

Network Working Group  
Internet-Draft  
Intended status: Standards Track  
Expires: June 17, 2016

C. Filsfils, Ed.  
S. Previdi, Ed.  
Cisco Systems, Inc.  
B. Decraene  
S. Litkowski  
Orange  
R. Shakir  
Jive Communications  
December 15, 2015

**Segment Routing Architecture**  
**draft-ietf-spring-segment-routing-07**

Abstract

Segment Routing (SR) leverages the source routing paradigm. A node steers a packet through an ordered list of instructions, called segments. 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.

Segment Routing can be directly applied to the MPLS architecture with no change on the forwarding plane. A segment is encoded as an MPLS label. An ordered list of segments is encoded as a stack of labels. The segment to process is on the top of the stack. Upon completion of a segment, the related label is popped from the stack.

Segment Routing can be applied to the IPv6 architecture, with a new type of routing header. A segment is encoded as an IPv6 address. An ordered list of segments is encoded as an ordered list of IPv6 addresses in the routing header. The active segment is indicated by the Destination Address of the packet. The next active segment is indicated by a pointer in the new routing 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 June 17, 2016.

Copyright Notice

Copyright (c) 2015 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](#) . . . . . [3](#)
- [1.1. Companion Documents](#) . . . . . [4](#)
- [2. Terminology](#) . . . . . [5](#)
- [3. Link-State IGP Segments](#) . . . . . [7](#)
- [3.1. IGP Segment, IGP SID](#) . . . . . [7](#)
- [3.2. IGP-Prefix Segment, Prefix-SID](#) . . . . . [7](#)
- [3.2.1. Prefix-SID Algorithm](#) . . . . . [7](#)
- [3.2.2. MPLS Dataplane](#) . . . . . [8](#)
- [3.2.3. IPv6 Dataplane](#) . . . . . [9](#)
- [3.3. IGP-Node Segment, Node-SID](#) . . . . . [10](#)
- [3.4. IGP-Anycast Segment, Anycast SID](#) . . . . . [10](#)
- [3.5. IGP-Adjacency Segment, Adj-SID](#) . . . . . [13](#)
- [3.5.1. Parallel Adjacencies](#) . . . . . [14](#)
- [3.5.2. LAN Adjacency Segments](#) . . . . . [15](#)
- [3.6. Binding Segment](#) . . . . . [16](#)
- [3.6.1. Mapping Server](#) . . . . . [16](#)
- [3.6.2. Tunnel Headend](#) . . . . . [16](#)
- [3.7. Inter-Area Considerations](#) . . . . . [16](#)
- [4. BGP Peering Segments](#) . . . . . [17](#)



<a href="#">5.</a>	IGP Mirroring Context Segment . . . . .	<a href="#">18</a>
<a href="#">6.</a>	Multicast . . . . .	<a href="#">18</a>
<a href="#">7.</a>	IANA Considerations . . . . .	<a href="#">19</a>
<a href="#">8.</a>	Security Considerations . . . . .	<a href="#">19</a>
<a href="#">9.</a>	Contributors . . . . .	<a href="#">19</a>
<a href="#">10.</a>	Acknowledgements . . . . .	<a href="#">20</a>
<a href="#">11.</a>	References . . . . .	<a href="#">20</a>
<a href="#">11.1.</a>	Normative References . . . . .	<a href="#">20</a>
<a href="#">11.2.</a>	Informative References . . . . .	<a href="#">20</a>
	Authors' Addresses . . . . .	<a href="#">23</a>

## [1.](#) Introduction

With Segment Routing (SR), a node steers a packet through an ordered list of instructions, called segments. 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 path and service chain while maintaining per-flow state only at the ingress node of the SR domain.

Segment Routing can be directly applied to the MPLS architecture ([\[RFC3031\]](#)) with no change on the forwarding plane. A segment is encoded as an MPLS label. An ordered list of segments is encoded as a stack of labels. The active segment is on the top of the stack. A completed segment is popped off the stack. The addition of a segment is performed with a push.

In the Segment Routing MPLS instantiation, a segment could be of several types:

- o an IGP segment,
- o a BGP Peering segments,
- o an LDP LSP segment,
- o an RSVP-TE LSP segment,
- o a BGP LSP segment.

The first two (IGP and BGP Peering segments) types of segments are defined in this document. The use of the last three types of segments is illustrated in [\[I-D.ietf-spring-segment-routing-mpls\]](#).

Segment Routing can be applied to the IPv6 architecture ([\[RFC2460\]](#)), with a new type of routing header. A segment is encoded as an IPv6 address. An ordered list of segments is encoded as an ordered list of IPv6 addresses in the routing header. The active segment is



indicated by the Destination Address of the packet. Upon completion of a segment, a pointer in the new routing header is incremented and indicates the next segment.

Numerous use-cases illustrate the benefits of source routing either for FRR, OAM or Traffic Engineering reasons.

This document defines a set of instructions (called segments) that are required to fulfill the described use-cases. These segments can either be used in isolation (one single segment defines the source route of the packet) or in combination (these segments are part of an ordered list of segments that define the source route of the packet).

