Cisco Systems

MPLS and VPN Architectures

Volume II

Master the latest MPLS VPN solutions to design, deploy, and troubleshoot advanced or large-scale networks

Ivan Pepelnjak, CCIE® No. 1354
Jim Guichard, CCIE No. 2069
Jeff Apcar

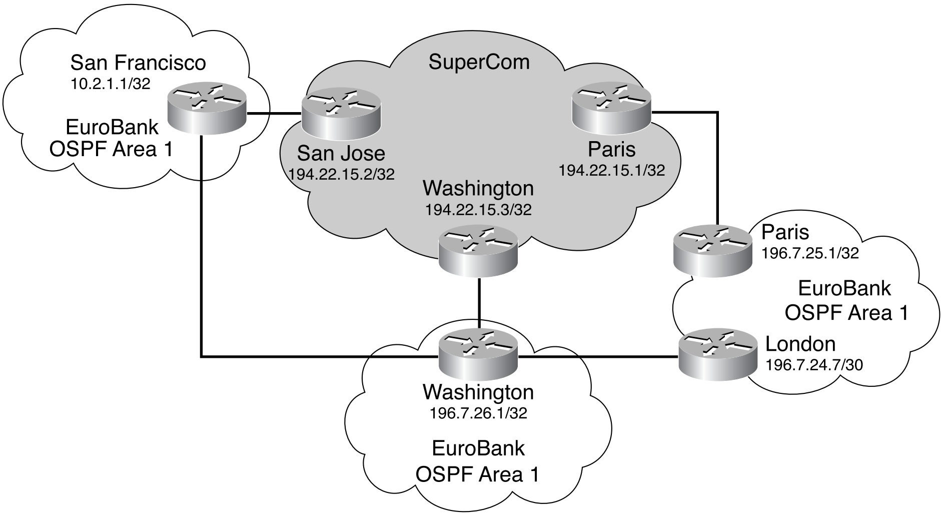
ciscopress.com

MPLS and VPN Architectures, Volume II

Jim Guichard, CCIE No. 2069
Ivan Pepelnjak, CCIE No. 1354
Jeff Apcar

Cisco Press
201 West 103rd Street
Indianapolis, IN 46290 USA

Figure 3-8 *OSPF Backdoor Links*



Example 3-10 *Backdoor Link Selection Example*

```
SanJose# show ip bgp v a 196.7.25.1
BGP routing table entry for 100:251:196.7.25.1/32, version 58
Paths: (3 available, best #2)
  Advertised to non peer-group peers:
  194.22.15.1 194.22.15.3
  Local
    194.22.15.3 (metric 30) from 194.22.15.3 (194.22.15.3)
      Origin incomplete, metric 22, localpref 100, valid, internal
      Extended Community: RT:1:793 OSPF DOMAIN ID:0.0.0.101 OSPF
      RT:1:2:0 OSPF 2
  Local
    192.168.2.13 from 0.0.0.0 (194.22.15.2)
      Origin incomplete, metric 86, localpref 100, weight 32768,
      valid, sourced, best
      Extended Community: RT:1:793 OSPF DOMAIN ID:0.0.0.101 OSPF
      RT:1:2:0 OSPF 2
  Local
    194.22.15.1 (metric 30) from 194.22.15.1 (194.22.15.1)
      Origin incomplete, metric 11, localpref 100, valid, internal
      Extended Community: RT:1:793 OSPF DOMAIN ID:0.0.0.101 OSPF
      RT:1:2:0 OSPF 2

SanJose# show ip route vrf EuroBank 196.7.25.1
Routing entry for 196.7.25.1/32
  Known via "ospf 101", distance 110, metric 86, type intra area
  Redistributing via bgp 215
  Advertised by bgp 215
  Last update from 192.168.2.13 on Serial0/0/0, 00:00:17 ago
  Routing Descriptor Blocks:
  * 192.168.2.13, from 192.168.2.14, 00:00:17 ago, via Serial0/0/0
      Route metric is 86, traffic share count is 1
```

Using this example, you can see that the 196.7.25.1/32 prefix (which is the loopback address of the EuroBank Paris CE router) is learned via Multiprotocol BGP from the Paris and Washington PE routers and is inserted locally into Multiprotocol BGP at the San Jose PE router. The locally generated route is considered the best path within Multiprotocol BGP. However, examination of the EuroBank VRF routing table shows that the selected path is learned via OSPF with a next-hop of 192.168.2.13, which is the EuroBank San Francisco CE router.

