

Network Working Group  
Internet Draft  
Expiration Date: March 2011  
Intended status: Standards Track

Luca Martini  
Chris Metz  
Cisco Systems Inc.

Thomas D. Nadeau  
LucidVision

Matthew Bocci  
Mustapha Aissaoui  
Alcatel-Lucent

September 16, 2010

## Segmented Pseudowire

[draft-ietf-pwe3-segmented-pw-18.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>.

This Internet-Draft will expire on Expiration Date: March 2011

### Abstract

This document describes how to connect pseudowires (PW) between different Packet Switched Network (PSN) domains or between two or more distinct PW control plane domains, where a control plane domain uses a common control plane protocol or instance of that protocol for

a given PW. The different PW control plane domains may belong to independent autonomous systems, or the PSN technology is heterogeneous, or a PW might need to be aggregated at a specific PSN point. The PW packet data units are simply switched from one PW to another without changing the PW payload.

Table of Contents

<a href="#">1</a>	Introduction .....	<a href="#">4</a>
<a href="#">2</a>	Specification of Requirements .....	<a href="#">6</a>
<a href="#">3</a>	Terminology .....	<a href="#">6</a>
<a href="#">4</a>	General Description .....	<a href="#">7</a>
<a href="#">5</a>	PW Switching and Attachment Circuit Type .....	<a href="#">10</a>
<a href="#">6</a>	Applicability .....	<a href="#">10</a>
<a href="#">7</a>	MPLS-PW to MPLS-PW Switching .....	<a href="#">10</a>
<a href="#">7.1</a>	Static Control plane switching .....	<a href="#">11</a>
<a href="#">7.2</a>	Two LDP control planes using the same FEC type .....	<a href="#">11</a>
<a href="#">7.2.1</a>	FEC 129 Active/Passive T-PE Election Procedure .....	<a href="#">12</a>
<a href="#">7.3</a>	LDP FEC 128 to LDP using the generalized FEC 129 .....	<a href="#">12</a>
<a href="#">7.4</a>	LDP Switching Point PE TLV .....	<a href="#">13</a>
<a href="#">7.4.1</a>	PW Switching Point Sub-TLVs .....	<a href="#">14</a>
<a href="#">7.4.2</a>	Adaptation of Interface Parameters .....	<a href="#">16</a>
<a href="#">7.5</a>	Group ID .....	<a href="#">16</a>
<a href="#">7.6</a>	PW Loop Detection .....	<a href="#">16</a>
<a href="#">8</a>	MPLS-PW to L2TPv3-PW Control Plane Switching .....	<a href="#">17</a>
<a href="#">8.1</a>	Static MPLS and L2TPv3 PWs .....	<a href="#">17</a>
<a href="#">8.2</a>	Static MPLS PW and Dynamic L2TPv3 PW .....	<a href="#">17</a>
<a href="#">8.3</a>	Static L2TPv3 PW and Dynamic LDP/MPLS PW .....	<a href="#">18</a>
<a href="#">8.4</a>	Dynamic LDP/MPLS and L2TPv3 PWs .....	<a href="#">18</a>
<a href="#">8.4.1</a>	Session Establishment .....	<a href="#">18</a>
<a href="#">8.4.2</a>	Adaptation of PW Status message .....	<a href="#">19</a>
<a href="#">8.4.3</a>	Session Tear Down .....	<a href="#">19</a>
<a href="#">8.5</a>	Adaptation of L2TPv3 AVPs to Interface Parameters ....	<a href="#">20</a>
<a href="#">8.6</a>	Switching Point TLV in L2TPv3 .....	<a href="#">21</a>
<a href="#">8.7</a>	L2TPv3 and MPLS PW Data Plane .....	<a href="#">21</a>
<a href="#">8.7.1</a>	Mapping the MPLS Control Word to L2TP .....	<a href="#">22</a>
<a href="#">9</a>	Operation And Management .....	<a href="#">23</a>
<a href="#">9.1</a>	Extensions to VCCV to Support MS-PWs .....	<a href="#">23</a>
<a href="#">9.2</a>	MPLS PW to L2TPv3 PW OAM .....	<a href="#">23</a>
<a href="#">9.3</a>	MPLS PW to MPLS PW OAM Data Plane Indication .....	<a href="#">23</a>
<a href="#">9.4</a>	Signaling OAM Capabilities for Switched Pseudowires ..	<a href="#">24</a>
<a href="#">9.5</a>	OAM Capability for MS-Pws Demultiplexed using MPLS ...	<a href="#">24</a>
<a href="#">9.5.1</a>	MS-PW and VCCV CC Type 1 .....	<a href="#">25</a>
<a href="#">9.5.2</a>	MS-PW and VCCV CC type 2 .....	<a href="#">25</a>
<a href="#">9.5.3</a>	MS-PW and VCCV CC type 3 .....	<a href="#">25</a>
<a href="#">9.6</a>	MS-PW VCCV Operations .....	<a href="#">25</a>
<a href="#">9.6.1</a>	VCCV Echo Message Processing .....	<a href="#">27</a>
<a href="#">9.6.1.1</a>	Sending a VCCV Echo Request .....	<a href="#">27</a>
<a href="#">9.6.1.2</a>	Receiving a VCCV Echo Request .....	<a href="#">27</a>
<a href="#">9.6.1.3</a>	Receiving a VCCV Echo Reply .....	<a href="#">28</a>
<a href="#">9.6.2</a>	Detailed VCCV procedures .....	<a href="#">28</a>
<a href="#">9.6.2.1</a>	End to End Connectivity Verification Between T-PEs ...	<a href="#">28</a>



<a href="#">9.6.2.2</a>	Partial Connectivity Verification from T-PE .....	<a href="#">29</a>
<a href="#">9.6.2.3</a>	Partial connectivity verification between S-PEs .....	<a href="#">30</a>
<a href="#">9.6.2.4</a>	MS-PW Path Verification .....	<a href="#">30</a>
<a href="#">9.6.2.5</a>	MS-PW Path Trace .....	<a href="#">31</a>
<a href="#">10</a>	Mapping Switched Pseudowire Status .....	<a href="#">32</a>
<a href="#">10.1</a>	S-PE initiated PW status messages .....	<a href="#">34</a>
<a href="#">10.1.1</a>	Local PW2 transmit direction fault .....	<a href="#">35</a>
<a href="#">10.1.2</a>	Local PW1 transmit direction fault .....	<a href="#">35</a>
<a href="#">10.1.3</a>	Local PW2 receive direction fault .....	<a href="#">36</a>
<a href="#">10.1.4</a>	Local PW1 receive direction fault .....	<a href="#">36</a>
<a href="#">10.1.5</a>	Clearing Faults .....	<a href="#">36</a>
<a href="#">10.2</a>	PW status messages and S-PE TLV processing .....	<a href="#">36</a>
<a href="#">10.3</a>	T-PE processing of PW status messages .....	<a href="#">37</a>
<a href="#">10.4</a>	Pseudowire Status Negotiation Procedures .....	<a href="#">37</a>
<a href="#">10.5</a>	Status Dampening .....	<a href="#">37</a>
<a href="#">11</a>	Peering Between Autonomous Systems .....	<a href="#">37</a>
<a href="#">12</a>	Congestion Considerations .....	<a href="#">37</a>
<a href="#">13</a>	Security Considerations .....	<a href="#">38</a>
<a href="#">13.1</a>	Data Plane Security .....	<a href="#">38</a>
<a href="#">13.1.1</a>	VCCV Security considerations .....	<a href="#">38</a>
<a href="#">13.2</a>	Control Protocol Security .....	<a href="#">38</a>
<a href="#">14</a>	IANA Considerations .....	<a href="#">40</a>
<a href="#">14.1</a>	L2TPv3 AVP .....	<a href="#">40</a>
<a href="#">14.2</a>	LDP TLV TYPE .....	<a href="#">40</a>
<a href="#">14.3</a>	LDP Status Codes .....	<a href="#">40</a>
<a href="#">14.4</a>	L2TPv3 Result Codes .....	<a href="#">40</a>
<a href="#">14.5</a>	New IANA Registries .....	<a href="#">41</a>
<a href="#">15</a>	Normative References .....	<a href="#">41</a>
<a href="#">16</a>	Informative References .....	<a href="#">42</a>
<a href="#">17</a>	Author's Addresses .....	<a href="#">43</a>
<a href="#">18</a>	Additional Authors .....	<a href="#">44</a>

## 1. Introduction

The PWE3 Architecture [[RFC3985](#)], defines the signaling and encapsulation techniques for establishing SS-PWs between a pair of terminating PEs. MS-PWs are most useful in two general cases:

- i. In some cases it is not possible, desirable or feasible to establish a PW control channel between the terminating source and destination PEs. At a minimum PW control channel establishment requires knowledge of and reachability to the remote (terminating) PE IP address. The local (terminating) PE may not have access to this information related to



topology, operational or security constraints.

An example is the inter-AS L2VPN scenario where the terminating PEs reside in different provider networks (ASes) and it is the practice to cryptographically sign all control traffic exchanged between two networks. Technically a SS-PW could be used but this would require to cryptographically sign on ALL terminating source and destination PE nodes. An MS-PW allows the providers to confine key administration to just the PW switching points connecting the two domains.

A second example might involve a single AS where the PW setup path between the terminating PEs is computed by an external entity. Assume a full mesh of PWE3 control channels established between PE-A, PE-B and PE-C. A client-layer L2 connection tunneled through a PW is required between terminating PE-A and PE-C. The external entity computes a PW setup path that passes through PE-B. This results in two discrete PW segments being built: one between PE-A and PE-B and one between PE-B and PE-C. The successful client-layer L2 connection between terminating PE-A and terminating PE-C requires that PE-B performs the PWE3 switching process.

A third example involves the use of PWs in hierarchical IP/MPLS networks. Access networks connected to a backbone use PWs to transport customer payloads between customer sites serviced by the same access network and up to the edge of the backbone where they can be terminated or switched onto a succeeding PW segment crossing the backbone. The use of PWE3 switching between the access and backbone networks can potentially reduce the PWE3 control channels and routing information processed by the access network T-PEs.

