Network Working Group
Request for Comments: 5085
Category: Standards Track

T. Nadeau, Ed.
C. Pignataro, Ed.
Cisco Systems, Inc.

December 2007

Pseudowire Virtual Circuit Connectivity Verification (VCCV): A Control Channel for Pseudowires

Status of This Memo

Abstract

This document describes Virtual Circuit Connectivity Verification (VCCV), which provides a control channel that is associated with a pseudowire (PW), as well as the corresponding operations and management functions (such as connectivity verification) to be used over that control channel. VCCV applies to all supported access circuit and transport types currently defined for PWs.

Table of Contents

1. Introduction

There is a need for fault detection and diagnostic mechanisms that can be used for end-to-end fault detection and diagnostics for a Pseudowire, as a means of determining the PW's true operational state. Operators have indicated in [RFC4377] and [RFC3916] that such a tool is required for PW operation and maintenance. This document defines a protocol called Virtual Circuit Connectivity Verification (VCCV) that satisfies these requirements. VCCV is, in its simplest description, a control channel between a pseudowire's ingress and egress points over which connectivity verification messages can be sent.

The Pseudowire Edge-to-Edge Emulation (PWE3) Working Group defines a mechanism that emulates the essential attributes of a telecommunications service (such as a T1 leased line or Frame Relay) over a variety of Packet Switched Network (PSN) types [RFC3985]. PWE3 is intended to provide only the minimum necessary functionality to emulate the service with the required degree of faithfulness for the given service definition. The required functions of PWs include encapsulating service-specific bit streams, cells, or PDUs arriving at an ingress port and carrying them across an IP path or MPLS tunnel. In some cases, it is necessary to perform other operations, such as managing their timing and order, to emulate the behavior and characteristics of the service to the required degree of faithfulness.

From the perspective of Customer Edge (CE) devices, the PW is characterized as an unshared link or circuit of the chosen service. In some cases, there may be deficiencies in the PW emulation that impact the traffic carried over a PW and therefore limit the applicability of this technology. These limitations must be fully described in the appropriate service-specific documentation.

For each service type, there will be one default mode of operation that all PEs offering that service type must support. However, optional modes have been defined to improve the faithfulness of the emulated service, as well as to offer a means by which older implementations may support these services.

Figure 1 depicts the architecture of a pseudowire as defined in [RFC3985]. It further depicts where the VCCV control channel resides within this architecture, which will be discussed in detail shortly.

