Possibilities

#CiscoLive

Intro to Segment Routing

Routing Protocol for SDN

Vinit Jain, Technical Leader
@vinugenie

DGTL-BRKRST-2124

cisco Live!

#CiscoLive

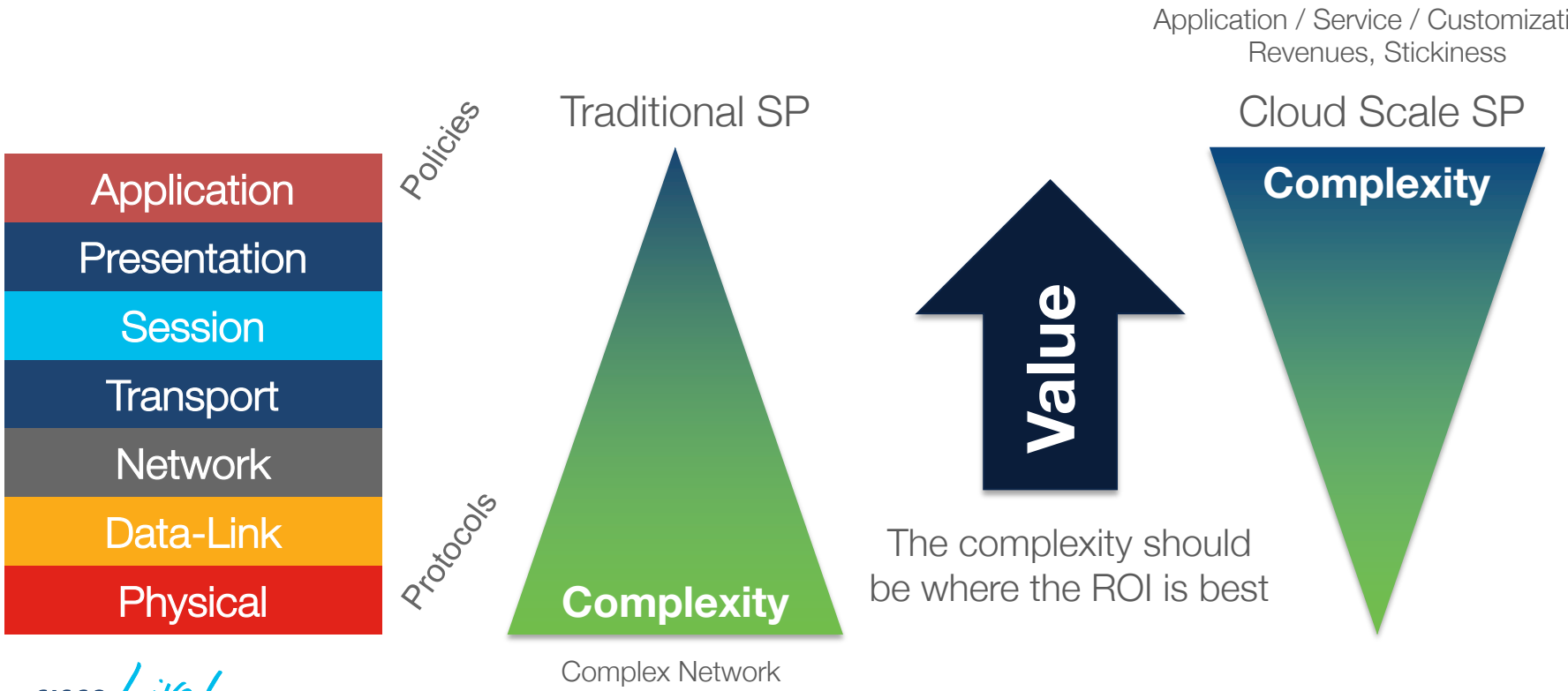
cisco

Agenda

- Introduction
- Technology Overview
- LDP to SR Migration
- Control Plane & Data Plane
- Traffic Protection - TI-LFA
- SRTE and SRTE Use Cases

Setting up the stage

SP Disruption: Complexity vs. Value



Infrastructure Simplification and Convergence Areas

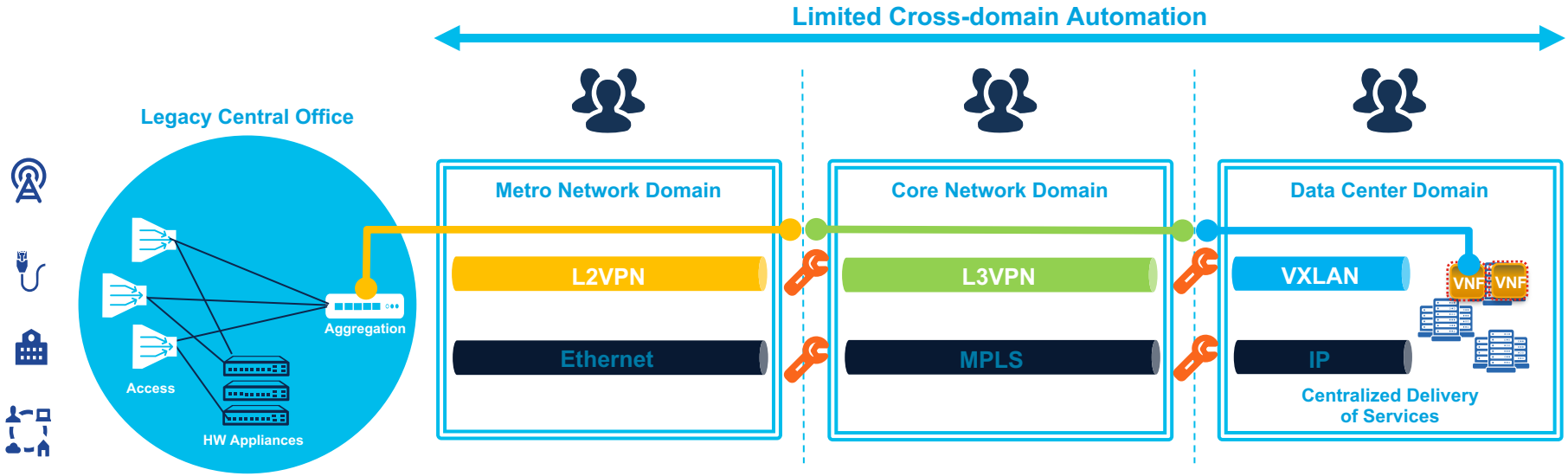
Delayering	Unified Infrastructure	Horizontal Integration
IP	Data	Access
ATM/Ethernet	SDH replacement	Aggregation
SDH/OTN	Video	Edge
WDM	Fixed Mobile	Core and DC

Transformation

SDN Driven Automation and Virtualization

IP WDM	Fixed Mobile Data Video Circuit	End to End IP Segment Routing

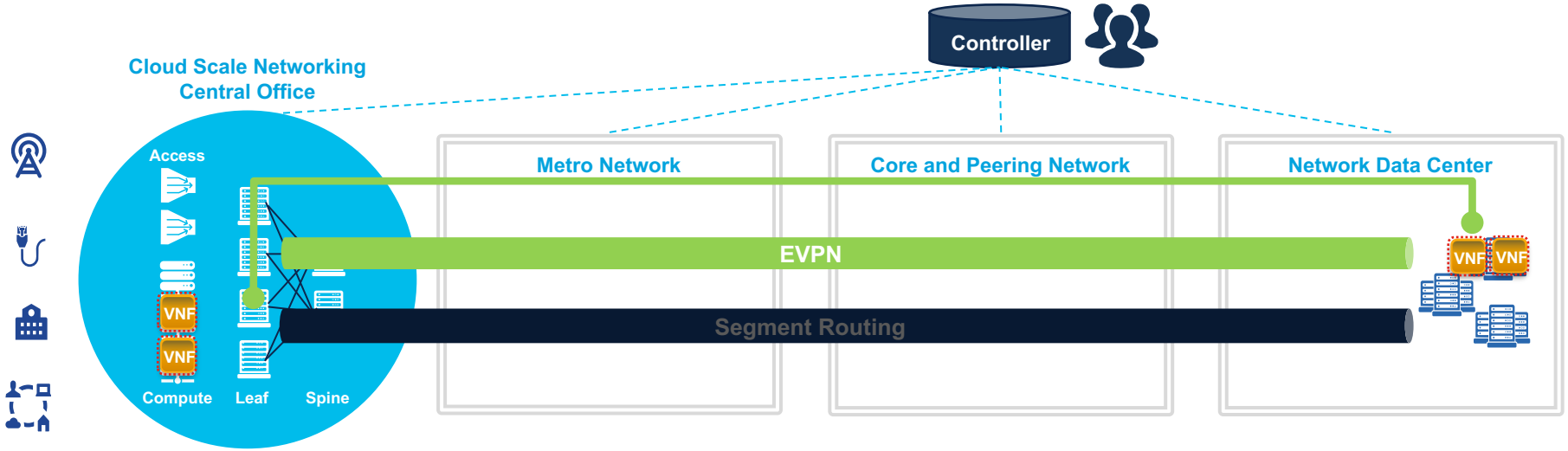
Challenges of Todays Service Creation



E2E service provisioning is lengthy and complex:

- ✓ Multiple network domains under different management teams
- ✓ Manual operations
- ✓ Heterogeneous Underlay and Overlay networks

Unified “Stateless Fabric” for Service Creation



Simplify — Unified underlay and overlay networks with segment routing and EVPN

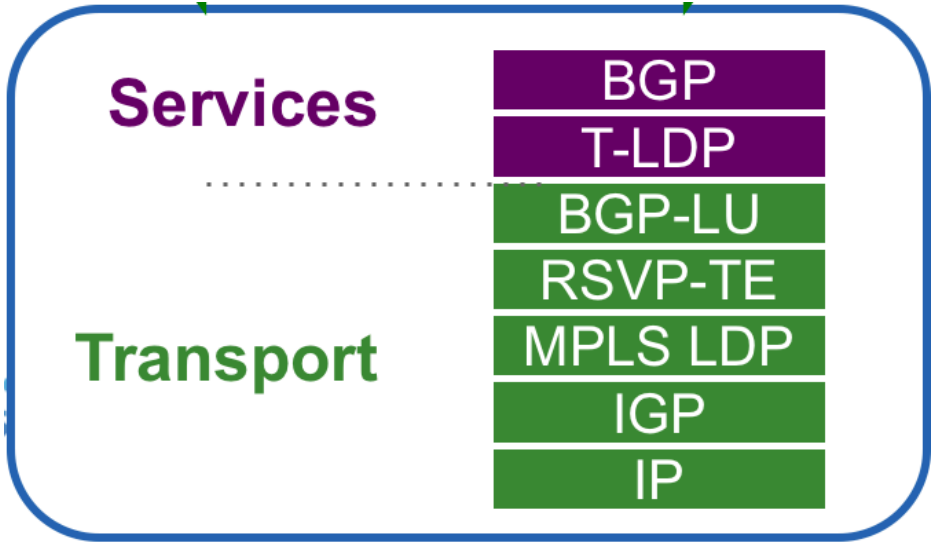
Automate — E2E Cross-domain automation with model-driven programmability and streaming telemetry

Virtualize — Transform the CO into a data center to enable distributed service delivery and speed up service creation

Network Transport Evolution

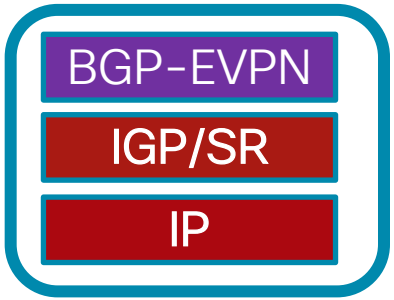
Simplify - Optimize - Enable

Unified MPLS



SR

Enabled Transport



Do more with less !!

Segment Routing Standardization

- IETF standardization in SPRING working group
- First RFC document – RFC 7855 (May 2016)
- Protocol extensions progressing in multiple groups
  - IS-IS
  - OSPF
  - PCE
  - IDR
  - 6MAN
  - BESS
- Broad vendor support
- Strong customer adoption and support
  - WEB, SP, Enterprise

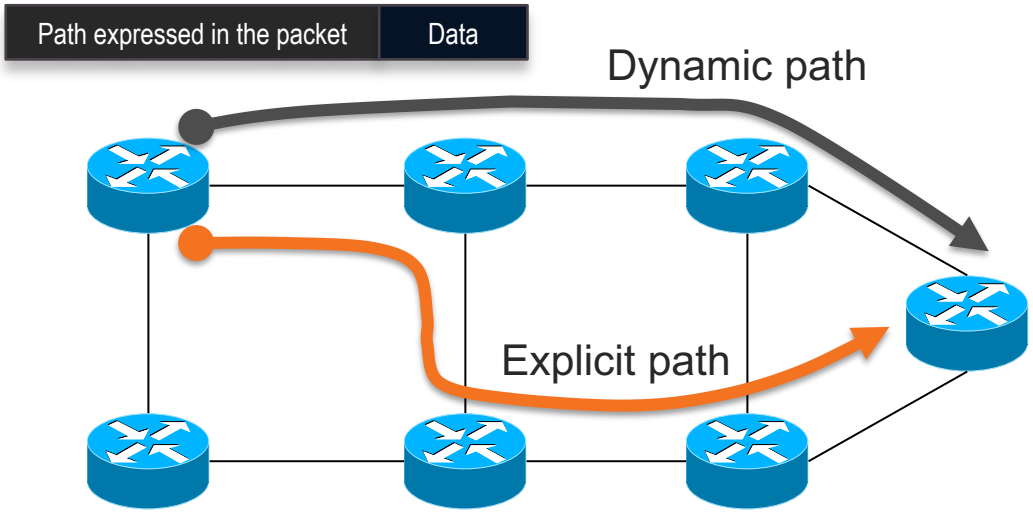
Sample IETF Documents
Problem Statement and Requirements (RFC 7855)
Segment Routing Architecture (draft-ietf-spring-segment-routing)
IPv6 SPRING Use Cases (draft-ietf-spring-ipv6-use-cases)
Segment Routing with MPLS data plane (draft-ietf-spring-segment-routing-mpls)
Topology Independent Fast Reroute using Segment Routing (draft-bashandy-rtgwg-segment-routing-ti-lfa)
IS-IS Extensions for Segment Routing (draft-ietf-isis-segment-routing-extensions)
OSPF Extensions for Segment Routing (draft-ietf-ospf-segment-routing-extensions)
PCEP Extensions for Segment Routing (draft-ietf-pce-segment-routing)

Close to 40 IETF drafts in progress

Technology Overview

Segment Routing

An IP and MPLS source-routing architecture that seeks the **right balance** between **distributed intelligence** and **centralized optimization**



Paths options

- Dynamic (Headend computation)
- Explicit (Operator / Controller)

Control Plane

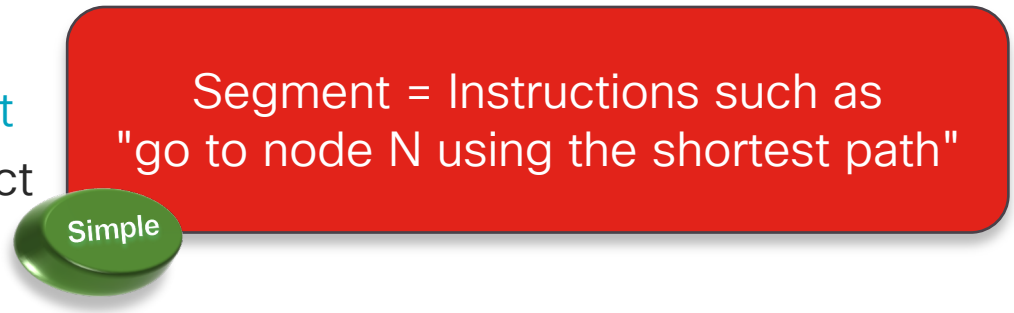
- Routing protocols with extensions (IS-IS,OSPF, BGP)
- SDN controller

