

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: December 19, 2013

S. Sivabalan
C. Filsfils
J. Medved
Cisco Systems, Inc.
E. Crabbe
Google, Inc.
R. Raszuk
NTT I3
June 17, 2013

PCE-Initiated Traffic Engineering Path Setup in Segment Routed Networks
[draft-sivabalan-pce-segment-routing-00.txt](#)

Abstract

Segment Routing (SR) enables any head-end node to select any path without relying on a hop-by-hop signaling technique (e.g., LDP or RSVP-TE). It depends only on "segments" that are advertised by Link-State Interior Gateway Protocols (IGPs). A Segment Routed Path can be derived from a variety of mechanisms, including an IGP Shortest Path Tree (SPT), explicit configuration, or a stateful Path Computation Element (PCE). This document specifies extensions to the Path Computation Element Protocol (PCEP) that allow a stateful PCE to signal and instantiate Traffic Engineering paths in SR networks.

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 [[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 December 19, 2013.

Copyright Notice

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

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

Table of Contents

1.	Introduction	3
2.	Terminology	4
3.	Overview of PCEP Operation in SR Networks	4
4.	SR Capability Negotiation	5
4.1.	Negotiating SR Capability	6
5.	SR-ERO objects	6
5.1.	The SR-ERO Subobject	6
5.2.	NAI Associated with SID	8
5.3.	SR-ERO Processing	10
6.	The SR-RR0 Object	10
7.	SR Specific PCEP Descriptors	11
7.1.	PCEP Descriptor for PCCreate Message	11
7.2.	PCEP Descriptor for PCRpt Message	11
7.3.	PCEP Descriptor for PCUpd Message	12
8.	Backward Compatibility	12
9.	Management Considerations	12
9.1.	Policy	12
9.2.	The PCEP Data Model	13
10.	Security Considerations	13
11.	IANA Considerations	13
12.	Acknowledgement	13
13.	References	13
13.1.	Normative References	13
13.2.	Informative References	15
	Authors' Addresses	15

1. Introduction

In SR networks, a path is chosen without relying on hop-by-hop signaling protocols such as LDP or RSVP-TE. Each path is specified as a set of "segments" advertised by link-state routing protocols (IS-IS or OSPF). [[SR-ARCH](#)] provides an introduction to SR technology. The corresponding IS-IS extension and OSPF extension are specified respectively in [[SR-ISIS](#)] and a document to be published soon. Two types of segments have been defined; nodal and adjacency segments. A nodal segment represents a path to a node, whereas an adjacency segment represents a specific adjacency to a node. A Segment Identifier (SID) is a 32-bit value and can be applied to both IP and MPLS data planes. In the case of the MPLS data plane, an SR path corresponds to an MPLS Label Switching Path (LSP).

A Segment Routed path (SR path) can be derived from an IGP Shortest Path Tree (SPT). Segment Routed Traffic Engineering paths (SR-TE paths) may not follow IGP SPT. Such paths may be chosen by a suitable network planning tool and provisioned on the source node of the SR-TE path.

[RFC5440] describes Path Computation Element Protocol (PCEP) for communication between a Path Computation Client (PCC) and a Path Control Element (PCE) or between one or a pair of PCEs. A PCE computes paths for MPLS Traffic Engineering LSPs (MPLS-TE LSPs) based on various constraints and optimization criteria.

[[I-D.ietf-pce-stateful-pce](#)] specifies extensions to PCEP that allow a stateful PCE to compute and recommend network paths in compliance with [[RFC4657](#)]. Stateful PCEP extensions provide synchronization of LSP state between a PCC and a PCE or between a pair of PCEs, delegation of LSP control, reporting of LSP state from a PCC to a PCE, controlling the setup and path routing of an LSP from a PCE to a PCC. Stateful PCEP extensions are intended for an operational model in which LSPs are configured on the PCC, and control over them is delegated to the PCE. [[I-D.crabbe-pce-stateful-pce-mpls-te](#)] builds on [[I-D.ietf-pce-stateful-pce](#)] to specify objects and TLVs for

MPLS-TE LSPs. Moreover, a mechanism to dynamically instantiate LSPs on a PCC based on the requests from a stateful PCE or a controller using stateful PCE is specified in [\[I-D.crabbe-pce-pce-initiated-lsp\]](#). Such mechanism is useful in Software Driven Networks (SDN) applications, such as demand engineering, or bandwidth calendaring.

