Network Working Group
Internet Draft
Intended Status: Proposed Standard
Expires: April 9, 2010

IJsbrand Wijnands
Eric C. Rosen
Cisco Systems, Inc.

Maria Napierala
AT&T

October 9, 2009

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

draft-wijnands-mpls-mldp-csc-02.txt

Status of this Memo

Copyright and License Notice

Abstract

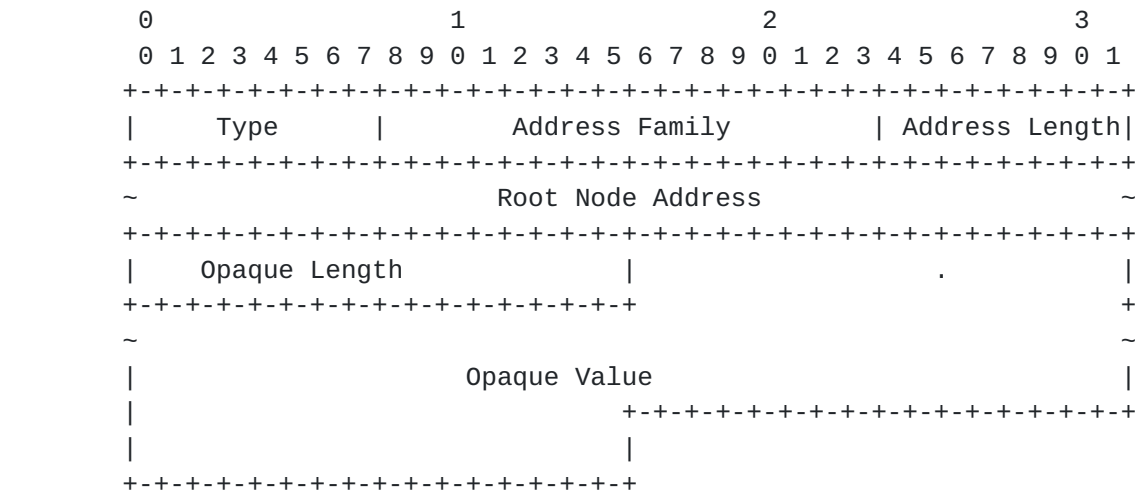
The control protocol used for constructing Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths ("MP LSPs") contains a field that identifies the address of a "root node". Intermediate nodes are expected to be able to look up that address in their routing tables. However, if the route 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

1. Introduction

[MLDP] defines several LDP FEC element encodings: P2MP, MP2MP Upstream, and MP2MP Downstream.

The encoding for these three FEC elements is shown in Figure 1.



MLDP FEC Element Encoding
Figure 1

Note that a P2MP or MP2MP label switched path ("MP LSP") is identified by the combination of a "root node" and a variable length "opaque value". The root node also plays a special role in the MLDP procedures - MLDP messages that are "about" a particular MP LSP are forwarded to the LDP adjacency that is the next hop on the route to the root node.

Sometimes it is desirable for a MP LSP to pass through a part of the network in which there is no route to the root node. For instance, consider the following topology:

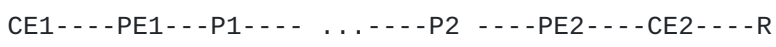


Figure 2

where CE1 and CE2 are "customer edge routers", PE1 and PE2 are "provider edge routers", but the provider's core is "BGP-free". That is, PE1 has a BGP-learned route for R, in which PE2 is the BGP next hop. However, the provider's interior routers (such as P1 and P2) do not have any BGP-learned routes, and in particular do not have any

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, and they also require PE2 to modify the MP FEC element before sending an MLDP message to CE2.

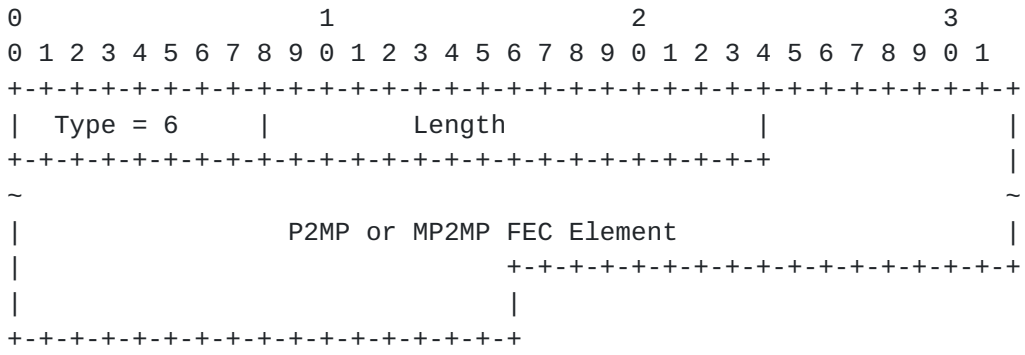
A slight variation on these procedures, also specified in this document, provides mLDP support for "Carrier's Carrier" MVPN service [VPN, MVPN].

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 is 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".

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

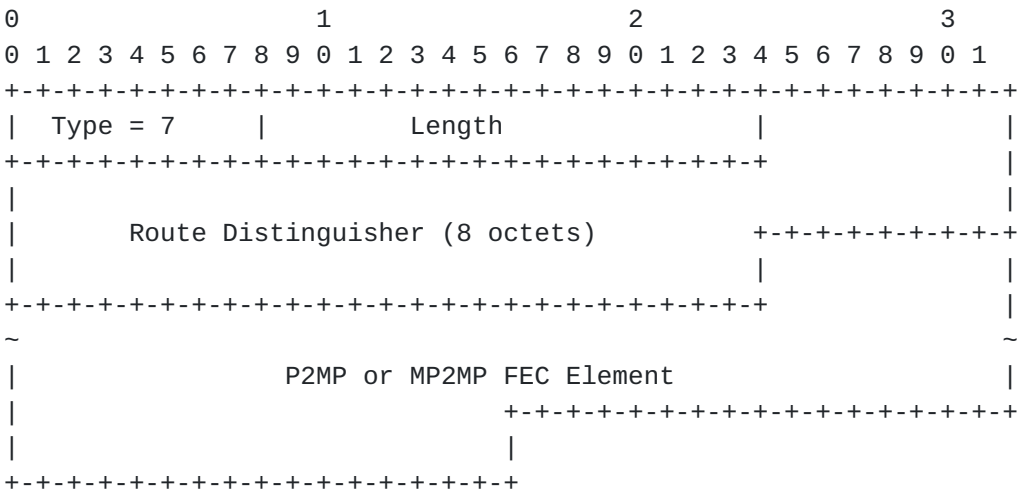
As far as the interior routers are concerned, they are being requested to build a MP LSP whose root node is PE2. They MUST NOT interpret the opaque value at all.

When PE2-FEC arrives at PE2, PE2 notes that it is the identified root node, and that the opaque value is a 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.

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 4

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 is 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

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 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 PE2-FEC arrives at PE2, PE2 notes that it is the identified root node, and that the opaque value is a type 7 opaque value. Therefore it MUST replace PE2-FEC with the contents of the type 7 opaque value (i.e., with CE1-FEC) before doing any further processing. It also finds the VRF associated with the identified RD, and MUST use that 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.

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

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-05.txt, May 2008

[MVPN] "Multicast in MPLS/BGP IP VPNs", Rosen, Aggarwal, et. al., draft-ietf-l3vpn-2547bis-mcast-08.txt, March 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