Segment Routing Proof of Concept for Business Services

Does it work for us

A Technical Paper prepared for SCTE•ISBE by

Elaine Yeo
Principal Engineer
Charter Communications
14810 Grasslands Drive | Englewood, CO 80112-7138
720.536.1357
Elaine.Yeo@charter.com

Table of Contents

Title **Page Number**

List of Figures

Title **Page Number**

List of Tables

Introduction

Segment Routing is a source based routing methodology that uses a list of unique segment IDs stacked in an order of arrival along a traffic path. It leverages the existing MPLS data plane by encoding a segment ID in each MPLS label, thus creating a stack of labels in the packet header which instructs each node along the path to execute the information within the labels upon receipt and forwards it along to the next node. This technology can be used to simplify and optimize the network, meeting key performance objectives such as latency and at the same time enhance operational efficiencies around capacity reporting, change control management via automation with the use of a controller. The interest lies in whether can Segment Routing be beneficial to business services, how does this simplify and affect a large service provider consisting of multiple network types, various vendor platforms and software. This technology has been chosen as a proof of concept that focuses on interoperability of existing multiple vendors within a service provider network, working together with a controller to maximize the benefits of segment routing, particularly the traffic engineering aspect. During the proof of concept, some observations were made and several key points were brought to attention where cascading effects from a system event were seen to have a potential impact that affects CAPEX, OPEX, and architectural processes. These findings from the proof of concept also provides information that can influence business decisions to move forward with Segment Routing or retain the existing mechanisms used today in the network.

Problem Statement

Charter was initially presented the opportunity to provide a solution to meet latency requirements per contractually agreed Service Level Agreement (SLA) for Cell Tower Backhaul (CTBH) services under a new architecture design. Existing services that had originally met the necessary SLAs must continue to meet the requirements if an architecture were to change to a Hub and Spoke design. The new design would cause traffic to double back to the direction it had traversed, thus increasing the latency to its final destination. For example, with a Hub and Spoke architecture, a cell tower's traffic will have to traverse east to a hub location before doubling back past the source over to the west where the end destination of the Mobile Switching Center (MSC) resides.

RSVP-TE is the common choice, given the history of successful deployment and the ability to provide fast-reroute. However, with the modern IP networks we have today, the need to keep up with the growing demands of network capacity and service quality makes it difficult to scale without compromising network resources to support traffic engineering and other pertinent applications. Segment Routing (SR), a fairly new mechanism simplifies the need of separate protocols such as IGP, LDP and RSVP-TE interacting in a single network and alleviates network resources to hold network state within the core. It serves to remove state from the core network and keeping it in the packet and the ingress node. These features offered by Segment Routing peaks the interest of a potential alternative to RSVP-TE. Segment Routing has the prospect to keep things simple and making room to scale for future enhancements.

In addition for the need to counter latency with a simple approach, other benefits such as using Topology Independent Loop Free Alternate TI-LFA with SR provides for 100% coverage of the network, making it possible to compute, instantiate traffic engineering paths and restore traffic optimally. This is best accomplished when coupled with a controller to provide path computation for optimal routes during a network failure with 100% visibility of the network. From a trending perspective, the lack of network visibility makes it difficult to determine usage patterns and flow characteristics. Without these trending information, it is challenging to learn and evolve our network while planing for growth. Operationally, route stability has been a manual process, potentially introducing human error, i.e configuration errors and delay reaction to act quickly to re-route traffic to an optimal path. Historically, service impacting outages are triggered by fiber cuts resulting in secondary root causing of sub-optimal back-up failover to latent or congested path in the absence of traffic protection and global path reoptimization.

To provide options of either RSVP-TE or SR-TE, a proof of concept was first pursued for Segment Routing alone to see if it would fit Charter's business model and simplify operational process. The proof of concept lab was built based on an example Charter market. To ensure that Segment Routing would fit into Charter's networks, the proof of concept was vetted against 3 existing major vendor platforms deployed across all legacy companies. This document outlines the differences and interoperability of the multiple platforms in addition to the functionality of Segment Routing and its effect on Charter.

1. Segment Routing

There are two types of SR technologies, SR over MPLS (SR-MPLS) and SR over IPv6 (SRv6). This document will focus on SR-MPLS. SR can be implemented over the existing MPLS architecture.

1.1. Concept

Segment Routing is a source based paradigm that uses an ordered list of segments appended to the packet header. These lists of segments serve as an abstraction instruction sets for the source node to process and execute on the path to take. A segment is encoded as an MPLS label and a list of segments are essentially a stack of labels. The segments are processed from top to bottom. With the list of segments holding the instructions on traffic path, the state is no longer held in the network but rather within the packet. Only the source node is required to compute and encode the instruction list, while the transit nodes simply reads the top most label before passing it along.

1.2. Operations

SR's top most segment is known as the Active Segment. The active segment is the segment that is processed by the receiving node. Segment Routing's segment list operations uses the same existing MPLS forwarding method. The table below details the correlation of the operation in a row within SR to the same row within MPLS label operations.

Table 1 – SR vs MPLS Label Operations

SR Segment List Operations	MPLS Label Stack Operations
PUSH Inserts an active segment over the list of segments pushes the label stack forward.	**PUSH** Injects a label over the label stack
CONTINUE The active segment is not completed, remains active and continues to next destination	**SWAP** Replace the top label with a new label
NEXT The active segment is completed. The next segment on the list is the active segment	**POP** Removes top label from label stack

1.3. Segment Types

There are many segment types in SR for specific functions. This document will use terms listed below that are relevant to only what was used in the SR Proof of Concept (POC) .

Segment Routing Global Block (SRGB)
Globally unique range of labels recognized within a node. Configured within IGP. The SRGB can be configured to the desired range. When the SRGB is set, a range of labels are set aside to be used, unique to all nodes within the SR domain. The beginning number of the range is known as the SRGB base.

Example:
Manually configured SRGB = 16000 – 19000. Therefore 16000 is the SRGB base.

SID
Segment Identifier (SID). The SID is encoded as an MPLS label that identifies the segment and sets it apart from other nodes or links.

Node SID
Allocated from the pool of SRGB. Globally significant and unique in within the SR domain. Similar to a router ID, it typically is attached to the loopback of the node. See Prefix-SID for how a Node SID is derived.

Adj-SID
Also known as IGP Adjacency Segment. A segment local only to the node. Dynamically allocated and advertised only to its direct neighbor via the adjacency link. The Adj-SID, when dynamically allocated uses the next label after the SRGB. Each Adj-SID is unique with the node only and mapped to its link to the next neighbor.

Example:
If SRGB = 16000 – 19000, then the first dynamically allocated Adj-SID is 19001.

Prefix-SID
Also known as prefix segment or Node SID. Global segment attached to a prefix.
Prefix-SID = SRGB base + Index table = Absolute value

Example:
SRGB base = 16000
Index = 3
Prefix-SID=16000 + 3 = 16003

Binding-SID
An outer SID that nests a segment list, commonly used to stitch across various domains. This can be thought as a form of label compression to reduce the lable stack depath

2. Proof of Concept Lab

The SR Proof of Concept (POC) lab consist of a virtual and physical environment which mocks up the Charter market. The purpose in using both environment is to eliminate the need for too many physical devices to make up a simulated network. A Nexus 5548UP switch was used to bridge between the virtual and the physical environment.

2.1. Virtual Environment

Cisco Modeling Lab (CML) was used as the virtual environment platform. Each node had a mix of a CRR or a DTR.

Platform – Cisco Modeling Lab
Platform Version – 1.3
Nodes – 12
Node Image – IOS-XRv9K
Image Software – 6.3.2
Roles – Distribution Routers (DTR) and Core Routers (CRR)

2.2. Physical Environment

Various models of devices were used from three main vendors. These vendors were selected since they were widely used across the former merged companies. The models used in the POC were a representative of the roles they play in current production.

Table 2 – Equipment Roles

Equipment	Software	Role
MX240	18.2	CER
7750 SR-7	15.1	MSC
7750 SR-7	15.1	MSC
QFX10K	18.2	DTR
QFX5K	18.3	DTR
ASR9001	6.4.1	DTR
ASR9001	6.4.1	DTR
ASR9001	6.4.1	DTR
ASR9001	6.4.1	DTR
7210 SAS-M	8.0	CPE

Device roles justification:

1. The Juniper MX series router was placed as a Commercial Edge Router (CER) in the SR POC which have a role as a PE router and an aggregation router for services.
2. The Nokia 7750 SR routers are used as Mobile Switching Center (MSC) routers to simulate CTBH production.
3. The Juniper QFX series are used as DTRs where a CER will connect to it.

2.3. SR POC Topology

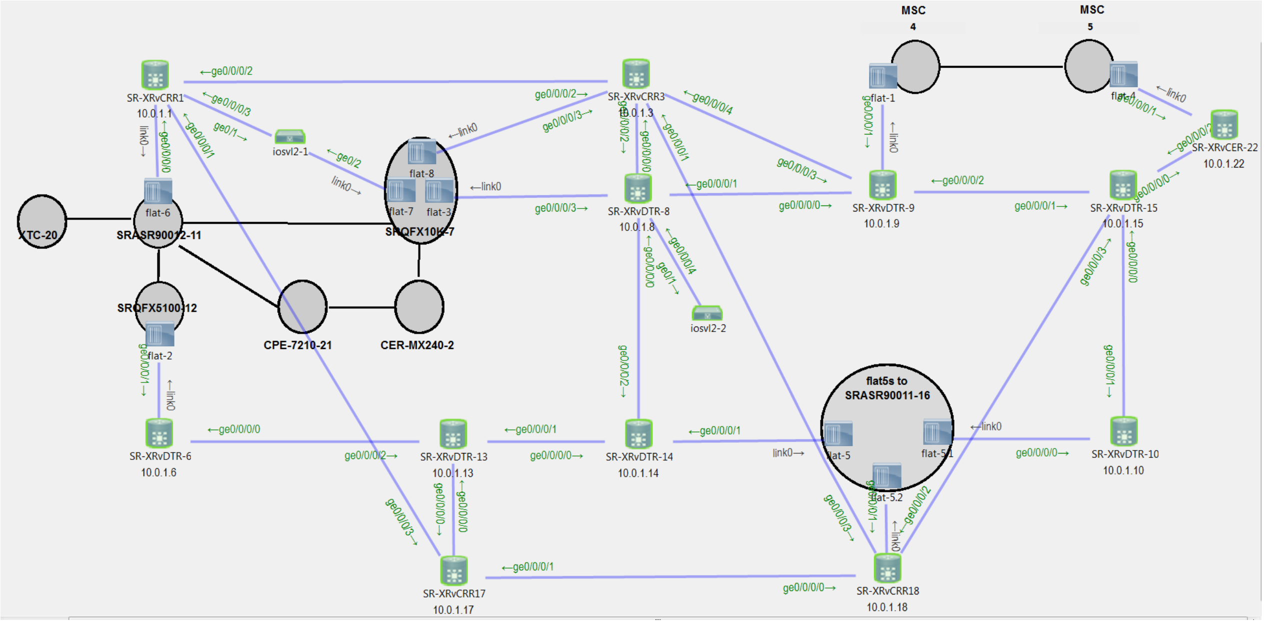


Figure 1 – CML Client View of SR POC

Diagram above is a layout of the SR topology in the CML client view.

▦ Represents an XRv9K node

○ Represents a physical equipment

The SR POC is designed and built based on an example Charter market with an added scenario of an East/West connection. The topology is a hybrid of the hub and spoke topology and the East West express route, where selected DTRs are assigned back to select CRRs (represented by the Hub and Spoke connections) and a high metric cost on the East West express route.

Below is the logical representation of the topology within CML client view.

Table 3 – Node Type and Role

Node Type	Role
XRv9K	CRR 1
MX240	CER 2
XRv9K	CRR 3
7750 SR-7	MSC 4
7750 SR-7	MSC 5
XRv9K	DTR 6
QFX10K	DTR 7
XRv9K	DTR 8
XRv9K	DTR 9
XRv9K	DTR 10
ASR9001	DTR 11
QFX5K	DTR 12
XRv9K	DTR 13
XRv9K	DTR 14
XRv9K	DTR 15
ASR9001	DTR 16
XRv9K	CRR 17
XRv9K	CRR 18
7210 SAS-M	CPE 21
XRv9K	CER 22

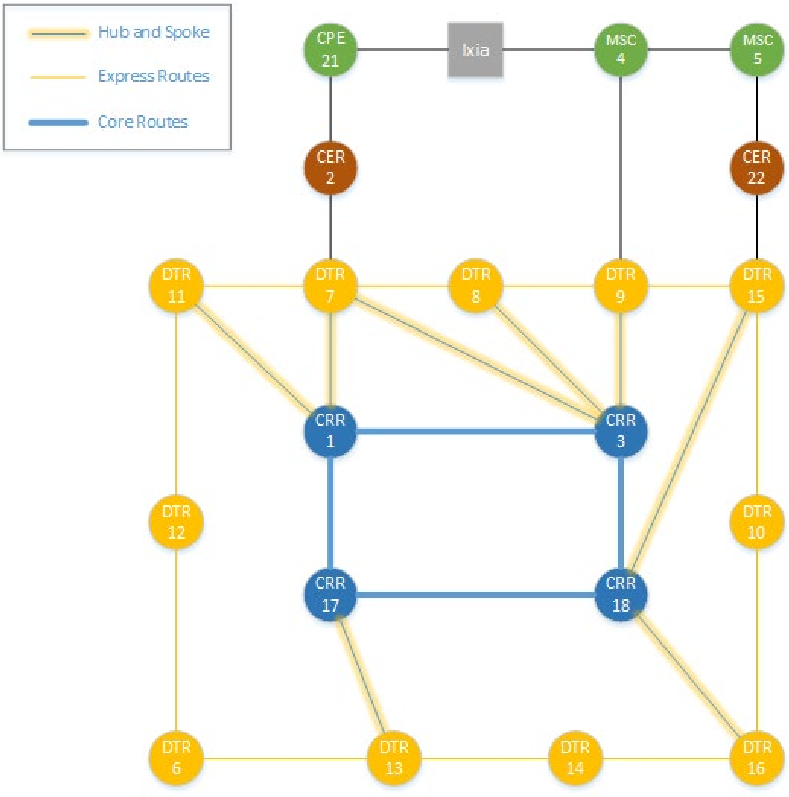


Figure 2 – Logical Topology of SR POC

A simpler topology was built in comparison to the typical hub and spoke design where only a few DTR nodes were homed back to select CRRs. Since traffic was focused from the CPE to the MSC node, only the DTR 7 node is designed to home back to two separate CRRs to allow for simulation of primary path failure.

2.4. Test Equipment, Traffic Analyzer, and Impairment Tools

Test Equipment

Ixia's IxNetwork was used to generate traffic to the SR POC. The type of traffic emulates the CTBH traffic from the subscriber in production today. Ixia was placed between the CPE and the MSC 4 node to allow for bi-directional traffic.

