Segment Routing in NGN

Sai Nyan Lynn Swe
CCIE # 38501 (R&S, SP and DC)
OPTIMITY Co Ltd.

Agenda

- Introduction
- What is Segment Routing
- Segment Routing Principles
- Basic Mechanics
- Segment Routing Global Block
- Consultants' Profiles

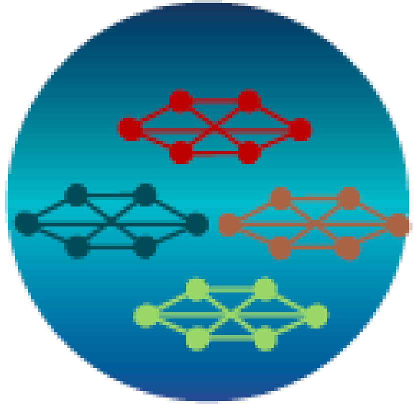
Introduction to Segment Routing

Current State of SP Network Deployments

Decades of Technical Evolution and Deployment

Vast Array of Technologies in Core, Edge, Access and Data Centers

Huge CAPEX Investment. Cannot be simply uprooted



Complex, multigenerational Networks

The Next Big Thing

1990s

IP

- Internet
- Flexibility
- Connectivity
- Content
- ...

2000s

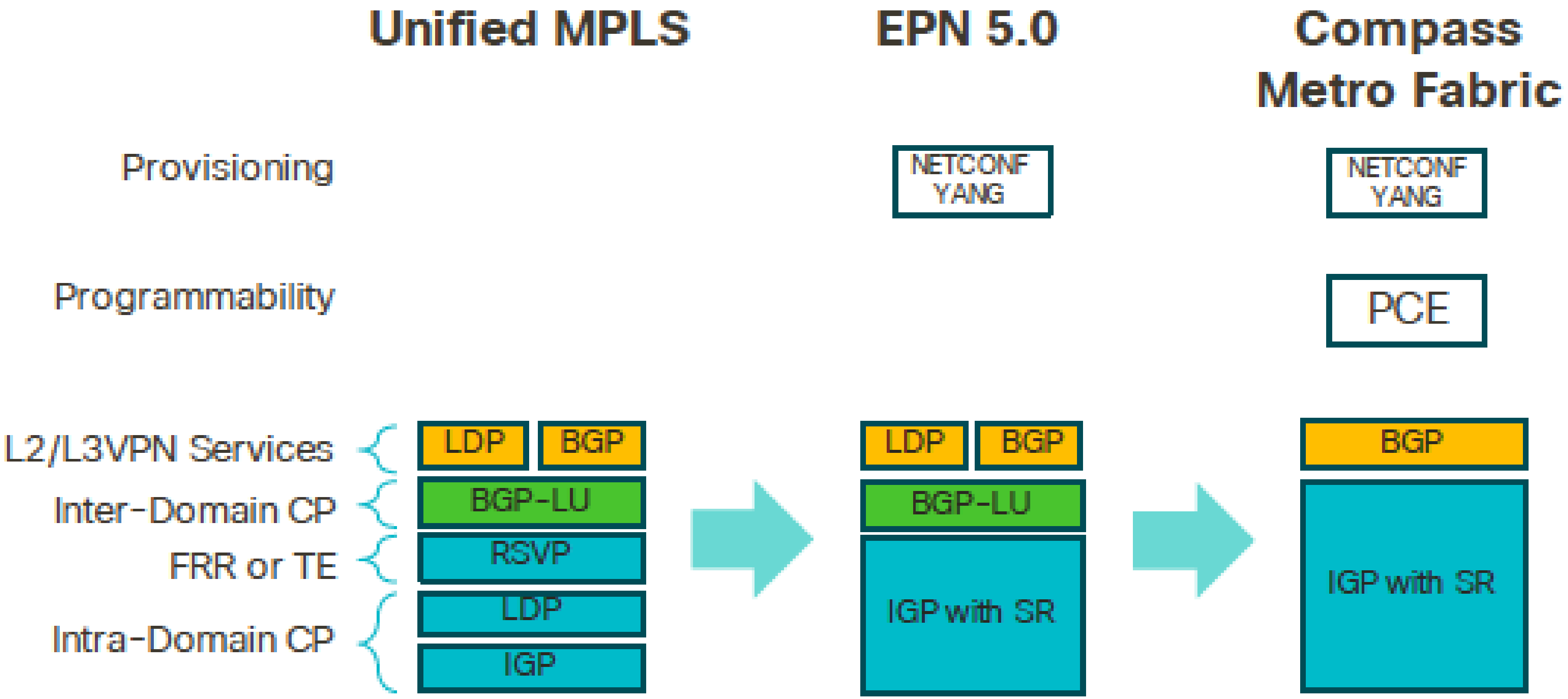
MPLS

- VPNs
- TE
- FRR
- SLA
- Large Scale
- ...

2010s

Simplification
Automation

Service Provider Network - Simplification Journey



Do more with less !!

https://xrdocs.github.io/design/

Service Provider Requirements

- Simplicity
- Scalability
- Automation
- End to end service.
  - End to end policy encoded by the application
- Fast Convergence
  - FRR convergence times
- Cloud integration and Virtualisation

Customers Ask For A Cloud-based solution

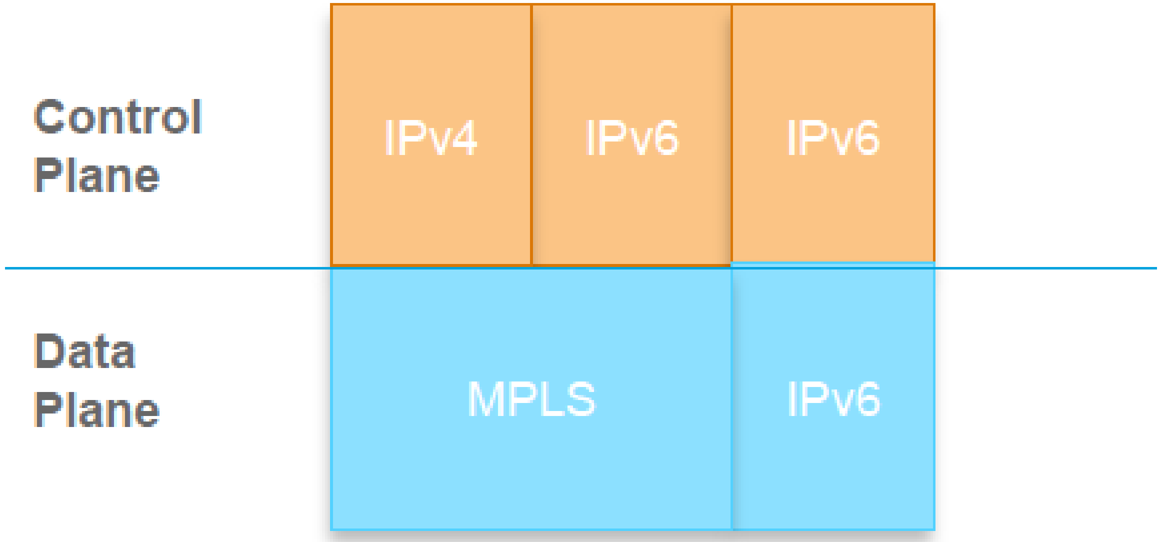
- Cloud-integrated
- Programmable network
- Applications must be able to define network paths
- The network must respond to application interaction
  - Rapidly-changing application requirements
  - Virtualisation
  - Guaranteed SLA and Network Efficiency
  - Fast recovery

Operators Ask For LDP/RSVP Improvement

- Simplicity
  - less protocols to operate
  - less protocol interactions to troubleshoot
  - less initial provisioning
  - avoid directed LDP sessions between core routers
  - deliver automated FRR for any topology
- Scale
  - avoid millions of labels in LDP database
  - avoid millions of TE LSP's in the network
  - avoid millions of tunnels to configure
  - avoid millions of bypass tunnels pre-signalled in the network
- End to end service

Segment Routing As A Solution

- **Simple**
- **Scalable**
- **Easy to provision**
- **MPLS**: an ordered list of segments is represented as a stack of labels
  - SR re-uses MPLS data plane without any change
- **IPv6**: an ordered list of segments is represented as a routing extension header



The need for SR...

Applications & Network Interaction - Implications for the Network Fabric

Many applications with dynamic and changing traffic patterns

IP Networks → **IP Networks & Traffic Engineering**

Shortest path with QoS

Traffic-engineered tunneling

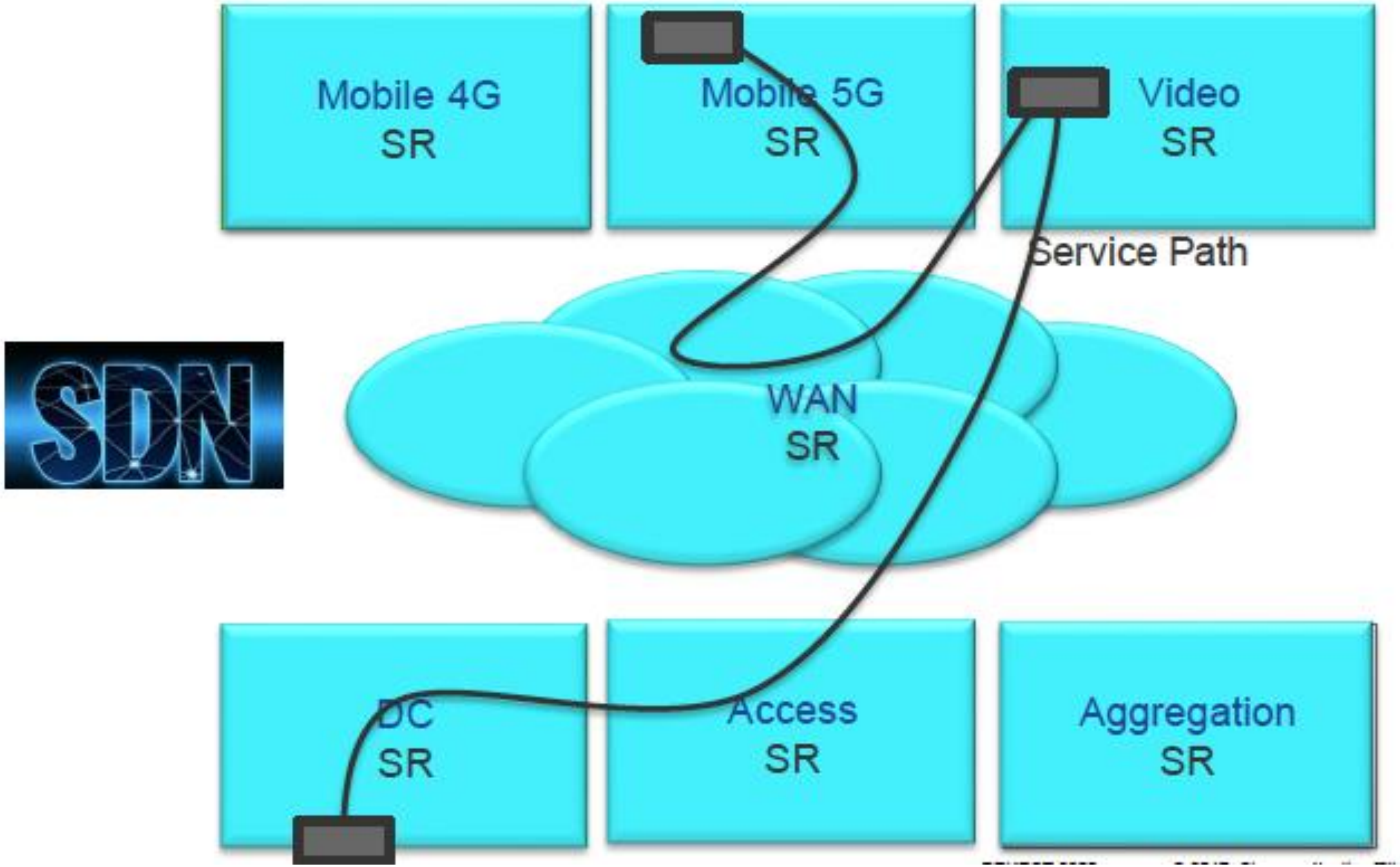
Limitations

- Limited to a single network domain
- Configuration & troubleshooting complexity
- States to be maintained in each network node

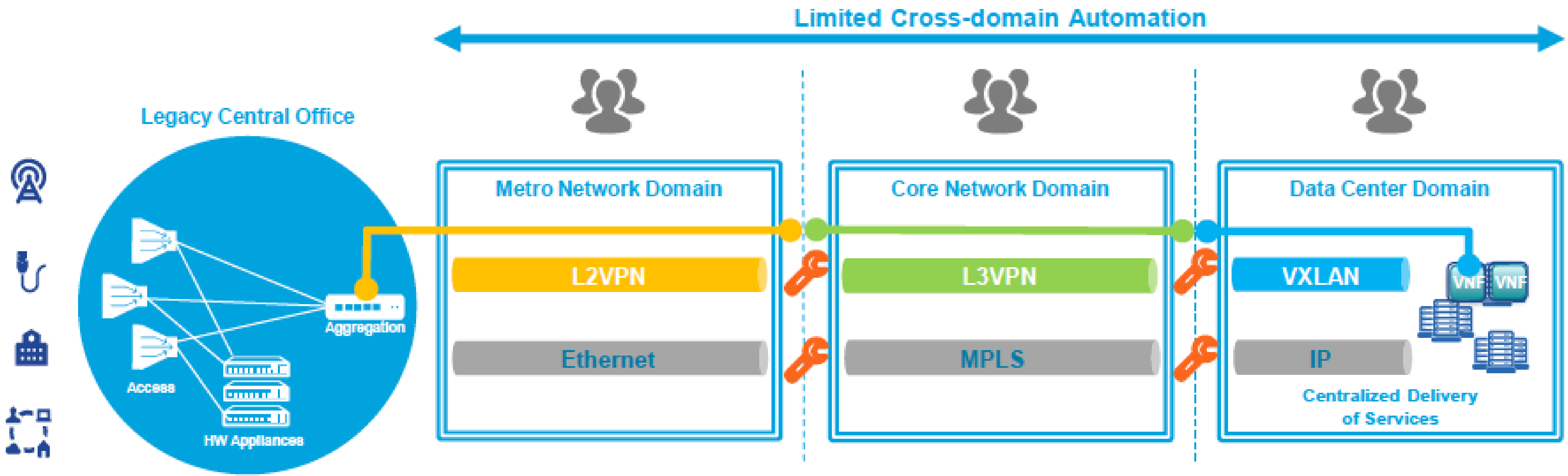
✓ **Major scalability issues**
✓ **Impediment to service creation**
✓ **Operational challenges**

SR solves Traffic Engineering Requirements for SPs without the complexity of RSVP-TE

Segment Routing enabled infrastructure
end to end connectivity, automation, protection, policy & SLA



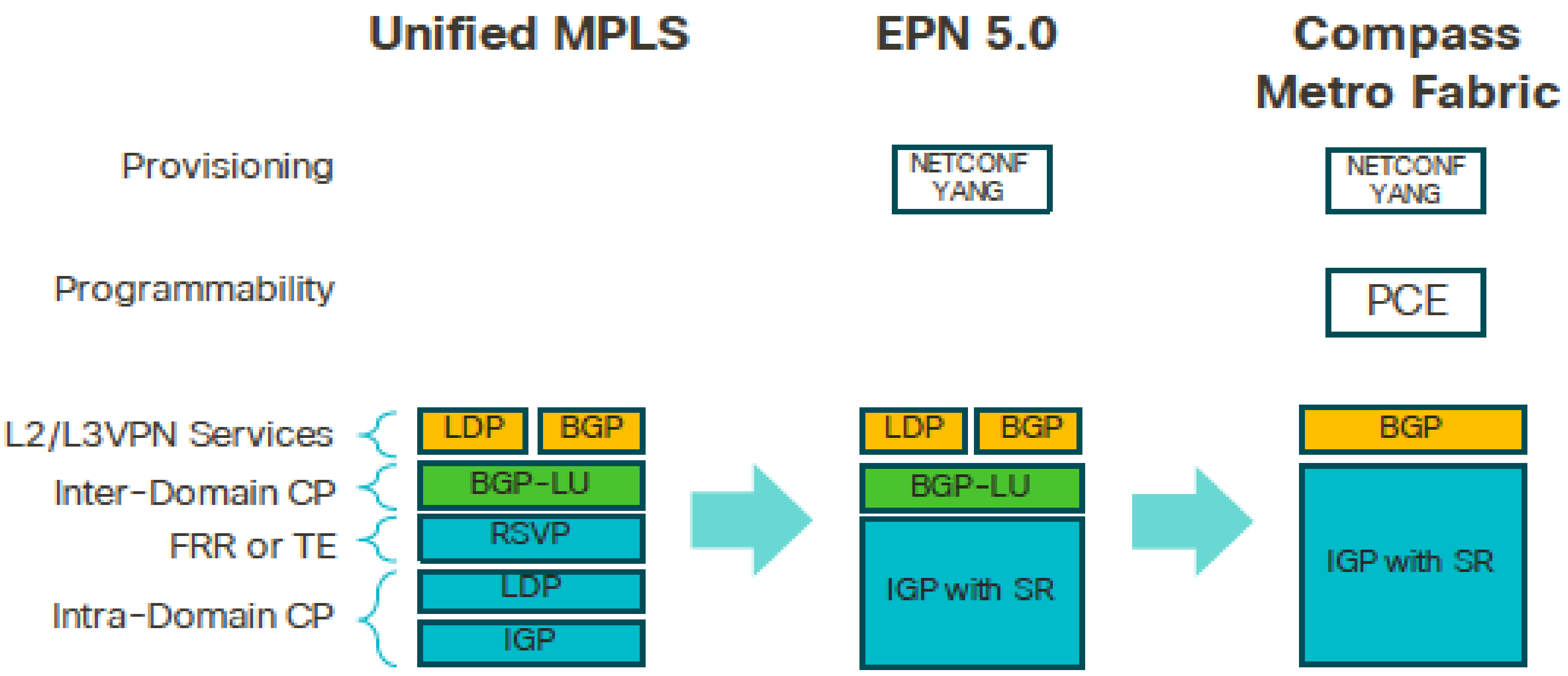
Understanding Todays Service Creation



E2E service provisioning is lengthy and complex:

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

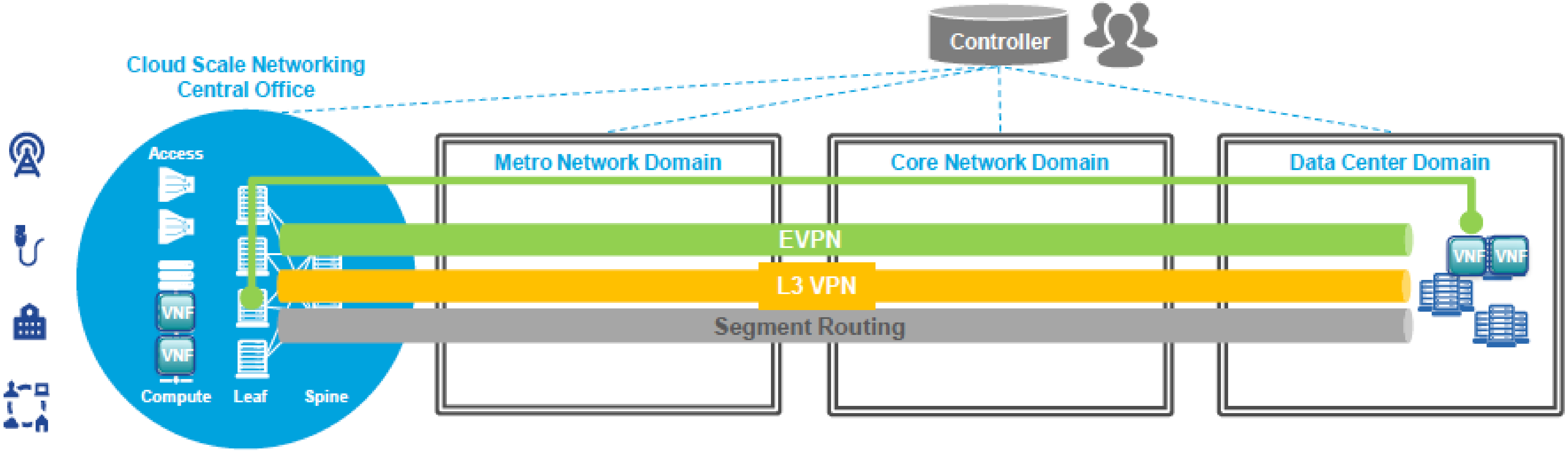
Service Provider Network - Simplification Journey



Do more with less !!

https://xrdocs.github.io/design/

Unified "Network as a 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

LDP/RSVP Label Distribution Comparison

	RSVP-TE	IGP+LDP	Segment routing MPLS
Basic mpls transport	Pre-signalled tunnels	IGP + LDP	IGP
Deployment	Tunnels must be manually set up	Automatic	Automatic
IGP/LDP synchronisation	N/A	Problem to manage	N/A
50msec FRR	Any topology	RLFA	IGP
ECMP-capability for TE	Yes	No	Yes
Seamless Interworking with classic MPLS and incremental deployment	Yes	N/A	Yes
Engineered for SDN	No	No	Yes

Segment Routing work at IETF

- An IETF Working Group called **SPRING** was created in the Routing Area (2013) to handle Segment Routing work at IETF
- SPRING stands for *Source Packet Routing In NetworkinG*
- Definition of Problem Statement, Use Cases and Requirements for Segment Routing Architecture are discussed in SPRING WG
- Protocol extensions required for SR are defined in the specific WG's:
  - ISIS, OSPF, IDR(Inter Domain Routing, for BGP), PCE (Path Computation Element), 6man (IPv6 maintenance)

Segment Routing Standardization

- RFC 7855 (May 2016), RFC 8354, RFC 8355, RFC 8402, RFC 8403
- Protocol extensions progressing in multiple groups
  - IS-IS
  - OSPF
  - PCE
  - IDR
  - 6MAN
  - BESS

Sample IETF Documents
Problem Statement and Requirements (RFC 7855)
Segment Routing Architecture (RFC 8402)
IPv6 SPRING Use Cases (RFC 8354)
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)

6 WG IETF drafts. 85 related drafts

Industry Eco-System

Public references

- Google, Facebook, Microsoft, Yandex, Apple, Amazon…
- Comcast, DT, Orange, BT, Telecom Italia…

Hidden but active participation

- Financial, large enterprises…

Multi-vendor support

- Cisco, Alcatel, Juniper, all cooperating on IETF standardization...

