Network Working Group
Internet Draft
Expires: June 2014
Intended status: Standards Track
Updates: 6073

Luca Martini (Ed.)
Cisco Systems Inc.

Matthew Bocci (Ed.)
Florin Balus (Ed.)
Alcatel-Lucent

December 2, 2013

Dynamic Placement of Multi-Segment Pseudowires

draft-ietf-pwe3-dynamic-ms-pw-20.txt

Status of this Memo

Abstract

RFC5254 describes the service provider requirements for extending the reach of pseudowires (PW) across multiple Packet Switched Network domains. A Multi-Segment PW is defined as a set of two or more contiguous PW segments that behave and function as a single point-to-point PW. This document describes extensions to the PW control protocol to dynamically place the segments of the multi-segment pseudowire among a set of Provider Edge (PE) routers. This document also updates RFC6073 as follows: it updates the

value of the length field of the PW Switching Point PE Sub-TLV Type 0x06 to 14.

Table of Contents

1. Introduction

1.1. Scope

[RFC5254] describes the service provider requirements for extending the reach of pseudowires across multiple Packet Switched Network (PSN) domains. This is achieved using a Multi-Segment Pseudowire (MS-PW). An MS-PW is defined as a set of two or more contiguous PW segments that behave and function as a single point-to-point PW. This architecture is described in [RFC5659].

The procedures for establishing PWs that extend across a single PSN domain are described in [RFC4447], while procedures for setting up PWs across multiple PSN domains, or control plane domains are described in [RFC6073].

The purpose of this document is to specify extensions to the pseudowire control protocol [RFC4447], and [RFC6073] procedures, to enable multi-segment PWs to be dynamically placed. The procedures follow the guidelines defined in [RFC5036] and enable the reuse of existing TLVs, and procedures defined for SS-PWs in [RFC4447]. Dynamic placement of point-to-multipoint (P2MP) PWs is for further study and outside the scope of this document.

1.2. Specification of Requirements

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 RFC 2119.

1.3. Terminology

[RFC5659] provides terminology for multi-segment pseudowires.

This document defines the following additional terms:

- Source Terminating Provider Edge (ST-PE). A Terminating Provider Edge (T-PE), which assumes the active signaling role and initiates the signaling for multi-segment PW.
- Target Terminating Provider Edge (TT-PE). A Terminating Provider Edge (T-PE) that assumes the passive signaling role. It waits and responds to the multi-segment PW signaling message in the reverse direction.

- Forward Direction: ST-PE to TT-PE.
- Reverse Direction: TT-PE to ST-PE.
- Pseudowire Routing (PW routing): The dynamic placement of the segments that compose an MS-PW, as well as the automatic selection of S-PEs.

1.4. Architecture Overview

The following figure shows the reference model, derived from [RFC5659], to support PW emulated services using multi-segment PWs.

```
       Native   |<-------------Pseudowire------------>|  Native
       Service  |                                     |  Service
        (AC)    |     |<-PSN1-->|     |<-PSN2-->|     |   (AC)
          |     V     V         V     V         V     V     |
          |     +-----+         +-----+         +-----+
   +----+ |     |T-PE1|=========|S-PE1|=========|T-PE2|     |    +----+
   |    |-------|.....PW.Seg't1........PW.Seg't3......|----------|    |
   | CE1| |     |     |         |     |         |     |     |    |CE2 |
   |    |-------|.....PW.Seg't2.......|PW.Seg't4......|----------|    |
   +----+ |     |     |=========|     |=========|     |     |    +----+
        ^       +-----+         +-----+         +-----+          ^
        |   Provider Edge 1        ^        Provider Edge 2      |
        |                          |                             |
        |                          |                             |
        |                  PW switching point                    |
        |                                                        |
        |<---------------- Emulated Service -------------------->|
```