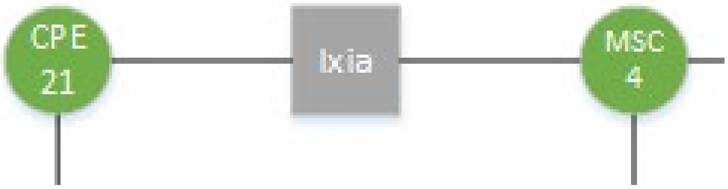
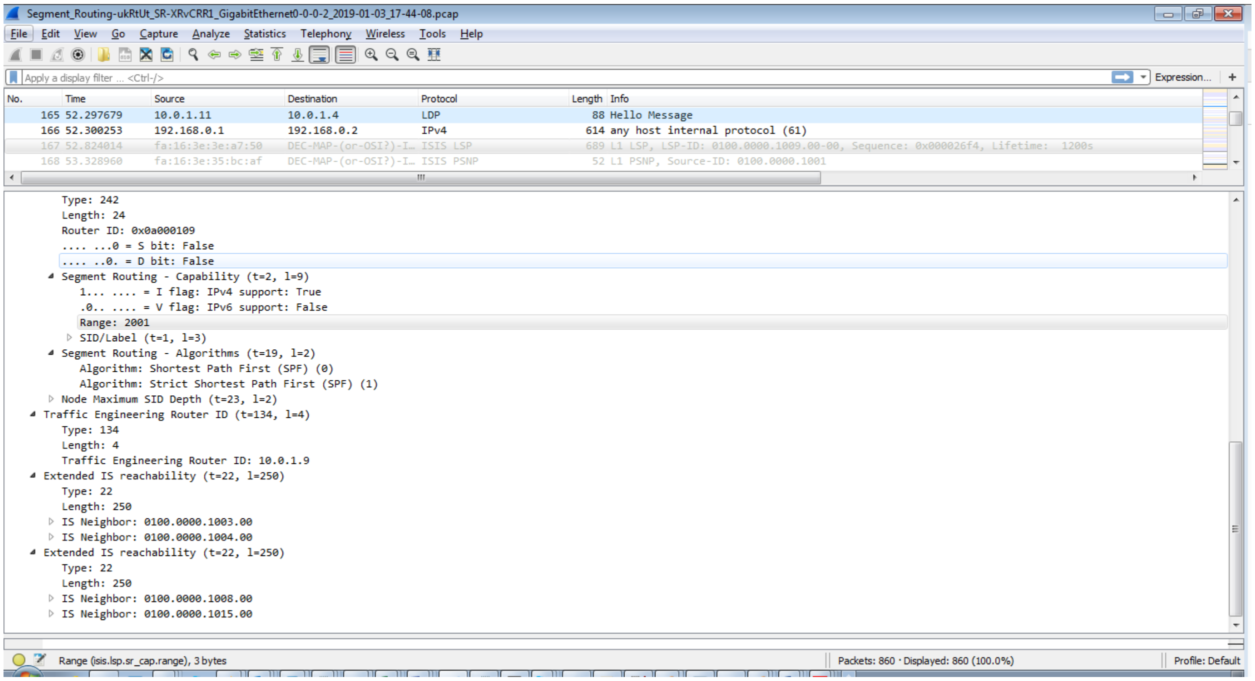


Figure 3 – Test Equipment Placement

Traffic Analyzer

The packet capture feature within CML was used to analyze and understand SR traffic. Since the feature is available within the virtual environment, traffic between physical equipment was not captured.



Segment_Routing-ukRtUt_SR-XRvCRR1_GigabitEthernet0-0-0-2_2019-01-03_17-44-08.pcap

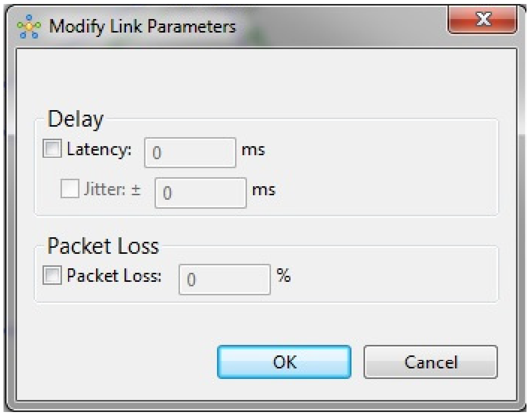
```
No.   Time       Source             Destination          Protocol   Length Info
  165 52.297679  10.0.1.11          10.0.1.4             LDP            88 Hello Message
  166 52.300253  192.168.0.1        192.168.0.2          IPv4          614 any host internal protocol (61)
  167 52.824014  fa:16:3e:3e:a7:50  DEC-MAP-(or-OSI?)-I… ISIS LSP      689 L1 LSP, LSP-ID: 0100.0000.1009.00-00, Sequence: 0x000026f4, Lifetime:  1200s
  168 53.328960  fa:16:3e:35:bc:af  DEC-MAP-(or-OSI?)-I… ISIS PSNP      52 L1 PSNP, Source-ID: 0100.0000.1001
```

```
            Type: 242
            Length: 24
            Router ID: 0x0a000109
            .... ...0 = S bit: False
            .... ..0. = D bit: False
        ◢ Segment Routing - Capability (t=2, l=9)
              1... .... = I flag: IPv4 support: True
              .0.. .... = V flag: IPv6 support: False
              Range: 2001
          ▷ SID/Label (t=1, l=3)
        ◢ Segment Routing - Algorithms (t=19, l=2)
              Algorithm: Shortest Path First (SPF) (0)
              Algorithm: Strict Shortest Path First (SPF) (1)
        ▷ Node Maximum SID Depth (t=23, l=2)
      ◢ Traffic Engineering Router ID (t=134, l=4)
            Type: 134
            Length: 4
            Traffic Engineering Router ID: 10.0.1.9
      ◢ Extended IS reachability (t=22, l=250)
            Type: 22
            Length: 250
          ▷ IS Neighbor: 0100.0000.1003.00
          ▷ IS Neighbor: 0100.0000.1004.00
      ◢ Extended IS reachability (t=22, l=250)
            Type: 22
            Length: 250
          ▷ IS Neighbor: 0100.0000.1008.00
          ▷ IS Neighbor: 0100.0000.1015.00
```

Range (isis.lsp.sr_cap.range), 3 bytes | Packets: 860 · Displayed: 860 (100.0%) | Profile: Default

Figure 4 – Sample Wireshark Capture from CML

Impairment Tools
CML provides the ability to apply latency, jitter, and packet loss at the link level parameters between the XRv9K nodes. Latency was used on the East West direction. The purpose of adding latency is to simulate a physical long path around the ring in the event of a failure on the shorter path from CPE to the MSC and understand the dynamics of SR-TE metric delay constraints.



Modify Link Parameters

Delay

□ Latency: 0 ms

□ Jitter: ± 0 ms

Packet Loss

□ Packet Loss: 0 %

Figure 5 – Impairment Parameters

2.5. Network, Services and Traffic Path

To simulate CTBH services across the TN market, Ethernet over MPLS were implemented from the MSC (7750s) across the IS-IS network to the CER (MX240) with a layer 2 hand-off to the CPE. Simulated CTBH Layer 2 traffic is generated with Ixia's IxNetwork.

Traffic between CPE 21 and MSC 4 will traverse via their local DTR hubs across the SR POC network. Topology below is a hybrid of logical and physical environment. The T-Hub and D-Hub locations are specified to show the physical location and direction of traffic. The DTR homes back to its assigned CRR for forwarding via the Hub and Spoke connections.

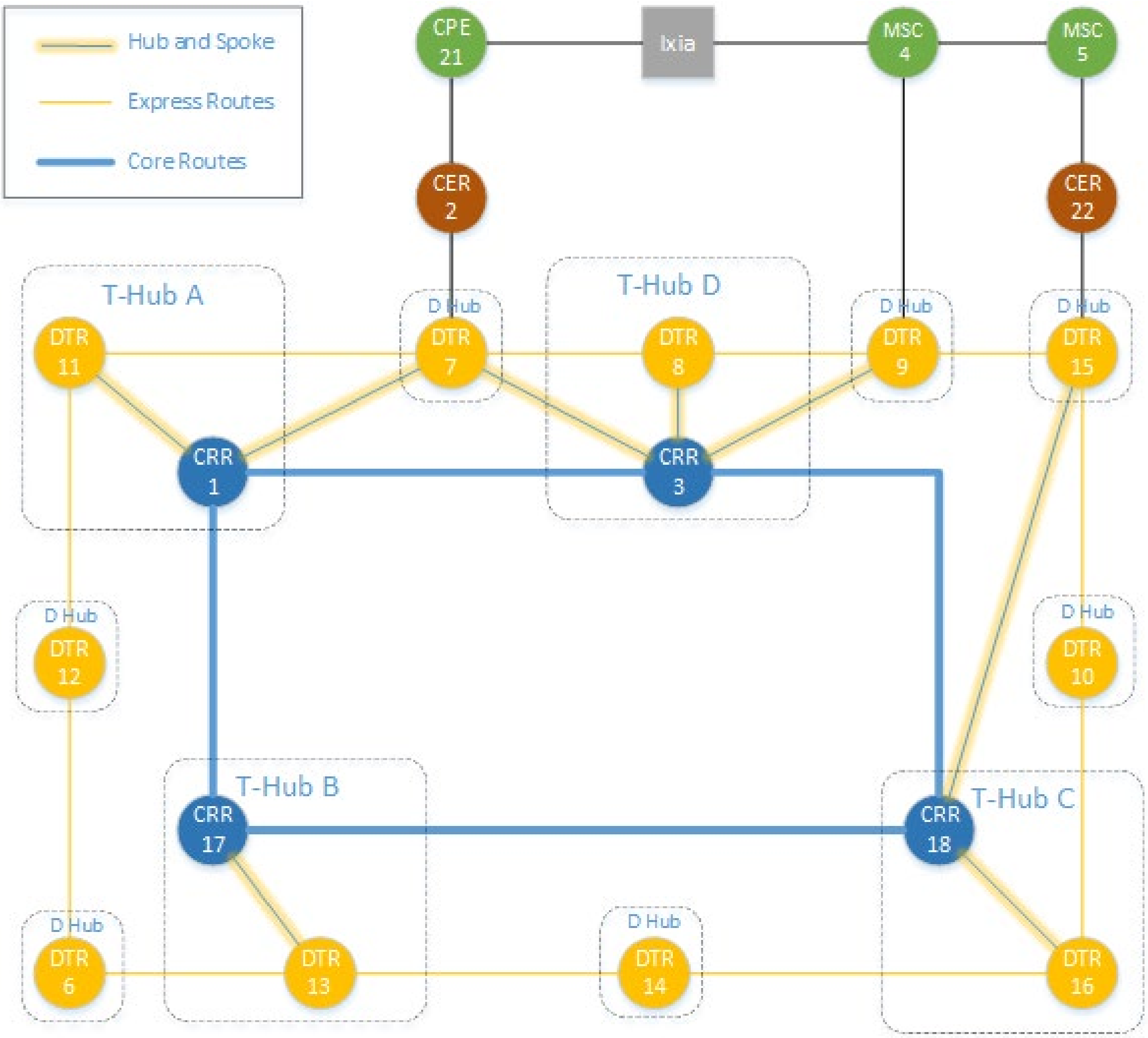


Figure 6 – Logical SR POC Topology with T and D Hub Specification

Test run was performed using the primary and back-up path as shown below to determine the SR characteristics and functionality during a failover. The traffic path is from CPE 21 to MSC 4 via CRR3 and vice versa based on dynamic SR

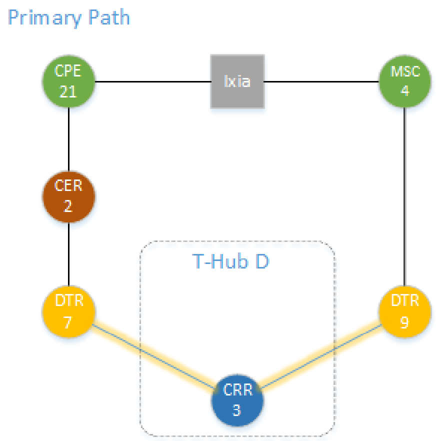


Figure 7 – Primary Path Traffic Pattern

The traffic path from CPE 21 to MSC 4 via CRR1, CRR 17, CRR 18, and CRR 3 in that order during a failover as depicted in the red line and vice versa based on dynamic SR

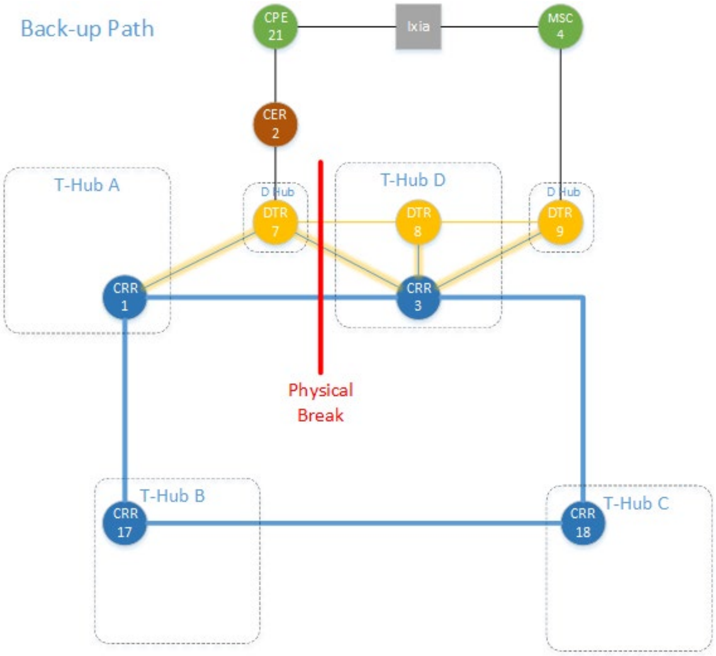


Figure 8 – Back-up Path Traffic Pattern

3. Vendor Platform Operations

Charter utilizes multiple vendor platforms in various networks. To understand how the different platform work and interoperate, three main vendors were selected for the proof of concept testing. These three vendors are Cisco, Juniper, and Nokia.

3.1. Maximum SID Depth (MSD)

MSD is the maximum number of SIDs supported by a node or link. Since each SID is encoded in an MPLS label, the MSD can be referred to as the maximum number of labels supported. The MSD of the device determines the Base MPLS Imposition (BMI), where BMI is the total number of labels imposed inclusive of all service and transport labels.

3.2. Vendor Operations and Comparison

The MSD is defined by the dataplane capability of each vendor platform. There is a variance of MSD supported across multiple vendor platforms due to the type of network processors (NPU) used. Depending on the vendor platforms, MSD at each node can be provisioned if not already set at the maximum. Table 1 below shows a range of MSD from SR capable line cards in production with their the label depth and operational limits. MSD in Table 4 is the BMI.

Table 4 – MSD Range per Vendor

Vendor	Platforms Role	MSD	PUSH	POP
Vendor X	PE	3-16	3-16	2-16
Vendor Y	Core, PE	10	10	10
Vendor Z	Core, PE, CPE	6-12	6-12	6-12

It introduces some complexity when trying to establish an SR LSP or SR-TE LSP along a path that consists of multiple vendors with varying MSDs as shown in Table 1 without exceeding the lowest supported MSD. Careful planning will be required when designing the LSP paths and manual analysis of node MSDs where the LSP traverses can be labor intensive along with the record keeping of MSDs per type of device. Additionally, MSDs can change after software upgrades requiring engineers to keep track of the changes.

Factors that introduces complexity:

1. Tracking of multiple vendor platforms used in the network with different MSDs
2. Inconsistent software on same device models within a network can lead to different MSDs
3. Node MSD and link MSD are not homogenous leading to additional leg work in tracking MSD types
4. Different line cards within a node supports different MSD

All the complexity mentioned above can be mitigated with the use of a controller as defined in following section.

3.3. Controller Use Case and Interoperability

When using a controller to compute the SR paths, the controller can learn the MSDs of each node and ensure the segment list depth does not exceed the MSD of the nodes on the computed path. The controller can receive the MSD of nodes via advertisement methods below.

Node MSD Advertisement Methods
There are 4 ways to advertise MSD capabilities.

1. IS-IS
   - Using the **Node MSD sub-TLV** within the **Router Capability TLV**
2. OSPF
   - Using the **Node MSD sub-TLV** within the LSA Type Opaque
3. Path Computation Element Protocol (PCEP)
   - Using the **SR-PCE-Capability sub-TLV** within the **Path-Setup-Type-Capability TLV**
4. BGP Link State
   - Using **Node Attribute TLV**

However, each controller's learning capability varies per vendor. See section "*Gap Analysis across vendor platforms*" for specific signaling protocols used for learning the MSD.

Gap Analysis across vendor platforms

Four vendor controllers and three vendor devices are selected as the subject for interoperability of MSD signaling between controllers and devices in this document. Currently, the learning and advertisement of MSD varies across multiple vendor platforms for both controller and nodes.

