CCNP DCCORE 350-601

▼ OSPFv2&OSPFv3

NOTE: To control the flooding rate of LSA updates in your network, you can use the LSA group
pacing feature. LSA group pacing can reduce high CPU or buffer usage. This feature groups
LSAs with similar link-state refresh times to allow OSPF to pack multiple LSAs into an OSPF
update message.

LSAs are removed from the link-state database if no LSA update has been received within a
set interval, called the MaxAge. Routers flood a repeat of the LSA every 30 minutes to prevent accurate link-state information from being aged out. The Cisco NX-OS operating system
supports the LSA grouping feature to prevent all LSAs from refreshing at the same time

In OSPFv3, LSA changed by creating a separation between prefixes and the SPF tree. There is no prefix information in LSA types 1 and 2. You find only topology adjacencies in these LSAs; you don't find any IPv6 prefixes in them. Prefixes are now advertised in type 9 LSAs, and the link-local addresses that are used for next hops are advertised in type 8 LSAs. Type 8 LSAs are flooded only on the local link, whereas type 9 LSAs are flooded within the area. The designers of OSPFv3 could have included link-local addresses in type 9 LSAs, but because these are only required on the local link, it would be a waste of resources.

Key Topic

Table 1-2 OSPFv2 and OSPFv3 LSAs Supported by Cisco NX-OS

Type	OSPFv2 Name	Description	OSPFv3 Name	Description
1	Router LSA	LSA sent by every router. This LSA includes the state and the cost of all links and a list of all OSPFv2 neighbors on the link. Router LSAs trigger an SPF recalculation. Router LSAs are flooded to the local OSPFv2 area.	Router LSA	LSA sent by every router. This LSA includes the state and cost of all links but does not include prefix information. Router LSAs trigger an SPF recalculation. Router LSAs are flooded to the local OSPFv3 area.
2	Network LSA	LSA sent by the DR. This LSA lists all routers in the multi-access network. Network LSAs trigger an SPF recalculation.	Network LSA	LSA sent by the DR. This LSA lists all routers in the multi-access network but does not include prefix information. Network LSAs trigger an SPF recalculation.
3	Network Summary LSA	LSA sent by the area border router to an external area for each destination in the local area. This LSA includes the link cost from the area border router to the local destination.	Inter-Area Prefix LSA	Same as OSPFv2; just the name changed.
4	ASBR Summary LSA	LSA sent by the area border router to an external area. This LSA advertises the link cost to the ASBR only.	Inter-Area Router LSA	Same as OSPFv2; just the name changed.
5	AS External LSA	LSA generated by the ASBR. This LSA includes the link cost to an external autonomous system destination. AS External LSAs are flooded throughout the autonomous system.	AS External LSA	Same as OSPFv2.

Type	OSPFv2 Name	Description	OSPFv3 Name	Description
7	NSSA External LSA	LSA generated by the ASBR within a not-so-stubby area (NSSA). This LSA includes the link cost to an external autonomous system destination. NSSA External LSAs are flooded only within the local NSSA.	NSSA External LSA	Same as OSPFv2.
8	N/A		Link LSA (New OSPFv3 LSA)	LSA sent by every router, using a link-local flooding scope. This LSA includes the link-local address and IPv6 prefixes for this link.
9	Opaque LSAs	LSA used to extend OSPF.	Intra-Area Prefix LSA	LSA sent by every router. This LSA includes any prefix or link state changes. Intra-Area Prefix LSAs are flooded to the local OSPFv3 area. This LSA does not trigger an SPF recalculation.
10	Opaque LSAs	LSA used to extend OSPF.	N/A	
11	Opaque LSAs	LSA used to extend OSPF.	Grace LSAs	LSA sent by a restarting router, using a link-local flooding scope. This LSA is used for a graceful restart of OSPFv3.

You can limit the amount of external routing information that floods an area by making it a
stub area. A stub area is an area that does not allow AS External (type 5) LSAs. These LSAs
are usually flooded throughout the local autonomous system to propagate external route
information. Stub areas have the following requirements:

- **All routers in the stub area are stub routers.**
- **No ASBR routers exist in the stub area.**
- **You cannot configure virtual links in the stub area.**

Figure 1-2 shows an example of an OSPF autonomous system where all routers in area 0.0.0.5 have to go through the ABR to reach external autonomous systems. Area 0.0.0.5 can be configured as a stub area.

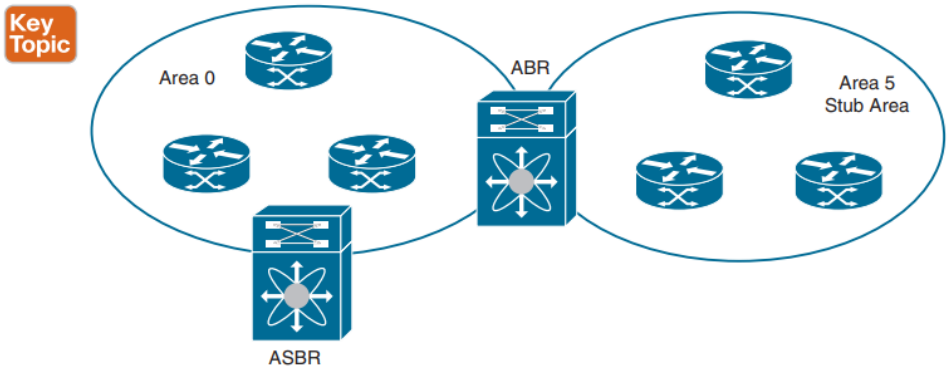


Figure 1-2 *OSPF Stub Area*

Stub areas use a default route for all traffic that needs to go through the backbone area to the external autonomous system. There is an option to allow OSPF to import autonomous system external routes within a stub area; this is a not-so-stubby area (NSSA). An NSSA is similar to a stub area, except that an NSSA allows you to import autonomous system (AS) external routes within an NSSA using redistribution. The NSSA ASBR redistributes these routes and generates NSSA External (type 7) LSAs that it floods throughout the NSSA. You can optionally configure the ABR that connects the NSSA to other areas to translate this NSSA External LSA to AS External (type 5) LSAs. The ABR then floods these AS External LSAs throughout the OSPF autonomous system. Summarization and filtering are supported during the translation.

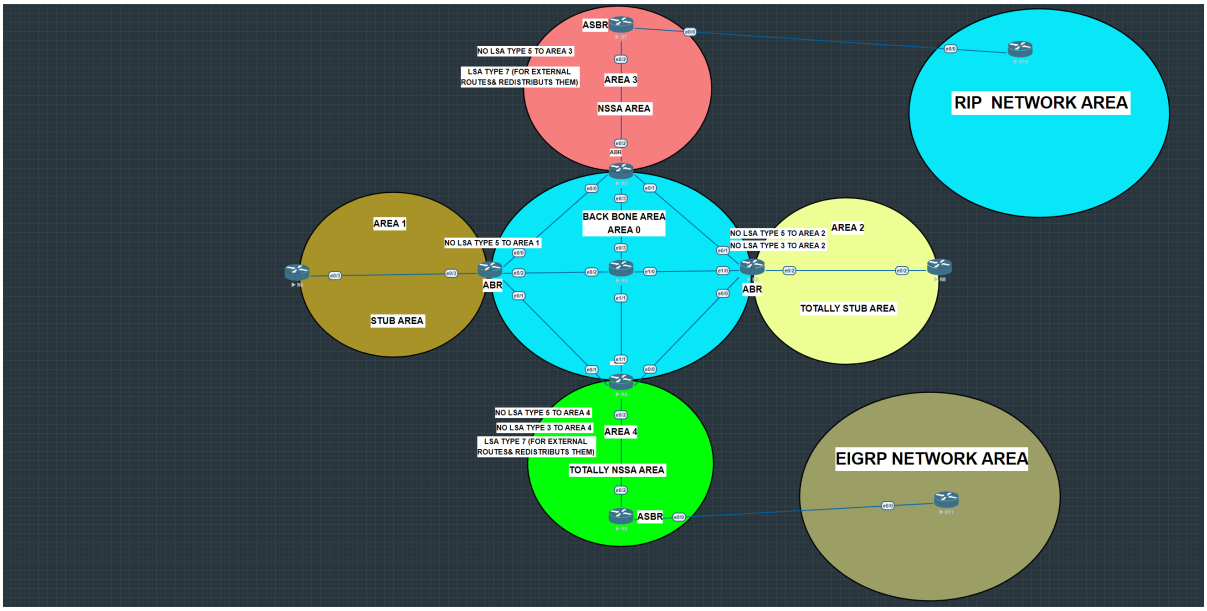


Table 1-4 Feature-Based Licenses for Cisco NX-OS OSPFv2 and OSPFv3

Platform	Feature License	Feature Name
Cisco Nexus 9000 Series Cisco Nexus 7000 Series	Enterprise Services Package LAN_ENTERPRISE_SERVICES_PKG	OSPF OSPFv3
Cisco Nexus 6000 Series Cisco Nexus 5600 Series Cisco Nexus 5500 Series Cisco Nexus 5000 Series	Layer 3 Base Services Package LAN_BASE_SERVICES_PKG	OSPF OSPFv3
Cisco Nexus 3600 Series	Layer 3 Enterprise Services Package LAN_ENTERPRISE_SERVICES_PK	OSPF OSPFv3
Cisco Nexus 3000 Series	Layer 3 Base Services Package LAN_BASE_SERVICES_PK	OSPF (limited routes)

OSPFv2 and OSPFv3 have the following configuration limitations:

- Cisco NX-OS displays areas in dotted-decimal notation regardless of whether you enter the area in decimal or dotted-decimal notation.
- The OSPFv3 router ID and area ID are 32-bit numbers with no relationship to IPv6 addresses.

