Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 24, 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 21, 2013

Segment Routing Architecture
draft-filsfils-rtgwg-segment-routing-01

Abstract

Segment Routing (SR) leverages the source routing paradigm. A 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. A segment can have a local semantic to an SR node or global within an SR domain. 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 dataplane with no change on the forwarding plane. IGP-based segments require minor extension to the existing link-state routing protocols. Segment Routing can also be applied to IPv6 with a new type of routing extension header.

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 24, 2014.

Copyright Notice

Table of Contents

1. Introduction

In this section, we illustrate the key properties of the SR architecture, introduce the companion documents to this note and relate SR to the MPLS and IPv6 architectures.

Section 2 defines the SR abstract routing model. Section 3 defines the IGP-based segments. Section 4 defines the Service Segments.

1.1. Illustration

In the context of Figure 1 where all the links have the same IGP cost, let us assume that a packet P enters the SR domain at an ingress edge router I and that the operator requests the following requirements for packet P:

The local service S offered by node B must be applied to packet P.

The links AB and CE cannot be used to transport the packet P.

Any node N along the journey of the packet should be able to determine where the packet P entered the SR domain and where it will exit. The intermediate node should be able to determine the paths from the ingress edge router to itself, and from itsef to the egress edge router.

Per-flow State for packet P should only be created at the ingress edge router.

State for packet P can only be created at the ingress edge router.

The operator can forbid, for security reasons, anyone outside the operator domain to exploit its intra-domain SR capabilities.

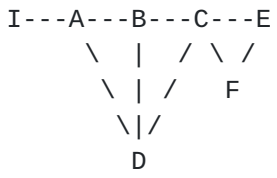

Figure 1: An illustration of SR properties

All these properties may be realized by instructing the ingress SR edge router I to push the following abstract SR header on the packet P.

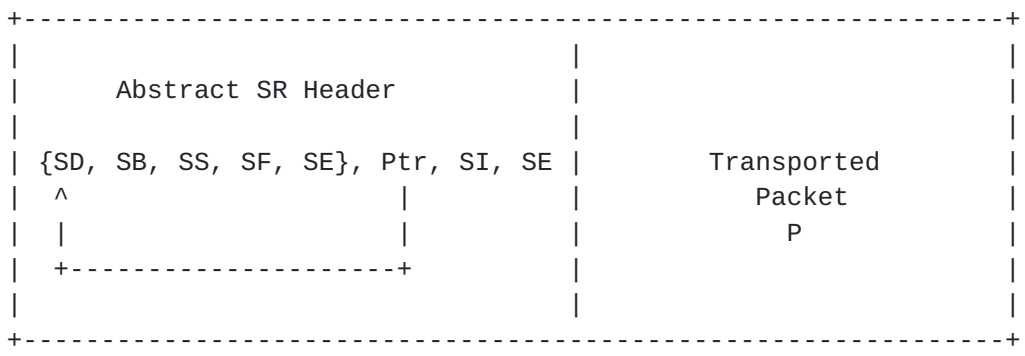

Figure 2: Packet P at node I

The abstract SR header contains a source route encoded as a list of segments {SD, SB, SS, SF, SE}, a pointer (Ptr) and the identification of the ingress and egress SR edge routers (segments SI and SE).

A segment is a 32-bit identification either for a topological instruction or a service instruction. A segment can either be global or local. The instruction associated with a global segment is recognized and executed by any SR-capable node in the domain. The instruction associated with a local segment is only supported by the specific node that originates it.

Let us assume some ISIS/OSPF extensions to define a "Node Segment" as a global instruction within the IGP domain to forward a packet along the shortest path to the specified node. Let us further assume that within the SR domain illustrated in Figure 1, segments SI, SD, SB, SE and SF respectively identify IGP node segments to I, D, B, E and F.

Let us assume that node B identifies its local service S with local segment SS.

With all of this in mind, let us describe the journey of the packet P.

The packet P reaches the ingress SR edge router. I pushes the SR header illustrated in Figure 2 and sets the pointer to the first segment of the list (SD).

SD is an instruction recognized by all the nodes in the SR domain which causes the packet to be forwarded along the shortest path to D.

Once at D, the pointer is incremented and the next segment is executed (SB).

SB is an instruction recognized by all the nodes in the SR domain which causes the packet to be forwarded along the shortest path to B.

Once at B, the pointer is incremented and the next segment is executed (SS).

SS is an instruction only recognized by node B which causes the packet to receive service S.

Once the service applied, the next segment is executed (SF) which causes the packet to be forwarded along the shortest path to F.

Once at F, the pointer is incremented and the next segment is executed (SE).

SE is an instruction recognized by all the nodes in the SR domain which causes the packet to be forwarded along the shortest path to E.

E then removes the SR header and the packet continues its journey outside the SR domain.

All of the requirements are met.

