Introduction 2
OSPF ABR and Loop Prevention 2
Loop Prevention in MPLS VPNs 5
Virtual Links and Sham Links 8
Multi-Homed Multi-Area OSPF Site 17
Appendix: Initial Configuration for Multi-Area Multi-Homed OSPF Scenario 37

Introduction

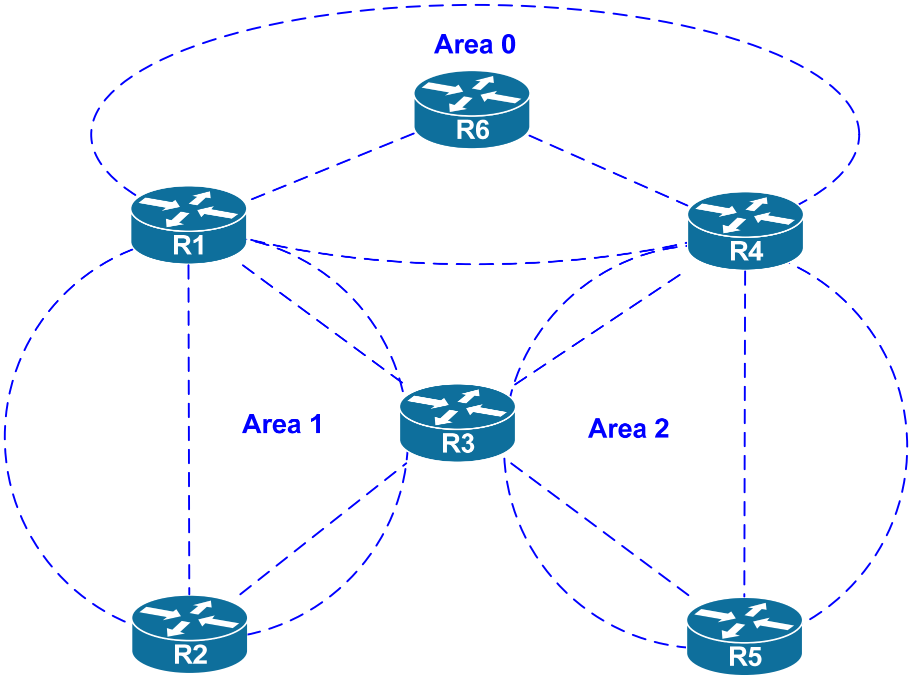
This document is presented as a series of Questions and Answers, discussing various aspects of OSPF protocol used to prevent inter-area routing loops. The discussion covers ABR functioning, Virtual-Links, OSPF Super-backbone, OSPF Sham-Links, BGP Cost Community. Reader is assumed to know these concepts already, as this publication focuses on complex interaction features arising in MPLS/BGP VPN scenarios. The discussion is culminated by analyzing a complex multi-area multi-homed OSPF site scenario in MPLS VPN environment.

OSPF ABR and Loop Prevention

Q: How does OSPF prevent routing loops when exchanging summary LSAs?
A: In OSPF, the backbone area is used for exchanging inter-area routes between all other areas. Since there is no common topology shared among different areas, loop prevention should be based on distance-vector principles. There are three main rules of generating and receiving inter-area routes (type-3 LSAs) in OSPF that prevent control-plane routing loops:

- Area Border Router (ABR) is a router that has at least one interface in Area 0 and this interface is NOT in DOWN state. ABR is distinguished by setting the B (border) bit in its router LSA to signal other routers in the same area of its ABR status. Only ABR is allowed to generate summary LSAs and inject them in the attached areas.
- ABR expects summary LSAs from Area 0 only. This means there should be at least one adjacency in FULL state built over Area 0 interface. In case if ABR *has such adjacency*, it will ignore summary-LSAs received over non-backbone areas. These LSAs will be installed in the database, but not used for SPF calculations.
- ABR will *accept and use* summary-LSAs learned over *non-backbone* area if it DOES NOT have a FULL adjacency built over an Area 0 interface. It is safe to do so, since the ABR will not be able to flood the summary back into Area 0 creating routing loops.

The use of above rules could be demonstrated using the topology below.



Q: Is R3 an ABR and will it generate summary LSAs for Areas 1 and 2?
A: Assuming that is has interfaces *only* in Area 1 and Area 2 but not Area 0, it will NOT summarize any routes from Area 1 into Area 2 and vice versa, because it does not consider itself an ABR. It will, however, learn the summary LSAs injected by both R1 and R4 and use them, as it is essentially a router internal to both Area 1 and Area 2.

Q: What if you want R2 and R5 to communicate via R3, as opposed to going via R1 and R4?
A: In this case, you need to make R3 and ABR. To accomplish this, create a new interface, e.g. a new Loopback, and advertise it into Area 0. There is no OSPF adjacency on this interface, so Area 0 will show up as “inactive” on R3. However, this is enough for R3 to start advertising itself as ABR and generating summary LSAs from Area 2 and from Area 1. After this configuration, R2 and R5 will both receive and process those summary LSAs and would be able to use R3 for transit.

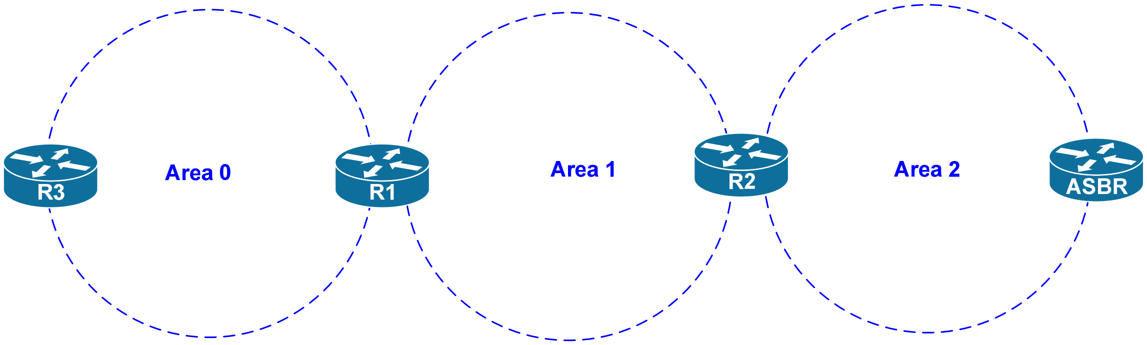
Q: So now that R3 is an ABR, will it reject the summary routes originated by R1 and R4?
A: No, like we mentioned previously R3 has to have a FULL adjacency in Area 0 to start ignoring the summary LSAs received over non-backbone areas.

Q: What about R1 and R4 – will they accept the summary LSAs generated by R3?
A: No, since R1 and R4 both have FULL adjacencies with another router in Area 0. They will receive the summary LSAs from R3 but not use them, preferring the same summaries learned via Area 0.

Q: What if I want R1 to reach R5 using the path via Area 1 and then Area 2, as opposed via Area 0 and Area 2?
A: You need a virtual link between R1 and R3. When virtual link is up, R1 and R3 will form Area 0 adjacency and R3 will advertise summary LSA for R5's networks into Area 0. Since this advertisement is now received over Area 0, R1 will accept it.

Q: But what would happen with R3 – now that it has an active adjacency in Area 0, it will start rejecting the summary LSAs R4 is sending to it?
A: Correct, it will prefer going over Area 1 to any destination in Area 0. If you want to resolve this issue, create another virtual-link between R3 and R4.

Q: Does the rule of blocking summary LSAs apply to Type-4 summaries as well?
A: Yes, the rule of ignoring summary-LSAs applies to both Type-3 (network summary) and Type-4 (router summary) LSA. Moreover, this will implicitly have effect on external routing information. Without having Type-4 LSA for an ASBR, routers will ignore all external routes injected by the corresponding ASBR, even though type-5 LSAs are flooded across the whole OSPF domain. Here is an example:



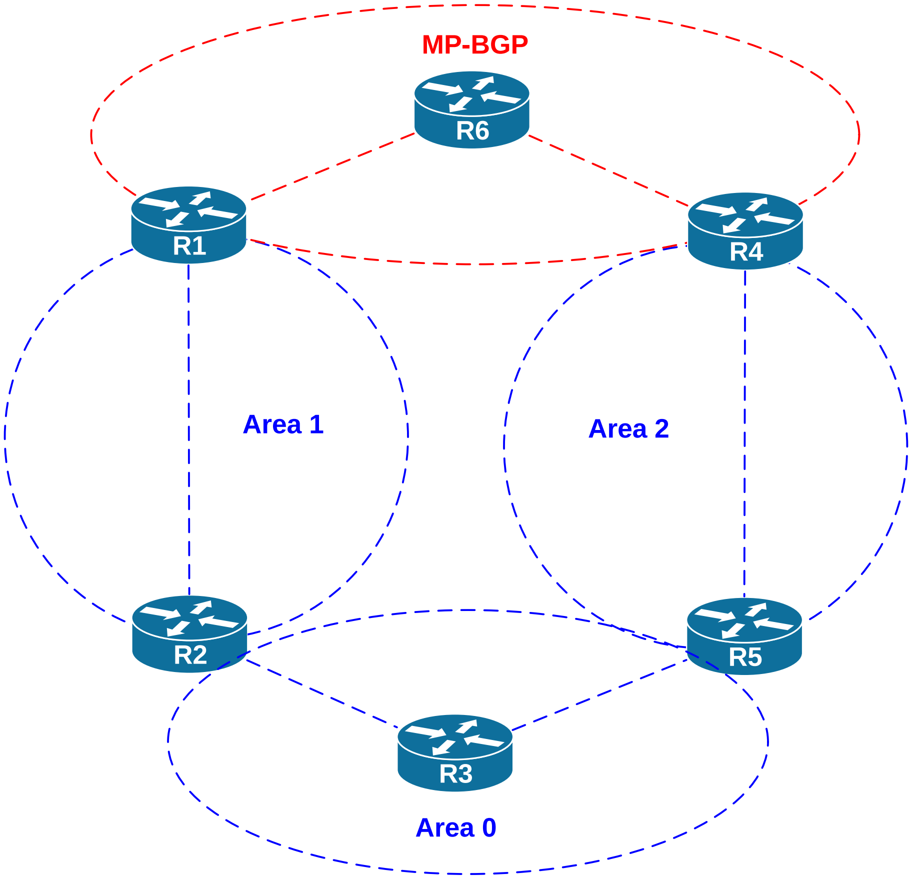
Let's assume that ASBR is injecting external information into OSPF domain. This information is flooded across every area and reaches to all routers using Type-5 LSAs. R2 should be configured as an ABR for Area 2, i.e. it should have at least one interface in Area 0. After this, it will generate a type-4 summary-LSA and propagate it to R1. However, R1 will ignore this summary LSA as received over a non-backbone area. Therefore, even though both R1 and R3 will have the type-5 LSAs, they will not be able to use them because there is no corresponding type-5 LSA for the ASBR. Those LSAs will appear in the database stating that "Adv. Router is not reachable".

In order to resolve this problem, a virtual-link should be deployed between R1 and R2. After the virtual link comes up, R2 will inject type-4 summary LSA into Area 0 and R1 will propagate it further down to R3. After this, all routers will be able to calculate the paths to the ASBR recursively via the advertising ABR.

Loop Prevention in MPLS VPNs

Q: How does this loop-prevention behavior apply in MPLS VPN scenarios?
A: PE devices running OSPF using the command **`router ospf X vrf XXX`** are assumed to have connection to OSPF super-backbone. The super-backbone is effectively an Area 0 created using redistribution into MP-BGP process. Every valid update received via MP-BGP translates into OSPF inter-area prefix and generate type-3 LSA at the PE.