Table 5 – Controller and Node MSD Signaling Protocols

Signaling Protocols	**Controller W**	**Controller X**	**Controller Y**	**Controller Z**
IS-IS	Yes			Future
OSPF	Yes			Future
PCEP		Yes	Yes	Yes
BGP-LS	Yes			Future

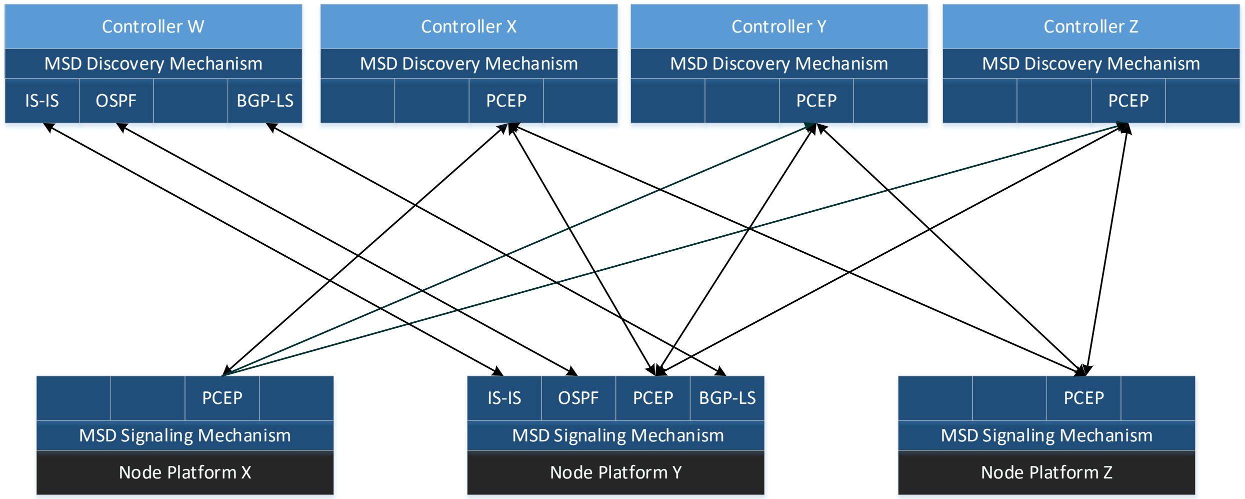


Figure 9 - Devices to Controller Interoperability for Node MSD Discover and Signaling

As shown above, not all vendor platforms support the same signaling protocol for MSD. While all controllers support learning of the MSD via PCEP except Controller W, which only supports IGP and BGP-LS, Controller Y additionally supports protocols such as IGP and BGP-LS.

Based on the diagram above, Controller W will have interoperability issues in discovering the MSD with Node X and Node Z device platform.

Transport and Service Labels with MSD

Since the MSD is the BMI that includes all transport and service labels, the controllers currently do not make a distinction between the types of labels. Establishing the MSD without that distinction between transport and service labels can result in the Path Computation Element (PCE) computing a sub-optimal path and/or returning a path that exceeds the MSD of a node without taking into account the service labels. Hence, a constraint can be set using the metric object in an exchange between the PCE and Path Computation Client (PCC) to reduce the label depth of a computed path.

A node can consist of the default MSD and the configured MSD. To signal a reduced label depth, the devices must allow configuration of the MSD.

Below is an example of Vendor Y's MSD signaling operations.

Option #1: Learning the default MSD

For the PCE to learn the MSD, the PCC will have a PCEP open session with the PCE. During the open session, the PCC will signal its default MSD X.

Option #2: Learning the configured MSD

PCC signals MSD Y during a path computation request via PCEP within the metric object. MSD Y overrides MSD X if the latter is already learned by the PCE.

Note: Vendor Y's MSD signaling operations complies with IETF Draft - https://tools.ietf.org/html/draft-ietf-pce-segment-routing-16

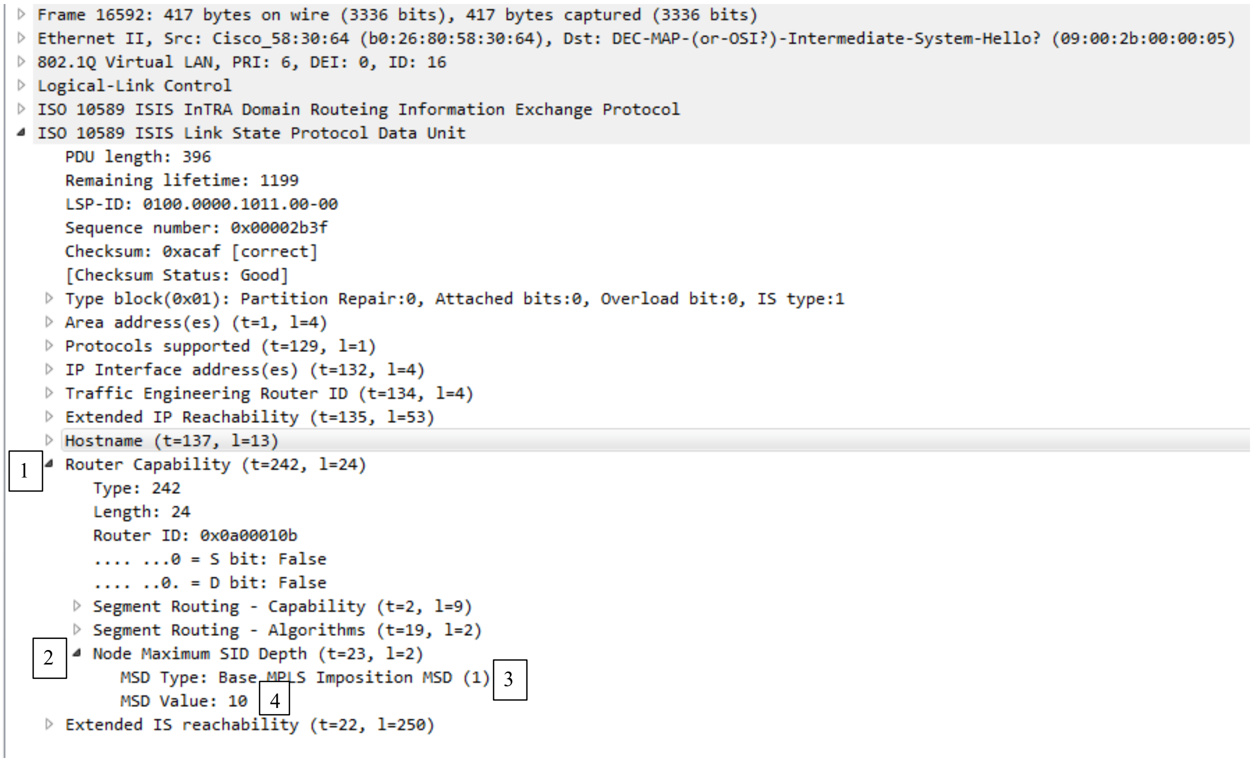
The table below summarizes the interaction of each vendor type controller with a different vendor device on what can be signaled. The exception would be Controller W which is not interoperable with node platform X and Z, but will learn the default MSD only from node platform Y.

Table 6 – Node and Controller Metric Change Capability

	Controller W	Controller X	Controller Y	Controller Z
Node X	Not Interoperable	Default/Configurable	Default/Configurable	Default/Configurable
Node Y	Default	Default/Configurable	Default/Configurable	Default/Configurable
Node Z	Not Interoperable	Default/Configurable	Default/Configurable	Default/Configurable

3.4. Wireshark Captures

Below is a capture of the Node MSD type 23 capability advertised by a Cisco ASR9K as specified in RFC8491.



```
▷ Frame 16592: 417 bytes on wire (3336 bits), 417 bytes captured (3336 bits)
▷ Ethernet II, Src: Cisco_58:30:64 (b0:26:80:58:30:64), Dst: DEC-MAP-(or-OSI?)-Intermediate-System-Hello? (09:00:2b:00:00:05)
▷ 802.1Q Virtual LAN, PRI: 6, DEI: 0, ID: 16
▷ Logical-Link Control
▷ ISO 10589 ISIS InTRA Domain Routeing Information Exchange Protocol
◢ ISO 10589 ISIS Link State Protocol Data Unit
    PDU length: 396
    Remaining lifetime: 1199
    LSP-ID: 0100.0000.1011.00-00
    Sequence number: 0x00002b3f
    Checksum: 0xacaf [correct]
    [Checksum Status: Good]
  ▷ Type block(0x01): Partition Repair:0, Attached bits:0, Overload bit:0, IS type:1
  ▷ Area address(es) (t=1, l=4)
  ▷ Protocols supported (t=129, l=1)
  ▷ IP Interface address(es) (t=132, l=4)
  ▷ Traffic Engineering Router ID (t=134, l=4)
  ▷ Extended IP Reachability (t=135, l=53)
  ▷ Hostname (t=137, l=13)
  ◢ Router Capability (t=242, l=24)
      Type: 242
      Length: 24
      Router ID: 0x0a00010b
      .... ...0 = S bit: False
      .... ..0. = D bit: False
    ▷ Segment Routing - Capability (t=2, l=9)
    ▷ Segment Routing - Algorithms (t=19, l=2)
    ◢ Node Maximum SID Depth (t=23, l=2)
        MSD Type: Base MPLS Imposition MSD (1)
        MSD Value: 10
  ▷ Extended IS reachability (t=22, l=250)
```

Figure 10 – Example capture of MSD TLV

1 All SR capabilities and information are stored in the IS-IS Router Capability TLV Type 242.

2 Sub-TLV for node MSD consisting of MSD Type and value.

3 Supported MSD Type of BMI MSD specifies that the MSD is based on the total amount the device imposition of labels.

4 MSD value of 10 indicates a supported Maximum Depth of 10 SIDs.

Note: Wireshark captures of Nokia and Juniper do not show MSD capabilities within the IS-IS sub-TLVs as they were not supported at the time of POC testing.

3.5. Segment Routing Global Block (SRGB)

Segment Routing Global Block (SRGB) is a local range of blocks recognized within a node. While SRGB is globally unique, the three vendors were found to have their own default SRGB or SRGB configurable block. The diagram below shows the difference between three vendors and the common SRGB space.

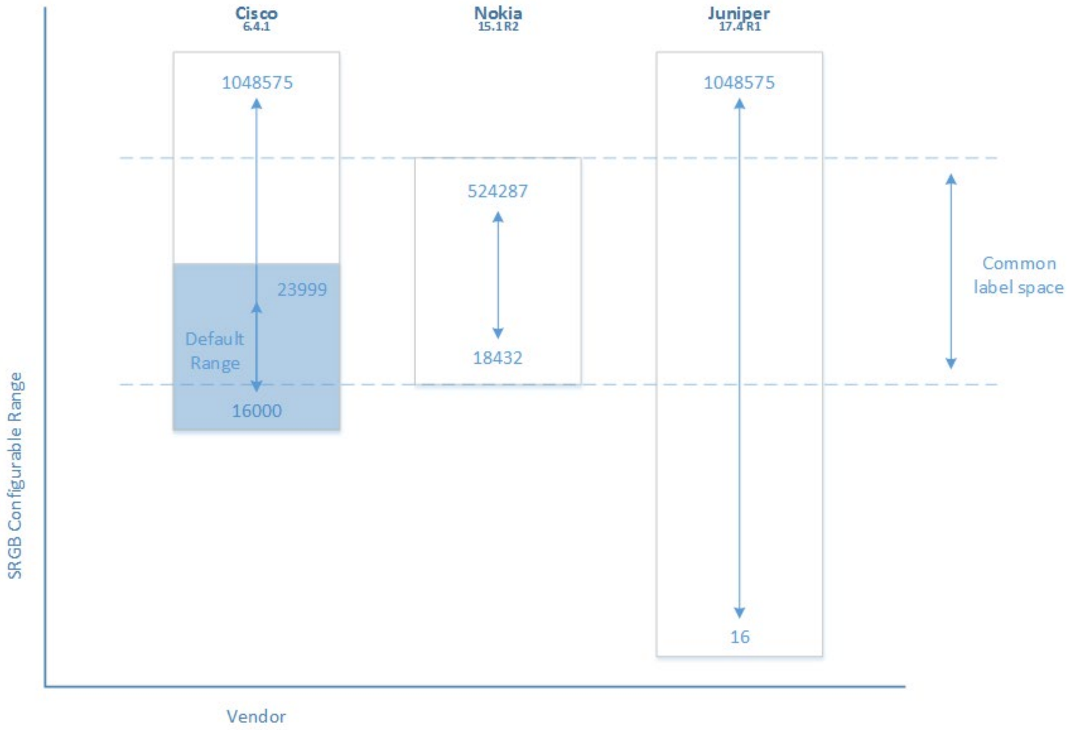
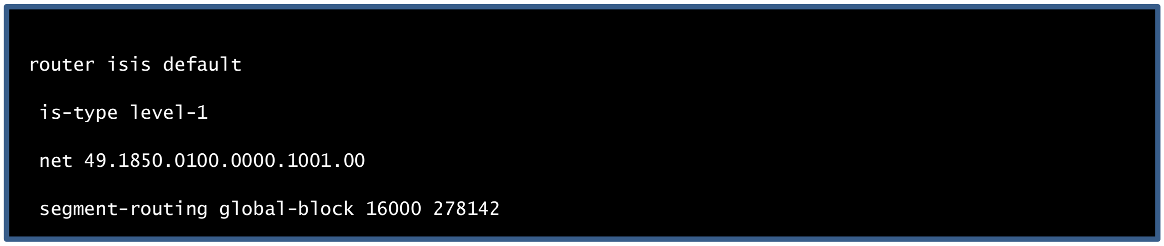


Figure 11 – SRGB Vendor Comparison

<u>Cisco</u>

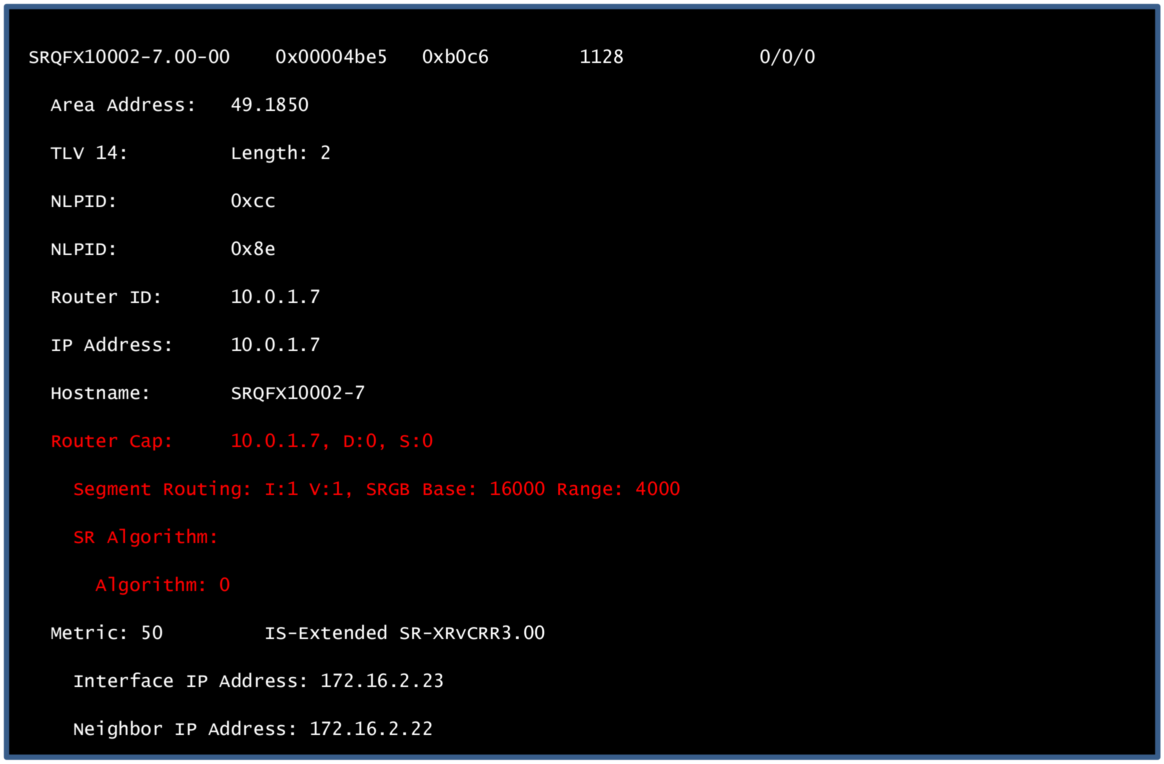
Cisco, by default have their SRGB set between 16000 and 23999. However, a different SRGB other than the default is configurable between 16000 and 1048575. Cisco's configuration of the SRGB resides within the IGP instance.



```
router isis default
 is-type level-1
 net 49.1850.0100.0000.1001.00
 segment-routing global-block 16000 278142
```

SRGB of nodes within the IGP database can be seen using the show command below. Within the CRR 1 node is an excerpt of the SRGB of the router capability of the QFX10K.

