

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 23, 2014

C. Filsfils, Ed.
S. Previdi, Ed.
A. Bashandy
Cisco Systems, Inc.
B. Decraene
S. Litkowski
Orange
M. Horneffer
Deutsche Telekom
I. Milojevic
Telekom Srbija
R. Shakir
British Telecom
S. Ytti
TDC Oy
W. Henderickx
Alcatel-Lucent
J. Tantsura
Ericsson
E. Crabbe
Google, Inc.
October 20, 2013

Segment Routing interoperability with LDP
draft-filsfils-spring-segment-routing-ldp-interop-00

Abstract

A Segment Routing (SR) node steers a packet through a controlled set of instructions, called segments, by prepending the packet with an SR header. A segment can represent any instruction, topological or service-based. SR allows to enforce a flow through any topological path and service chain while maintaining per-flow state only at the ingress node to the SR domain.

The Segment Routing architecture can be directly applied to the MPLS data plane with no change in the forwarding plane. This drafts describes how Segment Routing operates in a network where LDP is deployed and in the case where SR-capable and non-SR-capable nodes coexist.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC 2119](#) [[RFC2119](#)].

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on April 23, 2014.

Copyright Notice

Copyright (c) 2013 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1.	Introduction	4
2.	SR/LDP Ship-in-the-night coexistence	4
2.1.	MPLS2MPLS co-existence	6
2.2.	IP2MPLS co-existence	7
3.	Migration from LDP to SR	7
4.	SR and LDP Interworking	8
4.1.	LDP to SR	8
4.2.	SR to LDP	9
5.	Leveraging SR benefits for LDP-based traffic	10
5.1.	Eliminating Directed LDP Session	12
5.2.	Guaranteed FRR coverage	12
6.	Inter-AS Option C, Carrier's Carrier and Seamless MPLS	14
7.	IANA Considerations	14
8.	Manageability Considerations	14
9.	Security Considerations	14
10.	Acknowledgements	14
11.	References	15
11.1.	Normative References	15
11.2.	Informative References	15
	Authors' Addresses	15

1. Introduction

Segment Routing, as described in [\[I-D.filsfils-rtgwg-segment-routing\]](#), can be used on top of the MPLS data plane without any modification as described in [\[draft-filsfils-rtgwg-segment-routing-mpls-00\]](#).

Segment Routing control plane can co-exist with current label distribution protocols such as LDP.

This draft outlines the mechanisms through which SR provides interoperability with LDP in cases where a mix of SR-capable and non-SR-capable routers co-exist within the same network.

The first section describes the co-existence of SR with other MPLS Control Plane. The second section documents a method to migrate from LDP to SR-based MPLS tunneling. The third section documents the interworking of LDP and SR in the case of non-homogenous deployment. The fourth section describes how a partial SR deployment can be used to provide SR benefits to LDP-based traffic. The fifth section describes a possible application of SR in the context of inter-domain MPLS use-cases.

2. SR/LDP Ship-in-the-night coexistence

We call "MPLS Control Plane Client (MCC)" any control plane protocol installing forwarding entries in the MPLS data plane. SR, LDP, RSVP-TE, BGP 3107, VPNv4, etc. are examples of MCCs.

An MCC, operating at node N, must ensure that the incoming label it installs in the MPLS data plane of Node N has been uniquely allocated to himself.

Thanks to the defined segment allocation rule and specifically the notion of the SRGB, SR can co-exist with any other MCC.

This is clearly the case for the adjacency segment: it is a local label allocated by the label manager, as for any MCC.

This is clearly the case for the prefix segment: the label manager allocates the SRGB set of labels to the SR MCC client and the operator ensures the unique allocation of each global prefix segment/label within the allocated SRGB set.

Note that this static label allocation capability of the label manager has been existing for many years across several vendors and hence is not new. Furthermore, note that the label-manager ability

to statically allocate a range of labels to a specific application is not new either. This is required for MPLS-TP operation. In this case, the range is reserved by the label manager and it is the MPLS-TP NMS (acting as an MCC) that ensures the unique allocation of any label within the allocated range and the creation of the related MPLS forwarding entry.

Let us illustrate an example of ship-in-the-night (SIN) coexistence.

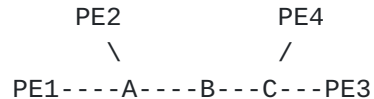


Figure 1: SIN coexistence

The EVEN VPN service is supported by PE2 and PE4 while the ODD VPN service is supported by PE1 and PE3. The operator wants to tunnel the ODD service via LDP and the EVEN service via SR.

