Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: February 8, 2015

P. Dutta
M. Bocci
Alcatel-Lucent
L. Martini
Cisco Systems

August 7, 2014

Explicit Path Routing for Dynamic Multi-Segment Pseudowires
draft-ietf-pwe3-mspw-er-05

Abstract

Dynamic Multi-Segment Pseudowire (MS-PW) setup through an explicit path may be required to provide a simple solution for 1:1 protection with diverse primary and backup MS-PWs for a service, or to enable controlled signaling (strict or loose) for special MS-PWs. This document specifies the extensions and procedures required to enable dynamic MS-PWs to be established along explicit paths.

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].

Table of Contents

1. Introduction

Procedures for dynamically establishing multi-segment pseudowires (MS-PWs), where their paths are automatically determined using a dynamic routing protocol, are defined in [RFC7267]. For 1:1 protection of MS-PWs with primary and backup paths, MS-PWs need to be established through a diverse set of S-PEs (Switching Provider-Edges) to avoid any single points of failure at the PW level. [RFC7267] allows this through BGP-based mechanisms. This document defines an additional mechanism that allows the ST-PE (Source Terminating PEs) to explicitly choose the path that a PW would take through the intervening S-PEs. Explicit path routing of dynamic MS-PWs may also be required for controlled set-up of dynamic MS-PWs and network resource management.

Note that in many deployments the ST-PE will not have a view of the topology of S-PEs and so the explicit route will need to be supplied

from a management application. How that management application determines the explicit route is outside the scope of this document.

2. Terminology

This document uses the terminology defined in [RFC7267], [RFC4447]and [RFC5036].

The following additional terminology is used:

Abstract Node: A group of nodes (S-PEs) representing an explicit hop along the path of an MS-PW. An abstract node is identified by an IPv4, IPv6 or S-PE address.

3. Explicit Path in MS-PW Signaling

This section describes the LDP (Label Distribution Protocol) extensions required for signaling explicit paths in dynamic MS-PW set-up messages. An explicitly routed MS-PW is set up using a Label Mapping message that carries an ordered list of the S-PEs which the MS-PW is expected to traverse. The ordered list is encoded as a series of Explicit Route (ER) Hop TLVs encoded in an ER-TLV that is carried in a Label Mapping message.

3.1. S-PE Addressing

An S-PE address is used to identify a given S-PE among the set of S-PEs belonging to the PSNs that may be used by an MS-PW. Each S-PE MUST be assigned an address as specified in [RFC7267] Section 3.2. An S-PE that is capable of dynamic MS-PW signaling, but has not been assigned an S-PE address, and that receives a Label Mapping message for a dynamic MS-PW MUST follow the procedures of [RFC7267] Section 3.2.

3.2. Explicit Route TLV (ER-TLV)

The ER-TLV specifies the path to be taken by the MS-PW being established. Each hop along the path is represented by an abstract node, which is a group of one or more S-PEs, identified by an IPv4, and IPv6 or an S-PE address. The ER-TLV format is as per Section 4.1 of [RFC3212].

The ER-TLV contains one or more Explicit Route Hop TLVs (ER-Hop TLVs) defined in Section 3.3.

3.3. Explicit Route Hop TLV (ER-Hop TLV)

The contents of an ER-TLV are a series of variable length ER-Hop TLVs. Each hop contains the identification of an "Abstract Node" that represents the hop to be traversed. The ER-Hop TLV format is as specified in Section 4.2 of [RFC3212].

[RFC3212] defines three ER-Hop TLV Types: IPv4 Prefix, IPv6 Prefix, and Autonomous System. This document specifies the following new ER-Hop TLV Type:

```
Value  Type
------ ------------------------
0x0805 L2 PW address of PW Switching Point
```

ER-Hop TLV

Details of ER Hop semantics are defined in Section 3.4.

3.4. ER-Hop Semantics

This section describes the various semantics associated with ER-Hop TLV.

3.4.1. ER-Hop Type: IPv4 Prefix

The semantics of the IPv4 ER-Hop TLV Type are specified in [RFC3212] Section 4.7.1.