Node - SR-XRvCRR1
Show Command: *show isis database verbose*

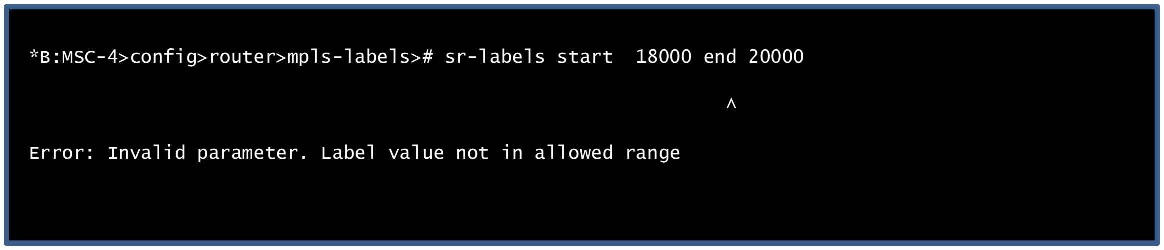


```
SRQFX10002-7.00-00    0x00004be5   0xb0c6        1128            0/0/0
  Area Address:   49.1850
  TLV 14:         Length: 2
  NLPID:          0xcc
  NLPID:          0x8e
  Router ID:      10.0.1.7
  IP Address:     10.0.1.7
  Hostname:       SRQFX10002-7
  Router Cap:     10.0.1.7, D:0, S:0
    Segment Routing: I:1 V:1, SRGB Base: 16000 Range: 4000
    SR Algorithm:
      Algorithm: 0
  Metric: 50         IS-Extended SR-XRvCRR3.00
    Interface IP Address: 172.16.2.23
    Neighbor IP Address: 172.16.2.22
```

Note the SRGB base of 16000 and Range of 4000 configured above.

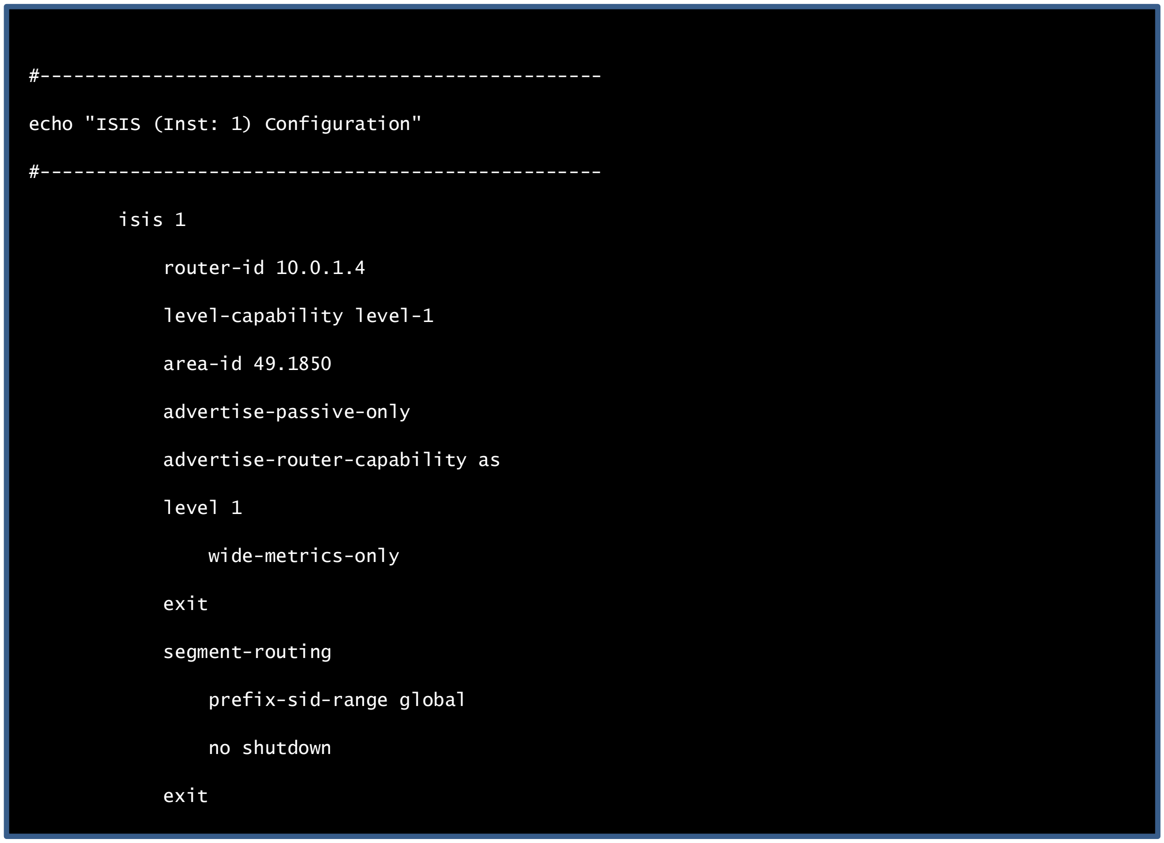
<u>Nokia</u>

Nokia does not have a default SRGB. A configurable value is allowed between 18432 and 524287 within the device's dynamic label range. Configuring outside of the range results in the error shown below.



```
*B:MSC-4>config>router>mpls-labels># sr-labels start  18000 end 20000
                                                              ^
Error: Invalid parameter. Label value not in allowed range
```

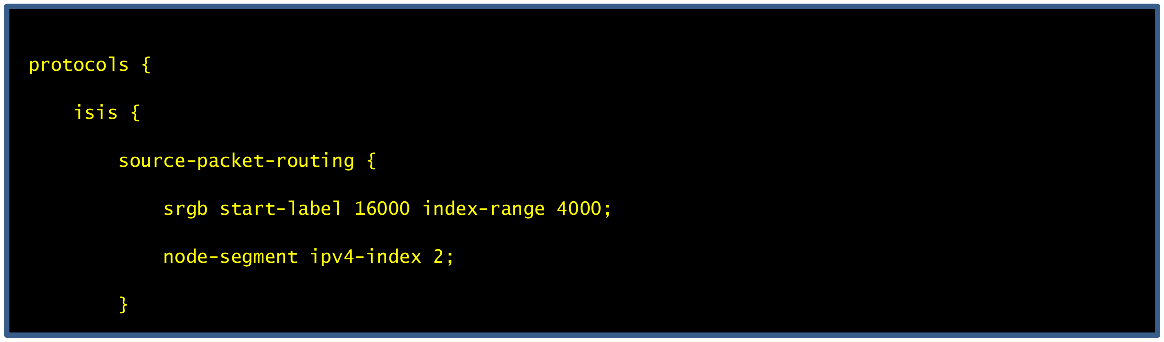
Configuring the SRGB differs from Juniper and Cisco. Rather than configuring the SRGB in the IGP instance, it is configured in a new category called "mpls-labels" under the router field as shown.
Note: Activating the SR is still required in the IGP instance.



```
#--------------------------------------------------
echo "ISIS (Inst: 1) Configuration"
#--------------------------------------------------
        isis 1
            router-id 10.0.1.4
            level-capability level-1
            area-id 49.1850
            advertise-passive-only
            advertise-router-capability as
            level 1
                wide-metrics-only
            exit
            segment-routing
                prefix-sid-range global
                no shutdown
            exit
```

Juniper

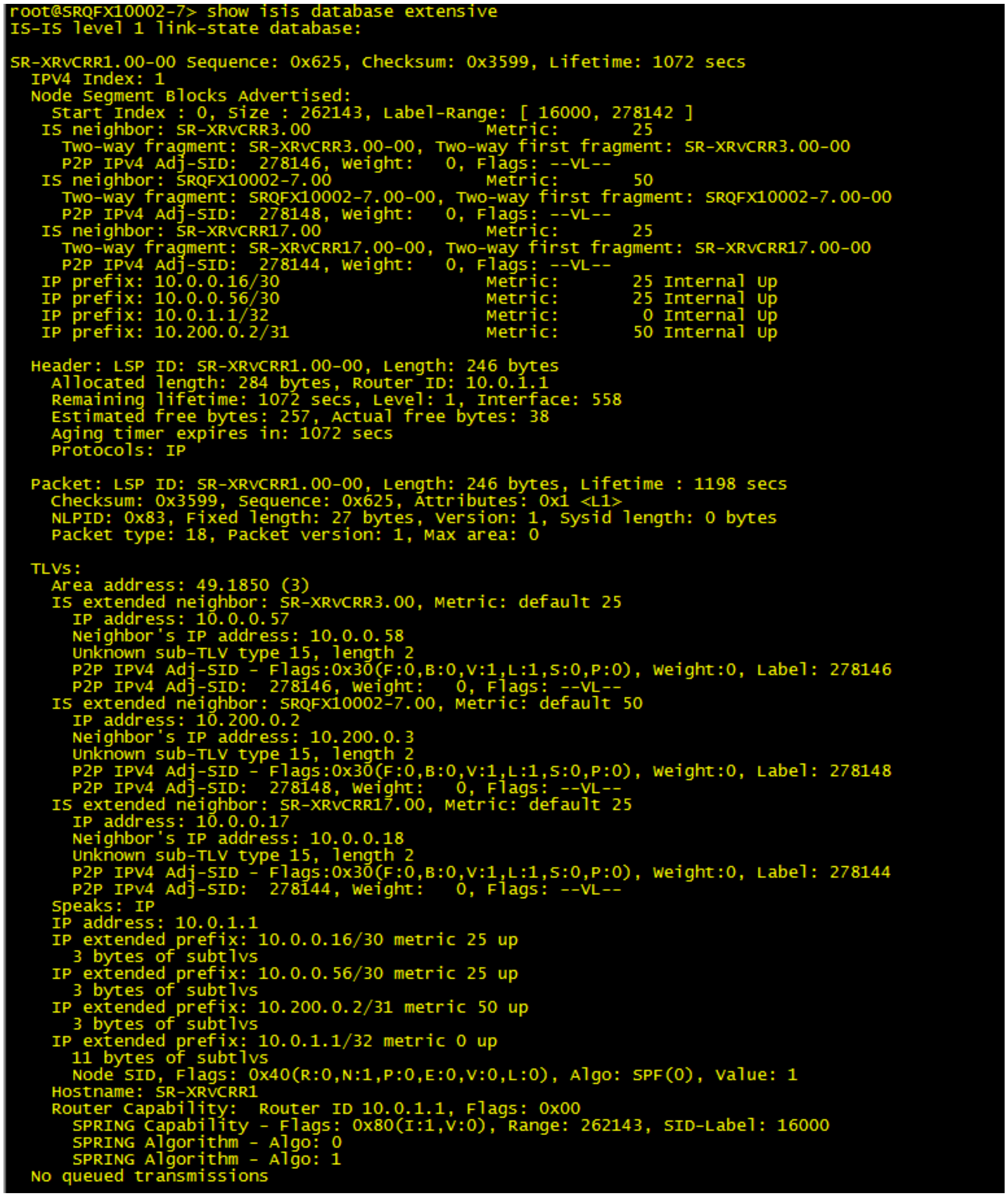
Juniper does not have a default SRGB. A configurable range between 16 and 1048575 is allowed. Juniper devices' SRGB configuration resides within the IGP instance.



```
protocols {
    isis {
        source-packet-routing {
            srgb start-label 16000 index-range 4000;
            node-segment ipv4-index 2;
        }
```

Node – SRQFX10002-7

Show Command: *show isis database database extensive*



```
root@SRQFX10002-7> show isis database extensive
IS-IS level 1 link-state database:

SR-XRvCRR1.00-00 Sequence: 0x625, Checksum: 0x3599, Lifetime: 1072 secs
  IPV4 Index: 1
  Node Segment Blocks Advertised:
    Start Index : 0, Size : 262143, Label-Range: [ 16000, 278142 ]
   IS neighbor: SR-XRvCRR3.00                   Metric:       25
     Two-way fragment: SR-XRvCRR3.00-00, Two-way first fragment: SR-XRvCRR3.00-00
     P2P IPv4 Adj-SID:  278146, Weight:   0, Flags: --VL--
   IS neighbor: SRQFX10002-7.00                 Metric:       50
     Two-way fragment: SRQFX10002-7.00-00, Two-way first fragment: SRQFX10002-7.00-00
     P2P IPv4 Adj-SID:  278148, Weight:   0, Flags: --VL--
   IS neighbor: SR-XRvCRR17.00                  Metric:       25
     Two-way fragment: SR-XRvCRR17.00-00, Two-way first fragment: SR-XRvCRR17.00-00
     P2P IPv4 Adj-SID:  278144, Weight:   0, Flags: --VL--
   IP prefix: 10.0.0.16/30                      Metric:       25 Internal Up
   IP prefix: 10.0.0.56/30                      Metric:       25 Internal Up
   IP prefix: 10.0.1.1/32                       Metric:        0 Internal Up
   IP prefix: 10.200.0.2/31                     Metric:       50 Internal Up

  Header: LSP ID: SR-XRvCRR1.00-00, Length: 246 bytes
    Allocated length: 284 bytes, Router ID: 10.0.1.1
    Remaining lifetime: 1072 secs, Level: 1, Interface: 558
    Estimated free bytes: 257, Actual free bytes: 38
    Aging timer expires in: 1072 secs
    Protocols: IP

  Packet: LSP ID: SR-XRvCRR1.00-00, Length: 246 bytes, Lifetime : 1198 secs
    Checksum: 0x3599, Sequence: 0x625, Attributes: 0x1 <L1>
    NLPID: 0x83, Fixed length: 27 bytes, Version: 1, Sysid length: 0 bytes
    Packet type: 18, Packet version: 1, Max area: 0

  TLVs:
    Area address: 49.1850 (3)
    IS extended neighbor: SR-XRvCRR3.00, Metric: default 25
      IP address: 10.0.0.57
      Neighbor's IP address: 10.0.0.58
      Unknown sub-TLV type 15, length 2
      P2P IPV4 Adj-SID - Flags:0x30(F:0,B:0,V:1,L:1,S:0,P:0), Weight:0, Label: 278146
      P2P IPv4 Adj-SID:  278146, Weight:   0, Flags: --VL--
    IS extended neighbor: SRQFX10002-7.00, Metric: default 50
      IP address: 10.200.0.2
      Neighbor's IP address: 10.200.0.3
      Unknown sub-TLV type 15, length 2
      P2P IPV4 Adj-SID - Flags:0x30(F:0,B:0,V:1,L:1,S:0,P:0), Weight:0, Label: 278148
      P2P IPv4 Adj-SID:  278148, Weight:   0, Flags: --VL--
    IS extended neighbor: SR-XRvCRR17.00, Metric: default 25
      IP address: 10.0.0.17
      Neighbor's IP address: 10.0.0.18
      Unknown sub-TLV type 15, length 2
      P2P IPV4 Adj-SID - Flags:0x30(F:0,B:0,V:1,L:1,S:0,P:0), Weight:0, Label: 278144
      P2P IPv4 Adj-SID:  278144, Weight:   0, Flags: --VL--
    Speaks: IP
    IP address: 10.0.1.1
    IP extended prefix: 10.0.0.16/30 metric 25 up
      3 bytes of subtlvs
    IP extended prefix: 10.0.0.56/30 metric 25 up
      3 bytes of subtlvs
    IP extended prefix: 10.200.0.2/31 metric 50 up
      3 bytes of subtlvs
    IP extended prefix: 10.0.1.1/32 metric 0 up
      11 bytes of subtlvs
      Node SID, Flags: 0x40(R:0,N:1,P:0,E:0,V:0,L:0), Algo: SPF(0), Value: 1
    Hostname: SR-XRvCRR1
    Router Capability:  Router ID 10.0.1.1, Flags: 0x00
      SPRING Capability - Flags: 0x80(I:1,V:0), Range: 262143, SID-Label: 16000
      SPRING Algorithm - Algo: 0
      SPRING Algorithm - Algo: 1
  No queued transmissions
```

Note: All non-SR nodes will start receiving SR router capabilities in their LSDB when there are SR nodes in the network.