This can be achieved in the following manner:

The operator configures PE1, PE2, PE3, PE4 with respective loopbacks 192.0.2.201/32, 192.0.2.202/32, 192.0.2.203/32, 192.0.2.204/32. These PE's advertised their VPN routes with next-hop set on their respective loopback address.

The operator configures A, B, C with respective loopbacks 192.0.2.1/32, 192.0.2.2/32, 192.0.2.3/32.

The operator configures PE2, A, B, C and PE4 with SRGB {100, 300}.

The operator attaches the respective Node-SIDs 202, 101, 102, 103 and 204 to the loopbacks of nodes PE2, A, B, C and PE4. The Node-SID's are configured to request penultimate-hop-popping.

PE1, A, B, C and PE3 are LDP capable.

PE1 and PE3 are not SR capable.

PE3 sends an ODD VPN route to PE1 with next-hop 192.0.2.203 and VPN label 10001.

From an LDP viewpoint: PE1 received an LDP label binding (1037) for FEC 192.0.2.203/32 from its nhop A. A received an LDP label binding (2048) for that FEC from its nhop B. B received an LDP label binding (3059) for that FEC from its nhop C. C received implicit-null LDP binding from its next-hop PE3.

As a result, PE1 sends its traffic to the ODD service route

advertised by PE3 to next-hop A with two labels: the top label is 1037 and the bottom label is 10001. A swaps 1037 with 2048 and forwards to B. B swaps 2048 with 3059 and forwards to C. C pops 3059 and forwards to PE3.

PE4 sends an EVEN VPN route to PE2 with next-hop 192.0.2.204 and VPN label 10002.

From an SR viewpoint: PE1 maps the IGP route 192.0.2.204/32 onto Node-SID 204; A swaps 204 with 204 and forwards to B; B swaps 204 with 204 and forwards to C; C pops 204 and forwards to PE4.

As a result, PE2 sends its traffic to the VPN service route advertised by PE4 to next-hop A with two labels: the top label is 204 and the bottom label is 10002. A swaps 204 with 204 and forwards to B. B swaps 204 with 204 and forwards to C. C pops 204 and forwards to PE4.

The two modes of MPLS tunneling co-exist.

The ODD service is tunneled from PE1 to PE3 through a continuous LDP LSP traversing A, B and C.

The EVEN service is tunneled from PE2 to PE4 through a continuous SR node segment traversing A, B and C.

2.1. MPLS2MPLS co-existence

We want to highlight that several MPLS2MPLS entries can be installed in the data plane for the same prefix.

Let us examine A's MPLS forwarding table as an example:

Incoming label: 1037

- outgoing label: 2048
- outgoing nhop: B
- Note: this entry is programmed by LDP for 192.0.2.203/32

Incoming label: 203

- outgoing label: 203
- outgoing nhop: B
- Note: this entry is programmed by SR for 192.0.2.203/32

These two entries can co-exist because their incoming label is unique. The uniqueness is guaranteed by the label manager allocation rules.

The same applies for the MPLS2IP forwarding entries.

[2.2.](#) IP2MPLS co-existence

By default, we propose that if both LDP and SR propose an IP2MPLS entry for the same IP prefix, then the LDP route is selected.

A local policy on a router **MUST** allow to prefer the SR-provided IP2MPLS entry.

[3.](#) Migration from LDP to SR

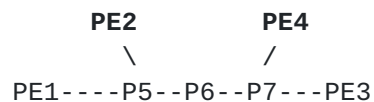


Figure 2: Migration

Several migration techniques are possible. We describe one technique inspired by the commonly used method to migrate from one IGP to another.

T0: all the routers run LDP. Any service is tunneled from an ingress PE to an egress PE over a continuous LDP LSP.

T1: all the routers are upgraded to SR. They are configured with the SRGB range (100, 200). PE1, PE2, PE3, PE4, P5, P6 and P7 are respectively configured with the node segments 101, 102, 103, 104, 105, 106 and 107 (attached to their service-recursing loopback).

At this time, the service traffic is still tunneled over LDP LSP. For example, PE1 has an SR node segment to PE3 and an LDP LSP to PE3 but by default, as seen earlier, the LDP IP2MPLS encapsulation is preferred.

T2: the operator enables the local policy at PE1 to prefer SR IP2MPLS encapsulation over LDP IP2MPLS.

The service from PE1 to any other PE is now riding over SR. All other service traffic is still transported over LDP LSP.

T3: gradually, the operator enables the preference for SR IP2MPLS encapsulation across all the edge routers.

All the service traffic is now transported over SR. LDP is still operational and services could be reverted to LDP.

T4: LDP is unconfigured from all routers.