First, the packet P has not used links AB and CE: the shortest-path from I to D is I-A-D, the shortest-path from D to B is D-B, the shortest-path from B to F is B-C-F and the shortest-path from F to E is F-E, hence the packet path through the SR domain is I-A-D-B-C-F-E and the links AB and CE have been avoided.

Second, the service S supported by B has been applied on packet P.

Third, any node along the packet path is able to identify the service and topological journey of the packet within the SR domain. For example, node C receives the packet illustrated in Figure 3 and hence is able to infer where the packet entered the SR domain (SI), how it got up to itself {SD, SB, SS, SE}, where it will exit the SR domain (SE) and how it will do so {SF, SE}.

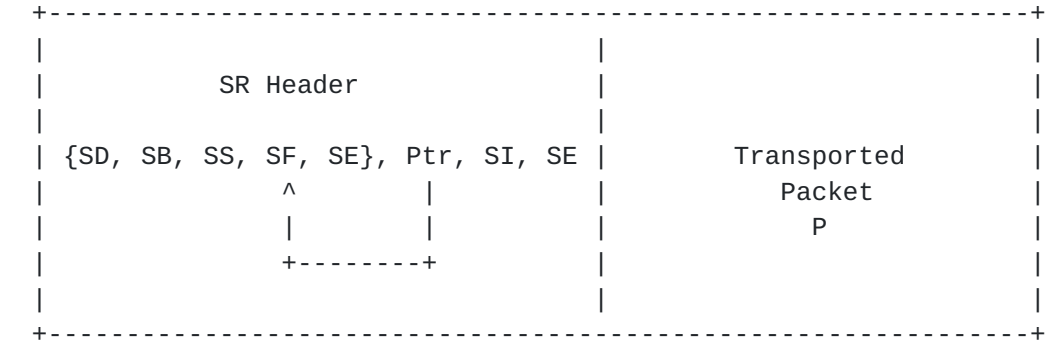

Figure 3: Packet P at node C

Fourth, only node I maintains per-flow state for packet P. The entire

program of topological and service instructions to be executed by the SR domain on packet P is encoded by the ingress edge router I in the SR header in the form of a list of segments where each segment identifies a specific instruction. No further per-flow state is required along the packet path. The per-flow state is in the SR header and travels with the packet. Intermediate nodes only hold states related to the IGP global node segments and the local IGP adjacency segments. These segments are not per-flow specific and hence scale very well. Typically, an intermediate node would maintain in the order of 100's to 1000's global node segments and in the order of 10's to 100 of local adjacency segments. Typically the SR IGP forwarding table is expected to be much less than 10000 entries.

Fifth, the SR header is inserted at the entrance to the domain and removed at the exit of the operator domain. For security reasons, the operator can forbid anyone outside its domain to use its intra-domain SR capability.

1.2. Terminology

The following terminology is defined:

Term	Definition
Segment	A segment that identifies an instruction
SID	A 32-bit identification for a segment
Segment List	Ordered list of segments encoding the topological and service source route of the packet
Active Segment	The segment that MUST be used by the receiving router to process the packet. It is identified by the pointer
SR-Pointer or pointer	In the SR header, it indicates the active segment in the segment list
Global Segment	The related instruction is supported by all the SR-capable nodes in the local domain
SRGB	SR Global Block: the set of global segments in the local SR domain
Local Segment	The related instruction is supported only by the node originating it

IGP Segment or IGP SID	The generic names for a segment attached to a piece of information advertised by a link-state IGP, e.g. an IGP prefix or an IGP adjacency
IGP-Prefix Segment or Prefix-SID	An IGP-Prefix Segment is an IGP segment attached to an IGP prefix. An IGP-Prefix Segment is always global within the SR/IGP domain and identifies the ECMP-aware shortest-path computed by the IGP to the related prefix. The Prefix-SID is the SID of the IGP-Prefix Segment
IGP-Node Segment or Node Segment or Node-SID	An IGP-Node Segment is a an IGP-Prefix Segment which identifies a specific router (e.g. a loopback). The terms "Node Segment" or Node-SID" are often used as an abbreviation
IGP-Anycast Segment or Anycast Segment or Anycast-SID	An IGP-Anycast Segment is an IGP-prefix segment which does not identify a specific router, but a set of routers. The terms "Anycast Segment" or "Anycast-SID" are often used as an abbreviation
IGP-Adjacency Segment or Adjacency Segment or Adj-SID	An IGP-Adjacency Segment is an IGP segment attached to an unidirectional adjacency or a set of unidirectional adjacencies. An IGP-Adjacency Segment is local to the node which advertises it
SRDB	The SR Database. Each entry is indexed by a segment value. Each entry must list the SR header operation to apply and the next-hop to forward the packet to
SR Header Operation	Push, Continue and Next are operations applied on the SR segment list

Table 1: Segment Routing Terminology

1.3. Properties

Assuming a packet flow F entering an SR domain at ingress SR edge router I, the properties offered by the SR architecture are:

Per-Flow state for F is only maintained by node I.

Any topological path through the SR domain can be enforced.

Any chain of services through the SR domain can be enforced.