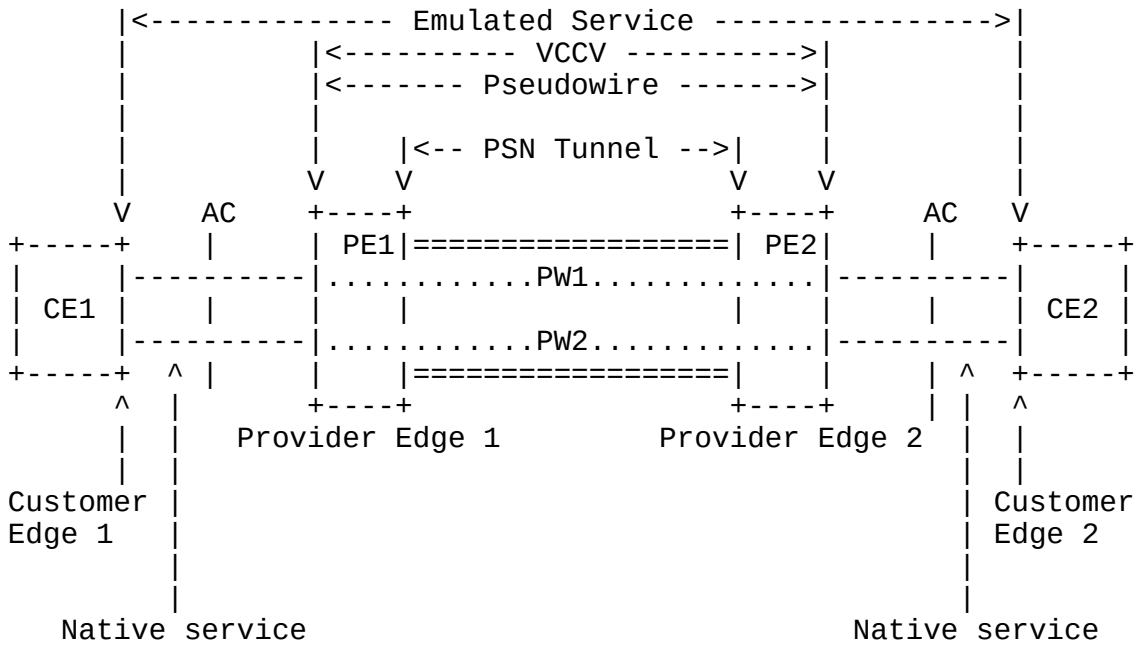


Figure 1: PWE3 VCCV Operation Reference Model

From Figure 1, Customer Edge (CE) routers CE1 and CE2 are attached to the emulated service via Attachment Circuits (ACs), and to each of the Provider Edge (PE) routers (PE1 and PE2, respectively). An AC can be a Frame Relay Data Link Connection Identifier (DLCI), an ATM Virtual Path Identifier / Virtual Channel Identifier (VPI/VCI), an Ethernet port, etc. The PE devices provide pseudowire emulation, enabling the CEs to communicate over the PSN. A pseudowire exists between these PEs traversing the provider network. VCCV provides several means of creating a control channel over the PW, between the PE routers that attach the PW.

Figure 2 depicts how the VCCV control channel is associated with the pseudowire protocol stack.

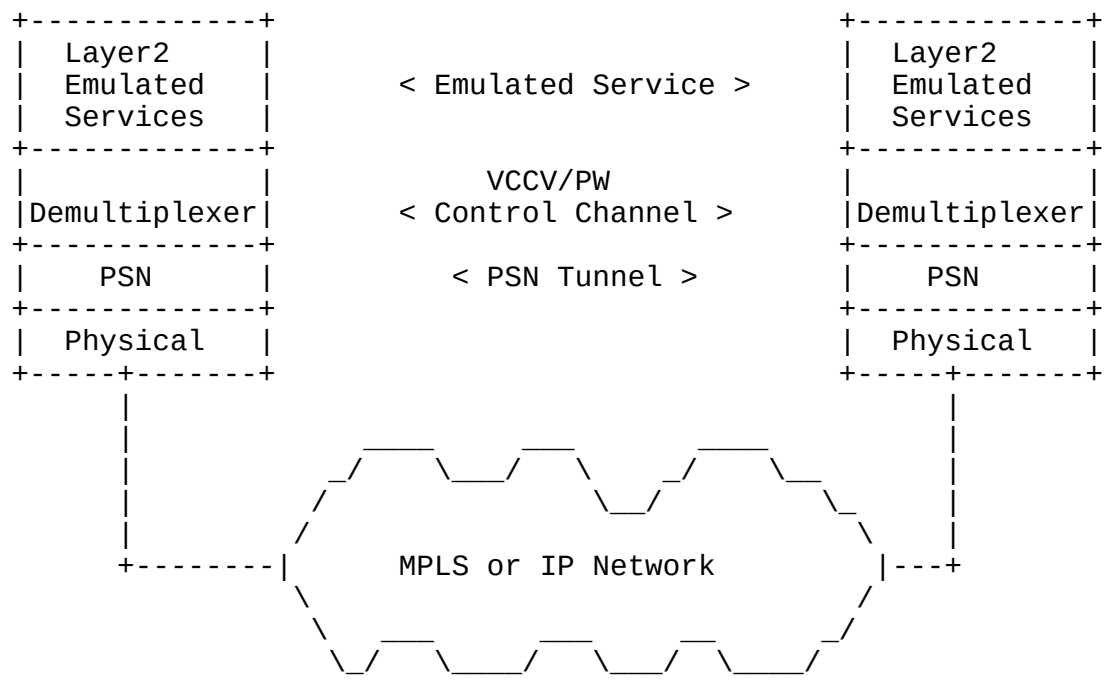


Figure 2: PWE3 Protocol Stack Reference Model including the VCCV Control Channel

VCCV messages are encapsulated using the PWE3 encapsulation as described in Sections 5 and 6, so that they are handled and processed in the same manner (or in some cases, a similar manner) as the PW PDUs for which they provide a control channel. These VCCV messages are exchanged only after the capability (expressed as two VCCV type spaces, namely the VCCV Control Channel and Connectivity Verification Types) and desire to exchange such traffic has been advertised between the PEs (see Sections 5.3 and 6.3), and VCCV types chosen.

1.1. 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].

2. Abbreviations

AC Attachment Circuit [RFC3985].

AVP Attribute Value Pair [RFC3931].

CC Control Channel (used as CC Type).

CE Customer Edge.

CV Connectivity Verification (used as CV Type).

CW Control Word [RFC3985].

L2SS L2-Specific Sublayer [RFC3931].

LCCE L2TP Control Connection Endpoint [RFC3931].

OAM Operation and Maintenance.

PE Provider Edge.

PSN Packet Switched Network [RFC3985].