Tables 1-5 through 1-8 describe the most used OSPFv2/v3 configuration commands. For

Parameters	Default
Hello interval	10 seconds
Dead interval	40 seconds
Graceful restart grace period	60 seconds
OSPFv2/OSPFv3 feature	Disabled
Stub router advertisement announce time	600 seconds
Reference bandwidth for link cost calculation	40 Gbps
LSA minimal arrival time	1000 milliseconds
LSA group pacing	240 seconds
SPF calculation initial delay time	200 milliseconds
SPF calculation maximum wait time	5000 milliseconds
SPF minimum hold time	1000 milliseconds

Cisco NX-OS is a modular system and requires a specific license to enable specific features. Table 1-4 covers the NX-OS feature licenses required for OSPFv2/OSPFv3. For more information, visit the Cisco NX-OS Licensing Guide.

Table 1-5 OSPF Global-Level Commands

Command	Purpose
feature ospf	Enables the OSPFv2 feature.
feature ospfv3	Enables the OSPFv3 feature.
router ospf *ospf-instance-tag*	Creates a new OSPFv2 routing instance.
router ospfv3 *ospf-instance-tag*	Creates a new OSPFv3 routing instance.

Table 1-6 OSPF Routing-Level Commands

Command	Purpose
router-id *ip-address*	(Optional) Configures a unique OSPFv2 or OSPFv3 router ID. *ip-address* must exist on a configured interface in the system.
area *area-id* **authentication [message-digest]**	Configures the authentication mode for an area.
area *area-id* **stub**	Creates this area as a stub area.

Command	Purpose
area *area-id* **nssa [no-redistribution] [default-information-originate]originate [route-map** *map-name*]] **[no-summary] [translate type7 {always | never } [suppress-fa]]**	Creates this area as an NSSA.
address-family ipv6 unicast	Enters IPv6 unicast address family mode.

Table 1-7 OSPF Interface-Level Commands

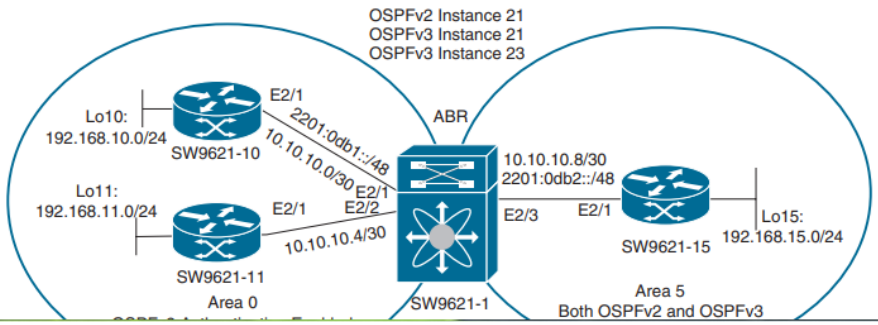
Command	Purpose
ip ospf *cost number*	(Optional) Configures the OSPFv2 cost metric for this interface. The default is to calculate the cost metric, based on reference bandwidth and interface bandwidth. The range is from 1 to 65,535.
ip ospf dead-interval *seconds*	(Optional) Configures the OSPFv2 dead interval in seconds. The range is from 1 to 65,535. The default is four times the hello interval in seconds.
ip ospf hello-interval *seconds*	(Optional) Configures the OSPFv2 hello interval in seconds. The range is from 1 to 65,535. The default is 10 seconds.
ip ospf mtu-ignore	(Optional) Configures OSPFv2 to ignore any IP maximum transmission unit (MTU) mismatch with a neighbor. The default is not to establish adjacency if the neighbor MTU does not match the local interface MTU.
ospf network {**broadcast** | **point-point** }	(Optional) Sets the OSPFv2 network type.
[**default** | **no**] **ip ospf passive-interface**	(Optional) Suppresses routing updates on the interface. This command overrides the router or VRF command mode configuration. The default option removes this interface mode command and reverts to the router or VRF configuration if present.
ip ospf priority *number*	(Optional) Configures the OSPFv2 priority used to determine the DR for an area. The range is from 0 to 255. The default is 1.
ip ospf shutdown	(Optional) Shuts down the OSPFv2 instance on this interface.
ip ospf message-digest-key *key-id* **md5** [**0** | **3**] *key*	Configures message digest authentication for this interface. Use this command if the authentication is set to message-digest. The *key-id* range is from 1 to 255. The MD5 option 0 configures the password in clear text and 3 configures the pass key as 3DES encrypted.

Table 1-8 OSPF Global-Level Verification and Process Clear Commands

Command	Purpose
show ip ospf [*instance-tag*] [**vrf** *vrf-name*]	Displays the OSPFv2 configuration.
show ip ospf interface [*instance-tag*] [*interface-type interface-number*] [**brief**] [**vrf** *vrf-name*]	Displays the OSPFv2 interface configuration.
show ip ospf route [*ospf-route*] [**summary**] [**vrf** { *vrf-name* | **all** | **default** | **management** }]	Displays the internal OSPFv2 routes.

Command	Purpose
show ip ospf virtual-links [**brief**] [**vrf** { *vrf-name* | **all** | **default** | **management** }]	Displays information about OSPFv2 virtual links.
show running-configuration ospf	Displays the current running OSPFv2 configuration.
show ip ospf statistics [**vrf** { *vrf-name* | **all** | **default** | **management** }]	Displays the OSPFv2 event counters.
show ip ospf traffic [*interface - type number*] [**vrf** { *vrf-name* | **all** | **default** | **management** }]	Displays the OSPFv2 packet counters.
clear ip ospf [**instance-tag**] **neighbor** {* | *neighbor-id* | *interface-type number* | *loopback number* | *port-channel number*} [**vrf** *vrf-name*]	Clears neighbor statistics and resets adjacencies for Open Shortest Path First (OSPFv2). **NOTE:** Clearing the OSPF **neighbor** command will reload the OSPF process, so take extra precaution before executing the command in a production environment.
show [**ipv6**] **ospfv3** [*instance-tag*] [**vrf** *vrf-name*]	Displays the OSPFv3 configuration.
show [**ipv6**] **ospfv3 interface** [*instance-tag*] [*interface-type interface-number*] [**brief**] [**vrf** *vrf-name*]	Displays the OSPFv3 interface configuration.
clear ospfv3 [*instance-tag*] **neighbor** {* | *neighbor-id* | **interface-type** *number* | **loopback number** | **port-channel** *number*} [**vrf** *vrf-name*]	Clears neighbor statistics and resets adjacencies for Open Shortest Path First (OSPFv3). **NOTE:** Clearing the OSPF **neighbor** command will reload the OSPF process, so take extra precaution before executing the command in a production environment.

Figure 1-4 shows the network topology for the configuration that follows, which demonstrates how to configure Nexus OSPF for IPv4 and IPv6.



▼ BGP_BORDER GATEWAY PROTOCOL

The Border Gateway Protocol (BGP) uses a path-vector routing algorithm to exchange

routing information between BGP speakers. Based on this information, each BGP speaker

determines a path to reach a particular destination while detecting and avoiding paths with
routing loops. The routing information includes the actual route prefix for a destination, the
path of autonomous systems to the destination, and additional path attributes

BGP selects a single path, by default, as the best path to a destination host or network. Each
path carries well-known mandatory, well-known discretionary, and optional transitive attributes that are used in BGP best-path analysis. You can influence BGP path selection by
altering some of these attributes by configuring BGP policies.

BGP also supports load balancing or equal-cost multipath (ECMP), where next-hop packet
forwarding to a single destination can occur over multiple “best paths” that tie for top place
in routing metric calculations. It potentially offers substantial increases in bandwidth by
load-balancing traffic over multiple paths.

Cisco NX-OS supports BGP version 4, which includes multiprotocol extensions that allow
BGP to carry routing information for IP multicast routes and multiple Layer 3 protocol address families. BGP uses TCP (Port 179) as a reliable transport protocol to create TCP
sessions with other BGP-enabled devices.

The BGP autonomous system (AS) is a network controlled by a single administration entity.
An autonomous system forms a routing domain with one or more Interior Gateway Protocols
(IGPs) and a consistent set of routing policies. BGP supports 16-bit and 32-bit autonomous
system numbers.

External BGP autonomous systems dynamically exchange routing information through
external BGP (eBGP) peering sessions. BGP speakers within the same autonomous

system
can exchange routing information through internal BGP (iBGP) peering sessions

▼ BGP PEERING

A BGP speaker does not discover and peer with another BGP speaker automatically. You
must configure the relationships between BGP speakers. A BGP peer is a BGP speaker that
has an active TCP connection to another BGP speaker.