The side effect is that OSPF process configured inside a VRF assumes that it ALWAYS has a valid adjacency in Area 0. Therefore, it will ignore any summary LSAs learned over a non-zero connected area. For example, look at the topology below:

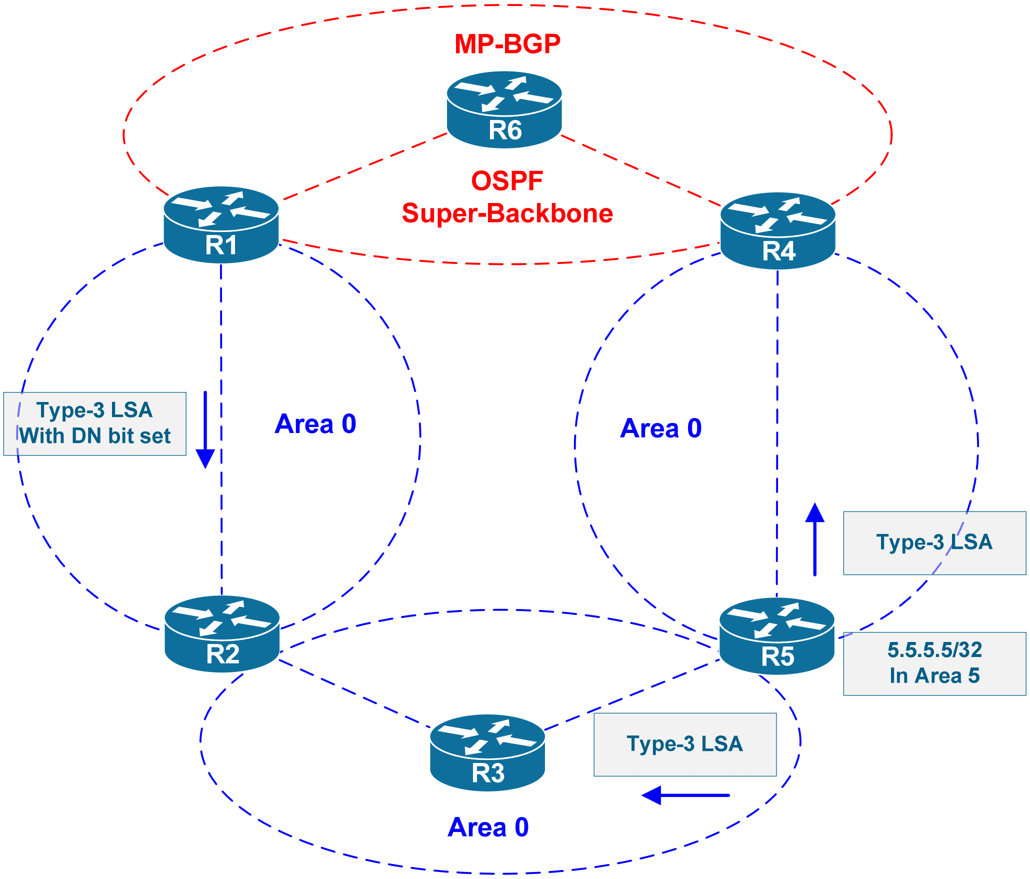


Here R1 and R4 are the PE devices, attached to OSPF super-backbone. What happens to the prefixes in Area 1? R2 advertise them as summary-LSA in Area 0 and R5 learns them via Area 0. R5 will then translate those LSAs to R4, but R4 will *ignore* them, since it has active connection to the super-backbone. Effectively, R4 will not learn the path to Area 1 destinations via R5 and will use only the path learned via MP-BGP.

Similar problem occurs on R2 and R5 – these two ABRs will ignore the summary LSAs that R1 and R4 inject into non-transit Area 1 and Area 2. The reason being, of course, is partitioned Area 0 – it is separated by non-transit Areas 1 and 2. In order to let R1 and R4 use the inter-area routes learned over Area 1 and Area 2 you need virtual-links running from R1 to R2 and from R4 to R5. This allows R4 and R1 to learn the summary-routes received over the virtual-links.

Q: What about the DN bit found in Type-3 LSAs? What's the purpose if the PE ABRs already reject summary routes?
A: Imagine you have dual-homed OSPF site that has area 0 spanning all routers (similar to the diagram below). In situation like this, you have a hierarchy of MP-BGP super-backbone and normal OSPF area 0. The routes that are leaked from MP-BGP into OSPF will be carried in type-3 summary LSAs, which ABRs must accept as they are learned over Area 0 adjacencies. This creates opening for routing loop, and this is why an additional mechanism is needed to detect routes leaking from super-backbone into regular backbone.

The DN bit is used to mark prefixes leaked from super-backbone. This allows the other PE routers connected to the same site properly detect control-plane routing loops. Look at the example below:



R3 receives two summary LSAs: one from R5, where the network 5.5.5.5/32 connects to. The other summary LSA is injected by R1 because of MP-BGP redistributed route. R3 accepts this summary LSA as it is learned over Area 0. When R5's connection to 5.5.5.5/32 fails, the following sequence of event happens:

1. R5 floods new summary LSA flushing its previous advertisement. This messages reaches both R3 and R4 (the PE device)
2. R3 selects the backup path to 5.5.5.5/32 via R2, since R1 advertised a summary-prefix for it, learned via MP-BGP.
3. R4 has the summary LSA for 5.5.5.5/32 advertised by R1, but it does not use it, since it has the DN bit set. Thus, R4 is not advertising this route into MP-BGP anymore.
4. The MP-BGP withdrawal message for 5.5.5.5/32 arrives to R1 and makes it stop advertising the summary LSA for 5.5.5.5/32.
5. R2 removes the invalid backup path and network stabilizes.

The critical piece of convergence process above was the DN bit, which allows R4 to properly distinguish inter-area route originated by another PE and not re-advertise it. If not this, R4 might have mistaken the path advertised by R1 as valid and re-injected it into MP-BGP.

Virtual Links and Sham Links

Q: What are virtual-links?
A: Virtual-links are point-to-point connections that belong to Area 0. They could be only created between two ABRs, and the virtual-link is advertised in router-LSA of the endpoint routers.

Virtual-links are only used for flooding Type-1, Type-2 and Type-3, 4 LSAs, e.g. they are purely a control-plane mechanism. There is no data-plane traffic being sent over the virtual-links, so these are not tunnels. Notice how this is different from sham-links that have to be deployed alongside some tunneling mechanism, such as MPLS.

Q: How are virtual-links being established? Why it is enough to configure just the ABR IDs using the command **`area X virtual-link <ID>`** without even specifying the tunnel endpoints?
A: Virtual-links establishment process is automated. When you configure the abovementioned command, the following set of steps occurs:

1. The initiating router looks up the router LSA for the endpoint ID configured. The LSA should exist and it should have the B (border) bit set, signalizing the other end is an ABR.
2. The initiating router performs path calculation from itself to the destination ABR. Since this is an intra-area calculation, this process reveals the links connecting the endpoints to the topology.
3. Using the identified links, the ABR may extract the IP addresses of associated routers to create tunnel endpoint. You may see these addresses using the command **`show ip ospf interface brief`**. This process is dynamic and may change the endpoint IP addresses should the selected links go down.
    a. Notice that this address discovery process may fail, if one or both routers are connected to the transit area using unnumbered interfaces.
    b. In OSPFv3, similar discovery failure may happen if the routers are connected to the transit area using interfaces with link-local addresses only.
4. Knowing the destination IP addresses, initiating router starts sending OSPF Hello packets using the locally configured Hello/Dead timers (10/40 by default) and Area 0 ID in the packets.

5. If the receiving side is configured for the virtual link, it recognizes the sending router ID and responds with Hello packets. A point-to-point adjacency is formed and databases are exchanged, flooding LSA types 1, 2, 3/4 but not type-5, as those are already flooded to all non-stub areas.

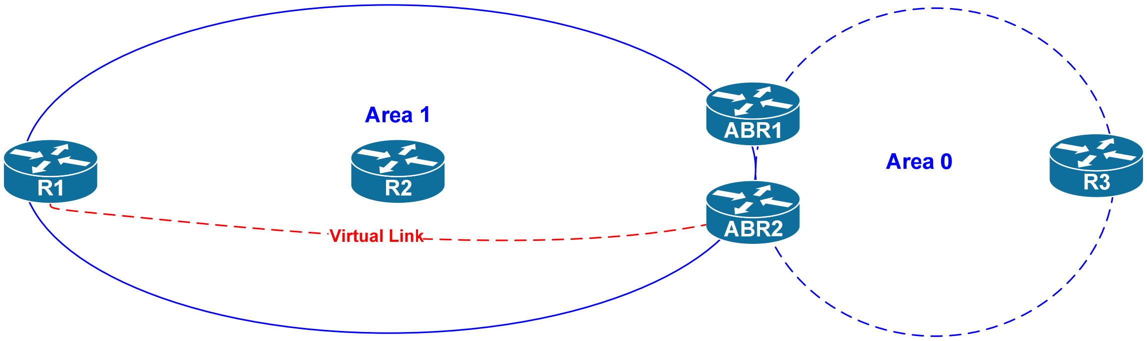
Q: How are virtual-links being monitored for availability?
A: After the database exchange has been completed, further Hello packets are suppressed on the virtual link. Effectively, the link is treated as a demand circuit. Only intermittent OSPF database changes are flooded, i.e. those triggered by topology changes. The reason for this is that virtual-link liveness is detected based on the reachability of the remote ABR, so Hello probing is not required. The LSAs exchanged over the virtual-link have DNA (do-not-age) bit set, which prevents their expiration in LSDB and requires no periodic flooding over virtual-links to refresh the LSAs.

Q: What if authentication is enabled on the virtual-link?
A: Authentication only has effect during the initial Hello packet exchange and any consequent LSAs exchanges. If you apply authentication after the virtual link went up and become fully adjacent, you will see link remaining in the up state, even if authentication settings do not match. However, in case of mismatching authentication the routers would not be able to exchange LSAs – the LS update packets will never get acknowledged. This effect could be observed using the command **`show ip ospf flood-list`** or **`show ip ospf retransmission-list`**.

The problem here is the retransmission process in OSPF. In never times out, a router keeps sending periodic LSA retransmits until it receives and acknowledgement back. Adjacency liveness is supposed to be validated using Hello process, which is suspended over virtual links. Furthermore, Cisco never implemented RFC 3883 – detecting inactive neighbor over demand circuits using LSA probing, and therefore cannot detect this sort of misconfigurations.

Q: How does OSPF determine virtual-link cost?
A: The cost of the virtual-link is calculated dynamically, based on the cost to transit the underlying intra-area topology. If this topology changes, so does the virtual-link cost. Sine virtual link is a normal OSPF link advertised in type-1 LSA, it could not have the cost above 65535. However, it is possible to the intra-area path to exceed this cost, e.g. as a result of cost manipulation or changing the reference bandwidth. In this situation, both virtual-link endpoints will declare it down and stop using. That is, if you have a virtual-link across the area, the maximum path cost in this area should not exceed 65535.

Q: How routes are being resolved over the virtual-link?



A: Look at the diagram above. R1 is learning summary LSAs for Area 0 routes from ABR1/ABR2 (as type 3 LSAs) and learning the links in area 0 via the virtual-link (type-1/2 LSAs). Consequently, R1, is injecting the summary LSAs for Area 0 into Area 1, covering all prefixes in area 0.

How R1 and R2 are supposed to make decision in this situation? Let's look at R1 first. There are two options for this ABR:

1. Area 1 has *transit capability* enabled on R1 (**`area 1 capability transit`**). This is the default behavior. R1 will compare the information learned in the summary LSAs from ABR1 and ABR2 and see if it matches the routing information constructed using the virtual-link connection to Area 0. After this, it will select the shortest path among the routes based purely on their metrics. For example, if ABR1 advertises better cost to some prefix in Area 0, R1 will prefer it. The routes for Area 0 destinations will still show up as intra-area on R1. Notice how the ABR is in fact using the LSAs received over the transit area to optimize routing decision.
2. Area 1 has *transit capability* disabled on R1 (**`no area 1 capability transit)`**. In this case, R1 will assume it has a single link connecting it to Area 0. It will use the virtual-link cost to calculate paths to Area 0 destinations and install in the local routing table as intra-area routes. It will resolve every destination via the next-hop of the advertising ABR, in our case ABR2. The prefixes advertised by ABR1 are ignored, even if they offer better path costs.

