BGP PIC (Prefix Independent Convergence) Edge for IP and MPLS-VPN

The BGP PIC (Prefix Independent Convergence) Edge for IP and MPLS-VPN feature improves BGP convergence after a network failure. This convergence is applicable to both core and edge failures and can be used in both IP and MPLS networks. The BGP PIC Edge for IP and MPLS-VPN feature creates and stores a backup or alternate path in the routing information base (RIB), forwarding information base (FIB), and Cisco Express Forwarding. When a failure is detected, the backup or alternate path immediately takes over, thus enabling fast failover.

Note In this document, the BGP PIC Edge for IP and MPLS-VPN feature is called by the short name BGP PIC.

- Prerequisites for BGP PIC, on page 1
- Restrictions for BGP PIC, on page 2
- Benefits, on page 2
- BGP Convergence, on page 2
- Improve Convergence, on page 2
- Detect a Failure, on page 4
- MPLS VPN–BGP Local Convergence, on page 5
- Enable BGP PIC, on page 5
- BGP PIC Scenario, on page 5
- Configure BGP PIC, on page 6

Prerequisites for BGP PIC

- Ensure that the Border Gateway Protocol (BGP) and the IP or Multiprotocol Label Switching (MPLS) network is up and running at the customer site that is connected to the provider site by more than one path (multihomed).
- Ensure that the backup or alternate path has a unique next hop that is not the same as the next hop of the best path.
- Enable the Bidirectional Forwarding Detection (BFD) protocol to quickly detect link failures of neighbors that are directly connected.

Restrictions for BGP PIC

- Unlabeled BGP PIC EDGE for global prefixes is not supported.
- TE, SR, SR-TE, flex-LSP are not supported.
- BVI as a core is not supported.
- Only one primary and one backup path is supported. No support for multiple primary paths and one backup path.
- PIC EDGE is supported for Global IPv4, IPv6 (6PE), and MPLS-VPN prefixes (VPNv4 and VPNv6).

Benefits

- An extra path for failover allows faster restoration of connectivity when a primary path is invalid or withdrawn.
- Reduction of traffic loss.
- Constant convergence time so that the switching time is the same for all prefixes.

BGP Convergence

Under normal circumstances, BGP can take several seconds to a few minutes to converge after a change in the network. At a high level, BGP goes through the steps of the following process:

1. BGP learns of failures through either Interior Gateway Protocol (IGP) or BFD events or interface events.
2. BGP withdraws the routes from the routing information base (RIB), and the RIB withdraws the routes from the forwarding information base (FIB) and distributed FIB (dFIB). This process clears the data path for the affected prefixes.
3. BGP sends withdrawn messages to its neighbors.
4. BGP calculates the next best path to the affected prefixes.
5. BGP inserts the next best path for affected prefixes into the RIB, and the RIB installs them in the FIB and dFIB.

This process may take from few seconds to a few minutes to complete. It depends on, the latency of the network, the convergence time across the network, and the local load on the devices. The data plane converges only after the control plane converges.

Improve Convergence

The BGP PIC functionality is achieved by an extra functionality in the BGP, RIB, Cisco Express Forwarding, and MPLS.

- BGP Functionality

BGP PIC affects prefixes under IPv4 and VPNv4 address families. For those prefixes, BGP calculates an extra second best path, along with the primary best path. (The second best path is called the backup or alternate path.) BGP installs the best and backup or alternate paths for the affected prefixes into the BGP RIB. The backup or alternate path provides a fast reroute mechanism to counter a singular network failure. BGP also includes the alternate or backup path in its application programming interface (API) to the IP RIB.

- RIB Functionality

For BGP PIC, RIB installs an alternate path per route if one is available. If the RIB selects a BGP route containing a backup or alternate path, it installs the backup or alternate path with the best path. The RIB also includes the alternate path in its API with the FIB.

- Cisco Express Forwarding Functionality

With BGP PIC, Cisco Express Forwarding stores an alternate path per prefix. When the primary path goes down, Cisco Express Forwarding searches for the backup or alternate path in a prefix-independent manner. Cisco Express Forwarding also listens to BFD events to rapidly detect local failures.

- MPLS Functionality

MPLS Forwarding is similar to Cisco Express Forwarding in that it stores alternate paths and switches to an alternate path if the primary path goes down.

When the BGP PIC feature is enabled, BGP calculates a backup or alternate path per prefix and installs it into BGP RIB, IP RIB, and FIB. This improves convergence after a network failure. There are two types of network failures that the BGP PIC feature detects:

- Core node or link failure (internal Border Gateway Protocol [iBGP] node failure): If a PE node or link fails, then the failure is detected through IGP convergence. IGP conveys the failure through the RIB to the FIB.
- Local link or immediate neighbor node failure (external Border Gateway Protocol [eBGP] node or link failure): To detect a local link failure or eBGP single-hop peer node failure in less than a second, you must enable BFD. Cisco Express Forwarding looks for BFD events to detect a failure of an eBGP single-hop peer.

