
cisco Live!
ALL IN

The bridge to possible

Segment Routing MPLS Introduction: Follow the Labels

Luc De Ghein –Technical Leader CX

BRKENT-2005

Cisco Webex App

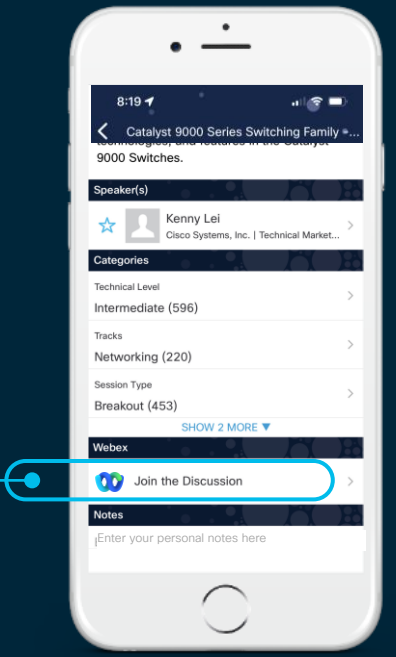
Questions?

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

Webex spaces will be moderated by the speaker until June 17, 2022.

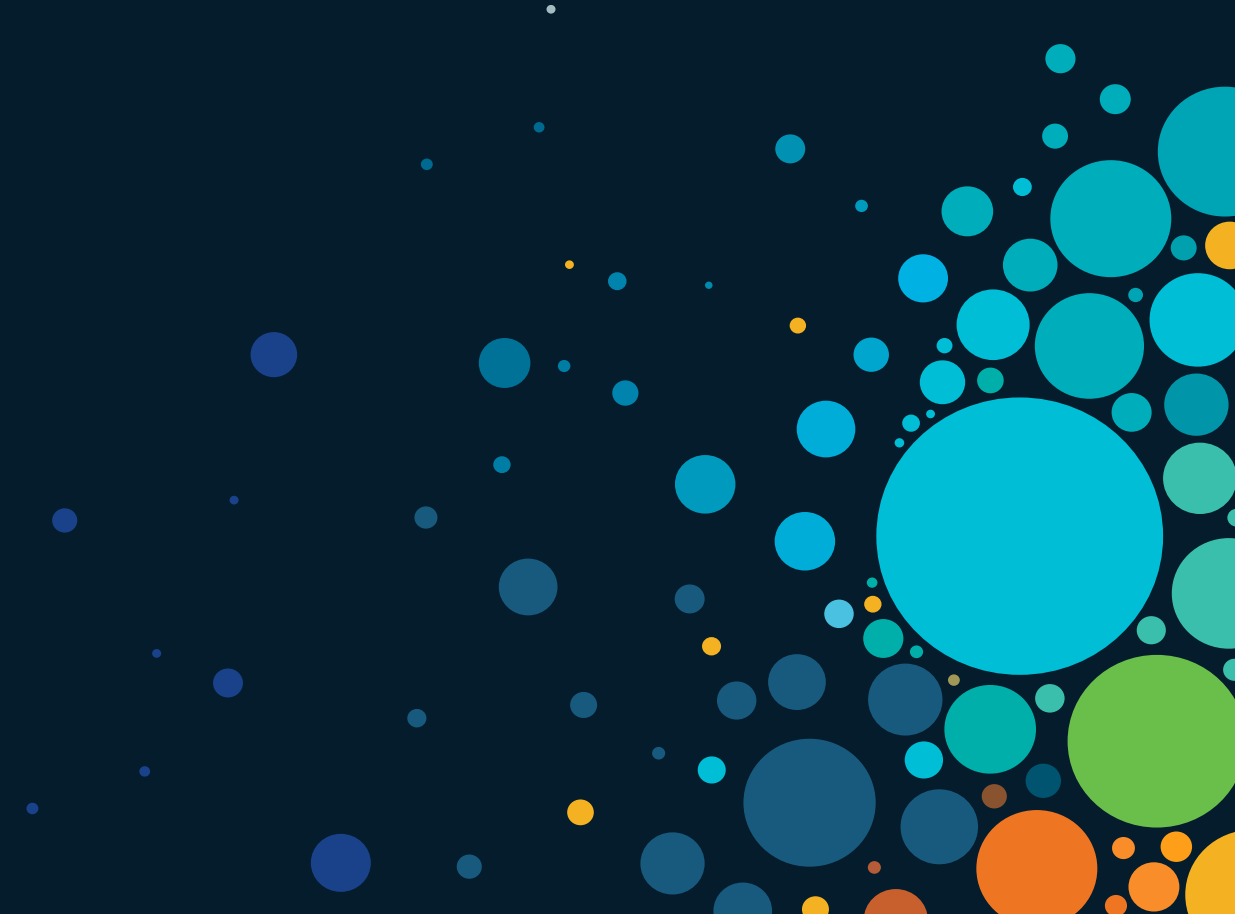


https://ciscolive.ciscoevents.com/ciscolivebot/#BRKENT-2005

Agenda

- Introduction
- The Labels
- MPLS Forwarding
- Ti-LFA
- SR Policy
- On Demand Next Hop (ODN)
- SR Operations, Administration, and Maintenance (OAM)
- Key Takeaways

Introduction



Stay Up-To-Date

http://www.segment-routing.net/

https://www.linkedin.com/groups/8266623

https://twitter.com/SegmentRouting

https://www.facebook.com/SegmentRouting/

Before We Get Started

- This session in on SR MPLS

- The services you are used, still work

Including Inter-AS Option A, B, and C

Why SR?

- No LDP
  - But interworking is still possible
- No RSVP-TE
- Load balancing by default
- Simplified forwarding
  - Based on color
  - Automated steering
- Simplified troubleshooting
- Better protection
- Many new features

Autoroute Announce (AA)

Autoroute Destination

Static route

Access-list Based Forwarding (ABF)

Path Disjointness (Multi-plane)

Real-Time Low Latency Services

Egress Peer Engineering (EPE)

Bandwidth Optimization

Path Disjointness (Multi-plane)

Micro-loop avoidance

Point-to-Multipoint delivery with Tree-SID: Multicast leveraging mVPN

Flex-Algo

On-Demand Next-Hop (ODN) + Automated steering (AS)

In a Nutshell

Data Plane

- MPLS (segment labels)
- IPv6 (+ SR extension header)

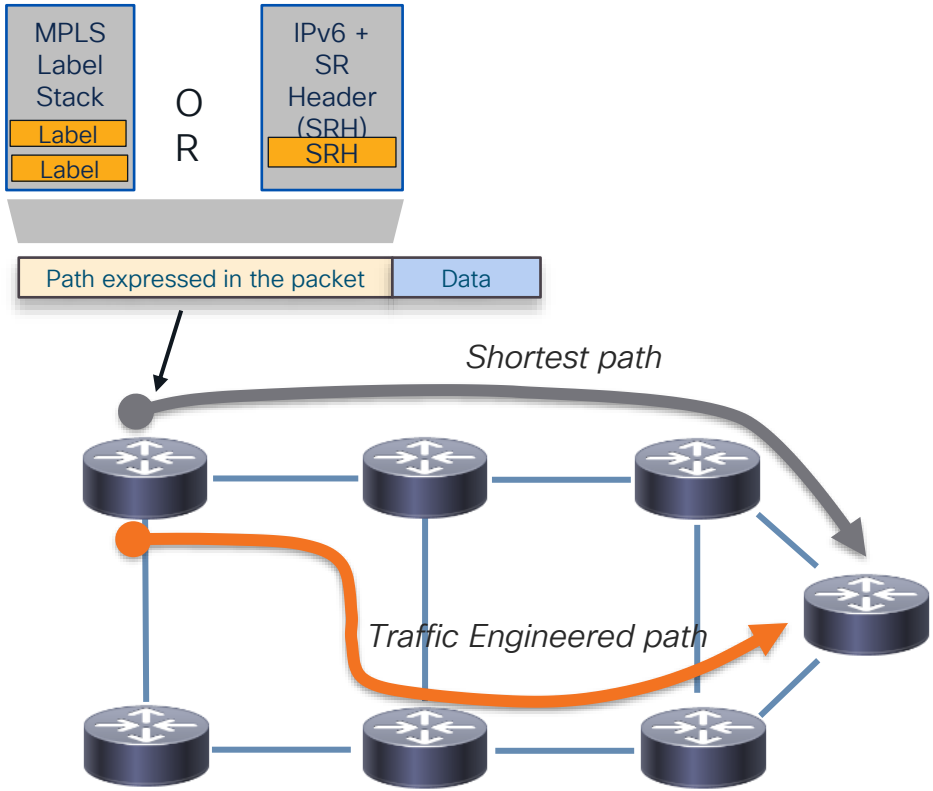
Control Plane

- Routing protocols with extensions (IS-IS, OSPF, BGP)
- SDN controller (BGP, PCEP, NETCONF/YANG)

