                dhcp
                    description "test"
```

```
                    lease-populate 1
                    no shutdown
                exit
-----------------------------------------
A:ALA-1>config>service>ies>if#
```

7.2.1.3.6 Configuring CPU protection

CPU Protection can be used to protect the SR OS in subscriber management scenarios. See the *7450 ESS, 7750 SR, 7950 XRS, and VSR System Management Guide* for information about CPU Protection operation and configuration.

7.2.1.4 Configuring web portal redirect

The generic CLI structure for defining and applying IP and MAC filters is described in the *7450 ESS, 7750 SR, 7950 XRS, and VSR Router Configuration Guide*.

The following example displays an IP filter entry configuration for web-portal redirect:

```
A:ALA-A>config>filter# info
-----------------------------------------
    ip-filter 10 create
        description "filter to forward DNS and web traffic to my portal; redirect al
l other web traffic to the portal and drop everything else"
        default-action drop
        entry 10 create
            description "allows DNS traffic"
                match protocol 17
                dst-port 53
            exit
            action forward
        exit
        entry 20 create
            description "allows web traffic destined to portal (IP address 10.0.0.1)
"
            match protocol 6
                dst-port eq 80
                dst-ip 10.0.0.1
            exit
            action forward
        exit
        entry 30
            description "redirects all web traffic to portal"
            match protocol 6
                dst-port eq 80
            exit
            action http-redirect http://www.myportal.com/defaultportal/
login.cgi?ip=$IP&mac=$MAC&orig_url=$URL&usb=$SUB
        exit
    exit
-----------------------------------------
A:ALA-A>config>filter#
```

- Filter entry 10 in the example output allows the customer to access DNS to get the IP address of the original website they are trying to view.
- Entry 20 allows HTTP packets destined for the captive portal itself to be forwarded.

Note: The actual IP address (a.b.c.d) must be entered, not the DNS name (“www.myportal.com”). The IP address can be easily resolved from the 7450 ESS or 7750 SR CLI using the **ping** command.

- Entry 30 (which is the last option that does not drop the customer packets) checks for HTTP protocol and then starts the redirection process:
  - The 7450 ESS or 7750 SR intercepts the HTTP GET from the subscriber and respond with an HTTP 302 (temporarily moved) with the URL configured in the filter entry. This URL can contain some variables, notably the customer IP and MAC addresses to allow the portal to create an entry for the customer. The original requested URL is also included to redirect the client site back to the original requested site when the process is done.
  - The client then closes the connection with the original IP address and open a connection to the redirected server. Entry 20 allows this connection.

The following displays how the redirection filter configured above is assigned to an ingress SAP:

```
A:ALA-A>config>service>vpls# info
----------------------------------------------
    vpls 3 customer 6 create
        description "VPLS with web portal redirection filter applied"
        sap 2/1/5:0 create
            ingress
                filter ip 10
            exit
        exit
        no shutdown
    exit
----------------------------------------------
A:ALA-A>config>service>vpls#
```

8 Triple Play multicast

8.1 Introduction to multicast

IP multicast provides an effective method of many-to-many communication. Delivering unicast datagrams is simple. Normally, IP packets are sent from a single source to a single recipient. The source inserts the address of the target host in the IP header destination field of an IP datagram, intermediate routers (if present) simply forward the datagram toward the target in accordance with their respective routing tables.

Sometimes distribution needs individual IP packets be delivered to multiple destinations (like audio or video streaming broadcasts). Multicast is a method of distributing datagrams sourced from one (or possibly more) hosts to a set of receivers that may be distributed over different (sub) networks. This makes delivery of multicast datagrams significantly more complex.

Multicast sources can send a single copy of data using a single address for the entire group of recipients. The routers between the source and recipients route the data using the group address route. Multicast packets are delivered to a multicast group. A multicast group specifies a set of recipients who are interested in a data stream and is represented by an IP address from a specified range. Data addressed to the IP address is forwarded to the members of the group. A source host sends data to a multicast group by specifying the multicast group address in datagram's destination IP address. A source does not have to register to send data to a group nor do they need to be a member of the group.

Routers and Layer 3 switches use the Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) to manage membership for a multicast session. When a host wants to receive one or more multicast sessions, it sends a join message for each multicast group it wants to join. When a host wants to leave a multicast group, it sends a leave message.

8.2 Multicast in the broadband service router

This section describes the multicast protocols employed when a Nokia router is used as a Broadband Service Router (BSR) in a Triple Play aggregation network.

The protocols used are:

- Internet Group Management Protocol (Internet Group Management Protocol)
- Multicast Listener Discovery (Multicast Listener Discovery)
- Source Specific Multicast Groups (Internet Group Management Protocol)
- Protocol Independent Multicast (Sparse Mode) (Protocol Independent Multicast Sparse Mode)

8.2.1 Internet Group Management Protocol

Internet Group Management Protocol (IGMP) is used by IPv4 hosts and routers to report their IP multicast group memberships to neighboring multicast routers. A multicast router keeps a list of multicast group memberships for each attached network, and a timer for each membership.

Multicast group memberships include at least one member of a multicast group on a given attached network, not a list of all the members. With respect to each of its attached networks, a multicast router can assume one of two roles, querier or non-querier. There is normally only one querier per physical network.

A querier issues two types of queries, a general query and a group-specific query. General queries are issued to solicit membership information about any multicast group. Group-specific queries are issued when a router receives a leave message from the node it perceives as the last group member remaining on that network segment.

Hosts wanting to receive a multicast session issue a multicast group membership report. These reports must be sent to all multicast enabled routers.

8.2.1.1 IGMP versions and interoperability requirements

If routers run different versions of IGMP, they negotiates the lowest common version of IGMP that is supported on their subnet and operate in that version.

Version 1

Specified in RFC 1112, *Host extensions for IP Multicasting*, was the first widely deployed version and the first version to become an Internet standard.

Version 2

Specified in RFC 2236, *Internet Group Management Protocol*, added support for low leave latency, that is, a reduction in the time it takes for a multicast router to learn that there are no longer any members of a group present on an attached network.

Version 3

Specified in RFC 3376, *Internet Group Management Protocol*, adds support for source filtering, that is, the ability for a system to report interest in receiving packets only from specific source addresses, as required to support Source-Specific Multicast (See Source Specific Multicast (SSM)), or from all but specific source addresses, sent to a multicast address.

IGMPv3 must keep state per group per attached network. This group state consists of a **filter-mode**, a list of sources, and various timers. For each attached network running IGMP, a multicast router records the necessary reception state for that network.

8.2.1.2 IGMP version transition

Nokia's SRs are capable of interoperating with routers and hosts running IGMPv1, IGMPv2, or IGMPv3. *Draft-ietf-magma-igmpv3-and-routing-0x.txt* explores some of the interoperability issues and how they affect the various routing protocols.

IGMP version 3 specifies that if at any point a router receives an older version query message on an interface that it must immediately switch into a compatibility mode with that earlier version. Because none of the previous versions of IGMP are source aware, should this occur and the interface switch to Version 1 or 2 compatibility mode, any previously learned group memberships with specific sources (learned from the IGMPv3 specific INCLUDE or EXCLUDE mechanisms) must be converted to non-source specific group memberships. The routing protocol then treats this as if there is no EXCLUDE definition present.

8.2.2 Multicast Listener Discovery

Multicast Listener Discovery (MLD) is used by IPv6 hosts and routers to report their IP multicast group memberships to neighboring multicast routers. A multicast router keeps a list of multicast group memberships for each attached network, and a timer for each membership. Multicast group memberships include at least one member of a multicast group on a specific attached network, not a list of all the members. With respect to each of its attached networks, a multicast router can assume one of two roles, querier or non-querier. There is normally only one querier per physical network.

A querier issues two types of queries, a general query and a group-specific query. General queries are issued to solicit membership information about any multicast group. Group-specific queries are issued when a router receives a leave message from the node it perceives as the last group member remaining on that network segment.

Hosts wanting to receive a multicast session issue a multicast group membership report. These reports must be sent to all multicast enabled routers.

8.2.2.1 MLD versions and interoperability requirements

If routers run different versions of MLD, they negotiates the lowest common version of MLD that is supported on their subnet and operate in that version.

Version 1

Specified in RFC 2710, *Multicast Listener Discovery (MLD) for IPv6*, was the first deployed version and included low leave latency, that is, a reduction in the time it takes for a multicast router to learn that there are no longer any members of a group present on an attached network.

Version 2

Specified in RFC 3810, *Multicast Listener Discovery Version 2 (MLDv2) for IPv6*, adds support for source filtering, that is, the ability for a system to report interest in receiving packets only from specific source addresses, as required to support Source-Specific Multicast.

Multicast (SSM)), or from all but specific source addresses, sent to a multicast address. MLDv2 must keep state per group per attached network. This group state consists of a filter mode, a list of sources, and various timers. For each attached network running MLD, a multicast router records the wanted reception state for that network.

8.2.2.2 Source-specific multicast groups

IGMPv3and MLDv2 allows a receiver to join a group and specify that it only wants to receive traffic for a group if that traffic comes from a specific source. If a receiver does this, and no other receiver on the LAN requires all the traffic for the group, then the Designated Router (DR) can omit performing a (*,G) join to set up the shared tree, and instead issue a source-specific (S,G) join only.

For IPv4, the range of multicast addresses from 232.0.0.0 to 232.255.255.255 is currently set aside for source-specific multicast. For groups in this range, receivers should only issue source specific IGMPv3 joins. If a PIM router receives a non-source-specific join for a group in this range, it should ignore it.

For IPv6, the multicast prefix FF3x::/32 is currently set aside for source-specific multicast. For groups in this range, receivers should only issue source specific MLDv2 joins. If a PIM router receives a non-source-specific join for a group in this range, it should ignore it.

A Nokia PIM router must silently ignore a received (*, G) PIM join message where G is a multicast group address from the multicast address group range that has been explicitly configured for SSM. This occurrence should generate an event. If configured, the IGMPv2 (MLDv1for IPv6) request can be translated into IGMPv3 (MLDv2 for IPv6). The SR allows for the conversion of an IGMPv2 (*,G) request into a IGMPv3 (S,G) request based on manual entries. A maximum of 32 SSM ranges is supported.

IGMPv3 and MLDv2 also allows a receiver to join a group and specify that it only wants to receive traffic for a group if that traffic does not come from a specific source or sources. In this case, the DR performs a (*,G) join as normal, but can combine this with a prune for each of the sources the receiver does not want to receive.

8.2.3 Protocol Independent Multicast Sparse Mode

Protocol Independent Multicast Sparse Mode (PIM-SM) leverages the unicast routing protocols that are used to create the unicast routing table: OSPF, IS-IS, BGP, and static routes. Because PIM uses this unicast routing information to perform the multicast forwarding function it is effectively IP protocol independent. Unlike DVMRP, PIM does not send multicast routing tables updates to its neighbors.

PIM-SM uses the unicast routing table to perform the Reverse Path Forwarding (RPF) check function instead of building up a completely independent multicast routing table.

PIM-SM only forwards data to network segments with active receivers that have explicitly requested the multicast group. PIM-SM in the ASM model initially uses a shared tree to distribute information about active sources. Depending on the configuration options, the traffic can remain on the shared tree or switch over to an optimized source distribution tree. As multicast traffic starts to flow down the shared tree, routers along the path determine if there is a better path to the source. If a more direct path exists, then the router closest to the receiver sends a join message toward the source and then reroutes the traffic along this path.

As stated above, PIM-SM relies on an underlying topology-gathering protocol to populate a routing table with routes. This routing table is called the Multicast Routing Information Base (MRIB). The routes in this table can be taken directly from the unicast routing table, or it can be different and provided by a separate routing protocol such as MBGP. Regardless of how it is created, the primary role of the MRIB in the PIM-SM protocol is to provide the next hop router along a multicast-capable path to each destination subnet. The MRIB is used to determine the next hop neighbor to whom any PIM join/prune message is sent. Data flows along the reverse path of the join messages. Thus, in contrast to the unicast RIB that specifies the next hop that a data packet would take to get to some subnet, the MRIB gives reverse-path information, and indicates the path that a multicast data packet would take from its origin subnet to the router that has the MRIB.

8.2.4 Ingress multicast Path Management (IMPM) enhancements

See the *Advanced Configuration Guide* for more information about IMPM as well as detailed configuration examples.

IMPM allows the system to manage Layer 2 and Layer 3 IP multicast flows by sorting them into the available multicast paths through the switch fabric. The ingress multicast manager tracks the amount of available multicast bandwidth per path and the amount of bandwidth used per IP multicast stream. The following traffic is managed by IMPM when enabled:

- IPv4 and IPv6 routed multicast traffic
- VPLS IGMP snooping traffic
- VPLS PIM snooping for IPv4 traffic

- VPLS PIM snooping for IPv6 traffic (only when sg-based forwarding is configured)

Two policies define how each path should be managed, the bandwidth policy, and how multicast channels compete for the available bandwidth, the multicast information policy.

Chassis multicast planes should not be confused with IOM/IMM multicast paths. The IOM/IMM uses multicast paths to reach multicast planes on the switch fabric. An IOM/IMM may have less or more multicast paths than the number of multicast planes available in the chassis.

Each IOM/IMM multicast path is either a primary or secondary path type. The path type indicates the multicast scheduling priority within the switch fabric. Multicast flows sent on primary paths are scheduled at multicast high priority while secondary paths are associated with multicast low priority.

The system determines the number of primary and secondary paths from each IOM/IMM forwarding plane and distributes them as equally as possible between the available switch fabric multicast planes. Each multicast plane may terminate multiple paths of both the primary and secondary types.

The system ingress multicast management module evaluates the ingress multicast flows from each ingress forwarding plane and determines the best multicast path for the flow. A specific path can be used until the terminating multicast plane is "maxed" out (based on the rate limit defined in the **per-mcast-plane-capacity** commands) at which time either flows are moved to other paths or potentially blackholed (flows with the lowest preference are dropped first). In this way, the system makes the best use of the available multicast capacity without congesting individual multicast planes.

The switch fabric is simultaneously handling both unicast and multicast flows. The switch fabric uses a weighted scheduling scheme between multicast high, unicast high, multicast low and unicast low when deciding which cell to forward to the egress forwarding plane next. The weighted mechanism allows some amount of unicast and lower priority multicast (secondary) to drain on the egress switch fabric links used by each multicast plane. The amount is variable based on the number of switch fabric planes available on the amount of traffic attempting to use the fabric planes. The per-mcast-plane-capacity commands allows the amount of managed multicast traffic to be tuned to compensate for the expected available egress multicast bandwidth per multicast plane. In conditions where it is highly desirable to prevent multicast plane congestion, the per-mcast-plane-capacity commands should be used to compensate for the non-multicast or secondary multicast switch fabric traffic.

8.3 Multicast in the BSA

IP Multicast is normally not a function of the Broadband Service Aggregator (BSA) in a Triple Play aggregation network being a Layer 2 device. However, the BSA does use IGMP snooping to optimize bandwidth utilization.

8.3.1 IGMP snooping

For most Layer 2 switches, multicast traffic is treated like an unknown MAC address or broadcast frame, which causes the incoming frame to be flooded out (broadcast) on every port within a VLAN. While this is acceptable behavior for unknowns and broadcasts, as IP Multicast hosts may join and be interested in only specific multicast groups, all this flooded traffic results in wasted bandwidth on network segments and end stations.

IGMP snooping entails using information in layer 3 protocol headers of multicast control messages to determine the processing at layer 2. By doing so, an IGMP snooping switch provides the benefit of conserving bandwidth on those segments of the network where no node has expressed interest in receiving packets addressed to the group address.

On the 7450 ESS and 7750 SR, IGMP snooping can be enabled in the context of VPLS services. The IGMP snooping feature allows for optimization of the multicast data flow for a group within a service to only those Service Access Points (SAPs) and Service Distribution Points (SDPs) that are members of the group. In fact, the 7450 ESS and 7750 SR implementation performs more than pure snooping of IGMP data, because it also summarizes upstream IGMP reports and responds to downstream queries.

The 7450 ESS and 7750 SR maintain several multicast databases:

- A port database on each SAP and SDP lists the multicast groups that are active on this SAP or SDP.
- All port databases are compiled into a central proxy database. Towards the multicast routers, summarized group membership reports are sent based on the information in the proxy database.
- The information in the different port databases is also used to compile the multicast forwarding information base (MFIB). This contains the active SAPs and SDPs for every combination of source router and group address (S,G), and is used for the actual multicast replication and forwarding.

When the router receives a join report from a host for a multicast group, it adds the group to the port database and (if it is a new group) to the proxy database. It also adds the SAP or SDP to existing (S,G) in the MFIB, or builds a new MFIB entry.

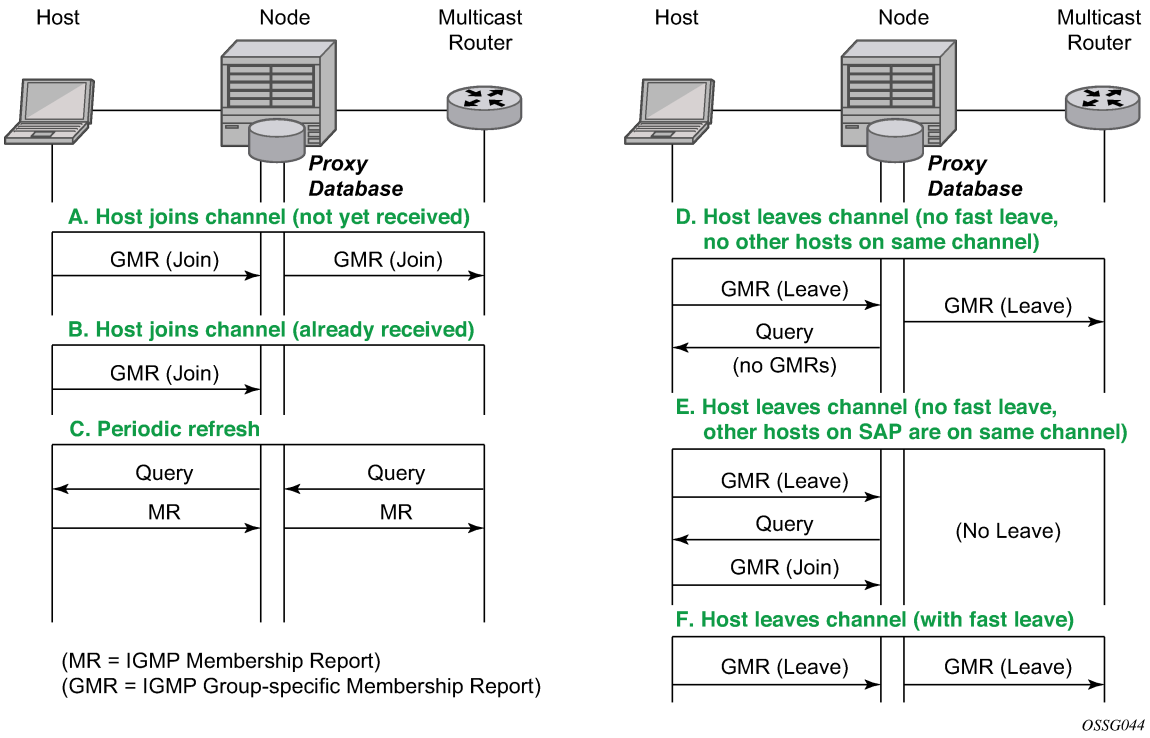
When the router receives a leave report from a host, it first checks if other devices on the SAP or SDP still want to receive the group (unless fast leave is enabled). Then it removes the group from the port database, and from the proxy database if it was the only receiver of the group. The router also deletes entries if it does not receive periodic membership confirmations from the hosts.

The fast leave feature finds its use in multicast TV delivery systems, for example. Fast Leave speeds up the membership leave process by terminating the multicast session immediately, instead of the standard procedure of issuing a group specific query to check if other group members are present on the SAP or SDP.

8.3.1.1 IGMP/MLD message processing

Figure 49: IGMP/MLD message processing illustrates the basic IGMP message processing in several situations.

Figure 49: IGMP/MLD message processing



8.3.1.2 IGMP message processing

Scenario A: A host joins a multicast group (TV channel) which is not yet being received by other hosts on the router, and therefore, is not yet present in the proxy database. The 7450 ESS or 7750 SR adds the group to the proxy database and sends a new IGMP Join group-specific membership report upstream to the multicast router.

Scenario B: A host joins a channel which is already being received by one or more hosts on the router, and is already present in the proxy database. No upstream IGMP report is generated by the router.

Scenario C: The multicast router periodically sends IGMP queries to the router, requesting it to respond with generic membership reports. Upon receiving such a query, the router compiles a report from its proxy database and send it back to the multicast router.

In addition, the router floods the received IGMP query to all hosts (on SAPs and spoke SDPs), and updates its proxy database based on the membership reports received back.

Scenario D: A host leaves a channel by sending an IGMP leave message. If fast-leave is not enabled, the router first checks whether there are other hosts on the same SAP or spoke SDP by sending a query. If no other host responds, the 7450 ESS or 7750 SR removes the channel from the SAP. In addition, if there are no other SAPs or spoke SDPs with hosts subscribing to the same channel, the channel is removed from the proxy database and an IGMP leave report is sent to the upstream Multicast Router.

Scenario E: A host leaves a channel by sending an IGMP leave message. If fast-leave is not enabled, the router checks whether there are other hosts on the same SAP or spoke SDP by sending a query. Another device on the same SAP or spoke SDP still wants to receive the channel and responds with a membership report. The router does not remove the channel from the SAP.

Scenario F: A host leaves a channel by sending an IGMP leave report. Fast-leave is enabled, so the router does not check whether there are other hosts on the same SAP or spoke SDP but immediately removes the group from the SAP. In addition, if there are no other SAPs or spoke SDPs with hosts subscribing to the same group, the group is removed from the proxy database and an IGMP leave report is sent to the upstream multicast router.

8.3.1.3 MLD message processing

MLD message processing differs from IGMP. An IPv6 host can have two WAN IPv6 addresses and an IPv6 prefix. MLD messages source address are link local addresses. This makes it difficult to know if the originating host is a WAN host or a PD host. By default, all requested IPv6 (s,g) are first associated with a WAN host. If this WAN host disconnects or ends its IPv6 session, the (s,g) is then associated with the remaining WAN host. If there are no more WAN hosts, the (s,g) is then associated with the remaining PD host. The (s,g) is always transferred to the remaining IPv6 host until there are no more report replies to corresponding to the queries. Scenarios A — F do not differ for IPv6 hosts.

8.3.1.4 IGMP/MLD filtering

A provider may want to block receive or transmit permission to individual hosts or a range of hosts. To this end, the 7450 ESS and 7750 SR support IGMP/MLD filtering. Two types of filter can be defined:

- Filter IGMP/MLD membership reports from a host or range of hosts. This is performed by importing an appropriately defined routing policy into the SAP or spoke SDP.
- Filter to prevent a host from transmitting multicast streams into the network. The operator can define a data-plane filter (ACL) which drops all multicast traffic, and apply this filter to a SAP or spoke SDP.

8.3.2 Multicast VPLS Registration (MVR)

Multicast VPLS Registration (MVR) is a bandwidth optimization method for multicast in a broadband services network. MVR allows a subscriber on a port to subscribe and unsubscribe to a multicast stream on one or more network-wide multicast VPLS instances.

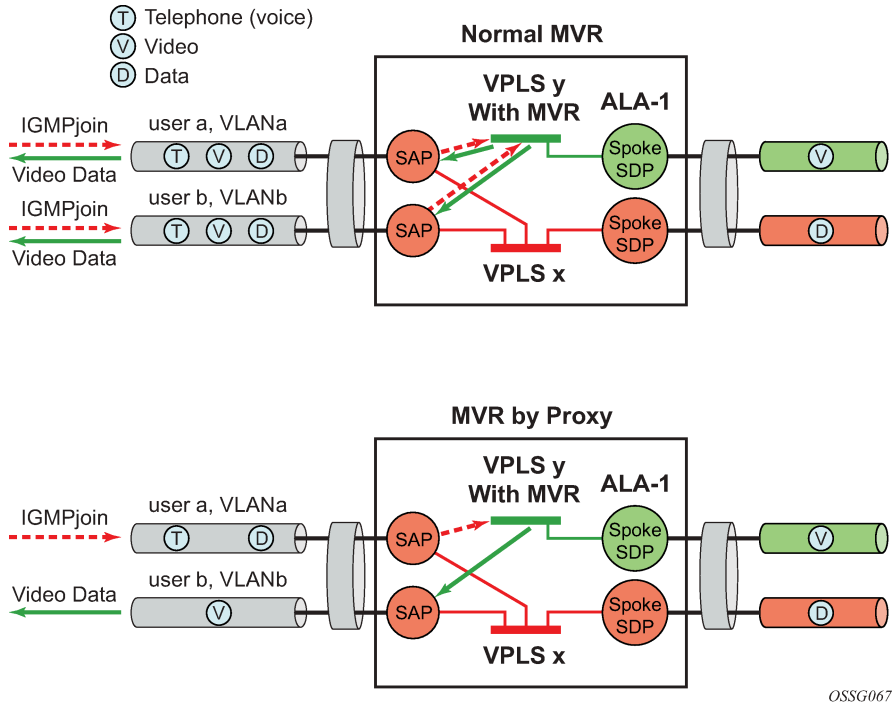
MVR assumes that subscribers join and leave multicast streams by sending IGMP join and leave messages. The IGMP leave and join message are sent inside the VPLS to which the subscriber port is assigned. The multicast VPLS is shared in the network while the subscribers remain in separate VPLS services. Using MVR, users on different VPLS cannot exchange any information between them, but still multicast services are provided.

On the MVR VPLS, IGMP snooping must be enabled. On the user VPLS, IGMP snooping and MVR work independently. If IGMP snooping and MVR are both enabled, MVR reacts only to join and leave messages from multicast groups configured under MVR. Join and leave messages from all other multicast groups are managed by IGMP snooping in the local VPLS. This way, potentially several MVR VPLS instances could be configured, each with its own set of multicast channels.

MVR by proxy — In some situations, the multicast traffic should not be copied from the MVR VPLS to the SAP on which the IGMP message was received (standard MVR behavior) but to another SAP. This is called MVR by proxy.

Figure 50: MVR and MVR by proxy shows a MVR and MVR by proxy configuration.

Figure 50: MVR and MVR by proxy



8.3.3 Layer 3 multicast load balancing

Layer 3 multicast load balancing establishes a more efficient distribution of Layer 3 multicast data over ECMP and LAG links. Operators have the option to redistribute multicast groups over ECMP or LAG links if the number of links changes either up or down.

When implementing this feature, there are several considerations. When multicast load balancing is not configured, the distribution remains as is. Multicast load balancing is based on the number of "s,g" groups. This means that bandwidth considerations are not considered. The multicast groups are distributed over the available links as joins are processed. When link failure occurs, the load is distributed on the failed channel to the remaining channels so multicast groups are evenly distributed over the remaining links. When a link is added (or failed link returned) all multicast joins on the added links are allocated until a balance is achieved.

When multicast load balancing is configured, but the channels are not found in the **multicast-info-policy**, then multicast load balancing is based on the number of "s,g" groups. This means that bandwidth considerations are not considered. The multicast groups are distributed over the available links as joins are processed. The multicast groups are evenly distributed over the remaining links. When link failure occurs, the load is distributed on the failed channel to the remaining channels. When a link is added (or failed link returned) all multicast joins on the added links are allocated until a balance is achieved. A manual redistribute command enables the operator to re-evaluate the current balance and, if required, move channels to different links to achieve a balance. A timed redistribute parameter allows the system to automatically, at regular intervals, redistribute multicast groups over available links. If no links have been added or removed from the ECMP/LAG interface, then no redistribution is attempted.

When multicast load balancing is configured, multicast groups are distributed over the available links as joins are processed based on bandwidth configured for the specified group address. If the bandwidth is not configured for the multicast stream then the configured default value is used.

If link failure occurs, the load is distributed on the failed channel to the remaining channels. The bandwidth required over each individual link is evenly distributed over the remaining links.

When an additional link is available for a specific multicast stream, then it is considered in all multicast stream additions applied to the interface. This multicast stream is included in the next scheduled automatic rebalance run. A rebalance run re-evaluates the current balance with regard to the bandwidth utilization and if required, move multicast streams to different links to achieve a balance.

A rebalance, either timed or executing the **mc-ecmp-rebalance** command, should be administered gradually to minimize the effect of the rebalancing process on the different multicast streams. If multicast re-balancing is disabled and subsequently enabled, keeping with the rebalance process, the gradual and least invasive method is used to minimize the effect of the changes to the customer.

By default multicast load balancing over ECMP links is enabled and set at 30 minutes.

The rebalance process can be executed as a low priority background task while control of the console is returned to the operator. When multicast load rebalancing is not enabled, then ECMP changes are not be optimized, however, when a link is added occurs an attempt is made to balance the number of multicast streams on the available ECMP links. This however may not result in balanced utilization of ECMP links.

Only a single **mc-ecmp-rebalance** command can be executed any specific time, if a rebalance is in progress and the command is entered, it is rejected with the message saying that a rebalance is already in progress. A low priority event is generated when an actual change for a specific multicast stream occurs as a result of the rebalance process.

8.3.4 IGMP state reporter

The target application for this feature is linear TV delivery. In some countries, wholesale Service Providers are obligated by the government regulation to provide information about channel viewership per subscriber to retailers.

A service provider (wholesaler or retailer) my use this information for:

- billing purposes
- market research/data mining to gain view into the most frequently watched channels, duration of the channel viewing, frequency of channel zapping by the time of the day, and so on

The information about channel viewership is based on IGMP states maintained per each subscriber host. Each event related to the IGMP state creation is recorded and formatted by the IGMP process. The formatted event is then sent to another task in the system (Exporter), which allocates a TX buffer and start a timer.

The event is then be written by the Exporter into the buffer. The buffer in essence corresponds to the packet that contains a single event or a set of events. Those events are transported as data records over UDP transport to an external collector node. The packet itself has a header followed by a set of TLV type data structures, each describing a unique filed within the IGMP event.

The packet is transmitted when it reaches a preconfigured size (1400 bytes), or when the timer expires, whichever comes first.

Note: The timer started when the buffer was initially created.

The receiving end (collector node) accepts the data on the destination UDP port. It must be aware of the data format so that it can interpret incoming data accordingly. The implementation details of the receiving node are outside of the scope of this description and are left to the network operator.

The IGMP state recording per subscriber host must be supported for hosts which are replicating multicast traffic directly as well as for those host that are only keeping track of IGMP states for the HQoS Adjustment purpose. The latter is implemented by redirection and not the Host Tracking (HT) feature as originally proposed. The IGMP reporting must differentiate events between direct replication and redirection.

It further distinguish events that are related to denial of IGMP state creation (because of filters, MCAC failure, and so on) and the ones that are related to removal of an already existing IGMP state in the system.

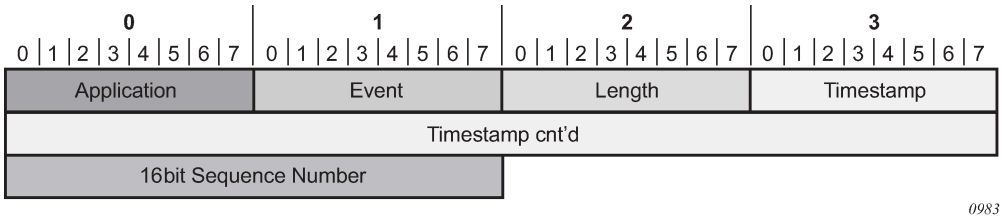
8.3.4.1 IGMP data records

Each IGMP state change generates a data record that is formatted by the IGMP task and written into the buffer. IGMP state transitions configured statically through CLI are not reported.

To minimize the size of the records when transported over the network, most fields in the data record are HEX coded (as opposed to ASCII descriptive strings).

Each data record has a common header as shown in Figure 51: Common IGMP data record header:

Figure 51: Common IGMP data record header



Application:

- 0x01 - IGMP
- 0x02 - IGMP Host Tracking Event:

Event:

- Related to denial of a new state creation:
  - **0x01**

    Join
  - **0x02 (Join_Deny_Filter)**

    Join denied because of filtering by an import policy
  - **0x03 (Join_Deny_CAC)**

    Join denied because of MCAC
  - **0x04 (Join_Deny_MaxGrps)**

    Join denied because of maximum groups per host limit reached
  - **0x05 (Join_Deny_MaxSrcs)**

    Join denied because of maximum sources limit reached

  - **0x06 Join (Join_Deny_SysErr)**

    Join denied because of an internal error (for example: out of memory)

- Related to removal of an existing IGMP state:
  - **0x07**
    (Drop_Leave_Rx) IGMP state is removed because of the Leave message
  - **0x08**

    (Drop_Expiry) IGMP state is removed because of the time out (by default 2*query_interval + query_response_interval = 260sec)
  - **0x09**

    (Drop_Filter) IGMP state is removed because of the filter (import policy) change
  - **0x0A**

    (Drop_CfgChange) IGMP state is removed because of a configuration change (clear grp, intf shutdown, PPPoE session goes unexpectedly down)
  - **0x0B**

    (Drop_CAC) an existing stream is stopped because of a configuration change in MCAC

Length: The length of the entire data record (including the header and TLVs) in octets.

16 bit Sequence Number

- Because IGMP Reporting is based on connectionless transport (UDP), a 16 bit sequence numbers are used in each data record so that data loss in the network can be tracked.
- The 16 bit sequence number is located after the time stamp field. The sequence numbers increases sequentially from 0 to 65535 and then rolls over back to 0.

Timestamp: Timestamp is in UNIX format (32 bit integer in seconds because 01/01/1970) plus an extra 8 bits for 10msec resolution.

TLVs describing the IGMP state record has the following structure shown in Figure 52: Data record field TLV structure and the data record field descriptions in Table 6: Data record field description:

Figure 52: Data record field TLV structure

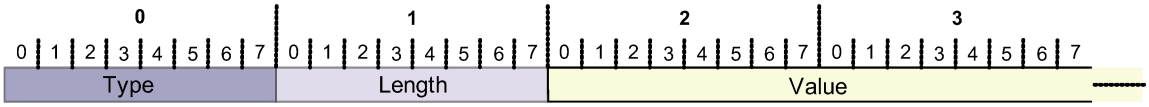


Table 6: Data record field description

Type	Description	Encoding/length	Mandatory/optional
0x02	Subscriber ID	ASCII	M
0x03	Sub Host IP	4 Bytes IPv4	M
0x04	Mcast Group IP	4 Bytes IPv4	M
0x05	Mcast Source IP	4 Bytes IPv4	M

Type	Description	Encoding/length	Mandatory/optional
0x06	Host MAC	6 Bytes	M
0x07	PPPoE Session-ID	2 Bytes	M
0x08	Service ID	4 Bytes	M
0x09	SAP ID	ASCII	M
0x0A	Redirection vRtrId	4 Bytes	M
0x0B	Redirection ifIndex	4 Bytes	M

The **redirection destination** TLV is a mandatory TLV that is sent only in cases where redirection is enabled. It contains two 32 bit integer numbers. The first number identifies the VRF where IGMPs are redirected; the second number identifies the interface index.

Optional fields can be included in the data records according to the configuration.

In IGMPv3, if an IGMP message (Join or Leave) contains multiple multicast groups or a multicast group contains multiple IP sources, only a single event is generated per group-source combination. In other words, data records are transmitted with a single source IP address and multiple mcast group addresses or a single multicast group address with multiple source IP addresses, depending on the content of the IGMP message. (*,G)

8.3.4.2 Transport mechanism

Data is transported by a UDP socket. The destination IP address, the destination port and the source IP address are configurable. The default UDP source and destination port number is 1037.

Upon the arrival of an IGMP event, the Exporter allocates a buffer for the packet (if not already allocated) and starts writing the events into the buffer (packet). Along with the initial buffer creation, a timer is started. The trigger for the transmission of the packet is either the TX buffer being filled up to 1400B (hard coded value), or the timer expiry, whichever comes first.

The source IP address is configurable within GRT (by default system IP), and the destination IP address can be reachable only by GRT. The source IP address is modified in the **system**>**security**>**source-address**>**application** CLI hierarchy.

The receiving end (the collector node) collects the data and process them according to the formatting rules defined in this document. The capturing and processing of the data on the collector node is outside of the context of this description.

The processing node should have sufficient resources to accept and process packets that contain information about every IGMP state change for every host from a set of network BRASes that are transporting data to this collector node.

Multicast Reporter traffic is marked as BE (all 6 DSCP bits are set to 0) exiting our system.

8.3.4.3 HA compliance

IGMP events are synchronized between two CPMs before they are transported out of the system.

8.3.4.4 QoS awareness

IGMP Reporter is a client of sgt-qos so that DSCP or dot1p bits can be appropriately be marked when egressing the system.

8.3.4.5 IGMP reporting restrictions

The following are not supported:

- Regular (non-subscriber) interfaces
- SAM support as the collector device

8.4 Multicast support over subscriber interfaces in Routed CO model

Applications for multicast over subscriber interfaces in Routed CO ESM model can be divided in two main categories:

Residential customers where the driver applications are:

- IPTV in an environment with legacy non-multicasting DSLAMs
- Internet multicast where users connect to a multicast stream sourced from the Internet.

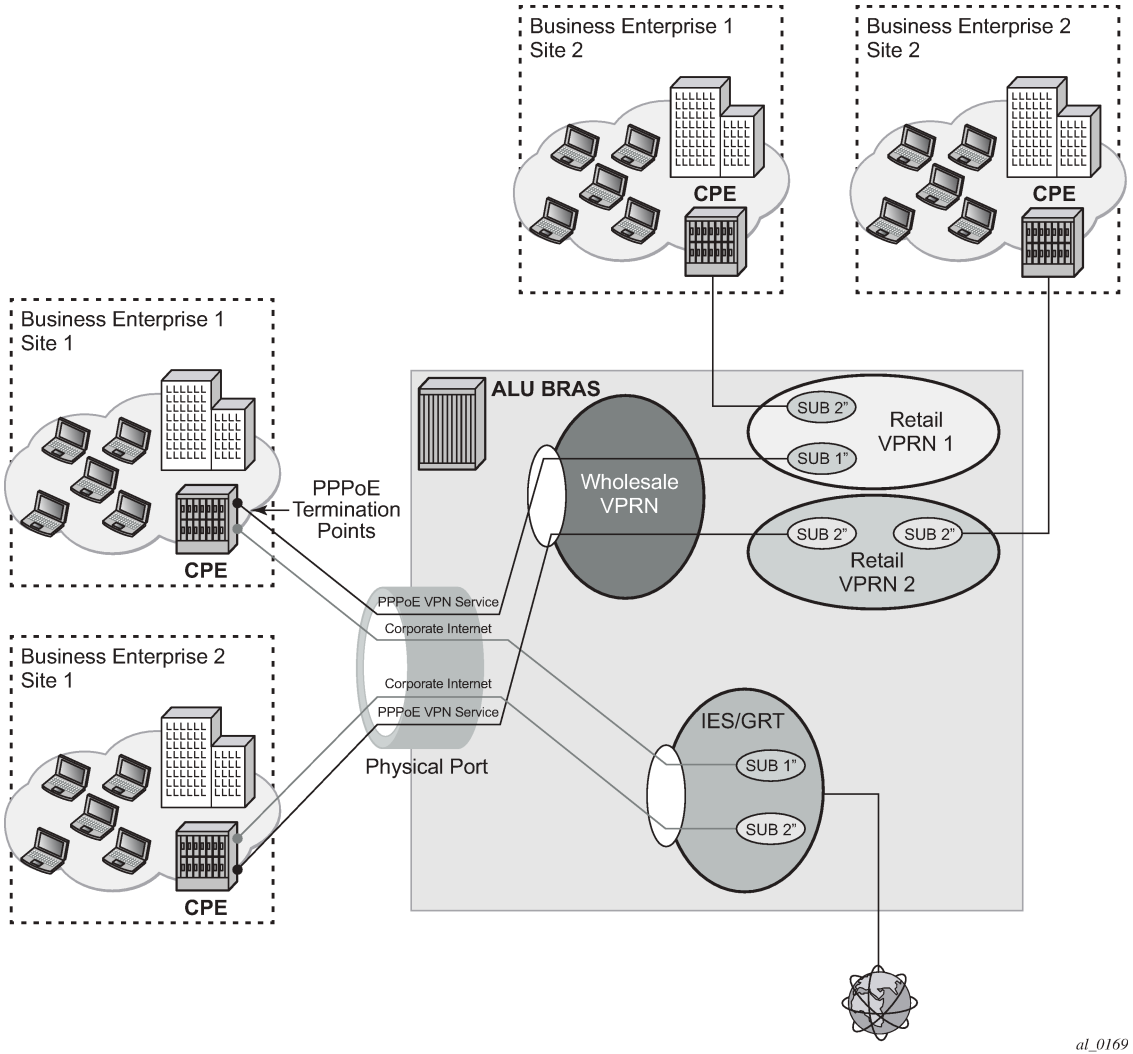
For the business customers, the main drivers are enterprise multicast and Internet multicast applications.

On multicast-capable ANs, a single copy of each multicast stream is delivered over a separate regular IP interface. AN would then perform the replication. This is how multicast would be deployed in Routed CO environment with the 7750 SR and 7450 ESS.

On legacy, non-multicast ANs, or in environments with low volume multicast traffic where it is not worth setting up a separate multicast topology (from BNG to AN), multicast replication is performed with subscriber-interfaces in the 7750 SR and 7450 ESS. There are differences in replicating multicast traffic on IPoE vs PPPoX which are described in subsequent sessions.

An example of a business connectivity model is shown in Figure 53: Typical business connectivity model.

Figure 53: Typical business connectivity model



In this example, HSI is terminated in a Global Routing Table (GRT) whereas VPRN services are terminated in Wholesale/Retail VPRN method, with each customer using a separate VPRN.

The actual connectivity model that is deployed depends on many operational aspects that are present in the customer environment.

Multicast over subscriber-interfaces in a Routed CO model is supported for both types of hosts, IPoE and PPPoE which can be simultaneously enabled on a shared SAP.

There are some fundamental differences in multicast behavior between two host types (IPoE and PPPoX). The differences are discussed further in the next sections.

8.4.1 Multicast over IPoE

There are several deployment scenarios for delivering multicast directly over subscriber hosts:

- 1:1 model (subscriber per VLAN/SAP) with the Access Node (AN) that is not IGMP/MLD aware.

- N:1 model (service per VLAN/SAP) with the AN in the Snooping mode.
- N:1 model with the AN in the Proxy mode.
- N:1 model with the AN that is not IGMP/MLD aware.

There are two modes of operation for subscriber multicast that can be chosen to address the above mentioned deployment scenarios:

1. Per SAP replication — A single multicast stream per group is forwarded on any given SAP. Even if the SAP has a multicast group (channel) that is registered to multiple hosts, only a single copy of the multicast stream is forwarded over this SAP. The multicast stream has a multicast destination MAC address (as opposed to unicast). IGMP/MLD states are maintained per host. This is the default mode of operation.
2. Per subscriber host replication in this mode of operation, multicast is replicated per subscriber host even if this means that multiple copies of the same stream that are forwarded over the same SAP. For example, if two hosts on the same SAP are registered to receive the same multicast group (channel), then this multicast channel is replicated twice on the same SAP. The streams have a unique unicast destination MAC address (otherwise it would not make sense to replicate the streams twice).

In all deployment scenarios and modes of operation the IGMP/MLD states per source IP address of the incoming IGMP/MLD message is maintained. This source IP address might represent a subscriber hosts or the AN (proxy mode).

For MLD, the source IP address is the host link local address. Therefore, the MLD message is associated with all IP address/prefix of the host.

8.4.1.1 Per SAP replication mode

In the per SAP replication mode a single copy of the multicast channel is forwarded per SAP. In other words, if a subscriber (in 1:1 mode) or a group of subscribers (in N:1 mode) have multiple hosts and all of them are subscribed to the same multicast group (watching the same channel), then only a single copy of the multicast stream for that group is sent. The destination MAC address is always a multicast MAC (there are no conversion to unicast MAC address).

IGMP/MLD states are maintained per subscriber host and per SAP.

8.4.1.1.1 Per SAP queue

Multicast traffic over subscribers in a per-SAP replication mode flows through a SAP queue which is outside of the subscriber queues context. Sending the multicast traffic over the default SAP queue is characterized by:

- the inability to classify multicast traffic into separate subscriber queues and therefore include it natively in the subscriber HQoS hierarchy; however, multicast traffic can be classified into specific (M-)SAP queues, assuming that such queues are enabled by (M-)SAP based QoS policy
- redirection of multicast traffic by internal queues in case the SAP queue in subscriber environment is disabled (**sub-sla-mgmt**>**single-sub-parameters**>**profiled-traffic-only**); this is applicable only to 1:1 subscriber model
- a possible necessity for HQoS Adjustment as multicast traffic is flowing outside of the subscriber queues
- de-coupling of the multicast forwarding statistics from the overall subscriber forwarding statistics obtained by subscriber specific **show** commands

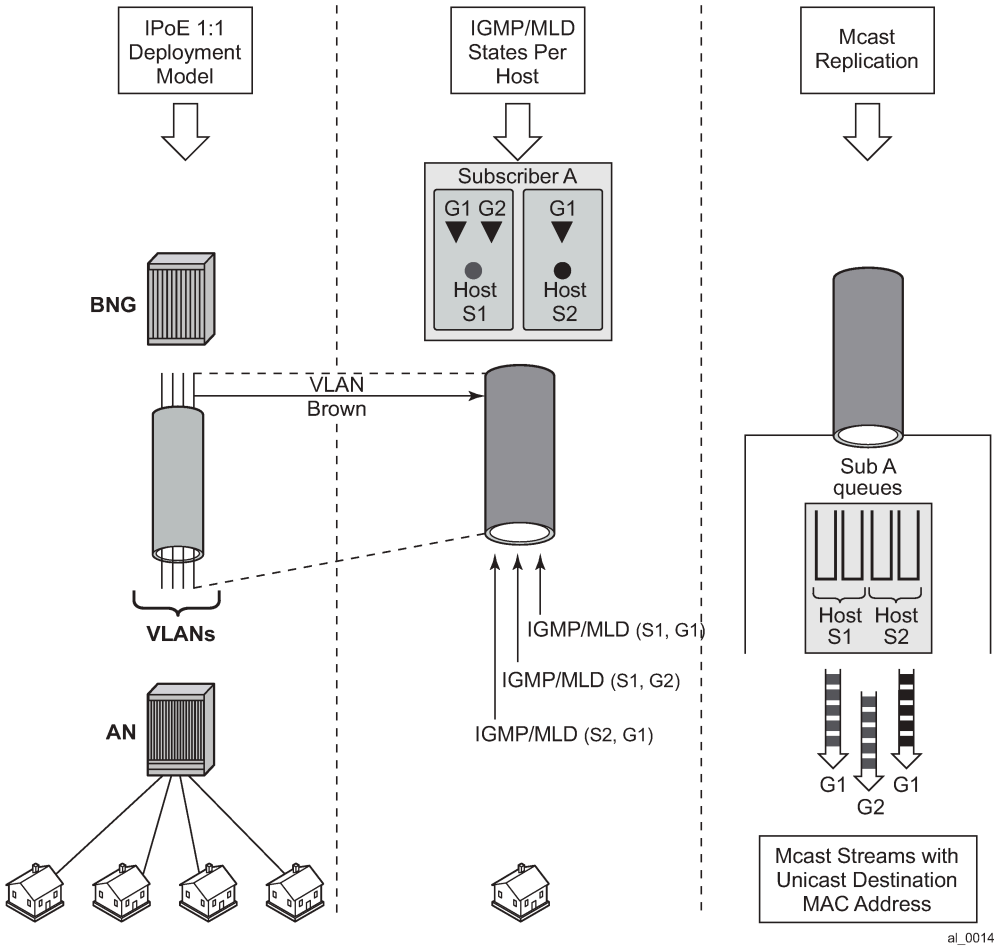
8.4.1.1.2 IPoE 1:1 model (subscriber per VLAN/SAP) — no IGMP/MLD in AN

The AN is not IGMP/MLD aware, all replications are performed in the BNG. From the BNG perspective this deployment model has the following characteristics:

- IGMP/MLD states are kept per hosts and SAPs. Each host can be registered to more than one group.
- MLD uses link-local as source address, which makes it difficult to associate with the originating host. MLD (S,G) are always first associated with a IPv6 WAN host if available. If this WAN host terminate its IPv6 session (by lease expiry, session terminate, and so on), the (S,G) is then associated with any remaining WAN host. Only when there are no more WAN hosts available is the (S,G) associated with the PD host, if any.
- IGMP/MLD Joins are accepted only from the active subscriber hosts as dictated by antispoofing.
- IGMP/MLD statistics can be displayed per host or per group.
- Multicast traffic for the subscriber is forwarded through the egress SAP queue. In case that the SAP queue is disabled (profiled-traffic-only command), multicast traffic flows through internal queues outside of the subscriber context.
- A single copy of any multicast stream is generated per SAP. This can be viewed as replication per unique multicast group per SAP, instead of the replication per host. In other words, the number of multicast streams on this SAP is equal to the number of unique groups across all hosts on this SAP (subscriber).
- Traffic statistics are kept per the SAP queue. Consequently multicast traffic stats are shown outside of the subscriber context.
- HQoS adjustment may be necessary.
- Traffic cannot be explicitly classified (forwarding classes and queue mappings) inside of the subscriber queues.
- Redirection to the common multicast VLAN (or Layer 3 interface) is supported.
- Multicast streams have multicast destination MAC.
- MLD uses link-local as source address, which makes it difficult to associate with the originating host. MLD (s,g) is associated with the IPv6 host. Therefore, if any WAN host or PD host end their IPv6 session (by lease expire, and so on), the (s,g) is associated with the remaining host address/prefix. The (s,g) is be delivered to the subscriber as long as a IPv6 address or prefix remains.

  This model is shown in Figure 54: 1:1 model.

Figure 54: 1:1 model



8.4.1.1.3 IPoE N:1 model (service per VLAN/SAP) — IGMP/MLD snooping in the AN

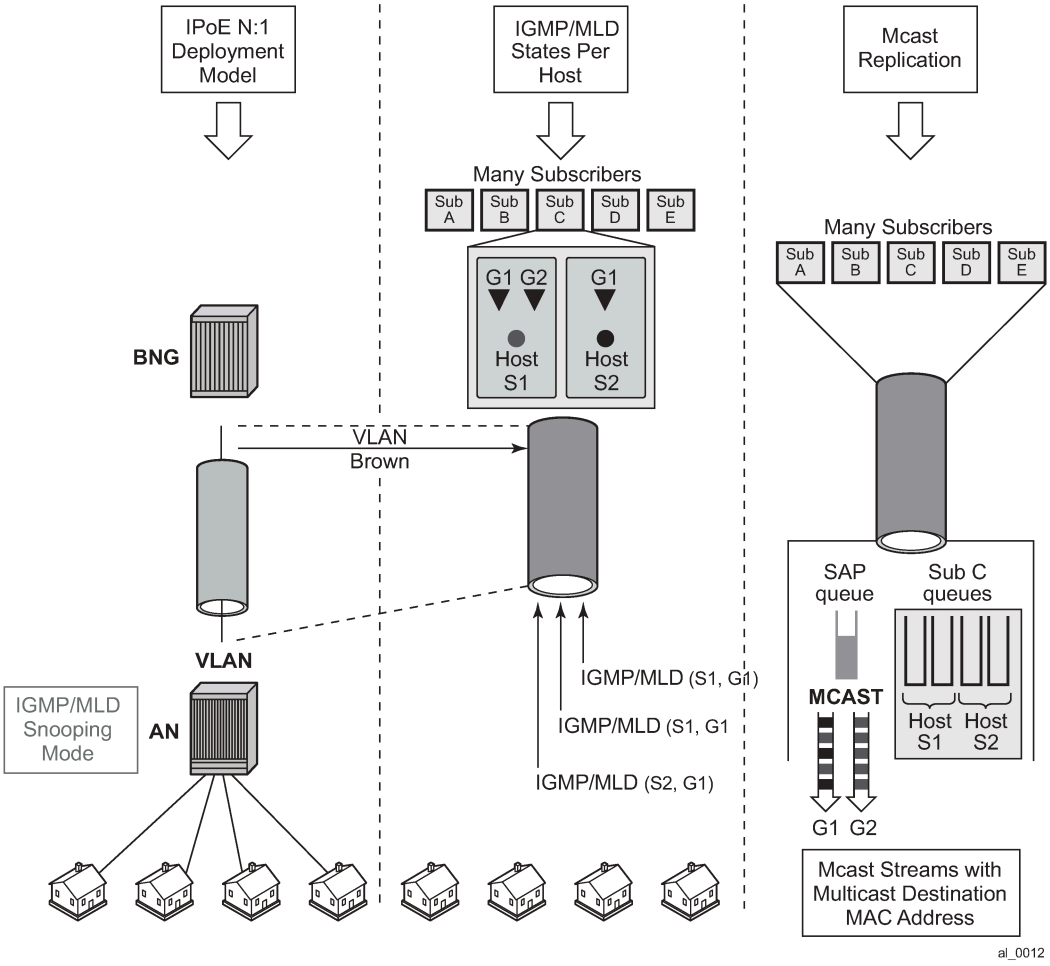
The AN is IGMP/MLD aware and is participating in multicast replication. From the BNG perspective this deployment model has the following characteristics:

- IGMP/MLD states are kept per hosts and SAPs. Each host can be registered to more than one group.
- IGMP/MLD Joins are accepted only from the active subscriber hosts as dictated by antispoofing.
- IGMP/MLD statistics are displayed per host, per group or per subscriber.
- Multicast traffic for all subscribers on this SAP is forwarded through the egress SAP queues.
- A single copy of any multicast stream is generated per SAP. This can be viewed as the replication per unique multicast group per SAP, instead of the replication per host or subscriber. In other words, the number of multicast streams on this SAP is equal to the number of unique groups across all hosts and subscribers on this SAP.

- The AN receives a single multicast stream and based on its own (AN) IGMP/MLD snooping information, it replicates the mcast stream to the appropriate subscribers.
- Traffic statistics are kept per the SAP queue. Consequently multicast traffic stats are shown on a per SAP basis (aggregate of all subscribers on this SAP).
- Traffic cannot be explicitly classified (forwarding classes and queue mappings) inside of the subscriber queues.
- Redirection to the common multicast VLAN is supported.
- Multicast streams have multicast destination MAC.
- IGMP Joins are accepted (src IP address) only for the sub hosts that are already created in the system. IGMP Joins coming from the hosts that are nonexistent in the system are rejected, unless this functionality is explicitly enabled by the sub-hosts-only command under the IGMP group-int CLI hierarchy level.
- MLD uses link-local as source address, which makes it difficult to associate with the originating host. MLD (S,G) are always first associated with a IPv6 WAN host, if available. If this WAN host terminate its IPv6 session (by lease expiry, session terminate, and so on), the (S,G) is then associated with any remaining WAN host. Only when there are no more WAN hosts available, is the (S,G) associated with the PD host, if any.
- MLD join are only accepted if it matches the subscriber host link local address. MLD Joins coming from the hosts that are nonexistent in the system is rejected, unless this functionality is explicitly enabled by the sub-hosts-only command under the MLD group-int CLI hierarchy level.

  This model is shown in Figure 55: N:1 model - AN in IGMP snooping mode.

Figure 55: N:1 model - AN in IGMP snooping mode



8.4.1.1.4 IPoE N:1 model (service per VLAN/SAP) — IGMP/MLD proxy in the AN

The AN is configured as IGMP/MLD Proxy node and is participating in downstream multicast replication.

For IPv4, IGMP messages from multiple sources (subscribers hosts) for the same multicast group are consolidated in the AN into a single IGMP messages. This single IGMP message has the source IP address of the AN.

For IPv6, MLD messages from multiple sources (subscribers hosts) for the same multicast group are consolidated in the AN into a single MLD messages. This single MLD message has the link-local IP address of the AN.

From the BNG perspective this deployment model has the following characteristics:

- Subscriber IGMP/MLD states are maintained in the AN.
- IGMP Joins are accepted from the source IP address that is different from any of the subscriber's IP addresses already existing in the BNG. This is controlled by an IGMP filter on a per group-interface

level assuming that the IGMP processing for subscriber hosts is disabled with the no **sub-hosts-only** command under the router/service **vprn**>**igmp**>**group-interface** CLI hierarchy. In this case all IGMP messages that cannot be related to existing hosts are treated in the context of the SAP while IGMP messages from the existing hosts are treated in the context of the subscriber hosts.

- MLD Joins are only accepted if the link-local address matches the subscriber' link local address. To allow processing of foreign link-local address such as the AN link local address, the MLD processing for subscriber hosts should be disabled with the no sub-hosts-only command under the **router/service vprn**>**mld**>**group-interface** CLI hierarchy. In this case all MLD messages that cannot be related to existing hosts are treated in the context of the SAP while MLD messages from the existing hosts are treated in the context of the subscriber hosts.
- IGMP/MLD statistics can be displayed per group-interface.
- Multicast traffic for all subscribers on this SAP is forwarded through the egress SAP queue.
- A single copy of any multicast stream is generated per SAP.
- The AN receives a single multicast stream. Based on the IGMP/MLD proxy information, the AN replicates the mcast stream to the appropriate subscribers.
- Traffic statistics are maintained per SAP queue.
- HQoS Adjustment is not useful because the per host/subscriber IGMP/MLD granularity is lost. IGMP/MLD states are aggregated per AN.
- Traffic can be explicitly classified into a specific SAP queues by a QoS policy applied under the SAP.
- Multicast streams have multicast destination MAC.

In the following example, IGMPs from the source IP address <ip> is accepted even though there is no subscriber-host with that IP addresses present in the system. An IGMP state is created under the SAP context (service per vlan, or N:1 model) for the **group** *pref-definition*. All other IGMP messages originated from non-subscriber hosts is rejected. IGMP messages for subscriber hosts are processed according to the IGMP policy applied to each subscriber host.

```
configure
     service vprn <id>
          igmp
               group-interface <ip-int-name>
                    import <policy-name>

configure
     router
          policy-options
               begin
                    prefix-list <pref-name>
                         prefix <pref-definition>

                    policy-statement proxy-policy
                entry 1
                    from
                          group-address <pref-name>
                            source-address <ip>
                            protocol igmp
                        exit
                   action accept
                   exit
                      exit
                      default-action reject
```

This functionality (accepting IGMP from non-subscriber hosts) can be disabled with the following flag.

```
configure
    service vprn <id>
        igmp
            group-interface <ip-int-name>
    sub-host-only
```

In this case, only per host IGMP processing is allowed.

In the following example, MLDs with foreign link-local-address is accepted even though there is no subscriber-host with that link local addresses present in the system. An MLD state is created under the SAP context (service per vlan, or N:1 model) for the **group** *pref-definition*. All other MLD messages originated from non-subscriber hosts are rejected. MLD messages for subscriber hosts are processed according to the IGMP policy applied to each subscriber host.

```
configure
     service vprn <id>
          mld
               group-interface <ip-int-name>
                    import <policy-name>

configure
     router
          policy-options
               begin
                    prefix-list <pref-name>
                         prefix <pref-definition>

                    policy-statement proxy-policy
                    entry 1
                        from
                              group-address <pref-name>
                            source-address <ip>
                            protocol igmp
                        exit
                       action accept
                       exit
                      exit
                      default-action reject
```

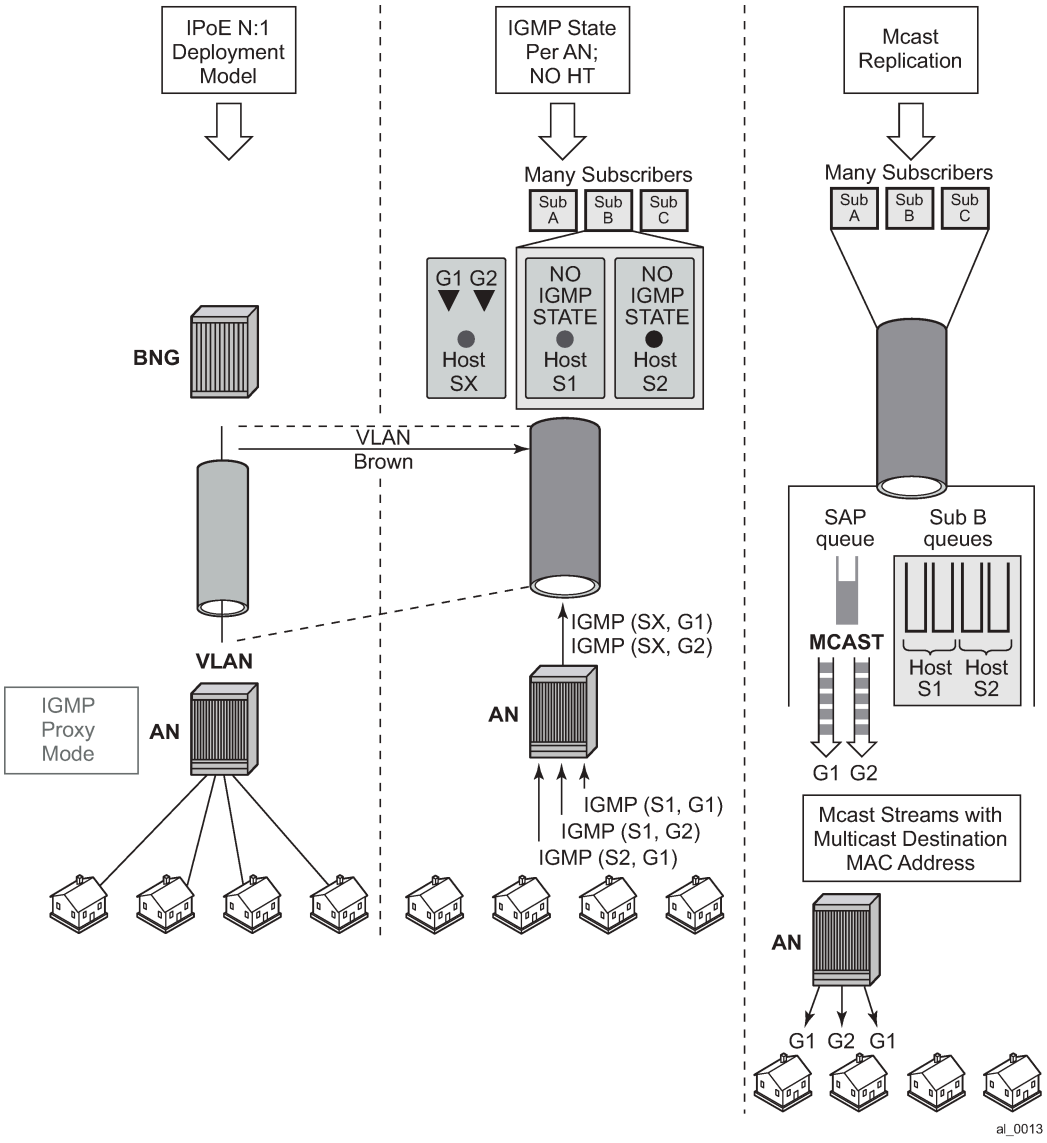
This functionality (accepting MLD from non-subscriber hosts) can be disabled with the following flag.

```
configure
    service vprn <id>
        mld
            group-interface <ip-int-name>
                sub-host-only
```

In this case, only per host MLD processing is allowed.

This model is shown in Figure 56: N:1 model - AN in proxy mode.

Figure 56: N:1 model - AN in proxy mode



8.4.1.2 Per subscriber host replication mode

In this mode a multicast stream is transmitted per subscriber hosts for each registered multicast group (channel). As a result, multiple copies of the same multicast stream destined for different destinations can be transmitted over the same SAP. In this case, traffic flows within the subscriber queues and consequently it is accounted in HQoS. As a result, HQoS Adjustment is not needed. Each copy of the same multicast stream have a unique unicast destination MAC addresses. The per host unicast MAC destination addresses are necessary to differentiate multiple copies between different receivers on the same SAP.

Per host replication mode can be enabled on a subscriber basis with the **per-host-replication** command in the **config**>**subscr-mgmt**>**igmp-policy** context.

For IPv6, the command is in the **config**>**subscr-mgmt**>**mld-policy** context.

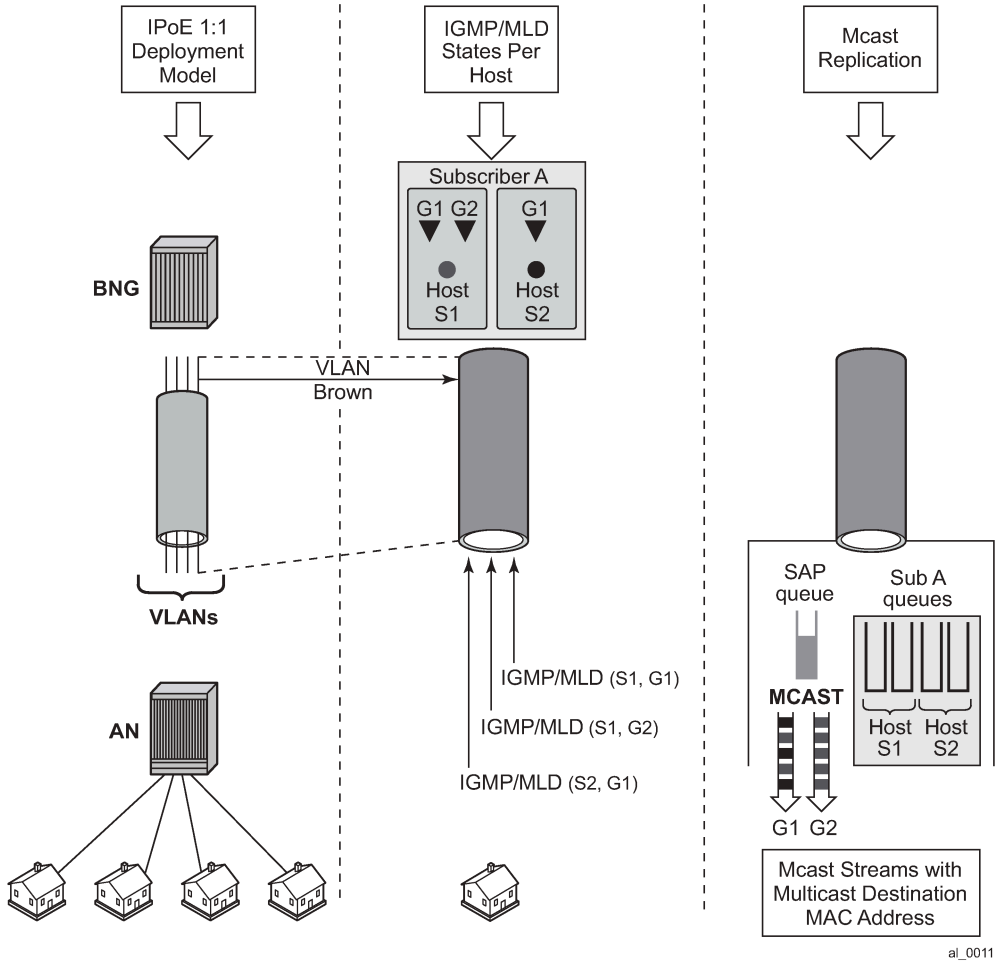
8.4.1.2.1 IPoE 1:1 model (subscriber per VLAN/SAP) — no IGMP/MLD in AN

The AN is not IGMP/MLD aware and multicast replication is performed in the BNG. Multicast streams are sent directly to the hosts using their unicast MAC addresses. HQoS adjustment is not needed as multicast traffic is flowing through subscriber queues. From the BNG perspective this deployment model has the following characteristics:

- IGMP/MLD states are kept per hosts. Each host can be registered to multiple IGMP/MLD groups.
- MLD uses link-local as source address, which makes it difficult to associate with the originating host. MLD (S,G) are always first associated with a IPv6 WAN host if available. If this WAN host terminate its IPv6 session (with lease expiry, session terminate, and so on), the (S,G) is then associated with any remaining WAN host. Only when there are no more WAN hosts available, are the (S,G) associated with the PD host, if any.
- IGMP/MLD Joins are accepted only from the active subscriber hosts. In other words antispoofing is in effect for IGMP/MLD messages.
- IGMP/MLD statistics can be displayed per host, per group or per subscriber.
- Multicast traffic is forwarded through subscriber queues using unicast destination MAC address of the destination host.
- Multiple copies of the same multicast stream can be generated per SAP. The number of copies depends on the number of hosts on the SAP that are registered to the same multicast group (channel). In other words, the number of multicast streams on the SAP is equal to the number of groups registered across all hosts on this SAP.
- Traffic statistics are kept per the host queue. In case that multicast statistics need to be separated from unicast, the multicast traffic should be classified in a subscriber separate queue.
- HQoS Adjustment is not needed as traffic is flowing within the subscriber queues and is automatically accounted in HQoS.
- Multicast traffic can be explicitly classified into forwarding classes and consequently directed into needed queues.
- MCAC is supported.
- profiled-traffic-only mode defined under sub-sla-mgmt is supported. This mode (profiled-traffic-only) is used to save the number of queues in 1:1 model (sub-sla-mgmt-> no multisub-SAP) by preventing the creation of the SAP queues. Because multicast traffic is not using the SAP queue, enabling this feature does not have any effect on the multicast operation.

  This model is shown in Figure 57: 1:1 model.

Figure 57: 1:1 model



8.4.1.2.2 IPoE N:1 model (service per VLAN/SAP) — no IGMP/MLD in the AN

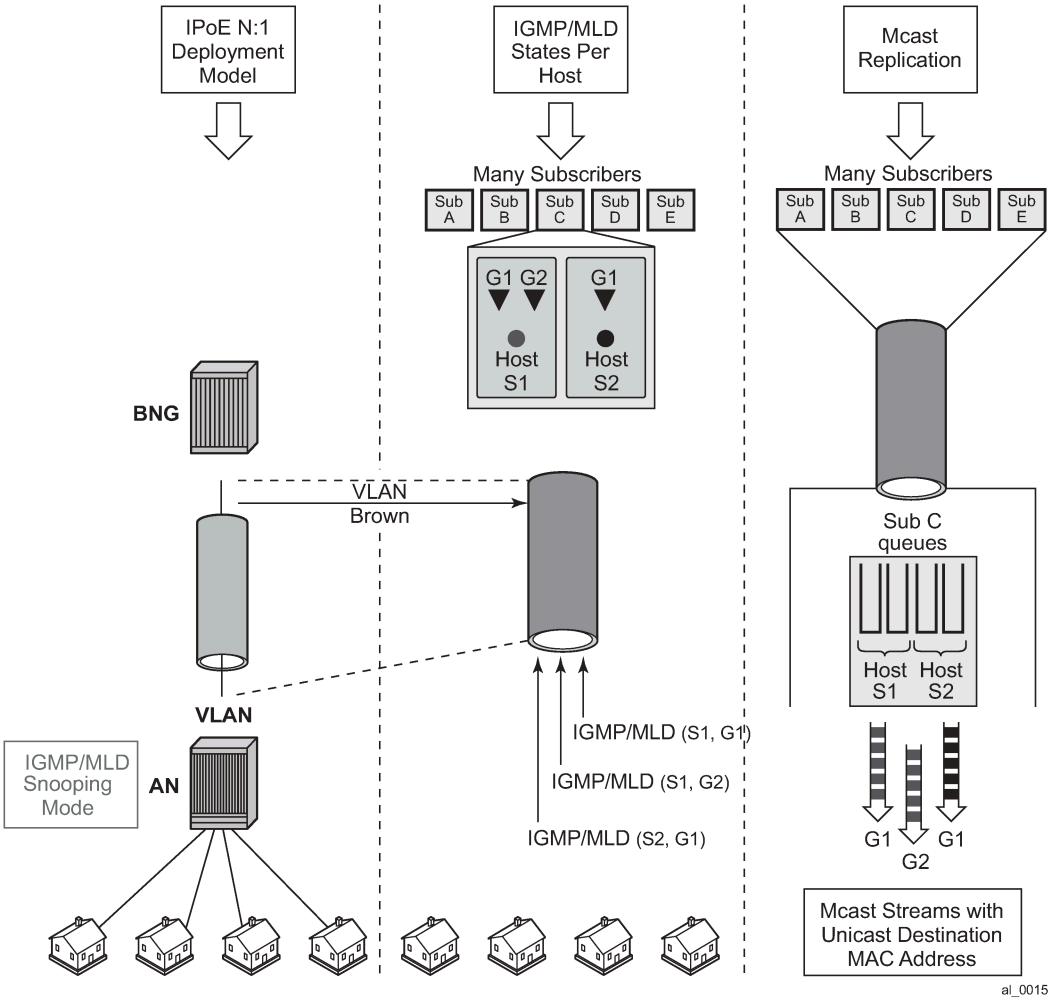
The AN is not IGMP/MLD aware and is not participating in multicast replication. From the BNG perspective this deployment model has the following characteristics:

- IGMP/MLD states are kept per hosts. Each host can be registered to multiple multicast groups.
- MLD uses link-local as source address, which makes it difficult to associate with the originating host. MLD (S,G) are always first associated with a IPv6 WAN host if available. If this WAN host terminate its IPv6 session (by lease expiry, session terminate, and so on), the (S,G) is then associated with any remaining WAN host. Only when there are no more WAN hosts available, are the (S,G) associated with the PD host, if any.
- IGMP/MLD Joins are accepted only from the active subscriber hosts, subject to antispoofing.
- IGMP/MLD statistics can be displayed per host, per group or per subscriber.

- Multicast traffic is forwarded through subscriber queues using unicast destination MAC address of the destination host.
- Multiple copies of the same multicast stream can be generated per SAP. The number of copies depends on the number of hosts on the SAP that are registered to the same multicast group (channel). In other words, the number of multicast streams on the SAP is equal to the number of groups registered across all hosts on this SAP.
- Traffic statistics are kept per the host queue. In case that multicast statistics need to be separated from unicast, the multicast traffic should be classified in a separate subscriber queue.
- HQoS Adjustment is not needed as traffic is flowing within the subscriber queues and is automatically accounted in HQoS.
- Multicast traffic can be explicitly classified into forwarding classes and consequently directed into needed queues.
- MCAC is supported.

  This model is shown in Figure 58: N:1 model — no IGMP/MLD in the AN.

Figure 58: N:1 model — no IGMP/MLD in the AN



8.4.1.3 Per-SLA profile instance replication mode

In the per-SLA Profile Instance (SPI) replication mode, a multicast stream is transmitted for the subscriber host SLA profile for each registered multicast group (channel). As a result, multiple copies of the same multicast stream are transmitted over the same SAP. The multicast packet replication depends on the number of SLA profile instances configured for the subscriber. For example, if the subscriber hosts use three SLA profiles, the (S,G) is replicated three times. But if the hosts use the same SLA profile, the (S,G) is replicated only once.

MCAC is supported for per-SPI replication.

Multicast traffic flows on the SAP queue and each copy of the multicast stream use the same multicast destination MAC address.

The per-SPI replication mode can be enabled per subscriber:

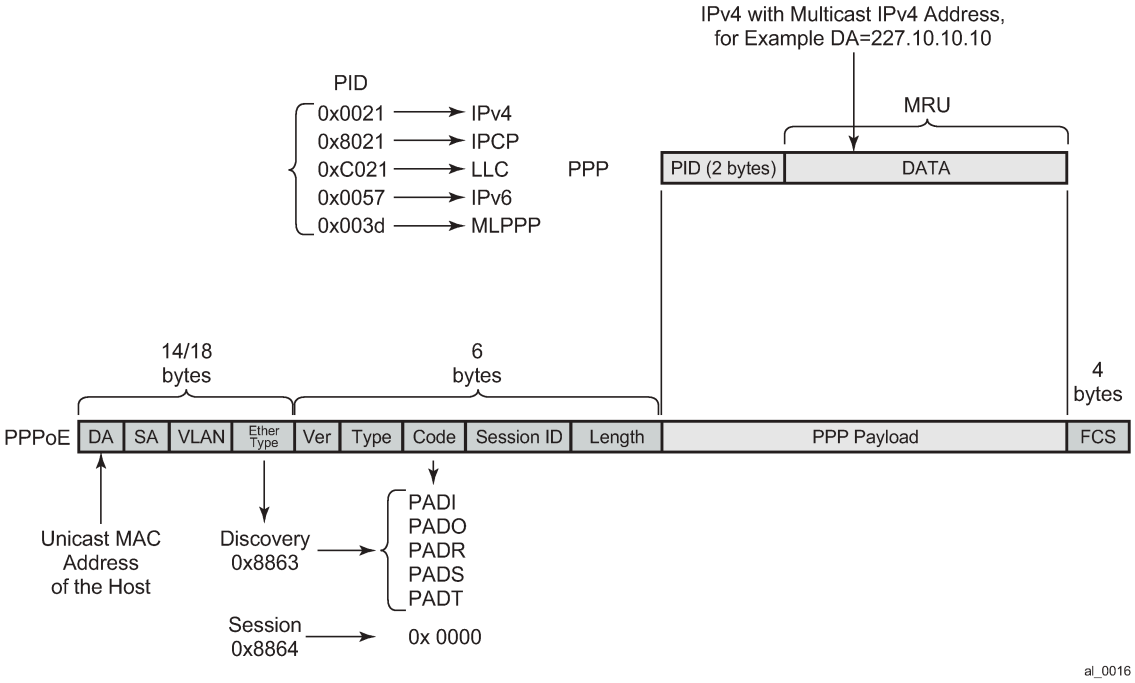
- in classic CLI, using the **per-spi-replication** command in the following contexts:
  - **config>subscriber-mgmt>igmp-policy**
  - **config>subscriber-mgmt>mld-policy** (IPv6)
- in MD-CLI, using the **replication** command in the following contexts:
  - **configure subscriber-mgmt igmp-policy**
  - **configure subscriber-mgmt mld-policy**

8.4.2 Multicast over PPPoE

In a PPPoE environment, multicast replication is performed per session (host) regardless of whether those sessions are shared per SAP or they reside on individual SAPs. This is because of the point-to-point nature of PPPoE sessions. HQoS adjustment is not needed as multicast is part of the PPPoE session traffic that is flowing through subscriber queues. Multicast packets are sent with unicast MAC address to each CPE. PPP protocol field is set to IP and the destination IP address is the multicast group address for each unique session ID.

This model is shown in Figure 59: Multicast IPv4 address and unicast MAC address in PPPoE subscriber multicast.

Figure 59: Multicast IPv4 address and unicast MAC address in PPPoE subscriber multicast



Note that in a SRRP setup, the multichassis active BNG (SRRP master state) uses the SRRP MAC address as the source MAC for both multicast and multicast control protocol packets. This is only applicable to PPPoE hosts.

8.4.3 IGMP flooding containment

The query function in IGMP can cause some unintended flooding in N:1 IPoE deployment model with AN in the IGMP snooping mode. By maintaining IGMP session states per host, it is assumed that the IGMP interaction between multicast receivers and the BNG are on a one-to-one basis. Upon arrival of an IGMP leave from a host for a specific multicast group, the IGMP querier would normally multicast a group-specific query (fast-leave). In N:1 model with SAP replication mode enabled, the 7750 SR and 7450 ESS sends a group-specific query (fast-leave) only when it receives the IGMP leave message for the last group shared amongst all subscribers on this SAP.

8.4.4 IGMP/MLD timers

IGMP/MLD timers are maintained under the following hierarchy:

IPv4:

config>router>igmp

config>service>vprn>igmp

IPv6:

config>router>mld

config>service>vprn>mld

The IGMP/MLD timers are controlled on a per routing instance (VRF or GRT) level.

The timer values are used to:

- determine the interval at which queries are transmitted (query-interval)
- determine the amount of time after which a join times out

However, the timers can be different for hosts and redirected interface in case that redirection between VRFs is enabled.

8.4.5 IGMP/MLD query intervals

IGMP/MLD query related intervals (query-interval, query-last-member-interval, query-response-interval, robust-count) are configured on a global router/vprn IGMP/MLD level. They are used to determine the IGMP/MLD timeout states and the rates at which queries are transmitted.

In case of redirection, the subscriber-host IGMP/MLD state determines the IGMP/MLD state on the redirected interface, assuming that IGMP/MLD messages are not directly received on the redirected interface (for example from the AN performing IGMP/MLD forking). For example, if the redirected interface is not receiving IGMP/MLD messages from the downstream node, then the IGMP/MLD state under the redirected interface is removed simultaneously with the removal of the IGMP/MLD state for the subscriber host (because of leave or a timeout).

If the redirected interface is receiving IGMP/MLD message directly from the downstream node, the IGMP/MLD states on that redirected interface are driven by those direct IGMP/MLD messages.

For example, an IGMP/MLD host in VRF1 has an expiry time of 60 seconds and the expiry time defined under the VRF2 where multicast traffic is redirected is set to 90 seconds. The IGMP/MLD state times out for the host in VRF1 after 60s, and if no host has joined the same multicast group in VRF2 (where redirected interface resides), the IGMP state is removed there too.

If a join was received directly on the redirection interface in VRF2, the IGMP/MLD state for that group is maintained for 90s, regardless of the IGMP/MLD state for the same group in VRF1.

8.4.6 HQoS adjustment

HQoS Adjustment is required in the scenarios where subscriber multicast traffic flow is disassociated from subscriber queues. In other words, the unicast traffic for the subscriber is flowing through the subscriber queues while at the same time multicast traffic for the same subscriber is explicitly (through redirection) or implicitly (per-sap replication mode) redirected through a separate non-subscriber queue. In this case HQoS Adjustment can be deployed where preconfigured multicast bandwidth per channel is artificially included in HQoS. For example, bandwidth consumption per multicast group must be known in advance and configured within the 7450 ESS and 7750 SR. By keeping the IGMP state per host, the bandwidth for the multicast group (channel) to which the host is registered is known and is deducted as consumed from the aggregate subscriber bandwidth.

The multicast bandwidth per channel must be known (this is always an approximation) and provisioned in the BNG node in advance.

In PPPoE and in IPoE per-host replication environment, HQoS Adjustment is not needed as multicast traffic is unicast to each subscriber and therefore is flowing through subscriber queues.

For HQoS Adjustment, the channel bandwidth definition and association with an interface is the same as in the MCAC case. This is a departure from the legacy HT channel bandwidth definition which is done by a multicast-info-policy.

Example of HQoS adjustment:

Channel definition:

```
configure
    router
        mcac
            policy <name>
                <channel definition>
```

Channel bandwidth definition policy can be applied under:

- group-interface

```
configure
    service vprn <id>
        igmp
            group-interface <ip-int-name>
                mcac
                    policy <mcac-policy-name>
```

- plain interface

```
configure
    router/service vprn
        igmp
            interface <ip-int-name>
                mcac
                    policy <mcac-policy-name>
```

- retailer group-interface

```
configure
    service vprn <id>
        igmp
            group-interface fwd-service <svc-id> <grp-if-name>
                mcac
  policy <mcac-policy-name>
```

Enabling HQoS adjustment:

```
configure
    subscriber-management
        igmp-policy <name>
            egress-rate-modify [egress-aggregate-rate-limit | scheduler <name>]
```

Applying HQoS adjustment to the subscriber:

```
configure
    subscriber-management
        sub-profile <subscriber-profile-name>
            igmp-policy <name>
```

To activate HQoS adjustment on the subscriber level, the sub-mcac-policy must be enabled under the subscriber with the following CLI:

```
configure
    subscriber-management
        sub-mcac-policy <pol-name>
            no shutdown

configure
    subscriber-management
        sub-profile <subscriber-profile-name>
            sub-mcac-policy <pol-name>
```

The adjusted bandwidth during operation can be verified with the following commands (depending whether agg-rate-limit or scheduler-policy is used):

```
*B:BNG-1# show service active-subscribers subscriber "sub-1" detail
===============================================================================
Active Subscribers
===============================================================================
-------------------------------------------------------------------------------
Subscriber sub-1
-------------------------------------------------------------------------------
I. Sched. Policy : up-silver
E. Sched. Policy : N/A                           E. Agg Rate Limit: 4000
I. Policer Ctrl. : N/A
E. Policer Ctrl. : N/A
Q Frame-Based Ac*: Disabled
Acct. Policy     : N/A                           Collect Stats    : Enabled
Rad. Acct. Pol.  : sub-1-acct
Dupl. Acct. Pol. : N/A
ANCP Pol.        : N/A
HostTrk Pol.     : N/A
IGMP Policy      : sub-1-IGMP-Pol
Sub. MCAC Policy : sub-1-MCAC
NAT Policy       : N/A
Def. Encap Offset: none                          Encap Offset Mode: none
Avg Frame Size   : N/A
Preference       : 5
Sub. ANCP-String : "sub-1"
Sub. Int Dest Id : ""
Igmp Rate Adj    : -2000
RADIUS Rate-Limit: N/A
Oper-Rate-Limit  : 2000
...
-------------------------------------------------------------------------------
*B:BNG-1#
```

Consider a different example with a scheduler instead of agg-rate-limit:

```
*A:Dut-C>config>subscr-mgmt>sub-prof# info
----------------------------------------------
            igmp-policy "pol1"
            sub-mcac-policy "smp"
            egress
                scheduler-policy "h1"
                    scheduler "t2" rate 30000
                exit
            exit
----------------------------------------------
```

```
*A:Dut-C>config>subscr-mgmt>igmp-policy# info
----------------------------------------------
            egress-rate-modify scheduler "t2"
            redirection-policy "mc_redir1"
----------------------------------------------
```

Assume that the subscriber joins a new channel with bandwidth of 1 Mb/s (1000 kb/s).

```
A:Dut-C>config>subscr-mgmt>sub-prof>egr>sched# show qos scheduler-hierarchy subscrib
er "sub_1" detail
===============================================================================
Scheduler Hierarchy - Subscriber sub_1
===============================================================================
Ingress Scheduler Policy:
Egress Scheduler Policy : h1
-------------------------------------------------------------------------------
Legend :
(*) real-time dynamic value
(w) Wire rates
B   Bytes
-------------------------------------------------------------------------------
Root (Ing)
|
No Active Members Found on slot 1

Root (Egr)
| slot(1)
|--(S) : t1
|   |    AdminPIR:90000      AdminCIR:10000
|   |
|   |
|   |    [Within CIR Level 0 Weight 0]
|   |    Assigned:0          Offered:0
|   |    Consumed:0
|   |
|   |    [Above CIR Level 0 Weight 0]
|   |    Assigned:0          Offered:0
|   |    Consumed:0
|   |    TotalConsumed:0
|   |    OperPIR:90000
|   |
|   |    [As Parent]
|   |    Rate:90000
|   |    ConsumedByChildren:0
|   |
|   |
|   |--(S) : t2
|   |   |    AdminPIR:29000      AdminCIR:10000(sum)       <==== bw 1000 from igmp s
ubstracted
|   |   |
|   |   |
|   |   |    [Within CIR Level 0 Weight 1]
|   |   |    Assigned:10000      Offered:0
|   |   |    Consumed:0
|   |   |
|   |   |    [Above CIR Level 1 Weight 1]
|   |   |    Assigned:29000      Offered:0                 <==== bw 1000 from igmp s
ubstracted
|   |   |    Consumed:0
|   |   |
|   |   |
|   |   |    TotalConsumed:0
|   |   |    OperPIR:29000                                 <==== bw 1000 from igmp s
```

```
ubstracted
|   |   |
|   |   |    [As Parent]
|   |   |    Rate:29000                                    <==== bw 1000 from igmp
substracted
|   |   |    ConsumedByChildren:0
|   |   |
|   |   |
|   |   |--(S) : t3
|   |   |   |    AdminPIR:70000      AdminCIR:10000
|   |   |   |
|   |   |   |
|   |   |   |    [Within CIR Level 0 Weight 1]
|   |   |   |    Assigned:10000      Offered:0
|   |   |   |    Consumed:0
|   |   |   |
|   |   |   |    [Above CIR Level 1 Weight 1]
|   |   |   |    Assigned:29000      Offered:0
|   |   |   |    Consumed:0
|   |   |   |
|   |   |   |
|   |   |   |    TotalConsumed:0
|   |   |   |    OperPIR:29000
|   |   |   |
|   |   |   |    [As Parent]
|   |   |   |    Rate:29000
|   |   |   |    ConsumedByChildren:0
|   |   |   |

*A:Dut-C>config>subscr-mgmt>igmp-policy# show service active-subscribers sub-mcac
===============================================================================
Active Subscribers Sub-MCAC
===============================================================================
Subscriber                          : sub_1
MCAC-policy                         : smp (inService)
In use mandatory bandwidth          : 1000
In use optional bandwidth           : 0
Available mandatory bandwidth       : 1147482647
Available optional bandwidth        : 1000000000
-------------------------------------------------------------------------------
Subscriber                          : sub_2
MCAC-policy                         : smp (inService)
In use mandatory bandwidth          : 0
In use optional bandwidth           : 0
Available mandatory bandwidth       : 1147483647
Available optional bandwidth        : 1000000000
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Number of Subscribers : 2
===============================================================================
*A:Dut-C#

*A:Dut-C# show service active-subscribers subscriber "sub_1" detail
===============================================================================
Active Subscribers
===============================================================================
-------------------------------------------------------------------------------
Subscriber sub_1 (1)
-------------------------------------------------------------------------------
I. Sched. Policy : N/A
E. Sched. Policy : h1                           E. Agg Rate Limit: Max
I. Policer Ctrl. : N/A
E. Policer Ctrl. : N/A
Q Frame-Based Ac*: Disabled
```

```
Acct. Policy     : N/A                                Collect Stats    : Disabled
Rad. Acct. Pol.  : N/A
Dupl. Acct. Pol. : N/A
ANCP Pol.        : N/A
HostTrk Pol.     : N/A
IGMP Policy      : pol1
Sub. MCAC Policy : smp
NAT Policy       : N/A
Def. Encap Offset: none                               Encap Offset Mode: none
Avg Frame Size   : N/A
Preference       : 5
Sub. ANCP-String : "sub_1"
Sub. Int Dest Id : ""
Igmp Rate Adj    : N/A
RADIUS Rate-Limit: N/A
Oper-Rate-Limit  : Maximum
...
===============================================================================
*A:Dut-C#


*A:Dut-C# show subscriber-mgmt igmp-policy "pol1"
===============================================================================
IGMP Policy pol1
===============================================================================
Import Policy                         :
Admin Version                         : 3
Num Subscribers                       : 2
Host Max Group                        : No Limit
Host Max Sources                      : No Limit
Fast Leave                            : yes
Redirection Policy                    : mc_redir1
Per Host Replication                  : no
Egress Rate Modify                    : "t2"
Mcast Reporting Destination Name      :
Mcast Reporting Admin State           : Disabled
===============================================================================
*A:Dut-C#
```

8.4.6.1 Host Tracking (HT) considerations

HT is a light version of HQoS Adjustment feature. The use of HQoS Adjustment functionality in place of HT is strongly encouraged.

When HT is enabled, the AN forks off (duplicates) the IGMP messages on the common mcast SAP to the subscriber SAP. IGMP states are not fully maintained per sub-host in the BNG, instead they are only tracked (less overhead) for bandwidth adjustment purposes.

Example of HT

```
Channel Definition:
configure
    mcast-management
        multicast-info-policy <policy-name>
            <channel to b/w mapping definition>
```

Applying channel definition policy on a router/VPRN global level:

```
config>router>multicast-info-policy <policy-name>
```

```
config>service>vprn>multicast-info-policy <policy-name>
```

Defining the rate object on which HT is applied:

```
configure
    subscriber-management
        host-tracking-policy <policy-name>
            egress-rate-modify [agg-rate-limit | scheduler <sch-name>]
```

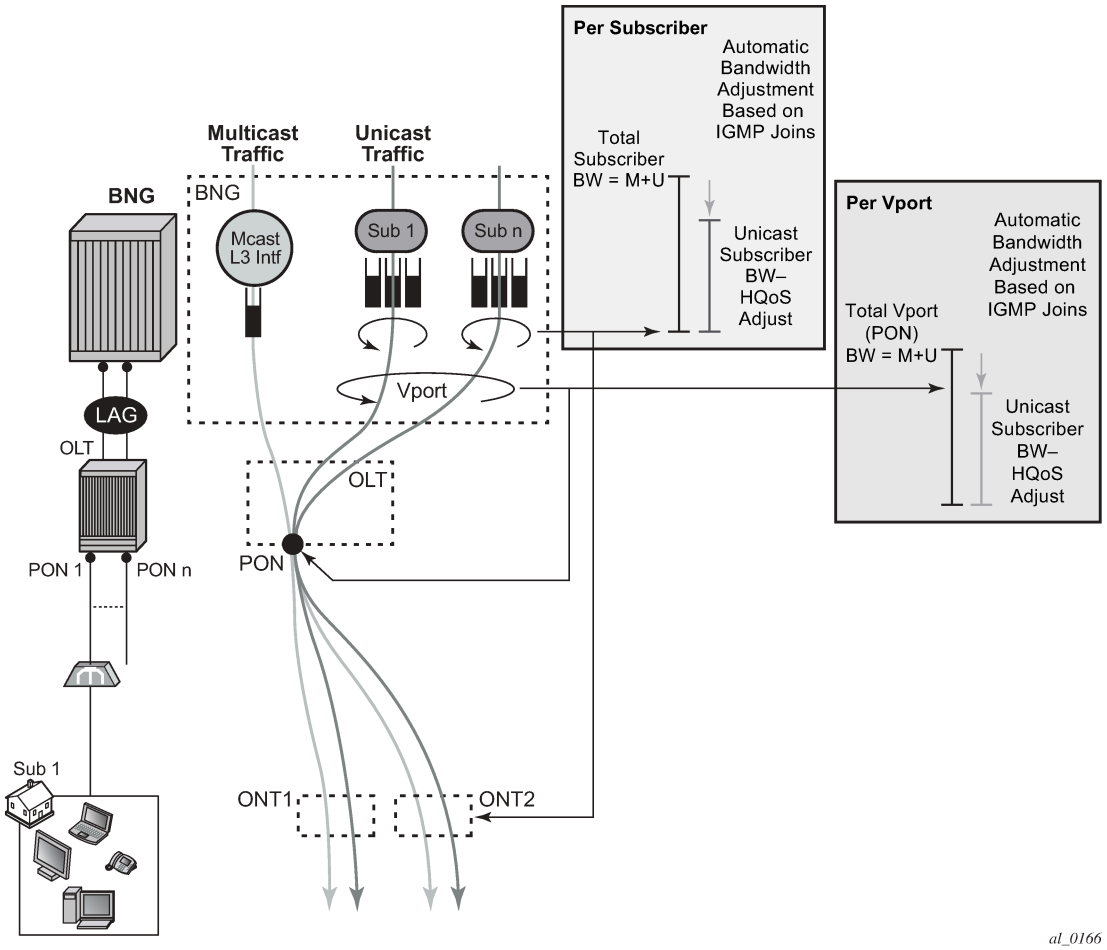
Applying the HT to the subscriber:

```
configure
    subscriber-management
        sub-profile <subscriber-profile-name>
            host-tracking-policy <policy-name>  => mutually exclusive with igmp-
policy
```

8.4.6.2 HQoS adjust per Vport

HQoS adjust per Vport can be used in environments where Vport represents a physical medium over which traffic for multiple subscribers is shared. Typical example of this scenario is shown in Figure 60: HQoS adjustment per subscriber and Vport. Multicast traffic within the router is taking a separate path from unicast traffic, only for the two traffic flows to merge later in the PON (represented by a Vport) and ONT (represented by the subscriber in the 7450 ESS and 7750 SR).

Figure 60: HQoS adjustment per subscriber and Vport



A single copy of each channel is replicated on the PON as long as there is at least one subscriber on that PON interested in this channel (has joined the IGMP/MLD group).

The 7450 ESS and 7750 SR monitors IGMP/MLD Joins at the subscriber level and consequently the channel bandwidth is subtracted from the current Vport rate limit only in the case that this is the first channel flowing through the corresponding PON. Otherwise, the Vport bandwidth is not modified. Similarly, when the channel is removed from the last subscriber on the PON, the channel bandwidth is returned to the Vport.

Association between the Vport and the subscriber is performed by an inter-destination-string or svlan during the subscriber setup phase. An inter-destination-string can be obtained either by RADIUS or LUDB. If the association between the Vport and the subscriber is performed based on the svlan (as specified in sub-sla-mgmt under the SAP or MSAP), then the destination string under the Vport must be a number matching the svlan.

The mcac-policy (channel definition bandwidth) can be applied on the group interface under which the subscribers are instantiated or in case of redirection under the redirected-interface.

In a LAG environment, the Vport instance is instantiated per member LAG link on the IOM. For accurate bandwidth control, it is prerequisite for this feature that subscriber traffic hashing is performed per Vport.

The CLI structure is as follows.

```
configure
    port <port-id>
        ethernet
            access
                egress
                    vport <name>
                        egress-rate-modify
                        agg-rate
                        host-match <destination-string>
                        port-scheduler-policy <port-scheduler-policy-name>

configure
    port <port-id>
        sonnet-sdh
            path [<sonnet-sdh-index>]
                access
                    egress
                        vport <name>
                            egress-rate-modify
                            agg-rate
                            host-match <destination-string>
                            port-scheduler-policy <port-scheduler-policy-name>
```

The Vport rate that is affected by this functionality depends on the configuration:

- When the agg-rate-limit within the Vport is configured, its value is modified based on the IGMP activity associated with the subscriber under this Vport.
- When the port-scheduler-policy within the Vport is referenced, the *max-rate* defined in the corresponding port scheduler policy is modified based on the IGMP activity associated with the subscriber under this Vport.

Note: HQoS adjust is not supported when a scheduler policy is configured under the Vport.

The Vport rates can be displayed with the following two commands:

- **show port 1/1/5 vport** *name*
- **qos scheduler-hierarchy port** *port-id* **vport** *vport-name*

As an example:

```
*A:system-1# show port 1/1/7 vport
===============================================================================
Port 1/1/7 Access Egress vport
===============================================================================
VPort Name    : isam1
Description   : (Not Specified)
Sched Policy  : 1
Rate Limit    : Max
Rate Modify   : enabled
Modify delta  : -14000
```

In this case, the configured Vport *aggregate-rate-limit* max value has been reduced by 14 Mb/s.

Similarly, if the Vport had a *port-scheduling-policy* applied, the *max-rate* value configured in the port-scheduling-policy would have been modified by the amount shown in the Modify delta output in the above command.

8.4.6.2.1 Multi-chassis redundancy

Modified Vport rate synchronization in multichassis environment relies on the synchronization of the subscriber IGMP/MLD states between the redundant nodes. Upon the switchover, the Vport rate on the newly active node is adjusted according to the current IGMP/MLD state of the subscribers associated with the Vport.

8.4.6.2.2 Scalability considerations

It is assumed that the rate of the IGMP/MLD state change on the Vport level is substantially lower than on the subscriber level.

The reason for this is that the IGMP/MLD Join/Leaves are shared amongst subscribers on the same Vport (PON for example) and therefore, the IGMP/MLD state on the Vport level is changed only for the first IGMP/MLD Join per channel and the last IGMP/MLD leave per channel.

8.4.7 Redirection

Two levels of MCAC can be enabled simultaneously and in such case this is referred as Hierarchical MCAC (H-MCAC). In case that redirection is enabled, H-MCAC per subscriber and the redirected interface is supported. However, mcac per group-interface in this case is not supported. Channel definition policy for the subscriber and the redirected interface is in this case referenced under the **igmp>interface** (redirected interface) CLI or for IPv6 **mld>interface**.

If redirection is disabled, H-MCAC for both, the subscriber and the group-interface is supported. The channel definition policy is in this case configured under the **config>router>igmp>group interface** context or for IPv6 **config>router>mld>group interface**.

There are two options in multicast redirection. The first option is to redirect all subscriber multicast traffic to a dedicated redirect interface.

Example:

Defining redirection action:

```
configure
    router
        policy-options
            begin
            policy-statement <name>
                default-action accept
                multicast-redirection [fwd-service <svc id>] <interface name>
                exit
            exit
        exit
    exit
```

The second option is to redirect only specific multicast groups to the redirect interface while the remaining groups remains on the subscriber SAP. This is applicable for both IPv4 and IPv6. For IPv6 host-ip for a policy statement is not supported.

Example:

Defining redirection action:

```
configure
    router
        policy-options
            begin
            prefix-list <name>
                prefix <IPv4 multicast groups>
                prefix <IPv4 multicast groups
            exit
            policy-statement <name>
                entry 1
                    from
                        group-address <prefix-list name>
                    action accept
                    multicast-redirection [fwd-service <svc id>] <interface name>
                exit
            exit
    exit
```

Applying redirection to the subscriber for IGMP and MLD respectively.

```
    configure
        subscr-mgmt
            igmp-policy <name>
                redirection-policy <name>
                exit
            exit
            mld-policy <name>
                redirection-policy <name>
```

Redirection that cross-connects GRT and VPRN is not supported. Redirection can be only performed between interfaces in the GRT, or between the interfaces in any of the VPRN (cross connecting VPRNs is allowed).

Redirection is also supported in a wholesaler/retailer VPRN model where redirected Layer 3 interface resides in the retailer VPRN.

8.4.8 Hierarchical Multicast CAC (H-MCAC)

MCAC is supported on three levels:

- per subscriber
- per group-interface
- per redirected interface

MCAC is supported for the following replication modes:

- per-host

- per-SAP
- per-SPI
- redirected interface

Two levels of MCAC can be enabled simultaneously and in such case this is referred as Hierarchical MCAC (H-MCAC). If redirection is enabled, H-MCAC per subscriber and the redirected interface is supported. However, MCAC per group-interface in this case is not supported. Channel definition policy for the subscriber and the redirected interface is in this case referenced under the **config>router>igmp>interface** (redirected interface) CLI hierarchy or IPv6 **config>router>mld>interface**.

If redirection is disabled, H-MCAC for both, the subscriber and the group-interface is supported. The channel definition policy is in this case configured under the **config>router>igmp>group-interface** CLI hierarchy or IPv6 **config>router>mld>group-interface**.

Examples

Note: The same channel definition and association with interfaces is used for MCAC/H-MCAC and HQoS Adjustment.

Channel definition:

```
configure
    router
        mcac
            policy <mcac-pol-name>
                bundle <bundle-name>
                    bandwidth <kbps>
                    channel <start-address> <end-address> bw <bw> [class {high|low}]
 [type {mandatory|optional}]
                    :
                    :
```

Channel bandwidth definition policy can be referenced under the:

- group-interface (this is used for subscribers when redirection is disabled)

```
configure
                    service vprn <id>
               igmp/mld
                    group-interface <ip-int-name>
                       mcac
                           policy <mcac-policy-name>
                           policy <mcac-policy-name>
```

- plain interface

```
configure
    router
        igmp/mld
            interface <ip-int-name>
                mcac
                    policy <mcac-policy-name>

configure
    service vprn <id>
        igmp/mld
            interface <if-name>
                mcac
```

```
                    unconstrained-bw <bandwidth> mandatory-bw <mandatory-bw>
```

- retailer’s VPRN reference the group-interface in the wholesaler’s VPRN

```
configure
    service vprn <id>
        igmp/mld
            group-interface fwd-service <svc-id> <ip-int-name>
                mcac
                    policy <mcac-policy-name>
```

Enabling MCAC per subscriber:

```
configure
    subscr-mgmt
        sub-mcac-policy <name>
            unconstrained-bw <bandwidth> mandatory-bw <mandatory-bw>

configure
    subscr-mgmt
        sub-profile <subscriber-profile-name>
            sub-mcac-policy <name>
```

Enabling MCAC per-group-interface

```
configure
        service vprn <id>
            igmp/mld
                group-interface <ip-int-name>
                    mcac
                        unconstrained-bw <bandwidth> mandatory-bw <mandatory-bw>
```

Enabling MCAC per redirected interface

```
configure
    router
        igmp/mld
            interface <ip-int-name>
                mcac
                    unconstrained-bw <bandwidth> mandatory-bw <mandatory-bw>

configure
    service vprn <id>
        igmp/mld
            interface <ip-int-name>
                mcac
                    unconstrained-bw <bandwidth> mandatory-bw <mandatory-bw>
```

8.4.8.1 MCAC bundle bandwidth limit considerations

In addition to multicast bandwidth limit that can be imposed on subscribers, group-interfaces or regular interfaces, there is another multicast bandwidth limit that can be imposed on a group of channels (channel bundle).

The MCAC policy, aside from the channel bandwidth definitions, could optionally contain this bandwidth cap for the group of channels:

```
config>router>mcac# info
----------------------------------------------
        policy "test"
            bundle "test" create
                bandwidth 100000
                channel 239.0.0.10 239.0.0.10 bw 10000 type mandatory
                channel 239.0.0.11 239.0.0.15 bw 5000 type mandatory
                channel 239.0.0.20 239.0.0.30 bw 5000 type optional
            exit
        exit
```

This can be used to prevent a single set of channels from monopolizing MCAC bandwidth allocated to the entire interface. The bandwidth of each individual bundle is capped to some value below the interface MCAC bandwidth limit, allowing each bundle to have its own share of the interface MCAC bandwidth.

In most cases, the bandwidth limit per bundle is not necessary to configure. The aggregate limit per all channels as defined under the subscriber or interface covers majority of scenarios. In case that one wants to explore the bundle bandwidth limits and how they affect MCAC behavior, the following information helps understanding this topic.

To further understand how various MCAC bandwidth limits are applied, one need to understand the concept of the mandatory bandwidth that is pre-allocated in the following way:

- Bandwidth of each mandatory channel in a bundle is pre-allocated. The artifacts of this are:
  - The total mandatory bandwidth in the bundle cannot exceed the bundle cap. For the sake of deterministic behavior, the configured bandwidth of each mandatory channel in the bundle is counted toward the total mandatory bandwidth only once. This means that only one replication of each mandatory channel is assumed. This is normal behavior on a regular interface with a single SAP under it. More than one replication of the same channel per regular interface (or SAP) would lead to packet duplication.
  - Optional (non-mandatory) channels can use only the difference in bandwidth between the bundle cap and total pre-allocated mandatory bandwidth. They cannot use more bandwidth than that even if the total pre-allocated mandatory bandwidth is not used up (mandatory channels are not being replicated).
- Mandatory bandwidth under the interface is pre-allocated and subtracted from the unconstrained bandwidth. In the configuration example below, 2 Mb/s is pre-allocated (guaranteed for mandatory channels) and the remaining 8 Mb/s can be used by the optional channels on a first come first serve basis.

```
config>router>igmp# info
----------------------------------------------
            interface "ge-1/1/1"
                mcac
                    unconstrained-bw 10000 mandatory-bw 2000
                exit
            exit
```

The bundle bandwidth limit poses a problem when the MCAC policy is applied under the group interface. The reason is that the group interface represents the aggregation point for the subscribers and their bandwidth. As such it is natural that the any aggregated bandwidth limit under the group interface be larger

than the bandwidth limit applied to any individual subscriber under it. Because the MCAC policy, along with the bundle bandwidth limit, is inherited by all subscribers under the group-interface, the exhaustion of the bundle bandwidth limit under the group interface coincides with the exhaustion of the bundle bandwidth limit of any individual subscriber. This results in a single subscriber starving out of multicast bandwidth the remaining subscribers under the same group interface. While it is perfectly acceptable for the subscribers to inherit the multicast channel definition from the group-interface, for the above reasons it is not acceptable that the subscriber inherit the bandwidth cap from the group-interface.

To resolve this situation, the MCAC bandwidth limits are independently configured for the group interface level (aggregated level) and the subscriber level with the **unconstrained-bw** *kbps* **mandatory-bw** *kbps* command. The unwanted bundle bandwidth cap in the MCAC policy is ignored under the group-interface and under the subscriber. However, the bundle bandwidth cap is applied automatically to each SAP under the group interface. A SAP is a natural place for a bundle bandwidth limit because each channel on a SAP can be replicated only once and therefore the amount of pre-allocated mandatory bandwidth can be pre-calculated. This is not the case for the group interface where single channel can be replicated multiple times (one per each SAP under the grp-if). Similarly, the same channel can be replicated multiple times for the same subscriber in per-host replication mode. Only subscribers in per-sap replication mode warrants a single replication per channel. Therefore, if bundle cap is configured, it is applied to limit the bandwidth of the bundle that is applied to a subscriber in a per-sap replication mode.

Figure 61: MCAC policy inheritance in per-SAP replication mode depicts MCAC related inheritances and MCAC bandwidth allocation model in per-sap replication mode. The MCAC policy is applied to the group interface and inherited by each subscriber as well as each SAP under the same group interface. However, the bundle bandwidth limit in the MCAC policy is ignored on the group-interface and under the subscriber (denoted by the X in the figure). The bundle limit is applied only to each sap under the group-interface.

Overall (non-bundle) MCAC bandwidth limits are independently applied to the group-interface and the subscribers. According to our example, 20 Mb/s of multicast bandwidth in total is allocated per group-interface. 6 Mb/s of the 20 Mb/s is allocated for mandatory channels. This leaves 14 Mb/s of multicast bandwidth for the optional channels combined served on a first come first serve basis. Each physical replication (multiple replications of the same channel can occur, one per each SAP), counts toward the respective group-interface bandwidth limits.

Similar logic applies to the subscriber MCAC bandwidth limits which are applied per sub-profile.

Finally, each SAP can optionally contain the bundle bandwidth limit.

Note: In a hierarchical MCAC fashion, if either of the bandwidth checks fails (SAP, sub or grp-if) the channel admission for the subscriber also fails.

In the example, six subscriber hosts watch the same channel but there are only three active replications (one per SAP). This would yield:

- 14 Mb/s of available multicast bandwidth under the group-interface. This bandwidth can be used for optional channels on a first come first serve basis. No reserved bandwidth is left.
- For Subscriber A, 3 Mb/s is still reserved for mandatory channels and 5 MB/s is available for optional channels (first come first serve). All this assume that the SAP and the grp-if bandwidth checks pass.
- For Subscriber B, 2 Mb/s is still reserved for mandatory channels and 6 Mb/s is available for optional channels (first come first serve). All this assume that the SAP and the grp-if bandwidth checks pass.
- For Subscriber C, no reserved bandwidth for mandatory channels is left. 3 Mb/s is still left for optional channels. All this assume that the SAP and the grp-if bandwidth checks pass.
- For SAPs, considering that 2 Mb/s are currently replicating (ch A, each SAP can still accept 1 Mb/s of the mandatory bandwidth (channel B and 7 Mb/s of the remaining optional channels.

To resolve this, the MCAC bandwidth limits are independently configured for the group interface (aggregated level) and the subscriber using the **unconstrained-bw mandatory-bw** command. The unwanted bundle bandwidth restriction in the MCAC policy is ignored for the group interface and under the subscriber. However, the bundle bandwidth maximum is applied automatically to each SAP for the group interface.

The bundle bandwidth limit is applied at the SAP level because each channel on the SAP is replicated only once and therefore the amount of pre-allocated mandatory bandwidth is precalculated. This is not the case for the group interface, where a single channel is replicated multiple times (once for each SAP in the group interface). Similarly, the same channel is replicated multiple times for the same subscriber in per-host replication and per-SPI replication modes. Only subscribers in per-SAP replication mode warrant a single replication per channel. Therefore, if the bundle maximum is configured, it limits the bandwidth of the bundle that is applied to a subscriber in per-SAP replication mode.

Figure 61: MCAC policy inheritance in per-SAP replication mode

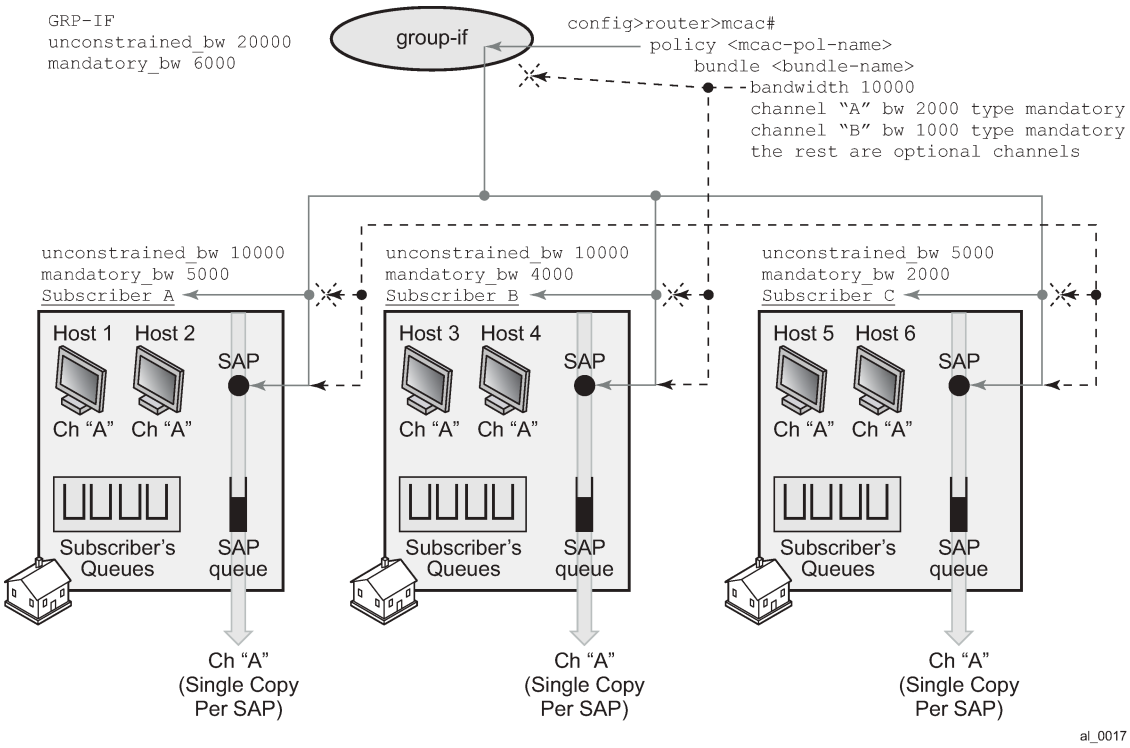


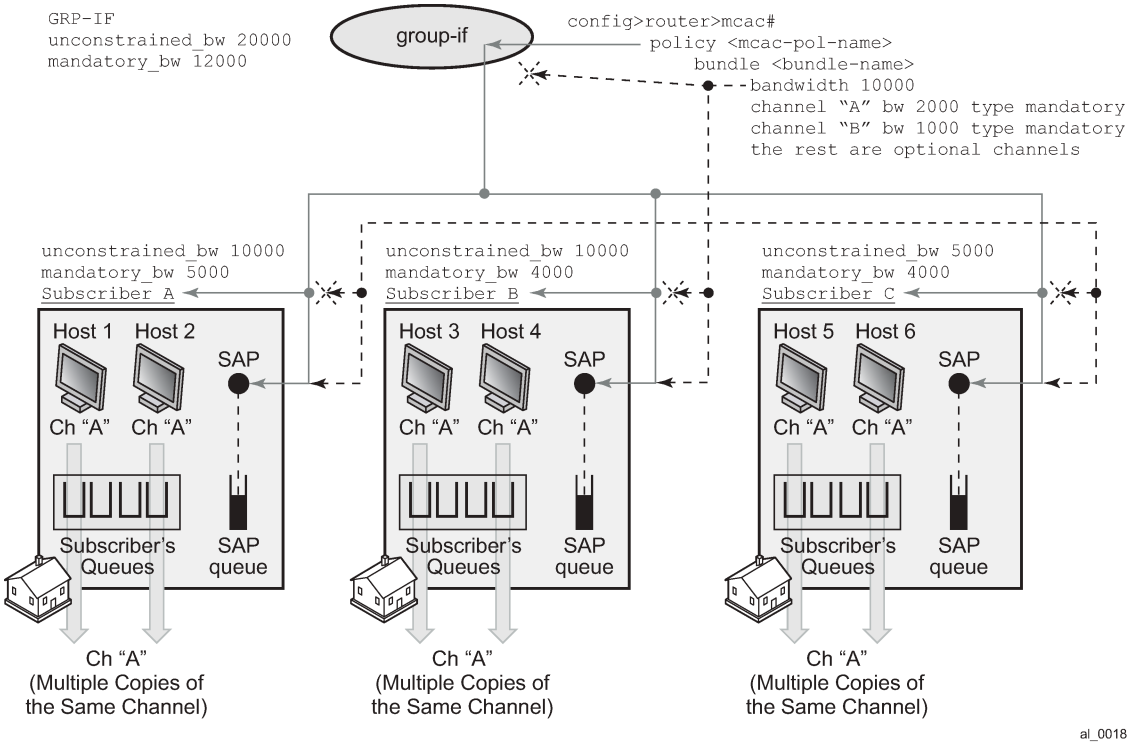
Figure 62: MCAC policy inheritance in per-HOST replication mode depicts behavior in per-host replication mode. MCAC policy inheritance flow is the same as in the previous example with the difference that the bundle limit has no effect at all. Each host generates its own copy of the same multicast stream that is flowing through subscriber queues and not the SAP queue. Because each of the copies counts toward the subscriber or group-interface bandwidth limits, the multicast bandwidth consumption is higher in this example. This needs to be reflected in the configured multicast bandwidth limits. For example, the group-interface mandatory bandwidth limit is increased to 12 Mb/s.

In our example, 6 subscriber hosts are still watching the same channel but now the number of replications is doubled from previous example. So the final tally for our MCAC bandwidth limit is as follows:

- For Subscriber A, 1 Mb/s is still reserved for mandatory channels and 5 Mb/s for optional channels (first come first serve). All this assume that SAP and grp-if bandwidth checks pass.

- For Subscriber B, no reserved mandatory bandwidth is left. 6 Mb/s is still left for optional channels (first come first serve). All this assume that SAP and grp-if bandwidth checks pass.
- For Subscriber C, no reserved mandatory bandwidth is left. 1 Mb/s is still left for optional channels (first come first serve). All this assume that SAP and grp-if bandwidth checks pass.
- No reserved bandwidth is left under the group-interface. 8 Mb/s of available multicast bandwidth under the group-interface is still left for optional channels on a first come first serve basis.
- Bundle limit on a SAP is irrelevant in this case.

Figure 62: MCAC policy inheritance in per-HOST replication mode



8.4.9 Determining MCAC policy in effect

Channel bandwidth definition (by a MCAC policy) can be applied under the interface level (group-interface or regular interface):

config>router/service>igmp/mld>interface/grp-if

The following configuration options can lead to the confusion as to which MCAC policy is in effect:

- The MCAC policy (channel bandwidth definition) can be applied under two different places (grp-if or regular intf).
- The same policy is used for (H)MCAC and HQoS Adjust with redirection enabled or disabled.

The general, the rule is that the MCAC policy under the group-interface is always be in effect in cases where redirection is disabled. This is valid for subscriber or group-interface MCAC, H-MCAC (subscriber and group-interface) and HQoS Adjust in per SAP replication mode.

If redirection is enabled, the MCAC policy under the group-interface is ignored.

If redirection is enabled, but there is no MCAC policy applied under the redirected interface (redirected interface is the interface to which IGMP/MLD Joins are redirected from subscriber hosts), regardless of whether the MCAC policy under the group-interface is applied or not, then:

- HQoS adjust has no effect.
- MCAC has no effect not only per redirected interface but also per subscriber.

8.4.10 Multicast filtering

Multicast filtering must be done per session (host) for IPoE and PPPoE. There are two types of filters that are supported:

- IGMP filters on access ingress. Those filters control the flow of IGMP messages between the host and the BNG. They are applied with the import statement in the igmp-policy. The same filters are used for multicast-redirection policy:

  For IPv4:

  ```
  configure
      subscr-mgmt
          igmp-policy <name>
              import <policy-name>
  ```

  For IPv6:

  ```
  configure
      subscr-mgmt
          mld-policy <name>
              import <policy-name>
  ```

  An example of the filter definition is shown below:

  ```
  configure
      router
          policy-options
              begin
              prefix-list <pref-name>
                  prefix <pref-definition>
              policy-statement <name>
                  entry 1
                      from
                          group-address <pref-name>
                          source-address <ip>
                          protocol igmp
                      exit
                      action accept
                      exit
                  exit
                  default-action reject
  ```

- Regular traffic filters where control multicast traffic flow can be controlled in both directions (ingress/egress). This is supported through ip-filters under the SLA profile.

For IPv6 specifically, up to 15 import policies can be applied to the subscriber IPv6 host. Each of the import policies can be configured as an IPTV package. Depending on the IPTV package the subscriber ordered, a list of import policies is applied at authentication.

At authentication, the first 14 import policies are applied either through LUDB or RADIUS using a list of VSAs. These 14 import policies are not read in any particular order and to make the list deterministic as a black list or a white list, the import policy inside the MLD policy must be configured. The import policy inside the MLD policy is always applied as the last policy which determines whether the list of 14 policies is a white list or a black list. For example, to make the list of 15 import policies a white list, the first 14 policies should have **action** configured as **accept** and the last import policy inside the MLD policy should have default action configured as **reject**. For a black list, the first 14 policies should have **action** configured as **reject** and the last import policy inside the MLD policy should have default action configured as **accept**. It is recommended to configure the MLD import policy with a default action without an import list which is used to determine if the list is a black or white list. If the last import policy is overridden, it changes the MLD policy which in turn overrides the subscriber profile.

It is possible to make a mixed black and white list, where the general rule is that the import policy inside the MLD policy is always the last to be applied to the subscriber.

The list of MLD import policies can be overridden either through RADIUS CoA or through the **tools**>**perform**>**subscriber-management**>**coa** command. There are two VSAs for overriding the list of import policies. The first VSA, Alc-Mld-Import-Policy, is used to completely override the first 14 MLD policies. A RADIUS CoA can contain up to 14 VSAs to override the existing list. Because of limitations with the length of a CLI command, up to five VSAs can be passed to override the existing list when using the **tools** command. The second VSA, Alc-Mld-Import-Policy-Modif, allows the addition or subtraction of a single MLD import policy. To add an import policy to the existing list, use a prefix of 'a:' and to remove an import policy, use a prefix of 's:', for example, 'a:sport-pack1-hd' or 's:movie-pack1-hd'. Multiple import policies can be added or removed within a single CoA by using multiple VSAs. Up to five import policies can be added or removed using a single CoA. The last import policy inside the MLD policy can be overridden using the subscriber profile.

If a subscriber is not associated with an MLD policy, the import policies applied either at authentication or overridden in mid-session are always stored while the subscriber is online. The import policy does not take effect as long as the subscriber is not associated with an MLD policy. However, as soon as the subscriber is associated with an MLD policy, the list of import policies stored is instantly applied.

8.4.11 Joining the multicast tree

The delivery of multicast to the subscribers-interface in a VPRN environment depends on the multicast deployment model (PIM, mBGP). In each model, a subscriber-interface is treated as a regular CE-PE interface that has registered v4 multicast listeners.

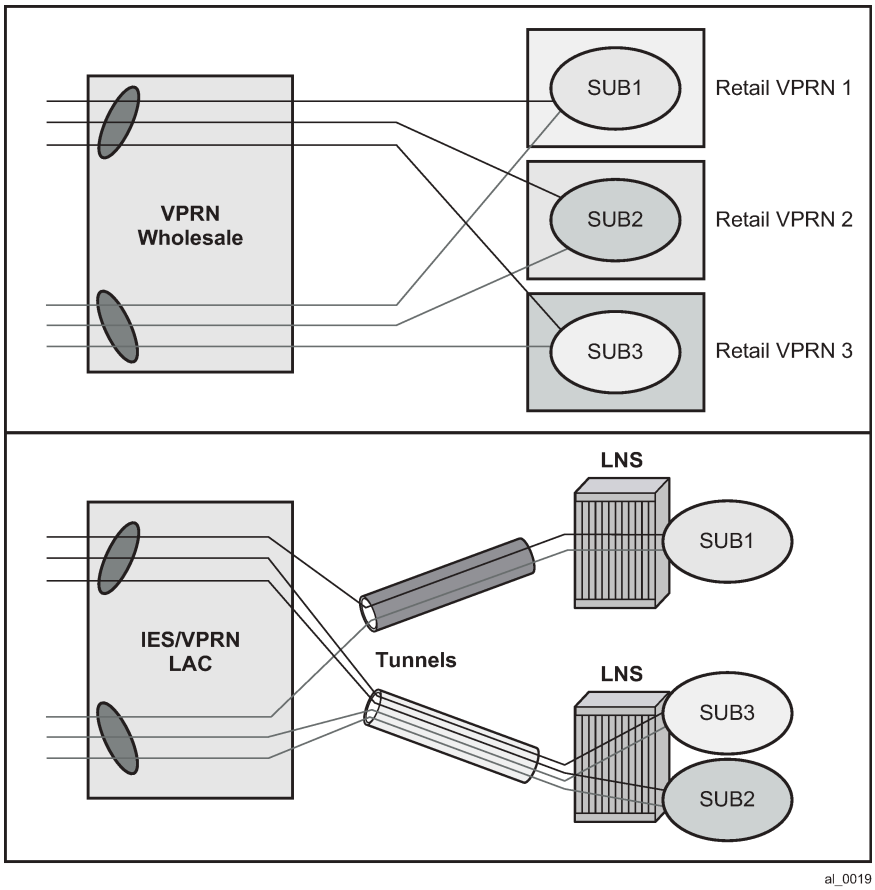
8.4.12 Wholesale/retail requirements

Multicast support on subscriber interfaces is supported in both wholesale/retail models:

- wholesale/retail VPRN (IPoE and PPPoE)
- LAC/LNS (PPPoE only)

  This model is shown in Figure 63: Wholesale/retail multicast support.

Figure 63: Wholesale/retail multicast support



The distinction between these two models is that in the case of LAC/LNS, the replication is performed further up in the network on an LNS node. This means that the traffic between LAC and LNS is multiplied by the amount of replications.

8.4.13 QoS considerations

In per-sap replication mode (which is applicable only to IPoE subscribers), multicast traffic is forwarded through the SAP queue which is outside of the subscriber's queues and therefore not accounted in subscriber aggregate rate limit. HQoS Adjust is used to remedy this situation.

If the SAP queue is removed from the static SAP in IPoE 1:1 model (with profiled-only-traffic command), multicast traffic flows through internal queues which cannot be tied into a port-scheduler as part of HQoS. Consequently, the **port-scheduler** *max-rate* as defined in the port scheduler policy is used only to rate limit unicast traffic. In other words, the *max-rate* value in the port scheduler policy must be lowered for the amount of anticipated multicast traffic that flows through the port where the port scheduler policy is applied.

A similar logic applies to the per-sap replication mode on dynamic SAPs (MSAPs) even if the SAP queue is not removed. Although the multicast traffic is flowing through the SAP queue in this case, the SAP QoS policy on MSAP cannot be changed from the default one. The default QoS policy on a SAP contains a single queue that is not parented by the port-scheduler.

Those restrictions do not apply to static SAPs where the SAP QoS policy can be customized and its queues consequently tied to the port-scheduler.

8.4.14 Redundancy considerations

The subscriber can receive multicast content through the subscriber SAP, the redirected interface, or a combination of both.

Multicast redundancy is only supported on a MC-LAG topology. Multicast traffic can be delivered over the subscriber SAP, the redirected interface, or a combination of both.

Subscriber IGMP states can be synchronized across multiple routers to ensure minimal interruption of (video delivery) service during network outages. The IGMP/MLD state of a subscriber-host in the system is tied to the state of the underlying MC-LAG protection mechanism. For example, IGMP states is activated only for subscribers that are anchored under the group-interfaces with master SRRP state or under an active MC-LAG port.

For multicast redirection on a MC-LAG topology, it must be ensured that the redirected interface (the interface to which multicast forwarding is redirected) is under the same MC-LAG as the subscriber. Otherwise, IGMP states on the redirected interface is derived independently of the IGMP states for the subscriber from which IGMP/MLD messages are redirected.

The IGMP/MLD synchronization process in conjunction with underlying access protection mechanisms works as follows:

IGMP/MLD states for the subscriber updates only if IGMP/MLD messages (Joins/Leaves/Reports, and so on) are received directly from the downstream node on an active MC-LAG link. This is valid irrespective of the IGMP querier status for the subscriber.

In all other cases, assuming that some protection mechanism in the access is present, the IGMP/MLD messages are discarded and consequently no IGMP/MLD state is updated. Similar logic applies to regular Layer 3 interfaces, where SRRP is replaced with VRRP.

- After the subscriber IGMP/MLD state is updated as a result of a directly-received IGMP/MLD message on an active subscriber (SRRP master state or active MC-LAG), the sync IGMP/MLD message is sent to the standby subscriber over the Multi-Chassis Synchronization (MCS) protocol. The synchronized IGMP state is populated in the MCS database in all pairing routers.
- If an IGMP/MLD sync (MCS) message is received from the peering node, the IGMP state for the standby subscriber is updated in the MCS database but it is not downloaded into the forwarding plane unless there is a switchover. If the IGMP/MLD sync message is received for the active subscriber, the message is discarded.
- IGMP/MLD queries are sent out only by IGMP/MLD querier. In an MC-LAG environment, this is the node with the active MC-LAG link.

  Note: MC-LAG is usually configured with SRRP and the SRRP state is derived from the MC-LAG.

- IGMP/MLD states from the MCS database are:
  - activated on non-querier subscriber in case that neither SRRP nor MC-LAG is deployed. It is assumed that the querier subscriber has received the original IGMP/MLD message and consequently sent the IGMP/MLD MCS Sync to the non-querier (standby). Non-querier interface accepts the MCS sync message and also it propagate these IGMP/MLD states to PIM.

    The querier subscriber does not accept the IGMP/MLD update from the MCS database.

  - aware of the state of MC-LAG. As soon as the standby MC-LAG becomes active, the IGMP/MLD states is activated and is propagated to PIM. Traffic is forwarded as soon as multicast streams are delivered to the node and the IGMP states under the subscriber are activated. On a standby MC-LAG, IGMP states are not propagated from the MCS database to PIM and consequently subscribers.
  - aware of the SRRP state. Because the subscriber with SRRP master state is considered active, the states are propagated to PIM as well. On standby SRRP, IGMP states are be propagated from MCS database to PIM and consequently to subscribers.
- For MC-lag setups, after the switchover is triggered by MC-LAG or SRRP, the IGMP/MLD states from MCS database on the newly active MC-LAG node or subscriber under the new SRRP instance in the master state is sent to PIM and consequently to the forwarding plane effectively turning on multicast forwarding.

An active and standby subscriber refers to the state of underlying protection mechanism (active MC-LAG).

Note:

- The subscribers themselves are always instantiated (or active) on both nodes. However, traffic forwarding over those subscribers is driven by the state of the underlying protection mechanism (MC-LAG). Hence the terms active and standby subscriber.
- In subscriber environment, SRRP should be always activated in dual-homing scenario. SRRP in subscriber environment ensures that downstream traffic is forwarded by the same node that is forwarding upstream traffic. In this fashion, accounting and QoS for the subscriber are consolidated within a single node.

To summarize, in multichassis environment with subscribers, IGMP synchronization enabled and an access layer protection mechanism in place (MC-LAG), the behavior for is the following:

- IGMP/MLD states are synchronized between the chassis.
- On a MC-LAG setup, only the SRRP instance in master state or active MC-LAG forwards downstream multicast traffic.
- Length of outage during the switchover is determined by the detection and recovery of the underlying protection mechanism (MC-LAG or MCS) in addition to local propagation of IGMP/MLD states from MCS database to PIM and consequently to forwarding plane.

Note: IGMP/MLD states can be statically configured on both redundant nodes to attract multicast traffic from upstream and therefore minimize outage during the switchover.

8.4.14.1 Redirection considerations

The redirection policy has two options: to redirect only a specific set of multicast groups to the redirect interface or redirect all multicast to the redirect interface. The redirect policy is source agnostic.

On a MC-LAG setup, for redirection and MCS to work simultaneously in predictable manner, the redirected interface and the corresponding subscribers have to be protected by the same MC-LAG. This binds the redirected interfaces and the subscriber-hosts to the same physical ports.

The following describe some guidelines for a MC-LAG setup:

- The active subscriber replicates its received IGMP/MLD message to the redirected Layer 3 interface. The Layer 3 redirected interface accepts this message:

  - independently of the corresponding VRRP state if MC-LAG is not used
  - only if the Layer 3 interface is IGMP querier
  - MC-LAG is used and in active state
- In all other cases the IGMP message under the Layer 3 redirected interface is be rejected.

  Note: Layer 3 redirected interface can also receive IGMP message directly from the downstream node in case that IGMP forking in the access node is activated.

- The Layer 3 redirected interface does not accept the IGMP state update from the MCS database unless the Layer 3 interface is a non-querier.
- In case that the Layer 3 redirected interface is part of MC-LAG, the IGMP state update sent to it by MCS database is accepted only during the transitioning phase from standby to active MC-LAG state.

  Briefly, IGMP states on Layer 3 interface are not VRRP aware. However, they are MC-LAG aware.

8.4.15 Query intervals for multicast

The query interval commands—**query-interval**, **query-last-listener-interval**, **query-last-member-interval**, and **query-response-interval** are configurable in the following command contexts:

- router and VPRN service IGMP and MLD command contexts
- router and VPRN service IGMP and MLD group interface command contexts
- IGMP policy and MLD policy command contexts

The IGMP policy and MLD policy query intervals are only applicable to ESM host-based queries. Group interface query intervals are only applicable when the **no sub-hosts-only** command is configured. This triggers all group interface SAPs to send multicast queries. IGMP policy, MLD policy, and group interface timers use the router IGMP and MLD timers as a default fallback (see ESM host-based queries and Group interface-based queries).

For IGMP, the **query-interval** must be configured to be longer than the **query-last-member-interval** and **query-response-interval**. For MLD, the **query-interval** must be configured to be longer than the **query-last-listener-interval** and **query-response-interval**.

8.4.15.1 ESM host-based queries

An ESM host associated with an IGMP policy or an MLD policy is sent host-based queries. A host-based query uses the intervals defined under the IGMP policy or MLD policy first, and if none are configured, it uses the interval values configured under the router IGMP or MLD context as a fallback. The router IGMP and MLD interval values are set to default values if none are configured. It is possible for the IGMP or MLD policy to only configure some of the query intervals while leaving other query intervals to use the values configured under the router IGMP or MLD context. For example, if the IGMP policy only has the **query-interval** configured to 11, the remaining **query-last-member-interval** and **query-response-interval** configurations uses values 1 and 10, respectively, defined under **router igmp**. It is also possible to configure an invalid combination of query timers when using both the IGMP policy configured intervals and intervals under **router igmp**. Using the example above, if the IGMP policy only sets the **query-interval** to 5, while the **query-response-interval** under **router igmp** is set to 10, this becomes an invalid combination because the response interval is longer than the query interval. When the system detects an invalid combination, it falls back and use all interval values configured under the **router igmp** or **router mld**

context and ignores the interval configured under the IGMP or MLD policy. It is highly recommended that all three query intervals be configured together under the IGMP or MLD policy. To see the active values that each host is using for queries, use the **show router igmp hosts detail** or **show router mld host** command.

8.4.15.2 Group interface-based queries

When the group interface is configured with **no sub-hosts-only** (applicable only to IPoE multicast), each SAP configured under that group interface sends out SAP-based queries. In this case, a single query is sent out of the SAP. The query intervals uses the values configured under the group interface. If there are no values configured under the group interface, the router IGMP or MLD values are used. When using the **no sub-hosts-only** command, IGMP and MLD policy are not required. To see the active values for the SAP-based queries, use the **show router igmp group-interface detail** or **show router mld group-interface detail** command.

8.4.16 ESM multicast replication modes

The following ESM multicast replication modes are supported:

- **per-host replication**

  This mode supports multicast packet replication per subscriber host. Each multicast packet destination uses an individual host source MAC address. The multicast traffic is unicast to the host by replacing the destination MAC from a multicast MAC with the host MAC address. A typical use case for per-host replication is for PPoE subscribers.

- **per-SAP replication**

  When all subscribers use this replication mode on a single SAP, the system ensures that no more than one multicast packet is sent per (S,G). This mode is used by default when another replication mode is not configured.

  Note: In the case of the wholesale and retail model (non-CUPS), if the SAP is configured on the VPRN and the subscriber is created on the IES, per-SAP replication is not supported and the system automatically defaults to per-host replication.

- **per-SPI replication**

  In this mode, the multicast packet replication depends on the number of SLA profile instances configured for the subscriber hosts. A single copy of the multicast packet is sent to hosts that use the same SLA profile, and the multicast packet is replicated for each host that uses a different SLA profile. For example, if all subscriber hosts use the same SLA profile, only one multicast packet is sent from a specific (S,G). However, if the subscriber hosts use three different SLA profiles, three multicast packets are sent.

- **multicast redirection**

  In this mode, the multicast packets are sent to all the subscribers using a dedicated VLAN.

8.5 ESM multicast on BNG CUPS UPF

The User Plane Function (UPF) of BNG Control and User Plane Separation (CUPS) deployments supports ESM multicast as described in Table 7: ESM multicast on BNG CUPS UPF .

Table 7: ESM multicast on BNG CUPS UPF

Multicast replication model	Regular ESM model	Acting as UPF for BNG CUPS
Per-host	Supported in: • VPRN service • IES The following apply: • When both the wholesale and retail are VPRNs, the multicast source must be injected into the retail VPRN. • When the wholesale is a VPRN and the retail is an IES, the multicast source must be injected into the retail IES.	Supported in: • VPRN service • IES When the wholesale and retail model is used, the following apply: • If retail is a VPRN, the multicast source must be injected into the retail VPRN. • If retail is an IES, the multicast source must be injected into the retail IES.
Per-SAP	Supported in: • VPRN service • IES When both the wholesale and retail are VPRNs, the multicast source must be injected into the wholesale VPRN.	Not Supported
Per-SPI	Supported in: • VPRN service • IES The following apply: • When both the wholesale and retail are VPRNs, the multicast source must be injected into the retail VPRN. • When the wholesale is a VPRN and the retail is an IES, the multicast source must be injected into the retail IES.	Supported in: • VPRN service • IES When the wholesale and retail model is used: • If retail is a VPRN, the multicast source must be injected into the retail VPRN. • If retail is an IES, the multicast source must be injected into the retail IES.
Mcast redirect	Supported in: • VPRN SAP • IES SAP	Supported in: • VPRN SAP • IES SAP

8.6 ESM multicast support on VSR

ESM multicast redirection is the only supported ESM multicast mode on VSR for both IPv4 and IPv6 subscribers. The subscriber can be in an IES or a VPRN routing instance.

The following features are unsupported with ESM multicast redirection:

- MCAC
- egress rate adjust
- MCS and redundancy with redirect interface
- wholesale and retail service including both VPRN and LNS

8.7 Configuring Triple Play multicast services with CLI

This section provides information to configure multicast parameters in a Triple Play network using the command line interface.

8.7.1 Configuring IGMP snooping in the BSA

8.7.1.1 Enabling IGMP snooping in a VPLS service

8.7.1.2 IGMPv3 multicast routers

When multicast routers use IGMPv3, it is enough to only enable IGMP snooping, without any further modification of parameters.

The following displays an example of an IGMP snooping configuration:

```
A:ALA-48>config>service>vpls# info
----------------------------------------------
            igmp-snooping
                no shutdown
            exit
            no shutdown
----------------------------------------------
A:ALA-48>config>service>vpls#
```

8.7.1.3 With IGMPv1/2 multicast routers

When the multicast routers do not support IGMPv3, some timing parameters must be configured locally in the SR-series routers.

Note: All routers in the multicast network must use the same values for these parameters.

The following displays an example of a modified IGMP snooping configuration:

```
A:ALA-48>config>service>vpls# info
----------------------------------------------
            stp
                shutdown
            exit
            igmp-snooping
                query-interval 60
                robust-count 5
                no shutdown
            exit
            no shutdown
----------------------------------------------
A:ALA-48>config>service>vpls#
```

8.7.1.4 Modifying IGMP snooping parameters

For interoperability with some multicast routers, the source IP address of IGMP group reports can be configured. Use the following CLI syntax to customize this IGMP snooping parameter:

The following displays an example of a modified IGMP snooping configuration:

```
A:ALA-48>config>service>vpls# info
----------------------------------------------
            stp
                shutdown
            exit
            igmp-snooping
                query-interval 60
                robust-count 5
                report-src-ip 10.20.20.20
                no shutdown
            exit
            no shutdown
----------------------------------------------
A:ALA-48>config>service>vpls#
```

8.7.1.5 Modifying IGMP snooping parameters for a SAP or SDP

Use the following CLI syntax to customize IGMP snooping parameters on an existing SAP. Commands for spoke or mesh SDPs are identical.

CLI syntax:

```
config>service# vpls service-id
    sap sap-id
        igmp-snooping
            fast-leave
            import policy-name
            last-member-query-interval interval
            max-num-groups max-num-groups
            mrouter-port
            query-interval interval
            query-response-interval interval
            robust-count count
            send-queries
```

To enable and customize sending of IGMP queries to the hosts:

Example:

```
config>service# vpls 1
    config>service>vpls# sap 1/1/3:0
    config>service>vpls>sap# igmp-snooping
    config>service>vpls>sap>snooping# send-queries
    config>service>vpls>sap>snooping# query-interval 100
    config>service>vpls>sap>snooping# query-response-interval 60
    config>service>vpls>sap>snooping# robust-count 5
    config>service>vpls>sap>snooping# exit
    config>service>vpls>sap# no shutdown
```

To customize the leave delay:

Example:

```
config>service# vpls 1
    config>service>vpls# sap 1/1/1:1
    config>service>vpls>sap# igmp-snooping
    config>service>vpls>sap>snooping# last-member-query-interval 10
    config>service>vpls>sap>snooping# no fast-leave
    config>service>vpls>sap>snooping# exit
    config>service>vpls>sap# exit
```

To enable Fast Leave:

Example:

```
config>service# vpls 1
    config>service>vpls# sap 1/1/1:1
    config>service>vpls>sap# igmp-snooping
    config>service>vpls>sap>snooping# no last-member-query-interval
    config>service>vpls>sap>snooping# fast-leave
    config>service>vpls>sap>snooping# exit
    config>service>vpls>sap# exit
```

To limit the number of streams that a host can join:

Example:

```
config>service# vpls 1
    config>service>vpls# sap 1/1/1:1
    config>service>vpls>sap# igmp-snooping
    config>service>vpls>sap>snooping# max-num-groups 4
    config>service>vpls>sap>snooping# exit
    config>service>vpls>sap# exit
```

To enable sending group reports on a SAP to standby multicast routers:

Example:

```
config>service# vpls 1
    config>service>vpls# sap 1/1/1:1
    config>service>vpls>sap# igmp-snooping
    config>service>vpls>sap>snooping# mrouter-port
    config>service>vpls>sap>snooping# exit
    config>service>vpls>sap# exit
```

The following example displays the modified IGMP snooping configuration on a SAP:

```
A:ALA-48>config>service>vpls>sap>snooping# info detail
----------------------------------------------
                    no fast-leave
                    no import
                    max-num-groups 4
                    last-member-query-interval 10
                    no mrouter-port
                    query-interval 100
                    query-response-interval 60
                    robust-count 5
                    send-queries
----------------------------------------------
A:ALA-48>config>service>vpls>sap>snooping#
```

8.7.2 Configuring static multicast groups on a SAP or SDP

Use the following CLI syntax to add static group membership entries on an existing SAP (commands for spoke or mesh SDPs are identical):

The following displays an example of a static IGMP snooping configuration on a SAP:

```
A:ALA-48>config>service>vpls>sap# info
----------------------------------------------
                max-nbr-mac-addr 4
                igmp-snooping
                    fast-leave
                    mrouter-port
                    static
                        group 239.0.10.10
                            source 10.10.10.1
                            source 10.10.10.2
                        exit
                    exit
                exit
----------------------------------------------
A:ALA-48>config>service>vpls>sap#
```

8.7.2.1 Enabling IGMP group membership report filtering

Routing policies can be defined to limit the multicast channels that can be joined by a host. For example, it is possible to define a policy listing a group of multicast streams (for example, 'basic' containing a basic set of TV channels or 'extended' containing a more extended set of TV channels), and to apply this policy to subscribers of IGMP snooping (SAPs or SDPs).

The following displays an example of a configuration to import a routing policy on a SAP:

```
A:ALA-48>config>service>vpls# info
----------------------------------------------
            stp
                shutdown
            exit
            igmp-snooping
                query-interval 60
                robust-count 5
                report-src-ip 10.20.20.20
```

```
                no shutdown
            exit
            sap 1/1/3:0 create
                igmp-snooping
                    query-interval 100
                    query-response-interval 60
                    robust-count 5
                    send-queries
                exit
            exit
            sap 1/1/3:22 create
                max-nbr-mac-addr 4
                igmp-snooping
                    fast-leave
                    import "test_policy"
                    mrouter-port
                    static
                        group 239.0.10.10
                            source 10.10.10.1
                            source 10.10.10.2
                        exit
                    exit
                exit
            exit
            no shutdown
-----------------------------------------------
A:ALA-48>config>service>vpls#
```

For details configuring a routing policy, see the Configuring Route Policies section in the 7450 ESS, 7750 SR, 7950 XRS, and VSR Router Configuration Guide.

The following shows a sample routing policy configuration accepting IGMP messages for only five multicast channels:

```
A:ALA-48>config>router>policy-options# info
-----------------------------------------------
            prefix-list "basic_channels"
                prefix 239.10.0.1/32 exact
                prefix 239.10.0.2/32 exact
                prefix 239.10.0.3/32 exact
                prefix 239.10.0.4/32 exact
                prefix 239.10.0.5/32 exact
            exit
            policy-statement "test_policy"
                description "basic set of 5 multicast channels"
                entry 1
                    from
                        group-address "basic_channels"
                    exit
                    action accept
                    exit
                exit
                default-action reject
            exit
-----------------------------------------------
A:ALA-48>config>router>policy-options#
```

8.7.2.1.1 Enabling IGMP traffic filtering

For security, it may be advisable to only allow multicast traffic into the SR-series routers from recognized multicast routers and servers. Multicast packets arriving on other interfaces (for example, customer-facing SAPs or spoke SDPs) can be filtered out by defining an appropriate IP filter policy.

For details on how to configure a filter policy, see section Creating an IP Filter Policy in the Router Configuration Guide.

The following example shows a sample IP filter policy configuration dropping all multicast traffic:

```
A:ALA-48>config>filter>ip-filter# info
----------------------------------------------
        ip-filter 1 create
            entry 1 create
                match
                    dst-ip 239.0.0.0/24
                exit
                action accept
            exit
            entry 2 create
                match
                    dst-ip 239.0.0.0/4
                exit
                action drop
            exit
        exit
----------------------------------------------
A:ALA-48>config>filter>ip-filter#
```

The following example shows how to apply this sample IP filter policy to a SAP:

```
A:ALA-48>config>service>vpls # info
----------------------------------------------
          sap 1/1/1:1
              ingress
                  filter ip 1
                  exit
              exit
          exit
----------------------------------------------
A:ALA-48>config>service>vpls>snooping#
```

8.7.3 Configuring multicast VPLS Registration

Use the following CLI syntax to configure Multicast VPLS Registration (MVR). The first step is to register a VPLS as a multicast VPLS.

CLI syntax:

```
config>service# vpls service-id
    igmp-snooping
        mvr
            no shutdown
            description description
            group-policy policy-name
```

Example:

```
config>service# vpls 1000
    config>service>vpls# igmp-snooping
    config>service>vpls>snooping# mvr
    config>service>vpls>snooping>mvr# no shutdown
    config>service>vpls>snooping>mvr# description "MVR VPLS"
    config>service>vpls>snooping>mvr# group-policy "basic_channels_policy"
```

The second step is to configure a SAP to take the multicast channels from the registered multicast VPLS.

CLI syntax:

```
config>service# vpls service-id
    sap sap-id
        igmp-snooping
            mvr
                from-vpls vpls-id
```

Example:

```
config>service# vpls 1
    config>service>vpls# sap 1/1/1:100
    config>service>vpls>sap# igmp-snooping
    config>service>vpls>snooping# mvr
    config>service>vpls>snooping>mvr# from-vpls 1000
```

For MVR by proxy, the destination SAP for the multicast channels should a;sp be configured.

CLI syntax:

```
config>service# vpls service-id
    sap sap-id
        igmp-snooping
            mvr
                from-vpls vpls-id
                to-sap sap-id
```

Example:

```
config>service# vpls 1
    config>service>vpls# sap 1/1/1:100
    config>service>vpls>sap# igmp-snooping
    config>service>vpls>snooping# mvr
    config>service>vpls>snooping>mvr# from-vpls 1000
    config>service>vpls>snooping>mvr# to-sap 1/1/1:200
```

8.7.4 Configuring IGMP, MLD, and PIM in the BSR

Note:
Configuring IGMP, MLD, and PIM in the BSR is supported by the 7750 SR only.

See the *7450 ESS, 7750 SR, 7950 XRS, and VSR Multicast Routing Protocols Guide* for information about multicast and the commands required to configure basic IGMP and PIM parameters.

8.7.4.1 Enabling IGMP

The following displays an example of enabled IGMP.

```
A:LAX>>config>router/service>vprn info detail
...
#------------------------------------------
echo "IGMP Configuration"
#------------------------------------------
        igmp
            query-interval 125
            query-last-member-interval 1
            query-response-interval 10
            robust-count 2
            no shutdown
        exit
#------------------------------------------
...
A:LAX>>config>system#
```

8.7.4.2 Configuring IGMP interface parameters

The following example displays an IGMP configuration:

```
A:LAX>config>router/service>vprn>igmp# info
----------------------------------------------
        interface "lax-sjc"
        exit
        interface "lax-vls"
        exit
        group-interface "Mcast-subscribers"
        exit
----------------------------------------------
A:LAX>config>router>igmp# exit
```

8.7.4.3 Configuring static parameters

The following example displays a configuration to add IGMP a static multicast source:

```
A:LAX>config>router/service>vprn>igmp# info
----------------------------------------------
        interface "lax-sjc"
        exit
        interface "lax-vls"
            static
                group 239.255.0.2
                    source 172.22.184.197
                exit
            exit
        exit
        group-interface "mcast-subscribers"
        interface "lax-vls"
        static
        group 239.255.0.2
        source 172.22.184.197
        exit
        exit
```

```
        exit
        exit----------------------------------------------
A:LAX>config>router>igmp#
```

:The following example displays the configuration to add a IGMP static starg entry:

```
A:LAX>config>router/service>vprn>igmp# info
----------------------------------------------
        interface "lax-sjc"
            static
                group 239.1.1.1
                    starg
                exit
            exit
        exit
        interface "lax-vls"
            static
                group 239.255.0.2
                    source 172.22.184.197
                exit
            exit
        exit
        group-interface "mcast-subscribers"
        static
        group 239.1.1.1
        starg
        exit
        exit
        exit
----------------------------------------------
A:LAX>config>router>igmp#
```

8.7.4.4 Configuring SSM translation

The following displays an SSM translation configuration:

```
A:LAX>config>router/service>vprn>igmp# info
----------------------------------------------
        ssm-translate
            grp-range 239.255.0.1 231.2.2.2
                source 10.1.1.1
            exit
        exit
        interface "lax-sjc"
            static
                group 239.1.1.1
                    starg
                exit
            exit
        exit
        interface "lax-vls"
            static
                group 239.255.0.2
                    source 172.22.184.197
                exit
            exit
        exit
        group-interface "mcast-subscribers"
        static
```

```
        group 239.1.1.1
        starg
        exit
        exit
        exit
-------------------------------------------
A:LAX>config>router/service>vprn>igmp# exit
```

8.7.4.5 Enabling MLD

The following displays an example of enabled MLD.

```
A:LAX>>config>router/service>vprn # info detail

#----------------------------------------
echo "MLD Configuration"
#----------------------------------------
mld
query-interval 125
query-last-member-interval 1
query-response-interval 10
robust-count 2
no shutdown
exit
#----------------------------------------
...
A:LAX>>config>router/service>vprn #
```

8.7.4.6 Configuring MLD interface parameters

The following example displays an MLD configuration:

```
A:LAX>config>router/service>vprn>mld# info
-------------------------------------------
interface "lax-sjc"
exit
interface "lax-vls"
exit
group-interface "mcast-subscribers"
exit
-------------------------------------------
A:LAX>config>router/service>vprn>mld#exit
```

8.7.4.7 Configuring static parameters

The following example displays a configuration to add MLD a static multicast source:

```
A:LAX>config>router/service>vprn>mld# info
-------------------------------------------
interface "lax-sjc"
exit
interface "lax-vls"
       static
               group ffe8::1
```

```
                        source 2001::1
                exit
        exit
exit
group-interface "mcast-subscribers"
        static
                group ffe8::1
                        source 2001::1
                exit
        exit
exit
-----------------------------------------------
A:LAX>config>router/service>vprn>mld#
```

The following example displays the configuration to add a MLD static starg entry:

```
A:LAX>config>router/service>vprn>mld# info
-----------------------------------------------
interface "lax-sjc"
        static
                group ffe8::2
                        starg
                        exit
                exit
        exit
interface "lax-vls"
        static
                group ffe8::1
                        source 2001::1
                        exit
                exit
        exit
group-interface "mcast-subscribers"
        static
                group ffe8::2
                        starg
                        exit
                exit
        exit
exit
-----------------------------------------------
A:LAX>config>router/service>vprn>mld#
```

8.7.4.8 Configuring SSM translation

The following displays an SSM translation configuration:

```
A:LAX>config>router/service>vprn>mld# info
-----------------------------------------------
ssm-translate
        grp-range ff31::1  ff32::1
                source 2001::2
                exit
        exit
interface "lax-sjc"
        static
                group ffe8::2
                        starg
                        exit
```

```
                exit
        exit
interface "lax-vls"
        static
                group ffe8::1
                       source 2001::1
                       exit
                exit
        exit
group-interface "mcast-subscribers"
        static
                group ff31::20
                       starg
                       exit
                exit
        exit
exit
-------------------------------------------
A:LAX>config>router/service>vprn>mld# exit
```

8.7.4.9 Configuring PIM

8.7.4.9.1 Enabling PIM

When configuring PIM, make sure to enable PIM on all interfaces for the routing instance, otherwise multicast routing errors can occur.

The following example displays detailed output when PIM is enabled.

```
A:LAX>>config>router# info detail
...
#---------------------------------------
echo "PIM Configuration"
#---------------------------------------
        pim
            no import join-policy
            no import register-policy
            apply-to none
            rp
                no bootstrap-import
                no bootstrap-export
                static
                exit
                bsr-candidate
                    shutdown
                    priority 0
                    hash-mask-len 30
                    no address
                exit
                rp-candidate
                    shutdown
                    no address
                    holdtime 150
                    priority 192
                exit
            exit
            no shutdown
        exit
```

```
#-----------------------------------------
...
A:LAX>>config>system#
```

8.7.4.9.2 Configuring PIM interface parameters

The following displays a PIM interface configuration:

```
A:LAX>config>router>pim# info
----------------------------------------------
            interface "system"
            exit
            interface "lax-vls"
            exit
            interface "lax-sjc"
            exit
            interface "p1-ix"
            exit
            rp
                static
                    address 10.22.187.237
                        group-prefix 239.24.24.24/32
                    exit
                    address 10.10.10.10
                    exit
                exit
                bsr-candidate
                    shutdown
                exit
                rp-candidate
                    shutdown
                exit
            exit
----------------------------------------------
A:LAX>config>router>pim#


A:SJC>config>router>pim# info
----------------------------------------------
            interface "system"
            exit
            interface "sjc-lax"
            exit
            interface "sjc-nyc"
            exit
            interface "sjc-sfo"
            exit
            rp
                static
                    address 10.22.187.237
                        group-prefix 239.24.24.24/32
                    exit
                exit
                bsr-candidate
                    shutdown
                exit
                rp-candidate
                    shutdown
                exit
            exit
----------------------------------------------
```

```
A:SJC>config>router>pim#


A:MV>config>router>pim# info
----------------------------------------------
            interface "system"
            exit
            interface "mv-sfo"
            exit
            interface "mv-vlc"
            exit
            interface "p3-ix"
            exit
            rp
                static
                    address 210.22.187.237
                        group-prefix 239.24.24.24/32
                    exit
                exit
                bsr-candidate
                    address 10.22.187.236
                    no shutdown
                exit
                rp-candidate
                    address 10.22.187.236
                    no shutdown
                exit
            exit
----------------------------------------------
A:MV>config>router>pim#


A:SFO>config>router>pim# info
----------------------------------------------
            interface "system"
            exit
            interface "sfo-sjc"
            exit
            interface "sfo-was"
            exit
            interface "sfo-mv"
            exit
            rp
                static
                    address 10.22.187.237
                        group-prefix 239.24.24.24/32
                    exit
                exit
                bsr-candidate
                    address 10.22.187.239
                    no shutdown
                exit
                rp-candidate
                    address 10.22.187.239
                    no shutdown
                exit
            exit
----------------------------------------------
A:SFO>config>router>pim#


A:WAS>config>router>pim# info
----------------------------------------------
            interface "system"
```

```
            exit
            interface "was-sfo"
            exit
            interface "was-vlc"
            exit
            interface "p4-ix"
            exit
            rp
                static
                    address 10.22.187.237
                        group-prefix 239.24.24.24/32
                    exit
                exit
                bsr-candidate
                    address 10.22.187.240
                    no shutdown
                exit
                rp-candidate
                    address 10.22.187.240
                    no shutdown
                exit
            exit
-------------------------------------------
A:WAS>config>router>pim#
```

8.7.4.9.3 Importing PIM join and register policies

The import command provides a mechanism to control the (*,g) and (sag) state that gets created on a router. Import policies are defined in the **config**>**router**>**policy-options** context. See Configuring PIM join and register policies.

Note: In the import policy, if an action is not specified in the entry then the default-action takes precedence. If no entry matches then the default-action also takes precedence. If no default-action is specified, then the **default-action** is executed.

The following example displays the command usage to apply the policy statement does not allow join messages for group 239.50.50.208/32 and source 192.168.0.0/16 but allows join messages for 192.168.0.0/16, 239.50.50.208:

Example:

```
config>router# pim
    config>router>pim# import join-policy "foo"
    config>router>pim# no shutdown
```

The following example displays the PIM configuration:

```
A:LAX>config>router>pim# info
-------------------------------------------
            import join-policy "foo"
            interface "system"
            exit
            interface "lax-vls"
            exit
            interface "lax-sjc"
            exit
            interface "p1-ix"
            exit
            rp
```

```
                static
                    address 10.22.187.237
                        group-prefix 239.24.24.24/3
                    exit
                    address 10.10.10.10
                    exit
                exit
                bsr-candidate
                    shutdown
                exit
                rp-candidate
                    shutdown
                exit
            exit
-------------------------------------------
A:LAX>config>router>pim#
```

8.7.4.9.4 Configuring PIM join and register policies

Join policies are used in Protocol Independent Multicast (PIM) configurations to prevent the transportation of multicast traffic across a network and the dropping of packets at a scope at the edge of the network. PIM Join filters reduce the potential for denial of service (DoS) attacks and PIM state explosion—large numbers of Joins forwarded to each router on the RPT, resulting in memory consumption.

*,g or s,g is the information used to forward unicast or multicast packets.

- **group-address** matches the group in join/prune messages

  group-address 239.55.150.208/32 exact
- **source-address** matches the source in join/prune messages

  source-address 192.168.0.0/16 longer
- **interface** matches any join message received on the specified interface:

  interface port 1/1/1
- **neighbor** matches any join message received from the specified neighbor:

  neighbor 1.1.1.1

The following configuration example does not allow join messages for group 239.50.50.208/32 and source 192.168.0.0/16 but allows join messages for 192.168.0.0/16, 239.50.50.208.

```
A:ALA-B>config>router>policy-options# info
-------------------------------------------
...
            policy-statement "foo"
                entry 10
                    from
                        group-address "239.50.50.208/32"
                        source-address 192.168.0.0
                    exit
                    action reject
                exit
            exit
            policy-statement "reg-pol"
                entry 10
                    from
                        group-address "239.0.0.0/8"
                    exit
                    action accept
```

```
                    exit
                exit
            exit
...
-----------------------------------------------
A:ALA-B>config>router>policy-options#
```

8.7.4.10 Configuring bootstrap message import and export policies

Bootstrap import and export policies are used to control the flow of bootstrap messages to and from the RP.

The following configuration example specifies that no BSR messages received or sent out of interface port 1/1/1.

```
:A:ALA-B>config>router>policy-options# policy-statement pim-import
:A:ALA-B>config>router>policy-options>policy-statement$ entry 10
:A:ALA-B>config>router>policy-options>policy-statement>entry$ from
:A:ALA-B>config>router>policy-options>policy-statement>entry>from$ interface port1/
1/1/
:A:ALA-B>config>router>policy-options>policy-statement>entry>from$ exit
:A:ALA-B>config>router>policy-options>policy-statement>entry# action reject
:A:ALA-B>config>router>policy-options>policy-statement>entry# exit
:A:ALA-B>config>router>policy-options>policy-statement# exit

:A:ALA-B>config>router>policy-options# policy-statement pim-export
:A:ALA-B>config>router>policy-options>policy-statement$ entry 10
:A:ALA-B>config>router>policy-options>policy-statement>entry$ to
:A:ALA-B>config>router>policy-options>policy-statement>entry>to$
```

9 Triple Play Enhanced Subscriber Management

9.1 Uniform RADIUS server configuration

9.1.1 RADIUS server configuration

The following two configuration methods coexist but are mutually exclusive:

- Uniform RADIUS server configuration (preferred)
- Legacy RADIUS server configuration

9.1.1.1 Uniform RADIUS server configuration (preferred)

This configuration method is preferred as it can be re-used amongst multiple applications (Subscriber authentication and accounting, L2TP tunnel accounting, WLAN gateway RADIUS proxy) and enables additional functionality not available in the legacy configuration method. For example:

- A RADIUS server policy operational state can be controlled by reception of accounting on or off responses.
- Buffering of accounting messages: When all servers in a **radius-server-policy** are unreachable, it is possible to buffer the acct-stop and acct-interim-update messages for up to 25 hours. When a RADIUS server becomes reachable again then the messages in the buffer are retransmitted.
- A configurable hold down time for accounting servers that are marked down and during which no new communication attempts are made (**hold-down-time**).
- A configurable maximum number of outstanding RADIUS requests for accounting servers (**pending-requests-limit**).
- Increased retry and timeout values for unsuccessful RADIUS communication.
- Enhanced RADIUS server statistics
- IPv6 RADIUS server

Note: A RADIUS server is marked down if it detects a few consecutive timeouts independent of the transaction ID or origin of request.

Where consecutive timeouts are defined by the number of retries configured below the RADIUS server policy servers.

The default number of retries is 3, meaning 1 initial try and 2 retries.

If, for example, the RADIUS server has "2 timeouts, 1 reply, 1 timeouts", whereby the timeouts are originated for the same host, the server is not marked down because intermediate replies were received.

To attach a RADIUS server policy to an authentication policy:

For example,

```
configure
    subscriber-mgmt
        authentication-policy "auth-policy-1" create
            radius-server-policy "aaa-server-policy-1"
        exit
    exit
```

Note: To avoid conflicts, the following CLI commands are ignored in the authentication policy when a **radius-server-policy** is attached:

- All commands in the **radius-authentication-server** context
- **accept-authorization-change**
- **coa-script-policy**
- **accept-script-policy**
- **request-script-policy**
- The **fallback-action** command specifies the action when no RADIUS server is available is configured direct in the **config**>**subscr-mgmt**>**auth-plcy** CLI context.

To attach a RADIUS server policy to a RADIUS accounting policy:

For example:

```
configure
    subscriber-mgmt
        radius-accounting-policy "acct-policy-1" create
            radius-server-policy "aaa-server-policy-1"
        exit
    exit
```

Note: To avoid conflicts, the following CLI commands are ignored in the RADIUS accounting policy when a **radius-server-policy** is attached:

- All commands in the **radius-accounting-server** context
- **acct-request-script-policy**

To configure the RADIUS servers in a RADIUS server policy:

For example:

```
configure
    aaa
        radius-server-policy "aaa-server-policy-1" create
            description "Radius AAA server policy"
            accept-script-policy "script-policy-2"
            acct-on-off oper-state-change
            acct-request-script-policy "script-policy-3"
            auth-request-script-policy "script-policy-1"
            no python-policy
            servers
                access-algorithm direct
                hold-down-time sec 30
                no ipv6-source-address
```

```
                retry 3
                router "Base"
                no source-address
                timeout sec 5
                buffering
                    acct-interim min 60 max 3600 lifetime 5
                    acct-stop min 60 max 3600 lifetime 5
                exit
                server 1 name "server-1"
                server 2 name "server-2"
            exit
        exit
    exit
```

To configure the RADIUS servers in the routing instance:

- In the Base routing instance: **config**>**router**>**radius-server**.
- In a VPRN routing instance: **config**>**service**>**vprn 10**>**radius-server**.
- In the management routing instance (out of band): **config**>**router management**>**radius-server**.

For example:

```
configure
    router
        radius-server
            server "server-1" address 172.16.1.1 secret <shared secret> hash2 create
                accept-coa
                coa-script-policy "script-policy-4"
                description "Radius server 1"
                pending-requests-limit 4096
                acct-port 1813
                auth-port 1812
            exit
            server "server-2" address 172.16.1.2 secret <shared secret> hash2 create
                accept-coa
                coa-script-policy "script-policy-4"
                description "Radius server 2"
                pending-requests-limit 4096
                acct-port 1813
                auth-port 1812
            exit
        exit
    exit
```

Note: To accept inbound RADIUS transactions, such as CoA or Disconnect Messages, configure a RADIUS server in the corresponding routing instance with the **accept-coa** command. To also enable outbound RADIUS transactions, such as authentication and accounting, associate the server with a RADIUS server policy. Association with a RADIUS server policy is not required for a CoA-only server.

9.1.1.2 Legacy RADIUS server configuration

Note: It is recommended to migrate to the uniform RADIUS server configuration as described above to have additional functionality enabled.

To configure a RADIUS server in an authentication policy:

```
configure
    subscriber-mgmt
        authentication-policy "auth-policy-1" create
            radius-authentication-server
                access-algorithm direct
                hold-down-time 30
                retry 3
                no source-address
                timeout 5
                router "Base"
                server 1 address 172.16.1.1 secret <shared secret> hash2 port 1812
                    pending-requests-limit 4096
                server 2 address 172.16.1.2 secret <shared secret> hash2 port 1812
                    pending-requests-limit 4096
            exit
            accept-authorization-change
            accept-script-policy "script-policy-2"
            coa-script-policy "script-policy-4"
            request-script-policy "script-policy-1"
        exit
    exit
```

Note: In a legacy RADIUS server configuration, to configure RADIUS CoA servers for use in Enhanced Subscriber Management, the server must be configured in the authentication policy with the **accept-authorization-change** command enabled. A CoA only server can be configured with the optional coa-only flag.

To configure a RADIUS server in a RADIUS accounting policy:

```
configure
    subscriber-mgmt
        radius-accounting-policy "acct-policy-1" create
            radius-accounting-server
                access-algorithm direct
                retry 3
                timeout 5
                no source-address
                router "Base"
                server 1 address 172.16.1.1 secret <shared secret> hash2 port 1813
                server 2 address 172.16.1.2 secret <shared secret> hash2 port 1813
            exit
            acct-request-script-policy "script-policy-3"
        exit
    exit
```

9.2 RADIUS authentication of subscriber sessions

This section describes the Nokia router acting as a Broadband Subscriber Aggregator (BSA).

Note:
In the TPSDA solutions, the Nokia 5750 Subscriber Services Controller (SSC) serves as the policy manager, DHCP and RADIUS server.

In this application, one of the required functions can be to authenticate users trying to gain access to the network. While sometimes the DHCP server (an SSC) can perform authentication, in most cases a RADIUS server (an SSC) is used to check the customer's credentials.

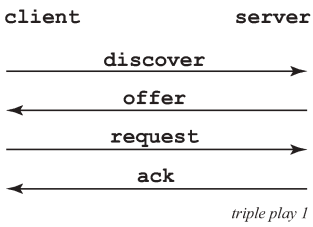
Note:
See the DHCP Management section for information about DHCP and DHCP Snooping.

For information about the RADIUS server selection algorithm, see the *7450 ESS, 7750 SR, 7950 XRS, and VSR System Management Guide*.

If authentication is enabled, the router temporarily holds any received DHCP discover message and sends an access-request message to a configured RADIUS server containing the client's MAC address or circuit-ID (from the Option 82 field) as the username. If access is granted by the RADIUS server, the router then forwards or relays the DHCP discover message to the DHCP server and allows an IP address to be assigned. If the RADIUS authentication request is denied, the DHCP message is dropped and an event is generated.

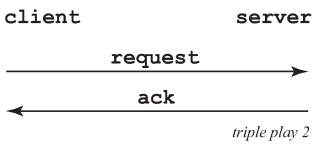
A typical initial DHCP scenario (after client bootup) is shown in Figure 64: Initial DHCP scenario.

Figure 64: Initial DHCP scenario



But, when the client already knows its IP address (when an existing lease is being renewed), it can skip straight to the request/ack phase, as shown in Figure 65: DHCP scenario with known IP address.

Figure 65: DHCP scenario with known IP address



In the first scenario, the DHCP discover triggers an authentication message to RADIUS and the DHCP request also triggers RADIUS authentication. The previous reply is cached for 10 seconds, the second DHCP packet does not result in a RADIUS request.

In the second scenario, the DHCP request triggers an authentication message to RADIUS.

If the optional subscriber management authentication policy re-authentication command is enabled, DHCP authentication is performed at every DHCP lease renew request. Only dynamic DHCP sessions are subject to remote authentication. Statically provisioned hosts are not authenticated.

9.2.1 RADIUS authentication extensions

This section describes an extension to RADIUS functionality in the subscriber management context. As part of subscriber host authentication, RADIUS can respond with access-response message, which, in the

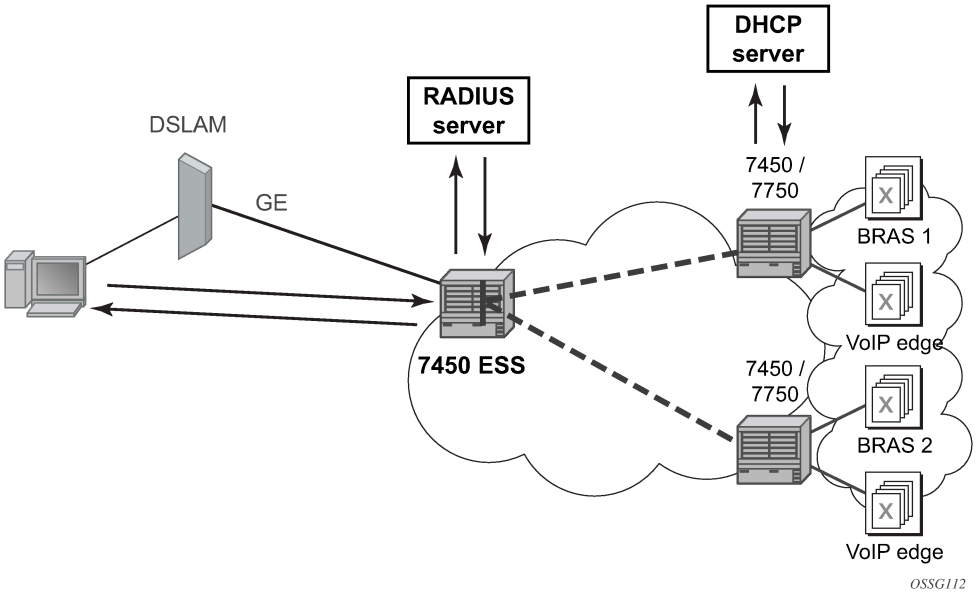
case of an accept, can include several RADIUS attributes (standard and vendor-specific) that allow correct provisioning of a given subscriber-host.

Change-of-Authorization (CoA) messages as defined by RFC 3576, *Dynamic Authorization Extensions to Remote Authentication Dial In User Service (RADIUS)*, are supported. The goal of CoA messages is to provide a mechanism for "mid-session change" support through RADIUS.

9.2.1.1 Triple Play network with RADIUS authentication

Figure 66: Triple Play aggregation network with RADIUS-based DHCP host authentication shows a flow of RADIUS authentication of DHCP hosts in the Triple Play aggregation environment. Besides granting the authentication of specified DHCP host, the RADIUS server can include RADIUS attributes (standard or Vendor-Specific Attributes (VSAs)) which are then used by the network element to provision objects related to a specified DHCP host.

Figure 66: Triple Play aggregation network with RADIUS-based DHCP host authentication



RADIUS is a distributed client/server concept that is used to protect networks against unauthorized access. In the context of the router's subscriber management in TPSDA, the RADIUS client running on nodes sends authentication requests to the SSC.

RADIUS can be used to perform three distinct services:

- Authentication determines whether a specific subscriber-host can access a specific service.
- Authorization associates connection attributes or characteristics with a specific subscriber host.
- Accounting tracks service use by individual subscribers.

The RADIUS protocol uses "attributes" to describe specific authentication, authorization, and accounting elements in a user profile (which are stored on the RADIUS server). RADIUS messages contain RADIUS attributes to communicate information between network elements running a RADIUS client and a RADIUS server.

RADIUS divides attributes into two groups, standard attributes and Vendor-Specific Attributes (VSAs). VSA is a concept allowing conveying vendor-specific configuration information in a RADIUS messages, as discussed in RFC 2865, *Remote Authentication Dial In User Service (RADIUS)*. It is up to the vendor to specify the exact format of the VSAs.

Nokia-specific VSAs are identified by vendor-id 6527.

9.2.2 RADIUS authorization extensions

The following sections define different functional extensions and list relevant RADIUS attributes.

Basic Provisioning of Authentication Extensions

To comply with RFC 4679, *DSL Forum Vendor-Specific RADIUS Attributes*, the software includes the following attributes in the authentication-request message:

- agent-circuit-id (as defined by DSL forum)
- agent-remote-id (as defined by DSL forum)

The following attributes can also be included if configured and provided by downstream equipment:

- actual-data-rate-upstream
- actual-data-rate-downstream
- minimum-data-rate-upstream
- minimum-data-rate-downstream
- access-loop-encapsulation

When the node is configured to insert (or replace) Option 82, the above mentioned attributes do have the content after this operation has been performed by the software.

In addition, the following standard RADIUS attributes are included in authentication request messages (subject to configuration):

- NAS-identifier — string containing system-name
- NAS-port-id
- NAS-port-type — Values: 32 (null encap), 33 (dot1q), 34 (QinQ), 15 (DHCP hosts), specified value (0 — 255)
- MAC-address (Nokia VSA 27)
- dhcp-vendor-class-id (Nokia VSA 36)
- calling-station-id

These are only be included in the access-request if they have been configured.

To provide the possibility to push new policies for currently active subscribers, the routers support commands to force re-authentication of the specified subscriber-host. After issuing such a command, the router sends a DHCP FORCERENEW packet, which causes the subscriber to renew its lease (provided it supports force-renew). The DHCP request and ACK are then authenticated and processed by the routers as they would be during a normal DHCP renew.

9.2.2.1 Calling station ID

A **calling-station-id** can be configured at SAP level and can be included in the RADIUS authentication and accounting messages. This attribute is used in legacy BRAS to identify the user (typically phone number used for RAS connection). In the broadband networks this was replaced by circuit-id in Option 82. However, the Option 82 format is highly dependent on access-node (AN) vendor, which makes interpretation in management servers (such as RADIUS) difficult. Some operators use the calling-station-id attribute as an attribute indicating the way the circuit-id should be interpreted. The **calling-station-id** attribute can be configured as a string which is be configured on the SAP. It can also be configured to use the **sap-id**, **remote-id**, or **mac-address**.

9.2.2.2 Subscriber session timeout

To limit the lifetime of a PPP session or DHCPv4 host to a fixed time interval, a timeout can be specified from RADIUS. By default, a PPP session or DHCPv4 host has no session timeout (infinite).

For PPP sessions, a session-timeout can be configured in the ppp-policy. A RADIUS specified session-timeout overrides the CLI configured value.

```
subscriber-mgmt
    ppp-policy "ppp-policy-1" create
        session-timeout 86400
    exit
exit
```

When the session timeout expires a PPP session is terminated and a DHCPv4 host deleted. For a DHCPv4 host, a DHCP release message is also sent to the server.

The following two attributes can be used in RADIUS Access-Accept and CoA messages to limit the PPP session or DHCPv4 host session time (Table 8: Subscriber session timeout):

Table 8: Subscriber session timeout

Attribute ID	Attribute name	Type	Limits	Purpose and format
27	Session-Timeout	integer	2147483647 seconds	0 = infinite (no session-timeout) (1 to 2147483647) in seconds For example: Session-Timeout = 3600
26-6527-160	Alc-Relative-Session-Timeout	integer	[0 to 2147483647] seconds	0 = infinite (no session-timeout) (1 to 2147483647) in seconds For example: Alc-Relative-Session-Timeout = 3600

When specified in a RADIUS Access-Accept message, both attributes specify an absolute value for session timeout. When specified in a RADIUS CoA message, attribute [26-6527-160] Alc-Relative-Session-Timeout specifies a relative session timeout value in addition to the current session time while attribute [27]

Session-Timeout specifies an absolute session timeout value. If the current session time is greater than the received Session-Timeout, a CoA NAK is sent with error cause "Invalid Attribute Value (407)".

Only one of the above attributes to specify a session timeout can be present in a single RADIUS message. An event is raised when both are specified in a single message.

The output of the **show service id** *service-id* **ppp session detail** command contains following fields related to session timeout for PPP sessions:

- Up Time: the PPP session uptime
- Session Time Left: the remaining time before the session is terminated
- RADIUS Session-TO: the RADIUS received session timeout value.

The output of the **show service id** *service-id* **dhcp lease-state detail** command contains following fields related to session timeout for DHCPv4 hosts:

- Up Time

  the DHCPv4 host uptime
- Remaining Lease Time

  the remaining time before the lease expires in the DHCP server. The client should renew its lease before this time.
- Remaining SessionTime

  the remaining time before the DHCPv4 host is deleted
- Session-Timeout

  the DHCPv4 host is deleted when its uptime reaches the Session-Timeout value.
- Lease-Time

  the lease time specified by the DHCPv4 server

Note: In a **radius-proxy** scenario or when a DHCPv4 host is created with a RADIUS CoA message, the RADIUS attribute [26-6527-174] Alc-Lease-Time attribute must be used to specify the lease time. If the [26-6527-174] Alc-Lease-Time is not present in these scenarios, then the RADIUS attribute [27] Session-Timeout is interpreted as DHCPv4 lease time.

9.2.2.2.1 Domain name in authentication

In many networks, the username has specific meaning with respect to the domain (ISP) where the user should be authenticated. To identify the user correctly, the username in an authentication-request message should contain a domain name. The domain name can be derived from different places. In PPPoE authentication the domain name is provided by the PPPoE client with the username used in PAP or CHAP authentication. For DHCP hosts similar functionality is implemented by a "pre-authentication" lookup in a local user database before performing the RADIUS request.

For example, it can be derived from option60 which contains the vendor-specific string identifying the ISP the set-box has been commissioned by.

To append a domain name to a DHCP host, the following configuration steps should be taken:

- Under the (group or IP) interface of the service, a local user database should be configured in the DHCP node and no authentication policy should be configured.

- In the local user database, there should be a host entry containing both the domain name to be appended and an authentication policy that should be used for RADIUS authentication of the host. The host entry should contain no other information needed for setting up the host (IP address, ESM string), otherwise the DHCP request is dropped.
- In the authentication policy, the **user-name-format** command should contain the parameter **append** *domain-name*.

9.2.2.3 RADIUS reply message for PPPoE PAP/CHAP

The string returned in a [18] Reply-Message attribute in a RADIUS Access-Accept is passed to the PPPoE client in the CHAP Success or PAP Authentication-Ack message.

The string returned in a [18] Reply-Message attribute in a RADIUS Access-Reject is passed to the PPPoE client in the CHAP Failure or PAP Authentication-Nak message.

When no [18] Reply-Message attribute is available, the SR OS default messages are used instead: "CHAP authentication success" or "CHAP authentication failure" for CHAP and "Login ok" or "Login incorrect" for PAP.

9.2.2.4 SHCV policy

SHCV policies are used to control subscriber host connectivity verification which verifies the host connectivity to the BNG. There are two types of SHCV: periodic and event triggered. Before Release 13.0.R4, some event triggered SHCV relied on the reference timer set by the host-connectivity-verify under the group interface while others had hard-coded values. Release 13.0.R4 introduced the SHCV policy that allows individual configuration of trigger SHCV timers and periodic SHCV timers depending on the application.

Under the group-interface, the **host-connectivity-verify** configuration was used as a reference timer for some event triggered SHCV while other used hard-coded values. The SHCV policy separated out every type of SHCV and allows each type to have their individual configurable timer values. Furthermore, individual SHCV trigger types can be shut down. The SHCV policy can be applied to one or more group interfaces and can be configured differently for IPv4 vs. IPv6 hosts. There are various types of triggered SHCV:

- **ip-conflict**

  This SHCV is sent in the following scenarios:

  – When the subscriber connects a new replacement residential gateway (RG) with a new MAC address on the same SAP on a system that still holds the subscriber's previous RG state. The new RG is assigned the same IP address or prefix as the previous RG (for example, with RADIUS address assignment). The system detects this IP conflict and triggers an SHCV to identify if the previous RG is still connected.

  – When the system does not detect an RG reboot and holds the subscriber's previous RG state, and the IPv4 address assignments are performed by an external DHCP server, the system detects an IP conflict when the DHCP server assigns a new IPv4 address that differs from the previously assigned IPv4 address. This conflict triggers an SHCV to identify if the previous IP address is still in use. This specific use case, where the same device is assigned two different IP addresses, is supported for IPv4 host only.

- **host-limit-exceeded**

Sent when a subscriber has exceeded a configured host or session limit. Host limits are set in the **sla-profile host-limits** and in the **sub-profile host-limits**. Session limits are set in the **group-interface ipoe-session sap-session-limit** and **session-limit**, in the **sla-profile session-limits** and in the **sub-profile session-limits** contexts.

- **inactivity**

  The **category-map** configured under **sla-profile** can trigger an SHCV when the subscriber host becomes idle.

- **mobility**

  Intended for mobility applications such as Wi-Fi. When a subscriber moves between SAPs and requests for the same IP address, a triggered SHCV is sent to verify if the old host is still connected before removing the old host entry.

- **mac-learning**

  For IP-only static-host MAC learning. The trigger SHCV is sent to learn the subscriber MAC when a **no shutdown** command is executed on the CLI for the static host.

Some SHCVs are triggered based on a host's DHCP messages. These DHCP messages are not buffered. The SHCV is used only to perform a verification check on an old host to verify if the host is still connected to the BNG. Therefore, the BNG still requires the new hosts to retransmit their DHCP messages after the SHCV removes the disconnected host.

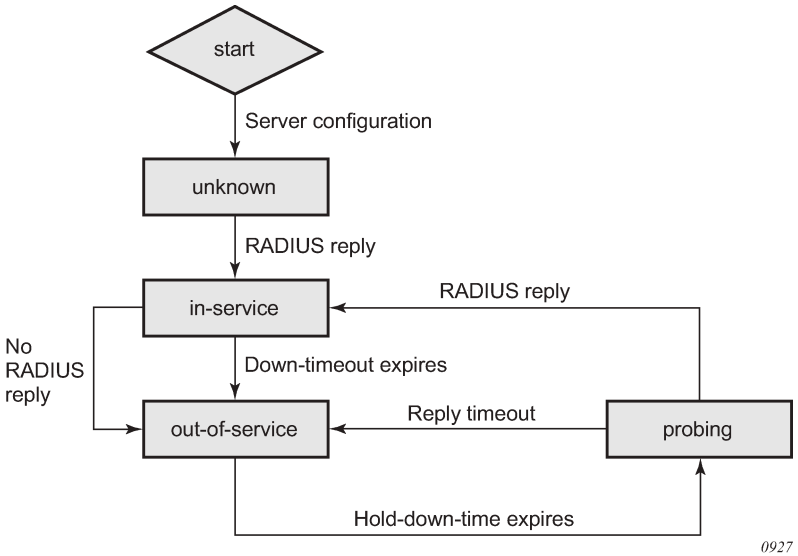
9.2.3 radius-server-policy retry attempt overview

This feature maximizes the use of the remaining healthy RADIUS servers for subscriber authentication and accounting. After the hold-down time expires, a single RADIUS message is used to determine the status of the RADIUS server. If the server remains unresponsive after waiting for a single timeout interval (without any retries), then it is placed back into the hold-down state. If the RADIUS server responds, then it is used for subscriber authentication and accounting with the rest of the healthy servers.

9.2.4 AAA RADIUS server operation status

The different operating states of a RADIUS server are shown in Figure 67: RADIUS server operating states. When a RADIUS server is first provisioned into the AAA using the **radius-server-policy** command, the operating state is "unknown". This state indicates that the RADIUS server has yet to receive a RADIUS request message. To send a request message, the **radius-server-policy** command provides three different access algorithms: direct, round-robin, and hash. With the direct algorithm, request messages are always sent to the in-service RADIUS server with the lowest configured server index. With the round-robin algorithm, the RADIUS requests are load-balanced in a round-robin manner. The hash algorithm offers a load-balanced alternative; the 7750 SR generates a hash-key based on the subscriber information, and the RADIUS request is then sent to a server based on the hash key. The hash method differs from the round-robin method in that, under normal working conditions, RADIUS requests from a particular subscriber are always forwarded to the same RADIUS server. When a server replies to a RADIUS request, it transitions from the operational state of "unknown" to "in-service". A server may transition from "unknown" to "out-of-service" if the server fails to respond to the initial RADIUS message.

Figure 67: RADIUS server operating states



A RADIUS server is declared "out-of-service" when the **down-timeout** timer expires. The router starts the **down-timeout** timer when an access-request is sent. The timer only resets to "0" when a reply is received from the RADIUS server. This means that the timer can be reset to "0" if a reply message is received for another subscriber. For example, the RADIUS server may miss a message but stay "in-service" if the server responds to another access request from a different subscriber or from a retry of the same subscriber, if the reply is received within the **down-timeout** interval.

Note: It is highly recommended that the **down-timeout** command be set to its default value.

The **down-timeout** default value is the timeout value multiplied by the number of retry attempts. The timeout value is the time that the router waits for the RADIUS server to reply, and the retry value is the number of attempts the 7750 SR makes to contact the RADIUS server. If the RADIUS server remains unresponsive, the timer continues to increment until it reaches the configured **down-timeout** value and the server is declared "out-of-service".

For RADIUS servers that do not respond to all RADIUS requests, a test user account can be optionally set up to periodically send RADIUS request messages to keep the server in service. Typically, a RADIUS server should always respond to all access requests. However, creating a test user account for periodic keep-alive may place an unnecessary load on the processor and may lower the overall scale of the router.

At the start of the out-of-service state, a **down-timeout** timer starts. The timer holds down the RADIUS server and prevents it from operating; no RADIUS messages are sent to an out-of-service server. This is beneficial for the following reasons.

- The server may be unresponsive because of excessive RADIUS message requests; holding it down allows the server to recover.
- Holding down an unresponsive server allows other healthy RADIUS servers to service new requests promptly.

After the **hold-down-time** timer expires, the server enters into the "probing" state. There must be multiple RADIUS servers and at least one healthy server for the server to enter the probing state. Probing is always performed by the test user account; actual subscriber requests are never used during probing.

If no test user account exists, an actual subscriber request is used to perform the probe. There are no retry attempts; only a single RADIUS message is used to probe a RADIUS server. If the RADIUS server responds, it is declared "in-service" immediately. If the RADIUS server fails to respond within the timeout value, it is declared "out-of-service" again and the **hold-down-time** timer restarts. Subscriber RADIUS messages used for probing are not cached, and if the server fails to respond, the subscriber is required to send the RADIUS message again by sending an address request; for example, DHCP, PPP, or Stateless Address Auto-Configuration (SLAAC) or by performing a **data-trigger**.

9.2.5 AAA RADIUS accounting server stickiness

Stickiness applies to the following subscriber RADIUS accounting sessions: start, interim, and stop. By default, the subscriber sticks with the server that served its last accounting message. For example, if server 1 served the subscriber an accounting start message, then the subsequent interim messages and stop message from the same subscriber is sent to server 1. If server 1 is out of service, server 2 is used for the subsequent interim and stop messages. When server 1 recovers, the interim and stop messages sticks with server 2. The RADIUS accounting messages are always be forwarded to the server that serviced the subscriber's last accounting message.

Typically, when using the direct access algorithm, the primary server (lowest configured server index) serves all RADIUS request messages. The other RADIUS servers are used for backup purposes only and may be using a lighter-weight processor. Therefore, it is best to revert to the primary server as soon as it is restored. This can be accomplished by disabling stickiness in direct mode; the RADIUS accounting messages are forwarded to the primary server after it is restored.

In a round-robin algorithm, while each subscriber session is assigned to a different server in round-robin order, a particular subscriber sticks with a server for the entire accounting session. Disabling stickiness sends a subscriber's RADIUS accounting messages to the list of configured RADIUS servers in a round-robin order.

9.2.6 AAA RADIUS authentication fallback action

The fallback action comes into effect when connectivity to all RADIUS servers is lost. The operating state of the RADIUS servers changes to either "out-of-service" or "probing". There are two configurable fallback actions: **accept** or **user-db**. An accept action without force-probing automatically accepts all authentication requests from all subscribers. A user-db action without force-probing uses the **local-user-db** for subscriber authentication.

Both **accept** and **user-db** can be combined with the **force-probing** command. Force-probing forces the out-of-service server to transition to the probing state immediately, bypassing the **hold-down-time** timer. Force-probing is a mechanism to promptly restore connectivity to a RADIUS server. A test user is not used to perform a force probe; only actual subscriber authentication is used to test the operating state of the RADIUS server. Probing only occurs when a server is out of service. If all servers are in the probing state, all new incoming authentication requests follow the fallback action immediately.

When probing with an actual subscriber authentication, the 7750 SR only waits for a reply for one timeout interval without any retries. During the wait, the server is in a probing state and no other subscribers are used to probe this server. The subscriber authentication request is not cached when used for probing. Therefore, to trigger authentication again, the subscriber is required to authenticate again with an address request or a data-trigger packet.

9.2.7 AAA test user account

A test user account is used in the rare case where a RADIUS server ignores RADIUS messages as mentioned in the AAA RADIUS server operation status section. Consequently, when messages are ignored, the router places the RADIUS server out of service. The test user account can keep a RADIUS server in service by periodically sending RADIUS requests to the server. The RADIUS server, while randomly ignoring other subscriber RADIUS requests, must respond to the test user requests. A RADIUS server is in service if it replies to RADIUS messages before the **down-timeout** timer expires. The default **down-timeout** default value is the timeout value multiplied by the retry value, but it is also configurable. The test user account has a configurable interval value, and it is recommended that this value be configured to be less than the **down-timeout** value for it to be useful. The test user account only applies to RADIUS authentication.

Typically, a RADIUS server always responds to all RADIUS requests, and therefore it is not recommended that a test user account be used unless it is absolutely necessary for specific types of servers. The test user account creates extra load for the processor and can affect scaling. The test user account can be used with a Python script (for example, adding additional attributes to the test user account during an **access-request** operation).

9.2.8 Troubleshooting the RADIUS server

The **tools**>**perform**>**security**>**authentication-server-check** command can be used to troubleshoot a RADIUS server by checking the connectivity and functional status of a RADIUS server for subscriber management operations. The command keyword **debug** can be specified to view more information about the access request. All VSAs sent and received from the RADIUS server, the hex dump, and all other debug information can be shown without the need to turn on system-wide debugging.

Additional attributes in an Access-Request message can be specified in an attribute file referenced with the command keyword **attr-from-file** *file-url*. Each attribute must be specified on a separate line in the text file in the following format shown in Table 9: authentication-server-check attribute file format .

Table 9: authentication-server-check attribute file format

Attribute file format	Description
<type> = <value>	Standard attribute
<vendor>,<type> = <value>	Vendor Specific Attribute
e,<type>,<ext-type> = <value>	Extended type attribute (RFC 6929)
evs, <type>, <vendor>, <vendortype> = <value>	Extended Vendor Specific attribute (RFC 6929)
le,<type>,<ext-type> = <value>	Long Extended type attribute (RFC6929)
evs, <type>, <vendor>, <vendortype> = <value>	Long Extended Vendor Specific attribute (RFC 6929)

9.2.9 Provisioning of Enhanced Subscriber Management (ESM) objects

In the ESM concept on network elements, a subscriber host is described by the following aspects:

- subscriber-id-string
- subscriber-profile-string
- sla-profile-string
- ancp-string
- intermediate-destination-identifier-string
- application-profile-string

This information is typically extracted from DHCP-ACK message using a Python script, and is used to provision subscriber-specific resources such as queues and filter entries. As an alternative to extracting this information from DHCP-ACK packet, provisioning from RADIUS server is supported.

As a part of this feature, the following VSAs have been defined:

- **alc-subscriber-id-string**

  Contains a string which is interpreted as a subscriber-id.
- **alc-subscriber-profile-string**

  Contains a string which is interpreted as a subscriber profile
- **alc-sla-profile-string**

  Contains string which is interpreted as an SLA profile.
- **alc-ancp-string**

  Contains string which is interpreted as an ANCP string.
- **alc-int-dest-id-string**

  Contains a string which is interpreted as an intermediate destination ID
- **alc-app-profile-string**

  Contains a string which is interpreted as an application profile

  Note that these strings can be changed in a CoA request.

When RADIUS authentication response messages contain the above VSAs, the information is used during processing of DHCP-ACK message as an input for the configuration of subscriber-host parameters, such as QoS and filter entries.

If ESM is not enabled on a specified SAP, information in the VSAs is ignored.

If ESM is enabled and the RADIUS response does not include all ESM-related VSAs (an ANCP string is not considered as a part of ESM attributes), only the **subscriber-id** is mandatory (the other ESM-related VSAs are not included). The remaining ESM information (sub-profile, sla-profile) is extracted from DHCP-ACK message according to normal flow (Python script, and so on).

If the profiles are missing from RADIUS, they are not extracted from the DHCP data with Python to prevent inconsistent information. Instead, the data reverts to the configured default values.

However, if the above case, a missing subscriber ID causes the DHCP request to be dropped. The DHCP server is not queried in that case.

When no DHCP server is configured, DHCP-discover/request messages are discarded.

9.2.9.1 Provisioning IP configuration of the host

The other aspect of subscriber-host authorization is providing IP configuration (ip-address, subnet-mask, default gateway and dns) through RADIUS directory instead of using centralized DHCP server. In this case, the node receiving following RADIUS attributes assumes the role of DHCP server in conversation with the client and provide the IP configuration received from RADIUS server.

These attributes are accepted only if the system is explicitly configured to perform DHCP server functionality on a specific interface.

The following RADIUS attributes are accepted from authentication-response messages:

- **framed-ip-address**

  The IP address to be configured for the subscriber-host
- **framed-ip-netmask**

  The IP network to be configured for the subscriber host If RADIUS does not return a netmask, the DHCP request is dropped
- **framed-pool**

  The pool on a local DHCP server from which a DHCP-provided IP address should be selected
- **alc-default-router**

  The address of the default gateway to be configured on the DHCP client
- **alc-primary-dns**

  The DNS address to be provided in DHCP configuration.
  - Juniper VSA for primary DNS
  - Redback VSA for primary DNS
- alc-secondary-dns
  - Juniper VSA for secondary DNS
  - Redback VSA for secondary DNS
- **alc-lease-time**

  Defines the lease time
- session-timeout — Defines the lease time in absence of the alc-lease-time attribute
- NetBIOS
  - **alc-primary-nbns**
  - **alc-secondary-nbns**

9.2.9.2 RADIUS-based authentication in wholesale environment

To support VRF selection, the following attributes are supported:

- Alc-Retail-Serv-Id
  Indicates the service ID of the required retail VPRN service configured on the system.
- Alc-Retail-Serv-Name
  Indicates the service name of the required retail VPRN service configured on the system.

Alc-Retail-Serv-Name takes precedence over Alc-Retail-Serv-Id if both are specified.

9.2.9.3 Change of authorization and disconnect-request

In a typical RADIUS environment, the network element serves as a RADIUS client, which means the messages are originated by a routers. In some cases, such as mid-session changes, it is desirable that the RADIUS server initiates a CoA request to impose a change in policies applicable to the subscriber, as defined by RFC 3576.

To configure a RADIUS server to accept CoA and Disconnect Messages is achieved in one of the following ways:

1. Configure up to 64 RADIUS CoA servers per routing instance:

```
config>router>radius-server#
config>service>vprn>radius-server#
 server "coa-1" address 10.1.1.1 secret <shared-secret> hash2 create
     accept-coa
 exit
```

   This is the preferred method.

2. Configure up to 16 RADIUS CoA servers per authentication policy.

```
config>subscr-mgmt>auth-plcy#
 accept-authorization-change
```

   The UDP port for CoA and Disconnect Messages is configurable per system with the command:

```
config>aaa#
radius-coa-port {1647|1700|1812|3799}
```

Note:
There is a priority in the functions that can be performed by CoA. The first matching one is performed:

- If the CoA packet contains a Force-Renew attribute, the subscriber gets a FORCERENEW DHCP packet. This function is not supported for PPPoE or ARP hosts.
- If the CoA packet contains a create-host attribute, a new lease-state is created. Only DHCP lease-states can be created by a CoA message. PPPoE sessions and ARP hosts cannot be created.
- Otherwise, the ESM strings are updated.

There are several reasons for using RADIUS initiated CoA messages:

1. Changing ESM attributes (SLA or subscriber profiles) or queues/policers/schedulers rates of the specific subscriber host. CoA messages containing the identification of the specified subscriber-host along with new ESM attributes.
2. Changing (or triggering the change) of IP configuration of the specified subscriber-host. CoA messages containing the identification of the specified subscriber-host along with VSA indicating request of FORCERENEW generation.
3. Configuring new subscriber-host. CoA messages containing the full configuration for the specific host.

If the changes to ESM attributes are required, the RADIUS server sends CoA messages to the network element requesting the change in attributes included in the CoA request:

- attributes to identify a single or multiple subscriber hosts: NAS-Port-Id + IP address/prefix or Acct-Session-Id or Alc-Subsc-ID-Str
  - Nas-Port-Id attribute + single IP address/prefix attribute:
    - Framed-IP-Address
    - Alc-Ipv6-Address
    - Framed-Ipv6-Prefix
    - Delegated-Ipv6-Prefix
    - Alc-Client-Hardware-Addr (Required for private-retail subnet. For more information, see the 7450 ESS, 7750 SR, and VSR RADIUS Attributes Reference Guide.)
  - Acct-Session-Id (number format)
  - Alc-Subsc-ID-Str
  - User-Name (Possible to use in combination with the following. For more information, see the *7450 ESS, 7750 SR, and VSR RADIUS Attributes Reference Guide*.)
    - Framed-Ip-Address
    - Alc-Ipv6-Address
    - Framed-Ipv6-Prefix
    - Delegated-Ipv6-Prefix
    - Alc-Client-Hardware-Addr
- alc-subscriber-profile-string
- alc-sla-profile-string
- alc-ancp-string
- alc-app-profile-string
- alc-int-dest-id-string
- alc-subscriber-id-string
- alc-subscriber-qos-override

Note: If the subscriber-id-string is changed while the ANCP string is explicitly set, the ANCP-string must be changed simultaneously. When changing the **alc-subscriber-id-string**, the lease state is temporarily duplicated, causing two identical ANCP-strings to be in the system at the same time. This is not allowed.

As a reaction to such message, the router changes the ESM settings applicable to the specified host.

If changes to the IP configuration (including the VRF-ID in the case of wholesaling) of the specified host are needed, the RADIUS server may send a CoA message containing VSA indicating request for FORCERENEW generation:

- attributes to identify a single or multiple subscriber hosts: "NAS-Port-Id + IP address/prefix" or "Acct-Session-Id" or "Alc-Subsc-ID-Str" or "user-name":
  - Nas-Port-Id attribute + single IP address/prefix attribute:

    - Framed-IP-Address
    - Alc-Ipv6-Address
    - Framed-Ipv6-Prefix
    - Delegated-Ipv6-Prefix
  - Acct-Session-Id (number format)
  - Alc-Subsc-ID-Str
  - User-Name
- alc-force-renew
- alc-force-nak

As a reaction to a message, router generates a DHCP FORCERENEW message for the specified subscriber host. Consequently, during the re-authentication, new configuration parameters can be populated based on attributes included in Authentication-response message. The force-NAK attribute has the same function as the Force-Renew attribute, but causes the BNG to reply with a NAK to the next DHCP renew. This invalidates the lease state on the BNG and force the client to completely recreate its lease, making it possible to update parameters that cannot be updated through normal CoA messages, such as IP address or address pool.

If the configuration of the new subscriber-host is required, RADIUS server sends a CoA message containing VSA request new host generation along with VSAs specifying all required parameters.

- **alc-create-host**
- **alc-subscriber-id-string**

  This attribute is mandatory in case ESM is enabled, and optional for new subscriber host creation otherwise.
- **NAS-port-id**

  This attribute indicates the SAP where the host should be created.
- **framed-ip-address**

  the framed IP address
- **alc-client-hw-address**

  A string in the xx:xx:xx:xx:xx:xx format. This attribute is mandatory for new subscriber-host creation.
- **alc-lease-time**

  Specifies the lease time. If both **session-timeout** and **alc-lease-time** are not present, then a default lease time of 7 days is used.
- **session-timeout**

  Specifies the lease time in absence of the **alc-lease-time** attribute. If both **session-timeout** and **alc-lease-time** are not present, then a default lease time of 7 days is used.
- **alc-retail-svc-id**

  This is only used in case of wholesaling for selection of the retail service. Indicates the *service-id* of the required retail VPRN service configured on the system.
- Optionally other VSAs describing a specified subscriber host. If the ESM is enabled, but the CoA message does not contain ESM attributes, the new host is not created.

After executing the requested action, the router element responds with an ACK or NAK message depending on the success/failure of the operation. In case of failure (and then, a NAK response), the element includes the error code in accordance with RFC 3576 definitions if an appropriate error code is available.

Supporting CoA messages has security risks as it essentially requires action to unsolicited messages from the RADIUS server. This can be primarily the case in an environment where RADIUS servers from multiple ISPs share the same aggregation network. To minimize the security risks, the following rules apply:

- Support of CoA messages is disabled by default. They can be enabled on a per RADIUS server or authentication-policy basis.
- When CoA is enabled, the node listens and react only to CoA messages received from RADIUS servers. In addition, CoA messages must be protected with the key corresponding to the specified RADIUS server. All other CoA messages are silently discarded.

In all cases (creation, modification, force-renew) subscriber host identification attributes are mandatory in the CoA request: "NAS-Port-Id + IP" or "Acct-Session-Id" or "Alc-Subsc-ID-Str" or "user-name".

- Nas-Port-Id + single IP address/prefix:
  - Nas-Port-Id
  - Framed-IP-Address
  - Alc-Ipv6-Address
  - Framed-Ipv6-Prefix
  - Delegated-Ipv6-Prefix
  - Alc-Client-Hardware-Addr (may be required for private retail subnet. For more information, see the *RADIUS Attributes Reference Guide*.)
- Acct-Session-Id (number format)
- Alc-Subsc-ID-Str
- User-Name (Possible to use in combination with the following. For more information, see the *7450 ESS, 7750 SR, and VSR RADIUS Attributes Reference Guide*.)
  - Framed-Ip-Address
  - Alc-Ipv6-Address
  - Framed-Ipv6-Prefix
  - Delegated-Ipv6-Prefix
  - Alc-Client-Hardware-Addr

When there are no subscriber host identification attributes present in the CoA, the message is NAK'd with corresponding error code.

- Receiving CoA message with the same attributes as currently applicable to the specified host responds with an ACK message.
- In case of dual homing (through SRRP), the RADIUS server should send CoA messages to both redundant nodes and this with all corresponding attributes (NAS-port-id with its local meaning to corresponding node).
- In the case of change requests, the node which has the specified host active (active SAP or SAP associated with a group interface in SRRP master state) processes the RADIUS message and reply to RADIUS. The standby node always replies with a NAK.

- In the case of create requests, the active node (the SAP described by NAS-port-id is active or associated with a group-interface in SRRP master state). Both nodes reply, but the standby NAKs the request.

The properties of an existing RADIUS-authenticated PPPoE session can be changed by sending a Change of Authorization (CoA) message from the RADIUS server. Processing of a CoA is done in the same way as for DHCP hosts, with the exception that only the ESM settings can be changed for a PPPoE session (the Force-Renew attribute is not supported for PPPoE sessions and a Create-Host CoA always generates a DHCP host).

For terminating PPPoE sessions from the RADIUS server, the disconnect-request message can be sent from the RADIUS server. This message triggers a shut down of the PPPoE session. The attributes needed to identify the PPPoE session are the same as for DHCP hosts.

9.2.9.3.1 Change of authorization using the tools command

A CoA can be triggered through the CLI by using a **tools** command that does not require a RADIUS authentication policy. The **tools** command can also be used to spoof a CoA from a configured server for purposes such as testing CoA python scripts. However, when spoofing the CoA from a RADIUS server, the configuration of a RADIUS authentication policy is required.

The **tools** command, **tools**>**perform**>**subscriber-mgmt**>**coa**, supports up to five different VSAs. If more than five VSAs are required, a file with more than five VSAs can be used for execution.

The **tools** command does not support lawful intercept attributes.

SNMP can also trigger the **tools** CoA command. However, SNMP cannot execute the command when it is referencing an on-board flash file. To execute from a file, the file must be non-local, such as using a URL specifying the location of the file on an FTP server.

Only one **tools** command, **tools**>**perform**>**subscriber-mgmt**>**coa** command can be performed at a time. The command must complete execution before processing a new one. If the **tools** command becomes unresponsive, CTRL-c can be used to break out of the CoA. In addition, a failsafe mechanism automatically terminates the **tools** command if it has not completed within a minute.

9.2.9.4 RADIUS-based accounting

When a router is configured to perform RADIUS-based accounting, at the creation of a subscriber-host, it generates an accounting-start packet describing the subscriber-host and sends it to the RADIUS accounting server. At the termination of the session, it generates an accounting-stop packet including accounting statistics for a specified host. The router can also be configured to send an interim-accounting message to provide updates for a subscriber-host.

The exact format of accounting messages, their types, and communication between client running on the routers and RADIUS accounting server is described in RFC 2866, *RADIUS Accounting*. The following describes a few specific configurations.

To identify a subscriber-host in accounting messages different RADIUS attributes can be included in the accounting-start, interim-accounting, and accounting-stop messages. The inclusion of the individual attributes is controlled by the following commands.

```
configure
    subscr-mgmt
        radius-accounting-policy <name>
            include-radius-attribute
```

```
                [no] acct-authentic
                [no] acct-delay-time
                [no] called-station-id
                [no] calling-station-id
                [no] circuit-id
                [no] delegated-ipv6-prefix
                [no] dhcp-vendor-class-id
                [no] framed-interface-id
                [no] framed-ip-addr
                [no] framed-ip-netmask
                [no] framed-ipv6-prefix
                [no] framed-route
                [no] framed-ipv6-route
                [no] ipv6-address
                [no] mac-address
                [no] nas-identifier
                [no] nas-port
                [no] nas-port-id
                [no] nas-port-type
                [no] nat-port-range
                [no] remote-id
                [no] sla-profile
                [no] sub-profile
                [no] subscriber-id
                [no] tunnel-server-attrs
                [no] user-name
                [no] wifi-rssi
                [no] alc-acct-triggered-reason
                [no] access-loop-options
                [no] all-authorized-session-addresses
                [no] detailed-acct-attributes
                [no] std-acct-attributes
                [no] v6-aggregate-stats
```

RADIUS volume accounting attributes are depending on the type of volume reporting and can be controlled with an **include-radius-attribute** CLI command. Multiple volume reporting types can be enabled simultaneously:

```
config
    subscr-mgmt
        radius-accounting-policy <name>
            include-radius-attribute
                [no] detailed-acct-attributes
                [no] std-acct-attributes
                [no] v6-aggregate-stats
```

where:

detailed-acct-attributes — Report detailed per queue and per policer counters using RADIUS VSAs (enabled by default). Each VSA contains a queue or policer ID followed by the stat-mode or 64 bit counter. The VSA's included in the Accounting messages is function of the context (policer or queue, stat-mode, MDA type, and so on):

[26-6527-107] Alc-Acct-I-statmode

[26-6527-127] Alc-Acct-O-statmode

[26-6527-19] Alc-Acct-I-Inprof-Octets-64

[26-6527-20] Alc-Acct-I-Outprof-Octets-64

[26-6527-21] Alc-Acct-O-Inprof-Octets-64
[26-6527-22] Alc-Acct-O-Outprof-Octets-64
[26-6527-23] Alc-Acct-I-Inprof-Pkts-64
[26-6527-24] Alc-Acct-I-Outprof-Pkts-64
[26-6527-25] Alc-Acct-O-Inprof-Pkts-64
[26-6527-26] Alc-Acct-O-Outprof-Pkts-64
[26-6527-39] Alc-Acct-OC-O-Inprof-Octets-64
[26-6527-40] Alc-Acct-OC-O-Outprof-Octets-64
[26-6527-43] Alc-Acct-OC-O-Inprof-Pkts-64
[26-6527-44] Alc-Acct-OC-O-Outprof-Pkts-64
[26-6527-69] Alc-Acct-I-High-Octets-Drop_64
[26-6527-70] Alc-Acct-I-Low-Octets-Drop_64
[26-6527-71] Alc-Acct-I-High-Pack-Drop_64
[26-6527-72] Alc-Acct-I-Low-Pack-Drop_64
[26-6527-73] Alc-Acct-I-High-Octets-Offer_64
[26-6527-74] Alc-Acct-I-Low-Octets-Offer_64
[26-6527-75] Alc-Acct-I-High-Pack-Offer_64
[26-6527-76] Alc-Acct-I-Low-Pack-Offer_64
[26-6527-77] Alc-Acct-I-Unc-Octets-Offer_64
[26-6527-78] Alc-Acct-I-Unc-Pack-Offer_64
[26-6527-81] Alc-Acct-O-Inprof-Pack-Drop_64
[26-6527-82] Alc-Acct-O-Outprof-Pack-Drop_64
[26-6527-83] Alc-Acct-O-Inprof-Octs-Drop_64
[26-6527-84] Alc-Acct-O-Outprof-Octs-Drop_64
[26-6527-91] Alc-Acct-OC-O-Inpr-Pack-Drop_64
[26-6527-92] Alc-Acct-OC-O-Outpr-Pack-Drop_64
[26-6527-93] Alc-Acct-OC-O-Inpr-Octs-Drop_64
[26-6527-94] Alc-Acct-OC-O-Outpr-Octs-Drop_64
[26-6527-108] Alc-Acct-I-Hiprio-Octets_64
[26-6527-109] Alc-Acct-I-Lowprio-Octets_64
[26-6527-110] Alc-Acct-O-Hiprio-Octets_64
[26-6527-111] Alc-Acct-O-Lowprio-Octets_64
[26-6527-112] Alc-Acct-I-Hiprio-Packets_64
[26-6527-113] Alc-Acct-I-Lowprio-Packets_64
[26-6527-114] Alc-Acct-O-Hiprio-Packets_64
[26-6527-115] Alc-Acct-O-Lowprio-Packets_64

[26-6527-116] Alc-Acct-I-All-Octets_64

[26-6527-117] Alc-Acct-O-All-Octets_64

[26-6527-118] Alc-Acct-I-All-Packets_64

[26-6527-119] Alc-Acct-O-All-Packets_64

std-acct-attributes

Report IPv4 and IPv6 aggregated forwarded counters using standard RADIUS attributes (disabled by default):

[42] Acct-Input-Octets

[43] Acct-Output-Octets

[47] Acct-Input-Packets

[48] Acct-Output-Packets

[52] Acct-Input-Gigawords

[53] Acct-Output-Gigawords

v6-aggregate-stats

Report IPv6 aggregated forwarded counters of queues and policers in stat-mode v4-v6 using RADIUS VSAs (disabled by default):

[26-6527-194] Alc-IPv6-Acct-Input-Packets

[26-6527-195] Alc-IPv6-Acct-Input-Octets

[26-6527-196] Alc-IPv6-Acct-Input-GigaWords

[26-6527-197] Alc-IPv6-Acct-Output-Packets

[26-6527-198] Alc-IPv6-Acct-Output-Octets

[26-6527-199] Alc-IPv6-Acct-Output-Gigawords

In addition to accounting-start, interim-accounting, and accounting-stop messages, a RADIUS client on a routers also sends accounting-on and accounting-off messages. An accounting-on message is sent when a specific RADIUS accounting policy is applied to a specified subscriber profile, or the first server is defined in the context of an already applied policy. The following attributes included are in these messages:

- NAS-identifier
- alc-subscriber-profile-string
- Accounting-session-id
- Event-timestamp

Accounting-off messages are sent at following events:

- An accounting policy has been removed from a sub-profile.
- The last RADIUS accounting server has been removed from an already applied accounting policy.

These messages contain following attributes:

- NAS-identifier
- alc-subscriber-profile-string
- Accounting-session-id
- Accounting-terminate-cause

- Event-timestamp

In case of dual homing, both nodes send RADIUS accounting messages for the host, with all attributes as it is locally configured. The RADIUS log files on both boxes need to be parsed to get aggregate accounting data for the specified subscriber host regardless the node used for forwarding.

For RADIUS-based accounting, a custom record can be defined to refine the data that is sent to the RADIUS server. See the "Configuring an Accounting Custom Record" in the *7450 ESS, 7750 SR, 7950 XRS, and VSR System Management Guide* for further information.

9.2.9.5 RADIUS accounting terminating cause

The VSA **acct-terminate-cause** attribute provides some termination information. Two additional attributes: [VSA 227] **alc-error-message** and [VSA 226] **alc-error-code** provide more information in both string and numeric format about the terminating cause of the subscriber session. The full list of error messages and their corresponding error codes may be viewed using the command **tools**>**dump**>**aaa**> **radius-acct-terminate-cause**.

If required, python can alter the content of both VSAs. The following is a python script example where the error codes are remapped from 123 to 8 and from 124 to 17:

```
import alc
import struct

ALU            = 6527
TERM_CAUSE     = 49
ALC_ERROR_CODE = 226

if (alc.radius.attributes.isSet(TERM_CAUSE) and
    alc.radius.attributes.isVSASet(ALU, ALC_ERROR_CODE)):

    error_code = alc.radius.attributes.getVSA(ALU, ALC_ERROR_CODE)
    error_code = struct.unpack('!i', error_code)[0]

    term_cause  = alc.radius.attributes.get(TERM_CAUSE)
    term_cause = struct.unpack('!i', term_cause)[0]

    #print "error code = ", error_code
    #print "term cause = ", term_cause

    # table with mapping from alc-error-code to the standard terminate cause
    # if no mapping is found, no transformation is performed
    error_map = {
        123 : 8,
        124 : 17
    }

    new_term_cause = error_map.get(error_code, term_cause)
    #print "new term_cause = ", new_term_cause

    alc.radius.attributes.set(TERM_CAUSE, struct.pack('!I', new_term_cause))
```

9.2.9.6 Accounting modes of operation

This section is applicable to the 7750 SR or the 7450 ESS. There are three basic accounting models:

- Per queue-instance
- Per Host
- Per Session

Each of the basic models can optionally be enabled to send interim-updates. Inclusion/exclusion of interim-updates depends on whether volume based (start/interim-updates/stop) or time-based (start/stop) accounting is required.

The difference between the three basic accounting models is in its core related to the processing of the acc-session-id for each model. The differences are related to:

- acct-session-id generation within each model
- outcome in response to the CoA action relative to the targeted acct-session-id

The counters for volume-based accounting are collected from queues or policers that are instantiated per SLA profile instance (SPI). This is true regardless of which model of accounting (or combination of models) is deployed. Within the accounting context, the SPI equates to queue-instance.

The following table summarizes the key differences between various accounting modes of operation that are supported. Interim-updates for each individual mode can be enabled or disabled through configuration (**keyword** as an extension to the commands that enable three basic modes of accounting). This is denoted by the IU-Config keyword under the 'I-U' column in the table. The table also shows that any two combinations of the three basic models (including their variants for volume and time- based accounting) can be enabled simultaneously.

Table 10: Accounting modes of operation

Accounting mode	Accounting entity	START	I-U	STOP	Acct-session-id	Acct-multi-session-id
queue-instance-accounting	queue-instance	X	IU-config	X	X	
	session					
	host					
session-accounting	queue-instance					
	session	X	IU-config	X	X	queue-instance
	host					
host-accounting	queue-instance					
	session					

Accounting mode	Accounting entity	START	I-U	STOP	Acct-session-id	Acct-multi-session-id
	host	X	IU-config	X	X	queue-instance
queue-instance-accounting + host-accounting	queue-instance	X	IU-config	X	X	queue-instance
	session					
	host	X	IU-config	X	X	queue-instance
queue-instance-accounting + session-accounting	queue-instance	X	IU-config	X	X	queue-instance
	session	X	IU-config	X	X	queue-instance
	host					
session-accounting + host-accounting	queue-instance					
	session	X	IU-config	X	X	queue-instance
	host	X	IU-config	X	X	session

Note: Hosts within the targeted CoA entity are affected as follows:

- If the CoA target is the session, both constituting members (IPv4 and IPv6) of the dual-stack host are affected.
- If the CoA target is the queuing-instance, up to 32 hosts that are sharing that SPI are affected.

The same principle applies to LI.

The accounting behavior (accounting messages and accounting attributes) in case that the SPI is changed with CoA depends on the accounting mode of operation. The behavior is the following:

- SPI change in conjunction with per queuing instance accounting triggers a STOP for the old SPI and a START for the new SPI with corresponding counters. Acct-session-id/Acct-Multi-Session-Id is unique per SPI. Note that Acct-Multi-Session-Id is only generated if per queuing-instance accounting mode of operation is combined with some other mode of operation (host or session).
- SPI change in conjunction with per host or per session accounting (no interim updates for either method) does not trigger any new accounting messages. In other words, SPI change goes unnoticed from the perspective of the accounting server until the host/session is terminated. When the host/session is terminated a STOP is sent with the VSA carrying the latest SPI name and the acct-multi-session-id attribute of the latest SPI. Acct-session-id stays the same during the lifetime of the host. Counters are not included in STOP (interim-update not enabled).

  SPI change in conjunction with per host accounting with interim-updates or per session accounting with interim-updates triggers two interim-update messages:

- One with the old counters (terminated queues) and the old SPI name VSA. This behavior is similar to the triggered STOP message in per queuing-instance accounting upon SPI change.
- One with the new counters (new queues instantiated), the VSA carrying the new SPI name and the new acct-multi-session-id referencing the new SPI. This behavior is similar to the triggered START message in per queuing-instance accounting when SPI is changed.

9.2.9.7 Per queue-instance accounting

In the per queue-instance accounting mode of operation, the accounting message stream (START/INTERIM-UPDATE/STOP) is generated per queue-instance (per SLA profile instance for non-HSQ cards).

An accounting message stream refers to a collection of accounting messages (START/INTERIM-UPDATE/STOP) sharing the same acct-session-id.

The following are the properties of the per queue-instance accounting model:

- A RADIUS accounting start message is sent when the queue instance (SLA profile instance) is created. It contains the IP address attribute of the host that caused the queue instance (SLA profile instance) to be created.
- Additional hosts may bind to the queue instance (SLA profile instance) at any time, but no additional accounting messages are sent during these events.
- If the original host disconnects then future accounting messages use an IP address of one of the remaining hosts.
- When the final host associated with a queue instance (SLA profile instance) disconnects an Accounting Stop message is sent.

9.2.9.8 Per host accounting

In the per host accounting mode of operation the accounting message stream (START/INTERIM-UPDATE/STOP) is generated per host.

An accounting message stream refers to a collection of accounting messages (START/INTERIM-UPDATE/STOP) sharing the same acct-session-id.

The following are the properties of the per host accounting model:

- A RADIUS accounting START message is sent each time a host is created in the system.
- Whenever a host disconnects, a RADIUS accounting STOP message is sent for that host.
- The accounting messages (START, INTERIM-UPDATE, STOP) carry the acct-multi-session-id attribute denoting the queue instance or session with which the host is associated (see Table 10: Accounting modes of operation).
- The counters are collected from the queues and policers instantiated through the queue instance (SLA profile instance). If multiple hosts share the same queue instance, the counters are aggregated. In other words, counters per individual hosts cannot be extracted from the aggregated count.

9.2.9.9 Per session accounting

In the per session accounting mode of operation, an accounting message stream (START/INTERIM-UPDATE/STOP) is generated per session. An accounting message stream refers to a collection of accounting messages (START/INTERIM-UPDATE/STOP) sharing the same acct-session-id.

- A PPPoE session is identified by the key {session-ID, mac}
- An IPoE session is identified by the configured session-key: {sap, mac} | {sap, mac, Circuit-ID} | {sap, mac, Remote-ID}

For a single stack session, the behavior defined in the per session accounting model is indistinguishable from the per host accounting model. The per session accounting model makes difference in behavior only for dual-stack sessions.

The following are the properties of the Per Session Accounting model:

- A single accounting session ID (acct-session-id) is generated per (IPoE or PPPoE) session and it can optionally be sent in RADIUS Access-Request message.
- This acct-session-id is synchronized through MCS in dual homing environment.
- The accounting messages (START, INTERIM-UPDATE, STOP) carry the acct-multi-session-id attribute denoting the queue instance (SLA profile instance) with which the session is associated.
- The counters are collected from the queues and policers instantiated through the queue instance (SLA Profile Instance). If multiple sessions are sharing the same queue instance, the counters are aggregated. In other words, counters per individual session cannot be extracted from the aggregated count.
- RADIUS-triggered changes and LI, targeted to the session's accounting session ID are applicable per session:
  - In queue and policer RADIUS overrides, parameters for the referenced queue and policer within the session are changed accordingly.
  - Subscriber aggregate rate limits, scheduler rates, and arbiter rates are changed accordingly.
  - CoA DISCONNECT brings down the entire session.
  - LI activation based on the session acct-session-id affects the hosts within the session (dual-stack).
  - An SLA profile instance change affects all hosts (or sessions) sharing the same sla-profile instance (SPI). Queues are re-instantiated and counters are reset.
- All applicable IP addresses (IPv4 and IPv6, including all IPv6 attributes; alc-ipv6-address, framed-ipv6-prefix, delegated-ipv6-prefix) are present in accounting messages for the session.

9.2.9.10 RADIUS session accounting with PD as a managed route

The Prefix Delegation (PD) prefix is included in the accounting messages using the VSA [99], Framed-IPv6-Route attribute with the string type "pd-host" appended to differentiate it from a regular framed IPv6 route; for example, FRAMED IPV6 ROUTE [99] 39 2001:1000::/64 :: 0 pref 0 type pd-host. PD as a managed route is applicable to both PPP and IPoE sessions and can point either to an IPv4 host or to an IPv6 WAN host.

Table 11: RADIUS accounting behavior describes the RADIUS accounting behavior based on the session type and the next-hop host.

Table 11: RADIUS accounting behavior

Session type and next-hop host	RADIUS accounting start	RADIUS accounting interims	RADIUS accounting stop
PPP session with IPv6 PD pointing to IPv4 host as the next hop	A PPP connection triggers an accounting start	A DHCP NA+PD solicit triggers an interim update for the PD host with interim reason "delegated-ipv6-prefix-up" and the	A PPP disconnect with only the IPv4 and IPv6 PD host triggers an accounting stop with the

Session type and next-hop host	RADIUS accounting start	RADIUS accounting interims	RADIUS accounting stop
		prefix included in the VSA Framed-IPv6-Route A DHCP PD solicit triggers an interim update for the PD host with interim reason delegated-ipv6-prefix-up and the prefix included in the VSA framed-ipv6-route Restriction: A DHCP PD lease expire triggers an interim update with interim reason "delegated-ipv6-prefix-down"; however, the VSA framed-ipv6-route is not included	prefix included in the VSA Framed-IPv6-Route Restriction: A PPP disconnect with the IPv4, NA, and PD host without **session-optimized-stop** enabled, is not include the VSA Framed-IPv6-Route
PPP session with IPv6 PD pointing to IPv6 NA host as the next hop	A PPP connection triggers an accounting start. It is possible to have a single-stack IPv6-only session	A DHCP NA+PD solicit triggers an interim update for the PD host with interim reason delegated-ipv6-prefix-up and the prefix included in the VSA framed-ipv6-route Restriction: A DHCP PD lease expire triggers an interim update with interim reason "delegated-ipv6-prefix-down"; however, the VSA FramedIPv6-Route is not included	A PPP subscriber disconnect triggers an accounting stop with the PD host prefix included in the VSA Framed-IPv6-Route
IPoE session with IPv6 PD pointing to IPv4 host as the next hop	A DHCPv4 or a DHCPv6 request (DHCPv6 always performs NA and PD requests together) triggers the accounting start	A DHCP PD is always performed together with NA. The PD is not in the start message but is included in the accounting interim update as a part of the host update. If the DHCPv4 lease expires, the interim update contains the PD prefix in the VSA framed-ipv6-route	If only the IPv4 host and PD host remain, the release of the DHCPv4 triggers an accounting stop with the PD host prefix included in the VSA Framed-IPv6-Route Restriction: If the DHCPv4 is released and an IPv6 NA host remains, the IPv6 lease release/expire is an

Session type and next-hop host	RADIUS accounting start	RADIUS accounting interims	RADIUS accounting stop
		Restriction: A DHCP PD lease expire triggers an interim update with interim reason "delegated-ipv6-prefix-down"; however, the VSA Framed-IPv6-Route is not included	interim update that does not include the prefix
IPoE session with IPv6 PD pointing to IPv6 NA host as the next hop	A DHCPv4 or a DHCPv6 request (DHCPv6 always performs NA and PD requests together) triggers the accounting start. It is possible to have a single-stack IPv6-only session	A DHCP PD is always performed together with NA. The PD is not in the start message but is included in the accounting interim update as a part of the host update. Restriction: A DHCP PD lease expire triggers an interim update with interim reason "delegated-ipv6-prefix-down"; however, the VSA Framed-IPv6-Route is not included	If only the IPv6 subscriber is left, the release of NA contains the prefix of the PD host Restriction: If the DHCPv6 is released and an IPv4 host remains, the IPv6 lease release/expire is an interim update that does not include the prefix

RADIUS per host accounting:

In SR OS, the accounting paradigm is based on SLA profile instances yet this is at odds with traditional RADIUS authentication and accounting which is host-centric. In previous SR OS releases, it was possible to have many hosts sharing a common SLA profile instance, and therefore accounting and QoS parameters. Complications arose with RADIUS accounting because Accounting-Start and Accounting-Stop are a function of **sla-profile** instance and not the hosts. This meant that some host-specific parameters (like framed-ip-address) would not be consistently included in RADIUS accounting.

Currently, dual-stack subscribers are really two different hosts sharing a single sla-profile instance. A new RADIUS accounting mode has been introduced to support multiple-host environments.

Under accounting-policy, a host-accounting command allows configurable behavior.

9.2.9.11 Reduction of host updates for session accounting start and stop

When **host-update** is enabled in session accounting, a dual-stack subscriber can generate multiple host update accounting messages at the start and end of a session (for example, one for the IPv4 host and two more for the IPv6 WAN and IPv6 PD hosts). Two features can be used to reduce the number of host update messages per subscriber.

The first feature delays the Start Accounting message by a configurable value and is applicable to both PPPoE and IPoE sessions. The command for configuring this feature is **config>subscr-mgmt>acct-plcy>delay-start-time**. The delay allows the full dual-stack address assignment to be completed before

triggering the accounting Start message. The Start message reports all the addresses and prefixes assigned to the subscriber at that time. Subsequent new or disconnected hosts triggers interim host updates if enabled.

The second feature is for PPPoE sessions only and is used to reduce the number of host update messages when a dual-stack PPP subscriber disconnects. The command for configuring this feature is **config**>**subscr-mgmt**>**sub-prof**>**rad-acct**>**session-optimized-stop**. A single accounting Stop message containing all the addresses and prefixes for the subscriber at the time is generated.

9.2.9.12 Accounting interim update message interval

The interval between two RADIUS Accounting Interim Update messages can be configured in the RADIUS accounting policy with the **update-interval** command, for example:

```
config
 subscr-mgmt
        radius-accounting-policy "acct-policy-1" create
            update-interval 60
            update-interval-jitter absolute 600
```

A RADIUS specified interim interval (attribute [85] Acct-Interim-Interval) overrides the CLI configured value.

By default, a random delay of 10% of the configured **update-interval** is added to the update-interval between two Accounting Interim Update messages. This jitter value can be configured with the **update-interval-jitter** to an absolute value in seconds between zero and 3600. The effective maximum random delay value is the minimum value of the configured absolute jitter value and 10% of the configured **update-interval**.

A value of zero sends the Accounting Interim Update message without introducing an additional random delay.

9.2.9.13 CoA triggered accounting interim update

The vendor-specific attribute (VSA) [228], Alc-Triggered-Acct-Interim, can be used in a Change of Authorization message to trigger an interim accounting message. This feature requires the accounting mode to have interim updates enabled. You can enable interim updates using, the **config**>**subscr-mgmt**>**radius-acct-plcy**>**host-accounting interim-update** command. The VSA can hold a string of up to 247 characters. The accounting interim echoes this string in the interim message under the same Alc-Triggered-Acct-Interim VSA along with Alc-Acct-Triggered-Reason = CoA-triggered. If the VSA is left blank, it still triggers the accounting interim message with Alc-Acct-Triggered-Reason = CoA-triggered (18), but without the Alc-Triggered-Acct-Interim attribute. If the subscriber session has multiple accounting policies or modes enabled, multiple interim messages are generated. Some CoAs, such as SLA profile or sub-profile changes, triggers accounting update messages to be generated automatically. These CoAs can automatically generate one or more accounting interim messages. If these CoAs also include the Alc-Triggered-Acct-Interim VSA, no additional interim accounting messages are generated. The last automatically-generated accounting interim message contain these reasons:

- the reason for the triggered interim message (such as an SLA start)
- the CoA-triggered (18) Alc-Triggered-Acct-Interim attribute that is echoed on the triggered accounting interim message if the VSA is not empty

9.2.9.14 Class attribute

The RADIUS class attribute helps to aid in user identification.

User identification is used to correlate RADIUS accounting messages with the specified user. During the authentication process, the RADIUS authentication server inserts a class attribute into the RADIUS authenticate response message and the router echoes this class attribute in all RADIUS accounting messages.

The 7750 SR can store up to six class attributes for both RADIUS and NASREQ. Each class VSA or AVP can have a maximum of 253 characters. If the VSA or AVP contains more than 253 characters, only the first 253 characters is stored. If there are more than six VSAs or AVPs, only the first six is stored. This functionality is also applicable to RADIUS authentication by the ISA.

9.2.9.15 Username

The username, which is used for user authentication (the "user-name" attribute in RADIUS authentication request), can be included in RADIUS accounting messages. Per RFC 2865, when a RADIUS server returns a (different) "user-name" attribute, the changed name is used in accounting and not the originally sent name.

9.2.9.16 Accounting-On and Accounting-Off

For RADIUS servers configured in a RADIUS server policy, the accounting on and off behavior is controlled with the **acct-on-off** command in the **radius-server-policy**.

By default, no Accounting-On or Accounting-Off messages are sent (**no acct-on-off**).

With the **acct-on-off** command configured in the **radius-server-policy**:

- An Accounting-On is sent for the following:
  - When the system is powered on
  - After a system reboots
  - When the **acct-on-off** command is added to the **radius-server-policy** configuration
  - User triggered with CLI: tools perform aaa acct-on
- An Accounting-Off is sent for the following:
  - Before a user initiated system reboot
  - When the **acct-on-off** command is removed from the **radius-server-policy** configuration
  - User triggered with CLI: tools perform aaa acct-off

The Accounting-On or Accounting-Off message is sent to the servers configured in the **radius-server-policy**, following the configured access-algorithm until an Accounting Response is received. If the first server responds, no message is sent to the other servers.

The Accounting-On message is repeated until an Accounting Response message is received from a RADIUS server: If after the configured retry or timeout timers for each RADIUS server in the RADIUS server no response is received then the process starts again after a fixed one minute wait interval.

The Accounting-Off message is attempted once: If after the configured retry or timeout timers for each RADIUS server in the RADIUS server policy no response is received then no new attempt is made.

It is possible to block a RADIUS server policy until an Accounting Response is received from one of the RADIUS servers in the RADIUS server policy that acknowledges the reception of an Accounting-On. The RADIUS server policy cannot be used by applications for sending RADIUS messages until the state becomes "Not Blocked". This is achieved with the optional "oper-state-change" flag, for example:

```
config
    aaa
        radius-server-policy "aaa-server-policy-1" create
            acct-on-off oper-state-change
            servers
                router "Base"
                server 1 name "server-1"
            exit
        exit
    exit
```

If multiple RADIUS server policies are in use for different applications (for example, authentication and accounting) and an Accounting-On must be send for only one RADIUS server policy, it is possible to tie the acct-on-off states of both policies together using an acct-on-off-group. With this configuration, it is possible to block the authentication servers until the accounting servers are available. An acct-on-off-group can be referenced by:

- a single RADIUS server policy as controller: the acct-on-off oper-state of the acct-on-off-group is set to the acct-on-off oper-state of the radius-server-policy
- multiple RADIUS server policies as monitor: the acct-on-off oper-state of the RADIUS server policy is inherited from the acct-on-off oper-state of the acct-on-off group.

```
config
    aaa
        acct-on-off-group "group-1" create
            description "Grouping of radius-server-policies acct-on-off"
        exit
        radius-server-policy "aaa-server-policy-1" create
            acct-on-off oper-state-change group "group-1"
            servers
                router "Base"
                server 1 name "server-1"
            exit
        exit
        radius-server-policy "aaa-server-policy-2" create
            acct-on-off monitor-group "group-1"
            servers
                router "Base"
                server 1 name "server-2"
            exit
        exit
```

It is possible to force an Accounting-On or Accounting-Off message for a RADIUS server policy with acct-on-off enabled using following CLI commands:

tools perform aaa acct-on [**radius-server-policy** *policy-name*] [**force**]

tools perform aaa acct-off [**radius-server-policy** *policy-name*] [**force**] [**acct-terminate-cause** *number*]

If an Accounting-On was sent to the **radius-server-policy** and it was acknowledged with an Accounting Response then a new Accounting-On can only be sent with the "force" flag.

If an Accounting-Off was sent to the **radius-server-policy** and it was acknowledged with an Accounting Response then a new Accounting-Off can only be sent with the "force" flag. The Acct-Terminate-Cause value in the Accounting-Off can be overwritten.

Use the following CLI command to display the Accounting On/Off information for a **radius-server-policy:**

```
# show aaa radius-server-policy "aaa-server-policy-3" acct-on-off
===============================================================================
RADIUS server policy "aaa-server-policy-3" AcctOnOff info
===============================================================================
Oper state                  : on
Session Id                  : 242FFF0000008F512A3985
Last state change           : 02/24/2013 16:06:41
Trigger                     : startUp
Server                      : "server-1"
===============================================================================
```

The operational state provides following state information: The sending of the Accounting-On or Accounting-Off message is ongoing (sendAcctOn, SendAcctOff), is successfully responded (on, off) or no response received (OffNoResp).

The Session-Id is a unique identifier for each RADIUS server policy accounting Accounting-On/Accounting-Off sequence.

The Trigger field shows what triggered the Accounting On or Accounting Off message. If the **radius-server-policy** is part of an acct-on-off group then the group name is shown in brackets.

The server field shows which server in the RADIUS server policy responded to the Accounting-On or Accounting-Off message.

To display the acct-on-off state of a **radius-server-policy**, use the command, for example:

```
# show aaa radius-server-policy "aaa-server-policy-3"
===============================================================================
RADIUS server policy "aaa-server-policy-3"
===============================================================================
Description                 : (Not Specified)
Acct Request script policy  : script-policy-1
Auth Request script policy  : script-policy-1
Accept script policy        : script-policy-1
Acct-On-Off                 : Enabled (state Blocked)
-------------------------------------------------------------------------------
RADIUS server settings
-------------------------------------------------------------------------------
Router                      : "Base"
Source address              : (Not Specified)
Access algorithm            : direct
Retry                       : 3
Timeout (s)                 : 5
Hold down time (s)          : 30
Last management change      : 02/20/2013 13:32:05
===============================================================================
===============================================================================
Servers for "aaa-server-policy-3"
===============================================================================
Idx Name                            Address         Port        Oper State
                                                    Auth/Acct
-------------------------------------------------------------------------------
1   server-3                        172.16.1.10     1812/1813   unknown
===============================================================================
```

The Acct-On-Off field indicates if the sending of Accounting-On and Accounting-Off messages is enabled or disabled. If enabled, the oper-state is displayed: state Blocked or state Not Blocked. When Blocked, the **radius-server-policy** cannot be used to send RADIUS messages.

To display acct-on-off-group information, use following command, for example:

```
# show aaa acct-on-off-group "group-1"
===============================================================================
Acct-On-Off-Group Information
===============================================================================
acct on off group name               : group-1
  - controlling Radius-Server-policy :
        aaa-server-policy-1
  - monitored by Radius-Serer-policy :
        aaa-server-policy-2

-------------------------------------------------------------------------------
Nbr of Acct-on-off-groups displayed : 1
-------------------------------------------------------------------------------
===============================================================================
```

9.2.9.17 RADIUS accounting message buffering

When all servers in a RADIUS server policy are unreachable, it is possible to buffer the Accounting Start, Accounting Stop, and Accounting Interim-Update messages for up to 25 hours. Accounting Start messages have a separate buffer from Accounting Interim-Update and Stop messages. When a RADIUS server becomes reachable again, the messages in the buffer are retransmitted. If, for the same accounting session, an Accounting Start message and an Accounting Interim-Update or Stop message is buffered, then the Accounting Start message is sent before the Interim-Update or Stop message.

RADIUS Accounting message buffering parameters can be configured per message type, for example:

```
config
    aaa
        radius-server-policy "aaa-server-policy-1" create
            servers
                router "Base"
                buffering
                    acct-start min 60 max 3600 lifetime 12
                    acct-interim min 60 max 3600 lifetime 12
                    acct-stop min 60 max 3600 lifetime 12
                exit
                server 1 name "server-1"
            exit
        exit
    exit
```

When RADIUS accounting message buffering is enabled:

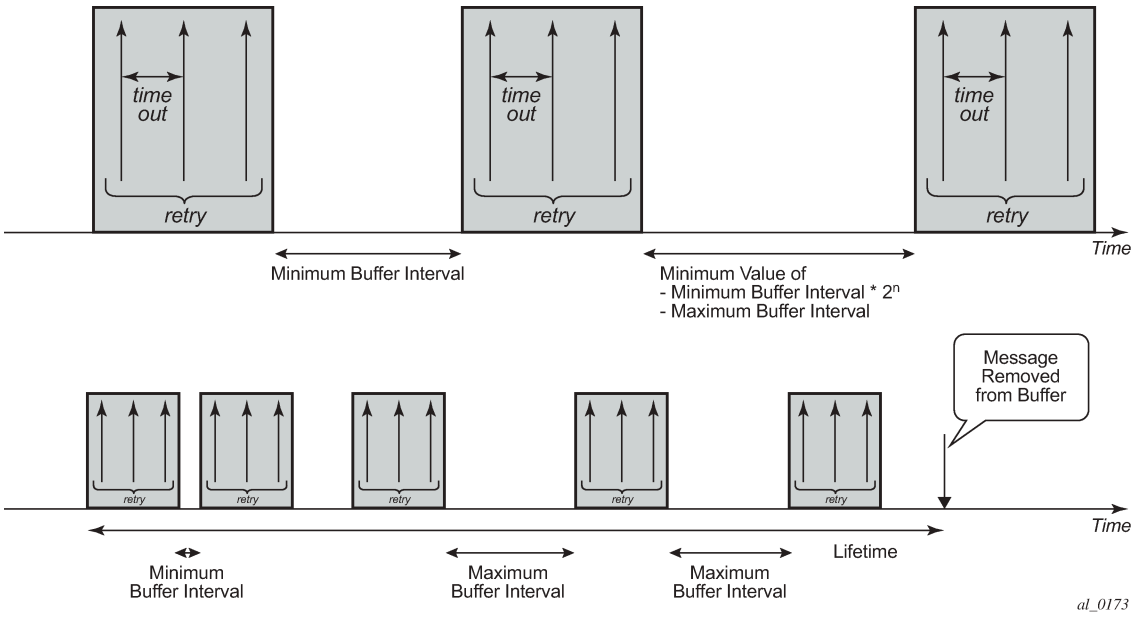
1. The message is stored in the buffer, a lifetime timer is started and the message is sent to the RADIUS server.
2. If, after **retry** *timeout* seconds, no RADIUS accounting response is received, then a new attempt to send the message is started after a minimum [(*min-val**2n), *max-val*] seconds. The *min-val* and *max-val* parameters are configurable and correspond to each accounting message type.
3. Repeat step 2 until one of the following events occurs and the message is purged from the buffer:
   a. RADIUS accounting response is received

b. the lifetime of the buffered message expires (as shown in Figure 68: Purging message from buffer)

c. (if the buffered message is an Accounting Interim-Update only) A new Accounting Interim-Update or an Accounting Stop or for the same accounting **session-id** and **radius-server-policy** is stored in the buffer

d. the message is manually purged from the message buffer with a clear command

Figure 68: Purging message from buffer



When Accounting Start message buffering is enabled:

- the Accounting Start message is stored in the buffer
- enabling Accounting Interim-Update and Stop message buffering with the same *lifetime* value is recommended. This guarantees the message ordering per accounting session. The RADIUS Accounting Start message is used to re-establish a connection to the RADIUS server. Therefore, when the connections to RADIUS servers are restored, Accounting Start messages are always sent first followed by the Accounting Interim-Update or Stop messages. In addition, when connection to the RADIUS server is restored, the system attempts to send the buffered Accounting Start messages first, as Accounting responses are received for the Accounting Start messages, Accounting Interim-Update or Stop messages for that particular subscriber session are sent.
- if, for the same accounting session, an Accounting Start message and an Accounting Interim-Update or Stop message are both buffered, it is possible for the Start message to be dropped from the accounting buffer because of *lifetime* expiry. As a result, when the connection to the RADIUS server is restored, only the Accounting Interim-Update or Stop message is sent.
- if the RADIUS server is unreachable for a prolonged period, it is possible for subscribers to have started and terminated more than one session. If buffering for Accounting Start is enabled, an Accounting Start message for each session is buffered.
- a Python script is applied only when the RADIUS Start message is sent. In other words, buffered RADIUS messages are never processed by Python. It is possible to alter the Python scripts when RADIUS messages are buffered and the message is subjected to the newest applied Python script.

When Accounting Interim-Update message buffering is enabled:

- only the last Accounting Interim-Update or Accounting Stop message (if enabled) is stored in the buffer. Accounting session events that are reported by a triggered Accounting Interim-Update, such as an SLA-Profile Change can be lost.
- enabling Accounting Stop message buffering is recommended. This guarantees message ordering per accounting session.

Use the following clear command to manually delete messages from the RADIUS accounting message buffer:

clear aaa radius-server-policy *policy-name* **msg**-buffer [**acct-session-id** *acct-session-id*]

When specifying the account session ID, only that specific message is deleted from the message buffer. If no account session ID is specified, all messages for that RADIUS server policy are deleted from the message buffer.

Use the following **show** commands to display the RADIUS accounting message buffer statistics:

```
# show aaa radius-server-policy "aaa-server-policy-1" msg-buffer-stats
===============================================================================
buffering acct-start        : enabled
  min interval (s)          : 60
  max interval (s)          : 300
  lifetime (hrs)            : 25
buffering acct-interim      : enabled
  min interval (s)          : 60
  max interval (s)          : 300
  lifetime (hrs)            : 25
buffering acct-stop         : enabled
  min interval (s)          : 60
  max interval (s)          : 300
  lifetime (hrs)            : 25
Statistics
-------------------------------------------------------------------------------
Total acct-start messages in buffer                     : 0
Total acct-interim messages in buffer                   : 0
Total acct-stop messages in buffer                      : 0
Total acct-start messages dropped (lifetime expired)    : 0
Total acct-interim messages dropped (lifetime expired)  : 0
Total acct-stop messages dropped (lifetime expired)     : 0
Last buffer clear time                                  : N/A
Last buffer statistics clear time                       : N/A
-------------------------------------------------------------------------------
===============================================================================
```

Use the following **clear** command to reset the RADIUS accounting message buffer statistics:

clear aaa radius-server-policy *policy-name* **statistics msg-buffer-only**

Use the following **tools** commands to display the RADIUS accounting message buffer content:

tools dump aaa radius-server-policy *policy-name* **msg-buffer** [**session-id** *acct-session-id*]

For example:

```
# tools dump aaa radius-server-policy "aaa-server-policy-1" msg-buffer
===============================================================================
RADIUS server policy "aaa-server-policy-1" message buffering
===============================================================================
message type Acct-Session-Id                            remaining lifetime
-------------------------------------------------------------------------------
acct-interim 242FFF0000009A512B36FC                     0d 11:58:54
```

```
acct-interim 242FFF0000009B512B36FC                          0d 11:58:48
acct-interim 242FFF0000009C512B36FC                          0d 11:58:30
acct-interim 242FFF0000009D512B36FC                          0d 11:58:29
acct-interim 242FFF0000009E512B36FC                          0d 11:59:05
-------------------------------------------------------------------------------
No. of messages in buffer: 5
===============================================================================
```

When specifying the Acct-Session-Id, the message details are displayed.

9.2.9.18 Multiple accounting policies

The subscriber profile allows the user to configure a primary accounting policy with an additional accounting policy. The accounting policies are independent of each other and each policy has its own accounting mode, update interval, and include attributes. The RADIUS VSA [85] Acct-Interim-Interval attribute changes both the primary and the duplicate accounting interim update interval.

9.2.9.19 Sending an accounting stop message upon a RADIUS authentication failure of a PPPoE session

In scenarios where PAP/CHAP RADIUS authentication is used for PPPoE sessions, an accounting stop message can be generated to notify the RADIUS servers in case of an authentication failure. This feature is not supported for PADI authentication.

The failure events are categorized in three categories:

- "**on-request-failure**"

  All failure conditions between the sending of an Access-Request and the reception of an Access-Accept or Access-Reject.

- "**on-reject**"

  When an Access-Reject is received.

- "**on-accept-failure**"

  All failure conditions that appear after receiving an Access-Accept and before successful instantiation of the host or session.

Each of the categories can be enabled separately in the RADIUS authentication policy.

In the Enhanced Subscriber Management (ESM) model, the RADIUS accounting server is found after authentication and host identification as part of the subscriber profile configuration. To report authentication failures to accounting servers, an alternative RADIUS accounting policy configuration is required: local user database pre-authentication can provide the RADIUS authentication policy to be used for authentication and the RADIUS accounting policy to be used for authentication failure reporting. A duplicate RADIUS accounting policy can be specified if the accounting stop resulting from a RADIUS authentication failure must also be sent to a second RADIUS destination.