Convergence in the Data Plane

Upon detecting a failure, Cisco Express Forwarding detects the alternate next hop for all prefixes that are affected by the failure. The data plane convergence is achieved in subseconds depending on whether the BGP PIC implementation exists in the software or hardware.

Convergence in the Control Plane

Upon detecting a failure, BGP learns about the failure through IGP convergence or BFD events and sends withdrawn messages for the prefixes, recalculating the best and backup or alternate paths, and advertising the next best path across the network.

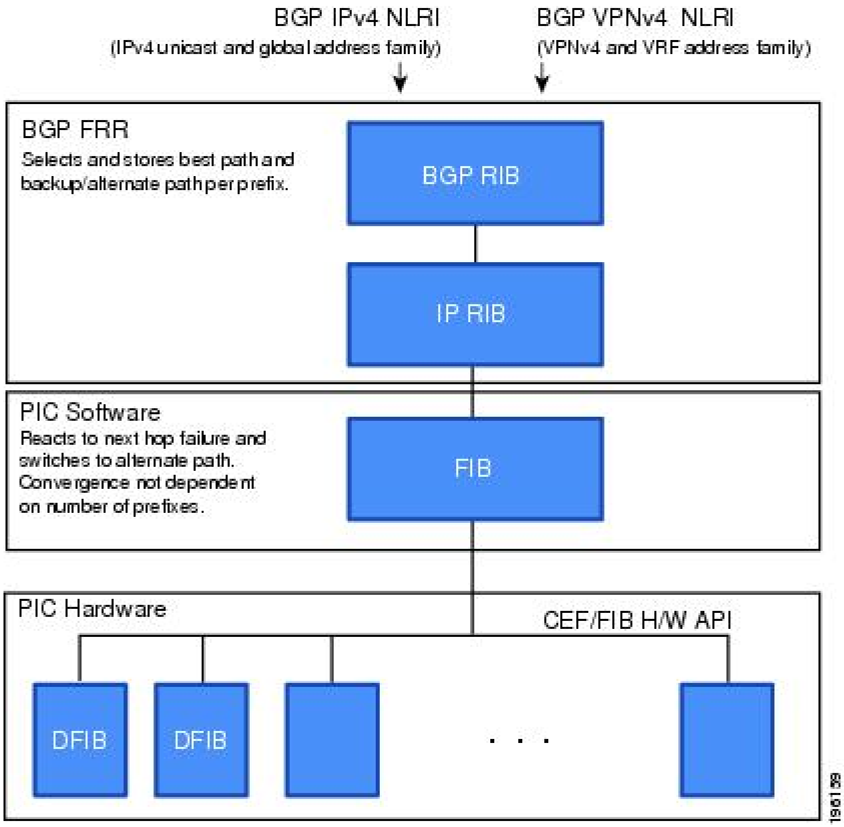
BGP Fast Reroute

BGP Fast Reroute (FRR) provides a best path and a backup or alternate path in BGP, RIB, and Cisco Express Forwarding. BGP FRR provides a fast reroute mechanism into the RIB and Cisco Express Forwarding (CEF) on the backup BGP next hop to reach a destination when the current best path is not available.

BGP FRR precomputes a second best path in BGP and gives it to the RIB and Cisco Express Forwarding as a backup or alternate path, and CEF programs it into line cards.

The BGP PIC feature provides the ability for CEF to quickly switch the traffic to the other egress ports if the current next hop or the link to this next hop goes down.

Figure 1: BGP PIC Edge and BGP FRR


BGP IPv4 NLRI
(IPv4 unicast and global address family)
BGP VPNv4 NLRI
(VPNv4 and VRF address family)
BGP FRR
Selects and stores best path and backup/alternate path per prefix.
BGP RIB
IP RIB
PIC Software
Reacts to next hop failure and switches to alternate path. Convergence not dependent on number of prefixes.
FIB
PIC Hardware
CEF/FIB H/W API
DFIB
DFIB
. . .
190159

Detect a Failure

IGP detects a failure in the iBGP (remote) peer; it may take a few seconds to detect the failure. Convergence can occur in subseconds or seconds, depending on whether PIC is enabled on the line cards.

If the failure is among the directly connected neighbors (eBGP), and if you use BFD to detect when a neighbor has gone down. Depending on whether PIC is enabled on the line cards, the detection may happen within subseconds and the convergence can occur in subseconds or few seconds.

MPLS VPN–BGP Local Convergence

The BGP PIC is an enhancement to the MPLS VPN–BGP Local Convergence feature. It provides a failover mechanism that recalculates the best path after a link failure. It then installs the new path in forwarding. To minimize traffic loss, the feature maintains the local label for 5 minutes to ensure that the traffic uses the backup or alternate path.

The BGP PIC improves the LoC time to under a second by calculating a backup or alternate path in advance. When a link failure occurs, the traffic is sent to the backup or alternate path.

When you configure BGP PIC, it overrides the functionality of the MPLS VPN--BGP Local Convergence feature. Do not remove the **protection local-prefixes** command from the configuration.