Support SR in open source projects

- Linux, OVS, ONF

SR TechField Day Presentations by Walmart, Microsoft Azure, Comcast...

https://www.youtube.com/playlist?list=PLinuRwpnsHacUlfUCrVstvpzURnK_M3iI&feature=view_all

Wide adoption of Segment Routing

Availability
Now
IOS XR
IOS XE
NexOS

Deployments
SP Core/Edge
SP Metro/Aggregation
WEB
Large Enterprises

Multi-vendor
Consensus
Interop testing
First at MPLS WC 2015

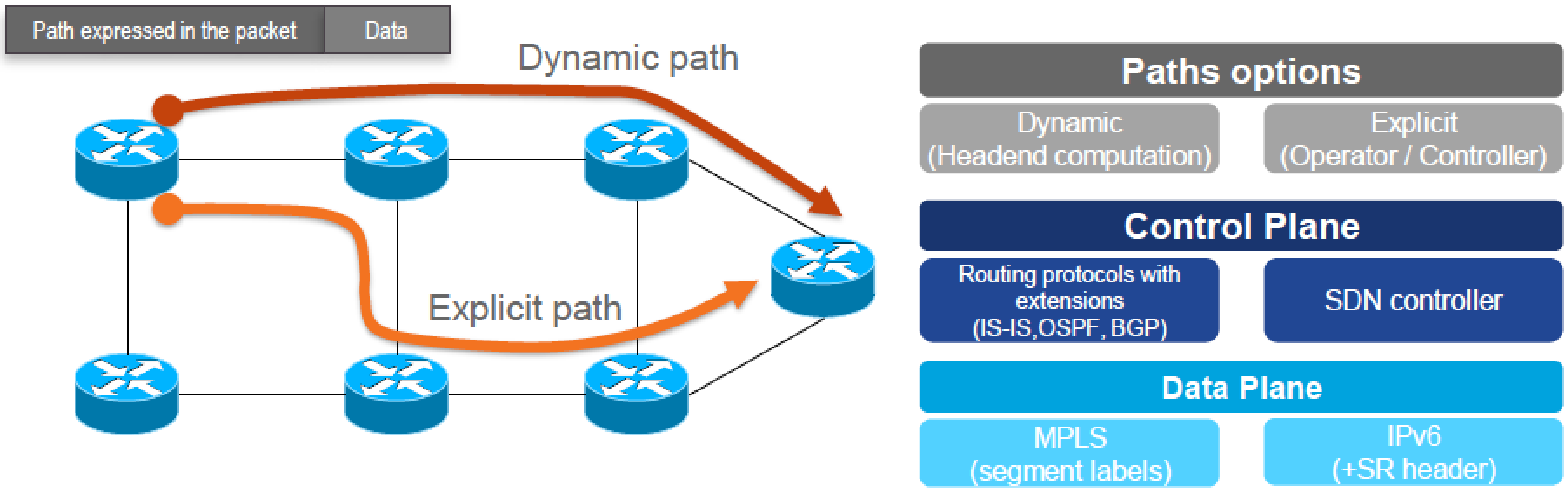
Segment Routing Product Support (2018)

- Cisco Platforms
  - **IOS-XR** (**ASR9000**, CRS, NCS5000, **NCS5500**, NCS540, NCS560, NCS6000, XRv9K)
  - **IOS-XE** (ASR1000, CSR1000v, ASR903, ASR907, ASR920, ISR4400)
  - **NX-OS** (N3K, N9K)
  - **Open Source** (FD.io/VPP, Linux Kernel, ODL, ONOS, OpenWRT)
  - **PCE** (WAN Automation Engine, SR-PCE)

Technology Overview

Segment Routing

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

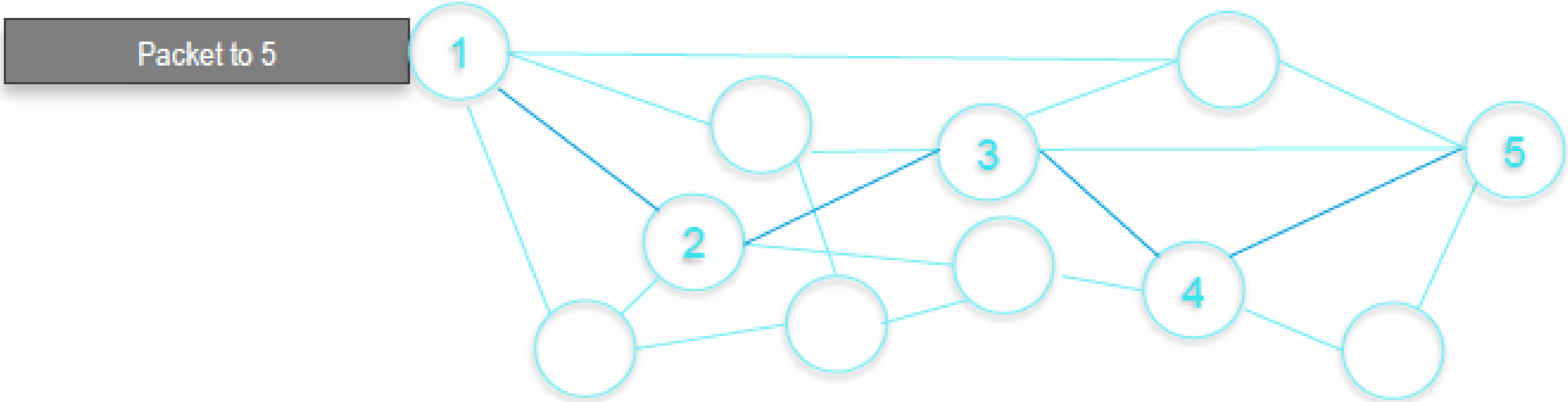


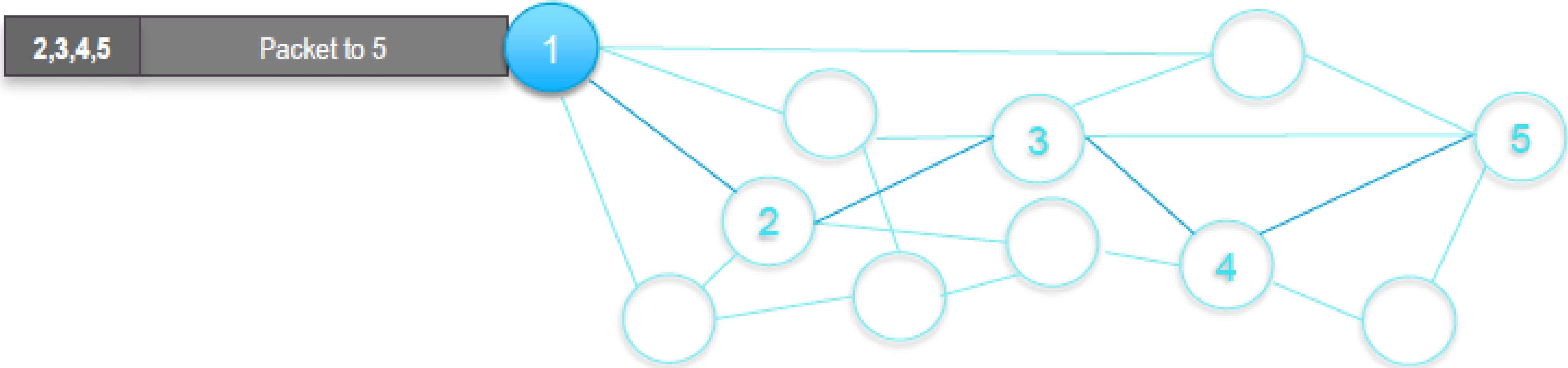
Segment Routing Overview

- **Scalable end-to-end policy:** less Label Databases, less TE LSP
  - Leverage MPLS services & hardware
- **Designed for IP and SDN**
  - Each engineered application flow is mapped on a path
  - A path is expressed as an ordered list of segments
  - The network maintains segments
- **Simple**: less Protocols, less Protocol interaction
  - No requirement for signalling protocols: RSVP, LDP
- **Forwarding** based on MPLS label (no change to MPLS forwarding plane)
- **Label distributed** by the IGP protocol with simple ISIS/OSPF extensions
- **50msec FRR** service level guarantees via LFA in any topology
- Service model intact

Segment Routing

- **Source Routing**
  - the source chooses a path and encodes it in the packet header as an ordered list of segments
  - the rest of the network executes the encoded instructions
- **Segment**: an identifier for any type of instruction
  - forwarding or service
- This presentation: **IGP-based forwarding construct**


Packet to 5
1
2
3
4
5


2,3,4,5
Packet to 5
1
2
3
4
5

Segment Routing – Forwarding Plane

- MPLS: an ordered list of segments is represented as a stack of labels
  - Segment Routing re-uses MPLS data plane without any change
  - Segment represented as MPLS label
  - Applicable to IPv4 and IPv6 address families
- IPv6: an ordered list of segments is encoded in a routing extension header
- This presentation: MPLS data plane

Segment Routing - cont.

- In case of MPLS a Segment is a MPLS label
  - A path with multiple segments is encoded as a stack of labels
- Segment Routing re-uses MPLS data plane without any change
    - > Label Push – Label Pop – Label Swap
- Applicable to IPv4 and IPv6 address families
- Segment (label) information are distributed using IGP or BGP
  - No need to have additional protocols like LDP or RSVP-TE

Segment Routing basic mechanics: IGP segments

Global and Local Segments

- Global Segment
  - Any node in SR domain can execute the associated instruction
  - Each node in SR domain installs the associated instruction in its forwarding table
  - MPLS label pool: Value in Segment Routing Global Block (SRGB)
- Local Segment
  - Only originating node can execute the associated instruction
  - MPLS label pool: locally allocated label

Global Segments – Global Label Indexes

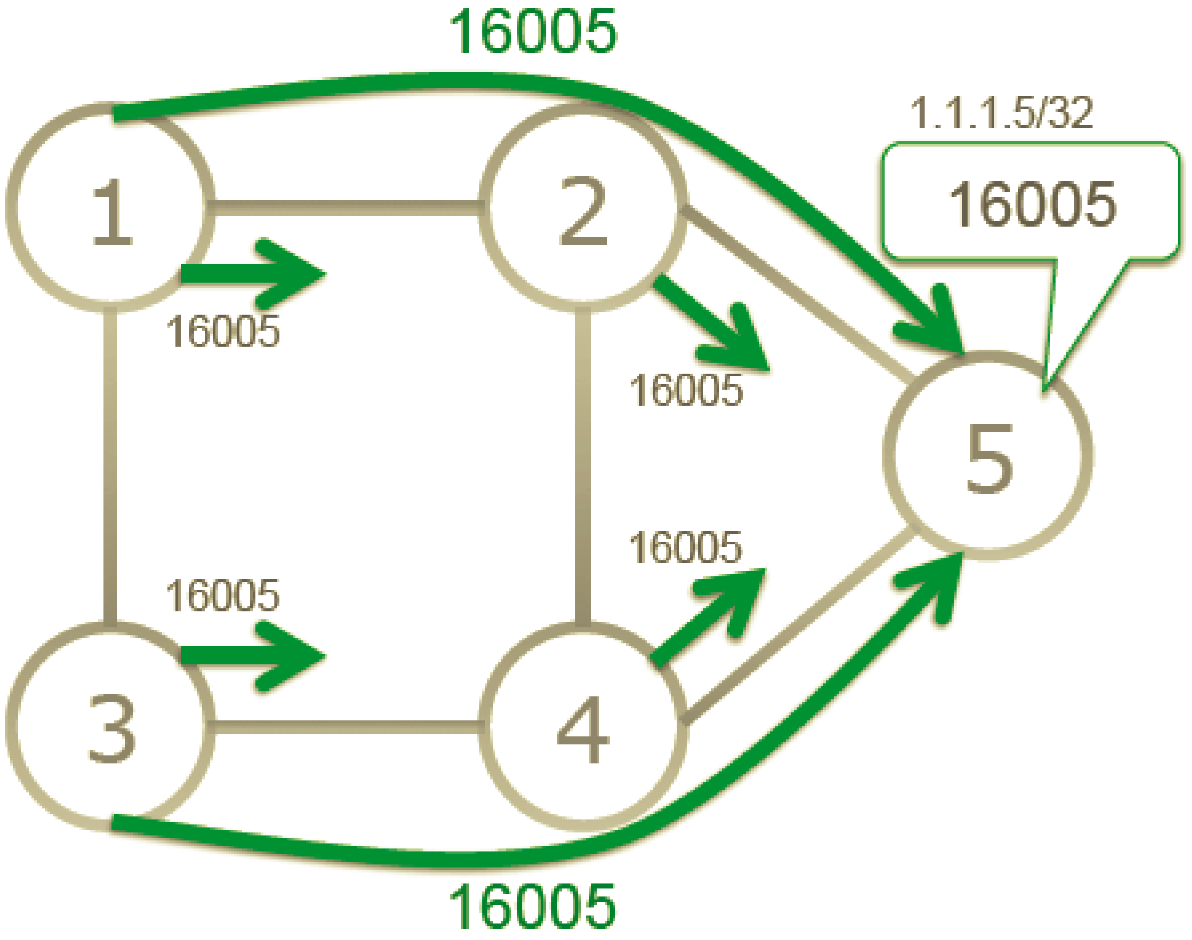
- Global Segments always distributed as a label range (SRGB) + Index
  - Index must be unique in Segment Routing Domain
- Best practice: same SRGB on all nodes
  - “Global model”, requested by all operators
  - Global Segments are global label values, simplifying network operations
  - Default SRGB: 16,000 – 23,999
    - Other vendors also use this label range

IGP segments

- Two basic building blocks distributed by IGP
  - Prefix Segments
  - Adjacency Segments

IGP Prefix Segment

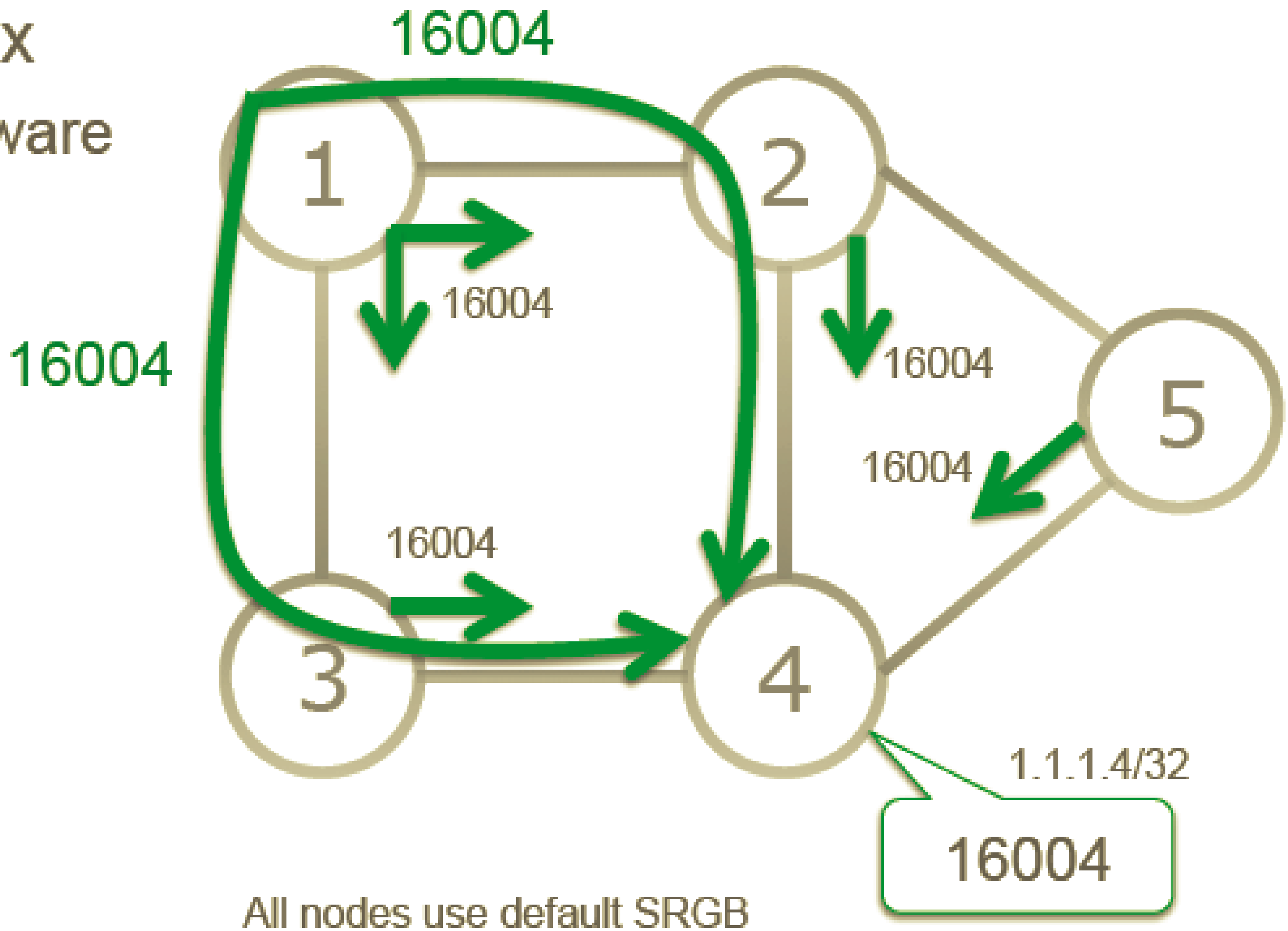
- Shortest-path to the IGP prefix
  - Equal Cost MultiPath (ECMP)-aware
- Global Segment
- Label = 16000 + Index
  - Advertised as index
- Distributed by ISIS/OSPF



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

IGP Prefix Segment

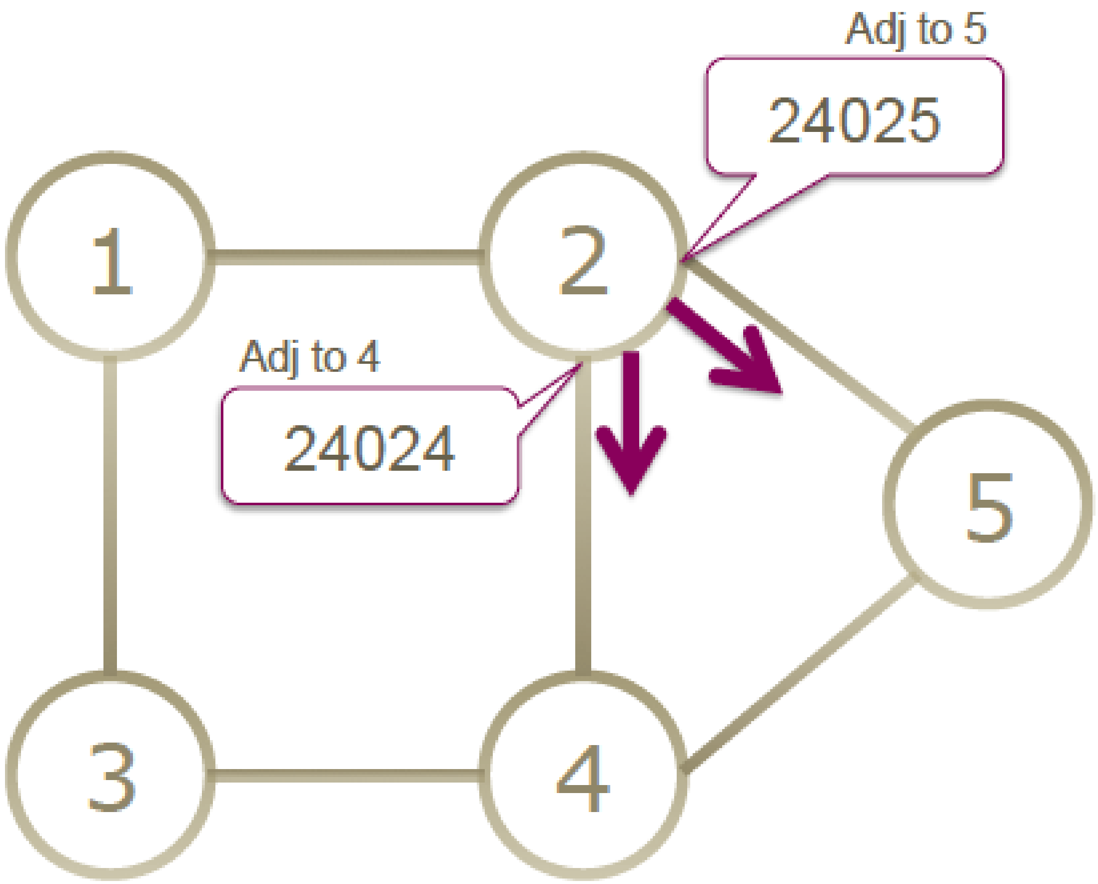
- Shortest-path to the IGP prefix
  - Equal Cost MultiPath (ECMP)-aware
- Global Segment
- Label = 16000 + Index
  - Advertised as index
- Distributed by ISIS/OSPF