Path Options

- Dynamic (Optimized CSPF computation)
- Explicit (expressed in the packet)

Source Routing paradigm
Stateless IP fabric



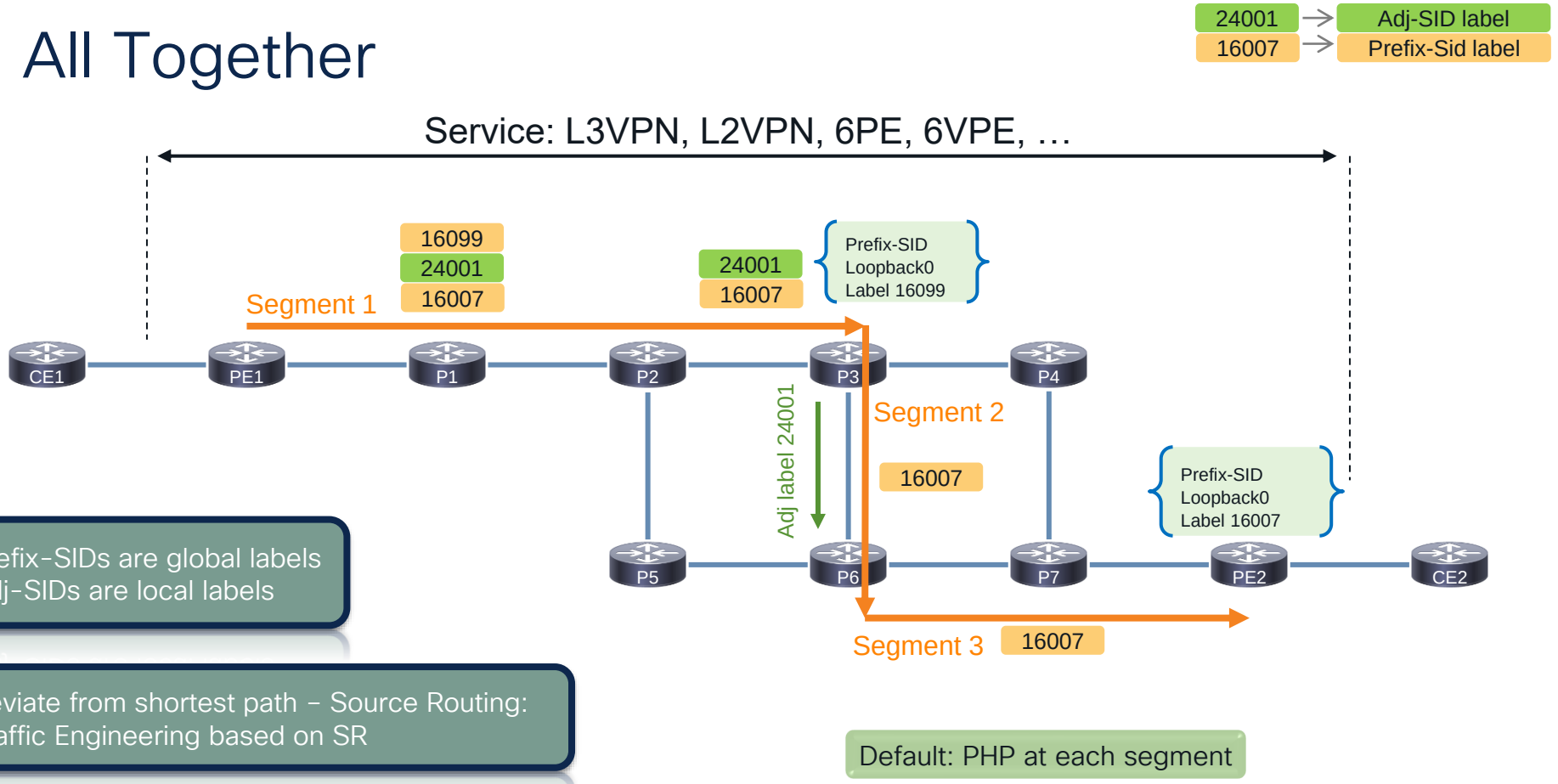
The Labels

A Segment

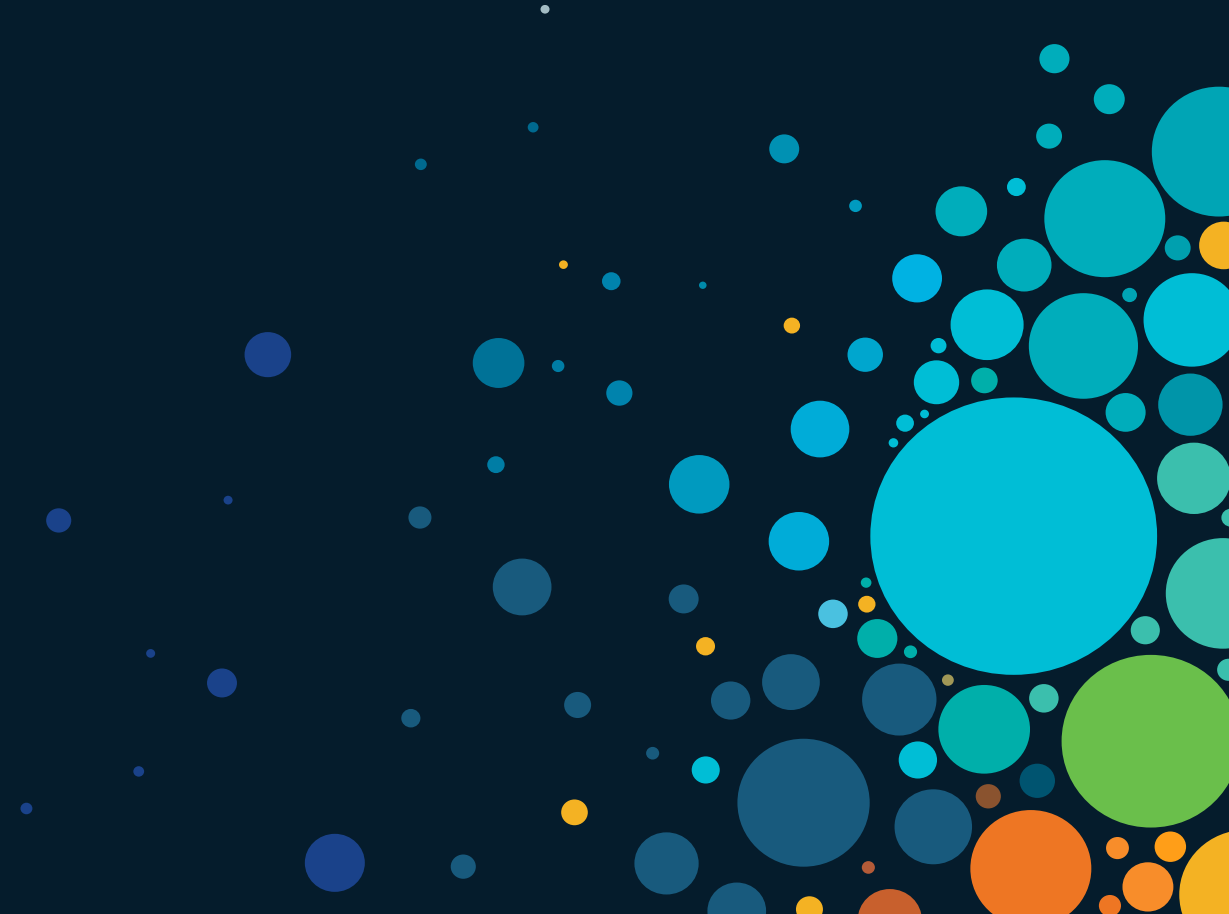
- A segment is an instruction
  - with MPLS forwarding: segment = label
- Link-state routing protocol is needed to advertise
  - Segments (Prefix-SID, Adjacency-SID)
  - MPLS Label
- Removing the signaling and state (no LDP/ no RSVP-TE)
- Controller/SDN can be used if/when needed

