DHCP Management

In This Chapter

This chapter provides information about using DHCP, including theory, supported features and configuration process overview.

Topics in this chapter include:

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 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 DHC 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 illustrated in Figure 14.

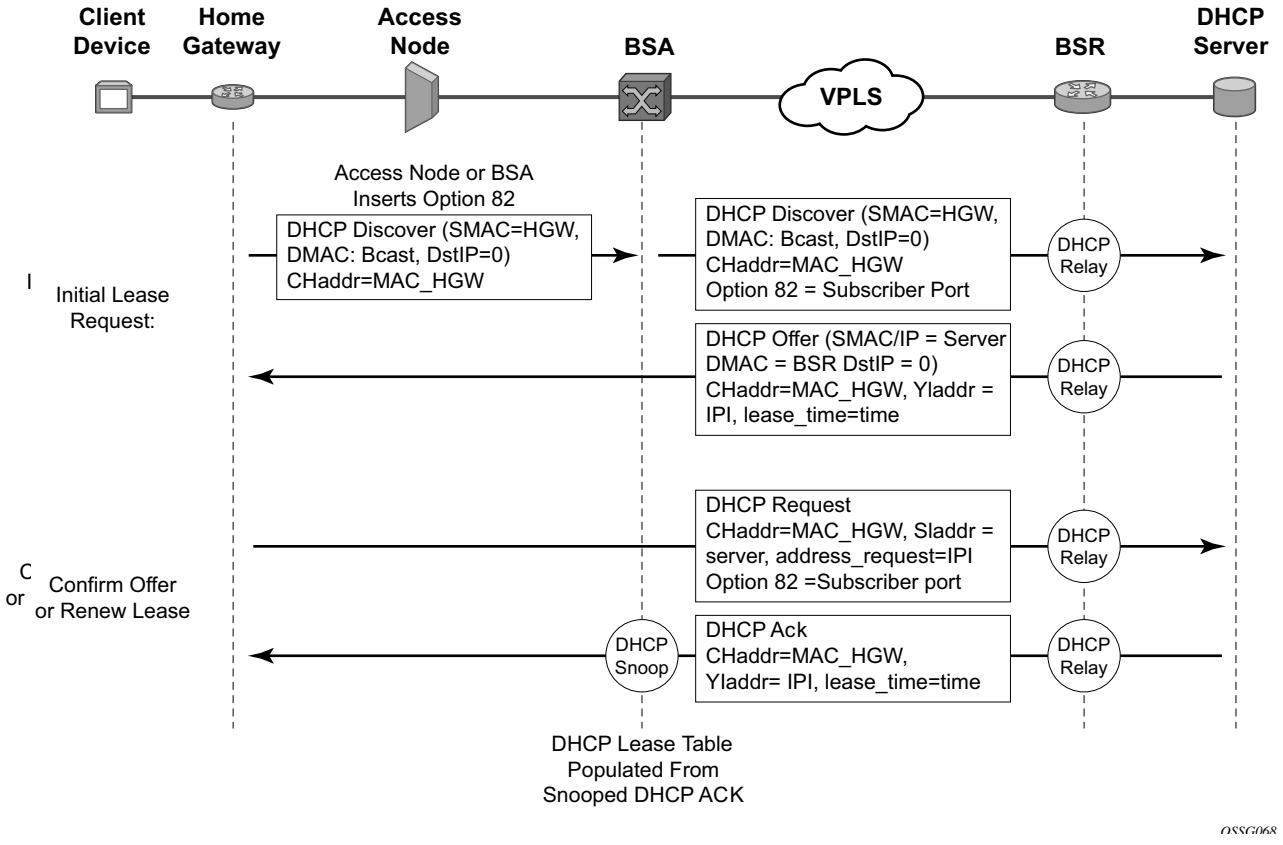


Figure 14: IP Address Assignment with DHCP

1. During boot-up, 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 — MAC of client device
   - → Client hardware address — MAC of 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 etc. 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 will relay the discover message as a unicast packet towards the configured DHCP server. DHCP relay is configured to insert the giaddr in order to indicate to the DHCP server in which subnet an address should be allocated.

2. The DHCP server will lookup 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 will be assigned and a DHCP offer message returned. The BSA or BSR will relay this back to the client device.
3. It is possible that the discover reached more than one DHCP server, and thus that more than one offer was returned. The client selects one of the offered IP addresses and confirms 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 will also contain the Lease Time of the IP address.

DHCP Features

DHCP Relay

Since DHCP requests are broadcast packets that normally will 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 (gateway IP address) 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 practical only in a Layer 3 environment, and thus is only supported in IES and VPRN services. On VPRN interfaces, however they will only forward to DHCP servers that also participate in that VPRN.

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, the subscriber interface's group interface DHCP relay is stateful.

DHCP Relay Enhancements

GRT-leaking can be used to relay DHCPv4 and DHCPv6 messages between a VRPN and the Global Routing Table (GRT) for network deployments where DHCP clients and server are in separate routing instances of which one is the Base routing instance.

In network deployments where DHCPv4 client subnets cannot be leaked in the DHCPv4 server routing instance, unicast renewals messages cannot be routed in the DHCPv4 server routing instance as illustrated in Figure 15:

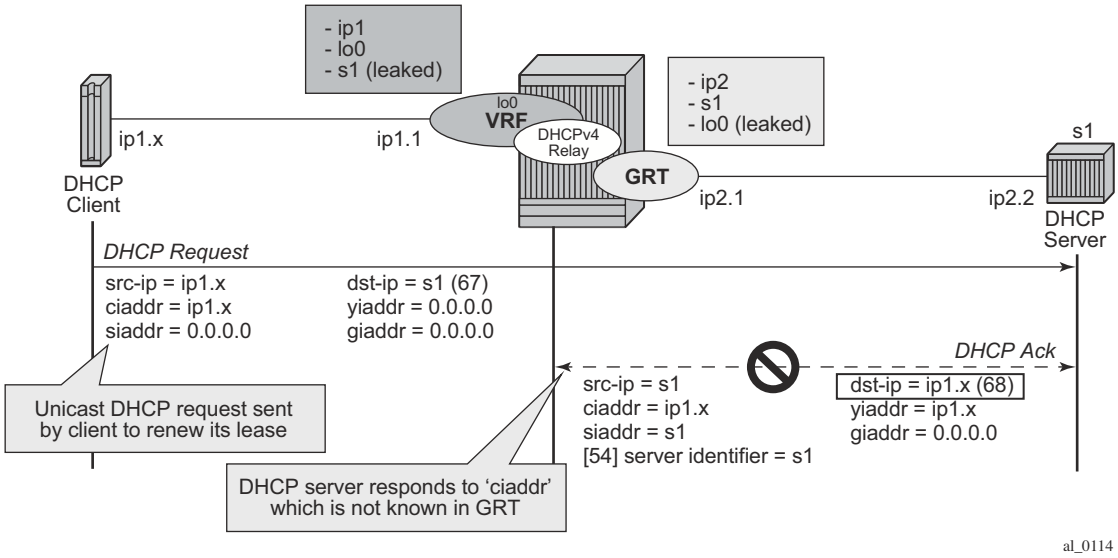


Figure 15: DHCPv4 Server Routing Instance

With the **relay-unicast-msg** command in the DHCPv4 relay on a regular interface or group-interface, it is possible to configure the gi-address of a DHCPv4 relayed message to any local address that is configured in the same routing instance. Unicast renewals are, in this case, relayed to the DHCPv4 server. In the upstream direction, update the source IP address and add the gateway IP address (gi-address) field before sending the message to the intended DHCP server (the message is not broadcasted to all configured DHCP servers). In the downstream direction, remove the gi-address and update the destination IP address to the value of the yiaddr (your IP address) field. By default, unicast DHCPv4 release messages are forwarded transparently. The optional **release-update-src-ip** flag, updates the source IP address with the value used for relayed DHCPv4 messages.