3.6. Segment Routing Mapping Server

Purpose: Interoperability between SR and LDP

Functions:

1. Creates a database of prefixes which are not SR capable for both mapping servers and clients.

2. Advertises prefix to SID mappings of non SR routers to SR routers.
3. A control plane mechanism.
4. Part of IGP extensions encoded in SID/Label Binding TLV and Extended Prefix range TLV for ISIS and OSPF, respectively.

Restrictions:

1. Mapping Server must have IGP adjacency to the network.
2. For a network that relies on mapping servers to interop between protocols, a redundant mapping server is recommended.
3. SID-mapping entries learned from one IGP process or instance, cannot be used to learn or calculate prefix-SIDs from another IGP process or instance. Each mapping server is required to be configured per IGP instance.
4. Does not support VRFs.
5. For traffic path from SR domain to LDP domain, a border router between both domains must be enabled with SR and LDP. This document refers to the border router as an SR/LDP border router.

Deployment Methods:

1. Dedicated physical device not inline
2. Inline device
3. Virtualized

Best practice:

1. No more than two mapping servers. Too many is counter-productive and results in having to track the Segment Routing Mapping Server (SRMS) prefix configurations to ensure they are the same across the board.
2. Placement of SRMS where the only two are in a single hub poses a single point of failure.

When is a Mapping Server (SRMS) required:

1. LDP to SR
    - Does not require SRMS
    - Why is SRMS not required?
        - When Independent Label Distribution Control Mode (RFC5036) is active on the router that is on the border of the LDP and SR domain.
        - Any node on the LDP to Segment Routing border automatically installs LDP-to-SR forwarding entries
2. SR to LDP
    - Requires SRMS
    - Why is SRMS required?

- Since LDP nodes are not capable of advertising a prefix SID, the SRMS acts as a translator for all other SR nodes by mapping prefixes of LDP nodes to SIDs.
- It advertises the prefix-to-SID mappings to all other SR nodes, which are mapping clients.

3.6.1. SR/LDP Domain Topology

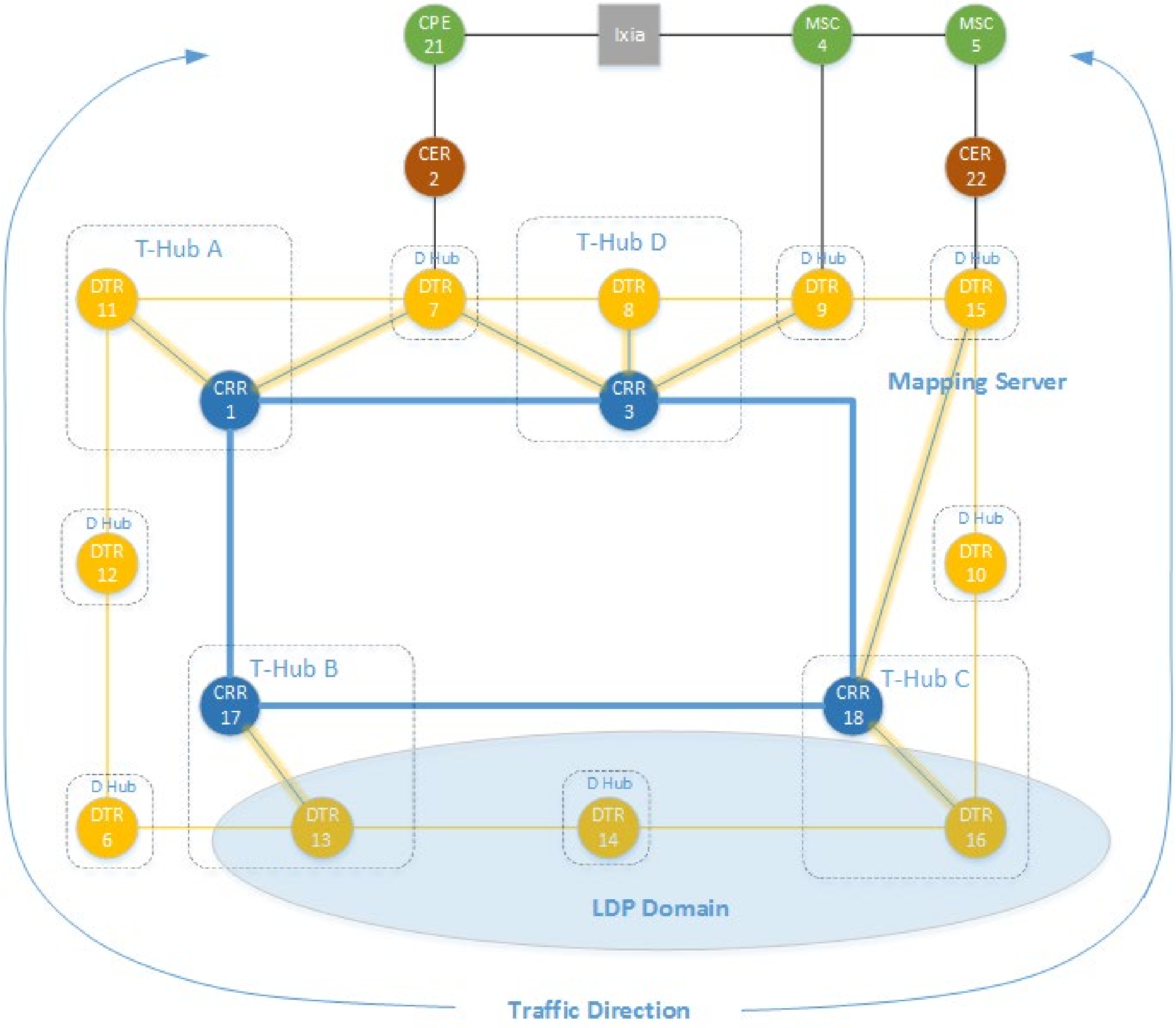


Figure 12 –SR-LDP-SR Topology

The LDP domain consist of Node 13, 14, and 16, while all other nodes are SR nodes. The exception pertains to DTR 9 and CER 2, where both nodes are the termination point of the L2VPN service. Hence SR and LDP are enabled for targeted LDP sessions. Traffic is steered around the outer ring of the topology as depicted in figure 12.

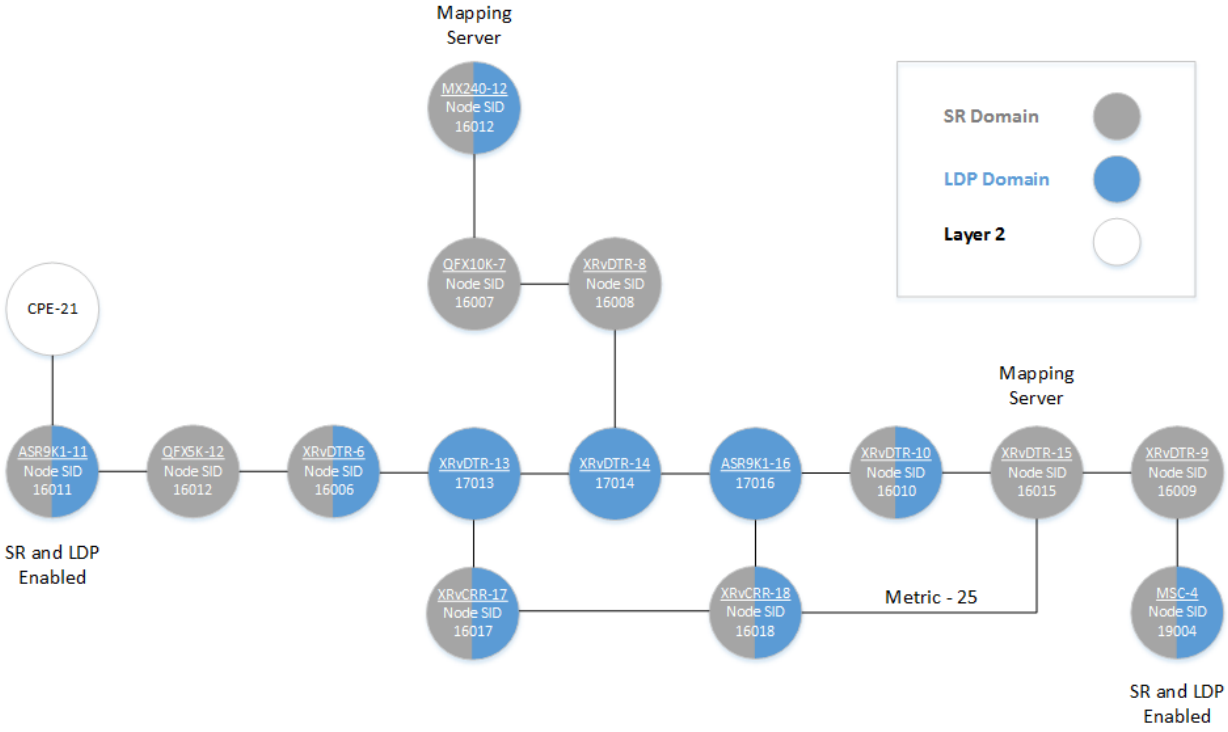


Figure 13 – SR/LDP with Node SIDs

Figure 13 represents SR/LDP topology with node SIDs. Index of 1013, 1014 and 1016 was used for prefix-to-SID mapping of LDP devices. A second mapping server is added for SRMS redundancy testing for a later test case.

3.6.2. Test Case 1 – SR-LDP-SR Interworking with vs without border routers

The purpose of this test case is to test the behavior of mapping server advertisements and mapping client interaction. Note that XRvCRR-17 node is not set as an SR/LDP border router.

(While all vendors documentation stipulate that the border between SR and LDP regions run both LDP and SR, acting as SRMS client and perform stitching; this test case was created to further understand SR/LDP interworking and scan for any unknowns.)

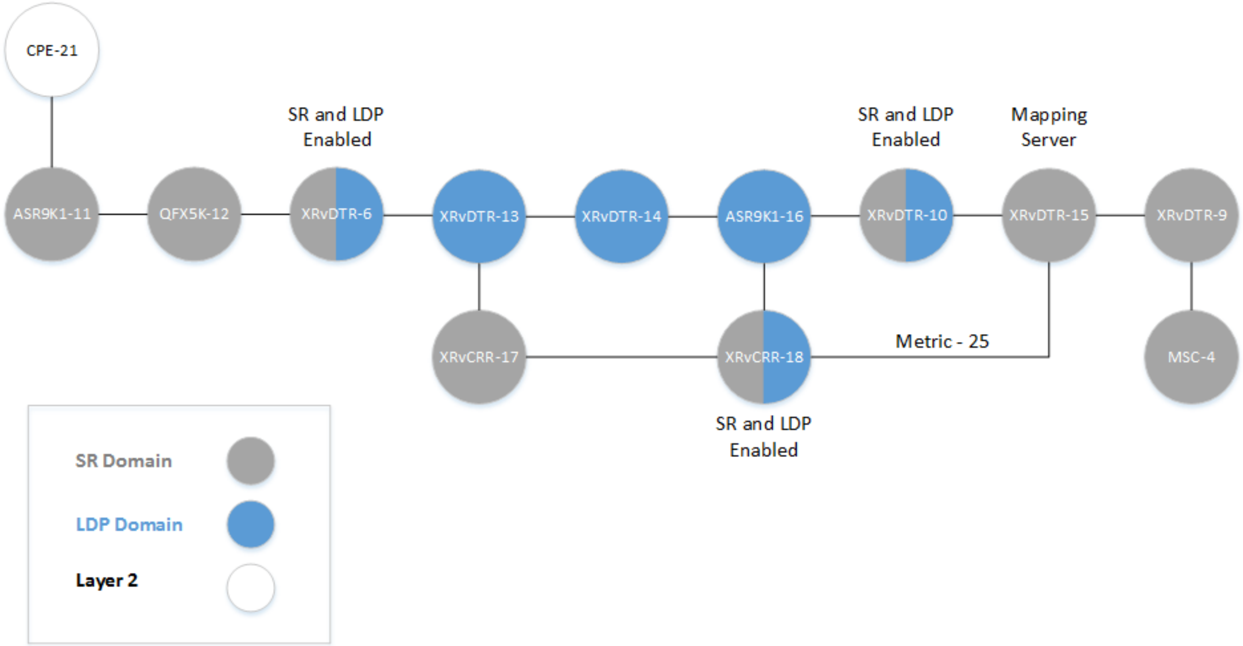


Figure 14 – SR Topology with and without SR/LDP border routers

1. Only XRvDTR-6, XRVDTR-10, and XRvDTR-18 have both SR and LDP enabled.
2. All other devices within SR domain are SR enabled only.
3. All other devices within LDP domain are LDP enabled only.
4. XRvDTR-15 is the mapping server, the rest of the nodes are mapping clients.
5. XRvCRR-17 is not an SR/LDP border router.

Outcome when traffic passes through a non SR/LDP border router:
With XRvCRR-17 set as an SR node instead of an SR/LDP border router, traffic failed when steered through the lower cost path as shown in step 4 below. Should Node 17 be enabled as SR/LDP border router, traffic will continue to be forwarded as shown in the following steps 5 through 7.

Traffic Path: Destination is ASR9K1-11 with a destination SR label of 16011.

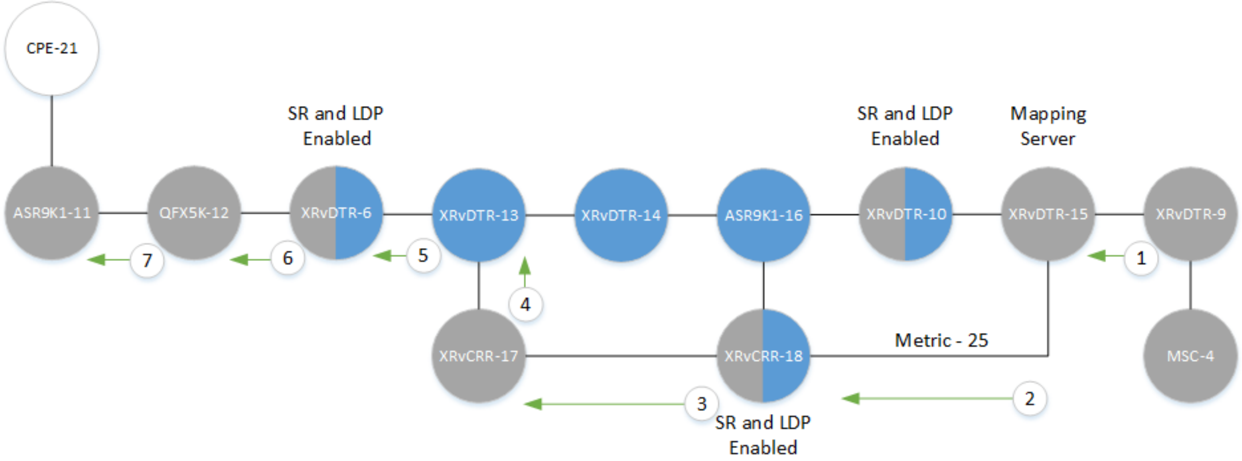
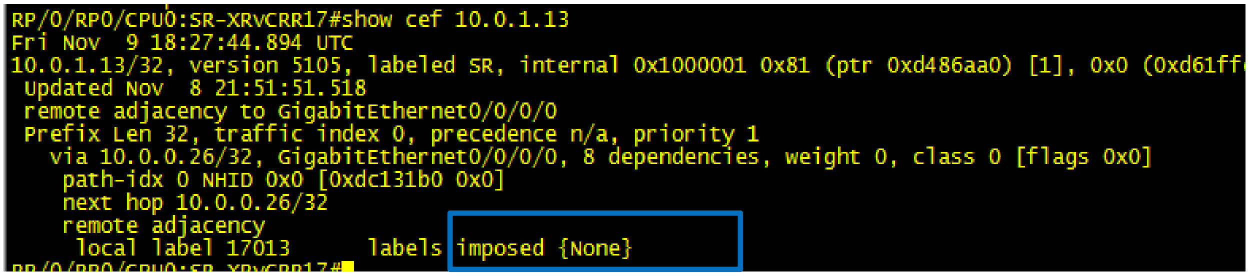


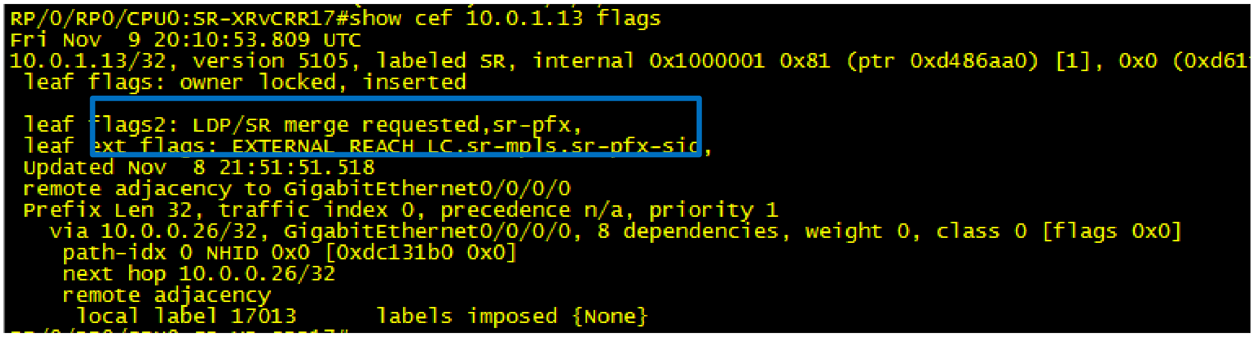
Figure 15 – Traffic Direction SR/LDP Topology without SR/LDP router

1. XRvDTR-9 pushes label 16011 which is ASR9K1-11’s Node SID.
2. XRvDTR-15 continues path to ASR9K1-11 using label 16011.
3. XRvCRR-18 continues path to ASR9K1-11 using label 16011.
4. XRvCRR-17, an SR only node did not receive a label binding FEC from XRvDTR-13, hence do not have label. This is because XRvCRR-17 is not enabled as a border router for SR/LDP. Even when node 17 has received a prefix-to-SID mapping label of 17013 to node 13 (10.0.1.13), no labels were imposed.