Data Plane

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

Segment Routing

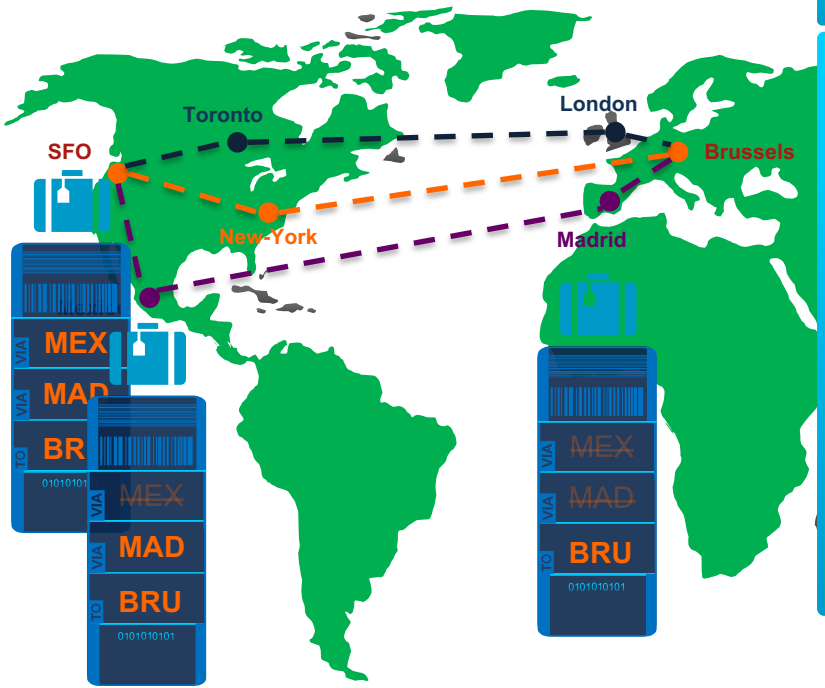
- Source Routing: the source chooses a path and encodes it in the packet header as an ordered list of segments
- Segment: an identifier for any type of instruction
  - Service
  - Context
  - Locator
  - IGP-based forwarding construct
  - BGP-based forwarding construct
  - Local value or Global Index



Segment Routing

Evolve MPLS with Segment Routing

Segment Routing



Mission – Route the luggage to Brussels via Mexico and Madrid

1. A unique and global luggage tag is attached to the luggage with the list of stops to the final destination
2. At each stop, the luggage is simply routed to the next hop listed on the luggage tag

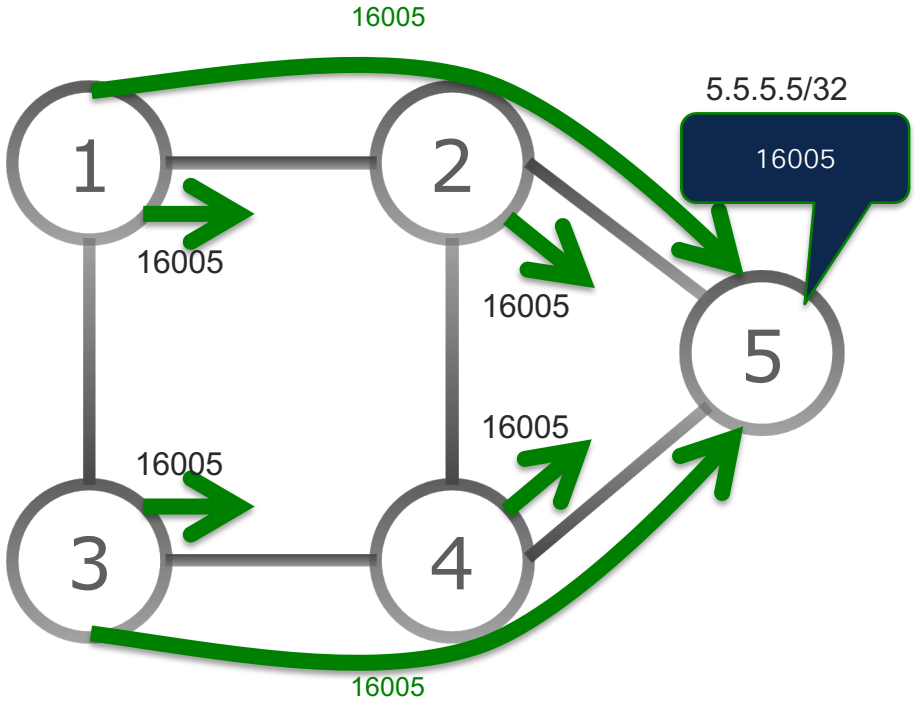
RESULT: **Path can be controlled**
Simple and scalable

Segment Routing – Forwarding Plane

- MPLS: an ordered list of segments is represented as a stack of labels
- IPv6: an ordered list of segments is encoded in a routing extension header
- This presentation: MPLS data plane
  - Segment → Label
  - Basic building blocks distributed by the IGP or BGP
- Two basic building blocks distributed by IGP
  - Prefix Segments
  - Adjacency Segments

IGP Prefix Segment

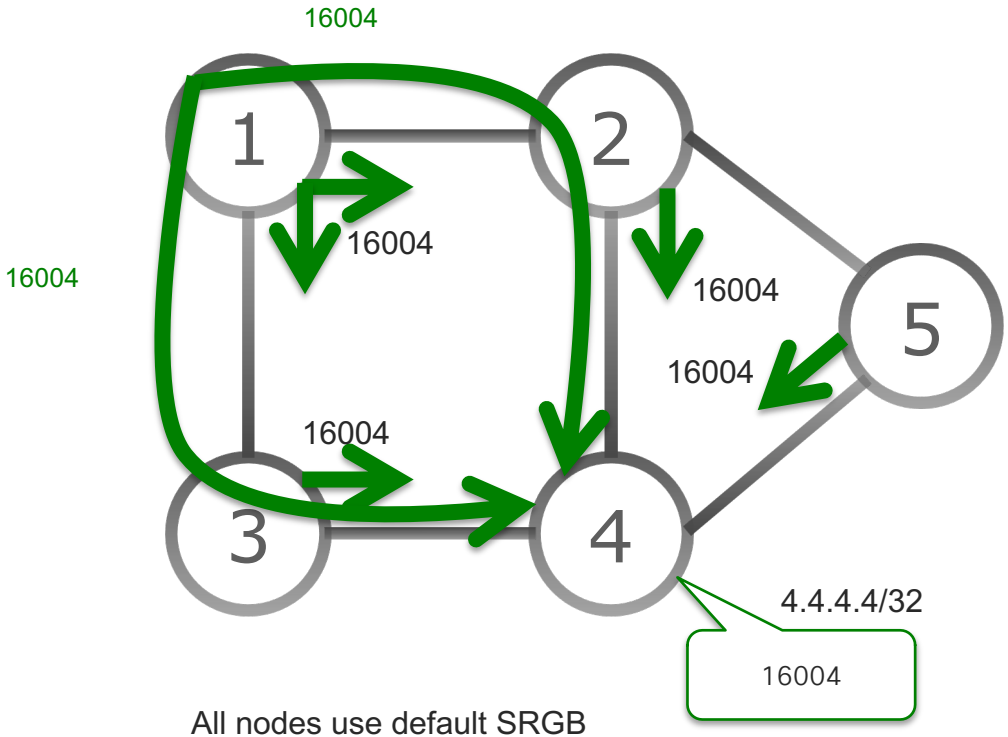
- Shortest-path to the IGP prefix
  - Equal Cost MultiPath (ECMP)-aware
- Global Segment
- Label = 16000 + Index
  - Advertised as index
- Distributed by ISIS/OSPF



All nodes use default SRGB
16,000 – 23,999

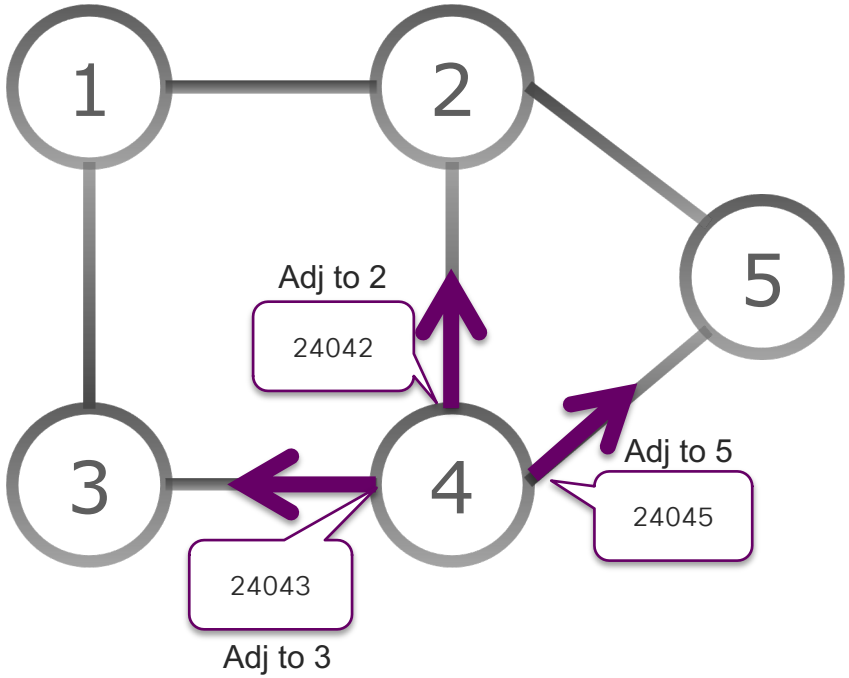
IGP Prefix Segment

- Shortest-path to the IGP prefix
  - Equal Cost MultiPath (ECMP)-aware
- Global Segment
- Label = 16000 + Index
  - Advertised as index
- Distributed by ISIS/OSPF



IGP Adjacency Segment

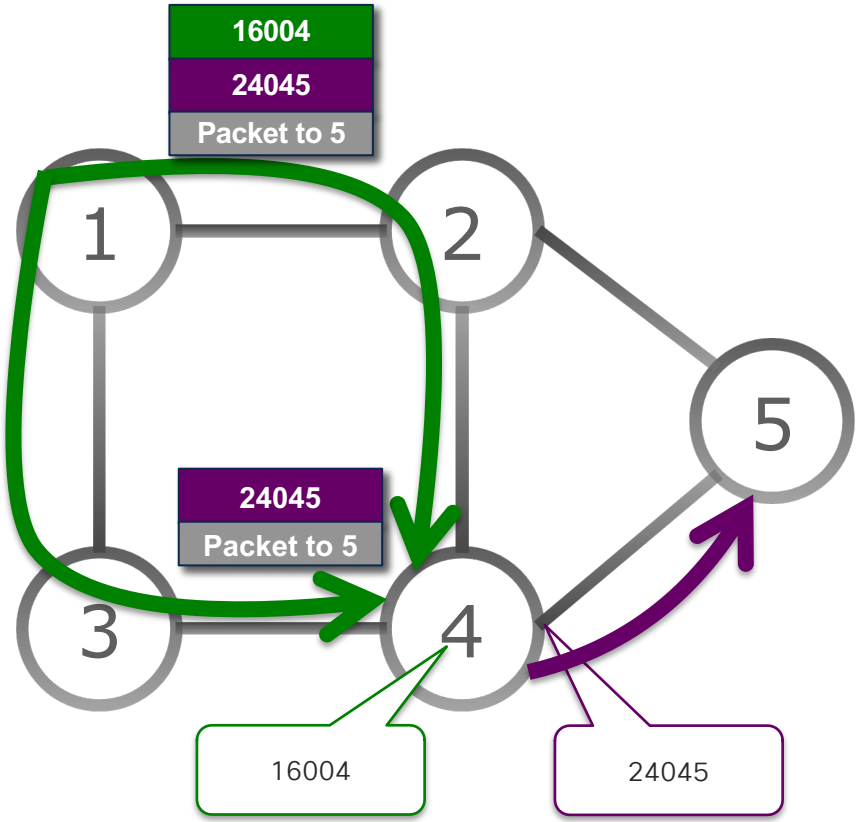
- Forward on the IGP adjacency
- Local Segment
- Advertised as label value
- Distributed by ISIS/OSPF



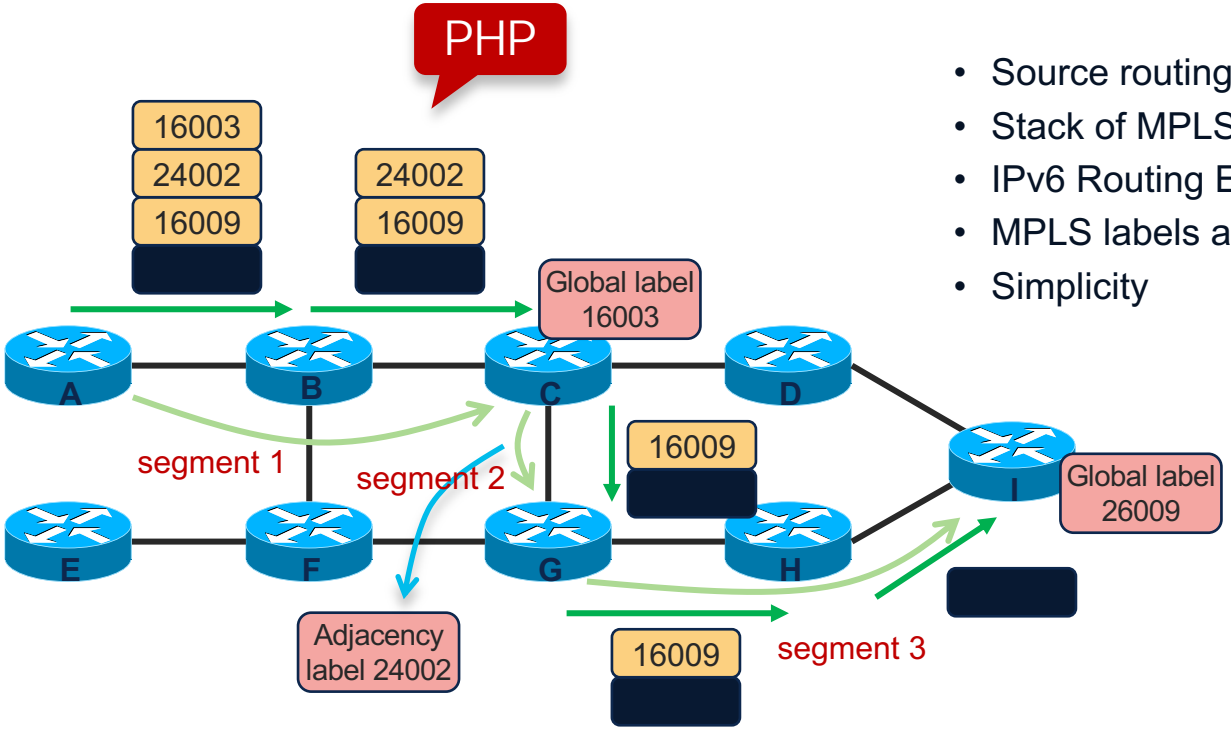
All nodes use default SRGB
16,000 – 23,999

Combining IGP Segments

- Steer traffic on any path through the network
- Path is specified by a stack of labels
- No path is signaled
- No per-flow state is created
- Single protocol: IS-IS or OSPF



Segment Routing – 3 Segments Example



- Source routing – ordered list of segments
- Stack of MPLS labels
- IPv6 Routing Extension
- MPLS labels are advertised by the IGP
- Simplicity

What happens if two devices have the same prefix SID?

Segment Routing Global Block (SRGB)

Segment Routing Global Block (SRGB)

- SRGB allocation based on Segment Routing Configuration
  - Default Range SRGB is 16000-23999
  - Dynamic Range starts at 16 (XE) or 24000 (XR)
  - If some labels are in use in the requested range SR_APP will periodically keep retrying to reserve the range
  - SR is disabled until range is reserved successfully
- A non-default SRGB can be configured
  - All protocols use the same SRGB
  - SRGB is allocated as a block of labels under control of SR-APP
- Modifying a SRGB configuration is disruptive for traffic
- Recommended to have same SRGB on all nodes

Segment Routing Global Block (SRGB)

```
IOS-XE
ONE(config)#segment-routing mpls
ONE(config-srmpls)#global-block 18000 19999
ONE(config-srmpls)#
```

Configure a non-default SRGB
18,000 – 19,999

Note “mpls” keyword. All config related to MPLS encap (for V4 or V6). In the future “ipv6 encap” may be available.