For retail subscriber interfaces, the **relay-unicast-msg** must be configured at the subscriber-interface dhcp CLI context as shown in the Example 1.

The **relay-unicast-msg** function is not supported in combination with a double DHCPv4 relay (L3 DHCPv4 relay in front of a 7750 DHCPv4 relay with "relay-unicast-msg" enabled).

Example 1

```
config>service>vprn>

            interface "lo0" create
                address 192.168.0.1/32
                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
```

```
                relay-unicast-msg release-update-src-ip
                gi-address 192.168.0.1 src-ip-addr
                no shutdown
            exit
        exit
    exit
```

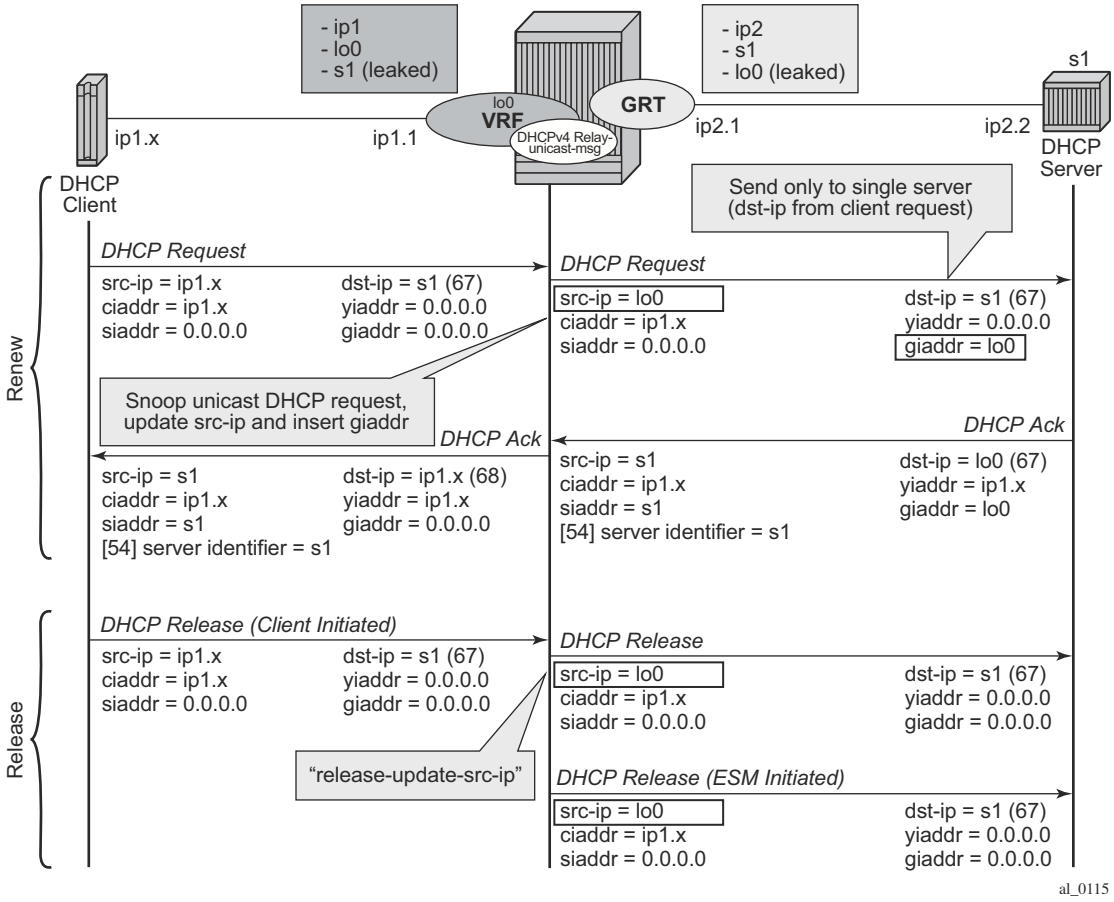


Figure 16: relay-unicast-msg Command in the DHCPv4 Relay

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 by the Alcatel-Lucent 7750 SR and can be used separately or together.

Inserting Option 82 information is supported independently of DHCP relay. However in a VPLS (when the 7750 SR is not configured for DHCP Relay), 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 by the 7750 SR, it can take following values:

- *sap-id:* The SAP index (only under a IES or VPRN service)
- *ifindex:* The index of the IP interface (only under a 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.

Note that 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.

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 the 7750 SR. On egress (towards 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 any additional 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 relay request will be forwarded without the Option 82 information. This packet size limitation exists to ensure that there will be 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 towards 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.

Trusted and Untrusted

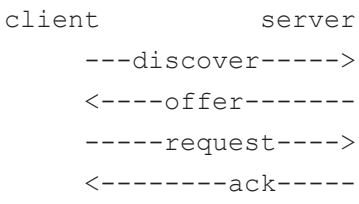
There is a case where the relay agent could receive a request where the downstream node added Option 82 information without also adding a giaddr (giaddr of 0). In this case the default behavior is for the router to drop the DHCP request. This behavior is in line with the RFC.

The 7750 SR supports a command `trusted`, which allows the router to forward the DHCP request even if it receives one with a giaddr of 0 and Option 82 information attached. This could occur with older access equipment. In this case the relay agent would modify the request's giaddr to be equal to the ingress interface. This only makes sense when the action in the information option is keep, and the service is IES or VPRN. In the case where the Option 82 information gets replaced by the relay agent, either through explicit configuration or the VPLS DHCP Relay case, the original Option 82 information is lost, and the reason for enabling the trusted option is lost.

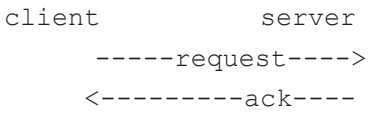
DHCP Snooping

This section discusses the Alcatel-Lucent 7750 SR acting as a Broadband Subscriber Aggregator (BSA) with Layer 2 aggregation towards a Broadband Subscriber Router (BSR).

A typical initial DHCP scenario is:



But, when the client already knows its IP address, it can skip the discover:



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 on page 341);
   - → to forward DHCP requests to a RADIUS server first, and not send them to the DHCP server unless the RADIUS server confirms positive identification. See section Anti-Spoofing Filters on page 650.

   For these applications, DHCP snooping must be enabled on the SAP towards 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 on page 346
   - → to perform Enhanced Subscriber Management, see section Triple Play Enhanced Subscriber Management on page 827

   For these applications, DHCP snooping must be enabled on both the SAP or SDP towards the network and the SAP towards 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 thus 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, etc.

DHCP Lease State Table

The DHCP lease state table has a central role in the BSA operation (Figure 17). For each SAP on each service it maintains the identities of the hosts that are allowed network access.