It should be noted that PWE3 switching does not help in any way to reduce the amount of PW state supported by each access network T-PE.

- ii. In some applications PWE3 edge to edge signaling and encapsulation protocols are different. The terminating PEs are connected to networks employing different PW signaling and encapsulation protocols. In this case it is not possible to use a SS-PW. A MS-PW with the appropriate interworking performed at the PW switching points can enable PW connectivity between the terminating PEs in this scenario.

A more detailed discussion of the requirements pertaining to MS-PWs can be found in [[RFC5254](#)].





There are four different mechanisms to establish PWs:

- i. Static configuration of the PW (MPLS or L2TPv3).
- ii. LDP using FEC 128
- iii. LDP using the generalized FEC 129
- iv. L2TPv3

While MS-PW are composed of PW segments, each PW segment cannot function independently, as the PW service is always instantiated across the complete MS-PW. Hence no PW segments, can be signaled, or be operational without the complete MS-PW being signaled at once.

## **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 [[RFC2119](#)].

## **3. Terminology**

- PW Terminating Provider Edge (T-PE). A PE where the customer-facing attachment circuits (ACs) are bound to a PW forwarder. A Terminating PE is present in the first and last segments of a MS-PW. This incorporates the functionality of a PE as defined in [[RFC3985](#)].
- Single-Segment Pseudowire (SS-PW). A PW set up directly between two T-PE devices. The PW label is unchanged between the originating and terminating T-PEs.
- Multi-Segment Pseudowire (MS-PW). A static or dynamically configured set of two or more contiguous PW segments that behave and function as a single point-to-point PW. Each end of a MS-PW by definition MUST terminate on a T-PE.
- PW Segment. A part of a single-segment or multi-segment PW, which traverses one PSN tunnel in each direction between two PE devices, T-PEs and/or S-PEs (switching PE).
- PW Switching Provider Edge (S-PE). A PE capable of switching the control and data planes of the preceding and succeeding PW segments in a MS-PW. The S-PE terminates the PSN tunnels of the preceding and succeeding segments of the MS-PW. It therefore includes a PW switching point for an MS-PW. A PW switching point is never the S-PE and the T-PE for the same MS-PW. A PW switching point runs necessary protocols to set up and manage PW segments with other PW switching points and terminating PEs. An S-PE can



exist anywhere a PW must be processed or policy applied. It is therefore not limited to the edge of a provider network.

- MS-PW path. The set of S-PEs that will be traversed in sequence to form the MS-PW.

#### 4. General Description

A pseudowire (PW) is a mechanism that carries the essential elements of an emulated service from one PE to one or more other PEs over a PSN as described in Figure 1 and in [RFC3985]. Many providers have deployed PWs as a means of migrating existing (or building new) L2VPN services (e.g., Frame Relay, ATM, or Ethernet) on to a PSN.

PWs may span multiple domains of the same or different provider networks. In these scenarios PW control channels (i.e. targeted LDP, L2TPv3) and PWs will cross AS boundaries.

Inter-AS L2VPN functionality is currently supported and several techniques employing MPLS encapsulation and LDP signaling have been documented [RFC4364]. It is also straightforward to support the same inter-AS L2VPN functionality employing L2TPv3. In this document we define methodology to switch a PW between different Packet Switched Network (PSN) domains or between two or more distinct PW control plane domains

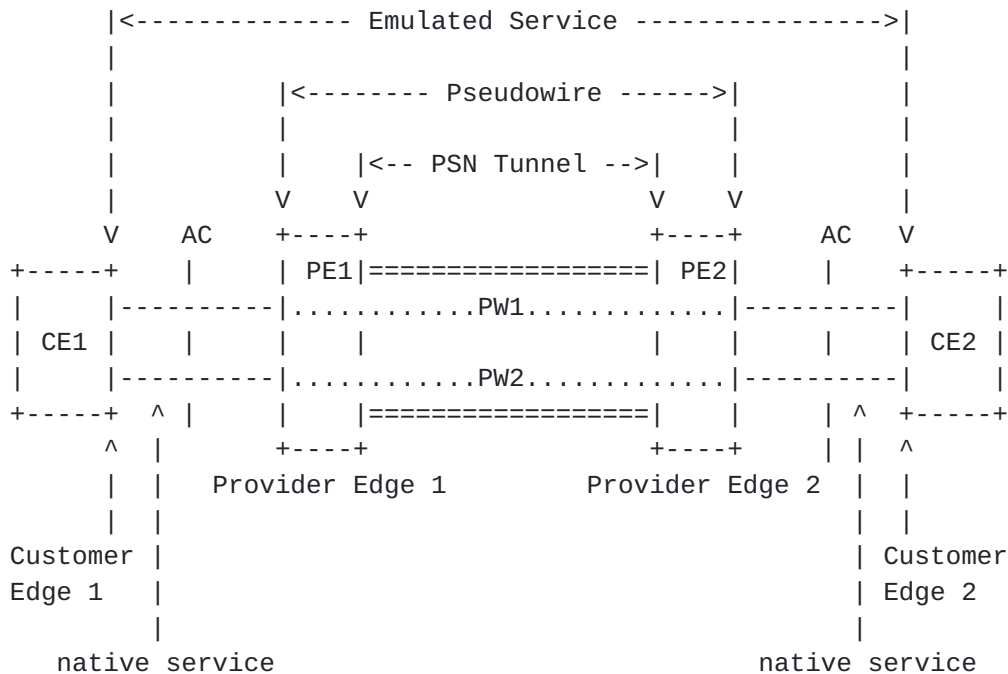


Figure 1: PWE3 Reference Model



There are two methods for switching a PW between two PW domains. In the first method (Figure 2), the two separate control plane domains terminate on different PEs.

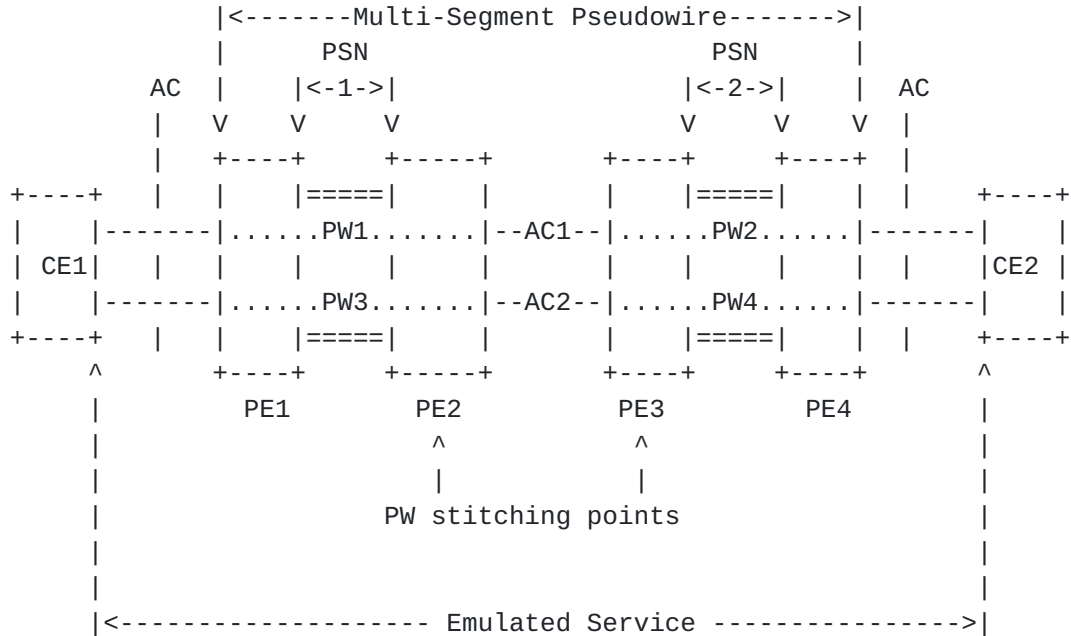


Figure 2: PW Switching using ACs Reference Model

In Figure 2, pseudowires in two separate PSNs are stitched together using native service attachment circuits. PE2 and PE3 only run the control plane for the PSN to which they are directly attached. At PE2 and PE3, PW1 and PW2 are connected using attachment circuit AC1, while PW3 and PW4 are connected using attachment circuit AC2.

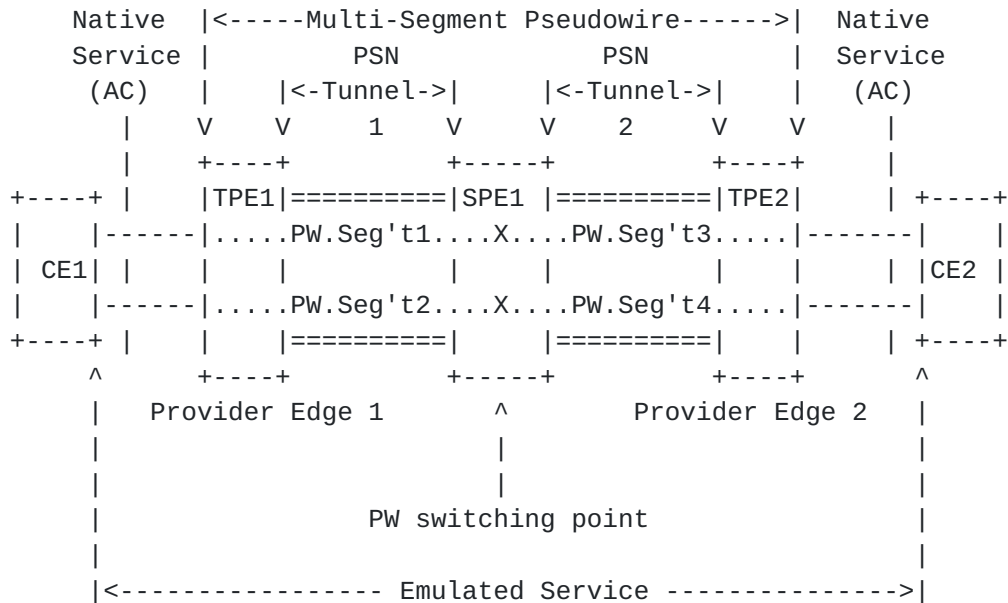




Figure 3: MS-PW Reference Model

In Figure 3 SPE1 runs two separate control planes: one toward TPE1, and one toward TPE2. The PW switching point (S-PE) is configured to connect PW Segment 1 and PW Segment 3 together to complete the Multi-Segment PW between TPE1 and TPE2. PW Segment 1 and PW Segment 3 MUST be of the same PW type, but PSN Tunnel 1 and PSN Tunnel 2 need not be the same technology. In the latter case, if the PW is switched to a different technology, the PEs must adapt the PDU encapsulation between the different PSN technologies. In the case where PSN Tunnel 1 and PSN Tunnel 2 are the same technology the PW PDU does not need to be modified, and PDUs are then switched between the pseudowires at the PW label level.

