



Segment Routing On Demand Next Hop for L3/L3VPN

When redistributing routing information across domains, provisioning of multi-domain services (L2VPN & L3VPN) has its own complexity and scalability issues. On Demand Next Hop (ODN) triggers delegation of computation of an end-to-end LSP to a PCE controller including constraints and policies without doing any redistribution. It then installs the replied multi-domain LSP for the duration of the service into the local forwarding information base (FIB).

- [Restrictions for Segment Routing On Demand Next Hop for L3/L3VPN, on page 1](#)
- [Information About Segment Routing On Demand Next Hop for L3/L3VPN, on page 2](#)
- [How to Configure Segment Routing On Demand Next Hop for L3/L3VPN, on page 2](#)
- [Verifying Segment Routing On Demand Next Hop for L3/L3VPN, on page 6](#)
- [Additional References for Segment Routing On Demand Next Hop for L3/L3VPN , on page 10](#)
- [Feature Information for Segment Routing On Demand Next Hop for L3/L3VPN, on page 10](#)

Restrictions for Segment Routing On Demand Next Hop for L3/L3VPN

- On Demand Next Hop (ODN) anycast SID is not supported.
- ODN for IPv6 is not supported.
- SR ODN tunnel is not supported with BGP Nonstop Routing (NSR). It is only supported with BGP Nonstop Forwarding (NSF).

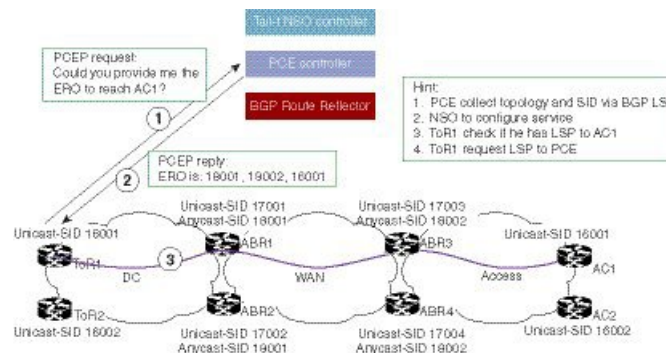
To enable BGP NSF, use the following command:

```
bgp grace-full restart  
neighbor 10.0.0.2 ha-mode graceful-restart
```

Information About Segment Routing On Demand Next Hop for L3/L3VPN

On Demand Next hop leverages upon BGP Dynamic SR-TE capabilities and adds the path computation (PCE) ability to find and download the end to end path based on the requirements. ODN triggers an SR-TE auto-tunnel based on the defined BGP policy. As shown in the below figure, an end to end path between ToR1 and AC1 can be established from both ends based on low latency or other criteria for VRF (L3VPN) or IPv4 services. The work-flow for ODN is summarized as follows:

Figure 1: ODN Operation



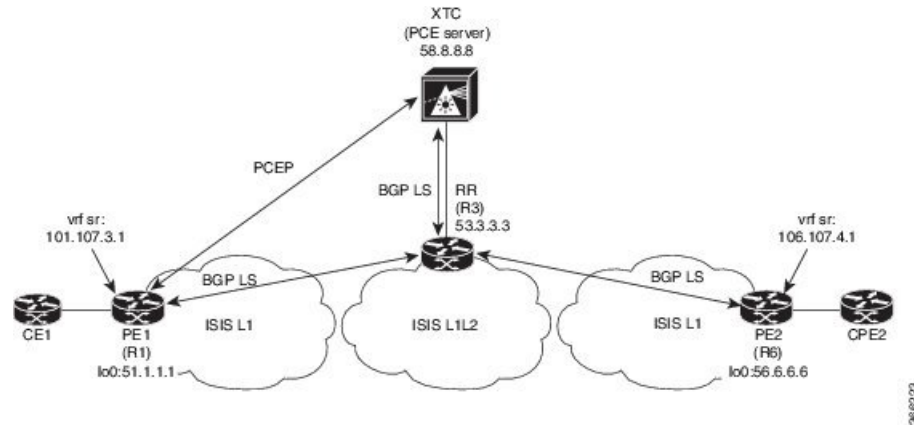
1. PCE controller collects topology and SIDs information via BGP Link State (BGP-LS). For more information on BGP-LS, refer [BGP Link-State](#).
2. If NSO controller is enable, it configures L3VPN VRF or IPv4 prefixes and requests are sent to ToR1 and AC1.
3. ToR1 and AC1 checks if a LSP towards each other exists. If not, a request is sent to the PCE controller to compute that SR-TE path that matches SR-TE policy that is carried via BGP.
4. PCE controller computes the path and replies with a label stack (18001, 18002, 16001, example in ToR1).
5. ToR1 and AC1 create a SR-TE auto-tunnel and reply back to the NSO controller indicating that the LSP for VRF or IPv4 is up and operational.