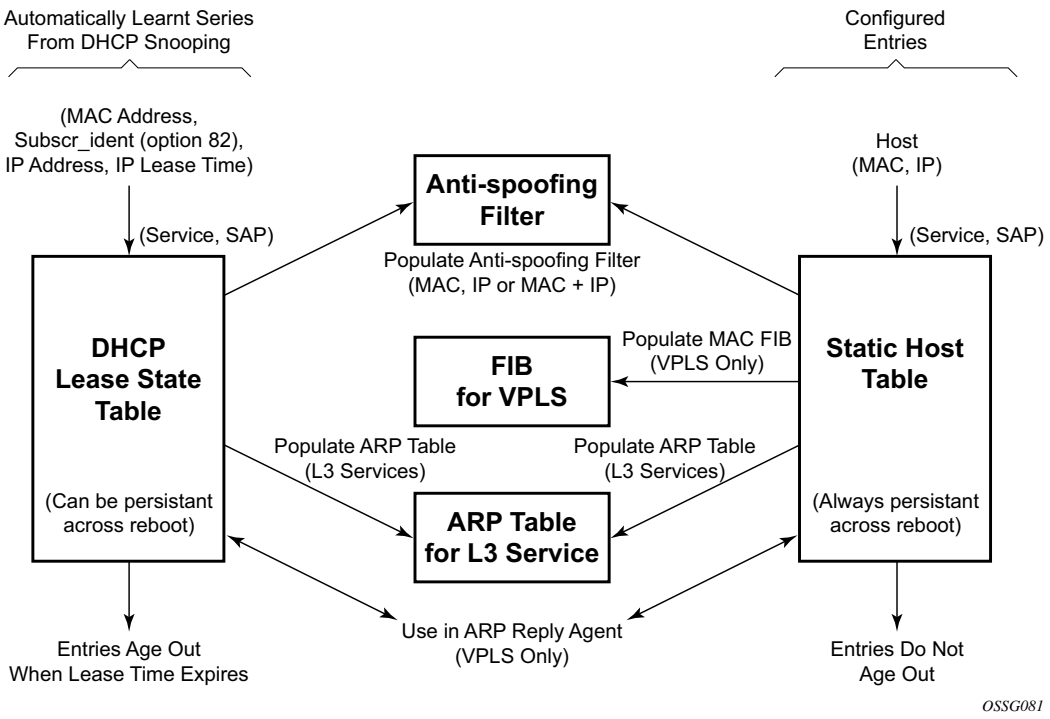


Figure 17: DHCP Lease State Table

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 above.

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 and/or SDP where DHCP messages requiring snooping ingress the VPLS instance. For IES and VPRN IP interfaces, 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 interfaces terminated on a SAP or VPRN IP interfaces terminated on a SAP.
- To populate the system's ARP cache using the **arp-populate** feature — **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 will automatically be populated into the VPLS forwarding database associated with the SAP on which the host is learned. The dynamic host MAC address will override any static MAC entries using the same MAC and prevent learning of the MAC on another interface. 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 Routed Central Office (CO) on page 1028.
- To support enhanced Subscriber management, see section RADIUS Authentication of Subscriber Sessions on page 833.

Note that 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.

DHCP and Layer 3 Aggregation

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 that 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 l2-header in the **lease-populate** configuration command at the DHCP level. Because SR-Series routers do not have the ability to verify the DHCP information (both the **src-ip** and **src-mac** of the packet will be those of the previous relay point) anti-spoofing must be performed at the access node prior to the SR-Series routers. This mode provides compatibility with MAC concentrator devices, and cable modem termination system (CMTS) and WiiMAX Access Controller (WAC).

A configuration example of a cable/wireless network together with subscriber management is shown in Figure 18. 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 will be 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 will be 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 will then use the configured MAC address when instantiating the DHCP lease states.

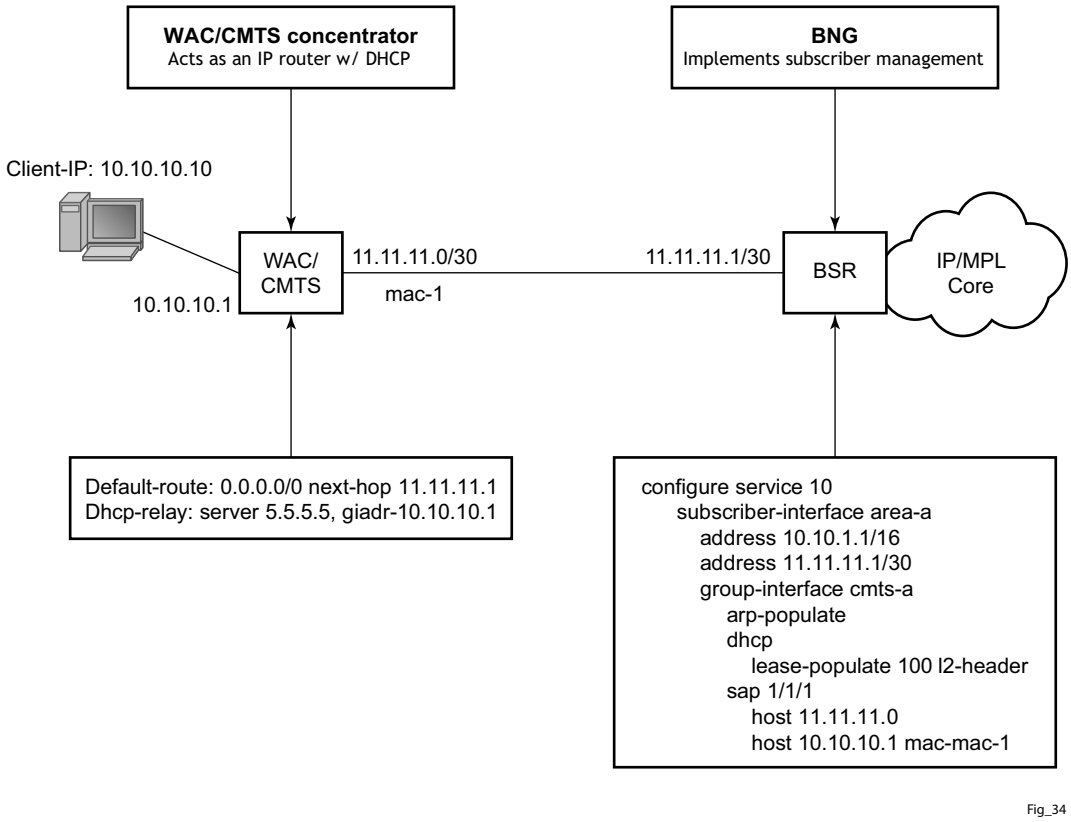


Figure 18: CMTS/WAC Network Configuration Example

Local DHCP Servers

Overview

The local-dhcp-server functions only if there is a relay (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.

A local DHCP server functions only if there is a relay agent (gateway) in front of it. 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 on one of the following valid addresses:

   → A fixed IP address — The IP address should not overlap with the address ranges configured in the local dhcp server.

   → A 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 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 once 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 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 once 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 will 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 will then be 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 certain combination of MAC and circuit-id, only one IP-address can be returned. If you ask 1 more, you get same address again. The same address will be returned if a re-request is made.

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, 7750 SR-12, and 7710 SR-Series models. Note that the 7450 ESS-Series 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 will receive addresses through a relay point in the customer's network.
- 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 will provide 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.

DHCPv6

In the stateful auto-configuration model, hosts obtain interface addresses and/or 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.

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.

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 will be assigned. For example, the delegating router can get a /48 prefix via DHCPv6 and assign /64 prefixes to multiple requesting routers. Prefix delegation is supported for the delegating router (not the requesting router).

Neighbor Resolution via DHCPv6 Relay

This is similar to ARP populate for IPv4. When it is turned on, an SR router 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 doesn’t 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.

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.

Local Proxy Neighbor Discovery

Local proxy neighbor discovery is similar to 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 subnet(s) unless disallowed by policy.
- Forwards traffic between hosts in the subnet(s) of the interface.
- Drops traffic between hosts if the link-layer address information for the IPv6 destination has not been learned.

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.

IPv6 Link-Address Based Pool Selection

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

IPv6 Address/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 of time, after the lease time has expired. A large configured timeout value allows for address and prefix “stickiness”. When a subscriber requests a lease via 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.

IPv4/v6 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 towards 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 circuit-id from DHCPv4 for authentication of DHCPv6 (or SLAAC) allows the 7750 SR to work around lack of support for LDRA on Access-nodes.

Host Connectivity Checks for IPv6

This feature provides support to perform SHCV checks on the global unicast address (assigned via SLAAC or DHCPv6 IA_NA) and link-local address of a L3 RG or a bridged host. SHCV uses IPv6 NS and NA messages. The configuration is similar to IPv4 support in SHCV. The **host-connectivity-check** command is extended to be configured for IPv6 or both IPv4 and IPv6.

