Configuring Policy-Based Routing

This chapter contains the following sections:

About Policy-Based Routing

With policy-based routing, you can configure a defined policy for IPv4 and IPv6 traffic flows that lessens the reliance on routes derived from routing protocols. All packets received on an interface with policy-based routing enabled are passed through enhanced packet filters or route maps. The route maps dictate the policy that determines where to forward packets.

Policy-based routing includes the following features:

- Source-based routing—Routes traffic that originates from different sets of users through different connections across the policy routers.
- Quality of Service (QoS)—Differentiates traffic by setting the precedence or type of service (ToS) values in the IP packet headers at the periphery of the network and leveraging queuing mechanisms to prioritize traffic in the core or backbone of the network (see the Cisco Nexus 9000 Series NX-OS Quality of Service Configuration Guide).
- Load sharing—Distributes traffic among multiple paths based on the traffic characteristics.

Policy Route Maps

Each entry in a route map contains a combination of match and set statements. The match statements define the criteria for whether appropriate packets meet the particular policy (that is, the conditions to be met). The set clauses explain how the packets should be routed once they have met the match criteria.

You can mark the route-map statements as permit or deny. You can interpret the statements as follows:

- If the statement is marked as permit and the packets meet the match criteria, the set clause is applied. One of these actions involves choosing the next hop.
- If a statement is marked as deny, the packets that meet the match criteria are sent back through the normal forwarding channels, and destination-based routing is performed.
- If the statement is marked as permit and the packets do not match any route-map statements, the packets are sent back through the normal forwarding channels, and destination-based routing is performed.

Note Policy routing is specified on the interface that receives the packets, not on the interface from which the packets are sent.

Set Criteria for Policy-Based Routing

The Cisco Nexus 9000 Series switches support the following **set** commands for route maps used in policy-based routing:

- **set** {**ip** | **ipv6**} **next-hop**
- **set interface null0**

These **set** commands are mutually exclusive within the route-map sequence.

In the first command, the IP address specifies the adjacent next-hop router in the path toward the destination to which the packets should be forwarded. The first IP address associated with a currently up connected interface is used to route the packets.

Note You can optionally configure this command for next-hop addresses to load balance traffic for up to 32 IP addresses. In this case, Cisco NX-OS sends all traffic for each IP flow to a particular IP next-hop address.

If the packets do not meet any of the defined match criteria, those packets are routed through the normal destination-based routing process.

Route-Map Processing Logic

When an interface with a route map receives a packet, the forwarding logic processes each route-map statement according to the sequence number.

If the route-map statement encountered is a route-map...permit statement, the packet is matched against the criteria in the **match** command. This command may refer to an ACL that has one or more access control entries (ACEs). If the packet matches the permit ACEs in the ACL, the policy-based routing logic executes the action that the **set** command specifies on the packet.

If the route-map statement encountered is a route-map... deny statement, the packet is matched against the criteria in the match command. This command may refer to an ACL that has one or more ACEs. If the packet matches the permit ACEs in the ACL, policy-based routing processing stops, and the packet is routed using the default IP routing table.

Note The **set** command has no effect inside a **route-map... deny** statement.

- If the route-map configuration does not contain a match statement, the policy-based routing logic executes the action specified by the **set** command on the packet. All packets are routed using policy-based routing.
- If the route-map configuration references a match statement but the match statement references a non-existing ACL or an existing ACL without any access control entries (ACEs), the packet is routed using the default routing table.
- If the next-hop specified in the **set** { **ip** | **ipv6**} **next-hop** command is down, is not reachable, or is removed, the packet is routed using the default routing table.

Beginning Cisco NX-OS Release 9.2(3), you can balance policy-based routing traffic if the next hop is recursive over ECMP paths using the **next-hop ip-address load-share** command. This situation is supported on the following switches, line cards, and modules:

- N9K-C9372TX
- N9K-X9564TX
- N9K-X9732C-EX

For all the next hop routing requests, the Routing Profile Manager (RPM) resolves them using unicast Routing Information Base (uRIB). RPM also programs all ECMP paths, which helps to uniformly load balance all the ECMP paths. PBR over ECMP is supported only on IPv4.