Figure 1: MS-PW Reference Model

T-PE1 and T-PE2 provide an emulated service to Customer Edge (CE) CE1 and CE2. These Provider Edge (PE) nodes reside in different PSNs. A PSN tunnel extends from T-PE1 to S-PE1 across PSN1, and a second PSN tunnel extends from S-PE1 to T-PE2 across PSN2. PWs are used to connect the attachment circuits (ACs) attached to T-PE1 to the corresponding AC attached to T-PE2. A PW segment on the tunnel across PSN1 is connected to a PW segment in the tunnel across PSN2 at S-PE1 to complete the multi-segment PW (MS-PW) between T-PE1 and T-PE2. S-PE1 is therefore the PW switching point and is referred to as the switching provider edge (S-PE). PW Segment 1 and PW Segment 3 are segments of the same MS-PW while PW Segment 2 and PW Segment 4 are

segments of another MS-PW. PW segments of the same MS-PW (e.g., PW segment 1 and PW segment 3) MUST be of the same PW type, and PSN tunnels (e.g., PSN1 and PSN2) can be of the same or a different technology. An S-PE switches an MS-PW from one segment to another based on the PW identifiers (PWid, or Attachment Individual Identifier (AII)). How the PW protocol data units (PDUs) are switched at the S-PE depends on the PSN tunnel technology: in case of a multiprotocol label switching (MPLS) PSN to another MPLS PSN, PW switching involves a standard MPLS label swap operation.

Note that although Figure 1 only shows a single S-PE, a PW may transit more than one S-PE along its path. For instance, in the multi-provider case, there can be an S-PE at the border of one provider domain and another S-PE at the border of the other provider domain.

2. Applicability

This document describes the case where the PSNs carrying the MS-PW are only MPLS PSNs using the Generalized PWID FEC element (also known as FEC129). Interactions with an IP PSN using L2TPv3 as described in [RFC6073] section 7.4 are for further study.

2.1. Changes to Existing PW Signaling

The procedures described in this document make use of existing LDP TLVs and related PW signaling procedures described in [RFC4447] and [RFC6073]. The following optional TLV is also defined:

- A Bandwidth TLV to address QoS Signaling requirements (see Section 6.2.1).

This document also updates the value of the length field of the PW Switching Point PE Sub-TLV Type 0x06 to 14.

3. PW Layer 2 Addressing

Single segment pseudowires on an MPLS PSN can use attachment circuit identifiers for a PW using FEC 129. In the case of a dynamically placed MS-PW, there is a requirement for the attachment circuit identifiers to be globally unique, for the purposes of reachability and manageability of the PW. Referencing figure 1 above, individual globally unique addresses MUST be allocated to all the ACs and S-PEs of an MS-PW.

3.1. Attachment Circuit Addressing

The attachment circuit addressing is derived from [RFC5003] AII type 2, shown here:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  AII Type=02  |    Length     |        Global ID              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       Global ID (contd.)      |        Prefix                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       Prefix (contd.)         |        AC ID                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      AC ID                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: AII Type 2 TLV Structure

The fields are defined in [RFC5003], Section 3.2.

AII type 2 based addressing schemes permit varying levels of AII summarization, thus reducing the scaling burden on PW routing. AII Type 2 based PW addressing is suitable for point-to-point provisioning models where auto-discovery of the address at the Target T-PE is not required. That is, it is known a-priori by provisioning.

Implementations of the following procedure MUST interpret the AII type to determine the meaning of the address format of the AII, irrespective of the number of segments in the MS-PW. All segments of the PW MUST be signaled with same AII Type.

A unique combination of Global ID, Prefix, and AC ID parts of the AII type 2 are assigned to each AC. In general, the same global ID and prefix are be assigned for all ACs belonging to the same T-PE. This is not a strict requirement, however. A particular T-PE might have more than one prefix assigned to it, and likewise a fully qualified AII with the same Global ID/Prefix but different AC IDs might belong to different T-PEs.

For the purpose of MS-PWs, the AII MUST be globally unique across all PSNs that are spanned by the MS-PW.

The AII for a local attachement circuit of a given T-PE of an MS-PW and the AII of the corresponding attachment circuit on a far-end T-PE (with respect to the LDP signaling) are known as the Source Attachment Individual Identifier (SAII) and Target Attachment

Individual Identifier (TAII) as per [RFC6074].

3.2. S-PE Addressing

Each S-PE MUST be assigned an address which uniquely identifies it from a pseudowire perspective, in order to populate the Switching Point PE (SP-PE) TLV specified in [RFC6073]. For this purpose, at least one Attachment Identifier (AI) address of the format similar to AII type 2 [RFC5003] composed of the Global ID, and Prefix part, only, MUST be assigned to each S-PE.