Any mix of topological paths and chain of services can be enforced.

Any node along the flow path can determine where flow entered the SR domain, how it got up to that node, where it will exit the SR domain and how it will get there.

1.4. Companion Documents

This document defines the SR architecture, its routing model, the IGP-based segments and the service segments.

Use cases are described in [I-D.filsfils-rtgwg-segment-routing-use-cases].

The support of SR by the MPLS dataplane is documented in [draft-filsfils-spring-segment-routing-mpls-00].

The support of SR on the Ipv6 dataplane will be documented in a future document.

IS-IS protocol extensions for Segment Routing are described in [I-D.previdi-isis-segment-routing-extensions].

OSPF protocol extensions for Segment Routing are described in [I-D.psenak-ospf-segment-routing-extensions] and [I-D.psenak-ospf-segment-routing-ospfv3-extension].

The FRR solution for SR is documented in [I-D.francois-sr-frr].

The PCEP protocol extensions for Segment Routing are defined in [I-D.sivabalan-pce-segment-routing].

The interaction between SR/MPLS with other MPLS Signaling planes is documented in [draft-filsfils-spring-segment-routing-ldp-interop-00].

1.5. Relationship with MPLS and IPv6

The source routing model is inherited from the one proposed by and [RFC1940] and [RFC2460].

The notion of abstract segment identifier which can represent any instruction is inherited from MPLS ([RFC3031]).

Deployment experiences has shown the need to limit the number of per-flow states maintained in the network while preserving information on the topological and service journey of a packet (e.g. the ingress to the domain for accounting/billing purpose).

The main differences from the IPv6 source route model are:

- The source route is encoded as an ordered list of segments instead of IP addresses.

- A segment can represent any instruction either a service or a topological path. Topologically, the path to an IP address is often limited to the shortest-path to that address. A segment can represent any path (e.g. an adjacency segment forces a packet to a nexthop through a specific adjacency even if the shortest-path to the next-hop does not use that adjacency).

- The ingress and egress egde routers are identified and always available, allowing for interesting accounting and policy applications.

- The source route functionality cannot be controlled from outside the SR domain.

The main differences from the current MPLS model are:

- Globally indexed segments are introduced (e.g. IGP Prefix segments).

- LDP and RSVP MPLS signaling protocols are not required. If present, SR can coexist and interwork with LDP and RSVP. [draft-filsfils-spring-segment-routing-ldp-interop-00].

- Per-flow states are only maintained at the ingress edge router.

SR can be instantiated on the IPv6 dataplane. A future document will detail the new routing extension header which carry all the elements of the abstract SR header. All the SR properties are preserved.

SR can be instantiated on the MPLS dataplane as detailed in [draft-filsfils-spring-segment-routing-mpls-00].

2. Abstract Routing Model

Segment Routing (SR) leverages the source routing paradigm.

At the entrance of the SR domain, the ingress SR edge router pushes

the SR header on top of the packet. At the exit of the SR domain, the egress SR edge router removes the SR header.

The SR header contains an ordered list of segments, a pointer identifying the next segment to process and the identifications of the ingress and egress SR edge routers on the path of this packet. The pointer identifies the segment that MUST be used by the receiving router to process the packet. This segment is called the active segment.

A property of the architecture is that the entire source route of the packet, including the identity of the ingress and egress edge routers is always available with the packet. This allows for interesting accounting and service applications.

We define three SR-header operations:

"PUSH": an SR header is pushed on an IP packet, or additional segments are added at the head of the segment list. The pointer is moved to the first entry of the added segments.

"NEXT": the active segment is completed, the pointer is moved to the next segment in the list.

"CONTINUE": the active segment is not completed, the pointer is left unchanged.

In the future, other SR-header management operations may be defined.

As the packet travels through the SR domain, the pointer is incremented through the ordered list of segments and the source route encoded by the SR ingress edge node is executed.

A node processes an incoming packet according to the instruction associated with the active segment.

Any instruction might be associated with a segment: for example, an intra or inter-domain topological strict or loose forwarding instruction, a service instruction, etc.

At minimum, a segment instruction must define two elements: the identity of the next-hop to forward the packet to (this could be the same node or a context within the node) and which SR-header management operation to execute.

Each segment is known in the network through a Segment Identifier (SID), a value allocated from the 32-bit Segment IDentifier space. The first 16 values are reserved. The terms "segment" and "SID" are

interchangeable.

Within an SR domain, all the SR-capable nodes are configured with the Segment Routing Global Block (SRGB). The SRGB is a subset of the 32-bit SID space. SRGB can be a non-contiguous set of segments.

All global segments must be allocated from the SRGB. Any SR capable node MUST be able to process any global segment advertised by any other node within the SR domain.

Any segment outside the SRGB has a local significance and is called a "local segment". An SR-capable node MUST be able to process the local segments it originates. An SR-capable node MUST NOT support the instruction associated with a local segment originated by a remote node.