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

IGP Adjacency Segment

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

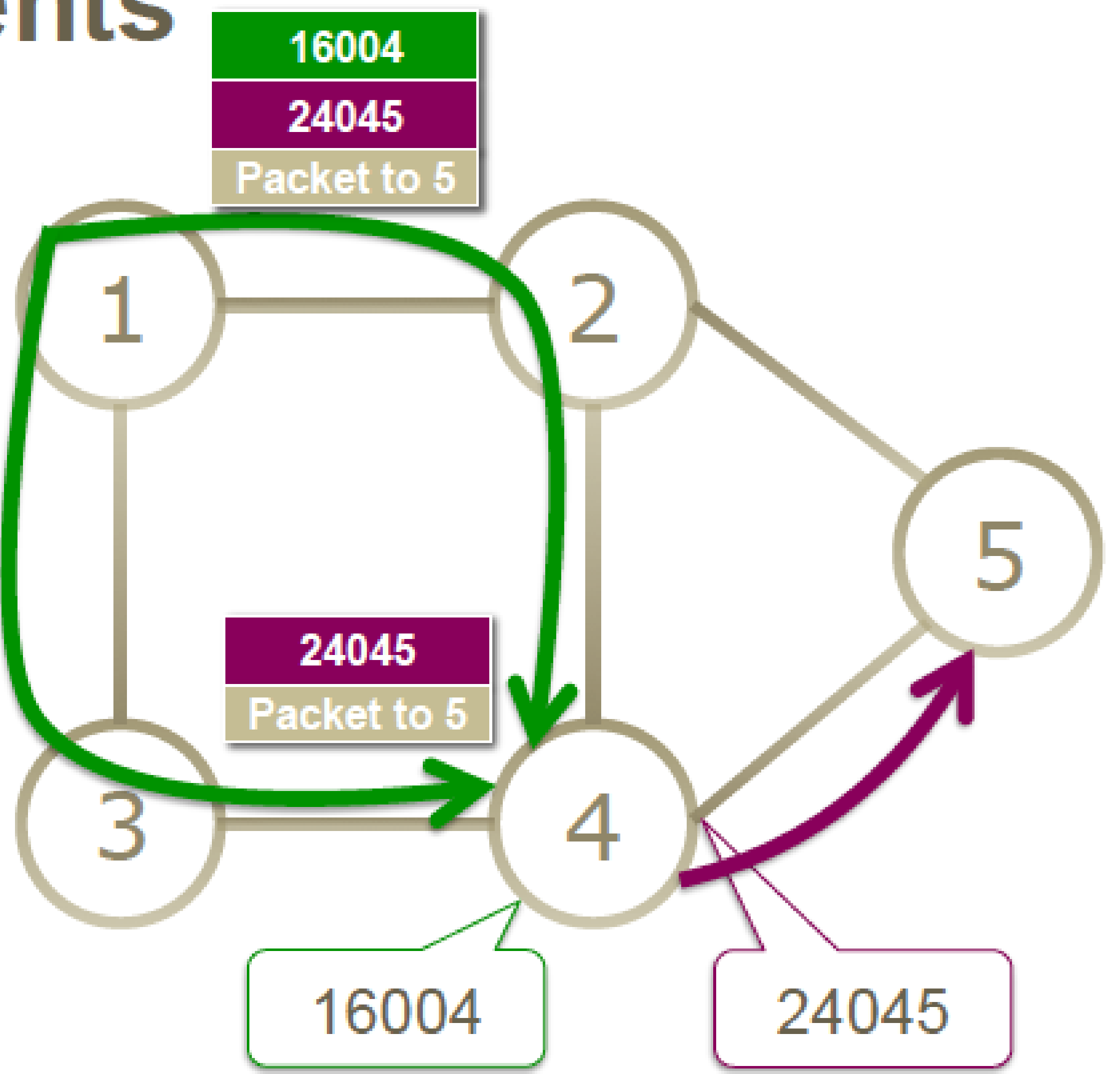


All nodes use default SRGB
16,000 – 23,999

Combining IGP Segments

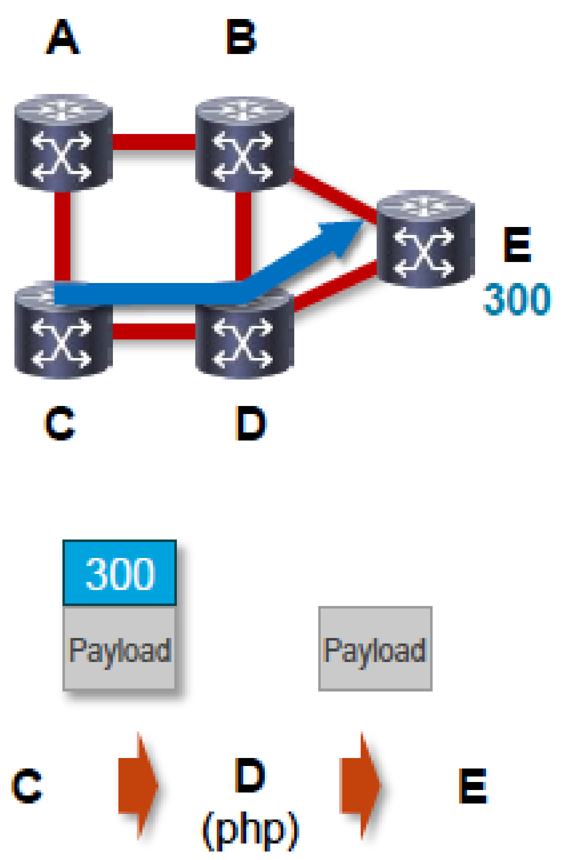
- Steer traffic on any path through the network
- Path is specified by list of segments in packet header, a stack of labels
- No path is signaled
- No per-flow state is created
- Single protocol: IS-IS or OSPF

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

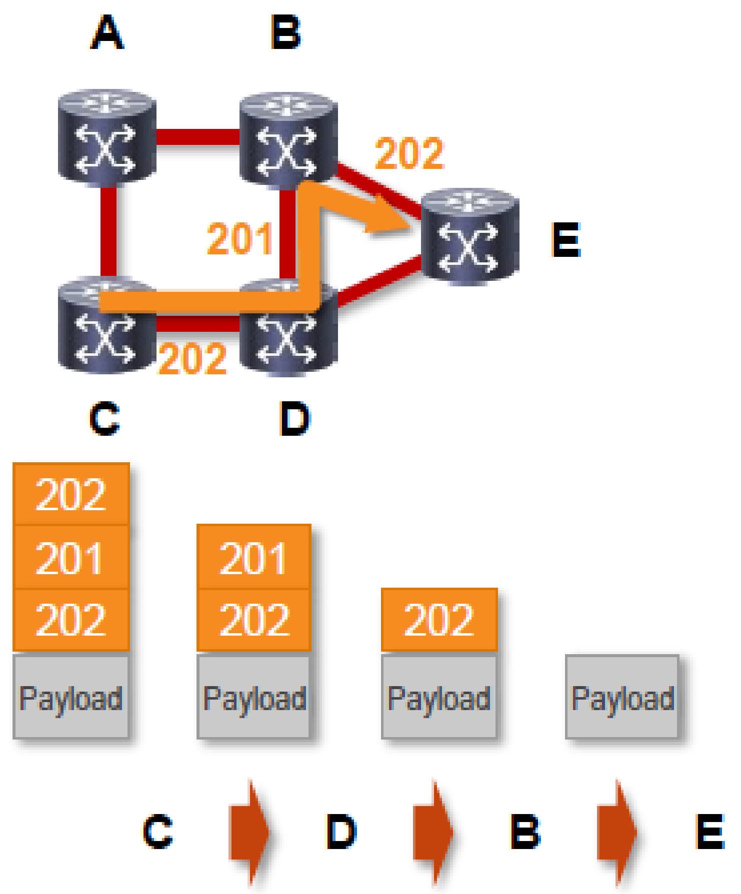


Segment Routing Forwarding Plane

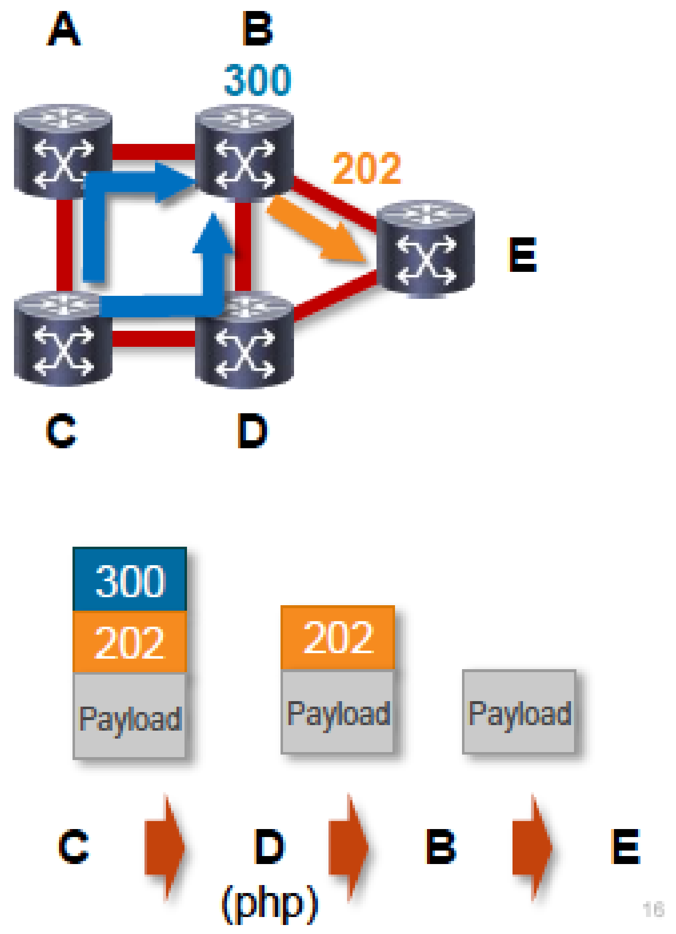
Node Path



Adjacency Path

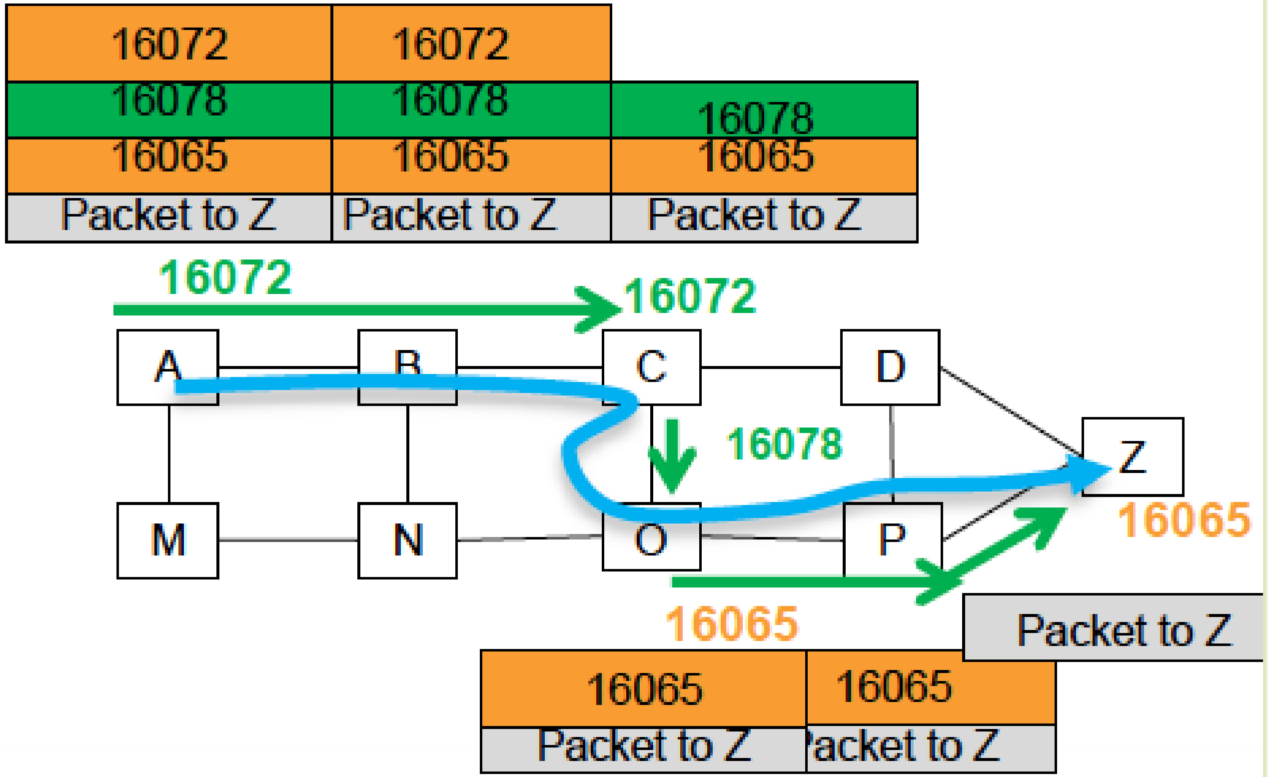


Combined Path

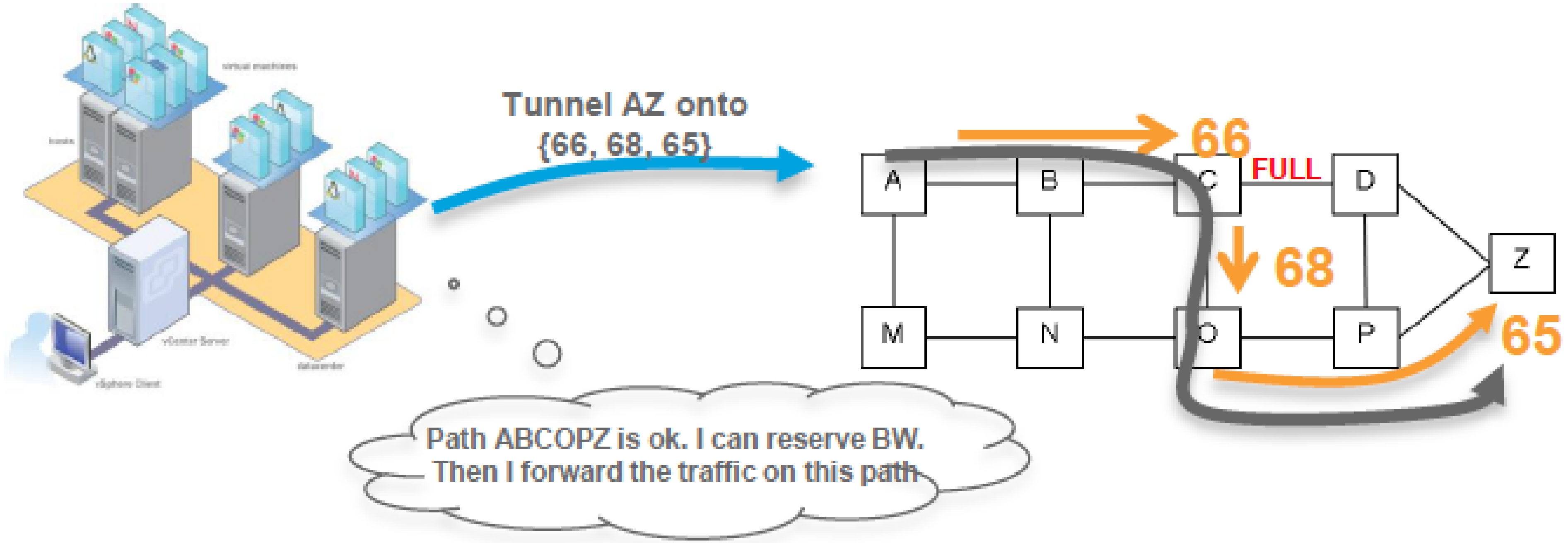


Explicit Path as Segment List

- ECMP
  - Node segment
- Per-flow state only at head-end
  - not at midpoints
- Source Routing
  - Source path can be programmed by application



Cloud Integration



The network is simple, highly programmable and responsive to rapid changes

Segment Routing Global Block
SRGB

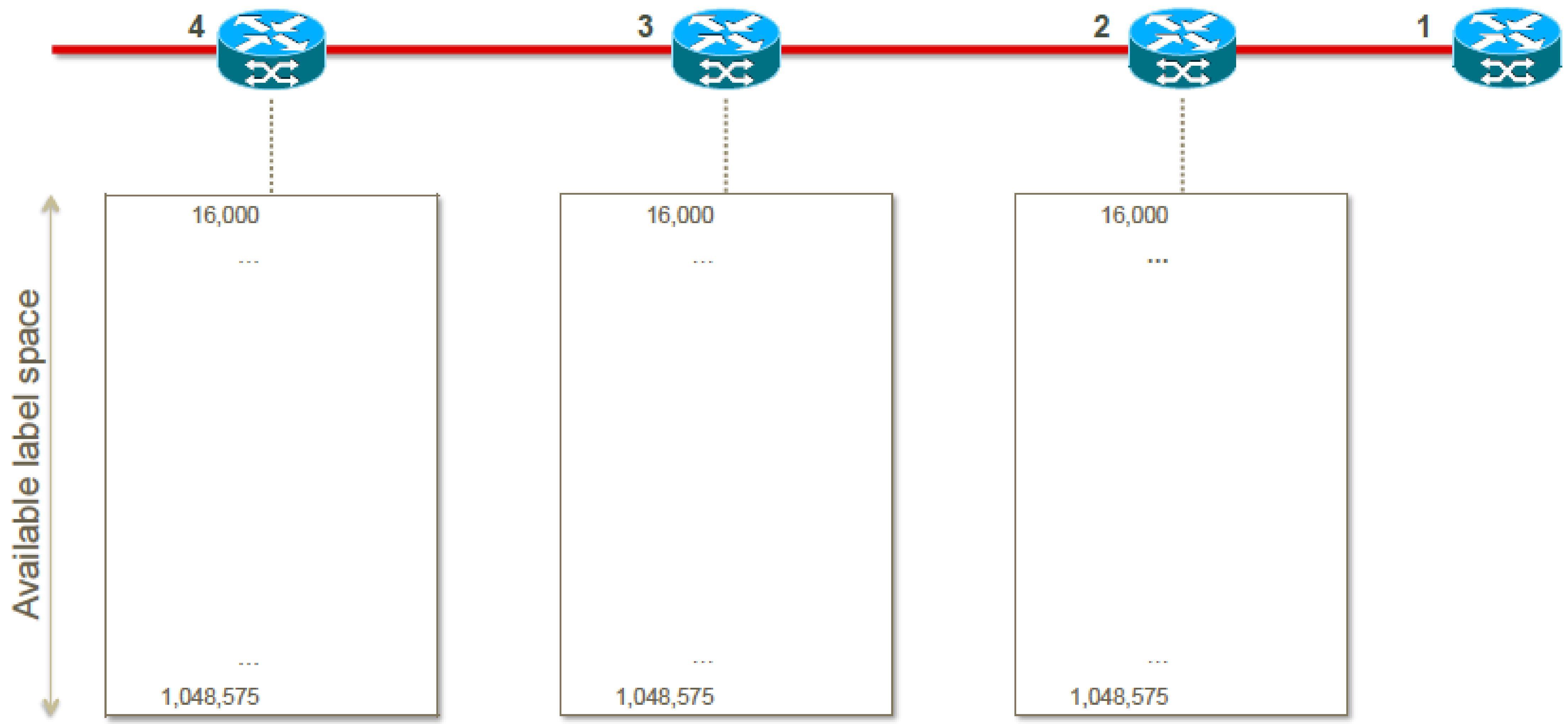
Segment Routing Global Block (SRGB)

- Segment Routing Global Block
  - Range of labels reserved for Segment Routing Global Segments
  - Default SRGB is 16,000 – 23,999
- A prefix-SID is advertised as a domain-wide unique index
- The Prefix-SID index points to a unique label within the SRGB
  - Index is zero based, i.e. first index = 0
  - Label = Prefix-SID index + SRGB base
  - E.g. Prefix 1.1.1.65/32 with prefix-SID index 65 gets label 16065