*SID = Segment Identifier

All Together

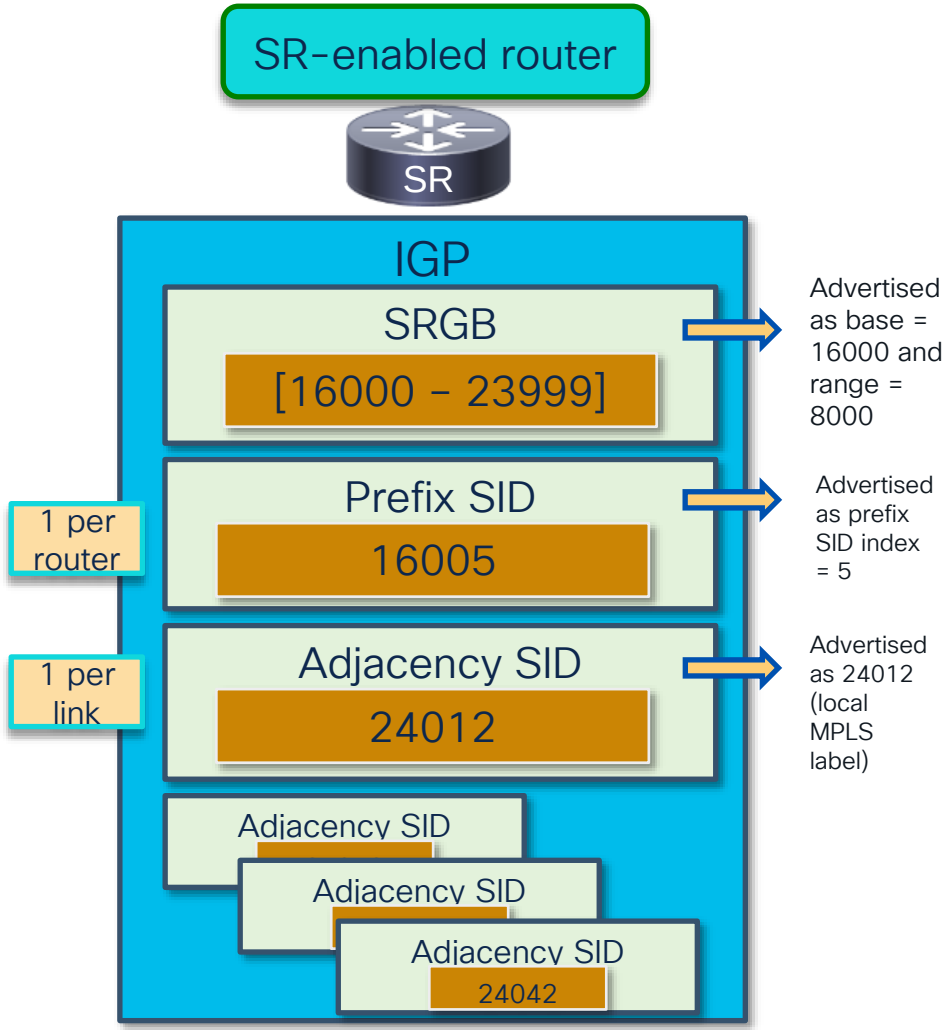


MPLS Forwarding



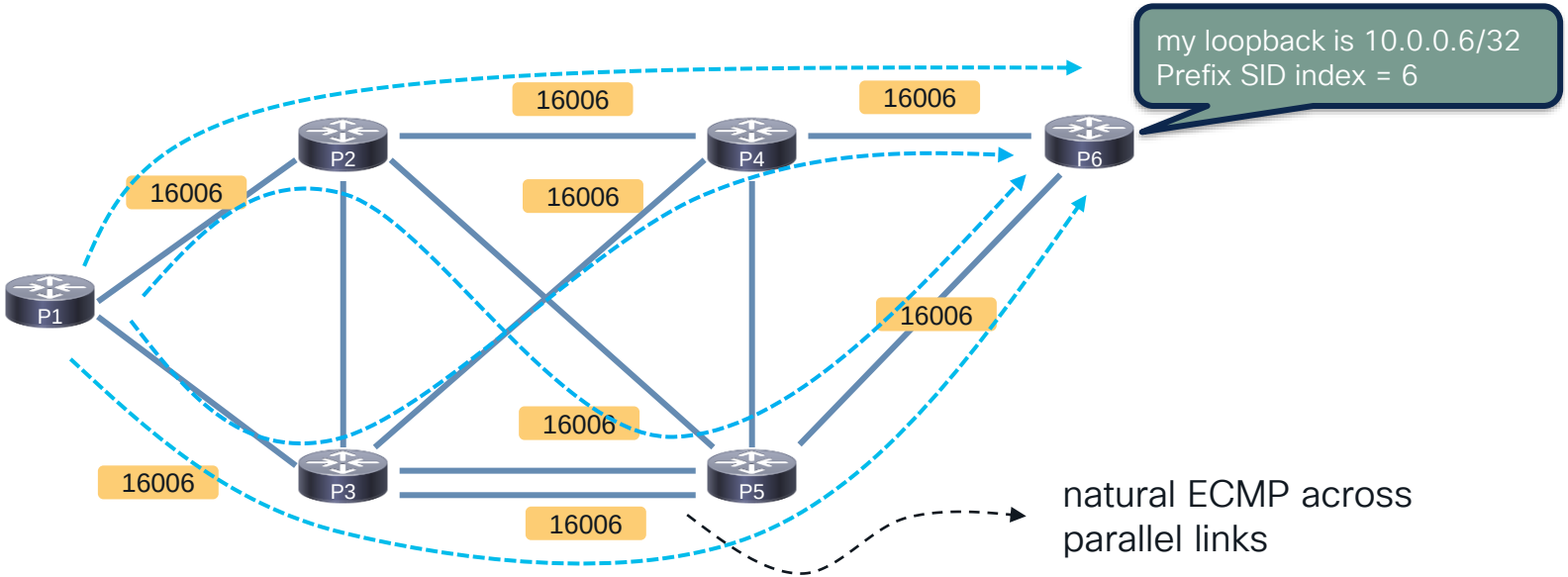
SIDs

- Prefix SID
  - SID encoded as an index
  - Index represents an offset from SRGB base
  - Index globally unique
  - SRGB may vary across LSRs
  - SRGB (base and range) advertised with router capabilities
- Adjacency SID
  - SID encoded as absolute (i.e. not indexed) value
  - Locally significant
  - Automatically allocated for each adjacency
  - From label range [24,000-max] used for dynamic label allocation



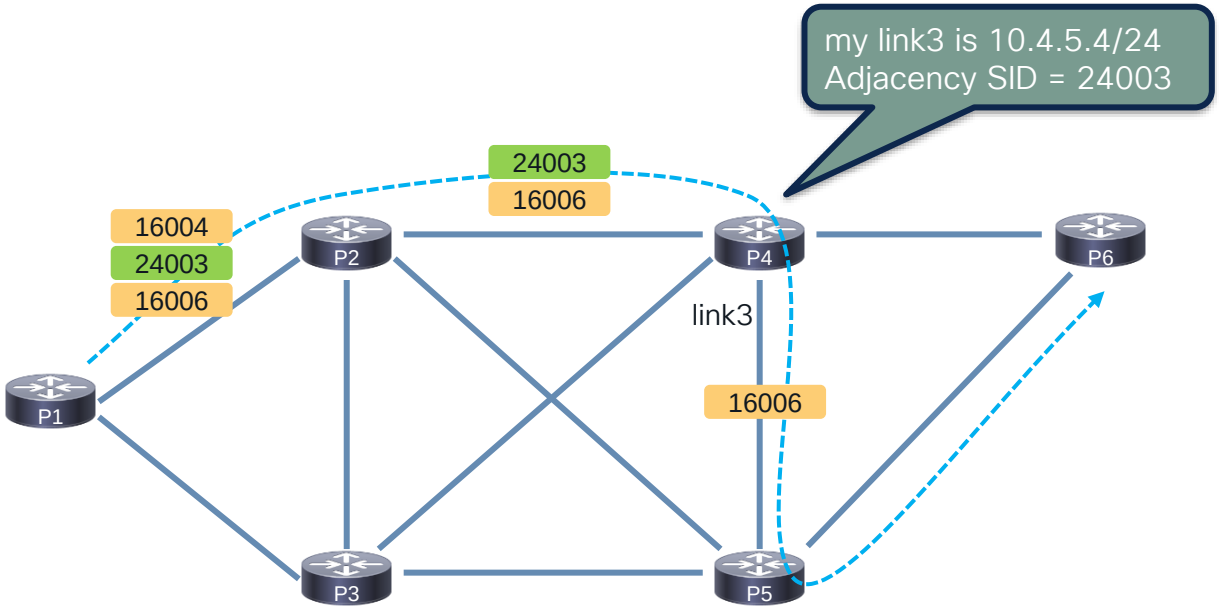
Global Labels

- Recommended same SRGB on all routers
- ‘Node’ label = get the traffic to me, by shortest route, possibly with ECMP
- A packet injected anywhere with top label 16006 will go to R6