DHCP Relay Enhancements

GRT-leaking can be used to relay DHCPv4 and DHCPv6 messages between a VRPN and the Global Routing Table (GRT) for network deployments where DHCP clients and server are in separate routing instances of which one is the Base routing instance.

In network deployments where DHCPv4 client subnets cannot be leaked in the DHCPv4 server routing instance, unicast renewals messages cannot be routed in the DHCPv4 server routing instance:

With the **relay-unicast-msg** command in the DHCPv4 relay on a regular interface or group-interface, it is possible to configure the gi-address of a DHCPv4 relayed message to any local address that is configured in the same routing instance. Unicast renewals are in this case relayed to the DHCPv4 server. In the upstream direction: update the source IP address and add the gateway IP address (gi-address) field before sending the message to the intended DHCP server (the message is not broadcasted to all configured DHCP servers). In the downstream direction: remove the gi-address and update the destination IP address to the value of the yiaddr (your IP address) field. By default, unicast DHCPv4 release messages are forwarded transparently. The optional **release-update-src-ip** flag, updates the source IP address with the value used for relayed DHCPv4 messages.

For retail subscriber interfaces, the “relay-unicast-msg” must be configured at the subscriber-interface dhcp CLI context.

```
The relay-unicast-msg function is not supported in combination with a double DHCPv4 relay
(L3 DHCPv4 relay in front of a 7750 DHCPv4 relay with relay-unicast-msg enabled).
config>service>vprn>
interface "lo0" create
                address 192.168.0.1/32
                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
                        relay-unicast-msg release-update-src-ip
                        gi-address 192.168.0.1 src-ip-addr
                        no shutdown
                     exit
                exit
            exit
```

Proxy DHCP Server

This section describes the implementation of a proxy DHCP server capability provide a standards-based DHCP server which will front end to downstream DHCP client and DHCP relay enabled devices and interface with RADIUS to authenticate the IP host and subscriber and will obtain 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.

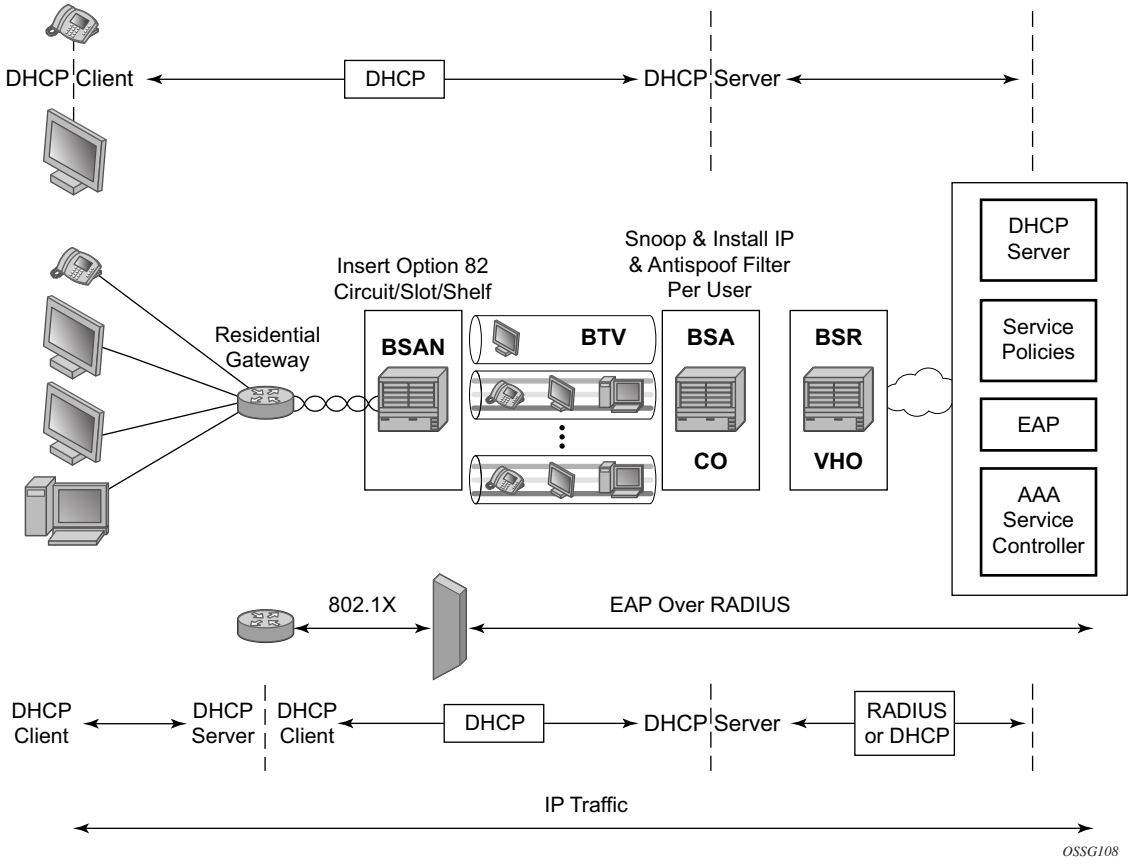


Figure 19: Typical DHCP Deployment Scenarios

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 stand alone 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 the means by which 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. The majority of 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 will be 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
- Ability to override the allocate IP address lease time, for improved IP address management.