PW Pseudowire [RFC3985].

PW-ACH PW Associated Channel Header [RFC4385].

VCCV Virtual Circuit Connectivity Verification.

3. Overview of VCCV

The goal of VCCV is to verify and further diagnose the pseudowire forwarding path. To this end, VCCV is comprised of different components:

- a means of signaling VCCV capabilities to a peer PE,
- an encapsulation for the VCCV control channel messages that allows the receiving PE to intercept, interpret, and process them locally as OAM messages, and
- specifications for the operation of the various VCCV operational modes transmitted within the VCCV messages.

When a pseudowire is first signaled using the Label Distribution Protocol (LDP) [RFC4447] or the Layer Two Tunneling Protocol version 3 (L2TPv3) [RFC3931], a message is sent from the initiating PE to the receiving PE requesting that a pseudowire be set up. This message has been extended to include VCCV capability information (see Section 4). The VCCV capability information indicates to the receiving PE which combinations of Control Channel (CC) and Connectivity Verification (CV) Types it is capable of receiving. If the receiving PE agrees to establish the PW, it will return its capabilities in the subsequent signaling message to indicate which CC

and CV Types it is capable of processing. Precedence rules for which CC and CV Type to choose in cases where more than one is specified in this message are defined in Section 7 of this document.

Once the PW is signaled, data for the PW will flow between the PEs terminating the PW. At this time, the PEs can begin transmitting VCCV messages based on the CC and CV Type combinations just discussed. To this end, VCCV defines an encapsulation for these messages that identifies them as belonging to the control channel for the PW. This encapsulation is designed to both allow the control channel to be processed functionally in the same manner as the data traffic for the PW in order to faithfully test the data plane for the PE, and allow the PE to intercept and process these VCCV messages instead of forwarding them out of the AC towards the CE as if they were data traffic. In this way, the most basic function of the VCCV control channel is to verify connectivity of the pseudowire and the data plane used to transport the data path for the pseudowire. It should be noted that because of the number of combinations of optional and mandatory data-plane encapsulations for PW data traffic, VCCV defines a number of Control Channel (CC) and Connectivity Verification (CV) types in order to support as many of these as possible. While designed to support most of the existing combinations (both mandatory and optional), VCCV does define a default CC and CV Type combination for each PW Demultiplexer type, as will be described in detail later in this document.

VCCV can be used both as a fault detection and/or a diagnostic tool for pseudowires. For example, an operator can periodically invoke VCCV on a timed, on-going basis for proactive connectivity verification on an active pseudowire, or on an ad hoc or as-needed basis as a means of manual connectivity verification. When invoking VCCV, the operator triggers a combination of one of its various CC Types and one of its various CV Types. The CV Types include LSP Ping [RFC4379] for MPLS PWs, and ICMP Ping [RFC0792] [RFC4443] for both MPLS and L2TPv3 PWs. We define a matrix of acceptable CC and CV Type combinations further in this specification.

The control channel maintained by VCCV can additionally carry fault detection status between the endpoints of the pseudowire. Furthermore, this information can then be translated into the native OAM status codes used by the native access technologies, such as ATM, Frame-Relay or Ethernet. The specific details of such status interworking is out of the scope of this document, and is only noted here to illustrate the utility of VCCV for such purposes. Complete details can be found in [MSG-MAP] and [RFC4447].

4. CC Types and CV Types

The VCCV Control Channel (CC) Type defines several possible types of control channel that VCCV can support. These control channels can in turn carry several types of protocols defined by the Connectivity Verification (CV) Type. VCCV potentially supports multiple CV Types concurrently, but it only supports the use of a single CC Type. The specific type or types of VCCV packets that can be accepted and sent by a router are indicated during capability advertisement as described in Sections 5.3 and 6.3. The various VCCV CV Types supported are used only when they apply to the context of the PW demultiplexer in use. For example, the LSP Ping CV Type should only be used when MPLS Labels are utilized as PW Demultiplexer.

Once a set of VCCV capabilities is received and advertised, a CC Type and CV Type(s) that match both the received and transmitted capabilities can be selected. That is, a PE router needs to only allow Types that are both received and advertised to be selected, performing a logical AND between the received and transmitted bitflag fields. The specific CC Type and CV Type(s) are then chosen within the constraints and rules specified in Section 7. Once a specific CC Type has been chosen (i.e., it matches both the transmitted and received VCCV CC capability), transmitted and replied to, this CC Type MUST be the only one used until such time as the pseudowire is re-signaled. In addition, based on these rules and the procedures defined in Section 5.2 of [RFC4447], the pseudowire MUST be re-signaled if a different set of capabilities types is desired. The relevant portion of Section 5.2 of [RFC4447] is:

> Interface Parameter Sub-TLV
>
> Note that as the "interface parameter sub-TLV" is part of the FEC, the rules of LDP make it impossible to change the interface parameters once the pseudowire has been set up.

