

Network Working Group
Internet Draft
Intended Status: Proposed Standard
Expires: October 16, 2010

IJsbrand Wijnands
Eric C. Rosen
Cisco Systems, Inc.

Maria Napierala
AT&T

Nicolai Leymann
Deutsche Telekom

April 16, 2010

Using mLDP through a Backbone where there is no Route to the Root

[draft-wijnands-mpis-ml dp-recurs-fec-01.txt](#)

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/1id-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

Copyright and License Notice

Copyright (c) 2010 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.

Abstract

The control protocol used for constructing Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths ("MP LSPs") contains a field that identifies the address of a "root node". Intermediate nodes are expected to be able to look up that address in their routing tables. However, if the route to the root node is a BGP route, and the intermediate nodes are part of a BGP-free core, this is not possible. This document specifies procedures which enable a MP LSP to be constructed through a BGP-free core. In these procedures, the root node address is temporarily replaced by an address which is known to the intermediate nodes.

Table of Contents

<u>1</u>	Introduction	<u>3</u>
<u>2</u>	The Recursive Opaque Value Type	<u>5</u>
<u>2.1</u>	Encoding	<u>5</u>
<u>2.2</u>	Procedures	<u>5</u>
<u>3</u>	The VPN-Recursive MP FEC Element	<u>6</u>
<u>3.1</u>	Encoding	<u>6</u>
<u>3.2</u>	Procedures	<u>7</u>
<u>3.2.1</u>	Unsegmented Inter-AS P-tunnels	<u>7</u>
<u>3.2.2</u>	Limited Carrier's Carrier Function	<u>9</u>
<u>4</u>	IANA Considerations	<u>10</u>
<u>5</u>	Security Considerations	<u>11</u>
<u>6</u>	Acknowledgments	<u>11</u>
<u>7</u>	Authors' Addresses	<u>11</u>
<u>8</u>	Normative References	<u>12</u>

routes to R.

In such an environment, data packets from CE1 address to R would get encapsulated by PE1, tunneled to PE2, decapsulated by PE2, and forwarded to CE2.

Suppose now that CE1 is trying to set up a MP LSP whose root is R, and the intention is that the provider's network will participate in the construction of the LSP. Then the MLDP messages identifying the LSP must be passed from CE1 to PE1, from PE1 to P1, ..., from P2 to PE2, from PE2 to CE2, and from CE2 to R.

To begin the process, CE1 creates a MP FEC element with the address of R as the root node address, and passes that FEC element via MLDP to PE1. However, PE1 cannot use this same FEC element to identify the LSP in the LDP messages it sends to P1, because P1 does not have a route to R.

However, PE1 does know that PE2 is the "BGP next hop" on the path to R. What is needed is a method whereby:

- PE1 can tell P1 to set up an LSP as if the root node were PE2, and
- PE2 can determine that the LSP in question is really rooted at R, not at PE2 itself,
- PE2 can determine the original FEC element that CE1 passed to PE1, so that PE2 can pass it on to CE2.

This document defines the procedures that allow CE1 to create an LSP rooted at R. These procedures require PE1 to modify the MP FEC element before sending an MLDP message to P1. The modified FEC element has PE2 as the root, and the original FEC element as the opaque value. This requires a new type of opaque value. Since the opaque value contains a FEC element, we call this a "Recursive Opaque Value". When PE2 sends an mLDP message to CE2, it replaces the FEC element with the opaque value, thus undoing the recursion. Details are in [section 2](#).

[Section 3](#) defines a "VPN Recursive Opaque Value". Whereas the "Recursive Opaque Value" carries the original FEC, the "VPN Recursive Opaque Value" carries the original FEC plus a Route Distinguisher (RD). This has several possible uses in an L3VPN context. Details are in [section 3](#).