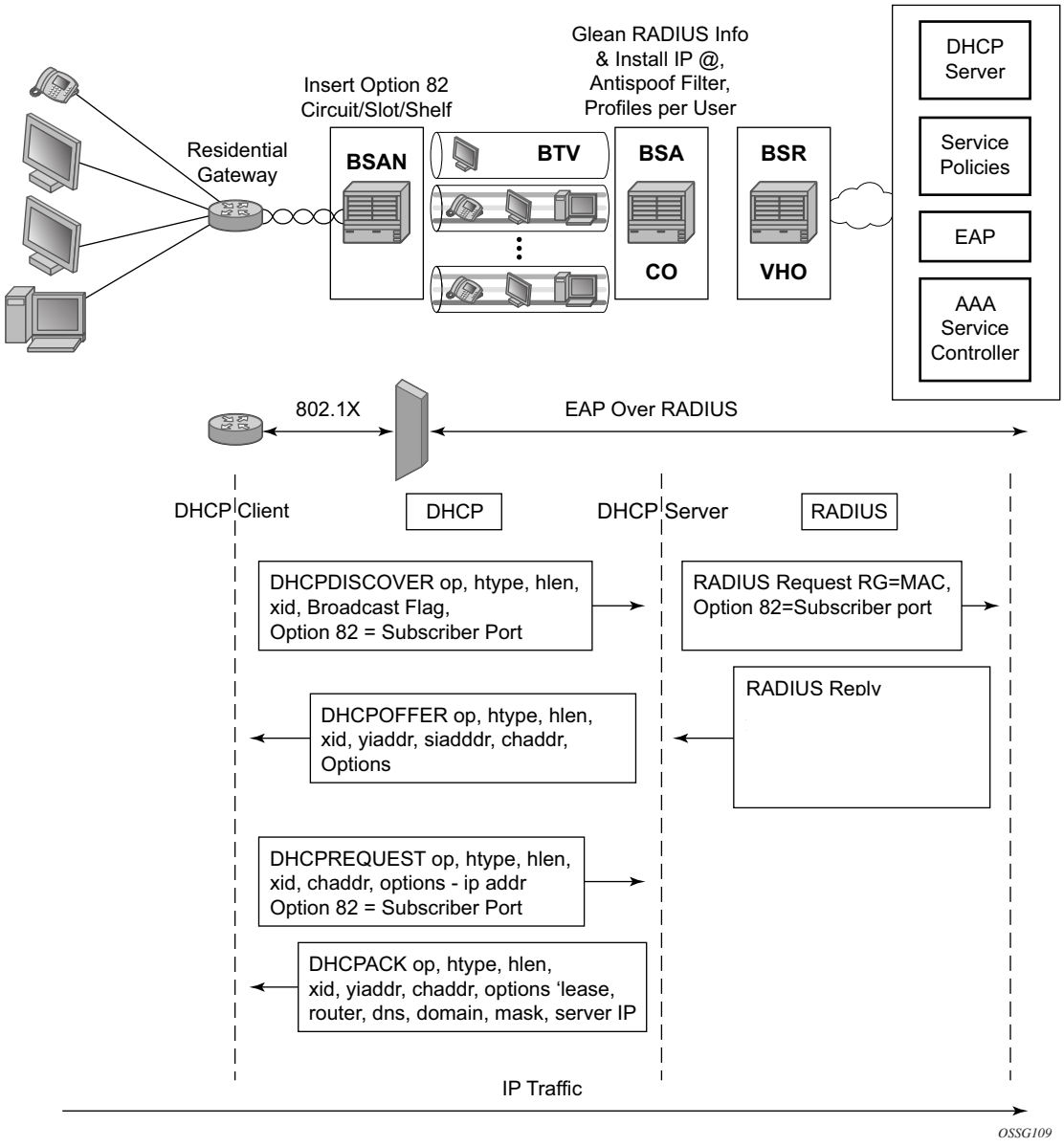


Figure 20: Example of Triple Play Aggregation Network With DHCP to RADIUS Authentication

Figure 20 displays a typical DHCP initial boot-up sequence with the addition of RADIUS authentication. The proxy DHCP server will interface with downstream DHCP client devices and then authenticate upstream using RADIUS to a providers subscriber management system.

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 given DHCP host — 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.

In order 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 TiMetra VSAs is available on a file included on the compact flash shipped with the image.

Local DHCP Servers

Terminology

- **Local 7x50 DHCP Server** – DHCP server instantiated on the local 7x50 node.
- **Remote 7x50 DHCP Server** – DHCP server instantiated on the remote 7x50 node (external to the local 7x50 node).
- **3rd party DHCP server** – DHCP server external to any 7x50 node and implemented outside of 7x50.
- **Intercommunication link** – the logical link between dual-homed 7x50 DHCP servers used for synchronizing DHCP lease states. Multi-chassis Synchronization (MCS) protocol runs over this link. Once 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/prefix** – refers to the local failover mode in which the IP address-range/prefix is configured in dual-homed DHCP environment. The local keyword does not refer to the locality (local vs remote) of the server on which the IP address-range/prefix in configured, but rather refers to the ownership of the IP address-range/prefix. The DHCP server on which the local IP address-range/prefix is configured, owns this IP address-range/prefix and consequently is allowed to delegate the IP addresses/prefixes from it at any time, regardless of the state of the intercommunication link.
- **Remote IP address-range/prefix** – refers to the remote failover mode in which the IP address-range/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/prefix is configured, but rather refers to the ownership of the IP address-range/prefix. The DHCP server on which the remote IP address-range/prefix is configured, but does not own this IP address-range/prefix during normal operation and consequently is NOT allowed to delegate the IP addresses/prefixes from it. Only when the intercommunication link between the two nodes transition into particular (failed) state, the DHCP server is allowed to start delegating new IP addresses from the remote IP address-range/prefix.
- **IP address-range/prefix ownership** – 7x50 DHCP server can delegate new leases from an IP address-range/prefix that it owns. For example, an IP address-range/prefix designated as remote is not owned by the DHCP server on which it is configured unless certain conditions are met. Those conditions are governed by the state of the intercommunication link.
- **IP address-range/prefix takeover** – a 7x50 DHCP server that does not own an IP address-range/prefix can take over the ownership of this IP address-range/prefix under certain conditions. Once the ownership is taken, the new IP addresses can start being delegated from this IP address-range/prefix. Only the remote IP address-range/prefix can be taken over. Note that the takeover of an IP address-range/prefix has only local significance – in other words, the ownership is not taken away from some other 7x50

DHCP server that has the same IP address-range/prefix designated as local. It only means that IP address-range/prefix that is configured as remote is available to takeover for new IP address delegation.

- **Local PPPoX Address Pools** – this term refers to the method of accessing an IPv4/v6 address pool in 7x50 DHCP4/6 server. For PPPoX clients, the IPv4/v6 addresses are allocated from those pools without the need for an intermediate DHCP relay-agent (7x50 internal DHCP relay-agent). Although those pools are part of the local DHCP server in 7x50, 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 via DHCP messaging
  - → PPPoE via internal API calls
- **Local PPPoX Pool Management** – IPv4 address allocation/management for PPPoX clients independent of DHCP process (DHCP lease state). An IPv4 address allocated via local PPPoX Pool Management is tied to the PPPoX session. It is without the need for an 7x50 internal DHCP relay-agent.

Overview

7x50 DHCP server multi-homing will ensure continuity of IP address/prefix assignment/renewal process in case that an entire 7x50 DHCP server fails or in case of a failure of the active link that connects clients to one of the 7x50 DHCP servers in the access part of the network. DHCP server multi-homing is an integral part of the overall subscriber management multi-chassis protection scheme.

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

Although the subscriber synchronization process and the DHCP lease states synchronization process use the same synchronization infrastructure within 7x50 (Multi Chassis Redundancy protocol), they are in essence 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/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 since the access-link failure coincides with the intercommunication link failure and vice versa. However, link-only failures become critical when it comes to their interpretation by the protection mechanisms (SRRP/MC-LAG/DHCP server multi-homing). Regardless of nature of the failure, overall DHCP server multi-chassis protection scheme must be devised in such fashion that the two 7x50 DHCP servers never allocate the same IP address/prefix to two different clients. Otherwise IP address/prefix duplication will ensue. Unique IP address/prefix allocation is achieved by making only one 7x50 DHCP server responsible for IP address/prefix delegation out of the shared IP address-range/prefix.

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

- Shared IP address-range/prefix is designated as local on one 7x50 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/prefix as well as the one with the remote IP address-range/prefix.

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

  The DHCP server configured with the remote IP address-range/prefix will start delegating new lease from it only when it declares that the redundant peer with the local IP address-range/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 chance the operator ample time to fix the problem.

For example, the DHCP server will take over the remote IP address-range/prefix once 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/prefix will not be possible until the preset timers expire. This is needed and justified in case that 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/prefix becomes truly unavailable, those additional restoration times will cause interruption in service since the new IP addresses from the remote IP address-range/prefix will not be immediately available for delegation.

Note that only new IP address delegation from the remote IP address-range/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/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/prefix space so that each node has an IP address-range/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 will be 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 will always be available to assign new leases without any overlap.

→ In case that 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 via configuration. In this case, a safe operation is warranted only if the operator is certain that the intercommunication link failure will be caused by the nodal failure, and not the physical link failure between the two nodes.

- Shared IP address-range/prefix is designated as access-driven on both 7x50 DHCP servers.

  In this scenario, the shared IP address-range/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 given time must be responsible for IP address assignment from this shared IP address-range/prefix.

This will be ensured by the access protection mechanism (SRRP/MC-LAG) that will provide 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/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 A sends DHCP Discover to the DHCP server A

→ 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 since 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.

→ Once the DHCP syn message is received from the peer, the DHCP server will realize that the IP lease already exist. In this case, the newer IP lease will override the older.

→ The result is that clients A and B use the same IP address and consequently the forwarding of the traffic will be impaired.

In access-driven model, the ESM subscriber host must be collocated 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 7x50 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 once 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/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.

In case that multiple IP address-ranges/prefixes are deployed, bypass of the timers is not needed since the local IP address-ranges/prefixes are available on both nodes.