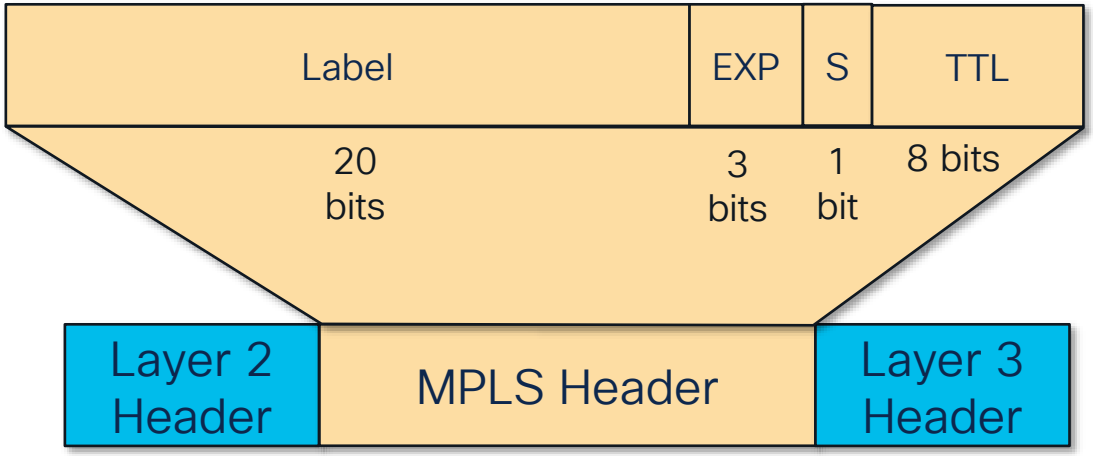
Local Adjacency SID Labels

- Force the packet to take a link at a router



No Change in MPLS Forwarding

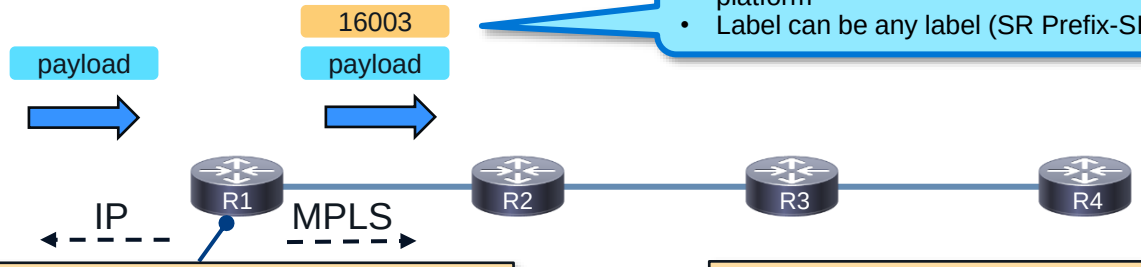
- MPLS label operations:
  - Push, Pop, and Swap
- We have, as before, …
  - Special labels {0 – 15}
    - PHP (default behavior, also for SR)
    - explicit-null for IPv4 and IPv6
  - QOS propagation (EXP bits)
    - Still uniform model, pipe, and short pipe model
  - TTL propagation as usual
  - Load balancing as before
    - FAT label support



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 (MPLS VPN, BGP-LU, etc.)
- Push can occur at intermediate MPLS (P) router
  - MPLS label(s) added in LFIB (active protection)

- 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
```

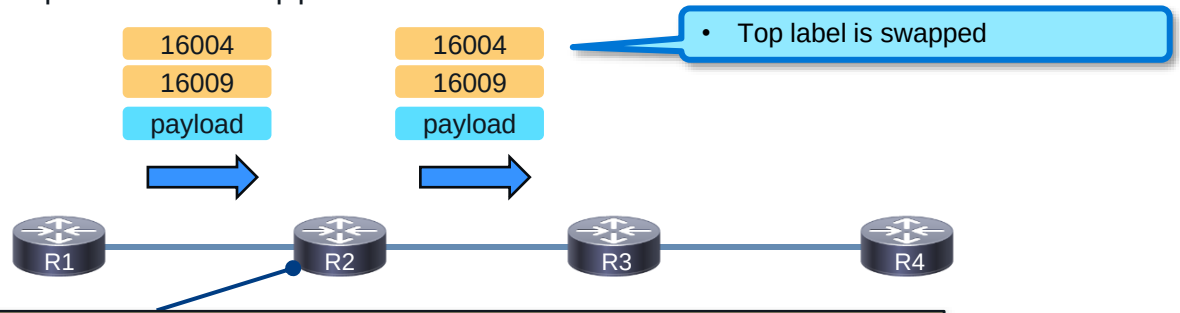
```
RP/0/0/CPU0:R1# show cef 10.100.1.3/32