It should be noted that it is possible to adapt one PSN technology to a different one, for example MPLS over an IP or GRE [[RFC4023](#)] encapsulation, but this is outside the scope of this document. Further, one could perform an interworking function on the PWs themselves at the S-PE, allowing conversion from one PW type to another, but this is also outside the scope of this document.

This document describes procedures for building multi-segment pseudowires using manual configuration of the switching point PE1. Other documents may build on this base specification to automate the configuration and selection of S-PE1. All elements of the establishment of end-to-end MS-PWs including routing and signaling are out of scope of this document and any discussion in this document serve purely as examples. It should also be noted that a PW can traverse multiple PW switching points along its path, and the edge PEs will not require any specific knowledge of how many S-PEs the PW has traversed (though this may be reported for troubleshooting purposes).

The general approach taken for MS-PWs is to connect the individual control planes by passing along any signaling information immediately upon reception. First the S-PE is configured to switch a PW segment from a specific peer to another PW segment destined for a different peer. No control messages are exchanged yet as the S-PE does not have enough information to actually initiate the PW setup messages. However, if a session does not already exist, a control protocol (LDP/L2TP) session MAY be setup. In this model the MS-PW setup is starting from the T-PE devices. Once the T-PE is configured it sends the PW control setup messages. These messages are received by the S-PE, and immediately used to form the PW setup messages for the next SS-PW of the MS-PW.





## **5. PW Switching and Attachment Circuit Type**

The PWs in each PSN are established independently, with each PSN being treated as a separate PW domain. For example, in Figure 2 for the case of MPLS PSNs, PW1 is setup between PE1 and PE2 using the LDP targeted session as described in [[RFC4447](#)], and at the same time a separate pseudowire, PW2, is setup between PE3 and PE4. The ACs for PW1 and PW2 at PE2 and PE3 MUST be configured such that they are the same PW type e.g. ATM VCC, Ethernet VLAN, etc.

## **6. Applicability**

The general applicability of MS-PWs and their relationship to L2VPNs is described in [[RFC5659](#)]. The applicability of a PW type, as specified in the relevant RFC for that encapsulation (e.g. [[RFC4717](#)] for ATM), applies to each segment. This section describes further applicability considerations.

As with SS-PWs, the performance of any segment will be limited by the performance of the underlying PSN. The performance may be further degraded by the emulation process, and performance degradation may be further increased by traversing multiple PW segments. Furthermore, the overall performance of an MS-PW is no better than the worst performing segment of that MS-PW.

Since different PSN types may be able to achieve different maximum performance objectives, it is necessary to carefully consider which PSN types are used along the path of a MS-PW.

## **7. MPLS-PW to MPLS-PW Switching**

Referencing Figure 3, T-PE1 set up PW segment 1 using the LDP targeted session as described in [[RFC4447](#)], at the same time a separate pseudowire PW segment 3 is setup to T-PE2. Each PW is configured independently on the PEs, but on S-PE1 pseudowire PW segment 1 is connected to pseudowire PW segment 3. PDUs are then switched between the pseudowires at the PW label level. Hence the data plane does not need any special knowledge of the specific pseudowire type. A simple standard MPLS label swap operation is sufficient to connect the two PWs, and in this case the PW adaptation function cannot be used. However when pushing a new PSN label the TTL SHOULD be set to 255, or some other locally configured fixed value.

This process can be repeated as many times as necessary, the only limitation to the number of S-PEs traversed is imposed by the TTL field of the PW MPLS Label. The setting of the TTL of the PW MPLS



label is a matter of local policy on the originating PE, but SHOULD be set to 255. However if the PW PDU contains an OAM packet then the TTL can be set to the required value as explained later in this document.

There are three different MPLS to MPLS PW setup mechanisms:

- i. Static configuration of the PW.
- ii. LDP using FEC 128
- iii. LDP using the generalized FEC 129

This results in four distinct PW switching situations that are significantly different, and must be considered in detail:

- i. PW Switching between two static control planes.
- ii. Static Control plane switching to LDP dynamic control plane.
- iii. Two LDP control planes using the same FEC type
- iv. LDP using FEC 128, to LDP using the generalized FEC 129

### **7.1. Static Control plane switching**

In the case of two static control planes the S-PE MUST be configured to direct the MPLS packets from one PW into the other. There is no control protocol involved in this case. When one of the control planes is a simple static PW configuration and the other control plane is a dynamic LDP FEC 128 or generalized PW FEC, then the static control plane should be considered similar to an attachment circuit (AC) in the reference model of Figure 1. The switching point PE SHOULD signal the appropriate PW status if it detects a failure in sending or receiving packets over the static PW segment. In the absence of a PW status communication mechanism when the PW is statically configured, the status communicated to the dynamic LDP PW will be limited to local interface failures. In this case, the S-PE PE behaves in a very similar manner to a T-PE, assuming an active signaling role. This means that the S-PE will immediately send the LDP Label Mapping message if the static PW is deemed to be UP.

### **7.2. Two LDP control planes using the same FEC type**

The S-PE SHOULD assume an initial passive role. This means that when independent PWs are configured on the switching point, the LSR does not advertise the LDP PW FEC mapping until it has received at least one of the two PW LDP FECs from a remote PE. This is necessary because the switching point LSR does not know a priori what the interface parameter field in the initial FEC advertisement will contain.

If one of the S-PEs doesn't accept an LDP Label Mapping message then



a Label Release message may be sent back to the originator T-PE depending on the cause of the error. LDP liberal label retention mode still applies, hence if a PE is simply not configured yet, the label mapping is stored for future use. A MS-PW is declared UP only when all the constituent SS-PWs are UP.

The Pseudowire Identifier (PWID), as defined in [RFC4447] is a unique number between each pair of PEs. Hence Each SS-PW that forms an MS-PW may have a different PWID. In the case of The Generalized PW FEC, the AGI/SAI/TAI may have to also be different for some, or sometimes all, SS-PWs.

### **7.2.1. FEC 129 Active/Passive T-PE Election Procedure**

When a MS-PW is signaled using FEC 129, each T-PE might independently start signaling the MS-PW. If the MS-PW path is not statically configured, in certain cases the signaling procedure could result in an attempt to setup each direction of the MS-PW through different S-PEs. If an operator wishes to avoid this situation one of the T-PE MUST start the PW signaling (active role), while the other waits to receive the LDP label mapping before sending the respective PW LDP label mapping message. (passive role). When the MS-PW path not statically configured, the Active T-PE (the ST-PE) and the passive T-PE (the TT-PE) MUST be identified before signaling is initiated for a given MS-PW.

The determination of which T-PE assume the active role SHOULD be done as follows:

The SAII and TAI are compared as unsigned integers, if the SAII is larger, then the T-PE assumes the active role.

The selection process to determine which T-PE assumes the active role MAY be superseded by manual provisioning. In this case one of the T-PEs MUST be set to active role, and the other one MUST be set to passive role.

### **7.3. LDP FEC 128 to LDP using the generalized FEC 129**

When a PE is using the generalized FEC 129, there are two distinct roles that a PE can assume: active and passive. A PE that assumes the active role will send the LDP PW setup message, while a passive role PE will simply reply to an incoming LDP PW setup message. The S-PE PE, will always remain passive until a PWID FEC 128 LDP message is received, which will cause the corresponding generalized PW FEC LDP message to be formed and sent. If a generalized FEC PW LDP message is



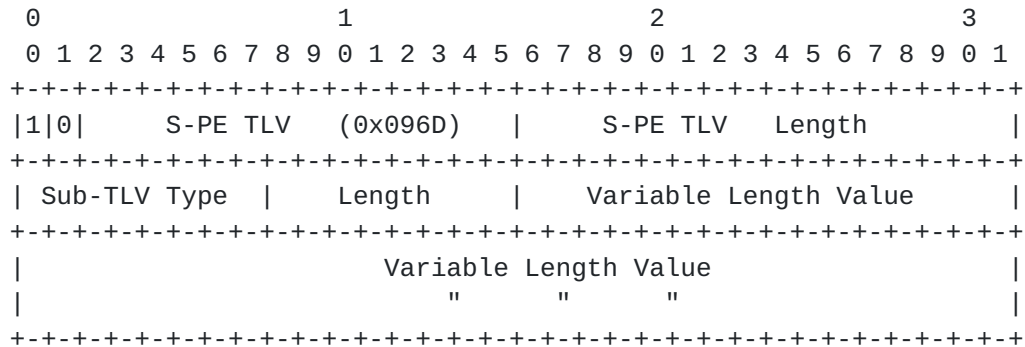
received while the switching point PE is in a passive role, the corresponding PW FEC 128 LDP message will be formed and sent.

PW IDs need to be mapped to the corresponding AGI/TAI/SAI and vice versa. This can be accomplished by local S-PE configuration, or by some other means, such as some form of auto discovery. Such other means are outside the scope of this document.

**7.4. LDP Switching Point PE TLV**

The edge to edge PW might traverse several switching points, in separate administrative domains. For management and troubleshooting reasons it is useful to record information about the switching points that the PW traverses. This is accomplished by using a PW switching Point TLV.