If an S-PE is capable of Dynamic MS-PW signaling, but is not assigned with an S-PE address, then on receiving a Dynamic MS-PW label mapping message the S-PE MUST return a Label Release with the "LDP_RESOURCES_UNAVAILABLE" (0x38)" status code.

4. Dynamic Placement of MS-PWs

[RFC6073] describes a procedure for concatenating multiple pseudowires together. This procedure requires each S-PE to be manually configured with the information required for each segment of the MS-PW. The procedures in the following sections describe a method to extend [RFC6073] by allowing the automatic selection of pre-defined S-PEs, and dynamically establishing a MS-PW between two T-PEs.

4.1. Pseudowire Routing Procedures

The AII type 2 described above contains a Global ID, Prefix, and AC ID. The Target Attachment Individual Identifier (TAII) is used by S-PEs to determine the next SS-PW destination for LDP signaling.

Once an S-PE receives a MS-PW label mapping message containing a TAII with an AII that is not locally present, the S-PE performs a lookup in a PW AII routing table. If this lookup results in an IP address for the next-hop PE with reachability information for the AII in question, then the S-PE will initiate the necessary LDP messaging procedure to set-up the next PW segment. If the PW AII routing table lookup does not result in a IP address for a next-hop PE, the destination AII has become unreachable, and the PW setup MUST fail. In this case the next PW segment is considered un-provisioned, and a label release MUST be returned to the T-PE with a status message of "AII Unreachable".

If the TAI of a MS-PW label mapping message received by a PE contains the prefix matching a locally-provisioned prefix on that PE, but an AC ID that is not provisioned, then the LDP liberal label retention procedures apply, and the label mapping message is retained.

To allow for dynamic end-to-end signaling of MS-PWs, information must be present in S-PEs to support the determination of the next PW signaling hop. Such information can be provisioned (equivalent to a static route) on each S-PE, or disseminated via regular routing protocols (e.g. BGP).

4.1.1. AII PW Routing Table Lookup Aggregation Rules

All PEs capable of dynamic MS-PW path selection MUST build a PW AII routing table to be used for PW next-hop selection.

The PW addressing scheme (AII type 2 in [RFC5003]) consists of a Global ID, a 32 bit prefix and a 32 bit Attachment Circuit ID.

An aggregation scheme similar to that used for classless IPv4 addresses can be employed. An (8 bits) length mask is specified as a number ranging from 0 to 96 that indicates which Most Significant Bits (MSB) are relevant in the address field when performing the PW address matching algorithm.

```
 0        31 32    63 64    95 (bits)
+-----------+--------+--------+
| Global ID | Prefix | AC ID  |
+-----------+--------+--------+
```

Figure 3: PW Addressing Scheme

During the signaling phase, the content of the (fully qualified) TAII type 2 field from the FEC129 TLV is compared against routes from the PW Routing table. Similar with the IPv4 case, the route with the longest match is selected, determining the next signaling hop and implicitly the next PW Segment to be signaled.

4.1.2. PW Static Route

For the purpose of determining the next signaling hop for a segment of the pseudowire, the PEs MAY be provisioned with fixed route entries in the PW next hop routing table. The static PW entries will follow all the addressing rules and aggregation rules described in the previous sections. The most common use of PW static provisioned

routes is this example of the "default" route entry as follows:

Global ID = 0 Prefix = 0 AC ID = 0 , Prefix Length = 0 Next Signaling Hop = {IP Address of next hop S-PE or T-PE}

4.1.3. Dynamic Advertisement with BGP

Any suitable routing protocol capable of carrying external routing information MAY be used to propagate MS-PW path information among S-PEs and T-PEs. However, T-PEs and S-PEs MAY choose to use Border Gateway Protocol (BGP) [RFC4271] with the Multiprotocol Extensions as defined in [RFC4760] to propagate PW address information throughout the PSN.

Contrary to layer 2 VPN signaling methods that use BGP [RFC6074] for auto discovery, in the case of the dynamically placed MS-PW, the source T-PE knows a-priori (by provisioning) the AC ID on the terminating T-PE that signaling should use. Hence there is no need to advertise a "fully qualified" 96 bit address on a per PW Attachment Circuit basis. Only the T-PE Global ID, Prefix, and prefix length needs to be advertised as part of well known BGP procedures - see [RFC4760].