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

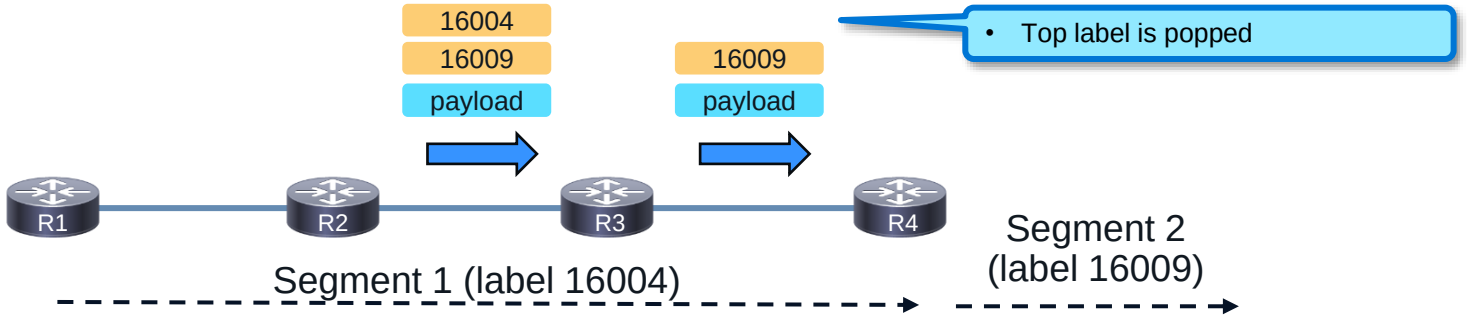



```
RP/0/0/CPU0:R2# show mpls forwarding labels 16004

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)

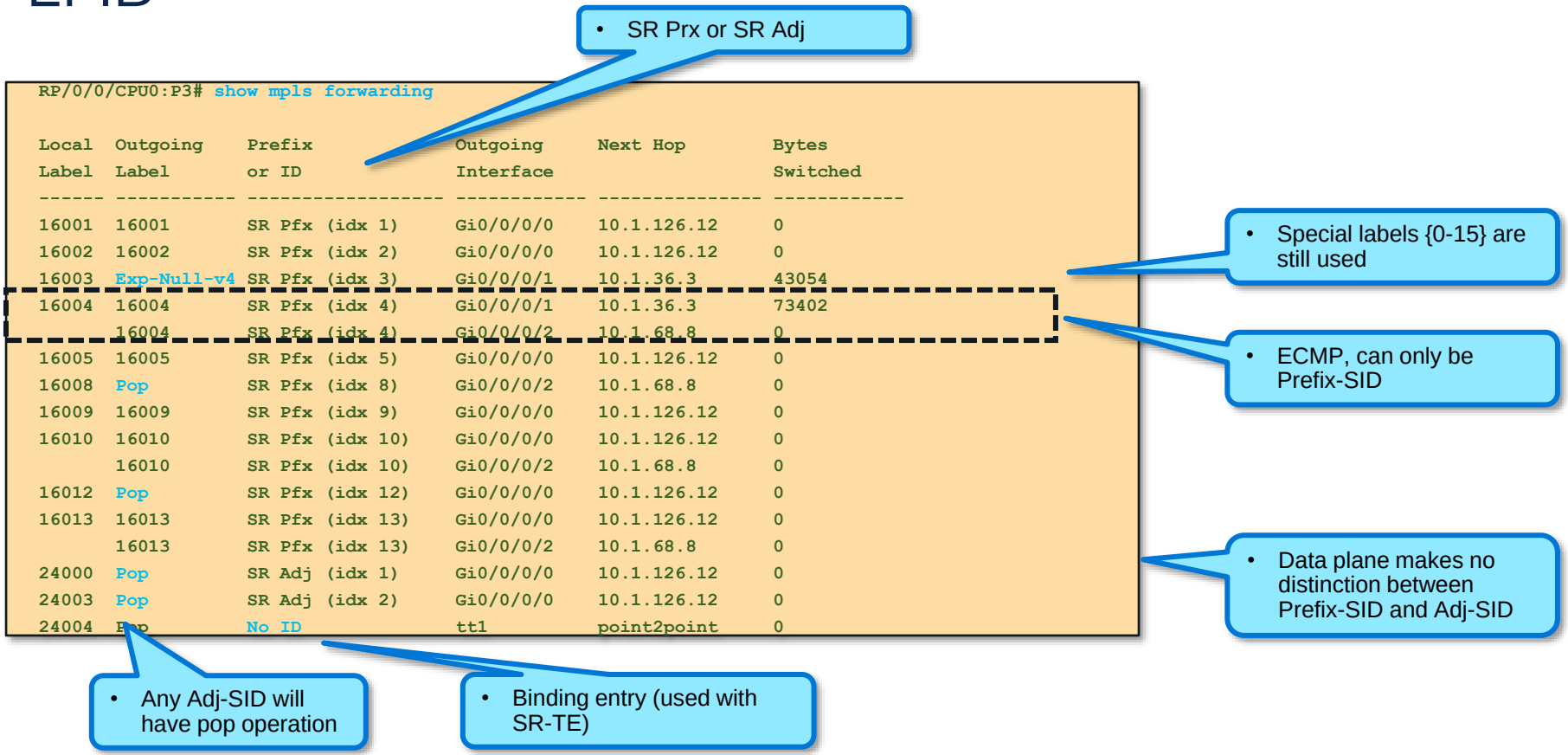
- Pop occurs at intermediate MPLS (P) router: top label is removed
- By default on penultimate router of one SR segment
  - Label stack could become unlabeled
  - Label stack can still have other labels
    - e.g. when packet is moved from one SR segment to another SR segment



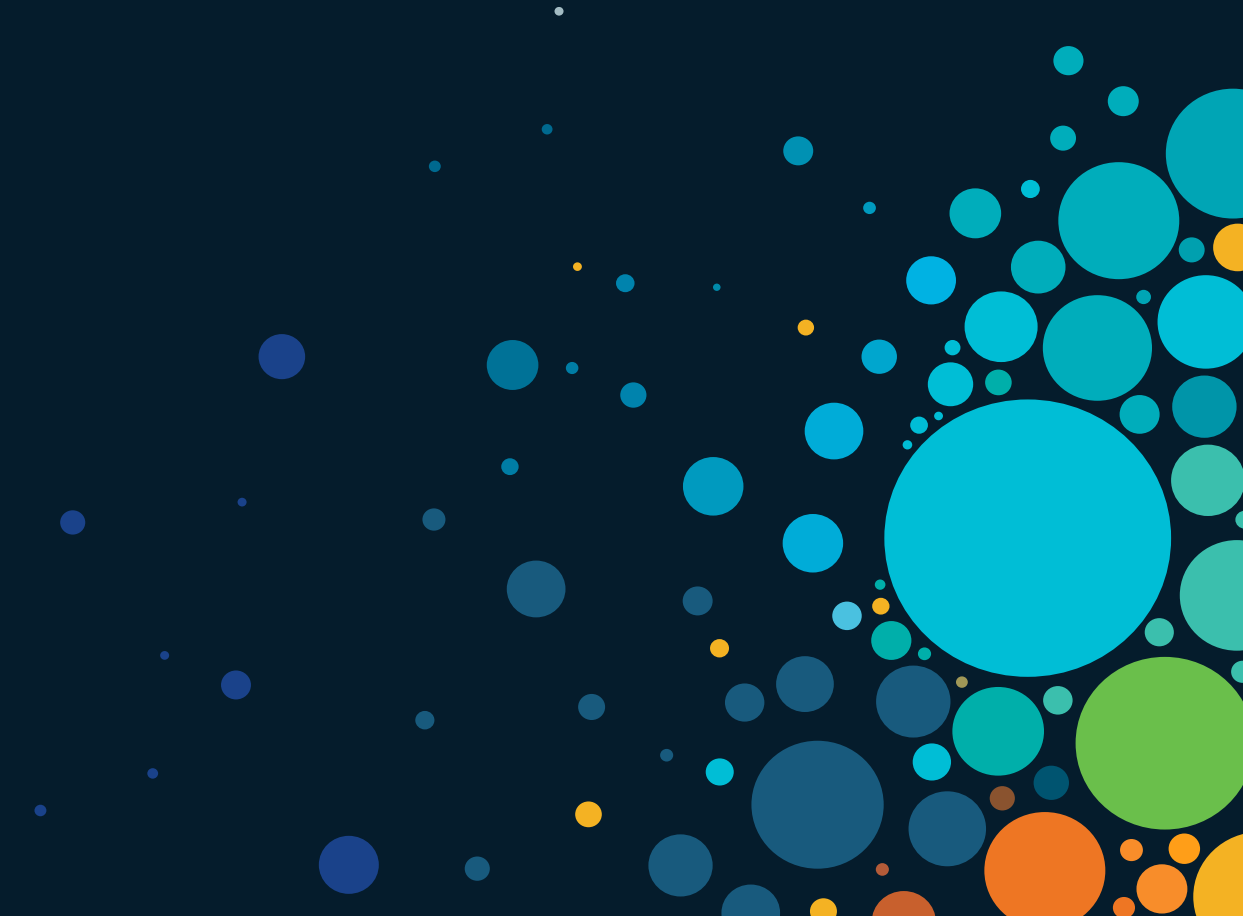

```
RP/0/0/CPU0:R3# show mpls forwarding labels 16004

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



Ti-LFA

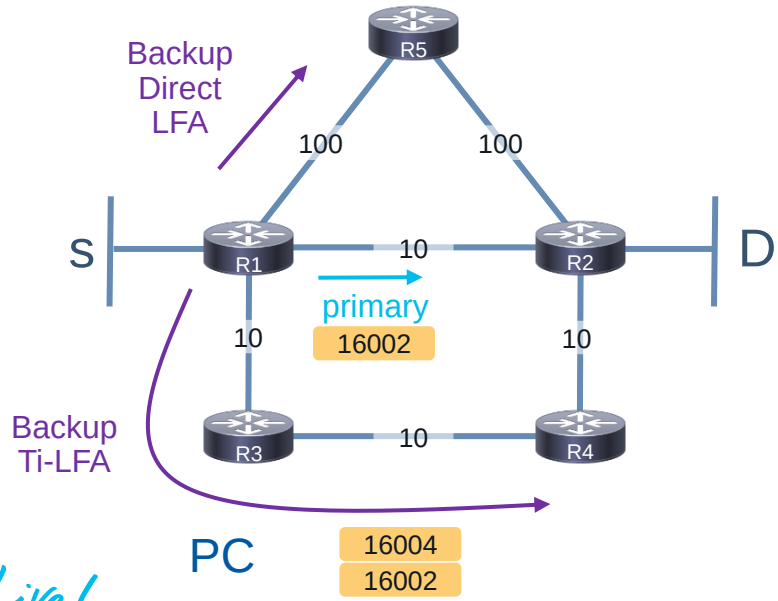


Ti-LFA

- It is LFA+
- No Signaling
- Link or node protection
- Protects IP and MPLS traffic
- Repair path can consist of Global Labels, and Adjacency SID Labels
- Makes LFA Topology Independent (Ti)
  - Algorithm, with similar tiebreakers
  - Using segments to force traffic over backup path
    - 100% coverage
  - Protected traffic is on Post-Convergence (PC) path
    - Avoiding another path move at regular convergence after failure
    - Not available with LFA

Ti-LFA Example

- Link State routing provides full topology view
- MPLS label stack can force the traffic to go anywhere
  - Without risk of (micro-)loop