- Access-driven model allows a single IP address-range/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).

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:

DHCP Discover ; DHCP client — DHCP server

DHCP Offer ; DHCP server — DHCP client

DHCP Request ; DHCP client — DHCP server

DHCP Sync Message ; DHCP server — via MCS to the peering DHCP server

DHCP Ack ; DHCP server— DHCP client

DHCP server failover mechanism in the local-remote IP address-range/prefix model relies on the detection of the failure of the link over which DHCP states are synchronized (via MCS protocol). This link is normally disjoint from the access links towards the clients. MCS protocol normally runs over a direct link between the two redundant nodes (1) or over backbone links (2) in case that the direct link is not present.

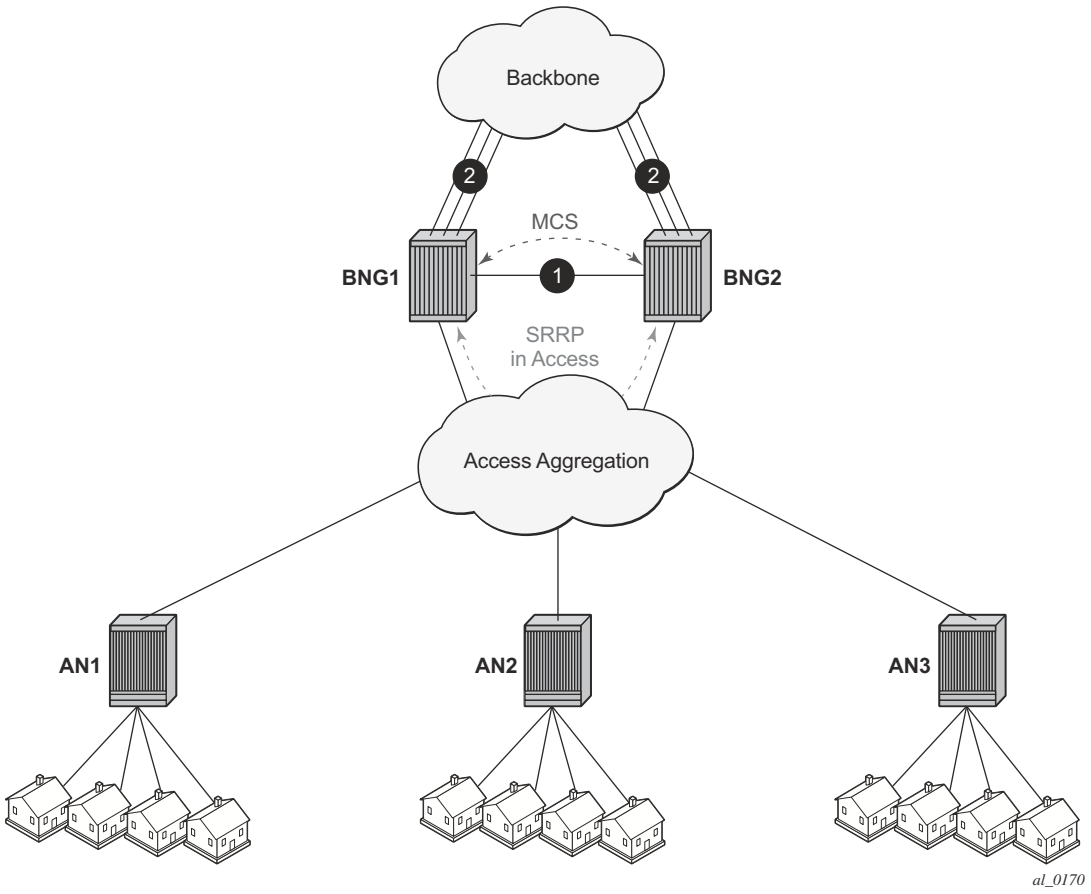


Figure 21: Redundancy Model

In the access-driven IP address-range/prefix model, the DHCP server address-ranges/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.

Intercommunication Link Failure Detection

7x50 DHCP Server is a client of Multi-chassis Synchronization (MCS) application with 7x50. Once MCS transitions into out-of-sync state, the 7750 DHCP server redundancy assumes that there is a failure in the network. In essence, 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 0.5sec, MCS will send its own keepalive to the peer. If a reply is NOT received within 3sec,

MCS will declare its operational state as DOWN and the DB Sync state as out-of-sync. MCS will consequently notify its clients (DHCP Server being one of them) of this condition.

In essence, it can take up to 3 seconds before the DHCP client realizes that the interclasses 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 utilize 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 will lead to certain failover state transitions on the DHCP server. The DHCP lease handling in the local-remote model will depend on the particular failover state on the DHCP server and the duration of each failover state is determined by preconfigured timers.

DHCP Server Failover States

DHCP Server when paired in redundant fashion can transition through several states:

- TRANSITION
- SHUTDOWN
- INIT
- STARTUP
- NORMAL – In this state, the 7x50 DHCP Server is serving all IP leases from the local and access-driven IP addresses-ranges/prefixes (assigning new leases and extending existing ones). Remote IP addresses-ranges/prefixes are not served (new or existing ones).
- COMMUNICATION INTERUPTED – IP addresses/prefixes under the local and access-driven IP address-range/prefix are served in the same fashion as in the NORMAL state. The IP addresses/prefixes under the remote IP address-range/prefix will be renewed. However, the new address/prefix from the remote address-range/prefix will not be 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 – Once the DHCP Server reaches this state, the remote IP address-

range/prefix is taken over and after an additional time period of MCLT (Maximum Client Lead Time), the new IP addresses/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 in case that all of 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/prefix on the both DHCP servers.

Lease Time Synchronization

7x50 DHCP server state synchronization is different from the lease state synchronization of the subscriber host itself.

- Each subscriber host lease state is synchronized via sub-mgmt client application. The host lease state can be seen via 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).
- On the other hand, the DHCP server lease states can be observed via the following CLI command:

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

The concern is with the latter, the 7x50 DHCP server lease state synchronization. To ensure uninterrupted IP lease renewal process after a failure in the network, the DHCP server lease time that is synchronized between the 7x50 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 22:

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

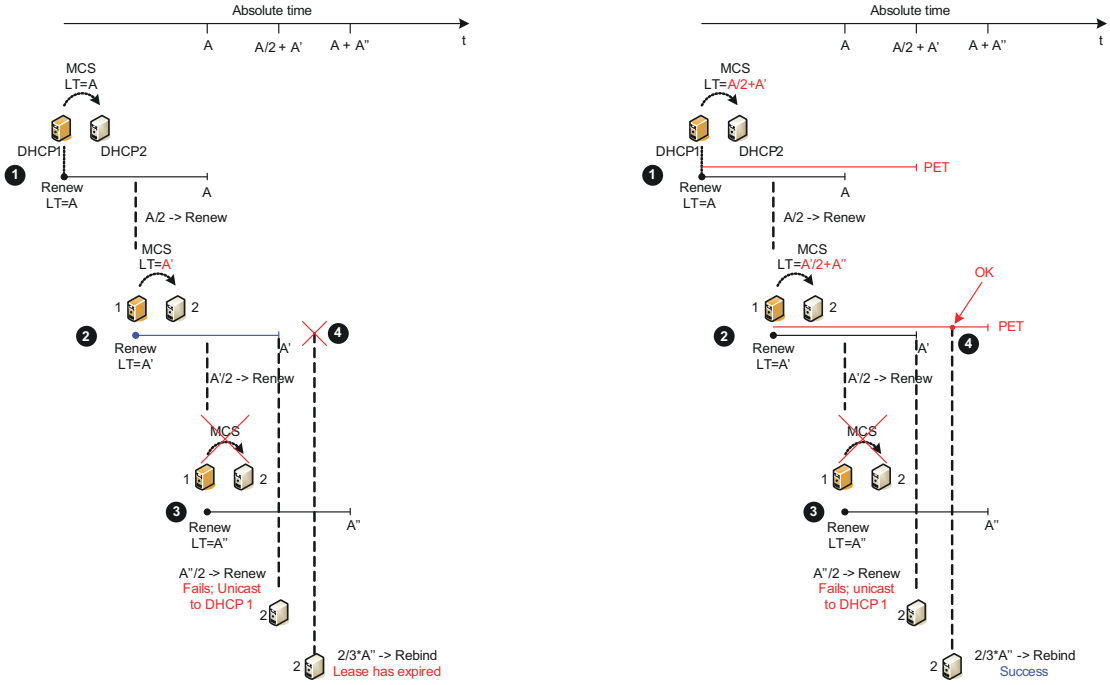


Figure 22: Potential Expiration Time

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 due to 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/prefix, the lease in the secondary server will be expired (4). As a result, the IP address/prefix will not be renewed.