Since PW Endpoints are provisioned in the T-PEs, the ST-PE will use this information to obtain the first S-PE hop (i.e., first BGP next hop) to where the first PW segment will be established. Any subsequent S-PEs will use the same information (i.e. the next BGP next-hop(s)) to obtain the next-signaling-hop(s) on the path to the TT-PE.

The PW dynamic path Network Layer Reachability Information (NLRI) is advertised in BGP UPDATE messages using the MP_REACH_NLRI and MP_UNREACH_NLRI attributes [RFC4760]. The {AFI, SAFI} value pair used to identify this NLRI is (AFI=25, SAFI=6 (pending IANA allocation)). A route target MAY also be advertised along with the NLRI.

The Next Hop field of the MP_REACH_NLRI attribute SHALL be interpreted as an IPv4 address, whenever the length of the NextHop address is 4 octets, and as a IPv6 address, whenever the length of the NextHop address is 16 octets.

The NLRI field in the MP_REACH_NLRI and MP_UNREACH_NLRI is a prefix comprising an 8-octet Route Distinguisher, the Global ID, the Prefix, and the AC-ID, and encoded as defined in section 4 of [RFC4760].

This NLRI is structured as follows:

```
Bit
0     7 8             71 72      103 104  135 136   167
+------+----------------+-----------+--------+--------+
|Length|  Route Dist    | Global ID | Prefix | AC ID  |
+------+----------------+-----------+--------+--------+
```

Figure 4: NLRI Field Structure

The Length field is the prefix length of the Route Distinguisher + Global ID + Prefix + AC-ID in bits.

Except for the default PW route, which is encoded as a 0 length prefix, the minimum value of the length field is 96 bits. Lengths of 128 bits to 159 bits are invalid as the AC ID field cannot be aggregated. The maximum value of the Length field is 160 bits. BGP advertisements received with invalid prefix lengths MUST be rejected as having a bad packet format.

4.2. LDP Signaling

The LDP signaling procedures are described in [RFC4447] and expanded in [RFC6073]. No new LDP signaling components are required for setting up a dynamically placed MS-PW. However, some optional signaling extensions are described below.

One of the requirements that MUST be met in order to enable the QoS objectives for a PW to be achieved on a segment is that a PSN tunnel MUST be selected that can support at least the required class of service and that has sufficient bandwidth available.

Such PSN tunnel selection can be achieved where the next hop for a PW segment is explicitly configured at each PE, whether the PE is a T-PE or an S-PE in the case of a segmented PW, without dynamic path selection (as per RFC6073). In these cases, it is possible to explicitly configure the bandwidth required for a PW so that the T-PE or S-PE can reserve that bandwidth on the PSN tunnel.

Where dynamic path selection is used and therefore the next-hop is not explicitly configured by the operator at the S-PE, a mechanism is required to signal the bandwidth for the PW from the T-PE to the S-PEs. This is accomplished by including an optional PW Bandwidth TLV. The PW Bandwidth TLV is specified 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|1|0|     PW BW TLV  (0x096E)   |          TLV  Length          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Forward SENDER_TSPEC                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Reverse SENDER_TSPEC                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 5: PW Bandwidth TLV Structure

The PW Bandwidth TLV fields are as follows:

- TLV Length: The length of the value fields in octets. Value = 64
- Forward SENDER_TSPEC = The SENDER_TSPEC for the forward direction of the PW, as defined in [RFC2210] section 3.1.
- Reverse SENDER_TSPEC = The SENDER_TSPEC for the reverse direction of the PW, as defined in [RFC2210] section 3.1.

The complete definitions of the content of the SENDER_TSPEC objects are found in [RFC2210] section 3.1. The forward SENDER_TSPEC refers to the data path in the direction of ST-PE to TT-PE. The reverse SENDER_TSPEC refers to the data path in the direction TT-PE to ST-PE.