It is possible to use a stateful PCE for computing one or more SR-TE paths taking into account various constraints and objective functions. Once a path is chosen, the stateful PCE can instantiate an SR-TE path on the PCC using PCEP extensions specified in [\[I-D.crabbe-pce-pce-initiated-lsp\]](#) along with the SR specific PCEP

extensions provided in this document.

[2.](#) Terminology

The following terminologies are used in this document:

ERO: Explicit Route Object.

IGP: Interior Gateway Protocol.

IS-IS: Intermediate System to Intermediate System.

LSR: Label Switching Router.

NAI: Node or Adjacency Identifier.

OSPF: Open Shortest Path First.

PCC: Path Computation Client.

PCE: Path Computation Element

PCEP: Path Computation Element Protocol.

RR0: Record Route Object.

SR: Segment Routing.

SR-ERO: Segment Routed Explicit Route Object.

SR Path: Segment Routed Path.

SR-RRO: Segment Routed Record Route Object.

SR-TE Path: Segment Routed Traffic Engineering Path.

[3.](#) Overview of PCEP Operation in SR Networks

In SR networks, an ingress node of an SR path appends all outgoing packets with an SR header consisting of a list of Segment IDs (SIDs). The header has all necessary information to guide the packets from the ingress node to the egress node of the path, and hence there is no need for any signaling protocol. A SID can represent a nodal segment representing a path to a node or adjacency segment representing path over a specific adjacency.

In a PCEP session, path information is carried in the Explicit Route Object (ERO), which consists of a sequence of subobjects. Various types of ERO subobjects have been specified in [\[RFC3209\]](#), [\[RFC3473\]](#), and [\[RFC3477\]](#). In SR networks, a PCE needs to specify ERO containing SIDs, and a PCC should be capable of processing such ERO. An ERO containing SIDs can be included in the PCEP Create message (PCCreate) defined in [\[I-D.crabbe-pce-pce-initiated-lsp\]](#), as well as in the PCEP Update (PCUpd) and PCEP Report (PCRpt) messages defined in [\[I-D.ietf-pce-stateful-pce\]](#).

When a PCEP session between a PCC and a PCE is established, both PCEP Speakers exchange information to indicate their ability to support SR-specific functionality. A PCEP session can carry EROs of different types. However, an ERO carrying SIDs MUST NOT include any other form of EROs, i.e., all subobjects within an ERO MUST represent SID. Furthermore, if an SR path is established using SR-ERO, subsequent PCEP Update and Report messages for that path MUST NOT contain other ERO types. This document specifies new error codes to handle these errors. Should the need to change the ERO type arise, the SR path must be deleted and re-created using a new ERO type.

A PCC can include an RRO object in a PCRpt message. In SR networks,

a PCC MAY learn the SR actual path actually taken by data packets and report that path to a PCE. For this purpose, we introduce a new RRO subobject type. However, methods used by a PCC to learn SR-TE paths are outside the scope of this document.

In summary, this document specifies a new PCEP capability, a new ERO subobject, a new RRO subobject, and new PCEP error codes.

4. SR Capability Negotiation

The format of the STATEFUL-PCE-CAPABILITY TLV is shown in the following figure:



Figure 1: STATEFUL-PCE-CAPABILITY TLV format

The type of the TLV is to be defined in [[I-D.ietf-pce-stateful-pce](#)]. The TLV length is 4 octets.

The value comprises a single field - Flags (32 bits). The U and S flags are defined in [[I-D.ietf-pce-stateful-pce](#)]. The I flag is defined in [[I-D.crabbe-pce-pce-initiated-lsp](#)].

A new flag - the "R bit" - is used to negotiate the SR capability between a PCE and a PCC.

4.1. Negotiating SR Capability

By setting the R-bit to 1, a PCEP Speaker indicates that it is SR-capable, i.e., it is able to perform SR related PCEP functions specified in this document. An SR-capable PCEP Speaker must be able to handle EROs containing SIDs (referred to as "SR-EROs" in this document). A PCEP Speaker MAY originate SR-ERO only if both PCEP Speakers in a PCEP session are SR-capable. A pair of SR-capable PCEP