2.1. Traffic Engineering with SR

An SR Traffic Engineering policy is composed of two elements: a flow classification and a segment-list to prepend on the packets of the flow.

In the SR architecture, this per-flow state only exists at the ingress egde router whether the policy is defined and the SR header is pushed.

It is outside the scope of the document to define the process that leads to the instantiation at a node N of an SR Traffic Engineering policy.

[I-D.filsfils-rtgwg-segment-routing-use-cases] illustrates various alternatives:

- N is deriving this policy automatically (e.g. FRR).
- N is provisioned explicitly by the operator.
- N is provisioned by a stateful PCE server.
- N is provisioned by the operator with a high-level policy which is mapped into a path thanks to a local CSPF-based computation (e.g. affinity/SRLG exclusion).

Any architecture that involves the insertion of information onto a packet involves performance consideration. [I-D.filsfils-rtgwg-segment-routing-use-cases]explains why the majority of use-cases require very short segment-lists.

A stateful PCE server, which desires to instantiate at node N an SR Traffic Engineering policy, collects the SR capability of node N such as to ensure that the policy meets its capability [I-D.sivabalan-pce-segment-routing].

2.2. Segment Routing Database

The Segment routing Database (SRDB) is a set of entries where each entry is identified by a segment value. The instruction associated with each entry at least defines the identity of the next-hop to which the packet should be forwarded and what operation should be performed on the SR header (PUSH, CONTINUE, NEXT).

Segment	Next-Hop	SR Header operation
Sk	M	CONTINUE
Sj	N	NEXT
Sl	NAT Srvc	NEXT
Sm	FW srvc	NEXT
Sn	Q	NEXT
etc.	etc.	etc.

Figure 4: SR Database

Each SR-capable node maintains its local SRDB. SRDB entries can either derive from local policy or or from protocol segment advertisement. The next section will detail segment advertisement by IGP protocols."

3. Link-State IGP Segments

Within a link-state IGP domain, an SR-capable IGP node advertises segments for its attached prefixes and adjacencies. These segments are called IGP segments or IGP SIDs. They play a key role in the Segment Routing architecture and use-cases [I-D.filsfils-rtgwg-segment-routing-use-cases] as they enable the expression of any topological path throughout the IGP domain. Such a topological path is either expressed as a single IGP segment or a list of multiple IGP segments.

In the first sub-section, we introduce a terminology for a set of IGP segments which are very frequently seen in the SR use-cases. The second sub-section details the IGP segment allocation and SRDB construction rules.

3.1. Illustration

Assuming the network diagram of Figure 5 and the IP address and IGP Segment allocation of Figure 6, the following examples can be constructed.

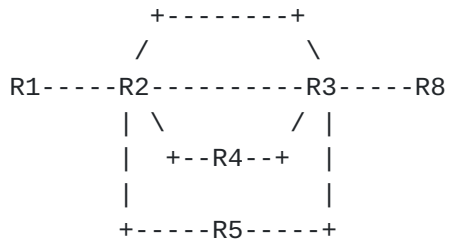

Figure 5: IGP Segments - Illustration

```
+-----------------------------------------------------------+
| IP address allocated by the operator:                     |
|                      192.0.2.1/32 as a loopback of R1     |
|                      192.0.2.2/32 as a loopback of R2     |
|                      192.0.2.3/32 as a loopback of R3     |
|                      192.0.2.4/32 as a loopback of R4     |
|                      192.0.2.5/32 as a loopback of R5     |
|                      192.0.2.8/32 as a loopback of R8     |
|              198.51.100.9/32 as an anycast loopback of R4 |
|              198.51.100.9/32 as an anycast loopback of R5 |
|                                                           |
| SRGB defined by the operator as 1000-5000                 |
|                                                           |
| Global IGP SID allocated by the operator:                 |
|                      1001 allocated to 192.0.2.1/32       |
|                      1002 allocated to 192.0.2.2/32       |
|                      1003 allocated to 192.0.2.3/32       |
|                      1004 allocated to 192.0.2.4/32       |
|                      1008 allocated to 192.0.2.8/32       |
|                      2009 allocated to 198.51.100.9/32    |
|                                                           |
| Local IGP SID allocated dynamically by R2                 |
|                     for its "north" adjacency to R3: 9001 |
|                     for its "north" adjacency to R3: 9003 |
|                     for its "south" adjacency to R3: 9002 |
|                     for its "south" adjacency to R3: 9003 |
+-----------------------------------------------------------+
```

Figure 6: IGP Address and Segment Allocation - Illustration

3.1.1. Example 1

R1 may send a packet P1 to R8 simply by pushing an SR header with segment list {1008}.

1008 is a global IGP segment attached to the IP prefix 192.0.2.8/32. Its semantic is global within the IGP domain: any router forwards a packet received with active segment 1008 to the next-hop along the ECMP-aware shortest-path to the related prefix.

In conclusion, the path followed by P1 is R1-R2--R3-R8. The ECMP-awareness ensures that the traffic be load-shared between any ECMP path, in this case the two north and south links between R2 and R3.