The CC and CV Type indicator fields are defined as 8-bit bitmasks used to indicate the specific CC or CV Type or Types (i.e., none, one, or more) of control channel packets that may be sent on the VCCV control channel. These values represent the numerical value corresponding to the actual bit being set in the bitfield. The definition of each CC and CV Type is dependent on the PW type context, either MPLS or L2TPv3, within which it is defined.

Control Channel (CC) Types:

The defined values for CC Types for MPLS PWs are:

MPLS Control Channel (CC) Types:

Bit (Value)	Description
Bit 0 (0x01)	Type 1: PWE3 Control Word with 0001b as first nibble (PW-ACH, see [RFC4385])
Bit 1 (0x02)	Type 2: MPLS Router Alert Label
Bit 2 (0x04)	Type 3: MPLS PW Label with TTL == 1
Bit 3 (0x08)	Reserved
Bit 4 (0x10)	Reserved
Bit 5 (0x20)	Reserved
Bit 6 (0x40)	Reserved
Bit 7 (0x80)	Reserved

The defined values for CC Types for L2TPv3 PWs are:

L2TPv3 Control Channel (CC) Types:

Bit (Value)	Description
Bit 0 (0x01)	L2-Specific Sublayer with V-bit set
Bit 1 (0x02)	Reserved
Bit 2 (0x04)	Reserved
Bit 3 (0x08)	Reserved
Bit 4 (0x10)	Reserved
Bit 5 (0x20)	Reserved
Bit 6 (0x40)	Reserved
Bit 7 (0x80)	Reserved

Connectivity Verification (CV) Types:

The defined values for CV Types for MPLS PWs are:

MPLS Connectivity Verification (CV) Types:

Bit (Value)	Description
Bit 0 (0x01)	ICMP Ping
Bit 1 (0x02)	LSP Ping
Bit 2 (0x04)	Reserved
Bit 3 (0x08)	Reserved
Bit 4 (0x10)	Reserved
Bit 5 (0x20)	Reserved

```
Bit 6 (0x40) - Reserved
Bit 7 (0x80) - Reserved
```

The defined values for CV Types for L2TPv3 PWs are:

L2TPv3 Connectivity Verification (CV) Types:

Bit (Value)	Description
Bit 0 (0x01)	ICMP Ping
Bit 1 (0x02)	Reserved
Bit 2 (0x04)	Reserved
Bit 3 (0x08)	Reserved
Bit 4 (0x10)	Reserved
Bit 5 (0x20)	Reserved
Bit 6 (0x40)	Reserved
Bit 7 (0x80)	Reserved

If none of the types above are supported, the entire CC and CV Type Indicator fields SHOULD be transmitted as 0x00 (i.e., all bits in the bitfield set to 0) to indicate this to the peer.

If no capability is signaled, then the peer MUST assume that the peer has no VCCV capability and follow the procedures specified in this document for this case.

5. VCCV Control Channel for MPLS PWs

When MPLS is used to transport PW packets, VCCV packets are carried over the MPLS LSP as defined in this section. In order to apply IP monitoring tools to a PW, an operator may configure VCCV as a control channel for the PW between the PE's endpoints [RFC3985]. Packets sent across this channel from the source PE towards the destination PE either as in-band traffic with the PW's data, or out-of-band. In all cases, the control channel traffic is not forwarded past the PE endpoints towards the Customer Edge (CE) devices; instead, VCCV messages are intercepted at the PE endpoints for exception processing.

5.1. VCCV Control Channel Types for MPLS

As already described in Section 4, the capability of which control channel types (CC Type) are supported is advertised by a PE. Once the receiving PE has chosen a CC Type mode to use, it MUST continue using this mode until such time as the PW is re-signaled. Thus, if a new CC Type is desired, the PW must be torn-down and re-established.

Ideally, such a control channel would be completely in-band (i.e., following the same data-plane faith as PW data). When a control word is present on the PW, it is possible to indicate the control channel by setting a bit in the control word header (see Section 5.1.1).

Section 5.1.1 through Section 5.1.3 describe each of the currently defined VCCV Control Channel Types (CC Types).

5.1.1. In-Band VCCV (Type 1)

CC Type 1 is also referred to as "PWE3 Control Word with 0001b as first nibble". It uses the PW Associated Channel Header (PW-ACH); see Section 5 of [RFC4385].

