Network Virtualization—Services Edge Design Guide

Cisco Validated Design

February 23, 2009

The centralization of access to shared services provides a common point of policy enforcement and control for all VPNs. This is referred to as the services edge functional area. Services edge has more of a logical than a physical meaning. In a specific network design, the point of policy enforcement can be physically located in a specific area of the network, but in certain cases, it might also be spread around the network.

For related information, see the following documents:

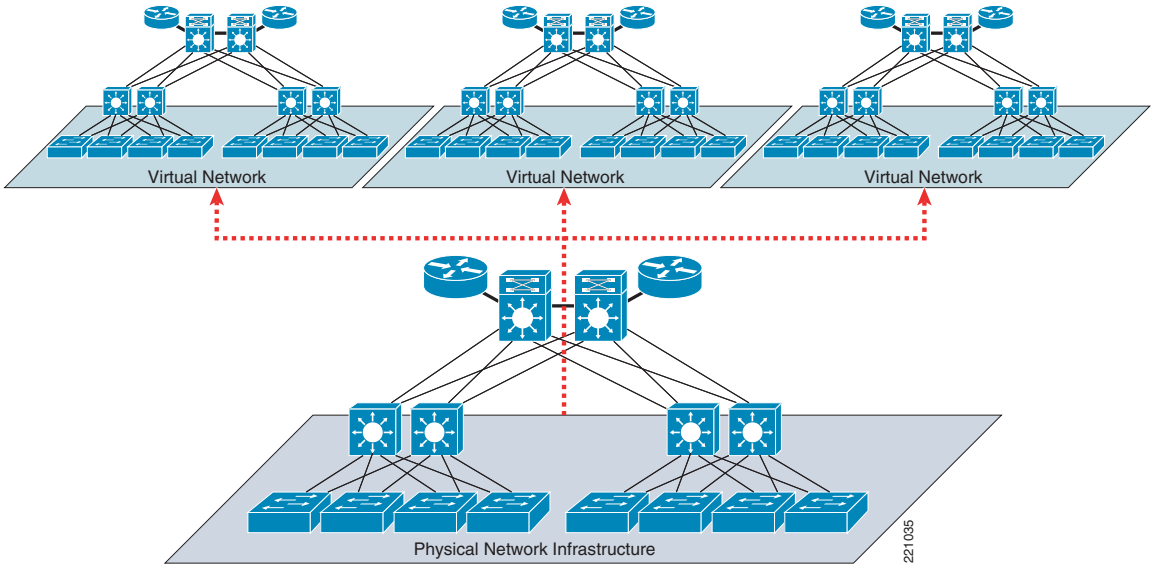
- *Network Virtualization—Guest and Partner Access Deployment Guide*— http://www.cisco.com/en/US/docs/solutions/Enterprise/Network_Virtualization/GuestAcc.html
- *Network Virtualization—Network Admission Control Deployment Guide*— http://www.cisco.com/en/US/docs/solutions/Enterprise/Network_Virtualization/NACDepl.html
- *Network Virtualization—Path Isolation Design Guide*— http://www.cisco.com/en/US/docs/solutions/Enterprise/Network_Virtualization/PathIsol.html

Introduction

The term *network virtualization* refers to the creation of logical isolated network partitions overlaid on top of a common enterprise physical network infrastructure, as shown in Figure 1.

Americas Headquarters:
Cisco Systems, Inc., 170 West Tasman Drive, San Jose, CA 95134-1706 USA

Figure 1 Network Virtualization



Each partition is logically isolated from the others and must provide the same services that would be available in a traditional dedicated enterprise network. This essentially means that the experience of the end user is that of being connected to a dedicated network that provides privacy, security, an independent set of policies, service level, and even routing decisions.

At the same time, the network administrator can easily create and modify virtual work environments for the various groups of users, and adapt to changing business requirements in a much easier way. The latter derives from the ability to create security zones that are governed by policies enforced centrally. Because policies are centrally enforced, adding users and services to or removing them from a VPN requires no policy reconfiguration. Meanwhile, new policies affecting an entire group can be deployed centrally at the VPN perimeter. Thus, virtualizing the enterprise network infrastructure provides the benefits of leveraging multiple networks but not the associated costs, because operationally they should behave like one network (reducing the relative operating expenses).

Network virtualization responds to both simple and complex business drivers. As an example of a simple scenario, an enterprise wants to provide Internet access to visitors (guest access). The stringent requirement in this case is to allow visitors external Internet access while preventing any possibility of connection to the enterprise internal resources and services. This can be achieved by dedicating a logical "virtual network" to handle the entire guest communications. A similar case is where Internet access can be combined with connectivity to a subset of the enterprise internal resources, as is typical in partner access deployments.

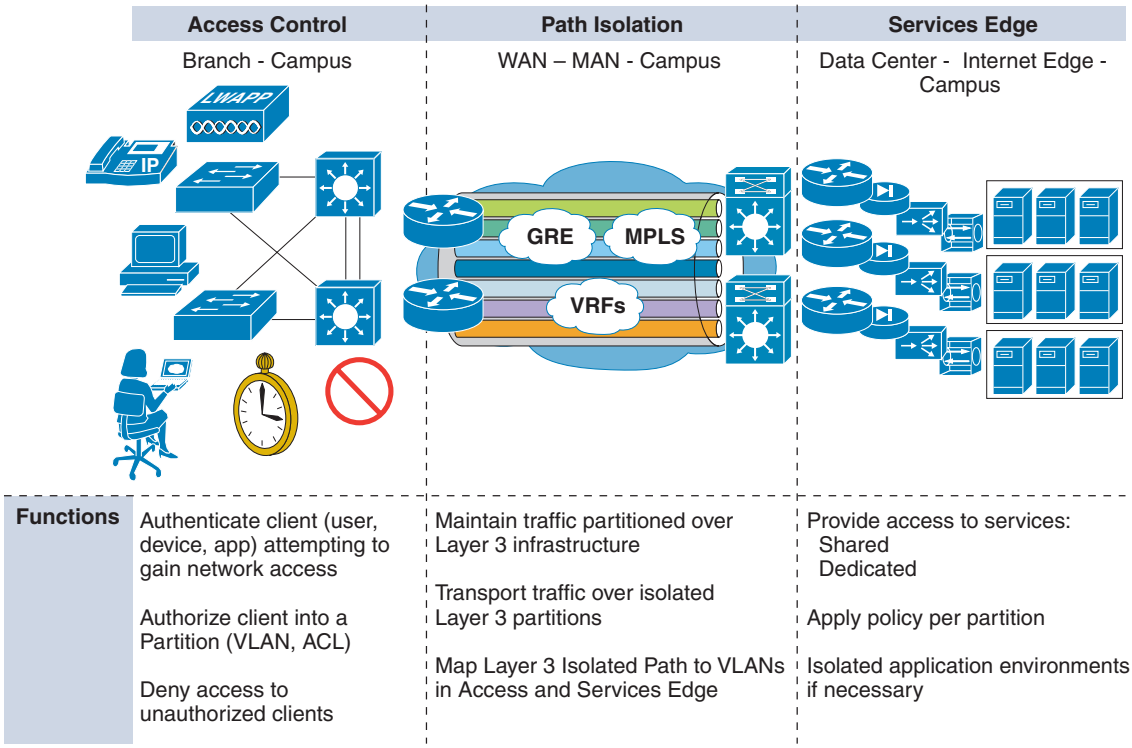
Another simple scenario is the creation of a logical partition to be dedicated to the machines that have been quarantined as a result of a Network Access Control (NAC) posture validation. In this case, it is essential to guarantee isolation of these devices in a remediation segment of the network, where only access to remediation servers is possible until the process of cleaning and patching the machine is successfully completed.

As an example of a more complex scenario, an enterprise IT department starts functioning as a service provider, offering access to the enterprise network to a variety of "customers" that need to be kept logically isolated from each other. Users belonging to each logical partition can communicate with each other and can access dedicated network resources, but inter-communication between groups is prohibited. A typical deployment scenario in this category involves retail stores that provide on-location network access for kiosks or hotspot providers.

The architecture of an end-to-end network virtualization solution that is targeted to satisfy the requirements listed above can be separated in three logical functional areas (see Figure 2):

- Access control
- Path isolation
- Services edge

Figure 2 Network Virtualization—Three Functional Areas



Each area performs several functions and interfaces with the other functional areas to provide a complete integrated end-to-end solution.

Each of these areas is discussed in great detail in a separate design guide. This document addresses the requirement of the services edge. For information on the other two functional areas, see the following guides:

- *Network Virtualization— Access Control Design Guide—* http://www.cisco.com/en/US/docs/solutions/Enterprise/Network_Virtualization/AccContr.html
- *Network Virtualization—Path Isolation Design Guide—* http://www.cisco.com/en/US/docs/solutions/Enterprise/Network_Virtualization/PathIsol.html

The virtualization of the enterprise network allows for the creation of a separate logical network that is placed on top of the physical infrastructure. The default state of these virtual networks (VPNs) is to be totally isolated from each other, in this way simulating separate physical networks.

This default behavior may need to be changed when the various VPNs need to share certain services, such as Internet access as well as network services such as DHCP and DNS and server farms.

This document presents alternative ways to accomplish this sharing of resources between various VPNs. The services that need to be shared are discussed, as well as the distinction between protected and unprotected services. This document broadly categorizes services that are shared by many VPNs as either protected or unprotected, depending on how they are accessed.

Various technologies are discussed that achieve the sharing of resources between different network partitions. To make good use of this document, note the following:

- The various technologies are discussed in the context of the network virtualization solution. This means that for these technologies, the details that have been validated and positioned as part of the network virtualization project to provide an answer to the business problems previously listed are discussed.
- Not all the technologies found in this design guide represent the right fit for each business problem. For example, there may be scenarios (such as guest access) where resources are dedicated to the specific virtual network and no sharing at all is required. To properly map the technologies discussed here with each specific business problem, reference the following deployment guides:
  - *Network Virtualization—Access Control Design Guide*—http://www.cisco.com/en/US/docs/solutions/Enterprise/Network_Virtualization/AccContr.html
  - *Network Virtualization—Guest and Partner Access Deployment Guide*—http://www.cisco.com/en/US/docs/solutions/Enterprise/Network_Virtualization/GuestAcc.html
  - *Network Virtualization—Network Admission Control Deployment Guide*—http://www.cisco.com/en/US/docs/solutions/Enterprise/Network_Virtualization/NACDepl.html
  - *Network Virtualization—Path Isolation Design Guide*—http://www.cisco.com/en/US/docs/solutions/Enterprise/Network_Virtualization/PathIsol.html

Services Edge—Document Scope

The services edge portion of the overall network virtualization process is where a large part of policy enforcement and traffic manipulation is done. Before the services edge is implemented, it is important to thoroughly understand which methodology is to be deployed and what the trade-offs are for selecting the methods described in this guide. It is also important for customers to understand their applications and their associated traffic flows to help in the overall network optimization process.

This guide accomplishes the following:

- Provides guidelines on how to accomplish the sharing of services between separate logical partitions, distinguishing two methods of access, unprotected and protected.
- Discuss the importance of services virtualization, as the Cisco Firewall Services Module (FWSM), in providing the services edge functionalities.
- Introduces a first technical option to integrate Unified Communication (UC) applications (voice, video) into a virtualized network environment. The scope of this solution is initially limited to campus deployments and would leverage the concept of a shared services area in the network to deploy UC services (like Cisco Call Manager, TFTP servers, etc.) that need to be accessed by network entities (IP phones, PCs, etc.) deployed in the context of separate virtual networks.
- Although this guide addresses many technical areas, it does not currently discuss the use of overlapping IP addresses in the different VPNs (IP address overlap may be addressed in the future). The use of overlapping IP addresses is usually discouraged because of the operational impacts that

this causes to customer networks in the operations and management aspects of the network infrastructure, and should be used only in scenarios (as for example merger and acquisitions) where it is not possible to avoid it.

Note In the rest of this document, the terms VPNs, virtual networks, and VRFs may be used interchangeably. In the context of this discussion, these terms all refer to the logical partitions that are deployed on top of the common physical network infrastructure.

It is also important to highlight that the services edge functionalities discussed in this document are independent from the access control strategy implemented to provide differentiated access to the various network entities and from the implemented path isolation technical alternatives. This is in line with the idea of creating a solution framework where the three functional areas (access control, path isolation, and services edge) can be deployed independently and interfaced with each other to provide the end-to-end solution.

Also, one of the main advantages of virtualizing a network infrastructure consists in the ability of creating flexible interfaces between the various virtual networks; this can be done to provide inter-VPN communications or share access to a specific set of resources.

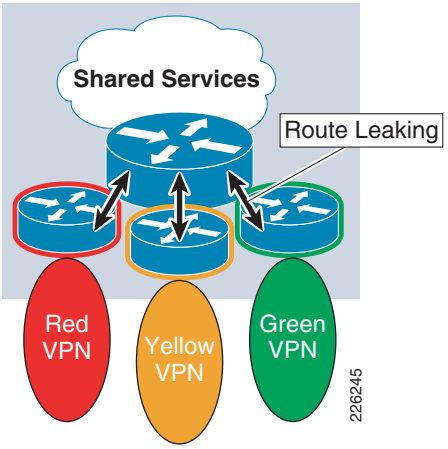
Finally, from a deployment point of view, services edge is more a logical and abstract entity than a physical one. The services edge functionalities may be physically deployed in different areas of the enterprise network: the data center, the Internet edge, a dedicated services block connected to the campus core, etc. It is also possible to deploy that functionality in multiple physical locations in order to increase the overall high availability of the solution; both the intra- and inter-site designs are discussed in the context of this document.

Unprotected Services Access

Unprotected services access means allowing communication to shared services without subjecting the traffic to any type of security check. An unprotected service is reachable from one or more VPNs without having a policy enforcement point between the service and the requesting host. As such, access to unprotected services usually is performed by following the best path route available in the routing table of the various VPNs, without enforcing the hair-pinning of traffic to a central location (we'll see this is the typical model used for protected access instead).

The technical solution to implement unprotected services access consists in leaking prefixes between the routing tables associated to each defined VPN. Remember that all these logical partitions aim to mimic separate physical networks but are actually deployed on a common underlying infrastructure, which allows leveraging several mechanism to open communication channels between them. Figure 3 provides a high-level view of this idea.

Figure 3 Unprotected Services Access



Because of its nature, this technical option is inherently unsecured and should be deployed carefully to avoid opening undesired back doors between VPNs. This implies that route leaking should not be used to provide peer-to-peer (inter-VPN) connectivity, but remain limited to the creation of an "Extranet VPN" which can communicate with different VPNs without providing a transit zone between them. The recommendation is to deploy unprotected services sharing in a limited fashion, for example to provide access to DHCP or DNS services to the various VPNs without adding an unnecessary load to the firewalls that are being used to control access to other shared services that must be protected.

Protected Services Access

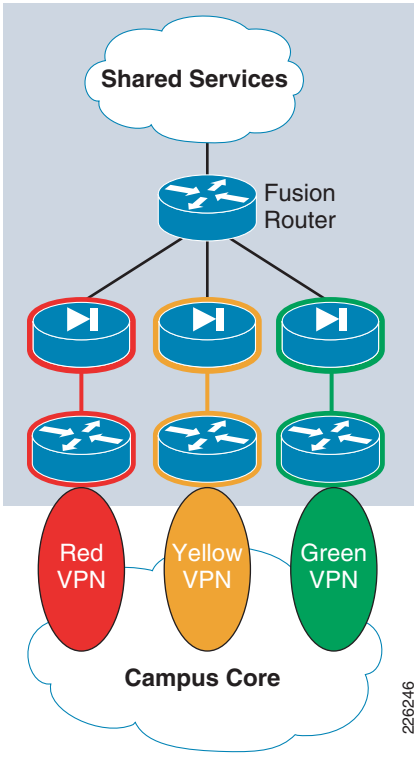
Protected services must be accessible from the VPNs, but only after specific security policies are enforced. To be able to enforce the necessary security policies in a manageable way, access to the services must go through a policy enforcement point. Thus, all traffic reaching the services must be routed through a common point of policy enforcement. As a result, the routing between a requesting host and a service can potentially be less than optimal. However, this is true only in very specific scenarios, such as when the shared services themselves are part of a VPN. In general, shared services that are to be protected are centrally located for optimal accessibility.

Examples of protected services include server farms and the Internet. When accessing the Internet, not only is it necessary to control access to the service from the VPNs, but it is also critical to control any access initiated from the service area towards the VPNs. Ideally, none of the VPNs should be accessed from the Internet unless the communication is initiated from inside the VPN; thus access into the VPNs from the services area is generally prohibited.

In cases where VPNs must communicate with each other in a controlled manner, the policies at the VPN perimeter can be changed to provide such access. In this particular inter-VPN connectivity application, the policies must be open to allow externally-initiated communication into the VPNs.

From an architectural point of view, the design allowing providing protected services access is shown in Figure 4.

Figure 4 Protected Services Edge



As noted above, each defined virtual network is front-ended by a security device (usually a firewall) that allows providing tight control on what type of communications are allowed in both inbound and outbound directions. Providing a separate firewall per VPN allows applying and managing the security policies for each virtual network independently, and it is hence the recommended deployment model. As we discuss further in the following sections, the virtualization of the firewall services available with both Cisco FWSM and Cisco ASA allows dedicating a virtual firewall instance (usually named context) to each logical partition.

The fusion router is the brain of the solution, responsible for properly routing traffic between each VPN and the shared pool of resources or even between separate VPNs; it is worth noticing how the shared pool of resources could in fact also be deployed as part of a dedicated VPN. The fusion router is usually aware of the prefixes available inside each VPN, either because of static routing configuration or through routing peering. When discussing the deployment of the services edge in more detail, we clarify how the chosen protocol may depend on the path isolation strategy adopted and on the firewall configuration (transparent or routed mode).

The fusion router functionality can be deployed in two different ways:

1. Dedicating to this function a physically separate router (or Layer 3 switch).
2. Defining a specific VRF to be used for this purpose.

The distinction between these two options would steer two different designs that in this document are named dual tier and single tier implementations.

Finally, the concept of shared services may have different meanings in different deployments. Common examples of shared services can be:

- Server farm
- The entire Internet

- The non-virtualized portion of the enterprise network (i.e., the “global table”).

Deploying Unprotected Shared Services

The deployment of unprotected shared services requires the deployment of some form of route leaking between the different VPNs. The details around this type of configuration are presented in the following section.

Leaking Routes between VRFs

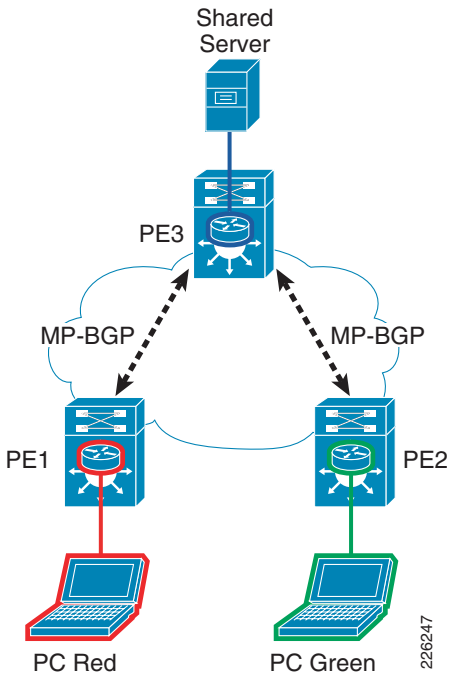
The basic element to perform route leaking between separate VRFs is the route-target BGP attribute. As a consequence, BGP needs to be enabled every time this mechanism is leveraged. Depending on the chosen path isolation technique, BGP may not be the deployed control plane protocol to implement routing functionalities inside each VPN. In the path isolation document, it is discussed how BGP is usually leveraged to exchange VPN routes for MPLS VPN deployments. At the same time, in scenarios where VRF-Lite is leveraged end-to-end across the network, the control protocol for each defined VRF is usually the same IGP (EIGRP or OSPF) already deployed in global table.

Note For more information on the different path isolation technologies, refer to: http://www.cisco.com/en/US/docs/solutions/Enterprise/Network_Virtualization/PathIsol.html

Because of the considerations above, it is possible to distinguish two main scenarios where BGP is used to perform route-leaking between VRFs:

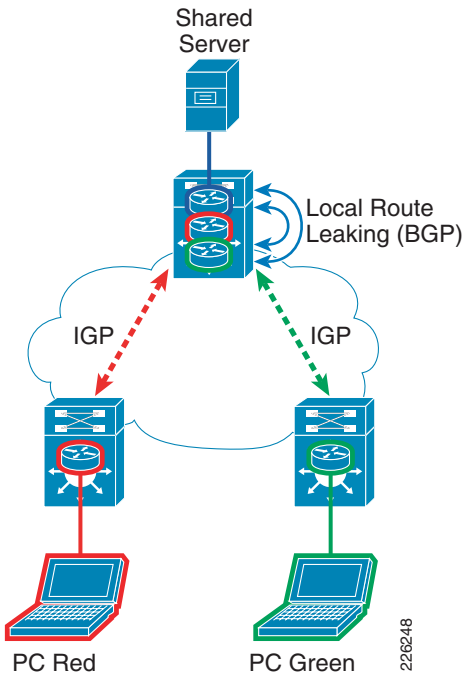
- Multi-device deployments—In this case, MPLS VPN is typically the chosen path isolation technique and MP-BGP is the control protocol leveraged to exchange the VPN routes between the defined PE devices (as shown in Figure 5).

Figure 5 Unprotected Access—Multi-Device Deployment



- Single device deployments—In these deployments, VRF-Lite is usually leveraged for providing path isolation capabilities and BGP is not used for the control plane (this functionality is performed by IGPs running in the context of each defined VPN). As a consequence, BGP is usually only enabled locally on the device where route-leaking must be performed, without requiring the establishment of any BGP neighborship with external devices. This concept is shown in Figure 6.

Figure 6 Unprotected Access—Single Device Deployment

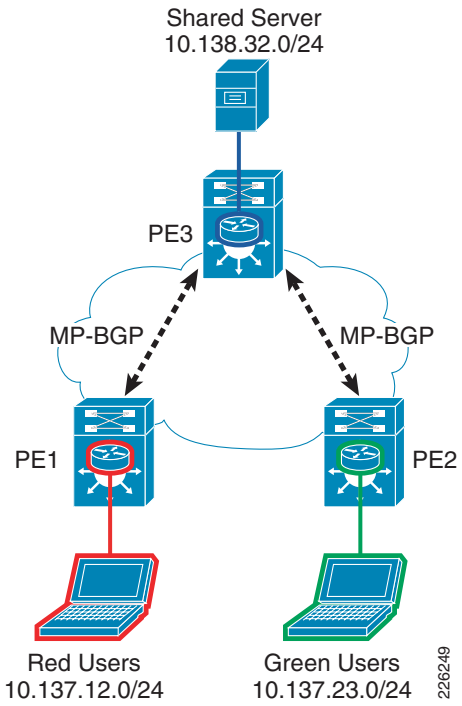


The following two sections highlight the required configuration steps to deploy both the multi-device and single device models.

Multi-Device Model Configuration

The configuration steps described below reference the specific example shown in Figure 7.

Figure 7 Multi-Device Deployment Example



In this scenario, a server directly connected to the PE3 device needs to be shared between Red and Green users who connect to the bottom PE1 and PE2 devices. This needs to be accomplished without compromising the logical isolation between Red and Green virtual networks.

- Initial conditions—The shared resource IP prefix is only known on PE3 in the context of the shared VRF and not in the Red and Green VRFs of PE1 and PE2.

  PE3

```
PE1#sh ip route vrf Shared 10.138.32.0
Routing entry for 10.138.32.0/24
  Known via "connected", distance 0, metric 0 (connected, via interface)
  Redistributing via eigrp 100
  Routing Descriptor Blocks:
  * directly connected, via Vlan32
      Route metric is 0, traffic share count is 1
```

  PE1

```
PE2#sh ip route vrf  Red 10.138.32.0
% Subnet not in table
```

  PE2

```
PE3#sh ip route vrf Green 10.138.32.0
% Subnet not in table
```

- The configuration on the three devices is properly modified to allow route-leaking of the shared prefix into the Red and Green VRF routing tables.

  1. Configure route-target attributes on PE3 to export the shared IP prefixes and import into the BGP table the Red and Green prefixes learned from the iBGP neighbor devices.

     PE3

```
ip vrf Shared
 rd 3:3
 route-target export 300:300
```

```
 route-target import 100:100
 route-target import 200:200
```

2. Configure route-target attributes on PE1 to export the Red IP prefixes and import into the BGP table the shared prefixes learned from PE3.

   PE1

```
ip vrf Red
 rd 1:1
 route-target export 100:100
 route-target import 300:300
     route-target import 100:100
```

3. Configure route-target attributes on PE2 to export the Green IP prefixes and import into the BGP table the shared prefixes learned from PE3.

   PE2

```
ip vrf Green
 rd 2:2
 route-target export 200:200
 route-target import 300:300
 route-target import 200:200
```

Note The last **route-target** commands in both PE1 and PE2 configuration samples (importing 100:100 and 200:200) are not required for providing access to the shared services but are shown above since they are usually used to provide any-to-any connectivity inside each VRF.

- Once the configuration is in place, it is possible to verify that the shared prefixes is now learned in the Red and Green VRFs (and displayed as "directly connected"); at the same time, the Red and Green subnets are injected in the shared VRF.

  PE1

```
PE1#sh ip route vrf Red 10.138.32.0
Routing entry for 10.138.32.0/24
  Known via "bgp 100", distance 200, metric 0, type internal
  Last update from 192.168.100.100 00:29:47 ago
  Routing Descriptor Blocks:
  * 192.168.100.100 (Default-IP-Routing-Table), from 192.168.100.100, 00:29:47 ago
      Route metric is 0, traffic share count is 1
      AS Hops 0
      MPLS label: 18
      MPLS Flags: MPLS Required
```

  PE2

```
PE2#sh ip route vrf Green 10.138.32.0
Routing entry for 10.138.32.0/24
  Known via "bgp 100", distance 200, metric 0, type internal
  Last update from 192.168.100.100 00:30:35 ago
  Routing Descriptor Blocks:
  * 192.168.100.100 (Default-IP-Routing-Table), from 192.168.100.100, 00:30:35 ago
      Route metric is 0, traffic share count is 1
      AS Hops 0
      MPLS label: 18
      MPLS Flags: MPLS Required
```

  PE3

```
PE3#sh ip route vrf Shared 10.137.12.0
```

```
Routing entry for 10.137.12.0/24
  Known via "bgp 100", distance 200, metric 0, type internal
  Last update from 192.168.100.1 00:31:51 ago
  Routing Descriptor Blocks:
  * 192.168.100.1 (Default-IP-Routing-Table), from 192.168.100.1, 00:31:51 ago
      Route metric is 0, traffic share count is 1
      AS Hops 0
      MPLS label: 16
      MPLS Flags: MPLS Required
PE3#sh ip route vrf Shared 10.137.23.0
Routing entry for 10.137.23.0/24
  Known via "bgp 100", distance 200, metric 0, type internal
  Last update from 192.168.100.2 00:31:59 ago
  Routing Descriptor Blocks:
  * 192.168.100.2 (Default-IP-Routing-Table), from 192.168.100.2, 00:31:59 ago
      Route metric is 0, traffic share count is 1
      AS Hops 0
      MPLS label: 17
      MPLS Flags: MPLS Required
```

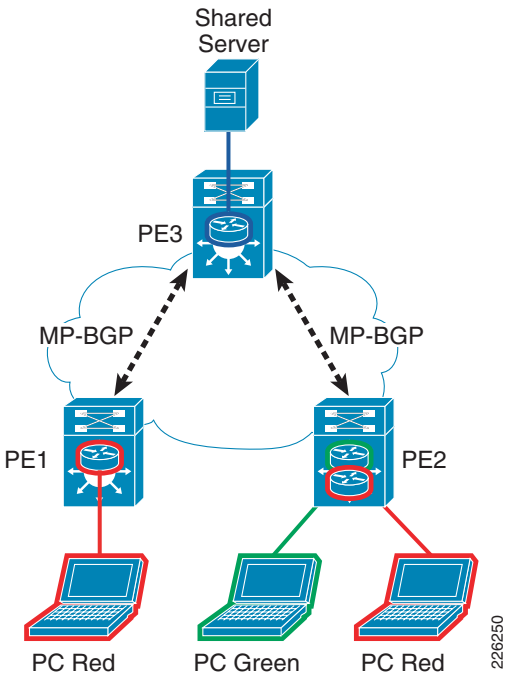
- As a result of the configuration steps above, Red and Green VRFs routing table contain routes for the shared services IP subnets (and vice versa), but the shared VRF does not act as a transit area between VPNs, maintaining the desired logical separation between the Red and Green virtual networks. This behavior is due to the non-transitive nature of iBGP; in order to clarify this concept, the route-target configuration for the shared VRF is shown again below:

  ```
  ip vrf Shared
   rd 3:3
   route-target export 300:300
   route-target import 100:100
   route-target import 200:200
  ```

  The Red and Green routes are imported into the shared VRF table (because of the route-target import commands) and are consequently **not exported** again (via the **route-target export** command) because of the non-transitive characteristics of BGP. If that was not the case, the routes would then be imported in the Red and Green VRFs at the remote PEs (PE1 and PE2), breaking the logical isolation between these virtual networks.

  Another implication of this BGP behavior is the fact that importing routes into a VRF on a specific physical router does not populate these routes in the routing table of other devices where the same VRF may be extended. For example, assume that the Red VRF is now defined also on the PE2 device, as shown in Figure 8.

Figure 8 Non Transitive Nature of MP-iBGP



The fact that the shared IP prefix is imported into the Red VRF on the PE1 device (based on the configuration steps previously discussed) does not make the same information available in the Red VRF routing table in PE2. In order for that to happen, the same route-target configuration applied to PE1 needs to be added on PE2, as shown below:

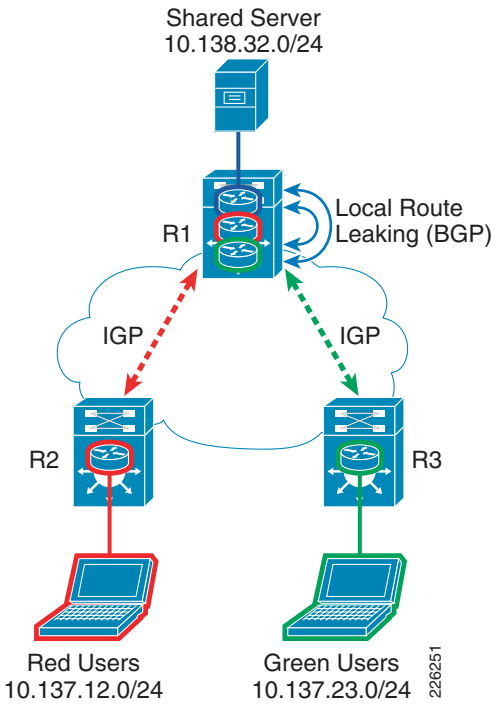
PE2

```
ip vrf Red
 rd 1:1
 route-target export 100:100
 route-target import 300:300
```

Single Device Model Configuration

The configuration steps described below reference the example shown in Figure 9.

Figure 9 Single Device Model Example



The objective of the design is identical to the one discussed in the multi-device model—providing access to the shared service to both Red and Green users without losing the logical separation between them.

- Initial condition—The shared resource IP prefix is only known on R1 in the context of the shared VRF and not in the Red and Green VRFs of routers R2 and R3.

  R1

```
R1#sh ip route vrf Shared 10.138.32.0
Routing entry for 10.138.32.0/24
  Known via "connected", distance 0, metric 0 (connected, via interface)
  Redistributing via eigrp 100
  Routing Descriptor Blocks:
  * directly connected, via Vlan32
      Route metric is 0, traffic share count is 1
```

  R2

```
R2#sh ip route vrf  Red 10.138.32.0
% Subnet not in table
```

  R3

```
R3#sh ip route vrf Green 10.138.32.0
% Subnet not in table
```

- Local route leaking is deployed on the device to which the resource is directly connected (R1 in our example). Since the initial assumption is that the single platform approach is used in VRF-Lite deployments (where an IGP and not MP-BGP is usually used as routing protocol), the following configuration steps need to be performed:

  1. Configure the route-target parameters for the VRFs. The shared IP prefix must be exported from the shared VRF and imported in the Red and Green user VRFs. At the same time, the user subnets exported from the Red and Green VRF need to be imported in the shared VRF in order to allow the establishment of two-way connectivity.

     R1

     ```
     ip vrf Red
      rd 1:1
      route-target export 100:100
      route-target import 300:300
     !
     ip vrf Green
      rd 2:2
      route-target export 200:200
      route-target import 300:300
     !
     ip vrf Shared
      rd 3:3
      route-target export 300:300
      route-target import 200:200
      route-target import 100:100
     ```

  2. Enable the BGP process to activate the exchange of prefixes by leveraging the route-target configuration. Note how there is no need to specify any neighbor relationships in the BGP configuration since the process has only a local function in this case. Also, the optional use of route-map statements allows for a tighter control on the IP prefixes that are leaked between the different routing tables.

     R1

     ```
     access-list 1 permit 10.138.32.0 0.0.0.255
     access-list 2 permit 10.137.12.0 0.0.0.255
     access-list 3 permit 10.137.23.0 0.0.0.255
     !
     route-map Allowed_Green_Users permit 10
      match ip address 3
     !
     route-map Allowed_Red_Users permit 10
      match ip address 2
     !
     route-map Shared_Services permit 10
      match ip address 1
     ```

     EIGRP specific configuration

     ```
     router bgp 100
      no synchronization
      bgp log-neighbor-changes
      no auto-summary
      !
      address-family ipv4 vrf Shared
       redistribute connected route-map Shared_Services
       no synchronization
      exit-address-family
      !
      address-family ipv4 vrf Green
       redistribute eigrp 100 route-map Allowed_Green_Users
     ```

```
 no synchronization
exit-address-family
!
address-family ipv4 vrf Red
 redistribute eigrp 100 route-map Allowed_Red_Users
 no synchronization
exit-address-family
```

OSPF specific configuration

```
router bgp 100
 no synchronization
 bgp log-neighbor-changes
 no auto-summary
 !
 address-family ipv4 vrf Shared
  redistribute connected route-map Shared_Services
  no synchronization
 exit-address-family
 !
 address-family ipv4 vrf Green
  redistribute ospf 2 route-map Allowed_Green_Users
  no synchronization
 exit-address-family
 !
 address-family ipv4 vrf Red
  redistribute ospf 1 route-map Allowed_Red_Users
  no synchronization
 exit-address-family
```

The user subnets (in VRF Red and Green) that need to have access to the shared services are instead injected into BGP by redistributing from the IGP that it is used to learn them (EIGRP or OSPF in this example). The prefix for the shared services is injected into BGP using the **redistribute connected** command, since in our example i tis directly connected to the Layer 3 device where the route leaking is performed. If the service was instead connected to a different Layer 3 device, it could be injected into BGP by redistributing the IGP used to learn it, similarly to what happens to the user subnets.

Note If the route-maps were not configured, the final results would be to leak all the Red and Green VRF prefixes into the shared VRF routing table. Note that this would not compromise the logical isolation between the Red and Green VPNs, as expected.

- Once the configuration is in place, it is possible to verify that the shared prefixes are now injected in the Red and Green VRFs routing tables (and displayed as “directly connected”); at the same time, the Red and Green subnets are injected in the shared VRF routing table.

  R1

```
R1#sh ip route vrf Red 10.138.32.0
Routing entry for 10.138.32.0/24
  Known via "bgp 100", distance 20, metric 0 (connected, via interface), type
external
  Redistributing via eigrp 100, bgp 100
  Routing Descriptor Blocks:
  * directly connected, via Vlan32
      Route metric is 0, traffic share count is 1
      AS Hops 0
      MPLS label: none

R1#sh ip route vrf Green 10.138.32.0
```

```
Routing entry for 10.138.32.0/24
  Known via "bgp 100", distance 20, metric 0 (connected, via interface), type
external
  Redistributing via eigrp 100, bgp 100
  Routing Descriptor Blocks:
  * directly connected, via Vlan32
      Route metric is 0, traffic share count is 1
      AS Hops 0
      MPLS label: none

R1#sh ip route vrf Shared 10.137.12.0
Routing entry for 10.137.12.0/24
  Known via "bgp 100", distance 20, metric 3840, type external
  Last update from 10.122.15.18 on GigabitEthernet1/3.412, 00:57:13 ago
  Routing Descriptor Blocks:
  * 10.122.15.18 (Red), from 0.0.0.0, 00:57:13 ago, via GigabitEthernet1/3.412
      Route metric is 3840, traffic share count is 1
      AS Hops 0
      MPLS label: none

R1#sh ip route vrf Shared 10.137.23.0
Routing entry for 10.137.23.0/24
  Known via "bgp 100", distance 20, metric 3840, type external
  Last update from 10.122.25.18 on GigabitEthernet1/3.422, 00:57:18 ago
  Routing Descriptor Blocks:
  * 10.122.25.18 (Green), from 0.0.0.0, 00:57:18 ago, via GigabitEthernet1/3.422
      Route metric is 3840, traffic share count is 1
      AS Hops 0
      MPLS label: none
```

However, communication between Red and Green users and the shared resource is still not achieved at this point. The reason is that the shared IP prefix has been leaked in the Red and Green VRF routing table, but only locally on the R1 device. R2 and R3 still do not have that information in their routing table. In order to propagate the shared prefix to R2 and R3, it is required to advertise that subnet into the IGP used as control plane for the Red and Green VRFs end-to-end across the network. This is accomplished by redistributing on R1 the shared prefix from BGP into the IGP, as shown in the configuration sample below:

EIGRP specific configuration

```
router eigrp 100
 address-family ipv4 vrf Green
  redistribute bgp 100 metric 100000 1 255 1 1500
!
 address-family ipv4 vrf Red
  redistribute bgp 100 metric 100000 1 255 1 1500
```

OSPF specific configuration

```
router ospf 1 vrf Red
 redistribute bgp 100 subnets
!
Router ospf 2 vrf Green
 redistribute bgp 100 subnets
```

At this point is possible to see how the shared prefix is learned via IGP on both R1 and R2, which allows successful communication between the clients and the server (the example below is valid for EIGRP deployments).