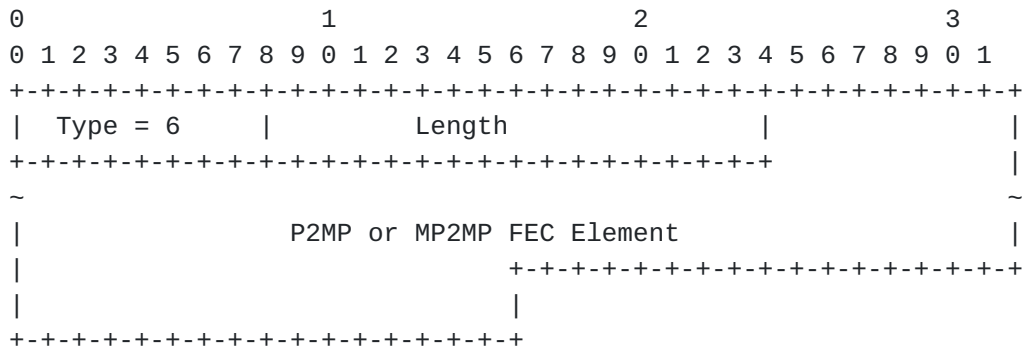
The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this

document are to be interpreted as described in [RFC2119].

2. The Recursive Opaque Value Type

2.1. Encoding

We define a new Opaque Value Type, the Recursive Opaque Value Type.



Recursive Opaque Value Type
Figure 3

The "opaque value" itself is a P2MP or MP2MP FEC element, encoded exactly as specified in [MLDP], with a type field, a length field, and value field of its own. The length field of the Recursive Opaque Value Type thus includes the type and length fields of the FEC element that is the value field.

2.2. Procedures

In the topology of Figure 2, let us suppose that CE1 sends PE1 an MP FEC element whose root node is R, and whose opaque value is Q. We will refer to this FEC element as "CE1-FEC". We may think of CE1-FEC as an ordered pair, as follows:

$$\text{CE1-FEC} = \langle \text{root}=R, \text{opaque_value}=Q \rangle.$$

PE1 determines that the root node R matches a BGP route, with a BGP next hop of PE2. PE1 also knows by its configuration that the interior routers on the path to PE2 are "BGP-free", and thus have no route to R.

PE1 therefore MUST create a new MP FEC element, whose root node address is the address of PE2, and whose opaque value is a Recursive

(type 6) Opaque Value whose value field contains CE1-FEC. We refer to this FEC element as PE2-FEC. PE1 then MUST send this FEC element to P1.

PE2-FEC = <root=PE2, opaque_value=CE1-FEC>, or

PE2-FEC = <root=PE2, opaque_value=<root=R,
opaque_value=Q>>

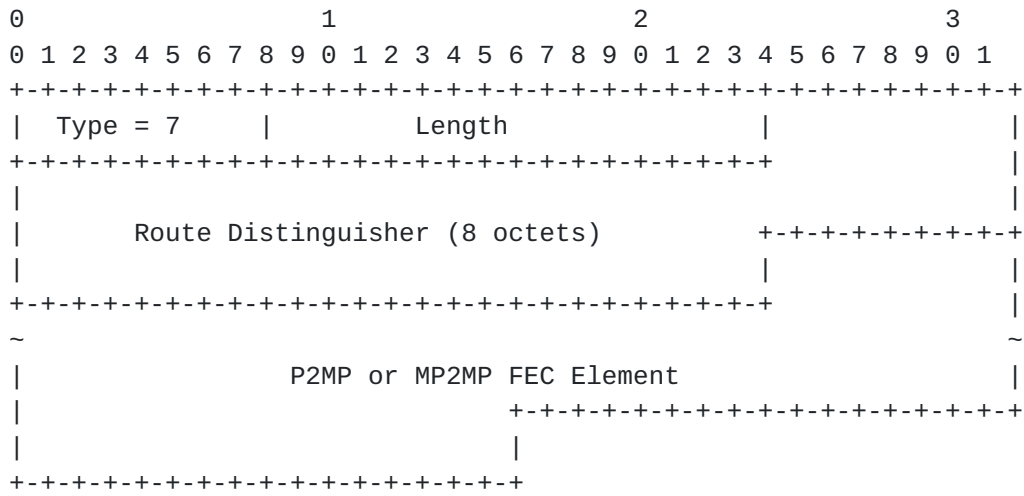
As far as the interior routers are concerned, they are being requested to build a MP LSP whose root node is PE2. They MUST NOT interpret the opaque value at all.

When PE2-FEC arrives at PE2, PE2 notes that it is the identified root node, and that the opaque value is a Recursive (type 6) opaque value. Therefore it MUST replace PE2-FEC with the contents of the type 6 opaque value (i.e., with CE1-FEC) before doing any further processing. This will result in CE1-FEC being sent on to CE2, and presumably further from CE2 to R. Note that CE1-FEC will contain the LSP root node specified by CE1; the presumption is that PE2 has a route to this root node.