4. SR and LDP Interworking

In this section, we analyze a use-case where SR is available in one part of the network and LDP is available in another part. We describe how a continuous MPLS tunnel can be built throughout the network.

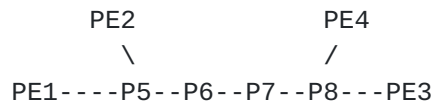


Figure 3: SR and LDP Interworking

Let us analyze the following example:

P6, P7, P8, PE4 and PE3 are LDP capable.

PE1, PE2, P5 and P6 are SR capable. PE1, PE2, P5 and P6 are configured with SRGB (100, 200) and respectively with node segments 101, 102, 105 and 106.

A service flow must be tunneled from PE1 to PE3 over a continuous MPLS tunnel encapsulation. We need SR and LDP to interwork.

4.1. LDP to SR

In this section, we analyze a right-to-left traffic flow.

PE3 has learned a service route whose nhop is PE1. PE3 has an LDP label binding from the nhop P8 for the FEC "PE1". Hence PE3 sends its service packet to P8 as per classic LDP behavior.

P8 has an LDP label binding from its nhop P7 for the FEC "PE1" and hence P8 forwards to P7 as per classic LDP behavior.

P7 has an LDP label binding from its nhop P6 for the FEC "PE1" and hence P7 forwards to P6 as per classic LDP behavior.

P6 does not have an LDP binding from its nhop P5 for the FEC "PE1". However P6 has an SR node segment to the IGP route "PE1". Hence, P6 forwards the packet to P5 and swaps its local LDP-label for FEC "PE1" by the equivalent node segment (i.e. 101).

P5 pops 101 (assuming PE1 advertised its node segment 101 with the penultimate-pop flag set) and forwards to PE1.

PE1 receives the tunneled packet and processes the service label.

The end-to-end MPLS tunnel is built from an LDP LSP from PE3 to P6

and the related node segment from P6 to PE1.

4.2. SR to LDP

In this section, we analyze the left-to-right traffic flow.

We assume that the operator configures P5 to act as a Segment Routing Mapping Server (SRMS) and advertise the following mappings: (P7, 107), (P8, 108), (PE3, 103) and (PE4, 104).

These mappings are advertised as Remote-Bundle SID with Flag TBD.

The mappings advertised by an SR mapping server result from local policy information configured by the operator. IF PE3 had been SR capable, the operator would have configured PE3 with node segment 103. Instead, as PE3 is not SR capable, the operator configures that policy at the SRMS and it is the latter which advertises the mapping. Multiple SRMS servers can be provisioned in a network for redundancy.

The mapping server advertisements are only understood by the SR capable routers. The SR capable routers install the related node segments in the MPLS data plane exactly like if the node segments had been advertised by the nodes themselves.

For example, PE1 installs the node segment 103 with nhop P5 exactly as if PE3 had advertised node segment 103.

PE1 has a service route whose nhop is PE3. PE1 has a node segment for that IGP route: 103 with nhop P5. Hence PE1 sends its service packet to P5 with two labels: the bottom label is the service label and the top label is 103.

P5 swaps 103 for 103 and forwards to P6.

P6's next-hop for the IGP route "PE3" is not SR capable (P7 does not advertise the SR capability). However, P6 has an LDP label binding from that next-hop for the same FEC (e.g. LDP label 1037). Hence, P6 swaps 103 for 1037 and forwards to P7.

P7 swaps this label with the LDP-label received from P8 and forwards to P8.

P8 pops the LDP label and forwards to PE3.

PE3 receives the tunneled packet and processes the service label.

The end-to-end MPLS tunnel is built from an SR node segment from PE1 to P6 and an LDP LSP from P6 to PE3.

Note: contrary to Prefix-SID, SR mappings do not allow for Penultimate Hop Popping. In the previous example, P6 requires the presence of the segment 103 such as to map it to the LDP label 1037. For that reason, the P flag available in the Prefix-SID is not available in the Remote-Bundle SID.

5. Leveraging SR benefits for LDP-based traffic

SR can be deployed such as to enhance LDP transport. The SR deployment can be limited to the network region where the SR benefits are most desired.

In Figure 4, let us assume:

All link costs are 10 except FG which is 30.

All routers are LDP capable.

X, Y and Z are PE's participating to an important service S.

The operator requires 50msec link-based FRR for service S.

A, B, C, D, E, F and G are SR capable.

X, Y, Z are not SR capable, e.g. as part of a staged migration from LDP to SR, the operator deploys SR first in a sub-part of the network and then everywhere.