R2

```
R2#sh ip route vrf Red 10.138.32.0
Routing entry for 10.138.32.0/24
  Known via "eigrp 100", distance 90, metric 3840, type internal
```

```
  Redistributing via eigrp 100
  Last update from 10.137.11.7 on GigabitEthernet5/2.612, 00:06:06 ago
  Routing Descriptor Blocks:
  * 10.137.11.7, from 10.137.11.7, 00:06:06 ago, via GigabitEthernet5/2.612
      Route metric is 3840, traffic share count is 1
      Total delay is 50 microseconds, minimum bandwidth is 1000000 Kbit
      Reliability 255/255, minimum MTU 1500 bytes
      Loading 1/255, Hops 4
```

R3

```
R3#sh ip route vrf Green 10.138.32.0
Routing entry for 10.138.32.0/24
  Known via "eigrp 100", distance 90, metric 3840, type internal
  Redistributing via eigrp 100
  Last update from 10.137.21.11 on Vlan123, 00:02:18 ago
  Routing Descriptor Blocks:
  * 10.137.21.11, from 10.137.21.11, 00:02:18 ago, via Vlan123
      Route metric is 3840, traffic share count is 1
      Total delay is 50 microseconds, minimum bandwidth is 1000000 Kbit
      Reliability 255/255, minimum MTU 1500 bytes
      Loading 1/255, Hops 4
```

- The end result of the configuration steps above is that only users belonging to the subnets 10.137.12.0 (in VRF Red) and 10.137.23.0 (in VRF Green) experience, or are provided with, two-way connectivity. The shared services IP prefix is distributed to all the Layer 3 devices where the VRF Red and Green are deployed; this would allow every user part of Red and Green VRF to be able to reach that shared subnet on R1. However, the return traffic is limited to the two subnets listed above, as a result of the application of the route-maps when redistributing the user prefixes into BGP.

  Even if the access to the shared resources has been defined as “unprotected”, it could still be possible to further restrict it by applying an ACL. For example, given the fact that the shared subnet is directly connected via VLAN 32, the following configuration would restrict connectivity only to the users in VRF Red and part of the 10.137.12.0 subnet.

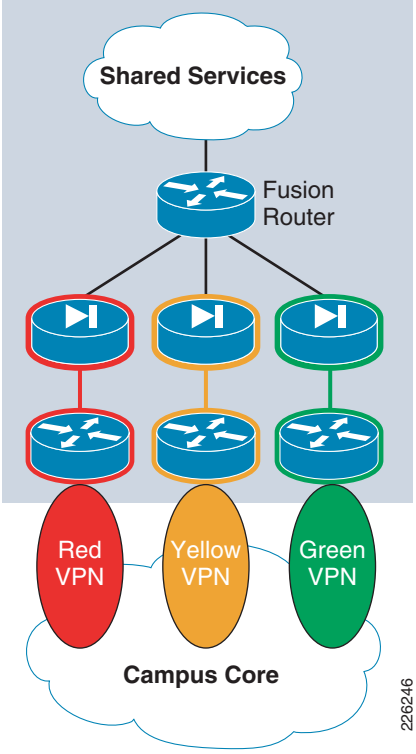
```
access-list 133 permit ip 10.137.12.0 0.0.0.255 any
!
interface Vlan32
 ip vrf forwarding Shared
 ip address 10.138.32.3 255.255.255.0
 ip access-group 133 out
```

  From an operational perspective this would obviously required more configuration and maintenance, but it would still benefit from the fact that this configuration is applied in a central location (compared to the traditional distributed ACL approach).

Deploying Protected Shared Services

When discussing the deployment of protected access to shared services (shown again in Figure 10), several variables of the design need to be taken into consideration and influence the specific deployed solution.

Figure 10 Protected Services Edge

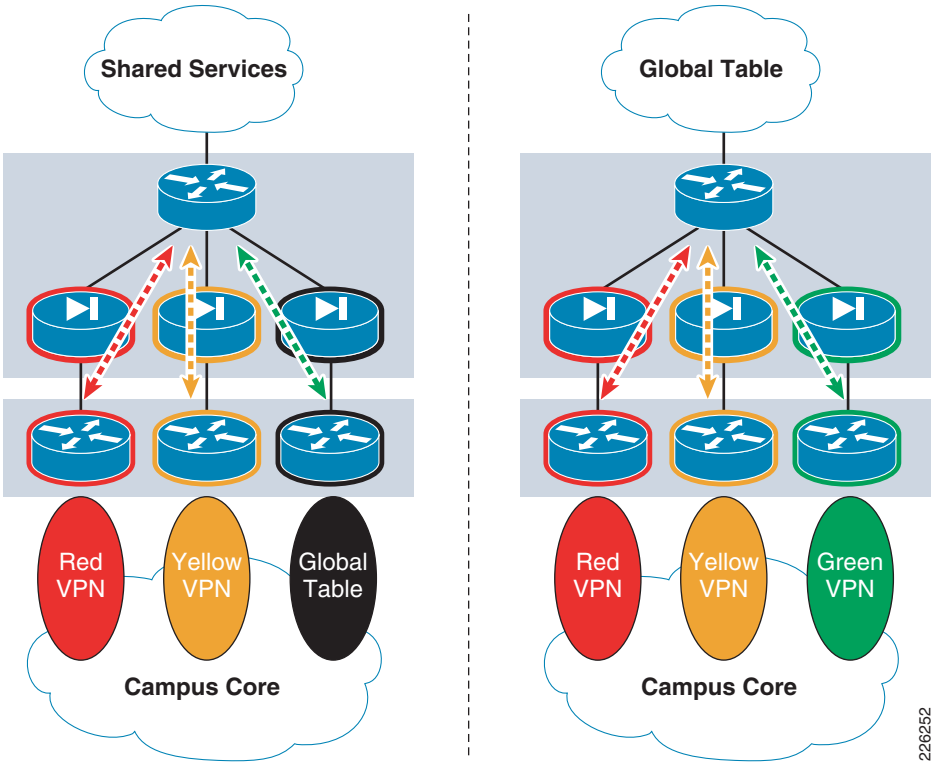


- The definition of shared services—In many deployments, the Internet represents a typical resource that needs to be shared between different virtual networks. In other scenarios, specific services (as for example a Call Manager) could represent the shared resource. In a few cases, the entire global table can also be seen as a shared service to which all the virtual networks may potentially connect.
- The security device front-ending each virtual network—This is usually a firewall device and it is common practice to leverage the virtualization capabilities of the Cisco Firewalls (FWSM, ASA, or PIX) in order to dedicate a virtual context (instead of a physical device) to each deployed logical partition.
- The working mode for the deployed firewall devices—Typically there are two options, transparent mode (where the firewall acts like a Layer 2 bridge device) or routed mode (the firewall becomes a routed hop).
- How connectivity to the non-virtualized portion of the network (the global table) could be achieved in this design. There are a couple of options that are commonly deployed:

  1. The global table is considered as another VPN (in fact can be usually considered the default VPN) and it is front-ended by its own security device.
  2. The global table is treated as a shared service; access to the global table from each VPN is subject to the policy enforcement provided by the services edge.

  These two options are shown in Figure 11.

Figure 11 Global Table Deployment Options



- How the fusion router functionality is deployed—The two most common options are dedicating a separate physical network device to perform this role or implementing the fusion router in the same switch as the services edge VPNs. This allows distinguishing two deployment scenarios—dual tier and single tier services edge models.

In the following sections we see how defining a fusion VRF may be a requirement for single tier deployments. In two tier implementations, the definition of a fusion VRF may not be required, given the fact that the fusion functionality is performed on an external device. However, in deployments where the shared services also need to be accessed from the global table, the use of a fusion VRF becomes mandatory. In order to keep the design generally applicable to all the scenarios, the rest of this document always discusses deployments leveraging the use of a fusion VRF.

The following sections presents some of the design variations that are possible based on the variables listed above, highlighting the pros and cons of each specific scenario. Two models are presented:

- Two tier
- Single tier

For each model, the following design options are available:

- Firewall contexts in transparent mode

  Routing peering options—EIGRP, OSPF and eBGP

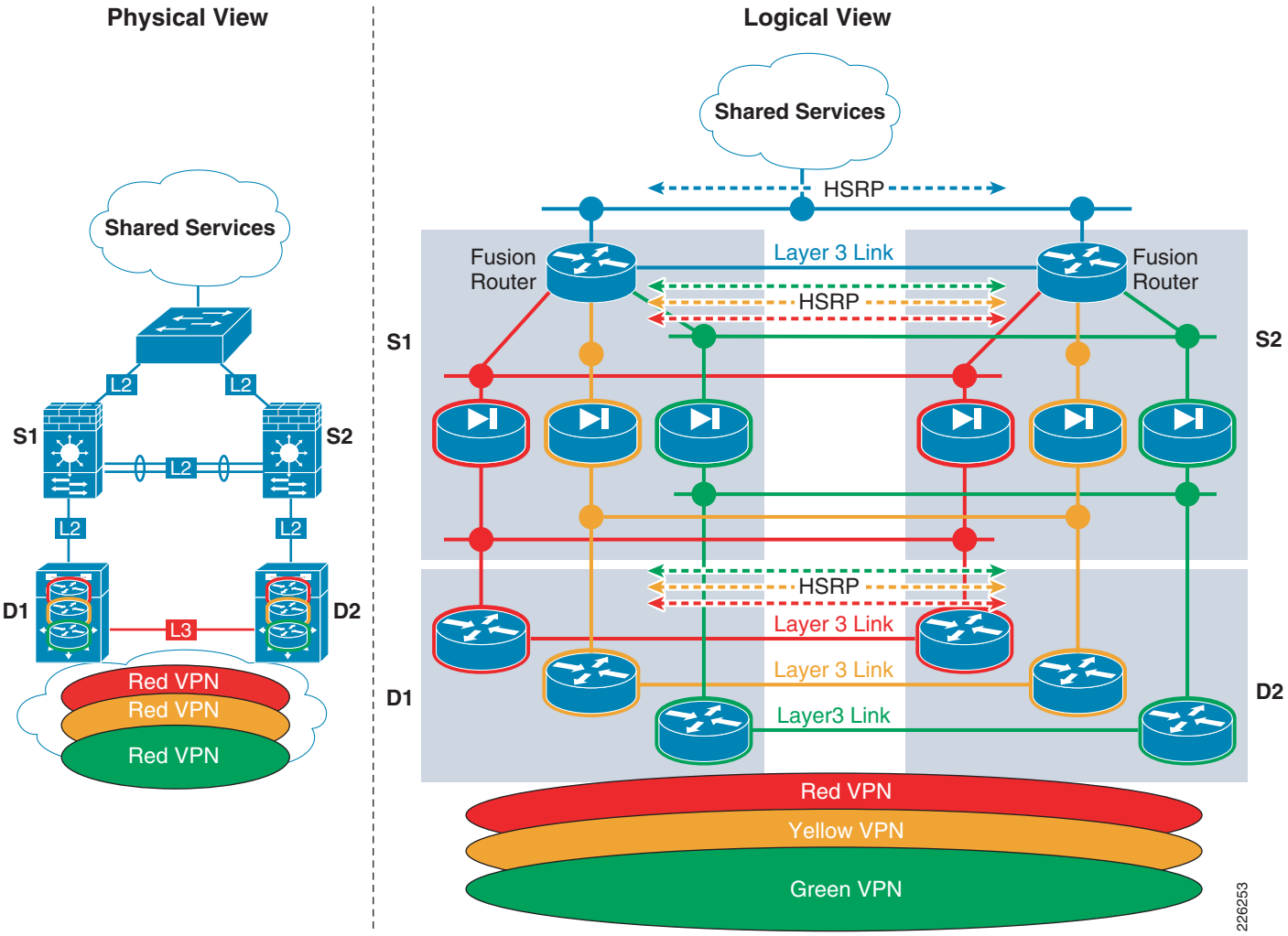
- Firewall contexts in routed mode

  Routing peering option—eBGP

Two Tier Implementation

The two tier service edge model is shown in Figure 12.

Figure 12 Two Tier Implementation Model



The D1 and D2 devices represent the distribution layer switches connected to the core of the network; as shown above, they are virtualized by defining the VRFs. The role they play in the virtualized network mostly depends on the deployed path isolation strategy:

- In MPLS VPN designs, these devices would be deployed as PEs.
- In a VRF-Lite End-to-End scenario, they would just connect to the core via virtualized links (either by using sub-interfaces or Layer 2 trunks with SVIs).
- If deploying VRF-Lite and GRE tunnels, these devices would most likely function as hubs aggregating the GRE tunnels originated from the remote spokes.

S1 and S2 are the devices functioning as fusion routers; in addition to that, they also perform firewall functionalities (usually by integrating a Firewall Services Module-FWSM). Because of these functionalities, S1 and S2 are named services switches in the context of this document.

The logical view in Figure 12 highlights how a separate firewall context is dedicated to front-ending each VPN. The different contexts function in active/standby mode (for the same VPN), but it is possible to alternate the active firewall on S1 and S2 for different contexts to achieve a better load-balancing of traffic.

Some additional considerations around the recommended two tier deployment:

- The fusion routers in the services switches and the VRFs defined in the distribution devices peer with a Layer 3 connection. For the fusion router this is achieved by dedicating a specific VLAN for this purpose; this VLAN is the carried over the Layer 2 port-channel trunk connecting the services switches. For the distribution VRFs, a physical routed link connects D1 and D2; this link is then virtualized (by leveraging sub-interfaces) in order to provide a Layer 3 connection to each VRF pair.

Note The use of a port-channel between the services switches is strongly recommended to offer a resilient connection between the two firewall modules.

- The inside interface for each firewall context is facing the fusion routers, whereas the outside interface is facing the distribution layer switches. This choice is dictated by the requirement for protecting the services deployed in the shared area.
- Both the firewall outside and inside VLANs are extended between the services switches (carried over the port-channel Layer 2 trunk), but that is not the case for the firewall inside VLANs. The design reason for this choice becomes clear when discussing the deployment of firewall in transparent mode.
- The distribution layer switches are connected to the services switches in a square topology leveraging Layer 2 trunks. The firewall outside VLANs are carried between the two tier devices over these trunk connections.
- The consequences of the design choices listed above is to create a loop-free topology between the distribution and services switches. Spanning tree is always recommended to be active (in order to prevent loops caused by configuration or cabling errors), but it is not responsible by design for blocking any link. This contributes to the resiliency of the overall services edge design.

Note that this type of deployment slightly differs from what usually recommended in data center deployments. For more information on data center design principles, refer to:

http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/dc_servchas/service-chassis_design.html

The following sections discuss two different deployments options for the two tier model, leveraging firewall contexts functioning in transparent or routed mode. The configuration samples and convergence results shown were achieved in a test bed with the following characteristics:

- Catalyst 6500 with Sup720 3BXL were deployed as services and distribution switches. The IOS release under test was 12.2(33)SXH4. Note that the use of this release is strongly recommended for the services edge deployment when leveraging eBGP as routing peering protocol between the distribution VRFs and the fusion router because of some fixes that were introduced (CSCsl39233 and CSCsu03167).
- Firewall Services Module (FWSM) deployed in multi-context mode running the 3.2(8) release. The use of this minimum release is required to have available an important fix (CSCsr11309) useful for firewall failover scenarios.

Note The documentation for the use of Cisco ASA for the same purpose is planned for a future release of this document.

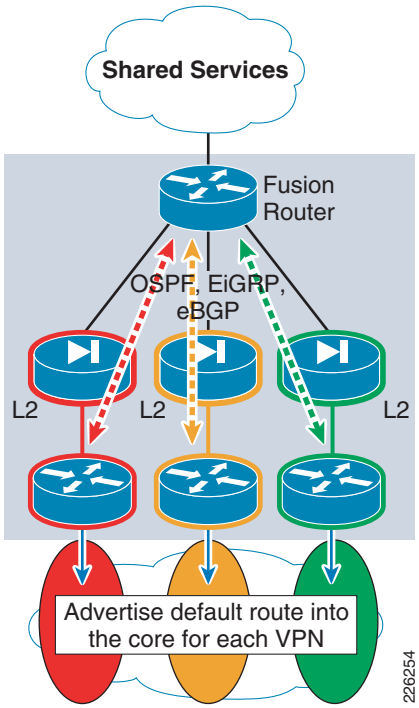
Firewall Contexts Deployed in Transparent Mode

Deploying the firewall contexts in transparent mode is a very common approach, since it allows inserting a firewall in the network without having to modify the IP addresses already in place. This is mainly because the firewall acts like a Layer 2 device, bridging traffic between the inside and outside interfaces.

Note When this document was written, only two interfaces can be defined when deploying a firewall context in transparent mode.

Another advantage when using transparent mode is that routing peering can be established between the VRFs defined on the bottom and the fusion router on top, as shown in Figure 13.

Figure 13 Routing Peering with Firewall Contexts in Transparent Mode

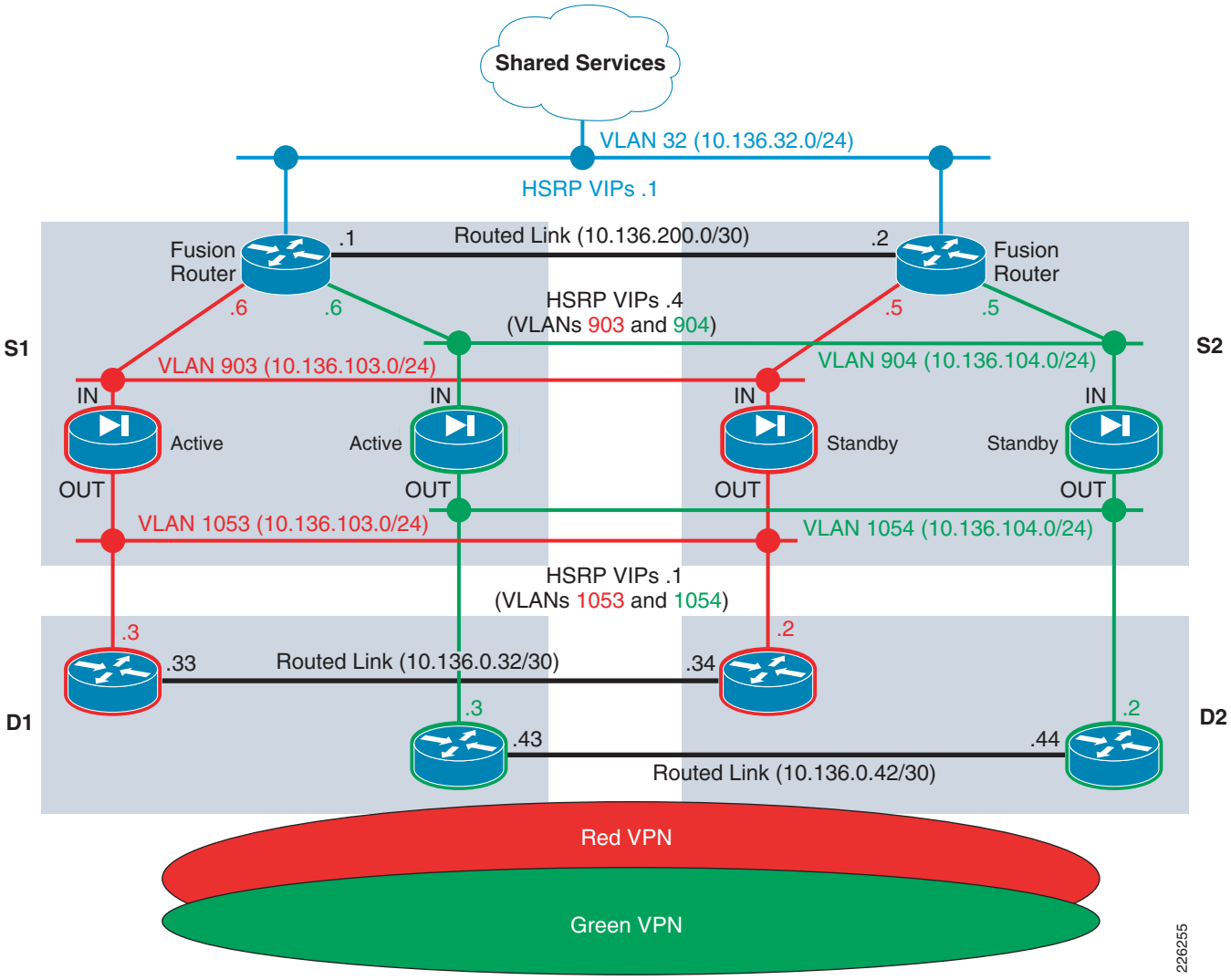


The type of routing protocol used for this functionality is usually dictated by the path isolation strategy deployed:

- For MPLS VPN deployments, the use of eBGP is advocated since iBGP is already in place between the PE devices exchanging VPN routes across the campus infrastructure.
- For VRF-Lite End-to-End (or VRF-Lite + GRE) deployments, typically the same IGP already used as control plane protocol inside each virtual network is also leveraged for this type of peering

The recommended deployment model for the services edge with firewall deployed in transparent mode is shown in Figure 14. This example shows the deployment of two VPNs (Red and Green). The specific VPN/IP subnet information shown in Figure 14 is used for the configuration samples shown in the remainder of this section.

Figure 14 Firewall Contexts in Transparent Mode Example



As shown in Figure 14, the Red and Green VPNs are front-ended by two firewall contexts deployed in transparent mode and hence bridging traffic from the VLANs deployed on the firewall outside and inside interfaces. Also, the firewall contexts inside S1 are active and the ones for the firewall in S2 are in standby mode. It is also possible to deploy an active-active model where the Red firewall context is active in S1 and the Green one is active in S2.

Some additional considerations around the recommended deployment model:

- HSRP is the First Hop Redundancy Protocol of choice in this case and it is leveraged to provide virtual gateway functionalities for the following interfaces:
  - Shared services subnet (VLAN 32)—This is under the assumption that the shared services are deployed in a subnet directly connected to the fusion devices.
  - Firewall inside subnets (VLAN 903 for the Red VPN and VLAN 904 for the Green VPN).
  - Firewall outside subnets (VLAN 1053 for the Red VPN and VLAN 1054 for the Green VPN).
- The fusion routers and the VRFs deployed in the distribution layer switches are also connected with a routed link. This is to provide a Layer 3 path to be used for re-routing traffic under specific failure conditions (as it is discussed later).

- The fusion router defines a separate interface (VLAN Interface in this case) to be dedicated to each campus VPN. In the example above, SVI 903 is leveraged to establish communication with the Red VPN (through the Red firewall context), whereas SVI 904 is used for the Green VPN.

When deploying the firewall in transparent mode, it is common best practice to define a specific ACL on each firewall context to allow BPDUs flow across. This is important in redundant scenarios (like the one depicted above) to be able to detect spanning tree loops that may be created when both the firewall contexts achieve an active state (because of the loss of keep-alive messages for example). The specific ACL (usually named an ethertype ACL) is shown below:

```
access-list BPDU ethertype permit bpdu
```

Since ethertype traffic is connectionless, it is then required to apply this ACL on both the inside and the outside interfaces:

```
access-group BPDU in interface outside-vrf-red
access-group BPDU in interface inside-vrf-red
```

Note Providing the entire firewall context configuration is out of the scope of this document. For specific information, see the data center design guides: http://www.cisco.com/en/US/netsol/ns743/networking_solutions_program_home.html

An interesting consequence of the configuration above is the fact that HSRP packets are also allowed through the firewall. This could represent an issue in our scenario because of the common subnet associated to the VLANs on the inside and outside of the firewall. In order for HSRP to function properly and exhibit the desired behavior, it is required to configure a different HSRP group for the services and distribution layer devices, as shown below.

Services switches (S1 and S2)

```
interface Vlan903  interface Vlan903
 description Firewall Inside Red VRF description Firewall Inside Red VRF
 ip vrf forwarding fusion ip vrf forwarding fusion
 ip address 10.136.103.6 255.255.255.0 ip address 10.136.103.5 255.255.255.0
 standby 1 ip 10.136.103.4 standby 1 ip 10.136.103.4
 standby 1 timers msec 250 msec 750 standby 1 timers msec 250 msec 750
 standby 1 priority 105
 standby 1 preempt delay minimum 180
```

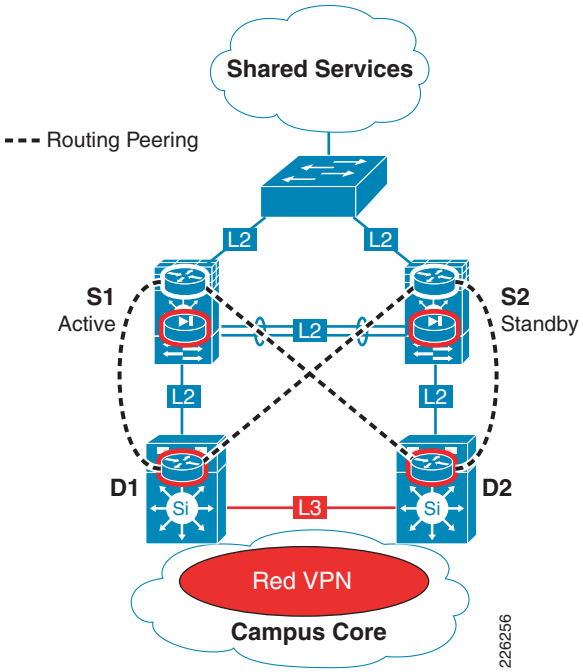
Distribution switches (D1 and D2)

```
interface Vlan1053interface Vlan1053
 description Firewall Outside Red VRF description Firewall Outside Red VRF
 ip vrf forwarding Red ip vrf forwarding Red
 ip address 10.136.103.3 255.255.255.0 ip address 10.136.103.2 255.255.255.0
 standby 2 ip 10.136.103.1 standby 2 ip 10.136.103.1
 standby 2 timers msec 250 msec 750 standby 2 timers msec 250 msec 750
 standby 2 priority 105
 standby 2 preempt delay minimum 180
```

The final result of this configuration is that S1 becomes the active HSRP device on the firewall inside subnet, whereas D1 takes the active HSRP role on the firewall outside subnet.

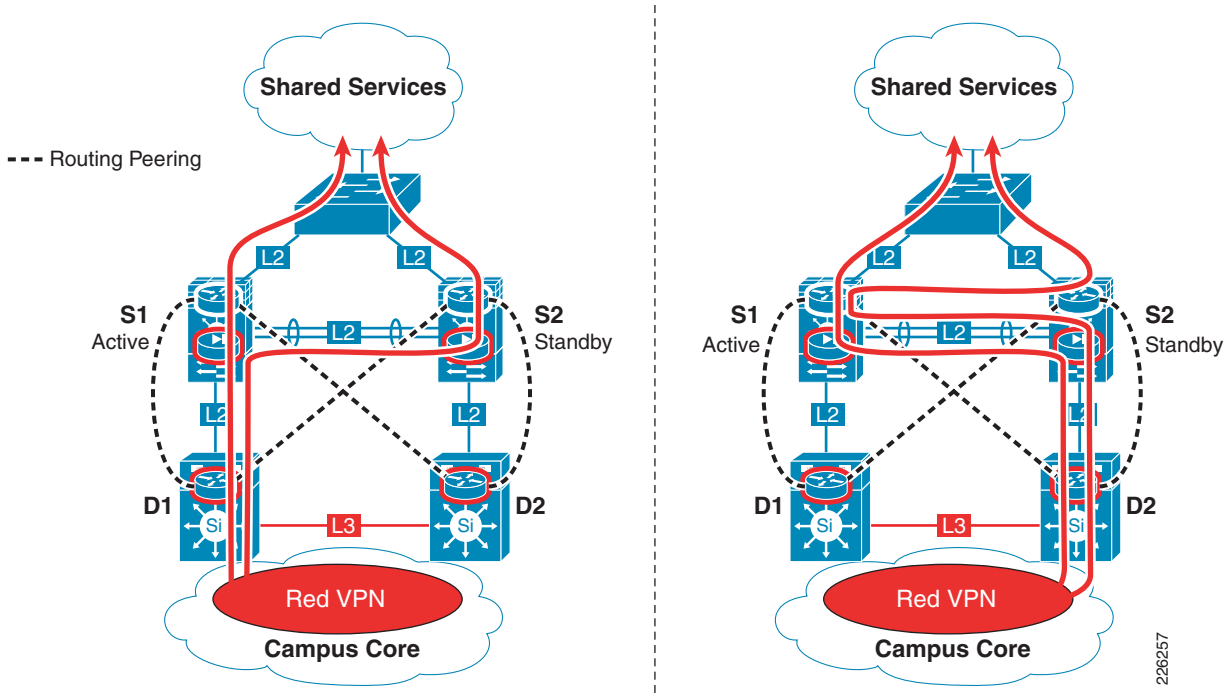
Assuming that the filters on each firewall context have been properly deployed so that the routing protocol of choice is allowed through (the configuration required for EIGRP, OSPF, and eBGP are discussed in the protocol specific sections later in the document), each VRF defined in the distribution switches peers with both fusion routers, as highlighted in Figure 15.

Figure 15 Full Mesh Routing Peering



Since the desire for each VPN is to steer the traffic to the services edge every time the destination is unknown in that specific routing domain, in a typical deployment the fusion routers just advertise a default route to each VRF defined in D1 and D2. At the same time, each VRF advertises to the fusion router the remote campus subnets. The consequence is that by default the traffic flows between the core and the shared services area follow the paths described in Figure 16.

Figure 16 Core to Services Traffic Flows

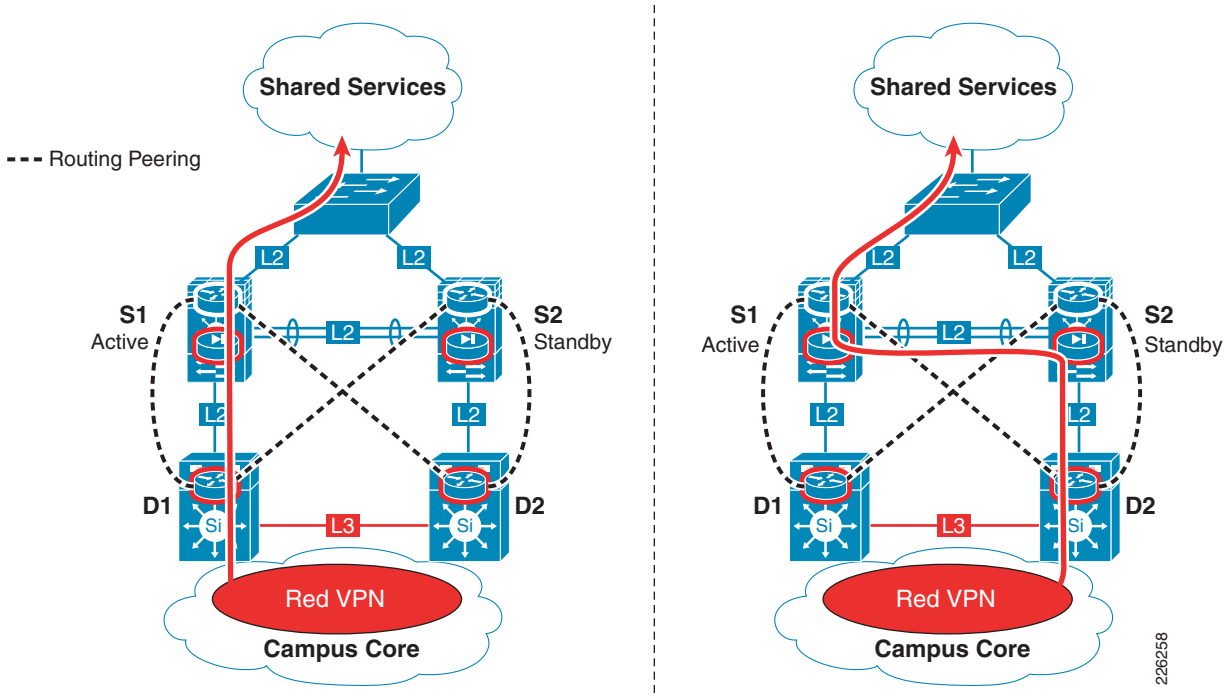


The reason behind this behavior is that by default each fusion router would advertise an equal cost default route to both VRFs. All the traffic flows are then required to go through the active firewall inside S1 independently from the fact that the next-hop is the fusion router inside S1 or in S2.

Note The same traffic flows are experienced if the more specific shared service subnet (instead that a generic default route) is advertised from the fusion router to the VRFs defined in the distribution layer switches.

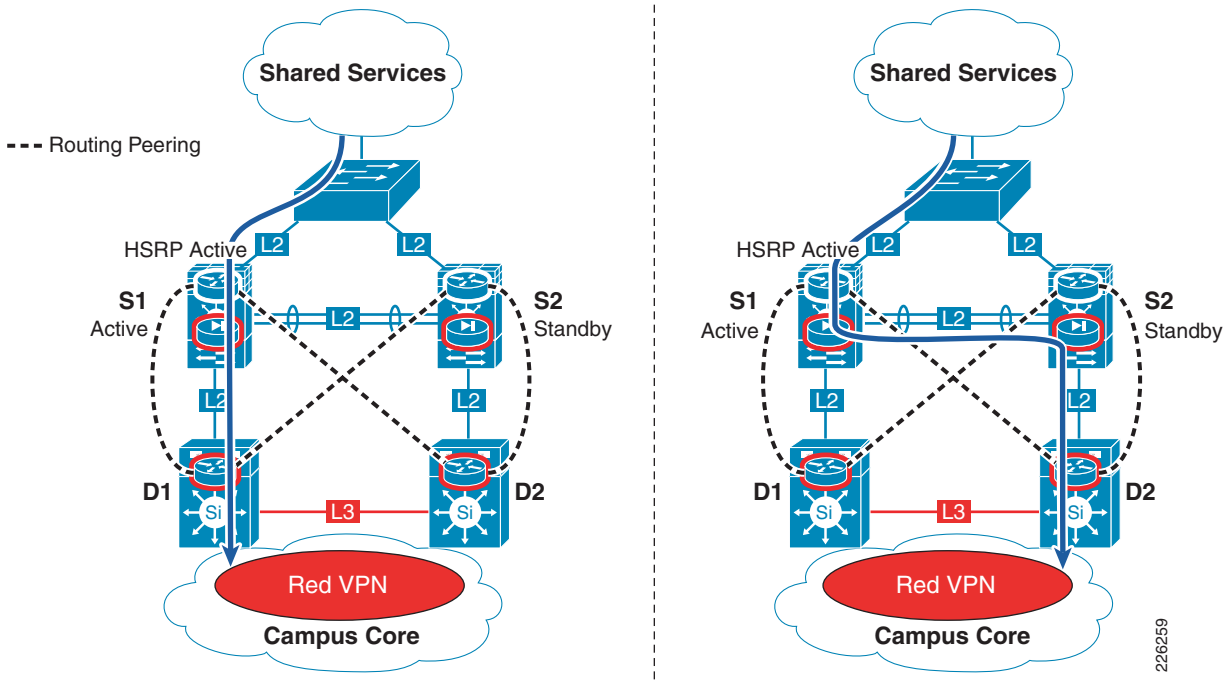
The flows shown in Figure 16 are sub-optimal and over utilize the transit link between the services switches. In order to optimize this scenario, the recommended solution is tuning the routing so that the fusion router inside S1 advertises a preferred metric for the default route to the distribution VRFs. The specifics on this tuning are discussed in the routing protocol specific sections below; the end result on the traffic flows is highlighted in Figure 17.

Figure 17 Optimizing Core to Services Traffic Flows



The flows in the different direction (shared services to campus core) are highlighted in Figure 18.

Figure 18 Shared Services to Core Traffic Flows



All the return traffic flows are initially sent to S1 because it is the HSRP active device on the shared services subnet (VLAN 32); this is assuming the shared services are deployed on a subnet directly connected to the fusion routers. At that point, 50% of the flows are sent to D1 (via the direct fiber link) and 50% are sent to D2 by leveraging the transit link between the services switches. This is because the Red VRFs defined in the distribution layer switches provide an equal cost path to the remote campus destinations. Note how the traffic flows are symmetrical to the ones shown in Figure 17.

In the following sections the recovery behaviors under different failure scenarios are discussed. The considerations around the traffic flows experienced after every failure are independent from the routing protocols deployed between the VRFs and the fusion routers. The specific recovery mechanisms and recovery times are discussed for each routing protocol in later sections.

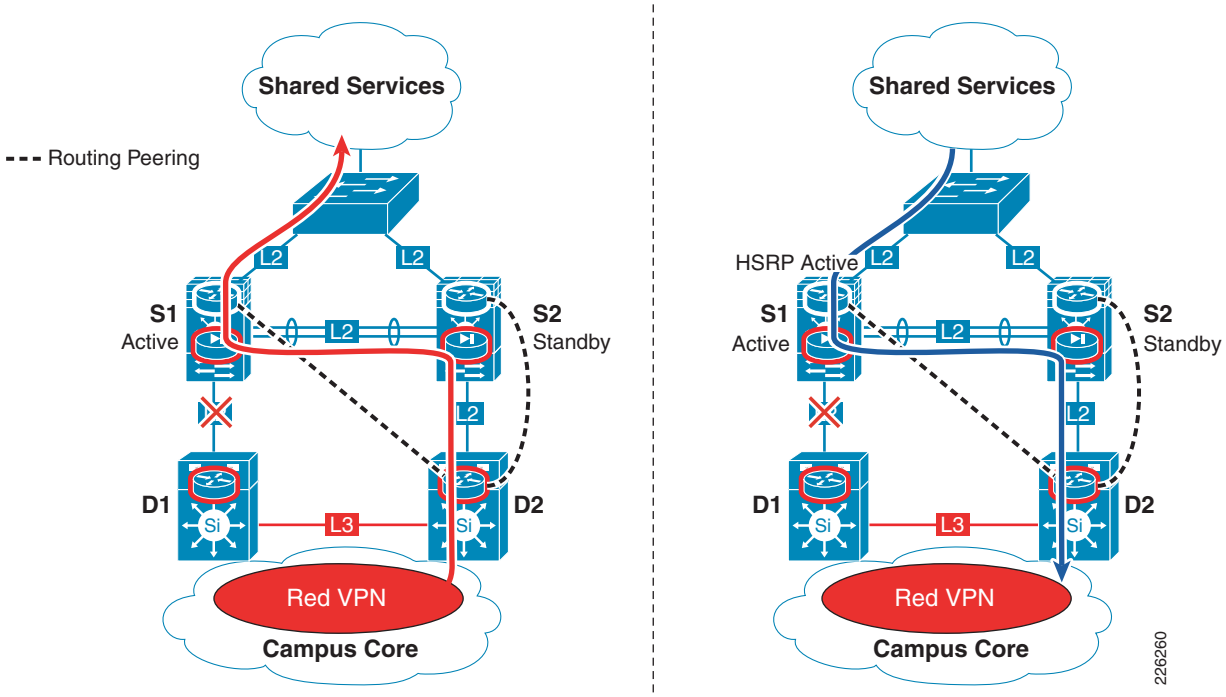
Note The redundancy discussion in this part of the document is focused on recovery mechanisms in a single site deployment. For redundant sites considerations, refer to Planning for Site Redundancy in a Virtualized Network.

