The bridge to possible

Troubleshooting Segment Routing

Luc De Ghein – Technical Leader

BRKSPG-3624

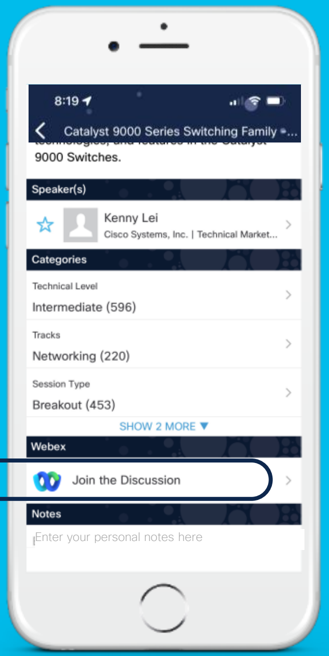
Cisco Webex App

Questions?

Use Cisco Webex App to chat
with the speaker after the session

How

1. Find this session in the Cisco Live Mobile App
2. Click “Join the Discussion”
3. Install the Webex App or go directly to the Webex space
4. Enter messages/questions in the Webex space



8:19

Catalyst 9000 Series Switching Family •...

9000 Switches.

Speaker(s)

Kenny Lei
Cisco Systems, Inc. | Technical Market...

Categories

Technical Level
Intermediate (596)

Tracks
Networking (220)

Session Type
Breakout (453)

SHOW 2 MORE ▼

Webex

Join the Discussion

Notes

Enter your personal notes here

Webex spaces will be moderated
until February 24, 2023.

Agenda

- Introduction
- Segment Routing (SR) Recap
- Troubleshooting Control Plane
- Troubleshooting Data Plane
- SR Policy
- SR PCE
- Tree-SID
- SR OAM
- Key Takeaways

Before We Get Started

- Fair basic knowledge on SR is required
- MPLS and IPv6 in data plane
  - This presentation mostly covers MPLS; there is some SRv6
- All is IOS-XR
  - Latest and greatest
  - Similar in IOS(-XE)
- Stay up-to-date http://www.segment-routing.net/

SR Recap

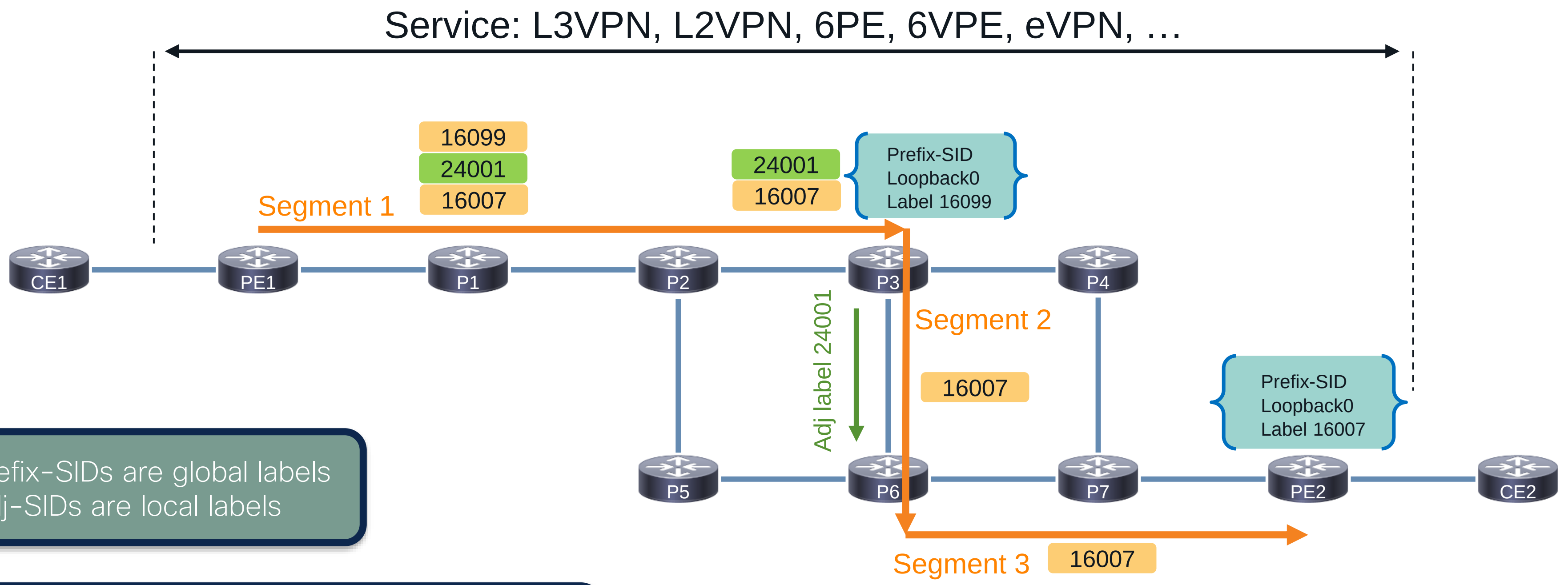
Introduction

- A segment is an instruction
  - with MPLS forwarding: segment = label
- Forwarding is done by MPLS or IPv6
- Link-state protocol is needed to advertise
- Segments (Prefix-SID, Adjacency-SID)
- MPLS Label
- Removing the signaling and state (no LDP/no RSVP-TE)
- Controller/SDN can be used if/when needed

*SID = Segment Identifier

SR Overview - Basic

24001	→	Adj-SID label
16007	→	Prefix-Sid label



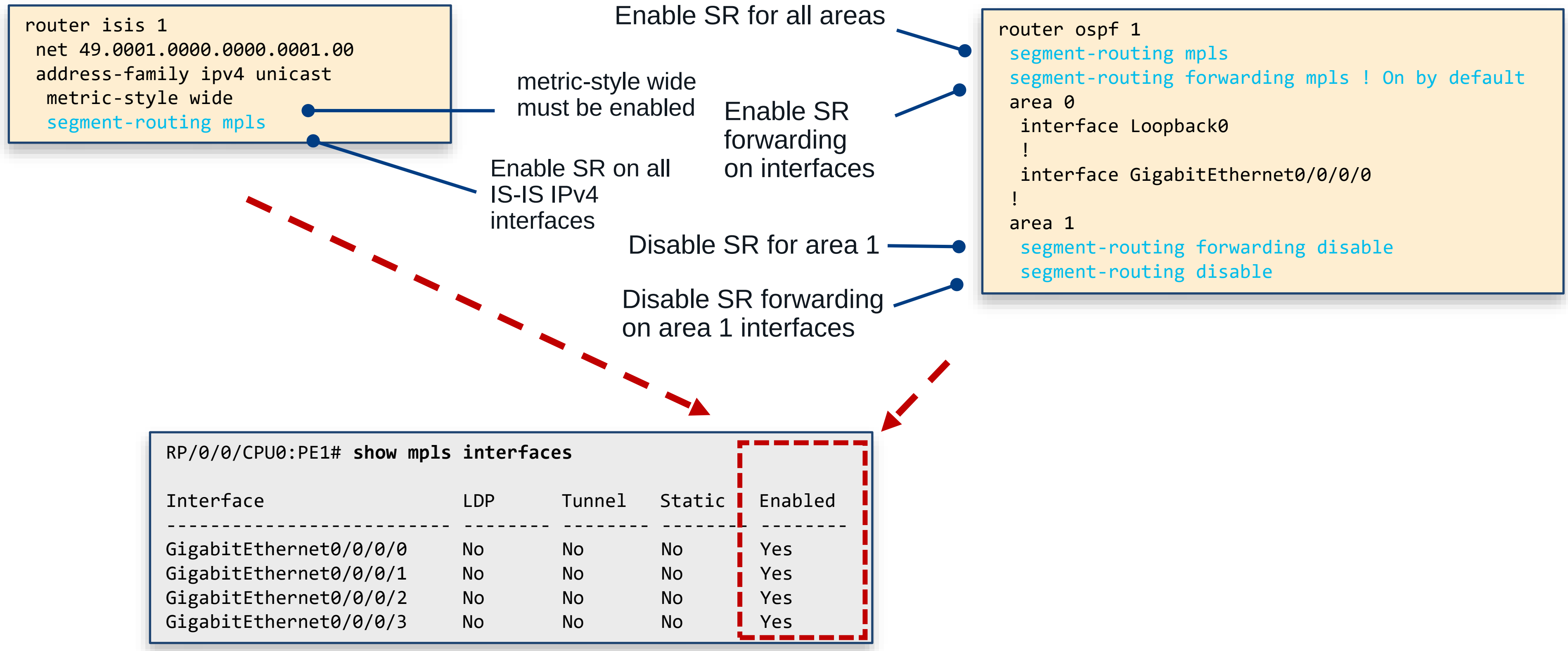
Prefix-SIDs are global labels
Adj-SIDs are local labels

Deviate from shortest path – Source Routing:
Traffic Engineering based on SR

Default: PHP at each segment

Troubleshooting – Control Plane

Bringing up Segment Routing



```
router isis 1
 net 49.0001.0000.0000.0001.00
 address-family ipv4 unicast
  metric-style wide
  segment-routing mpls
```

metric-style wide must be enabled

Enable SR on all IS-IS IPv4 interfaces

Enable SR for all areas

Enable SR forwarding on interfaces

Disable SR for area 1

Disable SR forwarding on area 1 interfaces

```
router ospf 1
 segment-routing mpls
 segment-routing forwarding mpls ! On by default
 area 0
  interface Loopback0
  !
  interface GigabitEthernet0/0/0/0
 !
 area 1
  segment-routing forwarding disable
  segment-routing disable
```

```
RP/0/0/CPU0:PE1# show mpls interfaces

Interface                  LDP      Tunnel   Static   Enabled
-------------------------- -------- -------- -------- --------
GigabitEthernet0/0/0/0     No       No       No       Yes
GigabitEthernet0/0/0/1     No       No       No       Yes
GigabitEthernet0/0/0/2     No       No       No       Yes
GigabitEthernet0/0/0/3     No       No       No       Yes
```

Verify the Labels

```
RP/0/0/CPU0:PE1# show mpls label table detail

Table Label   Owner                           State  Rewrite
----- ------- ------------------------------- ------ -------
0     0       LSD(A)                          InUse  Yes
0     1       LSD(A)                          InUse  Yes
0     2       LSD(A)                          InUse  Yes
0     13      LSD(A)                          InUse  Yes
0     1000    Application-Controller(A):XTC   InUse  Yes
  (TE Binding, vers:0, identifier=3, type=2)
0     2000    Application-Controller(A):XTC   InUse  Yes
  (TE Binding, vers:0, identifier=4, type=2)
0     15000   LSD(A)                          InUse  No
  (Lbl-blk SRLB, vers:0, (start_label=15000, size=1000, app_notify=0)

0     16000   ISIS(A):1                       InUse  No
  (Lbl-blk SRGB, vers:0, (start_label=16000, size=8000)
0     24000   ISIS(A):1                       InUse  Yes
  (SR Adj Segment IPv4, vers:0, index=1, type=0, intf=Gi0/0/0/1, nh=10.1.6.6)
0     24001   ISIS(A):1                       InUse  Yes
  (SR Adj Segment IPv4, vers:0, index=3, type=0, intf=Gi0/0/0/1, nh=10.1.6.6)
0     24002   ISIS(A):1                       InUse  Yes
  (SR Adj Segment IPv4, vers:0, index=1, type=0, intf=Gi0/0/0/3, nh=10.1.5.6)
```

default SR Global Block (SRGB): 16,000 – 23,999 (size 8,000)

labels for SR policies

SR Local Block (SRLB): 15,000 – 15,999

Manual allocation of Adj-SIDs , binding SIDs, BGP peering SIDs

first Adj-SID label (from dynamic range)

```
RP/0/0/CPU0:PE1# show mpls label range
Range for dynamic labels: Min/Max: 24000/1048575
```

dynamic range

IGP – LS Database

```
RP/0/0/CPU0:PE1# show isis database level 2 verbose P1.00

IS-IS 1 (Level-2) Link State Database
LSPID                 LSP Seq Num  LSP Checksum  LSP Holdtime  ATT/P/OL
P1.00-00              0x00000017   0xa223        674              0/0/0
  Area Address: 49.0001
  NLPID:         0xcc
  Hostname:      P1
  IP Address:    10.100.1.5
  Router Cap:    10.100.1.5, D:0, S:0
    Segment Routing: I:1 V:0, SRGB Base: 16000 Range: 8000
  Metric: 10         IS-Extended PE1.00
    Interface IP Address: 10.1.2.5
    Neighbor IP Address: 10.1.2.1
    ADJ-SID: F:0 B:1 V:1 L:1 S:0 weight:0 Adjacency-sid:24000
    ADJ-SID: F:0 B:0 V:1 L:1 S:0 weight:0 Adjacency-sid:24001

 Metric: 0         IP-Extended 10.100.1.5/32
   Prefix-SID Index: 5, Algorithm:0, R:0 N:1 P:0 E:0 V:0 L:0
```

Regular LSP header

SR Capabilities

I:1 IPv4 capable
V:0 no IPv6 capable
SRGB block

Adj-SID

F:0 address family IPv4
B:1 eligible for Backup
V:1 adj SID has a Value
L:1 Local significance
S:0 Set of adjacencies
Weight: 0 amount of load balancing (NA yet)

Prefix-SID

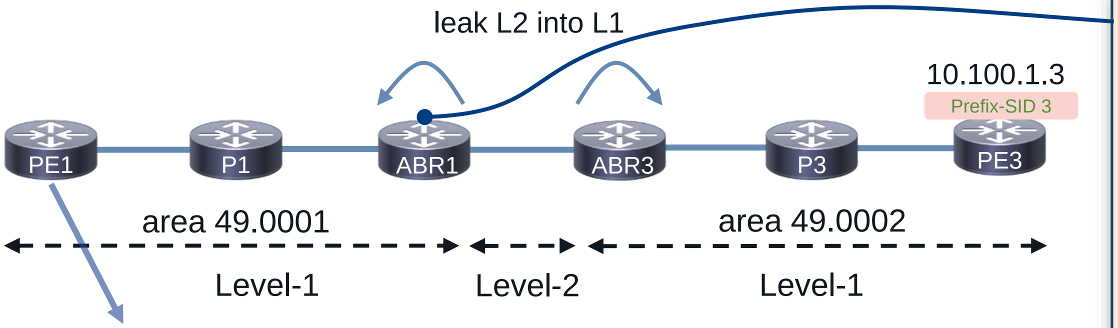
R:0 not Re-advertised
N:1 node SID
P:0 PHP on
E:0 explicit-NULL off
V:0 index
L:0 global significance

```
router isis 1
 interface Loopback0
  passive
  address-family ipv4 unicast
   prefix-sid index 5
```

```
router isis 1
 interface Loopback0
  passive
  address-family ipv4 unicast
   prefix-sid absolute 16005
```

*Prefix SID always advertised as relative index

Example of P and R Flags: Multi-Level ISIS



```
route-policy pass
  pass
end-policy
!
router isis 1
 net 49.0001.0000.0000.0009.00
 address-family ipv4 unicast
  metric-style wide
  mpls traffic-eng level-1-2
  router-id Loopback0
  propagate level 2 into level 1 route-policy pass
  segment-routing mpls
```

```
RP/0/0/CPU0:PE1# show isis database ABR1.00-00 verbose
..
 Metric: 40          IP-Extended-Interarea 10.100.1.3/32
    Prefix-SID Index: 3, Algorithm:0, R:1 N:1 P:1 E:0 V:0 L:0
    Prefix Attribute Flags: X:0 R:1 N:1
    Source Router ID: 10.100.1.3
…
```

```
RP/0/0/CPU0:PE1# trace 10.100.1.3

 1  10.1.15.5 [MPLS: Label 16003 Exp 0] 29 msec  29 msec  39 msec
 2  10.1.59.9 [MPLS: Label 16003 Exp 0] 29 msec  29 msec  39 msec
 3  10.1.129.12 [MPLS: Label 16003 Exp 0] 39 msec  49 msec  29 msec
 4  10.1.126.6 [MPLS: Label 16003 Exp 0] 19 msec  19 msec  39 msec
 5  10.1.36.3 29 msec  *  19 msec
```

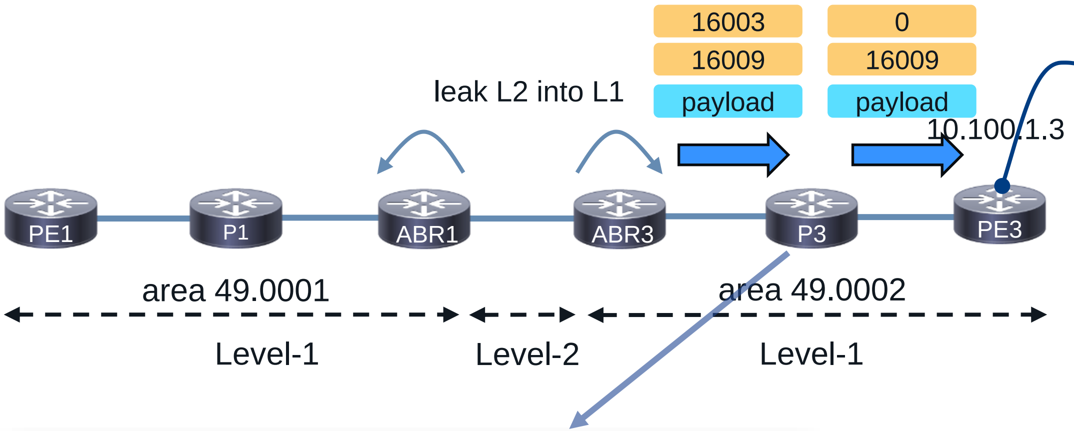
Prefix SID

R:1 Re-advertised
P:1 PHP off

The R-Flag MUST be set for prefixes that are not local to the router and either: advertised because of propagation (Level-1 into Level-2); advertised because of leaking (Level-2 into Level-1); advertised because of redistribution (e.g.: from another protocol).

no PHP at ABR = LSP is uninterrupted

Example of E Flag



```
router isis 1
 is-type level-1
 net 49.0002.0000.0000.0003.00
 address-family ipv4 unicast
  metric-style wide
  router-id Loopback0
  segment-routing mpls
 !
 interface Loopback0
  address-family ipv4 unicast
   prefix-sid index 3 explicit-null
```

```
RP/0/0/CPU0:P3# show isis database PE3.00-00 verbose
..
  Metric: 10          IP-Extended 10.100.1.3/32
    Prefix-SID Index: 3, Algorithm:0, R:0 N:1 P:1 E:1 V:0 L:0
    Prefix Attribute Flags: X:0 R:0 N:1
    Source Router ID: 10.100.1.3
```

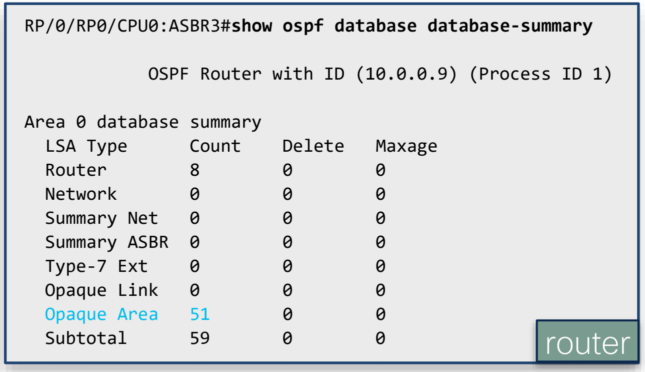
Prefix SID

E:1 explicit-NULL on

```
RP/0/0/CPU0:P3# show mpls forwarding labels 16003

Local  Outgoing    Prefix             Outgoing     Next Hop        Bytes
Label  Label       or ID              Interface                    Switched
------ ----------- ------------------ ------------ --------------- ------------
16003  Exp-Null-v4 SR Pfx (idx 3)     Gi0/0/0/1    10.1.36.3       10721
```

