CHAPTER 4

WAN Edge—MPLSoL2 Service

While Layer 3 VPN services are becoming increasing popular as a primary connection for the WAN, there are a much larger percentage of customers still using Layer 2 services such Frame-Relay (FR). A big factor in migrating to Layer 3 services is cost and bandwidth scalability and flexibility. There will be customers who may not want to migrate to Layer 3 services, but rather maintain their Layer 2 infrastructure (such as Financials) or are at least slow in moving towards it. For such customers, extending virtualization to the branches involves converting the branch routers into MPLS edge devices and enabling MPLS on the Layer 2 links. The WAN aggregation device is converted into a P router and connected directly to the MPLS network. Thus the branch routers (PE) are now part of the hub MPLS network.

The existing IGP can be used to distribute the PE and Router Reflector (RR) reachability information. The WAN aggregation router maintains LDP sessions with ever branch router to advertise label information for the PEs. The branch routers establish the MP-ibgp session with the core MPLS network RRs for VPN information. Since they are now part of the MPLS network, services such as MVPN can be extended to them as well.

To extend MPLS to the branches:

- Create loopback interfaces on the branch routers for MP-iBGP peering.
- Enable LDP on the Layer 2 links connecting the branches with the WAN aggregation hub.
- Ensure that the loopback addresses are advertised via IGP and that LDP labels are allocated for them.
- Configure the VRFs and place the user LAN or VLANs into appropriate VRF at each of the branches.
- Make the branch routers clients of the core RR for MP-BGP. This allows VPN information to be exchanged between the branch PEs and core PEs.

For network redundancy, the spoke could be dual homed with two PVCs to two aggregators. MPLS could be enabled on both the links and the spoke can load balance traffic destined to other PEs.

Note If the existing deployment uses IPSec encryption on the routers, then this model may present some challenges. Labeled packets cannot be encrypted by the routers.

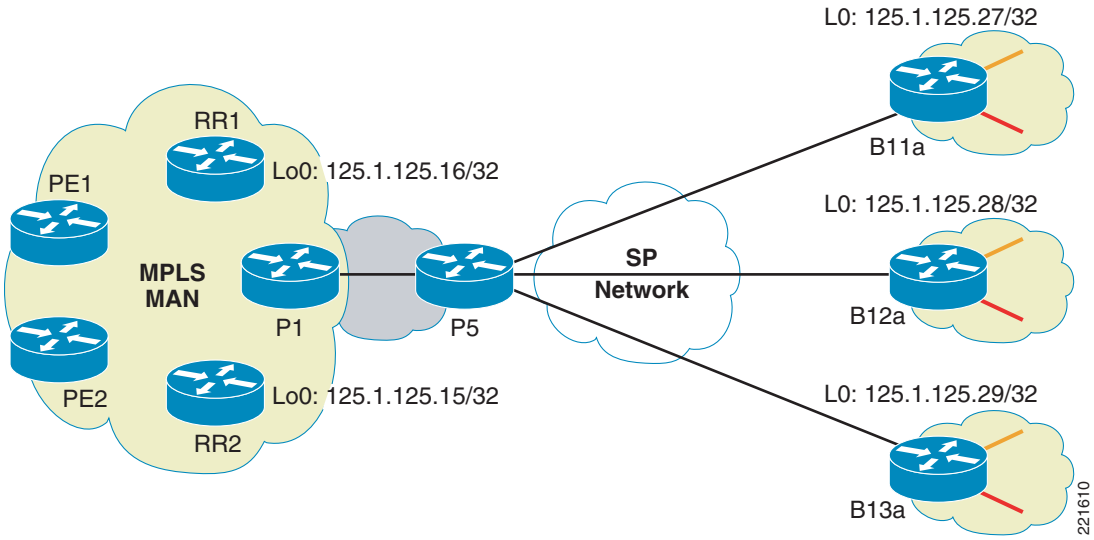
Platforms

The WAN aggregation hub could be any router that supports P functionality and meets the performance requirements, such as 12000, 7600, or 7200s. ISRs are typically recommended as spoke routers. The latest 12.4T images are recommended for the ISRs and 7200s used as branch routers. The image selection for GSR, 7600, and 7200 WAN aggregation routers need to be selected based on the feature, hardware requirement, and compatibilities.