In the forward direction, after a next hop selection is determined, a T/S-PE SHOULD reference the forward SENDER_TSPEC object to determine an appropriate PSN tunnel towards the next signaling hop. If such a tunnel exists, the MS-PW signaling procedures are invoked with the inclusion of the PW Bandwidth TLV. When the PE searches for a PSN tunnel, any tunnel which points to a next hop equivalent to the next hop selected will be included in the search (the LDP address TLV is used to determine the next hop equivalence)

When an S/T-PE receives a PW Bandwidth TLV, once the PW next hop is selected, the S/T-PE MUST request the appropriate resources from the PSN. The resources described in the reverse SENDER_TSPEC are allocated from the PSN toward the originator of the message or previous hop. When resources are allocated from the PSN for a specific PW, the SHOULD account for the usage of the resources by the PW.

In the case where PSN resources towards the previous hop are not

available, the following procedure MUST be followed:

- -i. The PSN MAY allocate more QoS resources, e.g. Bandwidth, to the PSN tunnel.
- -ii. The S-PE MAY attempt to setup another PSN tunnel to accommodate the new PW QoS requirements.
- -iii. If the S-PE cannot get enough resources to setup the segment in the MS-PW a label release MUST be returned to the previous hop with a status message of "Bandwidth resources unavailable"

In the latter case, the T-PE receiving the status message MUST also withdraw the corresponding PW label mapping for the opposite direction if it has already been successfully setup.

If an ST-PE receives a label mapping message the following procedure MUST be followed:

If the ST-PE has already sent a label mapping message for this PW then the ST-PE MUST check that this label mapping message originated from the same LDP peer to which the corresponding label mapping message for this particular PW was sent. If it is the same peer, the PW is established. If it is a different peer, then the ST-PE MUST send a label release message, with a status code of "Duplicate AII" to the PE that originate the LDP label mapping message.

If the PE has not yet sent a label mapping message for this particular PW , then it MUST send the label mapping message to this LDP peer, regardless of what the PW TAII routing lookup result is.

4.2.1. Equal Cost Multi Path (ECMP) in PW Routing

A next hop selection for a specific PW may find a match with a PW route that has multiple next hops associated with it. Multiple next hops may be either configured explicitly as static routes or may be learned through BGP routing procedures. Implementations at an S-PE or T-PE MAY use selection algorithms, such as CRC32 on the FEC TLV, or flow-aware transport PW [RFC6391], for load balancing of PWs across multiple next-hops. The details of such selection algorithms are outside the scope of this document.

4.2.2. Active/Passive T-PE Election Procedure

When a MS-PW is signaled, each T-PE might independently initiate signaling the MS-PW. This could result in a different path being used be each direction of the PW. To avoid this situation one T-PE MUST initiate PW signaling (i.e. take an active role), while the other T-

PE waits to receive the LDP label mapping message before sending the LDP label mapping message for the reverse direction of the PW (i.e. take a passive role). The Active T-PE (the ST-PE) and the Passive T-PE (the TT-PE) MUST be identified before signaling begins for a given MS-PW.

A T-PE SHOULD determine whether it assumes the active role or the passive role using procedures similar to those of [RFC5036] Section 2.5.2, Bullet 2. The T-PE compares the Source Attachment Individual Identifier (SAII) [RFC6074] with the Target Attachment Individual Identifier (TAII) [RFC6074] as unsigned integers, and if the SAII > TAII, the T-PE assumes the active role. Otherwise it assumes the passive role.

The following procedure for comparing the SAII and TAII as unsigned integers SHOULD be used:

- If the SAII Global ID > TAII Global ID, then the T-PE is active
  - else if the SAII Prefix > TAII Prefix, then the T-PE is active
    - else if the SAII AC-ID > TAII AC-ID, then the T-PE is active
      - else the T-PE is passive.

4.2.3. Detailed Signaling Procedures

On receiving a label mapping message, the S-PE MUST inspect the FEC TLV. If the receiving node has no local AII matching the TAII for that label mapping then the label mapping message should be forwarded on to another S-PE or T-PE. The S-PE will check if the FEC is already installed for the forward direction:

- If it is already installed, and the received mapping was received from the same LDP peer to which the forward LDP label mapping was sent, then this label mapping represents signaling in the reverse direction for this MS-PW segment.
- If it is already installed, and the received mapping was received from a different LDP peer to which the forward LDP label mapping was sent, then the received label mapping MUST be released with the status code of "PW_LOOP_DETECTED".
- If the FEC is not already installed, then this represents signaling in the forward direction.