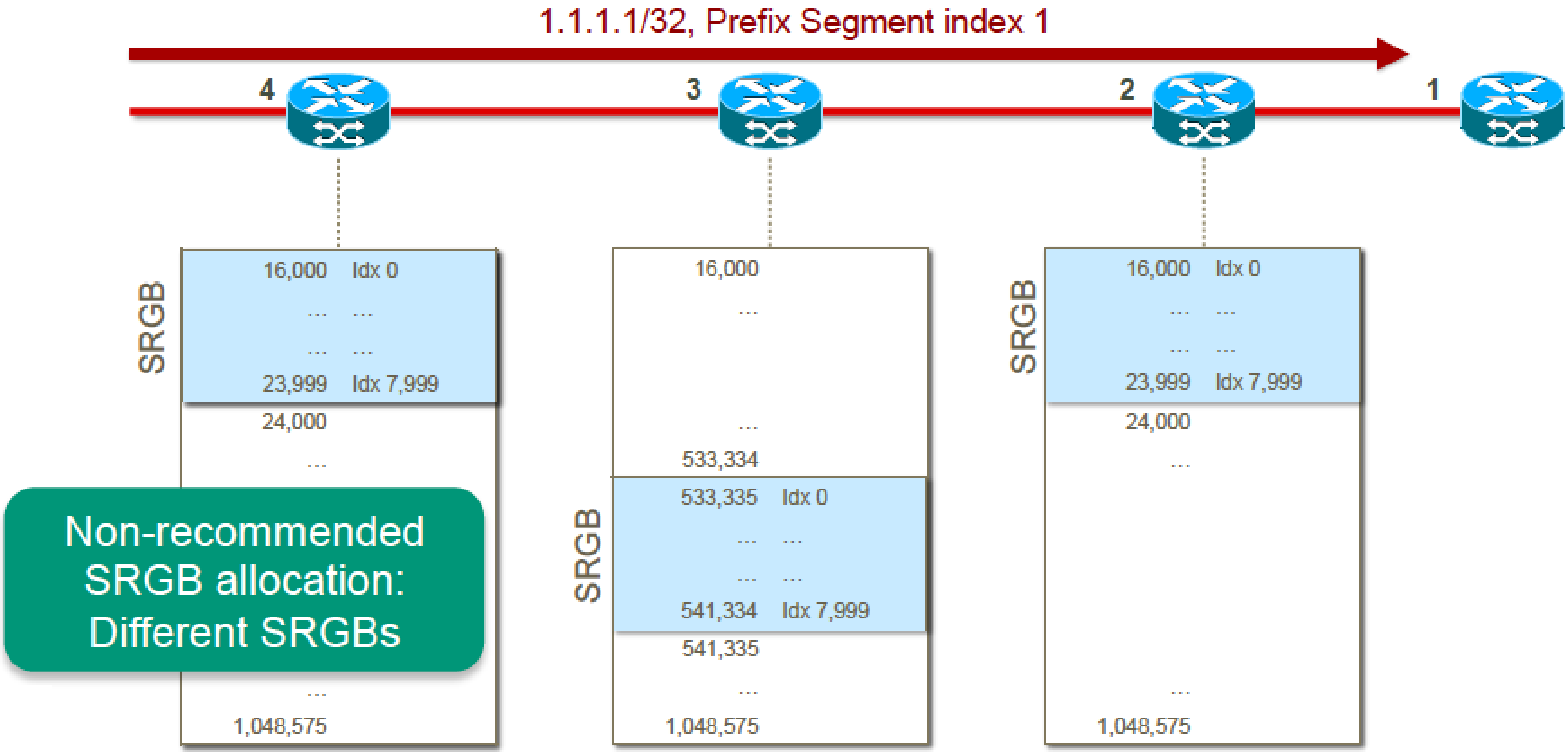
Segment Routing Global Block (SRGB)

- Strongly recommended to use same SRGB on all nodes
  - All operators asked for this deployment model
  - Simple, straightforward
  - Global Segment == Global Label value
  - Using different SRGBs is supported, but complicates operations for user
- A non-default SRGB can be allocated between 16,000 and 1,048,575
  - Or up to the platform limit, if any
- The size of the SRGB should be equal on all nodes
  - Current maximum size is 64k

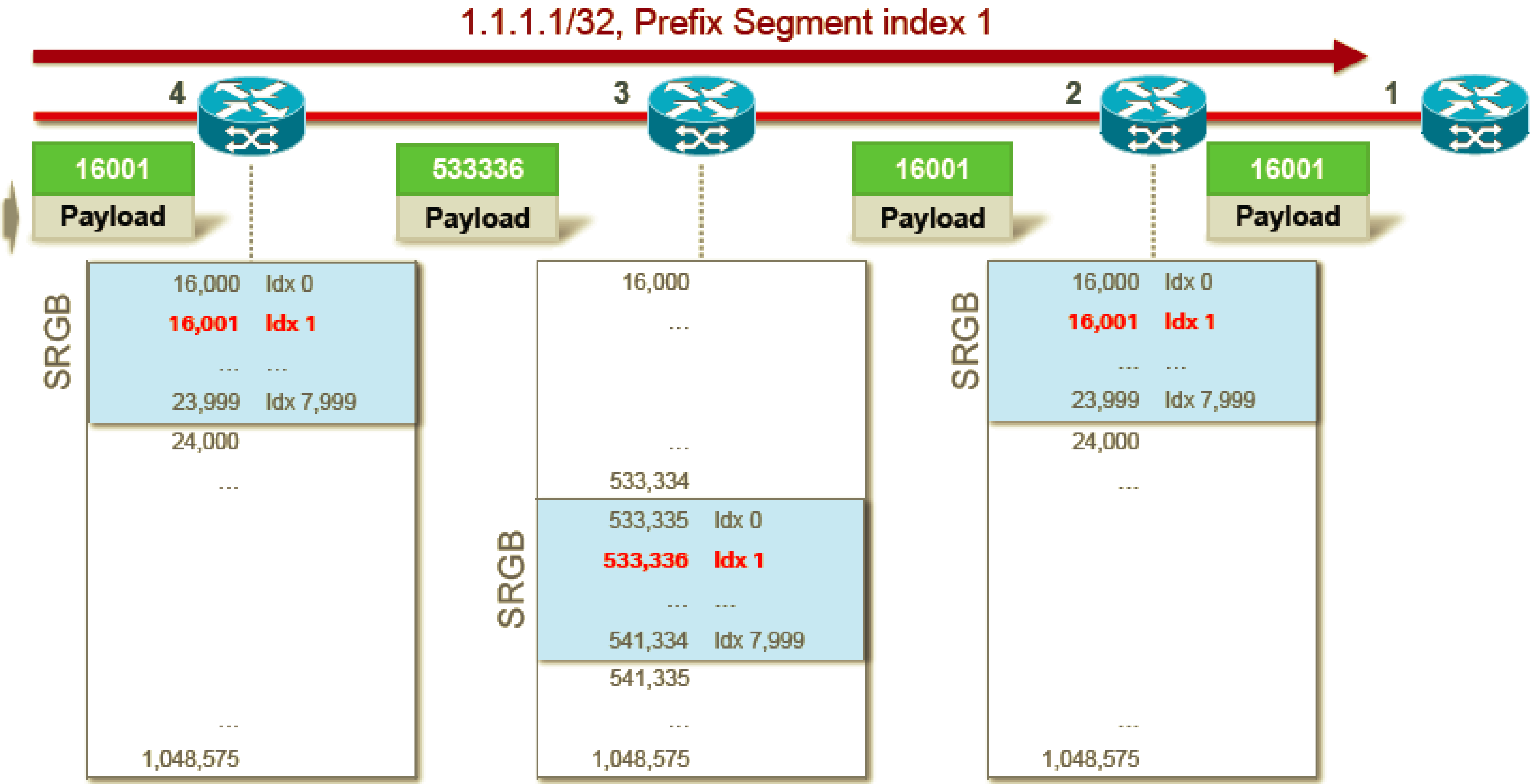
Segment Routing Global Block (SRGB)



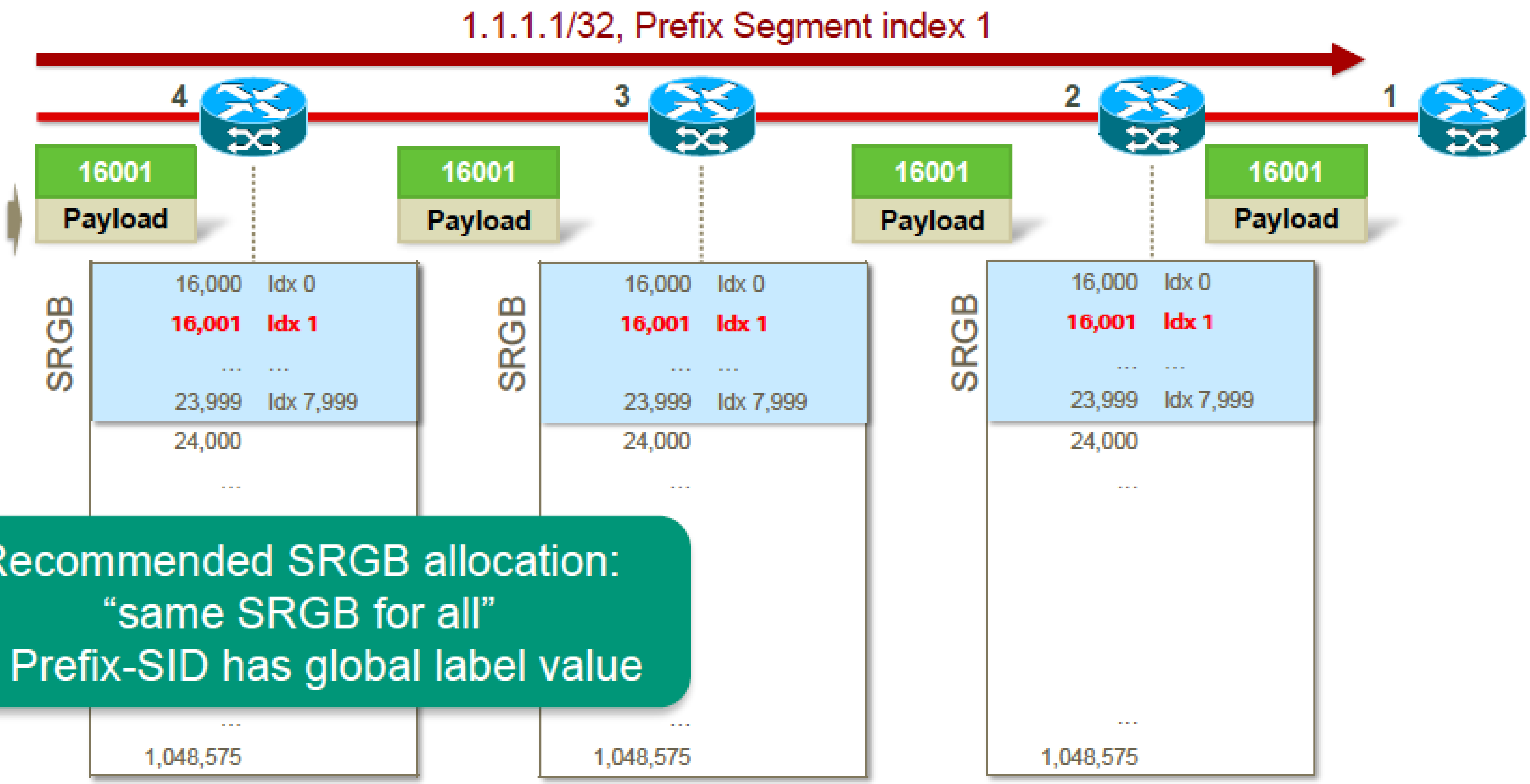
Not recommended, but possible SRGB allocation



Not recommended, but possible SRGB allocation

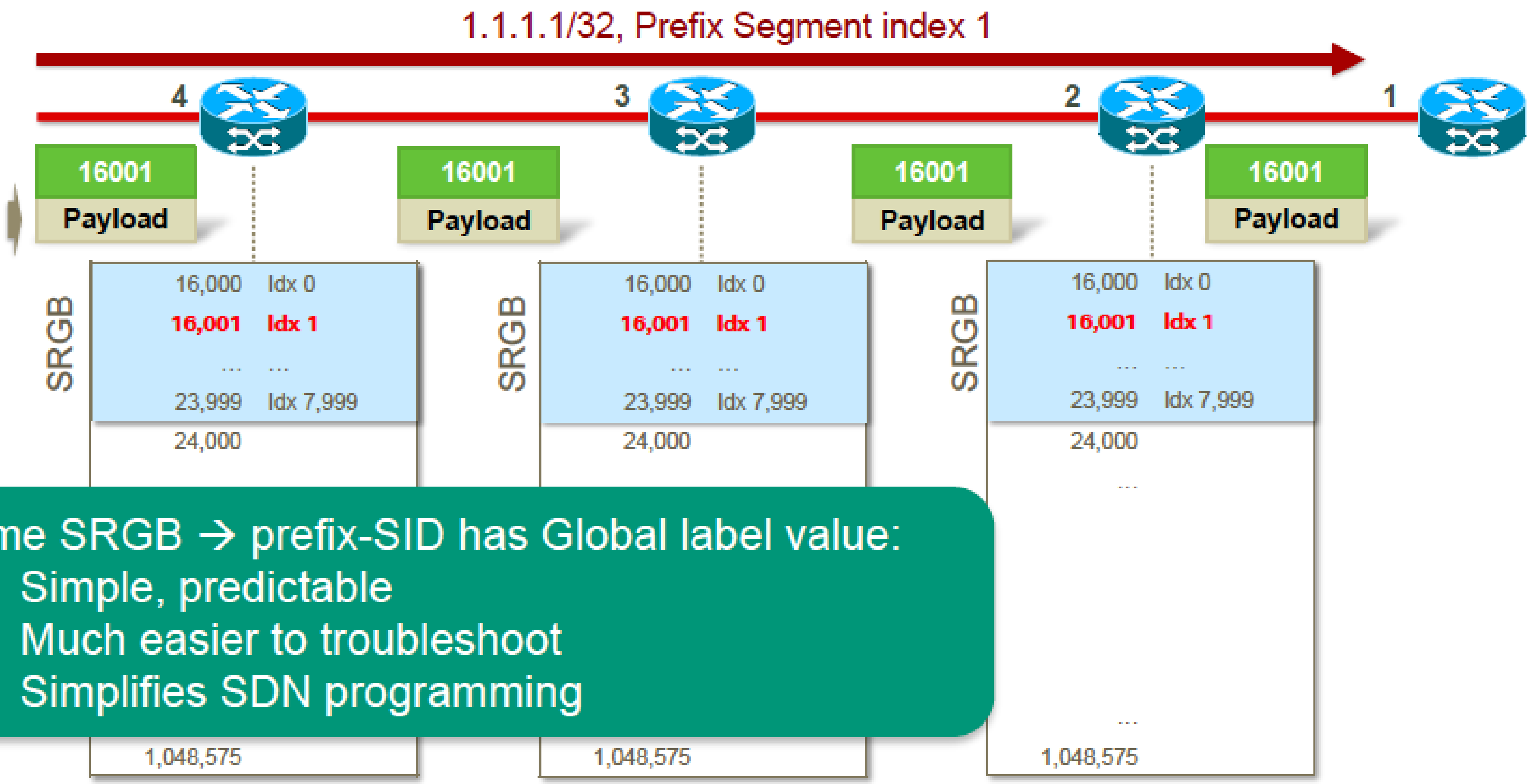


Recommended SRGB allocation



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

Label Switching Database (LSD)

- Local label allocation is managed by Label Switching Database (LSD)
- MPLS Applications must register as client with LSD to allocate labels
  - MPLS Applications are e.g. IGP, LDP, RSVP, MPLS static, …
- Label space carving of Segment Routing capable software release (even if Segment Routing is not enabled) :
  - Label range [0-15] reserved for special-purposes
  - Label range [16-15,999] reserved for static MPLS labels
  - Label range [16,000-23,999] preserved for SRGB
  - Label range [24,000-max] used for dynamic label allocation

SRGB label range preservation

- LSD preserves the default SRGB label range [16,000-23,999]
  - In any Segment Routing capable software release
  - Even if Segment Routing is not enabled
  - Except if the configured mpls label range includes this default range
- LSD allocates dynamic labels starting from 24,000
- If the configured mpls label range includes the default SRGB label range, the default preservation is disabled
  - E.g. `mpls label range 16000 1048575`

SRGB label range preservation

- Preservation of the default SRGB label range makes future Segment Routing activation possible without reboot
  - No labels are allocated from that preserved range. When enabling Segment Routing with default SRGB some time in the future, that label range is available and ready for use
  - See illustration further in this presentation

Segment Routing Global Block (SRGB)
Notes

- Multiple IGP instances can use the same SRGB or use different non-overlapping SRGBs
- Modifying a SRGB configuration is disruptive for traffic
  - And may require a reboot if the new SRGB is not (entirely) available

Segment Routing Global Block (SRGB)
Default SRGB

```
RP/0/0/CPU0:xrvr-1#show mpls label table detail
Table Label    Owner                               State
----- ------- ------------------------------- ------
<...snip...>
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=[illegible] type=0, intf=[illegible]/0/0/0, nh=10.0.0.2)
```

IS-IS SRGB

Start_label = 16,000

Size = 8,000

Default SRGB label block allocation for ISIS
[16,000 – 23,999]

Segment Routing Global Block (SRGB)
Non-default SRGB Example

```
router isis 1
 segment-routing global-block 18000 19999
```

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

```
RP/0/0/CPU0:xrvr-1#show mpls label table detail
Table Label   Owner                           State
----- ------- ------------------------------- ------
<...snip...>
0     18000   ISIS(A):1                       InUse  No
  (Lbl-blk SRGB, vers:0, (start_label=18000, size=2000)
0     24000   ISIS(A):1                       InUse  Yes
  (SR Adj Segment IPv4, vers:0, index=  type=0, intf=  /0/0/0, nh=10.0.0.2)
```

IS-IS SRGB

Start_label = 18,000

Size = 2,000

Non-default SRGB label block allocation for ISIS
[18,000 – 19,999]

Control Plane and Data Plane

Segment Routing – IGP Control plane

- Using IS-IS or OSPF to distribute segments
- Configuring Segment Routing under IGP
- Segment Routing in a multi-area, multi-level network
- Verifying Segment Routing advertisements

SR OSPF Control Plane Overview

- OSPF Segment Routing functionality
  - OSPFv2 control plane
  - Multi-area
  - IPv4 Prefix Segment ID (Prefix-SID) for host prefixes on loopback interfaces
  - Adjacency Segment ID (Adj-SIDs) for adjacencies
    - >Non-protected adj-SIDs and protected (since OSPF SR-TE release) adj-SIDs
  - MPLS penultimate hop popping (PHP) and explicit-null signaling

OSPF Extensions

- OSPF adds to the Router Information Opaque LSA (type 4):
  - SR-Algorithm TLV (8)
  - SID/Label Range TLV (9)
- OSPF defines new Opaque LSAs to advertise the SIDs
  - OSPFv2 Extended Prefix Opaque LSA (type 7)
    - >OSPFv2 Extended Prefix TLV (1)
      - Prefix SID Sub-TLV (2)
  - OSPFv2 Extended Link Opaque LSA (type 8)
    - >OSPFv2 Extended Link TLV (1)
      - Adj-SID Sub-TLV (2)
      - LAN Adj-SID Sub-TLV (3)

OSPF Configuration

- OSPFv2 control plane
- Required
  - Enable segment-routing under instance or area(s)
    - Command has area scope, usual inheritance applies
  - Enable segment-routing forwarding under instance, area(s) or interface(s)
    - Command has interface scope, usual inheritance applies
- Optional
  - Prefix-SID configured under loopback(s)
- MPLS forwarding enabled on all OSPF interfaces with segment-routing forwarding configured

OSPF Segment Routing Configuration
Recommended

```
router ospf 1
 segment-routing mpls
 segment-routing forwarding mpls
```

In a later release, SR forwarding will be enabled by default. This config line will no longer be required. (CSCuw93707)

- `segment-routing forwarding mpls` must be configured to install SIDs – received by OSPF – in the forwarding table
- MPLS forwarding is enabled on all `segment-routing forwarding` enabled OSPF interfaces
- Adjacency-SIDs are allocated and distributed for `segment-routing forwarding` enabled adjacencies
- Configuration under ospf instance is recommended, but can be customized

OSPF Segment Routing Configuration

```
router ospf 1
 area 0
  segment-routing mpls                    !! Area command
  segment-routing forwarding mpls         !! Interface command
  interface GigabitEthernet0/0/0/0
   segment-routing forwarding disable !! Interface command
```

- `segment-routing mpls` is an ospf area command, can be applied per area
  - Ospf inheritence rules are applicable
- `segment-routing forwarding mpls` is an ospf interface command, can be applied per interface
  - Ospf inheritence rules are applicable
- In the example, SR is enabled for all interfaces in area0, except Gi0/0/0/0

SR IS-IS Control Plane Overview

- Level 1, level 2 and multi-level routing
- Prefix Segment ID (Prefix-SID) for host prefixes on loopback interfaces
- Adjacency SIDs for adjacencies
- Prefix-to-SID mapping advertisements (mapping server)
- MPLS penultimate hop popping (PHP) signalling
- MPLS explicit-null label signalling

IS-IS TLV Extensions

- SR for IS-IS introduces support for the following (sub-)TLVs:

– SR Capability sub-TLV (2)	IS-IS Router Capability TLV (242)
– Prefix-SID sub-TLV (3)	Extended IP reachability TLV (135)
– Prefix-SID sub-TLV (3)	IPv6 IP reachability TLV (236)
– Prefix-SID sub-TLV (3)	Multitopology IPv6 IP reachability TLV (237)
– Prefix-SID sub-TLV (3)	SID/Label Binding TLV (149)
– Adjacency-SID sub-TLV (31)	Extended IS Reachability TLV (22)
– LAN-Adjacency-SID sub-TLV (32)	Extended IS Reachability TLV (22)
– Adjacency-SID sub-TLV (31)	Multitopology IS Reachability TLV (222)
– LAN-Adjacency-SID sub-TLV (32)	Multitopology IS Reachability TLV (222)
– SID/Label Binding TLV (149)	