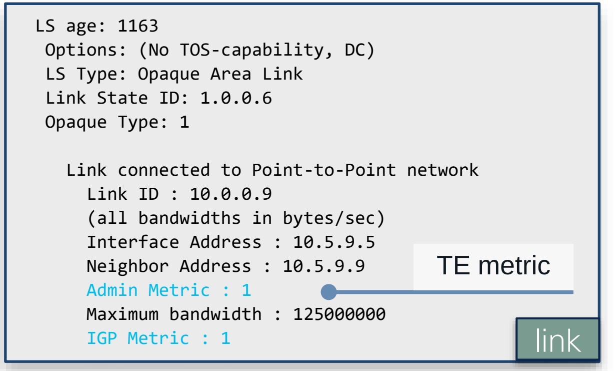
LFIB on previous router shows explicit-null label

OSPF LSAs



```
RP/0/RP0/CPU0:ASBR3#show ospf database database-summary

            OSPF Router with ID (10.0.0.9) (Process ID 1)

Area 0 database summary
  LSA Type      Count    Delete   Maxage
  Router        8        0        0
  Network       0        0        0
  Summary Net   0        0        0
  Summary ASBR  0        0        0
  Type-7 Ext    0        0        0
  Opaque Link   0        0        0
  Opaque Area   51       0        0
  Subtotal      59       0        0
```

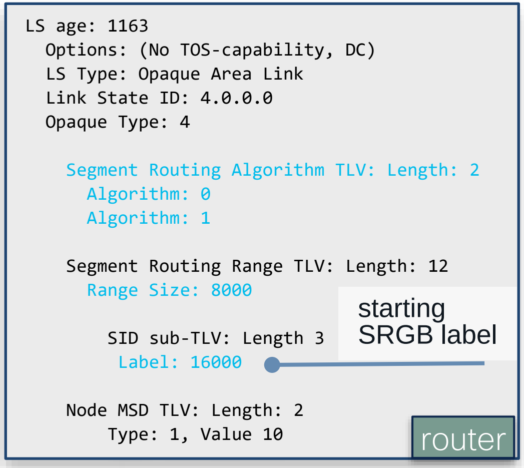
router



```
 LS age: 1163
  Options: (No TOS-capability, DC)
  LS Type: Opaque Area Link
  Link State ID: 1.0.0.6
  Opaque Type: 1

    Link connected to Point-to-Point network
      Link ID : 10.0.0.9
      (all bandwidths in bytes/sec)
      Interface Address : 10.5.9.5
      Neighbor Address : 10.5.9.9
      Admin Metric : 1
      Maximum bandwidth : 125000000
      IGP Metric : 1
```

TE metric

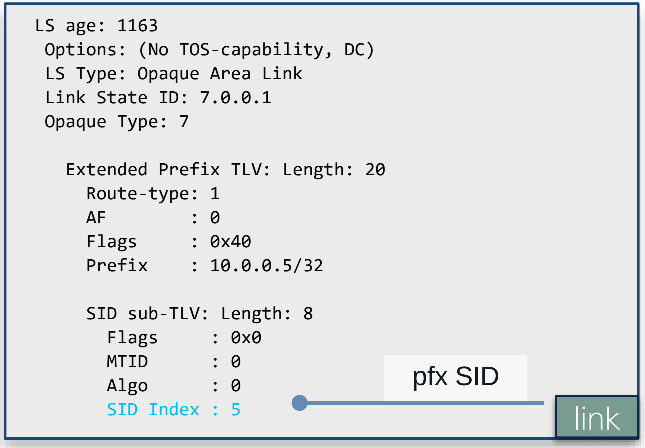
link



```
LS age: 1163
  Options: (No TOS-capability, DC)
  LS Type: Opaque Area Link
  Link State ID: 4.0.0.0
  Opaque Type: 4

    Segment Routing Algorithm TLV: Length: 2
      Algorithm: 0
      Algorithm: 1

    Segment Routing Range TLV: Length: 12
      Range Size: 8000

        SID sub-TLV: Length 3
          Label: 16000

    Node MSD TLV: Length: 2
        Type: 1, Value 10
```

starting SRGB label

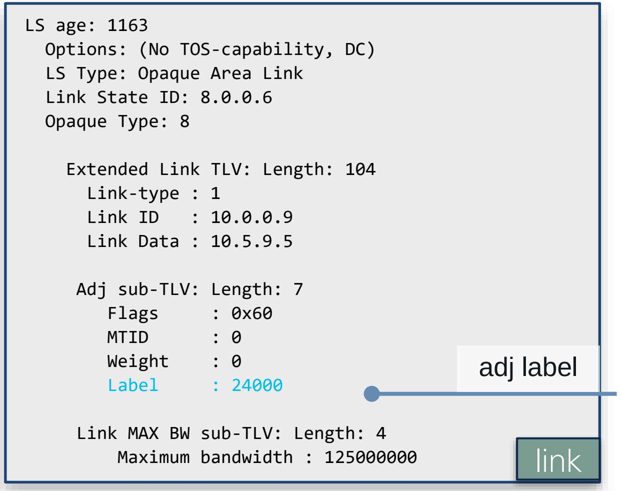
router



```
 LS age: 1163
  Options: (No TOS-capability, DC)
  LS Type: Opaque Area Link
  Link State ID: 7.0.0.1
  Opaque Type: 7

    Extended Prefix TLV: Length: 20
      Route-type: 1
      AF         : 0
      Flags      : 0x40
      Prefix     : 10.0.0.5/32

      SID sub-TLV: Length: 8
        Flags     : 0x0
        MTID      : 0
        Algo      : 0
        SID Index : 5
```

pfx SID

link



```
LS age: 1163
  Options: (No TOS-capability, DC)
  LS Type: Opaque Area Link
  Link State ID: 8.0.0.6
  Opaque Type: 8

    Extended Link TLV: Length: 104
      Link-type : 1
      Link ID   : 10.0.0.9
      Link Data : 10.5.9.5

     Adj sub-TLV: Length: 7
        Flags      : 0x60
        MTID       : 0
        Weight     : 0
        Label      : 24000

     Link MAX BW sub-TLV: Length: 4
         Maximum bandwidth : 125000000
```

adj label

link

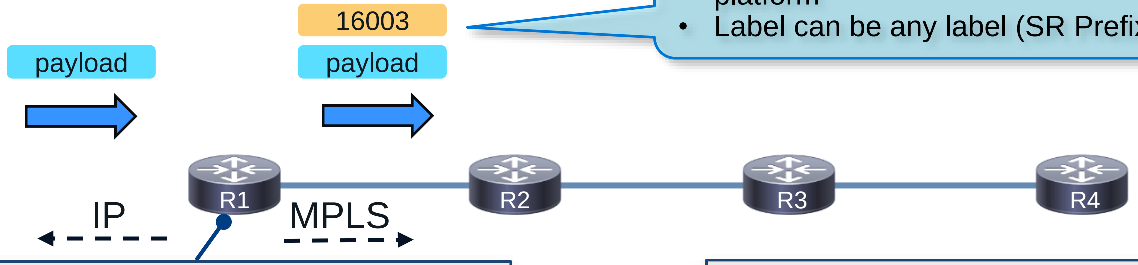
Troubleshooting – Data Plane

SR has Regular MPLS Data Plane

- SR has segments: 1 segment represented by 1 label in MPLS label stack
- SR uses the existing MPLS data plane
  - No exceptions
  - MPLS label operations:
    - Push, Pop, and Swap
  - We have
    - Special labels {0 - 15}
      - PHP (default behavior, also for SR)
      - Explicit-null for IPv4 and IPv6
    - Regular labels {16 - 1048575}
      - Static labels {16 - 4095}
      - SRGB {16000 - 23999} – Prefix-SIDs
      - Dynamic range {24000 - 1048575} – includes Adj-SIDs
    - QOS propagation (EXP bits)
      - Still uniform model, pipe, and short pipe model
    - TTL propagation as usual

MPLS Label Operation: Push Label(s)

- Push can occur at ingress of MPLS domain
  - MPLS label stack added in CEF (FIB) table
  - Top label is SR label; other labels can be service labels (e.g. MPLS VPN, BGP-LU, etc.)
- Push can occur at intermediate MPLS (P) router
  - MPLS label(s) added in LFIB (e.g. active FRR)



- Label stack added can be up to size supported by the platform
- Label can be any label (SR Prefix-SID, Adj-SID label, …)

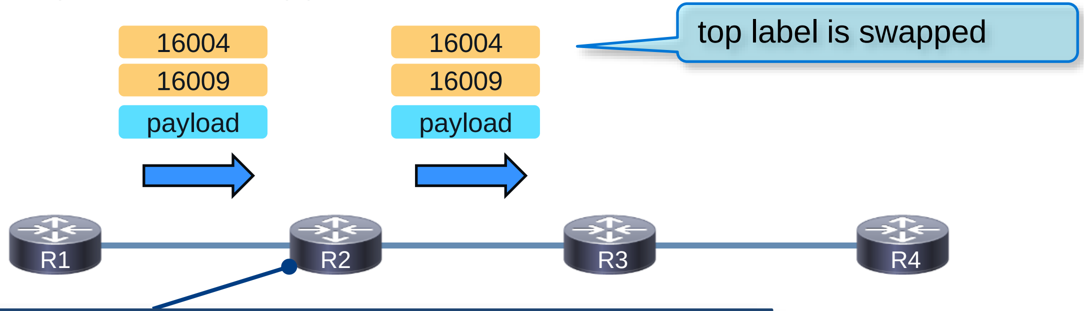
```
RP/0/0/CPU0:R1# show route 10.100.1.3/32

Routing entry for 10.100.1.3/32
  Known via "isis 1", … , labeled SR, …
   Routing Descriptor Blocks
    10.1.15.5, from 10.100.1.3, via GigabitEthernet0/0/0/0
      Route metric is 60
```

“labeled SR” must be present

```
RP/0/0/CPU0:R1# show cef …

10.100.1.3/32, …  labeled SR, …

 via 10.1.12.2/32, GigabitEthernet0/0/0/1,…
 local label 16003      labels imposed {16003}
```

MPLS Label Operation: Swap Label(s)

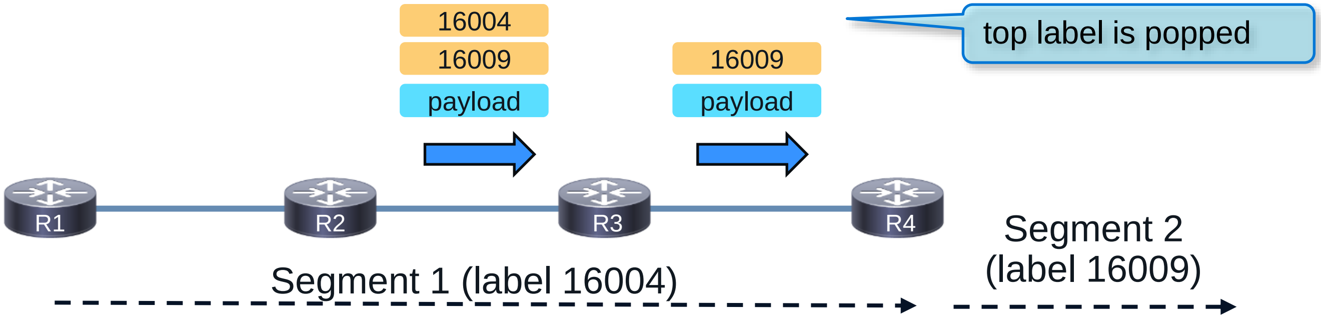
- Swap occurs at intermediate MPLS (P) router
  - Only top label is swapped
  - MPLS label is swapped in LFIB
  - Other labels are not touched (EXP bits, TTL)
  - Within one SR segment, top label is swapped with same label

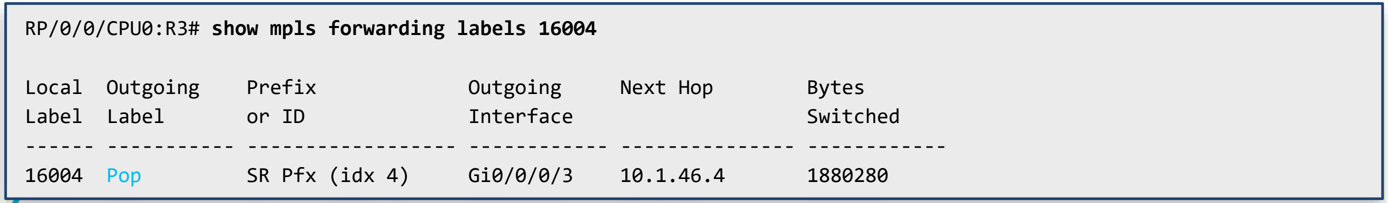


```
RP/0/0/CPU0:R2# show mpls forwarding labels 16004

Local  Outgoing    Prefix             Outgoing     Next Hop        Bytes
Label  Label       or ID              Interface                    Switched
------ ----------- ------------------ ------------ --------------- ------------
16004  16004       SR Pfx (idx 4)     Gi0/0/0/1    10.1.59.9       1420
```

MPLS Label Operation: Pop Label(s)

- Pop occurs at intermediate MPLS (P) router: top label is removed
- By default on penultimate router of one SR segment
  - Label stack could become unlabeled
  - Label stack can still have other labels
    - e.g. when packet is moved from one SR segment to another SR segment

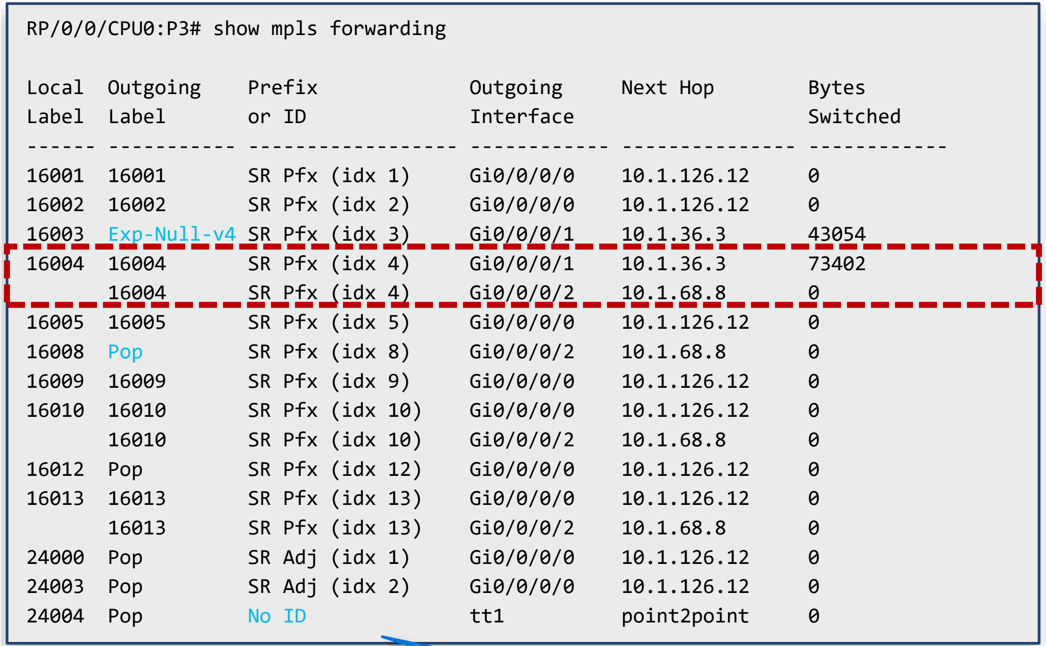




```
RP/0/0/CPU0:R3# show mpls forwarding labels 16004

Local  Outgoing    Prefix             Outgoing     Next Hop        Bytes
Label  Label       or ID              Interface                    Switched
------ ----------- ------------------ ------------ --------------- ------------
16004  Pop         SR Pfx (idx 4)     Gi0/0/0/3    10.1.46.4       1880280
```

LFIB

Nothing new!



```
RP/0/0/CPU0:P3# show mpls forwarding

Local  Outgoing    Prefix             Outgoing     Next Hop        Bytes
Label  Label       or ID              Interface                    Switched
------ ----------- ------------------ ------------ --------------- ------------
16001  16001       SR Pfx (idx 1)     Gi0/0/0/0    10.1.126.12     0
16002  16002       SR Pfx (idx 2)     Gi0/0/0/0    10.1.126.12     0
16003  Exp-Null-v4 SR Pfx (idx 3)     Gi0/0/0/1    10.1.36.3       43054
16004  16004       SR Pfx (idx 4)     Gi0/0/0/1    10.1.36.3       73402
       16004       SR Pfx (idx 4)     Gi0/0/0/2    10.1.68.8       0
16005  16005       SR Pfx (idx 5)     Gi0/0/0/0    10.1.126.12     0
16008  Pop         SR Pfx (idx 8)     Gi0/0/0/2    10.1.68.8       0
16009  16009       SR Pfx (idx 9)     Gi0/0/0/0    10.1.126.12     0
16010  16010       SR Pfx (idx 10)    Gi0/0/0/0    10.1.126.12     0
       16010       SR Pfx (idx 10)    Gi0/0/0/2    10.1.68.8       0
16012  Pop         SR Pfx (idx 12)    Gi0/0/0/0    10.1.126.12     0
16013  16013       SR Pfx (idx 13)    Gi0/0/0/0    10.1.126.12     0
       16013       SR Pfx (idx 13)    Gi0/0/0/2    10.1.68.8       0
24000  Pop         SR Adj (idx 1)     Gi0/0/0/0    10.1.126.12     0
24003  Pop         SR Adj (idx 2)     Gi0/0/0/0    10.1.126.12     0
24004  Pop         No ID              tt1          point2point     0
```

special labels {0-15} are still used

ECMP, can only be Prefix-SID

data plane makes no distinction between Prefix-SID and Adj-SID

any Adj-SID will have pop operation

Binding entry (used with SR-TE)

Load Balancing MPLS Traffic

Nothing new!

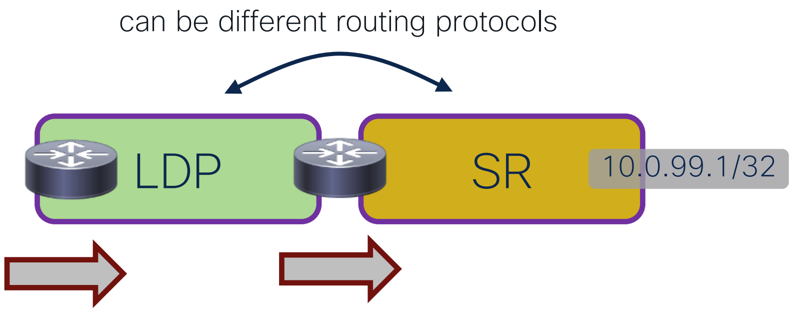
- Routers will try to load balance on the IP header, even when there is a label stack
  - IP traffic is best load-balanced by calculating hash over 3- or 7-tuple*
- PseudoWire traffic is load-balanced by calculating hash over the bottom label (PW service label)
  - Preserving per-flow load balancing
- Flow Aware Transport (FAT) Label can be used
  - Endpoints need to support this signaling
  - Endpoints classify traffic and pushes a unique flow label for each flow (each PW)
  - Load balancing on bottom (FAT) label
  - More granular than load balancing
- If many labels are present in label stack and the platform cannot look at/past last label, there is load balancing on higher label in the stack

Check platform!

* 3- Source IP, Destination IP, Router ID
7- Source IP, Destination IP, Router ID, Source port, Destination port, Protocol, Ingress interface handle

LDP to SR

LDP - SR Interworking



- Incoming label allocated by LDP for prefix 10.0.99.1/32
- Outgoing label: LDP label for 10.0.99.1/32

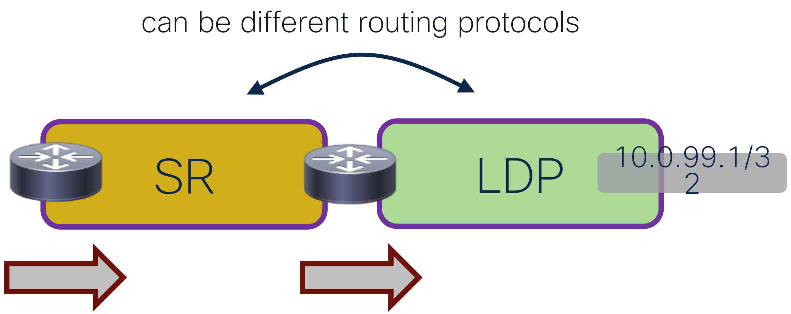
no further action needed

In Label	Out Label
24501	35311

- Incoming label allocated by LDP for prefix 10.0.99.1/32
- Outgoing label: prefix-SID label for 10.0.99.1/32

no further action needed

In Label	Out Label
24874	16066



- Incoming label ??
- Outgoing label: ??

solution needed

In Label	Out Label
??	??