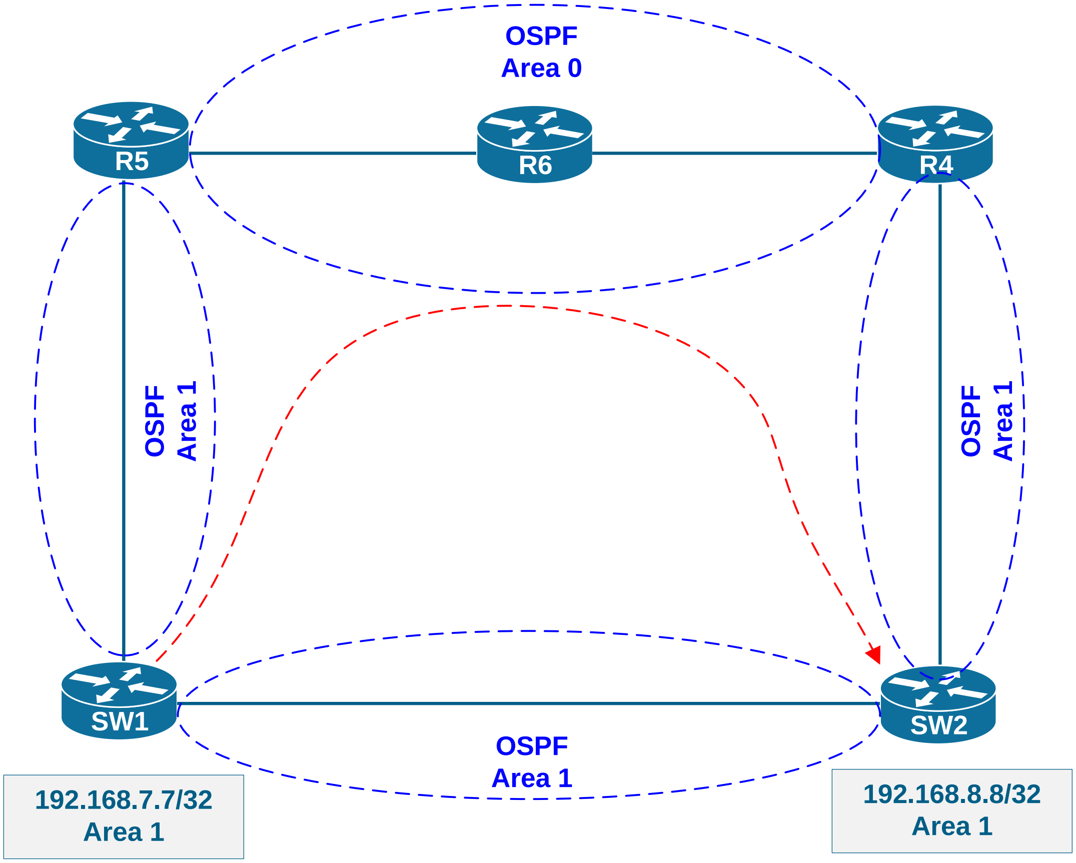
What about R2, the intra-area router in transit area? It will receive the summary advertisements from R1, ABR1 and ABR2. How will it choose the correct path? In fact, the cost associated with R1 advertisements will be higher than with the LSAs advertised by ABR1 and ABR2. The reason being, R1 advertises the cost of the virtual link in the summary LSAs, and the virtual-link goes across R2. Therefore, R2 will never select the summary LSAs advertised by R1, only those advertised by ABR1 and ABR2.

However, the intra-area path selection on R2 *could* be broken if ABR1 and ABR2 summarize Area 0 routes while advertising them into the area with virtual-link. In this case, R2 may prefer summary LSAs from R1 and create a routing loop, as R1 will kick the packet back to R2. This is why ABRs never summarize Area 0 prefixes when inject summary into an area having a virtual link. It is possible to summarize other, non backbone prefixes, into a transit area, but this is because those are not re-advertised on the other end of the virtual link.

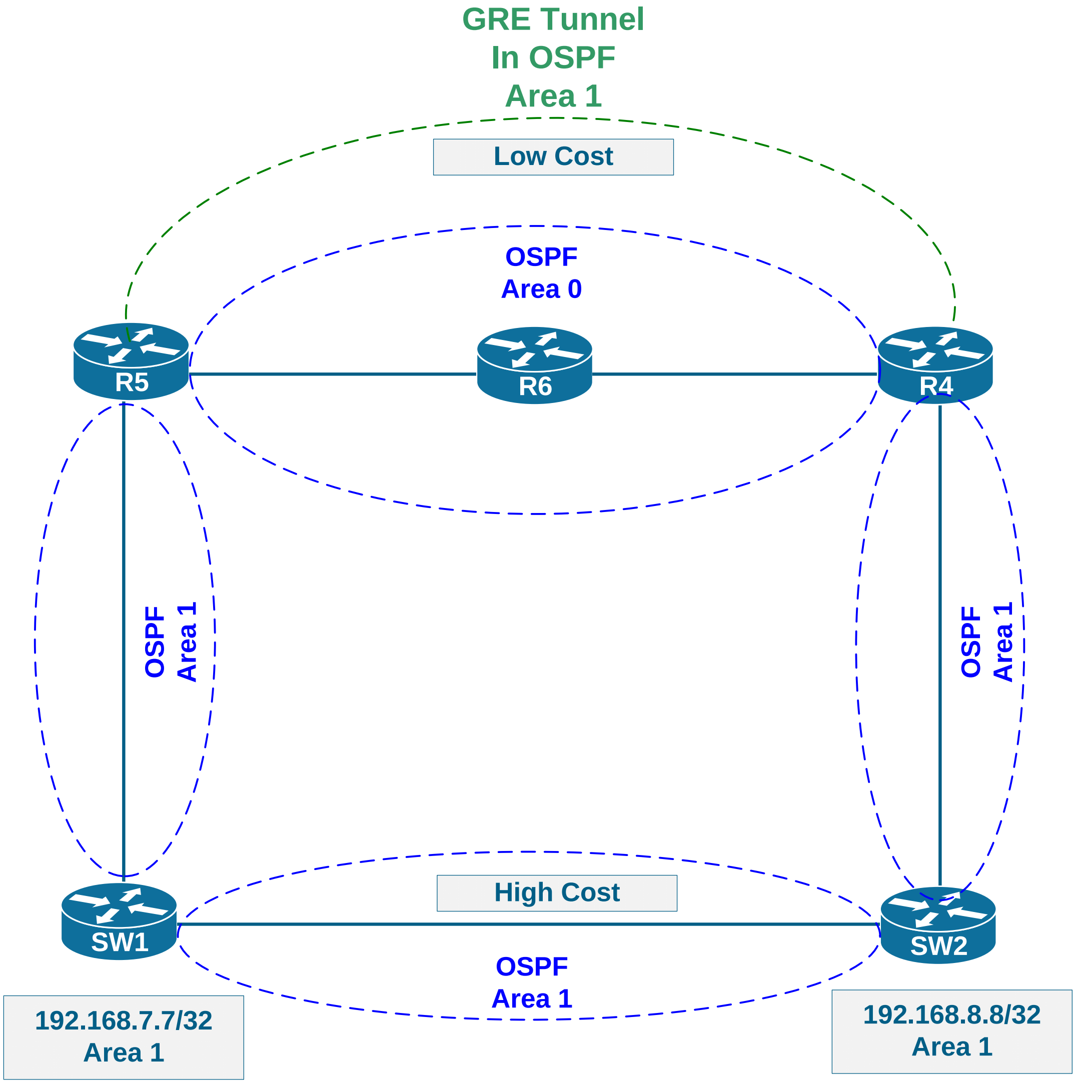
Q: Why I can't get virtual link up if my routers don't have an interface in area 0?
A: Like we mentioned previously, virtual link destination needs to be an ABR. Until an endpoint of the virtual link sees the other endpoint as ABR, it will not initiate the Hello message process.

If you configure the remote endpoint with a Loopback in area 0, it will advertise itself as an ABR and the other endpoint will start sending Hello packets. The fresh ABR will respond with Hellos, and the initiating router will move the virtual link out from DOWN state. This will immediately make both routers the ABRs, in turn. You may even remove the "seed" Loopback interface, and both routers will still remain ABR by the fact of having the virtual link up. However, if the virtual link would go down for some reason, it will not return back to functioning unless you have that seed interface in area 0.

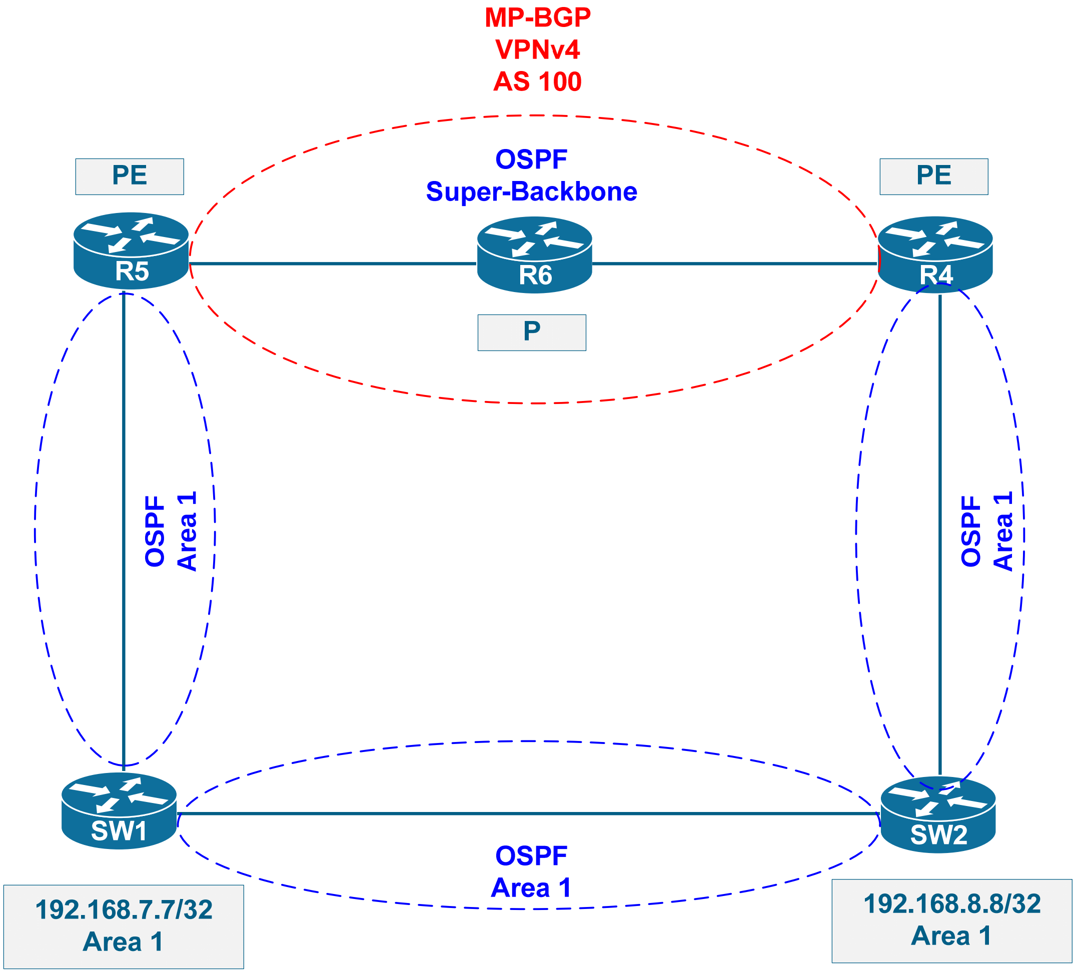
Q: What is the purpose of sham-links?
A: To understand the need for sham links, look at the "classic" OSPF scenario below first. We want traffic from 192.168.7.7 going to 192.168.8.8 to prefer the "longer" path via Area 0. SW1 does receive the summary LSA for 192.168.8.8/32 from R5 but never uses it, as it has intra-area path to the same destination. We may change the area numbers assigned to links for 192.168.7.7/32 and 192.168.8.8/32 to make the selection rules equal, but let's assume this is not possible.



If changing areas is not an option, we could deploy an IP tunnel between R4 and R5, using Area 0 addresses to source tunnel packets (this avoids recursive routing). We can then run OSPF over this tunnel and assign this link to OSPF Area 1. After this, all we need to do is adjust OSPF cost values on the tunnel and the backup link connecting SW1 and SW3.



This scenario could be easily converted to a multi-homed OSPF site connected to MPLS/BGP VPN backbone. From the viewpoint of OSPF process, MP-BGP clouds looks like an OSPF “super-backbone”, since it is used to pass OSPF inter-area routes from site to site.