- Implementation based on *draft-ietf-isis-segment-routing-extensions-02*

IS-IS Configuration

- Required
  - Wide metrics
  - SR enabled under address family IPv4 unicast
- Optional
  - Prefix-SID configured under loopback(s) AF IPv4
- MPLS forwarding enabled automatically on all (non-passive) IS-IS interfaces
- Adjacency-SIDs are automatically allocated for each adjacency

IS-IS Segment Routing Configuration

```
router isis 1
 address-family ipv4|ipv6 unicast
  metric-style wide
  segment-routing mpls
 !
```

enable SR capability

enable SR for the MPLS data plane

- MPLS forwarding is enabled on all non-passive IS-IS interfaces
- Adjacency-SIDs are allocated and distributed for all adjacencies
  - Non-protected adj-SIDs and protected (since IOS XR 5.3.2) adj-SIDs
  - See SR-TE section

IS-IS Segment Routing Configuration

```
router isis 1
 address-family ipv6 unicast
  metric-style wide
  segment-routing ipv6
 !
```

enable SR capability

enable SR for the IPv6 extension-header data plane

- SRv6 Extension Header data plane is outside the scope of this presentation

Segment Routing Global Block

- Default SRGB is [16,000-23,999]
  - Default SRGB configuration not shown in configuration
- Non-default SRGB can be configured per IGP instance
- Multiple IGP instances can use the same SRGB or use different non-overlapping SRGBs
- Segment Routing Global Block can be configured in global configuration (IOS XR 6.0)
  - SRGB under IGP instance has precedence over SRGB in global configuration

Segment Routing Global Block (SRGB)
Example

```
segment-routing
 global-block 18000 19999
!
router ospf 1
 segment-routing mpls
 !! no segment-routing global-block config
```

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

```
RP/0/0/CPU0:xrvr-1#show mpls label table detail
Table Label   Owner                               State
----- ------- ------------------------------- ------
<...snip...>
0     18000   OSPF(A):ospf-1                      InUse  No
  (Lbl-blk SRGB, vers:0, (start_label=18000, size=2000)
<...snip...>
```

OSPF SRGB

Start_label = 18,000

Size = 2,000

Non-default SRGB label block allocation for OSPF
[18,000 – 19,999]

Segment Routing Global Block (SRGB)
Example

```
!! no global segment-routing global-block config
router isis 1
 segment-routing mpls
 segment-routing global-block 18000 19999
```

Configure an IGP SRGB 18,000 – 19,999

```
RP/0/0/CPU0:xrvr-1#show mpls label table detail
Table Label   Owner                           State
----- ------- ------------------------------- ------
<...snip...>
0     18000   ISIS(A):1                       InUse  No
  (Lbl-blk SRGB, vers:0, (start_label=18000, size=2000)
<...snip...>
```

IS-IS SRGB

Start_label = 18,000

Size = 2,000

Non-default SRGB label block allocation for ISIS
[18,000 – 19,999]

Segment Routing Global Block (SRGB)
Example

```
segment-routing
 global-block 18000 19999
!
router ospf 1
 segment routing mpls
 segment-routing global-block 20000 21999
```

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

Configure an IGP SRGB 20,000 – 21,999

```
RP/0/0/CPU0:xrvr-1#show mpls label table detail
Table Label   Owner                             State
----- ------- --------------------------------- ------
<...snip...>
0     20000   OSPF(A):ospf-1                    InUse  No
  (Lbl-blk SRGB, vers:0, (start_label=20000, size=2000)
<...snip...>
```

OSPF SRGB

Start_label = 20,000

Size = 2,000

Non-default SRGB label block allocation for OSPF [20,000 – 21,999]

Prefix segment

- Global Segment – Global significance
  - Unique within SR domain
- Managed by routing protocol
  - IGP allocates a block of labels (SRGB) from Label Switching Database (LSD)
- Manually configured
  - Under IGP enabled loopback interface
  - Only /32 or /128 prefixes in global routing table
- Prefix-SIDs are assigned by the operator similar to e.g. assigning loopback addresses

Node segment

- Node segment is a Prefix segment associated with a host prefix that identifies a node
  - Equivalent to a router-id prefix, which is a prefix identifying a node
  - Node-SID is prefix-SID with N-flag set in advertisement
- By default, each configured prefix-SID is a node-SID
  - "regular" (i.e. non Node-SID) prefix-SID is configurable for IS-IS

Prefix-SID / Node-SID Configuration

```
router isis 1
 interface Loopback0
  address-family ipv4|ipv6 unicast
   prefix-sid {absolute|index} {<SID value>|<SID index>}
```

```
router ospf 1
 area 0
  interface Loopback0
   prefix-sid {absolute|index} {<SID value>|<SID index>}
```

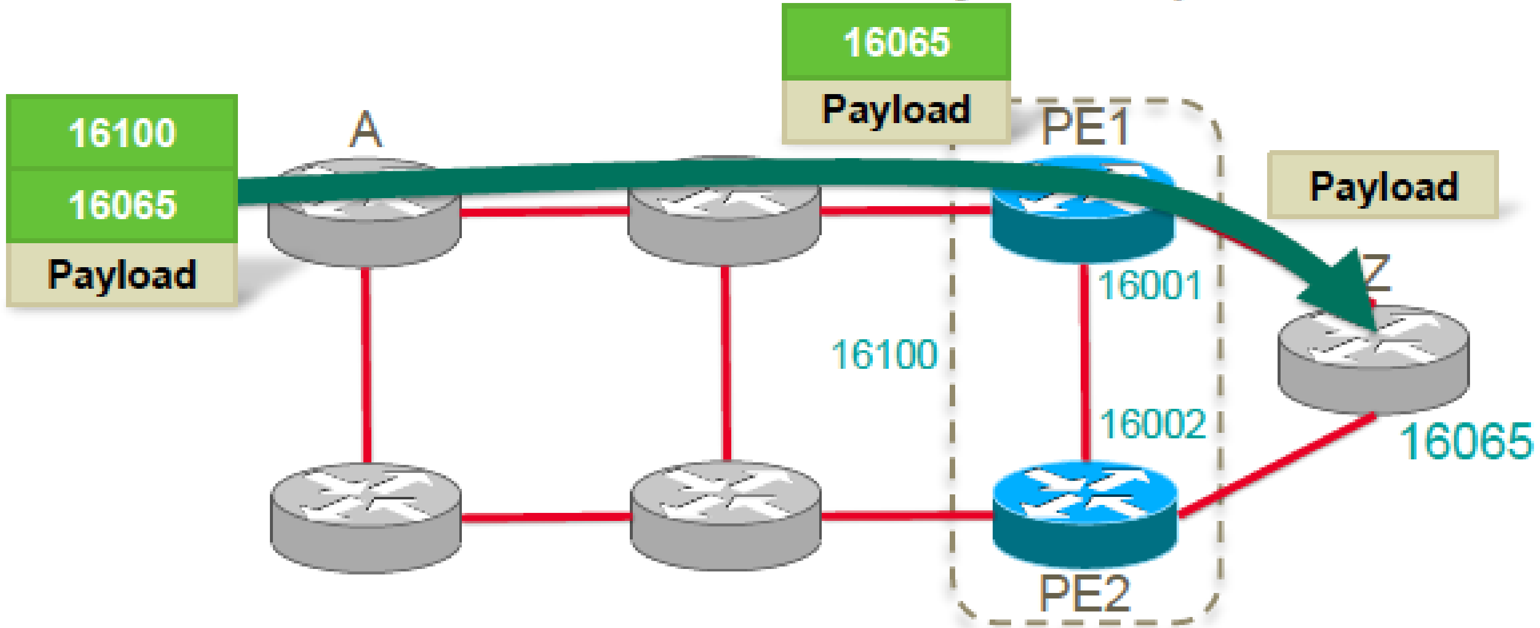
- Prefix-SID can be specified using:
  - an absolute value within the SRGB ("global mode")
  - *or* an index (offset) from the lower bound of the SRGB.

Anycast Prefix Segments

- Anycast prefixes: same prefix advertised by multiple nodes
- Anycast prefix-SID: prefix-SID associated with anycast prefix
  - Same prefix-SID for the same prefix!
- Traffic is forwarded to one of the Anycast prefix-SID originators based on best IGP path
- If primary node fails, traffic is auto re-routed to the other node
- Note: nodes advertising the same Anycast prefix-SID must have the same SRGB

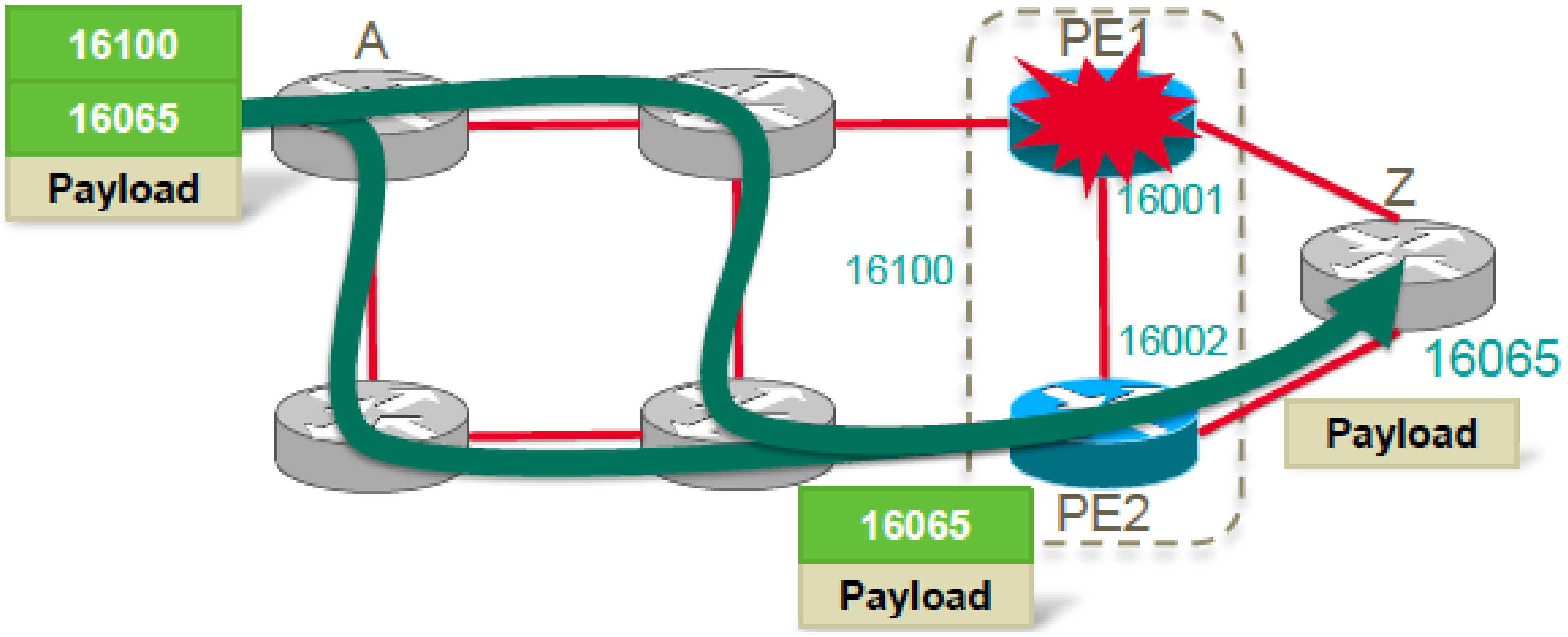
Anycast-SID – High Availability benefit

- PE1 and PE2 each advertise a prefix-SID, 16001 resp. 16002
- PE1 and PE2 both advertise an Anycast prefix-SID, 16100



Anycast-SID – High Availability benefit

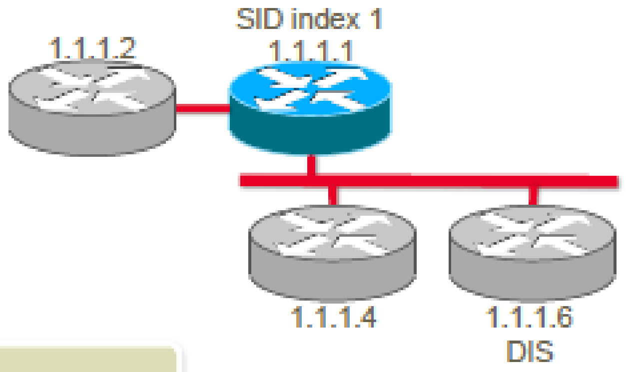
- If closest anycast node fails, traffic is auto re-routed to another node advertising the Anycast prefix-SID



Adjacency segments

- Local segment – Local significance
  - Local label, allocated from dynamic label pool
- Automatically allocated for each adjacency
  - Per adjacency: a protected and an unprotected adjacency-SID
    - >See SRTE presentation for more information
  - IS-IS: Different Adjacency-SID for L1 and L2 adjacencies between same neighbors
  - IS-IS: Different Adjacency-SID for IPv4 and IPv6 address-families
  - OSPF: Same Adjacency-SID in all areas of Multi-Area Adjacency (multiple adjacencies, each for a different area, over same interface)

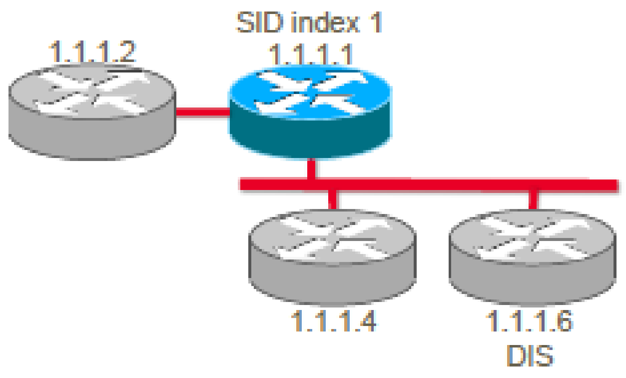
IS-IS Configuration – Example




```
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
  !
 !
<continue...>
```

- 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

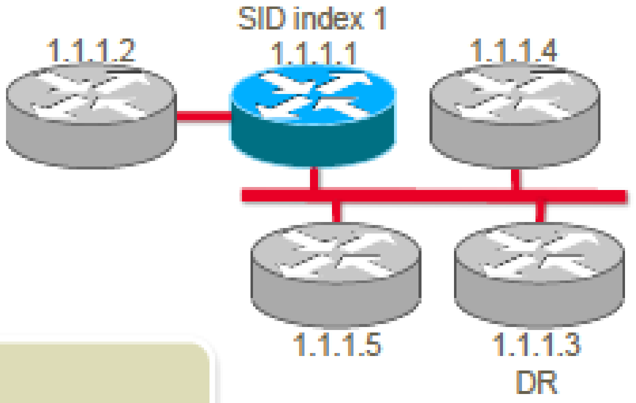
IS-IS Configuration – Example




```
<...continue>
 !
 interface TenGigE0/0/2/0
  point-to-point
  address-family ipv4 unicast
  !
  address-family ipv6 unicast
 !
 interface TenGigE0/0/3/0
  address-family ipv4 unicast
  !
  address-family ipv6 unicast
```

Adjacency-SIDs will automatically be allocated for all adjacencies

OSPF Configuration Example




```
router ospf 1
 router-id 1.1.1.1
 segment-routing mpls
 segment-routing forwarding mpls
 area 0
  interface Loopback0
   passive enable
   prefix-sid absolute 16001
  !
  interface GigabitEthernet0/0/0/0
   network point-to-point
  !
  interface GigabitEthernet0/0/0/1
  !
 !
!
```

Enable SR on all areas

Enable SR forwarding on all interfaces

Prefix-SID for loopback0

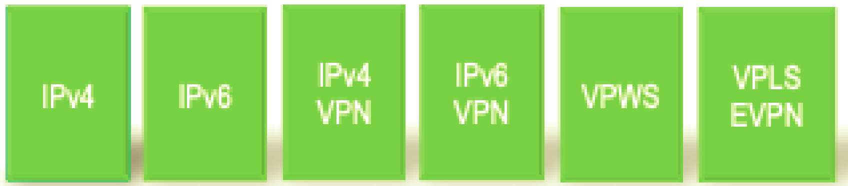
Adjacency-SIDs will automatically be allocated for adjacencies with SR forwarding enabled

Co-existence with Other MPLS Label Distribution Protocols

- The MPLS architecture permits concurrent usage of multiple label distribution protocols
  - LDP, RSVP-TE, BGP, static and SR control plane can co-exist without interaction
  - Easier to migrate from traditional services model to SDN
- Each node’s Label Manager
  - Reserves a label range (SRGB) for SR control-plane
  - Ensures that all dynamic labels are outside the SRGB block
  - Ensures that a dynamic label is uniquely allocated
- Each LSR must ensure that it can uniquely interpret its incoming labels
  - Adjacency segment: locally unique label allocated by the label manager
  - Prefix segment: operator ensures the unique allocation of each label within the allocated SRGB

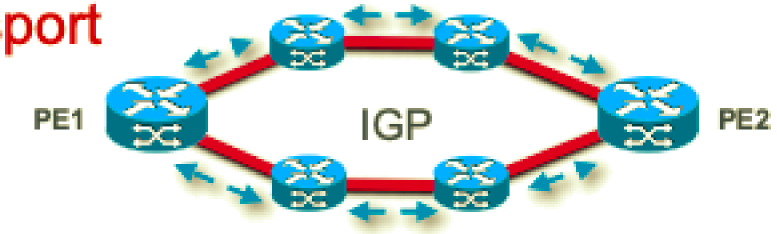
MPLS Control and Forwarding Operation with Segment Routing

Services



No changes to control or forwarding plane

Packet Transport



IGP label distribution for IPv4 and IPv6, same forwarding plane

Segment Routing Co-existence with LDP

Segment Routing – Co-existence with LDP

- Co-existence with LDP and other MPLS protocols
- Simple migration from LDP to Segment Routing

Co-existence with other MPLS label distribution protocols

- The MPLS architecture permits concurrent usage of multiple label distribution protocols
  - LDP, RSVP-TE, … and SR control plane can co-exist without interaction
- Each node's Label Manager
  - Reserves a label range (SRGB) for SR control-plane
  - Ensures that all dynamic labels are outside the SRGB block
  - Ensures that a dynamic label is uniquely allocated
- Each LSR must ensure that it can uniquely interpret its incoming labels
  - Adjacency segment: locally unique label allocated by the Label Manager
  - Prefix segment: operator ensures the unique allocation of each label within the allocated SRGB

MPLS-to-MPLS and MPLS-to-IP
label switching and label disposition

- For the MPLS2MPLS and MPLS2IP forwarding entries, SR and LDP can co-exist
  - These entries are indexed on a label
  - The local/incoming labels handled by LDP and SR (or other label distribution protocols) are unique
  - The outgoing label is only significant for the downstream neighbor, not for the local node
  - Multiple MPLS2MPLS and MPLS2IP entries can be programmed for the same prefix
    - cfr. LSP midpoint cross-connect

IP-to-MPLS – label imposition

- For IP2MPLS forwarding, LDP XOR SR entry can be inserted into FIB
  - Only one IP2MPLS entry can exists for each prefix path
- Default: LDP label imposition is preferred

```
router isis 1
 address-family ipv4|6 unicast
  segment-routing mpls sr-prefer
```

```
router ospf 1
 segment-routing mpls
 segment-routing sr-prefer
```

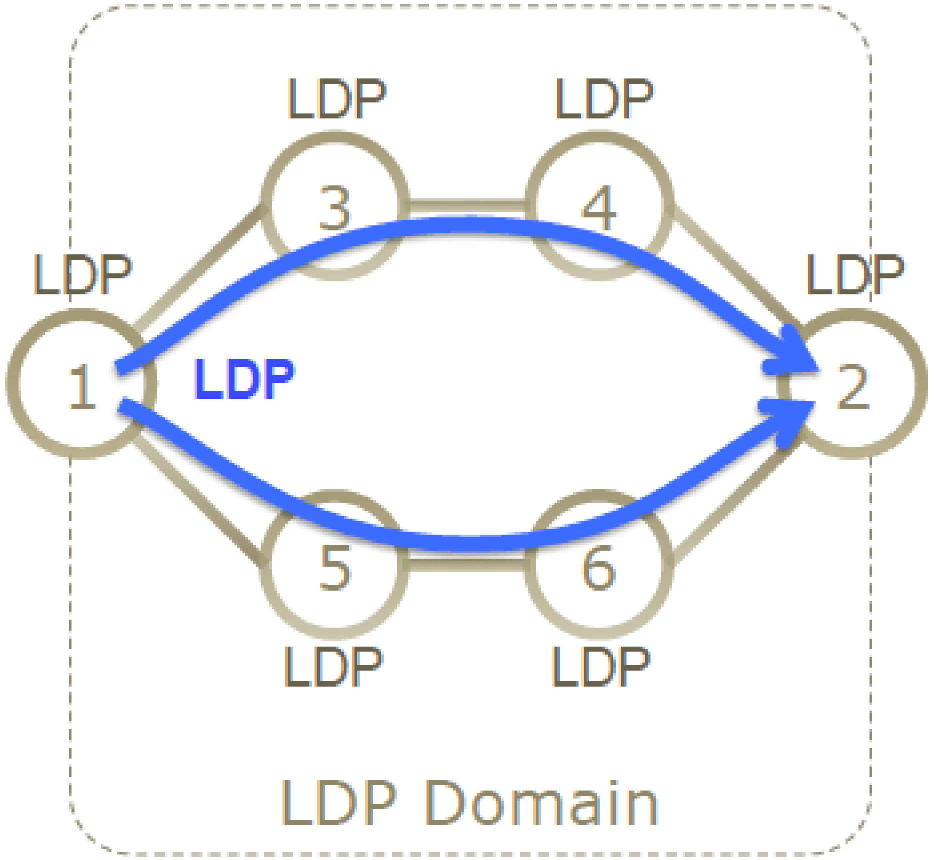
“Ships in the Night” Deployment Model

- LDP and SR are kept independent
  - continuous SR connectivity between SR PEs required
  - continuous LDP connectivity between LDP PEs required
  - no SR to LDP or LDP to SR interworking required
- Other deployment models are possible: see “SR/LDP interworking” section