3.1.2. Example 2

R1 may send a packet P2 to R8 by pushing an SR header with segment list {1002, 9001, 1008}.

1002 is a global IGP segment attached to the IP prefix 192.0.2.2/32. Its semantic is global within the IGP domain: any router forwards a packet received with active segment 1002 to the next-hop along the shortest-path to the related prefix.

9001 is a local IGP segment attached by node R2 to its north link to R3. Its semantic is local to node R2: R2 switches a packet received with active segment 9001 towards the north link to R3.

In conclusion, the path followed by P2 is R1-R2-north-link-R3-R8.

3.1.3. Example 3

R1 may send a packet P3 along the same exact path as P1 using a different segment list {1002, 9003, 1008}.

9003 is a local IGP segment attached by node R2 to both its north and south links to R3. Its semantic is local to node R2: R2 switches a packet received with active segment 9003 towards either the north or south links to R3 (e.g. per-flow loadbalancing decision).

In conclusion, the path followed by P3 is R1-R2-any-link-R3-R8.

3.1.4. Example 4

R1 may send a packet P4 to R8 while avoiding the links between R2 and R3 by pushing an SR header with segment list {1004, 1008}.

1004 is a global IGP segment attached to the IP prefix 192.0.2.4/32.

Its semantic is global within the IGP domain: any router forwards a packet received with active segment 1004 to the next-hop along the shortest-path to the related prefix.

In conclusion, the path followed by P4 is R1-R2-R4-R3-R8.

3.1.5. Example 5

R1 may send a packet P5 to R8 while avoiding the links between R2 and R3 while still benefitting from all the remaining shortest paths (via R4 and R5) by pushing an SR header with segment list {2009, 1008}.

2009 is a global IGP segment attached to the anycast IP prefix 198.51.100.9/32. Its semantic is global within the IGP domain: any router forwards a packet received with active segment 2009 to the next-hop along the shortest-path to the related prefix.

In conclusion, the path followed by P5 is either R1-R2-R4-R3-R8 or R1-R2-R5-R3-R8 .

3.2. IGP Segment Terminology

3.2.1. IGP Segment, IGP SID

The terms "IGP Segment" and "IGP SID" are the generic names for a segment attached to a piece of information advertised by a link-state IGP, e.g. an IGP prefix or an IGP adjacency.

The IGP signaling extension to advertise an IGP segment includes the G-Flag indicating whether the IGP segment is global or local.

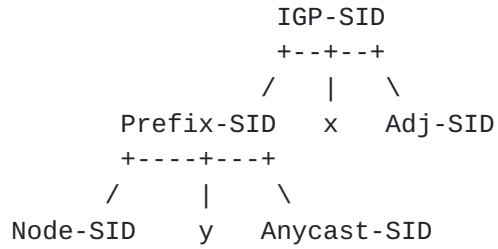

Figure 7: IGP SID Terminology

The IGP Segment terminology is introduced to ease the documentation of SR use-cases and hence does not propose a name for any possible variation of IGP segment supported by the architecture. For example, y in Figure 7 could represent a local IGP segment attached to an IGP Prefix. This variation, while supported by the SR architecture is not seen in the SR use-cases and hence does not receive a specific name.

In Figure 5 and Figure 6, SIDs 1001, 1002, 1003, 1004, 1008, 2009, 9001, 9002 and 9003 are called IGP SIDs.

3.2.2. IGP-Prefix Segment, Prefix-SID

An IGP-Prefix Segment is an IGP segment attached to an IGP prefix. An IGP-Prefix Segment is always global within the SR/IGP domain and identifies the ECMP-aware shortest-path computed by the IGP to the related prefix. The G-Flag MUST be set. The Prefix-SID is the SID of the IGP-Prefix Segment.

A packet injected anywhere within the SR/IGP domain with an active Prefix-SID will be forwarded along the shortest-path to that prefix.

The IGP signaling extension for IGP-Prefix segment includes the P-Flag. A Node N advertising a Prefix-SID SID-R for its attached prefix R resets the P-Flag to allow its connected neighbors to perform the NEXT operation while processing SID-R. This behavior is equivalent to Pen-ultimate Hop Popping in MPLS. When set, the neighbors of N must perform the CONTINUE operation while processing SID-R.

While the architecture allows to attach a local segment to an IGP prefix, we specifically assume that when the terms "IGP-Prefix Segment" and "Prefix-SID" are used then the segment is global (the SID is allocated from the SRGB). This is consistent with [I-D.filsfils-rtgwg-segment-routing-use-cases] as all the described use-cases require global segments attached to IGP prefix.

In Figure 5 and Figure 6, SIDs 1001, 1002, 1003, 1004, 1008, 2009 are called Prefix-SIDs.

3.2.3. IGP-Node Segment, Node-SID

An IGP-Node Segment is a an IGP-Prefix Segment which identifies a specific router (e.g. a loopback). The terms "Node Segment" or "Node-SID" are often used as an abbreviation.