Enable BGP PIC

BGP PIC Edge can be enabled on the following address families:

- IPv4
- IPv6
- VPNv4
- VPNv6

BGP PIC Scenario

You can configure the BGP PIC functionality to achieve fast convergence.

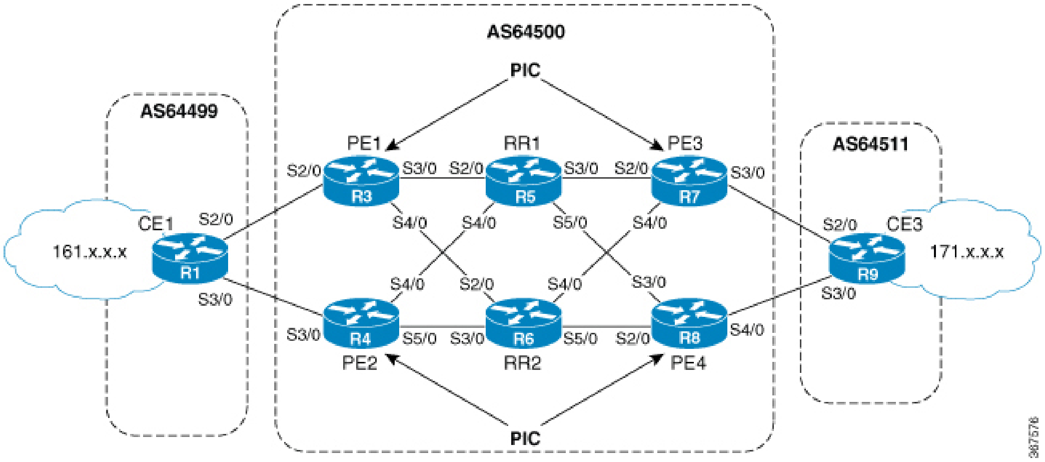
IP PE-CE Link and Node Protection

The network includes the following components:

- Traffic from CE1 (161.x.x.x) uses PE1 to reach network 171.x.x.x through router CE3. CE1 has two paths:
    - PE1 as the primary path.
    - PE2 as the backup or alternate path.

PE1, PE2, PE3, and PE4 are configured with the BGP PIC Edge feature. PE1 learns about prefixes 161.x.x.x from CE1. Also PE1 learns about the same prefix through PE2, from Route Reflectors (RR1 and RR2). PE1 installs primary and backup for prefix 161.x.x.x. When the link between PE1-CE1 goes down, PIC Edge is triggered on PE1, so the BGP PIC Edge becomes active and sends traffic to CE1 through PE2. This is BGP PIC Edge during a PE-CE link failure.

Figure 2: Using BGP PIC to Protect the PE-CE Link



- Similarly, PE1 has two paths to reach network 171.x.x.x through router CE3:
    - PE3 as the primary path.
    - PE4 as the backup or alternate path.

PE1 learns about prefixes 171.x.x.x from PE3 and PE4 through RR1 and RR2 and it installs primary and backup for this prefix. When PE3 goes down, BGP PIC Edge is triggered on PE1 and traffic is rerouted to PE4. This is BGP PIC Edge during a node failure.

Configure BGP PIC

Step 1 **cef encap-sharing disable**

Example:

```
RP/0/RP0/CPU0:router(config)# cef encap-sharing disable
```

By default, IPv4 global prefixes are installed with primary and backup path (if available) in the hardware. To install the protection in IPv6 (6 PE), VPNv4, and VPNv6 prefixes in the hardware, you must configure CLI **cef encap-sharing disable** command in global configuration mode.

Caution This CLI reprograms the CEF completely and impacts traffic. We recommend that you do it in the maintenance window.

Step 2 **router bgp** *as-number*

Example:

```
RP/0/RP0/CPU0:router(config)# router bgp 100
```

Specifies the autonomous system number and enters the BGP configuration mode, allowing you to configure the BGP routing process.

Step 3 **address-family** {**vpnv4 unicast** | **vpnv6 unicast** | **ipv4 unicast** | **ipv6 unicast**}

Example:

```
RP/0/RP0/CPU0:router(config-bgp)# address-family ipv4 unicast
```

```
address-family ipv4 unicast
  additional-paths receive
  additional-paths selection route-policy backup 1
  allocate-label all
!
```

Step 4 **additional-paths selection route-policy** *route-policy-name*

Example:

```
RP/0/RP0/CPU0:router(config-bgp-af)# additional-paths selection route-policy ap1
```

Configures extra paths selection mode for a prefix.

Note Use the **additional-paths selection** command with an appropriate route-policy to calculate backup paths and to enable Prefix-Independent Convergence (PIC) functionality.

The route-policy configuration is a prerequisite for configuring the additional-paths selection mode for a prefix. This is an example route-policy configuration to use with additional-selection command:

```
route-policy ap1
    set path-selection backup 1 install
  end-policy
```