Simplest migration LDP to SR

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

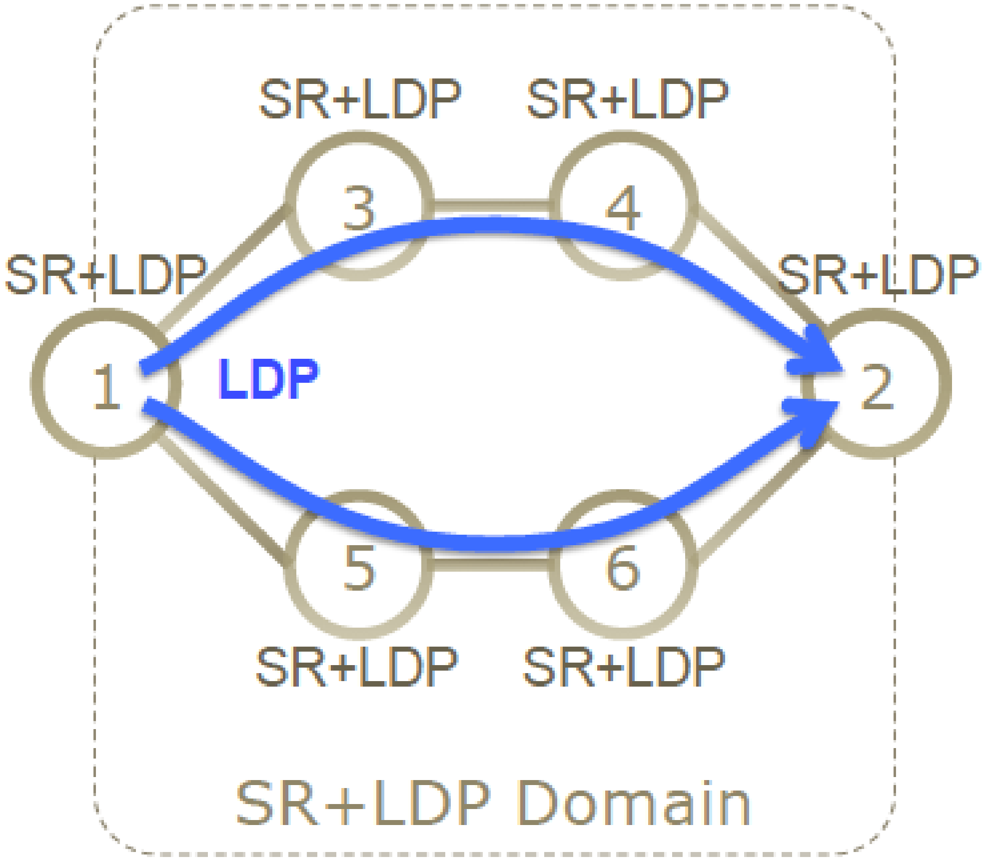
- Initial state: All nodes run LDP, not SR



Simplest migration LDP to SR

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

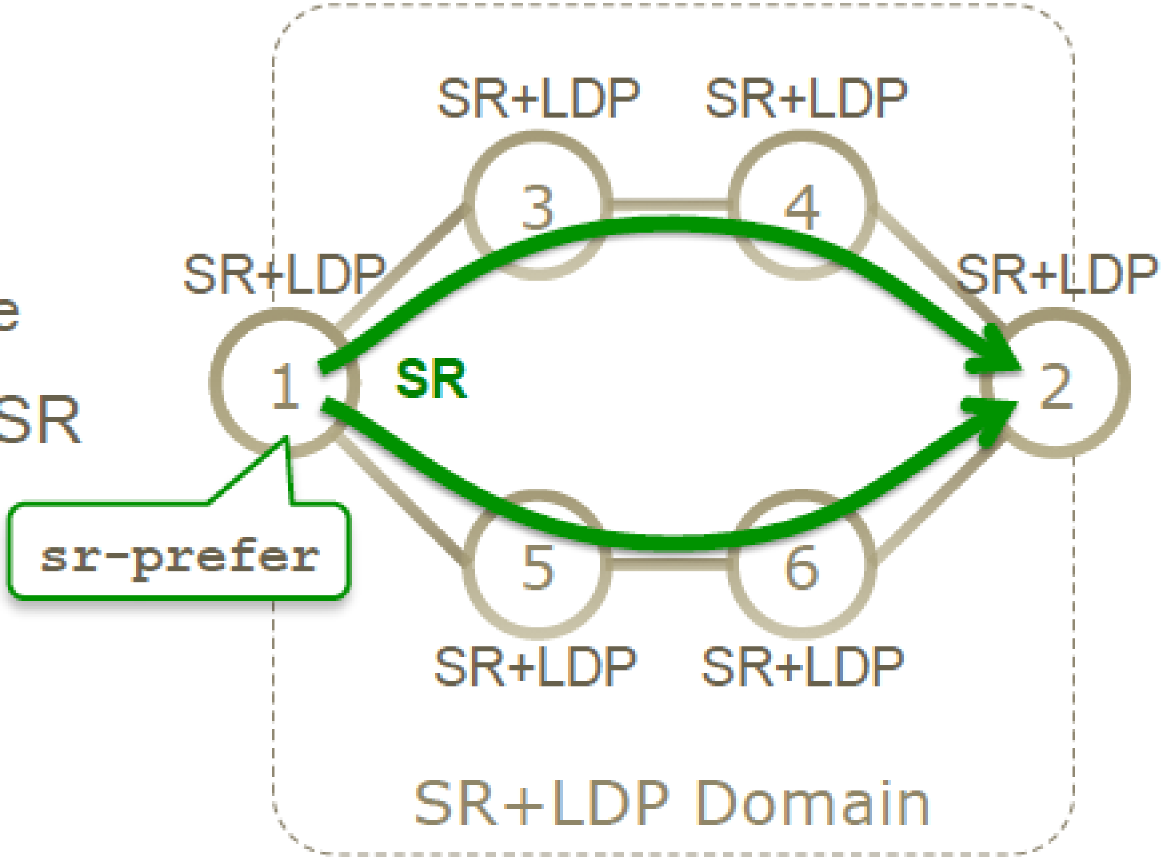
- Initial state: All nodes run LDP, not SR
- Step1: All nodes are upgraded to SR
  - In no particular order
  - leave default LDP label imposition preference



Simplest migration LDP to SR

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

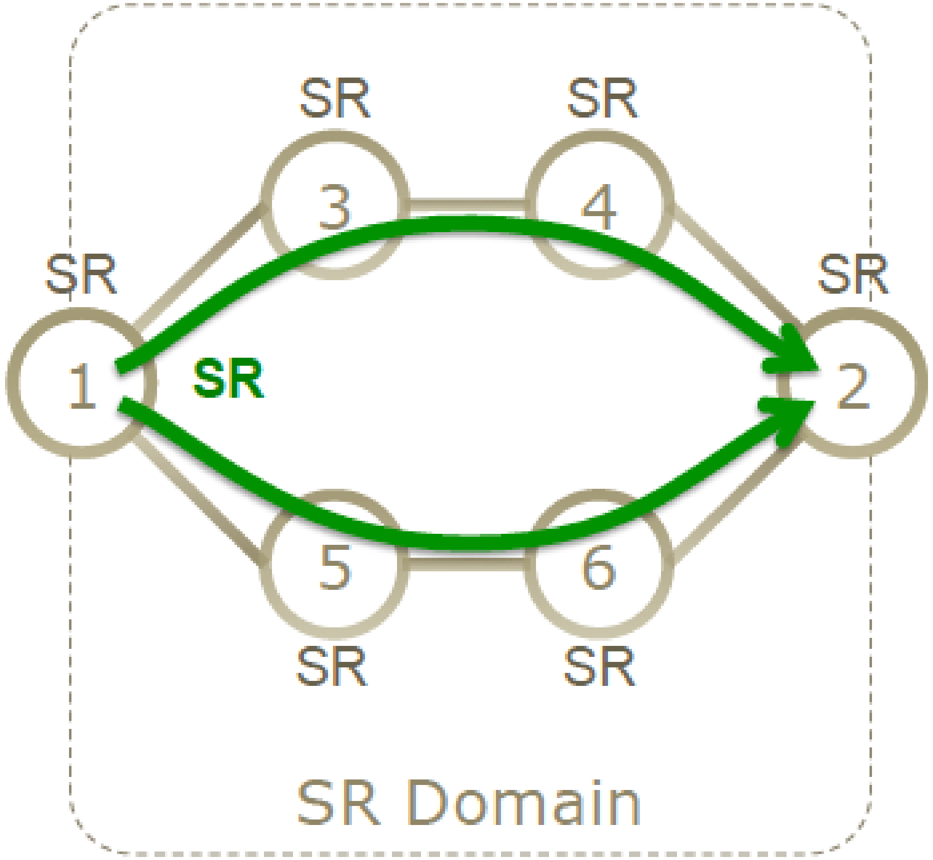
- Initial state: All nodes run LDP, not SR
- Step1: All nodes are upgraded to SR
  - In no particular order
  - leave default LDP label imposition preference
- Step2: All PEs are configured to prefer SR label imposition
  - In no particular order



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

Simplest migration LDP 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



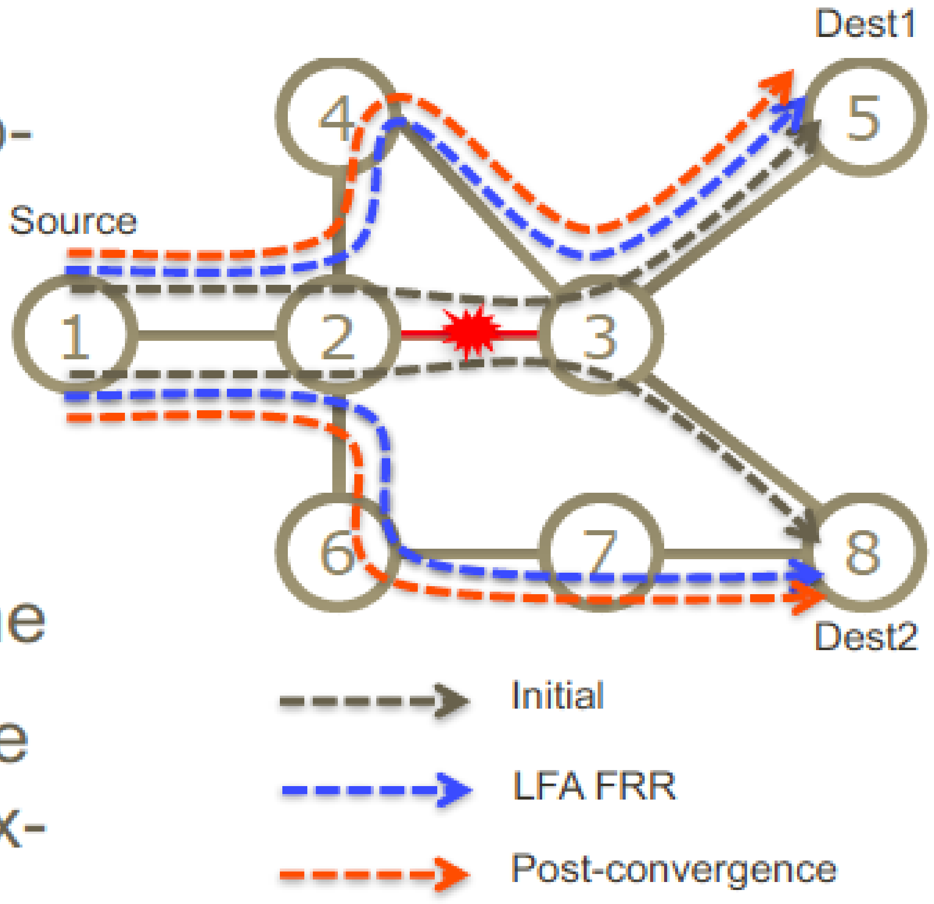
Segment Routing Topology Independent LFA (TI-LFA)

Topology Independent LFA (TI-LFA)

- Introduction to TI-LFA
- Simple, optimal and topology independent sub-50ms per-prefix protection
- Protects SR, LDP and IP traffic
- Examples of TI-LFA implementation

Classic Loop Free Alternate Fast ReRoute (LFA FRR)

- Per-prefix LFA: Simple, automatic, local, sub-50msec fast reroute technique
- IGP pre-computes a backup path per primary path per IGP destination: per-path IP optimality
- The backup path is pre-installed in data plane
- Upon local failure, all the backup paths of the impacted destinations are enabled in a prefix-independent manner (<50msec loss of connectivity)



Classic Per-Prefix LFA – disadvantages

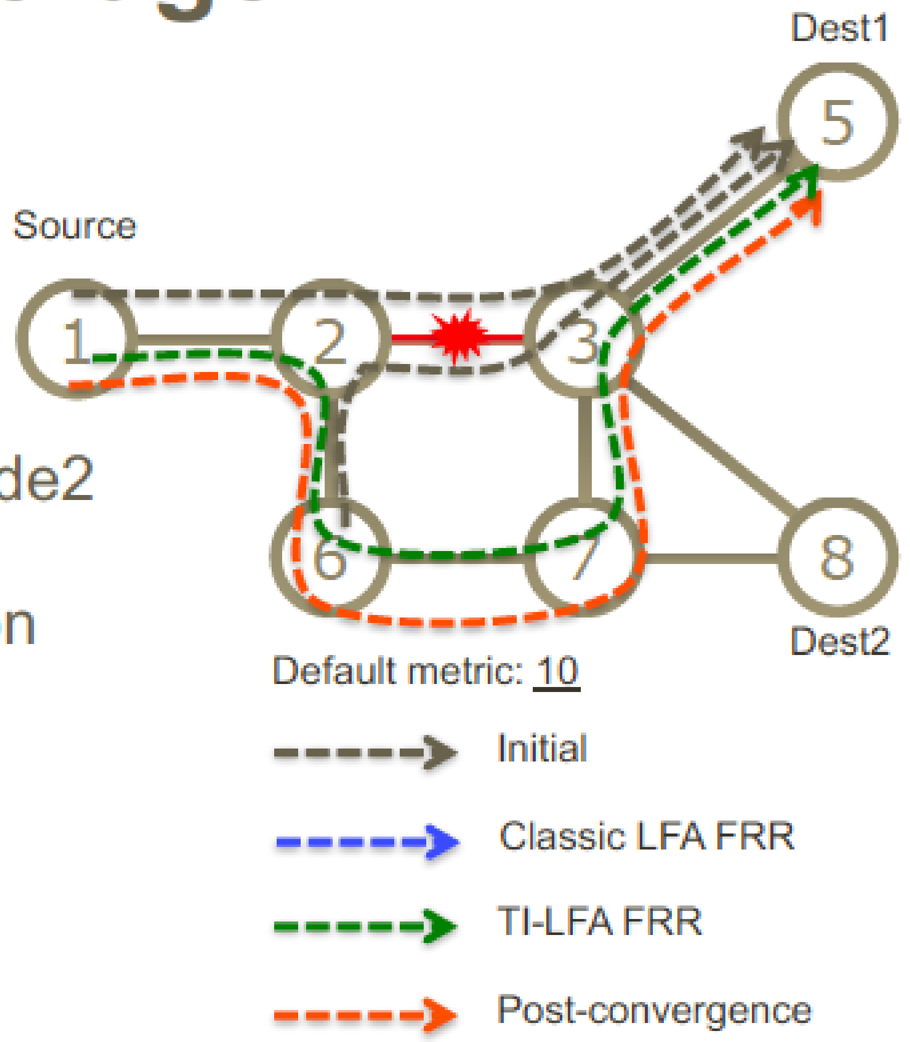
- Classic LFA has disadvantages:
  - Incomplete coverage, topology dependent
  - Not always providing most optimal backup path

→ Topology Independent LFA (TI-LFA) solves these issues

Classic LFA has partial coverage

- Classic LFA is topology dependent: not all topologies provide LFA for all destinations
  - Depends on network topology and metrics
  - E.g. Node6 is not an LFA for Dest1 (Node5) on Node2, packets would loop since Node6 uses Node2 to reach Dest1 (Node5)
    → Node2 does not have an LFA for this destination (no - - -> backup path in topology)

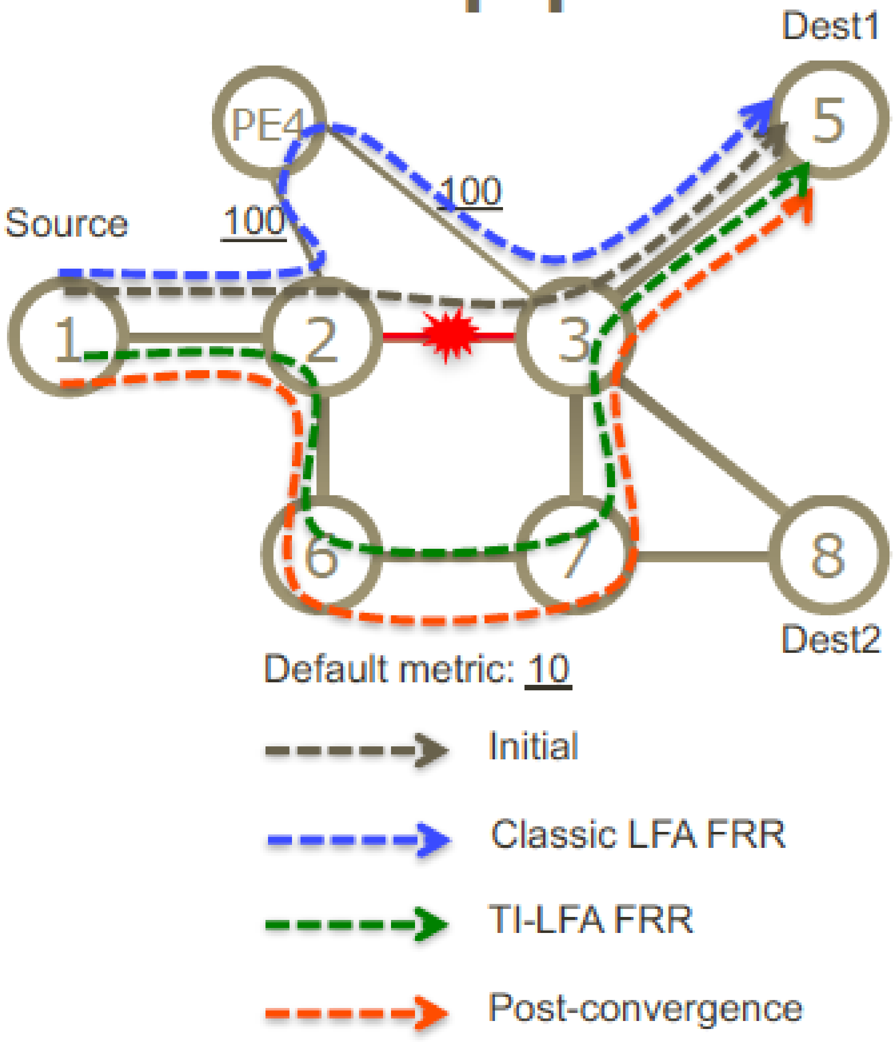
→ Topology Independent LFA (TI-LFA) provides 100% coverage



Classic LFA may provide suboptimal backup path

- Classic LFA may provide a suboptimal FRR backup path:
  - This backup path may not be planned for capacity, e.g. P node 2 would use PE4 to protect a core link, while a common planning rule is to avoid using Edge nodes for transit traffic
  - Additional case specific LFA configuration would be needed to avoid selecting undesired backup paths
  - Operator would prefer to use the post-convergence path as FRR backup path, aligned with the regular IGP convergence

