Using Multipoint LDP when the Backbone has no Route to the Root

draft-ietf-mpls-mldp-recurs-fec-04.txt

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 that is known to the intermediate nodes and is on the path to the true root node.

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."
1. Introduction

The document [mLDP] extends LDP ("Label Distribution Protocol", [LDP]) to support multipoint label-switched paths. These extensions are known as "Multipoint LDP", or more simply as "mLDP". [mLDP] defines several LDP "Forwarding Equivalence Class" (FEC) element encodings: "Point-to-Multipoint" (P2MP), "Multipoint-to-Multipoint (MP2MP) Upstream", and "MP2MP Downstream".

The encoding for these three FEC elements, as defined in [mLDP], is shown in Figure 1.

```
+-------+---------+----------------+---------+-------+-------+
| Type   | Address Family | Address Length |
|--------+----------------+---------+-------+-------+
|        | Root Node Address |       |       |       |
|--------+----------------+---------+-------+-------+
|        | Opaque Length    |         |       |       |
|--------+----------------+---------+-------+-------+
|        | Opaque Value     |         |       |       |
|--------+----------------+---------+-------+-------+
```

```
0                   1                   2                   3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      |        Address Family         | Address Length|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                       Root Node Address                       ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Opaque Length              |                   .           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                                                               ~
|.                     Opaque Value                              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                                                               ~

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 topology of Figure 2.

```
CE1----PE1---P1---- ...----P2 ----PE2----CE2----R

Figure 2
```
In Figure 2, CE1 and CE2 are "customer edge routers", R is a customer router at the same VPN site as CE2, PE1 and PE2 are "provider edge routers". Suppose that 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, unicast data packets from CE1 addressed 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 "Virtual Private Network (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 is applicable when MP LSPs are being used to carry the multicast traffic of a VPN [MVPN]. 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

2.1. Encoding

We define a new type of Opaque Value, the Recursive Opaque Value. This is a "basic type", identified by a one-octet type field.

```
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Type = TBD | Length |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                                                               ~
|                P2MP or MP2MP FEC Element                     |
|                                                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Recursive Opaque Value

Figure 3

The value field of the "Recursive Opaque Value" is 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 of the Recursive Opaque Value thus includes the lengths of the type, length, and value fields of the contained FEC element.

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:

```
CE1-FEC = <root=R, opaque_value=Q>
```
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 creates a new MP FEC element, whose root node address is the address of PE2, and whose opaque value is a Recursive Opaque Value whose value field contains CE1-FEC. We refer to this FEC element as PE2-FEC:

\[
PE2\text{-FEC} = <\text{root}=\text{PE2}, \text{opaque}\_\text{value}=\text{CE1}\text{-FEC}>, \text{ i.e.,}
\]

\[
\begin{align*}
PE2\text{-FEC} &= \langle \text{root}=\text{PE2}, \text{opaque}\_\text{value}=<\text{root}=\text{R}, \\
& \hspace{1cm} \text{opaque}\_\text{value}=\text{Q}>> \\
\end{align*}
\]

PE1 then sends 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 (PE2) is the identified root node, and that the opaque value is a Recursive Opaque Value. Therefore PE2 MUST replace PE2-FEC with the contents of the Recursive Opaque Value (i.e., with CE1-FEC) before doing any further processing. This will result in CE1-FEC being sent on to CE2, and 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 Opaque Value

3.1. Encoding

We define a new type of Opaque Value, the VPN-Recursive Opaque Value. This is a "basic type", identified by a one-octet type field.
The value field of the "VPN-Recursive 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 of the VPN-Recursive Opaque Value thus includes the 8 octets of RD plus the lengths of the type, length, and values fields of the contained FEC element.

3.2. Procedures

3.2.1. Non-segmented Inter-AS P-tunnels

Consider the Inter-AS VPN scenario depicted in Figure 5.

PE1 --- P1 ---- ASBR1 ... ASBR2 ---- P2 ---- PE2

Figure 5

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 "non-segmented 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. P1 has no route to PE2, and all PE1 knows about the route to PE2 is that ASBR1 is the BGP next hop. Since P1 has no root 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 as 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] through the use of the MVPN Join Attribute.

At PE1 in Figure 4, the LSP to be created is associated with a particular VPN Routing/Forwarding Table (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 Opaque Value. It parses the
VPN-Recursive Opaque value 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:

\[
\text{<root=PE2, opaque_value=Q>}
\]

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:

\[
\text{<root=ASBR2, opaque_value=<root=PE2, RD, opaque_value=Q>}}
\]

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 has
the address of PE2 as the root node address, and has a VPN-Recursive
Opaque Value. The value field of the VPN-Recursive Opaque Value
consists 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 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 non-segmented 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 Basic Type". This document requires the assignment of two new code points in this registry:

- Recursive Opaque Value: Type TBD (requested value: 7)

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

- VPN-Recursive Opaque Value: Type TBD (requested value: 8)

  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

The security considerations of [LDP] and [mLDP] apply.

Unauthorized modification of the FEC elements defined in this document can disrupt the creation of the multipoint LSPs, or can cause the multipoint LSPs to pass through parts of the network where they are not supposed to go. This could potentially be used as part of an attack to illegitimately insert or intercept multicast traffic. However, since the FEC elements defined in this document are not inherently more vulnerable to this form of attack than are the previously defined FEC elements, this document does not add new security vulnerabilities.

A description of general security issues for MPLS can be found in [RFC5920].

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


[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

9. Informative References