A "Node Segment" or "Node-SID" is fundamental to the SR architecture. From anywhere in the network, it enforces the ECMP-aware shortest-path forwarding of the packet towards the related node as explained in [I-D.filsfils-rtgwg-segment-routing-use-cases].

In Figure 5 and Figure 6, SIDs 1001, 1002, 1003, 1004 and 1008 are called Node-SIDs.

3.2.4. IGP-Anycast Segment, Anycast SID

An IGP-Anycast Segment is an IGP-prefix segment which does not identify a specific router, but a set of routers. The terms "Anycast Segment" or "Anycast-SID" are often used as an abbreviation.

An "Anycast Segment" or "Anycast SID" enforces the ECMP-aware shortest-path forwarding towards the closest node of the anycast set. This is useful to express macro-engineering policies as described in [I-D.filsfils-rtgwg-segment-routing-use-cases].

In Figure 5 and Figure 6, SID 2009 is called Anycast SID.

3.2.5. IGP-Adjacency Segment, Adj-SID

An IGP-Adjacency Segment is an IGP segment attached to an unidirectional adjacency or a set of unidirectional adjacencies. An IGP-Adjacency Segment is local to the node which advertises it. The SID of the IGP-Adjacency Segment is called the Adj-SID. The G-Flag must be reset.

The adjacency is formed by the local node (i.e.: the node advertising the adjacency in the IGP) and the remote node (i.e.: the other end of the adjacency). The local node MUST be an IGP node. The remote node MAY be:

An adjacent IGP node (i.e.: an IGP neighbor).

A non-adjacent neighbor (e.g.: a Forwarding Adjacency, [RFC4206]).

A virtual neighbor outside the IGP domain (e.g.: an interface connecting another AS) as defined in [RFC5316].

A packet injected anywhere within the SR/IGP domain with a segment list {SN, SNL}, where SN is the Node-SID of node N and SNL is an Adj-Sid attached by node N to its adjacency over link L, will be forwarded along the shortest-path to N and then be switched by N, without any IP shortest-path consideration, towards link L. If the Adj-Sid identifies a set of adjacencies, then the node N load-balances the traffic along the various members of the set.

An "IGP Adjacency Segment" or "Adj-SID" enforces the switching of the packet from a note towards a defined interface or set of interfaces. This is key to theoretically prove that any path can be expressed as a list of segments as explained in [I-D.filsfils-rtgwg-segment-routing-use-cases].

In Figure 5 and Figure 6, SIDs 9001, 9002 and 9003 are called Adj-

SIDs.

3.2.6. Finally

Figure 8 summarizes the different terms that can be used to refer to the SID's used in the example illustrated by Figure 5 and Figure 6. "Y" means that the term can be used to refer to the SID, "N" means that the term cannot be used to refer to the SID.

SID Value	IGP SID	Prefix-SID	Node-SID	Anycast SID	Adj-SID
1001	Y	Y	Y	N	N
1002	Y	Y	Y	N	N
1003	Y	Y	Y	N	N
1004	Y	Y	Y	N	N
1005	Y	Y	Y	N	N
1008	Y	Y	Y	N	N
2009	Y	Y	N	Y	N
9001	Y	N	N	N	Y
9002	Y	N	N	N	Y
9003	Y	N	N	N	Y

Figure 8: Terminology Example

3.3. IGP Segment Allocation, Advertisement and SRDB Maintenance

3.3.1. Prefix-SID

Multiple Prefix-SID's may be allocated to the same IGP Prefix (e.g. for class of service purpose). Typically a single Prefix-SID is allocated to an IGP Prefix.

A Prefix-SID is allocated from the SRGB according to a similar process to IP address allocation. Typically the Prefix-SID is allocated by policy by the operator (or NMS) and the SID very rarely changes.

The allocation process MUST NOT allocate the same Prefix-SID to different IP prefixes.

If a node learns a Prefix-SID having a value that falls outside the locally configured SRGB range, then the node MUST NOT use the Prefix-SID and SHOULD issue an error log warning for misconfiguration.

The required IGP protocol extensions are defined in [I-D.previdi-isis-segment-routing-extensions],

[I-D.psenak-ospf-segment-routing-extensions] and
[I-D.psenak-ospf-segment-routing-ospfv3-extension].

A node N attaching a Prefix-SID SID-R to its attached prefix R MUST
maintain the following SRDB entry:
Incoming Active Segment: SID-R
Ingress Operation: NEXT
Egress interface: NULL

A remote node M MUST maintain the following SRDB entry for any
learned Prefix-SID SID-R attached to IP prefix R:
Incoming Active Segment: SID-R
Ingress Operation:
        If the next-hop of R is the originator of R
        and instructed to remove the active segment: NEXT
        Else: CONTINUE
Egress interface: the interface towards the next-hop along
        the shortest-path to prefix R.