→ TI-LFA uses the post-convergence path as FRR backup path

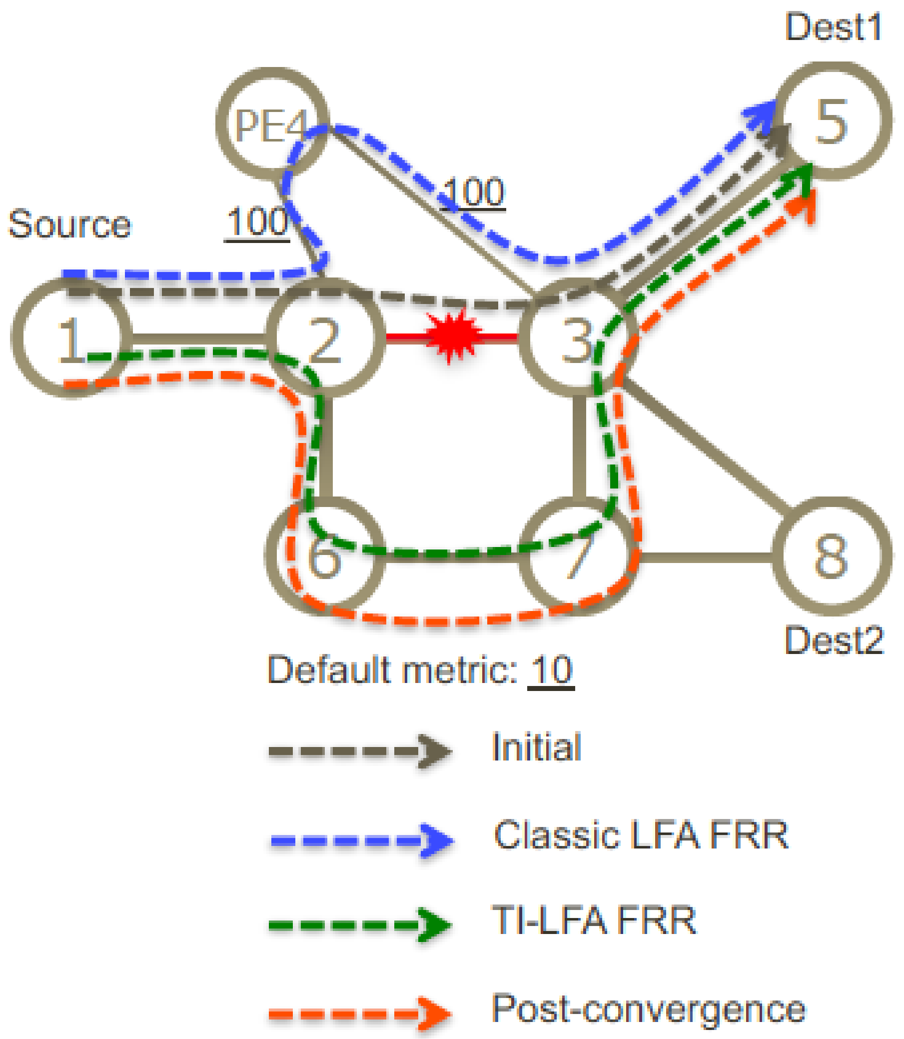


Topology Independent LFA (TI-LFA) – Benefits

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

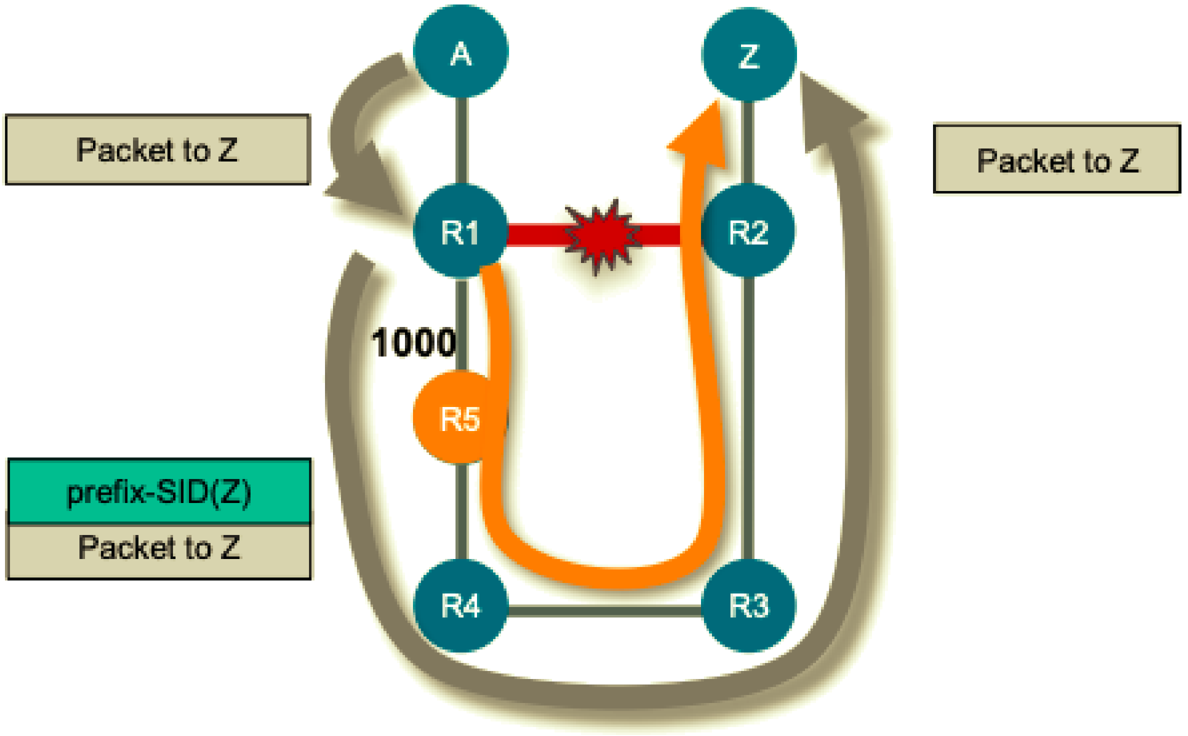
TI-LFA uses Post-Convergence Path
Optimality Benefit Example

- Protecting destination Node5 on Node2 against failure of link 2-3
- Classic LFA: Node2 switches all traffic destined to Node5 towards the edge node PE4
  - → Low BW (high metric) links and an edge node are used to protect the failure of a core link
  - → A common planning rule is to avoid Edge nodes for transit traffic
  - → Classic LFA does not respect this rule ✗
- TI-LFA: Node2 switches all traffic destined to Node5 via high BW core links: OK! ✓



TI-LFA – Zero-Segment Example

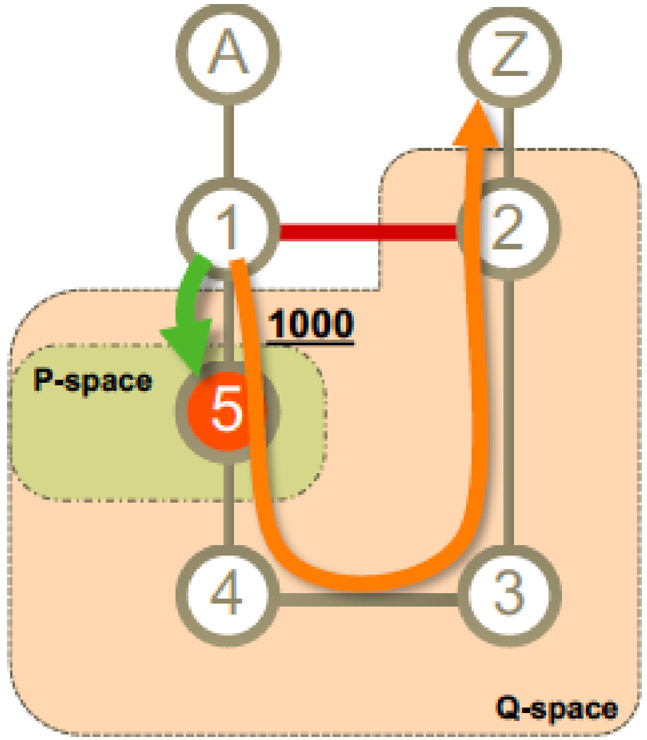
- TI-LFA for link R1R2 on R1
- Calculate LFA(s)
- Calculate post-convergence SPT
- Find LFA on post-convergence SPT
- R1 will steer the traffic towards LFA R5



Default metric:10

TI-LFA – zero-segment example

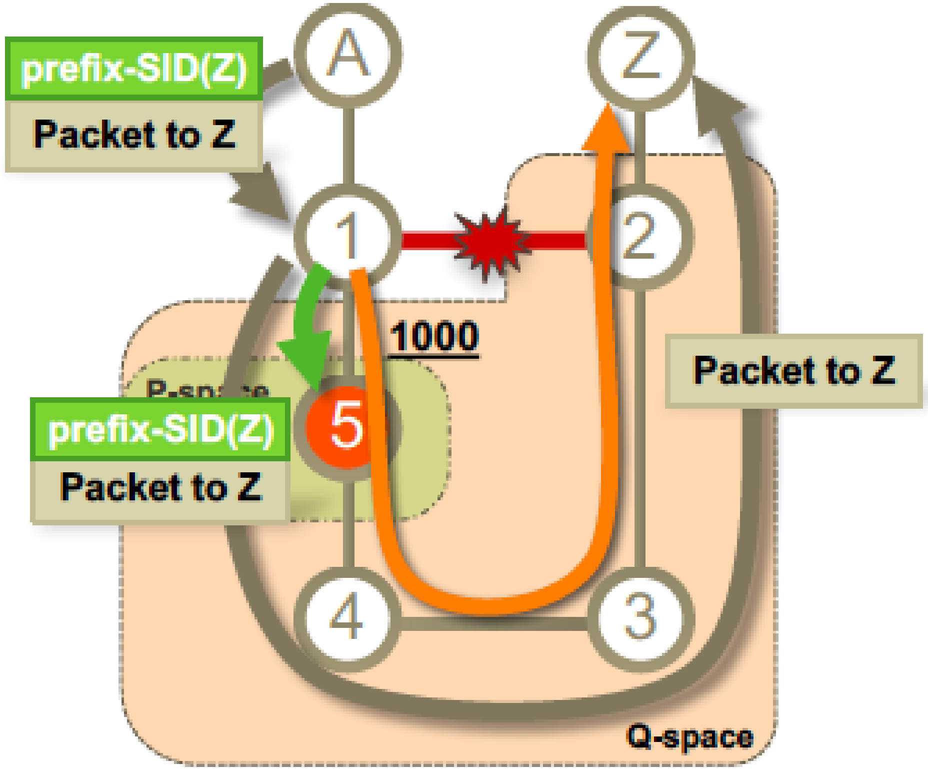
- For the destination Z, for the router R1, the primary link is R1R2. R1's TI-LFA computation for Z is:
  - Remove the primary link for Z (R1R2) and compute the SPF on the resulting topology. This gives us the post-convergence path from R1 to Z: <R5, R4, R3, R2>
  - R5 is in the P space (R1 can send a packet destined to R5 without any risk of having that packet flow back through the protected link R1R2)
  - R5 is in the Q space (R5 can send a packet to R2 without any risk of having this packet flow back through the protected link R1R2)
  - R5 is along the post-convergence path
  - Hence the TI-LFA backup computed by R1 for destination Z is "forward the packet to R5 without any additional segment"
- Note that this behavior is applied on a per-prefix basis and hence that for each prefix the primary link changes and the post-convergence path is computed accordingly together with the P and Q properties. The algorithm is proprietary (local behavior which is not in the scope of IETF standardization) and scales extremely well.



Default metric: 10

TI-LFA – zero-segment example

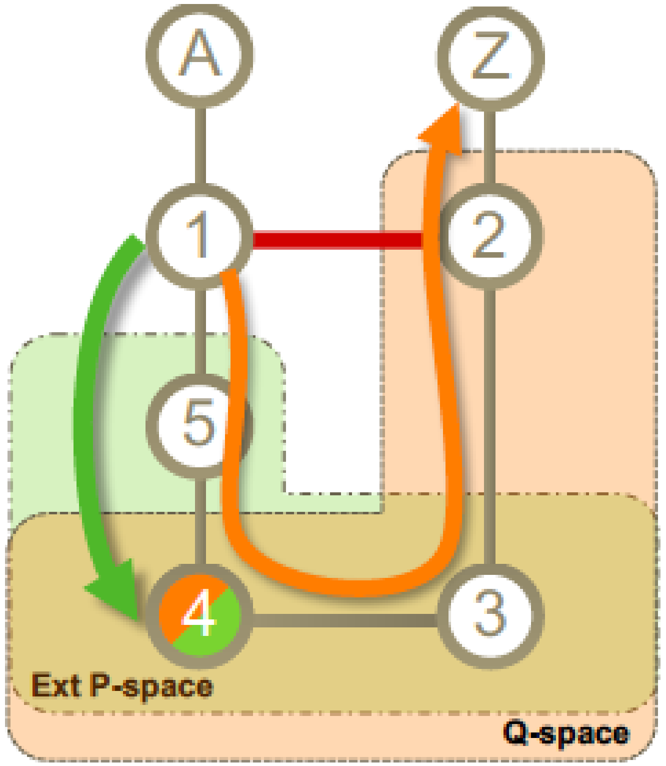
- To steer packets on the TI-LFA backup path:
  "forward the packet to R5 without any additional segment"



Default metric: 10

TI-LFA – single-segment example

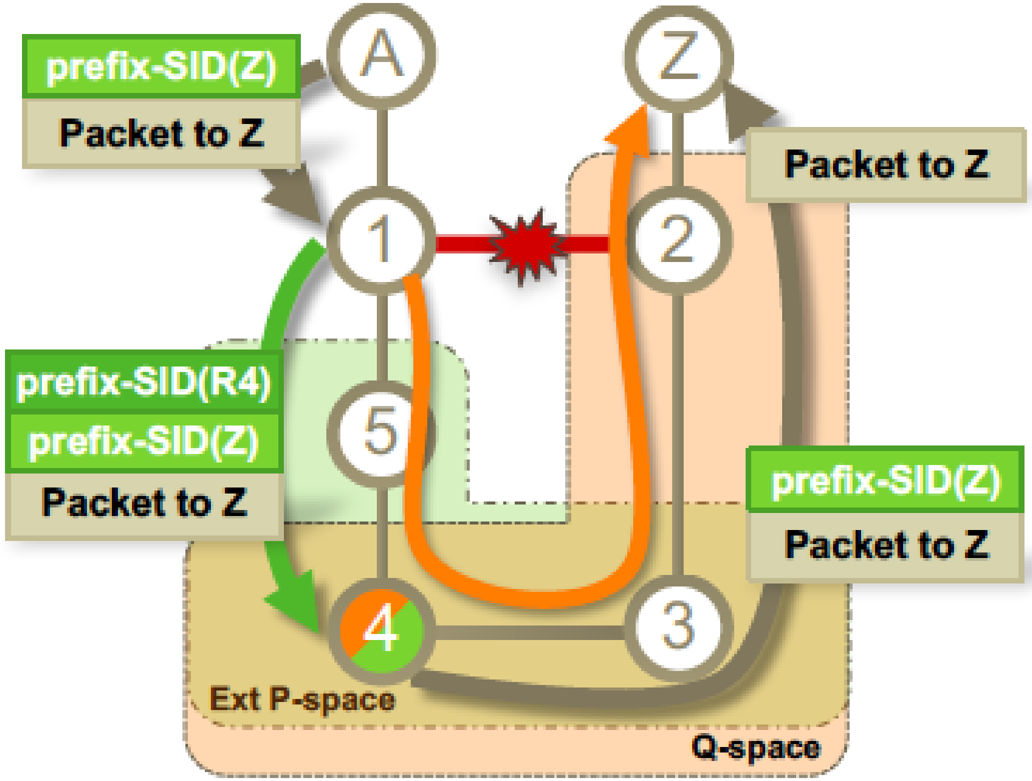
- For the destination Z, for the router R1, the primary link is R1R2. R1's TILFA computation for Z is:
  - Remove the primary link for Z (R1R2) and compute the SPF on the resulting topology. This gives us the post-convergence path from R1 to Z: <R5, R4, R3, R2>
  - R4 is in the P space (R1 can send a packet destined to R4 without any risk of having that packet flow back through the protected link R1R2)
  - R4 is in the Q space (R4 can send a packet to R2 without any risk of having this packet flow back through the protected link R1R2)
  - R4 is along the post-convergence path
  - Hence the TILFA backup computed by R1 for destination Z is "forward the packet on interface to R5 and push the segment R4"
- Note that this behavior is applied on a per-prefix basis and hence that for each prefix the primary link changes and the post-convergence path is computed accordingly together with the P and Q properties. The algorithm is proprietary (local behavior which is not in the scope of IETF standardization) and scales extremely well



Default metric:10

TI-LFA – single-segment example

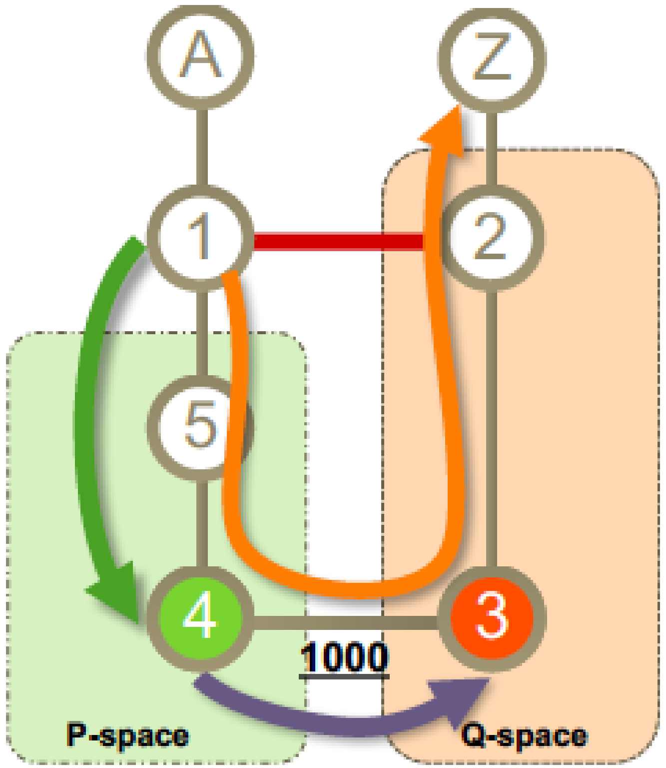
- To steer packets on the TI-LFA backup path:
  “forward the packet on interface to R5 and push the segment <prefix-SID(R4)>”



Default metric:10

TI-LFA – double-segment example

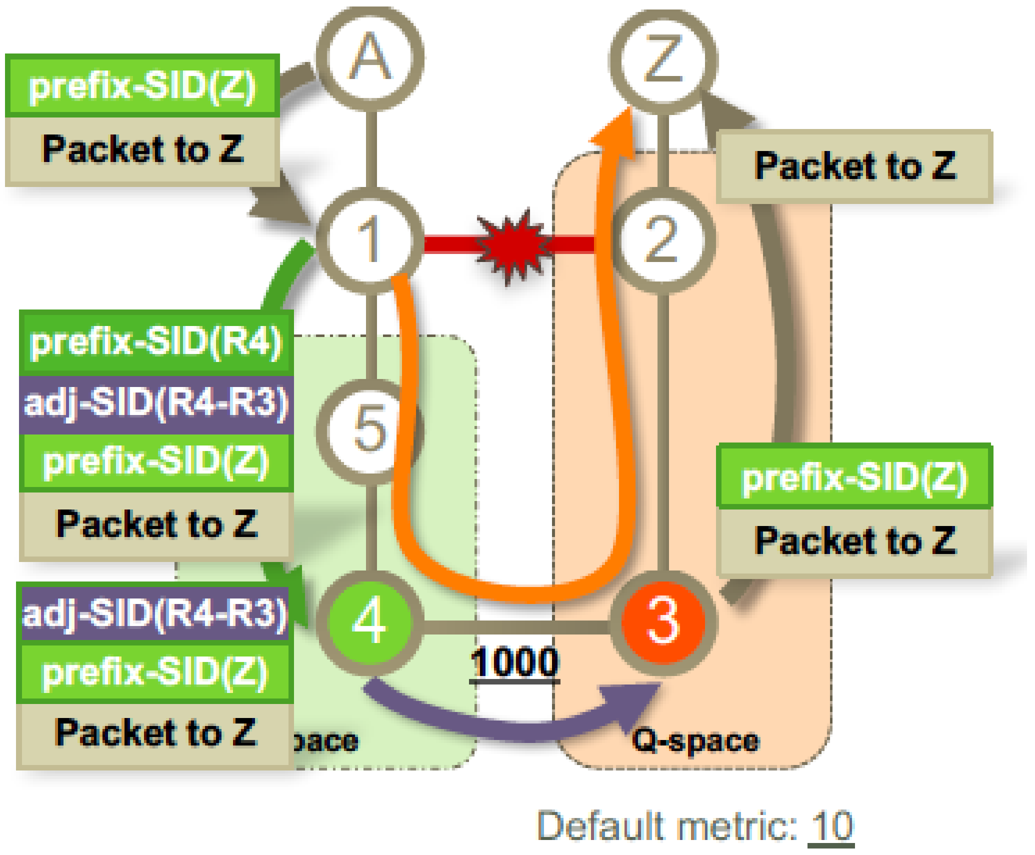
- For the destination Z, for the router R1, the primary link is R1R2. R1's TI-LFA computation for Z is:
  - Remove the primary link for Z (R1R2) and compute the SPF on the resulting topology. This gives us the post-convergence path from R1 to Z: <R5, R4, R3, R2>
  - R4 is in the P space (R1 can send a packet destined to R4 without any risk of having that packet flow back through the protected link R1R2)
  - R3 is in the Q space (R3 can send a packet to R2 without any risk of having this packet flow back through the protected link R1R2)
  - R4 and R3 are adjacent and along the post-convergence path
  - Hence the TI-LFA backup computed by R1 for destination Z is "forward the packet on interface to R5 and push the segments R4 and R4-R3"
- Note that this behavior is applied on a per-prefix basis and hence that for each prefix the primary link changes and the post-convergence path is computed accordingly together with the P and Q properties. The algorithm is proprietary (local behavior which is not in the scope of IETF standardization) and scales extremely well



Default metric: 10

TI-LFA – double-segment example

- To steer packets on the TI-LFA backup path:
  "forward the packet on interface to R5 and push the segments <prefix-SID(R4) and adj-SID(R4-R3)>"



Configuring Topology Independent Fast Reroute for IPv4 using Segment Routing and IS-IS (Cisco IOS XR)

```
router isis DEFAULT
 net 49.0001.1720.1625.5001.00
 address-family ipv4 unicast
  metric-style wide
  segment-routing mpls
 !
 interface Loopback0
  passive
  address-family ipv4 unicast
   prefix-sid absolute 16041
  !
 !
 interface GigabitEthernet0/0/0/0
  address-family ipv4 unicast
   fast-reroute per-prefix
   fast-reroute per-prefix ti-lfa
  !
 !
!
```

Enable TI-LFA for IPv4 prefixes on interface GigabitEthernet0/0/0/0

Configuring Topology Independent Fast Reroute for IPv6 using Segment Routing and IS-IS (Cisco IOS XR)

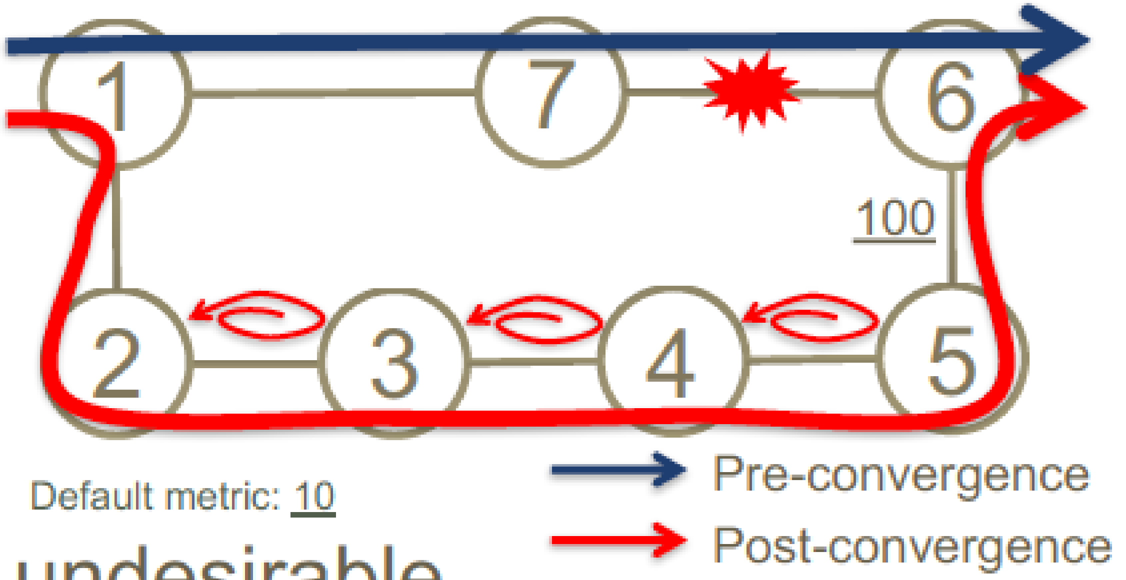
```
router isis DEFAULT
 net 49.0001.1720.1625.5001.00
 address-family ipv6 unicast
  metric-style wide
  segment-routing mpls
 !
 interface Loopback0
  passive
  address-family ipv6 unicast
   prefix-sid absolute 16061
  !
 !
 interface GigabitEthernet0/0/0/0
  address-family ipv6 unicast
   fast-reroute per-prefix
   fast-reroute per-prefix ti-lfa
  !
 !
!
```

Enable TI-LFA for IPv6 prefixes on interface GigabitEthernet0/0/0/0

Microloop avoidance

What is a microloop?

