Policy-based Routing

Typical routing systems and protocols forward traffic based on the destination address in the packet, which is used to look up an entry in a routing table. However, sometimes the traffic on your network requires a more hands-on approach. You might need to forward a packet based on the source address, the packet size, or other information in the packet header.

Policy-based routing (PBR) lets you make routing decisions based on filters that change the routing behavior of specific traffic so that you can override the routing table and influence where the traffic goes. For example, you can use PBR to help you reach the best bandwidth utilization for business-critical applications, isolate traffic for inspection or analysis, or manually load balance outbound traffic.

Policy-based routing is applied to incoming packets. All packets received on a PBR-enabled interface pass through enhanced packet filters that determine rules and specify where to forward the packets.

- You can create a maximum of 255 PBR match rules and 256 nexthop groups (this is the ECMP limit).
- You can apply only one PBR policy per input interface.
- You can match on source and destination IP address only.
- PBR is not supported for GRE or VXLAN tunneling.
- PBR is not supported on ethernet interfaces.
- A PBR rule cannot contain both IPv4 and IPv6 addresses.

Contents

- This topic describes ...
- Configure PBR
- Configuration Example
- Review Your Configuration
- Delete PBR Rules and Policies

Configure PBR

A PBR policy contains one or more policy maps. Each policy map:

- Is identified with a unique map name and sequence number. The sequence number is used to determine the relative order of the map within the policy.
- Contains a match source IP rule or a match destination IP rule, and a set rule.
  - To match on a source and destination address, a policy map can contain both match source and match destination IP rules.
  - A set rule determines the PBR nexthop for the policy. The set rule can contain a single nexthop IP address or it can contain a nexthop group. A nexthop group has more than one nexthop IP address so that you can use multiple interfaces to forward traffic.
  - To use ECMP, you configure a nexthop group.

To use PBR in Cumulus Linux, you define a PBR policy and apply it to the ingress interface (the interface must already have an IP address assigned). Traffic is matched against the match rules in sequential order and forwarded according to the set rule in the first match. Traffic that does not match any rule is passed onto the normal destination based routing mechanism.

For Tomahawk and Tomahawk+ platforms, you must configure the switch to operate in non-atomic mode, which offers better scaling as all TCAM resources are used to actively impact traffic. Add the line acl.non_atomic_update_mode = TRUE to the /etc/cumulus/switchd.conf file. For more information, see Nonatomic Update Mode vs. Atomic Update Mode.

To configure a PBR policy:

1. Configure the policy map with the net add pbr-map <name> seq <1-700> match dst-ip|src-ip <ip/prefixlen> command.
   The example commands below configure a policy map called map1 with sequence number 1, that matches on destination address 10.1.2.0/24 and source address 10.1.4.1/24.

   ```sh
   cumulus@switch:~$ net add pbr-map map1 seq 1 match dst-ip 10.1.2.0/24
   cumulus@switch:~$ net add pbr-map map1 seq 1 match src-ip 10.1.4.1/24
   ```

   If the IP address in the rule is 0.0.0.0/0 or ::/0, any IP address is a match. You cannot mix IPv4 and IPv6 addresses in a
2. Either apply a `nexthop` or a `nexthop group` to the policy map:
   - To apply a `nexthop` to the policy map, use the `net add pbr-map <name> seq <1-700> set nexthop <ipaddress> [ <interface> ] [ nexthop-vrf <vrfname> ]` command.
     The output interface and VRF are optional, however, you must specify the VRF you want to use for resolution if the nexthop is not in the default VRF.
     The example command below applies the nexthop 192.168.0.31 on the output interface swp2 and VRF `rocket` to the `map1` policy map:
     
     ```
     cumulus@switch:~$ net add pbr-map map1 seq 1 set nexthop 192.168.0.31 swp2 nexthop-vrf rocket
     ```
   - To apply a `nexthop group` (for ECMP) to the policy map, first create the nexthop group, then apply the group to the policy map:
     a. Create the `nexthop group` with the `net add nexthop-group <groupname> nexthop <ipaddress> [ <interface> ] [ nexthop-vrf <vrfname> ]` command.
        The output interface and VRF are optional. However, you must specify the VRF if the nexthop is not in the default VRF.
        The example commands below create a `nexthop group` `group1` that contains the nexthop 192.168.0.21 on output interface `swp1` and VRF `rocket`, and the nexthop 192.168.0.22.
        
        ```
        cumulus@switch:~$ net add nexthop-group group1 nexthop 192.168.0.21 swp1 nexthop-vrf rocket
        cumulus@switch:~$ net add nexthop-group group1 nexthop 192.168.0.22
        ```
     b. Apply the `nexthop group` to the policy map with the `net add pbr-map <name> seq <1-700> set nexthop-group <groupname>` command.
        The example command below applies the `nexthop group group1` to the `map1` policy map:
        
        ```
        cumulus@switch:~$ net add pbr-map map1 seq 1 set nexthop-group group1
        ```

3. Assign the PBR policy to an ingress interface with the `net add interface <interface> pbr-policy <name>` command.
   The example command below assigns the PBR policy `map1` to interface `swp51`:
   
   ```
   cumulus@switch:~$ net add interface swp51 pbr-policy map1
   cumulus@switch:~$ net pending
   cumulus@switch:~$ net commit
   ```

   You can only set one policy per interface.