- Incoming label: LDP label for 10.0.99.1/32
- Outgoing label: LDP label for 10.0.99.1/32

no further action needed

In Label	Out Label
24874	24635

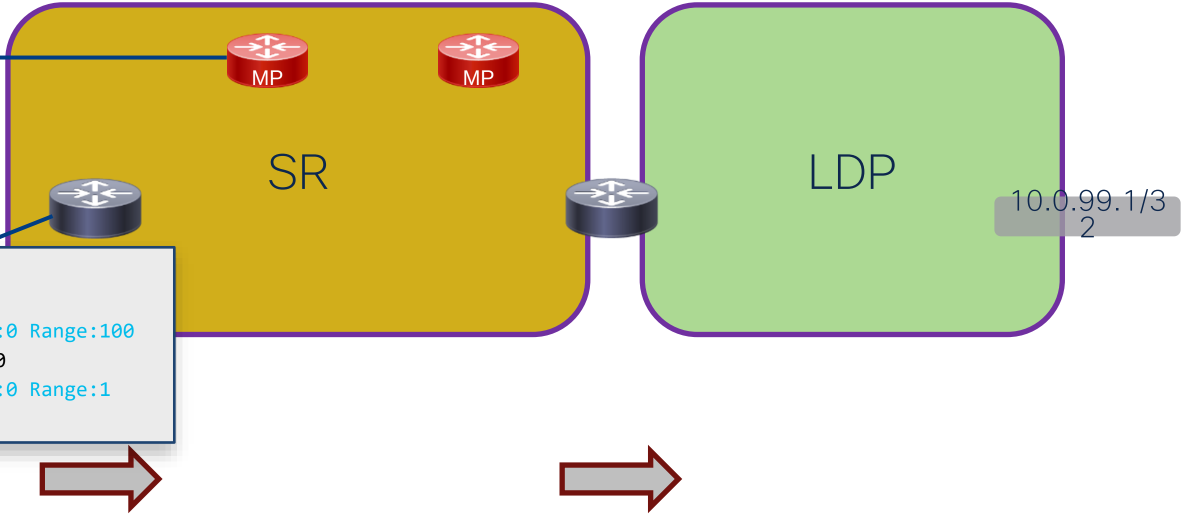
Mapping Servers

```
router isis 1
 net 49.0001.0000.0000.0008.00
 address-family ipv4 unicast
  metric-style wide
  segment-routing mpls
  segment-routing prefix-sid-map advertise-local
 !
segment-routing
 mapping-server
  prefix-sid-map
   address-family ipv4
    10.0.0.2/32 2
    10.0.99.1/32 1001 range 100
```

starting index

```
RP/0/RP0/CPU0:P1# show isis database MP.00-00 verbose

 SID Binding:    10.0.99.1/32 F:0 M:0 S:0 D:0 A:0 Weight:0 Range:100
    SID: Start:1001, Algorithm:0, R:0 N:0 P:0 E:0 V:0 L:0
  SID Binding:    10.0.0.2/32 F:0 M:0 S:0 D:0 A:0 Weight:0 Range:1
    SID: Start:2, Algorithm:0, R:0 N:0 P:0 E:0 V:0 L:0
```

mapping server (MS) advertises Prefix SIDs into SR domain for LDP prefixes



- Incoming label Prefix-SID label from MS Outgoing label: Prefix-SID label from MS

IN Label	Out Label
17001	17001

- Incoming label: prefix-SID label for 10.0.99.1/32
- Outgoing label: LDP label for 10.0.99.1/32

IN Label	Out Label
17001	24635

SR Policy

SR Policy

- An SR Policy is identified through the following tuple:
  - The head end where the policy is instantiated/implemented
  - The endpoint (i.e.: the destination of the policy)
  - The color (an arbitrary numerical value)
- At a given head end, an SR Policy is fully identified by the <color, endpoint> tuple
- An end point can be specified as an IPv4 or IPv6 address
- The result of path computation is always a list of segments, so SR labels
- Automated color-based steering
  - No Autoroute Announce, no ABF, no static routing
- PCE/PCC is possible (multi-domain)

SR-TE Configuration

```
segment-routing
 global-block 16000 23999
 traffic-eng
  logging
   policy status
  !
  segment-list explicit-to-ABR-1
   index 5 mpls adjacency 10.1.3.3
   index 10 mpls label 16007
   index 20 mpls label 16009
   index 30 mpls label 16005
  !
  policy to-ABR1
   binding-sid mpls 1000
   color 1000 end-point ipv4 10.0.0.5
   candidate-paths
    preference 100
     dynamic
      metric
       type igp
      !
     !
    !
    preference 200
     explicit segment-list explicit-to-ABR-1
```

SR-TE

Explicit path definition

SR-TE Policy path can be explicitly specified by configuring an ordered list of IPv4 or IPv6 addresses and/or label values

SR Policy

Dynamic path

candidate-paths with higher preference wins, if the path is valid

Explicit path

result of path computation is always a list of segments, so SR labels

first segment exception

If the first hop in the SR-TE Policy path is adjacency-SID or prefix-SID of adjacent node, then the label for this first hop is not added in SR-TE Policy rewrite label stack. This hop is only used to select outgoing interface(s).

cisco Live!

Verify SR Policy

```
RP/0/RP0/CPU0:PE1# show segment-routing traffic-eng policy name ?
  srte_c_1000_ep_10.0.0.5  Policy name (if contains space, enclose name with " ")
  WORD                     Policy name (if contains space, enclose name with " ")

RP/0/RP0/CPU0:PE1# show segment-routing traffic-eng policy name srte_c_1000_ep_10.0.0.5

SR-TE policy database
---------------------

Color: 1000, End-point: 10.0.0.5
  Name: srte_c_1000_ep_10.0.0.5
  Status:
    Admin: up  Operational: up for 00:12:09 (since Apr 30 06:20:19.182)
  Candidate-paths:
    Preference: 200 (configuration) (active)
      Name: to-ABR1
      Requested BSID: 1000
      Explicit: segment-list explicit-to-ABR-1 (valid)
        Weight: 1, Metric Type: TE
          24001 [Adjacency-SID, 10.1.3.1 - 10.1.3.3]
          16007
          16009
          16005
    Preference: 100 (configuration)
      Name: to-ABR1
      Requested BSID: 1000
        Maximum SID Depth: 10
      Dynamic (invalid)
        Metric Type: IGP,   Path Accumulated Metric: 20
  Attributes:
    Binding SID: 1000
    Forward Class: 0
    Steering labeled-services disabled: no
    Steering BGP disabled: no
    IPv6 caps enable: yes
```

lookup SR policy by name = *srte_color_endpoint*

CEF/FIB entry at the head end
Incoming label: 1000
Action: pop and push <16007, 16009, 16005>

SR-TE Database

```
RP/0/RP0/CPU0:PE1# show segment-routing traffic-eng ipv4 topology

SR-TE topology database
-----------------------

Node 1
  TE router ID: 10.0.0.1
  Host name: PE1
  ISIS system ID: 0000.0000.0001 level-1
  Prefix SID:
    Prefix 10.0.0.1, label 16001 (regular), flags: N:1, R:0, P:0, V:0, E:0, L:0

  Link[0]: local address 10.1.3.1, remote address 10.1.3.3
    Local node:
      ISIS system ID: 0000.0000.0001 level-1
    Remote node:
      Host name: P1
      ISIS system ID: 0000.0000.0003 level-1
    Metric: IGP 10, TE 10, Latency 10 microseconds
    Bandwidth: Total link 125000000, Reservable 0
    Admin-groups: 0x00000000
    Admin-groups-detail:
    Adj SID: 24001 (unprotected)

  Link[1]: local address 10.1.7.1, remote address 10.1.7.7
    Local node:
      ISIS system ID: 0000.0000.0001 level-1
    …
```

needed for TE topology

link entry

Adj-SID

```
router isis 1
 is-type level-1
 net 49.0001.0000.0000.0001.00
 distribute link-state
 address-family ipv4 unicast
  metric-style wide
  router-id Loopback0
  segment-routing mpls
 !
 interface Loopback0
  passive
  address-family ipv4 unicast
   prefix-sid index 1
  !
 !
 interface GigabitEthernet0/0/0/0
  point-to-point
  address-family ipv4 unicast
  !
```

Traffic To SR Policy: Automatic

```
segment-routing
 global-block 16000 23999
 traffic-eng
  logging
   policy status
  !
  segment-list explicit-to-PE2
   index 10 mpls label 16005
   index 20 mpls label 16002
  !
 policy to-PE2
   binding-sid mpls 2000
   color 2000 end-point ipv4 10.0.0.2
   candidate-paths
    preference 100
     explicit segment-list explicit-to-PE2
```

label 16005 to ABR

label 16002 to remote PE

next hop matches endpoint

prefix as received from remote PE

```
RP/0/RP0/CPU0:PE1# show bgp vpnv4 unicast vrf one 10.0.0.14/32

BGP routing table entry for 10.0.0.14/32, Route Distinguisher: 65000:1
Paths: (1 available, best #1)
  65002
    10.0.0.2 C:2000 (bsid:2000) (metric 20) from 10.0.0.12 (10.0.0.2)
      Received Label 24004
      Origin IGP, metric 0, localpref 100, valid, internal, best, group-
best, import-candidate, imported
      Received Path ID 0, Local Path ID 1, version 12
      Extended community: Color:2000 RT:65000:100
      Originator: 10.0.0.2, Cluster list: 10.0.0.12
      SR policy color 2000, up, not-registered, bsid 2000
```

```
RP/0/RP0/CPU0:PE1# show segment-routing traffic-eng policy color 2000

SR-TE policy database
---------------------
Color: 2000, End-point: 10.0.0.2
  Name: srte_c_2000_ep_10.0.0.2
  Status:
    Admin: up  Operational: up for 01:27:13 (since Apr 30 07:05:04.832)
  Candidate-paths:
    Preference: 100 (configuration) (active)
      Name: to-PE2
      Requested BSID: 2000
      Explicit: segment-list explicit-to-PE2 (valid)
        Weight: 1, Metric Type: TE
          16005 [Prefix-SID, 10.0.0.5]
          16002
```

matching color

Traffic To SR Policy: Automatic

```
RP/0/RP0/CPU0:PE1# show segment-routing traffic-eng policy color 2000

SR-TE policy database
---------------------
Color: 2000, End-point: 10.0.0.2
  Name: srte_c_2000_ep_10.0.0.2
  Status:
    Admin: up  Operational: up for 01:27:13 (since Apr 30
07:05:04.832)
  Candidate-paths:
    Preference: 100 (configuration) (active)
      Name: to-PE2
      Requested BSID: 2000
      Explicit: segment-list explicit-to-PE2 (valid)
        Weight: 1, Metric Type: TE
          16005 [Prefix-SID, 10.0.0.5]
          16002
```

```
RP/0/RP0/CPU0:PE1# show route vrf one 10.0.0.14/32

Routing entry for 10.0.0.14/32
  Known via "bgp 65000", distance 200, metric 0
   Routing Descriptor Blocks
    10.0.0.2, from 10.0.0.12
      Nexthop in Vrf: "default", Table: "default", IPv4 Unicast,
```

```
RP/0/RP0/CPU0:PE1# show cef vrf one 10.0.0.14/32

10.0.0.14/32, version 13, internal 0x5000001 0x0
 Prefix Len 32, traffic index 0, precedence n/a, priority 3
   via local-label 2000, 3 dependencies, recursive [flags 0x6000]
    path-idx 0 NHID 0x0 [0xd30fa50 0x0]
    recursion-via-label
    next hop VRF - 'default', table - 0xe0000000
    next hop via 2000/0/21
     next hop srte_c_2000_ labels imposed {ImplNull 24004}
```

next hop

labels used

or static

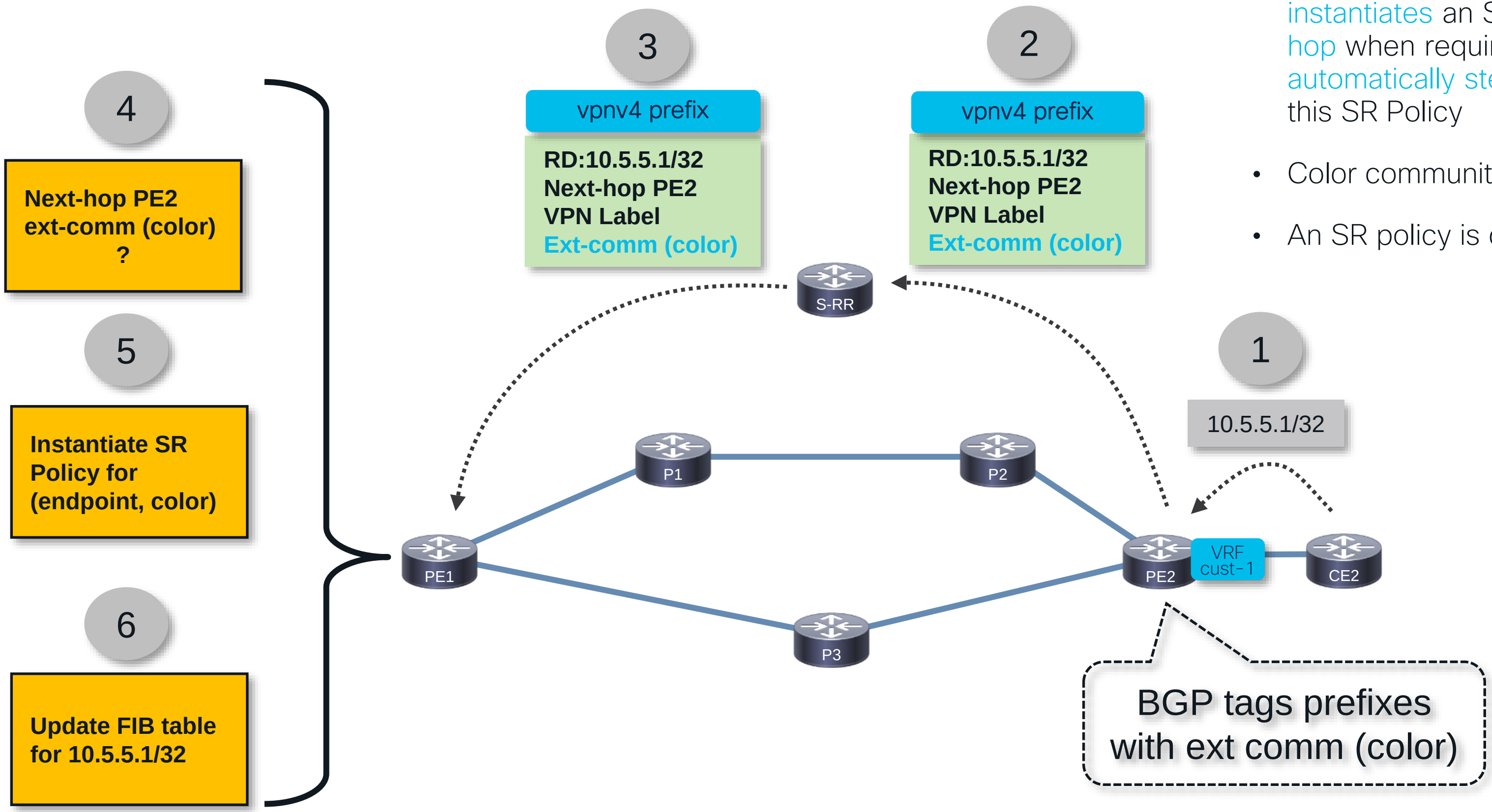
```
router static
!
 vrf one
  address-family ipv4 unicast
   10.0.0.14/32 sr-policy srte_c_2000_ep_10.0.0.2
```

```
RP/0/RP0/CPU0:PE1# traceroute vrf one 10.0.0.14

 1  10.1.3.3 [MPLS: Labels 16005/16002/24004 Exp 0] 12 msec  6 msec  9 msec
 2  10.3.5.5 [MPLS: Labels 16002/24004 Exp 0] 7 msec  6 msec  6 msec
 3  10.5.6.6 [MPLS: Labels 16002/24004 Exp 0] 9 msec  4 msec  7 msec
 4  10.4.6.4 [MPLS: Labels 16002/24004 Exp 0] 14 msec  60 msec  10 msec
 5  10.2.4.2 [MPLS: Label 24004 Exp 0] 8 msec  7 msec  4 msec
 6  10.2.14.14 5 msec  *  5 msec
```

inter-area traffic steering !

ODN (On-demand Next Hop)

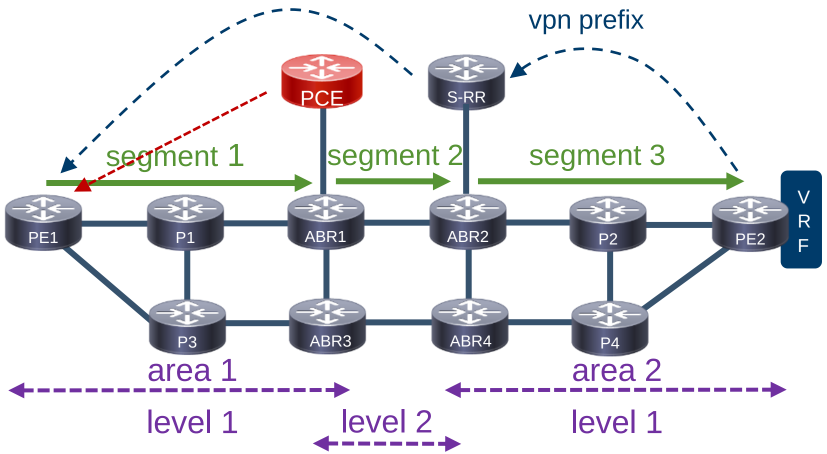
ODN Workflow



- A service head-end automatically instantiates an SR Policy to a BGP next hop when required (on-demand), automatically steering the BGP traffic into this SR Policy
- Color community is used as SLA indicator
- An SR policy is defined (color)

ODN Troubleshooting

- receive vpn prefix
- match the color
- next-hop is PE
- compute SR Policy
  by head end or by PCE



head end configuration (PE1)

```
segment-routing
 global-block 16000 23999
 traffic-eng
  logging
   policy status
  !
  on-demand color 100
   dynamic
    pcep
    !
    metric
     type igp
    !
   !
  !
  pcc
   source-address ipv4 10.0.0.1
   pce address ipv4 10.0.0.11
   !
```