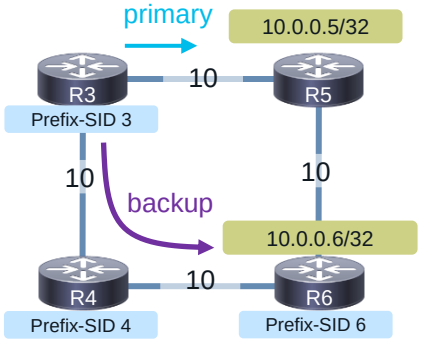


Ti-LFA protection always works by pushing extra MPLS labels
- Any number; rarely more than 2
- Any combination of Prefix SID or Adjacency SID labels

With Ti-LFA:
Ti-LFA uses PC path

Ti-LFA Example

Single-Segment Example



RIB

IP route, but MPLS forwarding entries are also protected

```
RP/0/RP0/CPU0:R3# show route 10.0.0.5/32

Routing entry for 10.0.0.5/32
  Known via "isis 1", distance 115, metric 10, labeled SR, type level-2
  Installed Apr 26 13:59:29.323 for 3d00h
  Routing Descriptor Blocks
    10.3.4.4, from 10.0.0.5, via GigabitEthernet0/0/0/2, Backup (TI-LFA)
      Repair Node(s): 10.0.0.6
      Route metric is 30
    10.3.5.5, from 10.0.0.5, via GigabitEthernet0/0/0/1, Protected
      Route metric is 10
  No advertising protos.
```

ISIS

```
RP/0/RP0/CPU0:R3# show isis fast-reroute 10.0.0.5/32

L2 10.0.0.5/32 [10/115]
     via 10.3.5.5, GigabitEthernet0/0/0/1, R5, SRGB Base: 16000, Weight: 0
       Backup path: TI-LFA (link), via 10.3.4.4, GigabitEthernet0/0/0/2 R4, SRGB Base: 16000, Weight: 0, Metric: 30
         P node: R6.00 [10.0.0.6], Label: 16006
         Prefix label: 16005
         Backup-src: R5.00
```

one additional label

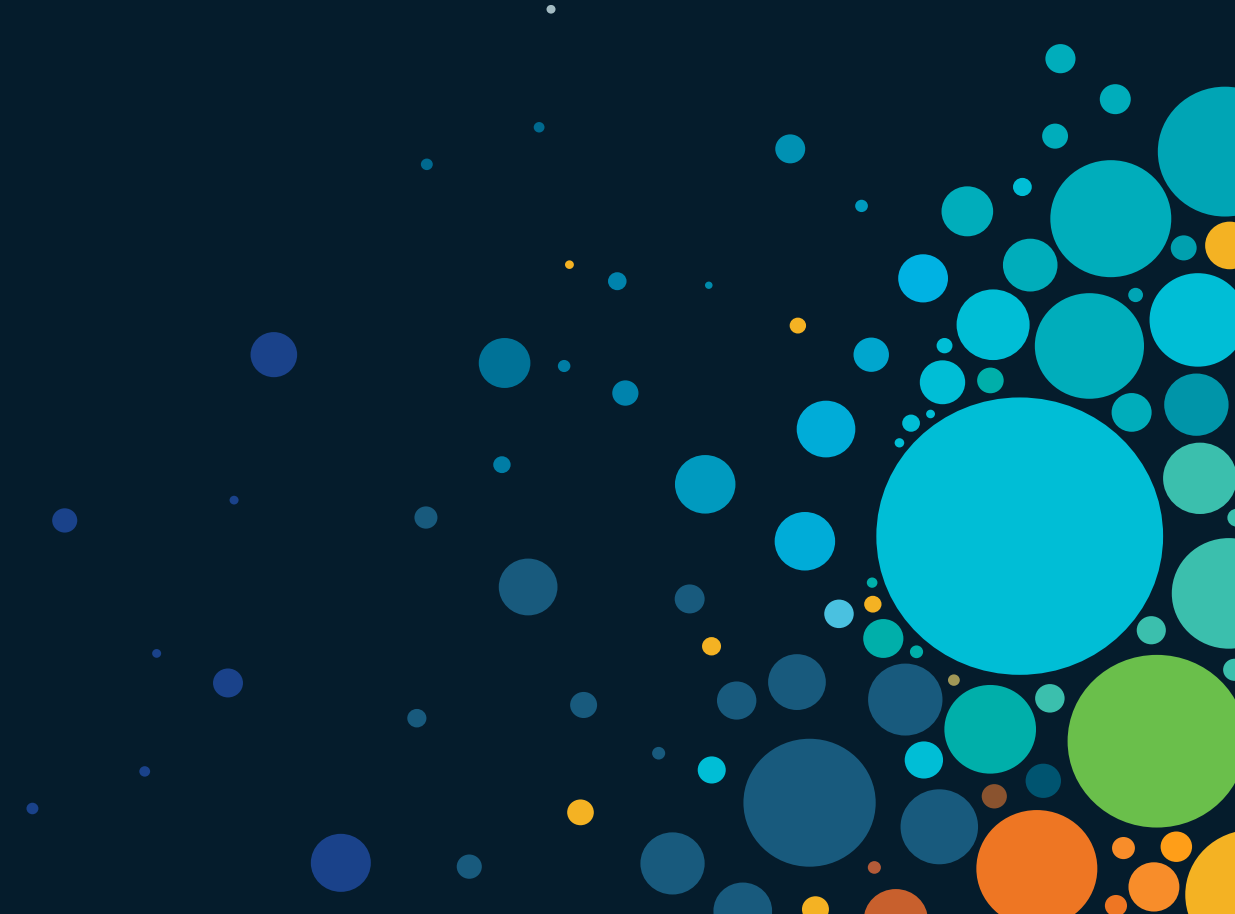
FIB

```
RP/0/RP0/CPU0:R3# show cef 10.0.0.5/32
10.0.0.5/32, version 212, labeled SR
 remote adjacency to GigabitEthernet0/0/0/1
 Prefix Len 32, traffic index 0, precedence n/a, priority 1
   via 10.3.4.4/32, GigabitEthernet0/0/0/2, 17 dependencies, weight 0, class 0, backup (TI-LFA) [flags 0xb00]
    path-idx 0 NHID 0x0 [0xf1244a0 0x0]
    next hop 10.3.4.4/32, Repair Node(s): 10.0.0.6
     local label 16005      labels imposed {16006 16005}
   via 10.3.5.5/32, GigabitEthernet0/0/0/1, 17 dependencies, weight 0, class 0, protected [flags 0x400]
    path-idx 1 bkup-idx 0 NHID 0x0 [0xf29e070 0xf29d0b0]
    next hop 10.3.5.5/32
     local label 16005      labels imposed {ImplNull}
```

repair node

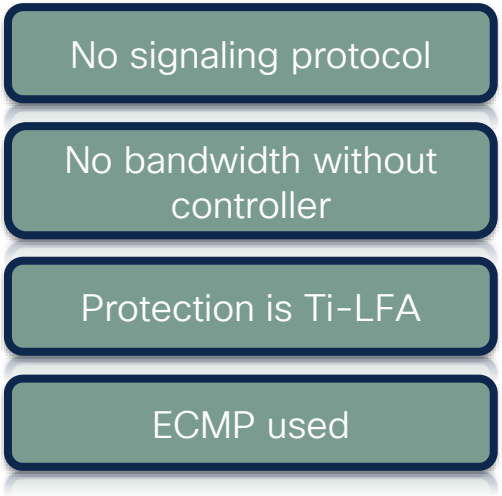
one additional label

SR Policy



What is an SR Policy?