How to Configure Segment Routing On Demand Next Hop for L3/L3VPN

Configuring Segment Routing On Demand Next Hop for L3/L3VPN

Perform the following steps to configure on-demand next hop for SR-TE. The below figure is used as a reference to explain the configuration steps.

Figure 2: ODN Auto-Tunnel Setup



1. Configure the router (R6 tail end) with VRF interface.

```
interface GigabitEthernet0/2/2
vrf forwarding sr
ip address 10.0.0.1 255.0.0.0
negotiation auto

interface Loopback0
ip address 192.168.0.1 255.255.0.0
ip router isis 1
```

2. Tags VRF prefix with BGP community on R6 (tail end).

```
route-map BGP_TE_MAP permit 9
match ip address traffic
set community 3276850

ip access-list extended traffic
permit ip 10.0.0.1 255.255.0.0 any
```

3. Enable BGP on R6 (tail end) and R1 (head end) to advertise and receive VRF SR prefix and match on community set on R6 (tail end).

```
router bgp 100
bgp router-id 172.16.0.1
bgp log-neighbor-changes
bgp graceful-restart
no bgp default ipv4-unicast
neighbor 10.0.0.2 remote-as 100
neighbor 10.0.0.2 update-source Loopback0

address-family ipv4
neighbor 10.0.0.2 activate
neighbor 10.0.0.2 send-community both
neighbor 10.0.0.2 next-hop-self
exit-address-family

address-family vpnv4
neighbor 10.0.0.2 activate
neighbor 10.0.0.2 send-community both
neighbor 10.0.0.2 route-map BGP_TE_MAP out
exit-address-family
```

```

address-family link-state link-state
  neighbor 10.0.0.2 activate
exit-address-family

address-family ipv4 vrf sr
  redistribute connected
exit-address-family

route-map BGP_TE_MAP permit 9
  match ip address traffic
  set community 3276850

ip access-list extended traffic
  permit ip 10.0.0.1 255.255.0.0 any

router bgp 100
  bgp router-id 192.168.0.2
  bgp log-neighbor-changes
  bgp graceful-restart
  no bgp default ipv4-unicast
  neighbor 10.0.0.2 remote-as 100
  neighbor 10.0.0.2 update-source Loopback0

address-family ipv4
  neighbor 10.0.0.2 activate
  neighbor 10.0.0.2 send-community both
  neighbor 10.0.0.2 next-hop-self
exit-address-family

address-family vpnv4
  neighbor 10.0.0.2 activate
  neighbor 10.0.0.2 send-community both
  neighbor 10.0.0.2 route-map BGP_TE_MAP in
exit-address-family

address-family link-state link-state
  neighbor 10.0.0.2 activate
exit-address-family

address-family ipv4 vrf sr
  redistribute connected
exit-address-family

route-map BGP_TE_MAP permit 9
  match community 1
  set attribute-set BGP_TE5555

ip community-list 1 permit 3276850

mpls traffic-eng lsp attributes BGP_TE5555
  path-selection metric igp
  pce

```

4. Enable route-map/attribute set on headend (R1).

```

route-map BGP_TE_MAP permit 9
  match community 1
  set attribute-set BGP_TE5555

ip community-list 1 permit 3276850

mpls traffic-eng lsp attributes BGP_TE5555

```

```

path-selection metric igp
pce

end

```

5. Enable PCE and auto-tunnel configurations on R1.

```

mpls traffic-eng tunnels
mpls traffic-eng pcc peer 10.0.0.3 source 10.0.0.4 precedence 255
mpls traffic-eng auto-tunnel p2p tunnel-num min 2000 max 5000

```

6. Enable all core links with SR-TE configurations and ensure that they are enabled as point to point interfaces.

```

mpls traffic-eng tunnels

interface GigabitEthernet0/2/0
ip address 101.102.6.1 255.255.255.0
ip router isis 1
mpls traffic-eng tunnels
isis network point-to-point

interface GigabitEthernet0/3/1
vrf forwarding sr
ip address 101.107.3.1 255.255.255.0
negotiation auto

end