```
configure
    subscriber-mgmt
        local-user-db "ludb-1" create
            ppp
                match-list username
                host "default" create
                    auth-policy "auth-policy-1"
```

```
                    acct-policy "acct-policy-1" duplicate "acct-policy-2"
                    no shutdown
                exit
            exit
            no shutdown
        exit
        authentication-policy "auth-policy-1" create
            pppoe-access-method pap-chap
            include-radius-attribute
               - - - snip - - -
            exit
            send-acct-stop-on-fail on-request-failure on-reject on-accept-failure
            radius-server-policy "aaa-server-policy-1"
        exit
        radius-accounting-policy "acct-policy-1" create
            - - - snip - - -
            radius-server-policy "aaa-server-policy-1"
        exit
        radius-accounting-policy "acct-policy-2" create
            - - - snip - - -
            radius-server-policy "aaa-server-policy-2"
        exit
```

To enable local user database pre-authentication, use the **user-db** configuration in the capture SAP and in the group interface. For example:

```
configure
    service
        vpls 10 customer 1 create
            sap 1/1/1:1.* capture-sap create
                trigger-packet pppoe
                pppoe-policy "ppp-policy-1"
                pppoe-user-db "ludb-1"
            exit
            no shutdown
        exit
        ies 1000 customer 1 create
            subscriber-interface "sub-int-1" create
               - - - snip - - -
                group-interface "group-int-1-1" create
                    - - - snip - - -
                    pppoe
                        policy "ppp-policy-1"
                        user-db "ludb-1"
                        no shutdown
                    exit
                exit
            exit
            no shutdown
        exit
```

9.2.9.20 Sending an accounting stop message upon an IPoE host creation failure

If IPoE host creation fails, the system can generate an accounting stop message. This feature is similar to the feature described in Sending an accounting stop message upon a RADIUS authentication failure of a PPPoE session. It allows the system to generate an accounting stop message for most host creation failure cases. For IPoE, only the failure event "on-accept-failure" is supported. This failure condition applies

when the host was successfully authenticated but the host creation failed (for example, a duplicate host IP address was detected on the new host).

Because RADIUS accounting starts only after the host is successfully created, a failed host cannot trigger a RADIUS accounting message. For this reason, similar to PPPoE, the local user database must be used to provide the RADIUS accounting server for reporting the failure.

The [26.6527.226] Alc-Error-Code and [26.6527.227] Alc-Error-Message attributes are used to report the failure in the RADIUS accounting stop message. The error code is a numeric value that represents the error, and the error message is a descriptive text **string** that describes the actual failure reason. For IPoE, the error code uses the 279 value (in decimal format) or 0x117 value (in hexadecimal format) "Failed to create subscriber host". The error message provides the same detailed reason for the host creation failure as the log message in log 99.

9.3 Enhanced subscriber management overview

9.3.1 Enhanced subscriber management basics

In residential broadband networks numerous subscribers can be provisioned that can require significant changes on a daily basis. Manually configuring the applicable parameters for each subscriber would be prohibitive. The Nokia 7450 ESS and 7750 SR have been designed to support fully dynamic provisioning of access, QoS and security aspects for residential subscribers using DHCP to obtain an IP address. Enabling Enhanced Subscriber Management (ESM) drastically reduces the configuration burden.

ESM in the 7450 ESS and 7750 SR supports many vendor's access nodes and network aggregation models, including VLAN per customer, per service or per access node.

9.3.1.1 Standard and Enhanced Subscriber Management

The system can switch between standard and enhanced subscriber management modes on a per SAP basis. The ESM mode is supported on the SR-7 and SR-12 chassis and on the ESS-7 chassis.

Some functions are common between the standard and enhanced modes. These include DHCP lease management, static subscriber host definitions and anti-spoofing. While the functions of these features may be similar between the two modes, the behavior is considerably different.

- **Standard mode**

  The system performs SLA enforcement functions on a per SAP basis, that is, the attachment to a SAP with DHCP lease management capabilities. The node can authenticate a subscriber session with RADIUS based on the MAC address, the circuit-id (from Option 82) or both. It then maintains the lease state in a persistent manner. It can install anti-spoofing filters and ARP entries based on the DHCP lease state. Static subscriber hosts are not required to have any SLA or subscriber profile associations and are not required to have a subscriber identification string defined.

- **Enhanced mode**

  When enabled on a SAP, the system expands the information it stores per subscriber host, allowing SLA enforcement and accounting features on a per subscriber basis. The operator can create a subscriber identification policy that includes a URL to a user-space script that assists with the subscriber host identification process.

- A subscriber host is identified by a subscriber identification string instead of the limited Option 82 values (although, the identification string is normally derived from string manipulation of the Option 82 fields). A subscriber identification policy is used to process the dynamic host DHCP events to manage the lease state information stored per subscriber host. The static subscriber hosts also must have subscriber identification strings associations to allow static and dynamic hosts to be grouped into subscriber contexts.
- Further processing by the subscriber identification policy derives the appropriate subscriber and SLA profiles used to define the hierarchical virtual schedulers for each subscriber and the unique queuing and filtering required for the hosts associated with each subscriber.
- The SLA profile information is used to identify which QoS policies and which queues/policers, and also which egress hierarchical virtual schedulers, is used for each subscriber host (dynamic or static).
- The system performs SLA enforcement functions on a per subscriber SLA profile instance basis. SLA enforcement functions include QoS (classification, filtering and queuing), security (filtering), and accounting.

When the enhanced mode is enabled on a SAP (see Subscriber SAPs), first, the router ensures that existing configurations on the SAP do not prevent correct enhanced mode operation. If any one of the following requirements is not met, enhanced mode operation is not allowed on the SAP:

- Anti-spoofing filters must be enabled and configured as IP+MAC matching.
- Any existing static subscriber hosts must have:
  - An assigned subscriber identification string.
  - An assigned subscriber profile name.
  - An assigned SLA profile name.
- The system must have sufficient resources to create the required SLA profile instances and schedulers.

When the router successfully enables the enhanced mode, the current dynamic subscriber hosts are not touched until a DHCP message event occurs that allows re-population of the dynamic host information. Thus, over time, the dynamic subscriber host entries are moved from SAP-based queuing and SAP-based filtering to subscriber-based queuing and filtering. If a dynamic host event cannot be processed because of insufficient resources, the DHCP ACK message is discarded and the previous host lease information is retained in the system.

9.3.2 ESM for IPv6

ESM for IPv6 is supported on the 7750 SR chassis or the 7450 ESS chassis. ESM for IPv6 is supported with RADIUS as the backend authentication and authorization mechanism.

9.3.2.1 Models

9.3.2.1.1 PPPoE host

For PPPoE, the BNG suggests the IPv6CP protocol to the client during the session setup phase if the appropriate attributes have been returned by the RADIUS server on authentication. The RADIUS attribute

that indicates the setup of a PPPoE host is Framed-IPv6-Prefix, which should contain a /64 prefix for the client.

When a PPPoE host has successfully completed the IPv6CP negotiation, the BNG transmits a Router Advertisement to the PPPoE host containing the suggested prefix and any other options that are configured. The client may use this information to pick one or more addresses from the suggested prefix; all addresses within the prefix are forwarded toward the client.

Alternatively, the Recursive DNS Server (RDNSS) option as defined in RFC 6106, *IPv6 Router Advertisement Options for DNS Configuration*, can be included in the IPv6 Router Advertisements for DNS name resolution of IPv6 SLAAC hosts. See also DNS and NBNS name server IP addresses for subscriber sessions.

9.3.2.1.2 PPPoE RG

Initially, a PPPoE RG follows the same procedure as a PPPoE host: the BNG receives a prefix from RADIUS (in this case through a Delegated-IPv6-Prefix attribute), which is used as a trigger to suggest the IPv6CP protocol to the client. The prefix that is suggested to the client should have the same prefix length as configured under the **subscriber>if>ipv6** *node* (delegated-prefix-length). This length should be between 48 and 64 bits, inclusive.

After the IPv6CP protocol has completed, however, the client should run the DHCPv6 protocol over its PPPoE tunnel to receive a Delegated Prefix (IA_PD) and optionally IPv6 DNS server information. This Delegated Prefix can then be subdivided by the client and distributed over its downstream interfaces. During DHCPv6, no extra RADIUS requests are made; the information is stored during the initial (PPPoE or PPP) authentication until the client starts DHCPv6.

Only after DHCPv6 has completed, the IPv6 subscriber host is instantiated and the BNG starts sending Router Advertisements (if configured.) The router advertisements do not contain any prefix information, which has already been provided by DHCPv6, but it is used as an indication to the client that its default gateway should be the BNG.

9.3.2.1.3 IPoE host/RG

Similar to an IPv4 DHCP client, a DHCPv6 client is authenticated at its Solicit message, where it can request one or more addresses or prefixes. The address and prefix types supported are IA_NA (Non-Temporary Address) through the Alc-IPv6-Address RADIUS attribute and IA_PD (Delegated Prefix) through the Delegated-IPv6-Prefix attribute. Contrary to the IPv4 case, the BNG always replies to a DHCPv6 request because the client may request more than one address or prefix simultaneously and not all of the requests may be honored.

The DHCPv6 protocol handling and Router Advertisement behavior are similar to the PPPoE RG case above, with the exception that for an IA_NA address, the entire /64 prefix containing the address is allocated to the client.

For SLAAC prefix assignment, authentication is triggered on router-solicit message. The SLAAC prefix can be assigned statically or dynamically. For a static SLAAC prefix, frame-ipv6-prefix, RADIUS attribute is used. For dynamic SLAAC prefix assignment from a local pool, Alc-slaac-ipv6-pool, RADIUS attribute is used.

9.3.2.2 Setup

IPv6 ESM hosts are only supported in the Routed CO model (both VPRN and IES).

At the IPv6 node under the subscriber interface level, the length of the prefixes that are offered is defined through the delegated-prefix-length option. This setting is fixed for the subscriber interface and cannot be changed after subscriber prefixes are defined.

Subscriber prefixes define the ranges of addresses that are offered on this subscriber interface. By default, only these subscriber prefixes are exported to the routing protocols to keep the routing tables small. There are three types of subscriber interfaces:

- **wan-host**

  A range of prefixes that are assigned to PPPoE hosts and as DHCPv6 IA_NA addresses. These prefixes are always /64.

- **pd**

  A range of prefixes that are assigned as DHCPv6 IA_PD prefixes for DHCPv6 IPoE clients and for PPPoE RGs. The length of these prefixes is defined by the delegated-prefix-length.

- **both**

  When both 'wan-host' and 'pd' are defined, the subscriber prefix is a range that can be used for both previous types. However, the delegated-prefix-length is restricted to /64 in this case.

The subscriber interface prefix can also be provisioned through RADIUS. The RADIUS VSA Alc-IPv6-Sub-If-Prefix requires a prefix and the prefix type. The prefix type can be **pd**, **wan**, or **both**. The prefix is then installed on the subscriber interface where the subscriber is instantiated. The prefix state is tied to the state of the subscriber. After the subscriber session ends, the prefix is removed from the subscriber interface and subsequently from both the FDB and the RIB. This feature can be used as an alternative to unnumbered subscriber interfaces, where the subscriber interface prefix does not need to be predetermined. However, by installing the prefix after authentication, the subscriber interface becomes numbered. In an unnumbered subscriber interface all subscriber routes are installed whereas in a numbered subscriber interface only the subscriber interface prefix is advertised, therefore reducing the number of advertised routes significantly. The RADIUS-installed prefix can then be advertised through a routing protocol. Subscriber interface prefixes are under the protocol **direct** type similar to other router interfaces. To advertise only the subscriber interface prefix installed by RADIUS, origin **aaa** can be used in the router policy.

The IPv6 node under the group interface contains the DHCPv6 proxy configuration and the router advertisement configuration.

9.3.2.3 64-bit and 128-bit WAN mode

Subscriber interfaces are created as 64-bit WAN mode interfaces by default. At the time of creation, the subscriber interface can also be created as a 128-bit WAN mode interface. After the subscriber interface is created, the WAN mode cannot be changed. To change the WAN mode, the 64-bit subscriber interface must be removed and then recreated as 128-bit. This section describes the differences between 64-bit and 128-bit WAN modes.

In a 64-bit WAN mode subscriber interface, the following rules apply.

- The system differentiates each subscriber using only the first 64 bits of the WAN address (each host must have a unique /64 prefix, with the exception of bridge host.) This differentiation includes the ability to identify individual subscribers and to apply different subscriber profile and SLA-profiles to each subscriber.

- For IPoE bridge hosts, when a group of hosts shares a prefix, all hosts must share the same SLA-profile.
- Each SLAAC subscriber must use a unique /64 prefix. IPoE bridge hosts can share the same SLAAC prefix and must also share the same SLA-profile.
- Each IPv6 data-trigger host must use a unique /64 prefix.
- The DHCPv6 server must be set up to assign each host with a unique /64 prefix. The 7750 SR local DHCP server assigns each subscriber with a unique /64 prefix by default.

64-bit WAN mode is applicable in deployment models where each subscriber is assigned a unique /64 WAN-prefix which can be used for DHCP or SLAAC.

In a 128-bit WAN mode subscriber interface, the following rules apply.

- It is not recommended to change a subscriber interface from unnumbered to numbered (or the other way around). If the subscriber interface must be changed, all ESM hosts under the subscriber should be removed first.
- The system can uniquely identify each subscriber using the full 128-bit WAN address. Each 128-bit WAN host can have its own unique SLA and Subscriber profile.
- When provisioning a numbered subscriber interface, an IPv6 address or a prefix can be assigned. The mask for the address can range from 32 to 127. If the mask is less than 96, an internal /96 route is generated when a WAN host is created (by DHCP IPv6 IANA). This automatically-generated /96 route is used for subscriber lookup. These /96 routes are visible in the RIB and occupy a route entry in the system forwarding table. A /96 route can serve approximately 4.2 billion WAN hosts. Therefore, a single /96 prefix or address should be able to accommodate all WAN hosts terminating on a subscriber interface. Nokia recommends, when using 128-bit WAN mode to configure subscriber interface, use addresses or prefixes with a mask length of 96 to 127. When using 128-bit WAN mode, it is not recommended to assign individual subscribers unique /64 prefixes, because the system generates an internal /96 route for each host, therefore overloading the routing table.
- Adding and removing a prefix or address from the subscriber interface in 128-bit WAN mode may trigger /96 routes to be generated or deleted, which can impact subscriber service. Nokia recommends performing this action during off-peak hours.
- The auto-generated /96 routes needed for subscriber lookup are tagged in the RIB as "Wan Mode 128 Route".

```
# show router 2000 route-table ipv6 2001:db8:2000:100::/96 extensive
===============================================================================
Route Table (Service: 2000)
===============================================================================
Dest Prefix             : 2001:db8:2000:100::/96
  Protocol              : LOCAL
  Age                   : 00h06m26s
  Preference            : 0
  Wan Mode 128 Route    : Yes
  Next-Hop              : N/A
    Interface           : sub-int-2
    QoS                 : Priority=n/c, FC=n/c
    Source-Class        : 0
    Dest-Class          : 0
    Metric              : 0
    ECMP-Weight         : N/A
-------------------------------------------------------------------------------
No. of Destinations: 1
===============================================================================
```

These routes can be leaked between local VPRN services on the same router using MP-BGP export and import policies as they are needed for subscriber lookup in extranet topologies. The auto-generated /96 routes are not advertised in the BGP RIB; instead, the prefix configured on the subscriber interface should be used in BGP.

- 128-bit WAN mode is supported in unnumbered subscriber interfaces. Each WAN host generates a /128 route.
- The **allow-unmatching-prefixes** command can be performed on a numbered subscriber interface in 128-bit WAN mode. This functionality can be used as a subnet migration tool, but must be performed without hosts under the subscriber interface. Changing a subscriber interface from numbered to unnumbered (or the other way around) impacts subscriber service.
- IPv6 DHCP IANA subscribers can be assigned incremental 128-bit addresses.
- IPoE-bridge mode is only recommended to be configured with 64-bit WAN mode. For 128-bit WAN mode, the system generates at least one /96 prefix per subscriber to help lookup. Because each subscriber interface has a limited number of allowed prefixes, generating a /96 per subscriber reduces scalability. If 128-bit WAN mode is used, each DHCP IANA host can have a distinct SLA profile. In 64-bit WAN mode, all DHCP IANA hosts must share the same SLA profile.
- For IPoE bridge SLAAC hosts, hosts sharing the same /64 prefix must share the same SLA profile. IPoE bridge SLAAC hosts do not differ from 128-bit or 64-bit WAN mode.
- Each IPv6 data-trigger host can use a unique /128 address.
- The DHCP IA_NA host for both PPPoE and IPoE hosts can assign incremental 128-bit addresses.
- For retail VPRN that requires 128-bit WAN mode support, the wholesale subscriber interface must also be configured with 128-bit WAN mode.
- SLAAC hosts and DHCP WAN hosts must not share the same prefix.
- The following host types are not supported:
  - GTP hosts
  - hybrid access hosts
  - WLAN hosts
  - default hosts

9.3.2.3.1 Migration from 64-bit to 128-bit WAN mode

It may be beneficial in some deployments for operators to migrate from 64-bit to 128-bit WAN mode. For example, the ability to assign consecutive 128-bit address and minimize the subnet required for Residential Gateway or Cable Modem IPv6 DHCP IANA WAN management address.

A 64-bit WAN mode subscriber interface cannot be changed into a 128-bit WAN mode subscriber interface in real time. To migrate to a 128-bit WAN mode subscriber interface, the 64-bit WAN mode subscriber interface must be a removed and re-created. The 64-bit configuration must be copied, shut down, and the configuration removed. The configuration can be pasted back with the 128-bit mode added to the subscriber interface. Below are some migration scenarios.

9.3.2.3.1.1 Migration of PPPoE and IPoE DHCP hosts on MSAPs

Change RADIUS, Diameter, and LUDB in advance of migrations to minimize service impact. Ensure that MSAP stickiness is disabled and idle sticky MSAPs are removed. Nokia recommends performing this migration during a maintenance window.

To prepare to migrate PPPoE and IPoE DHCP hosts on MSAPs, perform the following steps.

1. Create new subscriber and group interfaces for the 128-bit WAN mode.
2. Update the database for the host-related MSAP parameters. This includes AAA (both RADIUS and Diameter) and LUDB. The updated database directs subscribers to the new subscriber and group interface.

A migration can be performed for either PPPoE and LNS hosts or IPoE DHCP-based hosts. The migration is dependent on subscriber deletion.

For PPPoE and LNS hosts, when a host disconnects their session, the next session is migrated. To speed up the migration, and depending on the RG capability, manually clearing the session could trigger the RG to re-connect through PPPoE immediately, and migrate to the new interface.

When migrating IPoE DHCP-based hosts, Nokia recommends changing both the current DHCPv4 and DHCPv6 lease time and rebind times to one hour or more. It is important to migrate only a small sample size to control the number of DHCP renews. Subscribers are migrated in the following three ways.

- Subsets of subscribers are migrated to the new 128-bit interface without any end user action. For example, the end user does not need to reset their modem or RG. The new authentication establishes the host on the new interface. A maintenance window is not required.
- To migrate the remaining subscribers, enable the drain function on the DHCP server to stop all DHCP lease renewal. However, the DHCPv4 and DHCPv6 lease for a subscriber may end at different times. When both the DHCPv4 and DHCPv6 leases end, the subscriber is removed from the system. Depending on the RG/CM capability, it may send a DHCP request immediately for a new DHCPv4 and DHCPv6 addresses. The subscriber is now created on the new 128-bit subscriber interface without any action by the end user.
- For any remaining subscribers that have not been migrated, after the migration window has waited for the one hour lease time, action by the end user may be required. The operator must shut down both the subscriber interface and group interface, and clear all remaining hosts from those interfaces. The end user is then required to manually reboot the RG to send a DHCP discover/solicit.

9.3.2.3.1.2 Migration of PPPoE and IPoE DHCP hosts on static SAPs

Nokia recommends performing this migration during a maintenance window. This migration process is service-impacting.

When migrating IPoE DHCP-based hosts, Nokia recommends changing both the current DHCPv4 and DHCPv6 lease time and rebind time to one hour or more. It is also recommended to migrate only a small sample size to control the number of DHCP renews. After all leases have been changed to a shorter lease time, perform the following steps to prepare for the migration.

1. Remove the old subscriber and group interfaces. This may require manual clearing of subscriber sessions.
2. Re-create new subscriber and group interfaces for the 128-bit hosts (including SAPs).
3. Apply the appropriate AAA (RADIUS and Diameter) and LUDB changes (new 128-bit IPv6 addresses for all hosts), if any.

Following these preparation steps, a migration can be performed for either PPPoE and LNS hosts or IPoE DHCP-based hosts. The migration is dependent on subscriber deletion.

For PPPoE and LNS, when hosts disconnect their session, the RG may try to re-connect by PPPoE immediately and migrate to the new interface.

For IPoE hosts, new subscribers are automatically migrated upon logging in. Some end customers may be required to manually reboot the RG to send a DHCP discover/solicit.

9.3.2.3.1.3 Migration of data-trigger hosts

Nokia recommends performing this migration during a maintenance window. This migration process is service-impacting. Before the migration can begin, the data-trigger node on the group interface must be shut down, then all the data-triggered hosts must be cleared.

To migrate data-trigger hosts:

1. Remove the existing subscriber interface and recreate the new 128-bit subscriber interface, group interface, and related parameters. Removing the existing subscriber interface may require clearing data-triggered subscribers from the system.
2. Update the AAA (RADIUS or Diameter) or LUDB parameters related to the MSAP or SAP. If the end-subscriber is assigned a new IP address, all traffic using the old IP-address is dropped until the new IP address is in use.
3. To complete the migration, the subscriber must send a data packet for authentication using the new assigned IP address.

9.3.2.4 Behavior

9.3.2.4.1 Dual-stack

Clients may support both IPv4 and IPv6 simultaneously (dual-stack hosts.) In this case, one subscriber host entry is created for the IPv4 address family and one for the IPv6 instance. The scaling limits apply for all entries, regardless of address type.

For DHCP, these subscriber hosts are fully independent (as they are set up through different protocols), but for PPPoE hosts or RGs, the ESM information in both subscriber host entries is linked together through the PPPoE session.

9.3.2.4.2 Router Advertisements

Router Advertisement (RA) messages begin immediately after the subscriber host is instantiated and unsolicited messages are sent in the interval defined in the configuration. Apart from unsolicited RAs, the client may also send a router solicitation (RS) to explicitly request the information. RAs are throttled so that they are not sent more than once every three seconds.

The Router Advertisement Policy feature overrides the group interface RA configuration for hosts on a specific MAC on a specified SAP. The policy is applied directly to the sending instance where it sends periodic RAs. The policy can be applied at authentication or by CoA during the subscriber session. The RA policy can be used in the following ways.

- When applied to an IPoE session or a PPPoE session, the RA policy is applied to all IPv6 hosts within the session.
- When applied to a subscriber (regardless of whether it is session-based), the RA policy is applied to all IPv6 hosts within the subscriber.
- When applied to a host within a subscriber, the RA policy is applied only to the IPv6 hosts for the particular MAC (for IPoE session) or the particular MAC and PPP session (for PPPoE session).

The prefix option inside the RA policy allows independent prefix options for subscribers that use bridge hosts. The bridge hosts can consist of both DHCPv6 and SLAAC, and are represented as stateful and stateless within the policy respectively. Within the policy, the **autoconfig** flag is not configurable and is disabled by default for the DHCPv6 address and enabled by default for SLAAC. For SLAAC hosts, if the **autoconfig** flag is enabled inside the RA policy along with the SLAAC prefix, the **autoconfig** flag for the DHCPv6 address or prefix is not enabled as a result. The timers for either SLAAC and DHCPv6 prefixes can also be configured independently.

The router advertisement policy has a separate configuration for stateless and stateful operations. The general recommendation is to configure the valid and preferred lifetimes for longer than the minimum RA interval to ensure the subscriber has a valid address to use between each RA interval. If this general rule is not followed, the subscriber can deprecate the SLAAC prefix between each RA interval and experience service interruptions. As the minimum RA interval is approximately 15 minutes, the valid and preferred lifetime values should be at least 15 minutes. Shorter valid and preferred lifetime values can impact the system's scalability. The stateful RA has a static option and a dynamic option when configuring the valid and preferred lifetime values. If the static option is used, the valid and preferred lifetime values should be greater than the RA interval. For the dynamic option, the **auto-lifetimes** feature derives the valid and preferred lifetime values from the DHCPv6 lease. Therefore, the RA and DHCPv6 have the same valid and preferred lifetime values.

SLAAC hosts are assigned prefixes, where the full Global Unicast Address (GUA) is not known. Regardless of the **force-mcast** configuration, the destination IP address for an RA to an SLAAC host is always a multicast IP address, with one exception. If the feature **allow-multiple-wan-address** is enabled and the same host (same MAC on the same SAP and same device) has a DHCPv6 NA address, the NA address is used for the unicast RA. The MAC address can either be a multicast or unicast address, depending on the configuration of **force-mcast**.

Table 12: RA policy behavior describes the behavior of the system when the RA policy VSA is included in authentication, CoA, and re-authentication. The RA policy that is sent from RADIUS may not yet be provisioned in CLI, and therefore may not exist in the system.

Table 12: RA policy behavior

	Authentication	CoA/tools CoA	Re-authentication
BRG	An RA policy does not need to exist. The RA policy becomes active when a matching RA policy is provisioned. If an RA policy does not exist, the RA parameters configured under the group interface are used.	An RA policy must exist; otherwise, a NACK is sent in response to the CoA.	An RA policy must exist; otherwise, all VSAs and the RA policy are ignored. An SNMP trap is raised.

	Authentication	CoA/tools CoA	Re-authentication
Subscriber is session-based (for example, an IPoE session)	An RA policy does not need to exist for the IPv4 host. The RA policy becomes active when a matching RA policy is provisioned. ESM IPv6 host creation fails when a policy does not exist. If an RA policy does not exist, the RA parameters configured under the group interface are used.	An RA policy must exist; otherwise, a NACK is sent in response to the CoA.	An RA policy must exist; otherwise, all VSAs and the RA policy are ignored. An SNMP trap is raised.
Subscriber has a dual-stack host and is not session-based	An RA policy does not need to exist for the IPv4 host. The RA policy becomes active when a matching RA policy is provisioned. ESM IPv6 host creation fails when a policy does not exist. If an RA policy does not exist, the RA parameters configured under the group interface are used.	An RA policy must exist; otherwise, a NACK is sent in response to the CoA.	An RA policy must exist; otherwise, all VSAs and the RA policy are ignored. An SNMP trap is raised.
IPv4 host that is not session-based	An RA policy must exist. Otherwise, the subscriber setup is rejected.	An RA policy must exist; otherwise, a NACK is sent in response to the CoA.	An RA policy must exist; otherwise, all VSAs and the RA policy are ignored. An SNMP trap is raised.
Dual-stack host that is not session-based, where the CoA is targeted to an IPv4 host only	N/A	An RA policy must exist; otherwise, a NACK is sent in response to the CoA.	N/A
Dual-stack host that is not session-based, where the CoA is targeted to an IPv6 host only	N/A	An RA policy must exist; otherwise, a NACK is sent in response to the CoA.	N/A
IPoE linking (both session-based and not session-based)	An RA policy does not need to exist for the IPv4 host. The RA policy becomes active when a matching RA policy is provisioned.	An RA policy must exist; otherwise, a NACK is sent in response to the CoA.	An RA policy must exist; otherwise, all VSAs and the RA policy are ignored. An SNMP trap is raised.

	Authentication	CoA/tools CoA	Re-authentication
	ESM IPv6 host creation fails when a policy does not exist. If an RA policy does not exist, the RA parameters configured under the group interface are used.		
PD host as managed to IPv4 (both session-based and not session-based)	An RA policy does not need to exist for the IPv4 host. The RA policy becomes active when a matching RA policy is provisioned. ESM IPv6 host creation fails when a policy does not exist. If an RA policy does not exist, the RA parameters configured under the group interface are used.	An RA policy must exist; otherwise, a NACK is sent in response to the CoA.	An RA policy must exist; otherwise, all VSAs and the RA policy are ignored. An SNMP trap is raised.

9.3.2.4.2.1 Router Advertisement policy limitations

The following are RA policy limitations.

- As a result of a general limitation on CoA, a maximum of 32 bridge hosts can be updated. This limitation exists in the case of BRG.
- RA policies are not supported for DSM.
- RAs are configured to be sent at certain intervals. It is highly recommended that this interval is considered when configuring the prefix lifetimes. For example, if the interval is configured for one hour and the prefix lifetime is configured for 30 minutes, the (SLAAC) host is removed from the BNG before the next RA is sent.
- If the parameters within the RA policy are modified, the parameters only take effect on the next interval.

9.3.2.4.3 CoA and disconnect-request

For IPv6 subscriber hosts, RADIUS-triggered mid-session changes and session terminations may identify the subscriber host to be changed by the same address or prefix that was originally returned from RADIUS. Only one address attribute (framed-IP address, framed-IPv6-prefix, delegated-IPv6-prefix or Alc-IPv6-address) may be provided in a single request.

For PPPoE clients, changing either the IPv4 or IPv6 information results in both the v4 and v6 subscriber host being modified (if they are contained within the same PPPoE session).

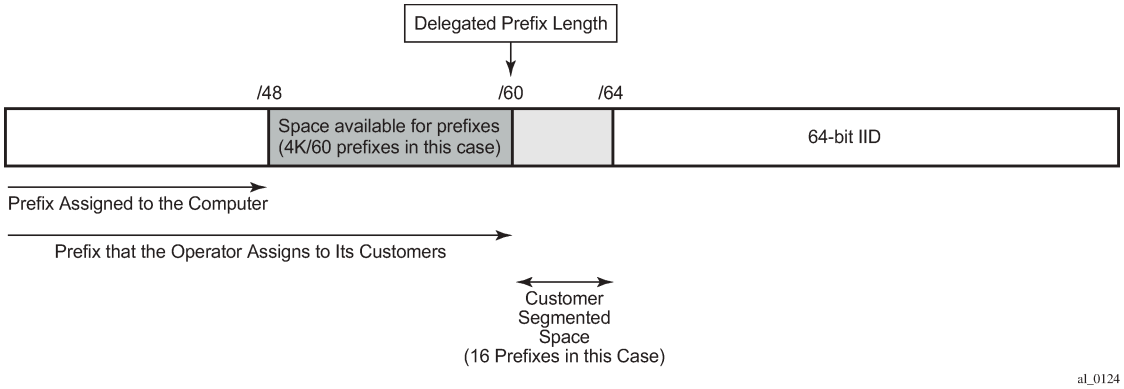
The only CoA action that is allowed for IPv6 hosts is a change of ESM strings; creation of new hosts and forcing a DHCPv6 RENEW is not supported.

9.3.2.5 Delegated prefix length

The delegated prefix length (DPL) is applicable to subscriber-hosts with IPv6 Prefix (IA-PD) assigned by the DHCPv6 Server. An IPv6 prefix is more similar to a route than it is to an IP address. The length of the prefix plays crucial role in forwarding decisions, antispoofing, and prefix assignment through DHCPv6 pools in the local DHCPv6 server.

The structure of an IPv6 prefix is shown in Figure 69: IPv6 prefix.

Figure 69: IPv6 prefix



For example, a DHCPv6 server prefix pool contains an aggregated (configured) IPv6 prefix from which the delegated prefixes are carved out. In Figure 69: IPv6 prefix this aggregated IPv6 prefix has length of /48. In addition, the DHCPv6 server needs to know the length of the delegated prefix (in the above case /60). These two values are marking the boundary within which a unique delegated prefix is selected.

The delegated prefix length can be obtained using:

- **RADIUS**
  - Delegated-IPv6-Prefix attribute that contains the prefix and the length (Delegated-IPv6-Prefix = AAAA:BBBB::/56). The DPL in this case is /56.
  - Alc-Delegated-IPv6-Prefix-Length VSA (to be used in conjunction with the DHCPv6 pool name - Alc-Delegated-IPv6-Pool VSA)
- **LUDB**

  Configured by LUDB per IPoEv6/PPPoEv6 host:

  This is to be used along with the DHCPv6 pool name (ipv6-delegated-prefix-pool) defined under the same CLI hierarchy.

  CLI syntax:

```
configure
subscriber-mgmt
    local-user-db <name>
        ipoe | ppp
            host <name>
                ipv6-delegated-prefix-length [48 to 64]
```

  Alternatively, the entire prefix, including the DPL can be returned by LUDB.

CLI syntax:

```
configure
    subscriber-mgmt
        local-user-db <name>
                ipoe | ppp
                    host <name>
                        ipv6-delegated-prefix <ipv6-prefix/prefix-length>
```

- **DHCPv6 server**

  Each DHCPv6 pool can optionally be configured with a DPL

  CLI syntax:

```
configure
     service/router
        dhcp6
            local-dhcp-server <name>
                pool <pool-name>
                    delegated-prefix-length [48 to 127]
```

Configured statically under the **ipv6** CLI node of subscriber interface. In this case, the DPL is fixed for all subscriber hosts under the subscriber interface.

CLI syntax:

```
configure
    service ies/vprn
        subscriber-interface <ip-int-name>
                ipv6
                    delegated-prefix-length [48 to 64] | variable
```

9.3.2.5.1 Order of preference for DPL

If the DPL is statically provisioned under the **sub-if**>**ipv6** hierarchy, all hosts under this subscriber interface inherits this fixed DPL. In case that the DPL is provided by LUDB or RADIUS in addition to static configuration under the subscriber interface then the LUDB or the RADIUS one not match the DPL that is statically provisioned under the subscriber-interface. Otherwise, the prefix instantiation in 7450 ESS and 7750 SR fails.

Note that the **no delegated-prefix-length** command under the **sub-if**>**ipv6** hierarchy means that the DPL is set to a default-value of 64.

When the **delegated-prefix-length** commands under the **sub-if**>**ipv6** hierarchy is set to variable, prefixes under such subscriber interface can have different lengths and the DPL can be configured by one of the following:

- LUDB
- RADIUS
- DHCP Server

9.3.2.5.2 DHCP server address utilization and delegated prefix length

If the delegated prefix length is variable, for each consecutive address allocation request for the specified delegated prefix, the DHCPv6 server allocates the prefix at the end of the last delegated lease with the same delegated prefix length. This minimizes the address space fragmentation within the configured prefix.

9.3.2.6 DHCPv6 Relay Agent

A DHCPv6 Relay Agent can support a 7450 ESS and 7750 SR DHCPv6 local server (same or remote chassis) and a third party DHCPv6 external server.

An incoming DHCPv6 client message is relayed within the Relay-Forward message specified in RFC 3315, *Dynamic Host Configuration Protocol for IPv6 (DHCPv6)*. If the server responds with a valid address/prefix, the ESM process attempts to install it. If it fails, the DHCPv6 Relay Agent sends an explicit RELEASE to the server. There is no retransmission of DHCPv6 Relay-Forwards in the case of failure, it requires the client to re-start or re-send the original DHCPv6 message.

A Lightweight DHCPv6 Relay Agent may insert Relay Agent Information including the Interface ID option between the DHCPv6 client and the DHCPv6 Relay Agent.

Additional Relay Agents (non-LDRA) between the DHCPv6 client and the DHCPv6 Relay Agent are not supported.

DHCPv6 Reconfigure messages received from an external DHCPv6 server are forwarded to the DHCP client, if a corresponding DHCPv6 lease exists. The Reconfigure message can be sent in a unicast message to the client or encapsulated in a Relay-Reply message to the DHCPv6 relay agent. The DHCPv6 Reconfigure message is dropped if no corresponding DHCPv6 lease exists.

9.3.2.6.1 Configuring a DHCPv6 Relay Agent

A DHCPv6 Relay Agent is configured in the IPv6 DHCP6 context of a group-interface:

```
config>service>vprn>sub-if>grp-if>ipv6>dhcp6# relay ?
config>service>ies>sub-if>grp-if>ipv6>dhcp6# relay ?
  - no relay
  - relay

 [no] client-applications - Configure the set of DHCP6 relay server client
                            applications
 [no] description     - Description for DHCPv6 relay
 [no] link-address    - Configure the link address of the DHCPv6 relay messages
 [no] option          + Configure the DHCPv6 Relay information options
 [no] server          - Configure the DHCPv6 server IPv6 address
 [no] shutdown        - Administratively enable/
disable DHCPv6 relay on this interface
 [no] source-address  -
 Configure the source IPv6 address of the DHCPv6 relay messages
```

Up to eight DHCPv6 servers can be provisioned to be served by a DHCPv6 Relay Agent. A Relay-Forward is send to all servers and the Relay-Replies from all servers are sent to the client.

The "client-applications" parameter specifies if the Relay Agent can be used for IPoE (dhcp) or PPP (ppp) hosts. Optional configuration parameters:

- **description**

A free configurable description string.

- **link-address**

  The link address field in the DHCPv6 Relay-Forward message header.

  The link address can be configured to enable link-address based pool selection in a 7450 ESS and 7750 SR DHCPv6 local server. The address must be one of the IPv6 prefixes configured at the ipv6 subscriber-prefixes context for a subscriber interface. If not configured, the system selects one of the prefixes.

- option: allows to configure following options to be inserted in the Relay-Forward message:
  - **Interface-Id [18]**

    The interface ID option identifies the interface on which the DHCPv6 client message is received. The format options are the following:

    - **ascii-tup**le:

      host-name | *service-id* | *group-interface-name* | *sap-id*

    - **ifindex**

      Interface index for the group interface

    - **sap-id**

      SAP identifier (port and VLANs)

    - **string** <*string*>

      A free configurable string (up to 80 characters)

  - **Remote-Id [37]**

    Relay Agent Remote Id option contains the DHCPv6 client DHCP Unique Identifier (DUID).

- source-address: the source-address of the Relay-Forward messages.

  If not configured, the outgoing interface IPv6 address is used. The source-address configuration is mandatory for a DHCP Relay Agent in a VPRN service when the DHCPv6 server is reachable by a tunneled next-hop (MPLS).

9.3.2.7 DHCPv6 Relay to third party DHCPv6 external server

When the DHCPv6 Relay Agent is relaying to a third party DHCPv6 external server, following conditions should be met:

- The third party DHCPv6 server must return a unique IA_PD IPv6 delegated prefix (/64 or lower) for each allocation. The length of the IA_PD IPv6 delegated prefix must match the delegated-prefix-len configured on the subscriber interface on the 7750 DHCP L3 relay. This length is also included in the Relay-Forward message as PFX_LEN option (3) in a Vendor-Specific-Information-Option (17).
- For IPv6oE routed CPEs, the 3rd party DHCPv6 server must return a unique IA_NA IPv6 address (/128) from a different /64 subnet for each allocation.
- For IPv6oE hosts behind bridged CPE's
  - the third party DHCPv6 server must return a unique IA_NA IPv6 address (/128) from a different /64 subnet for each allocation (host) that belongs to a different CPE.

– the third party DHCPv6 server may return a unique IA_NA IPv6 address (/128) from the same /64 subnet for allocations (hosts) that belong to the same CPE and that are attached to the same VLAN (SAP) on the BNG.

Following information is available to the third party DHCPv6 server in a Vendor-Specific-Information-Option (17) included in the Relay-Forward message:

- WAN_POOL option (1) contains the pool name from which the IA_NA IPv6 address should be allocated.
- PFX_POOL option (2) contains the pool name from which the IA_PD IPv6 delegated prefix should be allocated.
- PFX_LEN option (3): contains the IA_PD IPv6 delegated prefix length that should be allocated.

9.3.2.8 DHCPv6 local server

A local DHCPv6 pool server for both addresses (IA_NA) and prefixed (IA_PD) manages the address and prefixes sent to either routing gateways or hosts.

Because IPv6 home networks lack NAT, the IPv6 addresses delegated to a routing gateway are in turn assigned to hosts in the home. These addresses are assigned with reasonably long (but configurable) lifetimes so the loss of the WAN connection does not result in the IPv6 hosts in the LAN losing their IPv6 addresses. One consequence of these long lifetimes is that the IPv6 hosts retains any IPv6 address provided the valid-lifetime is greater than zero. If an operator delegates a prefix and then at a later time delegate a second IPv6 prefix, a host may end up with two or more valid prefixes. This situation affects IPv6 source address selection and may result in impaired service.

To overcome the problems of multiple IPv6 prefixes in the home, the operator must ensure that the individual subscriber has the same IPv6 prefix even across modem reboots (that is, if a subscriber session is destroyed and later re-created, an attempt should be made to use the previously delegated prefix). In Release 8.0, the operator used RADIUS for all address and prefix assignment, but in Release 9.0, with the introduction of the local DHCPv6 server, it requires the 7750 to process and maintain some state even after a session disconnects.

For the DHCPv6 local server to function, a DHCPv6 relay or proxy function must also operate alongside ESM. For the purposes of this document, to relay means to implement a DHCPv6 Relay as indicated in RFC 3315: a relay encapsulates the client DHCP message within a DHCP Relay-Forward message and unicasts it to a specified destination.

A proxy is an internal concept. Unlike a DHCPv6 relay, the DHCPv6 proxy does not encapsulate the client message in a Relay-Forward, nor does it send packets toward the Local DHCPv6 Server. The DHCPv6 proxy is exclusively used as an interface between the RADIUS Access-Accept or local user database lookup and the DHCPv6 client in the consumer device.

The use of the DHCPv6 relay or proxy function depends on the attributes returned from authentication phase (RADIUS or LUDB).

1. DHCPv6 proxy:

   If only the IPv6 address/prefix information is provided (Framed-IPv6-Prefix, Alc-IPv6-Address or Delegated-IPv6-Prefix).

2. DHCPv6 relay:

   - If no IPv6 address/prefix (Framed-IPv6-Prefix, Alc-IPv6-Address or Delegated-IPv6-Prefix) and no IPv6 pool (Framed-Pool, Delegated-Pool) information provided.

- If no IPv6 address/prefix (Framed-IPv6-Prefix, Alc-IPv6-Address or Delegated-IPv6-Prefix) and IPv6 pool (Framed-Pool, Delegated-Pool) information provided.

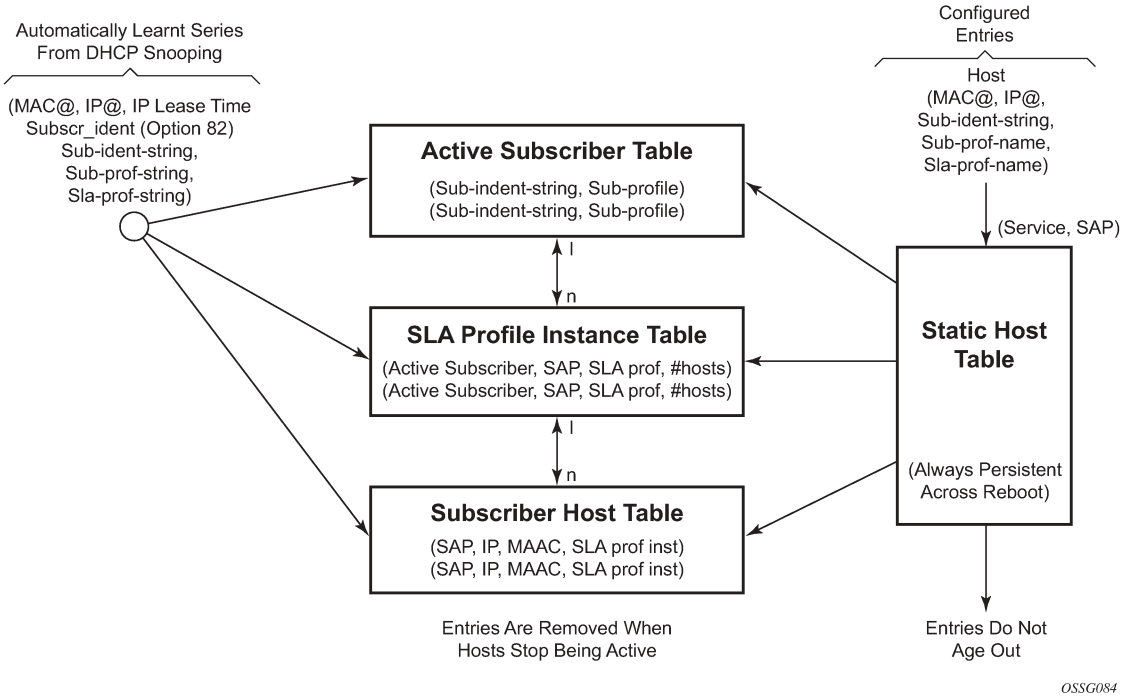
3. If both IPv6 address/prefix (Framed-IPv6-Prefix, Alc-IPv6-Address or Delegated-IPv6-Prefix) and IPv6 pool (Framed-Pool, Delegated-Pool) information are present, the DHCP packet is DROPPED.

9.3.3 Dynamic subscriber host processing

9.3.3.1 Dynamic tables

To support all processing for ESM, several tables are maintained in the router (Figure 70: ESM dynamic tables).

Figure 70: ESM dynamic tables



9.3.3.1.1 Active subscriber table

An entry is created in the active subscriber table when the first host (either dynamic or static) is created with a specific subscriber identification string. The entries are grouped by their subscriber identification string.

Fields for each entry in the active subscriber table include:

- The subscriber identification string (see Subscriber).
- In use subscriber profiles (see Subscriber profile).

9.3.3.1.2 SLA profile instance table

An entry is created in the SLA profile instance table when the first subscriber host on a certain SAP is created that uses a specific SLA profile. All subsequent hosts of the same subscriber on the same SAP that use the same SLA profile are associated with this entry. When the last host on this SAP, using this SLA profile disappears, the SLA profile instance is deleted from the table and the associated queues are removed.

SLA profile instances cannot span multiple subscriber SAPs. If subscriber hosts from the same subscriber exist on multiple SAPs and are associated with the same SLA profile template, a separate SLA profile instance is created for each SAP.

Fields for each entry in the SLA profile instance table include:

- Active subscriber
- SAP
- SLA profile
- Number of active subscriber hosts that share this instance

9.3.3.1.3 Subscriber host table

An entry is created in the subscriber host table if anti-spoofing is enabled as well as:

- The first host (dynamic or static) with a specific IP and MAC combination is created. If the anti-spoof is IP only, the MAC address is masked to all 0's. If anti-spoof is MAC, only the IP address is 0.0.0.0. All dynamic hosts and static hosts with the same IP and MAC combination are associated with the same subscriber host entry. If the anti-spoof type includes IP (IP-only or IP/MAC), there can be at most two hosts associated with the entry: one dynamic and one static. If the anti-spoof type is MAC-only, there can be a combination of several dynamic and static hosts associated with the entry.
- The non-prof-traffic is provisioned. Both IP and MAC address are all 0's.

Fields for each entry in the subscriber host table include:

- SAP
- IP address
- MAC address
- SLA profile instance (enhanced mode only)

9.3.3.1.4 DHCP lease state table

An entry in the DHCP lease state table is created for each dynamic host. Fields for each entry in the lease state table include:

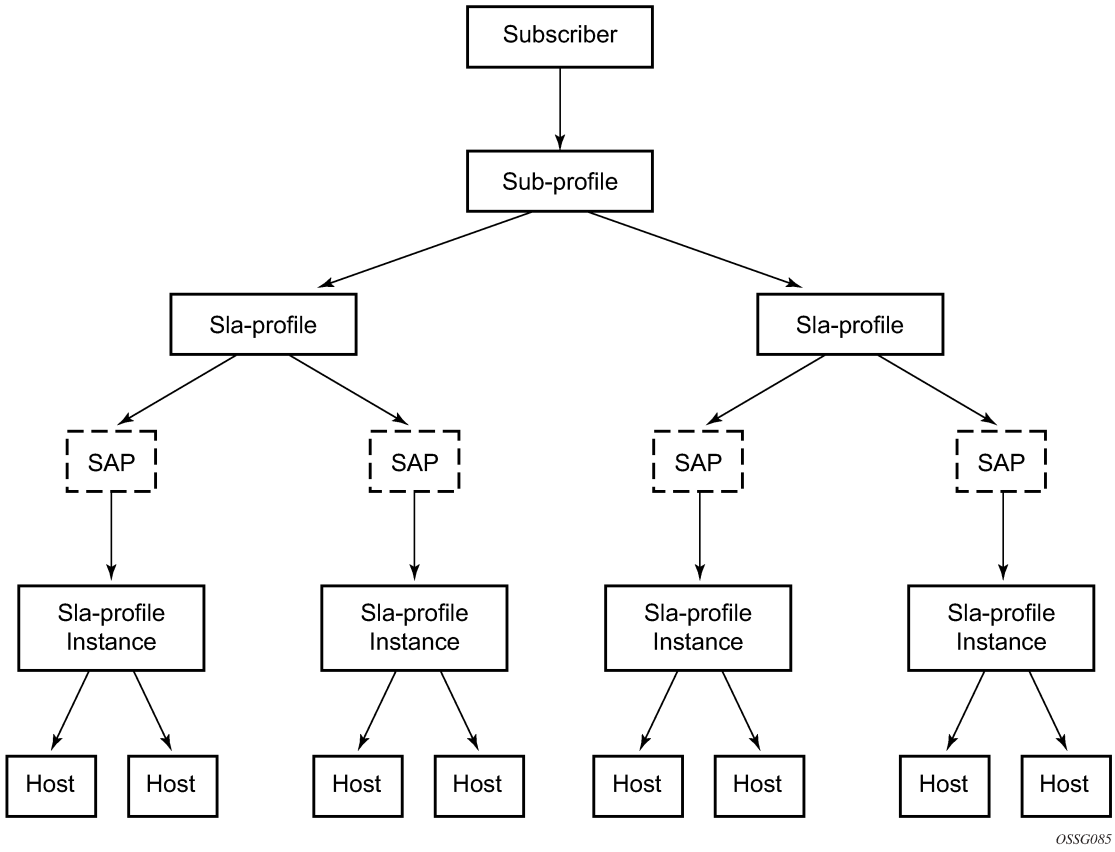
- Assigned IP address
- Assigned MAC address
- Persistence key

9.3.4 ESM entities

Figure 71: Relationship between ESM entities illustrates the relationship between the main entities in Enhanced Subscriber Management:

- A subscriber is associated with only one subscriber profile.
- A subscriber can be associated with one or more SLA profile (a VPLS service with 2 different SAPs can have different SLA profiles for the same subscriber).
- A maximum of one SLA profile instance is generated (including ingress and egress queues) per SAP per SLA profile.
- One or more hosts can be assigned to each SLA profile instance (these share the same queues).

Figure 71: Relationship between ESM entities



9.3.4.1 Instantiating a new host

When a DHCP ACK is received for a new subscriber host on a particular SAP:

- The ACK message is parsed using the appropriate script.
- An entry is generated in the subscriber host table with indexes:
  - The SAP on which the host resides

- The assigned IP address
- The assigned MAC address and as lookup parameters:
  - the subscriber profile
  - the SLA profile to be used (derived from using the script)

If this is the first host of a subscriber, an HQoS scheduler is instantiated using the ingress and egress scheduler policies referred to in the subscriber profile. Otherwise, if the subscriber profile of the new host equals the subscriber profile of the existing subscriber, the new host is linked to the existing scheduler. If the subscriber profile is different from the subscriber profile of the existing subscriber, a new scheduler is created and all the hosts belonging to that subscriber are linked to this new scheduler. The new subscriber profile does conflict with the subscriber profile provisioned for a static host or non-sub-traffic under the same SAP.

If this is the first host of a subscriber on a particular SAP using a particular SLA profile, an SLA profile instance is generated and added to the SLA profile instance table. This includes instantiating a number of queues, according to the ingress and egress QoS profiles referred to in SLA profile, optionally with some specific overrides defined in the SLA profile. Otherwise the host is linked to the existing SLA profile instance for this subscriber on this SAP.

Note:

- Any QoS and IP filter policies defined on the SAP are still processed even if Enhanced Subscriber Management is enabled on the SAP. For IPv4 traffic that is dropped because of anti-spoofing, counters, logging, and mirroring can be used. All other Layer 2 traffic that is never blocked by anti-spoofing can be processed by applying a QoS policy on the SAP and can still be classified differently, by the dot1p value.
- If insufficient hardware resources (queues) or software resources (profile instances) are available to support the new host, the DHCP ACK is dropped and an event is generated.

9.3.4.2 Packet processing for an existing host

Whenever an IP packet arrives on a subscriber-facing SAP on which Enhanced Subscriber Management (ESM) is enabled, a lookup is done in the subscriber host table using as the index the SAP, source IP address, and source MAC address.

- If there is no entry, this means that the host is not using the assigned IP address, so the packet is dropped.
- If there is an entry, this refers to the subscriber profile and SLA profile to be used.

9.3.5 ESM host lockout

This feature increasingly penalizes hosts that fail repeated login attempts within a configurable time interval. This is done by holding off on creation attempts for these hosts for a configured but adaptable time period. A transient failure, because of a misconfiguration, is quickly corrected and does not prevent the host from logging in within a reasonable amount of time. At the same time, a malicious client or a constantly misconfigured client is locked-out and does not take up resources impacting other clients.

A lockout time per host supports exponential back-off with each retry and failure cycle, starting with a configured minimum value and increasing up to a configured maximum. The lockout time can be reset to

the configured minimum value if there is no failed retry within a configured time threshold. The configurable values include:

CLI syntax:

```
lockout-reset-time seconds
lockout-time [minseconds] [maxseconds]
max-lockout-hosts hosts
```

If multiple retries/failure cycles occur within the lockout time, then lockout period is exponentially increased starting from configured minimum value up to the configured maximum value. The lockout is reset to the minimum value if there is no failed retry till this lockout time.

This mechanism is supported for both single and dual-stack PPPoE and IPoE (DHCP) hosts over 1:1 or N:1 static or managed SAPs. The hold-off timer maintenance is on a per host basis (as follows):

- For 1:1 VLAN (PPPoE or IPoE hosts) per <VLAN, MAC address>
- For N:1 VLAN (PPPoE or IPv4oE hosts) per <VLAN, agent-circuit-id, agent-remote-id, MAC@>
- For 1:1 VLAN (IPv6oE hosts) per <VLAN, DUID>

A show lockout state for hosts is supported, for one or more of <SAP, MAC@, agent-circuit-id, agent-remote-id>.

A clear lockout state is supported for hosts for one or more of <SAP, MAC@, agent-circuit-id, agent-remote-id>.

Any changes in configured lockout values do not apply to hosts currently under lockout and only applies after these hosts are out of lockout.

9.3.5.1 Functionality

ESM lockout is supported for dual-stack PPPoE hosts, L2TP LAC hosts, dual-stack IPoE hosts, and ARP hosts. ESM Lockout tracks the following:

- PPPoE PADI and PADR
- DHCPv4 discover, DHCPv4 request, DHCPv6 solicit, DHCPv6 request
- ARP Request
- PPPoE session disconnect after successful session establishment

During lockout, authentication and ESM host creation is suppressed. A lockout context is created when a client first enters lockout. The context maintains state and timeout parameters for the lockout. If a lockout policy is configured for the underlying SAP for a host that has failed authentication or host creation, the host enters lockout for the configured minimum time (1 to 86400 seconds). When the lockout time expires, normal authentication and ESM host creation is resumed on relevant PPP or DHCP messages. In case of another failure, the host again enters the lockout state. The lockout time for the host on each failure is exponentially increased up to the configured maximum time (1 to 86400 seconds). The lockout time for a client is reset to the configured minimum value, and the corresponding lockout context is deleted, if there is no authentication (and host creation) failure within a configured amount of time that needs to elapse after the client initially enters lockout. This time is called the **lockout-reset-time**.

The host identification for lockout includes <SAP, MAC@, circuit ID, remote ID>.

9.3.6 ANCP and GSMP

9.3.6.1 Access Node Control Protocol management

Access Node Control Protocol Management (ANCP) can provide the following information to the router:

- ANCP can communicate the current access line rate to the router. This allows the router to adjust the H-QoS subscriber scheduler with the correct rate or potentially change alarm when the rate goes below a set threshold. This allows a policy manager to change the entire policy when the rate drops below a minimal threshold value. The ANCP actual upstream synchronization rate is mapped to the ingress while ANCP actual downstream synchronization rate is mapped to the egress.
- The router can send DSL line OAM commands to complete an OAM test from a centralized point or when operational boundaries prevent direct access to the DSLAM.

When ANCP is used with ESM, the **ancp-string** string can be returned from the Python script or from RADIUS. If not returned it defaults to the subscriber ID.

ANCP version 0x31 and 0x32 are both supported and are autodetected at the start of each ANCP session. Within version 0x32, partitioning is also supported.

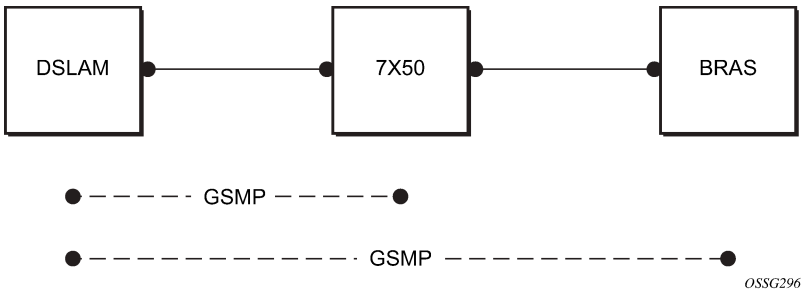
Multiple partitions from the same access node are also supported. If partitions are used, they are automatically detected during the start of an ANCP session.

9.3.6.1.1 Static ANCP management

As depicted in Figure 72: Static ANCP management example, a DSLAM is connected to an aggregation network that is connecting the DSLAM to a BRAS. ANCP is used to provide SAP level rate management. The DSLAM in this application maintains multiple ANCP connections. The primary connection is to the BRAS, providing rate and OAM capabilities while the secondary is to the router to provide rate management.

7750 SR and 7450 ESS:

Figure 72: Static ANCP management example



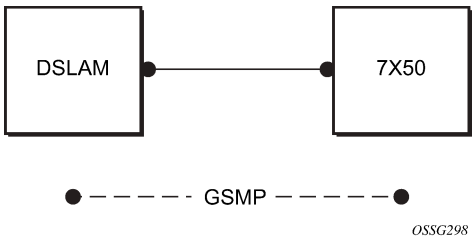
9.3.6.1.2 ESM dynamic ANCP

In this application ANCP is used between the DSLAM and the router to provide line control. There are multiple attributes defined as described below. Figure 73: ESM dynamic ANCP example depicts the connectivity model.

This application is used to communicate the following from the DSLAM to the router (the policy control point):

- Subscriber rate
- OAM

Figure 73: ESM dynamic ANCP example



9.3.6.1.3 ANCP string

To support node communication with the access device the line rate, OAM commands, and so on. the node can use an ANCP string that serves as a key in the out-of-band channel with the access node. The string can be either provisioned in the static case, retrieved from RADIUS or from the Python script.

9.3.6.1.4 ANCP persistency support

Persistency is available for subscriber’s ANCP attributes and is stored on the on-board compact flash card. ANCP data stays persistence during an ISSU as well as nodal reboots. During recovery, ANCP attributes are first restored fully from the persistence file and incoming ANCP sessions are temporarily on hold. Afterwards new ANCP data can overwrite any existing values. This new data is then stored into the compact flash in preparation for the next event.

9.3.6.2 General Switch Management Protocol Version 3

General Switch Management Protocol version 3 (GSMPv3) is a generic protocol that allows a switch controller node to establish and maintain connections with one or more nodes to exchange operational information. Several extensions to GSMPv3 exist in the context of broadband aggregation. These extensions were proposed to allow GSMPv3 to be used in a broadband environment as more information is needed to synchronize the control plane between access nodes (such as DSLAMs) and broadband network gateways (such as BRAS).

In the TPSDA framework, nodes fulfill some BRAS functionality, where per subscriber QoS enforcement is one of the most important aspects. To provide accurate per-subscriber QoS enforcement, the network element not only knows about the subscriber profile and its service level agreement but it is aware of the dynamic characteristics of the subscriber access circuit.

The most important parameters in this context are the subscriber-line capacity (DSL sync-rate) and the subscriber's channel viewership status (the actual number of BTV channels received by the subscriber in any point in time). This information can be then used to adjust parameters of aggregate scheduling policy.

Besides, the above-mentioned information, GSMPv3 can convey OAM information between a switch controller and access switch. The node can operate in two roles:

- **as the intermediate controller**

  The router terminates a connection from the DSLAM.

- **as the terminating controller**

  The router fulfills full the roll of BRAS.

The DSL forum working documents recommends that a dedicated Layer 2 path (such as, a VLAN in an Ethernet aggregation network) is used for this communication to provide a specific level of security. The actual connection between DSLAM and BRAS is established at TCP level, and then individual messages are transported.

9.3.6.3 DHCP client mobility

Client mobility allows the node to use host monitoring (SHCV, ANCP, split DHCP) to remove network and server state when a host is removed locally. This allows for MAC addressed learned and pinned to move based on policy parameters.

Subscriber Host Connectivity Verification (SHCV) configuration is mandatory. This allows clients to move from one SAP to another SAP in the same service. This is only applicable in a VPLS service and group interfaces.

The first DHCP message on the new SAP with same MAC address (and IP address for group-interfaces) triggers SHCV and is always discarded.

SHCV checks that the host is no longer present on the SAP where the lease is currently populated to prevent spoofing. When SHCV detects that the host is not present on the original SAP, the lease-state is removed. The next DHCP message on the new SAP can initiate the host.

9.3.6.4 DHCP lease control

DHCP lease control allows the node to be configured to present a different lease to the client. This can be used to monitor the health of the client.

9.3.7 Using scripts for dynamic recognition of subscribers

Whenever a host belonging to a subscriber is activated (when a PC or set-top box (STB) is turned on), the host typically requests an IP address from the network using DHCP. See the DHCP Management section for an explanation of DHCP and DHCP snooping in the router.

The DHCP ACK response from the DHCP server can be parsed and the contents of the message can be used to identify the class to which this host belongs, and therefore, the QoS and security settings to apply.

The information necessary to select these settings can be codified in, the IP address by the DHCP server and the Option 82 string inserted by the DSLAM or other access node.

9.3.7.1 Python Language and Programmable Subscriber Configuration Policy

Python Language and Programmable Subscriber Configuration Policy (PSCP) is an identification mechanism using the Python scripting language. The PSCP references a Python script that can use regular expressions to derive the *sub-ident-string*, *sub-profile-string* and *sla-profile-string* from the DHCP

response. A tutorial of regular expressions is beyond the scope of this guide, and can be found on the Internet (see https://docs.python.org/2/howto/regex.html).

A tutorial of Python is beyond the scope of this guide but can be found on the Internet (see http://www.python.org/).

Example scripts, using some regular expressions, can be found in Sample Python Scripts. See the Python Script Support for ESM section for more information about the service manager scripting language.

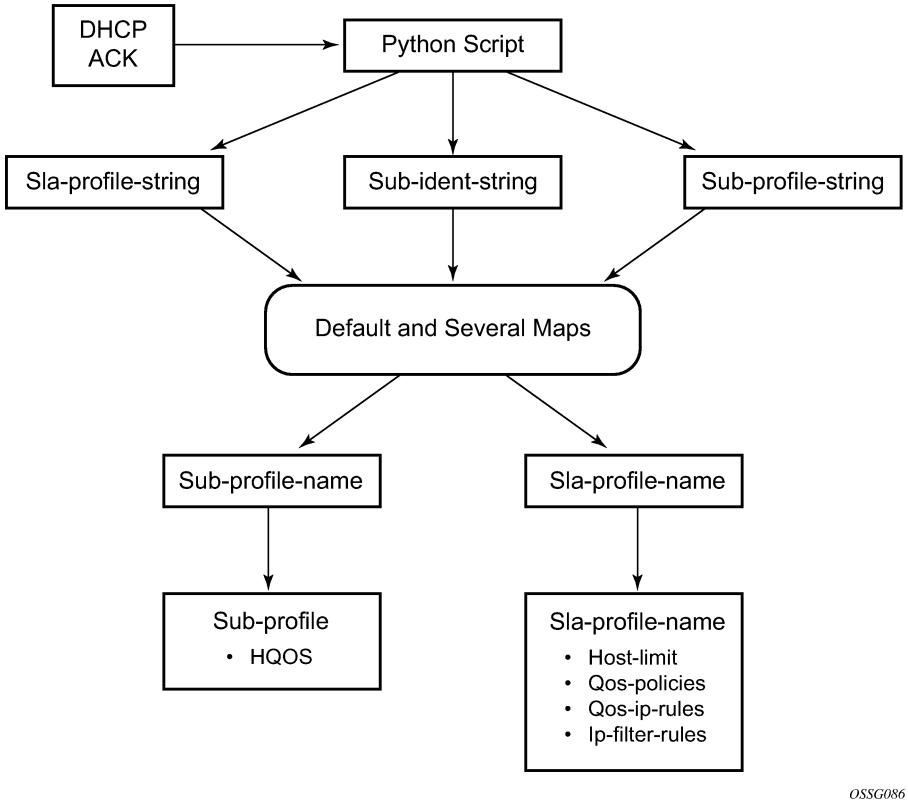
One or more scripts can be written by the operator and stored centrally on a server (in a location accessible by the router). They are loaded into each router at bootup.

Note that if a centrally stored script is changed, it is not automatically re-loaded onto the router. The reload must be forced by executing the **shutdown** and **no shutdown** commands on the affected URLs.

9.3.7.2 Determining the subscriber profile and SLA profile of a host

Figure 74: Data flow in determining subscriber profile and SLA profile describes the data flow while determining which subscriber profile and SLA profile to use for a specified subscriber host based on a snooped/relayed DHCP ACK for that subscriber host.

Figure 74: Data flow in determining subscriber profile and SLA profile



An incoming DHCP ACK (relayed or snooped) is processed by the script provisioned in the sub-ident-policy defined in the SAP on which the message arrived. This script outputs one or more of the following strings:

sub-ident

identifies the subscriber (always needed)

sub-profile

identifies the subscriber class (optional)

sla-profile

identifies the SLA Profile for this subscriber host (optional)

These strings are used for a lookup in one or more maps to find the names of the sub-profile and sla-profile to use. If none of the maps contained an entry for these strings, the names are determined based on a set of defaults.

Only when the names for both the sub-profile and sla-profile are known, the subscriber host can be instantiated. If even no default is found for either profile, the DHCP ACK is dropped and the host does not gain network access.

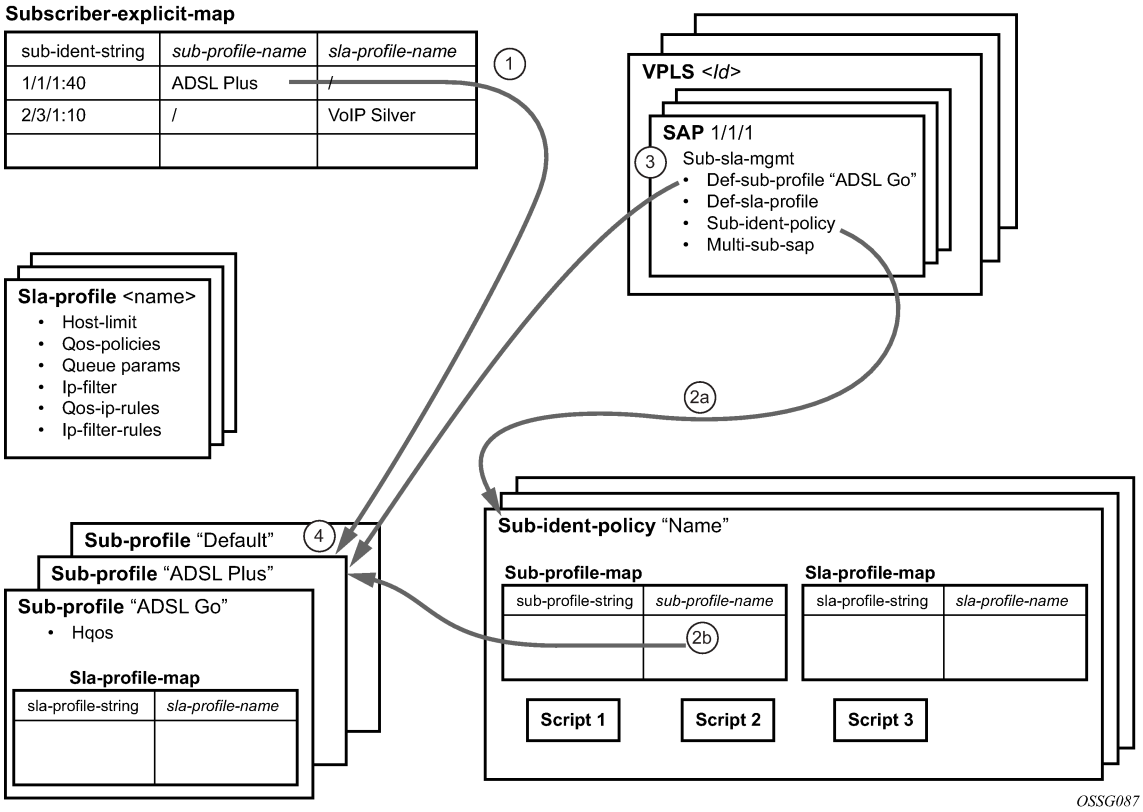
9.3.7.3 Determining the Subscriber Profile

All hosts (devices) belonging to the same subscriber are subject to the same HQoS processing. The HQoS processing is defined in the sub-profile. A sub-profile refers to an existing scheduler policy and offers the possibility to overrule the rate of individual schedulers within this policy.

Because all subscriber hosts of one subscriber use the same scheduler policy instance, they must all reside on the same I/O module.

Figure 75: Determining the subscriber profile shows how the *sub-profile* is derived, based on the *sub-ident* string, the *sub-profile* string and the provisioned data structures. The numbers associated with the arrows pointing toward the subscriber profiles indicate the precedence of the checks.

Figure 75: Determining the subscriber profile



1. A lookup in the **explicit-subscriber-map** is done with the sub-ident string returned by the script. If a matching entry is found, the sub-profile-name (if defined) is taken. Otherwise:
2. If a **sub-ident-policy** is defined on the SAP, a lookup is done on its **sub-profile-map** with the sub-profile string from the script. The sub-profile-name is taken from the entry.

   If no entry was found, then:
3. If provisioned, the **sub-profile-name** is taken from the **def-sub-profile** attribute on the SAP. If not provisioned, then:
4. The **sub-profile** with the name "default" is selected (if provisioned). If this is not provisioned, there are no other alternatives, the ACK is dropped, and the host does not gain access.

9.3.7.4 Determining the SLA profile

For each host that comes on-line, the router also needs to determine which SLA profile to use. The SLA profile determines for this host:

- The QoS-policies to use:
  - classification
  - queues/policers
  - queue mapping

- The egress scheduling policies use egress HQoS
- The IP filter to use.

The SLA profile also has **host-limits** and **session-limits** attributes that limit the number of hosts or sessions per SLA profile instance.

The classification and the queue mapping are shared by all the hosts on the same forwarding complex that use the same QoS policy (by their SLA profile).

The queues and policers are shared by all the hosts (of the same subscriber) on the same SAP that are using the same SLA profile. In other words, queues and policers are instantiated when, on a specific SAP, a host of a subscriber is the first to use a specific SLA profile. This instantiation is referred to as an SLA profile instance. Ingress queues can be parented to a scheduler referenced in the ingress of a subscriber profile. Egress policers and queues can be parented to a scheduler referenced in the egress of a subscriber or SLA profile, or to a port scheduler.

A scheduler policy can be applied to the egress an SLA profile, allowing its schedulers to be the parent for its queues and for its tier 1 schedulers to be parented to a scheduler in a scheduler policy applied to the egress of a subscriber profile or a Vport, or to a port scheduler applied to a port or Vport. Configuring scheduler overrides is allowed for SLA profile egress schedulers. The configuration of a scheduler policy in the egress of an SLA profile is supported for all host types only on Ethernet interfaces. It is not supported for ESM over MPLS pseudowires, nor is HQoS adjustment and host tracking supported on its schedulers.

The following show, monitor and clear commands are available related to the SLA profile scheduler:

```
show qos scheduler-hierarchy subscriber sub-ident-string sla-profile sla-profile-
name
sap sap-id [scheduler scheduler-name] [detail]
```

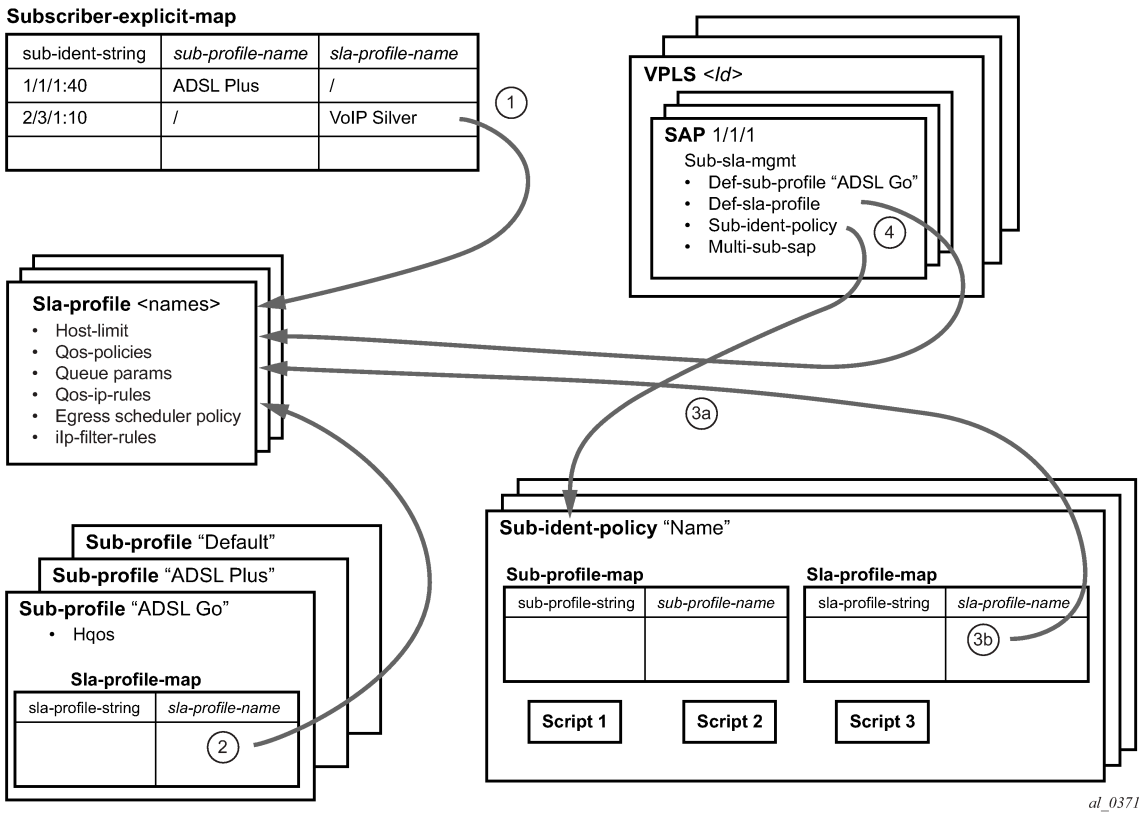
The **show qos scheduler-hierarchy subscriber** command (shown above) displays the scheduler hierarchy with the SLA profile scheduler as the root. Note that if the SLA profile scheduler is orphaned (that is when the scheduler has a parent which does not exist) then the hierarchy is only shown when the **show** command includes the **sla-profile** and **sap** parameters.

If the SLA profile scheduler is orphaned (that is when the scheduler has a parent which does not exist) then the hierarchy is only shown when the show command includes the sla-profile and SAP parameters.

```
monitor qos scheduler-stats subscriber sub-ident-string [interval seconds] [repeat
repeat] [absolute|rate] sap sap-id sla-profile sla-profile-name

show qos scheduler-stats subscriber sub-ident-string sap sap-id sla-profile sla-
profile-name [scheduler scheduler-name]

clear qos scheduler-stats subscriber sub-ident-string sap sap-id sla-
profile sla-profile-name [scheduler scheduler-name]
```

Figure 76: Determining the SLA profile shows a graphical description of how the SLA profile is derived based on the subscriber identification string, the SLA profile string and the provisioned data structures. The numbers on the arrows toward the SLA profile indicate the *priority* of the provisioning (the lower number means the higher priority).

Figure 76: Determining the SLA profile



1. A lookup is done with the sub-ident string returned by the script in the **explicit-subscriber-map**. If a matching entry is found, the sla-profile-name is taken from it – if defined. Otherwise:
2. A lookup with the sla-profile string from the script is done in the **sla-profile-map** of the sub-profile found earlier. The sla-profile-name from the found entry is taken. If no entry was found, then:
3. A lookup is done with the sla-profile string in the **sla-profile-map** of the **sub-ident-policy** configured on the SAP. The sla-profile-name from the found entry is taken. If no **sub-ident-policy** was configured on the SAP or no entry was found, then:
4. If provisioned, the sla-profile-name is taken from the **def-sla-profile** attribute on the SAP. If not provisioned, there are no more alternatives, the ACK is dropped, and the host does not gain access.

9.3.7.4.1 SLA profile instance sharing

Each subscriber host or session has an SLA Profile Instance (SPI) associated with it. The SPI, is by default, determined by the subscriber ID, the SLA profile name, and the SAP where the subscriber host or session is active. See Figure 71: Relationship between ESM entities.

SPIs with the same SLA profile name, have the same configuration, however, the following functions are effective per SPI:

- enforcing the different host limits
- instantiation of queues and policers

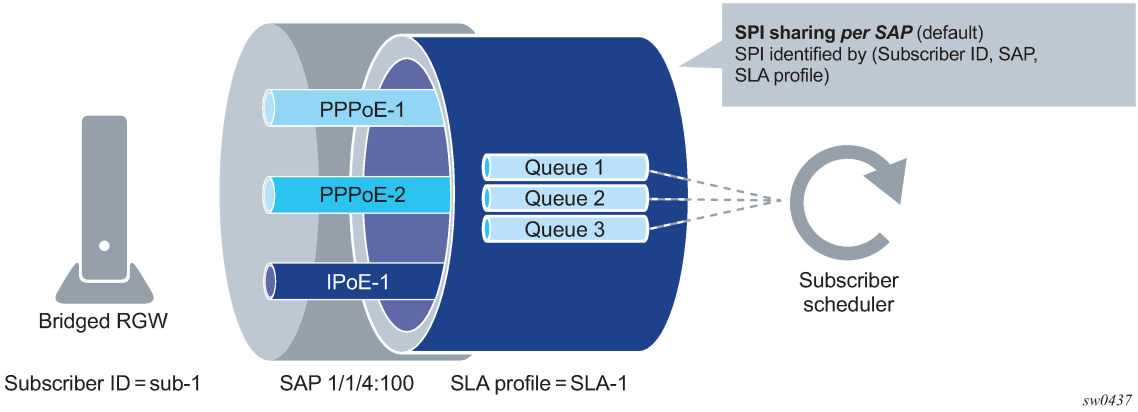
- accounting statistics
- credit control functions

For a bridged Residential Gateway deployment, typically multiple IPoE or PPPoE sessions per subscriber are active on the BNG. The next sections describe the different SPI sharing mechanisms that apply for multiple subscriber sessions from the same subscriber, that are active on the same SAP with the same SLA profile name assigned.

9.3.7.4.1.1 SPI sharing per SAP

By default, all subscriber sessions or hosts from the same subscriber, active on the same SAP and with the same SLA profile assigned, share an SPI. The default SPI sharing is per SAP, as depicted in Figure 77: SLA profile instance per SAP.

Figure 77: SLA profile instance per SAP



With SPI sharing per SAP, traffic from all subscriber sessions on a specific SAP and with the same SLA profile associated are mapped to the same set of queues and policers for QoS handling. Statistics from these queues and policers are also used in accounting. Per-host or per-session accounting modes cannot report counters for individual sessions unless their traffic is mapped in separate queues.

SPI sharing per SAP is the default configuration in an SLA profile and applies to PPPoE sessions, IPoE sessions (enabled on the group-interface) and IPoE hosts (IPoE sessions are disabled on the group-interface):

Example:

```
configure
    subscriber-mgmt
        sla-profile "sla-profile-1"
            def-instance-sharing per-sap
```

9.3.7.4.1.2 SPI sharing per session

If QoS handling or accounting per-IPoE or per-PPPoE session is required, then the SPI sharing is configured to per-session sharing in the SLA profile:

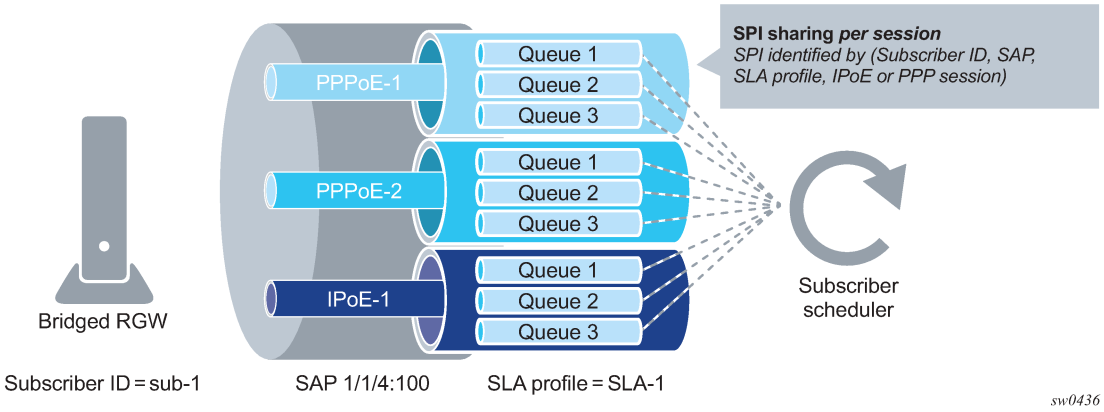
Example:

```
configure
    subscriber-mgmt
        sla-profile "sla-profile-1"
            def-instance-sharing per-session
```

Per-session sharing applies to PPPoE sessions and IPoE sessions (enabled on the group interface). An IPoE host setup fails when IPoE sessions are disabled on the group interface and per-session sharing is configured.

Each IPoE or PPPoE session from the same subscriber, active on the same SAP and having the same SLA profile assigned, has its own set of queues and policers. Per-session SPI sharing is depicted in Figure 78: SLA profile instance per session.

Figure 78: SLA profile instance per session



Note: SPI sharing per session is not supported on HS MDA and on HSQ with **hs-sla-mode single**.

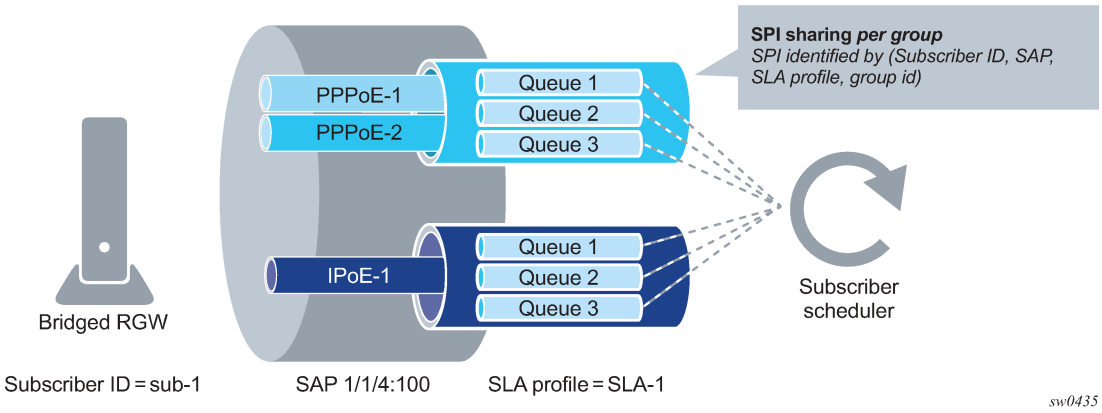
9.3.7.4.1.3 SPI per group

When even more granular control is needed over which sessions share an SPI, an SPI sharing group identifier can be specified during IPoE or PPPoE session authentication. This overrides the default SPI sharing method for that session as configured in the SLA profile.

Per-group SPI sharing is depicted in Figure 79: SLA profile instance per group. The same SPI is shared by all IPoE and PPPoE sessions from the same subscriber, active on the same SAP, having the same SLA Profile assigned and having the same SPI sharing group identifier.

Note:
SPI sharing per group is not supported on HSQ with **hs-sla-mode** single.

Figure 79: SLA profile instance per group



The SPI sharing group identifier is an integer value in the range 0 to 65535 and can be specified in authentication using:

- A local user database lookup:

```
configure
   subscriber-mgmt
      local-user-db local-user-db-name
         ipoe | ppp
            host host-name
               identification-strings
                  spi-sharing-group-id <group-id>
```

  Configure **no spi-sharing-group-id** to apply the **def-instance-sharing** method as configured in the SLA profile.

- RADIUS, by including the [241.26.6527.47] Alc-SPI-Sharing-Id VSA in an Access-Accept message:
  - Value "group:<group-id>" to enable SPI sharing per group identifier
  - Value "default" to apply the **def-instance-sharing** method as configured in the SLA profile

  See the 7450 ESS, 7750 SR, and VSR RADIUS Attributes Reference Guide for a detailed description of the attribute.

- Diameter NASREQ, by including the Vendor specific [NOKIA-1036] Alc-SPI-Sharing grouped AVP in an AA-Answer message:

```
Alc-SPI-Sharing ::= < AVP Header: 1036 >
                          {Alc-SPI-Sharing-Type}
                          [Alc-SPI-Sharing-Id]
```

  - To enable SPI sharing per group identifier
    - [NOKIA-1037] Alc-SPI-Sharing-Type = 2
    - [NOKIA-1038] Alc-SPI-Sharing-Id = <group-id>
  - To apply the **def-instance-sharing** method as configured in the SLA profile use [NOKIA-1037] Alc-SPI-Sharing-Type = 0

  See the Diameter and Diameter Applications, AA-Answer Message — Accepted Authorization AVPs section for a detailed description of the attribute.

- Diameter Gx, by including the Vendor specific [NOKIA-1036] Alc-SPI-Sharing grouped AVP in a CCA message:

```
Alc-SPI-Sharing ::= < AVP Header: 1036 >
                        {Alc-SPI-Sharing-Type}
                        [Alc-SPI-Sharing-Id]
```

  – To enable SPI sharing per group identifier
    - NOKIA-1037] Alc-SPI-Sharing-Type = 2
    - NOKIA-1038] Alc-SPI-Sharing-Id = <group-id>
  – To apply the **def-instance-sharing** method as configured in the SLA profile use [NOKIA-1037] Alc-SPI-Sharing-Type = 0

  See *7750 SR and VSR Gx AVPs Reference Guide* for a detailed description of the attribute.
- Python:
  – alc.dts.setESM module: alc.dtc.SpiSharingGroupId = <group-id>
  – alc.esm.set module: alc.esm.SpiSharingGroupId = <group-id>

  See the DHCP Management, ESM-Related Python Variables section for further details.

Per-group sharing applies to PPPoE sessions and IPoE sessions (enabled on the group interface). An IPoE host setup fails when IPoE sessions are disabled on the group interface and an SPI sharing group identifier is specified.

9.3.7.4.1.4 Dynamic changes of SLA profile and SPI sharing

During the lifetime of an IPoE or PPPoE session, the SLA profile and the SPI sharing can change. Such a dynamic change can be triggered by re-authentication, RADIUS CoA, or Diameter Gx RAR by specifying a new SLA profile and optionally, an SPI sharing group ID.

Table 13: Dynamic changes of SPI Sharing describes the different transitions in SPI sharing because of re-authentication, RADIUS CoA, or Diameter Gx RAR.

Table 13: Dynamic changes of SPI Sharing

from **-** **to**	**SLA profile and SPI sharing info provided for dynamic change**
per sap - per sap	SLA profile = <SLA profile name> SLA profile with "def-instance-sharing per-sap" [SPI sharing type = default] • Optional. Value must be default when present • SPI sharing ID must not be present
per session -	[SLA profile = <SLA profile name>] SLA profile with "def-instance-sharing per-session"

from - to	SLA profile and SPI sharing info provided for dynamic change
per session	[SPI sharing type = default] • Optional. Value must be default when present • SPI sharing ID must not be present
per group - per group	SLA profile = <SLA profile name> Optional. Current SLA profile name is used when not specified SPI sharing type = group SPI sharing ID = <group-id> Overrides the "def-instance-sharing" configured in the SLA profile
per sap - per group	SLA profile = <SLA profile name> Optional. Current SLA profile name is used when not specified SPI sharing type = group SPI sharing ID = <group-id> Overrides the "def-instance-sharing" configured in the SLA profile
per session - per group	SLA profile = <SLA profile name> Optional. Current SLA profile name is used when not specified SPI sharing type = group SPI sharing ID = <group-id> Overrides the "def-instance-sharing" configured in the SLA profile
per group - per sap	SLA profile = <SLA profile name> Optional. Current SLA profile name is used when not specified SPI sharing type = default SPI sharing ID must not be present ---------- SLA profile = <SLA profile name> SLA profile with "def-instance-sharing per-sap"
per group - per session	[SLA profile = <SLA profile name>] • Optional. Current SLA profile name is used when not specified • SLA profile with "def-instance-sharing per-session" SPI sharing type = default SPI sharing ID must not be present ----------

from - to	SLA profile and SPI sharing info provided for dynamic change
	SLA profile = <SLA profile name> SLA profile with "def-instance-sharing per-session"
per sap - per session	SLA profile = <SLA profile name> SLA profile with "def-instance-sharing per-session" [SPI sharing type = default] • Optional. Value must be default when present • SPI sharing ID must not be present
per session - per sap	SLA profile = <SLA profile name> SLA profile with "def-instance-sharing per-sap" [SPI sharing type = default] • Optional. Value must be default when present • SPI sharing ID must not be present

9.3.7.4.1.5 Identifying the SPI

An SPI is uniquely identified by the following characteristics:

- the subscriber identifier
- the SAP on which the subscriber session is active
- the SLA profile name
- An SPI sharing identifier that has two parts to support overlapping ids between groups and sessions:
  - SPI sharing type: per SAP, per IPoE session, per PPP session, per group
  - SPI sharing id:
    - an integer value determined by the system for SPI sharing per session
    - an integer value in the range 0 to 65535 specified by the user for SPI sharing per group
    - not required for SPI sharing per SAP

The following are examples for SPI representations in the system:

- SPI sharing per SAP

```
A:PE-1# show service active-subscribers detail
===============================================================================
Active Subscribers
===============================================================================
-------------------------------------------------------------------------------
Subscriber sub-01 (sub-profile-12)
-------------------------------------------------------------------------------
--- snip---
-------------------------------------------------------------------------------
```

```
(1) SLA Profile Instance
- sap:[1/1/4:1201.41] (IES 1000 - group-int-1-1)
- sla:sla-profile-12
-------------------------------------------------------------------------------
--- snip---
A:PE-1# show service active-subscribers hierarchy
===============================================================================
Active Subscribers Hierarchy
===============================================================================
-- sub-01 (sub-profile-12)
   |
   +-- sap:[1/1/4:1201.41] - sla:sla-profile-12
       |
       |-- PPP-session - mac:00:51:00:00:01:41 - sid:1 - svc:1000
       |   |
       |   +-- 10.1.1.141 - IPCP
       |
       |-- PPP-session - mac:00:51:00:00:01:42 - sid:2 - svc:1000
       |   |
       |   +-- 10.1.1.142 - IPCP
       |
       +-- PPP-session - mac:00:51:00:00:01:43 - sid:3 - svc:1000
           |
           +-- 10.1.1.143 - IPCP
-------------------------------------------------------------------------------
Number of active subscribers : 1
Flags: (N) = the host or the managed route is in non-forwarding state
===============================================================================
*A:PE-1# show qos scheduler-hierarchy subscriber "sub-01" detail
===============================================================================
Scheduler Hierarchy - Subscriber sub-01
===============================================================================
--- snip ---
Root (Egr)
| slot(1)
|--(Q) : Sub=sub-01:sla-profile-12 1000->1/1/4:1201.41->6  (Port 1/1/4)
|   |    AdminPIR:1500       AdminCIR:1500
--- snip ---
```

- SPI sharing per session

Note:
Although they can have the same value as in the following output, an SPI sharing ID is not the same as the PPP session ID.

```
A:PE-1# show service active-subscribers detail
===============================================================================
Active Subscribers
===============================================================================
-------------------------------------------------------------------------------
Subscriber sub-01 (sub-profile-12)
-------------------------------------------------------------------------------
---snip---
-------------------------------------------------------------------------------
(1) SLA Profile Instance
- sap:[1/1/4:1201.41] (IES 1000 - group-int-1-1)
- sla:sla-profile-12 PPP session:10
-------------------------------------------------------------------------------
---snip---
A:PE-1# show service active-subscribers hierarchy
===============================================================================
Active Subscribers Hierarchy
```

```
===============================================================================
-- sub-01 (sub-profile-12)
   |
   |-- sap:[1/1/4:1201.41] - sla:sla-profile-12 PPP session:10
   |   |
   |   |-- PPP-session - mac:00:51:00:00:01:41 - sid:10 - svc:1000
   |   |   |
   |   |   +-- 10.1.1.141 - IPCP
   |
   |-- sap:[1/1/4:1201.41] - sla:sla-profile-12 PPP session:11
   |   |
   |   |-- PPP-session - mac:00:51:00:00:01:42 - sid:11 - svc:1000
   |   |   |
   |   |   +-- 10.1.1.142 - IPCP
   |
   +-- sap:[1/1/4:1201.41] - sla:sla-profile-12 PPP session:12
       |
       +-- PPP-session - mac:00:51:00:00:01:43 - sid:12 - svc:1000
           |
           +-- 10.1.1.143 - IPCP
-------------------------------------------------------------------------------
Number of active subscribers : 1
Flags: (N) = the host or the managed route is in non-forwarding state
===============================================================================
A:PE-1# show qos scheduler-hierarchy subscriber "sub-01" detail
===============================================================================
Scheduler Hierarchy - Subscriber sub-01
===============================================================================
--- snip ---
Root (Egr)
| slot(1)
|--(Q) : Sub=sub-01:sla-profile-12:PPP-10 1000->1/1/4:1201.41->6  (Port 1/1/4)
|   |    AdminPIR:1500       AdminCIR:1500
--- snip ---
```

- SPI sharing per-group

```
A:PE-1# show service active-subscribers detail
===============================================================================
Active Subscribers
===============================================================================
-------------------------------------------------------------------------------
Subscriber sub-01 (sub-profile-12)
-------------------------------------------------------------------------------
---snip---
-------------------------------------------------------------------------------
(1) SLA Profile Instance
- sap:[1/1/4:1201.41] (IES 1000 - group-int-1-1)
- sla:sla-profile-12 group:100
-------------------------------------------------------------------------------
---snip---
A:PE-1# show service active-subscribers hierarchy
===============================================================================
Active Subscribers Hierarchy
===============================================================================
-- sub-01 (sub-profile-12)
   |
   |-- sap:[1/1/4:1201.41] - sla:sla-profile-12 group:100
   |   |
   |   |-- PPP-session - mac:00:51:00:00:01:41 - sid:15 - svc:1000
   |   |   |
   |   |   +-- 10.1.1.141 - IPCP
   |   |
   |   |-- PPP-session - mac:00:51:00:00:01:42 - sid:14 - svc:1000
```

```
   |   |   |
   |   |   +-- 10.1.1.142 - IPCP
   |   |
   |   +-- sap:[1/1/4:1201.41] - sla:sla-profile-12 group:200
   |       |
   |       +-- PPP-session - mac:00:51:00:00:01:43 - sid:13 - svc:1000
   |           |
   |           +-- 10.1.1.143 - IPCP
-------------------------------------------------------------------------------
Number of active subscribers : 1
Flags: (N) = the host or the managed route is in non-forwarding state
===============================================================================
A:PE-1#  show qos scheduler-hierarchy subscriber "sub-01" detail
===============================================================================
Scheduler Hierarchy - Subscriber sub-01
===============================================================================
---snip---
Root (Egr)
| slot(1)
|--(Q) : Sub=sub-01:sla-profile-12:Group-100 1000->1/1/4:1201.41->6  (Port 1/1/4)
|   |    AdminPIR:1500       AdminCIR:1500
---snip---
```

In RADIUS accounting messages, the SPI is uniquely defined by the following attributes:

Example:

```
configure
    subscriber-mgmt
        radius-accounting-policy "acct-policy-1"
            include-radius-attribute
                subscriber-id
                nas-port-id
                sla-profile
                spi-sharing
```

Example:

```
NAS PORT ID [87] 13 1/1/4:1201.41
   VSA [26] 40 Nokia(6527)
     SUBSC ID STR [11] 6 sub-01
     SLA PROF STR [13] 14 sla-profile-12

   # SPI sharing per SAP:
   VSA [241.26] 3 NOKIA(6527)
     SPI SHARING_ID [47] 3 SAP
   # SPI sharing per session:
   VSA [241.26] 14 NOKIA(6527)
     SPI SHARING_ID [47] 14 PPP session:12
   # SPI sharing per group:
   VSA [241.26] 14 NOKIA(6527)
     SPI SHARING_ID [47] 9 group:100
```

9.3.7.4.2 SLA-based egress QoS marking

The egress QoS marking for subscriber host traffic is derived from the SAP egress QoS policy associated with a corresponding SAP, instead of from the SLA profile associated with the corresponding subscriber host. Therefore, no egress QoS marking (Dot1p marking is set to 0, the dscp/prec field is kept unchanged)

is performed for traffic transmitted on a managed SAP because by default, sap-egress policy 1 is attached to every managed SAP.

The default value of the "qos-marking-from-sap" flag is enabled. This means that the qos-marking defined in the SAP egress QoS policy associated with the SAP is used. The default setting of this flag in a combination with managed-SAP results in the same behavior as in the current system (dot1p=0, dscp/prec is unchanged).

If the **no qos-marking-from-sap** command is executed, then both the Dot1p marking and DSCP marking are derived from the sla-profile.

Changing the flag setting in the SLA profile being used by any subscriber-hosts (this includes subscriber-hosts on managed-SAPs as well) is allowed.

The following MC traffic characteristics apply:

- On Layer 3 subscriber interfaces, MC is not supported so it is impossible to enable it at the SAP level or at the sla-instance level.
- On Layer 2 SAPs, IGMP snooping is supported while it is not supported on the SLA instance level. Therefore, any MC traffic transmitted at egress belongs to a SAP (meaning it uses SAP queues), instead of to an SLA instance.
- The special case are SAPs with a **profiled-traffic-only flag** enabled. Although it is possible to define an **sla-profile** applicable to a Layer 2 host, this is not taken as reference for marking mc-traffic, but rather SAP settings are used.

9.3.8 Sub-id and brg-id names with lengths between 32 and 64 characters

Beginning with Release 20.2.R1, the length of *sub-id* and *brg-id* names increased from 32 characters to 64 characters. These are referred to as long *sub-id* and long *brg-id*. The length of the corresponding RADIUS attributes, Alc-SubscID-Str and Alc-BRG-ID, that are mapped to the long *sub-id* and long *brg-id* are also increased to 64 characters.

As a result of this length increase, all MIB tables containing *sub-id* and *brg-id* names are affected. In a majority of those tables, the *sub-id* and *brg-id* name length is directly increased from 32 characters to 64 characters. However, tables where the MIB OID key contains a *sub-id* or *brg-id* as one of the fields do not increase the size of the *sub-id* and *brg-id* fields because the maximum key size of 128 characters could be exceeded when the *sub-id* and *brg-id* names are combined to form the key. Because the maximum size of the key in the MIB tables is limited to 128 characters, the *sub-id* and *brg-id* length in such tables remains limited to 32 characters. This ensures that the MIB key does not exceed the maximum size of 128 characters. This also means that an operator-defined *sub-id* and *brg-id* name that is greater than 32 characters must be internally translated (within the SR OS) into a 32-character identification. Instead of truncating *sub-id* and *brg-id* names that are greater than 32 characters to a 32 characters value (which could lead to duplicate *sub-id*s or *brg-id*s), an internal and unique 32-character length *sub-id* and *brg-id* is automatically generated by the system. These internally generated *sub-id* and *brg-id* names are used in the following tables where the long *sub-id* and *brg-id* (>32characters) could lead to violations of the maximum key size (128 characters). The affected tables are:

- TIMETRA-SUBSCRIBER-MGMT-MIB:
  - tmnxSLAProfInstOverridesEntry
  - tmnxSubSpiOvrEntry
  - tmnxSLAProfInstSubHostV2Entry
  - tmnxSubSpiHostEntry

  - tmnxSPICatEntry
  - tmnxSubSpiCatEntry
  - tmnxSpiEgrQosSchedStatsEntry
  - tmnxSPIEgrQosSchedStatsEntry
- TIMETRA-NAT-MIB:
  - tmnxNatL2AwHostPlcyEntry
  - tmnxNatFwlHostEntry
  - tmnxNatFwd2Entry

The operator-defined long *sub-id* and long *brg-id* names are listed in these tables and replaced with the internally-generated version with a 32-character length version.

Table 14: Potential violation of the maximum key size shows examples of *sub-id* and *brg-id* names where a long *sub-id* and long *brg-id* may lead to a violation of the maximum key size. The internally-generated ID begins with the _tmnx_ prefix:

Table 14: Potential violation of the maximum key size

tmnxSubInfoSubIdent	otherKey	Attributes
ABCshort	keyA1	existingInfoA1
ABCshort	keyA2	existingInfoA2
_tmnx_sub_123	keyB1	existingInfoB1
_tmnx_brg_123	keyB2	existingInfoB2
_tmnx_sub_456	keyC	existingInfoC
ghiShort	keyD	existingInfoD

The operator-defined *sub-id*s and *brg-id*s with lengths up to 32 characters are not affected by this change.

In most cases, operators are not concerned with the internal *sub-id* and *brg-id* which are only used by the system to access data in one of the 11 MIB tables where the long *sub-id* or long *brg-id* would otherwise violate the maximum length of the key. Therefore, the internal ID is not shown in the output of any show command.

An exception occurs when a SNMP table walk is performed in one of the 11 tables in which an entry of interest is found that contains an internal *sub-id* and *brg-id*, that needs to be connected with the real (long) subscriber identity.

This conversion can be performed and is aided by the MIB tables tmnxSubShortEntry (Table 15: tmnxSubShortEntry) and tmnxSubBrgShortEntry tables (Table 16: tmnxSubBrgShortEntry):

Table 15: tmnxSubShortEntry

tmnxSubShortId	tmnxSubLongId
ABCshort	ABCshort

tmnxSubShortId	tmnxSubLongId
_tmnx_sub_123	defLongstring
_tmnx_sub_456	JKLLongstring
ghiShort	ghiShort

Table 16: tmnxSubBrgShortEntry

tmnxSubBrgShortId	tmnxSubBrgLongId
MNPshort	ABCshort
_tmnx_brg_333	xyzLongstring
_tmnx_brg_222	OQRLongstring
WVZShort	WVZhort

The mapping from long to an internal ID can be retrieved from the tmnxSubscriberInfoEntry (Table 17: tmnxSubscriberInfoEntry) and tmnxSubBrgEntry (Table 18: tmnxSubBrgEntry) tables:

Table 17: tmnxSubscriberInfoEntry

tmnxSubInfoSubIdent	attributes	tmnxSubInfoShortId
ABCshort	subscrInfoA	ABCshort
JKLlongstring	subscrInfoC	_tmnx_456
defLongstring	subscrInfoB	_tmnx_123
ghiShort	subscrInfoD	ghiShort

Table 18: tmnxSubBrgEntry

tmnxSubBrgId	attributes	tmnxSubBrgIdShort
MNPshort	subscrInfoM	MNPshort
OQRlongstring	subscrInfoO	_tmnx_222
xyzLongstring	subscrInfoX	_tmnx_333
WVZShort	subscrInfoW	WVZShort

9.3.8.1 Online change of sub-id and brg-id

The *sub-id* and *brg-id* strings can be changed online with CLI and CoA/RAR (RADIUS and Diameter interfaces). Conversion between any combination of long and short *sub-id*s and short *brg-id*s is supported by moving each IPoE/PPPoE session or host under a new or renamed subscriber. This is performed by:

- CoA
- tools commands:
  - **tools>perform>subscr-mgmt>edit-ipoe-session subscriber**
  - **tools>perform>subscr-mgmt>edit-lease-state subscriber**
  - **tools>perform>subscr-mgmt>edit-ppp-session subscriber**
  - **tools>perform>subscr-mgmt>edit-slaac-host subscriber**

  The above commands must be run separately for each host or session of the subscriber that is renamed.

9.3.8.2 Usage notes

This section describes the usage of *sub-id* and *brg-id* criteria.

- Long *sub-id*s and long *brg-id*s are automatically enabled without having to explicitly enable them by provisioning additional CLI commands.
- Although, the internal ANCP strings are 64 characters long, some of the external ANCP strings are limited to 63 characters. This is the limitation of the ANCP protocol and some of these strings are:
  - Access-Loop-Circuit-ID TLV
  - Access-Loop-Remote-ID TLV
  - Access-Aggregation-Circuit-ID-ASCII

  Consequently, the maximum size of externally controlled ANCP parameters remains at 63 characters (such as the ANCP protocol, RADIUS, CLI). This means that when ANCP is used, the operators should restrict the *sub-id* string to a maximum of 63 characters, or always provide and explicit ANCP string that is a maximum of 63 characters in length.
- Multicast host-tracking is not supported with long *sub-id*s. This means that the hosts with a *sub-id* longer than 32 characters are excluded from host tracking. The *sub-id* key length for host-tracking related MIB tables remain unchanged at 32 characters and only contains subs with short *sub-id*s.

  The following MIBs are not supported for long *sub-id*s:
  - tmnxSubGrpTrkEntry
  - tmnxSubHostGrpTrkEntry
  - tmnxSubHostSapTrkEntry
  - tmnxSubHostTrkEntry
  - tmnxSubHostTrkStatsEntry
  - tmnxSubTrkPlcySubscriberEntry
  - tmnxSubTrkStatusEntry

  Multicast host tracking only works with short *sub-id*s and is configured as follows:

```
configure
  subscriber-management
      host-tracking-policy <policy-name>
            egress-rate-modify [agg-rate-limit | scheduler <sch-name>]

configure
  subscriber-management
```

```
      sub-profile <subscriber-profile-name>
          host-tracking-policy <policy-name>  => mutually exclusive with igmp-policy
```

However, multicast HQoS adjustment is supported with long *sub-id*s, and should be deployed as a replacement for legacy multicast host tracking. Multicast HQoS adjustment is configured as follows:

```
configure
  subscriber-management
      igmp-policy <policy-name>
              egress-rate-modify [egress-aggregate-rate-limit | scheduler <name>]

configure
  subscriber-management
      sub-profile <subscriber-profile-name>
              igmp-policy <policy-name>
```

- With the introduction of the long *sub-id* and long *brg-id* options, the **show** command output is rearranged by inserting additional line breaks and by wrapping the long *sub-id* and long *brg-id* at the end of the line. For example, the *sub-id* **subidlong.123456789_123456789_123456789 123456789_123456789_3333** is wrapped as follows:

```
===============================================================================
Subscriber                          : subidlong.123456789_123456789_12345678
                                      9 123456789_123456789_3333
===============================================================================
```

- In a multihoming environment, the internal short *sub-id*s and short *brg-id*s are not synchronized. This means that they are independently derived on each node, meaning that if they are needed by the operator, they have to be retrieved by the management system after every switchover.

  In most cases, the operator is not aware of the internal *sub-id* and *brg-id*. Their awareness is required only if an SNMP table walk is performed in one of the 11 MIB tables where the long *sub-id* and long *brg-id* causes the key to exceed the 128 character limit. If an entry with an internal *sub-id* or *brg-id* in one of the tables is found, then these internal values can be used to find the real (long) subscriber identity with a help of the conversion tables (tmnxSubShortEntry and tmnxSubBrgShortEntry).
- Internally generated *sub-id*s and *brg-id*s are not saved in a persistency file. The *sub-id*s and *brg-id*s change after a reboot. Similar to multihoming, if they are needed by the operator, they must be re-retrieved after a chassis reboot even when persistency is enabled.

9.3.9 Auto-sub ID

The subscriber ID name (*sub-id*) is a mandatory object that binds all hosts of a specific subscriber together. Briefly, the *sub-id* name represents a residential household. Many management/troubleshooting and even billing operations rely on the *sub-id* name entity. The *sub-id* name is required for the host creation process, and it can be supplied by any authentication source, such as RADIUS, Diameter, LUDB, or Python. It can be derived from the *sap-id* or can be statically provisioned in the form of a string.

In some ESM deployments, it is desirable that the *sub-id* is automatically generated within the router instead of burdening the OSS with this function. A typical application for auto sub-id is as follows:

- RADIUS server provides the SLA profile string and the sub-profile string but not the sub-id string.
- The *sub-id* name is automatically generated and formatted based on the configured options.

The following are the properties of auto sub-id generation:

The automatic generation of the subscriber ID name can be based on any combination of the following fields:

- MAC address
- sap-id
- circuit-id
- remote-id
- session-id

There can be only a single set of subscriber identification fields defined per host type (IPoE or PPPoE) per chassis. If the combination of the fields must be modified, the existing subscribers with an automatically generated subscriber ID must be manually terminated. Considering that remote termination of the IPoE subscribers by a DHCP server is not supported by all DHCP client vendors through the FORCERENEW DHCP message (RFC 3203, *DHCP reconfigure extension*), changing the subscriber fields while subscribers with automatically generated subscriber ID are active should be avoided.

The subscriber ID name automatic generation takes place at the end of the host initiation process (after the authentication phase is completed) and only in case whereby the subscriber ID had not been already provided by any other more specific means (RADIUS, Diameter, LUDB, or Python).

The format of the *sub-id* name can be either a 10-character encoded string (characters A to Z and 0 to 9) or a user- friendly string based on the subscriber identification fields. The maximum length of the subscriber ID name is 64 characters.

The subscriber ID name is not passed in the Access-Request to the RADIUS server because it is generated after the authentication phase.

The subscriber ID name can be automatically generated regardless of how the SLA or subscriber profile strings are obtained (RADIUS, LUDB, Python, or static).

The subscriber identification fields used in automatic generation of the subscriber ID name are enabled at the system level.

CLI syntax:

```
configure
subscriber-mgmt
    auto-sub-id-key
        ppp-sub-id-key [mac] [sap-id] [circuit-id] [remote-id] [session-id]
        ipoe-sub-id-key [mac] [sap-id] [circuit-id] [remote-id] [dual-stack-remote-id]
```

If no *sub-id-key* per host type is configured, the defaults are:

Table 19: Defaults

PPPoE host type:	mac, sap-id, session-id
IPoE host type:	<mac, sap-id>.

The order in which the fields are configured is important because the subscriber ID name potentially becomes a concatenated string of the subscriber host identifiers in the order in which they are provisioned. The subscriber ID cannot be longer than 64 characters.

- If the length of the concatenated fields for the subscriber ID name is longer than 64 characters, the host creation fails.

- If the circuit ID or remote ID is in the key and they contain non-printable characters, their place in subscriber ID name are formatted in hex instead of ASCII. ASCII printable characters can contain the byte values 0x20 to 0x7E. All other values are ASCII non-printable and therefore, are formatted in hex characters.

The following would generate a subscriber ID name: xx:xx:xx:xx:xx:xx|1/1/3:23|44. The length of such subscriber ID name would be 29B.

- mac: xx:xx:xx:xx:xx:xx
- sap: 1/1/3:23
- session-id: 44 (16bits length)

If the key contains the circuit ID as: 0x610163 (3 bytes), then the subscriber ID name is formatted as 610161, in hex, because 01 hex is non-printable in ASCII. Then the subscriber ID name's length is 6B.

However, if the circuit ID is 0x616263 (3 bytes), then the string is formatted as ASCII string abc (three characters). The subscriber ID name's length is 3B.

The assignment of the subscriber ID to dynamic hosts is performed in the following order:

1. From authentication sources: RADIUS, Diameter, LUDB, or Python.

   A subscriber ID name obtained from authentication sources can conflict with the format of an implicit auto-generated subscriber ID name. When this happens, the subscriber host or session setup fails. Therefore, when implicit subscriber ID name generation is enabled (the default), a 10-character string containing the characters A to Z and 0 to 9 should not be returned from authentication sources. Information in Step 3 describes information to disable the implicit automatic generation of a subscriber ID name.
2. An explicit configured default is configured as *def-sub-id*:
   - At the SAP level for static SAPs:

     ```
     configure
        service ies/vprn
           subscriber-interface <ip-int-name>
              group-interface <ip-int-name>
                 sap <sap-id>
                    sub-sla-mgmt
                       def-sub-id use-sap-id | use-auto-id | string <sub-id>
     ```
   - In the MSAP policy for managed SAPs:

     ```
     configure
        subscriber-mgmt
           msap-policy <name>
              sub-sla-mgmt
                 def-sub-id use-sap-id | use-auto-id | string <sub-id>
     ```

   where:

   - **use-sap-id**: the *sub-id* name is the SAP identifier
   - **use-auto-id**: the *sub-id* name is a combination of the identifiers specified in **auto-sub-id-key**.

     The *sub-id* name is in a readable format, that is, a concatenation of the fields in the **pppoe-sub-id-key** or **ipoe-sub-id-key** command separated by a "|" character.
   - **string**: the *sub-id* name is a user defined string
3. Implicitly generated default:

When no subscriber ID name is provided in authentication, and no explicit default is configured, then the system, by default, automatically generates a subscriber ID name, a 10-character string, using characters A to Z and 0 to 9, that is based on the fields defined in the:

- **ppp-sub-id-key** command for PPP host types. If no such fields are explicitly defined, the default are assumed: mac, sap-id, session-id.
- **ipoe-sub-id-key** command for IPoE host types. If no such fields are explicitly defined, the defaults are assumed: mac, sap-id.

The implicitly generated subscriber ID name is unique per chassis as well as in dual-homed environments.

The implicit automatic subscriber ID name generation can be disabled with the following command: **configure subscr-mgmt auto-sub-id-key no implicit-generation**.

- The implicit *auto-sub-id* name generation cannot be disabled when there are active subscribers in the system with an implicit automatically generated subscriber ID name.
- When disabled, the implicit **auto-sub-id** name generation cannot be enabled when there are active subscribers in the system.

With implicit subscriber ID generation disabled, the subscriber host or session setup fails when no subscriber ID name is provided in authentication, and no explicit default is configured. A 10-character subscriber ID name format, using the characters A to Z and 0 to 9, can be returned from authentication sources without risk of conflict.

Static subscribers are required to have the *sub-id* manually configured.

9.3.9.1 Sub-id identifiers

The sub-id can be based on any combination of the following identifiers:

- The sap-id, in combination with any other allowable identifier, is used as the search key. This assumes a 1:1 (subscriber per SAP) deployment model.
- The circuit-id, in combination with any other allowable identifier, is used to identify subscribers. This can be used in 1:1 deployment model, or in service per SAP deployment model. Circuit-id is applicable to IPoE v4 type hosts (option 82), to IPoE v6 type hosts (option 18 – interface-id) and PPPoE hosts (remote agent option signaled by PPPoE tags). The format of circuit-id is identical for IPv4 and IPv6 hosts.
- The remote-id, in combination with any other allowable identifier, is used to identify subscribers. This can be used in 1:1 deployment model, or in service per SAP deployment model. The remote-id is applicable to IPoE v4 type hosts (option 82), to IPoE v6 type hosts (option 37) and PPPoE hosts (remote agent option signaled by PPPoE tags).
- The **mac** address (in combination with any other allowable identifier is used to identify subscribers. This assumes a 1:1 deployment model.
- The PPPoE session ID, in combination with any other allowable identifier, is applicable only to PPPoE hosts. The session ID used is the first host that is instantiated for the subscriber.

9.3.9.2 Dual-stack hosts

Auto-generation of sub-id names for subscribers with a single dual-stack hosts (IPoE and PPPoE) is enabled by default by not explicitly provisioning anything for the def-sub-id. The sub-id name would be

semi-randomly generated based on the <mac, sap-id, session-id> for PPPoE hosts and the <mac, sap-id> combination for IPoE host.

9.3.9.3 Mixing hosts with auto-generated IDs and non-auto-generated IDs

Hosts with different sub-id names but identical auto-sub-id keys are not linked into the same subscriber. Such scenarios can arise with hosts with the same auto-sub-id keys but different methods for obtaining the sub-id name. For example, one host relying on auto-generated sub-id name while the other is using explicit configuration methods (sap-id, string, RADIUS or LUDB). If the auto-generated sub-id name and explicit sub-id name are the same, the host is tied into the same subscriber.

For example:

The default auto-sub-id for the following two hosts are <mac, sap-id>.

Host X on SAP 1/1/1:1 with MAC 00:00:00:00:00:01 obtains sub-id through RADIUS.

Host Y on SAP 1/1/1:1 with MAC 00:00:00:00:00:01 has sub-id auto-generated.

Regardless of which host comes up first, those two hosts at the end belong to different subscribers if their sub-ids are different.

9.3.9.4 Deployment considerations

The following is a deployment example scenario.

CLI syntax:

```
config
 subscriber-mgmt
    auto-sub-id-key
        ppp-sub-id-key sap-id
        ipoe-sub-id-key mac circuit-id
```

CLI syntax:

```
config
 service vprn 10
     subscriber-interface <ip-int-name>
     authentication-policy <auth-pol-name>
     group-interface <ip-int-name>
        sap 1
            sub-sla-mgmt
                def-sub-id  use-sap-id
                sub-ident-policy <ident-pol-name>
        sap 2
            sub-sla-mgmt
                def-sub-id auto-id
                sub-ident-policy <ident-pol-name>
        sap 3
            sub-sla-mgmt
                def-sub-id "sub3"
                sub-ident-policy <ident-pol-name>
        sap 4
            sub-sla-mgmt
                    sub-ident-policy <ident-pol-name>
```

Assume the following cases:

1. RADIUS returns the sub-id on all four SAPs.
2. RADIUS does not return the sub-id string on any of the SAPs.

In the first case where RADIUS returns the sub-id string, on all four SAPs, the sub-id string is assigned by the RADIUS server. Defaults have no effect, and neither do identifiers specified under the auto-sub-id-key node.

In the second case, the effects are the following:

- On SAP1 the *sub-id* name is the *sap-id* (1/1/1:3)
- On SAP 2 the *sub-id* name is the *sap-id* for PPPoE hosts and <mac>-<circuit-id> concatenation for IPoE type hosts.
- On SAP3 the *sub-id* name is the literal 'sub3' for PPPoE and IPoE hosts.
- On SAP4 the *sub-id* name is a semi-random value based on the *sap-id* for PPPoE hosts and the <mac, circuit-id> combination for IPoE hosts.

9.3.9.5 Restrictions

Only a single combination of the subscriber fields used to auto generate sub-id is allowed per host type (IPoE or PPPoE) and per chassis. In case that the combination of the fields needs to be changed, the existing subscribers with an auto-generated sub-id must be manually terminated. Considering that remote termination of the IPoE subscribers by DHCP server is not supported by all DHCP client vendors through FORCERENEW DHCP message (RFC 3203), changing the subscriber fields while subscribers with auto generated sub-id are active should be avoided.

9.3.10 Limiting subscribers, hosts, and sessions

This section provides an overview of the different configuration options in SR OS to restrict the number of subscribers, subscriber hosts, and subscriber sessions.

- **multi-sub-sap**

  limits the number of subscribers (dynamic and static) on a SAP

- **lease-populate**

  limits the number of dynamic and static hosts on a SAP

- **host-limits**

  limits the number of dynamic and static hosts per type and address family; enforced per SLA profile instance or per subscriber

- session-limits

  limits the number of IPoE, PPPoE and L2TP sessions per SLA profile instance or per subscriber

The setup of a new subscriber host or session fails if any of these limits is reached.

9.3.10.1 Limiting the number of IPoE sessions

The number of IPoE sessions per SAP is limited with the **sap-session-limit** command configured in the **group-interface ipoe-session** context

The number of IPoE sessions per group interface or retail subscriber interface is limited with the **session-limit** command configured in the **group-interface ipoe-session** or retail **subscriber-interface ipoe-session** context.

IPoE sessions and subscriber hosts associated with IPoE sessions are subject to the per SLA profile instance host and session limits configured in the **config**>**subscr-mgmt**>**sla-prof**>**host-limits** context and to the per subscriber host and session limits configured in the **config**>**subscr-mgmt**>**sub-prof** context. See Limiting the number of hosts and sessions per SLA profile instance and per subscriber for a detailed description.

9.3.10.2 Limiting the number of PPPoE sessions

The number of PPPoE sessions per SAP is limited with the **sap-session-limit** command configured in the **group-interface pppoe** context.

To limit the number of PPPoE sessions per group interface or retail subscriber interface use the **session-limit** command configured in the **group-interface pppoe** or retail **subscriber-interface pppoe** context.

PPPoE sessions and subscriber hosts associated with PPPoE sessions are subject to the per SLA profile instance host and session limits configured in the **config**>**subscr-mgmt**>**sla-prof**>**host-limits** context and to the per subscriber host and session limits configured in the **config**>**subscr-mgmt**>**sub-prof** context. See Limiting the number of hosts and sessions per SLA profile instance and per subscriber for a detailed description.

9.3.10.3 Limiting the number of hosts and sessions per SLA profile instance and per subscriber

Table 20: Host limits list the host limits and Table 22: Session limits lists the session limits that can be configured in the following profiles:

- **sla-profile**

  The limits are enforced per SLA profile instance.

- **sub-profile**

  The limits are enforced per subscriber.

Example

For a bridged RGW, allow one dual stack IPoE session (IPv4 and IPv6 IA-PD) per SLA profile instance and up to two sessions per subscriber.

```
configure
    subscriber-mgmt
        sla-profile "sla-profile-1"
            description "host and sessions limits per SLA Profile Instance"
            host-limits
                ipv4-overall 1
                ipv4-arp 0
                ipv4-dhcp 1
                ipv6-pd-overall 1
                ipv6-wan-overall 0
            exit
            session-limits
                ipoe 1
                pppoe-overall 0
                l2tp-overall 0
```

```
            exit
        exit
        sub-profile "sub-profile-1"
            description "host and session limits per subscriber"
            session-limits
                overall 2
```

Table 21: Host limit counters applicable per subscriber host type specifies the host-limits counters that are applicable for each of the different subscriber host types in SR OS. Table 23: Session limit counters applicable per subscriber session type specifies the session-limits counters that are applicable for each of the different subscriber session types in SR OS.

Host and session limits are checked when the host or session is created in the system. When a limit is reached, the host or session setup fails, and an error event is logged. For example:

```
6338 2020/09/24 09:48:50.612 UTC WARNING: DHCP #2005 Base Lease State Population Error

"Lease state table population error on SAP 1/1/4:2111.1 in service 1000 - sub-profile 'sub-
profile-1' : host-limit overall (1) exceeded for subscriber 'ipoe-001'"
```

When a host or session limit is reached for an ARP host, an IPoE host or an IPoE session, a host-limit-exceeded Subscriber Host Connectivity Verification (SHCV) can be triggered to clean up the state of disconnected devices.

Note: If the **remove-oldest** command is configured in the **host-limits** context and an IPv4 ARP host, IPv4 DHCP host, IPv4 host, or subscriber host limit is reached when a new DHCPv4 host or an ARP host connects, the oldest active host disconnects and the new host is granted access. The dynamic host with the least remaining lease time is considered the oldest host. The **remove-oldest** command is not applicable for PPPoE or IPv6 subscriber hosts.

Table 20: Host limits

Command name	Description
overall	Limits the total number of subscriber hosts
ipv4-overall	Limits the total number of IPv4 hosts
ipv4-arp	Limits the number of IPv4 ARP hosts
ipv4-dhcp	Limits the number of IPv4 DHCP hosts
ipv4-ppp	Limits the number of IPv4 PPP hosts
ipv6-overall	Limits the total number of IPv6 hosts
ipv6-pd-overall	Limits the total number of IPv6 DHCP Prefix Delegation hosts (IA-PD)
ipv6-pd-ipoe-dhcp	Limits the number of IPv6 IPoE DHCP Prefix Delegation hosts (IA-PD)
ipv6-pd-ppp-dhcp	Limits the number of IPv6 PPPoE DHCP Prefix Delegation hosts (IA-PD)
ipv6-wan-overall	Limits the total number of IPv6 WAN hosts
ipv6-wan-ipoe-dhcp	Limits the number of IPv6 IPoE DHCP WAN hosts (IA-NA)

Command name	Description
ipv6-wan-ipoe-slaac	Limits the number of IPv6 IPoE SLAAC WAN hosts
ipv6-wan-ppp-dhcp	Limits the number of IPv6 PPPoE DHCP WAN hosts (IA-NA).
ipv6-wan-ppp-slaac	Limits the number of IPv6 PPPoE SLAAC WAN hosts
lac-overall	Limits the total number of L2TP LAC hosts

Table 21: Host limit counters applicable per subscriber host type

Subscriber host type	Counts toward following host limits
IPv4 - PPP Hosts - IPCP	ipv4-ppp, ipv4-overall, overall
IPv4 - PPP Hosts - PFCP	—
IPv4 - IPOE Hosts - DHCP	ipv4-dhcp, ipv4-overall, overall
IPv4 - IPOE Hosts - ARP	ipv4-arp, ipv4-overall, overall
IPv4 - IPOE Hosts - Static	ipv4-overall, overall
IPv4 - IPOE Hosts - PFCP	—
IPv4 - IPOE Mngd Hosts - Data-trig	ipv4-overall, overall
IPv4 - IPOE Mngd Hosts - AAA	ipv4-overall, overall
IPv4 - IPOE Mngd Hosts - GTP	ipv4-overall, overall
IPv4 - IPOE Mngd Hosts - Bonding	ipv4-overall, overall
IPv4 - IPOE Hosts BSM - DHCP	—
IPv4 - IPOE Hosts BSM - Static	—
IPv4 - IPOE BSM - DHCP	—
IPv4 - IPOE BSM - Static	—
IPv6 - PPP Hosts - SLAAC	ipv6-wan-ppp-slaac, ipv6-wan-overall, ipv6-overall, overall
IPv6 - IPOE Hosts - DHCP6 (NA)	ipv6-wan-ppp-dhcp, ipv6-wan-overall, ipv6-overall, overall
IPv6 - PPP Hosts - DHCP6 (PD)	ipv6-pd-ppp-dhcp, ipv6-pd-overall, ipv6-overall, overall
IPv6 - PPP Mngd Routes - DHCP6 (PD)	—
IPv6 - PPP Hosts - PFCP (SLAAC)	—
IPv6 - PPP Hosts - PFCP (NA)	—
IPv6 - PPP Hosts - PFCP (PD)	—

Subscriber host type	Counts toward following host limits
IPv6 - IPOE Hosts - SLAAC	ipv6-wan-ipoe-slaac, ipv6-wan-overall, ipv6-overall, overall
IPv6 - IPOE Hosts - DHCP6 (NA)	ipv6-wan-ipoe-dhcp, ipv6-wan-overall, ipv6-overall, overall
IPv6 - IPOE Hosts - DHCP6 (PD)	ipv6-pd-ipoe-dhcp, ipv6-pd-overall, ipv6-overall, overall
IPv6 - IPOE Mngd Routes - DHCP6 (PD)	—
IPv6 - IPOE Hosts - Static (WAN)	ipv6-wan-overall, ipv6-overall, overall
IPv6 - IPOE Hosts - Static (Pfx)	ipv6-pd-overall, ipv6-overall, overall
IPv6 - IPOE Hosts - PFCP (SLAAC)	—
IPv6 - IPOE Hosts - PFCP (NA)	—
IPv6 - IPOE Hosts - PFCP (PD)	—
IPv6 - IPOE Mngd Hosts - Data-trig (WAN)	ipv6-wan-overall, ipv6-overall, overall
IPv6 - IPOE Mngd Hosts - Data-trig (Pfx)	ipv6-pd-overall, ipv6-overall, overall
IPv6 - IPOE Mngd Routes - Data-trig (Pfx)	—
IPv6 - IPOE Mngd Hosts - AAA	ipv6-wan-overall, ipv6-overall, overall
IPv6 - IPOE Mngd Hosts - GTP (SLAAC)	ipv6-pd-overall, ipv6-overall, overall
IPv6 - IPOE Mngd Hosts - Bonding	ipv6-pd-overall, ipv6-overall, overall
IPv6 - IPOE BSM - DHCP6 (NA)	—
IPv6 - IPOE BSM - DHCP6 (PD)	—
L2TP LAC Hosts	lac-overall, ipv4-overall, ipv6-overall, overall

Table 22: Session limits

Command name	Description
overall	Limits the total number of subscriber sessions
ipoe	Limits the number of IPoE sessions
pppoe-overall	Limits the total number of PPPoE sessions
pppoe-local	Limits the number of PPPoE local terminated sessions (PTA)
pppoe-lac	Limits the number of PPPoE L2TP LAC sessions
l2tp-overall	Limits the total number of L2TP sessions
l2tp-lns	Limits the number of L2TP LNS sessions

Command name	Description
l2tp-lts	Limits the number of L2TP LTS sessions

Table 23: Session limit counters applicable per subscriber session type

Subscriber session type	Counts toward following session limits
Local PPP Sessions - PPPoE	pppoe-local, pppoe-overall, overall
Local PPP Sessions - L2TP (LNS)	l2tp-lns, l2tp-overall, overall
LAC PPP Sessions - PPPoE	pppoe-lac, pppoe-overall, overall
LAC PPP Sessions - L2TP (LTS)	l2tp-lts, l2tp-overall, overall
IPOE Sessions	ipoe, overall
PFCP Sessions - PPP	—
PFCP Sessions - IPOE	—
PFCP Sessions - default tunnels	—

The host and session limits per SLA profile instance and per subscriber can be overridden at subscriber host or session creation by the following.

- Include the 245.26.6527.5 Alc-Spi-Host-And-Session-Limits and 245.26.6527.6 Alc-Sub-Host-And-Session-Limits VSAs in the RADIUS Access-Accept message during authentication.

  See the *7450 ESS, 7750 SR, and VSR RADIUS Attributes Reference Guide* for a detailed description of the VSAs.

- Include the NOKIA-1047 Alc-Spi-Host-And-Session-Limits and NOKIA-1048 Alc-Sub-Host-And-Session-Limits Vendor Specific Diameter AVPs in a CCA-I or CCA-U message during authentication.

  See the *7750 SR and VSR Gx AVPs Reference Guide* for a detailed description of the AVPs.

The combination of overrides and configured limits is only checked when the host or session is created.

Overrides are stored in the subscriber host and session ESM info and can be displayed using the following show commands:

- **show service id** *service-id* **dhcp lease-state detail**
- **show service id** *service-id* **dhcp6 lease-state detail**
- **show service id** *service-id* **slaac host detail**
- **show service id** *service-id* **arp-host detail**
- **show service id** *service-id* **ppp session detail**
- **show service id** *service-id* **pppoe session detail**
- **show service id** *service-id* **ipoe session detail**
- **show service id** *service-id* **managed-hosts type** {**aaa** | **bonding** | **data-triggered** | **gtp** | **wpp**}

For example:

```
# show service id 1000 ipoe session subscriber "ipoe-001" detail
===============================================================================
IPoE sessions for service 1000
===============================================================================
SAP                     : [1/1/4:2111.1]
Mac Address             : 0a:11:00:00:00:01
Circuit-Id              : pe1|1000|group-int-2-1|1/1/4:2111.1
Remote-Id               : 0a:11:00:00:00:01
Session Key             : sap-mac
--- snip ---
Subscriber Session Limit Overrides
 ipoe                   : 3
 pppoe-overall          : 0
 l2tp-overall           : 0
SLA Profile Instance Session Limit Overrides
 ipoe                   : 1
 pppoe-overall          : 0
 l2tp-overall           : 0
-------------------------------------------------------------------------------
Number of sessions : 1
===============================================================================
```

It is the operator's responsibility to keep consistency in the overrides that are stored per subscriber host and session by the following:

- ensure that all hosts and sessions that belong to the same SLA profile instance receive the same dynamic SLA profile instance limit overrides
- ensure that all hosts and sessions that belong to the same subscriber receive the same dynamic subscriber limit overrides

For existing hosts or sessions, this consistency can be achieved by a mid-session change, for example, by RADIUS CoA or Diameter Gx RAR or CCA-U.

Note: If different subscriber hosts or sessions that belong to the same SLA profile instance or subscriber have different override limits, an inconsistent behavior can occur when sessions are recovered from persistency or in case of Multi-Chassis Synchronization (MCS). This may occur because the order in which hosts recover from persistency and the order in which the hosts or sessions are synchronized through MCS, may be different from the order in which sessions were created in the system.

9.3.11 Static subscriber hosts

While it is typically preferred to have all hosts provisioned dynamically through DHCP snooping, it may be needed to provide static access for specific hosts (those that do not support DHCP).

Because a subscriber identification policy is not applicable to static subscriber hosts, the subscriber identification string, subscriber profile and SLA profile must be explicitly defined with the host's IP address and MAC address (if ESM is enabled).

If an SPI associated with the named SLA profile already exists on the SAP for the subscriber, the static subscriber host is placed into that SPI. If an SPI does not yet exist, one is created if possible. If the SLA profile cannot be created, or the host cannot be placed in the existing SPI (the **host-limits** was exceeded), the static host definition fails.

9.3.12 QoS for subscribers and hosts

9.3.12.1 QoS parameters in different profiles

QoS aspects for subscribers and hosts can be defined statically on a SAP or dynamically using.

ESM, for example, in a VLAN-per-service model, different services belonging to a single subscriber are split over different SAPs, and therefore the overall QoS (such as a scheduler policy) of this subscriber must be assigned using Enhanced Subscriber Management.

QoS parameters are shared among the subscriber profile and SLA profile as follows:

- The subscriber profile refers to HQoS ingress and egress scheduler policies which define the overall treatment for hosts of this subscriber when queues are used, or policers managed by HQoS are used at egress. If the subscriber is using policers, the subscriber profile also refers to CFHP ingress and egress policer-control-policies which define the overall treatment for hosts of this subscriber.
- The SLA profile refers to specific queue or policer settings for each host (BTV, VoIP, PC) using SAP ingress and SAP egress QoS policies. The SLA profile can also refer to an egress HQoS scheduler policy which defines the scheduling from the queues of the related host.

The primary use of the subscriber profile is to define the ingress and egress scheduler policies and policer control policies used to govern the aggregate SLA for all hosts associated with a subscriber. To be effective, the queues or policers defined in the SLA profile's QoS policies references a scheduler or arbiter from the scheduler policy or policer-control-policy respectively as their parent.

9.3.12.2 QoS policy overrides

Generic QoS queue or policer parameters can be specified for the SAP in a QoS policy and overridden for some customers by queue and policer parameters defined in the SLA profile. This allows for a single SAP ingress and SAP egress QoS policy to be used for many subscribers, while providing individual subscriber parameters for queue or policer operation.

9.3.13 ESM subscriber hierarchical traffic control

ESM subscribers can make use of both queues and or policers for both the ingress and egress traffic. The queue and policers are configured within SAP ingress and egress policies applied to the SLA profile. The policers (at egress only) and queues can parent to different levels and cir-levels with different weights and cir weights of a virtual scheduler configured within a scheduler policy, and to an egress port scheduler configured in a port scheduler policy, to achieve hierarchical traffic control. The policers can parent to different levels with different weights of an arbiter configured within a policer control policies to achieve hierarchical traffic control.

9.3.13.1 Subscriber HQoS

Hierarchical QoS (HQoS) corresponds to scheduling bandwidth distribution to policers, queues and schedulers and is applied using scheduler policies at ingress and egress of the subscriber profile for a subscriber, and at egress in the SLA profile for a host, together with a port scheduler at both the port and Vport level.

Each scheduler policy can contain up to three tiers of schedulers with lower level schedulers being able to parent to higher level schedulers in the same scheduler policy.

Policers and queues can parent to any scheduler in their related scheduler policy hierarchy (except Vport at egress) and also at the egress to a port scheduler.

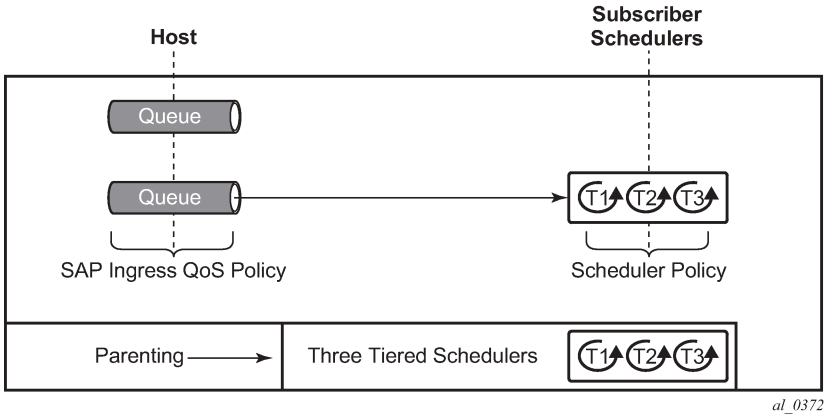
Schedulers can parent to any higher level scheduler in their related scheduler policy hierarchy and, at the egress to a port scheduler configured within the port or Vport. When an egress port scheduler is used, an aggregate rate limit can be applied at the subscriber profile and Vport levels instead of using a scheduler. To extend the hierarchy further at egress, a tier 1 scheduler within a scheduler policy can parent to any scheduler in a scheduler policy at a higher level.

The scheduling levels are composed of:

- ingress and egress queues
- egress policers
- egress SLA-profile schedulers
- ingress and egress subscriber profile schedulers
- egress Vport schedulers
- port schedulers

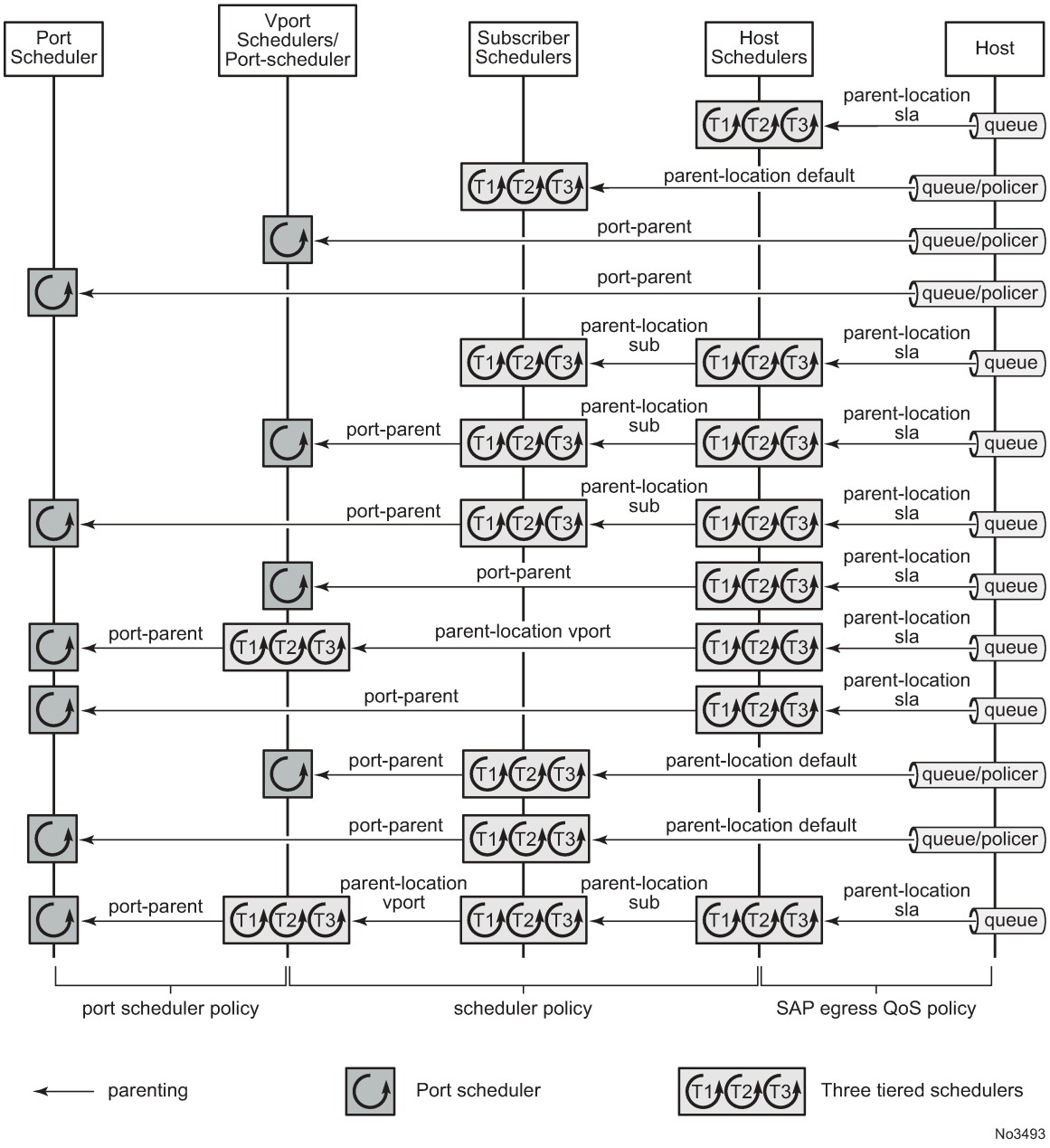
The ingress hierarchical parenting relationship options are shown in Figure 80: Ingress scheduling hierarchy options.

Figure 80: Ingress scheduling hierarchy options



The egress hierarchical parenting relationship options are shown in Figure 81: Egress scheduling hierarchy options. Not all combinations can be configured concurrently, and some uses of port parent could be equally achieved using a scheduler parent and a child parent-location.

Figure 81: Egress scheduling hierarchy options



The **parent** command is used to specify the name of the parent scheduler when parenting a queue or scheduler, together with the level/cir-level and weight/cir-weight at which to connect.

```
config>qos>sap-ingress>queue# parent
  - parent scheduler-name [weight weight] [level level]
                          [cir-weight cir-weight] [cir-level cir-level]

config>qos>sap-egress>queue# parent
- parent scheduler-name [weight weight] [level level]
                          [cir-weight cir-weight] [cir-level cir-level]
```

```
config>qos>scheduler-policy>tier>scheduler# parent
- parent scheduler-name [weight weight] [level level]
                        [cir-weight cir-weight] [cir-level cir-level]
```

The location of the parent scheduler (in which applied scheduler policy it exists) for a policer or queue defaults to a scheduler in the subscriber ingress or egress scheduler policy. Parents of schedulers themselves must be explicitly configured and by default must be within the same scheduler policy.

At egress, the scheduler parenting relationship is determined using the **parent-location** command:

By default, egress queues parent to any scheduler in subscriber egress scheduler policy.

```
config>qos>sap-egress# parent-location default
```

Egress queues can parent to any scheduler within the scheduler policy applied to the egress of an SLA profile (this is not supported for policers managed by HQoS).

```
config>qos>sap-egress# parent-location sla
```

By default, a tier 1 scheduler in the scheduler policy is not allowed to be parented to another scheduler.

```
config>qos>scheduler-policy>tier# parent-location none
```

A tier 1 scheduler in the scheduler policy applied to the egress of an SLA profile can parent to a scheduler applied to the egress of a subscriber profile.

```
config>qos>scheduler-policy>tier# parent-location sub
```

A tier 1 scheduler in the scheduler policy applied to the egress of a subscriber profile can parent to a scheduler applied to the egress of a Vport.

```
config>qos>scheduler-policy>tier# parent-location vport
```

The configuration of a **parent-location** and frame-based accounting in a scheduler policy is mutually exclusive to ensure consistency between the different scheduling levels.

Note that the **parent-location** command is supported only on Ethernet interfaces. It is not supported for ESM over MPLS pseudowires.

Both egress queues and egress schedulers can port parent using directly to different levels/cir-levels, with different weights/cir weights, to a port egress port scheduler. Egress schedulers can also port parent directly to different levels/cir-levels, with different weights/cir weights, to a Vport egress port scheduler.

9.3.13.2 Subscriber CFHP

Class Fair Hierarchical Policing (CFHP) corresponds to the policing control of traffic by policers/arbiters. This uses policer control policies and can be applied for ingress and egress capacity control for the subscriber in the subscriber profile.

Each policer control policy can contain up to three tiers of arbiters with lower level arbiters being able to parent to higher level arbiters in the same scheduler policy.

Policers can parent to any arbiter in their related policer control policy hierarchy.

The policing levels are composed of:

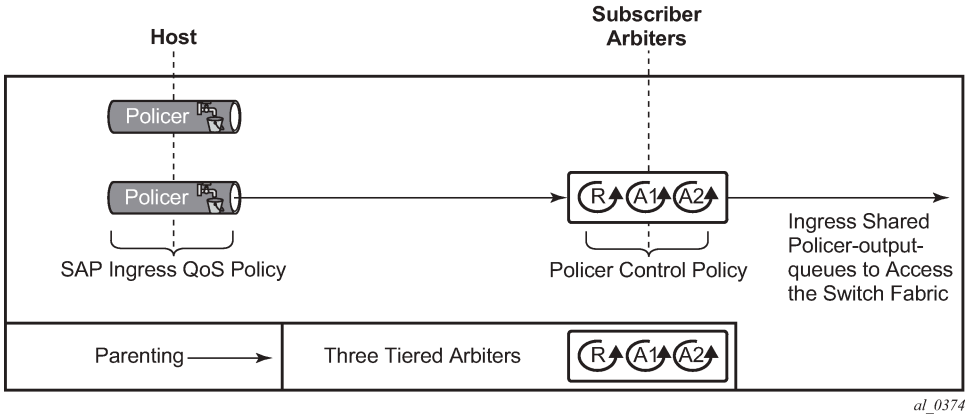
- Ingress and egress policers
- Ingress and egress subscriber arbiters

Note:

- Ingress policed traffic uses the shared policer-output-queues to access the switch fabric. At egress, the policed traffic accesses the egress port through a queue group queue (by default the policer-output-queues queue group, though user configurable queue groups can also be used) or a locally configured subscriber queue.
- Egress policers can also be managed by HQoS.

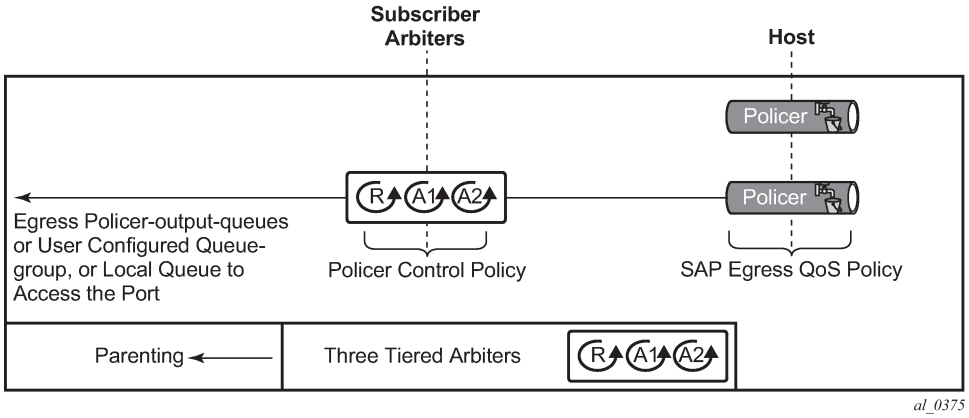
The ingress hierarchical parenting relationship options are shown in Figure 82: Ingress policing hierarchy options.

Figure 82: Ingress policing hierarchy options



The egress hierarchical parenting relationship options are shown in Figure 83: Egress policing hierarchy options.

Figure 83: Egress policing hierarchy options



The **parent** command is used to specify the name of the parent arbiter when parenting a policer or arbiter, together with the level and weight at which to connect.

```
config>qos>sap-ingress>policer$ parent
  - parent arbiter-name [weight weight-level] [level level]

config>qos>sap-egress>policer$ parent
- parent arbiter-name [weight weight-level] [level level]

config>qos>plcr-ctrl-plcy>tier>arbiter# parent
- parent arbiter-name [weight weight-level] [level level]
```

9.3.13.3 ATM/Ethernet last-mile aware QoS for broadband network gateway

This feature allows the user to perform hierarchical scheduling of subscriber host packets in a way that the packet encapsulation overhead and ATM bandwidth expansion (when applicable) because of the last mile for each type of broadband session, that is, PPPoEoA LLC/SNAP and VC-Mux, IPoE, IPoEoA LLC/SNAP and VC-Mux, and so on, is accounted for by the 7450 ESS and 7750 SR acting as the Broadband Network Gateway (BNG).

The intent is that the BNG distributes bandwidth among the subscriber host sessions fairly by accounting for the encapsulation overhead and bandwidth expansion of the last mile so the packets are less likely to be dropped downstream in the DSLAM DSL port.

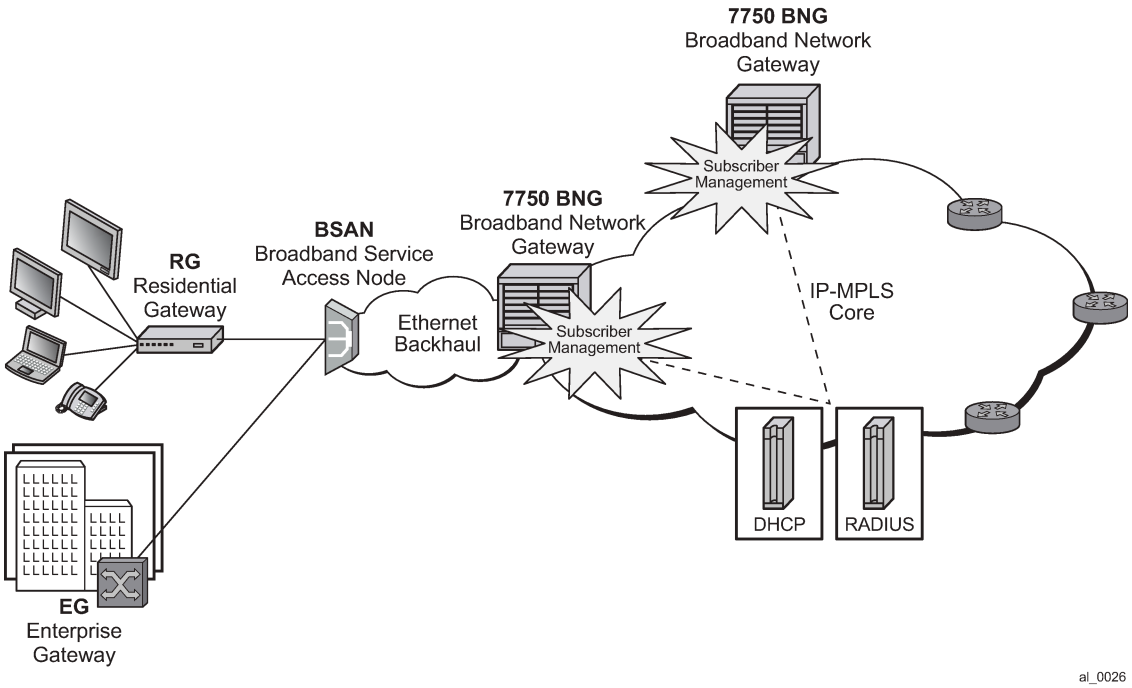
The last mile encapsulation type can be configured by the user or signaled using the Access-loop-encapsulation sub-TLV in the Vendor-Specific PPPoE Tags or DHCP Relay Options as per RFC 4679.

Furthermore, this feature allows the BNG to shape the aggregate rate of each subscriber and the aggregate rate of all subscribers destined for a specific DSLAM to prevent congestion of the DSLAM. The subscriber aggregate rate is adjusted for the last mile overhead. The shaping to the aggregate rate of all subscribers of a specific destination DSLAM is achieved by a new scheduling object, referred to as Virtual Port or Vport in CLI, which represents the DSLAM aggregation node in the BNG scheduling hierarchy

9.3.13.3.1 Broadband network gateway application

An application of this feature in a BNG is shown in Figure 84: BNG application.

Figure 84: BNG application



Residential and business subscribers use PPPoEoA, PPoA, IPoA, or IPoEoA based session over ATM/DSL lines. Each subscriber host can use a different type of session. Although Figure 84: BNG application illustrates ATM/DSL as the subscriber last mile, this feature supports both ATM and Ethernet in the last mile.

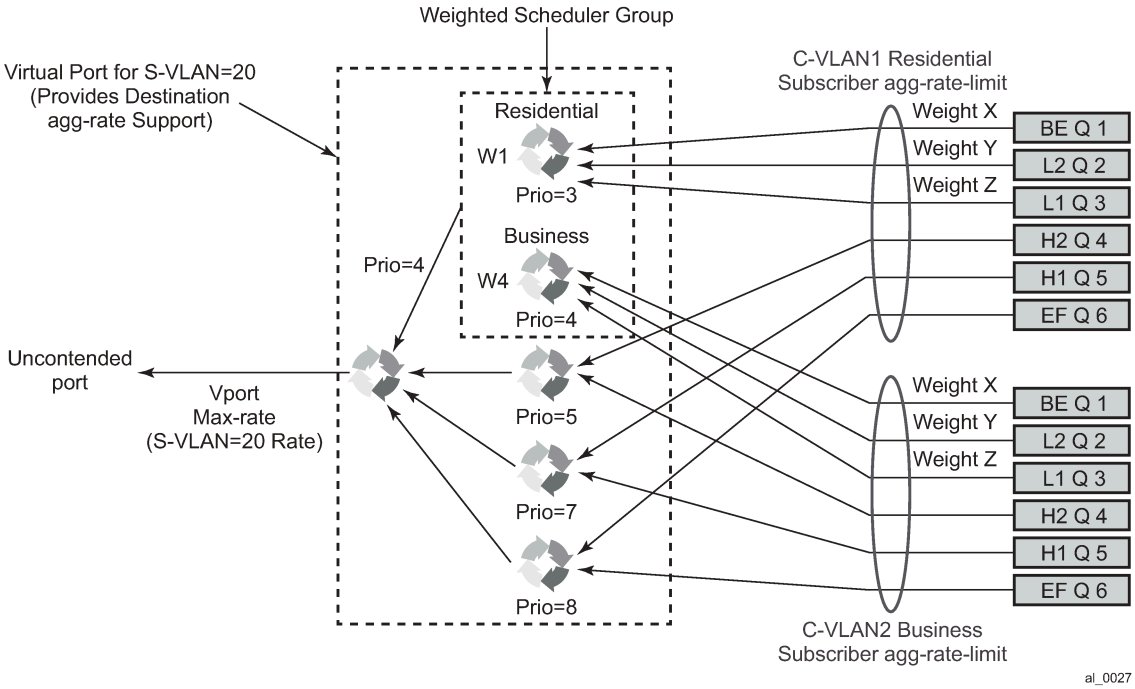
A subscriber SAP is auto-configured through DHCP or the RADIUS authentication process, or is statically configured, and uses a Q-in-Q SAP with the inner C-VLAN identifying the subscriber while the outer S-VLAN identifies the Broadband Service Access Node (BSAN) which services the subscriber, such as, the DSLAM. The SAP configuration is triggered by the first successfully validated subscriber host requesting a session. Within each subscriber SAP, there can be one or more hosts using any of the above session types. The subscriber SAP terminates on an IES or VPRN service on the BNG. It can also terminate on a VPLS instance.

When the 7750 BNG forwards IP packets from the IP-MPLS core network downstream toward the Residential Gateway (RG) or the Enterprise Gateway (EG), it adds the required PPP and Ethernet headers, including the SAP encapsulation with C-VLAN/S-VLAN. When the BSAN node receives the packet, it strips the S-VLAN tag, strips or overwrites the C-VLAN tag, and adds padding to minimum Ethernet size if required. It also adds the LLC/SNAP or VC-mux headers plus the fixed AAL5 trailer and variable AAL5 padding (to next multiple of 48 bytes) and then segments the resulting PDU into ATM cells when the last mile is ATM/DSL. Thus the packet size undergoes a fixed offset because of the encapsulation change and a variable expansion because of the AAL5 padding when applicable. Each type of subscriber host session requires a different amount of fixed offset and may require a per-packet variable expansion depending on the encapsulation used by the session. The BNG node learns the encapsulation type of each subscriber host session by inspecting the Access-loop-encapsulation sub-TLV in the Vendor-Specific PPPoE Tags as specified in RFC 4679. The BNG node must account for this overhead when shaping packets destined for subscriber.

9.3.13.3.2 Queue determination and scheduling

Figure 85: BNG queuing and scheduling model illustrates the queuing and scheduling model for a BNG using the Ethernet or ATM last-mile aware QoS feature.

Figure 85: BNG queuing and scheduling model



A set of per FC queues are applied to each subscriber host context to enforce the packet rate within each FC in the host session as specified in the subscriber's host SLA profile. A packet is stored in the queue corresponding the packet's FC as per the mapping of forwarding class to queue-id defined in the sap-egress QoS policy used by the host SLA profile. In the BNG application however, the host per FC queue packet rate is overridden by the rate provided in the RADIUS access-accept message. This rate represents the ATM rate that is seen on the last mile, that is, it includes the encapsulation offset and the per packet expansion because of ATM segmentation into cells at the BSAN.

To enforce the aggregate rate of each destination BSAN, a scheduling node, referred to as virtual port, and Vport is in the CLI. The Vport operates exactly like a port scheduler with the difference that multiple Vport objects can be configured on the egress context of an Ethernet port. The user adds a Vport to an Ethernet port using the following command:

CLI syntax:

```
config>port>ethernet>access>egress>vport vport-name create
```

The Vport is always configured at the port level even when a port is a member of a LAG. The *vport-name* is local to the port it is applied to but must be the same for all member ports of a LAG. It however does not need to be unique globally on a chassis.

CLI syntax:

```
config>port>ethernet>access>egress>vport vport-name create
```

The user applies a port scheduler policy to a Vport using the following command:

CLI syntax:

```
config>port>ethernet>access>egress>vport>port-scheduler-policy port-scheduler-policy-name
```

A Vport cannot be parented to the port scheduler when it is using a port scheduler policy itself. It is important the user ensures that the sum of the **max-rate** parameter value in the port scheduler policies of all Vport instances on a specific egress Ethernet port does not oversubscribe the port's hardware rate. If it does, the scheduling behavior degenerates to that of the H/W scheduler on that port. A Vport which uses an **agg-rate** can be parented to a port scheduler. This is described in Applying aggregate rate limit to a Vport. Note that the application of the **agg-rate**, **port-scheduler-policy** and **scheduler-policy** commands under a Vport configuration are mutually exclusive.

Each subscriber host queue is port parented to the Vport which corresponds to the destination BSAN using the existing **port-parent** command:

CLI syntax:

```
config>qos>sap-egress>queue>port-parent [weight weight] [level level] [cir-weight cir-weight]
 [cir-level cir-level]
```

This command can parent the queue to either a port or to a Vport. These operations are mutually exclusive in CLI as described above. When parenting to a Vport, the parent Vport for a subscriber host queue is not explicitly indicated in the above command. It is determined indirectly. The determination of the parent Vport for a specified subscriber host queue is described in Vport determination and evaluation.

Furthermore, the weight (cir-weight) of a queue is normalized to the sum of the weights (cir-weights) of all active subscriber host queues port-parented at the same priority level of the Vport or the port scheduler policy. Because packets of ESM subscriber host queues are sprayed among the link of a LAG port based on the subscriber-id, it is required that all subscribers host queues mapping to the same Vport, such as having the same destination BSAN, be on the same LAG link so that the aggregate rate toward the BSAN is enforced. The only way of achieving this is to operate the LAG port in active/standby mode with a single active link and a single standby link.

The aggregate rate of each subscriber must also be enforced. The user achieves this by applying the existing **agg-rate-limit** command to the egress context of the subscriber profile:

CLI syntax:

```
config>subscr-mgmt>sub-profile>egress>agg-rate-limit agg-rate
```

In the BNG application however, this rate is overridden by the rate provided in the RADIUS access-accept message. This rate represents the ATM rate that is seen on the last mile, that is, it includes the encapsulation offset and the per packet expansion because of ATM segmentation into cells at the BSAN.

9.3.13.3.3 Weighted scheduler group

The existing port scheduler policy defines a set of eight priority levels with no ability of grouping levels within a single priority. To allow for the application of a scheduling weight to groups of subscriber host queues competing at the same priority level of the port scheduler policy applied to the Vport, or to the Ethernet port, a new group object is defined under the port scheduler policy:

CLI syntax:

```
config>qos>port-scheduler-policy>group group-name rate pir-rate [cir cir-rate]
```

Up to eight groups can be defined within each port scheduler policy. One or more levels can map to the same group. A group has a rate and optionally a cir-rate and inherits the highest scheduling priority of its member levels. For example, the scheduler group shown in the Vport consists of level priority 3 and level priority 4. It therefore inherits priority 4 when competing for bandwidth with the standalone priority levels 8, 7, and 5.

In essence, a group receives bandwidth from the port or from the Vport and distributes it within the member levels of the group according to the weight of each level within the group. Each priority level competes for bandwidth within the group based on its weight under congestion situation. If there is no congestion, a priority level can achieve up to its rate (cir-rate) worth of bandwidth.

The mapping of a level to a group is performed as follows:

CLI syntax:

```
config>qos>port-scheduler-policy>level priority-level rate pir-rate [cir cir-rate] group group-name [weight weight-in-group]
```

Note: CLI enforces that mapping of levels to a group are contiguous. In other words, a user would not be able to add priority level to group unless the resulting set of priority levels is contiguous.

When a level is not explicitly mapped to any group, it maps directly to the root of the port scheduler at its own priority like in existing behavior.

9.3.13.3.4 Queue and subscriber aggregate rate configuration and adjustment

Software-Based Implementation

The subscriber aggregate rate is adjusted and based on an average frame size.

The user enables the use of this adjustment method by configuring the following option in the egress context of the subscriber profile:

CLI syntax:

```
config>subscr-mgmt>sub-profile>egress>encap-offset [type type]
```

This command allows the user to configure a default value to be used by all hosts of the subscriber in the absence of a valid signaled value. The following are configurable values:

pppoa-llc, pppoa-null, pppoeoa-llc, pppoeoa-llc-fcs, pppoeoa-llc-tagged, pppoeoa-llc-tagged-fcs, pppoeoa-null, pppoeoa-null-fcs, pppoeoa-null-tagged, pppoeoa-null-tagged-fcs, ipoa-llc, ipoa-null, ipoeoa-llc, ipoeoa-llc-fcs, ipoeoa-llc-tagged, ipoeoa-llc-tagged-fcs, ipoeoa-null, ipoeoa-null-fcs, ipoeoa-null-tagged, ipoeoa-null-tagged-fcs, pppoe, pppoe-tagged, ipoe, ipoe-tagged

Otherwise, the fixed packet offset is derived from the encapsulation type value signaled in the Access-loop-encapsulation sub-TLV in the Vendor-Specific PPPoE Tags as described in Section Signaling of Last Mile Encapsulation Type. Only signaling using PPPoE Tags is supported in the software based implementation. The last signaled valid value is then applied to all active hosts of this subscriber. If no value is signaled in the subscriber host session or the value in the fields of the Access-loop-encapsulation sub-TLV are invalid,

then the offset applied to the aggregate rate of this subscriber uses the last valid value signaled by a host of this subscriber if it exists, or the user entered default type value if configured, or no offset is applied.

Configure the average frame size value to be used for this adjustment:

CLI syntax:

```
config>subscr-mgmt>sub-profile>egress>avg-frame-size bytes
```

The entered value must include the FCS but not the Inter-Frame Gap (IFG) or the preamble. If the user does not explicitly configure a value for the **avg-frame-size** parameter, then it is also assumed the offset is zero regardless of the signaled or user-configured value.

The computation of the subscriber aggregate rate consists of taking the average frame size, adding the encapsulation fixed offset including the AAL5 trailer, and then adding the variable offset consisting of the AAL5 padding to next multiple of 48 bytes. The AverageFrameExpansionRatio is then derived as follows:

AverageFrameExpansionRatio = (53/48 x (AverageFrameSize + FixedEncapOffset + AAL5Padding)) / (AverageFrameSize + IFG + Preamble).

When the last mile is Ethernet, the formula simplifies to:

AverageFrameExpansionRatio = (AverageFrameSize + FixedEncapOffset + IFG + Preamble) / (AverageFrameSize + IFG + Preamble).

The following are the frame size and rate applied to the subscriber queue and scheduler:

Subscriber Host Queue (no change):

Size = ImmediateEgressEncap + Data

Rate = ImmediateEgressEncap + Data

Subscriber Aggregate Rate Scheduler:

Size = ImmediateEgressEncap + Data

Rate = sub-agg-rate / AverageFrameExpansionRatio

Note that the CPM applies the AverageFrameExpansionRatio adjustment to the various components used in the determination of the net subscriber operational aggregate rate. It then pushes these adjusted components to IOM which then makes the calculation of the net subscriber operational aggregate rate.

The formula used by the IOM for this determination is:

sub-oper-agg-rate = min(sub-policy-agg-rate/*AverageFrameExpansionRatio*, ancp_rate/*AverageFrameExpansionRatio*) + (igmp_rate_delta/*AverageFrameExpansionRatio*),

where *sub-policy-agg-rate* is either the value configured in the **agg-rate-limit** parameter in the subscriber profile or the resulting RADIUS override value. In both cases, the CPM uses an internal override to download the adjusted value to IOM.

The value of *sub-oper-agg-rate* is stored in the IOM's subscriber table.

The following are the procedures for handling signaling changes or configuration changes affecting the subscriber profile:

1. If a new RADIUS update comes in for the aggregate subscriber rate, then a new subscriber aggregate ATM adjusted rate is computed by CPM using the last configured **avg-frame-size** and then programmed to IOM.
2. If the user changes the value of the **avg-frame-size** parameter, enables/disables the **encap-offset** option, or changes the parameter value of the **encap-offset** option, the CPM immediately triggers a

re-evaluation of subscribers using the corresponding subscriber profile and an update the IOM with the new subscriber aggregate rate.

3. If the user changes the value of the **agg-rate-limit** parameter in a subscriber profile which has the **avg-frame-size** configured, this immediately triggers a re-evaluation of subscribers using the corresponding subscriber profile. An update to the subscriber aggregate rate is performed for those subscribers whose rate has not been previously overridden by RADIUS.
4. If the user changes the **type** value of the **encap-offset** command, this immediately triggers a re-evaluation of subscribers using the corresponding subscriber profile. An update to the subscriber aggregate rate is performed for those subscribers who are currently using the default value.
5. If two hosts of the same subscriber signal two different encapsulation types, the last one signaled gets used at the next opportunity to re-evaluate the subscriber profile.
6. If a subscriber has a DHCP host, a static host or an ARP host, the subscriber aggregate rate continues to use the user-configured default encapsulation type value or the last valid encapsulation value signaled in the PPPoE tags by other hosts of the same subscriber. If none was signaled or configured, then no rate adjustment is applied.

Hardware-Based Implementation — The datapath computes the adjusted frame size real-time for each serviced packet from a queue by adding the actual packet size to the fixed offset provided by CPM for this queue and variable AAL5 padding.

Like in the software based implementation, the user enables the use of the fixed offset and per packet variable expansion by configuring the following option in the egress context of the subscriber profile:

CLI syntax:

```
config>subscr-mgmt>sub-profile>egress>encap-offset [type type]
```

When this command is enabled, the fixed packet offset is derived from the encapsulation type value signaled in the Access-loop-encapsulation sub-TLV in the Vendor-Specific PPPoE Tags or DHCP Relay Options as described in Section Signaling of Last Mile Encapsulation Type.

If the user specifies an encapsulation type with the command, this value is used as the default value for all hosts of this subscriber until a host session signaled a valid value. The signaled value is applied to this host only and the remaining hosts of this subscriber continue to use the user entered default type value if configured, or no offset is applied. Hosts of the same subscriber using the same SLA profile and which are on the same SAP share the same instance of FC queues. In this case, the last valid encapsulation value signaled by a host of that same instance of the SAP egress QoS policy overrides any previous signaled or configured value.

The procedures for handling signaling changes or configuration changes affecting the subscriber profile are the same as in the software-based implementation with except for the following:

1. The **avg-frame-size** parameter in the subscriber profile is ignored.
2. If the user specifies an encapsulation type with the command, this value is used as the default value for all hosts of this subscriber until a host session signaled a valid value. The signaled value is applied to this host and other hosts of the same subscriber sharing the same SLA profile and which are on the same SAP. The remaining hosts of this subscriber continue to use the user entered default type value if configured, or no offset is applied.
3. If the user enables/disables the **encap-offset** option, or changes the parameter value of the **encap-offset** option, the CPM immediately triggers a re-evaluation of subscriber hosts using the corresponding subscriber profile and an update the IOM with the new fixed offset value.

4. If subscriber host session signals an encapsulation type at the session establishment time and subsequently sends a DHCP renewal message using a Layer 2 DHCP relay which does not insert option82 in a unicast message, the encapsulation type for this host does not change. TR-101 states that option82 is mandatory for DHCP broadcast messages).
5. If a subscriber has a static host or an ARP host, the subscriber host continues to use the user-configured default encapsulation type value or the last valid encapsulation value signaled in the PPPoE tags or DHCP relay options by other hosts of the same subscriber which use the same SLA profile instance. If none was signaled or configured, then no rate adjustment is applied.
6. The encapsulation type value signaled in DHCP relay options or PPPoE tags are not cross-checked against the host type. Thus, a host signaling PPPoA/LLC encapsulation type through DHCP relay options are not handled as if the packet included a PPPoE header when forwarded over the local Ethernet port. This results in applying an encap-offset in the datapath which assumes the PPPoE header is added to forwarded packets over the local Ethernet port.

The **encap-offset** option forces all the rates to be either last-mile frame over the wire or local port frame over the wire, described as **LM-FoW** and **FoW** respectively. The system maintains a running average frame expansion ratio for each queue to convert queue rates between these two formats as described in Frame size, rates, and running average frame expansion ratio. The following are details of the queue and scheduler operation:

1. When the **encap-offset** option is configured in the subscriber profile, the subscriber host queue rates, that is, CLI and operational PIR and CIR as well as queue bucket updates, the queue statistics, that is, forwarded, dropped, and HQoS offered counters use the **LM-FoW** format. The scheduler policy CLI and operational rates also use **LM-FoW** format. The port scheduler **max-rate** and the priority level rates and weights, if a Weighted Scheduler Group is used, are always entered in CLI and interpreted as **FoW** rates. The same is true for an **agg-rate-limit** applied to a Vport. Finally the subscriber **agg-rate-limit** is entered in CLI as **LM-FoW** rate. When converting between **LM-FoW** and **FoW** rates, the queue running average frame expansion ratio value is used.
   - If the user enabled **frame-based-accounting** in a scheduler policy or **queue-frame-based-accounting** with subscriber **agg-rate-limit** and a port scheduler policy, the queue operational rate is capped to a user configured **FoW** rate. The scheduler policy operational rates are also in the **FoW** format. A user-configured queue **avg-frame-overhead** value is ignored because the running average frame expansion ratio is what is used when the **encap-offset** option is enabled.
   - If the user configured queue **packet-byte-offset** value, it is ignored and is not accounted for in the net packet offset calculation.
2. When **no encap-offset** is configured in the subscriber profile, that is, default and pre-R9.0 behavior, queue CLI and operational PIR and CIR rates, as well as queue bucket updates, the queue statistics, use data format. The scheduler policy CLI and operational rates also use data format. The port scheduler **max-rate** and the priority level rates and weights, if a Weighted Scheduler Group is used, and the subscriber **agg-rate-limit** are entered in CLI and interpreted as **FoW** rates. When converting between **FoW** and data rates, the queue **avg-frame-overhead** value is used and because this an Ethernet port, it is not user-configurable but constant and is equal to +20 bytes (IFG and preamble).
   - If the user enabled **frame-based-accounting** in a scheduler policy or **queue-frame-based-accounting** with subscriber **agg-rate-limit** and a port scheduler policy, the queue operational rate is capped to a user configured FoW rate in CLI which is then converted into a data rate using the queue **avg-frame-overhead** constant value of +20 bytes. The scheduler policy operational rates are in the **FoW** format.
   - If the user configured queue **packet-byte-offset** value, it adjusts the immediate packet size. This means that the queue rates, that is, operational PIR and CIR, and queue bucket updates use the adjusted packet size. In addition, the queue statistics are also reflected the adjusted packet size.

Scheduler policy rates, which are data rates, use the adjusted packet size. The port scheduler **max-rate** and the priority level rates and weights, if a Weighted Scheduler Group is used, as well as the subscriber **agg-rate-limit** are always **FoW** rates and uses the actual frame size. Packet byte offset settings are not included in the applied rate when (queue) frame based accounting is configured, therefore, the offsets are applied to the statistics.

9.3.13.3.5 Frame size, rates, and running average frame expansion ratio

The following are the details of the rates and frame sizes applied to the subscriber host queues, the subscriber aggregate rate, and the Vport root scheduler for the scheduling model and when the **encap-offset** option is enabled in the subscriber profile.

Subscriber Host Queue:

Size = LastMileFrameOverWireEncap + Data

Rate = (48/53)* x (LastMileFrameOverWireEncap + Data)

*Applicable to ATM last-mile only.

Subscriber Aggregate Rate:

Size = LastMileFrameOverWireEncap + Data

Rate = (48/53)* x (LastMileFrameOverWireEncap + Data)

*Applicable to ATM last-mile only.

Vport/Port Scheduler and Weighted Scheduler Group

Size = FrameOverWireEncap + Data

Rate = FrameOverWireEncap + Data

When a frame arrives at the queue, the size is *ImmediateEgressEncap+Data*. This size is stored as the *OfferedFrameSize* so that the queue offered stats used in HQoS calculations are correct. See the HQoS-offered statistics as Offered.

This size is then adjusted by removing the *ImmediateEgressEncap* and adding the *LastMileFrameOverWireEncap*. This new adjusted frame size, referred as *LastMileOfferedFrameSize*, is then used for checking compliance of the frame against the queue PIR and CIR bucket sizes and for updating the queue forwarded and dropped stats.

The *LastMileOfferedFrameSize* value is computed dynamically for each packet serviced by the queue.

A new HQoS stat counter *OfferedLastMileAdjusted* is maintained for the purpose of calculating the running average frame expansion ratio, which is the ratio of the accumulated *OfferedLastMileAdjusted* and Offered of each queue:

RunningAverageFrameExpansionRatio = OfferedLastMileAdjusted / Offered

The **vport**/**port** *port-scheduler* hands out its **FoW** bandwidth in terms of Fair Information Rate (FIR) bandwidth to each subscriber queue. This queue FIR must be converted into **LM-FoW** format to cap it by the queue PIR (*adminPIR*) and to make sure the sum of *FIR*s of all queues of the same subscriber does not exceed the subscriber **agg-rate-limit** which is also expressed in **LM-FoW** format. The conversion between these two rates makes use of the cumulative *RunningAverageFrameExpansionRatio* value.

A queue **LM-FoW** AdminPIR value is always capped to the value of the local port **FoW** rate even if the conversion based on the current *RunningAverageFrameExpansionRatio* value indicates that a higher AdminPIR may be able to fill in the full line rate of the local port.

9.3.13.3.6 Vport determination and evaluation

In the BNG application, host queues of all subscribers destined for the same downstream BSAN, for example, all SAPs on the egress port matching the same S-VLAN tag value, are parented to the same Vport which matches the destination ID of the BSAN.

The BNG determines the parent Vport of a subscriber host queue, which has the **port-parent** option enabled, by matching the destination string associated with the subscriber with the string defined under a Vport on the port associated with the subscriber.

The user configures the dest string match under the egress Vport context of the Ethernet port associated with the subscriber:

CLI syntax:

```
config>port>ethernet>access>egress>vport>host-match dest string create
```

If a specific subscriber host queue does not have the **port-parent** option enabled, it is foster-parented to the Vport used by this subscriber and which is based on matching the **dest** string. If the subscriber could not be matched with a Vport on the egress port, the host queue is not bandwidth controlled and competes for bandwidth directly based on its own PIR and CIR parameters.

By default, a subscriber host queue with the **port-parent** option enabled is scheduled within the context of the port's port scheduler policy. To indicate the option to schedule the queue in the context of a port scheduler policy associated with a Vport, the user enters the following command in SLA profile used by the subscriber host:

CLI syntax:

```
config>subscr-mgmt>sla-profile>egress>qos sap-egress-qos-policy-id vport-scheduler
```

This command is persistent meaning that the user can re-enter the **qos** node without specifying the **vport-scheduler** argument each time and the system remembers it. The user can revert to the default setting without deleting the association of the SLA profile with the SAP egress QoS policy by explicitly re-entering the command with the following new argument:

CLI syntax:

```
config>subscr-mgmt>sla-profile>egress>qos sap-egress-qos-policy-id port-scheduler
```

9.3.13.3.7 Applying aggregate rate limit to a Vport

The user can apply an aggregate rate limit to the Vport and apply a port scheduler policy to the port.

This model allows the user to oversubscribe the Ethernet port. The application of the **agg-rate** option is mutually exclusive to the application of a port scheduler policy, or a scheduler policy to a Vport.

When using this model, a subscriber host queue with the **port-parent** option enabled is scheduled within the context of the port's port scheduler policy. However, the user must still indicate to the system that the queues are managed by the aggregate rate limit instance of a Vport by enabling the **vport-scheduler** option in the subscriber host SLA profile:

CLI syntax:

```
config>subscr-mgmt>sla-profile>egress>qos sap-egress-qos-policy-id vport-scheduler
```

A subscriber host-queue which is port-parented is parented to the port scheduler policy of the port used by the subscriber and aggregate rate limited within the instance of the Vport used by this subscriber and which is based on matching the **dest** string and **org** string.

If the specified subscriber host queue does not have the port-parent option enabled, it is foster-parented to the port used by this subscriber and aggregate rate limited within the instance of the Vport used by this subscriber. If the Vport exists but the port does not have a port scheduler policy applied, then the host queue is orphaned and no aggregate rate limit can be enforced.

9.3.13.3.8 Applying a scheduler policy to a Vport

The user can apply a scheduler policy to the Vport. This allows scheduling control of subscriber tier 1 schedulers in a scheduler policy applied to the egress of a subscriber or SLA profile, or to a PW SAP in an IES or VPRN service.

The advantage of using a scheduler policy under a Vport, compared to the use of a port scheduler (with or without an **agg-rate**), is that it allows a port parent to be configured at the Vport level.

Bandwidth distribution from an egress port scheduler to a Vport configured with a scheduler policy can be performed based on the level/cir-level and weight/cir-weight configured under the scheduler's port parent. The result is in allowing multiple Vports, for example representing different DSLAMs, to share the port bandwidth capacity in a flexible way that is under the control of the user.

The configuration of a scheduler policy under a Vport is mutually exclusive to the configuration of a port scheduler policy or an aggregate rate limit.

A scheduler policy is configured under a Vport as follows:

CLI syntax:

```
config>port>ethernet>access>egress>vport# scheduler-policy scheduler-policy-name
```

When using this model, a tier 1 scheduler in a scheduling policy applied to a subscriber profile or SLA profiles must be configured as follows:

CLI syntax:

```
config>qos>scheduler-policy>tier# parent-location vport
```

If the Vport exists, but port does not have a scheduler policy applied, then its schedulers are orphaned and no port level QOS control can be enforced.

The following **show**/**monitor**/**clear** commands are available related to the Vport scheduler:

```
show qos scheduler-hierarchy port port-id vport name [scheduler scheduler-name]
[detail]

show qos scheduler-stats port port-id vport name [scheduler scheduler-name] [detail]

monitor qos scheduler-stats port port-id vport name [interval seconds ] [repeat
repeat] [absolute | rate]

clear qos scheduler-stats port port-id vport name [scheduler scheduler-name] [detail]
```

HQoS adjustment and host tracking are not supported on schedulers that are configured in a scheduler policy on a Vport, so the configuration of a scheduler policy under a Vport is mutually exclusive to the configuration of the **egress-rate-modify** parameter.

ESM over MPLS pseudowires are not supported when a scheduler policy is configured on a Vport.

9.3.13.3.9 Signaling of last mile encapsulation type

A subscriber host session can signal one of many encapsulation types each with a different fixed offset in the last mile. These encapsulation types are described in RFC 4679 and are illustrated in Figure 86: Subscriber host session encapsulation types and Figure 87: Access-loop-encapsulation sub-TLV. The BNG node learns the encapsulation type of each subscriber host session by inspecting the Access-loop-encapsulation sub-TLV in the Vendor-Specific PPPoE Tags as specified in RFC 4679.When Ethernet is the last mile, the encapsulation type results in a fixed offset for all packet sizes. When ATM/DSL is the last mile, there is an additional expansion because of AAL5 padding to next multiple of 48 bytes and which varies depending on the packet size.

The software and hardware based implementations support both ATM and Ethernet access using PPP encapsulation options. Thus, both provide support for the Access-loop-encapsulation sub-TLV in the Vendor-Specific PPPoEv4/PPPoEv6 Tags with the ATM encapsulation values and Ethernet encapsulation values. ATM and Ethernet access using IP encapsulation are only supported using default encapsulation offset configuration in the subscriber profile in the software based implementation. Support for signaling the Access-loop-encapsulation sub-TLV in the DHCPv4/DHCPv6 Relay Options is included in the hardware based implementation. There is no support for DHCPv6 relay options.

Figure 86: Subscriber host session encapsulation types

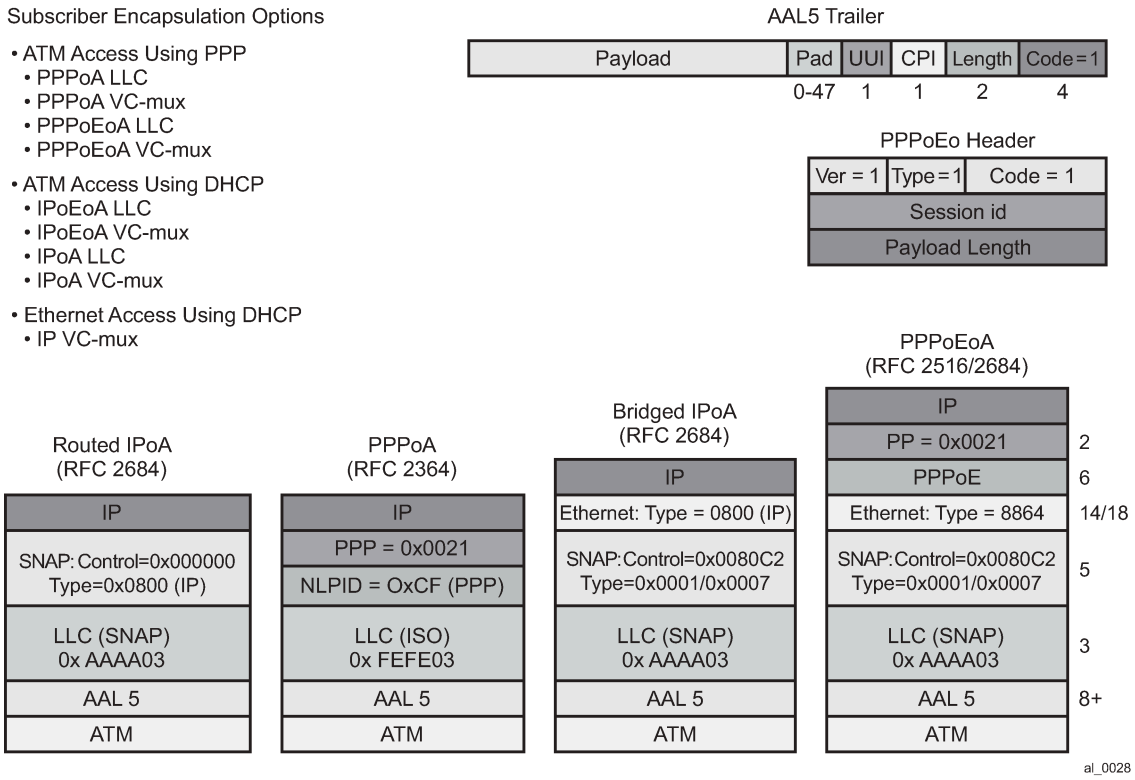


Figure 87: Access-loop-encapsulation sub-TLV

Encapsulation combinations (RFC 4679)

```
                Access-loop-encapsulation
0                   1                   2
      0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     |   Data Link   |    Encaps 1   |    Encaps 2   |
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

  Valid values for the sub-fields are as follows:

     Data Link

        0x00 AAL5
        0x01 Ethernet

     Encaps 1

        0x00 NA - Not Available
        0x01 Untagged Ethernet
        0x02 Single-Tagged Ethernet

     Encaps 2

        0x00 NA - Not Available
        0x01 PPPoA LLC
        0x02 PPPoA Null
        0x03 IPoA LLC
        0x04 IPoA Null
        0x05 Ethernet over AAL5 LLC with FCS
        0x06 Ethernet over AAL5 LLC without FCS
        0x07 Ethernet over AAL5 Null with FCS
        0x08 Ethernet over AAL5 Null without FCS
```

Encapsulation combinations

- AAL5
  - PPPoA LLC/Null
  - IPoA LLC/Null
  - Ethernet over ATM x 4
    - Tagged/Untagged PPP
    - Tagged/Untagged DHCP
  - Total of 20 AAL5combinations
- Ethernet
  - Tagged/Untagged PPP
  - Tagged/Untagged DHCP
  - Total of four Ethernet combinations
- *Total of 24 access combinations*

The operational last-mile values for hosts on the same SAP, having the same SLA profile are displayed in following the show command:

CLI syntax:

```
show>service active-subscribers>ale-adjust
```

The data-link can have values: **atm**, **other** and, **unknown**. If no offset is supplied it is set to **unknown**. **other** is used when the data-link is non-atm, otherwise it states **atm.**

Operational per-queue values can also be found in the show command:

CLI syntax:

```
show>qos>scheduler-hierarchy
```

The following is an example of displaying whether the queue is operating in last-mile mode.

Last mile ATM:

```
*A:Dut-C# /show service active-subscribers ale-adjust
===============================================================================
Active Subscriber Access Loop Encapsulation adjustment
===============================================================================
Subscriber
   SAP                                          SLA profile
   Data-link Offset(bytes)
-------------------------------------------------------------------------------
hpolSub81
```

```
   1/1/11:2000.1                               hpolSlaProf1
   atm       -10
-------------------------------------------------------------------------------
No. of Access Loop Encapsulation adjustments: 1
===============================================================================

*A:Dut-C# show qos scheduler-hierarchy subscriber "hpolSub81"
===============================================================================
Scheduler Hierarchy - Subscriber hpolSub81
===============================================================================
Ingress Scheduler Policy:
Egress Scheduler Policy :
-------------------------------------------------------------------------------
Root (Ing)
|
No Active Members Found on slot 1


Root (Egr)
| slot(1)
|--(Q) : Sub=hpolSub81:hpolSlaProf1 2000->1/1/11:2000.1->8->ATM  (Port 1/1/11)
|
|--(Q) : Sub=hpolSub81:hpolSlaProf1 2000->1/1/11:2000.1->7->ATM  (Port 1/1/11)
|
|--(Q) : Sub=hpolSub81:hpolSlaProf1 2000->1/1/11:2000.1->6->ATM  (Port 1/1/11)
|
|--(Q) : Sub=hpolSub81:hpolSlaProf1 2000->1/1/11:2000.1->5->ATM  (Port 1/1/11)
|
|--(Q) : Sub=hpolSub81:hpolSlaProf1 2000->1/1/11:2000.1->4->ATM  (Port 1/1/11)
|
|--(Q) : Sub=hpolSub81:hpolSlaProf1 2000->1/1/11:2000.1->3->ATM  (Port 1/1/11)
|
|--(Q) : Sub=hpolSub81:hpolSlaProf1 2000->1/1/11:2000.1->2->ATM  (Port 1/1/11)
|
|--(Q) : Sub=hpolSub81:hpolSlaProf1 2000->1/1/11:2000.1->1->ATM  (Port 1/1/11)
|
```

Last mile Ethernet:

```
*A:Dut-C# show service active-subscribers ale-adjust
===============================================================================
Active Subscriber Access Loop Encapsulation adjustment
===============================================================================
Subscriber
   SAP                                         SLA profile
   Data-link Offset(bytes)
-------------------------------------------------------------------------------
hpolSub81
   1/1/11:2000.1                               hpolSlaProf1
   other     +12
-------------------------------------------------------------------------------
No. of Access Loop Encapsulation adjustments: 1
===============================================================================

*A:Dut-C# show qos scheduler-hierarchy subscriber "hpolSub81"
===============================================================================
Scheduler Hierarchy - Subscriber hpolSub81
===============================================================================
Ingress Scheduler Policy:
Egress Scheduler Policy :
-------------------------------------------------------------------------------
Root (Ing)
```

```
|
No Active Members Found on slot 1


Root (Egr)
| slot(1)
|--(Q) : Sub=hpolSub81:hpolSlaProf1 2000->1/1/11:2000.1->8->Eth  (Port 1/1/11)
|
|--(Q) : Sub=hpolSub81:hpolSlaProf1 2000->1/1/11:2000.1->7->Eth  (Port 1/1/11)
|
|--(Q) : Sub=hpolSub81:hpolSlaProf1 2000->1/1/11:2000.1->6->Eth  (Port 1/1/11)
|
|--(Q) : Sub=hpolSub81:hpolSlaProf1 2000->1/1/11:2000.1->5->Eth  (Port 1/1/11)
|
|--(Q) : Sub=hpolSub81:hpolSlaProf1 2000->1/1/11:2000.1->4->Eth  (Port 1/1/11)
|
|--(Q) : Sub=hpolSub81:hpolSlaProf1 2000->1/1/11:2000.1->3->Eth  (Port 1/1/11)
|
|--(Q) : Sub=hpolSub81:hpolSlaProf1 2000->1/1/11:2000.1->2->Eth  (Port 1/1/11)
|
|--(Q) : Sub=hpolSub81:hpolSlaProf1 2000->1/1/11:2000.1->1->Eth  (Port 1/1/11)
|
```

9.3.13.3.10 Configuration example

The following CLI configuration achieves the specific use case shown in Figure 85: BNG queuing and scheduling model.

```
config
    qos
       port-scheduler-policy "dslam-vport-scheduler"
        group res-bus-be create
            rate 1000
        level 3 rate 1000 group res-bus-be weight w1
        level 4 rate 1000 group res-bus-be weight w4
        level 5 rate 1000 cir-rate 100
        level 7 rate 5000 cir-rate 5000
        level 8 rate 500 cir-rate 500
        max-rate 5000

       sap-egress 100                // residential policy
            queue 1                 // be-res
                port-parent weight x level 3
            queue 2                 // l2-res
                port-parent weight y level 3
            queue 3                 // l1-res
                port-parent weight z level 3
            queue 4                 // h2-res
                port-parent level 5
            queue 5                 // h1-res
                port-parent level 7
            queue 6                 // ef-res
                port-parent level 8
            fc be queue 1
            fc l2 queue 2
            fc l1 queue 3
            fc h2 queue 4
            fc h1 queue 5
            fc ef queue 6
        exit
        sap-egress 200               // business policy
```

```
            queue 1                 // be-bus
                     port-parent weight x level 4
            queue 2                 // l2-bus
                   port-parent weight y level 4
            queue 3                 // l1-bus
                      port-parent weight z level 4
            queue 4                 // h2-bus
                   port-parent level 5
            queue 5                 // h1-bus
                  port-parent level 7
            queue 6                 // ef-bus
                   port-parent level 8
            fc be queue 1
            fc l2 queue 2
            fc l1 queue 3
            fc h2 queue 4
            fc h1 queue 5
            fc ef queue 6
        exit
    exit

config
    sub-mgmt
        sla-profile "residential"
            egress
                qos 100 vport-scheduler
            exit
        exit
        sla-profile "business"
            egress
                qos 200 vport-scheduler
            exit
        exit
        sub-profile "residential"
            egress
               encap-offset
               avg-frame-size 1500
               agg-rate-limit 100
               exit
            exit
        exit
        sub-profile "business"
            egress
               encap-offset type pppoeoa-llc-tagged-fcs
               avg-frame-size 500
               agg-rate-limit 200
               exit
            exit
        exit
    exit

config
    port 1/1/1
        ethernet
            access
                egress
                    vport "dslam-1" create
                      port-scheduler-policy "dslam-vport-scheduler"
                        host-match dest “20” create
                        exit
                    exit
                exit
            exit
        exit
```

```
    exit
exit
```

9.3.14 Subscriber volume statistics

Subscriber volume statistics or octet and packet counters are available through the queues and policers that are instantiated for the subscriber. The queue and policer configuration is defined in the SLA profile using ingress and egress QoS policy associations with optional overrides. By default, subscriber hosts that belong to the same subscriber, that are active on the same SAP, and that have the same SLA profile share the set of queues and policers defined by that SPI. Alternatively, for bridged Residential Gateway scenarios, an SPI can be instantiated per subscriber session or per group identifier obtained during authentication. See SLA profile instance sharing for more details.

9.3.14.1 IP (Layer 3) volume accounting

Subscriber volume statistics by default count Layer 2 frame sizes optionally modified by configuration such as packet-byte-offset, last mile aware shaping, and so on.

To report subscriber volume statistics as Layer 3 (IP) packet sizes, the volume-stats-type can be configured to **ip** in the subscriber profile:

```
configure
    subscriber-mgmt
        sub-profile <subscriber-profile-name>
            volume-stats-type ip
```

volume-stats-type ip affects the subscriber statistics in SNMP, CLI, RADIUS accounting, XML accounting and Diameter Gx usage monitoring. Volume quota for RADIUS or Diameter Credit Control applications are interpreted as Layer 3 quota.

The following restrictions apply for **volume-stats-type ip**:

- Layer 3/IP accounting is not supported in combination with MLPPP
- Layer 3/IP accounting in combination with ESMoPW and last-mile-aware shaping may be inaccurate if the MPLS encapsulation overhead changes during the lifetime of a subscriber.
- Layer 3/IP accounting is restricted to a single encap per sla-profile instance (queue instance). The first host associated with the sla-profile instance (queue instance) determines the allowed encapsulation. Conflicting encapsulations are:
  - PPPoE and IPoE on regular Ethernet SAPs
  - PPPoE and IPoE on PW SAPs
- PPPoE keep alive packets do not contain IP payload and introduce an error in Layer 3/IP accounting when enabled in combination with L2TP-LAC. A workaround is to isolate the keep alives in a separate queue or policer.
- Padding of frames smaller than the Ethernet minimum frame size (64B) may introduce an inaccuracy in Layer 3/IP accounting.
- With ATM in the last mile, last-mile-aware shaping may introduce an inaccuracy in Layer 3/IP accounting.

- Packet-Byte-Offset (PBO) changes during the lifetime of a subscriber introduces an inaccuracy in Layer 3/IP accounting.

9.3.14.2 Separate IPv4 and IPv6 counters

IPv4 and IPv6 forwarded and dropped subscriber traffic can be counted separately by a **stat-mode v4-v6** command that is configured as a policer or queue qos override in the sla-profile. The **stat-mode v4-v6** command is only applicable for Enhanced Subscriber Management (ESM).

```
configure subscriber-mgmt
        sla-profile "sla-profile-1" create
            ingress
                qos 10
                    queue 1
                        stat-mode v4-v6
                    exit
                    policer 1
                        stat-mode v4-v6
                    exit
                exit
            exit
            egress
                qos 10
                    queue 1
                        stat-mode v4-v6
                    exit
                    policer 1
                        stat-mode v4-v6
                    exit
                exit
            exit
        exit
```

For policers, the **stat-mode** command overrides the policer stat-mode configuration as defined in the sap-ingress or sap-egress qos policy. For information about sap-ingress and sap-egress policer stat-mode, see the *7450 ESS, 7750 SR, 7950 XRS, and VSR Quality of Service Guide*. For a policer in stat-mode v4-v6, following counters are available:

- Offered IPv4 octets and packets
- Offered IPv6 octets and packets
- Dropped IPv4 octets and packets
- Dropped IPv6 octets and packets
- Forwarded IPv4 octets and packets
- Forwarded IPv6 octets and packets

When a policer's stat-mode is changed while the SLA profile is in use, any previous counter values are lost and any new counters are set to zero.

For queues, a stat-mode is only available for use in Enhanced Subscriber Management (ESM) context to enable separate IPv4/IPv6 counters. For a queue in stat-mode v4-v6, following counters are available:

- Offered High Priority, Low Priority, Uncolored, Managed octets and packets
- Dropped IPv4 octets and packets
- Dropped IPv6 octets and packets

- Forwarded IPv4 octets and packets
- Forwarded IPv6 octets and packets

A queue's stat-mode cannot be changed while the SLA profile is in use.

There are no in-profile or out-of-profile forwarded and dropped counters for policers and queues in **stat-mode v4-v6**.

Non-IP traffic (for example PPPoE LCP frames) is counted against the IPv4 counters.

The separate IPv4 and IPv6 forwarded and dropped counters are reported in

- SNMP
- CLI

```
show service active-subscribers detail
- - - snip - - -
-------------------------------------------------------------------------------
SLA Profile Instance statistics
-------------------------------------------------------------------------------
                         Packets                 Octets

Off. HiPrio           : 0                       0
Off. LowPrio          : 1102685                 1102685000
Off. Uncolor          : 0                       0
Off. Managed          : 0                       0

Queueing Stats (Ingress QoS Policy 10)
Dro. HiPrio           : 0                       0
Dro. LowPrio          : 0                       0
For. InProf           : 0                       0
For. OutProf          : 0                       0
Dro. V4               : 0                       0
Dro. V6               : 0                       0
For. V4               : 367543                  367543000
For. V6               : 735142                  735142000

Queueing Stats (Egress QoS Policy 10)
Dro. InProf           : 0                       0
Dro. OutProf          : 0                       0
For. InProf           : 0                       0
For. OutProf          : 0                       0
Dro. V4               : 0                       0
Dro. V6               : 0                       0
For. V4               : 367543                  367543000
For. V6               : 735088                  735088000

-------------------------------------------------------------------------------
SLA Profile Instance per Queue statistics
-------------------------------------------------------------------------------
                         Packets                 Octets

Ingress Queue 1 (Unicast) (Priority) (Stats mode: v4-v6)
Off. HiPrio           : 0                       0
Off. LowPrio          : 1102685                 1102685000
Dro. V4               : 0                       0
Dro. V6               : 0                       0
For. V4               : 367545                  367545000
For. V6               : 735146                  735146000

Egress Queue 1 (Stats mode: v4-v6)
Dro. V4               : 0                       0
Dro. V6               : 0                       0
```

```
For. V4               : 367547                  367547000
For. V6               : 735096                  735096000

-------------------------------------------------------------------------
SLA Profile Instance per Policer statistics
-------------------------------------------------------------------------
                        Packets                 Octets

Ingress Policer 1 (Stats mode: v4-v6)
Off. V4               : 0                       0
Off. V6               : 0                       0
Dro. V4               : 0                       0
Dro. V6               : 0                       0
For. V4               : 0                       0
For. V6               : 0                       0

Egress Policer 1 (Stats mode: v4-v6)
Off. V4               : 0                       0
Off. V6               : 0                       0
Dro. V4               : 0                       0
Dro. V6               : 0                       0
For. V4               : 0                       0
For. V6               : 0                       0
```

- **RADIUS accounting**

  When a queue or policer is configured in stat-mode v4-v6, existing VSA's are re-used in RADIUS detailed per queue or per policer accounting (**configure subscriber-mgmt radius-accounting-policy** *name* **include-radius-attribute detailed-acct-attributes**):

  – in-profile counter VSA's map to IPv4 octets/packets

  – ingress queue high priority dropped counter VSA's map to IPv4 octets/packets

  – out-of-profile counter VSA's map to IPv6 octets/packets

  – ingress queue low priority dropped counter VSA's map to IPv6 octets/packets

  In addition the [26-6527-107] Alc-Acct-I-statmode / [26-6527-127] Alc-Acct-O-statmode is sent with value set to "v4-v6".

  Optionally a set of VSAs can be included in RADIUS accounting to report the aggregate IPv6 forwarded octets and packets of queues and policers with stat-mode v4-v6 enabled (**configure subscriber-mgmt radius-accounting-policy** *name* **include-radius-attribute detailed-acct-attributes v6-aggregate-stats**):

  – [26-6527-194] Alc-IPv6-Acct-Input-Packets

  – [26-6527-195] Alc-IPv6-Acct-Input-Octets

  – [26-6527-196] Alc-IPv6-Acct-Input-GigaWords

  – [26-6527-197] Alc-IPv6-Acct-Output-Packets

  – [26-6527-198] Alc-IPv6-Acct-Output-Octets

  – [26-6527-199] Alc-IPv6-Acct-Output-Gigawords

  See the *7450 ESS, 7750 SR, and VSR RADIUS Attributes Reference Guide* for a detailed description of all counter attributes.

- **XML accounting**

The complete-subscriber-ingress-egress and custom-record-subscriber XML records use following fields to represent IPv4 and IPv6 forwarded/dropped octets and packets for queues or policers with **stat-mode v4-v6** enabled:

- v4po - IPv4PktsOffered (policer only)
- v4oo - IPv4OctetsOffered (policer only)
- v6po - IPv6PktsOffered (policer only)
- v6oo - IPv6OctetsOffered (policer only)
- v4pf - IPv4PktsForwarded
- v6pf - IPv6PktsForwarded
- v4pd - IPv4PktsDropped
- v6pd - IPv4PktsDropped
- v4of - IPv4OctetsForwarded
- v6of - IPv6OctetsForwarded
- v4od - IPv4OctetsDropped
- v6od - IPv4OctetsDropped

For custom records, the following CLI is re-used to include v4/v6 counters if the queue is configured in **stat-mode v4-v6**:

i-counters

- all-packets-offered-count # n/a
- all-octets-offered-count # n/a
- high-packets-offered-count # n/a
- low-packets-offered-count # n/a
- uncoloured-packets-offered-count # n/a
- high-octets-offered-count # n/a
- low-octets-offered-count # n/a
- uncoloured-octets-offered-count # n/a
- all-packets-offered-count # n/a
- all-octets-offered-count # n/a
- high-packets-discarded-count # IPv4
- low-packets-discarded-count # IPv6
- high-octets-discarded-count # IPv4
- low-octets-discarded-count # IPv6
- in-profile-packets-forwarded-count # IPv4
- out-profile-packets-forwarded-count # IPv6
- in-profile-octets-forwarded-count # IPv4
- out-profile-octets-forwarded-count # IPv6

e-counters

- in-profile-packets-forwarded-count # IPv4
- in-profile-packets-discarded-count # IPv4
- out-profile-packets-forwarded-count # IPv6
- out-profile-packets-discarded-count # IPv6
- in-profile-octets-forwarded-count # IPv4
- in-profile-octets-discarded-count # IPv4
- out-profile-octets-forwarded-count # IPv6
- out-profile-octets-discarded-count # IPv6

9.3.15 Configuring IP and IPv6 filter policies for subscriber hosts

Access Control Lists (ACLs) for subscriber traffic are defined as IP and IPv6 filter policies and are configured in the SLA-profile associated with the subscriber. For information about IP and IPv6 filter policy configurations, see the *7450 ESS, 7750 SR, 7950 XRS, and VSR Router Configuration Guide*.

```
config>subscr-mgmt>sla-prof
        sla-profile sla-profile-1 create
            ingress
                ip-filter 100
                ipv6-filter 300
            exit
            egress
                ip-filter 200
                ipv6-filter 400
            exit
       exit
```

Traffic from different subscriber hosts or sessions of a single subscriber and associated with the same sla-profile instance, is subject to the filter policies defined in the SLA profile.

Changing the IPv4 filter policy in an SLA profile in use by an active subscriber is allowed in the CLI, but not recommended. Changing the IPv6 filter policy in an SLA profile in use by an active subscriber is prevented in the CLI.

9.3.15.1 Dynamic updates of subscriber filter policies

The IP or IPv6 filter policy configuration of subscriber hosts can be dynamically updated using the mechanisms described in the next sections.

See the *7750 SR and VSR RADIUS Attributes Reference Guide* for a detailed description of the RADIUS attributes format.

See the *7750 SR and VSR Gx AVPs Reference Guide* for a detailed description of the Diameter AVP's format.

9.3.15.1.1 SLA profile change

Changing the SLA profile of a subscriber host or session, implicitly changes its associated IP and IPv6 filter policies. An SLA profile change can be done by, for example, a RADIUS CoA or Diameter Gx RAR

message. As the SLA profile also defines the QoS configuration for the subscriber hosts, this change may result in a discontinuity in accounting.

9.3.15.1.2 Override the IP and IPv6 filter policies

The ingress and egress IP and IPv6 filter policies can be overridden per subscriber host or session at creation time or mid-session:

- from RADIUS by including the [26.6527.134] Alc-Subscriber-Filter attribute or the [245.26.6527.7.x] Alc-Subscriber-Filter-Name sub-attributes in an Access-Accept or CoA message
- from Diameter Gx by including the Charging-Rule-Name AVP with the corresponding predefined name in a CCA or RAR message

Note:

- Irrelevant fields (for example, IPv4 filters for an IPv6 host) are ignored.
- If the ingress or egress field is missing in the VSA in a RADIUS CoA message, there is no change for that direction.
- If the ingress or egress field is missing in the VSA in a RADIUS Access-Accept message, the IP filters as specified in the SLA profile are active for that direction.
- An SLA profile IP filter override is applicable to all dynamic host types, including L2TP LNS but excluding L2TP LAC.
- Filter name and filter ID overrides must not be mixed during the lifetime of a subscriber host or session. A filter override specified as a filter name and installed with the Alc-Subscriber-Filter-Name VSA in RADIUS takes precedence over a filter override specified as a filter ID using the Alc-Subscriber-Filter VSA in RADIUS or the Charging-Rule-Name AVP in Diameter Gx. For example, a CoA with Alc-Subscriber-Filter cannot override a filter that was previously installed as an override specified as a filter name with Alc-Subscriber-Filter-Name.

9.3.15.1.3 Insert subscriber host-specific filter entries

A subscriber host specific entry is a filter entry where the match criteria is automatically extended with the subscriber host IP or IPv6 address as source (ingress) or destination (egress) IP. They represent a per host customization of a generic filter policy: only traffic to or from the subscriber host that match against these entries.

A subscriber host specific entry is dynamically created from

- a RADIUS Access-Accept or CoA message containing the [92] NAS-Filter-Rule or [26.6527.159] Alc-Ascend-Data-Filter-Host-Spec attribute
- a Diameter CCA or RAR message containing the [92] NAS-Filter-Rule AVP embedded in the [3GPP-1005] Charging-Rule-Name AVP

The format used to specify host specific filter entries ([92] NAS-Filter-Rule format or [26.6527.159] Alc-Ascend-Data-Filter-Host-Spec format) cannot change during the lifetime of the subscriber host. A RADIUS message can only contain a single format for host specific filter entries. US message can only contain a single format for host specific filter entries.

Up to 10 host-specific filter rules can be specified in a single RADIUS or Diameter message. Each new RADIUS CoA or Diameter CCA/RAR message containing host specific filter attributes overwrites the

previous subscriber host-specific filter entries for that host if there are enough free entries in the reserved range.

Subscriber host-specific filter entries can be removed with a [92] NAS-Filter-Rule attribute value equal to 0x00 or " "(a space).

When the subscriber host session terminates or is disconnected, then the corresponding subscriber host-specific filter entries are also deleted.

Note that subscriber host-specific filter entries are moved if the subscriber host filter policy is changed (new SLA profile or IP filter policy override) and the new filter policy contains enough free reserved entries (**sub-insert-radius**).

A range of entries must be reserved for subscriber host specific entries in a filter policy:

```
config>filter
        ip-filter 100 create
            sub-insert-radius start-entry 1000 count 100
```

High and low watermarks can be configured to raise an event when the thresholds of free entries in the reserved range are reached:

```
 *A:cses-nokia>config>filter>ip-filter$ sub-insert-wmark ?
    - no sub-insert-wmark
    - sub-insert-wmark low <low-watermark> high <high-watermark>

   <low-watermark>      : [0..100]
   <high-watermark>     : [0..100]
```

9.3.15.1.4 Insert shared filter entries

The target application for shared filter entries is operators that have a predefined limited number of different filter lists that each are shared with multiple subscriber hosts or sessions and that are to be managed and activated from RADIUS or Diameter at authentication.

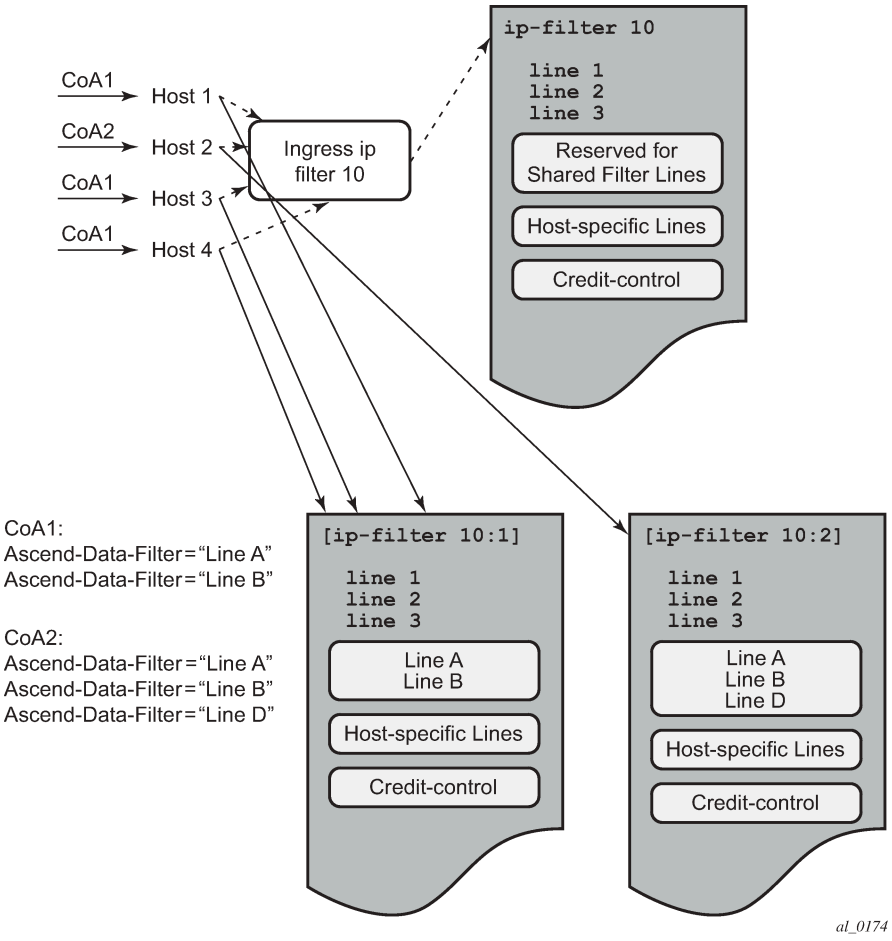
A local configured IP or IPv6 filter associated with a host or session (sla-profile or ip filter override) can be enhanced with dynamic filter entries that can be shared with multiple subscriber hosts or sessions. The shared dynamic filter entries are inserted with:

- a set of RADIUS attributes ([26.529.242] Ascend-Data-Filter or [26.6527.158] Alc-Nas-Filter-Rule-Shared) received in a RADIUS Access-Accept or CoA message. A CoA message containing a set of one of those attributes overrides the previous set of shared filter entries active for that subscriber host or session.
- a set of [6527-158] Alc-Nas-Filter-Rule-Shared AVP's embedded in the [3GPP-1005] Charging-Rule-Name AVP received in a Diameter CCA or RAR message. The last received set of attributes overrides the previous set of shared filter entries for that subscriber host or session.

For each unique set of dynamic filter entries received per type (IPv4 or IPv6) and direction (ingress or egress), a copy is made of the local filter with the dynamic entries included at a preconfigured insert point. If the same set of dynamic filter entries is sent to subscriber hosts or sessions that have the same associated local filter, then they share the same filter copy. When there are no more subscriber hosts associated with a filter copy, then the filter copy is deleted. A filter copy is identified as local filter id:number. For example: show filter ip 10:2.

Shared filter entries are moved if the subscriber host filter policy is changed (new SLA profile or ip filter policy override) and if the new filter policy contains enough free reserved entries.

Figure 88: Insert shared filters



A range of entries must be reserved for shared entries in a filter policy:

```
config>filter>ip-filter
   sub-insert-shared-radius start-entry 100 count 10
```

High and low watermarks can be configured to raise an event when the thresholds of dynamic filter copies are reached:

```
*A:cses-V22>config>filter>ip-filter# shared-radius-filter-wmark ?
  - no shared-radius-filter-wmark
  - shared-radius-filter-wmark low <low-watermark> high <high-watermark>
 <low-watermark>      : [0..7999]
 <high-watermark>     : [1..8000]
```

The format used to specify shared filter entries ([26.6527.158] Alc-Nas-Filter-Rule-Shared format or [26.529.242] Ascend-Data-Filter format) cannot change during the lifetime of the subscriber host or session. A RADIUS message can only contain a single format for shared filter entries.

Shared filter entries can be removed with [26.6527.158] Alc-Nas-Filter-Rule-Shared attribute value equal to 0x00 or " " (a space).

9.3.15.2 Checking filter policy details

Use following show commands to check filter policy details and the filter configuration for a subscriber host:

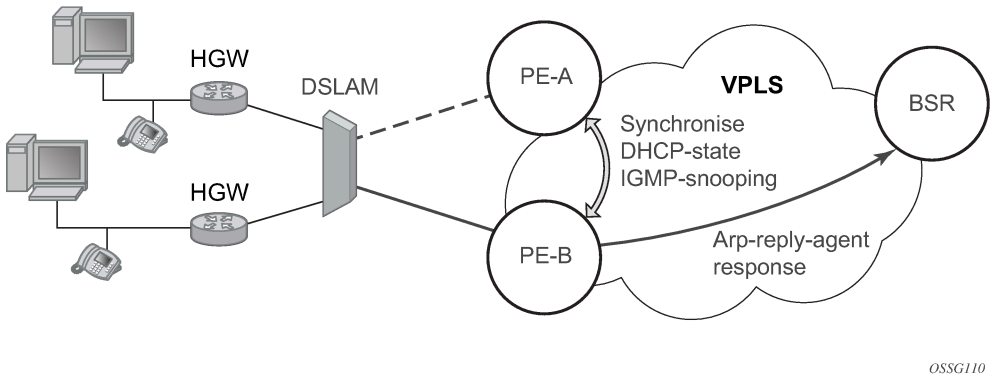
CLI syntax:

```
   show filter ip ip-filter-id detail
   show filter ipv6 ip-filter-id detail
   show filter ip ip-filter-id type entry-type
   show filter ipv6 ipv6-filter-id type entry-type
entry-type : fixed | radius-insert | credit-control-insert | radius-shared
   show service active-subscribers filter [subscriber sub-ident-string] [origin origin]
sub-ident-string : [64 chars max]
origin : radius | credit-control”
```

9.3.16 Multi-chassis synchronization

Figure 89: Dual-homing configuration shows the configuration under which synchronization of subscriber management information is performed. As depicted, a single access node aggregating several subscriber lines is dual- homed to redundant-pair of nodes.

Figure 89: Dual-homing configuration



Enabling subscriber management features (whether basic subscriber-management (BSM) or enhanced subscriber management (ESM)) causes the node to create and maintain state information related to a specific subscriber-host. This information is synchronized between redundant-pair nodes to secure non-stop service delivery in case of the switchover.

9.3.16.1 Overview

The synchronization process provides the means to manage distributed database (the Multi-Chassis Synchronization (MCS) database), which contains the dynamic state information created on any of the nodes by any application using its services. The individual entries in the MCS database are always paired by peering-relation, sync-tag and application-id. At any time the specified entry is related to the single redundant-pair objects (two SAPs on two different nodes) and therefore stored in a local MCS database of the respective nodes.

Internally, peering-relation and sync-tag are translated into a port and encapsulation value identifying the object (SAP) that the specified entry is associated with. The application-id then identifies the application which created the entry on one of the nodes. There are three basic operations that the application can perform on MCS database. The MCS database always synchronizes these operations with its respective peer for the specified entry.

The following principles apply:

- **add-operation**

  Any dynamic-state created in the application is pushed to the MCS database. MCS then creates and synchronizes with the corresponding peer provided (if configured). The application in the peer node is then notified as soon as the entry has been created. Similarly, the application in the local node (the node where the state has been created) is notified that entry has been synchronized (MCS is "in-sync" state). This operation is also used to modify existing MCS database entry.

- **local-delete**

  The MCS database entry is marked as no longer in use locally and this information is sent to the peer node. If the information is no longer used by applications on both nodes (the application in remote-node has already issued local-delete before), it is removed from database.

- **global-delete**

  The MCS database entry is removed from both nodes and from the application in the remote node.

The choice of the operation in corresponding situation is driven by the application. The following general guidelines are observed:

- An event which leads to a dynamic-state deletion on a standby chassis is handled as "local-delete".
- An event which leads to a dynamic-state deletion on an active chassis is handled as "global-delete".
- An exception to above the rules is an explicit **clear** command which is handled as global-delete regardless of where the command was executed.

As previously stated, the MCS process automatically synchronizes any database operation with the corresponding peer. During this time, the MCS process maintains state per peer indicating to the applications (and network operator) the current status, such as in-sync, synchronizing or sync_down. These states are indicated by corresponding traps.

9.3.16.1.1 Loss of Synchronization and Reconciliation

Each time the connection between the redundant pair nodes is established or re-established, the MCS database is re-synchronized. There are several levels of connectivity loss that can have different effects on amount of data lost. To prevent massive retransmissions when the synchronization connection experiences loss or excessive delay, the MCS process implementation takes provisions to ensure following:

- If a reboot of one or both nodes or establishing the peering for the first time, the full MCS database is reconciled.
- If the MCS communication is lost and then re-established but neither node rebooted during the connection loss, only the information not synchronized during this time is reconciled (using sequence numbers helps identify information which was not synchronized).
- If that MCS communication is lost because of excessive delay in ACK messages but no information has been effectively lost, the MCS process indicates a loss of synchronization but no reconciliation is performed.

9.3.16.2 DHCP lease state synchronization optimization

In a stateful BNG dual homing setup, Multi-Chassis Synchronization (MCS) is used to synchronize the subscriber state between the active and standby BNG, including the DHCP lease states, using **configure redundancy multi-chassis options sub-mgmt** configuration. For IPoE subscribers the synchronization includes DHCPv4 and DHCPv6 lease states and for PPPoE subscribers the synchronization includes DHCPv6 lease states.

The DHCP lease states are synchronized as follows:

- when the lease is created in the system
- at every DHCP renewal
- when the lease is removed from the system

With short lease times in a scaled deployment, MCS synchronization creates additional load on the control plane. Short least times typically occur when the lease-split feature is enabled because a short lease time is used between the DHCP client and DHCP relay agent and a long lease time is used between the DHCP relay agent and DHCP server. With lease split enabled, the MCS application synchronizes the DHCP renewals following the short lease speed.

To reduce the control plane load in scaled multichassis redundant BNG deployments with short DHCP leases, a DHCP lease time threshold can be configured to control the eligibility of a DHCP lease for MCS synchronization at renewal:

```
# configure redundancy multi-chassis options sub-mgmt dhcp-leasetime-threshold
  - dhcp-leasetime-threshold [days <days>] [hrs <hours>] [min <minutes>] [sec <seconds>]
  - no dhcp-leasetime-threshold
 <days>               : [0..1]
 <hours>              : [0..23]
 <minutes>            : [0..59]
 <seconds>            : [0..59]
```

An active DHCP lease time threshold per multichassis peer is determined as the smallest value configured on either of the redundant BNGs:

```
# show redundancy multi-chassis sync peer 192.0.2.1
===============================================================================
Multi-chassis Peer Table
===============================================================================
Peer
-------------------------------------------------------------------------------
Peer IP Address         : 192.0.2.1
Peer Name               : mcs-pe2-pe1
Description             : (Not Specified)
Authentication          : Disabled
Source IP Address       : 192.0.2.2
Admin State             : Enabled
Sub-mgmt options        :
  DHCP lease threshold  : Active (10 minutes)
    Local / Remote      : 10 minutes / 15 minutes
```

The DHCP lease time threshold is inactive when unconfigured or unsupported on at least one end of the multichassis peer:

```
# show redundancy multi-chassis sync peer 192.0.2.1
===============================================================================
Multi-chassis Peer Table
```

```
===============================================================================
Peer
-------------------------------------------------------------------------------
Peer IP Address         : 192.0.2.1
Peer Name               : mcs-pe2-pe1
Description             : (Not Specified)
Authentication          : Disabled
Source IP Address       : 192.0.2.2
Admin State             : Enabled
Sub-mgmt options        :
  DHCP lease threshold  : Inactive
    Local / Remote      : 10 minutes / --
```

DHCP leases with lease time committed by the DHCP server less than or equal to the active DHCP lease time threshold are not synchronized at renewal, if only the remaining lease time is changed.

When lease split is active, the following rules apply if the short lease time is less than or equal to the active DHCP lease time threshold:

- The DHCP lease is not synchronized when the DHCP client renewal or rebind is proxied, only the remaining short lease time is changed, and at least one DHCP server is reachable.
- The DHCP lease is always synchronized when the DHCP client renewal or rebind is relayed to the DHCP server.

After an MCS redundancy switchover, DHCP leases that are flagged to skip MCS synchronization are granted the full lease time in the new active BNG. This could lead to a temporary address conflict when a client disconnects ungracefully immediately after such a switchover as illustrated in the following scenario:

- A DHCP client lease with 15 minutes lease time is active on a redundant BNG pair with active DHCP lease time threshold equaling 20 minutes.
- Five minutes after the last DHCP client renewal, an SRRP switchover occurs. At the new active BNG, the DHCP lease is eligible to be extended to the DHCP server committed lease time of 15 minutes while the client and server have a remaining lease time of 10 minutes.
- If the client disconnects ungracefully before the next renewal (for example, by not sending a DHCP release), the state in the BNG is not cleared and the session lives longer than expected.
- The lease in the DHCP server expires 10 minutes after the switchover, while the lease in the BNG is still active. The DHCP server can allocate the same address or prefix to another user, which could create a temporary address conflict in the BNG.

Note:
The MCS DHCP lease time threshold is not applicable for DHCP server failover (using the **configure redundancy multi-chassis peer sync local-dhcp-server** context) and not applicable for DHCP snooping.

9.3.17 Subscriber Routed Redundancy Protocol

9.3.17.1 SRRP messaging

Subscriber Routed Redundancy Protocol (SRRP) uses the same messaging format as VRRP with slight modifications. The source IP address is derived from the system IP address assigned to the local router. The destination IP address and IP protocol are the same as VRRP (224.0.0.18 and 112, respectively).

The message type field is set to 1 (advertisement) and the protocol version is set to 8 to differentiate SRRP message processing from VRRP message processing.

The vr-id field has been expanded to support an SRRP instance ID of 32 bits.

Because of the large number of subnets backed up by SRRP, only one message every minute carries the gateway IP addresses associated with the SRRP instance. These gateway addresses are stored by the local SRRP instance and are compared with the gateway addresses associated with the local subscriber IP interface.

Unlike VRRP, only two nodes may participate in an SRRP instance because of the explicit association between the SRRP instance group IP interface, the associated redundant IP interface and the multichassis synchronization (MCS) peering. Because only two nodes are participating, the VRRP skew timer is not used when waiting to enter the SRRP master state. Also, SRRP always preempts when the local priority is better than the current SRRP master instance and the backup SRRP instance always inherits the SRRP master's instance advertisement interval from the SRRP advertisement messaging.

SRRP advertisement messages carry a *becoming-master* indicator flag. The *becoming-master* flag is set by a node that is attempting to usurp the master state from an existing SRRP master router. When receiving an SRRP advertisement message with a better priority and with the *becoming-master* flag set, the local SRRP master initiates the *becoming-backup* state, stops routing with the SRRP gateway MAC and sends an SRRP advertisement message with a priority set to zero. The new SRRP master continues to send SRRP advertisement messages with the *becoming-master* flag set until it either receives a return priority zero SRRP advertisement message from the previous SRRP master or its *becoming-master* state timer expires. The new backup node continues to send zero priority SRRP advertisement messages every time it receives an SRRP advertisement message with the *becoming-master* flag set. After the SRRP new master either receives the old SRRP master's priority zero SRRP advertisement message or the *become-master* state timer expires, it enters the SRRP *master* state. The *become-master* state timer is set to 10 seconds upon entering the *become-master* state.

The SRRP advertisement message is always evaluated to see if it has higher priority than the SRRP advertisement that would be sent by the local node. If the advertised priority is equal to the current local priority, the source IP address of the received SRRP advertisement is used as a tie breaker. The node with the lowest IP address is considered to have the highest priority.

The SRRP instance maintains the source IP address of the current SRRP master. If an advertisement is received with the current SRRP master's source IP address and the local priority is higher priority than the SRRP masters advertised priority, the local node immediately enters the *becoming-master* state unless the advertised priority is zero. If the advertised priority is zero, the local node bypasses the *becoming-master* state and immediately enters the SRRP *master* state. Priority zero is a special case and is sent when an SRRP instance is relinquishing the SRRP master state.

9.3.17.2 SRRP and multichassis synchronization

To take full advantage of SRRP resiliency and diagnostic capabilities, the SRRP instance should be tied to a MCS peering that terminates on the redundant node. The SRRP instance is tied to the peering using the **srrp** *srrp-id* command within the appropriate MCS peering configuration. After the peering is associated with the SRRP instance, MCS synchronizes the local information about the SRRP instance with the neighbor router. MCS automatically derives the MCS key for the SRRP instance based on the SRRP instance ID. For example, an SRRP instance ID of 1 would appear in the MCS peering database with a MCS-key srrp-0000000001.

The SRRP instance information stored and sent to the neighbor router consists of:

- The SRRP instance MCS key

- Containing service type and ID
- Containing subscriber IP interface name
- Subscriber subnet information
- Containing group IP interface information
- The SRRP group IP interface redundant IP interface name, IP address and mask
- The SRRP advertisement message SAP
- The local system IP address (SRRP advertisement message source IP address)
- The Group IP interface MAC address
- The SRRP gateway MAC address
- The SRRP instance administration state (up or down)
- The SRRP instance operational state (disabled/becoming-backup, becoming-master, master)
- The current SRRP priority
- Remote redundant IP interface availability (available or unavailable)
- Local receive SRRP advertisement SAP availability (available or unavailable)

9.3.17.3 SRRP instance

The SRRP instance uses the received information to verify provisioning and obtain operational status of the SRRP instance on the neighboring router.

9.3.17.3.1 SRRP instance MCS key

The SRRP instance MCS key ties the received MCS information to the local SRRP instance with the same MCS key. If the received key does not match an existing SRRP instance, the MCS information associated with the key is ignored. After an SRRP instance is created and mapped to an MCS peering, the SRRP instance evaluates received information with the same MCS key to verify it corresponds to the same peering. If the received MCS key is on a different peering than the local MCS key an SRRP peering mismatch event is generated detailing the SRRP instance ID, the IP address of the peering the MCS key is received on and the IP address to which the local MCS key is mapped. If the peering association mismatch is corrected, an SRRP peering mismatch clear event is generated.

9.3.17.3.2 Containing service type and ID

The containing service type is the service type (IES or VPRN) that contains the local SRRP instance. The containing service ID is the service ID of that service. This information is supplied for troubleshooting purposes only and is not required to be the same on both nodes.

9.3.17.3.3 Containing subscriber IP interface name

The containing subscriber IP interface name is the subscriber IP interface name that contains the SRRP instance and its group IP interface. This information is supplied for troubleshooting purposes only and is not required to be the same on both nodes.

9.3.17.3.4 Subscriber subnet information

The subscriber subnet information includes all subscriber subnets backed up by the SRRP instance. The information for each subnet includes the Owned IP address, the mask and the gateway IP address. If the received subscriber subnet information does not match the local subscriber subnet information, an SRRP Subscriber Subnet Mismatch event is generated describing the SRRP instance ID and the local and remote node IP addresses. After the subscriber subnet information matches, an SRRP Subscriber Subnet Mismatch Clear event is generated.

9.3.17.3.5 Containing group IP interface information

The containing group IP interface information is the information about the group IP interface that contains the SRRP instance. The information includes the name of the group IP interface, the list of all SAPs created on the group IP interface, the administrative and operational state of each SAP and the MCS key and the peering destination IP address associated with each SAP. To obtain the MCS information, the SRRP instance queries MCS to determine the peering association of the SRRP instance and then queries MCS for each SAP on the group IP interface. If the local SRRP instance is associated with a different MCS peering than any of the SAPs or if one or more SAPs are not tied to an MCS peering, an SRRP group interface SAP peering mismatch event is generated detailing the SRRP instance ID, and the group IP interface name.

When receiving the remote containing group IP interface information, the local node compares the received SAP information with the local group IP interface SAP information. If a local SAP is not included in the SAP information or a remote SAP is not included in the local group IP interface, an SRRP Remote SAP mismatch event is generated detailing the SRRP instance ID and the local and remote group IP interface names. If a received SAP’s MCS key does not match a local SAP's MCS Key, an SRRP SAP MCS key mismatch event is generated detailing the SRRP instance ID, the local and remote group IP interface names, the SAP-ID and the local and remote MCS keys.

9.3.17.3.6 Remote redundant IP interface mismatch

If the group IP remote redundant IP interface address space does not exist, is not within the local routing context for the SRRP instances group IP interface or is not on a redundant IP interface, the local node sends redundant IP interface unavailable to prevent the remote neighbor from using its redundant IP interface. An SRRP redundant IP interface mismatch event is generated for the SRRP instance detailing the SRRP instance, the local and remote system IP addresses, the local and remote group IP interface names and the local and remote redundant IP interface names and IP addresses and masks. The local redundant IP interface may still be used if the remote node is not sending redundant IP interface unavailable.

9.3.17.3.7 Remote sending redundant IP interface unavailable

If the remote node is sending redundant IP interface unavailable, the local node treats the local redundant IP interface associated with the SRRP instances group IP interface as down. A Local Redundant IP Interface Unavailable event is generated detailing the SRRP instance ID, the local and remote system IP addresses, the local group IP interface name, the local redundant IP interface name and the redundant IP interface IP address and mask.

9.3.17.3.8 Remote SRRP advertisement SAP non-existent

If the remote node's SRRP advertisement SAP does not exist on the local SRRP instances group IP interface, the local node sends local receive SRRP advertisement SAP unavailable to the remote node. An SRRP receive advertisement SAP non-existent event is generated detailing the SRRP instance ID, the local and remote system IP addresses, the local group IP interface name and the received remote SRRP advertisement SAP. Because SRRP advertisement messages cannot be received, the local node immediately becomes SRRP master if it has the lower system IP address.

9.3.17.3.9 Remote sending local receive SRRP advertisement SAP unavailable

If the local node is receiving local receive SRRP advertisements stating that the SAP is unavailable from the remote node, an SRRP Remote Receive advertisement SAP Unavailable event is generated. This details the SRRP instance ID, the local and remote system IP addresses, the remote group IP interface name and the local SRRP advertisement SAP. Because the remote node cannot receive SRRP advertisement messages, the local node immediately becomes SRRP master if it has the lower system IP address.

9.3.17.3.10 Local and remote dual SRRP master state detected

If both local and remote SRRP instances are in master states, then an SRRP dual master event is generated detailing the SRRP instance ID and the local, remote system IP addresses and the local and remote group IP interface names and port numbers.

9.3.17.4 Subscriber subnet-owned IP address connectivity

In order for the network to reliably reach the owned IP addresses on a subscriber subnet, the owning node must advertise the IP addresses as /32 host routes into the core. This is important because the subscriber subnet is advertised into the core by multiple routers and the network follows the shortest path to the closest available router which may not own the IP address if the /32 is not advertised within the IGP.

9.3.17.5 Subscriber subnet SRRP gateway IP address connectivity

The SRRP gateway IP addresses on the subscriber subnets cannot be advertised as /32 host routes because they may be active (SRRP master state) on multiple group IP interfaces on multiple SRRP routers. Without a /32 host route path, the network forwards any packet destined for an SRRP gateway IP address to the closest router advertising the subscriber subnet. While a case may be made that only a node that is currently forwarding for the gateway IP address in an SRRP master state should respond to ping or other diagnostic messages, the distribution of the subnet and the case of multiple SRRP master states make any resulting response or non-response inconclusive at best. To provide some ability to ping the SRRP gateway address from the network side reliably, any node receiving the ICMP ping request responds if the gateway IP address is defined on its subscriber subnet.

9.3.17.6 Receive SRRP advertisement SAP and anti-spoof

The group IP interface SAPs are designed to support subscriber hosts and perform an ingress anti-spoof function that ensures that any IP packet received on the group IP interface is coming in the correct SAP with the correct MAC address. If the IP and MAC are not registered as valid subscriber hosts on the SAP, the packet is silently discarded. Because the SRRP advertisement source IP addresses are not subscriber hosts, an anti-spoof entry cannot exist and SRRP advertisement messages would normally be silently discarded. To avoid this issue, when a group IP interface SAP is configured to send and receive SRRP advertisement messages, anti-spoof processing on the SAP is disabled. This precludes subscriber host management on the SRRP messaging SAP.

9.3.18 PPPoE MC redundancy

This feature minimizes the downtime for PPPoE clients in an ESM environment when a single node fails.

It is not necessary that an entire BNG fails before it triggers the corrective action. The solution described in this section includes protection against interfaces and line card failures within the BNG. The redundant (protective) entity, however, does not reside within the same BNG on which the failure occurs but instead it is on a separate BNG node.

The PPPoE MC Redundancy is based on SRRP and MC-LAG because SRRP is already established in ESM providing IPoE MC Redundancy. With some modifications, SRRP approach is adopted to PPPoE deployments.

9.3.18.1 SRRP considerations for PPPoE

SRRP is based on VRRP whose purpose is to provide a default gateway redundancy for clients sharing the transport medium such as Ethernet. IPoE would be a typical example of this where IPoE clients use a virtual IP and MAC address that is shared between two default gateway nodes in an active or standby configuration. SRRP supports only two nodes in a cluster but VRRP allows multiple nodes to be configured in a cluster with a priority that determines which node assumes the master state. Although it is mandatory for the correct operation of IPoE clients that the same SRRP IP address is shared between the two BNG nodes providing redundancy, having the same SRRP IP address is not necessary for the operation of SRRP itself. In other words, SRRP itself (Master or Backup states) works with different SRRP IP addresses on each node. Same is valid for MAC addressing. It is possible by configuration that the redundant BNG nodes use different IP/MAC addresses on a pair of SRRP instances.

Upon a switchover, a gratuitous ARP is sent from a newly selected active node so that each IPoE client can update the ARP table, if the MAC address has indeed changed (it does not have to). More importantly, if an Layer 2 aggregation network is in place between the BNG and the IPoE client, all intermediate Layer 2 devices must update their port-to-mac mappings (Layer 2 FDB). The above described process ensures correct packet addressing on the IPoE client side as well as the correct forwarding path through Layer 2 aggregation network to the newly activated BNG.

When considering PPPoE in conjunction with SRRP, keep in mind that PPP protocol (point-to-point protocol) is adopted for the Ethernet (shared medium) by enabling an extra Ethernet related layer in PPP that allows sharing of point-to-point sessions over Ethernet (shared medium). The result is a PPPoE protocol designed to 'tunnel' each PPP session over Ethernet.

PPPoE is not aware of ARP (Address Resolution Protocol) and it does not react to gratuitous ARP packets sent by a newly active BNG. The destination MAC address that PPPoE clients use when sending traffic is determined not by ARP but by the PPPoE Discovery phase at the beginning of the session establishment.

This originally discovered destination MAC is used throughout the lifetime of the session. This has a couple of consequences:

1. If SRRP is used for PPPoE then the 'SRRP' MAC address between the redundant BNG nodes must be shared. It is not allowed to use a unique 'SRRP' MAC address per BNG in the redundant pair of BNG nodes (as it can for IPoE). Every PADx conversation is based on the SRRP shared MAC address, that is, the PADO reply must have the shared SRRP MAC address as the source MAC. This has a significant impact on the operation of MSAP in conjunction with this feature.
2. Because PPPoE sessions are not ARP aware, the only purpose of the gratuitous ARP would be to update the Layer 2 FDB in the aggregation network (and not the PPPoE client destination MAC address). For IPoE, the gratuitous ARP is sent for all subnet gateway IP addresses found under the subscriber interface over either all SAPs (default) or top-tags only. For PPPoE, the gratuitous ARP is sent only for the system IP address. The purpose of the gratuitous ARP in PPPoE scenario is only to update Layer 2 network path which is otherwise IP unaware. It is not necessary to send the gratuitous ARP for every default-gateway address found under the subscriber-interface. Because this feature is only applicable to PPPoE deployments, therefore, only PPPoE is present under the group interface. This is indicated by the following command under the SRRP node:

```
group-interface <ip-int-name>
    srrp <id>
        one-garp-per-sap
```

9.3.18.1.1 SRRP fact checks

1. After Multi-chassis Synchronization (MCS) for subscriber management and SRRP are enabled, both BNG nodes, Active (SRRP master state) and Standby (SRRP backup state) forward packets (for subscribers) in both directions.
2. Traffic flows through an SRRP enabled node according to the entries in the SRRP sync database and the SRRP state of the node:
   - SRRP in backup state directs downstream traffic over the redundant-interface toward the active node (SRRP master state). If the redundant interface is unavailable, traffic is sent directly to the subscriber.
   - SRRP in master state always directly forwards the downstream traffic toward the subscriber.
   - In the upstream direction, the active SRRP node accepts subscriber traffic addressed either to the MAC address of the SRRP active group OR the native interface MAC address.
   - The standby node accepts in the upstream direction only packets addressed to its native interface MAC address.
3. If both SRRP nodes become active (SRRP master state), then both forward traffic to or from subscribers unaware of the link failure somewhere in the Layer 2 network. As a result, downstream traffic can be blackholed. Whether downstream traffic is lost depends on the native routing on the network side, which is unaware of the failures in the aggregation network.

9.3.18.2 State synchronization

PPPoE sessions are synchronized between the redundant BNG nodes. The subscriber synchronization is achieved through Multi-Chassis Synchronization (MCS) protocol in a similar way it is performed for IPoE.