BGP uses TCP port 179 to create a TCP session with a peer. When a TCP connection is
established between peers, each BGP peer initially exchanges all of its routes—the complete
BGP routing table—with the other peer. After this initial exchange, the BGP peers send
only incremental updates when a topology change occurs in the network or when a routing
policy change occurs. In the periods of inactivity between these updates, peers exchange
special messages called keepalives. The hold time is the maximum time limit that can elapse
between receiving consecutive BGP update or keepalive messages. Cisco NX-OS supports
the following peer configuration options:

■ Individual IPv4 or IPv4 address: BGP establishes a session with the BGP speaker that
matches the remote address and AS number.

■ IPv4 or IPv6 prefix peers for a single AS number: BGP establishes sessions with
BGP speakers that match the prefix and the AS number.

■ Dynamic AS number prefix peers: BGP establishes sessions with BGP speakers that
match the prefix and an AS number from a list of configured AS numbers

NOTE: The dynamic AS number prefix peer configuration overrides the individual AS number configuration that is inherited from a BGP template.

NOTE : BGP MUST HAVE A ROUTER_ID CONFIGURED MANUALLY OR CHOSEN BY AN ALGORITHM

▼ BGP PATH SELECTION

The best-path algorithm runs each time a path is added or withdrawn for a given network.

The best-path algorithm also runs if you change the BGP configuration. BGP selects the best
path from the set of valid paths available for a given network.

Cisco NX-OS implements the BGP best-path algorithm in the following steps.
Step 1: Comparing Pairs of Paths
This first step in the BGP best-path algorithm compares two paths to determine which path
is better. The following sequence describes the basic steps that Cisco NX-OS uses to compare two paths to determine the better path:

1. Cisco NX-OS chooses a valid path for comparison. (For example, a path that has an
   unreachable next-hop is not valid.)
2. Cisco NX-OS chooses the path with the highest weight.
3. Cisco NX-OS chooses the path with the highest local preference.
4. If one of the paths is locally originated, Cisco NX-OS chooses that path.
5. Cisco NX-OS chooses the path with the shorter AS-path.
   NOTE: **When calculating the length of the AS-path, Cisco NX-OS ignores confederation segments and counts AS sets as 1.**
6. Cisco NX-OS chooses the path with the lower origin. Interior Gateway Protocol (IGP)
   is considered lower than EGP.
   1
   ||||||||||||||||||||||||||||||||||||||||
   26 CCNP and CCIE Data Center Core DCCOR 350-601 Official Cert Guide

7. Cisco NX-OS chooses the path with the lower multi-exit discriminator (MED). You can configure a number of options that affect whether this step is performed. In general, Cisco NX-OS compares the MED of both paths if the paths were received from peers in the same autonomous system; otherwise, Cisco NX-OS skips the MED comparison.
   You can configure Cisco NX-OS to always perform the best-path algorithm MED comparison, regardless of the peer autonomous system in the paths. Otherwise, Cisco NX-OS will perform a MED comparison that depends on the AS-path attributes of the two paths being compared:
   a. If a path has no AS-path or the AS-path starts with an AS_SET, the path is internal, and Cisco NX-OS compares the MED to other internal paths.
   b. If the AS-path starts with an AS_SEQUENCE, the peer autonomous system is the first AS number in the sequence, and Cisco NX-OS compares the MED to other paths that have the same peer autonomous system.
   c. If the AS-path contains only confederation segments or starts with confederation segments followed by an AS_SET, the path is internal and Cisco NX-OS compares the MED to other internal paths.
   d. If the AS-path starts with confederation segments followed by an AS_SEQUENCE, the peer autonomous system is the first AS number in the AS_SEQUENCE, and Cisco NX-OS compares the MED to other paths that have the same peer autonomous system.
   NOTE: If Cisco NX-OS receives no MED attribute with the path, Cisco NX-OS considers the MED to be 0 unless you configure the best-path algorithm to set a missing MED to the highest possible value.

e. If the nondeterministic MED comparison feature is enabled, the best-path algorithm uses the Cisco IOS style of MED comparison.

8. If one path is from an internal peer and the other path is from an external peer, Cisco
   NX-OS chooses the path from the external peer.
9. If the paths have different IGP metrics to their next-hop addresses, Cisco NX-OS
   chooses the path with the lower IGP metric.
10. Cisco NX-OS uses the path that was selected by the best-path algorithm the last time
    that it was run.
    If all path parameters in step 1 through step 9 are the same, you can configure the
    best-path algorithm to compare the router IDs. If the path includes an originator
    attribute, Cisco NX-OS uses that attribute as the router ID to compare to; otherwise,
    Cisco NX-OS uses the router ID of the peer that sent the path. If the paths have
    different router IDs, Cisco NX-OS chooses the path with the lower router ID.
    NOTE When the attribute originator is used as the router ID, it is possible that two paths
    have the same router ID. It is also possible to have two BGP sessions with the same peer
    router, and therefore, you can receive two paths with the same router ID.
    ||||||||||||||||||||||||||||||||||||||||
    www.CareerCert.info
    Chapter 1: Implementing Routing in the Data Center 27
11. Cisco NX-OS selects the path with the shorter cluster length. If a path was not
    received with a cluster list attribute, the cluster length is 0.
12. Cisco NX-OS chooses the path received from the peer with the lower IP address.
    Locally generated paths (for example, redistributed paths) have a peer IP

address of 0.
NOTE Paths that are equal after step 9 can be used for multipath if you configure it.
Step 2: Determining the Order of Comparisons
The second step of the BGP best-path algorithm implementation is to determine the order in
which Cisco NX-OS compares the paths:

13. Cisco NX-OS partitions the paths into groups. Within each group, Cisco NX-OS
compares the MED among all paths. Cisco NX-OS uses the same rules as in step 1
to determine whether MED can be compared between any two paths. Typically, this
comparison results in one group being chosen for each neighbor autonomous system.
If you configure the bgp bestpath med always command, Cisco NX-OS chooses just
one group that contains all the paths.

14. Cisco NX-OS determines the best path in each group by iterating through all paths in
the group and keeping track of the best one so far. Cisco NX-OS compares each path
with the temporary best path found so far, and if the new path is better, it becomes
the new temporary best path, and Cisco NX-OS compares it with the next path in the
group.

15. Cisco NX-OS forms a set of paths that contain the best path selected from each group
in step 2. Cisco NX-OS selects the overall best path from this set of paths by going
through them as in step 2.
Step 3: Determining the Best-Path Change Suppression
The next part of the implementation is to determine whether Cisco NX-OS

will use the new
best path or suppress it. The router can continue to use the existing best path if the new one
is identical to the old path (if the router ID is the same). Cisco NX-OS continues to use the
existing best path to avoid route changes in the network.
You can turn off the suppression feature by configuring the best-path algorithm to compare
the router IDs. If you configure this feature, the new best path is always preferred to the
existing one.
You cannot suppress the best-path change if any of the following conditions occur:
■ The existing best path is no longer valid.
■ Either the existing or new best paths were received from internal (or confederation)
peers or were locally generated (for example, by redistribution).
■ The paths were received from the same peer (the paths have the same router ID).
■ The paths have different weights, local preferences, origins, or IGP metrics to their
next-hop addresses.
■ The paths have different MEDs.

16. NOTE: PATH VECTOR TELLS ALL THE ROUTES THAT THE PACKET IS GOING TO USE IN ORDER TO REACH ITS DESTINATION (DISPLAYS THE AS NUMBERS AND THE NUMBER OF ROUTES)

17. **NOTE: Prefer the path that was locally originated via a `network` or `aggregate` BGP subcommand or through redistribution from an IGP.**

Local paths that are sourced by the `network` or `redistribute` commands are preferred over local aggregates that are sourced by the `aggregate-address` command.

Weight
Local Preference
Originate
AS Path Length
Origin Type
Multi-Exit Discriminator (MED)
Paths
Router ID

We Love Oranges AS Oranges Mean Pure Refreshment

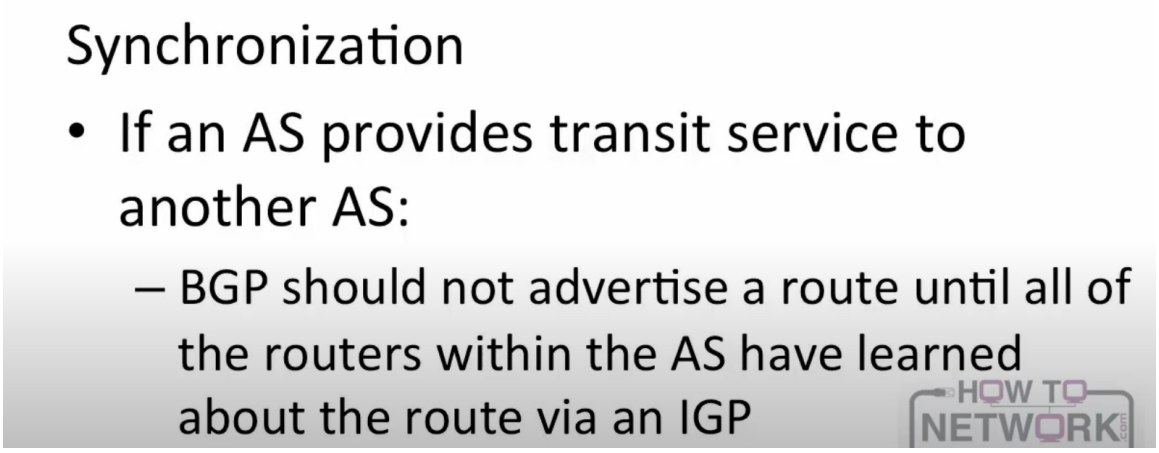
Path Selection Parameter	Description
Weight	A locally significant, Cisco-specific parameter that a router can set when receiving updates. A higher Weight is preferred. Commonly used to influence outbound routing decisions.
Local Preference	A parameter communicated throughout a single AS. A higher Local Preference is preferred. Commonly used to influence outbound routing decisions.