For the forward direction:

-i. Determine the next hop S-PE or T-PE according to the procedures above. If next-hop reachability is not found in the PW AII routing table in the S-PE then label release MUST be sent with status code "AII_UNREACHABLE". If the next-hop S-PE or T-PE is found and is the same LDP Peer that has sent

the label mapping message then a label Release MUST be returned with the status code "PW_LOOP_DETECTED". If the SAII in the received label mapping is local to the S-PE then a label released MUST be returned with status code "PW_LOOP_DETECTED".

-ii. Check that a PSN tunnel exists to the next hop S-PE or T-PE. If no tunnel exists to the next hop S-PE or T-PE the S-PE MAY attempt to setup a PSN tunnel.

-iii. Check that a PSN tunnel exists to the previous hop. If no tunnel exists to the previous hop S-PE or T-PE the S-PE MAY attempt to setup a PSN tunnel.

-iv. If the S-PE cannot get enough PSN resources to setup the segment to the next or previous S-PE or T-PE, a label release MUST be returned to the T-PE with a status message of "Resources Unavailable".

-v. If the label mapping message contains a Bandwidth TLV, allocate the required resources on the PSN tunnels in the forward and reverse directions according to the procedures above.

-vi. Allocate a new PW label for the forward direction.

-vii. Install the FEC for the forward direction.

-viii. Send the label mapping message with the new forward label and the FEC to the next hop S-PE/T-PE.

For the reverse direction:

-i. Install the received FEC for the reverse direction.

-ii. Determine the next signaling hop by referencing the LDP sessions used to setup the PW in the Forward direction.

-iii. Allocate a new PW label for the reverse direction.

-iv. Install the FEC for the reverse direction.

-v. Send the label mapping message with a new label and the FEC to the next hop S-PE/ST-PE.

5. Failure Handling Procedures

5.1. PSN Failures

Failures of the PSN tunnel MUST be handled by PSN mechanisms. An example of such a PSN mechanism is MPLS fast reroute [RFC4090]. If the PSN is unable to re-establish the PSN tunnel, then the S-PE SHOULD follow the procedures defined in Section 8 of [RFC6073].

5.2. S-PE Specific Failures

For defects in an S-PE, the procedures defined in [RFC6073] SHOULD be followed. A T-PE or S-PE may receive an unsolicited label release message from another S-PE or T-PE with various failure codes such "LOOP_DETECTED", "PW_LOOP_DETECTED", "RESOURCE_UNAVAILBALE", "BAD_STRICT_HOP", "AII_UNREACHABLE", etc. All these failure codes indicate a generic class of PW failures at an S-PE or T-PE.

When an unsolicited label release message with such a failure status code is received at T-PE then the T-PE MUST re-attempt to establish the PW immediately. However the T-PE MUST throttle its PW setup message retry attempts with an exponential backoff in situations where PW setup messages are being constantly released. It is also recommended that a T-PE detecting such a situation take action to notify an operator.

S-PEs that receive an unsolicited label release message with a failure status code should follow the following procedures:

- -i. If the label release is received from an S-PE or T-PE in the forward signaling direction then the S-PE MUST tear down both segments of the PW. The status code received in the label release message SHOULD be propagated when sending the label release for the next-segment.
- -ii. If the label release is received from an S-PE or T-PE in the reverse signaling direction, then then tear down both segments of the PW as described in i.

5.3. PW Reachability Changes

In general an established MS-PW will not be affected by next-hop changes in L2 PW reachability information.

If there is a change in next-hop of the L2 PW reachability information in the forward direction, the T-PE MAY elect to tear down the MS-PW by sending a label withdraw message to downstream S-PE or T-PE. The teardown MUST be also accompanied by a unsolicited label release message, and will be followed by and attempt to re-establish of the MS-PW by T-PE.

If there is a change in the L2 PW reachability information in the forward direction at S-PE, the S-PE MAY elect to tear down the MS-PW in both directions. A label withdrawal is sent on each direction followed by a unsolicited label release. The unsolicited label releases MUST be accompanied by the Status code "AII_UNREACHABLE". This procedure is OPTIONAL.

A change in L2 reachability information in the reverse direction has no effect on an MS-PW.