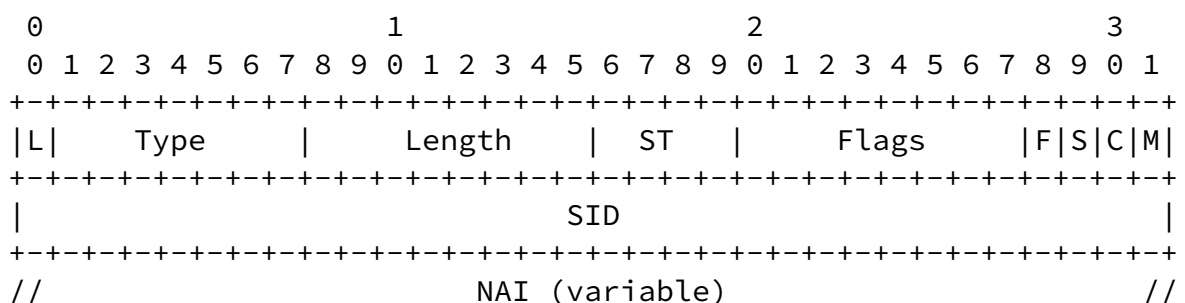
Speakers can exchange SR-EROs in PCCreate and PCUpd messages. Similarly, an SR-capable PCEP Speaker must be able to handle RROs containing SIDs (referred to as "SR-RRO" in this document). A pair of PCEP Speakers can exchange SR-RROs in PCRpt message.

5. SR-ERO objects

An SR-TE path consists of one or more SID(s) where each SID is associated with the identifier that represents the node or adjacency corresponding to the SID. This identifier is referred to as the 'Node or Adjacency Identifier' (NAI). As described later, a NAI can be represented in various formats (e.g., IPv4 address, IPv6 address, etc). Furthermore, a NAI is used only for troubleshooting purposes, and MUST not be used to replace or modify any fields in a data packet header. An SR-ERO object consists of one or more ERO subobjects described in the following section.

5.1. The SR-ERO Subobject

An SR-ERO subobject consists of a 32-bit header followed by the SID and the NAI associated with the SID. The SID is a 32-bit number. The size of the NAI depends on its respective type, as described in the following sections.



value represents an MPLS label stack entry as specified in [[RFC5462](#)], where all the entry's fields (Label, TC, S, and TTL) are specified by the PCE. However, a PCC MAY choose to override TC, S, and TTL values according its local policy and MPLS forwarding rules.

- * S: When this bit is set, the SID value in the subobject body is null. In this case, the PCC is responsible for choosing the SID value, e.g., by looking up its Traffic Engineering Database (TED) using node/adjacency identifier in the subobject body.
- * F: When this bit is set, the NAI value in the subobject body is null.

SID is the Segment Identifier.

NAI contains the NAI associated with the SID. Depending on the value of ST, the NAI can have different format as described in the following section.

[5.2](#). NAI Associated with SID

This document defines the following NAIs:

'IPv4 Node ID' is specified as an IPv4 address. In this case, ST and Length are 1 and 12 respectively.

'IPv6 Node ID' is specified as an IPv6 address. In this case, ST and Length are 2 and 20 respectively.

'IPv4 Adjacency' is specified as a pair of IPv4 addresses. In this case, ST and Length are 3 and 16 respectively, and the format of the NAI is shown in the following figure:

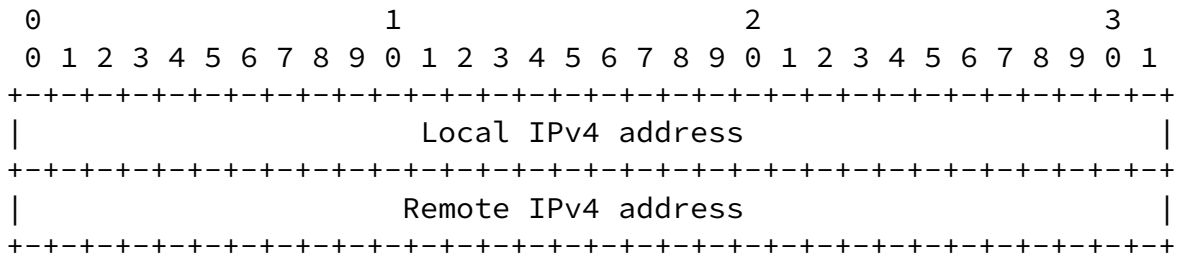


Figure 3: NAI for IPv4 Adjacency

'IPv6 Adjacency' is specified as a pair of IPv6 addresses. In this case, ST and Length are 4 and 34 respectively, and the format of the NAI is shown in the following figure:

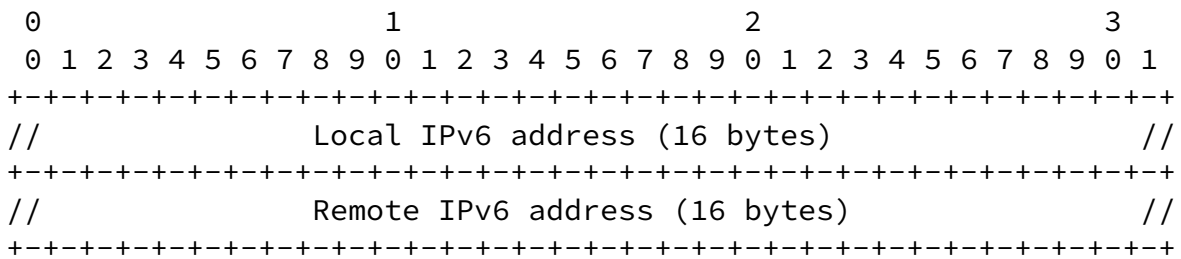


Figure 4: NAI for IPv6 adjacency

'Unnumbered Adjacency with IPv4 NodeIDs' is specified as a pair of Node ID / Interface ID tuples. In this case, ST and Length are 5 and 24 respectively, and the format of the NAI is shown in the following figure:

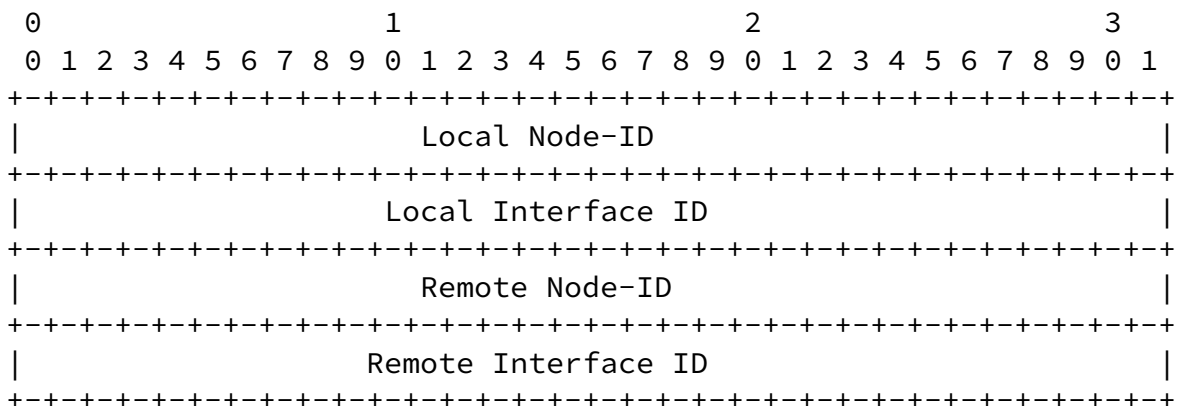


Figure 5: NAI for Unnumbered adjacency with IPv4 Node IDs

We are yet to decide if another SID subobject is required for

unnumbered adjacency with 128 bit node ID.

[5.3.](#) SR-ERO Processing

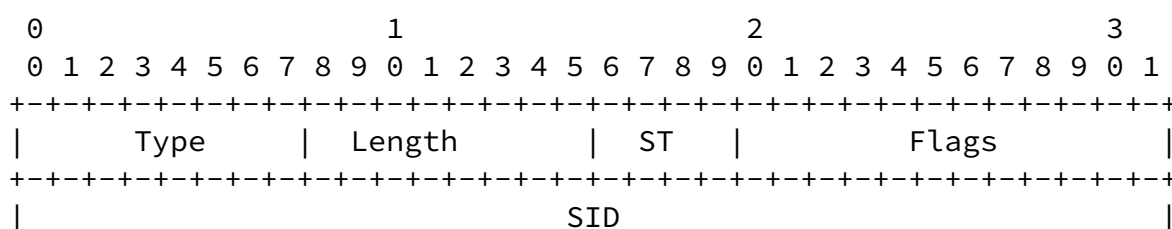
A PCEP Speaker that does not recognize the new ERO subobject in the PCCreate, PCUpd or PCRpt message MUST reject the entire PCEP message and MUST send a PCE error message with Error-Type="Unknown Object" or "Not supported object", defined in [[RFC5440](#)].