BGP SUMMERIZATION

NOTE1:to advertize a summerized address we use the command aggregated-address (the summerized address ,needs to be subbneted),to costomize the summerization by just advertizing the summerizeed address we add the command sommery-only

NOTE2:to advertise to the local network that this router is the way out to EBGP we use the commnad neighbor (the lan bgp neighbor) next-hop-self

NOTE: SYNCHRONIZATION IS DISABLED BY DEFAULT



IMPORTANT:for MED(multi-EXIT DESCRIMINATOR specification)

Multiple Exit Discriminator (MED)

The MUTLI_EXIT_DISC (MED) is an optional non-transitive attribute that provides a mechanism for the network administrator to convey to adjacent autonomous systems to optimal entry point in the local AS; Figure 1-11 illustrates this concept.

Figure
1-11: The Multiple Exit Discriminator

Here, AS 65200 is setting the MED on its T1 exit point to 100, and the MED on its OC3 exit point to 50, with the intended result that the OC3 connection be preferred. However, the problem with using the MED in this way becomes apparent with this simple example. First, AS 65100 will receive three paths to 10.1.1.0/24, one through AS 65300, and two through AS 65200. The MED of the path through AS 65100 and the paths through AS 65200 will not be compared, since their AS Path is not the same. If AS 65100 has set their BGP local preferences on router A, B, and C, to favor the path through AS 65300, then the MED from AS 65200 will have no impact, per MED is considered after local preference in the BGP decision algorithm.

NOTE

MEDs received from different autonomous systems are not compared as a default behavior, though many implementations provide a mechanism to enable comparing of MEDs between different autonomous systems. Benefits and offshoots of using MEDs, and comparing them between different ASes, will be discussed in later sections.

If the path through AS 65300 did not exist, or was not preferred over the path through AS 65200 for some other reason, the MEDs advertised by routers D and E might have some impact on the best path decision made by AS 65100. However, if AS 65100 sets some BGP metric with a higher degree of preference in the decision algorithm, such as the local preference, to prefer one path over the other, the MED would never be considered.

NOTE: Prefer the path that was locally originated via a `network` or `aggregate` BGP subcommand or through redistribution from an IGP.

Local paths that are sourced by the `network` or `redistribute` commands are preferred over local aggregates that are sourced by the `aggregate-address` command.

7. Cisco NX-OS chooses the path with the lower multi-exit discriminator (MED).

   You can configure a number of options that affect whether this step is performed. In general, Cisco NX-OS compares the MED of both paths if the paths were received from peers in the same autonomous system; otherwise, Cisco NX-OS skips the MED comparison.

   You can configure Cisco NX-OS to always perform the best-path algorithm MED comparison, regardless of the peer autonomous system in the paths. Otherwise, Cisco NX-OS will perform a MED comparison that depends on the AS-path attributes of the two paths being compared:

   a. If a path has no AS-path or the AS-path starts with an AS_SET, the path is internal, and Cisco NX-OS compares the MED to other internal paths.

   b. If the AS-path starts with an AS_SEQUENCE, the peer autonomous system is the first AS number in the sequence, and Cisco NX-OS compares the MED to other paths that have the same peer autonomous system.

   c. If the AS-path contains only confederation segments or starts with confederation segments followed by an AS_SET, the path is internal and Cisco NX-OS compares the MED to other internal paths.

   d. If the AS-path starts with confederation segments followed by an AS_SEQUENCE, the peer autonomous system is the first AS number in the AS_SEQUENCE, and Cisco NX-OS compares the MED to other paths that have the same peer autonomous system.

NOTE If Cisco NX-OS receives no MED attribute with the path, Cisco NX-OS considers the MED to be 0 unless you configure the best-path algorithm to set a missing MED to the highest possible value.

   e. If the nondeterministic MED comparison feature is enabled, the best-path algorithm uses the Cisco IOS style of MED comparison.

BGP Neighbor States

State	Typical Reasons
Idle	The BGP process is either administratively down or awaiting the next retry attempt.
Connect	The BGP process is waiting for the TCP connection to be completed.
Active	The TCP connection has been completed, but no BGP messages have been sent to the peer yet.
Opensent	The TCP connection exists, and a BGP Open message has been sent to the peer.
Openconfirm	An Open message has been sent to and received from the other router.
Established	All neighbor parameters match and peers can send Update messages.

BGP Clear Commands

Command	Hard or Soft	Neighbors	Direction
clear ip bgp *	Hard	all	both
clear ip bgp neighbor-id	Hard	one	both
clear ip bgp neighbor-id soft out	Soft	one	out
clear ip bgp neighbor-id soft in	Soft	one	in
clear ip bgp * soft	Soft	all	both
clear ip bgp neighbor-id soft	Soft	one	both

▼ Determining the order of comparision

The second step of the BGP best-path algorithm implementation is to determine the order in
which Cisco NX-OS compares the paths:

1. **Cisco NX-OS partitions the paths into groups**. **Within each group, Cisco NX-OS**

compares the MED among all paths. Cisco NX-OS uses the same rules as in step 1
to determine whether MED can be compared between any two paths. Typically, this
comparison results in one group being chosen for each neighbor autonomous system.
If you configure the bgp bestpath med always command, Cisco NX-OS chooses just
one group that contains all the paths.

2. Cisco NX-OS determines the best path in each group by iterating through all paths in
the group and keeping track of the best one so far. Cisco NX-OS compares each path
with the temporary best path found so far, and if the new path is better, it becomes
the new temporary best path, and Cisco NX-OS compares it with the next path in the
group.
3. Cisco NX-OS forms a set of paths that contain the best path selected from each group
in step 2. Cisco NX-OS selects the overall best path from this set of paths by going
through them as in step 2.

▼ Determining the best-path change suppression

The next part of the implementation is to determine whether Cisco NX-OS will use the new
best path or suppress it. The router can continue to use the existing best path if the new one
is identical to the old path (if the router ID is the same). Cisco NX-OS continues to use the
existing best path to avoid route changes in the network.
You can turn off the suppression feature by configuring the best-path algorithm to compare

the router IDs. **If you configure this feature, the new best path is always preferred to the**
existing one.
You cannot suppress the best-path change if any of the following conditions occur:

- **The existing best path is no longer valid.**
- **Either the existing or new best paths were received from internal (or confederation)**
peers or were locally generated (for example, by redistribution).
- **The paths were received from the same peer (the paths have the same router ID).**
- **The paths have different weights, local preferences, origins, or IGP metrics to their**
next-hop addresses.
- **The paths have different MEDs.**

NOTE :The order of comparison determined in Step 2 is important. Consider the case
where you have three paths—A, B, and C. When Cisco NX-OS compares A and B, it
chooses A. When Cisco NX-OS compares B and C, it chooses B. But when Cisco NX-OS
compares A and C, it might not choose A because some BGP metrics apply only among
paths from the same neighboring autonomous system and not among all paths.

The path selection uses the BGP AS-path attribute. The AS-path attribute includes the
list of autonomous system numbers (AS numbers) traversed in the advertised path. If you
subdivide your BGP autonomous system into a collection or confederation of autonomous
systems, the AS-path contains confederation segments that list these locally defined autonomous systems.

▼ MULTIPROTOCOL BGP

NOTE: **Reverse-path forwarding (RPF) is a technique used in modern routers for the purposes of ensuring loop-free forwarding of multicast packets in multicast routing and to help prevent IP address spoofing in unicast routing.**

NOTE Because Multicast BGP does not propagate multicast state information, you need a
multicast protocol, such as Protocol Independent Multicast (PIM).