3. The VPN-Recursive MP FEC Element

3.1. Encoding

We define a new Opaque Value Type, the VPN-Recursive Opaque Value Type.



VPN-Recursive Opaque Value Type
Figure 3

The "opaque value" consists of an eight-octet Route Distinguisher (RD), followed by a P2MP or MP2MP FEC element, encoded exactly as specified in [MLDP], with a type field, a length field, and value field of its own. The length field of the Recursive Opaque Value Type thus includes the 8 octets of RD plus the type and length fields of the FEC element that is the value field.

3.2. Procedures

3.2.1. Unsegmented Inter-AS P-tunnels

Consider the Inter-AS VPN scenario depicted in Figure 4.



Figure 4

Suppose this is an "option B" VPN interconnect ([VPN] [section 10](#)). This means that the Autonomous System Border Router (ASBR) in the first Autonomous System (i.e., ASBR1) does not have a route to PE routers in other ASes (such as PE2). Suppose also that the MVPN policy is to instantiate PMSIs [MVPN] using mLDP, and that "unsegmented inter-AS P-tunnels" [MVPN] are being used.

In this scenario, PE1 may need to join a P2MP or MP2MP LSP whose root is PE2. PE1 has no route to PE2, and all PE1 knows about the route to PE2 is that ASBR1 is the BGP next hop. Since PE1 has no route to PE2, PE1 needs to originate an mLDP message with a FEC element that identifies ASBR1 as the root. This FEC element must contain enough information to enable ASBR1 to find the next hop towards PE2 even though ASBR1 does not have a route to PE2.

Although ASBR1 does not have a route to PE2, it does have a BGP Intra-AS I-PMSI A-D route [[MVPN](#)] whose NLRI contains PE2's IP address together with a particular RD. PE1 also has this Inter-AS I-PMSI A-D route. The LSP needs to be set up along the path established by the Intra-AS I-PMSI A-D routes. Therefore one must use a Recursive FEC element that contains the RD as well as the address of PE2. The "VPN-Recursive FEC Element" defined herein is used for this purpose.

This enables us to provide the same functionality, for mLDP P-tunnels that is provided for PIM P-tunnels in section 8.1.3.2 of [[MVPN](#)] though the use of the MVPN Join Attribute.

At PE1 in Figure 4, the LSP to be created is associated with a particular VRF. PE1 looks up in that VRF the Intra-AS I-PMSI A-D route originated by PE2. It finds that the BGP next hop of that route is ASBR1. So it creates a P2MP or MP2MP FEC element whose root is ASBR1, and whose opaque value is a VPN-Recursive FEC element. The VPN-Recursive FEC element itself consists of a root, an RD, and an opaque value, set as follows:

- The root is PE2
- The RD is the RD from the NLRI of the Intra-AS A-D route originated by PE2.
- The opaque value is chosen (by some method outside the scope of this document) so as to be unique in the context of PE2. (E.g., it may have been specified in a PMSI tunnel attribute originated by PE2.) We will refer to this opaque value as "Q".

The resulting FEC element can be informally represented as

```
<root=ASBR1, opaque_value=<root=PE2, RD, opaque_value=Q>>.
```

PE1 can now begin setting up the LSP by using this FEC element in an LDP label mapping message sent towards ASBR1.

When ASBR1 receives, over a non-VRF interface, an mLDP label mapping message containing this FEC element, it sees that it is the root, and

that the opaque value is a VPN-Recursive (type 7) FEC element. It parses the VPN-Recursive FEC element and extracts the root value, PE2.

If ASBR1 has a route to PE2, it continues setting up the LSP by using the following FEC element:

```
<root=PE2, opaque_value=Q>
```

However, if ASBR1 does not have a route to PE2, it looks for an Intra-AS I-PMSI A-D route whose NLRI contains PE2's address along with the specified RD value. Say the BGP next hop of that route is ASBR2. Then ASBR1 continues setting up the LSP by using the following FEC element:

```
<root=ASBR2, opaque_value=<root=PE2, RD, opaque_value=Q>>.
```

Note that in this case, the root has changed from ASBR1 to ASBR2, but the opaque value is the unchanged VPN-Recursive FEC element.