Sending the PW switching Point TLV (S-PE TLV) is OPTIONAL, however the PE or S-PE MUST process the TLV upon reception. The "U" bit MUST be set for backward compatibility with T-PEs that do not support the MS-PW extensions described in the document. The S-PE TLV MAY appear only once for each switching point traversed, and it cannot be of length zero. The S-PE TLV is appended to the PW FEC at each switching point, and the order of the S-PE TLVs in the LDP message MUST be preserved. The S-PE TLV is necessary to support some of the VCCV functions for MS-PWs. See [Section 9.5](#) for more details. The S-PE TLV is encoded as follows:



[note] LDP TLV type is pending IANA approval.

- S-PE TLV Length

Specifies the total length of all the following S-PE TLV fields in octets





- Sub-TLV Type

Encodes how the Value field is to be interpreted.

- Length

Specifies the length of the Value field in octets.

- Value

Octet string of Length octets that encodes information to be interpreted as specified by the Type field.

PW Switching point sub-TLV Types are assigned by IANA according to the process defined in the "IANA Allocations" section below.

For local policy reasons, a particular S-PE can filter out all S-PE TLVs in a label mapping message that traverses it and not include its own S-PE TLV. In this case, from any upstream PE, it will appear as if this particular S-PE is the T-PE. This might be necessary, depending on local policy if the S-PE is at the service provider administrative boundary. It should also be noted that because there are no S-PE TLVs describing the path beyond the S-PE that removed them, VCCV will only work as far as that S-PE .

#### **7.4.1. PW Switching Point Sub-TLVs**

The S-PE TLV contains sub-TLVs that describe various characteristics of the S-PE traversed. The S-PE TLV MUST contain the appropriate mandatory sub-TLVs specified below. Below are the definitions of PW Switching Point Sub-TLVs defined in this document:

- PW ID of last PW segment traversed.

This is only applicable if the last PW segment traversed used LDP FEC 128 to signal the PW. This sub-TLV type contains a PW ID in the format of the PWID described in [[RFC4447](#)]. This is just a 32 bit unsigned integer number.

- PW Switching Point description string.

An OPTIONAL description string of text up to 80 characters long. Human-readable text MUST be provided in the UTF-8 character set using the Default Language [[RFC2277](#)].







- L2 PW address of PW Switching Point (recommended format).

This sub-TLV type contains a L2 PW address of PW Switching Point in the format described in [Section 3.2 of \[RFC5003\]](#). This includes the AII type field, and length, as well as the L2 PW address with the AC ID field set to zero.

#### **[7.4.2. Adaptation of Interface Parameters](#)**

[RFC4447] defines several interface parameters, which are used by the Network Service Processing (NSP) to adapt the PW to the Attachment Circuit (AC). The interface parameters are only used at the end points, and MUST be passed unchanged across the S-PE. However the following interface parameters MAY be modified as follows:

- 0x03 Optional Interface Description string This Interface parameter MAY be modified, or altogether removed from the FEC element depending on local configuration policies.
- 0x09 Fragmentation indicator This parameter MAY be inserted in the FEC by the switching point if it is capable of re-assembly of fragmented PW frames according to [\[RFC4623\]](#).
- 0x0C VCCV parameter This Parameter contains the CC type, and CV type bit fields. The CV type bit field MUST be reset to reflect the CV type supported by the S-PE. CC type bit field MUST have bit 1 "Type 2: MPLS Router Alert Label " set to 0. The other bit fields MUST be reset to reflect the CC type supported by the S-PE.

#### **[7.5. Group ID](#)**

The Group ID (GR ID) is used to reduce the number of status messages that need to be sent by the PE advertising the PW FEC. The GR ID has local significance only, and therefore MUST be mapped to a unique GR ID allocated by the S-PE PE.

#### **[7.6. PW Loop Detection](#)**

A switching point PE SHOULD inspect the PW switching Point TLV, to verify that it's own IP address does not appears in it. If the PE's IP address appears in a received PW switching Point TLV, the PE SHOULD break the loop, and send a label release message with the following error code:

Assignment E Description



0x0000003A 0 "PW Loop Detected"

[ note: error code pending IANA allocation ]

If an S-PE along the MS-PW removed all S-PE TLVs, as mentioned above, this loop detection method will fail.

## **8. MPLS-PW to L2TPv3-PW Control Plane Switching**

Both MPLS and L2TPv3 PWs may be static or dynamic. This results in four possibilities when switching between L2TPv3 and MPLS.

- i. Switching between MPLS and L2TPv3 static control planes.
- ii. Switching between a static MPLS PW and a dynamic L2TPv3 PW.
- iii. Switching between a static L2TPv3 PW and a dynamic MPLS PW.
- iv. Switching between a dynamic MPLS PW and a dynamic L2TPv3 PW.

### **8.1. Static MPLS and L2TPv3 PWs**

In the case of two static control planes, the S-PE MUST be configured to direct packets from one PW into the other. There is no control protocol involved in this case. The configuration MUST include which MPLS PW Label maps to which L2TPv3 Session ID (and associated Cookie, if present) as well as which MPLS Tunnel Label maps to which PE destination IP address.

### **8.2. Static MPLS PW and Dynamic L2TPv3 PW**

When a statically configured MPLS PW is switched to a dynamic L2TPv3 PW, the static control plane should be considered identical to an attachment circuit (AC) in the reference model of Figure 1. The switching point PE SHOULD signal the appropriate PW status if it detects a failure in sending or receiving packets over the static PW. Because the PW is statically configured, the status communicated to the dynamic L2TPv3 PW will be limited to local interface failures. In this case, the S-PE PE behaves in a very similar manner to a T-PE, assuming an active role.





### 8.3. Static L2TPv3 PW and Dynamic LDP/MPLS PW

When a statically configured L2TPv3 PW is switched to a dynamic LDP/MPLS PW, then the static control plane should be considered identical to an attachment circuit (AC) in the reference model of Figure 1. The switching point PE SHOULD signal the appropriate PW status (via an L2TPv3 SLI message) if it detects a failure in sending or receiving packets over the static PW. Because the PW is statically configured, the status communicated to the dynamic LDP/MPLS PW will be limited to local interface failures. In this case, the S-PE PE behaves in a very similar manner to a T-PE, assuming an active role.

### 8.4. Dynamic LDP/MPLS and L2TPv3 PWs

When switching between dynamic PWs, the switching point always assumes an initial passive role. Thus, it does not initiate an LDP/MPLS or L2TPv3 PW until it has received a connection request (Label Mapping or ICRQ) from one side of the node. Note that while MPLS PWs are made up of two unidirectional LSPs bonded together by FEC identifiers, L2TPv3 PWs are bidirectional in nature, setup via a 3-message exchange (ICRQ, ICRP and ICCN). Details of Session Establishment, Tear Down, and PW Status signaling are detailed below.

#### 8.4.1. Session Establishment

When the S-PE receives an L2TPv3 ICRQ message, the identifying AVPs included in the message are mapped to FEC identifiers and sent in an LDP label mapping message. Conversely, if an LDP Label Mapping message is received, it is either mapped to an ICRP message or causes an L2TPv3 session to be initiated by sending an ICRQ.

Following are two example exchanges of messages between LDP and L2TPv3. The first is a case where an L2TPv3 T-PE initiates an MS-PW, the second is a case where an MPLS T-PE initiates an MS-PW.

```

PE 1 (L2TPv3)      PW Switching Node      PE3 (MPLS/LDP)

AC "Up"
L2TPv3 ICRQ ---->
                    LDP Label Mapping ---->
                                AC "Up"
                                <--- LDP Label Mapping
                    <--- L2TPv3 ICRP
L2TPv3 ICCN ---->
<----- MS-PW Established ----->
    
```



```

PE 1 (MPLS/LDP)      PW Switching Node      PE3 (L2TPv3)

AC "Up"
LDP Label Mapping --->
                        L2TPv3 ICRQ --->
                                <--- L2TPv3 ICRP
<--- LDP Label Mapping
                        L2TPv3 ICCN --->
                                AC "Up"
<----- MS-PW Established ----->

```

### 8.4.2. Adaptation of PW Status message

L2TPv3 uses the SLI message to indicate a interface status change (such as the interface transitioning from "Up" or "Down"). MPLS/LDP PWs either signal this via an LDP Label Withdraw or the PW Status Notification message defined in [section 4.4 of \[RFC4447\]](#). The LDP status TLV bit SHOULD be mapped to the L2TPv3 equivalent Extended Circuit Status Values TLV specified in [\[RFC5641\]](#).

### 8.4.3. Session Tear Down

L2TPv3 uses a single message, CDN, to tear down a pseudowire. The CDN message translates to a Label Withdraw message in LDP. Following are two example exchanges of messages between LDP and L2TPv3. The first is a case where an L2TPv3 T-PE initiates the termination of an MS-PW, the second is a case where an MPLS T-PE initiates the termination of an MS-PW.

```

PE 1 (L2TPv3)      PW Switching Node      PE3 (MPLS/LDP)

AC "Down"
L2TPv3 CDN --->
                        LDP Label Withdraw --->
                                AC "Down"
                                <-- LDP Label Release
<----- MS-PW Data Path Down ----->

```



PE 1 (MPLS LDP)      PW Switching Node      PE3 (L2TPv3)

AC "Down"

LDP Label Withdraw    --->

                         L2TPv3 CDN -->

                         <-- LDP Label Release

   AC "Down"

<----- MS-PW Data Path Down ----->

### **8.5. Adaptation of L2TPv3 AVPs to Interface Parameters**

[RFC4447] defines several interface parameters which MUST be mapped to the equivalent AVPs in L2TPv3 setup messages.

\* Interface MTU

The Interface MTU parameter is mapped directly to the L2TP Interface MTU AVP defined in [\[RFC4667\]](#)

\* Max Number of Concatenated ATM cells

This interface parameter is mapped directly to the L2TP "ATM Maximum Concatenated Cells AVP" described in [section 6 of \[RFC4454\]](#).

\* PW Type

The PW Type defined in [\[RFC4446\]](#) is mapped to the L2TPv3 "PW Type" AVP defined in [\[RFC3931\]](#).

\* PW ID (FEC 128)

For FEC 128, the PW ID is mapped directly to the L2TPv3 "Remote End ID" AVP defined in [\[RFC3931\]](#).

\* Generalized FEC 129 SAI/TAI

[Section 4.3 of \[RFC4667\]](#) defines how to encode the SAI and TAI parameters. These can be mapped directly.