```
multi-chassis
            peer <IP@>create
                sync
                    local-dhcp-server
                    SRRP
                    sub-mgmt [ipoe | pppoe]
                 :
:
                    no shutdown
                exit
                no shutdown
            exit
```

Two keywords, **ipoe** and **pppoe** enable a more granular control over which type of subscribers the MCS should be enabled.

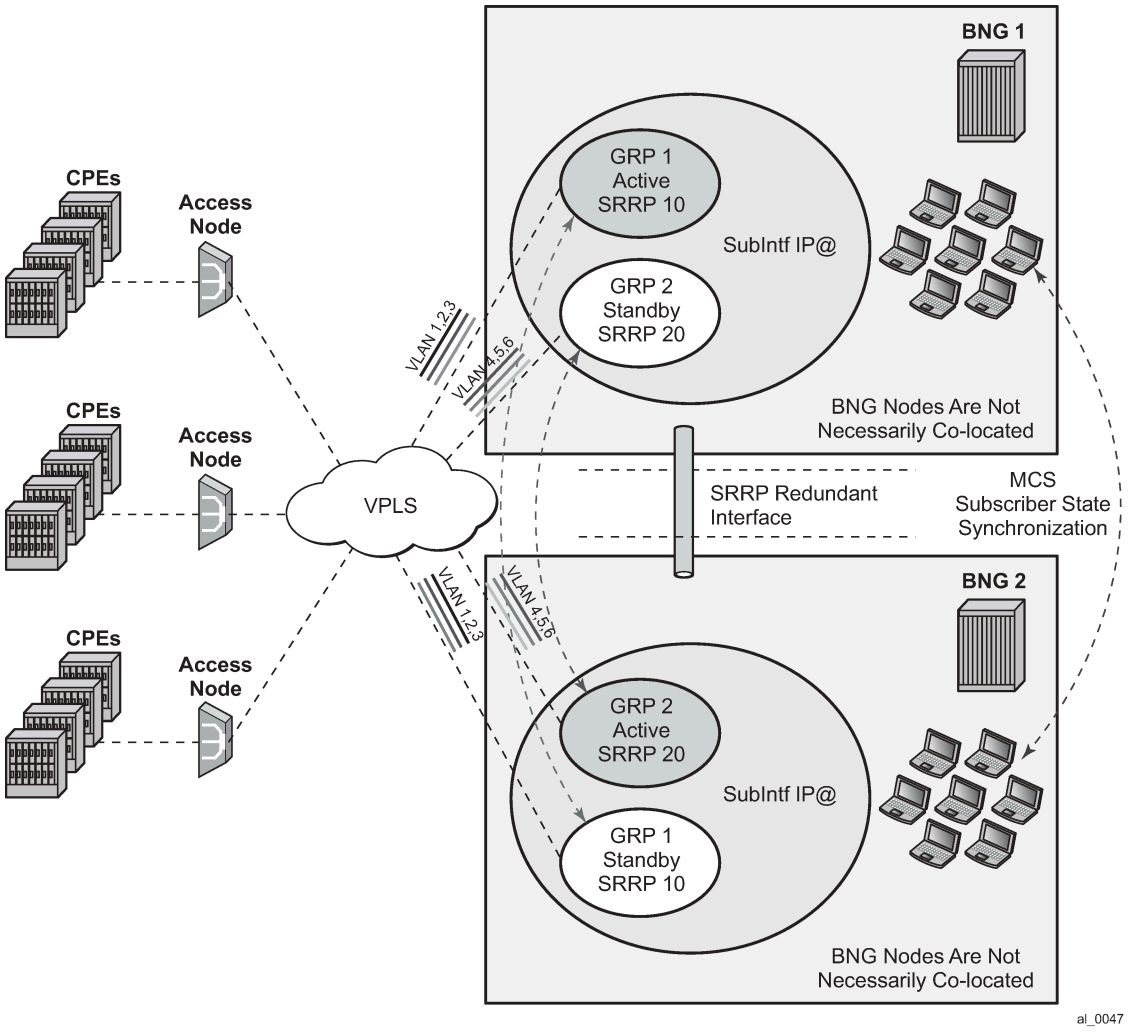
Subscriber synchronization is important for following reasons:

- Forwarding of downstream traffic between the redundant BNG nodes through a redundant interface is an artifact of how natural routing steers traffic through the network.
- Subscriber instantiation on the node which did not originally create subscriber session. This drastically reduces downtime during the SRRP switchover.
- Monitors operational aspects of the subscriber management through show commands.

9.3.18.2.1 PPPoE multichassis synchronization model

The PPPoE multichassis synchronization (MCS) model is based on SRRP synchronization and can be used in a centralized or distributed environment with or without Layer 2 aggregation network in-between access nodes and BNGs. The failure detection speed is dependent on SRRP timers. Traffic load can be balanced per SRRP group over the two links. In this model (Figure 90: Fully redundant "stateful 1:1" model), PPPoE states are synchronized between the redundant BNG nodes. If one BNG fails, the newly activated BNG sends out a 'MAC update' (gratuitous ARP) message prompting the intermediate Layer 2 nodes to update their forwarding tables so that forwarding can resume. The SRRP timers can be configured in the sub-second range. In reality, the limiting factor for timer values is the scale of the deployment, in particular the number of SRRP groups per node.

Figure 90: Fully redundant "stateful 1:1" model



9.3.18.3 Traffic control and redundant interface

To preserve QoS and Accounting, subscriber's traffic must flow in both directions through the multichassis active BNG node.

In the upstream direction, this is always true as traffic is steered to the active BNG (SRRP master state) node just by the virtue of SRRP operation.

In the downstream direction which represents bulk of traffic, SRRP cannot be relied up on to steer traffic through the active BNG (SRRP master state). This poses a problem in a very common environment where IP subnets are shared over multiple group interfaces with SRRP enabled. A particular subnet is advertised to the network side from both active and standby BNGs. Natural routing on the network side determines which BNG node receives subscriber's traffic in the downstream direction. If the standby BNG (SRRP backup state) node receives the traffic, it cannot simply send the traffic directly to the access network where the subscriber resides by just inserting the source MAC address of the SRRP instance in the outgoing packet. This would break the operation of SRRP. Instead, the standby BNG must send the

traffic to the active BNG through a redundant interface. The active BNG then forwards traffic directly to the subscriber. Source MAC address of this traffic is the MAC address of SRRP instance. This traffic shunting over the redundant interface can result in a substantial load on the link between the two BNGs.

The increase in shunted traffic can quickly become an issue if the redundant BNG nodes are not colocated. To minimize the shunt traffic, more granular routing information must be presented to the network core. This leads to more optimal routing where downstream subscriber traffic is directed toward the active BNG, without the need to cross the redundant interface. The disadvantage of this approach is that this further fragments the IP address space within the network core. In the extreme case where /32 (subscriber) IP addresses are advertised, the churn that /32s cause in the core routing can be unsustainable. In this case, routing updates in the core are triggered by subscribers coming on/off-line.

Optimal operation calls for the shunt traffic to be eliminated and at the same time, a high IP route aggregation on the network side is achieved. The existence of the shunt traffic stems from the fact that routing protocols advertise subscriber subnets into the network with no awareness of the SRRP master or backup state. To address this problem along with better aggregation of advertised subnets, two SRRP enhancements are introduced:

- SRRP fate-sharing
- SRRP aware routing

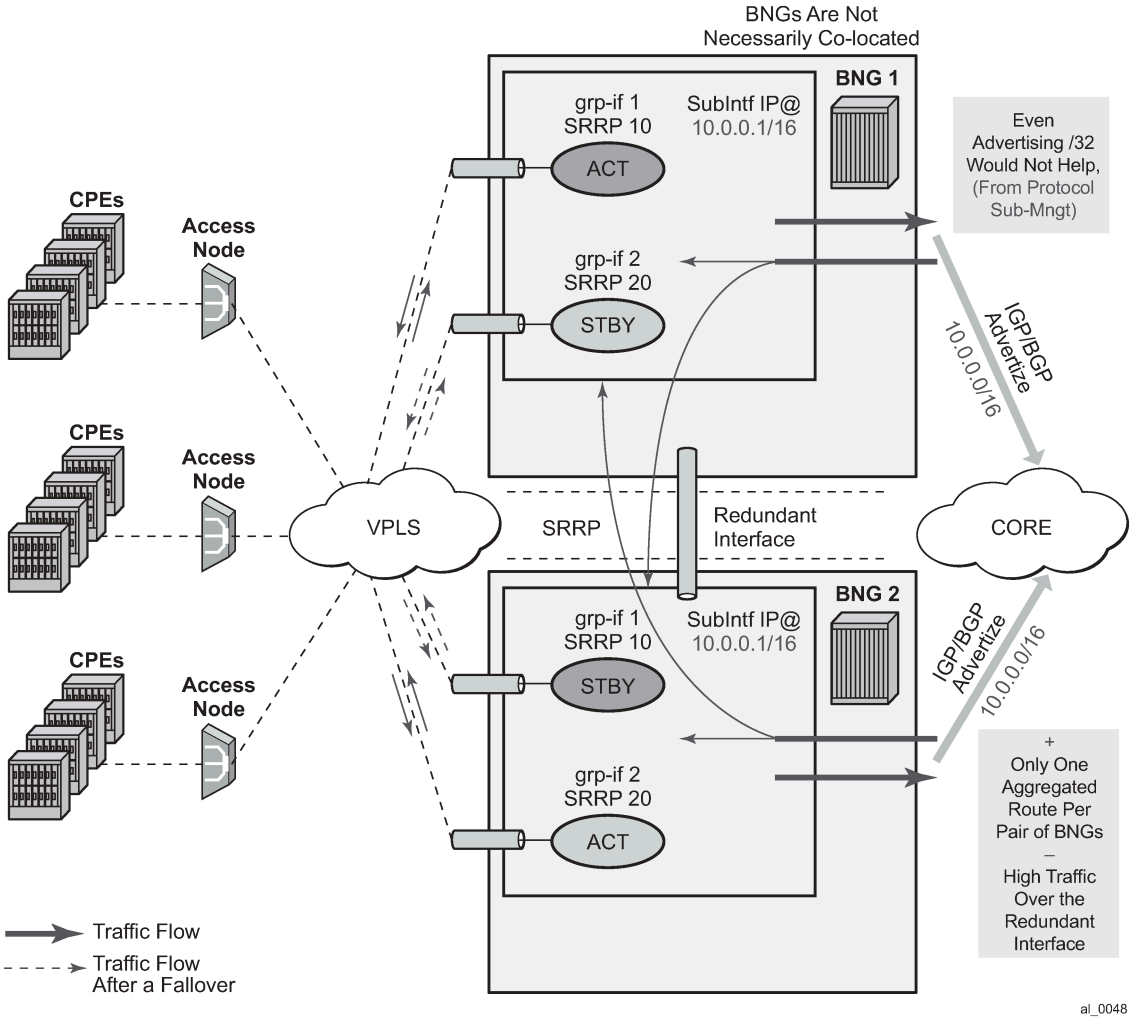
Both of these concepts are described in SRRP enhancement.

Traffic destined for or from the subscriber is forwarded under the condition that the subscriber-interface is operationally UP. This applies also to shunting of downstream subscriber traffic from the standby (SRRP backup state) to the active (SRRP master state) node. It is always necessary to keep the subscriber-interface operationally UP by configuring a dummy group interface with a **oper-up-while-empty** command under it. This is especially true for the MC-LAG which causes the messaging SAP on the STANDBY node always to be in the INIT state. In case that MSAPs are used on such group interfaces, the group interfaces would be also operationally DOWN, causing the subscriber-interface to be operationally DOWN.

9.3.18.3.1 Subnet assignment and advertisement - option A

A single IP subnet is used for all subscribers terminated within the redundant BNG nodes. The upside of the option A is that it offers aggregated IP addressing in the network core per pair of redundant BNG nodes. The downside is that the subscriber termination point (active BNG for the SRRP group) is hidden from the network core. Because both BNG nodes share the same IP subnet for the subscribers, the natural routing can cause downstream traffic to be sent to the standby BNG which must shunt the traffic to the active BNG. It is likely that half of the traffic is shunted over the redundant-interface with this approach. This scenario is shown in Figure 91: Shared subscriber IP space.

Figure 91: Shared subscriber IP space



9.3.18.3.2 Subnet assignment and advertisement - option B

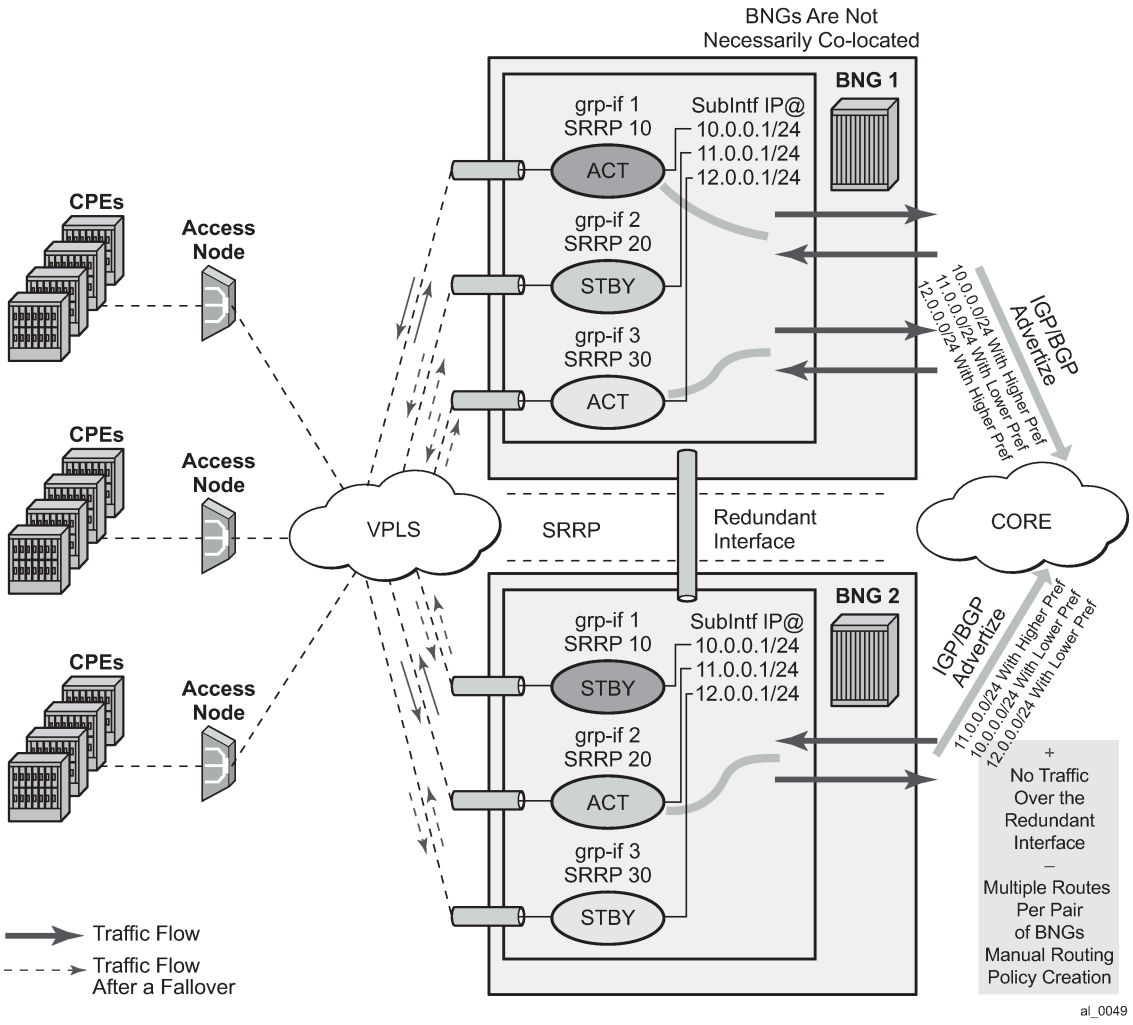
With the option B, an IP address pool (or subnet) can be allocated per group of SRRP instances that are in the SRRP master state. The routing decision on the network side is further influenced by the static increase of the metric of the advertised route on the BNG node hosting the active SRRP groups (Figure 92: Option B – IP subnet per active SRRP group).

This approach would cause greater IP space segmentation in the network core, but at the same time, it would indirectly provide more information about the subscriber whereabouts and therefore minimize or eliminate the shunt traffic during the normal operation. However, if an SRRP switchover occurs, the shunt traffic would ensue. The amount of the shunted traffic would depend on the scale of the failure. From the configuration displayed in Figure 92: Option B – IP subnet per active SRRP group, it can be concluded that:

- There is no shunted traffic.

- If any of the SRRP instances transitions out of the SRRP master state, traffic for an entire IP network associated with this failed SRRP instance would be shunted. The reason for this is that the advertised route metric is static and it does not follow changes in SRRP state.

Figure 92: Option B – IP subnet per active SRRP group



9.3.18.4 MSAP considerations

As per RFC 2516, *A Method for Transmitting PPP Over Ethernet (PPPoE)*, this has the implications on the operation of the capture SAP. In an IPoE environment, the initial DHCP traffic related to host establishment uses its native MAC of the physical port on the router. After the group interface is learned (later in the process, by RADIUS or msap-policy), the MAC address is switched to SRRP MAC address (virtual MAC). The IPoE client adapts easily to this change. On the contrary, for the correct operation of PPPoE with SRRP, the initial destination MAC address learned by the PPPoE client does not change during the lifetime of the session.

This is ensured by indirectly referencing the grp-if under the capture SAP:

```
config>service>vpls
    sap 1/1/1:1.* capture-sap
        track-srrp 10
    sap 1/1/1:2.* capture-sap
        track-srrp 20

config>service>vprn>
    subscriber-interface <ip-int-name>
        group-interface <ip-int-name>
            sap 1/1/1:1.1
                srrp 10
                message-path 1/1/1:1.1

    group-interface <ip-int-name>
        sap 1/1/1:2.1
            srrp 20
                message-path 1/1/1:2.1
```

With this approach the grp-if is nailed during the session initiation phase by referencing the SRRP instance in track-srrp statement (SRRP is a group interface-wide concept). RADIUS returned grp-if name must match the one on which referenced SRRP instance runs.

The capture SAP of the form

```
sap port-id:*.* capture-sap
    track-srrp X
```

assumes that there is only one grp-if associated with all MSAPs under this capture SAP.

A check is put in place to make sure that the MAC addresses associated with the SRRP instance is the same as the MAC address of the associated capture SAP. A log is raised if there is a discrepancy between the MAC addresses while the grp-if is operationally UP. If there is a MAC address change (user misconfiguration) then the existing PPPoE sessions time out and the new sessions fail to establish until the condition is corrected.

9.3.18.5 Unnumbered interface support

For unnumbered subscriber-interface support in PPPoE, the gateway IP address that is used to send gratuitous ARP is not available. For this reason, the system IP address is used to send gratuitous ARPs. Gratuitous ARP is used to update the Layer 2 network forwarding path toward the BNG node in the upstream direction.

The system IP address is used automatically if the subscriber interface is unnumbered.

9.3.18.6 Compatibility with MC-LAG

SRRP for PPPoE works in an environment where MC-LAG is enabled. For example, the standby MC-LAG link automatically puts the SRRP instance in a backup state and the active MC-LAG link puts the SRRP instance in a master state. It is important that the SRRP instance on the standby leg of the MC-LAG is forced into a SRRP backup state, or any other state that forces the downstream traffic to use the redundant interface.

Traffic destined for or from the subscriber is forwarded under the condition that the subscriber-interface is operationally UP. This applies also to shunting of downstream subscriber traffic from the standby (SRRP

backup state) to active (SRRP master state) node. It is always necessary to keep the subscriber-interface operationally UP by configuring a dummy group interface with a **oper-up-while-empty** command under it. This is especially true for the MC-LAG which causes the messaging SAP on the standby node always to be in the INIT state. If MSAPs are used on such group interfaces, the group interfaces would be also operationally DOWN, causing the subscriber-interface to be operationally DOWN.

9.3.18.7 IPv6 support

Prerequisite for MC IPv6 Redundancy is to synchronize PPPoEv6 and IPoEv6 subscribers between the nodes by MCS.

In PPPoE environment, SRRP is used to refresh the forwarding path (MAC addresses) in the access aggregation network (by gratuitous ARP). SRRP ensures that the upstream traffic is steered to the active BNG node (SRRP master state). In the downstream direction, the standby BNG directs traffic over to the active BNG node via a redundant-interface.

The IPv6 functionality currently relies on IPv4 based SRRP and IPv4 based redundant-interface. In other words, IPv4 is required to run on the access side as well as on the redundant-interface.

The redundant-interface is used in the downstream direction. Traffic arriving on the network links on the standby node is shunted over to the active node over the redundant-interface. This is required to ensure consistent QoS and accounting functionality across the nodes (upstream and downstream traffic on the access links for a subscriber must traverse the same BNG node). There is no IPv6 related CLI associated with the redundant-interface.

All IPv6 subscriber traffic that arrives on the standby node in the downstream direction is automatically shunted over the IPv4 redundant-interface to the active node. When IPv6 traffic arrives over the redundant-interface on the active node, it is either PPPoEv6 encapsulated or left as plain IPoEv6 before it is forwarded to the subscriber.

In the upstream direction (AN->BNG) the behavior is the following:

- PPPoEv6

  On the switchover, gratuitous ARPs (gARP) is sent from the new active BNG (SRRP master state) on each VLAN. The IP address in gARP is the IPv4 gw-ip address or the system IP if there are unnumbered interfaces. This updates the Layer 2 network path with the correct SRRP MAC address.

- IPoEv6

  IPv4 based SRRP is used to update the Layer 2 forwarding path in the case of a switchover. A gratuitous ARP is sent in the same way as it is used for IPoE v4 hosts. Router Advertisements (RA) are not sent out in the event of the switchover.

  However, the two BNG nodes share the same virtual Link Local (LL) IPv6 address. This address is used by the clients as a default-gw and only the active BNG (SRRP master state) advertises this LL address in RAs. RAs are suppressed on the standby BNG. As previously mentioned, RAs are not sent during the switchover. RAs are sent:

  – **When the client is first established**

    This is how the client learns its default-gw (in PPPoE case RA can also be used for SLAAC, stateless address configuration).

  – As a reply to Router Solicitations messages sent by the clients

  – Periodically to each client

  Note that RAs are unicast to each client.

Neighbor Advertisements (NA) used for address resolution are sent only from the active BNG. NA has the SRRP MAC address in the target link layer option on SRRP enabled group interfaces (on non-SRRP enabled group interfaces, NAs contains the group interface MAC address).

The LL IPv6 address must be the same on both nodes. In addition, the **gw-mac** address must be the same on both nodes. The IPv6 clients are not aware of the switchover and therefore they do not send NS to solicit the update of its neighbor cache with the possibly different **gw-mac** address.

The syntax to configure the LL address on the subscriber interface is as follows:

```
config>service>ies | vprn>
    subscriber-interface <ip-int-name>
        ipv6
            [no] link-local-address <ipv6-address>
```

```
        <ipv6-address>       : ipv6-address   - x:x:x:x:x:x:x:x
                                                x:x:x:x:x:x:d.d.d.d
                                                 x [0..FFFF]H
                                                 d [0..255]D
```

Note that the current version of SRRP relies only on IPv4 routes. The connection between SRRP and IPv4 routes is done with the subnets with gw IP addresses defined under the subscriber-interfaces in the ESM context. This connection is needed so that SRRP can send Gratuitous ARP properly.

These are the cases for PPPoEv6 MC Redundancy that are supported:

- unnumbered subscriber-interfaces (**config**>**service**>**subscriber-interface** hierarchy)
- numbered IPv4 subscriber-interfaces (**config**>**service**>**subscriber-interface** hierarchy)
- numbered IPv4 and IPv6 subscriber-interfaces (**config**>**service**>**subscriber-interface** and **config**>**service**>**sub-if**>**ipv6** hierarchy)

numbered IPv6 only subscriber-interfaces (**config**>**service**>**sub-if**>**ipv6** hierarchy) is not supported

9.3.18.8 Considerations with local DHCP server

When local DHCP server redundancy or synchronization is used, using address-range failover local | remote, in conjunction with PPPoE in multichassis environment, both DHCP servers must be referenced under the corresponding group interface on each node. For **address-range failover access-driven** configurations only one DHCP server must be referenced.

```
subscriber-interface <ip-int-name>
    group-interface <ip-int-name>
        dhcp
            server <local-dhcp-ip-address> <remote-dhcp-ip-address>
```

Otherwise, the PPPoE clients are not synchronized by MCS.

This is not the requirement in the IPoE environment. In the IPoE environment, it is enough that the DHCP server points to the IP address of the local DHCP server. If the IP lease is originally assigned by the peer DHCP server, the request for renewal is automatically forwarded to the remote DHCP server by the virtue of the IP address of the original DHCP server that is included in the renewal request.

It is necessary for the successful renewal of the IP address on the remote DHCP server, that the remote DHCP server has a valid return path back to the gi-address of the forwarder of the renewal request.

9.3.18.9 Redundant interface considerations

In PPPoE dual-chassis environment without the redundant-interface in place, SRRP aware routing should always be used. Otherwise, if the downstream traffic arrives on the standby node (SRRP backup state), it is forwarded directly to the client over the access network (assuming that the access network is operational) with the source MAC address of the group interface (instead of gw-mac). This grp-if MAC address is different from the MAC address (gw-mac) negotiated during the initial PPPoE phase, and therefore, this traffic is dropped by the client. It must be ensured that the downstream traffic is always attracted to the active node (SRRP master state) in the absence of redundancy.

9.3.18.10 IPCP address via DHCPv4 client considerations

When SRRP is in the INIT state on both ends of a multi-chassis redundant setup, such as in the case of directly connected access node failures, the PPPoE session in the BNG does not time out based on the LCP keepalives. When the IPCP address of the PPPoE session is allocated via the internal DHCPv4 client, the PPPoE session is maintained until the DHCPv4 client lease expires. The system deletes the PPPoE session and sends an accounting stop message. No link exists between the router that sends the accounting stop message and the router where the SRRP instance was last active.

9.3.19 Routed Central Office

9.3.19.1 Layer 3 subscriber interfaces

On regular interfaces in an IES or VPRN service, only one SAP can be associated. A group interface allows multiple SAPs to be configured as part of a single interface. All SAPs in a single group interface must be within the same port. Because broadcast is not allowed in this mode, forwarding to the subscriber is based on IP/MAC addresses information gathered by the subscriber management module and stored in the subscriber management table. These entries are based on both static and dynamic DHCP hosts. Routed Central Office (CO) must be used with standard subscriber management or enhanced subscriber management. DSLAMs are typically deployed with Ethernet interfaces.

This model is a combination of two key technologies, subscriber interfaces and group interfaces. While the subscriber interface define the subscriber subnets, the group interfaces are responsible for aggregating the SAPs.

As depicted in Figure 93: Subscriber interface in an IES/VPRN service, an operator can create a new subscriber interface in the IES or VPRN service. A subscriber interface allows for the creation of multiple group interfaces. The IP space is defined by the subnets of the subscriber interface's addresses. Figure 94: Details of a group interface shows the details of group interface A.

Figure 93: Subscriber interface in an IES/VPRN service

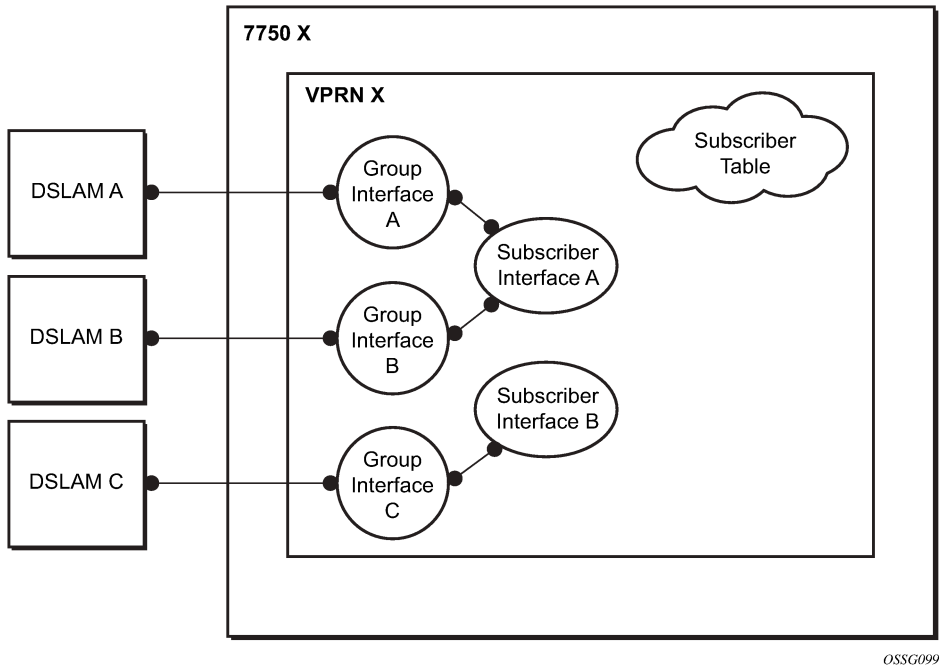


Figure 94: Details of a group interface

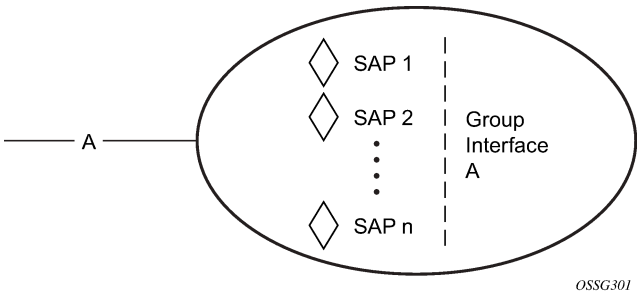
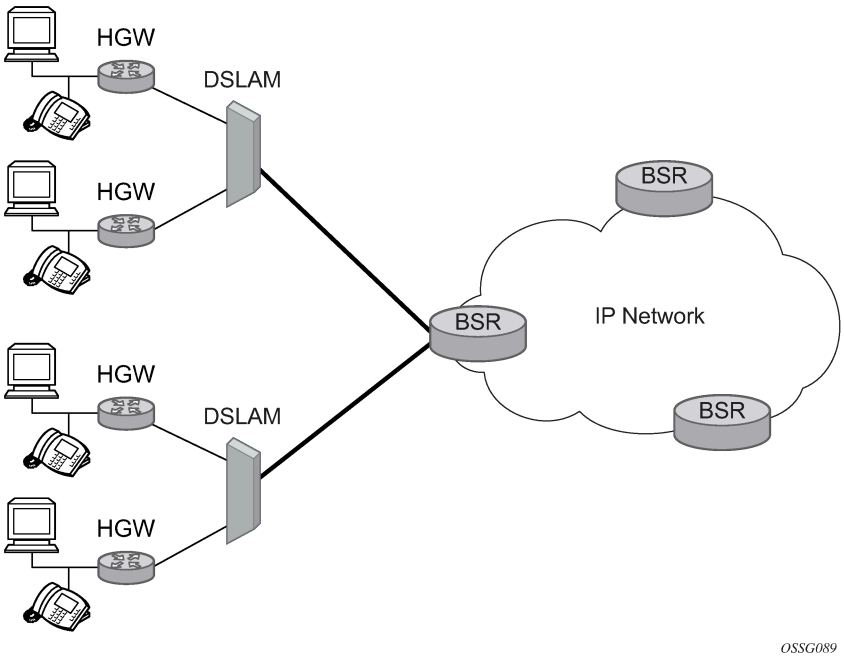


Figure 95: Aggregation network with direct DSLAM-BSR connection shows a network diagram where the DSLAM are connected directly to a Broadband Service Router (BSR) providing access to an IP subnet. Subscribers from multiple DSLAMs can be part of the same subnet. Note that BSR is also referred to as Broadband Network Gateway (BNG).

Figure 95: Aggregation network with direct DSLAM-BSR connection



The BSR can be configured with multiple subnets, allowing subscribers to be part of a single subnet as well as providing mechanisms for re-addressing or expanding existing services without affecting existing users.

Figure 96: Detailed view of configurable objects related to Layer 3 subscriber interfaces

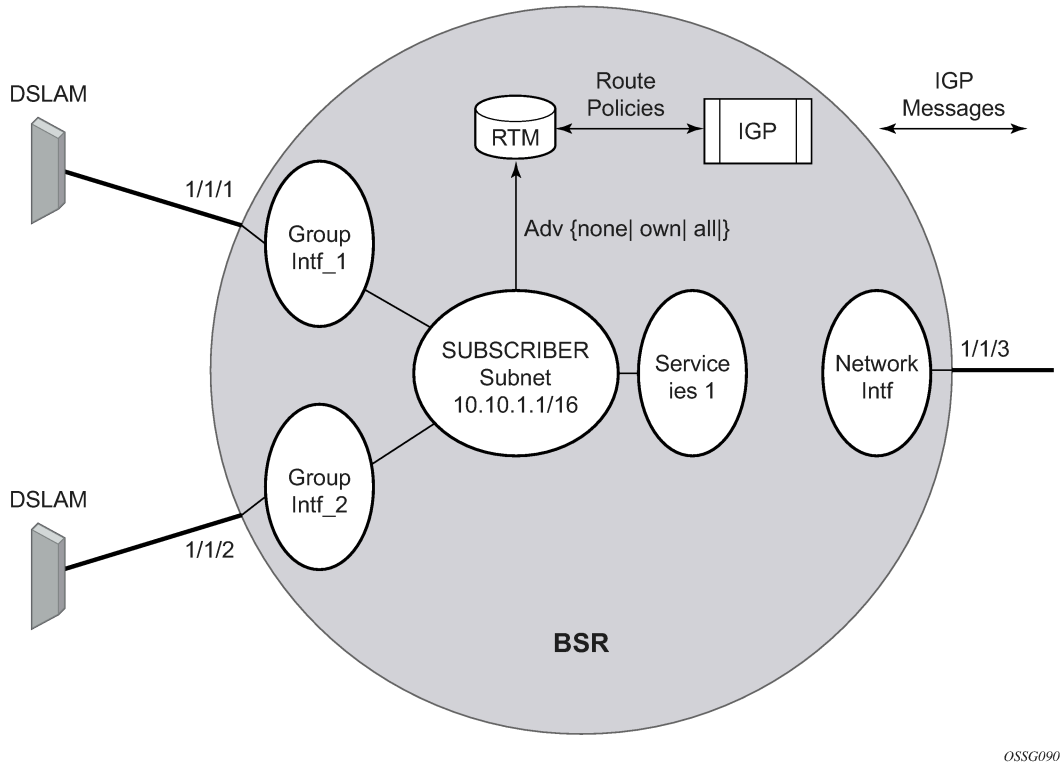


Figure 96: Detailed view of configurable objects related to Layer 3 subscriber interfaces shows a detailed view of a router and the configuration objects implemented to support Layer 3 subscriber interfaces.

- A subscriber service is defined by an IES Service. One or more IES services can be created.
- Each IES service concentrates a number of subscriber-interfaces. The operator can create multiple subscriber interfaces (represented as a subscriber subnet). A subscriber interface defines at least one subnet.
- A group interface is provisioned within the subscriber interface for each DSLAM connected. All group interfaces created under the subscriber interface share the same subnet (or subnets). Group interfaces (shown as intf_1 and intf_2 in Figure 96: Detailed view of configurable objects related to Layer 3 subscriber interfaces) are configured as unnumbered and are associated with the subscriber-interface under which they are configured.
- SAPs can be configured under the group interface. In a VLAN-per-DSLAM model only, one SAP per group interface is needed, while in the VLAN-per-subscriber model, a subscriber of the DSLAM requires its own SAP. All SAPs on a group interface must be on the same physical port or LAG.

The individual features related to subscribers, such as DHCP relay, DHCP snooping and anti-spoofing filters, are enabled at group interface level. For a Routed CO model of subscriber management, and when enhanced subscriber management (if sub-sla-mgmt is configured). Then, hashing is based on an internally assigned subscriber-ID. Having a unique subscriber ID configured in CLI ensures that each subscriber is assigned a unique internal subscriber ID.

It is assumed that individual end-user devices (further referred to as subscriber hosts) get their IP address assigned through either DHCP or static configuration. The management of individual subscriber hosts (such as creation, queue allocation, and so on) is performed by ESM.

The operator can provision how the system advertises routes. While most deployments advertise the full subnet it is possible to have the system advertise only the active, discovered or static host routes.

The distribution of this information into routing protocols is driven by import/export route policies configured by the operator.

9.3.19.1.1 DHCP interactions

The DHCP relay process has been enhanced to record incoming DHCP discover and request messages. Because forwarding to the SAPs is done by the information in the subscriber management table and multiple SAPs are allowed in one interface it was impossible to know which SAP is used to forward the DHCP replies. The node maintains a cache of the DHCP requests. The cache can be viewed using the **tools**>**dump**>**router**>**dhcp**>**group-if-mapping** command. The cache holds an entry for 30 seconds. If an ACK/NAK packet was not received from the server within the timeout the node discards the cache entry. The node can use the Option 82 circuit-id field as part of the temporary host entry. If used, the ACK must contain the same circuit-id field in Option 82 to be found in the cache only if the match-circuit-id is specified at the DHCP level of the group- interface. When the **match-circuit-id** command is enabled a check is performed for option 82 circuit-id.

9.3.19.1.2 Routed CO for IES service

The routed CO model depends on subscriber management to maintain the subscriber host information. To create a group interface the operator must first create a subscriber interface within the service (**config**>**service**>**ies**>**subscriber-interface** *ip-int-name*). The subscriber interface maintains up to 256 subscriber subnets and is configured with a host address for each subnet.

When a DHCP ACK is received the IP address provided to the client is verified to be in one of the subscriber subnets associated with the egress SAP. When DHCP snooping is enabled for regular IES interfaces the same rule applies.

The subscriber interface is an internal loopback interface. The operational state is driven from the child's group interface states and the configuration of an address in the RTM.

The group interface is an unnumbered interface. The interface is operationally up if it is in the **no shutdown** state and if at least one SAP has been defined and is up and the parent subscriber interface is administratively up. The first SAP defined determines the port for the group interface. If the user attempts to define a subsequent SAP that is on a different port results in an error. When the subscriber-interface or the group interface is in a shut down state no packets are delivered or received to or from the subscriber hosts but the subscriber hosts, both dynamic and static, are maintained based on the lease time.

In the routed CO model, the router acts as a DHCP relay agent and also serves as the subscriber-identification agent. The DHCP actions are defined in the group interface. All SAPs in that interface inherit these definitions. The group interface DHCP definition are a template for all SAPs.

Lease-populate is enabled by default with the number-of-entries set to 1. This enables DHCP lease state for each SAP in the group interface.

Because the group interface can aggregate subscribers in different subnets a GI address must be defined for the DHCP relay process. The address must be in one of the host addresses defined for the subscriber interface. The GI address can be defined at the subscriber interface level which causes all child group interface to inherit that route. The GI address can then be overridden at the group interface level. A GI address must be defined in order for DHCP relay to function.

Because of the nature of the group interface, local-proxy-arp, as well as arp-populate, should be enabled. This would allow the system to respond to subscriber ARP requests if the ARP request contains an IP address which is in the same subnet as one of the subscriber interface subnets.

When an authentication policy is specified for a SAP under a group interface, DHCP intercepts DHCP discover messages for RADIUS authentication. If the system is a DHCP-relay defined in a group interface and the GI address was not configured, the operational state of DHCP is down.

9.3.19.1.3 Routed CO for VPRN service

Much like in Routed CO for IES service, the Routed CO model for VPRN depends on subscriber management to maintain the subscriber host information. To create a group interface, the operator must first create a subscriber interface in the **config**>**service**>**vprn** context. The subscriber interface can maintain up to 256 subscriber subnets and can be configured with a host address for each subnet. The host IP address can be installed as a result of both relaying to a DHCP server and proxy to a RADIUS server. In both cases the host IP address must be in the subnet defined by the VPRN's subscriber interface.

Basic subscriber management is allowed only in a subscriber/SAP model and can be used only in a dedicated VPRN architecture. A RADIUS service selection (using Managed SAPs) requires Enhanced Subscriber Management. The subscriber interface's subnets are allowed to be advertised to both IGPs and BGP within a VPRN.

When an authentication policy is specified for a group interface, DHCP snooping must be enabled to intercept DHCP discover and renew messages for RADIUS authentication. Subscriber management RADIUS extensions are allowed if the operator chooses to have the RADIUS server return the subscriber identification, subscriber profile and sla-profile strings using RADIUS.

The node can be defined with both a DHCP relay or proxy function. If the user configures a DHCP relay, the **local-proxy-server** command enables DHCP split leases. In that configuration the node provides the configured DHCP lease to the client using either RADIUS or the real DHCP server as the source of the IP address to be provided.

The RADIUS server can send a Change of Authorization (CoA) message containing the DHCP FORCERENEW VSA which prompts the local-proxy-server to send a FORCERENEW message to the client. The node ACKs when the FORCERENEW messages has been sent, regardless of whether the subscriber responds. If the client fails to respond or if a new session cannot be established because of resource management issues or otherwise the node must respond with a NACK to the RADIUS server.

If the CoA message contains an IP address that is different than the configured IP address (when RADIUS was providing IP addresses) the node must send a FORCERENEW message to the client and NAK the request and provide a new IP address. If the node fails to receive a request, the CoA is ACK'd when the FORCERENEW message has been sent.

The operational state of group and subscriber interfaces are dependent on the state of active SAPs. A group interface can become operationally up only if at least one SAP is configured and is in an operationally up state. A subscriber interface becomes operationally up if at least one group interface is operationally up or the associated wholesale forwarding interface is operationally up. This ensures that, in a failure scenario that affects all group interfaces in a specific subscriber subnet, the node stops advertising the subnet to the network. The SRRP state affects this behavior as well and can cause the subnet to be removed if all group interfaces (and SRRP instances) are in backup state.

9.3.19.2 Wholesale retail Routed CO

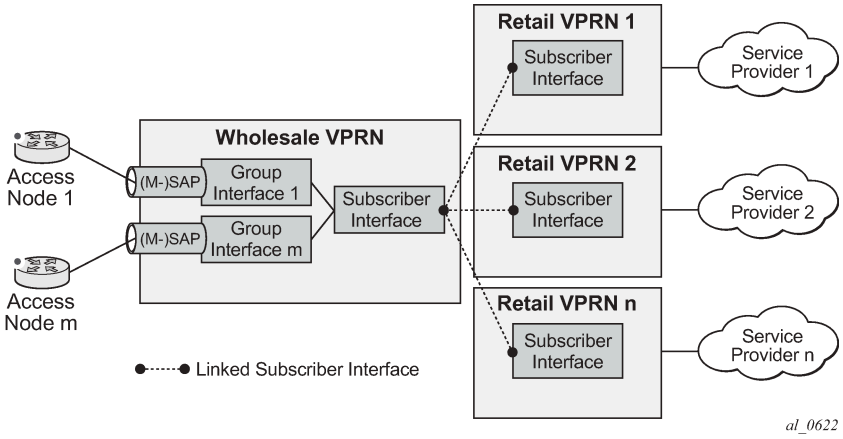
VPRN Routed CO allows a provider to resell wholesaler services (from a carrier) while providing direct DSLAM connectivity. An operator can create a VPRN service for the retailer and configure the access from subscribers as well as to the retailer network. Any further action acts as if the VPRN is a standalone router running the Routed CO model. All forwarding to these servers must be done within the VPRN service. The operator can leak routes from the base routing instance. In this model, the operator can use RADIUS for subscriber host authentication, DHCP relay and DHCP proxy. This provides maximum flexibility to the retailer while minimizing the involvement of the wholesaler. Access cannot be shared among retailers unless a subscriber SAP is used. This requires that the wholesaler maintain a different access node (DSLAM) for each retailer that does not scale well. The wholesale retail model described in this section overcomes these limitations.

9.3.19.2.1 Wholesale retail model

In the wholesale retail model (Figure 97: Wholesale retail model), the wholesaler instance connections that are common to the access nodes are distributed to many retail instances. A subscriber host attached to an access node connected in the wholesaler service can be instantiated in a retail service and obtain IP addresses from the retailers address space. The service context of the retailer is determined during the subscriber host authentication phase (for example, by the Alc-Retail-Serv-Id attribute or the Alc-Retail-Serv-Name attribute in RADIUS or the **retail-service-id** CLI command in the local user database).

Upstream subscriber traffic ingresses into the wholesaler instance and after identification is then forwarded into the retail instance. The reverse occurs for traffic in the downstream direction.

Figure 97: Wholesale retail model



In a wholesale retail model, two subscriber interfaces must be configured and linked together: one in the wholesale VPRN and one in the retail service.

The wholesale subscriber interface defines the IP subnets and host specific configuration parameters for subscriber hosts belonging to the wholesaler. There are associated group interfaces that contain the SAPs which connect to the access nodes.

The retail subscriber interface defines the IP subnets and host specific configuration parameters for subscriber hosts belonging to the retailer. The retail subscriber interface is linked to a wholesale subscriber interface for forwarding by explicit configuration. There are no associated group interfaces.

For example:

```
config>service
        vprn 1000 customer 1 create
            subscriber-interface "sub-int-ws-1" create
            # wholesale subscriber interface
                --- snip ---
                group-interface "group-int-1-1" create
                    --- snip ---
                    sap 1/1/1:1 create
                        --- snip ---
                    exit
                exit
            exit
        exit


       vprn 1001 customer 1 create
            subscriber-interface "sub-int-rt-1" fwd-service 1000    \\
                           fwd-subscriber-interface "sub-int-ws-1" create
            # linked retail subscriber interface
                --- snip ---
            exit
        exit
```

A retail subscriber interface can be linked to a single wholesale subscriber interface and context only. Subscriber interface chaining (linking a retail subscriber interface to another retail subscriber interface) is not supported. Multiple retail subscriber interfaces belonging to different retail contexts can be associated with a single wholesale subscriber interface. When a retail subscriber interface is linked to a wholesale context, all other retail subscriber interfaces from the same retailer must be linked to the same wholesale context.

9.3.19.2.2 Configuration and applicability

As described in the previous section, the wholesale retail model is provisioned with the linking of a subscriber interface in a retail service to a subscriber interface in the wholesale VPRN service.

Because a retail subscriber interface does not have a group interface context, some group interface-specific CLI parameters such as to configure dhcp relay are made available at the retail subscriber interface level. Other CLI parameters such as to provision RADIUS or local user database authentication are configured at the wholesale subscriber or group interface and apply to both wholesale and retail subscriber hosts.

The DHCP lease-populate configuration is special in wholesale retail as it is configured in both wholesale and retail context. The lease-populate value in the **wholesale group-interface dhcp** context controls the per SAP limits while the lease-populate value configured in the retail subscriber interface dhcp context controls the limits for the retailer subscriber interface. Both limits must be satisfied before a new subscriber host can be instantiated.

The sample configurations below enable dual-stack IPoE devices to connect to wholesale service VPRN 4000 and retail service VPRN 4001. Hosts connected in VPRN 4000 get their IP address assigned from RADIUS, therefore the proxy server configuration. Hosts connected in VPRN 4001 get their IP address from a DHCP server, therefore the DHCP relay configuration.

Only the service configurations are shown. They have to be completed with authentication policies and subscriber management configuration such as radius-server-policies, sub- and sla-profiles, and so on.

Sample configuration

Wholesale VPRN service:

```
config>service
        vprn 4000 customer 1 create
            autonomous-system 64500
            route-distinguisher 64500:4000
            auto-bind-tunnel
                resolution-filter
                    ldp
                    rsvp
                exit
                resolution filter
            exit
            vrf-target target:64500:4000
            subscriber-interface "sub-int-1" create
                address 10.10.1.254/24
                address 10.10.2.254/24
                ipv6
                    delegated-prefix-len variable
                    subscriber-prefixes
                        prefix 2001:db8:a:100::/56 wan-host
                        prefix 2001:db8:a001::/48 pd
                    exit
                exit
                group-interface "group-int-1" create
                    ipv6
                        router-advertisements
                            no shutdown
                        exit
                        dhcp6
                            proxy-server
                                no shutdown
                            exit
                        exit
                    exit
                    arp-populate
                    dhcp
                        proxy-server
                            emulated-server 10.10.1.254
                            no shutdown
                        exit
                        lease-populate 100
                        no shutdown
                    exit
                    authentication-policy "auth-policy-1"
                    sap 1/1/4:1201.27 create
                        sub-sla-mgmt
                            sub-ident-policy "sub-ident-1"
                            multi-sub-sap 100
                            no shutdown
                        exit
                    exit
                exit
            exit
            no shutdown
        exit
```

Sample configuration

Retail VPRN service:

```
config>service>
        vprn 4001 customer 1 create
            autonomous-system 64501
            route-distinguisher 64500:4001
            auto-bind-tunnel
                resolution-filter
                    ldp
                    rsvp
                exit
                resolution filter
            exit
            vrf-target target:64500:4001
            interface "int-loopback-1" create
                address 192.0.2.5/32
                ipv6
                    address 2001:db8::5/128
                exit
                loopback
            exit
            subscriber-interface "sub-int-rt-4000-1" fwd-service 4000 fwd-subscriber-
interface "sub-int-1" create
                address 10.10.11.254/24
                address 10.10.12.254/24
                dhcp
                    server 192.0.2.4
                    lease-populate 100
                    gi-address 10.10.11.254
                    no shutdown
                exit
                ipv6
                    subscriber-prefixes
                        prefix 2001:db8:b:100::/56 wan-host
                        prefix 2001:db8:b001::/48 pd
                    exit
                    dhcp6
                        relay
                            source-address 2001:db8::5
                            server 2001:db8::4
                            no shutdown
                        exit
                    exit
                    router-advertisements
                        no shutdown
                    exit
                exit
            exit
            no shutdown
        exit
```

The wholesale retail model applies to all IPoE, PPPoE PTA, IPv4 and IPv6 host types.

The wholesale service type must be VPRN. For IPoEv4 hosts, the retail service type must be a VPRN. For all other host types, the retail service type can be IES or VPRN.

Multicast-per-host replication can be enabled without support for multichassis redundancy.

The wholesale retail model can be deployed in combination with managed SAPs.

Overlapping subscriber subnets and prefixes in retail VPRN services associated with the same wholesale forwarding service are supported for PPPoE (IPv4 and IPv6) and IPoE (IPv4 and IPv6). This support is

enabled by configuring private retail subnets on the retail subscriber interface. Private retail subnets are supported when multichassis redundancy is needed.

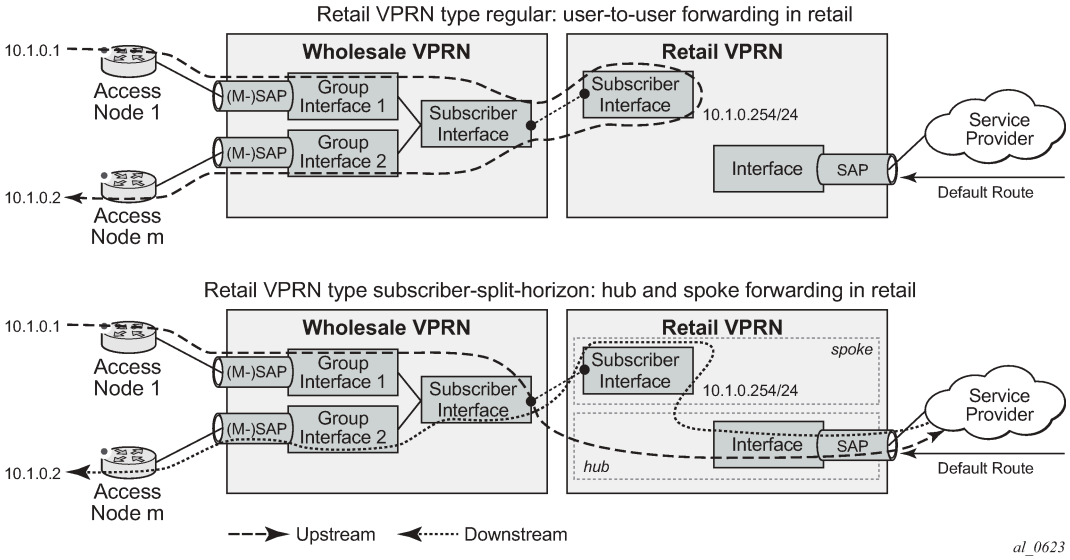
9.3.19.2.3 Hub-and-spoke forwarding

In some cases, hub-and-spoke-type forwarding is necessary for the retailer's VPRN. When the retailer expects all subscriber traffic to reach its router (for accounting, monitoring, wiretapping, and so on) normal best-hop behavior within the retailer VPRN is wanted. Any subscriber-to-subscriber traffic is forwarded within the VPRN preventing the retailer from receiving these packets. To force all subscriber packets to the retailer network, a hub-and-spoke topology is defined: **type subscriber-split-horizon**. It can be used to force all subscriber traffic (upstream) to the retailer's network. The system requires that the operator shut down the VPRN service to enable this flag.

With retail VPRN type configured to **subscriber-split-horizon**, routes learned from MBGP, IGP through a regular interface, static routes through regular interfaces and locally attached regular interface routes are considered hub routes and are used for upstream traffic forwarding. Subscriber subnets cannot be used for upstream traffic forwarding. Downstream traffic uses routes in both hub and spoke routing instances.

Figure 98: Wholesale retail – hub-and-spoke forwarding shows user-to-user traffic forwarding for both retail VPRN types: regular and subscriber-split-horizon.

Figure 98: Wholesale retail – hub-and-spoke forwarding



Hub-and-spoke forwarding can also be used in combination with wholesale unicast RPF (uRPF) check. The uRPF is performed on upstream traffic on spoke routes (subscriber subnets) and the forwarding uses hub routes only.

9.3.19.2.4 Static hosts in wholesale retail

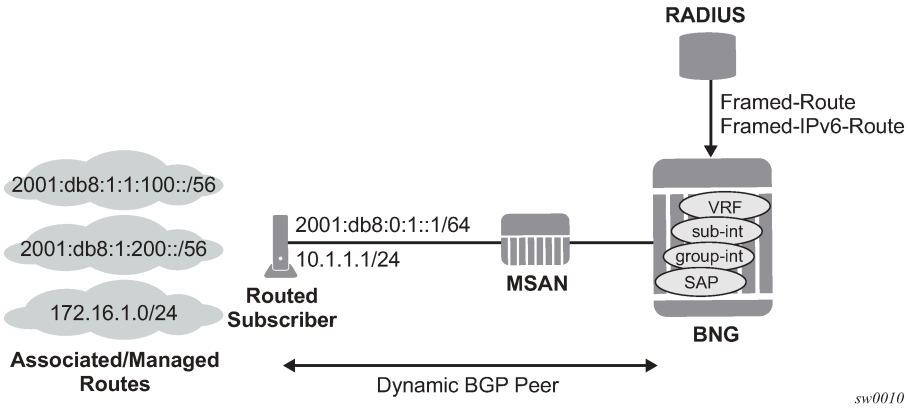
Static hosts configured in wholesale and retail models work as follows:

- In IPv4 static hosts with numbered retailer subscriber interface without private retail subnets or allow unmatching subnets, host addresses are automatically matched with the retailer subnet, and IP routes to hosts are created in the retailer VPRN.
- In IPv4 static hosts with retailer subscriber interface with unnumbered addresses, private retail subnets, or allows unmatching subnets, to create IP routes to hosts in the retailer VPRN, manual configuration of route export from wholesale to retail is required. On wholesale VPRN, only numbered subscriber interfaces are supported.
- In IPv6 static hosts, the **retail-svc-id** must be configured to specify the retailer VPRN.

9.3.19.3 Routed subscriber hosts

A routed subscriber host associated route, as shown in Figure 99: Routed subscriber hosts, is a global routable subnet/prefix behind a routed CPE or Home Gateway. The routed CPE is identified in the BNG as an ESM subscriber host: QoS, accounting and anti-spoofing is enforced per CPE. The associated routes are installed in the BNG route table with next-hop pointing to the routed subscriber host's WAN address.

Figure 99: Routed subscriber hosts



Routed subscriber host associated routes are supported on IES/VPRN subscriber interfaces in a routed CO configuration. To put a SAP or MSAP in routed subscriber mode, the anti-spoof type for the SAP or MSAP must be configured to nh-mac:

```
configure
    service ies/vprn <service-id>
        subscriber-interface <ip-int-name>
            group-interface <ip-int-name>
                sap <sap-id>
                           anti-spoof nh-mac

configure
    subscriber-mgmt
        msap-policy <msap-policy-name>
            ies-vprn-only-sap-parameters
                anti-spoof nh-mac
```

Routes associated with a routed subscriber host (known as managed routes) can be learned in the following ways:

- managed routes configuration for a static host
- the RADIUS or NASREQ authentication [22] Framed-Route and [99] Framed-IPv6-Route attributes
- advertised using an ESM dynamic BGP peer
- learned using a RIP listener neighbor (IPv4 routes only)
- the IPv6 Prefix Delegation prefix as a managed route

9.3.19.3.1 Static configured IPv4 managed route

The routes associated with a static host are populated in the routing table as "Remote Managed" routes. Up to 50 managed routes can be configured for a static host. The following examples show static IPv4 managed route configurations

Example: MD-CLI

```
[ex:/configure service ies subscriber-interface group-interface sap]
[ex:/configure service vprn subscriber-interface group-interface sap]
A:admin@node-2#
    static-host {
        ipv4 10.1.1.20 mac 00:00:00:00:00:00 {
            sub-profile "sub-profile-1"
            sla-profile "sla-profile-1"
            subscriber-id {
                string "static-host-1"
            }
            managed-route 172.20.1.0/24 {
                metric 10
                preference 5
                tag 100
}
            ...
managed-route 172.20.16.0/24 {
                metric 10
                preference 5
                tag 100
}
        }
    }
```

Example: classic CLI

```
config>service>ies>sub-if>grp-if>sap#
config>service>vprn>sub-if>grp-if>sap#
    static-host ip 10.1.1.20 create
        sla-profile "sla-profile-1"
        sub-profile "sub-profile-1"
        subscriber "static-host-1"
        managed-routes
            route-entry 172.20.1.0/24 create
                metric 10
                preference 5
                tag 100
            exit
...
route-entry 172.20.16.0/24 create
                metric 10
                preference 5
                tag 100
```

```
            exit
        exit
        no shutdown
exit
```

Use the following command to display the managed routes associated with a routed subscriber host.

```
show service id service-id static-host detail
```

9.3.19.3.2 Static configured IPv6 managed route

The routes associated with a static host are populated in the routing table as "Remote Managed" routes. Up to 50 managed routes can be configured for a static host. The following examples show static IPv6 managed route configurations.

Example: MD-CLI

```
[ex:/configure service ies subscriber-interface group-interface sap]
[ex:/configure service vprn subscriber-interface group-interface sap]
A:admin@node-2#
    static-host {
        ipv6 2001::1/128 mac 00:00:00:00:00:00 {
            sub-profile "sub-profile-1"
            sla-profile "sla-profile-1"
            subscriber-id {
                string "static-host-1"
            }
            managed-route 2000::/56 {
                metric 10
                preference 5
                tag 100
            }
            ...
            managed-route 3000::/56 {
                metric 10
                preference 5
                tag 100
            }
        }
    }
```

Example: classic CLI

```
config>service>ies>sub-if>grp-if>sap#
config>service>vprn>sub-if>grp-if>sap#
    anti-spoof nh-mac
    static-host ip 2001::1/128 create
        sla-profile "sla-profile-1"
        sub-profile "sub-profile-1"
        subscriber "static-host-1"
        managed-routes
        route-entry 2000::/56 create
            metric 10
            preference 5
            tag 100
        exit
          ...
        route-entry 3000::/56 create
            metric 10
```

```
            preference 5
            tag 100
        exit
exit
no shutdown
exit
```

Use the following command to display the managed routes associated with a routed subscriber host.

```
show service id service-id static-host detail
```

9.3.19.3.3 CPE connectivity check for managed routes

Verify the reachability of managed routes using CPE connectivity checks, which periodically send ICMP pings. If the ping fails a specified number of sequential times, the managed route is withdrawn, or command options of the route (metric, preference, or tag) are changed until the next successful ping.

Unlike the CPE connectivity check for static routes, the CPE connectivity check for managed routes supports the ping destination address within the managed route subnet to check the reachability of a specific address (for example, a LAN interface of the CPE) within the managed route. To avoid circular references between the ping and the managed route, a host route toward the ping destination address is installed in the routing table.

CPE checks for managed routes between IPv4 and IPv6 static hosts

An IPv4 address can be the target of a CPE check for managed routes between an IPv4 static host and an IPv6 static host. However, the IPv4 target address for a managed route on an IPv6 static host can be configured only when the same IPv4 target address is configured on the static IPv4 host on the same SAP.

9.3.19.3.4 ESM dynamic BGP peering

In enterprise IP VPNs, BGP is often used to exchange routing information, for example, between headquarter and branch offices. ESM dynamic BGP peering is needed when a residential access connection provides IP connectivity to the enterprise router.

An ESM dynamic BGP peer setup is automatic when a BGP peering policy attribute is received during RADIUS authentication of a routed subscriber host. The BGP peer is torn down and the associated routes removed from the routing table when the subscriber host is removed from the system (for example, are because of a lease timeout or log out).

For dual-stack routed subscriber sessions, an ESM dynamic BGP peer supports both IPv4 and IPv6 address families to exchange IPv4 and IPv6 routes. The BGP next hop of an IPv4 route received on an ESM dynamic BGPv6 peer should match the IPv4 subscriber host address. The BGP next hop of an IPv6 route received on an ESM dynamic BGPv4 peer should match the IPv6 WAN subscriber host address or an IPv6 address in the IPv6 PD subscriber prefix range.

ESM dynamic BGP peering is supported for routed subscriber hosts terminated in a VPRN service and is not supported for hosts terminated in an IES service. ESM dynamic BGP peering is supported in a wholesale and retail deployment.

The BGP learned routes scaling is limited by the BGP scaling limits. The routes learned by a dynamic BGP peer are populated in the routing table as remote BGP routes.

9.3.19.3.4.1 Configuring ESM dynamic BGPv4 peering

Pre-requisites:

- Enable subscriber management in the VPRN service.
- Configure **anti-spoof nh-mac** in the group interface SAP or MSAP policy, with **urpf-check** optionally configured in the group interface to compensate for IP anti-spoofing not being enabled.

The anti-spoof for SAPs is configured using the following commands.

```
configure service ies subscriber-interface group-interface sap anti-spoof
configure service vprn subscriber-interface group-interface sap anti-spoof
```

The anti-spoof for MSAPs is configured using the following command.

```
configure subscriber-mgmt msap-policy ies-vprn-only-sap-parameters anti-spoof
```

The urpf-check is configured using the following commands.

- **MD-CLI**

```
configure service ies subscriber-interface group-interface ipv4 urpf-check mode
configure service vprn subscriber-interface group-interface ipv4 urpf-check mode
```

- **classic CLI**

```
configure service ies subscriber-interface group-interface urpf-check mode
configure service vprn subscriber-interface group-interface urpf-check mode
```

An ESM dynamic BGPv4 peer is established for a routed subscriber host if the 26.6527.55 Alc-BGP-Policy VSA returned in a RADIUS Access-Accept message contains the name of a local configured BGP peering policy and an ESM dynamic peer group is configured in the VPRN BGP context, as shown in the following classic CLI example.

The following example shows the configuration of a BGP peering policy and ESM dynamic peer group.

Example: MD-CLI

```
[ex:/configure subscriber-mgmt]
A:admin@node-2# info
    bgp-peering-policy "bgpv4-policy-1" {
        local-address 10.3.2.254
        peer-as 65501
        type external
        local-as {
            as-number 65536
        }
    }

[ex:/configure service vprn "submgmt-vprn-2000"]
A:admin@node-2# info
    bgp {
        group "esm-dyn-peer-group-1" {
            admin-state enable
            static-group false
        }
    }
```

Example: classic CLI

```
A:node-2>config>subscr-mgmt# info
    bgp-peering-policy "bgpv4-policy-1" create
        local-address 10.3.2.254
        local-as 65536
        peer-as 65501
        type external
    exit

A:node-2>config>service>vprn$ info
    bgp
        group "esm-dyn-peer-group-1" esm-dynamic-peer
        exit
        no shutdown
    exit
```

The subscriber host IPv4 address is used as the BGP peer IP address.

The local address can be the subscriber interface IPv4 address (single hop BGP peer) or a loopback interface IPv4 address (multihop BGP peer).

Optionally, configure the address families (IPv4, IPv6) to negotiate for the ESM dynamic BGPv4 peer. When not configured, the router includes the AFI/SAFI for the IPv4 address family in the MP-BGP capability of its OPEN message.

See BGP peering command options for more information about how the other BGP peering parameters can be specified and Import and export policies for ESM dynamic BGP peers for more information about route policies.

Use the following commands to verify that an ESM dynamic BGPv4 peer is correctly.

```
show service id ipoe session detail
show service id ppp session detail
```

Output example

```
...
Bgp Peering Policy      : bgpv4-policy-1
Bgp Peer Status         : installed
...
```

Use the following command to verify the state of an ESM dynamic BGPv4 peer.

```
show router bgp summary
```

9.3.19.3.4.2 Configuring ESM dynamic BGPv6 peering

Prerequisites:

- Enable subscriber management in the VPRN service.
- Configure **anti-spoof nh-mac** in the **group-interface sap** or **msap-policy ies-vprn-only-sap-parameters** context, with **ipv6 urpf-check** optionally configured at the **group-interface** to compensate for IP anti-spoofing not being enabled.

The anti-spoof for SAPs is configured using the following commands.

```
configure service ies subscriber-interface group-interface sap anti-spoof
configure service vprn subscriber-interface group-interface sap anti-spoof
```

The anti-spoof for MSAPs is configured using the following command.

```
configure subscriber-mgmt msap-policy ies-vprn-only-sap-parameters anti-spoof
```

The urpf-check is configured using the following commands.

```
configure service ies subscriber-interface group-interface ipv6 urpf-check mode
configure service vprn subscriber-interface group-interface ipv6 urpf-check mode
```

An ESM dynamic BGPv6 peer is established for a routed subscriber host if the 26.6527.208 Alc-BGP-IPv6-Policy VSA returned in a RADIUS Access-Accept message contains the name of a local configured BGP peering policy and an ESM dynamic peer group is configured in the VPRN BGP context, as shown in the following classic CLI example.

The following example shows the configuration of a BGP peering policy and ESM dynamic peer group.

Example: MD-CLI

```
[ex:/configure subscriber-mgmt]
A:admin@node-2# info
    bgp-peering-policy "bgpv6-policy-1" {
        local-address 2001:db8:b002:201::1
        peer-as 65501
        type external
        local-as {
            as-number 65536
        }
    }

[ex:/configure service vprn "submgmt-vprn-2000"]
A:admin@node-2# info
    bgp {
        group "esm-dyn-peer-group-1" {
            admin-state enable
            static-group false
        }
    }
```

Example: classic CLI

```
A:node-2>config>subscr-mgmt# info
    bgp-peering-policy "bgpv6-policy-1" create
        local-address 2001:db8:b002:201::1
        local-as 65536
        peer-as 65501
        type external
    exit

A:node-2>config>subscr-mgmt# info
    bgp
        group "esm-dyn-peer-group-1" esm-dynamic-peer
        exit
        no shutdown
    exit
```

The subscriber host IPv6 WAN address is used as the BGP peer IP address. Both SLAAC and DHCPv6 IA_NA addresses are supported.

For a SLAAC host, the BGP mode on the subscriber side must be active, that is the router at the subscriber premises should initiate the BGP TCP connection, such that the BNG can snoop the TCP SYN and derive the /128 Global Unicast Address of the SLAAC host as the BGP peer address.

ESM dynamic BGP peering is not supported for a DHCPv6 IA_PD host.

The local address can be the subscriber interface IPv6 address (single hop BGP peer) or a loopback interface IPv6 address (multihop BGP peer).

Optionally, configure the address families (IPv4, IPv6) to negotiate for the ESM dynamic BGPv6 peer. When not configured, the router includes the AFI/SAFI for the IPv6 address family in the MP-BGP capability of its OPEN message.

See BGP peering command options for more information about how to configure other BGP peering parameters and Import and export policies for ESM dynamic BGP peers for more information about route policies.

Use the following commands to verify that an ESM dynamic BGPv6 peer is correctly installed.

```
show service id ipoe session detail
show service id ppp session detail
```

Output example

```
...
IPv6 Bgp Peering Policy : bgpv6-policy-1
IPv6 Bgp Peer Status    : installed
...
```

Use the following command to verify the state of an ESM dynamic BGPv6 peer.

```
show router bgp summary
```

9.3.19.3.4.3 BGP peering command options

ESM dynamic BGP peering command options can be specified from multiple sources:

- Use BGP peering command options returned in RADIUS VSAs; see Table 24: T1 dynamic BGP peering RADIUS VSAs and *RADIUS Attributes Reference Guide* for more details.
- If not available from RADIUS, use BGP peering command options configured in the BGP peering policy.
- If not configured in the BGP peering policy, use BGP peering command options configured for the **esm-dynamic-peer** group.
- If not configured in the **esm-dynamic-peer** group, use the BGP peering command options configured in the VPRN service BGP CLI context.
- If not configured in the VPRN service BGP CLI context, use the defaults.

Note: The address family for an ESM dynamic BGP peer can only be configured in the BGP peering policy. When not configured, the IPv4 address family is negotiated for ESM dynamic BGPv4 peers and the IPv6 address family is negotiated for ESM dynamic BGPv6 peers. The

MP-BGP address family configuration at the VPRN service BGP CLI context, and at the VPRN service BGP group CLI context is ignored for ESM dynamic BGP peers.

Table 24: T1 dynamic BGP peering RADIUS VSAs

Attribute-ID	Attribute name	Description
26-6527-55	Alc-BGP-Policy	Mandatory attribute to set up a dynamic BGP peer. References a BGP peering policy configured in the **configure subscriber-mgmt bgp-peering-policy** CLI context.
26-6527-208	Alc-BGP-IPv6-Policy	
26-6527-56	Alc-BGP-Auth-Keychain	Optional attribute reference for a keychain configured in the **configure system security keychain** CLI context.
26-6527-209	Alc-BGP-IPv6-Auth-Keychain	
26-6527-57	Alc-BGP-Auth-Key	Optional attribute for the MD5 authentication key used between BGP peers for BGP session establishment.
26-6527-210	Alc-BGP-IPv6-Auth-Key	
26-6527-58	Alc-BGP-Export-Policy	Optional attribute reference for a pre-configured BGP export routing policy.
26-6527-211	Alc-BGP-IPv6-Export-Policy	
26-6527-59	Alc-BGP-Import-Policy	Optional attribute reference for a pre-configured BGP import routing policy.
26-6527-212	Alc-BGP-IPv6-Import-Policy	
26-6527-60	Alc-BGP-PeerAS	Optional attribute for the Autonomous System number for the remote peer.
26-6527-213	Alc-BGP-IPv6-PeerAS	

9.3.19.3.4.3.1 Configuring per-address family prefix limits

A user can configure an optional prefix limit that specifies the maximum number of BGP routes of the specified address family that can be received from an ESM dynamic BGP peer before an administrative action is taken.

The following example displays prefix limits configuration for the IPv4 and IPv6 address families.

Example: MD-CLI

```
[ex: /configure subscriber-mgmt bgp-peering-policy "bgp-policy-1"]
A:admin@node-2# info
prefix-limits ipv4 {
        maximum 100
        log-only false
        threshold 90
     ## idle-timeout
        post-import false
    }
```

```
    prefix-limits ipv6 {
     maximum 100
        log-only false
        threshold 90
     ## idle-timeout
        post-import false
    }
```

Example: classic CLI

```
A:node-2>config>subscr-mgmt>bgp-prng-plcy# info
prefix-limits ipv4 100 threshold 90 idle-timeout 0
prefix-limits ipv6 100 threshold 90 idle-timeout 0
```

A log event is generated when the number of received routes reaches the threshold value (for example, a percentage of the configured maximum) as shown in the following example

Example

```
2023/05/22 13:46:00.144 UTC MINOR: BGP #2035 vprn2000 Peer 3: 10.3.2.201
"(ASN 65536) VR 3: Group esm-dyn-peer-group-1: Peer 10.3.2.201: number of routes learned
 has exceeded 90 percentage of the configured maximum (100) for ipv4 family"
```

After reaching the maximum number of routes, another log event is generated and by default the BGP session is taken down as shown in the following example.

Example

```
2023/05/22 13:46:00.144 UTC WARNING: BGP #2005 vprn2000 Peer 3: 10.3.2.201
"(ASN 65536) VR 3: Group esm-dyn-peer-group-1: Peer 10.3.2.201: sending notification: code
 CEASE subcode MAX_PFX_RCHD"

2023/05/22 13:46:00.144 UTC WARNING: BGP #2039 vprn2000 Peer 3: 10.3.2.201
"(ASN 65536) VR 3: Group esm-dyn-peer-group-1: Peer 10.3.2.201: moved from higher state
 ESTABLISHED to lower state IDLE due to event MAXPREFIX_EXCEEDED"

2023/05/22 13:46:00.144 UTC MINOR: BGP #2034 vprn2000 Peer 3: 10.3.2.201
"(ASN 65536) VR 3: Group esm-dyn-peer-group-1: Peer 10.3.2.201: number of routes learned
 has exceeded configured maximum (4) for ipv4 family"
```

The BGP session is automatically reestablished after a configured idle-timeout period. With no idle timeout configured, the user must manually reestablish the BGP session. Use the following command to clear the session.

```
clear router bgp neighbor
```

By configuring the **log-only** command option, the BGP session is not taken down after reaching the prefix limit.

By configuring the **post-import** command option, the limit is only applied to the number of routes that are accepted by import policies. Routes rejected by import policies are not counted against the configured limit.

The **prefix-limits** commands for an address family configured in the following context:

```
configure subscriber-mgmt bgp-peering-policy prefix-limits
```

have precedence over the **prefix-limit** commands for the same address family configured for the ESM dynamic peer group in this context:

```
configure service vprn bgp group prefix-limit
```

Use the following command to check the prefix limits that are active on the ESM dynamic BGP peer.

```
show router 2000 bgp neighbor 2001:db8:b002:201::aaa:1 detail
```

Output example

```
...
===============================================================================
Prefix Limits Per Address Family
===============================================================================
Family        Limit       IdleTimeout   TH    LogOnly   PostImport  ExcessInact
-------------------------------------------------------------------------------
ipv4          100         forever       90    Disabled  Disabled    N/A
ipv6          100         forever       90    Disabled  Disabled    N/A
===============================================================================
```

9.3.19.3.4.4 Import and export policies for ESM dynamic BGP peers

The import and export policies used for the ESM dynamic BGP peer are determined in the following priority order:

1. Use import or export policies returned in RADIUS VSAs. These are appended to the policies configured in the **bgp-peering-policy**. A single import and a single export policy can be returned from RADIUS. A maximum of 15 policies can be active per peer. When 15 policies are configured in the **bgp-peering-policy**, the last policy is replaced with the RADIUS returned policy.
2. If not available from RADIUS and not configured in the **bgp-peering-policy**, use the policies configured in the **esm-dynamic-peer** group.
3. If not configured in the **esm-dynamic-peer group**, use the policies configured in the VPRN service BGP CLI context.

To display the BGP learned routes associated with a routed subscriber host, use the BGP show commands; for example: **show router** *router-instance* **bgp neighbor** *ip-address* **received-routes**.

9.3.19.3.4.5 Fast failure detection for ESM dynamic BGP peers using BFD

BGP peer failure detection is by default based on the keep-alive and hold time. For cases where fast failure detection is needed, a BFD session can be used to control the operational state of the BGP peer. Fast failure detection for ESM dynamic BGP peers using BFD is supported for IPoE and PPPoE subscribers. It is not supported for L2TP LNS subscribers.

BFD for ESM dynamic BGP peers is enabled in the **bgp-peering-policy** in classic CLI.

```
config>subscr-mgmt
    bgp-peering-policy "bgpv4-policy-1" create
        bfd-enable
        local-address 10.3.2.254
        local-as 65536
        peer-as 65501
```

```
        type external
     exit
```

BFD for ESM dynamic BGP peers is enabled in the BGP peering policy in MD-CLI.

```
[pr:/configure subscriber-mgmt]
    bgp-peering-policy "bgpv4-policy-1" {
        bfd-liveness true
        local-address 10.3.2.254
        peer-as 65501
        type external
        local-as {
            as-number 65536
        }
    }
```

The parameters for the BFD sessions must be configured on the group interface or retail subscriber interface in classic CLI.

```
config>service>vprn>sub-if>grp-ifconfig>service>vprn>sub-if
    bfd 100 receive 100 multiplier 3
    ipv6
        bfd 100 receive 100 multiplier 3
    exit
```

The parameters for the BFD sessions must be configured on the group interface or retail subscriber interface in MD-CLI.

```
[pr:/configure service vprn "submgmt-vprn-2000" subscriber-interface "sub-int-1" group-
interface "group-int-1-1"]
    ipv4 {
        bfd {
            admin-state enable
            transmit-interval 100
            receive 100
            multiplier 3
        }
    }

    ipv6 {
        bfd {
            admin-state enable
            transmit-interval 100
            receive 100
            multiplier 3
        }
    }
```

The BFD session is always established as a single hop BFD session and therefore fast-failure detection using BFD works for single hop ESM dynamic BGP peers only. The local address for the BGP peer must be a local IPv4 or IPv6 address configured on the subscriber interface.

To verify the state of the BFD session, use the show commands in the following output examples:

```
A:pe2# show router 2000 bfd session
===============================================================================
Legend:
  Session Id = Interface Name | LSP Name | Prefix | RSVP Sess Name | Service Id
  wp = Working path   pp = Protecting path
```

```
===============================================================================
BFD Session
===============================================================================
Session Id                                      State      Tx Pkts    Rx Pkts
  Rem Addr/Info/SdpId:VcId                    Multipl     Tx Intvl   Rx Intvl
  Protocols                                      Type     LAG Port     LAG ID
  Loc Addr                                                           LAG name
-------------------------------------------------------------------------------
group-int-1-1                                    Up      3077793    3077716
  10.3.2.201                                      3          100        100
  bgp                                            iom         N/A        N/A
  10.3.2.254
group-int-1-1                                    Up      2988250    2988469
  2001:db8:b002:201::aaa:1                        3          100        100
  bgp                                            iom         N/A        N/A
  2001:db8:b002:201::1
-------------------------------------------------------------------------------
No. of BFD sessions: 2
===============================================================================

A:pe2# show router 2000 bfd session dest 2001:db8:b002:201::aaa:1 src 2001:db8:b002:201::1
===============================================================================
BFD Session
===============================================================================
Remote Address : 2001:db8:b002:201::aaa:1
Local Address  : 2001:db8:b002:201::1
Admin State    : Up                      Oper State       : Up
Protocols      : bgp
Rx Interval    : 100                     Tx Interval      : 100
Multiplier     : 3                       Echo Interval    : 0
Recd Msgs      : 2988898                 Sent Msgs        : 2988679
Up Time        : 2d 16:46:10             Up Transitions   : 1
Down Time      : None                    Down Transitions : 0
                                         Version Mismatch : 0
Forwarding Information
Local Discr    : 23                      Local State      : Up
Local Diag     : 0 (None)                Local Mode       : Async
Local Min Tx   : 100                     Local Mult       : 3
Last Sent      : 08/23/2021 08:18:43     Local Min Rx     : 100
Type           : iom
Remote Discr   : 19                      Remote State     : Up
Remote Diag    : 0 (None)                Remote Mode      : Async
Remote Min Tx  : 100                     Remote Mult      : 3
Remote C-flag  : 1
Last Recv      : 08/23/2021 08:18:43     Remote Min Rx    : 100
===============================================================================
===============================================================================
```

The following events are generated for a BFD protected ESM dynamic BGP peer.

- ESM dynamic BGP peer established:

```
2602 2021/08/20 15:32:32.609 UTC MINOR: BGP #2019 vprn2000 Peer 2:
2001:db8:b002:201::aaa:1 "(ASN 65501) VR 2: Group esm-dyn-peer-group-1: Peer
2001:db8:b002:201::aaa:1: moved into established state"
```

- BGP added as protocol to track by the BFD session:

```
2603 2021/08/20 15:32:32.610 UTC MINOR: VRTR #2064 vprn2000
2001:db8:b002:201::aaa:1 "The protocol(BGP) using BFD session on node
2001:db8:b002:201::aaa:1 has been added."
```

- BFD session to track the ESM dynamic BGP peer with peer address 2001:db8:b002:201::aaa:1 is up. This indicates that the ESM dynamic BGP peer is tracked by the BFD session and fast failure detection is enabled:

```
2604 2021/08/20 15:32:33.507 UTC MINOR: VRTR #2062 vprn2000
2001:db8:b002:201::aaa:1 "BFD: Local Discriminator 23 BFD session on node
2001:db8:b002:201::aaa:1 is up"
```

9.3.19.3.4.6 Dual homing for ESM dynamic BGP peering

Dual homing for ESM dynamic BGP peering is supported for IPoE DHCPv4 and IPoE DHCPv6 hosts.

Dual homing for ESM dynamic BGP peering is not supported for PPPoE sessions and IPoE data triggered hosts. For PPPoE sessions and IPoE IPv4 data triggered hosts, the BGP peering attributes are discarded when a **multi-chassis sync tag** is configured for the associated SAP. For IPoE IPv6 data triggered hosts, the BGP peering attributes are accepted but unsupported.

The Multi-Chassis Synchronization (MCS) subscriber management (**sub-mgmt**) application synchronizes the BGP peering policy and peering options together with the subscriber host information. BGP-learned routes are not synchronized via MCS.

Each BNG of a redundant pair establishes an independent BGP session toward the CPE. SRRP should not be enabled on the group interface such that traffic to BGP learned routes with a subscriber next-hop can be forwarded on each BNG of the redundant pair. Route advertisement, metrics associated with these routes, and BGP routing policies provide full control of the traffic forwarding over the links between the CPE and the redundant BNG pair.

Because SRRP is not enabled, RADIUS accounting messages are initiated by both BNGs. There is no default gateway IP address that moves to the redundant BNG when a link or BNG fails. The gateway address change must be managed separately in the CPE.

Note: To ensure subscriber prefix matching between the redundant BNG pair for IPv6 SLAAC hosts, both BNGs must use the same static provisioned prefix in the router advertisement.

9.3.19.3.5 RIP listener

If a routed subscriber host is associated with a RIP policy, the host's IPv4 routes can be learned over RIP. The BNG only supports RIP listener and does not support sending RIP routes to subscribers. To enable RIP for a subscriber, the subscriber must first be associated with a **rip-policy**. The group interface of the subscriber must also be configured as a RIP neighbor. The RIP policy can be associated with the subscriber during authentication from LUDB or by RADIUS. It can also be configured directly for static hosts. The RIP routes learned from a subscriber is removed as a subscriber is purged or shut down from the system. RIP listening for ESM host is supported on both IES and VPRN.

To display the RIP learned routes associated with a routed subscriber host, use the RIP commands. For example:

```
show router service-id rip neighbor interface advertised-routes
```

The group interface must be configured in the RIP CLI context of the routed instance where the subscriber host is created:

```
config>router/service vprn>rip
    group “rip-listener”
        neighbor “group-interface-01”
```

The RIP policy is configured in the subscriber-mgmt CLI context:

```
config>sub-mgmt
    rip-policy “rip-policy-01” create
```

A RIP neighbor is established for a subscriber host if the RADIUS attribute [26-6527-207] "Alc-RIP-Policy” is returned in the Access-Accept or in LUDB. RIP parameters such as **authentication key** and **type** can be specified in the RIP policy.

For more information about RIP, see the *7450 ESS, 7750 SR, 7950 XRS, and VSR Unicast Routing Protocols Guide*.

9.3.19.3.6 OSPF support

Subscriber interfaces can be configured as passive or active OSPF/OSPFv3 interfaces. Passive interfaces provide a method to advertise the subscriber interface through OSPF. Use the OSPF active interface to exchange OSPF routes with subscribers.

Subscriber interfaces added to both OSPF and OSPFv3 as broadcast interfaces would act as passive interfaces (meaning the subscriber interfaces do not participate in LSA exchanges). However, when an operator configures a subscriber interface as type point-to-multipoint non-broadcast multiple access (P2MP-NBMA), then the subscriber interface becomes an active OSPF interface for both OSPF and OSPFv3. P2MP NBMA is the most suitable interface type for a subscriber interface for the following reasons:

- P2MP ensures there is no DR or BDR election. No special precautions is required to prevent subscriber CPEs from becoming a DR or BDR. The SR will not rely on a CPE to distribute Network LSAs (LSA type 2) because the LSAs cannot be trusted by the SR and the number of LSAs exchanged may overload the CPE processing power.
- NBMA depicts the subscriber interface accurately. There is no bridging directly between subscribers and all traffic must cross through the subscriber SAP.

In order for a subscriber CPE to form an OSPF adjacency with the subscriber interface, the CPE must program its interface as a OSPF point-to-point interface.

For subscriber interfaces, the manual population of the OSPF NBMA neighbor list is not supported. All neighbors are auto-populated by the SR system via received OSPF Hello messages.

This feature is supported in base and VPRN routing instances.

Note:

- The OSPF MTU for the group interface and the subscriber SAP (port) must be configured with an adequate MTU size to accommodate the actual OSPF specified MTU size.
- The SR learns and populates the OSPF neighbor list via OSPF Hello messages with the following considerations:

  - For ESMv4 hosts, the host address and the OSPF source address must match.
  - For ESMv6 hosts, only DHCPv6 IA_NA and data-trigger hosts are supported. DHCPv6 PD-only hosts and SLAAC hosts are not supported. In addition, OSPFv3 Hello messages use IPv6 Link Local Addresses (LLA) that do not match the host global unique address (GUA). For an IPv6 data-trigger host, a OSPFv3 hello which use a link local address as the source IP must execute the following command to trigger authentication

    config service ies/vprn subscriber-interface/group-interface/data-trigger/accept-ipv6-link-local-address.

- For packets to route to and from OSPF routes announced by the CPE, configure the **anti-spoof** of the SAP to type **nh-mac**.
- ESM hosts and the OSPF applications are independent. Therefore, **clear**, **show**, and **tools** commands need to be applied individually.

Tip: The following log event occurs when a subscriber group interface is added to OSPFv3 and point-to-multipoint non-broadcast multiple access (P2MP-NBMA) is enabled.

```
LCL_RTR_ID <RTR ID>: Neighbor <Neighbor> on <svc name> router state changed to down
 (event KILL)
```

The group interface receives OSPFv3 packets with the host Link Local Addresses (LLA). The OSPFv3 packets are dropped and the OSPF adjacency is not established, if one the following commands is not configured:

```
configure service vprn subscriber-interface group-interface data-trigger accept-ipv6-link-local-address
```

```
configure service ies subscriber-interface group-interface data-trigger accept-ipv6-link-local-address
```

9.3.19.3.7 RADIUS: Framed-Route and Framed-IPv6-Route

RADIUS attribute [22] Framed-Route can be specified in a RADIUS Access-Accept message to associate an IPv4 route with an IPv4 routed subscriber host and Radius attribute [99] Framed-IPv6-Route can be used to associate an IPv6 route with an IPv6 routed subscriber wan host (DHCPv6 IA-NA or SLAAC). These routes are populated in the routing table as "Remote Managed" routes. Up to fifty managed routes can be installed for a routed subscriber host; this corresponds with up to fifty Framed-Routes and fifty Framed-IPv6-Routes for a dual-stack routed subscriber session. Framed-IPv6-Routes cannot be associated with a Prefix Delegation host (DHCP IA-PD).

The Framed-Route and Framed-IPv6-Route attributes should be formatted as:

"<ip-prefix>[/<prefix-length>] <space> <gateway-address> [<space> <metric>] [<space> tag <space> <tag-value>] [<space> pref <space> <preference-value>]"

where:

<space> is a white space or blank character.

<ip-prefix>[/prefix-length] is the managed route to be associated with the routed subscriber host. The prefix-length is optional for an IPv4 managed route. When not specified, a class-full class A,B or C subnet is assumed. The prefix-length is mandatory for an IPv6 managed route.

<gateway-address> must be the routed subscriber host IP address. "0.0.0.0" is automatically interpreted as the host IPv4 address for managed IPv4 routes.

"::" and "0:0:0:0:0:0:0:0" are automatically interpreted as the wan-host IPv6 address for managed IPv6 routes.

[<metric>] Optional. Installed in the routing table as the metric of the managed route. If not specified, metric zero is used. Value = [0 to 65535].

[tag <tag-value>] Optional. The managed route is tagged for use in routing policies. If not specified, or tag-value = 0, then the route is not tagged. Value = [0 to 4294967295].

[pref <preference-value>] Optional. Installed in the routing table as protocol preference for this managed route. If not specified, preference zero is used. Value = [0..255].

If the optional metrics (metric, tag, or preference) are specified in a wrong format or with out of range values, then the defaults are used for all metrics: metric=0, no tag and preference=0. No event is logged.

If the Framed-Route or Framed-IPv6-Route is invalid (for example because the gateway address specified does not match the host wan IP address or because the host bits are not zero) then the routed subscriber host is instantiated without the ill-defined managed route. An event is logged in this case.

Equal Cost Multi-Path (ECMP) is supported for Framed-Route and Framed-IPv6-Route:

The maximum number of equal cost paths in a routing instance is configured with:

```
config>router>
config>service>vprn>
        ecmp <max-ecmp-routes>
```

If an identical managed route is associated with different routed subscriber hosts in the context of the same IES/VPRN service, up to *max-ecmp-routes* managed routes are installed in the routing table. Candidate ECMP Framed-Routes/Framed-IPv6-Routes have:

- Identical prefix
- Equal lowest preference
- Equal lowest metric

A tie breaker determines if more candidate ECMP Framed-Routes/Framed-IPv6-Routes are available than the configured <max-ecmp-routes> is: Lowest ip next-hop.

Other identical managed routes are shadowed and an event is logged.

Note that Candidate ECMP Framed-Routes/Framed-IPv6-Routes can belong to hosts of the same or different subscriber.

Valid Framed-Routes and Framed-IPv6-Routes are persistent (stored in the persistency file for recovery after reboot) and synchronized in a Multi-Chassis Redundancy configuration.

RADIUS-learned Framed-Route/Framed-IPv6-Route and static host associated managed routes that are installed in the routing table can be identified in routing policies for redistribution as protocol "managed".

To display the managed routes associated with a routed subscriber host, use following commands:

show service id *service-id* **dhcp lease-state detail**

show service id *service-id* **dhcp6 lease-state detail**

show service id *service-id* **slaac host detail**

show service id *service-id* **ppp session detail**

show service id *service-id* **pppoe session detail**

show service id *service-id* **arp-host detail**

Valid RADIUS-learned managed routes can be included in RADIUS accounting messages with the following configuration:

```
configure
    subscriber-mgmt
        radius-accounting-policy <name>
            include-radius-attribute
                framed-route
                framed-ipv6-route
```

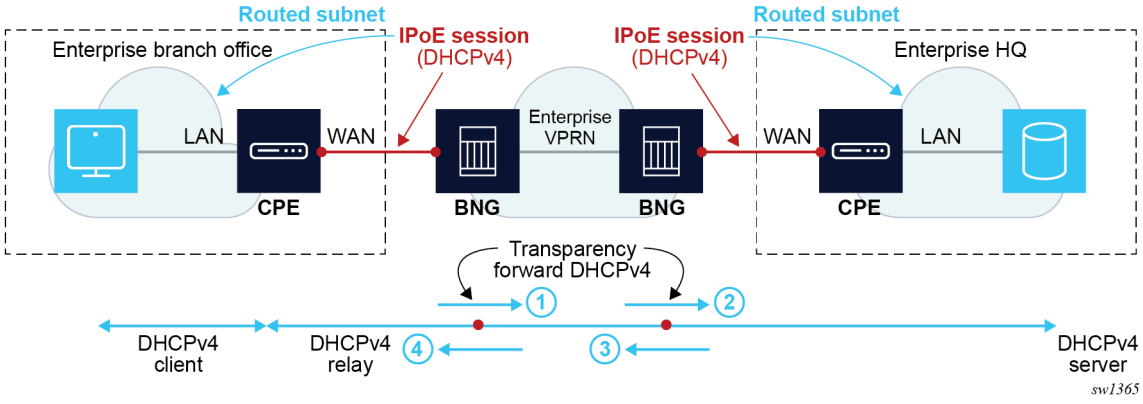
Associated managed routes for an instantiated routed subscriber host are included in RADIUS accounting messages independent of the state of the managed route (Installed, Shadowed, HostInactive, and so on).

For a PPP session, when a Framed-Route or Framed-IPv6-Route is available while the corresponding routed subscriber host is not yet instantiated, the managed route is in the state "notYetInstalled" and is not included in RADIUS accounting messages.

9.3.19.3.8 Transparent forwarding of DHCPv4 packets originated from or destined for routed subnets

In enterprise VPRN services, IPv4 address allocation for enterprise devices can be done via a DHCP server in one of the enterprise sites. For devices in branch offices connected to the network via a residential access network, the DHCP packets should be transparently forwarded over the routed subscriber host instead of triggering subscriber authentication in the BNG. Similarly, if the enterprise DHCPv4 server is in a site connected to the network via a residential access network, the BNG should transparently forward DHCP packets to and from the DHCP server over the routed subscriber host.

Figure 100: Transparent forwarding of DHCPv4 packets originated from or destined for routed subnets



PPPoE

DHCPv4 packets are, by default, transparently forwarded over a routed PPPoE session under the following conditions:

- The enterprise DHCPv4 client subnet is known in the BNG as a managed route with the PPPoE host as next hop.

- The enterprise DHCPv4 client is connected via a CPE that acts as a DHCPv4 Relay Agent.
- The DHCPv4 Relay Agent IP address (giaddr field) inserted by the CPE is part of the managed route. Do not use the PPPoE session IP address as the DHCPv4 Relay Agent IP address.
- Downstream DHCPv4 over PPPoE frames are forwarded via the PPPoE session's egress queues or policers and receives appropriate scheduling priority.
- Do not use a local server as enterprise DHCPv4 server on the router where the PPPoE session is terminated.

IPoE

For routed IPoE sessions or hosts, the transparent forwarding of DHCPv4 packets originated from routed subnets is enabled per routing instance ("Base" router or VPRN service) with the following configuration:

Example: MD-CLI

```
[ex:/configure service vprn "vprn-enterprise-1"]
A:admin@node2# info
    subscriber-mgmt {
        dhcpv4 {
            routed-subnet-transparent-forward true
        }
    }
```

Example: classic CLI

```
A:node-2>config>service>vprn# info
    subscriber-mgmt
        dhcpv4
            routed-subnet-transparent-forward
        exit
    exit
```

Note: In wholesale/retail deployments, the transparent forwarding of DHCPv4 packets for retail subscriber sessions must be configured in the wholesale VPRN instance. Transparent forwarding of DHCPv4 packets is not supported in combination with the **private-retail-subnets** command configured in the retail VPRN instance.

With the **routed-subnet-transparent-forward** command configured in the routing instance, DHCPv4 packets received on a subscriber interface are transparently forwarded over a routed IPoE session or host when the source IP address of the DHCPv4 packet is part of a routed subnet with the IPoE host as next hop.

Ensure the following conditions are met for DHCP clients in enterprise sites connected via a routed IPoE session or host:

- The DHCPv4 client subnet is known in the BNG as a routed subnet with the IPoE host as next hop.
- The DHCPv4 client is connected via a CPE that acts as a DHCP Relay Agent. The DHCP Relay Agent IP address (giaddr field) that is inserted by the CPE, and the source IP address of the relayed packet are part of the routed subnet. The giaddr is used by the DHCP server as a destination IP address of the DHCP Offer, ACK, or NAK packets. At the enterprise DHCP client site, these packets are transparently forwarded by the BNG.

Ensure the following conditions are met for an enterprise DHCP server connected via a routed IPoE session or host:

- The enterprise DHCPv4 server IP address is known in the BNG as a routed subnet with the IPoE host as next hop.
- The enterprise DHCP server uses the server IP address as source IP address.

Supported routed subnet types for transparent DHCPv4 forwarding are:

- RADIUS and NASREQ framed routes
- routes learned via an ESM dynamic BGP peer
- managed routes associated with a static host

Routes learned via a RIP listener neighbor are unsupported for transparent DHCPv4 forwarding.

DHCPv4 packets received on a subscriber interface and that are transparently forwarded must be marked by the CPE to provide adequate QoS treatment. The following applies. The packets:

- are not visible in ingress SAP nor subscriber queues or policers
- can be classified at ingress using the sla-profile sap-ingress QoS policy or ingress IP filter applied on the subscriber session
- cannot be remarked

Transparently forwarded DHCPv4 packets that are transmitted on a subscriber interface have the following characteristics. The packets:

- are forwarded via egress subscriber (M-)SAP queues or policers
- can be classified at egress using the sap-egress QoS policy applied on the subscriber SAP or MSAP
- can be remarked when forwarded via a policer using the sap-egress QoS policy applied on the subscriber (M-)SAP

Self Generated Traffic QoS (sgt-qos) for the DHCP application is not applicable for transparent forwarded DHCP packets.

DHCPv4 packets received on a subscriber interface and that are transparently forwarded are subject to CPU protection, distributed CPU protection and are processed by the DHCP overload protection in the system.

Transparently forwarded DHCPv4 packets are not visible in DHCP **debug** outputs and are not processed by DHCP Python. Per routing instance, statistics count the number of transparent forwarded DHCP Client Packets, Server Packets, or Other Opcode Packets received on a subscriber interface as seen in the following example **show** output.

Use the following command to display statistics information:

```
show router service-name "vprn-enterprise-1" dhcp statistics
```

Output example

```
===============================================================
DHCP Global Statistics (Service: 2501)
===============================================================
Rx Packets                          : 105
Tx Packets                          : 107
Rx Malformed Packets                : 0
Rx Untrusted Packets                : 0
Client Packets Discarded            : 2
Client Packets Relayed              : 2
Client Packets Snooped              : 51
Client Packets Proxied (RADIUS)     : 0
```

```
Client Packets Proxied (Diameter)    : 0
Client Packets Proxied (User-Db)     : 0
Client Packets Proxied (Lease-Split) : 0
Server Packets Discarded             : 0
Server Packets Relayed               : 2
Server Packets Snooped               : 50
DHCP RELEASEs Spoofed                : 0
DHCP FORCERENEWs Spoofed             : 0
Client packets streamed              : 0
Routed Subnet Transparent Forwarded
  Client Packets (BOOTREQUEST)       : 5
  Server Packets (BOOTREPLY)         : 0
  Other Opcode Packets               : 0
===============================================================
```

Note: With transparent forwarding enabled, the system does not check if the DHCP packet is valid and therefore DHCP Packets with Opcode field different from BOOTREQUEST and BOOTREPLY are forwarded.

9.3.19.4 Subscriber prefix leaking

This section describes VPRN leaking and GRT lookup and Routed CO in a VPRN.

9.3.19.4.1 VPRN leaking

Subscriber prefixes and prefix delegation, RADIUS, RIP, and BGP-managed routes with a subscriber prefix as next-hop can be leaked between VPRN services on the same router using MP-BGP import and export policies.

VPRN leaking enables the support of extranet topologies including hub-and-spoke for business services using residential access.

9.3.19.4.2 GRT lookup and Routed CO in a VPRN

GRT lookup allows routing from a VPRN to the GRT, and GRT leaking allows routing from the GRT to a VPRN. These features are particularly useful when VPRNs require routing to the Internet and the GRT already contains the Internet routing table. Wholesale/retail VPRNs and the routed CO VPRN have both GRT lookup and GRT leaking support.

The **config**>**service**>**vprn**>**grt-lookup**>**export-grt** command exports subscriber-related routes and protocols to the GRT. This allows traffic arriving from the network port to be routed downstream to the subscriber. The following configurations are supported in the downstream direction.

- For an IPv4 numbered subscriber interface inside a routed CO VPRN, an IPv4 subscriber subnet can be exported as a policy using the **config**>**service**>**vprn**>**grp-lookup**>**export-grt** command, where the policy is configured with the **config**>**router**>**policy-options**>**policy-statement** command. For an IPv6 numbered subscriber interface inside a routed CO VPRN, an IPv6 subscriber subnet/prefix can be exported as a policy using the **config**>service>vprn>grp-lookup>export-grt command, where the policy is configured with the **config**>**router**>**policy-options**>**policy-statement** command.
- Subscriber-related protocols (managed routes and subscriber management routes) inside a routed CO VPRN can be exported to the GRT.

- For an IPv4 unnumbered subscriber interface inside a routed CO VPRN, subscriber host /32 routes are exported individually to the GRT using the "sub-mgmt" protocol.
- For an IPv6 unnumbered subscriber interface inside a Routed CO VPRN, subscriber host /128 routes and prefixes are exported individually to the GRT using the "sub-mgmt" protocol.
- The retail VPRN supports all items listed above for a Routed CO VPRN (exporting subscriber routes and protocols to the GRT using the **config**>**service**>**vprn**>**grp-lookup**>**export-grt** command). The retail VPRN does not export host routes by default. Therefore, the **export-host-routes** command may be required for a retail VPRN unnumbered subscriber interface.
- The wholesale VPRN supports all items listed above for a Routed CO VPRN (exporting subscriber routes and protocols to the GRT by a route policy). However, retail-related routes that appear in the FDB of the wholesale VPRNs, cannot be exported. Retail-related routes must be exported individually within the VPRN to the GRT. This provides control over which retail VPRNs route to the GRT.

GRT lookup supports traffic from the subscriber to be routed upstream to the GRT. The following configurations are supported in an upstream direction:

- Routed CO VPRN and **grt-lookup**
- wholesale VPRN and **grt-lookup**
- retail VPRN and **grt-lookup**

Not Supported

- SRRP setup in Routed CO VPRN
- SRRP setup in wholesale/retail VPRN

9.3.20 Dual homing

All residential networks are based on two models: Layer 2 CO and Layer 3 CO. Dual homing methods for Layer 2 CO include MC-LAG and MC-Ring. Dual homing for Layer 3 CO is based on SRRP and can be done in ring-topologies (l3-mc-ring or with directly attached nodes. All methods use multichassis synchronization protocol to sync subscriber state.

9.3.20.1 Dual homing to two PEs (redundant-pair nodes) in Triple Play aggregation

Figure 101: Dual homing to two PEs

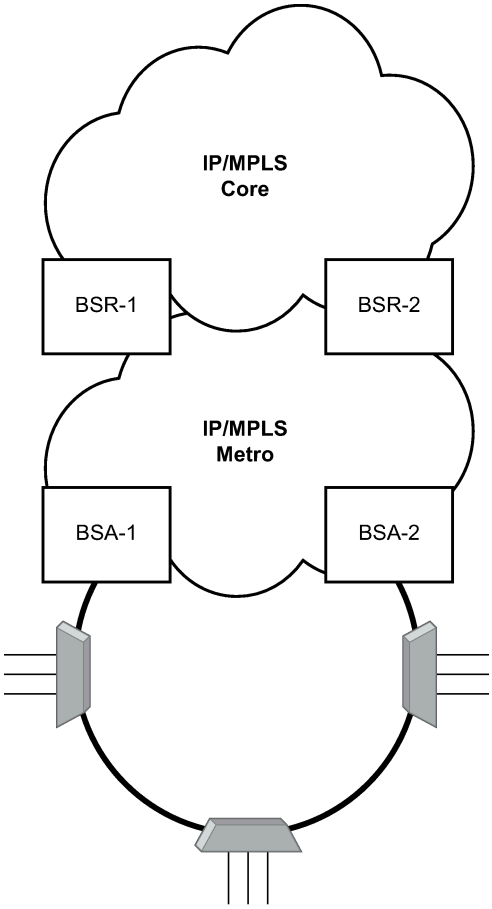


Figure 101: Dual homing to two PEs depicts dual homing to two different PE nodes. The actual architecture can be based on a single DSLAM having two connections to two different PEs (using MC-LAG) or ring of DSLAMs dual-connected to redundant pair of PEs.

Similarly to previous configuration, both aggregation models (VLAN-per-subscriber or VLAN-per-service) are applicable.

Configurations include:

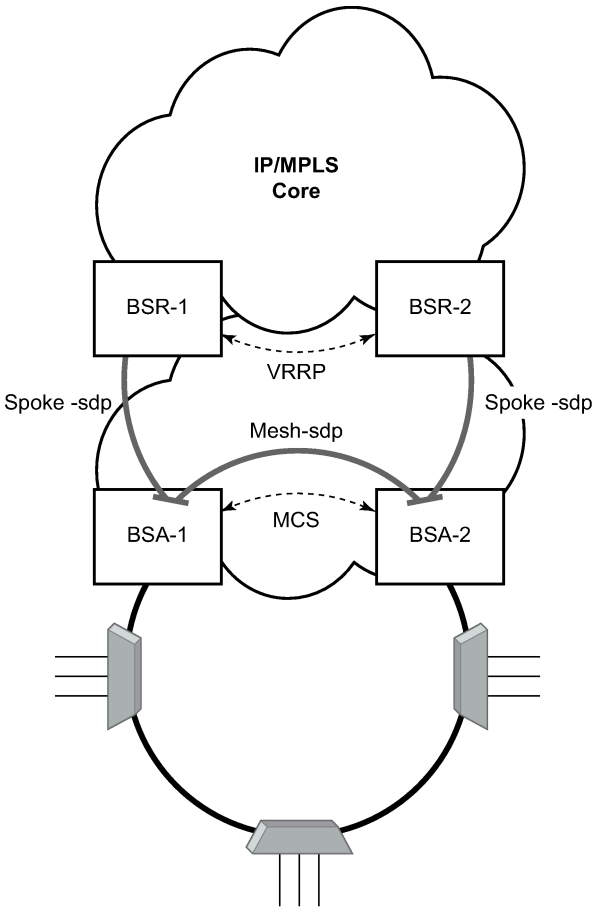
- Loop resolution and failure recovery — Can be based on MC-LAG or mVPLS.
- DHCP-lease-state persistency — Stores all required information to recover from node failure.
- DHCP-lease-state synchronization — A mechanism to synchronize the DHCP lease-state between two PE nodes in the scope of redundancy groups (a group of SAPs used for dual homing).
- IGMP snooping state synchronization — Similarly to DHCP lease-state synchronization, IGMP snooping state is synchronized to ensure fast switchover between PE nodes. In a VPLS network, a BTV stream is typically available in all PE nodes (the ring interconnecting all PEs with Mcast routers is typically used) so the switch over can be purely driven by RSTP or MC-LAG.

- **ARP reply agent responses**

  The ARP reply agent can response to ARP requests addressing a host behind the specific SAP if the SAP is in a forwarding state. This prevents the FDB table in the VPLS from being "poisoned" by ARP responses generated by the node with a SAP in a blocking state (see Figure 102: Layer 2 CO dual homing - network diagram).

Figure 102: Layer 2 CO dual homing - network diagram shows a typical configuration of network model based on Layer 2 CO model. Individual rings of access nodes are aggregated at BSA level in one (or multiple) VPLS services. At higher aggregation levels (the BSR), individual BSAs are connected to Layer 3 interfaces (IES or VPRN) by spoke SDP termination. Every Layer 3 interface at BSR level aggregates all subscribers in one subnet.

Figure 102: Layer 2 CO dual homing - network diagram



Typically, BTV service distribution is implemented in a separate VPLS service with a separate SAP per access-node. This extra VPLS is not explicitly indicated in Figure 102: Layer 2 CO dual homing - network diagram (and subsequent figures) but the descriptions refer to its presence.

From a configuration point of view in this model, it is assumed that all subscriber management features are enabled at the BSA level and that synchronization of the information (using multichassis synchronization) is configured between redundant pair nodes (BSA-1 and BSA-2 shown in Figure 102: Layer 2 CO dual homing - network diagram). The multichassis synchronization connection is used only for synchronizing

active subscriber host database and operates independently from dual-homing connectivity control. At the BSR level, there are no subscriber management features enabled.

The operation of redundancy at the BSR level through VRRP is the same as dual homing based on MC-LAG. The operation of dual homing at BSA level is based on two mechanisms. Ring control connection between two BSAs have two components, in-band and out-of-band communication. With in-band communication, BFD session between BSA-1 and BSA-2 running through the access ring and using dedicated IES/VPRN interface configured on both nodes. This connection uses a separate VLAN throughout the ring. The access nodes provides transparent bridging for this VLAN. The BFD session is used to continuously verify the integrity of the ring and to detect a failure somewhere in the ring.

With out-of-band communication, the communication channel is used by BSA nodes to exchange information about the reachability of individual access nodes as well as basic configurations to verify the consistency of the ring. The configuration information is synchronized through multichassis synchronization and therefore it is mandatory to enable multichassis synchronization between two nodes using the multichassis-ring concept.

In addition, the communication channel used by MC-LAG or MC-APS control protocol is used to exchange some event information. The use of this channel is transparent to the user.

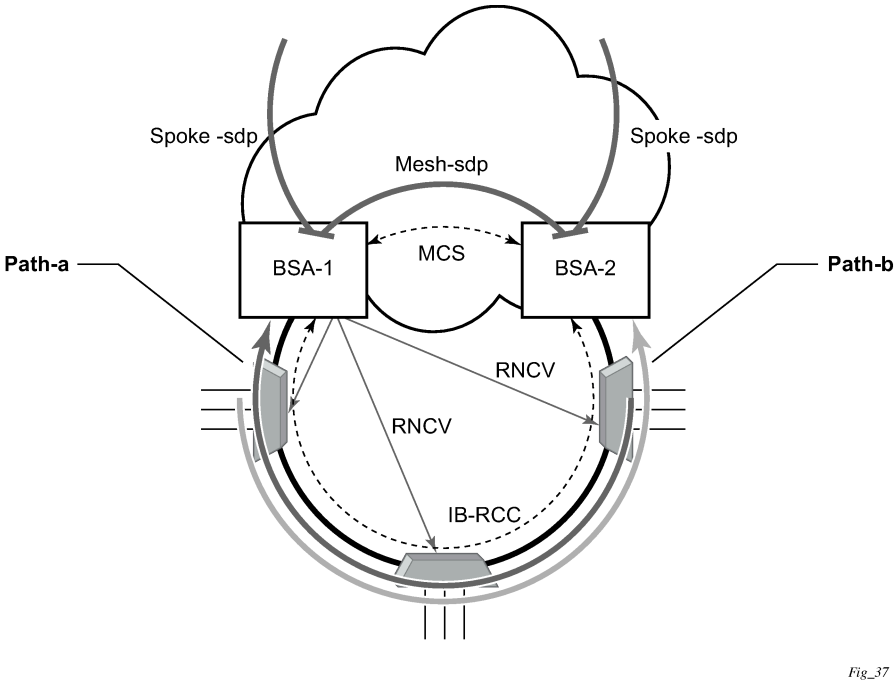
Ring node connectivity check continuously checks the reachability of individual access nodes in the ring. The session carrying the connection is conducted on separate VLAN, typically common for all access nodes. SHCV causes no interoperability problems.

9.3.20.2 Steady-state operation of dual homed ring

Figure 103: Dual homing ring under steady-state condition illustrates the operation of the dual-homed ring. The steady state is achieved when both nodes are configured in a consistent way and the peering relation is up. The multichassis ring must be provisioned consistently between two nodes.

In-Band Ring Control Connection (IB-RCC) is in an operationally UP state. Note that this connection is set up using a bidirectional forwarding session between IP interfaces on BSA-1 and BSA-2.

Figure 103: Dual homing ring under steady-state condition



In Figure 103: Dual homing ring under steady-state condition, the ring is fully closed and every access node has two possible paths toward the VPLS core. Figure 103: Dual homing ring under steady-state condition refers to these as **path-a** and **path-b**. To avoid the loop created by the ring, only one of the paths can be used by any specific ring node for any specified VLAN. The assignment of the individual VLANs to path-a or path-b, respectively, has to be provisioned on both BSAs.

The selection of the primary BSA for both paths is based on the IP address of the interface used for IB-RCC communication (bidirectional forwarding session). The BSA with the lower IP address of the interface used as IB-RCC channel becomes the primary for ring nodes and their respective VLANs assigned to path-a. The primary for path-b is the other BSA.

In this example, each path in the ring has a primary and standby BSA. The functionality of both devices in steady state are as follows:

In the primary BSA:

- All SAPs that belong to the path where the specific BSA is the primary node are operationally UP and all FDB entries of subscriber hosts associated with these SAPs point to their respective SAPs.
- The primary BSA of a path performs periodical Ring Node Connectivity Verification (RNCV) check to all ring nodes.
- In case of a RNCV failure, the respective alarm is raised. The loss of RNCV to the specified ring node does not trigger any switchover action even if the other BSA appears to have the connection to that ring node. If the BFD session is up, the ring is considered closed and the primary or standby behavior is driven solely by provisioning of the individual paths.
- The ARP reply agent replies to ARP requests addressing subscriber hosts for which the BSA is primary.

In the standby BSA:

All SAPs that belong to a BSA's path, the standby is operationally down and all FDB entries of subscriber hosts associated with those SAPs point toward the SDP connecting to the primary BSA (also called a shunt SDP).

In both primary and standby BSAs:

- The information about individual path assignments is exchanged between both BSAs through multichassis synchronization communication channel and conflicting SAPs (being assigned to different paths on both BSA nodes) are forced to path-a (the default behavior).
- For IGMP snooping, the corresponding multichassis IDs are targeting all subscriber-facing SAPs on both nodes. On the standby BSA node, the corresponding SAPs are in an operationally down state to prevent the MC traffic be injected on the ring twice.

9.3.20.3 Broken-ring operation and the transition to this state

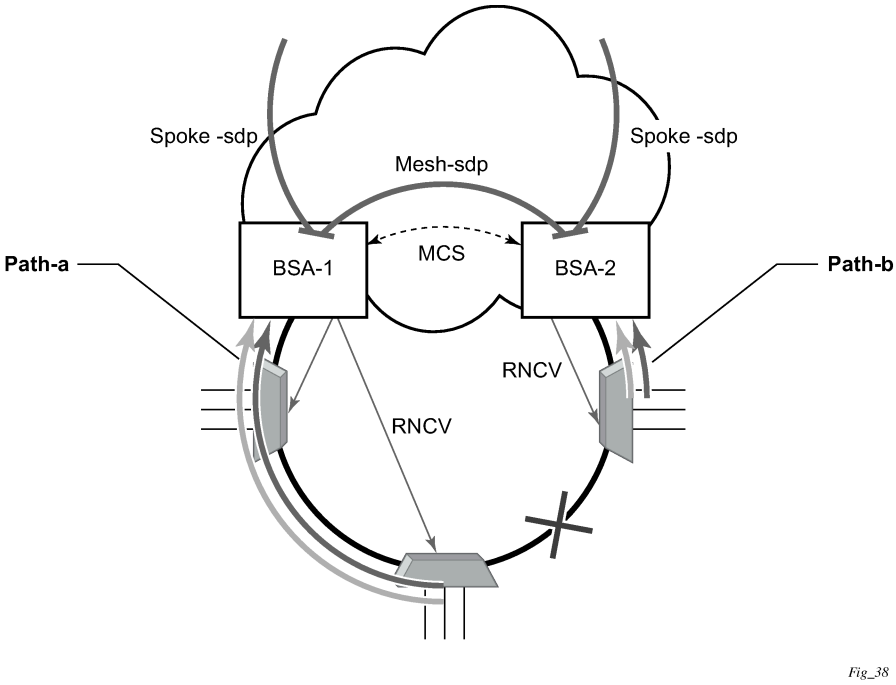
Figure 104: Broken ring state illustrates the model with a broken ring (link failure or ring node failure). This state is reached in following conditions:

- Both nodes are configured similarly.
- Peering is up.
- The multichassis ring is provisioned similarly between two nodes
- IB-RCC is operationally down.

In this scenario, every ring node has only one access path toward the VPLS core and therefore, the Path-a and Path-b notion has no meaning in this situation.

Functionally, both BSAs are now the primary BSA for the reachable ring nodes and act as described in Steady-state operation of dual homed ring. For all hosts behind the unreachable ring nodes, the corresponding subscriber host FDB entries point to the shunt SDP.

Figure 104: Broken ring state



The mapping of individual subscriber hosts into the individual ring nodes is complicated, especially in the VLAN-per-service model where a single SAP can represent all nodes on the ring. In this case, a specified BSA can have subscriber hosts associated with the specified SAP that are behind reachable ring nodes as well as subscriber hosts behind un-reachable ring nodes. This means that the specified SAP cannot be placed in an operationally down state (as in a closed ring state), but rather, selectively re-direct unreachable subscriber states to the shunt SDP.

All SAPs remain in an operationally up state if the ring remains broken. This mainly applies for BTV SAPs that do not have any subscriber hosts associated with and do not belong to any particular ring node.

To make the mapping of the subscriber-hosts on the specified ring node automatically provisioned, the ring node identity is extracted during subscriber authentication process from RADIUS or from a Python script. The subscriber hosts which are mapped to non-existing ring node remain attached to the SAP.

At the time both BSA detect the break in IB-RCC communication (if BFD session goes down) following actions are taken:

- Both nodes trigger a RNCV check toward all ring nodes. The node, which receives the reply first, assumes a primary role and informs the other BSA through an out-of-band channel. This way, the other node can immediately take actions related to the standby role without waiting for an RNCV timeout. Even if the other node receives an RNCV response from the specified ring node later, the primary role remains with the node that received the response first.
- After assuming the primary role for hosts associated with the specified SAPs, the node sends out FDB population messages to ensure that new path toward the VPLS core is established. The FDB population messages are sourced from the MAC address of the default gateway used by all subscriber hosts (such as the VRRP MAC address) which is provisioned at the service level.

9.3.20.4 Transition from broken to closed ring state

By its definition, the multichassis ring operates in a revertive mode. This means that whenever the ring connectivity is restored, the BSA with lower IP address in the IB-RCC communication channel become primary for the path-a and the other way around for path-b.

After restoration of BFD session, the primary role, as described in Steady-state operation of dual homed ring, is assumed by respective BSAs. The FDB tables are updated according to the primary/standby role of the specified BSA and FDB population messages are sent accordingly.

9.3.20.5 Provisioning aspects and error cases

The multichassis ring can operate only if both nodes similarly configured. The peering relation must be configured and both nodes must be reachable at IP level. The multichassis ring with a corresponding sync-tag as a ring-name identifying a local port ID must be provision on both nodes. And the BFD session and corresponding interfaces must be configured consistently.

If the multichassis rings are not provisioned consistently, the ring does not become operational and the SAP managed by it is in an operationally up state on both nodes.

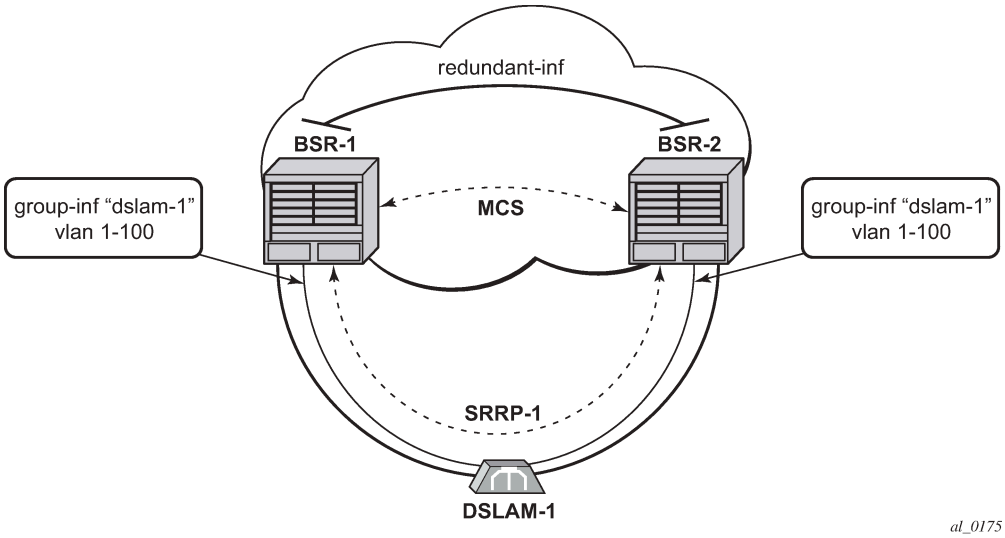
The assignment of individual SAPs to path-a and path-b is controlled by configuration of VLAN ranges according to the following rules:

- By default, all SAPs (and therefore all VLANs on the specified port) are assigned to path-a.
- An explicit statement defining the specified VLAN range assigns all SAPs falling into this range to the path-b.
- An explicit statement defining the specified VLAN range defines all SAPs that are excluded from the multichassis ring control.
- If a conflict in the configuration of VLAN ranges between two redundant nodes is detected, all SAPs falling into the "conflict-range” are assigned to path-a, on both nodes regardless the local configuration.
- For QinQ-encapsulated ports the VLAN range refers to the outer VLAN.

9.3.20.6 Dual homing to two BSR nodes

Figure 105: Low depicts a single DSLAM dual-homed to two BSRs.

Figure 105: Low



To provide dual-homing in the context of subscriber interfaces, the following items must be configured on both BSRs:

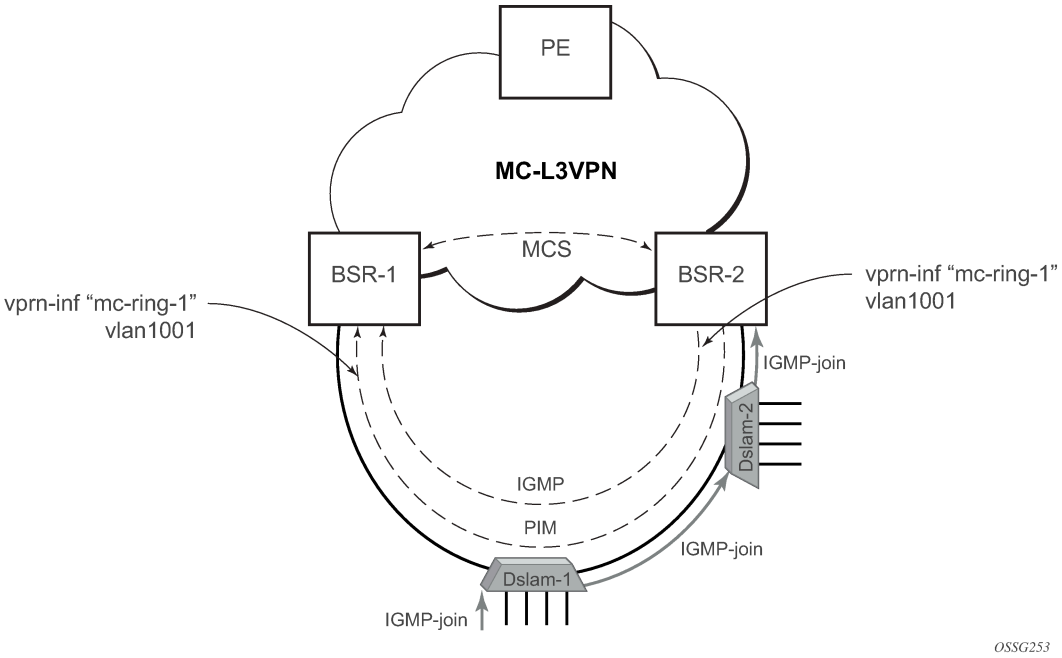
- Group interface (dslam-1) with corresponding SAPs (vlan 1-100)
- SRRP instance controlling a specific group interface
- Redundant interface between BSRs to provide "shunt" connectivity
- MCS connection to provide synchronization of dynamic subscriber-host entries

During the operation, BSR-1 and BSR-2 resolves active/standby relations and populates respective FDBs in such a way that, subscriber-host entries on the active node (SRRP master state) point to a corresponding group interface while subscriber-host entries on the standby node (SRRP backup state) point to the redundant interface. Note that the logical operation of the ring in the Layer 3 CO model is driven by SRRP. For more details on SRRP operation, see the Subscriber Routed Redundancy Protocol chapter.

9.3.20.7 MC services

The typical implementation of MC services at the network level is shown in Figure 106: MC services in a Layer 3-Ring topology (a).

Figure 106: MC services in a Layer 3-Ring topology (a)



The IGMP is used to register joins and leaves of the user. IGMP messaging between BSRs is used to determine which router performs the querier role (BSR2 in Figure 106: MC services in a Layer 3-Ring topology (a)). PIM is used to determine which router is the designated router and the router that sends MC streams on the ring.

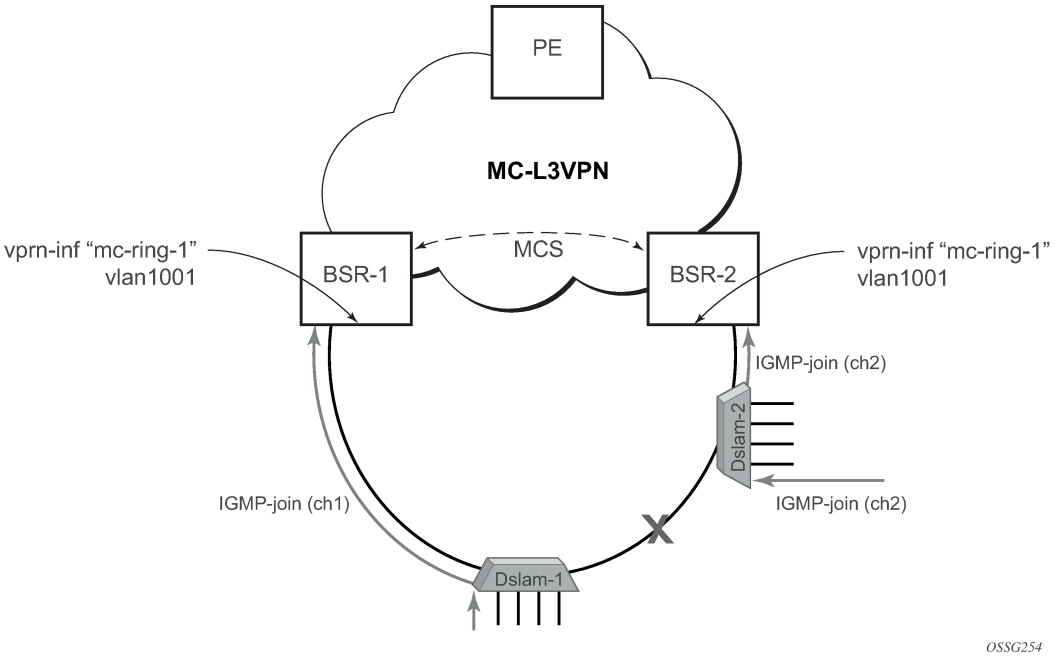
The access nodes have IGMP snooping enabled and from IGMP messaging between BSR, they are aware which router is the querier. In the most generic case, IGMP snooping agents (in access nodes) send the IGMP-joins messages only to IGMP-querier. The synchronization of the IGMP entries can then be performed through MCS. In some cases, access nodes can be configured in such a way that both ring ports are considered as m-router ports and IGMP joins are sent in both directions.

All of the above is a steady state operation which is transparent to the topology used in a Layer 2 domain.

A ring-broken state is shown in Figure 107: MC services on a Layer 3-ring topology (b).

In this case, IGMP and PIM messaging between BSRs is broken and both router assume role of querier and role of designated router. By the virtue of ring topology, both routers see only IGMP joins and leaves generated by host attached to a particular "half" of the ring. This means that both routers have "different" views on the dynamic IGMP state.

Figure 107: MC services on a Layer 3-ring topology (b)



In principle, MCS could be used to synchronize both routers, but in case of a Layer 2 ring, the implementation sends all IGMP messages to the primary BSR which then performs IGMP processing and consequently, MCS sync. As a result, any race conditions are avoided.

Another ring-specific aspect is related to ring healing. The ring continuity check is driven by BFD which then drives SRRP and PIM messaging. BFD is optimized for fast detection of ring-down events while ring-up events are announced more slowly. There is a time window when routers are not aware that the ring is recovered. In the case of MC, this means traffic is duplicated on the ring.

To avoid this, the implementation of BFD provides a "raw mode" which provides visibility on "ring-up" events. The protocols, such as SRRP and PIM, use this raw mode instead of the BFD API.

9.3.20.8 Routed CO dual homing

Routed CO dual homing is a solution that allows seamless failover between nodes for all models of routed CO. In the dual homed environment, only one node forwards downstream traffic to a specific subscriber at a time. Dual homing involves several components:

- **Redundant Interface**

  This is used to shunt traffic to the active node for a specific subscriber for downstream traffic.

- **SRRP**

  This is used to monitor the state of connectivity to the DSLAM. See the SRRP section for more detail.

- **MCS**

  This is used to exchange subscriber host and SRRP information between the dual homed nodes.

Routed CO dual homing can be configured for both wholesaling models. Dual homing is configured by creating a redundant interface that is associated with the protected group interfaces. The failure detection

mechanism can be SRRP. If SRRP is used, each node monitors the SRRP state to determine the priority of its own interface.

Dual homing is used to aggregate a large number of subscribers to support a redundancy mechanism that allows a seamless failover between nodes. Because of the Layer 3 nature of the model, forwarding is performed for the full subscriber subnet.

9.3.20.8.1 Redundant interfaces

In dual homing, a redundant interface must be created. A redundant interface is a Layer 3 spoke SDP-based interface that allows delivery of packets between the two nodes. The redundant interface is required to allow a node with a failed link to deliver packets destined for subscribers behind that link to the redundant node. Because subscriber subnets can span multiple ports it is not possible to stop advertising the subnet, therefore, without this interface the node would black hole.

The redundant interface is associated with one or more group interfaces. An interface in backup state uses the redundant interface to send traffic to the active interface (in the active node). The SAP structure under the group interface must be the same on both nodes as the synchronization of subscriber information is enabled on a group interface basis. Traffic can be forwarded through the redundant interface during normal operation even when there are no failed paths. See Figure 108: Dual homing example.

9.3.20.8.2 SRRP in dual homing

Subscriber Router Redundancy Protocol (SRRP) allows two separate connections to a DSLAM to operate in an active/standby fashion similar to how VRRP interfaces operate. Because the SRRP state is associated with the group interface, multiple group interfaces may be created for a specific port so some of the SAPs are active in one node and others active on the other node. While each SRRP pair is still allowed to be active/backup, the described configuration is allowed for load balancing between the nodes. In a failure scenario, subscriber bandwidth is affected. For more information about SRRP, see the Subscriber Routed Redundancy Protocol chapter.

If SRRP is configured before the redundant interface is up, and in backup state the router forwards packets to the access node using the backup interface but does not use the gateway MAC address. This applies to failures in the redundant interface as well. If the redundant interface exists and up the router sends downstream packets to the redundant interface and not use the backup group interface.

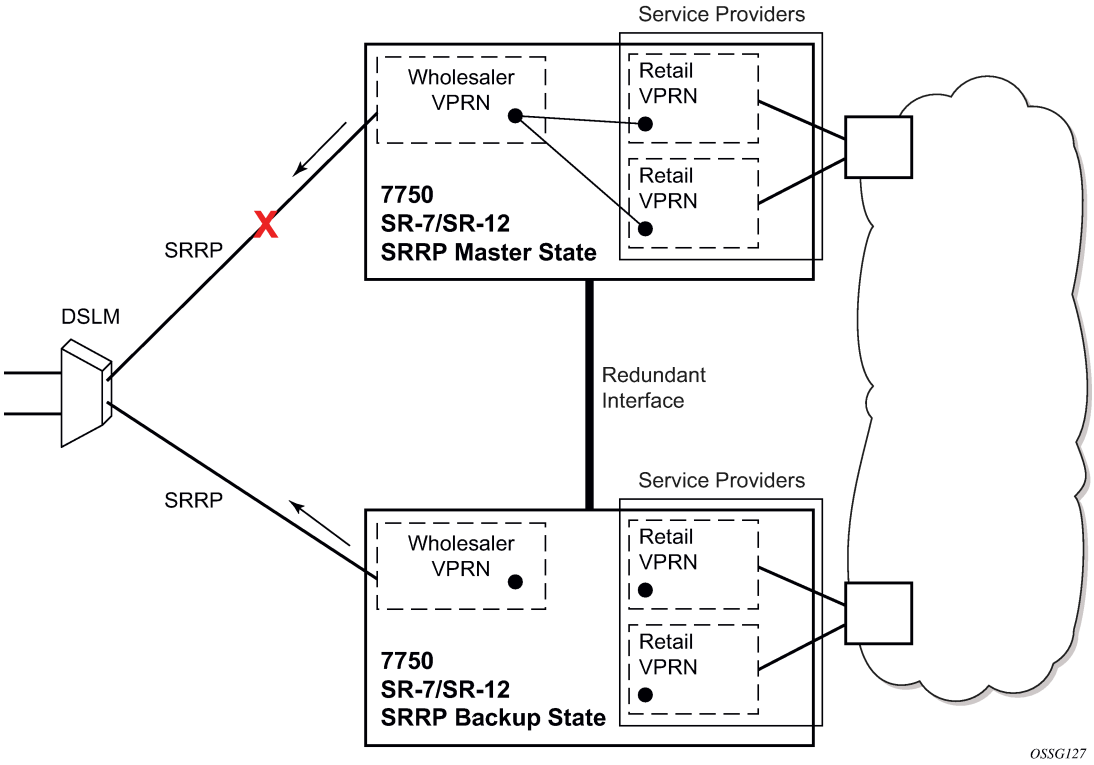
In a dual homing architecture the nodes must be configured with SRRP to support redundant paths to the access node. The nodes must also be configured to synchronize subscriber data and IGMP state. To facilitate data forwarding between the nodes in case some of the ports in a specific subscriber subnet are affected a redundant interface must be created and configured with a spoke. The redundant interface is associated with one or more group interfaces.

The service IDs for both the wholesale VPRN and the retailer VPRN must be the same in both nodes.

An interface in a backup state uses the redundant interface to send traffic to the active interface (in the active node). The SAP structure under the group interface must be the same on both nodes as the synchronization of subscriber information is enabled on a group interface basis.

SRRP is associated a group interface. Multiple group interfaces can be created for a specific port so that some of the SAPs are active in one node and others active on the other node. While every SRRP pair is still allowed to be active or backup the described configuration allows for load balancing between the nodes. In a failure scenario, subscriber bandwidth is affected.

Figure 108: Dual homing example



9.3.20.8.3 Synchronization

To establish subscriber state the nodes must synchronize subscriber information. See the *7450 ESS, 7750 SR, 7950 XRS, and VSR Basic System Configuration Guide* for multichassis synchronization configuration information. The operator must complete the configuration and the system must have data synchronized before the backup node may deliver downstream packets to the subscriber.

If dual homing is used with regular interfaces that run IGMP the nodes must be configured to synchronize the Layer 3 IGMP state.

The service IDs for both the wholesale VPRN and the retailer VPRN must be the same in both nodes.

9.3.20.8.4 Wholesale-retail multichassis redundancy

Multi-chassis redundancy for a retail service is enabled with the SRRP and redundant interface configuration on the wholesale group interface parented by the forwarding subscriber interface. The multichassis state (active or standby) of the retail subscriber host is determined from the SRRP state (master/non-master) of the group interface that parents the SAP of the retail subscriber host. The retail service ID must be equal on both nodes.

Sample wholesale service configuration:

```
vprn 3000 customer 1 create
          description "Wholesale service"
```

```
            route-distinguisher 64500:3000
            auto-bind-tunnel
                resolution-filter
                    ldp
                    rsvp
                exit
                resolution filter
            exit
            vrf-target import target:64500:3000
            redundant-interface "red-int-1" create
                address 192.168.100.0/31
                ip-mtu 1500
                spoke-sdp 12:3000 create
                    no shutdown
            exit
            subscriber-interface "sub-int-1" create
                address 10.1.1.253/24 gw-ip-address 10.1.1.254
                address 10.1.2.253/24 gw-ip-address 10.1.2.254
                group-interface "group-int-1-1" create
                    dhcp
                        --- snip ---
                    exit
                    redundant-interface "red-int-1"
                    sap 1/1/6:1.4001 create
                        description "SRRP 1 message path"
                    exit
                    srrp 1 create
                        message-path 1/1/6:1.4001
                        send-fib-population-packets outer-tag-only
                        no shutdown
                    exit
                    pppoe
                        --- snip ---
                    exit
                exit
                group-interface "group-int-1-2" create
                    dhcp
                        --- snip ---
                    exit
                    redundant-interface "red-int-1"
                    sap 1/1/6:2.4001 create
                        description "SRRP 2 message path"
                    exit
                    srrp 2 create
                        message-path 1/1/6:2.4001
                        priority 50
                        send-fib-population-packets outer-tag-only
                        no shutdown
                    exit
                    pppoe
                        --- snip ---
                    exit
                exit
            exit
            no shutdown
        exit
```

Sample retail service configuration:

```
vprn 3001 customer 1 create
            description "Retail service"
            route-distinguisher 64500:3001
            auto-bind-tunnel
```

```
                resolution-filter
                    ldp
                    rsvp
                exit
                resolution filter
            exit
            vrf-target target:64500:3001
            subscriber-interface "sub-int-rt-3000-1" fwd-service 3000 fwd-subscriber-
interface "sub-int-1" create
                address 10.1.11.253/24 gw-ip-address 10.1.11.254
                address 10.1.12.253/24 gw-ip-address 10.1.12.254
                dhcp
                    --- snip ---
                exit
                pppoe
                    --- snip ---
                exit
            exit
            no shutdown
        exit
```

Retail unnumbered host routes must be leaked in the wholesale service. Retail subscriber subnets and prefixes leaked in the wholesale service are needed to forward downstream shunted traffic over the redundant interface.

The address of an IPv4 unnumbered subscriber host (enabled with **unnumbered** {*ip-int-name* | *ip-address*} or **allow-unmatching-subnets** on the retail subscriber interface) is not contained in the subnets configured on the retail subscriber interface. The export of the IPv4 retail subscriber host routes to the wholesale service must be explicitly enabled with the **export-host-routes** command:

```
        vprn 3001 customer 1 create
            subscriber-interface "sub-int-rt-3000-1" fwd-service 3000 fwd-subscriber-
interface "sub-int-1" create
                allow-unmatching-subnets
                address 10.1.11.253/24 gw-ip-address 10.1.11.254
                address 10.1.12.253/24 gw-ip-address 10.1.12.254
                export-host-routes
                --- snip ---
```

The address of an IPv6 unnumbered subscriber host (enabled with **ipv6 allow-unmatching-prefixes** on the retail subscriber interface) is not contained in the IPv6 prefixes configured on the retail subscriber interface. IPv6 retail subscriber host routes are automatically exported to the wholesale service.

Downstream traffic arriving on a standby node (SRRP backup state) must be shunted over the redundant interface. Downstream traffic shunting can be reduced by advertising the retail subscriber subnets and prefixes from the active node (SRRP master state) with a more favorable metric using routing policies. To make retail subscriber subnets and prefixes SRRP state-aware, they have to be configured to track an SRRP instance that is active on a group interface of the connected wholesale subscriber interface:

```
        vprn 3001 customer 1 create
            subscriber-interface "sub-int-rt-3000-1" fwd-service 3000 fwd-subscriber-
interface "sub-int-1" create
                address 10.1.11.253/24 gw-ip-address 10.1.11.254 track-srrp 1
                address 10.1.12.253/24 gw-ip-address 10.1.12.254 track-srrp 2
                ---snip---
                ipv6
                    subscriber-prefixes
                        prefix 2001:db8:d001::/48 pd track-srrp 1
                        prefix 2001:db8:d002::/48 pd track-srrp 2
```

```
                        prefix 2001:db8:1:100::/56 wan-host track-srrp 1
                        prefix 2001:db8:1:200::/56 wan-host track-srrp 2
                    exit
                exit
```

Multi-chassis redundancy is supported for IPoE (IPv4 and IPv6) and PPPoE (IPv4 and IPv6) retail subscriber hosts and sessions.

Overlapping addresses on retail subscriber interfaces (enabled with **config**>**service**>**vprn**>**subscriber-interface**>**private-retail-subnets**) can be used in combination with multichassis redundancy.

When the **private-retail-subnets** command is enabled, downstream traffic arriving at retail services on a standby node (SRRP backup state) is shunted over to the redundant interface on a wholesale service. On a redundant interface, the service that each frame belongs to is identified by the source MAC address of the frame that includes service ID of a retailer service.

The service ID of each retailer service is synchronized over MCS. Therefore, service IDs for the retailer VPRN must be the same in both nodes.

Traffic shunting in the overlapping address scenario is supported for downstream traffic only.

9.3.20.9 SRRP and multichassis synchronization

To take full advantage of SRRP resiliency and diagnostic capabilities, the SRRP instance is tied to a MCS peering that terminates on the redundant node. After the peering is associated with the SRRP instance, MCS synchronizes the local information about the SRRP instance with the neighbor router. MCS automatically derives the MCS key for the SRRP instance based on the SRRP instance ID. An SRRP instance ID of 1 would appear in the MCS peering database with a MCS-key srrp-0000000001.

The SRRP instance information stored and sent to the neighbor router contains the following:

- SRRP instance MCS key
- service type and ID
- subscriber IP interface name
- subscriber subnet information
- group IP interface information
- SRRP group IP interface redundant IP interface name, IP address and mask
- SRRP advertisement message SAP
- local system IP address (SRRP advertisement message source IP address)
- group IP interface MAC address
- SRRP gateway MAC address
- SRRP instance administration state (up or down)
- SRRP instance operational state (disabled/becoming-backup/backup/becoming-master/ master)
- current SRRP priority
- remote redundant IP interface availability (available or unavailable)
- local receive SRRP advertisement SAP availability (available or unavailable)

9.3.20.10 Dual homing and ANCP

The routers provide a feature related to exchange of control information between DSLAM and BRAS (BSA is described in this model). This exchange of information is implemented by in-band control connection between DSLAM and BSA, also referred to as ANCP connection.

In case of dual homing, two separate connections are set. As a consequence, there is no need to provide synchronization of ANCP state. Instead every node of the redundant-pair obtains this information from the DSLAM and creates corresponding an ANCP state independently.

9.3.21 SRRP enhancement

The SRRP enhancements addressed in this section is to reduce the need for redundant-interface between the pair of redundant nodes without sacrificing the subnet aggregation on the back-end.

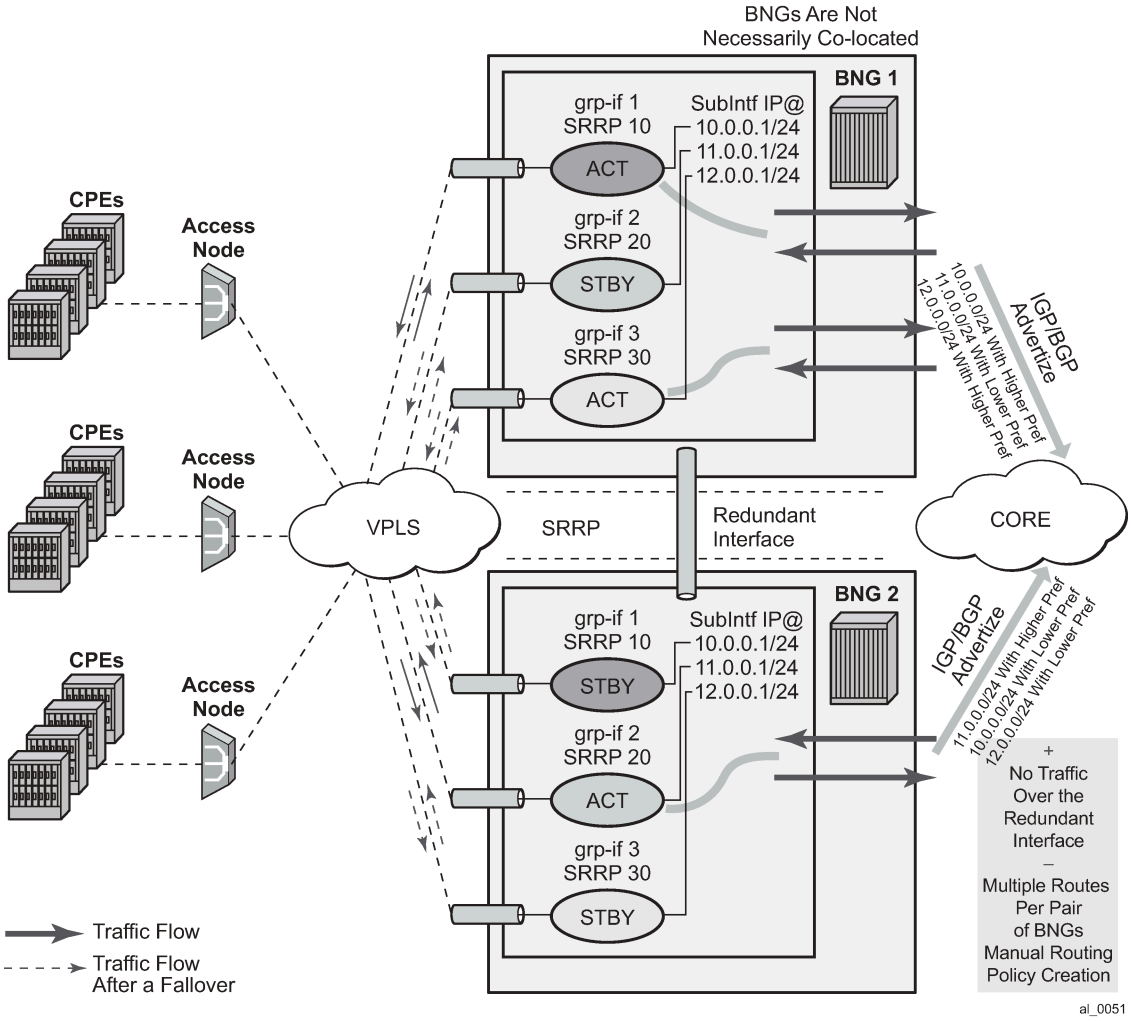
Redundant BNG nodes are not always colocated. This means that the logical link associated with the redundant (shunt) interfaces is taking the uplink path therefore wasting valuable bandwidth (downstream traffic that arrives to the standby (SRRP backup state) node is routed by uplinks for the second time over to the active (SRRP master state) node).

To meet the requirement to reduce the existence of shunted traffic only to the short transitioning period between SRRP switchovers while the routing on the network side is converging, the following was required (referring to Figure 109: IP subnet per SRRP master group):

1. Share IP subnets over multiple SRRP instances. This is not mandatory, but it would help to load balance traffic over the two nodes. For example, IP subnets 10 and 11 can be shared over SRRP instances 10 and 20 on node 1, and the IP subnet 12 can be associated with the SRRP instance 30 on node 2.
2. **SRRP aware routing**

   This allows to dynamically increase routing metric on the IP subnets advertised from the Master SRRP node in comparison to the Standby SRRP node. It also allows to advertise and withdraw routes from a routing protocol based on the SRRP state. In this way, downstream traffic is routed in a predictable manner toward the active node (SRRP master state).
3. SRRP Fate Sharing for SRRP instances 10 and 11. This ensures congruency of SRRP states on the same node. This is a necessary step toward SRRP-aware routing.

Figure 109: IP subnet per SRRP master group

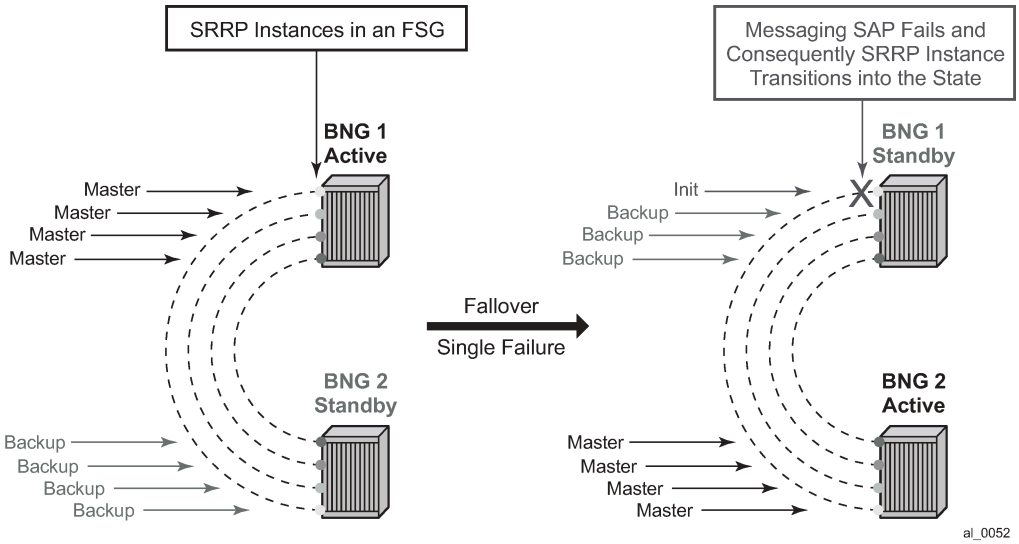


9.3.21.1 SRRP Fate Sharing

SRRP Fate Sharing is a concept in which a group of SRRP instances track a single operational-object composed of SRRP messaging SAPs. The SRRP instances behave as one (in the single failure case) with regards to SRRP state (init/master/backup). The group of SRRP instances that are sharing fate on a paired node are referred as a Fate Sharing Group (FSG).

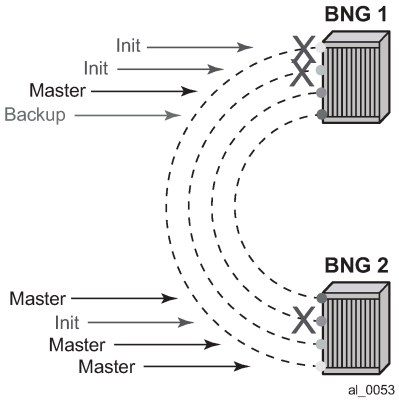
Transition of a single messaging SAP within the FSG into a DOWN state forces the SRRP instance on top of it into the INIT state. Consequently, all other SRRP instances within the same FSG transitions into a Backup state. In other words, SRRP instances within the FSG all share the same fate as the failed SRRP instance as shown in Figure 110: FSG — single network failure. SRRP Fate Sharing provides optimal protection in the context of a single failure in the network.

Figure 110: FSG — single network failure



In the event of multiple network failures, the concept of the FSG breaks as there is a possibility that a 'FSG' contains SRRP instances that are in any of the three possible SRRP states: master, backup, or init. This Fate Sharing feature may not provide optimal protection when there are multiple network failures distributed over both redundant nodes.

Figure 111: Multiple network failures



The whereabouts of the failure in the network path that SRRP is designed to monitor are not always clearly reflected through SRRP states. For example, if the network failure is somewhere in the aggregation network beyond the direct reach of our BNG, SRRP instances on both BNG nodes can reach the SRRP master state. This is a faulty condition and the reason why solely monitoring of the SRRP states is not enough to protect against failures. On the other hand, the SRRP messaging SAP states are more indicative of the network failure because they can be tied into Eth-OAM.

After a single network failure is detected and as a result an SRRP instance transitions into a non-master state, the remaining SRRP instances in the FSG are forced into a backup state. This is achieved by changing the priority of each individual SRRP instance in the FSG.

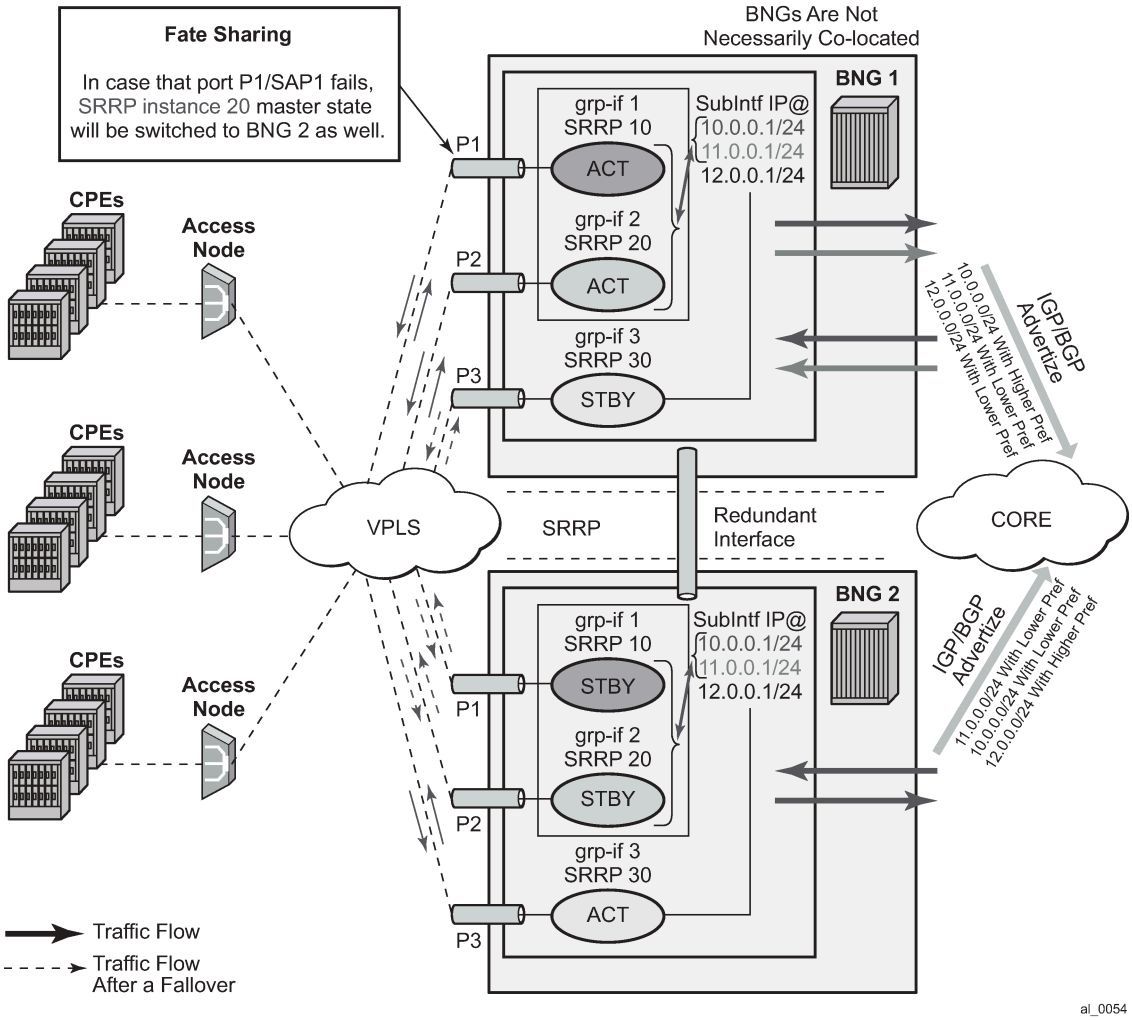
When there are simultaneous multiple failures (multiple ports fail at the same time), it is possible that the SRRP instances within the FSG settle in any of the three possible SRRP states: Master, Backup, or Init. In such scenario, shunted traffic ensues.

In the premise of SRRP Fate Sharing, the network failure is reflected in the operational state of the messaging SAP over which SRRP runs. This is the case if the failure is localized to the BNG (somewhere on the directly connected link). In the case of non-localized failure (beyond the direct reach of the BNG node), Eth-OAM may be needed in to detect the remote end failure and consequently bring the SAP operationally into a DOWN state.

After the single network failure is detected, all instance within the FSG transitions into an SRRP non-Master state.

If there are no failures in the network, all SAPs are UP and SRRP instances within the FSG are in a homogeneous and deterministic state based on their configured priorities.

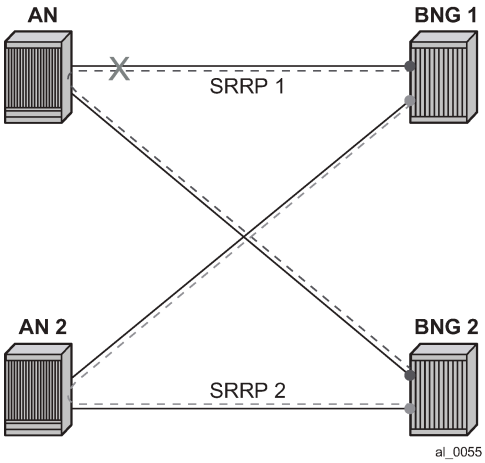
Figure 112: SRRP Fate Sharing



Failure Detection in a Fate Sharing Group

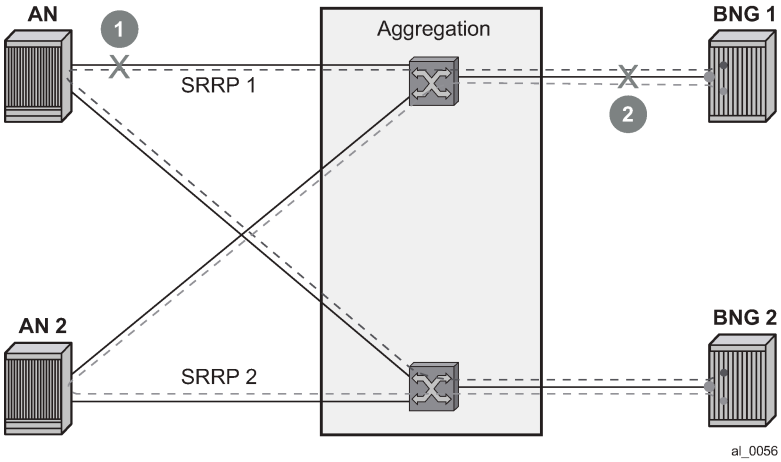
- Dual homing over directly connected ports. No Eth-OAM is needed, AN is directly connected to the BNG.

  Figure 113: Scenario 1



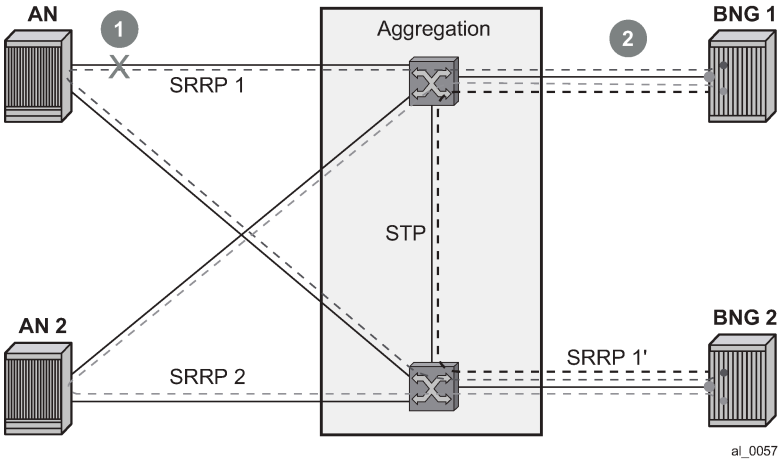
- Dual homing with aggregation network - aggregation network has no redundancy between Layer 2 switches (STP). To determine whereabouts of failure at point 1 in Figure 114: Scenario 2, Eth-OAM is needed.

  Figure 114: Scenario 2



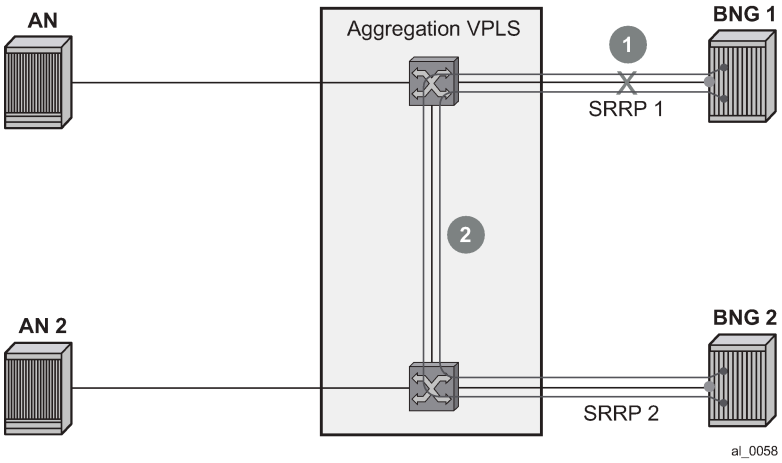
- Dual homing with aggregation network - aggregation network with redundancy between Layer 2 switches (STP). No Eth-OAM is needed in this case for successful operation. However, the failure detection is based on the failure of the directly attached ports.

Figure 115: Scenario 3



- Single homing with aggregation network. In this case, SRRP can protect only against direct failures. Any remote failure leaves a part of the network isolated from the subscriber point of view.

Figure 116: Scenario 4



9.3.21.2 Fate sharing algorithm

Fate Sharing Group (FSG) is relaying on tracking the state of messaging SAPs over which SRRP instances run. An SRRP instance with the messaging SAP operationally DOWN transitions into the Init state.

The transitioning of any messaging SAP in a FSG into an UP/DOWN state triggers SRRP priority adjustment within the FSG. The SRRP priorities should be chosen carefully to achieve the wanted behavior. They are modified dynamically as the SAP states change. The range in which SRRP priorities can be modified is from 1 to the SRRP priority that is initially configured under the SRRP node. Here are some general guidelines for choosing SRRP priorities in a FSG:

- Initially configured SRRP priorities for all SRRP instance within the FSG within the node should be the same.
- Initially configured SRRP priorities should be different between pairing FSGs. For example, SRRP instances in the BNG node A within an FSG all have the same SRRP priority 'X', while corresponding SRRP instances on the paired node within corresponding FSG all have SRRP priority 'Y'. This ensures that the SRRP master state is clearly defined between the two BNG nodes. This step is not mandatory as SRRP naturally breaks the master state election tie in the case that all SRRP priorities are the same. However, following this step may provide a clearer view from an operational perspective.
- The **priority-step** parameter used for dynamic SRRP priority adjustment must be greater than the difference in initially configured SRRP priorities between two BNG nodes. This ensures that a single failure event triggers the SRRP switchover. Otherwise, if the dynamically lowered SRRP priority is still greater than the one from the SRRP peer, the switchover would not be triggered. Therefore, the fate sharing concept would not function as intended.
- Initially configured SRRP priority of each SRRP instance should be greater than the (anticipated) number of SRRP instances in a FSG multiplied by the SRRP **priority-step**. This ensures that the dynamically priority never tries to go below 1. There is a code check that prevents SRRP priority going below 1. Nonetheless, it is recommended not to get into a situation where this needs to be enforced in the code.

The priorities can never be less than 1 or greater than initially configured SRRP priority.

Example scenarios:

Assume 3 SRRP instances in a FSG. The SRRP instances in the FSG-1 on BNG 1 have the priority of 100, while the SRRP instances in the FSG-2 on BNG 2 have the priority of 95. The **priority-step** is 6. The SRRP instances and underlying messaging SAPs are referred to as SRRP 1, 2, 3 and SAP 1,2,3, respectively.

Initialization:

Scenario 1 – all SAPs are operationally UP.

BNG 1 boots up and all messaging SAPs transition into the UP state. When the first SRRP instance in FSG-1 comes up, it looks under the FSG to finds out how many messaging SAPs are operationally UP. Because all messaging SAPs are operationally UP, this first SRRP instance assumes its initially configured priority of 100. The other two SRRP instances in the same FSG follows the same sequence of events.

BNG 2 follows the same flow of events. As a result, all SRRP instances within the corresponding FSG are in the SRRP master state on BNG 1.

Scenario 2 – messaging SAP 1 is operationally DOWN on BNG 1, the rest of the messaging SAPs are operationally UP.

SRRP 2 and 3, during the initialization, pick up SRRP priority of 94 (100 – 1*priority-step).

On BNG 2, all messaging SAPs are UP and consequently all SRRP instances within the FSG on BNG 2 have SRRP priority of 95. The SRRP instances are in the SRRP master state on BNG 2.

Scenario 3 – Continuing from scenario 2, the SAP 1 on BNG 1 transitions into the UP state. SRRP priority of each SRRP instance in FSG-1 is increased by 6, bringing it to 100, enough to assume Mastership.

Adding a New Instance into an FSG

To introduce minimal network disruption, first create messaging SAPs in both BNG nodes and ensure that both SAPs are operationally UP. Then a new SRRP 4 instance should be created on both BNG nodes. The next step would be to include this new messaging SAP into a SAP monitoring group. And finally, the SRRP-4 is added into the FSG (1 and 2). This triggers the recalculation of SRRP priorities for the existing

FSG-1 and FSG-2. Because all SRRP priorities are at the maximum (initially configured priority), nothing changes.

There are more disruptive ways of adding an SRRP instance into a FSG. One such example would be in the case where SRRP priorities are not at their maximum (initially configured) priority. If an SRRP instance is first added into an FSG that is in a backup state, this would increase the FSG priority and potentially cause a switchover. If the SRRP instances is then added in a FSG on the peer BNG (previously SRRP master state), the priority of this FSG would be increased again and the switchover would unnecessarily occur for the second time. The new SRRP instances, when operational, should always be added in the FSG with SRRP master state first.

SRRP priority re-calculation within the FSG is triggered by the following events:

- SRRP initialization
- addition of a SAP under the monitoring group
- messaging SAP failure

This priority calculation looks into how many SAPs are in the DOWN state within the monitored SAP group. Based on this number, the priority is calculated as follows:

SRRP priority = configured-priority – priority-step * num_down_SAPs.

9.3.21.3 SRRP aware routing - IPv4/IPv6 route advertisement based on SRRP state

There are three cases with its own specifics:

- subscriber interface routes (IPv4/IPv6)
- managed routes
- subscriber management Routes (/32 IPv4 hosts routes and IPv6 PD wan-host routes)

Depending on the route type, the action is to either modify the route metric based on the SRRP state that the route is tracking, or to advertise/withdraw the route based on the SRRP state that the route is tracking. The action is defined in the routing policy and it is based on the new attributes with which the routes are associated.

To achieve a better granularity of the routes that are advertised, an origin attribute is added to the subscriber management routes (/32 IPv4 routes and IPv6 PD wan-host) with three possible values:

aaa

IPv4

subscriber-management /32 host routes that are originated through RADIUS framed-ip-address VSA other than 255.255.255.254. The 255.255.255.254 returned by the RADIUS indicates that the BNG (NAS) should assign an IP address from its own pool.

IPv6

subscriber-management routes that are originated through framed-ipv6-prefix (SLAAC), delegated-ipv6-prefix (IA_PD) or alc-ipv6-address (IA_NA) RADIUS attributes. This is valid for IPoE and PPPoE type host.

dynamic

IPv4

subscriber-management /32 host routes that are originated through the DHCP server (local or remote) and also RADIUS framed-ip-address=255.255.255.254 (RFC 2865).

IPv6

subscriber-management routes that are assigned through the local DHCPv6 server pools whose name is obtained through Alc-Delegated-IPv6-Pool (PD pool) and Framed-IPv6-Pool (NA pool) RADIUS attributes. This is valid for IPoE and PPPoE type hosts.

In addition, for IPoEv6 only, the pool name can be also obtained through the ipv6-delegated-prefix-pool (PD pool) and ipv6-wan-address-pool (NA pool) from LUDB.

static

IPv4

subscriber-management /32 host routes that are originated through LUDB. This also covers RADIUS fallback category (RADIUS falls back to system-defaults or to LUDB).

IPv6

subscriber-management routes obtained from LUDB through the ipv6-address (IA_NA) or ipv6-prefix (IA_PD). This is supported only for IPoE.

Overall, the following new route attribute is added:

state: srrp-master, srrp-non-master

The existing origin attribute is expanded to contain the following values:

origin: aaa, dynamic, static

These two attribute types are applied:

The state attribute is applied to all three route types: subscriber interface routes, managed routes and subscriber management routes. Each route listens to the SRRP state.

If an attribute is defined in the routing policy as a match condition (from statement) but the route itself does not have this attribute, the route is evaluated into a non-match condition.

The origin attribute is always applied only to subscriber management routes. No additional statement is needed to explicitly apply this attribute as it may be the case for the state attribute.

Every time there is a change in the attribute associated with the route, the route is re-evaluated in the RTM by the routing policy and corresponding action is taken.

9.3.21.3.1 Subscriber interface routes (IPv4 and IPv6)

Optimized routing and elimination of downstream shunt traffic during normal operation can be achieved by statically favoring the routes on the network side that are advertised with an increased metric by active nodes (SRRP master state).

The downside of this static approach is that during the port or card failure and consequently a SRRP switchover, the node with the failed port or card continues to advertise routes with the same high metric if the subscriber interface is in the 'UP' state (or a single SAP under it). That is, the network side is not aware of the switchover. It continues to forward traffic to the standby node, and as a result, heavy shunt traffic ensues. To effectively deal with this, the network side must be aware of the routing change that occurred in the access layer.

When failure is detected, the metric for the route is changed automatically based on the following configuration:

```
configure
    service <type> <id>
        subscriber-interface <ip-int-name>
```

```
            address <ip-address> gw-ip-address <gw-address> track-srrp <srrp-inst> holdup-time
<msec>
            ipv6
            subscriber-prefixes
                prefix <ipv6-prefix> pd track-srrp <srrp-id> holdup-time <msec>
                prefix <ipv6-prefix> wan-host track-srrp <srrp-id> holdup-time <msec>


   policy-options
       begin
          policy-statement <name>
           entry 1
               from
                   protocol direct
                   state 'srrp-master'
                   exit
              action accept
               metric set 100
               exit
           exit
           entry 2
               from
                   protocol direct
                   state 'srrp-non-master'
                   exit
               action accept
               metric subtract 10
               exit
           exit
           entry 3
               from
                   protocol direct
                   exit
               action accept
               exit
           exit
```

This configuration ensures that the route metric is changed for the subscriber interface routes based on the SRRP state while the other, non-subscriber directly attached routes are unaffected by SRRP.

The route advertisement based on SRRP State requirement is applicable to BGP and IGP.

The routing policy also provides the flexibility to prevent route advertisement (**action** *reject*) instead of changing the route metric.

Although this feature is designed to minimize or eliminate the use of the redundant interface, it is important to note that the redundant interfaces are still used in the case of transient conditions. An example of such condition would be:

1. Messaging SAP Fails
2. SRRP switches over
3. Stale routing in the core is still in the effect while the metric is being propagated (or the route is being advertised or withdrawn). During this time, traffic is flowing over the redundant interface.
4. Network convergence is complete
5. Traffic in the network core is redirected to the new active node (SRRP master state)

9.3.21.3.2 Managed routes

Only the state attribute is applicable to managed routes, and only to the ones that are synchronized (static and RADIUS obtained – framed-route and framed-ipv6-route). The managed routes obtained by BGP are not synchronized and this feature is not applicable to them.

Based on the SRRP state, the managed route can be either advertised with a modified metric or be withdrawn altogether.

For example:

Managed routes that are tracking SRRP state are only advertised from the active node (SRRP master state) and denied from standby node (SRRP backup state). All other managed routes that are not tracking SRRP state are advertised regardless of the SRRP state.

```
policy-options
    begin
        policy-statement <name>
            entry 1
                from
                    protocol managed
                    state ‘srrp-master’
                    exit
                action accept
                exit
            exit
            entry 2
                from
                    protocol managed
                    state ‘srrp-non-master’
                    exit
                action reject
                exit
            exit
            entry 3
                from
                    protocol managed
                    exit
                action accept
                exit
            exit
```

9.3.21.3.3 Subscriber management routes (/32 IPv4 host routes, IPv6 PD WAN host routes)

Both attributes (state and origin) are applicable to the subscriber management routes.

For example:

A service provider wants to advertise only subscriber management routes with the origin dynamic and AAA from the active node (SRRP master state). Routes with the LUDB origin are not advertised. The standby node is not advertising any /32 subscriber management routes.

```
policy-options
    begin
        policy-statement <name>
            entry 1
                from
                    origin dynamic
                    state ‘srrp-master’
```

```
                    exit
            action accept
            exit
        exit
    exit
           entry 2
                from
                    origin dynamic
                    state 'srrp-master'
                    exit
            action accept
            exit
        exit
```

Default action is reject.

9.3.21.3.4 Activating SRRP state tracking

The SRRP state tracking by routes is turned on only when needed.

For subscriber-interface routes (IPv4 and IPv6), this is performed explicitly.

```
subscriber-interface <ip-int-name>
    address <ip-address> gw-ip-address <gw-address> track-srrp <inst-name> holdup-
time <msec>
        ipv6
        subscriber-prefixes
            prefix <ipv6-prefix> pd track-srrp <srrp-id> holdup-time <msec>
            prefix <ipv6-prefix> wan-host track-srrp <srrp-id> holdup-time <msec>
```

For managed and subscriber management routes, this is explicitly enabled under the group interface:

```
group-interface <ip-int-name>
    srrp-enabled-routing holdup-time <msec>
```

9.3.21.4 SRRP in conjunction with a PW in ESM environment – use case

In specific cases, subscriber traffic is terminated on the BNG by an Epipe. In this case, the subscriber traffic can be offloaded onto a plain Ethernet port by a VSM module (a 'loop') so that it can be terminated in ESM. Epipes can be configured in A/S configuration and terminated on two BNG nodes in multihomed environment.

In these multihomed environment with Epipes and 'loops', the ESM itself detaches from the Epipe, which brings the subscriber traffic to the BNG. Because of that, the ESM would not know if the PW's state is active or standby. As a result, in the downstream direction, traffic could end up being forwarded toward the standby PW, effectively being black-holed.

To overcome this, SRRP can be used in conjunction with an additional mechanism to help monitor the activity of the PWs. This monitoring mechanism is very similar to Fate-sharing. The difference in this case is that the messaging SAP (instead of SRRP instance) is monitoring the activity of the PW. As a result, the SRRP messaging SAP reflects the state of the PW. For example, the PW in a Standby mode would cause the messaging SAP to be in the DOWN state while the PW Active state would cause the messaging SAP to be in the UP state. That is, the SRRP instance reflects the operational state of the messaging SAP. SRRP is indirectly tied into PW state.

Modifying the priority of SRRP instance based on PW's state as a mean of tying the SRRP master state to the active PW would not help here as SRRP messages are not flowing over standby PWs. This is why SRRP state must be enforced by the messaging SAP.

Fate-sharing for PW termination in conjunction with SRRP is not supported.

Metric adjustment for the subscriber routes is supported. After the tracked SRRP instance transitions into a non-master SRRP state, the state attribute of the route changes and the appropriate action defined in the routing policy is taken.

9.3.21.5 Group monitor

The failure detection mechanism to trigger an action within FSG relies on the operational state of the messaging SAP. Such failure detection mechanism is referred as a group monitor.

Group monitor can also be used to detect the state change of the PW. PW state change is reflected in the messaging SAP which in turn triggers the state change of an SRRP instance.

All this is implemented through an oper-group object which is described in the *7450 ESS, 7750 SR, 7950 XRS, and VSR Layer 3 Services Guide: IES and VPRN*. All entities that needs to be monitored (messaging SAPs and PWs) are associated with this oper-group object. Finally, an SRRP instance (in case of FSG) or a messaging SAP (in case of PW) is instructed to monitor the entities in the oper-group object. State transitions of objects in a oper-group object trigger state transitions of entities that are monitoring them (messaging SAPs and SRRP instances). State transitions of monitored objects in a oper-group cause the following actions:

- With FSG, priorities of SRRP instances are recalculated.
- With PW termination on BNG, the operational state of the messaging SAP is changed.

The following is an overview of the CLI syntax showing the principles to create an operational group (**oper-group**). For command descriptions and full syntax, see the *7450 ESS, 7750 SR, 7950 XRS, and VSR Classic CLI Command Reference Guide* and *7450 ESS, 7750 SR, 7950 XRS, and VSR MD-CLI Command Reference Guide*.

- Create an oper-group.

```
config>service
    oper-group <name> [create]
```

- Add a SAP to the oper-group.

```
config>service(ies | vprn)>sub-if>grp-if>sap
    oper-group <name>
```

- Link the state of an oper-group to the SAP. A messaging SAP can monitor the state of a PW.

```
config>service(ies | vprn)>sub-if>grp-if>sap
    monitor-oper-group <name>
```

- Link the state of an oper-group to the SRRP instance. A state transition of an object in the oper-group (not the state of the oper-group itself) triggers an SRRP priority recalculation. When an object within the oper-group goes down, the SRRP priority is lowered by a priority step. The SRRP priority is adjusted on every state transition of an oper-group member object.

```
config>service(ies | vprn)>sub-if>grp-if>srrp <srrp-id>
```

```
        monitor-oper-group <name> priority-step [0-253]
```

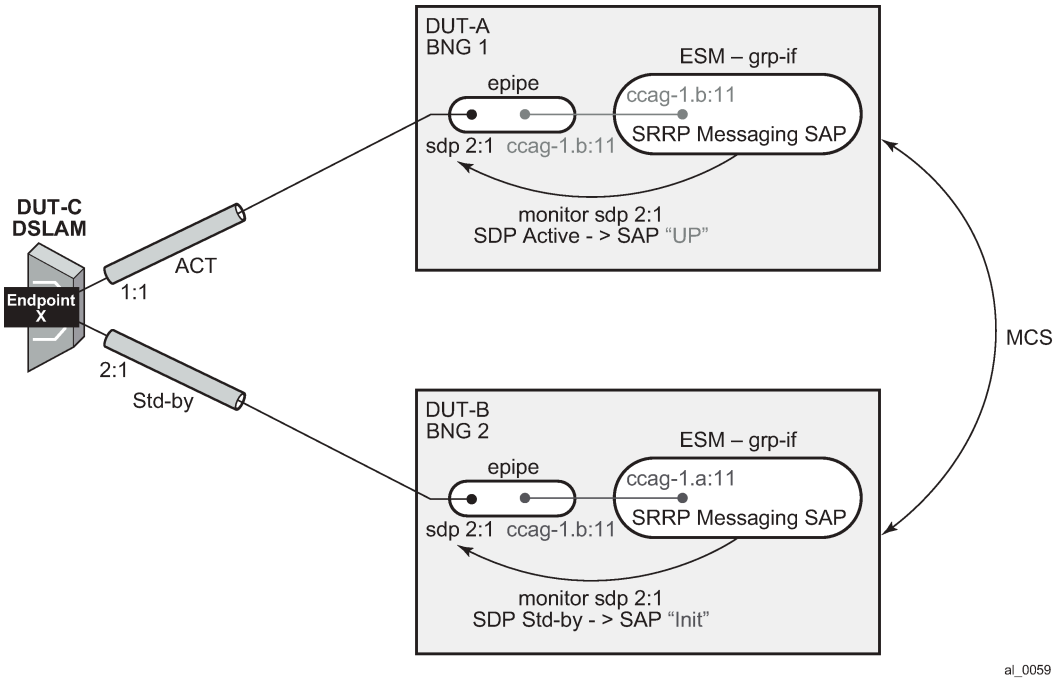
- Add a PW to the oper-group. A messaging SAP monitors the oper-group and assumes the state according to the state of the PW in the oper-group. A standby or a down PW state causes the messaging SAP to assume a down state. Otherwise, the messaging SAP would be in the UP state. In order for the SAP to assume the down state, both RX and TX sides of the PW must be shut down. In other words, a PW in standby mode also must have the local TX disabled by the via the 'slave' flag (**standby-signaling-slave** command in the spoke SDP context). Without the TX disabled, the SAP monitoring the PW does not transition in the down state.

```
    config>service>epipe>spoke-sdp
        oper-group <name>
```

A hold timer is provided within the **oper-group** command to suppress flapping of the monitored object (SAP or pseudowire).

Figure 117: Pseudowire example shows an example with ESM over pseudowire through a VSM loop.

Figure 117: Pseudowire example




```
*A:Dut-C>config>service>epipe# info
----------------------------------------------
            endpoint "x" create
                standby-signaling-master
            exit
            sap 1/1/7:1 create
            exit
            spoke-sdp 1:1 endpoint "x" create
                precedence primary
                no shutdown
            exit
            spoke-sdp 2:1 endpoint "x" create
```

```
                no shutdown
            exit
            no shutdown
-----------------------------------------------

*A:Dut-A>config>service>epipe# info
-----------------------------------------------
            sap ccag-1.b:11 create
            exit
            spoke-sdp 2:1 create
                standby-signaling-slave
                oper-group "1"
                no shutdown
            exit
            no shutdown
-----------------------------------------------
*A:Dut-B>config>service>epipe# info
-----------------------------------------------
            sap ccag-1.b:11 create
            exit
            spoke-sdp 2:1 create
                standby-signaling-slave
                oper-group "1"
                no shutdown
            exit
            no shutdown
-----------------------------------------------
*A:Dut-A>config>service>ies# info
-----------------------------------------------
            redundant-interface "redif11" create
                address 10.1.1.2/24 remote-ip 10.1.1.4
                spoke-sdp 1:1 create
                    no shutdown
                exit
            exit
            subscriber-interface "subif_1" create
                shutdown
                address 10.1.1.2/24 gw-ip-address 10.1.1.100
                group-interface "grpif_1_2" create
                    shutdown
                    redundant-interface "redif11"
                exit
            exit
            subscriber-interface "subTest" create
                address 10.1.1.2/24 gw-ip-address 10.1.1.254
                group-interface "grpTest" create
                    redundant-interface "redif11"
                    sap ccag-1.a:1 create
                    exit
                    sap ccag-1.a:11 create
                        monitor-oper-group "1"
                    exit
                    srrp 11 create
                        message-path ccag-1.a:11
                        no shutdown
                    exit
                exit
            exit
            no shutdown
-----------------------------------------------
*A:Dut-B>config>service>ies# info
-----------------------------------------------
            redundant-interface "redif11" create
                address 10.1.1.4/24 remote-ip 10.1.1.2
```

```
                spoke-sdp 1:1 create
                    no shutdown
                exit
            exit
            subscriber-interface "subif_1" create
                shutdown
                address 10.1.1.4/24 gw-ip-address 10.1.1.100
            exit
            subscriber-interface "subTest" create
                address 10.1.1.4/24 gw-ip-address 10.1.1.254
                group-interface "grpTest" create
                    redundant-interface "redif11"
                    sap ccag-1.a:1 create
                    exit
                    sap ccag-1.a:11 create
                        monitor-oper-group "1"
                    exit
                    srrp 11 create
                        message-path ccag-1.a:11
                        no shutdown
                    exit
                exit
            exit
            no shutdown
-------------------------------------------
*A:Dut-B>config>service>ies# show srrp
===============================================================================
SRRP Table
===============================================================================
ID        Service        Group Interface                Admin     Oper
-------------------------------------------------------------------------------
11        1              grpTest                        Up        initialize
-------------------------------------------------------------------------------
No. of SRRP Entries: 1
===============================================================================
*A:Dut-A>config>service>ies# show srrp
*A:Dut-A>config>service>ies#
===============================================================================
SRRP Table
===============================================================================
ID        Service        Group Interface                Admin     Oper
-------------------------------------------------------------------------------
11        1              grpTest                        Up        master
-------------------------------------------------------------------------------
No. of SRRP Entries: 1
===============================================================================
```

9.3.22 Subscriber QoS overrides

Subscriber QoS overrides enable per-subscriber and per-SLA Profile Instance QoS parameter customization to reduce the number of sub-profiles and sla-profiles that must be configured on the router to cover all needed service level combinations.

Subscriber QoS overrides can be installed at subscriber host or session creation:

- with an Alc-Subscriber-QoS-Override VSA in a RADIUS Access-Accept message
- with a Charging-Rule-Install/Charging-Rule-Definition/QoS-Information AVP in a DIAMETER Gx CCA message

- with a Qos-Information AVP at command level in a DIAMETER Gx CCA message and including the APN-Aggregate-Max-Bitrate-UL/DL AVPs or the Extended-APN-AMBR-UL/DL AVPs. The mapping of the APN-AMBR AVPs to a QoS override must be configured in the DIAMETER Gx application policy:

  MD-CLI:

```
configure subscriber-mgmt diameter-gx-policy "diam-gx-1" gx three-gpp-qos-mapping
    apn-ambr-dl {
        ## ignore-override
        ## arbiter
        ## policer
        ## queue
        ## scheduler
        ## aggregate-rate
        ## hs-sla-agg-rate
    }
    apn-ambr-ul {
        ## ignore-override
        ## arbiter
        ## policer
        ## queue
        ## scheduler
    }
```

  classic CLI:

```
configure subscriber-mgmt diameter-application-policy "diam-gx-1" gx 3gpp-qos-mapping ?
- 3gpp-qos-mapping

[no] apn-ambr-dl - Configure the APN-AMBR mapping for the downlink
[no] apn-ambr-ul - Configure the APN-AMBR mapping for the uplink
```

The APN-Aggregate-Max-Bitrate-UL AVP or Extended-APN-AMBR-UL AVP can be configured to override an ingress policer PIR rate, an ingress queue PIR rate, an ingress arbiter rate, an ingress user scheduler rate, or to ignore the override. The APN-Aggregate-Max-Bitrate-DL AVP or Extended-APN-AMBR-DL AVP can be configured to override an egress policer PIR rate, an egress queue PIR rate, an egress arbiter rate, an egress user scheduler rate, an egress aggregate rate, an egress high-scale SLA aggregate rate, or to ignore the override.

Note:

– The primary use for APN-AMBR AVPs, is to apply upstream and downstream QoS for GTP hosts. However, the APN-AMBR AVPs can also be used for other host types, such as IPoE and PPPoE sessions.

– Either the APN-Aggregate-Max-Bitrate-UL/DL AVPs (with rates in bits-per-second) or the Extended-APN-AMBR-UL/DL AVPs (with rates in kilobits-per-second) should be used. Support for the Extended-APN-AMBR-UL/DL AVPs must be enabled in the DIAMETER Gx application policy:
  MD-CLI:

```
configure subscriber-mgmt diameter-gx-policy "diam-gx-1"
    gx {
        include-avp {
            supported-features true
        }
        features {
            extended-bandwidth true
        }
```

```
    }
```

classic CLI:

```
configure subscriber-mgmt diameter-application-policy "diam-gx-1"
    gx
        features
            extended-bw
        exit
        include-avp
            supported-features
        exit
    exit
```

Subscriber QoS overrides can be installed, updated or removed in a mid-session change with a RADIUS CoA, a DIAMETER Gx RAR or a DIAMETER Gx CCA message using the same attributes as for a subscriber host or session creation.

Subscriber QoS overrides can also be activated using subscriber services. See QoS override-based subscriber service for details.

The format of the [26.6527.126] Alc-Subscriber-QoS-Override VSA is described in the *7450 ESS, 7750 SR, and VSR RADIUS Attributes Reference Guide*.

The format of QoS Overrides AVP's in the 3GPP-1016 QoS-Information AVP are described in the *7750 SR and VSR Gx AVPs Reference Guide*.

The following SLA profile instance QoS parameters can be overridden:

- ingress and egress queue: pir, cir, mbs, cbs
- ingress and egress policer: pir, cir, mbs, cbs
- egress queue class weight (applies to HSQ card only)
- egress queue wrr weight (applies to HSQ card only)
- egress aggregate rate (applies to HSQ card only)
- egress wrr group: rate, class weight (applies to HSQ card only)

The following subscriber QoS parameters can be overridden:

- egress aggregate rate
- ingress and egress root arbiter rate
- ingress and egress intermediate arbiter rate
- ingress and egress user scheduler: rate, cir

  The ingress and egress user scheduler overrides through DIAMETER Gx can only be performed using APN-Aggregate-Max-Bitrate-UL and APN-Aggregate-Max-Bitrate-DL AVPs and requires the following 3gpp-qos-mapping in the DIAMETER Gx Application policy

```
>config>subscr-mgmt>diam-appl-plcy>gx>
        3gpp-qos-mapping
             apn-ambr-dl scheduler <scheduler-name>
             apn-ambr-ul scheduler <scheduler-name>
         exit
```

The operational value of some of the QoS parameters can be derived from different sources.

For queue and policer QoS parameters, the following hierarchy applies (highest priority is listed first):

- Credit Control overrides
- Subscriber Services QoS overrides
- Subscriber QoS overrides (RADIUS, DIAMETER)
- Overrides configured at sla-profile level
- Queue parameters set in QoS policy level

For scheduler and arbiter overrides, the following hierarchy applies:

- ANCP overrides
- Subscriber Services QoS overrides
- Subscriber QoS overrides (RADIUS, DIAMETER)
- Overrides configured at sub-profile level
- Scheduler/arbiter parameters as configured in scheduling/policer-control-policy

Up to 18 QoS overrides can be installed per subscriber host or session. A new set of QoS overrides received using a mid-session change replaces the previous set of QoS overrides.

QoS overrides are always stored as part of the subscriber host or session data but are only applied when the override is valid in the active QoS configuration. For example:

An egress queue 5 PIR rate override is stored with the subscriber session but not applied when the sap-egress QoS policy has no queue 5 defined

RADIUS or DIAMETER Gx initiated QoS overrides can be displayed with the following show commands:

- **show service id** *service-id* | *name* **ipoe session detail**
- **show service id** *service-id* | *name* **ppp session detail**
- **show service id** *service-id* | *name* **dhcp lease-state detail**
- **show service id** *service-id* | *name* **dhcp6 lease-state detail**
- **show service id** *service-id* | *name* **arp-host detail**
- **show service id** *service-id* | *name* **slaac-host detail**

Subscriber services initiated QoS overrides can be displayed with:

show service sub-services

The active QoS overrides per-subscriber and per-SLA Profile Instance can be displayed with:

show service active-subscribers detail

The number of allocated and free Subscriber SLA Profile Instance QoS overrides, QoS Intermediate Arbiter Overrides and QoS User Scheduler Overrides per-line card can be monitored with the **tools dump resource-usage card** CLI command.

Subscriber QoS overrides are synchronized through MCS in a dual-homing environment. QoS overrides are not stored in the subscriber-mgmt application persistence file.

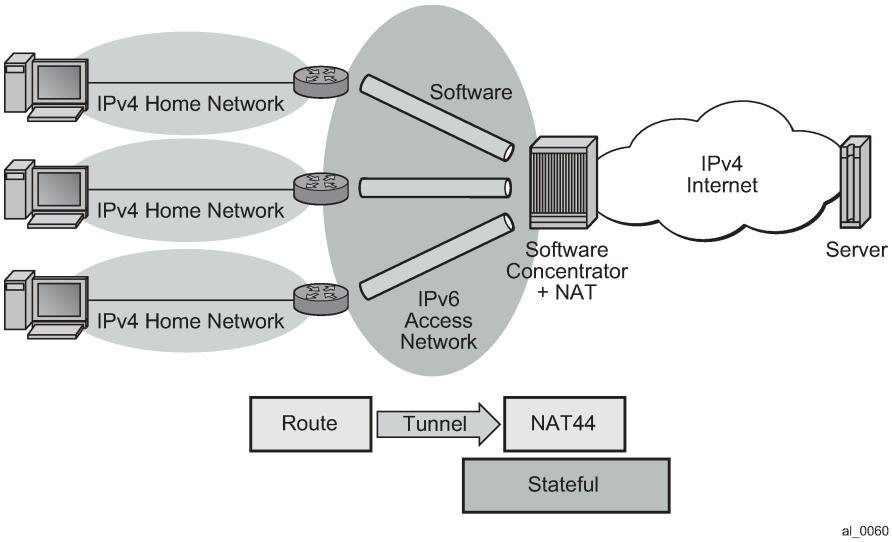
9.3.23 Dual-Stack Lite

The DS-Lite feature is supported on the 7710 SR-Series in combination with the MS-ISA to function as a DS-Lite Address Family Transition Router (AFTR).

DS-Lite is an IPv6 transition technique that allows tunneling of IPv4 traffic across an IPv6-only network. Dual-stack IPv6 transition strategies allow service providers to offer IPv4 and IPv6 services and save on OPEX by allowing the use of a single IPv6 access network instead of running concurrent IPv6 and IPv4 access networks. DS-Lite has two components: the client in the customer network, known as the Basic Bridging BroadBand element (B4) and an Address Family Transition Router (AFTR) deployed in the service provider network.

DS-Lite leverages a network address and port translation (NAPT) function in the service-provider AFTR element to translate traffic tunneled from the private addresses in the home network into public addresses maintained by the service provider. On the 7750 SR, this is facilitated through the Carrier Grade NAT function.

Figure 118: Dual-Stack Lite



As shown in Figure 118: Dual-Stack Lite, DS-Lite has two components, a softwire initiator in the RG and a softwire concentrator, deployed in the service provider network, where control-less IP-in-IP (using protocol 4 - IPv4 in IPv6) is used for tunneling. When using control-less protocol, packets are sent on the wire for the remote softwire endpoint without prior setup of a tunnel.

The softwire initiator in the home network is combined with a routing function, where the default route is passed to the softwire pseudo-interface. Note that there is no NAT function, therefor, the private IP addresses of the home network are encapsulated without source address modification, and forwarded to the softwire concentrator where all NAT is performed. The softwire pseudo-interface unicasts all IPv4 traffic to the IPv6 address of the softwire concentrator, which was pre-configured.

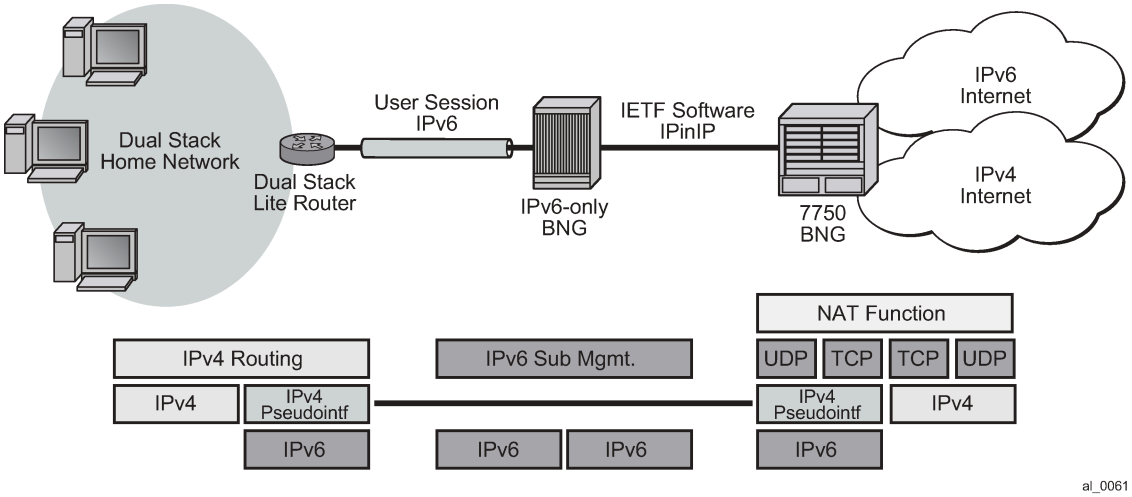
When encapsulated traffic reaches the softwire concentrator, the device treats the source-IP of the tunnel to represent a unique subscriber. The softwire concentrator performs IPv4 network address and port translation on the embedded packet by re-using Large Scale NAT and Layer 2–aware NAT concepts.

9.3.23.1 IP-in-IP

As shown in Figure 119: IP-in-IP, IP-in-IP uses IP protocol 4 (IPv4) to encapsulate IPv4 traffic from the home network across an IPv6 access network. The IPv4 traffic tunneling is treated as best-effort with no subscriber management or policy, and does not use ESM. The scale is dependent only on the internal

structures of the MS-ISA and CPM, that is, the IP-in-IP model can support more subscribers than an ESM-based approach.

Figure 119: IP-in-IP



DS-Lite IP-in-IP is configured through the existing **nat** command that is inside the CLI statements that are within the base router or VPRN. A service performing large scale NAT supports DS-Lite.

DS-Lite expects a routing (non-NATing) gateway in the home, where many different IPv4 inside addresses exist for each subscriber. These inside addresses may overlap other subscriber's address, especially with the heavy use of RFC 1918 address space.

The lack of control of protocol for the IP-in-IP tunnels simplifies the functional model, because any received IPv4 packet to the ISA DS-Lite address can be:

- checked for protocol 4 in the IPv6 header
- checked that the embedded IP packet is IPv4
- processed as if it were L2-Aware, where the source-IP of the tunnel (the source IPv6 address) is used as the subscriber identifier

Note that the inside IP address in the NAT, tables must not be the IPv6 address of the tunnel, but the true IPv4 address of any host within the home. The *subscriber-id* must be the literal IPv6 address (appreciating this may be 34 characters in length).

9.3.23.2 Configuring DS-Lite

DS-Lite is configured on an inside service and uses the existing Large Scale NAT policies and outside pools. DS-Lite and NAT44 Large Scale NAT can operate concurrently on the same inside and outside services.

DS-Lite is configured with the following CLI:

```
configure {router | service vprn service-id}
- [no] nat
- inside
- [no] dual-stack-lite
- [no] *address ipv6-address
```

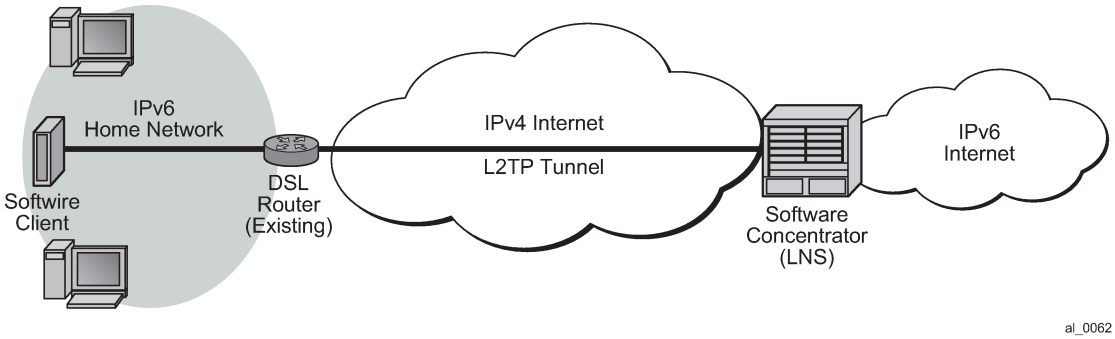
9.3.23.3 L2TP over IPv6

In this mode, L2TP provides the transport for IPv4 that allows full ESM capabilities on the 7750 SR. From the node's perspective, the L2TP tunnel is no different in capability to those already supported. Only the underlying transport (IPv6 instead of IPv4) distinguishes this approach.

To support legacy IPv4 access, L2TP over IPv6 is combined with the existing Layer 2–aware NAT feature as shown in Figure 120: L2TP over IPv6.

As ESM is used, scale is limited by the number of ESM hosts supported on a chassis and any associated resources like queues.

Figure 120: L2TP over IPv6



L2TP LNS over IPv6 is supported in both the base routing instance and VPRN that has 6VPE configured.

Like the LNS implementation, tunnels are terminated on any routing interface, including loopback, SAP, or network port. A single interface simultaneously supports IPv4 and IPv6 L2TP tunnel termination by having two different addresses configured.

For greater scalability, L2TP tunnel and session count per chassis are increased to allow 1 tunnel per session.

NAT capabilities are supported by existing Layer 2–aware NAT methods. Note that the L2TP LNS over IPv6 may be used without NAT as well and the L2TP sessions may be either IPv6-only or dual-stack.

9.3.24 Call trace

Call trace is an enhanced debugging feature that allows control plane messages for a single session to be monitored. When call trace is enabled, all protocols related to this session are captured. Operators can use this information to easily debug entire problematic sessions instead of debugging and verifying separate protocols such as DHCP, ARP, or RADIUS.

Call trace also logs some events that are not directly associated with a protocol, such as LUDB access.

Call trace can present the captured packets for further processing in one of the following ways:

- Call trace can send the captured packets as live output over a UDP tunnel to an external monitoring device.
- Call trace can store the captured packets as PCAP on a pre-provisioned compact flash. If the system already uses the compact flash intensively, such as for ESM persistency, then this method is not recommended for use in a live network.

- Call trace can display decoded packets in debug output. This method is mainly for use in low-scale debugging of a few sessions; use on large-scale live networks may impact overall performance.

Generated traces contain the original packets, encapsulated in a custom header that contains metadata. To decode the metadata and extract the packet, a Nokia-specific Wireshark plug-in is required. Contact the Nokia Technical Assistance Center (TAC) for information.

In general, call trace does not include packets that are common between sessions. Where it is necessary to indicate failure or progress, an event is generated. This is done to guarantee consistency between session traces, independent of timing or session setup order. For example, for PPPoE LAC sessions, L2TP tunnel setup messages are not reflected in call trace, but an event is generated to indicate whether an L2TP tunnel was set up successfully. Subsequent L2TP session setup messages are traced in context of the PPPoE LAC session.

When storing call trace results on a compact flash, files are not automatically synchronized to the standby CPM.

Call trace distinguishes between traces and trace jobs. A trace consists of a set of matching criteria and additional parameters such as a trace profile and a name. Each session that matches a trace creates a trace job if system resources are available. Trace jobs can either be stopped individually or by removing the original trace. By default, existing sessions do not create a trace job when a new trace is enabled; this functionality must be explicitly enabled.

9.3.25 DNS and NBNS name server IP addresses for subscriber sessions

The Domain Name System (DNS) in the Internet provides translation services between human readable names, such as nokia.com, that are used by underlying IP protocols in the URLs of web browsers and IP addresses. A DNS name server or resolver answers IPv4 A-record and/or IPv6 AAAA-record queries from DNS clients such as subscriber sessions. The DNS name server address can be an IPv4 or IPv6 address and must be provisioned in the client. A DNS name server reachable via an IPv4 address can also answer IPv6 AAAA-record queries and the other way around.

Similar, a NetBIOS Name Service (NBNS) provides services to register and lookup computer names on a network that uses NetBIOS as a naming service. NBNS name resolution is IPv4 only. An NBNS name server answers NBNS queries from clients such as subscriber sessions. The NBNS name server address is always an IPv4 address and must be provisioned in the client.

9.3.25.1 DNS and NBNS name server origins

The IPv4 and IPv6 addresses of DNS name servers and the IPv4 addresses of NBNS name servers can be dynamically assigned to subscriber sessions from different authentication origins as listed in Table 25: DNS and NBNS name server authentication origins . The default subscriber management authentication origin priority determines the relative priority when DNS and NBNS name server IP addresses are obtained from multiple origins as illustrated in Figure 121: DNS and NBNS name server authentication origins.

Table 25: DNS and NBNS name server authentication origins

Default origin priority[3]	ESM authentication origin	Name server	DNS/NBNS name server IP address configuration
1	Python – alc.dtc.setESM() – alc.esm.set()	NBNS	alc.dtc.primNbns, alc.esm.primNbns alc.dtc.secNbns, alc.esm.secNbns
		IPv4 DNS	alc.dtc.ipv4PrimDns, alc.esm.ipv4PrimDns alc.dtc.ipv4SecDns, alc.esm.ipv4SecDns
		IPv6 DNS	alc.dtc.ipv6PrimDns, alc.esm.ipv6PrimDns alc.dtc.ipv6SecDns, alc.esm.ipv6SecDns
3	Local user database[4] (LUDB)	NBNS	options netbios-name-server <ip-address> [<ip-address>...(up to 4 max)]
		IPv4 DNS	options dns-server <ip-address> [<ip-address>...(up to 4 max)]
		IPv6 DNS	options6 dns-server <ipv6-address> [<ipv6-address>...(up to 4 max)]
4	RADIUS (AAA)	NBNS	26.6527.29 Alc-Primary-Nbns 26.6527.30 Alc-Secondary-Nbns
		IPv4 DNS	26.6527.9 Alc-Primary-Dns 26.6527.10 Alc-Secondary-Dns
		IPv6 DNS	26.6527.105 Alc-Ipv6-Primary-Dns 26.6527.106 Alc-Ipv6-Secondary-Dns
5	Diameter NASREQ (AAA)	NBNS	26.6527.29 Alc-Primary-Nbns 26.6527.30 Alc-Secondary-Nbns
		IPv4 DNS	26.6527.9 Alc-Primary-Dns 26.6527.10 Alc-Secondary-Dns
		IPv6 DNS	26.6527.105 Alc-Ipv6-Primary-Dns 26.6527.106 Alc-Ipv6-Secondary-Dns
6	Local Address Assignment	NBNS	dhcp local-dhcp-server pool options netbios-name-server <ip-address> [<ip-address>...(up to 4 max)]

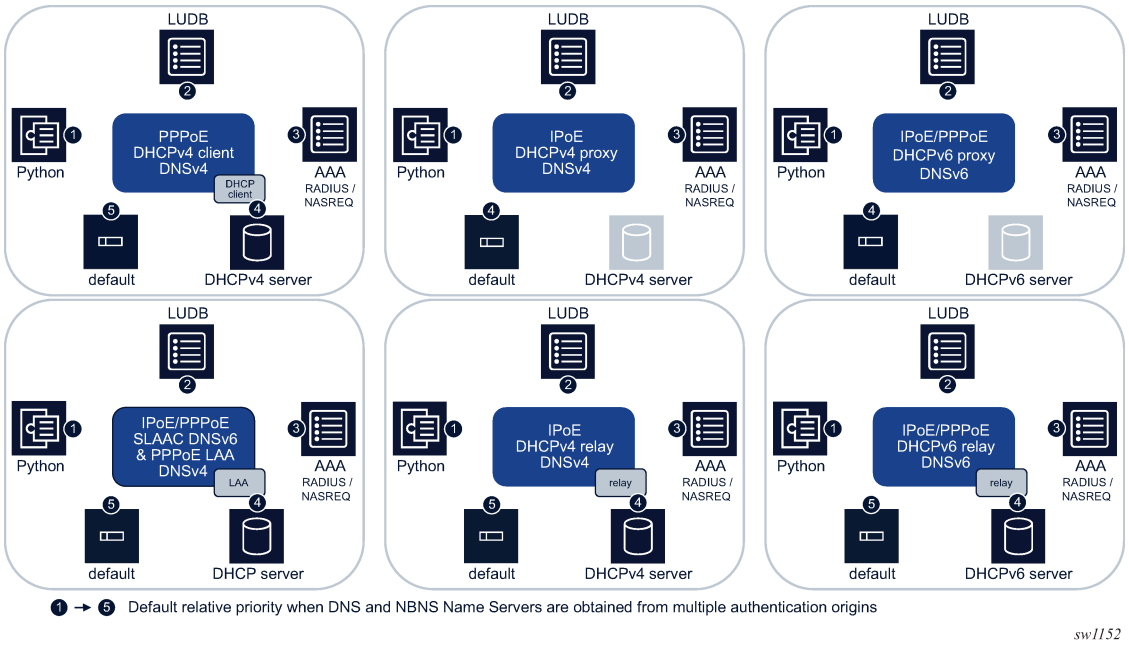
3 **show subscriber-mgmt authentication-origin** outputs the operational priority.

4 Only the first two DNS and NBNS name servers are returned when accessing the local user database during subscriber session authentication.

Default origin priority[3]	ESM authentication origin	Name server	DNS/NBNS name server IP address configuration
		IPv4 DNS	dhcp local-dhcp-server pool options dns-server <ip-address> [<ip-address>...(up to 4 max)]
		IPv6 DNS	dhcp6 local-dhcp-server defaults options dns-server <ipv6-address> [<ipv6-address>...(up to 4 max)] dhcp6 local-dhcp-server pool options dns-server <ipv6-address> [<ipv6-address>...(up to 4 max)] dhcp6 local-dhcp-server pool prefix options dns-server <ipv6-address> [<ipv6-address>... (up to 4 max)]
8	DHCP	NBNS	dhcp local-dhcp-server pool options netbios-name-server <ip-address> [<ip-address>...(up to 4 max)]
		IPv4 DNS	dhcp local-dhcp-server pool options dns-server <ip-address> [<ip-address>...(up to 4 max)]
		IPv6 DNS	dhcp6 local-dhcp-server defaults options dns-server <ipv6-address> [<ipv6-address>...(up to 4 max)] dhcp6 local-dhcp-server pool options dns-server <ipv6-address> [<ipv6-address>...(up to 4 max)] dhcp6 local-dhcp-server pool prefix options dns-server <ipv6-address> [<ipv6-address>... (up to 4 max)]
last resort	Defaults IES/VPRN subscriber-interface	NBNS	No defaults
		IPv4 DNS	default-dns <ip-address> [secondary <secondary-ip-address>]
		IPv6 DNS	ipv6 default-dns <ipv6-address> [secondary <ipv6-address>]

[3] **show subscriber-mgmt authentication-origin** outputs the operational priority.

Figure 121: DNS and NBNS name server authentication origins



9.3.25.2 Primary, secondary, and extended name servers

For redundancy purposes multiple name servers can be associated with a subscriber session:

- **for NBNS name servers**

  a primary and secondary address

- **for IPv4 and IPv6 DNS name servers**

  a primary, a secondary, and extended addresses

  The extended DNS name servers are an ordered list of addresses beyond primary and secondary and can be provisioned using an SR OS or third party DHCP server only. Extended addresses are not applicable for PPPoE IPCP subscriber hosts.

The order of preference in which the name servers are sent to the client is:

1. primary
2. secondary
3. extended

A client typically contacts the name servers in order of preference.

Typically, all name servers are obtained from the same authentication origin, for example RADIUS, but this is not enforced in SR OS. For each subscriber session, primary, secondary, and extended name servers are independently determined based on the authentication origin priorities.

For example, DNS name server IP addresses obtained from different authentication origins for an IPoE DHCPv4 host (relay):

- a primary DNS server (10.1.1.1) is configured in the Local User Database (LUDB)

- a primary (10.1.2.1) and secondary (10.1.2.2) DNS server is received in a RADIUS Access-Accept message
- the DHCP Offer or Ack message received from the DHCP server contains a domain name server option that includes four DNS servers (10.1.3.1, 10.1.3.2, 10.1.3.3 and 10.1.3.4)

Using default authentication origin priorities, the following DNS name server IP addresses are associated with the subscriber session and included in a domain name server option in the DHCP Ack message sent to the client:

- Primary DNS = 10.1.1.1 (origin = LUDB, highest priority for primary DNS)
- Secondary DNS = 10.1.2.2 (origin = RADIUS, highest priority for secondary DNS)
- Extended DNS 1 = 10.1.3.3 (origin = DHCP)
- Extended DNS 2 = 10.1.3.4 (origin = DHCP)

Note: Extended DNS name servers are handled as a set: they should come from the same authentication origin (only DHCP in current release) and all extended DNS name servers are updated when changed mid-session.

9.3.25.3 Assigning DNS and NBNS name servers to subscriber sessions

9.3.25.3.1 Initial subscriber host or session creation

After authentication of the first host of a subscriber session, primary, secondary, and extended DNSv4 and DNSv6 name servers and primary and secondary NBNS name servers of the highest authentication origin priority are associated with the subscriber session. The name servers of the authenticating host's IP stack are sent to the client. The same happens when a new host is associated with an existing session and re-authentication is performed.

When a new host is associated with an existing session and no re-authentication is performed, the name servers of the new host's IP stack that are associated with the subscriber sessions are sent to the client. In the case of DHCP relay, the name servers obtained from the DHCP servers are used if a corresponding name server obtained from a higher priority authentication origin is not associated with the session. Also when the DHCP server does not provide name servers, the configured subscriber interface defaults are associated with the session.

9.3.25.3.2 Changing DNS and NBNS name servers mid-session

DNS and NBNS name servers can be updated mid-session as follows:

- for authenticated renewals of IPoE DHCP hosts, such as a DHCP host renewal of an IPoE session for which the configured minimum authentication interval has expired — primary, secondary, and extended DNSv4 and DNSv6 name servers and primary and secondary NBNS name servers of the highest authentication origin priority are associated with the subscriber session. The name servers of the authenticating host's IP stack are sent to the client.
- for unauthenticated renewals of IPoE DHCP hosts and PPPoE DHCPv6 hosts — if the name servers of the renewing host's IP stack that are associated with the session were obtained from DHCP or defaults, the name servers committed by the DHCP server (or the defaults) are sent to the client. Otherwise, the name servers of the renewing host's IP stack that are associated with the session are sent to the client.

- for RADIUS CoA — the name servers received in a CoA are immediately associated with the subscriber session and sent to the client at the next unauthenticated DHCP renewal. For SLAAC hosts, an unsolicited Router Advertisement is sent if the DNSv6 name server addresses in the CoA are different from those stored in the session.

  When updating DNS or NBNS name servers with a CoA, it is important to also update all authentication sources such that when the subscriber session re-authenticates, the correct name servers are assigned. For example:

  – A DHCPv6 subscriber host connects and obtains primary and secondary DNSv6 name server addresses from the DHCP server. The corresponding IPoE session has a minimum authentication interval of 24 hours. The lease time is one hour.

  – The subscriber signs up for a parental control service which requires an update of its DNSv6 name servers. These servers are provided from RADIUS which takes up to 24 hours to update, as defined by the **min-auth-interval** command configured for the IPoE session.

  – To speed up the activation of the parental control subscription, a CoA is sent to the subscriber session which updates the DNS name servers associated with the session. At the next unauthenticated renew, the updated DNS name servers are sent to the client. This takes 30 minutes maximum (or half the lease time). At the same time, the RADIUS database is updated such that the updated DNS name servers is returned for that subscriber.

  – At the next authenticated renewal, the DNS name servers returned in the RADIUS Access Accept have priority over the DHCP server returned DNS name servers and are sent to the client.

9.3.25.3.3 Verifying the DNS and NBNS name servers stored for a subscriber session

The following show commands are used to verify the DNS and NBNS name servers stored for a subscriber session:

- **show service id** *service-name* **ppp session detail**
- **show service id** *service-name* **pppoe session detail**
- **show service id** *service-name* **ipoe session detail**
- **show service id** *service-name* **dhcp lease-state detail**
- **show service id** *service-name* **dhcp6 lease-state detail**
- **show service id** *service-name* **slaac host detail**

In the following sample example only DNS and NBNS name servers output is shown:

```
IPv4 NBNS Primary       : N/A
IPv4 NBNS Secondary     : N/A
IPv4 DNS Primary        : 10.1.2.1
IPv4 DNS Secondary      : 10.1.2.2
IPv4 DNS Extended 1     : 10.1.4.3
IPv4 DNS Extended 2     : 10.1.4.4
IPv6 DNS Primary        : 2001:db8:dddd::3:1
IPv6 DNS Secondary      : 2001:db8:dddd::3:2
IPv6 DNS Extended 1     : 2001:db8:dddd::4:3
IPv6 DNS Extended 2     : 2001:db8:dddd::4:4
```

The primary and secondary DNS and NBNS fields are always shown. When no IP address is available, they are shown as N/A (Not Applicable). The Extended DNS fields are only present when the corresponding name server IP addresses are stored in the session state.

In exceptional cases, the DNS name servers stored for a subscriber session do not match the DNS name servers sent to the client. For example, when the DNS name servers were not requested in an Option Request Option (6) for a DHCPv6 host, the DNS name servers are stored in the subscriber session but not sent to the client.

9.3.25.3.4 Deployment model specific notes

9.3.25.3.4.1 IPoE DHCPv4

A group interface configured as DHCPv4 Relay or DHCPv4 Proxy ignores the Parameter Request List Option (55) in DHCPv4 client messages and always inserts a Domain Name Server Option (6), a NetBIOS Name Server Option (44), or both in the DHCP Offer and DHCP Ack message when at least one DNS or NBNS name server IP address or both is received during authentication.

9.3.25.3.4.2 IPoE and PPPoE DHCPv6 IA-NA and IA-PD

A group interface configured as DHCPv6 Relay or DHCPv6 Proxy only inserts a DNS Recursive Name Server Option (23) in the DHCP Advertise and DHCP Reply message when requested by the DHCPv6 client in the Option Request Option (6) and at least one DNS name server IP address is received during authentication.

An SR OS DHCPv6 server (DHCPv6 relay) and a DHCPv6 proxy server insert the DNS Recursive Name Server Option as a global DHCPv6 option.

9.3.25.3.4.3 PPPoE IPCP

A PPPoE client obtains DNS and NBNS name servers by including following configuration options in its IPCP Configure Request:

- Primary DNS Server Address (129)
- Primary NBNS Server Address (130)
- Secondary DNS Server Address (131)
- Secondary NBNS Server Address (132)

Note: A PPPoE DHCPv4 client does not include a Parameter Request List Option (55) in its DHCP messages. The Domain Name Server Option (6), the NetBIOS Name Server Option (44), or both that is returned by the DHCP server are evaluated according to the authentication origin priority to determine the DNS and NBNS name server IP address assigned to the PPPoE session.

Mid-session changes are not supported for PPPoE DNSv4 and NBNS name server updates.

9.3.25.3.4.4 IPoE and PPPoE SLAAC

There are two mechanisms to assign a DNSv6 name server to an IPv6 SLAAC hosts:

- Stateless DHCPv6

The client starts a stateless DHCPv6 transaction by sending an Information Request message.

– PPPoE session

An Information Request message is always authenticated:

- When no DNSv6 name servers are received during authentication, then DHCPv6 relay is performed irrespective of whether the DNSv6 name servers are associated with the PPP session or not. The DNSv6 name servers in the Reply message from the DHCP server (or defaults if not available from DHCP) are sent to the client. These DNSv6 name servers are not associated with the PPP session.
- When DNSv6 name servers are received during authentication, DHCPv6 proxy is performed and the DNSv6 name servers are included in a DNS Recursive Name Server Option (23) of the Reply message sent on behalf of a DHCPv6 server. The DNSv6 name servers are not associated with the PPP session.

Because the Information Request for PPP SLAAC hosts are always authenticated, a mid-session change of DNSv6 name servers using CoA is not supported. Instead, the DNSv6 name servers can be included in the RADIUS Access Accept message.

– IPoE session

An Information Request message is authenticated based on the **ipoe-session min-auth-interval** value. When IPoE sessions are disabled, the authentication is based on the **re-authentication** command in the RADIUS authentication policy.

- Unauthenticated Information Request

  When DNSv6 name servers different from defaults are associated with the IPoE session, DHCPv6 proxy is performed and the DNSv6 name servers are included in a DNS Recursive Name Server Option (23) of the Information Reply message sent on behalf of a DHCPv6 server.

  Extended DNSv6 name servers saved in the IPoE session are not included in the Reply message to the client.

  When default or no DNSv6 name servers are associated with the IPoE session, DHCPv6 relay is performed. The DNSv6 name servers in the Reply message from the DHCP server (or defaults if not available from DHCP) are sent to the client. These DNSv6 name servers are now associated with the IPoE session.

  Mid-session change of DNSv6 name servers using CoA is supported: the DNSv6 name servers in the CoA are associated with the IPoE session and included in the reply to the next unauthenticated Information Request (proxy-server).

- Authenticated Information Request

  When no DNSv6 name servers are received during authentication, then DHCPv6 relay is performed irrespective of whether DNSv6 name servers are associated with the IPoE session or not. The DNSv6 name servers in the Reply message from the DHCP server (or defaults if not available from DHCP) are sent to the client. These DNSv6 name servers are now associated with the IPoE session.

  When DNSv6 name servers are received during authentication, then DHCPv6 proxy is performed and the DNSv6 name servers are included in a DNS Recursive Name Server Option (23) of the Reply message sent on behalf of a DHCPv6 server. These DNSv6 name servers are now associated with the IPoE session.

Mid-session change of DNSv6 name servers using CoA is not supported for authenticated Information Requests. Instead, the DNSv6 name servers can be included in the RADIUS Access Accept message

- Router Advertisements

  A Recursive DNS Server (RDNSS) option as defined in RFC 6106, IPv6 Router Advertisement Options for DNS Configuration, is included in the Router Advertisement sent to the IPv6 SLAAC host.

  The following CLI command includes the RDNSS option in IPv6 Router Advertisements for SLAAC hosts and sets the RDNSS lifetime:

```
config>service>ies>sub-if>grp-if>ipv6>rtr-adv
config>service>vprn>sub-if>grp-if>ipv6>rtr-adv
config>subscr-mgmt>rtr-adv-plcy
dns-options
    [no] include-dns - Set/reset inclusion of the RDNSS server
                       option 25 on this group-interface
    [no] rdnss-lifetime - Maximum time the RDNSS address is valid
                          in this group-interface
```

  The configuration at the group interface level is common to all subscriber sessions active on the interface. The configuration in a router advertisement policy overrides the group interface configuration for the sessions associated with the policy.

9.3.25.3.4.5 IPoE sessions

As a result of the single authentication for dual stack IPoE sessions, DNS and NBNS name servers for both IPv4 and IPv6 should be provided irrespective of the IP stack that triggers the authentication or re-authentication.

The result of a Python alc.dtc.setESM() or alc.esm.set() to set the DNS or NBNS name servers is ignored when the IPoE session is not re-authenticated.

9.3.25.3.4.6 SR OS DHCP server

An SR OS DHCPv4 server does not check the Parameter Request List Option (55) and always includes the configured options for the matched pool. Likewise, an SR OS DHCPv6 server does not check the Option Request Option (6) and always includes the configured options for the matched pool.

An SR OS DHCPv6 server inserts the DNS Recursive name server Option as a global DHCPv6 option.

9.3.25.4 Alternative ways to specify DNS and NBNS name servers

Table 25: DNS and NBNS name server authentication origins lists the DNS and NBNS name server configuration options for the different authentication origins.

Alternatively, the SR OS features described in this section can also be used to send DNS and NBNS name server IP addresses to the subscriber sessions. When using these mechanisms, the authentication origin priorities are overruled, and the name servers associated with the session in the BNG do not correspond with the name servers sent to the client.

9.3.25.4.1 To client options

DHCP options can be specified in RADIUS or Local User Database (LUDB) and then appended to the options present in DHCP messages to the client:

- from RADIUS, using the 26.6527.103 Alc-ToClient-Dhcp-Options and 26.6527.192 Alc-ToClient-Dhcp6-Options in an Access-Accept message
- in LUDB, by configuring **to-client-options**:

```
config>subscr-mgmt>loc-user-db>ipoe>host
      to-client-options
         ipv4
           option <option-number>
         exit
         ipv6
           option <option-number>
         exit
       exit
config>subscr-mgmt>loc-user-db>ppp>host
      to-client-options
         ipv6
           option <option-number>
         exit
       exit
```

When a DHCPv4 Domain Name Server Option (6), a DHCPv4 NetBIOS Name Server Option (44), or a DHCPv6 DNS Recursive Name Server option (23) is included using the described To Client Options methods, these options are appended in the outgoing DHCP message to the client, irrespective of whether DNS or NBNS options were already present. The name servers included in the DHCP options with the To Client Options method are not associated with the subscriber session as primary, secondary, or extended DNS servers.

9.3.25.4.2 DHCP Python

The DHCPv4 and DHCPv6 Python API enables the manipulation of DHCP packets received from or sent to the client.

Inserting a DHCPv4 Domain Name Server Option (6) or NetBIOS Name Server Option (44) in the DHCPv4 Offer and Ack messages using the alc.dhcpv4.set() Python API or inserting a DHCPv6 DNS Recursive Name Server option (23) in the DHCPv6 Advertise and Reply messages using alc.dhcpv6.set() Python API overwrites the corresponding option in the message sent to the client. In this case, the name servers associated with the subscriber session in the BNG do not correspond with the name servers sent to the client.

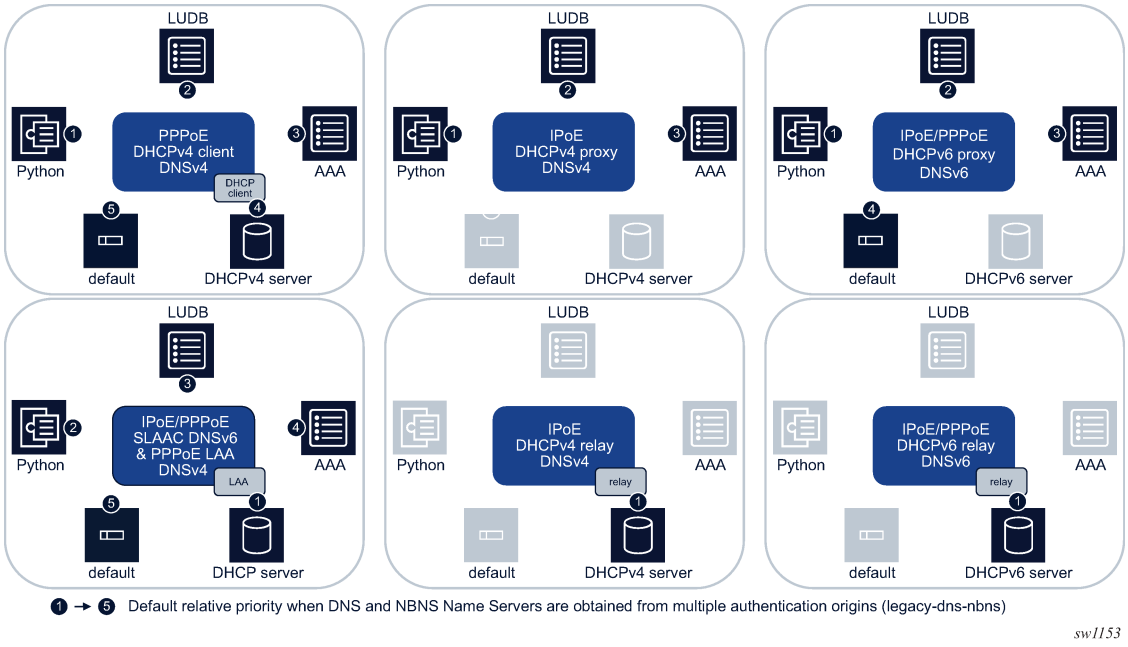
9.3.25.5 Legacy DNS and NBNS name server origins

Important changes occurred in the DNS and NBNS name server origin priorities in SR OS Release 21.7 which could result in different DNS and NBNS name server IP addresses being sent to a subscriber session after an upgrade, if the configurations of the DHCP and RADIUS servers are not simultaneously updated accordingly. To facilitate a smooth transition when the configuration of back-end systems cannot be changed at the time of the upgrade, the legacy behavior, which is backward compatible with SR OS Releases before 21.7 can be enabled using the following configuration:

```
configure subscriber-mgmt
```

```
    system-behavior
        legacy-dns-nbns
```

Figure 122: Legacy DNS and NBNS name server authentication origins



The following changes are enabled with the **legacy-dns-nbns** configuration:

- supported authentication origins and their relative priorities for DNS and NBNS name servers as illustrated in Figure 122: Legacy DNS and NBNS name server authentication origins:
  - **DHCPv4 and DHCPv6 relay**

    DNS and NBNS name servers can only be provided by the DHCP server
  - **DHCPv4 proxy**

    default DNS and NBNS name servers configured at the subscriber interface are not considered
  - **Local Address Assignment** (LAA)

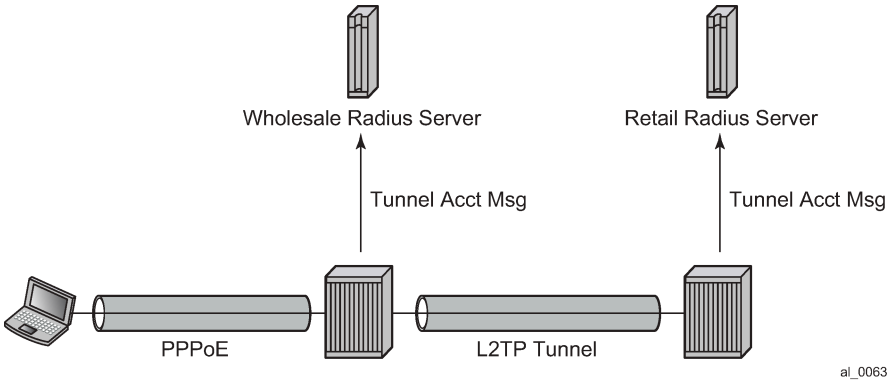
    DNS and NBNS name servers obtained from local address assignment (DHCP server options) have the highest origin priority
- mid-session changes for DNS and NBNS name servers at re-authentication as described in Changing DNS and NBNS name servers mid-session
- a group-interface configured as DHCPv6 Relay inserts a DNS Recursive name server Option (23) in the DHCP Advertise and DHCP Reply message without checking the Option Request Option (6) in the client message

Mid-session changes for DNS and NBNS name servers using RADIUS CoA are enabled by default and are not disabled with the **legacy-dns-nbns** configuration.

Note: The legacy system behavior for DNS and NBNS name servers is available as a temporary workaround. The recommended configuration is the default extended DNS and NBNS name server origin priorities (**no legacy-dns-nbns**).

9.4 L2TP tunnel RADIUS accounting

Figure 123: L2TP tunnel accounting



When L2TP tunnel accounting is enabled, except for **host** or **sla-profile-based** accounting packets and attributes, the following are additional accounting packets and attributes:

- Accounting packets: tunnel-start/stop/reject; tunnel-link-start/stop/reject — There are no interim updates for L2TP tunnel/session accounting.
- RADIUS accounting attributes:
  - Tunnel-Assignment-Id (LAC only)
  - Acct-Tunnel-Connection
  - Acct-Tunnel-Packets-Lost

These attributes were added into current account-start/stop/interim-update packets (host accounting/sla-profile accounting).

Tunnel level accounting and session level accounting can be enabled or disabled independently.

New accounting packets and related RADIUS attribute list are described in Table 26: L2TP tunnel accounting behavior .

Some considerations of RADIUS attributes are described in RADIUS attributes value considerations.

9.4.1 Accounting packets list

Table 26: L2TP tunnel accounting behavior describes L2TP tunnel accounting behavior along with some key RADIUS attributes (apply for both LAC and LNS):

Table 26: L2TP tunnel accounting behavior

Act-packet	When	Key attributes	Remark
Tunnel-Start	A new L2TP tunnel is created	Acct-Session-ID	—
		Event-Timestamp	—

Act-packet	When	Key attributes	Remark
		Tunnel-Type:0	—
		Tunnel-Medium-Type:0	—
		Tunnel-Assignment-Id:0	—
		Tunnel-Client-Endpoint:0	—
		Tunnel-Client-Auth-Id:0	—
		Tunnel-Server-Endpoint:0	—
		Tunnel-Server-Auth-Id:0	—
Tunnel-Reject	A new L2TP tunnel creation failed	Acct-Session-Id	—
		Event-Timestamp	—
		Tunnel-Type:0	—
		Tunnel-Medium-Type:0	—
		Tunnel-Assignment-Id:0	—
		Tunnel-Client-Endpoint:0	—
		Tunnel-Client-Auth-Id:0	—
		Tunnel-Server-Endpoint:0	—
		Acct-Terminate-Cause	—
Tunnel-Stop	An established L2TP tunnel is removed	Acct-Session-Id	—
		Event-Timestamp	—
		Tunnel-Type:0	—
		Tunnel-Medium-Type:0	—
		Tunnel-Assignment-Id:0	—
		Tunnel-Client-Endpoint:0	—
		Tunnel-Client-Auth-Id:0	—
		Tunnel-Server-Endpoint:0	—
		Tunnel-Server-Auth-Id:0	—
		Acct-Session-Time	—
		Acct-Input-Gigawords	—
		Acct-Input-Octets	—

Act-packet	When	Key attributes	Remark
		Acct-Output-Gigawords	—
		Acct-Output-Octets	—
		Acct-Input-Packets	—
		Acct-Output-Packets	—
		Acct-Terminate-Cause	—
Tunnel-Link-Start	An L2TP session is created	User-Name	—
		Acct-Session-Id	This is the same as Acct-Session-id in access-request of host auth
		Event-Timestamp	—
		Service-Type	Framed
		Class	—
		Tunnel-Type:0	—
		Tunnel-Medium-Type:0	—
		Tunnel-Assignment-Id:0	—
		Tunnel-Client-Endpoint:0	—
		Tunnel-Client-Auth-Id:0	—
		Tunnel-Server-Endpoint:0	—
		Tunnel-Server-Auth-Id:0	—
		Acct-Tunnel-Connection	See RADIUS attributes value considerations
Tunnel-Link-Reject	A new L2TP session creation is failed	Acct-Session-Id	Should be as same as Acct-Session-id in access-request of host auth
		Event-Timestamp	—
		Tunnel-Type:0	—
		Tunnel-Medium-Type:0	—
		Tunnel-Assignment-Id:0	—
		Tunnel-Client-Endpoint:0	—
		Tunnel-Client-Auth-Id:0	—
		Tunnel-Server-Endpoint:0	—

Act-packet	When	Key attributes	Remark
		Acct-Terminate-Cause	—
		Acct-Tunnel-Connection	—
Tunnel-Link-Stop	A established L2TP session is removed	User-Name	—
		Acct-Session-Id	Should be as same as Acct-Session-id in access-request of host auth
		Event-Timestamp	—
		Service-Type	Framed
		Class	—
		Tunnel-Type:0	—
		Tunnel-Medium-Type:0	—
		Tunnel-Assignment-Id:0	—
		Tunnel-Client-Endpoint:0	—
		Tunnel-Client-Auth-Id:0	—
		Tunnel-Server-Endpoint:0	—
		Tunnel-Server-Auth-Id:0	—
		Acct-Tunnel-Connection	—
		Acct-Session-Time	—
		Acct-Input-Gigawords	—
		Acct-Input-Octets	—
		Acct-Output-Gigawords	—
		Acct-Output-Octets	—
		Acct-Input-Packets	—
		Acct-Output-Packets	—
		Acct-Tunnel-Packets-Lost	—
		Acct-Terminate-Cause	—

Notes:

- Errors occur if there are multiple hosts sharing the same sla-profile instance and then these hosts go to different tunnel.

- 7750 SRs have an internal limitation of 500 pps for accounting messages. This feature shares the same limitation.

9.4.2 RADIUS attributes value considerations

- The value of Acct-Tunnel-Connection uniquely identify a L2TP session, and to match LAC and LNS accounting record, the value of Acct-Tunnel-Connection is determined by a method shared by LAC and LNS. This means for a specified L2TP session, Acct-Tunnel-Connection from the LAC and LNS are the same.
- Current ESM stats are used in Tunnel-Link and tunnel level accounting. This applies for both standard attribute and the 7750 SR's VSA.
- Tunnel level accounting stats need to aggregate all sessions stats that belong to the tunnel. There may be sessions that traverse before tunnel is down, so the system needs to remember the statistics of every session that has been created within the tunnel.

  This applies for both standard attribute and the 7750 SR's VSA.
- The value of Acct-Tunnel-Packets-Lost is the aggregation of all discarded packets on both ingress and egress.

9.4.3 Other optional RADIUS attributes

Table 27: Optional RADIUS attributes lists the optional attributes that could be optionally included in tunnel accounting packet, some of them are applied for link level accounting only.

Table 27: Optional RADIUS attributes

Attribute	Tunnel/link
nas-identifier	Both
nas-port	Link level only
nas-port-id	Link level only
nas-port-type	Link level only

9.4.4 RADIUS VSA to enable L2TP tunnel accounting

To support pure RADIUS-enabled L2TP tunnel accounting on LAC side, the following RADIUS VSA are supported:

Table 28: Supported RADIUS VSAs

VSA	Type	Value
Alc-Tunnel-Acct-Policy	String	Policy-name; if the name is **disable** then this means L2TP tunnel accounting is disabled for this tunnel

The Alc-Tunnel-Acct-Policy takes precedence over what is defined in CLI when Alc-Tunnel-Group is also returned.

9.4.5 MLPPP on the LNS side

With MLPPP, the counter on LNS side is only available for the bundle, not for each link, so the SR OS's behavior is:

- For each new link session system sends a tunnel-link-start.
- For each link session that is deleted system sends a tunnel-link-stop.
- For all link sessions except the last one system reports 0 for all counters.
- For the last link session, system reports the actual counters for the bundle.

9.5 RADIUS route download

The RADIUS route download mechanism periodically polls a RADIUS server for routes to download. The main objective of this feature is to download, in advance, customer-assigned subnets so that they can be re-advertised to the corresponding routing protocols. In this way, subscriber bring up can potentially be done faster (as the routes are already in place and advertised) and, most importantly, reduce the routing protocol churn as subscribers connect and disconnect. The routes being learned through this mechanism could be both managed routes/delegated prefixes as well as the WAN IP assigned to the subscriber in the case PPPoE and un-numbered interfaces are being used.

The route download process requests the routes to a configured RADIUS server by triggering an access-request message. The key identifier for this message is the username, which is a combination of the system's name (or an optionally configured value), appended by a dash ("-") and then a monotonically increasing integer. The download process sends an access request starting with 1 (such as "hostname-1") and the RADIUS server replies with an access-accept message and a number of routes embedded within the message. The system then increases the counter and sends another access request (this time being hostname-2) and receive a reply with the next batch of routes to download. The process continues, incrementing the counter by 1 each time until the system gets an access-reject or the maximum number of routes that can be downloaded is reached.

The routes to be accepted are in the following format:

[vrf {vprn-name | vprn-service-id}] prefix-mask {null0 | null 0 | black-hole} [metric] [tag tag-value]

The prefix-mask could be in any form as 'prefix/length', 'prefix mask' or 'prefix' (in the latter case, for IPv4 routes, the mask shall be derived from the IP class of the prefix).

The route formats are supported:

- Framed-Route (RADIUS attribute 22)

  Framed-Route = "192.168.3.0 255.255.255.0 null0"

  Framed-Route = "vrf vrfboston 192.168.1.0/24 null 0 0 tag 6"

  Framed-Route = "vrf 2001 192.168.10.0/24 black-hole 0 tag 8"

- Cisco-AVPair (Cisco VSA 26-1)

  cisco-avpair = "ip:route = 192.168.3.0 255.255.255.0 null0"

  cisco-avpair = "ip:route = vrf vrfboston 192.168.1.0/24 null 0 0 tag 6"

IPv6 routes are also supported. The format is based on using the IETF-defined IPv6 Framed-IPv6-Route (attribute 99). The following text shows the supported formats.

- Framed-IPv6-Route (RADIUS attribute 99)
- Framed-Route = "vrf 3000 2200:1bbbb:dead::/48 black-hole 0 tag 6"
- Framed-Route = "vrf vrfboston 2100:5aaa:dead:beaf::/96 null 0 0 tag 6"
- Framed-Route = "2001:100:bad:cafe::/64 null0"

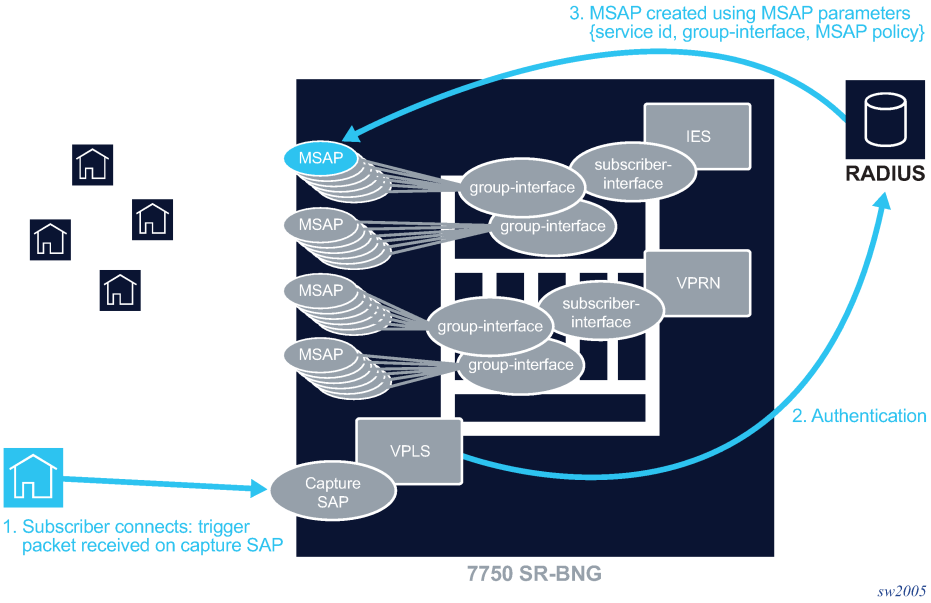
All the routes downloaded are a new protocol type "**periodic**". The download process re-starts the AAA requests after a specific interval (a configurable value but target refresh rate is 15 minutes) and routes shall be updated according to the following process:

- When the router initiates a new download process, the routes are kept in a temporary table until the download process completes (receives an access-reject from the AAA). The temporary download table is then checked for errors and finally, any changes reflected to the actual routing table.
- Routes no longer present in the download are removed from the routing table.
- If the AAA server responds with an access-reject for the first username (that is, an implicit empty route-download table), all routes are removed from the routing table.
- If there are any protocol errors (at the RADIUS level), such as time-out, no response, bad record format, too many records, and so on, the download process is suspended and retried after a configurable timer. The minimum retry timer is at least 1 minute and the light load this represents control-plane-wise (concurrent downloads are not supported) the retries can continue infinitely until the next refresh period occurs, where the download restarts from the beginning. An exponential backoff algorithm with a configured minimum and maximum delay is used to determine the retry timer.
- The routes are only purged from the routing table after a complete download process was achieved (properly terminated with an access-reject message). Under any other failure condition, the routes shall remain active. Shutting down the download process should not remove the downloaded routes. A **clear** command clears the periodic routes.
- All the imported routes (blackholes) are imported into the line-card FIBs to avoid the routing loops caused by announcing the prefixes but not installing the actual blackholes.

9.6 Managed SAPs

Subscriber sessions are created on a subscriber SAP. For a shared VLAN deployment model, these SAPs are usually statically configured as the limited number of VLANs are known. For a VLAN per-subscriber deployment model, it is advantageous that the subscriber SAPs are automatically created and deleted when subscriber sessions connect or disconnect. These are called Managed SAPs (MSAPs).

Figure 124: Managed SAP example configuration



The reception of a valid trigger packet on a capture SAP initiates a RADIUS, DIAMETER, or local user database authentication to provide the service context where the MSAP should be created. The VLAN of the created MSAP is the same as the authenticated trigger packet. An MSAP functions like a regular SAP but its configuration is not user editable and not maintained in the configuration file. By default, an MSAP is deleted from the system when the last subscriber session active on the MSAP disconnects.

9.6.1 Capture SAP

The following trigger types are supported on a capture SAP:

- **dhcp**

  DHCPv4 client messages
- **ppoe**

  PPPoE PADI messages from PPPoE clients

  The MSAP is created after the IP address is provided. A short temporary state handles packets between the PADO and ACK.
- **arp**

  ARP-Request from an ARP host with static configured IPv4 address
- **dhcp6**

  DHCPv6 client messages
- **rtr-solicit**

  Router Solicitation messages from a SLAAC hosts
- **data**

  An ARP-Request, IPv4 or IPv6 packet received from a data-triggered host

Multiple trigger types can be enabled on a single capture SAP. The **data** and **arp** trigger types are mutually exclusive.

A capture SAP is created in a VPLS service by specifying the **capture-sap** parameter. A capture SAP does not forward traffic but captures received trigger packets for authentication. Similar to a default SAP, at least one of the Q-tags of a capture SAP must be a wildcard *, meaning any tag value. See the following example configuration.