This seemingly illogical route selection is made because the intra-area path is preferred over the interarea path generated by the San Jose PE router. In addition, OSPF has a lower administrative distance than Internal BGP. This clearly shows that the MPLS VPN backbone will not be used for any intersite traffic, which will be carried exclusively by the backdoor links between the EuroBank sites. This default behavior is acceptable if the purpose of the connectivity into the MPLS VPN backbone is for backup purposes only. However, because this is generally not the case, the default behavior is not normally acceptable. To overcome this issue, an extra (logical) intra-area link between the PE routers is introduced to the topology. This link, known as a sham-link, is established between the VRF loopback interfaces in the PE routers, and it is treated as an OSPF demand circuit that has no periodic flooding across the link.

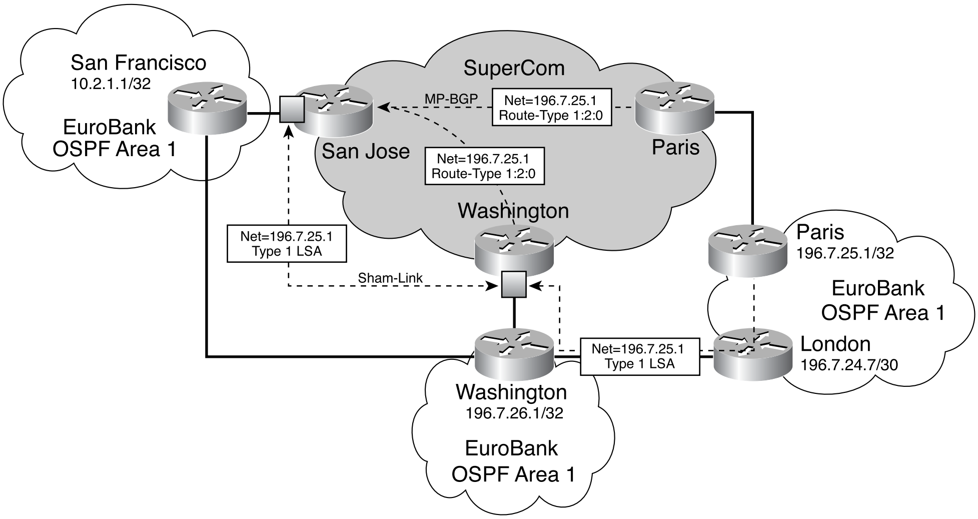
OSPF PE-CE Sham-Link Support

The sham-link provides virtual intra-area connectivity across the MPLS VPN Super-backbone so that traffic can be attracted to the backbone rather than taking the backdoor link between sites. As previously stated, this logical link runs within VRFs of the same VPN between PE routers. An OSPF adjacency is created and database exchange (for the particular OSPF process) occurs across the link. This means that the PE router can flood Type 1 and Type 2 LSAs between sites across the MPLS VPN backbone, thereby creating the desired intra-area connectivity.

With a sham-link configured between PE routers, if the PE router receives an update via Multiprotocol BGP for a particular prefix, it will prefer the intra-area path for the same prefix, which is still learned across the sham-link. Therefore, the traffic will flow across the MPLS VPN backbone.

In our example topology of Figure 3-8, the EuroBank customer has backdoor links between most of its sites; therefore, sham-links are necessary to prevent intersite traffic from crossing the backdoor links. Because backdoor links exist between the San Francisco and Washington CE routers and the Washington and London CE routers, you should deploy sham-links between the PE routers to which the CE routers attach. In our example, this means that a sham-link is required between the San Jose and Washington PE routers and the Washington and Paris PE routers. Figure 3-9 shows the use of the sham-link functionality, but only between the San Jose and Washington PE routers for ease of illustration.