Other interface parameter mappings are unsupported when switching between LDP/MPLS and L2TPv3 PWs.



### **8.6. Switching Point TLV in L2TPv3**

When translating between LDP and L2TPv3 control messages, the PW Switching Point TLV described earlier in this document is carried in a single variable length L2TP AVP present in the ICRQ, ICRP messages, and optionally in the ICCN message.

The L2TP "Switching Point AVP" is Attribute Type TBA-L2TP-AVP-1. The AVP MAY be hidden (the L2TP AVP H-bit may be 0 or 1), the length of the AVP is 6 plus the length of the series of Switching Point sub-TLVs included in the AVP, and the AVP MUST NOT be marked Mandatory (the L2TP AVP M-bit MUST be 0).

### **8.7. L2TPv3 and MPLS PW Data Plane**

When switching between an MPLS and L2TP PW, packets are sent in their entirety from one PW to the other, replacing the MPLS label stack with the L2TPv3 and IP header or vice versa.

[Section 5.4 of \[RFC3985\]](#) discusses the purpose of the various shim headers necessary for enabling a pseudowire over an IP or MPLS PSN. For L2TPv3, the Payload Convergence and Sequencing function is carried out via the Default L2-Specific Sublayer defined in [\[RFC3931\]](#). For MPLS, these two functions (together with PSN Convergence) are carried out via the MPLS Control Word. Since these functions are different between MPLS and L2TPv3, interworking between the two may be necessary.

The L2TP L2-Specific Sublayer and MPLS Control Word are shim headers which in some cases are not necessary to be present at all. For example, an Ethernet PW with sequencing disabled will generally not require an MPLS Control Word or L2TP Default L2-Specific Sublayer to be present at all. In this case, Ethernet frames are simply sent from one PW to the other without any modification beyond the MPLS and L2TP/IP encapsulation and decapsulation.

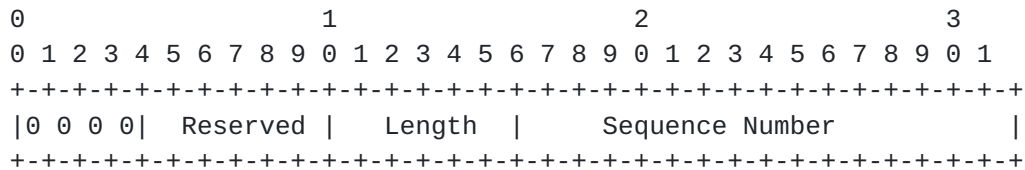
The following section offers guidelines for how to interwork between L2TP and MPLS for those cases where the Payload Convergence, Sequencing, or PSN Convergence functions are necessary on one or both sides of the switching node.





**8.7.1. Mapping the MPLS Control Word to L2TP**

The MPLS Control Word consists of (from left to right):



- i. These bits are always zero in an MPLS PW PDU. It is not necessary map them to L2TP.
- ii. These six bits may be used for Payload Convergence depending on the PW type. For ATM, the first four of these bits are defined in [[RFC4717](#)]. These map directly to the bits defined in [[RFC4454](#)]. For Frame Relay, these bits indicate how to set the bits in the Frame Relay header which must be regenerated for L2TP as it carries the Frame Relay header intact.
- iii. L2TP determines its payload length from IP. Thus, this Length field need not be carried directly to L2TP. This Length field will have to be calculated and inserted for MPLS when necessary.
- iv. The Default L2-Specific Sublayer has a sequence number with different semantics than that of the MPLS Control Word. This difference eliminates the possibility of supporting sequencing across the MS-PW by simply carrying the sequence number through the switching point transparently. As such, sequence numbers MAY be supported by checking the sequence numbers of packets arriving at the switching point and regenerating a new sequence number in the appropriate format for the PW on egress. If this type of sequence interworking at the switching node is not supported, and a T-PE requests sequencing of all packets via the L2TP control channel during session setup, the switching node SHOULD NOT allow the session to be established by sending a CDN message with Result Code set to 17 "sequencing not supported" (subject to IANA Assignment).



## **9. Operation And Management**

### **9.1. Extensions to VCCV to Support MS-PWs**

Single-segment pseudowires are signaled using the Virtual Circuit Connectivity Verification (VCCV) parameter included in the interface parameter field of the PW ID FEC TLV or the interface parameter sub-TLV of the Generalized PW ID FEC TLV as described in [[RFC5085](#)]. When a switching point exist between PE nodes, it is required to be able to continue operating VCCV end-to-end across a switching point and to provide the ability to trace the path of the MS-PW over any number of segments.

This document provides a method for achieving these two objectives. This method is based on re-using the existing VCCV CW and decrementing the TTL of the PW label at each S-PE in the path of the MS-PW.

### **9.2. MPLS PW to L2TPv3 PW OAM**

When an MS-PW includes SS-PWs that use the L2TPv3 the MPLS PW OAM MUST, be terminated at the S-PE connecting the L2TPv3 and MPLS segments. Status information received in a particular PW segment can then be used to generate the appropriate status messages on on the following PW segment. In the case of L2TPv3 the status bits in the circuit status AVP defined in [section 5.4.5 of \[RFC3931\]](#) and Extended Circuit Status Values defined in [[RFC5641](#)] can be directly be mapped to the PW status bits defined in [[RFC4447](#)] [section 5.4.3](#).

VCCV messages are specific to the MPLS data plane, and cannot be used for an L2TPv3 PW segment. Therefore, the S-PE MUST NOT send the VCCV parameter included in the interface parameter field of the PW ID FEC TLV or the sub-TLV interface parameter of the Generalized PW ID FEC TLV. It might be possible to translate VCCV messages from L2TPv3 PW segments to MPLS PW segments and vice versa, however this topic is left for further study.

### **9.3. MPLS PW to MPLS PW OAM Data Plane Indication**

As stated above the S-PE MUST perform a standard MPLS label swap operation on the MPLS PW label. By the rules defined in [[RFC3032](#)] the PW label TTL MUST be decreased at every S-PE. Once the PW label TTL reaches the value of 0, the packet is sent to the control plane to be processed. Hence, by controlling the PW TTL value of the PW label it is possible to select exactly which S-PE will respond to the VCCV



packet.

#### **9.4. Signaling OAM Capabilities for Switched Pseudowires**

Similarly to SS-PW, MS-PW VCCV capabilities are signaled using the VCCV parameter included in the interface parameter field of the PW ID FEC TLV or the sub-TLV interface parameter of the Generalized PW ID FEC TLV as described in [[RFC5085](#)].

In Figure 3 T-PE1 uses the VCCV parameter included in the interface parameter field of the PW ID FEC TLV or the sub-TLV interface parameter of the Generalized PW ID FEC TLV to indicate to the far end T-PE2 what VCCV capabilities T-PE1 supports. This is the same VCCV parameter as would be used if T-PE1 and T-PE2 were connected directly. S-PE2, which is a PW switching point, as part of the adaptation function for interface parameters, processes locally the VCCV parameter then passes it to T-PE2. If there were multiple S-PEs on the path between T-PE1 and T-PE2, each would carry out the same processing, passing along the VCCV parameter. The local processing of the VCCV parameter removes CC Types specified by the originating T-PE that are not supported on the S-PE. For example, if T-PE1 indicates supports CC Types 1,2,3 and the Then the S-PE removes the Router Alert CC Type=2, leaving the rest of the TLV unchanged, and passes the modified VCCV parameter to the next S-PE along the path.

The far end T-PE (T-PE2) receives the VCCV parameter indicating only the CC types that are supported by the initial T-PE (T-PE1) and all S-PEs along the PW path.

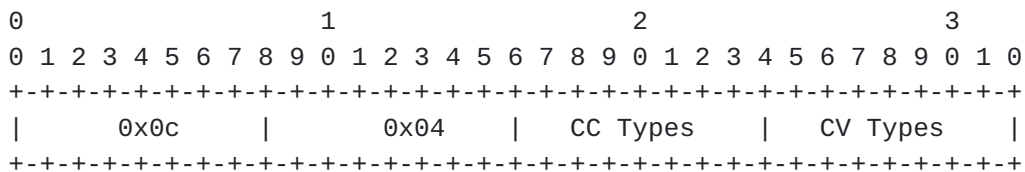
#### **9.5. OAM Capability for MS-Pws Demultiplexed using MPLS**

The VCCV parameter ID is defined as follows in [[RFC4446](#)]:

Parameter ID	Length	Description
0x0c	4	VCCV

The format of the VCCV parameter field is as follows:





- 0x01 Type 1: PWE3 control word with 0001b as first nibble as defined in [RFC4385].
- 0x02 Type 2: MPLS Router Alert Label.
- 0x04 Type 3: MPLS PW Demultiplexor Label TTL = 1 (Type 3).

**9.5.1. MS-PW and VCCV CC Type 1**

VCCV CC type 1 can be used for MS-PWs. However, if the CW is enabled on user packets, VCCV CC type 1 MUST be used according to the rules in [RFC5085]. When using CC type 1 for MS-PWs the PE transmitting the VCCV packet MUST set the TTL to the appropriate value to reach the destination S-PE. However if the packet is destined for the T-PE, the TTL can be set to any value that is sufficient for the packet to reach the T-PE.

**9.5.2. MS-PW and VCCV CC type 2**

VCCV CC type 2 is not supported for MS-PWs and MUST be removed from a VCCV parameter field by the S-PE.

**9.5.3. MS-PW and VCCV CC type 3**

VCCV CC type 3 can be used for MS-PWs, however if the CW is enabled VCCV type 1 is preferred according to the rules in [RFC5085]. Note that for using the VCCV type 3, TTL method, the PE will set the PW label TTL to the appropriate value necessary to reach the target PE, otherwise the VCCV packet might be forwarded over the AC to the CPE.