Prerequisites for Policy-Based Routing

Policy-based routing has the following prerequisites:

- Assign an IP address on the interface and bring the interface up before you apply a route map on the interface for policy-based routing.

Guidelines and Limitations for Policy-Based Routing

Policy-based routing has the following configuration guidelines and limitations:

- Cisco Nexus 9500 platform switches with 9700-EX/FX line cards do not support PBR IPv6 Default Next hop for FIB Miss traffic.
- The following switches support IPv4 and IPv6 policy-based routing:
  - Cisco Nexus 9200 platform switches
  - Cisco Nexus 9300-EX/FX/FX2/FX3/GX/H1/H2R platform switches
  - Cisco Nexus 9508 switches with 9636C-R, 9636C-RX, and 9636Q-R line cards (For these line cards, PBR policy has a higher priority over attached and local routes. Explicit white listing might be required if protocol neighbors are directly attached.)

- A policy-based routing route map can have only one match statement per route-map statement.
- A policy-based routing route map can have only one set statement per route-map statement, unless you are using IP SLA policy-based routing. For information on IP SLA policy-based routing, see the *Cisco Nexus 9000 Series NX-OS IP SLAs Configuration Guide*.

Note Cisco Nexus 9508 switches with 9636C-R, 9636C-RX, and 9636Q-R line cards do not support IP SLA.

- A match command cannot refer to more than one ACL in a route map used for policy-based routing.
- The same route map can be shared among different interfaces for policy-based routing as long as the interfaces belong to the same virtual routing and forwarding (VRF) instance.
- Using a prefix list as a match criteria is not supported. Do not use a prefix list in a policy-based routing route map.
- Policy-based routing supports only unicast traffic. Multicast traffic is not supported.
- Policy-based routing is not supported with inbound traffic on FEX ports.
- Policy-based routing is not supported on FEX ports for Cisco Nexus 9300-EX platform switches.
- Only Cisco Nexus 9508 switches with 9636C-R, 9636C-RX, and 9636Q-R line cards support policy-based routing with Layer 3 port-channel subinterfaces.
- Beginning with Cisco NX-OS Release 10.1(2), policy-based routing with Layer 3 port-channel subinterfaces are supported on Cisco Nexus 9300-X Cloud Scale Switches.
- An ACL used in a policy-based routing route map cannot include deny access control entries (ACEs).
- Policy-based routing is supported only in the default system routing mode.
- The Cisco Nexus 9000 Series switches do not support the **set vrf** and **set default next-hop** commands.
- When you configure multiple features on an interface (such as PBR and ingress ACL), the ACLs for those features are merged for TCAM optimization. As a result, statistics are not supported.
- For PBR with VXLAN, the load-share keyword is not required.

Note Cisco Nexus 9500 platform switches with the 9700-EX/FX line cards support IPv4/IPv6 policy-based routing over VXLAN. Cisco Nexus 9508 switches with 9636C-R, 9636C-RX, and 9636Q-R line cards do not support policy-based routing over VXLAN.

- The Cisco Nexus 9000 Series switches support policy-based ACLs (PBACLs), also referred to as object-group ACLs. For more information, see the *Cisco Nexus 9000 Series NX-OS Security Configuration Guide*.

Note Cisco Nexus 9508 switches with 9636C-R, 9636C-RX, and 9636Q-R line cards do not support PBACLs.