Convergence Analysis

Fiber Failure between Distribution and Services Switches

Following the failure of the fiber connection between D1 and S1, the peering between the VRF in D1 and the fusion routers are removed since D1 no longer has a Layer 2 path active to the newly activated firewall module. The consequence on the traffic flows is highlighted in Figure 19.

Figure 19 Traffic Flows after Fiber Failure between D1 and S1



- Core to services flows—D1 stops receiving the default route (or the more specific route for the shared services subnet) from the fusion routers and starts receiving the same information from D2 via the routed link connecting the distribution switches. As a consequence, it starts advertising that information to the campus core with a worse metric than D2, so that all the traffic from the core

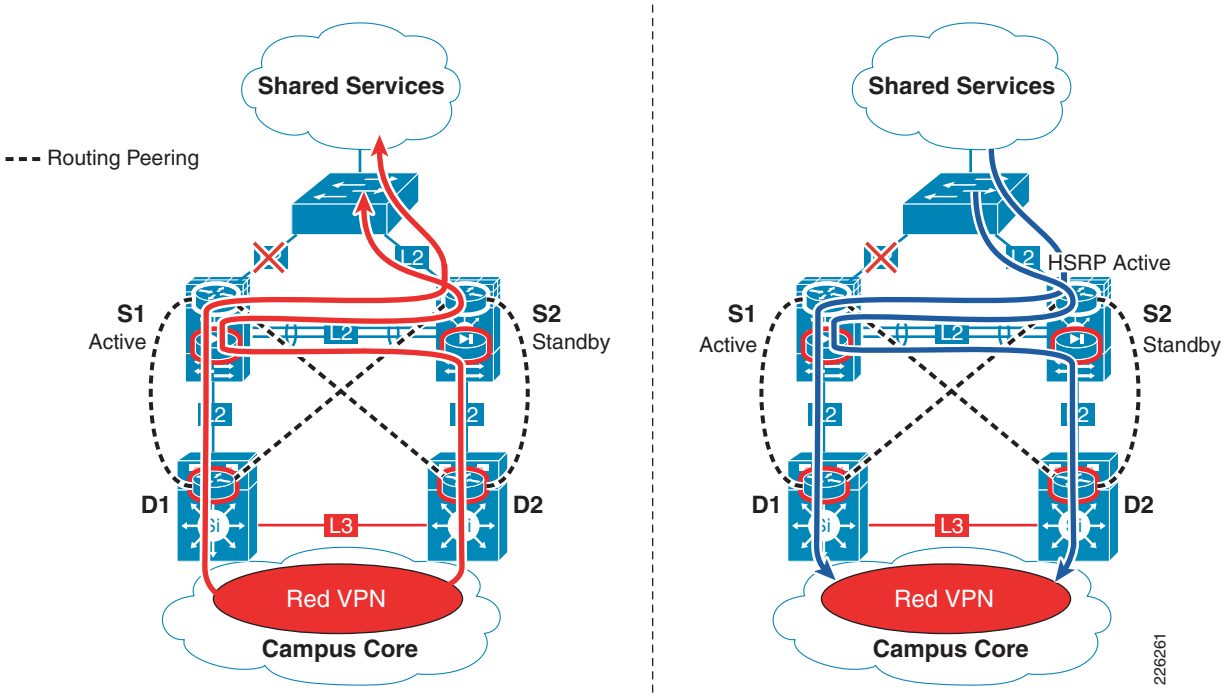
directed to the shared services area is now sent to D2. The convergence experienced in this case is based on how fast D1 detects the fiber failure; assuming that the link failure detection works, this would be very fast and cause sub-second outage.

- Services to core flows—Independently from the fiber failure, the fusion router in S1 maintains the active HSRP role. However, until the routing peering between the fusion routers and the VRF in D1 goes down, the fusion router would try to send VRF 50% of the flows destined for remote campus locations in that VRF to D1, causing a black-hole for these traffic flows. Once the peering is removed, the fusion router in S1 starts using only the VRF in D2 as the next hop to the remote campus destinations. This implies that these flows need to reach D2 via the transit link between the services switches. From a convergence perspective, the main factor dictating the length of the outage here is the time required for the fusion router in S1 to tear down the routing peering with the VRF in D1; this is usually directly related to the configured routing protocol hold-time.

Fiber Failure between Services Switches and Shared Services Area

The failure of the fiber link connecting the services switch S1 to the shared services area does not cause any change in the routing peering between VRFs and fusion routers. The traffic flows become the ones depicted in Figure 20.

Figure 20 Traffic Flows after Fiber Failure between S1 and Shared Services Area



- Core to services flows—Depending on how the fusion router in S1 learns the default route to advertise to the VRFs in the distribution layer, two different scenarios for the traffic flows are possible:
  - If the static route is not learned from a next-hop device but locally generated (this is achieved in different ways depending on the deployed routing protocol, as it is discussed later), then the fusion router in S1 keeps advertising the default route even after the fiber failure. This means that traffic needs to be re-routed across the routed link shown in Figure 14 (10.136.200.0/24 subnet) in order to be delivered to the shared services.

  - If the static route is normally learned from a next-hop device in the shared services area, S1 stops learning this information and stops advertising it to the VRFs. However the active firewall remains inside S1, causing the traffic flows to look exactly the same as the scenario in the previous bullet point.

- Services to core flows—The fusion router in S2 becomes the HSRP active device on the shared subnet (HSRP messages were exchanged via the connections to the switch in the services area) and has to send all the traffic flows across the transit link to S1 since the active firewall is in that device. The traffic flows look identical in the opposite direction. From a convergence perspective, the outage is dictated by the time required for the fusion router in S2 to become the HSRP active device on the shared services subnet (VLAN 32). In order to minimize this value, it is recommended to configure sub-second HRSP timers as shown below:

  S1

```
interface Vlan32
 description Shared Services
 ip vrf forwarding fusion
 ip address 10.136.32.3 255.255.255.0
 standby 1 ip 10.136.32.1
 standby 1 timers msec 250 msec 750
 standby 1 priority 105
 standby 1 preempt delay minimum 180
```

  S2

```
interface Vlan32
 description Shared Services
 ip vrf forwarding fusion
 ip address 10.136.32.2 255.255.255.0
 ip flow ingress
 standby 1 ip 10.136.32.1
 standby 1 timers msec 250 msec 750
```

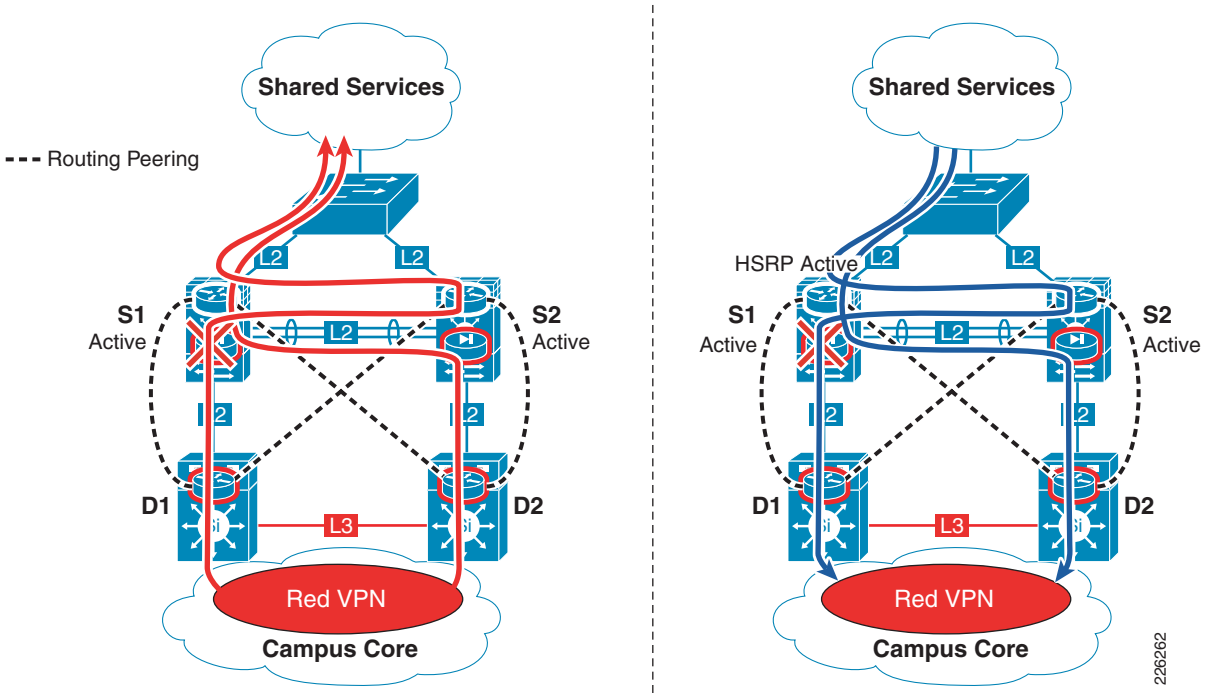
  The use of subsecond HSRP timers is recommended for deployments with 150 VLANs or less. For more information, refer to the campus design guides: http://www.cisco.com/en/US/netsol/ns815/networking_solutions_program_home.html

Note In order to avoid the suboptimal path across the transit link shown in Figure 20, when possible it is recommended to deploy a port-channel between each services switch and the shared services area.

Active Firewall Module Failure

After the failure of the firewall module in S1, the VRFs and the fusion routers maintain their routing peering through the newly activated firewall inside S2. The consequent traffic flows are shown in Figure 21.

Figure 21 Traffic Flows after Firewall Failure



- Core to services flows—D1 and D2 keep learning the best route through the fusion router in S1. As a consequence, 50% of the traffic flows experience the sub-optimal path shown in Figure 21, crossing the transit link between the services switches twice. The overall outage before these flows are established is mainly affected by the time required for the firewall in S2 to become active. This is true under the assumption that the configured routing protocol hold-time is higher than this time; if not, on top of the firewall failover time, we need to consider the time required to re-establish the routing protocol peering and advertise the routing information (more considerations can be found in the protocol specific sections below). Note that the time required by the firewall module in S2 to detect the failure of the other firewall and become the active device depends on the configured hold-time between the firewalls; FWSM does not allow to configure hold-time values lower than three seconds, as shown below:

```
FWSM(config)# failover polltime unit msec 500 holdtime ?
configure mode commands/options:
  <3-45>  Hold time in seconds, default is 3 X poll time but minimum is 3
          Seconds
```

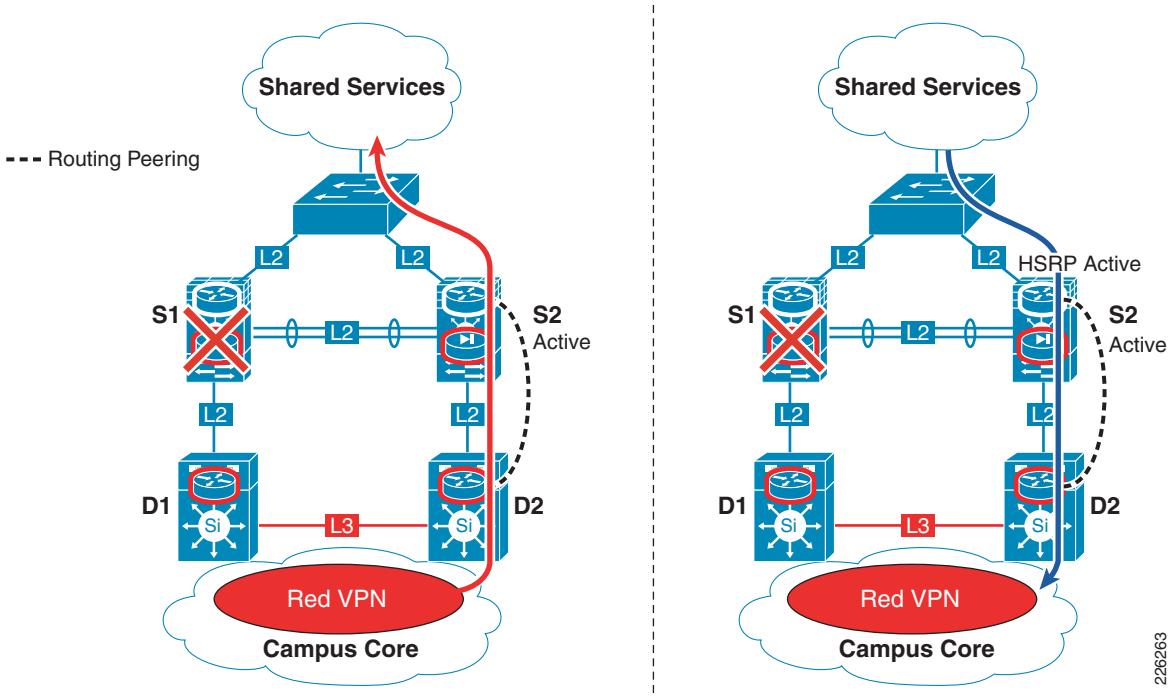
- Services to core flows—Independently of the firewall failure, S1 remains the active HSRP device, so all the traffic from the shared services area is directed to it. S1 still learns the remote campus destinations via both the VRFs in the distribution layer, so 50% of the traffic follows the same sub-optimal path discussed above. For the recovery time experienced in this direction, the same considerations made in the previous bullet point are still valid and convergence depends on how fast the firewall can failover and if the peering between devices is lost or not in the meantime.

Design recommendation—Since the suboptimal traffic path shown in Figure 21 happens when the firewall module is active inside S2, it is recommended to configure firewall preemption so that the firewall inside S1 (when up and running) always take the active role.

Services Switch Failure

This failure scenario is highlighted in Figure 22.

Figure 22 Traffic Flows after Services Switch Failure



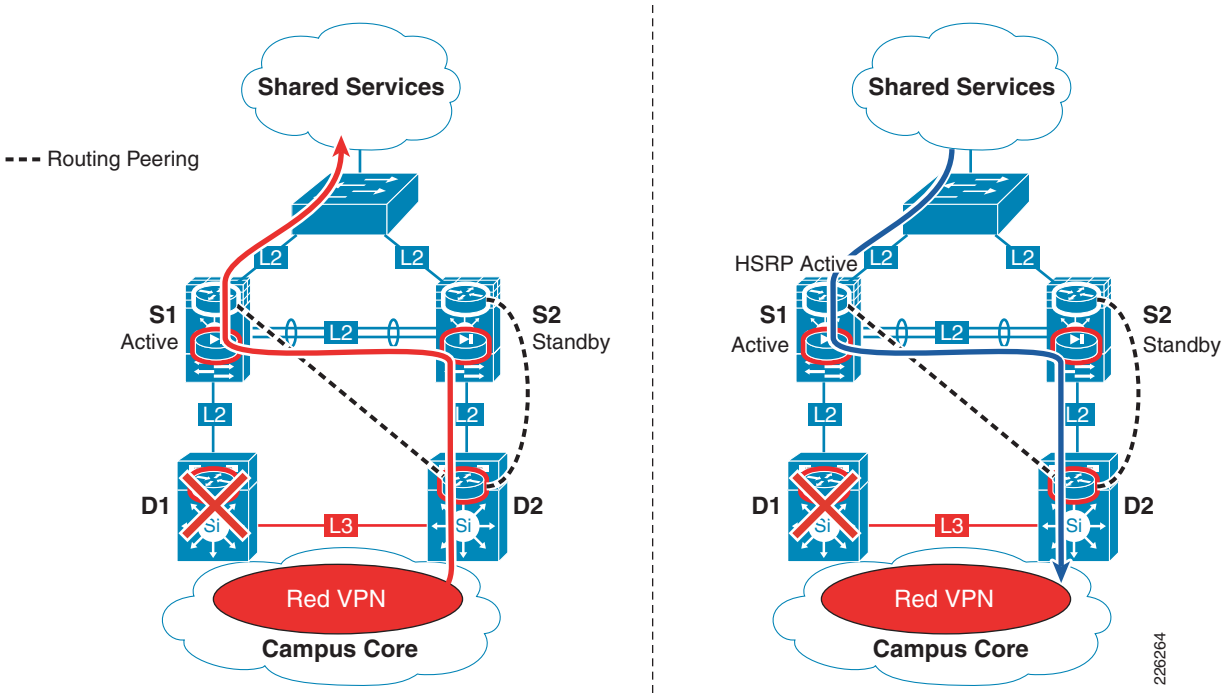
Because of the failure of the entire services switch S1, the peering between the VRFs and the fusion router in S1 is removed. Also, since D1 no longer has an active Layer 2 path to the newly activated firewall module, only the VRF in D2 is capable of establishing a routing peering with the fusion router in S2. The consequences on the traffic flows are:

- Core to services flows—The link between D2 and S2 is the only available path from the core to the shared services. From a convergence perspective, the outage is dictated by how fast the firewall becomes active, allowing the successive establishment of routing peering. We can expect the entity of this outage to be the same as the one discussed in the section for the firewall failure scenario.
- Services to core flows—The fusion router only has a valid path to the remote campus destinations in the VRF via D2. S2 becomes the HSRP active device because of the failure of the S1 switch. The recovery mechanism is identical to the one described in the previous bullet point.

Distribution Switch Failure

The last relevant scenario is the failure of the distribution switch, as shown in Figure 23.

Figure 23 Traffic Flows after Distribution Switch Failure



From a traffic flow and routing peering perspective, this scenario is very similar to the one relative to the fiber failure between D1 and S1.

- Core to services flows—Since the entire D1 switch fails, the core devices remove the valid path through it and re-route all the traffic via D2. Once again, assuming the link failure detection works, this is an ECMP re-route that causes only sub-second outages.
- Services to core flows—The fusion router does not have any way to detect that the distribution switch fails (since there is the firewall between them). As a consequence, it keeps sending traffic to the VRF in D1 until the routing peering is removed; this is the main mechanism dictating the convergence in this case.

The following sections discuss the specific design and configuration considerations for different routing protocol deployments. In each case, this includes convergence results achieved for all the failure scenarios previously discussed.

Use of EIGRP between VRFs and Fusion Router

The first option available for peering between the VRFs defined on the distribution switches (D1 and D2) and the fusion routers is to leverage EIGRP. This is specifically recommended for VRF-Lite End-to-End deployments that leverage EIGRP in the context of each defined virtual network.

The required configuration steps are:

1. Allow establishment of the EIGRP adjacencies across the firewall context—The assumption is that EIGRP is already enabled for each VRF in the distribution layer switches, so it is now required to enable it on the services switches. This is usually the case when VRF-Lite End-to-End is leveraged to provide cross-campus connectivity for each VPN.

   The required configuration is shown below (valid for both S1 and S2):

```
router eigrp 100
!
```

```
address-family ipv4 vrf fusion
  network 10.0.0.0
  no auto-summary
  autonomous-system 100
 exit-address-family
```

Note For a more detailed discussion of deployment of EIGRP in a VRF-Lite environment, see: http://www.cisco.com/en/US/docs/solutions/Enterprise/Network_Virtualization/PathIsol.html

Since traffic recovery in different failure scenarios is affected by the configured hold-time, it is recommended to tune the EIGRP timers to lower the hello and hold-time timers to the minimum values. This can be done with the configuration shown below that needs to be applied to SVIs 903 on the services switches and SVIs 1053 on the distribution switches.

```
interface Vlan1053
 ip vrf forwarding Red
 ip address 10.136.103.3 255.255.255.0
 ip hello-interval eigrp 100 1
 ip hold-time eigrp 100 3
```

Note that the EIGRP protocol is automatically allowed through the firewall running in transparent mode once the specific ethertype below is applied to the inside and outside interfaces:

```
access-list BPDU ethertype permit bpdu
```

As a consequence, the EIGRP adjacencies are established between the fusion router in S1 and the VRFs in the distribution layer switches, as highlighted below:

D1

```
D1#sh ip eigrp vrf Red neighbors
IP-EIGRP neighbors for process 100
H   Address                 Interface       Hold Uptime   SRTT   RTO  Q  Seq
                                            (sec)         (ms)       Cnt Num
2   10.136.103.2            Vl1053             2 00:04:06    4   200  0  41751490
1   10.136.103.5            Vl1053             2 00:04:06    6   200  0  50
0   10.136.103.6            Vl1053             2 00:04:06    3   200  0  114
5   10.136.0.33             Te1/3.532         13 15:44:26    1   200  0  41751489
4   10.122.35.34            Te1/1.632          7 15:45:08    1   200  0  4926138
3   10.122.35.38            Te1/2.732          2 15:45:09    1   200  0  51781844
```

D2

```
D2#sh ip eigrp vrf Red neighbors
IP-EIGRP neighbors for process 100
H   Address                 Interface       Hold Uptime   SRTT   RTO  Q  Seq
                                            (sec)         (ms)       Cnt Num
1   10.136.103.3            Vl1053             2 00:05:39   10   200  0  1254
5   10.136.0.32             Te1/3.532         12 15:45:59    1   200  0  1255
4   10.122.35.36            Te1/1.732          7 15:46:02    1   200  0  4926137
3   10.122.35.40            Te1/2.632          2 15:46:03    1   200  0  51781842
2   10.136.103.6            Vl1053             2 15:46:04    1   200  0  114
0   10.136.103.5            Vl1053             2 15:46:04    1   200  0  50
```

The Red VRFs peer with each other and with the fusion routers in S1 and S2 via SVI 1053. They also peer to each other and to the devices deployed in the core via routed links deployed defining sub-interfaces of Ten1/1, Ten1/2, and Ten1/3.

S1

```
S1#sh ip eigrp vrf fusion neighbors
IP-EIGRP neighbors for process 100
```

```
H   Address                 Interface       Hold Uptime   SRTT   RTO  Q  Seq
                                            (sec)         (ms)       Cnt Num
3   10.136.200.2            Vl200             13 00:00:34    1   200  0  61
0   10.136.103.3            Vl903              2 00:13:51    4   200  0  1254
1   10.136.103.2            Vl903              2 15:54:16    1   200  0  41751490
2   10.136.103.5            Vl903              2 15:58:18    1   200  0  62
```

S2

```
S2#sh ip eigrp vrf fusion neighbors
IP-EIGRP neighbors for process 100
H   Address                 Interface       Hold Uptime   SRTT   RTO  Q  Seq
                                            (sec)         (ms)       Cnt Num
3   10.136.200.1            Vl200             11 00:01:16    1   200  0  124
0   10.136.103.3            Vl903              2 00:14:34    4   200  0  1258
2   10.136.103.2            Vl903              2 15:54:58    1   200  0  41751493
1   10.136.103.6            Vl903              2 15:59:00    1   200  0  122
```

The fusion routers peer with the VRFs defined on the distribution layer devices D1 and D2 and with themselves on VLAN 903. A separate SVI 200 is also defined for establishing a Layer 3 peering between the services switches. This peering may be used for re-routing traffic from the shared services directed to the campus core under specific failure scenarios.

2. Configure the fusion router to advertise a default route into each VPN—As previously mentioned, the goal is for each defined VPN to steer traffic to this central location every time the destination is unknown in the context of a specific VPN. The assumption here is that a default route is present in the fusion router routing table. This could be because the fusion router learns it from a next-hop router (this for example would be the case for Internet edge deployments) or because it is statically defined. In order to optimize the traffic flows it is required to tune the routing protocol so that by design the fusion router inside S1 advertises a better metric for the default route. The configuration to achieve the desired behavior is shown below (for both S1 and S2 devices):

S1

```
router eigrp 100
 !
 address-family ipv4 vrf fusion
  redistribute static
  network 10.0.0.0
  default-metric 100000 100 255 1 1500
  distribute-list Default out Vlan903
  no auto-summary
  autonomous-system 100
  eigrp router-id 10.136.200.1
 exit-address-family
!
ip access-list standard Default
 permit 0.0.0.0
```

S2

```
router eigrp 100
 offset-list Default out 1000 Vlan903
 !
 address-family ipv4 vrf fusion
  redistribute static
  network 10.0.0.0
  default-metric 100000 100 255 1 1500
  distribute-list Default out Vlan903
  no auto-summary
  autonomous-system 100
  eigrp router-id 10.136.200.2
 exit-address-family
```

```
!
ip access-list standard Default
 permit 0.0.0.0
```

Some considerations valid for the example above:

- The **redistribute static** command is used to inject the default route into the routing table. This would not be required in scenarios where S1 and S2 learn the default from a next-hop device.
- An offset-list is configured on S2 to increase the metric of the routes advertised out of SVI 903, so that the distribution VRFs would always prefer the routes advertised by the fusion router in S1. Note how the **offset-list** command is listed under the global EIGRP configuration space, despite the fact that it applies to an interface mapped to the fusion VRF.
- A distribute-list is applied to filter the routes that S1 and S2 advertise to the VRFs in the distribution switches; only the default route is allowed in this specific example. The use of a distribute-list is strongly recommended; without it, the fusion router would also advertise to the Red VRFs the routing information for the other VPNs (remember that the fusion router has a complete cross-VPN view of all the campus prefixes). Even if this does not establish by default inter-VPN communications (a specific policy must be configured on each firewall context for that to happen), it is not a desirable behavior.

  The end result is that both D1 and D2 learn the default route from S1, as highlighted below:

  D1

```
D1#show ip route vrf Red supernets-only
Routing Table: Red
Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route
Gateway of last resort is 10.136.103.6 to network 0.0.0.0
D*EX 0.0.0.0/0 [170/51456] via 10.136.103.6, 1d00h, Vlan1053
```

  D2

```
D2#show ip route vrf Red supernets-only
Routing Table: Red
Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route

Gateway of last resort is 10.136.103.6 to network 0.0.0.0
D*EX 0.0.0.0/0 [170/51456] via 10.136.103.6, 5d22h, Vlan1053
```

The resulting traffic flows from the core to the shared services areas were shown in Figure 17.

EIGRP Convergence Results

The convergence results achieved in the different failure scenarios previously discussed are summarized below:

Test 1 (fiber failure between S1 and D1)

- Core to services flows: <1 sec.

This value is determined by how fast the D1 device detects the failure of the fiber link connecting to S1; this causes the SVI 1053 to go into a DOWN state, thus removing the default route learned via that interface from the routing table. This event is usually driven by the link failure detection mechanism and is therefore very fast (sub-second convergence).

- Services to core flows: ~3 sec.

  This value is determined by how fast the fusion router can remove from the routing table the routes learned from D1. Traffic is black-holed until that happens, since as mentioned above the SVI is now in a DOWN state on D1. The main factor for this detection is the configured EIGRP hold-time, which justify the recommendation to tune it aggressively to three seconds. This behavior is shown in the output below captured on the fusion router in S1:

```
Dec  9 16:13:10.414 EST: %LINEPROTO-5-UPDOWN: Line protocol on Interface
GigabitEthernet2/23, changed state to down
Dec  9 16:13:10.426 EST: %LINK-3-UPDOWN: Interface GigabitEthernet2/23, changed state
to down
Dec  9 16:13:10.418 EST: %LINEPROTO-SP-5-UPDOWN: Line protocol on Interface
GigabitEthernet2/23, changed state to down
Dec  9 16:13:10.422 EST: %LINK-SP-3-UPDOWN: Interface GigabitEthernet2/23, changed
state to down
Dec  9 16:13:12.894 EST: %DUAL-5-NBRCHANGE: IP-EIGRP(1) 100: Neighbor 10.136.103.3
(Vlan903) is down: holding time expired
Dec  9 16:13:12.894 EST: RT(fusion): delete route to 10.137.32.0 via 10.136.103.3,
eigrp metric [90/3840]
```

  As noted above, the route to the remote destination (10.137.32.0) is removed after the EIGRP peering between the fusion router in S1 and the VRF in D1 is removed because of the expiration of the EIGRP hold-time.

Test 2 (fiber failure between S1 and shared services area)

- Core to services flows: <1 sec.

  All the traffic flows directed to the fusion router in S1 need to be re-routed across the transit link toward S2. The re-routing usually causes sub-second outage.

- Services to core flows: <1 sec.

  Traffic originated in the shared services area is black-holed until the fusion router in S2 becomes the active HSRP device on the shared services subnet (VLAN 32). Configuring sub-second HSRP timers lets us keep this convergence event sub-second.

Test 3 (firewall failure)

- Core to services flows: ~4-5 sec.
- Services to core flows: ~4-5 sec.

  Two main components contribute to the traffic outages shown above:

  - Time required for the firewall in S2 to take on the active role—This cannot currently be configured lower than three seconds in a FWSM deployment.
  - Assuming the EIGRP timers are set aggressively, it is possible that the adjacencies will be reset (given the three seconds required for the new firewall to start passing traffic). The time required to re-establishing the EIGRP peering needs to be taken into consideration and that leads to the overall 4-5 second outage. In order to avoid the EIGRP adjacencies from being reset, the EIGRP timers could be set less aggressively. Keep in mind that this would affect the recovery times in failure scenarios (as, for example, the fiber failure one) where the EIGRP hold-time is a main factor in determining the overall traffic outage.

Test 4 (services switch failure)

- Core to services flows: ~4-5 sec.
- Services to core flows: ~4-5 sec.

The same consideration made in the firewall failure scenario applies here as well.

Test 5 (distribution switch failure)

- Core to services flows: <1 sec.

  The recovery mechanism in this case is ECMP from the core devices once the distribution switch fails. This is very fast, assuming that the link detection mechanism works properly on the core devices directly connected to the failing switch.

- Services to core flows: ~3 sec.

  The recovery time is dependant on the configure EIGRP hold-time because the fusion router on S1 (active HSRP device) keeps sending traffic to the VRF inside the failed switch until it removes the EIGRP peering, similar to the test 1 scenario above.

Use of OSPF between VRFs and Fusion Router

The use of OSPF for peering between the VRFs defined on the distribution switches (D1 and D2) and the fusion router is specifically recommended for VRF-Lite End-to-End deployments that leverage OSPF in the context of each defined virtual network.

The following are the required configuration steps:

1. Allow establishment of the OSPF adjacencies across the firewall context—The assumption is that OSPF is already enabled in the context of each VRF in the distribution layer switches, so it is now required to enable it on the services switches. The required configuration is shown below (valid for both S1 and S2, with the exception of the router-id value):

```
router ospf 100 vrf fusion
 router-id 10.136.100.1
 log-adjacency-changes
 auto-cost reference-bandwidth 10000
 capability vrf-lite
 timers throttle spf 10 100 5000
 timers throttle lsa all 10 100 5000
 timers lsa arrival 80
 passive-interface default
 no passive-interface Vlan200
 no passive-interface Vlan903
 network 10.136.0.0 0.0.255.255 area 136
```

Note For a more detailed discussion around deployment of OSPF in a VRF-Lite environment, see: http://www.cisco.com/en/US/docs/solutions/Enterprise/Network_Virtualization/PathIsol.html

Since traffic recovery in different failure scenarios is affected by the configured hold-time, it is recommended to tune the OSPF timers to lower the hello and hold-time timers. This can be done with the configuration shown below that needs to be applied to SVIs 903 on the services switches and SVIs 1053 on the distribution switches.

```
interface Vlan1053
 ip vrf forwarding Red
 ip address 10.136.103.3 255.255.255.0
 ip ospf hello-interval 1
```

The configuration of the hello-interval also implicitly set the dead timers to a value four times greater, as highlighted below:

```
D1#sh ip ospf 4 interface vlan1054
Vlan1054 is up, line protocol is up
  <snip>
  Timer intervals configured, Hello 1, Dead 4, Wait 4, Retransmit 5
```

Note Configuring sub-second OSPF timers is not recommended.

Note that, unlike the EIGRP protocol, OSPF packets are not allowed through the firewall running in transparent mode by the ethertype ACL. A specific ACE is required to allow OSPF through the firewall, as shown below:

```
access-list <ACL_NAME> extended permit ospf any any
```

As a consequence of the previous configuration steps, the OSPF adjacencies are established between the fusion router in S1 and the VRFs in the distribution layer switches, as highlighted below:

D1

```
D1 #sh ip ospf 4 neighbor
Neighbor ID     Pri   State           Dead Time   Address         Interface
10.122.233.4      1   FULL/BDR        00:00:03    10.122.35.34
TenGigabitEthernet1/1.632
10.136.4.2        1   FULL/BDR        00:00:03    10.136.103.2    Vlan1053
10.136.100.2      1   FULL/DR         00:00:03    10.136.103.5    Vlan1053
10.136.4.2        1   FULL/DR         00:00:03    10.136.0.33
TenGigabitEthernet1/3.532
10.122.233.5      1   FULL/DR         00:00:03    10.122.35.38
TenGigabitEthernet1/2.732
```

D2

```
D2 #sh ip ospf 4 neighbor
Neighbor ID     Pri   State           Dead Time   Address         Interface
10.122.233.4      1   FULL/BDR        00:00:03    10.122.35.36
TenGigabitEthernet1/1.732
10.136.4.1        1   FULL/DROTHER    00:00:03    10.136.103.3    Vlan1053
10.136.100.2      1   FULL/DR         00:00:03    10.136.103.5    Vlan1053
10.136.4.1        1   FULL/BDR        00:00:03    10.136.0.32
TenGigabitEthernet1/3.532
10.122.233.5      1   FULL/DR         00:00:03    10.122.35.38
TenGigabitEthernet1/2.632
```

Note how the VRFs peer with each other and with the fusion router in S1 via SVI 1053. They also peer to each other and to the devices deployed in the core (via sub-interfaces of Ten1/1, Ten1/2, and Ten1/3).

S1

```
S1#sh ip ospf 100 neighbor
Neighbor ID     Pri   State           Dead Time   Address         Interface
10.136.100.2      1   FULL/BDR        00:00:03    10.136.200.2    Vlan200
10.136.3.1        1   2WAY/DROTHER    00:00:03    10.136.103.3    Vlan903
10.136.3.2        1   FULL/BDR        00:00:03    10.136.103.2    Vlan903
10.136.100.2      1   FULL/DR         00:00:03    10.136.103.5    Vlan903
```

S2

```
S2#sh ip ospf 100 neighbor
Neighbor ID     Pri   State           Dead Time   Address         Interface
10.136.100.1      1   FULL/DR         00:00:03    10.136.200.1    Vlan200
10.136.3.1        1   FULL/DROTHER    00:00:03    10.136.103.3    Vlan903
10.136.3.2        1   FULL/BDR        00:00:03    10.136.103.2    Vlan903
```

```
10.136.100.1      1   FULL/DROTHER    00:00:03    10.136.103.6    Vlan903
```

S1 and S2 switches peer with the VRFs defined on the distribution layer devices D1 and D2 via SVI 903. A separate SVI 200 is defined for establishing a Layer 3 peering between the services switches. This peering may be used for re-routing traffic from the shared services directed to the campus core under specific failure scenarios.

2. Configure the fusion router to advertise a default route into each VPN—As previously mentioned, the goal is for each defined VPN to steer traffic to this central location every time the destination is unknown in the context of a specific VPN. The assumption here is that a default route is present in the fusion router routing table. This could be because the fusion router learns it from a next-hop router (this for example would be the case for Internet edge deployments) or because it is statically defined. The configuration to achieve the desired behavior is shown below (this applies to both S1 and S2 devices).

S1

```
router ospf 100 vrf fusion
 router-id 10.136.100.1
 default-information originate metric 10 metric-type 1
```

S2

```
router ospf 100 vrf fusion
 router-id 10.136.100.2
 default-information originate metric 20 metric-type 1
```

The example above leverages the **default-information originate** command to inject the default route. Note that this command also provides an always keyword that would allow to advertise the default route also when it is not actually present in the fusion router routing table. The result of the specific configuration above on the distribution switches is shown below:

D1

```
D1#show ip route vrf Red supernets-only
Routing Table: Red
Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route
Gateway of last resort is 10.136.103.6 to network 0.0.0.0
O*E1 0.0.0.0/0 [110/20] via 10.136.103.6, 00:09:13, Vlan1053
```

D2

```
D2#show ip route vrf Red supernets-only
Routing Table: Red
Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route

Gateway of last resort is 10.136.103.6 to network 0.0.0.0
O*E1 0.0.0.0/0 [110/20] via 10.136.103.6, 00:00:00, Vlan1053
```

As noted, D1 and D2 install in their routing tables the default route received from the fusion router in S1. The resulting traffic flows were shown in Figure 17.

It is worth noticing that the distribution VRFs also receive from the fusion routers the prefix information for the other VPNs. As already mentioned when discussing the EIGRP deployment option, the fusion router has a complete cross-VPN view of the campus prefixes. OSPF requires all routers belonging to a same area to sync up the OSPF databases, so it is not possible to filter the routing information advertised by the fusion routers. The recommendation is to configure a distribute-list on the receiving distribute VRFs, so that only selected routes (in our example below only the default route) are imported into the routing table (the configuration sample below is valid for both D1 and D2):

```
router ospf 4 vrf Red
 distribute-list Default in Vlan1053
!
ip access-list standard Default
 permit 0.0.0.0
```

OSPF Convergence Results

The convergence results achieved in the different failure scenarios previously discussed are summarized below:

Test 1 (fiber failure between S1 and D1)

- Core to services flows: <1 sec.

  This value is determined by how fast the D1 device detects the failure of the fiber link connecting to S1. This causes the SVI 1053 to go into a DOWN state, thus removing the default route learned via that interface from the routing table.

- Services to core flows: <1 sec.

  Differently from what observed with EIGRP, the fusion router in S1 is capable of removing the route learned from the VRF in D1 before the routing peering expires. This allows us to keep this convergence event sub-second, as highlighted below:

```
Dec  9 16:13:10.414 EST: %LINEPROTO-5-UPDOWN: Line protocol on Interface
GigabitEthernet2/23, changed state to down
Dec  9 16:13:10.418 EST: %LINEPROTO-SP-5-UPDOWN: Line protocol on Interface
GigabitEthernet2/23, changed state to down
Dec  9 16:13:10.422 EST: %LINK-SP-3-UPDOWN: Interface GigabitEthernet2/23, changed
state to down
Dec  9 16:13:10.426 EST: %LINK-3-UPDOWN: Interface GigabitEthernet2/23, changed state
to down
Dec  9 16:13:10.806 EST: RT(fusion): del 10.137.42.0/24 via 10.136.104.3, ospf metric
[110/23]
Dec  9 16:13:13.374 EST: %OSPF-5-ADJCHG: Process 100, Nbr 10.136.4.1 on Vlan904 from
2WAY to DOWN, Neighbor Down: Dead timer expired
```

Test 2 (fiber failure between S1 and shared services area)

- Core to services flows: <1 sec.