```

7. Enable R3 (RR) to advertise TED to the PCE server via BGP-LS.

```

router isis 1
net 49.0002.0000.0000.0003.00
ispf level-1-2
metric-style wide
nsf cisco
nsf interval 0
distribute link-state
segment-routing mpls
segment-routing prefix-sid-map advertise-local
redistribute static ip level-1-2
mpls traffic-eng router-id Loopback0
mpls traffic-eng level-1
mpls traffic-eng level-2

router bgp 100
bgp router-id 10.0.0.2
bgp log-neighbor-changes
bgp graceful-restart
no bgp default ipv4-unicast
neighbor 10.0.0.3 remote-as 100
neighbor 10.0.0.3 update-source Loopback0

address-family ipv4
neighbor 10.0.0.3 activate
exit-address-family

```

8. Enable PCE server configuration and verify BGP-LS session is properly established with RR.

```

Device# sh bgp li li summary
BGP router identifier 10.0.0.3, local AS number 100
BGP generic scan interval 60 secs
Non-stop routing is enabled
BGP table state: Active
Table ID: 0x0 RD version: 1436

```

```

BGP main routing table version 1436
BGP NSR Initial initsync version 1 (Reached)
BGP NSR/ISSU Sync-Group versions 0/0
BGP scan interval 60 secs
BGP is operating in STANDALONE mode.
Process          RcvTblVer   bRIB/RIB   LabelVer   ImportVer   SendTblVer   StandbyVer
Speaker          1436       1436       1436       1436        1436         1436
0

Neighbor        Spk    AS MsgRcvd  MsgSent   TblVer   InQ   OutQ   Up/Down   St/PfxRcd
10.0.0.2         0      100  19923     17437    1436   0      0
1w2d             103

Device# sh pce ipv4 topo | b Node 3
Node 3
  TE router ID: 10.0.0.2
  Host name: R3
  ISIS system ID: 0000.0000.0003 level-1

  ISIS system ID: 0000.0000.0003 level-2
  Prefix SID:
    Prefix 10.0.0.2, label 20011 (regular)

```

Verifying Segment Routing On Demand Next Hop for L3/L3VPN

The ODN verifications are based on L3VPN VRF prefixes.

1. Verify that PCEP session between R1 (headend and PCE server) is established.

```

Device# sh pce client peer
PCC's peer database:
-----
Peer address: 10.0.0.3 (best PCE)
  State up
  Capabilities: Stateful, Update, Segment-Routing

```

2. Verify that PCEP session is established between all the peers (PCCs).

```

Device# sh pce ipv4 peer
PCE's peer database:
-----
Peer address: 10.0.0.4
  State: Up
  Capabilities: Stateful, Segment-Routing, Update
Peer address: 172.16.0.5
  State: Up
  Capabilities: Stateful, Segment-Routing, Update

```

3. Verify that R1 (headend) has no visibility to R6 loopback address.

```

Device# sh ip route 192.168.0.1
% Network not in table

```

4. Verify that VRF prefix is injected via MP-BGP in R1 VRF SR routing table.

```

Device# sh ip route vrf sr
Routing Table: sr
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route

```

```

o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
a - application route
+ - replicated route, % - next hop override, p - overrides from PFR
Gateway of last resort is not set
  10.0.0.6/8 is variably subnetted, 2 subnets, 2 masks
C    10.0.0.7/24 is directly connected, GigabitEthernet0/3/1
L    10.0.0.7/32 is directly connected, GigabitEthernet0/3/1
    10.0.0.8/24 is subnetted, 1 subnets
B    10.0.0.9 [200/0] via binding label: 865, 4d21h

```

5. Verify that BGP is associating properly the policy and binding SID with the VRF prefix.

```

Device# sh ip bgp vpv4 vrf sr 106.107.4.0
BGP routing table entry for 100:100:106.107.4.0/24, version 3011
Paths: (1 available, best #1, table sr)
  Not advertised to any peer
  Refresh Epoch 4
  Local
    192.168.0.1 (metric 10) (via default) from 10.0.0.2 (10.0.0.2)
      Origin incomplete, metric 0, localpref 100, valid, internal, best
      Community: 3276850
      Extended Community: RT:100:100
      Originator: 192.168.0.1, Cluster list: 10.0.0.2
      mpls labels in/out nlabel/1085
      binding SID: 865 (BGP_TE5555)
      rx pathid: 0, tx pathid: 0x0