- The following guidelines and limitations apply to PBR over VXLAN EVPN:
  - PBR over VXLAN EVPN is supported only for Cisco Nexus 9300-EX/FX/FX2/FX3/GX platform switches.
  - PBR over VXLAN EVPN does not support the following features: VTEP ECMP and the load-share keyword in the **set {ip | ipv6} next-hop** *ip-address* command.
- The following guidelines and limitations apply to PBR fast convergence:
  - PBR fast convergence is supported only for policies that have route-map sequences defined with multiple alternate next-hops, without load-share option, and with SLA probes for tracking next-hop availability.
  - Simultaneous failures of primary and back-up next-hops are not handled in the fast path. In such events, the system will fall back to control plane updates.
  - PBR fast convergence is primarily supported in events where adjacency loss is detected.
  - PBR fast convergence is not supported for next-hops reachable over VXLAN.
  - PBR fast convergence should not be used when next-hops are specified with millisecond SLAs/tracks to track availability.

    For more information about SLA, see the *Cisco Nexus 9000 Series NX-OS IP SLAs Configuration Guide*.
  - When PBR fast convergence is disabled, the number of ACL redirect entries is proportional to the number of unique primary next-hops across the PBR policies. When PBR fast convergence is enabled, the system may require ACL redirect entries per port-slice that is proportional to the number of unique combinations of primary and back-up next-hops configured across the route-map sequences in the PBR policies.
  - The following platforms support PBR fast convergence: N9K-C93180YC-FX, N9K-C93180YC2-FX, N9K-C93180YC-FX-24, N9K-C93108TC-FX, N9K-C93108TC2-FX, N9K-C93108TC-FX-24, N9K-C9336C-FX2, N9K-C93240YC-FX2, N9K-C93360YC-FX2, N9K-C93216TC-FX2, N9K-C9336C-FX2-E, N9K-C9316D-GX, N9K-C93600CD-GX, N9K-C9364C-GX.

Default Settings for Policy-Based Routing

Table 1: Default Policy-Based Routing Parameters

Parameters	Default
Policy-based routing	Disabled

Configuring Policy-Based Routing

Enabling the Policy-Based Routing Feature

You must enable the policy-based routing feature before you can configure a route policy.

Procedure

	Command or Action	Purpose

Step 1	**configure terminal**
	Example:
	`switch# configure terminal`
	`switch(config)#`
		Enters global configuration mode.

Step 2	[**no**] **feature pbr** **Example:** `switch(config)# feature pbr`	Enables the policy-based routing feature. Use the **no** form of this command to disable the policy-based routing feature. **Note** The **no feature pbr** command removes the policies applied under the interfaces. It does not remove the ACL or route-map configuration nor does it create a system checkpoint.
Step 3	(Optional) **show feature** **Example:** `switch(config)# show feature`	Displays enabled and disabled features.
Step 4	(Optional) **copy running-config startup-config** **Example:** `switch(config)# copy running-config startup-config`	Copies the running configuration to the startup configuration.

Enabling the Policy-Based Routing over ECMP

PBR over ECMP is not enabled by default. You must enable the policy-based routing feature before you can configure a route policy.

Procedure

	Command or Action	Purpose

Step 1	**configure terminal**
	Example:
	`switch# configure terminal`
	`switch(config)#`
		Enters global configuration mode.

Step 2	[**no**] **feature pbr** **Example:** `switch(config)# feature pbr`	Enables the policy-based routing feature. Use the **no** form of this command to disable the policy-based routing feature. **Note** The **no feature pbr** command removes the policies applied under the interfaces. It does not remove the ACL or route-map configuration nor does it create a system checkpoint.
Step 3	(Optional) **show feature** **Example:** `switch(config)# show feature`	Displays enabled and disabled features.

Step 4	**[no] hardware profile pbr ecmp paths** *<maxpath>*
	Example:
	`switch(config)# hardware profile pbr ecmp paths 12`
	`Warning!!: The pbr ecmp path limits have been changed.`
	`Please reload the switch now for the change to take effect.`
	`switch(config)#`
	`switch(config)# no hardware profile pbr ecmp paths 12`
	`Warning!!: The pbr ecmp path limits have been changed.`
	`Please reload the switch now for the change to take effect.`
	`switch(config)#`
		Configure the number of ECMP paths for IP next hop. However, the traffic may not go through all the paths unless you explicitly configure the load share in the set IP next hop. Whenever you remove or modify the PBR ECMP paths, the changes will take effect only after next reload. The range is from 1 through 64.