- Simple, automated, and scalable
  - No core state: state in the packet header
  - Traffic engineered “SR Policy”
  - No headend a-priori configuration: on-demand policy instantiation
  - No headend a-priori steering: on-demand-steering
- Dynamic or explicit path
  - Explicit-path is defined as list of segments:
    - All hops are IP addresses (link/node = loopback)
    - All hops are MPLS labels
    - Mix of IP addresses/MPLS labels
- PCE/PCC is possible (multi-domain)



Definition of a 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 endpoint can be specified as an IPv4 or IPv6 address

- SR Policy can be defined on the head end or on the SR-PCE

SR Policy Breakdown

- Candidate path with preference
  - Higher preference is preferred, default = 100
- Explicit or dynamic path
  - Dynamic = calculated by PCC or PCE
  - Explicit = segment-list (labels) defined on PCC or PCE!
- Metric
  - Type = IGP, hopcount, latency, TE
  - Margin
- Constraints
  - Affinity
  - Disjoint-path
  - Bounds (cumulative metric)
  - Segments (protection type or SID algorithm)

SR Policy Examples

```
segment-routing
 traffic-eng
  logging
   policy status
  !
  segment-list explicit-to-ABR-1
   index 5 address ipv4 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
```

```
segment-routing
 traffic-eng
 !
 policy policy-1
   color 1000 end-point ipv4 10.0.0.2
   candidate-paths
    preference 100
     dynamic
      pcep
      !
      metric
       type te
```

```
segment-routing
 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
```

```
segment-routing
 traffic-eng
  policy policy-1
   candidate-paths
    preference 100
     dynamic
      metric
       sid-limit 5
       margin absolute 100
```

```
segment-routing
 traffic-eng
  policy POLICY_1
   end-point ipv4 1-.1.1.3 color 1
    candidate-paths
     preference 50
     dynamic mpls pce
      unprotected
      invalidation drop
```

```
segment-routing
 traffic-eng
  policy test
   candidate-paths
    preference 100
     constraints
      disjoint-path group-id 100 type {link | node| srlg | srlg-node}
```

```
segment-routing
 traffic-eng
  policy test
   candidate-paths
    preference 100
     constraints
      affinity {exclude-any | exclude-group | exclude-item | include-all| include-any}
```

```
segment-routing
 traffic-eng
  policy test
   candidate-paths
    preference 100
     constraints
      bounds cumulative type {hopcount | igp | latency | te}
```

Binding-SID

- Head end receives a packet with Binding Segment label and steers packet into SR policy
- Binding-SID is incoming label in LFIB
- Binding SID is automatically associated with every SR policy (overwritten if configured)

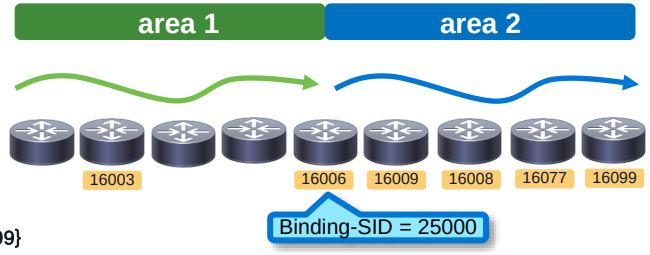
```
segment-routing
 global-block 16000 23999
 traffic-eng
  logging
   policy status
  !
  policy to-ABR1
   binding-sid mpls 1000
   color 1000 end-point ipv4 10.0.0.5
   candidate-paths
   …
```

```
FIB @ headend
Incoming label: 1000
Action: pop and push <label stack of policy>
```

Hierarchy of SR Policies

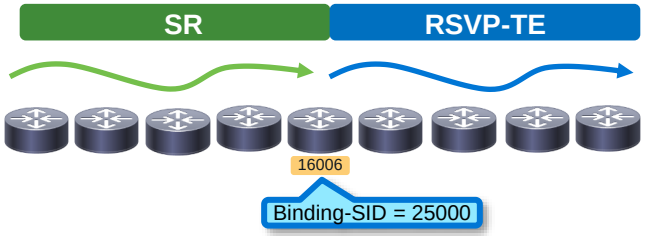
Shorter SID list and isolation of churn

{16003, 16006, 25000}
instead of
{16003, 16006, 16009, 16008, 16077, 16099}



SR Policy to RSVP-TE

{16006, 25000}



Automated Steering

per-destination

automatically steers service routes on their matching (color + endpoint) SR Policy

- Static into policy is possible

```
segment-routing
 traffic-eng
  policy c100
   color 100 end-point ipv4 10.0.0.3
   candidate-paths
    preference 100
     dynamic
      metric
       type igp
      !
```

Matching color and endpoint

```
router static
 address-family ipv4 unicast
  10.0.0.12/32 sr-policy srte_c_100_ep_10.0.0.6
```

per-flow

automatically steers service routes on their matching (color + endpoint) SR Policy per Forward Class

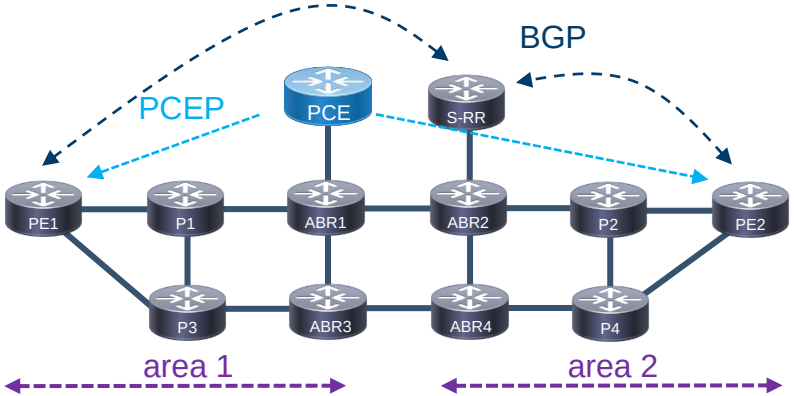
- Forward Class is internal to router
- COS can be mapped to FC
- Up to 8 ways

```
segment-routing
 traffic-eng
  policy c100
   color 100 end-point ipv4 10.0.0.3
   candidate-paths
    preference 100
     dynamic
      metric
       type igp
      !
policy c101
   color 101 end-point ipv4 10.0.0.3
   candidate-paths
    preference 100
     dynamic
      metric
       type hopcount
!
policy c1000
   color 1000 end-point ipv4 10.0.0.3
   candidate-paths
    preference 100
     per-flow
      forward-class 0 color 100
      forward-class 1 color 101
```

Each color is separately defined for same endpoint

On Demand Next Hop (ODN)

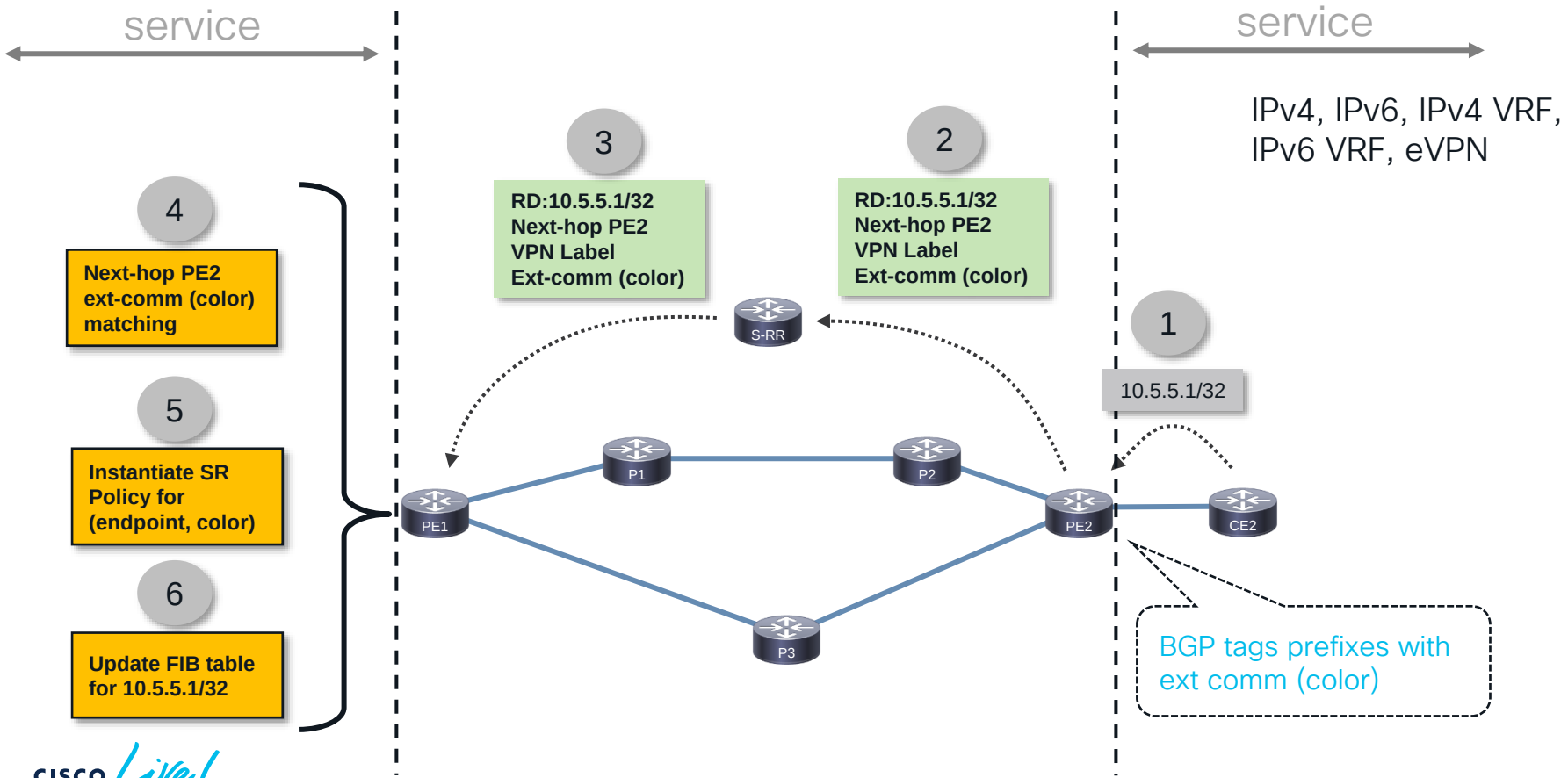
ODN



Automated SR Policy
Inter-AS & Seamless MPLS: no need for BGP-LU (RFC3107)
SLA-aware BGP service