**9.6. MS-PW VCCV Operations**

This document specifies four VCCV operations:

- i. End-to-end MS-PW connectivity verification. This operation enables the connectivity of the MS-PW to be tested from source T-PE to destination T-PE. In order to do this, the sending T-PE must include the FEC used in the last segment of the MS-PW to the destination T-PE in the VCCV-Ping echo request. This information is either configured at the





- sending T-PE or is obtained by processing the corresponding sub-TLVs of the optional S-PE TLV, as described below.
- ii. Partial MS-PW connectivity verification. This operation enables the connectivity of any contiguous subset of the segments of a MS-PW to be tested from the source T-PE or a source S-PE to a destination S-PE or T-PE. Again, the FEC used on the last segment to be tested must be included in the VCCV-Ping echo request message. This information is determined by the sending T-PE or S-PE as in (i) above.
  - iii. MS-PW path verification. This operation verifies the path of the MS-PW, as returned by the S-PE TLV, against the actual data path of the MS-PW. The sending T-PE or S-PE iteratively sends a VCCV echo request to each S-PE along the MS-PW path, using the FEC for the corresponding MS-PW segment in the switching point TLV. If the S-PE TLV information is correct, then a VCCV echo reply showing that this is a valid router for the FEC will be received. However, if the switching point TLV information is incorrect, then this operation enables the first incorrect switching point to be determined, but not the actual path of the MS-PW beyond that. This operation cannot be used when the MS-PW is statically configured or when the S-PE TLV is not supported. The processing of the PW switching TLV used for this operation is described below. This operation is OPTIONAL.
  - iv. MS-PW path trace. This operation traces the data path of the MS-PW using FECs included in the Target FEC stack TLV [[RFC4379](#)] returned by S-PEs or T-PEs in an echo reply message. The sending T-PE or S-PE uses this information to recursively test each S-PE along the path of the MS-PW in a single operation in a similar manner to LSP trace. This operation is able to determine the actual data path of the MS-PW, and can be used for both statically configured and signaled MS-PWs. Support for this operation is OPTIONAL.

Note that the above operations rely on intermediate S-PEs and/or the destination T-PE to include the switching point TLV as a part of the MS-PW setup process, or to include the Target FEC stack TLV in the VCCV echo reply message. For various reasons, e.g. privacy or security of the S-PE/T-PE, this information may not be available to the source T-PE. In these cases, manual configuration of the FEC MAY still be used.



### **9.6.1. VCCV Echo Message Processing**

The challenge for the control plane is to be able to build the VCCV echo request packet with the necessary information to reach the desired S-PE or T-PE, for example the target FEC 128 PW sub-TLV of the downstream PW segment that the packet is destined for. This could be even more difficult in situations in which the MS-PW spans different providers and Autonomous Systems.

For example, in Figure 3, T-PE1 has the FEC128 of the segment, PW segment 1, but it does not readily have the information required to compose the FEC128 of the following segment, PW segment 3, if a VCCV echo request to be sent to T-PE2. This can be achieved by the methods described in the following subsections.

#### **9.6.1.1. Sending a VCCV Echo Request**

When performing a partial or end-to-end connectivity or path verification, the sender of the echo request message requires the FEC of the last segment to the target S-PE/T-PE node. This information can either be configured manually or be obtained by inspecting the corresponding sub-TLVs of the PW switching point TLV.

The necessary S-PE sub-TLVs are :

Type	Description
0x01	PW ID of last PW segment traversed
0x03	Local IP address of PW Switching Point
0x04	Remote IP address of last PW Switching Point traversed or of the T-PE

When performing an OPTIONAL MS-PW path trace operation, the T-PE will automatically learn the target FEC by probing, one by one, the S-PEs of the MS-PW path, using the FEC returned in the Target FEC stack of the previous VCCV echo reply.

#### **9.6.1.2. Receiving a VCCV Echo Request**

Upon receiving a VCCV echo request the control plane on S-PEs (or the target node of each segment of the MS-PW) validates the request and responds to the request with an echo reply consisting of a return code of 8 (label switched at stack-depth) indicating that it is an S-PE and not the egress router for the MS-PW.

S-PEs that wish to reveal their downstream next-hop in a trace



operation should include the FEC of the downstream PW segment in the Target FEC stack (as per Sections 3.2 and 4.5 of [RFC4379]) of the echo reply message. FEC128 PWs MUST use the format shown in [Section 3.2.09 of \[RFC4379\]](#) for the sub-TLV in the Target FEC stack, while FEC129 PWs MUST use the format shown in [Section 3.2.10 of \[RFC4379\]](#) for the sub-TLV in the Target FEC stack. Note that an S-PE MUST NOT include this FEC information in the reply if it has been configured not to do so for administrative reasons, or for reasons explained previously.

If the node is the T-PE or the egress node of the MS-PW, it responds to the echo request with an echo reply with a return code of 3 (egress router).

#### **[9.6.1.3. Receiving a VCCV Echo Reply](#)**

The operation to be taken by the node receiving the echo reply in response to an echo request depends on the VCCV mode of operation described above. See [Section 9.5.2](#) for detailed procedures.

#### **[9.6.2. Detailed VCCV procedures](#)**

There are two similar methods of verifying the MS-PW path: Path Trace, and Path Verification. Path Trace does not use the LDP control plane to obtain information on the path to verify, so this method is well suited if portions of the MS-PW are statically configured SS-PWs. The Path Verification method relies on information obtained from the LDP control plane, and hence offers better verification of the current forwarding behavior compared to the LDP signaled forwarding information of the MS-PW path. However in the case where there are statically signaled SS-PW in the MS-PW path, the path information is unavailable and must be programmed manually.

##### **[9.6.2.1. End to End Connectivity Verification Between T-PEs](#)**

In Figure 3, if T-PE1, S-PE and T-PE2 support Control Word, the PW control plane will automatically negotiate the use of the CW. VCCV CC type 3 will function correctly whether the CW is enable or not on the PW. However VCCV type 1 for (which can be use for end to end verification only), is only supported if the CW is enabled.

At the S-PE the data path operations include an outer label pop, inner label swap and new outer label push. Note that there is no requirement for the S-PE to inspect the CW. Thus, the end-to-end connectivity of the multi-segment pseudowire can be verified by



performing all of the following steps:

- i. T-PE forms a VCCV-ping echo request message with the FEC matching that of the last segment PW to the destination T-PE.
- ii. T-PE sets the inner PW label TTL to the exact value to allow the packet to reach the far end T-PE. ( the value is determined by counting the number of S-PEs from the control plane information ) Alternatively, if CC type 1 is supported the packet can be encapsulated according to CC type 1 in [\[RFC5085\]](#)
- iii. T-PE sends a VCCV packet that will follow the exact same data path at each S-PE as that taken by data packets.
- iv. S-PE may performs an outer label pop, if PHP is disabled, and will perform an inner label swap with TTL decrement, and new outer label push.
- v. There is no requirement for the S-PE to inspect the CW.
- vi. The VCCV packet is diverted to VCCV control processing at the destination T-PE.
- vii. Destination T-PE replies using the specified reply mode, i.e., reverse PW path or IP path.

#### **9.6.2.2. Partial Connectivity Verification from T-PE**

In order to trace part of the multi-segment pseudowire, the TTL of the PW label may be used to force the VCCV message to 'pop out' at an intermediate node. When the TTL expires, the S-PE can determine that the packet is a VCCV packet by either checking the control word (CW), or if the CW is not in use by checking for a valid IP header with UDP destination port 3503. The packet should then be diverted to VCCV processing.

In Figure 3, if T-PE1 sends a VCCV message with the TTL of the PW label equal to 1, the TTL will expire at the S-PE. T-PE1 can thus verify the first segment of the pseudowire.

The VCCV packet is built according to [\[RFC4379\] section 3.2.9](#) for FEC 128, or 3.2.10 for a FEC 129 PW. All the information necessary to build the VCCV LSP ping packet is collected by inspecting the S-PE TLVs.

Note that this use of the TTL is subject to the caution expressed in





[RFC5085]. If a penultimate LSR between S-PEs or between an S-PE and a T-PE manipulates the PW label TTL, the VCCV message may not emerge from the MS-PW at the correct S-PE.

#### **9.6.2.3. Partial connectivity verification between S-PEs**

Assuming that all nodes along an MS-PW support the Control Word CC Type 3, VCCV between S-PEs may be accomplished using the PW label TTL as described above. In Figure 3, the S-PE may verify the path between it and T-PE2 by sending a VCCV message with the PW label TTL set to 1. Given a more complex network with multiple S-PEs, an S-PE may verify the connectivity between it and an S-PE two segments away by sending a VCCV message with the PW label TTL set to 2. Thus, an S-PE can diagnose connectivity problems by successively increasing the TTL. All the information needed to build the proper VCCV echo request packet as described in [\[RFC4379\] section 3.2.9](#) or 3.2.10 is obtained automatically from the LDP label mapping that contains S-PE TLVs.

#### **9.6.2.4. MS-PW Path Verification**

As an example, in Figure 3, VCCV trace can be performed on the MS-PW originating from T-PE1 by a single operational command. The following process ensues:

- i. T-PE1 sends a VCCV echo request with TTL set to 1 and a FEC containing the pseudowire information of the first segment (PW1 between T-PE1 and S-PE) to S-PE for validation. If FEC Stack Validation is enabled, the request may also include additional sub-TLV such as LDP Prefix and/or RSVP LSP dependent on the type of transport tunnel the segmented PW is riding on.
- ii. S-PE validates the echo request with the FEC. Since it is a switching point between the first and second segment it builds an echo reply with a return code of 8 and sends the echo reply back to T-PE1.
- iii. T-PE1 builds a second VCCV echo request based on the information obtained from the control plane (S-PE TLV). It then increments the TTL and sends it out to T-PE2. Note that the VCCV echo request packet is switched at the S-PE data path and forwarded to the next downstream segment without any involvement from the control plane.



- iv. T-PE2 receives and validates the echo request with the FEC. Since T-PE2 is the destination node or the egress node of the MS-PW it replies to T-PE1 with an echo reply with a return code of 3 (Egress Router).
- v. T-PE1 receives the echo reply from T-PE2. T-PE1 is made aware that T-PE2 is the destination of the MS-PW because the echo reply has a return code of is 3. The trace process is completed.

If no echo reply is received, or an error code is received from a particular PE, the trace process MUST stop immediately, and packets MUST NOT be sent further along the MS-PW.