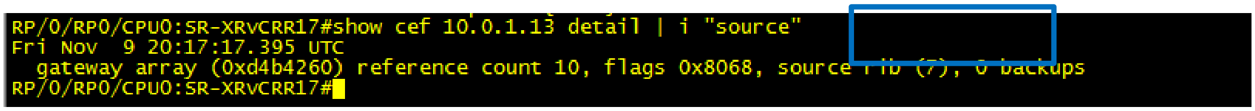
```
RP/0/RP0/CPU0:SR-XRvCRR17#show cef 10.0.1.13
Fri Nov  9 18:27:44.894 UTC
10.0.1.13/32, version 5105, labeled SR, internal 0x1000001 0x81 (ptr 0xd486aa0) [1], 0x0 (0xd61ff
 Updated Nov  8 21:51:51.518
 remote adjacency to GigabitEthernet0/0/0/0
 Prefix Len 32, traffic index 0, precedence n/a, priority 1
   via 10.0.0.26/32, GigabitEthernet0/0/0/0, 8 dependencies, weight 0, class 0 [flags 0x0]
    path-idx 0 NHID 0x0 [0xdc131b0 0x0]
    next hop 10.0.0.26/32
    remote adjacency
     local label 17013      labels imposed {None}
RP/0/RP0/CPU0:SR-XRvCRR17#
```

Looking further, SR/LDP merge is requested but has no active flag. There were no operations to replace the SR label with an LDP label.



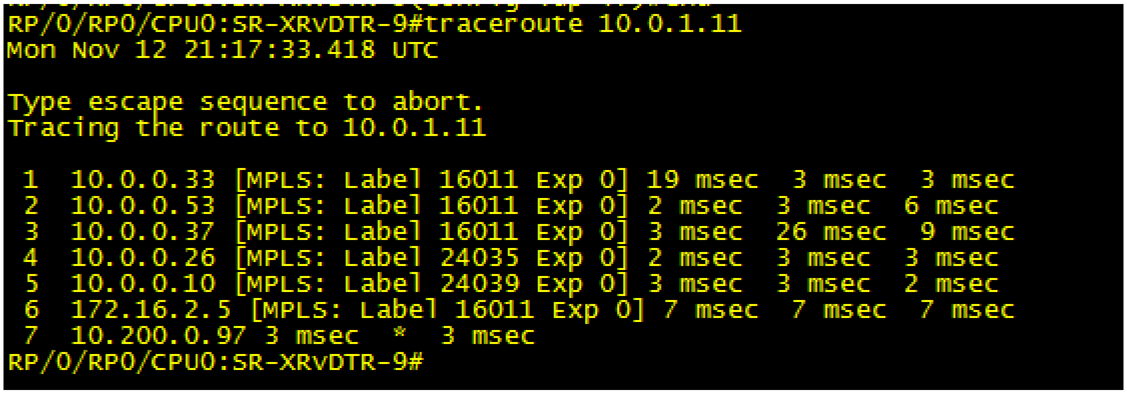
```
RP/0/RP0/CPU0:SR-XRvCRR17#show cef 10.0.1.13 flags
Fri Nov  9 20:10:53.809 UTC
10.0.1.13/32, version 5105, labeled SR, internal 0x1000001 0x81 (ptr 0xd486aa0) [1], 0x0 (0xd61
 leaf flags: owner locked, inserted

 leaf flags2: LDP/SR merge requested,sr-pfx,
 leaf ext flags: EXTERNAL REACH LC,sr-mpls,sr-pfx-sid,
 Updated Nov  8 21:51:51.518
 remote adjacency to GigabitEthernet0/0/0/0
 Prefix Len 32, traffic index 0, precedence n/a, priority 1
   via 10.0.0.26/32, GigabitEthernet0/0/0/0, 8 dependencies, weight 0, class 0 [flags 0x0]
    path-idx 0 NHID 0x0 [0xdc131b0 0x0]
    next hop 10.0.0.26/32
    remote adjacency
     local label 17013      labels imposed {None}
```

To demonstrate further that the mapping client (CRR-17 node) is receiving the mapping policy from the SRMS, the following command shows the local label's source is from the RIB but not the LSD which helps provide the operations to replace the SR label with an LDP label.



```
RP/0/RP0/CPU0:SR-XRvCRR17#show cef 10.0.1.13 detail | i "source"
Fri Nov  9 20:17:17.395 UTC
  gateway array (0xd4b4260) reference count 10, flags 0x8068, source rib (7), 0 backups
RP/0/RP0/CPU0:SR-XRvCRR17#
```

5. Assuming CRR-17 is an SR/LDP border router, XRvDTR-13 pushes LDP label 24039 towards XRvDTR-6.

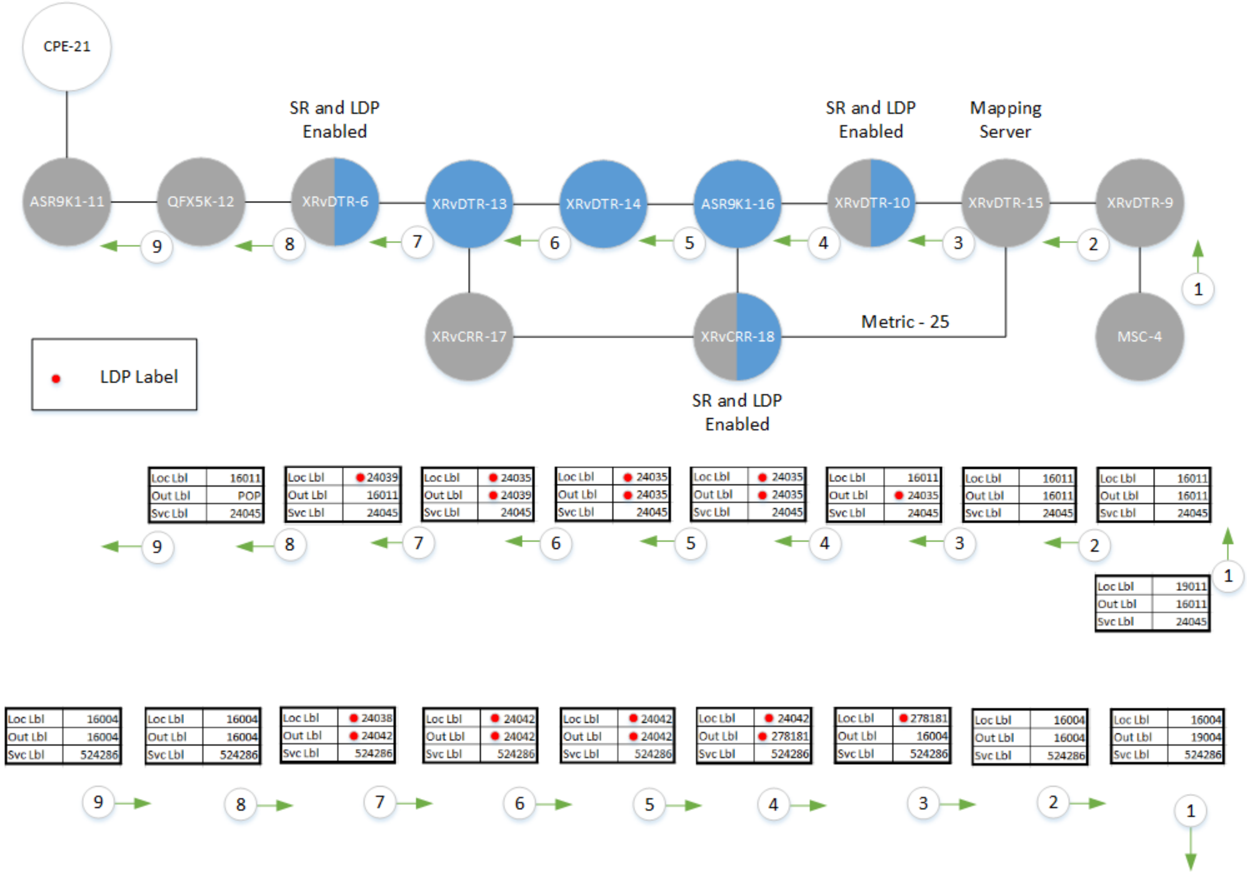


```
RP/0/RP0/CPU0:SR-XRvDTR-9#traceroute 10.0.1.11
Mon Nov 12 21:17:33.418 UTC

Type escape sequence to abort.
Tracing the route to 10.0.1.11

 1  10.0.0.33 [MPLS: Label 16011 Exp 0] 19 msec  3 msec  3 msec
 2  10.0.0.53 [MPLS: Label 16011 Exp 0] 2 msec  3 msec  6 msec
 3  10.0.0.37 [MPLS: Label 16011 Exp 0] 3 msec  26 msec  9 msec
 4  10.0.0.26 [MPLS: Label 24035 Exp 0] 2 msec  3 msec  3 msec
 5  10.0.0.10 [MPLS: Label 24039 Exp 0] 3 msec  3 msec  2 msec
 6  172.16.2.5 [MPLS: Label 16011 Exp 0] 7 msec  7 msec  7 msec
 7  10.200.0.97 3 msec  *  3 msec
RP/0/RP0/CPU0:SR-XRvDTR-9#
```

6. XRvDTR-6, being the SR/LDP border router, swaps LDP label 24039 with node SID label 16011 towards SR node QFX5K-12.
7. QFX5K-12 pops label 16011 upon receipt and sends traffic towards ASR9K1-11.

Outcome when traffic passes through an SR/LDP border router:

Note that each hop is assigned a label. Traffic path goes through SR/LDP border routers, XRvDTR-10 and XRvDTR-6, which allows for SR-to-LDP forwarding and vice versa.



	9	8	7	6	5	4	3	2	1
Loc Lbl	16011	● 24039	● 24035	● 24035	● 24035	16011	16011	16011	19011
Out Lbl	POP	16011	● 24039	● 24035	● 24035	● 24035	16011	16011	16011
Svc Lbl	24045	24045	24045	24045	24045	24045	24045	24045	24045

	9	8	7	6	5	4	3	2	1
Loc Lbl	16004	16004	● 24038	● 24042	● 24042	● 24042	● 278181	16004	16004
Out Lbl	16004	16004	● 24042	● 24042	● 24042	● 278181	16004	16004	19004
Svc Lbl	524286	524286	524286	524286	524286	524286	524286	524286	524286

Figure 16 – Traffic direction with SR and LDP Labels

Looking at how XRvDTR-10 works when receiving it receives an SR labeled traffic, any incoming SR label is stitched to an LDP label.

**Diagram below is a format taken out of the Segment Routing Part 1 book by Clarence Filsfils, Kris Michielsen, Ketan Talaulikar to illustrate operations of an incoming SR label transition to an LDP label within the SR POC Lab. Labels relevant to the SR POC Lab were replaced.*

Function of an SR/LDP border router:

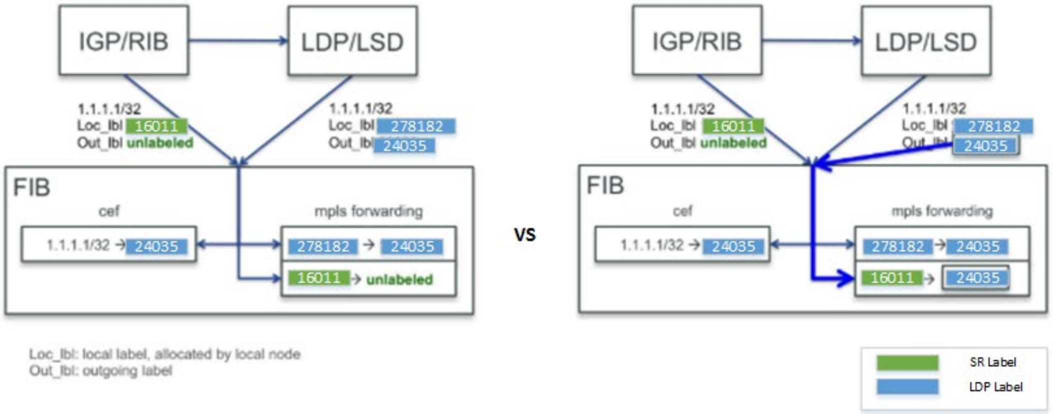
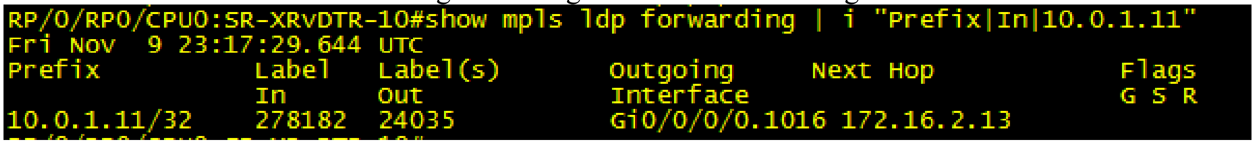


Figure 17 – SR to LDP Operations

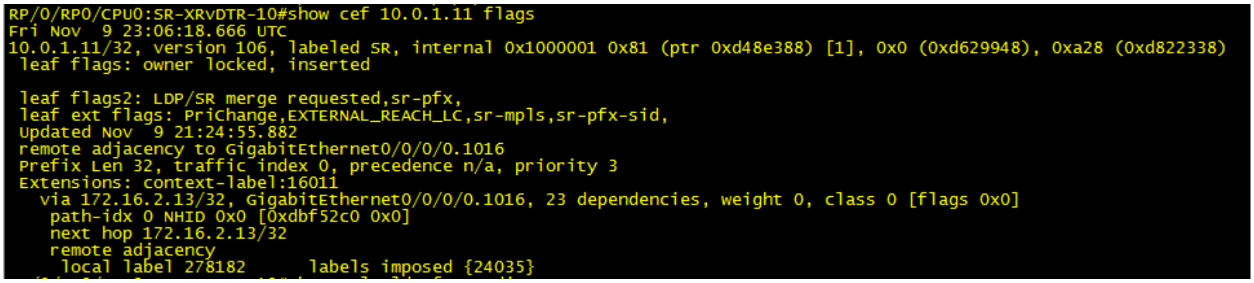
If ASR9K1-16 is not SR enabled, and XRvDTR-10 is not SR/LDP enabled, the latter will not get get an outgoing Prefix-SID label for ASR9K1-11 (10.0.1.11). Without knowing a label to reach 10.0.1.11, XRvDTR-10 provides "unlabeled" outgoing label in mpls forwarding entry for the 10.0.1.11 prefix.