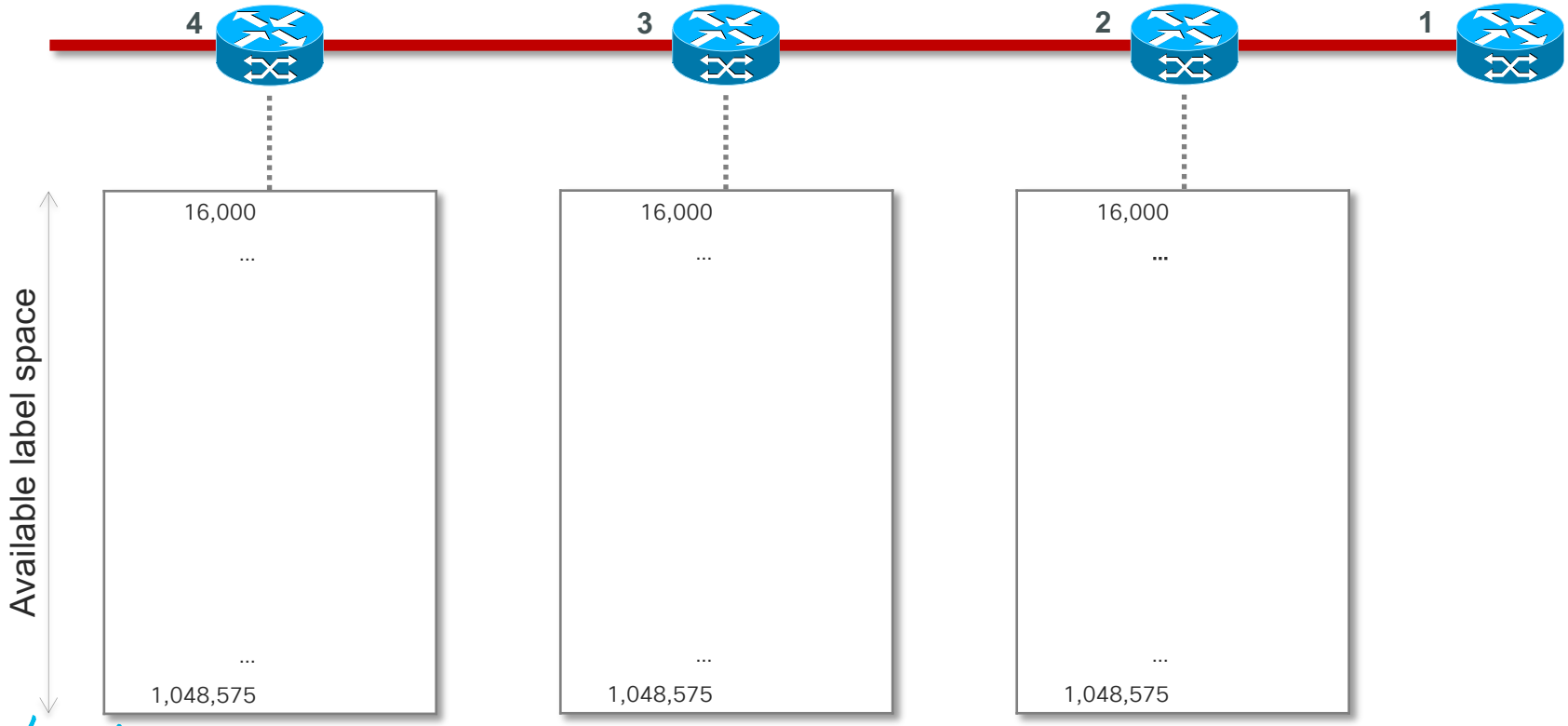
```
IOS-XR
RP/0/0/CPU0:XR-1(config)#segment-routing
RP/0/0/CPU0:XR-1(config-sr)#global-block 18000 19999
```

SRGB

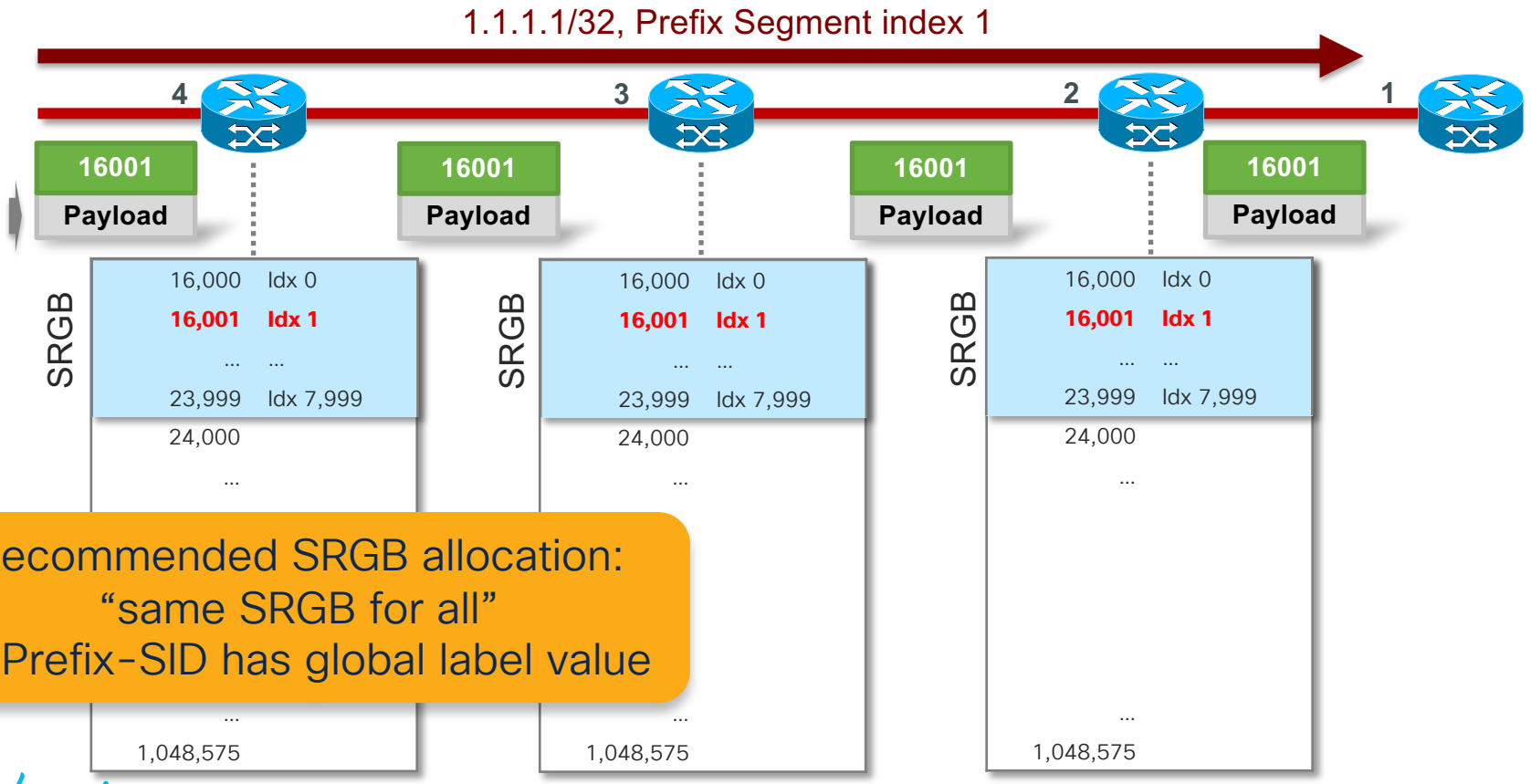
Modifying SRGB

- If SRGB allocation fails, no SR labels will be installed incl.
- IGP re-downloads all prefixes with new label values based on the new SRGB
- Disruptive !!
  - All labeled traffic will be dropped until IGP routes are re-downloaded with new labels

Segment Routing Global Block (SRGB)

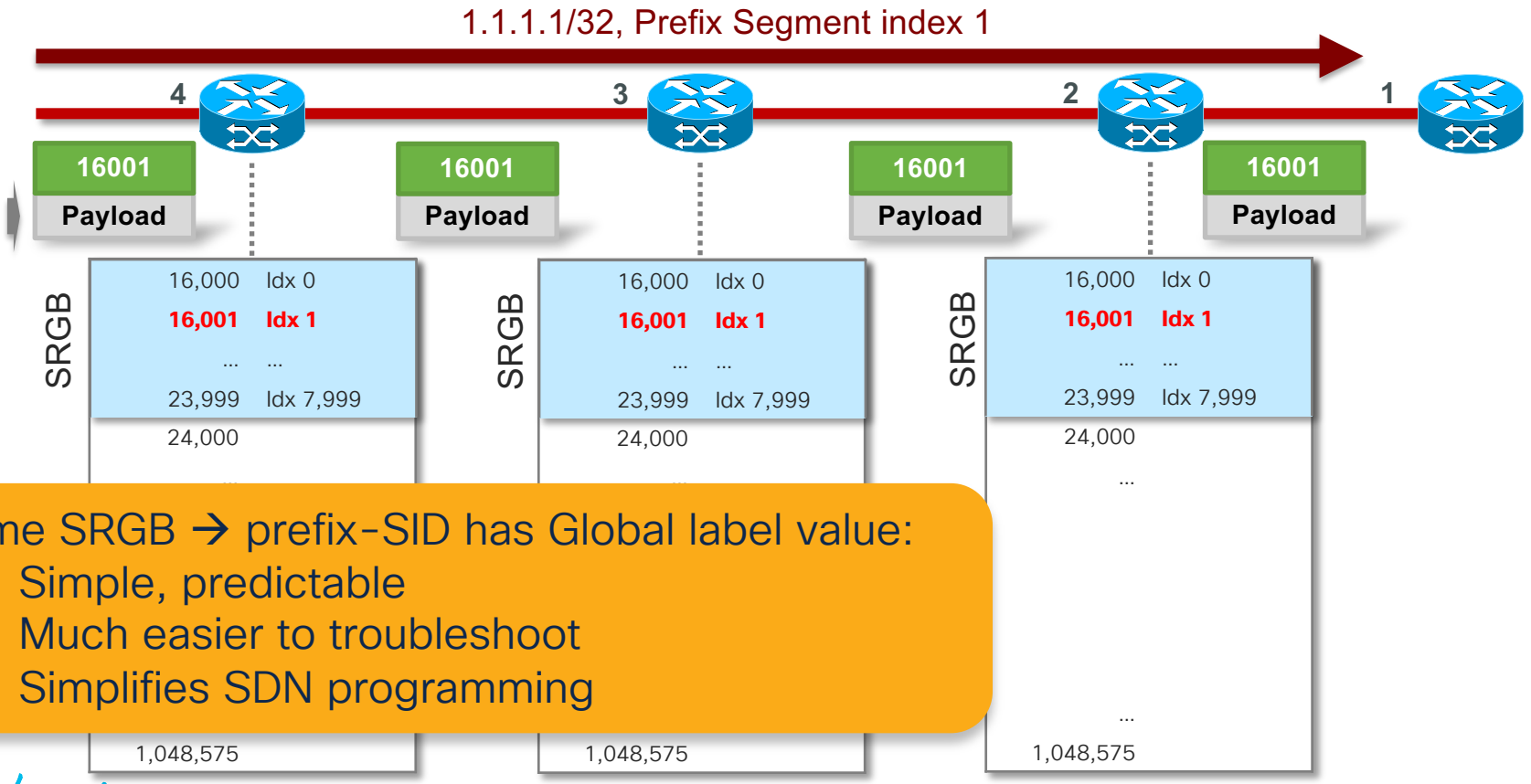


Recommended SRGB allocation



Recommended SRGB allocation:
"same SRGB for all"
→ Prefix-SID has global label value

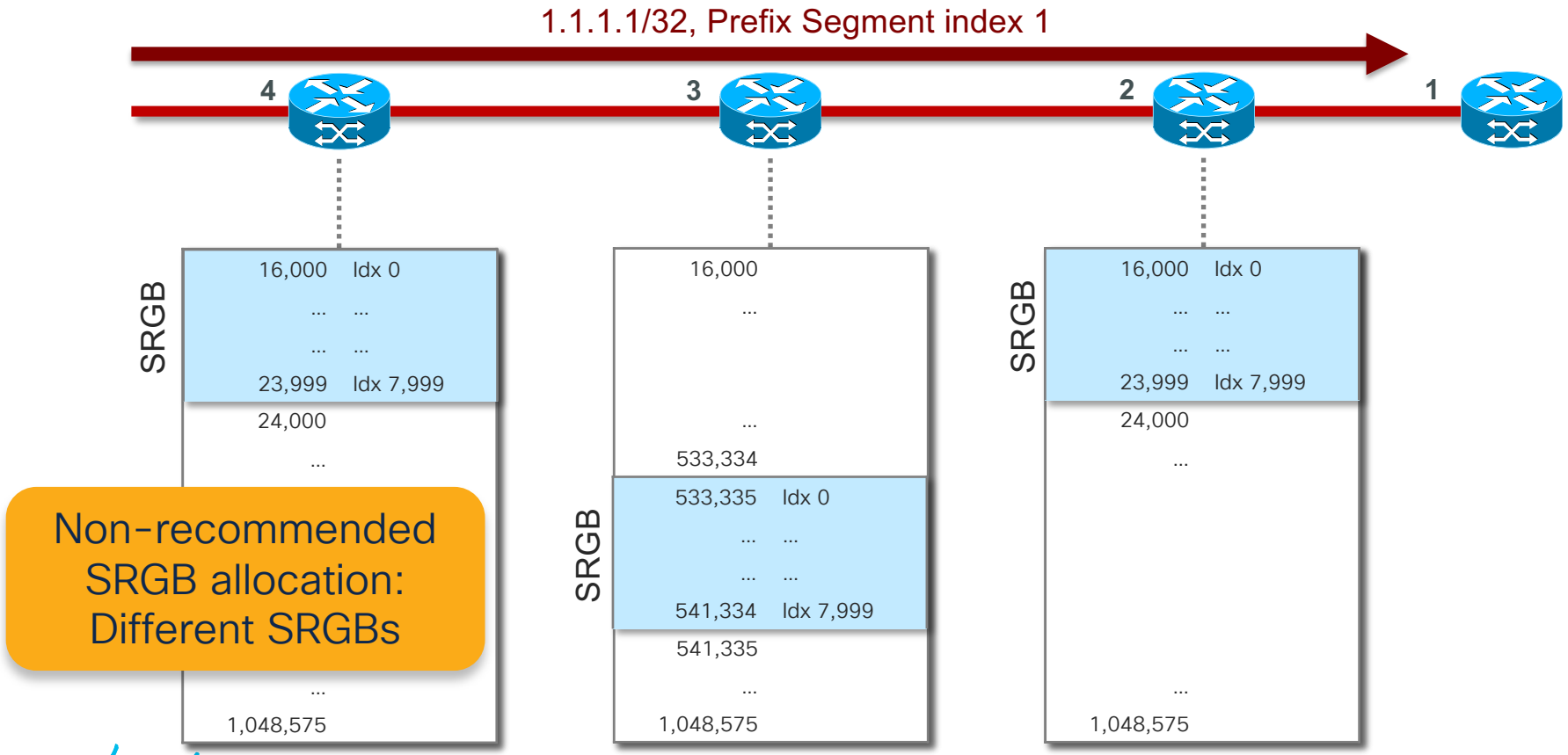
Recommended SRGB allocation



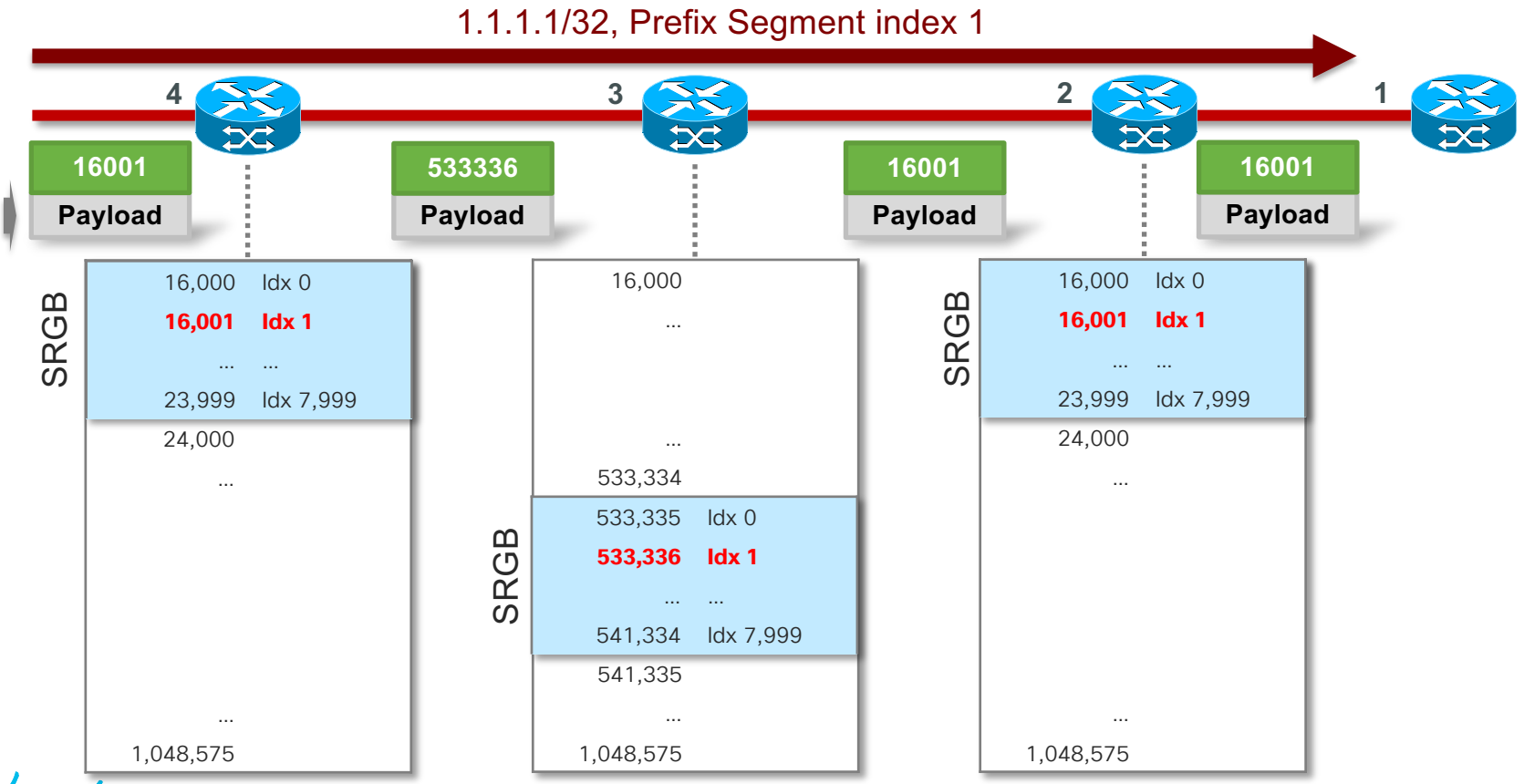
Same SRGB → prefix-SID has Global label value:

- Simple, predictable
- Much easier to troubleshoot
- Simplifies SDN programming

Not recommended, but possible SRGB allocation



Not recommended, but possible SRGB allocation

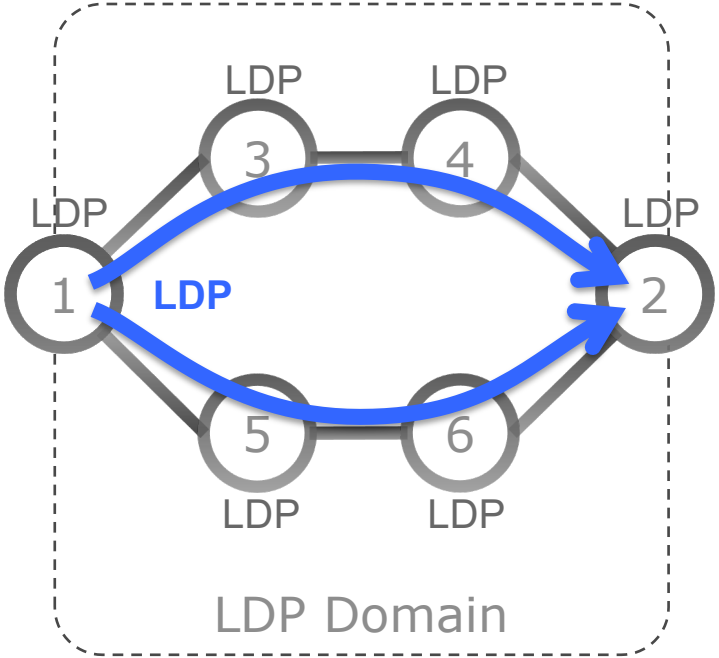


LDP to SR Migration

Simplest migration LDP to SR

Assumptions:
- all the nodes can be upgraded to SR
- all the services can be upgraded to SR

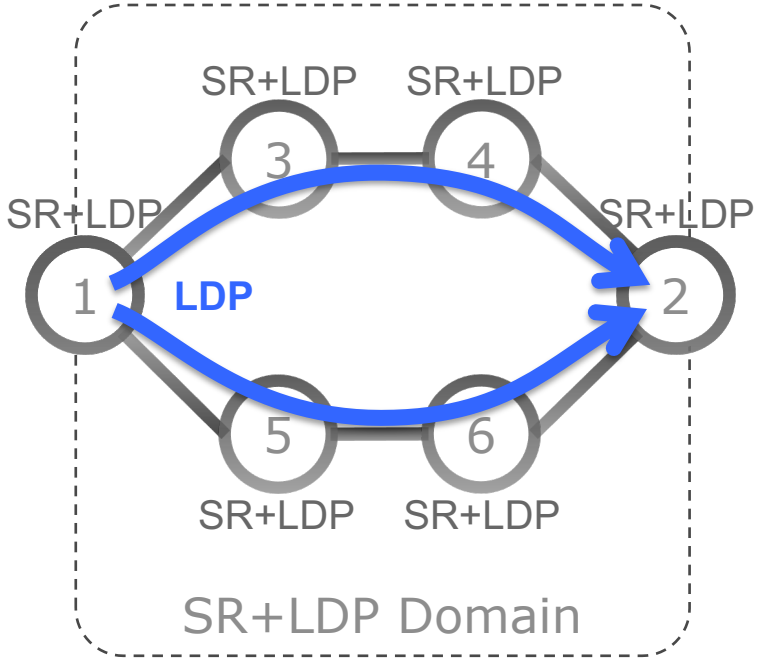
- Initial state: All nodes run LDP, not SR



Simplest migration LDP to SR

Assumptions:
- all the nodes can be upgraded to SR
- all the services can be upgraded to SR

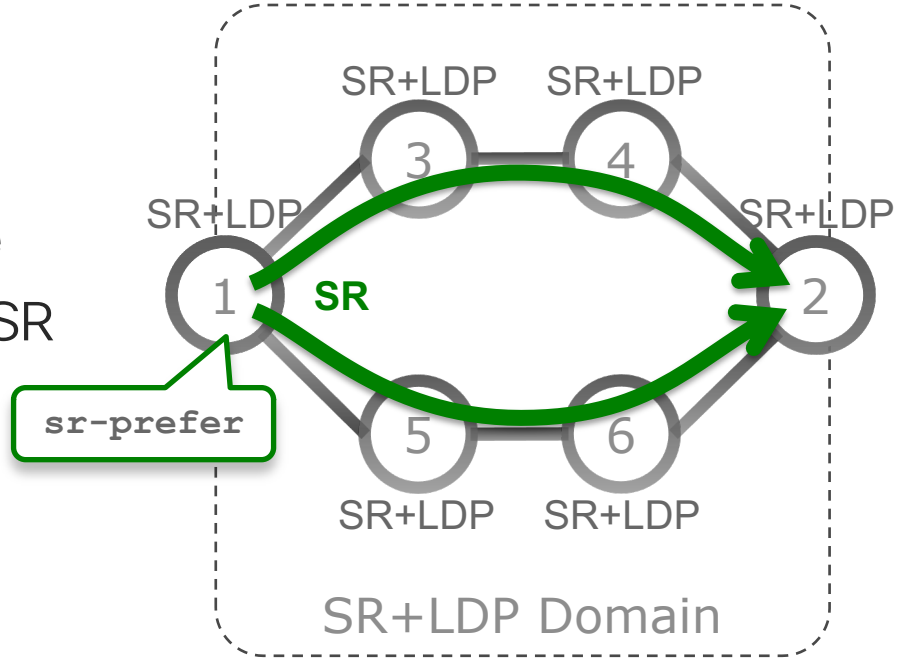
- Initial state: All nodes run LDP, not SR
- Step1: All nodes are upgraded to SR
  - In no particular order
  - leave default LDP label imposition preference



Simplest migration LDP to SR

Assumptions:
- all the nodes can be upgraded to SR
- all the services can be upgraded to SR

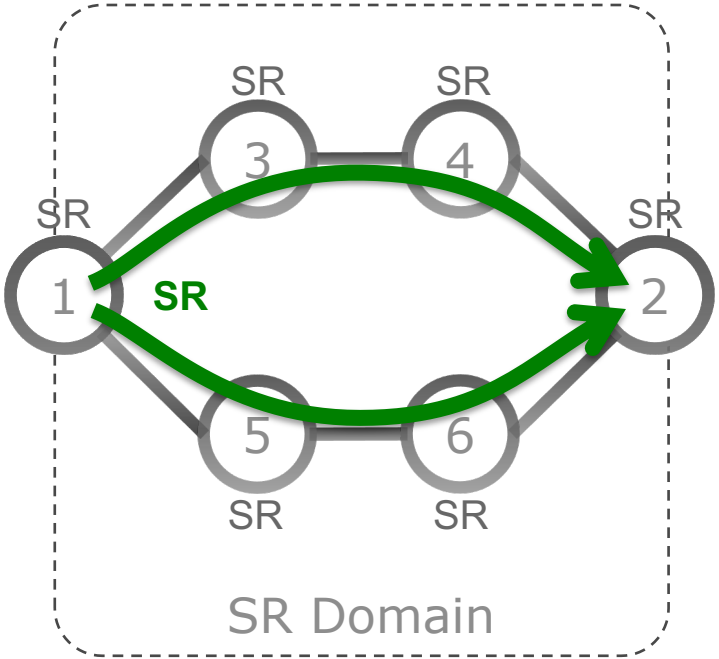
- Initial state: All nodes run LDP, not SR
- Step1: All nodes are upgraded to SR
  - In no particular order
  - leave default LDP label imposition preference
- Step2: All PEs are configured to prefer SR label imposition
  - In no particular order



Simplest migration LDP to SR

Assumptions:
- all the nodes can be upgraded to SR
- all the services can be upgraded to SR

- Initial state: All nodes run LDP, not SR
- Step1: All nodes are upgraded to SR
  - In no particular order
  - leave default LDP label imposition preference
- Step2: All PEs are configured to prefer SR label imposition
  - In no particular order
- Step3: LDP is removed from the nodes in the network
  - In no particular order
- Final state: All nodes run SR, not LDP



Enabling Segment Routing – XR and XE

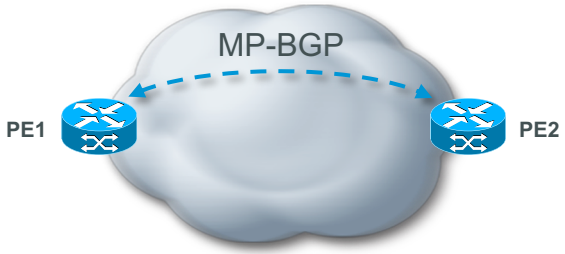
```
IOS-XR
segment-routing
!
router isis SR-AS-1
 address-family ipv4 unicast
 segment-routing mpls
 !
 interface Loopback0
 address-family ipv4 unicast
  prefix-sid absolute 16001
 !
commit
```

```
IOS-XE
XE-2(config)#segment-routing mpls
XE-2(config-srmpls)#connected-prefix-sid-map
XE-2(config-srmpls-conn)#address-family ipv4
XE-2(config-srmpls-conn-af)#2.2.2.2/32 absolute 16002 range 1
XE-2(config-srmpls-conn-af)#exit
XE-2(config-srmpls-conn)#exit
XE-2(config-srmpls)#exit
XE-2(config)#router isis SR-AS-1
XE-2(config-router)#segment-routing mpls
```

Understanding SR Control and Data Plane

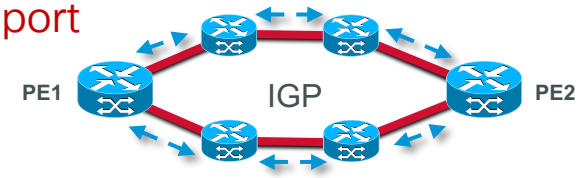
MPLS Control and Forwarding Operation with Segment Routing

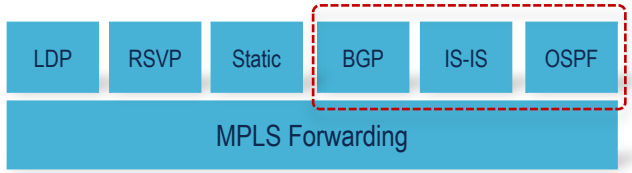
Services



No changes to control or forwarding plane

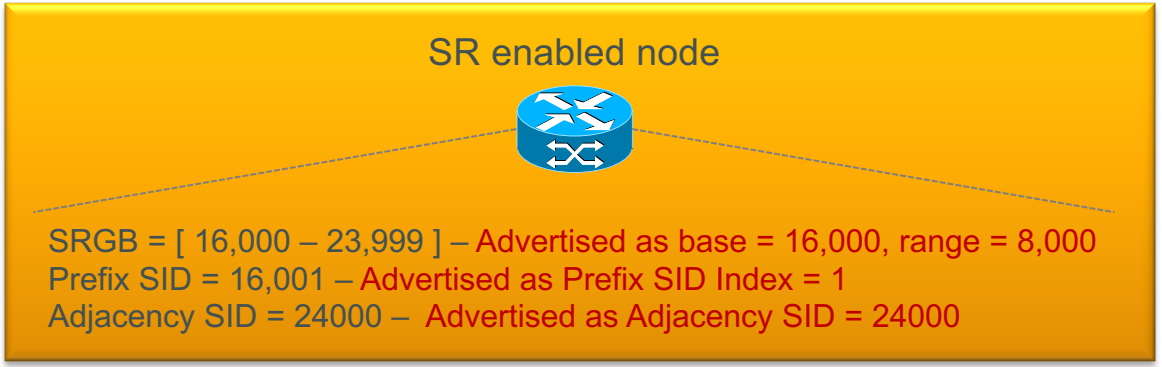
Packet Transport





IGP or BGP label distribution for IPv4 and IPv6. Forwarding plane remains the same

SID Encoding



- Prefix SID
  - **Label** form SR Global Block (SRGB)
  - **SRGB** advertised within IGP via TLV
  - In the configuration, Prefix-SID can be configured as an **absolute** value or an **index**
  - In the protocol advertisement, Prefix-SID is always encoded as **a globally unique index**

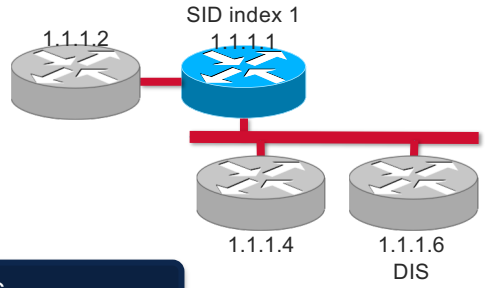
    Index represents an **offset** from SRGB base, zero-based numbering, i.e. 0 is 1st index

    E.g. index **1** → SID is 16,000 + **1** = 16,001
- Adjacency SID
  - **Locally significant**
  - **Automatically allocated** by the IGP for each adjacency
  - Always encoded as an absolute (i.e. not indexed) value

SR IS-IS Control Plane Summary

- IPv4 and IPv6 control plane
- Level 1, level 2 and multi-level routing
- Prefix Segment ID (Prefix-SID) for host prefixes on loopback interfaces
- Adjacency Segment IDs (Adj-SIDs) for adjacencies
- Prefix-to-SID mapping advertisements (mapping server)
- MPLS penultimate hop popping (PHP) and explicit-null signaling

IS-IS Configuration – Example




```
router isis 1
 address-family ipv4 unicast
  metric-style wide
  segment-routing mpls
 !
 address-family ipv6 unicast
  metric-style wide
  segment-routing mpls
 !
 interface Loopback0
  passive
  address-family ipv4 unicast
   prefix-sid absolute 16001
  !
  address-family ipv6 unicast
   prefix-sid absolute 20001
  !
 !
```

Wide metrics

enable SR IPv4 control plane and SR MPLS data plane on all ipv4 interfaces in this IS-IS instance

Wide metrics

enable SR IPv6 control plane and SR MPLS data plane on all ipv6 interfaces in this IS-IS instance

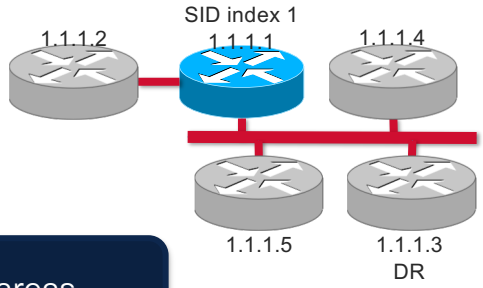
Ipv4 Prefix-SID value for loopback0

Ipv6 Prefix-SID value for loopback0

SR OSPF Control Plane Summary

- OSPFv2 control plane
- Multi-area
- IPv4 Prefix Segment ID (Prefix-SID) for host prefixes on loopback interfaces
- Adjacency Segment ID (Adj-SIDs) for adjacencies
- Prefix-to-SID mapping advertisements (mapping server)
- MPLS penultimate hop popping (PHP) and explicit-null signaling

OSPF Configuration Example



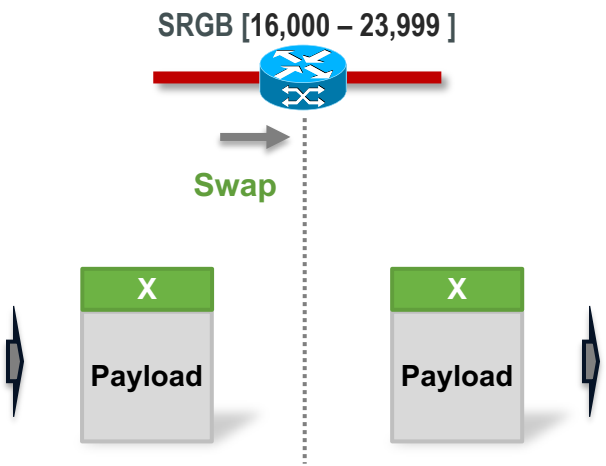

```
router ospf 1
 router-id 1.1.1.1
 segment-routing mpls
area 0
  interface Loopback0
   passive enable
   prefix-sid absolute 16001
  !
 !
!
```