```
vpls 10 customer 1 create
    sap 1/1/1:*.* capture-sap create
        description "capture sap"
        trigger-packet arp dhcp dhcp6 pppoe
        authentication-policy "auth-policy-1"
    exit
    no shutdown
exit
```

A capture SAP and default SAP cannot be configured simultaneously on a dot1q- encapsulated port. A capture SAP and default SAP cannot be configured simultaneously on a QinQ-encapsulated port when the outer tag is the same.

A SAP lookup based on the outer and inner tags is performed when a packet is received on a port. When no corresponding SAP or MSAP is found, the packet is handled by the capture SAP, meaning that the trigger packets are sent to the CPM and all other packets are dropped.

An ingress VLAN ID (VID) type **mac** filter can be configured on a capture SAP to have additional control on the VLANs that are allowed to initiate a host setup. Other filter types are not supported on a capture SAP.

For a capture SAP on a dot1q encapsulated port:

<port-id>:* Matches any valid single tagged trigger packet on a <port-id> for which no more specific SAP or MSAP is found. A single Q-tag (<port-id>:tag) is available for authentication. The corresponding MSAP is created as: <port-id>:tag

For a capture SAP on a QinQ-encapsulated port:

- **<port-id>:*.***

  Matches any valid double tagged trigger packet on a <port-id> for which no more specific SAP or MSAP is found.

  Both Q-tags (<port-id>:tag1.tag2) are available for authentication.

  The corresponding MSAP is created as: <port-id>:tag1.tag2.

  The optional **allow-dot1q-msaps** command configured at the capture SAP enables additional support for single-tagged trigger packets:

  – Valid single-tagged trigger packets for which no more specific SAP or MSAP is found are matched on <port-id>

  – A single Q-tag is available for authentication, the second tag is set to zero (<port-id>:tag.0)

  – The corresponding MSAP is created as: <port-id>:tag.0

  – The **config**>**system**>**ethernet**>**new-qinq-untagged-sap** command should be configured where a combination of <port-id>:tag1.0 and <port-id>:tag1.tag2 MSAPs coexist. When not configured, <port-id>:tag1.0 MSAPs attract double-tagged <port-id>:tag1.tag2 encapsulated traffic which is either dropped (IPoE traffic) or handled as single tagged traffic causing PPPoE sessions to fail.

- <port-id>:tag1.*

Matches any valid double-tagged trigger packet with and outer tag equaling tag1 on <port-id> and for which no more specific SAP or MSAP is found.

Both Q-tags (<port-id>:tag1.tag2) are available for authentication.

The corresponding MSAP is created as: <port-id>:tag1.tag2.

The optional **allow-dot1q-msaps** command configured at the capture SAP enables additional support for single-tagged trigger packets:

- Valid single-tagged trigger packets with tag equaling tag1 and for which no more specific SAP or MSAP is found are matched on <port-id>
- A single Q-tag is available for authentication, the second tag is set to zero (<port-id>:tag1.0)
- The corresponding MSAP is created as: <port-id>:tag1.0
- It is a prerequisite to have the **config**>**system**>**ethernet**>**new-qinq-untagged-sap** command configured to enable both <port-id>:tag1.* capture-sap and <port-id>:tag1.0 MSAP to coexist. The <port-id>:tag1.0 capture-sap cannot be created when not configured.

- **<port-id>:*.tag2**

  Matches any valid double-tagged trigger packet with inner tag tag2 on <port-id> for which no more specific SAP or MSAP is found.

  Both Q-tags (<port-id>:tag1.tag2) are available for authentication.

  The corresponding MSAP is created as: <port-id>:tag1.tag2.

  This is an inverse capture SAP that matches on a fixed inner tag with the outer tag identifying the user. The following restrictions apply when an inverse capture SAP is configured on a port:

  - Ethernet ports only
  - It is not possible to create y.* SAPs when there is a *.x capture SAP present on the port (y=0,1..4094,* and x=1..4094).
  - It is not possible to create a y.* network interface when there is a *.x capture SAP present on the port (y=0,1..4094,* and x=1..4094).
  - There is no support for single-tagged MSAP creation.

To enable the creation of single-tagged and double-tagged MSAPs by a QinQ-encapsulated capture SAP, enable the **allow-dot1q-msap** command in the capture SAP context:

```
config service
    vpls 10 customer 1 create
        sap 1/1/1:*.* capture-sap create
            allow-dot1q-msaps
```

In addition, the **new-qinq-untagged-sap** command should be configured for scenarios as described previously:

```
config system
    ethernet
        new-qinq-untagged-sap
```

Be aware that enabling the **new-qinq-untagged-sap** command affects the behavior of existing <port-id>:tag1.0 SAPs.

Valid single-tagged trigger packets result in the creation of a <port-id>:tag.0 MSAP. With the **encap-tag-range** matching in a local user database, it is possible to specify different MSAP defaults for single or double tagged MSAPs. For example:

```
config subscriber-mgmt
    local-user-db "ludb-1" create
        ipoe
            host "single-tagged" create
                host-identification
                    encap-tag-range start-tag *.0 end-tag *.0
                exit
                msap-defaults # defaults for dot1q MSAPs
                    group-interface "group-int-2"
                    policy "msap-policy-2"
                    service 2000
                exit
                no shutdown
            exit
        exit
config service
    vpls 10 customer 1 create
        sap 1/1/1:*.* capture-sap create
            trigger-packet dhcp dhcp6
            allow-dot1q-msaps
            ipoe-session
                ipoe-session-policy "ipoe-policy-1"
                user-db "ludb-1"
                no shutdown
            exit
            msap-defaults # defaults for qinq MSAPs
                group-interface "group-int-1"
                policy "msap-policy-1"
                service 1000
            exit
        exit
```

9.6.2 MSAP parameters

A set of mandatory parameters must be provisioned for MSAP creation:

- **Service ID**

  The service context in which the MSAP is created.

- **Interface name**

  The name of the group interface context in which the MSAP is created. The group interface must exist in the provided service in order for the MSAP to be installed (in a routed CO scenario only).

- **MSAP policy**

  The name of the policy that defines the MSAP parameters. The policy must exist in the subscriber-mgmt context.

MSAP parameters can be obtained from multiple sources with the following order of preference:

- Explicit MSAP parameters, specified during subscriber host or session authentication

  1. Local user database lookup

  2. RADIUS or DIAMETER authentication

- Implicit MSAP parameters

  3. Defaults configured in the capture SAP context

9.6.2.1 Explicit MSAP parameters from local user database

The local user database should be configured at the capture SAP and group interface context. For example:

- IPoE sessions:

  config>**service**>**vpls**>**sap**# ipoe-session user-db local-user-db-name

  config>service>ies>sub-if>grp-if# ipoe-session user-db local-user-db-name

- PPPoE sessions:

  config>service>vpls>sap# pppoe-user-db local-user-db-name

  config>service>ies>sub-if>grp-if>pppoe# user-db local-user-db-name

When RADIUS or DIAMETER authentication is also required after local user database authentication, then the authentication policy must be specified in the local user database. In this case, no authentication policy can be configured at the group-interface context. For example:

- IPoE sessions

  config>**subscr-mgmt**>**loc-user-db**>ipoe>**host**# **auth-policy** *policy-name*

  or

  config>**subscr-mgmt**>**loc-user-db**>ipoe>**host**# **diameter-auth-policy** *policy-name*

- PPP sessions

  config>**subscr-mgmt**>**loc-user-db**>**ppp**>**host**# **auth-policy** *policy-name*

  or

  config>**subscr-mgmt**>**loc-user-db**>**ppp**>**host**# **diameter-auth-policy** policy-name

The MSAP parameters are configured at the local user database host context. For example:

```
config>subscr-mgmt>loc-user-db>ipoe>host# msap-defaults
config>subscr-mgmt>loc-user-db>ppp>host# msap-defaults

  - msap-defaults

[no] group-interface - Configure the group interface
[no] policy - Configure the MSAP policy
[no] service - Configure the service
```

9.6.2.2 Explicit MSAP parameters from RADIUS or DIAMETER authentication

When RADIUS or DIAMETER authentication is required to return the MSAP parameters without prior local user database authentication, then the authentication policy should be configured at the capture SAP context. In a Bridged CO model, the authentication policy specified in the capture SAP is also used for the MSAP in the VPLS service. In a Routed CO model, the same authentication policy must also be configured at the group-interface context. For example:

```
config>service>vpls>sap# authentication-policy <auth-policy-name>
```

```
config>service>ies>sub-if>grp-if# authentication-policy <auth-policy-name>
or
config>service>vpls>sap# diameter-auth-policy <auth-policy-name>
config>service>ies>sub-if>grp-if# diameter-auth-policy <auth-policy-name>
```

The MSAP is not created if the group interface name returned from RADIUS or DIAMETER has a different authentication policy than the authentication policy configured at the capture SAP.

Table 29: RADIUS attributes/DIAMETER AVPs for MSAP parameters lists the RADIUS attributes (VSAs) and DIAMETER AVPs required to obtain MSAP parameters in the authentication phase.

Table 29: RADIUS attributes/DIAMETER AVPs for MSAP parameters

Attribute name	Type	Purpose and format
Alc-MSAP-Serv-Id [26-6527-31]	Integer	The service ID of the service context in which the MSAP is created.
Alc-MSAP-Policy [26-6527-32]	String	The name of the policy that defines the MSAP parameters.
Alc-MSAP-Interface [26-6527-33]	String	The name of the group interface context in which the MSAP is created.
Alc-MSAP-Serv-Name [241.26.6527.90]	String	(RADIUS only) The service name of the service context in which the MSAP is created. Alc-MSAP-Serv-Name takes precedence over Alc-MSAP-Serv-Id if both are specified.

9.6.2.3 Implicit MSAP parameters specified at the capture SAP

MSAP parameters that are not obtained from a local user database lookup, and that are not returned from RADIUS or DIAMETER can be specified in the **msap-defaults** section of the capture SAP context (this is a last resort scenario):

```
config>service>vpls>sap# msap-defaults ?
  - msap-defaults
[no] group-interface - Configure the group interface
[no] policy - Configure the MSAP policy
[no] service - Configure the service
```

9.6.3 MSAP creation

MSAPs can be created in IES or VPRN group interfaces (Routed CO model) and in a VPLS service (Bridged CO model).

An MSAP is persistent when subscriber-mgmt persistence is enabled. The MSAP parameters are part of the subscriber record.

If local user database, RADIUS, or DIAMETER authentication did not provide all the required information to create the subscriber host or session (no IP address for example), then the MSAP is created with a short

timer while waiting for the host to acquire the missing information. If no host is instantiated when the timer expires, the MSAP is deleted.

Multiple subscribers, subscriber hosts or sessions can share a single MSAP. The MSAP is created with the first instantiated subscriber host or session and deleted when the last associated subscriber host or session is removed from the system. Note that only a single MSAP policy can be specified for a MSAP. An attempt to change the MSAP policy by a new subscriber host or session for an existing MSAP results in a host or session setup failure.

MSAPs can be created in a wholesale VPRN service while the corresponding subscriber host or session is terminated in a retail VPRN or IES service. Both wholesale MSAP parameters (service, group interface, and policy) and the retail service ID must be provided during authentication.

9.6.4 MSAP QoS configuration

MSAPs are always used in combination with subscriber management. Subscriber traffic QoS models are defined in policies associated with the sla-profile and sub-profile and result in the instantiation of subscriber queues and policers used for subscriber traffic forwarding. The default QoS policies associated with MSAPs instantiate a single ingress and a single egress queue per MSAP for IES and VPRN services. For VPLS services, an additional ingress multipoint queue is instantiated per MSAP.

These MSAP queues have limited use and can be suppressed in most cases. For single-subscriber MSAPs, the MSAP queues can be suppressed with the **sub-sla-mgmt single-sub-parameters profiled-traffic-only** CLI command.

The default QoS policy associated with MSAPs may need to be changed to accommodate different scenarios. For example:

- Saving queue resources when **profiled-traffic-only** cannot be used, such as when more than one subscriber is active on an MSAP

  By mapping all forwarding classes to a policer in the QoS policy associated with an MSAP, a single policer instead of a queue is instantiated on the MSAP.

- Providing adequate QoS treatment for multicast traffic in a per MSAP replication mode

  Egress multicast traffic in per MSAP replication mode is forwarded by the MSAP queues or policers. Multicast traffic can be mapped into a dedicated queue or policer. The MSAP queue can be port-parented to provide scheduling priority on the port level.

The QoS policies associated with an MSAP are configured in the MSAP policy.

9.6.5 Sticky MSAP

After a subscriber session ends, the MSAP is removed from the system and the historical data of the subscriber is deleted. Sticky MSAP allows the MSAP to remain even when the subscriber session ends. This feature is only recommended for service providers who do not oversubscribe MSAPs in the network.

Sticky MSAP provides the following benefits.

- Because the sticky MSAP is never deleted, the subscriber can start a session faster; processing time is reduced because the MSAP does not have to be recreated.
- The MSAP may contain valuable historical information for the service provider. Keeping the MSAP provides a means for the service provider to look up subscriber historical data.

The MSAP is only eligible for stickiness if it was successfully created. The sticky MSAP introduces a new state: idle. An idle MSAP indicates that the subscriber on the MSAP has disconnected and the MSAP is ready for a new subscriber connection. An example is shown below:

```
A:BNG> # show service sap-using
===============================================================================
Service Access Points
===============================================================================
PortId                          SvcId      Ing.  Ing.    Egr.  Egr.   Adm  Opr
                                           QoS   Fltr    QoS   Fltr
-------------------------------------------------------------------------------
 [1/1/20:1841](I)                1000       1     none    1     none   Up   Up
 1/1/1:4000                      1000       1     none    1     none   Up   Up
-------------------------------------------------------------------------------
Number of SAPs : 20
-------------------------------------------------------------------------------
Number of Managed SAPs : 1, indicated by [<sap-id>]
Flags : (I) = Idle MSAP
-------------------------------------------------------------------------------
===============================================================================
```

There are two ways to remove sticky MSAPs from the system:

- **Manually**

  The **clear service id** *id* **msap** command removes MSAPs. This command can remove MSAPs with active subscribers. To clear only MSAPs without any active subscriber, use the keyword **idle-only**.

- **Automatically**

  Sticky MSAPs can be removed with the **sticky-msaps-idle-timeout** command if they are idle for longer than the specified time. This can be used to keep only MSAPs that are used by regular subscribers and free the system from consuming MSAPs resources used by occasional subscribers.

The **clear service id** *id* **msap** command removes MSAPs.

Note:

- This command can remove MSAPs with active subscribers. To clear only MSAPs without any active subscriber, use the keyword **idle-only**.
- Persistence restoration relies on configured **msap-defaults** parameters under capture SAP (**config**>**service**>**vpls**>**sap**>**msap-defaults**)

  With persistence enabled, it is generally recommended to avoid changing the default after the system has created hosts with these **msap-defaults** values. The hosts are not restored by the system as the **msap-defaults** values are no longer the same.

  The Sticky MSAP feature keeps the failed MSAPs on the system and consumes system resources. These MSAPs can be cleared with the **clear**>**service**>*id service-id*>**msap** command.

9.7 ESM identification process

9.7.1 SAP-ID ESM identifier

Providers migrating from Basic Subscriber Management (BSM) can assign a subscriber to a SAP. The SAP ID ESM identifier makes the transition easier by allowing the operator to continue using the *sap-id* as a subscriber-ID.

An ESM SAP ID provides the system the ability to:

- Provide access to the SAP ID string in the Python script.
- Allow the automatic assignment of the SAP ID to a static subscriber or subscriber host.

9.7.2 DSLAM-ID

A DSLAM ID provides the system the ability to define a DSLAM-ID string provided through the Python script, RADIUS, or local user database. If the DSLAM-ID was provided, but the subscriber host is instantiated on a regular MDA, the DSLAM-ID is ignored.

The ability to aggregate subscribers into DSLAMs for the purpose of QoS, can use the SAP ID to identify subscribers and associated DSLAMs.

9.8 Default subscriber

This feature provides a default subscriber definition under the SAP. If the object was configured the operator may use ESM without enabling a processing script or a RADIUS authentication policy. In the event both have been disabled any host that was installed for the SAP is installed with the configured default subscriber ID. If a RADIUS policy was used or if a script was enabled but a subscriber ID was not returned the default subscriber ID is used.

9.9 Subscriber mirroring

This section describes mirroring based on a subscriber match. Enhanced subscriber management provides the mechanism to associate subscriber hosts with queuing and filtering resources in a shared SAP environment. Mirroring used in subscriber aggregation networks for lawful intercept and debugging is required. With this feature, the mirroring capability allows the match criteria to include a subscriber-id.

Subscriber mirroring provides the ability to create a mirror source with subscriber information as match criteria. Specific subscriber packets can be mirrored mirror when using ESM with a shared SAP without prior knowledge of their IP or MAC addresses and without concern that they may change. The subscriber mirroring decision is more specific than a SAP. If a SAP (or port) is placed in a mirror and a subscriber host of which a mirror was configured is mirrored on that SAP, packets matching the subscriber host are mirrored to the subscriber mirror destination.

The mirroring configuration can be limited to specific forwarding classes used by the subscriber. When a forwarding class (FC) map is placed on the mirror only packets that match the specified FCs are mirrored. A subscriber can be referenced in maximum 2 different mirror-destinations: 1 for ingress and 1 for egress.

9.10 Multicast management

The multicast-management CLI node contains the **bandwidth-policy** and **multicast-info-policy** definitions. The bandwidth-policy is used to manage the ingress multicast paths into the switch fabric. The multicast-info-policy is used to define how each multicast channel is handled by the system. The policy may be used by the ingress multicast bandwidth manager, the ECMP path manager and the egress multicast CAC manager.

9.11 Volume and time-based accounting

Volume and time-based accounting includes the following components:

- Metering function performing stateful monitoring of the service delivery to the subscriber.
- Communication with an external management system that gets and updates credit per subscriber, notifications of credit exhaustion, and so on.
- Action on credit exhaustion takes pre-defined action when the credit has been exhausted.

9.11.1 Metering

Metering represents the core of time and volume-based accounting. Service usage is typically measured by performing an accounting of the traffic passing through corresponding subscriber-host queues (volume usage) or by keeping lease-state while the specified subscriber host is connected to the network (time usage).

- **Statefullness**

  The accounting information is compared with pre-defined credit expressed in terms of time or volume to monitor service usage.

- **Sensitivity**

  Defining so called activity-threshold allows distinction between subscriber-host being connected and subscriber-host effectively using the service. This is particularly of interest in cases of time-based charging.

- **Aggregated usage per-category per-subscriber-host**
  Accounting information can be reported on **per-queue per-sla** instance of the specified subscriber. In many situations, a specific level of aggregation (such as a per-subscriber or HSI ingress and egress traffic) is required to perform meaningful mechanism for pre-paid services.

9.11.1.1 Category map and categories

This feature uses an object **category-map** which defines individual aggregates (such as data in and out, video and data, and so on) and their mapping to individual forwarding queues.

The following output depicts a category-map configured in the subscriber management context.

```
*A:ALA-48>config>subscr-mgmt# info
----------------------------------------------
...
        category-map "triple-play" create
            category "data" create
```

```
                queue 1 ingress-egress
            exit
            category "video" create
                queue 2 egress-only
            exit
            category "voice" create
                queue 3 ingress-egress
            exit
        exit
        category-map "aggr-subscriber-service" create
            category "data-services" create
                queue 1 ingress-egress
                queue 3 egress-only
            exit
        exit
...
-----------------------------------------------
*A:ALA-48>config>subscr-mgmt#
```

Based on a category-map the system gathers usage information (volume/time) on a per-sla-instance-per-category basis. To do so, statistics of all queues and policers forming the category of the specified sla-instance are aggregated.

- **Single subscriber host** (routed CPE)

  Single SLA instance.

- **Multiple subscriber hosts on the same SAP** (bridged CPE)

  Single SLA instance. Several hosts use the same credit and the renewal of one causes renewal for all.

- **Multiple subscriber hosts on different SAP** (bridged CPE)

  SLA instance per host.

The per-category usage gathered as described above is compared with per-subscriber-host-per-category credit and when credit is exhausted several actions can be taken.

There are several category-maps pre-configured on the system. The category-map applicable to a specific subscriber-host is derived at the host creation from the RADIUS VSA in an authentication-response, Python script, or static configuration in the local-user-database. All subscriber-hosts belonging to the same subscriber and created on the same SAP (therefore, sharing the same sla-instance) must use the same category-map. In case of conflict, (an existing subscriber host has a different category-map than the one derived for the new host) the category-map of the last host is applied to a specific sla-instance. As a consequence, all previous information related to the status of the credit is lost.

There can be multiple queues and policers aggregated into one category. There can be up to sixteen categories in a category map.

9.11.1.2 Quota consumption

There are two types of quota (credit), volume and time. In volume usage monitoring, the system accumulates byte counters per category-sla-instance and compares it with the assigned quota. After the credit is exhausted (or threshold for renewal is met) the system attempts to renew it with corresponding management system.

In time-based credit, the distinction between active-usage and active-connection is made by defining an activity-threshold, where an object defines an average data rate under which the subscriber-host is considered silent.

If the effective rate of the application usage does not exceed the rate defined by the activity-threshold, the specified subscriber host is considered silent and its corresponding credit is not used. If the application usage exceeds the rate, the application-credit is consumed (in terms of time).

9.11.1.3 Minimum credit control quota values

The minimum credit control quota values are one second for time quota and one byte for volume quota. These minimum values are not realistic deployment values for multiple reasons such as effective sampling periods, statistics processing time, RADIUS message load, subscriber scale, and so on.

For typical deployment scenarios it is not recommended to implement Credit Control quota values smaller than 60 seconds for time quota and for volume quota the volume that can be consumed in 60 seconds for that category (function of number of queues/policers monitored and their respective rates).

9.11.1.4 RADIUS VSA Alc-Credit-Control-Quota

The quota in the RADIUS VSA Credit-Control-Quota uses this fixed format:

Alc-Credit-Control-Quota = "<*volume-quota*>|<*time-quota*>|<*category name*>"

- *volume-quota* is specified in bytes (B), in kilobytes (K or KB), in megabytes (M or MB), in gigabytes (G or GB)
- *time-quota* is specified in seconds (s), in minutes (m), in hours (h), in days (d) or a combination (5m30s). A lower unity may never exceed the higher unity: 5m60s is not allowed and should be specified as 6m.

Both volume and time quota should be specified in the attribute but only one credit type (volume or time) is applied per category. The credit-type of a category is configured in the category-map CLI context.

For example, use Alc-Credit-Control-Quota = "0|1h30m|cat1" to grant time quota and Alc-Credit-Control-Quota = "1G|0|cat2" to grant volume quota.

9.11.2 Credit negotiation mechanisms

The per-subscriber per-category credit can be obtained by several ways:

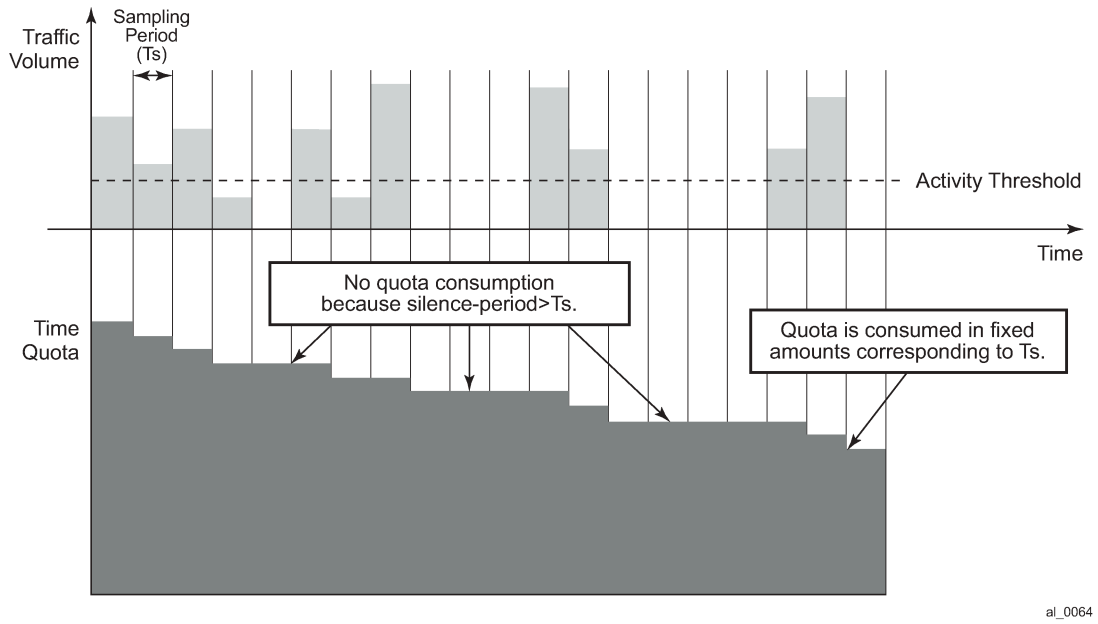
- RADIUS during authentication process.
- Static configuration - configured in the **config**>**subscr-mgmt**>**category-map**>**category** context.

Credit can be expressed by either

- Volume
- Time

The renewal of the credit using RADIUS authentication is triggered by credit exhaustion or (if configured) by depletion of the credit to exhausted-credit-threshold level. If this occurs, the system sends a RADIUS authentication message indicating the corresponding category and usage. The following are several possibilities for the RADIUS server response (as shown in Figure 125: Threshold configured/not configured):

Figure 125: Threshold configured/not configured



1. **No authentication response**

   The system installs out-of-credit action after the original credit has been used.

2. **Authentication response with reject**

   The corresponding host is removed after the original credit has been used.

3. **Authentication response with accept and no credit VSA included**

   The system installs out-of-credit action.

4. **Authentication response with accept and credit VSA included**

   The credit is installed.

   The new installed credit is reduced by the amount of credit consumed during credit renewal (in other words, between the start of the credit renewal and reception of an authentication-response). In case the new received credit is less than the credit consumed during credit renewal, then the out-of-credit-action is installed instead.

5. **Authentication response with accept and credit VSA included**

   The out-of-credit installed. The new credit is always reduced by the amount of credit consumed in time between renewal has been initiated and authentication-respond has been received. In case of a negative result (the newly receive credit is smaller than the amount consumed in the meantime) the test is installed.

To identify that the specified RADIUS-auth request is related to credit renewal instead of plain authentication, the node includes empty credit VSAs, depending on categories which has been exhausted. The RADIUS server can identify which category has requested credit renewal.

9.11.3 Action on credit exhaustion

System supports configurable actions after the credit for the subscriber is exhausted:

- Sends an SNMP trap and continue (the credit-usage counter is reset).
- Disconnect.
- Changes to a pre-defined service level (such as adjusting the queue rate).
- Blocks the category.

9.11.4 Action on error-conditions

During credit negotiation, the number of errors can occur which can lead to a specific subscriber-host category with no new credit renewed. This is different from credit exhaustion where a separate configurable action is taken. The following occurs:

- sends an SNMP trap and continues
- sends a trap and blocks the category

9.11.5 Applicability of volume and time-based accounting

Volume and time-based accounting are applicable to the ESM mode of operation only. Using credit control concept is not mutually exclusive to other accounting methods. In many network implementations the more traditional accounting methods such as XML file or RADIUS accounting is still used in a combination with the credit concept but with larger intervals. This is helpful when providing overviews of the average usage and service utilization.

9.12 Subscriber host idle timeout

An idle timeout is the maximum time that a subscriber session can be idle before the session is terminated or a connectivity check is started. Idle timeout applies to PPPoE and IPoE hosts.

The time/volume based accounting model is used to configure an idle timeout.

Create a category-map (see Category map and categories)

- Define a category with queues and policers to be monitored for activity (packets being forwarded).
- An activity threshold (in kb/s) must be configured for idle timeout to take effect. The activity threshold suppresses background traffic (for example control flows) from activity monitoring.

The following in an example of a category map configuration:

```
config>subscr-mgmt
        category-map "idle-timeout" create
            activity-threshold 25
            category "cat-1" create
                queue 1 ingress-egress
            exit
        exit
```

In the sla-profile, associate the category-map and optionally define

- An idle-timeout (60 to 15552000 seconds). The default is infinite (no idle-timeout).

  The idle-timeout can also be specified from RADIUS in an access-accept or CoA message with the [28] Idle-Timeout attribute. A RADIUS specified idle-timeout overrides the CLI- configured value. The values outside the limits are accepted but rounded to these boundaries.

Table 30: Idle-timeout attribute

Attribute ID	Attribute name	Type	Limits	Purpose and format
28	Idle-Timeout	integer	60 to 15552000 seconds	0 = infinite (no idle-timeout) 60 to 15552000, in seconds For example: Idle-Timeout = 3600

- An idle-action:

  shcv-check

  Perform a subscriber host connectivity check (IPoE hosts only). Host connectivity verification should be enabled on the corresponding group-interface for the **idle-timeout-action shcv-check** to take effect:

  configure>**service ies** | **vprn** *service-id* **subscriber-interface** *ip-int-name* >**group-interface** *ip-int-name*>**host-connectivity-verify**

  If the SHCV check is successful, the subscriber host is not disconnected, and the idle-timeout timer is reset to zero. If the SHCV check fails, the subscriber host is disconnected (same as terminate).

  For PPP hosts, the **idle-timeout-action shcv-check** is ignored and has the same effect as **idle-timeout-action terminate**.

  IPoE:

  Terminate (default): disconnect the subscriber hosts

  – Delete the subscriber host

  – Send a DHCP release message to the DHCP server

  – Send an Accounting Stop message to the RADIUS accounting server

  PPP:

  – Delete the subscriber host

  – Send a terminate request message to the CPE

  – Send an Accounting Stop message to the RADIUS accounting server

Example

```
config>subscr-mgmt
        sla-profile "sla-profile-1" create
            category-map "idle-timeout"
                category "cat-1" create
                    idle-timeout 3600
                    idle-timeout-action terminate
                exit
            exit
        exit
```

At host instantiation, a timer is initialized to the *idle-timeout* value (one timer per sla-profile instance). Each queue or policer in the category is monitored for activity over a fixed polling interval:

During the polling interval:

- If the forwarding rate falls below the configured activity threshold then the timer is deducted by the polling interval (time elapsed).
- If the forwarding rate is above the configured activity threshold then the timer is initialized to the idle-timeout value.

When the timer becomes zero, the **idle-timeout-action** is performed for all hosts associated with the SLA-profile-instance (all hosts from a subscriber on a single sap and that share the same sla-profile).

9.13 Web portal authentication

9.13.1 HTTP-redirect (captive portal)

A captive portal service can be created with an HTTP redirect action in an IP filter. The customer's request to the intended recipient is blocked and the customer is forced to connect to the service's portal server. See HTTP-redirect (Captive Portal) in the *7450 ESS, 7750 SR, 7950 XRS, and VSR Router Configuration Guide* for details.

9.13.2 One-time HTTP redirection overview

With one-time HTTP redirection enabled, after an ESM host is created, only the first HTTP request from the host is redirected to a configured URL with specified command options. Subsequent HTTP requests go through without being redirected.

Service providers can use this feature to push a web page to broadband users for the purpose of advertisement, announcements, and such.

A **one-time-http-redirection** filter configured in the **sla-profile** is installed as an ingress IP filter of a subscriber host until the first HTTP request is redirected. The ingress IP filter configured in the **sla-profile** or an active filter override replaces the associated ingress IP filter.

The [26.6527.136] Alc-Onetime-Http-Redirection-Filter-Id and [245.26.6527.7.5] Alc-Sub-Ipv4-Onetime-Http-Redirect-Filter-Name RADIUS attributes included in an Access-Accept or CoA message override the **one-time-http-redirection** filter configured in the **sla-profile**. In case of a CoA message, if the **one-time-http-filter** of the host was already replaced, the system ignores subsequent Alc-Onetime-Http-Redirection-Filter-Id and Alc-Sub-Ipv4-Onetime-Http-Redirect-Filter-Name overrides. For more information, see the *7450 ESS, 7750 SR, and VSR RADIUS Attributes Reference Guide*.

If the router receives shared or host specific filter inserts in a CoA or Access-Accept message when a **one-time-http-redirection** filter is still active, the new filter entries are applied to the ingress filter, but only installed for the subscriber host after the first HTTP redirection.

This feature only supports IPv4 filters.

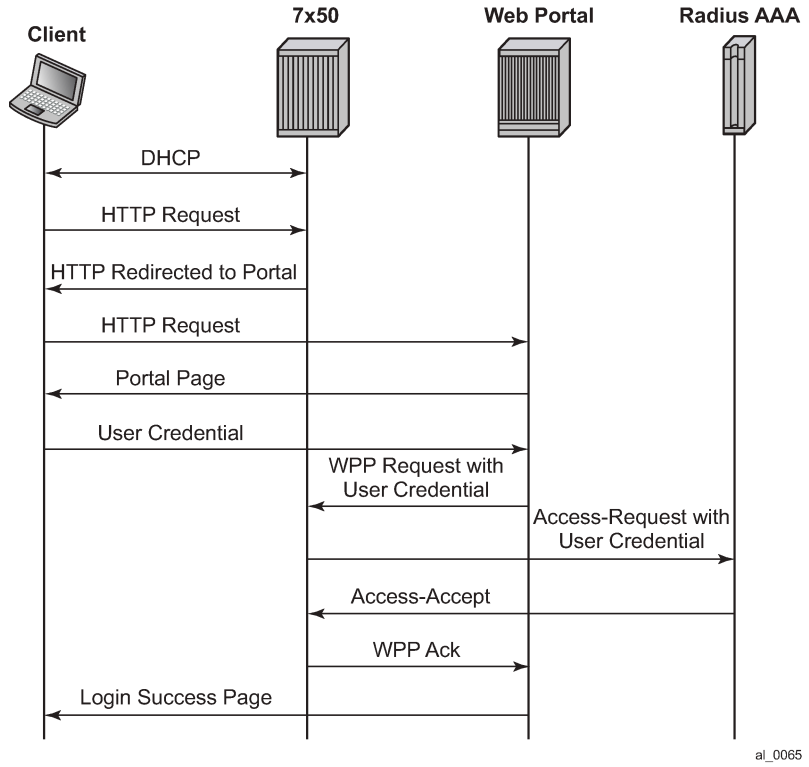
Note: Filter name ([245.26.6527.7.5] Alc-Sub-Ipv4-Onetime-Http-Redirect-Filter-Name) and filter ID ([26.6527.136] Alc-Onetime-Http-Redirection-Filter-Id) overrides must not be mixed during the lifetime of a subscriber host or session.

9.13.3 Web authentication Protocol (WPP)

The Web Authentication Protocol (WPP) is a protocol running between a BNG and a Web portal server. WPP is used for web portal authentication of WLAN users (DHCP Host). It can function like a web portal that can trigger BNG to perform RADIUS authentication for WLAN users, or send user disconnection notification to BNG.

The Figure 126: WPP authentication illustrates high level of call flow of WPP authentication.

Figure 126: WPP authentication



The following describes WPP authentication call flow:

1. When the WLAN user starts a DHCP exchange with a 7750 SR, the router creates a DHCP host from following configurations:
   - Sub-id is the default subscriber ID configured in the **sap**>**sub-sla-mgmt** context.
   - sla-profile/sub-profile/aa-profile takes the configuration from CLI command **grp-if**>**wpp**>**initial-sla-profile/initial-sub-profile/initial-app-profile**.
   - IP address from local or external DHCP server is assigned to the host.
2. When the user sends an HTTP request to visit a website by browser, the router redirects the HTTP request to the web portal.
3. The portal server sends an authentication page to the WLAN user.
4. WLAN user enters username and password in the authentication page and submit to the portal server.
5. The portal server sends a WPP request to router together with the user credentials.

6. The 7750 SR sends an access-request to RADIUS server with user credentials.
7. RADIUS returns an access-accept if authentication succeeds.
8. The 7750 SR returns a WPP ACK to the portal server.
9. If it was access-accept, then the router can optionally override the following host properties:
   - **sub-id**

     This is the subscriber ID from RADIUS. If there is no sub-id from RADIUS, then the host keeps using current sub-id.
   - **sla-profile, sub-profile, or aa-profile**

     The system uses the RADIUS server returned values. If the RADIUS server did not return these then the system tries to use the LUDB (in local DHCP server) return values if they are available. If not, the system tries to use the default values configured under SAP.

9.13.3.1 WPP configurations

A minimal WPP configurations must include the following:

- WPP portal server — Specifies the name and IP address of the WPP portal server.
- Enable WPP under the group-interface:
  - WPP portal server that system should listen to.
  - **authentication-policy** on **group-interface** that specifies address of RADIUS server.
  - **def-sub-id** under **sap>sub-sla-mgmt** that is used for DHCP host before user is authenticated by portal server.
  - **initial-sla-profile** and **initial-sub-profile** that are used for the DHCP host before user is authenticated by portal server. The **initial-sla-profile** should include a ingress filter that has **http-redirection** entry.

The following is an example configuration:

```
#--------------------------------------------------
echo "Web Portal Protocol Configuration"
#--------------------------------------------------
        wpp
            portals
                portal "portal-1" address 10.9.9.9 create
                    no shutdown
                exit
            exit
            no shutdown
        exit
config>service>vprn# info
----------------------------------------------------

            subscriber-interface "sub-if" create
                address 192.168.10.1/24
                group-interface "grp-if" create
                    dhcp
                        server 10.1.1.1
                        gi-address 192.168.10.1
                        no shutdown
                    exit
                    authentication-policy "radius-auth"
```

```
                    sap 1/1/9 create
                        sub-sla-mgmt
                            def-sub-id "WLAN-User-Unauth"
                            no shutdown
                        exit
                    wpp
                        initial-sla-profile "webportal"
                        initial-sub-profile "webportal"
                        portal router "Base" name "portal-1"
                        no shutdown
                    exit
                exit
            exit
…
----------------------------------------------
```

9.13.3.2 WPP triggered host creation

In some cases, a 7750 SR can sit behind a Layer 3 device (such as an CMTS), where the router does not participate in client's DHCP process. Such a use case is different from a normal WPP use case where the routers rely on getting client's DHCP request to create an initial ESM host.

This feature allows the system to create an ESM host upon successful WPP authentication without creating an initial host.

In the above use case (behind a Layer 3 device) the user also needs to configure one or more default hosts on the SAP to allow HTTP redirection without an ESM host. The default-host subnet is the user's source subnet and the next hop address is the Layer 3 device's interface address that connect to the SAP. Users also need to configure the **lease-populate l2-header** command in the **grp-if**>**dhcp** context to make HTTP redirection with default-host work. The **grp-if**>**dhcp** context could be shut down in the meantime.

This feature does not work with wholesale/retail.

9.13.3.2.1 LUDB support for WPP

The SR OS supports LUDB lookup for WPP authentication. Users can optionally configure LUDB using the **grp-if**>**wpp** context to return the WPP-related configuration attributes (such as a portal name, initial-sla-profile, initial-sub-profile, and so on) for an IPoE host. The system can access LUDB when creating the initial host before WPP authentication. The LUDB returned attribute overrides the corresponding configuration under the **group-interface** context.

A LUDB lookup is performed by the system in the following cases.

- when a host is created
- when the system restores a host after a disconnect

If the WPP LUDB lookup returns an authentication policy, it is used for WPP RADIUS authentication. When WPP LUDB is configured, the authentication policy on **group-interface** is optional and only used by the WPP if there is no authentication policy returned from the WPP LUDB lookup.

9.13.3.3 WPP multichassis redundancy support

The SR OS supports multichassis redundancy to WPP. This can be achieved by doing following:

- Create a loopback interface on both 7750 SRs with the same IP address X.

- Use the **track-srrp** parameter while configuring address X to track the corresponding SRRP instance.
- Configure a portal with the same name and same service-id on both nodes to send WPP packets to the destination address.
- Use an route-policy to export X to the routing protocol. The metric the route X can be set is based on the specified SRRP state (master or non-master) so that the active node can advertise route X with a better metric. Then the WPP packet from the portal is attracted to the active node.
- Only the active node process WPP packet, however in case of standby node receives (such as routing is still re-converging) the WPP packet, then the standby shunts the WPP back to the active node (SRRP master state).
- WPP hosts are synced by MCS.

9.13.3.4 WPP portal group

A WPP portal group allows users to configure up to eight WPP portals in a portal group. The system can receive portal-initiated WPP request packets from any configured portal in the portal group. When the system must initiate a WPP NTF_LOGOUT message, it sends a NTF_LOGOUT message to all configured portals in the portal group, and the first received ACK_LOGOUT stops retransmission of the NTF_LOGOUT message.

A WPP portal group can be used to achieve WPP portal redundancy:

- Each portal is only allowed to be configured in a single portal group
- Each can be in a portal group and be used as an individual portal simultaneously
- Mixed WPP versions (version 1 and version 2) of portals are allowed in the same portal group

This feature is also supported for WPP triggered hosts and SRRP/MCS.

9.13.4 WPP support for IPv6

WPP support for IPv6 includes the following:

- WPP over IPv6; for example, a BNG and portal can exchange WPP messages over IPv6
- User client access portal over IPv6
- Dual-stack IPoE sessions. This means that the user client can be a dual-stack device, such as:
  - IPv4 only
  - IPv6 only
  - IPv4 and IPv6

  For IPv6, the client can use a DHCPv6 IANA or SLAAC address to access the portal.

  Only the first address assignment of an IPoE session triggers WPP authentication. Subsequent address assignments in the same session do not require authentication. If IPoE reauthentication is enabled, when the system requires reauthentication of the client, the system restores the SLA profile or subscriber profile session to the initial WPP profile. This causes the client to be redirected to the portal again to authenticate.
- WPP portal redundancy (**portal-group**) for IPv6 portal and dual-stack IPoE session
- WPP LUDB support for dual-stack IPoE sessions

- WPP MCS redundancy for dual-stack IPoE sessions
- Triggered dual-stack IPoE sessions for Layer 3 access

  For these types of sessions, by default, each address of same client creates a separate IPoE session unless the **ipoe-session-policy circuit-id-from-auth** command is enabled and the RADIUS server returns a circuit ID during WPP-triggered RADIUS authentication.

9.13.5 WPP other details

Nokia recommends using an IPoE session with WPP for the best support. In addition to the feature difference between a normal IPoE host and session, IPv6 is only supported for WPP IPoE session, not for an IPoE host.

For IPoE (non-session) hosts, a DHCP renew triggers reauthentication of the host.

9.14 ESM over MPLS pseudowires

This feature allows IPoE and PPPoE (terminated or L2TP tunneled) subscriber sessions to be backhauled through an Ethernet aggregation network using MPLS pseudowires terminating directly on the BNG. The MPLS pseudowire originates from the first hop aggregation PE (referred to as access PE) upstream of the AN (or directly from a multiservice AN), and terminates on the BNG. Multiple subscriber sessions from a specific access-port on the Access-PE can be backhauled over a single P2P MPLS pseudowire toward the BNG. This capability allows the network to scale and does not require a MPLS pseudowire per subscriber between Access-PE and the BNG. The access-port on the Access-PE can be dot1q, q-in-q, or NULL encapsulated. The BNG terminates the MPLS pseudowire, decapsulates the received frames, and provides ESM functions including HQoS, without requiring an internal or external loopback. Each MPLS pseudowire is represented on the BNG as a "PW port" for which SAPs are created. A PW port can be configured with capture SAP. Both static and managed SAPs are supported. The underlying Ethernet port is required to be in hybrid mode. The feature set is supported for FP3 and later. This feature is supported on the 7750 SR and 7450 ESS.

Figure 127: ESM over MPLS pseudowire example

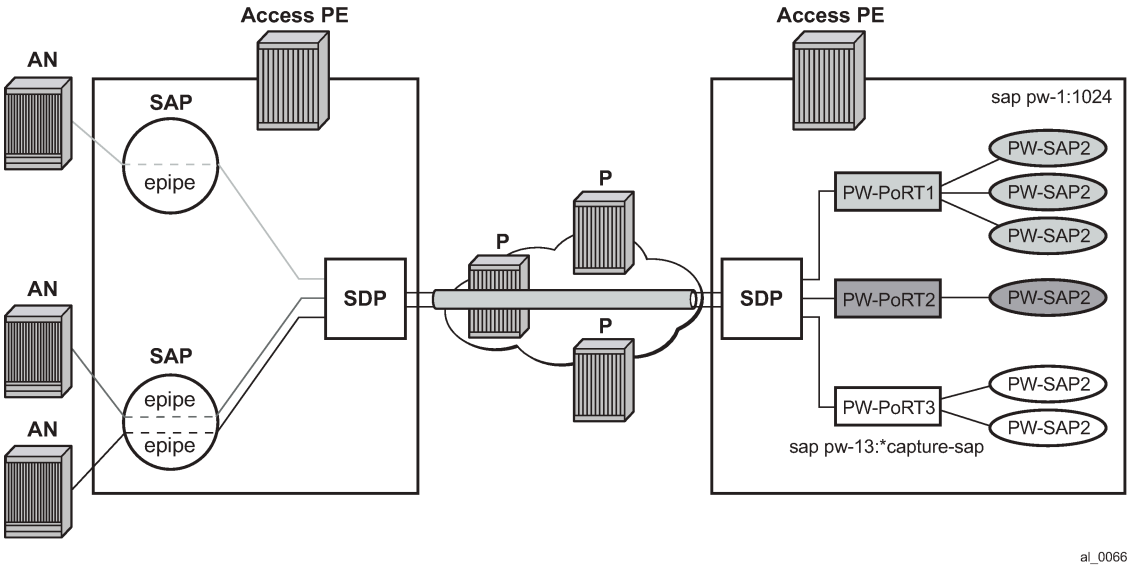
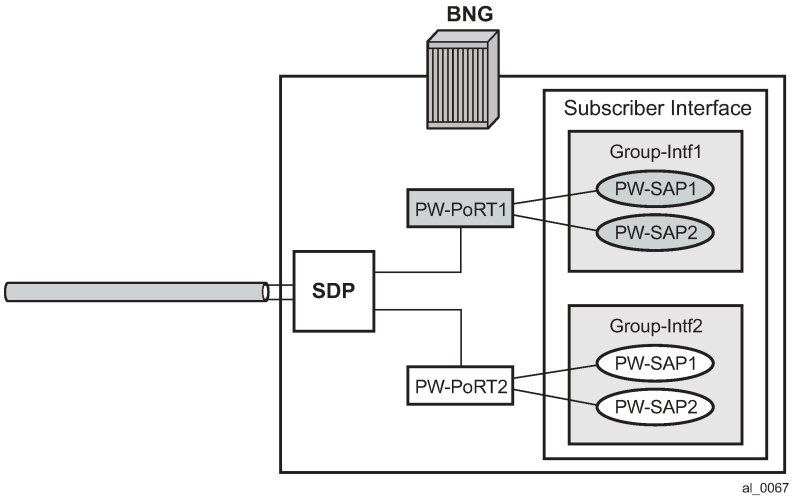


Figure 128: Group interface example



9.14.1 ESM over PW ports

In BNG deployments, tunneled traffic is typically terminated on PW ports where the payload is extracted and processed by ESM on PW SAPs. There are two modes of operation for PW ports in SR OS:

- When a PW port is bound to a specific physical port. A successful mapping between the tunnel and the PW port requires that the tunnel terminates on the same faceplate port (I/O port) to which the PW port is bound. In this mode of operation, PW ports do not support rerouting of tunnels between the I/O ports. For example, if a tunnel is rerouted to an alternate physical port because of a network failure, the PW port becomes non-operational. The only supported tunnel on a fixed PW port is an MPLS based PW.
- When a PW port is independent of the faceplate port (I/O port) on which the tunnel is terminated. The PXC based PW port is anchored or bound on the termination side of a port cross connect PXC (which can be either a PXC sub-port or a PXC LAG) under an internally created forward path extension (FPE) PW. The benefit of this type of PW port is that it provides services where the tunnel is switched between the I/O ports because of a network failure.

For more information about PW ports, see the Pseudowire Ports section in the *7450 ESS, 7750 SR, 7950 XRS, and VSR Layer 2 Services and EVPN Guide*.

The router supports the MPLS entropy label (as specified in RFC 6790, *The Use of Entropy Labels in MPLS Forwarding*) on fixed PW ports. This allows LSR nodes in a network to load-balance labeled packets in a more granular fashion than allowed by simply hashing on the standard label stack. For more information, see the *7450 ESS, 7750 SR, 7950 XRS, and VSR MPLS Guide*, *Entropy Label*.

9.14.1.1 ESM on PW port bound to a physical port

9.14.1.1.1 QoS support

QoS is supported for ESM over PW SAPs as with ESM over regular SAPs, and includes currently supported models.

- FC to queue mapping
- H-QoS
  - Per-subscriber H-QoS (service scheduler child to port-scheduler parent).
  - PW SAP queues attached to H-QoS scheduler by a **parent** statement.
  - Scheduler attached to port scheduler by a **port-parent** statement.
- Direct service queue to port-scheduler.

  Aggregate-rate-limit.

9.14.1.1.1.1 Bandwidth control at the PW port level by Vport

Bandwidth control per PW port (per AN or per AN/ per service) by Vport.

- The Vport can be created on the binding port.
- The Vport can be associated with the PW port either by static assignment or dynamic selection by inter-dest-id (returned from RADIUS or DHCP for a host).

- Aggregate-rate-limit can be configured to shape the egress traffic across all hosts associated with the Vport by inter-dest-sting match or by static association of underlying PW port with the Vport.

The following output displays a dynamic Vport selection based on an inter-dest-id configuration.

```
config>
   port 1/1/1
      ethernet
         mode hybrid
         encap-type dot1Q
         mtu 1540
         access
            egress
            vport “v1” create
               agg-rate
                     rate 1000
               host-match dest “dslam-1”    #### hosts will be associated with
            exit                            #### vport based on inter-dest-id
          exit
      exit
   exit

config>service>sdp>binding
   pw-port 11 vc-id 11 create
       egress
          shaping int-dest-id “dslam-1”   #### dynamic vport selection based on
                                          #### int-dest-id.
```

The following output displays a static assignment of PW port to Vport configuration.

```
config>
   port 1/1/2
      ethernet
         mode hybrid
         encap-type dot1Q
         access
            egress
                vport "v2" create
                    agg-rate
                        rate 1000
                    exit
                exit
            exit
        exit
    exit
config>service>sdp>binding
    pw-port 20 vc-id 20 create
       egress
          shaping vport “v2”    #### static assignment of pw-port to vport.
       exit
    exit
```

9.14.1.1.1.1.1 Last mile shaping

With normal Ethernet aggregation in the next-mile, when last-mile shaping is on, fixed encapsulation-offset is calculate based on the last-mile encapsulation type and the next-mile encapsulation (26 bytes with QinQ). This offset is applied to the frame, and the ATM overhead is then dynamically calculated on the adjusted size. The resulting dynamically calculated overhead in the datapath is then applied to the queue-rates and the subscriber aggregate-rate.

With this feature of backhauling subscriber sessions using MPLS PW in the aggregation network. The last mile does not see any MPLS PW overhead. The next-mile includes overhead because of the PW encapsulation. Therefore, when last mile shaping is enabled, the fixed encapsulation-offset is calculated based on the difference between last-mile encapsulation type and next-mile encapsulation. The next-mile encapsulation considers the additional PW overhead, which includes:

14B Ethernet header + [4B] (optional network interface Q-tag) + MPLS Labels (variable)

In the datapath the actual PW encapsulation overhead, considering the MPLS labels which could be variable (with FRR or PHP) is tracked, and is applied to the computed "encapsulation offset". This adjusted "encapsulation offset" is applied to the frame. The ATM overhead is then dynamically calculated on the adjusted size and applied for last mile shaping (to queue-rates and subscriber-aggregate-rate). Note that there is no change from ESM over normal SAPs, in how last-mile shaping is triggered or how the last mile encapsulation type is determined (by configuration in the egress context of the subscriber profile or dynamically learned from Access-Loop-Encapsulation sub-TLV in vendor specific PPPoE tags).

9.14.1.1.2 BNG redundancy with ESM over pseudowire

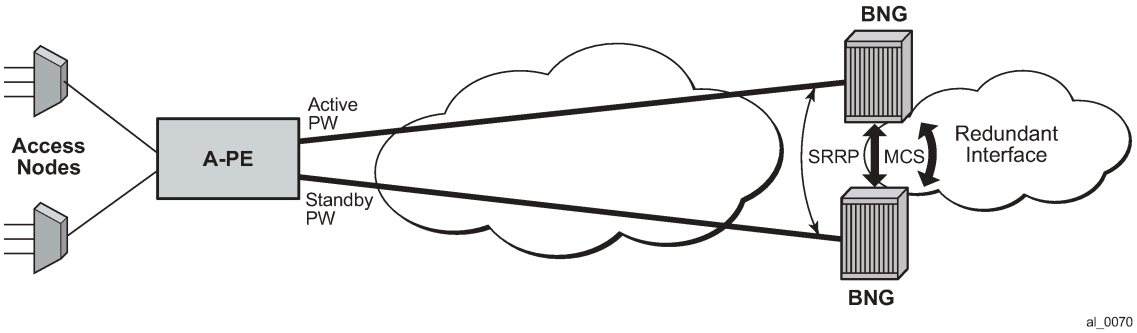
This feature provides support for stateful BNG redundancy. Such as when the far-end aggregation PE (A-PE) is dual-homed to two BNGs terminating subscriber sessions over MPLS pseudowires (PWs) that are initiated from the A-PE. The subscriber state between the BNG pair is synced using MCS.

9.14.1.1.2.1 Epipe-based aggregation service

For an Epipe based aggregation service, the redundancy is based on active/standby PWs from A-PE to the redundant BNG pair. A-PE signals active and standby pseudowire status to the BGN pair. An SRRP instance per PW port (group interface) is required on the BNG with a messaging SAP on each PW port. The SRRP instance on a PW port is in the SRRP master state when the PW is active and in the SRRP backup state when the PW is standby. This is achieved by tying the SRRP state to the state of the messaging SAP. The messaging SAP goes down when the underlying PW port goes down based on the PW status bit signaled by the A-PE.

In this model, there is no SRRP message exchange between the BNG pair, as there is no Layer 2 path between them. The purpose of SRRP is to get SRRP-aware routing for subscriber routes and managed routes, or to be able to use the redundant (shunt) interface. Downstream traffic for a subscriber that ingresses the standby BNG can only be shunted to the active BNG, if the corresponding subscriber interface on the standby BNG is operationally UP. This is achieved by creating a second empty group interface (without SAPs) on the same subscriber interface with the **oper-up-while-empty** command configured. Multiple PWs with endpoint configuration is not supported on the BNG.

Figure 129: BNG redundancy based on active/standby PW signaling



Sample configuration on an active BNG

```
config>
    pw-port 2 create
    exit

config>redundancy#
     multi-chassis
         peer 10.20.1.3 create
             source-address 10.20.1.2
                sync
                    srrp
                    sub-mgmt ipoe pppoe
                    port pw-2 sync-tag "tag2" create
                    exit
                    no shutdown
                exit
                no shutdown
            exit
        exit
     exit

config>service>ies#
       redundant-interface "redundant-interface" create
          address 10.10.30.2/24 remote-ip 10.10.30.3
          spoke-sdp 23:1000 create
             no shutdown
          exit
       exit
config>service#
    sdp 1 mpls create
       far-end 10.20.1.2
       ldp
       keep-alive
          shutdown
       exit
       binding
          port 1/1/1
          pw-port 2 vc-id 2 create
             vc-type vlan        #### default encaps-type dot1Q
             no shutdown
          exit
       exit
       no shutdown
    exit
config>service#
       subscriber-interface "subif" create
```

```
           address 10.11.1.2/16 gw-ip-address 10.11.1.1 populate-host-routes
                group-interface "grpif" create
                    authentication-policy "base_authpolicy"
                    redundant-interface "redundant-interface"
                    sap pw-2:1000 create
                        description "sap-grp-3"
                    exit
                    srrp 1 create
                        message-path pw-2:1000
                        no shutdown
                    exit
                    arp-host
                        host-limit overall 8000
                        min-auth-interval 1
                        no shutdown
                    exit
                exit
            exit
        exit
```

Sample configuration on a standby BNG

```
config>
    pw-port 2 create
    exit
config>redundancy#
     multi-chassis
         peer 10.20.1.2 create
             source-address 10.20.1.3
                sync
                    srrp
                    sub-mgmt ipoe pppoe
                    port pw-2 sync-tag "tag2" create
                    exit
                exit
                no shutdown
            exit
        exit
     exit
config>service>ies#
       redundant-interface "redundant-interface" create
          address 10.10.30.3/24 remote-ip 10.10.30.2
          spoke-sdp 32:1000 create
             no shutdown
          exit
       exit
config>service#
    sdp 1 mpls create
       far-end 10.20.1.2
       ldp
       keep-alive
          shutdown
       exit
       binding
          port 1/1/1
          pw-port 2 vc-id 2 create
             vc-type vlan        #### default encaps-type dot1Q
             no shutdown
          exit
       exit
       no shutdown
    exit
config>service#
```

```
       subscriber-interface "subif" create
           address 10.11.1.3/16 gw-ip-address 10.11.1.1 populate-host-routes
                group-interface "grpif" create
                    authentication-policy "base_authpolicy"
                    redundant-interface "redundant-interface"
                    sap pw-2:1000 create
                        description "sap-grp-3"
                    exit
                    srrp 1 create
                        keep-alive-interval 1
                        message-path pw-2:1000
                        no shutdown
                    exit
                    arp-host
                        host-limit 8000
                        min-auth-interval 1
                        no shutdown
                    exit
                exit
                group-interface "dummy" create
                    oper-up-while-empty
                exit
            exit
        exit
```

9.14.1.1.2.2 Sample configuration on A-PE

```
config>service>epipe#
     description "Default epipe description for service id 103"
     service-mtu 1492
     service-name "XYZ Epipe 103"
     endpoint "x" create
         standby-signaling-master
     exit
     sap 1/1/3 create
          description "Default sap description for service id 103"
     exit
     spoke-sdp 1:2 endpoint "x" create
          description "Description for Sdp Bind 1 for Svc ID 103"
          precedence primary
          no shutdown
     exit
     spoke-sdp 2:2 endpoint "x" create
          description "Description for Sdp Bind 2 for Svc ID 103"
          no shutdown
     exit
     no shutdown
```

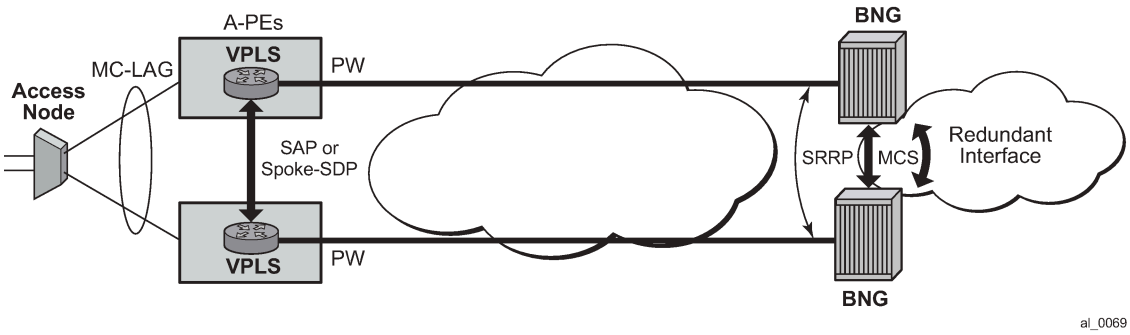
9.14.1.1.2.2.1 VPLS-based aggregation service

With VPLS based aggregation service from A-PE, normal SRRP message exchange can take place between the active and standby BNGs. An active or standby decision and switch-over is based on the SRRP state. An SRRP instance is configured per group-interface corresponding to PW port. Fate-sharing groups (FSG) can be configured for a set of SRRP instances (for example, SRRP instances corresponding to PW ports sharing the same subnet). A standard **oper-group** *grp-id* should be configured with messaging SAPs for all PW ports that are in the same FSG, and **monitor-oper-group** *grp-id* should be configured

under each SRRP instance in same FSG. Existing SRRP support defined in Triple-play services guide for ESM over regular group-interfaces and subscriber SAPs is applicable identically to ESM over PW ports and PW SAPs.

With ESM over PW, redundancy in the aggregation network based on MC-LAG between A-PE and dual BNGs is not supported.

Figure 130: BNG redundancy with VPLS based aggregation service



9.14.1.1.2.2.2 Sample BNG redundancy configuration with VPLS service on A-PE

```
config>
    pw-port 1 create
    exit

config>redundancy#
     multi-chassis
         peer 10.20.1.2 create
             source-address 10.20.1.3
                sync
                    srrp
                    sub-mgmt ipoe pppoe
                    port pw-1 sync-tag "tag1" create
                    exit
                exit
                no shutdown
            exit
        exit
     exit

 config>service>ies
       redundant-interface "red-1-1" create
           address 10.1.1.2/24 remote-ip 10.1.1.1
               spoke-sdp 1:1 create
                  no shutdown
               exit
           exit

       subscriber-interface "sub-1-1" create
            address 10.1.2.2/16 gw-ip-address 10.1.255.254 track-srrp 1
            address 10.2.2.2/16 gw-ip-address 10.2.255.254 track-srrp 2
            dhcp
                gi-address 10.1.2.2
            exit
            group-interface "grp-1-1-1" create
                 srrp-enabled-routing
```

```
                arp-populate
                dhcp
                    server 10.20.1.2
                    trusted
                    lease-populate 32767
                    client-applications dhcp ppp
                    gi-address 10.1.2.2
                    no shutdown
                exit
                   authentication-policy "iesAuthPol"
                   redundant-interface "red-1-1"

                  sap pw-1:1.1 create
                      sub-sla-mgmt
                          def-sub-profile “sub_prof_1”
                          def-sla-profile “sla_prof_1”
                          no shutdown
                  exit
                  sap pw-1:4000.1 create
                       oper-group "1"
                  exit

                  srrp 1 create
                       gw-mac 00:00:5e:00:01:01
                       keep-alive-interval 50
                       message-path pw-1:4000.1
                       monitor-oper-group "1" priority-step 10
                       no shutdown
                   exit
               exit
```

9.14.1.1.2.2.3 Sample BNG redundancy configuration with VPLS service on A-PE

```
config>service
     customer 1 create
         description "Default customer"
     exit
     sdp 1000 mpls create
         far-end 10.20.1.2
         lsp "lsp_1"
         path-mtu 1600
         keep-alive
         no shutdown
      exit
      sdp 1002 mpls create
         far-end 10.20.1.3
         lsp "lsp_3"
         path-mtu 1600
         keep-alive
         no shutdown
      exit
      vpls 1 customer 1 create
          service-mtu 1600
          stp
          sap 1/1/2 create   // to Access-Node
          exit
        sap 1/1/3 create; //to A-PE2
        exit
          spoke-sdp 1000:1 create  // to BNG1
```

```
            no shutdown
         exit
         no shutdown

       exit
    exit
```

9.14.1.1.2.2.4 A-PE configuration with VPLS aggregation service (A-PE2)

```
config>service
     customer 1 create
         description "Default customer"
     exit
     sdp 1002 mpls create
         far-end 10.20.1.3
         lsp "lsp_2"
         path-mtu 1600
         keep-alive
         no shutdown
      exit

      vpls 1 customer 1 create
          service-mtu 1600
          stp
          sap 1/1/2 create   // to Access-Node
          exit
        sap 1/1/3 create; //to A-PE1
        exit
          spoke-sdp 1002:1 create  // to BNG2
              no shutdown
          exit
          no shutdown
       exit
    exit
```

9.14.1.1.2.3 Show commands related to active/standby pseudowire on dual BNGs

The following example shows SRRP status, subscriber host, and routing information about the active BNG (SRRP master state):

```
A:Dut-B>config>redundancy# show srrp 1
===============================================================================
SRRP Instance 1
===============================================================================
Description         : (Not Specified)
Admin State         : Up                Oper State       : master
Preempt             : yes               One GARP per SAP : no
Monitor Oper Group  : None
System IP           : 10.20.1.2
Service ID          : VPRN 3
Group If            : grpif             MAC Address      : 1c:85:ff:00:00:00
Grp If Description  : N/A
Grp If Admin State  : Up                Grp If Oper State: Up
Subscriber If       : subif
Sub If Admin State  : Up                Sub If Oper State: Up
```

```
Address             : 10.11.1.2/16       Gateway IP       : 10.11.1.1
Redundant If        : redundant-interfa*
Red If Admin State  : Up                 Red If Oper State: Up
Address             : 10.10.30.2/24
Red Spoke-sdp       : 23:1000
Msg Path SAP        : pw-2:1000
Admin Gateway MAC   :                    Oper Gateway MAC : 00:00:5e:00:01:01
Config Priority     : 100                In-use Priority  : 100
Master Priority     : 100
Keep-alive Interval : 1 deci-seconds     Master Since     : 05/29/2012 07:22:26
Fib Population Mode : all
VRRP Policy 1       : None               VRRP Policy 2    : None
===============================================================================
* indicates that the corresponding row element may have been truncated.

A:Dut-B>config>redundancy# show service id 3 arp-host
===============================================================================
ARP host table, service 3
===============================================================================
IP Address      Mac Address       Sap Id              Remaining           MC
                                                      Time                Stdby
-------------------------------------------------------------------------------
10.11.1.11      00:80:00:00:00:01 [pw-2:11]           03h35m47s
10.11.1.12      00:80:00:00:00:02 [pw-2:12]           03h35m47s
-------------------------------------------------------------------------------
Number of ARP hosts : 2
===============================================================================

A:Dut-B>config>redundancy# show router 3 route-table 10.11.1.11
===============================================================================
Route Table (Service: 3)
===============================================================================
Dest Prefix[Flags]                            Type    Proto    Age         Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.11.1.11/32                                 Remote  Sub Mgmt 00h24m26s   0
       [grpif]                                                      0
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: L = LFA nexthop available    B = BGP backup route available
       n = Number of times nexthop is repeated
===============================================================================
A:Dut-B>config>service>vprn#
```

The following shows SRRP status, subscriber host, and routing info in standby BNG (SRRP init state):

```
A:Dut-C>config>redundancy# show srrp 1
===============================================================================
SRRP Instance 1
===============================================================================
Description         : (Not Specified)
Admin State         : Up                 Oper State       : initialize
Preempt             : yes                One GARP per SAP : no
Monitor Oper Group  : None
System IP           : 10.20.1.3
Service ID          : VPRN 3
Group If            : grpif              MAC Address      : 1c:87:ff:00:00:00
Grp If Description  : N/A
Grp If Admin State  : Up                 Grp If Oper State: Down
Subscriber If       : subif
Sub If Admin State  : Up                 Sub If Oper State: Up
Address             : 10.11.1.3/16       Gateway IP       : 10.11.1.1
Redundant If        : redundant-interfa*
```

```
Red If Admin State  : Up                 Red If Oper State: Up
Address             : 10.10.30.3/24
Red Spoke-sdp       : 32:1000
Msg Path SAP        : pw-2:1000
Admin Gateway MAC   :                    Oper Gateway MAC : 00:00:5e:00:01:01
Config Priority     : 1                  In-use Priority  : 1
Master Priority     : 1
Keep-alive Interval : 1 deci-seconds     Master Since     : 05/29/2012 07:22:26
Master Down Interval: 0.000 sec (Expires in 0.000 sec)
Fib Population Mode : all
VRRP Policy 1       : None               VRRP Policy 2    : None
===============================================================================
* indicates that the corresponding row element may have been truncated.

A:Dut-C>config>redundancy# show service id 3 arp-host
===============================================================================
ARP host table, service 3
===============================================================================
IP Address      Mac Address       Sap Id              Remaining           MC
                                                      Time                Stdby
-------------------------------------------------------------------------------
10.11.1.11      00:80:00:00:00:01 [pw-2:11]           03h38m01s           Yes
10.11.1.12      00:80:00:00:00:02 [pw-2:12]           03h38m02s           Yes
-------------------------------------------------------------------------------
Number of ARP hosts : 2
===============================================================================
A:Dut-C>config>redundancy# show router 3 route-table 10.11.1.11
===============================================================================
Route Table (Service: 3)
===============================================================================
Dest Prefix[Flags]                            Type    Proto    Age         Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
10.11.1.11/32                                 Remote  Sub Mgmt 00h22m03s   0
       [redundant-interface]                                        0
-------------------------------------------------------------------------------
No. of Routes: 1
Flags: L = LFA nexthop available    B = BGP backup route available
       n = Number of times nexthop is repeated
===============================================================================
```

9.14.1.2 ESM on PXC-based PW ports

This section provides an example to configure PW port based capture SAP that is used in ESM. For more information about PXC Based PW ports, see the *7450 ESS, 7750 SR, 7950 XRS, and VSR Layer 2 Services and EVPN Guide*.

PXC Configuration

The following is a PXC configuration example:

```
configure
   port-xc
      pxc 1 create
         port 1/1/1
         no shutdown
      pxc 2 create
         port 2/1/1
         no shutdown
```

With this configuration, ports 1/1/1 and 2/1/1 are auto-provisioned in hybrid mode operating as individual loopback ports. The SR OS system automatically creates a pair of sub-ports per PXC. Those sub-ports are, by default, are in the shutdown state, and must be explicitly enabled (**no shutdown**) by the operator.

```
configure
   port pxc-1.a
      shutdown
   port pxc-1.b
      shutdown

   port pxc-2.a
      shutdown
   port pxc-2.a
      shutdown
```

For redundancy purposes or increased bandwidth, the PXC sub-ports are aggregated in a LAG:

```
configure
    lag 1 create
        port pxc-1.a
        port pxc-2.a
    lag 2 create
        port pxc-1.b
        port pxc-2.b
```

FPE Configuration

A PW port is associated with PXC by an FPE configuration. FPE configuration facilitates creation of an internal tunnel over PXC. This tunnel is used to map the external PW to the PW port. For more information about FPE, see *7450 ESS, 7750 SR, 7950 XRS, and VSR Interface Configuration Guide*.

The following is an FPE configuration example:

```
configure
    fwd-path-ext
        fpe 1 create
            path xc-a lag-1 xc-b lag-2
            pw-port
```

The association between xc-a/b (cross-connects) and LAG IDs is performed arbitrarily by the operator. For example, it can associate xc-a with lag-id 2 (which includes PXC sub-ports on the .b side) and xc-b with lag-id 1 (which includes PXC sub-ports on the .a side). However, the FPE always assigns the .a side of the pxc sub-ports to the transit side of the cross-connect, while the .b side of the pxc subport is assigned to the termination side of the cross-connect. See the PXC Based PW port sections in the *7450 ESS, 7750 SR, 7950 XRS, and VSR Layer 2 Services and EVPN Guide*, for further information about transit/termination side of the cross-connect.

PW port creation

The PW port must be explicitly created in the SR OS, before mapping between PW and PW port can be performed.

```
pw-port 100 create
    encap-type qinq
```

SDP creation for the external PW

The following displays an SDP configuration for the external PW.

```
configure
    service
        sdp 1 create
            signaling tldp
            far-end 10.1.1.1
```

The PW can be static or dynamically signaled, with MPLS or GRE transport.

Mapping between the external PW and the PW port

The stitching of the external PW and PW port is configured through an Epipe in vc-switching mode.

```
configure
    service
        epipe 10 customer 1 vc-switching create
            spoke-sdp 1:100 create
            pw-port 100 fpe 1
```

Capture PW-SAP creation

PW-SAPs can be configured as capture SAPs. In this example, a capture PW-SAP with S-tag 3 is created on the pw-port 100.

```
configure
    service vpls 2 create
        sap pw-port-100:3.* capture-sap create
```

From here, ESM functionality is applied to the PW-SAP in the same manner as on any other regular SAP.

Cross-connecting SAPs to PW ports

In addition to PW termination, a PW port can become a terminating point for a regular SAP. For example:

```
configure
    service
        epipe 10 customer 1 vc-switching create
            sap 1/1/1:10.* create
            pw-port 100 fpe 1
```

In this example, the outer VLAN tag 10 in the payload is removed on ingress and the payload is delivered to the PW port where it can be mapped to a capture PW-SAP. This scenario allows traffic distribution from a single I/O port to different EMS termination points (anchor line cards) based on outer VLANs.

9.14.1.3 ESM multichassis redundancy with PXC-based PW ports and EVPN VPWS

Redundant BNGs with EVPN VPWS in the access area of the network rely on the EVPN Single-Active (SA) multihoming concept with PW ports in Ethernet Segments (ES). A PW port on one side in the ES is elected as the Designated Forwarder (DF) and the other side as the non-Designated Forwarder (NDF). The ES with the PW port as DF is operationally up, and conversely, the ES with the PW port as NDF is operationally down. The DF side is the active side while the NDF side is the standby side. SRRP, as part of subscriber management redundancy scheme, indirectly tracks ES states to determine which BNG side is active and which is standby. With multiple EVPN VPWS instances, the load is distributed between the redundant BNGs where one BNG can be active for one set of EVPN VPWS while the other BNG can be active for another set of EVPN VPWS instances. The operator can influence the selection of the active side (DF side) for each EVPN VPWS by configuring a higher preference number on the preferred DF side.

```
config>service>system>bgp-evpn>eth-seg>service-carving$ manual preference <number>
```

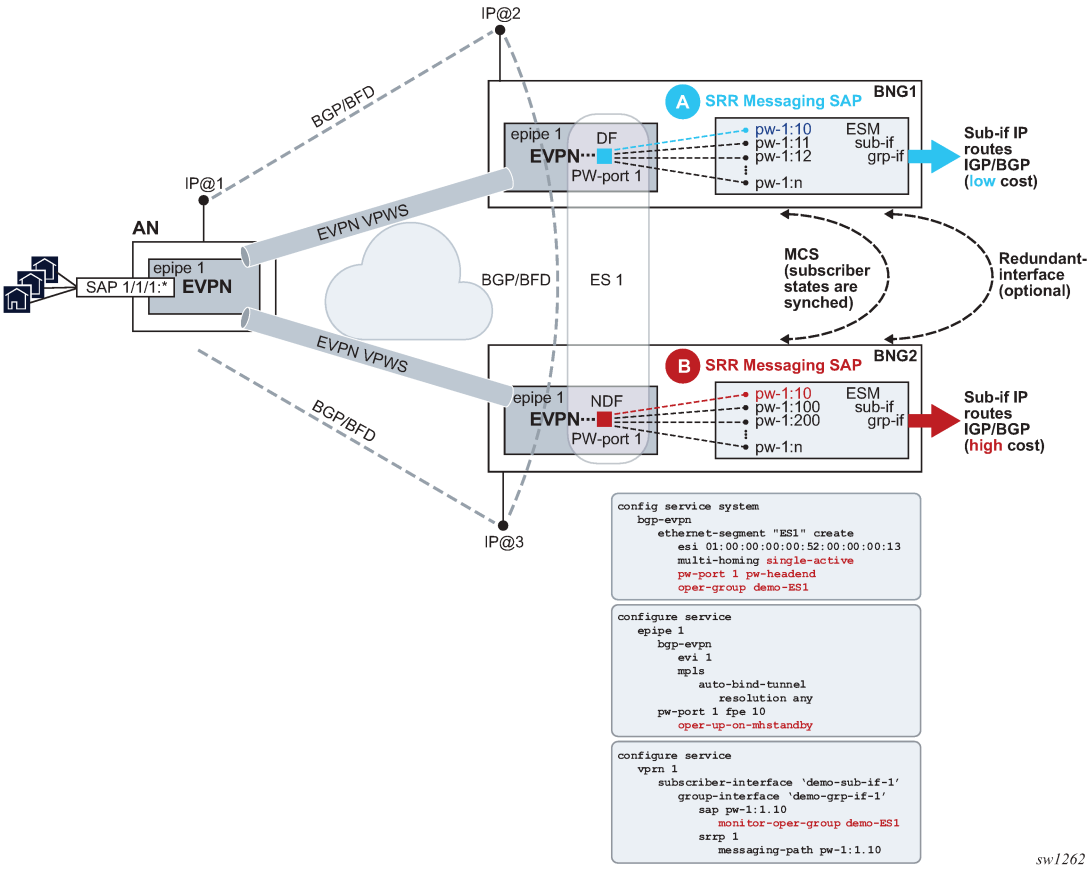
In a typical ESM environment, a PW port contains thousands of PW SAPs, with each PW SAP representing a subscriber. To minimize the outage time during failures, the operator (through configuration) can optionally keep those PW SAPs operationally up even if the underlying PW port is in the NDF state. This reduces the failover time otherwise required to bring all the PW SAPs operationally up.

The SRRP state must transition into a standby state on the NDF side even if the PW-port is operationally up. To achieve this, the SRRP messaging PW SAP goes through an oper-group that tracks the state of the ES, whose operational state is up on the DF side and down on the NDF side

The basic concept of this approach, where the messaging SRRP PW-SAP is tracking the state of the ES, is shown in Figure 131: BNG multichassis redundancy with EVPN VPWS . There are two key concepts introduced:

- The **oper-up-on-mhstandby** CLI flag ensures that the PW port is operationally up even while it is the NDF.
- The SRRP messaging SAP is tracking the state of the corresponding ES through an oper-group (“demo-ES2”). This ensures that the SRRP follows the activity state of the EVPN VPWS, while the PW port remains operationally up on both BNGs (active and standby).

Figure 131: BNG multichassis redundancy with EVPN VPWS



Within an SR node, a collection of separate entities work together to detect a failure in the network and divert the traffic around it. Those entities are:

- ESM where subscribers are synchronized between the chassis
- EVPN VPWS in the access area of the network
- Routing that is advertising subscriber routes into the network
- Oper-groups used to interconnect operational states between the entities
- Various network failure detection mechanisms such as BFD to quickly detect failure path between BGP peers

The following is a detailed description of the setup with a single EVPN VPWS and two BNGs (Figure 131: BNG multichassis redundancy with EVPN VPWS).

- EVPN VPWS is configured as SA) multihoming. SA is crucial in ESM as it drives the SRRP state which must always be in an active or standby state between the redundant pair of BNGs.
- BNGs are connected to the AN through an EVPN VPWS.
- One BNG in the EVPN is selected as the DF (BNG1), the other BNG (BNG2) is the NDF.
- BNG2 bring its ES down.
- Only BNG1 advertises its AD route toward the access node.

Consequently, the AN does not send any traffic to BNG2 (NDF). Instead, the AN sends all traffic only to BNG1 (DF).

- The ES is part of an oper-group (OG) which is monitored from the ESM side.
- The stitching Epipe on BNG2 does not change its status. Neither does the PW port in it. The PW port stays up despite the MHStandby flag being raised. Normally, the MHStandby flag would cause the PW port to go down, but because of the **oper-up-on-mhstandby** configuration option, this behavior is overridden.
- ESM subscribers are synchronized between the chassis through MCS and are using SRRP on the access side. With EVPN in the access, SRRP is not relying on its own keepalives to check the health of the network path, but instead, it follows the state of the PW port or the ES. If the PW port is operationally up, the messaging PW SAP is up, and therefore the SRRP is active. Conversely, if the PW port is operationally down, the messaging PW SAP is down and consequently SRRP is in the backup state. This is the expected behavior when the EVPN MPLS destination (network bind) goes down.

However, in the SA multihoming scenario, when the EVPN MPLS destination is not down, the PW port remains up even if the PW port is an NDF. Instead of relying on the PW port state, the SRRP messaging PW SAP monitors the state of the ES through an oper-group. When the oper-group changes its state to down, so does the SRRP messaging PW SAP, which then forces the SRRP into an INIT state (which is equivalent to a standby state).

- On the network side, the state of the SRRP controls the advertisement of the subscriber IP routes into the network. Subscribers routes are advertised with a lower cost from the active SRRP node than they are from the standby SRRP node.
- The solution described above protects against failures in the access part of the network or BNG node failure. Optionally, network side ports can be placed in an oper-group that can be monitored from the EVPN side. This can be used to protect against network port failures.

9.15 Logical Link Identifier (LLID)

This feature enables service providers to track subscribers based on a virtual-port known as logical line ID (LLID). The LLID (an alphanumeric string) is a logical identification of a subscriber line. Mapping of physical line of a subscriber to LLID is performed by pre-authentication with a separate AAA server than the AAA server used for authenticating the subscriber session during normal access authentication.

LLID serves the purpose of abstracting the physical line of the user from the ISP. If the user moves to a new physical line, the RADIUS server database maintaining the physical line of the subscriber to LLID is updated. Because a subscriber's LLID remains same regardless of subscriber's physical location, using LLID gives service provider a stable and secure identifier for tracking subscriber.

The local user database assigned to the PPPoE node under the group interface can have both a pre-authentication policy and an authentication policy. The purpose of the pre-authentication policy is to retrieve the LLID from the AAA server. The pre-authentication only extracts the calling-station-id attribute (0x31) which is used as the LLID, anything else returned during pre-authentication are simply ignored. If the pre-authentication is missing the LLID, the session moves on to the authentication policy. In the authentication policy that follows, it is possible to use the LLID as the calling-station ID.

It is possible to convey LLID from the LAC to the LNS. The LLID is retrieved through PPPoE pre-authentication where the returned RADIUS attribute calling station ID is used as the LLID. This LLID is selectable attribute in L2TP as a calling-number (AVP 22) to be passed from LAC to LNS. At the LNS, the subscriber calling station number is retrieved from AVP 22 and can be included as an attribute during authentication.

9.16 PADI authentication policy for managed SAP (MSAP)

The PADI Authentication Policy feature enables PADI authentication that retrieves MSAP parameters before pre-authentication and PPPoE authentication.

With this feature, authentication occurs in the following manner.

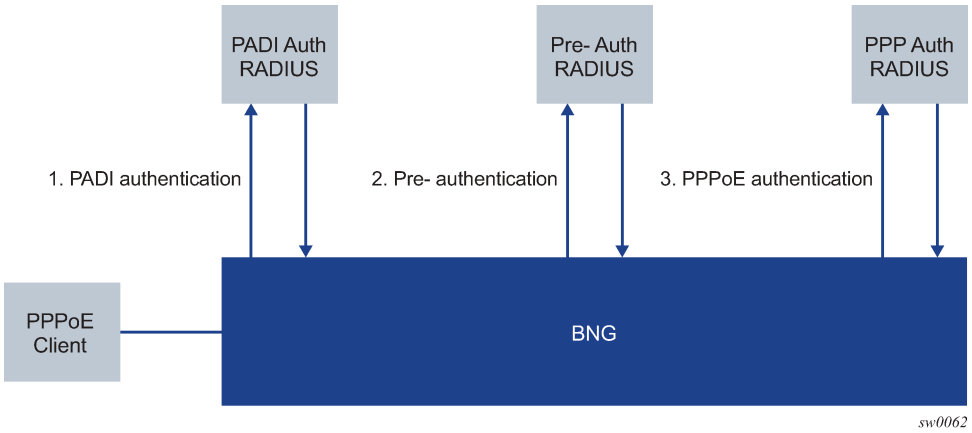
1. PADI authentication and MSAP authorization:
   a. A capture SAP receives a PPPoE PADI packet.
   b. A check verifies that the LUDB is configured as **pppoe-user-db** on the capture SAP.

      The LUDB must be configured on the capture SAP. Without the LUDB specified, the existing functionality is performed for MSAP creation.
   c. The LUDB host entry is matched to the entry that has the PADI authentication policy.
   d. RADIUS authentication and MSAP authorization occurs.
   e. An MSAP is created if the authentication policy is not configured for PAP/CHAP.
2. (Optional) LLID pre-authentication:
   a. Look up the LLID pre-authentication policy under the LUDB host entry.
   b. Perform RADIUS pre-authentication to obtain an LLID.
3. LCP authentication and L2TP tunnel authorization:
   a. Trigger RADIUS authentication defined by the authentication policy configured in the LUDB host entry in the previous step.

b. Create the MSAP, finish the LCP authentication with the PPP client, and establish an L2TP tunnel/session.

Figure 132: Triple authentication with MSAP authentication policy shows triple authentication with MSAP authentication policy.

Figure 132: Triple authentication with MSAP authentication policy



The CLI command **config>subscr-mgmt>loc-user-db>ppp>host>padi-auth-policy** configures a PADI authentication policy used for PADI authentication with MSAP authorization as shown in the CLI example below:

```
configure
    subscriber-mgmt
        local-user-db “ludb-1” create
            ppp
                host “default” create
                    padi-auth-policy “padi_auth”
```

9.17 Open authentication model for DHCP and PPPoE hosts

9.17.1 Terminology

- **LUDB**

  Local User Database configured within the 7450 ESS and 7750 SR.

- **IP Address Assignment with DHCP Relay**

  IP address assignment request (DHCP or IPCP) from the host is relayed to an internal or external DHCP server. A gi-address must be present in this relayed request while the pool name is optional. The internal DHCP server may select the IP address from is local pool based on the gi-address or based on the pool-name present in the request. The IP address selection method is configuration dependent. Third party DHCP servers may consider additional fields in IP address selection process (mac address, circuit-id, and so on).

- **IP Address Assignment with DHCP Proxy**

  A preconfigured IP address in LUDB or RADIUS server is handed out to the host using a DHCP proxy function. This proxy function responds natively using DHCP protocol to the IPoE host. Although PPPoE hosts are not utilizing DHCP protocol, the DHCP proxy functionality within the server is still needed for successful IP address delegation to PPPoE hosts.

9.17.2 Prioritization of authentication sources

ESM parameters (ESM strings and other IP parameters) obtained during authentication and re-authentication phases are combined from various sources with a specific preference order as follows:

1. ESM Python (set.esm function)
2. Diameter/Gx
3. LUDB
4. RADIUS
5. Diameter/Nasreq
6. LocalAddressAssignment
7. GTP
8. DHCP
   - DHCP parameters that came from standard DHCP options returned by the DHCP server directly
   - Information extracted from options (strings-from-options). This is applicable for IPoE and PPPoE (DHCP client) that use a local DHCP server with LUDB.
   - DHCP ACK Python
9. defaults, if any

For example, if the same ESM parameter is provided through both authentication sources, LUDB and RADIUS, the ESM parameter from LUDB always overrides the ESM parameter obtained from RADIUS.

SR OS allows the priority of LUDB and RADIUS sources to be reversed. This prioritization of authentication sources, where RADIUS is granted priority over LUDB, ensures that parameters from LUDB are used as a backup, only in cases where the same ESM parameters are not provided by RADIUS.

The settings that allow swapping of the LUDB and RADIUS priorities as authentication sources are configured on the system level as follows.

Classic CLI:

```
subscriber-mgmt
   authentication-origin
       [no] priority <id> source <string>
   exit
exit
```

The only accepted configuration option is *id* 3 and RADIUS as the source *string*. This configuration moves RADIUS to position 3 and shifts everything from the previous position 3 downward.

The defaults are restored by using the **no** form of the **priority** command.

The active order of priorities can be displayed in the output of the **show**>**subscr-mgmt**>**authentication-origin** command:

```
*A:cses-V26>config>subscr-mgmt>auth-orig# show subscriber-mgmt authentication-origin
===============================================================================
Authentication Origins
===============================================================================
Priority                        Source
-------------------------------------------------------------------------------
1                               python
2                               diameterGx
3                               ludb
4                               radius
5                               diameterNasreq
6                               localAddressAssignment
7                               gtp
8                               dhcp
-------------------------------------------------------------------------------
Number of Authentication Origins : 8
===============================================================================
*A:cses-V26>config>subscr-mgmt>auth-orig# priority 3 source radius
*A:cses-V26>config>subscr-mgmt>auth-orig# show subscriber-mgmt authentication-origin
===============================================================================
Authentication Origins
===============================================================================
Priority                        Source
-------------------------------------------------------------------------------
1                               python
2                               diameterGx
3                               radius
4                               ludb
5                               diameterNasreq
6                               localAddressAssignment
7                               gtp
8                               dhcp
-------------------------------------------------------------------------------
Number of Authentication Origins : 8
===============================================================================
*A:cses-V26>config>subscr-mgmt>auth-orig#
```

The following describes the configuration logic where both LUDB and RADIUS are accessed during authentication phase:

- LUDB is referenced under the capture SAP (if the capture SAP is deployed) and the group interface (DHCP, DHCP6, or PPPoE).
- The authentication policy is referenced in LUDB (and not under the capture SAP and group interface).

With this approach, LUDB is accessed first and subscribers can be authenticated based on generic criteria, such as a range of VLANs or a default user. The ESM parameters obtained in this step are stored.

After LUDB authentication, RADIUS is accessed when authentication on subscriber-specific authentication fields is performed (for example, based on a username, circuit-id, MAC address, and so on). During this RADIUS authentication phase, another set of ESM parameters more tailored for the specific user is obtained, effectively overriding the overlapping parameters from LUDB.

9.17.2.1 Authentication source — session versus host ESM Model

For IPoE host-based deployments, such as when the **ipoe-session** is disabled (meaning that clients are treated as hosts instead of sessions), the re-authentication option in the authentication policy must be enabled when RADIUS is prioritized over LUDB as the authentication source. Enabling re-authentication is only required in IPoE host-based deployments and not required for IPoE or PPPoE sessions.

9.17.3 No authentication

IPoE and PPPoE v4/v6 hosts on static SAPs can be instantiated without the need to access LUDB or RADIUS server. In this case, the default subscriber host parameters (sla-profile, sub-profile, subscriber-id) must be provisioned statically under the SAP. The IP address assignment is provided by internal or external DHCP server. The IP address selection on the router based DHCP server is based on the gi-address while third party DHCP servers may provide additional means to select the IP address (*mac-address*, *circuit-id*, and so on).

A DHCP pool name cannot be provided by an SR-series router DHCP relay agent, because the LUDB and RADIUS are not used.

This model does not support IP address delegation by DHCP Proxy function because there is no LUDB or RADIUS server available that can supply pre-configured IP address.

Host instantiation without LUDB or RADIUS access on dynamic VLANs (capture SAP and consequently mSAP) is not supported.

9.17.4 LUDB only access

Subscriber-host authentication, identification and IP address assignment can be performed by LUDB without the need to access the RADIUS server.

The LUDB is normally configured under the **group-interface**>**ppp**/**dhcp** hierarchy and can provide subscriber-identification parameters as well as IP addressing parameters:

Pool names for DHCP relay function (IPv4, IPv6 IA-NA, IPv6 IA-PD)

Fixed IP addresses – IPv4, IPv6 IA-NA, IPv6 IA-PD and IPv6 SLAAC prefix.

In case of capture SAP, the LUDB name configured under the capture SAP must match the LUDB name under the **group-interface**>**dhcp**/**ppp** hierarchy. If the LUDB names do not match, the subscriber-host instantiation fails.

9.17.5 LUDB access by DHCPv4 server

If the IPv4 addressing assignment is facilitated by the DHCPv4 relay and an internal DHCPv4 server, the DHCPv4 server itself can query the LUDB for IPv4 address information. LUDB can provide a v4 pool name and IPv4 DHCP options to the DHCPv4 server or it can instruct it to use the gi-address as the IPv4 address selection mechanism.

ESM strings can also be provided by LUDB queried by the DHCPv4 server.

If LUDB access by DHCPv4 server is provided in addition to other authentication means (another LUDB under the group-interface, or RADIUS server), the ESM strings from the LUDB under the grp-interface or from the RADIUS server has priority over the ESM strings configured under the LUDB accessed by the DHCPv4 server. On the other hand, the IPv4 addressing information has the highest priority from the LUDB accessed directly by the DHCPv4 server.

Accessing LUDB directly by DHCPv4 server should be used in rare and exceptional cases.

LUDB access under the group-interface, possibly complemented by the RADIUS server provides necessary means for subscriber-host instantiation in majority of use cases.

9.17.6 RADIUS only access

Like LUDB-only access, RADIUS server can provide all the necessary information for subscriber-host instantiation, including the IP addressing parameters (pool names or IP addresses/prefixes). Authentication-policy which defines the RADIUS access must be applied to the group-interface.

In case of capture SAP, the authentication policy must be applied under the capture SAP. This authentication policy name must match the authentication policy name that is configured under the group-interface. Otherwise, the host instantiation fails.

9.17.7 Consecutive access to LUDB and RADIUS

LUDB and RADIUS access can be combined during subscriber-host instantiation phase.

Configuration-wise, LUDB must be referenced under the **group-interface**>**dhcp**/**ppp**/**pppoe** hierarchy (and possibly under the capture SAP), while the authentication-policy is specified within the LUDB. In this fashion, LUDB access is followed by RADIUS access. The subscriber-host parameters retrieved from both sources are combined with LUDB parameters being prioritized over RADIUS parameters in case that both sources return the same parameters.

If LUDB and authentication policy are configured simultaneously under the group-interface (and possibly under the capture SAP), the RADIUS authentication policy evaluates and LUDB is ignored.

9.17.8 RADIUS fallback

If RADIUS server is not accessible (non-responsive), the host instantiation phase can be:

- Terminated if there is no fallback action within authentication policy specified.
- Continued within LUDB if the fallback action within the authentication-policy references LUDB.
- Continued without any response from RADIUS. Subscriber-host is instantiated if defaults parameters are statically configured or the instantiation fails if the defaults are not available.

The fallback action takes effect after the preconfigured RADIUS timeout period expires.

RADIUS fallback is not supported for DHCPv6 hosts for non-IPoE sessions but is supported for IPoE sessions.

9.18 Flexible subscriber-interface addressing (unnumbered subscriber-interfaces)

9.18.1 Terminology

Subscriber host

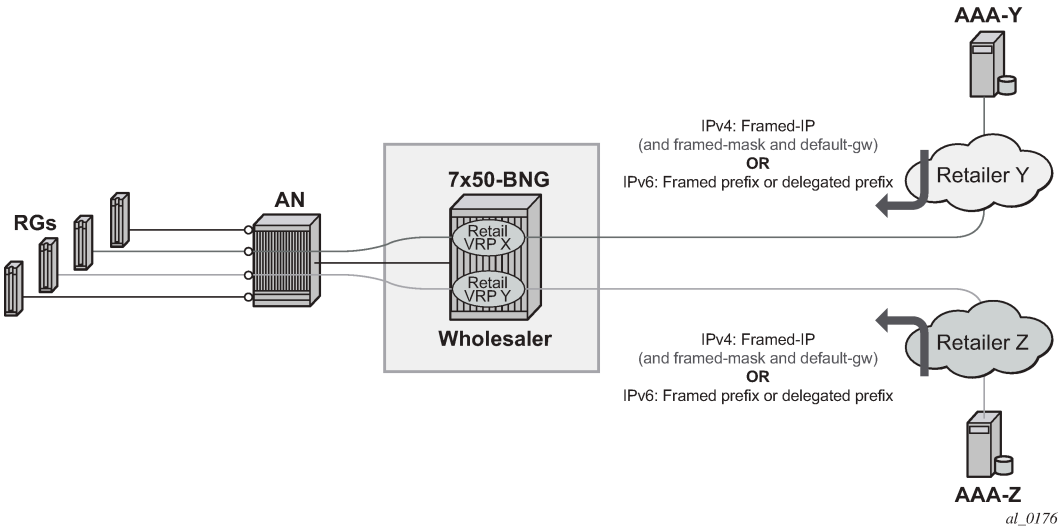
A representation of an external host requesting a service. Each such host is fully instantiated within the 7450 ESS and 7750 SR for the purpose of providing traffic control and billing services (for example, QoS, filtering, antispoofing, accounting). The external hosts may represent variety of devices such as regular PCs, STBs, residential gateways, CPEs, VoIP devices. In most cases, the external host runs a DHCPv4/v6 or PPPoEv4/v6 client. DHCP and PPPoE initiation messages from such clients triggers host instantiation within the router. For this the subscriber host term can be interchangeably used with a term DHCP client or PPPoE client.

9.18.2 Flexible subscriber-interface addressing for IPOE/PPPoE v4/v6 subscribers

In various wholesale or retail environments, the wholesale provider that own the 7450 ESS and 7750 SR BNG does not know the IP addresses that the retailers assigns to their clients in advance. For this reason, wholesaler's BNG must accept any IP address from retailers and consequently pass it to the client during subscriber-host initiation phase.

Figure 133: Use case for flexible IP addressing model shows a use case for flexible IP addressing mode.

Figure 133: Use case for flexible IP addressing model



Flexible addressing of the subscriber-interface assumes two deployment scenarios:

1. **Subscriber-interface is unnumbered**

   For example, there is no explicit assigned IP address. Instead the subscriber-interface borrows the IP address from an existing interface that is operationally UP and is located in the same routing instance (router or vprn).

   An interface must have an IP address assigned to be operationally UP. Therefore, an unnumbered subscriber-interface must reference another existing interface that is operationally UP in the same routing instance. The subscriber-interface borrows the IP address from the referenced interface.

   In this case any IP address can be assigned to the subscriber host under the unnumbered subscriber-interface. The subscriber IPv4 address is installed in the FDB as /32 route while IPv6 address is installed as an entry of the length anywhere between 64 and 128 bits.

2. **Subscriber-interface is numbered**

   The IP address/prefix is explicitly configured and solely owned by the subscriber-interface.

   In this case, all subscriber IP addresses/prefixes that fall under the subnet/prefix dictated by the configured subscriber-interface IP address/prefix is directly aggregated under the subscriber-interface subnet. They occupy a single entry in the FDB. The rest of the subscriber hosts with IP addresses/prefixes that fall outside of the configured range are installed in the FDB as individual entries (/32 for IPv4 and an entry of the length anywhere between 64 and 128 bits for IPv6 hosts).

9.18.3 Default gateway in IPv4 flexible addressing

In scenarios where subscriber host IPv4 address lies within the configured subscriber-interface subnet, the default-gw IPv4 address for the host is one of the subscriber-interface IPv4 addresses. In this case, the service provider is aware of the IPv4 addressing scheme in the BNG and supplies the DHCP client with the appropriate default-gw IPv4 address by LUDB, RADIUS or DHCP server (in that order of priority).

In scenarios where the retail service provider wants to maintain independence from the IPv4 addressing scheme deployed in the BNG (that is controlled by wholesaler), the retailer can always supply its own IPv4 address, the subnet mask and the default-gw IPv4 address. But if the default-gw IPv4 address and subnet mask is not supplied by the retailer, then they are auto-generated by the BNG. After the default-gw IPv4 address is auto-generated, it is sent to the requesting DHCP client by DHCP offer in option 3 (RFC 2132, Router Option, section 3.5). There is no additional configuration needed for this action. The BNG automatically detects whether the default-gw IPv4 address is supplied by LUDB, RADIUS or DHCP server and acts correspondingly.

The default-gw IPv4 address is auto-generated based on the assigned IPv4 address/mask by setting the last bit of the assigned host IPv4 address to binary 01 or binary 10. For example if the subscriber host's assigned IPv4 address is 10.10.10.10 255.255.255.0, then the default-gw IPv4 address is set to 10.10.10.1. If the assigned IPv4 address is 10.10.10.1 255.255.255.0, then the auto-generated default gateway IPv4 is set to 10.10.10.2.

The default gateway IPv4 address always must be within the subscriber's subnet. If it is not, the behavior may be inconsistent. For example:

RADIUS (or DHCP) returns IP@, mask and def-gw:

- IP 10.10.10.1
- Def-gw 10.10.0.254
- Subnet mask 255.255.255.0

The subscriber is successfully instantiated in the BNG, but the client may not ARP for a default-gw outside of its configured subnet. Whether the client does or does not ARP for a default-gw outside of its configured subnet depends on the implementation in the RG and CPE.

RADIUS returns IP@ and subnet mask.

In this case the auto-generated default-gw IPv4 address is always within subscriber's subnet.

Flexible IPv4 addressing with auto-generated default-gw is supported only in Routed Central Office (RCO) model with routed residential gateways (RGs) or CPEs. In RCO model with bridged residential gateways or CPEs, the default-gw IPv4 addresses and the assigned IPv4 addresses may overlap. After the IPv4 address of the default-gw is auto-generated, it is possible that the second host behind the bridged residential gateway or CPE is assigned the same IPv4 address as the IPv4 address of the default gateway of the first host. Such hosts would not be able to communicate with outside world.

For example:

RADIUS or DHCP server assigns IPv4 address and subnet mask to the first host in a bridged environment:

IP1: 10.10.10.1

Auto-generated default-gw IPv4 address: 10.10.10.2

Because the RADIUS and DHCP server are not aware of the auto-generated default-gw, they may assign the following IPv4 address to the second host that comes on-line:

IP 2: 10.10.10.2 (same IPv4 address as the default-gw IPv4 address of the first host)

Auto-generated default-gw IPv4 address: 10.10.10.1

Now the first host forwards all traffic outside of the configured subnet to the second hosts which discards this traffic, effectively rendering this operation model non-deployable. And the other way around.

9.18.4 IPv4 subnet sharing

Subnet sharing between the hosts in flexible IPv4 addressing model is supported. In other words, in flexible IPv4 addressing model the operator can assign all IPv4 addresses (minus one, the default-gw IPv4 address) from a specific subnet. In this fashion, all subscribers (routed RGs or CPEs) within a single subnet can share the same default gateway.

For example, if the operator owns the IPv4 subnet 10.10.10.0/24, then one IPv4 address can be set aside for the default-gw (for example 10.10.10.254) and the remaining addresses can be assigned to the subscriber (routed RGs or CPEs). An example would be:

RG1: IP=10.10.10.1/24 def-gw 10.10.10.254

RG2: IP=10.10.10.2/24 def-gw 10.10.10.254

RG3: IP=10.10.10.3/24 def-gw 10.10.10.254

:

RG100: IP=10.10.10.100/24 def-gw 10.10.10.254

The subnet sharing is also supported in conjunction with auto-generated default-gw IPv4 address. The implication of this is that the IPv4 address of the default-gw can collide with the same IPv4 address already assigned to an existing subscriber. This is not an issue for routed RGs or CPEs because the BNG always answers ARPs for the IPv4 address of the default-gw with its own MAC address. However, local-proxy ARP functionality in the 7450 ESS and 7750 SR BNG must be enabled to support this.

This behavior can be further clarified with the following example.

Let’s assume that we have scenario with two routed RGs:

RG-1, IP=10.10.10.0/24, default-gw IP=10.10.10.1

RG-2, IP=10.10.10.1/24, default-gw IP=10.10.10.0

After RG-1 ARPs for its default gateway of 10.10.10.1, the BNG replies with its own MAC address.

Now that host RG-1 has resolved ARP for it default-gw (MAC address pointing to the router), it can send traffic to the outside world by the BNG. When such traffic arrives to the router, the destination IPv4 address of the received packet determines the forwarding decision within the router. If the destination IPv4 address matches the IPv4 address of any subscriber (RG) instantiated within the system, the traffic is forwarded to the that RG. This also includes the case where the destination IPv4 address is the default-gw IPv4 address (10.10.10.1), which represents just another RG within the router. The traffic is consequently passed from RG-1 by 7450 ESS and 7750 SR to RG-2.

9.18.5 IPv4 subnet mask auto-generation

The subnet mask corresponding to the IPv4 address assigned to the subscriber is auto-generated in case that the IPv4 addressing authority (LUDB, RADIUS or DHCP server) does not supply it. The subnet mask is derived from the IPv4 address of the subscriber and possibly the default-gw IPv4 address and it is the smallest subnet that contains both, the IPv4 address of the subscriber and the default-gw.

For example, if the RADIUS received IPv4 address is 10.10.10.138 and the received default –gw IPv4 address is 10.10.10.170, then the subnet mask is auto-generated and set to 255.255.255.192 (/26).

138 = 10001010

170 = 10101010

192 = 11000000

In case that neither the subnet mask nor the default-gw are returned, then both would be auto-generated:

1. Subnet mask would be set to /31
2. Default-gateway which must belong to the subscriber’s subnet would be set to 10.10.10.139.

In cases where the host IPv4 address and the default-gw are directly supplied by the addressing authority but the subnet mask is missing, the subnet mask auto-generation may cause the host part of the default-gw IPv4 address to become a broadcast IPv4 address. If this is an issue, then it can be avoided by directly providing the subnet mask by the addressing authority.

9.18.6 local-proxy-arp and arp-populate

local-proxy-arp and **arp-populate** are two commands that are relevant only to IPoEv4 hosts.

The **local-proxy-arp** command ensures that the router answers ARP Requests with its own MAC address for any active IPv4 address under the subnet on which the ARP request arrived. The active IPv4 address is considered the one that is assigned to an already instantiated hosts or the default-gw (even auto-generated).

In absence of **local-proxy-arp** command, the only ARP Request that the router’s answer is the one for the statically configured IPv4 addresses of the subscriber-interface. In flexible IPv4 addressing, the IPv4 address of the default-gw does not necessarily match any of the configured subscriber-interface IPv4 addresses. The ARP Request for such default-gw IPv4 address would go unanswered. Consequently,

the subscriber hosts would not be able to communicate with outside world. Therefore, the flexible IPv4 addressing requires that the **local-proxy-arp** command is configured.

The **arp-populate** command disables dynamic learning of ARP entries (IPv4<->MAC mapping) on an interface based on the ARP protocol. In this case, the ARP table is populated based on the DHCPv4 lease state table which contains IPv4<->MAC mappings obtained through DHCP processing during the host instantiation phase. Arp-populate functionality is highly desirable in case of flexible IPv4 addressing.

When the **arp-populate** command is disabled the ARP entries are dynamically learned based on the ARP protocol. This, in conjunction with flexible IPv4 addressing may cause issues. Consider the following example:

- The subscriber host is instantiated in the 7450 ESS and 7750 SR
- The subscriber interface is unnumbered
- The ARP table does not contain an ARP entry for the subscriber-host

In this case, downstream traffic toward the subscriber host triggers the router to send ARP request for the subscriber host IPv4 address. The router must know the MAC address of the subscriber-host to forward traffic. Because the subscriber interface is unnumbered, the source IPv4 address of the ARP request is unknown and consequently, the ARP request are not sent. As a result, downstream traffic is dropped.

However, the above example is an unlikely scenario. If the subscriber host sends the ARP request for the default-gw first, the router would create an entry in the ARP table for it and the issue would be resolved. This is the most likely outcome because the subscriber host always tries to initiate communication with the outside and therefor ARP for the IPv4 address of the default-gw (which is a 7450 ESS and 7750 SR).

9.18.7 Gi-address configuration consideration

With flexible IPv4 address assignment, the gi-address can be configured as any IPv4 address that is already assigned to an interface (loopback interface, regular interface attached to physical port or subscriber interface) within the same routing instance (VRF or GRT).

9.18.8 PPPoE considerations

PPPoE subscriber hosts do not have the concept of default-gw. Consequently, the default-gw auto-generation concept does not apply to PPPoE hosts.

9.18.9 IPoEv4 considerations

Unnumbered subscriber hosts are instantiated on a subscriber interface that is configured to be unnumbered or to allow unmatching subnets. The IPv4 address of an unnumbered host falls outside the subnets configured on the subscriber interface.

By default, Subscriber Host Connectivity Verification (SHCV) ARP requests for unnumbered hosts are sent with an all-zeros (0.0.0.0) source IP address (also known as the sender IP address). The source IP address can be changed to any unicast IPv4 address in the SHCV policy:

```
configure subscriber-mgmt
        shcv-policy "shcv-policy-1" create
            layer-3
                unnumbered-source-ip 192.0.2.1
            exit
```

```
        exit
```

9.18.10 IPoEv6 considerations

The default-gw for IPoEv6 hosts is link-local IPv6 address. Because this address is always present, there is no need for auto-generation during the subscriber instantiation time.

SLAAC hosts are installed as /64 entries, the length of the installed DHCP-PD prefix is dictated by the prefix-length and the DHCP-NA hosts are installed as /128 entries.

9.18.11 General configuration guidelines for flexible IP address assignment

Flexible IP addressing for IPoE/PPPoE v4 and v6 hosts is by default disabled. In other words, the subscriber hosts are instantiated in the BNG with ability to forward traffic only if their assigned IP addresses belong to one of the configured subnets/prefixes that are associated with the subscriber-interfaces. IPv4 and IPv6 cases are be examined separately:

IPv4:

By default, IPoE and PPPoE subscriber host creation fails in the following two cases:

1. The subscriber-interface does not have an IPv4 address configured, and therefore it is operationally down. This configuration is also known as unnumbered subscriber-interface.
2. The subscriber-interface does have an IPv4 address configured but the IPv4 address assigned to the subscriber host itself is outside of the subscriber-interface configured subnets. In such case, the host is instantiated, but the forwarding is disabled.

Subscriber host instantiation and forwarding can be explicitly enabled for both cases above with flexible IP addressing functionality.

For case 1, this can be achieved by borrowing an IP address for the subscriber-interface from any interface that is operationally up within the routing context. This functionality can be enabled with the **configure service ies** | **vprn** <*service-id*>**subscriber-interface** <*ip-int-name*> **unnumbered** <*ip-int-name* | *ip-address*> command.

To enable forwarding for the subscribers whose IP address falls outside of the configured subnet under the subscriber-interface (case 2), the **configure service ies** | **vprn** <*service-id*> **subscriber-interface** <*ip-int-name*> **allow-unmatching-subnets** command must be entered.

The above commands (**unnumbered** and **allow-unmatching-subnets**) are mutually exclusive. In addition, the **unnumbered** command can be configured only if the subscriber interface does not have an IP address already configured. Otherwise the execution of this command fails.

In both of these cases the host is installed in the routing table as /32.

IPv6:

For IPv6 there is a single command that enables flexible IP addressing for both cases:

1. IPv6 prefixes are not configured under the **sub-if>ipv6** node
2. IPv6 prefixes are configured but the actual address or prefix assigned to the subscriber (by DHCP, LUDB or RADIUS) is outside any prefix that is configured under the **sub-if>ipv6** hierarchy.

This single command is **configure service ies** | **vprn** <*service-id*> **subscriber-interface** <*ip-int-name*> **ipv6 allow-unmatching-prefixes**.

To summarize, the following scenarios are possible:

- PPPoEv4
  - An IPv4 address under the subscriber-interface is configured
    - By default, hosts outside of the sub-intf subnet are instantiated but they are in a non-forwarding-state. Traffic is dropped.
    - **allow-unmatching-subnets** is configured. This command is allowed only if subscriber-interface has also configured its own IPv4 address(es). In this case the IP address for IPCP negotiation is one of the sub-intf addresses. Hosts outside of the sub-intf subnets are instantiated and forwarded.
    - The **unnumbered** <*ip-address* | *ip-int-name*> command is not allowed in this scenario.
  - An IPv4 address under the subscriber-interface is not configured
    - By default, the subscriber-interface is operationally down. Subscribers cannot be instantiated.
    - The **allow-unmatching-subnets** command has no effect because a subscriber-interface does not have an IPv4 address configured and is therefore operationally down. No subscribers can be instantiated.
    - The **unnumbered** <*ip-address* | *ip-int-name*> command is the only viable option in this case. The subscriber-interface borrows an IPv4 address from another interface that is operationally UP and consequently this allows subscribers to be instantiated. This command is mutually exclusive with **allow-unmatching-subnets**. In addition, this command can only be configured if the subscriber interface itself does not have explicitly configured an IPv4 address.
- IPoEv4

Like the PPPoE case above.

IPoEv6 and PPPoEv6

The **allow-unmatching-prefixes** command is independent of any IPv4 command related to flexible IP address assignment (**unnumbered** or **allow-unmatching-subnets**). This command can always be enabled, regardless of the v6 prefixes configured under the **sub-if>ipv6** hierarchy. Any subscriber, regardless of the subscriber interface prefix configuration is instantiated and forwarded.

9.18.12 Restrictions

- Auto-generation of the default-gw IPv4 address is supported only in RCO model with routed RGs/CPEs. Bridged RGs/CPEs are not supported.
- A configured IPv4 address cannot be removed from the subscriber-interface when DHCPv4 hosts under the corresponding subnet are instantiated in the system.
- An IPv4 address cannot be configured under the subscriber-interface while (unnumbered) DHCPv4 hosts under that subnet are already instantiated.
- Executing the **no allow-unmatching-subnets** command is only allowed when there are no unnumbered DHCPv4 hosts instantiated under the subscriber-interface.
- An IPoEv4 subscriber host must be either numbered or unnumbered on both the active and standby BNGs in a multichassis redundant setup. If an IPoEv4 subscriber host is numbered on one BNG and unnumbered on the other, multichassis synchronization fails and displays the **Numbered mode does not match** message as the delete reason in the output of the **tools>dump> redundancy>multi-chassis>sync-database detail** command. An IPoEv4 subscriber host is unnumbered when there is no

matching IPv4 subnet configured on a subscriber interface with **allow-unmatching-subnets** (IPv4) or **unnumbered** (IPv4) enabled.

9.19 uRPF for subscriber management

uRPF is supported for IPv4 and IPv6 dual-stack subscribers with framed routes.

For IPv4, uRPF is supported on group interfaces using anti-spoofing filters. A group interface configured for NATed subscribers is configured with MAC/IP/PPPoE Session-ID anti-spoofing filters.

IPv6 subscribers, which are non-NAT, are always treated as being on a local subnet. For such subscribers, a BNG installs an FDB entry for local routes that match either the wan-host prefix, or the delegated prefix, or both. In strict mode for IPv6 ESM, the uRPF check checks not just that the route matching the SA (which should be a local route, such as a subnet) would route the packet back out of the interface it came in on, but in addition that we would route the packet out to the same SAP it was received on.

SR OS supports the ability to configure a NH-MAC anti-spoof type for non-NATed subscribers. When configured, the datapath performs ingress anti-spoofing based on source MAC address and egress anti-spoof (also referred to as egress subscriber-host look-up) based on the nh-ip address.

The NH-MAC anti-spoof type is configured under the following context:

config>service>vprn>if>sap

config>service>ies>sub-if>grp-if>sap

config>service>vprn>sub-if>grp-if>sap

config>subscr-mgmt>msap-policy

A uRPF check is also performed that prefixes delegated to a subscriber on that MAC address exist in the FDB.

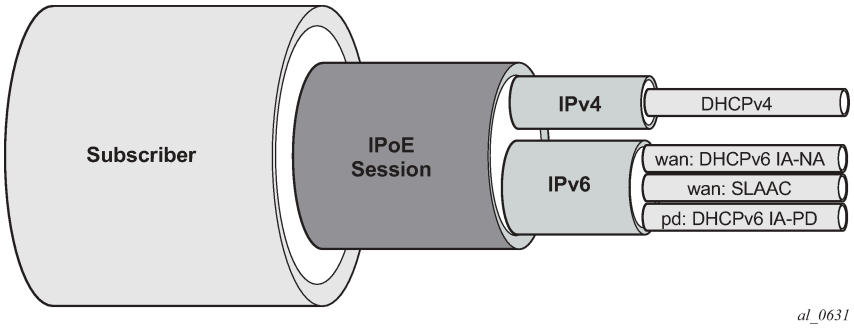
9.20 IPoE sessions

The IP stacks of dual-stack IPoE devices are set up and configured independently using different protocols such as DHCPv4, DHCPv6 or SLAAC. As opposed to PPPoE, there is no single protocol that binds the IP stacks from a single end device together.

To facilitate subscriber management of dual-stack IPoE devices as a single entity similar as for PPPoE sessions instead of handling individual IPoE subscriber hosts, there is a need for a logical IPoE session construct. An IPoE session enables single authentication, session accounting and policy management (mid-session changes) for dual-stack IPoE devices.

An IPoE session is a logical grouping of IPoEv4 and IPoEv6 subscriber hosts that represent the different IP stacks of a single end device and that share authentication data such as subscriber ID, subscriber and SLA profile, session-timeout, and so on. The grouping of subscriber hosts in an IPoE session is based on a configurable session key per group-interface. The IPoE session key includes by default the SAP identifier and MAC address and can be extended with Circuit-Id/Interface-Id or Remote-Id. For DHCPv6 Remote-Id, the enterprise number is excluded from the session-key. Circuit-id/Interface-Id or Remote-id should only be used in the IPoE session key if all subscriber host associated with the IPoE session have this field in their protocol trigger packets. The IPoE session creation (Figure 134: IPoE session) or subscriber host association to an IPoE session fails if the Circuit-Id/Interface-Id or Remote-id is not present in a trigger packet while the field is part of the session-key.

Figure 134: IPoE session



An IPoE session represents a single end device and can have following associated IP stacks:

- **IPv4**

  A single DHCPv4 host.

- **IPv6 WAN**

  One DHCPv6 IA-NA host and one SLAAC host.

- **IPv6 PD**

  One DHCPv6 IA-PD host or PD as managed route.

A violation of the above rules results in a setup failure of the subscriber host when an attempt is made to associate it to the IPoE session.

9.20.1 Enabling IPoE sessions

IPoE sessions are supported in a Routed CO environment with ESM enabled. To enable the IPoE session instantiation, the **ipoe-session** CLI context on the capture SAP (managed SAP scenario) and group-interface must be configured to no shutdown. See also the configuration steps below.

Important Notes:

- Enabling IPoE sessions on a group-interface with active subscriber hosts triggers a migration. Use one of the following CLI command to determine if there are active hosts on a group interface:
  - Check the number of subscriber hosts on a group interface:

    Note that DHCPv6 IA-PD modeled as a managed route are not counted with this command.

    show service id <*service-id*> **subscriber-hosts detail** | **match** <*group-int-name*> | **count**
  - Check the number of IP stacks (Client types) attached on a group-interface. **tools dump router** <*router-instance*> **ipoe-session migration interface** <*group-int-name*>

    The active number of IP stacks (Client types) on the group-interface are listed per type as well as if they are associated with an IPoE session or not.

    Note that DHCPv6 IA-PD modeled as a managed route is also counted in the DHCPv6 type counter.
  - Check the number of DHCPv4 leases, DHCPv6 leases and SLAAC hosts on the group interface that are not attached to a session: **show service id** <*service-id*> **dhcp4 lease-state interface** <*group-int-name*> session none

    show service id <*service-id*> **dhcp6 lease-state interface** <*group-int-name*> **session none**

show service id <*service-id*> **slaac host interface** <*group-int-name*> **session none**

DHCPv6 IA-PD modeled as a managed route is also counted in the DHCPv6 lease state counter.

If there are active hosts on the group interface, make sure you have read the "IPoE session migration" section before enabling IPoE sessions.

- Disabling IPoE sessions by executing an **ipoe-session shutdown** command or **no ipoe-session** command on a group interface deletes all active sessions and associated hosts on that group interface, resulting in service impact for these subscribers. Use one of the following CLI commands to determine if there are active ipoe-sessions on a group-interface:

```
# show service id <service-id> ipoe session interface <group-int-name>
# tools dump router <router-instance> ipoe-session migration interface <ip-int-name>
```

If there are active IPoE sessions on the group interface, be aware that disabling IPoE sessions on the group-interface results in service impact for those sessions.

9.20.2 IPoE session authentication

A single authentication is performed for all subscriber hosts that belong to the same IPoE session. Table 31: IPoE session authentication trigger packets lists the packets that trigger an IPoE session authentication.

Table 31: IPoE session authentication trigger packets

IP stack	Trigger packets
IPv4	DHCPv4 Discover
	DHCPv4 Request
IPv6 WAN	DHCPv6 Solicit
	DHCPv6 Request
	DHCPv6 Relay Forward (Solicit)
	DHCPv6 Relay Forward (Request)
	Router Solicitation
IPv6 PD	DHCPv6 Solicit
	DHCPv6 Request
	DHCPv6 Relay Forward (Solicit)
	DHCPv6 Relay Forward (Request)

When a trigger packet is received on a capture SAP or group-interface with IPoE sessions enabled, an IPoE session lookup is performed based on the configured IPoE session key:

- If no IPoE session is found, a new session is created and authenticated following the ESM authentication configuration such as local user database lookup, Radius or Diameter authentication,

defaults, and such. After successful authentication, the authentication data is stored in the IPoE session state. The subscriber host is created and associated with the session.

- If an IPoE session already exists, and no re-authentication must be performed then the subscriber host is created using the stored IPoE session data. The subscriber host is associated with the session.
- If an IPoE session already exists, and re-authentication must be performed then the session is re-authenticated. When successful, the authentication data for the IPoE session is updated and applied to all associated hosts. The subscriber host is created and associated with the session. When unsuccessful, existing hosts associated with the session are not impacted and the session data is kept unchanged.

Re-authentication is by default disabled for IPoE sessions. To enable re-authentication, a minimum authentication interval must be configured. The min-auth-interval CLI parameter configures the maximum frequency of re-authentications by specifying a minimum interval between two non-forced authentications for the same IPoE session. A re-authentication is triggered by the renewal of any host belonging to the IPoE session. Setting the min-auth-interval to zero seconds, always re-authenticates on each trigger packet. The **re-authentication** command in a RADIUS authentication policy is ignored for IPoE session authentication.

A forced authentication is performed when the Circuit-Id/Interface-Id or Remote-Id in the trigger packet has changed. An empty or absent Circuit-Id/Interface-Id or Remote-Id is not considered as a change. The default forced authentication behavior is changed with the **force-auth** command in the **group-interface**>**ipoe-session** context: only force authenticate on Circuit-Id/Interface-Id change or only force authenticate on Remote-Id change or disable forced authentications.

A new local user database config in the ipoe-session CLI context on a capture SAP or group interface ensures that all subscriber hosts associated with an IPoE session are using the same database and therefore common match criteria. The per subscriber host type user-db configurations, such as **ipv6 dhcp6 user-db**, **dhcp user-db**, and **rtr-solicit-user-db** are ignored when IPoE sessions are enabled.

9.20.3 IPoE session accounting

All RADIUS accounting modes can be enabled for IPoE sessions: queue instance accounting, host accounting or session accounting.

With session accounting, a RADIUS accounting start is generated when the first host of the session is created and an accounting stop when the last host of the session is deleted. The generation and interval of periodic interim updates can be configured. Optionally, triggered interim update messages can be generated when a host is deleted from the session or an additional host is associated.

A unique accounting session ID is generated for the IPoE session and is used in RADIUS session accounting. The IPoE session accounting session ID can be included in the RADIUS Access Request message the **config**>**subscr-mgmt**>**auth-plcy**> **include-radius-attribute acct-session-id session** command.

This accounting session ID can also be used in RADIUS CoA or Disconnect Messages to target the IPoE session.

9.20.4 IPoE session mid-session changes

Mid-session changes such as those initiated by RADIUS CoA or Diameter Gx RAR are applied to all hosts associated with the IPoE session.

A RADIUS CoA message targeting any host of an IPoE session has the same effect as a RADIUS CoA message targeting the IPoE session using the IPoE session Acct-Session-Id as key: all host of the session are targeted and the session state is updated with the new data.

The following **tools** commands are available to manually enforce a mid-session change:

```
# tools perform subscriber-mgmt edit-ipoe-session sap <sap-id> mac <mac-address>
[subscriber <sub-ident-string>] [sub-profile-string <sub-profile-string>] [sla-
profile-string <sla-profile-string>] [inter-dest-id <intermediate-destination-id>]
[ancp-string <ancp-string>] [app-profile-string <app-profile-string>] [circuit-id
<circuit-id>] [remote-id <remote-id>]

# tools perform subscriber-mgmt eval-ipoe-session [svc-id <service-id>] [sap <sap-id>]
[mac <mac-address>] [circuit-id <circuit-id>] [remote-id <remote-id>] [subscriber
<sub-ident-string>]
```

9.20.5 IPoE session termination

When the last subscriber host associated with an IPoE session is deleted from the system, then the IPoE session is also deleted.

An IPoE session and all associated subscriber hosts can be deleted by the following:

- CLI clear command: **clear service id** <*service-id*> **ipoe session**
- An **ipoe-session no shutdown** CLI command on a group-interface
- A **no ipoe-session** CLI command on a group-interface. This command resets to the default behavior, which is IPoE sessions disabled.
- Session timeout, configured in the IPoE session policy or obtained from AAA
- Idle timeout
- RADIUS Disconnect Message
- Diameter Gx session termination
- Credit Control: Radius or Diameter Gy

9.20.6 Limiting the number of IPoE sessions

See Limiting subscribers, hosts, and sessions for a detailed description of the configuration options to use to limit the number of IPoE sessions per SAP, per group-interface, per SLA profile instance, or per subscriber.

9.20.7 SAP session index

The system keeps track of the number of IPoE sessions active on a specified SAP and assign a per SAP session index to each so that the lowest free index is always assigned to the next active IPoE session. When RADIUS authentication is used, the SAP session index can be sent to, and received from, the RADIUS server using the [26-6527-180] Alc-SAP-Session-Index attribute.

It should only be used in a subscriber per VLAN model as the session index is per SAP.

The SAP session index allows IPoE sessions in a bridged RG environment to have their own set of queues for QoS and accounting purposes when using the same SLA profile name received from a RADIUS server. See Subscriber per PPPoE Session Index for further details.

Alternatively, this can be achieved by configuring per-session SPI sharing in the SLA profile as described in SLA profile instance sharing.

9.20.8 Resiliency

For non-redundant BNG deployments, the IPoE session state is stored in the subscriber-mgmt persistency file for recovery from Compact Flash after a node maintenance operation or failure. This is configured at the system persistence CLI context.

For multi chassis redundancy scenarios, the IPoE session state is synchronized by the "sub-mgmt ipoe" Multi Chassis Synchronization (MCS) application.

9.20.9 Notes

- Static hosts can be configured on a group-interface with IPoE sessions enabled. A static host is not associated with an IPoE session.
- Up to fifty Framed-Routes and fifty Framed-IPv6-Routes can be associated with an IPoE session.
- A fall back action (accept or local user database lookup) when no Radius servers are available for Radius authentication can be specified for IPoE sessions.
- Lawful Intercept sources initiated from Radius always include all IP stacks from the IPoE session regardless the targeted host in the CoA message.
- ARP hosts are not supported in an IPoE session and cannot be instantiated on a group-interface with IPoE sessions enabled.
- The creation of an IPv4 host using the Alc-Create-Host attribute in a Radius CoA message is not supported on a group-interface with IPoE session enabled.
- A local user database host identification based on option60 is ignored when authenticating an IPoE session.
- RADIUS authentication of an IPoE session fails when the **user-name-format** is configured to **mac-giaddr** or **ppp-user-name**.
- The DHCP Python module (alc.dhcp) used to derive subscriber host attributes from a DHCPv4 ACK message is not supported in combination with IPoE sessions.
- Subscriber Host Connectivity Verification (SHCV) continues to work on a per-stack basis. In other words, in a dual-stack scenario with SHCV action remove enabled for both stacks, a failure in IPv4 connectivity does not clean up the session unless the IPv4 subscriber host was the last associated host.

9.20.10 Configuration steps

To create an IPoE session policy:

```
config
    subscr-mgmt
        ipoe-session-policy "ipoe-policy-1" create
```

```
            description "Default IPoE session policy"
            session-key sap mac            # default
            no session-timeout             # default
        exit
```

Enable IPoE sessions on the capture SAP and group interface.

If IPoE sessions is enabled on a capture-sap, then it must also be enabled on the target group-interface. If an IPoE session local user database lookup is configured at the capture-sap, then the same local user database lookup must be configured at the target group-interface.

```
config
    service

        vpls 10 customer 1 create
            ---snip---
            sap 1/1/4:*.* capture-sap create
                ---snip---
                ipoe-session
                    description "IPoE sessions - capture-sap"
                    ipoe-session-policy "ipoe-policy-1"
                    user-db "ludb-1"
                    no shutdown
                exit

        ies 1000 customer 1 create
            subscriber-interface "sub-int-1" create
                ---snip---
                group-interface "group-int-1-1" create
                    ---snip---
                    ipoe-session
                        description "IPoE sessions - IES group-interface"
                        force-auth cid-change rid-change       # default
                        ipoe-session-policy "ipoe-policy-1"
                        min-auth-interval infinite             # default
                        sap-session-limit 1                    # default
                        session-limit 1000
                        user-db "ludb-1"
                        no shutdown
                    exit
```

To display the IPoE session state, use following command:

```
# show service id <service-id> ipoe session [detail]
```

9.20.11 IPoE session migration

This section is only applicable when enabling IPoE sessions on a group interface with active subscriber hosts. When there are no active subscriber hosts on a group interface, there is no need for a migration. Use one of the following CLI commands to determine if there are active hosts on a group interface:

- Check the number of subscriber hosts on a group interface: # **show service id** <*service-id*> **subscriber-hosts detail** | **match** <*group-int-name*> | **count**

  DHCPv6 IA-PDs modeled as a managed route are not counted with this command.

- Check the number of IP stacks (client types) attached on a group interface. # **tools dump router** <*router-instance*> **ipoe-session migration interface** <*group-int-name*>

  The number of active IP stacks (client types) on the group interface are listed per type whether they are associated with an IPoE session.

  DHCPv6 IA-PDs modeled as a managed route are also counted in the DHCPv6 type counter.

- Check the number of DHCPv4 leases, DHCPv6 leases, and SLAAC hosts on the group interface that are not attached to a session: # **show service id** <*service-id*> **dhcp4 lease-state interface** <*group-int-name*> **session none** # **show service id** <*service-id*> **dhcp6 lease-state interface** <*group-int-name*> **session none** # **show service id** <*service-id*> **slaac host interface** <*group-int-name*> **session none**

  DHCPv6 IA-PDs modeled as a managed route are also counted in the DHCPv6 lease state counter.

By default, IPoE sessions are disabled on a group interface (**ipoe-session shutdown**). Enabling IPoE sessions on a group interface with active subscriber hosts starts a migration process and should be planned carefully to allow a seamless migration.

A migration is required because of the nature of IPoE sessions: a single authentication is performed for all hosts (IP stacks) of a dual-stack end device. All hosts (IP stacks) in an IPoE session share the same MAC address, SAP, and optionally Circuit-ID / Interface-ID or Remote-ID which are configured as the **session-key** in the **ipoe-session-policy**. To determine if hosts (IP stacks) belong to a single session, a new trigger packet is required to obtain the session key.

To guarantee a correct IPoE session configuration and a correct authentication database, the migration is performed when the host state is renewed, and a new trigger packet is received:

- DHCPv4 renew or rebind for DHCPv4 hosts
- DHCPv6 renew or rebind for DHCPv6 hosts (IA-NA and IA-PD)
- DHCPv4 renew for IPoE linked SLAAC hosts
- Router Solicit for RS triggered SLAAC hosts

The duration of a migration is therefore dependent on the lease times for DHCPv4 and DHCPv6 hosts and for IPoE linked SLAAC hosts. If possible, the lease times could temporarily be reduced to a couple of hours to facilitate the migration process.

The actual migration is started by the arrival of a new trigger packet of an IP stack (host) that is not associated with an IPoE session. The IPoE session key is composed of the data in the trigger packet (MAC address and SAP, by default). If an IPoE session exists for the obtained IPoE session key, the corresponding session data is used for authentication. If no IPoE session exists for the obtained IPoE session key, authentication is performed, and based on the result, a new IPoE session is created. The old host state is deleted from the system and a trap is sent to indicate that this host is being migrated. A new host (IP stack) is created and associated with the IPoE session. When RADIUS accounting is enabled, this may result in an accounting start and stop depending on the accounting mode. For host accounting, an accounting stop is followed immediately by an accounting start. For queue instance accounting, an accounting stop is generated when the last host associated with the queue instance is migrated. An accounting start is generated when the first host is associated with the IPoE session.

The following notes must be considered for the migration procedure:

- For multichassis redundant nodes, IPoE sessions should be enabled first on the standby node and immediately thereafter on the active node.
- A renew as part of a DHCPv4 lease split operation does not trigger a migration to the IPoE session. The migration starts only when the renew is forwarded to the DHCP server.

- For DHCPv4 RADIUS proxy scenarios, it is recommended that the lease time be specified the with the [26-6527-174] Alc-Lease-Time RADIUS attribute instead of the [27] Session-Timeout attribute. After migration, the [27] Session-Timeout attribute is interpreted as the number of seconds before the session is terminated.
- DHCPv6 IA-PD modeled as a managed route may migrate separately from the IPv6 SLAAC host it is associated with for its next-hop. This could result in a temporary service impact until both the managed route and next-hop host are migrated.
- The migration of idle Router Solicit SLAAC hosts can be facilitated by specifying an inactivity timer.
- When the subscriber ID is auto-generated (**auto-sub-id**), then a new **sub-id** is be generated after migration. This may result in a temporary increase in used resources such as queues until all hosts from a subscriber are migrated.

Example high-level migration steps.

Important notes:

- It is recommended that a migration plan be built for the target network and validate the plan in advance in a lab environment.
- It is recommended that the migration be performed per group interface or capture SAP with all possible target group interfaces and that the next migration only be started when the previous one is successfully completed.
- When managed SAPs (MSAPs) are used, enabling an IPoE session on a group interface while not enabling IPoE sessions on the corresponding capture SAP, or enabling an IPoE session on a capture SAP while not enabling IPoE sessions on the target group interface, results in session setup failures for sessions where no MSAPs exist.

1. Using the CLI commands described at the beginning of this section, check if an IPoE session migration is applicable. A migration is not required when there are no active subscriber hosts on the target group interfaces.
2. Check if all preconditions are met:
   a. There are no conflicting requirements with IPoE sessions such as ARP host support on the same group interface or local user database authentication based on option 60. Check the Notes section above for a list of possible conflicts.
   b. IPoE session configuration is complete on the group interfaces and corresponding **capture-sap**: **ipoe-session-policy** (**session-key**) and on the optional local user database. On the group interfaces, the IPoE session limits should be configured as needed using the **session-limit** and **sap-session-limit** commands.
   c. Authentication servers are up to date to provide all required authentication data for a single dual-stack end device based on single authentication (for example, return both IP address and IPv6 prefix in a proxy scenario).
   d. Accounting servers are ready to deal with accounting stop/start when hosts migrate to an IPoE session.
3. Take a snapshot of the active hosts before the migration. Use the commands as described above. The following command provides a summary view: **tools**>**dump**>**router** <*router-instance*>**ipoe-session**>**migration**>**interface** <*group-int-name*>
4. Start the migration by enabling an IPoE session on the group interface and for MSAPs, by enabling an IPoE session on the capture SAP.

5. Monitor the progress during migration. Review the events (for example, by using the **show log log-id 99** command) and check the number of hosts migrated with the CLI command:

   tools>**dump**>**router** <*router-instance*>*ipoe-session*>**migration**>**interface** <*group-int-name*>

   The following event is generated when a host is deleted because of a migration:

   4 2015/06/29 19:37:57.47 UTC WARNING: SVCMGR #2559 Base IPoE session "IPoE session migration deleted host 2001:db8:2:101::1 on SAP 1/1/4:1201.2 in service 1000"

   2 2015/06/29 19:37:29.41 UTC WARNING: SVCMGR #2559 Base IPoE session "IPoE session migration deleted host 10.1.1.101 on SAP 1/1/4:1201.2 in service 1000"

   DHCP lease states and SLAAC host states associated with IPoE sessions can be found with:

   show service id <*service-id*> **dhcp4 lease-state interface** <*group-int-name*> **session ipoe**

   show service id <*service-id*> **dhcp6 lease-state interface** <*group-int-name*> **session ipoe**

   show service id <*service-id*> **slaac host interface** <*group-int-name*> **session ipoe**

   The migration is finished when all hosts are associated with an IPoE session. The counters in the column “Non IPoE session” should be all zero. For example:

```
# tools dump router ipoe-session migration interface group-int-1-1
===============================================================================
 Type session               Total     IPoE session     Non IPoE
===============================================================================
 Group-interface: group-int-1-1 (IPoE session enabled)
-------------------------------------------------------------------------------
 DHCPv4              16384     16384            0
 DHCPv6              16384     16384            0
 SLAAC               4096      4096             0
-------------------------------------------------------------------------
 IPoE sessions       20480
=========================================================================
```

6. Perform post migration steps. For example, verify that the number of users before and after the migration are in the same order of magnitude (users may connect and disconnect during the migration). Enable session accounting if required.

9.20.11.1 Additional notes for IPoE session migration of IPv4 hosts as a control channel for dynamic data services

During the migration of an IPv4 host as a control channel for Dynamic Data Services to an IPoE session as a control channel, the associated dynamic data services are deleted and recreated based on the IPoE session authentication data.

When IPoE sessions are enabled on the group interface, at the next DHCPv4 renew or rebind:

- the IPv4 host (control channel) is deleted
- the associated dynamic data services are deleted
- the IPv4 host is added in the IPoE session
- new dynamic data services are created based on IPoE session authentication data

9.21 Data-triggered subscriber management

Prerequisites

This feature allows the creation of ESM subscribers and hosts based on the receipt of upstream data packets.

Data-triggered host creation does not rely on protocol triggers (DHCP, PPPoE) or management triggers (static hosts) to create each host, and is especially useful in the following cases:

- BNG manages subscribers under Layer 3 nodes (BRAS, CMTS, GGSN/PGW, and so on) and is not on the DHCP message path.
- BNG needs to manage large numbers of static hosts and bulk provisioning is required.

BNG authenticates, creates, and deletes subscriber hosts as follows:

Procedure

Step 1. The subscriber SAP, including MSAP, receives a user packet that does not match existing anti-spoof table entries.

Step 2. BNG instantiates an IPoE session if there is no existing session with the same session key, and performs authentication using LUDB and RADIUS.

Note: Data-triggered ESM is supported only with IPoE sessions.

Step 3. A subscriber host is created with the ESM strings provided during authentication.

Step 4. The subscriber host is deleted when **session-timeout** or **idle-timeout** expires, CoA triggers a disconnect, SHCV check failure, or management (CLI, SNMP, and so on) triggers a host deletion.

Note: There are no automatic triggers to delete a host if **session-timeout** or **idle-timeout** and SHCV are not configured.

9.21.1 Provisioning data-triggered ESM

Data-triggered ESM can be enabled on a group interface. The following displays a sample configuration of data-triggered ESM:

```
subscriber-interface "SI1" create
     group-interface "GI1" create
          arp-populate
          ipoe-session
               ipoe-session-policy "IS1"
          exit
          data-trigger
               no shutdown
          exit
```

An IPoE session and ARP population are mandatory when configuring data-triggered ESM.

The following packets can start data-trigger processing:

- IPv4 data packet

- IPv6 data packet
- ARP

To terminate IPv6 hosts that send neighbor RS/NS before sending data packets, **auto-reply** must be configured.

```
subscriber-interface "SI1" create
     group-interface "GI1" create
          ipv6
               auto-reply
                    neighbor-solicitation
               exit
```

For MSAP, the "data" trigger packet type can accept data triggers.

Example: MD-CLI

```
configure service vpls "capture-vpls"
    capture-sap 1/1/2:200.* {
        trigger-packet {
            data true
        }
    }
```

Example: classic CLI

```
configure service vpls "10"
            sap 1/1/2:200.* capture-sap create
                trigger-packet data
                no shutdown
            exit
```

9.21.2 Authentication and host creation

Authentication of a data trigger can use LUDB configured in an IPoE session statement under a group interface.

To identify the source IPv4/IPv6 address of data-trigger packets, the IP prefix in the local user database can be configured with **host-identification**:

```
local-user-db "LUDB_DT" create
     ipoe
          match-list ip
          host "10.0.0.8" create
               host-identification
                    ip-prefix 10.0.0.8/29
               exit
          host "2001:1:b::1" create
               shutdown
               host-identification
                    ip-prefix 2001:a:b::1/128
```

Note: Only one IP prefix can be configured for each host. A dual-stack host requires two local user database host entries if the IP prefix needs to be used for host identification.

For RADIUS authentication, the circuit ID includes the source IPv4/IPv6 address of the data-trigger packet:

```
authentication-policy "AUTH1" create
     user-name-format circuit-id
     include-radius-attribute
          circuit-id
     exit
```

If IPoE session policy uses circuit ID to identify each session, a new IPoE session is created for each source IPv4/IPv6 address. However, RADIUS can return the circuit ID to merge multiple IPoE sessions with the same SAP, MAC, and circuit ID into a single session.

A host is created using the IPv4/IPv6 source address of the data trigger (a /32 address for IPv4 or a /128 address for IPv6), but IPv6 data-triggered hosts can be created as an IPv6 prefix by configuring **ipv6-delegated-address** in the local user database host entry.

RADIUS can return the following AVPs to model the address/prefix of the data-triggered host:

- Framed-IP-Address: /32 IPv4 address of the host
- Framed-Route: managed IPv4 route with the host as next hop
- Alc-IPv6-Address: /128 IPv6 address of the host
- Delegated-IPv6-Prefix: IPv6 prefix of the host
- Framed-IPv6-Route: managed IPv6 route with the host as next hop

Information on multiple hosts can be returned in a single Access-Accept message when the **nh-mac anti-spoof** command is configured. This is mandatory when provisioning dual-stack hosts with the same SAP and MAC addresses with **nh-mac anti-spoof** configured but is mutually exclusive with the CID key in the IPoE session policy.

9.21.3 DoS protection

To authenticate data triggers, only the first packet is used for further processing. Subsequent packets from the same source are discarded until ESM host creation.

Data trigger packets are classified as all-unspecified protocol by Distributed CPU Protection (DCP).

9.21.4 DHCP promotion

DHCP promotion allows data-triggered subscriber hosts to become DHCP hosts.

After a data-triggered host is created, DHCP packets sent by the client starts the DHCP promotion process as follows:

1. A DHCP Request/Renew/Rebind message comes from the data-triggered host.
2. Authentication using LUDB and RADIUS is performed. A RADIUS Access-Request message is sent if the authentication policy has re-authentication enabled. DHCP processing is done without authentication if re-authentication is not configured.
3. An Access-Accept message that contains ESM attributes is sent back from the AAA server. If an Access-Reject message is received, the data-triggered host is deleted.
4. A DHCP packet is relayed to the DHCP server.
5. A DHCP server replies with a DHCP Ack/Relay-Reply or Nak message.

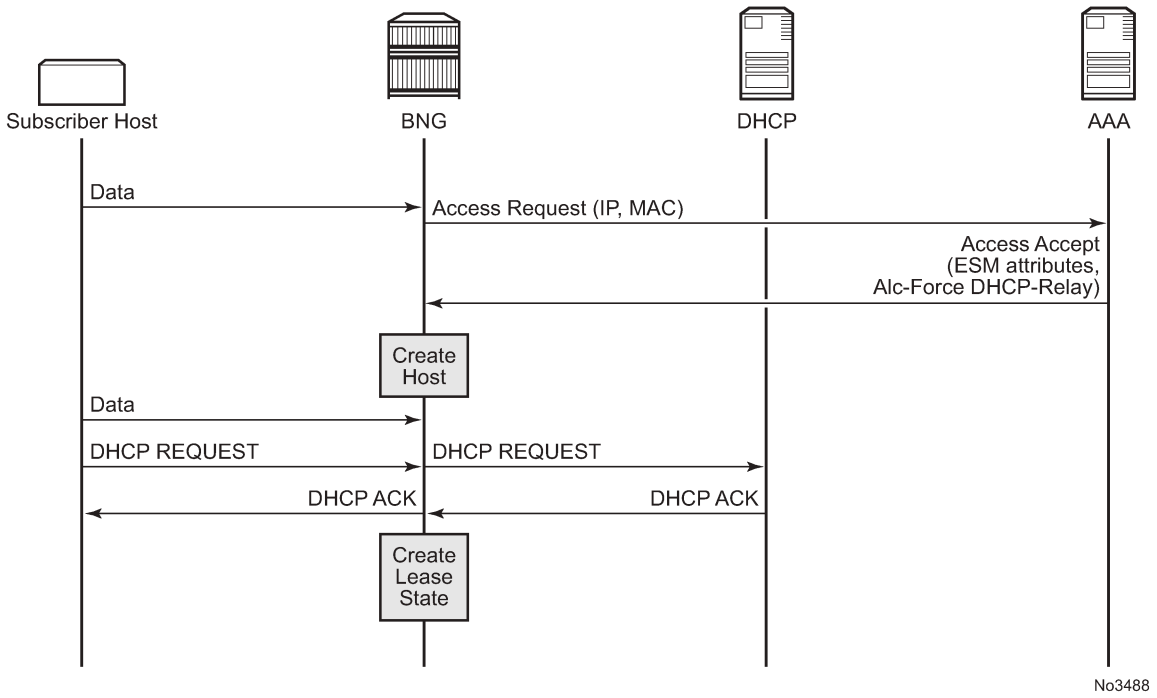
6. When an Ack/Relay-Reply message is received, a lease state is created, and the data-triggered host is promoted to the DHCP host.
7. If the DHCP server replies with Nak or with IP address information different from the existing data-triggered host, DHCP promotion fails and the data-triggered host is deleted.
8. With DHCP proxy, if the LUDB or AAA server returns IP information different from the data-triggered host, then DHCP promotion fails and the data-triggered host is deleted.
9. An interim accounting message is generated based on the configuration of the **radius-accounting-policy** command, as follows:
   a. **queue-instance-accounting**: interim-update with Alc-Acct-Triggered-Reason = Promotion of a Data-triggered host.
   b. **session-accounting**: interim-update with Alc-Acct-Triggered-Reason = Promotion of a Data-triggered host, if host-update is configured.
   c. **host-accounting**: interim-update with Alc-Acct-Triggered-Reason = Promotion of a Data-triggered host

Data-triggered hosts can be promoted to DHCP proxy hosts by default. To promote data-triggered hosts using DHCP relay to an internal or external DHCP server, the Alc-Force-DHCP-Relay VSA is included in Access-Accept messages to authenticate data-triggered hosts.

Figure 135: DHCP promotion with DHCP relay shows DHCP promotion with DHCP relay.

Note: DHCP relay promotion is only supported when using RADIUS. LUDB and NASreq is not supported.

Figure 135: DHCP promotion with DHCP relay



9.21.5 Data-triggered SLAAC hosts

A SLAAC host can be created instead of a data-triggered host on data-triggered authentication when the Framed-IPv6-Prefix is returned from the LUDB or AAA server and the IPv6 prefix in the AVP value matches the source address of the data-trigger.

9.21.5.1 Data-triggered subscriber management and LAA

The following process describes how data-triggered subscriber management works with LAA for SLAAC, configured in the **config**>**service**>**ies** | **vprn**>*sub-if*>**grp-if**>**lcl-addr-assign**>**ipv6**>**client-application ipoe-slaac** contexts.

- Data-triggered IPv6 ESM hosts are created with the prefix specified by LUDB, RADIUS, or from a source address of the data-trigger. The host creation fails if an overlapping address and prefix is found in the host table.
- Data-triggered IPv6 ESM host creation fails if AAA returns a Framed-IPv6-Pool AVP with no addressing information.
- Protocol-based IPv6 ESM host creation with LAA fails if LAA returns the address and prefix overlaps with an existing data-triggered ESM host.

LAA for **ipoe-wan** and **data-triggered** subscriber management cannot coexist on the same SAP. Data-triggered subscriber management, in the **config**>**service**>**ies** | **vprn**>*sub-if*>**grp-if**>**data-trigger**>**no shutdown** context, and LAA commands, in the **config**>**service**>**ies** | **vprn**>*sub-if*>**grp-if**>**lcl-addr-assign**>**ipv6**>**ipoe-wan** context, are mutually exclusive.

9.21.6 Stateful multichassis redundancy (MCS)

Data-triggered subscriber hosts can be protected with stateful multichassis redundancy. Subscriber management MCS applications, under **config**>**redundancy**>**multi-chassis**>**peer**>**sync**>**sub-mgmt** also include data-triggered subscriber host information.

9.21.7 Stateless multichassis redundancy

Stateless redundancy uses SRRP in the same form as the current implementation for active and standby selection and peer liveness detection, but does not need subscriber state synchronization using MCS that requires CPM/IOM resources to be on standby node.

Stateless redundancy has the following characteristics:

- IPoE sessions and their hosts (DHCP,DHCP6,SLAAC, and DT) are not synchronized by the subscriber management IPoE MCS client.
- SRRP is synchronized over MCS.
- Active BNG in the (SRRP master state) only processes the data-trigger and authenticates and creates a subscriber host state.
- Standby BNG in the (SRRP non-master state) discards all upstream packets.
- Shunt and redundant interfaces are not supported.
- After an SRRP switchover, the new active BNG starts processing subscriber traffic. The previous active BNG deletes all IPoE sessions and IPoE subscriber hosts on an SRRP switchover. Accounting stops are sent to indicate that this node became standby (non-master SRRP state). Accounting messages are different based on the **radius-accounting-policy** configuration.

  1. **queue-instance-accounting**:

     accounting-stop with Alc-Error-Code = Node has switched to stateless backup

2. **session-accounting**:

   interim-update with Alc-Acct-Triggered-Reason = Node has switched to stateless backup, upon each stack deletion if host-update is configured, and accounting-stop with Alc-Error-Code = Node has switched to stateless backup

3. **accounting**:

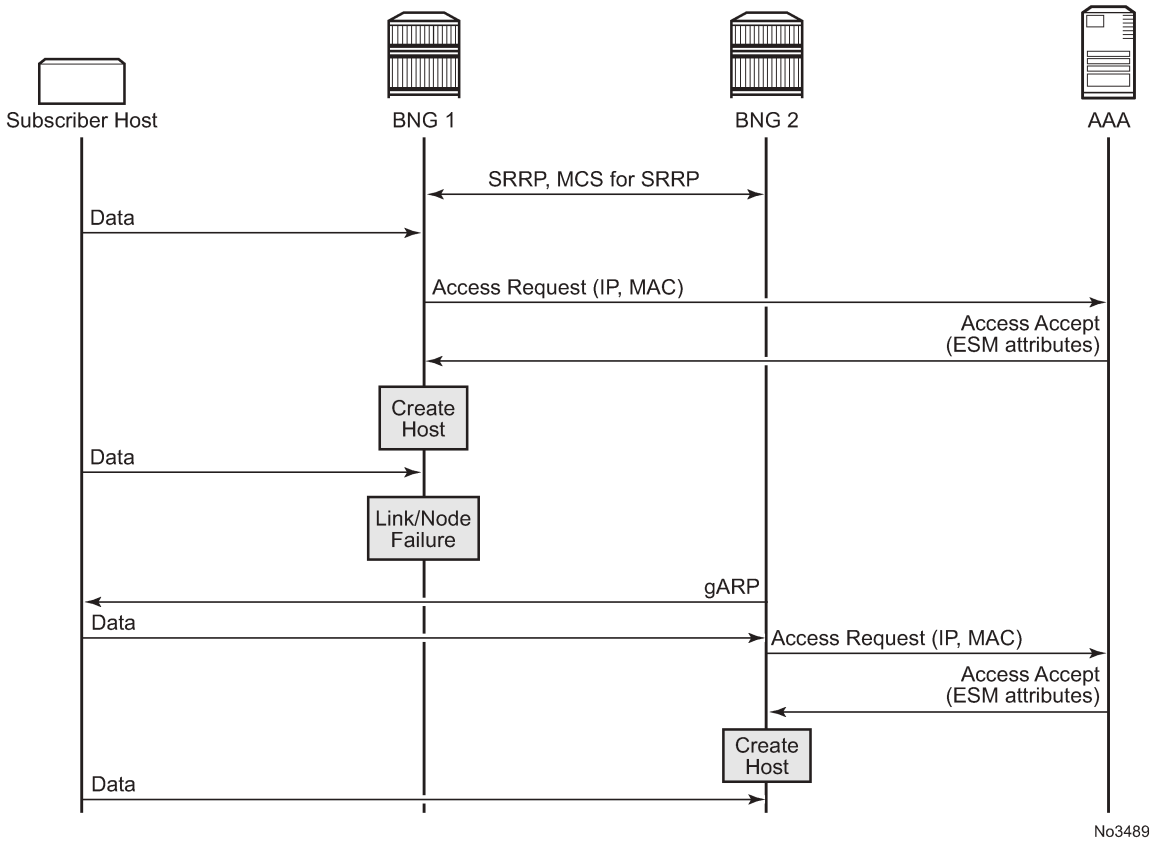
   accounting-stop with Alc-Error-Code = Node has switched to stateless backup, upon each host deletion

- The DHCP local server state can be synchronized for DHCP promotion.
- DHCP promotion for BRG requires lease state synchronization between redundant BNGs, which disables protocol-triggered IPoE ESM without an IPoE session. This requires the **config>service>ies | vprn>sub-if>grp-if>ipoe-session>stateless-redundancy** command to be configured.

  host-PPPoE (PTA and LAC) processing can be tied with an SRRP instance for stateless redundancy. Only the active BNG (SRRP master state) processes PPPoE/PPP control plane data. A standby BNG (SRRP non-master state) does not send LCP echo messages. After switchover, subscriber hosts retry to connect after echo timeout.
- Supports both static SAP and MSAP.

Figure 136: Stateless multichassis redundancy shows an example of stateless multichassis redundancy.

Figure 136: Stateless multichassis redundancy



9.21.7.1 MSAP support

Stateless redundancy does not have information about MSAP because MCS is not used for synchronizing subscriber host information.

Static SAPs must be configured to rewrite FDBs on Layer 2 switches with G-ARP in access or aggregation networks.

Use the following CLI commands to rewrite FDBs with G-ARP.

```
group-interface <ip-int-name>
 sap <sap-id> create
 no shutdown
```

This requires the aggregation network to re-learn MAC for multiple C-VLANs using a single G-ARP from a specific C-VLAN.

Two typical scenarios are supported:

- MSAP has QinQ encap, and aggregation switches have per SVLAN FDB
- MSAP has dot1q encap, and aggregation switches have per SVLAN FDB

  (Aggregation switches push SVID)

9.21.8 IPv6 prefix learning

IPv6 prefix learning enables the creation of data-triggered IPv6 prefix-based hosts by learning prefixes from source IPv6 addresses of data-trigger packets where the prefix lengths are specified in LUDB or AAA.

IPv6 prefix learning is enabled with the following LUDB configuration conditions.

- There are no **ipv6-address**, **ipv6-delegated-prefix**, or **ipv6-slaac-prefix** matching data-triggers in the LUDB host entry.
- The **ipv6-delegated-prefix-length** is defined in the LUDB host entry.

IP prefix learning from AAA is enabled with the following conditions.

- There are no Delegated-IPv6-Prefix, Framed-IPv6-Prefix, or Alc-IPv6-Address AVP-matching data-triggers in the Access-Accept or the AA-Answer responses.
- Delegated-IPv6-Prefix AVP exists in the Access-Accept or the AA-Answer responses.

DHCP promotion is also supported on the data-triggered host created with IPv6 prefix learning.

9.22 RADIUS subscriber services

RADIUS subscriber services enable the activation and deactivation of subscriber functions by RADIUS Access-Accept or CoA messages. Each subscriber service can have its own RADIUS accounting session.

The subscriber service functionality is built using the flexible RADIUS Python script interface to populate the subscriber service data structure using a parameter list received in subscriber service-specific RADIUS Vendor Specific Attributes (VSAs). The format and content of the VSA parameter list is defined by the operator. An accounting start/stop is sent when the subscriber service is activated/deactivated. Optionally, interim updates can be sent in intervals that can be specified per subscriber service instance. Accounting

interim updates and stop messages contain the subscriber service-related statistics (time or volume and time).

Subscriber services can be activated on a single-stack or dual-stack PPPoE or IPoE session or on a single-stack IPv4 host.

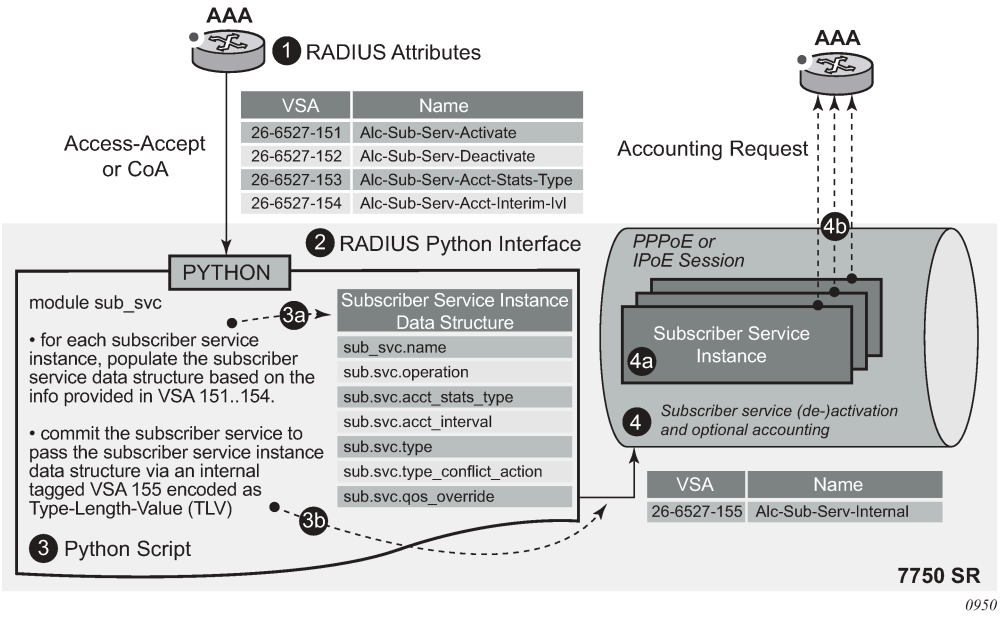
Subscriber service functionality can be built with:

- QoS overrides: changing queue or policer parameters (PIR/CIR rates and CBS/MBS burst sizes), adapting rates of a parent scheduler, root arbiter, or subscriber aggregate rate
- PCC rules: applying QoS or filter actions to a set of IP flows

9.22.1 Subscriber service building blocks

Figure 137: Subscriber services building blocks shows the building blocks required to activate or deactivate a subscriber service.

Figure 137: Subscriber services building blocks



Each of the building blocks is described in the following sections.

9.22.1.1 RADIUS access-accept or CoA message with subscriber service activate or deactivate VSAs

A subscriber service instance is activated from the RADIUS server by an Access-Accept or CoA message for a PPPoE or IPoE session. Deactivation of a subscriber service instance can be achieved by a RADIUS CoA message or is implicit when the associated subscriber session terminates.

To activate a subscriber service, the Alc-Sub-Serv-Activate (VSA) is used, and to deactivate a subscriber service, the Alc-Sub-Serv-Deactivate VSA is used. The formats of the Alc-Sub-Serv-Activate and Alc-Sub-

Serv-Deactivate VSAs can be freely defined by the operator if they match with the Python script that is used to commit the subscriber service instance activation or deactivation.

For example, to change the upstream and downstream bandwidth of an IPoE session, the following format can be defined:

```
rate-limit;<upstream_bw_in_kbps>;<downstream_bw_in_kbps>
```

To activate a subscriber service with an upstream bandwidth of 5 Mb/s and a downstream rate of 50 Mb/s, the following VSA can then be included in a RADIUS Access-Accept or CoA message:

```
Alc-Sub-Serv-Activate = "rate-limit;5120;30720"
```

To deactivate the same subscriber service and revert to the initial bandwidth, the following VSA can be included in a RADIUS CoA message:

```
Alc-Sub-Serv-Deactivate = "rate-limit;5120;30720"
```

To deactivate a subscriber service instance, its unique name must be used. In the example above, the name equals "rate-limit;5120;30720".

To start an accounting session when the subscriber service instance is activated, the following attributes can be included in the Access-Accept or CoA message:

```
Alc-Sub-Serv-Acct-Stats-Type = volume-time | time
Alc-Sub-Serv-Acct-Interim-Ivl = <update-interval>
```

For example, the Alc-Sub-Serv-Acct-Stats-Type attribute value is set to "volume-time" to include both the session time for time-based billing and standard counters for volume statistics collection. The Alc-Sub-Serv-Acct-Interim-Ivl attribute sets the interval for interim updates of the subscriber service instance accounting.

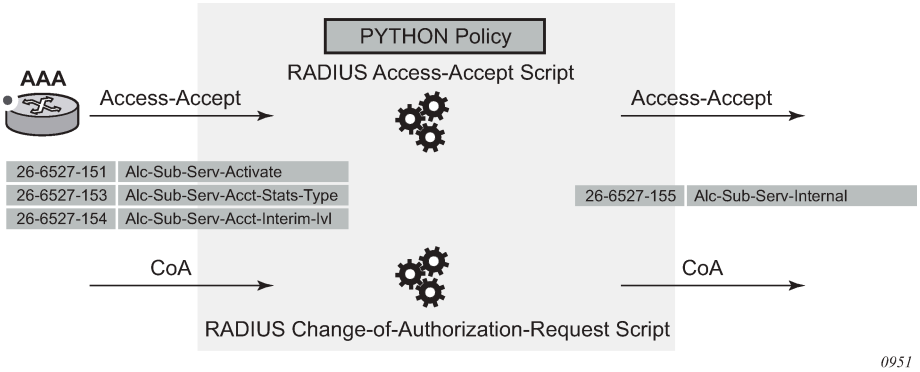
See the Subscriber services RADIUS VSAs section for details on RADIUS attributes.

See the Subscriber service RADIUS accounting section for details on subscriber service instance accounting.

9.22.1.2 RADIUS Python interface

The SR OS RADIUS Python interface is used to interpret the parameters specified in the subscriber service-specific VSAs and to generate an internal proprietary format VSA representing the subscriber service instance to activate or deactivate, as shown in Figure 138: RADIUS Python interface.

Figure 138: RADIUS Python interface



The SR OS only requires the Alc-Sub-Serv-Internal VSA [26-6527-155] to activate or deactivate the subscriber service. Subscriber service-specific VSAs [26-6527-151..154] are available to be used in the subscriber service Python script but are ignored in the SR OS.

See the Python script section for details on the subscriber services Python script functions and operation.

A Python script must be configured for RADIUS Access-Accept and CoA messages; for example:

```
config
    python
        python-script "subsvc-1" create
            primary-url "ftp://user:pwd@10.1.1.1/./py/subsvc-1.py"
            no shutdown
        exit
        python-policy "py-policy-subsvc-1" create
            radius access-accept direction ingress script "subsvc-1"
            radius change-of-authorization-request direction ingress script "subsvc-
1"
        exit
    exit
```

The Python policy must then be applied to the **radius-server-policy** to pass the Access-Accept messages to the Python script and to the RADIUS server to pass the CoA messages to the Python script. For example:

```
configure
    router
        radius-server
            server "server-1" address 10.1.1.2 secret <secret> create
                accept-coa
                python-policy "py-policy-subsvc-1"
            exit
        exit
    exit
    aaa
        radius-server-policy "aaa-server-policy-1" create
            python-policy "py-policy-subsvc-1"
            servers
                access-algorithm round-robin
                router "Base"
                server 1 name "server-1"
            exit
        exit
```

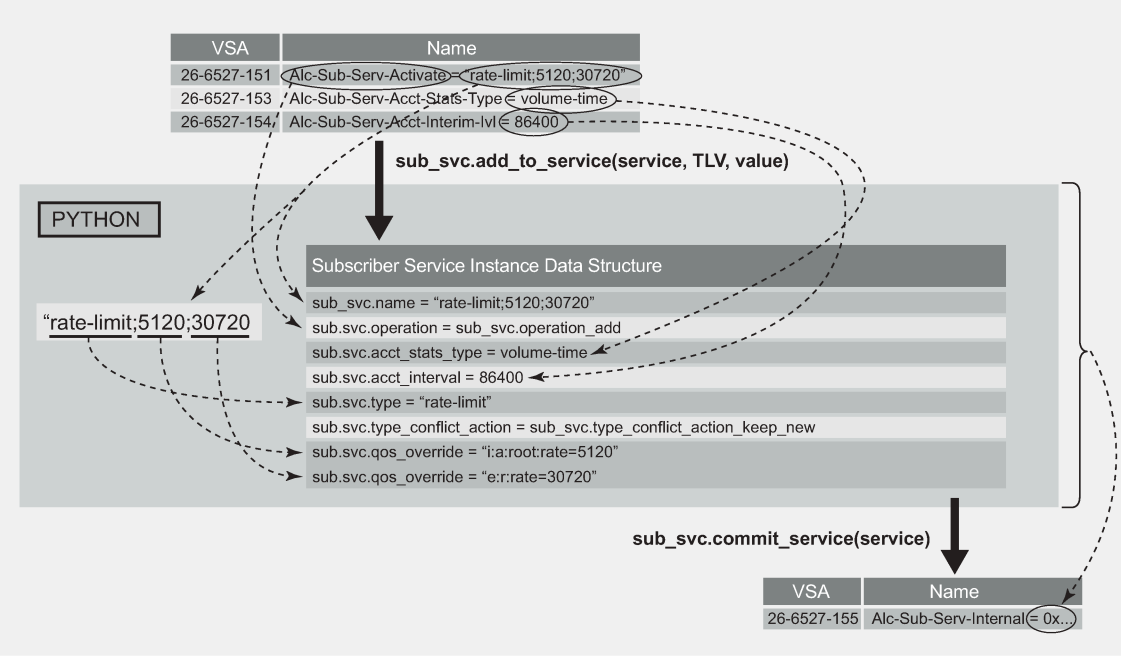
```
    exit
```

9.22.1.3 Python script

The RADIUS Access-Accept and CoA messages are passed to the configured RADIUS Python scripts. As shown in Figure 139: Subscriber service Python script operation, the function of the subscriber service Python script is to interpret the subscriber service-specific VSAs that contain the subscriber service instance parameters and to generate a new Alc-Sub-Serv-Internal VSA containing the information required to activate the actual subscriber service on the PPPoE or IPoE session.

This section covers the basics to understand the functionality of a subscriber service Python script. See the Subscriber services Python API section for a detailed description of the alc.sub_svc Python module containing functions and data structures used to define and activate a subscriber service instance.

Figure 139: Subscriber service Python script operation



The alc.sub_svc Python module contains the required functions and data structure to commit a subscriber service, including:

- TLVs to build a data structure describing the subscriber service functionality, such as QoS overrides or PCC rules
- functions to populate the data structure such as sub_svc.add_to_service(), sub_svc.pccrule.add_to_pccrule(), and sub_svc.flow.add_to_flow()
- a function to commit the subscriber service and create the internal VSA from the data structure: sub_svc.commit_service()

9.22.1.3.1 Python script example

In this section an example of a Python script is described that enables the activation or deactivation of a subscriber service.

In the example, it is assumed that only a single subscriber service is activated or deactivated per RADIUS message (no tagged VSAs are used) and that only a single Alc-Sub-Serv-Activate or Alc-Sub-Serv-Deactivate VSA is present (no concatenation of VSAs is required).

To change the upstream root arbiter rate and downstream aggregate rate bandwidth of an IPoE session, send the following parameters in the subscriber service activate VSA:

```
rate-limit;<upstream_bw_in_kbps>;<downstream_bw_in_kbps>
```

During the bandwidth change, the traffic should be accounted for in a separate accounting session.

To import the required modules in the Python script:

```
# Python - imports
import struct
from alc import radius
from alc import sub_svc
```

The alc.radius module provides the API access to the RADIUS VSAs in Access-Accept and CoA messages.

The alc.sub_svc module allows the API to activate and deactivate subscriber services.

The struct module is a Python module used in the example to convert data obtained from the RADIUS API as a string into Python integer values.

The following constants are used in the script:

```
# Python - constants
# VSA vendor ID
ALC = 6527
# ALC Radius VSA
SUB_SERVICE_ACTIVATE = 151
SUB_SERVICE_DEACTIVATE = 152
SUB_SERVICE_ACCT_STATS_TYPE = 153
SUB_SERVICE_ACCT_INTERIM_IVL = 154
```

The main flow in a subscriber service Python script is to first process the subscriber service deactivations, followed by the subscriber service activations. Optionally, the subscriber service-specific VSAs can be removed from the RADIUS message as they are not required for further processing in the SR OS:

```
# Python - main()
deactivate_services()
activate_services()
radius.attributes.clearVSA(ALC, SUB_SERVICE_ACTIVATE)
radius.attributes.clearVSA(ALC, SUB_SERVICE_DEACTIVATE)
radius.attributes.clearVSA(ALC, SUB_SERVICE_ACCT_STATS_TYPE)
radius.attributes.clearVSA(ALC, SUB_SERVICE_ACCT_INTERIM_IVL)
```

The function to deactivate a subscriber service executes the following steps:

- checks if an Alc-Sub-Serv-Deactivate VSA [26-6527-152] is present
- if present:

- builds the subscriber service data structure. For a subscriber service deactivation, only the subscriber service instance name is required.
- commits the subscriber service deactivation

```
# Python - deactivate_services()
def deactivate_services():
  value = radius.attributes.getVSA(ALC, SUB_SERVICE_DEACTIVATE)
  if value != '':
    service = []
    sub_svc.add_to_service(service, sub_svc.name, value);
    sub_svc.add_to_service(service, sub_svc.operation, sub_svc.operation_del);
    sub_svc.commit_service(service)
```

The function to activate a subscriber service executes the following steps:

- checks if an Alc-Sub-Serv-Activate VSA [26-6527-151] is present
- if present:
  - separates the parameters in the attribute value with a semicolon as the delimiter. The parameters are the subscriber service type, the upstream bandwidth in kb/s, and the downstream bandwidth in kb/s.
  - If the subscriber service type is rate-limit, builds the subscriber service data structure:
    - specifies that the service is added (activate)
    - adds the complete attribute value as the subscriber service instance name. This name should also be used to deactivate the subscriber service.
    - sets the subscriber service type and conflict action. With the conflict action set to "keep new", if a new subscriber service of the same type is activated, the old one is deactivated first.
    - configures the upstream and downstream bandwidth with a QoS override on the ingress root arbiter and egress aggregate rate
  - If the subscriber service type is something else (an unknown service type), a warning is printed for debugging purposes. No action is performed in this case.
  - If an Alc-Sub-Serv-Acct-Stats-Type VSA [26-6527-153] is present, adds the corresponding stats-type to the subscriber service data structure. The function also checks whether a Alc-Sub-Acct-Serv-Interim-Ivl VSA [26-6527-154] is present and adds the interval to the subscriber service data structure.

    Both the stats-type and interim interval must be specified as integers in the subscriber service data structure. As the RADIUS API returns an octet string, a conversion is required. The struct.unpack() function is used for this purpose.
  - commits the subscriber service activation

```
# Python - activate_services()
def activate_services():
  # Subscriber Service Activate VSA
  value = radius.attributes.getVSA(ALC, SUB_SERVICE_ACTIVATE)
  if value != '':
    values = value.split(';')
    if values[0] == "rate-limit":
      service = []
      sub_svc.add_to_service(service, sub_svc.operation, sub_svc.operation_add);
      sub_svc.add_to_service(service, sub_svc.name, value);
      sub_svc.add_to_service(service, sub_svc.type, 'rate-limit')
```

```
      sub_svc.add_to_service(service, sub_svc.type_conflict_action, sub_svc.type_con
flict_action_keep_new)
      sub_svc.add_to_service(service, sub_svc.qos_override, 'i:a:root:rate=' +
values[1])
      sub_svc.add_to_service(service, sub_svc.qos_override, 'e:r:rate=' + values[2])
    else:
      print "WARNING - Unknown service type :", values[0]
      return
    # Subscriber Service Accounting VSA
    stats_type = radius.attributes.getVSA(ALC, SUB_SERVICE_ACCT_STATS_TYPE)
    if stats_type != '':
      sub_svc.add_to_service(service, sub_svc.acct_stats_type, struct.unpack("!I", s
tats_type)[0]);
      interval = radius.attributes.getVSA(ALC, SUB_SERVICE_ACCT_INTERIM_IVL)
      if interval != '':
        sub_svc.add_to_service(service, sub_svc.acct_interval, struct.unpack("!I", i
nterval)[0]);
    # Activate the Subscriber Service
    sub_svc.commit_service(service)
```

The result of the sub_svc.commit_service() function is an Alc-Sub-Serv-Internal VSA [26-6527-155] that contains the required data for the SR OS to activate the corresponding subscriber services.

9.22.1.4 Subscriber service instance activation or deactivation with optional RADIUS accounting

When the SR OS receives an Alc-Sub-Serv-Internal VSA [26-6527-155] in an Access-Accept or CoA message as a result of the Python script sub_svc.commit_service() function, the corresponding subscriber services are activated or deactivated. Optionally, an accounting session can be started for each subscriber service instance.

The following is an example of the **show service sub-services** command.

```
# show service sub-services
===============================================================================
Subscriber service table
===============================================================================
Username        : cpe2@domain1.com
Subscriber      : cpe2@domain1.com                    Type     : PPP
SAP             : 1/1/4:1201.5                        Service  : 1000
MAC Address     : 00:51:00:00:01:01                   PPPoE-SID: 114
IP              : 10.1.1.3
Interface-ID    : 02:51:00:FF:FE:00:01:01
-------------------------------------------------------------------------------
  Service instance : rate-limit;5120;30720
  up time          : 0d 00:01:06
  type             : rate-limit
  acct sess id     : 144DFF00000A2556B757DA
  multi sess id    : 144DFF00000A1156B71E44
  acct type        : volume-time
  acct ivl         : 1d 00:00:00
  input octets     : 0
  input packets    : 0
  output octets    : 0
  output packets   : 0
  actions          :
  QoS-override
    ingress arbiter "root" rate 5120
  QoS-override
    egress aggregate rate limit 30720
```

```
-------------------------------------------------------------------------------
number of subscriber services found: 1
===============================================================================
```

The following rules apply for a subscriber service instance.

- A subscriber service instance can be activated on a:
  - single-stack or dual-stack PPPoE session
  - single-stack or dual-stack IPoE session
  - single-stack IPoEv4 host
- A subscriber service instance is uniquely identified by its name. To deactivate a subscriber service instance, the subscriber service name must be referenced.
- A subscriber service cannot be modified. To update a subscriber service instance, the old subscriber service must be deactivated and a new one activated.
- Multiple subscriber services can be activated or deactivated in a single Access-Accept or CoA message. Tagged subscriber service-specific VSAs are available for this purpose.
- A single subscriber service instance can have multiple actions, such as multiple QoS overrides or multiple PCC rules.
- The conflict action determines the behavior when multiple subscriber services of the same type are activated for the same PPPoE or IPoE session:
  - Conflict action = none

    Multiple instances of the same type can be activated.
  - Conflict action = keep new

    Only a single subscriber service instance of the same type is allowed.

    When a new subscriber service instance of the same type is activated on a single PPPoE or IPoE session, the old instance is deactivated, and the new subscriber service instance is activated.
  - Conflict action = keep old

    Only a single subscriber service instance of the same type is allowed.

    When a subscriber service instance of the same type is activated on a single PPPoE or IPoE session, the new subscriber service instance is rejected.
- Multi-Chassis Synchronization (MCS) is not supported for subscriber services.

9.22.2 Subscriber services RADIUS VSAs

Subscriber services VSAs 26-6527-151..154 are used as inputs for the subscriber service Python script and are ignored by the SR OS.

The subscriber service VSAs can be tagged to allow activation and deactivation of multiple subscriber service instances with a single RADIUS Access-Accept or CoA message.

Table 32: Subscriber services RADIUS VSAs lists the subscriber services RADIUS VSAs. See the RADIUS Attributes Reference Guide for a complete description of the subscriber services VSAs.

Table 32: Subscriber services RADIUS VSAs

Attribute ID	Attribute name	Description
26-6527-151	Alc-Sub-Serv-Activate (string)	Activate a subscriber service. The attribute contains parameters as input for the subscriber service Python script to define and activate a subscriber service. Multiple VSAs can be included per message if the parameter list does not fit in a single attribute.
26-6527-152	Alc-Sub-Serv-Deactivate (string)	Deactivate a subscriber service. The attribute contains parameters as input for the subscriber service Python script to deactivate a subscriber service. Multiple VSAs can be included per message if the parameter list does not fit in a single attribute.
26-6527-153	Alc-Sub-Serv-Acct-Stats-Type (integer)	Enable or disable subscriber service accounting and specify the statistics type: volume and time or time only. Values: 1=off, 2=volume-time and 3=time
26-6527-154	Alc-Sub-Serv-Acct-Interim-Ivl (integer)	The interim accounting interval in seconds at which Acct-Interim-Update messages should be generated for subscriber service accounting. A value of 0 (zero) corresponds to no interim update messages. The maximum value is 300 seconds (values 1 to 300).
26-6527-155	Alc-Sub-Serv-Internal	For internal use only. Its value is the result of the subscriber service commit function in Python. (sub_svc.commit_service).

9.22.3 Subscriber service RADIUS accounting

The configuration for subscriber service instance accounting sessions is obtained from the RADIUS accounting policies configured in the subscriber profile of the parent subscriber:

- The accounting destinations: RADIUS server configuration

  Subscriber service instance accounting is sent to multiple destinations when multiple RADIUS accounting policies are configured (for example, **config**>**subscr-mgmt**>**sub-prof# radius-accounting policy** *acct-policy-1 acct-policy-2*)

  Up to five accounting policies can be configured. Each policy is independent of the other, with its own accounting mode, update interval, and include attributes. Because resources are limited for sending out RADIUS account messages, contact your Nokia technical support representative for recommendations.

- The accounting attributes (**include-radius-attribute**) with some exceptions.

  Standard accounting attributes are always used for volume accounting, regardless of the configuration in the RADIUS accounting policy.

  An untagged [26-6527-151] Alc-Sub-Serv-Activate attribute is included in all subscriber service instance accounting messages. Its value is the subscriber service instance name (alc.sub_svc.name).

- Accounting interim updates for subscriber service instance accounting are controlled independently from the parent subscriber accounting session:
  - Enabling accounting interim updates for a subscriber service instance is achieved by setting the sub_svc.acct_interval TLV in the subscriber services Python script to a value greater than 0: sub_svc.add_to_service(service, sub_svc.acct_interval, 3600)
  - Disabling accounting interim updates for a subscriber service instance is achieved by setting the sub_svc.acct_interval TLV in the subscriber services Python script to 0: sub_svc.add_to_service(service, sub_svc.acct_interval, 0)

If the parent subscriber has no RADIUS accounting policy configured, subscriber service instance accounting cannot be enabled.

If the parent subscriber has a RADIUS accounting policy configured, subscriber service instance accounting can be disabled by setting the sub_svc.acct_stats_type TLV in the subscriber service Python script to 1 (Off):

```
sub_svc.add_to_service(service, sub_svc.acct_stats_type, 1)
```

Each subscriber service instance has a unique accounting session ID included as [44] Acct-Session-Id. The [50] Acct-Multi-Session-Id contains the accounting session ID of the parent PPPoE or IPoE session as shown in Table 33: Subscriber service accounting multisession ID.

Table 33: Subscriber service accounting multisession ID

Subscriber service parent	[50] Acct-Multi-Session-Id
PPPoE session	Session Acct-Session-Id of the PPPoE session: **show service id** <*service-id*> **pppoe session detail**
IPoE session	Session Acct-Session-Id of the IPoE session: **show service id** <*service-id*> **ipoe session detail**
IPoEv4 host	Host Acct-Session-Id of the IPoEv4 host: **show service id** <*service-id*> **subscriber-hosts detail**

The content of the volume counters when the subscriber service accounting statistics type equals **volume-time** is determined by the subscriber service action. For details, see the subscriber service sections that follow.

9.22.4 Accounting-only subscriber service

An accounting-only subscriber service has no specific action such as **qos-override** or **pccrule** defined and has subscriber service instance accounting enabled.

An example of when an accounting-only subscriber service would be used is if additional accounting data is needed for a specific time period in the lifetime of a PPPoE or IPoE session.

The sub_svc.acct_stats_type TLV in the subscriber services Python script must be set to a value different from 1 (Off) to enable subscriber service instance accounting. For example:

```
# Python example - Accounting only subscriber service
service = []
sub_svc.add_to_service(service, sub_svc.operation, sub_svc.operation_add)
sub_svc.add_to_service(service, sub_svc.name, 'subsvc-acct-1')
sub_svc.add_to_service(service, sub_svc.type, 'acct-only')
sub_svc.add_to_service(service, sub_svc.type_conflict_action, sub_svc.type_conflict_
action_none)
sub_svc.add_to_service(service, sub_svc.acct_stats_type, 2)
sub_svc.add_to_service(service, sub_svc.acct_interval, 3600)
sub_svc.commit_service(service)
```

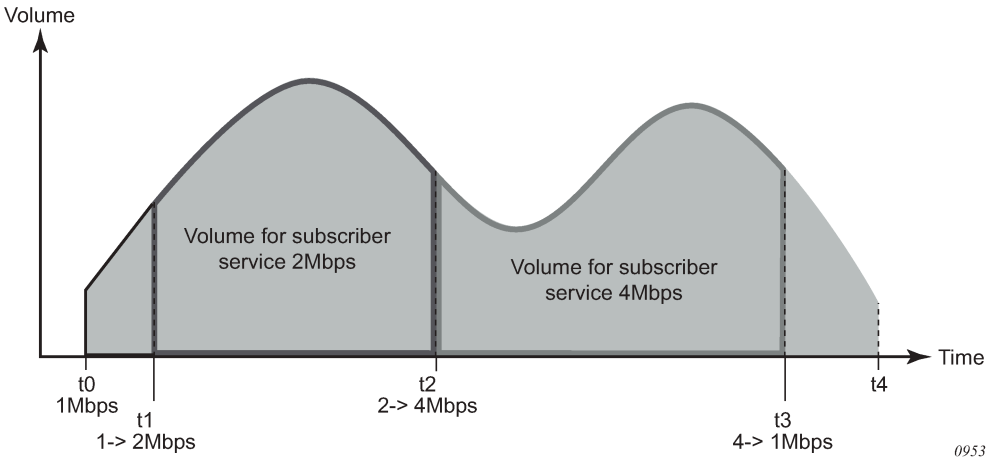
The volume counters for subscriber service statistics type **volume-time** contain the aggregate forwarded octets and packets of the parent PPPoE or IPoE session **sla-profile** instance from the start of the subscriber service.

9.22.5 QoS override-based subscriber service

A subscriber service instance with a **qos-override** action overrides queue or policer parameters (CIR, PIR, CBS, MBS) configured at the **sla-profile** level and hierarchical QoS parameters (aggregate rate, scheduler rate, or root and intermediate arbiter rate) configured at the **sub-profile level**.

An example of when a QoS override-based subscriber service would be used is to temporarily offer higher bandwidth and charge for the volume consumed during this period. Figure 140: Sample use case: QoS override-based subscriber services shows the volume statistics that are reported for the PPPoE or IPoE session accounting and for each of the subscriber service instances accounting sessions that were activated and deactivated.

Figure 140: Sample use case: QoS override-based subscriber services



- At t0, a subscriber connects and starts consuming bandwidth at the subscriber base egress rate of 1 Mb/s. An accounting session is started for the subscriber.

- At t1, a QoS override-based subscriber service instance is activated that boosts the download bandwidth to 2 Mb/s. A 2 Mb/s QoS override-based subscriber service instance accounting session is started.
- At t2, the 2 Mb/s QoS override-based subscriber service instance is deactivated and replaced with a new one that increases the download bandwidth to 4 Mb/s. The 2 Mb/s QoS override-based subscriber service accounting session is terminated, reporting the volume consumed during t1 to t2. A new 4 Mb/s QoS override-based subscriber service accounting session is started.
- At t3, the 4 Mb/s QoS override-based subscriber service instance is deactivated. The download bandwidth is reduced to the 1 Mb/s base rate again. The 4 Mb/s QoS override-based subscriber service accounting session is terminated, reporting the volume consumed during t2 to t3.
- At t4, the subscriber disconnects. The accounting session for the subscriber is terminated, reporting the volume consumed during t0 to t4.

The sub_svc.qos_override TLV in the subscriber services Python script adds a **qos-override** action. For example:

```
# Python example - QoS override subscriber service
service = []
sub_svc.add_to_service(service, sub_svc.operation, sub_svc.operation_add)
sub_svc.add_to_service(service, sub_svc.name, 'subsvc-qos-override-1')
sub_svc.add_to_service(service, sub_svc.type, 'qos-override')
sub_svc.add_to_service(service, sub_svc.type_conflict_action, sub_svc.type_conflict_
action_keep_new)
sub_svc.add_to_service(service, sub_svc.qos_override, 'e:q:1:pir=20000,cir=0')
sub_svc.add_to_service(service, sub_svc.qos_override, 'e:q:2:pir=5000,cir=5000')
sub_svc.add_to_service(service, sub_svc.qos_override, 'i:p:1:pir=1000,cir=0')
sub_svc.add_to_service(service, sub_svc.qos_override, 'e:r:rate=20000')
sub_svc.commit_service(service)
```

Actual values are used to populate the subscriber service data structure in this example; typically, these values are sent as parameters in subscriber service-specific VSAs.

Although a QoS override-based subscriber service instance is activated for a PPPoE or IPoE session, the overrides are applied at the SLA profile instance and subscriber level.

QoS override-based subscriber services have precedence over dynamic QoS overrides (RADIUS Alc-Subscriber-QoS-Override VSA or Gx Charging-Rule-Definition with QoS-Information) on the PPPoE or IPoE session.

- When a dynamic QoS override such as by the Alc-Subscriber-QoS-Override VSA is active on the parent PPPoE or IPoE session and a QoS override-based subscriber service is activated, the action from the QoS override-based subscriber service is installed. When the QoS override-based subscriber service is later deactivated, the original dynamic QoS override is restored.
- When a dynamic QoS override such as by the Alc-Subscriber-QoS-Override VSA is received for a parent PPPoE or IPoE session with an active QoS override-based subscriber service, the QoS override-based subscriber service remains installed. If the QoS override-based subscriber service is later deactivated, the previously received dynamic QoS override is installed.
- When both a QoS override-based subscriber service activation and a dynamic QoS override such as by the Alc-Subscriber-QoS-Override VSA is received in a single message for the parent PPPoE or IPoE session, the QoS override-based subscriber service action is installed and the dynamic QoS override is stored for later reference. When the QoS subscriber service is later deactivated, the previously received dynamic QoS override is installed.

The installed QoS override actions can be verified in the output of the **show service active-subscribers detail** CLI command.

The volume counters for subscriber service statistics type **volume-time** contain the aggregate forwarded octets and packets of the parent PPPoE or IPoE session **sla-profile** instance because of the start of the subscriber service.

QoS override-based subscriber services are stored in the subscriber-mgmt persistency file.

9.22.6 PCC rule-based subscriber services

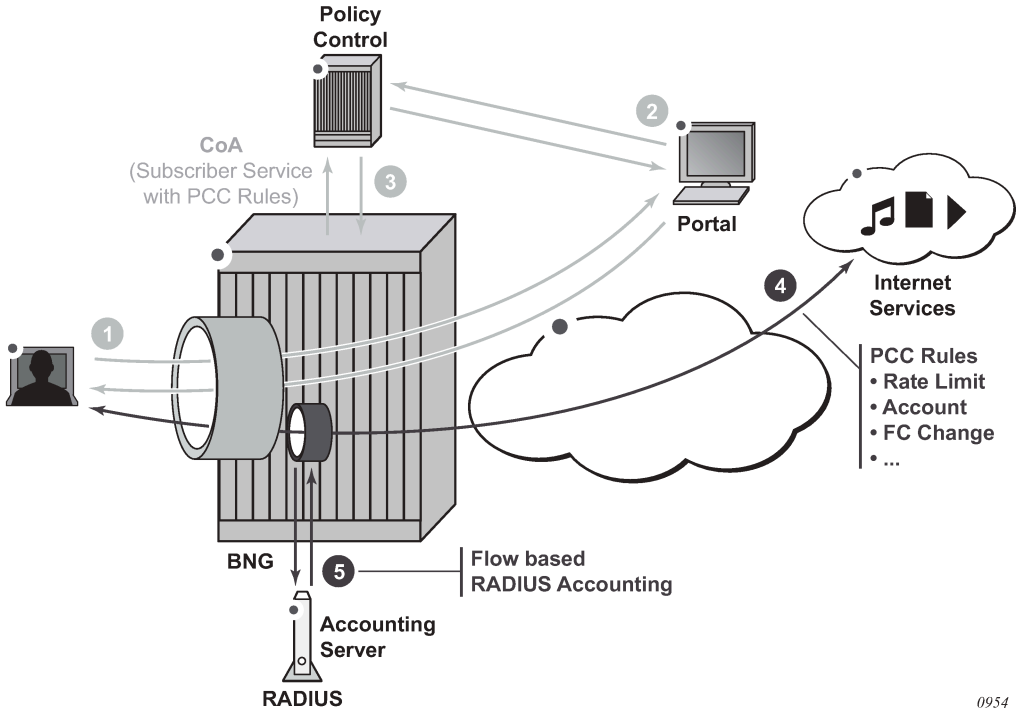
Policy and Charging Control (PCC) rules are defined in the 3GPP PCC architecture and used in the Diameter Gx application as a collection of parameters that enable IP traffic flows to be identified, QoS parameters and filtering actions to be applied to these flows, and charging to be performed on them. The use of PCC rules in policy management by the Diameter Gx interface is described in the PCC Rules.

The same PCC rule construct is used in RADIUS subscriber services to enable IP flow-based actions and accounting.

- Identify IP flows based on 5-tuple (protocol, source and destination IP address, source and destination port) and DSCP value.
- Apply QoS actions to these flows, such as rate limiting or change of forwarding class.
- Apply filter actions to these flows, such as forward, drop, HTTP redirect, redirect to a next hop or routing instance.
- Enable RADIUS accounting to report the forwarded octets and packets of the IP flows.

IP flow-based accounting can be used for a subscriber service using PCC rules. Activated by a self-service portal or as part of an Internet subscription package, applications identified by a 5-tuple receive specific treatment, such as bandwidth increase, expedited forwarding, or zero rating. Volume and time statistics for the application data is available in the subscriber service RADIUS accounting session. This is shown in Figure 141: Example: PCC rule-based subscriber service.

Figure 141: Example: PCC rule-based subscriber service



1. The user subscribes to a service by a web portal.
2. The policy control is informed of the new subscription.
3. The policy control instructs the BNG by RADIUS to activate a subscriber service with PCC rules.
4. The subscriber service is instantiated on the BNG. A dynamic policer is spawned to optionally rate-limit or count the application traffic.
5. The subscriber service instance accounting session reports the volume of traffic forwarded as part of this service.

9.22.6.1 PCC rule actions

A PCC rule is a unidirectional set of IP flows sharing a same set of actions. IPv4 and IPv6 flows can be combined within the same PCC rule.

A PCC rule name must be unique for each rule applied on a single PPPoE or IPoE session. For optimal PCC rule sharing, it is recommended that the same PCC rule name be used when its content is the same (that is, the same set of flows and same set of actions).

An IP flow is identified by a combination of:

- 5-tuple (IPv4 or IPv6): src-ip, src-port, dst-ip, dst-ip, protocol, next-header
- DSCP value

The CLI equivalent is:

```
match protocol | next-header <protocol>
       src-ip <ip-address>
```

```
        dst-ip <ip-address>
        src-port eq <port> | range <start-port> <end-port>
        dst-port eq <port> | range <start-port> <end-port>
        dscp <dscp>
    exit
```

Supported actions include forward, drop, redirect to ip next hop, redirect to a routing instance, HTTP redirect, forwarding class change, rate-limit, and account.

With a specified set of actions, PCC rules are instantiated in the SR OS by IP criteria or IPv6 criteria entries in SAP ingress or SAP egress QoS polices and in IP or IPv6 filter entries. A PCC rule precedence value determines the relative order of different PCC rules when inserted in the QoS or filter policy: a rule with a lower precedence value is be applied before a rule with a higher precedence value. Rules with the same precedence can be automatically optimized; the relative order in which they are applied is determined by the system for optimal sharing. Rules with no precedence are applied at the end and are also automatically optimized.

Table 34: Subscriber service PCC rule actions resulting in QoS policy changes and Table 35: Subscriber service PCC rule actions resulting in filter changes provide an overview of the PCC rule actions and where they apply.

Table 34: Subscriber service PCC rule actions resulting in QoS policy changes

Action	Direction	Description

Forwarding Class (FC) change	Ingress/Egress	changes the QoS forwarding class
		CLI equivalent:
		`config>qos`
		`  sap-ingress | sap-egress <id>`
		`create`
		`    ip-criteria | ipv6-criteria`
		`      entry <id> create`
		`        match`
		`          <5-tuple | dscp>`
		`        exit`
		`        action fc <fc>`
		`      exit`
		`    exit`

Rate-limit (PIR/CIR)	Ingress/Egress	rate-limit traffic matching the specified flows
		• spawn a dynamic policer per direction (if not already present)
		• set the PIR/CIR value of the dynamic policer
		• map the flows to the policer
		The forwarded octets and packets statistics of the dynamic policer are included in subscriber service accounting.
		CLI equivalent:
		`config>qos`
		`  sap-ingress | sap-egress <id>`
		`create`
		`    policer 1 # dynamic police`

Action	Direction	Description
		rate <pir> cir <cir> exit ip-criteria | ipv6-criteria entry <id> create match <5-tuple | dscp> exit action policer 1 exit exit
Account	Ingress / Egress	counts traffic matching the specified flows: • spawn a dynamic policer per direction (if not already present) • if no rate-limit action is specified in the PCC rule, set the PIR/CIR value of the dynamic policer to **max** • map the flows to the policer The forwarded octets and packets statistics of the dynamic policer are included in subscriber service accounting. CLI equivalent: config>qos sap-ingress | sap-egress <id> create policer 1 # dynamic policer rate max cir max exit ip-criteria | ipv6-criteria entry <id> create match <5-tuple | dscp> exit action policer 1 exit exit
Forward	Ingress/Egress	Creates an entry in the QoS policy to forward the traffic without explicit QoS action. Matching traffic does not match on the next entry (match and exit behavior). In case of overlapping flows, such as account all traffic except flow 1 and flow 2, the more specific flows must be explicitly forwarded. CLI equivalent: config>qos sap-ingress | sap-egress <id> create ip-criteria | ipv6-criteria entry <id> create

Action	Direction	Description
		match <5-tuple | dscp> exit action exit exit

Table 35: Subscriber service PCC rule actions resulting in filter changes

Action	Direction	Description
Forward/Drop	Ingress/Egress	Creates a filter entry to forward or drop the traffic CLI equivalent: config>filter ip-filter | ipv6-filter <id> create entry <id> create match <5-tuple | dscp> exit action forward | drop exit exit exit
Redirect to an IP next-hop	Ingress	redirect the traffic to the specified IP next-hop CLI equivalent: config>filter ip-filter | ipv6-filter <id> create entry <id> create match <5-tuple | dscp> exit action forward next-hop <ip-address> exit exit exit
Redirect to a routing instance	Ingress	redirect the traffic to the specified routing instance CLI equivalent: config>filter ip-filter | ipv6-filter <id> create entry <id> create match <5-tuple | dscp> exit action

Action	Direction	Description
		forward router <router-instance> exit exit exit
HTTP redirect	Ingress	HTTP redirection to the specified URL CLI equivalent: config>filter ip-filter | ipv6-filter <id> create entry <id> create match <5-tuple | dscp> exit action http-redirect <rdr-url-string> exit exit exit

Figure 142: Supported combinations of ingress PCC rule actions and Figure 143: Supported combinations of egress PCC rule actions show the actions that can be combined in a single ingress or egress PCC rule.

Figure 142: Supported combinations of ingress PCC rule actions

INGRESS supported combinations of PCC rule actions		QoS				Filter				Gate
		Forward	Rate limit	FC change	Account / UM	Forward	Drop	redirect next-hop or router	redirect URL	Flow Status (1)
QoS	Forward	✓	✓(3)	✓(3)	✓(3)	✓	–	✓	–	✓
	Rate limit	✓(3)	✓	✓	✓(4)	✓	–	✓	–	✓
	FC change	✓(3)	✓	✓	✓	✓	–	✓	–	✓
	Account / UM	✓(3)	✓(4)	✓	✓	✓	–	✓	–	✓
Filter	Forward	✓	✓	✓	✓	✓	–	✓(2)	✓(2)	✓
	Drop	–	–	–	–	–	✓	–	–	✓
	redirect next-hop or router	✓	✓	✓	✓	✓(2)	–	✓	–	✓
	redirect URL	–	–	–	–	✓(2)	–	–	✓	✓
Gate	Flow Status (1)	✓	✓	✓	✓	✓	✓	✓	✓	✓

0961

Figure 143: Supported combinations of egress PCC rule actions

EGRESS supported combinations of PCC rule actions		QoS				Filter		Gate
		Forward	Rate limit	FC change	Account / UM	Forward	Drop	Flow Status (1)
QoS	Forward	✓	✓(3)	✓(3)	✓(3)	✓	-	✓
	Rate limit	✓(3)	✓	✓	✓(4)	✓	-	✓
	FC change	✓(3)	✓	✓	✓	✓	-	✓
	Account / UM	✓(3)	✓(4)	✓	✓	✓	-	✓
Filter	Forward	✓	✓	✓	✓	✓	-	✓
	Drop	-	-	-	-	-	✓	✓
Gate	Flow Status (1)	✓	✓	✓	✓	✓	✓	✓

0962

Note:

Consider the following rules:

- The Flow Status can only be set by Gx.

  For RADIUS subscriber service-based PCC rules, the Flow Status is fixed to Enabled.
- The filter action forward is ignored (not installed) when combined with the following filter actions: redirect next-hop, redirect router, or http-redirect.
- The QoS action forward is ignored (not installed) when combined with the following QoS actions: rate-limit, FC change, account/Usage Monitoring (UM).
- When the QoS action rate-limit and QoS action account or Usage Monitoring (UM) are combined, only a single dynamic policer is installed, which is used for both rate-limiting and obtaining forward statistics for accounting or usage monitoring.

9.22.6.2 PCC rule instantiation

A PCC rule can result in one or more IPv4/IPv6 filter and QoS policy IPv4/IPv6 criteria entries. This is transparent to the operator.

- A PCC rule is split into IPv4 filter entries, IPv6 filter entries, SAP ingress QoS IP or IPv6 criteria, and SAP egress QoS IP or IPv6 criteria.

- Each entry is inserted into the corresponding policy at a reserved range for dynamic PCC rule inserts. Within the reserved range, the (optional) precedence value for the rule is considered for the relative order of different PCC rules.
- The QoS rate-limit and account actions spawn a dynamic policer from a reserved range in the QoS policy. A template configuration provides dynamic policer parameters such as hierarchical policer parent, burst sizes (MBS, CBS), statistics mode and packet byte offset. Each of the dynamic policer parameters configured in the template can be overridden per PCC rule in the subscriber service activation (see Table 45: PCC Rule TLVs in the alc.sub_svc.pccrule module). A maximum of one dynamic policer is instantiated per PCC rule. There is a maximum of 63 dynamic policers per direction and per SLA profile instance. The output queue for PCC rule traffic mapped in the dynamic policer is determined by a mechanism called forwarding class inheritance: the output queue is the same queue that would be used if a packet with the same forwarding class as the PCC rule packet was classified using the applied QoS policy. The resulting output queue can be a local subscriber queue (when the FC is mapped to a queue or when the FC is mapped to a policer at egress and the policer is mapped to a local queue), a shared queue (when the FC is mapped to a policer at ingress) or a queue-group queue (when the FC is mapped to a policer at egress).
- Optimal policy and rule sharing is achieved by QoS and filter policy cloning and internal PCC rule optimizations. The mechanisms are the same as for Gx-initiated PCC rules as described the Generic Policy Sharing and Rule Sharing and Gx Rule Ordering.

  PCC rule sharing can only happen when the content is the same: identical name, direction, precedence value, set of flows, and set of actions. PCC rules with the same content have the same PCC rule ID.

  Filter and QoS policy clones that result from PCC rule instantiation can be recognized by a filter ID or QoS policy ID in the format 1 to 65535:P1 to 4096; for example, filter ip 10:p3
- PCC rules can be inactive if the corresponding host type is not present. For example, a PCC rule-based subscriber service with an IPv6 filter action can be activated on an IPoE session while there is no IPv6 host instantiated on the session. When the IPv6 host is later created, the PCC rule is activated.

The following initial configuration is required before activating PCC rule-based subscriber services.

- To install PCC rule QoS actions, a non-default ingress and egress QoS policy with a **sub-insert-shared-pccrule** range configured must be associated with the IPoE or PPPoE session (the default QoS policy cannot be modified). If a rate-limit or account action is needed, a dynamic policer range must also be configured. Additional dynamic policer parameters are optional and can be overridden per PCC rule in the subscriber service activation (see Table 45: PCC Rule TLVs in the alc.sub_svc.pccrule module).

```
configure qos
    sap-ingress <policy-id> create
        sub-insert-shared-pccrule start-entry <entry-id> count <count>
        dynamic-policer
            range start-entry <policer-id> count <count>
            packet-byte-offset {add <add-bytes> | subtract <sub-bytes>}
            mbs <size> [bytes|kilobytes]
            cbs <size> [bytes|kilobytes]
            parent <arbiter-name> [weight <weight-level>] [level <level>]
            stat-mode <stat-mode>
        exit
    exit
```

- To install PCC rule filter actions, an IPv4 or IPv6 filter with **sub-insert-shared-pccrule** range configured must be associated with the IPoE or PPPoE session.

Although a PCC rule-based subscriber service is activated on a PPPoE or IPoE session, the actions are applied at the SLA profile instance and subscriber host level.

- PCC rule QoS actions result in QoS policy clones that are applied at the SLA profile instance level. Traffic from all subscriber hosts and sessions sharing the SLA profile instance is subject to the specified actions.
- PCC rule filter actions result in IPv4 or IPv6 filter clones that are applied at the subscriber host level. Only traffic from the subscriber host of the same type (IPv4 or IPv6) that belongs to the PPPoE or IPoE session is subject to the specified actions.

A PCC rule with flow match criteria that are not explicitly IPv4 or IPv6 results in both IPv4 and IPv6 match criteria being installed; for example, destination address = any.

Filter actions are executed before QoS actions. If an IP flow is rate-limited, it should pass the IPv4 or IPv6 filter first. Adding a QoS action rate limit to a PCC rule does not automatically insert a corresponding forward entry in an IP filter. When needed, this must be done explicitly by the operator with a filter forward action. For example, an IP filter with the default action drop and several explicit forward entries is applied to an IPoE session. A new IP flow must be rate-limited and accounted for. The PCC rule should include match criteria for the IP flow and a QoS action rate limit, QoS action account, and filter action forward. Without the filter action forward, the IP flow would be dropped by the default action in the filter policy.

```
config>filter
    ip-filter <filter-id> create
        sub-insert-shared-pccrule start-entry <entry-id> count <count>
    exit
    ipv6-filter <filter-id> create
        sub-insert-shared-pccrule start-entry <entry-id> count <count>
    exit
```

See Bulk Changes while Gx Rules are Active for information about the parameters that can be changed in the base filter and QoS policies when PCC rules are applied.

9.22.6.3 PCC rules in a subscriber service

The example below shows a pseudo-language representation of PCC rules in a subscriber service.

- A subscriber service can contain multiple PCC rules. Because a PCC rule is unidirectional, including an ingress and an egress PCC rule enables subscriber service accounting of bidirectional flows.
- A PCC rule can contain multiple flows. Flows in a PCC rule can be a mix of IPv4 and IPv6 flows.
- Per PCC rule dynamic policer parameters can optionally be specified. These parameters override the dynamic-policer configuration in the sap-ingress or sap-egress QoS policies.

```
subscriber-service {
    name = <subsvc name>
    operation = add | delete
    acct-stats-type = off | volume-time | time
    acct-interval = <value>
    type = <string>
    type-conflict-action = keep-old | keep-new | none
    pcc-rule {
        name = <name>
        direction = ingress | egress
        flow = <5-tuple> | <dscp>
        flow = <5-tuple> | <dscp>
           …
        flow = <5-tuple> | <dscp>
        action = <action>
        action = <action>
           …
```

```
        action = <action>
        precedence = <value>
        policer {
            parent-arbiter = <arbiter-name>
            parent-level = <level>
            parent-weight = <weight-level>
            mbs = <bytes | default>
            cbs = <bytes | default>
            stat-mode = <stat-mode>
            packet-byte-offset = <offset>
        }
    }
    …
    pcc-rule {
        name = <name>
        direction = ingress | egress
        flow = <5-tuple> | <dscp>
        flow = <5-tuple> | <dscp>
           …
        flow = <5-tuple> | <dscp>
        action = <action>
        action = <action>
           …
        action = <action>
        precedence = <value>
        policer {
            parent-arbiter = <arbiter-name>
            parent-level = <level>
            parent-weight = <weight-level>
            mbs = <bytes | default>
            cbs = <bytes | default>
            stat-mode = <stat-mode>
            packet-byte-offset = <offset>
        }
    }
}
```

The sub_svc.pccrule TLV in the subscriber services Python script adds a PCC rule to the subscriber service, as shown in the output example below:

Actual values are used to populate the subscriber service data structure in this example; typically, these values are sent as parameters in subscriber service-specific VSAs.

```
# Python example - PCC rules subscriber service
service = []
sub_svc.add_to_service(service, sub_svc.operation, sub_svc.operation_add)
sub_svc.add_to_service(service, sub_svc.name, 'subsvc-pccrule-1')
sub_svc.add_to_service(service, sub_svc.type, 'pccrule')
sub_svc.add_to_service(service, sub_svc.type_conflict_action, sub_svc.type_conflict_
action_none)
flow_i = []
sub_svc.flow.add_to_flow(flow_i, sub_svc.flow.dst_ip, '10.1.1.0/24')
rule_i = []
sub_svc.pccrule.add_to_pccrule(rule_i, sub_svc.pccrule.name, 'pcc-rule-1-i')
sub_svc.pccrule.add_to_pccrule(rule_i, sub_svc.pccrule.precedence, 10)
sub_svc.pccrule.add_to_pccrule(rule_i, sub_svc.pccrule.direction, sub_svc.pccrule.direction_
ingress)
sub_svc.pccrule.add_to_pccrule(rule_i, sub_svc.pccrule.flow, flow_i)
sub_svc.pccrule.add_to_pccrule(rule_i, sub_svc.pccrule.qos_action_rate_limit_pir, 10
00)
sub_svc.pccrule.add_to_pccrule(rule_i, sub_svc.pccrule.qos_action_account, True)
flow_e = []
sub_svc.flow.add_to_flow(flow_e, sub_svc.flow.src_ip, '10.1.1.0/24')
```

```
rule_e = []
sub_svc.pccrule.add_to_pccrule(rule_e, sub_svc.pccrule.name, 'pcc-rule-1-e')
sub_svc.pccrule.add_to_pccrule(rule_e, sub_svc.pccrule.precedence, 10)
sub_svc.pccrule.add_to_pccrule(rule_e, sub_svc.pccrule.direction, sub_svc.pccrule.direction_
egress)
sub_svc.pccrule.add_to_pccrule(rule_e, sub_svc.pccrule.flow, flow_e)
sub_svc.pccrule.add_to_pccrule(rule_e, sub_svc.pccrule.qos_action_rate_limit_pir, 5000)
sub_svc.pccrule.add_to_pccrule(rule_i, sub_svc.pccrule.qos_action_account, True)
sub_svc.add_to_service(service, sub_svc.pccrule, rule_i)
sub_svc.add_to_service(service, sub_svc.pccrule, rule_e)
sub_svc.add_to_service(service, sub_svc.acct_stats_type, 2)
sub_svc.add_to_service(service, sub_svc.acct_interval, 3600)
sub_svc.commit_service(service)
```

In the above subscriber service example, two PCC rules are installed, each with one flow:

- pcc-rule-1-i: ingress traffic to destination 10.1.1.0/24 is rate-limited to 1 Mb/s
- pcc-rule-1-e: egress traffic from 10.1.1.0/24 is rate-limited to 5 Mb/s

Subscriber service instance volume-time accounting is enabled. The volume counters include the forwarded octets and packets of the dynamic policers installed for the above rules and count the traffic matching the flows.

The counters reported in PCC rule-based subscriber services RADIUS accounting are determined by the PCC rule QoS account action.

- If at least one PCC rule in the subscriber service has the QoS account action enabled (pccrule.qos_action_account = True), then the volume counters contain the sum of the dynamic policer forwarded octets and packets statistics of all the PCC rules in the subscriber service with pccrule.qos_action_account = True.
- If all PCC rules in the subscriber service have the QoS account action disabled (pccrule.qos_action_account = False), then the volume counters contain the aggregated forwarded octets and packets of the parent PPPoE or IPoE session SLA profile instance because the start of the subscriber service.

PCC rule-based subscriber services are not stored in the subscriber-mgmt persistency file.

9.22.6.4 Interaction of the PPPoE or IPoE session QoS model and PCC rule-based subscriber services

PCC rule-based subscriber services with QoS actions interact with the classification and QoS forwarding mechanisms. This section describes how this affects the parent RADIUS accounting volume counters.

For subscriber service PCC rule QoS actions that do not result in the instantiation of a dynamic policer (such as a change of forwarding class or forward), the PCC rule matched traffic is included in the parent accounting session volume counters. This is shown in Figure 144: PCC rule-based subscriber service—QoS interaction: no dynamic policer, FC to queue, where the forwarding class is mapped to a subscriber queue, and in Figure 145: PCC rule-based subscriber service—QoS interaction: no dynamic policer, FC to policer to queue-group where the forwarding class is mapped to a subscriber policer.

Figure 144: PCC rule-based subscriber service—QoS interaction: no dynamic policer, FC to queue

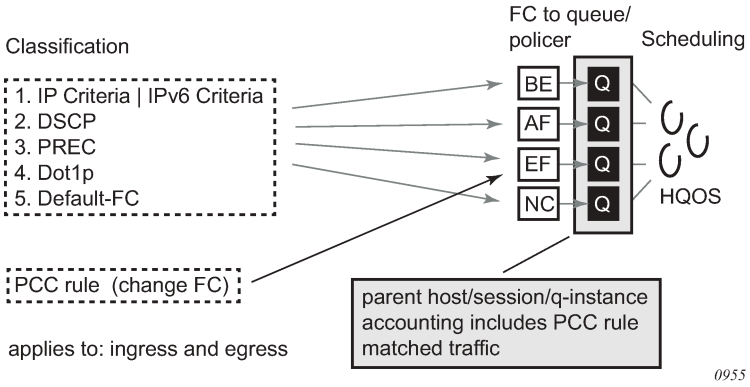
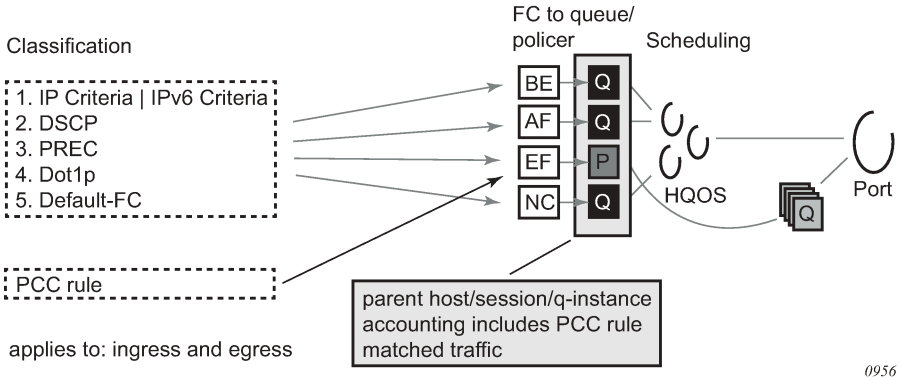


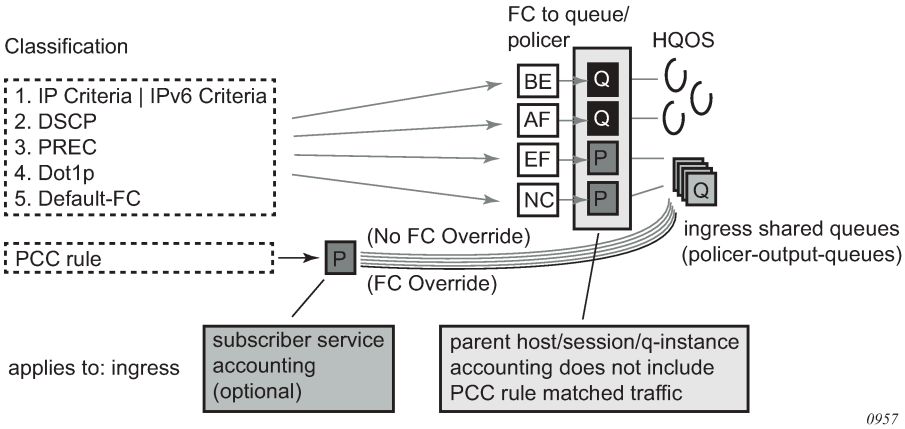
Figure 145: PCC rule-based subscriber service—QoS interaction: no dynamic policer, FC to policer to queue-group



For subscriber service PCC rule QoS actions that result in the instantiation of a dynamic policer (such as rate-limit or account), the dynamic policer counters are not included in the aggregate counters nor are they reported as separate detailed policer statistics. Instead, the traffic matching the PCC rules is counted in the output queues that correspond to the forwarding class of the packets.

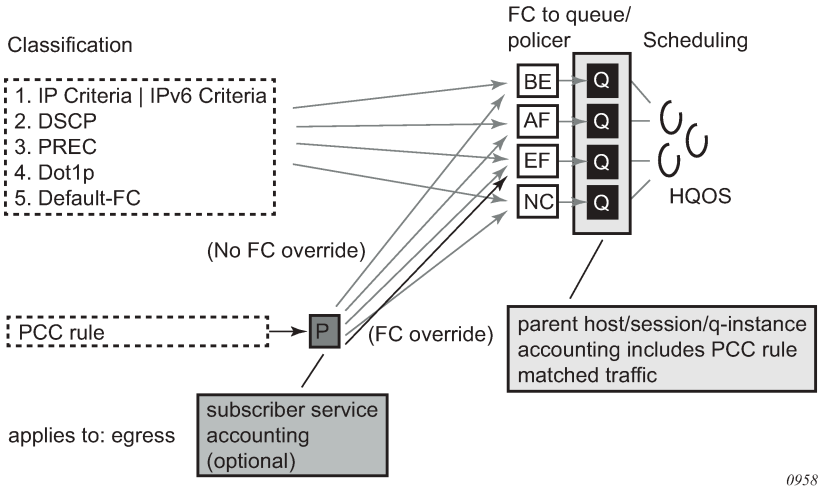
- On ingress, the dynamic policer PCC rule traffic is never included in the parent host, session, or queue instance accounting session counters. Ingress policed traffic always uses the ingress shared policer output queues, as shown in Figure 146: PCC rule-based subscriber service—QoS interaction: dynamic policer to ingress shared policer output queues.

Figure 146: PCC rule-based subscriber service—QoS interaction: dynamic policer to ingress shared policer output queues



- To include the egress dynamic policer PCC rule traffic in the parent host, session, or queue instance accounting session counters, the dynamic policer must use a local subscriber output queue, as shown in Figure 147: PCC rule-based subscriber service—QoS interaction: dynamic policer, FC to queue (egress).

Figure 147: PCC rule-based subscriber service—QoS interaction: dynamic policer, FC to queue (egress)



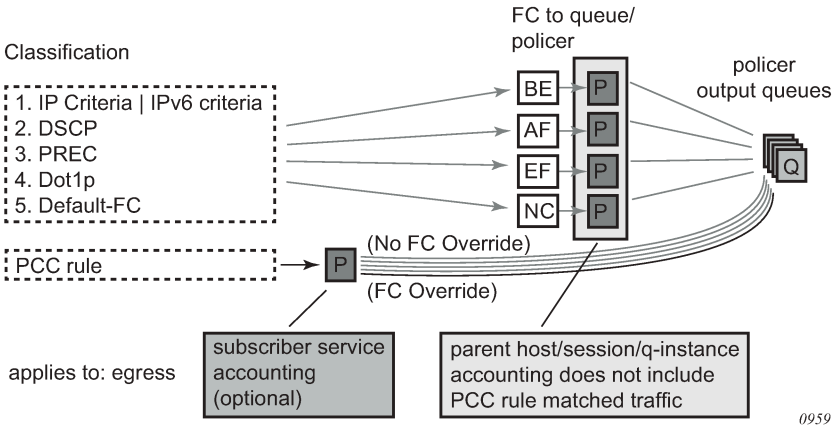
- To exclude the egress dynamic policer PCC rule traffic from the parent host, session, or queue instance accounting session counters, the dynamic policer must use a queue-group output queue, as shown in Figure 148: PCC rule-based subscriber service—QoS interaction: dynamic policer, FC to policer to policer output queues (egress). The dynamic policer traffic inherits the policer-to-output queue mapping from the static policer that corresponds to the forwarding class of the packet. The following SAP egress configuration example uses the default policer output queue-group:

```
config>qos
    sap-egress 10 create
       sub-insert-shared-pccrule start-entry 200 count 10
```

```
        dynamic-policer
            range start-entry 10 count 10
        exit
        policer 1 create
        exit
        fc be create
            policer 1
        exit
    exit
```

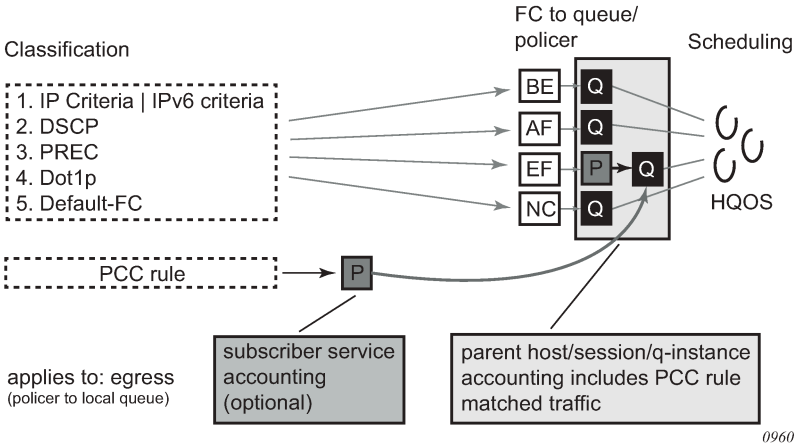
If a packet is classified as FC = Best Effort (BE) and matches a PCC rule with rate-limit action only (no FC change), the traffic hits the PCC rule dynamic policer and then the queue-group queue associated with policer 1 (the static policer for FC = BE).

Figure 148: PCC rule-based subscriber service—QoS interaction: dynamic policer, FC to policer to policer output queues (egress)



- A special case occurs when, at egress, the forwarding class maps to a static policer and then to a local subscriber queue. Traffic for that forwarding class hitting a dynamic policer uses the local subscriber queue as the output queue. In this case, the dynamic policer PCC rule traffic is included in the parent host, sessions or queue instance accounting session counters. This is shown in Figure 149: PCC rule-based subscriber service—QoS interaction: dynamic policer, FC to policer to local queue (egress).

Figure 149: PCC rule-based subscriber service—QoS interaction: dynamic policer, FC to policer to local queue (egress)



9.22.6.5 PCC rule-based subscriber service activation failures

PCC rule-based subscriber service activation failures can be categorized as failures detected in the subscriber services Python script as runtime errors (see Table 36: PCC rule-based subscriber service: Python runtime errors) and failures detected in the Enhanced Subscriber Management (ESM) processing (see Table 37: PCC rule-based subscriber service: ESM — RADIUS decoding failures and Table 38: PCC rule-based subscriber service: ESM — processing failures).

Table 36: PCC rule-based subscriber service: Python runtime errors

Python runtime errors
Pcc rule name type must be a string
Pcc rule name value too long
Pcc rule qos action account value must be True(1) or False(0)
Pcc rule filter action redirect to nexthop_v4/v6 type must be a string
Pcc rule filter action redirect to nexthop_v4/v6 value must be a valid IPv4/v6 address
Pcc rule filter action http redirect value is not a valid URL
Pcc rule filter action http redirect type must be a string
Pcc rule filter action http redirect value too long (> 255)
Pcc rule qos action fwd class change value is invalid
Pcc rule qos action rate limit type must be an int
Pcc rule precedence type must be an int

Python runtime errors
Pcc rule flow dscp match type must be a string
Pcc rule flow dscp match value is invalid
Pcc rule flow src/dst_ip match type must be a string: <ipv4-address>|<ipv6-address>|any
Pcc rule flow src/dst_ip match value is not a valid IP address: <ipv4-address>|<ipv6-address>|any
Pcc rule flow port match type must be a string: <port>[-<port>] with port [0..65535]
Pcc rule flow protocol match type must be an int
Pcc rule flow protocol match value must be less than 255
Pcc rule must have a name

Table 37: PCC rule-based subscriber service: ESM — RADIUS decoding failures

ESM — RADIUS decoding failures
The PCC rule precedence must be in the range 0 to 65535.
The PCC rule name has a maximum of 100 displayable characters.
The PCC rule redirect URL has a maximum of 255 displayable characters.
All flows in a PCC rule must have the same direction (ingress or egress).
A PCC rule has a maximum of 128 flows.
A PCC rule name can occur only one time in a RADIUS message.
A flow in a PCC rule cannot have a mix of IPv4 and IPv6 addresses (for example **src-ip** and **dst-ip**).

Table 38: PCC rule-based subscriber service: ESM — processing failures

ESM — processing failures
If a PCC rule contains a direction-specific action (such as a redirect), it must contain at least one flow in that direction.
If a PCC rule contains only IPv4 actions (such as a redirect to an IPv4 next hop), it must contain at least one IPv4 flow. This also applies to IPv6.
The combinations of PCC rule actions must be supported (see Figure 142: Supported combinations of ingress PCC rule actions and Figure 143: Supported combinations of egress PCC rule actions).
There must be at least one flow and at least one action per PCC rule.

ESM — processing failures
There is a maximum of 64 PCC rules per host or session.
There are not enough filter or QoS resources to create policy clones or apply them to the host or session.
The filter or QoS policy clone cannot be created (for example, the redirect service does not exist).
Simultaneous provisioning of PCC rules from Gx and RADIUS is operationally blocked per subscriber session/host.

9.22.7 Combined subscriber services

A subscriber service instance can be built with a combination of QoS override actions and PCC rules. In this case, the reported volume counters for subscriber service statistics type volume-time are determined by the PCC rule account action, as shown in Table 39: Subscriber service accounting—reported volume counters .

Table 39: Subscriber service accounting—reported volume counters

QoS override	PCC rule	Reported volume counters
No	No	Aggregated SLA profile queue and policer statistics
Yes	No	Aggregated SLA profile queue and policer statistics
No	Yes, "pccrule.qos_action_account = False" for all PCC rules	Aggregated SLA profile queue and policer statistics
No	Yes, "pccrule.qos_action_account = True" for at least one PCC rule	Dynamic policer statistics for PCC rules with "pccrule.qos_action_account = True"
Yes	Yes, "pccrule.qos_action_account = False" for all PCC rules	Aggregated SLA profile queue and policer statistics
Yes	Yes, "pccrule.qos_action_account = True" for at least one PCC rule	Dynamic policer statistics for PCC rules with "pccrule.qos_action_account = True"

9.22.8 Subscriber services Python API

The SR OS alc.sub_svc Python module offers functions and data structures to describe, activate, and deactivate a subscriber service.

9.22.8.1 Common subscriber services Python API

Table 40: Subscriber service functions in the alc.sub_svc module lists the subscriber service functions in the alc.sub_svc module.

Table 40: Subscriber service functions in the alc.sub_svc module

sub_svc functions	Description
sub_svc.add_to_service (svc, sub_svc TLV, value)	Appends a TLV to the service list. The service list describes the subscriber service and should be passed to the sub_svc.commit_service() to activate or deactivate the subscriber service. Parameters: svc (type = list): service list that describes the subscriber service. sub_svc TLVs are appended to this list with the sub_svc.add_to_service() function. sub_svc TLV (type int): TLV that is appended to the service list value (type as defined for the sub_svc TLV): the value of the sub_svc TLV that is appended to the service list
sub_svc.commit_service (svc)	Creates the required internal VSAs based on the TLVs provided in the svc list. Parameters: svc (type = list): service list that describes the subscriber service. sub_svc TLVs are appended to this list with the sub_svc.add_to_service() function. The service list should be passed to sub_svc.commit_service() to activate or deactivate the subscriber service.

Table 41: Subscriber service TLVs in the alc.sub_svc module lists the subscriber service TLVs in the alc.sub_svc module.

Table 41: Subscriber service TLVs in the alc.sub_svc module

sub_svc TLV	Activate/ deactivate	TLV details	
name (string)	M/M	Purpose	The unique subscriber service instance identifier (key). This field is matched for a subscriber service deactivate request. This field could, for example, be populated with the service-name and corresponding parameter list. This value is also echoed in the Alc-Sub-Serv-Session attribute in accounting messages. When not specified, the subscriber service activation fails and an event log is generated: WARNING: SVCMGR #2511 Base RADIUS CoA Error "Problem encountered in Subscriber Management, while processing a CoA request

sub_svc TLV	Activate/ deactivate	TLV details	
			from a RADIUS server: Could not decode RADIUS Attribute "Sub Service"".
		Value	Free format string (max. length = 1000 bytes)
		Default	Empty string
operation (int)	M/O	Purpose	Specifies if the referenced subscriber service should be activated or deactivated
		Value	operation_add (1): activate the subscriber service instance operation_del (2): deactivate the subscriber service instance
		Default	operation_del
acct_stats_type (int)	O/n.a.	Purpose	Defines if RADIUS accounting should be enabled for this subscriber service instance. If enabled, the accounting mode (Time or Volume and Time) is specified. Values as defined for Alc-Sub-Serv-Acct-Interim-Ivl VSA
		Value	Off (1) Volume-time (2) Time (3)
		Default	Off
acct_interval (int)	O/n.a.	Purpose	Defines the RADIUS interim accounting update interval for this subscriber service instance Values as defined for Alc-Sub-Serv-Acct-Interim-Ivl VSA
		Value	0 (no interim updates) 1 to 299 (rounded to a maximum 300 seconds) 300 to 15552000 (override the local configured update-interval for this subscriber service instance)
		Default	Update interval from the parent subscriber RADIUS accounting policy
type (string)	O/n.a.	Purpose	Grouping of subscriber service instances that belong to the same PPPoE or IPoE session
		Value	Free format string (max. length = 255 bytes)

sub_svc TLV	Activate/ deactivate	TLV details	
		Default	empty string
type_conflict_action (int)	O/n.a.	Purpose	Defines the action when another subscriber service instance of the same type is already activated for the same PPPoE or IPoE session.
		Value	type_conflict_action_keep_old (1): reject the new subscriber service instance type_conflict_action_keep_new (2): deactivate the old and activate the new subscriber service instance type_conflict_action_none (3): allow multiple subscriber service instances of this type
		Default	type_conflict_action_none

Note: M=Mandatory, O=Optional, n.a.=ignored

9.22.8.2 Subscriber service QoS override Python API

Table 42: QoS override TLVs in the alc.sub_svc module lists the QoS override TLVs in the alc.sub_svc module.

Table 42: QoS override TLVs in the alc.sub_svc module

sub_svc TLV	Activate/ deactivate	TLV details	
qos_override (string)	O/n.a.	Purpose	Adds a QoS override to the subscriber service. Multiple qos_override TLVs can be added in a single subscriber service instance.
		Value	As defined for the RADIUS Alc-Subscriber-QoS-Override VSA [26-6527-126]. See the *7450 ESS, 7750 SR, and VSR RADIUS Attributes Reference Guide* for details.
		Default	Not included

Note: M=Mandatory, O=Optional, n.a.=ignored

9.22.8.3 Subscriber service PCC rules Python API

Table 43: PCC Rule TLVs in the alc.sub_svc module lists the PCC rule PLVs in the alc.sub_svc module.

Table 43: PCC Rule TLVs in the alc.sub_svc module

sub_svc TLV	Activate/ deactivate	TLV details	
pccrule (list)	O/n.a.	Purpose	Adds a PCC rule to the subscriber service. Multiple PCC rule TLVs can be added in a single subscriber service instance.
		Value	A PCC rule list describing the PCC rule with PCC rule TLVs such as name, precedence, direction, flows, and actions PCC rule TLVs are appended to the PCC rule list with the sub_svc.pccrule.add_to_pccrule() function.
		Default	Not included

Note: M=Mandatory, O=Optional, n.a.=ignored

Table 44: alc.sub_svc.pccrule Function lists the alc.sub_svc.pccrule function.

Table 44: alc.sub_svc.pccrule Function

alc.sub_svc.pccrule function	Description
sub_svc.pccrule.add_to_pccrule (pccrule, pccrule TLV, value)	Appends a PCC rule TLV such as name, precedence, flow, or action to the PCC rule list. The PCC rule list describes the PCC rule and can be added to a subscriber service with the sub_svc.add_to_service() function. Parameters: pccrule (type = list): PCC rule list that describes the PCC rule. PCC rule TLVs are appended to this list with the sub_svc.pccrule.add_to_pccrule() function. pccrule TLV (type int): PCC rule TLV that is appended to the PCC rule list value (type as defined for the PCC rule TLV): the value of the PCC rule TLV that is appended to the PCC rule list

Table 45: PCC Rule TLVs in the alc.sub_svc.pccrule module lists the PCC rules TLVs in the alc.sub_svc.pccrule module.

Table 45: PCC Rule TLVs in the alc.sub_svc.pccrule module

PCC rule TLV	M|O	TLV details	
pccrule.name (String)	M	Purpose	Specifies the name of the PCC rule

PCC rule TLV	M|O	TLV details	
			A PCC rule with the same name and same or different content can only be applied one time on a single parent PPPoE or IPoE session. A PCC rule with the same name and same or different content can be applied on different parent PPPoE or IPoE sessions. Rules with the same name but different content gets a different PCC rule identifier (rule id).
pccrule.precedence (Integer)	O	Purpose	Specifies the precedence value for the PCC rule. The precedence defines a relative order of the different PCC rules: a rule with a lower precedence value is applied before a rule with a higher precedence value. Rules with the same precedence and rules without precedence can be automatically optimized; the relative order in which they are applied is determined by the system for optimal sharing.
		Value	0 to 65535
		Default	n/a These rules are applied at the end.
pccrule.direction (Integer)	M	Purpose	Specifies the direction of the PCC rule: ingress or egress
		Value	direction_ingress (1)
			direction_egress (2)
		Default	n/a
pccrule.flow (list)	M	Purpose	Adds a flow to the PCC rule. At least one flow must be added to a PCC rule. Multiple flow TLVs can be added to a PCC rule.
		Value	A flow list describing the flow with flow TLVs such as dscp, protocol, src-ip, dst-ip, src-port, and dts-port Flow TLVs are appended to the flow list with the sub_svc.flow.add_to_flow() function.
		Default	Not included
pccrule.qos_action_account (Boolean)	O (1)	Purpose	PCC rule action: account Can be applied on ingress and egress Results in IPv4 or IPv6 criteria entry in QoS policies. When set to True: if no rate-limit action is specified, a dynamic policer with pir=cir=max is instantiated for all flows in the PCC rule CLI equivalent: `policer 1 # dynamic policer`

PCC rule TLV	M|O	TLV details	
			rate max cir max exit entry 10 create match … exit action policer 1 exit The forwarded octets and packets statistics of the dynamic policer associated with this PCC rule are included in subscriber service accounting.
		Value	True (1) False (0)
		Default	False
pccrule.qos_action_ change_fc (string)	O (1)	Purpose	PCC rule action: change the forwarding class Can be applied on ingress and egress. Results in IPv4 or IPv6 criteria entry in QoS policies CLI equivalent: entry 10 create match … exit action fc <fc-name> exit
		Value	String with fixed format forwarding class name: "be", "l2", "af", "l1", "h2", "ef", "h1" or "nc"
		Default	n/a
pccrule.qos_action_ rate_limit_cir (Integer)	O (1)	Purpose	PCC rule action: rate-limit CIR • instantiate a dynamic policer for all flows in the PCC rule (if not already present) • set the CIR value Can be applied on ingress and egress. Results in IPv4 or IPv6 criteria entry in QoS policies CLI equivalent: policer 1 # dynamic policer rate … cir <cir> exit entry 10 create match … exit action policer 1

PCC rule TLV	M|O	TLV details	
			exit
		Value	0 to 2000000000 kb/s
		Default	n/a

pccrule.qos_action_ rate_limit_pir (integer)	O (1)	Purpose	PCC rule action: rate-limit PIR
			• Instantiate a dynamic policer for all flows in the PCC rule (if not already present)
			• Set the PIR value
			Can be applied on ingress and egress.
			Results in IPv4 or IPv6 criteria entry in QoS policies.
			CLI equivalent:
			`policer 1 # dynamic policer`
			`  rate <pir>`
			`exit`
			`entry 10 create`
			`  match`
			`  …`
			`  exit`
			`  action policer 1`
			`exit`

		Value	1 to 2000000000 kb/s
		Default	None

pccrule.qos_action (integer)	O (1)	Purpose	PCC rule action: QoS forward
			Can be applied on ingress and egress
			Results in IPv4 or IPv6 criteria entry in QoS policies
			CLI equivalent:
			`entry 10 create`
			`  match`
			`  …`
			`  exit`
			`  action`
			`exit`

		Value	pccrule.qos_action_forward (1)
		Default	n/a
pccrule.filter_action_ http_redirect (string)	O (1)	Purpose	PCC rule action: http-redirect Can be applied on ingress only Results in an IPv4 or IPv6 filter entry

PCC rule TLV	M|O	TLV details	

			CLI equivalent:
			`entry 10 create`
			`match next-header tcp`
			`...`
			`exit`
			`action`
			`http-redirect <rdr-url-string>`
			`exit`
			`exit`

		Value	http-redirect URL string (maximum 255 characters)
		Default	n/a

pccrule.filter_action_ redirect_to_nexthop_v4 (string)	O (1)	Purpose	PCC rule action: redirect to a next-hop IPv4 address
			Can be applied on ingress only
			Results in an IPv4 filter entry
			CLI equivalent:
			`entry 10 create`
			`match`
			`...`
			`exit`
			`action`
			`forward next-hop <ip-address>`
			`exit`
			`exit`

		Value	IPv4 address
		Default	n/a

pccrule.filter_action_ redirect_to_nexthop_v6 (string)	O (1)	Purpose	PCC rule action: redirect to a next-hop IPv6 address
			Can be applied on ingress only
			Results in an IPv6 filter entry. CLI equivalent:
			`entry 10 create`
			`match`
			`...`
			`exit`
			`action`
			`forward next-hop <ipv6-address>`
			`exit`
			`exit`

		Value	IPv6 address
		Default	None
pccrule.filter_action_ redirect_to_router_v4	O (1)	Purpose	PCC rule action: redirect to a routing instance Can be applied on ingress only

PCC rule TLV	M|O	TLV details	

(integer)			Results in an IPv4 filter entry
			CLI equivalent:
			`entry 10 create`
			`match`
			`...`
			`exit`
			`action`
			`forward router <router-instance>`
			`exit`
			`exit`

		Value	service-id
		Default	n/a

pccrule.filter_action_ redirect_to_router_v6
(Integer)
	O (1)	Purpose	PCC rule action: redirect to a routing instance
			Can be applied on ingress only
			Results in an IPv6 filter entry
			CLI equivalent:
			`entry 10 create`
			`match`
			`...`
			`exit`
			`action`
			`forward router <router-instance>`
			`exit`
			`exit`

		Value	*service-id*
		Default	n/a

pccrule.filter_action
(Integer)
	O (1)	Purpose	PCC rule action: Filter forward or drop
			Can be applied on ingress and egress
			Results in an IPv4 or IPv6 filter entry
			CLI equivalent:
			`entry 10 create`
			`match`
			`...`
			`exit`
			`action`
			`forward | drop`
			`exit`
			`exit`

		Value	pccrule.filter_action_forward (1) pccrule.filter_action_drop (2)

PCC rule TLV	M|O	TLV details	
		Default	n/a
pccrule.policer_parent_arbiter (String)	O	Purpose	Specifies the dynamic policer parent arbiter name for this PCC rule. The reserved value "_tmnx_no_parent" sets no arbiter parent for the dynamic policer used in this PCC rule. Overrides the dynamic policer value configured in the sap-ingress or sap-egress QoS policy.
		Value	Free format string (maximum length = 32 bytes) "_tmnx_no_parent" sets no parent arbiter
		Default	None
pccrule.policer_parent_level (Integer)	O	Purpose	Specifies the dynamic policer parent level for this PCC rule. Overrides the dynamic policer value configured in the sap-ingress or sap-egress QoS policy.
		Value	1 to 8
		Default	None
pccrule.policer_parent_weight (Integer)	O	Purpose	Specifies the dynamic policer parent weight for this PCC rule. Overrides the dynamic policer value configured in the sap-ingress or sap-egress QoS policy.
		Value	1 to 100
		Default	None
pccrule.policer_mbs (Integer)	O	Purpose	Specifies the dynamic policer MBS value in bytes or reset to the default MBS value for this PCC rule. Overrides the dynamic policer value configured in the sap-ingress or sap-egress QoS policy.
		Value	0 to 16777216 -1 sets the default MBS
		Default	None
pccrule.policer_cbs (Integer)	O	Purpose	Specifies the dynamic policer CBS value in bytes or resets to the default CBS value for this PCC rule. Overrides the dynamic policer value configured in the sap-ingress or sap-egress QoS policy.
		Value	0 to16777216

PCC rule TLV	M|O	TLV details	
			-1 sets the default CBS
		Default	None
pccrule.policer_stat_mode (Integer)	O	Purpose	Specifies the dynamic policer stat-mode for this PCC rule. Overrides the dynamic policer value configured in the sap-ingress or sap-egress QoS policy.
		Value	Note that integer values are mapped to each of the stats-mode. ingress: • 0 = pccrule.ingress_stat_mode_no_stats • 1 = pccrule.ingress_stat_mode_minimal • 2 = pccrule.ingress_stat_mode_offered_profile_no_cir • 3 = pccrule.ingress_stat_mode_offered_total_cir • 4 = pccrule.ingress_stat_mode_offered_priority_no_cir • 5 = pccrule.ingress_stat_mode_offered_profile_cir • 6 = pccrule.ingress_stat_mode_offered_priority_cir • 7 = pccrule.ingress_stat_mode_offered_limited_profile_cir • 8 = pccrule.ingress_stat_mode_offered_profile_capped_cir • 9 = pccrule.ingress_stat_mode_offered_limited_capped_cir egress: • 0 = pccrule.egress_stat_mode_no_stats • 1 = pccrule.egress_stat_mode_minimal • 2 = pccrule.egress_stat_mode_offered_profile_no_cir • 3 = pccrule.egress_stat_mode_offered_total_cir • 4 = pccrule.egress_stat_mode_offered_profile_cir • 5 = pccrule.egress_stat_mode_offered_limited_capped_cir • 6 = pccrule.egress_stat_mode_offered_profile_capped_cir • 8 = pccrule.egress_stat_mode_offered_total_cir_exceed • 9 = pccrule.egress_stat_mode_offered_four_profile_no_cir

PCC rule TLV	M|O	TLV details	
			• 10 = pccrule.egress_stat_mode_offered_total_cir_four_profile
		Default	None
pccrule.policer_packet_byte_offset (Integer)	O	Purpose	Specifies the dynamic policer packet-byte-offset for this PCC rule. Setting the value to zero (0) sets no packet-byte-offset. Overrides the dynamic policer value configured in the sap-ingress or sap-egress QoS policy.
		Value	ingress: -32 to +31 egress: -64 to +31
		Default	None

Notes:

- (1) At least one PCC rule action must be specified.
- M=Mandatory, O=Optional

Table 46: alc.sub_svc.flow functions lists the alc.sub_svc.flow function.

Table 46: alc.sub_svc.flow functions

alc.sub_svc.flow functions	Description
sub_svc.flow.add_to_flow (flow, flow TLV, value)	Appends a flow TLV such as dscp, protocol, src-ip, dst-ip, src-port, or dst-port to the flow list. The flow list defines matching criteria for an IP flow and can be added to a PCC rule with the sub_svc.pccrule.add_to_pccrule() function. Parameters: flow (type = list): list containing the match criteria (DSP, 5-tuple) that describes an IP flow. Flow TLVs are appended to this list with the sub_svc.flow.add_to_flow() function. The flow is added to a PCC rule with the sub_svc.pccrule.add_to_pccrule() function. flow TLV (type int): Flow TLV that is appended to the flow list. value (type as defined for the flow TLV): the value of the flow TLV that is appended to the flow list

Table 47: PCC rule TLVs alc.sub_svc.flow module lists the PCC Rule TLVs alc.sub_svc.flow Module.

Table 47: PCC rule TLVs alc.sub_svc.flow module

Flow TLV	M|O	TLV details	
flow.dscp (string)	O	Purpose	Specifies a DSCP flow match criterion

Flow TLV	M|O	TLV details	
		Value	Fixed DSCP name strings as in the output of **show qos dscp-table; f**or example, "be" or "ef". The DSCP name must be specified in lowercase.
		Default	n/a
flow.protocol (integer)	O	Purpose	Specifies a protocol number match criterion
		Value	0 to 255
		Default	n/a
flow.dst-ip (string)	O	Purpose	Specify a destination IPv4 or IPv6 match criterion
		Value	*ipv4-address* | *ipv6-address* | **any** where *ipv4-address*: d.d.d.d[/m] d [0 to 255] m [0 to 32] *ipv6-address*: x:x:x:x:x:x:x:x[/preflen] x: [0 to FFFF] preflen: 0 to 128
		Default	any
flow.dst-port (string)	O	Purpose	Specify a destination port match criterion
		Value	port or port range: *port*[-*port*] where *port*: 0 to 65535
		Default	n/a
flow.src-ip (string)	O	Purpose	Specify a source IP or IPv6 match criterion
		Value	*ipv4-address* | *ipv6-address* | **any** where *ipv4-address*: d.d.d.d[/m] d [0 to 255] m [0 to 2] *ipv6-address*: x:x:x:x:x:x:x:x[/preflen] x: [0 to FFFF] preflen: 0 to 128
		Default	any

Flow TLV	M|O	TLV details	
flow.src-port (string)	O	Purpose	Specifies a source port match criterion
		Value	port or port range: *port*[-*port*] where *port*: 0 to 65535
		Default	n/a

9.22.9 Operational commands

9.22.9.1 Show commands

To display the active subscriber services in the system, use the show service sub-services CLI command. The **sub-service-name** filter is a longest match.