Example:

As shown in Figure 4-1, sites B11, B12, and B13 have existing FR connections to WAN aggregation device (P5). The aggregation device is connected to the core P (P1). The three branch routers have loopbacks created that are advertised in the IGP—B11a (125.1.125.27/32), B12a (125.1.125.28/32), and B13a (125.1.125.29/32). The core has two RRs (125.1.125.15 and 16) to which each of the branch PEs is peering.

Figure 4-1 MPLSoL2 Deployment



P5:

```
mpls label protocol ldp
tag-switching tdp router-id Loopback0 force
!
interface Loopback0
 ip address 125.1.125.11 255.255.255.255
!
interface GigabitEthernet1/3
 description To P1 - intf G2/0/1
  ip address 125.1.100.102 255.255.255.252
  tag-switching ip
  mls qos trust dscp
!
interface Serial2/0/0
 mtu 1500
 no ip address
 encapsulation frame-relay
 dsu bandwidth 44210
 framing c-bit
 cablelength 10
 clock source internal
```

```
!
interface Serial2/0/0.1 point-to-point
 ip address 125.1.201.2 255.255.255.252
 ip pim sparse-mode
 tag-switching ip
 frame-relay interface-dlci 17
!
interface Serial2/0/0.2 point-to-point
 ip address 125.1.202.2 255.255.255.252
 ip pim sparse-mode
 tag-switching ip
 frame-relay interface-dlci 18
!
interface Serial2/0/0.3 point-to-point
 ip address 125.1.203.2 255.255.255.252
 ip pim sparse-mode
 tag-switching ip
 frame-relay interface-dlci 20
!
router ospf 10
 log-adjacency-changes
 network 125.1.201.0 0.0.0.3 area 0
 network 125.1.202.0 0.0.0.3 area 0
 network 125.1.203.0 0.0.0.3 area 0
 network 125.0.0.0 0.255.255.255 area 0
 maximum-paths 8
```

Note While the IGP configuration here shows all the spokes in Area 0 for simplicity, in practice any existing hierarchical design can be maintained as long as PE loopback addresses are never summarized and a label is allocated for each of the /32 addresses.

```
ip cef
mpls label protocol ldp
!
ip vrf red-data
 rd 10:1033
 route-target export 10:103
 route-target import 10:103
!
ip vrf red-voice
 rd 10:1043
 route-target export 10:104
 route-target import 10:104
!
interface Loopback0
 ip address 125.1.125.27 255.255.255.255
!
interface GigabitEthernet0/1.1
 encapsulation dot1Q 261
 ip vrf forwarding red-data
 ip address 125.1.20.1 255.255.255.0
!
interface GigabitEthernet0/1.2
 encapsulation dot1Q 262
 ip vrf forwarding red-voice
 ip address 125.1.20.1 255.255.255.0
!
interface Serial1/0/0
 no ip address
 encapsulation frame-relay
 load-interval 30
```

```
 clock rate 2000000
!
interface Serial1/0/0.1 point-to-point
 ip address 125.1.201.1 255.255.255.252
 mpls ip
 frame-relay interface-dlci 16
!
router ospf 10
 log-adjacency-changes
 network 125.1.125.27 0.0.0.0 area 0
 network 125.1.201.0 0.0.0.3 area 0
!
router bgp 1
 no bgp default ipv4-unicast
 bgp log-neighbor-changes
 neighbor 125.1.125.15 remote-as 1
 neighbor 125.1.125.15 update-source Loopback0
 neighbor 125.1.125.16 remote-as 1
 neighbor 125.1.125.16 update-source Loopback0
 !
 address-family vpnv4
 neighbor 125.1.125.15 activate
 neighbor 125.1.125.15 send-community extended
 neighbor 125.1.125.16 activate
 neighbor 125.1.125.16 send-community extended
 exit-address-family
 !
 address-family ipv4 vrf red-voice
 redistribute connected
 no synchronization
 exit-address-family
 !
 address-family ipv4 vrf red-data
 redistribute connected
 no synchronization
 exit-address-family
```

Multicast

As in a MPLS/Layer 3VPN network, mVPN is the technique to bring the multicast traffic of individual customer (or segmented user groups) across the core network MVRFs are configured for every VRF where multicast traffic is expected on the branch PEs. Default and Data MDTs are also configured if used within the core MPLS network.