Enable SR on all areas

Prefix-SID for loopback0

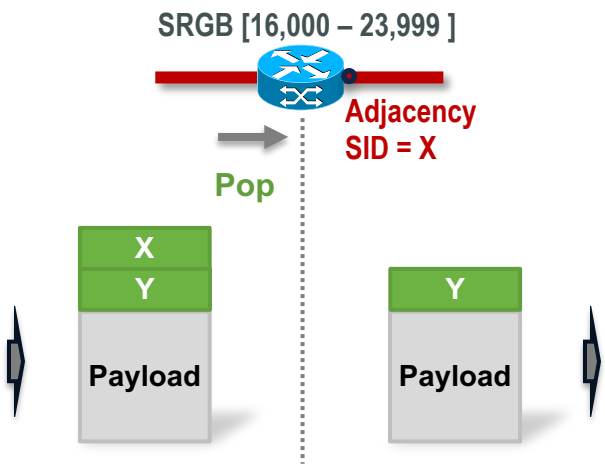
MPLS Data Plane Operation (labeled)

Prefix SID



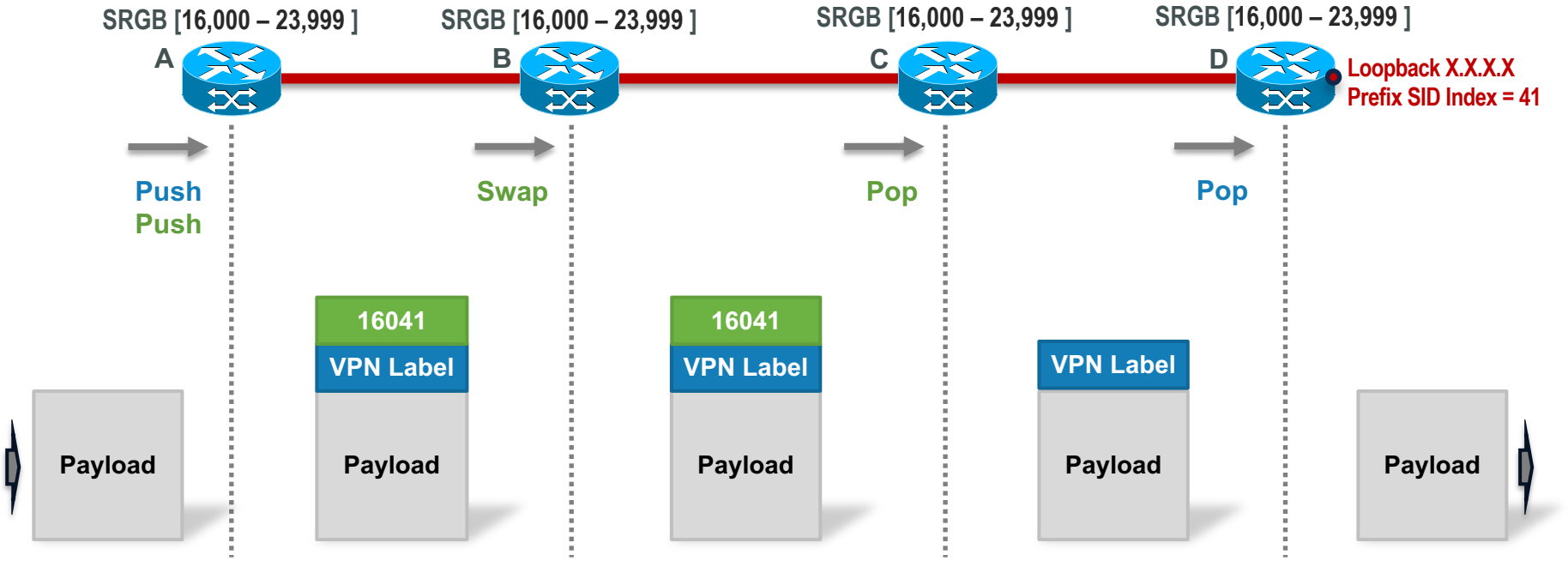
- Packet forwarded along IGP shortest path (ECMP)
- Swap operation performed on input label
- Same top label if same/similar SRGB
- PHP if signaled by egress LSR

Adjacency SID

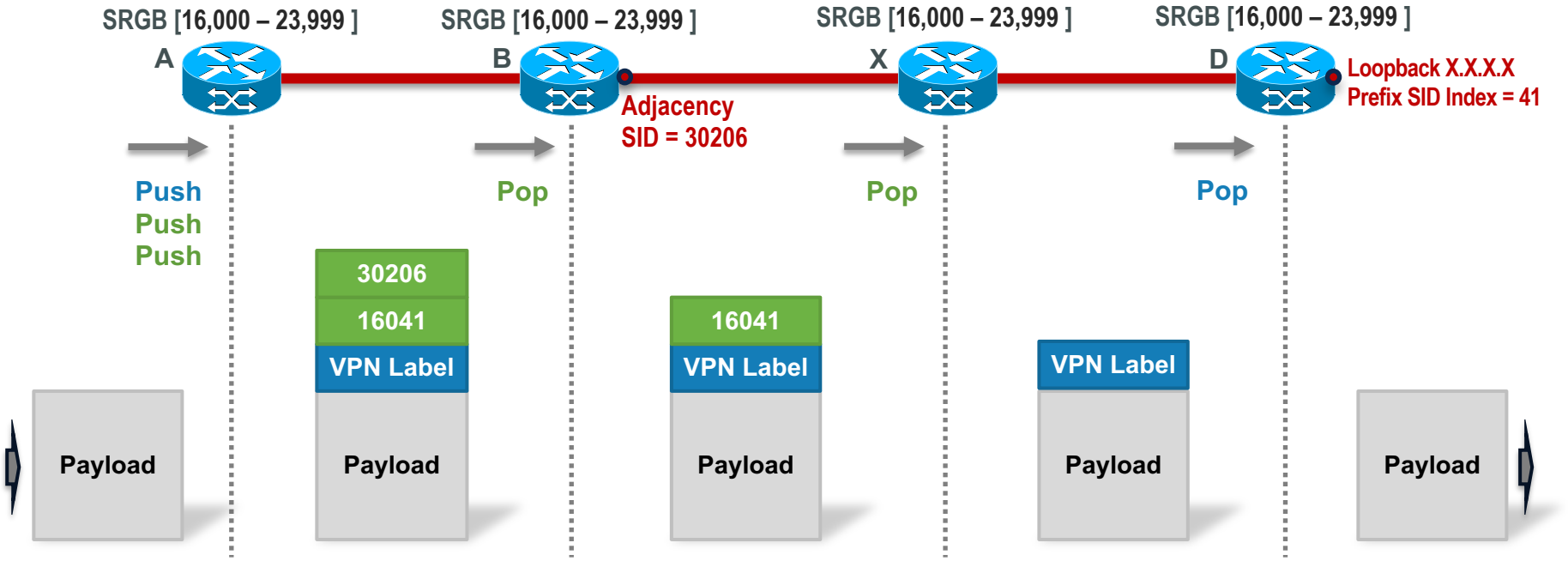


- Packet forwarded along IGP adjacency
- Pop operation performed on input label
- Top labels will likely differ
- Penultimate hop always pops last adjacency SID

MPLS Data Plane Operation (Prefix SID)



MPLS Data Plane Operation (Adjacency SIDs)



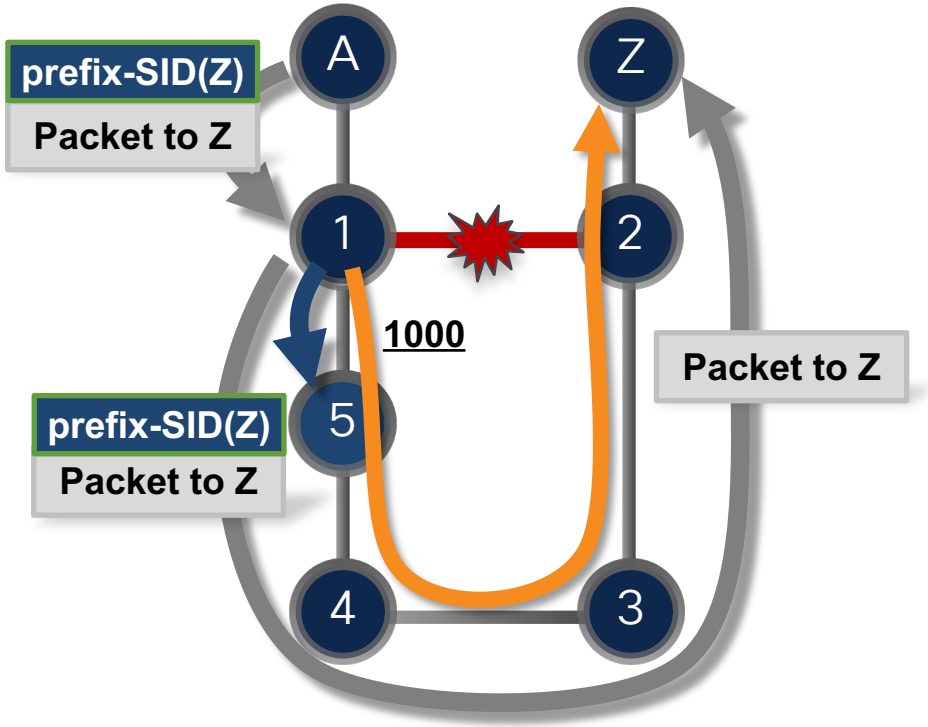
Traffic Protection – TI-LFA

Topology Independent LFA (TI-LFA) – Benefits

- 100%-coverage 50-msec link, node, and SRLG protection
- Simple to operate and understand
  - automatically computed by the IGP
- Prevents transient congestion and suboptimal routing
  - leverages the post-convergence path, planned to carry the traffic

TI-LFA – Zero-Segment Example

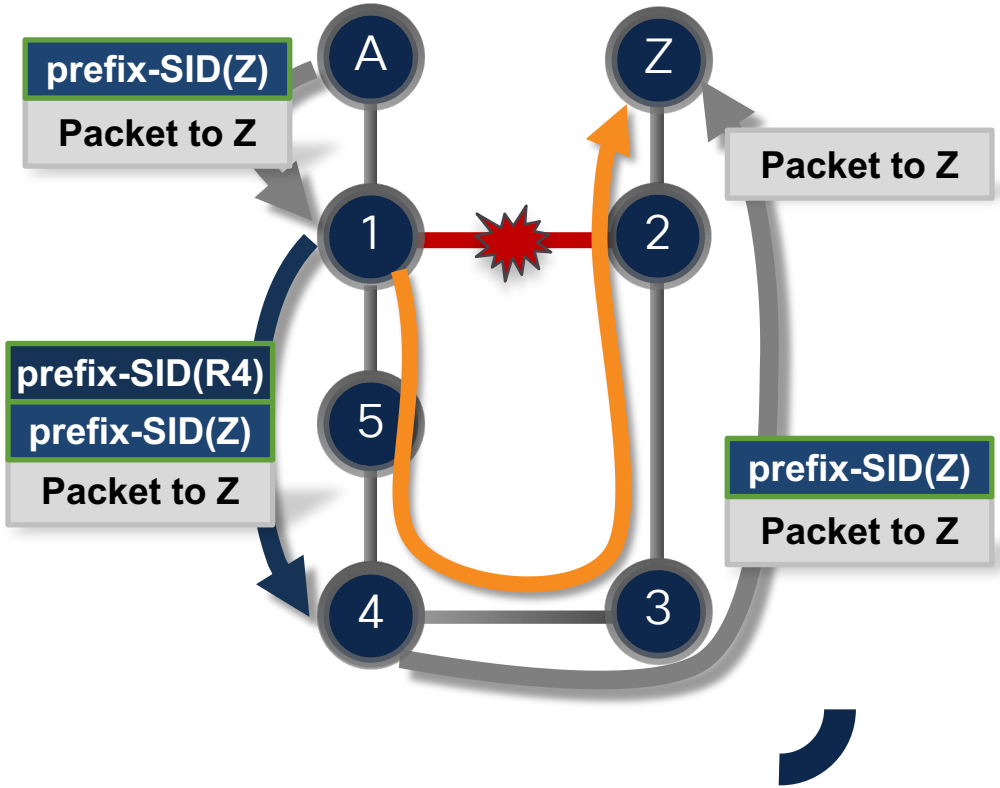
- TI-LFA for link R1R2 on R1
- Calculate post-convergence SPT
  - SPT with link R1R2 removed from topology
- Derive SID-list to steer traffic on post-convergence path
- R1 will steer the traffic towards LFA R5



Default metric: 10

TI-LFA – Single-Segment Example

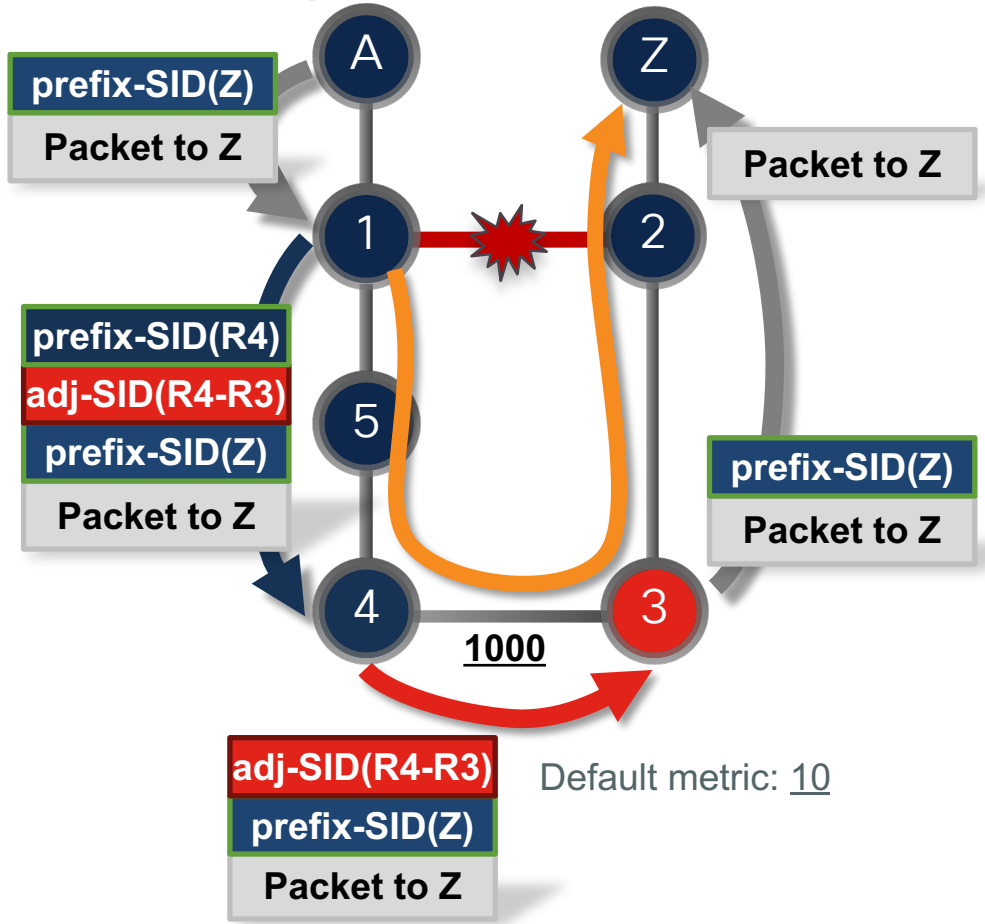
- TI-LFA for link R1R2 on R1
- Calculate post-convergence SPT
- Derive SID-list to steer traffic on post-convergence path → <Prefix-SID(R4)>
  - Also known as “PQ-node”
- R1 will push the prefix-SID of R4 on the backup path



Default metric:10

TI-LFA – Double-Segment Example

- TI-LFA for link R1R2 on R1
- Calculate post-convergence SPT
- Derive SID-list to steer traffic on post-convergence path → <Prefix-SID(R4), Adj-SID(R4-R3)
  - Also known as “P- and Q-node”
- R1 will push the prefix-SID of R4 and the adj-SID of R4-R3 link on the backup path