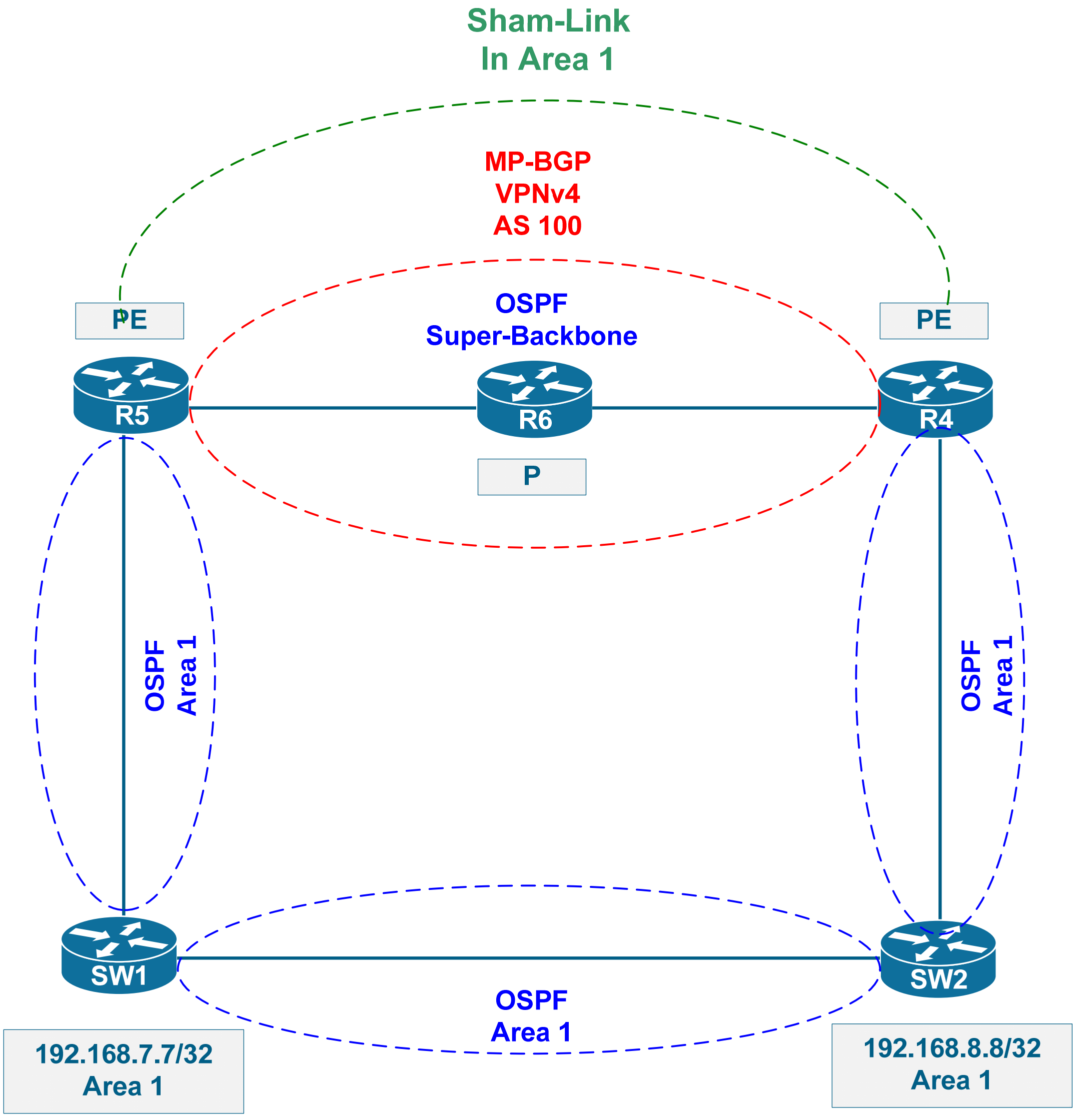


What happens here? Let’s look at R5 for example. R5 receives a router-LSA from SW1 and SW2 and learns about links with the prefixes 192.168.7.7/32 and 192.168.8.8/32. These are installed into the local routing table as OSPF routes and redistributed into MP-BGP process on R5. R4 learns the same router LSAs from SW1 and SW2 and creates OSPF routes for 192.168.8.8/32 and 192.168.7.7/32. These prefixes are redistributed into MP-BGP as well.

R5’s BGP process receives prefers the prefixes 192.168.7.7/32 and 192.168.8.8/32 injected locally since they have better weight and BGP prefers local prefixes per the best-path selection process. The paths advertised to the same prefixes from R4 are kept in the BGP table as “backup”. Therefore, R4 and R5 will always prefer locally learned OSPF paths due to BGP best-path selection process.

At this point, one may suggest using BGP Cost community to allow the paths learned from remote peer to override locally injected paths during BGP best-path selection. However, even if that would happen, OSPF would prefer the intra-area route constructed using intra-area topology over the summary LSA generate based on MP-BGP update. Thus, the trick that worked well with EIGRP will not work with OSPF. An alternate solution is required.

The solution is the same as using the “GRE” link we demonstrated before in “classic” OSPF scenario. However, this time, instead of GRE tunnel we’ll be using OSPF sham-link. Sham link is unicast control-plane tunnel that runs between the PE routers.



The main difference between sham-link and GRE tunnel is that sham-link is a control-plane only mechanism, similar to virtual-link. However, unlike virtual-link, the shame-link could be assigned into any area and could have its cost manually configured. The reason being is the fact that sham-link runs over MP-BGP cloud and that has no OSPF topology information available, hence automatic provisioning is not possible. You need to configured sham-link endpoint addresses and area manually to get it up and running. After OSPF processes hear each other on the sham-link by listening to Hello messages, they establish a point-to-point adjacency and perform LSDB synchronization. The sham-link is advertised in router LSA of the both routers as Type-1 link (point-to-point connection).

What happens after OSPF database synchronization over sham-link? Take R5 for instance, using the reference diagram above. Assume the backup link cost is *higher* than the sham-link cost.

1. R5 receives router LSA for SW2 via normal OSPF flooding process and learns about 192.168.8.8/32. However, now R5 has two intra-area (SPF calculated) paths to SW2: one across the backdoor link and another over the sham-link.
2. Based on the fact that sham-link cost is more preferable, R5 install intra-area OSPF route in the local table with the next-hop of sham-link remote endpoint.
3. Furthermore, BGP redistribution process DOES NOT redistribute OSPF routes with the next-hop address of the sham-link endpoint back into BGP. Based this, R5's BGP table will only have prefix/path entry for 192.168.8.8/32 that R4 injects into BGP based on its intra-area route.
4. Forwarding database for OSPF routes resolved over sham-link is populated based on corresponding BGP prefix information, even though the RIB contains the OSPF route. This allows for applying proper label stack and tunnel packets over the MPLS/BGP VPN backbone.

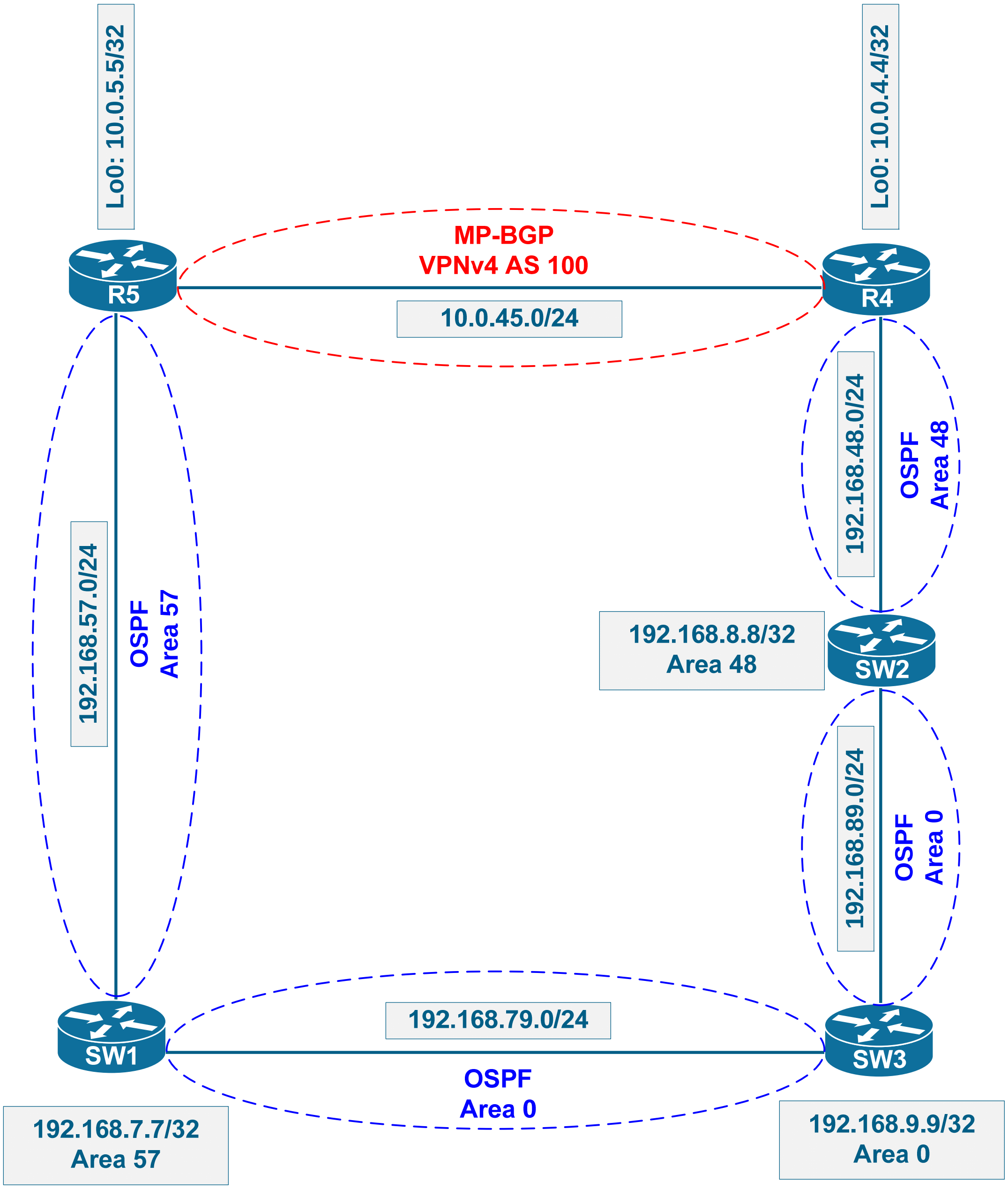
Effectively, sham-link creates a point-to-point OSPF adjacency that is used for control-plan information exchange. Data-plane forwarding is based on BGP information, i.e. BGP is used to propagate MPLS labels for OSPF prefixes. Notice that you would need sham-link even if you have Area 0 in the customer site, as OSPF super-backbone and OSPF Area 0 are not considered to be the same area.