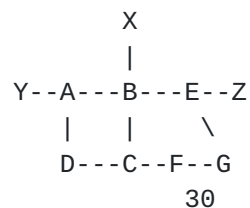


Figure 4: Leveraging SR benefits for LDP-based-traffic

The operator would like to resolve the following issues:

To protect the link BA along the shortest-path of the important flow XY, B requires an RLFA repair tunnel to D and hence a directed LDP session from B to D. The operator does not like these dynamically established multi-hop LDP sessions and would seek to eliminate them.

There is no LFA/RLFA solution to protect the link BE along the shortest path of the important flow XZ. The operator wants a guaranteed link-based FRR solution.

The operator can meet these objectives by deploying SR only on A, B, C, D, E and F:

The operator configures A, B, C, D, E, F and G with SRGB (100, 200) and respective node segments 101, 102, 103, 104, 105, 106 and 107.

The operator configures D as an SR Mapping Server with the following policy mapping: (X, 201), (Y, 202), (Z, 203}.

Each SR node automatically advertises local adjacency segment for its IGP adjacencies. Specifically, F advertises adjacency segment 9001 for its adjacency FG.

A, B, C, D, E, F and G keep their LDP capability and hence the flows XY and XZ are transported over end-to-end LDP LSP's.

For example, LDP at B installs the following MPLS data plane entries:

Incoming label: local LDB label bound by B for FEC Y

Outgoing label: LDP label bound by A for FEC Y

Outgoing nhop: A

Incoming label: local LDB label bound by B for FEC Z

Outgoing label: LDP label bound by E for FEC Z

Outgoing nhop: E

The novelty comes from how the backup chains are computed for these LDP-based entries. While LDP labels are used for the primary nhop and outgoing labels, SR information is used for the FRR construction. In steady state, the traffic is transported over LDP LSP. In transient FRR state, the traffic is backup thanks to the SR enhanced capabilities.

This helps meet the requirements of the operator:

Eliminate directed LDP session.

Guaranteed FRR coverage.

Keep the traffic over LDP LSP in steady state.

Partial SR deployment only where needed.

5.1. Eliminating Directed LDP Session

B's MPLS entry to Y becomes:

- Incoming label: local LDB label bound by B for FEC Y
- Outgoing label: LDP label bound by A for FEC Y
- Backup outgoing label: SR node segment for Y {202}
- Outgoing nhop: A
- Backup nhop: repair tunnel: node segment to D {104}
- with outgoing nhop: C

In steady-state, X sends its Y-destined traffic to B with a top label which is the LDP label bound by B for FEC Y. B swaps that top label for the LDP label bound by A for FEC Y and forwards to A. A pops the LDP label and forwards to Y.

Upon failure of the link BA, B swaps the incoming top-label with the node segment for Y (202) and sends the packet onto a repair tunnel to D (node segment 104). Thus, B sends the packet to C with the label stack {104, 202}. C pops the node segment 104 and forwards to D. D swaps 202 for 202 and forwards to A. A's nhop to Y is not SR capable and hence A swaps the incoming node segment 202 to the LDP label announced by its next-hop (in this case, implicit null).

After IGP convergence, B's MPLS entry to Y will become:

- Incoming label: local LDB label bound by B for FEC Y
- Outgoing label: LDP label bound by C for FEC Y
- Outgoing nhop: C

And the traffic XY travels again over the LDP LSP.

Conclusion: the operator has eliminated its first problem: directed LDP sessions are no longer required and the steady-state traffic is still transported over LDP. The SR deployment is confined to the area where these benefits are required.

5.2. Guaranteed FRR coverage

B's MPLS entry to Z becomes:

- Incoming label: local LDB label bound by B for FEC Z
Outgoing label: LDP label bound by E for FEC Z
Backup outgoing label: SR node segment for Z {203}
Outgoing nhop: E
Backup nhop: repair tunnel to G: {106, 9001}

G is reachable from B via the combination of a node segment to F {106} and an adjacency segment FG {9001}

Note that {106, 107} would have equally work. Indeed, in many case, P's shortest path to Q is over the link PQ. The adjacency segment from P to Q is required only in very rare topologies where the shortest-path from P to Q is not via the link PQ.

In steady-state, X sends its Z-destined traffic to B with a top label which is the LDP label bound by B for FEC Z. B swaps that top label for the LDP label bound by E for FEC Z and forwards to E. E pops the LDP label and forwards to Z.