For more detail on the format of the VCCV echo packet, refer to [\[RFC5085\]](#) and [\[RFC4379\]](#). The TTL here refers to that of the inner (PW) label TTL.

#### **[9.6.2.5](#). MS-PW Path Trace**

As an example, in Figure 3, VCCV trace can be performed on the MS-PW originating from T-PE1 by a single operational command. The following OPTIONAL process ensues:

- i. T-PE1 sends a VCCV echo request with TTL set to 1 and a FEC containing the pseudowire information of the first segment (PW1 between T-PE1 and S-PE) to S-PE for validation. If FEC Stack Validation is enabled, the request may also include additional sub-TLV such as LDP Prefix and/or RSVP LSP dependent on the type of transport tunnel the segmented PW is riding on.
- ii. The S-PE validates the echo request with the FEC.
- iii. The S-PE builds an echo reply with a return code of 8 and sends the echo reply back to T-PE1, appending the FEC128 information for the next segment along the MS-PW to the VCCV echo reply packet using the Target FEC stack TLV (as per Sections [3.2](#) and [4.5](#) of [\[RFC4379\]](#)).
- iv. T-PE1 builds a second VCCV echo request based on the information obtained from the FEC stack TLV received in the previous VCCV echo reply. It then increments the TTL and sends it out to T-PE2. Note that the VCCV echo request packet is switched at the S-PE data path and forwarded to the next downstream segment without any involvement from the



control plane.

- v. T-PE2 receives and validates the echo request with the FEC. Since T-PE2 is the destination node or the egress node of the MS-PW it replies to T-PE1 with an echo reply with a return code of 3 (Egress Router).
- vi. T-PE1 receives the echo reply from T-PE2. T-PE1 is made aware that T-PE2 is the destination of the MS-PW because the echo reply has a return code of is 3. The trace process is completed.

If no echo reply is received, or an error code is received from a particular PE, the trace process MUST stop immediately, and packets MUST NOT be sent further along the MS-PW.

For more detail on the format of the VCCV echo packet, refer to [\[RFC5085\]](#) and [\[RFC4379\]](#). The TTL here refers to that of the inner (PW) label TTL.

## **10. Mapping Switched Pseudowire Status**

In the PW switching with attachment circuits case (Figure 2), PW status messages indicating PW or attachment circuit faults MUST be mapped to fault indications or OAM messages on the connecting AC as defined in [\[PW-MSG-MAP\]](#).

In the PW control plane switching case (Figure 3), there is no attachment circuit at the S-PE, but the two PWs are connected together. Similarly, the status of the PWs are forwarded unchanged from one PW to the other by the control plane switching function. However, it may sometimes be necessary to communicate fault status of one of the locally attached PW segments at a S-PE. For LDP this can be accomplished by sending an LDP notification message containing the PW status TLV, as well as an OPTIONAL PW switching point TLV as follows:









- ii. When the local status for any of the PW segments is at fault, the S-PE always sends the local status bits regardless if the received status bits from the remote node indicated that an upstream fault has cleared. AC status bit are passed along unchanged.

### 10.1. S-PE initiated PW status messages

The PW fault directions are defined as follows:

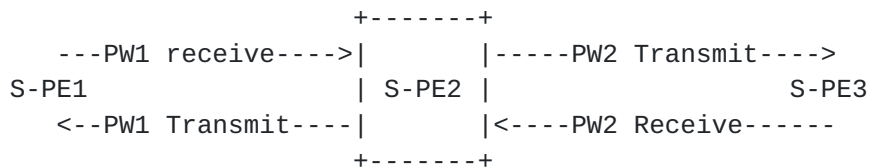


Figure 4. S-PE and PW tx/rx directions.

When a local fault is detected by the S-PE, a PW status message is sent in both directions along the PW. Since there are no attachment circuits on an S-PE, only the following status messages are relevant:

- 0x00000008 - Local PSN-facing PW (ingress) Receive Fault
- 0x00000010 - Local PSN-facing PW (egress) Transmit Fault

Each S-PE needs to store only two 32-bit PW status words for each PW segment: One for local failures, and one for remote failures (normally received from another PE). The first failure will set the appropriate bit in the 32-bit status word, and each subsequent failure will be ORed to the appropriate PW status word. In the case of the PW status word storing remote failures, this rule has the effect of a logical OR operation with the first failure received on the particular PW segment.

It should be noted that remote failures received on an S-PE are just passed along the MS-PW unchanged while local failures detected an S-PE are signalled on both PW segments.

A T-PE can receive multiple failures from S-PEs along the MS-PW, however only the failure from the remote closest S-PE will be stored (last pw status message received). The PW status word received are just ORed to any existing remote PW status already stored on the T-PE.

Given that there are two PW segments at a particular S-PE for a particular MS-PW, referring to figure 4, there are four possible



failure cases as follows:

- i. PW2 Transmit direction fault
- ii. PW1 Transmit direction fault
- iii. PW2 Receive direction fault
- iv. PW1 Receive direction fault

Once a PW status notification message is initiated at a S-PE for a particular PW status bit any further status message, for the same status bit, received from an upstream neighbor is processed locally and not forwarded until the S-PE original status error state is cleared.

Each S-PE along the MS-PW MUST store any PW status messages transiting it. If more than one status message with the same PW status bit set is received by a T-PE, or S-PE only the last PW status message is stored.

#### **10.1.1. Local PW2 transmit direction fault**

When this failure occurs the S-PE will take the following actions:

- \* Send a PW status message to S-PE3 containing "0x00000010 - Local PSN-facing PW (egress) Transmit Fault"
- \* Send a PW status message to S-PE1 containing "0x00000008 - Local PSN-facing PW (ingress) Receive Fault"
- \* Store 0x00000010 in the local PW status word for the PW segment toward S-PE3.

#### **10.1.2. Local PW1 transmit direction fault**

When this failure occurs the S-PE will take the following actions:

- \* Send a PW status message to S-PE1 containing "0x00000010 - Local PSN-facing PW (egress) Transmit Fault"
- \* Send a PW status message to S-PE3 containing "0x00000008 - Local PSN-facing PW (ingress) Receive Fault"
- \* Store 0x00000010 in the local PW status word for the PW segment toward S-PE1.



### **10.1.3. Local PW2 receive direction fault**

When this failure occurs the S-PE will take the following actions:

- \* Send a PW status message to S-PE3 containing "0x00000008 - Local PSN-facing PW (ingress) Receive Fault"
- \* Send a PW status message to S-PE1 containing "0x00000010 - Local PSN-facing PW (egress) Transmit Fault"
- \* Store 0x00000008 in the local PW status word for the PW segment toward S-PE3.

### **10.1.4. Local PW1 receive direction fault**

When this failure occurs the S-PE will take the following actions:

- \* Send a PW status message to S-PE1 containing "0x00000008 - Local PSN-facing PW (ingress) Receive Fault"
- \* Send a PW status message to S-PE3 containing "0x00000010 - Local PSN-facing PW (egress) Transmit Fault"
- \* Store 0x00000008 in the local PW status word for the PW segment toward S-PE1.

### **10.1.5. Clearing Faults**

Remote PW status fault clearing messages received by an S-PE will only be forwarded if there are no corresponding local faults on the S-PE. (local faults always supersede remote faults)

Once the local fault has cleared, and there is no corresponding (same PW status bit set) remote fault, a PW status messages is sent out to the adjacent PEs clearing the fault.

When a PW status fault clearing message is forwarded, the S-PE will always send the S-PE TLV associated with the PE which cleared the fault.

## **10.2. PW status messages and S-PE TLV processing**

When a PW status message is received that includes a S-PE TLV, the S-PE TLV information MAY be stored, along with the contents of the PW status Word according to the procedures described above. The S-PE TLV stored is always the S-PE TLV that is associated with the PE that set that particular last fault. If subsequent PW status message for the same PW status bit are received the S-PE TLV will overwrite the previously stored S-PE TLV.



### **10.3. T-PE processing of PW status messages**

The PW switching architecture is based on the concept that the T-PE should process the PW LDP messages in the same manner as if it was participating in the setup of a PW segment. However T-PE participating a MS-PW, SHOULD be able to process the S-PE TLV. Otherwise the processing of PW status messages, and other PW setup messages is exactly as described in [[RFC4447](#)].

### **10.4. Pseudowire Status Negotiation Procedures**

Pseudowire Status signaling methodology, defined in [[RFC4447](#)], SHOULD be transparent to the switching point.

### **10.5. Status Dampening**

When the PW control plane switching methodology is used to cross an administrative boundary it might be necessary to prevent excessive status signaling changes from being propagated across the administrative boundary. This can be achieved by using a similar method as commonly employed for the BGP protocol route advertisement dampening. The details of this OPTIONAL algorithm are a matter of implementation, and are outside the scope of this document.

## **11. Peering Between Autonomous Systems**

The procedures outlined in this document can be employed to provision and manage MS-PWs crossing AS boundaries. The use of more advanced mechanisms involving auto-discovery and ordered PWE3 MS-PW signaling will be covered in a separate document.

## **12. Congestion Considerations**

Each PSN carrying the PW may be subject to congestion. The Congestion considerations in [[RFC3985](#)] apply to PW segments as well. Each PW segment will handle any congestion experienced by the PW traffic independently of the other MS-PW segments. It is possible that passing knowledge of congestion between segments, and to the T-PEs can result in more efficient edge to edge congestion mitigation systems. However any specific methods of congestion mitigation are outside the scope of this document and left for further study.





## **13. Security Considerations**

This document specifies the LDP, L2TPv3, and VCCV extensions that are needed for setting up and maintaining pseudowires. The purpose of setting up pseudowires is to enable layer 2 frames to be encapsulated and transmitted from one end of a pseudowire to the other. Therefore we discuss the security considerations for both the data plane and the control plane in the following sections. The guidelines, and security considerations specified in [[RFC5920](#)], also apply to MS-PW when the PSN is MPLS.

### **13.1. Data Plane Security**

Data plane security consideration as discussed in [[RFC4447](#)], [[RFC3931](#)], and [[RFC3985](#)] apply to this extension without any changes.