Enabling TI-LFA

```
IOS-XR
router isis SR-AS-1
 interface GigabitEthernet0/0/0/0
 address-family ipv4 unicast
  fast-reroute per-prefix ti-lfa level 2
 !
 !
 interface GigabitEthernet0/0/0/1
 address-family ipv4 unicast
  fast-reroute per-prefix ti-lfa level 2
 !
 !
 interface GigabitEthernet0/0/0/3
 address-family ipv4 unicast
  fast-reroute per-prefix ti-lfa level 2
!
```

```
IOS-XE
router isis SR-AS-1
 fast-reroute ti-lfa level-2
```

TI-LFA Backup Coverage

IOS-XR

```
RP/0/0/CPU0:XR-1#show isis fast-reroute summary
IS-IS SR-AS-1 IPv4 Unicast FRR summary
                        Critical High     Medium   Low       Total
                        Priority Priority Priority Priority
Prefixes reachable in L2
 All paths protected    0        0        4        8         12
 Some paths protected    0         0         0         0         0
 Unprotected            0        0        0        0         0
 Protection coverage    0.00%    0.00%    100.00%  100.00%   100.00%
```

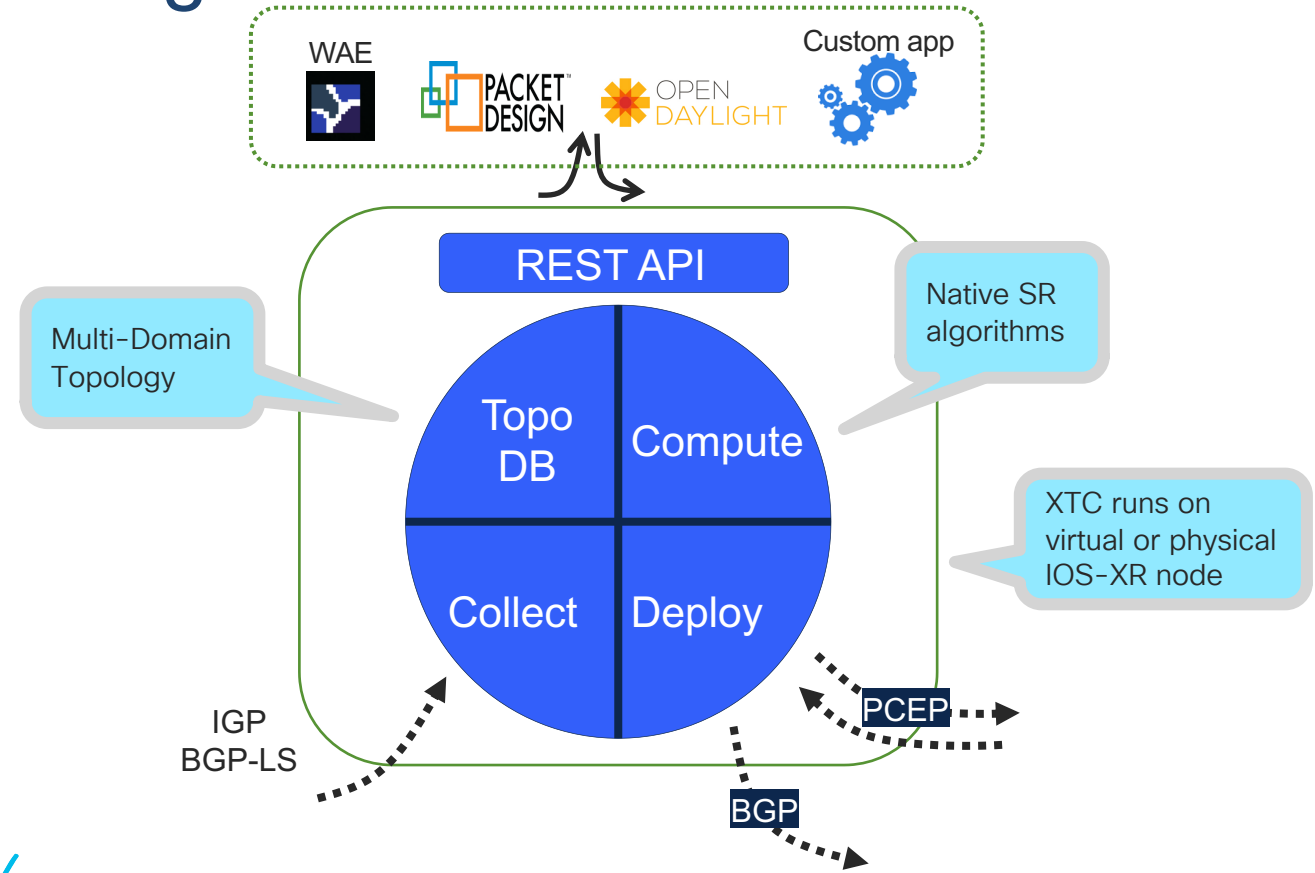
IOS-XE

```
XE-2#show isis fast-reroute summary
Tag SR-AS-1:
Microloop Avoidance State: Enabled for protected
Segment-Routing Microloop Avoidance State: Disabled
IPv4 Fast-Reroute Protection Summary:


 Prefix Counts:          Total      Protected   Coverage
  High priority:        0          0           0%
  Normal priority:      12        12          100%
  Total:                12        12          100%
```

ODN

XTC Building Blocks

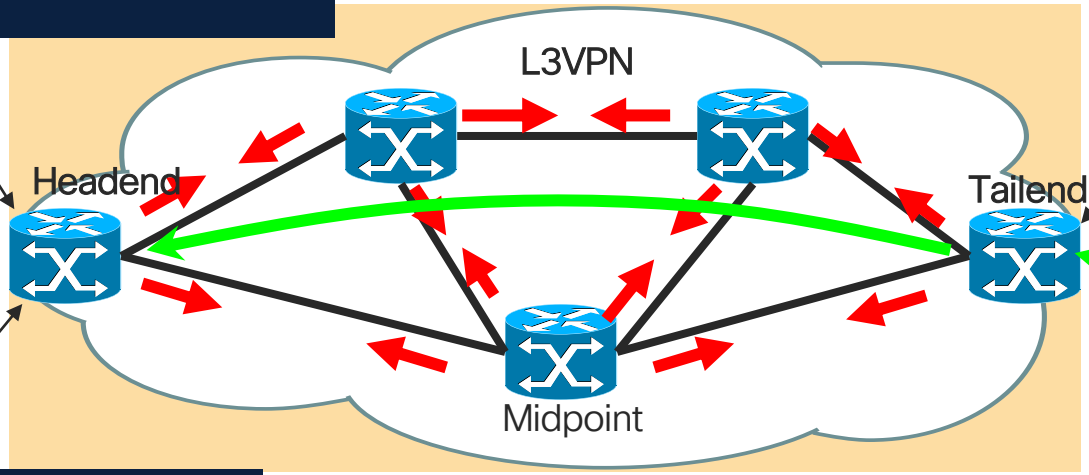


Dynamic BGP Traffic Engineering (BGP-TE)

- Headend must have **global auto tunnel** configuration
- Headend must have an **attribute-list** for TE specific configurations

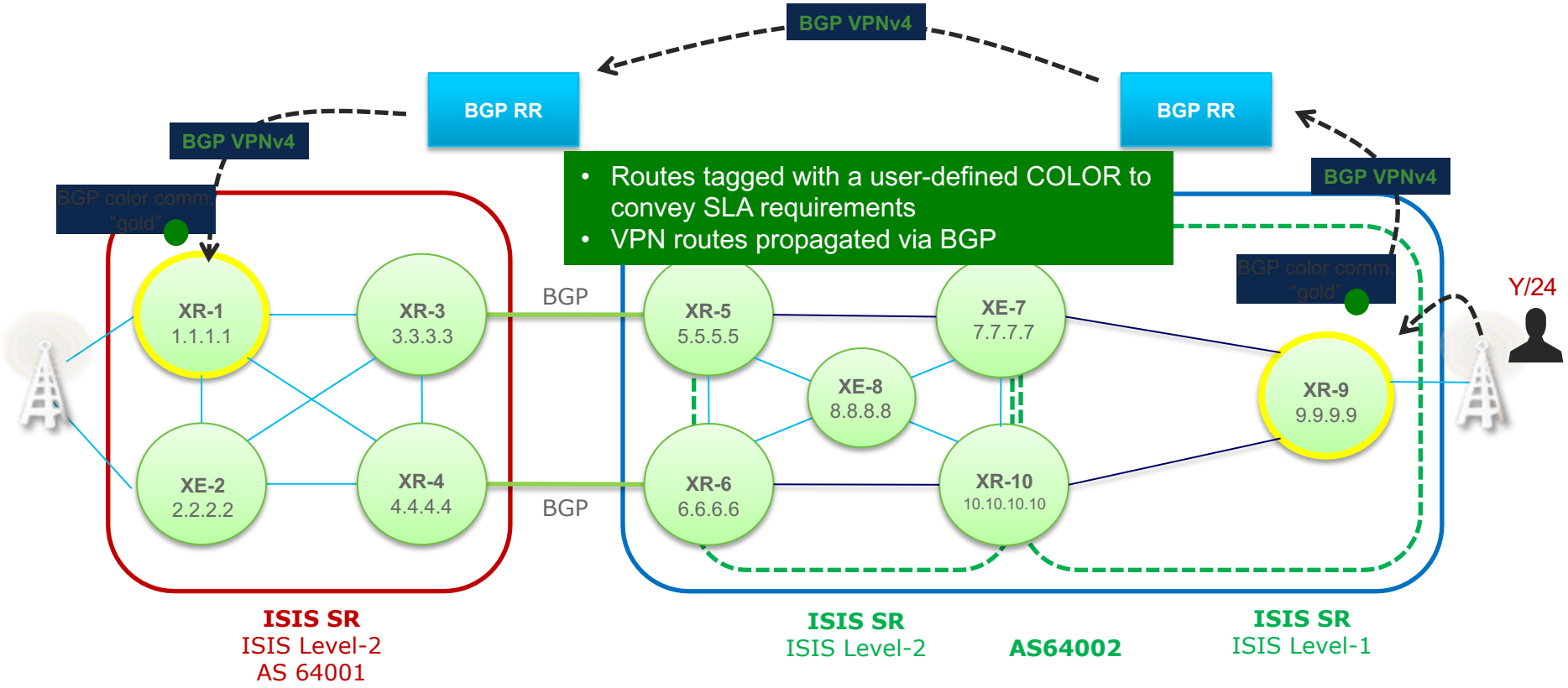
Step 1:
Setting Community Attribute:
Route-map matches customer prefix and sets unique community

Step 2:
Matching community Attribute: Route-map matches community attribute and sets attribute-list for the NLRI



Step 3:
Attribute-list Configuration- Has the tunnel related configurations

ODN Workflow

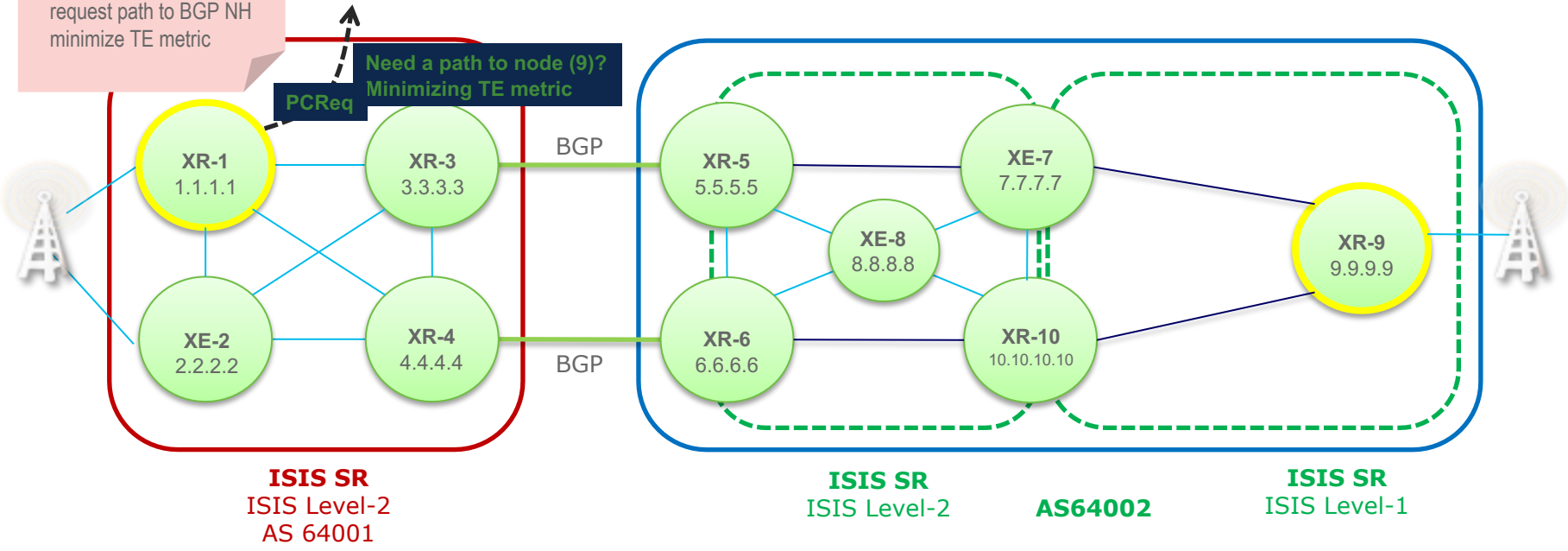


ODN Workflow

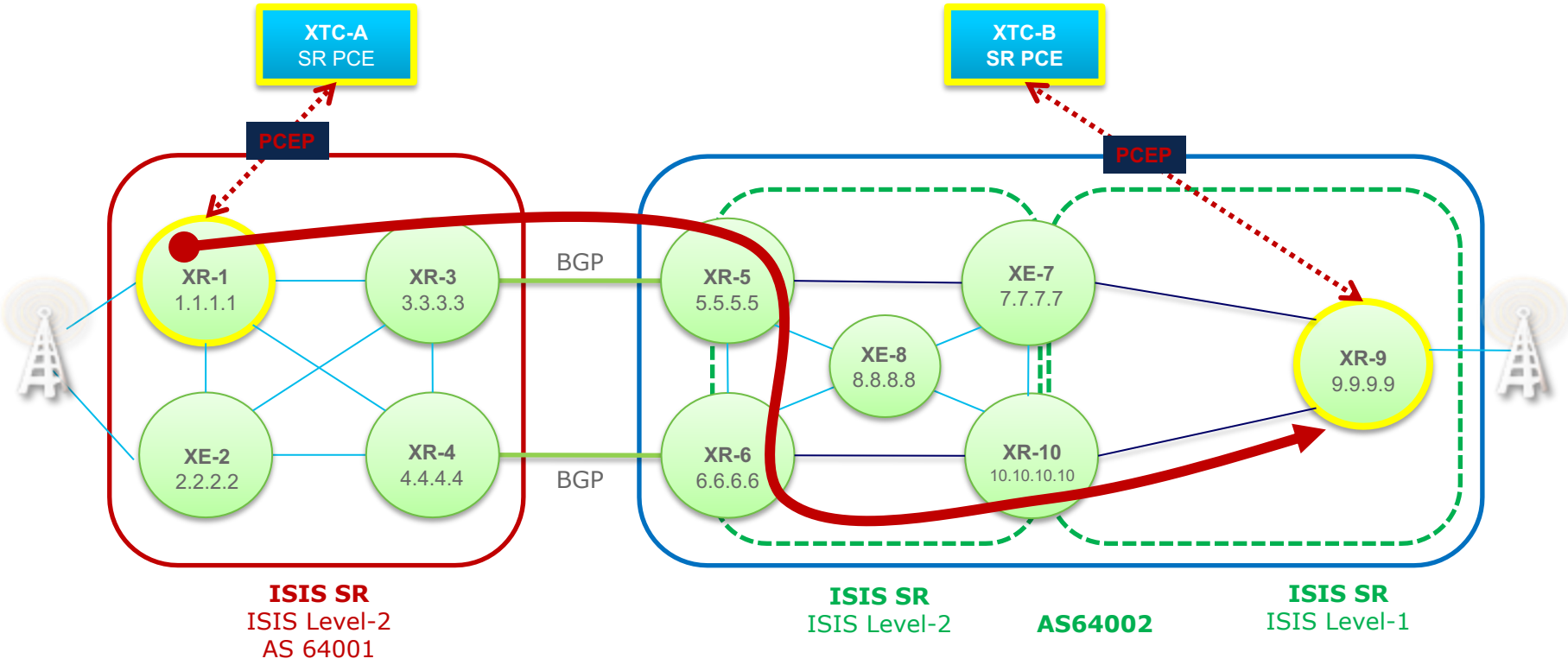
- Ingress PE matches on user-specified BGP “color” community
- Ingress PE enforces a “template” associated with the color community