Figure 3-9 *OSPF Sham-Link Deployment*



Creation of sham-links for the EuroBank customer results in two separate sham-links: one between the San Jose PE router and Washington PE router, and another between the Washington PE router and the Paris PE router. It is worth noting that no sham-link exists between the Paris PE router and the San Jose PE router. The reason for this is that no backdoor link exists between the EuroBank San Francisco and Paris sites; therefore, a sham-link is not strictly required. In practice, it might be easier from a provisioning and network management point of view to configure a sham-link between these two sites, thereby creating a full mesh of sham-links for this VPN. This results in several unnecessary sham-links, but it does relieve the service provider from the burden of understanding which VPN site has backdoor links with which other VPN sites.

NOTE A sham-link is required between any two sites that share a backdoor link. If no backdoor link exists between the sites, then a sham-link is not required. If the MPLS VPN backbone is to be used for connectivity, then the OSPF cost of the sham-link must be better than any other path via the backdoor links between the VPN sites.

OSPF Sham-Link Configuration

A separate loopback interface inside the VRF is required for each VRF that is to be connected to other PE routers using sham-links. This loopback interface is used as an endpoint address for the sham-link. The same loopback interface in a VRF can be used to terminate any number of sham links. (There is no requirement for a different loopback address per sham-link within the same VRF.) This loopback address should not be redistributed into Multiprotocol BGP because the sham-link endpoint address is distributed between PE routers using the OSPF extended community attribute with *route-type* 129. Example 3-11 shows the configuration necessary for the creation of a sham-link between the San Jose and Washington PE routers from Figure 3-9.

Example 3-11 *OSPF Sham-Link Configuration*

```
hostname SanJose
!
interface loopback 1
 description ** interface for sham-link to Washington
 ip vrf forwarding EuroBank
 ip address 10.2.1.2 255.255.255.255
!
router ospf 101 vrf EuroBank
 area 1 sham-link 10.2.1.2 196.7.26.2 cost 40

hostname Washington
!
interface loopback 1
 description ** interface for sham-link to San Jose
 ip vrf forwarding EuroBank
 ip address 196.7.26.2 255.255.255.255
!
router ospf 101 vrf EuroBank
 area 1 sham-link 196.7.26.2 10.2.1.2 cost 40
```

The **area** command is used to create the sham-link, and the source and destination loopback interfaces identify the two endpoints of the sham-link. An OSPF cost must be associated with the sham-link so that shortest path first (SPF) can use it to calculate the shortest path. If the cost of the sham-link is better than any backdoor link between sites, then intersite traffic flows across the MPLS VPN backbone. If the cost is higher than the backdoor link path, then intersite traffic flows across the backdoor link. This behavior provides flexibility within the OSPF topology. You can manipulate traffic flow based on administrator-controlled policy (OSPF costs) rather than on the intra-area versus interarea rule.

The **show ip ospf sham-link** command shown in Example 3-12 can now be used to check that the sham-link from Example 3-11 has been successfully created.

Example 3-12 *show ip ospf sham-link Command Output*

```
SanJose# show ip ospf sham-link

Sham Link OSPF_SL0 to address 196.7.26.2 is up
Area 1 source address 10.2.1.2
  Run as demand circuit
  DoNotAge LSA allowed. Cost of using 40 State POINT_TO_POINT,
  Timer intervals configured, Hello 10, Dead 40, Wait 40,
    Hello due in 00:00:04
    Adjacency State FULL (Hello suppressed)
```

continues

Example 3-12 *show ip ospf sham-link Command Output (Continued)*

```
    Index 2/2, retransmission queue length 4, number of
    retransmission 0
    First 0x63311F3C(205)/0x63311FE4(59) Next
    0x63311F3C(205)/0x63311FE4(59)
    Last retransmission scan length is 0, maximum is 0
    Last retransmission scan time is 0 msec, maximum is 0 msec
    Link State retransmission due in 360 msec
```

The output from the previous example confirms that the sham-link is active and that it runs as a demand circuit. (No period flooding occurs across the link, and hellos are suppressed.) The newly created sham-link is advertised within the PE routers Type 1 LSA as an unnumbered point-to-point connection between two PE routers. This is illustrated in Example 3-13.