Since the Layer 2 service is typically hub and spoke or partially meshed for larger branches, it is recommended to keep the multicast sources at or behind the hub as much as possible. In either case, the WAN aggregator at the hub would end up doing most of the multicast replication as it would be the last hop P for all the branch PEs that are receivers. Thus the multicast replication performance of the aggregator becomes critical to solution scaleability.

Use of Data MDTs with a very low threshold is highly recommended. This would limit the replication at the aggregator to only those branches that have sent a explicit join to the Data MDT. Keeping the threshold low ensures that the Data MDT is spawned for the more specific (S,G) as soon as possible to reduce the overhead at the aggregator.

Most deployments use either PIM SSM or PIM SM with RP in the MPLS core. PIM-SM tries to constrain data distribution so that a minimal number of routers in the network receive it. Packets are sent only if they are explicitly requested at the RP. By default, members of a multicast group receive data from senders to the group across a single data distribution tree rooted at the RP. PIM-SSM is similar to

PIM-SM with the additional ability to report interest in receiving packets from specific source addresses to an IP multicast address. It does not use RPs but uses source-based forwarding trees only. While PIM SSM provides the simplest implementation, it does create additional memory overhead since the number of mroutes now increases. This is not expected to be an issue in most enterprise deployments.

Example:

Continuing with our earlier example, we add MVPN to sites B11, B12, and B13. We have Data MDT setup with very low threshold (1kbps) to ensure that it gets initiated almost instantly for any stream. We will use PIM SSM for the Data MDTs.

B11a:

```
ip vrf red-data
 rd 10:1033
 route-target export 10:103
 route-target import 10:103
 mdt default 239.232.10.3
 mdt data 239.232.20.32 0.0.0.15 threshold 1
!
ip multicast-routing
ip multicast-routing vrf red-data
!
interface Loopback0
 ip address 125.1.125.27 255.255.255.255
 ip pim sparse-mode
!
interface GigabitEthernet0/1.1
 encapsulation dot1Q 261
 ip vrf forwarding red-data
 ip address 125.1.20.1 255.255.255.0
 ip pim sparse-mode
!
interface Serial1/0/0.1 point-to-point
 ip address 125.1.201.1 255.255.255.252
 ip pim sparse-mode
 mpls ip
!
ip pim ssm range 1
ip pim vrf red-data rp-address 3.3.3.11
!
access-list 1 permit 239.232.0.0 0.0.255.255
```

Note When the MPLS PE function is extended to the branch routers and the enterprise large campus is connected via an Inter-AS solution, multicast would be a problem if Inter-AS mVPN is not supported on the branch platforms. Currently the ISR platform does not support the Inter-AS mVPN feature, while c7200 does, including both NPE-G1 and NPE-G2. Consult the Cisco IOS feature navigator for up-to-date information.

QoS

Existing WAN Edge QoS models can still be implemented with MPLS WAN setup. At the headend the expectation is that a interface with high link speed is used (DS3 and up). At these speeds, link-efficiency policies such as LFI and cRTP are not required. The Enterprise QoS SRND recommends 5-11 classes at the WAN edge. At the branches, the PE could be configured to map the COS to DSCP, but in our example we assume that the packets are already marked with the appropriate DSCP. If the branches have slow/medium speed links (<T1), then a 3-5 class model is recommended.

The traffic classification has to be done based on EXP (3 bits). This restricts us to at the most 8 classes at the edge. The original IP packet DSCP are preserved and only the 3 bits of IP precedence are copied on to the outgoing EXP at the branches. There are three different variations of core QOS behaviors within a MPLS network:

- Uniform Mode—This is typically deployed when the core MPLS network and the VPNs are part of the same DiffServ domain as would be the case in enterprises. If policers or any other mechanisms re-mark the MPLS EXP values within the MPLS core, these marking changes are propagated to lower-level labels and eventually are propagated to the IP ToS field (MPLS EXP bits are mapped to IP Precedence values on the egress PE).
- Short Pipe Mode—This is typically deployed if the core MPLS network and the VPNs are part of different DiffServ domains. In the case of any re-marking occurrence within the core MPLS network, changes are limited to MPLS EXP re-marking only and are not propagated down to the underlying IP packet's ToS byte.
- Pipe Mode—The main difference between Short Pipe Mode and Pipe Mode is that the PE egress policies (toward the CEs) are provisioned according to the core network's explicit markings and re-markings, not the IP DiffServ markings used within the VPN (although these are preserved). As with Short Pipe Mode, any changes to label markings that occur within the core MPLS cloud do not get propagated to the IP ToS byte when the packet leaves the MPLS network.