SRTE
On-demand color “gold”
contact PCE
request path to BGP NH
minimize TE metric

XTC-A
SR PCE



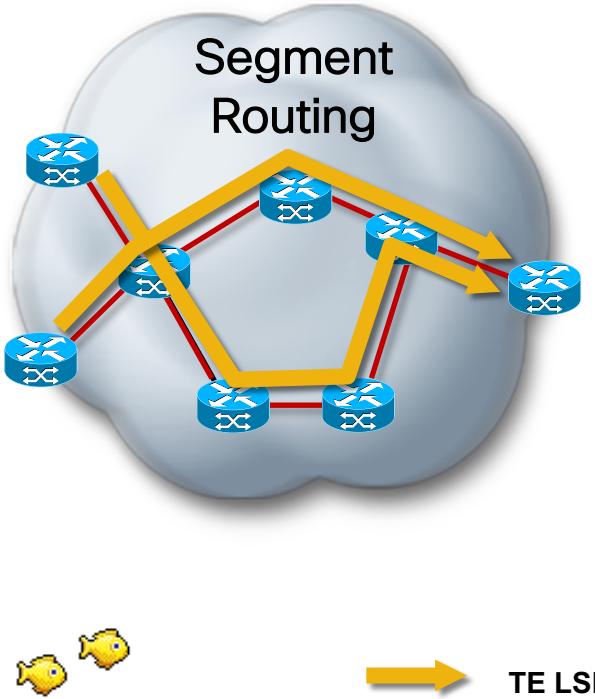
PCEP



SRTE

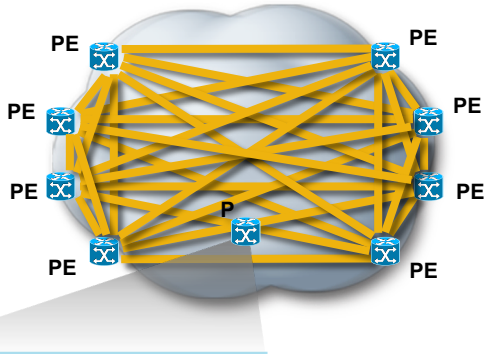
Traffic Engineering with Segment Routing

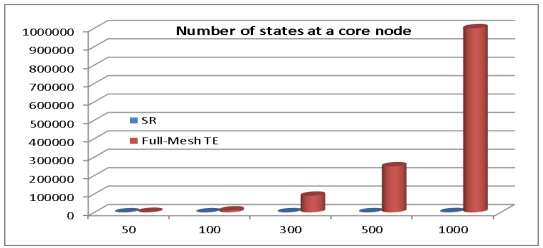
- Source-Based routing – State only at ingress PE
- Supports constraint-based routing
- Supports centralized admission control
- Uses existing ISIS / OSPF extensions to advertise link attributes
- No RSVP-TE to establish LSPs
- Supports ECMP



MPLS LFIB with Segment Routing

- LFIB populated by IGP (ISIS / OSPF)
- Forwarding table remains constant (Nodes + Adjacencies) regardless of number of paths

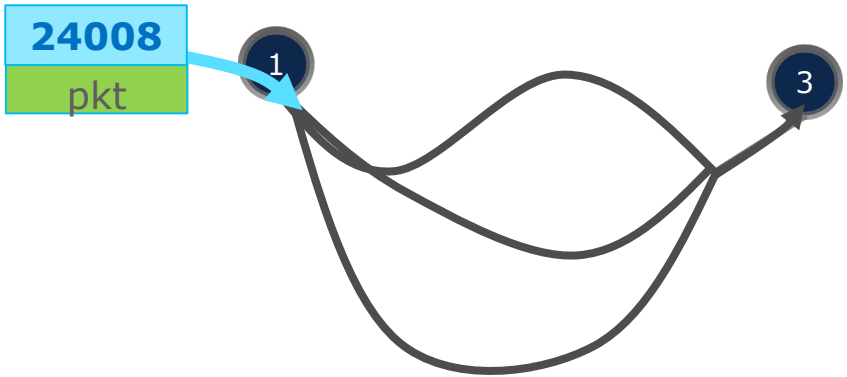




	In Label	Out Label	Out Interface
Network Node Segment Ids	L1	L1	Intf1
	L2	L2	Intf1
	…	…	…
	L8	L8	Intf4
	L9	L9	Intf2
Node Adjacency Segment Ids	L10	Pop	Intf2
	…	…	…
	Ln	Pop	Intf5

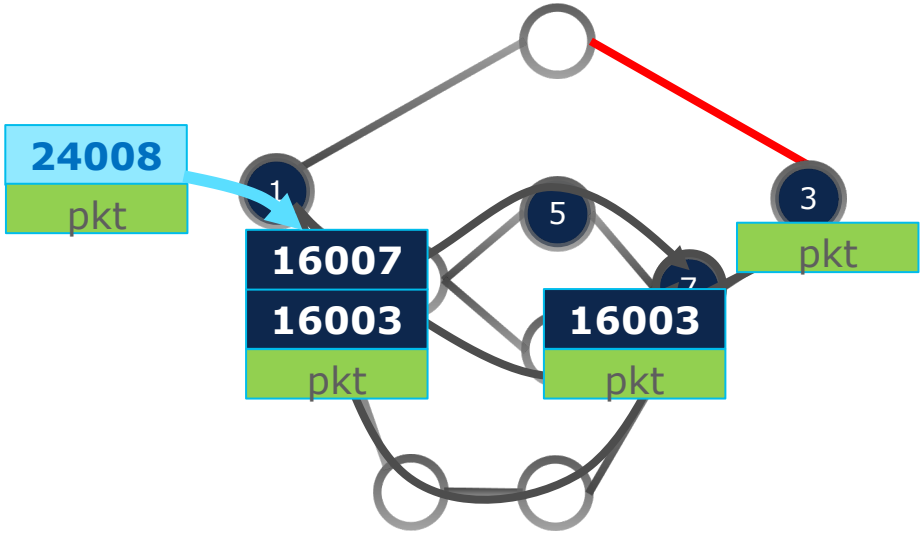
Forwarding table remains constant

Binding SID



- **Binding Segment is a fundamental building block of SRTE**
- The Binding Segment is a local segment
  - Has local significance
- A Binding-Segment ID identifies a SRTE Policy
  - Each SRTE Policy is associated 1-for-1 with a Binding-SID
- Packet received with Binding-SID as top label is steered into the SRTE Policy associated with the Binding-SID
  - Binding-SID label is popped, SRTE Policy's SID list is pushed
- Binding SID can be automatically assigned or statically configured as part of the SRTE policy

Binding SID



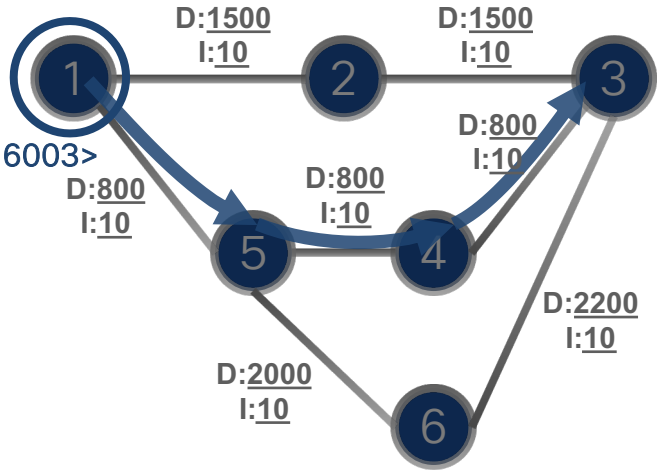
- **Binding Segment is a fundamental building block of SRTE**
- The Binding Segment is a local segment
  - Has local significance
- A Binding-Segment ID identifies a SRTE Policy
  - Each SRTE Policy is associated 1-for-1 with a Binding-SID
- Packets received with Binding-SID as top label are steered into the SRTE Policy associated with the Binding-SID
  - Binding-SID label is popped, SRTE Policy's SID list is pushed
- Binding SID can be automatically assigned or statically configured as part of the SRTE policy

SRTE Use Cases

Low- Delay path

Use-Case

SID-list: <16005, 16004, 16003>



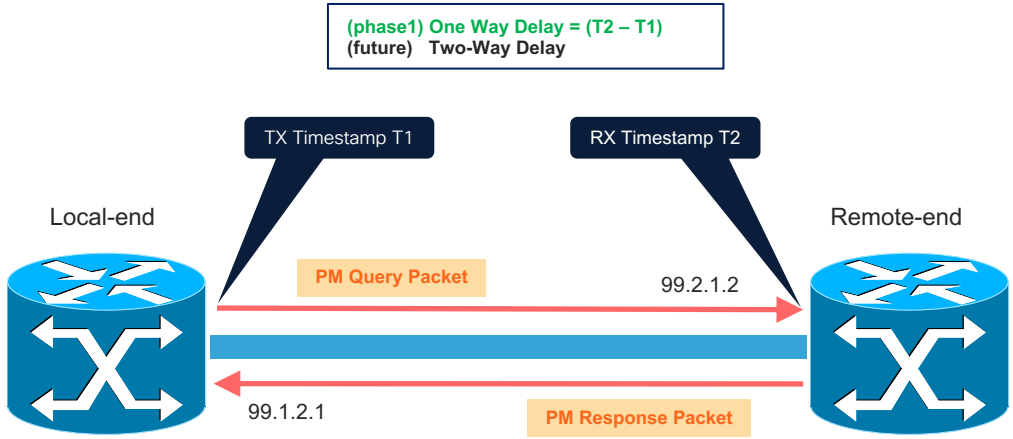
Node1

```
segment-routing
  traffic-eng
    policy POLICY1
      color 20 end-point ipv4 1.1.1.3
      candidate-paths
        preference 100
          dynamic mpls
            metric
              type delay
```

- Head-end computes a SID-list that expresses the shortest-path according to the selected metric **delay**

Link Delay Measurement Protocol

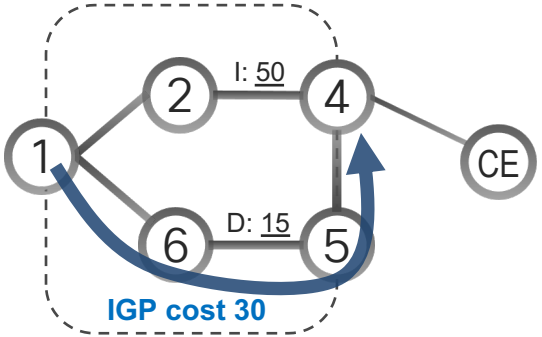
- PTP
  - Accurate time-stamp
- MPLS PM
  - using GAL/Gach defined in RFC 6374
- IGP and BGP support:
  - Extended TE Link Delay Metrics will be supported in ISIS (RFC 7810) and OSPF (RFC 7471)
  - BGP-LS (draft-ietf-idr-te-pm-bgp) Extended TE Link Delay Metrics
- No additional configuration in ISIS/OSPF/BGP-IS : Latency automatically flooded



XTC (PCE) view

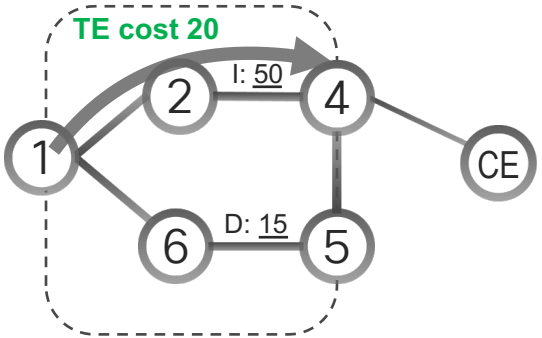
```
Link[0]: local address 99.1.2.1, remote address 99.2.1.2
    Local node:
      ISIS system ID: 0000.0000.6666 level-2 ASN: 64002
    Remote node:
      TE router ID: 5.5.5.5
      Host name: Napoli-5
      ISIS system ID: 0000.0000.5555 level-2 ASN: 64002
    Metric: IGP 1, TE 1,Delay 6000
    Bandwidth: Total 125000000, Reservable 0
    Adj SID: 24005 (protected) 24004 (unprotected)
    Excluded from CSPF: no
    Reverse link exists: yes
```

Different VPNs need different underlay SLA



Default IGP cost: I:10
Default Delay cost: D:10

Basic VPN should use lowest cost underlay path



Default IGP cost: I:10
Default Delay cost: D:10

Premium VPN should use lowest delay path

Objective: operationalize this service for simplicity, scale and performance

On-Demand SR Policy work-flow
Automatic LSP setup and steering

❺

```
router bgp 1
 neighbor 1.1.1.10
  address-family vpnv4 unicast
!
segment-routing
 traffic-eng
  on-demand color 20
   preference 100
    metric
     type delay
```

SR Policy template Low-Delay (color 20)

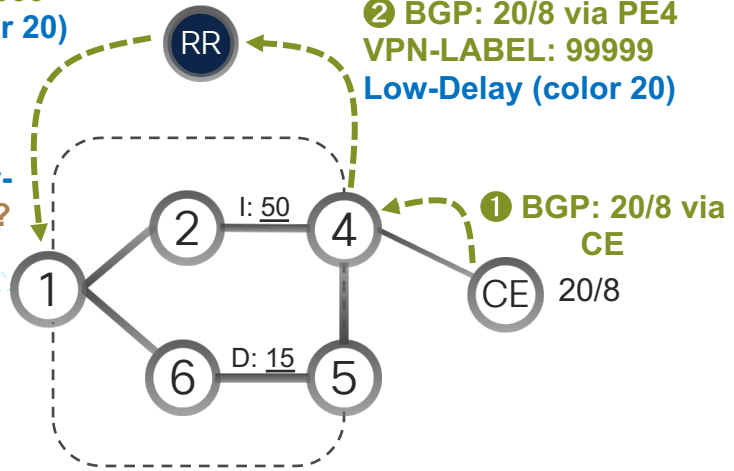
❸ BGP: 20/8 via PE4
VPN-LABEL: 99999
Low-Delay (color 20)

❷ BGP: 20/8 via PE4
VPN-LABEL: 99999
Low-Delay (color 20)

❹ PE4 with Low-Delay (color 20)?

❺ use template color 20

❻ → SID-list <16002, 30204>



Default IGP cost: I:10
Default Delay cost: D:10

Automated performant steering

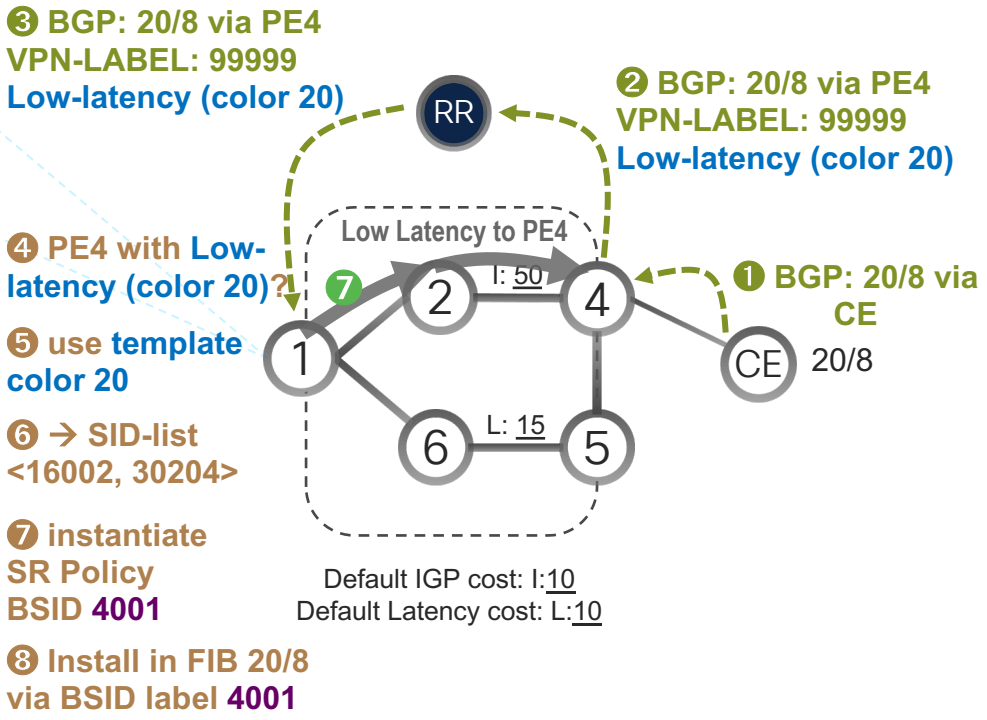
❼❽

FIB table at PE1
BGP: 20/8 via **4001**
SRTE: **4001**: Push <16002, 30204>

Automatically, the service route resolves on the Binding SID (4001) of the SR Policy it requires