ODN color 100

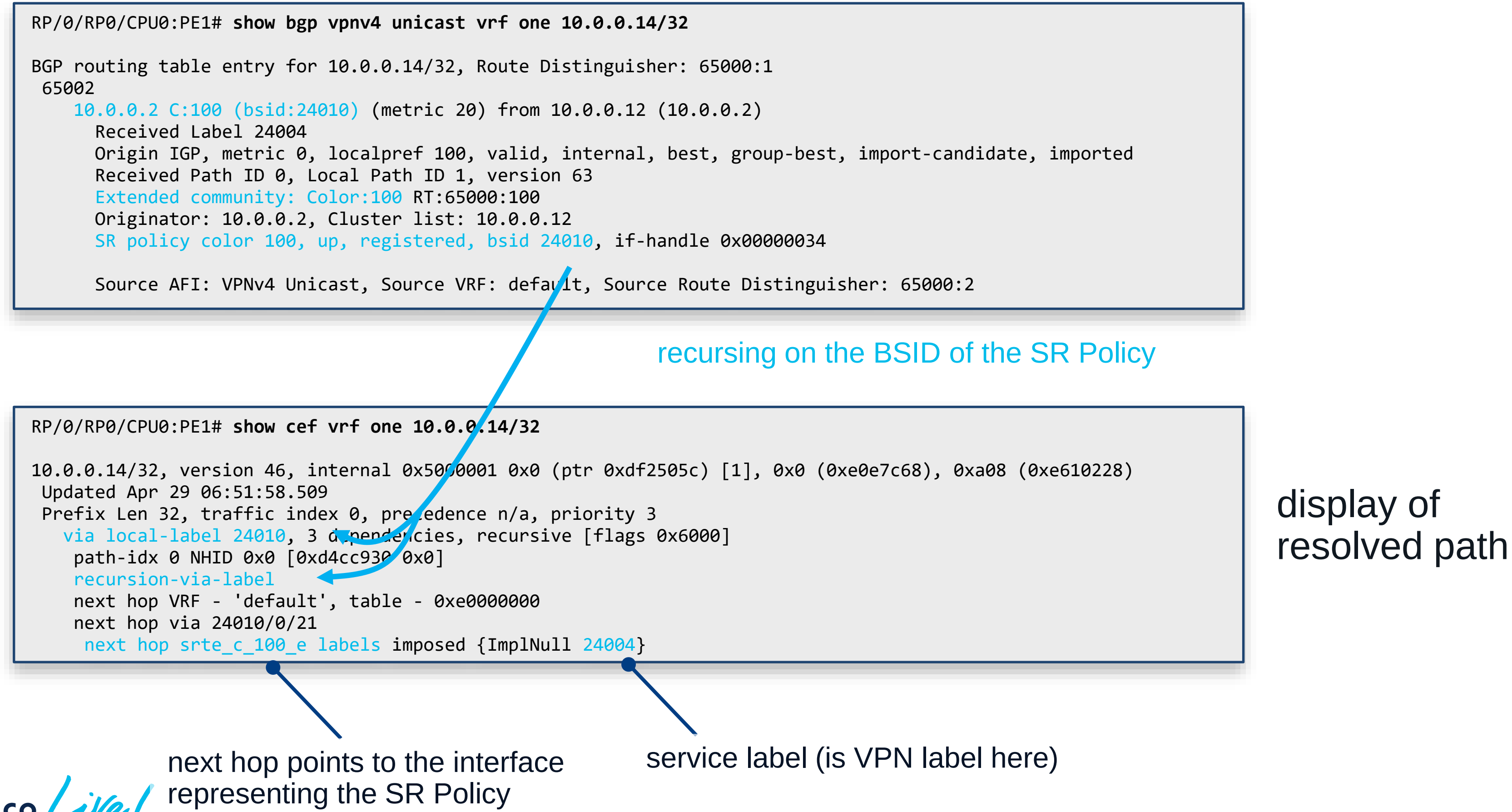
SR policy path calculated by PCE

RPL inbound to set color to green (100)

tail end configuration (PE2)

```
extcommunity-set opaque green
  100
!
route-policy ODN-CE2
  set extcommunity color green
!
router bgp 65000
!
 vrf one
  rd 65000:2
  address-family ipv4 unicast
  !
  neighbor 10.2.14.14
   remote-as 65002
   address-family ipv4 unicast
    route-policy ODN-CE2 in
    route-policy PASS out
```

Service Route



```
RP/0/RP0/CPU0:PE1# show bgp vpnv4 unicast vrf one 10.0.0.14/32

BGP routing table entry for 10.0.0.14/32, Route Distinguisher: 65000:1
 65002
    10.0.0.2 C:100 (bsid:24010) (metric 20) from 10.0.0.12 (10.0.0.2)
      Received Label 24004
      Origin IGP, metric 0, localpref 100, valid, internal, best, group-best, import-candidate, imported
      Received Path ID 0, Local Path ID 1, version 63
      Extended community: Color:100 RT:65000:100
      Originator: 10.0.0.2, Cluster list: 10.0.0.12
      SR policy color 100, up, registered, bsid 24010, if-handle 0x00000034

      Source AFI: VPNv4 Unicast, Source VRF: default, Source Route Distinguisher: 65000:2
```

recursing on the BSID of the SR Policy

```
RP/0/RP0/CPU0:PE1# show cef vrf one 10.0.0.14/32

10.0.0.14/32, version 46, internal 0x5000001 0x0 (ptr 0xdf2505c) [1], 0x0 (0xe0e7c68), 0xa08 (0xe610228)
 Updated Apr 29 06:51:58.509
 Prefix Len 32, traffic index 0, precedence n/a, priority 3
   via local-label 24010, 3 dependencies, recursive [flags 0x6000]
    path-idx 0 NHID 0x0 [0xd4cc930 0x0]
    recursion-via-label
    next hop VRF - 'default', table - 0xe0000000
    next hop via 24010/0/21
     next hop srte_c_100_e labels imposed {ImplNull 24004}
```

display of resolved path

next hop points to the interface representing the SR Policy

service label (is VPN label here)

SR Policy Path

head end (PE1)

```
RP/0/RP0/CPU0:PE1# show segment-routing traffic-eng policy color 100

SR-TE policy database
---------------------

Color: 100, End-point: 10.0.0.2
  Name: srte_c_100_ep_10.0.0.2
  Status:
    Admin: up  Operational: up for 01:26:32 (since Apr 29
06:03:25.176)
  Candidate-paths:
    Preference: 200 (BGP ODN) (shutdown)
      Requested BSID: dynamic
        Maximum SID Depth: 10
      Dynamic (invalid)
    Preference: 100 (BGP ODN) (active)
      Requested BSID: dynamic
      PCC info:
        Symbolic name: bgp_c_100_ep_10.0.0.2_discr_100
        PLSP-ID: 4
        Maximum SID Depth: 10
      Dynamic (pce 10.0.0.11) (valid)
        Metric Type: IGP,   Path Accumulated Metric: 50
          16005 [Prefix-SID, 10.0.0.5]
          16002 [Prefix-SID, 10.0.0.2]
  Attributes:
    Binding SID: 24010
    Forward Class: 0
    Steering labeled-services disabled: no
    Steering BGP disabled: no
    IPv6 caps enable: yes
```

engineered path per SR Policy (PCE)

PCE

```
RP/0/RP0/CPU0:PCE# show pce lsp detail

PCE's tunnel database:
----------------------
PCC 10.0.0.1:

Tunnel Name: bgp_c_100_ep_10.0.0.2_discr_100
 LSPs:
  LSP[0]:
   source 10.0.0.1, destination 10.0.0.2, tunnel ID 4, LSP ID 1
   State: Admin up, Operation up
   Setup type: Segment Routing
   Binding SID: 24010
   Maximum SID Depth: 10
   Bandwidth: signaled 0 kbps, applied 0 kbps
   PCEP information:
     PLSP-ID 0x4, flags: D:1 S:0 R:0 A:1 O:1 C:0
   LSP Role: Single LSP
   State-sync PCE: None
   PCC: 10.0.0.1
   LSP is subdelegated to: None
   Reported path:
     Metric type: IGP, Accumulated Metric 50
      SID[0]: Node, Label 16005, Address 10.0.0.5
      SID[1]: Node, Label 16002, Address 10.0.0.2
   Computed path: (Local PCE)
     Computed Time: Mon Apr 29 07:33:27 UTC 2019 (00:02:54 ago)
     Metric type: IGP, Accumulated Metric 50
      SID[0]: Node, Label 16005, Address 10.0.0.5
      SID[1]: Node, Label 16002, Address 10.0.0.2
```

```
CE1# trace 10.0.0.14 source loopback 0 numeric
Tracing the route to 10.0.0.14
VRF info: (vrf in name/id, vrf out name/id)
  1 10.1.13.1 2 msec 1 msec 1 msec
  2 10.1.3.3 [MPLS: Labels 16005/16002/24004 Exp 0] 14 msec 8 msec 9 msec
  3 10.3.5.5 [MPLS: Labels 16002/24004 Exp 0] 7 msec 8 msec 8 msec
  4 10.5.6.6 [MPLS: Labels 16002/24004 Exp 0] 7 msec 9 msec 8 msec
  5 10.4.6.4 [MPLS: Labels 16002/24004 Exp 0] 25 msec 8 msec 8 msec
  6 10.2.4.2 [MPLS: Label 24004 Exp 0] 41 msec 9 msec 12 msec
  7 10.2.14.14 6 msec *  8 msec
```

16005 is label to ABR1
16002 is label to PE2

BGP Soft Next Hop for ODN

- We relied on any route to be available to the BGP next-hop for the BGP service route
  - Next-hop route is not there for Seamless MPLS or Inter-AS
- Typical trick: configure static non-default route to Null0

```
router static
 address-family ipv4 unicast
  10.0.0.0/24 Null0
```

- This is the main command: it turns on the BGP soft next hop behavior. The RIB validation is not performed if there is an SR policy that is up for the next hop and color.

```
RP/0/RP0/CPU0:PE1(config-bgp)# nexthop validation color-extcomm sr-policy
```

- Instruct BGP to use the SR policy path metric instead of the IGP metric for the BGP best-path selection algorithm: consideration of the PCE/path admin and metric values.

```
RP/0/RP0/CPU0:PE1(config-bgp)# bgp bestpath igp-metric sr-policy
```

- More, much more here: Deploy BGP Soft Next-Hop in Cisco IOS XR

BGP soft next hop: you need both commands!

SR Policy Not Functional

- Inactive policy
  - No valid path found
- Invalid path
  - A Path is invalid as soon as it has no valid SID list
- Invalid SID list
  - It is empty
  - The headend is unable to resolve the first SID into one or more outgoing interface(s) and next hop(s)
  - The headend is unable to resolve any non-first SID expressed as an IP address
- Unreachable
  - The headend has no path to the SID in its SR-TE database

Most common: topology issues: check topology on router or PCE

```
RP/0/0/CPU0:PCE# show pce ipv4 topology | utility egrep -A5 -B5 10.100.1.13

  Link[2]: local address 10.1.113.10, remote address 10.1.113.13
    Local node:
      OSPF router ID: 10.100.1.10 area ID: 0
    Remote node:
      OSPF router ID: 10.100.1.13 area ID: 0
    Metric: IGP 1, TE 1
    Bandwidth: Total link 125000000, Reservable 0
    Adj SID: 24002 (protected) 24003 (unprotected)
    Excluded from CSPF: no
..
  Link[0]: local address 10.1.123.12, remote address 10.1.123.13
    Local node:
      OSPF router ID: 10.100.1.12 area ID: 0
    Remote node:
      OSPF router ID: 10.100.1.13 area ID: 0
    Metric: IGP 1, TE 1
    Bandwidth: Total link 125000000, Reservable 0
    Adj SID: 24002 (protected) 24003 (unprotected)
    Excluded from CSPF: no
```

The head end of an SR Policy updates the validity of a SID list upon network topological change.

Ti-LFA

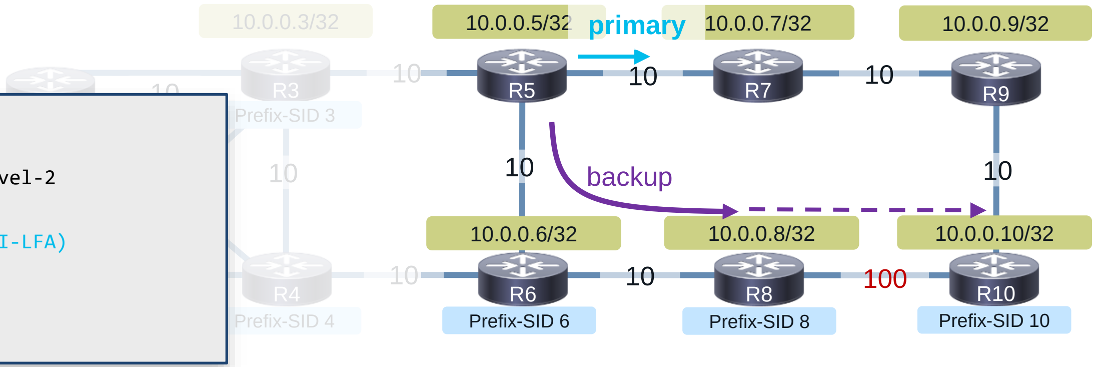
Topology Independent LFA (Ti-LFA)

- Loop Free Alternate (LFA), but topology independent
  - 100% coverage
  - Using segments to force traffic over backup path
  - Protected traffic is on Post-Convergence (PC) path
- Algorithm to calculate backup paths for IP (and MPLS)
- No signaling, link or node protection, and other tiebreakers
- Available for: SR-MPLS, SRv6, Tree-SID

```
router isis 1
 is-type level-2-only
 net 49.0001.0000.0000.0001.00
 address-family ipv4 unicast
  metric-style wide
  segment-routing mpls
 !
 interface Loopback0
  address-family ipv4 unicast
   prefix-sid absolute 16001
  !
 !
 interface GigabitEthernet0/0/0/0
  address-family ipv4 unicast
   fast-reroute per-prefix
   fast-reroute per-prefix ti-lfa
```

```
router ospf 1
 router-id 10.100.1.1
 fast-reroute per-prefix
 fast-reroute per-prefix ti-lfa enable
 address-family ipv4 unicast
 area 0
  segment-routing forwarding mpls !! On by default
  segment-routing mpls
  interface Loopback0
   prefix-sid absolute 16001
  !
  interface GigabitEthernet0/0/0/0
   network point-to-point
```

Ti-LFA – Double-Segment Example



RIB

```
RP/0/RP0/CPU0:R5# show route 10.0.0.7/32

Routing entry for 10.0.0.7/32
  Known via "isis 1", distance 115, metric 10, labeled SR, type level-2
  Installed Jun 11 17:40:41.894 for 00:00:39
  Routing Descriptor Blocks
    10.5.6.6, from 10.0.0.7, via GigabitEthernet0/0/0/2, Backup (TI-LFA)
      Repair Node(s): 10.0.0.8, 10.0.0.10
      Route metric is 140
    10.5.7.7, from 10.0.0.7, via GigabitEthernet0/0/0/1, Protected
      Route metric is 10
  No advertising protos.
```

ISIS

```
RP/0/RP0/CPU0:R5# show isis fast-reroute 10.0.0.7/32

L2 10.0.0.7/32 [10/115]
     via 10.5.7.7, GigabitEthernet0/0/0/1, R7, SRGB Base: 16000, Weight: 0
       Backup path: TI-LFA (link), via 10.5.6.6, GigabitEthernet0/0/0/2 R6, SRGB Base: 16000, Weight: 0, Metric: 140
         P node: R8.00 [10.0.0.8], Label: 16008
         Q node: R10.00 [10.0.0.10], Label: 24003
         Prefix label: 16007
         Backup-src: R7.00
```

P

Q

two additional labels

FIB

```
RP/0/RP0/CPU0:R5# show cef 10.0.0.7/32
10.0.0.7/32, version 138, labeled SR
  Prefix Len 32, traffic index 0, precedence n/a, priority 1
   via 10.5.6.6/32, GigabitEthernet0/0/0/2, 16 dependencies, weight 0, class 0, backup (TI-LFA) [flags 0xb00]
   next hop 10.5.6.6/32, Repair Node(s): 10.0.0.8, 10.0.0.10
    remote adjacency
     local label 16007      labels imposed {16008 24003 16007}
   via 10.5.7.7/32, GigabitEthernet0/0/0/1, 16 dependencies, weight 0, class 0, protected [flags 0x400]
    path-idx 1 bkup-idx 0 NHID 0x0 [0xf24da50 0xf24dc10]
    next hop 10.5.7.7/32
     local label 16007      labels imposed {ImplNull}
```

repair nodes

P

Q

two additional labels

SR PCE

Path Computation Element

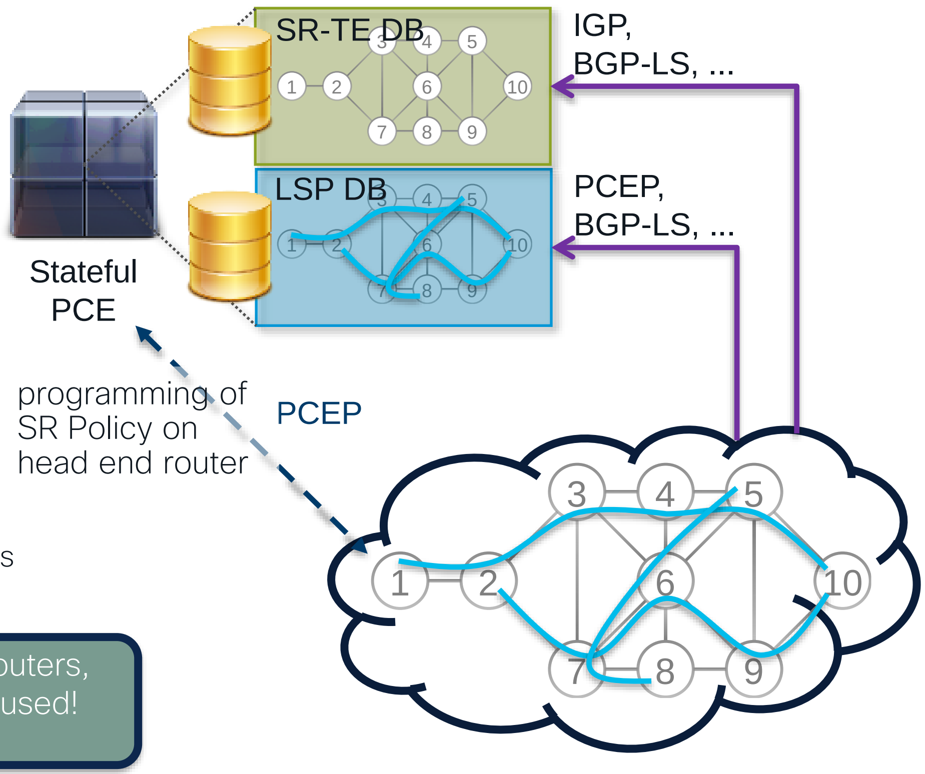
SR-TE DB = IGP Topology + TE link attributes + SIDs + SRGB

- PCE learns directly by participating in IGP or from BGP-LS

- PCE needed for inter-area, inter-AS
- Head end needs topology of all area's, domains

- PCCs need IP connectivity to PCE(s)
  - Important when doing inter-area/inter-AS
  - It's about getting all of the needed topology up to the PCE

- Direct vs BGP-LS
  - Direct = distribute IGP-LS info into BGP on PCE
  - BGP-LS = PCE has BGP-LS real-time feeds from key routers (usually RRs) in each area/AS/domain



programming of SR Policy on head end router

SR-TE needed on all intermediate routers, only if non-default TE attributes are used! No RSVP!

PCEP

1 PCReport

Requesting

End points constraints:

BW

Metric (IGP, TE, hop count)

Affinities

Priorities

RRO

…

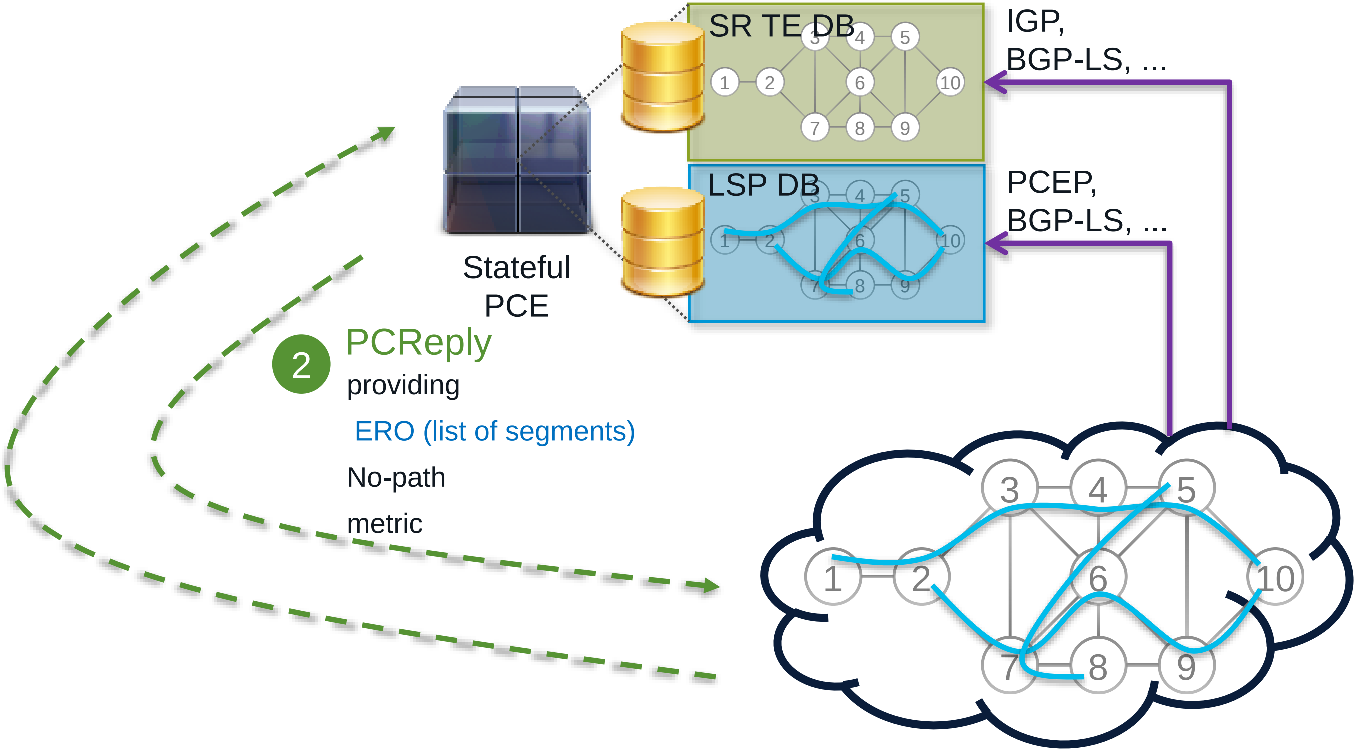
2 PCReply

providing

ERO (list of segments)

No-path

metric



SR PCE

SR PCE is IOS-XR router!