Example 3-13 *Sham-Link Representation Within the OSPF Database*

```
SanJose# show ip ospf data router 10.2.1.2

            OSPF Router with ID (10.2.1.2) (Process ID 101)

                Router Link States (Area 1)

  LS age: 527
  Options: (No TOS-capability, DC)
  LS Type: Router Links
  Link State ID: 10.2.1.2
  Advertising Router: 10.2.1.2
  LS Seq Number: 8000001F
  Checksum: 0x4CEB
  Length: 60
  Area Border Router
  AS Boundary Router
  Number of Links: 3

    Link connected to: another Router (point-to-point)
     (Link ID) Neighboring Router ID: 196.7.26.2
     (Link Data) Router Interface address: 0.0.0.18
      Number of TOS metrics: 0
       TOS 0 Metrics: 1
```

PE-CE Connectivity: Integrated IS-IS

Now that the enhancements to the OSPF protocol have been covered, it is time to introduce the first new protocol to be added to the list of PE-CE protocols: IS-IS. Although IS-IS is not expected to be one of the more widely deployed protocols for this type of connectivity due to its limited deployment within Enterprise networks, its availability as a PE-CE protocol might still be important in certain scenarios. For example, a VPN client might be running IS-IS on the internal network and might want to maintain the IS-IS topology when moving to an MPLS VPN environment. The primary reasons for this are similar to those discussed within the OSPF section:

- Avoidance of having to redistribute IS-IS information into other protocols such as BGP-4 or RIP version 2 at the CE routers
- Avoidance of having to learn/support another routing protocol such as BGP-4 at the network edge

Support for IS-IS is also important for the migration of an ISP, which uses IS-IS as its routing protocol toward the MPLS VPN backbone, such as in the Carrier's Carrier architecture. This architecture is explained in more detail in Chapter 6, "Large-Scale Routing and Multiple Service Provider Connectivity."

IS-IS PE-CE Connectivity Requirements

IS-IS, like OSPF, is a link-state routing protocol, and it is widely adopted within the service provider community. The technical details of how IS-IS operates are outside the scope of this publication. Readers who require this level of detail should refer to the Cisco Press book *IS-IS Network Design Solutions*, written by Abe Martey.

As with the OSPF protocol, IS-IS can split a routing domain into a series of areas where interarea connectivity is achieved by interconnection across a Level 2 backbone, partially overlaying the individual Level 1 areas. In general, small IS-IS topologies are built within a single area, and this area includes all the routers within the routing domain. As the network increases in size, it is split into a Level 2 backbone and a number of Level 1 areas. Routers establish Level 1 adjacencies to perform routing within a local area (intra-area routing) and Level 2 adjacencies to perform routing between Level 1 areas (interarea routing).

The IS-IS Level 2 backbone is created through the connection of all Level 2 routers from all areas, and local areas attach to the backbone via a Level 1-2 router. Within a local area, all routers know how to reach all other routers within the area, but they know nothing about routers in other areas.

The default behavior of a Cisco router for the first IS-IS process to be created is to act as a Level 1-2 router. This is basically a combination of Level 1 and Level 2. (The router establishes both Level 1 and Level 2 adjacencies and maintains two separate databases: one for the local Level 1 area and another for the Level 2 backbone.) You can configure the router to act as a Level 1 (intra-area) router only, as both a Level 1 router and a Level 2 (interarea) router (the default), or as an interarea router only. Because of this range of options, various combinations for connectivity can be established.

With the introduction of an MPLS VPN backbone between VPN sites, an additional Level of routing hierarchy (referred to as Level 3) above Level 2 has been added (similar to OSPF). This additional level is required so that VPN sites can run independent IS-IS processes and learn routes from other VPN sites without maintaining a direct adjacency with those sites. With this additional level, the routing hierarchy changes from Level 1/Level 2/Level 1 to Level 1/Level 2/Level 3/Level 2/Level 1. This gives various connectivity options between the PE routers and CE routers.