The PW set-up protocol [RFC4447] determines whether a PW uses a control word. When a control word is used, and that CW uses the "Generic PW MPLS Control Word" format (see Section 3 of [RFC4385]), a Control Channel for use of VCCV messages can be created by using the PW Associated Channel CW format (see Section 5 of [RFC4385]).

The PW Associated Channel for VCCV control channel traffic is defined in [RFC4385] as shown in Figure 3:

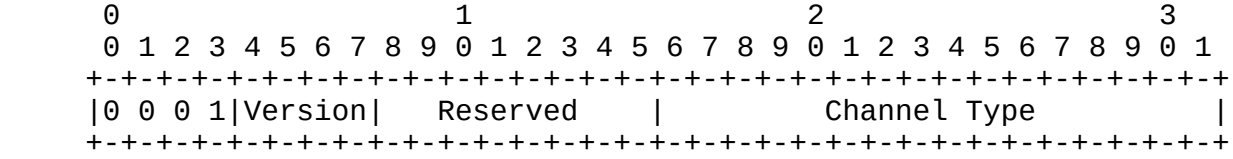


Figure 3: PW Associated Channel Header

The first nibble is set to 0001b to indicate a channel associated with a pseudowire (see Section 5 of [RFC4385] and Section 3.6 of [RFC4446]). The Version and the Reserved fields are set to 0, and the Channel Type is set to 0x0021 for IPv4 and 0x0057 for IPv6 payloads.

For example, Figure 4 shows how the Ethernet [RFC4448] PW-ACH would be received containing an LSP Ping payload corresponding to a choice of CC Type of 0x01 and a CV Type of 0x02:

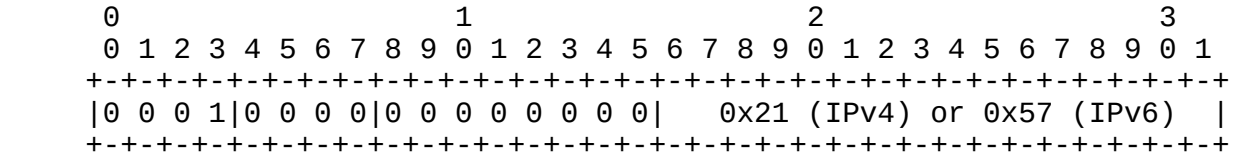


Figure 4: PW Associated Channel Header for VCCV

It should be noted that although some PW types are not required to carry the control word, this type of VCCV can only be used for those PW types that do employ the control word when it is in use. Further, this CC Type can only be used if the PW CW follows the "Generic PW MPLS Control Word" format. This mode of VCCV operation MUST be supported when the control word is present.

5.1.2. Out-of-Band VCCV (Type 2)

CC Type 2 is also referred to as "MPLS Router Alert Label".

A VCCV control channel can alternatively be created by using the MPLS router alert label [RFC3032] immediately above the PW label. It should be noted that this approach could result in a different Equal Cost Multi-Path (ECMP) hashing behavior than pseudowire PDUs, and thus result in the VCCV control channel traffic taking a path which differs from that of the actual data traffic under test. Please see Section 2 of [RFC4928].

CC Type 2 can be used whether the PW is set-up with a Control Word present or not.

This is the preferred mode of VCCV operation when the Control Word is not present.

If the Control Word is in use on this PW, it MUST also be included before the VCCV message. This is done to avoid the different ECMP hashing behavior. In this case, the CW uses the PW-ACH format described in Section 5.1.1 (see Figures 3 and 4). If the Control Word is not in use on this PW, the VCCV message follows the PW Label directly.

5.1.3. TTL Expiry VCCV (Type 3)

CC Type 3 is also referred to as "MPLS PW Label with TTL == 1".

The TTL of the PW label can be set to 1 to force the packet to be processed within the destination router’s control plane. This approach could also result in a different ECMP hashing behavior and VCCV messages taking a different path than the PW data traffic.

CC Type 3 can be used whether the PW is set-up with a Control Word present or not.

If the Control Word is in use on this PW, it MUST also be included before the VCCV message. This is done to avoid the different ECMP hashing behavior. In this case, the CW uses the PW-ACH format

described in Section 5.1.1 (see Figures 3 and 4). If the Control Word is not in use on this PW, the VCCV message follows the PW Label directly.

5.2. VCCV Connectivity Verification Types for MPLS

5.2.1. ICMP Ping

When this optional connectivity verification mode is used, an ICMP Echo packet using the encoding specified in [RFC0792] (ICMPv4) or [RFC4443] (ICMPv6) achieves connectivity verification. Implementations MUST use ICMPv4 [RFC0792] if the signaling for VCCV used IPv4 addresses, or ICMPv6 [RFC4443] if IPv6 addresses were used. If the pseudowire is set up statically, then the encoding MUST use that which was used for the pseudowire in the configuration.