Keep in mind the following practical configuration rules for sham-links:

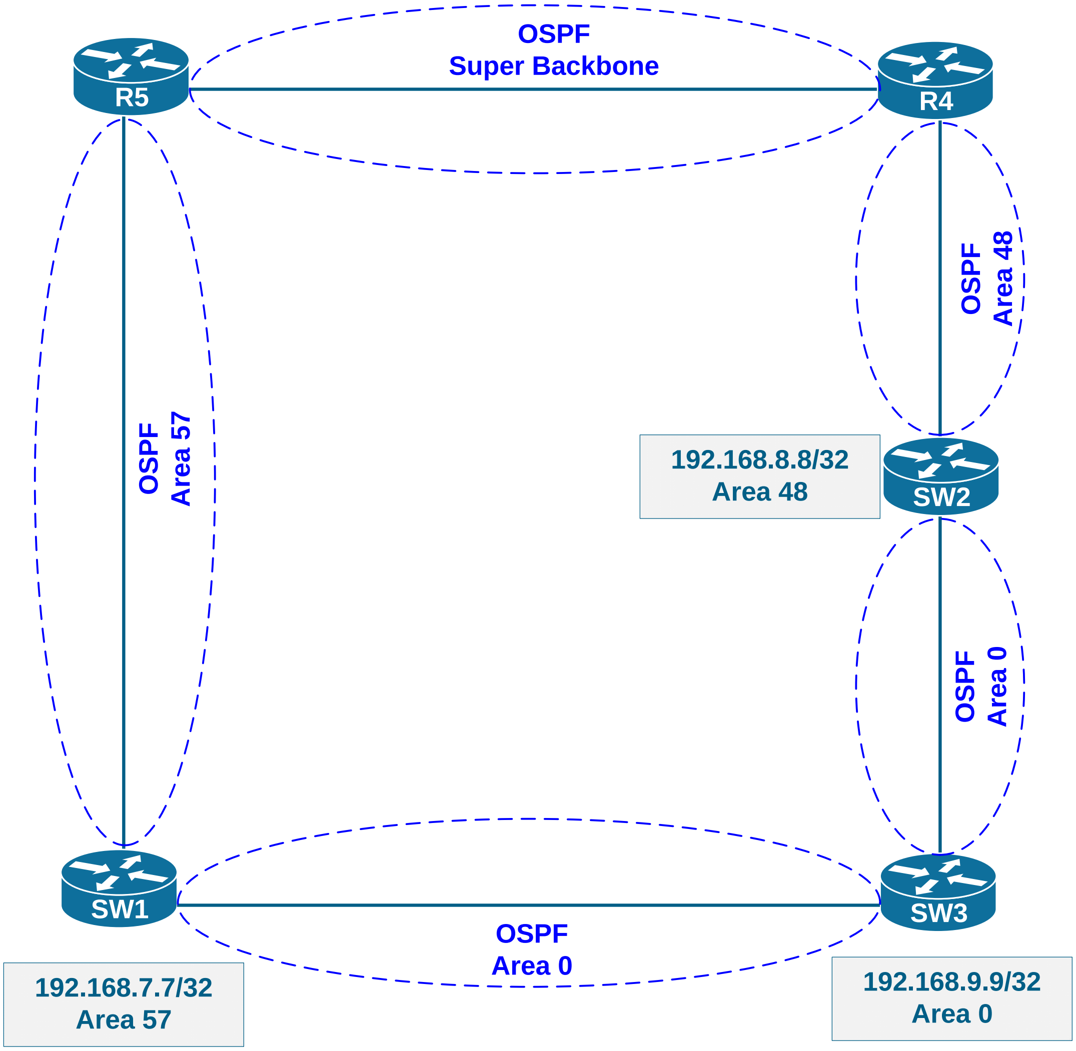
1. Sham-Link endpoints must be advertised as /32 host addresses.
2. Sham-Link endpoints must be advertised into BGP only.
3. Sham-Link cost could be manually configured, with the default being 1.

Multi-Homed Multi-Area OSPF Site

Q: What about a multi-homed customer site scenario with multiple OSPF areas deployed?
A: Traffic engineering in such configuration becomes more complicated, involving the use of both sham-links and virtual-links. Look at the topology below. You may find the IOS configuration snippets in the Appendix for this document.



This topology represents multi-homed OSPF site connected to MPLS/BGP VPN service provider. On the first sight, this topology looks normal, as both non-backbone areas are directly connected to on-site ("true") area 0. However, as you remember, MP-BGP cloud is treated as OSPF super-backbone, an additional layer of route-redistribution hierarchy. Therefore, from OSPF point of view, the topology actually looks like this:



And apparently there is an issue there: non-transit areas 57 and 48 separate the backbone area. The problem is that areas 57 and 48 could not be used for transit to reach inter-area routes. Let's have a look at prefix 192.168.7.7/32 on SW2:

```
SW2#show ip ospf database summary 192.168.7.7

            OSPF Router with ID (192.168.8.8) (Process ID 100)

                Summary Net Link States (Area 0)

  Routing Bit Set on this LSA
  LS age: 110
  Options: (No TOS-capability, DC, Upward)
  LS Type: Summary Links(Network)
  Link State ID: 192.168.7.7 (summary Network Number)
  Advertising Router: 192.168.7.7
  LS Seq Number: 80000001
  Checksum: 0xE666
  Length: 28
  Network Mask: /32
        TOS: 0  Metric: 1
```

This inter-area LSA has been advertised by SW1 into Area 0. Notice that routing bit is set on this LSA, which means it is being used for routing. SW2 is supposed to “translate” this information into all attached non-backbone areas.

```
                Summary Net Link States (Area 48)

  LS age: 37
  Options: (No TOS-capability, DC, Downward)
  LS Type: Summary Links(Network)
  Link State ID: 192.168.7.7 (summary Network Number)
  Advertising Router: 192.168.4.4
  LS Seq Number: 80000001
  Checksum: 0x9041
  Length: 28
  Network Mask: /32
        TOS: 0  Metric: 2
```

This is the LSA that R4 (the PE) sent to SW2. It has been constructed based on information that R4 learned via MP-BGP:

```
R4#show bgp vpnv4 unicast all 192.168.7.7
BGP routing table entry for 100:1:192.168.7.7/32, version 6
Paths: (1 available, best #1, table VPN_A)
Flag: 0x820
  Not advertised to any peer
  Local
    10.0.5.5 (metric 65) from 10.0.5.5 (10.0.5.5)
      Origin incomplete, metric 2, localpref 100, valid, internal, best
      Extended Community: RT:100:1 OSPF DOMAIN ID:0x0005:0x000000640200
        OSPF RT:0.0.0.57:2:0 OSPF ROUTER ID:192.168.5.5:0
      mpls labels in/out nolabel/17
```

Notice that R4 does not install the local path to 192.167.7.7/32 because it ignores the summary LSA that SW2 advertises (see below)

```
  LS age: 61
  Options: (No TOS-capability, DC, Upward)
  LS Type: Summary Links(Network)
  Link State ID: 192.168.7.7 (summary Network Number)
  Advertising Router: 192.168.8.8
  LS Seq Number: 80000001
  Checksum: 0xED5B
  Length: 28
  Network Mask: /32
        TOS: 0  Metric: 3
```

Finally, this is the LSA that SW2 originated into area 48 on its own. In fact, this LSA is the "translation" of the summary LSA found in Area 0 (see above).

If we shut down SW3' connection to SW1, SW2 will lose its route to 192.168.7.7/32 as it cannot use the summary LSA learned from R4 across non-transit area:

```
SW3(config)#interface fastEthernet 0/13
SW3(config-if)#shutdown

SW2#show ip route ospf
     192.168.9.0/32 is subnetted, 1 subnets
O       192.168.9.9 [110/2] via 192.168.89.9, 00:00:16,
FastEthernet0/16
```

```
SW2#show ip ospf database summary 192.168.7.7

            OSPF Router with ID (192.168.8.8) (Process ID 100)

                Summary Net Link States (Area 0)

  LS age: 1238
  Options: (No TOS-capability, DC, Upward)
  LS Type: Summary Links(Network)
  Link State ID: 192.168.7.7 (summary Network Number)
  Advertising Router: 192.168.7.7
  LS Seq Number: 80000001
  Checksum: 0xE666
  Length: 28
  Network Mask: /32
        TOS: 0  Metric: 1
```

This is SW'1 original LSA. Since we terminated connection to SW1, it never purged that LSA. However, there is no topological path to the advertising router, so this summary LSA learned via Area 0 could not be used, even though it is kept in the database.

```
                Summary Net Link States (Area 48)

  LS age: 1165
  Options: (No TOS-capability, DC, Downward)
  LS Type: Summary Links(Network)
  Link State ID: 192.168.7.7 (summary Network Number)
  Advertising Router: 192.168.4.4
  LS Seq Number: 80000001
  Checksum: 0x9041
  Length: 28
  Network Mask: /32
        TOS: 0  Metric: 2
```

This is the summary LSA learned from R4. It does not have routing bit set, because it has been learned via a non-transit area.

Let's see what happens when we eliminate SW2's adjacency in Area 0:

```
SW2(config)#interface fastEthernet 0/16
SW2(config-if)#shutdown
```

```
SW2#show ip ospf
 Routing Process "ospf 100" with ID 192.168.8.8
...
 Reference bandwidth unit is 100 mbps
    Area BACKBONE(0) (Inactive)
        Number of interfaces in this area is 1
        Area has no authentication
        SPF algorithm last executed 00:00:05.293 ago
        SPF algorithm executed 8 times
        Area ranges are
        Number of LSA 7. Checksum Sum 0x037365
        Number of opaque link LSA 0. Checksum Sum 0x000000
        Number of DCbitless LSA 0
        Number of indication LSA 0
        Number of DoNotAge LSA 0
        Flood list length 0
```

Now SW2 has the route to 192.168.7.7/32 via R4, as it lost its adjacency in Area 0:

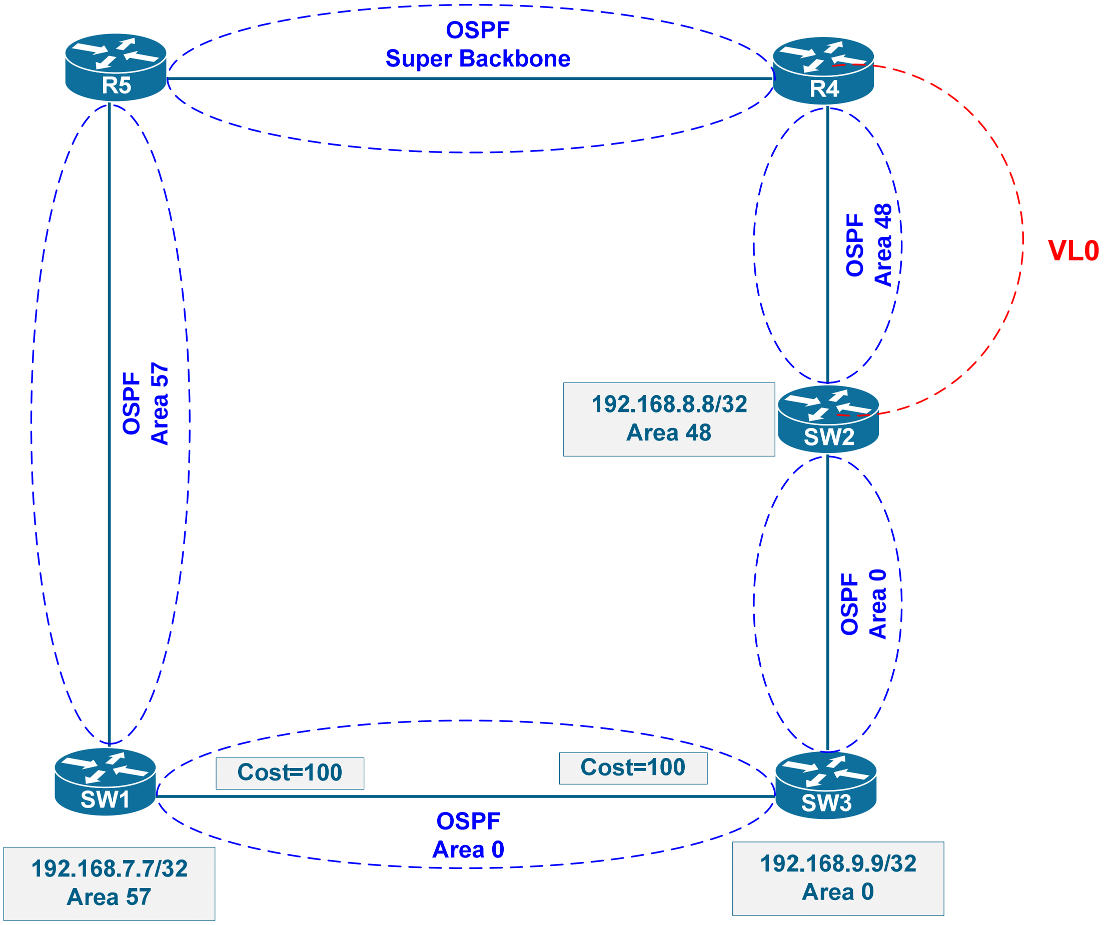
```
SW2#show ip route ospf
O IA 192.168.57.0/24 [110/2] via 192.168.48.4, 00:01:27,
FastEthernet0/4
     192.168.7.0/32 is subnetted, 1 subnets
O IA    192.168.7.7 [110/3] via 192.168.48.4, 00:01:27, FastEthernet0/4

SW2#show ip ospf database summary 192.168.7.7

            OSPF Router with ID (192.168.8.8) (Process ID 100)

                Summary Net Link States (Area 0)

  LS age: 1532
  Options: (No TOS-capability, DC, Upward)
  LS Type: Summary Links(Network)
  Link State ID: 192.168.7.7 (summary Network Number)
  Advertising Router: 192.168.7.7
  LS Seq Number: 80000001
  Checksum: 0xE666
  Length: 28
  Network Mask: /32
        TOS: 0  Metric: 1
```

```
              Summary Net Link States (Area 48)

  Routing Bit Set on this LSA
  LS age: 1458
  Options: (No TOS-capability, DC, Downward)
  LS Type: Summary Links(Network)
  Link State ID: 192.168.7.7 (summary Network Number)
  Advertising Router: 192.168.4.4
  LS Seq Number: 80000001
  Checksum: 0x9041
  Length: 28
  Network Mask: /32
        TOS: 0  Metric: 2
```

And routing bit has been finally set on the LSA.