  All the traffic flows directed to the fusion router in S1 need to be re-routed across the transit link toward S2. The re-routing usually causes sub-second outage.

- Services to core flows: <1 sec.

  Traffic originated in the shared services area is black-holed until the fusion router in S2 becomes the active HSRP device on the shared services subnet (VLAN 32). Configuring sub-second HSRP timers allows us to keep this convergence event sub-second.

Test 3 (firewall failure)

- Core to services flows: ~4-5 sec.

- Services to core flows: ~4-5 sec.

  Two main components contribute to the traffic outages shown above:

  - Time required for the firewall in S2 to take on the active role—This cannot currently be configured lower than three seconds in a FWSM deployment.
  - Assuming the OSPF timers are set aggressively, it is possible that the adjacencies will be reset (given the three seconds required for the new firewall to start passing traffic). The time required to re-establishing the OSPF peering needs to be taken into consideration and that leads to the overall 4-5 second outage. In order to avoid the OSPF adjacencies from being reset, the OSPF timers could be set less aggressively. Keep in mind that this would affect the recovery times in failure scenarios (as, for example, fiber failure one) where the OSPF dead time is a main factor in determining the overall traffic outage.

Test 4 (services switch failure)

- Core to services flows: ~4-5 sec.
- Services to core flows: ~4-5 sec.

  The same consideration made in the previous firewall failure scenario applies here as well.

Test 5 (distribution switch failure)

- Core to services flows: <1 sec.

  The recovery mechanism in this case is ECMP from the core devices once the distribution switch fails. This is very fast, assuming that the link detection mechanism works properly on the core devices directly connected to the failing switch.

- Services to core flows: ~4 sec.

  The recovery time is dependant on the configured OSPF dead timer. This is because the fusion router on S1 (active HSRP device) keeps sending traffic to the VRF inside the failed switch until it removes the OSPF peering, as shown below:

```
Dec  9 16:42:56.666 EST: %LINEPROTO-5-UPDOWN: Line protocol on Interface
GigabitEthernet2/23, changed state to down
Dec  9 16:42:56.674 EST: %LINK-3-UPDOWN: Interface GigabitEthernet2/23, changed state
to down
Dec  9 16:42:56.666 EST: %LINEPROTO-SP-5-UPDOWN: Line protocol on Interface
GigabitEthernet2/23, changed state to down
Dec  9 16:42:56.674 EST: %LINK-SP-3-UPDOWN: Interface GigabitEthernet2/23, changed
state to down
Dec  9 16:42:59.990 EST: %OSPF-5-ADJCHG: Process 100, Nbr 10.136.4.1 on Vlan904 from
2WAY to DOWN, Neighbor Down: Dead timer expired
Dec  9 16:43:00.014 EST: RT(fusion): del 10.137.32.0/24 via 10.136.103.3, ospf metric
[110/23]
```

Use of eBGP between VRFs and Fusion Router

The use of eBGP for peering between the VRFs defined on the distribution switches (D1 and D2) and the fusion routers is mostly used in conjunction with MPLS VPN path isolation deployments. The main reason is that with MPLS VPN, the MP-iBGP routing protocol is used between PE devices to exchange VPN routes information. In the network example shown in Figure 12, the distribution layer switches play the PE role, which means they have iBGP configured. It is hence easy to add the eBGP configuration to them to establish eBGP peering with the fusion routers and start exchanging routes (without the need for any routing protocol redistribution).

Note For a more detailed discussion around the deployment of MPLS VPN as a path isolation alternative, see: http://www.cisco.com/en/US/docs/solutions/Enterprise/Network_Virtualization/PathIsol.html

The following are the required configuration steps:

1. Allow establishment of the eBGP adjacencies across the firewall context—The required configurations for the fusion routers and the distribution VRFs are shown below:

   D1

```
router bgp 200
 timers bgp 2 10
 !
 address-family ipv4 vrf Red
  neighbor 10.136.103.5 remote-as 100
  neighbor 10.136.103.5 activate
  neighbor 10.136.103.6 remote-as 100
  neighbor 10.136.103.6 activate
  maximum-paths 2
  no synchronization
  bgp router-id 10.136.200.1
 exit-address-family
```

   D2

```
router bgp 200
 timers bgp 2 10
 !
 address-family ipv4 vrf Red
  neighbor 10.136.103.5 remote-as 100
  neighbor 10.136.103.5 activate
  neighbor 10.136.103.6 remote-as 100
  neighbor 10.136.103.6 activate
  maximum-paths 2
  no synchronization
  bgp router-id 10.136.200.2
 exit-address-family
```

   S1

```
router bgp 100
 timers bgp 2 10
 !
 address-family ipv4 vrf fusion
  neighbor 10.136.103.2 remote-as 200
  neighbor 10.136.103.2 activate
  neighbor 10.136.103.3 remote-as 200
  neighbor 10.136.103.3 activate
  neighbor 10.136.103.5 remote-as 100
  neighbor 10.136.103.5 activate
  maximum-paths 2
  no synchronization
  bgp router-id 10.136.100.1
 exit-address-family
```

   S2

```
router bgp 100
 timers bgp 2 10
 !
 address-family ipv4 vrf fusion
  neighbor 10.136.103.2 remote-as 200
  neighbor 10.136.103.2 activate
```

```
 neighbor 10.136.103.3 remote-as 200
 neighbor 10.136.103.3 activate
 neighbor 10.136.103.6 remote-as 100
 neighbor 10.136.103.6 activate
 maximum-paths 2
 no synchronization
 bgp router-id 10.136.100.2
exit-address-family
```

Some considerations on the configuration samples above:

- The eBGP sessions are established in a fully meshed fashion between the fusion routers and the distribution VRFs, following the design principle highlighted in Figure 15.
- An additional iBGP peering is established between the fusion routers. This is required to provide re-routing capabilities in a failure scenario where a fusion router loses direct connectivity to the shared services area.
- Similar to what was discussed in the EIGRP and OSPF scenario, a timer tuning is also required with eBGP to reduce the extent of outages in specific failure scenarios. The difference is that with eBGP it is critical to ensure that the timers are not tuned too low. Losing the established session (for example during a firewall failover scenario) would in fact cause a large outage (40 seconds or more) given the fact eBGP is slower in re-establishing peering sessions and exchanging routing information. The configuration above (2 seconds for keepalives and 10 seconds for hold-time) is conservative enough to shield from that event.
- The **maximum-paths 2** command is required to ensure that the fusion routers install in their routing tables the routing information received from both the distribution VRFs (and vice versa).

Before eBGP peering sessions can be successfully established, it is required to allow these packets through the firewall, as shown below:

```
access-list <ACL_NAME> extended permit tcp any any eq bgp
```

As a consequence of the previous configuration steps, eBGP sessions are established between the fusion router in S1 and the VRFs in the distribution layer switches, as highlighted below:

D1

```
D1#sh ip bgp vpnv4 vrf Red summary
BGP router identifier 10.136.200.1, local AS number 200
<snip>
Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
10.136.103.5    4   100   37058   37167  1167515    0    0 13:50:39        1
10.136.103.6    4   100   37051   37173  1167515    0    0 13:50:39        1
```

D2

```
D2#sh ip bgp vpnv4 vrf Red summary
BGP router identifier 10.136.200.2, local AS number 200
<snip>
Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
10.136.103.5    4   100   40552   40591  3596185    0    0 23:04:08        1
10.136.103.6    4   100  143455  143543  3596185    0    0 18:02:45        1
```

S1

```
S1#sh ip bgp vpnv4 vrf fusion summary
BGP router identifier 10.136.100.1, local AS number 100
<snip>
Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
10.136.103.2    4   200   38649   38597  2957809    0    0 18:06:11       32
10.136.103.3    4   200   38298   38203  2957809    0    0 14:07:05       32
```

```
10.136.103.5    4   100 2173448 2171218  2957809    0    0 18:06:08       89
```

S2

```
cr15-6500-2#sh ip bgp vpnv4 vrf fusion summary
BGP router identifier 10.136.100.2, local AS number 100
<snip>
Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
10.136.103.2    4   200   42475   42462  9446857    0    0 23:09:16        32
10.136.103.3    4   200  143736  143590  9446857    0    0 14:08:46        32
10.136.103.6    4   100 6534446 6510120  9446857    0    0 18:07:50        90
```

S1 and S2 switches peer with the VRFs defined on the distribution layer devices D1 and D2 via eBGP, but they also established a direct iBGP session between them. This latter peering may be used for re-routing traffic from the shared services directed to the campus core under specific failure scenarios.

2. Configure the fusion router to advertise a default route into each VPN—As previously mentioned, the goal is for each defined VPN to steer traffic to this central location every time the destination is unknown in the context of a specific VPN. The assumption here is that a default route is present in the fusion router routing table. This could be because the fusion router learns it from a next-hop router (this for example would be the case for Internet edge deployments) or because it is statically defined. The configuration to achieve the desired behavior is shown below.

S1

```
router bgp 100
 !
 address-family ipv4 vrf fusion
  neighbor 10.136.105.2 remote-as 200
  neighbor 10.136.105.2 activate
  neighbor 10.136.105.2 default-originate route-map default_only
  neighbor 10.136.105.2 route-map default_only out
  neighbor 10.136.105.3 remote-as 200
  neighbor 10.136.105.3 activate
  neighbor 10.136.105.3 default-originate route-map default_only
  neighbor 10.136.105.3 route-map default_only out
  exit-address-family
!
ip access-list standard Default
 permit 0.0.0.0
!
route-map default_only permit 10
 match ip address Default
 set metric 10
```

S2

```
router bgp 100
 !
 address-family ipv4 vrf fusion
  neighbor 10.136.105.2 remote-as 200
  neighbor 10.136.105.2 activate
  neighbor 10.136.105.2 default-originate route-map default_only
  neighbor 10.136.105.2 route-map default_only out
  neighbor 10.136.105.3 remote-as 200
  neighbor 10.136.105.3 activate
  neighbor 10.136.105.3 default-originate route-map default_only
  neighbor 10.136.105.3 route-map default_only out
exit-address-family
!
ip access-list standard Default
 permit 0.0.0.0
!
```

```
route-map default_only permit 10
 match ip address Default
 set metric 20
```

The example above leverages the **default-originate neighbor** command to inject the default route. Note that the additional route-map applied to each neighbors to ensure that only the default route is advertised to the distribution VRFs. Looking more carefully to the route-map default-only, it is possible to note how S2 also set an higher metric than S1 for the default route. This is done to ensure that S1 is always the preferred next hop toward the shared services area (refer to the design principle shown in Figure 17).

The result of the specific configuration above on the distribution switches is shown below:

D1

```
D1#show ip route vrf Red supernets-only
Routing Table: Red
Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route
Gateway of last resort is 10.136.103.6 to network 0.0.0.0
B*   0.0.0.0/0 [20/10] via 10.136.103.6, 00:20:11
```

D2

```
D2#show ip route vrf Red supernets-only
Routing Table: Red
Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route
Gateway of last resort is 10.136.103.6 to network 0.0.0.0
B*   0.0.0.0/0 [20/10] via 10.136.103.6, 00:20:11
```

As noted, D1 and D2 install in their routing tables the default route received from the fusion router in S1. The resulting traffic flows are shown in Figure 17.

eBGP Convergence Results

The convergence results achieved in the different failure scenarios previously discussed are summarized below:

Test 1 (fiber failure between S1 and D1)

- Core to services flows: <1 sec.

  This value is determined by how fast the D1 device detects the failure of the fiber link connecting to S1. This causes the SVI 1053 to go into a DOWN state, thus removing the default route learned via that interface from the routing table.

- Services to core flows: ~9-10 sec.

  This value is determined by how fast the fusion router can remove from the routing table the routes learned from D1. Traffic is black-holed until that happens, since as mentioned above the SVI is now in a DOWN state on D1. The main factor for this detection is the configured BGP hold-time. It is

not recommended to tune this lower than 10 seconds to avoid losing the peering under different failure scenarios (the consequence would be an outage above 40 seconds). This behavior is shown in the output below captured on the fusion router in S1:

```
Dec  9 16:53:57.780 EST: %LINEPROTO-5-UPDOWN: Line protocol on Interface
GigabitEthernet2/23, changed state to down
Dec  9 16:53:57.788 EST: %LINK-3-UPDOWN: Interface GigabitEthernet2/23, changed state
to down
Dec  9 16:53:57.784 EST: %LINEPROTO-SP-5-UPDOWN: Line protocol on Interface
GigabitEthernet2/23, changed state to down
Dec  9 16:53:57.788 EST: %LINK-SP-3-UPDOWN: Interface GigabitEthernet2/23, changed
state to down
Dec  9 16:54:07.840 EST: %BGP-5-ADJCHANGE: neighbor 10.136.103.3 vpn vrf fusion Down
BGP Notification sent
Dec  9 16:54:07.840 EST: %BGP-3-NOTIFICATION: sent to neighbor 10.136.103.3 4/0 (hold
time expired) 0 bytes
Dec  9 16:54:07.840 EST: RT(fusion): del 10.137.32.0/24 via 10.136.103.3, bgp metric
[20/3584]
```

As noted above, the route to a remote destination (10.137.32.0) is removed after the BGP peering between the fusion router in S1 and the VRF in D1 is removed because of the expiration of the BGP hold-time.

Test 2 (fiber failure between S1 and shared services area)

- Core to services flows: <1 sec.

  All the traffic flows directed to the fusion router in S1 need to be re-routed across the transit link toward S2. The re-routing usually causes sub-second outage.

- Services to core flows: <1 sec.

  Traffic originated in the shared services area is black-holed until the fusion router in S2 becomes the active HSRP device on the shared services subnet (VLAN 32). Configuring sub-second HSRP timers allows us to keep this convergence event sub-second.

Test 3 (firewall failure)

- Core to services flows: ~3-4 sec.
- Services to core flows: ~3-4 sec.

  The traffic outages shown above are due to the time required for the firewall in S2 to take on the active role. This cannot currently be configured lower than three seconds in a FWSM deployment.

Test 4 (services switch failure)

- Core to services flows: ~9-10 sec.

  This outage affects all the flows and is caused by the fact that the distribution VRFs keep sending traffic to the fusion router inside S1 until they realize it failed. The only indication of that condition is the expiration of the BGP held-time, which brings the recovery time to the value shown above.

- Services to core flows: ~9-10 sec.

  Half of the flows originated from the shared services (the ones directed to the VRF inside D1) area experience 9-10 seconds of outage because after the failure of S1 there is no longer a Layer 2 path available to deliver the traffic to D1. The other half of the flows are instead sent to the VRF inside D2 and only experience an outage of around 3-4 seconds, caused by the time required by the FWSM inside S2 to become active.

Test 5 (distribution switch failure)

- Core to services flows: <1 sec.

The recovery mechanism in this case is ECMP from the core devices once the distribution switch fails. This is very fast, assuming that the link detection mechanism works properly on the core devices directly connected to the failing switch.

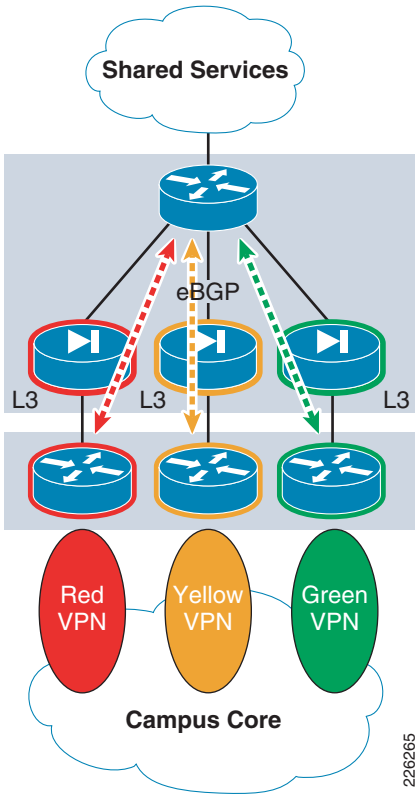
- Services to core flows: ~9-10 sec.

  The recovery time is once again dependant on the configure BGP hold-time. This is because the fusion router on S1 (active HSRP device) keeps sending traffic to the VRF inside the failed switch until it removes the eBGP peering.

Firewall Contexts Deployed in Routed Mode

When deploying the firewall contexts in routed mode, one current limitation is represented by the lack of routing protocol support on the contexts (a Cisco firewall supports routing protocols in routed mode only when non-virtualized, i.e., single context mode). The approach recommended in the context of this document is leveraging eBGP to allow the exchange of routing information between the fusion routers and the various VRFs defined at the distribution layer, as shown in Figure 24.

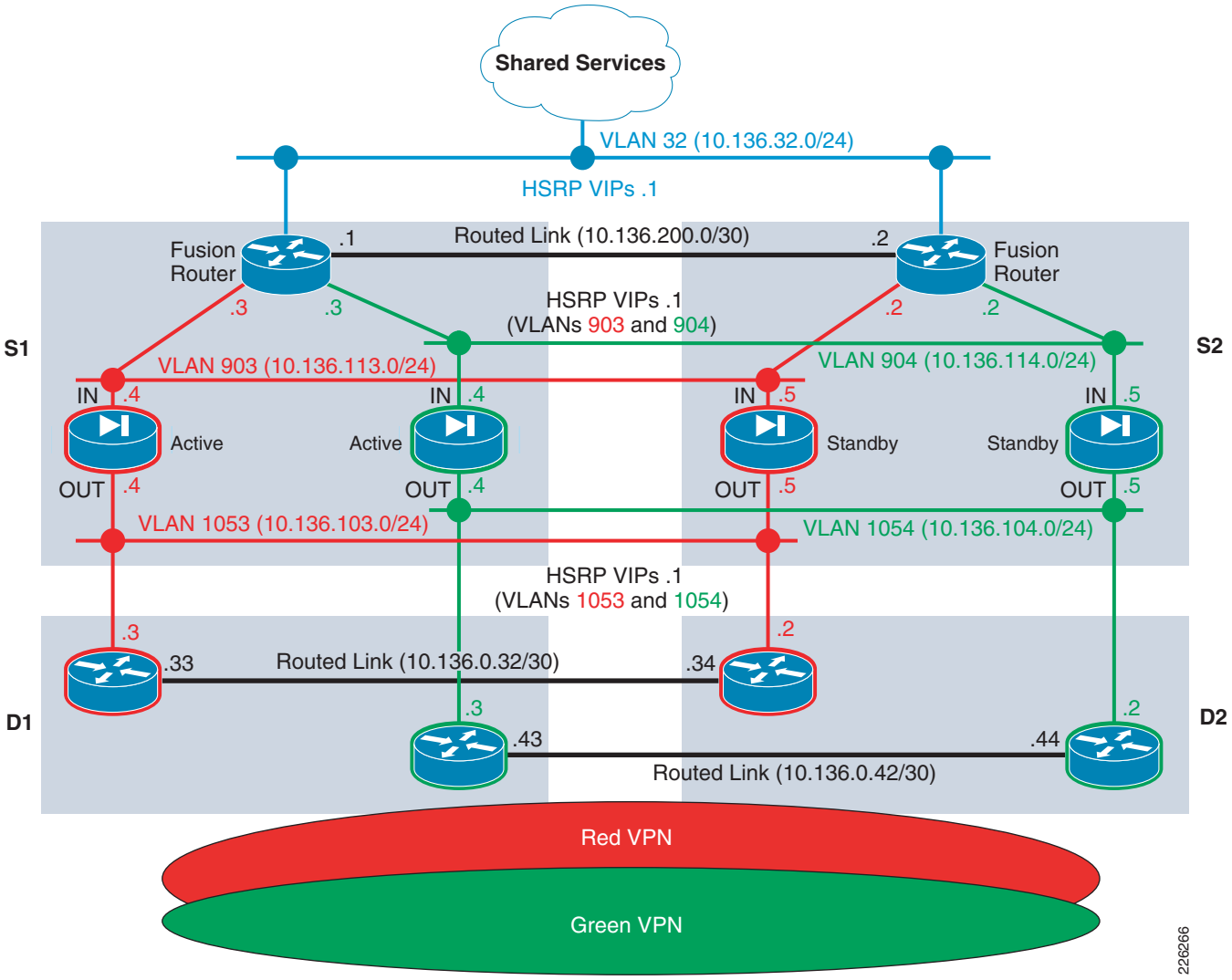
Figure 24 Routing Peering with Firewall Contexts in Routed Mode



Note An alternative approach would be to simply use static routing between the VRFs, firewall contexts, and fusion routers. This is not recommended because it may lead to black-holing traffic in specific failure scenarios, especially deploying the services edge function in redundant sites.

The recommended deployment model for the services edge with firewall deployed in routed mode is shown in Figure 25.

Figure 25 Firewall Contexts in Routed Mode Example



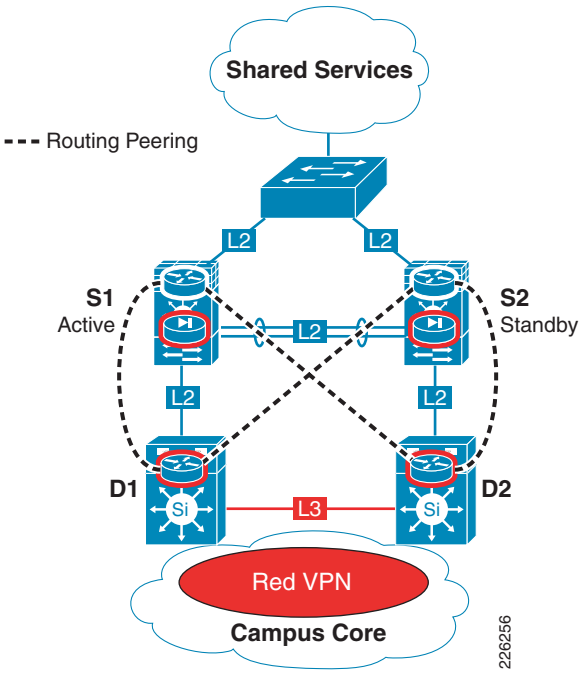
When comparing this network topology with the one in Figure 14 relative to the deployment of firewall contexts in transparent mode, note the following:

- Each firewall context is now a routed hop in the network. This implies that the firewall inside and outside VLANs now belong to two different IP subnets.
- HSRP is still the First Hop Redundancy Protocol of choice and it is leveraged to provide virtual gateway functionalities for the following interfaces:
  - Shared services subnet (VLAN 32)—This is under the assumption that the shared services are deployed in a subnet directly connected to the fusion devices.
  - Firewall inside subnets (VLAN 903 for the Red VPN and VLAN 904 for the Green VPN).
  - Firewall outside subnets (VLAN 1053 for the Red VPN and VLAN 1054 for the Green VPN).
- The fusion routers and the VRFs deployed in the distribution layer switches are still connected with a routed link. This is to provide a Layer 3 path to be used for re-routing traffic under specific failure conditions (as discussed later).

- As in the transparent firewall deployment, the fusion routers define a separate interface (or SVI) to be dedicated to each campus VPN.

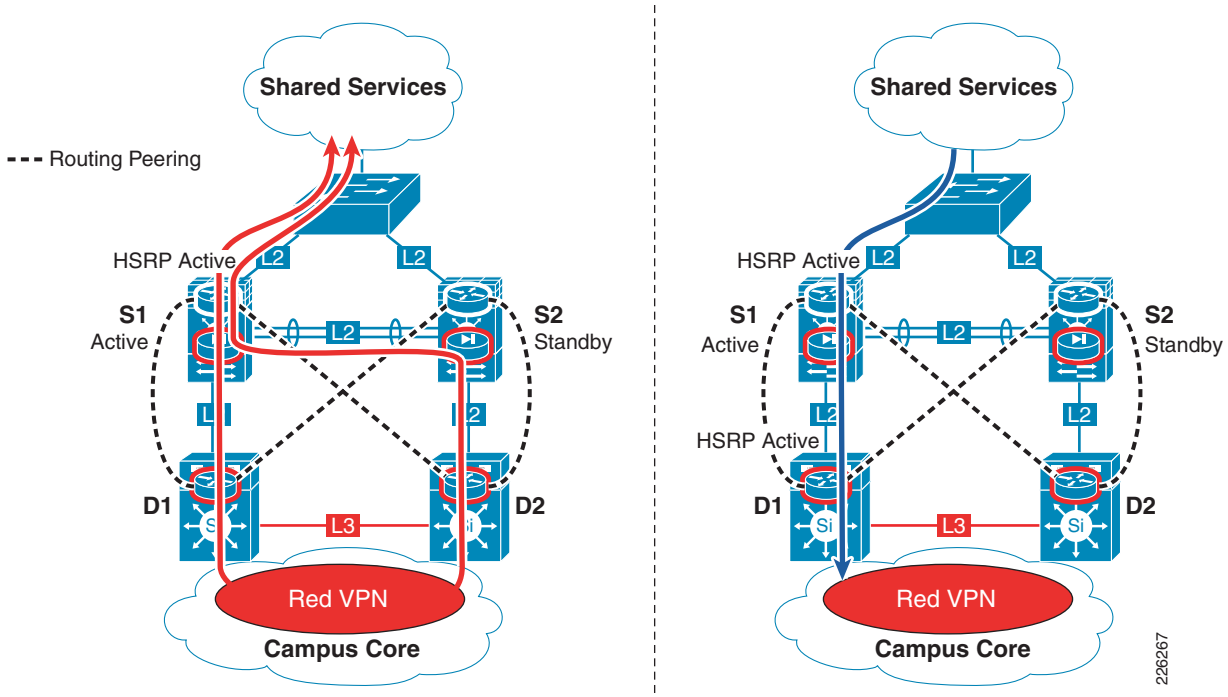
Assuming that the policies on each firewall context have been properly deployed so that the routing protocol of choice is allowed through (the configuration required for eBGP is discussed in the protocol specific section later in the document), each VRF defined in the distribution switches peers with both fusion routers, as highlighted in Figure 26.

Figure 26 Full Mesh Routing Peering



Since the desire for each VPN is to steer the traffic to the services edge every time the destination is unknown in that specific routing domain, in a typical deployment the fusion routers just advertise a default route to each VRF defined in D1 and D2. At the same time, each VRF advertises to the fusion router the remote campus subnets. The consequence is that by default the traffic flows between the core and the shared services area follow the paths described in Figure 27.

Figure 27 Default Traffic Flows



The reason behind this behavior is that the firewall module is now a routed hop in the network and routes the traffic based on static routing information that needs to be configured, as shown in the example below:

```
route outside-vrf6 10.137.0.0 255.255.0.0 10.136.113.1
route inside-vrf6 0.0.0.0 0.0.0.0 10.136.103.1
```

In the configuration example above, 10.137.0.0/16 represents all the address space used by the specific Red VPN inside the campus. The consequence is that every time the firewall receives a traffic flow originated in the campus core and directed to the shared services area, it uses the HSRP VIP address 10.136.113.1 (on VLAN 1053) as next-hop. Similarly, for traffic flows in the opposite direction it uses the HSRP VIP address 10.136.103.1 (on VLAN 903). The active devices on both subnets (S1 and D1 by design) would then be responsible to receive and route these flows.

In the following sections the behavior under different failure scenarios is discussed. The specific recovery mechanisms and recovery times for eBGP deployment are analyzed in the following protocol specific section.

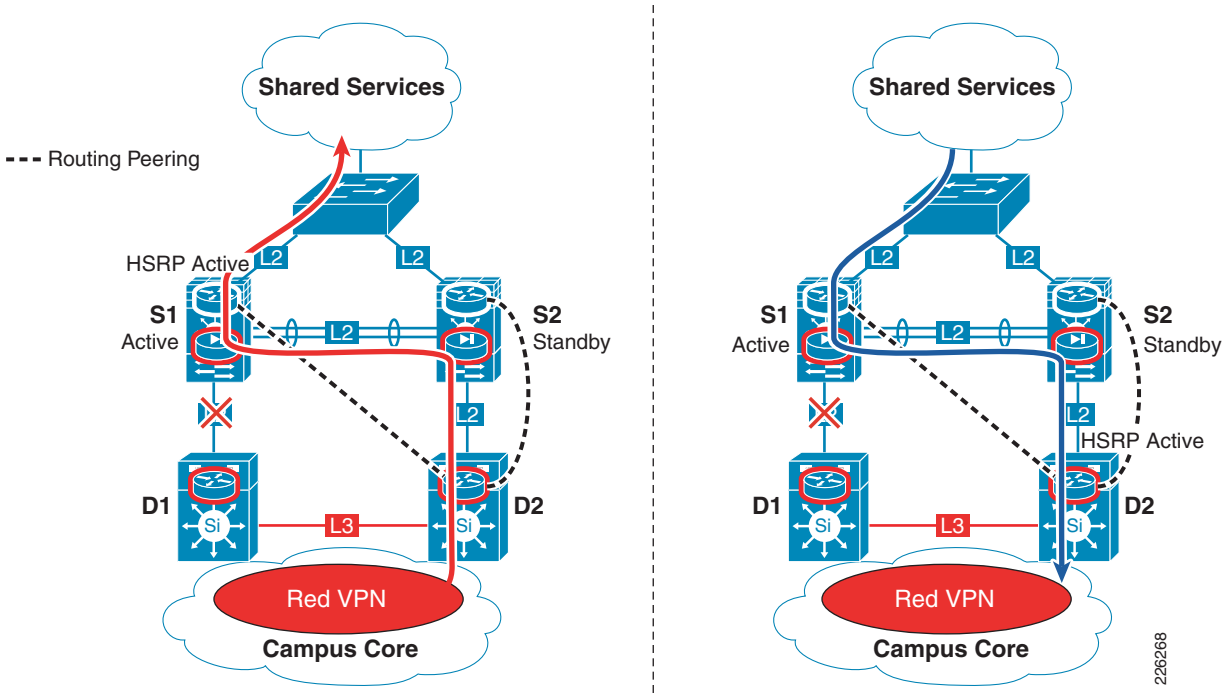
Note The redundancy discussion in this part of the document is focused on recovery mechanisms in a single site deployment. For redundant sites consideration, refer to Planning for Site Redundancy in a Virtualized Network.

Convergence Analysis

Fiber Failure between Distribution and Services Switches

Following the failure of the connection between D1 and S1, the peering between the VRF in D1 and the fusion routers are removed. In addition, D2 becomes the HSRP active device on the firewall outside VLAN 1053. The consequence on the traffic flows is highlighted in Figure 28.

Figure 28 Traffic Flows after Fiber Failure between D1 and S1



- Core to services flows—D1 stops receiving the default route (or the more specific route for the shared services subnet) from the fusion routers. As a consequence, it stops advertising that information to the campus core, with the consequence that all the traffic from the core directed to the shared services area is now sent to D2. The convergence experienced in this case is based on how fast D1 detects the fiber failure and stops advertising the routing information toward the core. Assuming that the link failure detection works, this would be very fast and cause sub-second outage.
- Services to core flows—Independently of the fiber failure, the fusion router in S1 maintains the active HSRP role on the shared services subnet (VLAN 32). Following the fiber failure, D2 becomes the active HSRP device on the firewall outside subnet (VLAN 1053) and from that moment the firewall starts sending all the traffic flows to the VRF inside D2. From a convergence perspective, the main factor dictating the length of the outage here is the time required for D2 to gain the HSRP active role. In order to keep this sub-second an aggressive setting of HSRP timers is required, as shown below:

  D1

```
interface Vlan1053
 description Firewall Outside VRF Red
 ip vrf forwarding Red
 ip address 10.136.103.3 255.255.255.0
 standby 1 ip 10.136.103.1
 standby 1 timers msec 250 msec 750
 standby 1 priority 105
 standby 1 preempt delay minimum 180
```

  D2

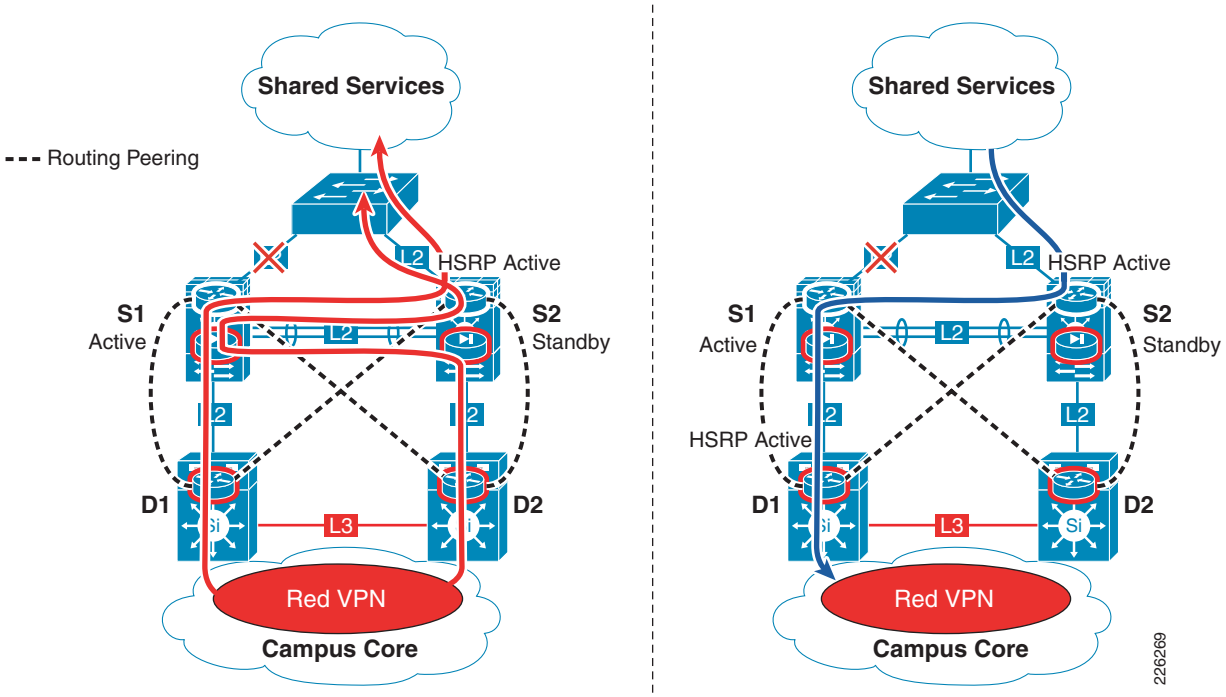
```
interface Vlan1053
 description Firewall Outside VRF Red
 ip vrf forwarding Red
 ip address 10.136.103.2 255.255.255.0
 standby 1 ip 10.136.103.1
 standby 2 timers msec 250 msec 750
```

Note The use of sub-second HRSP timers is recommended for deployments with 150 VLANs or less. For more information, refer to the campus design guides: http://www.cisco.com/en/US/netsol/ns815/networking_solutions_program_home.html

Fiber Failure between Services Switches and Shared Services Area

The failure of the fiber link connecting the services switch S1 to the shared services area does not cause any change in the routing peering between VRFs and fusion routers. The traffic flows become the ones depicted in Figure 29.

Figure 29 Traffic Flows after Fiber Failure between S1 and Shared Services Area



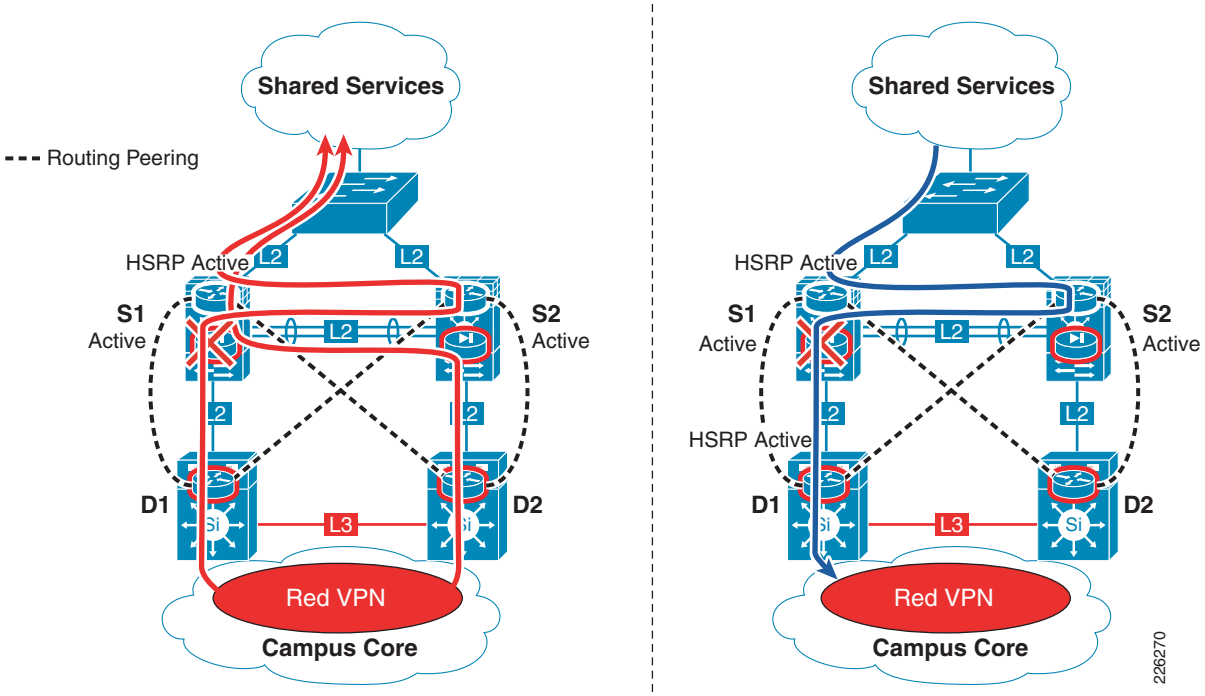
- Core to services flows—The fusion router in S1 remains the HSRP active device on the firewall inside subnet (VLAN 903) even after the fiber failure (this is because the HRSP messages are exchanged across the port-channel between the services switches). This means traffic needs to be re-routed across the transit link to reach the destination in the shared services area.
- Services to core flows—The fusion router in S2 becomes the HSRP active device on the shared services subnet (VLAN 32) and has to send all the traffic flows across the transit link to S1 since the active firewall is in that device. All the flows are then directed to the VRF in D1 that is the HSRP active device on the firewall outside subnet (VLAN 1053).