If XRvDTR-10 is SR/LDP enabled, it receives a valid LDP label advertised by ASR9K1-16 on how to reach 10.0.1.11. The LSD provides the received label 24035 to the FIB and replaces the "unlabeled" entry.

Additional illustration to show the ingress and egress labels related to figure 18.



```
RP/0/RP0/CPU0:SR-XRvDTR-10#show mpls ldp forwarding | i "Prefix|In|10.0.1.11"
Fri Nov  9 23:17:29.644 UTC
Prefix          Label   Label(s)       Outgoing     Next Hop            Flags
                In      Out            Interface                        G S R
10.0.1.11/32    278182  24035          Gi0/0/0/0.1016 172.16.2.13
```



```
RP/0/RP0/CPU0:SR-XRvDTR-10#show cef 10.0.1.11 flags
Fri Nov  9 23:06:18.666 UTC
10.0.1.11/32, version 106, labeled SR, internal 0x1000001 0x81 (ptr 0xd48e388) [1], 0x0 (0xd629948), 0xa28 (0xd822338)
 leaf flags: owner locked, inserted

 leaf flags2: LDP/SR merge requested,sr-pfx,
 leaf ext flags: PriChange,EXTERNAL_REACH_LC,sr-mpls,sr-pfx-sid,
 Updated Nov  9 21:24:55.882
 remote adjacency to GigabitEthernet0/0/0/0.1016
 Prefix Len 32, traffic index 0, precedence n/a, priority 3
 Extensions: context-label:16011
   via 172.16.2.13/32, GigabitEthernet0/0/0/0.1016, 23 dependencies, weight 0, class 0 [flags 0x0]
    path-idx 0 NHID 0x0 [0xdbf52c0 0x0]
    next hop 172.16.2.13/32
    remote adjacency
     local label 278182      labels imposed {24035}
```

There is no SR/LDP active because LDP provided a valid outgoing label to FIB.

Overall results show that interworking of SR/LDP is successful despite multiple vendor type used in a single topology.

Outcome when SRGB is out of range of other nodes:

During testing, several mapping clients failed in installing the entries advertised by the SRMS. Configuring a prefix-to-SID mapping index on the SRMS that exceeds the SRGB of the mapping clients, will prevent the installation of the mapping server policies in the clients forwarding table. Table 7 below shows the configured SRGB of each node. When a prefix-to-SID mapping index is configured on the SRMS, it adds to the based SRGB of the mapping clients.

Example:
SRGB base = 16000
Index = 17013
Prefix-to-SID mapping index = SRGB Base + Index = 16000 + 17013 = 33013

When the Prefix-to-SID mapping index falls outside of the mapping client's SRGB range, the entry will not be installed in the forwarding table as depicted in red in the table below.

Table 7 – Prefix-to-SID Index Comparison

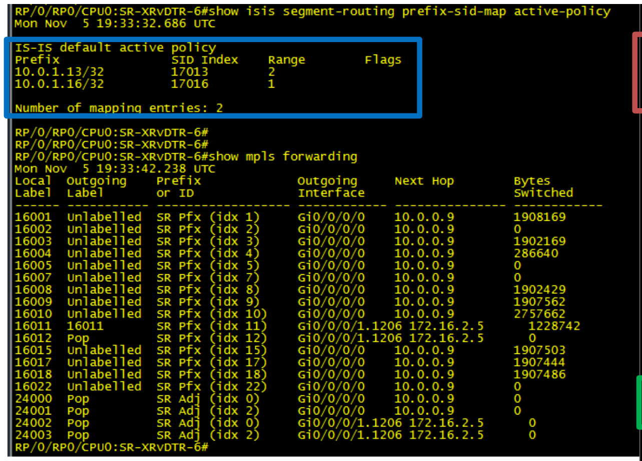
Nodes	SRGB	Index – 17013 (Prefix as seen on Clients)	Index – 1013 (Prefix as seen on Clients)	Index – 3013 (Prefix as seen on Clients)
XRvCER-22	16000 - 278142	33013	17013	19013
SR-XRvCRR1	16000 - 278142	33013	17013	19013
XRvCRR17	16000 - 278142	33013	17013	19013
XRvCRR18	16000 - 278142	33013	17013	19013
XRvCRR3	16000 - 278142	33013	17013	19013
XRvDTR-10	16000 - 278142	33013	17013	19013
XRvDTR-15	16000 - 18000	33013	17013	19013
XRvDTR-6	16000 - 18000	33013	17013	19013
XRvDTR-8	16000 - 18000	33013	17013	19013
MSC-4	19000 - 21000	52013	17013	22013
MSC-5	19000 - 21000	52013	17013	22013
ASR90012-11	16000 - 19000	33013	17013	19013

MX240-2	16000 - 19999	33013	17013	19013
SRQFX5100-12	16000 - 19999	33013	17013	19013
SRQFX10002-7	16000 - 20999	33013	17013	19013
XRvDTR-9	16000 - 18000	33013	17013	19013

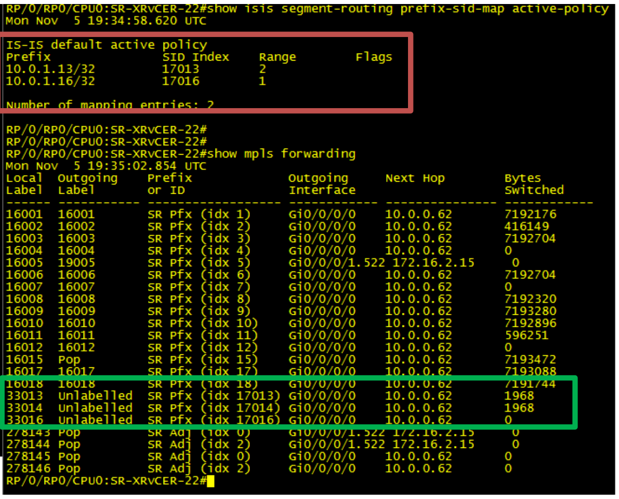
- delete this page since all vendors have BCP on the interworking between SR and LDP with border node between SR and LDP regions run both LDP and SR and act as SRMR clinet

Example Symptom:

Node 6 receives the mapping server advertisements of the prefix-to-SID mappings but does not install them in the forwarding table. See below in (A) and (B) for example.



```
RP/0/RP0/CPU0:SR-XRvDTR-6#show isis segment-routing prefix-sid-map active-policy
Mon Nov  5 19:33:32.686 UTC

IS-IS default active policy
Prefix                SID Index    Range        Flags
10.0.1.13/32          17013        2
10.0.1.16/32          17016        1

Number of mapping entries: 2

RP/0/RP0/CPU0:SR-XRvDTR-6#
RP/0/RP0/CPU0:SR-XRvDTR-6#
RP/0/RP0/CPU0:SR-XRvDTR-6#show mpls forwarding
Mon Nov  5 19:33:42.238 UTC
Local  Outgoing    Prefix             Outgoing     Next Hop        Bytes
Label  Label       or ID              Interface                    Switched
------ ----------- ------------------ ------------ --------------- ------------
16001  Unlabelled  SR Pfx (idx 1)     Gi0/0/0/0    10.0.0.9        1908169
16002  Unlabelled  SR Pfx (idx 2)     Gi0/0/0/0    10.0.0.9        0
16003  Unlabelled  SR Pfx (idx 3)     Gi0/0/0/0    10.0.0.9        1902169
16004  Unlabelled  SR Pfx (idx 4)     Gi0/0/0/0    10.0.0.9        286640
16005  Unlabelled  SR Pfx (idx 5)     Gi0/0/0/0    10.0.0.9        0
16007  Unlabelled  SR Pfx (idx 7)     Gi0/0/0/0    10.0.0.9        0
16008  Unlabelled  SR Pfx (idx 8)     Gi0/0/0/0    10.0.0.9        1902429
16009  Unlabelled  SR Pfx (idx 9)     Gi0/0/0/0    10.0.0.9        1907562
16010  Unlabelled  SR Pfx (idx 10)    Gi0/0/0/0    10.0.0.9        2757662
16011  16011       SR Pfx (idx 11)    Gi0/0/0/1.1206 172.16.2.5      1228742
16012  Pop         SR Pfx (idx 12)    Gi0/0/0/1.1206 172.16.2.5      0
16015  Unlabelled  SR Pfx (idx 15)    Gi0/0/0/0    10.0.0.9        1907503
16017  Unlabelled  SR Pfx (idx 17)    Gi0/0/0/0    10.0.0.9        1907444
16018  Unlabelled  SR Pfx (idx 18)    Gi0/0/0/0    10.0.0.9        1907486
16022  Unlabelled  SR Pfx (idx 22)    Gi0/0/0/0    10.0.0.9        0
24000  Pop         SR Adj (idx 0)     Gi0/0/0/0    10.0.0.9        0
24001  Pop         SR Adj (idx 2)     Gi0/0/0/0    10.0.0.9        0
24002  Pop         SR Adj (idx 0)     Gi0/0/0/1.1206 172.16.2.5      0
24003  Pop         SR Adj (idx 2)     Gi0/0/0/1.1206 172.16.2.5      0
RP/0/RP0/CPU0:SR-XRvDTR-6#
```



```
RP/0/RP0/CPU0:SR-XRvCER-22#show isis segment-routing prefix-sid-map active-policy
Mon Nov  5 19:34:58.620 UTC

IS-IS default active policy
Prefix                SID Index    Range        Flags
10.0.1.13/32          17013        2
10.0.1.16/32          17016        1

Number of mapping entries: 2

RP/0/RP0/CPU0:SR-XRvCER-22#
RP/0/RP0/CPU0:SR-XRvCER-22#
RP/0/RP0/CPU0:SR-XRvCER-22#show mpls forwarding
Mon Nov  5 19:35:02.854 UTC
Local  Outgoing    Prefix             Outgoing     Next Hop        Bytes
Label  Label       or ID              Interface                    Switched
------ ----------- ------------------ ------------ --------------- ------------
16001  16001       SR Pfx (idx 1)     Gi0/0/0/0    10.0.0.62       7192176
16002  16002       SR Pfx (idx 2)     Gi0/0/0/0    10.0.0.62       416149
16003  16003       SR Pfx (idx 3)     Gi0/0/0/0    10.0.0.62       7192704
16004  16004       SR Pfx (idx 4)     Gi0/0/0/0    10.0.0.62       0
16005  19005       SR Pfx (idx 5)     Gi0/0/0/1.522 172.16.2.15      0
16006  16006       SR Pfx (idx 6)     Gi0/0/0/0    10.0.0.62       7192704
16007  16007       SR Pfx (idx 7)     Gi0/0/0/0    10.0.0.62       0
16008  16008       SR Pfx (idx 8)     Gi0/0/0/0    10.0.0.62       7192320
16009  16009       SR Pfx (idx 9)     Gi0/0/0/0    10.0.0.62       7193280
16010  16010       SR Pfx (idx 10)    Gi0/0/0/0    10.0.0.62       7192896
16011  16011       SR Pfx (idx 11)    Gi0/0/0/0    10.0.0.62       596251
16012  16012       SR Pfx (idx 12)    Gi0/0/0/0    10.0.0.62       0
16015  Pop         SR Pfx (idx 15)    Gi0/0/0/0    10.0.0.62       7193472
16017  16017       SR Pfx (idx 17)    Gi0/0/0/0    10.0.0.62       7193088
16018  16018       SR Pfx (idx 18)    Gi0/0/0/0    10.0.0.62       7191744
33013  Unlabelled  SR Pfx (idx 17013) Gi0/0/0/0    10.0.0.62       1968
33014  Unlabelled  SR Pfx (idx 17014) Gi0/0/0/0    10.0.0.62       1968
33016  Unlabelled  SR Pfx (idx 17016) Gi0/0/0/0    10.0.0.62       0
278143 Pop         SR Adj (idx 0)     Gi0/0/0/1.522 172.16.2.15      0
278144 Pop         SR Adj (idx 2)     Gi0/0/0/1.522 172.16.2.15      0
278145 Pop         SR Adj (idx 0)     Gi0/0/0/0    10.0.0.62       0
278146 Pop         SR Adj (idx 2)     Gi0/0/0/0    10.0.0.62       0
RP/0/RP0/CPU0:SR-XRvCER-22#
```

(A) Non-working Mapping Client

Mapping policy received, highlighted in blue. However, when looking at the forwarding table under "*show mpls forwarding*" command, the entries for SID index 17013 and 17016 from the mapping policy are missing.

(B) Working Mapping Client

Mapping policy received, highlighted in red. When looking at the forwarding table under the "*show mpls forwarding*" command, the entries for SID index are installed, highlighted in green.

Note: While the working mapping client installs the SRMS policy in the forwarding table, the outgoing label is unlabeled as highlighted in green due to the lack of SR/LDP border router. Test Case 2 performs this testing further.

3.6.3. Test Cases 3 – Redundant Mapping Server

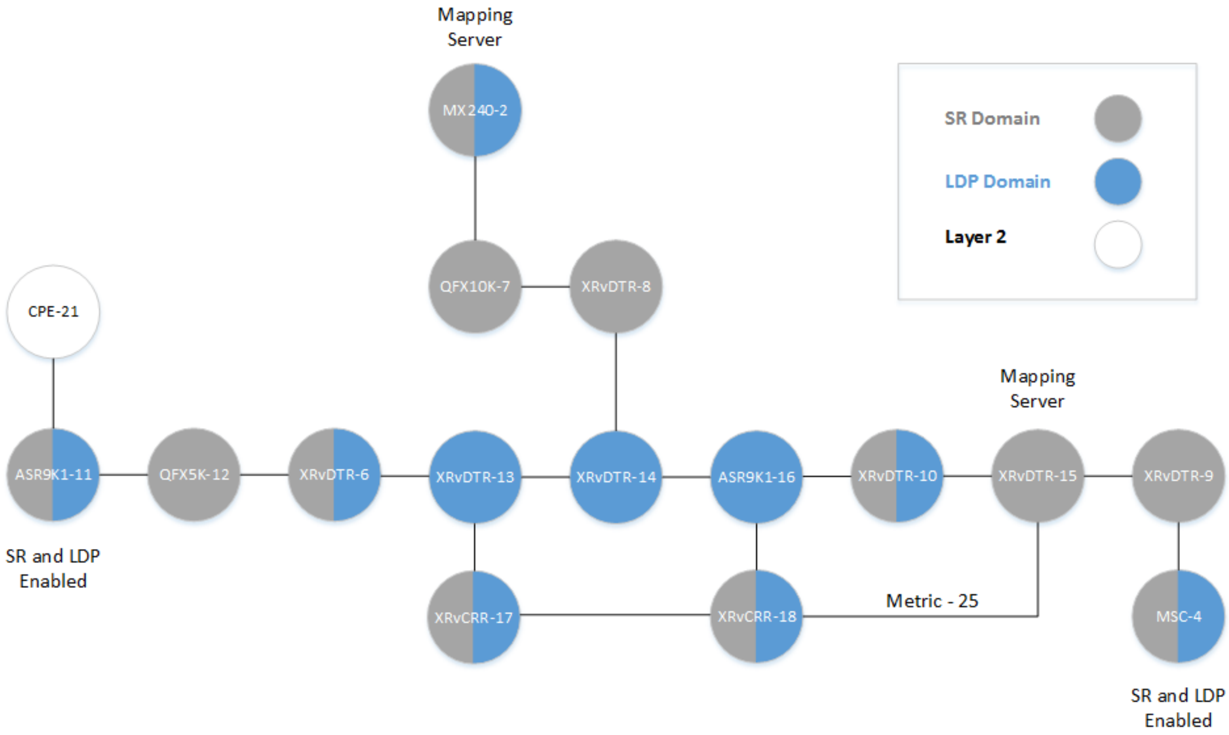
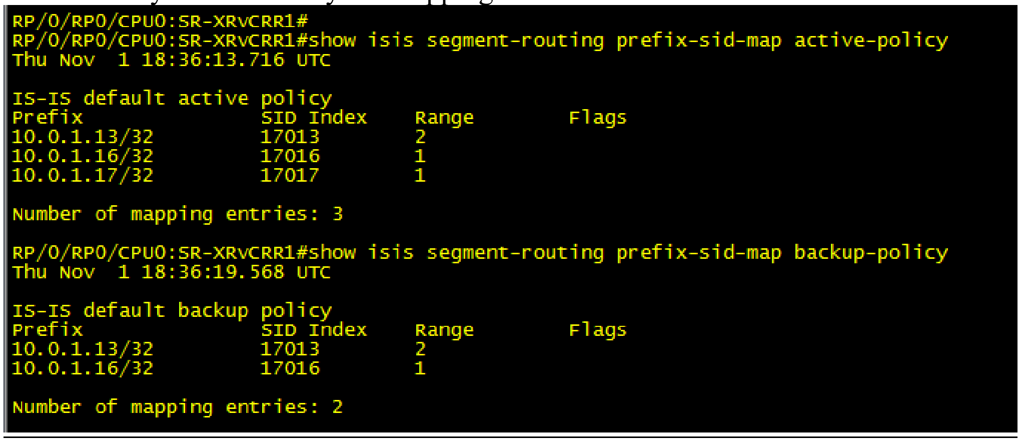


Figure 18 – Redundant Mapping Server Topology

1. All devices within SR domain are SR enabled only. Exceptions are ASR9K1-11 and MSC-4 where L2VPN terminates and the SR/LDP border routers.
2. All devices within LDP domain are LDP enabled only.
3. XRvDTR-15 is the original mapping server, the rest of the nodes are mapping clients.
4. MX240-2 is added as redundant mapping server.

