Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: November 14, 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.
May 13, 2014

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

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 extension 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 extension header. The segment to

process is indicated by a pointer in the routing extension header. Upon completion of a segment, the pointer is incremented.

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 will expire on November 14, 2014.

Copyright Notice

Table of Contents

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 (RFC 3031) 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.

Segment Routing can be applied to the IPv6 architecture (RFC2460), with a new type of routing extension 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 extension header. The

active segment is indicated by a pointer in the routing extension header. Upon completion of a segment, the pointer is incremented. A segment can be inserted in the list and the pointer is updated accordingly.

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.filsfils-spring-segment-routing-use-cases], [I-D.martin-spring-segment-routing-ipv6-use-cases], [I-D.francois-spring-resiliency-use-case], [I-D.geib-spring-oam-usecase] and [I-D.kumar-spring-sr-oam-requirement].

Segment Routing for MPLS dataplane is documented in [I-D.filsfils-spring-segment-routing-mpls].

Segment Routing for IPv6 dataplane is documented in [I-D.previdi-6man-segment-routing-header].

IGP protocol extensions for Segment Routing are described in [I-D.previdi-isis-segment-routing-extensions] and [I-D.psenak-ospf-segment-routing-extensions]

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

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

2. Terminology

Segment: a segment identifies an instruction

SID: a Segment Identifier

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.

Active segment: the segment that MUST be used by the receiving router to process the packet. It is identified by a pointer in the IPv6 architecture. It is the top label in the MPLS architecture.

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.

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. In the IPv6 architecture, it is the set of locally relevant IPv6 addresses.

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 is a link-local address.

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 always global within the SR/IGP domain and identifies an instruction to

forward the packet over 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-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. An IGP-Adjacency Segment is local 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 constrained-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.sivabalan-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 ([I-D.filsfils-spring-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.

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 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.

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

While SR 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, the segment is global (the SID is allocated from the SRGB). This is consistent with [I-D.filsfils-spring-segment-routing-use-cases] as all the described use-cases require global segments attached to IGP prefixes.

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 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.

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

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
```

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 shortest-path to prefix R.
```

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 N flag is set. The terms "Node Segment" or "Node-SID" are often used as an abbreviation.

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 as explained in [I-D.filsfils-spring-segment-routing-use-cases].

An IGP-Node-SID MUST NOT be associated with a prefix that is owned or advertised 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 as described in [I-D.filsfils-spring-segment-routing-use-cases].

The Anycast SID MUST be advertised with the N-flag unset.

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. 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 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.

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 as explained in [I-D.filsfils-spring-segment-routing-use-cases].

The encodings of the Adj-SID include the B-flag. When set, the Adj-SID benefits from a local protection.

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 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

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 [I-D.previdi-isis-segment-routing-extensions] and [I-D.psenak-ospf-segment-routing-extensions].

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.filsfils-spring-segment-routing-ldp-interop].

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

3.6.2. Tunnel Headend

The segment allocation and SRDB 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 would consist in representing tunnels as forwarding-adjacencies ([RFC4206]). The Remote-Binding SID is preferred as it allows to advertise the presence of a tunnel without influencing the LSDB and the SPF computation.

3.6.3. Mirroring Context

TBD.

3.7. Inter-Area Considerations

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

```
              !          !
              !          !
       B------C-----F----G-----K
      /       |          |     |
S---A/        |          |     |
     \        |          |     |
      \D------I----------J-----L----Z (192.0.2.1/32, Node-SID: 150)
              !          !
      Area-1  ! Backbone ! Area 2
              !   area   !
```

Figure 1: 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. Multicast

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

5. IANA Considerations

TBD

6. Manageability Considerations

TBD

7. Security Considerations

TBD

8. 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.

9. References

9.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[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.

9.2. Informative References

[I-D.filsfils-spring-segment-routing-ldp-interop]
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 interoperability with LDP", draft-filsfils-spring-segment-routing-ldp-interop-01 (work in progress), April 2014.

[I-D.filsfils-spring-segment-routing-mpls]
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 with MPLS data plane", draft-filsfils-spring-segment-routing-mpls-01 (work in progress), April 2014.

[I-D.filsfils-spring-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., Kini, S., and E. Crabbe, "Segment Routing Use Cases", draft-filsfils-spring-segment-routing-use-cases-00 (work in progress), March 2014.

[I-D.francois-segment-routing-ti-lfa]
Francois, P., Filsfils, C., Bashandy, A., and B. Decraene, "Topology Independent Fast Reroute using Segment Routing", draft-francois-segment-routing-ti-lfa-00 (work in progress), November 2013.

[I-D.francois-spring-resiliency-use-case]
Francois, P., Filsfils, C., Decraene, B., and R. Shakir, "Use-cases for Resiliency in SPRING", draft-francois-spring-resiliency-use-case-02 (work in progress), April 2014.

[I-D.geib-spring-oam-usecase]
Geib, R. and C. Filsfils, "Use case for a scalable and topology aware MPLS data plane monitoring system", draft-geib-spring-oam-usecase-01 (work in progress), February 2014.

[I-D.kumar-spring-sr-oam-requirement]
Kumar, N., Pignataro, C., Akiya, N., Geib, R., and G. Mirsky, "OAM Requirements for Segment Routing Network", draft-kumar-spring-sr-oam-requirement-00 (work in progress), February 2014.

[I-D.martin-spring-segment-routing-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-martin-spring-segment-routing-ipv6-use-cases-01 (work in progress), April 2014.

[I-D.previdi-6man-segment-routing-header]
Previdi, S., Filsfils, C., Field, B., and I. Leung, "IPv6 Segment Routing Header (SRH)", draft-previdi-6man-segment-routing-header-00 (work in progress), March 2014.

[I-D.previdi-isis-segment-routing-extensions]
Previdi, S., Filsfils, C., Bashandy, A., Gredler, H., Litkowski, S., and J. Tantsura, "IS-IS Extensions for Segment Routing", draft-previdi-isis-segment-routing-extensions-05 (work in progress), February 2014.

[I-D.psenak-ospf-segment-routing-extensions]
Psenak, P., Previdi, S., Filsfils, C., Gredler, H., Shakir, R., and W. Henderickx, "OSPF Extensions for Segment Routing", draft-psenak-ospf-segment-routing-extensions-04 (work in progress), February 2014.

[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.

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