### **1.1. Companion Documents**

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

Use cases are described in [[I-D.ietf-spring-problem-statement](#)], [[I-D.ietf-spring-segment-routing-central-epe](#)], [[I-D.ietf-spring-segment-routing-msdc](#)], [[I-D.filsfils-spring-large-scale-interconnect](#)], [[I-D.ietf-spring-ipv6-use-cases](#)], [[I-D.ietf-spring-resiliency-use-cases](#)], [[I-D.ietf-spring-oam-usecase](#)] and [[I-D.ietf-spring-sr-oam-requirement](#)].

Segment Routing for MPLS dataplane is documented in [[I-D.ietf-spring-segment-routing-mpls](#)].

Segment Routing for IPv6 dataplane is documented in [[I-D.ietf-6man-segment-routing-header](#)].

IGP protocol extensions for Segment Routing are described in [[I-D.ietf-isis-segment-routing-extensions](#)], [[I-D.ietf-ospf-segment-routing-extensions](#)] and [[I-D.ietf-ospf-ospfv3-segment-routing-extensions](#)] referred in this document as "IGP SR extensions documents".

The FRR solution for SR is documented in [[I-D.francois-rtgwg-segment-routing-ti-lfa](#)].

The PCEP protocol extensions for Segment Routing are defined in [[I-D.ietf-pce-segment-routing](#)].

The interaction between SR/MPLS with other MPLS Signaling planes is documented in [[I-D.ietf-spring-segment-routing-ldp-interop](#)].



## **2. Terminology**

Segment: an instruction a node executes on the incoming packet (e.g.: forward packet according to shortest path to destination, or, forward packet through a specific interface, or, deliver the packet to a given application/service instance).

SID: a Segment Identifier. Examples of SIDs are: a MPLS label, an index value in a MPLS label space, an IPv6 address. Other types of SIDs can be defined in the future.

Segment List: ordered list of SID's encoding the topological and service source route of the packet. It is a stack of labels in the MPLS architecture. It is an ordered list of IPv6 addresses in the IPv6 architecture.

Segment Routing Domain (SR Domain): the set of nodes participating into the source based routing model. These nodes may be connected to the same physical infrastructure (e.g.: a Service Provider's network) as well as nodes remotely connected to each other (e.g.: an enterprise VPN or an overlay). Note that a SR domain may also be confined within an IGP instance, in which case it is named SR-IGP Domain.

Active segment: the segment that MUST be used by the receiving router to process the packet. In the MPLS dataplane is the top label. In the IPv6 dataplane is the destination address of a packet having the Segment Routing Header as defined in [\[I-D.ietf-6man-segment-routing-header\]](#).

PUSH: the insertion of a segment at the head of the Segment list.

NEXT: the active segment is completed, the next segment becomes active.

CONTINUE: the active segment is not completed and hence remains active. The CONTINUE instruction is implemented as the SWAP instruction in the MPLS dataplane. In IPv6, this is the plain IPv6 forwarding action of a regular IPv6 packet according to its Destination Address.

SR Global Block (SRGB): local property of an SR node. In the MPLS architecture, SRGB is the set of local labels reserved for global segments. Using the same SRGB on all nodes within the SR domain ease operations and troubleshooting and is expected to be a deployment guideline. In the IPv6 architecture, the equivalent of the SRGB is in fact the set of addresses used as global segments. Since there are no restrictions on which IPv6 address can be used, the concept of





the SRGB includes all IPv6 global address space used within the SR domain.

**Global Segment:** the related instruction is supported by all the SR-capable nodes in the domain. In the MPLS architecture, a Global Segment has a globally-unique index. The related local label at a given node N is found by adding the globally-unique index to the SRGB of node N. In the IPv6 architecture, a global segment is a globally-unique IPv6 address.

**Local Segment:** the related instruction is supported only by the node originating it. In the MPLS architecture, this is a local label outside the SRGB. In the IPv6 architecture, this can be any IPv6 address whose reachability is not advertised in any routing protocol (hence, the segment is known only by the local node).

**IGP Segment:** the generic name 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, Prefix-SID:** an IGP-Prefix Segment is an IGP segment attached to an IGP prefix. An IGP-Prefix Segment is global (unless explicitly advertised otherwise) within the SR IGP instance/topology and identifies an instruction to forward the packet along the path computed using the algorithm field, in the topology and the IGP instance where it is advertised. The Prefix-SID is the SID of the IGP-Prefix Segment.

**IGP-Anycast:** 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:** an IGP-Adjacency Segment is an IGP segment attached to an unidirectional adjacency or a set of unidirectional adjacencies. By default, an IGP-Adjacency Segment is local (unless explicitly advertised otherwise) to the node that advertises it.

**IGP-Node:** an IGP-Node Segment is 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.

**SR Tunnel:** a list of segments to be pushed on the packets directed on the tunnel. The list of segments can be specified explicitly or implicitly via a set of abstract constraints (latency, affinity, SRLG, ...). In the latter case, a constraint-based path computation is used to determine the list of segments associated with the tunnel. The computation can be local or delegated to a PCE server. An SR tunnel can be configured by the operator, provisioned via netconf or



provisioned via PCEP. An SR tunnel can be used for traffic-engineering, OAM or FRR reasons.

Segment List Depth: the number of segments of an SR tunnel. The entity instantiating an SR Tunnel at a node N should be able to discover the depth insertion capability of the node N. The PCEP discovery capability is described in [[I-D.ietf-pce-segment-routing](#)].