```

6. Verify binding label association with VRF prefix.

```

Device# sh ip route vrf sr 106.107.4.0
Routing Table: sr
Routing entry for 106.107.4.0/24
  Known via "bgp 100", distance 200, metric 0, type internal
  Routing Descriptor Blocks:
  * Binding Label: 865, from 10.0.0.2, 4d22h ago
    Route metric is 0, traffic share count is 1
    AS Hops 0
    MPLS label: 1085
    MPLS Flags: NSF

```

7. Verify that VRF prefix is forwarded via ODN auto-tunnel.

```

Device# sh ip cef label-table
Label          Next Hop          Interface
0              no route
865           attached         Tunnel2000

Device# sh ip cef vrf sr 106.107.4.0 detail
10.0.0.8/24, epoch 15, flags [rib defined all labels]
  recursive via 865 label 1085
  attached to Tunnel2000

```

8. Verify ODN auto-tunnel status.

```

Device# sh mpls traffic-eng tunnels
P2P TUNNELS/LSPs:
Name: R1_t2000 (Tunnel2000) Destination: 192.168.0.1 Ifhandle: 0x6F5
(auto-tunnel for BGP TE)
Status:
  Admin: up      Oper: up      Path: valid      Signalling: connected---□
auto-tunnel 2000
  path option 1, (SEGMENT-ROUTING) (PCE) type dynamic (Basis for Setup, path weight 10)
Config Parameters:
  Bandwidth: 0 kbps (Global) Priority: 7 7 Affinity: 0x0/0xFFFF
  Metric Type: IGP (interface)

```

```

Path Selection:
  Protection: any (default)
Path-selection Tiebreaker:
  Global: not set   Tunnel Specific: not set   Effective: min-fill (default)
Hop Limit: disabled
Cost Limit: disabled
Path-invalidation timeout: 10000 msec (default), Action: Tear
AutoRoute: disabled LockDown: disabled Loadshare: 0 [0] bw-based
auto-bw: disabled
Attribute-set: BGP_TE5555--- attribute-set
Fault-OAM: disabled, Wrap-Protection: disabled, Wrap-Capable: No
Active Path Option Parameters:
  State: dynamic path option 1 is active
  BandwidthOverride: disabled LockDown: disabled Verbatim: disabled
PCEP Info:
  Delegation state: Working: yes   Protect: no
Working Path Info:
  Request status: processed
  Created via PCRep message from PCE server: 10.0.0.3--- via PCE server
  PCE metric: 30, type: IGP
Reported paths:
  Tunnel Name: Tunnel2000_w
  LSPs:
  LSP[0]:
    source 10.0.0.4, destination 192.168.0.1, tunnel ID 2000, LSP ID 1
    State: Admin up, Operation active
    Binding SID: 865
    Setup type: SR
    Bandwidth: requested 0, used 0
    LSP object:
      PLSP-ID 0x807D0, flags: D:0 S:0 R:0 A:1 O:2
    Metric type: IGP, Accumulated Metric 0
    ERO:
      SID[0]: Adj, Label 2377, NAI: local 101.102.6.1 remote 10.0.0.10
      SID[1]: Unspecified, Label 17, NAI: n/a
      SID[2]: Unspecified, Label 20, NAI: n/a
History:
  Tunnel:
    Time since created: 4 days, 22 hours, 21 minutes
    Time since path change: 4 days, 22 hours, 21 minutes
    Number of LSP IDs (Tun_Instances) used: 1
    Current LSP: [ID: 1]
    Uptime: 4 days, 22 hours, 21 minutes
  Tun_Instance: 1
  Segment-Routing Path Info (isis level-1)
    Segment0[Link]: 101.102.6.1 - 10.0.0.10, Label: 2377
    Segment1[ - ]: Label: 17
    Segment2[ - ]: Label: 20

```

9. Verify ODN auto-tunnel LSP status on R1 (headend).

```

Device# sh pce client lsp brief
PCC's tunnel database:
-----
Tunnel Name: Tunnel2000_w
  LSP ID 1
Tunnel Name: Tunnel2000_p

R1# sh pce client lsp detail
PCC's tunnel database:
-----
Tunnel Name: Tunnel2000_w
LSPs:
LSP[0]:
  source 10.0.0.4, destination 192.168.0.1, tunnel ID 2000, LSP ID 1