5.2.2. MPLS LSP Ping

The LSP Ping header MUST be used in accordance with [RFC4379] and MUST also contain the target FEC Stack containing the sub-TLV of sub-Type 8 for the "L2 VPN endpoint", 9 for "FEC 128 Pseudowire (deprecated)", 10 for "FEC 128 Pseudowire", or 11 for the "FEC 129 Pseudowire". The sub-TLV value indicates the PW to be verified.

5.3. VCCV Capability Advertisement for MPLS PWs

To permit the indication of the type or types of PW control channel(s) and connectivity verification mode or modes over a particular PW, a VCCV parameter is defined in Section 5.3.1 that is used as part of the PW establishment signaling. When a PE signals a PW and desires PW OAM for that PW, it MUST indicate this during PW establishment using the messages defined in Section 5.3.1. Specifically, the PE MUST include the VCCV interface parameter sub-TLV (0x0C) assigned in [RFC4446] in the PW set-up message [RFC4447].

The decision of the type of VCCV control channel is left completely to the receiving control entity, although the set of choices is given by the sender in that it indicates the control channels and connectivity verification type or types that it can understand. The receiver SHOULD choose a single Control Channel Type from the match between the choices sent and received, based on the capability advertisement selection specified in Section 7, and it MUST continue to use this type for the duration of the life of the control channel. Changing Control Channel Types after one has been established to be in use could potentially cause problems at the receiving end and could also lead to interoperability issues; thus, it is NOT RECOMMENDED.

When a PE sends a label mapping message for a PW, it uses the VCCV parameter to indicate the type of OAM control channels and connectivity verification type or types it is willing to receive and can send on that PW. A remote PE MUST NOT send VCCV messages before the capability of supporting the control channel(s) (and connectivity verification type(s) to be used over them) is signaled. Then, it can do so only on a control channel and using the connectivity verification type(s) from the ones indicated.

If a PE receives VCCV messages prior to advertising capability for this message, it MUST discard these messages and not reply to them. In this case, the PE SHOULD increment an error counter and optionally issue a system and/or SNMP notification to indicate to the system administrator that this condition exists.

When LDP is used as the PW signaling protocol, the requesting PE indicates its configured VCCV capability or capabilities to the remote PE by including the VCCV parameter with appropriate options in the VCCV interface parameter sub-TLV field of the PW ID FEC TLV (FEC 128) or in the interface parameter sub-TLV of the Generalized PW ID FEC TLV (FEC 129). These options indicate which control channel and connectivity verification types it supports. The requesting PE MAY indicate that it supports multiple control channel options, and in doing so, it agrees to support any and all indicated types if transmitted to it. However, it MUST do so in accordance with the rules stipulated in Section 5.3.1 (VCCV Capability Advertisement Sub-TLV.)

Local policy may direct the PE to support certain OAM capability and to indicate it. The absence of the VCCV parameter indicates that no OAM functions are supported by the requesting PE, and thus the receiving PE MUST NOT send any VCCV control channel traffic to it. The reception of a VCCV parameter with no options set MUST be ignored as if one is not transmitted at all.

The receiving PE similarly indicates its supported control channel types in the label mapping message. These may or may not be the same as the ones that were sent to it. The sender should examine the set that is returned to understand which control channels it may establish with the remote peer, as specified in Sections 4 and 7. Similarly, it MUST NOT send control channel traffic to the remote PE for which the remote PE has not indicated it supports.

5.3.1. VCCV Capability Advertisement LDP Sub-TLV

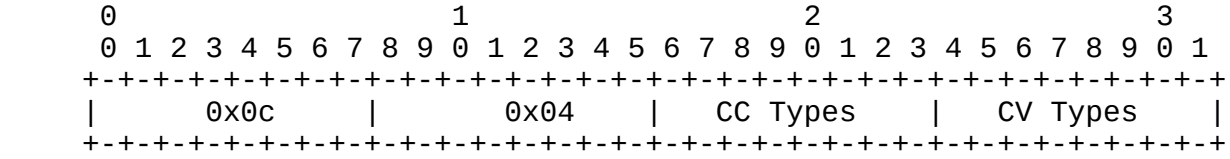
[RFC4447] defines an Interface Parameter Sub-TLV field in the LDP PW ID FEC (FEC 128) and an Interface Parameters TLV in the LDP Generalized PW ID FEC (FEC 129) to signal different capabilities for

specific PWs. An optional sub-TLV parameter is defined to indicate the capability of supporting none, one, or more control channel and connectivity verification types for VCCV. This is the VCCV parameter field. If FEC 128 is used, the VCCV parameter field is carried in the Interface Parameter sub-TLV field. If FEC 129 is used, it is carried as an Interface Parameter sub-TLV in the Interface Parameters TLV.