Restore the topology to normal state but un-shutting the interface. Let’s think what could be done to make SW2 prefer the path to 192.168.7.7/32 via R4 under normal conditions. Firstly, we know that in order to make area transit we need a virtual link across it. In order to make Area 48 transit we need a virtual link across it. This should join the super-backbone in the normal Area 0. In order to make the path via super-backbone more preferable, it seems like we could do that by changing the cost of the backdoor link between SW1 and SW3 to 100.

```
SW2#show ip ospf virtual-links
Virtual Link OSPF_VL0 to router 192.168.4.4 is up
  Run as demand circuit
  DoNotAge LSA allowed.
  Transit area 48, via interface FastEthernet0/4, Cost of using 1
  Transmit Delay is 1 sec, State POINT_TO_POINT,
  Timer intervals configured, Hello 10, Dead 40, Wait 40, Retransmit 5
    Hello due in 00:00:04
    Adjacency State FULL (Hello suppressed)
    Index 1/2, retransmission queue length 0, number of retransmission
0
    First 0x0(0)/0x0(0) Next 0x0(0)/0x0(0)
    Last retransmission scan length is 0, maximum is 0
    Last retransmission scan time is 0 msec, maximum is 0 msec
```



Check OSPF database on SW2 after these changes:

```
SW2#show ip ospf database summary 192.168.7.7

            OSPF Router with ID (192.168.8.8) (Process ID 100)

                Summary Net Link States (Area 0)

  Routing Bit Set on this LSA
  LS age: 170
  Options: (No TOS-capability, DC, Upward)
  LS Type: Summary Links(Network)
  Link State ID: 192.168.7.7 (summary Network Number)
  Advertising Router: 192.168.7.7
  LS Seq Number: 80000001
  Checksum: 0xE666
  Length: 28
  Network Mask: /32
        TOS: 0  Metric: 1
```

This is the LSA advertised by SW1 and used for routing by SW2. It has been flooded by SW1 into Area 0.

```
                Summary Net Link States (Area 48)

  LS age: 114
  Options: (No TOS-capability, DC, Upward)
  LS Type: Summary Links(Network)
  Link State ID: 192.168.7.7 (summary Network Number)
  Advertising Router: 192.168.8.8
  LS Seq Number: 80000001
  Checksum: 0xED5B
  Length: 28
  Network Mask: /32
        TOS: 0  Metric: 3
```

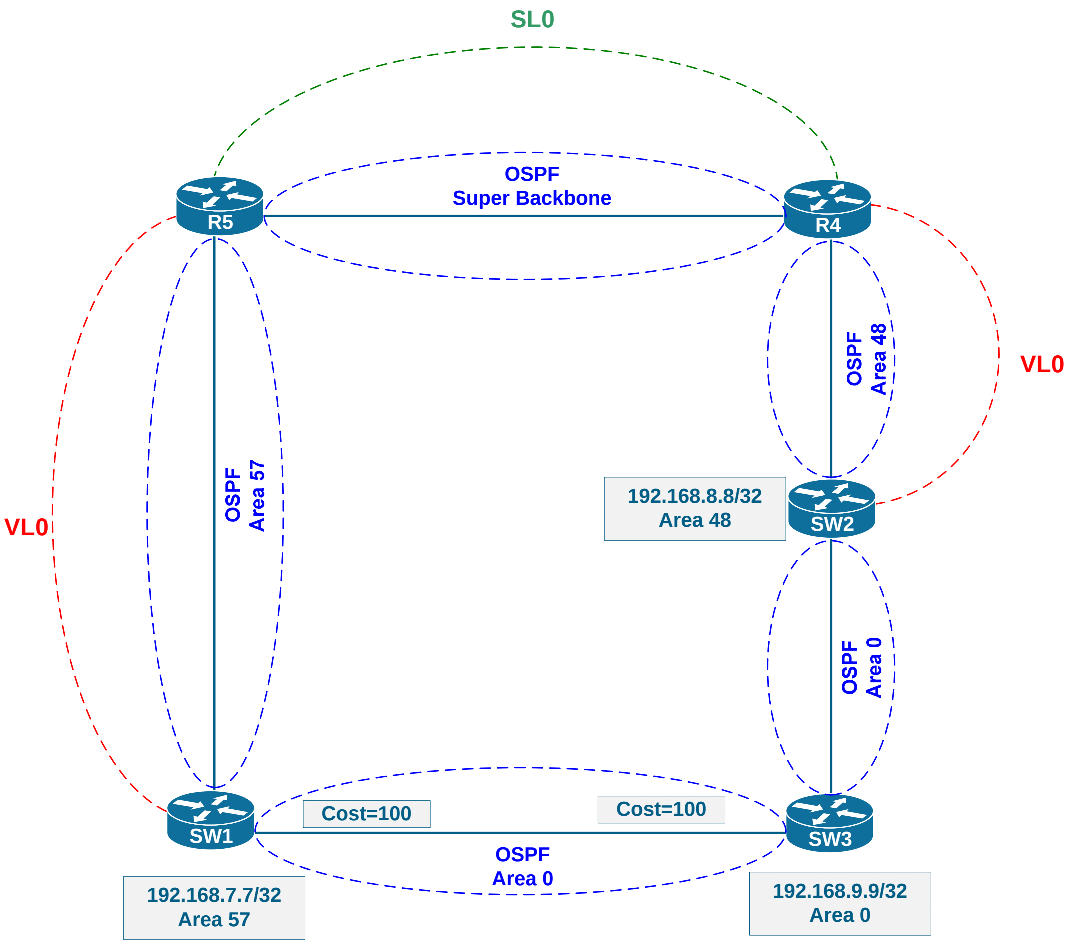
This is the LSA that SW2 generates into Area 48. But there is no longer any summary LSA for 192.168.7.7/32 from R4 any longer! What happened?

This behavior is due to the fact that R4 now accepts the summary LSA for 192.168.7.7/32 advertised by SW2 for routing. In result, this becomes the new MP-BGP best-path – locally sourced on R4. R4 compares the prefix received via MP-BGP session and the one sourced locally and the last one is always better by the virtue of being locally-sourced and having better weight.

```
R4#show bgp vpnv4 unicast all 192.168.7.7
BGP routing table entry for 100:1:192.168.7.7/32, version 20
Paths: (2 available, best #1, table VPN_A)
  Advertised to update-groups:
        1
  Local
    192.168.48.8 from 0.0.0.0 (10.0.4.4)
      Origin incomplete, metric 4, localpref 100, weight 32768, valid,
sourced, best
      Extended Community: RT:100:1 OSPF DOMAIN ID:0x0005:0x000000640200
        OSPF RT:0.0.0.0:3:0 OSPF ROUTER ID:192.168.4.4:0
      mpls labels in/out 23/nolabel
  Local
    10.0.5.5 (metric 65) from 10.0.5.5 (10.0.5.5)
      Origin incomplete, metric 2, localpref 100, valid, internal
      Extended Community: RT:100:1 OSPF DOMAIN ID:0x0005:0x000000640200
        OSPF RT:0.0.0.57:2:0 OSPF ROUTER ID:192.168.5.5:0
      mpls labels in/out 23/17
```

If we want R4 to prefer the MP-BGP path, we need to run a sham-link in area 0 over the super-backbone. This will let OSPF learn the 192.168.7.7/32 prefix as summary-LSA but received directly via OSPF adjacency. In result, OSPF will perform the selection based on the summary-LSA metrics, eliminating the BGP best-path decision step. Let’s create a sham-link:

```
R4#show ip ospf sham-links
Sham Link OSPF_SL0 to address 192.168.55.55 is up
Area 0 source address 192.168.44.44
  Run as demand circuit
  DoNotAge LSA allowed. Cost of using 1 State POINT_TO_POINT,
  Timer intervals configured, Hello 10, Dead 40, Wait 40,
    Hello due in 00:00:06
    Adjacency State FULL (Hello suppressed)
    Index 2/3, retransmission queue length 0, number of retransmission
0
    First 0x0(0)/0x0(0) Next 0x0(0)/0x0(0)
    Last retransmission scan length is 0, maximum is 0
    Last retransmission scan time is 0 msec, maximum is 0 msec
```



Now R4 has the summary LSA for 192.168.7.7/32 advertised by R5 (the ABR) over the sham-link adjacency:

```
R4#show ip ospf database summary 192.168.7.7

            OSPF Router with ID (192.168.4.4) (Process ID 100)

                Summary Net Link States (Area 0)

  Routing Bit Set on this LSA
  LS age: 433 (DoNotAge)
  Options: (No TOS-capability, DC, Upward)
  LS Type: Summary Links(Network)
  Link State ID: 192.168.7.7 (summary Network Number)
  Advertising Router: 192.168.5.5
  LS Seq Number: 80000004
  Checksum: 0x548
  Length: 28
  Network Mask: /32
        TOS: 0  Metric: 2
```

This is the LSA we received from R5 over the sham-link - notice the DNA bit. This path has the metric cost of “2” and is used for routing on R4. Pay attention to the fact that this LSA is marked as “Upward”, i.e. the DN bit is not set on it. The DN bit is only set when redistributing MP-BGP routes into OSPF.

```
  LS age: 644 (DoNotAge)
  Options: (No TOS-capability, DC, Upward)
  LS Type: Summary Links(Network)
  Link State ID: 192.168.7.7 (summary Network Number)
  Advertising Router: 192.168.7.7
  LS Seq Number: 8000000E
  Checksum: 0xCC73
  Length: 28
  Network Mask: /32
        TOS: 0  Metric: 1
```

What is this LSA? We have this LSA in the database because of the virtual-link connecting R4 to “true Area 0” (notice the DNA bit). However, the summary route metric is 1, so why this information is not preferred for routing over the summary route advertised by R5? Recall that we changed the cost on the backdoor link between SW1 and SW3. This creates the higher cost path to reach SW1 (the ABR) over the Area 0 topology:

```
R4#show ip ospf 100 border-routers

OSPF Process 100 internal Routing Table

Codes: i - Intra-area route, I - Inter-area route

i 192.168.5.5 [1] via 10.0.5.5, OSPF_SL0, ABR/ASBR, Area 0, SPF 11
i 192.168.7.7 [102] via 192.168.48.8, FastEthernet0/0, ABR, Area 0, SPF 11
i 192.168.8.8 [1] via 192.168.48.8, FastEthernet0/0, ABR, Area 0, SPF 11
i 192.168.8.8 [1] via 192.168.48.8, FastEthernet0/0, ABR, Area 48, SPF 12
```

```
                Summary Net Link States (Area 48)

  LS age: 198
  Options: (No TOS-capability, DC, Upward)
  LS Type: Summary Links(Network)
  Link State ID: 192.168.7.7 (summary Network Number)
  Advertising Router: 192.168.4.4
  LS Seq Number: 80000001
  Checksum: 0x222F
  Length: 28
  Network Mask: /32
        TOS: 0  Metric: 3
```

This final summary LSA is the one R4 injects in Area 48 translating the information it has learned from Area 0.