Design Recommendation—In order to avoid the suboptimal path across the transit link shown in Figure 20, when possible it is recommended to deploy a port-channel between each services switch and the shared services area.

Active Firewall Module Failure

After the failure of the firewall module in S1, the VRFs and the fusion routers maintain their routing peering through the newly activated firewall inside S2. The consequent traffic flows are shown in Figure 30.

Figure 30 Traffic Flows after Firewall Failure



- Core to services flows—Independently from the firewall failure, S1 retains the role of active HSRP on the firewall inside subnet (VLAN 903). This means all the traffic is still directed to the fusion router in S1, even if the firewall active is now inside S2, creating the traffic pattern shown in Figure 30. The overall outage before these flows are established is mainly affected by the time required for the firewall in S2 to become active. This is true under the assumption that the configured routing protocol hold-time is higher than this time. If not, on top of the firewall failover time, we need to consider the time required to re-establish the routing protocol peering and advertise the routing information (more considerations can be found in the protocol specific sections below). Note that the time required by the firewall module in S2 to detect the failure of the other firewall and become the active device depends on the configured hold-time between the firewalls. FWSM does not allow to configure hold-time values below three seconds, as shown below:

```
FWSM(config)# failover polltime unit msec 500 holdtime ?
configure mode commands/options:
  <3-45>  Hold time in seconds, default is 3 X poll time but minimum is 3
          Seconds
```

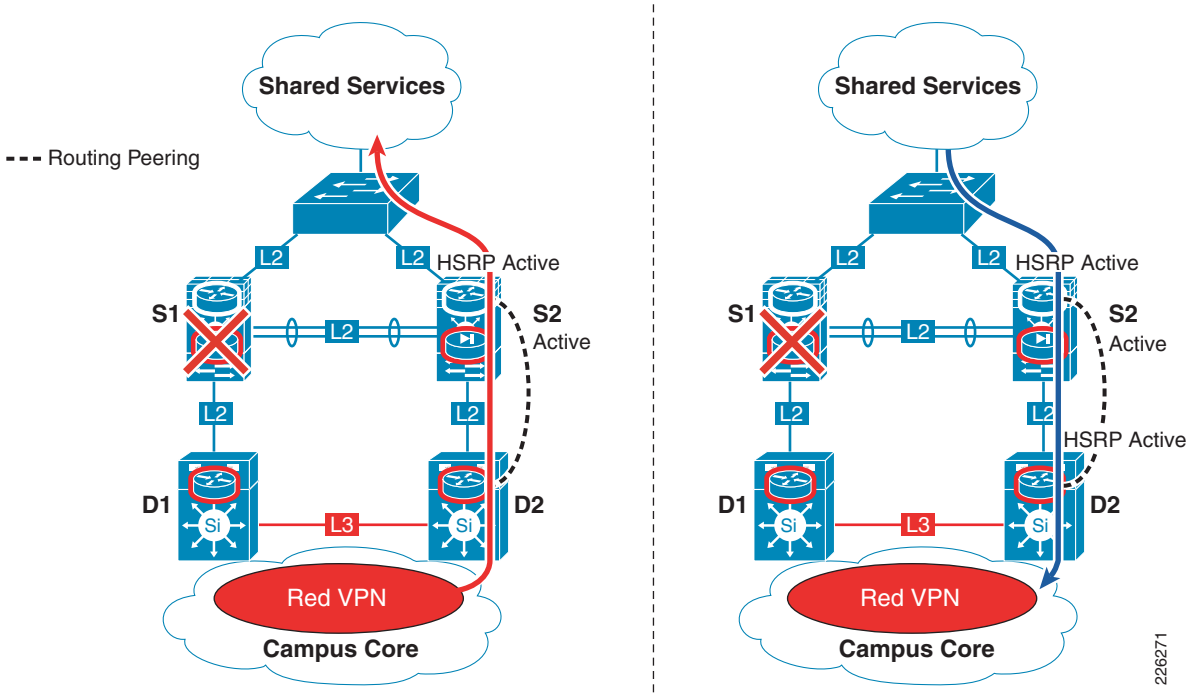
- Services to core flows—Independently from the firewall failure, S1 remains the active HSRP device on the shared services subnet (VLAN 32). At the same time, the VRF in D1 retains the active HSRP role on the firewall outside subnet (VLAN 1053), so all the traffic flows originated from the shared services area are directed to it. For the recovery time experienced in this direction, the same considerations made in the previous bullet point are still valid and convergence depends on how fast the firewall can failover and if the peering between devices is lost or not in the meantime.

Design Recommendation—Configure firewall preemption so that the firewall module inside S1 always takes the active role (when up and running), avoiding the suboptimal traffic paths shown above.

Services Switch Failure

This failure scenario is highlighted in Figure 31.

Figure 31 Traffic Flows after Services Switch Failure



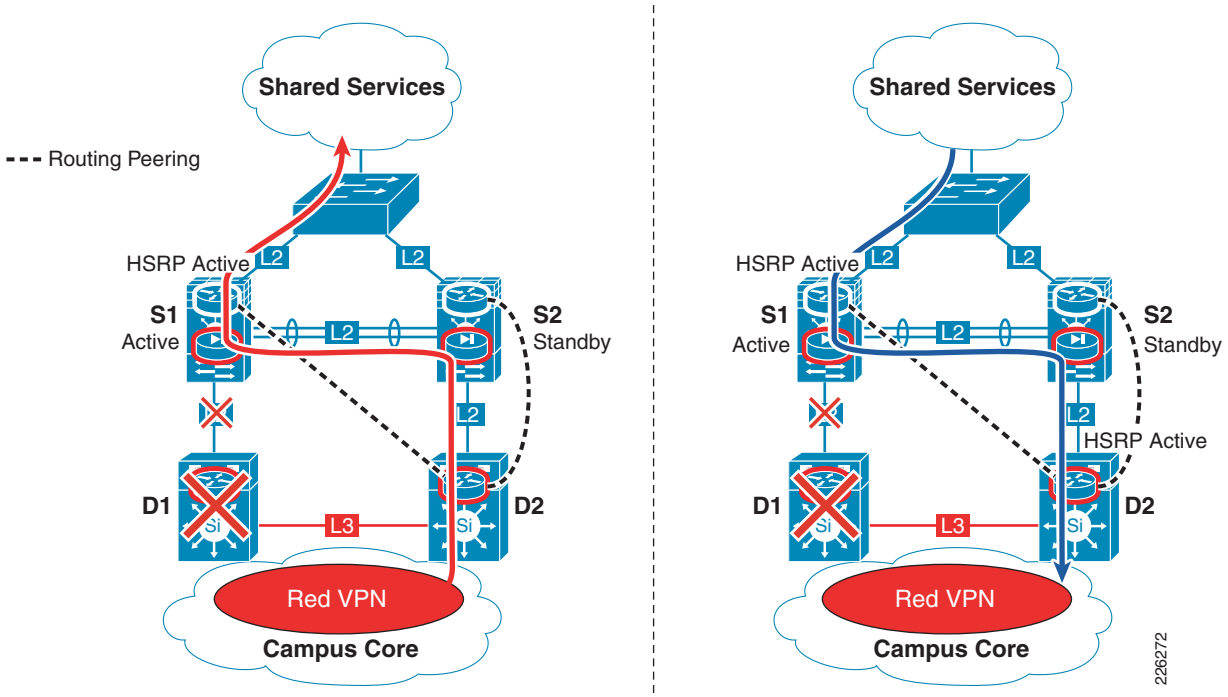
Because of the failure of the entire services switch S1, the peering between the VRFs and the fusion router in S1 is removed. The consequences on the traffic flows are:

- Core to services flows—The link between D2 and S2 is the only available path from the core to the shared services. From a convergence perspective, the outage is dictated by how fast the firewall becomes active, allowing the successive establishment of routing peering. We can expect the entity of this outage to be the same as the one discussed in the previous section for the firewall failure scenario.
- Services to core flows—The fusion router only has a valid path to the remote campus destinations via the VRF in D2. The fusion router in S2 becomes the HSRP active device on the shared services subnet (VLAN 32) because of the failure of the S1 switch. The recovery mechanism is identical to the one described in the previous bullet point.

Distribution Switch Failure

The last relevant scenario is the failure of the distribution switch, as shown in Figure 32.

Figure 32 Traffic Flows after Distribution Switch Failure



From a traffic flow and routing peering perspective, this scenario is very similar to the fiber failure one.

- Core to services flows—Since the entire D1 switch fails, the core devices remove the valid path through it and re-route all the traffic via D2. Once again, assuming the link failure detection works, this is an ECMP re-route that causes only sub-second outages.
- Services to core flows—The firewall inside S1 keeps sending all the traffic flows to the HSRP VIP on the firewall outside subnet (VLAN 1053). All these flows are picked up by the VRF in D2 once it becomes the HSRP active. The length of the outage depends on how long this process takes. Configuring sub-second timers allows us to keep this convergence event sub-second.

The following section discusses the specific design and configuration consideration for eBGP deployment, including the convergence results achieved for all the failure scenarios previously discussed.

Use of eBGP between VRFs and Fusion Router

The considerations around the use of eBGP for peering between the VRFs defined on the distribution switches (D1 and D2) and the fusion routers when deploying firewall contexts in routed mode are basically the same already made for firewall contexts in transparent mode (refer to the corresponding section for more details). The only difference is that there is no requirement to ensure that the fusion router in S1 advertises the best default route, since the routing decision is performed by the routed firewall based on the active HSRP device on the firewall inside subnet (VLAN 903).

eBGP Convergence Results

The convergence results achieved in the different failure scenarios are summarized below:

Test 1 (fiber failure between S1 and D1)

- Core to services flows: <1 sec.

This value is determined by how fast the D1 device detects the failure of the fiber link connecting to S1. This causes the SVI 1053 to go into a DOWN state, thus removing the default route learned via that interface from the routing table.

- Services to core flows: <1 sec.

  This value is determined by how fast the VRF on D2 can become the HSRP active for the firewall outside subnet (VLAN 1053). Configuring sub-second HSRP timers allow us to keep the recovery time sub-second.

Test 2 (fiber failure between S1 and shared services area)

- Core to services flows: <1 sec.

  All the traffic flows directed to the fusion router in S1 need to be re-routed across the transit link toward S2. The re-routing usually causes sub-second outage.

- Services to core flows: <1 sec.

  Traffic originated in the shared services area is black-holed until the fusion router in S2 becomes the active HSRP device on the shared services subnet (VLAN 32). Configuring sub-second HSRP timers allows us to keep this convergence event sub-second.

Test 3 (firewall failure)

- Core to services flows: ~3-4 sec.
- Services to core flows: ~3-4 sec.

  The traffic outages shown above are due to the time required for the firewall in S2 to take on the active role. This cannot currently be configured lower than three seconds in a FWSM deployment.

Test 4 (services switch failure)

- Core to services flows: ~3-4 sec.
- Services to core flows: ~3-4 sec.

  The traffic outages shown above are due to the time required for the firewall in S2 to take on the active role. Note how, unlike the eBGP deployment with firewall context in transparent mode, the configured BGP hold-time is not a factor dictating the convergence achieved in this case. This is because the firewall context is now a routed hop and always sends traffic to the HSRP VIP on both the firewall inside and outside subnets.

Test 5 (distribution switch failure)

- Core to services flows: <1 sec.

  The recovery mechanism in this case is ECMP from the core devices once the distribution switch fails. This is very fast, assuming that the link detection mechanism works properly on the core devices directly connected to the failing switch.

- Services to core flows: <1 sec.

  This outage is determined by the time required by the VRF in D2 to become the active HSRP device for the firewall outside subnet (VLAN 1053). Once again, assuming the use of sub-second HSRP timers, it is possible to contain the outage to a sub-second value.

Dual Tier Implementation—Summary and Design Recommendations

The dual tier implementation model discussed in this section is recommended for larger deployments, where it is beneficial to keep separated the functionalities of terminating the VRF (on the distribution layer switches) and providing firewall and fusion router service (on the services switches).

For the sake of the stability of the overall services edge design, it is recommended to avoid creating STP loops between the distribution layer and services switches. This is achieved by connecting these devices with Layer 2 trunks in a U-shape fashion, where the connection between the distribution layer switches is configured as a routed link.

Depending on the mode of operation of the firewall contexts, various routing protocols can be leveraged to allow a dynamic exchange of routing information between the fusion routers and the VRFs defined in the distribution layer:

- Firewall in transparent mode—EIGRP, OSPF, or eBGP
- Firewall in routed mode—eBGP

Table 1 summarizes the convergence results achieved under the following failure scenarios:

- Test 1—Fiber failure between services and distribution switches
- Test 2—Fiber failure between services switch and the shared services area
- Test 3—FWSM failure
- Test 4—Services switch failure
- Test 5—Distribution switch failure

Table 1 Two Tier Model Convergence Result[1]

	Test 1	Test 2	Test 3	Test 4	Test 5
FW Contexts in Transparent Mode					
EIGRP	**<1 sec.** *~3 sec.*	**<1 sec.** *<1 sec.*	**~4-5 sec.** *~4.5 sec.*	**~4-5 sec.** *~4-5 sec.*	**<1 sec.** *~3 sec.*
OSPF	**<1 sec.** *<1 sec.*	**<1 sec.** *<1 sec.*	**~4-5 sec.** *~4.5 sec.*	**~4-5 sec.** *~4-5 sec.*	**<1 sec.** *~4 sec.*
eBGP	**<1 sec.** *~9-10 sec.*	**<1 sec.** *<1 sec.*	**~3-4 sec.** *~3-4 sec.*	**~9-10 sec.** *~9-10 sec.*	**<1 sec.** *~9-10 sec.*
FW Contexts in Routed Mode					
eBGP	**<1 sec.** *<1 sec.*	**<1 sec.** *<1 sec.*	**~3-4 sec.** *~3-4 sec.*	**~3-4 sec.** *~3-4 sec.*	**<1 sec.** *<1 sec.*

1. **Bold** results are for core to services flows and *italic* results are for services to core flows.

From an analysis of the convergence results, it is possible to draw the following conclusions:

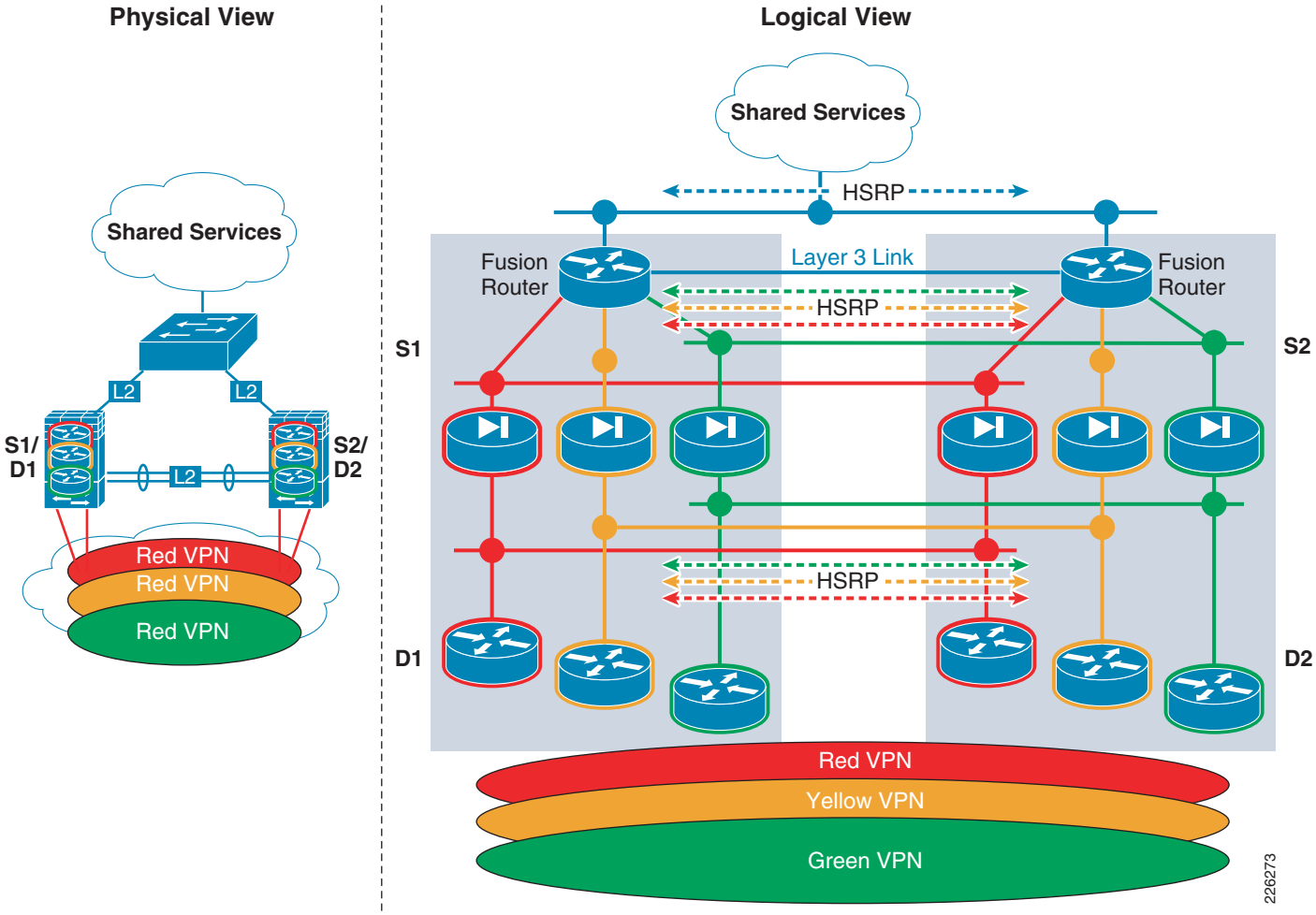
- The use of eBGP is recommended with firewall contexts deployed in routed mode. Recovery times are not sub-second in this scenario in only two circumstances:
  - Firewall failure—The main factor dictating the convergence in this case is the firewall hold-time. Currently, the minimum configurable value is three seconds for the FWSM, but this could be improved in future releases.
  - Services switch failure—The firewall hold-time is also the main factor responsible for convergence in this case, so the same considerations made above are valid here. In addition, it is possible to improve the specific device resiliency by deploying redundant power supplies, redundant supervisors, etc.

- The use of IGP (EIGRP or OSPF) is recommended with firewall in transparent mode. Similar considerations to those made above can be repeated here. In addition, the recovery time in a couple of failure scenarios (test 1 for EIGRP and test 4 for both EIGRP and OSPF) is dictated by the configured IGP hold-time. There is not much that can be done to shorten this, since the use of sub-second IGP timers is usually not a good idea for the sake of the stability of the overall design.

Single Tier Implementation

The second model for deploying protected services access proposes a single tier implementation as shown in Figure 33.

Figure 33 Single Tier Implementation Model



In this model, all the functionalities (VRF, firewall, and fusion routing) are performed in the same physical network device, labeled as S1/D1 since it performs both roles of distribution and services switch. The role these devices play in the virtualized network mostly depends on the deployed path isolation strategy:

- In MPLS VPN designs, these devices would be deployed as PEs.

- In a VRF-Lite End-to-End scenario, they would just connect to the core via virtualized links (either by using sub-interfaces or Layer 2 trunks with SVIs).
- If deploying VRF-Lite and GRE tunnels, these devices would most likely function as hubs aggregating the GRE tunnels originated on the remote spokes.

In the logical view note how there is no longer a Layer 3 link connecting the VRFs together. This was required in the two tier model in order to avoid the creation of a looped topology, but it is not useful here where the only Layer 2 link is the port-channel between the two switches.

Some additional considerations around the single tier deployment (some are similar to the ones already discussed for the two tier model):

- The fusion router functionality in this model must be implemented defining a dedicated VRF for that purpose, since it is not possible to have a separate physical device performing this role.

Note The default VRF (aka global table) could also be used to perform the fusion routing functionalities.

- The fusion VRFs still peer with a Layer 3 connection. This is achieved by dedicating a specific VLAN for this purpose, which is carried over the Layer 2 port-channel trunk connecting the services switches.

Note The use of a port-channel between the services switches is strongly recommended to offer a resilient connection between the two firewall modules.

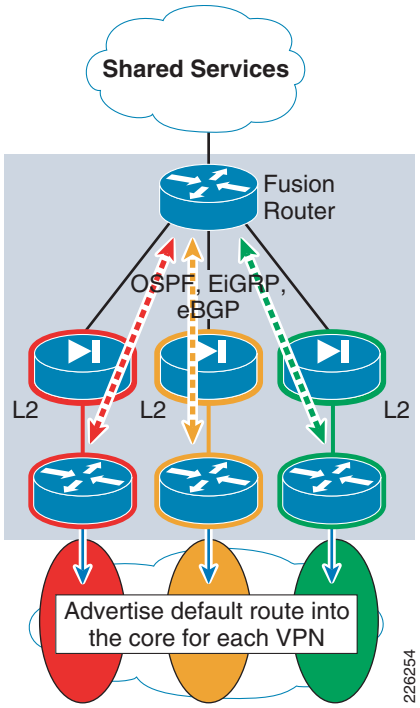
- The inside interface for each firewall context is facing the fusion routers, whereas the outside interface is facing the VRFs. This choice is dictated by the requirement for protecting the services deployed in the shared area.
- Both the firewall outside and inside VLANs are extended between the switches and HSRP is used to provide default gateway functionalities on both subnets.

The following sections highlight the design considerations and the configuration guidelines to deploy the single tier services edge model. Given the fact that the configuration is very similar to the scenarios already discussed for the two tires deployment, only the meaningful differences are highlighted.

Firewall Contexts Deployed in Transparent Mode

As already mentioned when discussing the two tier model, the deployment of firewall contexts in transparent mode allows us to insert them in the network without having to modify the IP addresses already in place. In addition, different types of routing peering can be established between the VRFs defined on the bottom and the fusion router on top, as shown in Figure 34.

Figure 34 Routing Peering with Firewall Contexts in Transparent Mode



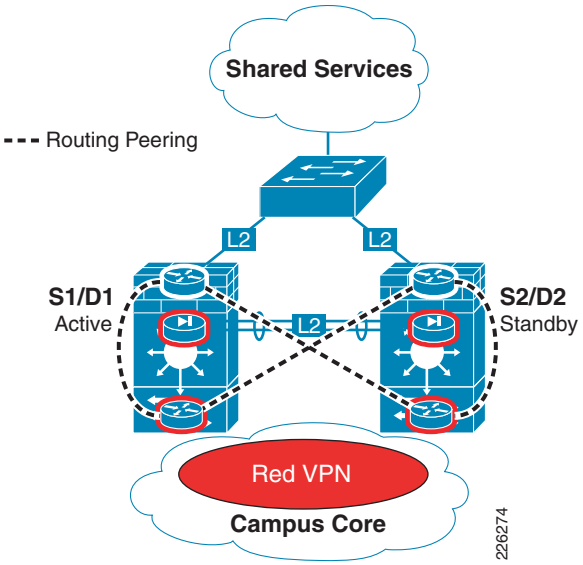
The type of routing protocol used for this functionality is usually dictated by the path isolation strategy deployed:

- For MPLS VPN deployments, the use of eBGP is advocated since iBGP is already in place between the PE devices exchanging VPN routes across the campus infrastructure.
- For VRF-Lite End-to-End (or VRF-Lite + GRE) deployments, typically the same IGP already used as control plane protocol inside each virtual network is also leveraged for this type of peering.

The considerations around traffic flows and recovery scenarios are similar to what was discussed in the two tier deployment. The traffic flows are actually simplified in this case, since there is only a Layer 2 link in the design (the transit link between services switches).

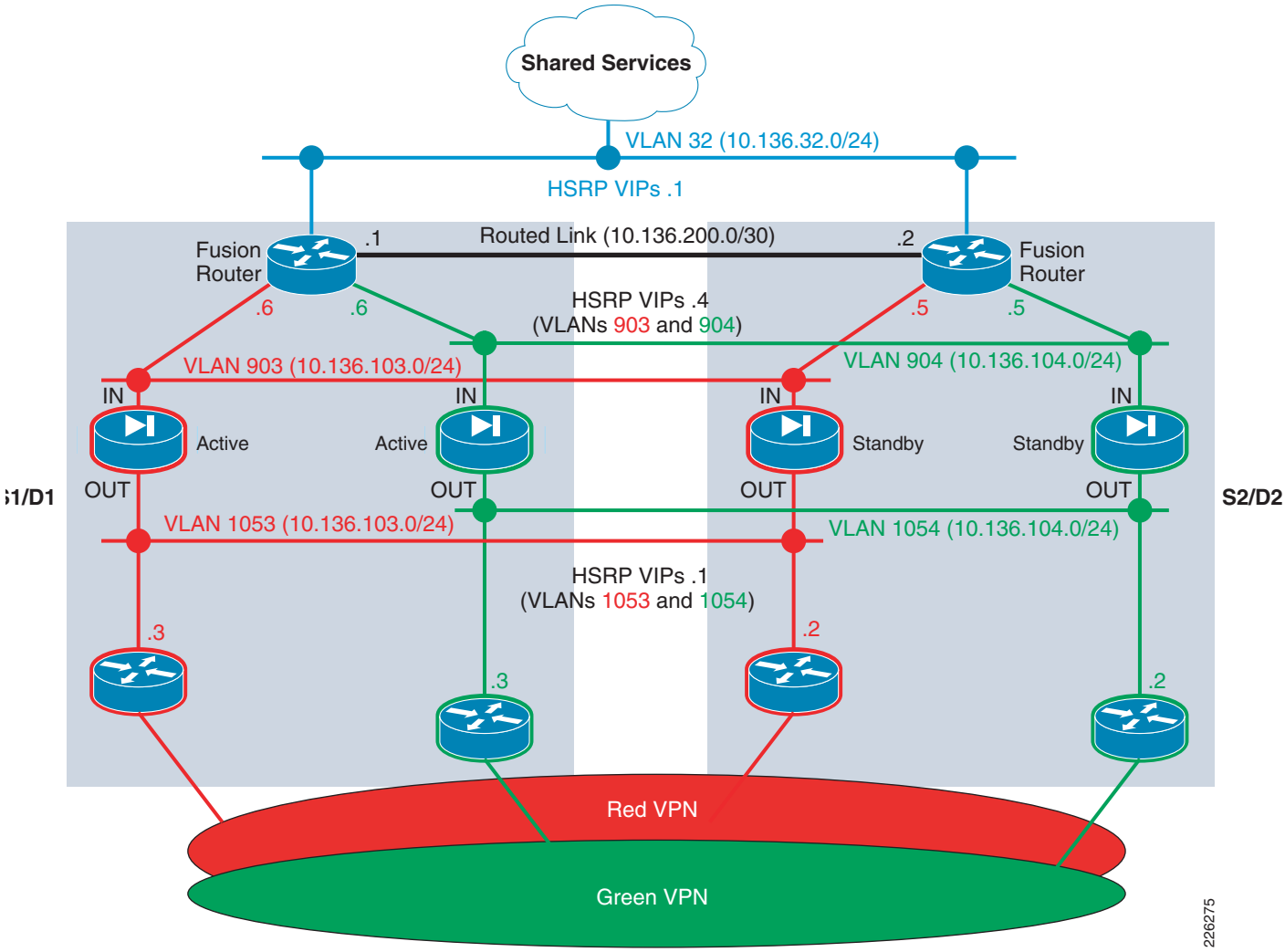
In the following sections, the differences from a configuration standpoint are pointed out to enable routing peering using EIGRP, OSPF, or eBGP. Similar to the two tier design, the goal is to create a full mesh peering between the fusion VRFs and the VRFs defined on the outside of the firewall, as shown in Figure 35.

Figure 35 Full Mesh Routing Peering



The configuration samples all reference the network topology in Figure 36, specifically focusing on the Red VPN.

Figure 36 Firewall Contexts in Transparent Mode Example



Before discussing the protocol specifics, there are some deployment considerations that apply to all three deployments (EIGRP, OSPF, and eBGP) and differentiate the single tier deployment from the dual tier one:

- The SVIs corresponding to the firewall inside and outside subnets need to be defined inside the same device (S1/D1 or S2/D2). This drives the requirement of mapping these SVIs to different VRFs. On the outside we have the Red VRF and on the inside we have the fusion VRFs (or optionally the use of the global table).
- When trying to establish connectivity between SVIs in the same physical device across a firewall context deployed in transparent mode, an additional configuration step is required in order to ensure that ARP can work properly. The root of the issue is that by default all the SVIs defined in a Catalyst 6500 device inherit the same MAC address, as shown in the example below:

```
S1/D1
S1/D1#sh int vlan 903
Vlan903 is up, line protocol is up
  Hardware is EtherSVI, address is 000b.4594.1c00 (bia 000b.4594.1c00)
  Description: Firewall_Inside_Red
<snip>
S1/D1#sh int vlan 1053
```

```
Vlan1053 is up, line protocol is up
  Hardware is EtherSVI, address is 000b.4594.1c00 (bia 000b.4594.1c00)
  Description: Firewall_Outside_Red
<snip>
```

The consequence is that ARP resolution would fail when a packet must be sent from the firewall outside to the firewall inside subnets (or vice versa). Enabling the debug for the ARP protocol unveils the root of the connectivity issue:

```
Dec 11 09:44:02.855 EST: IP ARP: creating incomplete entry for IP address:
10.136.103.6 interface Vlan1053
Dec 11 09:44:02.855 EST: IP ARP: sent req src 10.136.103.3 000b.4594.1c00,
                 dst 10.136.103.6 0000.0000.0000 Vlan1053
Dec 11 09:44:02.855 EST: IP ARP req filtered src 10.136.103.3 000b.4594.1c00, dst
10.136.103.6 0000.0000.0000 it's our address
```

In order to overcome this problem, it is recommended to manually configure the MAC addresses for the inside and outside SVIs, as shown in the configuration sample below:

S1/D1

```
interface Vlan903
 description Firewall_Inside_Red
 mac-address 0000.0000.0903
!
interface Vlan1053
 description Firewall_Outside_Red
 mac-address 0000.0000.1053
```

S2/D2

```
interface Vlan903
 description Firewall_Inside_Red
 mac-address 0000.0001.0903
!
interface Vlan1053
 description Firewall_Outside_Red
 mac-address 0000.0001.1053
```

The specific configuration and design consideration, depending on which routing protocol is used for establishing the routing peering, are discussed in the following sections.

Use of EIGRP between VRFs and Fusion Router

All the considerations discussed in the EIGRP specific section of the two tier deployment model are also valid in the single tier scenario. The only thing to keep in mind is that now the routing instances for the fusion VRFs and the outside VRFs are enabled in the same device, leading to the configuration shown below (this is valid for the S1/D1 device):

S1/D1

```
router eigrp 100
!
address-family ipv4 Red
  network 10.0.0.0
  no auto-summary
  autonomous-system 100
  eigrp router-id 10.136.203.1
 exit-address-family
 !
 address-family ipv4 vrf fusion
  redistribute static metric 200000 100 255 1 1500
  network 10.0.0.0
```

```
  distribute-list Default out Vlan903
  no auto-summary
  autonomous-system 100
  eigrp router-id 10.136.200.1
 exit-address-family
```

The same aggressive EIGRP timer setting is also recommended in this case, as shown in the example below for the firewall inside SVI:

S1/D1

```
interface Vlan903
 mac-address 0000.0000.0903
 ip vrf forwarding Red
 ip address 10.136.103.6 255.255.255.0
 ip hello-interval eigrp 100 1
 ip hold-time eigrp 100 3
```

Since the EIGRP protocol is automatically allowed through the firewall running in transparent mode once the ethertype ACL discussed in two tier model section is applied to the inside and outside interfaces, the EIGRP adjacencies are established between the fusion and outside VRFs, as highlighted below:

S1/D1

```
S1/D1#sh ip eigrp vrf Red neighbors
IP-EIGRP neighbors for process 100
H   Address                 Interface       Hold Uptime   SRTT   RTO  Q  Seq
                                            (sec)         (ms)       Cnt Num
4   10.136.103.2            Vl1053             2 15:45:20    1   200  0  474
3   10.136.103.6            Vl1053             2 15:47:41    4   200  0  22
2   10.136.103.5            Vl1053             2 15:47:41    1   200  0  1194
1   10.136.0.34             Te1/1.332         14 15:47:42  214  1284  0  598
0   10.136.0.38             Te1/2.432         11 15:47:42  137   822  0  41752890
S1/D1#sh ip eigrp vrf fusion neighbors
IP-EIGRP neighbors for process 100
H   Address                 Interface       Hold Uptime   SRTT   RTO  Q  Seq
                                            (sec)         (ms)       Cnt Num
3   10.136.200.2            Vl200             12 00:00:04  521  3126  0  1205
2   10.136.103.2            Vl903              2 15:48:04    2   200  0  475
1   10.136.103.3            Vl903              2 15:50:26    1   200  0  41
0   10.136.103.5            Vl903              2 15:50:26    1   200  0  1204
```

S2/D2

```
S2/D2#sh ip eigrp vrf Red neighbors
IP-EIGRP neighbors for process 100
H   Address                 Interface       Hold Uptime   SRTT   RTO  Q  Seq
                                            (sec)         (ms)       Cnt Num
2   10.136.103.6            Vl1053             2 15:49:00    4   200  0  31
1   10.136.103.3            Vl1053             2 15:49:00    3   200  0  41
0   10.136.103.5            Vl1053             2 15:49:00    4   200  0  1204
4   10.136.0.30             Te1/2.332         12 1d10h       1   200  0  41752889
3   10.136.0.36             Te1/1.432         12 1d10h       1   200  0  596
S2/D2#sh ip eigrp vrf fusion neighbors
IP-EIGRP neighbors for process 100
H   Address                 Interface       Hold Uptime   SRTT   RTO  Q  Seq
                                            (sec)         (ms)       Cnt Num
3   10.136.200.1            Vl200             11 00:02:32    1   200  0  30
0   10.136.103.2            Vl903              2 15:50:32    4   200  0  473
2   10.136.103.3            Vl903              2 15:52:54  170  1020  0  41
1   10.136.103.6            Vl903              2 15:52:54  137   822  0  31
```

Note how the outside VRFs peer with each other and with the fusion VRFs via SVI 1053. They also peer to the devices deployed in the core via routed links deployed defining sub-interfaces of Ten1/1 and Ten1/2. At the same time, the fusion VRFs peer with each other and with the outside VRFs via SVI 903. They also peer directly via SVI 200. This Layer 3 peering is required to provide re-routing in specifics failure scenarios.

Note Refer to the EIGRP specific section in the two tier deployment model for additional details on configuration and design recommendations.

Use of OSPF between VRFs and Fusion Router

Similar to EIGRP, for OSPF most of the configuration steps are identical to the two tier deployment. Once again, the OSPF processes associated to the fusion and outside VRFs are now enabled inside the same device, as highlighted below (this configuration sample is specifically valid for the S1/D1 device):

S1/D1

```
router ospf 4 vrf Red
 router-id 10.136.103.1
 log-adjacency-changes
 auto-cost reference-bandwidth 10000
 capability vrf-lite
 timers throttle spf 10 100 5000
 timers throttle lsa all 10 100 5000
 timers lsa arrival 80
 passive-interface default
 no passive-interface TenGigabitEthernet1/1.342
 no passive-interface TenGigabitEthernet1/2.442
 no passive-interface Vlan1053
 network 10.136.0.0 0.0.255.255 area 136
!
router ospf 100 vrf fusion
 router-id 10.136.100.1
 log-adjacency-changes
 auto-cost reference-bandwidth 10000
 capability vrf-lite
 timers throttle spf 10 100 5000
 timers throttle lsa all 10 100 5000
 timers lsa arrival 80
 passive-interface default
 no passive-interface Vlan903
 network 10.136.0.0 0.0.255.255 area 136
 default-information originate always metric 10 metric-type 1
```

The OSPF timer aggressive tuning is still recommended, as shown below for the inside SVI example:

```
interface Vlan903
 mac-address 0000.0000.0903
 ip vrf forwarding fusion
 ip address 10.136.103.6 255.255.255.0
 ip ospf hello-interval 1
```

Note that, unlike the EIGRP protocol, OSPF packets are not allowed through the firewall running in transparent mode by the ethertype ACL. A specific ACE is required to allow OSPF through the firewall, as shown below:

```
access-list <ACL_NAME> extended permit ospf any any
```

As a consequence of the configuration steps above, the OSPF adjacencies are established between the fusion VRFs and the outside VRFs, as highlighted below:

S1/D1

```
S1/D1 #sh ip ospf 4 neighbor
Neighbor ID     Pri   State           Dead Time   Address         Interface
10.136.100.1      1   FULL/DROTHER    00:00:03    10.136.104.6    Vlan1054
10.136.100.2      1   FULL/DROTHER    00:00:03    10.136.104.5    Vlan1054
10.136.104.2      1   FULL/DR         00:00:03    10.136.104.2    Vlan1054
10.136.4.2        1   FULL/BDR        00:00:38    10.136.0.48
TenGigabitEthernet1/2.442
10.136.4.1        1   FULL/BDR        00:00:38    10.136.0.44
TenGigabitEthernet1/1.342
S1/D1 #sh ip ospf 100 neighbor

Neighbor ID     Pri   State           Dead Time   Address         Interface
10.136.100.2      1   FULL/DR         00:00:03    10.136.200.2    Vlan200
10.136.100.2      1   2WAY/DROTHER    00:00:03    10.136.104.5    Vlan904
10.136.104.1      1   FULL/BDR        00:00:03    10.136.104.3    Vlan904
10.136.104.2      1   FULL/DR         00:00:03    10.136.104.2    Vlan904
```

S2/D2

```
S2/D2 #sh ip ospf 4 neighbor
Neighbor ID     Pri   State           Dead Time   Address         Interface
10.136.100.1      1   FULL/DROTHER    00:00:03    10.136.104.6    Vlan1054
10.136.100.2      1   FULL/DROTHER    00:00:03    10.136.104.5    Vlan1054
10.136.104.1      1   FULL/BDR        00:00:03    10.136.104.3    Vlan1054
10.136.4.2        1   FULL/DR         00:00:36    10.136.0.40
TenGigabitEthernet1/2.342
10.136.4.1        1   FULL/DR         00:00:36    10.136.0.46
TenGigabitEthernet1/1.442
S2/D2 #sh ip ospf 100 neighbor
Neighbor ID     Pri   State           Dead Time   Address         Interface
10.136.100.1      1   FULL/BDR        00:00:03    10.136.200.1    Vlan200
10.136.100.1      1   2WAY/DROTHER    00:00:03    10.136.104.6    Vlan904
10.136.104.1      1   FULL/BDR        00:00:03    10.136.104.3    Vlan904
10.136.104.2      1   FULL/DR         00:00:03    10.136.104.2    Vlan904
```

Note how the outside VRFs peer with each other and with the fusion VRFs via SVI 1053. They also peer to the devices deployed in the core (via sub-interfaces of Ten1/1 and Ten1/2 in the specific example above).