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:



The Control Channel Type field (CC Type) defines a bitmask used to indicate the type of control channel(s) (i.e., none, one, or more) that a router is capable of receiving control channel traffic on. If more than one control channel is specified, the router agrees to accept control traffic over either control channel; however, see the rules specified in Sections 4 and 7 for more details. If none of the types are supported, a CC Type Indicator of 0x00 SHOULD be transmitted to indicate this to the peer. However, if no capability is signaled, then the PE MUST assume that its peer is incapable of receiving any of the VCCV CC Types and MUST NOT send any OAM control channel traffic to it. Note that the CC and CV Types definitions are consistent regardless of the PW's transport or access circuit type. The CC and CV Type values are defined in Section 4.

6. VCCV Control Channel for L2TPv3/IP PWs

When L2TPv3 is used to set up a PW over an IP PSN, VCCV packets are carried over the L2TPv3 session as defined in this section. L2TPv3 provides a "Hello" keepalive mechanism for the L2TPv3 control plane that operates in-band over IP or UDP (see Section 4.4 of [RFC3931]). This built-in Hello facility provides dead peer and path detection only for the group of sessions associated with the L2TP Control Connection. VCCV, however, allows individual L2TP sessions to be tested. This provides a more granular mechanism which can be used to troubleshoot potential problems within the data plane of L2TP endpoints themselves, or to provide additional connection status of individual pseudowires.

The capability of which Control Channel Type (CC Type) to use is advertised by a PE to indicate which of the potentially various control channel types are supported. Once the receiving PE has chosen a mode to use, it MUST continue using this mode until such time as the PW is re-signaled. Thus, if a new CC Type is desired, the PW must be torn down and re-established.

An LCCE sends VCCV messages on an L2TPv3-signaled pseudowire for fault detection and diagnostic of the L2TPv3 session. The VCCV message travels in-band with the Session and follows the exact same path as the user data for the session, because the IP header and L2TPv3 Session header are identical. The egress LCCE of the L2TPv3 session intercepts and processes the VCCV message, and verifies the signaling and forwarding state of the pseudowire on reception of the VCCV message. It is to be noted that the VCCV mechanism for L2TPv3 is primarily targeted at verifying the pseudowire forwarding and signaling state at the egress LCCE. It also helps when L2TPv3 Control Connection and Session paths are not identical.

6.1. VCCV Control Channel Type for L2TPv3

In order to carry VCCV messages within an L2TPv3 session data packet, the PW MUST be established such that an L2-Specific Sublayer (L2SS) that defines the V-bit is present. This document defines the V-bit for the Default L2-Specific Sublayer [RFC3931] and the ATM-Specific Sublayer [RFC4454] using the Bit 0 position (see Sections 8.3.2 and 8.3.3). The L2-Specific Sublayer presence and type (either the Default or a PW-Specific L2SS) is signaled via the L2-Specific Sublayer AVP, Attribute Type 69, as defined in [RFC3931]. The V-bit within the L2-Specific Sublayer is used to identify that a VCCV message follows, and when the V-bit is set the L2SS has the format shown in Figure 5:

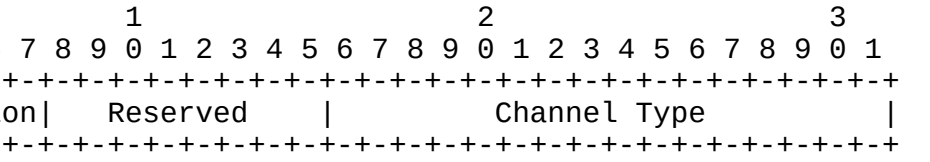


Figure 5: L2-Specific Sublayer Format when the V-bit (bit 0) is set

The VCCV messages are distinguished from user data by the V-bit. The V-bit is set to 1, indicating that a VCCV session message follows. The next three bits MUST be set to 0 when sending and ignored upon receipt. The remaining fields comprising 28 bits (i.e., Version, Reserved, and Channel Type) follow the same definition, format, and number registry from Section 5 of [RFC4385].

The Version and Reserved fields are set to 0. For the CV Type currently defined of ICMP Ping (0x01), the Channel Type can indicate IPv4 (0x0021) or IPv6 (0x0057) (see [RFC4385]) as the VCCV payload directly following the L2SS.