3.3.2. Adj-SID

The Adjacency Segment SID (Adj-SID) identifies a unidirectional
adjacency or a set of unidirectional adjacencies.
A node SHOULD allocate one Adj-SIDs for each of its adjacencies.
A node MAY allocate multiple Adj-SIDs to the same adjacency.
A node MAY allocate the same Adj-SID to multiple adjacencies.

Adjacency suppression MUST NOT be performed by the IGP.

A node MUST install an SRDB entry for any Adj-SID of value V attached
to data-link L:
Incoming Active Segment: V
Operation: NEXT
Egress Interface: L

When associated to a Forwarding Adjacency ([RFC4206]), the Adj-SID
MAY also include the necessary information in order to describe the
path to the remote end of the Forwarding Adjacency in the form of an
Explicit Route Object.

The Adj-SID implies, from the router advertising it, the forwarding
of the packet through the adjacency identified by the Adj-SID,
regardless its IGP/SPF cost. In other words, the use of Adjacency
Segments overrides the routing decision made by SPF algorithm.

3.3.2.1. Parallel Adjacencies

Adj-SIDs can be used in order to represent a set of parallel interfaces between two adjacent routers. For example, SID 9003 in figures 5 and 6 identify the set of interfaces between R2 and R3.

A node MUST install an SRDB entry for any locally originated Adjacency Segment (Adj-SID) of value W attached to a set of link B with:
Incoming Active Segment: W
Ingress Operation: NEXT
Egress interface: loadbalance between any data-link within set B

3.3.2.2. LAN Adjacency Segments

In LAN subnetworks, link-state protocols define the concept of Designated Router (DR, in OSPF) or Designated Intermediate System (DIS, in IS-IS) that conduct flooding in broadcast subnetworks and that describe the LAN topology in a special routing update (OSPF Type2 LSA or IS-IS Pseudonode LSP).

The difficulty with LANs is that each router only advertises its connectivity to the DR/DIS and not to each other individual nodes in the LAN. Therefore, additional protocol mechanisms (IS-IS and OSPF) are necessary in order for each router in the LAN to advertise an Adj-SID associated to each neighbor in the LAN. These extensions are defined in [I-D.previdi-isis-segment-routing-extensions], [I-D.psenak-ospf-segment-routing-extensions] and [I-D.psenak-ospf-segment-routing-ospfv3-extension].

3.3.2.3. External Adjacencies Considerations

IGPs have been extended in order to advertise virtual adjacencies that represent external links ([RFC5316]).

Segment Routing allows to allocate an Adj-SID to these external links.

Figure 9: External Adjacency Example

In the diagram above, C advertises in the IGP an adjacency to peer F of AS2 together with an associated Adj-SID. When S wants to force an inter-domain path to Z via the peering link CF, S encapsulates the packets with the list {Prefix-SID(C), Adj-SID(C,F, AS2)}.

[I-D.filsfils-rtgwg-segment-routing-use-cases] provides an external-adjacency use-case.

3.4. Inter-Area Considerations

In the following example diagram we assume an IGP deployed using areas and where SR has been deployed.

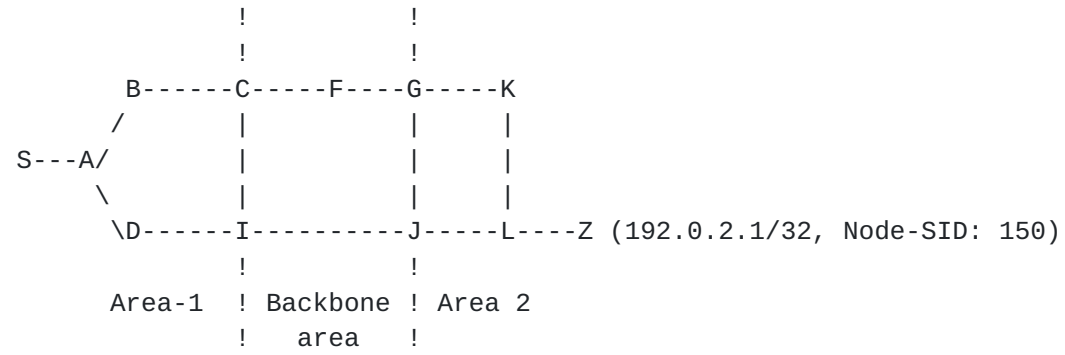

Figure 10: Inter-Area Topology Example

In area 2, node Z allocates Node-SID 150 to his local prefix 192.0.2.1/32. ABRs G and J will propagate the prefix into the backbone area by creating a new instance of the prefix according to normal inter-area/level IGP propagation rules.

Nodes C and I will apply the same behavior when leaking prefixes from the backbone area down to area 1. Therefore, node S will see prefix 192.0.2.1/32 with Prefix-SID 150 and advertised by nodes C and I.

It therefore results that a Prefix-SID remains attached to its

related IGP Prefix through the inter-area process.

When node S sends traffic to 192.0.2.1/32, it pushes Node-SID(150) as active segment and forward it to A.