**You need to use the router address-family and neighbor address-family configuration modes
to support Multiprotocol BGP configurations. MBGP maintains separate Routing Information Bases (RIBs) for each configured address family, such as a unicast RIB and a multicast
RIB for BGP.**

**A Multiprotocol BGP network is backward compatible, but BGP peers that do not support
multiprotocol extensions cannot forward routing information, such as address family identifier information, that the multiprotocol extensions carry.**

Table 1-9 Default BGP Parameters

Parameters	Default
BGP feature	Disabled
Keepalive interval	60 seconds
Hold timer	180 seconds
BGP PIC core	Enabled
Auto-summary	Always disabled
Synchronization	Always disabled

Table 1-10 Feature-Based Licenses for Cisco NX-OS

Platform	Feature License	Feature Name
Cisco Nexus 9000 Series Cisco Nexus 7000 Series	Enterprise Services Package LAN_ENTERPRISE_SERVICES_PKG	BGP
Cisco Nexus 6000 Series Cisco Nexus 5600 Series Cisco Nexus 5500 Series Cisco Nexus 5000 Series	Layer 3 Enterprise Services Package LAN_ENTERPRISE_SERVICES_PKG	BGP
Cisco Nexus 3600 Series	Layer 3 Enterprise Services Package LAN_ENTERPRISE_SERVICES_PKG	BGP
Cisco Nexus 3000 Series	Layer 3 Enterprise Services Package LAN_ENTERPRISE_SERVICES_PKG	BGP

BGP has the following configuration limitations:

- The dynamic AS number prefix peer configuration overrides the individual AS number configuration inherited from a BGP template.
- If you configure a dynamic AS number for prefix peers in an AS confederation, BGP establishes sessions with only the AS numbers in the local confederation.
- BGP sessions created through a dynamic AS number prefix peer ignore any configured eBGP multihop time-to-live (TTL) value or a disabled check for directly connected peers.
- Configure a router ID for BGP to avoid automatic router ID changes and session flaps.
- Use the maximum-prefix configuration option per peer to restrict the number of routes received and system resources used.
- Configure the update source to establish a session with BGP/eBGP multihop sessions.
- Specify a BGP policy if you configure redistribution.
- Define the BGP router ID within a VRF.
- If you decrease the keepalive and hold timer values, you might experience BGP session flaps.
- The BGP minimum route advertisement interval (MRAI) value for all iBGP and eBGP sessions is zero and is not configurable.

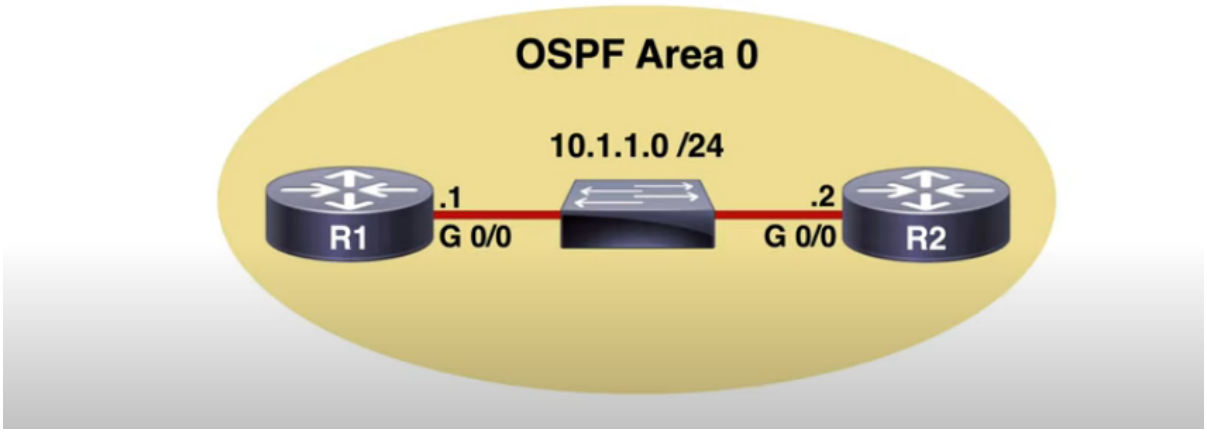
▼ bgp verificaiton commands

Command	Purpose
show bgp all [**summary**] [**vrf** *vrf-name*]	Displays the BGP information for all address families.
show {**ipv** | **ipv6**} **bgp** *options*	Displays the BGP status and configuration information. This command has multiple options. One important option is summary (**show ip bgp summary**).
show bgp convergence [**vrf** *vrf-name*]	Displays the BGP information for all address families.
show bgp {**ip** | **ipv6**} {**unicast** | **multicast**} [*ip-address* | *ipv6-prefix*] **community** {**regexp expression** | [**community**] [**no-advertise**] [**no-export**] [**no-export-subconfed**]} [**vrf** *vrf-name*]	Displays the BGP routes that match a BGP community.
show bgp *process*	Displays the BGP process information.
show running-configuration bgp	Displays the current running BGP configuration.
show bgp sessions [**vrf** *vrf-name*]	Displays the BGP sessions for all peers. You can use the **clear bgp sessions** command to clear these statistics.
show bgp statistics	Shows BGP statistics.
clear bgp all { **neighbor** | * | *as-number* | **peer-template** *name* | *prefix* } [**vrf** *vrf-name*]	Clears one or more neighbors from all address families. * clears all neighbors in all address families. The arguments are as follows: *neighbor*: IPv4 or IPv6 address of a neighbor. *as-number*: Autonomous system number. The AS number can be a 16-bit integer or a 32-bit integer in the form of higher 16-bit decimal number and a lower 16-bit decimal number in xx.xx format. *name*: Peer template name. The name can be any case-sensitive, alphanumeric string up to 64 characters. *prefix*: IPv4 or IPv6 prefix. All neighbors within that prefix are cleared. *vrf-name*: VRF name. All neighbors in that VRF are cleared. The name can be any case-sensitive, alphanumeric string up to 64 characters.
clear bgp all dampening [**vrf** *vrf-name*]	Clears route flap dampening networks in all address families. The *vrf-name* can be any case-sensitive, alphanumeric string up to 64 characters.
clear bgp all flap-statistics [**vrf** *vrf-name*]	Clears route flap statistics in all address families. The *vrf-name* can be any case-sensitive, alphanumeric string up to 64 characters.

▼ Biderctional Forwarding Detection

use cases:

Bidirectional Forwarding Detection



issue:

let's say that the phisical interface between R2 and the switch goes down what happens then .R2 will tear down the OSPF adjacency based on that link failure , however as far as R1 is concerned at least for some time the link between itself and the switch is still functioning ,so it has no awarness about the link failure event between R2 and the switch .R1 is going to continue to send traffic in a situation that we call black holing ,so the traffic goes over to the L2 switch even though it cannot reach the destinaiton of R2 and it will do that untill our configured protocol timers expiers and let R1 know that the OSPF neighbor is down(we do ofc have some options for lowering thows ospf timers to help detect failure much quicker but no where near as quicky as BFD).

Bidirectional Forwarding Detection (BFD) is a detection protocol designed to provide fast forwarding–path failure detection times **for media types, encapsulations, topologies, and routing protocols**. BFD provides subsecond failure detection between two adjacent devices and can be less CPU-intensive than protocol hello messages because some of the BFD load can be distributed onto the data plane on supported modules.

Cisco NX-OS supports the BFD asynchronous mode, which sends BFD control packets
between two adjacent devices to activate and maintain BFD neighbor sessions

between the
devices.You configure BFD on both devices (or BFD neighbors).once BFD has been enabled on the interfaces and on the appropiate protovold,Cisco NX-OS creates a BFD session, negotiates BFD session parameters , and begins to send BFD control packets to each BFD neighbor at the negotiated inteval. the BFD session parameters include the following:

Desired minimum transmit interval: the interval at which this device wants to send BFD hello messages .

Required minimum receive interval : the minimum interval at which this device can accept BDF hello messages from ano therBGF deveice .

Detect multiplier: the number of missing BFD hello messages from another BFD device before this local device detect a fault in the forwarding path .

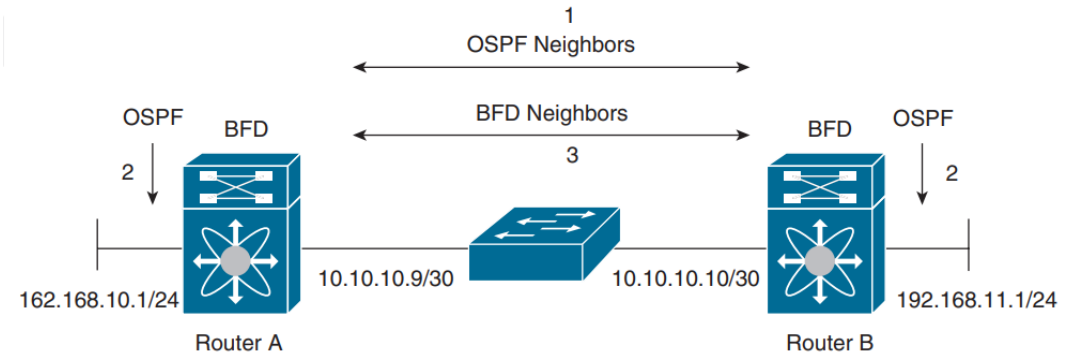


Figure 1-6 *Establishing a BFD Neighbor Relationship*

Rapid Detection of Failiures:

After a BFD session has been established and timer negotiations are comlete, BFD neighbors send BFD control packets that act in the same manner as an IGP hello protocol to detect liveliness , except at a more accelerated rate. BFD detects a failure , but the protocol must take action to bypass a failed peer

NOTE:BFD is a light weight udp protocol used for fast forwarding failure detection

it is a routing protocol independent(can be used with any kind of routing protocol)

NOTE2: BFD is a way more efficient than using the built in failure detection mechanisms , BDF offers failure detection in the ms range wich is very short range

Table 1-14 Default Settings for BFD Parameters

Parameters	Default
BFD feature	Disabled
Required minimum receive interval	50 milliseconds
Desired minimum transmit interval	50 milliseconds
Detect multiplier	3
Echo function	Enabled
Mode	Asynchronous
Port channel	Logical mode (one session per source-destination pair address)
Slow timer	2000 milliseconds
Subinterface optimization	Disabled

BFD has the following configuration limitations:

■ NX-OS supports BFD version 1.