```
# show service sub-services [id <service-id>] [sap <sap-id>] [ip <ip-prefix/prefix-length>]
 [mac <ieee-address>] [pppoe-session-id <pppoe-session-id>] [sub-service-name <sub-service-
name>] [sub-service-type <sub-service-type>] [summary|associations]
```

Sample output:

```
# show service sub-services
===============================================================================
Subscriber service table
===============================================================================
Username        : cpe2@domain1.com
Subscriber      : cpe2@domain1.com                    Type     : PPP
SAP             : 1/1/4:1201.5                        Service  : 1000
MAC Address     : 00:51:00:00:01:01                   PPPoE-SID: 114
IP              : 10.1.1.3
Interface-ID    : 02:51:00:FF:FE:00:01:01
-------------------------------------------------------------------------------
  Service instance : rate-limit;5120;30720
  up time          : 0d 00:05:23
  type             : rate-limit
  acct sess id     : 144DFF00000A2556B757DA
  multi sess id    : 144DFF00000A1156B71E44
  acct type        : volume-time
  acct ivl         : 1d 00:00:00
  input octets     : 0
  input packets    : 0
  output octets    : 0
  output packets   : 0
  actions          :
  QoS-override
    ingress arbiter "root" rate 5120
  QoS-override
    egress aggregate rate limit 30720
-------------------------------------------------------------------------------
number of subscriber services found: 1
===============================================================================
```

To display the active PCC rules in the system, use the **show service active-subscribers pcc-rule** CLI command. A PCC rule can be inactive when, for example, a PCC rule with filter actions on IPv6 flows is activated on an IPv4single-stack PPPoE or IPoE session.

```
# show service active-subscribers pcc-rule [subscriber <sub-ident-string>] [detail]
```

Sample output:

```
# show service active-subscribers pcc-rule subscriber "ipoe-bridged-001" detail
===============================================================================
Active Subscribers
===============================================================================
-------------------------------------------------------------------------------
Subscriber ipoe-bridged-001 (sub-profile-1)
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
(1) SLA Profile Instance sap:[1/1/4:1201.19] - sla:sla-profile-1
-------------------------------------------------------------------------------
IP Address
                MAC Address          Session        Origin       Svc        Fwd
-------------------------------------------------------------------------------
2001:db8:1:101::1/128
                00:51:00:00:00:05    IPoE           DHCP6        1000       Y
-------------------------------------------------------------------------------
Ingr Filter Override     : N/A
Egr  Filter Override     : N/A
Ingr Qos Policy Override : 10:P10
Egr  Qos Policy Override : 10:P9
===============================================================================
Precedence Rule Id   Rule Name                Subscriber Service Name
-------------------------------------------------------------------------------
10         19        rule-1                   subsvc_pcc-coa-1
10         20        rule-2                   subsvc_pcc-coa-1
===============================================================================
===============================================================================
PCC Rules
===============================================================================
PCC rule name         : rule-1
PCC rule id           : 19
Monitoring key        : -
Flow status           : Enabled
Nbr of Flows          : 2 (ingress)
HTTP-Redirect         : -
Next-Hop Redir. IPv4  : -
Next-Hop Redir. IPv6  : -
QoS Ingr. CIR/PIR     : - / 1000 kbps
QoS Egr. CIR/PIR      : - / -
FC change             : -
Account               : Enabled
-------------------------------------------------------------------------------
Flows
-------------------------------------------------------------------------------
Src. IP  : any                                  Src. Port: -
Dst. IP  : 172.16.1.1/32                        Dst. Port: -
Protocol : -                                    DSCP     : -
-------------------------------------------------------------------------------
Src. IP  : any                                  Src. Port: -
Dst. IP  : 2001:db8:aaa:1::1/128                Dst. Port: -
Protocol : -                                    DSCP     : -
-------------------------------------------------------------------------------
===============================================================================
PCC rule name         : rule-2
```

```
PCC rule id           : 20
Monitoring key        : -
Flow status           : Enabled
Nbr of Flows          : 2 (egress)
HTTP-Redirect         : -
Next-Hop Redir. IPv4  : -
Next-Hop Redir. IPv6  : -
QoS Ingr. CIR/PIR     : - / -
QoS Egr. CIR/PIR      : - / 5000 kbps
FC change             : -
Account               : Enabled
-------------------------------------------------------------------------------
Flows
-------------------------------------------------------------------------------
Src. IP  : 172.16.1.1/32                      Src. Port: -
Dst. IP  : any                                Dst. Port: -
Protocol : -                                  DSCP     : -
-------------------------------------------------------------------------------
Src. IP  : 2001:db8:aaa:1::1/128              Src. Port: -
Dst. IP  : any                                Dst. Port: -
Protocol : -                                  DSCP     : -
-------------------------------------------------------------------------------
===============================================================================
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
```

Use the following alternative command to check the PCC rules in the system:

```
# show subscriber-mgmt pcc-rule
  - pcc-rule
  - pcc-rule monitoring-key <key> detail
  - pcc-rule rule-id <id> detail
  - pcc-rule rule-name <rule-name>
  - pcc-rule rule-name <rule-name> detail
  - pcc-rule summary
  - pcc-rule monitoring-key <key>
 <rule-name>          : [100 chars max]
 <key>                : [80 chars max]
 <id>                 : [1..1023]
 <detail>             : keyword
```

The statistics of dynamic policers can be displayed with:

```
# show service active-subscribers subscriber "ipoe-bridged-001" detail
--- snip ---
-------------------------------------------------------------------------------
SLA Profile Instance per Policer statistics
-------------------------------------------------------------------------------
                        Packets                 Octets
--- snip ---
Ingress Policer 10 (Stats mode: minimal)
used by pcc-rule rule-1
Off. All              : 0                       0
Dro. All              : 0                       0
For. All              : 0                       0
--- snip ---
```

The details of the cloned QoS and filter policies as a result of PCC rule activation can be displayed with the following show commands:

```
# show qos sap-ingress 10:P1
# show qos sap-egress 10:P1
```

```
# show filter ip 10:P1
# show filter ipv6 "10:P1"
```

9.22.9.2 Debug commands

There are no specific RADIUS subscriber services debug commands. The debugging is part of the RADIUS and Python debug output; for example:

```
debug
    router "Base"
        radius
            packet-type authentication accounting coa
            detail-level high
        exit
    exit
    python
        python-script "subsvc-1"
            script-all-info
        exit
    exit
exit
```

9.22.9.3 Resource monitoring

For information about resource monitoring, see PCC Rules and Capacity Planning and PCC Rule Scaling Example.

The following CLI command provides an overview of the resource usage per line card, such as the number of ACL and ACL QoS entries, Filters, QoS policies, dynamic policers, and QoS overrides:

```
# tools dump resource-usage card [<slot-number>] all
```

These resource counters are available in SNMP and can be used in RMON to trigger threshold crossing alarms; for example:

```
configure system
    thresholds
        rmon
            alarm 1 variable-oid tFPResIngIPv6AclEntryAlloc.1.1.1 interval 10 rising-event 1
 rising-threshold 25000 falling-event 2 falling-threshold 24000
            event 1 description "Ingress IPv6 ACL Entries too high"
            event 2 description "Ingress IPv6 ACL Entries - below limit"
        exit
```

The summary output of the **show subscriber-mgmt pcc-rule** command lists the number of active PCC rules and the number of active combinations:

```
# show subscriber-mgmt pcc-rule summary
===============================================================================
PCC Rules Summary
===============================================================================
Total Nbr PCC Rules     : 2 / 1024
Nbr Active PCC Rules    : 2 / 1024
Nbr Active Combinations
  IPv4 Filter           : 0 / 4095
  IPv6 Filter           : 0 / 4095
```

```
  Egress Qos            : 1 / 4095
  Ingress Qos           : 1 / 4095
===============================================================================
```

9.23 Residential gateway replacement

Residential gateway (RG) replacements are performed for a variety of reasons such as upgrading hardware, replacing broken equipment, or relocating to a new home. However, the BNG's anti-spoof filter and host-limit features can sometime prevent immediate RG replacement. In some cases, a subscriber must wait for an old DHCP lease to expire before a new RG can connect to the BNG. For example, some service providers assign an IP address and prefix based on physical line, sap-id, circuit-id, or interface-id. Therefore, a home is always assigned the same IP address and prefix. On the BNG, an anti-spoof mechanism prevents different MAC addresses from using the same IP address. As a result, the new RG fails the anti-spoof filter and is denied an IP address and/or prefix. The subscriber in this case must wait for the DHCP lease of the RG to time out for the anti-spoof filter to remove its entry.

Two features, **lease-override** and **shcv-policy**, may help improve the RG replacement process. These features focus on minimizing service interruption and enhancing the end subscriber experience. RGs, in general, have no mechanisms to inform the BNG or the DHCP server that they have been disconnected from the network. Even if the BNG has periodic SHCV enabled, the detection may take some time. Often, when a subscriber plugs in a new RG, the BNG still has the old RG registered as a host. This has two consequences. First, if the new RG is assigned the same IP address as the old RG, then an IP-conflict occurs and fails the anti-spoof filter. Second, if the SAP has a host limit or a session limit provisioned, then exceeding the limit prevents the new RG from receiving an IP address or prefix.

Starting in Release 13.0R4, if an IP conflict occurs on the same SAP, then by default the new RG (MAC) immediately overrides the DHCP lease of the old device. This is known as **lease-override**. This is applicable to DHCP relay and proxy for both IPv4 and IPv6 hosts. Before Release 13.0.R4, **lease-override** only occurred for DHCPv4 relay. The **lease-override** is performed only when an IP conflict occurs within the same SAP.

The other option is to use trigger SHCV to check the connectivity status of the old RG before removing it and its lease. This is known as the ip-conflict-triggered SHCV under the SHCV policy. The SHCV is sent only when the BNG detects an IP address conflict on DHCP discovery. If the host does not respond within the configured timeout, both the host and lease are removed from the BNG. The new RG is required to perform a subsequent DHCP discovery or request to install a host. SHCV can help prevent malicious RGs from spoofing another RG IP address. Trigger SHCV for IP-conflict is available for DHCPv4/v6 relay and proxy, as well as ARP hosts. The following table specifies when the SHCV is sent for IP-conflict.

Table 48: IP-conflict SHCV triggering points

Configuration on group interface	Triggered on
DHCPv4 proxy	DHCP Discovery
DHCPv4 relay	DHCP Request
DHCPv6 proxy	DHCP Advertisement
DHCPv6 relay	DHCP Advertisement
ARP-host	Host's initial ARP

It is also possible that new RGs are denied service as a consequence of a set of host limits against the subscriber including sla-profile host-limits and session-limits, sub-profile host-limits and session-limits, ipoe session-limit, and ipoe sap-session limits. For example, setting a host limit of **overall 1** can ensure that each home only takes one IP address. As mentioned before, RGs do not inform the BNG of a disconnect. If SHCV is enabled, it may take some time before the BNG is informed of the disconnect. Therefore, when a new RG connects to the BNG, the BNG performs a host-limit check (if configured) against the subscriber. If the old host still has an entry on the BNG and there is a host-limit of **overall 1**, the new RG is denied an IP address and prefix assignment because it has exceeded the host limit. A trigger SHCV, "host-limit-exceeded" inside the SHCV policy can be configured against a group interface. This SHCV is triggered when an over limit is detected. If the existing host registered on the BNG does not respond within the configured timeout, both the host and its lease are removed from the BNG. The SHCV can only remove hosts from the BNG and the new RG is still required to perform a subsequent DHCP discoveries or requests to obtain an IP address.

By providing lease-override and various SHCV triggers in the SHCV policy, service providers have a variety of options to allow subscribers to perform quick and seamless RG replacements.

It is possible to use the host-connectivity check without the SHCV policy. The main function of the host-connectivity check is for periodic check. The trigger functions are performed through the SHCV policy.

9.24 ESM troubleshooting show command

Network operators are sometimes unable to turn on debug to troubleshoot customers issues on a live production network. Turning on debug may affect the BNG performance, and some support technicians may not have access to debug and configuration commands.

The **show subscriber-mgmt errors** command is a show command that captures detailed ESM errors. This command helps diagnose problems immediately without the need to turn on the debug function. Only DHCPv4 and PPPv4 supports this command; some support details are provided below. This command does not compromise the BNG performance and does not require debug or configuration commands.

- DHCPv4 support:
  - dropped protocol messages (host setup/renew)
  - protocol timeouts
- PPPoEv4 support:
  - dropped protocol messages
  - traps

IPv6 host setup errors may be captured in error logs.

All subscriber problems are first stored in a main circular buffer. The main circular buffer is then fed into smaller circular buffers, organized by MAC addresses. When the buffer is full, the first circular buffer purges the oldest message to make room for the newest message. The smaller circular buffers (one per MAC address) store a limited number of messages per MAC. Again, the smaller buffer deletes the oldest to make room for the newest. The circular buffer, per MAC, prevents one device from holding all the error messages in the buffer. The main circular buffer can hold 5,000 errors in total, while the smaller buffer can hold 10 log entries per MAC. The circular buffer supports CPM3 and higher.

The **show** command allows sorting by MAC, subscriber SAP, SDP, and unknown-origin (unknown SAP or SDP). The **show** command allows the input of a specific MAC, SAP, or SDP to directly search for particular subscriber issues or problems.

The circular buffer only logs drop reasons for DHCPv4 and PPPoEv4. Non-error reasons are never logged; for example, a drop because of being in SRRP standby is not logged. The circular buffer has a timestamp associated with each error and the errors are listed beginning with the most recent. Error logs are lost on a HA switchover, and persistency is not supported. There is no throttling mechanism for the same errors; it is possible to fill the circular buffer with the same error message from different MAC addresses.

9.25 Subscriber accumulated statistics

The accumulated statistics policy defines which statistics for queues and policers of a particular subscriber should be collected and stored in memory. At the end of the subscriber session, the queue and policer statistics are added to the statistics already stored in memory from previous sessions. This enables operators to view queue and policer statistics even if the subscriber is offline.

The accumulated statistics policy supports up to four ingress and four egress entries. For queue statistics, v4-v6 mode is not supported, where v4 and v6 statistics are always aggregated. For policer statistics, only min-mode is supported.

When a single subscriber has a list of bridge hosts, all hosts are forced to use the same statistics policy. If hosts use a different SLA profile and the operator wants to collect the statistics for all hosts, the statistics policy must encompass all queues and policers for various SLA profiles. If there are multiple SLA profile instances for the same subscriber, the statistics are summed up for each instance on a per policer or queue basis. These statistics are not exchanged between MCS peers. Therefore, for dual-homed hosts, the statistics need to be gathered from all the nodes and then summed up. If a queue or a policer is missing from the accumulated statistics policy in the current subscriber session and offline statistics exist for that entity from previous sessions, these offline stats are lost when the current subscriber session ends.

Cumulative statistics for a subscriber are not persistent; they are only stored in memory and are lost after node reboot (the statistics start at zero).

The **show subscriber-mgmt accu-stats subscriber** command displays the cumulative statistics for a subscriber. If the subscriber is online, the cumulative statistics consist of the sum of both the subscriber statistics while online and the offline statistics. If the subscriber is offline, the last subscriber session statistics are added to the offline statistics. Cumulative statistics for a subscriber are not persistent. They are only stored in memory and are lost after a node reboot (the statistics start at zero).

When the accumulated statistics policy changes through a subscriber profile override, either through the **tools** command or through RADIUS, the stored statistics can be affected. If the new SLA profile uses the same queues and policers already stored in memory, these statistics continue to accumulate. Only queues and policers that differ are added and start from zero.

If resources for capturing offline statistics are full, a trap is generated in log 99 to warn the operator. The command **show subscriber-mgmt status system** shows the number of subscribers using these accumulated statistics and a flag in this command shows whether the usage is at its peak value.

In the **show subscriber-mgmt accu-stats-subscribers** command, active subscribers with an accumulated statistics policy configured have both the *sub-profile-name* and **accu-stats-policy** fields populated. Active subscribers that are no longer associated with an accumulated statistics policy have **accu-stats-policy** populated with "**Unknown**". Inactive subscribers have "Unknown" for both the *sub-profile-name* and the accu-stats-policy fields. This command is useful when combined with the CLI **match** command, to search for subscribers with specified properties. For example, **match** "Unknown" displays the list of subscribers that are no longer associated with any accumulative statistics policy.

It is possible for an active subscriber to have offline statistics without an accumulated statistics policy if the accu-stats-policy was removed from the sub-profile.

When the active subscriber has an accumulated statistics policy, the subscriber's offline statistics can be deleted using the following command.

clear subscriber-mgmt accu-stats subscriber *subscriber-id*

The **show subscriber-mgmt accu-stats subscriber** *subscriber-id* command then displays only the subscriber's current statistics as no statistics remain in the offline storage.

If an active subscriber does not have an accumulated statistics policy, the subscriber's offline statistics can be deleted in one of the following ways.

To remove the offline statistics for all active subscribers that are no longer associated with an accumulated statistics policy, the following command can be used.

clear subscriber-mgmt accu-stats active-subs no-accu-stats-policy

To remove the offline statistics for a group of active subscribers that is no longer associated with an accumulated statistics policy and that has a defined subscriber profile (for example, if the accumulated statistics policy has been removed from the subscriber profile), the following command can be used.

clear subscriber-mgmt accu-stats active-subs sub-profile *profile-name*

It is also possible to remove the offline statistics for an inactive (offline) subscriber. To remove offline statistics for all inactive subscribers, use the following command.

clear subscriber-mgmt accu-stats inactive-subs

To remove offline statistics for a specific inactive subscriber, use the following command.

clear subscriber-mgmt accu-stats subscriber *subscriber-id*

9.26 Hybrid access

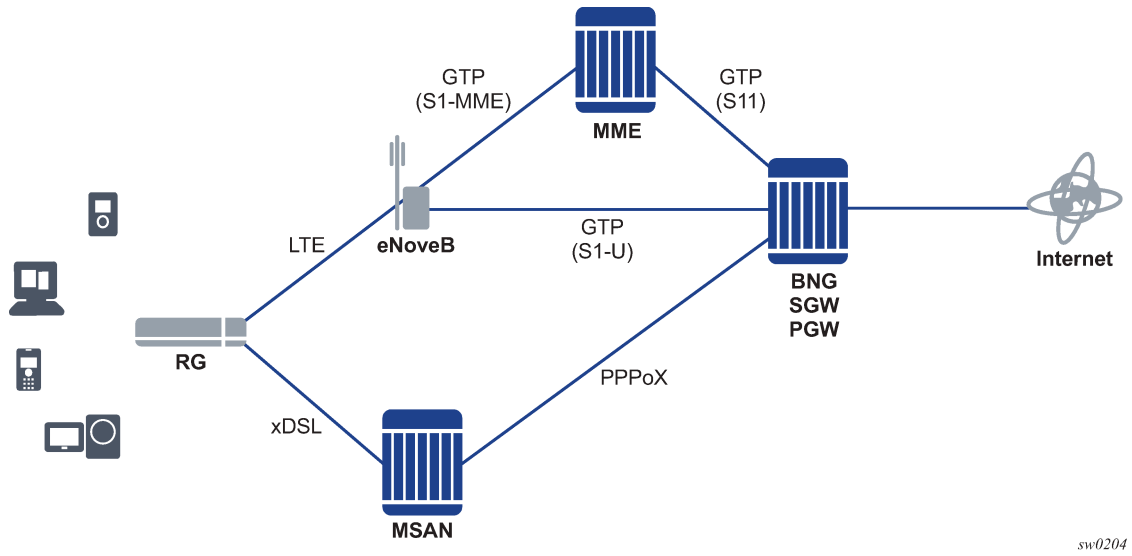
In a hybrid access deployment, a home's residential gateway (RG) is connected to the network by both a wired and a wireless link. Traffic can be split over these links by various mechanisms such as per-flow hashing, flow policies, or MP-TCP. Both access connections must be terminated in a common endpoint called the Hybrid Access Gateway (HAG). This gateway provides Internet connectivity to the home.

9.26.1 BNG-based HAG

SR OS TPSDA supports hybrid access deployments where the BNG acts as a Serving Gateway (SGW), PDN Gateway (PGW), and HAG. In this model, both the fixed access and wireless access links share the same Layer 3 IP/IPv6 address. The RG/HAG determines which Layer 2 connection should be used. To support this model, SR OS supports GTP termination and ESM connection bonding. The BNG is the Layer 3 gateway in this model and attracts all IP traffic. Multicast traffic is supported.

Figure 150: SGW/PGW/BNG integrated hybrid access gateway shows a sample hybrid access deployment with a BNG-based HAG.

Figure 150: SGW/PGW/BNG integrated hybrid access gateway

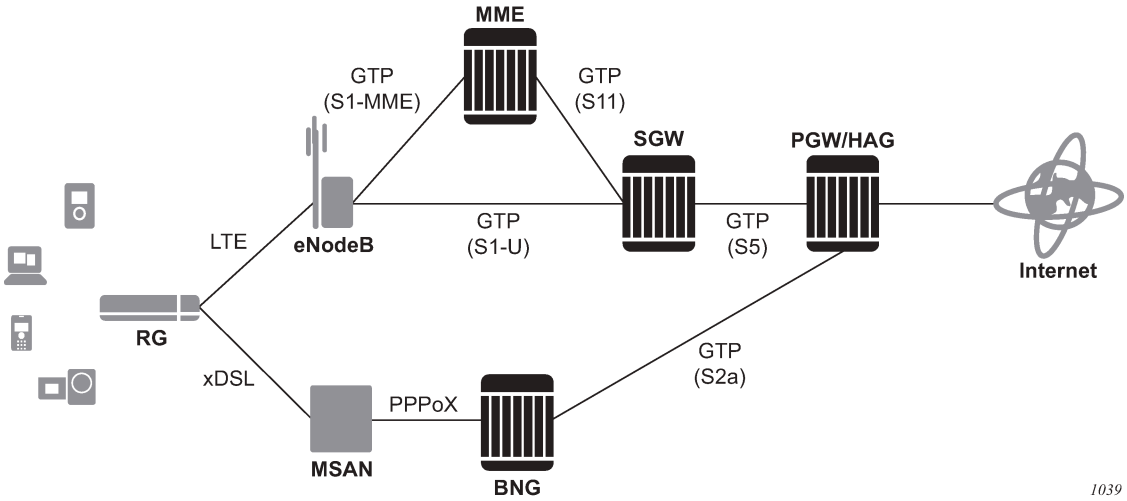


9.26.2 PGW-based HAG

SR OS TPSDA supports hybrid access deployments where a PGW acts as a HAG. In this model, both the fixed access and wireless access links share the same Layer 3 IP/IPv6 address. The RG/HAG determines which Layer 2 connection should be used. To support this model, SR OS IPoE session and PPPoE session functionality is extended to connect to a GTP uplink. When this uplink is active, all unicast traffic is forwarded by a GTP tunnel between the BNG and PGW. This way, the PGW acts as the Layer 3 gateway and the BNG does not attract subscriber traffic by regular routing. Multicast traffic is not supported and should be handled out-of-band.

Figure 151: Sample hybrid access deployment shows a sample Hybrid Access deployment with a PGW-based HAG.

Figure 151: Sample hybrid access deployment



For more details on GTP uplinks, see the GTP section.

9.27 Connection bonding

ESM connection bonding allows two Layer 2 access connections, for example, GTP and PPPoE, to combine to share a single Layer 3 IP connection. This allows the seamless use of two connections for additional bandwidth or resiliency without the need to manage multiple IP addresses.

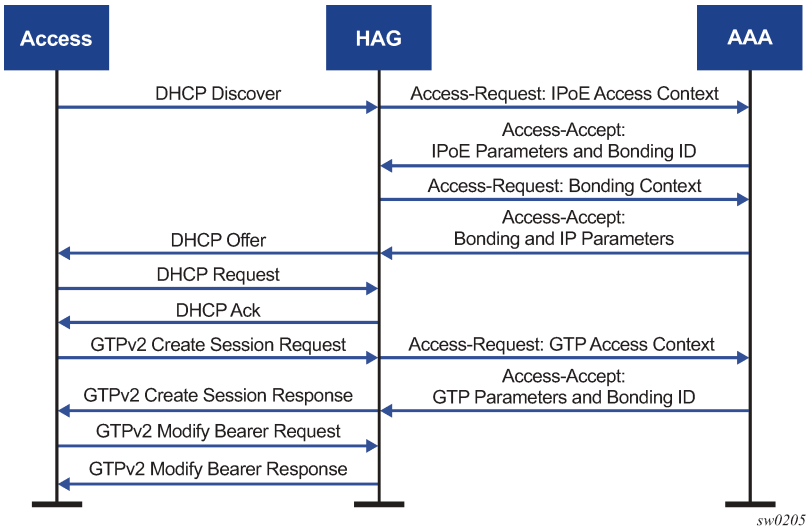
SR OS spreads downstream traffic, either based on fixed weights, a filter decision, a pure active/standby, or dynamic load-balancing that attempts to saturate one link before using another link. Upstream traffic can enter through either connection, but it is recommended to keep flows identified by 5-tuple on the same link to avoid reordering.

Connection bonding requires an FPE type of **sub-mgmt-extensions**.

9.27.1 Setup

During normal authentication, access connections can indicate they are part of a common bonding context by specifying a bonding identifier. When the first connect is set up, an additional authentication phase is started for the bonding context itself. Figure 152: Bonding authentication shows RADIUS-based authentication for bonding of an IPoE and GTP access connection. For simplicity, all access nodes, such as the residential gateway, MSAN, eNodeB, and MME have been identified as a single entity.

Figure 152: Bonding authentication



All address assignments and Layer 3 parameters are shared between the access connections and are therefore handled in the bonding context. The system supports either Local Address Assignment or AAA-provisioned IP addresses. Access connections cannot use any DHCPv6 relay or client mechanisms.

After the setup is complete, ESM subscriber resources are created for each context as follows.

- a unique subscriber with a single internal IPoE session to represent the bonding context

  This is the main context for functionality, including QoS, filters, and accounting. This subscriber cannot be reused for any other hosts, regardless if they are bonding or not.

- a subscriber per access connection with one or more hosts as needed, depending on the access type

  This subscriber must be distinct from the bonding subscriber. Other non-bonded hosts or sessions may be present under the same subscriber. A subset of ESM features (for example, QoS, filters, and accounting) are also available in this context; however, it is recommended that the most of the feature be configured in the bonding context.

All access and bonding ESM contexts need to be present in distinct VRFs. The bonding context must be created in a special group interface of type **bonding**. This group interface is not linked to any SAPs, however, an FPE construct ensures the link between the access and bonding context.

9.27.2 Downstream load balancing

By default, downstream packets are hashed over two connections on a per-flow basis, allowing packets of the same flow to follow the same path and avoid reordering issues. Flows are identified by the 5-tuple <src-ip, dst-ip, protocol, src-port, dst-port>. The hash weights of each connection are configurable. An IP-filter based selection may be used to override the hash-based connection selection.

The initial hash weights can also be dynamically adapted based on the load on the primary connection. The IOM periodically measures how much traffic is sent over the primary connection, comparing it to a predefined saturation rate. When the connection is saturated, the hash weights automatically change to send more traffic over the alternate link. Similarly, if the total rate of traffic decreases, the hashing weights change to send more traffic over the primary connection.

For dynamic load-balancing, the following must be defined:

- the primary connection, identified as connection 1 in the bonding context
- the reference rate in the primary connection, which defines the saturation point

  By default, the aggregate rate is used, but this can be changed during AAA authentication. The saturation point is further refined using thresholds that define when the rate change should be executed.
- a weight change value, specifying how much the weight percentages should change each time the saturation point is reached
- policers for all downstream traffic in the bonding context

  A single policer that directly feeds into multiple local queues can be used.

When only a single connection is active, all traffic is sent to this connection, regardless of hashing weights or filters.

If one of the two access connections is idle, then the system activates this connection before changing hashing weights. This sequence allows the system to avoid overflowing the packet buffer of the idle connection. For example, for an idle S11 GTP connection, the system reactivates the connection through a network-triggered service request before changing weights.

9.27.3 QoS

Regular ESM QoS is supported in both the access and bonding contexts; however, there is no direct feedback mechanism between the two contexts. Therefore, if an access connection drops a packet, it is not reflected in bonding statistics, nor does it cause backpressure on the bonding QoS algorithm.

When traffic passes over the FPE from the bonding context to the access context or from the access context to the bonding context, the system keeps the traffic classification and the in- and out-profile markings. Although this occurs automatically, bonding subscriber policies for ingress and egress should consider the following recommendations.

- Enable **de-mark** for access egress and map each FC to the dot1p as defined in Table 49: FC to dot1p mapping, therefore ensuring that the same classification is used in the access connection context.
- Perform classification for access ingress based on dot1p as defined in Table 49: FC to dot1p mapping and enable **in-profile** and **de-1-out-profile** for each FC, therefore ensuring the same classification is used as for the access connection context. A different classification scheme can be used if required, for example, based on DSCP or IP criteria.

Table 49: FC to dot1p mapping

FC	dot1p
be	0
l2	1
af	2
l1	3
h2	4
ef	5

FC	dot1p
h1	6
nc	7

To support load-balancing in the bonding context, the configured **stat-mode** of any egress policer in the bonding context is ignored. Instead, an internal stat-mode is used, which uses two counters (one per access connection), which is reflected in the in- and out-of-profile statistics.

9.27.4 Multicast

Multicast replication is supported in context of the access connections. By default, multicast streams are replicated in the connection where the corresponding IGMP/MLD join is received; however, this can be overridden to always force a specific connection.

When one connection fails, multicast replication automatically sets up in the alternate connection and does not require a new IGMP/MLD packet to arrive.

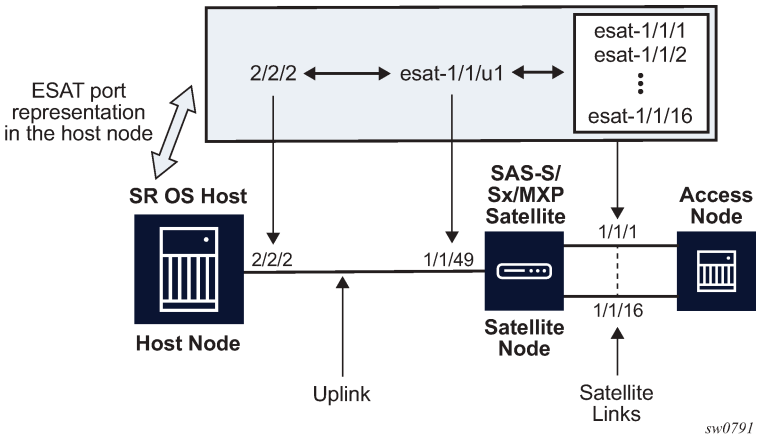
MCAC bandwidths must be configured equally on both access connections; otherwise, there may be unexpected drops of (S,G) pairs.

If a multicast stream is forwarded over the primary connection and **egress-rate-modify** is in use, any potential change of the reference rate is taken into consideration by the load-balancing mechanism for unicast traffic as described in Downstream load balancing. When using per-host replication for a bonded host, similar adaptations are made based on channel definitions in applicable MCAC policies.

9.28 Ethernet satellites with redundant uplinks

Ethernet satellite (ESAT) ports in the host node are logical ports (for example, esat-1/1/1) that represent physical ports on the remote (satellite) node. ESM in the host node treats ESAT ports similarly to all other physical ports in the system, even without knowing that those ports reside in a remote chassis. An example of an Ethernet satellite configuration is shown in Figure 153: Concept of ESAT.

Figure 153: Concept of ESAT



An SR OS ESAT complex (an SR OS host node with satellite nodes) supports multiple pairs of active and standby uplinks. The supported topologies are described in the following sections.

9.28.1 Single host, single satellite

The following describes a single host and a single satellite topology.

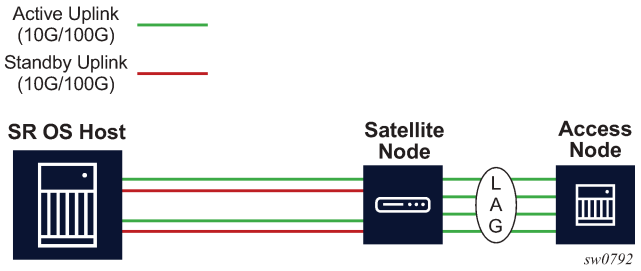
- A/S uplinks can be on the same forwarding path (FP) or forwarding complex or on a different FP. An FP refers to a set of chipsets on a line card used to simultaneously forward paired upstream and downstream traffic.
- Although a pair of uplinks can be active or standby for the same set of ports, both uplinks can be active for a set of different ports. For example:

Table 50: Active and standby ports

Ports	Active	Standby
esat-1/1/u1	satellite ports 1-12	satellite ports 13-24
esat-1/1/u2	satellite ports 13-24	satellite ports 1-12

Figure 154: Single host, single satellite displays an example of an SR OS host, satellite node and access node.

Figure 154: Single host, single satellite



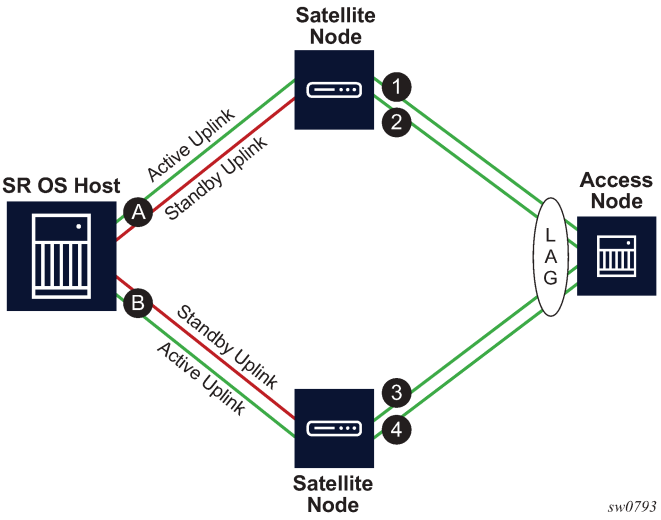
9.28.2 Single host node, dual satellite

The following describes a single host and dual satellite topology.

- Ports 1, 2, 3, and 4 are in the same LAG.
- Hashing decisions on ports 1, 2, 3, and 4 are made at ingress on the host node.
- Ports 1 and 2 can be reached only through side A.
- Ports 3 and 4 can be reached only through side B.
- If the active uplink on side A fails, the standby link on the side A becomes the new active uplink servicing the same satellite ports (ports 1 and 2).
- If both uplinks on the same side fail, then the corresponding satellite node becomes isolated.

Figure 155: Single host, dual satellite displays an example of an SR OS host, satellite node, and access node.

Figure 155: Single host, dual satellite



9.28.3 QoS

Queues and policers in ESM are created on a per SLA profile instance in the host node. A subscriber host resides in a host node on a SAP that is associated with a logical port (mapped to a user port on the satellite node) which is then associated with the physical uplink.

Subscriber aggregate rate and subscriber level schedulers are subscriber-level configurations and therefore, are independent of the ports. However, port schedulers and Vports (**agg-rate-limit** or **port-scheduler**) are port level features. They are also created in the host node, on a per-user-port basis (user ports are in the host node represented by a logical ports). They must be manually created per logical port in the host even though those ports may be LAG members.

Buffer pools are the only QoS configurations that are created on a per-physical uplink basis.

9.28.4 Preservation of statistics and accounting in ESM

ESM accounting is based on queue and policer statistics. Consider that queues are recreated and remapped every time a logical port (satellite port) or an uplink for the subscriber is changed, expect that the ESM accounting can be affected by this. If the new uplink is on the same FP as the old, then statistics are preserved. Otherwise, statistics are lost.

9.29 Multi-chassis synchronization of RADIUS usage counters

The following section describes multichassis synchronization of RADIUS usage counters.

9.29.1 Overview

SR OS supports synchronization of usage counters that can be reported through RADIUS accounting in a dual-homed BNG scenario.

The active (SRRP master state) node keeps the total number of statistics that are reported. The active node synchronizes those statistics in regular intervals with MCS to the standby node. This way, the main copy of the total statistics is maintained on both nodes and failure cases of link, node, and so on, these statistics can be recovered from the surviving node.

9.29.2 MCS interval

The statistics are synchronized at preconfigured intervals (the MCS interval) which are independent of interim-update intervals using the **config**>**subscr-mgmt**>**acct-plcy**>**mcs-interval** *minutes* | **use-update-interval** command. The MCS minutes interval value can also be the same as the interim-update-interval use-update-interval.

If there are multiple RADIUS accounting policies in a subscriber profile, the minimum value of all the configured MCS intervals in these RADIUS accounting policies is used for usage counter synchronization.

9.29.3 Usage counters synchronized

The following SPI-level counters are synchronized:

INPUT OCTETS [42]

OUTPUT OCTETS [43]

ACCT-INPUT-GIGAWORDS [52]

ACCT-OUTPUT-GIGAWORDS [53]

Alc-IPv6-Acct-Input-Octets [26-6527-195]

Alc-IPv6-Acct-Output-Octets [26-6527-198]

Alc-IPv6-Acct-Input-Gigawords [26-6527-196]

Alc-IPv6-Acct-Output-Gigawords [26-6527-199]

In addition to aggregate accounting counters, detailed per queue and policer counters are also synchronized.

The following accounting attributes are synchronized per queue and policer:

Alc-Acct-I-Inprof-Octets-64 [26-6527-19]

Alc-Acct-I-Outprof-Octets-64 [26-6527-20]

Alc-Acct-O-Inprof-Octets-64 [26-6527-21]

Alc-Acct-O-Outprof-Octets-64 [26-6527-22]

9.29.4 Incomplete MCS configuration

After the **mcs-interval** is configured, statistics are collected or baselined on the CPM even if the configuration to synchronize statistics is not complete (no peer, no sync tag, and so on).

9.29.5 Configuration mismatch

In misconfiguration scenarios when two nodes use different queues or policers, the local configuration wins and decides those queues that are stored or baselined on the CPM.

9.29.6 Switchover scenarios

The following are switchover scenarios.

- Node reboot — Because the main copy of the stats is maintained on both nodes, the stats are recovered from the remaining node.
- Split brain — A split brain scenario should never occur. If it does, the statistics are reported from both nodes (both in SRRP master state). After the MCS recovers, the statistics are re-synchronized at the regular MCS intervals.

  The active node accepts updates received from the other node only when the value of the counter is larger than the local total.
- SRRP switchovers— If a host is installed by MCS, and a SRRP switchover occurs before the statistics are retrieved from the active node, the RADIUS accounting messages going out before the statistics are retrieved have unsynced statistics (statistics on a time the node did not yet receive an answer from its peer and is still requesting statistics).

  After retrieving the stats, retrieved and unsynced counter values are compared and the higher counter value is chosen for accounting.

9.30 Configuring ESM with CLI

This section provides information to configure subscriber management features using the command line interface. It is assumed that the reader is familiar with VPLS and IES services.

9.30.1 Configuring RADIUS authentication of DHCP sessions

When RADIUS authentication for subscriber sessions is enabled, DHCP messages from subscribers are temporarily held by the BSA, while the user's credentials are checked on a RADIUS server.

Configuring RADIUS authentication for subscriber sessions is done in two steps:

- First define an authentication-policy in the **config**>**subscriber-mgmt**>**authentication-policy** context.
- Then apply the policy to one or more SAPs in the **config**>**service**>**vpls**>**sap**>**authentication-policy** *auth-plcy-name* context (for a VPLS service).

  Or apply the policy to one or more interfaces **config**>**service**>**ies**>**if**>**authentication-policy** *auth-plcy-name* context (for an IES service):

The following example displays a partial BSA configuration with RADIUS authentication:

```
A:ALA-1>config>service# info
----------------------------------------------
subscriber-management
    authentication-policy BSA_RADIUS create
```

```
        description "RADIUS policy for DHCP users Authentication"
        password "mysecretpassword"
        radius-authentication-server
            server 1 address 10.100.1.1 secret "radiuskey"
            retry 3
            timeout 10
        exit
        re-authentication
        user-name-format circuit-id
    exit
exit
...
vpls 800 customer 6001
    description "VPLS with RADIUS authentication”
    sap 2/1/4:100 split-horizon-group DSL-group create
        authentication-policy BSA_RADIUS
    exit
    sap 3/1/4:200 split-horizon-group DSL-group create
        authentication-policy BSA_RADIUS
    exit
    no shutdown
exit
...
-----------------------------------------------
A:ALA-1>config>service#
```

9.30.2 TCP MSS adjustment for ESM hosts

TCP MSS adjustment is supported to prevent fragmentation of TCP packets from/to ESM hosts. See the TCP MSS Adjustment for ESM Hosts section of the *7450 ESS, 7750 SR, and VSR Multiservice ISA and ESA Guide*.

9.30.3 Configuring ESM

9.30.3.1 Basic configurations

Configuring and applying the Enhanced Subscriber Management profiles and policies are optional. There are no default Profiles or policies.

The basic Enhanced Subscriber Management profiles and policies must conform to the following:

- Unique profile or policy names (IDs).
- Profiles and policies must be associated with a VPLS or IES service to facilitate Enhanced Subscriber Management.
- QoS and IP filter entries configured in ESM profiles and policies override the defaults and modified parameters or the default policies.
- The ESM profiles and policies must be configured within the context of VPLS or IES.

9.30.3.2 Subscriber interface configuration

The following output displays a basic subscriber interface configuration.

```
*A:ALA-48>config>service>ies>sub-if# info
----------------------------------------------
                description "Routed CO - Antwerp 2018"
                address 192.168.2.254/24
                address 192.168.3.254/24
                address 192.168.4.254/24
                address 192.168.5.254/24
                address 192.168.6.254/24
                group-interface "DSLAM_01" create
                    description "Routed CO - vlan / subscriber"
                    sap 1/1/2:1001 create
                        static-host ip 192.168.2.2 create
                        exit
                    sap 1/1/2:1002 create
                        static-host ip 192.168.2.2 create
                        exit
                    sap 1/1/2:1004 create
                        static-host ip 192.168.2.4 create
                        exit
                    sap 1/1/2:1100 create
                        static-host ip 192.168.2.100 create
                        exit
                    exit
                exit
----------------------------------------------
*A:ALA-48>config>service>ies>sub-if#
```

9.30.3.3 Configuring ESM entities

9.30.3.3.1 Configuring a subscriber identification policy

The following displays an example of a subscriber identification policy configuration:

```
A:ALA-48>config>subscr-mgmt# info
----------------------------------------------
...
        sub-ident-policy "Globocom" create
            description "Subscriber Identification Policy Id Globocom"
            sub-profile-map
                entry key "1/1/2" sub-profile "ADSL Business"
            exit
            sla-profile-map
                entry key "1/1/2" sla-profile "BE-Video"
            exit
            primary
                script-url "primaryscript.py"
                no shutdown
            exit
            secondary
                script-url "secundaryscript.py"
            exit
            tertiary
                script-url "tertiaryscript.py"
```

```
                no shutdown
            exit
        exit
...
-----------------------------------------------
A:ALA-48>config>subscr-mgmt#
```

9.30.3.3.2 Configuring a subscriber profile

ESM subscriber profile configurations specify existing QoS scheduler profiles. In the following example, "BE-Video-max100M" is specified in the sub-profile "ADSL Business" for the ingress-scheduler-policy. "Upload" is specified in the sub-profile egress-scheduler-policy.

```
#--------------------------------------------------
echo "QoS Policy Configuration"
#--------------------------------------------------
    qos
        scheduler-policy "BE-Video-max100M" create
            description "Scheduler Policy Id BE-Video-max100M"
            tier 1
                scheduler "tier1" create
                    description "Scheduler Policy Id BE-Video-max100M Tier 1 tier1"
                exit
            exit
        exit
        scheduler-policy "Upload" create
            description "Scheduler Policy Id Upload"
            tier 3
                scheduler "tier3" create
                    description "Scheduler Policy Id Upload Tier 3 tier3"
                exit
            exit
        exit
        sap-ingress 2 create
            description "Description for Sap-Ingress Policy id # 2"
            queue 1 create
            parent "tier1"
            exit
            queue 11 multipoint create
            parent "tier1"
            exit
        exit
        sap-egress 3 create
            description "Description for Sap-Egress Policy id # 3"
            queue 1 create
            parent "tier3"
            exit
        exit
    exit
#----------------------
```

The following displays an example of a subscriber identification policy configuration:

```
A:ALA-48>config>subscr-mgmt# info
----------------------------------------------
...
        sub-profile "ADSL Business" create
            description "Subscriber Profile Id ADSL Business"
            ingress-scheduler-policy "BE-Video-max100M"
```

```
                scheduler "tier1" rate 99
            exit
            egress-scheduler-policy "Upload"
                scheduler "tier3" rate 1 cir 1
            exit
            sla-profile-map
                entry key "1/1/3" sla-profile "BE-Video"
            exit
        exit
-------------------------------------------
A:ALA-48>config>subscr-mgmt#
```

9.30.3.3.3 Configuring an SLA profile

The following displays an example of a SLA Profile configuration:

```
A:ALA-48>config>subscr-mgmt# info
-------------------------------------------------
    subscriber-mgmt
        sla-profile "BE-Video" create
            description "SLA Profile Id BE-Video"
            ingress
                qos 2
                    queue 1
                    exit
                exit
            exit
            egress
                qos 3
                    queue 1
                    exit
                exit
            exit
        exit
-------------------------------------------
A:ALA-48>config>subscr-mgmt#
```

9.30.3.3.4 Configuring explicit mapping entries

The following displays an example of an explicit subscriber mapping:

```
A:ALA-7>config>subscr-mgmt# info
-------------------------------------------------
A:ALA-48>config>subscr-mgmt# info
-------------------------------------------
...
        explicit-subscriber-map
            entry key "1/1/1:1111" sub-profile "ADSL GO" alias "Sub-Ident-1/1/1:
1111" sla-profile "BE-Video"
        exit
...
-------------------------------------------
A:A:ALA-48>config>subscr-mgmt#
```

9.30.3.4 Routed CO with basic subscriber management features

The following displays the output of an IES service configured with and without enhanced subscriber management and only applies to the 7750 SR.

```
A:term17>config>service>ies# inf
------------------------------------------
            subscriber-interface "s2" create
                address 10.20.1.1/16
                dhcp
                    gi-address 10.20.1.1
                exit
                group-interface "g3" create
                    description "With Enhanced Subscriber Mgmt"
                    arp-populate
                    dhcp
                        server 10.1.1.1
                        trusted
                        lease-populate 8000
                        no shutdown
                    exit
                    sap lag-1:11 create
                        sub-sla-mgmt
                            def-sub-profile "subProf"
                            def-sla-profile "slaProf"
                            sub-ident-policy "foo"
                            multi-sub-sap
                            no shutdown
                        exit
                        host ip 10.20.1.10 mac 00:00:aa:aa:aa:dd subscriber "One" sub-
profile "subProf" sla-profile "slaProf"
                    exit
                exit
            exit
            subscriber-interface "s3" create
                address 10.39.1.1/16
                dhcp
                    gi-address 10.39.1.1
                exit
                group-interface "g5" create
                    description "Without Enhanced Subscriber Mgmt"
                    arp-populate
                    dhcp
                        server 10.1.1.1
                        trusted
                        lease-populate 8000
                        no shutdown
                    exit
                    sap 4/1/1:24.4094 create
                    exit
                exit
            exit
            no shutdown
------------------------------------------
A:term17>config>service>ies#
```

9.30.3.5 Applying the profiles and policies

Note: Subscriber interfaces operate only with basic (or enhanced) subscriber management. At the very least, a host, either statically configured or dynamically learned by DHCP must be present in order for the interface to be useful. This note applies to the 7750 SR only.

Apply the ESM profiles and policies to the SLA profile.

9.30.3.5.1 SLA profile

The following syntax applies to the 7450 ESS:

CLI syntax:

```
config>service>ies service-id
    interface ip-int-name
        sap sap-id
            host {[ip ip-address] [mac ieee-address} [subscriber sub-ident-string] [sub-
profile sub-profile-name] [sla-profile sla-profile-name]
```

The following syntax applies to the 7750 SR:

CLI syntax:

```
config>service>ies service-id
    interface ip-int-name
        sap sap-id
            host {[ip ip-address] [mac ieee-address} [subscriber sub-ident-string] [sub-
profile sub-profile-name] [sla-profile sla-profile-name]
            sub-sla-mgmt
                def-sla-profile default-sla-profile-name
                single-sub-parameters
                    non-sub-traffic sub-profile sub-profile-name sla-profile sla-profile-name
 [subscriber sub-ident-string]
    subscriber-interface ip-int-name
            group-interface ip-int-name
                sap sap-id
                    host ip ip-address [mac ieee-address] [subscriber sub-ident-string] [sub-
profile sub-profile-name] [sla-profile sla-profile-name]
                    sub-sla-mgmt
                        def-sla-profile default-sla-profile-name
                        single-sub-parameters
                            non-sub-traffic sub-profile sub-profile-name sla-profile sla-
profile-name[subscriber sub-ident-string]
```

The following syntax applies to the 7450 ESS and 7750 SR:

CLI syntax:

```
config>service>vpls service-id
    sap sap-id
        host {[ip ip-address] [mac ieee-address]} [subscriber sub-ident-string] [sub-profile
 sub-profile-name] [sla-profile sla-profile-name]
        sub-sla-mgmt
            def-sla-profile default-sla-profile-name
            single-sub-parameters
                non-sub-traffic sub-profile sub-profile-name sla-profile sla-profile-
name[subscriber sub-ident-string]
```

The following syntax applies to the 7750 SR:

CLI syntax:

```
config>service>vprn service-id
    interface ip-int-name
        sap sap-id
            host {[ip ip-address] [mac ieee-address]} [subscriber sub-ident-string] [sub-
profile sub-profile-name] [sla-profile sla-profile-name]
```

The following syntax applies to the 7450 ESS and 7750 SR:

CLI syntax:

```
config>subscriber-mgmt
    explicit-subscriber-map
        entry key sub-ident-string [sub-profile sub-profile-name] [alias sub-alias-string]
 [sla-profile sla-profile-name]
    sub-ident-policy sub-ident-policy-name
        sla-profile-map
            entry key sla-profile-string sla-profile sla-profile-name
    sub-profile sla-profile-map
        sla-profile-map
            entry key sla-profile-string sla-profile sla-profile
```

9.30.4 Configuring dual homing

The following displays an example of a dual homing configuration a. The configuration shows dual homing with a peer node with a system address of 10.1.1.23. The DHCP server returns a default route with a 10.21.1.3 next hop. This example only applies to the 7750 SR.

```
A:ALA-48#
#------------------------------------------------
echo "Redundancy Configuration"
#------------------------------------------------
    redundancy
        multi-chassis
            peer 10.1.1.23 create
                sync
                    srrp
                    sub-mgmt
                    port lag-100 sync-tag "Tag1" create
                    exit
                    no shutdown
                exit
                no shutdown
            exit
        exit
    exit
#------------------------------------------------
echo "Service Configuration"
#------------------------------------------------
    service
        customer 1 create
            description "Default customer"
        exit
        sdp 23 create
            far-end 10.1.1.23
            no shutdown
        exit
```

```
        ies 40 customer 1 create
            redundant-interface "r40-1" create
                address 10.1.1.1/31
                spoke-sdp 23:1 create
                exit
            exit
            subscriber-interface "s40-1" create
                address 10.21.1.1/16 gw-ip-address 10.21.1.3
                dhcp
                    gi-address 10.21.1.1
                exit
                group-interface "g40-1" create
                    dhcp
                        server 10.1.1.1
                        lease-populate 8000
                        no shutdown
                    exit
                    redundant-interface r40-1
                    remote-proxy-arp
                    sap lag-100:1 create
                        sub-sla-mgmt
                            def-sub-profile "subProf"
                            def-sla-profile "slaProf"
                            sub-ident-policy "subIdentPolicy"
                            multi-sub-sap
                            no shutdown
                        exit
                    exit
                    sap lag-100:4094 create
                    exit
                    srrp 1 create
                        message-path lag-100:4094
                        no shutdown
                    exit
                exit
            exit
            no shutdown
        exit
exit
...
-----------------------------------------------
A:ALA-48#
```

10 Oversubscribed multichassis redundancy (OMCR) in ESM

10.1 Overview

10.1.1 Terminology and abbreviations

- **OMCR**

  Oversubscribed Multi-Chassis Redundancy

- **Warm-Standby Node** or **Protecting Node**

  Refers to the oversubscribed node that offers the protection of subscriber hosts spread over multiple BNGs. During the normal operation, the protecting node maintains the subscriber host in the form of an MCS record (Multi-Chassis Synchronization Record) in the control plane. Only when the failure occurs and the protecting node becomes active, are the subscriber-hosts fully instantiated in the data and control plane. This node is sometimes referred to as N:1 node.

- **Active/Active (1:1) Model**

  This mode of operation refers to the model where subscribers host are fully synchronized between two chassis, regardless of the state of the underlying SRRP instance (Master/Standby). Each node can have MCS peering sessions with four other nodes where each peering session represent 1 to 1 mapping set of active subscriber hosts.

10.1.2 Restrictions

- The protecting node must use CPM-4 or higher (other protected nodes can continue to use CPM-3).
- The protecting node functionality is not supported in the 7450 ESS chassis.
- All nodes in the OMCR cluster (central standby and the protected nodes) must run at the minimum SR OS Release 12.0R1.
- Warm-standby mode is a chassis-wide property. In other words, while in warm-standby mode, the chassis cannot operate in 1:1 (active-active) redundancy mode.
- OMCR is supported only for IPoEv4/v6 subscribers. However, non-synchronized PPPoEv4/v6 subscribers are supported in the OMCR cluster. PPPoEv4/v6 PTA (locally terminated) non-synchronized subscribers and the OMCR synchronized IPoE subscriber must be instantiated under separate group-interfaces. On the other hand, non-synchronized PPPoEv4/v6 LAC sessions can be under the same group interface as the OMCR synchronized IPoE subscribers. Non-synchronized PPPoEv4/v6 subscriber hosts rely on ppp-keepalive timeouts to re-establish connectivity when the failure occurs.
- Preemption of already instantiated subscriber hosts in the protecting node by another subscriber hosts is not allowed.

- Redundant interface (shunting) is not supported for subscribers on the protecting node while they are not fully instantiated in the control/data plane (or while the underlying SRRP instance is in a non-master state on the protecting node).
- Persistency in multichassis environment must be disabled because redundant nodes are protecting each other and they maintain up-to-date lease states.
- The failover trigger is based on SRRP only (no MC-LAG support).
- Unnumbered subscriber-interface model is not supported in OMCR.
- The protecting node supports 10 MCS peers, while the protected node (in active/active mode of operation) supports 4 MCS peers.
- Synchronization of the following MCS clients is not supported:
  - Host tracking
  - MC ring
  - Layer 2 subscriber hosts
  - Layer 3 IGMP/MLD
  - Layer 2 IGMP/MLD
  - DHCP Server
  - PPPoE Clients
  - MC-LAG
  - MC-IPSEC
  - MC-ENDPOINT

10.2 Deploying oversubscribed multichassis redundancy

To optimize the cost, operators prefer an oversubscribed model in which a single central standby BNG (protecting BNG) supports multiple other BNGs in a semi-stateful fashion.

In Oversubscribed Multi-Chassis Redundancy (OMCR) model, many subscriber-hosts are backed up by a single central standby node. Standby subscriber-hosts within the protecting node are synchronized only within the control plane (CPM) in the form of a Multi-Chassis Synchronization (MCS) record. Such subscriber hosts are not instantiated in the data plane and therefore the data plane resources can be spared and used only on an as needed basis. This trait allows the protecting node to back up many subscribers that are scattered over multiple active BNG nodes at the expense of slower convergence.

Only a subset of the subscribers, up to the available resource capacity of the data plane in the protecting node, would be activated on the protecting node at any specific time during the failure.

The failover trigger is based on SRRP (no MC-LAG support). The subscriber hosts under the corresponding group-interface is switched over after the SRRP instance on the protecting node transitions into the SRRP master state.

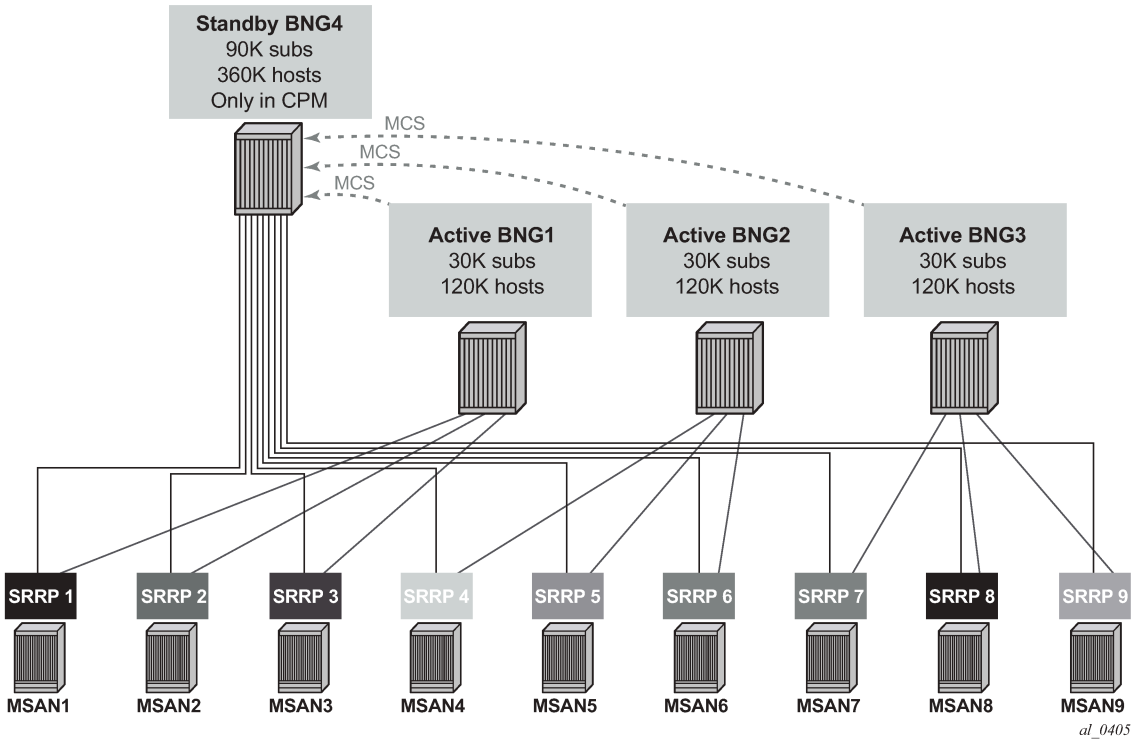
There are two possible models for this deployment:

1. Access nodes are directly connected to the BNGs. From the perspective of standby subscribers, in this model the line card is oversubscribed but the physical ports on it are not. For example, each of the 10 physical ports on the same line card can be directly connected to respective access nodes. Assume

that each physical port can support 64K subscriber hosts. Considering that the subscriber host limit per line card is also 64K (at the time of this writing), the oversubscription ratio in this case would be 10:1.

The concept of this deployment scenario is shown in Figure 156: OMCR scenario without an aggregation network.

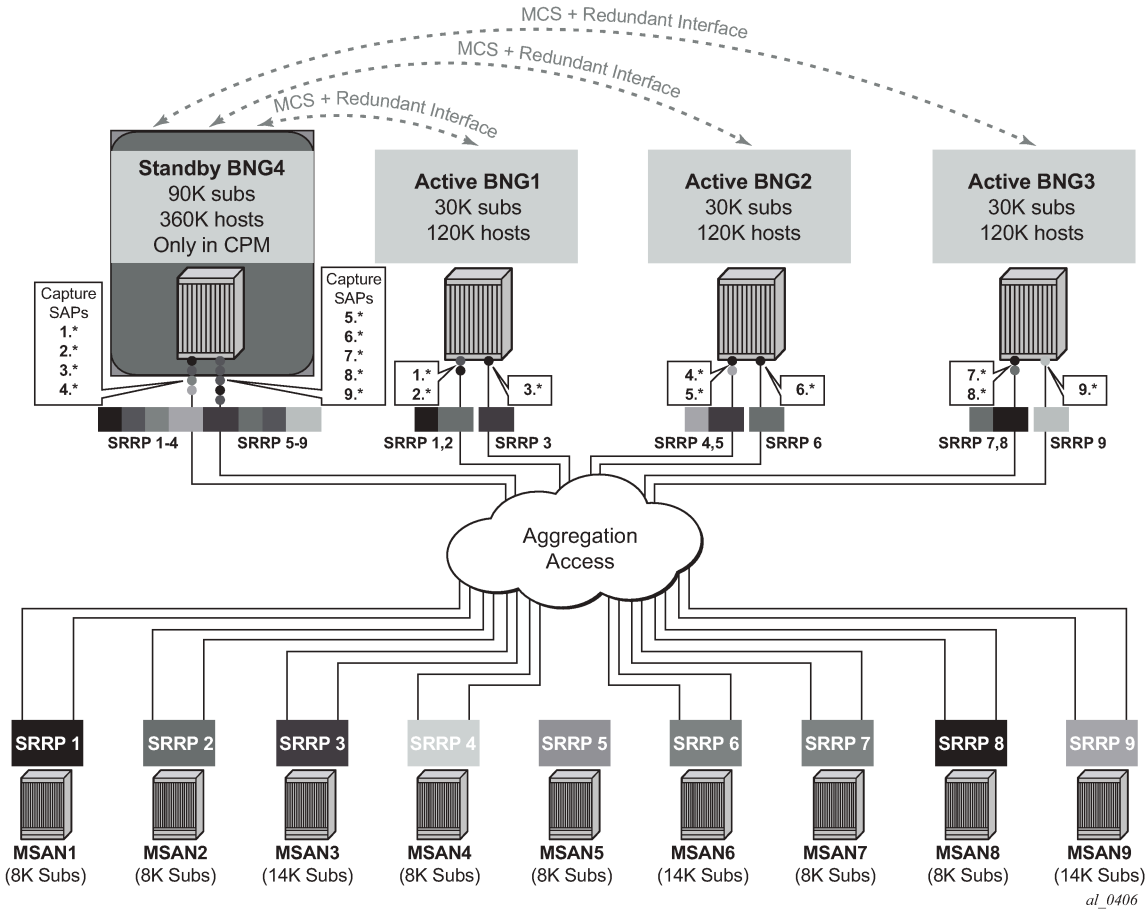
Figure 156: OMCR scenario without an aggregation network



2. Aggregation network in the access (double VLAN tags). In this case a line card and a physical port can be oversubscribed with standby subscribers. For example, multiple capture-saps (each capture SAP containing 4K c-vlans) can be created on a single physical port on the protecting BNG, for the total of >>64K subscribers per physical port.

   Conceptual model for this scenario is shown in Figure 157: OMCR scenario with an aggregation network (although the number of SRRP instances and capture-saps is in this figure is reduced for simplification).

Figure 157: OMCR scenario with an aggregation network



In both cases, a maximum of 64K subscribers per line card can be activated on the protecting BNG during the switchover. This is something that the operator should plan around, and consequently group the access nodes in a way so that the eventual number of active subscribers per line card on the protecting node does not exceed the maximum number of supported subscribers per line card.

Note: A deployment scenario can exist in which system-wide ESM capacity is oversubscribed but the line card capacity is not. For example, on chassis with 10 line cards, each line card can be reserved to protect a total host count of 64k. This yields a total of 640k protected hosts distributed across the 10 cards but only up to 256k hosts can be activated simultaneously if it is required because the SRRP transitions to master state.

10.2.1 Resource exhaustion notification and simultaneous failures

The protection success of the OMCR model relies on grouping protected entities (links and nodes) according to the likelihood of their failure within the time frame required for their restoration. For example, the same resource (IOM card or port) on the protecting node can be used to protect multiple entities in the network if their failures do not overlap in time. In other words, if one failure can be repaired before the next one contending for the same resource on the central standby node, the OMCR model serves the purpose.

Because the oversubscribed model does not offer any guarantees, it is possible that the protecting node in some cases runs out of resources and fails to offer protection. In this case, the protecting node generates an SNMP trap identifying the SRRP instance on which subscriber protection has failed. One SNMP trap is raised per SRRP instance in case where at least one subscriber under the corresponding group interface was not instantiated. The trap is cleared either when all subscribers become instantiated or when the SRRP transition into a non-master state.

The number of the subscriber hosts that failed to instantiate, can also be determined using the operational **show>redundancy>multi-chassis** all command. This command shows the number of subscribers that failed to instantiate along with SRRP instances on which the subscriber host are relaying for successful connectivity.

Pre-emption of already instantiated subscriber hosts in the protecting node by another subscriber hosts is not allowed.

10.2.2 Resource monitoring

Management and conservation of resources is of utmost importance in OMCR. The resources consumed by the subscriber host depend on the type and the size of subscriber parameters (the number of strings, length of strings, and so on).

For these reasons it is crucial that the operator has a view of the amount of memory in the CPM used by subscribers and the amount of free memory that can be used for additional subscribers. The MCS line is of interest in this output. In addition, the Subscriber Mgmt line shows memory utilization for active subscribers in the CPM.

The Available Memory line gives an indication about how much memory remains.

For example:

```
*A:right-21# show system memory-pools
===============================================================================
Memory Pools
===============================================================================
Name                Max Allowed    Current Size      Max So Far          In Use
-------------------------------------------------------------------------------
BFD                    No limit       6,291,456       6,291,456       5,509,872
BGP                    No limit       5,242,880       5,242,880       3,635,976
CFLOWD                 No limit       1,048,576       1,048,576          26,576
Cards & Ports          No limit      24,117,248      27,262,976      18,010,424
DHCP Server            No limit       2,097,152       2,097,152         173,680
ETH-CFM                No limit       6,291,456       9,437,184       4,016,128
ICC                 25,165,824       7,340,032      25,165,824       2,880,008
IGMP/MLD               No limit       1,048,576       1,048,576         166,216
IMSI Db Appl           No limit       1,048,576       1,048,576         793,984
IOM                    No limit       8,388,608       8,388,608       6,894,360
IP Stack               No limit      29,360,128      35,651,584      13,565,120
IS-IS                  No limit       2,097,152       2,097,152       1,095,360
ISA                    No limit       3,145,728       3,145,728       1,217,464
LDP                    No limit       6,291,456       6,291,456       5,607,240
Logging              411,041,792      6,291,456       6,291,456       3,473,024
MBUF               1,073,741,824      2,097,152       2,097,152         299,976
MCS                    No limit     454,033,408     454,033,408     416,753,472
MPLS/RSVP              No limit      49,283,072      69,206,016      42,947,776
MSCP                   No limit       2,097,152       2,097,152       1,022,848
MSDP                   No limit               0               0               0
Management             No limit      19,922,944      26,214,400       5,689,112
OAM                    No limit       1,048,576       1,048,576          86,080
OSPF                   No limit       8,388,608       8,388,608       4,975,824
```

```
OpenFlow               No limit       1,048,576       1,048,576         391,880
PIM                    No limit      19,922,944      19,922,944      15,755,792
PTP                    No limit       1,048,576       1,048,576           1,408
RIP                    No limit               0               0               0
RTM/Policies           No limit       9,437,184       9,437,184       7,002,648
Redundancy             No limit       9,437,184     424,673,280         703,160
SIM                    No limit       3,145,728      12,582,912             648
Services               No limit      25,165,824      25,165,824      18,128,056
Stats                  No limit       1,048,576       1,048,576           9,456
Subscriber Mgmt        No limit      24,117,248      41,943,040      14,846,512
System                 No limit     794,820,608     856,686,592     776,394,656
Traffic Eng            No limit       1,048,576       1,048,576         444,744
VRRP                   No limit       2,097,152       3,145,728         393,808
WEB Redirect          16,777,216      1,048,576       1,048,576         128,640
-------------------------------------------------------------------------------
Current Total Size :    1,540,358,144 bytes
Total In Use       :    1,373,041,928 bytes
Available Memory   :    5,778,702,336 bytes
===============================================================================
```

Similar output is shown for CPU utilization:

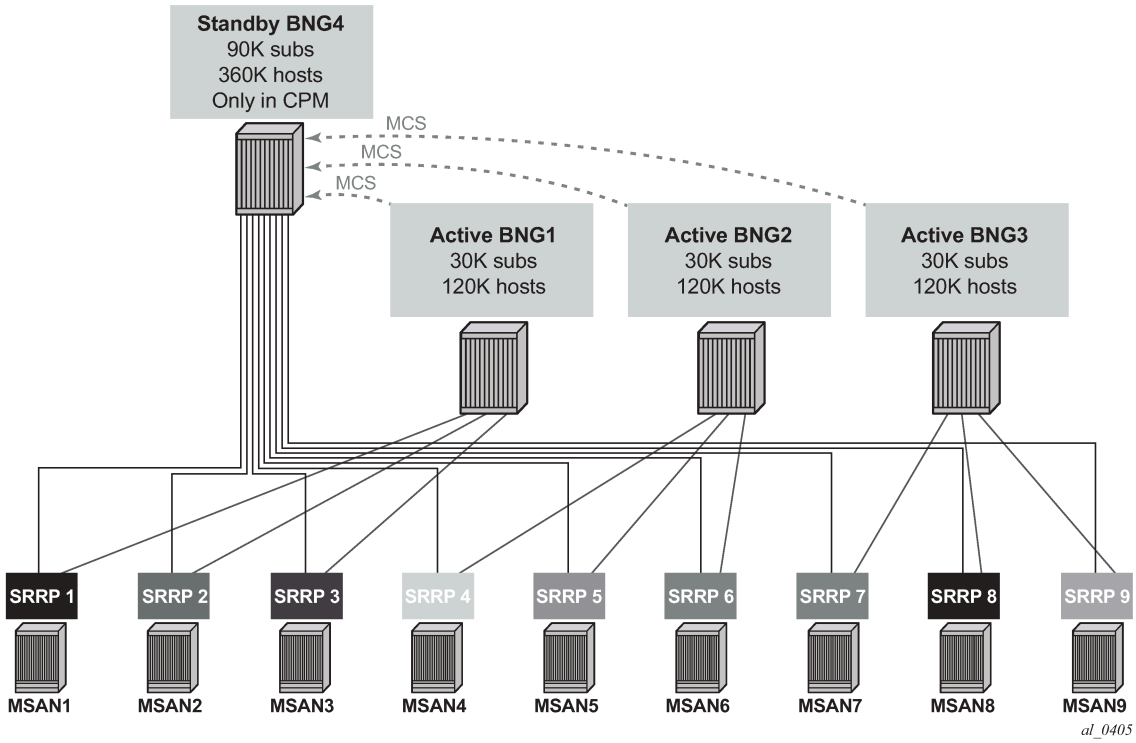
```
*A:right-21#  show system cpu
===============================================================================
CPU Utilization (Sample period: 1 second)
===============================================================================
Name                                    CPU Time       CPU Usage        Capacity
                                          (uSec)                          Usage
-------------------------------------------------------------------------------
BFD                                            0           0.00%           0.00%
BGP                                        2,504           0.03%           0.05%
BGP PE-CE                                      0           0.00%           0.00%
CFLOWD                                        25          ~0.00%          ~0.00%
Cards & Ports                             14,501           0.18%           0.13%
DHCP Server                                   30          ~0.00%          ~0.00%
ETH-CFM                                      614          ~0.00%           0.06%
ICC                                        1,803           0.02%           0.18%
IGMP/MLD                                     538          ~0.00%           0.05%
IMSI Db Appl                                  37          ~0.00%          ~0.00%
IOM                                            0           0.00%           0.00%
IP Stack                                   4,578           0.05%           0.24%
IS-IS                                        423          ~0.00%           0.02%
ISA                                        2,690           0.03%           0.10%
LDP                                           78          ~0.00%          ~0.00%
Logging                                       13          ~0.00%          ~0.00%
MBUF                                           0           0.00%           0.00%
MCS                                        2,718           0.03%           0.27%
MPLS/RSVP                                  1,137           0.01%           0.08%
MSCP                                           0           0.00%           0.00%
MSDP                                           0           0.00%           0.00%
Management                                 6,571           0.08%           0.19%
OAM                                        1,532           0.01%           0.09%
OSPF                                      18,397           0.23%           0.08%
OpenFlow                                      18          ~0.00%          ~0.00%
PIM                                            0           0.00%           0.00%
PTP                                           24          ~0.00%          ~0.00%
RIP                                            0           0.00%           0.00%
RTM/Policies                                   0           0.00%           0.00%
Redundancy                                 3,618           0.04%           0.19%
SIM                                       10,959           0.13%           1.08%
SNMP Daemon                                    0           0.00%           0.00%
Services                                   1,037           0.01%           0.03%
Stats                                          0           0.00%           0.00%
```

```
Subscriber Mgmt                          835          0.01%          0.03%
System                                29,863          0.37%          1.32%
Traffic Eng                                0          0.00%          0.00%
VRRP                                     970          0.01%          0.07%
WEB Redirect                              26         ~0.00%         ~0.00%
-------------------------------------------------------------------------------
Total                                7,975,383        100.00%
   Idle                              7,869,844         98.67%
   Usage                               105,539          1.32%
Busiest Core Utilization                33,264          3.33%
===============================================================================
*A:right-21#
```

10.2.3 Warm-standby mode of operation

The protecting node operates in a warm-standby mode. Warm-standby mode of operation is a property of the entire node. In other words, while in the central-standby mode of operation (warm-standby command), only subscribers under the SRRP instances that are in the master state is fully instantiated in the data plane on the central standby node (protecting node). All other subscribers (under the SRRP instances that are in the standby state) is synchronized only in the control plane. However, non-central standby nodes can have a peering connection with a protecting node (OMCR) and at the same time another peering connection with another active BNG node in active/active model. All nodes participating in the OMCR mode of operation must run SR OS Release 12.0 or higher. This model is shown in Figure 158: Network wide mixing of OMCR and active/active (1:1) model.

Figure 158: Network wide mixing of OMCR and active/active (1:1) model



The central backup property is configured with the following CLI:

```
configure
    redundancy
        multi-chassis
            peer 10.1.1.1
                warm-standby
```

The **warm-standby** keyword configures the chassis to be in the central standby mode of operation. Although the configuration option is configured per peer, the **warm-standby** functionality is applied per chassis.

Synchronization of IPoE subscribers (**config**>**redundancy**>**multi-chassis**>**peer**>**sync**>**sub-mgmt ipoe**) on the protecting node is only possible if all peers are configured for **warm-standby** or none are.

To transition from one mode to another (warm to hot), all peers must be administratively shutdown and the warm-standby keyword must be either removed or configured on all peers, depending on the direction of the transition.

Single-homed subscribers are supported in the central standby node, subject to resource limitations.

10.2.4 IPoE versus PPPoE

OMCR is supported only for IPoEv4/v6 subscribers. PPPoEv4/v6 subscriber hosts are not supported. However, non-synchronized PPPoE hosts can be hosted on the protecting node simultaneously with the protected IPoE subscribers. PPPoE PTA (locally terminated) non-synchronized subscribers and OMCR synchronized IPoE subscriber must not be configured under the same group-interfaces. On the other hand, non-synchronized PPPoE LAC sessions can be under the same group interface as the OMCR synchronized IPoE subscribers.

The recovery of PPPoE subscriber host in non-synchronized environment is based on the timeout of ppp-keepalives.

10.2.5 Persistency

Persistency in the multichassis environment must be disabled because redundant nodes are protecting each other and maintain up-to-date lease states. Otherwise, race conditions resulting in stale lease states caused by contention between MCS data and persistency data may occur.

10.2.6 Routing and redundant interface in OMCR

Support for redundant interface is limited and can be used only in cases where subscribers are activated in the protecting node. In other words, the shunting over the redundant interface cannot be used if subscriber hosts are not fully instantiated (in the data and control plane). For this reason, downstream traffic must not be attracted (via routing) to the protecting node while the subscriber hosts are in the standby mode (SRRP is in a backup state).

During the transient period while the switchover is in progress, subscriber hosts are being instantiated or withdrawn (depending on the direction of the switchover) in the data plane on the protecting node. The duration of this process is dependent on the number of the hosts that needs to be instantiated/withdrawn and it is proportional to the regular host setup/tear-down rates. The redundant interface in this case can be

used only for the hosts that are present in the data plane during the switchover transitioning period (from the moment that they are instantiated in the data plane when protecting node is assuming activity, or up the moment when they are withdrawn from the data plane when the protecting node is relinquishing activity).

The following routing models are supported:

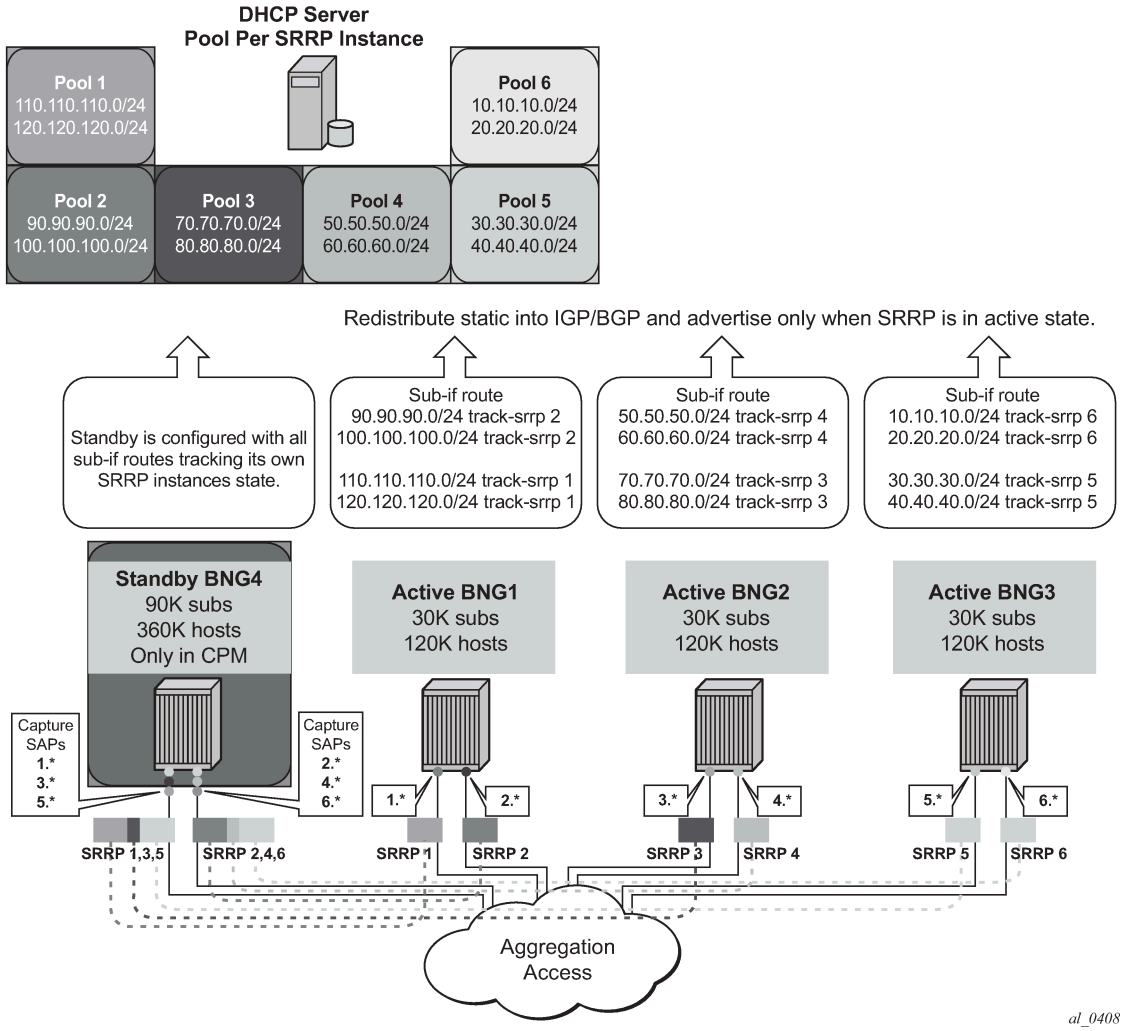
Case 1

SRRP-Aware routing where subnets can be assigned per group-interfaces (SRRP instances). In a steady state, the redundant interface is not needed because the downstream traffic is attracted to the active node (SRRP master state). During switchover periods (routing convergence transitioning periods), redundant interface can be used only for the subscriber hosts that are instantiated in the data plane (from the moment that they are instantiated in the data plane when protecting node is assuming activity, or up the moment when they are withdrawn from the data plane when the protecting node is relinquishing activity). See Figure 159: Subnet per group interface.

Case 2

SRRP-Aware routing where subnets spawn (are aggregated) over multiple group-interfaces (SRRP instances). In case of a switchover, /32 IPv4 addresses and /64 IPv6 addresses/prefixes are advertised from the protecting node. In a steady state, the redundant interface is not needed because the downstream traffic is attracted via more specific routes (/32s and /64s) to the active node (SRRP master state). During switchover periods (routing convergence transitioning periods), the redundant interface can be used only for the subscriber-hosts that are instantiated in the data plane (from the moment that they are instantiated in the data plane when protecting node is assuming activity, or up the moment when they are withdrawn from the data plane when the protecting node is relinquishing activity). To reduce the number of routes on the network side, /32s and /64s should only be activated on the protecting node.

Figure 159: Subnet per group interface



A deployment case that is not supported is the one where subnets spawn (are aggregated) over multiple group-interfaces (SRRP instances) and at the same time /32s are not allowed to be advertised from the protecting node. This scenario would require redundant interface support while subscriber-hosts are not necessarily instantiated in the protecting node.

10.2.7 Revertive behavior

In case that failure is repaired on the original active node (non-central standby node) while SRRP preemption (**preempt**) is configured, the corresponding active subscribers on the protecting node is withdrawn from the data plane and the activity (SRRP master state) is switched to the original node.

This behavior ensures that the resources in the central backup are freed upon failure restoration and are available for protection of other entities in the network (other links/nodes).

In the preemption case, the upstream traffic is steered toward the newly active BNG via gratuitous ARP (GARP). In other words, the virtual MAC is advertised from the newly active node, and consequently

the access and aggregation nodes update their Layer 2 forwarding entries. This action should cause no interruption in the upstream traffic.

In the downstream direction, the service interruption is equivalent to the time it takes to withdraw the routes from the network side on the standby node. In this case, there are two scenarios:

- A route per group interface (SRRP) is advertised in the network from the central standby node. In this case, downstream traffic interruption is a function of the convergence time of the routing protocol deployed on the network side. After the routing is converged, all downstream subscriber traffic is attracted to the newly active node. In the meantime, the redundant interface can be used to shunt traffic from the central standby node to the newly active node, but only for the subscriber hosts that have not yet been withdrawn from the data plane on the protecting node. This withdrawal process may take some time and therefore downstream traffic for some subscriber hosts is restored before the others during the routing convergence period.
- /32 subscriber-host routes are advertised from the protecting node. The total recovery time for downstream traffic depends on the routing convergence. The routing convergence may be slower than in the previous case because more routes (/32s) need to be withdrawn from the network. The redundant interface can be used in the meantime for the subscriber hosts that have not yet been withdrawn from the data plane in protecting node.

10.2.8 Service restoration times

Service restoration times depends on the scale of the outage. The factors that affect the restoration times are:

- Failure detection time based on SRRP (could be in a sub second range, also supported based on BFD).
- Time needed to instantiate/withdraw subscriber host in/from the data plane.
- Routing convergence (based on SRRP aware routing).

10.2.9 Processing of the SRRP flaps

When multiple SRRP instances fail at the same time, they is processed one at the time on first come first serve basis. The subscriber instantiation processing during the switchover is divided into 1seconds intervals. In-between those intervals, the state of the SRRP are checked to ensure that it has not changed while the subscriber instantiation is in progress. This mechanism breaks the inertia (snowball effect) that can be caused by SRRP instance flaps. Furthermore, an SRRP flap is handled by not requesting a withdrawal followed by an instantiation request for the same SRRP instance.

10.2.10 Accounting

The OMCR accounting follows the active/active (1:1) redundancy model.

One difference in accounting behavior between the OMCR model and 1:1 redundancy model is in the processing of the accounting session-time attribute which on the protecting node denotes the time when the host was instantiated on the protecting node.

In contrast, the session-time attribute in 1:1 redundancy model is recorded almost simultaneously on both BNG nodes at the time when the host is originally instantiated.

As a result, the session-time attribute is for the most part uninterrupted during the switchover in 1:1 model whereas in OMCR model, the session-time attribute is reset on the switchover to the protecting node.

10.2.11 Configuration guidelines

- For all protected SRRP instances, the protected node should be the preferred SRRP master. To achieve this, the SRRP priority should be higher in the protected node than in the protecting node. SRRP preemption is recommended in the protecting node to force it to be in the SRRP master state when possible.

  Note: An SRRP switch from non-master to master state in the protected node does not suffer the slow convergence observed when the non-master to master state transition takes place in the protecting node. This is because the protected node always has the hosts instantiated in the data plane.

- ARP hosts configuration is strongly discouraged in protected group interfaces, unless the operator is ready to tolerate an incomplete redundancy mechanism for these hosts.
- SRRP tracking is strongly recommended to expedite the routing convergence upon an SRRP transition from non-master to master state in the protecting node.
- It is recommended to have a 1:1 relationship between SRRP and subscriber subnets to have smooth routing advertisements based on SRRP state tracking.
- The use of M-SAPs should be preferred over the use of static SAPs. Static SAPs are supported in the OMCR mode but they are consuming resources in the protecting node even when the underlying SRRP instance is in a non-master state.
- It is recommended that the capture-sap configuration include the **track-srrp** statement (at least for the protecting node). With this configuration the CPM does not process trigger packets when the leases cannot be created because the SRRP is not in the master state. Configuring SRRP tracking at the capture-sap offloads the CPM from performing false authentication and MSAP creation attempts.
- Load balancing between active (SRRP master state) and standby (SRRP non-master state) via export policies for SRRP must not be configured as the hosts are not instantiated in the protecting node when the corresponding SRRP state is non-master.
- To minimize traffic impact in the event of node reboot, it is recommended to use **hold-time down ip** | **ipv6** *seconds* command under the subscriber-interface and allow enough time for the MCS database to reconcile. This is particularly important in the protected node. If the SRRP transitions to the master state in the protected node before the database has been reconciled, the protecting node removes the leases (SRRP non-master state) which have not been synchronized. This would create partial outage.
- To avoid SRRP collisions, lack of resources and partial subscriber host instantiation, the use of fate-sharing-groups is not recommended. If an SRRP instance can be served by the protected node, it is preferred to keep it in the SRRP master state in there, instead of switching it to the protecting node as part of the operation-group.

10.2.12 Troubleshooting commands

Some of the commands that can assist in troubleshooting are listed below.

Note: To get a summary view of SRRPs and their OMCR status use the following command as shown below (the **domain** concept is reserved for future use):

```
*A:right-21# show redundancy multi-chassis omcr all
===============================================================================
Domain Table
===============================================================================
Domain    Domain SRRP  SRRP    Domain Instan. Failed Failed Reason
name       state ID    State    Color  Failed  Hosts
-------------------------------------------------------------------------------
N/A       N/A    201   Standby N/A    not-act 0
N/A       N/A    202   Standby N/A    not-act 0
N/A       N/A    203   Standby N/A    not-act 0
N/A       N/A    204   Standby N/A    not-act 0
N/A       N/A    301   Standby N/A    not-act 0
N/A       N/A    302   Standby N/A    not-act 0
N/A       N/A    303   Standby N/A    not-act 0
N/A       N/A    304   Standby N/A    not-act 0
N/A       N/A    401   Standby N/A    not-act 0
N/A       N/A    402   Standby N/A    not-act 0
N/A       N/A    403   Standby N/A    not-act 0
N/A       N/A    404   Standby N/A    not-act 0
N/A       N/A    501   Standby N/A    not-act 0
N/A       N/A    601   Standby N/A    not-act 0
N/A       N/A    701   Standby N/A    not-act 0
N/A       N/A    801   Standby N/A    not-act 0
N/A       N/A    901   Standby N/A    not-act 0
N/A       N/A    1001  Standby N/A    not-act 0
N/A       N/A    1101  Standby N/A    not-act 0
-------------------------------------------------------------------------------
No. of Entries: 19
===============================================================================
*A:right-21#
```

Note: To obtain specific SRRP OMCR information, OMCR information has been added to the **show srrp x detail** command:

```
*A:right-21# show srrp 1001 detail
===============================================================================
SRRP Instance 1001
===============================================================================
Description         : (Not Specified)
Admin State         : Up                 Oper State       : backupRouting
Oper Flags          : subnetMismatch
Preempt             : yes                One GARP per SAP : no
Monitor Oper Group  : None
System IP           : 10.20.1.1
Service ID          : IES 2
Group If            : grp.Dut-J.1        MAC Address      : 00:00:61:ac:ac:0a
Grp If Description  : N/A
Grp If Admin State  : Up                 Grp If Oper State: Up
Subscriber If       : ies-sub-if-svc-2
Sub If Admin State  : Up                 Sub If Oper State: Up
Address             : 10.1.0.1/16       Gateway IP       : 10.1.0.3
Address             : 10.2.0.1/16       Gateway IP       : 10.2.0.3
Msg Path SAP        : 8/2/2:2.4094
Admin Gateway MAC   : 00:00:51:ac:0a:01  Oper Gateway MAC : 00:00:51:ac:0a:01
Config Priority     : 1                  In-use Priority  : 1
Master Priority     : 100
Keep-alive Interval : 10 deci-seconds    Master Since     : 02/11/2014 11:38:52
```

```
Master Down Interval: 3.000 sec (Expires in 2.700 sec)
Fib Population Mode : all
VRRP Policy 1       : None              VRRP Policy 2    : None
OMCR Client status  : Sub-mgmt-ipoe
Instantiation failed: not-act           Failed IPOE Hosts: 0
OMCR Reason         :
```

Note: To have a view of the MCS synchronization including OMCR standby records:

```
*A:right-21# show redundancy multi-chassis sync peer 10.20.1.6 detail
===============================================================================
Multi-chassis Peer Table
===============================================================================
Peer
-------------------------------------------------------------------------------
Peer IP Address         : 10.20.1.6
Description             : (Not Specified)
Authentication          : Disabled
Source IP Address       : 10.20.1.1
Admin State             : Enabled
Warm standby            : Yes
Remote warm standby     : No
-------------------------------------------------------------------------------
Sync-status
-------------------------------------------------------------------------------
Client Applications     : SUBMGMT-IPOE SRRP
Sync Admin State        : Up
Sync Oper State         : Up
Sync Oper Flags         :
DB Sync State           : inSync
Num Entries             : 64026
Lcl Deleted Entries     : 0
Alarm Entries           : 0
OMCR Standby Entries    : 64000
OMCR Alarm Entries      : 0
Rem Num Entries         : 64026
Rem Lcl Deleted Entries : 0
Rem Alarm Entries       : 0
Rem OMCR Standby Entries: 0
Rem OMCR Alarm Entries  : 0
===============================================================================
MCS Application Stats
===============================================================================
Application             : igmp
Num Entries             : 0
Lcl Deleted Entries     : 0
Alarm Entries           : 0
OMCR Standby Entries    : 0
OMCR Alarm Entries      : 0
-------------------------------------------------------------------------------
Rem Num Entries         : 0
Rem Lcl Deleted Entries : 0
Rem Alarm Entries       : 0
Rem OMCR Standby Entries: 0
Rem OMCR Alarm Entries  : 0
-------------------------------------------------------------------------------
Application             : igmpSnooping
Num Entries             : 0
Lcl Deleted Entries     : 0
Alarm Entries           : 0
OMCR Standby Entries    : 0
OMCR Alarm Entries      : 0
```

```
-------------------------------------------------------------------------------
Rem Num Entries         : 0
Rem Lcl Deleted Entries : 0
Rem Alarm Entries       : 0
Rem OMCR Standby Entries: 0
Rem OMCR Alarm Entries  : 0
-------------------------------------------------------------------------------
Application             : subMgmtIpoe
Num Entries             : 64000
Lcl Deleted Entries     : 0
Alarm Entries           : 0
OMCR Standby Entries    : 64000
OMCR Alarm Entries      : 0
-------------------------------------------------------------------------------
Rem Num Entries         : 64000
Rem Lcl Deleted Entries : 0
Rem Alarm Entries       : 0
Rem OMCR Standby Entries: 0
Rem OMCR Alarm Entries  : 0
-------------------------------------------------------------------------------
Application             : srrp
Num Entries             : 26
Lcl Deleted Entries     : 0
Alarm Entries           : 0
OMCR Standby Entries    : 0
OMCR Alarm Entries      : 0
-------------------------------------------------------------------------------
Rem Num Entries         : 26
Rem Lcl Deleted Entries : 0
Rem Alarm Entries       : 0
Rem OMCR Standby Entries: 0
Rem OMCR Alarm Entries  : 0
-------------------------------------------------------------------------------
Application             : mcRing
Num Entries             : 0
Lcl Deleted Entries     : 0
Alarm Entries           : 0
OMCR Standby Entries    : 0
OMCR Alarm Entries      : 0
-------------------------------------------------------------------------------
Rem Num Entries         : 0
Rem Lcl Deleted Entries : 0
Rem Alarm Entries       : 0
Rem OMCR Standby Entries: 0
Rem OMCR Alarm Entries  : 0
-------------------------------------------------------------------------------
Application             : mldSnooping
Num Entries             : 0
Lcl Deleted Entries     : 0
Alarm Entries           : 0
OMCR Standby Entries    : 0
OMCR Alarm Entries      : 0
-------------------------------------------------------------------------------
Rem Num Entries         : 0
Rem Lcl Deleted Entries : 0
Rem Alarm Entries       : 0
Rem OMCR Standby Entries: 0
Rem OMCR Alarm Entries  : 0
-------------------------------------------------------------------------------
Application             : dhcpServer
Num Entries             : 0
Lcl Deleted Entries     : 0
Alarm Entries           : 0
OMCR Standby Entries    : 0
```

```
OMCR Alarm Entries      : 0
-------------------------------------------------------------------------------
Rem Num Entries         : 0
Rem Lcl Deleted Entries : 0
Rem Alarm Entries       : 0
Rem OMCR Standby Entries: 0
Rem OMCR Alarm Entries  : 0
-------------------------------------------------------------------------------
Application             : subHostTrk
Num Entries             : 0
Lcl Deleted Entries     : 0
Alarm Entries           : 0
OMCR Standby Entries    : 0
OMCR Alarm Entries      : 0
-------------------------------------------------------------------------------
Rem Num Entries         : 0
Rem Lcl Deleted Entries : 0
Rem Alarm Entries       : 0
Rem OMCR Standby Entries: 0
Rem OMCR Alarm Entries  : 0
-------------------------------------------------------------------------------
Application             : subMgmtPppoe
Num Entries             : 0
Lcl Deleted Entries     : 0
Alarm Entries           : 0
OMCR Standby Entries    : 0
OMCR Alarm Entries      : 0
-------------------------------------------------------------------------------
Rem Num Entries         : 0
Rem Lcl Deleted Entries : 0
Rem Alarm Entries       : 0
Rem OMCR Standby Entries: 0
Rem OMCR Alarm Entries  : 0
-------------------------------------------------------------------------------
Application             : mcIpsec
Num Entries             : 0
Lcl Deleted Entries     : 0
Alarm Entries           : 0
OMCR Standby Entries    : 0
OMCR Alarm Entries      : 0
-------------------------------------------------------------------------------
Rem Num Entries         : 0
Rem Lcl Deleted Entries : 0
Rem Alarm Entries       : 0
Rem OMCR Standby Entries: 0
Rem OMCR Alarm Entries  : 0
-------------------------------------------------------------------------------
Application             : mld
Num Entries             : 0
Lcl Deleted Entries     : 0
Alarm Entries           : 0
OMCR Standby Entries    : 0
OMCR Alarm Entries      : 0
-------------------------------------------------------------------------------
Rem Num Entries         : 0
Rem Lcl Deleted Entries : 0
Rem Alarm Entries       : 0
Rem OMCR Standby Entries: 0
Rem OMCR Alarm Entries  : 0
-------------------------------------------------------------------------------
===============================================================================
===============================================================================
Ports synced on peer 10.20.1.6
===============================================================================
```

```
Port/Encap                    Tag
-------------------------------------------------------------------------------
4/2/2
  2.1-2.4094                  Dut-F.1
===============================================================================
===============================================================================
DHCP Server instances synced on peer 10.20.1.6
===============================================================================
Router-Name                       Server-Name
  Tag
-------------------------------------------------------------------------------
No instances found
===============================================================================
===============================================================================
*A:right-21#
```

Note: To have the MCS database view of the sync status including OMCR status use the following command syntax:

```
*A:right-21# tools dump redundancy multi-chassis sync-database application sub-mgmt-ipoe peer
 10.20.1.6
The following totals are for:
 peer ip 10.20.1.6, port/lag ALL, sync-tag ALL, application SUBMGMT-IPOE
Valid Entries:                   64000
Locally Deleted Entries:         0
Locally Deleted Alarmed Entries: 0
Pending Global Delete Entries:   0
Omcr Alarmed Entries:            0
Omcr Standby Entries:            64000
*A:right-21#
```

11 ESM on High Scale QoS IOM

11.1 Overview

This section describes High Scale QoS (HSQ) IOM as it applies to ESM. HSQ IOM features a high scale egress traffic manager with an expansive set of QoS capabilities that are particularly suitable for ESM applications. Refer to the *7450 ESS, 7750 SR, 7950 XRS, and VSR Quality of Service Guide* for detailed information about HSQ.

Subscriber queues on HSQ in the ESM context are always allocated in a set of eight queues per SLA profile instance (SPI). SPI has a particularly important role on HSQ that determines not only the scale of subscribers but also the subscriber's QoS modes as it pertains to shaping hierarchy. At this point it suffices to say that SPI is an instantiation of SLA profile and each subscriber host or a PPPoE or IPoE session must be associated with an SPI. An SLA-profile, with its reference to a QoS policy, indirectly defines queuing configuration along with classification for the subscriber host or session. Multiple subscriber hosts or sessions can share the same SPI. The number of SPIs per subscriber determines the SLA profile mode in which ESM on HSQ operates.

11.1.1 HSQ traffic manager overview

One of the principal functionalities of HSQ egress traffic manager are hierarchical rate limiting (with buffering or queuing) and scheduling functions. Rate limiters are implemented by a single or multiple threshold token buckets. The bucket thresholds define traffic burstiness. This implies that packets within each token bucket (rate limiter) are always sent at the full line rate, and only when one of the thresholds is reached (or bucket is filled), the object (queue, scheduling-class, aggregate-shaper) associated with the token bucket is stopped being served. The rate limiter allows traffic bursts (no traffic smoothing function), with a queue at the end of the shaping hierarchy. Because of the ability of the queue to buffer traffic, traffic limiters in this context are referred to as shapers.

The HSQ egress traffic manager supports a seven-tier shaping hierarchy with six levels of scheduling. Six levels of scheduling are realized via six scheduling classes that are served in a strict priority order by the port scheduler. Level six is the highest priority and one is the lowest. Scheduling classes should not be confused with a QoS policy driven forwarding class. Forwarding classes within the system are used between the ingress and egress forwarding complexes and help the system to map a packet to per-hop and per-domain behavior. Scheduling classes are slices of scheduling opportunity within a port scheduling context.

11.1.1.1 Shaping hierarchy

The following describes the seven tiers of hierarchical shaping on HSQ:

- First shaping tier:

  This is the rate control tier. At the bottom of the shaping hierarchy is a queue that represents the basic entity from which a packet is extracted and sent out of the port by the port scheduler. At the SPI level, a

queue can be attached either to a scheduler class or to a WRR group (but not to both simultaneously). Accordingly, the following shapers are supported at the bottom (first) tier:

  – Queue shaper — If a queue is directly attached to a scheduler class.
  – WRR group shaper – If a queue (possibly along with other queues) is attached to one of the two WRR groups at the SPI level. In this case, the queue shaper is ignored.

- Second shaping tier:

  This tier continues to operate at the SPI level. Eight queues form a queue set which is always allocated per SPI and the entire queue set can be shaped to an aggregate rate. In the ESM context, this tier is referred to as the SPI shaping tier. Depending on the SLA mode of operation, this tier can represent a subscriber aggregate rate (in a single SLA mode) or the aggregate rate for a subset of services within a subscriber (in an expanded SLA mode).

- Third shaping tier:

  Multiple SPIs within a single subscriber (supported only in expanded SLA mode) can be attached to a primary shaper which controls subscriber's aggregate rate. The aggregate rate of the primary shaper is the third shaping tier.

- Fourth shaping tier:

  This tier is enforced within the context of the secondary shaper. The secondary shaper represents an aggregation of subscribers, for example, a group of subscribers that are attached to the same aggregation node (DSLAM or OLT).

  Each scheduling class at the secondary shaper can be individually shaped.

- Fifth shaping tier:

  This tier is the aggregate rate of the secondary shaper.

- Sixth shaping tier:

  Sixth and seventh shaping tier are at the port level.

  At the port level, scheduling classes can be either:

  – individually shaped
  – collapsed into a single WRR group, in which the collection of scheduling classes (a WRR group) is shaped.
  – These two possibilities (individual or group shaping) are mutually exclusive by configuration, and whichever is chosen represents the sixth shaping tier.

- Seventh shaping tier:

  This tier is the aggregate shaping rate of the physical port.

11.1.1.2 Scheduling

Each port scheduler can be viewed as six separate schedulers, one for each scheduling class. At the port level, there are six strict priorities serviced exhaustively (priority level six is highest).

Scheduling classes at the port level and queues at the SPI level can be mapped to WRR groups. Within the WRR group, each WRR group member (a queue or a scheduling class) is assigned a weight that is used for relative scheduling importance of the queue or scheduler class to other active queues or scheduler classes at the same scheduling service level.

Scheduling opportunity at each tier within the scheduling class is governed by a weighted mechanism. By default, primary shaper members in secondary shaper lists and secondary shaper members in port scheduling class lists use a dynamic weighting function that increases a member's weight based on the amount of pending work associated with that member. This allows a primary shaper with 10,000 active queue sets (SPIs) to receive proportionately more scheduling opportunities at the secondary shaper level than another primary shaper with only 100 active queue sets (SPIs). The dynamic weight is based on actual activity and not simply based on number of potential (provisioned) membership.

A complete scheduling hierarchy must be maintained between each attached queue at the SPI level and its port priority. In other words, the HSQ has no provision to bypass the primary and secondary shaping and scheduling levels. Entities at the SPI level must be associated with a primary shaper and each primary shaper must be associated with a secondary shaper. All queues in the same queue set are mapped to the same primary shaper. All scheduling classes in the same primary shaper are mapped to the same secondary shaper. A default secondary shaper context is created for each port scheduler and a default primary shaper (this is called a managed bypass primary shaper) context is created for each secondary shaper context. Through normal provisioning events, other non-default secondary shapers and primary shapers can also be created.

11.1.2 HSQ and ESM SLA modes

HSQ supports two SLA modes for ESM, expanded and single. The SLA mode is configurable per subscriber profile.

In single mode, the subscriber aggregate rate and the SLA profile queuing are both performed using a single primary shaper context for the subscriber. This mode provides the highest subscriber scale.

In expanded mode, multiple SLA profile instances are supported by moving the subscriber aggregate shaping function to a primary shaper specifically created for the subscriber instance. The subscriber scale in this mode depends on the number of available primary shapers.

Both modes support routed and bridged homes (multiple subscriber hosts can be associated with a single SPI). However, multiple SPIs per subscribers support richer QoS granularity for bridged homes where a set of hosts using the same service can have their own queues and aggregate rate, separate from other services within the same home.

Figure 160: Single SPI mode and Figure 161: Expanded SLA mode display two ESM configurations that are designed to emphasize the distinction between the two SLA modes. These are examples to explore SLA modes with an arbitrary queue and WRR group assignments. For example, in both examples, the first four queues within the queue set (SPI) are assigned to two WRR groups at the SPI level, while scheduling classes three and four at the port level are collapsed into a single WRR group. Other configurations for queue, WRR group and scheduling classes are possible.

Both examples contain common parts generic to HSQ, independent of the SLA modes:

- Some of the queues in a queue set (SPI) are directly assigned to scheduling classes, and some are assigned indirectly through two WRR groups that are supported per each SPI on HSQ. Note that configuring WRR groups on HSQ is optional.
- On HSQ, only one queue or one WRR group can be mapped to a specific scheduling class. Queues either not mapped to a scheduling class or mapped to a WRR group that is not mapped to a scheduling class drops all frames incrementing the queue's discard counters.
- A queue's WRR weight is only used when the queue is attached to a WRR group. The WRR weight specifies the relative scheduling importance of the queue relative to other active queues on the same group. A queue's class attachment weight (CW) is only used when the queue is directly attached to a scheduler class. The CW of a WRR group specifies the relative scheduling importance of a queue or

WRR group at the scheduling class to queues or WRR groups belonging to other SPIs. Each individual queue or a WRR group can be shaped to a configurable rate.

- The secondary shaper in ESM represents an access node. A subscriber is associated with the secondary shaper by an *intermediate-destination-id* string which is supplied by RADIUS, LUDB, or Python during the authentication phase.
- At the port level, multiple scheduling classes can be collapsed through a WRR group (only one WRR group is supported on the port level) and, in this example, scheduling classes three and four are collapsed in WRR group one, with a single scheduling priority that is numerically between the scheduling priorities two and five. Essentially, the WRR group collapses the member scheduling priorities into a single scheduling priority using each class's WRR weight to manage scheduling opportunities between the classes. Because the member classes default priority levels are unused, the system simply maps the WRR group to one of the unused priorities (in this example, scheduling classes three or four). Because the unused priority levels are contiguous, and the remainder is unused, the priority chosen has no bearing on scheduler performance.

The following sections explore the two SLA modes focusing on the assignment of SPIs and the use of primary shapers as principal differentiators between the two SLA modes. Be aware that primary shapers on HSQ are not exposed as customer-defined objects and are explicitly created only when intermediate shaping is required by specific application (for example, as required by expanded SLA mode).

11.1.2.1 ESM single SLA mode

An example for single SLA mode is shown in Figure 160: Single SPI mode. In this mode, only a single SPI can be assigned to the subscriber. On the QoS level, the SPI represents a subscriber with the aggregate rate that is configured in the subscriber profile. In this mode of operation, the aggregate rate in the SLA profile should not be configured. If the aggregate rate is configured, the minimum of the rate configured in the subscriber profile and SLA profile are in effect for the SPI.

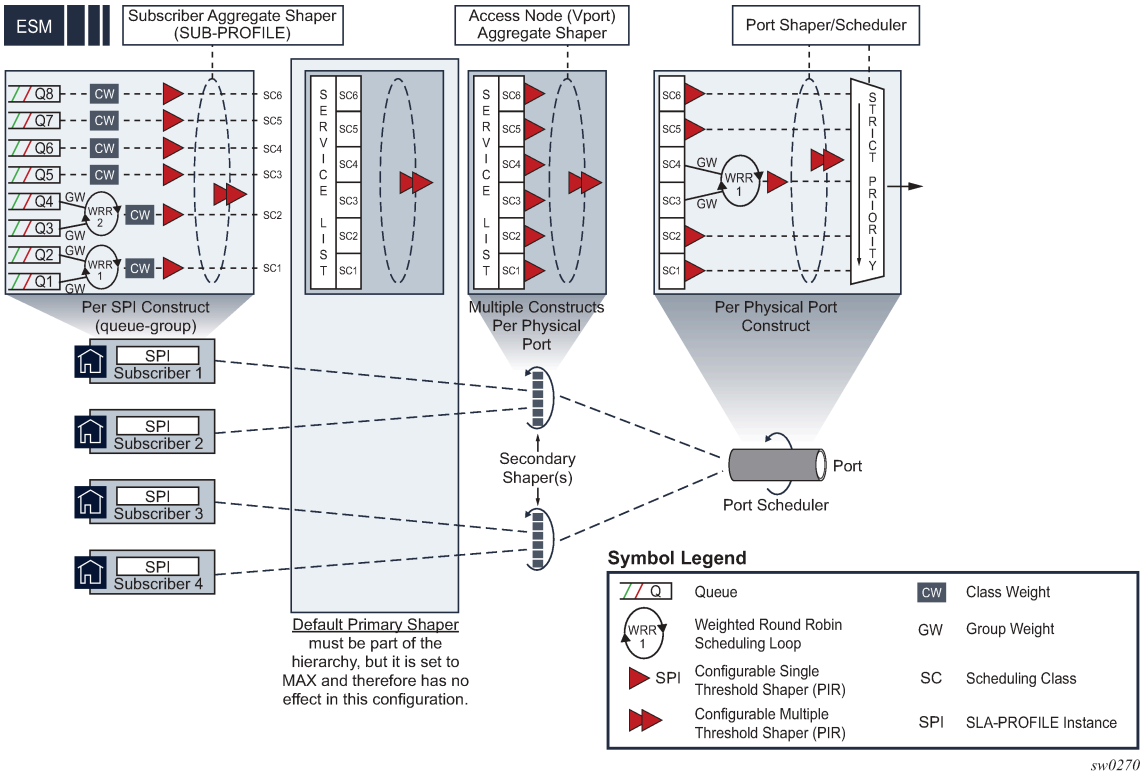
In single SLA mode, primary shapers have no functional role. However, because HSQ mandates the presence of the primary shaper in the hierarchy, the queue set under the SPI must be associated with a primary shaper construct. In such case, this enforced and futile primary shaper is implicitly created by the system. The number of primary shapers on an HSQ is finite, and consequently the number of implicitly created primary shapers should be tracked by the operator to ensure that the scaling limit of primary shapers in expanded SLA mode is not violated.

The system automatically creates the following primary shapers:

- one default primary shaper per physical port
- one managed bypass primary shaper per secondary shaper in single SLA mode

By default, the aggregate shaping rate of implicitly-created primary shapers is set to **max** which means that no rate limiting is performed at that level.

Figure 160: Single SPI mode



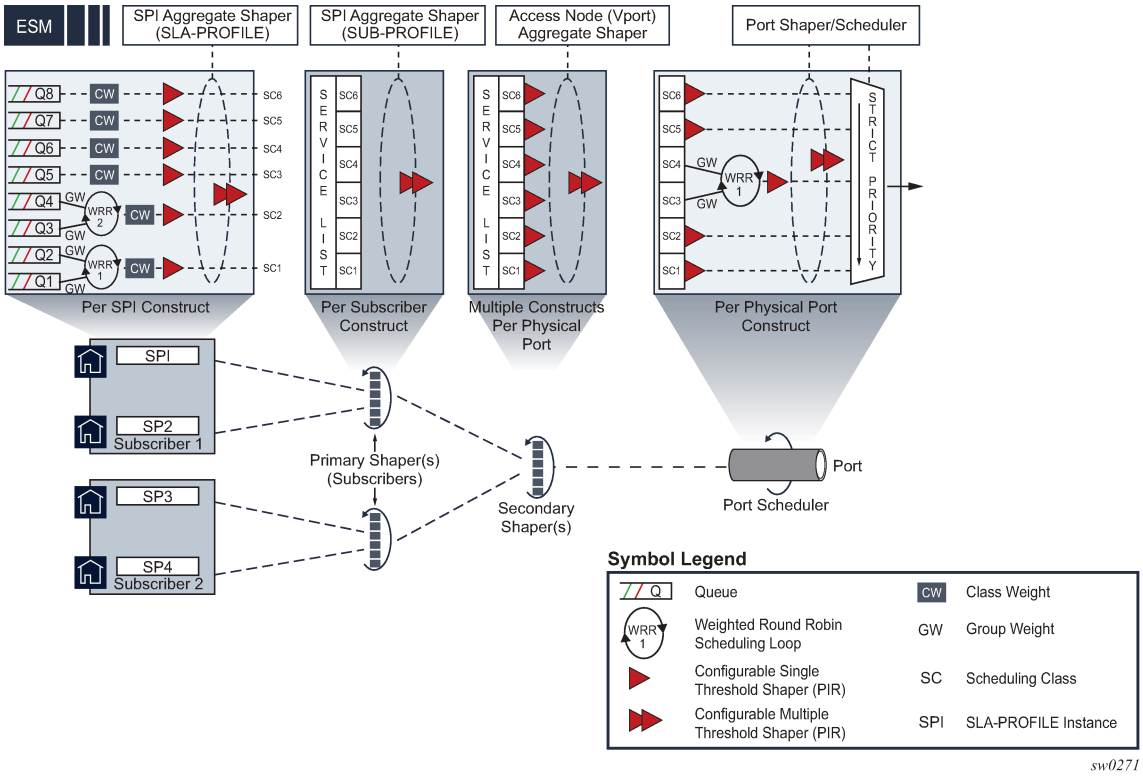
11.1.2.2 ESM expanded SLA mode

In expanded SLA mode, the primary shaper is used to limit the subscriber's aggregate rate. Multiple SPIs can be associated with a subscriber and each SPI can be individually rate limited. An example of this scenario is shown in Figure 161: Expanded SLA mode.

In this example, two SPIs are associated with the subscriber (only the details of the SPI1 for subscriber 1 are shown in Figure 161: Expanded SLA mode). Each SPI is shaped to an aggregate rate while the aggregate rate of the subscriber is shaped by the aggregate rate of the primary shaper. This scaling configuration is limited by the number of available primary shapers per line card. See the scaling figures in the SR OS Scaling Guides.

The access to the aggregate rate of the primary shaper through a subscriber profile in expanded SLA mode is an important distinction between the two modes of operation. In single SLA mode aggregate rate of the primary shaper is not accessible and is set to **max** (no rate limiting function is performed by the primary shaper).

Figure 161: Expanded SLA mode



11.1.3 Configuration steps

There are QoS configurations common to both SLA modes. The configuration difference is rooted in subscriber management-specific commands where the SLA mode and the aggregate rate of the primary shaper are configured.

The following are the common configuration QoS blocks for both SLA modes:

A **hs-scheduler-policy** defines parameters at the port level:

- – the maximum rate of the port scheduler
  - – the maximum rate of the WRR group (single WRR group is supported at this level)
  - – the mapping of scheduling classes to WRR group 1 and their weights within this group.
  - – the shaping rate of each scheduling class that is not mapped to a WRR group 1.

```
*A:BNG>config>qos>
   hs-scheduler-policy “hs1” create
   max-rate 10000
      group 1 rate 3000
      scheduling-class 1 rate 1000
      scheduling-class 2 rate 1000
      scheduling-class 3 group 1 weight 1
      scheduling-class 4 group 1 weight 1
      scheduling-class 5 rate 1000
```

```
      scheduling-class 6 rate 1000
```

- After it is defined, the **hs-scheduler-policy** is applied under the port. There is a single HSQ port scheduler per port.

```
*A:BNG>config>
    port 1/1/1
      ethernet
         egress
           hs-secondary-shaper "1" create
              aggregate
                rate 1000000
              exit
              class 1
                rate 10000
              exit
              class 2
                rate 10000
              exit
              class 3
                rate 10000
              exit
              class 4
                rate 10000
              exit
              class 5
                rate 10000
              exit
              class 6
                rate 10000
              exit
```

In this example, the **msap-policy** defines the default *inter-dest-id* to be the top Q-tag on the subscriber SAP. All subscriber hosts on a SAP with top Q-tag '1' are mapped to this secondary shaper.

```
*A:BNG>config>
   subscr-mgmt
      msap-policy "msaps" create
        sub-sla-mgmt
           def-inter-dest-id use-top-q
        exit
```

- An HSQ attachment policy defines how the eight individual queues within the SPI attach to either a scheduling class or to one of two WRR groups local to the SPI.

  The HSQ attachment policy also defines the maximum scheduling class (one through six) that uses the SPI aggregate shaper's low burst tolerance threshold. If a value of three is configured, classes one through three is associated with the low burst tolerance threshold while classes four through six are associated with the high burst threshold. In this example, if the SPI aggregate shaper becomes non-conforming, classes one through three shuts off first (associated queues removed from scheduler). This allows classes four to six to keep being served without buffering and therefore avoiding packet delays. If the rate of classes four through six is enough to cause the high burst tolerance threshold to be crossed, the remaining classes are also shutoff.

```
*A:BNG>config>qos
   hs-attachment-policy "attach-1" create
      low-burst-max-class 3
      queue 1 wrr-group 1
```

```
      queue 2 wrr-group 1
      queue 3 wrr-group 2
      queue 4 wrr-group 2
      queue 5 sched-class 3
      queue 6 sched-class 4
      queue 7 sched-class 5
      queue 8 sched-class 6
      wrr-group 1 sched-class 1
      wrr-group 2 sched-class 2
```

- The HSQ attachment policy described in the previous step is referenced in the SAP egress QoS policy that is associated with the subscriber host or session.

  The principal functionality provided by the SAP egress QoS policy is the following:

  – Forwarding class mappings that associate a forwarding class to one of the eight available egress SPI queues. A queue can service one, several, or all eight forwarding classes. A queue without a forwarding class mapping receives no user traffic.

  – WRR weight and class attachment weight parameters for each queue. A queue's WRR weight is only used when the queue is attached to a WRR group. In this configuration, the queues' WRR weight specifies the relative scheduling importance of the queue relative to other active queues in the same queue set. A queue's class attachment weight is only used when the queue is directly attached to a scheduler class.

  – Class attachment weights for the WRR groups. The class attachment weight specifies the relative scheduling importance of a queue or WRR group at the primary shaper compared to queues or WRR groups belonging to other SPIs.

  – Queue rate. The following is an example of queue rate configurations.

```
*A:BNG>config>qos>
sap-egress 10 create
-----------------------------------------------
            hs-attachment-policy "attach1"
            queue 1 create
                hs-wrr-weight 5
            exit
            queue 2 create
                hs-wrr-weight 10
            exit
            queue 3 create
                hs-wrr-weight 15
            exit
            queue 4 create
                hs-wrr-weight 20
            exit
            queue 5 create
            hs-class-weight 2
                rate 20000
            exit
            queue 6 create
                hs-class-weight 3
                rate 20000
            exit
            queue 7 create
                hs-class-weight 4
                rate 20000
            exit
            queue 8 create
                hs-class-weight 5
                rate 20000
```

```
            exit
                hs-wrr-group 1
           rate 10000
            hs-class-weight 8
            exit
hs-wrr-group 2
          rate 20000
                hs-class-weight 2
            exit
            fc af create
                queue 3
            exit
            fc be create
                queue 1
            exit
            fc ef create
                queue 7
            exit
            fc h1 create
                queue 5
            exit
            fc h2 create
                queue 6
            exit
            fc l1 create
                queue 4
            exit
            fc l2 create
                queue 2
            exit
            fc nc create
                queue 8
            exit
```

The following commands are specific to ESM:

SLA mode as configured in the subscriber profile:

```
*A:BNG>config>subscr-mgmt
   sub-profile “hs-sub” create
       hs-sla-mode {expanded | single}
```

Single SLA mode:

In single SLA mode, only the **sub-profile** *hs-aggregate-rate-limit* should be configured:

```
*A:BNG>config>subscr-mgmt
   sub-profile “hs-sub-single” create
       hs-sla-mode single
       egress
         hs-agg-rate-limit 20000
       exit
*A:BNG>config>subscr-mgmt
   sla-profile “hs-single” create
       ingress
          qos 5
          exit
       exit
       egress
          qos 5
          exit
       exit
```

When both, subscriber profile egress **hs-agg-rate-limit** and SLA profile **hs-agg-rate-limit** are configured, the system uses the minimum value to program the **hs-agg-rate** of the SPI shaper.

Expanded SLA mode:

In expanded mode, the **hs-agg-rate-limit** in the SLA profile determines the aggregate shaping rate of each SPI associated with the subscriber:

```
*A:BNG>config>subscr-mgmt>
   sla-profile “hs-expanded” create
       ingress
           qos 5
           exit
       exit
       egress
           qos 5
           exit
           hs-agg-rate-limit 1000
       exit
```

The aggregate rate of the subscriber (primary shaper) is configured in the subscriber profile:

```
*A:BNG>config>subscr-mgmt>
   sub-profile “hs-sub-expanded” create
       hs-sla-mode expanded
       egress
          hs-agg-rate-limit 20000
       exit
```

HSQ overrides:

```
*A:DUAL_HOME_2>config>subscr-mgmt>sla-prof# info
----------------------------------------------
            egress
                qos 2
                    queue 1
                        rate 100
                        mbs 1000 kilobytes
                        hs-class-weight 8
                        hs-wred-queue-policy "_tmnx_hs_default"
                        hs-wrr-weight 20
                    exit
                    queue 2
                        rate 100
                        mbs 1000 kilobytes
                        hs-class-weight 8
                        hs-wred-queue-policy "_tmnx_hs_default"
                        hs-wrr-weight 20
                    exit
                    queue 3
                        rate 100
                        mbs 1000 kilobytes
                        hs-class-weight 8
                        hs-wred-queue-policy "_tmnx_hs_default"
                        hs-wrr-weight 20
                    exit
                    queue 4
                        rate 100
                        mbs 1000 kilobytes
```

```
                        hs-class-weight 8
                        hs-wred-queue-policy "_tmnx_hs_default"
                        hs-wrr-weight 20
                    exit
                    queue 5
                        rate 100
                        mbs 1000 kilobytes
                        hs-class-weight 8
                        hs-wred-queue-policy "_tmnx_hs_default"
                        hs-wrr-weight 20
                    exit
                    queue 6
                        rate 100
                        mbs 1000 kilobytes
                        hs-class-weight 8
                        hs-wred-queue-policy "_tmnx_hs_default"
                        hs-wrr-weight 20
                    exit
                    queue 7
                        rate 100
                        mbs 1000 kilobytes
                        hs-class-weight 8
                        hs-wred-queue-policy "_tmnx_hs_default"
                        hs-wrr-weight 20
                    exit
                    queue 8
                        rate 100
                        mbs 1000 kilobytes
                        hs-class-weight 8
                        hs-wred-queue-policy "_tmnx_hs_default"
                        hs-wrr-weight 20
                    exit
                    hs-wrr-group 1
                        hs-class-weight 2
                        rate 100
                    exit
                    hs-wrr-group 2
                        hs-class-weight 4
                        rate 100
                    exit
                exit
    exit
-----------------------------------------------
```

11.1.4 Deployment considerations

For general restrictions for HSQ, see the HSQ Restrictions section in the *7450 ESS, 7750 SR, 7950 XRS, and VSR Quality of Service Guide*

Deploy ESM on HSQ with the following considerations:

- An on-line subscriber profile change (using CoA, RAR, re-authentication, or the **tools perform** command) for subscribers is allowed only for subscriber profiles with the same **hs-sla-modes**. In other words, an on-line change of the SLA mode for a subscriber on HSQ is not supported.
- Change of a SLA mode within the subscriber profile is allowed only for subscriber profiles that are not associated with subscribers.
- An on-line change of a SLA profile for a subscriber in single SLA mode with host-shared filters is not supported.

- An on-line change of a SLA profile for subscribers in single SLA mode with more than one non-session based hosts or more than one session, is not supported. When a single subscriber session with multiple hosts within the session where SLA change is allowed, all the hosts within the session are associated with the new SLA.
- An on-line change of an SLA profile for subscribers in single SLA mode with any static/l2 host is not supported.
- A single SLA mode supports a single SPI and a single SAP. Any attempt to create an additional SPI, or to add an SPI with a different SAP for the subscriber is rejected.
- Shared queueing on HSQ ingress is not supported and consequently the default **shared-queuing** setting in MSAP policy must be explicitly disabled.
- To achieve maximum scaling on ingress, it is recommended that policing is deployed on ingress. This implies that all FCs on ingress are explicitly mapped to ingress policers within the QoS policy and that **profiled-traffic-only** command in the **sap**>**sub-sla-mgmt** or **msap**>**sub-sla-mgmt** context under is enabled.

12 Wi-Fi aggregation and offload

12.1 Wi-Fi aggregation and offload overview

This solution set adds support for managing subscribers gaining network access over WLAN. The WLAN access enables a service provider to offer a mobile broadband service to its subscribers or to offload traffic on its or a partner's macro cellular (3G/4G) network. The WLAN access can be from public hot-spots (indoor or outdoor APs), venues, enterprises, or home-spots (with public SSID).

The 7750 SR serves as a WLAN Gateway (WLAN-GW) providing Layer 3 termination and ESM for these subscribers. The connectivity from WLAN AP or AC can be over any existing access technology (DSL, PON, Fiber, DOCSIS, and so on), with Ethernet based connectivity from the access node (DSLAM, OLT, Eth MTU, Layer 2 CMTS) to the WLAN-GW. WLAN-GW functions could be on a standalone 7750 as shown in Figure 162: Standalone WLAN-GW or could be an add-on functionality on existing 7750 based BNG as shown in Figure 163: WLAN-GW functions on existing BNG. WLAN connectivity to the WLAN-GW could be over a Layer 2 aggregation or an Layer 3 aggregation network (typical when WLAN-GW is upstream of an existing BNG or CMTS). In case of Layer 2 aggregation the connectivity to the WLAN-GW could be tagged or untagged Ethernet. In case of Layer 3 aggregation, supported connectivity option is Ethernet over GRE (or Eth-over-MPLS over GRE) tunnel originating from the AP/AC, and terminating on the WLAN-GW. The WLAN AP acts as a bridge, switching Ethernet frames into a GRE tunnel terminating on an MS-ISA in the WLAN-GW.

Figure 162: Standalone WLAN-GW

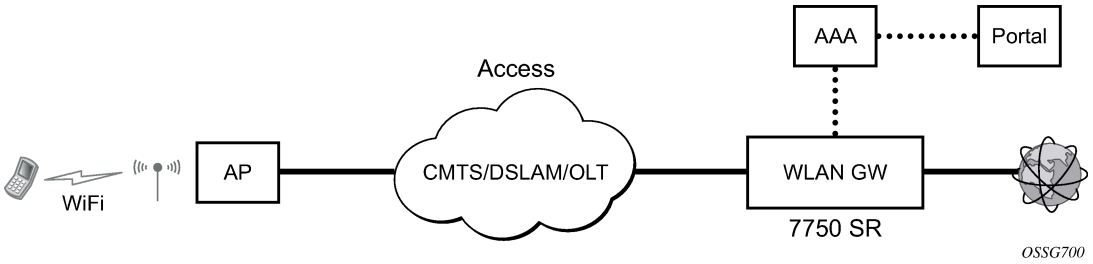
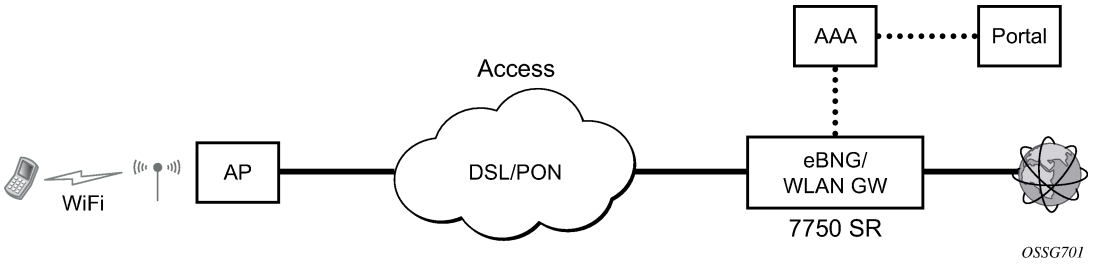


Figure 163: WLAN-GW functions on existing BNG



AP Connectivity to the WLAN-GW could be direct Ethernet (tagged or untagged) or could be Ethernet over GRE. With the bridged AP using GRE tunnels, the WLAN-GW solution elements are discussed in the following sections.

12.2 WLAN-GW group

To operate a Nokia Wi-Fi Gateway, a WLAN-GW group must be provisioned. This group contains a list of either Integrated Services Adapters (ISAs) or Extended Services Appliances (ESAs). A mix of both is not supported. The WLAN-GW functionality described in this chapter is executed on these ISAs or ESAs. WLAN-GW tunnels and sessions are automatically load-balanced over ESAs or ISAs to maximize resource usage.

ISAs and ESAs in a WLAN-GW group can be configured for NAT functions also. A WLAN-GW group can be used in any place where a NAT group is expected, but a NAT group cannot be used as a WLAN-GW group.

For resiliency, standby ESAs and ISAs can be provisioned. ISAs can be provisioned in either an IOM resiliency mode or MDA redundancy mode, where one MDA equals one ISA. ESAs can only be configured in MDA redundancy mode, but MDA refers to the ESA VM itself, not the MDA on which the ESA VM is connected.

All descriptions in the remainder of this chapter apply to both ISA and ESA, even when only the terms ISA and MS-ISA are used. Consult the *SR OS Release Notes* for an overview of which functionality is supported on each platform.

For more information about ISA and ESA, see the *7450 ESS, 7750 SR, and VSR Multiservice ISA and ESA Guide*.

12.2.1 IOM-based resiliency

The WLAN-GW supports N:M warm standby redundancy on an IOM level. One WLAN-GW group can be configured with a set of WLAN-GW IOMs, and a limit of active IOMs. Each WLAN-GW IOM can only contain ISA hardware with the ISA BB image loaded. Any combination with other hardware or ISAs with another image is not supported and MDA-based redundancy must be used instead.

In IOM resiliency mode, all ISAs on an IOM are either active or standby and the maximum number of active resources is also configured at the IOM level using the **active-iom-limit** command. IOMs beyond the active limit act as warm standby and take over the tunnel termination and session management functions from a failed IOM. It is not possible to change the **active-iom-limit** while a system is running.

12.2.2 MDA-based redundancy

The WLAN-GW supports N:M warm standby redundancy on an MDA level. With this feature, it is possible to configure separate MDAs in the WLAN-GW group without the restriction that all MDAs in an IOM be BB modules. Up to 14 BB modules can be configured and active at one time.

MDA-based redundancy makes hardware provisioning more flexible but does not guarantee that all ESM session states is recovered after an MDA failure. Because BB MDAs share IOM resources with another MDA in the same IOM, there is no guarantee that the same set of resources is available on the IOM of the new MDA after a failure. This restriction does not apply to DSM.

There are several ways to ensure that the system has adequate IOM resources following a failure.

- Use symmetric provisioning over all IOMs with BB modules. For example, do not create a 2:3 active:standby group by putting two BB MDAs on one IOM (IOM A) and one BB MDA on another IOM (IOM B). When the active ISAs are spread over IOM A and IOM B, the WLAN-GW has the full

resources of two IOMs. However, after a failure, it is possible that both active ISAs is on IOM A and limited to the resources of a single IOM.

- Combine BB MDAs with MDAs and services using predictable and fixed resources. For example, combining BB functionality with AA modules provides enhanced traffic services because AA modules typically have a fixed resource usage that allows you to predict resources following a failure.
- Create a safety buffer by leaving some resources unallocated. By not maximizing resource usage on a single IOM, it is more likely those resources are also available on a backup location.

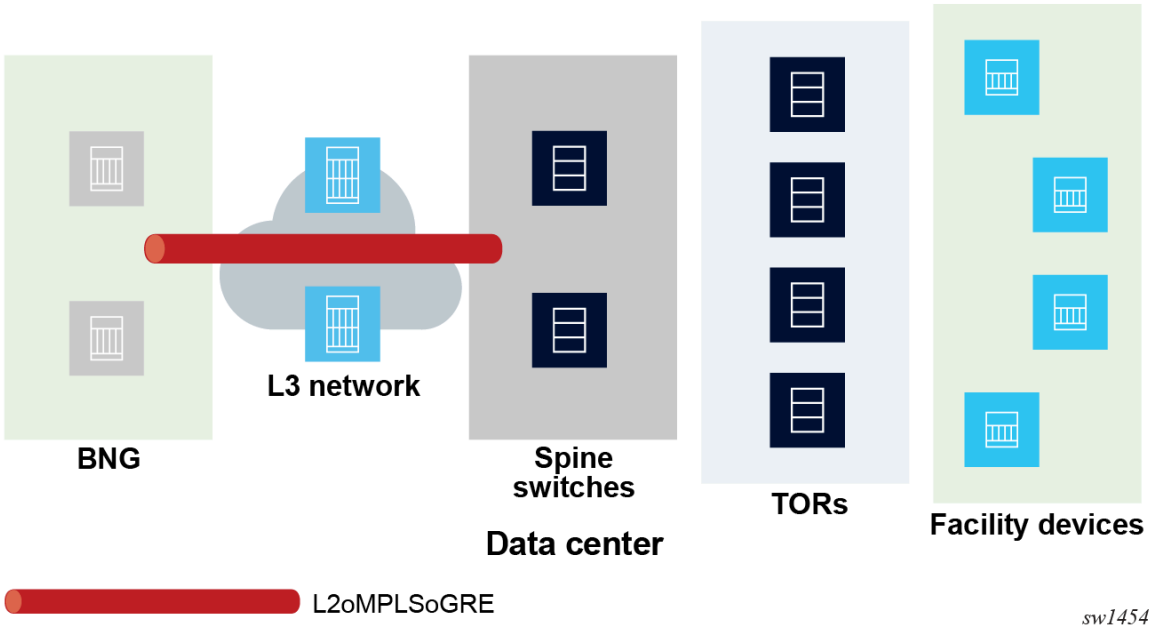
For more details, contact a Nokia representative.

12.3 ESM over soft-GRE for facility management devices

ESM over soft-GRE for facility management devices allows BNG to be used at the edge of a data center (DC) to provide subscriber management for facility management devices in the DC. Traffic related to discovery and communication between devices includes device-to-device IPv4/IPv6 unicast, IPv4 subnet-directed broadcasts, IPv4 network broadcasts, IPv4/IPv6 link-local multicast, and IPv6 solicited-node multicast traffic.

By default, the traffic originating from these devices is destined for other devices on the same network in the DC, not to the BNG. However, this traffic is not locally switched by access switches and instead forwarded to the BNG over L2oMPLSoGRE soft tunnels (see Encapsulation for more information.)

Figure 164: L2oMPLSoGRE soft tunnels



The BNG provides ESM over soft-GRE tunnels, including RADIUS authentication, IPv4/IPv6 filters, statistics, and LI.

Devices may have static IPv4/IPv6 addresses or get an initial address using DHCPv4/v6. Because all device-to-device traffic goes through the BNG, proxy-ARP must be enabled on the group interface. IPv6

proxy-ND is not supported. Therefore, IPv6 prefixes advertised in router advertisements are sent with the “on-link” bit reset.

The TOR switches aggregating upstream Ethernet traffic from the facility devices insert a Dot1q tag corresponding to an access port on the TOR. The TOR switch uses the Dot1q tag to identify the access port for forwarding downstream traffic to its connected devices. The spine switches insert a static VC-label and tunnel the L2 frame over a GRE tunnel terminating on the BNG. On an SR OS-based spine switch, this is done by configuring a GRE spoke SDP for each connected TOR, where the GRE SDP far-end IP address terminates on the BNG and corresponds to a configured soft-GRE tunnel endpoint on the BNG.

The BNG reflects the VC-label and .1q tag in the downstream-tunneled traffic toward the spine switch.

To process and perform ESM on any traffic received on the BNG that is not destined for the MAC or vMAC of the group interface, use one of the following commands.

```
configure service ies subscriber-interface group-interface wlan-gw promiscuous-mode
configure service vprn subscriber-interface group-interface wlan-gw promiscuous-mode
```

For devices that have static addresses, this feature relies on IPv4/IPv6 data-triggered, ARP-triggered, and NS-triggered ESM hosts over soft-GRE. The data-trigger host creation also includes traffic that is destined for multicast and broadcast addresses. For IPv6 traffic, the source address needs to be a global unicast address (GUA) for ESM host creation. The forwarding behavior of packets destined for multicast or broadcast addresses is as follows:

- Packets received from an ESM host that are destined for an IPv4 subnet-broadcast address or an IPv4 link-local multicast or an IPv6 link-local multicast or IPv6 solicited-node multicast address are replicated to all hosts that have an address from the same subnet as the sender, as configured on a subscriber interface.
- Traffic destined for IPv4 or IPv6 link-local multicast addresses corresponding to protocols that are relevant for local processing on the BNG as the first L3 hop (for example, RS, DHCPv6, OSPFv3, or PIM) are also forwarded to the CPM.
- Traffic from an ESM host destined for an IPv4 network broadcast address is forwarded to all hosts that are in the same service as the sender.

A RADIUS VSA (*Alc-Bcast-LL-Mcast-Replication)* in the Access-Accept message enables the host to act as either a sender or a receiver of the multicast and broadcast traffic replicated on the BNG. The VSA is defined in the 7450 ESS, 7750 SR, and VSR RADIUS Attributes Reference Guide. COA for this VSA is not supported.

Because each host is reachable over an L2oMPLSoGRE tunnel, BNG performs per-host replication. A copy of the packet destined for the receiving host is encapsulated over the corresponding L2oMPLSoGRE tunnel. The destination MAC of the replicated packet is the host unicast MAC address. The .1q tag and VC-label (as received in upstream traffic from the host) are reflected in the downstream forwarded packet.

ESM over soft-GRE supports 1:1 stateful gateway redundancy based on stateful synchronization of ESM hosts over L2oGRE/L2oMPLSoGRE soft tunnels between pairs of BNGs. Host state and tunnel encapsulation (including tunnel IP endpoints, VC-label, and .1q tag) are synchronized between CPMs of the active and backup BNG using the multichassis synchronization (MCS) protocol. An MCS peer must be configured using the following command.

```
configure service vprn subscriber-interface group-interface wlan-gw mcs-peer
```

Ensure the WLAN-GW node is administratively disabled before changing the peer address.

ESM over soft-GRE supports IPv4 and IPv6 L2oGRE tunnels. The two BNGs must have same number of tunnel ESAs in the WLAN-GW group. The backup BNG populates the host MAC and corresponding

tunnel encapsulation to the correct ESAs, where the packet from the host MAC is processed. The active and backup BNG determination and switchover uses existing WLAN-GW redundancy based on supported monitor and export route concepts as described in WLAN-GW 1:1 active-backup redundancy.

12.4 Layer 2 over soft-GRE tunnels

Soft-GRE refers to stateless GRE tunneling, whereby the AP forwards GRE encapsulated traffic to the WLAN-GW, and the gateway (GW) reflects the encapsulation in the downstream traffic toward the AP. WLAN-GW does not require any per-AP end-point IP address configuration. The WLAN-GW learns the encapsulation as part of creating the subscriber state on processing the encapsulated control and data traffic. Following are some of the advantages of soft-GRE:

- Resources are only consumed on the WLAN-GW if there is one or more active subscriber on the AP. Merely broadcasting an SSID from an AP does not result in any state on the WLAN-GW.
- No per-AP tunnel end-point configuration on WLAN-GW. This is important as the AP can get renumbered.
- No control protocol to setup and maintain tunnel state on WLAN-GW.

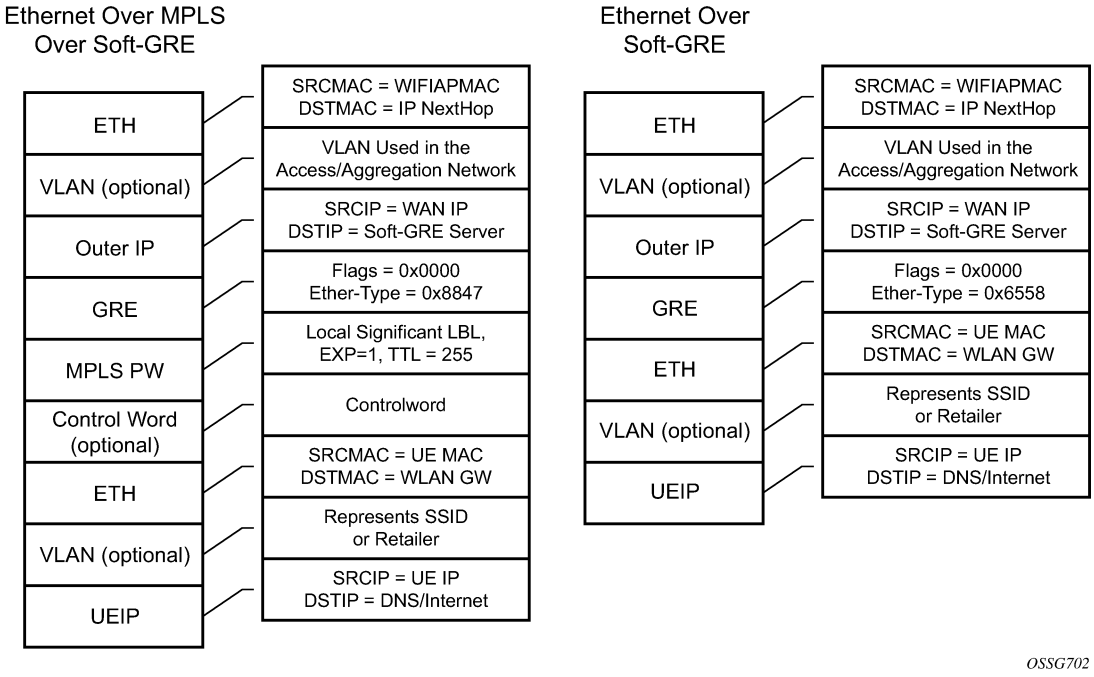
Soft-GRE tunnel termination is performed on dedicated MS-ISAs. MS-ISA provides tunnel encapsulation/decapsulation, bandwidth shaping per tunnel (or per-tunnel per SSID), and anchor point for inter-AP mobility. The ESM function such as per-subscriber anti-spoofing (IP and MAC), filters, hierarchical policing, and lawful intercept are provided on the carrier IOM corresponding to the ISA where the subscriber is anchored.

By default WLAN-GW uses IOM level resiliency which requires dedicated IOMs with only BB ISAs, these are also referred to as WLAN-GW IOMs. In this mode a single ISA failure causes a full backup IOM to take over, independent of the state of other ISAs. This mode is recommended in combination with ESM as it provides guaranteed resource recovery in failure cases. The WLAN-GW also supports MDA level resiliency as indicated in MDA-based redundancy.

12.4.1 Encapsulation

The GRE encapsulation is based on RFC 1701/2784, *Generic Routing Encapsulation (GRE)*, WLAN-GW encapsulates according to RFC 1701 with all the flag fields set to 0, and no optional fields present. WLAN-GW can receive both encapsulations specified in RFC 1701 and RFC 2784, with all flag fields set to 0, and no optional fields present in the header.

Figure 165: Encapsulation example



The encapsulation is built as follows:

- Outer Ethernet header: (14 bytes)
  - Source MAC: MAC address of the Wi-Fi AP/RG/HGW HW address
  - Destination MAC: MAC address of the first IP NH the Wi-Fi AP/RG/HGW is connected to (for example, CMTS, IP aggregation router, BNG, and so on)
- Outer VLAN: (4 bytes): optional, typically used for service delineation in the access or aggregation network.
- Outer IPv4 Header: (20 bytes)
  - Source IP — IP address used for WAN addressing which is retrieved by the AP/RG from the ISP through DHCP, PPPoX, and so on
  - Destination IP — Soft-GRE server address which can be retrieved by a DHCP Option, PPPoX option or configured by TR69 or configured statically in a boot file (in cable environment).
  - DSCP — Reflects QoS used in the access/aggregation network.
  - TTL — Should be set to 255 or should reflect the amount of IP hops in the access/aggregation network
- GRE: (4 bytes)
  - All flags are set to 0, such as checksum, sequence number and keys are not present.
  - The Ether-Type is set to 0x6558 for native Ethernet is used, and 0x8847 when MPLS encapsulation is used.
- MPLS Pseudowire Label (4 bytes)

  - Label Value, statically assigned in the Wi-Fi AP/Controller and reflected from the soft-GRE server to the Wi-Fi AP/Controller. The Label is unique within the context of the source IP address of the tunnel.
  - EXP: 0 (not used)
  - TTL: 255 (not used)
- Inner Ethernet header: (14 bytes)
  - Source MAC: MAC address of the UE
  - Destination MAC: MAC address of the soft-GRE server/WLAN-GW.
- Inner VLAN: (4 bytes): optional, inserted by AP/RG per unique SSID (typically, when the AP is providing SSID per retailer). WLAN-GW allows mapping the VLAN to a service context per retailer, in the data plane.
- Inner IPv4 Header: (20 bytes)
  - Source IP: Client's IP address obtained via DHCP (tunneled).
  - Destination IP: IP address of the destination client trying to reach.
  - DSCP: set by the client/application
  - TTL: set by the client/application

Soft-GRE tunnel termination is performed on dedicated IOMs with MS-ISAs (referred to as WLAN-GW IOM). Each WLAN-GW IOM requires both MS-ISAs to be plugged in for soft-GRE tunnel termination. MS-ISA provides tunnel encapsulation/decapsulation and anchor point for inter-AP mobility. The carrier IOMs of the ISA where the tunnel is terminated performs bandwidth shaping per tunnel (or per-tunnel per SSID). ESM function such as per-subscriber anti-spoofing (IP and MAC), filters, hierarchical policing, and lawful intercept are provided on the carrier IOM corresponding to the ISA where the subscriber is anchored.

An ESM and soft-gre configuration is required for wlan-gw functions. Subscriber and group interfaces are configured as part of normal ESM configuration. The group interface is enabled for wlan-gw by configuration. L2oGRE is the currently supported soft tunnel types. The wlan-gw related configuration includes the following:

- Tunnel end-point IP address.
- Service context for tunnel termination.
- TCP MSS segment size. This is set in TCP SYN and SYN-ACKs by WLAN-GW to adjust to the MTU on access/aggregation network to prevent fragmentation of upstream and downstream TCP packets.
- Mobility related configuration, including mobility trigger packet types (normal data or special Ethernet IAPP fame), and hold-down time between successive mobility triggers.
- VLAN to retailer mapping. The AP typically inserts a unique dot1Q tag per retail service provider in the Ethernet payload. The mapping of dot1Q tag to retail service context is configured under wlan-gw tunnel. The subscriber is then created in the configured retail service context. The retail service context can also be provided by AAA server in authentication-accept message based on subscriber credentials or SSID information contained in DHCP Option82.
- Egress QoS configuration for downstream traffic entering the wlan-gw module for tunnel encapsulation. This includes type of aggregate bandwidth shaping (per-tunnel or per-retailer), aggregate-rate-limit, egress QoS policy and scheduler policy. The tunnel shaping can be configured to be applied only when

there is more than one subscriber on the tunnel. By default the shaping if configured is applied when first subscriber on the tunnel logs in.

```
*B:Dut-C>config>service>vprn>sub-if>grp-if>wlan-gw# info detail
----------------------------------------------
                        authentication
                            no authentication-policy
                            hold-time sec 5
                        exit
                        no data-triggered-ue-creation
                        dhcp
                            shutdown
                            active-lease-time min 10
                            initial-lease-time min 10
                            no l2-aware-ip-address
                            no primary-dns
                            no primary-nbns
                            no secondary-dns
                            no secondary-nbns
                        exit
                        egress
                            no agg-rate-limit
                            no hold-time
                            qos 1
                            no scheduler-policy
                            no shape-multi-client-only
                            no shaping
                        exit
                        gw-addresses
                            address 10.1.1.4
                        exit
                        no http-redirect-policy
                        no nat-policy
                        mobility
                            hold-time 5
                            no trigger
                        exit
                        router 70
                        no tcp-mss-adjust
                        track-mobility
                            mac-format "aa:"
                            no radius-proxy-cache
                        exit
                        wlan-gw-group 3
                        vlan-tag-ranges
                            range start 0 end 100
                                authentication
                                    no authentication-policy
                                    hold-time sec 5
                                exit
                                no data-triggered-ue-creation
                                dhcp
                                    shutdown
                                    active-lease-time min 10
                                    initial-lease-time min 10
                                    no l2-aware-ip-address
                                    no primary-dns
                                    no primary-nbns
                                    no secondary-dns
                                    no secondary-nbns
                                exit
                                no http-redirect-policy
```

```
                                no nat-policy
                                retail-svc-id 35
                                track-mobility
                                    mac-format "aa:"
                                    no radius-proxy-cache
                                exit
                            exit
                        exit
                        no shutdown
```

12.4.2 Data path

In the upstream direction, the ingress IOM receiving the GRE tunneled packets from the Wi-Fi AP or AC, load-balances tunnel processing amongst the set of MS-ISAs on the active WLAN-GW IOMs in the WLAN-GW group. The load-balancing is based on a hash of source IP address in the outer IP header. The MS-ISA receiving the GRE encapsulated packets removes the tunnel encapsulation, and internally tunnels (MAC-in-MAC, using BVPLS) the packet to an anchor MS-ISA on the WLAN-GW IOM. All traffic from a specific UE is always forwarded to the same anchor MS-ISA based on hashing on UE's MAC address. The MS-ISA provides a mobility anchor point for the UE. The UE MAC's association to the GRE tunnel identifier is created or updated. The corresponding IOM provides ESM functions including ESM lookup, ingress ACLs and QoS. DHCP packets are forwarded to the CPM from the anchor IOM.

In the downstream direction, the IP packets are forwarded as normal from the network IOM (based on route lookup yielding subscriber subnet) to the IOM where the ESM host is anchored. ESM processing including per UE hierarchical policing and LI is performed on the anchor IOM. Configured MTU on the group-interface is enforced on the IOM, and if required packets are fragmented. The packets are then forwarded to the appropriate anchor MS-ISA housed by this IOM. Lookup based on UE's MAC address is performed to get the tunnel identification, and the packets are MAC-in-MAC tunneled to the MS-ISA terminating the GRE tunnel. Aggregate shaping on the tunneled traffic (per tunnel or per retailer) is performed on the carrier IOM housing the tunnel termination MS-ISA. The tunnel termination MS-ISA removes MAC-in-MAC encapsulation, and GRE encapsulates the Layer 2 packet, which exits on the Layer 3 SAP to the carrier IOM. The GRE tunneled packet is forwarded to the right access IOM toward the Wi-Fi AP-based on a routing lookup on IP DA in the outer header.

12.5 Wi-Fi SSIDs and VLAN ranges

A commonly supported feature of WLAN access points is to map a single SSID to an Ethernet VLAN tag. If the access point supports multiple simultaneous bands for the same SSID (for example, 2.4Ghz and 5Ghz) it can map these onto distinct VLANs. The Nokia WLAN-GW supports provisioning of distinct per-SSID parameters on the VLAN tag ranges. It may be desirable for example, to have different address pools or retail service IDs for different SSIDs.

In some cases, VLANs assigned to a single SSID may not be continuous and configuring them in one large range it would cause overlap with another SSID. For example, this can happen if a deployment starts with only one band for each SSID and later adds another band, creating a mapping as follows:

- SSID premium, 2.4Ghz Channel 6 to VLAN 10
- SSID basic, 2.4Ghz Channel 8 to VLAN 11
- SSID Premium, 5Ghz Channel 38 to VLAN 20
- SSID basic, 5Ghz Channel 48 to VLAN 21

To support this case, the WLAN-GW supports configuration of VLAN range extensions under each VLAN range. Functionally, these extensions are part of the same VLAN range and share configuration. However, each extension counts as a full VLAN range for system VLAN range scale limits.

12.6 Wi-Fi mobility anchor

7750 WLAN-GW supports seamless handling for UE mobility, when a UE moves from one AP to another, where the new AP is broadcasting the same SSID, and is anchored on the same WLAN-GW. In case of open SSID, when the UE re-associates with the same SSID on the new AP and already has an IP@ from association with previous AP, the UE can continue to send and receive data. The WLAN-GW learns the association of the UE's MAC address to the GRE tunnel corresponding to the new AP and updates its state on the MS-ISA as well as on the CPM. The UE continues to be anchored on the same anchor MS-ISA, thereby avoiding any disruption in ESM functions (SLA enforcement and accounting). State update based on data learning results in fast convergence after mobility and minimal packet loss. The data-triggered mobility can be turned on via configuration. Mobility trigger can be configured to be restricted to special Ethernet IAPP frame (originated by the AP with the source MAC of UE).

For 802.1x/EAP based SSIDs, by default the AP requires re-authentication to learn the new session keys (PMK). 7750 SR as WLAN-GW RADIUS proxy infers mobility from the re-authentication, and updates the ESM host to point to the new AP. The new AP's IP address is derived from the RADIUS attribute NAS-IP-address. The re-authentication also provides the new session keys to the AP in access-accept RADIUS response. In case the Wi-Fi AP or ACs are capable of PMK key caching or standard 802.11r (or OKC, the opportunistic key caching pre-802.11r), the re-authentication on re-association can be avoided. In this case the UE can continue to send data, and the WLAN-GW can provide fast data-triggered mobility as defined in context of open SSIDs.

The following output provides a mobility anchor configuration example.

```
config>service>ies>
config>service>vprn>
subscriber-interface <ip-int-name>
group-interface <ip-int-name> wlangw
wlan-gw
     [no] router (base | <vprn-id>) # tunnel service context
     [no] wlan-gw-group <group-id>
      ....snip
       mobility
         [no] trigger {data | iapp}
         [no] hold-time <seconds> // [0..255 secs]
       exit
     exit
exit
```

12.7 WLAN location enhancements

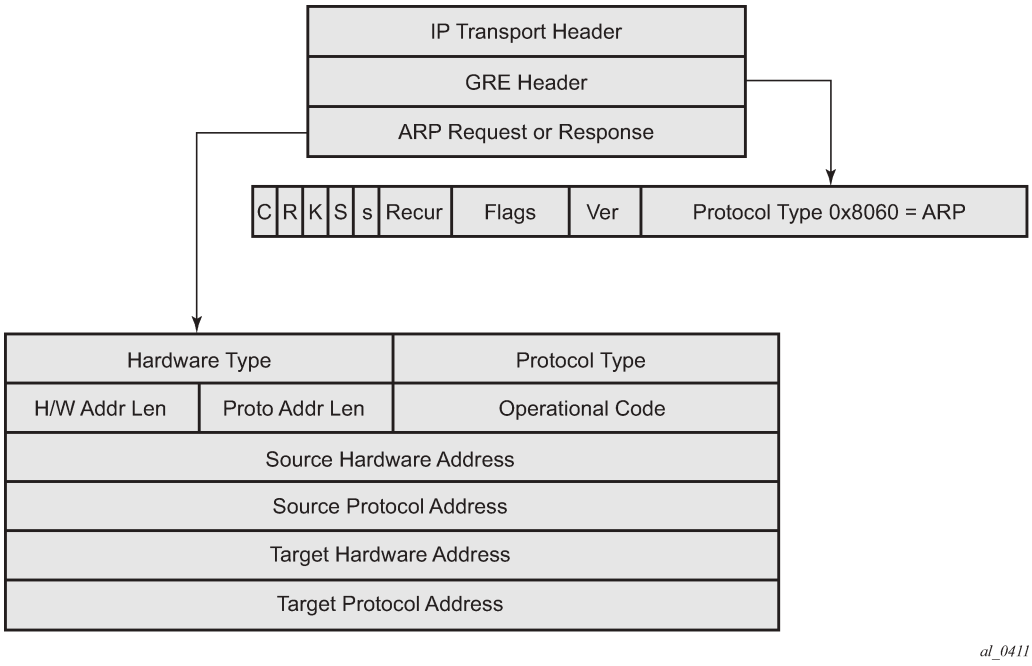
This feature adds configurable support for learning and reporting the AP's MAC address (which represents WLAN location of the UE), to the AAA server. Support is also added for triggered interim accounting-updates to report the AP's MAC@ to the AAA server.

12.7.1 Triggered interim accounting-updates

Using location based policy for Wi-Fi subscribers is important. The business logic in AAA could use the location of the subscriber. Therefore, it is important to notify location change of the subscriber to AAA. Standard way to do this is by generating an interim accounting update when the WLAN-GW learns of the location change for a subscriber. The location for a Wi-Fi subscriber can be inferred from MAC@ (preferred) or WAN IP@ of the AP.

For open-SSID, learning about mobility could be data-triggered or IAPP packet triggered. If triggered, interim accounting-update is configured via CLI, then on detecting a location change for the UE, an interim accounting-update is sent immediately to the AAA server with the new AP's MAC@ (if already known to WLAN-GW). The accounting-update contains NASP-port-id (which contains the AP's IP@), and circuit-id (from DHCP option-82) which contains AP's MAC@ and SSID. In case of data-triggered mobility, if the new AP's MAC@ is not already known to WLAN-GW, a GRE encapsulated ARP packet is generated toward the AP to learn the MAC@ of the AP. The AP is expected to reply with a GRE encapsulated ARP response containing its MAC@. The generation of ARP to learn the AP's MAC@ is controlled via CLI. The GRE encapsulated ARP packet is shown in Figure 166: GRE encapsulated ARP request.

Figure 166: GRE encapsulated ARP request



The standard ARP request must be formatted as follows:

- Hardware Type = Ethernet (1)
- Protocol Type = 0x0800 (IPv4)
- H/W Addr Len = 6
- Proto Addr Len = 4
- Operational code (1 = request)
- Source hardware address = WLAN-GW MAC@

- Source protocol address = Tunnel endpoint IP@ on WLAN-GW
- Target hardware address = Unknown
- Target protocol address = WAN IP@ of the AP (source IP in GRE packet)

The AP must generate a GRE encapsulated ARP response when it receives the GRE encapsulated ARP request for its WAN IP@ (that is used to source tunneled packets). The standard ARP response should be formatted as follows:

- Hardware Type = Ethernet (1)
- Protocol Type = 0x0800 (IPv4)
- H/W Addr Len = 6
- Proto Addr Len = 4
- Operational code (2 = response)
- Source hardware address = AP MAC@
- Source protocol address = WAN IP@ of AP (used for sourcing tunneled packets)
- Target hardware address = source hardware address from the request
- Target protocol address = source protocol address from ARP request

For 802.1x/EAP SSID, the location change (mobility) is learned from an interim-accounting update from the AP. The called-station-Id (containing the AP MAC@) is compared against the current stored called-station-Id that the subscriber is associated with. If the called-station-id is different than the received interim accounting update is immediately forwarded to the accounting server, if triggered interim accounting-update is configured via CLI. In previous releases, the interim-update received from the AP is not immediately forwarded by the accounting proxy. Only a regularly scheduled interim-update is sent.

12.7.2 Mobility triggered interim updates with counters

This feature supports configurable inclusion of counters for an ESM host (UE) in triggered interim updates, because of mobility, generated from WLAN-GW. Triggered interim updates provide a way for RADIUS to learn of the current location of the UE (such as AP MAC).

The feature also supports a configurable hold down time to protect against very frequent mobility events. The CPM must communicate with the IOM for each mobility event to obtain the counters. By default, when inclusion of counters is enabled, no hold-down time is imposed on mobility triggered interim updates. If a hold down time is configured, the first mobility event triggers an interim update with counters included and start the hold down timer. If the hold down timer has not expired, interim updates because of a mobility event are not sent.

The inclusion of counters and hold down time are applicable to all access-types (soft-GRE, soft-L2TPv3, L2-AP, SAP, PW-SAP, and so on), and for DHCP, data-triggered, and EAP authentication or re-authentication triggered mobility.

The configuration for counter inclusion subject to optional hold down time is only applicable to ESM. For counters to be included in the mobility-triggered interim updates, the general **std-acct-attributes** or **detailed-acct-attributes** in the include-attribute in the **radius-accounting-policy** must also be enabled. With DSM, counters are always included in mobility-triggered interim updates if the **isa-radius-policy** has inclusion for either or both **frame-counters** or **octet-counters** enabled. Unlike ESM, no new WLAN-GW-specific configuration is required for the counter inclusion.

The feature is supported at the maximum ESM host scale.

```
config>router>wlan-gw>mobility-triggered-acct>
     interim-update
     interim-update include-counters [hold-down <seconds>]
     no interim-update
<include-counters>   : keyword
<seconds>            : [60..86400]
```

12.7.3 Operational support

Following command shows if GRE encapsulated ARP request is enabled.

```
*A:Dut-C# show router 4 interface "grp-vprn_ue-2/1/2:50" detail
===============================================================================
Interface Table (Service: 4)
===============================================================================
-------------------------------------------------------------------------------
Interface
-------------------------------------------------------------------------------
If Name          : grp-vprn_ue-2/1/2:50
Sub If Name      : ies-4-20.0.0.1
Red If Name      :
Admin State      : Up                  Oper (v4/v6)      : Up/Up
Protocols        : None

WLAN Gateway details
Administrative state        : in-service
Router                      : 50
IP address                  : 10.1.1.3
IPv6 address                : 2001:db8::0
ISA group ID                : 1
Egr shaping                 : none
Egr shape multi UE only     : false
Egr qos policy ID           : (Not Specified)
Egr scheduler policy        : (Not Specified)
Egr agg rate limit (kbps)   : (Not Specified)
Egr qos resrc hold time (s) : 0
Mobility trigger            : data iapp
Mobility ARP AP             : enabled
Mobility hold time (s)      : 0
Default retailer service    : (Not Specified)
TCP MSS adjust              : (Not Specified)
Number of tunnels           : 0
Last management change      : 02/19/2014 17:48:52
```

12.8 Migrant user support

Migrant users are UEs that connect to an SSID but move out of the range of the access-point before initiating or completing authentication. For open-SSIDs, a migrant user may stay in the range of the access-point just enough to get a DHCP lease from the WLAN-GW. In real Wi-Fi deployments with portal authentication, it has been observed that a large percentage of users are migrant, such as get a DHCP lease but do not initiate or complete authentication. Before this feature, an ESM host is created when DHCP completes. This results in consumption of resources on both CPM and IOM, limiting the ESM

scale and performance for fully authenticated active users. This feature adds support to only create an ESM host after a user has been fully authenticated, either via web portal or with a AAA server based on completing EAP exchange. In addition, with this feature L2-aware NAPT is enabled on the ISA, such that each UE gets the same shared configured inside IP@ from the ISA via DHCP. Until a user is authenticated, forwarding of user traffic is constrained (via policy) to only access DNS and portal servers. Each user is allocated a small number of configured NAT outside ports to minimize public IP address consumption for unauthenticated users. After the user is successfully authenticated, as indicated via a RADIUS COA on successful portal authentication, an ESM host is created, and the Layer 2–aware NAT is applied via a normal per-subscriber NAT policy. The inside IP address of the user does not change. The outside IP pool used is as per the NAT policy, and the L2-aware NAT could be 1:1 or NAPT with larger number of outside ports than in the un-authenticated phase. If a user is already pre-authenticated (for example, if RADIUS server remembers the MAC@ of the UE from previous successful portal authentication), then the initial access-accept from RADIUS triggers the creation of the ESM host.

Migrant user support is only applicable to EAP based closed SSIDs when RADIUS-proxy is not enabled on WLAN-GW. This is described in Migrant user support with EAP authentication.

12.8.1 Portal authentication

12.8.1.1 DHCP

Based on DHCP and L2 NAT configuration on the ISA, IP address is assigned to the user via DHCP. A different DHCP lease-time can be configured for an unauthenticated user and an authenticated user for which an ESM host has been created. DHCP return options, for example, DNS and NBNS server addresses can be configured. This configuration can be per wlan-gw group interface or per VLAN range (where a VLAN tag corresponds to an SSID). After the DHCP ACK is sent back to the UE from the ISA, the UE is created on the ISA in "migrant (or unauthenticated) state". ARP requests coming from the UE in migrant state is responded to from the ISA. The authentication to RADIUS is triggered on receiving first L3 data-packet as opposed to on DHCP DISCOVER.

12.8.1.2 Authentication and forwarding

The authentication is initiated from the RADIUS client on the ISA anchoring the user, based on an ISA RADIUS policy (configured under AAA) and specified on the WLAN-GW group interface. The initial Access-Accept from RADIUS can indicate if a user needs to be portal authenticated or is a pre-authenticated user. The indication is based on inclusion of a "redirect policy" applicable to the user, in a VSA (Alc-Wlan-Portal-Redirect, type = string). The Access-Accept can also include a redirect URL VSA (Alc-Wlan-Portal-Url, type = string) for the user. An empty Alc-Wlan-Portal_redirect VSA forces the use of locally configured redirect policy. If neither of the two VSAs are included, this indicates a pre-authenticated user, and an ESM host is created for the subscriber with a subscriber profile and other subscriber configuration from the Access-Accept, and normal ESM-based forwarding occurs for the subscriber.

It is also possible to bypass RADIUS authentication directly using the following command:

- **MD-CLI**

```
configure service ies subscriber-interface group-interface wlan-gw vlan-range authentication
 local
```

```
configure service vprn subscriber-interface group-interface wlan-gw vlan-range
 authentication local
```

- **classic CLI**

```
configure service ies subscriber-interface group-interface wlan-gw vlan-tag-ranges
 authentication local
configure service vprn subscriber-interface group-interface wlan-gw vlan-tag-ranges
 authentication local
```

When you configure this option, the system immediately creates the UE in the DSM or portal state, using the DSM or portal parameters configured under the VLAN range.

However, if a user requires portal authentication (indicated in the Access-Accept), while the authentication is pending, forwarding is restricted to DNS and portal servers via the redirect policy. The redirect policy is an IP ACL that restricts forwarding based on IP destination, destination port, and protocol, and specifies HTTP redirect for HTTP traffic that does not match any of the forwarding rules. The URL for redirect is configured in the redirect policy or provided in the Authentication-Accept. A maximum of 16 redirect policies can be created in the system, with a maximum of 64 forwarding rules across all redirect policies. During the "authentication pending" phase, all forwarded traffic is subjected to Layer 2–aware NAT on the ISA. The NAT policy to use for these users is configured on the WLAN-GW interface or per VLAN range under the WLAN-GW interface. After an Access-Accept has been received from RADIUS for such a user, the next HTTP packet triggers a redirect function from the ISA, and an HTTP 302 is sent to the client. The client presents its credentials to the portal and after successful authentication, a CoA is generated from the RADIUS server (triggered by the portal). The CoA message triggers creation of an ESM host with the subscriber configuration contained in the CoA, such as the subscriber profile, SLA profile, NAT profile and application profile. From this point, normal ESM- based forwarding occurs for the subscriber.

You can configure the redirect URL with the following macro attributes, which are automatically replaced with values relevant to that UE:

- **$MAC**

  This is the MAC address of the UE in the format XX:XX:XX:XX:XX:XX.

- **$IP**

  This is the IPv4 or IPv6 IP address the UE uses to generate the original HTTP request.

- **$URL**

  This is the originally requested URL, which is included as specified without special encoding (for example, using base64). Nokia recommends to only append this URL to the end of the redirect URL, to avoid URL parameter conflicts.

  Example

  If you include the following URL:

  example.com?a=1&b=2

  In the following redirect URL:

  portal.omlogin?url=$URL&mac= &MAC

  This results in a new redirect URL such as the following:

  portal.com/login?url=example.com?a=1&b=2&mac=00:00:5e:00:53:1

  For the portal server, it is not clear whether the parameters "b" and "mac" are part of the original URL or their own URL, which could lead to parsing errors.

- **$NASIP**

This is the RADIUS IPv4 address assigned to the ISA/ESA-VM, either from the authentication policy or the CoA policy. It can be used as the destination address of a CoA.

- **$URLPRM**

  This attribute is replaced by the contents of the Alc-AA-Sub-Http-Url-Param RADIUS attribute, only when AA functionality is applied to the UE. The main reason for using this attribute is to include the redirect URL in DHCP, DHCPv6, and router advertisement (RA) messages.

The configured HTTP redirect URL is also signaled automatically in DHCP, DHCPv6, and ICMPv6 RA messages, as described in RFC 8910. In this case the $URL macro parameter is replaced by an empty string “” because there is no incoming URL. The $IP macro parameter is replaced with the following:

- IPv4 address in DHCP
- SLAAC prefix in RA
- NA address in DHCPv6

Note: Nokia recommends not using the $IP macro replacement because it leads to different URLs being signaled in different messages, which is discouraged by RFC 8910.

See Migrant User NAT Configuration for configuration information related to migrant users.

The WLAN-GW portal authentication framework can be used to support the captive portal architecture as outlined in RFC 8952. The simplest configuration involves provisioning a portal URL that points to a captive portal API server as described in RFC 8908. It is also possible to explicitly provision a URL for the captive portal API server using one of the following commands:

- **MD-CLI**

```
configure service ies subscriber-interface group-interface wlan-gw vlan-range dsm captive-
portal-api-url
configure service vprn subscriber-interface group-interface wlan-gw vlan-range dsm captive-
portal-api-url
```

- **classic CLI**

```
configure service ies subscriber-interface group-interface wlan-gw vlan-tag-ranges dsm
 captive-portal-api-url
configure service vprn subscriber-interface group-interface wlan-gw vlan-tag-ranges dsm
 captive-portal-api-url
```

When this command is configured, the RFC 8910 URL signaled in DHCP, DHCPv6, and RA becomes the configured URL. This URL also supports the same macro substitution options as the HTTP redirect URL described in the preceding information, with the same restrictions. This API server URL can be configured simultaneously with an AA or BB redirect-based captive portal, which keeps using the regular redirect URL. However, in this case, the portal redirect URL is not signaled in DHCP, DHCPv6, or RA.

12.8.2 Migrant user support with EAP authentication

Migrant user support can only be used for closed SSIDs when there is no RADIUS-proxy configured on WLAN-GW. If no RADIUS proxy is configured, then initial RADIUS request carrying EAP from the AP is normally forwarded to a RADIUS server. The RADIUS exchange is between AP and the AAA server, and no information from EAP authentication is cached on the WLAN-GW. The subsequent DHCP DISCOVER after successful EAP authentication is received on the ISA. However, for subscriber that needs to be GTP tunneled to PGW/GGSN, the DHCP is forwarded to the CPM, where it triggers a RADIUS authorization.

RADIUS correlates the MAC address with calling-station-id from EAP authentication for the user. GTP tunnel initiation, and ESM host creation then follows after receiving an Access-Accept. However, for a "local-breakout" subscriber DHCP and Layer 2–aware NAT is handled on the ISA (as in the case for migrant users with portal based authentication). Shared inside IP address can be handed out to each subscriber. The first L3 packet triggers MAC address based RADIUS authorization from the ISA. RADIUS server can correlate the EAP authentication with the MAC address of the user and send an access-accept. This triggers ESM host creation as normal.

For closed SSIDs with EAP authentication, if a RADIUS proxy function is configured on WLAN-GW, then the initial EAP authentication from the AP is processed by the RADIUS-proxy on the CPM and is forwarded to the RADIUS server based on configured authentication policy. Based on authentication response, ESM host creation with local DHCP address assignment or GTP tunnel initiation proceeds as usual.

12.8.3 Data-triggered subscriber creation

With **data-triggered-ue-creation** configured under wlan-gw group interface or per VLAN range (such as, per one or more SSIDs), the first UDP or TCP packet received on WLAN-GW ISA from an unknown subscriber (with no prior state, such as an unknown MAC address) triggers RADIUS authentication from the ISA. The authentication is based on configured **isa-radius-policy** (under the **aaa** context). If RADIUS authentication succeeds, then ESM host is created from the CPM. The ESM host can get deleted based on idle-timeout. Data-triggered authentication and subscriber creation enables stateless inter WLAN-GW redundancy, as shown in Figure 167: N:1 WLAN-GW redundancy based on data-triggered authentication and subscriber creation. If the AP is configured with a backup WLAN-GW address (or FQDN), it can tunnel subscriber traffic to the backup WLAN-GW, when it detects failure of the primary WLAN-GW (based on periodic liveness detection). With "data-triggered-ue-creation" configured, the first data packet results in authentication and ESM host creation on the backup WLAN-GW. If the subscriber had obtained an IP address via DHCP with Layer 2–aware NAT on the primary WLAN-GW, it can retain it with Layer 2–aware NAT on the backup WLAN-GW. The NAT outside pool for the subscriber changes on the backup WLAN-GW based on local configuration. For a subscriber that needs to be anchored on GGSN/PGW (as indicated via RADIUS access-accept), RADIUS server returns the IP address of PGW/GGSN where the UE was anchored before the switch-over. GTP tunnel is then signaled with "handover indication" set. The PGW/GGSN must return the requested IP address of the UE, which is the address with which the UE originated data packet that triggered authentication.

The same data-triggered authentication and subscriber creation is also used to support inter WLAN-GW mobility, such as when a UE moves form one AP to another AP such that the new AP is anchored on a different WLAN-GW. This is shown in Figure 167: N:1 WLAN-GW redundancy based on data-triggered authentication and subscriber creation.

Figure 167: N:1 WLAN-GW redundancy based on data-triggered authentication and subscriber creation

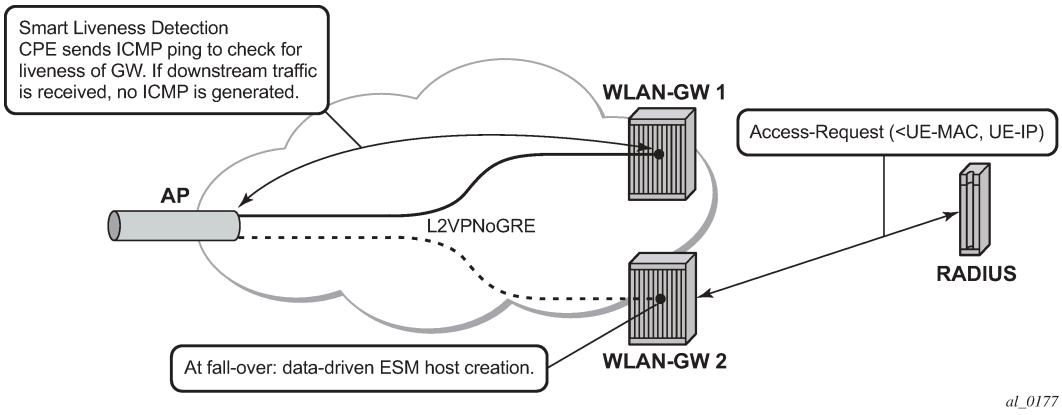
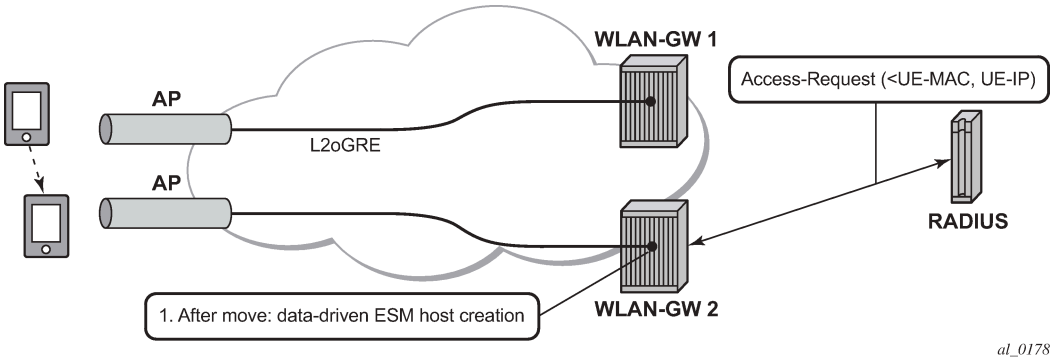


Figure 168: Inter WLAN-GW mobility based on data-triggered authentication and subscriber creation



The following output displays the configuration for migrant user support and "data-triggered" subscriber creation.

Migrant user NAT configuration

```
#--------------------------------------------------
 NAT configuration for migrant and authenticated users
#--------------------------------------------------
service

  vprn 300 customer 1 create

     nat
       inside
          l2-aware
               address 10.20.12/16
          exit
       exit
       outside
           pool "migrant_outside_pool" nat-group 1 type wlan-gw-anchor create
                address-range 10.22.0.0 10.22.0.255 create
                exit
                no shutdown
           exit
           pool "wifi_outside_pool" nat-group 1 type l2-aware create
                address-range 10.0.0.0 10.0.0.255 create
```

```
                exit
                no shutdown
            exit
        exit
     exit
  exit

  nat
   nat-policy "migrant_nat_300" create
        pool "migrant_outside_pool" router 300
        timeouts
             tcp-established min 1
        exit
   exit

   nat-policy "wifi_nat_300" create
        pool "wifi_outside_pool" router 300
   exit

 exit

#--------------------------------------------------------------------------------
echo "AAA Configuration" - ISA-RADIUS-Policy for authentication from WLAN-GW ISA
#--------------------------------------------------------------------------------
    aaa
        isa-radius-policy "wifi_isa_radius" create
            description "Default authentication policy for migrant users"
            password "i2KzVe9XPxgy4KN2UEIf6jKeMT3X4mT6JcUmnnPZIrw" hash2
            servers
                router "Base"
                source-address-range 10.100.100.4
                server 1 create
                    authentication
                    coa
                    ip-address 10.100.100.2
                    secret "ABIQRobhHXzq13ycwqS74FSrj.OdTwh5IdjhRB.yAF." hash2
                    no shutdown
                exit
            exit
        exit
        radius-server-policy "radius_server_policy" create
            servers
                router "Base"
                server 1 name "radius_server"
            exit
        exit
    exit

#-------------------------------------------------
echo "Subscriber-mgmt Configuration" - Redirect Policy
#-------------------------------------------------
    subscriber-mgmt
        http-redirect-policy "migrant_redirect" create
            url "portal.ipdtest.nokia.com:8081/start/?mac=$MAC&url=$URL&ip=$IP"
            portal-hold-time 10
            forward-entries
                dst-ip 10.8.8.1 protocol tcp dst-port 8081
                dst-ip 10.8.8.7 protocol tcp dst-port 8007
                dst-ip 10.8.8.8 protocol udp dst-port 53
            exit
        exit
     exit
service
```

```
#------------------------------------------------------------------
echo "migrant user configuration under wlan-gw group interface”
#------------------------------------------------------------------

  vprn 300 customer 1 create

    subscriber-interface "ies-4-20.10.1.1" create
        address 10.20.12/16

        group-interface "grp-vprn_ue-2/1/2:51" wlangw create
            sap-parameters
               sub-sla-mgmt
                   def-sla-profile "slaprof_1"
                   def-sub-profile "subprof_1"
                   sub-ident-policy "identprof"
                exit
             exit
             dhcp
                 proxy-server
                     emulated-server 10.20.12.12
                     no shutdown
                 exit
                 trusted
                 lease-populate 32767
                 user-db "radius_ludb"
                 no shutdown
              exit
              host-connectivity-verify interval 1000
              wlan-gw
                  gw-addresses
                     address 10.1.1.4
                  exit
                  mobility
                      hold-time 0
                      trigger data iapp
                  exit
                  router 50
                  wlan-gw-group 1
                  vlan-tag-ranges
                       range start 100 end 100
                            authentication
                                 authentication-policy "wifi_isa_radius"
                            exit
                            data-triggered-ue-creation
                            dhcp
                               l2-aware-ip-address 10.1.1.2
                               primary-dns 10.1.1.1
                               secondary-dns 10.1.1.1
                               no shutdown
                            exit
                            nat-policy "migrant_nat_4"
                        exit
    exit
                   no shutdown
                exit
          exit
     exit
exit
```

12.9 Distributed Subscriber Management

With Distributed Subscriber Management (DSM), after the UE is successfully authenticated (portal, auto-signed-in, or EAP), the corresponding subscriber can be created on the anchor ISA, and both control plane and forwarding plane for the subscriber are handled on the ISA. This mode of subscriber management is therefore referred to as Distributed Subscriber Management (DSM).

Before this feature, only ESM is supported for WLAN UEs, where the ESM host state is created on the IOM/IMMs from the CPM (triggered by the ISA on successful authentication). With ESM, the initial DHCP process and authentication could be triggered from the ISA (based on a per VLAN-range configuration for DHCP) under the group-interface with of type **wlangw**. However, control plane operations after the ESM host creation (such as accounting and DHCP renews) are handled on the CPM.

With DSM, in addition to initial DHCP and authentication, after the subscriber state exists on the anchor ISA, accounting and DHCP renews are also handled from the anchor ISA for the UE. This allows a higher UE scale and better control plane performance (including DHCP transactions per second, rate of authentications, and web redirects) because of load-balancing amongst set of ISAs in the WLAN-GW group. With DSM, the UE data-plane functions (such as per UE IP filtering, ingress/egress policing, legal intercept, per UE counters, and web-redirect) are performed on the ISA.

The decision to create an authenticated UE as an ESM or DSM UE can be controlled from RADIUS via inclusion of Alc-Wlan-Ue-Creation-Type VSA. The VSA can be included in access-accept for a UE that is auto-signed-in (for example, it does not need web redirect to portal), or in a COA message triggered to remove web redirect for a UE after successful portal authentication. The VSA is described in the *7450 ESS, 7750 SR, and VSR RADIUS Attributes Reference Guide*. If Alc-Wlan-Ue-Creation-Type is not present in access-accept (for auto-signed UE) or in the COA message (for UE creation of portal authenticated UE), then the UE is created as an ESM host. DSM is not supported for UEs which require a GTP host. If Alc-Wlan-Ue-Creation-Type indicates a DSM UE then any IPv6 or GTP related parameters in access-accept or COA is ignored, and the UE is created as a DSM host. Alc-Wlan-Ue-Creation-Type cannot be changed mid-session via COA. A COA containing Alc-Wlan-Ue-Creation-Type for an existing UE does not result in any change of state and is NACK'ed.

12.9.1 DHCP

Based on DHCP and L2 NAT configuration on the ISA, the configured IP address (**l2-aware-ip-address** configured under **vlan-tag-ranges range** start *vlan-id* **end** *vlan-id* or **vlan-tag-ranges range default**) is assigned to the user via DHCP. A different DHCP lease-time can be configured for an un-authenticated and an authenticated user for which an ESM or DSM host has been created. DHCP return options, for example, DNS and NBNS server addresses can be configured. This configuration can be per soft-wlan-gw group interface (by explicitly configuring it under **vlan-tag-ranges range default**), or per VLAN range (where a VLAN tag corresponds to an SSID). By default, for open SSIDs, DHCP DORA is completed, and authentication request is sent to AAA server only on reception of the first Layer 3 packet. However, with an **authenticate-on-dhcp** command configured under **vlan-tag-ranges range default** (default or specific range), authentication can be triggered on received DHCP DISCOVER or REQUEST when no UE state is present. If UE anchoring on GGSN/PGW is required, then **authenticate-on-dhcp** must be enabled, because the decision to setup GTP tunnel (in which case the IP@ for the UE comes from the GGSN/PGW) is based on RADIUS response.

To support unique inside IP addresses, the ISA Pool Manager can be used. Pools are allocated to each ISA in large blocks, requiring an IPv4 subnet with prefix-length 16. From this prefix, the subnet address (x.x.0.0), broadcast address (x.x.255.255), and gateway address (x.x.0.1) are reserved and not allocated to UEs, reducing the total number of available addresses by three. NAT pools are only available in non-

retailer subscriber interfaces. Any retail service ID derived from configuration or AAA is only used for IPv6 pool selection and is ignored for IPv4 NAT pools. Forwarding in a different VRF can be achieved by selecting a different NAT policy and outside VRF.

12.9.2 Authentication and accounting

The authentication is initiated from RADIUS client on the ISA anchoring the user, based on an isa-radius-policy (configured under **aaa**) and specified on the wlan-gw group-interface. This support exists in prior releases and is described in Authentication and forwarding. The auth-policy can contain up to ten servers, five of which can be for authentication and all ten can be COA servers.

To generate accounting updates for DSM UEs, an accounting policy (type isa-radius-policy) must be configured under the **aaa** node and specified under **vlan-range** (**default** or **specific range**) on the wlan-gw interface. Accounting for DSM UEs includes **accounting-start**, **accounting-stop**, and **interim-updates**. Interim-update interval is configurable under vlan-range on wlan-gw interface. The username format to be included in RADIUS messages is configurable in the auth-policy and accounting-policy via the **user-name-format** command. By default, the username contains the UE MAC address, but can be configured to include the UEs MAC address and IP address, or circuit-id or DHCP vendor options. If **authenticate-on-dhcp** is enabled, then the IP address for the UE is not known before authentication, and, if the username is configured to contain both MAC and IP address, then only the MAC address is included.

The accounting-policy can be configured with attributes to be included in the accounting messages. The details of the attributes are covered in the *7450 ESS, 7750 SR, and VSR RADIUS Attributes Reference Guide*. The attributes are included here for reference.

```
*A:Dut-1>config>aaa# info
----------------------------------------------
  isa-radius-policy "isaRadiusPol1" create
      user-name-format mac mac-format alu
     acct-include-attributes
           acct-delay-time
           acct-trigger-reason
           called-station-id
           calling-station-id
           circuit-id
           dhcp-options
           dhcp-vendor-class-id
           frame-counters
           framed-ip-addr
           framed-ip-netmask
           hardware-timestamp
           inside-service-id
           mac-address
           multi-session-id
           nas-identifier
           nas-port-id
           nas-port-type
           octet-counters
           outside-ip
           outside-service-id
           port-range-block
           release-reason
           remote-id
           session-time
           subscriber-id
           ue-creation-type
           user-name
           wifi-rssi
```

```
            wifi-ssid-vlan
        exit
```

The **isa-radius-policy** for auth/COA and accounting specifies the server selection method for the servers specified in the policy with respect to load-balancing and failure of one or more servers. The three methods implemented include:

- **direct**

  Specifies that the first server is used as primary for all RADIUS messages, the second server is used as secondary (that is, used for all RADIUS messages if primary server fails), and so on.

- **round-robin**

  RADIUS messages across accounting-sessions are distributed in a round-robin manner amongst the list of configured servers. All accounting messages for a given session are sent to the selected server for that session, until that server fails. If a server fails, then the sessions targeted to that server are distributed in a round-robin manner amongst the remaining servers. If the failed server comes back up, the sessions that were originally assigned to the failed server revert to the original server.

- **hash**

  Server is picked via hash on UE MAC. The hash list consists of all configured servers that are up. If a server fails, then the UEs hashed to that server are re-hashed over the remaining servers that are up.

If a response is not received for a RADIUS message from a particular server for a configurable timeout value (per server), and the time elapsed because the last packet received from this RADIUS server is longer than this configured timeout value, then the server is deemed to be down. Periodically an accounting-on message is sent to a server that is marked as down, to probe if it has become responsive. If a response is received then the server is marked as up.

```
*A:Dut-1>config>aaa# info
isa-radius-policy "isaRadiusPol1" create
            nas-ip-address-origin system-ip
            password "6mNsKxvTe.0.nNCTIpGFcu.rr/qtdijazQ3ED8WAFfk" hash2
            user-name-format mac mac-format alu release-reason
            servers
                access-algorithm hash-based
                retry 3
                router "Base"
                source-address-range 81.1.0.1
                timeout sec 5
                server 1 create
                    accounting port 1813
                    authentication port 1812
                    coa port 3799
                    ip-address 10.13.0.2
                    secret "3BmWbBfDO38hPY8DtLFn8bYDBaduy6w.ogeSUsouoHc" hash2
                    no shutdown
                exit
            exit
        exit
-------------------------------------------
*A:Dut-1>config>aaa#
```

12.9.2.1 DSM data-plane

NAT on the anchor ISA is required for forwarding of traffic to/from a DSM UE. There is no UE state in the IOM/IMM for a DSM UE. The downstream forwarding is based on FDB lookup that should match a route corresponding to the NAT outside pool and get the downstream traffic to the right anchor ISA, where NAT is performed for the UE. The inside IP address assigned to the UE is the configured **l2-aware-ip-address** on the vlan-range (default or specific range) under wlan-gw interface. Therefore every UE corresponding to the default or specific vlan-range gets the same inside IP@. The NAT is L2-aware and uses UE MAC to de-multiplex.

12.9.3 IP filtering

Filtering based on protocol, destination IP, destination port or any combination is supported for traffic to and from the UE. The match entries and corresponding actions can be specified within the **isa-filter** which can be created in the **config**>**subscr-mgmt** context. The filter can be associated with a vlan-range (default or specific vlan-range) on wlan-gw interface, in which case all subscribers associated with the vlan-range is associated with an instance of this filter.

The supported filter actions include drop and forward. The first match causes corresponding action to be executed and no further match entries is executed. If there is no match, or no action configured for a match, configurable default action for the filter is executed. The filter can be overridden on a per UE basis via RADIUS access-accept or COA. The new VSA Alc-Wlan-Dsm-Ip-Filter is defined for specifying the per UE filter from RADIUS. The VSA is defined in the RADIUS guide.

```
A:system>config>subscr-mgmt# info
----------------------------------------------
                isa-filter  “foo” type dsm create
                    default-action forward
                    entry 1 create
                        action drop
                        match protocol udp
                            dst-ip 239.0.113.0/32
                            dst-port  eq  53
                        exit
                    exit
                    ipv6
                        default-action forward
                        entry 1 create
                            action drop
                            match protocol tcp
                                dst-ip 2001:db8::/120
                                dst-port  eq  80
                            exit
                        exit
                    exit
----------------------------------------------


*A:vsim>config>service>vprn# info
----------------------------------------------
 subscriber-interface "s1" create
      group-interface "g1" wlangw create
             wlan-gw
                   vlan-tag-ranges
                          range default
                               distributed-sub-mgmt
                                    dsm-ip-filter “foo”
```

```
                            exit
                        exit
                    exit
                exit
-----------------------------------------------
```

12.9.3.1 HTTP redirect

DSM ISA filters allow for the specification of an HTTP Redirect action. Valid HTTP Request flows matching the entry are redirected using the specified URL. All other packets are dropped. The URL can be overridden by AAA authentication on a per UE basis, but this override applies to all redirect entries. The URL supports the $MAC, $IP, and $URL substitution variables.

The basic filter-based HTTP Redirect action can be combined with one-time redirect, where filter behavior has precedence. Traffic matching an entry with an **http-redirect** action is redirected using the specified URL. Traffic matching an entry with a **forward** action is redirected using the one-time URL and resets the one-time redirect override. One-time redirect is mutually exclusive to a dynamic URL override for filter-based redirection.

```
-----------------------------------------------
isa-filter "redirect-example" type dsm create
    default-action forward
    entry 10 create
        match protocol tcp
            dst-port  eq  80
        exit
        action http-redirect www.example.org?mac=$MAC
    exit
    entry 20 create
        match protocol tcp
            dst-port  eq  8080
        exit
        action http-redirect www.example.org/alternate_port/?ip=$IP
    exit
exit
-----------------------------------------------
```

12.9.4 Policing

Per UE policing for both ingress and egress direction is supported. Policers can be created in the **config>subscr-mgmt>isa-policer** context. The policers can be of type single-bucket (PIR) bandwidth limiting or dual-bucket (PIR and CIR) bandwidth limiting. Only policer action supported as permit-deny, non-conforming traffic is dropped, as opposed to marked out-of-profile. The administrative peak and committed rates and peak and committed burst sizes are configurable. For single-bucket bandwidth policers, CIR and CBS are not applicable, and only PIR and MBS are configurable.

```
*A:vsim>config>subscr-mgmt>isa-policer# info detail
-----------------------------------------------
                    no description
                    action permit-deny
                    cbs 100
                    mbs 200
                    rate 1000 cir 500
```

The policers can be associated with a vlan-range (default or specific vlan-range) on wlan-gw interface, in which case all subscribers associated with the vlan-range is associated with an instance of these policers. These ingress and egress policers can be overridden on per UE basis via RADIUS access-accept or COA. The new VSAs Alc-Wlan-Dsm-Ingress-Policer and Alc-Wlan-Dsm-Egress-Policer are defined for specifying the per UE policers from RADIUS. The VSAs are defined in the *7750 SR-OS RADIUS Attributes Reference Guide*. If the policers specified in access-accept are not found the message is dropped. If the policers specified in COA are not found, a NACK is sent back.

```
*A:vsim>config>service>vprn# info
----------------------------------------------
 subscriber-interface "s1" create
      group-interface "g1" wlangw create
             wlan-gw
                   vlan-tag-ranges
                          range default
                               distributed-sub-mgmt
                                    egress-policer "silver-egress"
                                    ingress-policer "silver-ingress"
                               exit
                          exit
                    exit
              exit
----------------------------------------------
```

12.9.5 Lawful Intercept (LI)

LI can be triggered for a DSM UE LI via CLI or RADIUS, and is performed post-NAT. Only routable encaps (IP/UDP/LI-shim) and IP-only mirror-dest are supported. A maximum of 2K DSM UEs per-chassis can be under LI simultaneously. LI mirror dest (service in which mirrored packets are injected) along with other required mirror information (mirror-dest type, encapsulation-type, ip-udp-shim, and encapsulation information, IP and UDP header information) is configurable. A DSM UE identified by its MAC address can be associated with the mirror destination (service in which mirrored packets for the host are injected) via the **li-source** command. For routable encapsulation (IP/UDP/LI-Shim), the session-id and transaction-id to be inserted in the LI-Shim are configured under **li-source**.

```
A:Dut-1>config>mirror# info
----------------------------------------------
        mirror-dest 60000 type ip-only create
            encap
                layer-3-encap ip-udp-shim create
                    gateway create
                        ip src 1.1.1.1 dest 2.2.2.2
                        udp src 2048 dest 2049
                    exit
                exit
            exit
            no shutdown
        exit

----------------------------------------------
A:Dut-1>config>li# info
----------------------------------------------
        li-source 60000
            wlan-gw
                dsm-subscriber mac 00:00:00:07:02:03
                    intercept-id 10000
```

```
                    session-id 20000
                exit
            exit
            no shutdown
        exit
```

LI can be enabled or disabled from RADIUS via inclusion of the Alc-LI-Action VSA in access-accept or COA. The Alc-LI-Destination VSA is required to indicate the mirror-dest service that the DSM UE under LI is associated with. The Intercept-Id and Session-Id for a DSM UE can be provided from RADIUS access-accept or COA via inclusion of Alc-LI-Intercept-Id and Alc-LI-Session-Id VSAs. These LI related VSAs are described in the *7450 ESS, 7750 SR, and VSR RADIUS Attributes Reference Guide*.

Information for a specific **li-source** and its associated **mirror-dest** can be shown via CLI.

12.9.6 Data-triggered UE creation

Like data-triggered UE creation with ESM, a DSM UE can also be created based on data-triggered authentication discussed in Data-triggered subscriber creation. The decision to create ESM versus DSM UE is based on the value of RADIUS VSA Alc-Wlan-Ue-Creation-Type present in the access-accept message. The data-triggered authentication and UE creation if configured provides for WLAN-GW IOM redundancy. The DSM UE is created on the standby ISA based on successful data-triggered authentication. Also, inter-chassis redundancy is supported for DSM UE based on data-triggered authentication and is identical to ESM (as described in Data-triggered subscriber creation).

If an IPv6 data trigger is received for which the source IP address matches a SLAAC prefix that is already reassigned to another UE, the WLAN-GW takes the following actions:

- allocates a new SLAAC prefix
- immediately sends a Router Advertisement (RA) message including the following two prefixes:
  - A new SLAAC prefix is included with a regular preferred and valid lifetime, as provisioned for the session.
  - A SLAAC prefix is derived from the data trigger source IP address with the preferred and valid lifetime set to zero. This deprecates the prefix and expedites the UE to stop using it, as per RFC 4862.

12.9.7 Idle-timeout and session-timeout management

The per UE idle-timeout value can be provided in RADIUS access-accept or COA for a DSM UE in standard Idle-Timeout attribute. The minimum idle-timeout allowed is 150 seconds. The idle-timeout is enforced on the ISA for a DSM UE. If there is no data to/from a UE for up to idle-timeout value, the UE is removed and accounting-stop is sent. Subsequently, if a UE re-associates and connects to an open SSID on an AP, and has an IP address with a valid lease, then the first data packet from the UE triggers authentication. Successful authentication results in creation of DSM UE.

To improve idle-timeout behavior an optional SHCV check can be performed after the idle-timeout expires. This check verifies connectivity to all DHCP, DHCPv6 and /128 SLAAC addresses using ARP or NDP. While the check is performed for every address, the result is applied to the whole UE. The UE is only deleted when verification of all addresses fails. When at least one connectivity verification succeeds the UE and all of the allocated addresses are kept and the idle-timeout process is restarted.

The per UE session timeout value can be provided in RADIUS access-accept or COA in standard Session-Timeout attribute. The value is interpreted as an absolute value, and the UE is unconditionally deleted regardless of activity. The minimum allowed value for session-timeout is 300 seconds.

12.9.8 Operational commands

The following shows the command usage to dump information about UE under LI (only allowed to users with LI privilege).

```
A:Dut-1# tools dump li wlan-gw ue
No sessions on Slot #2 MDA #1 match the query
===============================================================================
Matched 2 sessions on Slot #2 MDA #2
===============================================================================
UE-Mac          : 00:00:00:07:02:03     Mirror Service  : 60000
LI Intercept-Id : 10000                 LI Session-Id   : 20000
-------------------------------------------------------------------------------
UE-Mac          : 00:00:00:07:02:08     Mirror Service  : 60000
LI Intercept-Id : 42                    LI Session-Id   : 2013
-------------------------------------------------------------------------------
===============================================================================


A:Dut-1>show>li# li-source 60000
===============================================================================
Mirror Service
===============================================================================
Service Id       : 60000                Type          : ipOnly
-------------------------------------------------------------------------------
L3 encap type    : ip-udp-shim          Router        : Router: Base
                                        Direction bit : No
Primary gateway
Source IP        : 1.1.1.1              Dest IP       : 2.2.2.2
Source UDP port  : 2048                 Dest UDP port : 2049
===============================================================================
Local Sources
-------------------------------------------------------------------------------
Admin State      : Up
-------------------------------------------------------------------------------
WLAN Gateway LI sources
-------------------------------------------------------------------------------
MAC-Address                                          Intercept-Id Session-Id
-------------------------------------------------------------------------------
00:00:00:07:02:03                                    10000        20000
===============================================================================
```

12.9.9 Pool manager

To support allocations of unique IP addresses each ISA is assigned pools from a centralized pool manager on the CPM. The ISA can subsequently assign addresses from these pools to UEs, but this state is not synchronized back to the CPM. Different applications have different pools, for example, SLAAC and DHCPv6 IA_NA cannot share a single pool. To support Wholesale/Retail scenarios a pool-manager can be configured per subscriber interface.

The allocation of additional pools and freeing up unused pools is based on configurable high and low watermarks. When the usage level of all pools combined on an ISA reaches the high watermark, a new

pool is allocated. When the usage level of a single pool reaches zero and the usage level of the other pools combined is below the low watermark, this pool is freed.

With redundancy, the pool manager signals the pools that were allocated to the failed ISA back to the new active ISA. These pools can no longer be used to allocate new addresses because allocations are lost. However, these can still be used to forward traffic based on data-triggered UE creation. This is supported both for IOM redundancy and Active/Standby WLAN-GW redundancy. The new active ISA also receives new pools that it can use for new allocations.

The pool manager uses DHCPv6 Prefix Delegation to allocate pools to the ISAs. Each ISA is represented by a separate DHCPv6 Client ID. These clients request fixed prefix sizes to accommodate up to 64K UEs. With Active/Standby redundancy the Pool Manager uses a DHCPv6 Lease Query Message to retrieve the prefixes that were allocated to the failed WLAN-GW. To identify the correct PD leases in the DHCPv6 server, a configurable virtual-chassis-name is added to the DHCPv6 client-id, this value should be identical on both WLAN-GWs and unique otherwise. The Pool Manager always sends out a DHCPv6 Relay message and supports up to eight DHCPv6 servers.

IPv4 pools are supported by encoding an IPv4 subnet into an IPv6 prefix. The least significant 32 bits of the prefix are treated as an IPv4 address and the allocated prefix-length is subtracted with 96 to obtain an IPv4 prefix length. It is recommended to use the special IPv6 prefix "::ffff:/96" to provision these pools.

```
A:system>config>subscr-mgmt>wlan-gw# info
----------------------------------------------
            virtual-chassis-identifier "wlan_gw_pair"
----------------------------------------------
A:system>config>service>vprn>sub-if>wlan-gw# info
----------------------------------------------
                pool-manager
                    watermarks high 85 low 66
                    wlan-gw-group 1
                    dhcpv6-client
                        server 2001:db8::1
                        lease-query max-retry 2
                        slaac
                            pool-name "pool_ue_pd_v6_slaac"
                            no shutdown
                        exit
                        ia-na
                            pool-name "pool_ue_pd_v6_dhcp6"
                            no shutdown
                        exit
                        dhcpv4-nat
                            pool-name "pool_ue_pd_v6_dhcp4_nat"
                            no shutdown
                        exit
                    exit
                exit
----------------------------------------------
```

12.9.10 DHCPv6 and SLAAC

DHCPv6 and SLAAC support can be configured per VLAN range. Authentication can be triggered by a Router Solicit, DHCPv6, or DHCP packet but is only triggered only one time per UE. Each UE can be assigned a unique DHCPv6 IA_NA address (/128) or SLAAC prefix (/64) from the ISA pools. The IPv6 pools are installed by the centralized pool manager. SLAAC privacy extensions are supported and up to three /128 SLAAC addresses can be learned via either Duplicate Address Detection or upstream data.

Wholesale/retail is supported (IPv6 only) both by RADIUS and per vlan-range CLI, the applicable pool is selected from the retailer service. ESM and DSM IPv6 are not supported in the same *vlan-range* context.

```
A:system>config>service>vprn>sub-if>grp-if>wlan-gw>ranges>range# info
----------------------------------------------
                                ...
                                authenticate-on-dhcp
                                ...
                                dhcp6
                                    active-preferred-lifetime hrs 1
                                    active-valid-lifetime hrs 1
                                    no shutdown
                                exit
                                slaac
                                    active-preferred-lifetime hrs 1
                                    active-valid-lifetime hrs 1
                                    no shutdown
                                exit
                                ...
----------------------------------------------
```

Configuration of other DHCPv6/SLAAC parameters, such as server DUID and RA flags is taken from the **wlan-gw** *group-interface* configuration. For DSM only the configuration of the **wlan-gw** *group-interface* applies, the retailer interface cannot override this configuration.

```
A:system>config>service>vprn>sub-if>grp-if>ipv6# info
----------------------------------------------
                        router-advertisements
                            other-stateful-configuration
                            prefix-options
                                autonomous
                            exit
                        exit
                        dhcp6
                            proxy-server
                                server-id duid-en string "example_duid"
                            exit
                        exit
```

A subset of DHCPv6 options retrieved by the pool-manager in the PD process is reflected in DHCPv6 toward the client. For IA_NA leases these are included in the associated DHCPv6 messages. For SLAAC allocations, the DNS option can be reflected in the Router Advertisement and all options can also be reflected in a stateless DHCPv6 Information Reply message.

When using a captive portal, different valid/preferred lifetimes can be configured for authenticated and un-authenticated UEs. The DHCPv6 lease time is equal to the applied valid-lifetime and can be extended via the regular renew process. SLAAC lifetime is equal to the applied valid-lifetime and is extended when sending an RA including the SLAAC prefix. To avoid infinite SLAAC allocations, when sending an unsolicited RA, SHCV is performed for all learned /128 addresses. If SHCV fails for all addresses, the unsolicited RA is not contained the SLAAC prefix and the SLAAC lifetime is not extended.

In redundancy scenarios the new active ISA migrates the leases in the old pools as soon as possible to a lease in the new pools. For SLAAC this is done by sending an unsolicited RA to deprecate the old prefix (lifetimes 0) and include a new prefix. For DHCPv6 this is done during the first Renew, that again deprecates the old address (lifetimes 0) and include a new address in the same IA.

12.9.11 Application Assurance support

To support Application Assurance (AA) with DSM, an AA group needs to be linked to a WLAN-GW group. By default, every ISA configured in the WLAN-GW, including standby ISAs, requires a counterpart in the AA group. In case AA is only required for a subset of UEs, an oversubscription factor can be configured. With an oversubscription factor of N, each AA ISA can serve up to N WLAN-GW ISAs, including standby ISAs.

When oversubscription is used, the maximum number of AA-enabled UEs per WLAN-GW ISA is the maximum number of UEs for that ISA divided by the oversubscription factor.

AA is enabled per-UE by provisioning an application profile to the UE. A default profile can be provisioned under a **vlan-range**, or a specific profile can be signaled using RADIUS.

When an ISA AA fails, the corresponding BB ISAs continue to forward traffic without AA functionality. After the AA ISA recovers, AA is enabled again for those UEs.

For more information about how to configure AA, see the *7450 ESS, 7750 SR, and VSR Multiservice ISA and ESA Guide*.

12.9.12 Volume quota enforcement

Volume quota can be provided per UE using the RADIUS attribute Alc-Credit-Control-Quota. For DSM, the time quota of the attribute is not supported and must be zero. The quota categories are non-configurable and must be either soft or hard. Upon exhaustion of hard quota the UE is immediately disconnected.

The following non-configurable quota categories are supported:

- **hard quota**

  When a hard quota is exhausted, the UE is immediately disconnected.

- **soft quota**

  When a soft quota is exhausted, the WLAN-GW can apply several actions:

  – Send a triggered RADIUS Accounting Interim Update with the reason WLAN-Quota-Exhausted, if enabled under the **configure aaa isa-radius-policy** *name* **acct-update-triggers soft-quota-exhausted** context.

  – Apply a filter configured under the **vlan-tag-ranges range** *range* **distribute-sub-mgmt soft-quota-exhausted-filter** context. The filter replaces any previous filter applied to the UE. If the soft quota is extended by a CoA message, the filter is reverted to the previously applied filter for that UE.

Both soft and hard quotas can be independently extended by means of CoA messages, using the same attribute. Quotas received in a CoA message are only enforced from that moment onward and none of the UE traffic before the CoA message counts toward the new quota. This supports flexible combinations of hard and soft quotas, for example:

- Apply soft quota close to hard quota to offer some grace period before the session is terminated. During this grace period, the UE is notified when nearing quota exhaustion using an HTTP redirect installed by the **soft-quota-exhausted-filter** command and can buy new credit. After the UE buys new credit, a CoA message updates both hard and soft quota.
- Apply hard quota as the real limit but use soft quota to automatically trigger accounting updates instead of rely on periodic interim updates; for example, for each 5% of volume consumed. Upon each interim update, a CoA message is sent that resets only the soft quota and generates a new report.

- Apply hard quota as the real limit but use the **soft-quota-exhausted-filter** command (**configure service ies/vprn subscriber-interface group-interface wlan-gw vlan-tag-ranges range distributed-sub-mgmt**) to redirect to an advertisement page, for each amount of volume consumed. After the advertisement has been shown, a CoA message extends the soft volume to remove the filter again.

By default, quota is enforced on the combination of upstream and downstream traffic. This can be changed to only upstream or only downstream using the **volume-quota-direction** command in the **vlan-tag-ranges range** *range* **distribute-sub-mgmt volume-quota-direction** context.

12.10 Enhanced Subscriber Management

12.10.1 Authentication

The solution supports multiple authentication mechanisms. Type of authentication support depends on the Wi-Fi AP, UE capabilities and customer preference. In case of 802.1x/EAP capable Wi-Fi APs, supporting secure SSIDs via 802.11i/WPA2, various EAP based authentication such as SIM/uSIM based (SIM/AKA/AKA'), TTLS, PEAP, certs, and so on, are supported. The solution also supports web-portal based authentication with or without WISPr client on the UE. EAP and portal authentication works independent of the type of connectivity from the AP (tunneled or native IP).

The SR OS WLAN-GW uses the IPoE session concept to authenticate and manage UEs in ESM. Every WLAN-GW group interface uses a pre-defined default **ipoe-session-policy** that cannot be changed or disabled. The contents of the default policy also cannot be changed and always uses **sap** and **mac** as **session-key**. The **ipoe-session session-timeout** can optionally be ignored in a **wlan-gw** context. This is to support closed SSID authentication where the session-timeout is relative to the last re-authentication while for **ipoe-session** the timeout is absolute to the start of the session.

It can be useful to identify the AP and the SSID to which a UE is connected. Therefore, the AP MAC and SSID name can be learned as follows:

- From the called-station-id as defined in RFC 3580, *Secure/Multipurpose Internet Mail Extensions (S/MIME) Version 3.1 Certificate Handling*
- From DHCP circuit-Id or DHCPv6 interface-ID, if those options use the format specified below
- From ARP or ND over GRE as specified in section 11.10. This only identifies the AP MAC, not the SSID
- From the L2TPv3 cookie as specified in section 11.22. This only identifies the AP MAC, not the SSID

The format used for DHCP(v6) is AP-MAC;SSID-STRING;SSID-Type, where the AP-MAC should contain the AP MAC address in colon separated format (xx:xx:xx:xx:xx:xx), the SSID string should not contain the ";" delimiter and the SSID type is a single character "s" (secure) or "o" (open).

For example, if AP-MAC is "00:10:A4:23:19:C0", SSID is "SP1-Wi-Fi", and SSID-type is secure, then the value of circuit-id or interface-id would be the string "00:10:A4:23:19:C0;SP1-wifi;s".

12.10.1.1 EAP-based authentication

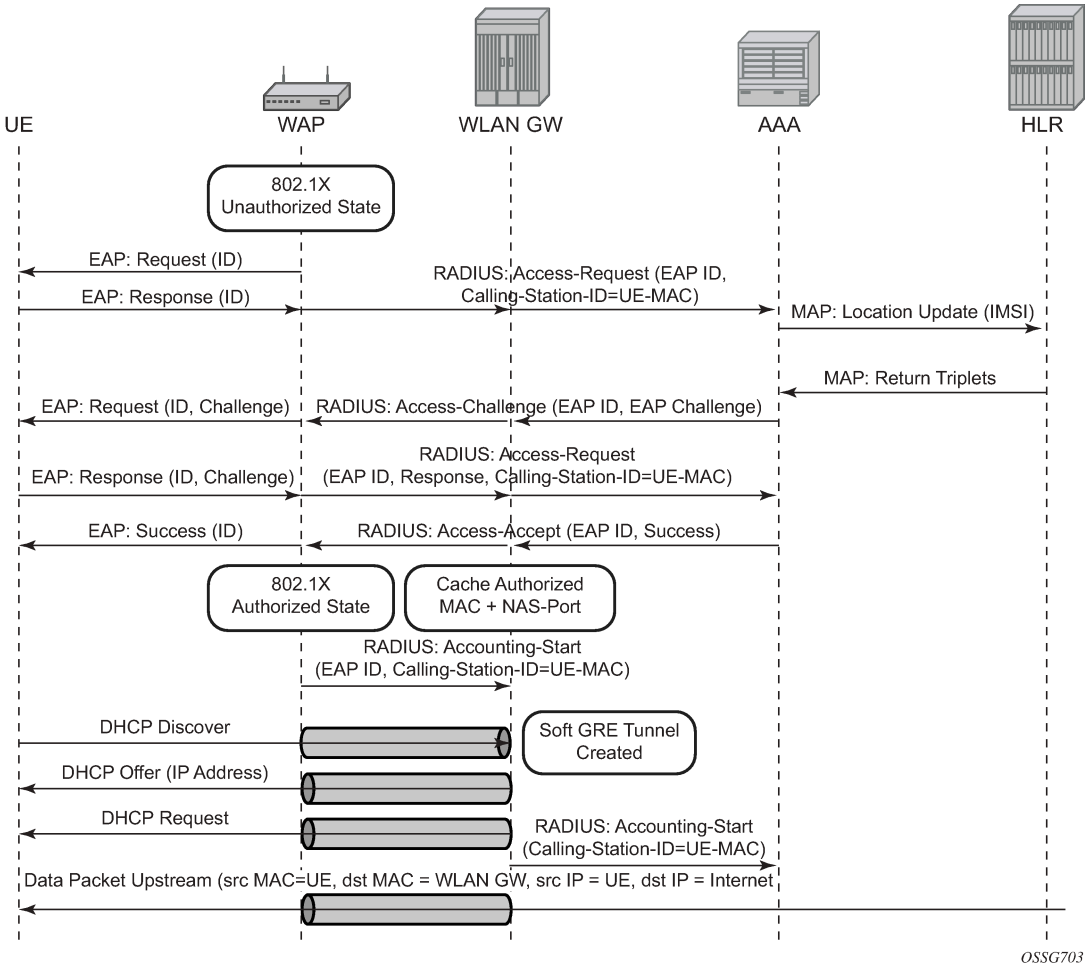
In this model the Wi-Fi AP supports a RADIUS client and originates RADIUS messages based on 802.1x/EAP exchange with the UE. It sends EAP payload in RADIUS messages toward the RADIUS server or RADIUS proxy. 7750 WLAN-GW can be configured as a RADIUS proxy for the Wi-Fi APs. The Wi-Fi AP should be configured with the IP address of the RADIUS proxy, and should send authentication

and accounting messages non-tunneled, natively routed to the RADIUS proxy. See Figure 169: EAP authentication call flow with WLAN-GW RADIUS proxy.

The RADIUS proxy function allows 7750 SR to look at the RADIUS authentication and accounting messages and create or update corresponding subscriber state. RADIUS proxy transparently forwards RADIUS messages between AP (authenticator) and the AAA server. The access-request message contains standard RADIUS attributes (including username), and the EAP payload. Standard authentication algorithms negotiated with EAP involve multiple round-trips (challenge/response) between AP (and UE) and the AAA server.

After authentication is complete, AAA server passes back subscriber related configuration parameters as well as the computed session keys (Pairwise Master Key (PMK)) for 802.11i to the AP. These keys are encrypted using shared secret between AP (authenticator) and the AAA server. 7750 SR WLAN-GW can optionally cache authentication information of the subscriber from access-request and access-accept messages. The cached information allows local authorization of subsequent DHCP messages from the UEs behind the AP against the cached state on the 7750 SR RADIUS proxy and avoids another trip to the RADIUS server.

Figure 169: EAP authentication call flow with WLAN-GW RADIUS proxy



12.10.1.1.1 RADIUS proxy

RADIUS proxy can be configured per service router (base or VPRN). The proxy acts as a server toward the Wi-Fi AP RADIUS clients, and as a client toward RADIUS servers. Therefore, both client and server parts of the RADIUS proxy need to be configured. The attribute from access-request or response message that serves as the key for the cache is configurable. The key configuration is mandatory for enabling the cache. Commonly the key is the MAC address of the UE, which is available in subsequent DHCP request, and used to locate the cache entry. The UE's MAC address is typically available in the Calling-station-Id attribute (31) in the RADIUS access-request message from the AP. The proxy can be configured for both authentication and accounting. The RADIUS server policies referred by RADIUS proxy are configured under "aaa" context. If caching is enabled in the RADIUS proxy, the subscriber attributes returned in access-accept are cached. These can include 802.1x credentials/keys, IP address or pool, DNS information, default gateway information, retail-service-id, SLA-profile, filter parameters, charging information, session keys (MS-MPPE-RECV-KEY, MS-MPPE-SEND-KEY) and so on. If subsequent DHCP DISCOVER is not received within the configured timeout, the cache entry is removed.

The following output displays a RADIUS proxy configuration.

```
config>service>ies>
config>service>vprn>
     description "Default Description For VPRN ID 50"
        interface "listening_radius_server" create
          address 10.9.9.9/32
          loopback
        exit

     radius-proxy
         server "radius_proxy" purpose accounting authentication create
             cache
                 key packet-type request attribute-type 31
                 timeout min 5
                 track-accounting stop interim-update accounting-on accounting-off
                 no shutdown
             exit
             default-accounting-server-policy "radius_acct_server_policy"
             default-authentication-server-policy "radius_Auth_server_policy"
             interface "listening_radius_server"
             load-balance-key attribute-type 102 vendor 5
             secret "AQepKzndDzjRI5g38L3LbbN3E8qualtn" hash2
             send-accounting-response
             no shutdown
         exit
```

12.10.1.1.1.1 RADIUS proxy — server load-balancing

RADIUS proxy can be configured for load-balancing to multiple authentication and accounting servers. Load-balancing can be round-robin or hash-based and is configured via access-algorithm under RADIUS policy. With round-robin the first RADIUS request is sent to the first server, the second request to the second server and so on. With hash, it is possible to load-balance subscribers across a set of servers. Based on the configured hash key, configured in the RADIUS proxy, it can be ensured that all RADIUS messages for a single subscriber are sent to the same server. The hash key can include any specified standard or vendor-specific RADIUS attribute. An example is calling-station-id which contains subscriber's MAC address).

If the hash lookup causes the request to be sent to a server that is currently known to be unresponsive, a second hash lookup is performed that only takes the servers into account that are not known to be unresponsive. This is done to maximize the likelihood that all requests end on the same server. If all configured servers are known to be unresponsive, the RADIUS proxy falls back to the round-robin algorithm with the starting point determined by the first hash lookup to maximize the chance of getting any response to the request.

The following output displays a RADIUS server and policy configuration for servers referred from the RADIUS proxy.

```
config>service>vprn
   radius-server
      server "radius_server" address 10.100.100.2 secret "9OkclHYDDbo9eHrzFmuxiaO/LAft3Pw"
                             hash2 port 1812 create
      exit
   exit

config>aaa
   radius-server-policy "radius_server_policy" create
      servers
         router 50
         access-algorithm hash-based
         source-address 10.1.1.1
         timeout min 1
         hold-down-time 2
         server 1 name "radius_server"
      exit
```

12.10.1.1.1.2 RADIUS proxy — cache lookup

Local-user-database can be programmed to associate a host match with the RADIUS proxy cache instance. The host-match criterion is configurable, based on a subscriber attribute from the DHCP request.

The following output displays a RADIUS proxy cache lookup configuration.

```
config>subscriber-mgmt
   local-user-db "radius_ludb" create
       ipoe
           match-list service-id
           host "default" create
           auth-policy "auth_policy_1"
           match-radius-proxy-cache
              fail-action continue
              match mac
              server router 50 name "radius_proxy"
           exit
              no shutdown
        exit
        no shutdown
    exit
exit
```

If caching is enabled in the RADIUS proxy, then the actions on receiving DHCP message for the authenticated client includes the following:

- A host lookup is done in the local-user-database to find the RADIUS proxy cache for the subscriber.

- The field used to lookup the cache is configurable. It can include circuit-id or remote-id (present in sub-option in DHCP option-82), MAC@ or one of the other options in the DHCP packet. If a match is not found, the configured fail-action is executed. The default match field is MAC@. If the configured fail-action is "drop", the DHCP DISCOVER is dropped. If the configured fail-action is "continue", then the ESM host creation proceeds based on the authentication policy configured under the group-interface on which the DHCP packet is received.
- If a match is found, the parameters from original authentication accept in the cache are used to create the ESM host. If the group-interface is wlan-gw, then the ESM host is associated with the wlan-gw tunnel the (AP's WAN IP@) and corresponding AP (MAC@ from the called-station-id in the authentication state).

12.10.1.1.1.3 RADIUS proxy — accounting

An ESM accounting-start is generated after the ESM host is created on successful authorization of DHCP against cached authentication state, and IP@ allocation is complete. The accounting-start contains information from locally cached 802.1x/EAP authentication such as calling-station-id, called-station-id, NAS-port-id, Subscriber-profile, SLA-profile, NAT port range for subscriber-aware NAT and so on.

If RADIUS proxy is configured as an accounting proxy in addition to authentication proxy, then the RADIUS proxy transparently forwards the accounting messages to the authentication servers referred from the RADIUS proxy and can also load-balance. If caching is enabled, then the proxy can be configured to also track and locally act on the accounting messages, while still transparently forwarding these messages. The possible actions if accounting messages are tracked include the following:

- **Accounting-start**

  The Wi-Fi AP RADIUS client generates an accounting-start when a UE has successfully authenticated and associated with the AP. In cases where after mobility, the new AP does not re-authenticate because of key caching. accounting-start can be used as a mobility trigger on the WLAN-GW. Also, in cases where a UE associates with a single AP but pre-authenticates with multiple APs in range, tracking mobility based on authentication can falsely associate a UE with incorrect AP. Mobility tracking based on authentication can be disabled via CLI (no track-authentication under radius-proxy cache), and instead be performed based on accounting-start. On receiving accounting-start, the RADIUS proxy on WLAN-GW finds the corresponding ESM host based on the calling-station-id attribute (typically the MAC@) of the subscriber) in accounting-start and associates the UE with the RADIUS client (for example, Wi-Fi AP).

- **Accounting-stop**

  The Wi-Fi AP RADIUS client generates an accounting stop if it detects the UE has disassociated or is deleted because of inactivity or session timeout. The RADIUS proxy finds the corresponding ESM host based on the calling-station-id (typically the MAC@) of the subscriber. Note that if the called-station-id is filled out this must also match with what is currently stored as a security measure. When a UE moves the called-station-id should get updated and therefore an accounting-stop from a previous AP cannot delete this UE anymore.

- The ESM host is deleted, an ESM accounting-sop message is sent, and the accounting-stop message from the AP is forwarded to the accounting-server.

- **Accounting-ON or Accounting-OFF**

  This would be received from the AP if the AP has restarted. The RADIUS proxy finds all the impacted subscribers for the AP based on the called-station-id attribute (the AP's MAC@) in the accounting message and delete all the corresponding ESM hosts.

- **Interim Accounting Updates**

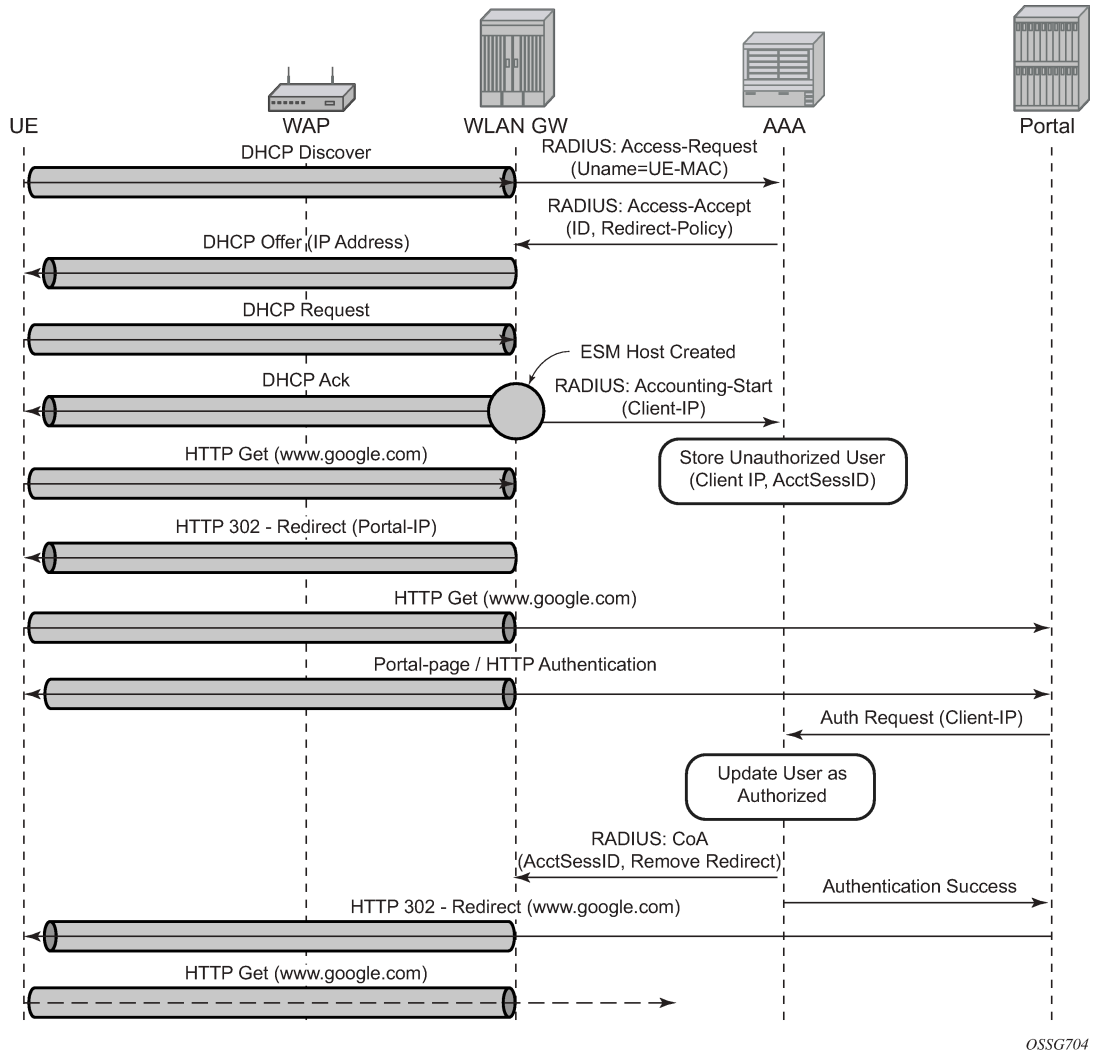
  If the client moves and re-associates with a new AP, the RADIUS client in the new AP generates interim-update. The RADIUS proxy locates the impacted ESM host and update its state to point to the new AP's MAC@ (as available in called-station-id in the accounting message). The ESM interim-updates to accounting servers are sent on scheduled interval configured in accounting-policy, but with the updated information from the interim updates received from the AP.

12.10.1.2 Portal authentication

For SSIDs without 802.11i/WPA2-based key exchange and encryption, it is common to authenticate the user by directing user's HTTP traffic to a portal, where the user is prompted for its credentials, which are verified against a subscriber database. The backend can optionally remember the MAC@ and subscriber credentials for a set time period such that subsequent logins of the user do not require portal redirection. Some UEs support a client application (aka WISPr client), which automatically posts subscriber credentials on redirect, and parse HTTP success or failure response from the portal server.

7750 WLAN-GW uses existing http-redirect action in IP filter to trigger redirect port-80 traffic. In case of open SSID, on receiving DHCP DISCOVER, MAC based authentication is performed with the RADIUS server as per configured authentication policy. The SLA-profile returned from RADIUS server in authentication-accept (or the default SLA-profile) contains the filter with http-redirect. Redirect via HTTP 302 message to the UE is triggered from the CPM. After the user posts its credentials, RADIUS server generates a CoA-request message removing the http-redirect by specifying an SLA-profile without redirect action. If the portal authentication fails, the RADIUS server generates a disconnect-request message to remove the ESM host. In case of wlan-gw tunnel from the AP, the DHCP messages and data are both tunneled to the WLAN-GW. See Figure 170: Portal authentication for open SSIDs.

Figure 170: Portal authentication for open SSIDs



The following output displays a portal authentication for open SSIDs configuration example.

```
config>subscriber-mgmt
      sla-profile "portal-redirect" create
          ingress
             ip-filter 10
          exit
      exit
   exit

system>config>filter
   ip-filter 10 create
       entry 1 create
            match protocol udp
                dst-port range 67 68
            exit
            action forward
       exit
       entry 2 create
```

```
            match protocol tcp
                dst-port eq 80
            exit
            action http-redirect "http://www.google.ca"
        exit
    exit
exit
```

It is possible to view the subscriber HTTP redirect statistics by using the **show service id** *id* **subscriber-hosts statistics** command. The statistics are captured per host and supports both IPv4 and IPv6. This command is only supported from CPM5 and up and SR-e platforms.

12.10.1.3 AA-based portal redirection

An application profile can be provided using RADIUS to be used during portal authentication. The WLAN-GW ISA redirects all traffic to a linked AA ISA, where all AA functionalities can be used to enable a more extensive portal redirect. The WLAN-GW redirect and filter are completely bypassed, but the UE is kept in a portal authentication state. The UE can subsequently be promoted to DSM or ESM with a CoA as with regular portal functionality. When promoting to DSM, AA is disabled unless a DSM application profile is provisioned using configuration or RADIUS.

See Application Assurance support for more information about enabling AA functionality for a WLAN-GW.

12.10.2 Address assignment

The address to the UEs can be assigned via local DHCP server from locally defined pools, or from RADIUS server via local DHCP proxy, or from an external DHCP server. Subscriber-interface and group-interface are configured as part of normal ESM configuration. In the case of wlan-gw, the group interface is wlan-gw enabled. Subnets on the subscriber interface are used for the pools from which the DHCP local server assigns addresses to UEs.

The following output displays an address assignment configuration example.

```
config>service>vprn
    dhcp
       local-dhcp-server "dhcp" create    #### create local DHCP server
           pool “1” create                #### define Pool
               options
                   dns-server 10.8.8.8 8.8.4.4
                   lease-time min 5
               exit
               subnet 203.0.113.255/30 create
                  options
                     subnet-mask 239.255.0.0
                     default-router 10.203.254.181
                  exit
                  address-range 128.203.254.182 128.203.254.183
               exit
             exit
         exit
     exit

    interface "DHCP-lb" create          #### loopback interface with DHCP server
        address 10.1.1.1/32
        local-dhcp-server "dhcp"
        loopback
```

```
    exit

    subscriber-interface "sub-int" create          #### subscriber interface
         address 172.31.235.235/30                #### Subnets out of which UE
         address 172.31.245.245/16                ######  addresses are allocated.
         group-interface "group-int" wlangw create
             sap-parameters
                 sub-sla-mgmt
                   def-sla-profile "sla_def"
                   def-sub-profile "sub_def"
                   sub-ident-policy "sub_ident"
                 exit
              exit
         exit
         dhcp
             proxy-server
                 emulated-server 10.10.0.1   #### proxy to get IP address from AAA
                 lease-time min 5           #### or from DHCP server. Can provide
                 no shutdown        #### split lease (shorter lease towards client,
              exit                 #### and longer lease towards AAA or DHCP server.
              no option
              server 10.1.1.1               #### DHCP local server
                  trusted
                  lease-populate 32000
                  gi-address 172.31.245.245
                  user-db "radius_ludb"      #### LUDB for proxy cache co-relation
                  no shutdown
               exit
        exit
```

12.10.3 Wholesale

With EAP the AAA server can look at the realm from the user credential (IMSI) in authentication request and appropriately provide the service context in retail-service-id, for the ESM host corresponding to the UE.

For open SSID, the decision can be made by the AAA server based on the SSID, as learned by any of the methods described in Authentication.

The SSID is passed to the AAA server in initial MAC based authentication on DHCP DISCOVER. The retail-service-id can be returned in access-accept. This assumes the AP broadcasts unique SSID per retail provider and inserts it in Option82 as a DHCP relay-agent. As an alternative to SSID in option-82, the AP can insert a unique dot1Q tag per retail provider, before tunneling the Ethernet frame, using single GRE tunnel per AP to the WLAN-GW. 7750 supports configuring a map of dot1Q tags to retail-service-id. Therefore, the determination of the retail provider for the subscriber can be made in the data plane when DHCP is received, and the subscriber state can be created and processed in the right service context.

The following output displays a wholesale configuration example.