### **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 Segment Routing and 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.

#### **3.1. 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.

#### **3.2. IGP-Prefix Segment, Prefix-SID**

An IGP-Prefix Segment is an IGP segment attached to an IGP prefix. An IGP-Prefix Segment is global (unless explicitly advertised otherwise) within the SR/IGP domain.

The required IGP protocol extensions are defined in IGP SR extensions documents.

##### **3.2.1. Prefix-SID Algorithm**

The IGP protocol extensions for Segment Routing define the Prefix-SID advertisement which includes a set of flags and the algorithm field. The algorithm field has the purpose of associating a given Prefix-SID to a routing algorithm.

In the context of an instance and a topology, multiple Prefix-SID's MAY be allocated to the same IGP Prefix as long as the algorithm value is different in each one.

Multiple instances and topologies are defined in IS-IS and OSPF in: [[RFC5120](#)], [[RFC6822](#)], [[RFC6549](#)] and [[RFC4915](#)].

Initially, two "algorithms" have been defined:



- o "Shortest Path": this algorithm is the default behavior. The packet is forwarded along the well known ECMP-aware SPF algorithm however it is explicitly allowed for a midpoint to implement another forwarding based on local policy.. The "Shortest Path" algorithm is in fact the default and current behavior of most of the networks where local policies may override the SPF decision.
- o "Strict Shortest Path": This algorithm mandates that the packet is forwarded according to ECMP-aware SPF algorithm and instruct any router in the path to ignore any possible local policy overriding SPF decision. The SID advertised with "Strict Shortest Path" algorithm ensures that the path the packet is going to take is the expected, and not altered, SPF path.

An IGP-Prefix Segment identifies the path, to the related prefix, along the path computed as per the algorithm field.

A packet injected anywhere within the SR/IGP domain with an active Prefix-SID will be forwarded along path computed by the algorithm expressed in the algorithm field.

The ingress node of an SR domain validates that the path to a prefix, advertised with a given algorithm, includes nodes all supporting the advertised algorithm. In other words, when computing paths for a given algorithm, the transit nodes MUST compute the algorithm X on the IGP topology, regardless of the support of the algorithm X by the nodes in that topology. As a consequence, if a node on the path does not support algorithm X, the IGP-Prefix segment will be interrupted and will drop packet on that node. It's the responsibility of the ingress node using a segment to check that all downstream nodes support the algorithm of the segment.

Details of the two defined algorithms are defined in [\[I-D.ietf-isis-segment-routing-extensions\]](#), [\[I-D.ietf-ospf-segment-routing-extensions\]](#) and [\[I-D.ietf-ospf-ospfv3-segment-routing-extensions\]](#).

### **3.2.2. MPLS Dataplane**

When SR is used over the MPLS dataplane:

- o 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 Penultimate Hop Popping in MPLS. When set, the neighbors of N must perform the CONTINUE operation while processing SID-R.



- o A Prefix-SID is allocated in the form of an index in the SRGB (or as a local MPLS label) according to a process similar to IP address allocation. Typically the Prefix-SID is allocated by policy by the operator (or NMS) and the SID very rarely changes.
- o While SR allows to attach a local segment to an IGP prefix (using the L-Flag), we specifically assume that when the terms "IGP-Prefix Segment" and "Prefix-SID" are used, the segment is global (the SID is allocated from the SRGB or as an index). This is consistent with all the described use-cases that require global segments attached to IGP prefixes.
- o The allocation process MUST NOT allocate the same Prefix-SID to different IP prefixes.
- o 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.
- o A node N attaching a Prefix-SID SID-R to its attached prefix R MUST maintain the following FIB entry:

```
Incoming Active Segment: SID-R
Ingress Operation: NEXT
Egress interface: NULL
```

- o A remote node M MUST maintain the following FIB 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
                  path computed using the algorithm advertised with
                  the SID toward prefix R.
```

### **3.2.3. IPv6 Dataplane**

When SR is used over the IPv6 dataplane:

- o The Prefix-SID is the prefix itself. No additional identifier is needed for Segment Routing over IPv6.
- o Any address belonging to any of the node's prefixes can be used as Prefix-SIDs.





- o An operator may want to explicitly indicate which of the node's prefixes can be used as Prefix-SIDs through the setting of a flag (e.g.: using the IGP prefix attribute defined in [[I-D.ietf-isis-prefix-attributes](#)]) in the routing protocol used for advertising the prefix.
- o A global SID is instantiated through any globally advertised IPv6 address.
- o A local SID is instantiated through a local IPv6 prefix not being advertised and therefore known only by the local node.

A node N advertising an IPv6 address R usable as a segment identifier MUST maintain the following FIB entry:

```
Incoming Active Segment: R
Ingress Operation: NEXT
Egress interface: NULL
```

Regardless Segment Routing, any remote IPv6 node will maintain a plain IPv6 FIB entry for any prefix, no matter if they represent a segment or not.

### **3.3. 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. The IGP SR extensions define a flag that identifies Node-SIDs.

A "Node Segment" or "Node-SID" is fundamental to SR. From anywhere in the network, it enforces the ECMP-aware shortest-path forwarding of the packet towards the related node.

An IGP Node-SID MUST NOT be associated with a prefix that is owned by more than one router within the same routing domain.

### **3.4. IGP-Anycast Segment, Anycast SID**

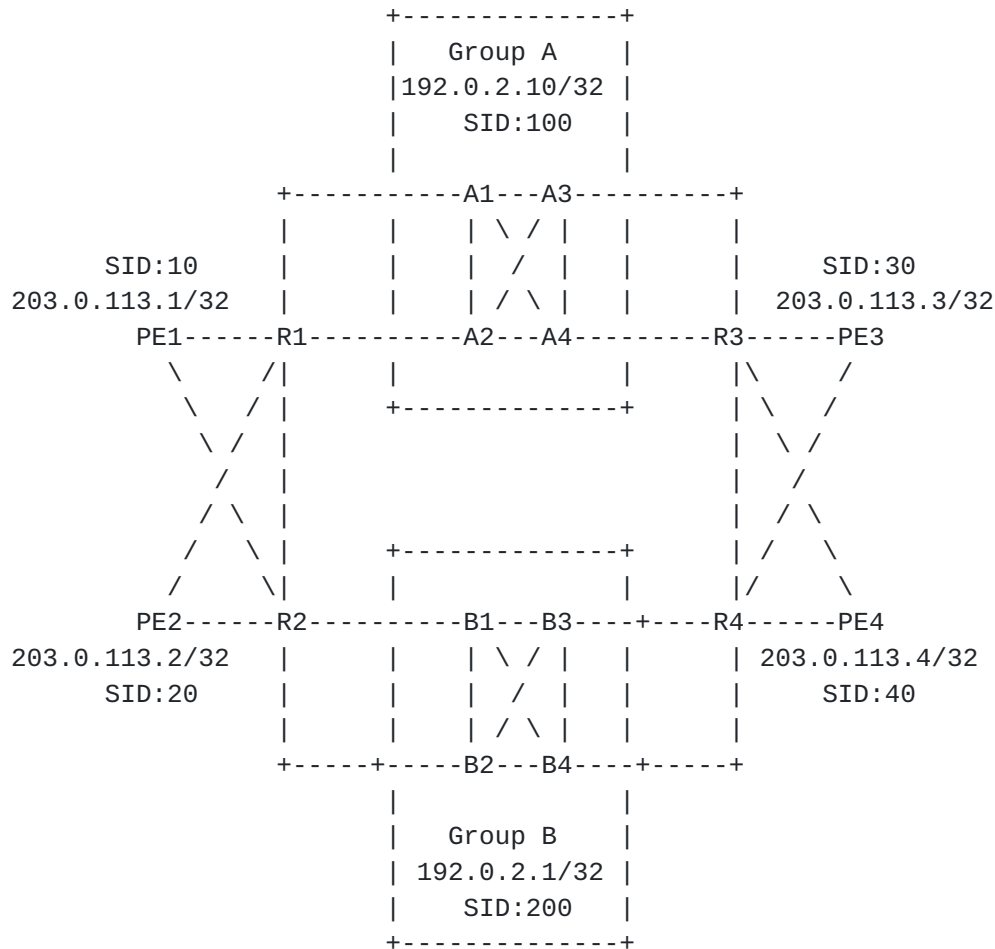
An IGP-Anycast Segment is an IGP-prefix segment which does not identify a specific router, but a set of routers. The terms "Anycast Segment" or "Anycast-SID" are often used as an abbreviation.

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 or protection mechanisms.



An IGP-Anycast Segment MUST NOT reference a particular node.

Within an anycast group, all routers MUST advertise the same prefix with the same SID value.



Transit device groups

The figure above describes a network example with two groups of transit devices. Group A consists of devices {A1, A2, A3 and A4}. They are all provisioned with the anycast address 192.0.2.10/32 and the anycast SID 100.

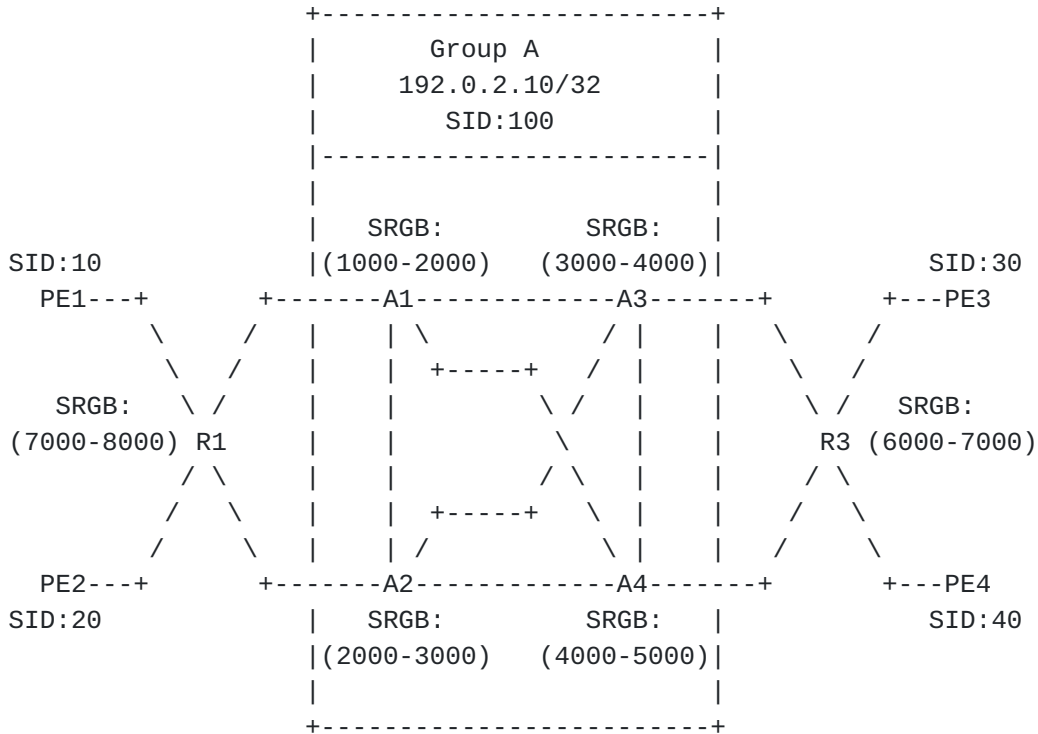
Similarly, group B consists of devices {B1, B2, B3 and B4} and are all provisioned with the anycast address 192.0.2.1/32, anycast SID 200. In the above network topology, each PE device is connected to two routers in each of the groups A and B.

PE1 can choose a particular transit device group when sending traffic to PE3 or PE4. This will be done by pushing the anycast SID of the group in the stack.



Processing the anycast, and subsequent segments, requires special care.

Obviously, the value of the SID following the anycast SID MUST be understood by all nodes advertising the same anycast segment.



Transit paths via anycast group A

Considering a MPLS deployment, in the above topology, if device PE1 (or PE2) requires to send a packet to the device PE3 (or PE4) it needs to encapsulate the packet in a MPLS payload with the following stack of labels.

- o Label allocated by R1 for anycast SID 100 (outer label).
- o Label allocated by the nearest router in group A for SID 30 (for destination PE3).

While the first label is easy to compute, in this case since there are more than one topologically nearest devices (A1 and A2), unless A1 and A2 allocated the same label value to the same prefix, determining the second label is impossible. Devices A1 and A2 may be devices from different hardware vendors. If both don't allocate the same label value for SID 30, it is impossible to use the anycast group "A" as a transit anycast group towards PE3. Hence, PE1 (or PE2) cannot compute an appropriate label stack to steer the packet



exclusively through the group A devices. Same holds true for devices PE3 and PE4 when trying to send a packet to PE1 or PE2.

To ease the use of anycast segment in a short term, it is recommended to configure the same SRGB on all nodes of a particular anycast group. Using this method, as mentioned above, computation of the label following the anycast segment is straightforward.

Using anycast segment without configuring the same SRGB on nodes belonging to the same device group may lead to misrouting (in a MPLS VPN deployment, some traffic may leak between VPNs).

### **3.5. IGP-Adjacency Segment, Adj-SID**

An IGP-Adjacency Segment is an IGP segment attached to a unidirectional adjacency or a set of unidirectional adjacencies. By default, an IGP-Adjacency Segment is local to the node which advertises it. However, an Adjacency Segment can be global if advertised by the IGP as such. The SID of the IGP-Adjacency Segment is called the Adj-SID.

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

A packet injected anywhere within the SR 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 among the various members of the set.

Similarly, when using a global Adj-SID, a packet injected anywhere within the SR domain with a segment list {SNL}, where SNL is a global 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 among the various members of the set. The use of global Adj-SID allows to reduce the size of the segment list when expressing a path at the cost of additional state (i.e.: the global Adj-SID will be inserted by all routers within the area in their forwarding table).





An "IGP Adjacency Segment" or "Adj-SID" enforces the switching of the packet from a node 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.

The encodings of the Adj-SID include the B-flag. When set, the Adj-SID refers to an adjacency that is eligible for protection (e.g.: using IPFRR or MPLS-FRR).

The encodings of the Adj-SID include the L-flag. When set, the Adj-SID has local significance. By default the L-flag is set.

A node SHOULD allocate one Adj-SIDs for each of its adjacencies.

A node MAY allocate multiple Adj-SIDs to the same adjacency. An example is where the adjacency is established over a bundle interface. Each bundle member MAY have its own Adj-SID.

A node MAY allocate the same Adj-SID to multiple adjacencies.

Adjacency suppression MUST NOT be performed by the IGP.

A node MUST install a FIB entry for any Adj-SID of value V attached to data-link L:

```
Incoming Active Segment: V
Operation: NEXT
Egress Interface: L
```

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.

### **3.5.1. Parallel Adjacencies**

Adj-SIDs can be used in order to represent a set of parallel interfaces between two adjacent routers.

A node MUST install a FIB 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
```

When parallel adjacencies are used and associated to the same Adj-SID, and in order to optimize the load balancing function, a "weight"



factor can be associated to the Adj-SID advertised with each adjacency. The weight tells the ingress (or a SDN/orchestration system) about the loadbalancing factor over the parallel adjacencies. As shown in Figure 1, A and B are connected through two parallel adjacencies

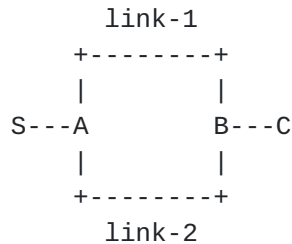


Figure 1: Parallel Links and Adj-SIDs

Node A advertises following Adj-SIDs and weights:

- o Link-1: Adj-SID 1000, weight: 1
- o Link-2: Adj-SID 1000, weight: 2

Node S receives the advertisements of the parallel adjacencies and understands that by using Adj-SID 1000 node A will loadbalance the traffic across the parallel links (link-1 and link-2) according to a 1:2 ratio.