Step 5	**show system internal rpm state**	Displays the currently configured and operational values of PBR ECMP paths.

Configuring PBR Fast Convergence

In the case of a failure of a next-hop that is currently in use in PBR, PBR fast convergence can reduce the traffic convergence time to sub-second. PBR fast convergence assists policies that have route-map sequences

defined with multiple alternate next-hops, without the load-share option, and with SLA probes for tracking next-hop availability.

PBR fast convergence is disabled on the switch by default. After configuring PBR fast convergence and saving the configuration, you must reload the switch to activate PBR fast convergence.

Before you begin

You must enable the policy-based routing feature before you can configure PBR fast convergence.

Procedure

	Command or Action	Purpose

Step 1	**configure terminal**
	Example:
	`switch# configure terminal`
	`switch(config)#`
		Enters global configuration mode.

Step 2	[**no**] **feature pbr** **Example:** `switch(config)# feature pbr`	Enables the policy-based routing feature.
Step 3	[**no**] **hardware profile pbr next-hop fast-convergence** **Example:** `switch(config)# hardware profile pbr next-hop fast-convergence`	Configures PBR fast convergence. Use the **no** form of this command to disable PBR fast convergence. **Note** Enabling or disabling PBR fast convergence takes effect after the switch is reloaded.
Step 4	**copy running-config startup-config** **Example:** `switch(config)# copy running-config startup-config`	Copies the running configuration to the startup configuration.

Example

This example enables PBR fast convergence and reloads the switch:

```
switch(config)# hardware profile pbr next-hop fast-convergence
Warning: Please save config and reload the system for the configuration to take effect.
switch(config)# copy running-config startup-config
switch(config)# reload
```

What to do next

After enabling or disabling PBR fast convergence and saving the configuration, reload the switch.

Configuring a Route Policy

You can use route maps in policy-based routing to assign routing policies to the inbound interface. Cisco NX-OS routes the packets when it finds a next hop and an interface.

Procedure

	Command or Action	Purpose

Step 1	**configure terminal**
	Example:
	`switch# configure terminal`
	`switch(config)#`
		Enters global configuration mode.

Step 2	**interface** *type slot/port*
	Example:
	`switch(config)# interface ethernet 1/2`
	`switch(config-if)#`
		Enters interface configuration mode.

Step 3	(Optional) **hardware profile pbr ecmp paths** *paths_limit* **Example:** `switch(config-if)# hardware profile pbr ecmp paths 64`	You can specify the **hardware profile pbr ecmp paths** command if you want to limit the number of egress-paths for every PBR IP next-hop to a maximum of 64 paths. The range is 1–64 paths. For example, with two PBR next hop IP addresses configured as **set ip next-hop***address1 address2*, IP1 can be resolved across 32+ next-hops and IP2 can also resolve across 32+ next-hop as well. The effective member count can go above 64 members, which exceeds the hardware limitation of 64 members per ECMP group.

Step 4	{**ip** | **ipv6**} **policy route-map** *map-name*
	Example:
	`switch(config-if)# ip policy route-map Testmap`
	`switch(config-route-map)#`
		Assigns a route map for IPv4 or IPv6 policy-based routing to the interface.

Step 5	**route-map map-name [permit | deny] [seq]**
	Example:
	`switch(config-if)# route-map Testmap`
	`switch(config-route-map)#`
		Creates a route map or enters route-map configuration mode for an existing route map. Use seq to order the entries in a route map.

Step 6	Required: **match** {**ip** | **ipv6**} **address access-list-name** *name* [*name...*] **Example:** `switch(config-route-map)# match ip address access-list-name ACL1`	Matches an IPv4 or IPv6 address against one or more IP or IPv6 access control lists (ACLs). This command is used for policy-based routing and is ignored by route filtering or redistribution.