Example:

We use a modified version of the 8 class model from the Enterprise QoS SRND (scavenger combined with bulk data) with dual LLQ for voice and video. Recall that MVPN encapsulates the multicast packets into GRE and forwards it as IP packets and not MPLS. So we must ensure that the MVPN packets are accounted for in a specific class, otherwise they would be dropped into the default class. We use the video class for interactive video, streaming video, and any other multicast traffic. A hierarchical policy is applied to shape all the traffic leaving the branch PE.

The sample below shows branch PE configuration for QoS. Similar configuration can be applied at the aggregator adjusted for the link speed.

B11a:

```
class-map match-any Bulk-Data
 match mpls experimental topmost 1
class-map match-any Video
 match mpls experimental topmost 4
 match ip precedence 4
class-map match-any Network-Control
 match mpls experimental topmost 6
 match mpls experimental topmost 7
class-map match-any Critical-Data
 match mpls experimental topmost 2
class-map match-any Call-Signaling
 match mpls experimental topmost 3
 match ip dscp af31
class-map match-any Voice
 match mpls experimental topmost 5
!
policy-map WAN-EDGE
 class Voice
  priority percent 18
 class Video
  priority percent 15
 class Call-Signaling
  bandwidth percent 5
 class Network-Control
  bandwidth percent 5
 class Critical-Data
```

```
  bandwidth percent 27
  random-detect dscp-based
 class Bulk-Data
  bandwidth percent 5
  random-detect dscp-based
 class class-default
  bandwidth percent 25
  random-detect
policy-map MQC-FRTS-1536
 class class-default
  shape average 1460000 14600 0
  service-policy WAN-EDGE
!
interface Serial1/0/0.1 point-to-point
 ip address 125.1.201.1 255.255.255.252
 mpls ip
 frame-relay interface-dlci 16
  class FR-MAP-CLASS-1536
!
map-class frame-relay FR-MAP-CLASS-1536
 service-policy output MQC-FRTS-1536
```