The fusion VRFs peer with each other and with the outside VRFs via SVI 903. They also peer with each other over SVI 200. Once again, this Layer 3 peering is required in specific failure recovery scenarios.

Note Refer to the OSPF specific section of the two tier deployment model for additional details on configuration and design recommendations.

Use of eBGP between VRFs and Fusion Router

The collapse of the fusion router, firewall, and VRF termination functionalities on a single box brings interesting challenges when trying to leverage eBGP for establishing routing peering sessions across the firewall contexts.

Two specific features are required for eBGP to work in a single box scenario:

- Providing a separate BGP router ID per VRF address family.
- Providing separate BGP router processes in the same device to allow the establishment of eBGP sessions.

Note Cisco IOS by default does not allow the creation of more than one BGP process in the same device.

New functionalities have been delivered for Catalyst 6500 platforms starting with software release 12.2(33)SXH that allow us to fulfill both requirements listed above. The required configuration to allow the establishment of the eBGP adjacencies across the firewall context is shown below:

S1/D1

```
router bgp 100
 timers bgp 2 10
 !
 address-family ipv4 vrf Red
  neighbor 10.136.103.5 remote-as 101
  neighbor 10.136.103.5 local-as 200 no-prepend replace-as
  neighbor 10.136.103.5 activate
  neighbor 10.136.103.6 remote-as 101
  neighbor 10.136.103.6 local-as 200 no-prepend replace-as
  neighbor 10.136.103.6 activate
  maximum-paths 2
  no synchronization
  bgp router-id 10.136.200.1
 exit-address-family
!
address-family ipv4 vrf fusion
  neighbor 10.136.103.2 remote-as 200
  neighbor 10.136.103.2 local-as 101 no-prepend replace-as
  neighbor 10.136.103.2 activate
  neighbor 10.136.103.2 default-originate route-map default_only
  neighbor 10.136.103.2 route-map default_only out
  neighbor 10.136.103.3 remote-as 200
  neighbor 10.136.103.3 local-as 101 no-prepend replace-as
  neighbor 10.136.103.3 activate
  neighbor 10.136.103.3 default-originate route-map default_only
  neighbor 10.136.103.3 route-map default_only out
  neighbor 10.136.103.5 remote-as 100
  neighbor 10.136.103.5 activate
  maximum-paths 2
  no synchronization
  bgp router-id 10.136.100.1
 exit-address-family
```

S2/D2

```
router bgp 100
 timers bgp 2 10
 !
 address-family ipv4 vrf Red
  neighbor 10.136.103.5 remote-as 101
  neighbor 10.136.103.5 local-as 200 no-prepend replace-as
  neighbor 10.136.103.5 activate
  neighbor 10.136.103.6 remote-as 101
  neighbor 10.136.103.6 local-as 200 no-prepend replace-as
  neighbor 10.136.103.6 activate
  maximum-paths 2
  no synchronization
  bgp router-id 10.136.200.2
 exit-address-family
 !
 address-family ipv4 vrf fusion
  neighbor 10.136.103.2 remote-as 200
  neighbor 10.136.103.2 local-as 101 no-prepend replace-as
  neighbor 10.136.103.2 activate
```

```
 neighbor 10.136.103.2 default-originate route-map default_only
 neighbor 10.136.103.2 route-map default_only out
 neighbor 10.136.103.3 remote-as 200
 neighbor 10.136.103.3 local-as 101 no-prepend replace-as
 neighbor 10.136.103.3 activate
 neighbor 10.136.103.3 default-originate route-map default_only
 neighbor 10.136.103.3 route-map default_only out
 neighbor 10.136.103.6 remote-as 100
 neighbor 10.136.103.6 activate
 maximum-paths 2
 no synchronization
 bgp router-id 10.136.100.2
exit-address-family
```

Some important considerations around the configuration samples above (these are in addition to the ones made in the eBGP section of the two tier model that are still valid here):

- Despite the fact that a single BGP process is still configured (router bgp 100), the configuration of BGP neighbors is done leveraging two different AS numbers (101 and 200) since the goal is to establish eBGP sessions between the fusion and Red VRFs. In order for this to work, it is required to have BGP support for dual AS configuration. This is achieved leveraging the **local-as** command. The result is that the BGP instance for the Red VRF tries to establish an eBGP neighborship with a system in the remote AS 101, but accepts neighborship requests for the local AS 200. The reverse is valid for the BGP instance in the fusion VRF.

Note For more information on the BGP support for dual AS configuration, refer to: http://www.cisco.com/en/US/docs/ios/12_3t/12_3t11/feature/guide/gtbgpdas.html

- A separate BGP router ID is configured for the fusion and Red VRFs address families.
- The eBGP sessions are established in a fully meshed fashion between the fusion and outside VRFs, following the design principle highlighted in Figure 35.
- An additional iBGP peering is established between the fusion VRFs. This is required to provide re-routing capabilities in a failure scenario where a fusion router loses direct connectivity to the shared services area. Note how the AS number specified for establishing this iBGP peering is the same of the overall BGP process configured in the box (AS 100).

Before eBGP peering sessions can be successfully established, these packets must be allowed through the firewall, as shown below:

```
access-list <ACL_NAME> extended permit tcp any any eq bgp
```

As a consequence of the configuration steps above, eBGP sessions are established between the fusion and the outside VRFs, as highlighted below:

S1/D1

```
S1/D1#sh ip bgp vpnv4 vrf Red summary
BGP router identifier 10.136.200.1, local AS number 100
<snip>
Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
10.136.103.5    4   101   29782   29805  3441362    0    0 16:56:28        1
10.136.103.6    4   101   29780   29805  3441362    0    0 16:56:29        1
S1/D1#sh ip bgp vpnv4 vrf fusion summary
BGP router identifier 10.136.100.1, local AS number 100
<snip>
Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
10.136.103.2    4   200   29906   29891  3453884    0    0 17:00:16       39
10.136.103.3    4   200   29916   29891  3453884    0    0 17:00:17       39
10.136.103.5    4   100 1451545 1351372  3453884    0    0 17:00:11      103
```

S2/D2

```
S2/D2#sh ip bgp vpnv4 vrf Red summary
BGP router identifier 10.136.200.2, local AS number 100
<snip>
Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
10.136.103.5    4   101   40920   41092  3533870    0    0 17:16:42        1
10.136.103.6    4   101   40292   40505  3533870    0    0 17:09:24        1
S2/D2#sh ip bgp vpnv4 vrf fusion summary
BGP router identifier 10.136.100.2, local AS number 100
<snip>
Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
10.136.103.2    4   200  105398  105151  3538990    0    0 17:18:15       39
10.136.103.3    4   200  194634  194165  3538990    0    0 17:10:57       39
10.136.103.6    4   100 8935809 9021284  3538990    0    0 17:10:53      104
```

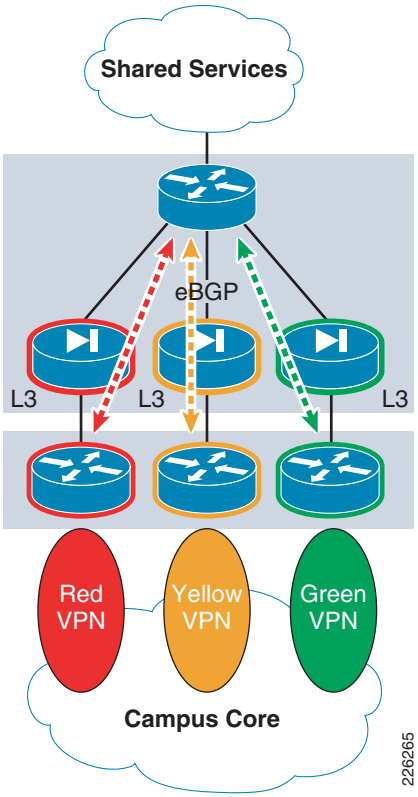
As noted above, the outside VRFs peer with the fusion VRFs, as these were in the remote AS 101. At the same time, the fusion VRFs peer with the outside VRFs, as they were in AS 200. The iBGP peering between fusion VRFs is instead performed with AS 100 (the "real" BGP AS).

Note Refer to the eBGP specific section of the two tier deployment model for additional details on configuration and design recommendations.

Firewall Contexts Deployed in Routed Mode

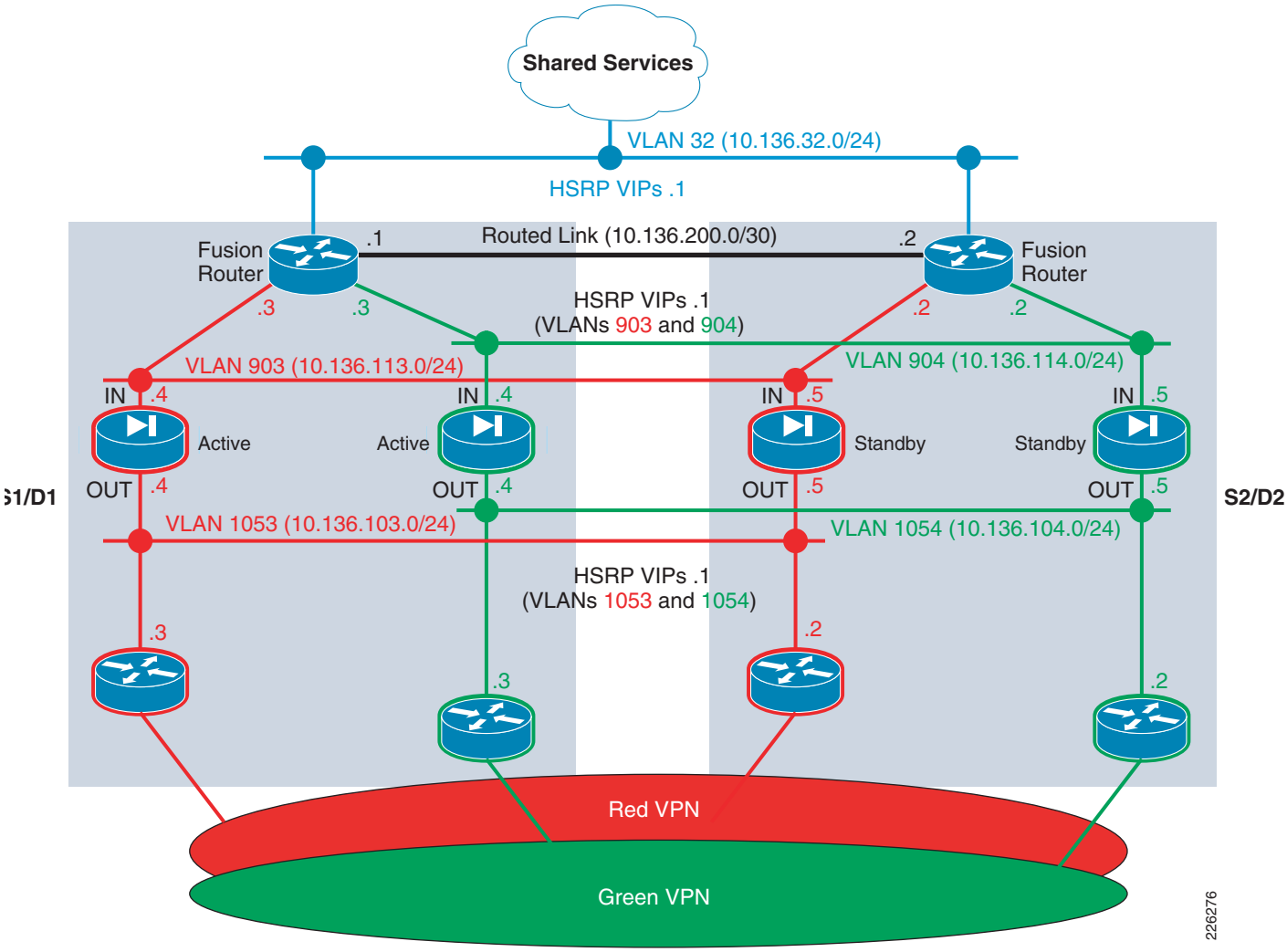
Similar to what was discussed for the dual tier deployment model, the recommended protocol of choice when deploying the firewall contexts in routed mode is eBGP, as shown in Figure 37.

Figure 37 Routing Peering with Firewall Contexts in Routed Mode



The corresponding recommended deployment model for a single tier scenario is shown in Figure 38.

Figure 38 Firewall Contexts in Routed Mode Example



The following considerations can be drawn from Figure 38:

- Each firewall context is now a routed hop in the network. This implies that the firewall inside and outside VLANs now belong to two different IP subnets.
- HSRP is still the First Hop Redundancy Protocol of choice and it is leveraged to provide virtual gateway functionalities for the following interfaces:
  - Shared services subnet (VLAN 32)—This is under the assumption that the shared services are deployed in a subnet directly connected to the fusion devices.
  - Firewall inside subnets (VLAN 903 for the Red VPN and VLAN 904 for the Green VPN).
  - Firewall outside subnets (VLAN 1053 for the Red VPN and VLAN 1054 for the Green VPN).
- The fusion VRFs are still connected with a routed link. This is to provide a Layer 3 path to be used for re-routing traffic under specific failure conditions.

Assuming that the policies on each firewall context have been properly deployed so that the routing protocol of choice is allowed through (the configuration required for eBGP is discussed in the protocol specific section later in this document), each VRF peers with both fusion VRFs, as highlighted in Figure 35.

The following section discusses the specific design and configuration consideration for eBGP deployment and also lists the convergence results achieved for all the failure scenarios previously discussed.

Use of eBGP between VRFs and Fusion Router

The same challenges of a single tier eBGP deployment already discussed for the firewall transparent mode deployment still applies here. The same solution, based on the use of a separate BGP router ID per VRF and the BGP support for dual AS configuration, can also be leveraged with the firewall working in routed mode. The corresponding configuration is highlighted below:

S1/D1

```
router bgp 100
 timers bgp 2 10
 !
 address-family ipv4 vrf Red
  neighbor 10.136.103.2 remote-as 101
  neighbor 10.136.103.2 local-as 200 no-prepend replace-as
  neighbor 10.136.103.2 ebgp-multihop 2
  neighbor 10.136.103.2 activate
  neighbor 10.136.103.3 remote-as 101
  neighbor 10.136.103.3 local-as 200 no-prepend replace-as
  neighbor 10.136.103.3 ebgp-multihop 2
  neighbor 10.136.103.3 activate
  maximum-paths 2
  no synchronization
  bgp router-id 10.136.206.1
 exit-address-family
!
address-family ipv4 vrf fusion
  neighbor 10.136.103.2 remote-as 100
  neighbor 10.136.103.2 activate
  neighbor 10.136.113.2 remote-as 200
  neighbor 10.136.113.2 local-as 101 no-prepend replace-as
  neighbor 10.136.113.2 ebgp-multihop 2
  neighbor 10.136.113.2 activate
  neighbor 10.136.113.2 default-originate route-map default_only
  neighbor 10.136.113.2 route-map default_only out
  neighbor 10.136.113.3 remote-as 200
  neighbor 10.136.113.3 local-as 101 no-prepend replace-as
  neighbor 10.136.113.3 ebgp-multihop 2
  neighbor 10.136.113.3 activate
  neighbor 10.136.113.3 default-originate route-map default_only
  neighbor 10.136.113.3 route-map default_only out
  maximum-paths 2
  no synchronization
  bgp router-id 10.136.100.1
 exit-address-family
```

S2/D2

```
router bgp 100
 timers bgp 2 10
 !
 address-family ipv4 vrf Red
  neighbor 10.136.103.2 remote-as 101
  neighbor 10.136.103.2 local-as 200 no-prepend replace-as
  neighbor 10.136.103.2 ebgp-multihop 2
  neighbor 10.136.103.2 activate
  neighbor 10.136.103.3 remote-as 101
  neighbor 10.136.103.3 local-as 200 no-prepend replace-as
  neighbor 10.136.103.3 ebgp-multihop 2
```

```
 neighbor 10.136.103.3 activate
 maximum-paths 2
 no synchronization
 bgp router-id 10.136.200.2
exit-address-family
!
address-family ipv4 vrf fusion
 neighbor 10.136.103.3 remote-as 100
 neighbor 10.136.103.3 activate
 neighbor 10.136.113.2 remote-as 200
 neighbor 10.136.113.2 local-as 101 no-prepend replace-as
 neighbor 10.136.113.2 ebgp-multihop 2
 neighbor 10.136.113.2 activate
 neighbor 10.136.113.2 default-originate route-map default_only
 neighbor 10.136.113.2 route-map default_only out
 neighbor 10.136.113.3 remote-as 200
 neighbor 10.136.113.3 local-as 101 no-prepend replace-as
 neighbor 10.136.113.3 ebgp-multihop 2
 neighbor 10.136.113.3 activate
 neighbor 10.136.113.3 default-originate route-map default_only
 neighbor 10.136.113.3 route-map default_only out
 maximum-paths 2
 no synchronization
 bgp router-id 10.136.100.2
exit-address-family
```

The end result of the configuration above is the establishment of eBGP peering, similar to the transparent deployment case:

S1/D1

```
S1/D1#sh ip bgp vpnv4 vrf Red summary
BGP router identifier 10.136.200.1, local AS number 100
<snip>
Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
10.136.106.2    4   101   33549 1195471  3879287    0    0 03:32:54        1
10.136.106.3    4   101   33559 1195843  3879287    0    0 03:33:19        1
S1/D1#sh ip bgp vpnv4 vrf fusion summary
BGP router identifier 10.136.100.1, local AS number 100
<snip>
Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
10.136.106.2    4   100 1632771 1520788  3886064    0    0 19:08:03      104
10.136.116.2    4   200 1205825   33618  3886095    0    0 03:35:10       39
10.136.116.3    4   200 1197863   33624  3886095    0    0 03:35:24       39
```

S2/D2

```
S2/D2#sh ip bgp vpnv4 vrf Red summary
BGP router identifier 10.136.200.2, local AS number 100
<snip>
Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
10.136.106.2    4   101   37514 1217035  3943814    0    0 03:34:45        1
10.136.106.3    4   101   37227 1213227  3943814    0    0 03:35:55        1
S2/D2#sh ip bgp vpnv4 vrf fusion summary
BGP router identifier 10.136.100.2, local AS number 100
<snip>
Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
10.136.106.3    4   100 4255452 4368163  3945041    0    0 19:09:09      104
10.136.116.2    4   200 5509274  186755  3945069    0    0 03:35:07       38
10.136.116.3    4   200 5239082  174719  3945069    0    0 03:36:06       38
```

Note Refer to the eBGP specific section of the two tier deployment model for additional details on configuration and design recommendations.

Single Tier Implementation—Summary and Design Recommendations

The single tier deployment model just discussed is recommended for smaller deployments, where keeping all the functionalities (fusion routing, firewall services, and VRF termination) in a single physical device can still be operationally feasible. Similar to the two tier scenario, both firewall contexts functioning in transparent or routed mode can be supported and the same routing protocols can be used in the two scenarios, as summarized below:

- Firewall in transparent mode—EIGRP, OSPF, or eBGP
- Firewall in routed mode—eBGP

From a deployment perspective, two things are specific for this single box implementation:

- The MAC address for the VLAN interface defined on the firewall inside and outside subnets needs to be manually configured. This is required to overcome the specific Catalyst behavior of assigning by default the same MAC address to all the defined SVIs.
- When leveraging eBGP as the routing protocol in the services edge, two additional functionalities are mandatory to ensure a successful eBGP peering between VRFs defined in the same physical device—defining a unique BGP route-id per VRF and BGP support for dual AS configuration. Both features are available for Catalyst 6500 platforms running IOS version 12.2(33)SXH and newer.

From a convergence perspective, the results achieved with the single tier deployment are similar to the two tier model. Refer to Table 1 for a summary of results under various failure scenarios (note that only a subset of these failure scenarios apply to the single tier model).

Planning for Site Redundancy in a Virtualized Network

The discussion around the deployment of the services edge to provide protected access to shared resources has been focused on a single-site deployment, where intra-site redundancy is offered by deploying redundant distribution an services switches, redundant fiber connections, redundant firewall modules, etc.

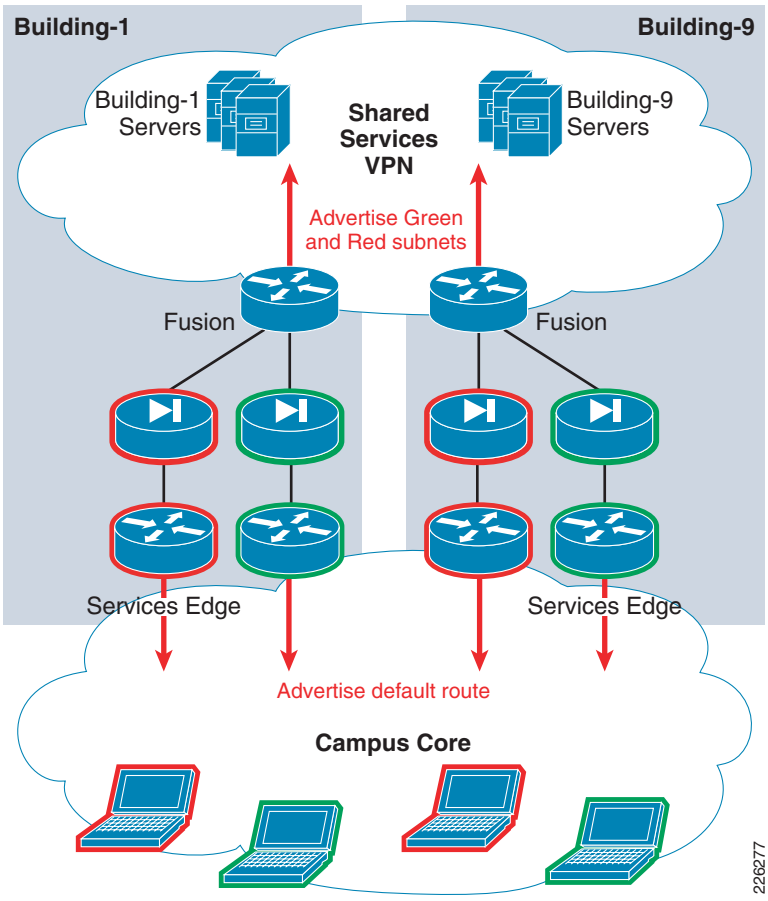
The focus of this section is on discussing the design considerations for providing inter-site redundancy. In this scenario, it is clarified how care needs to be taken to ensure that all traffic streams flowing through the firewalls in the services edge follow a symmetric routing path; that is, traffic to and from an endpoint has to flow through the same firewall so as to pass firewall inspection policies.

Default Redundant Site Configuration

Figure 39 shows a large campus network. There are many buildings throughout the campus, but Building-1 and Building-9 in this example provide the services edge functionalities. Each shared services site shows a single router and firewall for each client VPN, but the shared services design should follow the intra-site redundancy guidance discussed in the previous sections of this document. This means the Red and Green VPNs may be virtual instances on the same router and firewall and that there may be redundant routers and firewalls at each site.

The shared services VPN can be spanned between Building-1 (Site 1) and Building-9 (Site 2) by using a dedicated connection or via a virtualized connection over the same campus infrastructure. Note that a possible option is to leverage the global table (i.e., default VPN) to provide the shared services extension functionalities.

Figure 39 Redundant Services Edge Deployment



Although the actual method for the routing between the fusion routers and the services edge routers is not discussed here, the following can be reasonably assumed;

- Both fusion routers in the shared services VPN advertise routes to the Red and Green VPNs. As previously mentioned, it is typical to configure the fusion router to just advertise a default route to the defined outside VRFs. These default routes are then advertised to all the other campus devices belonging to the defined VPNs.
- The outside VRFs advertise routing information to the fusion router to provide connectivity to the remote campus subnets.

The shared services servers are typically deployed on subnets directly connected to the fusion router. The fusion router is normally their default gateway and it is logical to expect that:

- Shared servers in Building-1 that send traffic destined for the VPNs in the campus core are most likely to send it through the firewall contexts deployed in Building-1 services edge.
- Shared servers in Building-9 that send traffic destined for the VPNs in the campus core are most likely to send it through the firewall contexts deployed in Building-9 services edge.

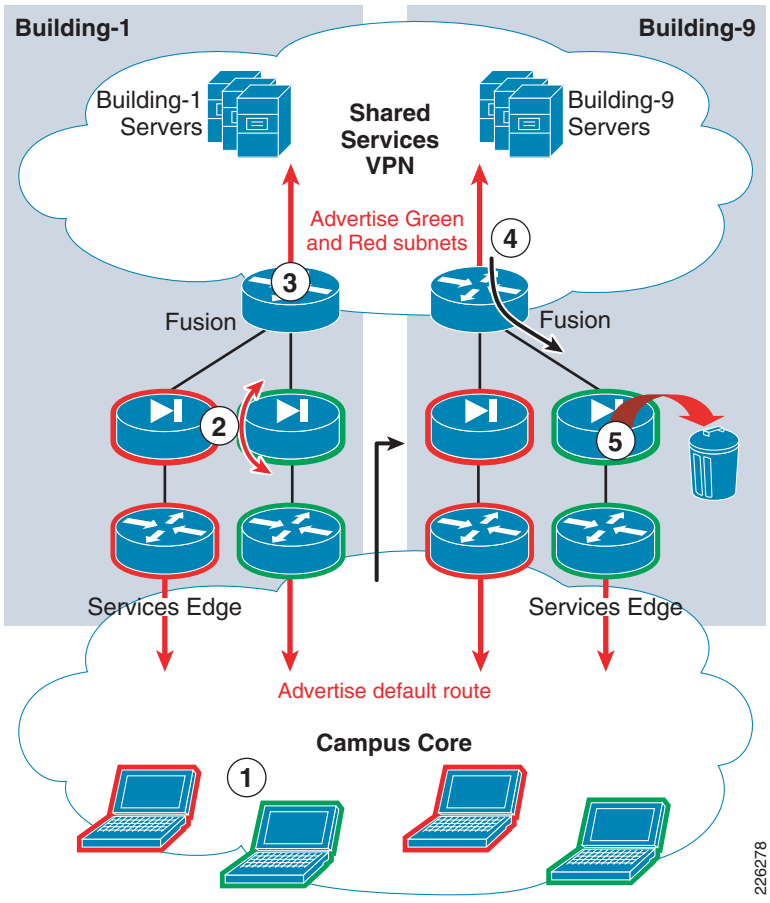
The default route from each of the services edge routers propagates through the network, so every router in the campus core learns the least cost route to the shared services VPN.

- Clients in the campus core that send traffic to the shared service VPN, or to other client VPNs, are routed to the services edge router with the least cost default route metric, as seen by the clients local router.

The Challenge with Asymmetric Routing Paths

Figure 40 illustrates the problem where traffic between the client and the shared services takes different paths.

Figure 40 Asymmetric Routing Paths



This example shows how a client in the campus core is unable to establish a TCP session with a server in the shared services VPN when the traffic from the client is routed to a different services edge than is traffic to the client. There can be many scenarios where this could happen; a specific one is highlighted in Integrating Voice Technologies in a Virtualized Network Infrastructure.

1. The traffic from the client in this example is in a part of the campus core where the default route from the services edge in Building-1 has a lower metric that the default route from Building-9. Traffic from the client follows the default route to the services edge in Building-1.
2. The firewall inspects the traffic and sees a SYN packet originated from the client as part of a TCP handshake

3. The fusion router in Building-1 routes the traffic across the shared services VPN to the destination server in Building-9.
4. The server in Building-9 routes traffic to the client VPN via its local services edge in Building-9.
5. Because the firewall in Building-9 did not see the TCP SYN from the client, it drops the SYN-ACK message originated from the server, preventing the establishment of the connection.

Ensuring Symmetric Routing Paths

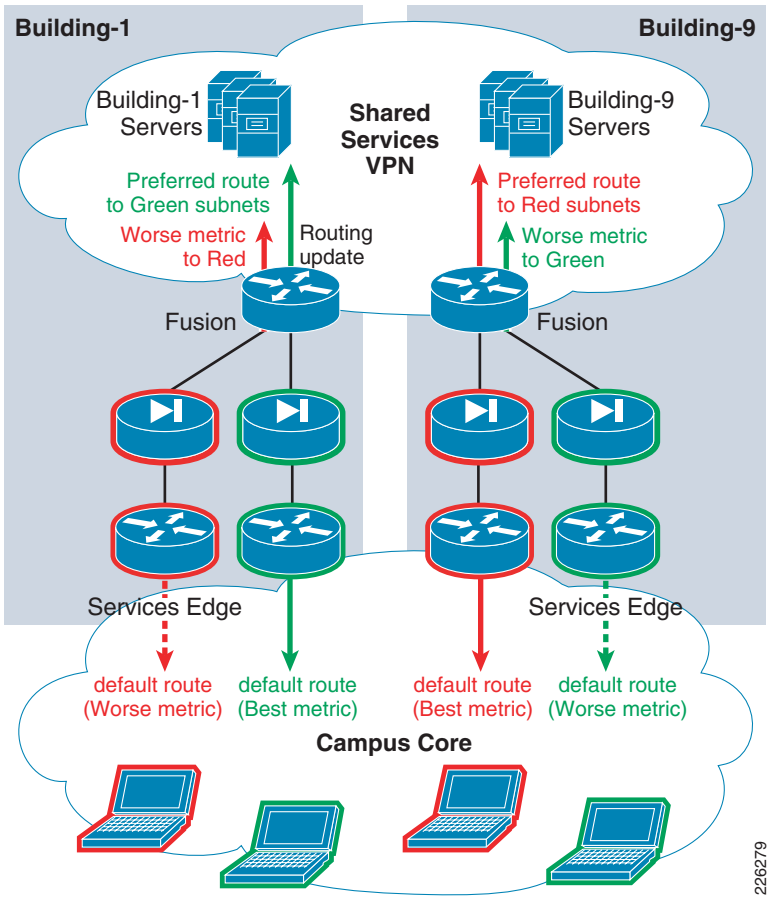
Figure 41 illustrates a solution to prevent asymmetric routing between the client and the shared services area. The goal with this design is to ensure that all packets between a client in the campus core and a server in the shared services VPN are always routed via the same physical services edge site. This ensures that firewall inspection sees bidirectional traffic.

In this design, the traffic has been load balanced so that one services edge site routes half the VPNs and the other site routes the other half. An alternative is to route all VPNs to one site and only use the other as backup. Either approach is valid, but the focus of this document is on the load-balanced design.

In Figure 41 the routing protocol has been tuned so that:

- The shared services VPN always sees a better route to the Green VPN via Building-1 and to the Red VPN via Building-2.
- The campus core always sees a better default route via the services edge in Building-1 for the Green VPN and via the services edge in Building-2 for the Red VPN.

Figure 41 Ensuring Symmetric Routing Paths



The consequences of this tuning on the traffic flows are:

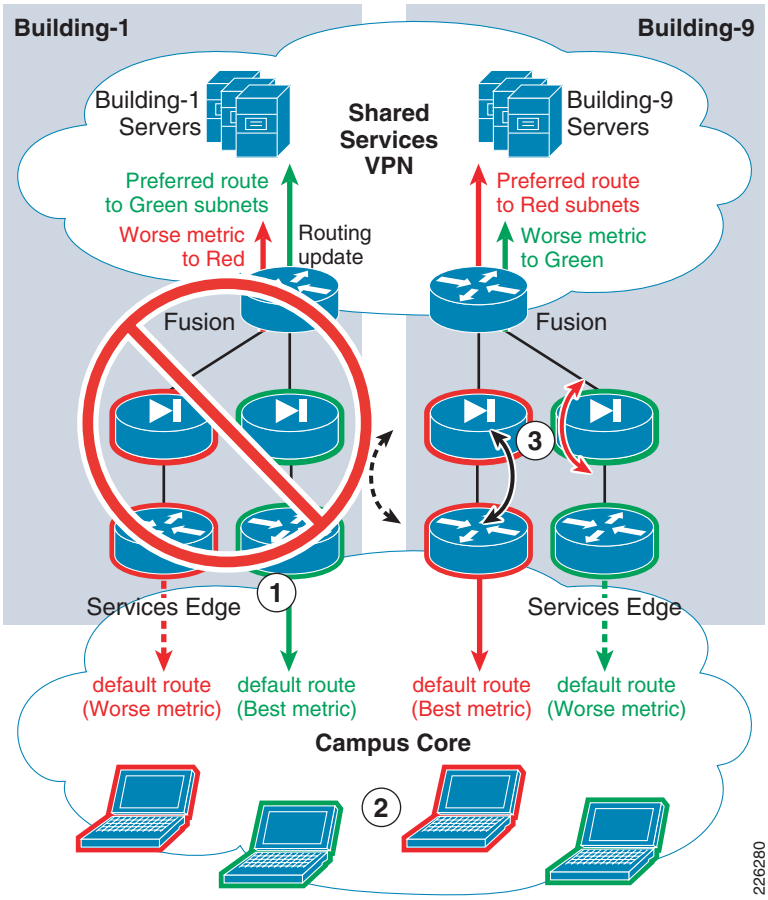
- The Green client sends traffic to a server in Building-9. The default route for this traffic steers it to the Building-1 services edge. From there, it is routed to and within the shared services VPN to the server in Building-9. The routers in the shared services VPN have the lowest cost route to the Green VPN via Building 1, so all traffic to the client goes back via the same path.
- The firewall is able to inspect bidirectional traffic and sees the full TCP handshake and all packets between the client and the server are permitted.

Symmetric Routing Failover

The deployment of intra-site redundancy could provide failure recovery in the event of link or device failure within a site. In this section we look at inter-site redundancy where failure recover is provided in the event of the failure of an entire site.

In this example, highlighted in Figure 42, we see how the routing protocols ensure that the network can survive a failure of the services edge at Building-1.

Figure 42 Service Edge Site Failure

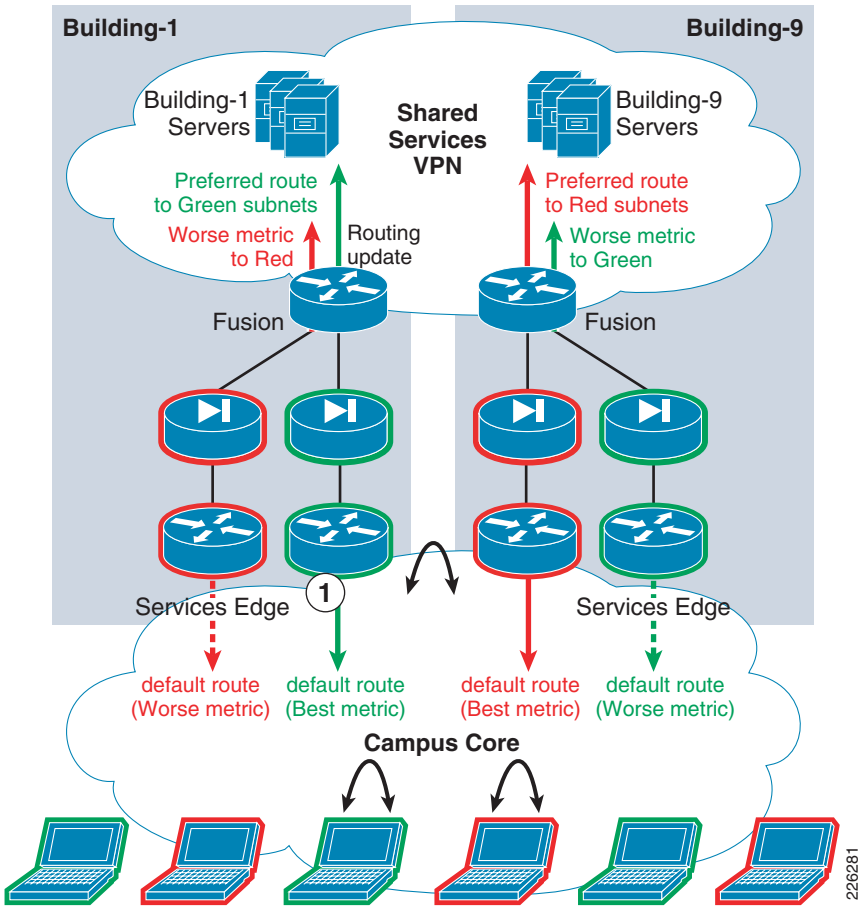


- The client was following the tuned routing paths to and from Building-1 to get to the server in Building-9. Network connectivity to Building-1 is lost.
- The routing protocol re-converges on Building-9 shared services.
  - The fusion router in Building-1 stops advertising routes to the Green VPN and the routes from the fusion router in Building-9 are used instead.
  - The services edge router in Building-1 stops advertising the default route and the default route advertised by the services edge router in Building-9 is used instead.
- The route to and from the client now goes through the services edge in Building-9 and the firewall is able to inspect and permit the traffic.

Symmetric Routing Client Traffic Flows

Figure 43 shows the traffic patterns that could be expected form a redundant network that has had its routing tuned to ensure symmetric traffic paths. In this example we have a Green VPN and a Red VPN.

Figure 43 Symmetric Traffic Flows



- Traffic between any clients in the campus core that are part of the same VPN is unaffected by changes to the inter-VPN routing.
- Traffic between clients in different VPNs has been tuned so that all traffic to and from the Green VPN must go via Site-1 and all traffic to and from the Red VPN must go via Site-2.

In Figure 43, the Green firewall in Site-1 inspects all traffic to or from the Green VPN and the Red firewall in Site-2 does the same thing for traffic to or from the Green VPN.

Configuring Site Redundancy in a Virtualized Network

This section of the document provides configuration guidance for tuning the routing protocols so that traffic to or from a given campus core VPN always uses the same services edge site (ensuring symmetric routing paths).

The specific configuration to be used mainly depends on two factors:

- The firewall context mode of operation—The firewall can be deployed in transparent (or Layer 2 bridging) mode or routed (Layer 3) mode.
- The routing protocol used for exchanging routes between each VRF and the fusion router—This is also dependant on the firewall mode of operation. With firewall in transparent mode all options (EIGRP, OSPF, and eBGP) are possible, whereas with firewall in routed mode eBGP is generally the recommended approach.