■ NX-OS supports IPv4 only.

■ BFD supports single-hop BFD; BFD for BGP supports single-hop EBGP and iBGP peers.

■ BFD depends on Layer 3 adjacency information to discover topology changes, including Layer 2 topology changes. A BFD session on a VLAN interface (SVI) may not be
up after the convergence of the Layer 2 topology if no Layer 3 adjacency information is available.

■ For port channels used by BFD, you must enable the Link Aggregation Control Protocol (LACP) on the port channel.

■ HSRP for IPv4 is supported with BFD. HSRP for IPv6 is not supported with BFD.

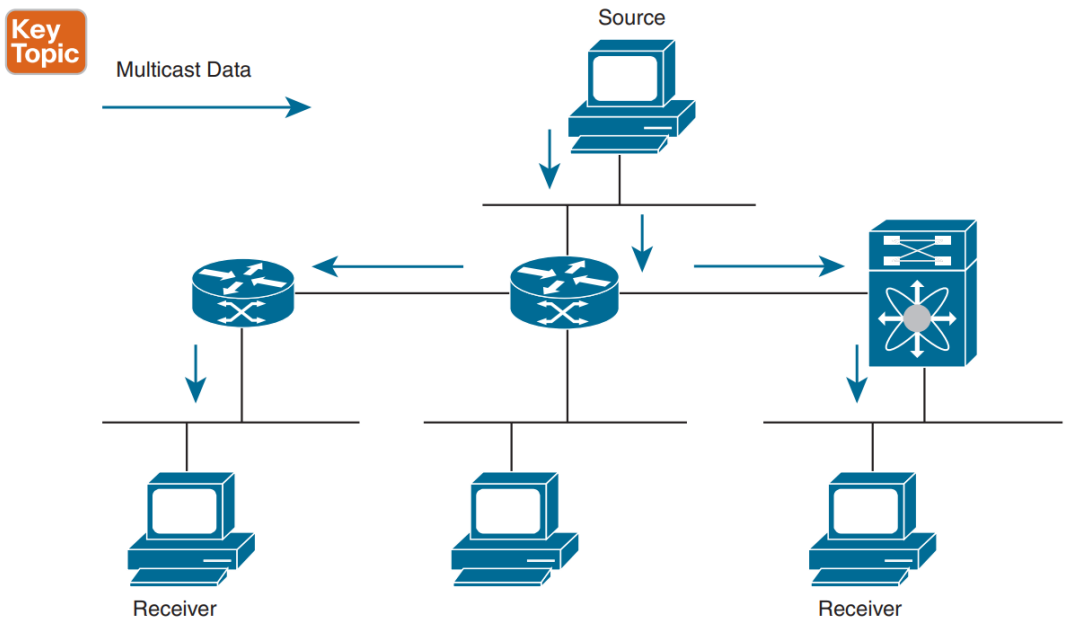
NOTE: **bfd slow-timer ,the slow timer used in the echo function . this value determines ho fast BFD starts up a new ession and at what speed the asynchronous sessions usee for BFD control packets whem the echo function is enabled.** The slow-timer value is used as the new
control packet interval, while the echo packets use the configured BFD intervals. **The echo packets are used for link failure detection, while the control packets at the slower rate maintain the BFD session**. The range is from 1000 to 30,000 milliseconds. The
default is 2000

NOTE: for integrating BFD alongside with BGP we use the command **router bgp as_number neighbor ip_add remote-as as_number bfd**

NOTE: You can optimize BFD on subinterfaces. BFD creates sessions for all configured subinterfaces. BFD setsthe subinterface with the lowest configured VLAN ID as the master subinterface and that subinterfaceuses the BFD session parameters of the parent interface. The remaining subinterfaces use the slow timer.If the optimized subinterface session detects an error, BFD marks all subinterfaces on that physicalinterface as down.

▼ MULTICAST

Multicast IP routing protocols used to distribute data to multiple recipients in a single session (for example, audio/video streaming broadcasts). Multicast IP can send IP data to
a group of interested receivers in a single transmission. Multicast IP addresses are called
groups. A multicast address that includes a group and source IP address is often referred
to as a channel. You can use multicast in both IPv4 and IPv6 networks to provide efficient
delivery of data to multiple destinations.
The Internet Assigned Numbers Authority (IANA) has assigned 224.0.0.0 through 239.255.255.255 as IPv4 multicast addresses, and IPv6 multicast addresses begin with 0xFF.



IGMP:Internet Group Management Protocol

The IGMP is **used by hosts that want to receive multucast data to request membership in multicasst gourp**. Onece the group membership is established , multicast data for the group is directed to the LAN segment of the requseting host . IGMP protocols is an ipv4 protocol that a host uses to request data for particular group. Using the info obtained through the IGMP, the software maintains a list of multicast groups or channel memberships on a oer-interface basis. **The systems that receive these IGMP packets send multicast data that they receive for requested groups or channels out the network segment of the known receivers.**

NX-OS supports IGMPv2 and IGMPv3. By default, NX-OS enables IGMPv2 when it starts
the IGMP process. You can enable IGMPv3 on interfaces where you want its capabilities.

IGMPv3 includes the following key changes from IGMPv2:

■ IGMPv3 supports source-specific multicast (SSM), which builds shortest path trees from each receiver to the source, through the following features:

■ Host messages that can specify both the group and the source.

■ The multicast state that is maintained for groups and sources, not just for groups as
in IGMPv2.

■ Hosts no longer perform report suppression, which means that hosts always send IGMP membership reports when an IGMP query message is received.

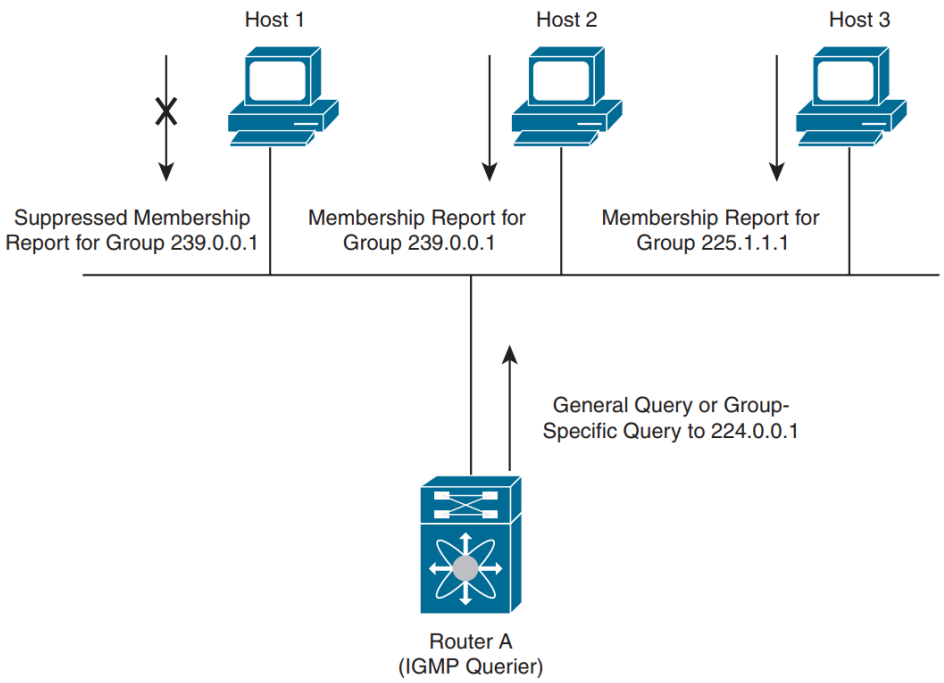


Figure 1-10 *IGMPv1 and IGMPv2 Query-Response Process*

In Figure 1-10, router A, which is the IGMP designated querier on the subnet, sends query messages to the all-hosts multicast group at 224.0.0.1 periodically to discover whether any hosts want to receive multicast data. You can configure the group membership timeout value that the router uses to determine that no members of a group or source exist on the subnet.

The software elects a router as the IGMP querier on a subnet if it has the lowest IP address.

As long as a router continues to receive query messages from a router with a lower IP address, it resets a timer that is based on its querier timeout value.

If the querier timer of a router expires, it becomes the designated querier. If that router later receives a host query message from a router with a lower IP address, it drops its role as the designated querier and sets its querier timer again.