#### **13.1.1. VCCV Security considerations**

The VCCV technology for MS-PW offers a method for the service provider to verify the data path of a specific PW. This involves sending a packet to a specific PE and receiving an answer which either confirms, or indicates that the information contained in the packet is incorrect. This is a very similar process to the commonly used IP ICMP ping, and TTL expired methods for IP networks. It should be noted that when using VCCV Type 3 for PW when the CW is not enabled, if a packet is crafted with a TTL greater than the number of hops along the MS-PW path, or an S-PE along the path mis-processes the TTL, the packet could mistakenly be forwarded out the attachment circuit as a native PW packet. This packet would most likely be treated as an error packet by the CE. However if this possibility is not acceptable, the CW should be enabled to guarantee that a VCCV packet will never be mistakenly forwarded to the AC.

### **13.2. Control Protocol Security**

General security considerations with regard to the use of LDP are specified in [section 5 of RFC 5036](#). Security considerations with regard to the L2TPv3 control plane are specified in [[RFC3931](#)]. These considerations apply as well to the case where LDP or L2TPv3 is used to set up PWs.

A Pseudowire connects two attachment circuits. It is important to make sure that LDP connections are not arbitrarily accepted from anywhere, or else a local attachment circuit might get connected to an arbitrary remote attachment circuit. Therefore an incoming session



request MUST NOT be accepted unless its IP source address is known to be the source of an "eligible" peer. The set of eligible peers could be pre-configured (either as a list of IP addresses, or as a list of address/mask combinations), or it could be discovered dynamically via an auto-discovery protocol which is itself trusted. (Note that if the auto-discovery protocol were not trusted, the set of "eligible peers" it produces could not be trusted.)

Even if a connection request appears to come from an eligible peer, its source address may have been spoofed. So some means of preventing source address spoofing must be in place. For example, if all the eligible peers are in the same network, source address filtering at the border routers of that network could eliminate the possibility of source address spoofing.

For a greater degree of security, the LDP authentication option, as described in [section 2.9 of \[RFC5036\]](#), or the Control Message Authentication option of [\[RFC3931\]](#) MAY be used. This provides integrity and authentication for the control messages, and eliminates the possibility of source address spoofing. Use of the message authentication option does not provide privacy, but privacy of control messages are not usually considered to be highly important. Both the LDP and L2TPv3 message authentication options rely on the configuration of pre-shared keys, making it difficult to deploy when the set of eligible neighbors is determined by an auto-configuration protocol.

The protocol described in this document relies on the LDP MD5 authentication key option, as described in [section 2.9 of \[RFC5036\]](#), to provide integrity and authentication for the LDP messages and protect against source address spoofing. This mechanism relies on the configuration of pre-shared keys, which typically introduces some fragility. In the specific case of MS-PW, the number of links that leave an organization will be limited in practice, so the reliance on pre-shared keys should be manageable.

When the Generalized ID FEC Element is used, it is possible that a particular peer may be one of the eligible peers, but may not be the right one to connect to the particular attachment circuit identified by the particular instance of the Generalized ID FEC element. However, given that the peer is known to be one of the eligible peers (as discussed above), this would be the result of a configuration error, rather than a security problem. Nevertheless, it may be advisable for a PE to associate each of its local attachment circuits with a set of eligible peers, rather than having just a single set of eligible peers associated with the PE as a whole.



## **14. IANA Considerations**

### **14.1. L2TPv3 AVP**

This document uses a new L2TP parameter, IANA already maintains a registry of name "Control Message Attribute Value Pair" defined by [[RFC3438](#)]. The following new value is required:

TBA-L2TP-AVP-1 - PW Switching Point AVP

### **14.2. LDP TLV TYPE**

This document uses a new LDP TLV types, IANA already maintains a registry of name "TLV TYPE NAME SPACE" defined by [RFC5036](#). The following value is suggested for assignment:

TLV type	Description
0x096D	Pseudowire Switching Point PE TLV

### **14.3. LDP Status Codes**

This document uses a new LDP status codes, IANA already maintains a registry of name "STATUS CODE NAME SPACE" defined by [RFC5036](#). The following value is suggested for assignment:

Assignment	E	Description
0x0000003A	0	"PW Loop Detected"

### **14.4. L2TPv3 Result Codes**

This document uses a new L2TPv3 Result Code for the CDN message, to be assigned IANA in the "Result Code AVP (Attribute Type 1) Values" registry as follows:

Registry Name: Result Code AVP (Attribute Type 1) Values Defined  
Result Code values for the CDN message are:

Assignment	Description
TBD	Sequencing not supported



#### **14.5. New IANA Registries**

IANA needs to set up a registry of "Pseudowire Switching Point PE sub-TLV Type". These are 8-bit values. Type value 1 through 6 are defined in this document. Type values 7 through 64 are to be assigned by IANA using the "Expert Review" policy defined in [RFC5226](#). Type values 65 through 127, 0 and 255 are to be allocated using the IETF consensus policy defined in [[RFC5226](#)]. Type values 128 through 254 are reserved for vendor proprietary extensions and are to be assigned by IANA, using the "First Come First Served" policy defined in [RFC5226](#).

The Type Values are assigned as follows:

Type	Length	Description
0x01	4	PW ID of last PW segment traversed
0x02	variable	PW Switching Point description string
0x03	4/16	Local IP address of PW Switching Point
0x04	4/16	Remote IP address of last PW Switching Point traversed or of the T-PE
0x05	variable	FEC Element of last PW segment traversed
0x06	12	L2 PW address of PW Switching Point

#### **15. Normative References**

- [RFC4385] " Pseudowire Emulation Edge-to-Edge (PWE3) Control Word for Use over an MPLS PSN", S. Bryant, et al., [RFC4385](#), February 2006.
- [RFC4446] "IANA Allocations for Pseudowire Edge to Edge Emulation (PWE3)", L. Martini, [RFC4446](#), April 2006.
- [RFC4447] "Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP)", Martini, L., et al., [rfc4447](#) April 2006.
- [RFC4364] "BGP/MPLS IP Virtual Private Networks (VPNs)", Rosen, E, Rekhter, Y., [RFC4364](#), February 2006 October 2004.
- [RFC3931] "Layer Two Tunneling Protocol (Version 3)", J. Lau, M. Townsley, I. Goyret, [RFC3931](#)
- [RFC5085] Nadeau, T., et al. "Pseudo Wire Virtual Circuit Connection Verification (VCCV), A Control Channel for Pseudowires", [RFC5085](#) December 2007.





[RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations section in RFCs", [BCP 26](#), [RFC 5226](#), May 2008

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

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

[RFC4379] K. Kompella, G. Swallow, "Detecting Multi-Protocol Label Switched (MPLS) Data Plane Failures", [RFC4379](#), September 2007

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

[RFC2277] H. Alvestrand, "IETF Policy on Character Sets and Languages", [RFC 2277](#), January 1998

[RFC5641] N. McGill, C. Pignataro, "Layer 2 Tunneling Protocol Version 3 (L2TPv3) Extended Circuit Status Values", August 2009

## **16. Informative References**

[RFC4023] "Encapsulating MPLS in IP or Generic Routing Encapsulation (GRE)", Rosen, E, Rekhter, Y. [RFC4023](#), March 2005.

[RFC3985] "Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture" Bryant, et al., [RFC 3985](#), March 2005.

[RFC4623] "Pseudowire Emulation Edge-to-Edge (PWE3) Fragmentation and Reassembly", A. Malis, W. M. Townsley, [RFC 4623](#), August 2006

[RFC4667] "Layer 2 Virtual Private Network (L2VPN) Extensions for Layer 2 Tunneling Protocol (L2TP)", Luo, Wei, [RFC4667](#), W. Luo, September 2006

[RFC4454] "Asynchronous Transfer Mode (ATM) over Layer 2 Tunneling Protocol Version 3 (L2TPv3)", Singh, Townsley, Pignataro, [RFC4454](#), May 2006  
( work in progress ), March 2004.

[RFC4717] "Encapsulation Methods for Transport of (ATM) MPLS Networks", Martini et al., [RFC4717](#), December 2006



- [RFC3438] W. M. Townsley, "Layer Two Tunneling Protocol (L2TP) Internet Assigned Numbers Authority (IANA) Considerations Update", December 2002, [RFC3438](#)
- [PW-MSG-MAP] "Pseudo Wire (PW) OAM Message Mapping", Nadeau et al, [draft-ietf-pwe3-oam-msg-map-12.txt](#), (work in progress), March 2010
- [RFC3032] "MPLS Label Stack Encoding", [RFC3032](#), January 2001
- [RFC5659] "An Architecture for Multi-Segment Pseudo Wire Emulation Edge-to-Edge", Bocci et al, [RFC5659](#), October 2009.
- [RFC5254] "Requirements for Multi-Segment Pseudowire Emulation Edge-to-Edge (PWE3)", N. Bitar, M. Bocci, L. Martini, [RFC5254](#), October 2008
- [RFC5920] Luyuan Fang, Ed., "Security Framework for MPLS and GMPLS Networks", [RFC5920](#), July 2010

## **17. Author's Addresses**

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

Thomas D. Nadeau  
e-mail: [tnadeau@lucidvision.com](mailto:tnadeau@lucidvision.com)

Chris Metz  
Cisco Systems, Inc.  
e-mail: [chmetz@cisco.com](mailto:chmetz@cisco.com)

Matthew Bocci  
Alcatel-Lucent  
Grove House, Waltham Road Rd  
White Waltham, Berks, UK. SL6 3TN  
e-mail: [matthew.bocci@alcatel-lucent.co.uk](mailto:matthew.bocci@alcatel-lucent.co.uk)



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

## **18. Additional Authors**

The following people also contributed text to this document:

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

Mike Duckett  
Bellsouth  
Lindbergh Center  
D481  
575 Morosgo Dr  
Atlanta, GA 30324  
e-mail: mduckett@bellsouth.net

## Copyright 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.

This document may contain material from IETF Documents or IETF



Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.

#### Acknowledgments

The authors wish to acknowledge the contributions of Satoru Matsushima, Wei Luo, Neil McGill, Skip Booth, Neil Hart, Michael Hua, and Tiberiu Grigoriu.

Expiration Date: March 2011