The high-level design principle to be implemented in order to achieve the desired behavior is shown in Figure 43. Looking at the specific example above, the fusion router in Building-1 has to:

- Inject a higher metric into the shared VPN for the Red subnets.
- Advertise routes with a higher metric to the Red distribution VRFs (usually only a default route is sent from the fusion router toward the distribution VRFs).

The fusion router in Building-9 has to perform the same operations for the Green VRF.

The following sections provide an example of routing tuning valid for a specific deployment leveraging EIGRP as routing protocol between the fusion router and the distribution VRFs. In this specific example the two following assumptions are also made:

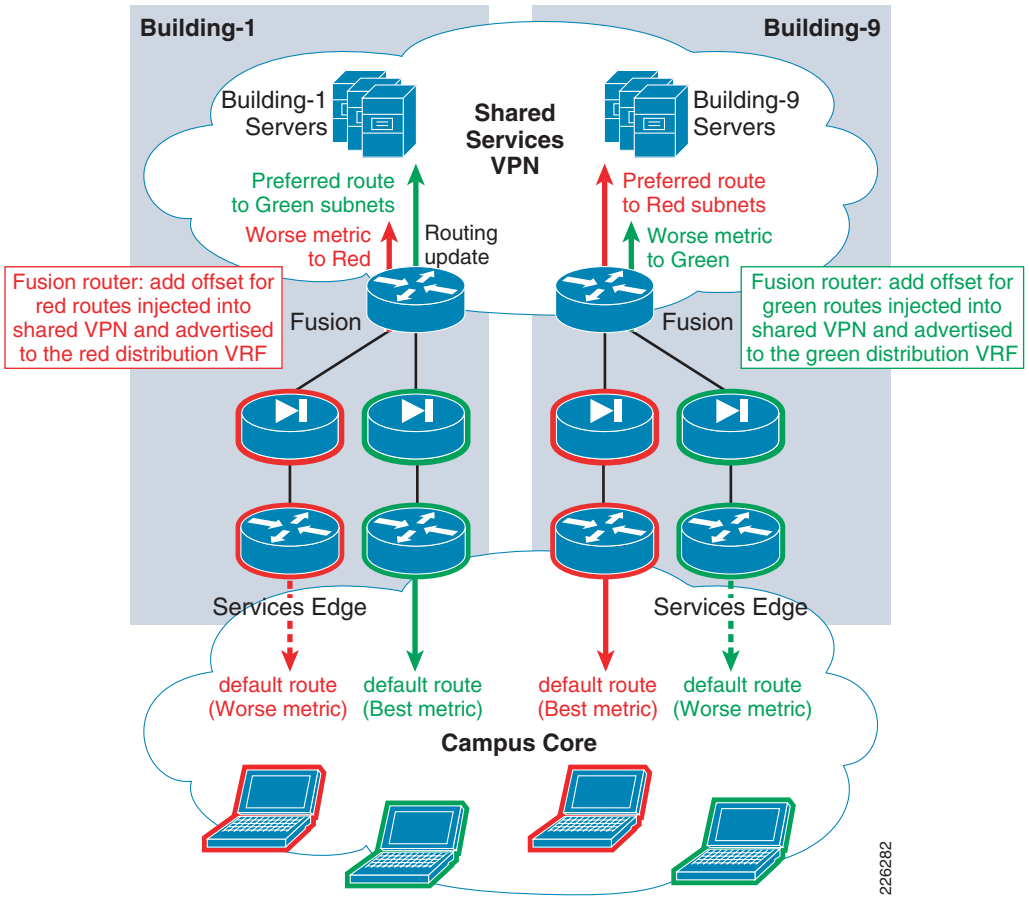
- The firewall are deployed in transparent mode.
- EIGRP is the routing protocol also leveraged inside the shared VPN or across the campus for each specific VPN (i.e., it is a VRF-Lite End-to-End design).

EIGRP—Using Offset-list to Tune Routing

When the firewall is in transparent mode, it bridges traffic between the services edge router and the fusion router. This means that routing protocols on the fusion and the services edge routers can establish neighbor relationships and exchange routing updates with each other.

The scenario leveraging EIGRP as protocol for exchanging routes between the fusion router and the VRFs is highlighted in Figure 44.

Figure 44 Using EIGRP Offset-List



The section below configures offset on one of the router interfaces in the fusion routing instance. The offset needs to be added to all the interfaces to less preferred VPNs in the fusion routing interface. Considering as an example the fusion router in Building-1, it would mean applying the offset-list to the SVI corresponding to the Red firewall inside subnet and to the interface connecting to the shared VPN.

1. Network configuration before configuring the offset-list.

   The output below shows a portion of the routing table from a router in the core of the network. From this we see that there are two equal-cost default routes for each Red and Green VPN. Each of these default routes points to a services edge router in a different site.

```
c1#sh ip route vrf Red
Routing Table: Red
<...>
D*EX 0.0.0.0/0 [170/3072] via 10.1.0.7, 00:00:37, GigabitEthernet1/4.53
               [170/3072] via 10.1.0.5, 00:00:37, GigabitEthernet1/3.63
c1#sh ip route vrf Green
Routing Table: Green
<...>
D*EX 0.0.0.0/0 [170/3072] via 10.2.0.7, 00:00:37, GigabitEthernet1/4.54
               [170/3072] via 10.2.0.5, 00:00:37, GigabitEthernet1/3.64
```

2. Configuring the offset-list.

This is how the change was made. The **offset-list** command adds a positive offset of 1000 to metrics for networks matching the access list on the interface to which it is applied. Note how a different offset-list must be applied to the two interfaces preciously mentioned, given that a default route is usually advertised to the distribution VRFs, whereas the entire address space available in a each given VPN must be injected into the shared VPN.

Fusion router Building-1

```
ip access-list standard Default
permit 0.0.0.0
!
ip access-list standard routes_into_shared_VPN
permit 10.1.0.0 0.0.255.255
!
router eigrp 100
!
address-family ipv4 vrf fusion
 offset-list Default out 1000 VLAN23
 offset-list routes_into_shared_VPN out 1000 vlan23
```

Fusion router Building-9

```
ip access-list standard Default
permit 0.0.0.0
!
ip access-list standard routes_into_shared_VPN
permit 10.2.0.0 0.0.255.255
!
router eigrp 100
!
address-family ipv4 vrf fusion
 offset-list Default out 1000 VLAN24
 offset-list routes_into_shared_VPN out 1000 vlan24
```

Note VLAN 23 in the example above is the SVI associated to the Red firewall inside subnet (in Building-1), whereas VLAN 24 is the SVI associated to the Green firewall inside subnet (in Building-9). 10.1.0.0/16 is the summary for all the routes defined in the campus for the Red VPN, whereas 10.2.0.0/16 is the summary for all the routes defined in the campus for the Green VPN.

3. Results of configuring the offset-list.

This is a portion of the routing table from a router in the core of the network after the change has been made. Now the route with the additional offset applied is no longer shown in the routing table.

```
c1#sh ip route vrf Red
Routing Table: Red
<...>
D*EX 0.0.0.0/0* [170/3072] via 10.1.0.5, 00:03:51, GigabitEthernet1/3.63
c1#sh ip route vrf Green
Routing Table: Green
<...>
D*EX 0.0.0.0/0* [170/3072] via 10.2.0.7, 00:03:51, GigabitEthernet1/4.54
```

As noted above, the default route for the two different VPNs points to two different directions toward Building-1 (for the Green VPN) and Building-9 (for the Red VPN). Assuming also that the shared VPN is extended via the same campus core, it is also possible to note how the return traffic (directed to the specific VPN subnets) is steered:

```
c1#sh ip route vrf fusion
Routing Table: Red
```

```
<...>
D*EX 10.1.0.0/16* [170/3072] via 12.1.0.5, 00:03:51, GigabitEthernet1/3.65
D*EX 10.2.0.0/16* [170/3072] via 12.1.0.7, 00:03:51, GigabitEthernet1/4.55
```

Traffic directed to campus subnets belonging to the Red VPN (10.1.0.0/16 address space) is steered toward Building-9, whereas traffic for Green destinations (10.2.0.0/16 address space) is steered toward Building-1. The overall traffic behavior becomes therefore that depicted in Figure 43.

Virtualized Network Site Redundancy Summary

Virtualization can be deployed with all the normal redundancy provisions. In the first part of this document we discussed how to provide intra-site redundancy for a services edge deployment by leveraging redundant firewall module, redundant switches, and fiber connections.

The ability to utilize redundant services edge sites (i.e., inter-site redundancy) is also available, but care needs to be taken to ensure that all traffic streams follow a symmetric routing path; that is, traffic to and from an endpoint has to flow through the firewall(s) at the same services edge site so as to pass firewall inspection policies.

The routing protocols used at the services edge routers and at the fusion routers need to be tuned so that traffic to or from a given VPN in the campus core always uses the same services edge site.

The design chosen depends on how the firewall between the services edge router is configured and what routing protocol is chosen for exchanging routes between the fusion routers and the distribution VRFs. The firewall mode of operations with the corresponding recommended routing protocols are:

- Transparent firewall mode—EIGRP, OSPF, and eBGP
- Routed firewall mode—eBGP

In the specific case of EIGRP deployments, we have seen how the use of offset-list can represent a simple way to affecting the routing of traffic between a specific VPN and the shared services area (or a different VPN).

Integrating Voice Technologies in a Virtualized Network Infrastructure

Businesses use network virtualization to segment different parts of the organization from each other. This section provides guidance to ensure that all portions of the organization can continue to communicate and collaborate across the campus network, even after network virtualization has been deployed.

Prior to the current version of this document, Unified Communications applications were not tested on a virtualized network infrastructure and the recommendation was to deploy Unified Communications applications in the global table. The current version of this document allows UC to operate outside of the global table, providing an architecture whereby Unified Communications technology such as voice, video, and collaboration can occur within and between different virtual segments (or VPNs) of the organization.

Note The Unified Communications section of the current version of this document focuses on campus deployments. Branch deployments are more challenging as it is undesirable to route inter-VPN VoIP traffic over a WAN link to the services edge. A future revision will address the challenges posed by branch deployments.

Virtualized Unified Communications Traffic Flow Overview

The goal of this section is to provide a network design that allows unified communications endpoints to be fully integrated into a virtualized campus network. Past guidance has been to place all unified communication endpoints and infrastructure into the global table. This chapter shows how the network can be designed to permit unified communications endpoints to exist within multiple VPNs and still be able to securely communicate with each other.

In this design, unified communications servers such as the Cisco Unified Communications Manager are deployed as centralized resources shared between all the defined VPNs. As previously discussed in Deploying Protected Shared Services, firewalls can be used to front-end each VPN and provide tight security policies to control access to the shared area. The figures in this section show individual firewalls, but they may be deployed as virtual instances (contexts) within a single physical firewall.

The following are some initial design assumptions for leveraging this protected services edge model to integrate Unified Communication applications in the virtualized network infrastructure:

- The outside interface of the firewalls are configured with some very specific filters to permit unified communications traffic from the IP telephony endpoint VPNs to reach the servers in the shared services.
- The inside interface of the firewalls permits all traffic from the shared services VPN to the IP telephony endpoint VPNs.
- Traffic that is permitted by the firewall interface filter is inspected using the firewalls security enforcement technologies that include protocol anomaly detection and application and protocol state tracking.

The key technical functionality at the base of the design proposed in this section is the capability of Cisco firewalls to perform traffic inspection. This allows the firewall to inspect call signaling protocols and to dynamically open firewall pinholes to permit traffic between two different VPNs. The firewall dynamically closes the pinhole when the call ends. This functionality is critical because it simplifies the firewall configuration and reduces the number of static holes that would otherwise be required to allow inter-VPN voice streams.

Given the nature of the proposed solution (based on firewall inspection), all the considerations previously made about the need to avoid asymmetric routing in redundant services edge deployments are specifically relevant now. It is critical for the proposed solution to work that the signaling and media traffic for endpoints belonging to a specific VPN always traverses the same firewall context (located in the same services edge physical location). Refer to Planning for Site Redundancy in a Virtualized Network for more details on how to achieve this behavior.

As previously discussed, a fusion router is used in the services edge design to provide IP routed connectivity between the shared services VPN and the other VPNs or even to route traffic between VPNs. In the specific context of the voice integration solution, the fusion router performs the following:

- The fusion router redistributes all routes from the IP telephony endpoint VPNs into the shared services VPN.
- The fusion router advertises a single default route into the IP telephony endpoint VPNs. These VPNs only have routes for subnets within their own VPN and use the default route to the fusion router to reach the shared services VPN or to reach an IP endpoint in any other VPN.

Figure 45 Firewall Inspection for Voice Integration

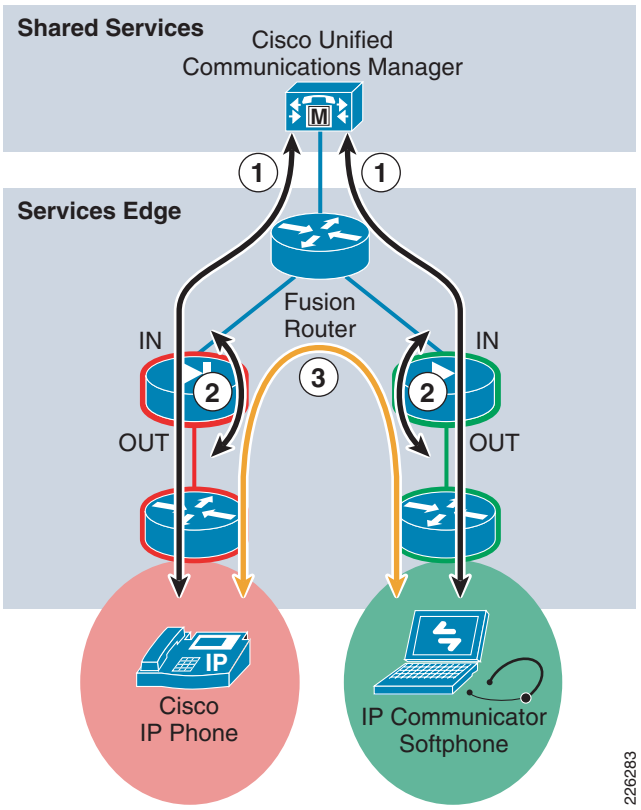


Figure 45 shows the traffic flows and inspection points for a voice call between an IP phone in a voice VPN and a PC-based softphone in a data VPN.

1. When the Red IP phone dials the number for the Green IP phone, signaling messaging is exchanged between it and CUCM. CUCM then signals both the Red and Green phone with the IP address and port that should be used to send RTP voice media directly between the phones.
2. Both the Red and Green Cisco firewalls are able to inspect the signaling and are thus aware of the IP address and UDP port used for the voice RTP media stream. The firewall opens a pinhole to allow the communications between the two endpoints for the duration of the call. The pinhole is dynamically closed when the call terminates.
3. Bidirectional RTP media streams between the two endpoints can now be established via the fusion router.

Desktop Unified Communications Client Virtualization

In this section, the traffic flows and firewall filter requirements are examined in detail for each of the UC applications that were tested for this document. The focus of this section is on unified communications applications, such as softphones and IP video endpoints, that might be deployed in a data VPN and which need to talk to IP phones that might be deployed in a voice VPN. The following UC applications and endpoints were tested and documented:

- Cisco IP phones and Cisco IP Communicator
- Cisco Unified Video Advantage
- Cisco Unified Personal Communicator

- Cisco Unity
- Cisco PSTN gateway

For each of these applications, we examine the signaling and IP data flows required to complete a call. We determine what filters are required on the services edge firewalls to permit the signaling and look at how firewall inspection is able to inspect the signaling and dynamically open and close firewall pinholes to permit the voice or video media between the endpoints.

Cisco Unified IP Phones 7900 Series Phone and the Cisco IP Communicator

In this section we describe testing of the Cisco Unified IP Phones 7900 series and the Cisco IP Communicator softphone. The Cisco IP Communicator emulates the IP protocols used by Cisco Unified IP Phones 7900 series phones. The call signaling and call setup data flows are identical for the Cisco IP Communicator and the Cisco Unified IP Phones 7900 series phones, so they are discussed together.

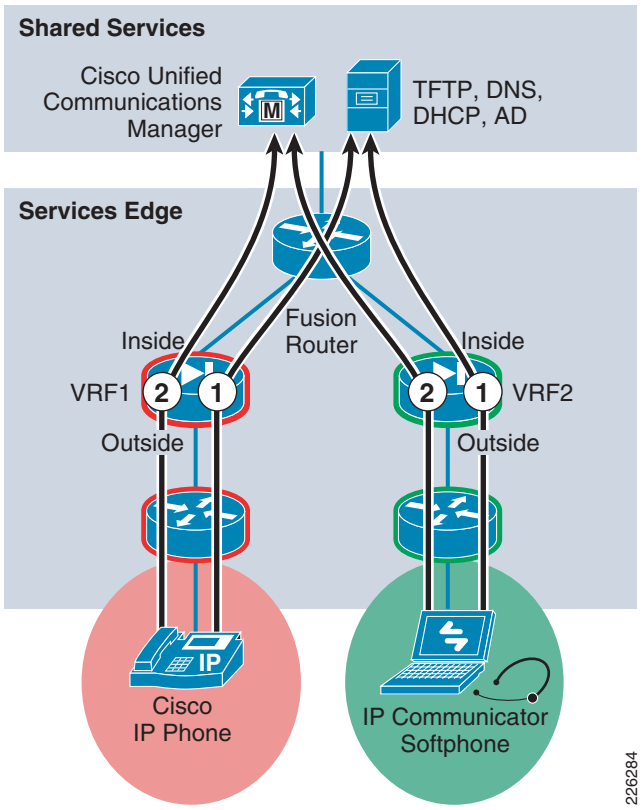
We look at the data flows in three parts:

- The signaling necessary for the IP communicator or 7900 phone to register with the CUCM.
- Data flows between two IP communicators or 7900 phones that are in the same VPN.
- Data flows between two IP communicators or 7900 phones that are in the different VPNs.

Endpoint Initialization and CUCM Registration

Figure 46 shows the signaling necessary for a Cisco IP Communicator or a Cisco Unified IP Phones 7900 series phone to boot up and register with CUCM.

Figure 46 Voice Endpoints Boot-up Process



In Figure 46, we see a 7960 desk phone in the Red VRF and an IP communicator in the Green VRF. Figure 46 highlights the protocol flows that occur when the phones initially boot up. The boot up process is identical for both phones:

1. The phones usually receive a valid IP address via DHCP. Part of the information returned by the DHCP server is option 150, which is the IP address of their TFTP server. This information allows the phones to download their configuration from the specified TFTP server.

Note If the CUCM name in the configuration is a name rather than an IP address, the phones resolve the CUCM name into an IP address using DNS. This would require an additional filter permitting DNS traffic to UDP port 53.

2. The phones register with CUCM. In this example, it uses the skinny (SCCP) protocol. If SIP was used instead, the access filter below would need to permit TCP and UDP port 5060 instead of TCP port 2000.

   The filters required in the firewall for these operations are shown in the configuration sample below. Note that these filters are applied on the interface marked outside in Figure 46.

```
!Define names used in IP access lists
name 10.13.100.70 CUPS description CUPS Server
name 10.13.100.5 DNS_DHCP_AD_Main
name 10.13.100.20 CUCM_pub description CUCM publisher
name 10.14.100.20 CUCM_sub description CUCM subscriber
!
! Permit DHCP and DNS
access-list outside-vrf4-ACL extended permit udp any host DNS_DHCP_AD_Main eq bootps
```

```
access-list outside-vrf4-ACL extended permit udp any host DNS_DHCP_AD_Main eq domain
!
! Permit normal phone bootup - TFTP (UDP/69) and skinny signaling (TCP/2000)
access-list outside-vrf1-ACL extended permit udp any host CUCM_pub eq tftp
access-list outside-vrf1-ACL extended permit tcp any host CUCM_pub eq 2000
```

Call Flows between IP Telephony Endpoints in the Same VPN

In this section we examine the steps necessary to establish a phone call between IP telephony endpoints deployed inside the same VPN.

Figure 47 Intra-VPN Voice Stream

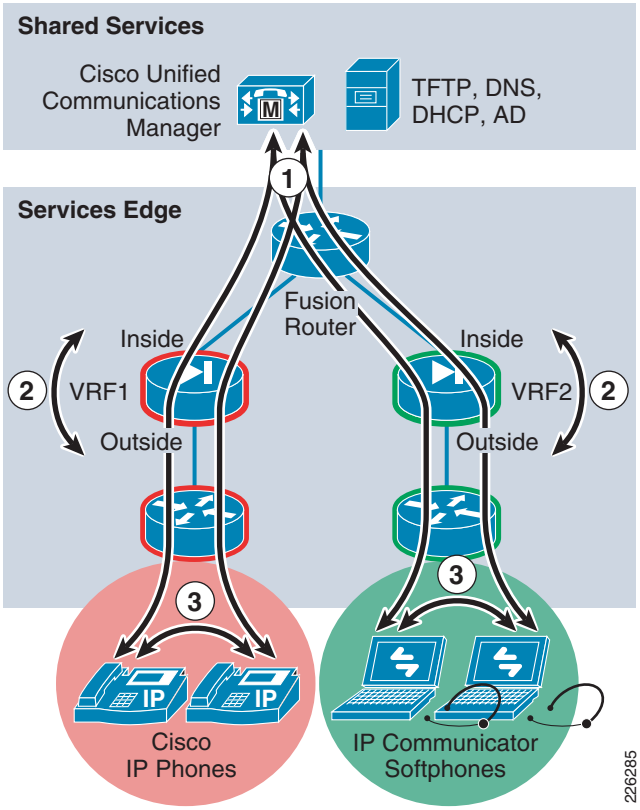


Figure 47 shows the call flows involved in a call between two IP telephony endpoints in the same VPN. The following endpoints are shown in this example:

- Two Cisco Unified IP Phones 7900 series phones in the same VPN (Red)
- Two Cisco IP Communicators in the same VPN (Green)

The step-by-step procedure is:

1. When one IP telephony endpoint calls another, call signaling information is exchanged between the calling phone and the CUCM. In the case of Cisco SCCP signaling, the following messages are transmitted:
   - The calling phone sends the dialed number to CUCM.
   - CUCM signals both phones to ring and to display the calling and called party information.

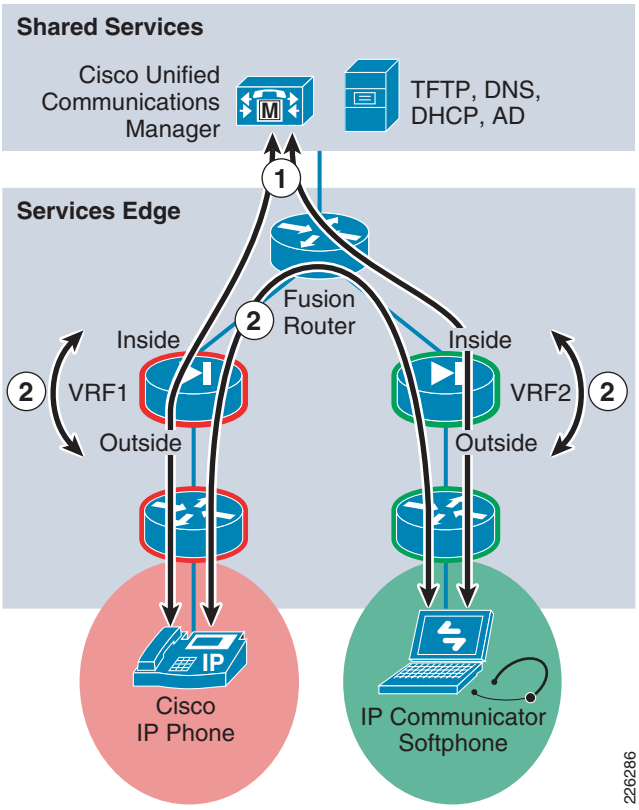
- When the called phone answers, the CUCM sends both phones a StationOpenReceiveChannel, which provides information about the RTP media stream and asks the phone to open a UDP port to receive the RTP media stream form the other phone.
- Each phone responds with a StationOpenReceiveChannelAck, which provides the IP address and port number that the IP phone is listening on for RTP packets.
- When CUCM receives the StationOpenReceiveChannelAck from each phone, it sends a StationStartMediaTransmission to the other phone telling it to start streaming the voice RTP media stream to the IP address and Port that was received in the StationOpenReceiveChannelAck.

2. The firewall filter has permitted the call signaling protocol and the firewall inspects it to perform protocol anomaly detection and application and protocol state tracking. Because the calls being established are between endpoints belonging to the same VPN, the media traffic does not have to traverse the firewall and no pinholes are opened between the inside and outside firewall interfaces.
3. The two IP telephony endpoints use the routing within their own VPN to send the peer-to-peer voice RTP media to each other.

Call Flows between IP Telephony Endpoints in Different VPNs

Figure 48 shows the common scenario where the IP Communicator is in the Green data VPN and the desk IP phone is in the red voice VPN.

Figure 48 Inter-VPN Voice Stream



In order for the voice RTP data to flow between two phones in different VPNs, it must traverse the firewall and must be routed by the fusion router. This is the required sequence of events:

1. When one IP telephony endpoint calls another, call signaling information is exchanged between the calling phone and the CUCM. The call signaling is the same as was described in the example where the two phones calling were in the same VPN:
   - The calling phone sends the dialed number to CUCM.
   - CUCM signals both phones to ring and to display the calling and called party information.
   - When the called phone answers, the CUCM sends both phones a StationOpenReceiveChannel, which provides information about the RTP media stream, and asks the phone to open a UDP port to receive the RTP media stream form the other phone.
   - Each phone responds with a StationOpenReceiveChannelAck, which provides the IP address and port number that the IP phone is listening on for RTP packets.
   - When CUCM receives the StationOpenReceiveChannelAck from each phone, it sends a StationStartMediaTransmission to the other phone telling it to start streaming the voice RTP media stream to the IP address and Port that was received in the StationOpenReceiveChannelAck.
2. The firewall filter has permitted the call signaling protocol and the firewall inspects it to perform protocol anomaly detection and application and protocol state tracking.
   - Each firewall sees that the call media flows through the firewall, to-and-from the phones VPN and the fusion router in the shared services VPN.
   - The firewall opens a pinhole that permits traffic to-and-from the IP addresses and UDP port it observed in the StationOpenReceiveChannelAck sent by the phone to CUCM and the StationStartMediaTransmission sent from CUCM to the phone.
3. The two IP telephony endpoints use the default routes advertised by the fusion router and injected into each VPN to send the traffic through the firewall pinhole to the fusion router. The fusion router forwards the traffic through the other firewall to the IP telephony endpoint in the other VPN.

The key point is that in this scenario firewall holes do not need to be configured to allow the RTP media data traffic to flow between outside and inside interfaces of each firewall context. This is because the SIP or SCCP signaling that sets up the call is inspected by the firewall, allowing it to then dynamically open conduits for the RTP media data traffic. The firewall pinhole used for that traffic is dynamically closed when the call ends.

SIP and Skinny inspection is enabled by default on the Cisco Firewall Service Module and ASA. The configuration sample below shows the statements responsible for SIP and SCCP inspection:

```
policy-map global_policy
class inspection_default
inspect dns preset_dns_map
inspect ftp
inspect h323 h225
inspect h323 ras
inspect netbios
inspect rsh
inspect rtsp
inspect skinny
inspect esmtp
inspect sqlnet
inspect sunrpc
inspect tftp
inspect sip
inspect xdmcp
inspect icmp
!
```

```
service-policy global_policy global
```

Summary—Cisco Unified IP Phones 7900 Series Phone and the Cisco IP Communicator

For calls between Cisco Unified IP Phones 7900 series phone and Cisco IP Communicators, the firewall needs to be configured with filters to allow the call signaling protocol through and permit the voice endpoints to successfully register with the CUCM. At this point two scenarios are possible:

- If the voice call is between two IP telephony endpoints in the same VPN, call routing stays within the VPN and the media streams do not need to flow through the firewall.
- If the call is between two IP telephony endpoints in different VPNs, the default behavior of the Cisco firewall is to inspect the call signaling and dynamically open a pinhole to allow the traffic to traverse each firewall as it is routed by the fusion router between VPNs.

Cisco Unified Video Advantage

Cisco VT Advantage brings video telephony functionality to the Cisco IP Phone 7940, 7960, and 7970 (and later) models. Cisco VT Advantage software coupled with the Cisco VT Camera (a USB camera) allows a PC connected to a Cisco IP phone to add video to telephone calls without requiring any extra button-pushing or mouse-clicking. When registered to Cisco Call Manager, the Cisco VT Advantage-enabled Cisco IP phone has the features and functionality of a full-featured IP video phone. Supplementary services, such as call forward, transfer, hold, and mute, are also available for video calls and are all initiated through the Cisco IP phone.

Cisco VT Advantage is intended for desktop-to-desktop IP video telephony environments, not as a general-purpose video conferencing solution for use in conference rooms. Users can choose to place Cisco Unified Video Advantage video calls with either a Cisco Unified IP phone or Cisco IP Communicator.

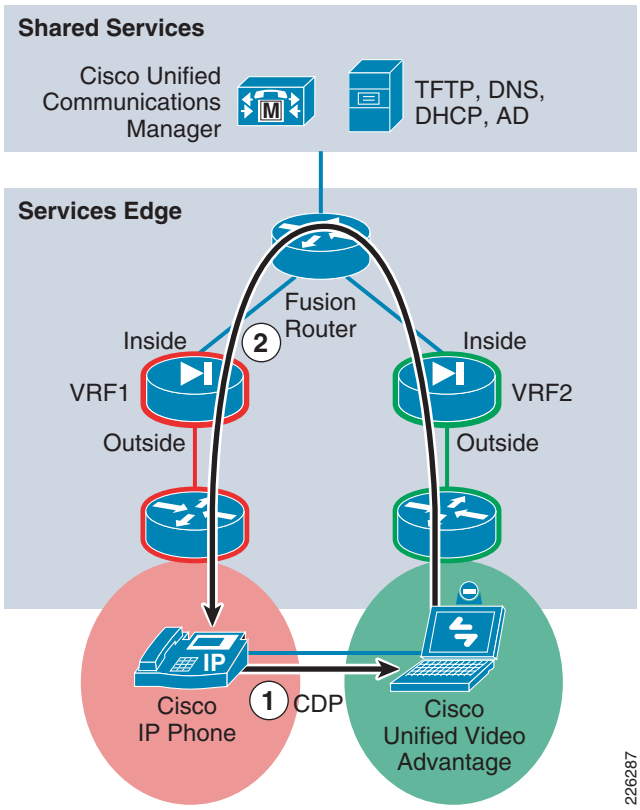
In normal use, Video Advantage runs minimized in the PC system tray. When a phone call is placed with another video Advantage user, video is automatically displayed when the voice call is connected. The sequence of events allowing the establishment of a CUVA call is listed below:

- CUVA initialization
- CUVA call signaling
- CUVA call media flows—intra-VPN
- CUVA call media flows—inter-VPN

CUVA Initialization

Figure 49 shows the protocol flows required for Video Advantage to initialize and become operational. For this we assume that the Cisco IP phone has already initialized and registered with the CUCM. The CUCM has been configured with a check-box denoting that this phone is video-enabled.

Figure 49 Initialization of the CUVA Client



Note Cisco Unified Video Advantage may be used with a Cisco Unified IP phone or with Cisco IP Communicator.

CUVA Desk Phone Usage

In Figure 49, Cisco Unified Video Advantage is shown directly connected to a 7960. The video Advantage is in the Green data VRF and the desk phone is in the Red voice VRF.

- Video Advantage does not require configuration on the PC, so it needs to learn the IP address of the phone used for the voice call. It learns this by listening to the link-layer CDP advertisements sent by the directly connected desk phone. Being a Layer 2 protocol, CDP can be sent directly between the endpoints even thought they are in different virtualized VPNs; virtualization provides Layer 3 segmentation, not Layer 2 segmentation.
- Once Video Advantage has learned the IP address of its directly attached IP phone, it establishes a TCP session with the phone using the Cisco Audio Session Tunnel (CAST) protocol. Because video Advantage and the desk phone are in different VPNs, the CAST session must traverse the firewalls and be routed by the fusion router.

Note The PC running CUVA needs to be directly connected to the IP phone used in desk phone mode to be able to receive the link-layer CDP frames it needs to register.

- A filter is required on the outside interface of the firewall context front-ending the Green data VPN in order to allow port 4224 for the CAST protocol to pass between the two VPNs. The CAST communication is initiated by the CUVA client, so the filter only needs to be applied on outside interface of the firewall on the CUVA VPN. Once the CAST protocol is routed to the IP phone VPN, it passes from the inside, where there is usually a permit-all filter. This establishes a connection dynamically permitting the return traffic. The filter used to permit CAST on the outside interface of the CUVA VPN is shown in the configuration sample below:

```
access-list outside-vrf2-ACL remark Permit CUVA (Video Advantage) CAST protocol (TCP
Port 4224)
access-list outside-vrf2-ACL extended permit tcp any any eq 4224
```

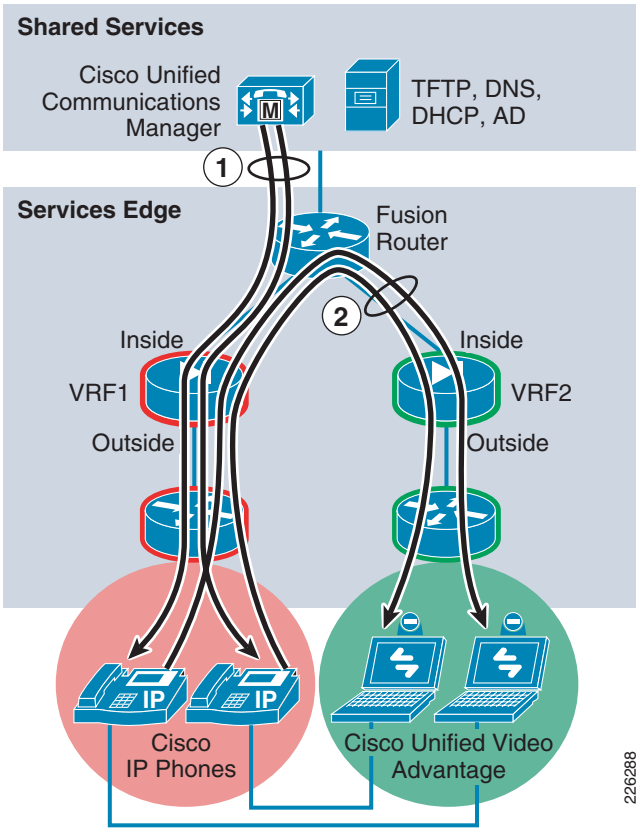
CUVA Softphone Usage

CUVA can be used with the Cisco IP communicator instead of a physical Cisco Unified IP Phone 7900 series phone. In this case, CUVA and IP Communicator are both running as software applications on the same PC. Because of this, the communication between CUVA and IP communicator is internal and the traffic flows shown in Figure 49 do not occur on the network.

CUVA Call Signaling

Figure 50 shows the video advantage signaling used to set up a call.

Figure 50 CUVA Signaling



In Figure 50, there are two Cisco IP phones in the Red voice VRF with two directly-connected PCs in the Green data VRF. Figure 50 shows what happens when one of the Cisco IP phones calls the other.

For clarity sake, Figure 50 does not show the call signaling that established the audio call between the two phones. The call signaling for the audio portion of the call is the same as was covered in Cisco Unified IP Phones 7900 Series Phone and the Cisco IP Communicator.

The Cisco Unified IP phone acts as a signaling proxy between CUCM and CUVA. Video signaling is sent via SIP or SCCP to the phone, which relays the signaling to the CUVA client via the CAST protocol. Note that both IP phones need to be configured in CUCM with video capabilities set to enabled. Video signaling between the phone and the CUCM is sent via SIP or SCCP. Acting as a video signaling proxy, each phone then relays the video signaling between itself and the CUVA PC using the CAST protocol

Because the phone and PC are in different VPNs, the CAST protocol messages have to traverse the firewalls to be routed between the VPNs by the fusion router.

The video signaling is similar to voice signaling and call establishment culminates in the following;

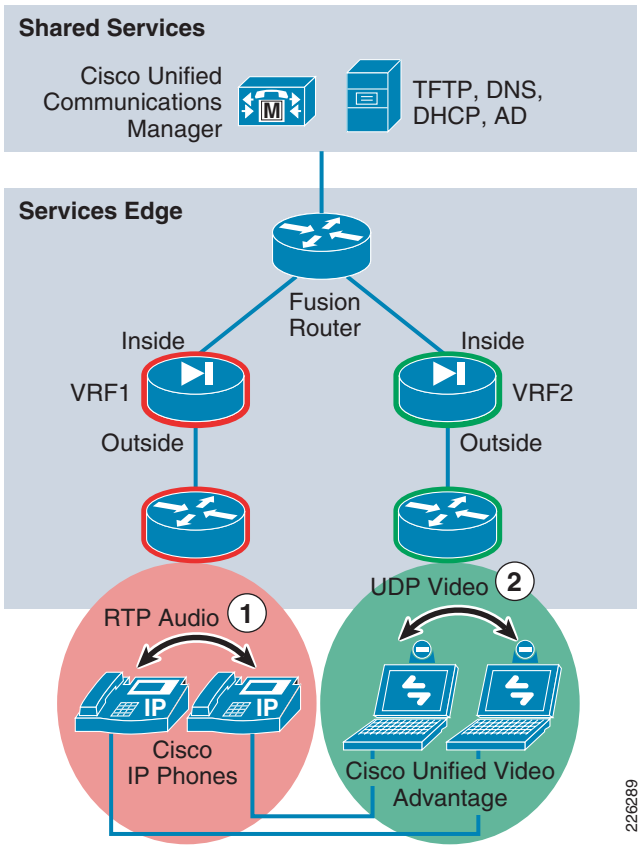
- OpenMultiMediaReceiveChannelMessage being sent using SIP or SCCP from the CUCM to the phone. The phone proxies this message to the CUVA endpoint using the CAST protocol.
- The CUVA endpoint sends an OpenMultiMediaReceiveChannelAckMessage to the phone using CAST protocol. The phone proxies this to the CUCM using the SIP or SCCP protocol. The OpenMultiMediaReceiveChannelAckMessage contains the IP address and UDP port the CUVA endpoint has opened to receive video media from the peer CUVA endpoint.
- When the CUCM receives the OpenMultiMediaReceiveChannelAckMessage from one endpoint, it sends a StartMultiMediaTransmissionMessage to the other endpoint telling it to begin sending the RTP video media stream to the specified IP address and UDP port (port 5445 is always used for CUVA video media). This message is sent from the CUCM to the phone via SIP or Skinny and relayed to the CUVA endpoint using the CAST protocol.