```
config>service>ies>
config>service>vprn>
subscriber-interface <ip-int-name>
group-interface <ip-int-name> wlangw
wlan-gw
[no] router (base | <vprn-id>) # tunnel service context
[no] wlan-gw-group <group-id>
....snip
vlan-tag-ranges # Precedence for retail-service-id:
# RADIUS, vlan-retail-service-map, default-retail-svc
  [no] vlan start <start-tag> end <end-tag> retail-svc-id <svc-id>
  [no] default-retail-svc-id
```

```
exit
exit
exit
```

12.10.4 3G/4G interworking

The WLAN-GW supports 3G/4G interworking based on a per-UE GTP tunnel from WLAN-GW to the mobile packet core (GGSN or P-GW). For more details on the GTP uplink setup, see the GTP section. The bridged Wi-Fi AP connectivity with the WLAN-GW can be WLAN-GW-based (soft-tunnel/l2-ap), or it can be a native Layer 2 (regular group-interface).

12.10.4.1 Signaling call flow

The decision to setup a GTP tunnel for a subscriber or locally breakout subscriber's traffic is AAA based and received in authentication response. If the traffic is to be tunneled to the PGW or GGSN, the signaling interface or PGW/GGSN interface would be provided via AAA. Absence of these attributes in the authentication response implicitly signifies local-breakout.

12.10.4.2 GTP setup with EAP authentication

After the EAP authentication completes as described in the section on authentication, the RADIUS proxy caches the authentication response, including any attributes related to GTP signaling. Subsequently DHCP is initiated from the UE. On receiving DHCP DISCOVER, the RADIUS proxy cache is matched to get the AAA parameters related to the UE from the original authentication response. If PGW/GGSN (mobile gateway) IP address is not present in cached authentication, DNS resolution as described in section 1.2 is initiated for the WLAN APN obtained from AAA (in the cache) or for locally configured APN in the service associated with the UE. The DNS resolution provides a set of IP addresses for the mobile gateways. The GTP tunnel setup is attempted to the selected mobile gateway. The IP address provided by PGW/GGSN in the GTP response is returned in DHCP offer to the UE. The WLAN-GW acts as a DHCP to GTP proxy. The WLAN-GW is the default-GW for the UE. Any packets from the UE are then GTP tunneled to the mobile gateway. If the UE requests an IP address (for which it may have an existing lease on one of its interface) via DHCP option 50 in the DHCP request, then WLAN-GW sets the "handover bit" in the GTP session create message, and indicates the requested address in the PDN Address Allocation (PAA) field. This allows the PGW to look for existing session corresponding to the signaled IMSI and APN (with potentially different RAT-Type) and return its existing IP address in session create response. The old session and bearer is deleted by the PGW. The signaling of "handover bit" is supported with S2a and S2b (Release 10.0 and beyond). The IP address cannot be preserved over the Gn interface. The call flow in Figure 168: Inter WLAN-GW mobility based on data-triggered authentication and subscriber creation shows basic GTP setup (with S2a), the output provided on Figure 168: Inter WLAN-GW mobility based on data-triggered authentication and subscriber creation shows IP address preservation across inter-access (WIFI <-> 4G) moves.

DHCP release or lease timeout on WLAN-GW results in deletion of the GTP tunnel corresponding to the UE. The session or PDP context deactivation from PGW/GGSN also results in removal of the GTP state for the UE and the corresponding ESM host on WLAN-GW. Only the default bearer (or primary PDP context) for single default APN is handled over Wi-Fi. GTP path-management messages (echo request and reply) are supported. Mandatory IEs are supported in GTP signaling. Hard coded default values are signaled for QoS and charging related IEs. For GTPv2, the bearer is signaled as non-GBR bearer with QCI value of 8, and MBR/GBR values of 0. APN-AMBR default values signaled are 20 Mb/s / 10 Mb/s downstream/

upstream. For GTPv1, reliability and priority classes default to "best-effort", allocation/retention priority defaults to 1, and the default peak-rate corresponds to class 9 (bit-wise 1001) which is slightly over 2 Mb/s. Charging characteristics IE which contains a 16 bit flag defaults to 0. In the future, RADIUS returned values or locally configurable values is signaled in QoS and charging IEs.

The IP address is returned in the create PDP context response or Create session response. The DNS server addresses for the UE are returned in IP control protocol (IPCP) option in a PCO IE in the response. The default gateway address provided to the UE in DHCP is auto-generated algorithmically on the WLAN-GW from the IP address returned by the PGW/GGSN for the UE. The Wi-Fi AP is required to provide a split-horizon function, where there is no local switching on the AP, and all communication to/from any AP is via WLAN-GW. The WLAN-GW implements proxy-ARP and forwards all received traffic from the UE into the GTP tunnel. In the future, the default-GW address to be returned to the UE could be obtained in a PCO from the PGW/GGSN. The GTP-U processing of data packets is done in the IOM.

12.10.4.3 Location notification in S2a

This feature adds support on WLAN-GW for reporting UE's WLAN location (TWAN Identifier IE) and cellular location (ULI IE) over S2a interface to PGW and UE's cellular location (ULI IE) to GGSN (over Gn interface). Location information is useful for charging on PGW/GGSN.

12.10.4.3.1 WLAN location over S2a

The WLAN location information consists of the *TWAN Identifier IE* as described in 29 274 V11.6.0 (2013-04) and is sent in GTPv2 "create session request" message. If present, this IE carries BSSID (MAC address of the AP) and the SSID. WLAN-GW learns the AP's MAC@ from calling-station-id attribute in the RADIUS messages from the AP (both authentication and accounting messages) or from circuit-id in DHCP DISCOVER or REQUEST messages. The IE is only sent at session creation time. Therefore, it reports location on initial attach, on handover from LTE to Wi-Fi, and on AP mobility across WLAN-GWs. Mobility across APs anchored on the same WLAN-GW does not result in location update. 3GPP Release 11 does not define location update mechanism for S2a.

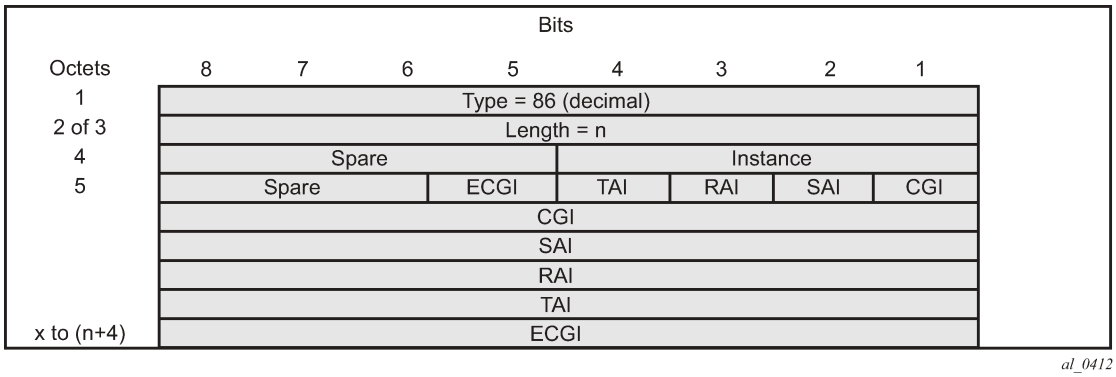
By default, location is not reported. It can be enabled with CLI.

```
config>subscr-mgmt>gtp
   peer-profile "pgw-west-mn01"
      [no] report-wlan-location
```

12.10.4.3.2 Cellular location over S2a

The "User Location Info" IE is included in "Create Session Request and is described in 3GPP TS 29.274 version 8.1.1 Release 8. The encoding for individual location identifiers (CGI, SAI, RAI, TAI, and ECGI) is also defined in the same reference (as shown in Figure 171: User location information).

Figure 171: User location information

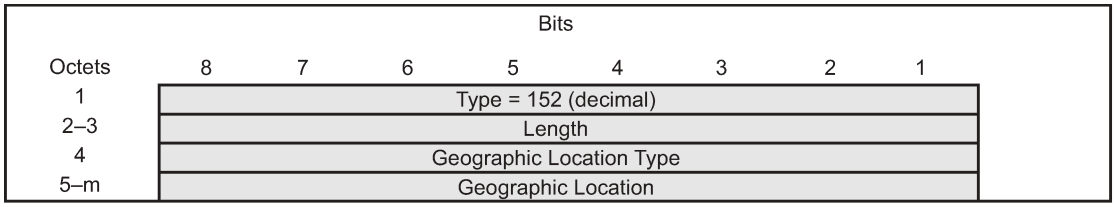


The AP's MAC@ and IP@ are provided to AAA server in RADIUS messages during EAP authentication and accounting. If AAA provides the cellular location (corresponding to this AP) in 3GPP attribute *3GPP-User-Location-Info* in access-accept, and location reporting is enabled via CLI. The ULI IE is included in GTPv2 "create session request". The *3GPP-User-Location-Info* attribute is described in 3GPP TS 29.061 v9.3.0.

12.10.4.3.3 Cellular location over Gn interface

The "User Location Info" IE (as shown in Figure 172: User location information IE) can be included in create-pdp-context message as described in 3GPP TS 29.060 V10.1.0. The geographic location type field describes the type of location included in the "Geographic Location" field that follows. The location can be CGI (cell global identification), SAI (service area identity), or RAI (routing area identity). The formats for these location identifiers are defined in the same reference *3GPP TS 29.060 V10.1.0*.

Figure 172: User location information IE



AP MAC address and SSID is reported to AAA (including changes on mobility). AAA can then specify the ULI IE contents based on static mapping of AP's MAC address to one of the cellular location types (CGI, SAI or RAI). AAA should provide the cellular location in 3GPP attribute 3GPP-User-Location-Info (below) in access-accept. The attribute is described in 3GPP TS 29.061 v9.3.0.

In case a UE moves to a different WLAN-GW, UE is authenticated based on data-trigger. In this case, the AAA server can provide the WLAN location (AP's MAC@ and SSID) in called-station-ID attribute and cellular location in 3GPP-User-Location-Info attribute. The WLAN location is then encoded in TWAN identifier in "create session request" message, and the cellular location is encoded in the ULI IE.

12.10.4.3.4 Operational support

The following command shows state of location reporting (enabled/disabled).

```
*A:Dut-C>config>subscr-mgmt>gtp>peer-profile$ /show subscriber-mgmt gtp
peer-profile "test"

===============================================================
GTP peer profile "test"
===============================================================
Description                 : (Not Specified)
Retransmit timeout (s)      : 5
Retransmit retries          : 3
Keepalive interval (s)      : 60
Keepalive retries           : 4
Keepalive retry timeout (s) : 5
Time to live                : 255
Interface type              : s2a
Charging char home UE       : (None)
Charging char roaming UE    : (None)
Session hold time (s)       : 30
Report WLAN location        : enabled
Last management change      : 02/21/2014 16:31:12
GGSN uplink GBR (Kbps)      : 5000
GGSN uplink MBR (Kbps)      : 5000
GGSN downlink GBR (Kbps)    : 2000
GGSN downlink MBR (Kbps)    : 2000
GGSN Alloc/Retention Prio   : 1
GGSN last management change : 02/19/2014 17:31:55
PGW uplink GBR (Kbps)       : 0
PGW uplink MBR (Kbps)       : 0
PGW downlink GBR (Kbps)     : 0
PGW downlink MBR (Kbps)     : 0
PGW Alloc/Retention Prio    : 1
PGW Qos Class ID            : 8
PGW last management change  : 02/19/2014 17:31:55
===============================================================
```

12.10.5 CGN on WLAN-GW

Both LSN and Layer 2–aware NAT for Wi-Fi subscribers over wlan-gw tunnels is supported. NAT on WLAN-GW is only supported for locally terminated subscribers and not for GTP tunneled subscribers. NAT can be performed on the same set of ISAs that are used for WLAN-GW functions, by referring to the WLAN-GW ISA group from NAT configuration. Alternatively, dedicated set of ISAs can be used for NAT function by creating and referencing a separate NAT-group. Configuration related to LSN and Layer 2–aware NAT is provided in the *7450 ESS, 7750 SR, and VSR Multiservice ISA and ESA Guide*.

12.10.6 Lawful Intercept on WLAN-GW

Mirroring traffic for Wi-Fi subscribers to a mediation device, when the subscriber is under legal intercept is supported. The mirroring function is performed on the anchor IOM where the subscriber is anchored. Both Ether and IP-only mirror is supported. With Ether mirror, VLAN tags which are part of internal SAP between ISA and IOM, are included in the mirrored Ethernet frame of the subscriber. IP-only mirror includes the IP header and the payload. Conventional IP-only mirror service can be used with direct p2p or MPLS (for remote mirroring) connection to the mediation device. Routable-encapsulation is also supported. Both IP/

UDP encapsulation with optional shim-header for subscriber correlation on the mediation device, and IP/GRE encapsulation is supported with routable-encapsulation of mirrored data. LI can be triggered via CLI, SNMPv3 or RADIUS, as supported with ESM. RADIUS triggered LI can be via LI related VSAs in access-accept or in CoA. The CoA is keyed on accounting-session-id. LI is supported for both local and GTP tunneled subscribers.

Existing LI support with ESM is described in the *7450 ESS, 7750 SR, 7950 XRS, and VSR OAM and Diagnostics Guide*.

12.10.7 Tunnel level egress QoS

Downstream traffic can be subjected to aggregate rate-limit per tunnel or per tunnel and per retailer combination (in case of wholesale). Typically a unique SSID is used per retailer for wholesale on the AP and is reflected via unique dot1Q tag. With a wlan-gw tunnel per AP, the tunnel encapsulation is performed on the tunnel ISA. The downstream traffic on the tunnel IOM is received over B-VPLS from the anchor IOM, and is MAC-in-MAC (802.1ah) encapsulated. I-SID in the packet represents the GRE tunnel or tunnel and retailer combination. SAP-egress QoS policy defining queues (with rates), and FC to queue mapping, can be specified under the wlan-gw interface. This policy is applicable to all tunnels (or tunnel and SSID combinations) associated with the wlan-gw interface and is attached to corresponding I-SIDs on the B-VPLS SAP. Traffic is shaped into these queues based on configured queue rates. An aggregate rate-limit applied across queues on an I-SID (representing tunnel or tunnel and retailer combination) can be configured under the wlan-gw interface (represented by the wlan-gw node under the group-interface configuration). The aggregate rate-limit works in conjunction with a port-scheduler. The port-scheduler corresponds to the internal port between tunnel ISA and its carrier IOM and is specified at the wlan-gw IOM group level. The rate-limit includes the B-VPLS encapsulation overhead. The configuration is shown in Figure 169: EAP authentication call flow with WLAN-GW RADIUS proxy. Queues per I-SID also work with virtual-scheduler (with or without a port scheduler). Virtual-scheduling and aggregate-rate enforcement are mutually exclusive. Configuration is shown in Figure 170: Portal authentication for open SSIDs. Egress SAP QoS policy, aggregate rate-limit, port-scheduler, and virtual-schedulers are described in the *7450 ESS, 7750 SR, 7950 XRS, and VSR Quality of Service Guide*. The SAP egress QoS policy associated with a wlan-gw interface implicitly creates queues (and scheduler association) on ISIDs as corresponding wlan-gw tunnels are created. General ISID queuing and shaping is defined in the SR OS Services Guide.

A configuration node under wlan-gw interface (egress) controls where the egress shaping is applied and can specify either tunnel or retailer (tunnel and retailer combination in case of wholesale). Per I-SID shaping resources can be held after the last subscriber on the tunnel is deleted, for a configurable amount of time (hold-time) configured under the wlan-gw interface. During ISA or IOM failover the tunnel resources on the IOM kept because the hold-time are reclaimed. ISID shaping can be configured (via knob shape-multi-client) to be applied only when there is more than one UE on the corresponding tunnel (or tunnel and retailer combination). A total of 40,000 shaped tunnels (or shaped tunnel and retailer combinations) are supported per WLAN-GW IOM. Hardware resources for tunnel (ISID) shapers are shared with subscribers. With 3 WLAN-GW IOMs per chassis, a maximum of 98,000 (3 *64K / 2) shaped tunnels and subscribers can be supported per chassis.

The following output depicts per tunnel or per tunnel/SSID egress QoS (with aggregate-rate and port-scheduler).

// Port-scheduler

```
config>qos#
      port-scheduler-policy “lo-gre-port-sched”
           max-rate 5000
```

```
           level 1 rate 1000 cir-rate 1000
           level 8 rate 500 cir-rate 500
      exit
exit
```

// Egress queues (per ISID) parented by port-scheduler specified under associated wlan-gw interface

```
config>qos>
   sap-egress 3 create
       queue 1 create
          rate 300
          port-parent level 1 weight 10 cir-level 1 weight 10
       exit
       queue 2 create
          rate 100
          port-parent level 8 weight 10 cir-level 8 weight 10
       fc af create
           dot1p 2
           de-markweight
       exit
       fc be create
          queue 1
          dot1p 0
          de-mark
       exit
       fc ef create
           queue 2
           dot1p 5
           de-mark
       exit
   exit
exit
```

// The wlan-gw interface refers to SAP egress QoS policy and aggregate rate-limit for associated ISIDs

```
config>service>ies>sub-if>grp-if>wlan-gw>egress
      agg-rate-limit 2000
      hold-time 300
      qos 3
      shaping per-tunnel
      shape-multi-client
exit
```

// Port-scheduler parenting queues (per ISID)

Per Tunnel or Per Tunnel/SSID Egress QoS (with aggregate-rate and port-scheduler)

```
config>isa>wlan-gw-group#
      active-iom-limit 1
      tunnel-port-policy "lo-gre-port-sched"
      iom 2
      iom 3
      no shutdown
exit
```

The following output depicts per tunnel or per tunnel/SSID egress QoS (with virtual-scheduler).

```
// hierarchical virtual scheduler
config>qos#
      scheduler-policy “virtual-sched-policy”
           tier1
```

```
                scheduler "all-traffic" create
                     rate 10000
                exit
            exit
            tier2
                scheduler "non-voice" create
                    parent all-traffic cir-level 1
                    rate 9000
                exit
                scheduler "voice" create
                   parent all-traffic level 2 cir-level 2
                   rate 3000
                exit
            exit

        exit
```

// egress queues (per ISID) parented by virtual scheduler

```
config>qos>
   sap-egress 3 create
       queue 1 create
          parent "non-voice"
          rate 2000 cir 1000
       exit
       queue 2 create
          parent "voice"
          rate 500 cir-rate 500
       fc be create
          queue 1
          dot1p 0
          de-mark
       exit
       fc ef create
           queue 2
           dot1p 5
           de-mark
       exit
   exit
exit
```

// A wlan-gw interface refers to SAP egress QoS policy and hierarchical scheduler for associated ISIDs

```
config>service>ies>sub-if>grp-if>wlan-gw>egress
      hold-time 300
      qos 3
      scheduler-policy "virt-sched-policy"
      shaping per-tunnel
      shape-multi-client
exit
```

12.10.7.1 QoS overrides

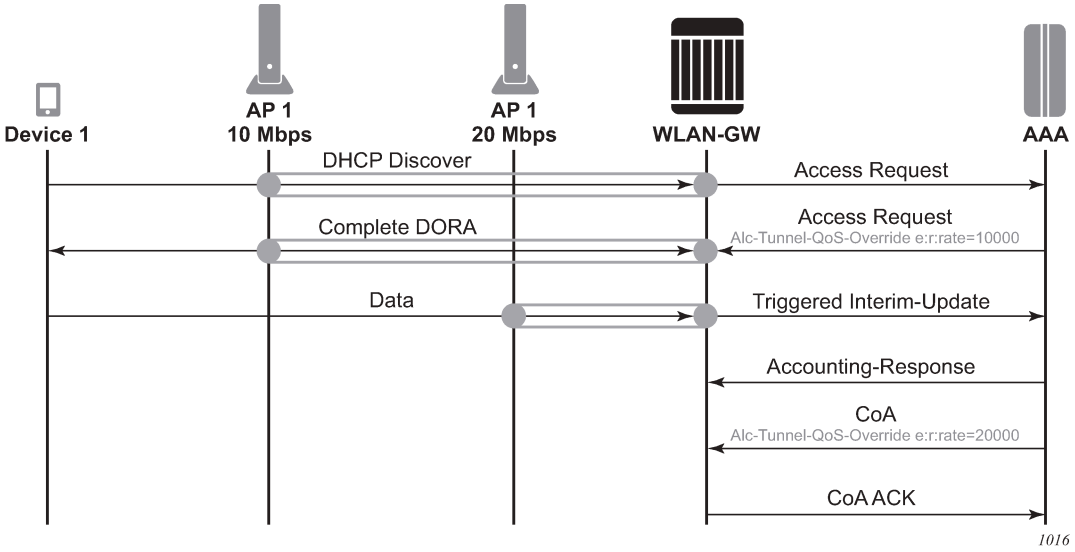
Per-tunnel QoS overrides can be provided to have more specific values for specific tunnels. This allows the WLAN gateway to accommodate a heterogeneous set of access points in one Wi-Fi network; for example, provisioning different bandwidth for homespots and hotspots. For a detailed list of allowed overrides and values, see the *7450 ESS, 7750 SR, and VSR RADIUS Attributes Reference Guide*.

QoS overrides can be signaled via RADIUS in UE authentication and in UE CoA. The override is applied to the tunnel where the UE is currently active. Therefore, two scenarios should be considered:

- If the first UE on a tunnel is also a new UE that is authenticated, any tunnel QoS override can be signaled in the Access-Accept message for that UE.
- If the first UE on a tunnel is a previously authenticated UE that moved from another tunnel, a mobility triggered Interim Update can be sent and used as a trigger to send a CoA with the per-tunnel QoS overrides.

Figure 173: QoS override flows shows flows for both scenarios.

Figure 173: QoS override flows



To guarantee that the CoA is applied to the correct tunnel, use the CoA key consisting of the *nas-port-id* and the UE IP address; both can be derived from the interim update. In the WLAN gateway, the *nas-port-id* identifies the tunnel to which the UE is linked, and, if a mismatch is detected, the CoA is not applied. Other CoA keys are also supported but do not offer this guarantee.

12.10.7.2 Operational commands

Egress per tunnel (or per tunnel, per SSID) QoS with aggregate rate-limit and port-scheduler.

```
show router 50 wlan-gw soft-gre-tunnels detail
===============================================================================
Soft GRE tunnels
===============================================================================
Remote IP address           : 239.1.1.2
Local IP address            : 10.1.1.1
ISA group ID                : 1
ISA group member ID         : 1
Time established            : 2012/06/19 20:31:36
Number of UE                : 1

Tunnel QoS
----------
Operational state           : active
```

```
Number of UE                : 1
Remaining hold time (s)     : N/A
Service Access Points(SAP)
===============================================================================
Service Id         : 2147483650
SAP                : 2/1/lo-gre:1            Encap            : q-tag
Description        : Internal SAP
Admin State        : Up                      Oper State       : Up
Flags              : None
Multi Svc Site     : None
Last Status Change : 06/19/2012 07:13:31
Last Mgmt Change   : 06/19/2012 20:30:24
-------------------------------------------------------------------------------
Encap Group Specifics
-------------------------------------------------------------------------------
Encap Group Name   : _tmnx_SHAPER_GR000       Group Type       : ISID
Qos-per-member     : TRUE
Members            :1
-------------------------------------------------------------------------------
QOS
-------------------------------------------------------------------------------
E. qos-policy      : 3                       Q Frame-Based Acct: Disabled
E. Sched Policy    :                         E. Agg-limit     : 4000
-------------------------------------------------------------------------------
Encap Group Member 1 Base Statistics
-------------------------------------------------------------------------------
Last Cleared Time     : N/A

Forwarding Engine Stats
                        Packets                 Octets
For. InProf           : 0                       0
For. OutProf          : 0                       0
Dro. InProf           : 0                       0
Dro. OutProf          : 0                       0
-------------------------------------------------------------------------------
Encap Group Member 1 Queue Statistics
-------------------------------------------------------------------------------
                        Packets                 Octets
Egress Queue 1
For. InProf           : 0                       0
For. OutProf          : 0                       0
Dro. InProf           : 0                       0
Dro. OutProf          : 0                       0
===============================================================================
-------------------------------------------------------------------------------
No. of tunnels: 1
===============================================================================


show qos scheduler-hierarchy sap 2/1/lo-gre:1 encap-group "_tmnx_SHAPER_GR000" member 1 detail
===============================================================================
Scheduler Hierarchy - Sap 2/1/lo-gre:1
===============================================================================
Egress Scheduler Policy :
-------------------------------------------------------------------------------
Legend :
(*) real-time dynamic value
(w) Wire rates
B   Bytes
-------------------------------------------------------------------------------
Root (Egr)
| slot(2)
|--(Q) : -2147483646->2/1/lo-gre:1->EG(_tmnx_SHAPER_GR000):1->1  (Port 2/1/lo-gre Orphan)
|   |    AdminPIR:10000000   AdminCIR:0
```

```
|   |    AvgFrmOv:100.00
|   |    AdminPIR:10000000(w) AdminCIR:0(w)
|   |    CBS:0 B              MBS:12582912 B
|   |    Depth:0 B            HiPrio:1376256 B
|   |    MaxAggRate:4000(w)    CurAggRate:0(w)
|   |
|   |    [Within CIR Level 0 Weight 0]
|   |    Assigned:0(w)        Offered:0(w)
|   |    Consumed:0(w)
|   |
|   |    [Above CIR Level 1 Weight 0]
|   |    Assigned:4000(w)     Offered:0(w)
|   |    Consumed:0(w)
|   |
|   |    TotalConsumed:0
|   |    OperPIR:4000         OperCIR:0
|   |
|   |    PktByteOffset:add 0*
|   |    OnTheWireRates:false
|   |    ATMOnTheWireRates:false
|   |    LastMileOnTheWireRates:false
```

Egress per tunnel (or per tunnel, per SSID) QoS with hierarchical virtual scheduler.

```
show router 50 wlan-gw soft-gre-tunnels detail
===============================================================================
Soft GRE tunnels
===============================================================================
Remote IP address           : 239.1.1.2
Local IP address            : 10.1.1.1
ISA group ID                : 1
ISA group member ID         : 1
Time established            : 2012/06/19 20:43:03
Number of UE                : 1

Tunnel QoS
----------
Operational state           : active
Number of UE                : 1
Remaining hold time (s)     : N/A
Service Access Points(SAP)
===============================================================================
Service Id         : 2147483650
SAP                : 2/1/lo-gre:1             Encap             : q-tag
Description        : Internal SAP
Admin State        : Up                       Oper State        : Up
Flags              : None
Multi Svc Site     : None
Last Status Change : 06/19/2012 07:13:31
Last Mgmt Change   : 06/19/2012 20:30:24
-------------------------------------------------------------------------------
Encap Group Specifics
-------------------------------------------------------------------------------
Encap Group Name   : _tmnx_SHAPER_GR000       Group Type        : ISID
Qos-per-member     : TRUE
Members            :
1
-------------------------------------------------------------------------------
QOS
-------------------------------------------------------------------------------
E. qos-policy      : 3                        Q Frame-Based Acct: Disabled
E. Sched Policy    : virtual_scheduler_policy E. Agg-limit      : -1
```

```
-------------------------------------------------------------------------------
Encap Group Member 1 Base Statistics
-------------------------------------------------------------------------------
Last Cleared Time     : N/A

Forwarding Engine Stats
                        Packets                 Octets

For. InProf           : 2                       752
For. OutProf          : 0                       0
Dro. InProf           : 0                       0
Dro. OutProf          : 0                       0
-------------------------------------------------------------------------------
Encap Group Member 1 Queue Statistics
-------------------------------------------------------------------------------
                        Packets                 Octets

Egress Queue 1
For. InProf           : 2                       752
For. OutProf          : 0                       0
Dro. InProf           : 0                       0
Dro. OutProf          : 0                       0
===============================================================================
-------------------------------------------------------------------------------
No. of tunnels: 1
===============================================================================


show qos scheduler-hierarchy sap 2/1/lo-gre:1 encap-group "_tmnx_SHAPER_GR000" member 1 detail
===============================================================================
Scheduler Hierarchy - Sap 2/1/lo-gre:1
===============================================================================
Egress Scheduler Policy :
-------------------------------------------------------------------------------
Legend :
(*) real-time dynamic value
(w) Wire rates
B   Bytes
-------------------------------------------------------------------------------
Root (Egr)
| slot(2)
|--(S) : virtual_scheduler (Port 2/1/lo-gre)
|   |    AdminPIR:4000       AdminCIR:0(sum)
|   |
|   |    AvgFrmOv:105.31(*)
|   |    AdminPIR:4212(w)    AdminCIR:0(w)
|   |
|   |    [Within CIR Level 0 Weight 0]
|   |    Assigned:0(w)       Offered:0(w)
|   |    Consumed:0(w)
|   |
|   |    [Above CIR Level 1 Weight 1]
|   |    Assigned:4212(w)    Offered:0(w)
|   |    Consumed:0(w)
|   |
|   |
|   |    TotalConsumed:0(w)
|   |    OperPIR:3999
|   |
|   |    [As Parent]
|   |    Rate:3999
|   |    ConsumedByChildren:0
|   |
|   |
```

```
|   |--(Q) : -2147483646->2/1/lo-gre:1->EG(_tmnx_SHAPER_GR000):1->1
|   |   |    AdminPIR:10000000   AdminCIR:0
|   |   |    AvgFrmOv:105.31(*)
|   |   |    CBS:0 B             MBS:12582912 B
|   |   |    Depth:0 B           HiPrio:1376256 B
|   |   |
|   |   |    [Within CIR Level 0 Weight 1]
|   |   |    Assigned:0          Offered:0
|   |   |    Consumed:0
|   |   |
|   |   |    [Above CIR Level 1 Weight 1]
|   |   |    Assigned:3999       Offered:0
|   |   |    Consumed:0
|   |   |
|   |   |    TotalConsumed:0
|   |   |    OperPIR:4000        OperCIR:0
|   |   |
|   |   |    PktByteOffset:add 0*
|   |   |    OnTheWireRates:false
|   |   |    ATMOnTheWireRates:false
|   |   |    LastMileOnTheWireRates:false
```

12.11 Call trace

Call trace is supported for use on the ISA and the CPM. WLAN-GW-specific commands apply only to ISA tracing. To enable tracing on the CPM, IPoE session commands must be used without a SAP, circuit ID, or remote ID.

See Call Trace in the *Triple Play Enhanced Subscriber Management* chapter for more information.

12.12 Distributed RADIUS proxy

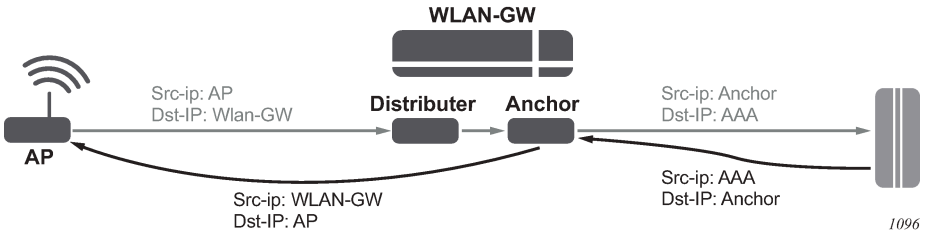
The distributed RADIUS proxy acts just like the regular RADIUS proxy but runs on an ISA and is designed for high scale and high performance. It can handle a high number of RADIUS transactions; therefore it is able to keep up with EAP authentications that consists of many RADIUS transactions (EAP-PEAP) and all the accounting messages sent by an Access Point for a specific UE. The distributed RADIUS proxy is designed to handle the scale and performance of Distributed Subscriber Management (DSM) but can also be used as a performance improvement for Enhanced Subscriber Management (ESM). All common server-selection mechanisms are supported (direct, round-robin, hash-based) and both IPv4 and IPv6 RADIUS clients can communicate with the proxy. Important difference with the CPM based proxy is no IPv6 support toward the RADIUS server.

The distributed proxy also supports caching an access-accept to aid authentication of Layer 3 setup (DHCP/SLAAC/DHCPv6). After UE creation the cache supports tracking of both accounting and authentication messages. Contrary to the CPM-based RADIUS proxy the key used in the cache is always the calling-station-id attribute and it is expected to contain the UE MAC address, as specified in RFC 3580, *IEEE 802.1X Remote Authentication Dial In User Service (RADIUS) Usage Guidelines*. Accounting-on and accounting-off messages are not supported. The RADIUS proxy cache works with both ESM and DSM UEs.

For caching to work, the distributed proxy makes sure that all packets are routed via the anchor ISA tied to the UE. An AP sends a RADIUS packet to the RADIUS proxy IP address shared by all ISAs, the WLAN-GW forwards the packet to a distributor ISA based on the source IP address of the radius packet. That

ISA then looks for the calling-station-id and forwards the packet to the correct anchor-isa to handle proxy functionality and caching. If no calling-station-id is found (such as acct-on/acct-off), the packet is always forwarded to a fixed ISA that is chosen at startup. The chosen ISA forwards the packet with a per-ISA IP as source-ip, this source-ip is assigned at startup from the range configured under **config>aaa>isa-radius-policy** *policy-name*. From server to client the packet is sent back to that IP address and therefore immediately arrives at the correct anchor ISA, which subsequently forwards the packet straight to the AP without an additional ISA pass through.

Figure 174: Distributed RADIUS packet forwarding



The following is a distributed proxy configuration example.

```
#--------------------------------------------------
/configure service vprn 50 radius-proxy
----------------------------------------------
    server "distributed_radius_proxy" purpose accounting authentication wlan-gw-group 1 create
        cache
            key packet-type request attribute-type 31
            timeout min 5
            no track-accounting
            track-authentication accept
            track-delete-hold-time 0
            no shutdown
        exit
        default-accounting-server-policy "wlangw_isa_radius"
        default-authentication-server-policy "wlangw_isa_radius"
        no description
        no load-balance-key
        no python-policy
        secret "BLoAGDmsLt/Rs9LLU5/lESjjqZa/ssWnEIMJNvgBwmo" hash2
        send-accounting-response
        wlan-gw
            address 10.1.10.1
            ipv6-address 2001:db8::0
        exit
        no shutdown
    exit
----------------------------------------------
/configure aaa isa-radius-policy "wlangw_isa_radius"
----------------------------------------------
            password "rNPEv/V0j095N0Qy4rnekTVbF89OIlVj" hash2
            servers
                router "Base"
                source-address-range 10.100.100.4
                server 1 create
                    authentication
                    ip-address 10.100.100.2
                    secret "rNPEv/V0j095N0Qy4rnekPU0fmH2TwEl" hash2
                    no shutdown
                exit
            exit
```

```
-------------------------------------------
```

12.12.1 ESM

For ESM support **authenticate-on-dhcp** must be enabled under **configure service ies** | **vprn** *service-id* **subscriber-interface** *ip-int-name* **group-interface** *ip-int-name* **wlan-gw vlan-tag-ranges range start start end** *end*. When receiving DHCPv4 the ISA sends the DHCP message and the cached access-accept to the CPM which further processes the setup sequence. On the CPM a regular RADIUS authentication policy should be picked up for the UE either through configuration on the group interface or via the LUDB. Typically this policy reflects the ISA policy. This policy is used as a context to store the access-accept on the CPM for 10s.

IPv6 hosts are supported but can only be authenticated after DHCPv4 has triggered the promote from ISA to CPM. When IPoE linking is enabled a SLAAC host is created together with the DHCPv4 host as usual. If an additional IPv6 host would arrive after the 10s timeout, a regular RADIUS authentication is started from the CPM using the previously mentioned RADIUS policy.

When tracking is enabled, the radius messages are handled on the ISA and specific tracking actions (mobility, delete) are sent directly to the CPM.

12.12.2 Distributed subscriber management

For DSM support the radius-proxy cache is directly tied to the UE record on the anchor ISA and is automatically used during UE creation. Tracking immediately executes the associated actions (mobility, timed host-delete) on the UE record. If a cached accept would time out before DHCP is received, a regular RADIUS authentication is used using the configuration under **configure service ies** | **vprn** *service-id* **subscriber-interface** *ip-int-name* **group-interface** *ip-int-name* **wlan-gw** *vlan-tag-ranges* **range start** *start* **end** *end* **authentication**.

12.12.3 VLAN awareness

The distributed RADIUS proxy optionally allows inclusion of the Alc-WLAN-SSID-VLAN attribute in an Access-Accept message. When this attribute is received, any subsequent traffic, including control plane traffic, must match the VLAN or VLAN-range, otherwise it is dropped. When an Access-Accept message with this attribute is received for a UE, one of the following scenarios may occur.

- The VLAN in authentication matches the current UE VLAN or VLAN-range. This is assumed to be a re-authentication within the same SSID, and no further action is taken.
- The VLAN in authentication does not match the current VLAN or VLAN-range. It is assumed that the UE moved between SSIDs and that this is a new initial authentication. The UE is returned to an authenticated-only state from which the UE can be recreated with the correct new SSID parameters. If the UE was in the ESM or DSM state, it is gracefully removed. Any allocated IP addresses are released, and final accounting data is sent.

The DRP checks on the exact VLAN or VLAN range based on configuration of inter-VLAN mobility. When inter-VLAN mobility is enabled, it is assumed that an SSID maps to a VLAN range instead of a single VLAN, and all comparisons are adapted accordingly. Additionally, when receiving an Access-Accept message for a UE with a different VLAN in the same VLAN range, the UE starts using the new VLAN to send downstream traffic.

This is recommended for deployment models where there are multiple SSIDs. Without specifying the VLAN, every authentication for a UE would be considered as a re-authentication for the same SSID and would make no changes to the WLAN-GW state. After authentication, the UE would send control-plane messages on the new SSID with a different VLAN or VLAN-range. The WLAN-GW would see this as a non-seamless change and trigger SHCV to remove the UE. Only after the UE is removed would control-plane traffic on the new VLAN or VLAN-range succeed, and it would also require an additional authentication trigger.

12.12.4 Operational commands

The following commands display all statistics related to the radius-proxy, both for communication toward the client and for communication toward the server.

show router *router-id* **radius-proxy-server** *server-name* **statistics**

clear router *router-id* **radius-proxy-server** server-name **statistics**

Example output:

```
*A:Dut-C# show router 50 radius-proxy-server "radius_proxy_isa" statistics
...
Group 1 member 3
-------------------------------------------------------------
Rx packet                                             : 2
Rx Access-Request                                     : 2
Rx Accounting-Request                                 : 0
Rx dropped                                            : 0
  Retransmit                                          : 0
  Wrong purpose                                       : 0
  No UE MAC to cache                                  : 0
  Client context limit reached                        : 0
  No ISA RADIUS policy configured                     : 0
  Invalid attribute encoding                          : 0
  Invalid password                                    : 0
  Accounting-Request with invalid Acct-Status-Type    : 0
  Accounting-Request with no Acct-Status-Type         : 0
  Invalid accounting Authenticator                    : 0
  Invalid Message-Authenticator                       : 0
  Management core overload                            : 0

Tx Access-Accept                                      : 1
Tx Access-Reject                                      : 0
Tx Access-Challenge                                   : 1
Tx Accounting-Response                                : 0
Tx dropped                                            : 0
  Server timeout                                      : 0
  Invalid response Authenticator                      : 0
  Invalid Message-Authenticator                       : 0
  Invalid attribute encoding                          : 0
  RADIUS server send failure                          : 0
...
```

The following RADIUS proxy messages sent to the server using this policy are also counted here.

show aaa isa-radius-policy *policy-name*

clear aaa isa-radius-policy *policy-name* **statistics**

Example output:

```
*A:Dut-C# show aaa isa-radius-policy "wifi_isa_radius"
...
Server 1, group 1, member 3
-------------------------------------------------------------------------------
Purposes Up                                  : accounting authentication
Source IP address                            : 10.100.100.6
Acct Tx Requests                             : 0
Acct Tx Retries                              : 0
Acct Tx Timeouts                             : 0
Acct Rx Replies                              : 0
Auth Tx Requests                             : 2
Auth Tx Retries                              : 0
Auth Tx Timeouts                             : 0
Auth Rx Replies                              : 2
CoA Rx Requests                              : 0
...
```

12.13 WLAN-GW 1:1 active-backup redundancy

This feature provides support for 1:1 inter WLAN-GW active-backup redundancy. The failure detection and switchover mechanism is contained in WLAN-GWs, and there is no dependency on the AP to detect failure of WLAN-GW and switch traffic to tunnel endpoint on a different WLAN-GW. There is also no dependency on NAT or a particular type of NAT on WLAN-GW. If local DHCP servers are used for address allocation, then DHCP leases in the server are synchronized to the backup WLAN-GW via MCS. However, ESM state for the UE is created on the backup WLAN-GW based on data-triggered authentication after switchover. The granularity of switchover is subscriber-interface. Both WLAN-GWs are required to be configured with the same tunnel endpoint address. Also, the subscriber interfaces on both WLAN-GW must be configured with the same subnets. Only the WLAN-GW that is deemed as active announces the tunnel endpoint address in routing toward the APs.

Active-backup decision is based on monitor and export route concept (same as what is used with NAT redundancy). Monitor and export routes are configured on the subscriber-interface on both WLAN-GWs. These should be complementary with respect to the ones on the other WLAN-GW. When WLAN-GW group goes up operationally, check is made in the FDB for presence of monitor route (which is the route exported by the other WLAN-GW). If it is not found, then the WLAN-GW assumes active state with respect to ownership of the tunnel end-point address, and the tunnel end-point address is announced in IGP toward the AP (subject to configured IGP and routing policy). The active WLAN-GW also announces the aggregate subscriber subnets upstream in routing. When WLAN-GW group comes up operationally, and detects the monitor route in the FDB, it assumes standby state with respect to the tunnel endpoint address. It does not announce the tunnel endpoint or the subscriber subnets in routing.

Each WLAN-GW needs to track the monitor route in the FDB. If the monitor route is no longer in the FDB, and the WLAN-GW is in standby state, it transitions to active, and announce the tunnel end-point toward APs, and subscriber subnets upstream. This draws the traffic from the AP to the backup WLAN-GW. Redundancy is non-revertive. The monitor and export routes are configured on the subscriber-interface.

```
config>service>ies>sub-if
wlan-gw
redundancy
        [no] export <ip-prefix/length>
        [no] monitor <ip-prefix/length>
   exit
exit
```

If the number of operationally up WLAN-GW IOMs in wlan-gw group drops below the number of active IOMs configured, the WLAN-GW group is brought down (based on the **oper-down-on-group-degrade** command under the WLAN-GW interface), and switchover procedures for the subscriber-interface are triggered (export route, tunnel endpoint address and subscriber subnets are withdrawn from routing).

```
config>service>vprn>sub-if>grp-if
config>service>ies>sub-if>grp-if
wlan-gw
[no] oper-down-on-group-degrade
```

The switchover can also be triggered administratively on per subscriber-interface basis using the **tools perform** command.

```
*A:vsim-07-cpm# tools perform wlan-gw redundancy force-switchover service <service-id>
 interface <ip-int-name>
```

12.13.1 DHCP server redundancy

1:1 redundancy provided with this feature only handles complete failure of WLAN-GW (either because of a chassis reboot or because of the number of operational WLAN-GW IOMs in WLAN-GW group falling below the number of active WLAN-GW IOMs, which operationally brings down the WLAN-GW group, and trigger switchover). For any partial failures (port, MDA or IOM failure), it is assumed there is network level redundancy, such that the soft-GRE tunnel is re-routed to the primary WLAN-GW. This ensures there is only one active WLAN-GW owning the subnets defined on the two WLAN-GWs (that is, allows local/local subnets). The DHCP servers state is synchronized between the two WLAN-GWs using MCS.

Supported access includes:

- DHCPv4 relay to external server
- DHCPv4 relay to local server
  - Pool name could be returned by AAA (framed-pool) in access-accept
  - Pool name could come from LUDB (as relay we would set use-pool-from-client). LUDB could be specified under group-interface or under DHCP server. LUDB or AAA returned pool allows support for per SSID pool selection. SSID is contained in circuit-id.
  - Local pool selection based on giaddr
- DHCPv4 proxy (IP@ from AAA or IP@ from PGW/GGSN)

The unnumbered case works in relay and proxy scenarios. IPv6 is not supported (data-triggered auth and subscriber creation for IPv6 is not supported). Therefore, DHCPv6 server synchronization is not applicable. Also, IPv4 address from LUDB is not supported (as data-triggered authentication against LUDB is not supported).

12.13.2 Subscriber creation after switchover

When standby WLAN-GW transitions to an active state and receives data on the anchor ISA there is not a UE state on the anchor ISA. Data-triggered authentication Data-triggered subscriber creation is used to

create the subscriber. To infer how the UE originally obtained the IP@ (DHCP relay versus proxy, such as AAA or GTP), the following holds:

1. If any GTP related parameters are returned in access-accept, then it is assumed the IP@ comes from GGSN/PGW, and the origin for the IP@ is assumed to be GTP.
2. If no GTP parameter is returned, and access-accept contains framed-IP, then proxy case is assumed (that is, the origin as AAA).
3. If no GTP parameter or framed-IP is returned, then DHCP relay is assumed. The remaining lease time is set to initial lease-time (if it was originally provided from AAA on primary WLAN-GW, it could be provided in access-accept for data-triggered auth on backup WLAN-GW). If AAA does not provide it, then it is initialized to default value of 7 days.

If authentication indicates GTP for the subscriber, then create-session-request is signaled with Handover indication. However, for dual-sack subscriber over soft-GRE, if AAA returns the SLAAC prefix in access-accept (in response to IPv4 data-triggered auth), and linking is configured, RA message is sent (unicast to client's MAC@), and a SLAAC host is created.

12.14 WLAN-GW triggered stateless redundancy (N:1)

Existing stateless redundancy, described in Data-triggered subscriber creation, is enhanced to support WLAN-GW based failure detection and switchover based on monitor and export route mechanism described above. The AP is not required to be configured with different tunnel endpoint addresses for active and standby WLAN-GWs. Single tunnel endpoint address is configured on the APs. The tunnel endpoint address is only announced in routing by the primary WLAN-GW as described in the section above. This form of redundancy as described in Data-triggered subscriber creation, required Layer 2–aware NAT. After failure, the subscriber on the standby WLAN-GW that transitions to primary is based on data-triggered authentication. This is supported for both ESM and DSM.

12.15 AP triggered stateless WLAN-GW redundancy (N:1)

Existing AP controlled redundancy, described in Data-triggered subscriber creation, is enhanced to trigger switchover on primary WLAN-GW if the number of WLAN-GW IOMs in the WLAN-GW group fall below number of active WLAN-GW IOMs. Based on a configuration using a **configure service ies** | **vprn** *service-id* **subscriber-interface** *ip-int-name* **group-interface** *ip-int-name* **wlan-gw** command, the WLAN-GW group is operationally brought down if a WLAN-GW IOM fails and the number of WLAN-GW IOMs fall below number of active WLAN-GW IOMs configured for the WLAN-GW group. This results in loss of route to the tunnel endpoint from the active WLAN-GW. The AP detects this as WLAN-GW failure, and start tunneling the data to a configured backup WLAN-GW, where the subscriber is created based on data-triggered authentication. This is supported for both ESM and DSM.

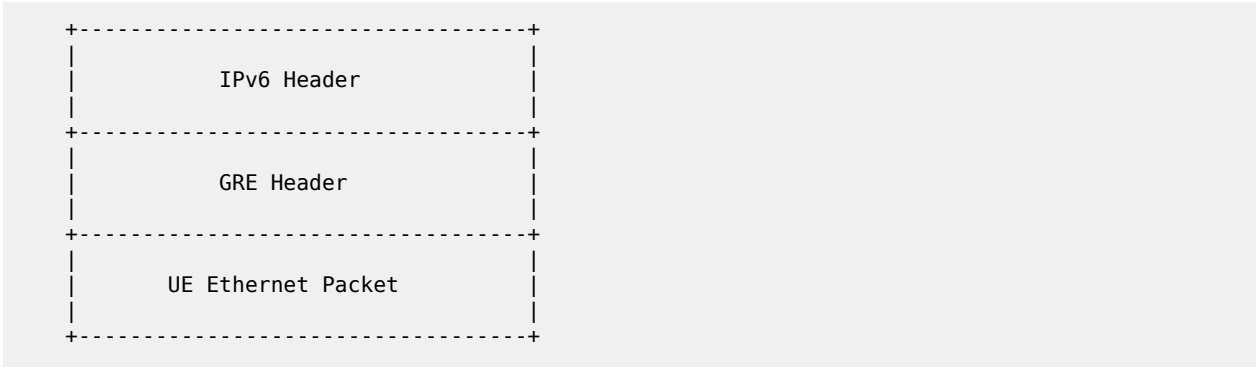
12.16 IPv6-only access

To accommodate IPv6 only AP/CPEs, IPv6 wlan-gw tunnel transport, and IPv6 client-side support for RADIUS-proxy have been added.

12.16.1 IPv6 GRE tunnels

Support for IPv6 GRE tunnels require configuration of local IPv6 tunnel end-point address under wlan-gw configuration on the group-interface. The transport for L2oGRE (or L2VPNoGRE) packet is IPv6 as shown below. The outer IPv6 header contains the value 0x2F (GRE) in its Next Header field. GRE header contains protocol Ethernet (0x6558) or Ethernet-over-MPLS (0x8847) as in the case IPv4 GRE.

IPv6 Transport for L2oGRE Packet:



A single wlan-gw endpoint instance on the group-interface can have both IPv4 and IPv6 address configured as shown below. Inter-AP mobility between IPv4 and IPv6 only APs is supported in this scenario.

IPv6 Endpoint Configuration for WLAN-GW:

```
service
    vprn 300 customer 1 create
        group-interface "grp-intf-1" wlangw create
            wlan-gw
                gw-addresses
                   address 10.1.1.4
                   address 2001:db8:
                exit
                gw-ipv6-address 2001:db8::0
                mobility
                   hold-time 0
                   trigger data iapp
                exit
                egress
                   shaping per-tunnel
                exit
                tcp-mss-adjust 1000
                vlan-tag-ranges
                    range start 100 end 100
                        data-triggered-ue-creation
                        retail-svc-id 402
                    exit
                 exit
                 router 30
                 wlan-gw-group 1
                 no shutdown
              exit
           exit
      exit
exit
```

The datapath for IPv6 GRE tunneled packets, including load-balancing of tunneled packets amongst set of ISAs in the WLAN-GW group, and anchoring after tunnel decapsulation remains unchanged. Per tunnel traffic shaping is supported like IPv4 tunnels. All existing per tunnel configuration on the group-interface described in previous sections (including mobility, egress shaping, VLAN ranges, and so on) is supported identically for IPv6 tunnels. Tunnel reassembly for upstream tunneled traffic is not supported for IPv6 tunnels. TCP mss-adjust is supported for IPv6 tunnels and is configurable under wlan-gw mode on group-interface. APs must use globally routable addresses for GRE IPv6 transport. Packets with extension headers are dropped.

12.16.2 IPv6 client-side RADIUS proxy

RADIUS proxy is extended to listen for incoming IPv6 RADIUS messages from IPv6 RADIUS clients on AP/CPEs. The listening interface that the RADIUS proxy binds to must be configured with an IPv6 address as shown below. The IPv6 RADIUS proxy is solely for DHCPv4-based UEs behind IPv6 only AP/CPEs (IPv6-capable UEs are not supported). All RADIUS-proxy functions (including caching, correlation with DHCPv4, and mobility tracking) are supported identically to existing IPv4 client-side RADIUS-proxy.

Configuration for IPv6 Client-Side RADIUS proxy:

```
service
    vprn 300 customer 1 create
        shutdown
        interface "listening_radius_server" create
             address 10.9.9.9/32
             ipv6
                 address 2001:db8::0
             exit
             loopback
         exit
     radius-proxy
         server "radius-proxy" purpose accounting authentication create
             shutdown
             cache
                 key packet-type request attribute-type 31
                 track-accounting stop interim-update accounting-on accounting-off
                 no shutdown
             exit
             default-accounting-server-policy "radius_server_policy"
             default-authentication-server-policy "radius_server_policy"
             interface "listening_radius_server"
             load-balance-key attribute-type 102 vendor 5
             secret "AQepKzndDzjRI5g38L3LbbN3E8qualtn" hash2
             send-accounting-response
             no shutdown
         exit
     exit
```

12.16.3 Dual-stack UEs over WLAN-GW

This feature adds support for dual-stack UEs over wlan-gw. Each dual-stack UE appears to WLAN-GW as a bridged client. Dual-stack UE support includes both SLAAC and DHCPv6, with and without linking with DHCPv4. Handling of DHCPv6, RS/RA, and NS/NA messages over wlan-gw has been added. WLAN-GW can assign /128 GUA to the UE via DHCPv6 or assign a /64 prefix in SLAAC to each UE. Each UE can be handed via DHCPv6, a /128 IA_NA from a unique /64 prefix, with the "on-link" flag is off in the RA message. This is because the public Wi-Fi users are distinct subscribers, and the communication must

always be via WLAN-GW. The CPE must prevent local-switching on the Wi-Fi link even if the /64 prefix is signaled as on-link or if the UEs are handed out /128 from the same /64 prefix.

Existing ESMv6 support on normal group-interface is applicable to wlan-gw group-interface and is already documented in the ESMv6 sections in this guide. There are a few exceptions that are mentioned in following sections.

12.16.3.1 SLAAC prefix assignment

SLAAC prefix assignment to the UE can be from local prefix pool, where pool name can come from RADIUS in Alc-SLAAC-IPv6-Pool VSA or from LUDB (see general section on ESMv6 SLAAC pool assignment). Alternatively, the SLAAC prefix can be provided from RADIUS (in standard Framed-IPv6-Prefix attribute) or from LUDB. SLAAC with stateless DHCPv6 (DHCPv6 information-request) is supported. DNS can be sent in RA messages (per RFC 6106). RS authentication (based on MAC address) can be configured (as described in general ESMv6 section on SLAAC only ESM hosts). SLAAC host is created on successful RS authentication. For successfully authenticated SLAAC host, an RA is sent in response to every received RS message (subject to a configured min-auth-interval). RA messages are sent to unicast MAC address of the UE.

SLAAC host creation can be linked to DHCPv4 by configuring **ipoe-linking** under **group-interface**. With **ipoe-linking** enabled, any received RS messages are dropped till DHCPv4 successfully authenticates and ESMv4 host is created. If **gratuitous-rtr-adv** is configured under ipoe-linking context then an RA is sent when ESMv4 host is created. If available, the SLAAC-prefix is included in the RA message. **shared-circuit-id** command under wlan-gw is not supported on wlan-gw interfaces. The O-Bit (other-stateful-configuration) is configurable on the group-interface.

12.16.3.2 DHCPv6 IA_NA assignment

If UE requests DHCPv6 IA_NA, a /128 address can be provided from a unique /64 prefix per UE from a local-pool. The pool name can be provided from LUDB or from RADIUS (in Framed-IPv6-Pool attribute). The address could also be provided via LUDB or RADIUS (in Alc-IPv6-Address VSA). DHCPv6 can also be linked with DHCPv4 by enabling **ipoe-linking** command. The M-bit in RA message is configurable. DHCPv6 IA_NA is allowed if it is received after a SLAAC host exists, if **allow-multiple-wan-addresses** is enabled under group-interface ipv6 configuration. In previous releases, this is precluded. This however consumes two hosts (one each for IA_NA and SLAAC) per UE. Based on a configuration command **override-slaac**, SLAAC host can be deleted if DHCPv6 IA_NA host is successfully created. Prefix-delegation is not supported with DHCPv6 on WLAN-GW interfaces.

12.16.3.3 Migrant user support

Migrant user support is only applicable to IPv4. However, if linking is configured for SLAAC or DHCPv6 with DHCPv4 then RS and DHCPv6 messages are dropped till IPv4 ESM host exists (that is, the UE is out of migrant state). After the IPv6 ESM host exists, that is, UE is out of migrant state, RA is sent to the UE (unicast MAC), and subsequent RS or DHCPv6 messages can result in creation of IPv6 ESM host. Therefore, with migrant UEs, linking should be enabled. SLA-profile instance accounting (with interim-updates), and per-host accounting (w/ interim-updates) are supported.

12.16.3.4 Accounting

Per SLA-profile instance accounting (with interim-updates) and per SLA-profile instance accounting (with interim-updates) with host accounting enabled is supported. The interim-updates are scheduled updates and carry IPv4 address and IPv6 address or prefix are assigned to the UE.

A sample sequence with per SLA-profile instance accounting (with interim-updates) is shown below:

0. IPv4oE host created based on DHCPv4.

1. Acct-start generated (contains framed-ip-address).

2. SLAAC host comes up.

3. Next scheduled interim-update (contains framed-ip-address and framed-IPv6-Prefix, that is, SLAAC-prefix).

4. DHCPv6 IA_NA gets assigned and corresponding host is created.

5. Next Scheduled interim-update (contains framed-ip-address, framed-IPv6-Prefix and Alc-Ipv6-Address).

6. SLAAC host times out.

7. Next Scheduled interim-update (contains only Alc-IPv6-Address and does not contain framed-IPv6-Prefix).

8. DHCPv6 IA_NA lease times out.

9. Next Scheduled interim-update (contains only framed-ip-address).

A sample sequence with per SLA-profile instance accounting (with interim-updates) with host accounting enabled is shown below:

0. IPv4oE host created based on DHCPv4.

1. Acct-start for sla-profile instance generated (contains framed-ip-address).

2. Acct-start for DHCPv4 host is generated (contains framed-ip-address).

3. SLAAC host comes up.

4. Acct-start for SLAAC host is generated (this should contain framed-IPv6-Prefix, that is, SLAAC-prefix)

5. Next scheduled interim-update for sla-profile instance accounting (contains framed-ip-address and framed-IPv6-Prefix, that is, SLAAC-prefix).

6. DHCPv6 IA_NA gets assigned and corresponding host is created.

7. Acct-start for DHCPv6 IA_NA host is generated (contains Alc-Ipv6-Address).

8. Next Scheduled interim-update (contains framed-ip-address, framed-IPv6-Prefix and Alc-Ipv6-Address).

9. SLAAC host times out.

10. Acct-stop (SLAAC-host-acct-session-id) is generated.

11. Next Scheduled interim-update for sla-profile instance accounting (contains only Alc-IPv6-Address).

12. DHCPv6 IA_NA lease times out.

13. Acct-stop (DHCPv6-IA_NA-host-acct-session-id) is generated.

14. Next Scheduled interim-update for sla-profile instance accounting (contains framed-ip-address).

15. DHCPv4 lease times out.

16. Acct-stop (DHCPv4-host-acct-session-id) is generated.

17. Acct-stop for sla-profile instance accounting is generated.

12.17 Layer 2 wholesale

This feature adds support for mapping a UE to a VPLS instance based on configuration. The mapping is explicitly created by assigning a VPLS instance to an SSID that the UE is connected to. The SSID is represented by the dot1q tag in the received Layer 2 frames from the UE. A VPLS instance is configured per vlan-range on wlan-gw group-interface (as shown in Figure 162: Standalone WLAN-GW). This feature therefore enables Layer 2 wholesale, where traffic from all UEs on an SSID is transparently forwarded into the corresponding VPLS instance associated with the retail ISP. UE authentication, address assignment, Layer 3 classification and QOS are managed by the retail provider terminating the subscriber. There is no local-switching on the WLAN-GW providing the wholesale service. When a VPLS instance is configured under a VLAN-range, an internal SAP is implicitly created in the VPLS instance between each ISA and corresponding carrier IOM in the WLAN-GW group. The internal SAP is associated with an implicitly created SHG to constrain broadcast and multicast traffic received from UEs, such that it is not forwarded back on the SAP. Layer 2 wholesale and Layer 3 termination are possible simultaneously on same wlan-gw interface, because Layer 2 wholesale or Layer 3 termination is a per SSID decision. UE state on the ISA is removed when the UE MAC in the VPLS instance ages out based on **local-age** configured under VPLS service.

A **vpls-sap-template** command (described in the *7450 ESS, 7750 SR, 7950 XRS, and VSR Layer 2 Services and EVPN Guide*) can be defined under **config**>**service**>**template** and associated with the VPLS service for Layer 2 wholesale via **config**>**service**>**vpls**>**wlan-gw**>**sap-template** command. Ingress and egress filter and QoS specified in the **vpls-sap-template** for the VPLS service is applied to the implicitly created internal SAP (between ISA and carrier IOM) in the VPLS service.

```
*A:vsim>config>service>vprn# info
----------------------------------------------
 subscriber-interface "s1" create
       group-interface "g1" wlangw create
              wlan-gw
                    vlan-tag-ranges
                           range start 100 end 100
                                l2-service 600
                                      no shutdown
                                exit
                           exit
                     exit
               exit
----------------------------------------------

*A:vsim>config>service>vpls# info
----------------------------------------------
 wlan-gw
     shutdown
     sap-template "foo"
 exit
----------------------------------------------
```

12.18 VLAN to WLAN-GW IOM/IMM steering via internal Epipe

This feature provides the steering of traffic received on an access VLAN or spoke SDP from a Wi-Fi AP/AC to a WLAN-GW IOM/IMM via an internal Epipe. The benefit of this internal steering is that all existing features available with native soft GRE tunnels on WLAN-GW IOM/IMM are now available to pure Layer

2 access via VLANs or spoke SDPs. The access SAP can be null, .q, or q-in-q. Access SAPs aggregating Wi-Fi APs or ACs can and be configured in the **config**>**service**>**ies**>**sub-if**>**grp-if**>**wlan-gw**>**l2-access-points**>**l2-ap** or **config**>**service**>**vprn**>**sub-if**>**grp-if**>**wlan-gw**>**l2-access-points**>**l2-ap** context.

The aggregation network can insert up to two AP identifying VLAN tags, and the AP can insert a .1q tag (typically for identifying the SSID). The number of AP identifying tags sent on the internal Epipe depends on the encapsulation on the access SAP. For example, if an aggregation network inserts two AP identifying tags, and an access SAP is configured with null encapsulations, then the traffic sent on the internal Epipe carries two AP identifying tags. The number of AP identifying tags in the frame forwarded over the internal Epipe must be configured via the **l2-ap-encap-type** command.

```
configure service (vprn|ies) <svc-id> subscriber-interface <ip-int-name> group-interface <ip-
int-name> wlan-gw
    l2-access-points
        [no] l2-ap <sap-id> [create]
            [no] encap-type {default|null|dot1q|qinq}
            [no] epipe-sap-template <name>
            [no] shutdown
        exit
    exit
    [no] l2-ap-encap-type {null|dot1q|qinq}
exit
```

The traffic on an internal Epipe is load-balanced among ISAs in the WLAN-GW group. The load balancing uses a hash based on AP identifying tags that remain on the frame after being received on the access SAP (based on the SAP encapsulation). This ensures all traffic from a particular AP is Epiped to the same ISA. Ingress and egress QoS and filters can be defined in an **epipe-sap-template** as displayed in the configuration shown below and associated with the access sap or spoke SDP. IP filters and DSCP remarking are not supported if more than two tags are present in the frame ingressing the SAP. Also, downstream filters and DSCP remarking is not applied if a retail tag is present. Both Layer 3 ESM and DSM as well as Layer 2 wholesale are supported for steered traffic.

```
configure service template epipe-sap-template <name> [create]
    egress
        [no] filter
            [no] ip <filter-id>
            [no] ipv6 <filter-id>
            [no] mac <filter-id>
        exit
        [no] qos <policy-id>
    exit
    ingress
        [no] filter
            [no] ip <filter-id>
            [no] ipv6 <filter-id>
            [no] mac <filter-id>
        exit
        [no] qos <policy-id> {shared-queuing|multipoint-shared}
    exit
exit
```

Mobility from an AP that is reached over a VLAN or spoke SDP to an AP that is reached over a soft GRE or soft L2TPv3 tunnels are not supported. Each internal Epipe takes away two SAPs on each WLAN-GW IOM (one per ISA) in WLAN-GW group. With 64K SAPs per IOM, the maximum number of internal Epipes supported per chassis is 32K.

12.19 Soft-L2TPv3 tunnels

This feature adds support for Layer 2 over soft-L2TPv3 tunnels. L2TPv3 is over UDP and both IPv4 and IPv6 transport is supported. The encapsulation with UDP allows NAT traversal. Soft-L2TPv3 tunnels are terminated on WLAN-GW IOM/IMM. All features supported with soft-GRE tunnels are supported identically with soft-L2TPv3 tunnels. L2TPv3 tunnels are stateless and there is no support for control channel, dynamic exchange of session-id and cookie, and L2-specific sublayer for sequencing. Received cookie in L2TPv3 is reflected. The AP can encode it's MAC address in 8-byte cookie. Based on configuration, the cookie can be ignored and just reflected, or parsed to interpret AP-MAC from the least significant 6 bytes. Both L2TPv3 over IP and L2TPv3 over UDP encapsulation is supported. L2TPv3 tunnels are load-balanced from ingress IOMs to WLAN-GW IOMs based on source IP address. Figure 175: L2TPv3 over UDP (IPv6 transport) and Figure 176: L2TPv3 over IP (IPv6 transport) shows these encapsulations with IPv6.

Figure 175: L2TPv3 over UDP (IPv6 transport)

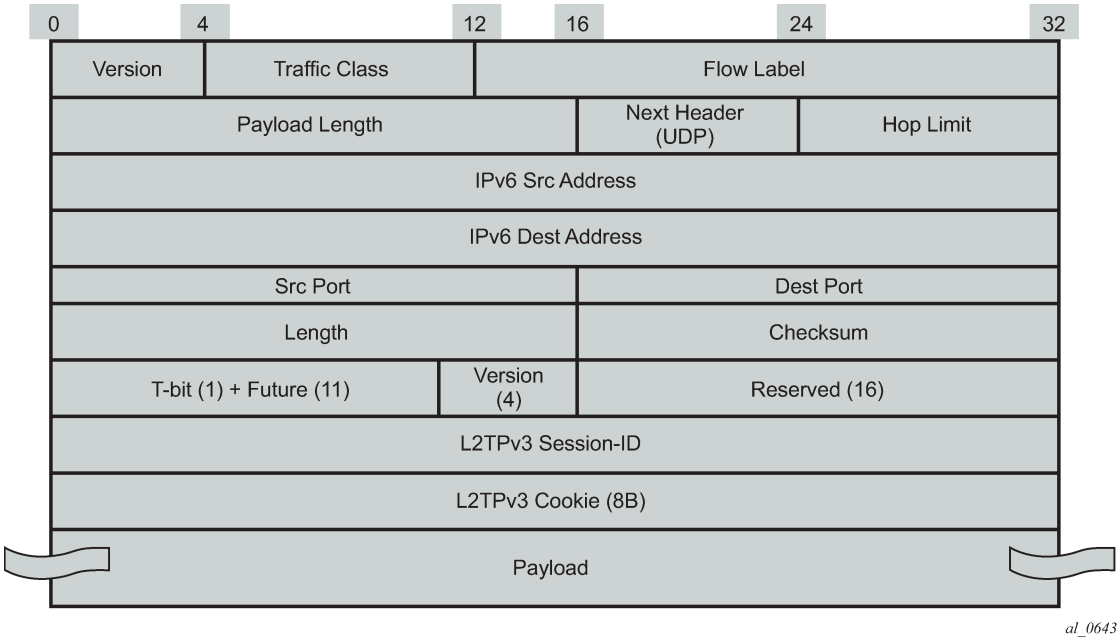
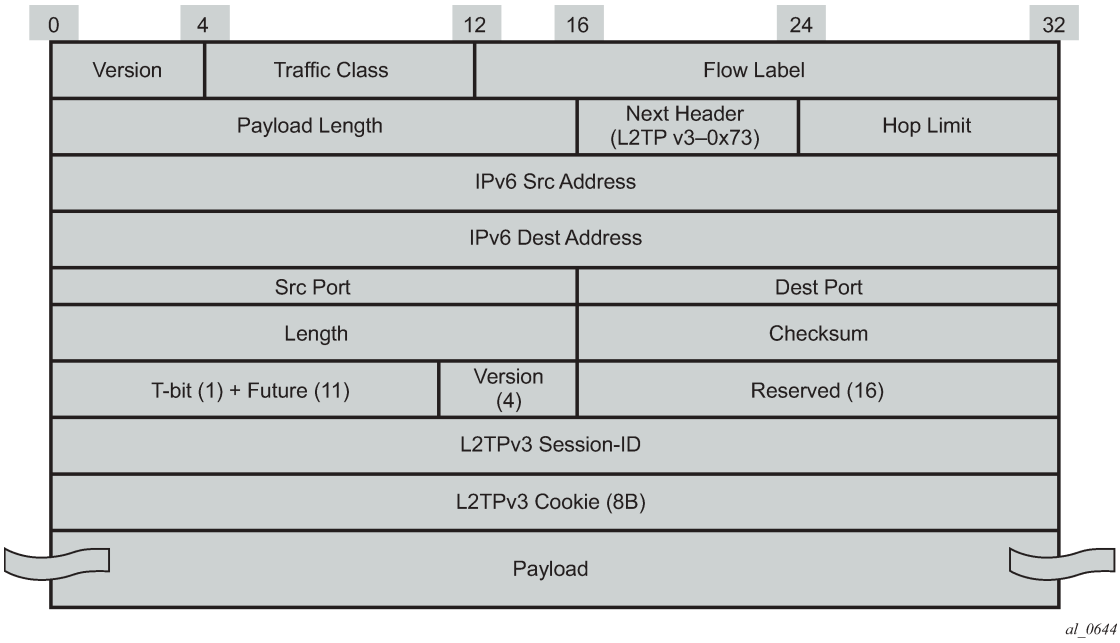


Figure 176: L2TPv3 over IP (IPv6 transport)



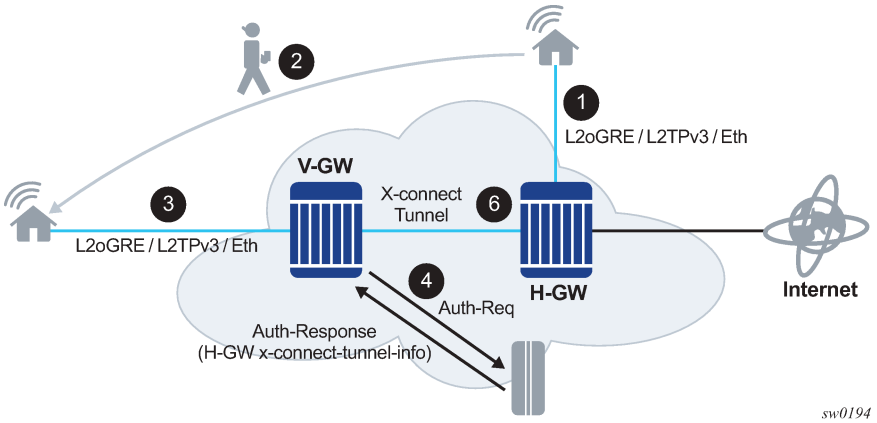
Enabling **multi-tunnel-type** on a **wlan-gw** *group-interface* allows multiple tunnel types (such as soft-GRE and L2TPv3) to the same gateway tunnel endpoint. Mobility between APs reachable with soft-L2TPv3 tunnels and APs reachable by soft-GRE tunnels is supported. There is feature and scale parity between soft-GRE and soft-L2TPv3 tunnels. The local tunnel gateway endpoint and other configurations parameters are shown below.

```
A:Dut-C>config>service>vprn>sub-if>grp-if>wlan-gw# info
----------------------------------------------
                        gw-addresses
                            address 10.1.1.3
                            address 2001:db8::
                        exit
                        gw-ipv6-address 2001:db8::0
                        learn-ap-mac
                        mobility
                            hold-time 0
                            multi-tunnel-type
                            trigger data iapp
                        exit
                        tunnel-encaps
                            learn-l2tp-cookie always
                        exit
                        wlan-gw-group 1
                        no shutdown
----------------------------------------------
```

12.20 WLAN-GW — Dynamic tunnel x-connect for seamless inter-WLAN-GW mobility

This feature adds support for seamless inter WLAN-GW mobility when a UE roams from an AP behind a WLAN-GW to an AP behind another WLAN-GW. Before this feature, inter-WLAN-GW mobility was supported by creating a UE state on the target WLAN-GW based on a data-trigger and carries over its Layer 2–aware NAT inside IP address to the target WLAN-GW. However, the outside NAT pool (and therefore the outside IP) changes, and any traffic for existing sessions is dropped. This feature introduces the concept of "home WLAN-GW" (H-GW) that provides an anchor point for a UE when it roams to an AP behind a different WLAN-GW, referred to as "visited WLAN-GW" (V-GW). With anchoring on the H-GW, the UE retails both its NAT inside and outside IP addresses. Existing NAT flows on the H-GW stay intact, and forwarding is seamless, as shown in Figure 177: Dynamic tunnel x-connect for seamless inter-WLAN-GW mobility. The V-GW tunnels the UE traffic to the H-GW that anchors the UE.

Figure 177: Dynamic tunnel x-connect for seamless inter-WLAN-GW mobility



12.20.1 Processing on the V-GW

A control or data packet is received on the V-GW when a UE moves to an AP behind the V-GW. The packet can be received on any supported access type (L2oGRE, L2TPv3 tunnel, or internal VLAN anchor). The packet, after tunnel decapsulation, is forwarded to the anchor ISA on the V-GW. Authentication is triggered from the anchor ISA based on the configured ISA authentication policy. The **authenticate-on-dhcp** command must be enabled on V-GW for DHCP-triggered UE state creation scenarios to work.

If the access-accept message contains any VSAs related to x-connect tunneling, then the UE is created in the x-connect mode, and any subsequent Layer 2 frames received in the V-GW from the UE are transparently tunneled to H-GW. By default, the x-connect tunnel used between the V-GW and the H-GW is of same type (LoGRE or L2TPv3) as an access tunnel (for example, the tunnel from the AP to the V-GW). If the access type is L2-AP (such as an internal VLAN anchor), then the x-connect tunnel used is L2oGRE. An x-connect tunnel type can be overridden from the AAA in an access-accept message by a VSA (Alc-Xconnect-Tunnel-Type). IPv6 transport is used for x-connect tunnels.

The tunnel destination for an IPv6 address on the H-GW is returned in a VSA in an access-accept message (Alc-Xconnect-Tunnel-Ipv6). The service in which the tunnel is routed is also returned in a VSA (Alc-Xconnect-Tunnel-Service). The anchor ISA on the V-GW allocates a unique IPv6 address per UE for the x-connect tunnel source. This address is allocated from a contiguous IPv6 address range assigned to

the ISA (based on the configured /64 prefix for the WLAN-GW group supporting x-connect tunneling). If the UE moves from one AP to another on the V-GW, then the anchor ISA allocates a new IPv6 address to use as tunnel sources for the x-connect tunnel. This change in x-connect tunnel source services as a UE mobility indication on the H-GW.

12.20.2 Processing on H-GW

Normal UE processing for WLAN-GW is performed on the H-GW. After tunnel decapsulation, the UE is received on the anchor ISA. If the UE state already exists on the anchor ISA, then it is treated as a mobility event, and the tunnel information in the UE (to the tunnel toward the V-GW). A mobility-triggered interim-update, if configured, is generated. The AP's MAC address must be reflected in **called-station-ID** attribute. The AP MAC is learned from the DHCP circuit-ID. However, if the mobility trigger is a data packet and the tunnel type is L2oGRE, then an ARPoGRE is generated (if configured) to learn the AP MAC. The V-GW forwards ARPoGRE is received on x-connect tunnel to the access tunnel toward the AP and reflects the received response back on x-connect tunnel. If the x-connect tunnel is type L2TPv3, then the AP MAC address can be carried in the cookie field. If the UE state does not exist, then normal UE processing leads to authentication and subsequent UE creation (as DSM or ESM UE).

IPv6 tunnel reassembly is not supported.

12.20.3 Idle timeout handling

An idle timeout (with optional SHCV) is independently enforced on the H-GW and V-GW. The idle timeout value can be provided by AAA. The UE state is deleted if the idle timeout triggers and there is no response to SHCV.

12.20.4 Distributed RADIUS proxy for closed SSID

An existing distributed RADIUS proxy (as described in Distributed RADIUS proxy) can be enabled on both the H-GW and V-GW. Authentication and accounting messages are proxied to configured AAA servers. When the UE state exists on the V-GW, any authentication and accounting messages are visible on the H-GW and are proxied directly to the configured AAA server from the V-GW. Data-triggered mobility must be enabled on the H-GW, as it cannot rely on authentication and accounting messages to infer UE mobility when the UE is on V-GW.

12.20.5 H-GW redundancy

Existing WLAN-GW controlled redundancy (active or backup decision based on monitor or export route tracking mechanism, as described in section WLAN-GW 1:1 active-backup redundancy) is supported. The backup H-GW advertises the tunnel endpoint IP in routing when it becomes the active to draw the traffic from the V-GW. The UE state is created on the new active based on data-triggered authentication.

12.21 ISA operational commands and key performance indicators

This section provides an overview of operational commands that can be used to monitor ISA behavior and performance. This focuses on WLAN-GW functionality. If NAT is also used on the same ISA, see the NAT guides for similar commands.

12.21.1 ISA resources

Resources on ISA BB are statically allocated at boot time and therefore strict have limits. These limits depend on both the provisioned ISA type used and on specific configuration flags. Use the command **tools dump wlan-gw isa resources mda** *mda-id* to generate a list of all resources, together with the currently in-use, free, maximum, and peak usage counters.

Most resources are directly related to configuration, and the corresponding configuration fails if no resources are available. However, specific resources, such as the number of UEs on an ISA, are allocated dynamically. To facilitate operational maintenance, configure watermarks on those resources using the command **configure isa wlan-gw-group** *wlan-gw-group-id* **watermarks**. When the corresponding high watermark is reached, a notification is generated, which can prompt an operator to investigate the high resource consumption and take appropriate steps, such as optimizing load or adding additional resources.

12.21.2 ISA load

Use the command **tools dump wlan-gw isa performance mda** *mda-id* **last** *time-span time-unit* to generate a high-level overview of ISA processing load and total data processed in a specified time span with a polling granularity depending on that time span. The following time spans and granularity are supported:

- last minute with second granularity
- last hour with minute granularity
- last day with hour granularity
- last month with day granularity

12.21.3 Query-based UE and tunnel states

The query-based state is a mechanism to fetch the states of UEs and tunnels in large-scale environments using ISA-only features such as migrant users, l2-wholesale, or DSM. In all these cases, UE and corresponding tunnel states are only created on the ISA and can scale into the millions.

To retrieve specific data without going through all UEs and tunnels, configure a query under the **config**>**subscriber-mgmt**>**wlan-gw**>**ue-query** command or **config**>**subscriber-mgmt**>**wlan-gw**>**tunnel-query** command. This query specifies match criteria for the state output. After a query has been created, use it to retrieve the matching results through any state interface. The **show**>**subscriber-mgmt**>**wlan-gw**>**ue**>**query-results** and **show**>**subscriber-mgmt**>**wlan-gw**> **tunnels**>**query-results** commands display the results in CLI.

By retrieving only the number of records that match the specified criteria, a set of very specific custom counters can be created; for example, "count all DSM UEs with IPv6 addresses" or "count all tunnels with more than three UEs". For UE queries, this count is always available. For tunnel queries this count must be explicitly enabled by using the command **calculate-counts**. Do not use tunnel counters when the expected

number of tunnels is greater than 1000, because retrieving an exact count for such data sets may take too much time to complete.

```
ue-query 1 name "by_mac" create
    mac-address 00:00:5e:00:53:11
exit
tunnel-query 1 name "min_3_dsm_ue" create
    min-num-ue 3
    ue-state
        dsm
    exit
exit
```

A UE can be assigned up to four custom user groups during RADIUS authentication using the Alc-Wlan-Custom-User-Group attribute. The UE and tunnel queries use this group to filter specific sets of UEs or tunnels with these UEs. The groups for a UE are updated using a RADIUS CoA message and the current applicable groups for a UE can be included in the RADIUS accounting messages.

12.21.4 Packet statistics

The ISA BB keeps extensive counters for each packet that is sent, terminated, or forwarded. These statistics are displayed using the command **show isa wlan-gw-group** *wlan-gw-group-id* **member** *member-id* **statistics**. Most statistics are straight-forward operational counters, such number of DHCP messages received or sent; however, there is also an extensive set of error counters available. When a WLAN-GW shows unexpected behavior, these counters can provide preliminary information about the issue. During debugging, it is recommended to first reset statistics to reset old and potentially unrelated error counters.

If a specified statistic displays an unexpected increase, it is not always easy to identify which UE or which behavior in a UE caused this increase. In this case, enable debugging of a specific statistic using the command **debug wlan-gw group** *wlan-gw-group-id* **statistic**. This command generates a hex dump of the first packet causing the statistic to increase in the debug logs. This packet can then be further investigated to determine the root cause and can be used to derive a UE MAC to start UE-based debugging.

12.22 Dynamic VPLS service

The dynamic VPLS service feature differs from the Layer 2 over soft-GRE tunnels feature described in Layer 2 over soft-GRE tunnels, in which VPLS service assignment is based on SSID information. With the dynamic VPLS service feature, the VPLS service assignment is based on the RADIUS access-accept attribute, Alc-L2-Service-Name. Use one of the following commands respectively to enable the dynamic VPLS service feature for a VPRN or IES service.

- **MD-CLI**

```
configure service vprn subscriber-interface group-interface wlan-gw vlan-range dynamic-service
configure service ies subscriber-interface group-interface wlan-gw vlan-range dynamic-service
```

- **classic CLI**

```
configure service vprn subscriber-interface group-interface wlan-gw vlan-tag-ranges range
 dynamic-service
```

```
configure service ies subscriber-interface group-interface wlan-gw vlan-tag-ranges range
 dynamic-service
```

The VPLS service, to which the WLAN-GW subscriber is assigned, must have a reference to the WLAN-GW group. Use the following command to configure an already-existing WLAN-GW group as the endpoint within the VPLS.

```
configure service vpls wlan-gw wlan-gw-group
```

The following packets from the UE can be used to trigger dynamic VPLS service creation:

- ARP
- DHCP
- DHCPv6
- RS (for SLAAC)
- OSPFv2 and OSPFv3 packets
- IP TCP or UDP data trigger packets

The WLAN-GW can reply to neighbor solicitation. However, you must use one the following commands respectively to enable neighbor solicitation for the VPRN or IES service.

```
configure service vprn subscriber-interface group-interface ipv6 auto-reply neighbor-
solicitation
configure service ies subscriber-interface group-interface ipv6 auto-reply neighbor-
solicitation
```

Enabling neighbor solicitation allows the UE to resolve IPv6 neighbor entries and forward traffic, which can be used as IP TCP or UDP data trigger packets.

A UE context is only created on the WLAN-GW ISA application on successful authentication. No ESM subscriber entry is created for the dynamic VPLS service. When the WLAN-GW endpoint is created on the VPLS, the UE context is only removed if the MAC address times out in the Layer 2 FDB.

Use the following command to remove a VPLS WLAN-GW endpoint.

```
tools perform wlan-gw clear-ue
```

Note: Clearing the VPLS FDB removes both the VPLS WLAN-GW endpoint and the UE context on the ISA application, as well as all other endpoints or SAPs and SDPs created by the WLAN-GW in the service.

RADIUS temporarily rejects the authentication of the UE to prevent the UE from being created on the VPLS service.

13 GTP

13.1 GTP uplink

SR OS supports subscriber traffic forwarding over an uplink GTP tunnel toward a GGSN or P-GW. This requires a per-subscriber GTP tunnel based on authentication to be configured. Each subscriber may access only a single APN. Both GTPv1 (Gn) and GTPv2 (S2a/S2b) are supported. A single primary PDP context per subscriber is supported on the Gn interface (3GPP TS 29.060 Release 8) from SR OS to the GGSN. A single default-bearer per subscriber is supported on the S2b interface (3GPPTS 29.274 Release 10), and S2a interface (3GPP TS 29.274 Release 11) from SR OS to the P-GW.

13.1.1 Identification attributes

GTP requires at least an IMSI and an APN to set up a connection. The IMSI is required to identify the user, and the APN is required to identify the network the user is connecting to.

There is a 1:1 relationship between IMSI and subscriber ID. It is possible to provision only one of these and the other will be accepted as the same value. If both are provisioned, they must be equal. Therefore, it is not possible to set up more than one GTP tunnel per subscriber.

APN can be provisioned explicitly per subscriber, or a default APN can be provisioned per VRF. If this APN does not contain an Operator Identifier (OI), it will be added automatically based on the IMSI.

13.1.2 P-GW/GGSN selection

The initial address of the P-GW or GGSN can either be provided during authentication, or, in the absence of authentication, resolved dynamically via DNS. For DNS, a FQDN is generated based on the APN as specified in 3GPP 29.303. This FQDN always consists of both the Network-ID (NI) portion and the OI portion, and is formatted as NI.apn.epc.mnc<MNC>.mcc<MCC>.3gppnetwork.org. The system performs an S-NAPTR lookup with this FQDN.

When multiple gateway addresses are returned as part of this lookup, load-balancing is performed according to regular NAPTR and SRV rules. If no addresses are returned, or the S-NAPTR lookup failed, the system tries a regular A host lookup with the same FQDN. In this case, SR OS load-balances over multiple gateway addresses using a round-robin mechanism. DNS servers can be configured per VPRN, except for the base router where the servers defined in the BOF are used.

```
config>router
config>service>vprn
    apn "internet.mno1.apn"
config>service>vprn>dns
    primary-dns 10.1.1.1
    secondary-dns 10.1.1.2
    tertiary-dns 10.1.1.3
    ipv4-source-address 10.1.1.1
exit
```

After initial GTP setup, it is possible for the P-GW or GGSN to return another address as a GTP-C or GTP-U destination. All data plane traffic is sent using the signaled GTP-U address. All subsequent control plane traffic is forwarded to the new GTP-C address.

The VRF used for GTP tunneling can be selected via provisioning a retain service ID for the subscriber. The source IP for GTP tunneling is taken from a loopback interface with the name **system** in that VRF. If no such interface is present, the tunnel setup fails.

13.1.3 Configuration

Profiles with signaling-related configuration per mobile gateway can be created locally on the SR OS router. These profiles include configuration for the interface type used between the router and the mobile gateway, path management parameters, retransmission parameters, and default values for GTP information elements such as AMBR. Each profile can be mapped to a specific gateway IP address or subnet per VRF. Most of the per-session/context parameters can be overridden via RADIUS authentication. See the *7450 ESS, 7750 SR, and VSR RADIUS Attributes Reference Guide* for more information.

```
config>subscr-mgmt>gtp>peer-profile
    no description
    interface-type s2a
    ip-ttl 255
    keep-alive interval 60 retry-count 4 timeout 5
    message-retransmit timeout 5 retry-count 3
    protocol-configuration-options apco
    no python-policy
    rat-type wlan
    no report-wlan-location
    session-hold-time 30
    charging-characteristics
        no home
        no roaming
    exit
    ggsn
        qos
            no ambr
            arp 1
            down-link gbr 2000 mbr 2000
            up-link gbr 5000 mbr 5000
        exit
    exit
    pgw
        qos
            ambr down-link 20000 up-link 10000
            arp 1
            down-link gbr 0 mbr 0
            qci 8
            up-link gbr 0 mbr 0
        exit
    exit

config>router
config>service>vprn
    gtp uplink peer-profile-map
        address 198.51.100.1/32 "pgw-policy"
        address 239.0.113.0/24 "ggsn-policy"
    exit
```

13.1.4 QoS support

SR OS provides appropriate traffic treatment and remarking based on DSCP bits in the outer and inner header in GTP packets.

Downstream from PGW/GGSN, the DSCP bits from the outer header in a GTP packet can be mapped to a forwarding class on network ingress, which can be preserved through the chassis as the packet passes to the egress IOM. On egress, reclassification can be done based on either the inner or outer DSCP bits, depending on the configuration value of the **use-ingress-l2tp-dscp** option in the SLA profile.

In the upstream direction, regular ESM FC classification is used. This FC is carried through the IOM to the egress complex. In the egress complex, this FC can be used for remarking of the outer DSCP values.

DSCP and default FC values for egress GTP-C packets can be configured under **sgt-qos**.

It is possible to signal the subscriber's aggregate rate or the rate of a specific scheduler in the downlink AMBR IE in both GTPv1-C and GTPv2-C. This uses the **report-rate** configuration of the SLA profile; the pppoe-actual-rate and rfc5515-actual-rate values are not applicable for GTP. This value can be subtracted with a value signaled during authentication to consider an average use for selective breakout.

```
config>subscr-mgmt>sla-prof>egress
    report-rate
```

Other signed QoS IE values are taken from static configuration or values signaled in authentication.

13.1.5 GTP session hold

Deletion of an IPoE or PPPoE session also triggers deletion of any corresponding GTP sessions. This deletion is subject to a configurable hold time. When the subscriber returns with the same GGSN/P-GW parameters within the hold time, the GTP connection is not re-signaled. This avoids releasing resources (such as IP addresses) too quickly on the GGSN/P-GW. This is useful in the following cases:

- **non-seamless control plane mobility with WLAN-GW**

  Because of SHCV usage, the UE is completely removed, but should reconnect quickly, preferably with the same IP address. Having a short hold time avoids the P-GW releasing the IP address.

- **IP handover from Wi-Fi-GW to RAN**

  In some cases, Wi-Fi connectivity is lost before wireless data connectivity is established. By holding the GTP session active on the WLAN-GW, the IP address is also held on the GGSN/P-GW until the RAN connection is complete.

While a GTP session is in hold, all downstream traffic is dropped, but no error indication messages are sent.

13.1.6 Selective breakout

This feature adds support for selecting a subset of traffic from a host (via IP filter) for local forwarding, while tunneling the remaining traffic to GGSN/PGW. This allows selected traffic to bypass the mobile packet core. The IP address for the host still comes from the GGSN/PGW during GTP session setup. Therefore, the selected traffic for local breakout from SR OS requires NAT functionality to draw the return traffic back to the router. To support address overlap within GTP, the NAT functionality is L2-aware. The selection of traffic for local breakout (local forwarding and NAT) is based on a net action in an upstream Ip filter applied to the host.

Selective breakout can be enabled on a per-host basis via RADIUS VSA (Alc-GTP-Local-Breakout) in access-accept. It is not possible to change this during a host's lifetime, such as via CoA. AA functionality is supported for local breakout traffic. Also, LI (after NAT) is supported for local breakout traffic, and is enabled via existing secure CLI, as stated in the *OAM and Diagnostics Guide*.

```
system>config>filter
    ip-filter 10 create
        entry 1 create
            match protocol udp
                dst-port eq 4000
            exit
            action gtp-local-breakout
        exit
```

On traffic ingress from the host UE, normal ESM host lookup and CAM lookup with the ingress host filter is performed. If there is a match in the filter indicating "gtp-local-breakout", the traffic is forwarded within the chassis to an ISA-BB, where is it subjected to Layer 2–aware NAT function, and afterwards is forwarded using regular routing in the NAT outside VRF. The inside IP address is the address returned in GTP, and may not match a NAT L2-aware inside prefix. The outside IP is an address belonging to the NAT outside IP address range on the ISA. If the filter action results in a "forward" action (default), the traffic is GTP-tunneled without performing NAT functionality. The traffic received from the network can be a normal L3 packet or a GTP encapsulated packet. The normal Layer 3 packet is expected to be destined for the NAT outside IP and is normally routed to the NAT ISA.

By default, per-host accounting includes counters that are aggregated across GTP and local breakout traffic. Separate counters can be obtained by directing the GTP and local breakout traffic into different queues associated with the corresponding ESM host based on QoS IP classification. NAT information (outside IP and port range) associated with an ESM host subjected to selective breakout is included in accounting-updates.

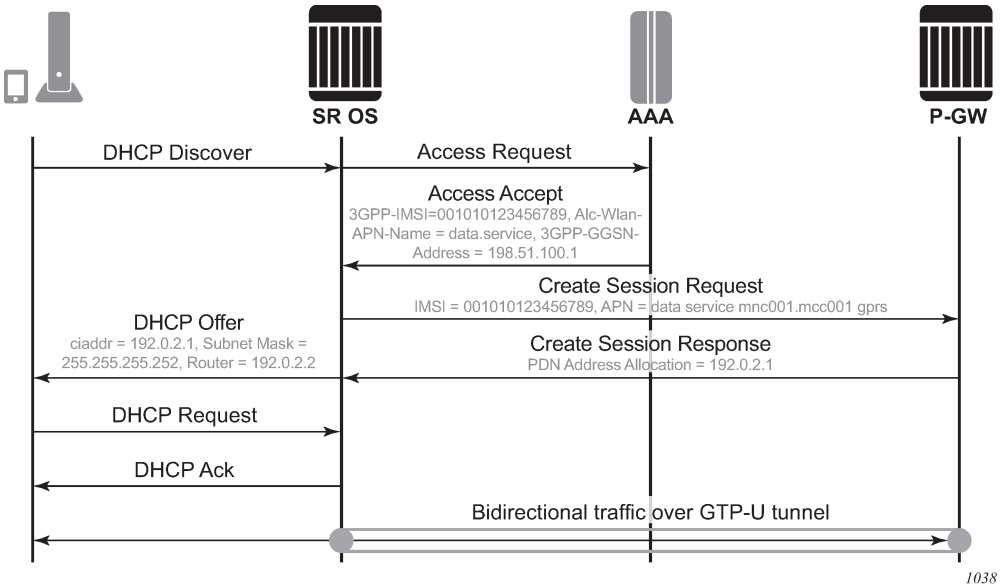
Selective breakout is supported for IPv4 only.

13.1.7 IPoE support

A GTPv2 session or GTPv1 PDP context is set up when IPoE session authentication includes any GTP parameters. The GTP session provides the IPv4 and IPv6 address used for the connecting host. Currently only DHCPv4 and SLAAC are supported to deliver these addresses back to the client. If DHCP is used, SR OS automatically derives a standards-compliant subnet mask and default gateway from the signaled IP address. It is important that all GTP subscribers are in a shared split-horizon domain and that there is no Layer 2 switching between GTP subscribers. Only a single IPoE session is supported per GTP subscriber. Additionally, DNS and NBNS can be signaled via GTP (A)PCO and reflected in DHCP, SLAAC, and stateless DHCPv6. Control plane packets such as DHCP and ICMPv6 RS are always terminated on the BNG and are not forwarded over GTP.

Figure 178: IPoE session shows a sample IPoE session for GTP.

Figure 178: IPoE session



GTP without an IPoE session is available for IPv4 DHCP leases only for backwards compatibility. It may not be used for new deployments; existing deployments should move to the IPoE session concept.

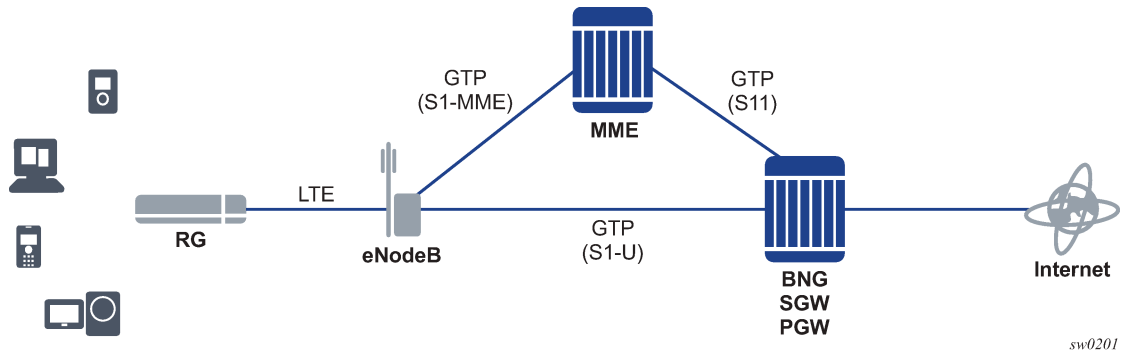
13.1.8 PPPoE support

A GTPv2 session or GTPv1 PDP context is set up when PPPoE session authentication includes any GTP parameters. The GTP session provides the IPv4 and IPv6 address to be used for the connecting host. IPCP and IPv6CP with SLAAC are supported to signal these addresses to the client. Only a single PPPoE session is supported per GTP subscriber. Additionally, DNS and NBNS can be signaled via GTP (A)PCO and reflected in IPCP, SLAAC, and stateless DHCPv6. Control plane packets such as ICMPv6 RS are always terminated on the BNG and are not forwarded over GTP.

13.2 GTP access

SR OS supports TPSDA functionality over GTP access funnels toward eNodeBs using the S11 (GTPv2-C) and S1-U (GTP-U) interfaces toward MME and eNodeB. In this setup, as shown in Figure 179: GTP access, the SR OS router performs the roles of BNG, SGW, and PGW. Both IPv4 and IPv6 connectivity over GTP is supported. GTP session authentication can be performed using LUDB, RADIUS, or NASREQ. GTP access termination is built on top of forwarding path extensions.

Figure 179: GTP access



13.2.1 GTP termination

GTP S11 termination can be enabled on interfaces in the base router and VPRNs. The configuration is linked to an APN policy that lists all supported APNs and the authentication mechanism (for example, LUDB or RADIUS) to be used per APN. The configured APN string should match the value signaled in GTP; however, fallback configuration is supported for any unknown APNs.

```
*A:FWGW>config>subscr-mgmt>gtp# info
----------------------------------------------
            apn-policy apn_demo create
                apn demo.mnc001.mcc001.gprs create
                    radius-auth-policy "gtp_auth"
                exit
            exit
----------------------------------------------
*A:FWGW>config>service>vprn# info
----------------------------------------------
            gtp
                s11
                    interface gtp_endpoint create
                        apn-policy "apn_demo"
                    exit
                exit
            exit
            interface "gtp_endpoint" create
                description "Tunnel endpoint IP"
                address 192.0.2.10/32
                loopback
            exit
----------------------------------------------
```

A GTP peer profile defines specific signaling parameters such as TTL values, keepalive timers, retransmit timers, and default values for information elements. By default, an automatically-created profile with the name "default_s11" is used. A more specific profile can be configured and mapped to a peer by a per-VRT mapping of IP address or prefix to that profile. To map all peers within the same VRF to the same profile, it is possible to use prefix 0.0.0.0/0.

```
*A:FWGW>config>subscr-mgmt>gtp# info
----------------------------------------------
            peer-profile "s11_peers" create
                interface-type s11
```

```
                keep-alive interval 180 retry-count 10 timeout 20
                message-retransmit timeout 3 retry-count 1
                mme
                    qos
                        ambr down-link 50000 up-link 10000
                    exit
                exit
            exit
-----------------------------------------------
*A:FWGW>config>service>vprn>gtp>s11# info
-----------------------------------------------
                    peer-profile-map
                        address 0.0.0.0/0 peer-profile s11_peers
                    exit
-----------------------------------------------
```

When an S11 session is set up, the accompanying S1-U bearer is terminated in the same VRF, but it is directly linked to a group interface in either the same or a different VRF. A default group interface can be configured per APN, which can be overridden during S11 session authentication. The group interfaces are of type **gtp** and require an FPE construct to operate. The traffic takes two passes through the forwarding plane. For upstream data, in the first pass, the GTP-U header is stripped; in the second pass, it is inserted into the group interface for regular ESM processing, based on existing IPoE functionality. For downstream data, in the first pass, regular ESM processing is performed and traffic egresses over the group interface; in the second pass, GTP-U encapsulation occurs.

Active GTP sessions support the S1 Release, UE triggered service request, and network-triggered service request procedures as defined in TS 23.401 to support connection idling and paging.

Both IPv4 and IPv6 are supported for GTP termination. GTP is enabled for the primary IPv4 and IPv6 addresses configured on the S11 interface. If both IPv4 and IPv6 address are configured, GTP supports dual-stack operations as follows:

- For GTP-C, a stack is chosen based on the stack of the incoming GTP-C message. Any subsequent GTP-C transactions initiated by the FWA gateway uses this stack. Subsequent GTP-C transactions initiated by the MME may change the IP stack.
- For GTP-U, a downstream peer is selected based on the information in the S1 eNodeB F-TEID IE. If that peer also contains two addresses, IPv6 is preferred. Upstream traffic can be received on both the IPv4 and IPv6 address simultaneously.

13.2.1.1 Multiple APNs

SR OS supports multiple APN connectivity for the same IMSI. When an additional session setup with a different APN is triggered, the system treats this as a completely new session that goes through APN selection and authentication again. PDN session user plane can be instantiated in the same VRF as previous sessions or in different VRFs. Sessions for the same IMSI can belong to the same ESM subscriber by using the same subscriber-id, but are not required to do so. S11 messages defined with UE granularity in 3GPP TS 29.274 apply to all sessions of the same IMSI.

13.2.2 GTP session setup

When a GTPv2 Create Session Request message arrives, SR OS looks up the required authentication mechanism in the applicable APN policy. To aid in the identification process, you can configure both RADIUS and NASREQ with GTP-specific include attributes such as IMSI, MSISDN, and IMEI. If a PAP message is present in the PCO IE of the Create Session request, the system uses that username and

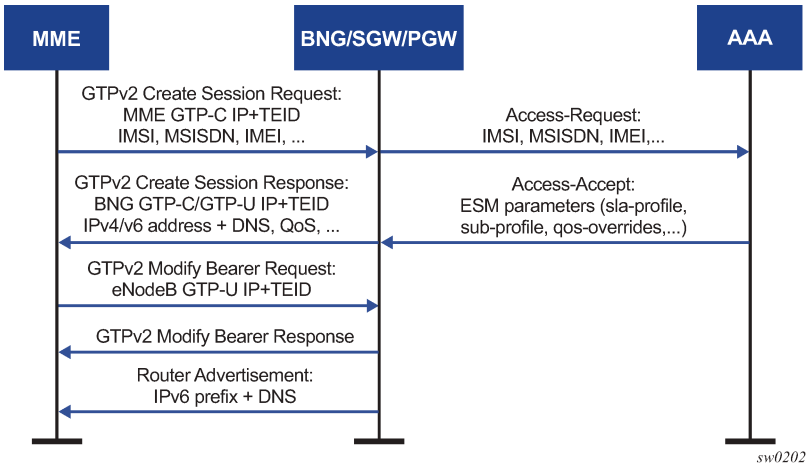
password for authentication; if not, it falls back to the username and password configured in the **configure subscriber-mgmt authentication-policy** context. For LUDB-based authentication, it is recommended to use **derived-id** for identification values.

Authentication is performed per GTP session and not per IP stack (host). Therefore, the initial authentication returns parameters for all stacks that need to be set up.

After a successful authentication, a Create Session Response message is sent, which includes all relevant parameters including assigned addresses, DNS servers, and applicable QoS values. The Create Session Response message is followed by an initial Modify Bearer Request message. When the host setup is completed with a Modify Bearer Response message, downstream data can then flow toward the eNodeB. If IPv6 is enabled, an unsolicited RA is sent.

Figure 180: High-level example of GTP access setup shows a high-level overview of the setup call flow using RADIUS authentication.

Figure 180: High-level example of GTP access setup



13.2.2.1 Supported IP stacks

IPv4 and IPv6 SLAAC are supported via dual-stack bearers. IP addresses can be provisioned during authentication or through Local Address Assignment. Any DNS (IPv4 and IPv6) and NBNS (IPv4 only) addresses are signaled in the PCO IE. An unsolicited RA is sent after the first Modify Bearer Request message is received. An RS-triggered RA is also supported and either RA can be configured to contain DNS servers. Stateless DHCPv6 can also be used to retrieve the DNS configuration.

13.2.3 Mobility and location tracking

GTP access supports X2 handover and Tracking Area Update (TAU) without SGW relocation procedures as defined in TS 23.401.

GTP also supports subscriptions to location reporting. The required level of location reporting (for example, ECGI, TAI, ECGI+TAI) can be configured and overridden by AAA. Location reporting is only enabled if support is signaled by an MME. If an MME changes as part of a TAU procedure, change reporting is disabled unless the new MME also explicitly signals support.

Updated ULI can be learned in any regular procedure such as X2 handover, TAU, and UE-triggered service request. Additionally, the Location Change Reporting procedure, as defined in TS 23.401, is supported to signal location changes in absence of any other change.

Any AAA messages sent by the router always contain the latest ULI for the related GTP session. Additionally, RADIUS Accounting and Gx support triggered updates whenever the ULI changes. For RADIUS, this support is configured locally in the **radius-accounting-policy**. For Gx, the PCRF needs to subscribe to the USER_LOCATION_CHANGE, RAT_CHANGE, or ECGI_CHANGE event as specified in TS 29.214. In case of RAT/ECGI event subscription, this is directly reflected in GTP Change Reporting Action, overriding any local configuration. If only a generic USER_LOCATION_CHANGE subscription is requested, the GTP signaled action depends on the local configuration, requiring **change-reporting-action** to be configured in the applicable **peer-profile**.

13.2.4 QoS

ESM QoS using subscriber and SLA profiles also applies to GTP hosts. SR OS provides various methods to align internal QoS objects with 3GPP signaling of UL/DL APN-AMBR values:

- The incoming APN-AMBR from GTP can be reflected in RADIUS and Diameter. Additionally, this can be mapped to an SR OS QoS override using the command **configure subscriber-mgmt gtp apn-policy apn ambr-qos-mapping**. This uses generic QoS overrides, meaning that a subsequent QoS-override from AAA, for example, removes this mapping again.
- The APN-AMBR IE signaled in Gx can similarly be mapped to QoS overrides using the command **configure subscriber-mgmt diameter-application-policy gx 3gpp-qos-mapping**. This has precedence over the similar GTP command.
- The APN-AMBR IE signaled in outgoing GTP messages is derived from either local QoS objects, a Gx/radius signaled value, the incoming GTP APN-AMBR, or default values under **configure subscriber-mgmt gtp peer-profile mme qos ambr**, in that priority order. The mapping of local QoS objects to APN-AMBR is done with the command **configure subscriber-mgmt sla-profile egress/ingress report-rate**.

Values for QCI and ARP can be reflected from AAA or default values can be configured under **configure subscriber-mgmt gtp peer-profile mme qos arp/qci**. There is no interaction with local QoS objects.

SR OS maintains the FC classification and in- and out-of-profile markings consistent over the two data plane passes. The egress and ingress QoS policies in the SLA profile should be configured with the following considerations.

- Enable **de-mark** for access egress and map each FC to the dot1p as defined in Table 51: FC to dot1p mapping, therefore ensuring that the GTP-U encapsulated packet uses the same classification as the ESM context.
- Perform classification for access ingress based on dot1p as defined in Table 51: FC to dot1p mapping and enable **in-profile** and **de-1-out-of-profile** for each FC, therefore ensuring that the ESM context uses the same classification as determined for the GTP-U packet. However, a different classification scheme can be used, if required; for example, based on DSCP or IP criteria.

Table 51: FC to dot1p mapping

FC	dot1p
be	0

FC	dot1p
l2	1
af	2
l1	3
h2	4
ef	5
h1	6
nc	7

13.2.5 Multicast

GTP supports the reception of MLD and IGMP, with either a redirect-interface or per-host replication enabled. Per-SAP replication is not applicable and is ignored.

13.3 DHCP over GTP-u

A routed gateway (RG) with a fixed-wireless WAN link can get an IPv4 address and an IPv6 /128 address using DHCPv4 and DHCPv6 IA_NA, respectively. The RG can also get an IPv6 prefix using DHCPv6 PD for its LAN. A fixed-wireless RG can be an RG with integrated LTE or 5G modem, or connected to an external 5G modem using Ethernet or WLAN. After the default bearer and PDU session is created using NAS signaling, DHCP messages can be forwarded by the RG (with integrated modem), or the standalone modem (referred to as user equipment (UE)), over the default bearer toward the BNG, which functions as fixed-wireless gateway and terminates the GTP tunnel. DHCP messages are received and sent by the BNG over the GTP tunnel.

13.3.1 Address management related PCOs

Without DHCPv4, the fixed-wireless RG can only get an IPv4 address during initial attach using NAS SM (session management) procedures. As part of NAS signaling, the BNG acting as a fixed-wireless gateway allocates an IPv4 address and returns it in a PDN address allocation (PAA) IE in the GTP session creation response.

- If an RG requires IPv4 address allocation using DHCPv4 instead of address allocation during initial attach using NAS signaling, then it can signal a PCO "deferred address allocation" (value 0x000b) in GTP session setup. If this PCO is present in GTP setup request, then the BNG does not allocate an IPv4 address during GTP session setup. The GTP setup completes with IPv4 address of 0.0.0.0 returned in a PAA to the RG. The BNG only allocates and assign an IPv4 address using subsequent DHCPv4 from the RG, which is received over the GTP-u tunnel (S1-U interface) on the BNG.
- If PCO "allocation via NAS" (value 0x000a) is present in a GTP setup request, then the IPv4 address is allocated during GTP setup and returned in a GTP session creation response.

- If neither PCO is present, then by default, anIPv4 address is allocated on the GTP setup, and returned in the GTP session creation response. This can be overridden using the **skip-gtp-ipv4-alloc** CLI command. In this case, no IPv4 address is allocated during GTP setup and is only allocated using a subsequent DHCPv4 exchange.

The **skip-gtp-ipv4-alloc** configuration is a per-APN configuration. This can be overridden on a per-subscriber basis using a RADIUS VSA (`Alc-Gtp-Skip-Ipv4-Alloc-Override`).

13.3.2 Address allocation modes

DHCPv6 IA_NA can be used for IPv6 WAN address assignment. IPv6 WAN prefix can also be supported using SLAAC.

For DHCPv4, the following address allocation modes are supported.

- **DHCPv4 proxy**

  The IPv4 address offered in DHCPv4 from the RG can be provided from AAA during initial authentication in the framed-IP-address RADIUS attribute.

- **DHCPv4 relay to a local or external DHCPv4 server for IPv4 address allocation**

  The pool can be provided from AAA during initial authentication in the framed-pool RADIUS attribute, and can be added in the relayed DHCP packets with a DHCP Option 82 vendor-specific sub-option.

For DHCPv6, the following address allocation modes are supported.

- **DHCPv6 proxy**

  The /128 IPv6 address for RG WAN using DHCPv6 IA_NA can be provided from AAA during initial authentication in Alc-IPv6-address VSA. The PD prefix requested by RG using DHCPv6 IA_PD can be provided from AAA in the Delegated-IPv6-Prefix RADIUS attribute.

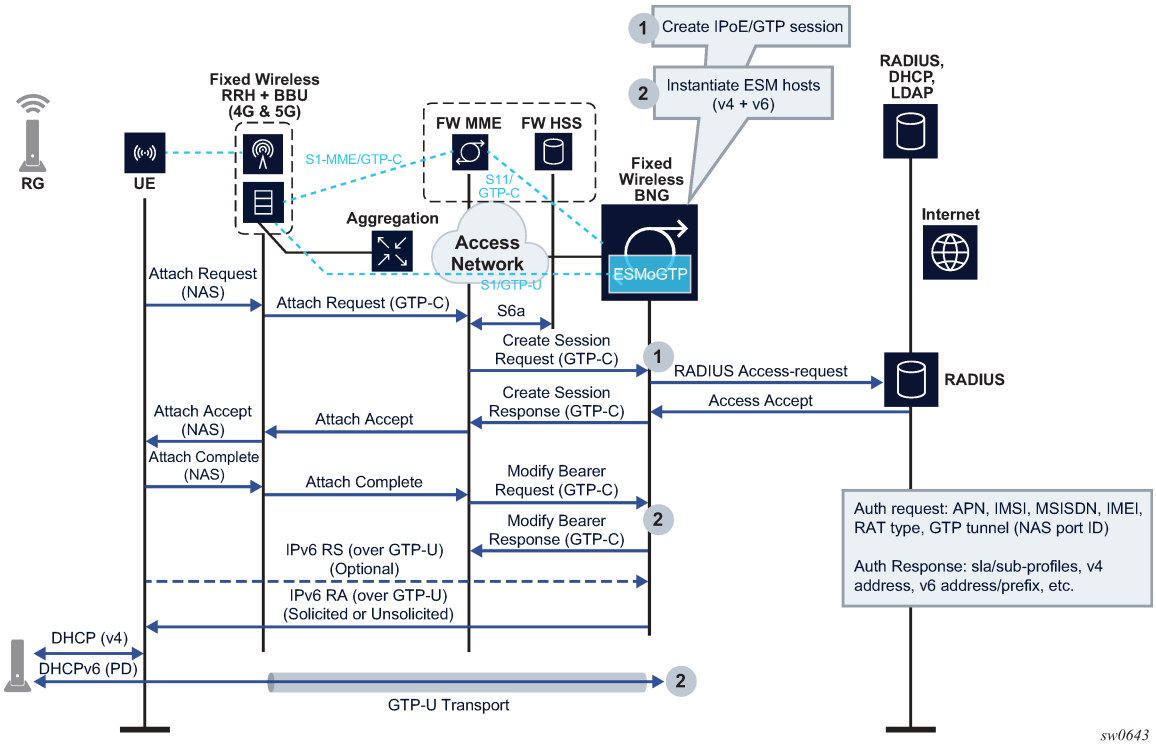
- **DHCPv6 relay to a local or external DHCPv6 server for IPv6 address and prefix allocation**

  The pool for /128 IPv6 address allocation for DHCPv6 IA_NA can be provided by AAA during initial authentication in the framed-ipv6-pool attribute. Pool for PD prefix can be provided by AAA during initial authentication in the Alc-Delegated-IPv6-Pool VSA. These are signaled by DHCPv6 relay to the server.

Authentication is performed only during GTP session setup. All the RADIUS attributes related to DHCP and IP stack (host) setup are received during this authentication. No subsequent authentication is performed on the receiving DHCP. A single IPoE session containing all the hosts (DHCPv4, DHCPv6 IA_NA, and SLAAC) are created per GTP session. PD prefix should be configured to be added as a managed-route if both DHCPv6 IA_NA and SLAAC are in use and creating respective hosts.

Figure 181: Call flow for GTP session setup shows the call flow for GTP session setup, subsequent DHCP exchange, and host creation in case of RG with a standalone modem (user equipment is noted as UE in the diagram).

Figure 181: Call flow for GTP session setup



DHCP over GTP-u tunnel is not supported in conjunction with bonding (hybrid-access).

DHCP over GTP-u tunnel is supported in conjunction with multiple APNs. The APNs can have different address management (for example, in case of two APNs). It is possible to use WAN address assignment using NAS for one APN and DHCPv4 for another APN.

Broadcast flag should not be set in DHCP messages from the client.

13.4 GTP peering

SR OS tracks each GTP-C peer for which it has at least a single GTP session or PDP context active. It tracks the peer's operational state with the following mechanisms:

- Regular GTP echo messages and parameters are configurable on a per-mgw-profile basis. When the echo mechanism fails, the peer is considered down.
- Active route entries toward the peer are monitored. If no route toward the peer is available, the peer is considered down.
- The Restart counter value of the peer is monitored. This is initially learned when the first active session or context is created. If the value is not available in regular messaging, an echo request is sent out immediately to learn the correct value. If the Restart counter is incremented during any later messaging exchange, the peer is considered rebooted.

When a peer is considered down or rebooted, all active GTP sessions and PDP contexts are forcefully removed.

SR OS also keeps a recovery counter in a persistent state, and increments this value on every reboot. This value is kept in the `restcntr.txt` file on CF3 and may not be modified or removed. This value is included in every control plane message.

SR OS responds to GTP echo messages for both active peers and unknown sources. This can be restricted using CPM filters if required. An incoming echo request from an unknown source does not create a peer; this can only be done by setting up GTP sessions or PDP contexts.

14 Virtual Residential Gateway

This section describes virtualized residential gateway (vRGW).

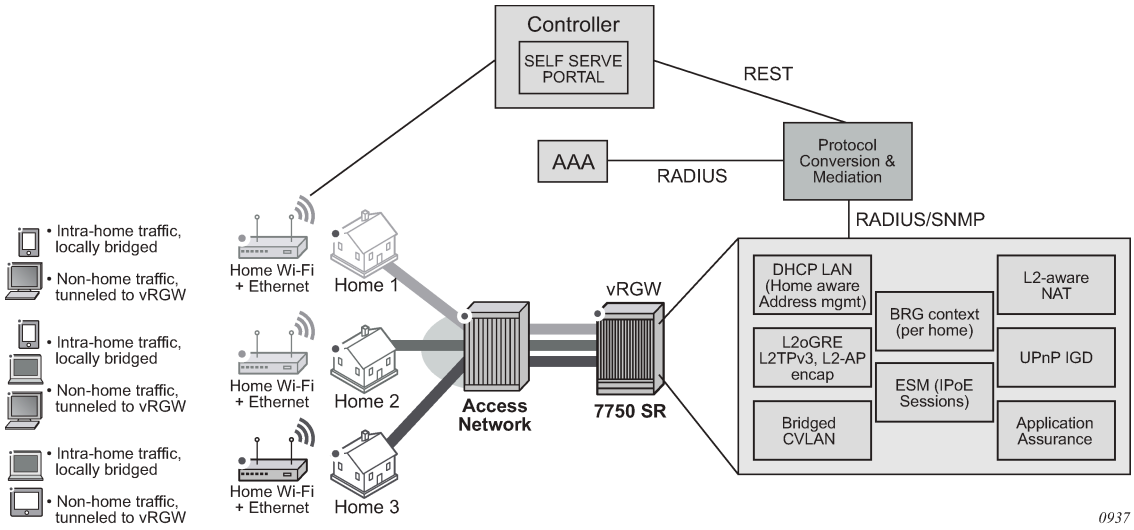
14.1 Overview

A Virtualized Residential Gateway (vRGW) transforms the Layer 3 routed Residential Gateway (RGW) in the home into a Bridged Residential Gateway (BRG), by moving the Layer 3 functions of an RGW into the network resides on a node (vRGW). The Layer 3 functionalities (such as address management, routing, Internet connectivity, NAT, UPnP, firewall, and application awareness) provides functions such as URL filtering and parental control and is moved to the network and resides on the vRGW. The RGW then operates in a bridged mode and performs local switching of intra-home traffic that originates and terminates on devices within the home. Bridged traffic destined for outside the home (such as to the Internet, the provider's content, or another home) over the WAN toward the vRGW is hereby referred to as a BRG.

This mode of operation allows an operator visibility of the connected devices on the home LAN, instead of just a single IP address per home as is the case for a Layer 3 RGW. This allows operational improvements through per-device control and troubleshooting, and the ability to offer new services faster and on a more granular device specific-basis.

The router, as a dedicated gateway or as a BNG, serves as a vRGW providing Layer 3 termination and ESM functionality for bridged homes, as shown in Figure 182: Virtualized Residential Gateway (vRGW).

Figure 182: Virtualized Residential Gateway (vRGW)



14.1.1 Access modes

All vRGW functionality is supported on both regular group interfaces (SAPs, LAGs, PW-SAPs, and so on) and WLAN-GW group interfaces (soft-GRE, soft-L2TPv3, VLAN).

A new configuration sub-node for the BRG is provided under the **group-interface** context for regular group interfaces and under the **vlan-range** context for WLAN-GW group interfaces. A regular group interface with a BRG sub-node does not support any non-BRG configuration and must operate in the **ipoe-bridged** mode.

In a WLAN-GW group interface, the BRG is configured in the **vlan-range** level. With a **vlan-range** it is not possible to mix BRG and other existing functionalities, but it is possible to mix BRG and other functionalities (such as WLAN-GW) on the same group interface. If a BRG also supports public Wi-Fi, the expectation is that the BRG has different SSIDs for public Wi-Fi and for private home traffic on the BRG, each represented by a different VLAN tag.

Contrary to WLAN-GW UEs, which require anchoring based on their MAC addresses (for mobility), devices associated with a BRG are anchored based on the tunnel source IP address of the BRG. The system therefore load-balances on a per-BRG granularity basis across a set of configured ISAs. Anchoring based on the source IP address of the BRG allows all devices in the home to be anchored on the same ISA and IOM. This enables aggregate QoS functionality within a single home.

Tunnel QoS is not supported as this is performed by regular subscriber QoS in the BRG scenario. WLAN-GW IOM (N:M) redundancy is supported. Data-triggered authentication (IPv4 only) is supported. All WLAN-GW access types (GRE, L2TPv3, L2-AP) are supported.

14.1.2 Home context on the vRGW

The system keeps a management context for each connected home or BRG, which is identified by a BRG ID. This context is used for authentication, configuration, and retrieval of operational data. Authentication can be performed explicitly by the BRG, in which case the vRGW acts as a RADIUS proxy. Alternatively, the vRGW can implicitly authenticate on the BRG's behalf when the first host from the BRG connects.

Initial configuration parameters are provided as RADIUS attributes during authentication. Configuration parameters provided on the home level is used as defaults on the host level can be overridden on a per-host basis. Home-level configuration can be dynamically overridden by means of a RADIUS CoA message. See the "Virtual Residential Gateway" section in the *7450 ESS, 7750 SR, and VSR RADIUS Attributes Reference Guide*.

Operational data provides information about the BRG, such as the number of connected devices, how long the BRG has been active, and which configuration parameters were provided. This can be retrieved via MIBs or show commands.

The vRGW communicates with a classic AAA server mainly to perform the authentication and with a separate configuration or operational platform. This separate configuration/operational platform is referred to as a "controller" and can be combined in a single management control system.

14.1.2.1 Implicit home authentication

With implicit home authentication (see Figure 183: Implicit home authentication), the vRGW authenticates a new BRG when the first associated device connects. To avoid restrictions on the connectivity model, the vRGW initially attempts to identify a BRG with a BRG ID. The vRGW always performs authentication of a new host, and this authentication should return the BRG ID with which this host is associated. If this BRG ID is not yet known, the vRGW triggers BRG-level authentication. This allows an operator flexibility to

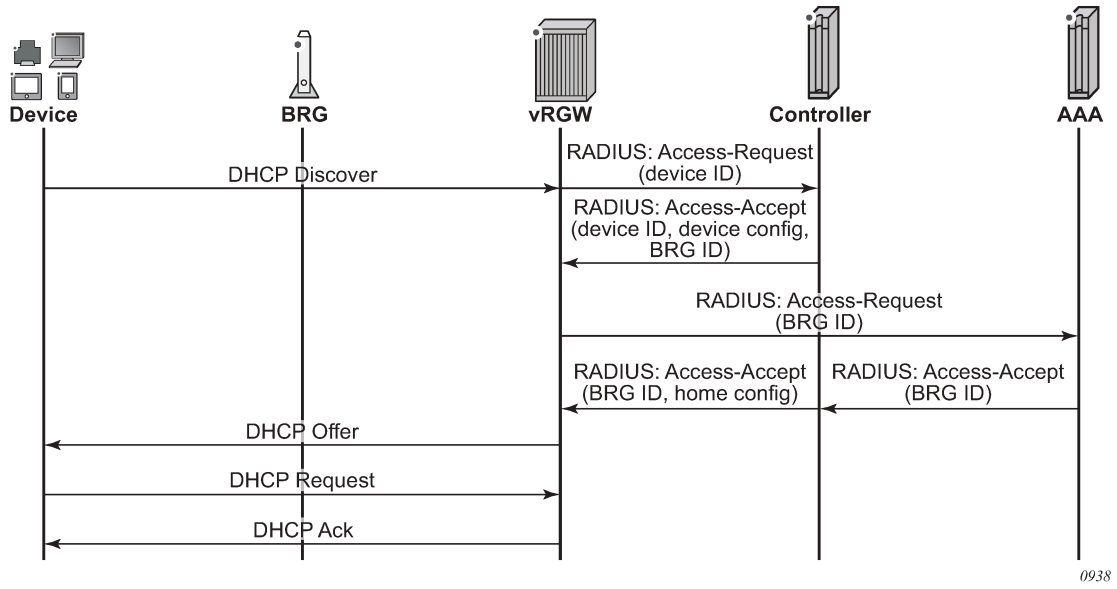
identify a home. For example, one deployment may use CVLANs as an identifier while another may use a BRG MAC as an identifier.

When using the BRG MAC, this can be learned using the same methods as the AP MAC for WLAN-GW, as described in Wi-Fi Aggregation and Offload, Authentication. The learned MAC can subsequently be reflected in authentication of the new device and mapped to a suitable BRG-ID by the AAA server or controller.

While a home can be optionally authenticated with an AAA server, each device in the home, is typically only authorized with a controller (via regular RADIUS messages and configured authentication policy) to get its configuration. This authorization is handled by the controller, which identifies and returns the associated BRG ID and, optionally, any device-level configuration.

Per-home authentication is forwarded to the AAA server to be fully authenticated. In this case, the controller typically performs an AAA proxy functionality so it can insert home configuration data in the final Access-Accept message. After both the device and BRG authentication are completed, the resulting RADIUS attributes are used to set up all required ESM objects (hosts, subscribers, SLA profile instances, filters, and so on).

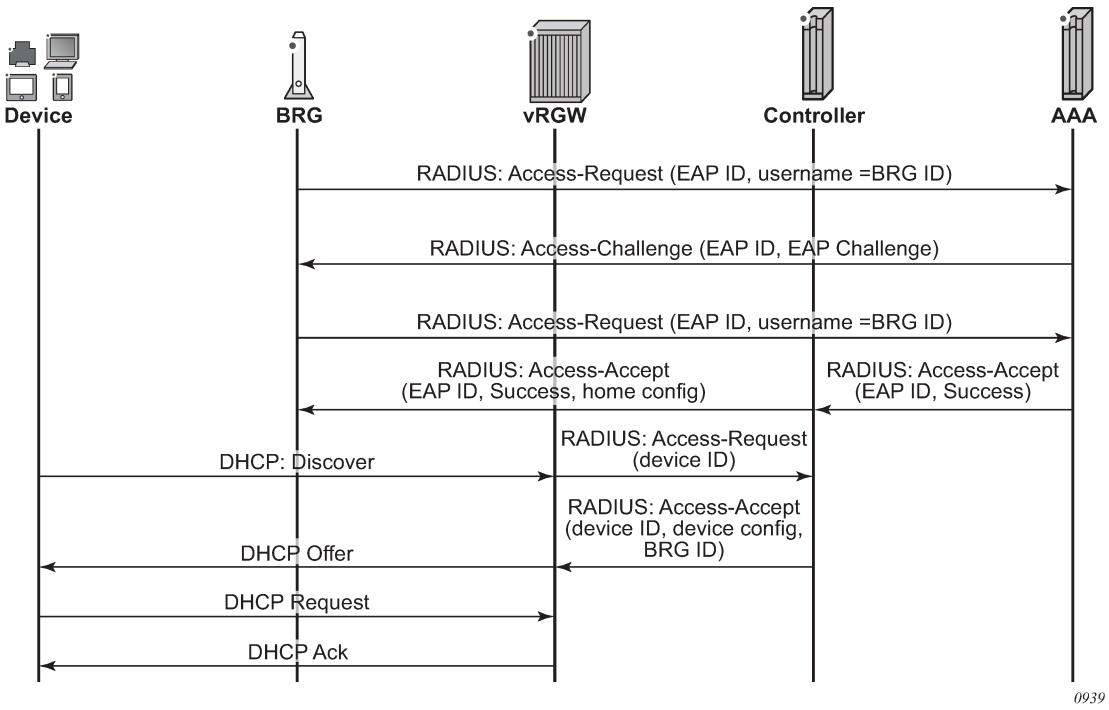
Figure 183: Implicit home authentication



14.1.2.2 Explicit home authentication

With explicit home authentication (see Figure 184: Explicit home authentication), the BRG authenticates itself at boot time before allowing connectivity of any devices. The vRGW acts as a proxy for this authentication, and upon receipt of a successful authentication, the BRG context is created and stores all home parameters. As with implicit authentication, the vRGW performs authentication for each device and expects the BRG ID as a result.

Figure 184: Explicit home authentication



On the router, this functionality is obtained by including a RADIUS proxy in the vRGW configuration. This RADIUS proxy can be used for both WLAN-GW and vRGW scenarios and determines the difference based on the presence of a BRG ID attribute in the Access-Accept message. Proxy transactions in the context of a BRG are not cached. Instead, a fully functional BRG context is already created. Track accounting is not supported in the context of a BRG. Re-authentication is supported and behaves the same as on a CoA at the BRG level.

14.1.2.3 Change of configuration

RADIUS CoA can be used to change the configuration on a home level. The CoA can use the BRG ID as a key and contain attributes for all new or updated parameters. For all correctly specified parameters, the vRGW overwrites the existing configuration on the home level and updates all devices connected to the BRG.

14.1.2.4 Home lifetime

A configurable ping based on a smart connectivity-verification mechanism is provided to detect whether a BRG is alive. The BRG state is removed if it is not deemed alive (subject to a configurable hold time). A BRG is always considered to be alive if there are connected non-static devices. Static devices are not considered because there is no control plane to track whether they are alive. When no sessions are present, the home context can be in one of the following conditions.

- On explicit authentication, where there is no session connection, the BRG state should be kept alive. To control this, an initial hold time applies during which the BRG context cannot be removed. Any

connected device cancels the timer. If the timer expires and there are still no connections present, the BRG fall backs to the behavior as if the last session was removed.

- If the last session was removed and connectivity verification is configured, the vRGW tries to ping the BRG. This ping mechanism uses ICMP or ICMPv6 ping toward the tunnel source IP or, if not present, toward the RADIUS source IP address. If the pinging is successful, the BRG context is kept. If the pinging is unsuccessful or impossible to perform (for example, if no IP address is known), the BRG falls back to the hold time. If, during the connectivity verification, a host connects to the BRG, the verification is stopped and the BRG is assumed to be alive again.
- If the last session was removed and connectivity verification failed or was not configured, a configurable hold time applies, during which time the BRG is not removed. If a new device connects during a hold time, the timer is canceled. If the timer is not canceled, the BRG is removed when the timer expires. The configured hold time can be zero and can be different from the initial hold time.

14.1.3 Device context on the vRGW

The router maps every device in a home to an IPoE session, which can contain one or more ESM hosts, depending on the number of IP stacks active on the device. The vRGW authorizes every single device and stores the resulting configuration data with the IPoE session. Whenever the session is created or updated, the vRGW combines the configuration data of the home context with the configuration data of the device. When the same configuration object is present on both levels, the device level applies. The resulting combined configuration data is used to install or update the ESM objects.

14.1.4 Dynamic configuration changes

Use a RADIUS CoA to dynamically change the configuration on both the home level and the device level. Use the BRG ID as the key on the CoA on the home level. The included attributes overwrites the existing stored configuration on the home level and subsequently update all devices connected to the home. All devices pick up the new home-level configuration unless a more specific configuration exists on the device level.

A CoA on the device level can use all existing ESM keys as detailed in the *7450 ESS, 7750 SR, and VSR RADIUS Attributes Reference Guide*. The keys update the existing configuration on the device level. If the device inherits the specified parameter from the home level, this parameter is updated on the device level; the existing home-level configuration remains intact and is used as the default for other devices.

In specific cases, the configuration can be removed on the device level to be able to fall back to the home level, supported by the RADIUS attribute Alc-Remove-Override. This attribute lists which overrides must be removed on the session level and fall back (revert) to the home-level configuration. If the home parameter changes, the device also picks up the new configuration.

For more information about RADIUS configuration attributes, see the *7450 ESS, 7750 SR, and VSR RADIUS Attributes Reference Guide*.

14.1.5 Per-home pool management and Layer 2–aware NAT

This feature allows the provisioning of a DHCP pool per home in a vRGW context. The addresses in a per-home pool are unique, so local bridging can occur within the home functions. Devices in different homes can, however, be allocated the same address. Layer 2–aware NAT is then used to handle address translation and connectivity toward the network and between homes. Per-home default gateway, subnet,

and address range (out of which addresses are allocated to home devices) can be configured via the CLI. The address range can be overridden from RADIUS. The subnet associated with the home pool must be within the configured L2-Aware inside subnet. L2-Aware source NAT can optionally be combined with destination NAT to support enhanced traffic redirection (such as for stateful DNS overwrite function). More details on network address translation forwarding can be found in the *7450 ESS, 7750 SR, and VSR Multiservice ISA and ESA Guide*. In addition to the IPv4 addresses from the home pool, the following hosts can be set up:

- IPv4 DHCP proxy hosts using an AAA-provisioned address

  Only non L2-Aware hosts can receive a framed IP address from RADIUS.
- IPv6 SLAAC hosts using an AAA-provisioned address or a local address assignment
- IPv6 IA_NA hosts using DHCPv6 relay, an AAA-provisioned address, or a local address assignment

14.1.5.1 Sticky IP addresses

The vRGW can be programmed with a list of sticky IP addresses per BRG. A sticky IP address is an IPv4 address from the home pool allocation range that is set aside for a specific device based on the MAC address and is never allocated to any other device. This can be used for devices where a persistent IP address is desirable but configuring a static IP address on the device is too cumbersome (such as a network printer). The device still uses DHCP to gain connectivity but is always assigned the same IP address. Public (non-NAT) sticky IP addresses are not directly supported in pool management but can be provisioned in the following manner.

1. Create a local DHCP server and pool to manage the public address space.
2. For each public sticky IP address that must be provisioned, create a sticky lease in this pool. See DHCP Address Reservation for Sticky Leases for further details. Both the user identification and hostname can be the MAC address of the associated device.
3. When the device linked to the sticky lease authorizes with the controller, it returns the allocated address as a framed IP address. This address takes precedence over any home pool configuration and a DHCP proxy host is installed.

14.1.5.2 Managed static IPv4 addresses

Like sticky IP addresses, the vRGW can be programmed with a list of static IP addresses per BRG. However, for static IP addresses, DHCP is not expected from clients and is dropped. The vRGW automatically installs all the addresses as IPv4 hosts contained in IPoE sessions. Because every host must be linked to a specific point of access (such as a SAP or tunnel) the vRGW needs to wait on a data trigger or the first non-static host in a home before static hosts can be created. Like regular (DHCP-triggered) hosts, a device-level authorization is performed for each static host to retrieve the per-device level configuration. This type of static host can cover both private (inside the home subnet) and public (non-Layer 2–aware NAT) addresses. The created hosts can only be explicitly removed by the management interface, and mechanisms such as session-timeout and idle-timeout are not supported.

14.1.5.3 DMZ

A single DMZ address can be provisioned per home via RADIUS (VSA Alc-DMZ-Address), which can be any address inside the home subnet except for the default gateway. When a host exists with this address, the DMZ mode is activated. DMZ is be activated if a host exists with the address and the subscriber uses

only one port range per IP. Without DMZ mode enabled, any traffic arriving for the NAT outside IP that does not match an existing flow, pinhole, or port-forward, is dropped. With DMZ mode enabled, this traffic is forwarded to the provisioned DMZ host.

14.1.6 IPv6

The vRGW supports the following IPv6 host types:

- IPv6 SLAAC hosts using an AAA-provisioned address or a local address assignment
- IPv6 IA_NA hosts using DHCPv6 relay, an AAA-provisioned address, or a local address assignment

The vRGW only operates in IPoE bridge mode. For regular group interfaces, the IPoE bridge mode must be explicitly enabled before the BRG can be enabled. For WLAN-GW group interfaces, the IPoE bridge mode is implicitly assumed for a VLAN with BRG enabled. This has the following consequences:

- SLAAC hosts from the same home can share a /64 prefix
- when a local address assignment is used, SLAAC hosts from the same home are automatically assigned the same /64 prefix
- IA_NA hosts from the same BRG can receive unique addresses from a shared /64 prefix. This prefix is automatically signaled in an RA

14.1.7 QoS and filter support

The vRGW is based on existing ESM QoS configurations on the router where each home maps to a single subscriber instance. Home-level bandwidth can be provided by an aggregate rate or scheduler policy on the subscriber level. Groups of home devices (such as telephones, computers, televisions, and security systems) can be given a shared bandwidth by assigning a different SLA profile per such group. Individual device-level QoS can be provisioned by mapping a specific device to a specific forwarding class using IP classifiers based on the device IP. This forwarding class can then be mapped to its own individual queue. Dynamic QoS overrides can be provisioned on both the home level and the device level. While QoS overrides are represented by a single RADIUS attribute, the device level does not override the whole attribute but only the QoS objects specified on that level. For example, on the home level, the scheduler bandwidth is overridden, and on the device level, queue bandwidth is overridden. The resulting override is a combination of both.

14.1.8 Data-triggered authentication

ISA-based IPv4 data-triggered authentication (Figure 185: Data-triggered authentication) and host creation is supported on WLAN-GW group interfaces. When authenticating a data-triggered device, connection data provides less data for the controller to derive the BRG, unlike DHCP where the BRG can insert its identifier, such as a circuit ID. For pure Layer 2 access, a BRG ID can be hard-coded to a port and VLANs, although for tunneled access, this is not always possible as the corresponding value would be the tunnel source IP address. This IP address can be dynamically assigned and changed with a BRG reboot. The following alternatives are suggested.

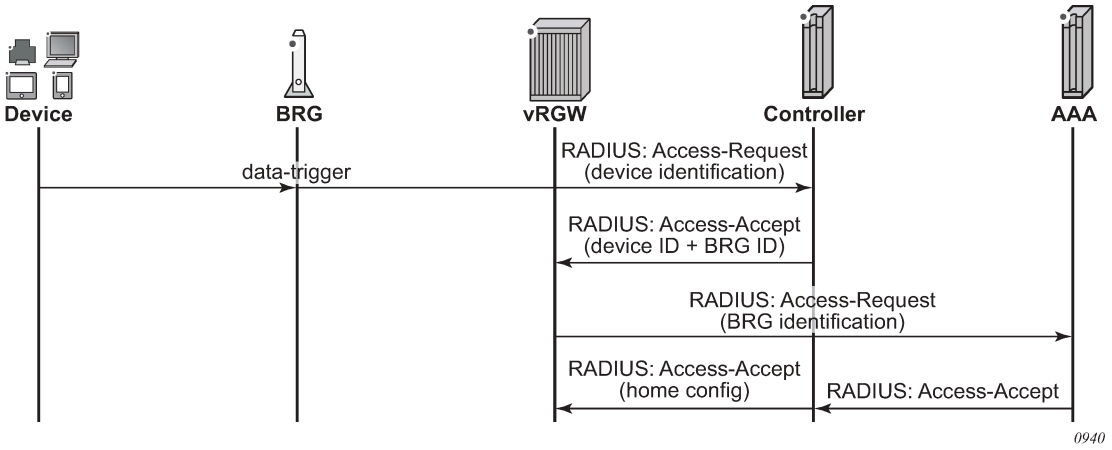
- Use AP MAC as the identifier. This identifier can be signaled in DHCP and DHCPv6 as specified in Wi-Fi Aggregation and Offload. For data-trigger purposes, it can also be sent as part of the L2TPv3 header, or if a GRE, it can be learned upon data-trigger via an ARPoGRE/NDoGRE message.

- Use a custom identifier that is sent in DHCP in a circuit ID option or a vendor-specific option. To handle the data trigger while a DHCP lease is active, a controller keeps its state to map the device MAC to the BRG identifier.
- If it the data trigger was for a static IP address (for example, when the static device is the first to send upstream data in the home), the triggering static host and any other provisioned static hosts are installed.

  If the data-triggered device is the first device to come up in the home and the BRG did not perform explicit authentication, the vRGW also triggers an implicit authentication. After authentication, the data-triggered host can be installed by one of the following methods.
- If the trigger was sent for a dynamic host (sticky/not sticky), (for example, when connection with a device was lost (based on an idle-timeout) but the lease was still valid), a DHCP lease is re-created using the provisioned lease time on the home level. This installed lease time is usually excessive compared to the actual remaining lease time on the device, but this is corrected when the lease performs DHCP renew or rebind procedures.

  The actual remaining lease time is used if known. If a host goes idle and sends a data trigger, the actual remaining lease time is used.

Figure 185: Data-triggered authentication



14.1.9 Per-host NAT port ranges

Carriers want to offer opt-in value-added services (VAS) through dedicated DPI-based appliances or VMs in data centers. This functionality requires vRGW support to forward traffic (ideally, only for subscribers who have subscribed to the VAS) to the external appliance. The appliance implements per-subscriber and per-device policies, and must be able to determine the subscriber and device from the received packets. Because address space across homes can overlap, subscriber-aware NAT is a requirement in the vRGW architecture. When subscriber-aware NAT is used, the outside IP address is unique and corresponds to the subscriber but, by default, the device information is lost. However, the device can be determined for the external appliance from the Layer 3 packet if a unique NAT outside port range is used per device on the vRGW.

By default, the subscriber-aware NAT allows the entire port range (other than the port range for static port forwarding) to be available for dynamic NAT flows and dynamic port forwarding (via UPnP). The feature adds support for allocating per-host NAT outside port ranges, and reporting per-host port-range allocation

and deallocation in RADIUS accounting. External VAS appliances can then track RADIUS accounting to determine device to port-range mapping.

The port range for a host is allocated and deallocated when the host is created and deleted, respectively. A single port range per host is supported. The RADIUS attribute **Alc-Per-Host-Port-Range** provides the count of ports per host for a subscriber. See the *7450 ESS, 7750 SR, and VSR RADIUS Attributes Reference Guide* for information about the attribute format. In case of multiple NAT policies per subscriber, the attribute value is required to be the same for all policies.

The presence of the VSA implicitly enables the per-host NAT port-range allocation mode. The **ports-per-host** mode is only enabled (via the VSA) if vRGW is enabled (as indicated by the presence of BRG in **no shutdown**) under the VLAN range on the WLAN-GW interface, or on the group interface. The VSA can be present in **access-accept** for BRG authentication (implicit or explicit), and in CoA (with **Alc-BRG-Id** as the key). If a CoA is received with the **Alc-Per-Host-Port-Range** set to 0, it indicates the disabling of the per-host port-range mode.

If the *Alc-Per-Host-Port-Range* VSA is changed, the flows in the overlapping region between the new and old port ranges remain intact; any remaining flows are removed.

- When the **ports-per-host mode** is disabled (via the VSA), the new port range includes all ports except the SPFs, and in accordance with the preceding rule, no ports are deleted.
- When **ports-per-host** mode is enabled mid-session (via the VSA), the new port range falls within the old port range, where the old range contained all ports except the SPFs.

The ISA is updated from the CPM for all hosts when the **ports-per-host** mode changes, and cleanup occurs in accordance with the preceding rule.

The **per-host** *port-range* is included in the *Alc-Nat-Port-Range* attribute in per-host or per-session RADIUS accounting (in accordance with the currently supported format of the *Alc-Nat-Port-Range* VSA for Layer 2–aware NATs).

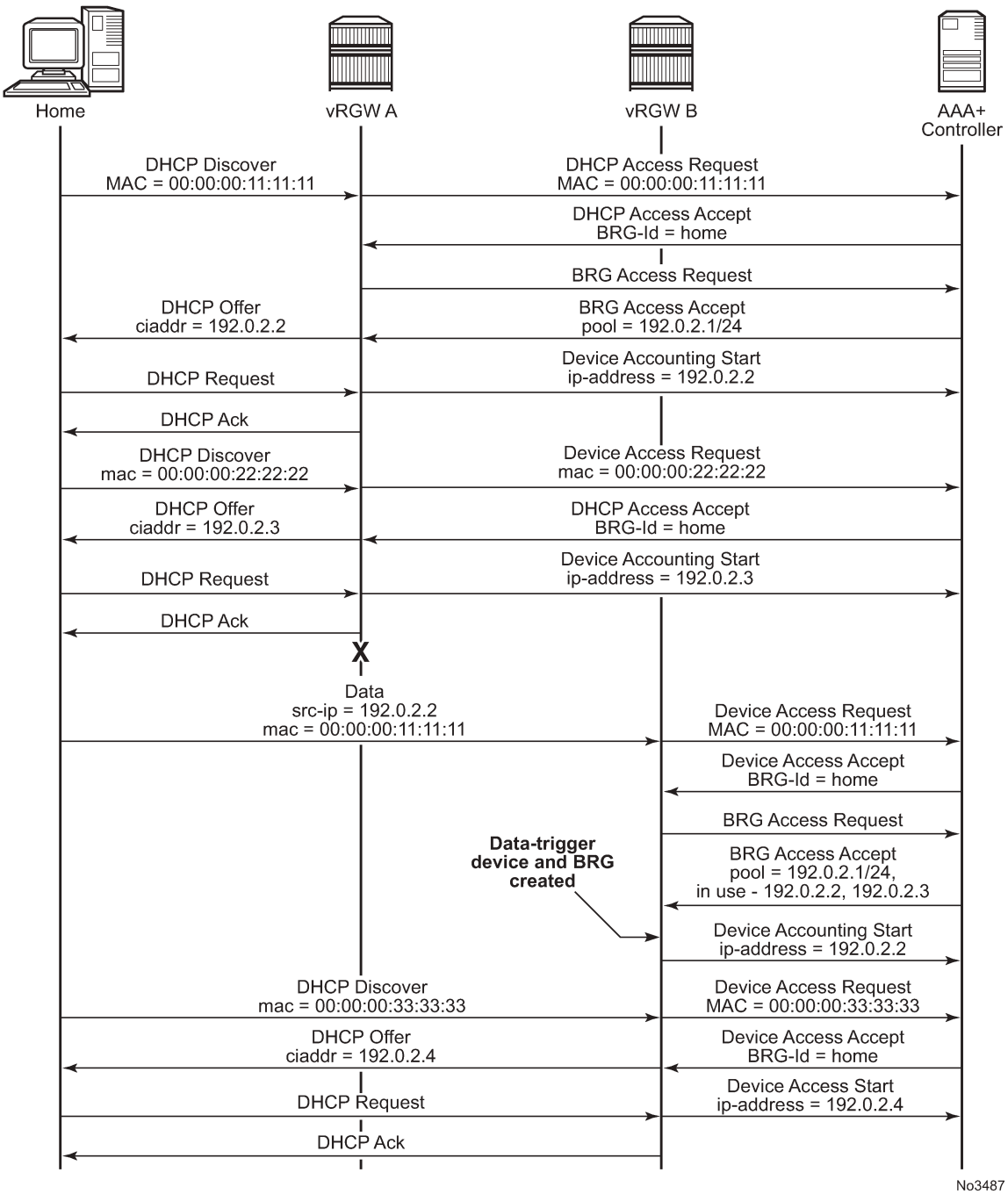
When a new port block for a host is allocated or freed, an interim-update message with a *Nat-Port-Range-Event* reason is sent. The interim update is sent only when interim updates are enabled and the configured **include-attributes** contain the *alc-port-range* VSA.

The port range for a host remains allocated for the lifetime of the host (unless explicitly removed using the VSA). Per-host reserved ports (for prioritized sessions) and watermarks to indicate exhaustion of **per-host** *port-range* are not supported.

14.1.10 Inter-chassis redundancy

vRGW supports stateless inter-chassis redundancy, where the home and device state is not explicitly synchronized between the two vRGWs. After a redundancy switchover event, Data Triggered Authentication (DTA) is used to download the correct home and device state back from the controller. The controller should always have the latest configuration state of a specific home, including any overrides after the home was activated. Additionally, the controller maintains a minimal state for DHCP pool management that can be reflected to the vRGW in case of a redundancy event, as shown in Figure 186: Inter-chassis redundancy.

Figure 186: Inter-chassis redundancy



14.1.10.1 Pool state synchronization

Because there is no stateful synchronization between pairs of redundant gateways, all pool state information is lost in a switchover event. The new gateway creates a new pool with state information for

all configured reserved and static address, but has no information about dynamic addresses. When a data-triggered host is authenticated, the pool tries to recreate the lease corresponding with this source IP address. However, this is subject to race conditions. If a data trigger is delayed (for example, the device is not active upon switchover) the DHCP pool may already have assigned the IP address to a new device in the home. This not only blocks the data-trigger for the old device, but may lead to address duplication issues inside the home. To solve this issue, the controller keeps track of allocated dynamic IP addresses. After a switchover event, these addresses can be sent to the vRGW as part of the BRG configuration. In this case, the per-home pool only allocates these addresses to data-triggered devices or DHCP renew triggered devices that request this specific IP address. New devices receive addresses not present on this list. After a configurable timeout, the pool relinquishes any unclaimed IP addresses back to the pool for general use. The currently unclaimed addresses can be displayed using the following command.

```
show subscriber-mgmt brg gateway brg-id "00:00:5e:00:53:05" standby-ip-addresses
===============================================================================
Bridged Residential Gateway home-aware pool standby IP addresses
===============================================================================
Home-aware pool : 00:00:5e:00:53:05
-------------------------------------------------------------------------------
192.0.2.10
192.0.2.11
192.0.2.12
-------------------------------------------------------------------------------
No. of standby IP addresses: 3
===============================================================================
```

It is recommended to keep the pool state on the controller basic, by marking an address as used when a device is connected and unused when it is disconnected. Periodically, explicitly synchronize the controller with the vRGW to remove addresses for which a device disconnect notification may have been lost.

14.1.10.2 Regular group interfaces

On regular group interfaces, redundancy is supported by enabling SRRP and IPoE session stateless redundancy. For more information about stateless redundancy, see Stateful Multi-Chassis Redundancy (MCS). The backup router is always in the backup-routing mode. There is no support for shunting traffic via a redundant interface from the standby to the active router.

IPv6 traffic and public IPv4 is attracted to the correct active (SRRP master state) router by the regular track-SRRP mechanism. Layer 2–aware NAT outside pools cannot use track-SRRP and should be configured distinctly on both routers making up the redundant pair. After a redundancy event, each home gains a new outside IP address on the new active router.

BRGs send out Gratuitous ARP (GARP) messages, which direct traffic to the correct active router. These GARPs update the FDB of the dual-homed Layer 2 aggregation nodes. GARPs are not supported for L2-Aware inside prefixes. At least one subscriber interface subnet must be explicitly configured to send a GARP.

When using MSAP deployments, the managed SAPs are not synchronized and must be recreated on the new active router via data-triggered authentication. Because there are no managed SAPs present, no GARPs can be sent when a standby router becomes active. If the aggregation node has a shared FDB for all C-VLANs (FDB per S-VLAN) or for all S-VLANs (FDB per switch), it is recommended to configure a single static SAP per S-VLAN or per port over which a single ARP can be sent.

14.1.10.3 WLAN-GW group interfaces

Redundancy on WLAN-GW group interfaces reuses the WLAN-GW monitor route mechanism to determine the active or standby state of each router in a pair of redundant routers. For more information see WLAN-GW 1:1 Active-Backup Redundancy. When using the BRG MAC as a BRG identifier, an ARPoGRE (IPv4 GRE tunnel) or NDoGRE (IPv6 GRE tunnel) is triggered when receiving a data-trigger for an unknown device. Authentication of the data trigger is delayed until the ARP/ND is answered or timed out.

```
config>service>vprn>sub-if>grp-if>wlan-gw
        learn-ap-mac delay-auth
```

IPv6 and public IPv4 traffic from the network is attracted to the active router because a subscriber interface on a standby router is operationally disabled and its associated subnets are not exported to the route table. Layer 2–aware NAT outside pools stay active on both routers and should be configured independently on both routers, making up the redundant pair. After a redundancy event, each home gains a new outside IP address on the new active router.

Traffic from the BRG is directed by regular routing, as only the active router exports the configured tunnel endpoints.

Access via the Layer 2 access point is not supported in a redundant setup.

Alternatively, vRGW with tunnel access also supports active and active redundancy using two different tunnel endpoint addresses. In this setup, each vRGW acts independently and does not synchronize any state. Each BRG is responsible for determining which vRGW is considered active; for example, by sending ICMP/ICMP6 echo requests toward the tunnel endpoints. After detecting a failover, the BRG is direct traffic toward the alternate tunnel endpoint. The state on this vRGW is re-created using the data triggers. This is like the monitor-route active or backup redundancy as discussed in WLAN-GW group interfaces. Explicit BRG authentication is supported in this use case if the BRG keeps track of a pair of tunnel endpoints and RADIUS server (proxy) addresses.

Redundancy of public IPv4 addresses and IPv6 addresses is not supported with an active/active deployment. After a failover event, it is not possible to send traffic for these addresses until the BRG restores connectivity to its original vRGW.

14.1.11 BRG and vRG restrictions

- Static IPv6 hosts are not supported in vRGW. It is possible to provide a static IPv4 host with a SLAAC prefix and use IPoE linking to automatically create an associated IPv6 host.
- Wholesale/retail is not supported in vRGW.
- Subnets provisioned for BRG pool management must lie in a pre-configured L2-Aware NAT inside prefix. The dynamic range of a BRG pool must not contain the configured Layer 2–aware NAT inside IP address.
- BRG connectivity verification is limited to a maximum of 50,000 BRGs. If more BRGs are connected, connectivity verification is performed for the excess numbers and only hold-time applies before the BRG instance is deleted from the vRGW.
- On regular group interfaces, only a single BRG is supported per SAP.
- There is a maximum of one SLAAC prefix per BRG.
- There is a maximum of 128 IPoE sessions per BRG, and a maximum of 8 static hosts per BRG.

- The idle timeout is based on SLA profile instance, not per host. For hosts under the same BRG sharing an SLA profile, it is not possible to detect early disconnect of a single host.
- All SLAAC hosts under a BRG sharing the same prefix uses a common forwarding context downstream. For predictable behavior, the same SLA profile should be used for each SLAAC host.

14.1.12 External allocation of Layer 2–aware NAT outside IP addresses

Certain residential deployments use multiple NAT outside IP addresses on routed physical raw (unknown encapsulation) packets, typically one per service. This feature provides similar support for multiple NAT outside IP addresses (up to four) per BRG (in other words, per subscriber) on vRGW. These addresses fall within locally-defined NAT pools on vRGW, but are assigned and managed by an external backend system. Each NAT outside IP address typically corresponds to a service, for instance, an HSI service may be NAT'ed to a different outside IP address than the voice service.

The NAT outside IP address and the corresponding NAT policy associated with the subscriber is provided by a VSA (Alc-Nat-Outside-IPs) in the RADIUS access-accept message or the CoA. Up to four instances of this VSA can be included in the RADIUS access-accept message or the CoA, which provides multiple NAT outside IP addresses for a BRG. The NAT pool referred to within the NAT policy must be configured for external assignment. If the provided outside IP address in the VSA does not fall in the NAT pool referenced within the corresponding NAT policy, then the outside IP address (and the mapping) is ignored. If the outside IP address falls within the NAT pool, and is not already allocated, then it is assigned to the L2-Aware subscriber. If the NAT policy contained in the VSA refers to a NAT pool that is not configured for external assignment, then the host setup fails. The external system is responsible for ensuring the stickiness of the outside IP addresses for the subscriber, if needed.

The system internally chooses an ISA in the NAT group to anchor an L2-Aware subscriber, such as the ISA, to where the NAT flows for the subscriber are created, based on upstream data or static port forwards. If a NAT outside IP address does not fall within the address block owned by the NAT ISA that anchors the L2-Aware subscriber, then the NAT outside IP address (/32) is added to the FDB with the ISA as the next hop. Otherwise, the downstream traffic forwarding to the NAT ISA for the L2-Aware subscriber follows the aggregate route corresponding to the address block owned by the NAT ISA.

The NAT outside IP address to use for a flow on the NAT ISA is based on a destination IP address lookup in a **nat-prefix-list** specified in the **sub-profile** for the subscriber. The **nat-prefix-list** contains a list of IP prefixes to NAT-policy mappings. The destination IP lookup in the **nat-prefix-list** provides the NAT policy to use. The NAT outside IP address that is associated with this NAT-policy is then used as the translated source IP.

```
config>service
     nat
       nat-policy "nat-policy-voice" create
           pool "voice-pool" router 401
       exit
       nat-prefix-list "nat-prefix-list-voice" application l2-aware-dest-to-policy create
           prefix 203.0.113.235/16 nat-policy "nat-policy-voice"
       exit
     exit
config>subscr-mgmt
        sub-profile "sub-prof-hsi-voice" create
            nat-policy "default-policy"
            nat-prefix-list "nat-prefix-list-voice"
        exit
config>service>vprn>nat>outside
    pool "voice-pool" nat-group 1 type l2-aware create
    port-reservation blocks 1
```

```
    external-assignment
    exit
```

To add a new NAT outside IP address for a subscriber by CoA, AAA must provide updated NAT outside IP-to-nat-policy mappings. AAA associates the correct **sub-profile** (containing the **nat-prefix-list** to map a destination IP prefix with the NAT policy that contains the new NAT outside IP address) with the subscriber; that is, the sub-profile associated with the subscriber needs to be changed using CoA. The CoA must always contain all the <NAT outside IP to-nat-policy> mappings associated with the subscriber (because it is cumulative). When the entire port range is available to the subscriber, the **port-reservation blocks** *num-blocks* command should be configured as 1 if **external-assignment** is enabled on the NAT pool.

All associated NAT outside IP addresses and corresponding NAT policies can be displayed via the **show service nat l2-aware-subscribers** command.

```
A:Dut-A# show service nat l2-aware-subscribers
===============================================================================
Layer-2-Aware NAT subscribers
===============================================================================
Subscriber                          : sub-2-4-ext
-------------------------------------------------------------------------------
ISA NAT group                       : 1
ISA NAT group member                : 1
UPnP policy                         : (None)
Default NAT policy                  : nat-policy-hsi
Per-host port block size            : N/A
Firewall policy                     : (None)

Policy                              : nat-policy-hsi
Purpose                             : nat
Outside router                      : vprn100
Outside IP                          : 198.51.100.235
DNAT default IP address override    : (Not Specified)
DNAT disabled by override           : false
Ports                               : 1024-5119

Policy                              : nat-policy-voice
Purpose                             : nat
Outside router                      : vprn100
Outside IP                          : 198.51.100.245
DNAT default IP address override    : (Not Specified)
DNAT disabled by override           : false
Ports                               : 1024-5119

-------------------------------------------------------------------------------
No. of subscribers: 1
```

The **nat pool** show command output shows the attribute that controls external assignment.

```
show router 70 nat pool "vprn l2aw"
===============================================================================
NAT Pool vprn l2aw
===============================================================================
Description                         : (Not Specified)
ISA NAT Group                       : 4
Pool type                           : l2Aware
Applications                        : (None)
Admin state                         : outOfService
Mode                                : auto (napt)
Port forwarding dyn blocks reserved : 0
Port forwarding range               : 1 - 1023
Port reservation                    : 128 blocks
```

```
Block usage High Watermark (%)        : (Not Specified)
Block usage Low Watermark (%)         : (Not Specified)
Block usage (%)                       : < 1
External assignment                   : true
Last Mgmt Change                      : 05/17/2016 13:41:04
===============================================================================
```

14.1.13 PPPoE client

The vRGW allows operators to start a PPPoE client per BRG to a remote BNG. This PPPoE client forwards traffic that goes through NAT or firewall. The NAT-outside IP address and SLAAC prefix used are negotiated via the PPPoE client.

The PPPoE client runs directly on the ISA and encapsulates packets directly after NAT or firewall functionality. Encapsulated packets are sent out over an Epipe that is directly linked to the NAT group during configuration. Multiple clients can share a single Epipe if they have unique MAC addresses. Clients are automatically hashed over multiple ISAs in the NAT-group. By default, hashing is based on source and destination MAC; however, if many Epipes are being used and there is a high reuse of MAC addresses, Epipes should be configured to use per-service hashing instead.

A vRGW with PPPoE client enabled can still support routed hosts outside NAT or firewall.

14.1.13.1 PPPoE client setup

PPPoE client setup is triggered when a BRG is created with any PPPoE parameter provided during authentication. At least one PPPoE Epipe service ID and one PPPoE client policy name must be provided. A MAC address for the client should also be provided, but, if absent, the system attempts to parse the BRG ID as a MAC address. An optional service name can be configured that are reflected in PPPoE setup. LCP keepalives are defined in the PPPoE client policy. The client supports either no authentication, PAP, or CHAP. The username and password can be transmitted as parameters during BRG authentication.

The PPPoE client allows negotiation of both IPCP and IPv6CP with SLAAC and, optionally, stateless DHCPv6. The negotiated IPv4 address is used as a NAT-outside address and is not signaled to the clients in the home. The received IPv6 SLAAC prefix is reused as a SLAAC prefix toward the home. The Preferred and Valid lifetimes signaled toward home are not synchronized with the values received with the BRAS, but are only subject to local configuration. DHCPv6 hosts in the home cannot be subject to the PPPoE client.

The BRG can use the PPPoE client for address assignments instead of traditional sources such as AAA, LAA, or NAT pool. PPPoE client is mutually exclusive to other address sources. The used NAT and firewall policies must be configured in **l2-outside** mode. DNS, DNSv6, and NBNS can be provided by both the PPPoE client (IPCP, RA, stateless DHCPv6) or by vRGW AAA. Nokia recommends operators do not mix sources, because there is no strict precedence and only the last update is kept.

Setup of the BRG and devices linked to the BRG is blocked until setup of the PPPoE client is complete. If, after a timeout (configurable in the BRG profile) the client does not completely come up, setup of the BRG and devices continues with the information available. This allows, for example, IPv4 to continue setup even if IPv6 negotiation fails.

Aside from the MAC and service ID, most PPPoE client parameters can be changed while BRG is operational. Changing any PPPoE parameters causes the PPPoE client to restart, which impacts data traffic in the same way as a PPPoE client failure.

Non-LCP PPPoE control plane packets are always sent within the NC forwarding class and dot1p 0. Egress remarking can be used to change dot1p values.

14.1.13.2 PPPoE client failure

If there is a partial failure of the PPPoE client, no BRG or in-home device states are removed. All affected stacks no longer forwards traffic and any flow states may be removed. Setup of IPv4 hosts and routed (non-PPPoE client) devices continues. Existing SLAAC hosts are not removed, but a deprecation RA (with 0 lifetimes) is sent and no new SLAAC hosts are created.

The PPPoE client attempts to re-establish the connection, subject to the configured retry and backoff timers. Traffic is allowed again after the connection is restored. If a new IPv4 address is assigned, this is immediately applied as the new NAT outside address. If a new IPv6 prefix is assigned, IPv6 prefix replacement is triggered for existing SLAAC hosts.

If the peer MAC address changes during a reconnection, it may be hashed to a different NAT ISA. This is not supported by the system, and if it occurs, the entire BRG is removed.

14.1.13.3 LCP keepalive

LCP keepalive intervals are configured in the PPPoE client policy. While the regular PPPoE control plane is initiated from the CPM, keepalives are handled directly on the ISA. Because the ISA is aware of any traffic, this handling is optimized so that echo requests are not sent if active traffic is received from the BRAS. The keepalive timer is reset upon each received echo-request. These two mechanisms significantly reduce PPP keepalive messages without compromising liveness detection. Incoming echo requests are always responded to, if LCP is in the LCP Opened state.

LCP PPPoE packets are always sent within the NC forwarding class and dot1p 7. Egress remarking can be used to change dot1p values.

14.1.13.4 MRU/MTU

Each PPPoE client policy allows configuration of both MRU and MTU. The MRU is sent toward the peer during the discovery phase. If any of the MRU or MTU is bigger than 1492, the Max-Payload tag is included and is set to the maximum of the configured MRU and MTU values.

Upstream data packets are subject to the smallest of either the received MRU or the configured MTU. The node fragments IPv4 packets without the DF flag set and drops IPv6 packets and IPv4 packets with the DF flag set. For the packets with the DF flag set, appropriate ICMP or ICMP v6 error messages are sent.

14.1.14 SLAAC prefix replacement

SLAAC prefix deprecation is supported with vRGW. Contrary to generic prefix replacement, vRGW uses radius attribute Framed-IPv6-Prefix to change the prefix in a BRG-level CoA or re-authentication. For PPPoE clients, prefix replacement is performed whenever the SLAAC prefix is signaled via PPPoE changes. See SLAAC for more information about SLAAC prefix deprecation.

When a SLAAC prefix is assigned to a BRG that previously had no prefix, an unsolicited RA is generated for all sessions currently active for that BRG. Data-triggered host creation should be enabled to allow subsequent setup of IPv6 hosts.

SLAAC prefix deprecation is supported for a PPPoE client. Whenever the state of the SLAAC prefix is lost, such as when a PPPoE client goes down, an RA with 0 lifetimes is sent out and no new SLAAC hosts can

be set up. Existing hosts remain in the system but are no longer forwarded. When the same SLAAC prefix is reassigned, existing hosts begin forwarding traffic again.

14.2 Home LAN Extension

This section describes the Home LAN Extension (HLE) feature.

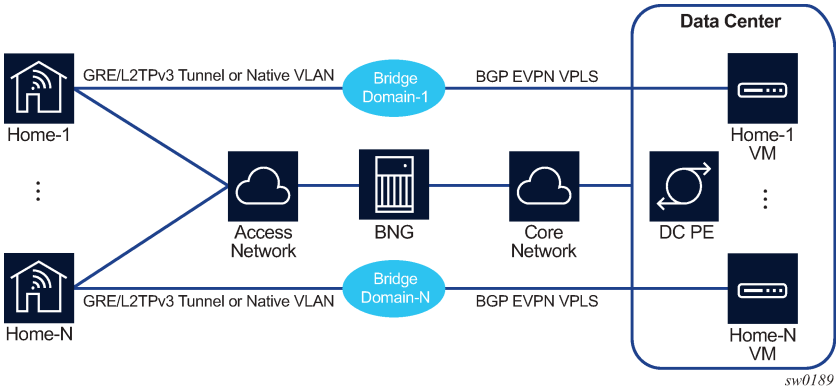
14.2.1 Overview

HLE allows an operator to extend the home network of the broadband user to a WAN network, such as a data center, by creating a per-home Bridge Domain (BD) on a Broadband Network Gateway (BNG). This feature provides the operator with the capability to deploy new services in a data center that have full Layer 2 reachability to the home and visibility to each individual host.

The per-home BD is created on a WLAN-GW ISA, and the system uses standard BGP EVPN VPLS services to extend the BD to remote networks.

Figure 187: Home LAN extension displays an example of an HLE configuration.

Figure 187: Home LAN extension



HLE requires the BNG to have Layer 2 access to the home network. The BNG supports the following types of access:

- soft GRE or L2TPv3 tunnel from a home gateway, which encapsulates Ethernet traffic from the host into a GRE/L2TPv3 tunnel
- native VLAN access that is terminated on the WLAN-GW group ISA using L2-AP access

A unique BD is created on the ISA for each home. The BD bridges traffic between the following connections:

- **access-facing connection (for example, home)**

  GRE/L2TPv3/L2-AP
- **network-facing connection (for example, DC)**

  BGP EVPN tunnel
- **ESM SAP-facing connection**

each home has its own ESM SAP

Each BD has a unique ID, which is a number returned by the RADIUS server as the Alc-Bridge-Id attribute during home authentication.

With HLE services, each home host is a WLAN-GW UE object and an ESM host object. Each network host (such as a VM in a data center) is a WLAN-GW UE object but not a ESM host. This means that a network host cannot use the BNG as the default router for other non-home-facing traffic.

HLE relies on SR OS vRGW functions, which means that the vRGW BRG in the same VLAN range of the WLAN-GW group interface must be enabled for HLE.

14.2.2 Authentication and authorization

Because HLE relies on the vRGW, it is carried in conjunction with vRGW home authentication and supports both implicit and explicit vRGW home authentication. Enabling the HLE service for a specific home (ESM subscriber) is triggered by a RADIUS server that returns an Alc-Bridge-Id message during home authentication.

The Alc-Bridge-Id attribute must be returned in both home and session authentication. The returned value must be equal on both authentication levels.

The Alc-Bridge-Id is the ID of the per-subscriber BD on the ISA. It is different from the Alc-BRG-ID, which is the ID of the BRG.

In addition to the Alc-Bridge-Id, the following optional attributes can be returned during BRG authentication:

- HLE BGP EVPN route target: Alc-RT
- HLE BGP EVPN route distinguisher: Alc-RD
- HLE BGP EVPN VXLAN VNI: Alc-Vxlan-VNI

14.2.3 Data plane tables

HLE is an essential Layer 2 bridge service. The data plane consists of a per- subscriber BD that contains the following tables:

- **MAC table**

  This table includes the learned MAC address by access and network connections.

- **flood table**

  This table includes flood destinations for broadcast, unknown unicast, multicast (BUM) traffic and typically contains access, and network connections.

- **IPv4 ARP table**

  This table includes learned ARP entries by access and network connections.

- **IPv6 neighbor table**

  This table includes learned neighbor entries by access and network connections.

The ARP and neighbor tables are populated when Assistive Address Resolution (AAR) is enabled.

When MAT is disabled, the ISA forwards packets based on the destination MAC address by performing a lookup in the MAC table and floods BUM traffic to destinations in the flood table.

For flooding traffic, the system does not send a to-be-flooded packet back to where it was received. If the packet is received from one remote network connection, the system does not forward it to another remote network connection.

14.2.4 BGP EVPN VPLS

HLE uses standard BGP EVPN VPLS services to extend the BD to the remote network. A BGP EVPN VPLS service is automatically created when a BD is created on the ISA.

HLE only supports VXLAN as a BGP EVPN data plane. When the next-hop field of the MP_REACH_NLRI is the VXLAN VTEP, it supports the following BGP EVPN route types:

- Type 2 MAC or IP advertisement route:
  - ESI: 0
  - Ethernet-tag: 0
  - IP address field:
    - sending: not used
    - receiving: used to populate the BD's ARP/neighbor table
  - Label-1: VNI

    Label-2: not used
  - BGP encapsulation extended community with tunnel type set to VXLAN
  - MAC mobility extended community: seq=0, sticky=1 (only when MAC NAT is enabled)
- Type 3 inclusive multicast Ethernet tag route: originating router address: system interface address
- BGP encapsulation extended community route with tunnel type set to VXLAN

PMSI tunnel attributes include:

- flags: 0
- tunnel type: ingress replication
- MPLS label:
  - sending: local VNI
  - receiving: remote VNI
- tunnel ID: VXLAN VTEP
  - sending: VTEP of the corresponding ISA
  - receiving: remote flooding destination

The VXLAN packet is forwarded in the base routing instance

14.2.5 Assistive Address Resolution

Assistive Address Resolution (AAR) is an optional HLE feature used to avoid sending ARP/ND requests from the home across a WAN to a remote network, or sending ARP/ND requests from the remote network across an access network to a home. The system responds to the ARP/ND request from the network or home with the learned ARP/ND entries, instead of flooding it.

With AAR, the system populates the ARP and neighbor tables with the learned ARP and neighbor entries, using the following methods:

- BGP EVPN MAC routes that contain an IP address
- (G)ARP/ND/NS packets

When the ISA receives an ARP/ND request, it performs a lookup in the BD's ARP and neighbor tables by using a target IP address as the key. The following conditions apply:

- If the requested IP address is over a different connection, then the ISA generates an ARP/ND reply with the match result without flooding the request.
- If the requested IP address is over same connection, then the ISA drops the request. If there is no match, then the ISA floods the address.

14.2.6 MAC Address Translation

MAC address translation (MAT) is an optional HLE feature that translates the access or network host's MAC address into a single BD MAC address. This feature decreases the number of BGP EVPN MAC routes-to-advertise per subscriber to one and eliminates the BGP update when hosts are created and removed which increases BGP's stability.

MAT is performed on traffic in both directions:

- **traffic from access to network**

  The system changes the source MAC address to BD MAC address before forwarding the packet to the network direction and change the destination MAC address into a real MAC address of the network host based on a ARP table lookup by using the destination IP address as key.

- **traffic from network to access**

  The system changes the destination MAC address (BD MAC) to the real host MAC address based on the ARP table or neighbor table lookup using the IP address as key, and change the source MAC address to a BD MAC address before forwarding the packet to the access direction.

With MAT enabled, upon receiving an ARP or neighbor request, the system performs a lookup in the ARP or neighbor table with the target IP address which determines the next action:

- When the target IP address is known, the system responds with a BD MAC address.
- When the target IP address is unknown, the request is flooded and the ARP or neighbor table is populated with the response.

Because of the ARP processing, the destination MAC of the received unicast packets would be BD MAC. The system only advertises BD MAC in EVPN when MAT is enabled. For unicast data packets received, if the destination MAC address is a BD MAC address, the ISA performs a lookup in the ARP or neighbor table by using the target IP address as key.

- If there is a match, the destination MAC address is changed to the MAC address that resulted from the ARP or neighbor table lookup and change the source MAC address to the BD MAC.
- If there is no match if it is unicast ARP/ND packet, then it is flooded, otherwise, it is forwarded to the BD's ESM SAP.

If the destination MAC address is a unicast MAC address, but not a BD MAC address, the destination MAC address is forwarded based on the MAC table lookup.

If the destination MAC address is a broadcast or multicast MAC address, then the packet is flooded.

MAT requires assistive address resolution to be enabled and with MAT enabled, the system also changes the source and target hardware addresses in ARP/ND requests and replies.

14.2.7 Configuring HLE

The following is an example of a configuration to enable HLE in the system with implicit home authentication.

1. Configure the vRGW BRG.
2. Enable HLE in the base routing instance:

```
config>router>vrgw# info
----------------------------------------------
            lanext
                vxlan-vtep-range start 198.51.100.235 end 198.51.100.245
                wlan-gw-group 1
                no shutdown
            exit
----------------------------------------------
```

3. Configure the HLE EVPN route target number:

```
config>subscr-mgmt>vrgw>lanext# info
----------------------------------------------
                router-target-as-number 100
----------------------------------------------
```

4. Configure the maximum number of BDs under the group interface:

```
config>service>vprn>sub-if>grp-if>wlan-gw
                max-lanext-bd 100
```

5. Configure BGP in the base routing instance:

```
config>router>bgp# info
----------------------------------------------
            group "evpn"
                family evpn
                type internal
                neighbor 2.2.2.2
                exit
            exit
            no shutdown
----------------------------------------------
```

6. Configure HLE under the VLAN range:

```
config>service>vprn>sub-if>grp-if>wlan-gw>ranges>range# info
----------------------------------------------
          authentication
              authentication-policy "isa-rad"
          exit
          authenticate-on-dhcp
          vrgw
              brg
                  default-brg-profile "brgp-1"
                  no shutdown
              exit
```

```
            lanext
                assistive-address-resolution
                bd-mac-prefix AA:BB:CC
                mac-translation
                no shutdown
            exit
        exit
-----------------------------------------------
```

7. Provision the RADIUS server to include Alc-Bridge-Id in the host and BRG records.

14.2.8 Traffic handling

The system allows an ISA policer to rate the limit per tunnel for ingress traffic on any of the following HLE connection types:

- Access
- Network
- Cross-connect

The policer name can be derived from either the local CLI configuration or from a RADIUS server during BRG authentication.

14.3 AP agnostic access for multiple dwelling units

14.3.1 Overview

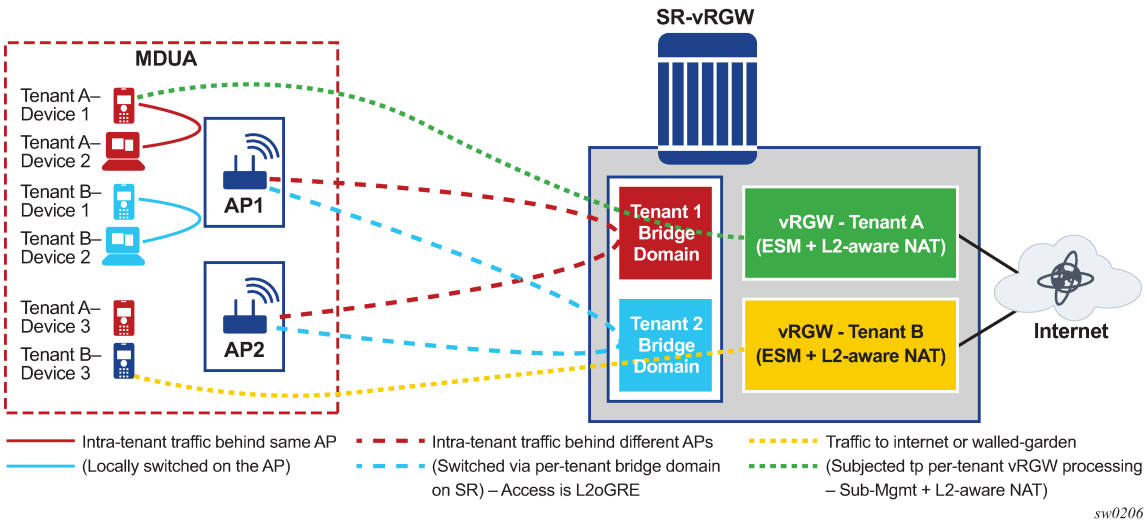
In a typical vRGW deployment, including HLE, the subscriber's BRG instance and BD on vRGW are tied to an access circuit (such as a soft GRE or soft L2TPv3 tunnel) from a single bridged Access Point (AP) or a residential gateway (RG). This feature adds support for subscriber to be AP agnostic. This means that a subscriber's BRG instance and BD are not tied to a single bridged AP or RG. This is particularly useful when the customer premise is an multidwelling (MDU) unit inhabited by multiple independent tenants where these tenants within the building can obtain connectivity from any bridged AP in the building. Bridged Wi-Fi AP and RGs can be installed in various parts of the building and are not owned or operated by specific tenants. Each AP can be provisioned with a common SSID (for example, an operator-branded SSID providing bulk Internet access and intra-MDU connectivity). Each AP accesses the network by an L2oGRE or L2TPv3 tunnel terminating on a gateway (vRGW) that provides integrated bridging and vRGW processing. The existing vRGW functionality is defined in Virtual Residential Gateway. The per tenant (access) bridging function on vRGW is described in Home LAN Extension. With this AP agnostic access feature, the traffic flow is handled as follows (Figure 188: AP agnostic access – integrated bridging and vRGW processing).

- Traffic between two devices of a tenant behind the same AP is locally switched by the AP.
- Traffic between two devices belonging to the same tenant connecting from two different APs is not locally switched by the APs but is tunneled to the vRGW or gateway that provides a bridged domain per tenant. This traffic is bridged by the gateway using the per-tenant bridge-domain (BD).
- Traffic between two devices belonging to different tenants connecting from the same or different AP in the building are tunneled to the gateway or vRGW. This traffic is subject to vRGW processing, such

as ESM followed by Layer 2–aware NAT. The isolation between tenants is provided by separate BRG contexts per tenant on the vRGW.

- Traffic from a tenant device, connecting from any AP in the building, that is destined for any destination on the Internet is forwarded by the AP over the tunnel and is subject to vRGW processing; for example, ESM followed by NAT.

Figure 188: AP agnostic access – integrated bridging and vRGW processing



14.3.2 Bridge domain and BRG identification

The tenant's bridge domain (BD) and BRG instance (for vRGW processing) is not identified by the tunnel source's IP address. Each tenant is provided a unique dot1q tag, which is maintained by the back-end system. The traffic from a device that is tunneled to vRGW carries this unique dot1q tag. The AP is aware of the tag per tenant and can provide traffic isolation between tenants. The BD and BRG instance for the tenant on vRGW are identified by the combination of the dot1q tag and the tunnel destination (that is, a gateway address configured under the WLAN-GW group-interface). Therefore, devices belonging to a tenant can connect from any AP, as the tenant context is not identified based on tunnel source (that is, the AP's IP address). This AP agnostic connectivity can be enabled on a per-VLAN range basis (in the **vrgw**>**lanext** context) as shown in the following example.

```
config>service>vprn>sub-if>grp-if>wlan-gw>ranges>range>
     vrgw
       lanext
          access
            [no] multi-access
          exit
       exit
     exit
```

The **multi-access** keyword indicates that the connecting device is not tied to a single AP and implicitly enables the identification of the BD and BRG instance for this device based on dot1q tag in the frame, and the gateway address, such as the L2oGRE or L2TPv3 tunnel endpoint IP address.

14.3.3 ARP handling

Split-horizon semantics are maintained for ARP handling with the per-tenant BD and ARP table. If the target IP of an incoming ARP request on the tunnel is in the ARP table for the tenant, and the tunnel source of the ARP request matches the tunnel source associated with the target IP, then no ARP response is generated. A lack of ARP response implies that both the source and target are behind the same AP and the ARP is handled on a local WLAN on the AP. If the target IP in the ARP request is not found in the tenant's ARP table, then the ARP request is flooded to all the access tunnels, such as to all known tunnel sources for the MAC addresses in the bridge-domain.

14.3.4 Mobility

An existing UE belonging to a tenant can connect from a different AP than its initial connection. This can be either because of mobility or a change in the signal strength of those APs in range. Both control and data-triggered mobility within context of the dot1q tag are supported. Control mobility includes DHCPv4, DHCPv6, and RS messages. All UEs for a specific tenant are anchored on the same ISA. When mobility is detected (by a change in the tunnel source for the UE), the impacted UE MAC address in the BD is updated to point to the new tunnel. Triggered interim accounting updates are generated, if enabled. vRGW functions in the SR OS are supported for this AP agnostic access.

14.4 Per-host DNS override

The DNS address override is enabled or disabled using a RADIUS VSA (Alc-Host-DNAT-Override), which can be present in access-accept or the COA corresponding to a host. This provides the operator with the capability to force DNS packets of a device in vRGW to the DNS servers of its choice. DNS override is achieved by subjecting DNS traffic to destination NAT. The traffic that is subjected to destination NAT is selected by applying a NAT classifier.

Overriding of the DNS address on a per-host basis also adds support for RADIUS VSA (Alc-Host-DNAT-Default-Address-Override) to specify a per-host default address for overriding the address in DNS packets. This per-host default address is used if no IP address is configured as part of the DNAT action within the classifier entry. This per-host default address can also be overridden by the RADIUS COA directed at the host. The per-host default address from RADIUS using the VSA can only be associated with a single NAT policy.

The per-host DNAT override can be removed using the Alc-Remove-Override attribute, in which the host inherits the DNAT override state at the BRG level. The per-host default address for DNS address override can also be removed using Alc-Remove-Override.

Additionally, the per-host DNS override extends existing NAT classifier match criteria to include foreign IP addresses in the match for selecting traffic that goes through the DNAT. A foreign IP address is a destination IP address on the NAT inside service before translation. Following is a sample configuration:

```
*A:vSIM>config>service>nat# info
----------------------------------------------
            nat-classifier "dns-override" create
                entry 1 create
                    action dnat ip-address 12.12.12.3
                    match protocol udp
                        dst-port-range start 53 end 53
                        foreign-ip 8.8.8.8
                    exit
```

```
                exit
                entry 2 create
                    action dnat ip-address 12.12.12.4
                    match protocol udp
                        dst-port-range start 53 end 53
                        foreign-ip 8.8.8.9
                    exit
                exit
            exit
```

The show command for L2-aware hosts is extended to include the per-host DNS override and default DNAT address set using the RADIUS VSA.

```
===============================================================================
Layer-2-Aware NAT hosts
===============================================================================
Subscriber                  : sub-1-1
Inside IP address           : 10.1.1.1
-------------------------------------------------------------------------------
Policy                      : pol 2
Bypassing                   : false
VAS filter                  : N/A
Override DNAT               : enable
DNAT default addr. override : 2.2.2.2
Outside router              : 101
Outside IP address          : 40.101.2.1
Port block                  : N/A
-------------------------------------------------------------------------------
No. of hosts: 1
===============================================================================
```

15 Service chaining for ESM hosts with Layer 2–aware NAT

This section describes service chaining for ESM hosts with Layer 2–aware NAT.

15.1 Steering to service chains for ESM hosts with Layer 2–aware NAT

This feature provides the steering of traffic flows for ESM hosts with Layer 2–aware NAT (typically used in vRGW and WLAN-GW), to service-functions (SF) in a data-center, reachable over an IP underlay network with a tunnel. The supported tunnel encapsulation is VXLAN. The steering function on the gateway (vRGW or WLAN-GW) is configured through a PBR action in an ISA filter (also referred to as VAS filter) associated with a Layer 2–aware NAT host. The PBR action specifies the IP address of the SF, the EVPN service instance through which the SF is reached, and optionally, an ESI. A controller in the data-center, and the subscriber edge (a gateway such as vRGW, WLAN-GW, or BNG) acting as a service-function-classifier (SFC) participate in BGP-EVPN to exchange reachability information for the SF in the DC, and NAT pools on the gateway. The gateway resolves the PBR target, in other words, SF IP@ in the EVPN service configured in the VAS filter via BGP EVPN routes received from the controller in the DC. The network virtualization edge (NVE) in the DC, such as a host stack running the SF in a VM can act as a bridge or a router. Nuage VRS or VSG is an example of an NVE.

The ISA filter used for steering can also be configured with an optional action to insert a network services header (NSH) in steered traffic, as described in RFC 8300, *Network Service Header (NSH)*.

A group of NAT ISAs that provide per-host or per-flow steering functions for L2-aware NAT hosts on the gateway are configurable under the base router. The steered traffic from the gateway to the VAS is VXLAN encapsulated on these ISAs. A range of IP addresses used for local VXLAN VTEPs on ISAs is configured (and routable) under the base router.

15.1.1 Terminology

Controller

A network element in the DC that peers with the subscriber edge to exchange BGP EVPN routes. The controller can advertise type-1 (ESI), type-2 (MAC advertisement, with an optional IP address), and type-5 (an IP prefix route containing an SF's IP address or subnet) routes for each SF. The controller also learns EVPN type-2 and type-5 routes from the subscriber edge for reachability of NAT pools on the ISAs. The controller can program the NVE with these routes.

EVPN import-mode

A configurable attribute of the EVPN service on the subscriber edge that controls the type of EVPN route information that is imported or tracked from the EVPN peer and exported or advertised to the EVPN peer. The possible values are **bridged**, **routed**, or **none**. For import mode **bridged**, the subscriber edge imports EVPN type-2 routes containing the SF's MAC and IP addresses and the associated VXLAN VTEP to reach the SF. With import mode **routed**, the EVPN type-5 route containing the SF's IP address or subnet and the type-2 route containing the NVE's IP and MAC address are imported or tracked. The NVE, in this case, routes traffic to or from the SF. Also, with **import-mode routed**, the subscriber edge advertises type-2 and

type-5 routes for NAT prefixes (outside pools). If no import-mode is configured, then only type-2 and type-5 routes for NAT prefixes are advertised to the peer (no routes are imported). See EVPN route updates and tracking.

NSH

Network Services Header. Extra encapsulation carried in the steered traffic over VXLAN tunnels that contains information about the packet handling through the chain of value added service functions in the DC. In addition to the information about the service chain that the steered traffic traverses, it also contains optional meta-data (for example, the subscriber-id). More details on NSH usage and format are defined in VAS filters on the ISA.

NVE

Network virtualization edge. A network element in the DC (for example, TOR or virtual switch or router such as the Nuage VRS) that provides connectivity to the SFs (VM or appliance). The NVE can perform bridging or routing functions for traffic to or from the SF. EVPN service instances (including R-VPLS tied to a VPRN) can be configured on the NVE, depending on the configured forwarding mode (bridging or routing to connected SFs).

Service-chaining service

An EVPN control plane instance on the subscriber edge that is used to learn and track EVPN routes for reachability to the SFs in the DC. It is also used to advertise EVPN routes for NAT pools on the subscriber edge to the controller in the DC.

SF

Service function. The value-added service (for example, a parental control service) running in a VM or on an appliance in the Data Center (DC). The SF can be part of a chain of SFs providing opt-in value added services.

SFC

Service function classifier. The component on the vRGW/WLAN-GW/BNG (subscriber edge) that matches flows belonging to the Layer 2–aware NAT host that has opted-in for value-added services, and performs steering (PBR) functions to the first service in the service chain that is applicable to the host. SFC supports optional NSH insertion in the steered traffic. Flow matching and steering is specified in VAS filters that are applied upstream and downstream flows on the ISA. SFC also tracks EVPN routes (type-1, type-2, and type-5) to resolve the SF's IP address, and optionally, ESI in an EVPN service, all of which is specified in the steering action in the VAS filters.

15.2 VAS filters on the ISA

A VAS filter can be associated with individual hosts of a Layer 2–aware NAT subscriber either at creation time or via CoA. The name of the VAS filter associated with the host is provided by the AAA server. This VAS filter is applied on each packet that creates a new flow. This results in a pair of actions which are then used for all subsequent upstream and downstream packets in that flow. These actions allow forwarding different kinds of traffic to different service chains in both upstream and downstream directions.

15.2.1 Matching

A VAS filter contains a set of ordered entries. Each entry contains a single match criteria and either an upstream action, a downstream action, or both. Matching occurs independently for both directions,

only considering the entries for which an action is configured in that direction. Matching stops at the first matching entry. Actions from different entries for either direction can result.

The VAS filter entries can match on:

- a foreign IP or subnet (for example, a destination IP and subnet for upstream traffic, and source IP and subnet for downstream traffic match)
- a foreign port (for example, a destination port for upstream and a source port for downstream traffic)
- a protocol

15.2.2 Forwarding

The resulting forwarding action can be one of the following:

- **action forward**

  The packet is forwarded as usual to its original destination.

- **action forward sf-ip** *sf-ip* [**esi** *esi*] **service-id** *svc-id*

  The packet is tunneled toward the **sf-ip** over EVPN/VXLAN overlay.

  If the **sf-ip** cannot be resolved (see EVPN route updates and tracking), the fail-action is used. This can be either forward or drop. The default is forward. Optionally, an NSH header can be inserted.

15.2.3 NSH insertion

NSH insertion is an optional action of the VAS filter when forwarding to the service chain. An additional header is inserted between the tunneling Ethernet header and the IP payload, as defined in RFC 8300, *Network Service Header (NSH)*. VXLAN-GPE is not supported.

IP => UDP (port 4789) => VXLAN => Ethernet =>NSH => IP

The NSH header is populated with following values:

- TTL in the base header is set to 63.
- The **service-path-id** (24-bit number) and **service-index** (8-bit number) are filled in from the filter entry's action.
- The MD-Type is set to 1 and the 16-byte metadata is filled in. The source of this metadata is discussed below. If no metadata is provided, this field is zero.

The metadata can be specified in the filter entry's action. The metadata can either be a 16-byte opaque data hex string (zero-padded if it is smaller than 16 bytes), or it can be derived from the subscriber string (in the Alc-Subsc-Id-Str VSA). In the latter case, the subscriber string is truncated after the first 16 bytes, and therefore, these first 16 bytes should be unique.

Alternatively, the opaque data string can be provided by AAA. This source has precedence over the filter entry's action.

15.2.4 Configuration

Modifications to an existing VAS filter are applicable to all hosts using that VAS filter.

- Adding filter entries is always possible and the new entry is considered on the next filter application at flow creation. Existing flows are not re-evaluated against the modified filter.
- Removing filter entries is only possible when the filter entry is shut down.
- Modifying the match criteria on a filter entry is only possible when the filter entry is shut down.
- Adding or removing a filter entry action is only possible when the filter entry is shut down.
- Modifying an existing filter entry's action is always possible and takes effect immediately for all flows currently matched against that filter entry action.

When a filter entry is shut down, all flows matched against any of the filter entry actions is killed. The **clear subscriber-mgmt isa-service-chaining vas-filter** *filter-name* command can be executed to clear all flows of all hosts associated with the filter. This command can be used if all flows needing re-evaluation against a modified set of matching criteria.

A maximum of 512 VAS filters with up to 64 entries per filter are supported.

15.3 EVPN route updates and tracking

15.3.1 NVE bridging to SF

In this scenario, the NVE acts as a bridge for upstream traffic. The service chain (EVPN) service on subscriber edge is configured with import-mode bridged. In this mode, the SF IP address in the VAS filter action is resolved to obtain SF's MAC address and VXLAN VTEP/VNI to reach it. If the filter action only specifies SF IP address and no ESI, then EVPN route type-2 (MAC route with MAC and IP address of SF) is used to resolve SF's IP address. If a route is found, then the SF MAC address and VXLAN VTEP/VNI from the type-2 route are used. If ESI is specified in the filter action, then the ESI is resolved using EVPN type-1 route. In this case, the MAC address is taken from route type-2, and VXLAN and VNI are taken from route type-1. The resolution is downloaded to the ISA. The upstream packet from the host is VXLAN encapsulated. The destination MAC in the inner Ethernet frame is the SF's MAC address. The source MAC is ISA's MAC address.

```
config>sub-mgmt>service-chaining
     service-chain 500 import-mode bridged create
         bgp
            route-distinguisher 65001:500
            route-target import 65000:500
         no shut
```

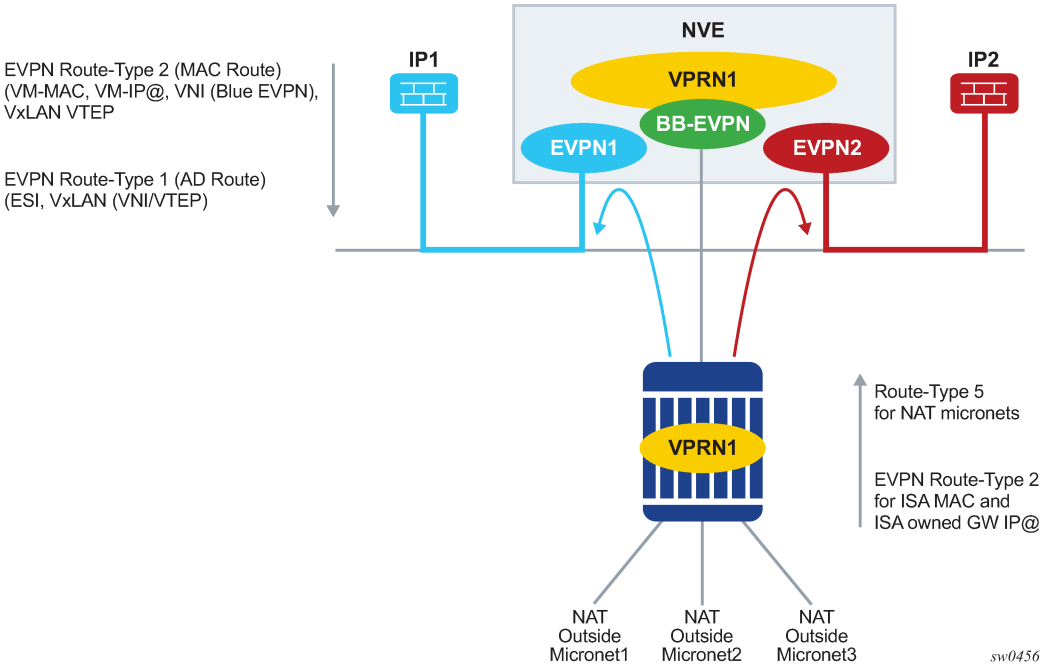
A single separate EVPN service (a backbone EVPN) can be configured between the controller and subscriber edge. This is used to advertise type-5 routes for NAT pools on the subscriber edge.

```
config>sub-mgmt>service-chaining
       service-chain 600 create
          bgp
             route-distinguisher 65001:600
             route-target export 65000:600
          evpn
             vxlan vni 10
             nat-group 1
             gw-address-range start 120.1.1.10 end 120.1.1.4
             ip-advertise-routes ipv4 nat
                 outside-svc 400 outside-pool pool-1
```

```
                outside-svc 400 outside-pool pool-2
```

A Layer 3 domain (overlay VPRN) must be configured on the NVE. The subscriber edge is connected to this Layer 3 domain on the NVE with an EVPN (the backbone EVPN or R-VPLS on NVE). The R-VPLS interface on the NVE (into this VPRN) must share the same subnet with the subscriber edge. NVE receives the type-5 routes in the backbone EVPN (R-VPLS), and adds the received type-5 prefixes to the FIB in the VPRN that the R-VPLS is connected to. The gateway IP address in the EVPN type-5 routes sent by the subscriber edge must be on the same subnet as the R-VPLS interface on NVE. A range of gateway IP addresses in this subnet are configured under the EVPN service on the subscriber edge, such that each individual ISA gets a gateway IP address to use for exported type-5 routes. NVE acts as a router for downstream traffic from the SF that is destined for NAT outside IP address of the subscriber (Figure 189: NVE bridging traffic to the SF).

Figure 189: NVE bridging traffic to the SF

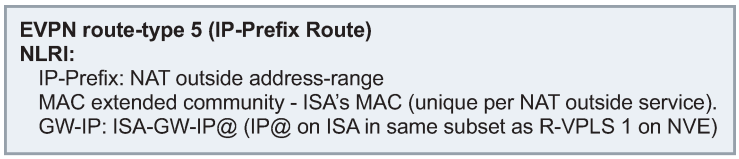


15.3.2 NVE routing to SF

In this situation, the NVE acts as a router for both upstream (subscriber edge to SF) and downstream traffic (SF to the subscriber edge). The SF must be configured with NVE as the default router.

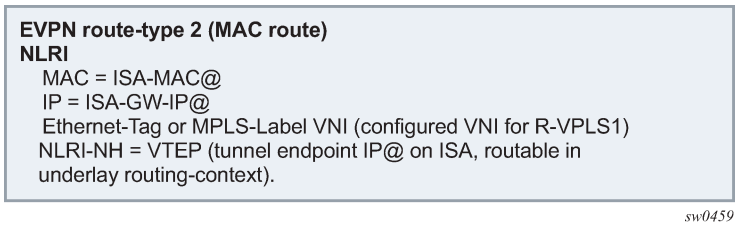
The service-chain (EVPN) service on subscriber edge is configured with import-mode **routed**. NAT pools are advertised in EVPN type-5 routes. The ISA's gateway IP address is sent in a type-5 route (Figure 190: EVPN route type-5).

Figure 190: EVPN route type-5



A type-2 router advertising ISA's MAC address and gateway IP address is generated (Figure 191: EVPN route type-2).

Figure 191: EVPN route type-2

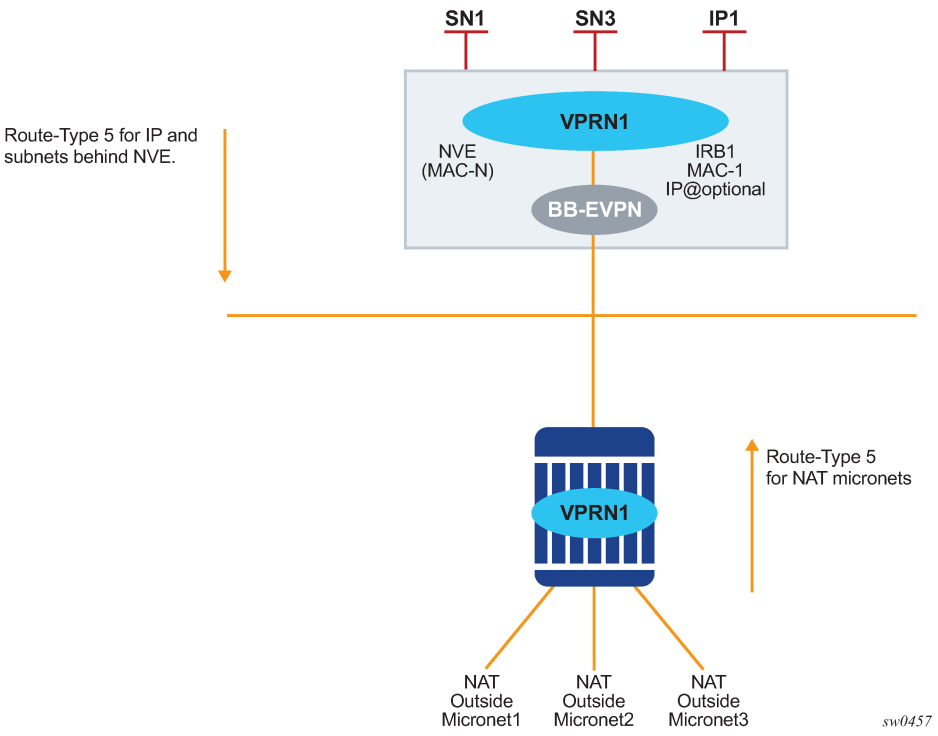


This EVPN service (called a backbone EVPN) is configured on NVE as an R-VPLS that ties to a VPRN as described in NVE routing to SF. NVE adds the received type-5 prefixes to the FIB in VPRN.

The controller in the data-center also generates type-5 routes carrying IP address or subnet for the SF. The MAC address in a type-5 route must be the R-VPLS interface MAC or NVE's system MAC address (because the NVE is routing traffic). NVE should also generate an EVPN route type-2 to advertise the MAC of the R-VPLS interface (or single MAC address of the NVE), and optionally the IP address of the R-VPLS interface. This is shown in Figure 192: NVE routing traffic to and from the SF.

On subscriber edge, SF's IP address in the filter action is resolved in the configured service in the filter action via a best match IP lookup against EVPN route type-5. If the resolved route type-5 has nonzero GW-IP, then a recursive lookup (exact match) is done in the service. If it is resolved by EVPN route type-2, then the next hop MAC (DA MAC that is used in inner Ethernet header) and the VXLAN VNI/VTEP are taken from the route type-2. If GW-IP in route type-5 is zero, then the dest MAC and next hop (VTEP) is take from the route type-5.

Figure 192: NVE routing traffic to and from the SF



15.4 Data path on the subscriber edge

15.4.1 Upstream traffic (access to network)

In this scenario, Layer 2–aware NAT is performed. Traffic is subjected to VAS filtering associated with the host. Based on the match entry, the output of the action contains an SF IP address, EVPN service instance, optional ESI, and optional NSH parameters (**service-path-id**, **service-index**, and optional **metadata**). SF IP address and optional ESI are resolved in the indicated EVPN service as per the configured import-mode of the EVPN service (described in the previous sections). The result of resolution is SF MAC address, VXLAN VTEP and VNI. The upstream packet is encapsulated as shown below:

With an optional NSH insertion, the encapsulation used is VXLAN, where the Ethernet header following the VXLAN carries Ether-Type NSH, as shown below. VXLAN-GPE is not supported.

IP => UDP (port 4789) => VXLAN => Ethernet => NSH => IP

- Outer IP source address: Local VTEP (ISA's local IP address)
- Outer IP destination address: Remote VTEP (from EVPN route, as described in previous sections)
- Destination MAC in Ethernet header: SF MAC address or NVE's MAC address (depending on bridging or routing on NVE).
- Source MAC in Ethernet header: MAC address of ISA (generated based on configured MAC prefix)
- VNI in VXLAN: VNI from EVPN route resolution (as defined in previous sections)

- Inner IP (payload) source address: NAT outside IP address
- Inner IP (payload) destination address: original destination on the Internet
- Without NSH the encapsulation is standard VXLAN, for example, IP => UDP => VXLAN => Ethernet => IP. The parameters in outer IP, inner Ethernet, and IP are as defined about (in other words, the same as for the encapsulation with NSH).

15.4.2 Downstream traffic — from network

The destination IP address in the downstream traffic is the subscriber's NAT outside IP address. The traffic is revived on the right NAT ISA. NAT flow is looked up. Associated VAS filter action is executed. The action can be steering (with or without NSH), with steering parameters, such as an SF IP address, EVPN service instance, (optional) ESI, and (optional) NSH parameters (**service-path-id**, **service-index**, **optional meta-data**). SF IP address and optional ESI are resolved in the indicated EVPN service as per the configured import-mode of the EVPN service (described in previous sections). The result of resolution is SF MAC address, VXLAN VTEP and VNI. The downstream packet steered to the SF is encapsulated similarly to upstream traffic described in NVE bridging to SF.

15.5 Data path on NVE

Upstream traffic from the subscriber edge is either bridged or routed by the NVE. The EVPN service is located based on the VNI. The destination MAC address in the Ethernet frame is either the SF's MAC address, in which case the traffic is bridged to the SF behind the NVE, or the destination MAC matches the local R-VPLS interface MAC address or single MAC address of the NVE. Then, the packet is routed (based on FIB lookup on the NVE) to the right SF.

Downstream traffic (network to subscriber edge to vas) after being processed by the VAS chain contains the NAT outside IP address in the IP destination address. SFs are configured with NVE as the default router. The packet is received on the NVE from the SF. A lookup is performed in the FIB on NVE (which contains IP prefixes populated by EVPN route type-5 updates). The next-hop IP address is the IP address of the ISA (GW-IP field in route type-5). An ARP and L2FDB lookup respectively provides the destination MAC address, and VXLAN VTEP and VNI. Traffic is VXLAN encapsulated and forwarded in the underlay toward the subscriber edge.

16 Dynamic data services

16.1 Introduction to dynamic data services

Dynamic data services enables a zero-touch, single-ended provisioning model for business services. Two deployment models are available:

- **With control channel**

  Triggered by the authentication of a single or dual-stack PPPoE or IPoE session as a business CPE control channel, parameters are passed in a RADIUS Access-Accept or CoA message to set up a Layer 2 or Layer 3 data service.

- **Without control channel**

  RADIUS or local authentication of a dynamic data service data trigger provides the required parameters to set up a Layer 2 or Layer 3 data service.

Dynamic Data Services support includes:

- Epipe VLL services
- Epipe VLL services with dynamic MS-PWs (FEC 129) or spoke SDP
- EVPN based Epipe service
- VPLS services
- VPLS services with BGP-AD pseudowire
- VPLS service with spoke SDP or mesh SDP
- EVPN based VPLS service
- IES and VPRN services
- MVPN services
- Routing policy, filter, operational groups and OAM-PM provisioning

The full list of supported configuration commands can be displayed with the **tools dump service dynamic-services command-list** CLI command. Dynamic data service SAPs must be located on dot1q- or QinQ-encapsulated Ethernet, anchoring or satellite ports and can be part of a LAG.

A Python script interface adds a flexible abstraction layer that reduces the OSS integration cost; only the business user specific service parameters, such as service type, IP address, QoS, and filter parameters, are required from RADIUS or local authentication. These parameters are then used in the Python script to generate a CLI template to set up the target Dynamic Data Service.

Dynamic data services configuration can be updated via a RADIUS CoA message.

Both XML accounting and RADIUS accounting for up to two different RADIUS destinations can be activated on a dynamic data service SAP.

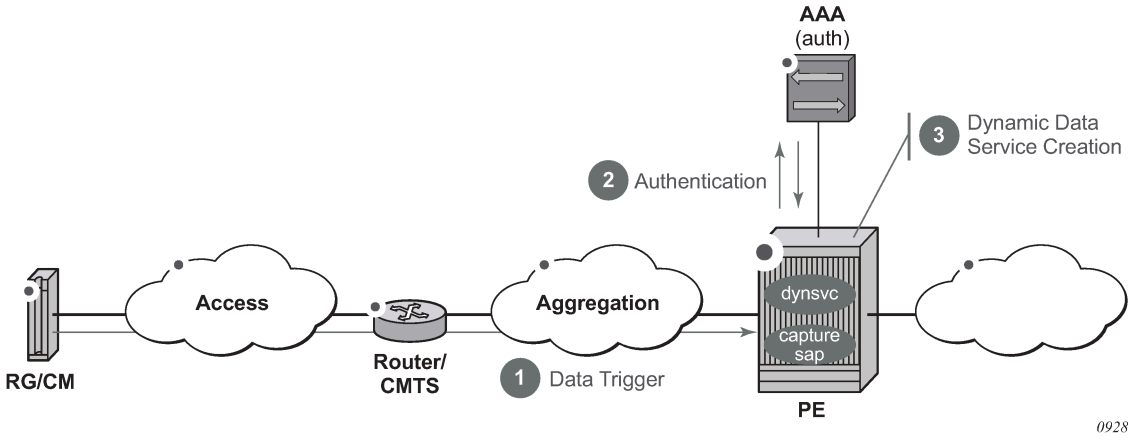
16.2 RADIUS-triggered dynamic data services associated with a PPPoE or IPoE session as control channel

See the "RADIUS-Triggered Dynamic Data Service Provisioning" section in the *Advanced Configuration Guide* for a detailed description.

16.3 Data-triggered dynamic data services

In the data-triggered dynamic data services model, as shown in Figure 193: Data-triggered dynamic data services, any frame arriving on a dynamic data services capture SAP can result in RADIUS or local authentication. The dynamic data service is then created from a Python script that generates CLI snippets using parameters obtained from authentication.

Figure 193: Data-triggered dynamic data services



16.3.1 Data trigger

A dynamic services data trigger is an object that is created when a frame received on a dynamic services capture SAP is sent to the control plane for authentication. A dynamic services data trigger is uniquely identified with its SAP ID. If the dynamic services data trigger was received on capture SAP x / y / z:*.* with outer-tag = a and inner-tag = b, then the dynamic service data trigger SAP ID is "x / y / z:a.b". For each dynamic services data trigger, the following information is stored.

Table 52: Data trigger information

Data trigger information	Description
Acct-Session-ID	The RADIUS accounting session ID for this dynamic services data trigger. This accounting session ID is used as accounting multisession ID in RADIUS accounting for associated dynamic services. It can also be used as a key in CoA or Disconnect Messages to set up or terminate associated dynamic services.

Data trigger information	Description
MAC	The MAC address learned to set up this dynamic service data trigger. The MAC address is included in the Access-Request message for RADIUS authentication.
IP	The IPv4 or IPv6 address learned to set up this dynamic service data trigger. If the data trigger packet was not an IP packet, then this field is empty. When available, the IP address is included in the RADIUS authentication and accounting messages.
State	The current state of the dynamic service data trigger: Pending: (initial state) data trigger received and authentication started Accepted: (transient state) authentication succeeded; dynsvc script started but not yet completed sapCreated: (final state) corresponding dynamic services SAP created

The dynamic services data trigger information can be displayed as follows:

```
# show service dynamic-services data-triggers
===============================================================================
Dynamic Services Data-triggers
===============================================================================
SAP                         : 1/1/4:1214.101
-------------------------------------------------------------------------------
Acct session-ID             : 144DFF0000009156A24138
MAC                         : 00:51:00:dd:01:01
IP                          :
State                       : sapCreated
-------------------------------------------------------------------------------
No. of Data-triggers: 1
===============================================================================
```

For a data-triggered dynamic data service to be successfully set up, a dynamic services SAP equal to the data trigger SAP ID must be created.

In the same way as the control channel model, multiple dynamic data services can be associated with a single dynamic services data trigger: up to 32 dynamic services during data trigger authentication or up to 4000 in total through CoA. When the dynamic services SAP that corresponds to a data trigger is deleted (teardown action), then all dynamic services associated with that dynamic services data trigger are deleted (teardown action).

16.3.2 Dynamic services data trigger capture SAP

In the same way as an Enhanced Subscriber Management (ESM) capture SAP is configured, a dynamic services data trigger capture SAP is configured in a VPLS service and captures frames for authentication. A dynamic service data trigger capture SAP does not forward traffic within the VPLS service, and no MAC learning occurs.

A VPLS capture SAP becomes a dynamic services data trigger capture SAP when a dynamic services policy is configured and the **dynamic-services** context is enabled (**no shutdown**). The dynamic services policy specifies the authentication mechanism to be used as detailed in the authentication sections below.

For example:

```
configure
    service
        vpls 10 customer 1 create
            sap 1/1/4:1214.* capture-sap create
                description "Dynamic Services Data Trigger capture-sap"
                dynamic-services
                    dynamic-services-policy "dyn-svc-1"
                    no shutdown
                exit
                no shutdown
            exit
            no shutdown
        exit
    exit
```

The **trigger-packet** type needs to be configured. A dynamic services data trigger **capture-sap** captures any valid Ethernet frames, including non-IP frames.

Valid dynamic services data trigger capture SAPs are:

- dot1q-encapsulated Ethernet, anchoring, satellite ports or LAGs
  - sap x/y/z:*
  - lag-x:*
  - pxc-x.a:*
  - esat-x/y/z:*
- QinQ-encapsulated Ethernet, anchoring, satellite ports or LAGs:
  - sap x/y/z:*.*
  - sap lag-x:*.*
  - sap x/y/z:tag.*
  - sap x/y/z:*.tag
  - sap lag-x:tag.*
  - sap lag-x:*.tag
  - pxc-x.a:*.*
  - pxc-x.a:tag.*
  - esat-x/y/z:*.*
  - esat-x/y/z:tag.*

Enhanced Subscriber Management (ESM) and dynamic services data trigger cannot be enabled simultaneously on a single capture SAP: a **no shutdown** command of the **dynamic-services** CLI context is mutually exclusive when configuring an ESM trigger-packet type.

Use the following CLI commands to display capture SAP statistics:

- To display the number of data trigger packets forwarded to the CPM and dropped on the IOM:

```
# show service id <service-id> sap <sap-id> sap-stats
===============================================================================
Service Access Points(SAP)
===============================================================================
--- snip ---
```

```
-------------------------------------------------------------------------------
Sap Statistics
-------------------------------------------------------------------------------
Last Cleared Time     : N/A
                        Packets                 Octets
CPM Ingress           : 108                     6696
Forwarding Engine Stats
Dropped               : 6144                    405504
```

Table 53: Data trigger packets forwarded/dropped counters

Counter	Description
CPM Ingress	Number of dynamic service data triggers received on the capture SAP that are forwarded to CPM
Forwarding Engine Stats Dropped	Number of dynamic service data triggers is received on the capture SAP that are dropped on IOM

- To display the dynamic services capture SAP control plane statistics (data triggers received and data trigger drop reason counters):

```
# show service id 10 dynamic-services capture-sap 1/1/4:1214.* statistics
===============================================================================
Dynamic Services Capture SAP 1/1/4:1214.* Statistics
===============================================================================
Data packets received by SAP                          : 3
-------------------------------------------------------------------------------
Drop Reason Counters
-------------------------------------------------------------------------------
No policy configured at capture SAP level             : 0
No authentication configured in policy                : 0
Data-trigger already exists                           : 0
Lockout is active                                     : 0
Reached data-trigger system limit                     : 0
No memory available                                   : 0
Unsuccessful authentication                           : 1
No data-trigger SAP-id in authentication              : 0
Corresponding dynamic SAP is not created              : 0
===============================================================================
```

Table 54: Data triggers received/dropped reason counters

Counter	Description
Data packets received by SAP	Number of dynamic service data triggers received on the **capture-sap** that reached the CPM
No policy configured at capture SAP level	No dynamic services policy configured at the **capture-sap**; required to determine the authentication destination.
No authentication configured in policy	The authentication section in the specified dynamic services policy is missing or incomplete.
Data-trigger already exists	A new data trigger frame is received for an existing data trigger that is authenticated, but the corresponding

Counter	Description
	dynamic SAP is not yet created. The new data trigger packet is dropped.
Lockout is active	The data trigger for this managed SAP is currently in a lockout state because of previous authentication failures.
Reached data-trigger system limit	The maximum number of dynamic service data triggers supported on the system is reached. Additional data triggers are dropped.
No memory available	There is not enough system memory available to process the data trigger.
Unsuccessful authentication	The authentication for a data trigger on this capture SAP failed or timed out.
No data-trigger SAP-id in authentication	The dynamic services data trigger SAP ID is not provided in authentication. This is a mandatory parameter.
Corresponding dynamic SAP is not created	The data trigger successfully authenticated but the corresponding dynamic SAP was not created. This is typically caused by a dynamic services script error.

- to clear the dynamic services capture SAP control plane statistics:

```
# clear service id <service-id> dynamic-services capture-sap <sap-id> statistics
```

16.3.3 RADIUS authentication

When a valid Ethernet frame is received on a dynamic services data trigger capture SAP, it is sent to the control plane for authentication. The dynamic services policy configured at the capture SAP specifies the RADIUS authentication parameters, as shown in the following example:

```
configure service
        vpls 10 customer 1 create
            sap 1/1/4:1214.* capture-sap create
                description "Dynamic Services Data Trigger capture-sap"
                dynamic-services
                    dynamic-services-policy "dyn-svc-1"
                    no shutdown
                exit
                no shutdown
            exit
            no shutdown
        exit
        dynamic-services
            dynamic-services-policy "dyn-svc-1" create
                ---snip---
                authentication
                    password "RwXx4x0jao776C3CGlDBKVaNOd//ySXw" hash2
                    server-policy "aaa-server-policy-1"
                exit
                ---snip---
            exit
```

```
        exit
```

Local authentication and RADIUS authentication are mutually exclusive and cannot be configured simultaneously in a **config**>**service**>**dynsvc**>**plcy**>**authentication** context.

The **server-policy** CLI command references the **config**>**aaa**>**radius-server-policy** *policy-name* to be used for authentication.

The **password** CLI command specifies the password that is used in all RADIUS Access-Request messages.

Table 55: RADIUS access-request message attributes specifies the attributes that are included in the RADIUS Access-Request message for dynamic services data triggers.

Table 55: RADIUS access-request message attributes

RADIUS attribute	Description
[1] User-Name	The username format for dynamic services data trigger authentication is fixed to *nas-port-id* (SAP).
[2] Password	The password as configured in the authentication section of the **dynamic-services-policy**.
[4] NAS-IP-Address	The outband management interface or system interface IPv4 address. Only included if the RADIUS server is reachable via an IPv4 address.
[95] NAS-IPv6-Address	The outband management interface or system interface IPv6 address. Only included if the RADIUS server is reachable via an IPv6 address.
[44] Acct-Session-Id	A unique accounting session ID (number format) per dynamic service data trigger. Included as [50] Acct-Multi-Session-Id in radius accounting for all dynamic services that are associated with this data trigger.
[87] NAS-Port-Id	The dynamic service data trigger **sap-id**
[32] NAS-Identifier	The system name of the router
[26-6527-27] Alc-Client-Hardware-Addr	The MAC address of the data trigger frame that resulted in the authentication. Fixed format (xx:xx:xx:xx:xx:xx)
[8] Framed-IP-Address	The IPv4 source address of the IPv4 data trigger frame that resulted in the authentication. Not included if the data trigger frame is not an IPv4 packet.
[26-6527-99] Alc-Ipv6-Address	The IPv6 source address of the IPv6 data trigger frame that resulted in the authentication. Not included if the data trigger frame is not an IPv6 packet.

The attributes that must be returned in the Access-Accept message are the same as for RADIUS-triggered Dynamic Data Services associated with an IPoE or PPPoE session as a control channel.

16.3.4 Local authentication

Local authentication is available for data-triggered dynamic services deployments where RADIUS is used for accounting and dynamic changes (CoA) but cannot provide the actual service provisioning parameters.

When a valid Ethernet frame is received on a dynamic services data trigger capture SAP, it is sent to the control plane for authentication. The dynamic services policy configured at the capture SAP specifies the local authentication parameters, as shown in the following example:

```
configure service
        vpls 10 customer 1 create
            sap 1/1/4:1214.* capture-sap create
                description "Dynamic Services Data Trigger capture-sap"
                dynamic-services
                    dynamic-services-policy "dyn-svc-2"
                    no shutdown
                exit
                no shutdown
            exit
            no shutdown
        exit
        dynamic-services
            dynamic-services-policy "dyn-svc-2" create
                ---snip---
                authentication
                    local-auth-db "dynsvc-db-1"
                exit
                ---snip---
            exit
        exit
```

Local authentication and RADIUS authentication are mutually exclusive and cannot be configured simultaneously in a **config**>**service**>**dynsvc**>**plcy**>**authentication** context.

The **local-auth-db** CLI command references the local authentication database to be used for authentication, as shown in the following example:

```
configure service
        dynamic-services
            local-auth-db "dynsvc-db-1" create
                user-name "1/1/4:1214.101" create
                    description "dynsvc: epipe"
                    index 1 create
                        dynamic-services-policy "dyn-svc-2"
                        sap-id "1/1/4:1214.101"
                        script-parameters-1 "epipe_1={'t':('dynsvc-epipe-1',None,None,10,11)}"
                        accounting 1 create
                            stats-type volume-time
                            update-interval min 30
                        exit
                    exit
                    no shutdown
                exit
                no shutdown
            exit
        exit
```

A username is used as a key for a lookup in the local authentication database. The username format for dynamic service data triggers is fixed to the SAP ID of the data trigger. For each username entry (data

trigger **sap-id**), multiple dynamic service SAPs can be specified (indexes). The index enables multiple dynamic service SAPs to be associated with a single dynamic service data trigger.

The following data can be specified for each index (dynamic service SAP) in a **user-name** entry:

- dynamic service **sap-id** (mandatory)

  The dynamic service SAP ID that is created. The SAP ID of one of the indexes must match the dynamic service data trigger **sap-id**.

- dynamic-services-policy **dynsrv-policy-name** (optional)

  Specifies the policy to use for setting up the dynamic service. If not specified, the policy provisioned at the dynamic service data trigger **capture-sap** is used.

- **script-parameters** (mandatory)

  Script parameters are used as input to the dynamic data service Python script. They are specified as four strings of up to 250 characters each. The concatenation of all four script parameter strings are passed to the Python script and must be formatted as **function-key** *dictionary*. The **function-key** specifies which Python functions is called, and *dictionary* contains the actual parameters in a Python dictionary structure format. The format should match the format of the [26-6527-165] Alc-Dyn-Serv-Script-Params attribute when RADIUS authentication is used.

- accounting overrides (optional)

  For each of the two RADIUS accounting destinations, the **stats-type** and **update-interval** can be specified. These parameters override the configured value in the dynamic services policy:

  – **stats-type** specifies if dynamic service RADIUS accounting should be enabled or disabled. RADIUS accounting is enabled by specifying the statistics type: volume and time or time only. RADIUS accounting is disabled when **no stats-type** is specified.

  – **update-interval** specifies the time between each dynamic data service accounting interim update. The generation of interim accounting updates is disabled when **no update-interval** is specified.

A local authentication database can only be used to authenticate a dynamic service data trigger and provide parameters to set up associated dynamic services. The script action cannot be specified and is always set to "setup".

The setup timeout for Access=Accept (CLI command: **configure service dynamic-services timers setup-timeout access-accept** *timeout*) also applies for local authenticated dynamic services.

16.3.5 Data-triggered dynamic service provisioning

After authentication, the mechanism to set up, modify, and tear down a data triggered dynamic service is the same as for RADIUS-triggered Dynamic Data Services associated with an IPoE or PPPoE session as a control channel.

The auto-provisioning of a data-triggered dynamic service is initiated by the RADIUS messages or local authentication as listed in Table 56: Dynamic service script actions.

Notes:

- Using the Nas-Port-Id as a key in a CoA or Disconnect-Message targets the corresponding dynamic services SAP; this also occurs when the Nas-Port-Id corresponds with the SAP ID of a data trigger.
- Using the Acct-Session-Id as a key in a CoA or Disconnect-Message targets:
  – the corresponding dynamic services SAP if the Acct-Session-Id belongs to a dynamic services SAP that is not a dynamic services data trigger SAP

  - the data trigger SAP if the Acct-Session-Id belongs to the dynamic services data trigger SAP
- In the event of a tear down, if the dynamic services SAP ID is a dynamic service data trigger SAP ID, all dynamic services associated with that dynamic services data trigger are also removed.

Table 56: Dynamic service script actions

Action	Dynamic service script action	Comments
Rx Access-Accept or local authentication (dynamic services data trigger authentication)	Setup	Up to 32 dynamic data services SAPs in a single message The dynamic services SAP that corresponds with the data trigger (also referred to as the **dynamic services data trigger sap-id**) must be part of this list. The Alc-Dyn-Serv-Script-Action VSA is optional for RADIUS authentication.
	Modify / Teardown	Not supported
Rx CoA (Nas-Port-Id or Acct-Session-Id of a dynamic service SAP different from the data trigger)	Setup	Not supported
	Modify	Only a single dynamic data service per message Mandatory VSAs: Alc-Dyn-Serv-Script-Action Alc-Dyn-Serv-Script-Params
	Teardown	Tear down the dynamic service of the dynamic services SAP identified by the Acct-Session-Id or Nas-Port-Id. Alc-Dyn-Serv-Script-Action VSA is mandatory
Rx CoA (Nas-Port-Id of a data trigger)	Setup	Not supported. Nas-Port-Id always targets the dynamic services SAP and not the data trigger.
	Modify	Only a single dynamic data service per message Mandatory VSAs: Alc-Dyn-Serv-Script-Action Alc-Dyn-Serv-Script-Params
	Teardown	Tear down the dynamic service of the dynamic services SAP identified by the Nas-Port-Id. Because this is the data trigger SAP that is deleted, all dynamic services SAPs associated with the data trigger are also deleted. Alc-Dyn-Serv-Script-Action VSA is mandatory
Rx CoA	Setup	Only a single dynamic service SAP per message When successful, the dynamic services SAP is associated with the data trigger identified by the specified Acct-Session-Id.

Action	Dynamic service script action	Comments
(Acct-Session-Id of a data trigger)		Mandatory VSAs: Alc-Dyn-Serv-Script-Action Alc-Dyn-Serv-SAP-Id Alc-Dyn-Serv-Script-Params Alc-Dyn-Serv-Policy (if no "default" policy configured)
	Modify	Only a single dynamic service per message Modify the dynamic service of the dynamic services SAP identified by the Alc-Dyn-Serv-SAP-Id. The dynamic services SAP must be associated with the data trigger identified with the specified Acct-Session-Id. Mandatory VSAs: Alc-Dyn-Serv-Script-Action Alc-Dyn-Serv-SAP-Id Alc-Dyn-Serv-Script-Params
	Teardown	Tear down the dynamic service of the dynamic services SAP identified by the Alc-Dyn-Serv-SAP-Id. The dynamic services SAP must be associated with the data trigger identified with the specified Acct-Session-Id. If the dynamic services SAP identified by the Alc-Dyn-Serv-SAP-Id is a data trigger sap, then teardown the dynamic services of all dynamic services saps associated with that data trigger Mandatory VSAs: Alc-Dyn-Serv-Script-Action Alc-Dyn-Serv-SAP-Id
Rx Disconnect Message (Nas-Port-Id or Acct-Session-Id of a dynamic service SAP different from a data trigger)	N/A	Tear down the dynamic service of the dynamic services SAP identified by the Acct-Session-Id or Nas-Port-Id
Rx Disconnect Message (Nas-Port-Id or Acct-Session-Id of a data trigger)	N/A	Tear down the dynamic services of all dynamic services SAPs associated with the data trigger identified by the Acct-Session-Id or Nas-Port-Id

A data-triggered dynamic service must be explicitly removed by one of the following:

- with a RADIUS Disconnect message containing the Acct-Session-Id or NAS-Port-Id as key
- with a RADIUS CoA message containing the Acct-Session-Id or NAS-Port-Id as key and Alc-Dyn-Serv-Script-Action VSA with value 3 (teardown)
- with a CLI **clear** command: **clear**>**service**>**dynamic-services**>**data-trigger sap** *sap-id*

  All dynamic service SAPs associated with the dynamic services data trigger is removed.
- with a CLI tools command: **tools**>**perform**>**service**>**dynamic-services**> **evaluate-script sap** *sap-id* **control-session** *acct-session-id action teardown*

  The control session accounting session ID corresponds to the dynamic services data trigger accounting session ID.

The removal of a dynamic service SAP that is a data trigger SAP results in the removal (teardown) of all dynamic service SAPs associated with that dynamic services data trigger.

To prevent a data-triggered dynamic service from being immediately set up again after it was removed (because traffic is still being received), the following procedures can be used:

- Authentication failure
  - Update the configuration of the RADIUS server or local authentication such that the authentication for the dynamic service data trigger fails
  - Tear down the dynamic service
  - The dynamic service is not set up again because the data trigger authentication fails, resulting in a host-lockout when provisioned.
- VID filter on the **capture-sap**
  - Add the data trigger encapsulation to a VID filter (ingress MAC filter of type **vid**) that is applied on the data trigger **capture-sap**
  - Tear down the dynamic service
  - The dynamic service is not set up again because the data trigger is now dropped by the VID filter applied on the **capture-sap**

16.3.6 Control plane protection

As a dynamic services data trigger **capture-sap** potentially forward all valid Ethernet frames for authentication to the control plane, control plane protection mechanisms are required to prevent overload conditions.

1. **capture-sap** data trigger packet throttling (frames dropped at IOM)

   The number of data trigger packets sent to the control plane via the ingress forwarding complex is rate-limited based on a hash using the sap ID, outer tag, and inner tag as the key. The per-hash result is that a maximum of 1 frame is forwarded to the control plane.

   This throttling mechanism is always enabled and has no configuration options. It guarantees fairness between different encapsulation while limiting the frame rate sent to the control plane.
2. Blocking VLANs from authenticating (frames dropped at the IOM)

   This can be achieved by applying ingress MAC filters of type VID with the **capture-sap** command. In the example below, frames with encap 1/1/4:1214.20 is dropped by the VID filter.

```
configure
```

```
service
    vpls <service-id> customer <customer-id>
        sap 1/1/4:1214.* capture-sap
            dynamic-services
                dynamic-services-policy <dynsvc-policy-name>
                no shutdown
            exit
            ingress
                filter mac 10
            exit
        exit
    exit
exit
filter
    mac-filter 10 create
        default-action forward
        type vid
        entry 10 create
            match frame-type ethernet_II
                outer-tag 20 4095
            exit
            action
                drop
            exit
        exit
    exit
exit
```

3. Dynamic service data trigger rate limiting in the control plane (frames dropped at the CPM)

   An overall rate limit of dynamic service data triggers limits the number of frames that come from the different IOMs to an acceptable rate for the control plane to handle.

   Control plane rate-limiting for dynamic service data triggers is always enabled and has no configuration options.

4. Using a lockout mechanism to protect the RADIUS infrastructure from overload because of permanent failing authentications of Python script errors.

   The existing Enhanced Subscriber Management (ESM) host lockout mechanism can be enabled on a dynamic services data trigger **capture-sap**.

   The dynamic services data trigger **sap-id** is used as a key for the **host-lockout** context. A **host-lockout** context is created whenever a data trigger is deleted:

   - Authentication failures: RADIUS Access Reject, RADIUS Access-Accept with a wrong or missing data trigger SAP ID, timeout, local authentication lookup failure, local authentication returns a wrong or **no data-trigger sap-id**.
   - No data trigger SAP created within the configured dynamic services **access-accept setup timeout** (30 seconds by default):

     configure service dynamic-services timers setup-timeout access-accept <*timeout*>
   - A dynamic services Python script failure
   - A **clear** command on a data trigger:

     clear service dynamic-services data-trigger sap <*sap-id*>
   - A tear down of a data trigger initiated via a CoA or Disconnect Message

The last two bullets are tear down operations that occur infrequently and should not result in an actual lockout; only the host lockout context is created and should disappear again when the **lockout-reset-time** expires. This is configured in the **host-lockout-policy**:

```
configure subscriber-mgmt host-lockout-policy <policy-name> lockout-reset-time
<seconds>   (default 60 seconds).
```

To enable host lockout for dynamic services data trigger, configure a **host-lockout-policy** at the dynamic services data trigger **capture-sap**:

```
configure service
        vpls <service-id> customer <customer-id>
            sap <sap-id> capture-sap
                dynamic-services
                    dynamic-services-policy <dynsvc-policy-name>
                    no shutdown
                exit
                host-lockout-policy <policy-name>
            exit
        exit
```

The **host-lockout-policy** is configured in the **config**>**subscr-mgm**t CLI context. For data-triggered dynamic services, the **host-key** in the policy must be set to **all** which is achieved with the **no host-key** CLI command (default). Configuring the host key to **mac** in a host lockout policy associated with a dynamic service data trigger **capture-sap** is a configuration error and results in a faulty host lockout behavior.

Use the following show command to display active dynamic services data trigger capture SAP lockouts (the capture SAP must be used as the **sap-id**):

```
# show subscriber-mgmt host-lockout-policy "host-lockout-1" all sap 1/1/4:1214.*
===============================================================================
Host Lockout Policy "host-lockout-1"
===============================================================================
Description                     Host lockout policy
Last Mgmt Change                01/25/2016 16:37:47
Lockout time min                10
Lockout time max                3600
Lockout reset time              60
Max lockout hosts               100
Host key                        all
-------------------------------------------------------------------------------
Active Lockouts for SAP: 1/1/4:1214.*
===============================================================================
circuit-id/                     elapsed  current  elapsed  next     nr
  mac/                          reset    lock     lock     lock     of
remote-id                       time (s) time (s) time (s) time (s) lockouts
-------------------------------------------------------------------------------
:1214.101                       0        10       1        20       1
-------------------------------------------------------------------------------
Nr of active lockouts           1
Nr of lockouts in grace period  0
Nr of total lockouts:           1
-------------------------------------------------------------------------------
Totals for Host Lockout Policy "host-lockout-1"
-------------------------------------------------------------------------------
Nr of active lockouts           1
Nr of lockouts in grace period  0
Nr of total lockouts:           1
```

```
===============================================================================
```

Use the following command to clear active dynamic services data trigger capture SAP lockouts (the capture SAP must be used as the **sap-id**):

```
# clear subscriber-mgmt host-lockout-policy [policy <host-lockout-policy-
name>] <lockout-state>
# clear subscriber-mgmt host-lockout-policy sap <sap-id> [lockout-state]
```

5. Use the **cpu-protection** and **dist-cpu-protection** commands on the platforms that support it.

   When using the **cpu-protection** command, a **cpu-protection** *policy-id* is needed to apply on the capture SAP.

```
configure
    system
        security
            cpu-protection
                policy 200 create
                    overall-rate 100
                    out-profile-rate 50
                exit
            exit
        exit
    exit
    service
        vpls <service-id> customer <customer-id>
            sap 1/1/1:*.* capture-sap create
                cpu-protection 200
                dynamic-services
                    dynamic-services-policy "dynServPolicy"
                    no shutdown
                exit
                no shutdown
            exit
        exit
    exit
```

   When using the **dist-cpu-protection** command, a **dist-cpu-protection** *policy-id* is needed to apply on the capture SAP.