PCE must have LS database from all area’s
e.g. multi-area link from ABR router to PCE

```
pce
 address ipv4 10.0.0.11
 segment-routing
  traffic-eng
   peer ipv4 10.0.0.1
   !
…
router ospf 1
 distribute link-state instance-id 33
```

Local LS

There is no need for
BGP-LS sessions

```
pce
 address ipv4 10.0.0.11
 segment-routing
  traffic-eng
   peer ipv4 10.0.0.1
   !
…
router ospf 1
 distribute link-state instance-id 33
```

Remote LS: through BGP-LS peering

Instance-id needed for multi-domain (one instance-id per domain)

```
router bgp 65000
 address-family ipv4 unicast
 !
 address-family link-state link-state
 !
 neighbor 10.100.1.9
  address-family link-state link-state
```

There are BGP-LS sessions

Path Computation Client (PCC)

```
segment-routing
 global-block 16000 23999
 traffic-eng

 pcc
   pce address ipv4 10.0.0.11
    precedence 0
   !
   pce address ipv4 10.0.0.12
    precedence 10
```

SR-TE is used

Lower precedence is more preferred PCE

SR Policy on PCC

```
segment-routing
 global-block 16000 23999
 traffic-eng
 !
  policy policy-1
   color 1000 end-point ipv4 10.0.0.2
   candidate-paths
    preference 100
     dynamic
      pcep
      !
      metric
       type igp
!
```

Path computed by PCE

No constraints in this example

PCEP Session Verification – PCE/PCC

```
RP/0/0/CPU0:PCE# show pce ipv4 peer detail

PCE's peer database:
--------------------
Peer address: 10.100.1.1
  State: Up
  Capabilities: Stateful, Segment-Routing, Update
  PCEP has been up for: 02:06:56
  PCEP session ID: local 0, remote 0
  Sending KA every 30 seconds
  Minimum acceptable KA interval: 20 seconds
  Peer timeout after 120 seconds
  Statistics:
    Keepalive messages: rx      2 tx     2
    Request messages:   rx    810 tx     0
    Reply messages:     rx      0 tx   810
    Error messages:     rx      0 tx     0
    Open messages:      rx      1 tx     1
    Report messages:    rx      5 tx     0
    Update messages:    rx      0 tx     0
    Initiate messages:  rx      0 tx     0
  Last PCError:
    Received: None
    Sent: None

Peer address: 10.100.1.9
…
```

```
RP/0/RP0/CPU0:PE1# show segment-routing traffic-eng pcc ipv4
peer detail

PCC's peer database:
--------------------

Peer address: 10.0.0.11 (best PCE)
  State up
  Capabilities: Stateful, Update, Segment-Routing,
Instantiation
  PCEP has been up for: 00:30:01
  Local keepalive timer is 30 seconds
  Remote keepalive timer is 30 seconds
  Local dead timer is 120 seconds
  Remote dead timer is 120 seconds
  Statistics:
    Open messages:      rx 1        |  tx 1
    Close messages:     rx 0        |  tx 0
    Keepalive messages: rx 61       |  tx 61
    Error messages:     rx 0        |  tx 0
    Report messages:    rx 0        |  tx 0
    Update messages:    rx 0        |  tx 0
```

Notice the difference in command for checking the peering if done on PCE or PCC!

Verify Topology on PCE

```
RP/0/RP0/CPU0:PCE# show pce ipv4 topology isis summary

PCE's topology database summary:
--------------------------------
Showing summary data for ISIS

Topology nodes:                 7
Prefixes:                      14
Prefix SIDs:
  Total:                       14
  Regular:                     14
  Strict:                       0
Links:
  Total:                       28
  EPE:                          0
Adjacency SIDs:
  Total:                       28
  Unprotected:                 28
  Protected:                    0
  EPE:                          0

Private Information:
Lookup Nodes                   14
Consistent                    yes

Update Stats (from IGP and/or BGP):
  Noded added:                 26
  Noded deleted:                0
  Links added:                 54
  Links deleted:                0
  Prefix added:               121
  Prefix deleted:               0
```

Verify presence of all:

- Nodes
- Prefix-SIDs
- Adj-SIDs

Check the session status and # of prefixes

If BGP is used: verify sessions

```
RP/0/0/CPU0:PCE# show bgp link-state link-state summary

Process       RcvTblVer   bRIB/RIB   LabelVer  ImportVer  SendTblVer  StandbyVer
Speaker             332        332        332        332         332           0

Neighbor        Spk    AS MsgRcvd MsgSent   TblVer  InQ OutQ  Up/Down  St/PfxRcd
10.100.1.9        0 65000      92      75      332    0    0 00:17:16        164
10.100.1.10       0 65000      89      54      332    0    0 00:11:25        164
10.100.1.12       0 65000      89      54      332    0    0 00:11:23        164
10.100.1.13       0 65000      73       4      332    0    0 00:00:08        164
```

BGP LS: Nodes, Links, and Prefixes

You can copy/paste the complete NLRI part in the show command

```
RP/0/0/CPU0:PCE# show bgp link-state link-state
```

Node information

OSPF

ASN

BGP ID (0 if not running BGP)

area-id

router-id

```
*>i[V][O][I0x0][N[c65000][b0.0.0.0][a0.0.0.2][r10.100.1.6]]/376
```

Link information

Link between neighbors 10.100.1.10 and 10.100.1.13 (link IP addresses: 10.1.113.10 and 10.1.113.13)

```
*>i[E][O][I0x0][N[c65000][b0.0.0.0][a0.0.0.0][r10.100.1.10]][R[c65000][b0.0.0.0][a0.0.0.0][r10.100.1.13]][L[i10.1.113.10][n10.
1.113.13]]/792
```

IP reachable route

OSPF route type 2

Prefix 10.100.1.6/32

```
*>i[T][O][I0x0][N[c65000][b0.0.0.0][a0.0.0.1][r10.100.1.9]][P[o0x02][p10.100.1.6/32]]/488
```

Attached to the Node, Link, Prefix LS prefixes is: SRGB, SR Algorithm, SID Index, Adj SID, Flags, MPLS label, Link Attributes

→

All the info to rebuild the topology

PCE: Verify LSP Paths

Verify SR-TE policies on PCE

```
RP/0/RP0/CPU0:PCE# show pce lsp pcc ipv4 10.0.0.1 detail

PCE's tunnel database:
----------------------
PCC 10.0.0.1:

Tunnel Name: cfg_to-PE2-PCE_discr_100
 LSPs:
  LSP[0]:
   source 10.0.0.1, destination 10.0.0.2, tunnel ID 4, LSP ID 1
   State: Admin up, Operation up
   Setup type: Segment Routing
   Binding SID: 1234
   Maximum SID Depth: 10
   Bandwidth: signaled 0 kbps, applied 0 kbps
   PCEP information:
     PLSP-ID 0x2, flags: D:1 S:0 R:0 A:1 O:1 C:0
   LSP Role: Single LSP
   State-sync PCE: None
   PCC: 10.0.0.1
   LSP is subdelegated to: None
   Reported path:
     Metric type: IGP, Accumulated Metric 50
      SID[0]: Node, Label 16005, Address 10.0.0.5
      SID[1]: Node, Label 16002, Address 10.0.0.2
   Computed path: (Local PCE)
     Computed Time: Tue Apr 30 13:34:51 UTC 2019 (00:02:22 ago)
     Metric type: IGP, Accumulated Metric 50
      SID[0]: Node, Label 16005, Address 10.0.0.5
      SID[1]: Node, Label 16002, Address 10.0.0.2
```

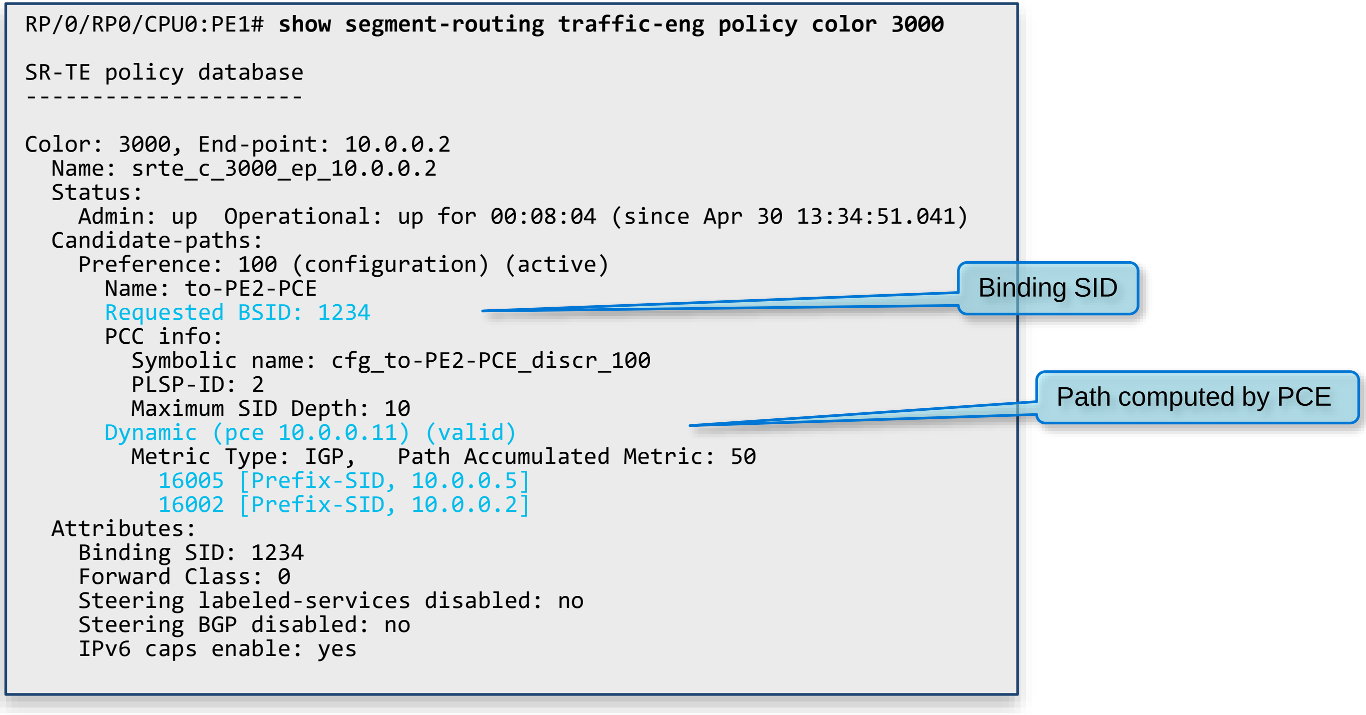
Filter on head end router

Path from source to destination

This command can be used on PCE for any source and destination and provides the path
No LSP/policy needs to be actually requested/present!

```
RP/0/RP0/CPU0:PCE# show pce ipv4 path source
10.0.0.1 destination 10.0.0.2

Path:
----:
Hop0: 10.1.3.1
Hop1: 10.3.5.3
Hop2: 10.5.6.5
Hop3: 10.4.6.6
Hop4: 10.2.4.4
```

PCC: Verify SR-TE Policy



```
RP/0/RP0/CPU0:PE1# show segment-routing traffic-eng policy color 3000

SR-TE policy database
---------------------

Color: 3000, End-point: 10.0.0.2
  Name: srte_c_3000_ep_10.0.0.2
  Status:
    Admin: up  Operational: up for 00:08:04 (since Apr 30 13:34:51.041)
  Candidate-paths:
    Preference: 100 (configuration) (active)
      Name: to-PE2-PCE
      Requested BSID: 1234
      PCC info:
        Symbolic name: cfg_to-PE2-PCE_discr_100
        PLSP-ID: 2
        Maximum SID Depth: 10
      Dynamic (pce 10.0.0.11) (valid)
        Metric Type: IGP,   Path Accumulated Metric: 50
          16005 [Prefix-SID, 10.0.0.5]
          16002 [Prefix-SID, 10.0.0.2]
  Attributes:
    Binding SID: 1234
    Forward Class: 0
    Steering labeled-services disabled: no
    Steering BGP disabled: no
    IPv6 caps enable: yes
```

Dynamic PCE SR Policy and Binding SID

```
segment-routing
 global-block 16000 23999
 traffic-eng
  logging
   policy status
  !
 policy to-PE2-PCE
   binding-sid mpls 1234
   color 3000 end-point ipv4 10.0.0.2
   candidate-paths
    preference 100
     dynamic
      pcep
      !
      metric
       type igp
```

Binding-SID

Local label for the SR Policy

```
RP/0/RP0/CPU0:PE1# show mpls forwarding

Local  Outgoing    Prefix             Outgoing     Next Hop        Bytes
Label  Label       or ID              Interface                    Switched
------ ----------- ------------------ ------------ --------------- ------------
1000   Pop         No ID              srte_c_1000_ point2point     0
1234   Pop         No ID              srte_c_3000_ point2point     0
2000   Pop         No ID              srte_c_2000_ point2point     1192
16003  Pop         SR Pfx (idx 3)     Gi0/0/0/0    10.1.3.3        0
       Unlabelled  SR Pfx (idx 3)     Gi0/0/0/2    10.1.7.7        0           (!)
24004  Aggregate   one: Per-VRF Aggr[V]   \
                                      one                          14868
24005  16005       SR TE: 4 [TE-INT]  Gi0/0/0/0    10.1.3.3        0
24006  16007       SR TE: 1 [TE-INT]  Gi0/0/0/0    10.1.3.3        0
```

```
RP/0/RP0/CPU0:PE1# show mpls forwarding labels 1234 detail
Local  Outgoing    Prefix             Outgoing     Next Hop        Bytes
Label  Label       or ID              Interface                    Switched
------ ----------- ------------------ ------------ --------------- --------
1234   Pop         No ID              srte_c_3000_ point2point     0
     Label Stack (Top -> Bottom): { Unlabelled Imp-Null }
     Outgoing Interface: srte_c_3000_ep_10.0.0.2 (ifhandle 0x00000034)
     Packets Switched: 0
```

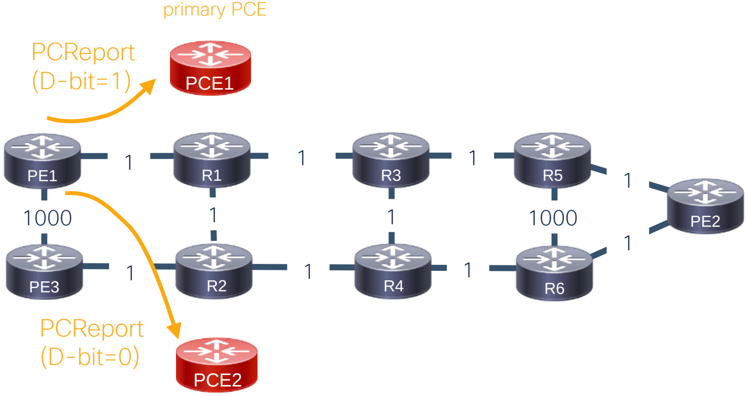
```
RP/0/RP0/CPU0:PE1# show mpls forwarding labels 24005 detail
Local  Outgoing    Prefix             Outgoing     Next Hop        Bytes
Label  Label       or ID              Interface                    Switched
------ ----------- ------------------ ------------ --------------- --------
24005  16005       SR TE: 4 [TE-INT]  Gi0/0/0/0    10.1.3.3        0
     Label Stack (Top -> Bottom): { 16005 16002 }
     Outgoing Interface: GigabitEthernet0/0/0/0 (ifhandle 0x01000018)
     Packets Switched: 0
```

SR-PCE High Availability (HA)

- Head end can send PCEP Reports for its SR Policies to all connected SR-PCEs
- The Delegate (D) flag indicates the primary SR-PCE
- If primary SR-PCE Fails: delegation to other SR-PCE by head end router

PE1

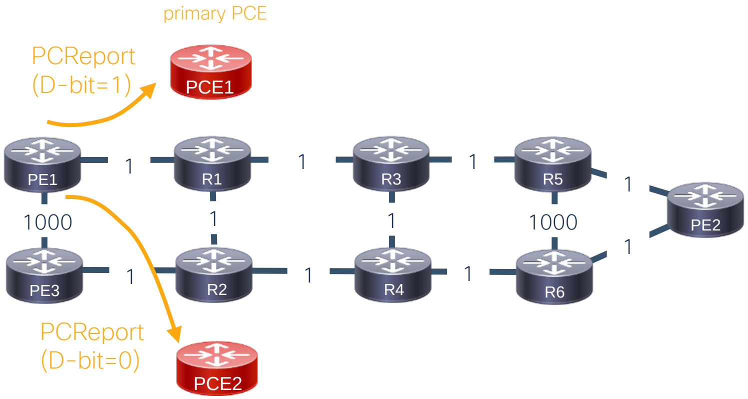
```
segment-routing
 traffic-eng
  pcc
   pce address ipv4 10.0.0.12
    precedence 10
   !
   pce address ipv4 10.0.0.13
    precedence 5
```



* PCC sends report to all connected PCEs

SR-PCE Failover

- Head sends reports to both PCEs
  - Delegates to primary PCE only (lowest precedence)
- Primary SR-PCE fails
- Head end detects PCEP session going down (dead timer)
- Re-delegation timer starts
  - Here 0 sec (PCC-centric model)
- When re-delegation timer expires, head end fails over to backup SR-PCE
  - If no more PCEs are seen by PCC:
    - Remove SR Policy LSP
- If no other PCE is found: LSP stays active for 10 min (default)



can be changed

PE1

```
segment-routing
 traffic-eng
  pcc
   pce address ipv4 10.0.0.12
    precedence 10
   !
   pce address ipv4 10.0.0.13
    precedence 20
   !
   report-all
   timers initiated state 60
```

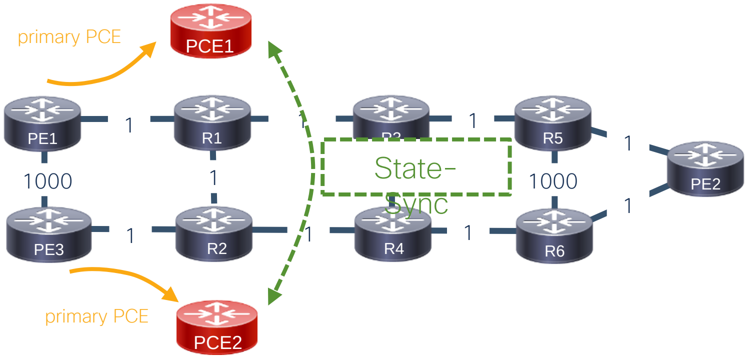
report-all: all local policies are sent to PCEs

State-Sync

- 2 different PCCs delegate policies to different PCEs
  - PCEs are unaware of each others' policies
- Disjoint path requested : the PCEs perform path computation independently: paths computed without disjointness

aka split-brain

- State-Sync session between the 2 PCEs



* configured on both PCEs: state-sync command gives direction of advertised state

PCE2

```
pce
 address ipv4 10.0.0.13
 state-sync ipv4 10.0.0.12
 segment-routing
  traffic-eng
```

PE3

```
segment-routing
 traffic-eng
  pcc
   pce address ipv4 10.0.0.12
    precedence 10
   !
   pce address ipv4 10.0.0.13
    precedence 5
```

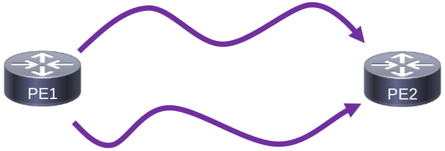
```
RP/0/RP0/CPU0:PCE1# show pce lsp private

PCC 10.0.0.4:
Tunnel Name: cfg_policy-1_discr_100
Color: 1000
Interface Name: srte_c_1000_ep_10.0.0.2
 State-sync PCE: 10.0.0.13

Event history (oldest first):
  Time                    Event
 Dec 15 21:52:10.981     Path computation (Exclude LSP): LSP-ID: 7, delegated: false,
source: 10.0.0.4, destination: 10.0.0.2, status: Disjoint Path Success, computed path
(cache hit: false, algorithm: Suurballe, duration: 1, metric type: 2 value: 40, max-
delay: 0)): 24007, 24014, 24032, 24031
```

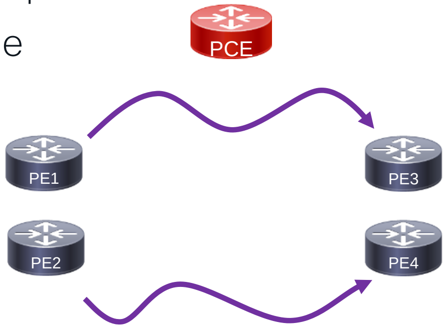
copied over by PCEP State -Sync

D-flag not set on this PCE

What is Disjointness?