What happened to the summary LSA for 192.168.7.7/32 that SW2 is supposed to inject into virtual-link? It has been superseded by the summary LSA that SW2 has learned over the virtual-link from R4:

```
SW2#show ip ospf database summary 192.168.7.7

            OSPF Router with ID (192.168.8.8) (Process ID 100)

                Summary Net Link States (Area 0)

  Routing Bit Set on this LSA
  LS age: 434 (DoNotAge)
  Options: (No TOS-capability, DC, Upward)
  LS Type: Summary Links(Network)
  Link State ID: 192.168.7.7 (summary Network Number)
  Advertising Router: 192.168.5.5
  LS Seq Number: 80000004
  Checksum: 0x548
  Length: 28
  Network Mask: /32
        TOS: 0  Metric: 2
```

This is the summary LSA that R5 has originated over the sham-link and that has reached SW2 over the virtual-link, effectively traversing both virtual connections. Notice the metric value of 2 and the “routing bit” set on the LSA.

```
  Routing Bit Set on this LSA
  LS age: 1983
  Options: (No TOS-capability, DC, Upward)
  LS Type: Summary Links(Network)
  Link State ID: 192.168.7.7 (summary Network Number)
  Advertising Router: 192.168.7.7
  LS Seq Number: 8000000E
  Checksum: 0xCC73
  Length: 28
  Network Mask: /32
        TOS: 0  Metric: 1
```

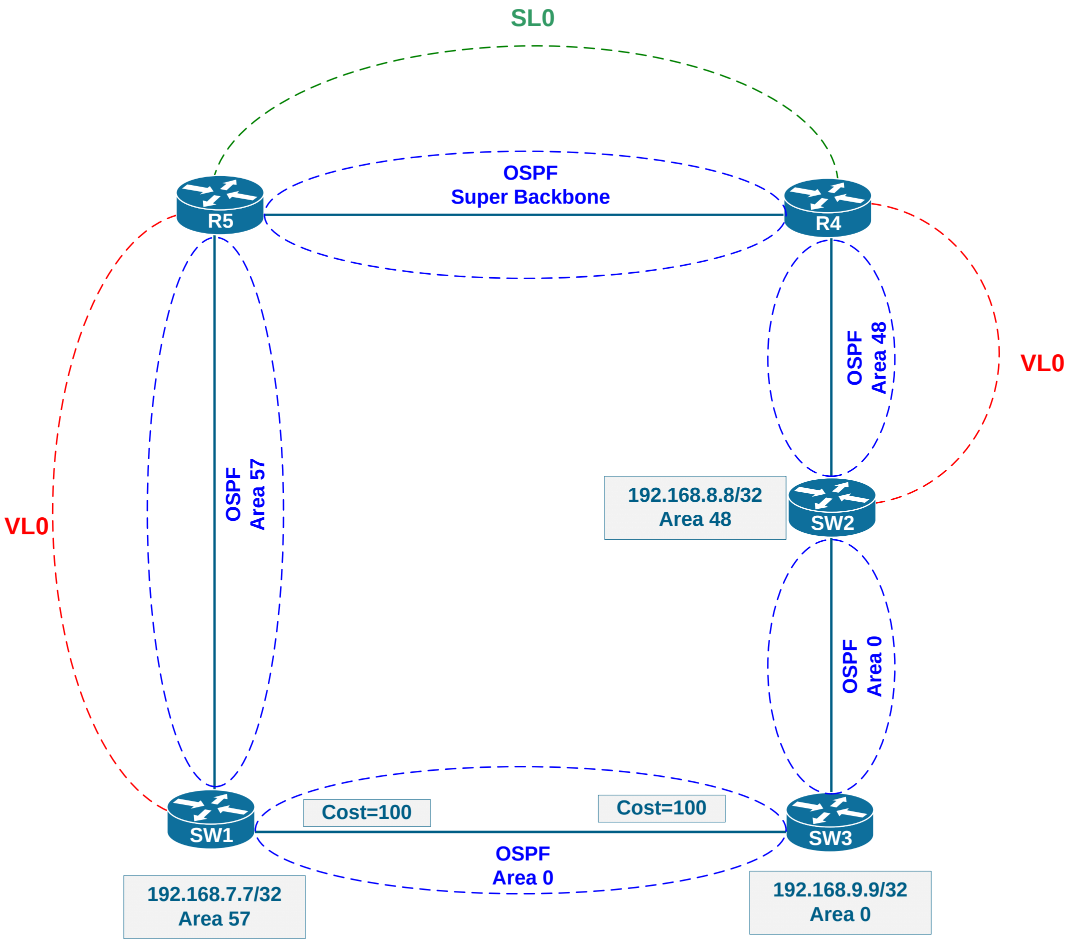
This is usable summary route information that SW1 has injected into Area 0 and that has reached SW2. The metric value here is “1”, so how comes SW2 prefers the first LSA? The reason being, again, the cost to reach the boundary routers:

```
SW2#show ip ospf 100 border-routers

OSPF Process 100 internal Routing Table

Codes: i - Intra-area route, I - Inter-area route

i 192.168.4.4 [1] via 192.168.48.4, FastEthernet0/4, ABR/ASBR, Area 0, SPF 17
i 192.168.4.4 [1] via 192.168.48.4, FastEthernet0/4, ABR/ASBR, Area 48, SPF 13
i 192.168.5.5 [2] via 192.168.48.4, FastEthernet0/4, ABR/ASBR, Area 0, SPF 17
i 192.168.7.7 [101] via 192.168.89.9, FastEthernet0/16, ABR, Area 0, SPF 17
```

```
                Summary Net Link States (Area 48)

  LS age: 873
  Options: (No TOS-capability, DC, Upward)
  LS Type: Summary Links(Network)
  Link State ID: 192.168.7.7 (summary Network Number)
  Advertising Router: 192.168.4.4
  LS Seq Number: 80000001
  Checksum: 0x222F
  Length: 28
  Network Mask: /32
        TOS: 0  Metric: 3
```

This it the summary LSA injected into Area 48 by R4. It is not being used by SW2 as the router has active adjacency in Area 0. However, it may be used for transit route optimization, though that is irrelevant in this scenario.

To complete the scenario, we still need a virtual-link across Area 57, to make it possible for SW1 to accept summary LSA originated at R5.



This allows SW1 reaching the prefix 192.168.8.8/32 via the MP-BGP path, since we increased the backdoor link cost.

```
SW1#show ip ospf virtual-links
Virtual Link OSPF_VL1 to router 192.168.5.5 is up
  Run as demand circuit
  DoNotAge LSA allowed.
  Transit area 57, via interface FastEthernet0/5, Cost of using 1
  Transmit Delay is 1 sec, State POINT_TO_POINT,
  Timer intervals configured, Hello 10, Dead 40, Wait 40, Retransmit 5
    Hello due in 00:00:09
    Adjacency State FULL (Hello suppressed)
    Index 2/3, retransmission queue length 0, number of retransmission
0
    First 0x0(0)/0x0(0) Next 0x0(0)/0x0(0)
    Last retransmission scan length is 0, maximum is 0
    Last retransmission scan time is 0 msec, maximum is 0 msec
```

Check OSPF summary LSAs for 192.168.8.8/32 on SW1 next. First, take note that SW1 has two paths to SW2 via Area 0: the first one via the “true” Area 0

(SW3) and another via the set of virtual and sham-links across the MP-BGP cloud.

```
SW1#show ip ospf database summary 192.168.8.8

            OSPF Router with ID (192.168.7.7) (Process ID 100)

                Summary Net Link States (Area 0)

  Routing Bit Set on this LSA
  LS age: 7 (DoNotAge)
  Options: (No TOS-capability, DC, Upward)
  LS Type: Summary Links(Network)
  Link State ID: 192.168.8.8 (summary Network Number)
  Advertising Router: 192.168.4.4
  LS Seq Number: 80000001
  Checksum: 0x34D
  Length: 28
  Network Mask: /32
        TOS: 0  Metric: 2
```

This is the summary-LSA advertised by R4, which is an ABR connected to Area 48. R4 hears SW2 advertising the prefix 192.168.8.8/32 in router LSAs for Area 48 and constructs corresponding summary LSA, injecting it into Area 0.

```
  Routing Bit Set on this LSA
  LS age: 1206
  Options: (No TOS-capability, DC, Upward)
  LS Type: Summary Links(Network)
  Link State ID: 192.168.8.8 (summary Network Number)
  Advertising Router: 192.168.8.8
  LS Seq Number: 8000002C
  Checksum: 0x6EAF
  Length: 28
  Network Mask: /32
        TOS: 0  Metric: 1
```

This is the summary LSA originated into Area 0 by SW2 itself.

```
                Summary Net Link States (Area 57)

  LS age: 490
  Options: (No TOS-capability, DC, Upward)
  LS Type: Summary Links(Network)
  Link State ID: 192.168.8.8 (summary Network Number)
  Advertising Router: 192.168.5.5
  LS Seq Number: 80000024
  Checksum: 0xB970
  Length: 28
  Network Mask: /32
        TOS: 0  Metric: 3
```

This last LSA is the one received from R5. R5 learns the summary LSA for 192.168.8.8/32 via Area 0 and translates this summary into a new summary injected into the attached area 57

As a result, due to backdoor link cost manipulation previously, SW1 selects the path via R5:

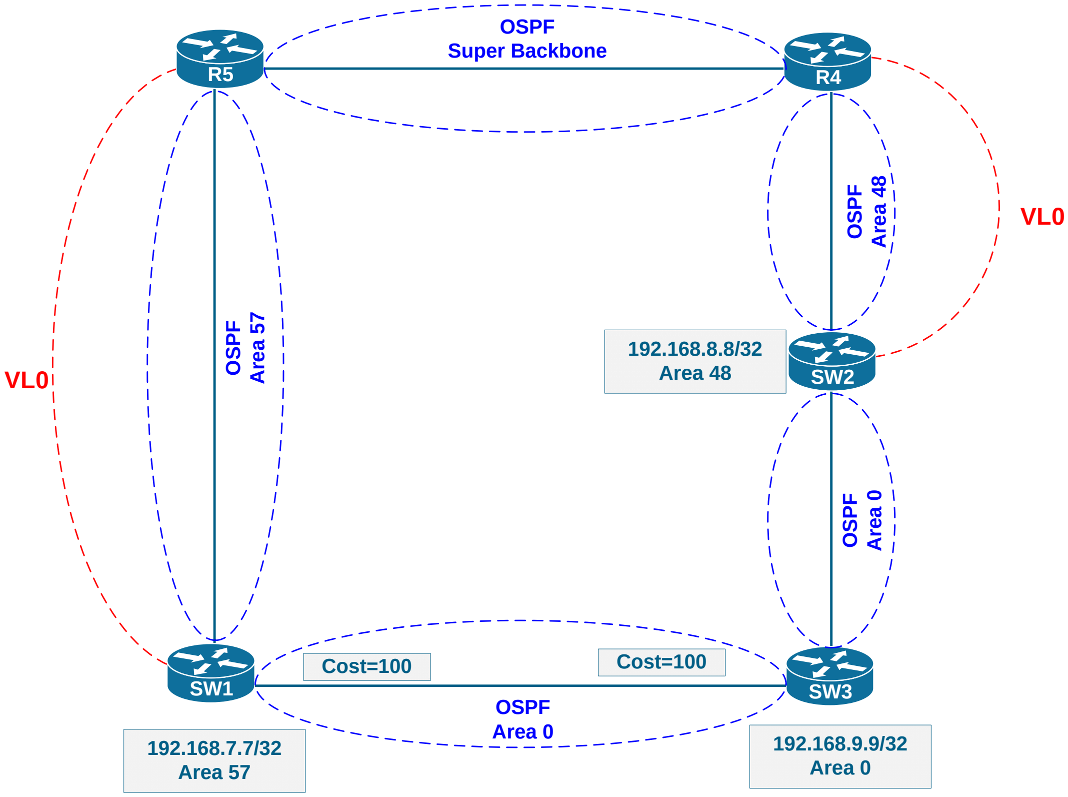
```
SW1#show ip route ospf
O    192.168.89.0/24 [110/4] via 192.168.57.5, 00:00:55, FastEthernet0/5
     192.168.44.0/32 is subnetted, 1 subnets
O E2    192.168.44.44 [110/1] via 192.168.57.5, 00:00:55, FastEthernet0/5
     192.168.8.0/32 is subnetted, 1 subnets
O IA    192.168.8.8 [110/4] via 192.168.57.5, 00:00:55, FastEthernet0/5
     192.168.9.0/32 is subnetted, 1 subnets
O       192.168.9.9 [110/5] via 192.168.57.5, 00:00:55, FastEthernet0/5
     192.168.55.0/32 is subnetted, 1 subnets
O E2    192.168.55.55 [110/1] via 192.168.57.5, 00:00:55, FastEthernet0/5
O IA 192.168.48.0/24 [110/3] via 192.168.57.5, 00:00:55, FastEthernet0/5

SW1#show ip ospf border-routers

OSPF Process 100 internal Routing Table

Codes: i - Intra-area route, I - Inter-area route

i 192.168.4.4 [2] via 192.168.57.5, FastEthernet0/5, ABR/ASBR, Area 0, SPF 19
i 192.168.5.5 [1] via 192.168.57.5, FastEthernet0/5, ABR/ASBR, Area 0, SPF 19
i 192.168.5.5 [1] via 192.168.57.5, FastEthernet0/5, ABR/ASBR, Area 57, SPF 13
i 192.168.8.8 [3] via 192.168.57.5, FastEthernet0/5, ABR, Area 0, SPF 19
```

Interestingly enough, SW1 sees SW2, the ABR, as being reachable over the MP-BGP core. This is due to the cost-manipulation on the backdoor link and OSPF sham-link configured over the MP-BGP core.