```
configure
    system
        security
            dist-cpu-protection
                policy "distCpuProt" create
                    static-policer "dcpuStatPol" create
                        rate packets 50 within 1
                        exceed-action discard
                        detection-time 5
                    exit
                    protocol all-unspecified create
                        enforcement static "dcpuStatPol"
                    exit
                exit
            exit
        exit
    exit
    service
        vpls <service-id> customer <customer-id>
            sap 1/1/1:*.* capture-sap create
```

```
                dynamic-services
                    dynamic-services-policy "dynServPolicy"
                    no shutdown
                exit
                dist-cpu-protection "distCpuProt"
                no shutdown
            exit
        exit
    exit
```

16.3.7 Debugging

To enable dynamic services data trigger debugging, use the following debug commands:

```
debug
    dynamic-services
        data-triggers
            # multiple capture-sap commands allowed
            capture-sap <sap-id> [encap-val <qtag[.qtag]>] [mode <mode>]
            capture-sap <sap-id> [encap-val <qtag[.qtag]>] [mode <mode>]
        exit
    exit
```

During debugging, the system logs data trigger events such as:

- data trigger received
- data trigger authentication events
- data trigger SAP created
- dynamic service SAP created
- dropped data trigger with drop reason such as data trigger already exists or lockout active

The **encap-val** command limits the debug output to data trigger events for specific encapsulation values.

The mode specifies which data trigger events are logged: all events or dropped data trigger events only.

16.4 Dynamic data services Python API

Table 57: Dynamic data services functions in the alc.dyn module describes the functions available in the SR OS alc.dyn Python module to create dynamic data services.

Table 57: Dynamic data services functions in the alc.dyn module

alc.dyn functions	Description
dyn.action (dictionary)	Executes the setup, modify or teardown function that is found with a lookup in the specified dictionary using the function-key present in the script parameters. The script-parameters can be obtained from RADIUS in the Alc-Dyn-Serv-Script-Params VSA or can be configured in the local authentication database. The action (setup, modify or teardown) is determined from the Alc-Dyn-Script-Action VSA. Returns: None (no value)

alc.dyn functions	Description
	dictionary - Python dictionary with format: dictionary = {function-key-1 : (setup-function-1, modify-function-1|None, revert-function-1|None, teardown-function-1), …, function-key-n : (setup-function-n, modify-function-n|None, revert-function-n|None, teardown-function-n) } where: • setup-function-n: name of the function invoked when Alc-Dyn-Serv-Script-Action=setup • modify-function-n: name of the function invoked when Alc-Dyn-Serv-Script-Action=modify. The modify function is optional. When not specified, the keyword "None" should be used in the dictionary. • revert-function-n: name of the function invoked when modify script execution failed. This function must undo all changes from the modify function. The revert function is optional. When not specified, the keyword "None" should be used in the dictionary. • teardown-function-n: name of the function invoked when Alc-Dyn-Serv-Script-Action=teardown
dyn.add_cli (string)	Executes the specified CLI commands to create the dynamic data service Returns: None (no value) *string* - CLI commands. The string can span multiple lines in the script when enclosed in three double quotes ("""").
dyn.get_circuit-id()	Returns a string containing the control channel circuit-id (DHCP relay agent option 82 or PPP tags)
dyn.get_remote-id()	Returns a string containing the control channel remote-id (DHCP relay agent option 82 or PPP tags)
dyn.get_sap()	Returns a string containing the dynamic data service *sap-id*: the value of the Alc-DynServ-SAP-ID VSA. A wildcard hash (#) in the VSA value is replaced with the corresponding control channel or data trigger port or vlan-id field.
dyn.reference (function-key, reference-id, dictionary)	Create a dynamic reference to another function in the script. Typically used to create n:1 relations between dynamic data services, such as multiple SAPs in a VPLS service or multiple services for the same customer. Dynamic referencing is only allowed in setup functions. Corresponding teardown functions automatically dereference. Returns a dictionary provided by the setup script matching the function-key: *function-key* is the key in the action dictionary (see dyn.action) to find the corresponding setup or teardown functions. *reference-id* is the unique reference ID string that identifies the dynamic reference (for example, all SAPs from a VPLS service would have the same reference id).

alc.dyn functions	Description
	dictionary is a Python dictionary containing parameters for use in the setup or teardown function, such as to generate CLI output.
dyn.select_free_id (type)	Returns a string representing a free identifier of the specified type. *type* is identifier type that is returned: • "*service-id*" obtains a free service ID in the configured dynamic-services service-range • "*ip-filter-id*" generates a new ip-filter identifier • "*ipv6-filter-id*" generates a new ipv6-filter identifier • "*mac-filter-id*" generates a new mac-filter identifier The corresponding filters are shown in the system as "_tmnx_dyn_<number>"

17 Diameter and diameter applications

17.1 Restrictions

Diameter-Based Restrictions:

- Accounting (RFC 6733, *Diameter Base Protocol*) via Diameter is not supported in this release.
- Accounting-Request (ACR), Accounting-Answer (ACA), Session-Termination-Requests (STR) and Session-Termination-Answer (STA) messages are not supported.
- SCTP and IPsec as transport protocols are not supported. TCP is supported.

Gx-Based Restrictions:

- Static hosts, MLPPP sessions, and LAC (L2TP) hosts are not supported in Gx.
- Diameter Gx policy management and usage monitoring can be enabled for Layer 2–aware NAT hosts with following restrictions: there is no NAT information reported in Gx messages, there are no NAT-triggered Gx messages and there is no support for NAT information received in Gx messages from PCRF.
- Bridged homes and AA subscribers

  Because there is no concept of a subscriber host in AA, the last AA policy submitted by Gx for any ESM subscriber host within the home is applied to the AA subscriber as a whole and overwrite any previously active AA policy.
- The <SAP, MAC> combination must be unique for each host (single stack or dual-stack) if ipoe-session or ppp-sessions are not used or enabled.
- When an SLA profile instance contains multiple subscriber hosts, it is mandatory that all hosts have the same PCC rules applied.
- The Charging-Rule-Name within the Charging-Rule-Definition cannot contain a double colon (::) set of characters in the name string. The use of double colon in the name string itself is reserved for future use.
- Reporting about successful rule activation on the node (3GPP 29.212, §4.5.2) is not supported. The rule report is sent only if the rule instantiation fails.
- Time-based Usage-Monitoring is not supported.
- Gx persistency is not supported. However, upon reboot with ESM persistency enabled, the node re-initiates Gx sessions (new CCR-I is generated for each Gx enabled host).
- Gy and Usage-Monitoring cannot be enabled for the same host and the same category-map at the same time. Gy is pre-configured at the time of the host instantiation. If a Usage-Monitoring request is received while Gy is enabled, the node ignores the Usage-Monitoring request.
- Each ESM host can have up to 16 Usage-Monitoring entities enabled simultaneously. Static configured categories and the internal category required for session level Usage-Monitoring are counted against this limit. The instantiation of the internal category for session level Usage-Monitoring is controlled with the **gx-session-level-usage** CLI command in the category map. If 16 categories are configured,

then Usage-Monitoring cannot be enabled per session (host) because this would exceed the limit of 16 Usage-Monitoring entities per host.

17.2 Terminology

The term Gx interface (or simply Gx) refers to the implementation of the Gx reference point on the node. Gx reference points are defined in the 3GPP 29.212 document.

The ESM subscriber is a host or a collection of hosts instantiated on the Broadband Network Gateway. The ESM subscriber represents a household or a business entity for which various services with committed Service Level Agreements (SLA) can be delivered.

AA Subscriber is a representation of ESM subscriber in MS-ISA for the purpose of managing its traffic based on applications (Layer 7 awareness). An AA subscriber has no concepts of ESM hosts.

BNG refers to the network element on which a Gx interface is implemented and policy rules are enforced (PCEF).

Flow

A flow in Gx context represents traffic matching criteria (traffic classification or traffic identification) based on any combination of the following fields:

- source IP address
- destination IP address
- source port or port ranges
- destination port or port ranges
- protocol field
- DSCP bits

A Gx flow is defined in the Flow-Information AVP:

```
Flow-Information ::= < AVP Header: 1058 >  3GPP 29.212 §5.3.53
          [ Flow-Description ]       3GPP 29.214 §5.3.8
          [ ToS-Traffic-Class ]      3GPP 29.212 §5.3.15
          [ Flow-Direction ]              3GPP 29.212 §5.3.65
          *[ AVP ]
```

Gx flows are similar to dynamically created filter or ip-criteria (QoS) entries and are inserted within the entry range configured for the base filter/qos-policy.

IP criterion

These fields are used in IPv4/v6 packet header used as a match criterion. The supported fields are DSCP bits and 5 tuple. This is part of traffic classification (or traffic identification) within the PCC rule or within the static qos-policy/filter entry.

Gx policy rule

There are three types of Gx policy rules supported:

- **Gx based overrides**

  subscriber-related overrides (sub/sla/aa-profile, subscriber-id, QoS, filter, category-map, and so on).

- **NAS filter entry inserts via Gx**

Basic permit/deny entries, similar to ACL filter entries

- PCC rules or IP-criterion based rules which are fully constructed Policy and Charging Control (PCC) rules with multiple QoS/filter actions per rule and its own traffic classification

PCC rule represents a single or a set of IP based classifiers (DSCP bits or 5 tuple) associated with a single or multiple actions.

For example:

Each PCC rule can be removed via Gx from the node by referencing its name (Charging-Rule-Name AVP).

The PCC rule can contain a combination of QoS and IPv4/v6 filter actions as they pertain to the node.

PCC Rule Classifier

A flow-based (5 tuple) or a DSCP classifier defined in the Flow-Information AVP within the PCC Rule. There can be a single classifier or multiple classifiers (Flow-Information AVPs) within a single PCC rule. A PCC classifier (Flow-Information AVP) corresponds to an entry (match criteria) within the filter/ip-criteria definition.

CAM entry

A single entry in the CAM that counts toward the CAM scaling limit. For example a match condition within ip-criteria in a filter or qos-policy can evaluate into a single CAM entry or into multiple entries (where port-ranges are configured in the classifier, or where matching against UDP and TCP protocols are enabled simultaneously).

Subscriber Host Policy

A collection of PCC rules (classifiers and actions), Gx overrides, NAS filter inserts and statically configured rules (CLI defined QoS or filter) that are together applied to the subscriber host.

17.3 Diameter base

An updated Diameter base has been implemented to improve future enhancements. The advantages that the Diameter base implementation brings about are as follows:

- Future proof extensibility and better compliance with RFC 6733, *Diameter Base Protocol*
- CLI alignment with current AAA framework
- SMP (multicore) capabilities
- Performance, the ability to process high volume of messages
- Diameter routing

The legacy Diameter base continues to be supported along with the new implementation, although the two cannot be used simultaneously for an application (NASREQ/Gx/Gy) within a given Diameter node instance within the SR OS. Operators are advised to deploy the new Diameter base as there are no plans going forward to add new functionality in the legacy Diameter implementation. Furthermore, the legacy Diameter implementation is scheduled to be discontinued in a future SR OS release, in accordance with the NOKIA software discontinuation policy.

This section describes the operation of the updated Diameter base. The principal difference between the two implementations, from the operator’s point of view, is not only in the functional behavior but also in the CLI. For this reason, the CLI descriptions are provided for both implementations. Some of the CLI commands remain the same for both implementations, and some of the commands are specific to their

respective implementation (new or legacy). Each command that is specific to an implementation contains an explicit statement about the implementation to which it pertains.

17.3.1 Diameter base protocol

The Diameter base protocol is used to provide reliable and secure communication for Diameter applications between and across the Diameter peers (diameter clients, agents, or servers). Its routing capabilities to transport traffic across Diameter nodes rely on Diameter identities that are composed of host and realm names. Diameter applications supported in SR OS are NASREQ, Gx, and Gy.

In the SR OS, Diameter base protocols run over TCP. It starts by establishing a TCP connection with peers, followed by capability exchanges. A data exchange occurs in the form of Attribute Value Pairs (AVPs). Some of these AVPs are used by the base protocol itself, while others are used by Diameter applications that are layered on top of it.

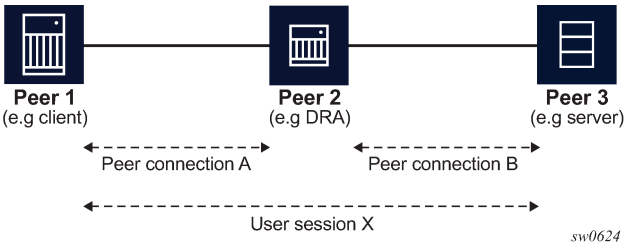
17.3.2 Diameter peers and the role of a diameter node in SR OS

Diameter peers exchange messages using a transport level connection (TCP in SR OS). Each peer is identified by a unique Fully Qualified Domain Name (FQDN). On the IP level, Diameter peers are always directly connected.

Applications (NASREQ, Gx, and Gy) relaying on Diameter base protocol use a concept of a session to communicate between the two end nodes that may be adjacent to each other or they may be multiple Diameter hops away.

This concept is shown in Figure 194: Diameter peers.

Figure 194: Diameter peers



SR OS supports Diameter client functionality where it performs access control for a device running on the edge of the network. A Diameter client in an SR node can initiate and accept connection requests to and from its peers. A connection request is only accepted from another SR node in Diameter multichassis redundancy setup. This peering connection between the two redundant Diameter SR nodes is referred to as inter-peering connection and its use is described in Diameter multichassis redundancy.

A Diameter server hosting multiple applications must support all applications over a single peering connection. In other words, such a server must not open a separate peering connection per application because peers in the forwarding table are not application aware.

17.3.3 Capability Exchange

Upon TCP connectivity establishment, Diameter peers in an SR node exchange Capability Exchange messages. These messages contain information about the peer such as the peer's identity, supported applications, protocol version, and security mechanisms. Capability Exchange messages do not traverse Diameter nodes but instead they are exchanged only between the peers (immediate Diameter next hops).

In this phase, an SR node performs peer authorization where the peer identity received from the peer through a Capability Exchange message (origin-host in CEA) is compared to the locally configured peer identity. This peer identity (the peer name) comparison in SR OS is case insensitive. If the compared peer names do not match, the TCP session is closed.

For example:

- The configured peer name for SR OS is: peer-1.acme. This identifies the peer that the SR is allowed to establish a connection with.
- The origin strings received from this peer in CEA are:
  - Origin-host: peer-1
  - Origin-realm: realm-x

In this scenario, SR rejects the connection from this peer because the configured peer name in the SR (peer-1.acme) does not match the origin-host received from the same peer in CEA (peer-1).

An SR node always advertises all three supported applications (NASREQ, Gx, and Gy) in Capability Exchange, regardless of whether these applications are configured in an SR node.

If no common applications are negotiated, the peer (the receiver of a Capability Exchange Request - CER) must return a Capability Exchange Answer (CEA) with the Result-Code AVP set to DIAMETER_NO_COMMON_APPLICATION and should disconnect the transport layer connection.

If the application ID of 0xffffffff (Diameter Relay Agent - DRA) is received by an SR node, then this is interpreted as having a common application with the peer.

Diameter names and realms learned from peers through Capability Exchange are used in SR nodes to populate local peer and realm routing tables that are then used for forwarding and routing of the Diameter messages.

17.3.4 Connection termination

The SR OS may initiate termination of the peering connection because of the following reasons:

- **peer shutdown**

  This refers to graceful termination by initiating a Disconnect Peer Request (DPR) message from either side of the connection.

- **watchdog timeouts**

  The peer is unresponsive and consequently the TCP connection is forcefully closed.

- **message authorization failure during the Capability Exchange phase with the peer**

  There is no support for common application or hostnames mismatch.

- **failure to verify the following AVPs in incoming Diameter Base messages**
  - CEA: Origin-Host
  - DWR: Origin-Host, Origin-Realm, Origin-State-Id

– DWA: Origin-Host, Origin-Realm, Origin-State-Id
– DPR: Origin-Host, Origin-Realm
– DPA: Origin-Host, Origin-Realm

These values of these AVPs are checked against:

– Origin-Host: peer name configuration
– Origin-Realm: learned from CEA
– Origin-State-Id: learned the first time seen in either CEA, DWR, DWA

17.3.5 Diameter hosts and realms

Diameter realms are administrative domains that have a relationship with the user account. For example, a Diameter client can be common to multiple departments with different domain realm names, for example xyz.com and wvu.com. In some cases, realms can represent different geographical regions. Regardless, the realm names are maintained in Diameter nodes (as routes) and are used for routing of diameter messages. In general, realms can be thought of as a string in NAI (Network Access Identifier) that immediately follows the first "." character in the string (for example, jdoe.example.com).

Host names identify a host within a realm. Together, the host and realm names form a Diameter Identity (DI). Every Diameter message must contain origin-host and origin realm AVP.

17.3.5.1 Forwarding and routing of application messages in Diameter

Forwarding and routing of the application level (NASREQ/Gx/Gy) messages in Diameter depends on the message type (requests or answers). Requests messages are forwarded or routed based on destination hosts and realms. Forwarding and routing of request messages is based on the Diameter hosts and realms and is dependent on the two tables maintained in each node between the source and destination:

- A peer table contains a list of active peers (adjacent Diameter nodes). The peers in the peer table are populated though configuration and are not learned dynamically. The peer table contains only the peer name (no realms).
- A realm-based routing table contains mapping of realms and applications to peers. A realm-based routing table is in the SR OS created per Diameter node. Realms in the routing table are learned dynamically from peers from the origin-realm AVP in Capability Exchange Answer (CEA) messages and cannot be statically configured.

The selection of a peer to which a request message is sent can be a two step process. First, if the request message contains the destination-host AVP, the peer table is checked to find out if there is a matching peer available. If so (the destination is directly connected), the request message is forwarded to it, without further checking the routing table for the next hop (this is called forwarding phase). If the matching host is not present in the peer table (the destination is not directly connected), or the destination-host AVP is not present in the request message (in other words, CCR-I), then the diameter routing table is checked for the matching realm. This is referred to as routing phase. Each routable request message must contain the destination realm AVP and the destination realm is a mandatory configuration parameter in an SR node for a Diameter application. However, on the application level, a Gx and Gy session can also learn the destination realm from received application messages. In this case, the learned value overwrites the configured value.

If the Diameter realm table does not contain an entry for the destination realm and the application, the request message is forwarded to the default peer, if one is configured. There can be only one default peer per Diameter node. Unlike a realm routing entry in the realm table, the default peer is application (NASREQ, Gx, Gy) unaware, and it is used as a route of last resort, regardless whether the selected path leads to a server supporting this application.

If the default peer is not configured, messages destined for a realm that is not known in the realm routing table, are dropped.

Application level answer messages do not rely on the routing or forwarding table. They are forwarded in reverse direction of the matching requests in the transactional cache of each traversed Diameter node and the Hop-by-Hop AVP. The Hop-by-Hop AVP is set to a locally unique value by each Diameter node that forwards request messages. This AVP is then used to match request and answer messages in the reverse direction. This allows a Diameter response to follow the same route as the corresponding Diameter request.

17.3.5.2 Static Diameter realm routes

Static Diameter realm routes are used to reach remote realms, which are not directly connected to the local realm configured on the SR OS.

They are configured under the Diameter peer:

```
configure
   aaa
      diameter
           node <origin-host-string> [origin-realm <origin-realm-string>]
                peer index <index> [<destination-host-string>]
                    route index <index> [realm <name>] [application <name>] [create]
                        preference <1-100>
```

There are two preference parameters that can be configured under this CLI hierarchy:

- Peer preference is configured directly under the peer and applies only to routes learned using the capability exchange phase and not the static routes. This parameter determines the directly-connected preference of the learned routes relative to the static routes. Currently, there can be only one learned realm route per peer.
- Static route preference is configured for each static route. The static route with the numerically lowest preference is used.

The order of the evaluation for realm routes is:

- The route with the lowest preference is used. The preference of static routes is individually configured, whereas the preference of the learned routes is configured by a peer.
- In case of a preference tie, the route configured by the peer with the lowest index is chosen.

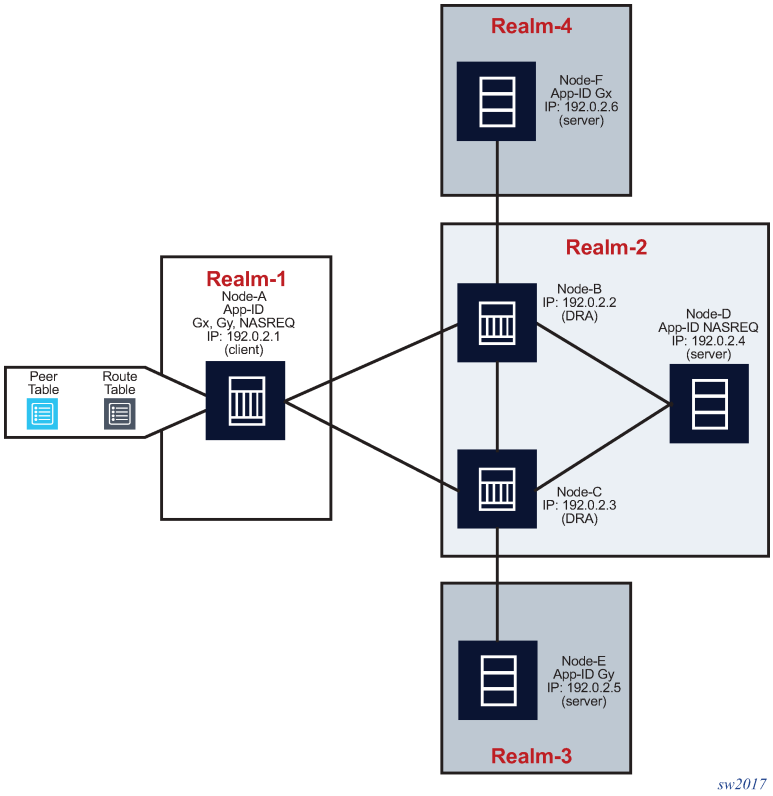
A static route is installed in the realm table only if the peer in which it has been configured is in an open state.

17.3.5.2.1 Example with static realm routes

In this example, Gx and Gy traffic, destined for the remote realms Realm-3 and Realm-4, is balanced over two DRAs. Gx traffic destined for Realm-4 flows through Node-B, while Gy traffic destined for Realm-3

flows through Node-C, according to the preference of the static routes. If the connection to one of the peers fails, the surviving connection takes over as shown in Figure 195: Reaching remote realms.

Figure 195: Reaching remote realms



NASREQ traffic, which is destined for the directly-connected Realm-2 flows through Node-B because it has a lower peer index, while the peer preference is the same as on peer Node-C as described in Table 58: Peer table and Table 59: Realm routing table . Only when the communication to this peer fails does Node-C be selected as the new peer.

```
>config>aaa# info
----------------------------------------------
        diameter
            node "sr" origin-realm "realm-1" create
                peer index 1 "node-b" create
                    address 192.0.2.2
                    preference 5
                    route index 1 realm realm-3 application gx
                          preference 10
                       route index 2 realm realm-4 application gy
                          preference 20
                       no shutdown
                exit
                peer index 2 "node-c" create
                    address 192.0.2.3
                    preference 5
                    route index 1 realm realm-3 application gx
                          preference 20
                    route index 2 realm realm-4 application gy
                          preference 10
```

```
                no shutdown
            exit
        exit
    exit
```

Table 58: Peer table

Host identity	IP address
Node-B	192.0.2.2
Node-C	192.0.2.3

Table 59: Realm routing table

Realm name	App ID	Peer	Entry	Preference
Realm-2	0xffffffff	Node-B	dynamic (learned)	5
Realm-2	0xffffffff	Node-C	dynamic (learned)	5
Realm-3	Gx	Node-B	static	10
Realm-3	Gx	Node-C	static	20
Realm-4	Gy	Node-C	static	20
Realm-4	Gy	Node-B	static	10

The NASREQ routing in this case can be configured in several ways:

- with the same preference for both peers (as in the example above), where the peer with the lowest index prevails
- with different peer preferences as shown in the example below

```
*A:right-b4>config>aaa# info
----------------------------------------------
        diameter
            node "sr" origin-realm "realm-1" create
                peer index 1 "node-b" create
                    address 192.0.2.2
                    preference 10
                    no shutdown
                exit
                peer index 2 "node-c" create
                    address 192.0.2.3
                    preference 5
                    no shutdown
                exit
            exit
        exit
```

In this case, the preferred NASREQ path is through Node-C.

- with static routing

```
*A:right-b4>config>aaa# info
----------------------------------------------
        diameter
            node "sr" origin-realm "realm-1" create
                peer index 1 "node-b" create
                    address 192.0.2.2
                    preference 10
                    no shutdown
                exit
                peer index 2 "node-c" create
                    address 192.0.2.3
                    preference 10
                    route index <index> realm realm-2 application nasreq
                        preference 5
                    no shutdown
                exit
            exit
        exit
```

In this case, the NASREQ path is through Node-C because the preference of the static route is the lowest.

17.3.5.2.2 Default peer

If the Diameter realm table does not contain an entry for the destination realm and the application, the request message is forwarded to the default peer, if one is configured. There can be only one default peer per Diameter node. Unlike a realm routing entry in the realm table, the default peer is application (NASREQ/Gx/Gy) unaware, and therefore it is used as a route of last resort, regardless of whether the selected path leads to a server supporting this application or not. If the default peer is not configured, messages destined for a realm that is not listed in the realm routing table are dropped.

17.3.5.3 Configuration of hostnames and realms in SR OS

Host names and realm names are essential in Diameter routing. They are conveyed in Diameter messages through the following AVPs:

- Origin-Host
- Origin-Realm
- Destination-Host
- Destination-Realm

Origin-host and origin-realm AVPs are used to identify nodes in Diameter networks. Every Diameter message must carry these two AVPs. In SR OS, the origin hostname must be statically configured through CLI. The configuration of the origin realm is optional. If the origin realm is not configured, the origin realm is extracted from the origin hostname. The first ".” in the configured origin hostname is used as a delimiter between the host portion and the realm portion. If there is no ".” in the configured origin hostname, then the origin hostname is used as both, the origin hostname and the origin realm name. Origin host and realm names do not support online changes (once they are configured, because they cannot be changed without first deleting the Diameter node hierarchy).

The origin-host and origin-realm names received from peers, as a result of successful Capability Exchange negotiation, are used to populate peer and realm tables in SR nodes.

While origin-host and origin realms are used for identification of Diameter nodes within realms, the destination-hosts and destination-realms are used in Diameter request messages for forwarding and routing purposes. They steer the message through the Diameter network toward the destination.

The destination hostname that is used by applications (NASREQ, Gx, Gy) cannot be configured in an SR node and is only learned from the incoming application level messages (the origin-host AVP in the incoming request or answer messages). within each Diameter session.

On the other hand, the destination realm that is used in application request messages is a mandatory configuration parameter in an SR node. However, its value can optionally be learned from the incoming application messages and any newly learned value overwrites the previous one. This learning is performed on a per-message basis. The destination realm value (configured or learned) is not used to populate the realm table but is only used on the application level to populate the destination-realm AVP in every application level request message.

Because the destination-host AVP is optional and the destination-realm is mandatory in application request messages, the destination-realm can be used without the destination-host in forwarding and routing decisions to guide the request message to its destination. The destination-realm used in initial request message (CCR-I) is a configured one, while the destination-host is not used in the same message because it has not been learned yet.

Destination-host or destination-realm AVPs are never present in:

- Diameter messages that are not to be proxied or relayed (such as CER)
- Diameter answer messages

17.3.5.3.1 New base configuration

Origins are creation time parameters in the Diameter node CLI configuration. Multiple unique diameter nodes can be instantiated within the same SR node.

```
config>aaa>diam#
    node <origin-host-string> [origin-realm <origin-realm-string>]
```

Applications (NASREQ, Gx, and Gy) are associated with the Diameter node (and the origins) through the diameter-application-policy:

```
config>subscr-mgmt>diam-appl-plcy#
diameter-node <origin-host-string> destination-realm <destination-realm-string>
```

The destination-realm referenced in the Diameter node above is used to populate the destination-realm AVP in the application level request messages. This destination realm can be optionally learned from the incoming application level messages and any newly learned value overwrites the previous one. This learning is performed on a per-message basis.

17.3.5.3.2 Legacy configuration

The centerpiece of legacy in the Diameter base configuration in an SR node is a **diameter-peer-policy** where all the Diameter base parameters related to communication with other external Diameter peers are defined (peering connections, DiameterIdentity, timers, and so on). This **diameter-peer-policy** is then associated with Diameter applications (NASREQ, Gx, Gy) which rely on it to interact outside the network.

In legacy Diameter base, the origins are configured under the **diameter-peer-policy**:

```
config>aaa>diam-peer-plcy#
     origin-host <origin-host-string>
     origin-realm <origin-realm-string>
```

while the destinations are configured one level below, under the peers:

```
config>aaa>diam>diameter-peer-plcy>peer#
     destination-host <destination-host-string>
     destination-realm <destination-realm-string>
```

If these parameters are not configured, the **diameter-peer-policy** is operationally down.

Applications (NASREQ, Gx, and Gy) are associated with the **diameter-peer-policy** through the diameter-application:

```
config>subscr-mgmt>diam-appl-plcy#
     diameter-peer-policy <name>
```

17.3.6 Dynamically learned parameters

The Diameter entities that are dynamically learned from a peer using Capability Exchange are:

- the received origin realm name (via origin-realm AVP)
- the App-ID

These entities are used to populate the realm table (<realm, app-id> to peer mapping) on which the routing decisions are based.

The origin-host AVP from CEA is used to cross check the peer name in the peer table. A mismatch between the configured and learned (origin-host AVP in CEA) peer names causes the TCP connection to close.

The destination hostname is initially learned on the application level (Gx and Gy) using CCA-I messages and the destination hostname can be re-learned through subsequent incoming application level messages. For example, the origin hostname in incoming application level messages become the destination hostname in outgoing application level messages.

The destination realm, although initially configured, can also be learned on the application level using incoming application level messages (Gx and Gy). Dynamic learning is enabled per message, and the learned value is used in subsequent request messages.

17.3.7 Diameter routing loop avoidance

Routing loop prevention in an SR node relies on the evaluation of a set of candidate next hop peers that are returned by each forwarding lookup (which is a realm routing lookup based on destination DI in the request message). The following peers are excluded from this set as next-hop candidates:

- Peer from which the request is received
- A peer from which a DIAMETER_UNABLE_TO_DELIVER (3002) or DIAMTER_TOO_BUSY (3004) answer was received for the request.

If an error message with the result-code AVP of DIAMETER_UNABLE_TO_DELIVER (3002) or DIAMTER_TOO_BUSY (3004) is received, the Diameter client should use an alternate route to deliver this message (§7, example relevant to Figure 7 in RFC 6733, *Diameter Base Protocol*).

- Any peer identified in the Record-Route AVP of received request message (see the Diameter Multi-Chassis Redundancy section).

The remaining peers are evaluated and the one with the numerically lowest preference value is selected as the next-hop peer.

17.3.8 Retransmissions and message timers

Application level Diameter request messages originated from an SR node can be retransmitted in response to either an error messages received from the Diameter peer or server, or because of a lack of acknowledgment (Diameter Answer message or a TCP ACK) received from a Diameter peer or server.

Message retransmissions can occur on multiple levels:

- TCP level where a TCP ACK is not received from a peer. Such retransmission cannot traverse Diameter nodes (for example between the originating node and the destination node that are separated by one or more intermediate Diameter nodes). They work only between Diameter peers.

  The lack of TCP ACK can be caused by some anomaly in transmission lines or congestion in the network or the peering node where the TCP packets are dropped.

- The Diameter Base level where they are triggered by the peer failover procedures that can be caused by the peer failure or the receipt of DIAMETER_UNABLE_TO_DELIVER (3002) and DIAMETER_TOO_BUSY (3004) error messages. The peer failover (and retransmission on the Diameter base level) is performed only for the messages that directly solicit 3002/3004 error response.
- On the Diameter application level (NASREQ, Gx, Gy) retransmissions are sent only if the server failover functionality is enabled:

```
*A:node1>config>subscr-mgmt>diam-appl-plcy# on-failure
    failover {enabled|disabled}
    handling {continue|retry-and-terminate|terminate}
```

  On the Diameter application level, the request message is retransmitted only one time. Retransmissions on the Diameter application level are triggered by:

  – Answer messages that have the E-bit set (protocol error). The only two exceptions are DIAMETER_UNABLE_TO_DELIVER (3002) and DIAMETER_TOO_BUSY (3004) error messages that trigger retransmissions on the Diameter Base level until all viable Diameter paths are exhausted, or the Tx timer expires (whichever event occurs first). Only after all viable paths are exhausted, or the Tx timer expires, the handling is passed to the Diameter application which has the option to retransmit the message one last time, with the destination-host AVP cleared.

  – Notification by the Diameter base that the path to the destination-realm is unavailable. For example, because of Tx-timer timeout, the answer to an original request message is not received within the time window determined by this configurable timer.

Answer messages are never retransmitted on the Diameter application level. However, Answer messages can be retransmitted on a TCP level.

Retransmission of non-routable Diameter request messages is driven by an internal timer which is set to a fixed value of 10 seconds. Non-routable Diameter messages are messages required for maintenance of

the peering connection. These messages never propagate beyond the peer. This internal timer is used in the following cases:

- Failure to initiate TCP peering connection
- **Incomplete Capability Exchange**

  If the response to the CER message (CEA) is not received within 10 seconds, the peer connection is closed and the connection-timer is restarted. The expiry of the connection-timer triggers a new connection request.

  Peer Disconnect Request

  If the response to the DPR message is not received (DPA) within 10 seconds, the peer connection is closed.

Diameter Watchdog messages are an exception and they are not retransmitted. Watchdog messages are used to detect the liveliness of a peer. A Diameter Watchdog message is sent in absence of any traffic on the peer level and only when the peer level watchdog-timer expires.

17.3.8.1 Clearing the destination-host AVP in the retransmitted messages

The destination-host AVP is cleared in all Diameter request messages that are retransmitted by the Diameter application (NASREQ, Gx, Gy). Messages retransmitted by the Diameter Base (peer failover related retransmissions) do not have the destination-host AVP cleared.

The Diameter application retransmits the message with the assumption that the path to the destination-host, or the destination-host itself is unavailable. Clearing of the destination-host AVP provides a new chance to for the Diameter Base to deliver the message to an alternate destination in the same realm. This alternate destination must be in a geo-redundant setup where sessions are synchronized between the geo-redundant nodes.

17.3.8.2 Retransmission bit (T-Bit)

The T-bit can be set in a retransmitted Diameter request message as a direct indication that the message may be a duplicate. For example, if the original request message is received by the server, but the answer is lost on its way back to the sender. If the original request message is not received by the server, then the message is not a duplicate from the server point of view. Detection of the duplicates on the application server side is important, especially in credit control applications. Duplicate requests received by the server should not be interpreted as new requests for credit, or new reports for usage, otherwise, double accounting or billing may occur.

On the other hand, the indirect indication of a duplicate Diameter request message on the server side relies on the comparison of the origin-host and end-to-end identifier fields that must always be the same in original and retransmitted messages. This requires the server to keep a history of received messages (or some metadata after it has sent the answer). Normally, after the answer is sent, reference to end-to-end identifier is lost and the duplicate detection is not possible. While some servers may implement such logic to rely on implicit identification of duplicate messages, not all of them are expected to implement such logic. For this reason, setting the T-bit explicitly is a more reliable mechanism to detects duplicates on the server side.

The T-bit in an SR is set on retransmitted messages that are triggered by a timeout (via application) as well as on retransmitted messages triggered by an imminent peer failure.

The T-bit is not set on any message that is retransmitted in response to any explicit error message (3002, 3004, and so on). An explicit error message is an indication that the original request message has reached some Diameter node and solicited a specific response from that node, and therefore is not considered a duplicate.

17.3.9 Handling of Diameter_Unable_To_Deliver (3002) error message

Diameter_Unable_to_Deliver (3002) is a protocol error message with E-bit set that is handled on a hop-per-hop basis (the node receiving this message must handle it and not just transparently pass it to the next node). 3002 is a response to an application request message, and its purpose is to notify the sender (the first upstream Diameter node) that the intended destination for this particular application and within this realm is not available on that path. This should trigger the node that receives such a response to retransmit the request message over an alternate path, that is the next peer found in the realm routing table.

After all eligible next hop peers are exhausted, or the Tx timer expires, the Diameter application is notified that the message request has failed. A configurable Tx timer is started after the first peer failover is performed (reaction to the first 3002 or 3004 error message received). All this is in the context of a single application request message. At this point the Diameter application determines, based on the server failover configuration, whether to retransmit this message. The destination-host AVP is cleared in retransmitted message.

For example:

- A 3002 message is received by the Diameter Base in the SR
- The TX timer is started
- The Diameter Base denylists the peer from which this message is received
- The request message in the pending queue is immediately re-transmitted to the next eligible peer
- This process continues until a viable peer is found (a request message is successfully delivered to the destination), the last eligible peer is tried, or the Tx timer expires, whichever event occurs first. The subsequent eligible peers can be only found in the realm routing table (and not the peering table).
- If the message is not successfully delivered within the Tx time, the Diameter Base notifies the application layer (NASREQ, Gx, Gy).
- The application executes the server failover procedure and potentially retransmit the request.
- At this point the application clears the destination-host AVP and the message is delivered to any server in the same realm supporting this application. This assumption is that the application servers support geo-redundancy.

17.3.10 Response to Diameter_Too_Busy (3004) error message

The processing of the Diameter_Too_Busy error message is the same as the processing of Diameter_Unable_To-Deliver (3002) error message.

The SR attempts to retransmit the message to an alternate peer by a Diameter Base, and only when the delivery fails on all peers, the Diameter application (NASREQ, Gx, Gy) are notified for potentially one last retransmission attempt in which the destination-host AVP is cleared.

In both of these cases (Diameter_Unable_To_Deliver and Diameter_Too_Busy), an agent or server tells a downstream peer that it is either too busy to handle a request or unable to route a request to an upstream

destination, perhaps because the destination itself is overloaded to the point of unavailability (RFC 7068, §4, *Diameter Overload Control Requirements*).

17.3.11 An SR as a transit Diameter node

In addition to Diameter client functionality where messages are originated and terminated, an SR node can also act as a transit node, routing messages between the peers. This functionality is primarily used in redundant Diameter multichassis configuration, where Diameter messages are sent between the two redundant SR nodes over inter-peer connection.

Transit Diameter request messages are never retransmitted. However, they are kept in a pending queue until a matching answer is received. If the answer to a request messages in the pending queue is not received within 10 seconds, the message is discarded. This 10 second timer is the same internal timer mentioned in section Retransmissions and message timers, which is also used for non-routable messages (such as CER/CEA, DPR/DPA).

17.3.12 Python support

Python processing of the Diameter messages is supported in Diameter base where the processing can be enabled per message type per direction.

17.4 3GPP-Based Diameter Credit Control Application (DCCA) - Online Charging

The 3GPP-based Diameter Credit Control Online Charging applications allow the control of subscriber access to services based on a pre-paid credit. The volume and time accounting on the node supports online charging using the Diameter Credit-Control Application (DCCA). The node supports Session Charging with Unit Reservation (SCUR), allowing the node to reserve volume and time quotas for rating-groups. Furthermore, the node supports centralized unit determination and centralized rating: it requests quotas and reports usage against the quota provided by the Online Charging Server (OCS). Credit control is always on a per-rating group basis. A rating group maps to a category inside a category map of the node volume-based and time-based accounting function.

The following are the basic configuration steps:

1. Configure a diameter policy.

   In the **config**>**aaa** CLI context, configure a Diameter peer policy with one or multiple Diameter peers.

```
configure
    aaa
        diameter-peer-policy "diameter-peer-policy-1" create
            description "Diameter peer policy"
            applications gy
            connection-timer 5
            origin-host "bng.domain.com"
            origin-realm "domain.com"
            source-address 10.0.0.1
            peer "peer-1" create
                address 10.1.0.1
                destination-host "server.domain.com"
```

```
                destination-realm "domain.com"
                no shutdown
            exit
        exit
    exit
```

When the diameter peer is reachable from IPv6, then the peer address should be specified as an IPv6 address. Optionally, an IPv6 source address can be specified:

```
configure
    aaa
        diameter-peer-policy "diameter-peer-policy-1" create
            ---snip---
            ipv6-source-address 2001:db8::1
            peer "peer-1" create
                address 2001:db8:100::1
                ---snip---
```

2. Configure a diameter application policy.

   In the **config**>**subscr-mgmt** CLI context, configure a diameter application policy:

   a. Set the application to Gy (Diameter Credit Control Application)

   b. Specify the Diameter peer policy to use and optionally specific additional Gy application specific parameters (for example AVP format).

```
configure
    subscriber-mgmt
        diameter-application-policy "diameter-gy-policy-1" create
            description "Diameter Gy policy"
            application gy
            diameter-peer-policy "diameter-peer-policy-1"
            gy
                avp-subscription-id subscriber-id type e164
                include-avp
                    radius-user-name
                exit
            exit
        exit
    exit
```

3. Create a category-map and define:

   - the credit type (time or volume)
   - a category defining the queues to monitor for quota consumption and the rating-group this category maps to in DCCA

```
configure
    subscriber-mgmt
        category-map "cat-map-1" create
            description "Category Map"
            credit-type time
            category "cat-1" create
                rating-group 1
```

```
                queue 1 ingress-egress
                exhausted-credit-service-level
                    pir 256
                exit
            exit
        exit
    exit
```

4. Create a credit control policy.

   Define the credit control servers to use by specifying the diameter application policy. Optionally, specify the default-category-map and an out-of-credit-action.

```
configure
    subscriber-mgmt
        credit-control-policy "cc-policy-1" create
            description "Credit Control Policy"
            credit-control-server diameter "diameter-gy-policy-1"
            default-category-map "cat-map-1"
            out-of-credit-action change-service-level
        exit
    exit
```

5. Configure the diameter credit-control-policy in the sla-profile of the subscriber host for which credit control should be activated.

```
configure
    subscriber-mgmt
        sla-profile "sla-profile-3" create
            description "SLA profile"
            credit-control-policy "cc-policy-1"
        exit
    exit
```

The following are examples of Diameter online charging flows:

- **scenario 1**

  Scenario 1 depicts a redirect use-case:

  When the quota is depleted, the subscriber is redirected to a web portal. When the credit is refilled, the OCS server notifies the BNG and provides a new quota. The configured out-of-credit-action when receiving a Final Unit Indication with action different from terminate is installed. See Figure 196: Online Charging scenario 1 - redirect (1/2) and Figure 197: Online Charging scenario 1 - redirect (2/2).

Figure 196: Online Charging scenario 1 - redirect (1/2)

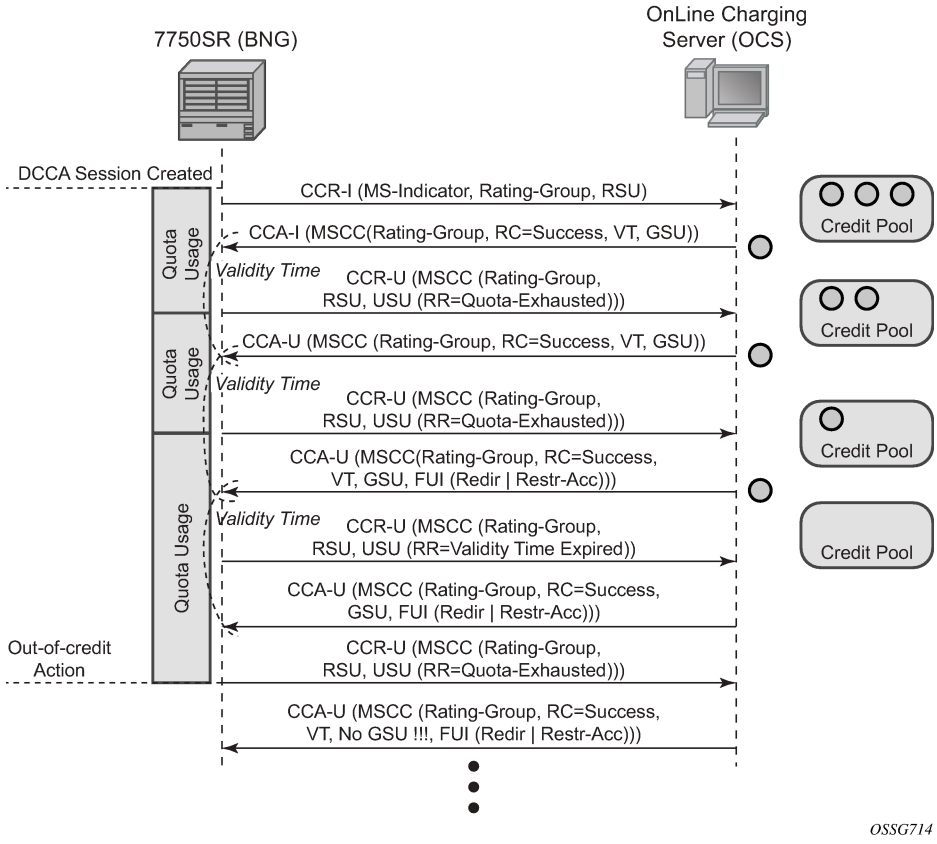
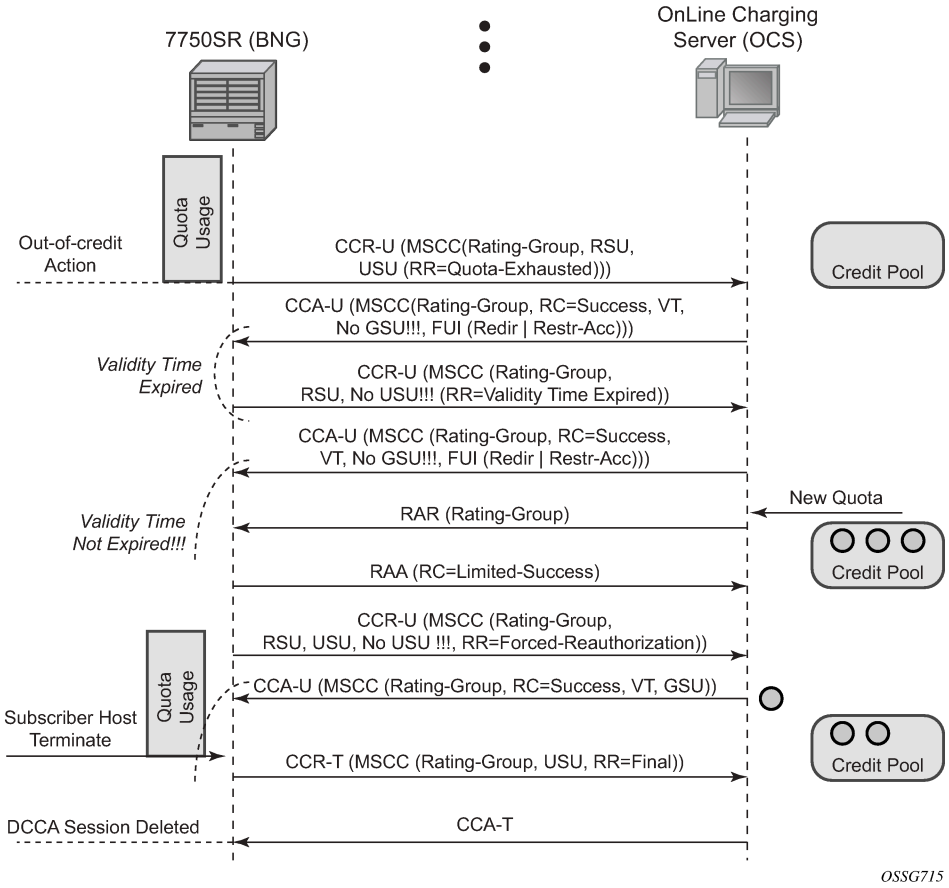


Figure 197: Online Charging scenario 1 - redirect (2/2)

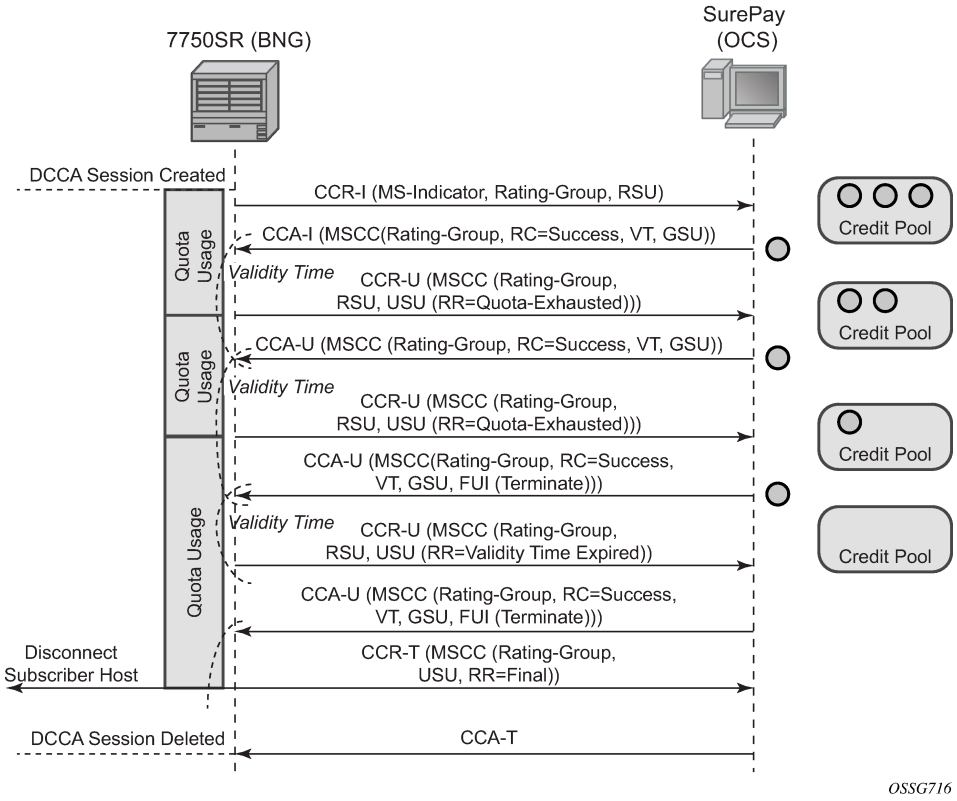


- **scenario 2**

  Scenario 2 depicts a terminate use case:

  When the quota is depleted after reception of a Final Unit Indication with action set to Terminate, the subscriber host is disconnected. The configured out-of-credit-action is ignored in this case. See Figure 198: Online Charging scenario 2 – terminate.

Figure 198: Online Charging scenario 2 – terminate



Abbreviations used in the previous drawings:

Table 60: Online Charging scenario legend

Abbreviation	Expansion
CCR	Credit Control Request (-Initial, -Update, -Terminate)
CCA	Credit Control Answer (-Initial, -Update, -Terminate)
RAR	Re-Authentication Request
RAA	Re-Authentication Answer
MSCC	Multiple Services Credit Control
GSU	Granted Service Unit
RSU	Requested Service Unit
USU	Used Service Unit
RC	Result Code
RR	Reporting Reason

Abbreviation	Expansion
VT	Validity Time

17.4.1 Diameter Gy out of credit actions

When all quota of a Diameter Gy credit control rating group is consumed and no additional quota is granted by the server, the out of credit action as configured in the **credit-control-policy** or **category-map** is executed.

The out of credit action is one of the following:

- **continue**

  Traffic corresponding to the rating group is no longer accounted for by credit control

  .

- **disconnect-host**

  The subscriber host or session is disconnected if any one of the categories within the category-map has expired credit

  .

- **block-category**

  Traffic corresponding to the rating group is blocked.

- **change-service-level**

  The service level is modified for the rating-group by applying one of the following:

  – a new PIR value to the corresponding queues or policers

  – an ingress or egress filter policy which could contain a filter to provide an HTTP redirect

17.4.1.1 Graceful service termination

RFC 4006, *Diameter Credit Control Application*, specifies a graceful service termination mechanism using the Final-Unit-Indication to indicate that an Answer message contains the final units for the service. When the final units are consumed, the action is specified with the Final-Unit-Action AVP.

```
Final-Unit-Indication ::= < AVP Header: 430 >
          { Final-Unit-Action }
         *[ Restriction-Filter-Rule ]
         *[ Filter-Id ]
          [ Redirect-Server ]
```

In SR OS, with the Final-Unit-Action AVP = TERMINATE, the subscriber host or session is disconnected. With the Final-Unit-Action = REDIRECT or RESTRICT_ACCESS, the **out-of-credit-action** as specified in the **credit-control-policy** or **category-map** is executed. In the case of REDIRECT, the URL specified in the Redirect-Server AVP is used when IPv4 HTTP redirect is enabled as the out of credit action and the following conditions are met:

- a Final-Unit-Indication AVP is present in the Multiple-Services-Credit-Control AVP of a CCA message
- the Final-Unit-Action AVP is set to REDIRECT (1)

- a Redirect-Server AVP is included with the following:
  - the Redirect-Address-Type AVP set to URL (2)
  - the Redirect-Server-Address AVP containing the URL to use for this rating group (**category-map**)
- the out of credit action for the corresponding rating group is set to **change-service-level** using one of the following CLI commands:
  - **configure subscriber-mgmt credit-control-policy** *policy-name* **out-of-credit-action change-service-level**
  - **configure subscriber-mgmt category-map** *category-map-name* **category** *category-name* **out-of-credit-action-override change-service-level**
- an IPv4 HTTP redirection action with **allow-override** is specified in the exhausted credit service level context for the corresponding rating group using the command **configure subscriber-mgmt category-map** *category-map-name* **category category-name exhausted-credit-service-level ingress-ip-filter-entries entry** *entry-id* **action http-redirect url allow-override**.

In all other cases, the URL specified in the Redirect-Server-Address AVP is ignored and the configured URL is used if HTTP redirect is enabled as the out of credit action.

The Restriction-Filter-Rule and Filter-Id AVPs included in the Final-Unit-Indication AVP are ignored.

Use the following **show** commands to find the active URL when an **out-of-credit-action change-service-level** is triggered that includes an IPv4 HTTP redirect action in the credit control ingress IP filter entries:

- **show service active-subscribers filter** [**subscriber** *sub-ident-string*]

  This command displays the active IPv4 ingress filter identifier.

- **show filter ip** *ip-filter-id*

  With the filter identifier found in the previous command, this command displays the credit control inserted entries (Origin = "Inserted on ingress by Credit Control"), including the IPv4 HTTP redirect action with the static configured URL. The optional **allow-radius-override** flag may be shown. This flag is common for RADIUS and Diameter-based overrides, where the flag indicates that the URL may have been overridden by the Diameter Credit Control server. Use the next command to determine if an override was specified.

- **show service active-subscribers credit-control**

  This command displays the URL received from the Diameter Credit Control server for a rating group or category with "Out of Credit Action = ChangeServiceLevel" active. The URL is displayed in the "HTTP Rdr URL Override" field.

17.4.2 Extended Failure Handling (EFH)

In a Diameter Gy application, that is, a Diameter Credit Control Application (DCCA), Credit Control Failure Handling (CCFH) determines the behavior of a credit control client in fault situations. When the CCFH value is set to CONTINUE and a failure occurs, the credit control client first attempts a failover procedure. If failover is not enabled or not supported by both client and server, or the failover is not successful, the client deletes the credit control session and continues the service to the end user without the Diameter Gy session until the user disconnects.

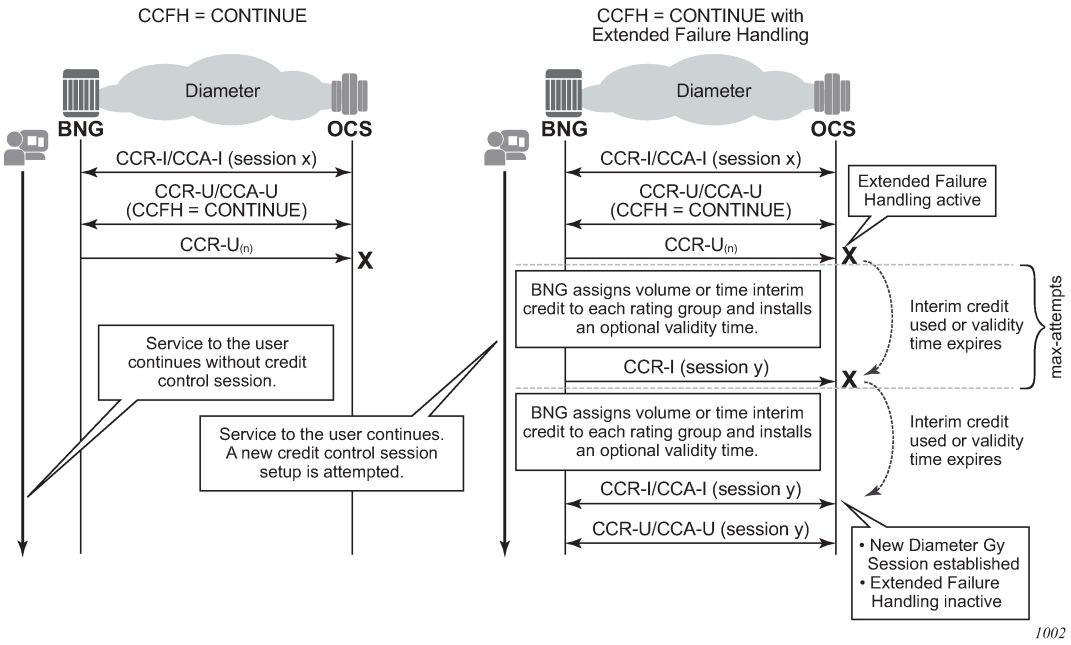
Extended Failure Handling (EFH) enables the credit control client to establish a new Diameter Gy session with the Online Charging Server (OCS) in failure situations where CCFH is triggered and the CCFH value is set to CONTINUE.

The following occurs when EFH is enabled and the CCFH value is set to CONTINUE.

- Service to the end user continues during failures (such as lost connectivity to the OCS).
- A new Diameter Gy session is established when the failure is restored.
- Usage information is kept up-to-date for reporting purposes (optional).

Figure 199: Extended Failure Handling (EFH) shows an example of EFH.

Figure 199: Extended Failure Handling (EFH)



If a failure occurs when EFH is active, a preconfigured time interim credit or volume interim credit with an optional validity time is assigned to all rating groups. A new Diameter Gy session setup is attempted each time the interim credit is used or the validity time expires. The following occurs when an attempt to re-establish the Diameter Gy session is made.

- The user session continues normally and EFH becomes inactive when the Diameter Gy session is successfully established with the OCS.
- A new interim credit with optional validity time is assigned to all rating groups if the Diameter Gy session is not established with the OCS.
- The user session is terminated if the Diameter Gy session is still not established with the OCS after a configurable maximum number of attempts.

17.4.2.1 EFH example call flow

Figure 200: EFH - example call flow

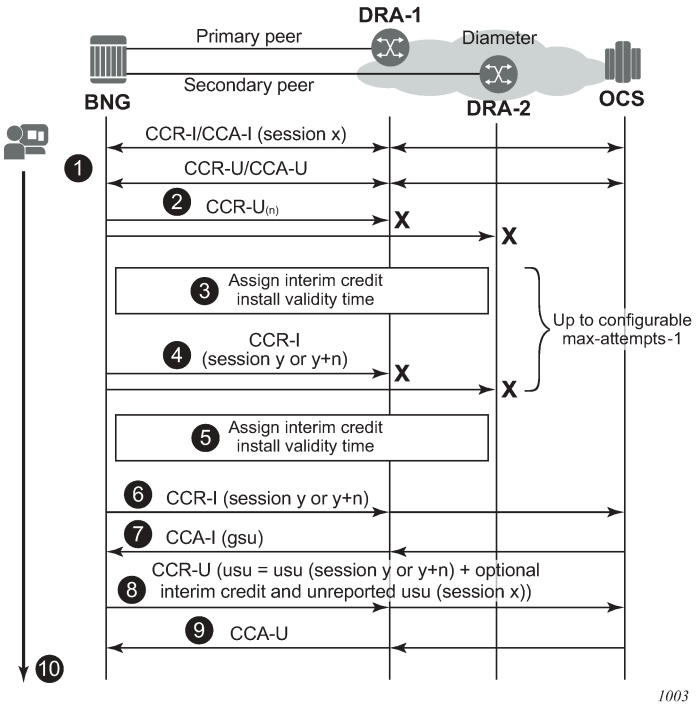


Figure 200: EFH - example call flow shows a sample call flow with EFH enabled. The following describes the call flow.

1. A Diameter Gy session is established between the Broadband Network Gateway (BNG), or credit control client, and the OCS, or credit control server.
2. A Credit Control Request Update (CCR-U) message is sent via the primary peer but no answer (CCA-U) is received. A timeout occurs which triggers a failover to the secondary peer. The same CCR-U is sent via the secondary peer. Again no answer is received. Because the CCFH value is set to CONTINUE for this session, EFH is activated.
3. The following EFH actions occur:
   - Service to the user continues uninterrupted.
   - If this is the first attempt to re-establish a Diameter Gy session:
     a. The failed Diameter Gy session (session ID x) is terminated without sending a CCR-T (Terminate) message.
     b. The unreported used credits for each rating group are stored in an EFH unreported credit counter.
     c. A new Diameter Gy session (new session ID y) is created internally but is not yet established with the OCS; the CCR-I (Initial) message is sent later.
   - For all subsequent attempts to re-establish a Diameter Gy session:
     a. The failed Diameter Gy session (session ID y or y+n) is terminated without sending a CCR-T (Terminate) message.

b. The unreported used interim credit for each rating group is added to the EFH unreported credit counter.

c. A new Diameter Gy session is created internally but not yet established with the OCS: the CCR-I (Initial) message is sent later. The internal Diameter session is created with the same session ID (y) or a new session ID (y+n) based on a configuration knob.

- Pre-configured interim credit is assigned to all rating groups and an optional validity time is installed.

4. If either all interim credit is used or the validity time expires for one of the rating group, an attempt is made to establish the new Diameter Gy session (session ID y or y+n) with the OCS.

A CCR-I message is sent via the primary peer but no answer (CCA-I) is received. A timeout occurs which triggers a failover to the secondary peer. The same CCR-I is sent via the secondary peer. Again no answer is received.

EFH stays active for this user session.

Steps 3 and 4 can be repeated multiple times until the maximum number of interim credit allocations is reached and the user session is terminated (not shown in this example call flow).

5. The EFH actions are as follows:

- Service to the user continues uninterrupted.
- The failed Diameter Gy session (session ID y or y+n) is terminated without sending a CCR-T (Terminate) message.
- The unreported used interim credit for each rating group is added to the EFH unreported credit counter.
- A new Diameter Gy session is created internally but is not yet established with the OCS: the CCR-I (Initial) message is sent later. The Diameter session is created with the same session ID (y) or a new session ID (y+n) based on the configuration.
- Pre-configured interim credits are assigned to all rating groups and an optional validity time is installed.

6. If either all interim credits are used or the validity time expires for one of the rating groups, an attempt is made to establish the new Diameter Gy session (session ID y or y+n) with the OCS.

A CCR-I message is sent via the primary peer.

7. An answer (CCA-I) is received with new granted service units (credit). Because communication with the OCS is restored, EFH becomes inactive.

8. The new Diameter Gy session resumes normal operation. Optionally, the EFH unreported credit usage is reported together with the usage from the newly granted credit in the next CCR-U credit negotiation for the rating group.

9. An answer (CCA-U) is received.

10. The service to the user continues and is uninterrupted during the OCS connectivity failure.

17.4.2.2 EFH triggers

For EFH to become active, the Credit Control Failure Handling (CCFH) value must be set to CONTINUE.

For a new session, the CCFH value is set in the configuration:

```
configure
    subscriber-mgmt
        diameter-application-policy <application-policy-name> [create]
            on-failure [failover {enabled|disabled}] handling continue
```

For ongoing sessions, the CCFH value is determined from the configuration or can be overridden by the OCS by including the following AVP in an answer message (CCA-I or CCA-U):

```
[427] Credit-Control-Failure-Handling AVP = CONTINUE (1)
```

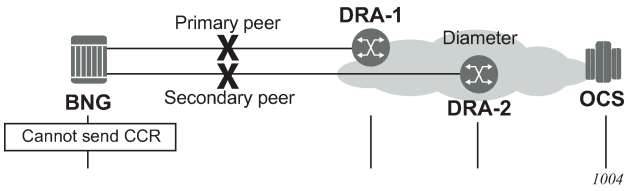
EFH is triggered when the CCFH value is set to CONTINUE and one of the following conditions occurs:

- **transmit failure**

  failure to send a CCR-I or CCR-U message

  Figure 201: EFH trigger: transmit failure shows an example of a transmit failure.

  Figure 201: EFH trigger: transmit failure

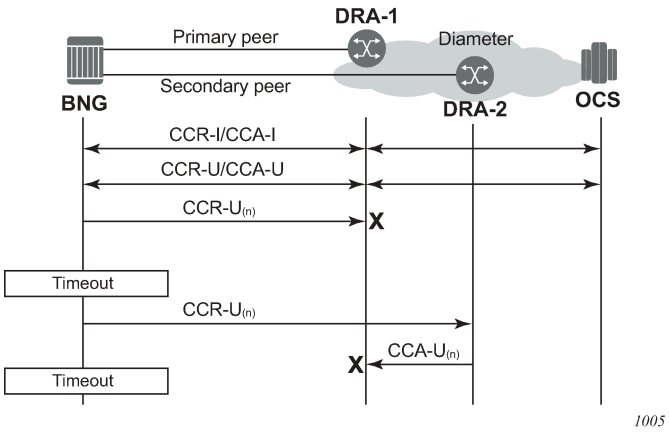


- **timeout**

  failure to receive an answer (CCA-I or CCA-U) within the configured timeout

  Figure 202: EFH trigger: timeout shows an example of a timeout.

  Figure 202: EFH trigger: timeout

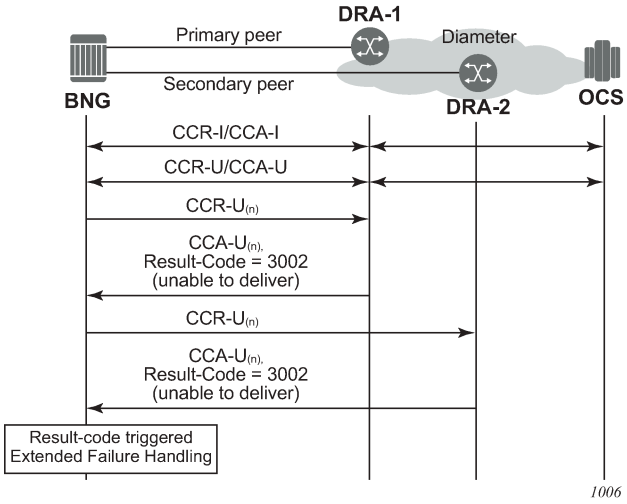


- **protocol error**

  failure because of a protocol error; seen as a Result-Code at command level in an answer message (CCA-I or CCA-U)

  Figure 203: EFH trigger: protocol error shows an example of a protocol error.

Figure 203: EFH trigger: protocol error



- **failure**

  reception of unknown Result-Code values in a Credit Control Answer message. Table 61: Diameter Gy known transient and permanent failures result-code values lists the known Transient and Permanent Failures Result-Code values.

Table 61: Diameter Gy known transient and permanent failures result-code values

Result code		Command level		MCSCC level
		CCA-I	CCA-U	
4001	DIAMETER_AUTHENTICATION_REJECTED	known	known	unknown
4010	DIAMETER_END_USER_SERVICE_DENIED	unknown	known	known
4011	DIAMETER_CREDIT_CONTROL_NOT_APPLICABLE	known	known	known
4012	DIAMETER_CREDIT_LIMIT_REACHED	unknown	known	known
5003	DIAMETER_AUTHORISATION_REJECTED	known	known	known
5030	DIAMETER_USER_UNKNOWN	known	known	unknown
5031	DIAMETER_RATING_FAILED	unknown	known	known

- a Diameter Gy message decoding error, for example (this is not an exhaustive list):
  - a missing Result-Code AVP
  - an unknown command code received
  - an incorrect session ID, origin host or origin realm in CCA
  - quota received for unexpected rating group
  - quota received with Result-Code 4012 (Credit Limit Reached)

17.4.2.3 Assigning interim credit

EFH interim credit can be specified in two ways:

- as a single volume interim credit value that is assigned to all rating groups of the Diameter Gy session with active EFH and that have no default credits configured
- ```
  configure subscriber-mgmt
     diameter-application-policy <application-policy-name> create
        gy
           extended-failure-handling
              interim-credit
                 volume <credits> {bytes|kilobytes|megabytes|gigabytes}
  ```
- as a single volume or time interim credit value per rating group (default credit)

  The activity threshold configured in the category map also applies to all rating groups that have time-based interim credit assigned while EFH is active.
- ```
  configure subscriber-mgmt
     category-map <category-map-name> create
        activity-threshold <kilobits-per-second>
        category <category-name> create
           default-credit time <seconds>
           default-credit volume <credits> <bytes|kilobytes|megabytes|gigabytes>
  ```

A single validity time value can be specified and applied to all rating groups that have interim credit assigned, regardless of whether the interim credit is configured for all rating groups in the Diameter application policy or per rating group in the category map.

```
configure subscriber-mgmt
   diameter-application-policy <application-policy-name> create
      gy
         extended-failure-handling
            interim-credit
                 validity-time <seconds>
```

The maximum number of times that interim credit is assigned to all rating groups of a Diameter Gy session when EFH is active can be limited in the configuration. The **max-attempts** value also corresponds with the maximum number of attempts to establish a new Diameter Gy session with the OCS (the default maximum attempts = 10).

```
configure subscriber-mgmt
   diameter-application-policy <application-policy-name> create
      gy
         extended-failure-handling
            interim-credit
               max-attempts <count>
               max-attempts infinite
```

An attempt to establish a new Diameter Gy session with the OCS is made when one of the following conditions occurs.

- The assigned interim credit for one of the rating groups is used.
- The validity time of the interim credit for one of the rating groups expires.

When the maximum number of attempts is reached and a new Diameter Gy session is not yet successfully established, then the user session is terminated; that is, the corresponding subscriber hosts are deleted from the system.

The reporting of used EFH interim credit can be enabled using the following CLI command:

```
configure subscriber-mgmt
   diameter-application-policy <application-policy-name> create
      gy
         extended-failure-handling
            interim-credit
               [no] reporting
```

With reporting enabled, the following accumulated used credit is reported when a new Diameter Gy session is established with the OCS and usage reporting is triggered for a rating group:

- unreported used credit granted via the initial Diameter Gy session that caused the EFH activation
- used interim credit when EFH was active
- used credit granted via the new Diameter Gy session

By default, reporting is disabled and all the used credit from the initial Gy session and all used interim credit are discarded.

17.4.2.4 Enabling EFH

EFH is enabled by specifying **no shutdown** in the **extended-failure-handling** CLI context in a Gy Diameter application policy.

```
configure subscriber-mgmt
   diameter-application-policy <application-policy-name> create
      gy
         extended-failure-handling
            no shutdown
```

Table 62: EFH states lists the EFH states for a Diameter Gy session.

Table 62: EFH states

EFH state	Description
Disabled	EFH is disabled (**shutdown**). EFH cannot be triggered for the Diameter Gy session.
Enabled - active	EFH is enabled (**no shutdown**) and active. A failure event occurred that triggered EFH. Interim credit is assigned to all rating groups and when either the validity time expires or the interim credit is used for a rating group, a new attempt is made to establish a Diameter Gy session with the OCS.
Enabled - inactive	EFH is enabled (**no shutdown**) and inactive.

EFH state	Description
	A Diameter Gy session with the OCS is established. EFH is activated for the Diameter Gy session if a trigger condition occurs.

17.4.2.5 Configuration example 1 - single volume interim credit value

```
configure subscriber-mgmt
    diameter-application-policy gy-1 create
        --- snip ---
        on-failure handling continue
        gy
            --- snip ---
            extended-failure-handling
                no new-session-id          # default
                interim-credit
                    no reporting           # default
                    volume 100 megabytes
                    validity-time 900
                    max-attempts 96
                exit
                no shutdown
            exit
        exit
    exit
```

In this example, EFH is enabled and when active, 100 Mbytes of volume interim credit is assigned to all rating groups of the Diameter Gy session with a validity time of 900 s. The maximum number of attempts to establish a Diameter Gy session with the OCS is 96.

Therefore, a maximum of 96 x 100 Mbytes or 9.6 Gbytes can be consumed before the user session is terminated (that is, the subscriber host is deleted from the system) when the OCS remains unreachable. Alternatively, when less than 100 Mbytes per rating group is consumed every 15 minutes (900 s), the user is disconnected after 900 s x 96 = 24 hours when the OCS remains unreachable.

17.4.2.6 Configuration example 2 - interim credit values per rating group

```
configure subscriber-mgmt
    category-map "cat-map-1" create
        activity-threshold 10
        category "cat-1" create
            description "Time interim credit per category"
            default-credit time 900
            rating-group 1
            queue 1 ingress-egress
        exit
        category "cat-2" create
            description "Volume interim credit per category"
            default-credit volume 10 megabytes
            rating-group 2
            queue 2 ingress-egress
        exit
        category "cat-3" create
            no default-credit
            rating-group 3
```

```
            queue 3 ingress-egress
        exit
    exit
    diameter-application-policy gy-1 create
        --- snip ---
        on-failure handling continue
        gy
            --- snip ---
            extended-failure-handling
                new-session-id
                interim-credit
                    reporting
                    volume 100 megabytes
                    validity-time 900
                    max-attempts 96
                exit
                no shutdown
            exit
        exit
    exit
```

In this example, EFH is enabled and when active, the following interim credit is assigned:

- 900 s of time interim credit to rating group 1 (category cat-1) with a validity time of 900 s. The activity threshold applies. When the usage stays below 10 kbytes/s, no time credit is used.
- 10 Mbytes of volume interim credit to rating group 2 (category cat-2) with a validity time of 900 s
- 100 Mbytes of volume interim credit to rating group 3 (category cat-3) with a validity time of 900 s

For inactive users, the validity time ensures that new Diameter Gy session connection attempts with the OCS are made at regular intervals.

For each attempt to establish a Diameter Gy session with the OCS, a new session ID is used.

When a new Diameter Gy session is successfully established, the EFH unreported credit for each rating group is included in the used service units when reporting is triggered for that rating group on the new Diameter Gy session.

When no new Diameter Gy session is established after 96 attempts, the user session is terminated; that is, the subscriber hosts are deleted from the system.

17.4.2.7 Monitoring the EFH state

Subscribers with Diameter Gy sessions that have EFH enabled can be displayed with the following CLI command:

```
# show service active-subscribers credit-control extended-failure-handling [state] [summary]
```

where the EFH state can be set to:

- **active**

  shows subscribers with Diameter Gy sessions that have EFH enabled and EFH is active
- **inactive**

  shows subscribers with Diameter Gy sessions that have EFH enabled and EFH is inactive
- **all**

  shows subscribers with Diameter Gy sessions that have EFH enabled and EFH is active or inactive.

Example output:

```
A:BNG-1# show service active-subscribers credit-control extended-failure-handling
===============================================================================
Active Subscribers
===============================================================================
-------------------------------------------------------------------------------
Subscriber ipoe-msap-002 (sub-profile-1)
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
(1) SLA Profile Instance sap:[1/1/4:1201.2] - sla:sla-profile-3
-------------------------------------------------------------------------------
Credit Control Policy: cc-policy-1
Category Map         : cat-map-1
Diameter Session Gy  : bng.domain.com;1464610029;840
CC Failure Handling  : continue
Extended Failure Handling (EFH)
  State              : active
  Attempts           : 1                     Maximum Attempts   : 10
  Active time        : 0d 00:00:10
  Total Active time  : 0d 00:00:10
  Total Active Count : 1
Category Name        : cat-1-time
Ingress Queues       : 1
Egress Queues        : 1
Ingress Policers     :
Egress Policers      :
Credit Volume Used   : 0                   Credit Time Used     : 11
Credit Volume Avail. : 0                   Credit Time Avail.   : 589
Credit Volume Thres. : 0                   Credit Time Thres.   : 0
Credit Expired       : False               Credit Negotiating   : False
Out Of Credit Action : None                Quota Holding Time   : 0
Validity Time Used   : 0                   Validity Time Avail. : 0
EFH Unreported Credit
  Current Volume     : 0                     Current Time       : 601
  Total Volume       : 0                     Total Time         : 601
Category Name        : cat-2-volume
Ingress Queues       : 4
Egress Queues        : 4
Ingress Policers     :
Egress Policers      :
Credit Volume Used   : 38400               Credit Time Used     : 0
Credit Volume Avail. : 10447360            Credit Time Avail.   : 0
Credit Volume Thres. : 0                   Credit Time Thres.   : 0
Credit Expired       : False               Credit Negotiating   : False
Out Of Credit Action : None                Quota Holding Time   : 0
Validity Time Used   : 13                  Validity Time Avail. : 587
EFH Unreported Credit
  Current Volume     : 1527600               Current Time       : 0
  Total Volume       : 1527600               Total Time         : 0
-------------------------------------------------------------------------------
IP Address
                MAC Address          Session        Origin       Svc        Fwd
-------------------------------------------------------------------------------
10.1.1.101
                00:51:00:00:00:02    IPoE           DHCP         1000       Y
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Number of active subscribers : 1
===============================================================================
```

The following information is displayed In the Extended Failure Handling (EFH) section of the example:

```
Extended Failure Handling (EFH)
  State              : active
```

State indicates that EFH is enabled and active. Another possible state is "inactive". When EFH is disabled, no EFH information is included.

```
Attempts           : 1                     Maximum Attempts   : 10
```

Attempts indicates the number of times interim credit has been assigned to all categories followed by an attempt to establish a new Diameter Gy session with the OCS. When the attempt to establish a new Diameter Gy session with the OCS is still failing after the Maximum Attempts value is reached, then the user session is terminated (that is, the subscriber hosts are deleted from the system).

```
Active time        : 0d 00:00:10
```

Active time indicates the time because the EFH state became active for this subscriber session.

```
Total Active time  : 0d 00:00:10
  Total Active Count : 1
```

Total Active time indicates the accumulative time of all occurrences that EFH was active during the lifetime of this subscriber session.

Total Active Count indicates the number of times that EFH was active during the lifetime of this subscriber session.

For each category (rating group), the EFH Unreported Credit is displayed:

```
EFH Unreported Credit
  Current Volume     : 0                     Current Time       : 601
  Total Volume       : 0                     Total Time         : 601
```

The Current Volume and Current Time counters contain, respectively, the unreported volume and time credit for the current occurrence of the EFH in an active state. These counters include the unreported used credit for the initial Diameter Gy session that caused the EFH state to become active and the unreported interim credit from previous attempts. Used interim credit for the current attempt per category (rating group) is shown in the following counters:

```
Credit Volume Used   : 0                   Credit Time Used     : 11
Credit Volume Avail. : 0                   Credit Time Avail.   : 589
```

The Total Volume and Total Time counters contain respectively the accumulated total unreported volume and time credit for the previous occurrences of EFH active state. The total counters are updated when the EFH state toggles from active to inactive. When interim credit reporting is enabled, the counters are reset to zero when the actual usage reporting happens for that rating group. When interim credit reporting is disabled, the counters are accumulating the total unreported credit during the lifetime of the subscriber session.

Current and Total Volume EFH Unreported Credit counters are the sum of used ingress and egress octets.

For each category (rating group), the validity time is displayed:

```
Validity Time Used   : 13                  Validity Time Avail. : 587
```

The following fields are only displayed when the EFH state is active:

- Extended Failure Handling (EFH): "Attempts", "Max Attempts" and "Active time"
- EFH Unreported Credit: "Current Volume" and "Current Time"

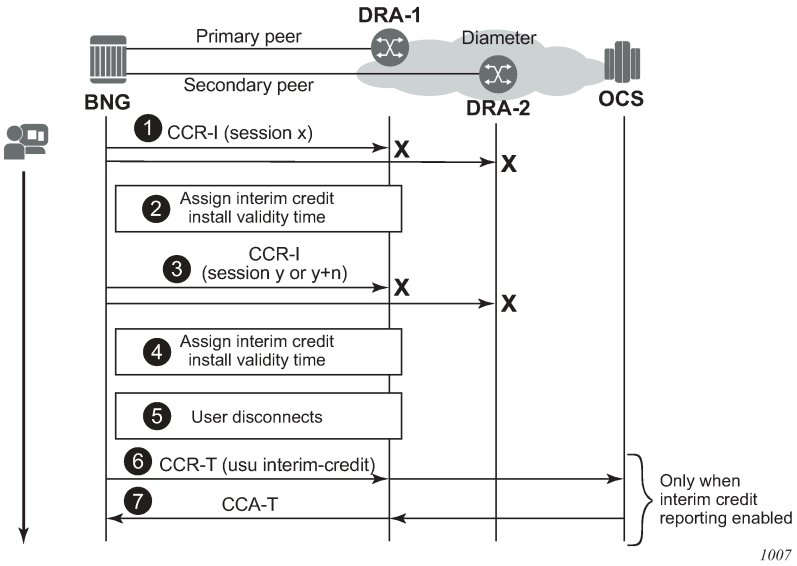
When EFH is disabled (shutdown), then the EFH information is not included in the credit control output.

17.4.2.8 Additional call flow examples

17.4.2.8.1 User disconnects while EFH is active

The call flow in Figure 204: EFH call flow - user disconnects during EFH shows a scenario where EFH is activated before the session is established with the OCS. The scenario is similar when EFH is activated by a CCR-U message.

Figure 204: EFH call flow - user disconnects during EFH



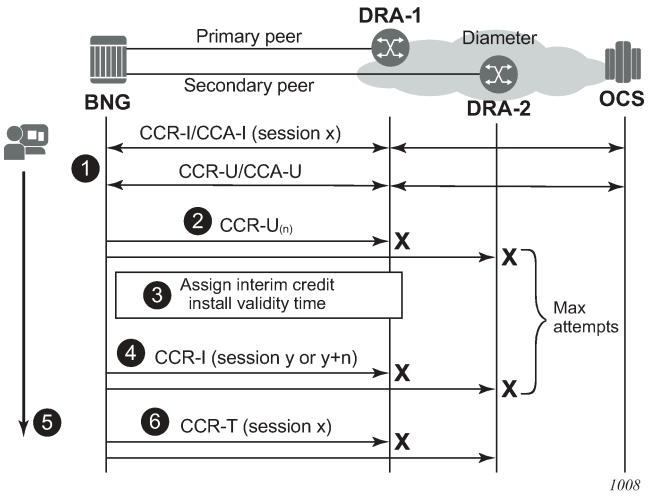
1. The initial Diameter Gy session setup fails: the CCR-I message sent to the primary peer times out or an error condition occurs that triggers Diameter Gy EFH. If failover is enabled, the CCR-I message is resent on the secondary peer. The CCR-I message sent to the secondary peer times out or an error condition occurs that triggers Diameter Gy EFH.
2. EFH becomes active for this user session. Interim credit is assigned to all rating groups and, optionally, a validity time is installed.
3. Interim credit is exhausted for a rating group or a validity time expires. A new attempt is made to establish the Diameter Gy session with the OCS; a CCR-I message is sent to the primary peer. A timeout or an error condition occurs that triggers Diameter Gy EFH to become active. If failover is enabled, the CCR-I message is resent on the secondary peer. The CCR-I message sent to the secondary peer times out or an error condition occurs that triggers EFH.
4. Interim credit is assigned to all rating groups and, optionally, a validity time is installed.
5. The user disconnects, resulting in a termination of the user session.

6. If interim credit reporting is enabled, a CCR-T is sent reporting the accumulated consumed interim credit.
7. If the OCS becomes reachable, a CCA-T may be received. Because the Diameter Gy session was not established with the OCS, the result code is DIAMETER_UNKNOWN_SESSION_ID.

17.4.2.8.2 Maximum number of attempts is reached

The call flow in Figure 205: EFH call flow - maximum attempts reached shows a scenario where the maximum attempts is reached to establish a Diameter Gy session with the OCS.

Figure 205: EFH call flow - maximum attempts reached



1. A Diameter Gy session is in progress.
2. An update message sent to the primary peer times out or an error condition occurs that triggers Diameter Gy EFH. If failover is enabled, the CCR-U message is resent on the secondary peer. The CCR-U message sent to the secondary peer times out or an error condition occurs that triggers Diameter Gy EFH.
3. EFH becomes active for this user session. Interim credit is assigned to all rating groups and, optionally, a validity time is installed.
4. Interim credit is exhausted for a rating group or a validity time expires. A new attempt is made to establish the Diameter Gy session with the OCS; a CCR-I message is sent to the primary peer. A new session ID is used for the first attempt. For subsequent attempts, a configuration command determines if a new or the same session ID is used. A timeout or an error condition occurs that triggers Diameter Gy EFH to become active. If failover is enabled, the CCR-I message is resent on the secondary peer. The CCR-I message sent to the secondary peer times out or an error condition occurs that triggers EFH.

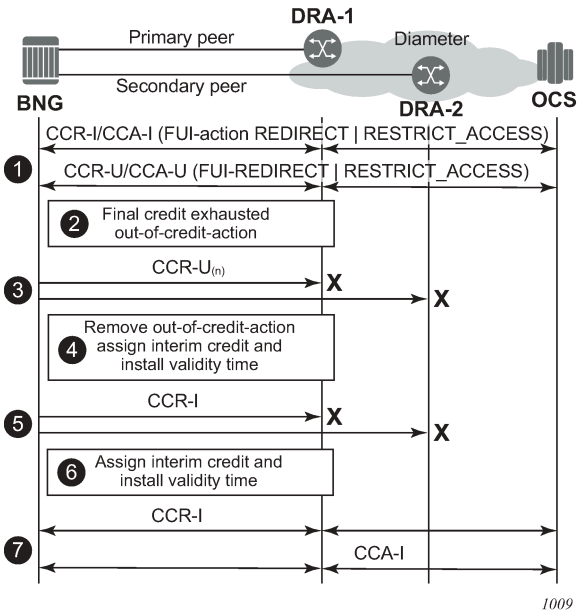
   Steps 3 and 4 are repeated until the maximum attempts for interim credit is reached.
5. The user session is disconnected; that is, the subscriber hosts are deleted from the system.
6. When interim credit reporting is enabled, a CCR-T is sent, reporting the accumulated consumed interim credit: the unreported quota from the initial session x and the accumulated interim credit.

17.4.2.8.3 EFH activation triggered during Final Unit Indication

This section describes two scenarios where EFH is activated during a graceful service termination initiated by the OCS with a Final Unit Indication (FUI) AVP. A graceful service termination with FUI action equal to REDIRECT or RESTRICT_ACCESS relies on a validity time or RAR to trigger a new credit negotiation with the OCS. Because the OCS is unreachable, it cannot be verified if a new quota has been granted. With EFH enabled, interim credit is assigned to guarantee service to the user until the connectivity with OCS is restored.

In the first scenario shown in Figure 206: EFH call flow - FUI scenario 1, the OCS initiates the graceful service termination with the Final-Unit-Action AVP = REDIRECT or RESTRICT_ACCESS. EFH is activated immediately after the out-of-credit action is installed.

Figure 206: EFH call flow - FUI scenario 1



1. A Diameter Gy session is in progress. Either the CCA-I or a CCA-U contains the Final-Unit-Indication AVP and the Final-Unit-Action AVP is set to REDIRECT or RESTRICT_ACCESS.
2. The final assigned credit is exhausted and because the Final-Unit-Action AVP is different from TERMINATE, the provisioned out-of-credit action is installed. A CCR-U is sent to the OCS to notify it that the Final-Unit-Action has started.
3. The CCR-U message sent to the primary peer times out or an error condition occurs that triggers diameter Gy EFH. If failover is enabled, the CCR-U message is resent on the secondary peer. The CCR-U message sent to the secondary peer times out or an error condition occurs that triggers Diameter Gy EFH.
4. EFH becomes active for this user session. Interim credit is assigned to all rating groups and, optionally, a validity time is installed. The out-of-credit action is removed.
5. Interim credit is exhausted for a rating group or a validity time expires. A new attempt is made to establish the Diameter Gy session with the OCS and a CCR-I message (new session ID) is sent to the primary peer. A timeout or an error condition occurs that triggers Diameter Gy EFH to become active.

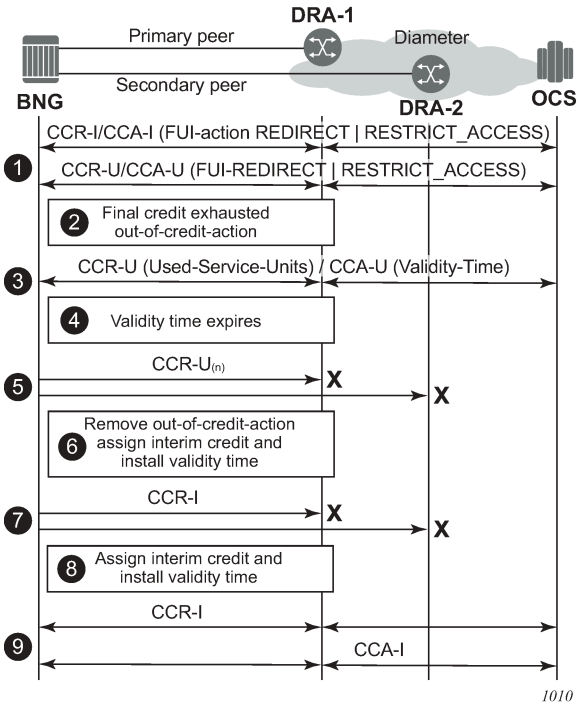
If failover is enabled, the CCR-I message is resent on the secondary peer. The CCR-I message sent to the secondary peer times out or an error condition occurs that triggers EFH.

6. A new interim credit is assigned to all rating groups and, optionally, a validity time is installed.
7. Interim credit is exhausted for a rating group or a validity time expires. A new attempt is made to establish the Diameter Gy session with the OCS. A CCR-I is sent with the same or a new session ID to the primary peer. Whether the same or a new session ID is used is determined by a configuration command.

   A CCA-I answer message is received. EFH becomes inactive and the user session continues with the new established credit control session. Optionally, the used interim credit can be reported via the new established session.

In the second scenario shown in Figure 207: EFH call flow - FUI scenario 2, the OCS initiates the graceful service termination with the Final-Unit-Action AVP = REDIRECT or RESTRICT_ACCESS. EFH is activated after the FUI validity time expires.

Figure 207: EFH call flow - FUI scenario 2



1. A diameter Gy session is in progress. Either the CCA-I or a CCA-U, contains the FUI AVP and Final-Unit-Action AVP = REDIRECT or RESTRICT_ACCESS.
2. The final assigned credit is exhausted and because the Final-Unit-Action is different from TERMINATE, the configured out-of-credit action is installed.
3. A CCR-U is sent to the OCS to notify that the Final-Unit-Action has started. The server responds with CCA-U containing a validity time.
4. The validity time expires.
5. The subsequent CCR-U message sent to the primary peer times out or an error condition occurs that triggers Diameter Gy EFH. If failover is enabled, the CCR-U message is resent on the secondary peer.

The CCR-U message sent to the secondary peer times out or an error condition occurs that triggers Diameter Gy EFH.

6. EFH becomes active for this user session. Interim credit is assigned to all rating groups and, optionally, a validity time is installed. The out-of-credit action is removed.
7. Interim credit is exhausted for a rating group or a validity time expires. A new attempt is made to establish the Diameter Gy session with the OCS; a CCR-I message is sent to the primary peer. For the first attempt, a new session ID is used. For subsequent attempts, a configuration command determines if a new or the same session ID is used. A timeout or an error condition occurs that triggers Diameter Gy EFH to become active. If failover is enabled, the CCR-I message is resent on the secondary peer. The CCR-I message sent to the secondary peer times out or an error condition occurs that triggers EFH.
8. A new interim credit is assigned to all rating groups and, optionally, a validity time is installed.
9. Interim credit is exhausted for a rating group or a validity time expires. A new attempt is made to establish the Diameter Gy session with the OCS, a CCR-I message is sent to the primary peer.

   A CCA-I answer message is received. EFH becomes inactive and the user session continues with the new established credit control session. Optionally, the used interim credit can be reported via the new established session.

17.4.3 Gy CCR-T replay

In diameter credit control, if the OCS is unreachable, EFH can be enabled to re-establish a Gy session with the OCS when the server becomes reachable again. This guarantees that no usage information is lost for billing. EFH applies to new and ongoing subscriber sessions.

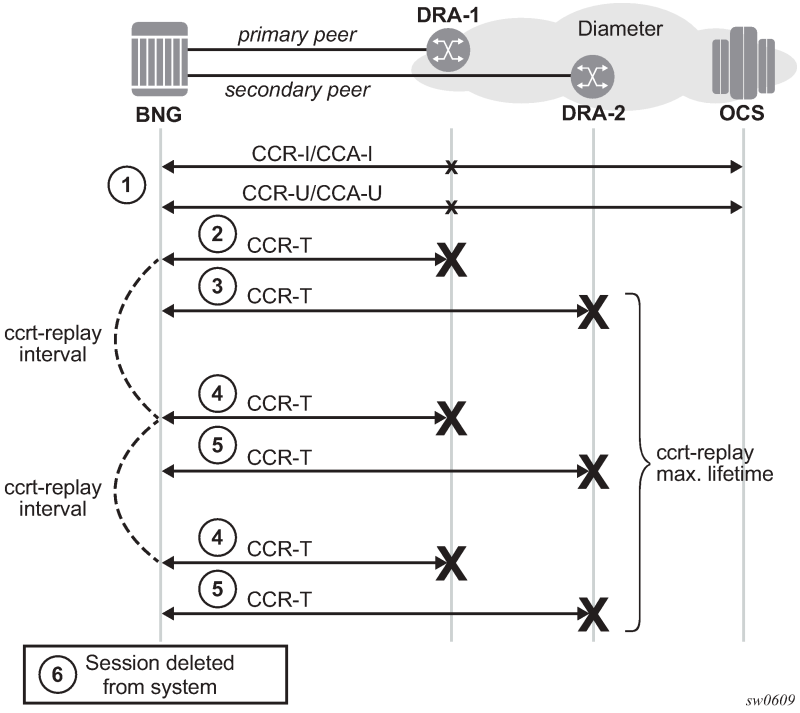
If the OCS is unreachable after subscribers disconnect, the CCR-T that contains the final usage reporting is lost. CCR-T replay is a mechanism to recover final usage reporting data during OCS failure by replaying the CCR-T at configured intervals for a configured maximum lifetime or until an answer (CCA-t) is received from the OCS.

When enabled, CCR-T replay is triggered after:

- a CCR-T transmit failure occurs. If session failover is enabled, the CCR-T transmit is retried on the secondary peer.
- no CCA-t is received within the configured timeout. If session failover is enabled, the CCR-T is retried on the secondary peer.

Figure 208: CCR-T replay illustrates CCR-T replay in action when the OCS is unreachable.

Figure 208: CCR-T replay



Where:

1. A diameter Gy session is in progress.
2. The subscriber disconnects, triggering the diameter Gy session to terminate. A CCR-T sent to the primary peer times out or an error condition occurs that triggers **ccrt-replay**.
3. If failover is enabled, the CCR-T is retried on the secondary peer. The CCR-T to the secondary peer times out or an error condition occurs that triggers **ccrt-replay**.
4. The CCR-T is replayed at the configured **ccrt-replay interval** until a valid answer is received or until the configured **ccrt-replay max-lifetime** is reached.
5. If failover is enabled, CCR-T replay is retried on the secondary peer.
6. The diameter Gy session in CCR-T replay is deleted when **crrt-replay max-lifetime** expires. An event is generated containing the Diameter Gx session ID.

The following configuration example enables CCR-T replay for Diameter Gy sessions:

Example:

```
configure subscriber-mgmt
    diameter-application-policy diam-gy-1 create
        gy
            ccrt-replay
                interval 1800
                max-lifetime 12
                no shutdown
            exit
        exit
    exit
```

The **ccrt-replay interval** value can be configured in seconds between 60 seconds (1 minute) and 86,400 seconds (24 hours).

The **ccrt-replay max-lifetime** value can be configured in hours between 1 hour and 24 hours.

Use the **show subscriber-mgmt diameter-session ccrt-replay** command to display Diameter Gy sessions that are in CCR-T replay mode.

Use the **show subscriber-mgmt diameter-session ccrt-replay diameter-application-policy** *name* command to display Diameter Gy sessions from the specified diameter application policy that are in CCR-T replay mode and per-diameter application policy statistics for those sessions.

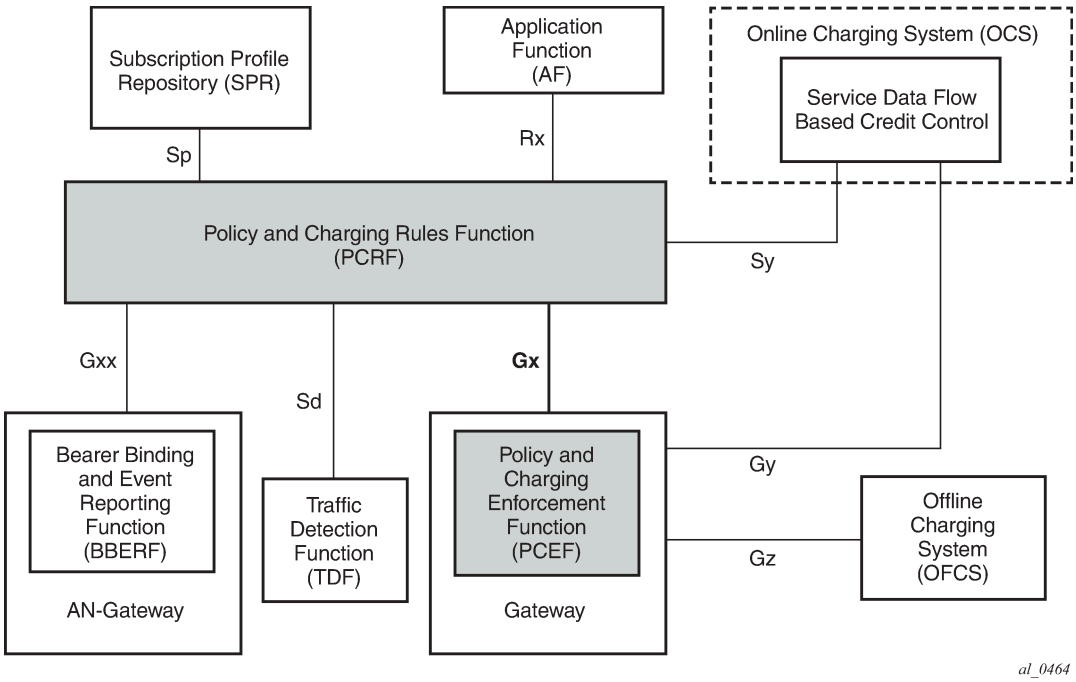
Use the **clear subscriber-mgmt diameter-session ccrt-replay diameter-application-policy** *name sessions* command to drop all Diameter Gy sessions that are in CCR-T replay mode.

Use the **clear subscriber-mgmt diameter-session ccrt-replay diameter-application-policy** *name* **statistics** command to clear the CCR-T replay statistics and update the "Statistics last cleared time" timestamp.

17.5 Policy management via Gx interface

Gx is a reference point in the network architecture model describing mobile service delivery. The network elements are described in various technical documents under the umbrella of 3GPP and are used to deliver, manage, report on and charge end-user traffic for mobile users. The Gx reference point is used for policy control and charging control. As shown in Figure 209: Gx reference point, it is placed between a policy server (Policy and Rule Charging Function (PCRF)) and a traffic forwarding node (Policy and Charging Enforcement Function) that enforces rules set by the policy server.

Figure 209: Gx reference point



The Gx reference point is defined in the Policy and Charging Control (PCC) architecture within 3GPP standardization body. The PCC architecture is defined in the document 23.203 while the Gx functionality is defined in the document 29.212. Gx is an application of the Diameter protocol (RFC 3588/6733).

Although the Gx reference point is defined within 3GPP standardization body (spurred by mobile/wireless industry) its applicability has spread to wire-line operation as well. For example, mobile operators that also have fixed line customers (residential plus business) would like to streamline policy management in their mobile and non-mobile domains with a single and already existing Gx based policy management infrastructure. In other words they want to integrate policy management of nodes serving fixed line subscribers into the system that is currently managing mobile subscriber base.

In such mixed environments, the node plays the role of a PCEF with an integrated TDF (Traffic Detection Function, or Application Awareness [AA] in ALU terminology) where policies and charging rules can be managed via PCRF.

With Wi-Fi Offload as a new emerging application, supporting Gx reference point on nodes is becoming even more important.

The Gx interface on the node encompasses the following functionality:

- Per subscriber host policy management
- Usage-Monitoring

Gx is applicable to ESM as well as to AA.

17.5.1 Gx protocol

The Gx application is defined as a vendor specific Diameter application, where the vendor is 3GPP and the Application-ID for Gx application is 16777238. The vendor identifier assigned to 3GPP by IANA is 10415.

When a Diameter protocol defined over the Gx interface, the node (PCEF) acts as a Diameter Client and the PCRF acts as a Diameter server. The Gx Diameter application uses existing Diameter command codes from the Diameter Base Protocol (RFC 6733) and Diameter Credit Control Application (RFC 4006), both of which are already implemented on the node.

Gx is using Attribute-Value Pairs (AVPs) for data representation within its messaging structures (command codes). AVPs in Gx come from various sources:

- Gx-specific AVPs defined in 3GPP Gx specification TS 29.212.
- Reused AVPs from other existing Diameter applications (RFC 4006, RFC 4005, and so on), other 3GPP specifications, ETSI, and so on.
- RADIUS reused attributes (AVP codes 0-255 are reserved for RADIUS re-used attributes)
- Vendor-specific AVPs

The initialization and maintenance of the connection between the node (PCEF) and the PCRF is defined by the underlying Diameter protocol as defined in RFC 3588/6733.

17.5.2 Policy assignment models

Subscriber and AA policies on the node (PCEF with integrated TDF) is assigned through the Gx protocol from the policy server (PCRF).

There are two modes of operation:

- **Pull mode**

The policy creation or modification is solicited by the node.

- **Push mode**

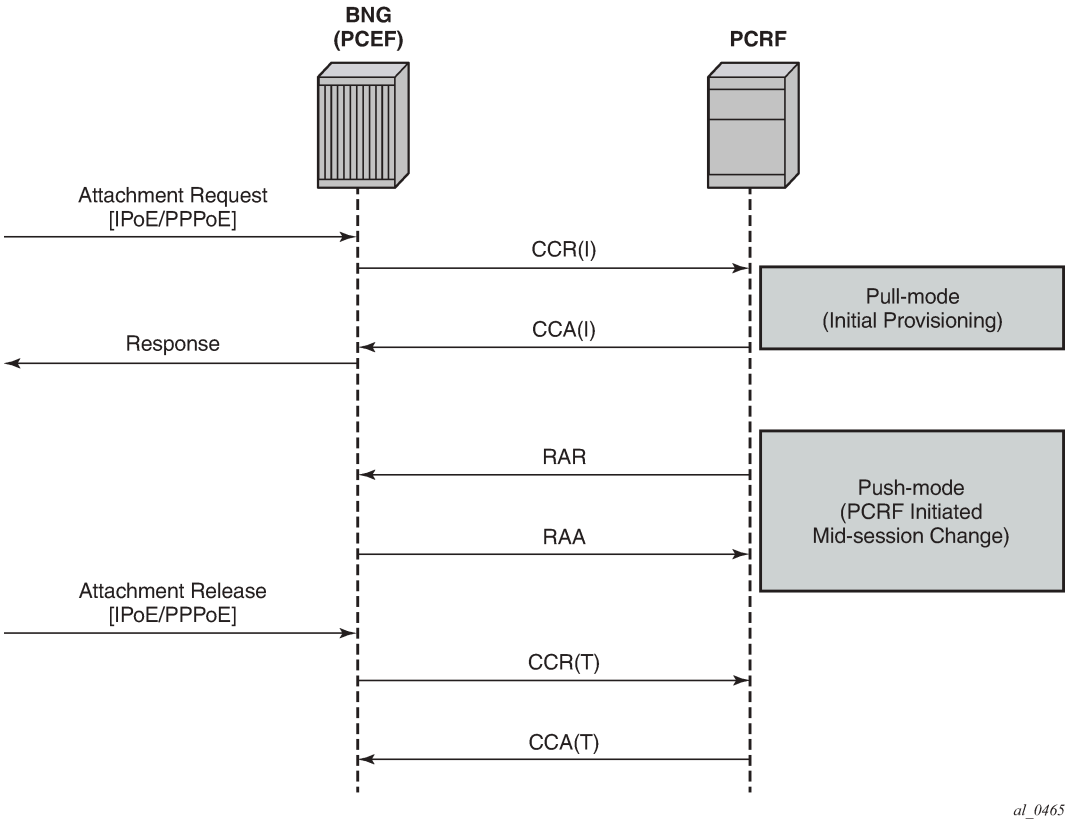
  The policy change is unsolicited by the node.

In the pull mode, during the host creation process, a user is authenticated by the AAA server. This process is independent from PCRF. After the user is authenticated and the IP address is allocated to it, the node sends a request for policies to the PCRF via CCR-I messages (initialization request message). This communication occurs via the Gx interface. The subscriber-host must be uniquely identified in this request toward the PCRF. This sub-identification over the Gx interface could be by means of IP address, username, SAP ID, and so on.

Based on the user identification, PCRF submits policies to the node. Those policies can range from subscriber strings (sub/sla-profiles/AA-profiles) to QoS and filter-related parameters.

In the push mode, the PCRF initiates the mid-session policy change through the Re-Authentication Request (RAR) message (Figure 210: Policy assignment models).

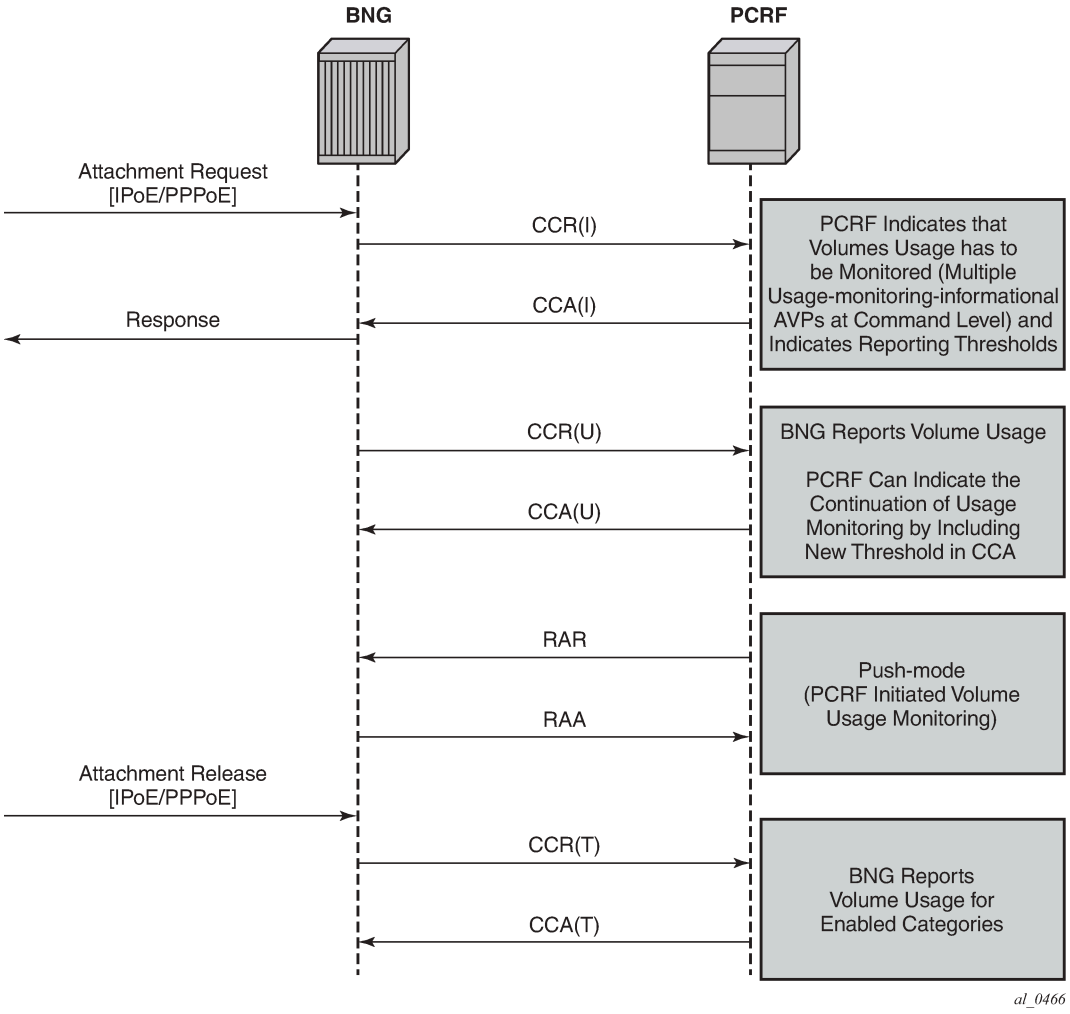
Figure 210: Policy assignment models



If that Usage-Monitoring is requested, the PCRF submitted policy changes are triggered by the Credit Control Request (Update) messages. This is based on ESM or AA Usage Monitoring. After the specified usage threshold is reached on the session-level, credit-category level, pcc rule level or application level on the node, the Usage-Monitoring is reported from the node to the PCRF in the CCR-U message. See the *7450 ESS, 7750 SR, and VSR Multiservice ISA and ESA Guide* for details on AA based Usage-Monitoring (Figure 211: On-demand usage reporting).

Alternatively, PCRF can request usage reporting on-demand via the **rar** command.

Figure 211: On-demand usage reporting



17.5.3 IP-CAN session – Gx session identification

The IP Connectivity Access Network (IP-CAN) session is a concept that has roots in mobile applications. A policy rule via the Gx interface can be applied/modified to an entity that is identified as an IP-CAN session (in addition to individual bearers within the IP-CAN session, the bearer concept is currently not applicable to the BNG). For example, a UE (user interface or simply a mobile phone) can host several services, each of which appears as a separate IP-CAN session associated with the same IP address. For example, in mobile technologies, an IP-CAN session can be defined as <IP_address, APN, IMSI>, where:

- APN (Access Point Name) is the service identifier
- IMSI (International Mobile Subscriber Identification) is the UE identifier (SIM Card)

In wireline environment (ESM deployments), an IP-CAN session identifies an entity to which the policy can be applied or modified. In SR OS, this can be a single or dual stack IPoE host, IPoE session, or PPPoE session.

For the purpose of identifying the subscriber host or session in the SR OS node in all Gx-related transactions, the SR OS node generates a unique Gx session-id AVP (RFC 6733, §8.8) per single or dual stack IPoE host, IPoE session, or PPPoE session. Note that the Gx Session-Id AVP is not the same as the Acct-Session-Id attribute used in RADIUS accounting.

If the IPoE session is enabled, the Gx session key can be based on one of the following combinations (configuration dependent):

- {SAP,MAC}
- {SAP,MAC,Circuit-Id}; the Circuit-Id must be present in the packet
- {SAP,MAC,Remote-Id}; the Remote-Id must be present in the packet

If the IPoE session is disabled, the Gx session key for IPoE hosts (dual or single stack) is, by default, based on the {SAP,MAC} combination.

In an environment where a Layer 3 node is in front of the BNG, the MAC address of arriving packets are that of the Layer 3 node. Then, it is not possible to differentiate between IPoE hosts on the same SAP unless the IPoE session concept is enabled in SR OS. Each IPoE session must have a Circuit-id or Remote-id as a differentiator.

A session concept is native to PPPoE where the Gx session equates to a single or dual stack PPPoE session. This means that the Gx session key is based on a {MAC,SAP, PPPoE session-id} combination.

17.5.3.1 User identification in PCRF

The following identification related AVPs are sent to the PCRF through Gx messages to aid in IP-CAN session identification:

- **subscription-id** AVP (RFC 4006, §8,46)

  This can be used to identify the subscribers on the PCRF. For the supported fields within the subscription-id AVP, see the *7750 SR and VSR Gx AVPs Reference Guide*.
- **NAS-Port-Id** AVP (RFC 2869 / §5.17; RFC 4005 / §4.3)
- **AN-GW-Address** AVP (3GPP 29.212 / §5.3.49)
- **Calling-Station-ID** AVP (RFC 4005 / §4.6)
- **user-equipment-info** AVP (RFC 4006, §8.49)
- **logical-access-id** AVP (ETSI TS 283 034)

  This contains the circuit-id from DHCPv4 Option (82,1) or interface-id from DHCPv6 option 18. The vendor-id is set to ETSI (13019).
- **physical-access-id** AVP (ETSI TS 283 034)

  This contains remote-id from DHCPv4 option (82,2) or DHCPv6 option 37. The vendor-id is set to ETSI (13019).

Physical and logical access IDs are also defined in BBF TR-134 (§7.1.4.1).

Table 63: PDP to PEP direction parameters shows PDP to PEP direction parameters.

Table 63: PDP to PEP direction parameters

Parameter	Category	Type	Description
Logical access ID	User identification	Octet String	The identity of the logical access to which the user device is connected. It is stored temporarily in the AAA function connected to PDP. This corresponds to the Agent ID in case of IPv4 and to the interface ID of DHCP option 82 for IPv6.
Physical Access ID	User identification	UTF8String	The identity of the physical access to which the user device is connected. It is stored temporarily in the AAA function connected to the PDP. This corresponds to the Agent Remote ID.

A Subscription-id AVP is most commonly used to identify the subscriber but any combination of the above listed parameters can be used to uniquely identify the IP-CAN session on PCRF and consequently identify the subscriber.

In addition, NAS-Port, NAS-Port-Type, and Called-Station-ID AVPs from RFC 4005 (§4.2, §4.4, §4.5) can be passed to the PCRF.

17.5.3.2 NAS-Port-Id as subscription-id

The node allows the NAS-Port-Id to be carried within the Subscription-Id AVP. Because the NAS-Port-Id may not be unique network-wide (two independent nodes may use the same NAS-Port-Id), there is a need for another identifier in conjunction with NAS-Port-Id to make the Subscription-Id unique across the network. This additional identifier is a custom string that can be appended to the NAS-Port-Id. It is defined when the NAS-Port-Id is configured for inclusion in Gx messages. See the *7450 ESS, 7750 SR, and VSR RADIUS Attributes Reference Guide* for information to format NAS-Port-Id AVP on the node.

The string can be used to identify the location of the node. For example:

```
@ALU-MOV-SITE-1
```

An example of the augmented NAS-Port-Id would look like:

```
NAS-Port-Id = lag-1.1/1/2:23.2000@ALU-MOV-SITE-1
```

where: 'lag-1.1/1/2:23.2000' is the part referencing the SAP on the node (port + vlan tags) and the '*@ALU-MOV-SITE-1*' is the node itself.

The NAS-Port-Id can be then inserted in the Subscription-Id AVP.

17.5.4 Gx interface and ESM subscriber instantiation

Policy management via Gx enables operators to consolidate policy management systems used in wireline (mostly based on RADIUS/CoA) and wireless environment (PCRF/Gx) into a single system (PCRF/Gx).

The model for policy instantiation/modification via Gx is similar to the one using RADIUS CoA. The authentication and IP address assignment is provided outside of Gx while the policy management function is provided via Gx.

Some PCRFs may require that the IP address information is passed from the node in CCR-I. This assumes that the IP address assignment phase (via LUDB, RADIUS or DHCP Server) is completed before the PCRF is contacted via CCR-I. Message flow for various protocols (DHCP, AAA, Gx) related to IPv4 subscriber-host instantiation phase is shown in Figure 212: Message slow during DHCPv4 host instantiation phase.

CCR-I message is sent to the PCRF after DHCP Ack is received from the DHCP server. Relaying DHCP Ack to the client in the final phase of the host instantiation process depends on the answer from the PCRF and the configuration settings of the fallback function if the answer is not received.

This model allows the IP address of the host to be sent in the CCR-I message, even though the subscriber-host is not fully instantiated at the time when the CCR-I message is generated.

AAA/LUDB must still be used for authentication and assignment of parameters necessary to place the subscriber host in the correct routing context (service-id, grp-id, msap-policy).

The start of the accounting process fits into this model because the host is not instantiated until the policy information from all sources (Gx, AAA, defaults) is known. After the final sub-profile (containing the acct-policy) is known, the host is instantiated and accounting can consequently be activated.

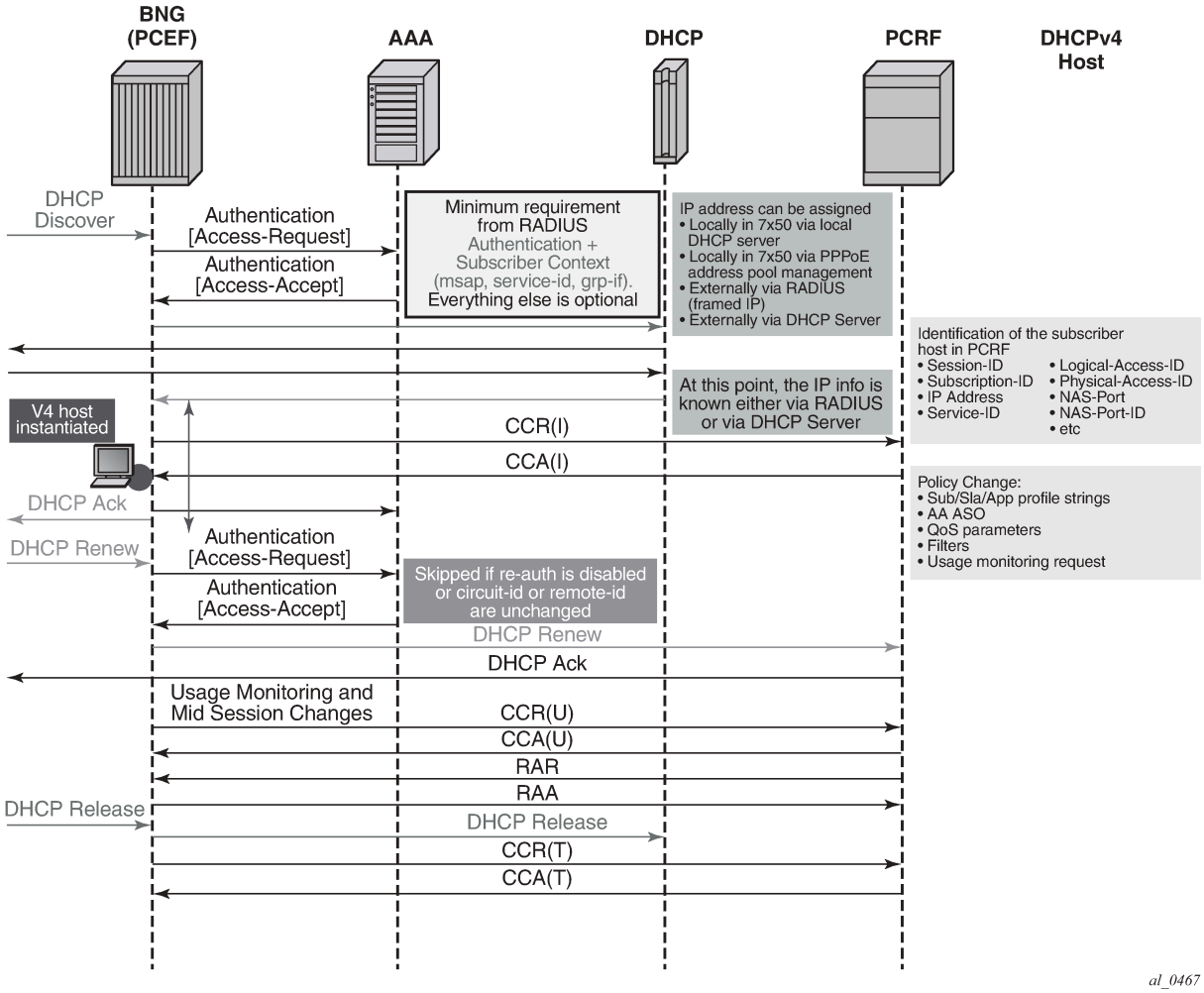
If the IP address cannot be assigned via Gx, and this functionality is outside of the Gx scope (3GPP TS 23.203 Rel12, Annex S, IP-CAN Session Establishment section).

The purpose of the CCR-I message is the following:

- To notify the PCRF that the sub-host was about to be instantiated on the node. Consequently, the PCRF creates a Gx session for the subscriber host if the CCR-I is successfully processed by PCRF.
- To identify the subscriber host in the PCRF, the PCRF uses the subscriber host identification information to identify the policy (for the subscriber host) that needs to be submitted to the node. The policy rules can be sent via CCA-i immediately following the initial CCR-I or they can be modified at any time during the subscriber-host lifetime through RAR messages.

Figure 212: Message slow during DHCPv4 host instantiation phase shows the message flow during the DHCPv4 host instantiation phase.

Figure 212: Message slow during DHCPv4 host instantiation phase



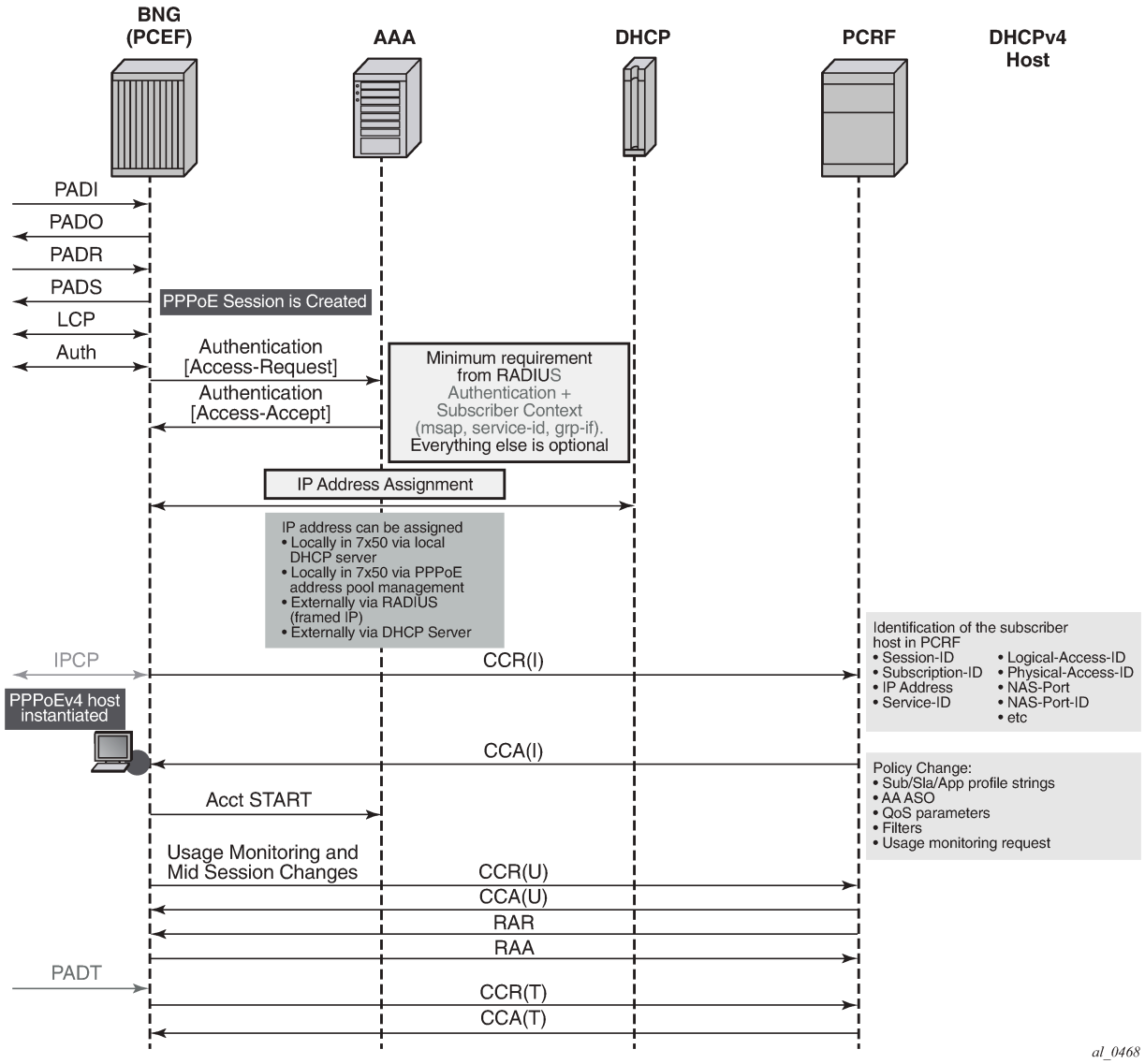
Message flow for PPPoEv4 host is similar. The host is instantiated after the answer from PCRF is received.

However, IPCP negotiation and Gx negotiation (CCR/CCA) are performed in parallel, independently of each other, and therefore the node does not wait for the Gx session to be established before the last IPCP ConfAck is sent (as it is for DHCP ACK).

After the host is instantiated on the node (after the CCA-i is received or as defined by the fallback action if the PCRF is not available), the Accounting-Start message is sent by the node (assuming that accounting is enabled).

The message flow is shown in Figure 213: Message flow during PPPoEv4 host instantiation phase.

Figure 213: Message flow during PPPoEv4 host instantiation phase



The host is created when the Gx session is established and therefore the subscriber host transitions into the traffic forwarding state when the Gx processing is completed. If the PCRF is unavailable or unresponsive, the host creation or termination is driven by the fallback configuration.

17.5.4.1 Gx and dual-stack hosts

Dual-stack (DS) hosts are treated as a single session from the Gx perspective, regardless whether an IPoE session concept is enabled. The difference between the IPoE session and non-IPoE session concept is related to the identification of the hosts for identification and policy management purposes.

If the IPoE session concept is disabled, a dual-stack host is identified by its {MAC,SAP} combination. Consequently, the key used for Gx session creation is based on the same {MAC,SAP} combination.

In some scenarios, this is not sufficient to differentiate hosts on the same SAP for Gx policy management purposes. As a result. a single Gx session is created for all hosts on that SAP. To avoid this creation, the IPoE session concept must be enabled in SR OS, where a Circuit-id or a Remote-id can be used for further differentiation between the Gx managed entities (subscriber-hosts).

The PCRF submits the policy rules per Gx session. The rules are then be applied to the underlying entity that is managed by this Gx session (single or dual stack IPoE host or IPoE session). For PPPoE deployment, the Gx session is automatically tied to the PPPoE session. A notion of a session is native to PPPoE (there is a session ID in PPPoE), and therefore, it is more natural to conceptualize the relationship between a PPPoE session (for single of dual-stack hosts) and a Gx session. By contrast, the concept of a session for IPoE is artificial, and which is why additional consideration is required for IPoE hosts, as described above.

The following example examines a case where IPoE session concept is disabled and consequently the IPoE dual-stack host is tied by a {MAC,SAP} combination.

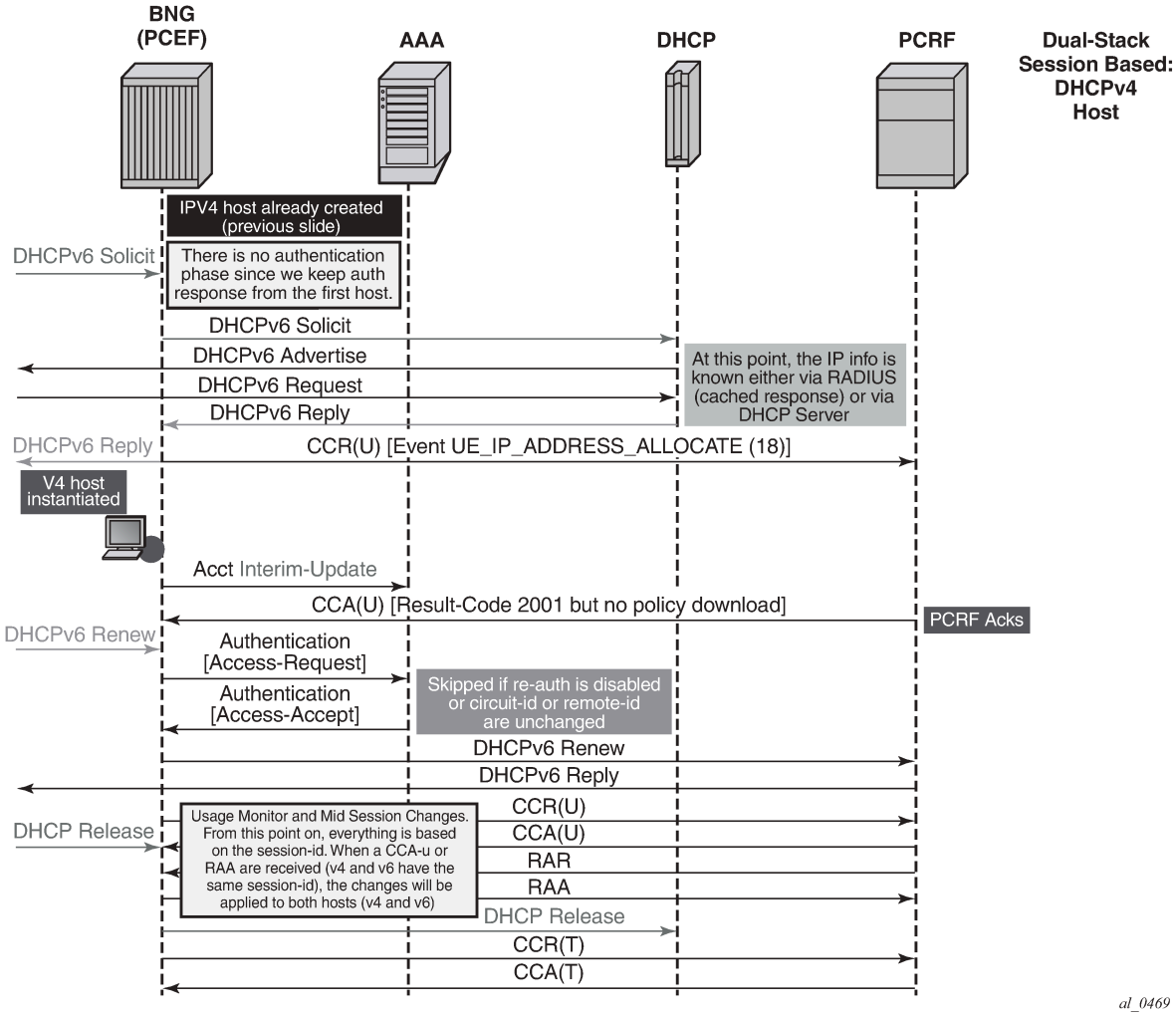
The CCR-I contains the IP address that was first allocated (meaning, the first IP address that triggered the session creation). The request for the second IP address family triggers (if enabled by configuration) an additional CCR-U that carries the IP address allocation update to the PCRF along with the UE_IP_ADDRESS_ALLOCATE (18) event.

Separately, the CCR-U content mirrors the content of the CCR-I with exception of the pre- allocated IP address or addresses. A single Gx message (CCR-I or CCR-U) carries the update for the DHCPv6 IA-NA+IA-PD and DHCPv6/PPPoE NA+PD address/prefix. This assumes that the NA+PD is requested in a single DHCP message.

Similarly, for the Gx session teardown, the CCR-U messages are sent carrying the UE_IP_ADDRESS_RELEASE event, followed by the CCR-T message.

The message flow is depicted in Figure 214: Gx and DS session instantiation.

Figure 214: Gx and DS session instantiation

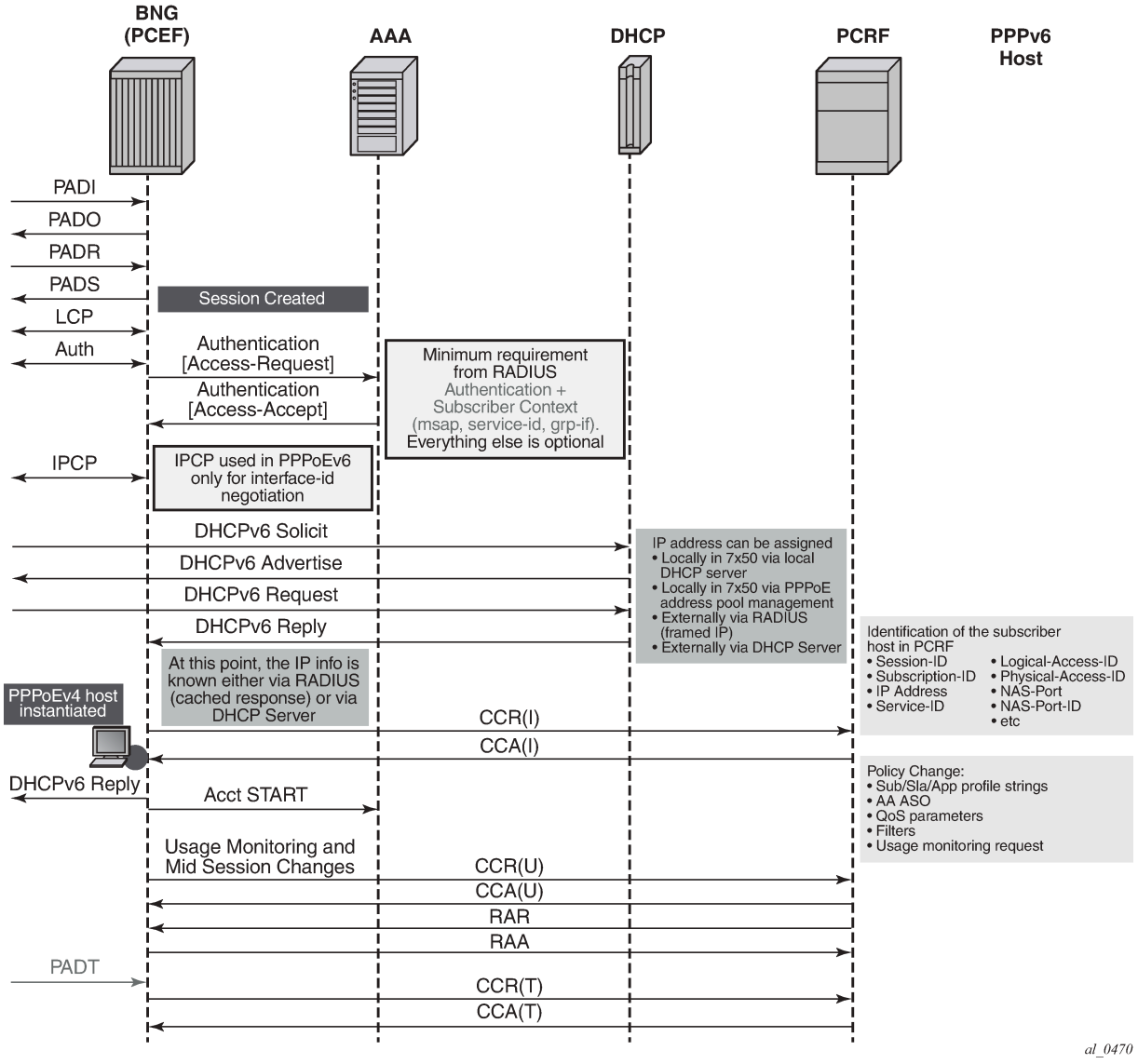


For Dual-Stack PPPoE host, the CCR-I is sent when the first IP address is assigned to the host. In the example in Figure 214: Gx and DS session instantiation, processing of the DHCPv6 Replay and CCR-U messages is performed in parallel. In other words, sending the DHCPv6 Reply message to the client is not delayed until the response from the PCRF is received. The reason being is that the Gx session is already established (triggered by the IPv4 host in our example) and all parameters for IPv4 and IPv6 are already known as received in CCA-i. Then, the CCR-U message is simply a notification message, informing the PCRF about the new IPv6 address/prefix being assigned to an existing client.

17.5.4.2 Gx and PPPoEv6-DHCP

For PPPoE v6 hosts, the IPv6 address is not obtained during the IPCP phase (only the interface-id is negotiated). Then, the node waits until the IPv6 address/prefix is allocated to the IPv6 hosts before it sends the CCR-I message. Otherwise, the IP address would not be available in CCR-I. This is shown in Figure 215: Gx and PPPoEv6 host instantiation.

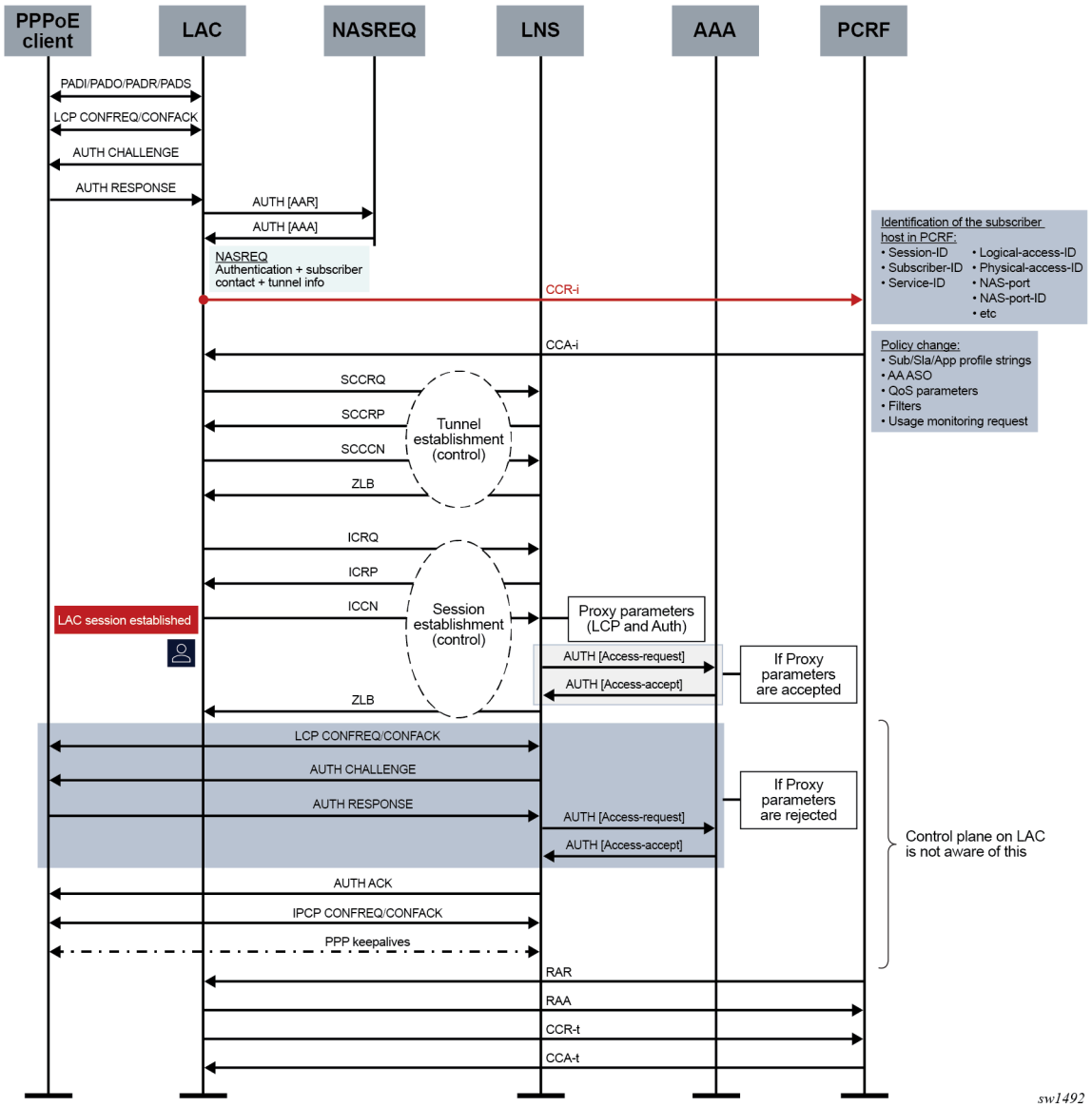
Figure 215: Gx and PPPoEv6 host instantiation



17.5.4.3 Gx on LAC

The Layer 2 access concentrator (LAC) does not have the knowledge of the IP address of the subscriber, as it is assigned by the L2TP Network Server (LNS). Consequently, the CCR-I messages originating from the LAC lacks the IP address of the subscriber. The following figure shows the Gx flow on the LAC.

Figure 216: Gx on LAC



17.5.5 Gx fallback function

The Gx fallback functionality refers to the behavior related to the subscriber host instantiation in situations where the PCRF is unresponsive while peering connections are up or the PCRF is unavailable with all peering connections down. This functionality affects only Gx session processing related to CCR-I messages on the node and has no effect on already established Gx sessions.

The fallback behavior can be controlled via the local configuration on the node or can be controlled via AVPs provided by PCRF.

PCRF-provided AVPs that control fallback behavior are:

- CC-Session-Failover AVP with the following values:
  - FAILOVER_NOT_SUPPORTED
  - FAILOVER_SUPPORTED
- Credit-Control-Failure-Handling AVP with the following values:
  - TERMINATE
  - CONTINUE
  - RETRY_AND_TERMINATE

If the fallback-related AVPs are not provided via PCRF, the node can provide a local configuration option to define the fallback behavior. If the response from the PCRF cannot be obtained, the local configuration can allow the subscriber host to be instantiated with default parameters, or alternatively the local configuration can deny subscriber host instantiation.

PCRF provided AVPs overrule the local configuration.

The local configuration that defines Gx fallback behavior can be found under the following CLI hierarchy:

```
config
    subscr-mgmt
      diam-appl-plcy
        on-failure
          failover {enabled|disabled}
            handling {continue|retry-and-terminate|terminate}
```

The **failover** configuration option (equivalent to CC-Session-Failover AVP) controls whether the secondary peer is used in if the primary peer is unresponsive. The unresponsiveness is determined by the timeout of the previously sent message.

The **handling** configuration option (equivalent to Credit-Control-Failure-Handling AVP) controls whether the subscriber is terminated or instantiated with default parameters if the PCRF is unresponsive.

Table 64: Handling options

	Handling: CONTINUE	**Handling: RETRY-AND-TERMINATE**	**Handling: TERMINATE**
Failover: ENABLED			
After the message sent to the primary peer times out, the secondary peer (and consecutive peers after that) is attempted. After the message times out after it has been sent to all available peers, the HANDLING action is examined to determine whether to terminate the host instantiation attempt or whether to use default	After the message times out after it has been sent to all available peers, the subscriber host is instantiated with default parameters (if they are configured)	After the message times out after it has been sent to all available peers, the subscriber host instantiation is terminated.	After the message sent to the primary peer times out, the subscriber host instantiation is terminated.

	Handling: CONTINUE	Handling: RETRY-AND-TERMINATE	Handling: TERMINATE
parameters to instantiate the host.			
Failover: DISABLED			
After the message sent to the primary peer times out, the HANDLING action is examined to determine whether to terminate the host instantiation attempt or whether to use default parameters to instantiate the host.	After the message sent to the primary server times out, the subscriber host is instantiated with default parameters (if they are configured)	After the message sent to the primary peer times out, the subscriber host is terminated.	After the message sent to the primary peer times out, the subscriber host is terminated.

The CCR retransmissions are controlled by the **tx-timer** command under **diameter-application-policy**.

If all peers are down (no connections are open), the **handling** action determines the behavior. If the action is set to **continue**, the subscriber host is immediately instantiated with the default-settings (provided that the defaults are available). In all other action cases, the host instantiation is immediately terminated.

17.5.6 Gx CCR-I replays

As described in the previous section, the subscriber host can optionally be (configuration controlled) established with default settings (sla-profile, sub-profile, app-profile) if the PCRF is not available to answer the CCR-I. This results in a subscriber-host state mismatch between the node and the PCRF, where the subscriber-host is established on the node but there is no corresponding Gx session established in the PCRF.

To resolve this situation, ESM periodically sends CCR-I for the Gx **orphaned subscriber-host** until the response from PCRF is received. The CCR-I is periodically retransmitted every 60 seconds.

17.5.7 Gx CCR-T replays

Termination of the subscriber-host on the node without termination of the corresponding Gx session in PCRF results in a state mismatch between the node and the PCRF, whereby the host Gx session is present in the PCRF while it is removed from the node.

Some PCRFs can cope with such out-of-sync condition by periodically auditing all existing Gx sessions. For example, a **probing** RAR can be sent periodically for each active Gx session. The sole purpose of this probing RAR is to solicit a response from the PCEF (node) and provide indication on whether the corresponding Gx session is alive on the node or is vanished. The 'probing' RAR can contain an Event-Trigger that is already applied on the node, or if none is applied, then the Event-Trigger can contain NO_EVENT_TRIGGER. In either case, the probing RAR does not cause any specific action to be taken in the node and it is used only to solicit reply from PCRF.

To minimize the impact on performance, **probing** RARs are sent infrequently; therefore, it may take days to discover a stale Gx session on the PCRF. The node supports a mechanism that can clear the stale session in the PCRF sooner. It does this by replaying CCR-T messages until the correct response from the

PCRF is received (CCA-t). The CCR-T messages are replayed up to 24 hours. Both the interval at which the CCR-T messages are replayed and the max-lifetime period are configurable. If the max-lifetime period expires before a valid answer is received, the CCR-T is deleted and a log is generated. The log contains Gx session ID.

The T-bit (retransmission bit) is set in all replayed CCR-T messages.

The following command clears all orphaned sessions on the node for a specified Diameter application policy:

```
clear subscriber-mgmt diameter-session ccrt-replay diameter-application-policy gx-policy-name
 sessions
```

17.5.7.1 RAR and CCR-T replay

Certain scenarios allow the PCRF to send a RAR message to an orphaned Gx session running CCR-T replays on the node. The ESM host associated with the orphaned Gx session does not exist and therefore RAR cannot be applied.

In this scenario, the node replies with RAA carrying Result-Code= DIAMETER_UNKNOWN_ SESSION_ID (5002).

17.5.7.2 CCR-T replay and multichassis redundancy

The first CCR-T reply for each Gx session is synchronized, but the consecutive CCR-T replays for the same Gx sessions are not synchronized. When the answer (CCA-t) is received, the CCR-T replay is terminated and this event (deletion of CCR-T replay) is synchronized to the other node.

CCR-T replays are sent from the node that was in SRRP active state at the time when the CCR-T was initiated. They continue to be sent from the same node even if the SRRP is switched over in the meantime.

This entire process can be thought of as if the CCR-T initiating node (active SRRP) armed its MCS peer with CCR-T replay for a Gx session. This occurs at the very beginning, when a CCR-T replay is first initiated for a Gx session. The armed node stays silent until the MCS peer that is actively sending CCR-T replays for a Gx session, reboots. Only when the MCS peer reboots, the armed node starts sending CCR-T replays for a Gx session in the following fashion: first message is sent with cleared T-bit, followed by replays at the configured replay interval and a fresh 24 hour lifetime.

17.5.7.3 CCR-T replay and high availability

On CPM switchover, the newly active CPM resumes the CCR-T replay with the T-bit set until the lifetime, which is synchronized between the CPMs, expires.

17.5.8 Automatic updates for IP address allocation/de-allocation

During the subscriber-host setup phase, the first allocated IP address is sent in the CCR-I message from the node to the PCRF.

Each subsequent IP address allocation or de-allocation for the same host can optionally trigger a CCR-U, notifying the PCRF of the IP address allocation/de-allocation event.

This behavior can be enabled via the following CLI command:

```
configure
    subscriber-mgmt
      diameter-application-policy <pol-name>
        gx
          [no] report-ip-addr-event
```

The IP address allocation/de-allocation event driven CCR-U message carries the respective event code [UE_IP_ADDRESS_ALLOCATE(18) or UE_IP_ADDRESS_RELEASE(19)] along with the corresponding IP address.

The IP address allocation/de-allocation events are applicable to the following addresses:

- Framed-IP-Address (AVP Code 8) IPv4
- Framed-IPv6-Prefix (AVP Code 97) SLAAC
- Delegated-IPv6Prefix (AVP Code 123) IA-PD
- Alc-IPv6-Address (AVP Code 1023) IA-NA

These event-codes are only sent in CCR-U messages and not in CCR-I and CCR-T messages (when the host is instantiated and terminated).

Examples:

- IPv6 attachment request arrives with two IP addresses: IA-NA and IA-PD. This is a new host. CCR-I is generated with two IP addresses included (IA-NA and IA-PD, assuming that request for their allocation is carried in the same DHCPv6 message).
- Afterward, the attachment request for an IPv4 address arrives on the same host. CCR-U is generated with the event UI_IP_ADDRESS_ALLOCATE and corresponding AVP (framed-address) is sent to the PCRF. No IP address other than this new IPv4 address is sent.
- RAR is received for the (any) policy change. The node replies with RAA and it contains all three IP addresses (AVPs) that have been allocated to the host.

If the IP address notification event is enabled, the node-originated Gx message carries all known IP addresses/prefixes associated with the subscriber-host (Gx session), unless those messages contain one of the two event codes:

UE_IP_ADDRESS_ALLOCATE(18) or UE_IP_ADDRESS_RELEASE(19).

If one of those two events is present in the Gx message, the IP address/prefix carried in that message is only relevant to the event contained in the message (address/prefix allocated or released).

If the IP address notification event is disabled, the node only sends the IP address from the first host. This IP address is included in all messages related to the Gx session. If this IP address is removed (de-allocated) mid-session from the dual-stack host, the node stops advertising it, or any other address, from Gx messages for that particular session.

17.5.9 DHCPv4/v6 re-authentication and RADIUS CoA interactions with Gx

If re-authentication for DHCPv4/v6 hosts is enabled, any policy changes that may be submitted during re-authentication (for example sla-profile update via Access-Accept) overwrites the one previously applied, regardless of the source of the policy update. For example, if that the Gx policy is applied to a subscriber host via RAR (mid-session policy update) and then later an overlapping policy with different values is submitted via RADIUS or LUDB during the re-authentication phase, the RADIUS/LUDB submitted policy

overwrites the one applied via Gx. In other words, the origin of the current policy in effect is not maintained internally in the system and therefore the overlapping policy update cannot be prioritized according to the source of the policy.

The following guidelines should be followed if the policy is provided via Gx:

- If LUDB access is enabled, there should be no overlap between the LUDB provided parameters and Gx provided parameters. LUDB is accessed during every DHCP lease renew process and consequently parameters configured via LUDB would overwrite parameters provided by Gx.
- If LUDB access is enabled, there should be no overlap between the LUDB provided parameters and Gx provided parameters. LUDB is accessed during every DHCP lease renew process and consequently parameters configured via LUDB would overwrite parameters provided by Gx.
- If re-authentication is enabled, there should be no overlap between the RADIUS provided parameters and Gx provided parameters. With re-authentication enabled, RADIUS is contacted during every DHCP lease renew process and consequently parameters configured via RADIUS would overwrite parameters provided by Gx.

These guidelines are not applicable for PPPoE subscriber-hosts because re-authentication cannot be enabled for PPPoE hosts. Consequently, LUDB or RADIUS parameters cannot override Gx provided parameters.

Coexistence of RADIUS CoA and Gx for the same host is allowed. The two policy change mechanisms are independent of each other and therefore, they can override each other. For example, if the RADIUS CoA for policy change for the host is received, the policy is updated but the PCRF (Gx) is not notified of the change. If both policy management mechanisms are deployed simultaneously, then it is the operator's responsibility to synchronize the actions between the two.

17.5.10 Gx, ESM and AA

Although the ESM subscriber and the AA subscriber are two separate instantiations on the node, their policy management and Usage-Monitoring are handled uniformly through a single Gx session.

17.5.10.1 ESM subscriber-host vs AA subscriber

Because ESM and AA modules are part of integrated service offering (ESM with residential AA on the same node), they share the same subscriber-id string. However, Gx interface in ESM is primarily applicable to hosts (basic entity to which policy is applied) while AA has no awareness of hosts. AA is only aware of subscribers (which is, in broader terms, a collection of hosts within a residence). See the *7450 ESS, 7750 SR, and VSR Multiservice ISA and ESA Guide* for details on Application Assurance concepts.

17.5.10.2 AA subscriber state

AA subscriber state must exist for App-profiles and ASO overrides to be applied.

The app-profile for the aa-sub is applied explicitly by a CCR-I or RAR message with an AVP AA-App-Profile-Name.

App-profiles interact with ASO characteristics in this way:

- The AA-App-Service-Options AVP within the app-profile assignment is optional at subscriber instantiation time and may be used later to modify the policy.

- The newly submitted AA-App-Profile-Name AVP overwrites the one that is already applied. Any ASO AVPs that is received within the Gx message is applied.

Note: If an app-prof AVP is present, even if it is the same app-profile as currently applied, all previous ASO override policies are removed for the sub.

The state of the subscriber policy attributes is modified by ASO AVPs in this way:

- The app profile can define one or more ASO characteristics attributed to a subscriber.
- If there are multiple ASO AVPs for the same characteristic in the message, the first one takes effect.
- There is no explicit delete of ASO overrides (the PCRF can always resend or change the app-profile to delete all overrides).

17.5.11 Policy management via Gx

A policy change can be implicitly requested by the node at IP-Can session establishment time via the CCR-I command. The node supplies user identification attributes to the PCRF so that the PCRF can identify rules to be applied. However, the node does not explicitly request specific policy update, for example via Event-Trigger = RESOURCE_MODIFICATION_REQUEST.

Another way to request a policy update on the node is via the **rar** command in a PUSH model.

Gx policies on the node can be enforced via these three mechanisms:

- **Gx-based overrides**

  This refers to subscriber related overrides (sub/sla/aa-profile, subscriber-id, QoS, filter, category-map, and so on).

- **PCC rules or IP-criterion based rules**

  This refers to fully constructed Policy and Charging Control (PCC) rules with multiple qos/filter actions per rule and its own traffic classification.

- **NAS filter entry inserts via Gx**

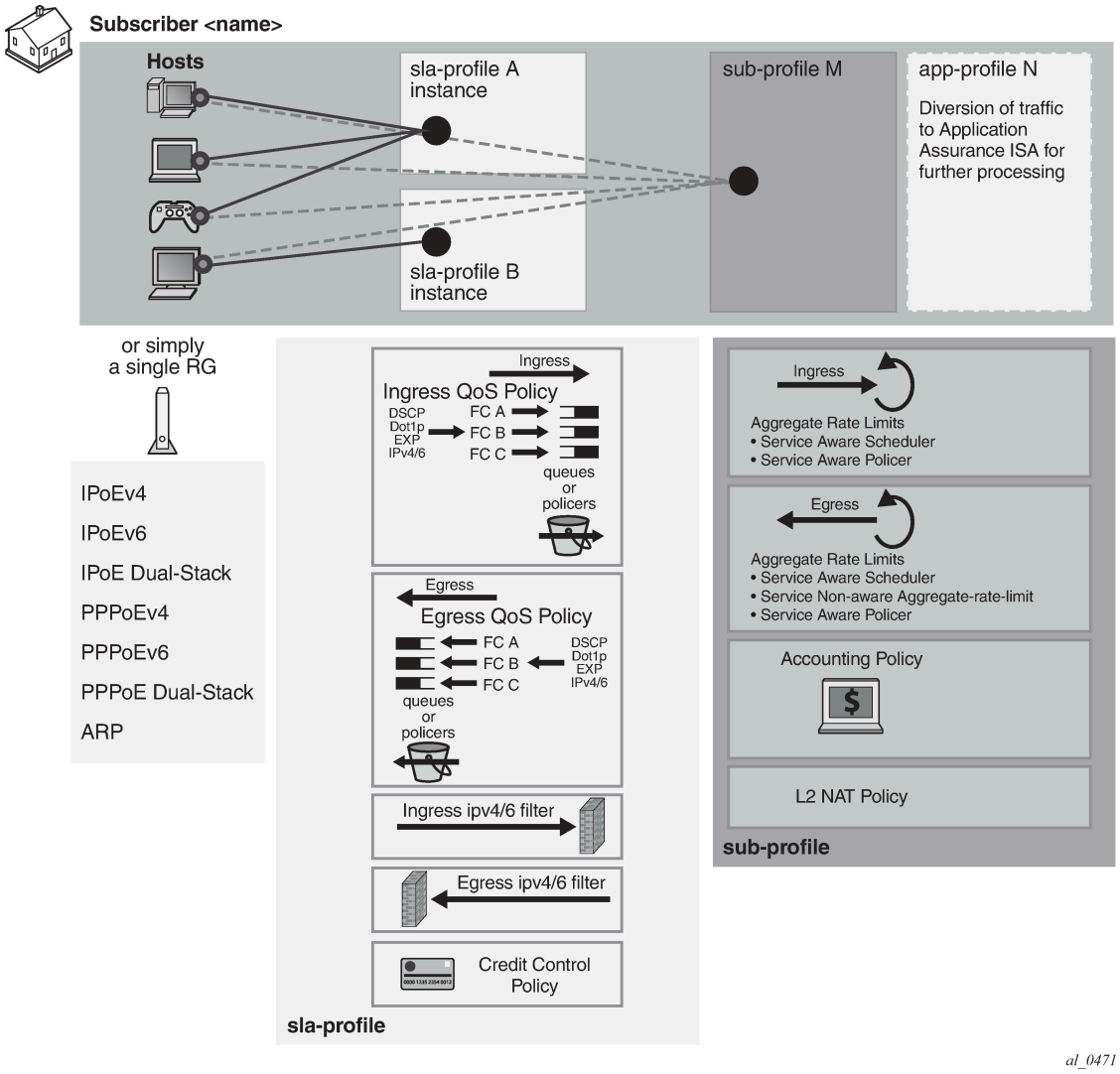
  This refers to basic permits or denies entries, similar to ACL filter entries.

17.5.12 Gx-based overrides

Gx-based overrides refer to the activation or modification of the existing subscriber host-related objects on the node.

Subscriber host-related objects are shown in Figure 217: ESM objects managed via various policies and profiles. A subscriber represents a residence or home and it is identified by Subscriber-Id string on the node. A subscriber on the node can consist of multiple hosts in a bridged home environment or a single host in a routed home environment.

Figure 217: ESM objects managed via various policies and profiles



The two basic concepts in ESM context are sla-profile with its associated objects and sub-profile with its associated objects.

- Sla-profile defines a service level (rates, queues, filters) for a group of hosts sharing the same sla-profile instance within the subscriber. There can be multiple sla-profile instances per subscriber.
- Sub-profile defines aggregate rate of the subscriber along with accounting policy. There is only one sub-profile per subscriber.

17.5.12.1 Instantiation of Gx overrides

For a list of Gx related AVPs supported on the node, see the *7750 SR and VSR Gx AVPs Reference Guide*.

Gx overrides are installed via Charging-Rule-Install AVP (for ESM or AA) or ADC-Rule-Install AVP (for AA only – 3GPP Release 11) sent from the PCRF toward the node.

AVP Format:

```
Charging-Rule-Install ::=       < AVP Header: 1001 >
                                      *[ Charging-Rule-Definition ]
                                      *[ Charging-Rule-Name ]
                                      *[ AVP ]

ADC-Rule-Install ::=       < AVP Header: 1092 >
                                      *[ ADC-Rule-Definition ]
                                      *[ ADC-Rule-Name ]
                                      *[ AVP ]
```

Every Gx override must have a Charging-Rule-Name (ESM) or ADC-Rule-Name (AA - 3GPP Release 11 and Release 12) associated with it. This is important for returning the override status from the node to the PCRF upon the override instantiation.

The objects (subscriber-hosts) to which the new overrides are applied must exist on the node; otherwise, the override installation fails.

The parameters defining a new override simply replaces the existing parameters that are already applied to the subscriber-host, without the need to remove the previously installed parameters.

There are four types of overrides that are currently supported via Gx:

- ESM string overrides (sla/sub/app-profiles, subscriber-id)
- update of subscriber host QoS information (queue rate change)
- filter override for the subscriber host (including one-time http redirect)
- category-map installation/override
- HTTP redirect override within the filter

A Charging-Rule-Name AVP within the Charging-Rule-Install grouped AVP can have several meanings:

- It can directly reference a preconfigured filter in the system.
- It can directly reference a preconfigured subscriber profile (sla and sub).
- It can directly reference a preconfigured category map.
- It can directly reference an ESM string (such as an inter-destination-string used to associate the subscriber host with a Vport configuration).
- Or, if there is an HTTP redirect override, it serves as a special container within which the HTTP redirect URL is defined. Note that HTTP redirect override is not to be confused with PCC rule-based HTTP redirect.

In all of the above cases, the existing objects applied to the subscriber-host is replaced with the referenced object.

It is important to distinguish two locations for invoking the Charging-Rule-Name AVP for overrides:

- **directly under the Charging-Rule-Install AVP**

  Then, the Charging-Rule-Name references the predefined structures (profiles, filter-ids, cat-maps, and so on) on the node. The type of the structure is contained within the Charging-Rule-Name AVP in the form of a reserved keyword that has to be prepended to the identifier of structure:

  Filter installation/overrides:

  – Charging-Rule-Name = Ingr-v4:<id>

  – Charging-Rule-Name = Ingr-v6:<id>

– Charging-Rule-Name = Egr-v4:<id>
– Charging-Rule-Name = Egr-v6:<id>
– Charging-Rule-Name = In-Othr-v4:<id> (othr - one-time-http-redirect)
– Charging-Rule-Name = In-Othr-v6:<id> (othr - one-time-http-redirect)

Subscriber-Id override:

– Charging-Rule-Name = Sub-Id:sub-id-string

Profile installation/overrides:

– Charging-Rule-Name = Sla-Profile:sla-profile-name
– Charging-Rule-Name = Sub-Profile:sub-profile-name

Inter-destination string override:

– Charging-Rule-Name = Inter-Dest:inter-dest-string

Usage-Monitoring:

– Charging-Rule-Name = Cat-Map:category-map-name

AA:

– Charging-Rule-Name = AA-UM:<string-name>
– Charging-Rule-Name=AA-Functions:<string-name>

In summary, the reserved prefixes ingr-v4:,ingr-v6:, egr-v4:, egr-v6:, in-othr-v4:, in-othr-v6:, sub-id:, sla-profile:, sub-profile:, inter-dest:, cat-map:, aa-um: and aa-functions: have special meaning within the Charging-Rule-Name AVP on the node.

One exception to this is HTTP redirect override, which is described separately.

- **under the Charging-Rule-Install**

  Charging-Rule-Definition AVP. In this case, the override itself is not pre-provisioned on the node but instead, directly defined in the Charging-Rule-Definition. Part of the override definition is the name assignment via Charging-Rule-Name AVP. The Charging-Rule-Name AVP is used to report on the override status.

  For example, the Charging-Rule-Name AVP for QoS overrides is an arbitrary name. This AVP is part of Charging-Rule-Definition AVP in which QoS-Information is provided. Such Charging-Rule-Name is used to report errors related to instantiation of the override.

  Another example of the override defined within the Charging-Rule-Definition AVP is HTTP redirect override:

```
Charging-Rule-Name = v4-http-url:<name>
Redirect-Information ::= < AVP Header: 1085 >

[ Redirect-Address-Type ]
[ Redirect-Server-Address ]
```

  and

```
Charging-Rule-Name = v6-http-url:<name>
Redirect-Information ::= < AVP Header: 1085 >
[ Redirect-Address-Type ]
 [ Redirect-Server-Address ]
```

The ADC-Rule-Name AVP within the ADC-Rule-Install grouped AVP handles application policy related processing (AA). This AVP is applicable under the ADC-Rule-Install—ADC-Rule-Definition AVP. In this case the ADC rule itself is not pre-provisioned on the node but instead directly defined in the ADC-Rule-Definition. In AA, such rule definition can define AA overrides that are applied to the subscriber. In other words, the existing objects for the subscriber are replaced with the ones referenced in the rule. Part of the ADC rule definition is the ADC rule name assignment via ADC-Rule-Name AVP. The ADC-Rule-Name defined in such manner is used to report on the rule status.

The AA-Functions: prefix in the ADC rule name is reserved for ADC rule definitions applicable to AA-functions (namely, app-profile and ASOs):

ADC-Rule-Name = AA-Functions:aa-rule-name

Then the aa-rule-name is an arbitrary name that is used in rule status reporting.

If that ADC-Rule-Name is used in AA Usage-Monitoring, the "AA-Functions:" prefix must not be present (Usage-Monitoring in AA is covered in detail in the *7450 ESS, 7750 SR, and VSR Multiservice ISA and ESA Guide*).

Note: AA-Function AVP and AA-Usage-Monitoring cannot coexist in the same ADC rule.

The Charging-Rule-Definition AVP (AVP code 1003, 3GPP 29.212 §5.3.4) is of type Grouped, and it defines the override sent by the PCRF to the node.

The Charging-Rule-Name in this AVP can be arbitrarily set and it is used to uniquely identify the override in error reporting.

```
Charging-Rule-Definition ::= < AVP Header: 1003 >
                                { Charging-Rule-Name }
                                [ QoS-Information ]
                                [ Nas-Filter-Rule]
                                [ Alc-NAS-Filter-Rule-Shared]
                                *[ AVP ]
```

The ADC-Rule-Definition AVP (AVP code 1094, 3GPP 29.212 §5.3.87) is of type Grouped, and it defines the ADC override sent by the PCRF to the node. The ADC-Rule-Name AVP within the ADC-Rule-Definition AVP uniquely identifies the ADC policy rule and it is used to reference to a policy rule in communication between the node and the PCRF within one IP CAN session.

```
ADC-Rule-Definition ::= < AVP Header: 1094 >
              { ADC-Rule-Name }
              [AA-Functions]
              *[ AVP ]
```

17.5.12.2 HTTP redirect override

HTTP redirect override submitted via the Gx interface overrides the current URL string defined in the filter that is currently applied to the subscriber-host. The override is implemented through the standard Redirect-Information AVP nested within the Charging-Rule-Definition (CRD) AVP.

IPv4:

```
Charging-Rule-Install
   Charging-Rule-Definition
```

```
      Charging-Rule-Name = v4-http-url:<name>
         Redirect-Information ::= < AVP Header: 1085 >
           [ [ Redirect-Address-Type ]
           [ [ Redirect-Server-Address ]
```

IPv6:

```
Charging-Rule-Install
   Charging-Rule-Definition
      Charging-Rule-Name = v6-http-url:<name>
         Redirect-Information ::= < AVP Header: 1085 >
           [ Redirect-Address-Type ]
           [ Redirect-Server-Address ]
```

The keywords **v4-http-url** and **v6-http-url** are special keywords that must be part of the Charging-Rule-Name (CRN) AVP. These keywords can be followed by an arbitrary string *name*.

The purpose of the <name> string in the CRN AVP is for the PCRF to differentiate between different HTTP redirect overrides. However, the *name* string in the context of the **http url host override** command in a filter has no meaning on the node, and therefore it is ignored. This means that there can be only one HTTP redirect override per host and per address family on a node.

The outcome of this Gx directive (Redirect-Information AVP without the Flow-Information AVP within the Charging-Rule-Definition AVP) is the override of the HTTP redirect URL in the currently applied subscriber-host filter. The filter definition must explicitly allow overrides via the **allow-radius-override** keyword.

As long as the override rule is present in the system (meaning, it has been submitted via the Gx and has not been explicitly removed since), the override tries to enforce itself when both of the following two conditions are met:

- A filter with the action **http-redirect** is applied to the subscriber-host. This is valid whether the filter was there when the rule was submitted, or whether the filter was applied after the rule was submitted.
- The filter definition allows **http-redirect** overrides (filter with the **allow-radius-override** keyword).

If the above conditions are not met, the override is accepted (the node responds with RAA=OK) and stored by the system, although it is not applied until the above conditions are met.

For the HTTP URL host, the CRD directive must not contain any flow information or any other action besides the Redirect-Information AVP. Otherwise, the Diameter encoding fails and an error response is generated for RAR while CCR-I is silently dropped.

17.5.12.3 Removal of overrides

With the exception of HTTP redirect override, overrides cannot be removed by the Charging-Rule-Remove AVP. They can only be overridden, and consequently the Charging-Rule-Remove AVP is ignored. It is ignored only for regular overrides and not for PCC rules (see PCC rules) or for HTTP redirect override. An HTTP redirect override can be removed whether it is active (a filter with HTTP redirect action is applied) or inactive (a filter without HTTP redirection is applied).

```
Charging-Rule-Remove ::= < AVP Header: 1002 >
   Charging-Rule-Name = v4-http-url:<name>
   Charging-Rule-Name = v6-http-url:<name>
```

The *name* string in the CRN AVP is ignored in the context of HTTP redirect override. This means that the removal of HTTP redirect override with any name removes the currently installed HTTP redirect override.

Similarly, the installation of the HTTP redirect override replaces any currently installed HTTP redirect override, regardless of the *name* string (implicit removal of the current HTTP redirect override, followed by the installation of the new one).

The node replies with RAA=OK if a properly formatted Charging-Rule-Remove directive with any name is received for HTTP redirect override.

17.5.12.4 Examples of Gx overrides

The instantiation of HTTP redirect overrides via the Gx can be summarized as:

- The Central AVP for the Gx overrides in SR OS is the Charging-Rule-Install AVP. Multiple overrides can be submitted to an node via a single Charging-Rule-Install AVP, or each Gx override can be submitted via its own Charging-Rule-Install AVP.
- Gx override is identified by the Charging-Rule-Name AVP. This AVP is also used to report on the status of Gx override. The Charging-Rule-Name can reference a pre-configured construct on the node (profiles, cat-maps, filters) or it can be assigned by PCRF to identify the PCRF defined override (QoS policy modifications, HTTP redirect AA ASO modifications, and so on).
- With the exception of the HTTP redirect override, overrides cannot be removed by the Charging-Rule-Remove AVP. They can only be overridden. The Charging-Rule-Remove AVP is ignored for Gx overrides.

The following is an example of Gx override instantiation where all Gx overrides are submitted under a single Charging-Rule-Install AVP. The AVPs in this example can be included in the CCA-i, CCA-u or RAR messages sent from the PCRF.

The outcome of the override is the following:

- The subscriber-id is overridden with the new name ‘residence-1’.
- New ingress v4 and egress v6 filters are installed.
- New sub-profiles and sla-profiles are applied to the subscriber host.
- The subscriber host is associated with the Vport named *vport-AN-1*.
- The existing HTTP redirect URL is overwritten with a new URL.
- Characteristics for queues 5 and 7 are overridden.
- The egress rate limit for the subscriber is overridden.
- ASOs are overridden.

```
Charging-Rule-Install ::= <AVP Header: 1001>
Charging-Rule-Name <AVP Header: 1005> = “sub-id:residence-1”
Charging-Rule-Name <AVP Header: 1005> = “ingr-v4:7”
Charging-Rule-Name <AVP Header: 1005> = “eggr-v6:5”
Charging-Rule-Name <AVP Header: 1005> = “Sub-Profile:prem”
Charging-Rule-Name <AVP Header: 1005> = “Sla-Profile:voip+data”
Charging-Rule-Name <AVP Header: 1005> = “Inter-Dest:vport-AN-1”

Charging-Rule-Definition <AVP Header: 1003>
Charging-Rule-Name <AVP Header: 1005> “v4-http-url:http-redir-1”
Redirect-Information < AVP Header: 1085 >
  Redirect-Address-Type < AVP Header: 433 > = 2 -> URL type
    Redirect-Server-Address < AVP Header: 435 > =
“http://www.operator.com/portal.php?mac=$MAC”
```

```
Charging-Rule-Definition <AVP Header: 1003>
Charging-Rule-Name <AVP Header: 1005> = “premium-service”
QoS-Information <AVP Header: 1016>
Alc-Queue <AVP Header; vnd ALU; 1016>
 Alc-Queue-id <AVP Header; vnd ALU; 1007> = 5
 Max-Requested-Bandwidth-UL <AVP Header: 516> = 10000
         Max-Requested-Bandwidth-DL <AVP Header: 515> = 100000
         Guaranteed-Bitrate-UL <AVP Header: 1026> = 5000
             Guaranteed-Bitrate-DL <AVP Header: 1027> = 50000
 Alc-Committed-Burst-Size-UL <AVP Header; vnd ALU; 1008> = 1000
     Alc-Maximum-Burst-Size-UL <AVP Header; vnd ALU; 1009> = 2000
     Alc-Committed-Burst-Size-DL <AVP Header; vnd ALU; 1010> = 1000
     Alc-Maximum-Burst-Size-DL <AVP Header; vnd ALU; 1011> = 2000

 Alc-Queue <AVP Header; vnd ALU; 1006>
Alc-Queue-id <AVP Header; vnd ALU; 1007> = 7
 Max-Requested-Bandwidth-UL <AVP Header: 516> = 10000
         Max-Requested-Bandwidth-DL <AVP Header: 515> = 100000
         Guaranteed-Bitrate-UL <AVP Header: 1026> = 5000
        Guaranteed-Bitrate-DL <AVP Header: 1027> = 50000
 Alc-Committed-Burst-Size-UL <AVP Header; vnd ALU; 1008> = 1000
     Alc-Maximum-Burst-Size-UL <AVP Header; vnd ALU; 1009> = 2000
     Alc-Committed-Burst-Size-DL <AVP Header; vnd ALU; 1010> = 1000
     Alc-Maximum-Burst-Size-DL <AVP Header; vnd ALU; 1011> = 2000

Alc-Sub-Egress-Rate-Limit <AVP Header; vnd ALU; 1016> = 10000



ADC-Rule-Install ::= <AVP Header: 1092>
ADC-Rule-Definition <AVP Header: 1094>
ADC-Rule-Name <AVP Header: 1096> = “AA-Functions:apps”
AA-Functions
AA-App-Profile-Name = “apps-prof”
AA-App-Service-Options
AA-App-Serv-Options-Name = “bitttorent”
AA-App-Serv-Options-Value = “low-prio-1mbps”
AA-App-Service-Options
AA-App-Service-Options-Name =  “ftp”
AA-App-Service-Options-Value = “hi-prio”
```

In the following example, all Gx overrides are submitted via a separate Charging-Rule-Install AVP:

```
Charging-Rule-Install ::= <AVP Header: 1001>
     Charging-Rule-Name <AVP Header: 1005> = “sub-id:residence-1”

Charging-Rule-Install ::= <AVP Header: 1001>
     Charging-Rule-Name <AVP Header: 1005> = “ingr-v4:7”

Charging-Rule-Install ::= <AVP Header: 1001>
     Charging-Rule-Name <AVP Header: 1005> = “eggr-v6:5”

Charging-Rule-Install ::= <AVP Header: 1001>
     Charging-Rule-Name <AVP Header: 1005> = “Sub-Profile:prem”

Charging-Rule-Install ::= <AVP Header: 1001>
     Charging-Rule-Name <AVP Header: 1005> = “Sla-Profile:voip+data”

Charging-Rule-Install ::= <AVP Header: 1001>
     Charging-Rule-Name <AVP Header: 1005> = “Inter-Dest:Gx-inserted-string”

Charging-Rule-Install ::= <AVP Header: 1001>
```

```
   Charging-Rule-Definition <AVP Header: 1003>
      Charging-Rule-Name <AVP Header: 1005> "v4-http-url:http-redir-1"
         Redirect-Information < AVP Header: 1085 >
            Redirect-Address-Type < AVP Header: 433 > = 2  # URL type
            Redirect-Server-Address < AVP Header: 435 > =
                 "http://www.operator.com/portal.php?mac=$MAC"
Charging-Rule-Install ::= <AVP Header: 1001>
     Charging-Rule-Definition <AVP Header: 1003>
          Charging-Rule-Name <AVP Header: 1005> = "premium-service"
          QoS-Information <AVP Header: 1016>
               Alc-Queue <AVP Header; vnd ALU; 1016>
                 Alc-Queue-id <AVP Header; vnd ALU; 1007> = 5
                          Max-Requested-Bandwidth-UL <AVP Header: 516> = 10000
                   Max-Requested-Bandwidth-DL <AVP Header: 515> = 100000
           Guaranteed-Bitrate-UL <AVP Header: 1026> = 5000
           Guaranteed-Bitrate-DL <AVP Header: 1027> = 50000
                          Alc-Committed-Burst-Size-UL <AVP Header; vnd ALU; 1008> = 1000
                    Alc-Maximum-Burst-Size-UL <AVP Header; vnd ALU; 1009> = 2000
                    Alc-Committed-Burst-Size-DL <AVP Header; vnd ALU; 1010> = 1000
              Alc-Maximum-Burst-Size-DL <AVP Header; vnd ALU; 1011> = 2000

               Alc-Queue <AVP Header; vnd ALU; 1006>
                          Alc-Queue-id <AVP Header; vnd ALU; 1007> = 7
                          Max-Requested-Bandwidth-UL <AVP Header: 516> = 10000
           Max-Requested-Bandwidth-DL <AVP Header: 515> = 100000
           Guaranteed-Bitrate-UL <AVP Header: 1026> = 5000
           Guaranteed-Bitrate-DL <AVP Header: 1027> = 50000
                          Alc-Committed-Burst-Size-UL <AVP Header; vnd ALU; 1008> = 1000
              Alc-Maximum-Burst-Size-UL <AVP Header; vnd ALU; 1009> = 2000
              Alc-Committed-Burst-Size-DL <AVP Header; vnd ALU; 1010> = 1000
              Alc-Maximum-Burst-Size-DL <AVP Header; vnd ALU; 1011> = 2000

               Alc-Sub-Egress-Rate-Limit <AVP Header; vnd ALU; 1016> = 10000

ADC-Rule-Install ::= <AVP Header: 1092>
     ADC-Rule-Definition <AVP Header: 1094>
          ADC-Rule-Name <AVP Header: 1096> = "AA-Functions:apps"
          AA-Functions
               AA-App-Profile-Name = "apps-prof"
               AA-App-Service-Options
                    AA-App-Service-Options-Name = "bitttorent"
                    AA-App-Service-Options-Value = "low-prio-1mbps"
               AA-App-Service-Options
                    AA-App-Service-Options-Name =  "ftp"
                    AA-App-Service-Options-Value = "hi-prio"
```

Gx overrides (QoS rates, sub/sla-profiles, filters, and so on) can be examined individually with subscriber specific operational commands. In the example below, fields in bold can be overridden.

```
show service active-subscribers detail
===============================================================================
Active Subscribers
===============================================================================
-------------------------------------------------------------------------------
Subscriber residence-1 (prem)
-------------------------------------------------------------------------------
I. Sched. Policy : basic_policy
E. Sched. Policy : N/A               E. Agg Rate Limit: 10000
I. Policer Ctrl. : N/A
E. Policer Ctrl. : N/A
Q Frame-Based Ac*: Disabled
Acct. Policy     : N/A               Collect Stats    : Disabled
```

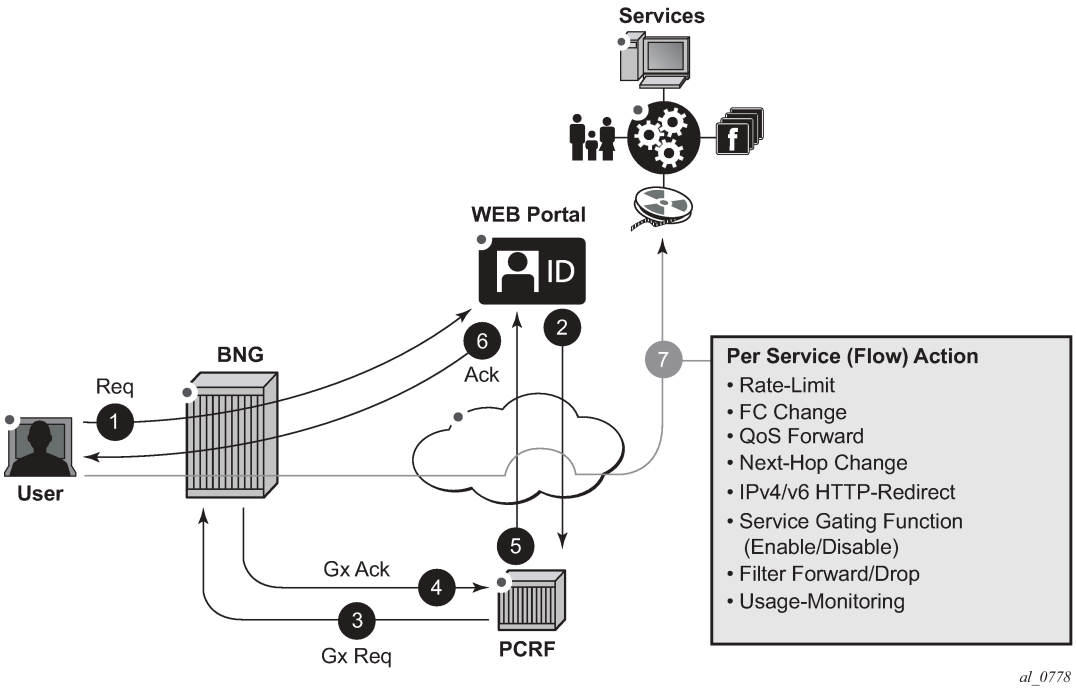
```
Rad. Acct. Pol.  : N/A
Dupl. Acct. Pol. : N/A
ANCP Pol.        : N/A
HostTrk Pol.     : N/A
IGMP Policy      : N/A
MLD Policy       : N/A
Sub. MCAC Policy : N/A
NAT Policy       : N/A
Def. Encap Offset: none             Encap Offset Mode: none
Avg Frame Size   : N/A
Vol stats type   : full
Preference       : 5
Sub. ANCP-String : "iope-left-dupl"
Sub. Int Dest Id : "Gx-inserted-string"
Igmp Rate Adj    : N/A
RADIUS Rate-Limit: N/A
Oper-Rate-Limit  : 10000


show service id 10 subscriber-hosts detail
===============================================================
Subscriber Host table
===============================================================
Sap                    Subscriber
  IP Address
    MAC Address          PPPoE-SID Origin       Fwding State
---------------------------------------------------------------
[1/1/5:6.2]            iope-sub
  192.168.0.11
    00:00:65:06:02:01    N/A       DHCP         Fwding
---------------------------------------------------------------
Subscriber-interface  : int1
Group-interface       : g1
Sub Profile           : prem
SLA Profile           : voip+data
App Profile           : apps-prof
Egress Q-Group        : N/A
Egress Vport          : N/A
Acct-Session-Id       : D896FF0000001852EDFF46
Acct-Q-Inst-Session-Id: D896FF0000001952EDFF51
Address Origin        : DHCP
OT HTTP Rdr IP-FltrId : N/A
OT HTTP Rdr Status    : N/A
OT HTTP Rdr Fltr Src  : N/A
HTTP Rdr URL Override : N/A
GTP local break-out   : No
```

17.5.13 PCC rules

A generic use case for flow-based dynamic policy is related to customized network level treatment of on-demand services. Such services can represent a wide range of applications, such as video-on-demand or access to a specific application in the network. The service can be identified by traffic destination parameters or DSCP bits. After the service is identified, a set of actions can be applied to the service (rate change, forwarding-class change, Usage-Monitoring, and so on).

Typical flow of events for service activation is shown in Figure 218: Generic use case for IP criterion-based policy change via Gx:

Figure 218: Generic use case for IP criterion-based policy change via Gx



1) An established user subscribes to a service in the network via a Web portal at any time.

2) After the authentication/payment is accepted, the back end (for example, the Web portal integrated in OSS) identifies the service and submits the parameters defining the network delivery of the offered service to the PCRF.

3) The PCRF converts those parameters into rules and submits those rules to the subscriber-host on the BNG via the Gx. The rules identify the service on the network level (destination IP@ and port) along with the needed action.

4, 5, and 6) Before the service can be started, the action of individual policy management elements must be acknowledged to ensure that the resources for the service delivery are available and instantiated before the service is delivered to the subscriber.

7) The service traffic can be started from the subscriber side. Network requirements for the successful service delivery are enforced on a per flow or DSCP basis as defined by the PCC rule.

17.5.13.1 PCC rule concept

A PCC rule consists of traffic classifiers (Flow-Information AVPs) required for traffic identification, and one or more actions associated with such classified traffic. PCC rules are unidirectional, which means that each rule is applied on ingress or egress. They are provisioned from PCRF via Gx interface.

Traffic classification is based on:

- 5 Tuple (IPv4 and IPv6)

- DSCP bits. When the content hosting device cannot be identified by the IP address, port and protocol, the DSCP marking can be used instead. Then, the DSCP marking is set by the client application and the markings should be preserved throughout the network until they reach the BNG.

Supported actions are:

- rate limiting (in | out)
- forwarding class (FC) change (in | out)
- QoS forward (in | out)
- next hop redirect (in)
- service ID redirect (in)
- HTTP redirect (in)
- service gating function (in | out)
- filter forward/drop (in | out)
- Usage-Monitoring (in | out)

17.5.13.2 PCC rule instantiation

A PCC rule that is submitted to the node via PCRF is internally instantiated using two basic policy constructs, QoS policy and filter policy (ACL). This internal division is transparent to the operator at the time of the rule provisioning. The operator perceives Gx as a unified method for provisioning policy rules, whether the rule is QoS-related or filter-related.

The type of action within the PCC rule determines whether the PCC rule is split between the QoS policy and the filter policy.

Rules with actions:

- rate limit
- forwarding class change
- QoS forward; matching traffic is forwarded without QoS actions and do not match on the next entry (match and exit behavior). This is equivalent to an allowlist entry.
- Usage-Monitoring

are converted into a QoS policy while the rules with actions:

- next hop redirect
- service ID redirect
- HTTP redirect
- gate function (enable/drop)
- filter forward/drop

are converted into filter rules.

The operator should be aware of this division for dimensioning (scaling) purposes. Operational commands can be used to reveal resource consumption on the node.

A PCC rule is addressed to a subscriber-host (single stack or dual stack) via the diameter session-id. However, qos-policy-related entries are applied per sla-profile instance because the qos resources are allocated per sla-instance. An sla-profile instance and sla-profile are two distinct concepts. An sla-profile

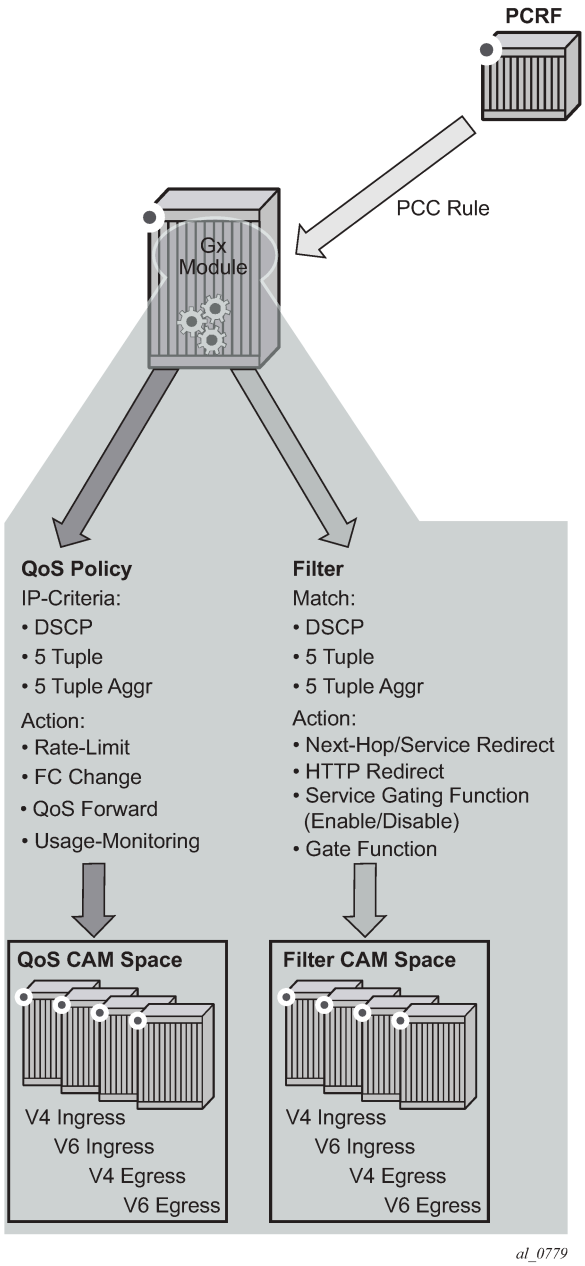
instance is an instantiation of the sla-profile which is a configuration concept in which parameters are defined. An sla-profile is instantiated per a subscriber-host, or multiple subscriber-hosts can share an sla-profile instance as long as they belong to the same SAP and have the same subscriber ID.

This means that all hosts sharing the same sla-profile instance inherits the change. The **sla-profile** *instance* and **sla-profile** are two distinct concepts. The **sla-profile** *instance* is an instantiation of the **sla-profile** which is a configuration concept in which parameters are defined. The **sla-profile** is instantiated per a subscriber-host, or multiple subscriber-hosts can share an **sla-profile** *instance* as long as they belong to the same SAP and have the same subscriber-id.

Filter-related entries are applied per each subscriber-host, whether the hosts are sharing or not sharing an sla-profile instance.

The concept of splitting the rules is shown in Figure 219: PCC rule conversion on the node.

Figure 219: PCC rule conversion on the node



The PCC rule instantiation fails if a PCC rule contains only actions without any classification, or if it contains only classification without any actions.

17.5.13.3 Base QoS-policy and base filter

Subscriber host must have an explicit static (or base) filter or qos-policy before any dynamic entries can be inserted via Gx. For example, a base filter/qos-policy can be referenced by a sla-profile when

the subscriber is instantiated. However, the parameters in the base qos-policy and base filter cannot be modified via Gx.

In the absence of explicitly defined qos-policy for the subscriber host, the default **qos-policy 1** is in effect. Then, PCC rules with a QoS-related action cannot be applied.

PCC rule entries can be inserted in specifically allocated range in the base filter or qos-policy. The insertion point is controlled by the operator. This is shown in Figure 220: Static and dynamic QoS-policy/filter entries. The entries reserved for PCC rules start at the beginning of the range specified by the following CLI command:

```
sub-insert-shared-pccrule start-entry <entry-id> count <count>
```

under the following CLI hierarchy:

```
config>filter>ip-filter>
config>filter>ipv6-filter
config>qos>sap-ingress>
config>qos>sap-egress>
```

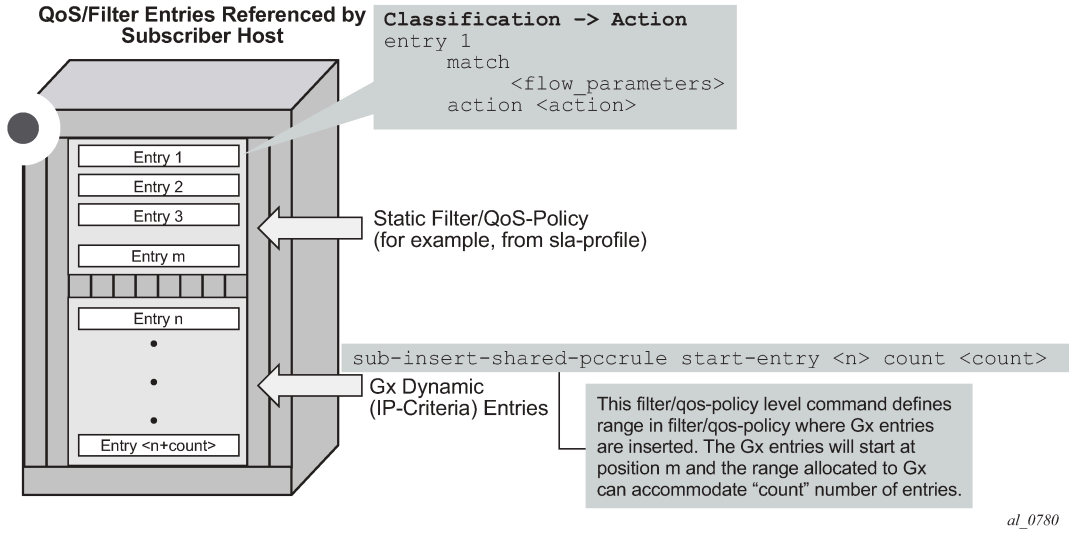
An entry corresponds to a Flow-Information AVP and is equivalent to a match condition defined as any combination of the following parameters under a filter or **qos-policy ip-criteria**:

- source IP address
- destination IP address
- source port or port range
- destination port or port range
- protocol
- DSCP

Such defined entry maps into a single CAM entry with exception of port range configured as match criteria whereby a single port range command can expand into multiple CAM entries.

Static entries in filter/qos-policy can be inserted before and after the range reserved for PCC rules.

Figure 220: Static and dynamic QoS-policy/filter entries

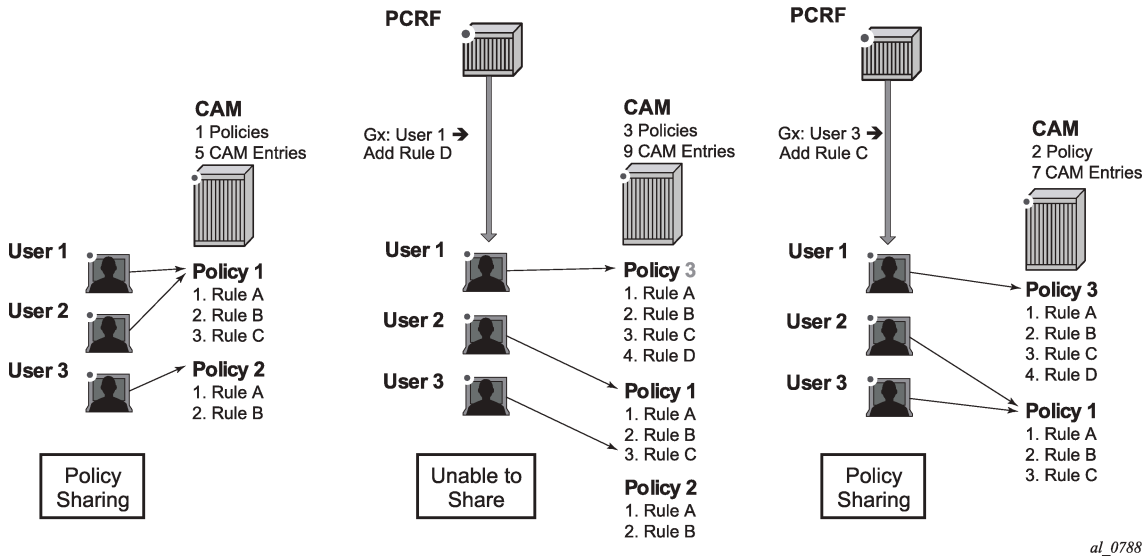


17.5.13.4 Generic policy sharing and rule sharing

Policy (defined in this context as a collection of static and dynamic rules) sharing between the subscriber hosts is depicted in Figure 221: Policy cloning. To simplify CAM scaling explanations, the examples in this section assume that one rule within the policy occupies exactly one CAM entry. For simplicity, only PCC rules are shown but in reality a subscriber-host policy consist of PCC rules together with the base qos-policy/filter.

A policy, as a set of rules, can be shared amongst the subscriber-hosts. However, when a new rule is added to one of the subscriber-host, the newly created set of rules for this host becomes unique. Hence, a new policy for the subscriber-host is instantiated. This new policy consumes additional resources for all the old rules (clone of the old policy) along with the new rule. Figure below shows that a new policy (3) is instantiated when rule D is added to User 1, even though the rules A, B and C remain the same for Users 1 and 2. Policy 3 is a newly cloned with the same rules as Policy 1, and then Rule D is added onto it. On the other hand, when the rule C is applied to User 3, the set of rules becomes identical to the set of rules for User 2. Thus the two can start sharing rules and therefore the resources are freed.

Figure 221: Policy cloning



17.5.13.5 PCC rule name and PCC rule removal

Each PCC rule has a subscriber-host scope and it is referred to it by its name which is assigned by the operator on PCRF. The rules with exactly the same content but different rule names are evaluated into separate rules. To optimize performance and maximize scale, it is recommended that the rules sharing the same content have the same name (as provisioned in the PCRF).

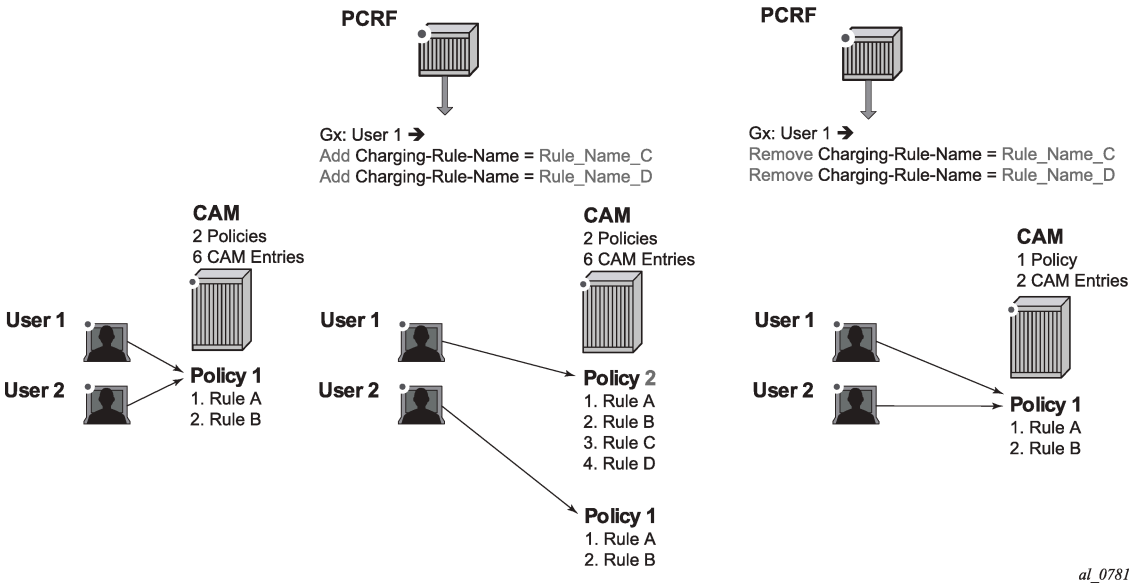
PCC rules can be removed from the node via a Gx directive by referencing the PCC rule name. The rule name is supplied via the Charging-Rule-Name AVP at the time of the rule submission to the node by the PCRF. There is no Gx mechanism that would remove all PCC rules at once. Each PCC rule must be removed individually.

The AVP used to remove the rule from the node is:

```
Charging-Rule-Remove ::= < AVP Header: 1002 >
                                  *[ Charging-Rule-Name ]
```

An example of rule instantiation and rule removal is shown in Figure 222: Policy removal by name.

Figure 222: Policy removal by name



17.5.13.6 Gx rule ordering

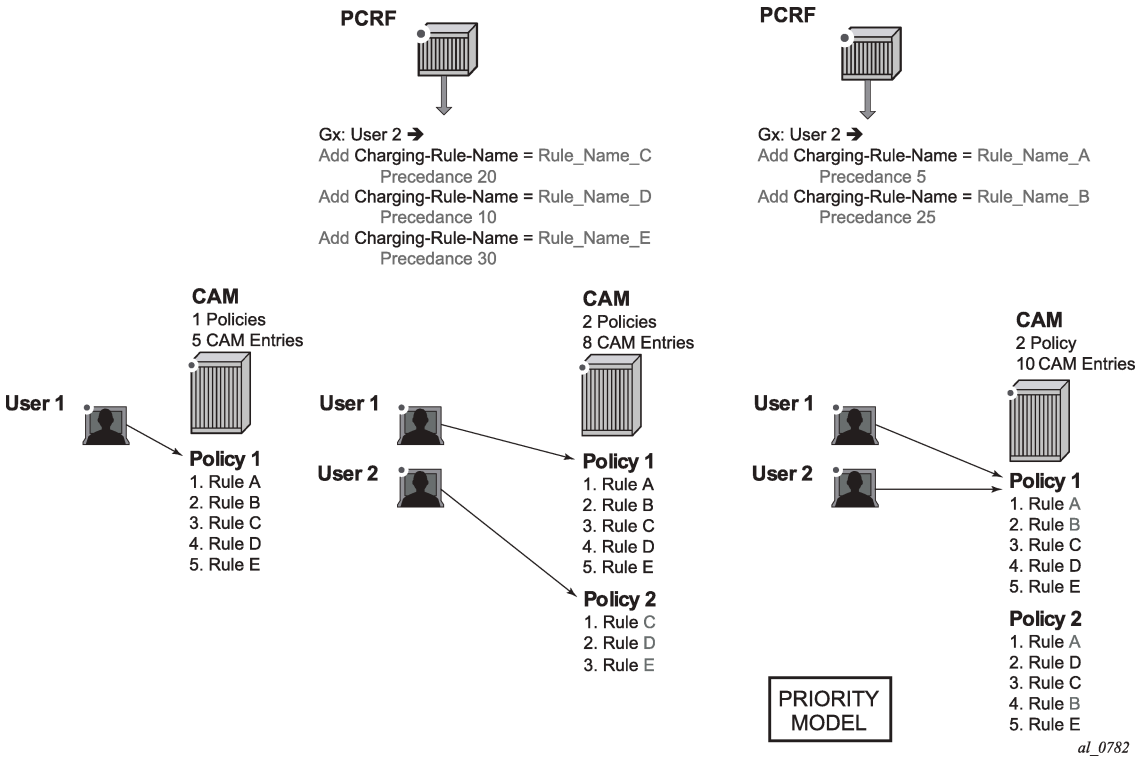
Entries in IPv4/v6 filter and QoS policy created via CLI are ordered according to the numerical value associated with each entry command (which corresponds to the match condition) within the policy. CLI rules can be re-ordered with the **renum** command (in filters and QoS policies).

On the other hand, the PCC rules are ordered in one of the two ways. The difference between ordering of the entries with the rules, and the ordering of the rules themselves is:

- PCC rules are prioritized according to the Precedence AVP within the Charging-Rule-Definition AVP. They are inserted within the subscriber-host policy, according to the Precedence AVP relative to the other PCC rules already applied to the subscriber-host. A PCC rule with a lower Precedence value is applied before a rule with a higher Precedence value.

  The ordering behavior according to the Precedence is shown in Figure 223: PCC rule ordering - priority model.

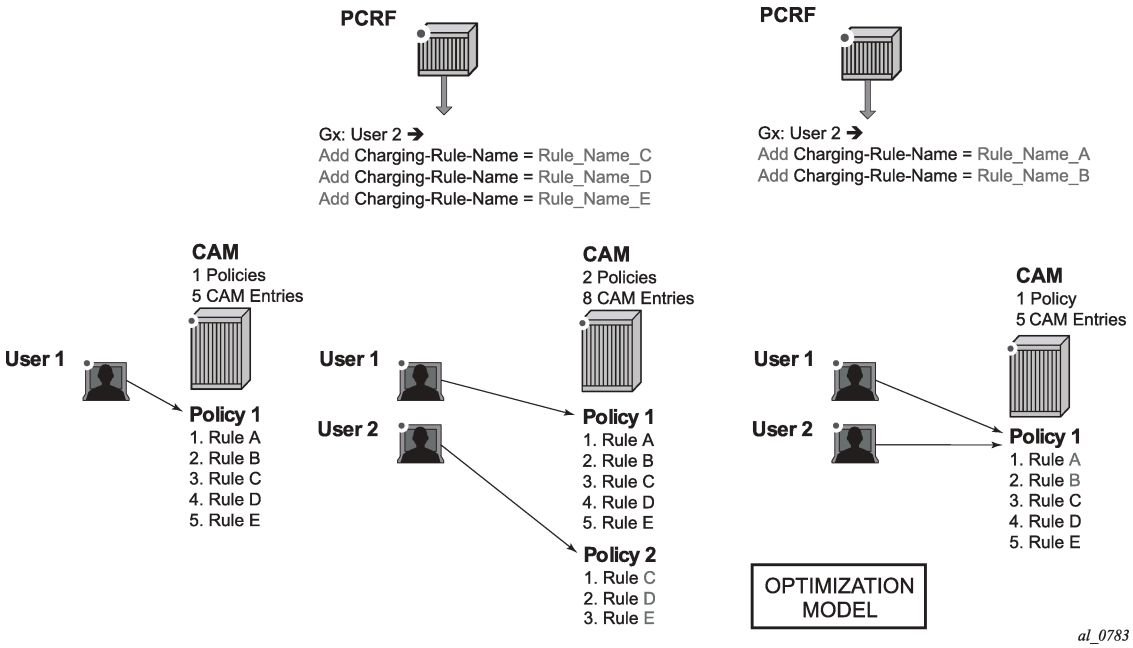
Figure 223: PCC rule ordering - priority model



- Automatic optimization of PCC rules. Automatic optimization is used in cases where the PCC rule order is not important to the operator. Then, the node optimizes the rule ordering to achieve the best possible scale by maximizing rule sharing. This optimization (or internal rule ordering) is performed for PCC rules without the Precedence AVP, or for PCC rules with the same Precedence value.

  The premise of the automatic rule optimization is shown in Figure 223: PCC rule ordering - priority model.

Figure 224: Rule ordering - optimization model



The ordering of PCC rules has no effect on the ordering of the static entries in the base qos-policy or filter.

Mixing of the PCC rules with the Precedence AVP and without Precedence AVP is allowed for the same subscriber-host. PCC rules without the Precedence AVP are inserted at the end of all PCC rules that do have the Precedence value set explicitly. In other words, the Precedence value for PCC rules without the explicitly configured Precedence AVP is assumed to be the highest. The PCC rules without the Precedence value are automatically inserted at the bottom of the PCC rule range.

A distinction should be made between the order of PCC rules in a PCC rule set and the order between the entries within each PCC rule. A PCC rule contains a group of classifiers that are all associated with the same actions. Therefore, the order of the entries (equivalent to match conditions) within any specific PCC rule does not matter (all entries result in the same action). For this reason, PCC rules with identical name and identical entries but different order of the entries are automatically ordered in a way that would allow more optimal sharing of the rules between different subscribers.

17.5.13.7 PCC rule override

A PCC rule applied to a subscriber-host on the node can be overridden by re-submitting the PCC rule with the same name but different contents.

If at least one new flow is sent in the PCC rule update, then the existing flows are removed and replaced with the new flow. If no new flows are submitted, then the existing flows stay in place.

If there are conflicting parameters between the existing rule and the modified rule (for example the combination of the unsupported actions), the PCC rule override fails.

17.5.13.8 Aggregation of IP-criterion

An action with a PCC rule can be applied for a set of IP-criterion.

For example, a single policer can be instantiated for a set of flows for rate-limiting purposes.

A pseudo Gx directive would look like this:

```
Charging-Rule-Install – Directive to install the rule in 7750 SR
          Charging-Rule-Definition – PCC rule definition created on PCRF
               Charging-Rule-Name = Rule-1 – PCC rule name
                    Flow1     – match-criteria for flow 1
                    Flow2     – match-criteria for flow 2
                    Flow 2    – match-criteria for flow 3
                    Rate-limit    – rate-limiting action applicable as an aggregate action for
 all 3 flows
```

All three flows are fed into the same rate limiter (policer).

17.5.13.9 Combining IPv4 and IPv6 entries within the rule

IPv4 flow entries and IPv6 flows entries can be combined within the same PCC rule. The actions that carry the IP address are address-type-specific (for example **next-hop-redirect**). All other actions (rate-limit, FC change, and so on) are universal and it are applied to both flow types (IPv4 and IPv6). The node automatically sorts out flow types (IPv4 and IPv6) within the rule and apply corresponding actions.

If the rule contains a mismatching flow type and actions (for example, IPv4 flows and IPv6 specific actions), the rule is rejected. It is the operator's responsibility to ensure that the address-type-specific actions in the rule have corresponding flows to which they can be applied.

17.5.13.10 Gx rules with multiple actions and action sharing

PCC rules can contain multiple actions. For the list of support action combinations, see the PCC rules.

17.5.13.11 Alc-NAS-Filter-Rule-Shared AVP vs Flow-Information AVP

A Gx rule (as defined in a single Charging-Rule-Definition AVP) can contain either Flow-Information AVP or Alc-NAS-Filter-Rule-Shared AVP, but not both simultaneously.

Presence of either AVP within the Charging-Rule-Definition AVP determines the mode of operation for the rule:

Alc-NAS-Filter-Rule-Shared AVP indicates the mode of operation in which the permit or deny action is part of the flow definition itself (Alc-NAS-Filter-Rule-Shared AVP). This mode of operation is referred as NAS filter inserts. The basic format of the AVP is the following (RFC 4849 and 4005; AVP Code 400):

```
<action> <direction> <protocol>  from <source> to <destination> <options>.
```

There can be multiple ip-criteria definitions within the rule per subscriber-host, and each ip-criteria carries its own permit/deny action. There can be only one such rule (Charging-Rule-Definition) per subscriber-host. The rule entries are installed within the filter range defined by the following command:

```
sub-insert-shared-radius start-entry <entry-id> count <count>
under the following CLI hierarchy:
```

```
config>filter>ip-filter>
config>filter>ipv6-filter
```

Such rule cannot be removed by the Charging-Rule-Remove directive referencing the rule name. Instead, each such Gx rule overwrites the previous one.

Flow-Information AVP indicates the mode of operation whereby all the flows in the rule share the same actions carried in separate AVPs. This mode of operation is referred to as PCC rule inserts. The rule entries are installed within the filter or qos-policy range defined by the following command:

```
sub-insert-shared-pccrule start-entry <entry-id> count <count>
under the following CLI hierarchy:
config>filter>ip-filter>
config>filter>ipv6-filter>
config>qos>sap-ingress>
config>qos>sap-egress>
```

There can be multiple flow based rules present in an orderly fashion and each rule can be individually removed by referencing its name.

Both modes of operation are supported simultaneously for the subscriber host.

17.5.13.12 RADIUS and Gx interaction

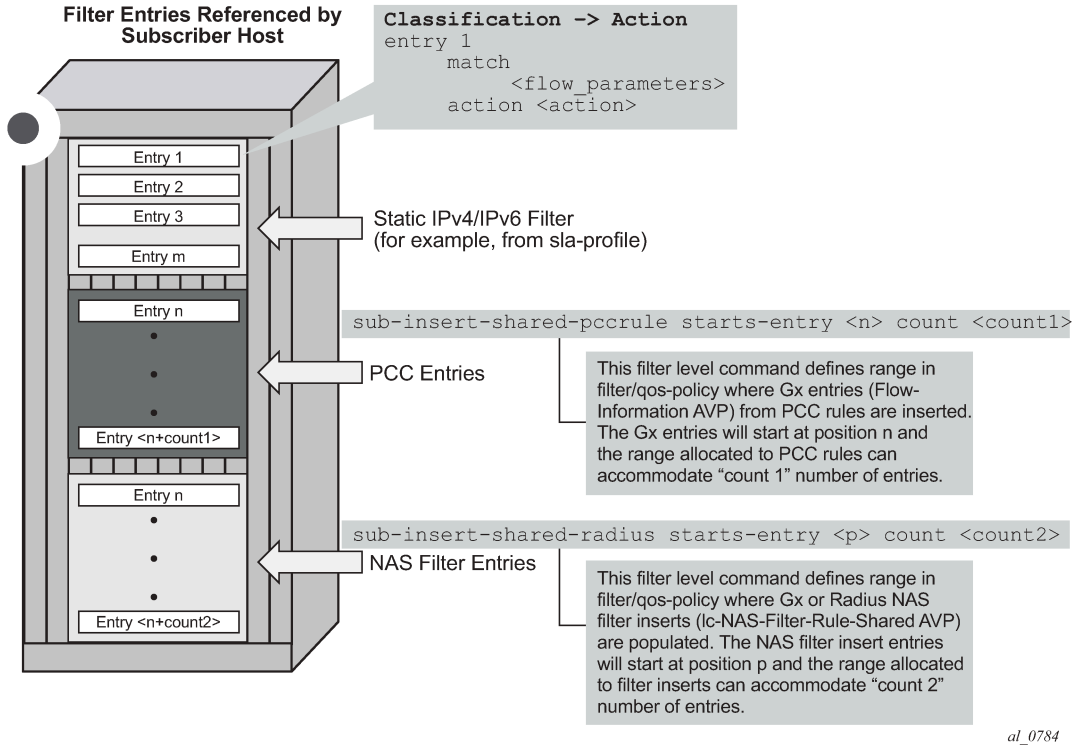
Gx and RADIUS (CoA) policy management interfaces are simultaneously supported for the same subscriber-host.

RADIUS and Gx share the same entries for filter entry inserts (NAS-Filter-Rules and Alc-NAS-Filter-Rule-Shared) and therefore the most recent insert overrides the previous one. Similar logic applies to subscriber-string overrides and QoS overrides, where the most recent source, overrides the previous one.

However, PCC rules (IP-criteria based Gx rules) are provisioned in a separate filter 'entry' space from RADIUS and Gx filter inserts and therefore the PCC rules and RADIUS/Gx based filter inserts can independently coexist.

Filter/QoS-policy entry order is shown in Figure 225: CAM table population. The order of configuration blocks (static, PCC rules or NAS filter inserts) is configurable. For example, an operator can specify that static filter entries are populated before PCC rules which are then populated before NAS filter inserts.

Figure 225: CAM table population



17.5.13.13 Bulk changes while Gx rules are active

When PCC rules are applied to a subscriber-host, the operator can modify some of the CLI parameters in the base QoS or filter policies. For example, the operator can add and remove terms in the base ACL filter.

Table 65: CLI modifiable parameters in base QoS policy that contains clones lists the parameters in the QoS policy that can be changed. Adding or removing queue and policer, re-mapping of FC, modifying the DSCP map or adding or removing static IPv4 or IPv6 criteria entries with action different from FC, profile, or priority is not allowed.

Modified parameters in the base QoS policy or filter referenced in the sla-profile affects all subscribers using this sla-profile. Replacing the base QoS or filter policy in sla-profile is not allowed for any subscriber-host if a clone of the base QoS or filter policy exist anywhere in the system.

However, replacing the base filter ID for a host using CoA or Gx override is allowed. Then, only the targeted host is affected and all existing PCC rules for this host are merged with the new filter.

Table 65: CLI modifiable parameters in base QoS policy that contains clones

CLI
config>qos>sap-ingress>queue
[no] cbs - Specify CBS
drop-tail low [no] percent-reduction-from-mbs - Specifies the drop tail for out-of-profile packets

CLI
[no] mbs - Specify MBS
[no] packet-byte-of* - Specify packet byte offset
[no] parent - Specify the scheduler to which this queue feeds
[no] percent-rate - Specify percent rates (CIR and PIR)
[no] rate - Specify rates (CIR and PIR)
config>qos>sap-egress>queue
[no] cbs - Specify CBS rate
drop-tail low [no] percent-reduction-from-mbs - Specifies the drop tail for out of profile packets
drop-tail exceed [no] percent-reduction-from-mbs - Specifies the drop tail for exceed profile packets
drop-tail high [no] percent-reduction-from-mbs - Specifies the drop tail for in profile packets
drop-tail highplus [no] percent-reduction-from-mbs - Specifies the drop tail for inplus profile packets
[no] mbs - Specify MBS rate
[no] parent - Specify the scheduler to which this queue feeds
[no] percent-rate - Specify percent rates (CIR and PIR)
[no] port-parent - Specify the port-scheduler to which this queue feeds
[no] rate - Specify rates (CIR and PIR)
[no] packet-byte-of* - Specify packet byte offset
config>qos>sap-ingress>policer
[no] cbs - Specify CBS
[no] high-prio-only - Specify high priority only percent-of-mbs
[no] mbs - Specify MBS
[no] packet-byte-of* - Specify packet byte offset
[no] parent - Specify the arbiter to which this policer feeds
[no] percent-rate - Specify percent rates (CIR and PIR)
[no] rate - Specify rates (CIR and PIR)
config>qos>sap-egress>policer
[no] cbs - Specify Cbs
[no] high-prio-only - Specify high priority only percent-of-mbs

CLI
[no] mbs - Specify Mbs
[no] packet-byte-of* - Specify packet byte offset
[no] parent - Specify the scheduler to which this policer feeds
[no] percent-rate - Specify percent rates (CIR and PIR)
[no] rate - Specify rates (CIR and PIR)
config>qos>sap-ingress>ip-criteria>entry>match
config>qos>sap-egress>ip-criteria>entry>match
config>qos>sap-ingress>ipv6-criteria>entry>match
config>qos>sap-egress>ipv6-criteria>entry>match

17.5.13.14 PCC rule direction

PCC rules are unidirectional. The PCC rule direction is determined based on the value of the Flow-Direction AVP within the Flow-Information AVP. In the absence of the Flow-Direction AVP, the PCC rule direction is determined based on the Flow-Description AVP (as part of IPFilterRule direction field). Both of these AVPs (Flow-Direction and Flow-Description) are part of the PCC rule definition.

If the action within the PCC rule is in conflict with the direction of the flow, the PCC rule instantiation fails. For example, an error is raised if the flow direction is UPSTREAM, while the action is 'Max-Requested-Bandwidth-DL' (downstream bandwidth limit).

17.5.13.15 Action

A PCC rule may contain multiple actions. Each action is carried in a separate, action specific AVP. The action specified in the **flow-description>ipfilter rule** data type is ignored. If rule contains multiple instances of the same action, each with a different value, the last occurrence of the action value is in effect.

Not all of the action types can be applied at the same time. The allowed combination of the actions per direction is described in Table 65: CLI modifiable parameters in base QoS policy that contains clones and Table 67: Failure reporting .

17.5.13.16 Rate-limiting action (ingress, egress)

Rate-limiting action is implemented via policers. The policer is dynamically created at the PCC rule instantiation time. The rate can be enforced based on Layer 2 rates or Layer 3 rates.

Dynamically instantiated policers have their own policer ID range to avoid the conflict with static policers.

The dynamically created policers shares common properties configured under the dynamic-policer CLI hierarchy:

```
configure
   qos
```

```
      sap-ingress/egress
         dynamic-policer
            stat-mode <stat_mode>
            parent <arbiter-name> [weight <weight-level>] [level <level>
            mbs <size> [bytes | kilobytes]
            cbs  <size> [bytes | kilobytes]
            packet-byte-offset {add <add-bytes> | subtract <sub-bytes>}
            range start-entry <entry-id> count <count>
```

The configured dynamic policer parameters can be overridden per PCC rule by including the Alc-Dynamic-Policer grouped AVP in the QoS-Information AVP. All AVPs are optional and when specified override the configured value:

- ingress PCC rules

```
+--1046 Alc-Dynamic-Policer
   +--1039 Alc-Policer-Parent
      +--1022 Alc-Arbiter-Name
      +--1040 Alc-Parent-Level
      +--1041 Alc-Parent-Weight
   +--1008 Alc-Committed-Burst-Size-UL
   +--1009 Alc-Maximum-Burst-Size-UL
   +--1042 Alc-Stat-Mode-UL
   +--1044 Alc-Packet-Byte-Offset-UL
```

- egress PCC rules

```
+--1046 Alc-Dynamic-Policer
   +--1039 Alc-Policer-Parent
      +--1022 Alc-Arbiter-Name
      +--1040 Alc-Parent-Level
      +--1041 Alc-Parent-Weight
   +--1010 Alc-Committed-Burst-Size-DL
   +--1011 Alc-Maximum-Burst-Size-DL
   +--1043 Alc-Stat-Mode-DL
   +--1045 Alc-Packet-Byte-Offset-DL
```

The policer rates are part of PCC rule itself and are not part of static configuration.

The generic Gx directive for rate-limiting action is:

```
Charging-Rule-Install ::=      <AVP Header: 1001>
     Charging-Rule-Definition <AVP Header: 1003>
          Charging-Rule-Name <AVP Header: 1005>
          QoS-Information <AVP Header: 1016>
               Max-Requested-Bandwidth-UL <AVP Header: 516> [bps] 3GPP 29.214 §5.3.15
        Max-Requested-Bandwidth-DL <AVP Header: 515> [bps] 3GPP 29.214 §5.3.14
        Guaranteed-Bitrate-UL <AVP Header: 1026> [bps] 3GPP 29.214 §5.3.26
        Guaranteed-Bitrate-DL <AVP Header: 1027> [bps] 3GPP 29.214 §5.3.25
```

The above rate limits refer to PIR and CIR rates of the dynamic policer in the respective direction.

17.5.13.16.1 Dynamic policers and queue mappings

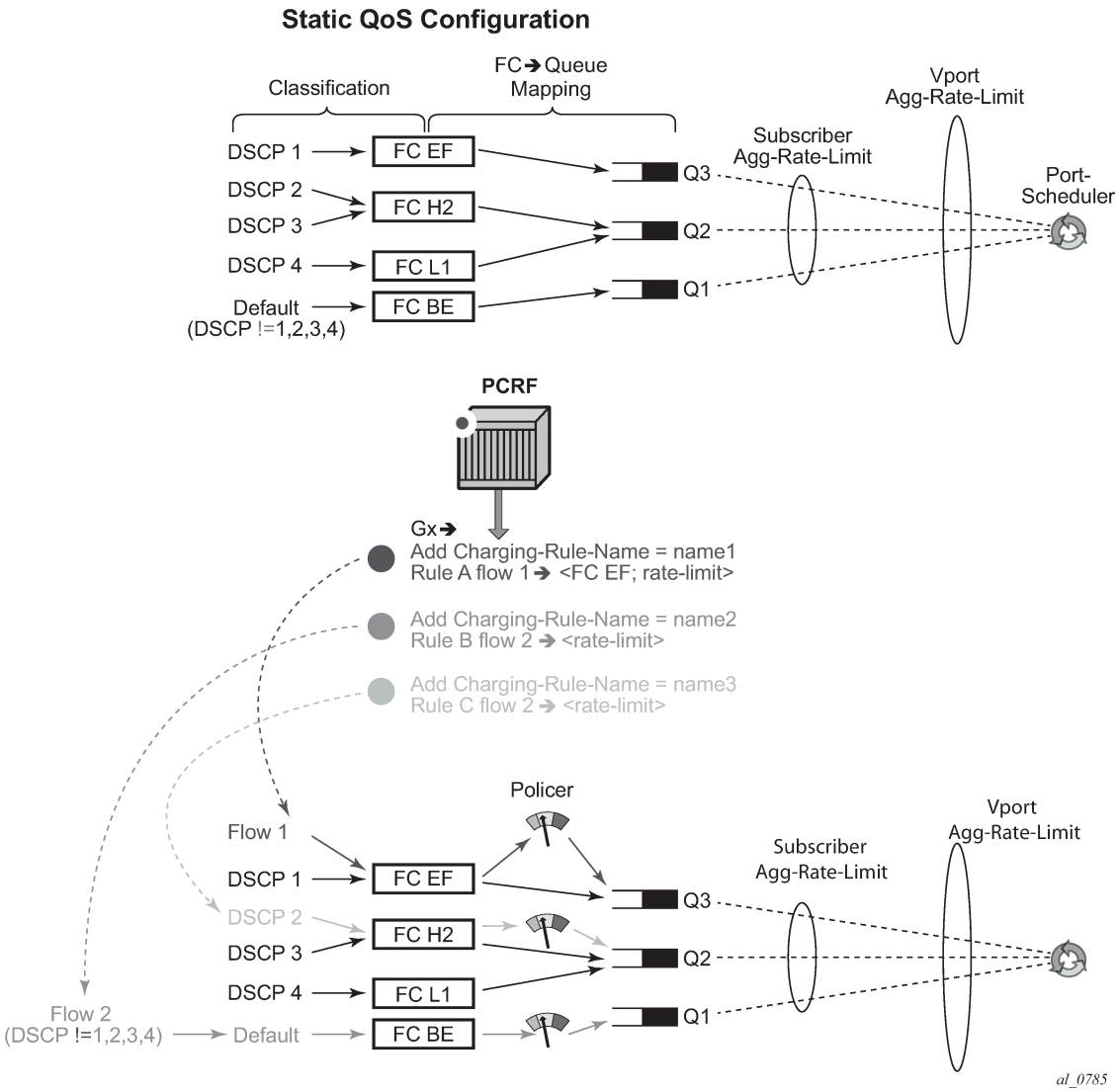
Once traffic is processed by the dynamic policers on ingress, the traffic flows through the policer-output-queues shared queues. Traffic through dynamic policers always bypass subscriber queues or policers on ingress that are statically configured in the base QoS policy.

Similar behavior is exhibited when static policers are configured on egress. Traffic outputting dynamic policer is never mapped to another static policer. Instead, such traffic is mapped to the corresponding shared queue in a queue-group. By default, this queue-group is the policer-output-queue group. However, the selection of the queue-group is configurable.

In contrast to the above, traffic processed by dynamic policers can be fed into statically configured subscriber (local) queues on egress. Dynamic policers and subscriber queues are tied through the forwarding-class.

The policer to local queue mapping and inheritance of the forwarding-class is shown. In this example, the mapping of traffic —> forwarding-class in rule 2 (flow 2) depends on the DSCP bits in the traffic flow. If the DSCP value in this traffic flow are different from the explicitly configured DSCP values in the static (base) QoS policy, then traffic is mapped to the default forwarding-class.

Figure 226: FC inheritance for dynamically instantiated policers on egress



17.5.13.16.2 Dynamic policer rates and accounting statistics

By default, policer rates are configured based on Layer 2 frame length (for example, the Ethernet header plus the IP packet). This can be changed by the **packet-byte-offset** (PBO) command under the policer. If the policer is fed into a local queue, the PBO of the policer does not affect the PBO of the local queue it is feeding.

The rates for local subscriber queues can be independently measured based on Layer 2 or Layer 1 frame length and the queue statistics can be measured based on Layer 1, Layer 2 or Layer 3 (IP-only) frame length. The IP-only stats for queues can be configured in the **sub-prof>volume-stats-type {ip|default}**.

Dynamic policer (instantiated because of rate limiting or usage-monitoring action in PCC rules) statistics are not reported in RADIUS-based accounting. On egress, this has no effect on volume counters in RADIUS-based accounting, because the dynamic policers are normally fed into local queues whose statistics are reported in RADIUS-based accounting. However, on ingress, the dynamic policers are always fed into the queue-group queues which are excluded from RADIUS based accounting. The consequence is that the ingress RADIUS-based accounting lacks statistics for the traffic that is flowing via dynamic policers.

If the dynamic policer is feeding a local queue, the aggregate statistics in show commands for such queue are not reported to avoid double counting (because the traffic statistics in **show** commands are already reported for the dynamic policer). However, the per-queue statistics are reported in **show** commands, irrespective of whether the policer is mapped to the local queue or not.

To avoid losing aggregate SAP or subscriber stats in show commands, the recommendation is to have policers feed into local queues which are not already mapped to an FC. For example:

```
queue 4 create //Not counted since policer 2 is feeding it
exit
policer 2 create
exit
fc be create
     queue 4     //Not counted
exit
fc l1 create
     queue 4     //Not counted
exit
fc ef create
     policer 2 queue 4
exit
```

FC BE, FC2 L1 —> queue 4

FC EF —> policer 2, queue 4

Then, traffic from queue 4 is not counted in aggregate stats at all and consequently the aggregate accounting information is lost for FC BE and FC L1.

17.5.13.17 Forwarding-class change (ingress, egress)

Traffic can be re-prioritized via PCC rule by re-classification into a different forwarding class. The forwarding-class can be changed in several cases.

The original static mapping between traffic type, forwarding-class and the queue/policer in the base qos-policy is configured outside of the ip-criteria CLI hierarchy.

For example:

```
    config>qos>sap-egress#
         dscp af11 fc "af"
```

Such mapping is configured outside of CAM and therefore it has lower evaluation priority than the mapping configured via PCC rule which is installed in CAM.

The original static mapping is provisioned in the base qos-policy via ip-criteria CLI hierarchy.

For example:

```
    config>qos>sap-egress>#
          ip-criteria
               entry 40000 create
                    match
                         dscp af11
                exit
           action fc "af"
     exit
    exit
```

Then, the configured entry range for PCC rules must precede the static entry (match criteria) in which the original forwarding-class is configured. The insertion point (entry) is controlled via configuration: **sub-insert-shared-pccrule start-entry** <*entry-id*> **count** <*count*> command under the qos-policy.

In both of the above cases, the following PCC rule would override forwarding-class for traffic with DSCP value of 10 (af11 traffic class) from value af to h2.

```
Charging-Rule-Install
     Charging-Rule-Definition
          Charging-Rule-Name = fc-change
          Flow-Information
               ToS-Traffic-Class = 00101000 11111100
               Flow-Direction = 1
          QoS-Information
               QoS-Class-Identifier  = 2
```

The eight forwarding classes are mapped to QCIs (3GPP TS 23.203 §6.1.7.2) in the manner displayed in Table 66: Forwarding class and QCI mapping.

Table 66: Forwarding class and QCI mapping

Forwarding class	QCI
BE	QCI 8
L2	QCI 7
AF	QCI 6
L1	QCI 4
H2	QCI 2

Forwarding class	QCI
EF	QCI 3
H1	QCI 1
NC	QCI 5

The generic Gx directive for forwarding-class change:

```
Charging-Rule-Install ::=      <AVP Header: 1001>
     Charging-Rule-Definition <AVP Header: 1003>
          Charging-Rule-Name <AVP Header: 1005>
               QoS-Information <AVP Header: 1016>              3GPP 29.212 §5.3.16
                    QoS-Class-Identifier <AVP Header: 1028>     3GPP 29.212 §5.3.17
```

17.5.13.18 QoS forward (ingress and egress)

Create an **ip-criteria** or **ipv6-criteria** entry with no action specified. Matching traffic is forwarded without a QoS action and not match on a next entry (match and exit behavior). This is equivalent to an allowlist entry. CLI equivalent:.

```
config>qos
  sap-ingress | sap-egress <id> create
    ip-criteria | ipv6-criteria
      entry <id> create
        match
          <5-tuple | dscp>
        exit
        action
      exit
    exit
```

The generic Gx directive for QoS forwarding:

```
Charging-Rule-Install ::= <AVP Header: 1001>
    Charging-Rule-Definition <AVP Header: 1003>
        Charging-Rule-Name <AVP Header: 1005>
            Alc-QoS-Action = Forward (1) <AVP Header: 1028>
```

17.5.13.19 Next hop redirect (ingress)

The next hop redirection explicitly or implicitly changes the next hop for the traffic flow within the same service ID (routing context) or a different service ID (routing context).

If the next hop is not explicitly provided, the next hop is selected automatically, according to the routing lookup in the referenced service ID.

The generic Gx directive:

```
Charging-Rule-Install <AVP Header: 1001>
     Charging-Rule-Definition <AVP Header: 1003>
          Charging-Rule-Name <AVP Header: 1005>
               Alc-Next-Hop :: <AVP Header: 1023>
```

```
                      Alc-Next-Hop-IP <AVP Header: 1024>
Alc-V4-Next-Hop-Service-id <AVP Header: 1025>
                     Alc-V6-Next-Hop-Service-id <AVP Header: 1026>
```

This action overwrites the routing table lookup based on the destination IP and sets the next hop to the:

- IPv4/6 address within the same service ID
- IPv4/6 address within a different service ID

The next hop search is indirect, which means that if the explicitly provided next hop in the PCC rule cannot be found in the routing table, then an additional routing table lookup is performed to find the path (next hop) to the indirect next hop from the PCC rule.

If only the service-id is specified in PCC rule (without the next hop), then the next hop is selected from the specified service-id based on the destination IP address of the packet.

17.5.13.20 HTTP redirect (ingress)

HTTP redirect uses Redirect-Information AVP from 3GPP 29.212, §5.3.82.

The generic Gx directive:

```
Redirect-Information < AVP Header: 1085 >
     Redirect-Support < AVP Header: 1086 >
     Redirect-Address-Type < AVP Header: 433 >
     Redirect-Server-Address < AVP Header: 435 >
```

17.5.13.21 Filter forward/drop (ingress and egress)

This action is used to control traffic flow within a PCC rule by using ALU specific AVP. PCC rules are utilizing filters and QoS policies as distinct building blocks. This action within a PCC rule creates an IP or IPv6 filter entry with an action **forward** or **drop**.

The CLI equivalent follows:

```
config>filter
    ip-filter | ipv6-filter <id> create
        entry <id> create
            match
                <5-tuple | dscp>
            exit
            action
                forward | drop
            exit
        exit
    exit
```

The generic Gx directive for filter forward/drop is implemented through ALU specific AVP:

```
Charging-Rule-Install ::= <AVP Header: 1001>
    Charging-Rule-Definition <AVP Header: 1003>
        Charging-Rule-Name <AVP Header: 1005>
            Alc-Filter-Action = Forward (1) <AVP Header: 1027>
            Alc-Filter-Action = Drop (2)
```

17.5.13.22 Service gating function

The service gating function is used to enable or disable the service that is represented by the PCC rule. This action is enforced through a Flow-Status AVP (AVP code 511) - 3GPP 29.214, §5.3.11. The system supports the following values (actions) for the Flow-Status:

enabled (2)

enables all actions specified within the PCC rule. Note that although Flow-Status (service gating function) is considered an action, in this context it is used to enable all other actions that are explicitly set within the PCC rule.

disabled (2)

disable all action specified within the PCC rule and drops the traffic

The service gating function is applicable in the direction that is associated with the rule (PCC rules in the system are unidirectional).

Flow-Status is by default enabled (2) (if the Flow-Status AVP is not explicitly specified within the PCC rule). Flow-Status=Enabled must be accompanied by one or more additional actions in the same PCC rule (see Gx rules with multiple actions and action sharing for a list of allowed simultaneous actions), otherwise the PCC rule instantiation in the node fails.

If the Flow-Status is set to disabled (3), all other actions within the same rule loses their meaning because the packet is dropped. The disabled directive disables the flow of packet through the system. A disabled Flow-Status is equivalent to the Alc-Filter-Action = Drop (2).

This AVP is carried inside of Charging-Rule-Definition (3GPP 29.212, §5.3.5):

```
Charging-Rule-Definition ::= < AVP Header: 1003 >
                              { Charging-Rule-Name }
                             *[ Flow-Information ]
                              [ Flow-Status ]
                              [ QoS-Information ]
                              [ Precedence ]
                             *[ Flows ]
                              [ Monitoring-Key]
                              [ Redirect-Information ]
                             *[ AVP ]
```

17.5.13.23 PCC rule provisioning example

The following is an example of PCC rule provisioning in a CCA-I message:

```
<CC-Answer> ::=  < Diameter Header: 272, PXY >
     < Session-Id >
     { Auth-Application-Id }
     { Origin-Host }
     { Origin-Realm }
     [ Result-Code ]
     [ Experimental-Result ]
     { CC-Request-Type }
     { CC-Request-Number }
     *[ Supported-Features ]
     *[ Event-Trigger ]
     [ Origin-State-Id ]

     Charging-Rule-Install ::= <AVP Header: 1001> -> host instantiation
```

```
          Charging-Rule-Name = "ingr-v4:7"
          Charging-Rule-Name = "eggr-v6:5"
          Charging-Rule-Name = "Sub-Profile:prem"
          Charging-Rule-Name = "Sla-Profile:voip+data"
          Charging-Rule-Name = "Inter-Dest:vport-AN-1"

     Charging-Rule-Install  -> service instantiation
          Charging-Rule-Definition
               Charging-Rule-Name = "service-1"  -> should be able to remove the rule by name
 later on
                    Flow-Information      -> traffic flow definition
                    Flow-Description = "permit in 6(TCP) from any to ip 10.10.10.10/32 40000-
40010"
                    ToS-Traffic-Class = 00101000 11111100]  -> DSCP definition (value mask). In
 case of the DSCP, Flow-Direction (1080) AVP must be included.
                    Flow-Direction = UPSTREAM  -> traffic flow direction
                    QoS-Information <AVP Header: 1016>
                         Max-Requested-Bandwidth-UL = 10000000  -> UPSTREAM rate definition
 (not downstream, since the traffic flow direction is IN)
                         QoS-Class-Identifier  = 3      -> EF forwarding class; in general one
 of 8 forwarding-classes (FC) in 7450 ESS and 7750 SR (be|l2|af|l1|h2|ef|h1|nc). This is used
 for re-prioritization of the traffic.
```

In this example the host is instantiated using the two Charging-Rule-Install AVPs. The first is used to instantiate the host. The second is used to instantiate the IP-criterion based service named service-1. Service-1 is defined as the upstream traffic flow with traffic class AF11, destined for the TCP port range 40000-40010 on the node with IP address 10.10.10.10/32.

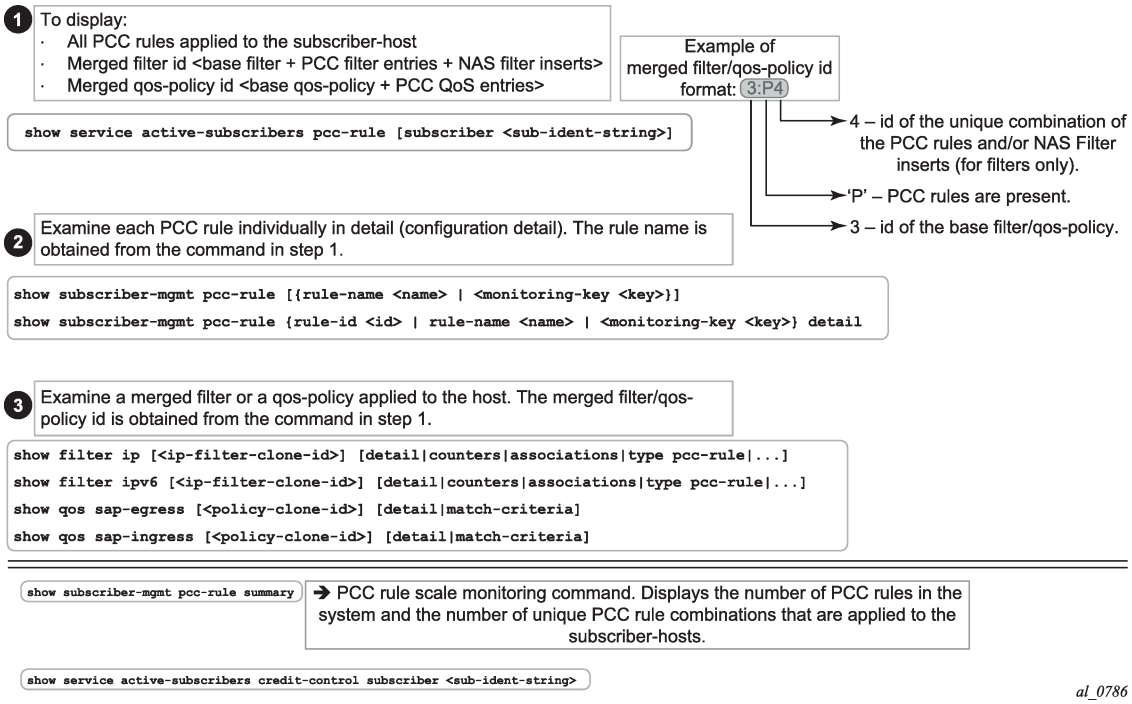
The actions for this traffic flow are:

- rate-limit of 10M
- change forwarding-class to AF11

17.5.13.24 Operational aspects

The commands used to examine dynamic rules and NAS filter inserts associated with the subscriber hosts are shown in Figure 227: Overview of PCC rule-related operational commands.

Figure 227: Overview of PCC rule-related operational commands



17.5.13.25 PCC rules and capacity planning

One of the most important factors to be considered for capacity planning with PCC rules is the number of unique policies that are applied to subscribers.

A unique policy constitutes a base a QoS policy or filter ID along with all PCC rules that are applied to a subscriber or a set of subscribers.

Now examine an example where there are 'n' PCC rules in the system ('n' qos rules and 'n' ACL filter rules). Those rules are applied to IPv4 traffic in ingress direction. Further, assume that the PCC rules do not have defined Precedence AVP, which means that the system can optimize their order for maximum sharing and maximized scale. Then, 'n' PCC rules can be combined by various permutation into 2^n-1 unique combinations Next assumption is that there are five possible base qos-policies for IPv4 traffic in ingress direction and five possible base filters for IPv4 traffic in the ingress direction.

Given the above, the unique PCC rule combinations (2^n-1) together with five base QoS polices produce 5*(2^n-1) unique qos-policies per ingress IPv4. Same logic can be applied for ingress IPv4 filters.

This exercise must be repeated for egress direction as well as for IPv6 type traffic, by taking into consideration the number of respective base qos-policies/filters and the number of PCC rules.

When the number of unique policy combinations is determined and ensured that it is within the system limits, each policy must be further evaluated to determine the number of entries it takes in CAM.

17.5.13.26 PCC rule scaling example

Figure 228: Example of the scaling limits for PCC rules depicts an example relevant to capacity planning with focusing on understanding the scaling limits pertaining to the number of PCC rules and their mutual combinations when they are applied to the subscriber hosts.

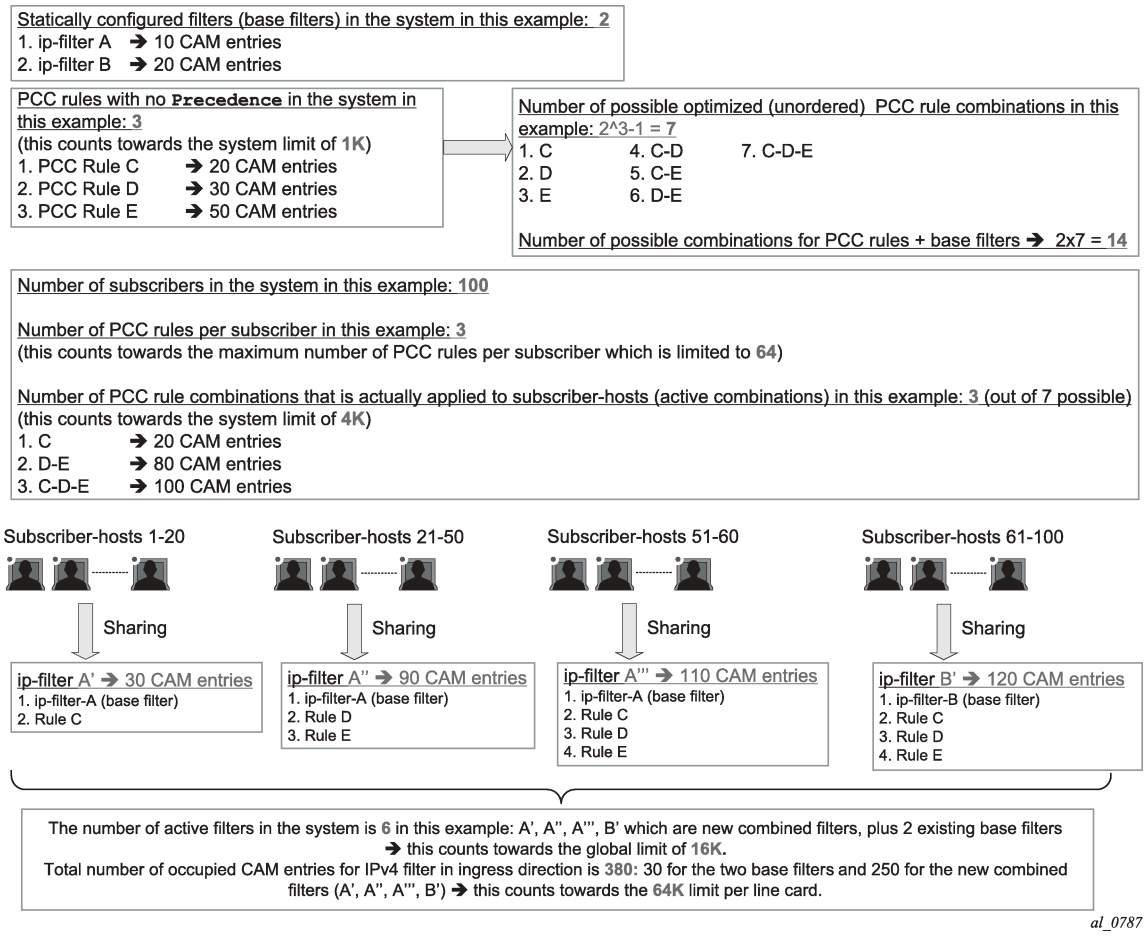
This example is focuses on an IPv4 filter applied in ingress direction but similar logic can be used in understanding other policy types (QoS, egress, IPv6).

The system/line card limits in this example are set to the following values for illustration purposes only:

- Number of PCC rules per system is 1K.
- Max number of rules per subscriber host is 64.
- Max number of combinations of the PCC rules (or services) that are active (applied to the subscriber-hosts) per system is 4K.
- Max ingress IPv4 filter CAM entries per FP2/FP3 is 64k.
- Max filters per system is 16k.

Note: The actual CAM limits vary per policy type (filter/QoS), direction and IP address type (v4 vs v6). The actual scaling limits can be found in the Scaling Guides for the relevant software release.

Figure 228: Example of the scaling limits for PCC rules



17.5.14 NAS filter inserts

Gx filter entries inserted via the NAS-Filter-Rule are subscriber-host-specific entries. This means that in the upstream direction, the source IP address in the NAS-Filter-Rule is always internally set by the node to the IP address of the subscriber host itself. Similarly, in the downstream direction, the destination IP address in the NAS-Filter-Rule is set by the node to be the IP address of the subscriber-host itself.

On the other hand, the entries in the Alc-NAS-Filter-Rule-Shared AVP are processed as received without any modifications. This means that such entries can be shared with all the hosts that have the same Alc-NAS-Filter-Rule-Shared applied.

Similar to QoS overrides, NAS filter entries are not predefined on the node but instead they are defined under the Charging-Rule-Install — Charging-Rule-Definition AVP.

The Charging-Rule-Name AVP for NAS filter inserts is an arbitrary name that is part of Charging-Rule-Definition AVP in which NAS-Filter-Rule AVP or Alc-NAS-Filter-Rule-Shared is provided. Such Charging-Rule-Name is used to report errors related to instantiation of the inserts.

17.5.14.1 Examples of NAS entry inserts

The following AVPs identify NAS filter inserts that are applied to a subscriber host. Those AVPs can be included in CCA-i, CCA-u or RAR message sent from the PCRF.

```
Charging-Rule-Install ::= <AVP Header: 1001>
Charging-Rule-Definition <AVP Header: 1003>
            Charging-Rule-Name <AVP Header: 1005> = “allow-all”
            Alc-NAS-Filter-Rule-Shared <AVP Header: 158> = "permit in ip from any to any "ASCII
 NUL" permit out ip
from any to any"
```

In this example, the filter entry defined in Alc-NAS-Filter-Rule-Shared AVP is inserted in the clone of the existing base filter for the subscribers.

17.5.15 Error handing and rule failure reporting in ESM

The Gx rule (overrides, PCC rules or NAS filter inserts) instantiation failure can occur on two levels:

- **AVP decoding level in the Diameter**

  The Gx message contains an unrecognized AVP with the M-bit set. Then, all Gx rules (ESM, UM and AA) in the message are rejected and the CCR-U with the Charging-Rule-Report AVP (rule status) and Error-Massage AVP (failure description) or an RAA message with the appropriate Result-Code AVP (fail – 5xxx), Error-Message AVP (description) and the Failed-AVP AVP are sent to the PCRF.

  Invalid content of a supported AVP with the M-bit set triggers the same response. Invalid content of an AVP refers to the malformed syntax of an AVP that carries the type of the AVP value implicitly embedded in the AVP value itself. Consider sla-profile:rule-name-1 string. then, the sla-profile: refers to the type of the value carried in the Charging-Rule-Name AVP. The value that the Charging-Rule-Name carries is rule-name-1 and this value represents the sla-profile name already configured on the node (as opposed to filter, sub-profile, category-map, and so on). If the sla-profile is misspelled (type is unrecognized), the entire AVP is un-decodable.

- **Gx rule instantiation level in ESM, UM or AA**

  Each module (ESM, UM or AA) would fail all rules destined for it. The failure of a Gx rule within a module can be caused by referencing non-existing profile (for example sla-profile:unknown-name) or a lack of resources on the node. Then, a CCR-U message from the respective module is sent with the Rule-Report-Status AVP listing all the rules destined for this module and the corresponding Error-Message AVP describing the cause for the failure.

  Another example that can cause all rules with the ESM module to fail would be invalid combination of actions within the rule.

17.5.15.1 AVP decoding failure in Gx

Reporting an AVP decoding problem in Gx is described in the following example:

A Gx directive is received to install two overrides on the node. The two overrides are supposed to change the sla-profiles and sub-profiles for the subscriber host. The AVP that is used to change the sla-profile is miss-formatted. The predefined sla-profile keyword in the Charging-Rule-Install AVP is misspelled as spa-profile instead of sla-profile.

```
Charging-Rule-Install ::= <AVP Header: 1001>
```

```
    Charging-Rule-Name <AVP Header: 1005> = "Sub-Profile:prem"
    Charging-Rule-Name <AVP Header: 1005> = "Spa-Profile:voip+data"
```

Because the Charging-Rule-Name AVP has the M-bit set, the whole message fails and an error is reported. No rules within this Gx message is installed (not even the valid ones, then this would be the Charging-Rule-Name = "**Sub-Profile**:prem").

Note: If the M-bit was clear in the Charging-Rule-Name AVP, the erroneous AVP would be simply ignored and proceed with installation of the remaining, correctly formatted rules.

The nature of the error depends on the original directive sent by the PCRF (RAR or CCA – push or pull model).

If the directive from the PCRF is passed with the **cca** command, the response is CCR-U with the following error related AVPs:

```
[ Error-Message ] – "Invalid value spa-profile:voip+data"

Charging-Rule-Report ::= < AVP Header: 1018 >
    *[ Charging-Rule-Name ] – Spa-Profile:voip+data
    [ PCC-Rule-Status ]    – INACTIVE (1)
    [ Rule-Failure-Code ]  – GW/PCEF_MALFUNCTION (4)


Charging-Rule-Report ::= < AVP Header: 1018 >
    *[ Charging-Rule-Name ]      –  Sub-Profile:prem
    [ PCC-Rule-Status ]          –  INACTIVE (1)
    [ Rule-Failure-Code ]        –  GW/PCEF_MALFUNCTION (4)

Failed-AVP ::= < AVP Header: 279 >
    Charging-Rule-Name = Spa-Profile:voip+data
```

If the directive is passed to the node through RAR, the node responds with the following RAA message:

```
Failed-AVP ::= < AVP Header: 279 >
    Charging-Rule-Name = Spa-Profile:voip+data

Result-Code ::= < AVP Header: 268 > = DIAMETER_INVALID_AVP_VALUE (5004)
```

Similarly, if the number of filter entries for each entry type (NAS-Filter-Rule — host-specific or Alc-NAS-Filter-Rule-Shared — shared) exceeds the maximum supported number (see the *7750 SR and VSR Gx AVPs Reference Guide*), the whole message fails the decoding phase.

The reason that the Result-Code AVP is present in the RAA message and not in the CCR-U message is that this code is only allowed to be present in the answer messages, according to the standard.

17.5.15.2 ESM rule-installation failure

This assumes that the rule installation directives are successfully passed from the Gx module to the ESM module and the failure to install rules occurs in the ESM module.

In the Gx override example below, the referenced sla-profile is unknown. Then, all directives passed to the ESM module fails and consequently no rules or overrides are installed. The sub-profile change fails as well although the prem sub-profile is known in the system.