In Figure 1-10, host 1's membership report is suppressed, and host 2 sends its membership report for group 239.0.0.1 first. Host 1 receives the report from host 2. **Because only one membership report per group needs to be sent to the router, other hosts suppress their reports to reduce network traffic**. Each host waits for a random time interval to avoid sending reports at the same time. **You can configure the query maximum response time parameter to control the interval in which hosts randomize their responses.**

NOTE IGMPv2 membership report suppression occurs only on hosts that are connected to the same port.

"IGMPv2 membership report suppression" refers to a feature or mechanism within IGMPv2 that allows the suppression or blocking of membership reports. In IGMP, hosts (devices) can send membership reports to indicate their interest in receiving multicast traffic from a specific group.

"Occurs only on hosts that are connected to the same port" means that this suppression of membership reports happens only for hosts that are connected to a particular network port. In other words, if multiple hosts are connected to the same network port, this feature is relevant for them.(same port here means the same subnet)

In Figure 1-11, router A sends the IGMPv3 group-and-source-specific query to the LAN. Hosts 2 and 3 respond to the query with membership reports that indicate that they want to receive data from the advertised group and source. This IGMPv3 feature supports SSM.

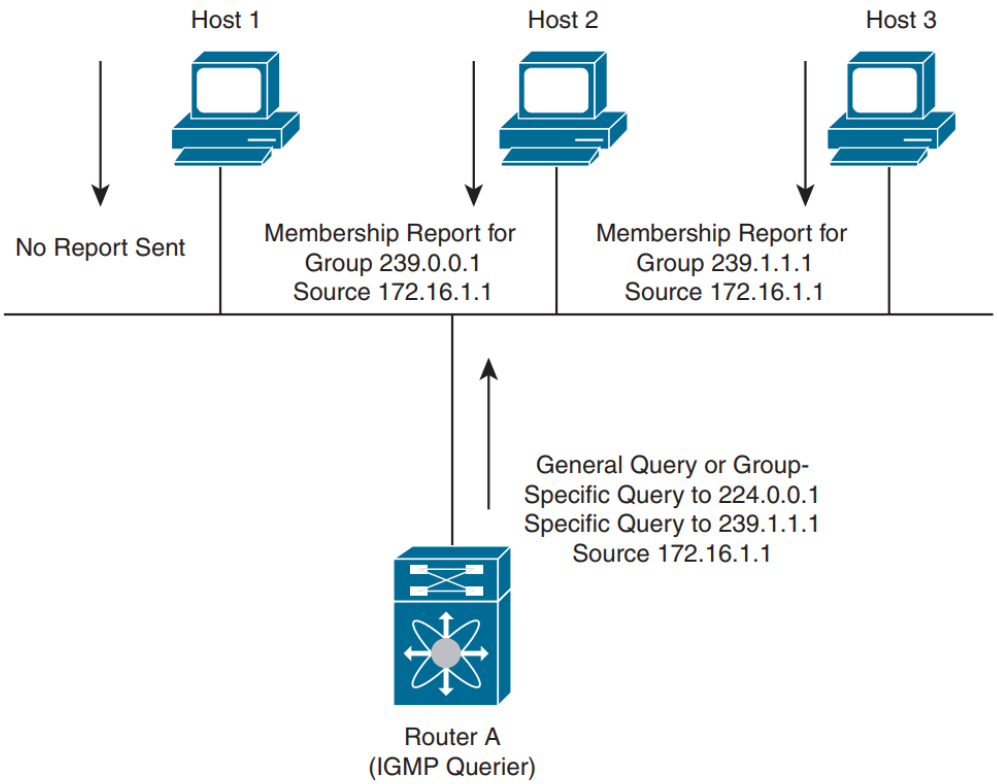


Figure 1-11 *IGMPv3 Group-and-Source-Specific Query*

NOTE IGMPv3 hosts do not perform IGMP membership report suppression.

IGMP messages sent by the designated querier have a time-to-live (TTL) value of 1, which
means that the messages are not forwarded by the directly connected routers on the subnet("Have a time-to-live (TTL) value of 1" means that the IGMP messages sent by the designated querier are assigned a Time-to-Live value of 1. The TTL value is a field in IP packets that indicates how many hops (routers) the packet can traverse before it is discarded. In this case, the TTL value of 1 means that the IGMP messages are limited to the local subnet and should not be forwarded beyond the directly connected routers on that subnet.).

You can configure the frequency and number of query messages sent specifically for IGMP
start-up, and you can configure a short query interval at start-up so that the group state is
established as quickly as possible.

Although usually unnecessary, you can tune the query
interval used after start-up to a value that balances the responsiveness to host group membership messages and the traffic created on the network("You can tune the query interval used after start-up" means that there is an option to modify the query interval duration after the initial start-up of the IGMP protocol. The query interval refers to the time duration between successive queries sent by the designated querier to hosts in order to check their group membership status.).

"To a value that balances the responsiveness to host group membership messages and the traffic created on the network" implies that the query interval can be adjusted to find a balance between two factors:

1. Responsiveness to host group membership messages: This refers to how quickly the network can detect changes in group membership. A shorter query interval allows for more frequent queries, resulting in quicker detection of host group membership changes.
2. Traffic created on the network: A longer query interval reduces the frequency of queries, which in turn reduces the amount of IGMP-related network traffic generated. This can help in optimizing network resources and reducing unnecessary traffic.

NOTE Changing the query interval can severely impact multicast forwarding.

IMPORTANT:By tuning the query interval, network administrators have the flexibility to adjust the timing between queries to find an appropriate balance. The objective is to ensure that the network remains responsive to host group membership changes while minimizing any unnecessary network traffic associated with IGMP query messages.It's important to note that adjusting the query interval should be done judiciously and with consideration for the specific network requirements, as making the interval too short or too long can have implications on network performance and responsiveness.

When a multicast host leaves a group, a host that runs IGMPv2 or later sends an IGMP
leave message. **To check if this host is the last host to leave the group, the software sends**
an IGMP query message and starts a timer that you can configure, called the last member

query response interval. If no reports are received before the timer expires, the software
removes the group state. The router continues to send multicast traffic for a group until its
state is removed.

You can configure a robustness value to compensate for packet loss on a congested network.
The robustness value is used by the IGMP software to determine the number of times to
send messages.
Link-local addresses in the range 224.0.0.0/24 are reserved by the Internet Assigned Numbers
Authority (IANA). Network protocols on a local network segment use these addresses; routers do not forward these addresses because they have a TTL of 1.
By default, the IGMP process sends membership reports only for nonlink-local addresses, but you can configure the
software to send reports for link-local addresses.

Switch IGMP Snooping:

IGMP snooping is a feature that limits multicast traffic on VLANs to the subset of ports
that have known receivers. The IGMP snooping software examines IGMP protocol messages
within a VLAN to discover which interfaces are connected to hosts or other devices interested in receiving this traffic. Using the interface information, IGMP snooping can reduce
bandwidth consumption in a multi-access LAN environment to avoid flooding the entire
VLAN. The IGMP snooping feature tracks which ports are attached to multicast-capable
routers to help it manage the forwarding of IGMP membership reports. Multicast traffic is
sent only to VLAN ports on which interested hosts reside.The IGMP snooping software

responds to topology change notifications.
NOTE: By default, IGMP snooping is enabled on the Cisco NX-OS system.

Multicast Listener Discovery:

Multicast Listener Discovery (MLD) **is an IPv6 protocol** that a host uses to request multicast
data for a particular group. Using the information obtained through MLD, the software
maintains a list of multicast group or channel memberships on a per-interface basis("Send the multicast data that they receive for requested groups or channels" means that when these devices receive multicast data (data sent to a specific multicast group or channel), they forward or send that data to the network segment where the known receivers are located. This implies that the devices act as intermediaries or routers for the multicast data.).

MLDv1 is derived from IGMPv2, and MLDv2 is derived from IGMPv3. IGMP uses **IP Protocol 2 message types**, whereas MLD uses **IP Protocol 58 message types**, which is a subset of
the ICMPv6 messages.