To remedy this situation, the primary server must synchronize the lease time as the current RENEW time + the next lease-time. In this fashion, as depicted in point 3, when the REBIND reaches server 2, the lease in its DB will still be active (4) and the server 2 will be able to extend the lease for the client.

Maximum Client Lead Time (MCLT)

Maximum Client Lead Time (MCLT) is the maximum time that the 7x50 DHCP Server can extend the lease time to its clients beyond the lease time currently know by the 7x50 DHCP partner node. By default this time is relatively short (10min).

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

The local 7x50 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 7x50 DHCP server failure, the remote 7x50 DHCP server is not aware of the IP address/prefix that has been allocated on the local 7x50 DHCP server. This condition creates the possibility that the remote 7x50 DHCP server allocates the same address/prefix to another client. This would cause IP address/prefix duplication. MCLT is put in place to prevent this scenario.

MCLT based solution is shown in Figure 23.

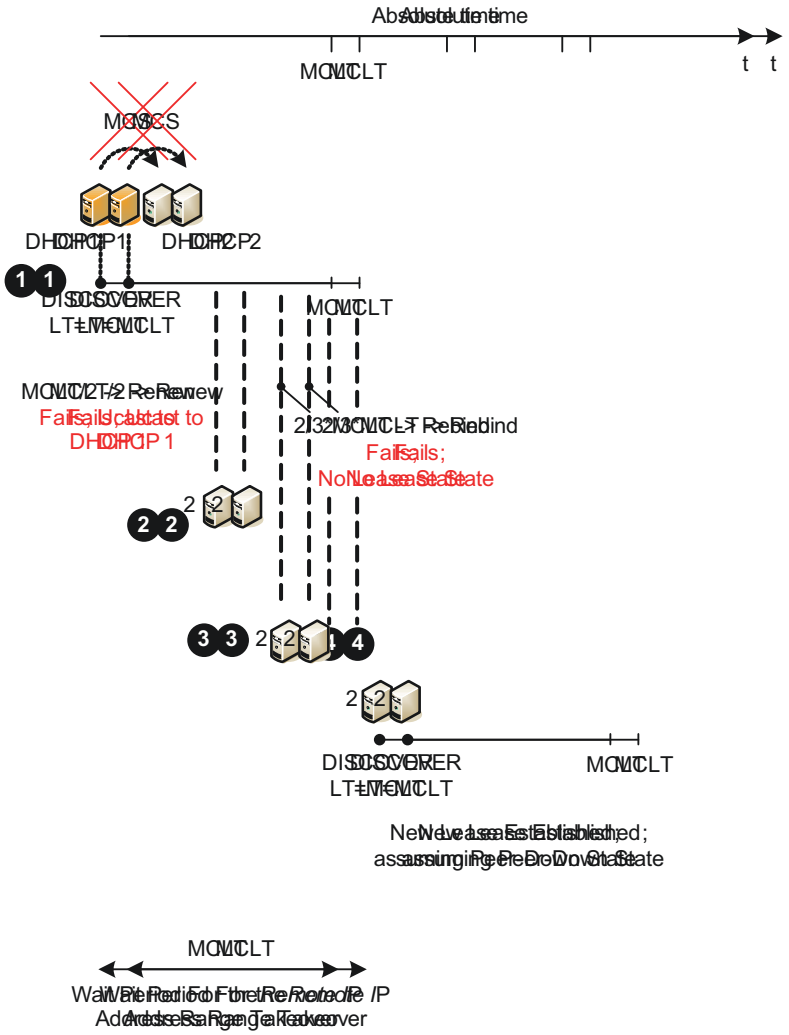


Figure 23: Address/Prefix Allocation without Synchronization

The sequence of events is the following:

1. 7x50 DHCP Server 1 is the local DHCP server (with the local address-range/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 7x50 DHCP server fails before it gets a chance to synchronize the lease state with the 7x50 DHCP Server 2 (remote 7x50 DHCP server with the remote address-range/prefix).
2. The remote 7x50 DHCP server transitions into the PARTNER-DOWN state (assuming that the partner-down timer is 0). In this state the remote 7x50 DHCP server can extend the lease time to the existing clients but it can NOT assign a new lease for a period of MCLT. In MCLT/2 a new RENEW request is sent directly to the local 7x50 DHCP server. This server is DOWN and therefore it cannot reply.
3. The client broadcasts a REBIND request that reaches the remote 7x50 DHCP server. The remote 7x50 DHCP server has no knowledge of the requested lease and therefore it does not reply.
4. The lease for the client will expire and the client will have to reinitiate the IP address/prefix assignment process.

Since the remote 7x50 DHCP server is not aware of the lease state that was assigned by the local 7x50 DHCP server, there is a chance that the remote 7x50 DHCP server assigns to the new client the same IP address/prefix already allocated by the local 7x50 DHCP server just before it crashed. This is why the remote 7x50 DHCP server needs to wait for the MCLT time to expire so that the IP addresses/prefixes allocated (but never synchronized) by the local 7x50 DHCP server can time out.

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

1. 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 7x50 DHCP server only renews the leases but does not delegate new ones (primary is DOWN).

     The local 7x50 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 7x50 DHCP server starts delegating new IP addresses/prefixes from the remote address-range/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.

2. The communication channel is down but the remote 7x50 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 7x50 DHCP server keeps renewing existing leases but it does not delegate the new leases. The chances that this will happen are low as the clients will keep sending RENEWS via unicast to the local 7x50 DHCP server which is still active. The non-synched leases in the remote 7x50 DHCP server will time out. The local 7x50 DHCP server will start trickling down the lease time to MCLT.

   The local 7x50 DHCP server will keep delegating the new leases, although with the MCLT lease time.

   → PARTNER-DOWN State
   The local 7x50 DHCP server will keep extending the existing leases but it will also start delegating new IP leases once the initial MCLT elapses.

   Note that both local and remote 7x50 DHCP server will delegate new leases in PARTNER-DOWN state (although the remote 7x50 DHCP server in PARTNER-DOWN state will have to wait an additional MCLT period before it can start delegating new leases).

Sharing IPv4 Address-Range or IPv6 Prefix Between Redundant 7x50 DHCP Servers in Access-Driven Mode

Access-driven DHCP server redundancy model will ensure uninterrupted IP address assignment service from a single IP address-range/prefix in case that 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 7x50 DHCP servers in redundant configuration. This single active path is ensured via a protection mechanism in dual-homed environment in the access side of the network. The supported protection mechanisms in the access in 7x50 are SRRP or MC-LAG.

In access-driven DHCP redundancy model, the DHCP relay in each 7x50 node must point only to the IP address of the local DHCP server. In other words, the DHCP messages received on one DHCP server should never be relayed to the other. Since the IP address-ranges/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 will allow 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 via SRRP awareness. The end 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 will be advertised to the network side only from the actively serving node (SRRP Master or active MC-LAG node). This will be performed indirectly via the corresponding subscriber-interface route that is aligned (by configuration) with the DHCP address range or prefix in access-driven mode.