The same firewall access-list on the outside interface of the CUVA VPN used to permit CAST for CUVA initialization is used to permit CAST for call signaling.

CUVA Call Media Flows—Intra VPN

Figure 51 shows how the RTP voice and video media flow between the IP phones and the CUVA endpoints involved in a CUVA call, in the specific case where both entities are deployed inside the same VPN.

Figure 51 Intra-VPN CUVA Call

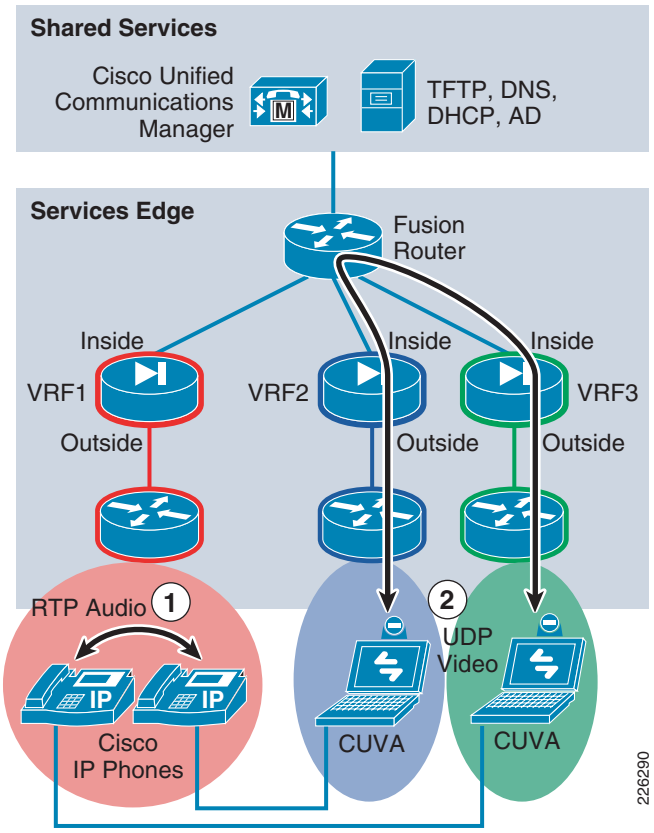


- The two phones are in the same VPN, so the RTP audio packets between them are routed within the Red VPN (and do not need to be inspected by the firewall or routed by the fusion router).
- Similarly, in this example, the two PCs are in the same data VPN and the RTP video packets between them are routed within the Green VPN.

CUVA Call Media Flows—Inter VPN

Figure 52 shows the call media flows for a slightly more complex CUVA deployment. In this example, there is one voice VPN corporate-wide, but there are multiple data VPNs.

Figure 52 Inter VPN CUVA Call



- Within the Red voice VPN, voice RTP packets flow directly just as in the previous example.
- The video endpoints associated with the Cisco IP phones that established the call are in different data VPNs. The RTP video media in this case needs to flow through the firewall and be routed by the fusion router. For this to happen another filter needs to be added to the outside interface of the firewall contexts protecting both the data VPNs to permit the UDP port used by the video Advantage RTP data. Fortunately Video Advantage only uses a single port as the source and destination of its media traffic, so only UDP port 5445 needs to be opened on the firewalls.

Cisco firewalls do not currently inspect the CAST protocol, so the filter required for CUVA initiated Video media needs to be manually configured. Each CUVA endpoint establishes a unidirectional RTP video media stream to the other endpoint, so the filter needs to be applied to the outside interface of both CUVA client VPN firewalls. The firewall ACL required to allow CUVA video RTP data between VPNs is:

```
! Permit CUVA video data between VPNs
access-list outside-vrf3-ACL extended permit udp any any eq 5445
```

CUVA Summary

When a call is placed between two phones that are configured as video enabled and that have a workstation with CUVA attached to them, the phone plays the role of the voice endpoint, whereas the data device represents the video endpoint.

All CUCM signaling is sent to and from the phone. The phone proxies the video signaling to and from the workstation via the CAST protocol. Cisco firewalls do not currently inspect the CAST protocol, so the filter required for CUVA-initiated video media needs to be manually configured.

Voice media flows directly between the voice endpoints and video media flows directly between the video endpoints.

Cisco Unified Personal Communicator

Cisco Unified Personal Communicator transparently integrates frequently used communications applications and services into a single unified client. From an easy-to-use interface on a PC or Mac, it provides quick and easy access to powerful communications tools—softphone, presence, instant messaging, visual voice mail, click to call, employee directory, communication history, video, and Web conferencing.

When being used for IP Telephony, CUPC can operate in either of two modes:

- Softphone mode

  In softphone mode, the CUPC is an IP telephony endpoint, using SIP signaling to communicate with CUCP and establishing RTP voice media streams with other voice endpoints.

- Deskphone mode

  In deskphone mode, the CUPC is used to control a desktop Cisco Unified IP phone to make, receive, or merge calls.

Next we look at the following stages of CUVA call establishment;

- CUPC initialization
- CUPC signaling and call flows—deskphone mode
- CUPC signaling and call flows—softphone mode
- CUPC signaling and call flows—softphone video call
- CUPC signaling and call flows—softphone video call, inter-VPN
- CUPC signaling and call flows—instant messaging

CUPC Initialization

In this section we look at the data flows that occur when CUPC is started on a workstation.

Figure 53 CUPC Initialization Process

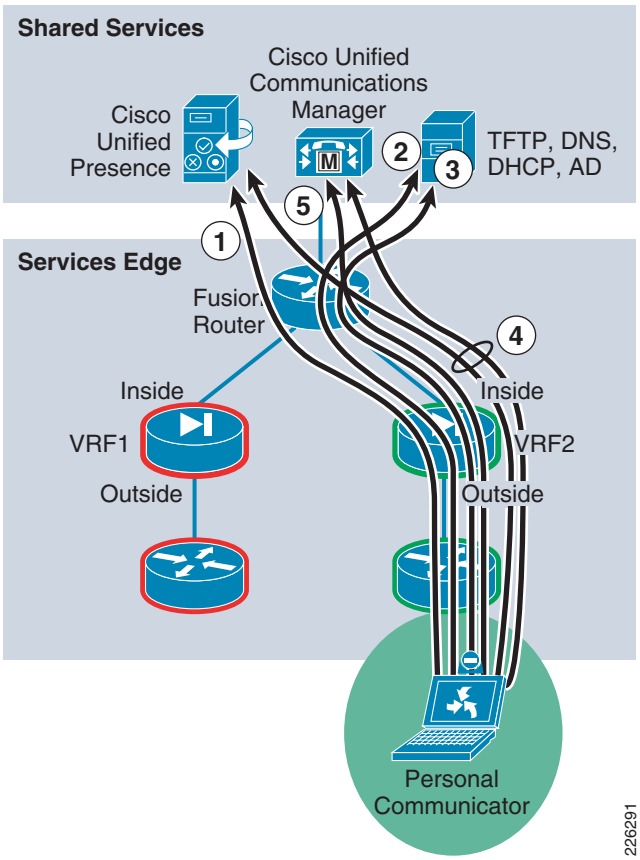


Figure 53 highlights the steps required for CUPC to boot up completely:

1. CUPC uses HTTPS to securely log into the CUPS server. CUPC learns the IP address of the TFTP server from the presence server.
2. CUPC then TFTP downloads its configuration file just like any other phone.
3. Cisco Unified Personal Communicator access corporate Lightweight Directory Access Protocol (LDAP) version 3 directory to provide additional contact information (first name, last name, phone numbers, and so forth) through directory searches for each contact in the contact list.
4. CUPC uses the SIP protocol to register with the CUCM and with the CUPS server for presence information. CUPC uses SIP signaling protocol; it can not use Skinny (SCCP).
5. If the CUPC client is being used in deskphone mode, a CTI connection is established with CUCM. The CTI protocol allows the CUPC to control a desktop Cisco Unified IP Phone to make, receive, or merge calls. The controlled phone does not have to be directly attached, but it does need to be associated with the CUPC user in the CUCM configuration.

Note Cisco Unified Presence delivers a Session Initiation Protocol (SIP) presence engine and SIP proxy server functionality to Cisco Unified Personal Communicator. The presence engine provides Cisco Unified Personal Communicator with the infrastructure for user and device status information (for example, available, away, do not disturb) by using SIP Instant Messaging and Presence Leveraging Extensions (SIMPLE).

The presence engine stores information about the preferred communication method (instant message, E-mail, voice, video) and the contact list for each user. Cisco Unified Presence also maintains login

authentication for, and provides configuration information to, Cisco Unified Personal Communicator by using Simple Object Access Protocol (SOAP) over HTTP and HTTPS. The proxy server provides both registration and routing support for its clients, all of which are SIP-based. The Cisco Unified Personal Communicator sends SIP messages to and receives SIP messages from this proxy server. These SIP messages are for presence information and database change notifications. Cisco Unified Personal Communicator also sends SIP messages (via the proxy) to other Cisco Unified Personal Communicator clients for instant messaging.

The configuration sample below shows all the filters that need to be added to the outside interface of the firewall context front-ending the CUPC VPN in order to allow CUPC to initialize:

```
name 10.13.100.70 CUPS description CUPS Server
name 10.13.100.5 DNS_DHCP_AD_Main
name 10.13.100.20 CUCM_pub description CUCM publisher
name 10.14.100.20 CUCM_sub description CUCM subscriber
!
object-group protocol TCPUDP
protocol-object udp
protocol-object tcp
!
! 1. Permit HTTPS for CUPC login to CUPS (TCP port 443)
access-list outside-vrf1-ACL extended permit tcp any host CUPS eq https
! 2. Permit IP phone TFTP configuration download (UDP port 69)
access-list outside-vrf1-ACL extended permit udp any host CUCM_pub eq tftp
! 3. Permit LDAP communication between CUPC and LDAP DB (TCP port 389)
access-list outside-vrf1-ACL extended permit tcp any host DNS_DHCP_AD eq ldap
! 4. Permit SIP signaling between CUPC and CUCM & CUPS (TCP/UDP port 5060)
access-list outside-vrf1-ACL extended permit object-group TCPUDP any host CUCM_pub eq sip
access-list outside-vrf1-ACL extended permit object-group TCPUDP any host CUPS eq sip
! 5. Permit CTI communication between CUPC in deskphone mode and CCM (TCP port 2748)
access-list outside-vrf1-ACL extended permit tcp any host CUCM_pub eq ctiqbe
```

CUPC Signaling and Call Flows—Deskphone Mode

This section presents the flows that occur when CUPC is used to control a desktop Cisco Unified IP phone to make a call.

Figure 54 CUPC in Deskphone Mode

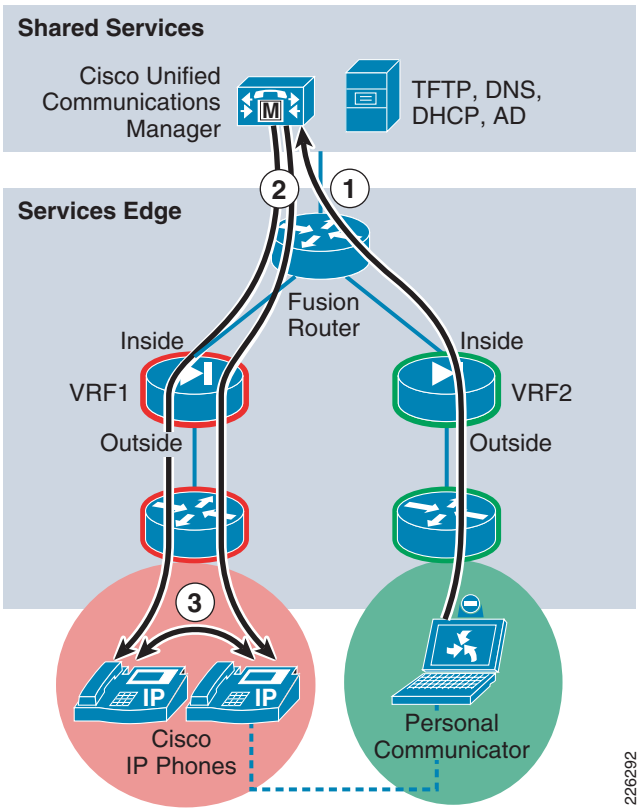


Figure 54 shows the signaling and call flows when a CUPC is used to control a deskphone. An example of this use case is using the click-to-dial functionality provided by the CUPC Outlook toolbar. When outlook click-to-dial is used on a CUPC workstation, CTI signaling is sent to CUCM to cause the local deskphone and the remote phone to ring simultaneously. When the remote phone is answered, the deskphone is used for the call and CUPC is no longer involved.

1. To achieve this the CUPC signals via Computer Telephony Integration (CTI) to CUCM.
2. CUCM then signals the Cisco IP phone being called via Skinny (or SIP).
3. Once connected, the call media flows directly between the two IP phones if deployed as part of the same voice VRF as shown in this example.

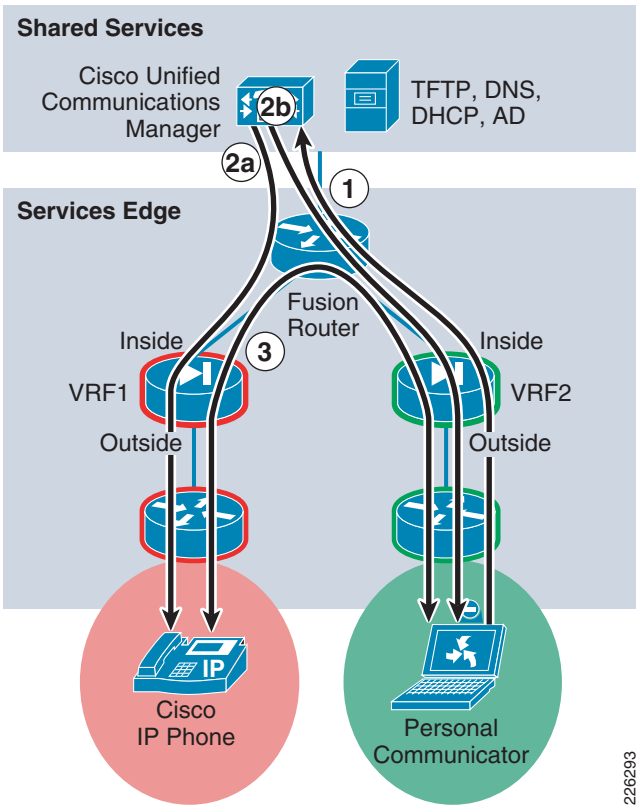
One additional thing to note here is that with CUPC the relationship between CUPC and the deskphone it controls is configured in CUCM, there is no need for a physical Ethernet connection between the two. This logical relationship is shown as a dotted line in Figure 54.

The CTI flow signaling is let through the firewalls by the filter numbered five in the initialization section. Additionally, the SCCP or SIP signaling used to control the deskphones is inspected by the firewall, but in this example the two deskphones are in the same VPN and their RTP voice media traffic does not need to traverse a firewall.

CUPC Signaling and Call Flows—Softphone Mode

In softphone mode the CUPC is an IP telephony endpoint, using SIP signaling to communicate with CUCP, and establishing RTP voice media streams with other voice endpoints.

Figure 55 CUPC in Softphone Mode



In Figure 55, CUPC is being used in stand-alone softphone mode.

1. The CUPC signals to CUCM using SIP. CUPC only uses SIP for signaling; Skinny is not currently supported.
2. CUCM signals CUPC and the deskphone it is calling. CUCM uses SIP to the CUPC and can use either SIP or Skinny to the deskphone.
3. Finally the RTP media stream flows between the two voice entities.

Because the call is between a deskphone in the Red voice VPN and a CUPC in the Green data VPN, the media flows have to go via the firewalls and the fusion router.

In this case the SIP and Skinny flows were let through the firewalls by the filters defined in the initialization section. Additionally, the SIP and SCCP signaling is inspected and the RTP data flows are dynamically permitted for the life of the call.

CUPC Signaling and Call Flows—Intra VPN Softphone Video Call

CUPC can be used as a video phone when used on workstations equipped with a Webcam. This section examines the flows that occur when a video call is placed between to CUPC workstations in the same VPN.

Figure 56 Intra VPN Video Call with CUPC in Softphone Mode

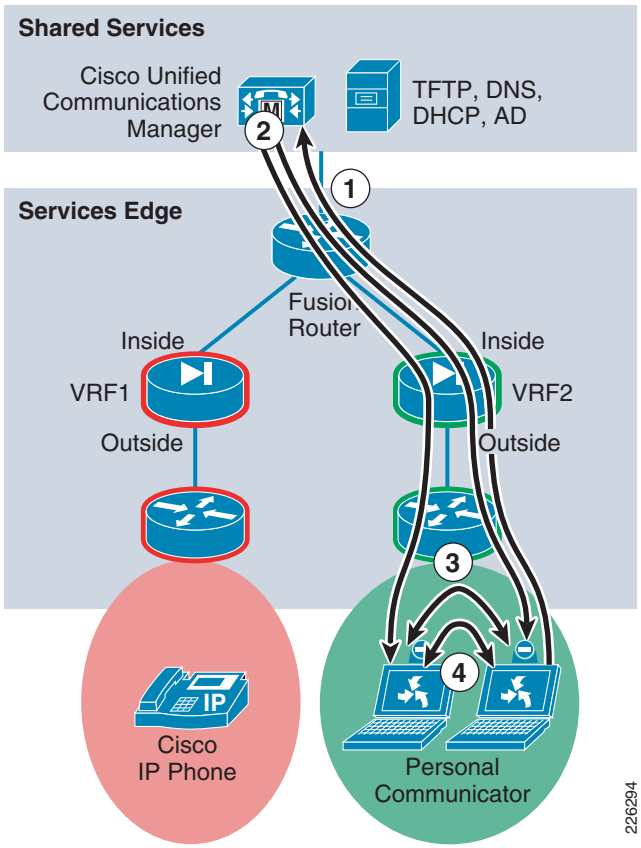


Figure 56 shows a softphone call between two CUPC clients within the same Green VRF.

1. One CUPC client calls the other. SIP signaling is sent to CUCM requesting call up.
2. SIP signaling flows between the CUCM and the two CUPC endpoints setting up the voice call.
3. Because both CUPCs are within the same Green data VPN, the RTP audio between the two CUPCs is routed within the VPN.
4. CUCM recognizes that the call is between two video-enabled clients and instructs the CUPC clients to begin sending video to each other. The video is launched and is also routed within the VPN.

Because voice and video RTP streams are routed within the same VPN, the data streams do not traverse the services edge and no traffic inspection or additional filters are required.

CUPC Signaling and Call Flows—Inter VPN Softphone Video Call

This section is similar to the previous except the CUPC endpoints are in two different VPNs.

Figure 57 Inter VPN Video Call with CUPC in Softphone Mode

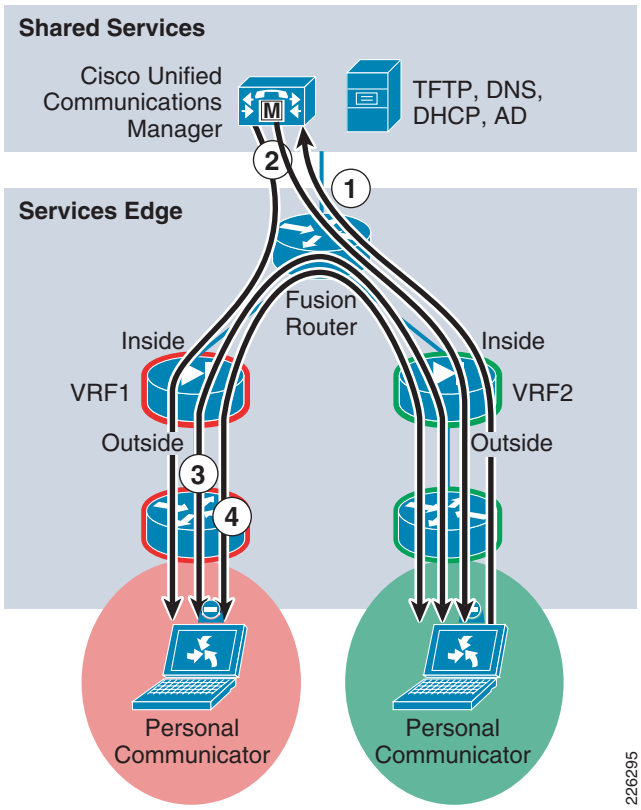


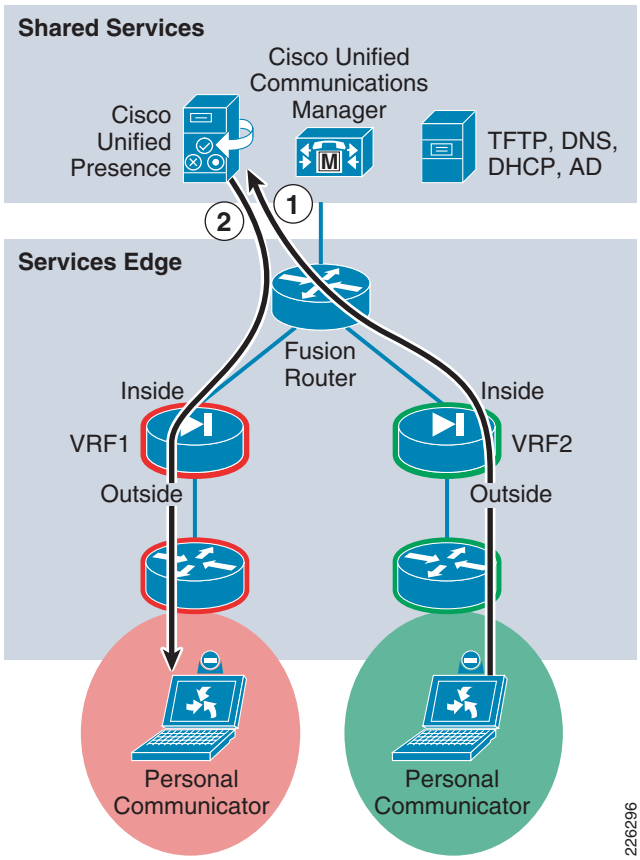
Figure 57 shows the two CUPC clients deployed in different VPNs.

1. One CUPC client calls the other. SIP signaling is sent to CUCM requesting call up.
2. SIP signaling flows between the CUCM and the two CUPC endpoints setting up the voice call.
3. The firewall in each VPN is inspecting the SIP signaling and dynamically opens a pinhole to allow the UDP ports used for the voice RTP media stream to pass through.
4. Both endpoints are configured in CUCM as video-enabled. CUCM sends SIP signaling to the CUPC clients instructing them to initiate video between each other. The firewall in each VPN is inspecting the SIP signaling and dynamically opens a pinhole to allow the UDP ports used for the video RTP media stream to pass through.

CUPC Signaling and Call Flows—Instant Messaging

Instant messaging support enables CUPC users to chat in real time with other Cisco Unified Personal Communicator users to save time and reduce phone tag.

Figure 58 Instant Messaging with CUPC



Cisco Unified Personal Communicator sends instant messages to other CUPC users by using the SIMPLE protocol support on the Cisco Unified Presence server to relay the messages between CUPC clients.

1. CUPC one in the Green VRF sends SIP/SIMPLE instant messages to the SIP proxy function in the Cisco Unified Presence server.
2. The CUPS server then relays the instant message down to CUPC two in the Red VRF.

The SIP protocol is used for instant messaging between CUPC clients. SIP instant messaging is allowed and inspected as a result of the SIP filters defined for CUPC initialization.

CUPC Summary

Cisco Unified Personal Communicator transparently integrates frequently-used communications applications and services into a single, unified client. It provides quick and easy access to powerful communications tools—softphone, presence, instant messaging, visual voice mail, click to call, employee directory, communication history, video, and Web conferencing.

This section looked at CUPC message flows and noted that the following statically-defined filters permit CUPC clients to operate successfully:

- Permit HTTPS, TFTP, SIP, LDAP, and CTI traffic from the CUPC VPN to specific server destinations in the shared services VPN.
- Firewall inspection of the SIP signaling is sufficient to permit the applications tested to communicate within and between voice endpoints in different VPNs.

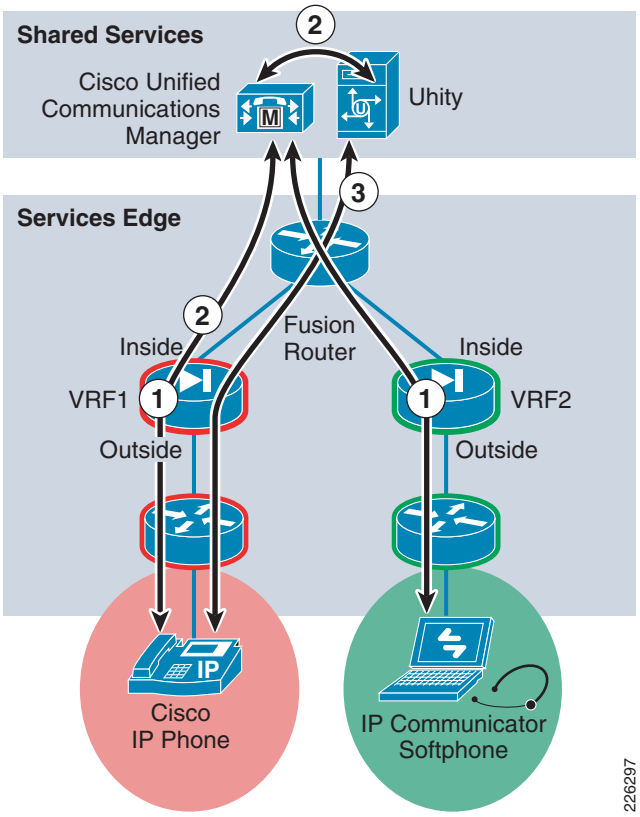
Note CUPC version 1.2 supports two features which were not tested for this document:
- Web conferencing—Launch a Web conferencing session at a moment's notice to share content, such as a presentation, with others
- Voice messages—Access Cisco Unity® or Cisco Unity Connection voice mail messages—view, play back, sort, and delete messages—from within the application. These features use the following TCP ports which would need additional filters if they were to be enabled: TCP 143 (Web conferencing) and TCP 80 (Voice messages).

Also, CUPC version 7.0 has been released since testing was completed. CUPC 7.0 permits secure voice mail messages to be decrypted and played back by CUPC. This requires the CUPC to be able to communicate with the following ports, for which additional filters would be required:
- 7993—Special IMAP port (Cisco Unity Connection)
- 993—SSL (Exchange-IMAP)
- 443—HTTPS (Cisco Unity)

Unity Voicemail

Accessing Unity voicemail is primarily a matter of placing a voice call to the Unity server. The Unity server is located in the shared services VPN with the CUCM. Co-locating Unity with CUCM provides consistent call control and voice messaging access to all voice clients.

Figure 59 Integrating Unity Voicemail



1. The IP phone places a call to the softphone. SIP or Skinny call signaling between the endpoints and CUCM causes the phones to ring.

2. When the called phone does not answer, CUCM is configured to redirect the call to Unity Voicemail. CUCM uses SCCP signaling to Unity and CTI to signaling to Unity Express. Because CUCM and Unity are co-located in the same VPN, this signaling does not need to be configured or inspected in the services edge firewalls.
3. CUCM uses standard SCCP or SIP signaling between the calling endpoint and Unity to establish an RTP voice media call.
4. The call proceeds. Unity prompts and message playback are sent from the Unity server to the voice endpoint as RTP voice media packets. End-user key-pad responses to Unity voice prompts are sent to Unity as dual-tone multi-frequency (DTMF) touch-tones. DTMF touch-tones between the calling phone and unity are typically relayed out-of-band to CUCM via the SCCP or SIP call signaling protocol, but an option exists to carry them within the RTP voice data stream using a separate RTP payload type.

Firewall holes do not need to be configured to allow IP phones to access Unity Voicemail. This is because the SIP or Skinny signaling that set up the call is inspected by the firewall which then dynamically opens firewall for the specific IP addresses and UDP ports used in the call. The firewall pinhole used for that traffic is dynamically closed when the call ends.

PSTN Gateway Access

The PSTN gateway is used to place and receive calls to phones on, or attached to, the public telephone network. The PSTN gateway translates between the call signaling and transport protocols used by the IP network and those used on the PSTN. MGCP and H323 are the most common control protocols used for signaling between CUCM and the PSTN gateway. Both MGCP and H323 were tested for this document.

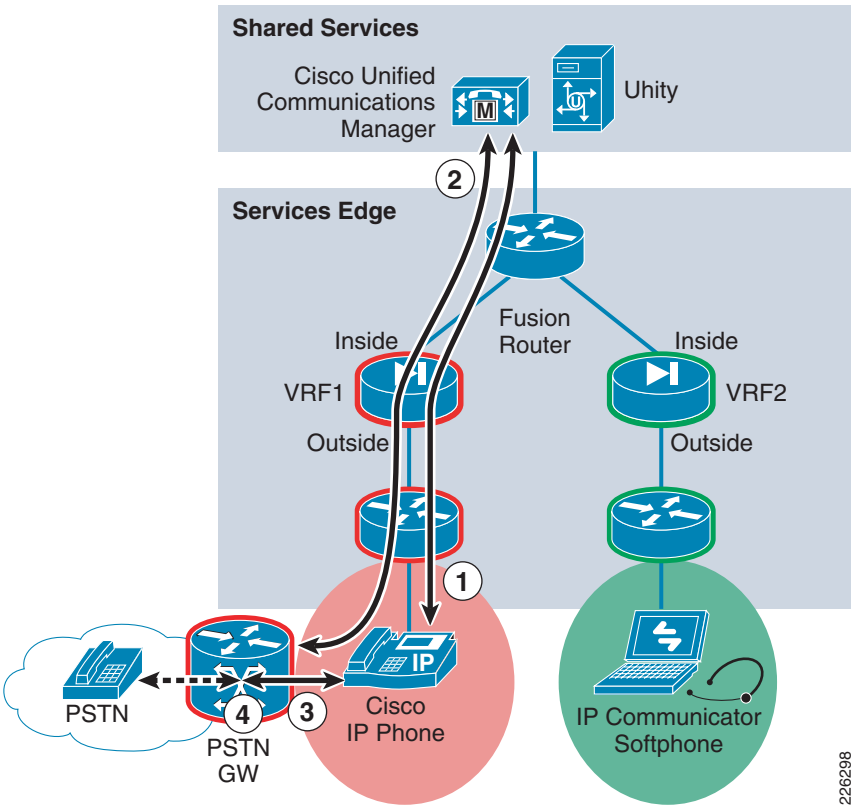
Note Our network design does not use a VRF-aware PSTN gateway. Instead we connect the PSTN gateway with an Ethernet interface that belongs to the same VLAN/VRF as the physical Cisco IP phones. Because physical IP phones are in the same VRF as the PSTN gateway, they can send RTP voice packets directly to and from the PSTN gateways. Voice endpoints in other VPNs need to traverse the services edge to communicate with the PSTN gateway.

This section looks at the following use cases:

- PSTN gateway access from a physical IP phone in the same VPN as the PSTN gateway
- PSTN gateway access from a softphone in a different VPN to the PSTN gateway

PSTN Gateway Access—Intra VPN

Figure 60 PSTN Gateway Access (Intra VPN)

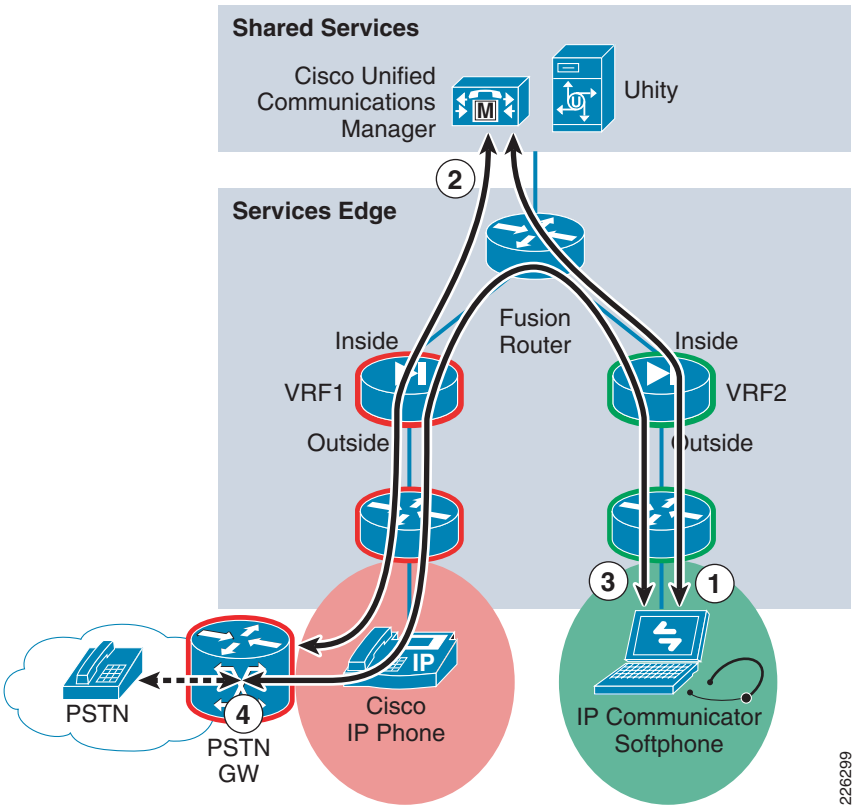


1. The Red phone dials the phone number of an external phone on the PSTN. It uses Skinny or SIP signaling to CUCM to set up the call.
2. Call manager dial plan determines that the number dialed is reachable through the PSTN gateway and uses either MGCP or H323 signaling to the PSTN gateway to set up a call between the Red phone and the external phone. The PSTN gateway acts as a signaling proxy when setting up the call to the external phone.
3. The Red phone and the PSTN gateway use the call setup information signaled by CUCM to establish the call between each other. Because both the IP phone and the PSTN gateway are in the same VPN, no firewall traversal is required.
4. The PSTN gateway acts as a proxy between relaying the call between the IP network and the PSTN network.

End-user key-pad responses to voice prompts are sent to Unity as DTMF touch-tones. By default, the gateway sends these tones within the voice RTP stream. When voice is sent uncompressed, these tones arrive in their original state. However, when voice is compressed, such as with the G.729 codec, the tones might be distorted or part of the signal lost. DTMF relay addresses this by separating these tones from the rest of the voice and sending them in a different way. DTMF tones can be carried in-band, within the RTP voice data stream using a separate RTP payload type, or they can be carried within the MGCP or H323 signaling. Refer to IOS gateway documentation for more information on DTMF relay.

PSTN Gateway Access—Inter VPN

Figure 61 PSTN Gateway Access (Inter VPN)



The inter-VPN case very similar to the intra-VPN one just discussed. The only difference is that the RTP voice media between the PSTN gateway and the IP phone now has to traverse the firewalls in the services edge. Inspection of the signaling between the CUCM and the gateway and IP phone permit the firewalls to open pinholes for the IP address and UDP ports used by the RTP media stream.

Filters Required for a PSTN Gateway

For calls from the IP network, these filters are not required. Because packets are not filtered on the inside interface of the firewall, the MGCP or H323 filtering passes through the firewall from the CUCM to the PSTN gateway. Return traffic for outgoing sessions is automatically permitted by the firewall.

For calls from the PSTN, filters for the MGCP or H323 signaling need to be explicitly added to the outside interface of the VPN to which the PSTN gateway belongs. The filters required are highlighted in the configuration sample below.

- MGCP gateway

```
! MGCP inspection is not enabled by default and needs to be turned on
policy-map global_policy
 class inspection_default
  inspect mgcp
!
name 10.13.100.20 CUCM_pub description CUCM publisher
!
access-list outside-vrf2-ACL remark MGCP access-list outside-vrf2-ACL extended permit
tcp any host CUCM_pub eq 2428
```

```
access-list outside-vrf2-ACL extended permit udp any host CUCM_pub eq 2427
```

- H323 gateway

```
name 10.13.100.20 CUCM_pub description CUCM publisher
!
access-list outside-vrf2-ACL remark H323
access-list outside-vrf2-ACL extended permit tcp any host CUCM_pub eq h323
```

PSTN Summary

In this section we looked at MGCP and H323 gateway usage in a virtualized network. Our design considered the gateway to be a specialized form of IP telephony endpoint and so placed it in the IP telephony VPN. This had the effect of not requiring PSTN traffic from physical IP phones to be routed via the services edge and the fusion router.

SIP PSTN gateways were not tested, but it is expected that they would work as long as SIP signaling is permitted through the outside firewall interface on the VPN connected to the SIP PSTN gateway.

The filters required for receiving calls from the PSTN gateway were provided.

Services Edge—Summary

The services edge portion of the overall network virtualization process is where a large part of policy enforcement and traffic manipulation is done. The purpose of this document is to provide design guidance around various methodologies to be deployed to achieve this purpose.

The initial sections of the document defined the concept of shared services, distinguishing between two scenarios:

- Unprotected services access achieved by performing route leaking between different VRF routing tables.
- Protected services access, representing the recommended approach and usually deployed by front-ending each defined virtual network with a security device (firewall or firewall context).

For the specific scenario of protected services access, two deployment models (single and dual tier) were introduced, discussing in depth the design principles and the configuration steps required for their implementation, together with a detailed convergence analysis under different failure scenarios.

Finally, as a specific application of the services edge deployment, we discussed a possible technical solution to the problem of integrating voice applications into a virtualized network environment. The scope of this solution is currently limited to campus deployments and leverages the concept of a shared services area in the network to deploy UC services (like Cisco Call Manager, TFTP servers, etc.) that need to be accessed by network entities (IP phones, PCs running softphone applications, etc.) deployed in the context of separate virtual networks.

Cisco Validated Design

The Cisco Validated Design Program consists of systems and solutions designed, tested, and documented to facilitate faster, more reliable, and more predictable customer deployments. For more information visit www.cisco.com/go/validateddesigns.

ALL DESIGNS, SPECIFICATIONS, STATEMENTS, INFORMATION, AND RECOMMENDATIONS (COLLECTIVELY, "DESIGNS") IN THIS MANUAL ARE PRESENTED "AS IS," WITH ALL FAULTS. CISCO AND ITS SUPPLIERS DISCLAIM ALL WARRANTIES, INCLUDING, WITHOUT