- On-demand Next-hop automates and simplifies the service head end configuration
  - No SR Policy config on the head end router
  - No complex/explicit steering on the service head end for the service
    - For example: no autoroute-announce, no static routes
  - No need for full path knowledge on head end router is SR-PCE is used
- The SR Policies deployed when needed
  - The learning of the service route, initiates the SR Policy, and traffic-to-SR Policy mapping
  - Example of a service route: vpnv4 route

ODN Behavior



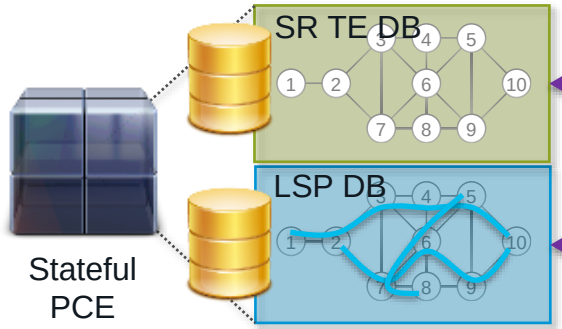
SR-PCE

Path Computation Element (PCE) (controller)

PCEP

Path Computation Client (PCC) (router)

2 Computation



1 PCReport

SR policy status: delegate flag set

End points

Constraints:

Metric (IGP, TE, hop count)

Affinities

Path setup type: SR

Segment list: empty

3 PCUpdate

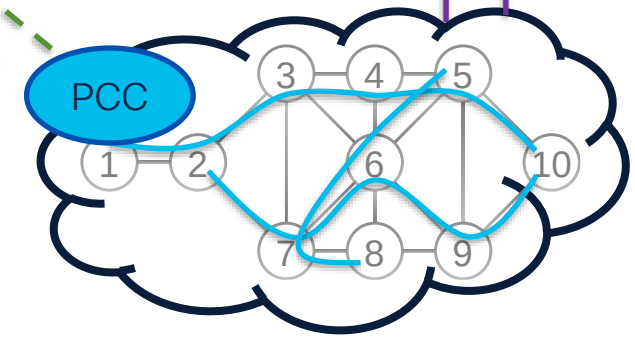
providing

ERO (list of segments)

metric

5 PCReport

4 SR policy programming



SR Operations, Administration, and Maintenance (OAM)

SR MPLS 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>
```

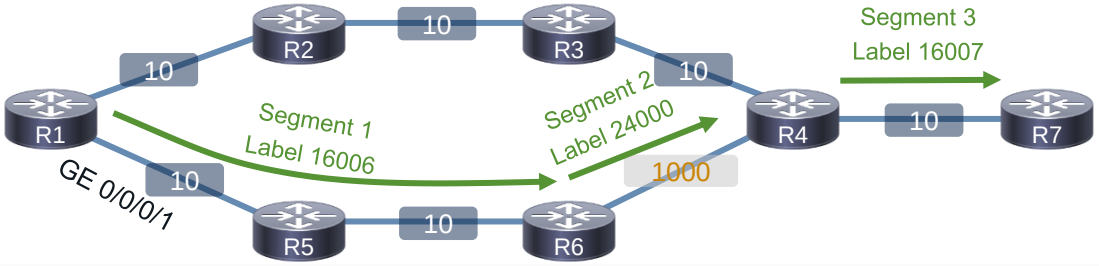
```
RP/0/RP0/CPU0:PE1# trace sr-mpls policy ?
  binding-sid  Specify the binding-sid of the SR
policy
  name         Specify the name of the SR policy
```

NIL-FEC

- Nil FEC (defined in RFC4379), specifies that no explicit FEC (Control Plane) is associated with the label
- Typically used to carry labels in reserved range (explicit-null or router alert) for diagnostic purpose
- Ping and traceroute
- But very powerful tool to check any combination of segments on any path!
- Does not carry any information to identify the intended target
  - The packet may be forwarded wrongly somewhere, but still make it
  - No control plane validation is performed at originator or responder
- This was an interim solution
- Can force traffic over non-least cost path

NIL-FEC Example

User specifies:
- Outgoing label stack (one or more labels)
- Outgoing interface
- Next-hop interface address



Specify outgoing interface and next hop

```
RP/0/0/CPU0:R1# trace mpls nil-fec labels 16006,24000,16007 output interface gigabitEthernet 0/0/0/1 nexthop 10.1.5.5

Tracing MPLS Label Switched Path with Nil FEC with labels [16006,28097,16007], timeout is 2 seconds

Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
  'L' - labeled output interface, 'B' - unlabeled output interface,
  'D' - DS Map mismatch, 'F' - no FEC mapping, 'f' - FEC mismatch,
  'M' - malformed request, 'm' - unsupported tlvs, 'N' - no rx label,
  'P' - no rx intf label prot, 'p' - premature termination of LSP,
  'R' - transit router, 'I' - unknown upstream index,
  'X' - unknown return code, 'x' - return code 0

Type escape sequence to abort.

 0 10.1.5.1 MRU 1500 [Labels: 16006/24000/16007/explicit-null Exp: 0/0/0/0]
L 1 10.1.5.5 MRU 1500 [Labels: implicit-null/24000/16007/explicit-null Exp: 0/0/0/0] 12 ms
L 2 10.5.6.6 MRU 1500 [Labels: implicit-null/16007/explicit-null Exp: 0/0/0] 19 ms
L 3 10.4.6.4 MRU 1500 [Labels: implicit-null/explicit-null Exp: 0/0] 13 ms
! 4 10.4.7.7 41 ms
```

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

24000 is adj-SID label from R6 to R4

Key Takeaways

- SR is simpler and easier to troubleshoot than LDP or RSVP-TE
- 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

Technical Session Surveys

- Attendees who fill out a minimum of four session surveys and the overall event survey will get Cisco Live branded socks!
- Attendees will also earn 100 points in the Cisco Live Game for every survey completed.
- These points help you get on the leaderboard and increase your chances of winning daily and grand prizes.

Cisco Learning and Certifications

From technology training and team development to Cisco certifications and learning plans, let us help you empower your business and career. www.cisco.com/go/certs

Pay for Learning with Cisco Learning Credits

(CLCs) are prepaid training vouchers redeemed directly with Cisco.

Learn

Cisco U.
IT learning hub that guides teams and learners toward their goals

Cisco Digital Learning
Subscription-based product, technology, and certification training

Cisco Modeling Labs
Network simulation platform for design, testing, and troubleshooting

Cisco Learning Network
Resource community portal for certifications and learning

Train

Cisco Training Bootcamps
Intensive team & individual automation and technology training programs

Cisco Learning Partner Program
Authorized training partners supporting Cisco technology and career certifications

Cisco Instructor-led and Virtual Instructor-led training
Accelerated curriculum of product, technology, and certification courses

Certify

Cisco Certifications and Specialist Certifications
Award-winning certification program empowers students and IT Professionals to advance their technical careers

Cisco Guided Study Groups
180-day certification prep program with learning and support

Cisco Continuing Education Program
Recertification training options for Cisco certified individuals

Here at the event? Visit us at **The Learning and Certifications lounge at the World of Solutions**

Continue your education

- Visit the Cisco Showcase for related demos
- Book your one-on-one Meet the Engineer meeting
- Attend the interactive education with DevNet, Capture the Flag, and Walk-in Labs
- Visit the On-Demand Library for more sessions at www.CiscoLive.com/on-demand

cisco Live!

Thank you


cisco Live!
ALL IN