When the SID represents an MPLS label (i.e. the M bit is set), its value (20 most significant bits) MUST be larger than 15, unless it is special purpose label, such as an Entropy Label Indicator (ELI) or an Entropy Label (EL). If a PCEP Speaker receives a label ERO subobject with an invalid value, it MUST send the PCE error message with Error-Type = "Reception of an invalid object" and Error-Value = "Bad label value". If both M and C bits of an ERO subobject are set, and if a PCEP Speaker finds erroneous setting in one or more of TC, S, and TTL fields, it MUST send a PCE error with Error-Type = "Reception of an invalid object" and Error-Value = "Bad label format".

If a PCC receives a stack of SR-ERO subobjects, and the number of stack exceeds the maximum number of SIDs that the PCC can impose on the packet, it MAY send a PCE error with Error-Type = "Reception of an invalid object" and Error-Value = "Unsupported number of Segment ERO subobjects".

[6.](#) The SR-RR0 Object

An SR-RR0 objects consists of one or more subobject(s), each carrying a SID and the associated NAI. The format of the SR-RR0 subobject is the same as the ERO subobject, except for L-flag which does not exists in the SR-RR0 Subobject. Also, there are no flags defined for the SR-RR0 subobject.



```

+-----+-----+-----+-----+-----+-----+-----+-----+-----+
//                                         NAI (variable)                               //
+-----+-----+-----+-----+-----+-----+-----+-----+-----+

```

Figure 6: SR-RR0 Subobject format

7. SR Specific PCEP Descriptors

As defined in [RFC5440], a PCEP message consists of a common header followed by a variable length body made up of mandatory and/or optional objects. [I-D.ietf-pce-stateful-pce] and [I-D.crabbe-pce-stateful-pce-mpls-te] respectively describe the stateful PCEP descriptors for general state control and MPLS-TE specific state control. Finally, [I-D.crabbe-pce-pce-initiated-lsp] specifies the PCEP descriptors for PCE-based LSP instantiation. This document specifies the SR specific PCEP descriptors.

7.1. PCEP Descriptor for PCCreate Message

The format of the PCCreate message is as follows:

```

<PCCreate Message> ::= <Common Header>
                        <sr-path-instantiation-list>

```

Where:

```

<sr-path-instantiation-list> ::=
    <sr-path-instantiation-request> [<sr-path-instantiation-list>]

<sr-path-instantiation-request> ::= <LSP>
                                     <SID-ERO> <----- New subobject

```

The LSP object in the Common Header MUST include the SYMBOLIC-PATH-NAME TLV.

7.2. PCEP Descriptor for PCRpt Message

The format of the PCRpt message is as follows:

In the PCUpd message, only the LSP object is considered mandatory. All other objects are optional, and may be included for synchronizing SR-TE paths when PCEP session is re-established.

8. Backward Compatibility

An LSR that does not support the SR PCEP capability negotiation cannot recognize the SR-ERO subobjects. As such, it shall send a PCEP error with Error-Type = 4 (Not supported object) and Error-Value = 2 (Not supported object Type) as per [[RFC5440](#)].

9. Management Considerations

9.1. Policy

PCEP implementation:

- o Can enable SR-PCEP capability either by default or via explicit configuration.

Sivabalan, et al. Expires December 19, 2013 [Page 12]

Internet-Draft PCEP Extensions for Segment Routing June 2013

- o May generate PCEP error due to unsupported number of SR-ERO subobjects either by default or via explicit configuration.

9.2. The PCEP Data Model

A PCEP MIB module is defined in [[I-D.ietf-pce-pcep-mib](#)] needs be extended to cover additional functionality provided by [[RFC5440](#)] and [[I-D.crabbe-pce-pce-initiated-lsp](#)]. Such extension will cover the new functionality specified in this document.

10. Security Considerations

The security considerations described in [[RFC5440](#)] and [[I-D.crabbe-pce-pce-initiated-lsp](#)] are applicable to this specification. No additional security measure is required.

[11.](#) IANA Considerations

IANA is requested to allocate a ERO subobject type (recommended value = 5) for the SR-ERO subobject.

This document also defines new Error-Values for the following new error conditions:

Error-Type	Meaning
10	Reception of an invalid object
	Error-value=2: Bad label value
	Error-value=3: Unsupported number of Segment ERO subobjects

[12.](#) Acknowledgement

We like to thank George Swallow for the valuable comments.