```



```

State: Admin up, Operation active
Binding SID: 865
Setup type: SR
Bandwidth: requested 0, used 0
LSP object:
  PLSP-ID 0x807D0, flags: D:0 S:0 R:0 A:1 O:2
Metric type: IGP, Accumulated Metric 0
ERO:
  SID[0]: Adj, Label 2377, NAI: local 101.102.6.1 remote 10.0.0.10
  SID[1]: Unspecified, Label 17, NAI: n/a
  SID[2]: Unspecified, Label 20, NAI: n/a

```

10. Verify ODN LSP status on the PCE server.

```

Device# sh pce lsp summ

PCE's LSP database summary:
-----
All peers:
Number of LSPs:          1
Operational: Up:         1 Down:          0
Admin state: Up:         1 Down:          0
Setup type: RSVP:        0 Segment routing: 1

Peer 10.0.0.4:
Number of LSPs:          1
Operational: Up:         1 Down:          0
Admin state: Up:         1 Down:          0
Setup type: RSVP:        0 Segment routing: 1

```

11. Verify detailed LSP information on the PCE server.

```

Device# sh pce lsp det
PCE's tunnel database:
-----
PCC 10.0.0.4:
Tunnel Name: Tunnel2000_w
LSPs:
LSP[0]:
  source 10.0.0.4, destination 192.168.0.1, tunnel ID 2000, LSP ID 48
  State: Admin up, Operation active
  Binding SID: 872
  PCEP information:
    plsp-id 526288, flags: D:1 S:0 R:0 A:1 O:2
  Reported path:
    Metric type: IGP, Accumulated Metric 0
    SID[0]: Adj, Label 885, Address: local 10.0.0.9 remote 10.0.0.10
    SID[1]: Unknown, Label 17,
    SID[2]: Unknown, Label 20,
  Computed path:
    Computed Time: Tue Dec 20 13:12:57 2016 (00:11:53 ago)
    Metric type: IGP, Accumulated Metric 30
    SID[0]: Adj, Label 885, Address: local 10.0.0.9 remote 10.0.0.10
    SID[1]: Adj, Label 17, Address: local 10.0.0.12 remote 10.0.0.13
    SID[2]: Adj, Label 20, Address: local 10.0.0.14 remote 10.0.0.14
  Recorded path:
    None

```

12. Shutdown the interface that is connected to VRF SR so that the prefix is no longer advertised by MP-BGP.

```

Device# int gig0/2/2
Device(config-if)#shut

```

13. Verify that VRF prefix is no longer advertised to R1 (headend) via R6 (tailend).

```

Device# sh ip route vrf sr
Routing Table: sr
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       a - application route
       + - replicated route, % - next hop override, p - overrides from PfR
Gateway of last resort is not set
  10.0.0.6/8 is variably subnetted, 2 subnets, 2 masks
C       10.0.0.7/24 is directly connected, GigabitEthernet0/3/1
L       10.0.0.8/32 is directly connected, GigabitEthernet0/3/1

```

14. Verify that no ODN auto-tunnel exists.

```

Device# sh mpls traffic-eng tunnels
P2P TUNNELS/LSPs:
P2MP TUNNELS:
P2MP SUB-LSPS:

```

Additional References for Segment Routing On Demand Next Hop for L3/L3VPN

Related Documents

Related Topic	Document Title
Cisco IOS Commands	Cisco IOS Master Command List, All Releases

Feature Information for Segment Routing On Demand Next Hop for L3/L3VPN

The following table provides release information about the feature or features described in this module. This table lists only the software release that introduced support for a given feature in a given software release train. Unless noted otherwise, subsequent releases of that software release train also support that feature.

Use Cisco Feature Navigator to find information about platform support and Cisco software image support. To access Cisco Feature Navigator, go to www.cisco.com/go/cfn. An account on Cisco.com is not required.

Table 1: Feature Information for Segment Routing On Demand Next Hop for L3/L3VPN

Feature Name	Releases	Feature Information
Segment Routing On Demand Next Hop for L3/L3VPN	Cisco IOS XE Everest 16.5.1b	<p>On-Demand Next Hop (ODN) triggers delegation of computation of an end-to-end LSP to a PCE controller including constraints and policies without doing any redistribution.</p> <p>The following commands were introduced or modified:</p> <p>route-map BGP_TE_MAP permit, mpls traffic-eng tunnels, sh bgp li li summary, sh pce client peer, sh pce ipv4 peer, sh ip route vrf sr, sh ip bgp vpnv4 vrf sr, sh ip cef label-table, sh mpls traffic-eng tunnels, sh pce client lsp brief, sh pce lsp summ, sh pce lsp det, routing-default-optimize</p>