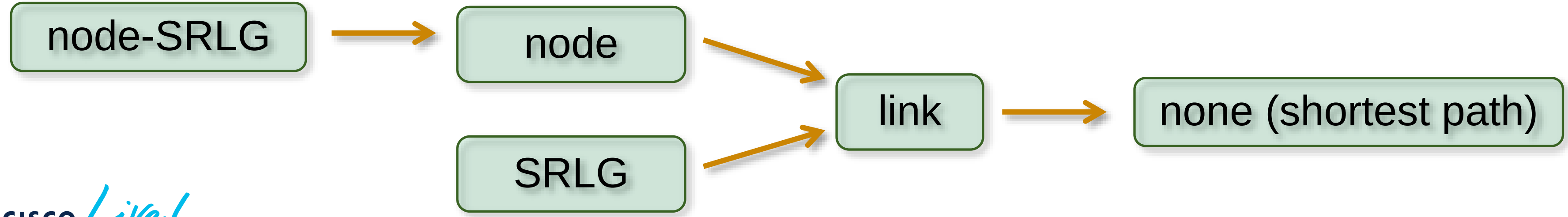
- Disjoint paths do not share any (or limited) network resources
  - It guarantees service resilience, live-live, or primary-backup
  - Link, node, SRLG, and node+SRLG disjointness is possible





different head and/or tail end routers

- Fallback if no paths adhering to disjoint constraints are found:



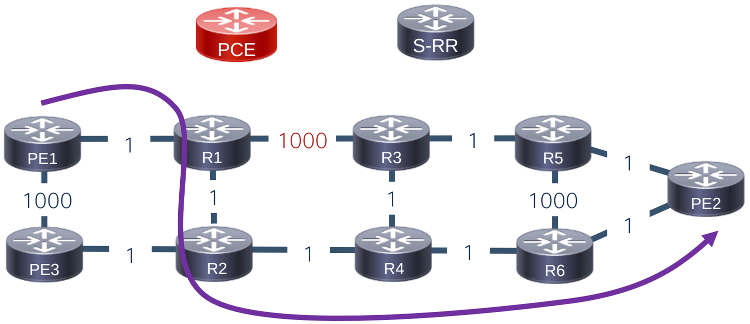
Disjointness

PE1

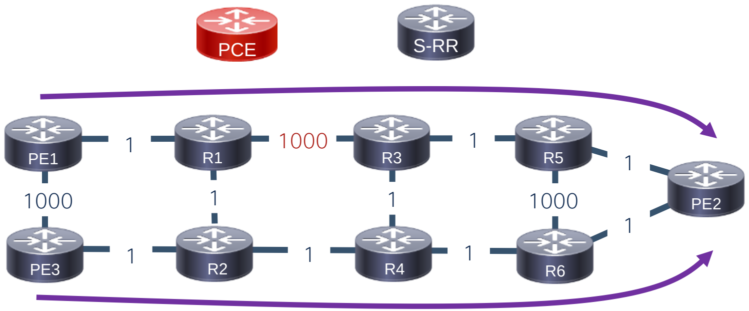
```
segment-routing
 traffic-eng
  candidate-paths
  !
  policy policy-77
   color 1000 end-point ipv4 10.0.0.2
   candidate-paths
    preference 100
     dynamic
      pcep
      !
      metric
       type te
      !
      constraints
      disjoint-path group-id 1 type node
```

PE3

```
segment-routing
 traffic-eng
  candidate-paths
  !
  policy policy-88
   color 1000 end-point ipv4 10.0.0.2
   candidate-paths
    preference 100
     dynamic
      pcep
      !
      metric
       type te
      !
      constraints
      disjoint-path group-id 1 type node
```



first path is moved, due to second SR Policy with node disjointness



- More constraints … more segments
- PCE and constraints: make sure all constraints are network-wide unique

Flex-Algo

Flex-Algo

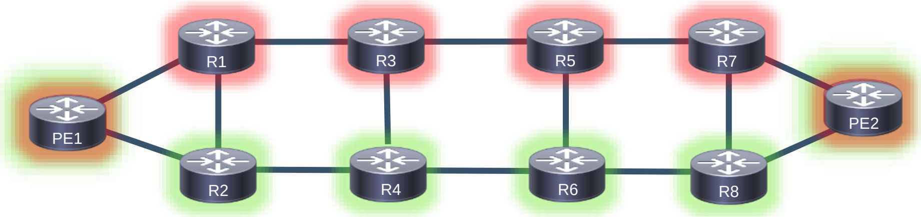
- New Prefix-Segments with specific optimization objective and constraints: custom IGP algorithm (SR-MPLS and SRv6)
  - Minimize igp-metric or delay or te-metric
  - Avoid SRLG or affinity
  - Example“: operator defines Flex-Algo 128 as “minimize delay metric” with Prefix SID 18004 associated to Lo0
- A router installed SIDs (labels) for each prefix for each algo it is part of (not the others!)
- Advertised in SR algorithm sub TLV by every router
  - Can be advertised as additional prefix-SID’s of the existing loopback address
  - Automated steering onto flex algo

Flex-Algo – Dual Plane

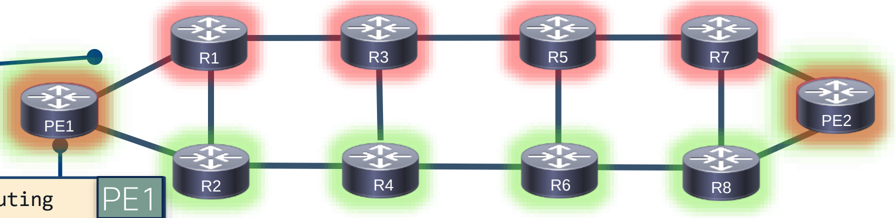
- PE routers: red + green
- P routers: red (flex algo 128)
- P routers: green (flex algo 129)

- Algo 128 & 129 both min TE metric
- Each router advertises a Prefix SID for supported algo:
  - Base 16000 for grey (algo 0)
  - Base 17000 for red (algo 128)
  - Base 18000 for green (algo 129)

PE1 is part of algo 0, 128, and 129
R1 is part of algo 0, and 128
R2 is part of algo 0, and 129



Flex-Algo – Dual Plane



R1

```
router isis 1
 net 49.0001.0000.0000.0003.00
 address-family ipv4 unicast
  metric-style wide
  segment-routing mpls
 !
 flex-algo 128
  advertise-definition
 !
 interface Loopback0
  address-family ipv4 unicast
   prefix-sid index 3
   prefix-sid algorithm 128 index 1003
```

flex algo definition should be advertised by at least two routers

16000 + 1003

PE1

```
segment-routing
 traffic-eng
  candidate-paths
  !
  on-demand color 1001
   dynamic
    pcep
    !
    sid-algorithm 128
```

```
RP/0/RP0/CPU0:PE1# show isis database R1.00-00 verbose
LSPID                 LSP Seq Num  LSP Checksum  LSP Holdtime/Rcvd  ATT/P/OL
R1.00-00              0x0000005b   0x7ec7        773  /1200         0/0/0

 Router Cap:     10.0.0.3 D:0 S:0
    Segment Routing: I:1 V:0, SRGB Base: 16000 Range: 8000
    SR Algorithm:
      Algorithm: 0
      Algorithm: 1
      Algorithm: 128
    Flex-Algo Definition:
      Algorithm: 128 Metric-Type: 0 Alg-type: 0 Priority: 128

Metric: 10         IP-Extended 10.0.0.3/32
    Prefix-SID Index: 3, Algorithm:0, R:0 N:1 P:0 E:0 V:0 L:0
    Prefix-SID Index: 1003, Algorithm:128, R:0 N:1 P:0 E:0 V:0 L:0
    Prefix Attribute Flags: X:0 R:0 N:1 E:0 A:0
    Source Router ID: 10.0.0.3
```

advertised in ISIS LSP

```
RP/0/RP0/CPU0:PE1# show segment-routing traffic-eng policy color 1001

Color: 1001, End-point: 10.0.0.2
  Name: srte_c_1001_ep_10.0.0.2
  Status:
    Admin: up  Operational: up for 00:26:31
    Candidate-paths:
    Preference: 100 (BGP ODN) (active)
      Requested BSID: dynamic
      Constraints:
        Prefix-SID Algorithm: 128
        Protection Type: protected-preferred
        Maximum SID Depth: 10
      Dynamic (pce 10.0.0.12) (valid)
        Metric Type: IGP,    Path Accumulated Metric: 50
          SID[0]: 17002 [Prefix-SID, 10.0.0.2]
```

ODN on head end

```
RP/0/RP0/CPU0:PE1# trace sr-mpls policy name srte_c_1001_ep_10.0.0.2 lsp-end-point 10.0.0.2
  0 10.1.3.1 MRU 1500 [Labels: 17002 Exp: 0]
L 1 10.1.3.3 MRU 1500 [Labels: 17002 Exp: 0] 31 ms
L 2 10.3.7.7 MRU 1500 [Labels: 17002 Exp: 0] 50 ms
L 3 10.7.9.9 MRU 1500 [Labels: 17002 Exp: 0] 48 ms
L 4 10.5.9.5 MRU 1500 [Labels: implicit-null Exp: 0] 91 ms
! 5 10.2.5.2 67 ms
```

Tree-SID

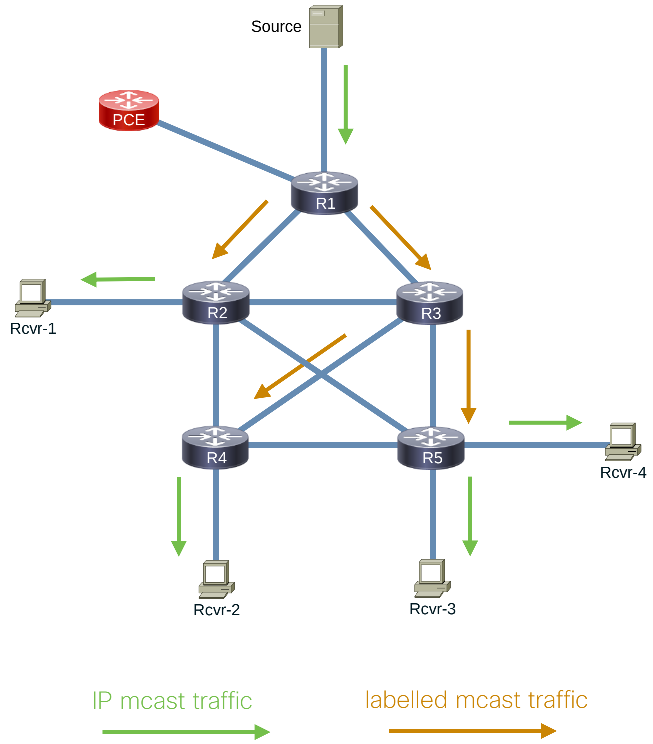
Tree-SID?

- Tree – multicast tree

  +

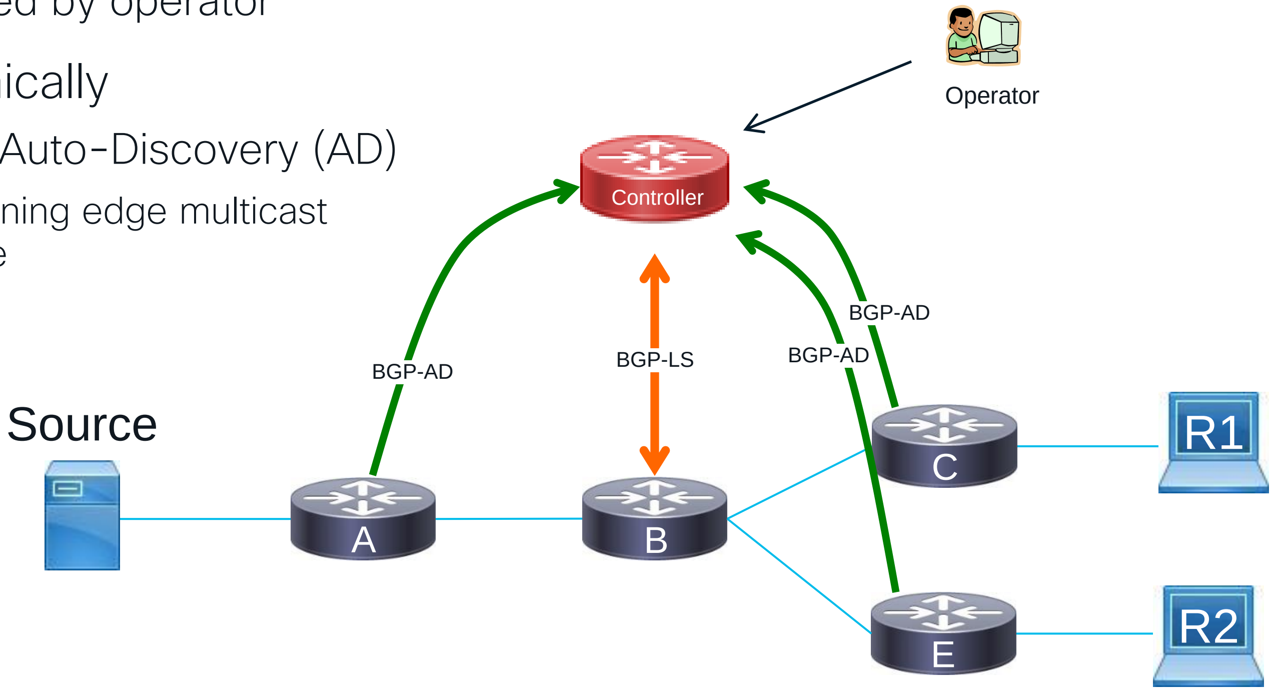
- SID – Segment IDentifier, hence Segment Routing

- Multicast packet replication on core routers
- No signalling in the core for multicast
- Applicable to both SR-MPLS and SRv6
- Today: MPLS forwarding: P2MP LSPs
  - 1 label per p2mp tree, assigned by PCE, advertised in PCEP to the routers
- Supports Inter-AS, TE capabilities, Ti-LFA protection

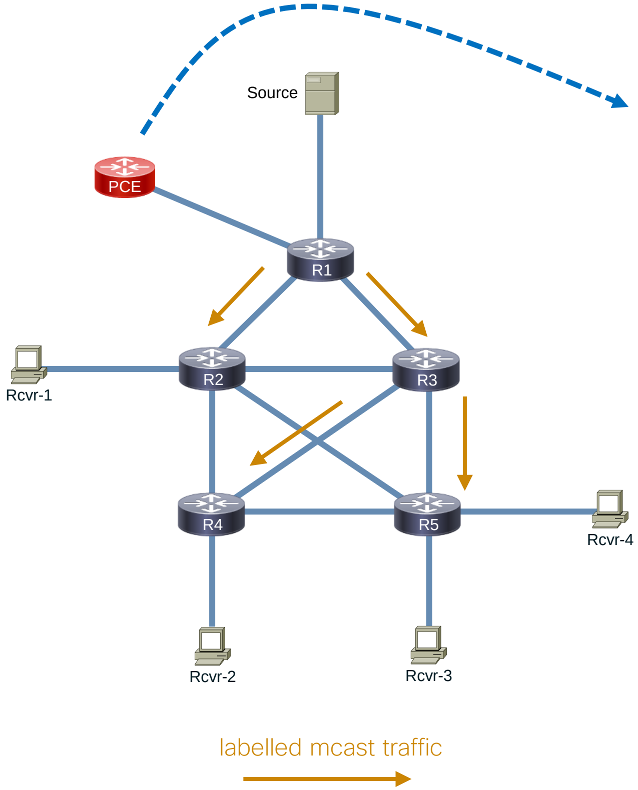


The Controller – Learning the Tree: Root and Leafs

- Statically
  - Defined by operator
- Dynamically
  - BGP-Auto-Discovery (AD)
    - Learning edge multicast state



SR-PCE Configuration - Static



```
pce
 address ipv4 10.0.0.6
 segment-routing
  traffic-eng
   p2mp
    endpoint-set R2-R4-R5
     ipv4 10.0.0.2
     ipv4 10.0.0.4
     ipv4 10.0.0.5
    !
    policy Tree-SID-Policy-1
     source ipv4 10.0.0.1
     color 1001 endpoint-set R2-R4-R5
     treesid mpls 23001
     candidate-paths
      preference 100
       dynamic
        metric
         type te
        !
```

p2mp for Tree-SID

End points: routers with receivers behind

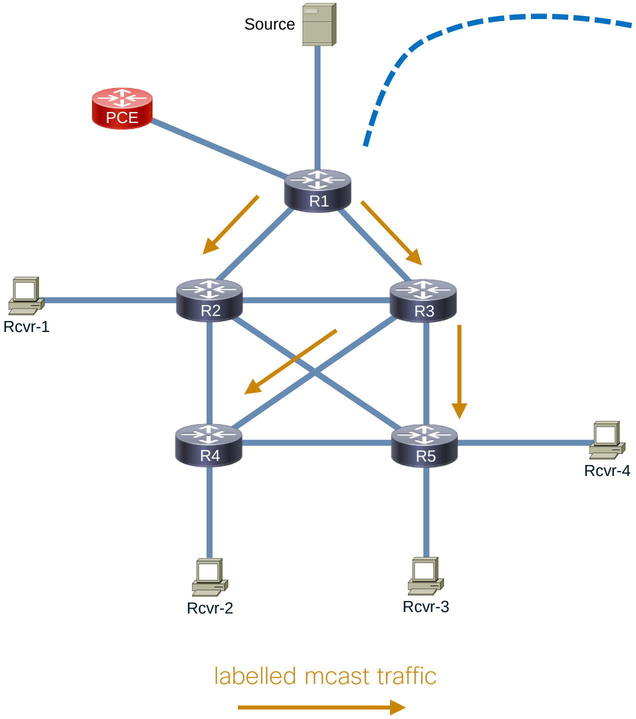
Router with source (root) connected

Color and end points

Chose one label per p2mp LSP

Other attributes like affinity/color can be configured

Headend Router Configuration - Static



```
ipv4 access-list ssm
 10 permit ipv4 232.0.0.0/8 any
!
route-policy sr-p2mp-core-tree
  set core-tree sr-p2mp
end-policy
```

mcast group must be SSM

set core tree

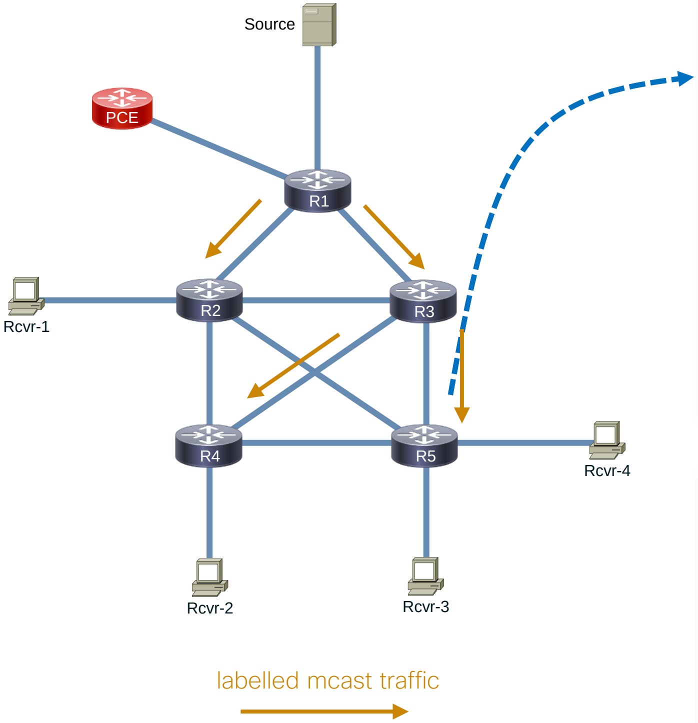
```
router pim
  interface Loopback0
   enable
  !
 !
 vrf one
  address-family ipv4
   rpf topology route-policy sr-p2mp-core-tree
 !
   ssm range ssm
   sr-p2mp-policy Tree-SID-Policy-1
    static-group 232.1.1.1 10.1.7.7
```

set RPF to p2mp core tree

must match P2MP policy name on PCE!