To summarize the above, we encountered two problems:

1. Disjoint Area 0 that prevents “true ABRs” such as SW2 from using the summary route transported over the super-backbone. This problem has been fixed using virtual-link to connect the ABR to super-backbone.
2. MP-BGP best-path selection process that prefers locally sourced routes over theoretically better (in OSPF metric) paths learned via iBGP from remote PE. This problem has been resolved by running a sham-link that allows native OSPF LSA exchange and OSPF-only path selection process for summary LSAs.

Q: Is it possible to make MP-BGP selecting non-locally sourced paths without using OSPF sham-link?
A: Indeed, the situation where you have local inter-area routes preferred in MP-BGP looks very similar to the problem we encounter running EIGRP for multi-homed site. It is possible to use BGP Cost community to make BGP prefer the route based on its IGP cost. With EIGRP, the Cost community is automatically generated based on the redistributed route's IGP metric. For OSPF, we need to do this manually. Look at the reference diagram first. There is no sham-link here, but we still need virtual-links to make areas transit.



Here are the initial configuration steps:

1. Configure R5 to attach BGP Cost community of 10 to SW1's prefix 192.168.7.7/32 and BGP Cost community of 100 to the prefix 192.168.8.8/32 when redistributing OSPF into BGP. This correctly reflects the "relative" path costs that R5 has for these prefixes.
2. Symmetrically, configure R4 to attach BGP cost community value of 10 to prefix 192.168.8.8/32 and cost of 100 to prefix 192.168.7.7/32 when redistributing OSPF into BGP.
3. Notice that you cannot use VRF **`export-map`** to accomplish this, this map is only used for setting RT values. You need to attach a route map when redistributing OSPF into BGP.

R4:

```
ip prefix-list NET77 seq 5 permit 192.168.7.7/32
!
ip prefix-list NET88 seq 5 permit 192.168.8.8/32
!
route-map OSPF_TO_BGP permit 10
 match ip address prefix-list NET88
 set extcommunity cost pre-bestpath 128 10
!
route-map OSPF_TO_BGP permit 20
 match ip address prefix-list NET77
 set extcommunity cost pre-bestpath 128 100
!
route-map OSPF_TO_BGP permit 1000
!
router bgp 100
 address-family ipv4 vrf VPN_A
  redistribute ospf 100 vrf VPN_A match internal external 1 external 2
route-map OSPF_TO_BGP
```

R5:

```
ip prefix-list NET77 seq 5 permit 192.168.7.7/32
!
ip prefix-list NET88 seq 5 permit 192.168.8.8/32
!
route-map OSPF_TO_BGP permit 10
 match ip address prefix-list NET88
 set extcommunity cost pre-bestpath 128 100
!
route-map OSPF_TO_BGP permit 20
 match ip address prefix-list NET77
 set extcommunity cost pre-bestpath 128 10
!
route-map OSPF_TO_BGP permit 1000
!
router bgp 100
 address-family ipv4 vrf VPN_A
  redistribute ospf 100 vrf VPN_A match internal external 1 external 2
route-map OSPF_TO_BGP
```

This allows the BGP process to properly select best-paths based on IGP cost values, ignoring normal BGP best-path selection process, because we selected POI value of “pre-bestpath”. The solution is not working, however. Look at R4’s BGP table for prefix 192.168.7.7/32:

```
R4#show bgp vpnv4 unicast all 192.168.7.7
BGP routing table entry for 100:1:192.168.7.7/32, version 160
Paths: (2 available, best #1, table VPN_A, RIB-failure(17))
Flag: 0x820
  Not advertised to any peer
  Local
    10.0.5.5 (metric 65) from 10.0.5.5 (10.0.5.5)
      Origin incomplete, metric 2, localpref 100, valid, internal, best
      Extended Community: RT:100:1 OSPF DOMAIN ID:0x0005:0x000000640200
        Cost:pre-bestpath:128:10 OSPF RT:0.0.0.57:2:0
        OSPF ROUTER ID:192.168.5.5:0
      mpls labels in/out nolabel/21
  Local
    192.168.48.8 from 0.0.0.0 (10.0.4.4)
      Origin incomplete, metric 103, localpref 100, weight 32768,
valid, sourced
      Extended Community: RT:100:1 OSPF DOMAIN ID:0x0005:0x000000640200
        Cost:pre-bestpath:128:100 OSPF RT:0.0.0.0:3:0
        OSPF ROUTER ID:192.168.4.4:0
```

The BGP process has indeed selected the path advertised by R5 over the one sourced locally. However, the BGP route is not making it into RIB table, as the same route already exists there injected by OSPF with better AD:

```
R4#show ip route vrf VPN_A 192.168.7.7
Routing entry for 192.168.7.7/32
  Known via "ospf 100", distance 110, metric 103, type inter area
  Redistributing via bgp 100
  Advertised by bgp 100 match internal external 1 & 2 route-map
OSPF_TO_BGP
  Last update from 192.168.48.8 on FastEthernet0/0, 00:01:57 ago
  Routing Descriptor Blocks:
  * 192.168.48.8, from 192.168.7.7, 00:01:57 ago, via FastEthernet0/0
      Route metric is 103, traffic share count is 1
```

In order to solve this problem, we should change the administrative distance for either OSPF or BGP. It is not possible to change BGP AD for VPNv4 prefixes in Cisco IOS, so we have to resort to changing OSPF distance to a value above the default iBGP distance:

```
R4 & R5:
router ospf 100
 distance 201
```

Now all routers perform path selection just as we wanted originally, e.g. SW2 prefers reaching 192.168.7.7/32 over the MP-BGP core network.. Notice that R4 now has just one BGP path, because the local OSPF route is no longer in RIB, due to AD manipulation.

```
R4#show bgp vpnv4 unicast all 192.168.7.7
BGP routing table entry for 100:1:192.168.7.7/32, version 187
Paths: (1 available, best #1, table VPN_A)
Flag: 0x820
  Not advertised to any peer
  Local
    10.0.5.5 (metric 65) from 10.0.5.5 (10.0.5.5)
      Origin incomplete, metric 2, localpref 100, valid, internal, best
      Extended Community: RT:100:1 OSPF DOMAIN ID:0x0005:0x000000640200
        Cost:pre-bestpath:128:10 OSPF RT:0.0.0.57:2:0
        OSPF ROUTER ID:192.168.5.5:0
      mpls labels in/out nolabel/21

R4#show ip route vrf VPN_A 192.168.7.7
Routing entry for 192.168.7.7/32
  Known via "bgp 100", distance 200, metric 2, type internal
  Redistributing via ospf 100
  Advertised by ospf 100 subnets
                bgp 100 (self originated)
  Last update from 10.0.5.5 00:00:36 ago
  Routing Descriptor Blocks:
  * 10.0.5.5 (Default-IP-Routing-Table), from 10.0.5.5, 00:00:36 ago
      Route metric is 2, traffic share count is 1
      AS Hops 0

SW2#show ip route 192.168.7.7
Routing entry for 192.168.7.7/32
  Known via "ospf 100", distance 110, metric 3, type inter area
  Last update from 192.168.48.4 on FastEthernet0/4, 00:00:51 ago
  Routing Descriptor Blocks:
  * 192.168.48.4, from 192.168.4.4, 00:00:51 ago, via FastEthernet0/4
      Route metric is 3, traffic share count is 1
```

It is important to notice that manipulating BGP Cost community is only helpful when routes are “comparable” in IGP. In our case, this was possible because prefixes for 192.168.7.7/32 and 192.168.8.8/32 were known as OSPF inter-area routes on the PE routers.

Appendix: Initial Configuration for Multi-Area Multi-Homed OSPF Scenario

R4:

```
no ip domain-lookup
ip tcp synwait-time 5
no service timestamps
!
ip vrf VPN_A
 rd 100:1
 route-target both 100:1
!
interface Serial 0/1/0
 no shutdown
 ip address 10.0.45.4 255.255.255.0
 mpls ip
!
interface Loopback 0
 ip address 10.0.4.4 255.255.255.255
!
interface Fa 0/0
 no shut
 ip vrf forwarding VPN_A
 ip address 192.168.48.4 255.255.255.0
!
router ospf 100
 network 10.0.0.0 0.0.255.255 area 0
!
router ospf 10000 vrf VPN_A
 network 192.168.0.0 0.0.255.255 area 48
 redistribute bgp 100 subnets
!
router bgp 100
 no bgp default ipv4-unicast
  neighbor 10.0.5.5 remote-as 100
  neighbor 10.0.5.5 update-source Loopback 0
  address-family vpnv4 unicast
  neighbor 10.0.5.5 activate
  address-family ipv4 vrf VPN_A
   redistribute ospf 100 match internal external
!
line con 0
 privilege level 15
 exec-timeout 0 0
 no login
 loggin synch
```

R5:

```
no ip domain-lookup
ip tcp synwait-time 5
no service timestamps
!
ip vrf VPN_A
 rd 100:1
 route-target both 100:1
!
interface Serial 0/1/0
 no shutdown
 ip address 10.0.45.5 255.255.255.0
 mpls ip
!
interface Loopback 0
 ip address 10.0.5.5 255.255.255.255
!
interface Fa 0/0
 no shut
 ip vrf forwarding VPN_A
 ip address 192.168.57.5 255.255.255.0
!
router ospf 100
 network 10.0.0.0 0.0.255.255 area 0
!
router ospf 10000 vrf VPN_A
 network 192.168.0.0 0.0.255.255 area 57
 redistribute bgp 100 subnets
!
router bgp 100
 no bgp default ipv4-unicast
  neighbor 10.0.4.4 remote-as 100
  neighbor 10.0.4.4 update-source Loopback 0
  address-family vpnv4 unicast
  neighbor 10.0.4.4 activate
  address-family ipv4 vrf VPN_A
   redistribute ospf 100 match internal external
!
line con 0
 privilege level 15
 exec-timeout 0 0
 no login
 loggin synch
```

SW1:

```
hostname SW1
!
no ip domain-lookup
ip tcp synwait-time 5
no service timestamps
!
line con 0
 privilege level 15
 exec-timeout 0 0
 no login
 loggin synch
!
interface Fa 0/5
 no switch
 ip address 192.168.57.7 255.255.255.0
 !
interface Fa 0/16
 no switch
 ip address 192.168.79.7 255.255.255.0
!
interface Loopback 0
 ip address 192.168.7.7 255.255.255.255
  ip ospf 1 area 57
!
router ospf 100
 network 192.168.57.7 0.0.0.0 area 57
 network 192.168.79.7 0.0.0.0 area 0
```

SW2:

```
hostname SW2
!
no ip domain-lookup
ip tcp synwait-time 5
no service timestamps
!
line con 0
 privilege level 15
 exec-timeout 0 0
 no login
 loggin synch
!
interface Fa 0/4
 no switch
 ip address 192.168.48.8 255.255.255.0
 !
interface Fa 0/16
 no switch
 ip address 192.168.89.8 255.255.255.0
!
interface Loopback 0
 ip address 192.168.8.8 255.255.255.255
  ip ospf 1 area 48
!
router ospf 100
 network 192.168.48.8 0.0.0.0 area 48
 network 192.168.89.8 0.0.0.0 area 0
```

SW3:

```
hostname SW1
!
no ip domain-lookup
ip tcp synwait-time 5
no service timestamps
!
line con 0
 privilege level 15
 exec-timeout 0 0
 no login
 loggin synch
!
interface Loopback0
 ip address 192.168.9.9 255.255.255.0
  ip ospf 1 area 0
!
interface Fa 0/16
 no switchport
  ip address 192.168.89.9 255.255.255.0
!
interface FastEthernet 0/13
 no switchport
  ip address 192.168.79.9 255.255.255.0
!
router ospf 100
 network 192.168.89.9 0.0.0.0 area 0
 network 192.168.79.9 0.0.0.0 area 0
```