6. Operations and Maintenance (OAM)

The OAM procedures defined in [RFC6073] may be used also for MS-PWs. A PW switching point PE TLV is used [RFC6073] to record the switching points that the PW traverses.

In the case of a MS-PW where the PW Endpoints are identified though using a globally unique, FEC 129-based AII addresses, there is no PWID defined on a per segment basis. Each individual PW segment is identified by the address of adjacent S-PE(s) in conjunction with the SAI and TAI. In this case, the following TLV type (0x06) MUST be used in place of type 0x01 in the PW switching point PE TLV:

Type	Length	Description
0x06	14	L2 PW address of PW Switching Point

The above field MUST be included together with type 0x02 in the TLV once per individual PW Switching Point following the same rules and procedures as described in [RFC6073]. A more detailed description of this field is also in setion 7.4.1 of [RFC6073]. However, the length value MUST be set to 14.

7. Security Considerations

This document specifies only extensions to the protocols already defined in [RFC4447], and [RFC6073]. The extensions defined in this document do not affect the security considerations for those protocols. Note that the protocols for dynamically distributing PW Layer 2 reachability information may have their own security issues, however those protocols specifications are outside the scope of this document.

8. IANA Considerations

8.1. Corrections

IANA is requested to correct a minor error in the registry "Pseudowire Switching Point PE sub-TLV Type". The entry 0x06 "L2 PW address of the PW Switching Point" should have Length 14.

8.2. LDP TLV TYPE NAME SPACE

This document defines one new LDP TLV types. IANA already maintains a registry for LDP TLV types called "Type, Length, and Value (TLV) Type Name Space" within the "Label Distribution Protocol (LDP) Parameters" as defined by RFC5036. IANA is requested to assign on permanent basis the value (0x096E) that has been assigned to this document by early allocation (TEMPORARY - Expires 2008-11-21).:

Value	Description	Reference	Notes/Registration Date
0x096E	Bandwidth TLV	This document	

8.3. LDP Status Codes

This document defines three new LDP status codes. IANA maintains a registry of these called the "STATUS CODE NAME SPACE" in the "Label Distribution Protocol (LDP) Parameters" as defined by RFC5036. The IANA is requested to assign on permanent basis the values that has been assigned to this document by early allocation (TEMPORARY - Expires 2008-11-21):

Range/Value	E	Description	Reference
0x00000037	0	Bandwidth resources unavailable	This document
0x00000038	0	Resources Unavailable	This document
0x00000039	0	AII Unreachable	This document

8.4. BGP SAFI

IANA needs to allocate a new BGP SAFI for "Network Layer Reachability Information used for Dynamic Placement of Multi-Segment Pseudowires" from the IANA "Subsequence Address Family Identifiers (SAFI)" registry. The IANA is requested to assign on permanent basis the values that has been assigned to this document by early allocation (TEMPORARY - Expires 2008-11-21)::

Value	Description	Reference
6	Network Layer Reachability Information used for Dynamic Placement of Multi-Segment Pseudowires	This document

9. References

9.1. Normative References

[RFC6073] Martini et.al. "Segmented Pseudowire", RFC6073, January 2011

[RFC2210] Wroclawski, J. "The Use of RSVP with IETF Integrated Services", RFC 2210, September 1997

[RFC5036] Andersson, Minei, Thomas. "LDP Specification" RFC5036, October 2007

[RFC4447] "Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP)", Martini L.,et al, RFC 4447, June 2005.

[RFC5003] "Attachment Individual Identifier (AII) Types for Aggregation", Metz, et al, RFC5003, September 2007

9.2. Informative References

[RFC5254] Martini et al, "Requirements for Multi-Segment Pseudowire Emulation Edge-to-Edge (PWE3)", RFC5254, Bitar, Martini, Bocci, October 2008

[RFC5659] Bocci at al, "An Architecture for Multi-Segment Pseudo Wire Emulation Edge-to-Edge", RFC5659,October 2009.

[RFC4760] Bates, T., Rekhter, Y., Chandra, R. and D. Katz, "Multiprotocol Extensions for BGP-4", RFC 4760, January 2007.