Upon failure of the link BE, B swaps the incoming top-label with the node segment for Z (203) and sends the packet onto a repair tunnel to G (node segment 106 followed by adjacency segment 9001). Thus, B sends the packet to C with the label stack {106, 9001, 203}. C pops the node segment 106 and forwards to F. F pops the adjacency segment 9001 and forwards to G. G swaps 203 for 203 and forwards to E. E's nhop to Z is not SR capable and hence E swaps the incoming node segment 203 for the LDP label announced by its next-hop (in this case, implicit null).

After IGP convergence, B's MPLS entry to Z will become:

- Incoming label: local LDB label bound by B for FEC Z
Outgoing label: LDP label bound by C for FEC Z
Outgoing nhop: C

And the traffic XZ travels again over the LDP LSP.

Conclusion: the operator has eliminated its second problem: guaranteed FRR coverage is provided. The steady-state traffic is still transported over LDP. The SR deployment is confined to the area where these benefits are required.

6. Inter-AS Option C, Carrier's Carrier and Seamless MPLS

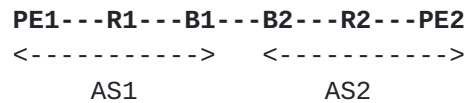


Figure 5: Inter-AS Option C

In Inter-AS Option C [[RFC4364](#)], B2 advertises to B1 a BGP3107 route for PE2 and B1 reflects it to its internal peers, such as PE1. PE1 learns from a service route reflector a service route whose nhop is PE2. PE1 resolves that service route on the BGP3107 route to PE2. That BGP3107 route to PE2 is itself resolved on the AS1 IGP route to B1.

If AS1 operates SR, then the tunnel from PE1 to B1 is provided by the node segment from PE1 to B1.

PE1 sends a service packet with three labels: the top one is the node segment to B1, the next-one is the BGP3107 label provided by B1 for the route "PE2" and the bottom one is the service label allocated by PE2.

The same straightforward SR applicability is derived for CsC and Seamless MPLS ([\[I-D.ietf-mpls-seamless-mpls\]](#)).

7. IANA Considerations

TBD

8. Manageability Considerations

TBD

9. Security Considerations

TBD

10. Acknowledgements

We would like to thank Pierre Francois and Ruediger Geib for their contribution to the content of this document.

11. References

11.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", [RFC 4364](#), February 2006.

11.2. Informative References

- [I-D.filsfils-rtgwg-segment-routing]
Filsfils, C., Previdi, S., Bashandy, A., Decraene, B., Litkowski, S., Horneffer, M., Milojevic, I., Shakir, R., Ytti, S., Henderickx, W., Tantsura, J., and E. Crabbe, "Segment Routing Architecture", [draft-filsfils-rtgwg-segment-routing-00](#) (work in progress), June 2013.
- [I-D.ietf-mpls-seamless-mpls]
Leymann, N., Decraene, B., Filsfils, C., Konstantynowicz, M., and D. Steinberg, "Seamless MPLS Architecture", [draft-ietf-mpls-seamless-mpls-04](#) (work in progress), July 2013.
- [[draft-filsfils-rtgwg-segment-routing-mpls-00](#)]
Filsfils, C. and S. Previdi, "Segment Routing with MPLS data plane", October 2013.

Authors' Addresses

Clarence Filsfils (editor)
Cisco Systems, Inc.
Brussels,
BE

Email: cfilsfil@cisco.com

Stefano Previdi (editor)
Cisco Systems, Inc.
Via Del Serafico, 200
Rome 00142
Italy

Email: sprevidi@cisco.com

Ahmed Bashandy
Cisco Systems, Inc.
170, West Tasman Drive
San Jose, CA 95134
US

Email: bashandy@cisco.com

Bruno Decraene
Orange
FR

Email: bruno.decraene@orange.com

Stephane Litkowski
Orange
FR

Email: stephane.litkowski@orange.com

Martin Horneffer
Deutsche Telekom
Hammer Str. 216-226
Muenster 48153
DE

Email: Martin.Horneffer@telekom.de

Igor Milojevic
Telekom Srbija
Takovska 2
Belgrade
RS

Email: igormilojevic@telekom.rs

Rob Shakir
British Telecom
London
UK

Email: rob.shakir@bt.com

Saku Ytti
TDC Oy
Mechelininkatu 1a
TDC 00094
FI

Email: saku@ytti.fi

Wim Henderickx
Alcatel-Lucent
Copernicuslaan 50
Antwerp 2018
BE

Email: wim.henderickx@alcatel-lucent.com

Jeff Tantsura
Ericsson
300 Holger Way
San Jose, CA 95134
US

Email: Jeff.Tantsura@ericsson.com

Edward Crabbe
Google, Inc.
1600 Amphitheatre Parkway
Mountain View, CA 94043
US

Email: edc@google.com