(S,G) to forward over Tree-SID

Tailend Router Configuration - Static



```
multicast-routing
 address-family ipv4
  interface Loopback0
   enable
  !
 !
 vrf one
  address-family ipv4
   mdt source Loopback0
   interface all enable
   static sr-policy Tree-SID-Policy-1
   mdt static segment-routing
  !
 !
```

must match P2MP policy name on PCE!

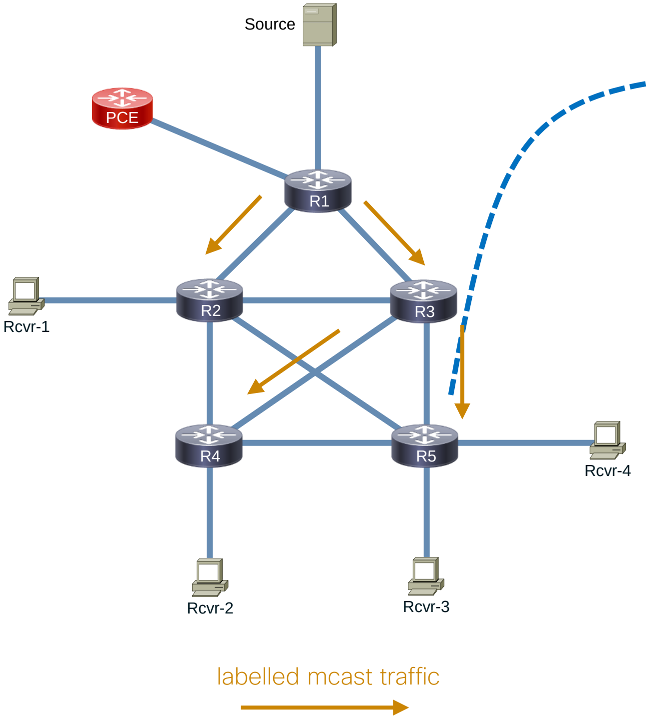
Static Tree-SID

```
router igmp
 vrf one
  interface HundredGigE0/0/0/0
   static-group 232.1.1.1 10.1.7.7
  !
  interface HundredGigE0/1/0/0
   static-group 232.1.1.1 10.1.7.7
  !
```

static forwarding requires IGMP on tailend router

statically defined forwarding onto specified interfaces

Tailend Router Configuration - Static



```
route-policy sr-p2mp-core-tree
  set core-tree sr-p2mp
end-policy
```

```
router pim
 address-family ipv4
  interface Loopback0
   enable
  !
 !
 vrf one
  address-family ipv4
   rpf topology route-policy sr-p2mp-core-tree
   ssm range ssm
  !
```

specify where the mcast traffic comes from and define the RPF

Verify Path on PCE

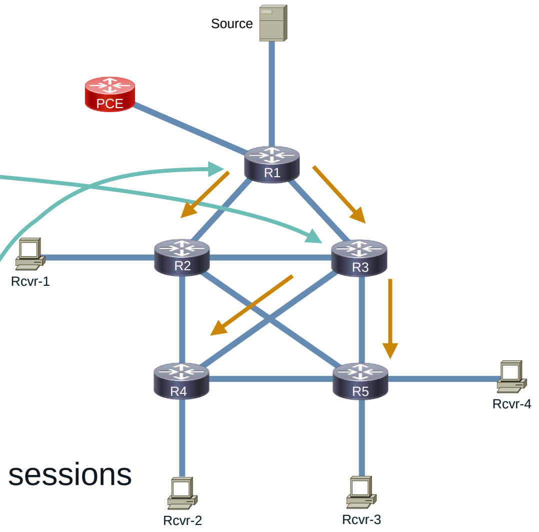
```
RP/0/RSP0/CPU0:SR-PCE# show pce lsp p2mp Tree-SID-Policy-1 private

Tree: Tree-SID-Policy-1
 Label:    23001     Operational: up  Admin: up
 Local LFA FRR: Disabled
 Metric Type: TE
 Transition count: 1
 Uptime: 03:32:17 (since Mon May 04 11:33:28 UTC 2020)
 Source: 10.0.0.1
 Destinations: 10.0.0.2, 10.0.0.4, 10.0.0.5
 Nodes:
  Node[0]: 10.0.0.3 (R3)
   Role: Transit
  Endpoints: 10.0.0.5 10.0.0.4
   Hops:
    Incoming: 23001 CC-ID: 1
    Outgoing: 23001 CC-ID: 1 (10.3.4.4) [R4:10.0.0.4]
      Endpoints: 10.0.0.4
    Outgoing: 23001 CC-ID: 1 (10.3.5.5) [R5:10.0.0.5]
      Endpoints: 10.0.0.5
  Node[1]: 10.0.0.1 (R1)
   Role: Ingress
  Endpoints: 10.0.0.2 10.0.0.5 10.0.0.4
   Hops:
    Incoming: 23001 CC-ID: 2
    Outgoing: 23001 CC-ID: 2 (10.1.3.3) [R3:10.0.0.3]
      Endpoints: 10.0.0.5 10.0.0.4
    Outgoing: 23001 CC-ID: 2 (10.1.2.2) [R2:10.0.0.2]
      Endpoints: 10.0.0.2
…
```

'private' shows the event history for this p2mp LSP

End points

End points behind this router on the P2MP tree

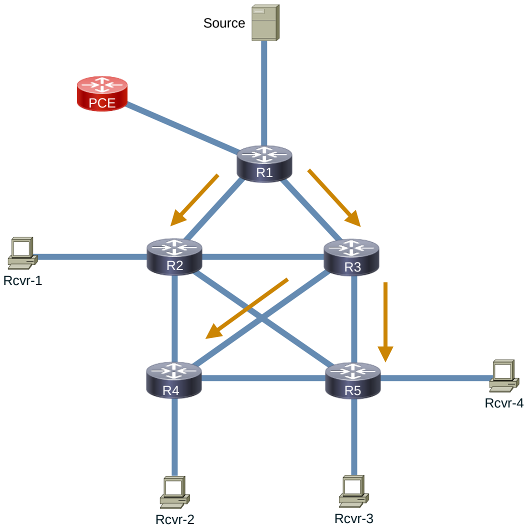
If BGP is used: verify sessions



CC-ID: Central Controller Identifier – unique ID for the instruction in PCEP

PCEP session from PCE to every router!

Verify Policy on Routers



```
RP/0/RSP0/CPU0:R3# show segment-routing traffic-eng p2mp policy

SR-TE P2MP policy database:
----------------------
! - Replications with Fast Re-route

Policy: Tree-SID-Policy-1  LSM-ID: 0x1
 Role: Transit
 Replication:
  Incoming label: 23001 CC-ID: 1
  Interface: HundredGigE0/0/0/3 [10.3.4.4]  Outgoing label: 23001 CC-ID: 1
  Interface: HundredGigE0/0/0/2 [10.3.5.5]  Outgoing label: 23001 CC-ID: 1
```

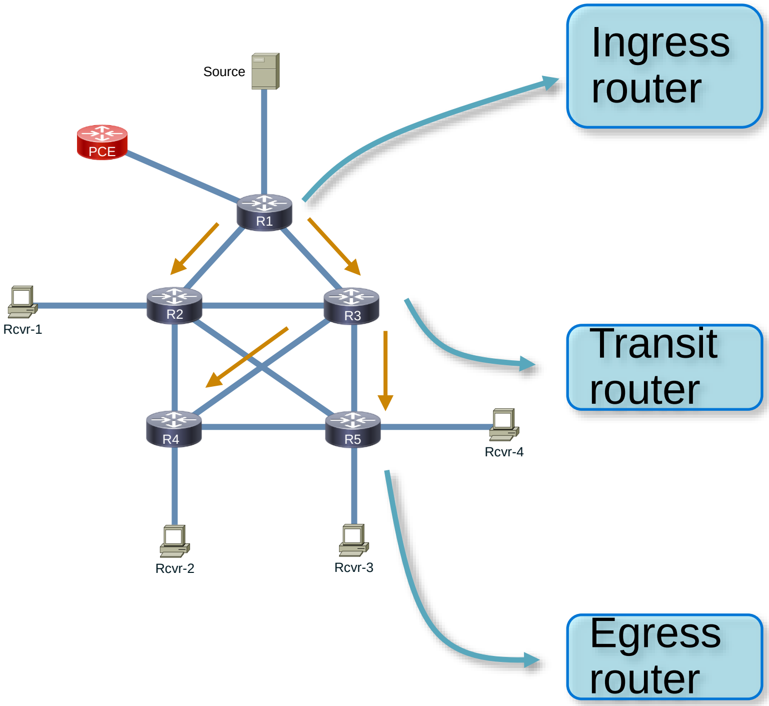
Role is Root, Transit, Bud-node, or Leaf

```
RP/0/RSP0/CPU0:R3# show segment-routing traffic-eng pcc lsp

PCC's SR policy database:
-------------------------
Symbolic Name: Tree-SID-Policy-1
 LSP[0]:
  Source 10.0.0.6, Destination 10.0.0.6, Tunnel ID 1, LSP ID 0
  State: Admin up, Operation up
  Setup type: SR
  Binding SID: 0
```

controller address

```
RP/0/RSP0/CPU0:R3# show segment-routing traffic-eng pcc lsp detail

PCC's SR policy database:
-------------------------
Symbolic Name: Tree-SID-Policy-1
 LSP[0]:
  Source 10.0.0.6, Destination 10.0.0.6, Tunnel ID 1, LSP ID 0
  State: Admin up, Operation up
  Setup type: SR
  Bandwidth: requested 0, used 0
  LSP object:
    PLSP-ID 0x1, flags: D:0 S:0 R:0 A:1 O:1 C:0
  Metric type: TE, Accumulated Metric 0
```

Verify Data Plane on Routers



Ingress router

Transit router

Egress router

```
RP/0/RSP0/CPU0:R1# show mpls forwarding p2mp
Local  Outgoing    Prefix             Outgoing     Next Hop        Bytes
Label  Label       or ID              Interface                    Switched
------ ----------- ------------------ ------------ --------------- ------------
23001  23001       mLDP/IR: 0x00000   Hu0/0/0/2    10.1.2.2        6515600
       23001       mLDP/IR: 0x00000   Hu0/0/0/1    10.1.3.3        6515600
```

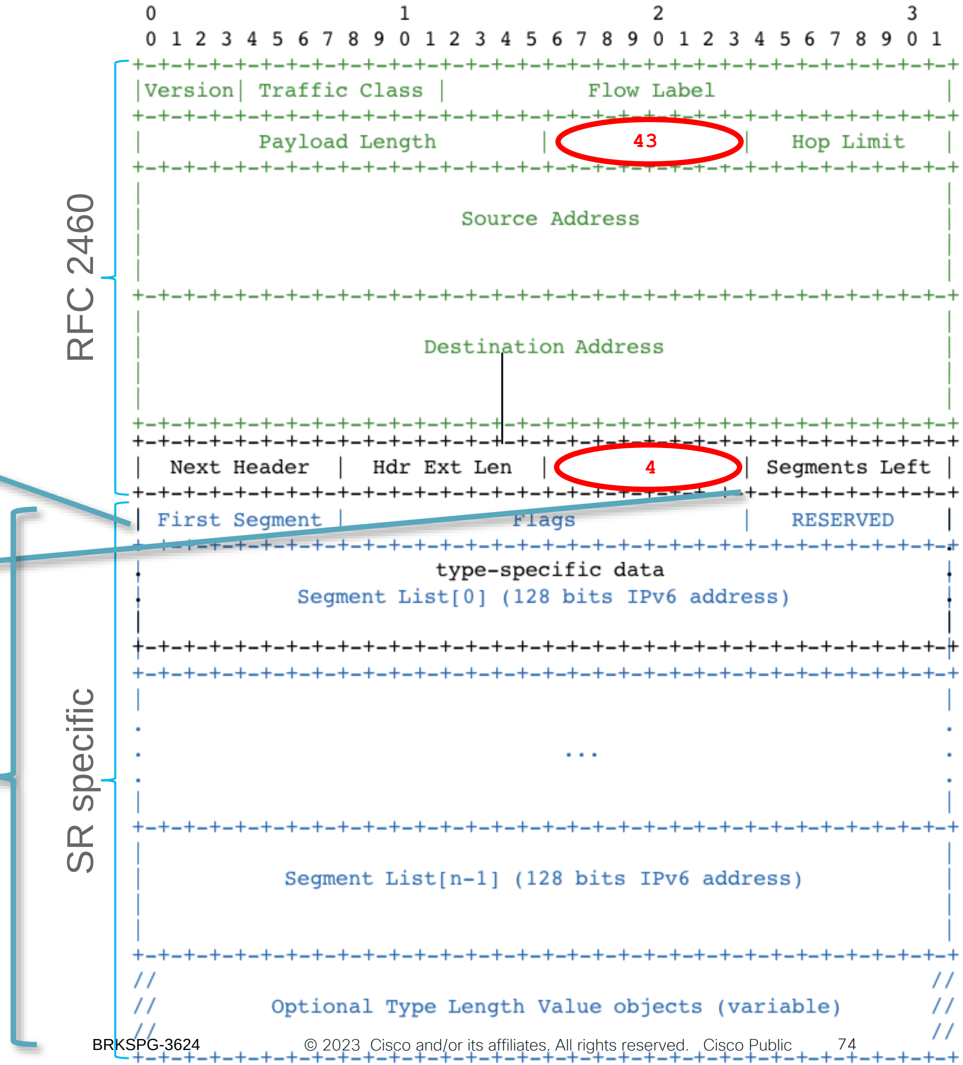
```
RP/0/RSP0/CPU0:R3# show mpls forwarding p2mp
Local  Outgoing    Prefix             Outgoing     Next Hop        Bytes
Label  Label       or ID              Interface                    Switched
------ ----------- ------------------ ------------ --------------- ------------
23001  23001       mLDP/IR: 0x00000   Hu0/0/0/3    10.3.4.4        6546568
       23001       mLDP/IR: 0x00000   Hu0/0/0/2    10.3.5.5        6546568
```

```
RP/0/RSP0/CPU0:R5# show mpls forwarding p2mp
Local  Outgoing    Prefix             Outgoing     Next Hop        Bytes
Label  Label       or ID              Interface                    Switched
------ ----------- ------------------ ------------ --------------- ------------
23001  Unlabelled  mLDP/IR: 0x00000
```