- Microloops are a day-one IP drawback
- IP hop-by-hop routing may induce microloop at any topology transition
  - Link up/down, metric up/down
- E.g. Microloops can occur after failure of link 6-7
- Microloops can increase packet loss, which is especially undesirable when FRR is used.



SR microloop avoidance

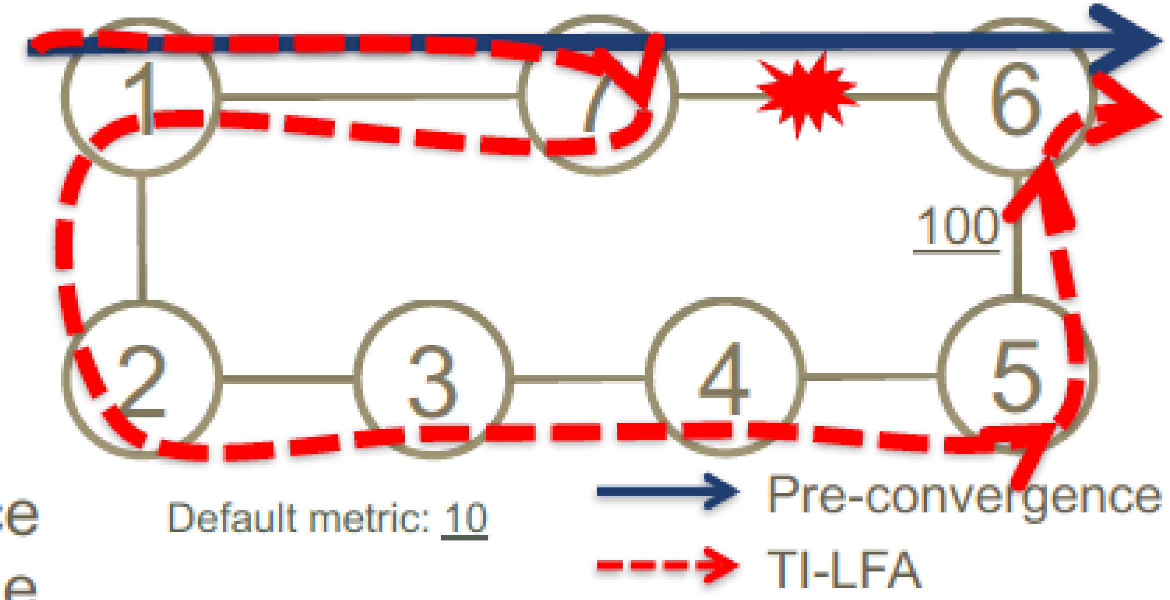
- Prevent any microloop upon an isolated convergence due to
  - link up/down event
  - metric increase/decrease event
- If multiple back-to-back convergences, fall back to native IP convergence
- Configuration:

```
router isis 1
 address-family ipv4 unicast
  microloop avoidance segment-routing
```

```
router ospf 1
 microloop avoidance segment-routing
```

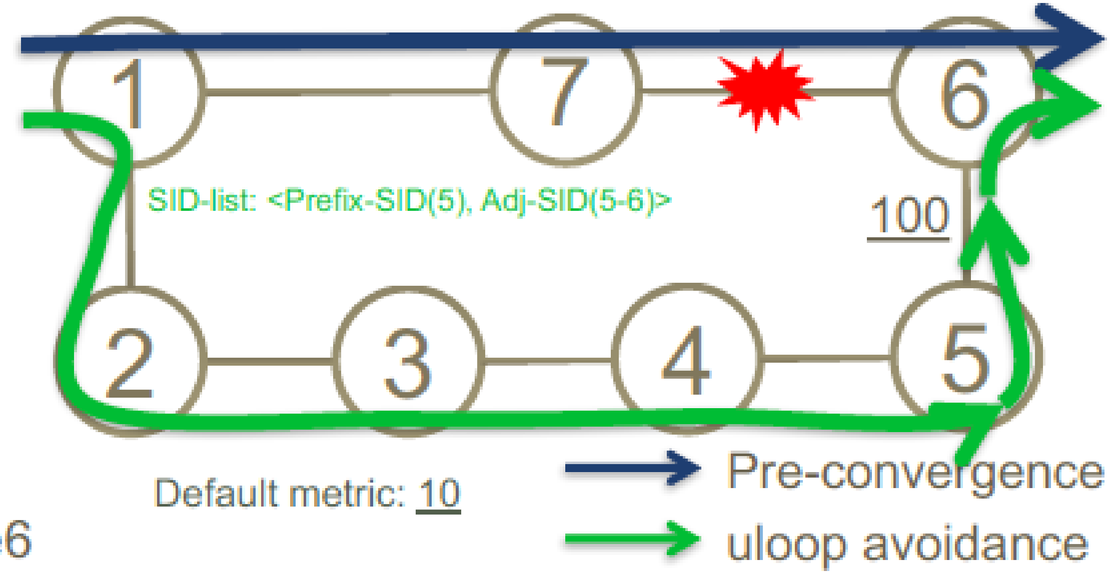
SR microloop avoidance – workflow

- TI-LFA protection kicks in on Node7, repairing the traffic to Node6 via Node5 and link Node5-Node6
- All nodes are notified of the topology change due to the failure
- E.g. Node1 computes the post-convergence SPT and detects possible microloops on the post-convergence paths for any destination, such as Node6
- If microloops are possible on the post-convergence path for a destination, then a SID-list is constructed to steer the traffic to that destination loop-free over the post-convergence path; in this example: <Prefix-SID(5), Adj-SID(5-6)> for destination Node6



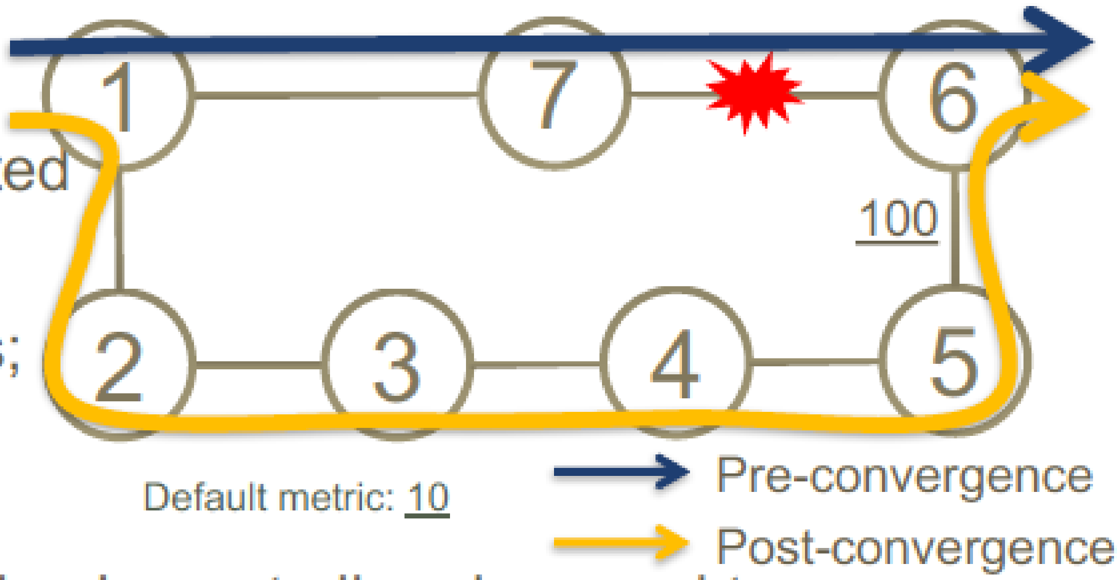
SR microloop avoidance – workflow

- IGP on Node1 updates the forwarding table and installs the SID-list imposition entries for those destinations with possible microloops, such as destination Node6
  - Node1 imposes SID-list <Prefix-SID(5), Adj-SID(5-6)> on packets to Node6



- All nodes converge and update their forwarding tables, using SID-lists where needed

SR microloop avoidance – workflow



- Some time later, the new topology is applied and no more microloops are expected
- IGP updates the forwarding table, removing the microloop avoidance SID-lists; traffic now natively follows the post-convergence path
- Note: SR microloop avoidance is a local behavior, not all nodes need to implement it to get the benefits
- There is incremental benefit for each node that has it implemented
  - E.g. if only Node1 has SR microloop avoidance, then e.g. traffic entering Node2 (not from Node1) to Node6 would still see microloops
  - When enabling SR microloop avoidance on Node2, then e.g. traffic entering Node3 (not from Node2) to Node6 would still see microloops, etc.

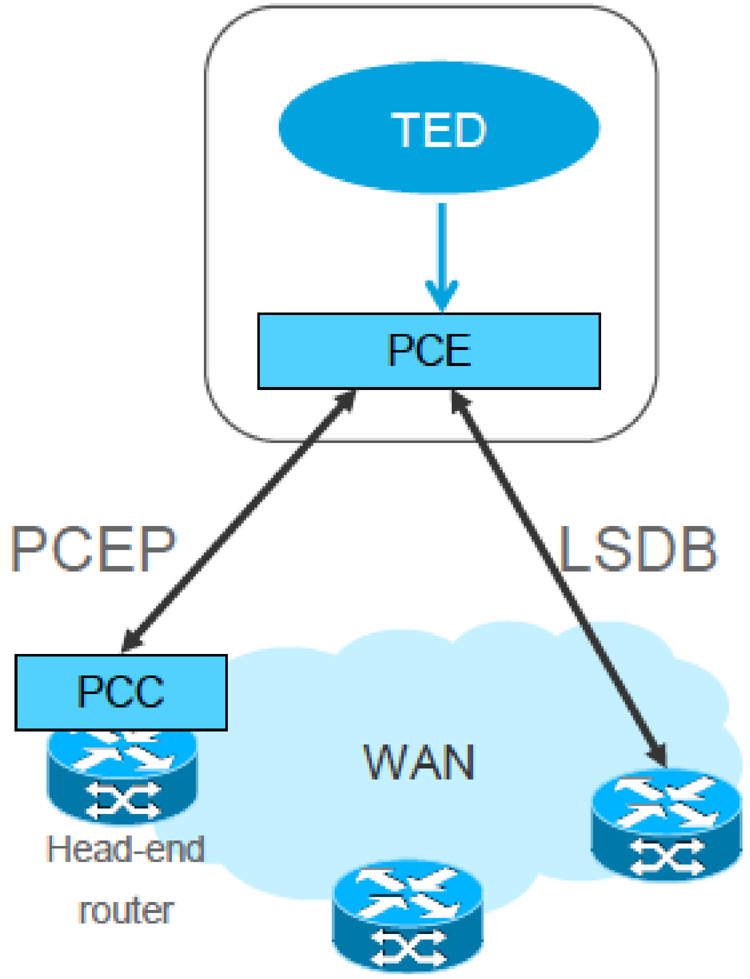
SDN Integration

Path Computation Element

- “.. entity that is capable of computing a network path or route based on a network graph, and of applying computational constraints during the computation..” (RFC4655)
- Main function is to compute paths
- Generally resides on a server platform, but can be on a router as well
- Computed Path might be:
  - Explicit route identifying a contiguous set of strict hops(adjacency SIDs) between the source and destination
  - Combination of strict/loose hops(adjacency and node SIDs) between the source and destination

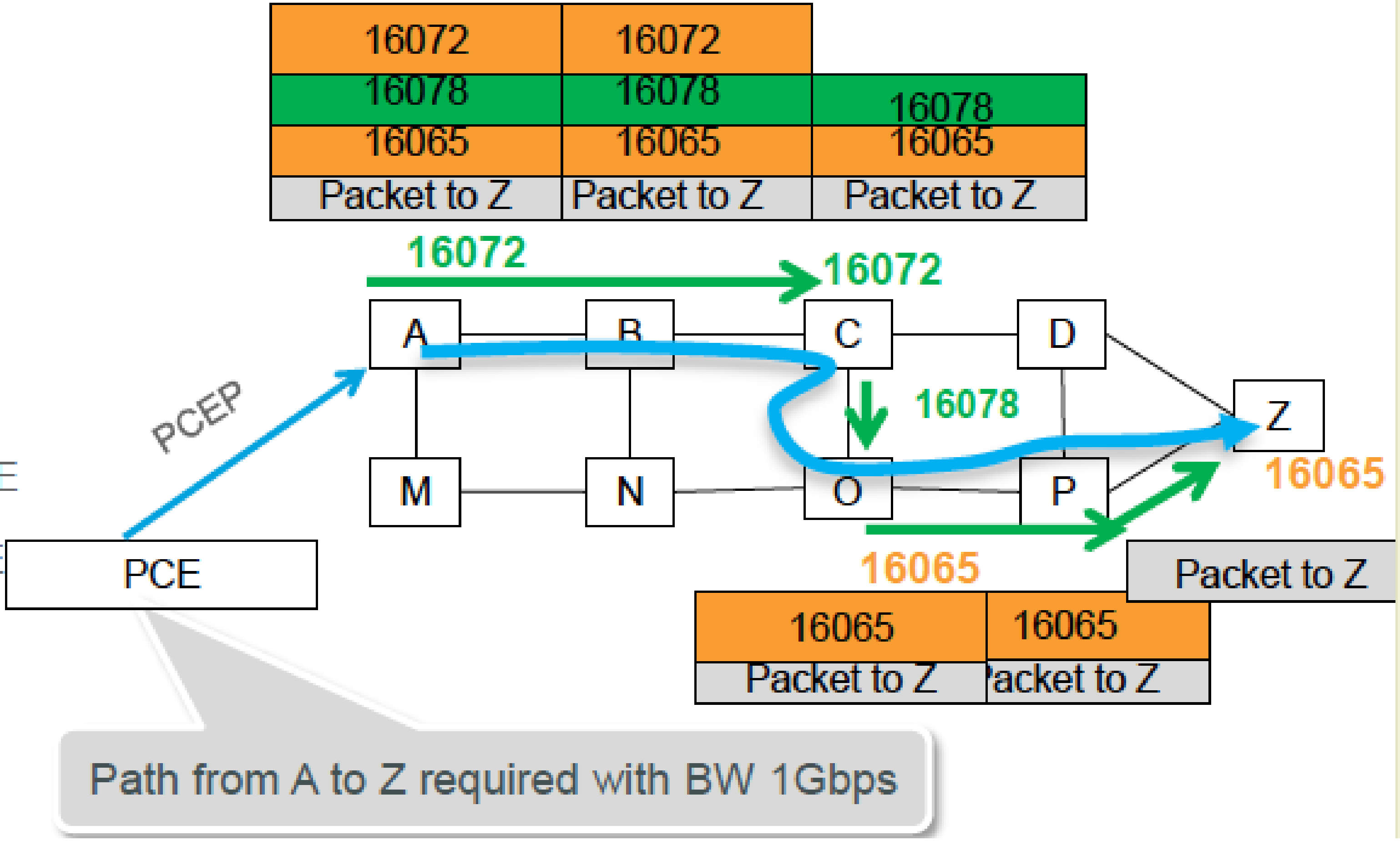
PCE Definitions

- Traffic Engineering Database (TED)
  - Contains topology and resource information
  - Inputs are IGP LSDB and by other means
- PCE Server (PCS)
- Path Computation Client (PCC)
  - Agent on router(s) that interact with PCE Server
- PCE Protocol (PCEP)
  - Protocol that runs between PCC on router and PCE server



SR with SDN WAN Orchestration

- Segment Routing – Traffic Engineering (SR-TE) paths may not follow IGP-computed shortest path
  - PCE the key enabler for SR-TE
- PCE and SR-TE functions:
  - Global visibility and optimisation
  - Single touch-point at the Source Node (ingress router)
  - Set-up time is O(PCEP session latency) – NO RSVP-TE signalling latency
  - Abstracts away details of underlying SR-TE network – applications developed over IP/MPLS TE network will work over SR-TE networks
- draft-sivabalan-pce-segment-routing



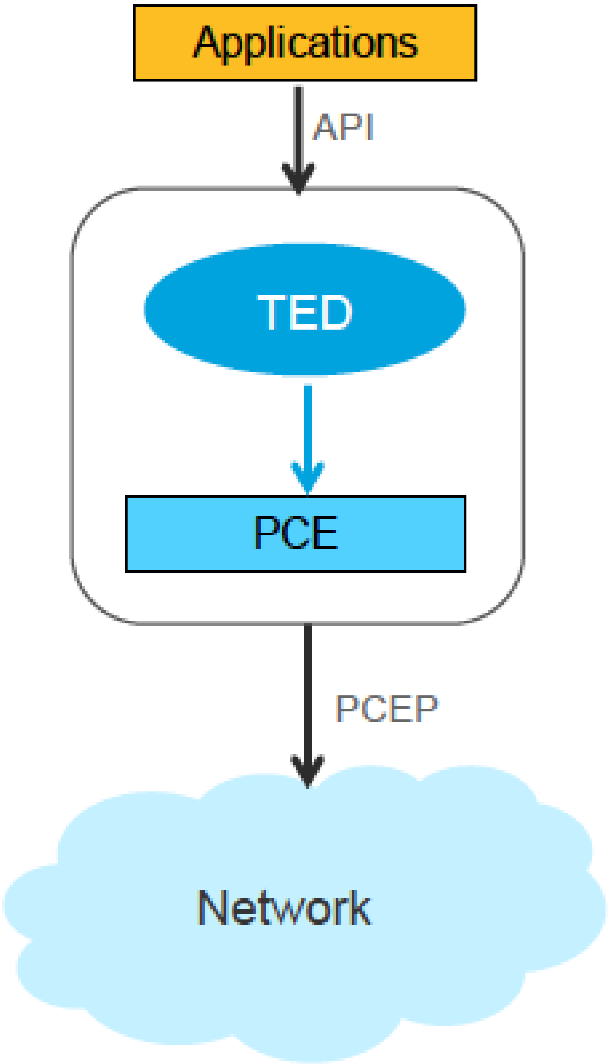
The Benefits of Centralised TE

Centralised Traffic Engineering

- Better optimum
- Better predictability
- Faster convergence
- Better suited for Application Programmability (Nbound-API)
- Network Programmability (Sbound-API, PCEP)

Centralised TE with Segment Routing

- Controller expresses path as segment list
- Network maintains segments and provide FRR for them
- ECMP-awareness
- No signalling and per-flow state at midpoint



Summary

- Simple routing extensions to implement source routing
- Packet path determined by prepended segment identifiers (one or more)
- Data plane agnostic (MPLS, IPv6)
- Network scalability and agility by reducing network state and simplifying control plane
- Traffic protection with 100% coverage with more optimal routing
- Interworking capabilities with LDP-only devices
- SDN ready topologies and easy to migrate

OPTIMITY

Customers' satisfaction and trust is our most valuable ASSET