3.4.2. ER-Hop Type: IPv6 Prefix

The semantics of the IPv6 ER-Hop TLV Type are specified in [RFC3212] Section 4.7.2.

3.4.3. ER-Hop Type: L2 PW Address

The semantics of the L2 PW Address ER-Hop TLV Type, which contains the L2 PW Address derived from the Generalized PWid FEC AII type 2 defined in [RFC5003], are as follows.

```
  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 2 3 4 5 6 7 8 9 0 1
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |U|F|      ER Hop Type          |      Length = 18              |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |L|             Reserved                        |    PreLen     |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |  AII Type=02  |    Length     |        Global ID              |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |       Global ID (contd.)      |        Prefix                 |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |       Prefix (contd.)         |        AC ID                  |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |      AC ID                    |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

U/F

These bits MUST be set to zero and the procedures of [RFC5036] followed when the TLV is not known to the receiving node.

Type

A fourteen-bit field carrying the value of the ER-Hop 3, L2 PW Address, Value = TBD

Length

Specifies the length of the value field in bytes = 18.

L Bit

Set to indicate Loose hop.
Cleared to indicate a strict hop.

Reserved

Zero on transmission. Ignored on receipt.

PreLen

Prefix Length 1-96 (including the length of the Global ID, Prefix and AC ID fields).

All other fields (AII Type, Length, Global ID, Prefix, and AC ID) define the L2 PW Address and are to be set and interpreted as defined in Section 3.2 of [RFC5003].

4. Explicit Route TLV Processing

4.1. Next-Hop Selection

A PW Label Mapping Message containing an explicit route TLV specifies the next hop for a given MS-PW path. Selection of this next hop may involve a selection from a set of possible alternatives. The mechanism for making a selection from this set is implementation specific and is outside of the scope of this document. The mechanism used to select a particular path is also outside of the scope of this document, but each node MUST attempt to determine a loop-free path. A noted in Section 1, in many deployments the ST-PE will not have a view of the topology of S-PEs and so the path will need to be supplied from a management application.

If a loop free path cannot be found, then a node MUST NOT attempt to signal the MS-PW. For an S-PE, if it cannot determine a loop free path, then the received Label Mapping MUST be released with a status code of "PW Loop Detected" as per Section 4.2.3 of [RFC7267].

To determine the next hop for the MS-PW path, a node performs the following steps. Note that these procedures assume that a valid S-PE address has been assigned to the node, as per Section 3.1, above.

1. The node receiving the Label Mapping Message that contains an ER-TLV MUST evaluate the first ER Hop. If the L bit is not set in the first ER Hop and if the node is not part of the abstract node described by the first ER Hop (i.e it does not lie within the prefix as determined by the prefix length specified in the ER-Hop TLV), it has received the message in error, and MUST reply with a Label Release Message with a "Bad Initial ER Hop Error" (0x04000004) status code. If the L bit is set and the local node is not part of the abstract node described by the first ER Hop, the node selects a next hop that is along the path to the abstract node described by the first ER Hop. If there is no ER-Hop TLV contained in the ER-TLV, the message is also in error and the node should return a "Bad Explicit Routing TLV Error" (0x04000001) status code in a Label Release Message sent to upstream node. Note that this statement does not preclude a Label mapping message with no ER-TLV. If a Label Mapping message with no ER-TLV is received, then it MUST be processed as per [RFC7267].

2. If there are no further ER-Hop TLVs following the first ER-Hop TLV, this indicates the end of the explicit route. The Explicit Route TLV MUST be removed from the Label Mapping message. This node may or may not be the end of the PW. Processing continues

as per Section 4.2, where a new explicit route TLV MAY be added to the Label Mapping Message.