The weight value is advertised with the Adj-SID as defined in IGP SR extensions documents.

### 3.5.2. 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 IGP SR extensions documents.



### 3.6. Binding Segment

#### 3.6.1. Mapping Server

A Remote-Binding SID S advertised by the mapping server M for remote prefix R attached to non-SR-capable node N signals the same information as if N had advertised S as a Prefix-SID. Further details are described in the SR/LDP interworking procedures ([I-D.ietf-spring-segment-routing-ldp-interop]).

The segment allocation and SRGB Maintenance rules are the same as those defined for Prefix-SID.

#### 3.6.2. Tunnel Headend

The segment allocation and SRGB Maintenance rules are the same as those defined for Adj-SID. A tunnel attached to a head-end H acts as an adjacency attached to H.

Note: an alternative consists of representing tunnels as forwarding-adjacencies ([RFC4206]). In such case, the tunnel is presented to the routing area as a routing adjacency and is considered as such by all area routers. The Remote-Binding SID is preferred as it allows to advertise the presence of a tunnel without influencing the LSDB and the SPF computation.

### 3.7. Inter-Area Considerations

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

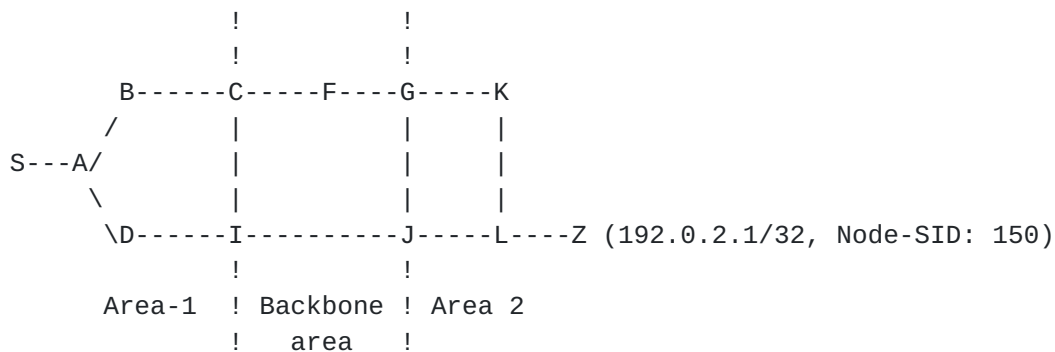


Figure 2: Inter-Area Topology Example

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.

#### **4. BGP Peering Segments**

In the context of BGP Egress Peer Engineering (EPE), as described in [[I-D.ietf-spring-segment-routing-central-epe](#)], an EPE enabled Egress PE node MAY advertise segments corresponding to its attached peers. These segments are called BGP peering segments or BGP Peering SIDs. They enable the expression of source-routed inter-domain paths.

An ingress border router of an AS may compose a list of segments to steer a flow along a selected path within the AS, towards a selected egress border router C of the AS and through a specific peer. At minimum, a BGP Peering Engineering policy applied at an ingress PE involves two segments: the Node SID of the chosen egress PE and then the BGP Peering Segment for the chosen egress PE peer or peering interface.

Hereafter, we will define three types of BGP peering segments/SID's: PeerNodeSID, PeerAdjSID and PeerSetSID.

- o PeerNode SID. A BGP PeerNode segment/SID is a local segment. At the BGP node advertising it, its semantics is:
  - \* SR header operation: NEXT.
  - \* Next-Hop: the connected peering node to which the segment is related.
- o PeerAdj SID: A BGP PeerAdj segment/SID is a local segment. At the BGP node advertising it, its semantics is:





- \* SR header operation: NEXT.
  - \* Next-Hop: the peer connected through the interface to which the segment is related.
- o PeerSet SID. A BGP PeerSet segment/SID is a local segment. At the BGP node advertising it, its semantics is:
- \* SR header operation: NEXT.
  - \* Next-Hop: loadbalance across any connected interface to any peer in the related group.

A peer set could be all the connected peers from the same AS or a subset of these. A group could also span across AS. The group definition is a policy set by the operator.

The BGP extensions necessary in order to signal these BGP peering segments will be defined in a separate document.

## 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.

When a given node B wants to provide egress node A protection, it advertises a segment identifying node's A context. Such segment is called "Mirror Context Segment" and identified by the Mirror SID.

The Mirror SID is advertised using the Binding Segment defined in SR IGP protocol extensions ( [[I-D.ietf-isis-segment-routing-extensions](#)], [[I-D.ietf-ospf-segment-routing-extensions](#)] and [[I-D.ietf-ospf-ospfv3-segment-routing-extensions](#)]).

In the event of a failure, a point of local repair (PLR) diverting traffic from A to B does a PUSH of the Mirror SID on the protected traffic. B, when receiving the traffic with the Mirror SID as the active segment, uses that segment and process underlying segments in the context of A.