▼ LABS SECTION:

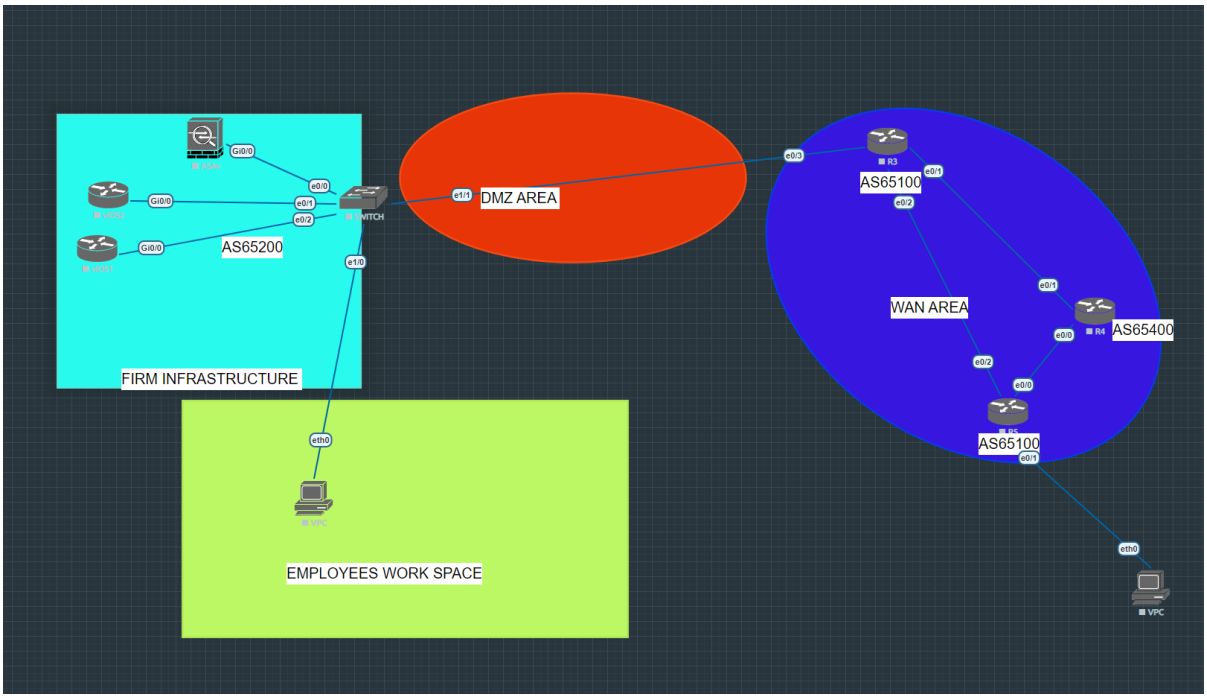
VRF

GRE

GRE-OVER-IPSEC

LISP

BGP



LAB file to test :

eBGp.unl

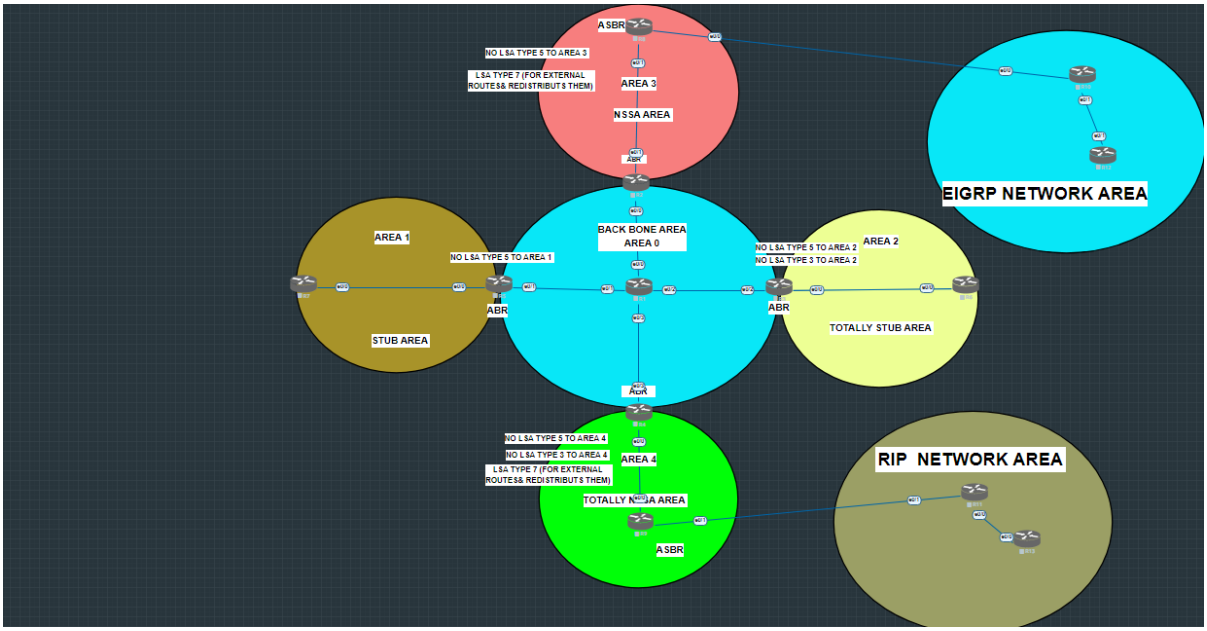
VXLAN

MPLS

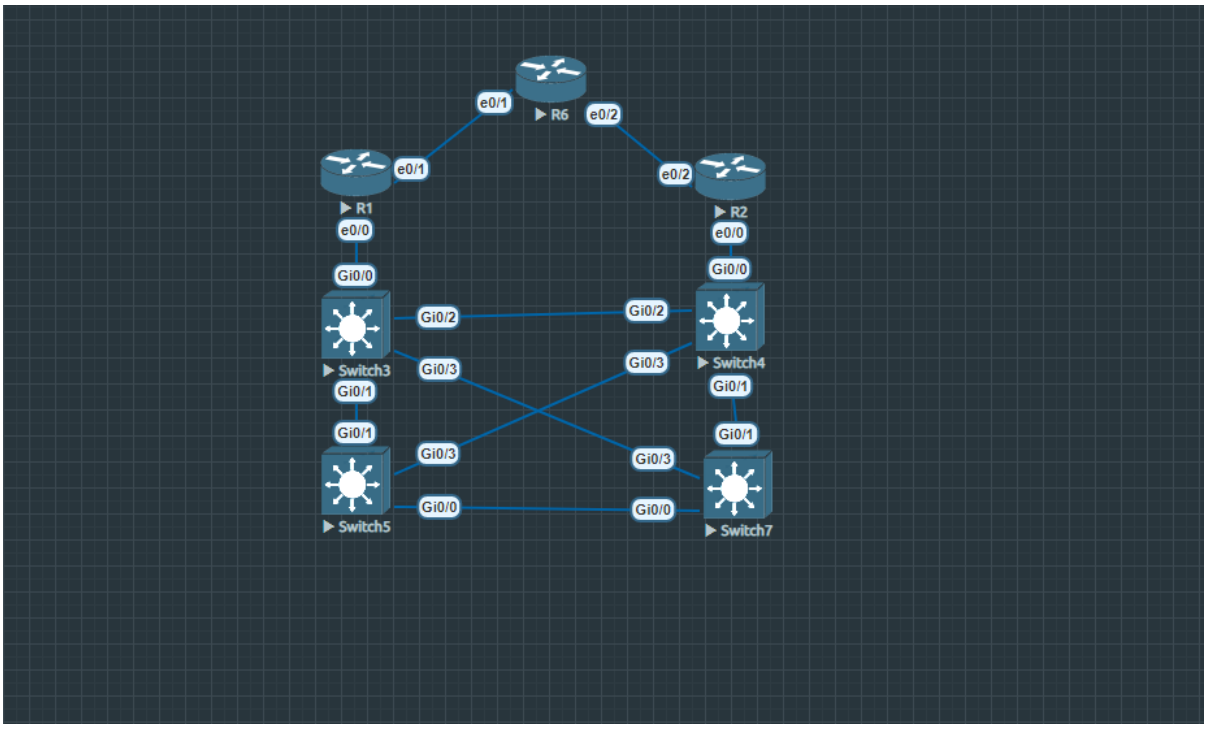
OPSFV3-MULTI-AREA(along with EIGRP&RIP)

1. WE START BY CONFIGURING THE BACK BONE AREA 0 OF THE NETWORK

2. WE MOVE TO THE AREA 1 (STUB AREA) USING THE COMMAND area 1 stub IN THE OSPFV3 PROCCESS CONFIGURED IN THAT AREA ROUTERS (OSPDV3 DOESN'T USE NETWORK COMMAND WE HAVE TO ENTER THE INTERFACE GLOBAL CONFIG MODE TO INCLUDE THE ROUTER TO PARTICIPATE IN THE OSPFV3 PROTOCOL)
3. WE MOVE TO THE AREA 2 (TOTALLY STUB AREA) USING THE COMMAND area 2 stub no-summary
4. WE MOVE TO AREA 3 (NSSA) USING THE COMMAND area 3 nssa AND WE CONFIGURE THE ROUTER ON THE EDGE TO ACT AS ASBR ROUTER SO WE NEED TO CONFIGURE EIGRP USING ipv6 router eigrp 1 AND CHOOSE A ROUTER-ID USING eigrp router-id
5. WE CONFIGURE ROUTER REDISTRIBUTION ON OSPF PROCESS USING redistribute eigrp 1 (THIS IS THE BASIC OSPF REDISTRIBUTION CONFIGUREATION)
6. ALSO WE NEED TO CONFIGURE EIGRP REDISTRIBUTION USING redisribute ospf 8
    a. WE MOVE TO AREA 4 (TOTALLY NSSA AREA) USING THE COMMAND area 4 nssa no-summary
    b. AND ALSO CONFIGURE THE ASBR REDISTRIBUTION



BFD WITH OSPF



Advanced Techniques

Check out this Notion Editor 101 guide for more advanced tips and how-to's.