Outcome with second mapping server:

All devices receive the prefix-to-SID advertisement from the second SRMS, in this case MX240-2. The redundant mapping server is configured with one less entry. Note figure below shows backup-policy with one less entry as received by the mapping client.



```
RP/0/RP0/CPU0:SR-XRvCRR1#
RP/0/RP0/CPU0:SR-XRvCRR1#show isis segment-routing prefix-sid-map active-policy
Thu Nov  1 18:36:13.716 UTC

IS-IS default active policy
Prefix              SID Index    Range        Flags
10.0.1.13/32        17013        2
10.0.1.16/32        17016        1
10.0.1.17/32        17017        1

Number of mapping entries: 3

RP/0/RP0/CPU0:SR-XRvCRR1#show isis segment-routing prefix-sid-map backup-policy
Thu Nov  1 18:36:19.568 UTC

IS-IS default backup policy
Prefix              SID Index    Range        Flags
10.0.1.13/32        17013        2
10.0.1.16/32        17016        1

Number of mapping entries: 2
```

3.6.4. Test Case 4 – Services over SR/LDP Domain

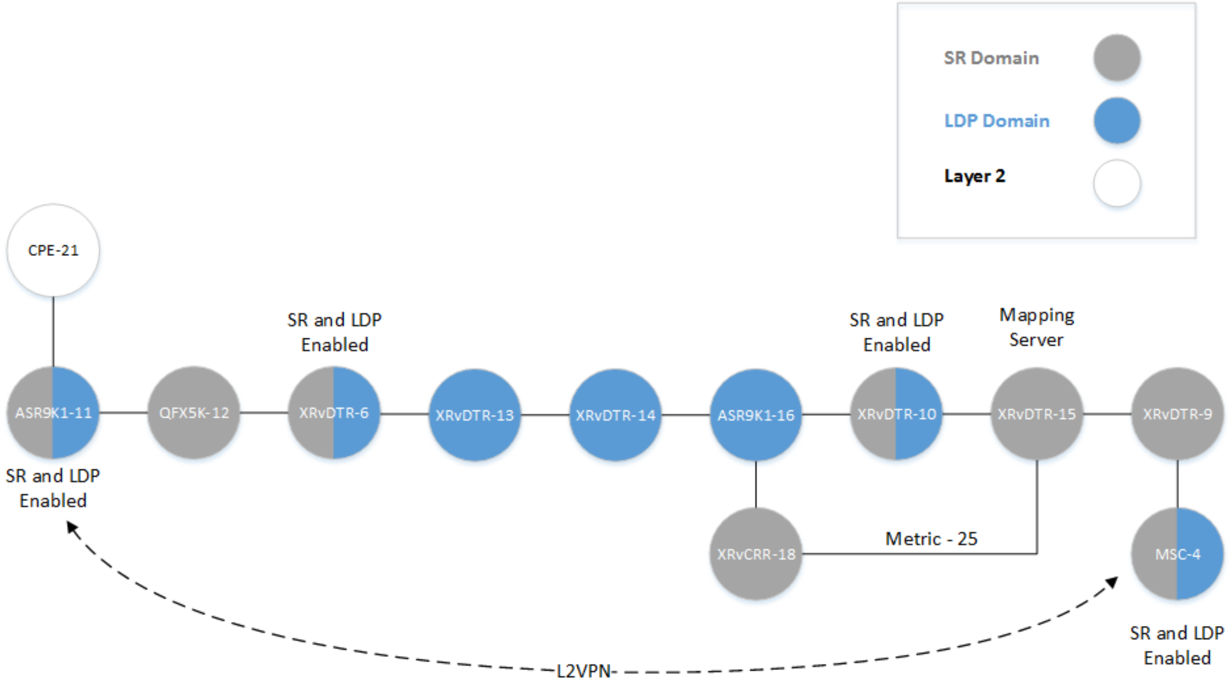


Figure 19 – L2VPN service over SR/LDP domain

1. All devices within SR domain are SR enabled only. Exceptions are ASR9K1-11 and MSC-4 where L2VPN terminates for targeted LDP session and the SR/LDP border routers.
2. All devices within LDP domain are LDP enabled only.

3. XRvDTR-15 is the mapping server, the rest of the nodes are mapping clients.
4. MX240-2 is added as redundant mapping server now shown in figure 19.

Outcome:

L2VPN service successfully traversed across the LDP/SR domain.

3.6.5. Assessment of SRMS and SR/LDP Interworking with multiple vendors

Cautionary Practices and Deployment

1. Prefix-to-SID-mapping range
    - Rather than configuring many prefix-sid-mapping entries for each node, it is easier to use the prefix, range, and start index command to represent a wide range of prefixes in a single entry. This works best if production loopback addresses are in contiguous fashion.
2. Any LDP only nodes must have all their direct SR neighbors to be an SR/LDP border router. Leaving out one node as SR/LDP could cause traffic loss should that node be the best IGP path after a primary failure.

Mapping Server and Client Risks

1. Risks of Conflicts and Overlapping prefix-to-SID-mappings
    - Forwarding loops can occur
    - Traffic blackholes
2. Traffic loss and service impact could occur if different vendor platforms interpret and perform conflict resolution differently as this could lead to inconsistent forwarding state across the network. (See Table 8 for vendor differences in accordance with IETF draft)
3. Vendor Z does appear to perform conflict resolutions for SID conflicts while it supports Prefix conflicts.
4. Troubleshooting commands are limited.
5. Using different SRGB values and ranges would require tracking of every node’s SRGB since a mapping server could advertise an index that is outside of the receiving mapping client’s SRGB range. Therefore it is better to use the common label space across three platforms.

Table 8 – Conflict Resolution Preference Rules Comparison

Node Platform X	Node Platform Y	Node Platform Z
Largest router-id (OSPF) or system-id (ISIS) is preferred	Largest router-id (OSPF) or system-id (ISIS) is preferred	
	Smallest area-id (OSPF) or level (ISIS) is preferred	
	IPv4 range is preferred over IPv6 range	IPv4 range is preferred
Smallest prefix length is preferred	Smallest prefix length is preferred	
Smallest IP address is preferred	Smallest IP address is preferred	Smallest IP address is preferred (IPv4 Only)
Smallest SID index is preferred	Smallest SID index is preferred	Smallest SID index is preferred
Smallest range is preferred	Smallest range is preferred	Smallest range is preferred
	First received range is preferred	

Mapping ranges conflict - https://tools.ietf.org/html/draft-ietf-spring-conflict-resolution-05

4. Design Considerations

4.1. MPLS MTU

While SR with IGP path forwarding with just the transport label do not impose additional overhead, it is the other mechanisms like Traffic Engineering or TI-LFA that should be taken into consideration when designing Maximum Transmission Unit (MTU) across the network. When using Traffic Engineering or TI-LFA, the MTU size could grow as the segment label stack increases.

With the MTU conditions of a baseline IP and layer 2 header with a single label, it will be of no if the network is configured with jumbo frames of 9000 or higher. For certain scenarios that require packets with higher MTU together with Traffic Engineering or TI-LFA, it will be necessary to evaluate the traffic path and the MTU across the network.

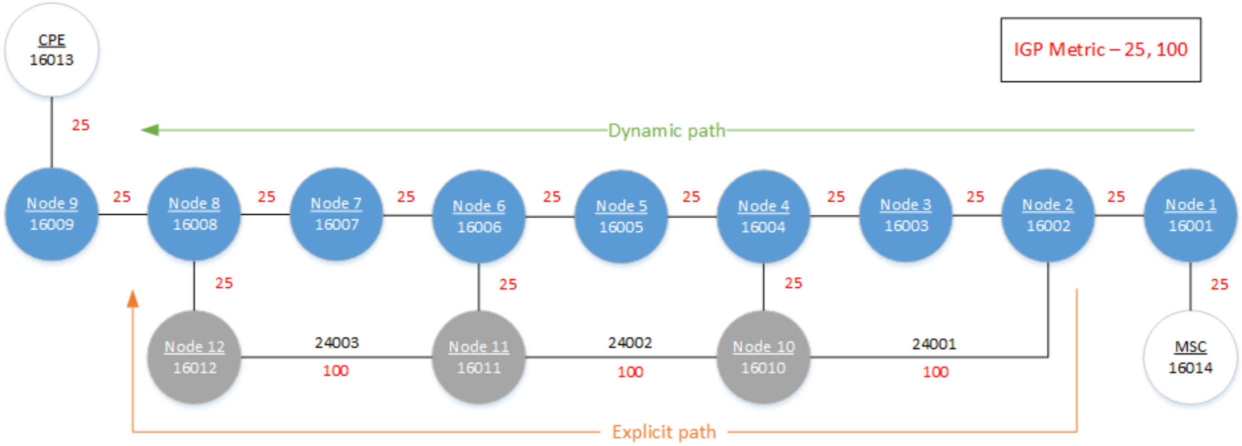


Figure 20 – Dynamic and Explicit SR Path

Based on the topology in figure 1, the following shows the multiple scenarios of how SR labels could impact MTU sizing. The scenarios below are built on the following assumptions where the payload is the typical customer service frame with C-VLAN tag.

Low MTU, dynamically forwarded

1522	4	4
Service Frame	Service Label	Transport Label 16013

As every SR node is aware of other SR node's unique SID, only the destination node SID 16013 is imposed on the label stack. This scenario has no impact to Charter's MTU restrictions

Low MTU, explicitly forwarded

1522	4	4	4	4	4	4	4
Service Frame	Service Label	Transport Label 6 16013	Transport Label 5 24003	Transport Label 4 16011	Transport Label 3 24002	Transport Label 2 16010	Transport Label 1 24001

Steering this frame along the higher metric path calls for an increased segment list of six labels. Adjacency SIDs 24001, 24002, and 24003 were added to ensure traffic flows through nodes 11 through 12. Despite the growing label stack, the low MTU size of the service frame keeps the entire Ethernet frame well under 2000. This scenario has little to no impact to Charter's MTU restrictions.

High MTU, dynamically forwarded

9216	4	4
Service Frame	Service Label	Transport Label 16013

Similarly to the low MTU size, dynamically forwarded scenario, only a single label is imposed for transport. While in this scenario the service frame is significantly larger in MTU, adding a single label still has little impact as it simply replaces a previously used transport label such as LDP or RSVP.

High MTU, explicitly forwarded

9216	4	4	4	4	4	4	4
Service Frame	Service Label	Transport Label 6 16013	Transport Label 5 24003	Transport Label 4 16011	Transport Label 3 24002	Transport Label 2 16010	Transport Label 1 24001

In a scenario where a high MTU service frame is steered, the label stack can potentially grow. This adds to the MTU size of the entire Ethernet frame. Adjacency SIDs 24001, 24002 and 24003 were added to ensure traffic flows from node 10 through 12, thus inflating the label stack.

Recommendations:
For best practice of MTU size considerations in SR-TE, it is recommended to account for the needed headroom against the maximum label stack and MTU restrictions across the network when setting up a service end to end. This maximum label stack or MSD is defined by each vendor platform. Please see MSD section for more information.

Conclusion

The intent was to review the option for SR-TE as an alternate traffic engineering solution for business services. In the first phase of planning for an SR POC Lab, an audit of Charter's network was performed to better account for all types of major hardware and linecards used in production so that the POC is built according to deployment. During this audit, it was discovered that 50% of two vendor platforms' hardware/linecard would require replacement to support a full adoption of SR. However, with the successful testing of utilizing mapping servers in a mix of SR and LDP domains within a network, the impact of hardware replacement can be lessened. Caveat to the previous statement would be if the PE falls into the 50% replacement of hardware, deeming it difficult to depoly SR since the PE would be used as the SR headend node.

During the testing of SR, there were multiple instances where some features were not supported prompting several upgrades. Careful planning of what is required for a successful SR deployment is imperative. Even the smallest detail or a feature that didn't seem relevant at that point but is crucial to enabling the bigger feature can be easily overlooked. To date, there are still some vendors that may not have certain features available to align with another vendor that may be capable. SR may have been introduced for a few years, but some vendors are still catching up to the latest spec making it difficult to harmonize all vendors in a single deployment for the same feature. This is particularly important to providers that use multiple vendors in their network.

It is observed that some features could use a controller to alleviate some of the operational work such as visibility of MSD across the network and assist in reducing user errors. The major assist in having a controller would be path computation for optimal routing and constraints to meet the customer SLA.

Overall view of SR appears to be feasible given the multiple vendor platforms for baseline SR deployment. Though CAPEX would be seemingly high at initial SR rollout due to hardware support, the trade-offs are operational expense with less complexity to maintain the network with traffic engineering, lesser penalty fees for missing MTTR, and scalability to use a controller for full visibility of a network which includes efficient capacity planning, trend reporting, telemetry, and change control modeling. This also introduces opportunities for automation. An incremental or partial deployment of SR, or even a greenfield market, would be a better and more cost effective approach by Charter so as not to dive into a full investment while continuing to use existing transport methodology until SR is fully baked-in by all vendors. Charter is still pursuing further in-depth testing, particularly with the use of controllers and a higher subset of different interdomain networks operating together.

Abbreviations

BMI	Base MPLS Imposition
CER	commercial edge router
CTBH	cell tower backhaul
IGP	Interior Gateway Protocol
LIB	label information base
LSD	label switching database
MPLS	multi-protocol label switching
MSC	Mobile Switching Center
MSD	Maximum SID Depth
MTTR	mean time to repair
MTU	maximum transmission unit
PCC	path computation client
PCE	path computation element
PCEP	Path Computation Element Protocol
PE	provider edge
POC	proof of concept
RSVP-TE	resource reservation protocol – traffic engineering
SLA	service level agreement
SID	segment identifier
SR	segment routing
SRGB	segment routing global block
SRMS	Segment Routing Mapping Server
SR-TE	Segment Routing Traffic Engineering
TI-LFA	topology independent – loop free alternate

Bibliography & References

PCEP Extensions for Segment Routing - https://tools.ietf.org/html/draft-ietf-pce-segment-routing-16

Segment Routing MPLS Conflict Resolution - https://tools.ietf.org/html/draft-ietf-spring-conflict-resolution-05

LDP Specification - https://tools.ietf.org/html/rfc5036

Segment Routing, Part 1; Clarence Filsfils, Kris Michielsen, Ketan Talaulikar