## 6. Multicast

Segment Routing is defined for unicast. The application of the source-route concept to Multicast is not in the scope of this document.



## **7. IANA Considerations**

This document does not require any action from IANA.

## **8. Security Considerations**

This document doesn't introduce new security considerations when applied to the MPLS dataplane.

There are a number of security concerns with source routing at the IPv6 dataplane [[RFC5095](#)]. The new IPv6-based segment routing header defined in [[I-D.ietf-6man-segment-routing-header](#)] and its associated security measures address these concerns. The IPv6 Segment Routing Header is defined in a way that blind attacks are never possible, i.e., attackers will be unable to send source routed packets that get successfully processed, without being part of the negotiations for setting up the source routes or being able to eavesdrop legitimate source routed packets. In some networks this base level security may be complemented with other mechanisms, such as packet filtering, cryptographic security, etc.

## **9. Contributors**

The following people have substantially contributed to the definition of the Segment Routing architecture and to the editing of this document:

Ahmed Bashandy  
Cisco Systems, Inc.  
Email: bashandy@cisco.com

Martin Horneffer  
Deutsche Telekom  
Email: Martin.Horneffer@telekom.de

Wim Henderickx  
Alcatel-Lucent  
Email: wim.henderickx@alcatel-lucent.com

Jeff Tantsura  
Ericsson  
Email: Jeff.Tantsura@ericsson.com

Edward Crabbe  
Individual  
Email: edward.crabbe@gmail.com



Igor Milojevic  
Email: milojevicigor@gmail.com

Saku Ytti  
TDC  
Email: saku@ytti.fi

## **10. Acknowledgements**

We would like to thank Dave Ward, Dan Frost, Stewart Bryant, Pierre Francois, Thomas Telkamp, Les Ginsberg, Ruediger Geib, Hannes Gredler, Pushpasis Sarkar, Eric Rosen and Chris Bowers for their comments and review 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](#), DOI 10.17487/RFC2119, March 1997, <<http://www.rfc-editor.org/info/rfc2119>>.
- [RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", [RFC 2460](#), DOI 10.17487/RFC2460, December 1998, <<http://www.rfc-editor.org/info/rfc2460>>.
- [RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", [RFC 3031](#), DOI 10.17487/RFC3031, January 2001, <<http://www.rfc-editor.org/info/rfc3031>>.
- [RFC4206] Kompella, K. and Y. Rekhter, "Label Switched Paths (LSP) Hierarchy with Generalized Multi-Protocol Label Switching (GMPLS) Traffic Engineering (TE)", [RFC 4206](#), DOI 10.17487/RFC4206, October 2005, <<http://www.rfc-editor.org/info/rfc4206>>.

### **11.2. Informative References**

- [I-D.filsfils-spring-large-scale-interconnect]  
Filsfils, C., Cai, D., Previdi, S., Henderickx, W., Shakir, R., Cooper, D., Ferguson, F., Laberge, T., Lin, S., Decraene, B., Jalil, L., and J. Tantsura, "Interconnecting Millions Of Endpoints With Segment Routing", [draft-filsfils-spring-large-scale-interconnect-01](#) (work in progress), November 2015.



[I-D.francois-rtgwg-segment-routing-ti-lfa]

Francois, P., Filsfils, C., Bashandy, A., and B. Decraene, "Topology Independent Fast Reroute using Segment Routing", [draft-francois-rtgwg-segment-routing-ti-lfa-00](#) (work in progress), August 2015.

[I-D.ietf-6man-segment-routing-header]

Previdi, S., Filsfils, C., Field, B., Leung, I., Linkova, J., Kosugi, T., Vyncke, E., and D. Lebrun, "IPv6 Segment Routing Header (SRH)", [draft-ietf-6man-segment-routing-header-00](#) (work in progress), December 2015.

[I-D.ietf-isis-prefix-attributes]

Ginsberg, L., Decraene, B., Filsfils, C., Litkowski, S., Previdi, S., Xu, X., and U. Chunduri, "IS-IS Prefix Attributes for Extended IP and IPv6 Reachability", [draft-ietf-isis-prefix-attributes-02](#) (work in progress), December 2015.

[I-D.ietf-isis-segment-routing-extensions]

Previdi, S., Filsfils, C., Bashandy, A., Gredler, H., Litkowski, S., Decraene, B., and J. Tantsura, "IS-IS Extensions for Segment Routing", [draft-ietf-isis-segment-routing-extensions-06](#) (work in progress), December 2015.

[I-D.ietf-ospf-ospfv3-segment-routing-extensions]

Psenak, P., Previdi, S., Filsfils, C., Gredler, H., Shakir, R., Henderickx, W., and J. Tantsura, "OSPFv3 Extensions for Segment Routing", [draft-ietf-ospf-ospfv3-segment-routing-extensions-03](#) (work in progress), June 2015.

[I-D.ietf-ospf-segment-routing-extensions]

Psenak, P., Previdi, S., Filsfils, C., Gredler, H., Shakir, R., Henderickx, W., and J. Tantsura, "OSPF Extensions for Segment Routing", [draft-ietf-ospf-segment-routing-extensions-05](#) (work in progress), June 2015.