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

Note that Multi-chassis redundancy relies on MCS to synchronize various client applications. (subscriber states, DHCP states, IGMP states, etc) between the two chassis. Therefore the links over which MCS peering session is established must be highly redundant. Failed MCS peering session will render 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:

Table 8: DHCP Failover Scenarios

Access Related Failure	Inter-chassis Link State	Number of Failures	Action
None	None	None	Only the MASTER SRRP BNG will grant IP leases. The subnet tied to the SRRP instance is advertised to the network side on the MASTER SRRP node.
None	COMM-INT	Possibly multiple	Only the MASTER SRRP BNG will grant 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 + 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 particular scenario.

Table 8: DHCP Failover Scenarios (Continued)

Access Related Failure	Inter-chassis Link State	Number of Failures	Action
Access Link Towards the Master SRRP	NORMAL	Single	SRRP will switch over. The new IP lease grants will continue from new SRRP Master using the same IP address range. IP leases will be synched OK. Subnets will be advertised from new Master via SRRP awareness.
Access Link Towards the Master SRRP	COMM-INT	Multiple	SRRP will switch over. The new leases can be handed from the DHCP server on the newly SRRP Master node. However, this DHCP server may not have its lease state table up to date since the inter-chassis communication link is non-operational. Consequently, new SRRP Master node may start handing out leases that are already allocated on the node with the failed link. In this case, IP address duplication would ensue. This is why it of utmost importance that the intercommunication chassis link is well protected and the only event that causes it to 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 Master SRRP	PARTNER-DOWN	Multiple	Same as above.
Access Link Towards the Standby SRRP	NORMAL	Single	No effect on the operation since everything is active on the Master SRRP node anyway.
Access Link Towards the Standby SRRP	COMM-INT	Multiple	Intercommunication link is broken. The Master SRRP node will continue handing out the new leases and renewing the old ones. However, they will not be synchronized to the peering node.
Access Link Towards the Standby SRRP	PARTNER-DOWN	Multiple	Same as above.

Table 8: DHCP Failover Scenarios (Continued)

Access Related Failure	Inter-chassis Link State	Number of Failures	Action
Entire Master Node	COMM-INT	Single	SRRP will switch over. However, lease duplication may occur on the newly Master since the intercommunication link is broken and this newly SRRP Master is not aware of DHCP leases that the peer (failed node) may have allocated to the clients while the intercommunication-link was broken.
Entire Master Node	PARTNER-DOWN	Single	Same as above.
Entire Standby Node	COMM-INT	Single	Operation continues OK but multi-chassis redundancy is lost.
Entire Standby Node	PARTNER-DOWN	Single	Same as above.

A possible deployment scenario is shown in Figure 24.

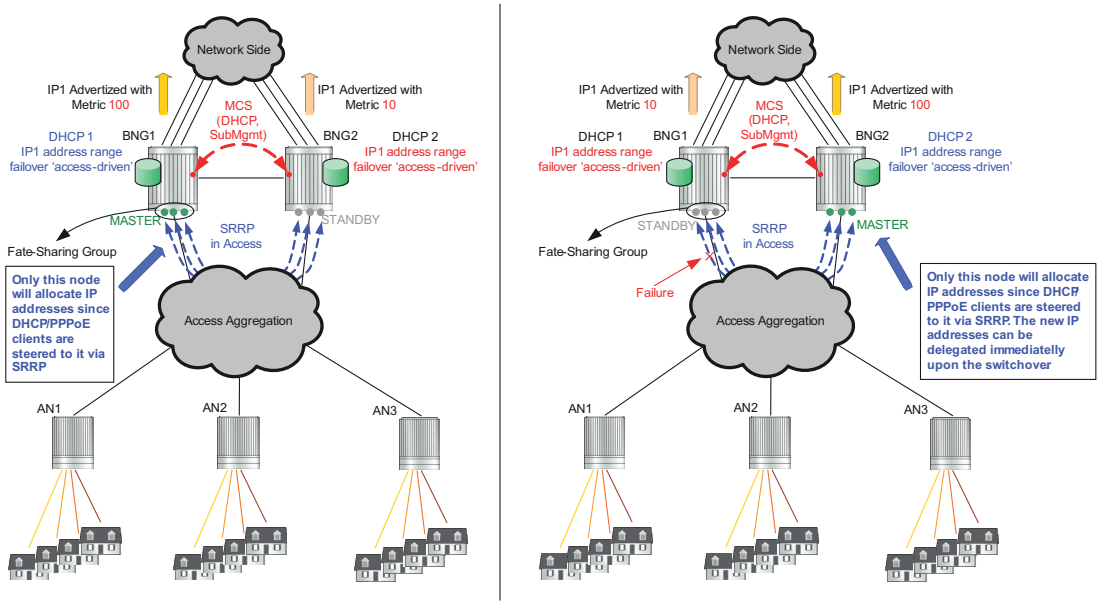


Figure 24: Failover Scenario with SRRP and DHCP in Access-Driven Mode

Fast-Switchover of IP Address/Prefix Delegation For ‘Remote’ IP Address/ Prefix Range

Some deployments require that the ‘remote’ IP address/prefix range starts delegating new IP addresses/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/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:

- partner-down-delay must be set to 0. This will cause the intercommunication link to bypass the COMM-INT state upon the failure and transition straight into the PARTNER-DOWN state. ‘Remote’ IP address-range/prefix can be taken over only in PARTNER-DOWN state, once the MCLT expires.
- ignore-mclt-on-takeover flag must be enabled. With this flag enabled, ‘Remote’ IP address/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. Note that by setting this flag, the lease times of the existing DHCP clients, while the intercommunication link is in the PARTNER-DOWN state, will still be reduced to the MCLT over time and all new lease times will be set to MCLT this behavior remain the same as originally intended for MCLT. this functionality must be exercised with caution. One needs to keep in mind 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 due to the failure of intercommunication link. These timers (partner-down-delay and MCLT) would ensure that during their duration, the new IP addresses/ prefixes are delegated only from one node – the one with local IP address-range/prefix. The drawback is of course that the new IP address delegation is delayed and thus service is impacted.

But if one could ensure that the intercommunication link is always available, then the DHCP nodes would stay in sync and the two timers would not be needed. This is why it is of utmost importance 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 since in this case only one DHCP node is available to provide new IP addresses/prefixes.

DHCP Server Synchronization and Local PPPoX Pools

Since there is no classical DHCP Lease state maintained for local PPPoX pools, the IP addresses will not be synchronized via DHCP Server. Instead they will be synchronized via PPPoX clients. Once the PPPoX subscriber is synchronized, the respective IP address lease will be updated in the respective local pool.

For example:

- a PPPoE client is created on 7x50 ‘A’
- the IP address is assigned from the local poll on 7x50 ‘A’
- the PPPoE client will be synchronized to the peering node 7x50 ‘B’
- once the client is synchronized in 7x50 ‘B’, the IP address assignment will be synchronized by our internal PPPoE process on 7x50 ‘B’ with the local pool.

One artifact of this behavior (IP address assignment in local DHCP pools is synchronized via PPPoX clients and not via DHCP server synchronization mechanism) is that during the node boot, the DHCP server must wait for the completion of PPPoX subscriber synchronization via MCS so that it learns which addresses/prefixes are already allocated on the peering node. Since the DHCP server can theoretically start assigning IP addresses before the PPPoX sync is completed, a duplicate address assignment my occur. For example an IP address lease can be granted via DHCP local pools while PPPoX sync is still in progress. Once 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 will be kept and the other will be removed from both systems. To prevent this scenario, a configurable timer can be set to an arbitrary value that will render 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.