[13.](#) References

[13.1.](#) Normative References

[I-D.crabbe-pce-pce-initiated-lsp]

Crabbe, E., Minei, I., Sivabalan, S., and R. Varga, "PCEP Extensions for PCE-initiated LSP Setup in a Stateful PCE Model", [draft-crabbe-pce-pce-initiated-lsp-00](#) (work in progress), October 2012.

Sivabalan, et al. Expires December 19, 2013 [Page 13]

Internet-Draft PCEP Extensions for Segment Routing June 2013

[I-D.crabbe-pce-stateful-pce-mpls-te]

Crabbe, E., Medved, J., Minei, I., and R. Varga, "Stateful PCE extensions for MPLS-TE LSPs", [draft-crabbe-pce-stateful-pce-mpls-te-01](#) (work in progress), May 2013.

[I-D.ietf-pce-pcep-mib]

Koushik, K., Stephan, E., Zhao, Q., King, D., and J. Hardwick, "PCE communication protocol (PCEP) Management Information Base", [draft-ietf-pce-pcep-mib-04](#) (work in progress), February 2013.

[I-D.ietf-pce-stateful-pce]

Crabbe, E., Medved, J., Minei, I., and R. Varga, "PCEP Extensions for Stateful PCE", [draft-ietf-pce-stateful-pce-04](#) (work in progress), May 2013.

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

[RFC5088] Le Roux, J.L., Vasseur, J.P., Ikejiri, Y., and R. Zhang, "OSPF Protocol Extensions for Path Computation Element (PCE) Discovery", [RFC 5088](#), January 2008.

[RFC5089] Le Roux, J.L., Vasseur, J.P., Ikejiri, Y., and R. Zhang, "IS-IS Protocol Extensions for Path Computation Element (PCE) Discovery", [RFC 5089](#), January 2008.

[RFC5440] Vasseur, J.P. and J.L. Le Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)", [RFC 5440](#), March 2009.

[RFC5462] Andersson, L. and R. Asati, "Multiprotocol Label Switching (MPLS) Label Stack Entry: "EXP" Field Renamed to "Traffic Class" Field", [RFC 5462](#), February 2009.

[SR-ARCH] Filsfils, C., Previdi, S., Bashandy, A., and M. Horneffer, "Segment Routing Architecture", [draft-filsfils-previdi-segment-routing-00.txt](#) (work in progress), June 2013.

[SR-ISIS] Previdi, S., Filsfils, C., Bashandy, A., and M. Horneffer, "Segment Routing with IS-IS Routing Protocol", [draft-previdi-filsfils-isis-segment-routing-00.txt](#) (work in progress), September 2013.

[13.2.](#) Informative References

[RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP

Tunnels", [RFC 3209](#), December 2001.

- [RFC3473] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReSerVation Protocol-Traffic Engineering (RSVP-TE) Extensions", [RFC 3473](#), January 2003.
- [RFC3477] Kompella, K. and Y. Rekhter, "Signalling Unnumbered Links in Resource ReSerVation Protocol - Traffic Engineering (RSVP-TE)", [RFC 3477](#), January 2003.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", [RFC 4655](#), August 2006.
- [RFC4657] Ash, J. and J. Le Roux, "Path Computation Element (PCE) Communication Protocol Generic Requirements", [RFC 4657](#), September 2006.
- [RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", [RFC 5305](#), October 2008.
- [RFC5394] Bryskin, I., Papadimitriou, D., Berger, L., and J. Ash, "Policy-Enabled Path Computation Framework", [RFC 5394](#), December 2008.
- [RFC5557] Lee, Y., Le Roux, J.L., King, D., and E. Oki, "Path Computation Element Communication Protocol (PCEP) Requirements and Protocol Extensions in Support of Global Concurrent Optimization", [RFC 5557](#), July 2009.

Authors' Addresses

Siva Sivabalan
Cisco Systems, Inc.
2000 Innovation Drive
Kanata, Ontario K2K 3E8
Canada

Email: msiva@cisco.com

Clarence Filsfils
Cisco Systems, Inc.
Pegasus Parc
De kleetlaan 6a, DIEGEM BRABANT 1831
BELGIUM

Email: cfilsfil@cisco.com

Jan Medved
Cisco Systems, Inc.
170 West Tasman Dr.
San Jose, CA 95134
US

Email: jmedved@cisco.com

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

Email: edward.crabbe@gmail.com

Robert Raszuk
NTT I3
101 S. Ellsworth Ave
San Mateo, CA 94401
US

Email: robert@raszuk.net