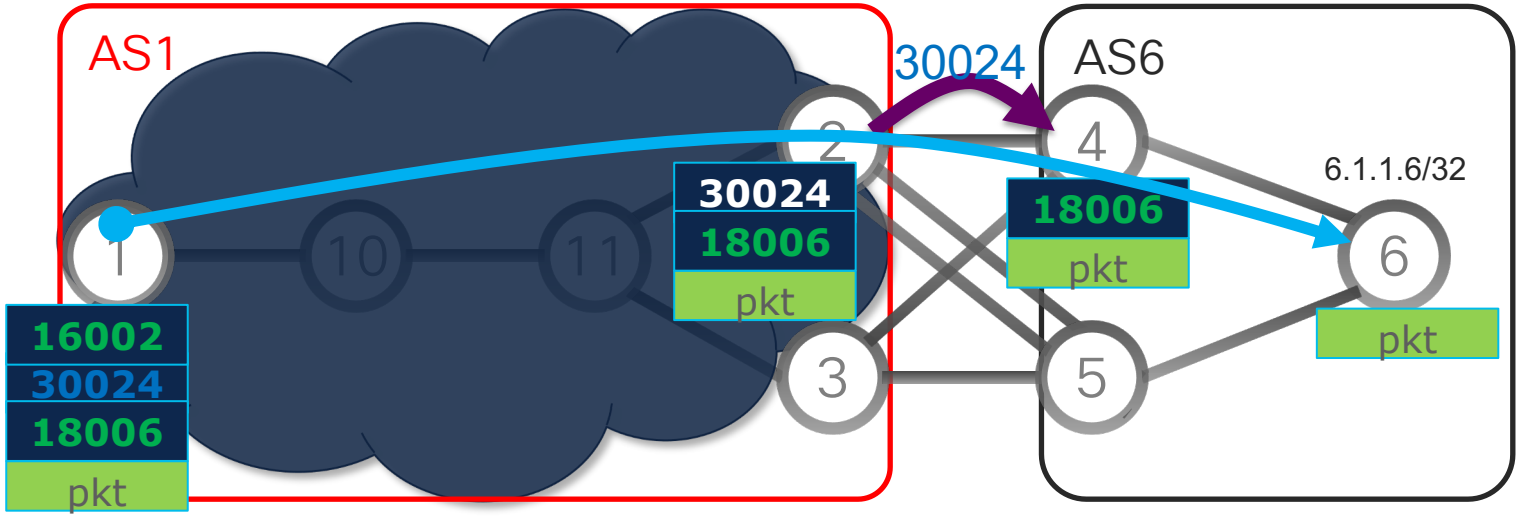
Simplicity and Performance

No route-policy required. No complex PBR to configure, no PBR performance tax



SR-TE Use Cases
Inter domain connectivity with SLA

Crossing the AS border: BGP Peering Segment



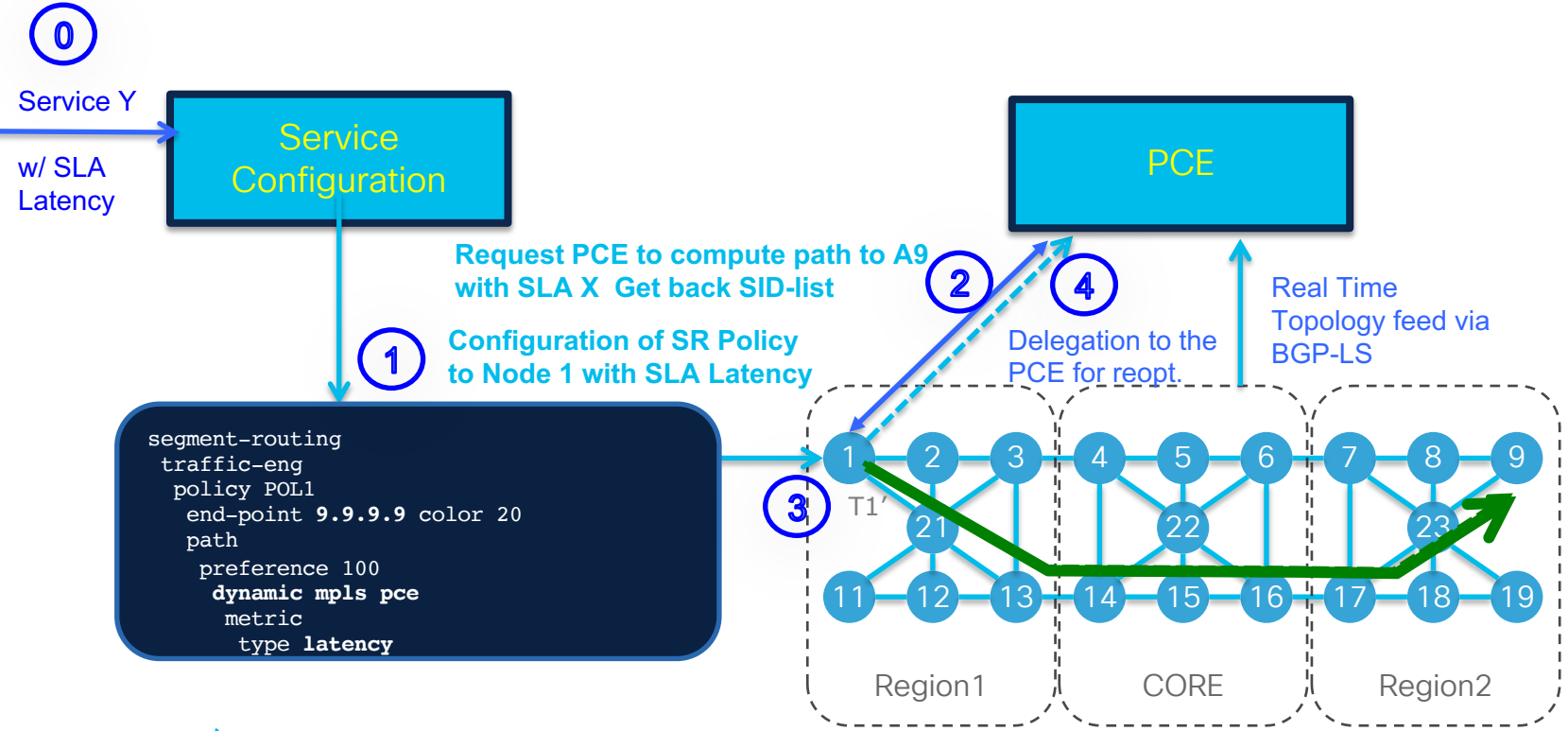
- "Pop
- Local
  - Dyn

What is missing?

How do I get topology info of external domains?

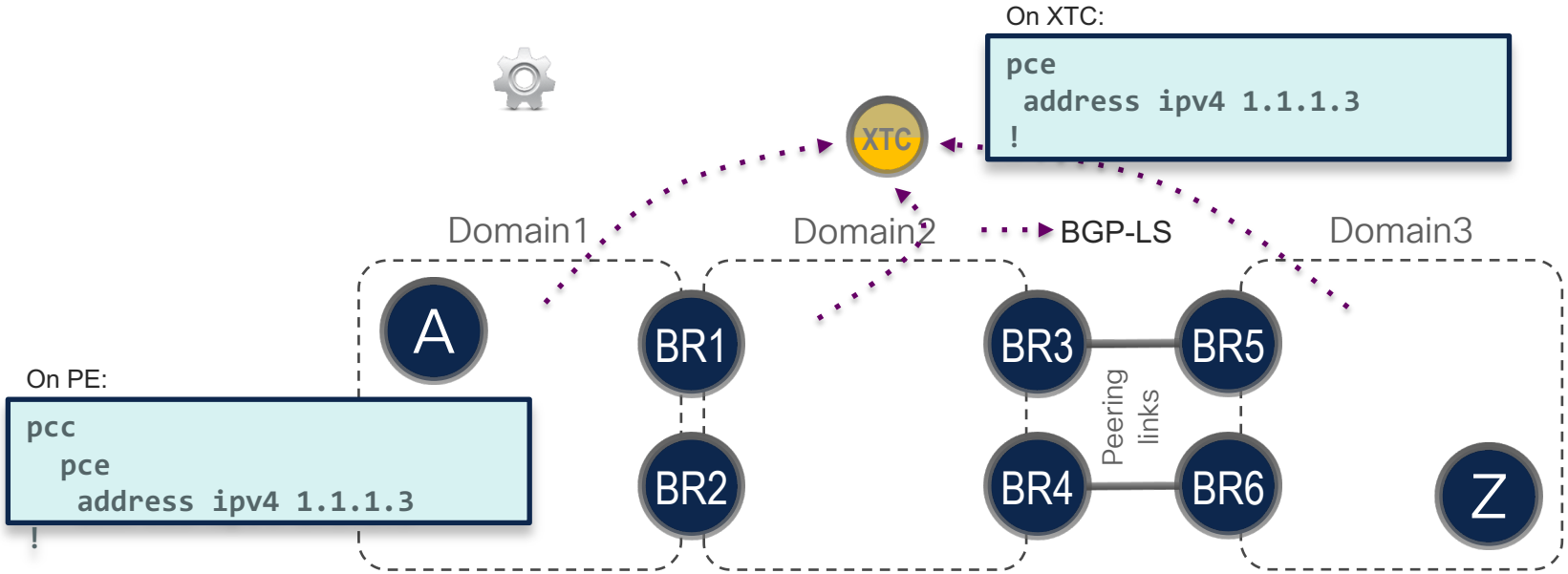
Inter Area Path Computation with SLA

Ask: Provide latency optimized path across multiple AS's from a source to a destination



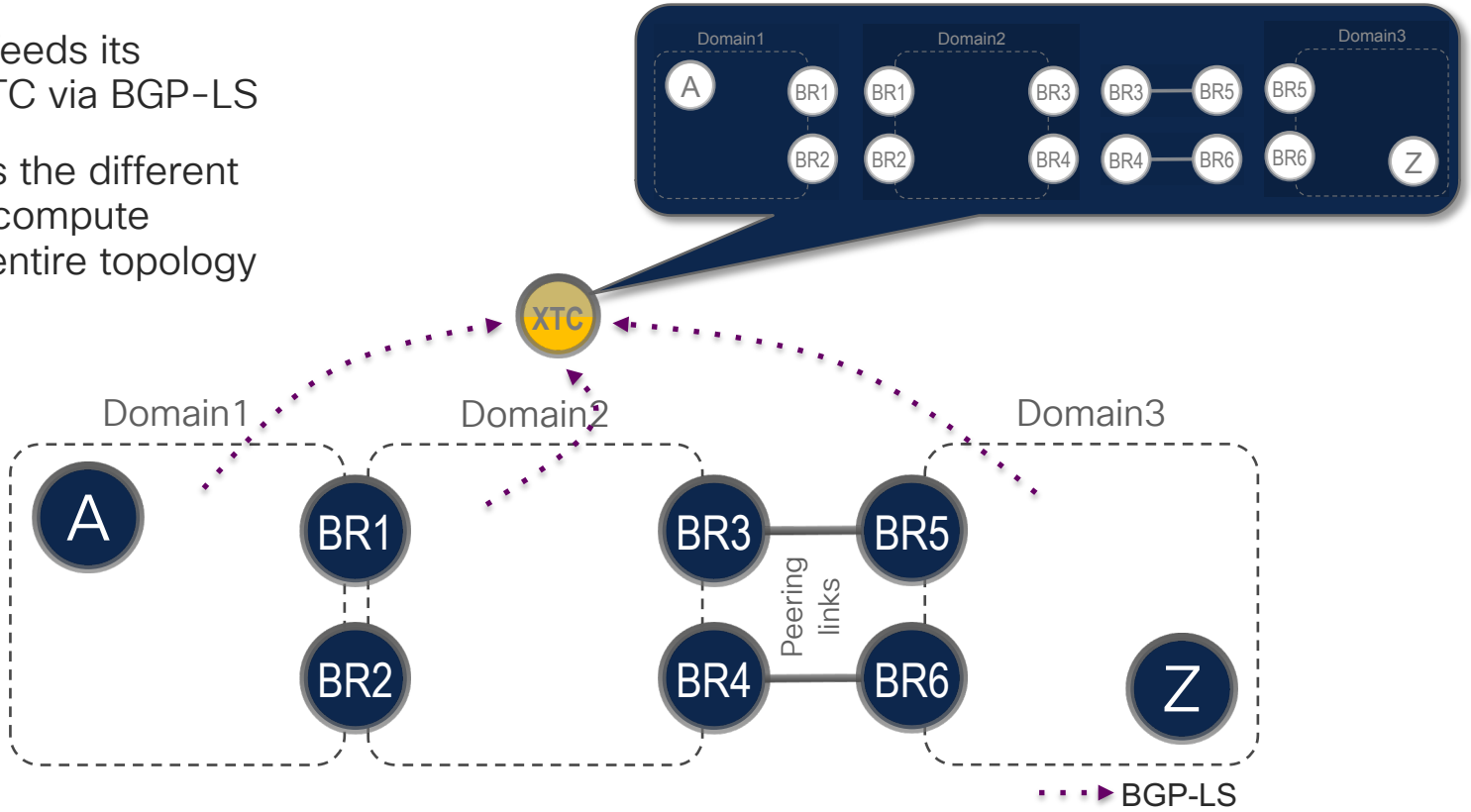
XR Transport Controller

- XTC is an IOS XR multi-domain stateful SR Path Computation Element (PCE)
  - Fundamentally Distributed (RR-like Deployment)
  - Supports RSVP-TE



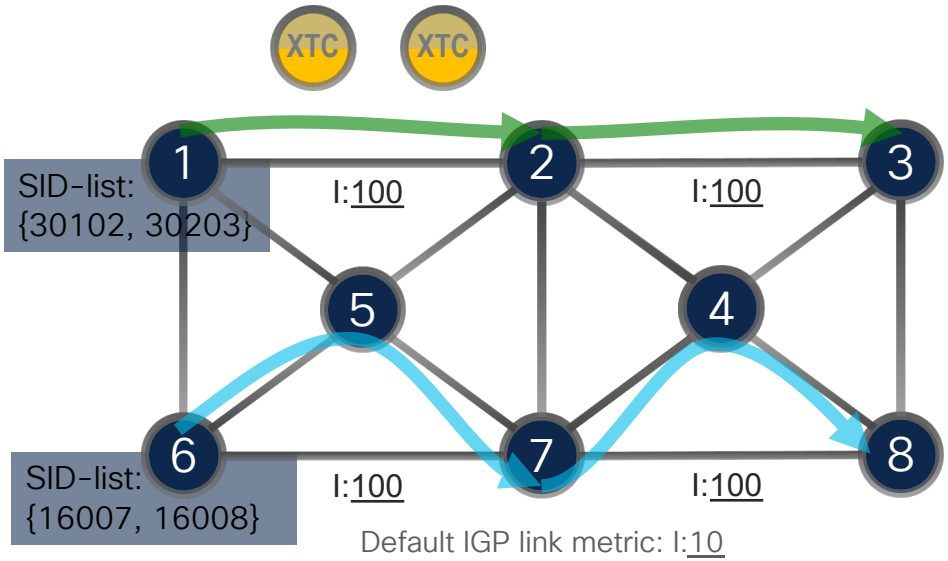
XTC Receives & Consolidates Multiple Topologies

- Each domain feeds its topology to XTC via BGP-LS
- XTC combines the different topologies to compute paths across entire topology



Service Disjointness

Intra and inter domain



Node1

```
segment-routing
 traffic-eng
  policy POLICY1
   color 20 end-point ipv4 1.1.1.3
    candidate-paths
     preference 100
      dynamic mpls pce
       metric
        type igp
       association group 1 type node
```

Node6

```
segment-routing
 traffic-eng
  policy POLICY2
   color 20 end-point ipv4 1.1.1.8
    candidate-paths
     preference 100
      dynamic mpls pce
       metric
        type igp
       association group 1 type node
```

- Two dynamic paths between two different pairs of (head-end, end-point) must be disjoint from each other

Path Computation

Distributed or Centralized ?

Policy	Single-Domain	Multi-Domain
Reachability	IGP's	Centralized
Low Latency	Distributed or Centralized	Centralized
Disjoint from same node	Distributed or Centralized	Centralized
Disjoint from different node	Centralized	Centralized
Avoiding resources	Distributed or Centralized	Centralized
Capacity optimization	Centralized	Centralized
Multi Layer	Centralized	Centralized

SR-TE

- Simple, Automated and Scalable
  - No state in the network: state in the packet header
  - No tunnel interface: “SR Policy”
  - No head-end a-priori configuration: on-demand policy instantiation
  - No head-end a-priori steering: automated steering
- Multi-Domain
  - XR Traffic Controller (XTC) for compute
- Lots of Functionality and flexibility
  - Designed with lead operators along their use-cases

Thank you

Possibilities

#CiscoLive