	Command or Action	Purpose

Step 7	(Optional) **set** {**ip** | **ipv6**} **next-hop** *address1* [*address2...*]**[load-share] [drop-on-fail] [force-order]**
	Example:
	`switch(config-route-map)# set ip next-hop 192.0.2.1`
	`switch(config-route-map)# set ipv6 next-hop 2001:0DB8::1`
		Sets the IPv4 or IPv6 next-hop address for policy-based routing. This command uses the first valid next-hop address if multiple addresses are configured.
		Use the optional **load-share** keyword to load balance traffic across a maximum of 32 next-hop addresses.
		Use the optional **force-order** keyword to enable next-hop ordering as specified in the CLI.
		Use the optional **drop-on-fail** keyword to drop packets instead of using default routing when the configured next hop becomes unreachable. This option is supported for Cisco Nexus 9200, 9300-EX/FX/FX2 and 9364C platform switches and Cisco Nexus 9500 platform switches with -EX/FX line cards.

Step 8	(Optional) **set** {**ip** | **ipv6**} **next-hop verify-availability** *next-hop-address* **track** *object* **Example:** `switch(config-route-map)# set ip next-hop verify-availability 192.0.2.2 track 1`	Use this command to configure policy routing to verify the reachability of the next hop of a route map before the switch performs policy routing to that next hop. Repeat this step to configure the route map to verify the reachability of other tracked objects. **Note** For additional information about object tracking, see the *Cisco Nexus 9000 Series NX-OS IP SLAs Configuration Guide*.
Step 9	(Optional) **set interface null0** **Example:** `switch(config-route-map)# set interface null0`	Sets the interface that is used for routing. Use the **null0** interface to drop packets.
Step 10	(Optional) **copy running-config startup-config** **Example:** `switch(config)# copy running-config startup-config`	Copies the running configuration to the startup configuration.

Verifying the Policy-Based Routing Configuration

To display policy-based routing configuration information, perform one of the following tasks:

Command	Purpose
show [**ip** | **ipv6**] **policy** [*name*]	Displays information about an IPv4 or IPv6 policy.
show route-map [*name*] **pbr-statistics**	Displays policy statistics.

Use the **route-map** *map-name* **pbr-statistics** command to enable policy statistics. Use the **clear route-map** *map-name* **pbr-statistics** command to clear these policy statistics.

Configuration Examples for Policy-Based Routing

This example shows how to configure a simple route policy on an interface:

```
feature pbr
ip access-list pbr-sample_1
  permit tcp host 10.1.1.1 host 192.168.2.1 eq 80
ip access-list pbr-sample_2
  permit tcp host 10.1.1.2 host 192.168.2.2 eq 80
!
route-map pbr-sample permit 10
match ip address pbr-sample_1
set ip next-hop 192.168.1.1
route-map pbr-sample permit 20
match ip address pbr-sample_2
set ip next-hop 192.168.1.2
!
route-map pbr-sample pbr-statistics

interface ethernet 1/2
  ip policy route-map pbr-sample
```

The following output verifies this configuration:

```
switch# show route-map pbr-sample

route-map pbr-sample, permit, sequence 10
 Match clauses:
   ip address (access-lists): pbr-sample_1
Set clauses:
   ip next-hop 192.168.1.1
route-map pbr-sample, permit, sequence 20
 Match clauses:
   ip address (access-lists): pbr-sample_2
 Set clauses:
   ip next-hop 192.168.1.2

switch# show route-map pbr-sample pbr-statistics

route-map pbr-sample, permit, sequence 10
Policy routing matches: 84 packets

route-map pbr-sample, permit, sequence 20
Policy routing matches: 94 packets

Default routing: 233 packets
```

Note **Policy routing matches** shown against every route-map sequence contains the number of packets in the incoming data traffic that has a match with the sequence in the route-map. This counter increments irrespective of whether the PBR redirection ('set' command of that sequence) is resolved or not. Correspondingly, in the example shown above, policy routing matches is shown against two route-map sequence (sequence 10 and 20) in the show route-map pbr-statistics pbr-sample output.