[RFC6074] E. Rosen, W. Luo, B. Davie, V. Radoaca, "Provisioning, Autodiscovery, and Signaling in L2VPNs", RFC6074, January 2011

[RFC4271] Rekhter, Y., et al, "A Border Gateway Protocol 4 (BGP-4)", RFC4271, January 2006

[RFC6391] Bryant, S., et al, "Flow-Aware Transport of Pseudowires over an MPLS Packet Switched Network", RFC6391, November 2011

[RFC4090] Pan, P., et al, "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", RFC4090, May 2005

10. Major Co-authors

The editors gratefully acknowledge the following additional co-authors: Mustapha Aissaoui, Nabil Bitar, Mike Loomis, David McDysan, Chris Metz, Andy Malis, Jason Rusmeisel, Himanshu Shah, Jeff Sugimoto.

11. Acknowledgements

The editors also gratefully acknowledge the input of the following people: Mike Duckett, Paul Doolan, Prayson Pate, Ping Pan, Vasile Radoaca, Yeongil Seo, Yetik Serbest, Yuichiro Wada.

12. Author's Addresses

Luca Martini
Cisco Systems, Inc.
9155 East Nichols Avenue, Suite 400
Englewood, CO, 80112
e-mail: lmartini@cisco.com

Matthew Bocci
Alcatel-Lucent,
Voyager Place
Shoppenhangers Road
Maidenhead
Berks, UK
e-mail: matthew.bocci@alcatel-lucent.com

Florin Balus
Alcatel-Lucent
701 E. Middlefield Rd.
Mountain View, CA 94043
e-mail: florin.balus@alcatel-lucent.com

Nabil Bitar
Verizon
40 Sylvan Road
Waltham, MA 02145
e-mail: nabil.bitar@verizon.com

Himanshu Shah
Ciena Corp
35 Nagog Park,
Acton, MA 01720
e-mail: hshah@ciena.com

Mustapha Aissaoui
Alcatel-Lucent
600 March Road
Kanata
ON, Canada
e-mail: mustapha.aissaoui@alcatel-lucent.com

Jason Rusmisel
Alcatel-Lucent
600 March Road
Kanata
ON, Canada
e-mail: Jason.rusmisel@alcatel-lucent.com

Yetik Serbest
AT&T Labs
9505 Arboretum Blvd.
Austin, TX 78759
e-mail: yetik.serbest@labs.att.com

Andrew G. Malis
Verizon
117 West St.
Waltham, MA 02451
e-mail: andrew.g.malis@verizon.com

Chris Metz
Cisco Systems, Inc.
3700 Cisco Way
San Jose, Ca. 95134
e-mail: chmetz@cisco.com

David McDysan
Verizon
22001 Loudoun County Pkwy
Ashburn, VA, USA 20147
e-mail: dave.mcdysan@verizon.com

Jeff Sugimoto
Alcatel-Lucent
701 E. Middlefield Rd.
Mountain View, CA 94043
e-mail: jeffery.sugimoto@alcatel-lucent.com

Mike Duckett
ATT
Lindbergh Center D481
575 Morosgo Dr
Atlanta, GA 30324
e-mail: md9308@att.com

Mike Loomis
Alcatel-Lucent
701 E. Middlefield Rd.
Mountain View, CA 94043
e-mail: mike.loomis@alcatel-lucent.com

Paul Doolan
Coriant GmbH & Co. KG
St Martin Str. 76
81541 Munich
e-mail: paul.doolan@coriant.com

Ping Pan
Infinera
e-mail: ppan@infinera.com

Prayson Pate
Overture Networks, Inc.
507 Airport Blvd, Suite 111
Morrisville, NC, USA 27560
e-mail: prayson.pate@overturenetworks.com

Vasile Radoaca
Alcatel-Lucent
388 NINGQIAO RD
PUDONG JINQIAO
SHANGHAI 201206
CHINA
email: vasile.radoaca@alcatel-lucent.com

Yuichiro Wada
NTT
3-9-11 Midori-cho
Musashino-shi
Tokyo 180-8585
Japan
e-mail: wada.yuichiro@lab.ntt.co.jp

Yeong-il Seo
Korea Telecom Corp.
463-1 Jeonmin-dong, Yusung-gu
Daejeon, Korea
e-mail: yohan.seo@kt.com

Copyright Notice

Expiration Date: June 2014