3.2.2. Limited Carrier's Carrier Function

Another possible use of the VPN recursive FEC is to provide a limited version of "Carrier's Carrier Service". Referring again to the topology of Figure 2, suppose that PE1/PE2 are offering "Carrier's Carrier VPN Service" [VPN] to CE1/CE2. CE1 sends PE1 an MP FEC element whose root node is R, and whose opaque value is Q. We will refer to this FEC element as "CE1-FEC". However, PE1's route to R will be in a VRF ("Virtual Routing and Forwarding Table"). Therefore the FEC-element created by PE1 must contain some identifier that PE2 can use to find the proper VRF in which to look up the address of R.

When PE1 looks up the address of R in a VRF, it will find a route in the VPN-IP address family. The next hop will be PE2, but there will also be a Route Distinguisher (RD) as part of that NLRI of the matching route. In this case, the new FEC element created by PE1 MUST have the address of PE2 as the root node address, and MUST have a VPN-Recursive (type 7) opaque value. The value field of the type 7 opaque value MUST consist of the 8-octet RD followed by CE1-FEC.

As far as the interior routers are concerned, they are being requested to build a MP LSP whose root node is PE2. They MUST NOT interpret the opaque value at all.

When an mLDP label mapping message containing PE2-FEC arrives at PE2 over a VRF interface, PE2 notes that it is the identified root node,

and that the opaque value is a VPN-recursive (type 7) opaque value. Therefore it MUST replace PE2-FEC with the contents of the VPN-recursive opaque value (i.e., with CE1-FEC) before doing any further processing. It uses the VRF to lookup up the path to R. This will result in CE1-FEC being sent on to CE2, and presumably further from CE2 to R.

In this scenario, the RD in the VPN-Recursive Opaque Value also ensures uniqueness of the FEC Element within the inner carrier's network.

This way of providing Carrier's Carrier service has limited applicability, as it only works under the following conditions:

- Both the inner carrier and the outer carrier are using unsegmented mLDP P-tunnels
- The inner carrier is not aggregating the P-tunnels of the outer carrier, but is content to carry each such P-tunnel in a single P-tunnel of its own.

The carrier's carrier scenario can be distinguished from the inter-AS scenario by the fact that in the former, the mLDP messages are being exchanged on VRF interfaces.

4. IANA Considerations

[MLDP] defines a registry for "The LDP MP Opaque Value Element Type". This document requires the assignment of two new code points in this registry:

- Type 6.

An opaque value of this type is itself a TLV that encodes an mLDP FEC type, as defined in [MLDP].

- Type 7

An opaque value of this type consists of an eight-octet Route Distinguisher as defined in [VPN], followed by a TLV that encodes an mLDP FEC type, as defined in [MLDP].

5. Security Considerations

TBD

6. Acknowledgments

The authors wish to thank Toerless Eckert for his contribution to this work.

7. Authors' Addresses

IJsbrand Wijnands
Cisco Systems, Inc.
De kleetlaan 6a Diegem 1831
Belgium
E-mail: ice@cisco.com

Eric C. Rosen
Cisco Systems, Inc.
1414 Massachusetts Avenue
Boxborough, MA, 01719
E-mail: erosen@cisco.com

Maria Napierala
AT&T Labs
200 Laurel Avenue, Middletown, NJ 07748
E-mail: mnapierala@att.com

Nicolai Leymann
Deutsche Telekom
Winterfeldtstrasse 21
Berlin 10781
Germany
E-mail: n.leymann@telekom.de

8. Normative References

[MLDP] "Label Distribution Protocol Extensions for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths", Minei, Kompella, Wijnands, Thomas, [draft-ietf-mpls-ldp-p2mp-08.txt](#), October 2009

[MVPN] "Multicast in MPLS/BGP IP VPNs", Rosen, Aggarwal, et. al., [draft-ietf-l3vpn-2547bis-mcast-10.txt](#), January 2009

[RFC2119] "Key words for use in RFCs to Indicate Requirement Levels.", Bradner, March 1997

[VPN] "BGP/MPLS IP Virtual Private Networks (VPNs)", Rosen, Rekhter, [RFC 4364](#), February 2006