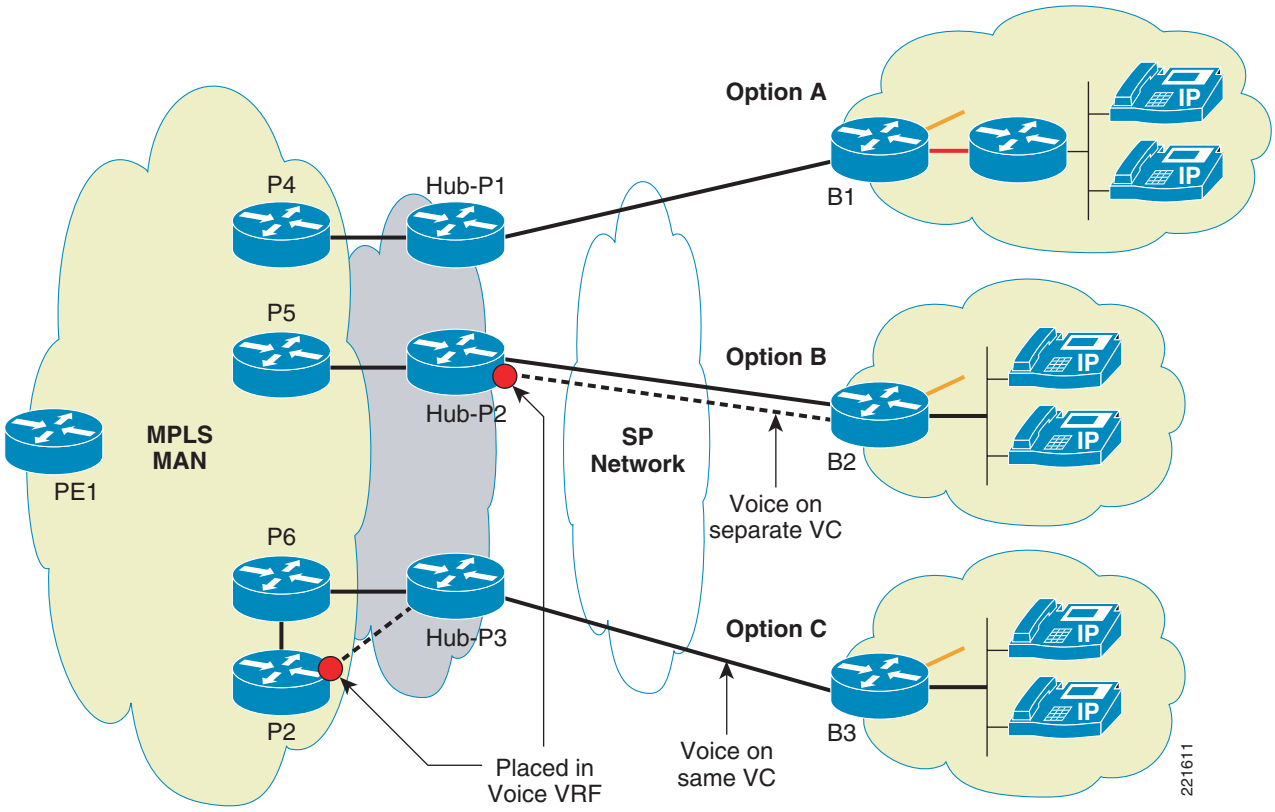
Voice and VRFs

Typically voice traffic has no dependency on the network type since they are just transported as IP packets and require correct QoS behavior applied to them. An exception is when routers are used as gateways for voice services because a lot of voice features and protocols deployed at the branches are not VRF aware (for example, SRST, CME, etc.). Thus just getting the voice traffic in a VRF could be a challenge. This is apart from larger issues of having the voice in a VRF—while you can have the IP phones within a VRF, other services such as softphones VT advantage may be in a different VRF. There are challenges in implementing Inter-VRF IP communications, but they are not discussed here as it is part of the larger virtualization architecture issue. The current recommendation is to keep voice within the global space especially at the branches. At the hub they could remain in the global space or would have to be placed within its own VRF. We look at both options, getting the voice in the VRF at the branch as well keeping it in the global table at the branch.

Voice in a VRF at the Branch

If we need to put the voice in the VRF and still want to use voice features such as CME, then the only way to currently do this is by having two separate routers at the branch. The branch edge router still has a voice VRF configured but treats it like any other VRF. It has a second router (such as a low end ISR) connected to its voice VRF VLAN. The CCME is implemented in the second router, as shown in Figure 4-2 (option A), has all the phones attached to it. Cost might be an issue with this approach as it requires two routers at every such branch site.

Figure 4-2 MPLSoL2—Voice and VRFs



Voice Global at the Branch

If we choose to keep the voice in the global space at the branch, then a single router would be sufficient. The voice VLAN is connected to the branch router but remains in the global space. If the voice is gong to be kept in the global space within the hub network as well, then it can be transported over the existing connection to the hub (MPLS-switched traffic and IP-forwarded traffic share the same link). But at the hub if this traffic needs to be placed within its own VRF, then we would need a separate logical link between the hub and the spoke. This link would be in the global space at the spoke, but be placed within the voice VRF at the hub as shown in Figure 4-2 (option B). The reason we need a separate logical link is that the MPLS link at the hub cannot be placed in a VRF since its configured with “mpls ip” for tag switching. This can potentially increase the circuit cost for the Layer 2 service.

A third option as shown in Figure 4-2 (option C) is to have a separate link at the headend to a PE device which puts the traffic into a VRF. We would need proper routing mechanisms at the hub including route filtering to control the route advertisement within the core network as well as voice VRF.

System Scale and Performance Considerations

Some of the considerations that need to be accounted for from a system scale and performance perspective:

- The WAN aggregator now has IGP and LDP sessions to all the branch routers; the number of peers that it can support can affect the system

- Typically there are large number of branches (into the thousands) and with each one peering directly to the core RR, a more distributed RR design may need to be adopted depending on number of peers supported on a platform.
- In case of MVPN the headend is expected to replicate multicast packets for every spoke receiver and this performance bottleneck can affect the scale (number of branches terminated on each WAN aggregator).
- Converting all the branch routers to PEs increases the footprint of the MPLS network exponentially, from a few 10s within the core network to potentially thousands. This can present a management challenge if the right tools are not used.

Note Inter-AS is another option mentioned in the architecture chapter that will be addressed in the future phases of the solution.