SRv6

SRv6 Encap

- Next Header = 43 (Routing Extension)
- Type = 4 (SR)



```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Version| Traffic Class |           Flow Label                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Payload Length        |      43       |   Hop Limit   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|                        Source Address                         |
|                                                               |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|                     Destination Address                       |
|                                                               |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Next Header   |  Hdr Ext Len  |       4       | Segments Left |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| First Segment |         Flags                 |   RESERVED    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
.                      type-specific data                       .
.           Segment List[0] (128 bits IPv6 address)             .
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
.                                                               .
.                              ...                              .
.                                                               .
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|          Segment List[n-1] (128 bits IPv6 address)            |
|                                                               |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                                                             //
//       Optional Type Length Value objects (variable)         //
//                                                             //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

First Segment: offset in the SR Header (SRH), not including the first 8 octets and expressed in 16-octet units, pointing to the last element of the Segment List (i.e.: that contains the first segment of the path).

Segments Left: index, in the Segment List, of the current active segment in the SRH. Decremented at each segment endpoint.

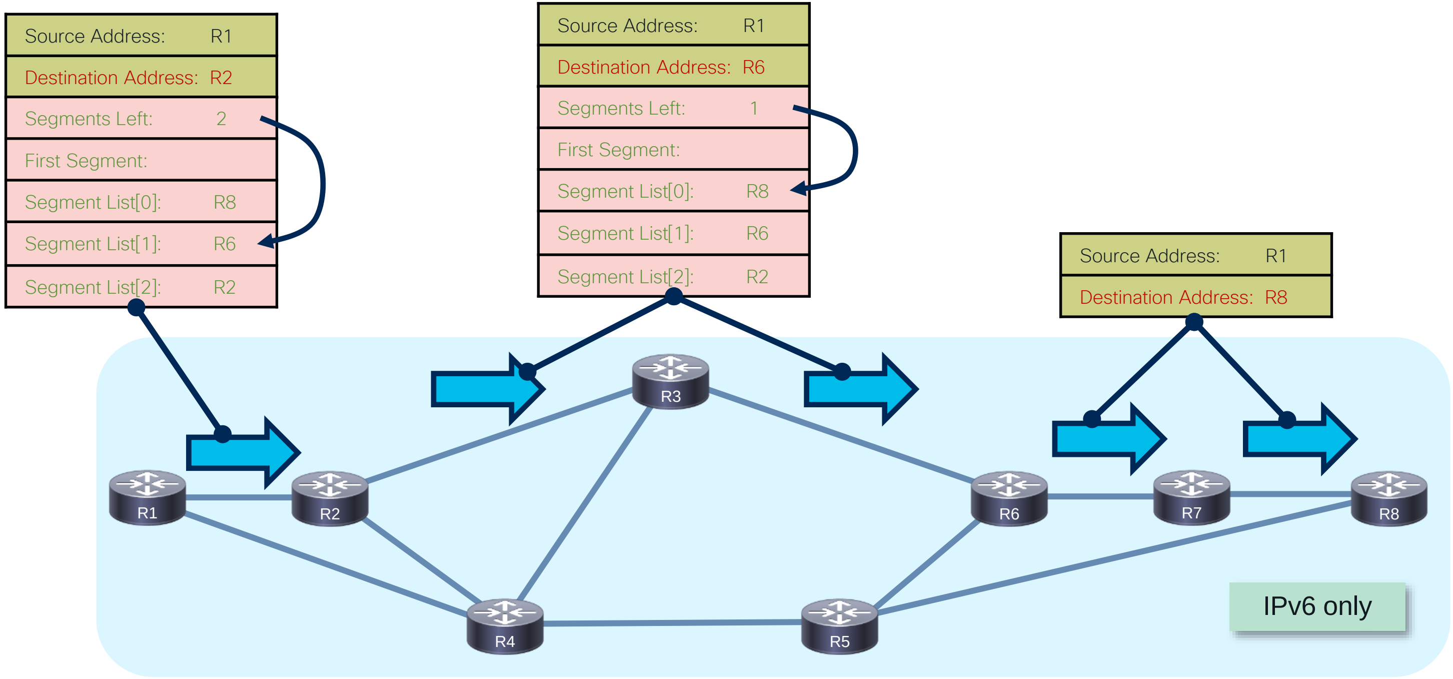
SRH: Segment List[n]: 128 bit IPv6 addresses representing each segment of the path. The segment list is encoded in the reverse order of the path: the last segment is in the first position of the list and the first segment is in the last position.

The payload can be anything

SRv6 Forwarding

Forwarding rule:

if DA is myself, and if Segments Left > 0, then update DA with segment list[Segments Left] and decrement Segments Left

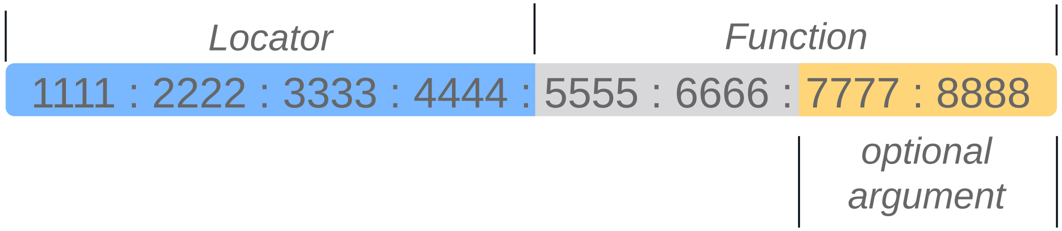


Source Address:	R1
Destination Address:	R2
Segments Left:	2
First Segment:	
Segment List[0]:	R8
Segment List[1]:	R6
Segment List[2]:	R2

Source Address:	R1
Destination Address:	R6
Segments Left:	1
First Segment:	
Segment List[0]:	R8
Segment List[1]:	R6
Segment List[2]:	R2

Source Address:	R1
Destination Address:	R8

Locator & Function



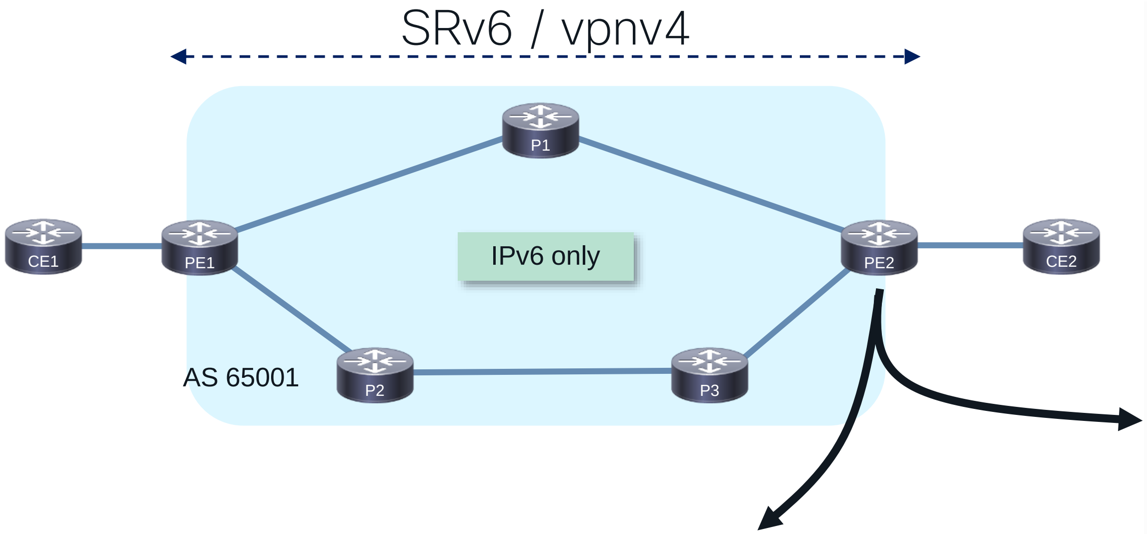
- 128-bit SRv6 SID
  - Locator: routed to the node performing the function
  - Function: any possible function (optional argument)

  - Flexible bit-length selection

Sid Functions

Function	Meaning	Comments
End	Endpoint function	SRv6 instantiation of a prefix SID
End.X	Endpoint with Layer-3 cross-connect	SRv6 instantiation of a Adj SID
End.DX4	Endpoint with decaps and IPv4 cross-connect	IPv4-L3VPN (equivalent to per-CE VPN label)
End.DT4	Endpoint with decaps and IPv4 table lookup	IPv4-L3VPN (equivalent to per-VRF VPN label)
End.B6.Insert	Endpoint bound to an SRv6 policy	SRv6 instantiation of a Binding SID
End.DT2U	Endpoint with decaps and unicast MAC L2table lookup	EVPN Bridging unicast use-cases
…		

SRv6 Based L3 vpnv4: Configuration

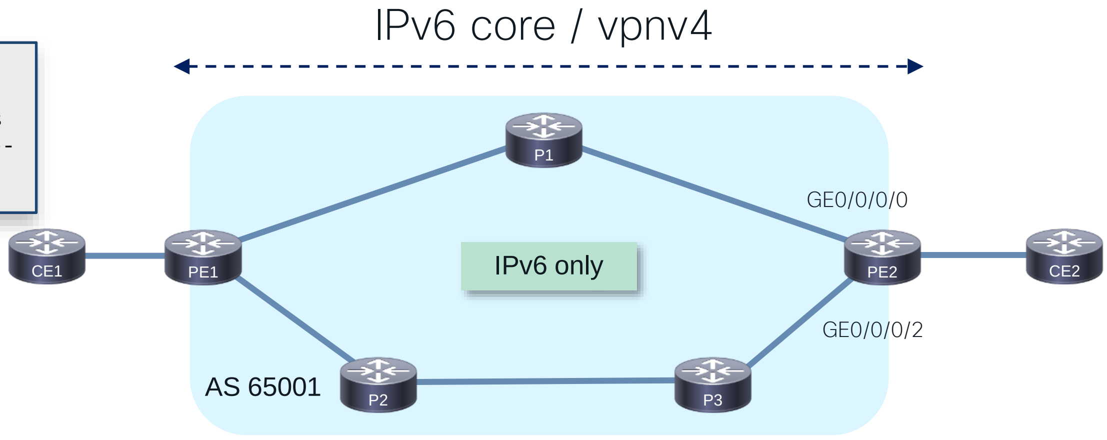


```
segment-routing
 srv6
  encapsulation
   source-address 2001:100::2
  !
  locators
   locator default
    prefix 2001:db8:2::/64
```

```
router bgp 65001
 bgp router-id 10.100.1.2
 address-family ipv4 unicast
 !
 address-family vpnv4 unicast
  segment-routing srv6
   locator default
  !
 !
 neighbor 2001:db8:100::1     # PE1
  remote-as 65001
  update-source Loopback0
  address-family vpnv4 unicast
  !
 vrf one
  address-family ipv4 unicast
   segment-routing srv6
    alloc mode per-ce  # default mode
   !
   redistribute connected
  !
  address-family ipv6 unicast
  !
  neighbor 10.2.7.7      # CE2
   remote-as 65003
   address-family ipv4 unicast
```

SRv6 Based L3 vpnv4: Verification

```
RP/0/RP0/CPU0:PE2# show segment-routing srv6 locator

Name                  ID       Prefix                   Status
--------------------  -------  ------------------------ -------
default*              1        2001:db8:2::/64          Up
```



```
RP/0/RP0/CPU0:PE2# show segment-routing srv6 sid all
*** Locator: 'default' ***

SID                         Function     Context                         Owner               State  RW
--------------------------  -----------  ------------------------------  ------------------  -----  --
2001:db8:2:0:1::            End (PSP)    'default':1                     sidmgr              InUse  Y
2001:db8:2:0:11::           End.OP       'default'                       sidmgr              InUse  Y
2001:db8:2:0:40::           End.X (PSP)  [Gi0/0/0/2, Link-Local]         isis-1              InUse  Y
2001:db8:2:0:41::           End.X (PSP)  [Gi0/0/0/0, Link-Local]         isis-1              InUse  Y
2001:db8:2:0:42::           End.DT4      'one'                           bgp-65001           InUse  Y
2001:db8:2:0:43::           End.DX4      'one':2                         bgp-65001           InUse  Y
```

used for IP lookup in VRF context

used for forwarding to destinaton on CE2

PSP: Penultimate Segment Pop of the SRH
USP: Ultimate Segment Pop of the SRH
USD: Ultimate Segment Decapsulation

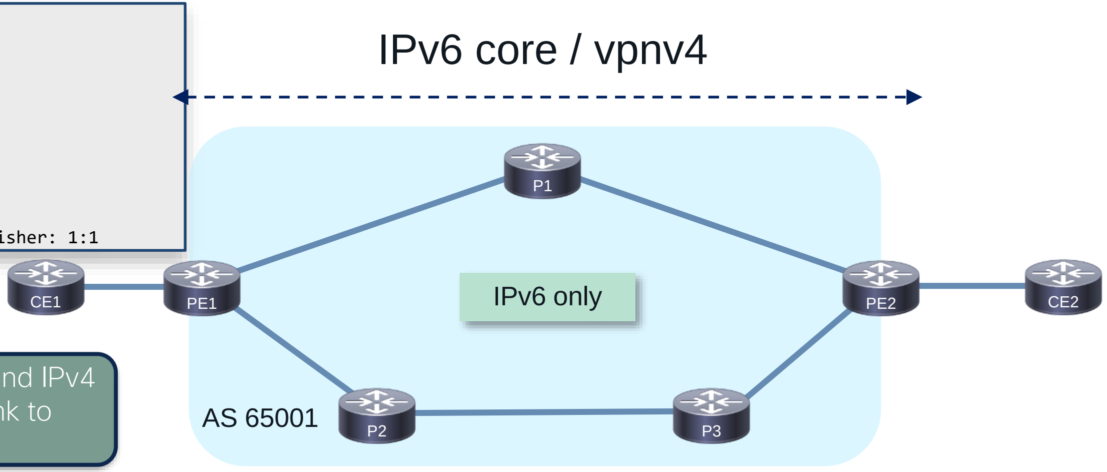
SRv6 Based L3 vpnv4: Verification

```
RP/0/RP0/CPU0:PE1# show bgp vpnv4 unicast vrf one 10.100.1.7/32
BGP routing table entry for 10.100.1.7/32, Route Distinguisher: 1:1
Paths: (1 available, best #1)
 65003
    2001:db8:100::2 (metric 20) from 2001:db8:100::2 (10.100.1.2)
      Received Label 3
      Origin IGP, metric 0, localpref 100, valid, internal, best
      Received Path ID 0, Local Path ID 1, version 34
      Extended community: RT:1:1
      SRv6-VPN-SID: T1-2001:db8:2:0:43:: [total 1]
      Source AFI: VPNv4 Unicast, Source VRF: one, Source Route Distinguisher: 1:1
```

NH = loopback PE2

End.DX4 'one':2

meaning: end point with decaps and IPv4 cross-connect -> forward onto link to CE2



```
RP/0/RP0/CPU0:PE1# show cef vrf one 10.100.1.7/32
10.100.1.7/32, version 22, SRv6 Transit
  Prefix Len 32, traffic index 0, precedence n/a, priority 3
   via 2001:db8:2::/128, 3 dependencies, recursive [flags 0x6000]
    path-idx 0 NHID 0x0 [0xe014724 0x0]
    next hop VRF - 'default', table - 0xe0800000
    next hop 2001:db8:2::/128 via 2001:db8:2::/64
     SRv6 T.Encaps.Red SID-list {2001:db8:2:0:43::}
```

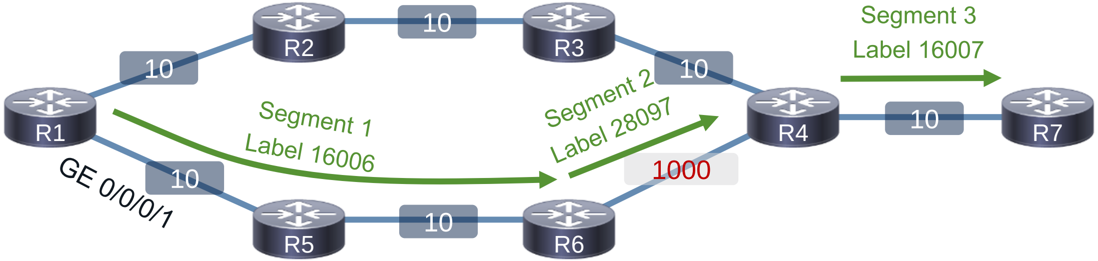
End.DX4 'one':2

T	Transit behavior
T.Insert	Transit behavior with insertion of an SRv6 policy
T.Insert.Red	Transit behavior with reduced insert of an SRv6 policy
T.Encaps	Transit behavior with encapsulation in an SRv6 policy
T.Encaps.Red	Transit behavior with reduced encaps in an SRv6 policy
T.Encaps.L2	T.Encaps applied to received L2 frames
T.Encaps.L2.Red	T.Encaps.Red applied to received L2 frames

SR Operations, Administration, and Maintenance (OAM)

NIL-FEC Example

- For any kind of MPLS OAM: you need “mpls oam” configured on every router!
- Ping and traceroute
- But very powerful tool to check any combination of segments on any path, including non-least cost path!



specify segments as list of labels in comma separated list (first label is top label)

specify outgoing interface and next hop

```
RP/0/0/CPU0:R1# trace mpls nil-fec labels 16006,28097,16007 output interface gigabitEthernet 0/0/0/1 nexthop 10.1.15.5

  0 10.1.15.1 MRU 1500 [Labels: 16006/28097/16007/explicit-null Exp: 0/0/0/0]
L 1 10.1.15.5 MRU 1500 [Labels: implicit-null/28097/16007/explicit-null Exp: 0/0/0/0] 10 ms
L 2 10.1.56.6 MRU 1500 [Labels: implicit-null/16007/explicit-null Exp: 0/0/0] 0 ms
L 3 10.1.46.4 MRU 1500 [Labels: implicit-null/explicit-null Exp: 0/0] 10 ms
! 4 10.1.47.7 10 ms
```

28097 is adj-SID label from R6 to R4

SR OAM

- Ping, traceroute for Prefix SIDs and Adj-SIDs for IGP
- Regular MPLS OAM works for SR
- OAM gives you extra (above normal ping and traceroute):
  - Consistency check
  - Path discovery
  - MPLS traffic black hole
  - Path divergence detection
  - Premature IP header exposition
  - Can detect inconsistencies between control plane and forwarding
- OAM was expanded with SR OAM
  - Only prefix-SID for now
  - Only new Target FEC Stack TLV for SR is added

```
ping mpls ipv4 10.1.1.1/32
traceroute mpls ipv4 10.1.1.1/32
```

```
ping mpls ipv4 10.1.1.1/32 fec-type generic
traceroute mpls ipv4 10.1.1.1/32 fec-type generic
```

```
ping sr-mpls 10.1.1.1/32 fec-type igp <isis/ospf>
traceroute sr-mpls 10.1.1.1/32 fec-type igp <isis/ospf>
```

verify the SR Policy

```
RP/0/RP0/CPU0:PE1# trace sr-mpls policy ?
  binding-sid  Specify the binding-sid of the SR policy
  color        Specify the color of the SR policy
  name         Specify the name of the SR policy
```

Debugging SR OAM

```
lspv_server[1113]: DBG-TLV: : Echo Hdr encode: version:[1], msg type:[1], reply mode:[2],
lspv_server[1113]: DBG-TLV: :   return_code:[0], return_subcode:[0], sender handle:[5b9d40ec],
lspv_server[1113]: DBG-TLV: :   sequence number:[2],
lspv_server[1113]: DBG-TLV: :   timestamp sent:[E092609D.9D5A9BC6 (12:22:53.614 UTC Fri May 24 2019)],
lspv_server[1113]: DBG-TLV: :   timestamp rcvd:[00000000.00000000 (00:00:00.000 UTC Thu Jan  1 1970)]
lspv_server[1113]: DBG-TLV: : Cisco ext subTLV encode: type:[1], length:[4], tlv revision:[0x4]
lspv_server[1113]: DBG-TLV: : SR IGP IPv4 Prefix SID encode: destaddr 10.0.0.2/32, protocol OSPF
lspv_server[1113]: DBG-Pkt: : TFS TLV added for request (sender_handle:[0x5b9d40ec])
lspv_server[1113]: DBG-TLV: : DSMAP encode:
lspv_server[1113]: DBG-TLV: :   addr_type:[1], rtr_id:[10.3.5.5], mtu:[1500],
lspv_server[1113]: DBG-TLV: :   intf_addr:[10.3.5.5], flags:[0x0], hashkey:[0], depth limit:[0],
lspv_server[1113]: DBG-TLV: :   multipath length:[0], [16002]
lspv_server[1113]: DBG-Pkt: : DSMAP TLV added for request (sender_handle:[0x5b9d40ec])
lspv_server[1113]: DBG-Pkt: : UDP checksum:[0xcfa9] <10.1.3.1,3503> -> <127.0.0.1,3503>, len:[96]
lspv_server[1113]: DBG-Pkt: : Echo packet built successfully for request (sender_handle:[0x5b9d40ec]), pak
size:[124]

lspv_server[1113]: DBG-Pkt: : Processing received ipv4 packet
lspv_server[1113]: DBG-Pkt: : Echo packet received: rx interface:[0x1000018], src:[10.3.5.5], dst:[10.1.3.1],
lspv_server[1113]: DBG-Pkt: : Getting rx info
lspv_server[1113]: DBG-TLV: : Echo Hdr decode: version:[1], msg type:[2], reply mode:[2],
lspv_server[1113]: DBG-TLV: :   return_code:[8], return_subcode:[1], sender handle:[5b9d40ec],
lspv_server[1113]: DBG-TLV: :   sequence number:[2],
lspv_server[1113]: DBG-TLV: :   timestamp sent:[E092609D.9D5A9BC6 (12:22:53.614 UTC Fri May 24 2019)],
lspv_server[1113]: DBG-TLV: :   timestamp rcvd:[E092609E.3DD46742 (12:22:54.241 UTC Fri May 24 2019)]
lspv_server[1113]: DBG-TLV: : Downstream Mapping, tlvtype:[0x2], tlvlength:[0x14]
lspv_server[1113]: DBG-TLV: :   epkt->dsmap_tlvs.echo_reply_dsmap_queue inited
lspv_server[1113]: DBG-TLV: : DSMAP decode:
lspv_server[1113]: DBG-TLV: :     addr_type:[1], rtr_id:[10.5.6.6], mtu:[1500],
lspv_server[1113]: DBG-TLV: :     intf_addr:[10.5.6.6], flags:[0x0], hashkey:[0],
lspv_server[1113]: DBG-TLV: :     depth limit:[0], multipath length:[0],
lspv_server[1113]: DBG-TLV: :     Labels:[16002]
lspv_server[1113]: DBG-TLV: :   DSMAP enqueued to epkt->dsmap_tlvs.echo_reply_dsmap_queue
lspv_server[1113]: DBG-TLV: : Echo packet decoded assuming tlv version:[4]
```

debug mpls oam packet

debug mpls oam tlv

Key Takeaways

Key Takeaway’s

- SR is simpler and easier to troubleshoot than LDP
- No changes in MPLS forwarding
- Ti-LFA
    - Built from same fundaments as LFA
    - But much better and much easier
- SR Policy (SR-TE) is simpler than RSVP-TE
- Controller
    - Understand basic PCEP
    - Understand BGP-LS if used
- Tree-SID: SR-MPLS + multicast!

Complete your Session Survey

- Please complete your session survey after each session. Your feedback is important.
- Complete a minimum of 4 session surveys and the Overall Conference survey (open from Thursday) to receive your Cisco Live t-shirt.

- All surveys can be taken in the Cisco Events Mobile App or by logging in to the Session Catalog and clicking the "Attendee Dashboard” at https://www.ciscolive.com/emea/learn/sessions/session-catalog.html

Continue Your Education

Visit the Cisco Showcase for related demos.

Book your one-on-one Meet the Engineer meeting.

Attend any of the related sessions at the DevNet, Capture the Flag, and Walk-in Labs zones.

Visit the On-Demand Library for more sessions at ciscolive.com/on-demand.

The bridge to possible

Thank you

cisco Live!

cisco Live!

ALL IN