[I-D.ietf-pce-segment-routing]

Sivabalan, S., Medved, J., Filsfils, C., Crabbe, E., Lopez, V., Tantsura, J., Henderickx, W., and J. Hardwick, "PCEP Extensions for Segment Routing", [draft-ietf-pce-segment-routing-06](#) (work in progress), August 2015.





[I-D.ietf-spring-ipv6-use-cases]

Brzozowski, J., Leddy, J., Leung, I., Previdi, S., Townsley, W., Martin, C., Filsfils, C., and R. Maglione, "IPv6 SPRING Use Cases", [draft-ietf-spring-ipv6-use-cases-05](#) (work in progress), September 2015.

[I-D.ietf-spring-oam-usecase]

Geib, R., Filsfils, C., Pignataro, C., and N. Kumar, "Use Case for a Scalable and Topology-Aware Segment Routing MPLS Data Plane Monitoring System", [draft-ietf-spring-oam-usecase-01](#) (work in progress), October 2015.

[I-D.ietf-spring-problem-statement]

Previdi, S., Filsfils, C., Decraene, B., Litkowski, S., Horneffer, M., and r. rjs@rob.sh, "SPRING Problem Statement and Requirements", [draft-ietf-spring-problem-statement-06](#) (work in progress), December 2015.

[I-D.ietf-spring-resiliency-use-cases]

Francois, P., Filsfils, C., Decraene, B., and r. rjs@rob.sh, "Use-cases for Resiliency in SPRING", [draft-ietf-spring-resiliency-use-cases-02](#) (work in progress), December 2015.

[I-D.ietf-spring-segment-routing-central-epe]

Filsfils, C., Previdi, S., Ginsburg, D., and D. Afanasiev, "Segment Routing Centralized Egress Peer Engineering", [draft-ietf-spring-segment-routing-central-epe-00](#) (work in progress), October 2015.

[I-D.ietf-spring-segment-routing-ldp-interop]

Filsfils, C., Previdi, S., Bashandy, A., Decraene, B., and S. Litkowski, "Segment Routing interoperability with LDP", [draft-ietf-spring-segment-routing-ldp-interop-00](#) (work in progress), October 2015.

[I-D.ietf-spring-segment-routing-mpls]

Filsfils, C., Previdi, S., Bashandy, A., Decraene, B., Litkowski, S., Horneffer, M., rjs@rob.sh, r., Tantsura, J., and E. Crabbe, "Segment Routing with MPLS data plane", [draft-ietf-spring-segment-routing-mpls-02](#) (work in progress), October 2015.

[I-D.ietf-spring-segment-routing-msdc]

Filsfils, C., Previdi, S., Mitchell, J., and P. Lapukhov, "BGP-Prefix Segment in large-scale data centers", [draft-ietf-spring-segment-routing-msdc-00](#) (work in progress), October 2015.



## [I-D.ietf-spring-sr-oam-requirement]

Kumar, N., Pignataro, C., Akiya, N., Geib, R., Mirsky, G., and S. Litkowski, "OAM Requirements for Segment Routing Network", [draft-ietf-spring-sr-oam-requirement-00](#) (work in progress), June 2015.

[RFC4915] Psenak, P., Mirtorabi, S., Roy, A., Nguyen, L., and P. Pillay-Esnault, "Multi-Topology (MT) Routing in OSPF", [RFC 4915](#), DOI 10.17487/RFC4915, June 2007, <<http://www.rfc-editor.org/info/rfc4915>>.

[RFC5095] Abley, J., Savola, P., and G. Neville-Neil, "Deprecation of Type 0 Routing Headers in IPv6", [RFC 5095](#), DOI 10.17487/RFC5095, December 2007, <<http://www.rfc-editor.org/info/rfc5095>>.

[RFC5120] Przygienda, T., Shen, N., and N. Sheth, "M-ISIS: Multi Topology (MT) Routing in Intermediate System to Intermediate Systems (IS-ISs)", [RFC 5120](#), DOI 10.17487/RFC5120, February 2008, <<http://www.rfc-editor.org/info/rfc5120>>.

[RFC6549] Lindem, A., Roy, A., and S. Mirtorabi, "OSPFv2 Multi-Instance Extensions", [RFC 6549](#), DOI 10.17487/RFC6549, March 2012, <<http://www.rfc-editor.org/info/rfc6549>>.

[RFC6822] Previdi, S., Ed., Ginsberg, L., Shand, M., Roy, A., and D. Ward, "IS-IS Multi-Instance", [RFC 6822](#), DOI 10.17487/RFC6822, December 2012, <<http://www.rfc-editor.org/info/rfc6822>>.

## Authors' Addresses

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

Email: [cfilsfil@cisco.com](mailto:cfilsfil@cisco.com)

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

Email: [sprevidi@cisco.com](mailto:sprevidi@cisco.com)



Bruno Decraene  
Orange  
FR

Email: [bruno.decraene@orange.com](mailto:bruno.decraene@orange.com)

Stephane Litkowski  
Orange  
FR

Email: [stephane.litkowski@orange.com](mailto:stephane.litkowski@orange.com)

Rob Shakir  
Jive Communications, Inc.  
1275 West 1600 North, Suite 100  
Orem, UT 84057

Email: [rjs@rob.sh](mailto:rjs@rob.sh)