**Configuration Example**

In the following example, the PBR-enabled switch has a PBR policy to route all traffic from the Internet to a server that performs anti-DDOS. The traffic returns to the PBR-enabled switch after being cleaned and is then passed onto the regular destination based routing mechanism.

The configuration for the example above is:
cumulus@switch:~$ net add pbr-map map1 seq 1 match src-ip 0.0.0.0/0
cumulus@switch:~$ net add pbr-map map1 seq 1 set nexthop 192.168.0.32
cumulus@switch:~$ net add interface swp51 pbr-policy map1
cumulus@switch:~$ net pending
cumulus@switch:~$ net commit

These commands produce the following snippet in the /etc/frr/frr.conf file.

```plaintext
interface swp51
  pbr-policy map1
  pbr-map map1 seq 1
  match src-ip 0.0.0.0/0
  set nexthop 192.168.0.32
```

Review Your Configuration

Use the following commands to see the configured PBR policies.

To see the policies applied to all interfaces on the switch, use the `net show pbr interface` command. For example:

```
cumulus@switch:~$ net show pbr interface
swp55s3(67) with pbr-policy map1
```

To see the policies applied to a specific interface on the switch, add the interface name at the end of the command; for example, `net show pbr interface swp51`.

To see information about all policies, including mapped table and rule numbers, use the `net show pbr map` command. If the rule is not set, you see a reason why.

```
cumulus@switch:~$ net show pbr map
pbr-map map1 valid: 1
  Seq: 700 rule: 999 Installed: 1(1) Reason: Valid
  SRC Match: 10.0.0.1/32
  nexthop 192.168.0.32
  Installed: 1(1) Tableid: 10003
Seq: 701 rule: 1000 Installed: 1(2) Reason: Valid
  SRC Match: 90.70.0.1/32
  nexthop 192.168.0.32
  Installed: 1(1) Tableid: 10004
```

To see information about a specific policy, what it matches, and with which interface it is associated, add the map name at the end of the command; for example, `net show pbr map map1`.

To see information about all nexthop groups, run the `net show pbr nexthop-group` command:
cumulus@switch:~$ net show pbr nexthop-group
Nexthop-Group: map1701 Table: 10004 Valid: 1 Installed: 1
  Valid: 1 nexthop 10.1.1.2
Nexthop-Group: map1700 Table: 10003 Valid: 1 Installed: 1
  Valid: 1 nexthop 10.1.1.2
Nexthop-Group: group1 Table: 10000 Valid: 1 Installed: 1
  Valid: 1 nexthop 10.1.1.2
Nexthop-Group: group1 Table: 10001 Valid: 1 Installed: 1
  Valid: 1 nexthop 192.168.10.0 bond1
    Valid: 1 nexthop 192.168.10.2
    Valid: 1 nexthop 192.168.10.3 vlan70
Nexthop-Group: group2 Table: 10001 Valid: 1 Installed: 1
  Valid: 1 nexthop 192.168.8.1
    Valid: 1 nexthop 192.168.8.2
    Valid: 1 nexthop 192.168.8.3

To see information about a specific nexthop group, add the group name at the end of the command; for example, net show pbr nexthop-group group1.

A new Linux routing table ID is used for each nexthop and nexthop group.

Delete PBR Rules and Policies

You can delete a PBR rule, a nexthop group, or a policy with the net del command. The following commands provide examples.

Use caution when deleting PBR rules and nexthop groups, as you might create an incorrect configuration for the PBR policy.

The following example shows how to delete a PBR rule:

cumulus@switch:~$ net del pbr-map map1
cumulus@switch:~$ net pending
cumulus@switch:~$ net commit

The following example shows how to delete a PBR rule match:

cumulus@switch:~$ net del pbr-map map1 seq 1 match dst-ip 10.1.2.0/24
cumulus@switch:~$ net pending
cumulus@switch:~$ net commit

The following example shows how to delete a nexthop group:

cumulus@switch:~$ net del nexthop-group group1
cumulus@switch:~$ net pending
cumulus@switch:~$ net commit

The following example shows how to delete a nexthop from a group:

cumulus@switch:~$ net del nexthop 192.168.10.3

cumulus@switch:~$ net pending
cumulus@switch:~$ net commit
The following example shows how to delete a PBR policy so that the PBR interface is no longer receiving PBR traffic:

```bash
cumulus@switch:~$ net del interface swp3 pbr-policy map1
cumulus@switch:~$ net pending
cumulus@switch:~$ net commit
```