```
Charging-Rule-Install ::= <AVP Header: 1001>
    Charging-Rule-Name <AVP Header: 1005> = "Sub-Profile:prem"
            Charging-Rule-Name <AVP Header: 1005> = "Sla-Profile:unknown"
```

The error reporting flow is as follows:

- If the directives are passed using the CCA command, the response is CCR-U command with the following error related AVPs:

```
[ Error-Message ]                    – "sla-profile 'unknown' lookup failed"

Charging-Rule-Report ::= < AVP Header: 1018 >
    *[ Charging-Rule-Name ]                –  Sla-Profile:unknown
    [ PCC-Rule-Status ]                    –  INACTIVE (1)
    [ Rule-Failure-Code ]                  –  GW/PCEF_MALFUNCTION (4)


Charging-Rule-Report ::= < AVP Header: 1018 >
    *[ Charging-Rule-Name ]         –  Sub-Profile:prem
    [ PCC-Rule-Status ]             –  INACTIVE (1)
    [ Rule-Failure-Code ]           –  GW/PCEF_MALFUNCTION (4)
```

- If the directive is passed to the node with RAR, the node responds with the following messages:

  RAA = OK because the Gx module successfully processed the AVP parsing.

  The RAA is followed by CCR-U, triggered by the rule instantiation failure in ESM module. CCR-U contains the following AVP related to the rule status:

```
[ Error-Message ]                –   "sla-profile 'unknown' lookup failed"

Charging-Rule-Report ::= < AVP Header: 1018 >
    *[ Charging-Rule-Name ]       –  Sla-Profile:unknown
    [ PCC-Rule-Status ]           –  INACTIVE (1)
    [ Rule-Failure-Code ]         –  GW/PCEF_MALFUNCTION (4)


Charging-Rule-Report ::= < AVP Header: 1018 >
    *[ Charging-Rule-Name ]       –  Sub-Profile:prem
    [ PCC-Rule-Status ]           –  INACTIVE (1)
    [ Rule-Failure-Code ]         –  GW/PCEF_MALFUNCTION (4)
```

Similar behavior would be exhibited if the directive is sent to the UM or AA modules. However, ESM, UM, and AA are separate modules and failure to install rules in one module does not affect rule installation in another.

17.5.15.3 Failure reporting in AA

Failure reporting in AA is performed in similar fashion as in ESM.

Instead of Charging-Rule-Report AVP, the ADC-Rule-Report is used:

```
ADC-Rule-Report ::= < AVP Header: 1097 >
```

```
                        *[ ADC-Rule-Name ]
                         [ PCC-Rule-Status ]
                         [ Rule-Failure-Code ]
                        *[ AVP ]
```

17.5.15.4 Summary of failure reporting

Table 67: Failure reporting summarizes Gx failure reporting on the node.

Table 67: Failure reporting

Failure event	Gx message received via CCA (pull model)	Gx message received via RAR (push model)
AVP decoding/interpreting failure; M-bit cleared	Ignore AVP	Ignore AVP
AVP decoding/interpreting failure; M-bit set	CCR-U is sent by the node. CCR-U contains: • Charging-Rule-Report AVP for all rules (all rules inactive) • First failed AVP in Failed-AVP AVP • Error-Message AVP at the top level describing the reason for the failure. No rules within the message is instantiated on the node.	RAA is sent by the node. RAA contains: • Result-Code AVP [DIAMETER_INVALUID_AVP_VALUE (5004), DIAMETER_AVP_UNSUPPORTED (5001), DIAMETR_UNABLE_TO_COMPLY (5012) • First failed AVP in Failed-AVP AVP No rules within the message is instantiated on the node.
Rule failure in ESM	CCR-U is sent by the node. CCR-U contains: • Charging-Rule-Report AVP for all rules (all rules inactive) • Error-Message AVP at the top level describing the reason for the failure. No rules is instantiated in the ESM module.	RAA with the Result-Code AVP 'success' (2001) is sent by the node, followed by a CCR-U. CCR-U contains: • Charging-Rule-Report AVP for all rules (all rules inactive) • Error-Message AVP at the top level describing the reason for the failure. No rules is instantiated in the ESM module.
Rule failure in Usage-Monitoring (UM)	CCR-U is sent by the node. CCR-U contains: • Charging-Rule-Report AVP for all rules (all rules inactive) • Error-Message AVP at the top level describing the reason for the failure.	RAA with the Result-Code AVP 'success' (2001) is sent by the node, followed by a CCR-U. CCR-U contains: • Charging-Rule-Report AVP for all rules (all rules inactive) • Error-Message AVP at the top level describing the reason for the failure.

Failure event	Gx message received via CCA (pull model)	Gx message received via RAR (push model)
	No rules is instantiated in the UM module.	No rules is instantiated in the UM module.
Rule failure in AA	CCR-U is sent by node. CCR-U contains: • ADC-Rule-Report AVP for all rules (all rules inactive) • Error-Message AVP at the top level describing the reason for the failure. No AA rules is instantiated in the AA module.	RAA with the Result-Code AVP 'success' (2001) is sent by the node, followed by CCR-U. CCR-U contains: • ADC-Rule-Report AVP for all rules (all rules inactive) • Error-Message AVP at the top level describing the reason for the failure. No rules is instantiated in the AA module.

17.5.16 Usage-Monitoring and reporting

Usage-Monitoring and reporting refers to the collection and reporting of octets (volume) that a service or application on the node has consumed during a specific period. The usage on the node is reported via the Gx interface to the PCRF. Based on this information, the PCRF can apply a specific action (policy change) to the entity being monitored. For example, QoS can be modified, or the service can be blocked when specific thresholds are reached.

Usage-Monitoring and reporting is performed over a single Gx session for the ESM/AA subscriber. In other words, there is only a single session for an ESM subscriber-host and corresponding AA subscriber. Via this single Gx session, Usage-Monitoring can be requested simultaneously in ESM context (PCC rule level, credit-category or IP-CAN session) and AA context (application based Usage-Monitoring).

For volume based usage monitoring in dual-homed system, see Gx Usage Monitoring in dual-homed systems.

17.5.16.1 ESM Usage-Monitoring - what is being monitored

In the ESM context, volume consumption (octets - 3GPP 23.203 §4.4) can be monitored on three levels:

- per entire IP-CAN session
- per credit-category
- per PCC rule

Usage-Monitoring can be monitored simultaneously on all three levels.

An IP-CAN session on the node represents a subscriber-host whose service types are determined by the sla-profile instance. In per IP-CAN session volume monitoring, the aggregated queue or policer counters are reported per direction (in or out). This includes dynamic policers that are instantiated as a result of a Gx action; for example, rate-limiting.

The following configuration is necessary to allow per IP-CAN session level Usage-Monitoring to be enabled for sessions associated with the category map:

```
configure
    subscriber-mgmt
        category-map <category-map-name> [create]
            gx-session-level-usage
```

If the sla-profile instance changes mid-session, the counters are reset.

One obvious difference between regular RADIUS accounting and Gx Usage-Monitoring is that in RADIUS accounting the cumulative byte number for sla-profile instance is presented in each report (interim-updates or stop acct messages), while in Usage-Monitoring this count is reset between the two reports (when the quota is reached, the usage report is triggered).

Per credit-category monitoring refers to volume monitoring of a single queue/policer or a set of queues/policers within the sla-profile instance. Each queue/policer (or set of queues/policers as a subset of the sla-profile instance) represents a service for which the Usage-Monitoring is required. Those queues/policers (services) are organized on the node in credit categories.

```
*A:7750>config>subscr-mgmt>cat-map# info
----------------------------------------------
            activity-threshold 1
            credit-exhaust-threshold 50
            category "queue1" create
                queue 1 ingress-egress
            exit
            category "queue3-5" create
                queue 3 ingress-egress
                queue 5 ingress-egress
            exit
            category "rest-queues" create
                queue 2 egress-only
                queue 4 egress-only
                queue 6 egress-only
                queue 7 egress-only
                queue 8 egress-only
            exit
----------------------------------------------
```

Each service category has a name that is used to reference the category in Usage-Monitoring and reporting.

The category-map (predefined on the node) that is used in Usage-Monitoring can be associated with the subscriber-host through the following methods (in the order of priority):

- PCRF — Charging-Rule-Install AVP that references the category map in Charging-Rule-Name = cat-map:<cat-map-name>
- LUDB
- RADIUS
- Python script

PCC rule Usage-Monitoring reports volume usage per flow or set of flows. PCC rule Usage-Monitoring is described in a separate section below.

Usage-Monitoring for the subscriber host can be configured on the node, but it is not active until it is turned on by the PCRF either via CCA-i, CCA-u or RAR.

Usage-Monitoring can be enabled per ingress or egress direction or as total count. However monitoring the total count and per direction count are mutually exclusive. For example, total Usage-Monitoring cannot be enabled simultaneously with ingress (or egress) Usage-Monitoring for the same monitoring entity (session or category).

17.5.16.2 AA Usage-Monitoring – what is being monitored

In AA, charging groups (CG), application groups (AG) and applications are monitored. See the *7450 ESS, 7750 SR, and VSR Multiservice ISA and ESA Guide* for details.

17.5.16.3 Requesting Usage-Monitoring in ESM

Gx Usage-Monitoring is activated explicitly from the PCRF via CCA-I, CCA-U or RAR. It is triggered via the Usage-Monitoring-Information AVP along with the event-trigger = usage-report (33). The Usage-Monitoring-Information AVP contains the following AVPs:

```
Usage-Monitoring-Information::= < AVP Header: 1067 >
                                  [ Monitoring-Key ]
                            0,2   [ Granted-Service-Unit ]
                            0,2   [ Used-Service-Unit ]
                                  [ Usage-Monitoring-Level ]
                                  [ Usage-Monitoring-Report ]
                                  [ Usage-Monitoring-Support ]
```

There could be multiple instances of Usage-Monitoring-Information AVP present in a single CCA or RAR messages. For example, simultaneous Usage-Monitoring for IP-CAN session level, credit-category level or pcc rule level can be requested.

Usage-Monitoring-Level for IP-CAN session is set to SESSION_LEVEL (0)

Usage-Monitoring-Level for category-map is set to PCC_RULE_LEVEL (1)

Usage-Monitoring-Level for PCC rules is set to PCC_RULE_LEVEL (1)

17.5.16.4 Reporting accumulated usage

The node reports usage information to the PCRF under the following conditions:

- when a usage threshold is reached
- when all pcc rules associated with the monitoring are removed or deactivated
- when Usage-Monitoring is explicitly disabled by PCRF
- when a session is terminated
- when requested by PCRF (on demand)

To report accumulated usage for a specific monitoring-key, the node sends a CCR with the Usage-Monitoring-Information AVP containing the accumulated usage information since the last report. For each of the enabled monitoring-keys, the Usage-Monitoring-Information AVP includes the Monitoring-Key AVP and the accumulated volume usage in the Used-Service-Unit AVP.

A usage report on the node can be triggered by reaching the usage threshold communicated to the node by the PCRF in the CCR-U message carrying accumulated usage for that monitoring entity along with the Event-Trigger AVP set to USAGE_REPORT.

In response to the CCR-U message, the PCRF communicates to the node via a CCA-u message whether the Usage-Monitoring should continue:

- If the new thresholds for the currently monitored entity/levels are provided in Granted-Service-Units AVP, the Usage-Monitoring continues.
- If the thresholds are not included in Granted-Service-Units AVP, the Usage-Monitoring stops.

Thresholds are incremental. For example, if the quota of 100 MB is submitted to the node, the usage should be reported when that quota is reached. At that point, the user can be granted another 100 MB. The new usage report on the node is triggered when another 100 MB are accumulated. Absence of the threshold for an entity in the CCA-u message is an indication that the Usage-Monitoring should stop.

When the PCRF informs the node that Usage-Monitoring should stop (by not including thresholds in CCA-u), the node does not report usage which has accumulated between sending the CCR and receiving the CCA.

Another possibility of usage reporting is on-demand. In this scenario, usage for one or more monitoring keys is reported whether the usage threshold has been reached. This is achieved by sending the node the Usage-Monitoring-Report AVP (within the Usage-Monitoring-Information AVP) set to USAGE-MONITORING_REPORT_REQUIRED. If the Monitoring-Key AVP is omitted in such a request, Usage-Monitoring for all enabled entities is reported to the PCRF.

If that the credit-category is removed from the subscriber host (the sla-profile instance referencing the category-map is changed for the subscriber host), the node reports the outstanding usage in a CCR-U message with the Event-Trigger set to USAGE_REPORT.

17.5.16.5 Disabling Usage-Monitoring

When the PCRF explicitly disables Usage-Monitoring on the node, the node reports the accumulated usage which has occurred while Usage-Monitoring was enabled.

To disable Usage-Monitoring for an entity, the PCRF sends the Usage-Monitoring-Information AVP referencing only the applicable monitoring entity with the Monitoring-Key AVP and the Usage-Monitoring-Support AVP set to USAGE_MONITORING_DISABLED.

When the PCRF disables Usage-Monitoring in a RAR or CCA command, the node sends new a CCR-U with the Event-Trigger AVP set to "USAGE_REPORT" to report the accumulated usage for the disabled Usage-Monitoring entities.

17.5.16.6 Usage-Monitoring for PCC rules

Each PCC rule for which Usage-Monitoring is required, contains Monitoring-Key AVP.

```
Charging-Rule-Definition ::= < AVP Header: 1003 >
                              { Charging-Rule-Name }
                             *[ Flow-Information ]
                              [ Flow-Status ]
                              [ QoS-Information ]
                              [ Precedence ]
                              [ Monitoring-Key]
                             *[ AVP ]
```

Usage-Monitoring for PCC rules is implemented through a dynamic policer. The policer is instantiated at the time when the PCC rule with Monitoring-Key AVP is installed.

The same monitoring-key can be used in multiple PCC rules assuming that these rules are for the same direction. In other words, the charging rule is rejected if the same monitoring-key is used for ingress and egress.

17.5.16.7 Session termination

At IP-CAN session termination, the node sends the accumulated usage information for all entities for which Usage-Monitoring is enabled in the CCR-T.

17.5.16.8 Usage Monitoring when multiple subscriber hosts or sessions share an SLA profile instance

A Diameter Gx session from which Usage Monitoring is started controls the Usage Monitoring for the entire SLA profile instance. Only one Diameter Gx session can control the Usage Monitoring per SPI at a specific time. That is, Usage Monitoring can only be started from a single Gx session when multiple subscriber hosts or sessions share an SLA profile instance.

The session ID of the Diameter Gx session that controls the Usage Monitoring is displayed as the "Diameter Session Gx" field in the output of the following **show** command.

```
# show service active-subscribers credit-control [subscriber <sub-ident-string>]
===============================================================================
Active Subscribers
===============================================================================
-------------------------------------------------------------------------------
Subscriber sub-1 (sub-profile-1)
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
(1) SLA Profile Instance sap:1/1/2:10 - sla:sla-profile-1
-------------------------------------------------------------------------------
Category Map         : catmap-1
Diameter Session Gx  : bng1.domain.com;1490015297;7
Number of categories
static               : 2
gx-session           : 0
gx-pcc               : 1
Category Name        : cat-1
Ingress Queues       : 1
Egress Queues        : 1
Ingress Policers     : (Not Specified)
Egress Policers      : (Not Specified)
No monitoring
HTTP Rdr URL Override: (Not Specified)
Category Name        : cat-2
Ingress Queues       : 2
Egress Queues        : 2
Ingress Policers     : (Not Specified)
Egress Policers      : (Not Specified)
No monitoring
HTTP Rdr URL Override: (Not Specified)
Category Name        : fl_um1_up
Volume Used          : 0                    Volume Available     : 100000
```

```
HTTP Rdr URL Override: (Not Specified)
-------------------------------------------------------------------------------
```

A Diameter Gx session stops being the controlling Gx session for Usage Monitoring of the SLA profile instance in the following situations:

- The Gx session controlling Usage Monitoring for an SLA profile instance is terminated, for example, because the associated subscriber session is disconnected. A CCR-T is sent immediately reporting the usage.
- All Usage Monitoring is disabled form the controlling Gx session. A CCR-U is sent immediately reporting the usage.

When the Usage Monitoring terminates, the usage of subscriber hosts or sessions sharing the SLA profile instance is no longer accounted for. Usage Monitoring can now be started from another Diameter Gx session, which then becomes the controlling Gx session for Usage Monitoring on the SLA profile instance.

17.5.16.9 Usage-Monitoring examples

For the description of the specific AVP, see the *7750 SR and VSR Gx AVPs Reference Guide*.

IP-CAN session Usage-Monitoring

A **category-map** with **gx-session-level-usage-monitoring** must be associated with the subscriber host or session:

```
configure
    subscriber-mgmt
        category-map cat-map-1 create
            gx-session-level-usage
```

PCRF in RAR sends the following AVPs (among all the other mandatory ones: session-id, and so on.)

```
Usage-Monitoring-Information
        Monitoring-Key = “any-string”
        Granted-Service-Unit
            CC-Input-Octets = 1000000
              CC-Output-Octets = 1000000
        Usage-Monitoring-Level = session_level(0)

Event-Trigger = USAGE_REPORT
```

The node reports usage when the thresholds are reached sometime later in the CCR-U. The usage is monitored internally on the node based on the current sla-profile instance.

```
Usage-Monitoring-Information
        Monitoring-Key = “any-string”
        Used-Service-Unit
        CC-Input-Octets = 1000000
                 CC-Output-Octets = 1000000
```

The PCRF instructs the node to continue Usage-Monitoring with the new thresholds in the CCA-U:

```
Usage-Monitoring-Information
        Monitoring-Key = “any-string”
        Granted-Service-Unit
            CC-Input-Octets = 1000000
```

```
            CC-Output-Octets = 1000000
        Usage-Monitoring-Level = session_level(0)
```

Category Usage-Monitoring

Assume that the following category-map is associated with the subscriber host:

```
*A:7750>config>subscr-mgmt>cat-map# info
----------------------------------------------
            activity-threshold 1
            credit-exhaust-threshold 50
            category "queue1" create
                queue 1 ingress-egress
            exit
            category "queue3-5" create
                queue 3 ingress-egress
                queue 5 ingress-egress
            exit
            category "rest-queues" create
                queue 2 egress-only
                queue 4 egress-only
                queue 6 egress-only
                queue 7 egress-only
                queue 8 egress-only
            exit
```

The PCRF sends the following AVPs in the RAR message (among all the other mandatory ones: session-id, and so on.)

```
Charging-Rule-Install
        Charging-Rule-Name = Cat-Map:cat1 – cat-map rule install

Usage-Monitoring-Information
        Monitoring-Key = “queue-1”
        Granted-Service-Unit
            CC-Input-Octets = 1000000
            CC-Output-Octets = 1000000
            Usage-Monitoring-Level = PCC_RULE_LEVEL (1)
Usage-Monitoring-Information
        Monitoring-Key = “queue-3-5”
        Granted-Service-Unit
            CC-Input-Octets = 2000000
            CC-Output-Octets = 2000000
            Usage-Monitoring-Level = PCC_RULE_LEVEL (1)
Event-Trigger = USAGE_REPORT
```

The node reports usage when the thresholds are reached sometime later in the CCR-U:

```
Usage-Monitoring-Information
        Monitoring-Key = “queue-1”
        Used-Service-Unit
               CC-Input-Octets = 1000000
               CC-Output-Octets = 1000000
Usage-Monitoring-Information
        Monitoring-Key = “queue-3-5”
        Used-Service-Unit
               CC-Input-Octets = 2000000
               CC-Output-Octets = 2000000
```

The PCRF instructs the node to continue Usage-Monitoring with the new thresholds in the CCA-U:

```
Usage-Monitoring-Information
        Monitoring-Key = “queue-1”
        Granted-Service-Unit
            CC-Input-Octets = 1000000
            CC-Output-Octets = 1000000
            Usage-Monitoring-Level = PCC_RULE_LEVEL (1)
Usage-Monitoring-Information
        Monitoring-Key = “queue-3-5”
        Granted-Service-Unit
            CC-Input-Octets = 2000000
            CC-Output-Octets = 2000000
            Usage-Monitoring-Level = PCC_RULE_LEVEL (1)
```

17.5.17 Event triggers

The PCRF may subscribe to an event trigger on the node. The PCRF subscribes to new event triggers or removes armed event triggers unsolicited at any time. When an event matching the event trigger occurs, the node reports the event to the PCRF. The event triggers that are required in procedures are unconditionally reported (for example, IP address allocation/de-allocation) from the node, while the PCRF may subscribe to the remaining events (for example Usage-Monitoring).

When sent from the PCRF to the node, the Event Trigger AVP indicates an event that triggers an action in the node. When sent from the node to the PCRF, the Event Trigger AVP indicates that the corresponding event has occurred. If no Event Trigger AVP is included in a CCA or RAR operation, any previously provisioned event trigger is still applicable.

The PCRF may remove all previously provided event triggers by providing the Event-Trigger AVP set to the value NO_EVENT_TRIGGERS. When an Event-Trigger AVP is provided with this value, no other Event-Trigger AVP is provided in the **cca** or **rar** command. Upon reception of an Event-Trigger AVP with this value, the node does not inform the PCRF of any event except for those events that are always reported and do not require provisioning from the PCRF.

When the PCRF subscribes to one or more event triggers by using the **rar** command, the node sends the corresponding currently applicable values to the PCRF in the RAA if available, and in this case, the Event-Trigger AVPs are not included.

For a list of the supported events on the node, see the *7750 SR and VSR Gx AVPs Reference Guide*.

17.5.18 Subscriber verification

At any time, the PCRF can query the node for the presence of the subscriber-host via a RAR message.

The node responds with the following result-codes in RAA:

- DIAMETER_SUCCES (2001) — subscriber active
- DIAMETER_UNKNOWN_SESSION_ID (5002) — subscriber does not exist

17.5.19 Subscriber termination

The PCRF can request IP-CAN session termination on the node via two messages:

- a RAR directive with the Session-Release-Cause AVP to the node
- ASR

Upon the arrival of either of those messages, the node starts the IP-CAN session termination procedure (CCR-T with a corresponding Termination-Cause AVP is sent to the PCRF). This is described in the 3GPP 29.212 document, §4.5.9.

For a list of the supported Termination-Cause AVP values on the node, see the *7750 SR and VSR Gx AVPs Reference Guide*.

17.5.20 Mobility support in Wi-Fi

When a Wi-Fi subscriber moves between the access points (APs), a CCR-U message is triggered on the node, carrying the Called-Station-Id AVP. The Called-Station-Id AVP carries the MAC IP address of the new AP. This functionality allows the PCRF to make a location-based policy decision.

This functionality is enabled via event trigger USER_LOCATION_CHANGE (13) [3GPP 29.212, §5.3.7] sent to the node by a PCRF in a CCA or RAR message.

The same event is reported back from the node to the PCRF in a CCR-U message when the user location changes.

17.5.20.1 Redundancy

Redundancy in Gx relies on the Diameter redundancy mechanisms described in Diameter redundancy.

17.5.21 Persistency and Origin-State-ID AVP

Persistency and Origin-State-ID AVP (RFC 6733, §8.6 and §8.16).

Persistency (saving the state of IPoE hosts on the compact flash) for Gx sessions is not supported. This means that, on reboot, the node restores the subscriber-hosts from the persistency but the Gx session awareness for the recovered hosts is lost. Any previously applied QoS or filter overrides are lost. However, subscriber-strings (subscriber-id, sub-profile, sla-profile, aa-profile) can be made persistent and can be preserved across reboots.

The Origin-State-Id (OSI) AVP is not stored in persistency. If the node reboots, the Origin-State-ID AVP is set to boot time (UTC).

The Origin-State-Id AVP is contained in the CER messages and application messages that are sent from the node to the PCRF/DRA. In the other direction, sent by the PCRF to the node, the OSI is ignored.

To restore a lost session after the reboot, the node initiates a CCR-I message for every host that has recovered from persistency. The CCR-I contains the new session-id and origin-state-id. Based on this CCR-I, it is expected that the PCRF returns the most current policy for the host.

17.5.22 Overload protection

Each SR OS node has a receiving queue per Gx application (ESM, UM, AA). Each queue can hold 10,000 messages. While the queue is in the overloaded state, the SR OS node replies to every new RAR message with the RAA (ACK) immediately followed by a CCR-U message containing the error-message with the description 'Overload'. This can be considered as explicit signaling toward the PCRF notifying it of the condition on the SR OS node.

If the messages in the overwhelmed queue do not require sending an answer (in case that the overwhelmed queue contains CCA-I/U messages), the TCP window fills up, TCP ACKs are not sent and consequently this IS an implicit notification to the PCRF to slow down.

If the SR OS node receives a response from an overloaded PCRF (Result-Code = DIAMETER_TOO_BUSY), the SR OS node timeouts (tx-timer) the originally sent message. After the message is timed out, the configuration settings (on-failure) determines whether to trigger the peer-failover procedure or not (Peer-failover based on DIAMETER_TOO_BUSY Result-Code is recommended in RFC6733, §7.1.3.

17.6 Supported-Features AVP in Gx

The Supported-Features AVP is used to negotiate common supported features between PCEF and PCRF on a Gx session. The list of supported features on SR OS and their AVP format is described in the *7750 SR and VSR Gx AVPs Reference Guide*.

17.6.1 Extended bandwidth 5G new radio feature

The Gx specification Release 15 (3GPP doc 29.212) and the Rx specification (3GPP doc 29.214) define extended AVPs for higher bandwidth values which are used to accommodate the need for higher speeds offered to the subscribers in 5G-enabled networks.

The difference between the non-extended and extended AVPs carrying the bandwidth value is in their units. The bandwidth unit for non-extended bandwidth AVPs is defined in b/s, while extended bandwidth AVPs are in kb/s, increasing the upper limit of the supported bandwidth by a scale of 1000. The value for QoS rates carried in those AVPs is a 32-bit value. In b/s units, this amounts to a maximum supported rate of 2^{32} - 1 = 4 294 967 295 b/s or approximately 4.3 Gb/s. In kb/s, this equals over 4 Tb/s.

Support for the extended AVPs that can carry rate information in kb/s is enabled through the negotiated Extended Bandwidth 5G New Radio (Extended-BW-NR) feature (§5.4.1, 3GPP doc 29212). This feature is negotiated during Gx session establishment phase (CCR-I/CCA-I), between the PCEF and PCRF.

Extended-BW-NR feature is defined as:

- Vendor-Id = 10415 (3GPP)
- Feature-List-Id = 2
- Feature-List = bit 7 (bitmask 00000000 00000000 00000000 10000000 to 0x80) with the AVP flag rule set to *O* (Optional)

Feature-List-Id = 2 is included in the CCR-I message only if the following conditions are met:

- Application is on Gx.

- Extended-BW-NR feature is enabled.

```
config
      subscr-mgmt
        diam-appl-plcy
           gx
             features
                extended-bw
```

- Supported-Features AVP is enabled.

```
config
  subscr-mgmt
     diam-appl-plcy
        gx
          include-avp
                supported-features
```

After the Extended-BW-NR feature is successfully negotiated, the following AVPs with units in kb/s are supported:

- Extended-GBR-DL (AVP code 2850)
- Extended-GBR-UL (AVP code 2851)
- Extended-Max-Requested-BW-DL (AVP code 554)
- Extended-Max-Requested-BW-UL (AVP code 555)
- Extended-APN-AMBR-DL(AVP code 2848)
- Extended-APN-AMBR-UL (AVP code 2849)

17.6.2 Transmission of extended bandwidth AVPs during Gx session initiation

The node initiating the Gx session (sending the CCR-I message) with support for extended bandwidth AVPs is not aware of whether the peering node supports extended bandwidth AVPs at the session initiation time. The extended-BW-NR feature flag is negotiated from both sides during session negotiation.

For this reason, if the bandwidth information carrying AVPs needs to be transmitted in the CCR-I message during session initiation and the bandwidth value is greater than 2^{32} - 1, both AVPs, non-extended and extended, must be transmitted.

- The non-extended bandwidth AVP is set to the value of 2^{32} - 1 (maximum value).
- The extended bandwidth AVP is set to the real bandwidth value.

If the receiving node supports extended AVPs, it processes the received extended bandwidth AVP with the real value; otherwise it processes the non-extended bandwidth AVP with the maximum value.

In SR OS, only APN-related bandwidth AVPs are transmitted during session negotiation and, therefore, their processing should adhere to the above described behavior.

17.6.3 Processing the extended bandwidth AVPs

When both AVP types (standard and extended) are received, while SR OS is supporting extended type (Extended-BW-NR), the extended AVP is processed and the standard one is ignored.

If the extended AVP type is not supported, that is if the Extended-BW-NR is not enabled, and a message with extended bandwidth AVPs is received, one of the following occurs:

- If the M-bit in the extended type AVP is set, SR OS returns an error.
- If the M-bit in the extended type AVP is not set, SR OS ignores the extended type and processes the standard type.

17.7 Diameter NASREQ application

The Diameter NASREQ application is used for Authentication, Authorization, and Accounting services in the Network Access Server (NAS) environment. The SR OS supports a stateless operation of NASREQ authentication and authorization, interacting with a NASREQ server that does not maintain session state.

Subscriber host or session authentication results in an AA-Request (AAR) message being sent to the Diameter NASREQ server. An Auth-Session-State AVP with value equal to 1 (No State Maintained) is included in the AAR to inform the server of the stateless mode. The server responds with an AA-Answer (AAA) message and must include the Auth-Session-State AVP with value equal to 1 (No State Maintained), together with the authorization AVPs.

Diameter NASREQ accounting is not supported.

Table 68: Supported Diameter NASREQ messages lists the supported Diameter NASREQ messages. Vendor-specific AVPs are shown as: v-<vendor-id>-<AVP id>.

Table 68: Supported Diameter NASREQ messages

Diameter message		Code
AAR	AA-Request	265
AAA	AA-Answer	265

Diameter NASREQ authentication is supported for IPoE hosts and sessions, PPPoE PTA PAP/CHAP, and L2TP LAC/LNS authentication.

NASREQ and RADIUS authentication cannot be configured simultaneously on a capture-sap, local-user-database, or group-interface. They have the same priority in the hierarchy of different sources (such as local user database, Gx, defaults, and so on) for obtaining the subscriber host or session authorization parameters.

Multi-chassis redundancy is supported via separate Diameter NASREQ peers on each redundant node. Each node of the multichassis redundancy pair has its own Diameter Identity (origin host or realm). The subscriber host or session is authenticated on the BNG where it is initially connected. Because of the stateless operation, there is no need to synchronize NASREQ session state. Alternatively, Diameter Multi-Chassis Redundancy can be deployed as described in Diameter redundancy.

The following rules apply for stateless NASREQ re-authentication:

- For PPPoE sessions, there is no re-authentication.
- For the IPoE host model, only forced re-authentication of DHCP renews when the circuit ID, interface ID, or remote ID has changed.

- For the IPoE session model, re-authentication of DHCP renews when the ipoe-session min-auth-interval expired or forced re-authentication of DHCP renews when the circuit ID, interface ID, or remote ID has changed.

Stateless NASREQ authentication can be complemented with Diameter Gx policy management for policy control and mid-session changes. Diameter NASREQ and Gx applications are supported simultaneously on a single Diameter peer.

Figure 229: Sample Diameter NASREQ call flow shows a sample call flow for a subscriber using Diameter NASREQ for authentication and Diameter Gx for policy management.

Figure 229: Sample Diameter NASREQ call flow

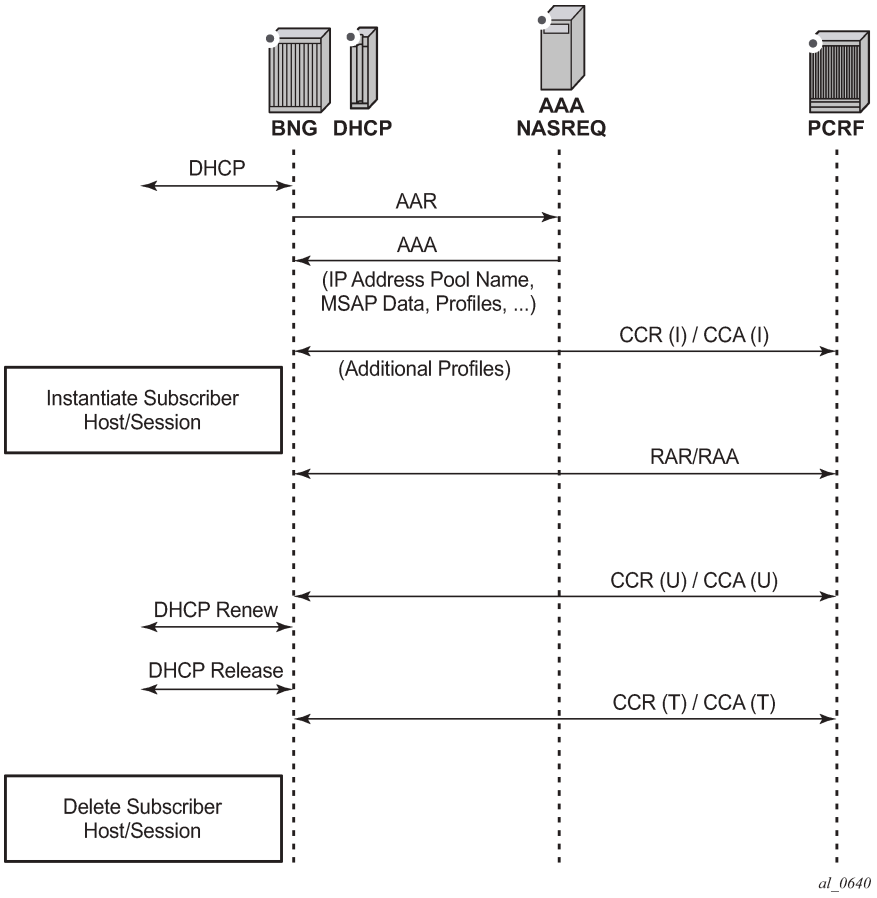


Table 69: AA-Answer message — accepted authorization AVPs lists the authorization AVPs that are accepted in a Diameter NASREQ AA-Answer message. Vendor-specific AVPs are shown in the table as: v-<vendor-id>-<AVP-id>.

Table 69: AA-Answer message — accepted authorization AVPs

AVP ID	AVP name	Description
1	User-Name	Overrides the "Radius User-Name".
8	Framed-IP-Address	The IPv4 address of the subscriber host.

AVP ID	AVP name	Description
9	Framed-IP-Netmask	The IPv4 netmask of the subscriber host.
22	Framed-Route	IPv4 managed route to be configured on the NAS for a routed subscriber host.
25	Class	Opaque value; echoed in RADIUS accounting.
88	Framed-Pool	The name of an IPv4 address pool.
97	Framed-IPv6-Prefix	SLAAC IPv6 prefix (wan-host).
99	Framed-IPv6-Route	IPv6 managed route to be configured on the NAS for a v6 routed subscriber host.
100	Framed-IPv6-Pool	The name of an IPv6 IA-NA address pool (wan-host).
123	Delegated-IPv6-Prefix	DHCPv6 IA-PD IPv6 prefix (pd-host).
26.6527.9	Alc-Primary-Dns	The IPv4 address of the primary DNS server.
26.6527.10	Alc-Secondary-Dns	The IPv4 address of the secondary DNS server.
26.6527.11	Alc-Subsc-ID-Str	Unique subscriber ID string.
26.6527.12	Alc-Subsc-Prof-Str	Subscriber profile string.
26.6527.13	Alc-SLA-Prof-Str	SLA profile string.
26.6527.16	Alc-ANCP-Str	ANCP string.
26.6527.17	Alc-Retail-Serv-Id	The service-id of the retailer to which this subscriber host belongs.
26.6527.18	Alc-Default-Router	The default gateway for the user (DHCP option [3] default-router for a DHCPv4 proxy)
26.6527.28	Alc-Int-Dest-Id-Str	Intermediate destination ID string.
26.6527.29	Alc-Primary-Nbns	The IPv4 address of the primary NetBios Name Server (NBNS).
26.6527.30	Alc-Secondary-Nbns	The IPv4 address of the secondary NetBios Name Server (NBNS).
26.6527.31	Alc-MSAP-Serv-Id	Service ID where the managed SAP is to be created.
26.6527.32	Alc-MSAP-Policy	Managed SAP policy used to create the MSAP.
26.6527.33	Alc-MSAP-Interface	Group-interface name where the managed SAP is to be created.
26.6527.45	Alc-App-Prof-Str	Application profile string.

AVP ID	AVP name	Description

26.6527.46	Alc-Tunnel-Group	The tunnel-group configured in the following contexts.
		`configure router l2tp group`
		`configure service vprn service-id l2tp group`

241.26.6527.91	Alc-Tunnel-Serv-Name	The service name of the routing instance that initiates a tunnel. **Note:** Alc-Tunnel-Serv-Name takes precedence over Alc-Tunnel-Id if both are specified.
26.6527.99	Alc-Ipv6-Address	DHCPv6 IA-NA IPv6 address (wan-host).
26.6527.104	Alc-Tunnel-Serv-Id	The service ID of the routing instance that initiates a tunnel. **Note:** Alc-Tunnel-Serv-Name takes precedence over Alc-Tunnel-Id if both are specified.
26.6527.105	Alc-Ipv6-Primary-Dns	The IPv6 address of the primary DNSv6 server.
26.6527.106	Alc-Ipv6-Secondary-Dns	The IPv6 address of the secondary DNSv6 server.
26.6527.131	Alc-Delegated-Ipv6-Pool	The name of an IPv6 IA-PD prefix pool (pd-host).
26.6527.161	Alc-Delegated-Ipv6-Prefix-Length	DHCPv6 IA-PD prefix length (pd-host).
26.6527.174	Alc-Lease-Time	The lease-time for proxy, in seconds.
26.6527.181	Alc-SLAAC-IPv6-Pool	The name of an IPv6 SLAAC prefix pool (wan-host).
26.6527.1036	Alc-SPI-Sharing	grouped AVP Sets the SLA Profile Instance (SPI) sharing method for this subscriber session to SPI sharing per group or default.
26.6527.1037	Alc-SPI-Sharing-Type	Must be included in an Alc-SPI-Sharing grouped AVP. Sets the SPI sharing method. value 0 = default as specified in the SLA profile with **def-instance-sharing**. The Alc-SPI-Sharing-Id AVP should not be present. value 2 = per group; the group identifier is specified with the Alc-SPI-Sharing-Id AV.
26.6527.1038	Alc-SPI-Sharing-Id	Must be included in an Alc-SPI-Sharing grouped AVP. Specifies the group identifier when SPI sharing is per group.

17.7.1 Sample configuration steps

To specify the peers to reach the Diameter NASREQ server in a diameter peer policy:

```
configure
    aaa
        diameter-peer-policy "diameter-peer-policy-1" create
            description "Diameter NASREQ peer policy"
            applications nasreq
            origin-host "bng@nokia.com"
            origin-realm "nokia.com"
            peer "peer-1" create
                address 172.16.3.1
                destination-realm "myDSCRealm.com"
                no shutdown
            exit
        exit
```

To specify the Diameter NASREQ application specific parameters, such as AVP format and values, in a Diameter application policy:

```
configure
    subscriber-mgmt
        diameter-application-policy "diameter-nasreq-policy-1" create
            description "Diameter NASREQ application policy"
            application nasreq
            diameter-peer-policy "diameter-peer-policy-1"
            nasreq
                user-name-format mac
                include-avp
                    circuit-id
                    nas-port-id
                    nas-port-type
                    remote-id
                exit
            exit
        exit
```

To apply the Diameter NASREQ application policy as Diameter authentication policy at a VPLS capture SAP, at an IES/VPRN group-interface or at a local user database:

Note: A Diameter authentication policy cannot be configured simultaneously with a RADIUS authentication policy on the same group-interface or capture SAP, nor for the same host in a local user database.

```
configure
    service
        vpls 10 customer 1 create
            sap 1/1/4:*.* capture-sap create
                ---snip---
                diameter-auth-policy "diameter-nasreq-policy-1"
        ies 1000 customer 1 create
            subscriber-interface "sub-int-1" create
                ---snip---
                group-interface "group-int-1-1" create
                ---snip---
                    diameter-auth-policy "diameter-nasreq-policy-1"
        vprn 2000 customer 1 create
```

```
            subscriber-interface "sub-int-1" create
                ---snip---
                group-interface "group-int-1-1" create
                ---snip---
                    diameter-auth-policy "diameter-nasreq-policy-1"

configure
    subscriber-mgmt
        local-user-db "ludb-1" create
            ipoe
                host "ipoe-host-1" create
                    ---snip---
                    diameter-auth-policy "diameter-nasreq-policy-1"
            ppp
                host "ppp-host-1" create
                    diameter-auth-policy "diameter-nasreq-policy-1"
```

If no AA-Answer message is received from the primary or secondary Diameter peer, then the host or session can be instantiated with the configured defaults. This is achieved by the following NASREQ application policy configuration:

```
configure
    subscriber-mgmt
        diameter-application-policy "diameter-nasreq-policy-1" create
            on-failure failover enabled handling continue
```

To enable flexible integration with different NASREQ servers, a Python policy can be configured on the Diameter node. The Python script can interact on the AVPs present in the AA-Request and AA-Answer messages.

```
configure
    python
        python-policy "py-policy-nasreq-1" create
            diameter aar direction egress script "NasreqAar"
            diameter aaa direction ingress script "NasreqAaa"
configure
    aaa
        diameter-peer-policy "diameter-peer-policy-1" create
            ---snip---
            python-policy "py-policy-nasreq-1"
```

17.8 Diameter redundancy

Diameter redundancy is supported on multiple levels:

- **High Availability (HA)**

  This refers to control plane redundancy with dual Control Plane Modules (CPMs) in a single chassis configuration. User sessions for Diameter applications are fully synchronized between CPMs.

  However, the TCP connections with peers are not synchronized and after a CPM switchover, the TCP connections need to be re-established. Once the TCP peering connections are re-established, communication on the user level (NASREQ/Gx/Gy) is resumed without the loss of any user sessions.

- **peer and server failover**

An existing user session can be rerouted to an alternate peer or server if the currently used peer or server fails.

- **multichassis redundancy**

  Diameter sessions are synchronized on the application level (through ESM if Gx and NASREQ) between two redundant SR nodes. Only one of those two nodes maintain a peering connection with the agent or server. As a result, the agent or server interacts with the redundant pair through a single peering connection.

17.8.1 Diameter peer and server failover

Peer and server failover mechanisms in the SR OS are concerned with retransmission of Diameter request messages and selection of alternative network paths toward the destinations during network failures. Both mechanisms (peer and server failover) assume that there are multiple peers, paths, and possibly redundant servers available in the network that can serve requests redundantly.

Peer failover involves the selection of the next best peer for a Diameter request message that have failed to be delivered over the current peer because of an explicit error notification or simply because of a peer failure. The selection of the next best peer in the SR node is based on forwarding and routing lookups performed in the Diameter base.

However, server failover is concerned with Diameter application servers (NASREQ, Gx, Gy) that may or may not be directly connected to an SR node. From an SR node viewpoint, the server failover procedure involves retransmission of request messages that are not successfully acknowledged by the Diameter application servers (NASREQ, Gx, Gy). Such unacknowledged request messages can be attributed to the loss of a specific Diameter application server and as a result be optionally retransmitted in a way that allows delivery of the retransmitted message to an alternate and possibly redundant Diameter application server.

Application answer messages do not rely on peer or server failover procedures because their forwarding is governed on a hop-by-hop basis (the exact reverse path of the request message).

17.8.1.1 Diameter peer failover

Each configured Diameter node in SR OS can support several peers that are simultaneously open. Only one of those peers is used to forward application messages for a specific user session. If there are multiple peers for the same realm, then the next-hop peer with the numerically lowest *preference* value is selected. If all peers have same preference for a specific realm, then the peer index is used to break the tie.

Peer failover is performed by a Diameter base protocol and is supported only for application request messages (for example, a peer failover does not apply to a CER message). Events that trigger peer failover can be categorized as:

- explicit notifications from a peer through an error message, informing the recipient of the error message that the peer cannot deliver the request message to the destination

  The receipt of the DIAMETER_UNABLE_TO_DELIVER (3002) or DIAMETER_TOO_BUSY (3004) protocol error messages triggers a retransmission of the original request messages to the next best peer. The retransmission bit (T-bit) in the retransmitted message is set. At the same time, the Tx timer (a configurable parameter in the SR OS) is started. Continued errored replies (3002/3004) for the retransmitted messages over newly selected peers continue to trigger the selection of the next best peer until:

– a valid peer is found
– all eligible peers are attempted without success
– the Tx timer expires

If there are no viable peers available to deliver the request message (3002/3004 errors received from all attempted peers), or the Tx timer expires, the Diameter base notifies the application layer (NASREQ, Gx, Gy) which may invoke a server failover procedure (if enabled on the application level by configuration).

- termination of an active peering connection

  A peering connection can be explicitly closed by either side because of a Disconnect-Peer-Request message attributed to the timeout of the watchdog messages or any other error on the TCP transport level. Termination of an active peer triggers a search for the next best peer through forwarding/routing for all messages in the transmit queue toward that peer. The T-bit is set on all retransmitted messages.

  If no eligible peers are left, the Diameter base notifies the application layer (NASREQ, Gx, Gy) which may invoke a server failover procedure (if enabled on the application level through configuration).

The destination-host AVP in the retransmitted messages because of the peer failover remains the same as in the original request message.

17.8.1.2 Diameter server failover

After the Diameter application is notified by the Diameter base that it is unable to deliver the message to the destination, the Diameter application can invoke a server failover procedure. A server failover procedure is enabled through configuration in the Diameter application policy (failure handling).

The server failover procedure on the Diameter application level provides the opportunity for one last retransmission of the original Diameter request messages, but this time with the destination-host AVP cleared in the retransmitted message. The T-bit in the retransmitted messages is set.

Clearing the destination-host AVP allows delivery of the application message to any server in the realm supporting that application. The retransmitted application request message is successfully processed on such server only if the applications servers are redundant and synchronized.

17.8.2 Diameter multichassis redundancy

Diameter multichassis redundancy offers a protection mechanism against network failures where the Diameter application sessions are synchronized and preserved after the chassis switchover.

Two redundant Diameter SR nodes appear as a single Diameter entity, while maintaining a single peering connection with each network peer (in other words, only one BNG in the pair forms a peering connection with a Diameter peer on the network side). With this configuration, each network node (agents and servers) maintains a single, unambiguous path toward the pair of redundant SR nodes. The SR node selected to maintain this peering connection is chosen through the Multi-Chassis Synchronization (MCS) protocol that runs between the two nodes. The decision about which node opens the peering connection is made on a per-peer basis, which means that the networks peering connections can be distributed over the pair of redundant SR nodes.

In a multichassis (or a dual-homed) environment, an ESM subscriber is instantiated in both SR nodes. However, traffic for the subscriber is steered (in upstream and downstream direction) through the SR node with the SRRP in the Active state. Because the Diameter Base in the SR OS is not aware of the ESM

subscribers and their (SRRP) states, an inter-peering Diameter connection is mandatory between the two redundant SR nodes. This peering connection is required to relay Diameter messages between the two redundant SR nodes if the SRRP is active on one node while the Diameter peering connection is open on the other node.

The concept of Diameter dual-homing is shown in Figure 230: Multi-chassis redundancy in DRA environment and Figure 231: Multi-chassis redundancy with directly connected servers.

Figure 230: Multi-chassis redundancy in DRA environment

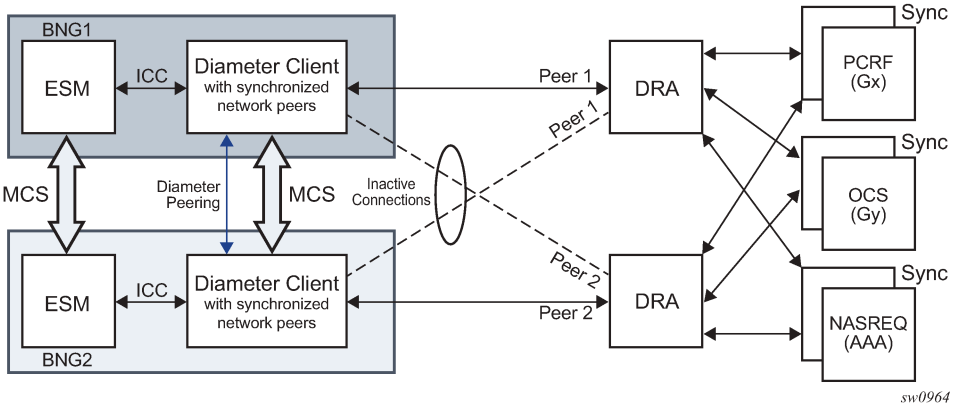
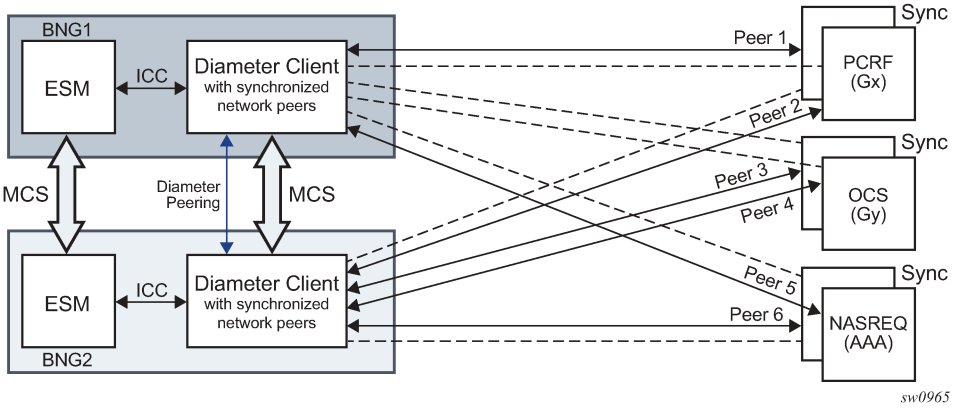


Figure 231: Multi-chassis redundancy with directly connected servers



17.8.2.1 Single Diameter Identity (DI) per a pair of redundant diameter nodes

Both SR nodes in the redundant pair are configured with the same Diameter Identity (DI), the origin host and origin realm. This configuration is performed per Diameter node configuration context, and there can be multiple Diameter node configuration contexts per an SR node or a pair of redundant SR nodes.

17.8.2.2 Single peering connection per redundant pair

Only one node in the SR redundant pair is instructed to open a peering connection, while the other node keeps the connection to the same peer closed (inactive). A single peering connection from a pair

of redundant SR nodes to the same Diameter network peer is essential for unambiguous transport of Diameter messages in an environment with:

- a single realm spanning multiple redundant pairs
- a pair of redundant SR nodes sharing the same Diameter Identity (DI)

If the open peering connection fails (such as a node reboot or a peer is closed or unavailable), the redundant SR node, upon detection of this event, immediately starts initiating the connection to the same peer.

17.8.2.3 Inter-peering connection

Inter-peering Diameter connection between the two redundant SR nodes is required to transport Diameter messages directly between them if the SR node with an active SRRP instance does not have a direct Diameter path to the agent or server. Instead, the messages must be sent over the inter-peering connection to the other SR node which is directly connected to the Diameter agent or server.

The synchronization mechanism between the two redundant nodes ensures that only one SR node initiates this inter-peer connection. The other (peering) SR node accepts this request but does not initiate it.

The watchdog timer for this peering connection should be set to the lowest configurable value.

17.8.2.4 Inter-peer as a default peer

If the Diameter request message is originated from a node (SRRP Active) that does not have direct connections to the agent or relevant Diameter servers, the request message must be sent to the inter peer which does have such connections. This is accomplished by designating the inter-peer as a default peer. A default peer is a peer used for all Diameter traffic that does not have an exact match for the destination-realm and application in the realm routing table. A default-peer is application (NASREQ/Gx/Gy) unaware. For routing loop avoidance information, see Diameter routing loop avoidance .

Because both redundant SR nodes have their default peers pointing to each other, a mechanism for routing loop prevention must be in place. This means that a Diameter message received from a peer, is never transmitted in the reverse direction back to this same peer (in other words, looped around).

17.8.2.5 Handling RARs

In a Diameter multichassis configuration, the SR node with SRRP in an Active state (where the subscriber is active) and the SR node with open relevant Diameter peering connections may not be co-located. As a result, it is possible that the RAR message is received on the SR node where the Diameter application (for example, Gx) is not active. In this case, the RAR message is re-routed from the default peer to the peering SR node where the SRRP is in the Active state, and the Diameter application is active.

If the default peer is not available, the RAR message is dropped and a DIAMETER_UNABLE_TO_DELIVER error message is sent back to the peer from which the RAR message was received.

17.8.2.6 Handling of the Route-Record AVP

An SR inserts the Route-Record AVP to all CCR messages that are passed through it. The Route-Record AVP is not added on the SR originating the message, but instead only on the transit SR, where messages

are received from the inter-peer and passed to the network, or when they are received from the network and passed to the inter-peer.

The value inserted in the Route-Record AVP is the origin-host of the peer from which the message is received.

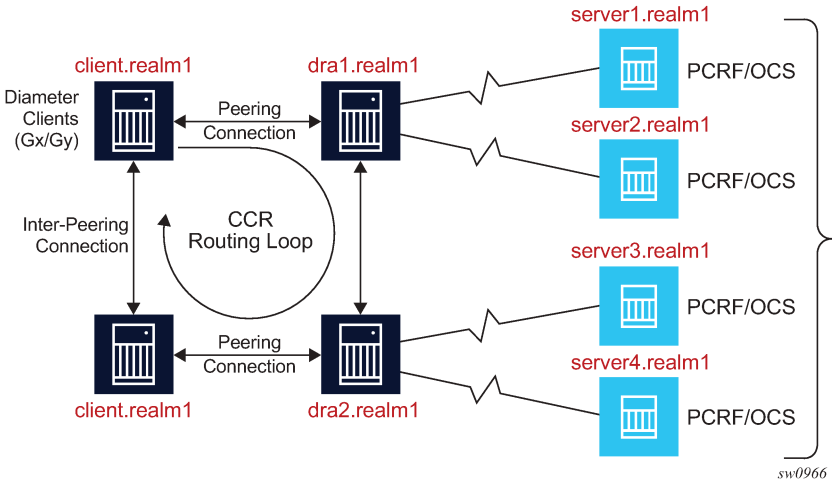
Checks that are performed on the Route-Record AVP are the following:

- If the peer's identity in any of the Route-Record AVPs in the received request messages matches the local host identity, then this means that a routing loop is detected (Figure 232: Diameter routing loop detection). The SR answers with the Result-Code AVP set to DIAMETER_LOOP_DETECTED.
- Route-Record AVPs are examined in received request messages to determine whether the message should or should not be forwarded. Only the next-hop (peers) candidates for the route that are not in any of the Route-Record AVPs of the received message, are allowed to be used. If there are no such next-hop candidates, an error message with the Result-Code AVP set to DIAMETER_UNABLE_TO_DELIVER must be sent back to the peer.

Scenarios in which an SR can receive a request message are:

- RAR from the inter-peer
- RAR from a network peer
- CCR-I/U/T message received from the inter-peer
- CCR-I/U/T message is received from a network peer

Figure 232: Diameter routing loop detection



17.8.2.7 SRRP switchover

SRRP activity and Diameter peer activity are independent of each other, and consequently, an SRRP switchover that is because of an access link or path failure does not cause Diameter peers to switchover. However, a SRRP switchover can trigger a Diameter application to notify the application server (for example, PCRF) of this event. This is performed through the AN-GW-IP AVP which is reported in CCR-U message sent to the PCRF. AN-GW-IP AVP carries the IP address of the newly active SRRP SR node, where the application session is now active. This functionality is enabled on the application level by arming the Diameter client on SR node with the event-trigger ID 13 (USER_LOCATION_CHANGE) from the PCRF.

17.8.2.8 Unsupported failures

The following two scenarios are not protected by the Diameter multichassis redundancy:

- Loss of the intra-peering (MCS) link. This scenario causes a split-brain scenario where both nodes act as stand-alone nodes and they both try to open peering connections to the Diameter servers.
- SRRP Active-Active state, where the SRRP messaging path between the two SR nodes is broken and consequently, both SR nodes own the Diameter application session (for example, Gx session).

These two scenarios do not meet the minimum requirements (MCS and SRRP must be operational) for successful Diameter multichassis operation, and therefore, are not supported. However, after the valid control channel states are restored (intra-chassis link is restored or SRRP states become stable, Active or Standby), the expectation is that the redundant system brings itself up to a valid protecting state.

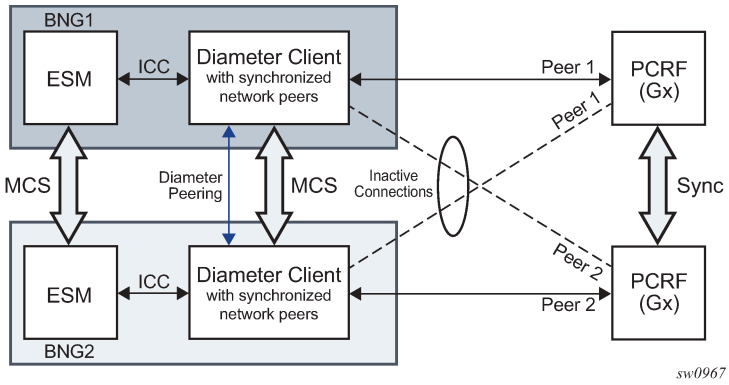
17.8.2.9 Peer preference in multichassis setup

Peer preference is a configuration parameter used to break the tie between multiple peers in the realm routing table leading to the same Diameter realm. Peer preference is a local configuration parameter and it does not transfer across the chassis. The peer with the numerically lowest preference value is preferred. As a result, peer preference may lose its meaning in the multichassis environment because the peers leading to the same realm may be distributed across both chassis.

The scenario can be clarified by the following example:

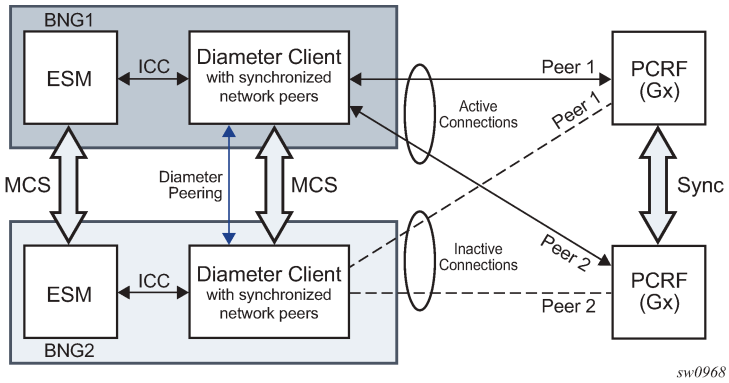
- Peer 1 and Peer 2 are redundant peers and Peer 1 has a lower preference than Peer 2 on both SR nodes.
- Because each SR node in Figure 233: Distributed peer connections has peering connection open to its own PCRF, the local peer preference loses its meaning. Peer 1 is used on BNG1 and Peer 2 is used on BNG2, although it may be needed that Peer 1 is always preferred over Peer 2.

Figure 233: Distributed peer connections



If the peering connections are colocated (Figure 234: Colocated peer connections), the preference parameter regains its meaning. In this case, Peer 1 is always selected over Peer 2.

Figure 234: Colocated peer connections



17.8.3 Gx Usage Monitoring in dual-homed systems

Gx Usage Monitoring (UM) relies on the Credit Control Instance (CCI) for the collection of byte counters. CCI is instantiated through a credit-category map (monitoring-key) and associated with the SLA profile instance (SPI) of the DHCP/PPPoE session. The bytes collected through CCI are periodically synchronized between the nodes. The synchronization of UM is automatically enabled for Gx/UM enabled subscribers in a dual-homed environment.

The direction of synchronization depends on the activity of a Gx sessions in ESM because UM is an integral part of Gx. Gx sessions in a dual-homed ESM environment follow the active or standby models, where the activity of Gx sessions can be distributed between the nodes, according to SRRP activity. SRRP as a basis of dual-homing in subscriber management is used to determine which node owns a Gx session. A Gx session maps to a IPoE/PPPoE session in a 1:1 manner. If multiple IPoE/PPPoE sessions share the same SLA profile instance (SPI), and with this the CCI, then the corresponding Gx sessions all share the same CCI. A Gx session under the SRRP instance in a master state is used to communicate with the PCRF servers, while the mirrored Gx session on the peering node with SRRP instance in the standby mode remains silent. In this context, Gx session can be considered as active or standby.

Although it is possible to have multiple Gx sessions per SPI, only one of the sessions can control UM at any time. In terms of UM, this Gx session can be referred to as a UM controlling Gx session. This is possible when there are multiple hosts or IPoE/PPPoE sessions per SPI, and each of them is associated with its own Gx session. Only when the UM is explicitly stopped via a session (RAR/CCA, or the UM controlling Gx session is terminated via CCR-T) can another Gx session take control over UM.

The initial quota for UM sent by the PCRF is always received by the active Gx session and the quota is periodically synchronized to the standby Gx session on the peering node. This is achieved with cooperation between ESM and Diameter in a dual-homed environment. When the quota is exhausted, only the active Gx session reports the exhausted quota to the server.

Gx communication with the DRA/PCRF is performed through Diameter Base and this functionality is described in Diameter base and Diameter multichassis redundancy.

17.8.3.1 Synchronization frequency

Synchronization frequency is configurable per diameter application policy and the minimum configurable value is five minutes. This interval is maintained per each individual Gx session associated with the relevant diameter application policy.

The configured synchronization interval governs the accuracy of UM in dual-homed systems where the switchover is caused by the sudden unavailability (crash, reboot, and so on) of the node on which a UM session is active.

Online configuration changes (while UM sessions are active) for synchronization interval are activated only for the new sessions that are established after the configuration is made.

17.8.3.2 Switchover triggered synchronization

A switchover, other than the one caused by a node loss, triggers an immediate synchronization of counters. This ensures better UM accuracy by synchronizing data up to the moment of the switchover, instead of up to the moment of the last periodic sync interval.

17.8.3.3 What is being synchronized

Credit quotas as received by the PCRF and UM counters are synchronized between dual homed nodes. However, the configuration is not synchronized or checked at the time of synchronization. This means that it is left to the operator to ensure that the UM configuration (for example, credit categories) is identical on both nodes.

17.8.3.4 Loss of inter-chassis link

Inter-chassis link (ICL) is used to transfer synchronization information. Any logical connectivity serves this purpose, but it is crucial that this link is redundant. The loss of this link is a faulty scenario where the control channel for UM synchronization between the two chassis is broken. In this case, both nodes remain operational in a stand-alone mode and they continue to forward traffic. After the ICL is recovered, the UM counters are re-synchronized from the active to standby node automatically within the configured synchronization interval.

17.8.3.5 Master-to-Master SRRP scenario

In a Master-to-Master scenario, the SRRP connectivity between the two nodes is lost and consequently both nodes transition into a master SRRP state. This means that in the downstream direction both nodes forward traffic (this depends on how the traffic is attracted on each node via routing). Similarly, in the upstream direction both nodes forward traffic, but the node selection is guided by the learning of VMAC (SRRP MAC) in the Layer 2 access network (the last received downstream packet determines the upstream path). The Master-to-Master state is, from the operational standpoint, an invalid and erroneous state where both nodes are sending synchronization messages to each other and at the same time they are rejecting them. The system recovers within one synchronization periodic interval after the valid SRRP state combination is restored (master/backup).

17.8.3.6 Usage counter collection with no credit grants received

Collecting usage counters without quota can occur when the original quota is exhausted and reported by the BNG, but the new quota is not given by Policy and Rule Charging Function (PCRF) while UM is still active (not explicitly disabled).

In this case, BNG continues counting. When the new quota is submitted, all consumed bytes up to this point are counted toward the new quota. For example, if 200 units are consumed since the last report, and then a new quota of 1000 units is issued, the 200 already consumed units are counted toward the new quota. For this reason, operators must avoid enabling UM without the quota. This can be achieved by:

- disabling UM if it is not needed
- submit a new quota by PCRF on every quota report exhaustion generated by the BNG

17.8.3.7 ISSU

The correct sequence of events during the (M)ISSU from a software release which does not support UM MCS (old release), to a software release that does support UM MCS (new release) is the following:

- The old release has ESM and Gx synchronization enabled but the UM does not. This is because UM synchronization is not supported on older releases.
- (M)ISSU is performed
- The new software release boots up with UM still disabled. UM must be enabled via Gx directives send from PCRF.
- The UM can now be enabled in the new release. When enabled, synchronization automatically takes place.

17.9 Diameter troubleshooting

The following sections describe the Diameter troubleshooting capabilities.

17.9.1 Operational commands

The following command displays the active Diameter Gx and Diameter Gy session IDs.

Note: The Diameter NASREQ application is stateless in SR OS and therefore a NASREQ session ID is not included in the output.

```
show service active-subscribers hierarchy diameter
```

Output example

```
===============================================================================
Active Subscribers Hierarchy (diameter information)
===============================================================================
-- ipoe-1
   (sub-profile-1)
   |
   +-- sap:[1/x1/1/c1/4:2211.2005] - sla:sla-profile-1
```

```
       |   Gy-diam-session-id:bng-gy.realm-2.com;1668427651;23
       |
       +-- IPOE-Session - mac:00:bb:02:00:00:05
           |   Gx-diam-session-id:bng-nasreq-gx.realm-1.com;1668427651;22
           |
           |-- 10.2.1.13
           |
           +-- 2001:db8:b002:100::1/128
               |
               +-- 2001:db8:b002:101::/64

-------------------------------------------------------------------------------
Number of active subscribers : 1
Flags: (N) = the host or the managed route is in non-forwarding state
===============================================================================
```

To display diameter sessions that are kept in the system for CCRT replay and that have no active subscriber session, use the following command.

```
show subscriber-mgmt diameter-session ccrt-replay
```

18 Python script support for ESM

18.1 Python script support for ESM

To provide programmable flexibility in ESM applications, the SR OS provides the following features with Python script support:

- sub-ident-policy
- radius-script-policy
- python-policy

18.2 Python in SR OS overview

SR OS supports both Python2 and Python3 for Python scripts. Python3 only supports python-policy.

The SR OS Python2 script support is based on Python version 2.4.2. Python2 has a set of language features (such as functions, lists, and dictionaries) and a large set of packages that provide most of the Python2 functionality. The SR OS keeps the language features intact and reduces the number of packages available to provide the operator with a small, but flexible, scripting language. The APIs described in this user guide are Python2 APIs.

The only feature removed from the Python2 language is unicode support. The only packages provided to the operator are:

- **alc**

  The SR OS-provided packages provide access to various ESM objects such as DHCPv4, DHCPv6, or RADIUS packets.

- **binascii**

  The binascii package contains common ASCII decoding, like base64.

- **re**

  The re package contains regular expression support.

- **struct**

  The struct package parses and manipulates binary strings

  .

- **md5**

  The MD5 package provides the MD5 message digest algorithm.

The MD5 package provides the MD5 message digest algorithm.

The SR OS Python3 script support is based on MicroPython version 3.4. Even though the majority of Python3 TPSDA APIs are similar to Python2 TPSDA API, there are some differences. For details about the Python3 TPSDA API, see SR BNG TPSDA Python3 API.

In addition to Python3 TPSDA API, the following Python3 libraries are also available:

- MicroPython libraries:
  - uarray
  - ubinascii
  - ucollections
  - uhashlib
  - uio
  - ure
  - ustruct
  - utime (Nokia modified the implementation of this module. For usage information, see *Nokia pySROS API*.)
- Standard libraries:
  - datetime
  - ipaddress
  - sys

18.2.1 Python policy – GTPv1-C API

The system provides a Python object for input GTPv1-C packet alc.gtp1c. The alc.gtp1c packet has following attributes to represent the GTPv1-C header fields as displayed in Table 70: GTPv1-C header fields .

Table 70: GTPv1-C header fields

Class attributes	GTPv1-C header field	Access
alc.gtp1c.version	Integer, version field in the header	Read
alc.gtp1c.ptbit	boolean, true mean 1, false means 0; PT bit in the header	Read
alc.gtp1c.ebit	boolean, E bit in the header	Read
alc.gtp1c.sbit	boolean, S bit in the header	Read
alc.gtp1c.pnbit	boolean, PN bit in the header	Read
alc.gtp1c.type	Integer message type field in the header	Read
alc.gtp1c.len	Integer message length field in the header	Read
alc.gtp1c.teid	Unsigned integer, or None if TEID does not exist	Read
alc.gtp1c.seq	Integer sequence number in the header	Read

Class attributes	GTPv1-C header field	Access
alc.gtp1c.npdu_number	Integer N-PDU Number field in the header	Read
alc.gtp1c.next_ext_type	Integer Next Extension Type field in the header	Read
alc.gtp1c.ext_list	Tuple, a tuple of extension headers; each element in the tuple represent one extension header: (ext_type, ext_content)	Read

The following is a list of alc.gtp1c functions:

- **alc.gtp1c.drop()**

  The system drops the resulting packet.
- **alc.gtp1c.getIEList()**

  The system returns a tuple of ie-type, each element represent one IE instance in the packet.
- **alc.gtp1c.get(ie_type)**

  The system returns a tuple of str, each str is the IE data of specified ie_type instances. If the specified IE does not exist, then return (); if the IE length=0, then return "" for this IE instance.
- **alc.gtp1c.set(ie_type, ie_tuple)**

  This function removes all instances of ie_type and insert a list of new IE instances. Each element in ie_tuple is a string, represent one instance of IE to be inserted. For each inserted IE instance, the type is ie_type. System uses the TV format for type<128 and TLV format for type>=128.
- **alc.gtp1c.clear(ie_type)**

  This function removes all existing instances of ie_type.

18.2.2 Python policy – GTPv2-C API

The system provides a Python object for input GTPv2-C packet: alc.gtp2c.

Shown in Table 71: GTPv2-C API , the alc.gtp2c packet has following attributes to represent the GTPv2-C header fields:

Table 71: GTPv2-C API

Class attributes	GTPv2-C header field	Access
alc.gtp2c.version	Integer version field in the header	Read
alc.gtp2c.pbit	boolean, true mean 1, false means 0; P bit in the header	Read
alc.gtp2c.tbit	boolean, T bit in the header	Read
alc.gtp2c.type	Integer message type field in the header	Read
alc.gtp2c.len	Integer message length field in the header	Read

Class attributes	GTPv2-C header field	Access
alc.gtp2c.teid	Unsigned integer, or None if TEID does not exist	Read
alc.gtp2c.seq	Integer sequence number in the header	Read

The following is a list of alc.gtp2c functions:

- **alc.gtp2c.drop()**

  The system drops the resulting packet.

- **alc.gtp2c.getIEList()**

  The system returns a tuple of 2-element tuple (ie-type, ie-instance). Each 2-element tuple represent one IE instance in the packet. For example: a packet with the following IEs:

  – type:2, instance: 1

  – type:2, instance: 1

  – type:74, instance: 0

  The system returns ((2,1),(2,1),(74,0))

  – **alc.gtp2c.get(ie_type, instance_id)**

  The system returns a tuple of str. Each str is the IE data of specified (ie-type, instance-id). If the IE is a grouped IE, then the str is also the content of whole grouped IE. If the specified (ie_type, instance_id) does not exist, then return (); if the IE length=0, then return "" for this IE instance.

  – **alc.gtp2c.set(ie_type, instance_id, ie_tuple)**

  This function removes all instances of (ie_type, instance_id) and inserts a list of new IE instances. Each element in an ie_tuple is a string, represent one instance of IE to be inserted. For each inserted IE instance, the type is ie_type, the instance is instance_id.

  – **alc.gtp2c.clear(ie_type, instance_id)**

  This function removes all existing instances of (ie_type, instance-id).

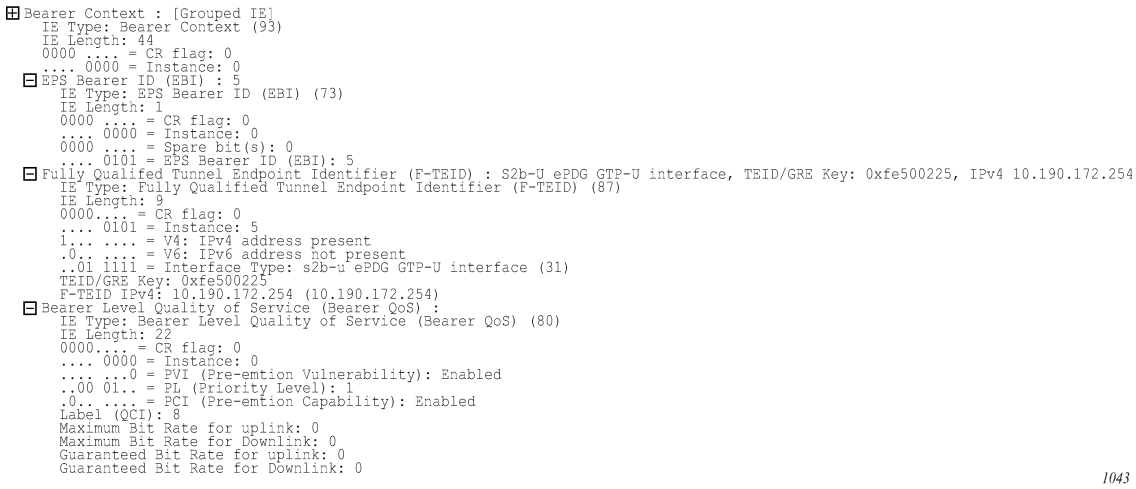
  – **alc.gtp2c.str_to_groupedIE (parent_ie_str)**

  Parses the parent_ie_str and return a tuple of embedded IEs. Parent_ie_str is a string of data field of parent IE; each element of returned tuple is a tuple of (ie_type, ie_instance, ie_data), which represent one embedded IE. For example, with the following code:

  - ie_str = alc.gtp2c.get(93,0)[0]
  - ie_tuple = alc.gtp2c.str_to_groupedIE(ie_str)

  The above code parses the first instance of the bearer context with an instance=0. Assume the bearer context looks like that shown in Figure 235: Bearer context example:

Figure 235: Bearer context example



Then, the ie_tuple is

((5,0,'\x05'),(87,5'\x9f\xfe\x50\x02\x25\x0a\xbe\xac\xfe'),
(80,0,'\x04\x08\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00'))

An exception is raised if the system cannot parse the parent_ie_str into a ie_tuple.

– **alc.gtp2c.groupedIE_to_str(grouped_ie_tuple)**

Converts the grouped_ie_tuple into a str and return the str (see the description of alc.gtp2c.str_to_groupedIE for the format grouped_ie_tuple). An exception is raised if the conversion fails. The following is an example:

- ie_str = alc.gtp2c.get(93,0)[0]
- ie_tuple = alc.gtp2c.str_to_groupedIE(ie_str)
- fteid = list(ie_tuple[1])
- fteid[1]=0
- ie_str = alc.gtp2c.groupdedIE_to_str((ie_tuple[0],tuple(ftied),ie_tuple[2]))
- alc.gtp2c.set(93,0,(ie_str,))

With the previous example packet, the code changes the instance of the embedded F-TEID to 0.

Be aware that the alc.gtp2c.str_to_groupedIE (parent_ie_str) and alc.gtp2c.groupedIE_to_str(grouped_ie_tuple) do not check whether the input parameter is a grouped IE. The user should verify that the input parameter is a grouped IE before using these two functions.

18.2.3 Python changes

The following changes have been made to Python for it to run on an embedded system:

- No files or sockets can be opened from inside Python scripts.
- No system calls can be made from inside Python scripts nor is the posix package available.
- The maximum recursion depth is fixed to twenty.

- The total amount of dynamic memory available for Python itself and Python scripts (excluding the Python cache) is capped at 8 Mb for 32-bit SR OS and 16 Mb for 64-bit SR OS on a CPM.
- The size of the Python script source file should be less than 60 kb.

18.3 Python support in sub-ident-policy

Legacy Python scripts used in subscriber identification policy can output these values:

- Subscriber identification string
- Subscriber profile string that needs to be mapped to the actual subscriber-profile name
- SLA profile string that needs to be mapped to the actual SLA-profile name
- ANCP string

Whenever a DHCP session belonging to a subscriber is activated (when a device is turned on), the device typically requests an IP address from the network. The DHCP ACK response from the DHCP server can be parsed and the contents of the message can be used to identify the class to which this session belongs, and therefore, the QoS and security settings to apply. The information necessary to select these settings can be codified in the DHCP options inserted by the access node.

Up to three URLs can be defined per subscriber identification policy. These are designated as primary, secondary, and tertiary. Each URL can be individually enabled or disabled. Only one script (the URL with the highest priority active script) is used at any one time to process DHCP ACK messages. If the system detects an error with a specified script, the URL is placed in an operationally down state. A script that is operationally or administratively down is considered inactive. The system automatically reverts to the highest priority active script. If a script becomes operationally down, it must be cycled through the administratively down and then the administratively up states for the system to attempt to reactivate the script.

The alc package contains a DHCP object, and has the following members (Table 72: DHCP object members).

Table 72: DHCP object members

Name	Read	Write	Class
htype	✓		integer
hlen	✓		integer
hops	✓		integer
flags	✓		integer
ciaddr	✓		integer
yiaddr	✓		integer
siaddr	✓		integer
giaddr	✓		integer
chaddr	✓		string

Name	Read	Write	Class
sname	✓		string
file	✓		string
options	✓		TLV
sub_ident		✓	string
sub_profile_string		✓	string
sla_profile_string		✓	string
ancp_string		✓	string
app_profile_string		✓	string
category_map_name		✓	string
int_dest_id		✓	string

The TLV type provides easy access to the value part of a stream of type-length-value variables, as is the case for the DHCP option field. In the example on Configuration, the circuit-ID is accessed as alc.dhcp.options[82][1].

Some DHCP servers do not echo the relay agent option (option 82) when the DHCP message was snooped instead of relayed. For the convenience of the operator, the relay agent option from the request message is returned when alc.dhcp.options[82] is called.

18.3.1 Configuration

Consider using script us5.py which sets the subscriber ID (sub_ident) based on the circuit-id in the DHCP message.

Configure the URL to this script in a **sub-ident-policy** as follows:

```
-----------------------------------------------
    sub-ident-policy "DSLAM" create
        description "Parse circuit IDs from different DSLAMs"
        primary

            script-url "ftp://198.51.100.0/py/us5.py"
            no shutdown
        exit
    exit
-----------------------------------------------
```

And attach this sub-ident-policy to the sub-sla-mgmt from a SAP:

```
A:dut-A>config>service>vpls>sap# info
-----------------------------------------------
    dhcp
        description "client side
        lease-populate 50
```

```
        no shutdown
    exit
    anti-spoof ip-mac
    sub-sla-mgmt
        sub-ident-policy "DSLAM"
        no shutdown
    exit
-----------------------------------------------
```

Note: DHCP snooping and relaying should be configured correctly for this to work.

18.3.2 Operator debugging

Verbose debug output is sent to debug-trace on compile errors, execution errors, execution output and the exported result variables.

```
A:dut-A>config>subscr-mgmt>sub-ident-pol>primary# script-url "ftp://xxx.xxx.xx.xx/py/
parsefail1.py"
A:dut-A>config>subscr-mgmt>sub-ident-pol>primary# no shutdown

1 2006/07/30 01:17:33.14 UTC MINOR: DEBUG #2001 - Python Compile Error
"Python Compile Error: parsefail1.py
    File "ftp://xxx.xxx.xx.xx/py/parsefail1.py", line 2 def invalid_function():
        ^
IndentationError: expected an indented block
"

A:dut-A>config>subscr-mgmt>sub-ident-pol>primary# script-url "ftp://xxx.xxx.xx.xx/py/
dump.py"

2 2006/07/30 01:24:55.50 UTC MINOR: DEBUG #2001 - Python Output
"Python Output: dump.py htype   = 0
hlen    = 0 hops    = 0 flags   = 0
ciaddr  = '0.0.0.0' yiaddr  = '0.0.0.0' siaddr  = '0.0.0.0' giaddr  = '0.0.0.0' chaddr  = ''
sname   = ''
file    = ''
options = '5\x01\x056\x04\n\x01\x07\n3\x04\x00\x00\x00\xb4\x01\x04\xff\xff\xff
\x00\x1c\x04\n\x02\x02\xffR\x0f\x01\rdut-A|1|1/1/1\xff' "

3 2006/07/30 01:24:55.50 UTC MINOR: DEBUG #2001 - Python Result
"Python Result: dump.py
"

A:dut-A>config>subscr-mgmt>sub-ident-pol>primary# script-url "ftp://xxx.xxx.xx.xx/py/end-
 less.py"

4 2006/07/30 01:30:17.27 UTC MINOR: DEBUG #2001 - Python Output
"Python Output: endless.py
"

5 2006/07/30 01:30:17.27 UTC MINOR: DEBUG #2001 - Python Error
"Python Error: endless.py
Traceback (most recent call last):
File "ftp://xxx.xxx.xx.xx/py/endless.py", line 2, in ? FatalError: script interrupted (timeout)
"
```

Note: All the Python Result events are empty because none of the scripts set any of the output variables.

18.3.3 Python scripts

Note:
The scripts in this section are test scripts and not scripts that the operator would normally use.

dump.py from alc import dhcp

```
def print_field(key, value):
print '%-8s = %r' % (key, value)

def ipaddr2a(ipaddr):
return '%d.%d.%d.%d' % (
(ipaddr & 0xFF000000) >> 24, (ipaddr & 0x00FF0000) >> 16, (ipaddr & 0x0000FF00) >> 8, (ipaddr &
 0x000000FF))

print_field('htype',   dhcp.htype) print_field('hlen',    dhcp.hlen) print_field('hops',
 dhcp.hops) print_field('flags',   dhcp.flags) print_field('ciaddr',  ipaddr2a(dhcp.ciaddr))
 print_field('yiaddr',  ipaddr2a(dhcp.yiaddr)) print_field('siaddr',  ipaddr2a(dhcp.siaddr))
 print_field('giaddr',  ipaddr2a(dhcp.giaddr)) print_field('chaddr',  dhcp.chaddr) print_
field('sname',   dhcp.sname) print_field('file',    dhcp.file) print_field('options',
 str(dhcp.options))
```

18.3.4 Sample Python scripts

This section provides examples to show how the script can be used in the context of Enhanced Subscriber Management.

Note: These scripts are included for informational purposes only. The operator must customize the script to match their own network and processes.

18.3.4.1 Example

This script uses the IP address assigned by the DHCP server to derive both *sub_ident* and *sla_profile_string*.

Script:

```
1.  import alc
2.  yiaddr = alc.dhcp.yiaddr
3.  # Subscriber ID equals full client IP address.
4.  # Note: IP address 10.10.10.10 yields 'sub-168430090'
5.  # and not 'sub-10.10.10.10'
6.  alc.dhcp.sub_ident = 'sub-' + str(yiaddr)
7.  # DHCP server is configured such that the third byte (field) of the IP
8.  # address indicates the session Profile ID.
9.  alc.dhcp.sla_profile_string = 'sp-' + str((yiaddr & 0x0000FF00) >> 8)
```

Explanation:

Line 1 imports the library "alc" – Library imports can reside anywhere in the script if the items are imported before they are used.

Line 2 assigns the decimal value of the host's IP address to a temporary variable "yiaddr". Line 6: The text "sub_" followed by yiaddr is assigned to "sub_ident" string.

Line 9 the text "sp-" followed with the third byte of the IP address is assigned to the "sla- profile" string.

If this script is run, for example, with a DHCP-assigned IP address of:

```
yiaddr = 10.10.0.2
```

The following variables are returned:

```
sub_ident: sub-168427522(hex = A0A00002 = 10.10.0.2)
sla_ident: sp-0
```

18.3.4.2 Example

This script returns the *sub_profile_string* and *sla_profile_string*, which are coded directly in the Option 82 string.

Script:

```
1.  import re
2.  import alc
3.  # option 82 formatted as follows:
4.  # "<subscriber Profile>-<sla-profile>"
5.  ident = str(alc.dhcp.options[82][1])
6.  alc.dhcp.sub_ident = ident
7.  tmp = re.match("(?P<sub>.+)-(?P<sla>.+)", str(ident))
8.  alc.dhcp.sub_profile_string = tmp.group("sub")
9.  alc.dhcp.sla_profile_string = tmp.group("sla")
```

Explanation:

Line 1-2 import the libraries "re" and "alc". Library imports can reside anywhere in the script if the items are imported before they are used.

Line 6 assigns the full contents of the DHCP Option 82 field to the "sub_ident" variable.

Line 7 splits the options 82 string into two parts, separated by "-".

Line 8 assigns the first part of the string to the variable "sub_profile_string".

Line 9 assigns the second part of the string to the variable "sla_profile_string".

If this script is run, for example, with DHCP option field:

```
options = \x52\x0D\x01\0x0Bmydsl-video
```

The following variables are returned:

```
sub_ident: mydsl-video
sub_profile_string: mydsl
sla_profile_string: video
```

18.3.5 Limitations

'%' operator

While %f is supported, %g and %e are not supported.

Floating Point Arithmetic

The floating point arithmetic precision on the box is less than the precision required by the regression suites of Python.

1024.*1024.*1024. until five numbers after the point instead of seven and sqrt(9) equals to 3. for the first seven numbers after the point.

Using the round operator fixes these problems. For example, round(pow(2., 30)) equals round(1024.*1024.*1024.) and round(sqrt(9)) equals 3.

18.4 RADIUS script policy overview

Python scripts for RADIUS AAA packets support manipulation in subscriber management application. This feature is supported on the 7750 SR and 7450 ESS routers.

A Python script can be executed in following cases:

- Before the system sends an access-request packet.
- After the system receives an access-accept packet.
- After the system receives an CoA-request packet.
- Before the system sends an accounting-request packet.

The input of the script is the corresponding original packet; and the output of the packet is used as the new corresponding packet for further ESM AAA process.

The **radius-script-policy** contains URLs of a primary and an optional secondary Python script, which could be a local CF file path or a FTP URL. The configured **radius-script-policy** could be used in different ESM polices like authentication-policy or radius-accounting-policy.

The following operations are supported within the script:

- Obtain the value of an existing attribute or VSA.
- Modify the value of an existing attribute or VSA.
- Add a new attribute or VSA.
- Remove an existing attribute or VSA.

Note: The following RADIUS attributes or VSA are read-only to Python script:

- Message-Authenticator
- Alc-LI-Action
- Alc-LI-Direction
- Alc-LI-Destination
- Alc-LI-FC
- Alc-LI-Intercept-Id

- Alc-LI-Session-Id

After Release 12.0R1, users should use a Python policy (instead of a RADIUS script policy) for RADIUS packet manipulation.

18.4.1 Python RADIUS API

The following new Python objects, alc.radius, have the following methods:

- drop(): Drop the packet
- header(): Return the dictionary object that includes RADIUS header information

alc.radius also provides methods to manipulate radius attributes via alc.radius.attributes, which has the following methods:

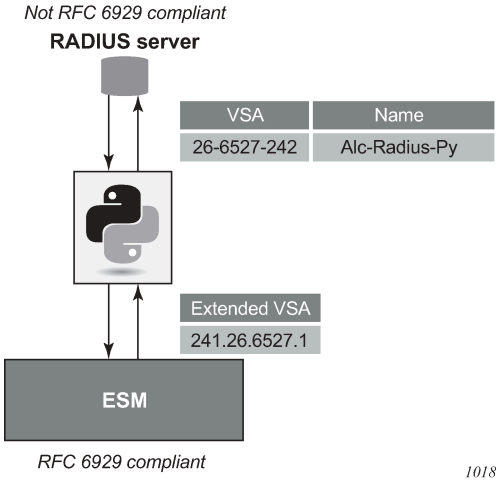
- get(type): Return the first attribute with specified type as a string
- getTuple(type): The same as above but returns a tuple of strings
- getVSA(vendor, type): Return the first VSA as a string
- getVSATuple(vendor, type): The same as above but returns a tuple of strings
- getExt(type, ext-type): Return a string containing the value of the first Extended Type attribute (RFC 6929, *Remote Authentication Dial In User Service (RADIUS) Protocol Extensions*) matching the specified type and extended type
- getExtTuple(type, ext-type): Return a tuple of strings. For each occurrence of an Extended Type attribute (RFC 6929) matching the specified type and extended type, the attribute value is included as a string entry in the tuple
- getExtVSA(type, vendor, vendor-type): Return a string containing the value of the first Extended Vendor Specific attribute (RFC 6929) matching the specified type, vendor and vendor-type
- getExtVSATuple(type, vendor, vendor-type): Return a tuple of strings. For each occurrence of an Extended Vendor Specific attribute (RFC 6929) matching the specified type, vendor and vendor-type, the attribute value is included as a string entry in the tuple.
- getLongExtTuple(type, ext-type): Return a tuple of strings. For each occurrence of a Long Extended Type attribute (RFC 6929) matching the specified type and extended type, the attribute value is included as a string entry in the tuple. The value of all fragments of the same Long Extended Type attribute are concatenated in a single string entry.
- getLongExtVSATuple(type, vendor, vendor-type): Return a tuple of strings. For each occurrence of a Long Extended Vendor Specific attribute (RFC 6929) matching the specified type, vendor and vendor-type, the attribute value is included as a string entry in the tuple. The value of all fragments of the same Long Extended Vendor Specific attribute are concatenated in a single string entry.
- set(type, value): Set the specified attribute to the value. The value must be either a string or a tuple of strings.
- setVSA(vendor, type, value): Set the specified VSA to the value. The value must be either a string or a tuple of strings.
- setExt(type, ext-type, value): Set the specified Extended Type attribute value (RFC 6929). The value can be specified as a string (single attribute) or a tuple of strings (multiple attributes).

- setExtVSA(type, vendor, vendor-type, value): Set the specified Extended Vendor Specific attribute value (RFC 6929). The value can be specified as a string (single attribute) or a tuple of strings (multiple attributes).
- setLongExt(type, ext-type, value): Set the specified Long Extended Type attribute value (RFC 6929). The value can be specified as a string (single attribute) or a tuple of strings (multiple attributes). When the string value length is too long to fit in a single attribute (more than 251 octets), then the value field is fragmented across multiple attributes. The M-bit (more) is set in all but the last fragment.
- setLongExtVSA(type, vendor, vendor-type, value): Set the specified Long Extended Vendor Specific attribute value (RFC 6929). The value can be specified as a string (single attribute) or a tuple of strings (multiple attributes). When the string value length is too long to fit in a single attribute (more than 246 octets), then the value field is fragmented across multiple attributes. The M-bit (more) is set in all but the last fragment.
- clear(type): Remove the specified attribute
- clearVSA(vendor, type): Remove the specified VSA
- clearExt(type, ext-type): Remove all Extended Type or Long Extended Type attributes (RFC 6929) matching the specified type and extended type
- clearExtVSA(type, vendor, vendor-type): Remove all Extended Vendor Specific or Long Extended Vendor Specific attributes (RFC 6929) matching the specified type, vendor and vendor-type
- isSet(type): Return True if the specified attribute exists, False otherwise
- isVSASet(vendor, type): Return True if the specified VSA exists, False otherwise
- isExtSet(type, ext-type): Return True if an Extended Type or Long Extended Type attribute (RFC 6929) matching the specified type and extended type is present. Else return False
- isExtVSASet(type, vendor, vendor-type): Return True if an Extended Vendor Specific or Long Extended Vendor Specific attribute (RFC 6929) matching the specified type, vendor and vendor-type is present. Else return False.

Extended attribute type, Long Extended attribute type, and Extended Vendor Specific data type are specified in RFC 6929 RADIUS Protocol Extensions.

Both the RADIUS client and RADIUS server must support RFC 6929 RADIUS Protocol Extensions to successfully use the Extended and Long Extended attribute types and the Extended Vendor Specific data type. Therefore, it is recommended to check the RADIUS servers for RFC 6929 support and upgrade if needed. As shown in Figure 236: RADIUS Python: interworking with RADIUS server with no RFC 6929 support, RADIUS Python can be used as a workaround to interwork with a RADIUS server that does not support these extensions. Use the general purpose [26-6527-242] Alc-Radius-Py VSA of type "octets" for this purpose.

Figure 236: RADIUS Python: interworking with RADIUS server with no RFC 6929 support



18.4.2 Sample script

From alc import radius

#1. Get the value of an existing Attribute

```
Username=radius.attributes.get(1)
```

#2. Modify an existing attribute

```
radius.attributes.set(1, 'Tom')
```

#3. Remove an existing attribute

```
radius.attributes.clear(1)
```

#4. Add a new attribute

```
radius.attributes.set(126,”WIFI-operator”)
```

18.5 Python policy overview

The Python policy represents a general framework to support all existing and new python features. A Python policy allows users to configure a Python script for a specified ESM packet type (such as DHCP and RADIUS) in a specified direction (ingress or egress). The system executes the configured script when sending or receiving the specified type of packet.

The corresponding original packet is used as input within the script. The user can use the system-provided API to manipulate the input packet (such as add, change, remove option or attribute). The changed packet is used as the output for further ESM processing. With a DHCP transaction cache, the script can return ESM attributes.

Python policies support the following ESM packet types and applications:

- DHCP Transaction Cache on CPM
- DHCPv4 on CPM
- DHCPv6 on CPM
- Diameter on CPM
- GTPv1-C/GTPv2-C
- Python cache on CPM
- RADIUS on CPM
- RADIUS on ISA

There are two types of Python policies:

- Centralized: Runs on the CPM and handles above protocol packets processed on the CPM.
- Distributed: Runs on the ISA and handles packets processed locally on the ISA (such as the DSM). Currently in RADIUS only.

The user must specify the type when creating the Python policy.

The following is an example configuration of a specified group interface. The system executes cf1:/dhcpv4.py after receiving DHCPv4 discovery and before the system forwards a DHCPv4 request packet.

```
config>python# info
----------------------------------------------
        python-script "dhcpv4" create
            primary-url "cf1:/dhcpv4.py"
            no shutdown
        exit
        python-policy "dhcp" create
            dhcp discover direction ingress script "dhcpv4"
            dhcp request direction egress script "dhcpv4"
        exit
----------------------------------------------
config>service>vprn>sub-if>grp-if>dhcp# info
----------------------------------------------
                        python-policy "dhcp"
                        server 10.9.9.9
                        lease-populate 100
                        gi-address 192.168.100.1
                        no shutdown
----------------------------------------------
```

18.5.1 Python policy – RADIUS API

The RADIUS API in Python policy uses the same API of the radius-script-policy.

18.5.2 Python policy – DHCPv4 API

- The system provides a Python object for input DHCPv4 packet: alc.dhcpv4.
- alc.dhcpv4 has following attributes to represent the DHCPv4 header fields:

  Table 73: alc.dhcpv4 attributes displays alc.dhcpv4 attributes.

Table 73: alc.dhcpv4 attributes

Class attributes	DHCPv4 header field	Access
alc.dhcpv4.pkt_len	Integer Total length of original DHCPv4 packet (UDP/IP header excluded, including pad option) in bytes	read
alc.dhcpv4.pkt_netlen	Integer Total length of original DHCPv4 packet (UDP/IP header excluded, pad option in the "options" field excluded) in bytes	read
alc.dhcpv4.op	op	read
alc.dhcpv4.htype	htype	read/write
alc.dhcpv4.hlen	hlen	read/write
alc.dhcpv4.hops	hops	read/write
alc.dhcpv4.xid	xid	read
alc.dhcpv4.secs	secs	read/write
alc.dhcpv4.flags	flags	read/write
alc.dhcpv4.ciaddr	ciaddr	read/write
alc.dhcpv4.yiaddr	yiaddr	read/write
alc.dhcpv4.siaddr	siaddr	read/write
alc.dhcpv4.giaddr	giaddr	read/write
alc.dhcpv4.chaddr	chaddr	read/write
alc.dhcpv4.sname	sname	read/write
alc.dhcpv4.file	file	read/write

All attributes, except alc.dhcpv4.pkt_len and alc.dhcpv4.pkt_netlen, are string with value of the actual bytes in the header.

The following is a list all functions of alc.dhcpv4:

- alc.dhcpv4.drop()
- alc.dhcpv4. getOptionList()
- alc.dhcpv4.pad(min_size=300)
- alc.dhcpv4. get(op_code)
- alc.dhcpv4. set(op_code,valTuple)
- alc.dhcpv4. clear(op_code)
- alc.dhcpv4.get_relayagent() / alc.dhcpv4.set_relayagent(D4OL)
- alc.dhcpv4.get_vendorspecific() / alc.dhcpv4.set_vendorspecific (D4OL)

DHCPv4 allows using sname and file fields to store options. However all DHCPv4 functions only operates with the "options" field. If a customer wants to manipulate options in the sname/file field, they need to do the parsing work in the script. (extended string.tlvxy method could help here)

- alc.dhcpv4. drop(): The system drops the result packet
- alc.dhcpv4. getOptionList(): Returns a tuple that includes the option-code of the existing top level DHCPv4 options in the packet.
  - The order of the element in the tuple is as same as the options that appear in the packet.
  - If there are multiple instances of the same option, then each instance is one element in the tuple.
  - Pad option(0) is excluded.
  - End option(255) is included
  - Example: A DHCP discovery packet with msg-type/lease-time/request-addr/parameter-request-list/ agent-info/end returns (53,51,50,55,82,255)
- alc.dhcpv4.pad(min_size=300): This function pads the resulting DHCPv4 packet to the specified min_size with pad option(0) after executing the whole script. Padding is not added if the result packet is already >=min_size. The default value of min_size is 300. Although not defined in DHCPv4 RFC, many DHCPv4 implementation has a minimal length requirement of 300 bytes. So this function could pad the result packet to the specified min_size.
- alc.dhcpv4. get(op_code): Returns a tuple that includes all instances of the specified top level option as a string. The value of this string is the exact bytes of the option as it appears in packet(excludes option-code and option-len).
  - If the specified option does not exist, then the function returns ()
  - If the instance of the specified option does not have the value (len=0 or does not have len and value part), then the function returns "" for that instance in the tuple.
  - Example: There is an address lease time option(51) in the packet, with value 60, then alc.dhcpv4.get(51) should return: ('\x00\x00\x00\x3c',)
- alc.dhcpv4. set(op-code,valTuple): This function removes all the existing instances of specified top level option and insert a list of new options. Each element in valTuple is a string, representing one instance of the new option to be inserted; For each new option, the option-code is specified in op-code. The option-len is the length of the element. The rest of option is the element itself.

  Example: To insert an address lease time option(51) in the packet, with the value 60; use alc.dhcpv4.set(51, ('\x00\x00\x00\x3c',))
- alc.dhcpv4. clear(op-code): This function removes all the existing instances of specified top level option.
- Although alc.dhcpv4.get() and alc.dhcpv4.set() provide a generic way to manipulate DHCPv4 top level options, but some DHCPv4 options have a complex/hierarchical structure like option82 and option43. To provide a friendly access to these kinds of options, the system provides the following options' specific functions:
  - alc.dhcpv4.get_relayagent() / alc.dhcpv4.set_relayagent(D4OL)
  - alc.dhcpv4.get_vendorspecific() / alc.dhcpv4.set_vendorspecific (D4OL)

  All alc.dhcpv4.get_XXX() returns a data structure:"D4OL" (DHCPv4 Option List)
  - D4OL is a list. Each element in the list represents an instance of that option. For example, if there are 3 option82 in the "options" field of packet, then get_relayagent() returns a list of 3 elements. Each element represents one instance of the option in the packet.
  - Each element in D4OL is a dict (called dict as "D4O" in this example):

The key of D4O is the sub-option- code. The value is a list of strings of sub-option-value of all instance of the sub-option.

All alc.dhcpv4.set_XX(OPDL) accepts D4OL as the parameter. Remove all existing instances of the corresponding options and then insert the new options represented by specified D4OL.

Examples:

For a packet with an option82 like following

```
    Option: (82) Agent Information Option
        Length: 22
        Option (82) Suboption: (1) Agent Circuit ID
            Length: 8
            Agent Circuit ID: 4a616e737656e73
        Option 82 Suboption: (2) Agent Remote ID
            Length: 10
            Agent Remote ID 62617369632364617461
```

The option-data is (hex formatted) "01:08:4a:61:6e:73:73:65:6e:73:02:0a:62:61:73:69:63:23:64:61:74:61”

The following is an example script:

```
import alc

option82_list=alc.dhcpv4.get_relayagent()
#option82_list will be [{1:['\x4a\x61\x6e:73\x73\x65\x6e\x73',],2:['\x62\x61\x73\x69\x63\x23\
x64\x61\x74\x61',]},]
Option82_list[0][2][0]='basic#video' #change remote-id to 'basic#video'
alc.dhcpv4.set_relayagent(option82_list)#update the option82
```

18.5.3 Python policy – DHCPv6 API

- The system provides a Python object for input DHCPv6 packet: alc.dhcpv6
- alc.dhcpv6 has following attributes to represent the DHCPv6 header fields:

  Table 74: DHCPv6 header fields displays DHCPv6 header fields.

Table 74: DHCPv6 header fields

Class attributes	DHCPv6 header field	Client/server msg	Relay msg	Access
alc.dhcpv6.pkt_len	Integer, Total length of original DHCPv4 packet(UDP/IP header excluded) in bytes	✓	✓	Read
alc.dhcpv6.msg_type	msg-type	✓	✓	Read
alc.dhcpv6.transaction_id	transaction-id	✓		Read
alc.dhcpv6.hop_count	hop-count		✓	read/write
alc.dhcpv6.link_addr	link-address		✓	read/write

Class attributes	DHCPv6 header field	Client/server msg	Relay msg	Access
alc.dhcpv6.peer_addr	peer-address		✓	read/write

All header fields (as the attribute of alc.dhcpv6 class) are strings (except pkt_len) with exact bytes as it appears in the packet.

If the attribute does not exist in the specified msg-type, for example if the link_attr does not exist in client/server message(C/S msg), then its value should be **None**.

- The following is a list of all functions in the class:
  - alc.dhcpv6.drop()
  - alc.dhcpv6. getOptionList()
  - alc.dhcpv6. get(op-code)
  - alc.dhcpv6. set(op-code,valTuple)
  - alc.dhcpv6. clear(op-code)
  - alc.dhcpv6.get_iana() / alc.dhcpv6.set_iana(OPDL)
  - alc.dhcpv6.get_iata() / alc.dhcpv6.set_iata(OPDL)
  - alc.dhcpv6.get_vendoropts() / alc.dhcpv6.set_vendoropts(OPDL)
  - alc.dhcpv6.get_iapd() / alc.dhcpv6.set_iapd(OPDL)
  - alc.dhcpv6.get_relaymsg()
  - alc.dhcpv6.set_relaymsg(packet)
- alc.dhcpv6. drop(): The system drops the resulting packet.
- alc.dhcpv6. getOptionList(): Returns a tuple that includes the option-code of the existing top level DHCPv6 options in the packet. The order of the element in the tuple is as same as the options appear in the packet. If there are multiple instances of same option, then each instance is one element in the tuple. For example:
  - A C/S Msg with Elapsed Time/Client Identifier/IANA/FQDN/Vendor Class/Option Request returns (8,1,3,39,16,6).
  - A Relay Msg with Relay Message option only returns (9)
- alc.dhcpv6. get(op-code): Returns a tuple that includes all instances of the specified top level option as string The value of this string is the exact bytes of the option as it appears in packet (excludes option-code and option-len). If the specified option does not exist in the input packet, then it returns ().

  Examples:

  If there is a Status Code option in the packet, status-code 0 and status-msg:”Address Assigned”; then alc.dhcpv6.get(13) should return: (‘\x00\x00Address Assigned’,)
- alc.dhcpv6. set(op-code,valTuple): This function removes all the existing instances of the specified top level option and insert a list of new options. Each element in valTuple is a string, representing one instance of the new option to be inserted. For each new option, the option-code is specified in op-code, the option-len is the length of the element, reset of option is the element itself.

  To insert a Status Code options with status-code 0 and status-msg:”Address Assigned”; use alc.dhcpv6.set(13, (‘\x00\x00Address Assigned’,))
- alc.dhcpv6. clear(op-code): This function removes all the existing instances of specified top level option.

- Although alc.dhcpv6.get() and alc.dhcpv6.set() provide a generic way to manipulate DHCPv6 top level options, but some DHCPv6 options have more complex/hierarchical structure like IA_NA/IA_TA, and so on. To provide a friendly access to these kinds of options, the system provides the following options specific functions:
  - alc.dhcpv6.get_iana() / alc.dhcpv6.set_iana(OPDL)
  - alc.dhcpv6.get_iata() / alc.dhcpv6.set_iata(OPDL)
  - alc.dhcpv6.get_vendoropts() / alc.dhcpv6.set_vendoropts(OPDL)
  - alc.dhcpv6.get_iapd() / alc.dhcpv6.set_iapd(OPDL)

  All alc.dhcpv6.get_XXX() returns a data structure:"OPDL" (Option Data structure List)
  - OPDL is a list. Each element in the list represents an instance of that option. For example, if there are 3 IANA in the packet, then get_iana() returns a list of 3 elements, each element represent one instance of IANA option in the packet.
  - Each element in OPDL is a list, referred to as "OPD" in this list.
    - Each element in OPD represent one field in the option(option-code and option-len are not included), the order of the element is as same as the fields appear in the option
    - For field that could be parsed into sub-option by RFC, then the element is a dict, the key of this dict is the sub-option type, if sub-option is one of following supported-sub-option, the value to the key is a sub-option_OPDL represent the list of that specific sub-option
    - IAADDR(5)
    - Status Code(13)
    - IAPREFIX(26)

    Else, if the sub-option is not one of the above, then the value to the key is a list of string of sub-option-data, each string represent one instance of the sub-option.
    - The structure of sub-option_OPDL of IAADDR is: [[v6_addr,prefer_lifetime, valid_lifetime,sub-option_OPDL],]
    - The structure of sub-option_OPDL of Status Code is: [[status-code,status-msg], and so on]
    - The structure of sub-option_OPDL of IAPREFIX is: [[prefer_lifetime,valid_lifetime,prefix-len,v6prefix,sub-option_OPDL],..]
    - For the field (by RFC definition) could be parsed into sub-options, but it does not actually exist, then the dict is empty {}
    - For field that cannot be parsed into sub-option by RFC, the element is a string of exact bytes of that field

All alc.dhcpv6.set_XX(OPDL) accepts an OPDL as the parameter. Remove all existing instances of the corresponding options and then insert new options represented by the specified OPDL.

- alc.dhcpv6.get_iana()/alc.dhcpv6.get_iana(OPDL)
- The general OPDL structure for these two functions is: [[IAID_val,T1_val,T2_val, sub-option_dict]]
- The structure of sub-option_dict is:{sub-option-type:sub-option_val}
- If sub-option is supported-sub-option, then sub-option_val is a sub-option_OPDL
- For all other sub-options, the sub-option_val is a list of string of sub-option-data

Examples:

For a packet with an IANA option as in the following:

```
⊟ Identity Association for Non-temporary Address
    Option: Identity Association for the Non-temporary Address (3)
    Length: 40
    Value: 0ff0def10002a3000004380000050018200105586045004 7...
    IAID: 0ff0def1
    T1: 172800
    T2: 276480
  ⊟ IA Address: 2001:558:6045:47:45cc:d9f2:5727:eae0
      Option: IA Address (5)
      Length: 24
      Value: 200105586045004745ccd9f25727eae00005460000054600
      IPv6 address: 2001:558:6045:47:45cc:d9f2:5727:eae0
      Preferred lifetime: 345600
      Valid lifetime: 345600
```

sw0005

The option-data is (hex formatted)

```
"0f:f0:de:f1:00:02:a3:00:00:04:38:00:00:05:00:18:20:01:05:58:60:45:00:47:45:cc:d9:f2:57:27:ea:e0:00:05:46:00:
```

The following is an example script:

```
import alc

iana_list=alc.dhcpv6.get_iana()
#iana_list will be [['\x0f\xf0\xde\xf1','\x00\x02\xa3\x00','\x00\x04\\x38\x00',{5:[['\x20\x01\
x05\x58\x60\x45\x00\x47\x45\xcc\xd9\xf2\x57\x27\xea\xe0','\x00\x05\x46\x00','\x00\x05\x46\x00',
{}]]}]]
iana_list[0][1]='\x00\x00\x04\xb0' #change T1 to 1200
alc.dhcpv6.set_iana(iana_list)#update the iana
```

- alc.dhcpv6.get_iata()/alc.dhcpv6.get_iata(OPDL)
  - The general OPDL structure for these two functions is: [[IAID_val, sub-option_dict]]
  - The structure of sub-option_dict is:{sub-option-type:sub-option_val}
  - If sub-option is supported-sub-option, then sub-option_val is a sub-option_OPDL
  - For all other sub-options, the sub-option_val is a list of string of sub-option-data

  Examples: These two function are very similar with IANA, so the examples are skipped here.
- alc.dhcpv6.get_iapd()/alc.dhcpv6.get_iapd(OPDL)
- The general OPDL structure for these two functions is: [[IAID_val,T1_va1,T2_val, sub-option_dict]]
- The structure of sub-option_dict is:{sub-option-type:sub-option_val}
- If sub-option is supported-sub-option, then sub-option_val is a sub-option_OPDL
- For all other sub-options, the sub-option_val is a list of string of sub-option-data

  Examples: For a packet with IA_PD as in the following:

```
⊟ Identity Association for Prefix Delegation
    Option: Identity Association for Prefix Delegation (25)
    Length: 41
    Value: 0000000100000708000000b40001a001900000e1000015180...
    IAID: 00000001
    T1: 1800
    T2: 2880
  ⊟ IA Prefix
      Option: IA Prefix (26)
      Length: 25
      Value: 00000e10000151803820010db80002000000000000000000...
      Preferred lifetime: 3600
      Valid lifetime: 86400
      Prefix length: 56
      Prefix address: 2001:db8:2::
```

sw0006

The option-data is (hex formatted)

```
"00:00:00:01:00:00:07:08:00:00:0b:40:00:1a:00:19:00:00:0e:10:00:01:51:80:38:20:01:0d:b8:00:02:00:00:00:00:
```

- Following is an example script:

```
import alc

iapd_list=alc.dhcpv6.get_iapd()
#iapd_list will be [['\x00\x00\x00\x01','\x00\x00\x07\x08','\x00\x00\x0b\x40',{26:[['\x00\
x00\x0e\x10','\x00\x01\x51\x80','\x38', '\x20\x01\x0d\xb8\x00\x02\x00\x 00\x00\x00\x00\x00\
x00\x00\x00\x00',{}]]}]]
iapd_list[0][1]='\x00\x00\x04\xb0' #change T1 to 1200
alc.dhcpv6.set_iapd(iapd_list)#update the iapd
```

- alc.dhcpv6.get_vendoropts()/alc.dhcpv6.get_vendoropts (OPDL)
  - The general OPDL structure for these two functions is: [[enterpriseid_val, sub-option_dict]]
  - The structure of sub-option_dict is:{sub-option-type:sub-option_val}
  - If sub-option is supported-sub-option, then sub-option_val is a sub-option_OPDL
  - For all other sub-options, the sub-option_val is a list of string of sub-option-data

Examples: For a packet with vendor options as in the following:

```
⊟ Vendor-specific Information
    Option: Vendor-specific Information (17)
    Length: 40
    Value: 0000197f0001000969612d6e615f30303100020009696 12d...
    Enterprise ID: Panthera Networks, Inc. (6527)
  ⊟ option
      option code: 1
      option length: 9
      option-data
  ⊟ option
      option code: 2
      option length: 9
      option-data
  ⊟ option
      option code: 3
      option length: 1
      option-data
  ⊟ option
      option code: 4
      option length: 1
      option-data
```

sw0007

The option-data is (hex formatted)

```
"00:00:19:7f:00:01:00:09:69:61:2d:6e:61:5f:30:30:31:00:02:00:09:69:61:2d:70:64:5f:30:30:31:00:03:00:01:38:00:
```

The following is an example script:

```
import alc

vendoropts_list=alc.dhcpv6.get_vendoropts()
# vendoropts_list will be [['\x00\x00\x19\x7f', {1:['\x69\x61\x2d\x6e\x61\x5f\x30\x30\x31'],2:
[ '\x69\x61\x2d\x70\x64\x5f\x30\x30\x31'],3:['\x38'], 4:['\x40']}]]
iapd_list[0][1][4][0]='\x60' #change sub-option 4’s value to 0x60
alc.dhcpv6.set_vendoropts(vendoropts_list)#update the vendor options
```

- For DHCPv6 relay message, the "Relay Message” option embedded a full DHCPv6 packet and the embedded packet could itself have a "Replay Message” option which embedded another DHCPv6 packet.

  To provide direct access to embedded DHCPv6 packet, the system provides following functions:

  – alc.dhcpv6.get_relaymsg()

  – alc.dhcpv6.set_relaymsg(packet)
- alc.dhcpv6. get_relaymsg(): This function returns a populated alc.dhcpv6 object. which means the returned object was initialized with the DHCPv6 packet embedded in "Relay Message” option as the input.
- alc.dhcpv6. set_relaymsg(packet): This function accepts an alc.dhcpv6 object as a parameter. This object replaces existing "Relay Message” option.

Example-1 script for single relay message:

```
import alc
#input packet is a relay-reply msg
embed_dhcpv6_packet=alc.dhcpv6.get_relaymsg()
iana_list=embed_dhcpv6_packet.get_iana()
iana_list[0][1]='\x00\x00\x04\xb0' #change T1 to 1200
embed_dhcpv6_packet.set_iana(iana_list)#update the iana of the embedded packet
alc.dhcpv6.set_relaymsg(embed_dhcpv6_packet)#update the Relay Message option
```

Example-2 script for double relay message (relay of relay):

```
import alc
#input packet is a relay-reply msg
embed_lv1_packet=alc.dhcpv6.get_relaymsg() #get the level=1 embedded packet
embed_lv2_packet= embed_lv1_packet.get_relaymsg()#get level-2 packet embedded in level-1 packet
iana_list=embed_dhcpv6_packet.get_iana()
iana_list[0][1]='\x00\x00\x04\xb0' #change T1 to 1200
embed_dhcpv6_packet.set_iana(iana_list)#update the iana
embed_lv1_packet.set_relaymsg(embed_lv2_packet)#update the Relay Message option of lv1 msg
alc.dhcpv6.set_relaymsg(embed_lv1_packet)#update the Relay Message option of the top level msg
```

18.5.4 Python policy – Diameter API

The alc.diameter Python module provides an API for Diameter message manipulation.

Terminology used in the API description:

- **top-level-AVP**

AVP appearing at the top level in a Diameter message, in other words, not embedded in the Data field of a grouped AVP

- **embedded-AVP**

  AVP embedded in the Data field of a grouped AVP. An embedded AVP can be a grouped AVP. This is called nesting.

- **AVP-tuple**

  Python tuple with following values:

  (AVP code, vendor ID, flags)

  – AVP code: integer, AVP header field

  – Vendor ID: integer, AVP header field

  – flags: string, AVP header field

- **AVP-value-tuple**

  Python tuple with following values:

  (flags, data)

  – flags: string, AVP header field

  – data: string, AVP data field

- **AVP-key-tuple**

  Python tuple with following values:

  (AVP code, Vendor ID)

  – AVP code: integer, AVP header field

  – Vendor ID : integer, AVP header field

- **grouped-AVP-value-tuple**

  Python tuple with following values:

  (flags, grouped-AVP-dictionary)

  flags: string, AVP header field

- **grouped-AVP-dictionary**

  Python dictionary with following key:value pairs:

  {AVP-key-tuple: [AVP-value-tuple or grouped-AVP-value-tuple, …], ... }

  – key = AVP-key-tuple

  – value = list of AVP-value-tuples or list of grouped-AVP-value-tuples

- **grouped-AVP-decode-tuple**

  Python tuple with the following values (AVP-key-tuple, …)

  tuple of AVP-key-tuples

- **AVP-order-tuple**

  Python tuple with the following values (AVP-key-tuple, …)

  tuple of AVP-key-tuples

Table 75: Diameter message header alc.diameter attributes displays attributes available in **alc.diameter** module providing access to the Diameter message header:

Table 75: Diameter message header alc.diameter attributes

Attribute	Description	Type	Access
application_id	Diameter message header field	string	Read/Write
code	Diameter message header field	string	Read/Write
end_to_end_id	Diameter message header field	string	Read/Write
flags	Diameter message header field	string	Read/Write
hop_by_hop_id	Diameter message header field	string	Read/Write
msg_length	Diameter message header field. The value is the message length of the original diameter message.	string	Read
version	Diameter message header field	string	Read/Write

Table 76: Message and AVP manipulation functionality alc.diameter methods displays methods available in **alc.diameter** module providing message and AVP manipulation functionality:

Table 76: Message and AVP manipulation functionality alc.diameter methods

Method	Description
clear_avps(AVP code, vendor id) AVP code, vendor id = top-level-AVP	Remove all instances of the specified AVP from the message. Applies to top-level-AVP's only. If the specified AVP is not present, no python error is generated. Vendor ID value zero matches top-level-AVP's without vendor ID field. Return value: None For example: diameter.clear_avps(256, 12645)
drop()	Drop the Diameter message. Packet is consumed at TCP level (ack send). A drop triggers retransmits on Diameter level. Return value: None For example: diameter.drop()
get_avps(AVP code, vendor id) AVP code, vendor id = top-level-AVP	Returns a list of AVP-value-tuples. Each AVP-value-tuple represents an instance of the specified AVP in the message. Applies to top-level-AVP's only. The position in the list corresponds with the position of the AVP instance in the message at that stage in the script. When executed before any clear or set AVP method, the list order corresponds with the AVP order in the received message. If the specified AVP is a grouped AVP, then the data contains all the embedded-AVP's. An empty list is returned

Method	Description
	if the specified AVP is not present. Vendor ID value zero matches top-level-AVP's without vendor ID field. For example: diameter.get_avps(263, 0) [('@', 'bng.nokia.com;1398156449;28')]
get_avps_list()	Returns a list of AVP-tuples. Each AVP-tuple represents an instance of an AVP in the message. Applies to top-level-AVP's only. The position in the list corresponds with the position of the AVP in the message at that stage in the script. When executed before any clear or set AVP method, the list order corresponds with the AVP order in the received message. When multiple instances of an AVP are present in the message, then there are multiple instances in the list. The vendor ID has value zero when not present. Grouped AVPs cannot be distinguished from other AVPs in the list. For example: diameter.get_avps_list() [(263, 0, '@'), (264, 0, '@'), (296, 0, '@'), (258, 0, '@'), (416, 0, '@'), (415, 0, '@'), (268, 0, '@'), (55, 0, '@'), (456, 0, '@'), (456, 0, '@'), (456, 0, '@'), (293, 0, '@'), (256, 12645, '\x80')]
get_grouped_avps(AVP code, vendor id, grouped-AVP-decode-tuple) AVP code, vendor id = top-level-AVP	Returns a list of grouped-AVP-value-tuples with each grouped-AVP-dictionary entry representing an embedded AVP. Each grouped-AVP-value-tuple represents an instance of the specified AVP in the message. Applies to top-level-AVPs of type grouped only. The position in the list corresponds with the position of the grouped AVP instance in the message at that stage in the script. When executed before any clear or set AVP method, the list order corresponds with the AVP order in the received message. If the grouped-AVP-decode-tuple is empty, only the specified top-level-AVP is expanded in a grouped-AVP-value-tuple, with each grouped-AVP-dictionary entry representing an embedded AVP and all dictionary values of type "list of AVP-value-tuples" To expand nested AVPs (grouped AVPs embedded in a grouped AVP), the grouped top-level-AVP and grouped embedded-AVP to expand must be added to the grouped-AVP-decode-tuple. All grouped AVPs in the grouped-AVP-decode-tuple are expanded in a list of grouped-AVP-value-tuples provided that their embedding AVP is also in the list. The position of the embedded AVPs in the grouped-AVP-dictionary does not correspond with the position in the grouped AVP. If the specified top-level-AVP is not a grouped AVP, then a Python error is generated: "ValueError: malformed diameter message". For example: To expand the Multiple Services Credit Control (456) grouped top level AVP:

Method	Description
	diameter.get_grouped_avps(456,0,()) [('@', {(432, 0): [('@', '\x00\x00\x00\x01')], (431, 0): [('@', '\x00\x00\x01\xa4@\x00\x00\x0c\x00\x00\x00d')], (448, 0): [('@', '\x00\x00\x04\xb0')], (268, 0): [('@', '\x00\x00\x07\xd1')]}), ('@', {(432, 0): [('@', '\x00\x00\x00\x02')], (431, 0): [('@', '\x00\x00\x01\xa4@\x00\x00\x0c\x00\x00\x03\x84')], (448, 0): [('@', '\x00\x00\x04\xb0')], (268, 0): [('@', '\x00\x00\x07\xd1')]}), ('@', {(432, 0): [('@', '\x00\x00\x00\x03')], (431, 0): [('@', '\x00\x00\x01\xa4@\x00\x00\x0c\x00\x00\x00<')], (448, 0): [('@', '\x00\x00\x04\xb0')], (268, 0): [('@', '\x00\x00\x07\xd1')]})] To expand the nested Granted-Service-Unit AVP (code 431) in the grouped Multiple Services Credit Control top level AVP (code 456): diameter.get_grouped_avps(456,0,((456,0),(431,0)))
	[('@', {(432, 0): [('@', '\x00\x00\x00\x01')], (431, 0): [('@', {(420, 0): [('@', '\x00\x00\x00d')]})], (448, 0): [('@', '\x00\x00\x04\xb0')], (268, 0): [('@', '\x00\x00\x07\xd1')]}), ('@', {(432, 0): [('@', '\x00\x00\x00\x02')], (431, 0): [('@', {(420, 0): [('@', '\x00\x00\x03\x84')]})], (448, 0): [('@', '\x00\x00\x04\xb0')], (268, 0): [('@', '\x00\x00\x07\xd1')]}), ('@', {(432, 0): [('@', '\x00\x00\x00\x03')], (431, 0): [('@', {(420, 0): [('@', '\x00\x00\x00<')]})], (448, 0): [('@', '\x00\x00\x04\xb0')], (268, 0): [('@', '\x00\x00\x07\xd1')]})]
set_avps(AVP code, vendor id, list of AVP-value-tuples) AVP code, vendor id = top-level-AVP	Remove all instances of the specified top-level-AVP from the message. For each entry in the AVP-value-tuple list, a top-level-AVP instance is inserted. If the specified vendor ID value is zero, then no vendor ID field is inserted and setting the Vendor-Specific bit in the flags field of the AVP value tuple results in a Python error: "ValueError: no vendor ID but vendor flag set". If the specified vendor ID value is non-zero, then a vendor ID field is inserted and not setting the Vendor-Specific bit in the flags field of the AVP value tuple results in a Python error: "ValueError: vendor ID but vendor flag not set". Padding between AVPs, AVP length and Diameter message length are adapted accordingly by the system. Return value is None. For example: diameter.set_avps(461,0,[('\x40', 'Pytho.n-1'), ('\x40', 'Python-2')])
set_fixed_position_avps(AVP-order-tuple)	Put the specified top-level-AVPs at the beginning of the message in the order as specified in the AVP-order-tuple. This method overrides the order of the top-level-AVPs in the resulting Diameter message. If for example the session-id AVP must appear as first in the message, then the corresponding AVP-key-tuple must be included in the first position of the AVP-order-tuple. AVPs not present in the message but specified in the AVP-order-tuple are ignored.

Method	Description
	AVPs present in the message and not specified in the AVP-order-tuple are included in the final message after the AVPs listed in the AVP-order-tuple. The order is deterministic but implementation specific. This method can appear at any point in the script. The last call overrides the previous one. From a black box viewpoint, this method is executed at the end of the script: the result of the call is not reflected in the list returned by a subsequent get_avp_list() call. Return value: None For example: diameter.set_fixed_position_avps(((263,0), (264,0), (296,0), (268,0)
set_grouped_avps(AVP code, vendor id, list of grouped-AVP-value-tuples) AVP code, vendor id = top-level-AVP	Remove all instances of the specified grouped top-level-AVP from the message. For each entry in the grouped-AVP-value-tuples list, a grouped top-level-AVP instance is inserted. The order of the embedded-AVPs in the grouped AVP cannot be specified. If the specified vendor ID value is zero, then no vendor ID field is inserted and setting the Vendor-Specific bit in the flags field of the AVP value tuple will then result in a Python error: "ValueError: no vendor ID but vendor flag set". If the specified vendor ID value is non-zero, then a vendor ID field is inserted and not setting the Vendor-Specific bit in the flags field of the AVP value tuple will result in a Python error: "ValueError: vendor ID but vendor flag not set". Padding between AVPs, AVP length and Diameter message length are adapted accordingly. Return value is None. For example: diameter.set_grouped_avps(456,0,[('@', {(432, 0): [('@', '\x00\x00\x00\x01')], (431, 0): [('@', {(420, 0): [('@', '\x00\x00\x00\x2b')]})], (448, 0): [('@', '\x00\x00\x04\xb0')], (268, 0): [('@', '\x00\x00\x07\xd1')]}), ('@', {(432, 0): [('@', '\x00\x00\x00\x03')], (431, 0): [('@', {(420, 0): [('@', '\x00\x00\x00\x53')]})], (448, 0): [('@', '\x00\x00\x04\xb0')], (268, 0): [('@', '\x00\x00\x07\xd1')]})])

To enable Diameter message manipulation using Python, a Python policy must be configured in the diameter-peer-policy. For example:

The Python policy specifies the Python script to use for each Diameter message type received or transmitted on a Diameter peer. In the ingress direction, the Python script is executed when the corresponding packet type is received on the Diameter peer but before further processing in the system. In the egress direction, the Python script is executed prior to sending the corresponding packet type on the Diameter peer. For example:

The Python script specifies the location of the script and optional protection mechanism. For example:

```
configure
    python
        python-script "diameter-1" create
            primary-url "ftp://usr:pwd@192.0.2.1/./py/diam-1.py"
            no shutdown
        exit
    exit
```

As an example, the diam-1.py script, clears the M-bit from the Event-Timestamp AVP (code 55):

```
    from alc import diameter
    avp55=diameter.get_avps(55,0)
    diameter.set_avps(55,0,[('\x00',avp55[0][1])])
```

18.5.5 Python policy – DHCP transaction cache API

A DHCP transaction cache (DTC) is a short-lived cache during DHCPv4/v6 transaction. The cache could be used to store user-chosen information or return ESM attributes via a Python script. The DTC's lifetime is only during a single DHCP transaction (for example, only between discovery-offer, request-reply). This includes both alc.dtc.store() data and alc.dtc.setESM() data. DTC is also a transaction specific cache, which means the cached information could only be accessed by the Python script running in same DHCP transaction.

The following are the DTC APIs:

- alc.dtc.derivedId: A string used as a LUDB lookup key
- alc.dtc.store(cache-key, cache-value): Store the value with the specified cache-key in DTC, both key and value are string.
- alc.dtc.retrieve(cache-key): Returns the cached value string according to the specified cache-key, raise exception if specified key does not exist.
- alc.dtc.setESM(ESM-key, value): Sets the specified ESM attribute, which could be used when system creating the ESM host.

  Note:
  Because of the short-lived nature of DTC, setESM should be used in final DHCP transaction before system create ESM host., such as DHCPv4 REQUEST-ACK.

- The following is a list of supported ESM-keys and corresponding Python type:
  - alc.dtc.accountingPolicy:str
  - alc.dtc.ancpString:str
  - alc.dtc.appProfileString:str
  - alc.dtc.catMapString:str
  - alc.dtc.defGw:str
  - alc.dtc.dhcpv4GIAddr:str
  - alc.dtc.dhcpv4Pool:str

- alc.dtc.dhcpv4ServerAddr:str
- alc.dtc.dhcpv6LinkAddr:str
- alc.dtc.dhcpv6PreferredLifetime
- alc.dtc.dhcpv6RebindTimer
- alc.dtc.dhcpv6RenewTimer
- alc.dtc.dhcpv6ServerAddr:str
- alc.dtc.dhcpv6ValidLifetime
- alc.dtc.intDestId:str
- alc.dtc.ipAddress:str
- alc.dtc.ipv4LeaseTime:int
- alc.dtc.ipv4PrimDns:str
- alc.dtc.ipv4SecDns:str
- alc.dtc.ipv6Address:str
- alc.dtc.ipv6DelegatedPrefix:str
- alc.dtc.ipv6DelegatedPrefixLength:int
- alc.dtc.ipv6PrefixPool:str
- alc.dtc.ipv6PrimDns:str
- alc.dtc.ipv6SecDns:str
- alc.dtc.ipv6SlaacPrefix:str
- alc.dtc.ipv6WanPool:str
- alc.dtc.msapGroupInterface:str
- alc.dtc.msapPolicy:str
- alc.dtc.msapServiceId:str
- alc.dtc.primNbns:str
- alc.dtc.primNbns:str
- alc.dtc.retailServiceId:str
- alc.dtc.secNbns:str
- alc.dtc.slaProfileString:str
- alc.dtc.subIdent:str
- alc.dtc.subnetMask:str
- alc.dtc.subProfileString:str

18.5.6 Python for PPPoE API

The following are the system-provided Python API PPPoE packets:

- The system provides a Python object for input PPPoE packet: alc.pppoe.

- alc.pppoe class is always available even when Python script is trigger by other type of packet that encapsulated in PPPoE, such as LCP/IPCP/PAP/CHAP, and so on.
- alc.pppoe has the following attributes to represent the PPPoE header fields.

Table 77: System provided Python API PPPoE packets displays information about system-provided Python API PPPoE packets.

Table 77: System provided Python API PPPoE packets

Class attributes	PPPoE header field	Access
alc.pppoe.dest_mac	Str, destination MAC address; the format is "xx:xx:xx:xx:xx:xx"	Read
alc.pppoe.src_mac	Str, source MAC address; the format is "xx:xx:xx:xx:xx:xx"	Read
alc.pppoe.port_id	Str, the port-id of the PPPoE session like "1/1/3"	Read
alc.pppoe.vlan_tag	Str, the vlan tag of the PPPoE session like "100" or "100.200"	Read
alc.pppoe.ver	Integer, ver field in PPPoE header	Read
alc.pppoe.type	Integer, type field in PPPoE header	Read
alc.pppoe.code	Integer, code field in PPPoE header	Read
alc.pppoe.session_id	Integer, session_id field in PPPoE header	Read
alc.pppoe.len	Integer, length field in PPPoE header	Read

The following is a list of functions of alc.pppoe:

- **alc.pppoe.drop()**

  The system drops the resulting packet.
- **alc.pppoe.getTagList()**

  The system returns a tuple that includes tag-type of existing PPPoE tags in the packet. The order of the element in the tuple is as same as the tags appear in the packet. If there are multiple instances of same option, then each instance is one element in the tuple.
- **alc.pppoe.get(tag-type)**

  The system returns a tuple of strings, each string represent one instance of the specified tags, the value of this string is exact bytes of the tag_value as it appears in packet (excludes tag_type and tag_len); if tag does not exist, return (); if a specific instance's tag_length=0, return "" for this instance.
- **alc.pppoe.set(tag-type,valTuple)**

  This function removes all the existing instances of specified tags and insert a list of new tags. Each element in valTuple is a string, represent one instance of the new tag to be inserted. For each new option, the type is specified in tag-type, the option-len is the length of the element, reset of option is the element itself.

- **alc.pppoe.clear(tag-type)**

  This function removes all the existing instances of specified tags. In fact this function is a shortcut version of alc.pppoe.set(tag-type,()).

- **alc.pppoe.getPPP()**

  This function returns an alc.ppp object which represent the LCP/IPCP/IP6CP packet encapsulated inside the PPPoE packet. It returns none if there is no such packet.

- **alc.pppoe.getPAP()**

  This function returns an alc.pap object which represents the PAP packet encapsulated inside the PPPoE packet. It returns none if there is no PAP packet.

- **alc.pppoe.setPAP(pap_obj)**

  This function accepts an alc.pap object as the parameter and replace the existing PAP packet that encapsulated in the PPPoE packet. An exception is raised when there is no existing PAP packet.

- **alc.pppoe.getCHAP()**

  This function returns an alc.chap object, which represent the CHAP packet encapsulated inside the PPPoE packet. It returns none if there is no CHAP packet.

- **alc.pppoe.setCHAP(chap_obj)**

  This function accepts an alc.chap object as the parameter and replace the existing CHAP packet that encapsulated in the PPPoE packet. An exception is raised when there is no existing CHAP packet.

18.5.7 Python API for PPP packet

The system provides a Python object for PPP packet encapsulated in PPPoE packets: alc.ppp.

Use alc.pppoe.getPPP() to get the alc.ppp object in the input PPPoE packet. This class only supports following control level protocol:

- LCP (read-only)
- IPCP (read-only)
- IP6CP(read-only)
- alc.ppp has following attributes to represent the PPP header fields:

Table 78: PPP header field attributes displays information about PPP header field attributes.

Table 78: PPP header field attributes

Class attributes	PPP header field	Access
alc.ppp.protocol	Integer, protocol field in ppp header	Read
alc.ppp.code	Integer, code field in ppp protocol header	Read
alc.ppp.id	Integer, identifier field in ppp protocol header	Read
alc.ppp.len	Integer, length field in ppp protocol header	Read

The following is a list of functions of alc.ppp:

- **alc.ppp.getOptionList()**

  Returns a tuple that includes type of existing PPP options in the packet. The order of the element in the tuple is as same as the options appear in the packet. If there are multiple instances of same option, then each instance is one element in the tuple.

- **alc.ppp.get(op-type)**

  Returns a tuple of strings, each string represent one instance of the specified option, the value of this string is exact bytes of the option_data as it appears in packet (excludes option_type and option_len); if the option does not exist, return (); if a specific instance's option_length=0, return "" for this instance.

18.5.8 Python API for PPP PAP

The system provides a python object for PAP packet encapsulated in the input PPPoE packets: alc.pap.

Use alc.pppoe.getPAP() to get the alc.pap object in the input PPPoE packet.

To apply changes to PAP packet, alc.pppoe.setPAP(new_alc.pap_obj) must be called.

Table 79: PAP header fields provides PAP header field information. alc.pap has following attributes to represent the PAP header fields:

Table 79: PAP header fields

Class attributes	PAP field	Access
alc.pap.code	Integer, code field in pap protocol header	Read
alc.pap.id	Integer, identifier field in pap protocol header	Read
alc.pap.len	Integer, length field in pap protocol header	Read

The following is a list of functions in the class:

- **alc.pap.getCred()**

  Only apply to Authentication-Request packet. Returns a tuple: (peer-id, password), both are str. If either of them has 0 length, it is ""; if the system failed to parse the packet, such as the length < 6 bytes or the packet is not an authentication-request, then the system raises an exception. If the packet has a wrong type it returns None (that is, no exception is raised).

- **alc.pap.setCred(cred_tuple)**

  Only apply to Authentication-Request packet. Set the peer-id and password to cred_tuple; cred_tuple is a tuple:(peer-id, password) both are str. "" is allowed as the value for either of them. If the packet is not authentication-request, then system raises an exception.

- **alc.pap.getMsg()**

  Only apply to Authentication-Ack/Nak packet. Return the message in the packet as a str; return "" if len of message is 0; raise an exception for wrong type of packet (such as an auth-request). If the packet has a wrong type it returns None (that is, no exception is raised).

- **alc.pap.setMsg(msg)**

  Only apply to Authentication-Ack/Nak packet. set the message in the packet to parameter msg; msg is a str; "" is allowed; raise an exception for wrong type of packet (such as an auth-request).

18.5.9 Python API for PPP CHAP

The system provides a python object for CHAP packet encapsulated in the input PPPoE packets: alc.chap.

Use alc.pppoe.getCHAP() to get the alc.chap object in the input PPPoE packet.

To apply changes to PAP packet, alc.pppoe.setCHAP(new_alc.pap_obj) need to be called.

Table 80: CHAP header fields provides CHAP header field information. alc.chap has following attributes to represent the CHAP header fields:

Table 80: CHAP header fields

Class attributes	PAP field	Access
alc.chap.code	Integer, code field in chap protocol header	Read
alc.chap.id	Integer, identifier field in chap protocol header	Read
alc.chap.len	Integer, length field in chap protocol header	Read

The following is a list of functions in the class:

- **alc.chap.getCred()**

  Only apply to challenge/response packet. Returns a tuple: (name, challenge, response), both are str. If either of them has 0 length, it is be ""; if system failed to parse the packet, such as the length < minimal_length or the packet is not challenge/response, then system raise an exception. If the packet has a wrong type it returns None (that is, no exception is raised). The second value in the returned tuple either contains the challenge (in a challenge packet) or the response (in a response packet) value.

- **alc.chap.setCred(chap_cred_tuple)**

  Only apply to challenge/response packets. Set the name and challenge to chap_cred_tuple; chap_cred_tuple is a tuple: (name, challenge, response) both are str. "" is allowed as the value for either of them. If packet is not challenge/response, then system raises an exception. The second value in the supplied tuple either contains the challenge (in a challenge packet) or the response (in a response packet) value.

- **alc.chap.getMsg()**

  Only apply to Success/Failure packet. Return the message in the packet as a str; return "" if len of message is 0; raise an exception for wrong type of packet (such as challenge). If the packet has a wrong type it returns None (that is, no exception is raised).

- **alc.chap.setMsg(msg)**

  Only apply to Success/Failure packet. set the message in the packet to parameter msg; msg is a str; "" is allowed; raise an exception for wrong type of packet (such as challenge).

18.5.10 Python ESM API

The system provides a python object: alc.esm, to provide flexible LUDB lookup and return ESM attributes directly from script.

alc.esm support both DHCPv4/v6 and PPPoE hosts.

alc.esm has following attribute and function:

- alc.esm.derivedId: A string used as LUDB lookup key
- alc.esm.set(ESM-key, value): Set the value of specified ESM attribute, it is used for ESM host creation.

alc.esm.derivedId should be set for all types of ingress packets that system used to access LUDB; for example PPPoE PADI and PADR, DHCP discovery and request.

The following is a list of supported ESM-key and its corresponding Python type:

- alc.esm.accountingPolicy:str
- alc.esm.ancpString:str
- alc.esm.appProfileString:str
- alc.esm.catMapString:str
- alc.esm.defGw:str
- alc.esm.dhcpv4GIAddr:str
- alc.esm.dhcpv4Pool:str
- alc.esm.dhcpv4ServerAddr:str
- alc.esm.dhcpv6LinkAddr:str
- alc.esm.dhcpv6ServerAddr:str
- alc.esm.intDestId:str
- alc.esm.ipAddress:str
- alc.esm.ipv4LeaseTime:int
- alc.esm.ipv4PrimDns:str
- alc.esm.ipv4SecDns:str
- alc.esm.ipv6Address:str
- alc.esm.dhcpv6PreferredLifetime
- alc.esm. dhcpv6RebindTimer
- alc.esm. dhcpv6RenewTimer
- alc.esm. dhcpv6ValidLifetime
- alc.esm.ipv6DelegatedPrefix:str
- alc.esm.ipv6DelegatedPrefixLength:int
- alc.esm.ipv6PrefixPool:str
- alc.esm.ipv6PrimDns:str
- alc.esm.ipv6SecDns:str
- alc.esm.ipv6SlaacPrefix:str
- alc.esm.ipv6WanPool:str
- alc.esm.msapGroupInterface:str
- alc.esm.msapPolicy:str
- alc.esm.msapServiceId:str
- alc.esm.primNbns:str

- alc.esm.retailServiceId:str
- alc.esm.secNbns:str
- alc.esm.slaProfileString:str
- alc.esm.spiSharingGroupId:int
- alc.esm.subIdent:str
- alc.esm.subProfileString:str
- alc.esm.subnetMask:str
- alc.dtc.dhcpv6PreferredLifetime
- alc.dtc.dhcpv6RebindTimer
- alc.dtc.dhcpv6RenewTimer
- alc.dtc.dhcpv6ValidLifetime

The following are PPPoE-specific keys:

- alc.esm.padoDelay: int
- alc.esm.pppAuthProtocol: str, {'pap'|'chap'|'pref-chap'|'pref-pap'}
- alc.esm.pppMTU: int

With PPPoE, alc.esm.set() could only be called ingress discovery (PADI, PADR) and authentication (PAP Authenticate, CHAP Response) packets. If it is called multiple times, the first one wins except the derived-id, while the alc.esm.padoDelay only applies when the system receives PADI. alc.esm.pppAuthProtocol and alc.esm.pppMTU can only be set when a PADR is received.

alc.esm.set() is blocked during LCP/IPCP/IP6CP phase.

The system raises an exception if alc.esm.set() is called at wrong time. Setting the ESM keys that do not apply to the host type are ignored; for example, setting PPP(oE) keys for DHCP hosts is ignored.

The value returned by alc.esm.set() take highest precedence than other sources like CLI/LUDB.

For alc.esm.set(ESM-key, value)

- PPPoE can only be called for ingress discovery (PADI, PADR) and authentication (PAP Authenticate, CHAP Response) packets. An exception is raised otherwise (for example, wrong time, including egress discovery PADO and PADS, and authentication PAP Ack/Nak and CHAP Success/Failure).
- The PPPoE key alc.esm.padoDelay can only be set when PADI is received (otherwise, it is too late).
- Setting ESM keys for a PPPoE host which do not apply is ignored (that is, no exception is raised).
- PPP keys alc.esm.pppAuthProtocol and alc.esm.pppMTU can only be set when PADR is received. It has no effect when PADI is received.
- Setting PPP(oE) keys for a DHCP host is ignored.

18.5.11 Python cache support

Python cache support allows information to be shared across different run times of the same python script or even different python scripts in a programmatic way. It essentially provides a central memory cache and a set of APIs which allows the user store and retrieve strings. For example, a DHCP python-script could store a DHCP option into cache and later a RADIUS python-script could retrieve stored string and add it into access-request.

Each cached entry in the cache is a tuple of (**key**, **val**). **key** is used as entry ID. **val** is the string to be cached. Both **key** and **val** are strings. The max length of the key is 512 bytes. Future more, the combine length of key+val is limited by the configured value of **entry-size** *size* command in the python-policy.

The Python cache can be enabled per python-policy. Each python policy has its own cache memory which script in other python-policy cannot access. This also implies that the key of a cached entry in different a python policy could overlap.

The user can also specify the max number cache entry per python policy the command **max-entries** command. System has a global limit for python cache memory of 256MB.

The cached entries could be made persistent by saving it to CF card. This can be enabled with the **persistence** command in the Python policy.

Note: From memory consumption point of view, with MCS enabled, each cached entry has a corresponding MCS record, so each entry consumes twice amount of memory.

The system also supports syncing the python cache across chassis with MCS. This can be configured per python policy with the **mcs-peer** command in the python policy.

Each cached entry has a remaining lifetime. If it decreases over time, the system removes the cached entry if its remaining lifetime is 0. The remaining lifetime can be changed using a system- provided API. The initial lifetime of a newly created cache entry is 600 seconds. Python cache is supported only with CPM-based Python policies.

The following are the Python cache APIs in a module alc.cache:

- **alc.cache.save(val, key)**

  Saves the val identified by the key into the cache. If there is an existing cache entry with same key, it is overwritten with the new val. In such overwritten cases, the lifetime of the entry is also reset to the default value of 600 seconds. An exception is raised if the save failed (for example, because the maximum number of entries is exceeded).

- **alc.cache.retrieve(key)**

  Returns the stored entry's val identified by the key. A KeyError exception is raised if the specified entry does not exist.

- **alc.cache.clear(key)**

  Removes the cached entry identified by the key. Raise KeyError exception if the specified entry does not exist.

- **cache.get_lifetime(key)**

  The system returns a integer as seconds of remaining lifetime of the specified entry. It returns none if the specified entry does not exist. An exception is raised for any other error.

- **cache.set_lifetime(key,new_lifetime)**

  The new_lifetime value is an integer. The system sets the remaining lifetime of the specified entry to the number of seconds of the new_lifetime. An exception is raised for any error including specified entry does not exist. If the new_lifetime>=max_lifetime(configurable using the **max-entry-life** command in the python policy), then the system sets the actual lifetime to the max_lifetime.

18.5.12 Applying a Python policy

The following is a list of places that a Python policy could be applied:

- **Under capture SAP**

  Apply to the DHCPv4/v6 packets sent/received on the capture SAP
- **Under group-interface**

  Apply to DHCPv4/v6 packets sent/received on the group-interface
- **Under subscriber-interface**

  Apply to DHCPv4 packets on the retail subscriber interface
- **In the radius-server-policy**

  Apply to the RADIUS packets sent, received to, or from the RADIUS servers configured in the radius-server-policy
- **In the radius-proxy-server**

  Apply to the RADIUS packets on the client side of proxy
- **In the diameter-peer-policy**

  Apply to the Diameter packets sent or received on the Diameter peers configured in the policy

18.5.13 Python script protection

To protect the Python script from unintended changes, the SR OS supports a new Python script file format:SRPY. SRPY includes a key based hash (HMAC) of the original script content. When the system loads a script with SRPY format, a hash is computed by using a configured key and script content. The result hash is compared to the embedded hash. If it is the same, then this script is considered valid. Otherwise, the system aborts with a warning message.

Users can configure **protection hmac-sha256 key** *key* within a Python script. To mandate, all configured scripts must be in SRPY format.

The system provides a tools command (**tools perform python-script protect**) to convert a Python script into SRPY format.

There are also running time limitations for Python scripts to prevent DoS attacks:

- Centralized:
  - Initial run: 100 ms
  - Subsequent run: 10ms
- Distributed:
  - Initial run: 10 ms
  - Subsequent run: 1ms

18.6 Tips and tricks

- Use xrange() instead of range().
- Avoid too many string operations. The following scripts provide the same output:

```
# This script takes 2.5 seconds. s = ""
```

```
for c in 'A'*50000:
s += str(ord(c)) + ', ' print '[' + s[:-2] + ']'

# This script takes 0.1 seconds. print map(ord, 'A'*50000)
```

19 Aggregated forwarding statistics in subscriber management

The aggregated number of forwarding bytes and packets per direction in the subscriber management environment is supported on the following levels:

- outer VLAN
- subscriber interface
- group interface

In addition to forwarding bytes and packets, the outer VLAN statistics also include the number of subscribers hosts active on that VLAN.

Aggregated statistics collection is supported for the following scenarios:

- all subscriber types such as IPoE, PPPoE, LAC/LNS, static, ARP, RS, data-triggered
- all SAP types such as LAG, PW-SAP, MSAP, static SAPs
- wholesaler and retailer models where the statistics are collected on retailer interfaces
- subscribers on direct faceplate ports (breakouts or not), satellite ports, LAGs or PW-ports (fixed or FPE-based)

Statistics are aggregated for combined IPv4 and IPv6 traffic and are based on the subscriber queue and policer statistics. SAP queue statistics are excluded from the count.

When policers are deployed on egress, traffic flowing through them is also traversing a queue or a queue group to which the policer is associated. Policers are always associated with a queue or a queue group. To avoid double counting, statistics must be gathered only from a single entity in the chain (a policer or the next queue), but not both. The following is a list of supported egress deployments with policers that produce accurate aggregate statistics collection with no double counting:

- policers attached to the default queue-group
- policers attached to a specific queue-group
- policers attached to the subscriber (local) queues

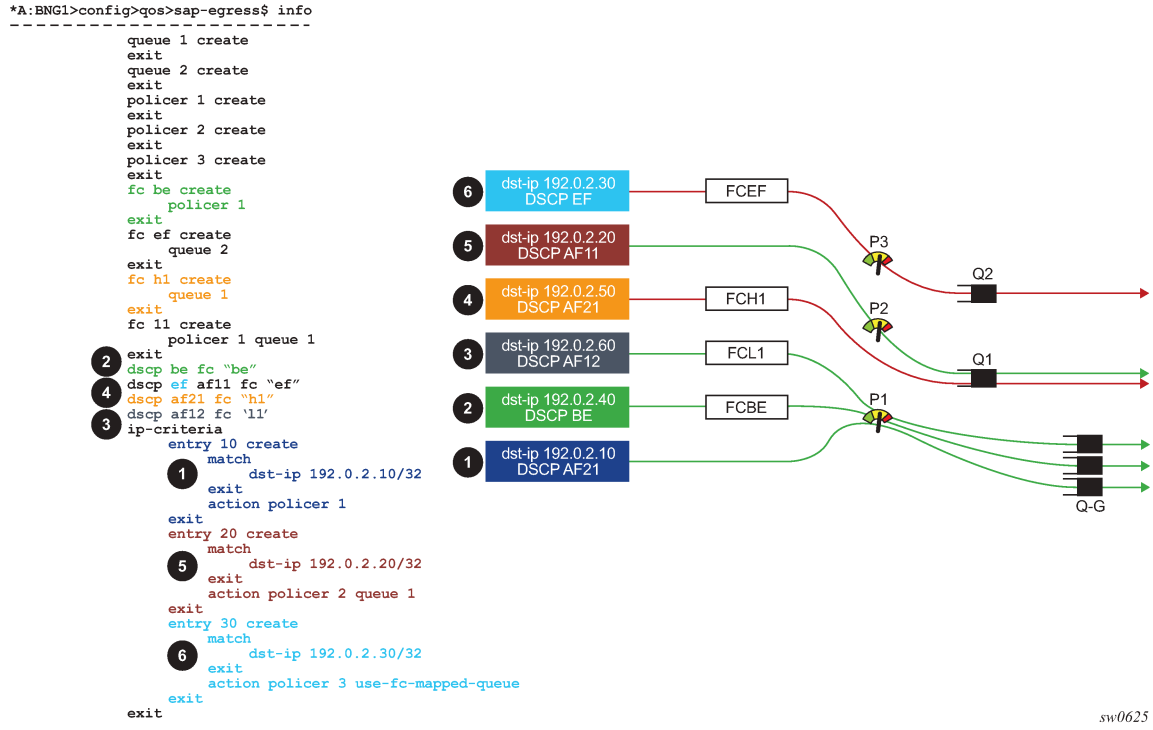
  In this case, only policer statistics are included in aggregated statistics while the post-policer queue statistics are ignored. This means that if traffic is sent directly to this queue without first going through a policer (for example, traffic mapped through an FC directly to a queue), the traffic is omitted from the aggregated statistics.
- HQoS-managed policers attached to the queue group (default or a specific group)

  Traffic traversing HQoS-managed policers is accounted for in HQoS, while traffic traversing non-HQoS-managed policers is not included in HQoS-managed calculations. The queue groups must configure the **no queues-hqos-manageable** command to avoid double counting.

The **no queues-hqos-manageable** command prevents HQoS from using the queue group statistics in its calculation, and therefore, avoids double counting.

Figure 237: Egress statistics in relation to QoS deployment models shows an example of how egress statistics are used.

Figure 237: Egress statistics in relation to QoS deployment models



In Figure 237: Egress statistics in relation to QoS deployment models , the right side of the diagram represents traffic streams and their mapping to policers and queues according to the configuration statement shown on the left side. The four green lines represent traffic streams that are counted properly, and the two red lines represent the two streams that are counted incorrectly (they are either double counted, or not counted at all). The colored boxes numbered 1 through 6 represent a traffic stream with relevant classification fields. For example, the traffic stream in box 1 has the destination IP address set to 192.0.2.10 and DSCP value set to AF21.

1. The **dark blue** traffic stream (box 1) is classified by IP criteria (dst-ip 192.0.2.10 in entry 10), which has a higher evaluation priority than the classification using DSCP (DSCP=AF21). Traffic is matched directly to policer 1 and a default queue group after that, and not to queue 1 as DSCP=AF21 would suggest. Drops on the queue-group are reflected in forwarded statistics for policer 1.
2. The **green** traffic stream (box 2) is not classified by IP criteria (dst-ip 192.0.2.40 is outside any configured entry in the IP criteria). Instead, this traffic is classified using DSCP=BE which is mapped to FC 'BE' and then to policer 1. In terms of statistics, this case is identical to the previous case (1).
3. The **grey** traffic stream (box 3) is not classified by IP criteria (dst-ip 192.0.2.60 is outside any configured entry in the ip-criteria). Instead, this traffic is classified using DSCP=AF12 which is mapped to FC 'l1' and then to policer 1 followed by a local queue 1. In this case, only policer statistics are counted toward aggregated statistics. This is expected for accurate accounting (no double accounting).
4. The **orange** (box 4) traffic stream is not accounted for in the aggregated statistics. Classification is performed based on DSCP=AF21 which is mapped to FC 'h1' and then to queue 1. For the purpose of aggregated statistics collection, counters on queue 1 in this example are not collected because this queue is an explicitly configured post-policer queue (both policer 1 and 2 are fed into this queue by explicit configuration). This excludes queue counters from being counted in the aggregated statistics. While this is the wanted behavior when traffic is passed through a policer first, that is not the case here.

5. The **brown** (box 5) traffic steam is classified by IP criteria (entry 20). Traffic is mapped to policer 2 and then to local queue 1. Similar to case 3, aggregated statistics are properly counted in this example. Drops on the queue 1 are reflected in forwarded statistics for policer 2.
6. The **light blue** (box 6) traffic stream is classified by IP criteria (entry 30). Traffic is mapped to the policer 3 and the local post-policer queue which is derived from FC mapping (DSCP=EF to FC 'ef' to queue 2). This case is similar to dynamic policers. Aggregated statistics are not counted properly, because the mapping between the policer and the post-policer local queue is not explicit using the configuration but instead it is implicitly derived using FC. As a result, double counting occurs.

19.1 Statistics retention

The SR OS node preserves statistics from a subscriber even when the subscriber is disconnected, and the subscriber's policers or queues are released. This prevents statistics fluctuation in relation to the subscriber's presence and ensures that a statistic counter remains stable. After being counted on a VLAN, subscriber interface, or a group interface, the octet or packet remains accounted for during the life time of that VLAN, subscriber interface, or group interface. However, statistics can be manually cleared with the **clear**>**subscr-mgmt**>**interface-statistics** command.

Reporting absolute (or cumulative) counts in aggregated statistics allows smooth measurement of bandwidth per VLAN, subscriber interface, or group interface without dips caused by departing subscribers. The rate measurement can be performed externally by calculating the difference in byte count between two consecutive statistics polls, divided by the collection interval.

Sudden changes in rates can give an indication of a path failure in the network.

19.2 Simultaneous statistical monitoring for multiple entities

Statistics monitoring can be simultaneously enabled for all of the following entities:

- VLANs
- subscriber interfaces
- group interfaces

The sum of monitored entities must not exceed 2,000.

19.3 Enabling aggregate statistics collection

Aggregate statistics collection is enabled by a configuration flag at a global level. The following shows the CLI configuration.

```
configure
    subscriber-mgmt
        svlan-statistics
            [no] shutdown
        subscriber-interface-statistics
            [no] shutdown
        group-interface-statistics
            [no] shutdown
```

After aggregate statistics collection is enabled, the MIB table is populated automatically with the current VLAN, subscriber interface, or group interface entries, up to the supported limit.

19.4 MIBs

Aggregated statistics are provided in the form of a read-only MIB table for the currently active VLANs, subscriber interfaces, and group interfaces.

19.4.1 VLAN MIBs

The keys for the VLAN MIB table are the port ID (which can also be a LAG ID or PW port ID) and the VLAN ID. After the VLAN is instantiated (either statically or through MSAP), an entry is created in the MIB table.

When the VLAN is no longer present in the system, it is automatically removed from the table.

Each VLAN can be queried through SNMP either directly or as an SNMP walk, in which case, all entries in the table are read.

The MIB table name is tmnxSubSvlanStatsTable and has the format shown in Table 81: tmnxSubSvlanStatsEntry objects with up to 2000 tmnxSubSvlanStatsEntry entries.

Table 81: tmnxSubSvlanStatsEntry objects

	Entry objects
Entry Keys Index	tmnxSubSVlanStatsPort The port ID
	tmnxSubSVlanStatsVlan The VLAN ID
Statistics	tmnxSubSVlanStatsLastCleared The most recent time when the stats were cleared
	tmnxSubSVlanStatsIngPkts The number of packets forwarded on ingress
Statistics (cont.)	tmnxSubSVlanStatsIngOctets The number of octets forwarded on ingress
	tmnxSubSVlanStatsEgrPkts The number of packets forwarded on egress
	tmnxSubSVlanStatsEgrOctets The number of octets forwarded on egress
	tmnxSubSVlanStatsActiveSubHosts The number of subscriber hosts

Figure 238: Output captured from SNMP tool

No.	OID-Name	Syntax	Value	Full_OID	Access	Request Time
1	tmnxSubSVlanStatsLastCleared.35717120.51	SNMPv2-TC:TimeStamp	293294	1.3.6.1.4.1.6527.3.1.2.33.1.1001.7.1.3.35717120.51	read-only	11:07:27
2	tmnxSubSVlanStatsIngPkts.35717120.51	Counter64	5	1.3.6.1.4.1.6527.3.1.2.33.1.1001.7.1.4.35717120.51	read-only	11:07:27
3	tmnxSubSVlanStatsIngOctets.35717120.51	Counter64	670	1.3.6.1.4.1.6527.3.1.2.33.1.1001.7.1.5.35717120.51	read-only	11:07:27
4	tmnxSubSVlanStatsEgrPkts.35717120.51	Counter64	5	1.3.6.1.4.1.6527.3.1.2.33.1.1001.7.1.6.35717120.51	read-only	11:07:27
5	tmnxSubSVlanStatsEgrOctets.35717120.51	Counter64	670	1.3.6.1.4.1.6527.3.1.2.33.1.1001.7.1.7.35717120.51	read-only	11:07:27
6	tmnxSubSVlanStatsActiveSubHosts.35717120.51	Gauge32	1	1.3.6.1.4.1.6527.3.1.2.33.1.1001.7.1.8.35717120.51	read-only	11:07:27

19.4.2 Subscriber interface and group interface MIBs

The key for aggregate statistics MIB table are the subscriber interface and group interface name and the router ID.

After the subscriber or group interface is created, it is placed as an entry in the MIB table. After the subscriber or group interface is no longer present in the system (removed or deleted), it is automatically deleted from the table.

Each subscriber or group interface can be directly queried through SNMP or an SNMP walk can be issued to read all the entries in the table.

The MIB table name for the aggregated statistics under the subscriber and group interface is svcIfSubStatsTable and is described in Table 82: svcIfSubStatsTable objects .

Table 82: svcIfSubStatsTable objects

	Entry objects
Entry Keys Index	iesIfIndex The subscriber or group interface name
	svcId The service ID
Statistics	svcIfSubStatsLastCleared The most recent time when the stats were cleared
	svcIfSubStatsIngPkts The number of packets forwarded on ingress
	svcIfSubStatsIngOctets The number of octets forwarded on ingress
	svcIfSubStatsEgrPkts The number of packets forwarded on egress

	Entry objects
	svcIfSubStatsEgrOctets The number of octets forwarded on egress

Figure 239: Output captured with SNMP tool

No.	Object	OID	Syntax	Access	Status	Devfel	Misc
1	svcIfSubStatsLastCleared	1.3.6.1.4.1.6527.3.1.2.4.2.267.1.3	SNMPv2-TC:TimeStamp	read-only	current	N/A	N/A
2	svcIfSubStatsingPkts	1.3.6.1.4.1.6527.3.1.2.4.2.267.1.4	Counter64	read-only	current	N/A	N/A
3	svcIfSubStatsingOctets	1.3.6.1.4.1.6527.3.1.2.4.2.267.1.5	Counter64	read-only	current	N/A	N/A
4	svcIfSubStatsEgrPkts	1.3.6.1.4.1.6527.3.1.2.4.2.267.1.6	Counter64	read-only	current	N/A	N/A
5	svcIfSubStatsEgrOctets	1.3.6.1.4.1.6527.3.1.2.4.2.267.1.7	Counter64	read-only	current	N/A	N/A

sw1166

20 Appendix: Subscriber management in distributed access aggregation devices

This model requires an aggregation network with high-capacity Ethernet-centric aggregation devices that are highly scalable in terms of filter policies, queues, policers, QoS scheduler, and accounting capabilities, which are applied to subscribers as part of subscriber management. These nodes are referred to as Broadband Service Aggregators and are highly distributed, connecting large number of access nodes to the VPLS aggregation network. Sometimes this deployment is referred to as a "bridged model in distributed service edge".

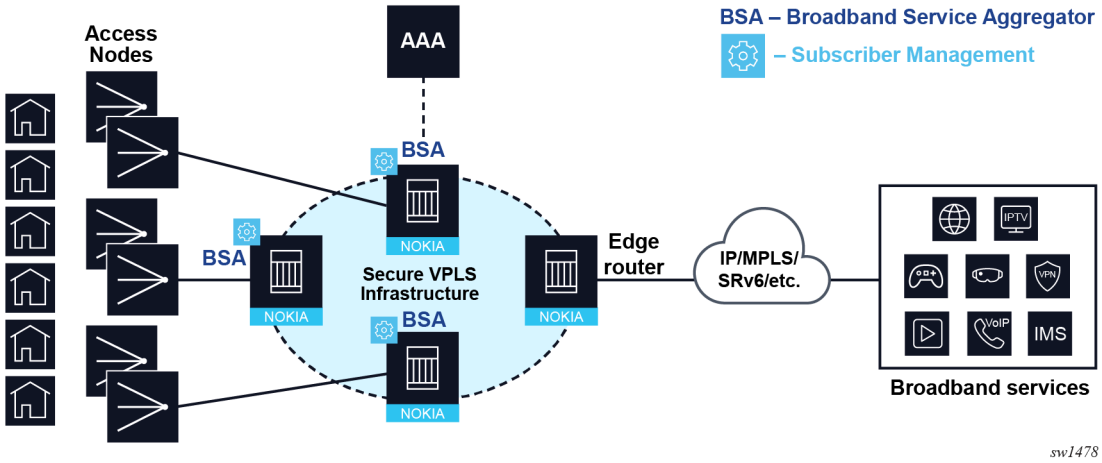
Through Layer 2 aggregation switches, subscriber traffic is aggregated toward the Layer 3 edge router. This edge router terminates Layer 2 access and routes subscriber traffic over various transport technologies (such as IP, MPLS, VXLAN, and SRv6) with the full set of routing protocols, including multicast routing. The edge router supports sophisticated QoS for per service and per-content or source differentiation.

The connectivity between BSAs and the edge router is a secure VPLS infrastructure shown in Figure 240: Bridged Broadband Service Delivery Architecture. This VPLS forms a multipoint Ethernet network with security extensions to prevent unauthorized communication, denial of service, and theft of service. One of the advantages of using VPLS for this application is that VPLS instances can be automatically established over both "hub and spoke" and ring topologies providing sub-50 ms resilience. Regardless of the fiber plant layout, VPLS enables a full mesh to be created between the BSAs and edge router, ensuring efficient traffic distribution and resilience to node or fiber failure.

Other unique features of the BSA and the edge router that contribute to this secure VPLS infrastructure are:

- Using Residential Split Horizon Groups (RSHG), direct user-user bridging is automatically prohibited, without the need for address-specific ACLs.
- RSHG combined with the ARP reply agent perform ARP and broadcast suppression to ensure that addressing information is restricted.
- Protection against theft of service and denial of service is provided by MAC or IP filters automatically populated using DHCP snooping, and by MAC pinning.
- Using the RADIUS interface, it is possible to perform RADIUS authentication of users before allowing a DHCP discover to progress into the network.

Figure 240: Bridged Broadband Service Delivery Architecture



This model, subscriber management in distributed access aggregation devices, is considered a legacy model that relies on DHCPv4 snooping functions in BSA for subscriber identification and, therefore, is supported only for IPoEv4 devices. Because IPoEv6 and PPPoEv4/v6 devices are not supported in this model its use has been limited and replaced by the RCO model, where subscribers are terminated on the BNG in a central office location.

20.1 Configuration example

This configuration example focuses on subscriber management in a distributed aggregation network implemented in a VPLS on BSA. In such environment, the key configuration concepts in subscriber management, such as SLA profile, subscriber profile, subscriber ID, and subscriber identification policy, are optional. If these configuration concepts are omitted, a subscriber with all of its devices is equated to a SAP where DHCP lease state of the hosts is snooped, and consequently managed in the Layer 2 environment. This configuration example deliberately excludes the setup of these optional concepts, which is explored in more detail in Subscriber management concepts , particularly within the context of the commonly employed RCO model.

BSA Configuration

Note: Only relevant configuration in the context of subscriber management is provided. Details about generic concepts, such as QoS and filter policies, or how to configure SDPs, is not part of this configuration example.

VPLS configuration:

- DHCP snooping is enabled on the subscriber-facing SAP and the spoke SDP facing the edge router.
- This example includes only one SAP, which represents a subscriber. For multiple subscribers in the network, a SAP per subscriber must be configured.
- All SAPs are part of a split-horizon group, preventing direct communication between the subscribers. For inter-subscriber communication, traffic must flow through the BSR for security purposes.
- Antispoof is set to subscriber MAC and IP address.

- Subscribers are authenticated based on DHCPv4 Option82 Circuit-Id.
- The BSA replies to the ARP request from subscribers, utilizing a distributed environment, instead of propagating ARP communication to a central point – edge router.
- MAC addresses and IP addresses of the subscribers are learned through DHCP snooping (mac-pinning is used to populate MAC address into VPLS learned through DHCP).
- Filters and QoS setting are configured on ingress and egress for this subscriber.

The following example shows a VPLS configuration.

Example: VPLS configuration (MD-CLI)

```
[ex:/configure service vpls "demo-bridged"]
A:admin@node-2# info
    service-id 10
    customer "1"
    split-horizon-group "RSHG" {
        residential true
    }
    spoke-sdp 100:50 {
        dhcp {
            snoop true
        }
    }
    sap 1/2/6:10 {
        anti-spoof source-ip-and-mac-addr
        split-horizon-group "RSHG"
        arp-reply-agent true
        radius-auth-policy "radius-1"
        ingress {
            qos {
                sap-ingress {
                    policy-name "demo-ingress"
                }
            }
            filter {
                ip "sub1-ingress"
            }
        }
        egress {
            qos {
                sap-egress {
                    policy-name "demo-egress"
                }
            }
            filter {
                ip "sub1-egress"
            }
        }
        stp {
            admin-state disable
        }
        fdb {
            mac-pinning true
        }
        dhcp {
            snoop true
            option-82 {
                action keep
            }
        }
```

```
    }
```

Example: VPLS configuration (classic CLI)

```
A:node-2>config>service# info
----------------------------------------------
         vpls 10 name "demo-bridged" customer 1 create
            split-horizon-group "RSHG" residential-group create
            exit
            stp
                shutdown
            exit
            sap 1/x1/1/c1/4:10.10 split-horizon-group "RSHG" create
                arp-reply-agent
                dhcp
                    shutdown
                    snoop
                exit
                authentication-policy "radius-1"
                anti-spoof ip-mac
                ingress
                    qos 10
                    filter ip 10
                exit
                egress
                    qos 20
                    filter ip 20
                exit
                no shutdown
            exit
            spoke-sdp 100:50 create
                dhcp
                    snoop
                exit
                no shutdown
            exit
   no shutdown
        exit
----------------------------------------------
```

The following example shows a persistency configuration for DHCP lease states on BSR.

Example: persistency configuration for DHCP lease states on BSR (MD-CLI)

```
[ex:/configure system persistence]
A:admin@node-2# info
    subscriber-mgmt {
        location cf3
    }
```

Example: persistency configuration for DHCP lease states on BSR (classic CLI)

```
A:node-2>config>system>persistence# info
----------------------------------------------
            subscriber-mgmt
                location cf3:
            exit
----------------------------------------------
```

The following example shows RADIUS configuration for authentication.

Example: RADIUS server configuration for authentication (MD-CLI)

```
[ex:/configure router "Base" radius]
A:admin@node-2# info
    server "free-radius-1" {
        address 192.168.114.2
        secret "98a4Z4jC9e5KtqVzEN24jTYa+9lOnwQask5IqOdfshV0BB7of64BFUcEdsr88XvwRNV2
 hash2"
    }
```

Example: RADIUS server configuration for authentication (classic CLI)

```
A:node-2>config>router>radius-server# info
----------------------------------------------
            server "free-radius-1" address 192.168.114.2 secret "98a4Z4jC9e5KtqVzEN24jTYa
+9lOnwQask5IqOdfshV0BB7of64BFUcEdsr88XvwRNV2" hash2 create
            exit
----------------------------------------------
```

Example: RADIUS configuration for authentication (MD-CLI)

```
[ex:/configure subscriber-mgmt radius-authentication-policy "radius-1"]
A:admin@node-2# info
    password "ncd8qyrNUMhYfa2SfrUqHMDZ9IXn3sVSmYBzbw== hash2"
    pppoe-access-method pap-chap
    radius-server-policy "radius-server-1"
    user-name {
        format circuit-id
    }
    include-radius-attribute {
        circuit-id true
        nas-identifier true
    }
```

Example: RADIUS configuration for authentication (classic CLI)

```
A:node-2>config>subscr-mgmt# info
----------------------------------------------
        authentication-policy "radius-1" create
            password "ncd8qyrNUMhYfa2SfrUqHMDZ9IXn3sVSmYBzbw==" hash2
            user-name-format circuit-id
            pppoe-access-method pap-chap
            include-radius-attribute
                circuit-id
                nas-identifier
            exit
            radius-server-policy "radius-server-1"
        exit
----------------------------------------------
```

Example: RADIUS configuration for AAA (MD-CLI)

```
[ex:/configure aaa]
A:admin@node-2# info
    radius {
        server-policy "radius-server-1" {
            servers {
                router-instance "Base"
                server 1 {
```

```
                    server-name "free-radius-1"
                }
            }
        }
    }
```

Example: RADIUS configuration for AAA (classic CLI)

```
A:node-2>config>aaa# info
----------------------------------------------
         radius-server-policy "radius-server-1" create
            servers
                router "Base"
                server 1 name "free-radius-1"
            exit
        exit
----------------------------------------------
```

Edge router configuration

The edge router acts as the DHCPv4 relay, which is sending DHCP messages to the server. DHCPv4 snooping is not enabled on the edge router. In addition to the remote DHCPv4 server, only the BSA maintains lease state in this environment.

Example: Edge Router Configuration (MD-CLI)

```
[ex:/configure service ies "BSR" interface "towards-vpls"]
A:admin@node-2# info
    spoke-sdp 101:34 {
    }
    ipv4 {
        primary {
            address 192.0.2.254
            prefix-length 24
        }
        dhcp {
            server [172.16.0.254]
            trusted true
```

Example: Edge Router Configuration (classic CLI)

```
A:node-2>config>service>ies>if# info
----------------------------------------------
                address 192.0.2.254/24
                dhcp
                    shutdown
                    server 172.16.0.254
                    trusted
                exit
                spoke-sdp 101:34 create
                    no shutdown
                exit
----------------------------------------------
```

21 Standards and protocol support

Note:
The information provided in this chapter is subject to change without notice and may not apply to all platforms.

Nokia assumes no responsibility for inaccuracies.

21.1 Access Node Control Protocol (ANCP)

draft-ietf-ancp-protocol-02, *Protocol for Access Node Control Mechanism in Broadband Networks*

RFC 5851, *Framework and Requirements for an Access Node Control Mechanism in Broadband Multi-Service Networks*

21.2 Bidirectional Forwarding Detection (BFD)

draft-ietf-lsr-ospf-bfd-strict-mode-10, *OSPF BFD Strict-Mode*

RFC 5880, *Bidirectional Forwarding Detection (BFD)*

RFC 5881, *Bidirectional Forwarding Detection (BFD) IPv4 and IPv6 (Single Hop)*

RFC 5882, *Generic Application of Bidirectional Forwarding Detection (BFD)*

RFC 5883, *Bidirectional Forwarding Detection (BFD) for Multihop Paths*

RFC 7130, *Bidirectional Forwarding Detection (BFD) on Link Aggregation Group (LAG) Interfaces*

RFC 7880, *Seamless Bidirectional Forwarding Detection (S-BFD)*

RFC 7881, *Seamless Bidirectional Forwarding Detection (S-BFD) for IPv4, IPv6, and MPLS*

RFC 7883, *Advertising Seamless Bidirectional Forwarding Detection (S-BFD) Discriminators in IS-IS*

RFC 7884, *OSPF Extensions to Advertise Seamless Bidirectional Forwarding Detection (S-BFD) Target Discriminators*

RFC 9247, *BGP - Link State (BGP-LS) Extensions for Seamless Bidirectional Forwarding Detection (S-BFD)*

21.3 Border Gateway Protocol (BGP)

draft-gredler-idr-bgplu-epe-14, *Egress Peer Engineering using BGP-LU*

draft-hares-idr-update-attrib-low-bits-fix-01, *Update Attribute Flag Low Bits Clarification*

draft-ietf-idr-add-paths-guidelines-08, *Best Practices for Advertisement of Multiple Paths in IBGP*

draft-ietf-idr-best-external-03, *Advertisement of the best external route in BGP*

draft-ietf-idr-bgp-flowspec-oid-03, *Revised Validation Procedure for BGP Flow Specifications*

draft-ietf-idr-bgp-gr-notification-01, *Notification Message support for BGP Graceful Restart*

draft-ietf-idr-bgp-ls-app-specific-attr-16, *Application-Specific Attributes Advertisement with BGP Link-State*

draft-ietf-idr-bgp-ls-flex-algo-06, *Flexible Algorithm Definition Advertisement with BGP Link-State*

draft-ietf-idr-bgp-optimal-route-reflection-10, *BGP Optimal Route Reflection (BGP-ORR)*

draft-ietf-idr-error-handling-03, *Revised Error Handling for BGP UPDATE Messages*

draft-ietf-idr-flowspec-interfaceset-03, *Applying BGP flowspec rules on a specific interface set*

draft-ietf-idr-flowspec-path-redirect-05, *Flowspec Indirection-id Redirect* – localised ID

draft-ietf-idr-flowspec-redirect-ip-02, *BGP Flow-Spec Redirect to IP Action*

draft-ietf-idr-link-bandwidth-03, *BGP Link Bandwidth Extended Community*

draft-ietf-idr-long-lived-gr-00, *Support for Long-lived BGP Graceful Restart*

RFC 1772, *Application of the Border Gateway Protocol in the Internet*

RFC 1997, *BGP Communities Attribute*

RFC 2385, *Protection of BGP Sessions via the TCP MD5 Signature Option*

RFC 2439, *BGP Route Flap Damping*

RFC 2545, *Use of BGP-4 Multiprotocol Extensions for IPv6 Inter-Domain Routing*

RFC 2858, *Multiprotocol Extensions for BGP-4*

RFC 2918, *Route Refresh Capability for BGP-4*

RFC 4271, *A Border Gateway Protocol 4 (BGP-4)*

RFC 4360, *BGP Extended Communities Attribute*

RFC 4364, *BGP/MPLS IP Virtual Private Networks (VPNs)*

RFC 4456, *BGP Route Reflection: An Alternative to Full Mesh Internal BGP (IBGP)*

RFC 4486, *Subcodes for BGP Cease Notification Message*

RFC 4659, *BGP-MPLS IP Virtual Private Network (VPN) Extension for IPv6 VPN*

RFC 4684, *Constrained Route Distribution for Border Gateway Protocol/MultiProtocol Label Switching (BGP/MPLS) Internet Protocol (IP) Virtual Private Networks (VPNs)*

RFC 4724, *Graceful Restart Mechanism for BGP* – helper mode

RFC 4760, *Multiprotocol Extensions for BGP-4*

RFC 4798, *Connecting IPv6 Islands over IPv4 MPLS Using IPv6 Provider Edge Routers (6PE)*

RFC 5004, *Avoid BGP Best Path Transitions from One External to Another*

RFC 5065, *Autonomous System Confederations for BGP*

RFC 5291, *Outbound Route Filtering Capability for BGP-4*

RFC 5396, *Textual Representation of Autonomous System (AS) Numbers* – asplain

RFC 5492, *Capabilities Advertisement with BGP-4*

RFC 5668, *4-Octet AS Specific BGP Extended Community*

RFC 6286, *Autonomous-System-Wide Unique BGP Identifier for BGP-4*

RFC 6368, *Internal BGP as the Provider/Customer Edge Protocol for BGP/MPLS IP Virtual Private Networks (VPNs)*

RFC 6793, *BGP Support for Four-Octet Autonomous System (AS) Number Space*

RFC 6810, *The Resource Public Key Infrastructure (RPKI) to Router Protocol*

RFC 6811, *Prefix Origin Validation*

RFC 6996, *Autonomous System (AS) Reservation for Private Use*

RFC 7311, *The Accumulated IGP Metric Attribute for BGP*

RFC 7606, *Revised Error Handling for BGP UPDATE Messages*

RFC 7607, *Codification of AS 0 Processing*

RFC 7674, *Clarification of the Flowspec Redirect Extended Community*

RFC 7752, *North-Bound Distribution of Link-State and Traffic Engineering (TE) Information Using BGP*

RFC 7854, *BGP Monitoring Protocol (BMP)*

RFC 7911, *Advertisement of Multiple Paths in BGP*

RFC 7999, *BLACKHOLE Community*

RFC 8092, *BGP Large Communities Attribute*

RFC 8097, *BGP Prefix Origin Validation State Extended Community*

RFC 8212, *Default External BGP (EBGP) Route Propagation Behavior without Policies*

RFC 8277, *Using BGP to Bind MPLS Labels to Address Prefixes*

RFC 8571, *BGP - Link State (BGP-LS) Advertisement of IGP Traffic Engineering Performance Metric Extensions*

RFC 8950, *Advertising IPv4 Network Layer Reachability Information (NLRI) with an IPv6 Next Hop*

RFC 8955, *Dissemination of Flow Specification Rules*

RFC 8956, *Dissemination of Flow Specification Rules for IPv6*

RFC 9086, *Border Gateway Protocol - Link State (BGP-LS) Extensions for Segment Routing BGP Egress Peer Engineering*

21.4 Bridging and management

IEEE 802.1AB, *Station and Media Access Control Connectivity Discovery*

IEEE 802.1ad, *Provider Bridges*

IEEE 802.1ag, *Connectivity Fault Management*

IEEE 802.1ah, *Provider Backbone Bridges*

IEEE 802.1ak, *Multiple Registration Protocol*

IEEE 802.1aq, *Shortest Path Bridging*

IEEE 802.1AX, *Link Aggregation*

IEEE 802.1D, *MAC Bridges*

IEEE 802.1p, *Traffic Class Expediting*

IEEE 802.1Q, *Virtual LANs*

IEEE 802.1s, *Multiple Spanning Trees*

IEEE 802.1w, *Rapid Reconfiguration of Spanning Tree*

IEEE 802.1X, *Port Based Network Access Control*

21.5 Broadband Network Gateway (BNG) Control and User Plane Separation (CUPS)

3GPP TS 23.003, *Numbering, addressing and identification*

3GPP TS 23.007, *Restoration procedures*

3GPP TS 23.402, *Architecture enhancements for non-3GPP accesses* – S2a roaming based on GPRS

3GPP TS 23.501, *System architecture for the 5G System (5GS)*

3GPP TS 23.502, *Procedures for the 5G System (5GS)*

3GPP TS 23.503, *Policy and charging control framework for the 5G System (5GS)*

3GPP TS 24.501, *Non-Access-Stratum (NAS) protocol for 5G System (5GS)*

3GPP TS 29.244, *Interface between the Control Plane and the User Plane nodes*

3GPP TS 29.281, *General Packet Radio System (GPRS) Tunnelling Protocol User Plane (GTPv1-U)*

3GPP TS 29.500, *Technical Realization of Service Based Architecture*

3GPP TS 29.501, *Principles and Guidelines for Services Definition*

3GPP TS 29.502, *Session Management Services*

3GPP TS 29.503, *Unified Data Management Services*

3GPP TS 29.512, *Session Management Policy Control Service*

3GPP TS 29.518, *Access and Mobility Management Services*

3GPP TS 32.255, *5G data connectivity domain charging*

3GPP TS 32.290, *Services, operations and procedures of charging using Service Based Interface (SBI)*

3GPP TS 32.291, *5G system, charging service*

BBF TR-459, *Control and User Plane Separation for a Disaggregated BNG*

BBF TR-459.2, *Multi-Service Disaggregated BNG with CUPS: Integrated Carrier Grade NAT function*

RFC 8300, *Network Service Header (NSH)*

RFC 8910, *Captive-Portal Identification in DHCP and Router Advertisements (RAs)*

21.6 Certificate management

RFC 4210, *Internet X.509 Public Key Infrastructure Certificate Management Protocol (CMP)*

RFC 4211, *Internet X.509 Public Key Infrastructure Certificate Request Message Format (CRMF)*

RFC 5280, *Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile*

RFC 6712, *Internet X.509 Public Key Infrastructure -- HTTP Transfer for the Certificate Management Protocol (CMP)*

RFC 7030, *Enrollment over Secure Transport*

RFC 7468, *Textual Encodings of PKIX, PKCS, and CMS Structures*

21.7 Circuit emulation

RFC 4553, *Structure-Agnostic Time Division Multiplexing (TDM) over Packet (SAToP)*

RFC 5086, *Structure-Aware Time Division Multiplexed (TDM) Circuit Emulation Service over Packet Switched Network (CESoPSN)*

RFC 5287, *Control Protocol Extensions for the Setup of Time-Division Multiplexing (TDM) Pseudowires in MPLS Networks*

21.8 Ethernet

IEEE 802.3ah, *Media Access Control Parameters, Physical Layers, and Management Parameters for Subscriber Access Networks*

IEEE 802.3x, *Ethernet Flow Control*

ITU-T G.8031/Y.1342, *Ethernet Linear Protection Switching*

ITU-T G.8032/Y.1344, *Ethernet Ring Protection Switching*

ITU-T Y.1731, *OAM functions and mechanisms for Ethernet based networks*

21.9 Ethernet VPN (EVPN)

draft-ietf-bess-bgp-srv6-args-00, *SRv6 Argument Signaling for BGP Services*

draft-ietf-bess-evpn-ip-aliasing-00, *EVPN Support for L3 Fast Convergence and Aliasing/Backup Path* – IP Prefix routes

draft-ietf-bess-evpn-ipvpn-interworking-06, *EVPN Interworking with IPVPN*

draft-ietf-bess-evpn-irb-mcast-09, *EVPN Optimized Inter-Subnet Multicast (OISM) Forwarding* – ingress replication and mLDP

draft-ietf-bess-evpn-pref-df-06, *Preference-based EVPN DF Election*

draft-ietf-bess-evpn-unequal-lb-16, *Weighted Multi-Path Procedures for EVPN Multi-Homing* – section 9

draft-ietf-bess-evpn-virtual-eth-segment-06, *EVPN Virtual Ethernet Segment*

draft-ietf-bess-pbb-evpn-isid-cmacflush-00, *PBB-EVPN ISID-based CMAC-Flush*

draft-sr-bess-evpn-vpws-gateway-03, *Ethernet VPN Virtual Private Wire Services Gateway Solution*

RFC 7432, *BGP MPLS-Based Ethernet VPN*

RFC 7623, *Provider Backbone Bridging Combined with Ethernet VPN (PBB-EVPN)*

RFC 8214, *Virtual Private Wire Service Support in Ethernet VPN*

RFC 8317, *Ethernet-Tree (E-Tree) Support in Ethernet VPN (EVPN) an Provider Backbone Bridging EVPN (PBB-EVPN)*

RFC 8365, *A Network Virtualization Overlay Solution Using Ethernet VPN (EVPN)*

RFC 8560, *Seamless Integration of Ethernet VPN (EVPN) with Virtual Private LAN Service (VPLS) and Their Provider Backbone Bridge (PBB) Equivalents*

RFC 8584, *DF Election and AC-influenced DF Election*

RFC 9047, *Propagation of ARP/ND Flags in an Ethernet Virtual Private Network (EVPN)*

RFC 9135, *Integrated Routing and Bridging in Ethernet VPN (EVPN)* – Asymmetric IRB Procedures and Mobility Procedure

RFC 9136, *IP Prefix Advertisement in Ethernet VPN (EVPN)*

RFC 9161, *Operational Aspects of Proxy ARP/ND in Ethernet Virtual Private Networks*

RFC 9251, *Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Proxies for Ethernet VPN (EVPN)*

21.10 gRPC Remote Procedure Calls (gRPC)

cert.proto version 0.1.0, *gRPC Network Operations Interface (gNOI) Certificate Management Service*

file.proto version 0.1.0, *gRPC Network Operations Interface (gNOI) File Service*

gnmi.proto version 0.8.0, *gRPC Network Management Interface (gNMI) Service Specification*

PROTOCOL-HTTP2, *gRPC over HTTP2*

system.proto Version 1.0.0, *gRPC Network Operations Interface (gNOI) System Service*

21.11 Intermediate System to Intermediate System (IS-IS)

draft-ietf-isis-mi-02, *IS-IS Multi-Instance*

draft-ietf-lsr-igp-ureach-prefix-announce-01, *IGP Unreachable Prefix Announcement* – without U-Flag and UP-Flag

draft-kaplan-isis-ext-eth-02, *Extended Ethernet Frame Size Support*

ISO/IEC 10589:2002 Second Edition, *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 (ISO 8473)*

RFC 1195, *Use of OSI IS-IS for Routing in TCP/IP and Dual Environments*

RFC 2973, *IS-IS Mesh Groups*

RFC 3359, *Reserved Type, Length and Value (TLV) Codepoints in Intermediate System to Intermediate System*

RFC 3719, *Recommendations for Interoperable Networks using Intermediate System to Intermediate System (IS-IS)*

RFC 3787, *Recommendations for Interoperable IP Networks using Intermediate System to Intermediate System (IS-IS)*

RFC 5120, *M-ISIS: Multi Topology (MT) Routing in IS-IS*

RFC 5130, *A Policy Control Mechanism in IS-IS Using Administrative Tags*

RFC 5301, *Dynamic Hostname Exchange Mechanism for IS-IS*

RFC 5302, *Domain-wide Prefix Distribution with Two-Level IS-IS*

RFC 5303, *Three-Way Handshake for IS-IS Point-to-Point Adjacencies*

RFC 5304, *IS-IS Cryptographic Authentication*

RFC 5305, *IS-IS Extensions for Traffic Engineering TE*

RFC 5306, *Restart Signaling for IS-IS* – helper mode

RFC 5308, *Routing IPv6 with IS-IS*

RFC 5309, *Point-to-Point Operation over LAN in Link State Routing Protocols*

RFC 5310, *IS-IS Generic Cryptographic Authentication*

RFC 6119, *IPv6 Traffic Engineering in IS-IS*

RFC 6213, *IS-IS BFD-Enabled TLV*

RFC 6232, *Purge Originator Identification TLV for IS-IS*

RFC 6233, *IS-IS Registry Extension for Purges*

RFC 6329, *IS-IS Extensions Supporting IEEE 802.1aq Shortest Path Bridging*

RFC 7775, *IS-IS Route Preference for Extended IP and IPv6 Reachability*

RFC 7794, *IS-IS Prefix Attributes for Extended IPv4 and IPv6 Reachability* – sections 2.1 and 2.3

RFC 7981, *IS-IS Extensions for Advertising Router Information*

RFC 7987, *IS-IS Minimum Remaining Lifetime*

RFC 8202, *IS-IS Multi-Instance* – single topology

RFC 8570, *IS-IS Traffic Engineering (TE) Metric Extensions* – Min/Max Unidirectional Link Delay metric for flex-algo, RSVP, SR-TE

RFC 8919, *IS-IS Application-Specific Link Attributes*

21.12 Internet Protocol (IP) Fast Reroute (FRR)

draft-ietf-rtgwg-lfa-manageability-08, *Operational management of Loop Free Alternates*

RFC 5286, *Basic Specification for IP Fast Reroute: Loop-Free Alternates*

RFC 7431, *Multicast-Only Fast Reroute*

RFC 7490, *Remote Loop-Free Alternate (LFA) Fast Reroute (FRR)*

RFC 8518, *Selection of Loop-Free Alternates for Multi-Homed Prefixes*

21.13 Internet Protocol (IP) general

draft-grant-tacacs-02, *The TACACS+ Protocol*

RFC 768, *User Datagram Protocol*

RFC 793, *Transmission Control Protocol*

RFC 854, *Telnet Protocol Specifications*

RFC 1350, *The TFTP Protocol (revision 2)*

RFC 2347, *TFTP Option Extension*

RFC 2348, *TFTP Blocksize Option*

RFC 2349, *TFTP Timeout Interval and Transfer Size Options*

RFC 2428, *FTP Extensions for IPv6 and NATs*

RFC 2617, *HTTP Authentication: Basic and Digest Access Authentication*

RFC 2784, *Generic Routing Encapsulation (GRE)*

RFC 2818, *HTTP Over TLS*

RFC 2890, *Key and Sequence Number Extensions to GRE*

RFC 3164, *The BSD syslog Protocol*

RFC 4250, *The Secure Shell (SSH) Protocol Assigned Numbers*

RFC 4251, *The Secure Shell (SSH) Protocol Architecture*

RFC 4252, *The Secure Shell (SSH) Authentication Protocol* – publickey, password

RFC 4253, *The Secure Shell (SSH) Transport Layer Protocol*

RFC 4254, *The Secure Shell (SSH) Connection Protocol*

RFC 4511, *Lightweight Directory Access Protocol (LDAP): The Protocol*

RFC 4513, *Lightweight Directory Access Protocol (LDAP): Authentication Methods and Security Mechanisms* – TLS

RFC 4632, *Classless Inter-domain Routing (CIDR): The Internet Address Assignment and Aggregation Plan*

RFC 5082, *The Generalized TTL Security Mechanism (GTSM)*

RFC 5246, *The Transport Layer Security (TLS) Protocol Version 1.2* – TLS client, RSA public key

RFC 5425, *Transport Layer Security (TLS) Transport Mapping for Syslog* – RFC 3164 with TLS

RFC 5656, *Elliptic Curve Algorithm Integration in the Secure Shell Transport Layer* – ECDSA

RFC 5925, *The TCP Authentication Option*

RFC 5926, *Cryptographic Algorithms for the TCP Authentication Option (TCP-AO)*

RFC 6398, *IP Router Alert Considerations and Usage* – MLD

RFC 6528, *Defending against Sequence Number Attacks*

RFC 7011, *Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of Flow Information*

RFC 7012, *Information Model for IP Flow Information Export*

RFC 7230, *Hypertext Transfer Protocol (HTTP/1.1): Message Syntax and Routing*

RFC 7231, *Hypertext Transfer Protocol (HTTP/1.1): Semantics and Content*

RFC 7232, *Hypertext Transfer Protocol (HTTP/1.1): Conditional Requests*

RFC 7301, *Transport Layer Security (TLS) Application Layer Protocol Negotiation Extension*

RFC 7616, *HTTP Digest Access Authentication*

RFC 8446, *The Transport Layer Security (TLS) Protocol Version 1.3*

21.14 Internet Protocol (IP) multicast

cisco-ipmulticast/pim-autorp-spec01, *Auto-RP: Automatic discovery of Group-to-RP mappings for IP multicast* – version 1

draft-ietf-bier-pim-signaling-08, *PIM Signaling Through BIER Core*

draft-ietf-idmr-traceroute-ipm-07, *A "traceroute" facility for IP Multicast*

draft-ietf-l2vpn-vpls-pim-snooping-07, *Protocol Independent Multicast (PIM) over Virtual Private LAN Service (VPLS)*

RFC 1112, *Host Extensions for IP Multicasting*

RFC 2236, *Internet Group Management Protocol, Version 2*

RFC 2365, *Administratively Scoped IP Multicast*

RFC 2375, *IPv6 Multicast Address Assignments*

RFC 2710, *Multicast Listener Discovery (MLD) for IPv6*

RFC 3306, *Unicast-Prefix-based IPv6 Multicast Addresses*

RFC 3376, *Internet Group Management Protocol, Version 3*

RFC 3446, *Anycast Rendevous Point (RP) mechanism using Protocol Independent Multicast (PIM) and Multicast Source Discovery Protocol (MSDP)*

RFC 3590, *Source Address Selection for the Multicast Listener Discovery (MLD) Protocol*

RFC 3618, *Multicast Source Discovery Protocol (MSDP)*

RFC 3810, *Multicast Listener Discovery Version 2 (MLDv2) for IPv6*

RFC 3956, *Embedding the Rendezvous Point (RP) Address in an IPv6 Multicast Address*

RFC 3973, *Protocol Independent Multicast - Dense Mode (PIM-DM): Protocol Specification (Revised)* – auto-RP groups

RFC 4541, *Considerations for Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Snooping Switches*

RFC 4604, *Using Internet Group Management Protocol Version 3 (IGMPv3) and Multicast Listener Discovery Protocol Version 2 (MLDv2) for Source-Specific Multicast*

RFC 4607, *Source-Specific Multicast for IP*

RFC 4608, *Source-Specific Protocol Independent Multicast in 232/8*

RFC 4610, *Anycast-RP Using Protocol Independent Multicast (PIM)*

RFC 4611, *Multicast Source Discovery Protocol (MSDP) Deployment Scenarios*

RFC 5059, *Bootstrap Router (BSR) Mechanism for Protocol Independent Multicast (PIM)*

RFC 5186, *Internet Group Management Protocol Version 3 (IGMPv3) / Multicast Listener Discovery Version 2 (MLDv2) and Multicast Routing Protocol Interaction*

RFC 5384, *The Protocol Independent Multicast (PIM) Join Attribute Format*

RFC 5496, *The Reverse Path Forwarding (RPF) Vector TLV*

RFC 6037, *Cisco Systems' Solution for Multicast in MPLS/BGP IP VPNs*

RFC 6512, *Using Multipoint LDP When the Backbone Has No Route to the Root*

RFC 6513, *Multicast in MPLS/BGP IP VPNs*

RFC 6514, *BGP Encodings and Procedures for Multicast in MPLS/IP VPNs*

RFC 6515, *IPv4 and IPv6 Infrastructure Addresses in BGP Updates for Multicast VPNs*

RFC 6516, *IPv6 Multicast VPN (MVPN) Support Using PIM Control Plane and Selective Provider Multicast Service Interface (S-PMSI) Join Messages*

RFC 6625, *Wildcards in Multicast VPN Auto-Discover Routes*

RFC 6826, *Multipoint LDP In-Band Signaling for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Path*

RFC 7246, *Multipoint Label Distribution Protocol In-Band Signaling in a Virtual Routing and Forwarding (VRF) Table Context*

RFC 7385, *IANA Registry for P-Multicast Service Interface (PMSI) Tunnel Type Code Points*

RFC 7716, *Global Table Multicast with BGP Multicast VPN (BGP-MVPN) Procedures*

RFC 7761, *Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)*

RFC 8279, *Multicast Using Bit Index Explicit Replication (BIER)*

RFC 8296, *Encapsulation for Bit Index Explicit Replication (BIER) in MPLS and Non-MPLS Networks* – MPLS encapsulation

RFC 8401, *Bit Index Explicit Replication (BIER) Support via IS-IS*

RFC 8444, *OSPFv2 Extensions for Bit Index Explicit Replication (BIER)*

RFC 8487, *Mtrace Version 2: Traceroute Facility for IP Multicast*

RFC 8534, *Explicit Tracking with Wildcard Routes in Multicast VPN* – (C-*,C-*) wildcard

RFC 8556, *Multicast VPN Using Bit Index Explicit Replication (BIER)*

21.15 Internet Protocol (IP) version 4

RFC 791, *Internet Protocol*

RFC 792, *Internet Control Message Protocol*

RFC 826, *An Ethernet Address Resolution Protocol*

RFC 951, *Bootstrap Protocol (BOOTP)* – relay

RFC 1034, *Domain Names - Concepts and Facilities*

RFC 1035, *Domain Names - Implementation and Specification*

RFC 1191, *Path MTU Discovery* – router specification

RFC 1519, *Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy*

RFC 1534, *Interoperation between DHCP and BOOTP*

RFC 1542, *Clarifications and Extensions for the Bootstrap Protocol*

RFC 1812, *Requirements for IPv4 Routers*

RFC 1918, *Address Allocation for Private Internets*

RFC 2003, *IP Encapsulation within IP*

RFC 2131, *Dynamic Host Configuration Protocol*

RFC 2132, *DHCP Options and BOOTP Vendor Extensions*

RFC 2401, *Security Architecture for Internet Protocol*

RFC 3021, *Using 31-Bit Prefixes on IPv4 Point-to-Point Links*

RFC 3046, *DHCP Relay Agent Information Option (Option 82)*

RFC 3768, *Virtual Router Redundancy Protocol (VRRP)*

RFC 4884, *Extended ICMP to Support Multi-Part Messages* – ICMPv4 and ICMPv6 Time Exceeded

21.16 Internet Protocol (IP) version 6

RFC 2464, *Transmission of IPv6 Packets over Ethernet Networks*

RFC 2529, *Transmission of IPv6 over IPv4 Domains without Explicit Tunnels*

RFC 3122, *Extensions to IPv6 Neighbor Discovery for Inverse Discovery Specification*

RFC 3315, *Dynamic Host Configuration Protocol for IPv6 (DHCPv6)*

RFC 3587, *IPv6 Global Unicast Address Format*

RFC 3596, *DNS Extensions to Support IP version 6*

RFC 3633, *IPv6 Prefix Options for Dynamic Host Configuration Protocol (DHCP) version 6*

RFC 3646, *DNS Configuration options for Dynamic Host Configuration Protocol for IPv6 (DHCPv6)*

RFC 3736, *Stateless Dynamic Host Configuration Protocol (DHCP) Service for IPv6*

RFC 3971, *SEcure Neighbor Discovery (SEND)*

RFC 3972, *Cryptographically Generated Addresses (CGA)*

RFC 4007, *IPv6 Scoped Address Architecture*

RFC 4191, *Default Router Preferences and More-Specific Routes* – Default Router Preference

RFC 4193, *Unique Local IPv6 Unicast Addresses*

RFC 4291, *Internet Protocol Version 6 (IPv6) Addressing Architecture*

RFC 4443, *Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification*

RFC 4861, *Neighbor Discovery for IP version 6 (IPv6)*

RFC 4862, *IPv6 Stateless Address Autoconfiguration* – router functions

RFC 4890, *Recommendations for Filtering ICMPv6 Messages in Firewalls*

RFC 4941, *Privacy Extensions for Stateless Address Autoconfiguration in IPv6*

RFC 5007, *DHCPv6 Leasequery*

RFC 5095, *Deprecation of Type 0 Routing Headers in IPv6*

RFC 5722, *Handling of Overlapping IPv6 Fragments*

RFC 5798, *Virtual Router Redundancy Protocol (VRRP) Version 3 for IPv4 and IPv6* – IPv6

RFC 5952, *A Recommendation for IPv6 Address Text Representation*

RFC 6092, *Recommended Simple Security Capabilities in Customer Premises Equipment (CPE) for Providing Residential IPv6 Internet Service* – Internet Control and Management, Upper-Layer Transport Protocols, UDP Filters, IPsec and Internet Key Exchange (IKE), TCP Filters

RFC 6106, *IPv6 Router Advertisement Options for DNS Configuration*

RFC 6164, *Using 127-Bit IPv6 Prefixes on Inter-Router Links*

RFC 6221, *Lightweight DHCPv6 Relay Agent*

RFC 6437, *IPv6 Flow Label Specification*

RFC 6603, *Prefix Exclude Option for DHCPv6-based Prefix Delegation*

RFC 8021, *Generation of IPv6 Atomic Fragments Considered Harmful*

RFC 8200, *Internet Protocol, Version 6 (IPv6) Specification*

RFC 8201, *Path MTU Discovery for IP version 6*

21.17 Internet Protocol Security (IPsec)

draft-ietf-ipsec-isakmp-mode-cfg-05, *The ISAKMP Configuration Method*

draft-ietf-ipsec-isakmp-xauth-06, *Extended Authentication within ISAKMP/Oakley (XAUTH)*

RFC 2401, *Security Architecture for the Internet Protocol*

RFC 2403, *The Use of HMAC-MD5-96 within ESP and AH*

RFC 2404, *The Use of HMAC-SHA-1-96 within ESP and AH*

RFC 2405, *The ESP DES-CBC Cipher Algorithm With Explicit IV*

RFC 2406, *IP Encapsulating Security Payload (ESP)*

RFC 2407, *IPsec Domain of Interpretation for ISAKMP (IPsec DoI)*

RFC 2408, *Internet Security Association and Key Management Protocol (ISAKMP)*

RFC 2409, *The Internet Key Exchange (IKE)*

RFC 2410, *The NULL Encryption Algorithm and Its Use With IPsec*

RFC 2560, *X.509 Internet Public Key Infrastructure Online Certificate Status Protocol - OCSP*

RFC 3526, *More Modular Exponential (MODP) Diffie-Hellman group for Internet Key Exchange (IKE)*

RFC 3566, *The AES-XCBC-MAC-96 Algorithm and Its Use With IPsec*

RFC 3602, *The AES-CBC Cipher Algorithm and Its Use with IPsec*

RFC 3706, *A Traffic-Based Method of Detecting Dead Internet Key Exchange (IKE) Peers*

RFC 3947, *Negotiation of NAT-Traversal in the IKE*

RFC 3948, *UDP Encapsulation of IPsec ESP Packets*

RFC 4106, *The Use of Galois/Counter Mode (GCM) in IPsec ESP*

RFC 4109, *Algorithms for Internet Key Exchange version 1 (IKEv1)*

RFC 4301, *Security Architecture for the Internet Protocol*

RFC 4303, *IP Encapsulating Security Payload*

RFC 4307, *Cryptographic Algorithms for Use in the Internet Key Exchange Version 2 (IKEv2)*

RFC 4308, *Cryptographic Suites for IPsec*

RFC 4434, *The AES-XCBC-PRF-128 Algorithm for the Internet Key Exchange Protocol (IKE)*

RFC 4543, *The Use of Galois Message Authentication Code (GMAC) in IPsec ESP and AH*

RFC 4754, *IKE and IKEv2 Authentication Using the Elliptic Curve Digital Signature Algorithm (ECDSA)*

RFC 4835, *Cryptographic Algorithm Implementation Requirements for Encapsulating Security Payload (ESP) and Authentication Header (AH)*

RFC 4868, *Using HMAC-SHA-256, HMAC-SHA-384, and HMAC-SHA-512 with IPsec*

RFC 4945, *The Internet IP Security PKI Profile of IKEv1/ISAKMP, IKEv2 and PKIX*

RFC 5019, *The Lightweight Online Certificate Status Protocol (OCSP) Profile for High-Volume Environments*

RFC 5282, *Using Authenticated Encryption Algorithms with the Encrypted Payload of the IKEv2 Protocol*

RFC 5903, *ECP Groups for IKE and IKEv2*

RFC 5996, *Internet Key Exchange Protocol Version 2 (IKEv2)*

RFC 5998, *An Extension for EAP-Only Authentication in IKEv2*

RFC 6379, *Suite B Cryptographic Suites for IPsec*

RFC 6380, *Suite B Profile for Internet Protocol Security (IPsec)*

RFC 6960, *X.509 Internet Public Key Infrastructure Online Certificate Status Protocol - OCSP*

RFC 7296, *Internet Key Exchange Protocol Version 2 (IKEv2)*

RFC 7321, *Cryptographic Algorithm Implementation Requirements and Usage Guidance for Encapsulating Security Payload (ESP) and Authentication Header (AH)*

RFC 7383, *Internet Key Exchange Protocol Version 2 (IKEv2) Message Fragmentation*

RFC 7427, *Signature Authentication in the Internet Key Exchange Version 2 (IKEv2)*

21.18 Label Distribution Protocol (LDP)

draft-pdutta-mpls-ldp-adj-capability-00, *LDP Adjacency Capabilities*

draft-pdutta-mpls-ldp-v2-00, *LDP Version 2*

draft-pdutta-mpls-mldp-up-redundancy-00, *Upstream LSR Redundancy for Multi-point LDP Tunnels*

draft-pdutta-mpls-multi-ldp-instance-00, *Multiple LDP Instances*

draft-pdutta-mpls-tldp-hello-reduce-04, *Targeted LDP Hello Reduction*

RFC 3037, *LDP Applicability*

RFC 3478, *Graceful Restart Mechanism for Label Distribution Protocol* – helper mode

RFC 5036, *LDP Specification*

RFC 5283, *LDP Extension for Inter-Area Label Switched Paths (LSPs)*

RFC 5443, *LDP IGP Synchronization*

RFC 5561, *LDP Capabilities*

RFC 5919, *Signaling LDP Label Advertisement Completion*

RFC 6388, *Label Distribution Protocol Extensions for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths*

RFC 6512, *Using Multipoint LDP When the Backbone Has No Route to the Root*

RFC 6826, *Multipoint LDP in-band signaling for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths*

RFC 7032, *LDP Downstream-on-Demand in Seamless MPLS*

RFC 7473, *Controlling State Advertisements of Non-negotiated LDP Applications*

RFC 7552, *Updates to LDP for IPv6*

21.19 Layer Two Tunneling Protocol (L2TP) Network Server (LNS)

draft-mammoliti-l2tp-accessline-avp-04, *Layer 2 Tunneling Protocol (L2TP) Access Line Information Attribute Value Pair (AVP) Extensions*

RFC 2661, *Layer Two Tunneling Protocol "L2TP"*

RFC 2809, *Implementation of L2TP Compulsory Tunneling via RADIUS*

RFC 3438, *Layer Two Tunneling Protocol (L2TP) Internet Assigned Numbers: Internet Assigned Numbers Authority (IANA) Considerations Update*

RFC 3931, *Layer Two Tunneling Protocol - Version 3 (L2TPv3)*

RFC 4719, *Transport of Ethernet Frames over Layer 2 Tunneling Protocol Version 3 (L2TPv3)*

RFC 4951, *Fail Over Extensions for Layer 2 Tunneling Protocol (L2TP) "failover"*

21.20 Multiprotocol Label Switching (MPLS)

draft-ietf-mpls-lsp-ping-ospfv3-codepoint-02, *OSPFv3 CodePoint for MPLS LSP Ping*

RFC 3031, *Multiprotocol Label Switching Architecture*

RFC 3032, *MPLS Label Stack Encoding*

RFC 3270, *Multi-Protocol Label Switching (MPLS) Support of Differentiated Services* – E-LSP

RFC 3443, *Time To Live (TTL) Processing in Multi-Protocol Label Switching (MPLS) Networks*

RFC 4023, *Encapsulating MPLS in IP or Generic Routing Encapsulation (GRE)*

RFC 4182, *Removing a Restriction on the use of MPLS Explicit NULL*

RFC 4950, *ICMP Extensions for Multiprotocol Label Switching*

RFC 5332, *MPLS Multicast Encapsulations*

RFC 5884, *Bidirectional Forwarding Detection (BFD) for MPLS Label Switched Paths (LSPs)*

RFC 6374, *Packet Loss and Delay Measurement for MPLS Networks* – Delay Measurement, Channel Type 0x000C

RFC 6424, *Mechanism for Performing Label Switched Path Ping (LSP Ping) over MPLS Tunnels*

RFC 6425, *Detecting Data Plane Failures in Point-to-Multipoint Multiprotocol Label Switching (MPLS) - Extensions to LSP Ping*

RFC 6790, *The Use of Entropy Labels in MPLS Forwarding*

RFC 7308, *Extended Administrative Groups in MPLS Traffic Engineering (MPLS-TE)*

RFC 7510, *Encapsulating MPLS in UDP*

RFC 7746, *Label Switched Path (LSP) Self-Ping*

RFC 7876, *UDP Return Path for Packet Loss and Delay Measurement for MPLS Networks* – Delay Measurement

RFC 8029, *Detecting Multiprotocol Label Switched (MPLS) Data-Plane Failures*

21.21 Multiprotocol Label Switching - Transport Profile (MPLS-TP)

RFC 5586, *MPLS Generic Associated Channel*

RFC 5921, *A Framework for MPLS in Transport Networks*

RFC 5960, *MPLS Transport Profile Data Plane Architecture*

RFC 6370, *MPLS Transport Profile (MPLS-TP) Identifiers*

RFC 6378, *MPLS Transport Profile (MPLS-TP) Linear Protection*

RFC 6426, *MPLS On-Demand Connectivity and Route Tracing*

RFC 6427, *MPLS Fault Management Operations, Administration, and Maintenance (OAM)*

RFC 6428, *Proactive Connectivity Verification, Continuity Check and Remote Defect indication for MPLS Transport Profile*

RFC 6478, *Pseudowire Status for Static Pseudowires*

RFC 7213, *MPLS Transport Profile (MPLS-TP) Next-Hop Ethernet Addressing*

21.22 Network Address Translation (NAT)

draft-ietf-behave-address-format-10, *IPv6 Addressing of IPv4/IPv6 Translators*

draft-ietf-behave-v6v4-xlate-23, *IP/ICMP Translation Algorithm*

draft-miles-behave-l2nat-00, *Layer2-Aware NAT*

draft-nishitani-cgn-02, *Common Functions of Large Scale NAT (LSN)*

RFC 4787, *Network Address Translation (NAT) Behavioral Requirements for Unicast UDP*

RFC 5382, *NAT Behavioral Requirements for TCP*

RFC 5508, *NAT Behavioral Requirements for ICMP*

RFC 6146, *Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers*

RFC 6333, *Dual-Stack Lite Broadband Deployments Following IPv4 Exhaustion*

RFC 6334, *Dynamic Host Configuration Protocol for IPv6 (DHCPv6) Option for Dual-Stack Lite*

RFC 6887, *Port Control Protocol (PCP)*

RFC 6888, *Common Requirements For Carrier-Grade NATs (CGNs)*

RFC 7753, *Port Control Protocol (PCP) Extension for Port-Set Allocation*

RFC 7915, *IP/ICMP Translation Algorithm*

21.23 Network Configuration Protocol (NETCONF)

RFC 5277, *NETCONF Event Notifications*

RFC 6020, *YANG - A Data Modeling Language for the Network Configuration Protocol (NETCONF)*

RFC 6022, *YANG Module for NETCONF Monitoring*

RFC 6241, *Network Configuration Protocol (NETCONF)*

RFC 6242, *Using the NETCONF Protocol over Secure Shell (SSH)*

RFC 6243, *With-defaults Capability for NETCONF*

RFC 8342, *Network Management Datastore Architecture (NMDA)* – Startup, Candidate, Running and Intended datastores

RFC 8525, *YANG Library*

RFC 8526, *NETCONF Extensions to Support the Network Management Datastore Architecture* – <get-data> operation

21.24 Open Shortest Path First (OSPF)

RFC 1765, *OSPF Database Overflow*

RFC 2328, *OSPF Version 2*

RFC 3101, *The OSPF Not-So-Stubby Area (NSSA) Option*

RFC 3509, *Alternative Implementations of OSPF Area Border Routers*

RFC 3623, *Graceful OSPF Restart Graceful OSPF Restart* – helper mode

RFC 3630, *Traffic Engineering (TE) Extensions to OSPF Version 2*

RFC 4222, *Prioritized Treatment of Specific OSPF Version 2 Packets and Congestion Avoidance*

RFC 4552, *Authentication/Confidentiality for OSPFv3*

RFC 4576, *Using a Link State Advertisement (LSA) Options Bit to Prevent Looping in BGP/MPLS IP Virtual Private Networks (VPNs)*

RFC 4577, *OSPF as the Provider/Customer Edge Protocol for BGP/MPLS IP Virtual Private Networks (VPNs)*

RFC 5185, *OSPF Multi-Area Adjacency*

RFC 5187, *OSPFv3 Graceful Restart* – helper mode

RFC 5243, *OSPF Database Exchange Summary List Optimization*

RFC 5250, *The OSPF Opaque LSA Option*

RFC 5309, *Point-to-Point Operation over LAN in Link State Routing Protocols*

RFC 5340, *OSPF for IPv6*

RFC 5642, *Dynamic Hostname Exchange Mechanism for OSPF*

RFC 5709, *OSPFv2 HMAC-SHA Cryptographic Authentication*

RFC 5838, *Support of Address Families in OSPFv3*

RFC 6549, *OSPFv2 Multi-Instance Extensions*

RFC 6987, *OSPF Stub Router Advertisement*

RFC 7471, *OSPF Traffic Engineering (TE) Metric Extensions* – Min/Max Unidirectional Link Delay metric for flex-algo, RSVP, SR-TE

RFC 7684, *OSPFv2 Prefix/Link Attribute Advertisement*

RFC 7770, *Extensions to OSPF for Advertising Optional Router Capabilities*

RFC 8362, *OSPFv3 Link State Advertisement (LSA) Extensibility*

RFC 8920, *OSPF Application-Specific Link Attributes*

21.25 OpenFlow

TS-007 Version 1.3.1, *OpenFlow Switch Specification* – OpenFlow-hybrid switches

21.26 Path Computation Element Protocol (PCEP)

draft-alvarez-pce-path-profiles-04, *PCE Path Profiles*

draft-dhs-spring-pce-sr-p2mp-policy-00, *PCEP extensions for p2mp sr policy*

draft-ietf-pce-binding-label-sid-15, *Carrying Binding Label/Segment Identifier (SID) in PCE-based Networks.* – MPLS binding SIDs

draft-ietf-pce-pceps-tls13-04, *Updates for PCEPS: TLS Connection Establishment Restrictions*

RFC 5440, *Path Computation Element (PCE) Communication Protocol (PCEP)*

RFC 8231, *Path Computation Element Communication Protocol (PCEP) Extensions for Stateful PCE*

RFC 8253, *PCEPS: Usage of TLS to Provide a Secure Transport for the Path Computation Element Communication Protocol (PCEP)*

RFC 8281, *PCEP Extensions for PCE-initiated LSP Setup in a Stateful PCE Model*

RFC 8408, *Conveying Path Setup Type in PCE Communication Protocol (PCEP) Messages*

RFC 8664, *Path Computation Element Communication Protocol (PCEP) Extensions for Segment Routing*

21.27 Point-to-Point Protocol (PPP)

RFC 1332, *The PPP Internet Protocol Control Protocol (IPCP)*

RFC 1990, *The PPP Multilink Protocol (MP)*

RFC 1994, *PPP Challenge Handshake Authentication Protocol (CHAP)*

RFC 2516, *A Method for Transmitting PPP Over Ethernet (PPPoE)*

RFC 4638, *Accommodating a Maximum Transit Unit/Maximum Receive Unit (MTU/MRU) Greater Than 1492 in the Point-to-Point Protocol over Ethernet (PPPoE)*

RFC 5072, *IP Version 6 over PPP*

21.28 Policy management and credit control

3GPP TS 29.212 Release 11, *Policy and Charging Control (PCC); Reference points* – Gx support as it applies to wireline environment (BNG)

RFC 4006, *Diameter Credit-Control Application*

RFC 6733, *Diameter Base Protocol*

21.29 Pseudowire (PW)

draft-ietf-l2vpn-vpws-iw-oam-04, *OAM Procedures for VPWS Interworking*

MFA Forum 12.0.0, *Multiservice Interworking - Ethernet over MPLS*

MFA Forum 13.0.0, *Fault Management for Multiservice Interworking v1.0*

MFA Forum 16.0.0, *Multiservice Interworking - IP over MPLS*

RFC 3916, *Requirements for Pseudo-Wire Emulation Edge-to-Edge (PWE3)*

RFC 3985, *Pseudo Wire Emulation Edge-to-Edge (PWE3)*

RFC 4385, *Pseudo Wire Emulation Edge-to-Edge (PWE3) Control Word for Use over an MPLS PSN*

RFC 4446, *IANA Allocations for Pseudowire Edge to Edge Emulation (PWE3)*

RFC 4447, *Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP)*

RFC 4448, *Encapsulation Methods for Transport of Ethernet over MPLS Networks*

RFC 5085, *Pseudowire Virtual Circuit Connectivity Verification (VCCV): A Control Channel for Pseudowires*

RFC 5659, *An Architecture for Multi-Segment Pseudowire Emulation Edge-to-Edge*

RFC 5885, *Bidirectional Forwarding Detection (BFD) for the Pseudowire Virtual Circuit Connectivity Verification (VCCV)*

RFC 6073, *Segmented Pseudowire*

RFC 6310, *Pseudowire (PW) Operations, Administration, and Maintenance (OAM) Message Mapping*

RFC 6391, *Flow-Aware Transport of Pseudowires over an MPLS Packet Switched Network*

RFC 6575, *Address Resolution Protocol (ARP) Mediation for IP Interworking of Layer 2 VPNs*

RFC 6718, *Pseudowire Redundancy*

RFC 6829, *Label Switched Path (LSP) Ping for Pseudowire Forwarding Equivalence Classes (FECs) Advertised over IPv6*

RFC 6870, *Pseudowire Preferential Forwarding Status bit*

RFC 7023, *MPLS and Ethernet Operations, Administration, and Maintenance (OAM) Interworking*

RFC 7267, *Dynamic Placement of Multi-Segment Pseudowires*

RFC 7392, *Explicit Path Routing for Dynamic Multi-Segment Pseudowires* – ER-TLV and ER-HOP IPv4 Prefix

RFC 8395, *Extensions to BGP-Signaled Pseudowires to Support Flow-Aware Transport Labels*

21.30 Quality of Service (QoS)

RFC 2430, *A Provider Architecture for Differentiated Services and Traffic Engineering (PASTE)*

RFC 2474, *Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers*

RFC 2597, *Assured Forwarding PHB Group*

RFC 3140, *Per Hop Behavior Identification Codes*

RFC 3246, *An Expedited Forwarding PHB (Per-Hop Behavior)*

21.31 Remote Authentication Dial In User Service (RADIUS)

draft-oscca-cfrg-sm3-02, *The SM3 Cryptographic Hash Function*

RFC 2865, *Remote Authentication Dial In User Service (RADIUS)*

RFC 2866, *RADIUS Accounting*

RFC 2867, *RADIUS Accounting Modifications for Tunnel Protocol Support*

RFC 2868, *RADIUS Attributes for Tunnel Protocol Support*

RFC 2869, *RADIUS Extensions*

RFC 3162, *RADIUS and IPv6*

RFC 4818, *RADIUS Delegated-IPv6-Prefix Attribute*

RFC 5176, *Dynamic Authorization Extensions to RADIUS*

RFC 6613, *RADIUS over TCP* – with TLS

RFC 6614, *Transport Layer Security (TLS) Encryption for RADIUS*

RFC 6929, *Remote Authentication Dial-In User Service (RADIUS) Protocol Extensions*

RFC 6911, *RADIUS attributes for IPv6 Access Networks*

21.32 Resource Reservation Protocol - Traffic Engineering (RSVP-TE)

draft-newton-mpls-te-dynamic-overbooking-00, *A Diffserv-TE Implementation Model to dynamically change booking factors during failure events*

RFC 2702, *Requirements for Traffic Engineering over MPLS*

RFC 2747, *RSVP Cryptographic Authentication*

RFC 2961, *RSVP Refresh Overhead Reduction Extensions*

RFC 3097, *RSVP Cryptographic Authentication -- Updated Message Type Value*

RFC 3209, *RSVP-TE: Extensions to RSVP for LSP Tunnels*

RFC 3477, *Signalling Unnumbered Links in Resource ReSerVation Protocol - Traffic Engineering (RSVP-TE)*

RFC 3564, *Requirements for Support of Differentiated Services-aware MPLS Traffic Engineering*

RFC 3906, *Calculating Interior Gateway Protocol (IGP) Routes Over Traffic Engineering Tunnels*

RFC 4090, *Fast Reroute Extensions to RSVP-TE for LSP Tunnels*

RFC 4124, *Protocol Extensions for Support of Diffserv-aware MPLS Traffic Engineering*

RFC 4125, *Maximum Allocation Bandwidth Constraints Model for Diffserv-aware MPLS Traffic Engineering*

RFC 4127, *Russian Dolls Bandwidth Constraints Model for Diffserv-aware MPLS Traffic Engineering*

RFC 4561, *Definition of a Record Route Object (RRO) Node-Id Sub-Object*

RFC 4875, *Extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs)*

RFC 5712, *MPLS Traffic Engineering Soft Preemption*

RFC 5817, *Graceful Shutdown in MPLS and Generalized MPLS Traffic Engineering Networks*

21.33 Routing Information Protocol (RIP)

RFC 1058, *Routing Information Protocol*

RFC 2080, *RIPng for IPv6*

RFC 2082, *RIP-2 MD5 Authentication*

RFC 2453, *RIP Version 2*

21.34 Segment Routing (SR)

draft-ietf-bess-mvpn-evpn-sr-p2mp-07, *Multicast and Ethernet VPN with Segment Routing P2MP and Ingress Replication* – MVPN

draft-bashandy-rtgwg-segment-routing-uloop-15, *Loop avoidance using Segment Routing*

draft-filsfils-spring-net-pgm-extension-srv6-usid-15, *Network Programming extension: SRv6 uSID instruction*

draft-filsfils-spring-srv6-net-pgm-insertion-08, *SRv6 NET-PGM extension: Insertion*

draft-ietf-idr-bgpls-srv6-ext-14, *BGP Link State Extensions for SRv6*

draft-ietf-idr-segment-routing-te-policy-23, *Advertising Segment Routing Policies in BGP*

draft-ietf-idr-ts-flowspec-srv6-policy-03, *Traffic Steering using BGP FlowSpec with SR Policy*

draft-ietf-pim-p2mp-policy-ping-03, *P2MP Policy Ping*

draft-ietf-pim-sr-p2mp-policy-06, *Segment Routing Point-to-Multipoint Policy* – MPLS

draft-ietf-rtgwg-segment-routing-ti-lfa-11, *Topology Independent Fast Reroute using Segment Routing*

draft-ietf-spring-conflict-resolution-05, *Segment Routing MPLS Conflict Resolution*

draft-ietf-spring-sr-replication-segment-16, *SR Replication segment for Multi-point Service Delivery* – MPLS

draft-ietf-spring-srv6-srh-compression-xx, *Compressed SRv6 Segment List Encoding in SRH*

draft-voyer-6man-extension-header-insertion-10, *Deployments With Insertion of IPv6 Segment Routing Headers*

RFC 8287, *Label Switched Path (LSP) Ping/Traceroute for Segment Routing (SR) IGP-Prefix and IGP-Adjacency Segment Identifiers (SIDs) with MPLS Data Planes*

RFC 8426, *Recommendations for RSVP-TE and Segment Routing (SR) Label Switched Path (LSP) Coexistence*

RFC 8476, *Signaling Maximum SID Depth (MSD) Using OSPF* – node MSD

RFC 8491, *Signaling Maximum SID Depth (MSD) Using IS-IS* – node MSD

RFC 8660, *Segment Routing with the MPLS Data Plane*

RFC 8661, *Segment Routing MPLS Interworking with LDP*

RFC 8663, *MPLS Segment Routing over IP* – BGP SR with SR-MPLS-over-UDP/IP

RFC 8665, *OSPF Extensions for Segment Routing*

RFC 8666, *OSPFv3 Extensions for Segment Routing*

RFC 8667, *IS-IS Extensions for Segment Routing*

RFC 8669, *Segment Routing Prefix Segment Identifier Extensions for BGP*

RFC 8754, *IPv6 Segment Routing Header (SRH)*

RFC 8814, *Signaling Maximum SID Depth (MSD) Using the Border Gateway Protocol - Link State*

RFC 8986, *Segment Routing over IPv6 (SRv6) Network Programming*

RFC 9085, *Border Gateway Protocol - Link State (BGP-LS) Extensions for Segment Routing*

RFC 9088, *Signaling Entropy Label Capability and Entropy Readable Label Depth Using IS-IS* – advertising ELC

RFC 9089, *Signaling Entropy Label Capability and Entropy Readable Label Depth Using OSPF* – advertising ELC

RFC 9252, *BGP Overlay Services Based on Segment Routing over IPv6 (SRv6)*

RFC 9256, *Segment Routing Policy Architecture*

RFC 9259, *Operations, Administration, and Maintenance (OAM) in Segment Routing over IPv6 (SRv6)*

RFC 9350, *IGP Flexible Algorithm*

RFC 9352, *IS-IS Extensions to Support Segment Routing over the IPv6 Data Plane*

21.35 Simple Network Management Protocol (SNMP)

draft-blumenthal-aes-usm-04, *The AES Cipher Algorithm in the SNMP's User-based Security Model* – CFB128-AES-192 and CFB128-AES-256

draft-ietf-isis-wg-mib-06, *Management Information Base for Intermediate System to Intermediate System (IS-IS)*

draft-ietf-mboned-msdp-mib-01, *Multicast Source Discovery protocol MIB*

draft-ietf-mpls-ldp-mib-07, *Definitions of Managed Objects for the Multiprotocol Label Switching, Label Distribution Protocol (LDP)*

draft-ietf-mpls-lsr-mib-06, *Multiprotocol Label Switching (MPLS) Label Switching Router (LSR) Management Information Base Using SMIv2*

draft-ietf-mpls-te-mib-04, *Multiprotocol Label Switching (MPLS) Traffic Engineering Management Information Base*

draft-ietf-ospf-mib-update-08, *OSPF Version 2 Management Information Base*

draft-ietf-vrrp-unified-mib-06, *Definitions of Managed Objects for the VRRP over IPv4 and IPv6* – IPv6

ESO-CONSORTIUM-MIB revision 200406230000Z, *esoConsortiumMIB*

IANA-ADDRESS-FAMILY-NUMBERS-MIB revision 200203140000Z, *ianaAddressFamilyNumbers*

IANAifType-MIB revision 200505270000Z, *ianaifType*

IANA-RTPROTO-MIB revision 200009260000Z, *ianaRtProtoMIB*

IEEE8021-CFM-MIB revision 200706100000Z, *ieee8021CfmMib*

IEEE8021-PAE-MIB revision 200101160000Z, *ieee8021paeMIB*

IEEE8023-LAG-MIB revision 200006270000Z, *lagMIB*

LLDP-MIB revision 200505060000Z, *lldpMIB*

RFC 1157, *A Simple Network Management Protocol (SNMP)*

RFC 1212, *Concise MIB Definitions*

RFC 1215, *A Convention for Defining Traps for use with the SNMP*

RFC 1724, *RIP Version 2 MIB Extension*

RFC 1901, *Introduction to Community-based SNMPv2*

RFC 2021, *Remote Network Monitoring Management Information Base Version 2 using SMIv2*

RFC 2206, *RSVP Management Information Base using SMIv2*

RFC 2213, *Integrated Services Management Information Base using SMIv2*

RFC 2494, *Definitions of Managed Objects for the DS0 and DS0 Bundle Interface Type*

RFC 2578, *Structure of Management Information Version 2 (SMIv2)*

RFC 2579, *Textual Conventions for SMIv2*

RFC 2580, *Conformance Statements for SMIv2*

RFC 2787, *Definitions of Managed Objects for the Virtual Router Redundancy Protocol*

RFC 2819, *Remote Network Monitoring Management Information Base*

RFC 2856, *Textual Conventions for Additional High Capacity Data Types*

RFC 2863, *The Interfaces Group MIB*

RFC 2864, *The Inverted Stack Table Extension to the Interfaces Group MIB*

RFC 2933, *Internet Group Management Protocol MIB*

RFC 3014, *Notification Log MIB*

RFC 3165, *Definitions of Managed Objects for the Delegation of Management Scripts*

RFC 3231, *Definitions of Managed Objects for Scheduling Management Operations*

RFC 3273, *Remote Network Monitoring Management Information Base for High Capacity Networks*

RFC 3410, *Introduction and Applicability Statements for Internet Standard Management Framework*

RFC 3411, *An Architecture for Describing Simple Network Management Protocol (SNMP) Management Frameworks*

RFC 3412, *Message Processing and Dispatching for the Simple Network Management Protocol (SNMP)*

RFC 3413, *Simple Network Management Protocol (SNMP) Applications*

RFC 3414, *User-based Security Model (USM) for version 3 of the Simple Network Management Protocol (SNMPv3)*

RFC 3415, *View-based Access Control Model (VACM) for the Simple Network Management Protocol (SNMP)*

RFC 3416, *Version 2 of the Protocol Operations for the Simple Network Management Protocol (SNMP)*

RFC 3417, *Transport Mappings for the Simple Network Management Protocol (SNMP)* – SNMP over UDP over IPv4

RFC 3418, *Management Information Base (MIB) for the Simple Network Management Protocol (SNMP)*

RFC 3419, *Textual Conventions for Transport Addresses*

RFC 3498, *Definitions of Managed Objects for Synchronous Optical Network (SONET) Linear Automatic Protection Switching (APS) Architectures*

RFC 3584, *Coexistence between Version 1, Version 2, and Version 3 of the Internet-standard Network Management Framework*

RFC 3592, *Definitions of Managed Objects for the Synchronous Optical Network/Synchronous Digital Hierarchy (SONET/SDH) Interface Type*

RFC 3593, *Textual Conventions for MIB Modules Using Performance History Based on 15 Minute Intervals*

RFC 3635, *Definitions of Managed Objects for the Ethernet-like Interface Types*

RFC 3637, *Definitions of Managed Objects for the Ethernet WAN Interface Sublayer*

RFC 3826, *The Advanced Encryption Standard (AES) Cipher Algorithm in the SNMP User-based Security Model*

RFC 3877, *Alarm Management Information Base (MIB)*

RFC 3895, *Definitions of Managed Objects for the DS1, E1, DS2, and E2 Interface Types*

RFC 3896, *Definitions of Managed Objects for the DS3/E3 Interface Type*

RFC 4001, *Textual Conventions for Internet Network Addresses*

RFC 4022, *Management Information Base for the Transmission Control Protocol (TCP)*

RFC 4113, *Management Information Base for the User Datagram Protocol (UDP)*

RFC 4220, *Traffic Engineering Link Management Information Base*

RFC 4273, *Definitions of Managed Objects for BGP-4*

RFC 4292, *IP Forwarding Table MIB*

RFC 4293, *Management Information Base for the Internet Protocol (IP)*

RFC 4631, *Link Management Protocol (LMP) Management Information Base (MIB)*

RFC 4878, *Definitions and Managed Objects for Operations, Administration, and Maintenance (OAM) Functions on Ethernet-Like Interfaces*

RFC 7420, *Path Computation Element Communication Protocol (PCEP) Management Information Base (MIB) Module*

RFC 7630, *HMAC-SHA-2 Authentication Protocols in the User-based Security Model (USM) for SNMPv3*

SFLOW-MIB revision 200309240000Z, *sFlowMIB*

21.36 Timing

GR-1244-CORE Issue 3, *Clocks for the Synchronized Network: Common Generic Criteria*

GR-253-CORE Issue 3, *SONET Transport Systems: Common Generic Criteria*

IEEE 1588-2008, *IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems*

ITU-T G.781, *Synchronization layer functions*

ITU-T G.811, *Timing characteristics of primary reference clocks*

ITU-T G.813, *Timing characteristics of SDH equipment slave clocks (SEC)*

ITU-T G.8261, *Timing and synchronization aspects in packet networks*

ITU-T G.8262, *Timing characteristics of synchronous Ethernet equipment slave clock (EEC)*

ITU-T G.8262.1, *Timing characteristics of an enhanced synchronous Ethernet equipment slave clock (eEEC)*

ITU-T G.8264, *Distribution of timing information through packet networks*

ITU-T G.8265.1, *Precision time protocol telecom profile for frequency synchronization*

ITU-T G.8272 , *Timing characteristics of primary reference time clocks* – PRTC-A, PRTC-B

ITU-T G.8275.1, *Precision time protocol telecom profile for phase/time synchronization with full timing support from the network*

ITU-T G.8275.2, *Precision time protocol telecom profile for phase/time synchronization with partial timing support from the network*

RFC 3339, *Date and Time on the Internet: Timestamps*

RFC 5905, *Network Time Protocol Version 4: Protocol and Algorithms Specification*

RFC 8573, *Message Authentication Code for the Network Time Protocol*

21.37 Two-Way Active Measurement Protocol (TWAMP)

RFC 5357, *A Two-Way Active Measurement Protocol (TWAMP)* – server, unauthenticated mode

RFC 5938, *Individual Session Control Feature for the Two-Way Active Measurement Protocol (TWAMP)*

RFC 6038, *Two-Way Active Measurement Protocol (TWAMP) Reflect Octets and Symmetrical Size Features*

RFC 8545, *Well-Known Port Assignments for the One-Way Active Measurement Protocol (OWAMP) and the Two-Way Active Measurement Protocol (TWAMP)* – TWAMP

RFC 8762, *Simple Two-Way Active Measurement Protocol* – unauthenticated

RFC 8972, *Simple Two-Way Active Measurement Protocol Optional Extensions* – unauthenticated

21.38 Virtual Private LAN Service (VPLS)

RFC 4761, *Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling*

RFC 4762, *Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling*

RFC 5501, *Requirements for Multicast Support in Virtual Private LAN Services*

RFC 6074, *Provisioning, Auto-Discovery, and Signaling in Layer 2 Virtual Private Networks (L2VPNs)*

RFC 7041, *Extensions to the Virtual Private LAN Service (VPLS) Provider Edge (PE) Model for Provider Backbone Bridging*

RFC 7117, *Multicast in Virtual Private LAN Service (VPLS)*

21.39 Voice and video

DVB BlueBook A86, *Transport of MPEG-2 TS Based DVB Services over IP Based Networks*

ETSI TS 101 329-5 Annex E, *QoS Measurement for VoIP - Method for determining an Equipment Impairment Factor using Passive Monitoring*

ITU-T G.1020 Appendix I, *Performance Parameter Definitions for Quality of Speech and other Voiceband Applications Utilizing IP Networks - Mean Absolute Packet Delay Variation & Markov Models*

ITU-T G.107, *The E Model - A computational model for use in planning*

ITU-T P.564, *Conformance testing for voice over IP transmission quality assessment models*

RFC 3550, *RTP: A Transport Protocol for Real-Time Applications* – Appendix A.8

RFC 4585, *Extended RTP Profile for Real-time Transport Control Protocol (RTCP)-Based Feedback (RTP/AVPF)*

RFC 4588, *RTP Retransmission Payload Format*

21.40 Yet Another Next Generation (YANG)

RFC 6991, *Common YANG Data Types*

RFC 7950, *The YANG 1.1 Data Modeling Language*

RFC 7951, *JSON Encoding of Data Modeled with YANG*

21.41 Yet Another Next Generation (YANG) OpenConfig Models

openconfig-aaa.yang version 0.4.0, *OpenConfig AAA Model*

openconfig-aaa-radius.yang version 0.3.0, *OpenConfig AAA RADIUS Model*

openconfig-aaa-tacacs.yang version 0.3.0, *OpenConfig AAA TACACS+ Model*

openconfig-acl.yang version 1.0.0, *OpenConfig ACL Model*

openconfig-alarms.yang version0.3.2, *OpenConfig System Alarms Model*

openconfig-bfd.yang version 0.2.2, *OpenConfig BFD Model*

openconfig-bgp.yang version 6.1.0, *OpenConfig BGP Model*

openconfig-bgp-common.yang version 6.0.0, *OpenConfig BGP Common Model*

openconfig-bgp-common-multiprotocol.yang version 6.0.0, *OpenConfig BGP Common Multiprotocol Model*

openconfig-bgp-common-structure.yang version 6.0.0, *OpenConfig BGP Common Structure Model*

openconfig-bgp-global.yang version 6.0.0, *OpenConfig BGP Global Model*

openconfig-bgp-neighbor.yang version 6.1.0, *OpenConfig BGP Neighbor Model*

openconfig-bgp-peer-group.yang version 6.1.0, *OpenConfig BGP Peer Group Model*

openconfig-bgp-policy.yang version 4.0.1, *OpenConfig BGP Policy Model*

openconfig-if-aggregate.yang version 2.4.3, *OpenConfig Interfaces Aggregated Model*

openconfig-if-ethernet.yang version 2.12.1, *OpenConfig Interfaces Ethernet Model*

openconfig-if-ip.yang version 3.1.0, *OpenConfig Interfaces IP Model*

openconfig-if-ip-ext.yang version 2.3.1, *OpenConfig Interfaces IP Extensions Model*

openconfig-igmp.yang version 0.2.0, *OpenConfig IGMP Model*

openconfig-interfaces.yang version 3.0.0, *OpenConfig Interfaces Model*

openconfig-isis.yang version 1.1.0, *OpenConfig IS-IS Model*

openconfig-isis-policy.yang version 0.5.0, *OpenConfig IS-IS Policy Model*

openconfig-isis-routing.yang version 1.1.0, *OpenConfig IS-IS Routing Model*

openconfig-lacp.yang version 1.3.0, *OpenConfig LACP Model*

openconfig-lldp.yang version 0.1.0, *OpenConfig LLDP Model*

openconfig-local-routing.yang version 1.2.0, *OpenConfig Local Routing Model*

openconfig-mpls.yang version 2.3.0, *OpenConfig MPLS Model*

openconfig-mpls-ldp.yang version 3.0.2, *OpenConfig MPLS LDP Model*

openconfig-mpls-rsvp.yang version 2.3.0, *OpenConfig MPLS RSVP Model*

openconfig-mpls-te.yang version 2.3.0, *OpenConfig MPLS TE Model*

openconfig-network-instance.yang version 1.1.0, *OpenConfig Network Instance Model*

openconfig-network-instance-l3.yang version 0.11.1, *OpenConfig L3 Network Instance Model* – static routes

openconfig-ospfv2.yang version 0.4.0, *OpenConfig OSPFv2 Model*

openconfig-ospfv2-area.yang version 0.4.0, *OpenConfig OSPFv2 Area Model*

openconfig-ospfv2-area-interface.yang version 0.4.0, *OpenConfig OSPFv2 Area Interface Model*

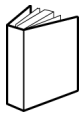
openconfig-ospfv2-common.yang version 0.4.0, *OpenConfig OSPFv2 Common Model*

openconfig-ospfv2-global.yang version 0.4.0, *OpenConfig OSPFv2 Global Model*

openconfig-packet-match.yang version 1.0.0, *OpenConfig Packet Match Model*

openconfig-pim.yang version 0.2.0, *OpenConfig PIM Model*
openconfig-platform.yang version 0.15.0, *OpenConfig Platform Model*
openconfig-platform-fan.yang version 0.1.1, *OpenConfig Platform Fan Model*
openconfig-platform-linecard.yang version 0.1.2, *OpenConfig Platform Linecard Model*
openconfig-platform-port.yang version 0.4.2, *OpenConfig Port Model*
openconfig-platform-transceiver.yang version 0.9.0, *OpenConfig Transceiver Model*
openconfig-procmon.yang version 0.4.0, *OpenConfig Process Monitoring Model*
openconfig-relay-agent.yang version 0.1.0, *OpenConfig Relay Agent Model*
openconfig-routing-policy.yang version 3.0.0, *OpenConfig Routing Policy Model*
openconfig-rsvp-sr-ext.yang version 0.1.0, *OpenConfig RSVP-TE and SR Extensions Model*
openconfig-system.yang version 0.10.1, *OpenConfig System Model*
openconfig-system-grpc.yang version 1.0.0, *OpenConfig System gRPC Model*
openconfig-system-logging.yang version 0.3.1, *OpenConfig System Logging Model*
openconfig-system-terminal.yang version 0.3.0, *OpenConfig System Terminal Model*
openconfig-telemetry.yang version 0.5.0, *OpenConfig Telemetry Model*
openconfig-terminal-device.yang version 1.9.0, *OpenConfig Terminal Optics Device Model*
openconfig-vlan.yang version 2.0.0, *OpenConfig VLAN Model*

Customer document and product support



Customer documentation

Customer documentation welcome page

Technical support

Product support portal

Documentation feedback

Customer documentation feedback