When packet arrives at ABR I (or C), the ABR forwards the packet according to the active segment (Node-SID(150)). Forwarding continues across area borders, using the same Node-SID(150), until the packet reaches its destination.

When an ABR propagates a prefix from one area to another it MUST set the R-Flag.

3.5. IGP Mirroring Context Segment

It is beneficial for an IGP node to be able to advertise its ability to process traffic originally destined to another IGP node, called the Mirrored node and identified by an IP address or a Node-SID, provided that a "Mirroring Context" segment be inserted in the segment list prior to any service segment local to the mirrored node.

[I-D.filsfils-rtgwg-segment-routing-use-cases] illustrates such a use-case where two IGP nodes offer the same set of services (e.g. BGP VPN) and mirror each other upon their failure. A similar behavior is described in [I-D.minto-rsvp-lsp-egress-fast-protection].

IS-IS and OSPF Router Capability extensions are described in [I-D.previdi-isis-segment-routing-extensions], [I-D.psenak-ospf-segment-routing-extensions] and [I-D.psenak-ospf-segment-routing-ospfv3-extension].

4. Service Segments

A service segment refers to a service offered by a node (e.g. firewall, vpn, etc.).

Further informations will be included in future revisions.

5. OAM

SR offers an interesting capability to monitor SR domains:

Any path can be monitored by setting the segment list accordingly.

A path can be expressed with ECMP-awareness or not.

The probe travels along the desired path while staying at the forwarding level.

A monitoring system is able to check any element of the entire SR domain, even if it located multiple hops away.

Some elements of the SR/OAM functionality will require standardization and a related independent draft will eventually be submitted.

SR/OAM use-cases are described in [I-D.filsfils-rtgwg-segment-routing-use-cases].

6. Multicast

The text will be added in future revision.

7. IANA Considerations

TBD

8. Manageability Considerations

TBD

9. Security Considerations

TBD

10. Acknowledgements

We would like to thank Dave Ward, Dan Frost, Stewart Bryant, Pierre Francois, Thomas Telkamp, Les Ginsberg, Ruediger Geib and Hannes Gredler 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.

[RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, December 1998.

[RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, January 2001.

[RFC4206] Kompella, K. and Y. Rekhter, "Label Switched Paths (LSP) Hierarchy with Generalized Multi-Protocol Label Switching (GMPLS) Traffic Engineering (TE)", RFC 4206, October 2005.

[RFC5316] Chen, M., Zhang, R., and X. Duan, "ISIS Extensions in Support of Inter-Autonomous System (AS) MPLS and GMPLS Traffic Engineering", RFC 5316, December 2008.

11.2. Informative References

[I-D.filsfils-rtgwg-segment-routing-use-cases]
Filsfils, C., Francois, P., Previdi, S., Decraene, B., Litkowski, S., Horneffer, M., Milojevic, I., Shakir, R., Ytti, S., Henderickx, W., Tantsura, J., and E. Crabbe, "Segment Routing Use Cases", draft-filsfils-rtgwg-segment-routing-use-cases-01 (work in progress), July 2013.

[I-D.francois-sr-frr]
Francois, P., Filsfils, C., Bashandy, A., Previdi, S., and B. Decraene, "Segment Routing Fast Reroute", draft-francois-sr-frr-00 (work in progress), July 2013.

[I-D.minto-rsvp-lsp-egress-fast-protection]
Jeganathan, J., Gredler, H., and Y. Shen, "RSVP-TE LSP egress fast-protection", draft-minto-rsvp-lsp-egress-fast-protection-02 (work in progress), April 2013.

[I-D.previdi-isis-segment-routing-extensions]
Previdi, S., Filsfils, C., Bashandy, A., Gredler, H., and S. Litkowski, "IS-IS Extensions for Segment Routing", draft-previdi-isis-segment-routing-extensions-02 (work in progress), July 2013.

[I-D.psenak-ospf-segment-routing-extensions]
Psenak, P., Previdi, S., Filsfils, C., Gredler, H., and R.

Shakir, "OSPF Extensions for Segment Routing", draft-psenak-ospf-segment-routing-extensions-02 (work in progress), July 2013.

[I-D.psenak-ospf-segment-routing-ospfv3-extension]
Psenak, P. and S. Previdi, "OSPFv3 Extensions for Segment Routing", October 2013.

[I-D.sivabalan-pce-segment-routing]
Sivabalan, S., Medved, J., Filsfils, C., Crabbe, E., and R. Raszuk, "PCEP Extensions for Segment Routing", draft-sivabalan-pce-segment-routing-02 (work in progress), October 2013.

[RFC1940] Estrin, D., Li, T., Rekhter, Y., Varadhan, K., and D. Zappala, "Source Demand Routing: Packet Format and Forwarding Specification (Version 1)", RFC 1940, May 1996.

[draft-filsfils-spring-segment-routing-ldp-interop-00]
Filsfils, C. and S. Previdi, "Segment Routing interoperability with LDP", October 2013.

[draft-filsfils-spring-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