3. If a second ER Hop TV does exist, and the node is also a part of the abstract node described by the second ER-Hop, then the node deletes the first ER-Hop and continues processing with step 2, above. Note that this makes the second ER Hop into the first ER Hop for the iteration for the next PW segment.

4. The node determines if it is topologically adjacent to the abstract node described by the second ER Hop. That is, it is directly connected to the next node by a PW control plane adjacency. If so, the node selects a particular next hop which is a member of the abstract node. The node then deletes the first ER-Hop and continues processing as per Section 4.2, below.

5. Next, the node selects a next hop within the abstract node of the first ER Hop that is along the path to the abstract node of the second ER Hop. If no such path exists then there are two cases:

   A. If the second ER Hop is a strict ER Hop, then there is an error and the node MUST return a Label Release Message to upstream node with "Bad Strict Node Error" (0x04000002) status code.

   B. Otherwise, if the second ER Hop is a loose ER Hop, then the node selects any next hop that is along the path to the next abstract node. If no path exists within the MPLS domain, then there is an error, and the node MUST return a Label Release Message to upstream node with "Bad Loose Node Error" (0x04000003) status code.

6. Finally, the node replaces the first ER Hop with any ER Hop that denotes an abstract node containing the next hop. This is necessary so that when the explicit route is received by the next hop, it will be accepted.

7. Progress the Label Mapping Message to the next hop.

4.2. Adding ER Hops to the Explicit Route TLV

After selecting a next hop, the node may alter the explicit route in the following ways.

If, as part of executing the algorithm in Section 4.1, the explicit route TLV is removed, the node may add a new explicit route TLV.

Otherwise, if the node is a member of the abstract node for the first ER-Hop, then a series of ER Hops may be inserted before the First ER Hop or may replace the first ER Hop. Each ER Hop in this series must denote an abstract node that is a subset of the current abstract node.

Alternately, if the first ER-Hop is a loose ER Hop, an arbitrary series of ER Hops may be inserted prior to the first ER-Hop.

5. IANA Considerations

RFC5036 [RFC5036] defines the LDP TLV name space which is maintained by IANA as "LDP TLV Registry". TLV types for the Explicit Route TLV, IPv4 Prefix ER-Hop TLV, and the IPv6 Prefix ER-Hop TLV are already defined in the LDP TLV Registry.

IANA is requested to assign a further code point from the IETF consesus portion of this registry as follows:

TLV Type	Value	Reference
L2 PW Address of Switching Point	TBD	This Document

A value of 0x0805 is requested.

6. Security Considerations

This document introduces no new security considerations over [RFC5036], [RFC4447] and [RFC7267]. The security considerations detailed in those documents apply to the protocol extensions described in this RFC.

7. Acknowledgements

The authors gratefully acknowledge the contribution of the [RFC3212] RFC3212 authors through the specification of TLVs, which are reused by this document. The authors also gratefully acknowledge the input of Lizhong Jin.

8. Normative References

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

[RFC3212] Jamoussi, B., Andersson, L., Callon, R., Dantu, R., Wu, L., Doolan, P., Worster, T., Feldman, N., Fredette, A., Girish, M., Gray, E., Heinanen, J., Kilty, T., and A. Malis, "Constraint-Based LSP Setup using LDP", RFC 3212, January 2002.

[RFC4447] Martini, L., Rosen, E., El-Aawar, N., Smith, T., and G. Heron, "Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP)", RFC 4447, April 2006.

[RFC5003] Metz, C., Martini, L., Balus, F., and J. Sugimoto, "Attachment Individual Identifier (AII) Types for Aggregation", RFC 5003, September 2007.

[RFC5036] Andersson, L., Minei, I., and B. Thomas, "LDP Specification", RFC 5036, October 2007.

[RFC7267] Martini, L., Bocci, M., and F. Balus, "Dynamic Placement of Multi-Segment Pseudowires", RFC 7267, June 2014.

Authors' Addresses

Pranjal Kumar Dutta
Alcatel-Lucent
701 E Middlefield Road
Mountain View, California 94043
USA

Email: pranjal.dutta@alcatel-lucent.com

Matthew Bocci
Alcatel-Lucent
Voyager Place, Shoppenhangers Road
Maidenhead, Berks SL6 2PJ
UK

Email: matthew.bocci@alcatel-lucent.com

Luca Martini
Cisco Systems
9155 East Nichols Avenue, Suite 400
Englewood, Colorado 80112
USA

Email: lmartini@cisco.com