Note **Default routing** contains the number of packets in the incoming data traffic that has no match with any of the sequence in the route-map. Correspondingly, in the example shown above, default routing is shown only once at the end in the show route-map pbr-statistics pbr-sample output.

This example shows load sharing between ECMP and non ECMP paths:

```
switch# show run rpm
!Command: show running-config rpm
!Running configuration last done at: Sun Dec 23 16:02:32 2018
!Time: Sun Dec 23 16:06:13 2018

version 9.2(3) Bios:version 08.35
feature pbr

route-map policy1 pbr-statistics
route-map policy1 permit 10
  match ip address acl2
  set ip next-hop 131.1.1.2 load-share
route-map policy2 pbr-statistics
route-map policy2 permit 10
  match ip address acl2
  set ip next-hop verify-availability 131.1.1.2 track 1
  set ip next-hop verify-availability 30.1.1.2 track 2 load-share


interface Ethernet1/31
  ip policy route-map policy2
```

This example displays information about next hop routing request:

```
switch# show system internal rpm pbr ip nexthop
PBR IPv4 nexthop table for vrf default

30.1.1.2 Usable
  via 28.1.1.2 Ethernet1/18 a46c.2ae3.02a7

131.1.1.2 Usable
  via 111.1.1.2 Vlan81 8478.ac58.afc1
Usable
  via 112.1.1.2 Vlan82 8478.ac58.afc1
Usable
  via 113.1.1.2 Vlan83 8478.ac58.afc1
Usable
  via 114.1.1.2 Vlan84 8478.ac58.afc1
Usable
  via 115.1.1.2 Vlan85 8478.ac58.afc1
Usable
  via 116.1.1.2 Vlan86 8478.ac58.afc1
```

```
Usable
  via 117.1.1.2 Vlan87 8478.ac58.afc1
Usable
  via 118.1.1.2 Vlan88 8478.ac58.afc1
```

This example display routes from the unicast RIB:

```
switch# show ip route 130.1.1.2
IP Route Table for VRF "default"
'*' denotes best ucast next-hop
'**' denotes best mcast next-hop
'[x/y]' denotes [preference/metric]
'%<string>' in via output denotes VRF <string>

130.1.1.0/24, ubest/mbest: 8/0
    *via 111.1.1.2, Vlan81, [110/120], 00:07:57, ospf-1, inter
    *via 112.1.1.2, Vlan82, [110/120], 00:07:57, ospf-1, inter
    *via 113.1.1.2, Vlan83, [110/120], 00:07:57, ospf-1, inter
    *via 114.1.1.2, Vlan84, [110/120], 00:07:57, ospf-1, inter
    *via 115.1.1.2, Vlan85, [110/120], 00:07:57, ospf-1, inter
    *via 116.1.1.2, Vlan86, [110/120], 00:07:57, ospf-1, inter
    *via 117.1.1.2, Vlan87, [110/120], 00:07:57, ospf-1, inter
    *via 118.1.1.2, Vlan88, [110/120], 00:07:57, ospf-1, inter


switch# show ip route 30.1.1.2
IP Route Table for VRF "default"
'*' denotes best ucast next-hop
'**' denotes best mcast next-hop
'[x/y]' denotes [preference/metric]
'%<string>' in via output denotes VRF <string>

30.1.1.0/24, ubest/mbest: 1/0
    *via 28.1.1.2, [1/0], 00:38:36, static
```

Related Documents for Policy-Based Routing

Related Topic	Document Title
IP SLA PBR object tracking	Cisco Nexus 9000 Series NX-OS IP SLAs Configuration Guide
Troubleshooting information	Cisco Nexus 9000 Series NX-OS Troubleshooting Guide