6.2. VCCV Connectivity Verification Type for L2TPv3

The VCCV message over L2TPv3 directly follows the L2-Specific Sublayer with the V-bit set. It MUST contain an ICMP Echo packet as described in Section 6.2.1.

6.2.1. L2TPv3 VCCV using ICMP Ping

When this connectivity verification mode is used, an ICMP Echo packet using the encoding specified in [RFC0792] for (ICMPv4) or [RFC4443] (for ICMPv6) achieves connectivity verification. Implementations MUST use ICMPv4 [RFC0792] if the signaling for the L2TPv3 PW used IPv4 addresses, or ICMPv6 [RFC4443] if IPv6 addresses were used. If the pseudowire is set-up statically, then the encoding MUST use that which was used for the pseudowire in the configuration.

The ICMP Ping packet directly follows the L2SS with the V-bit set. In the ICMP Echo request, the IP Header fields MUST have the following values: the destination IP address is set to the remote LCCE's IP address for the tunnel endpoint, the source IP address is set to the local LCCE's IP address for the tunnel endpoint, and the TTL or Hop Limit is set to 1.

6.3. L2TPv3 VCCV Capability Advertisement for L2TPv3

A new optional AVP is defined in Section 6.3.1 to indicate the VCCV capabilities during session establishment. An LCCE MUST signal its desire to use connectivity verification for a particular L2TPv3 session and its VCCV capabilities using the VCCV Capability AVP.

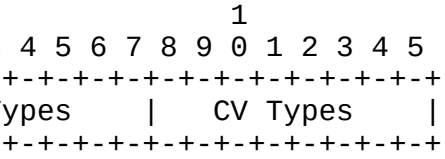
An LCCE MUST NOT send VCCV packets on an L2TPv3 session unless it has received VCCV capability by means of the VCCV Capability AVP from the remote end. If an LCCE receives VCCV packets and it is not VCCV capable or it has not sent VCCV capability indication to the remote end, it MUST discard these messages. It should also increment an error counter. In this case the LCCE MAY optionally issue a system and/or SNMP notification.

6.3.1. L2TPv3 VCCV Capability AVP

The "VCCV Capability AVP", Attribute Type 96, specifies the VCCV capabilities as a pair of bitflags for the Control Channel (CC) and Connectivity Verification (CV) Types. This AVP is exchanged during

session establishment (in ICRQ (Incoming-Call-Request), ICRP (Incoming-Call-Reply), OCRQ (Outgoing-Call-Request), or OCRP (Outgoing-Call-Reply) messages). The value field has the following format:

VCCV Capability AVP (ICRQ, ICRP, OCRQ, OCRP)



CC Types:

The Control Channel (CC) Types field defines a bitmask used to indicate the type of control channel(s) that may be used to receive OAM traffic on for the given Session. The router agrees to accept VCCV traffic at any time over any of the signaled VCCV control channel types. CC Type values are defined in Section 4. Although there is only one value defined in this document, the CC Types field is included for forward compatibility should further CC Types need to be defined in the future.

A CC Type of 0x01 may only be requested when there is an L2-Specific Sublayer that defines the V-bit present. If a CC Type of 0x01 is requested without requesting an L2-Specific Sublayer AVP with an L2SS type that defines the V-bit, the session MUST be disconnected with a Call-Disconnect-Notify (CDN) message.

If no CC Type is supported, a CC Type Indicator of 0x00 SHOULD be sent.

CV Types:

The Connectivity Verification (CV) Types field defines a bitmask used to indicate the specific type or types (i.e., none, one, or more) of control packets that may be sent on the specified VCCV control channel. CV Type values are defined in Section 4.

If no VCCV Capability AVP is signaled, then the LCCE MUST assume that the peer is incapable of receiving VCCV and MUST NOT send any OAM control channel traffic to it.

All L2TP AVPs have an M (Mandatory) bit, H (Hidden) bit, Length, and Vendor ID. The Vendor ID for the VCCV Capability AVP MUST be 0, indicating that this is an IETF-defined AVP. This AVP MAY be hidden (the H bit MAY be 0 or 1). The M bit for this AVP SHOULD be set to 0. The Length (before hiding) of this AVP is 8.

7. Capability Advertisement Selection

When a PE receives a VCCV capability advertisement, the advertisement may potentially contain more than one CC or CV Type. Only matching capabilities can be selected. When multiple capabilities match, only one CC Type MUST be used.

In particular, as already specified, once a valid CC Type is used by a PE (traffic sent using that encapsulation), the PE MUST NOT send any traffic down another CC Type control channel.

For cases where multiple CC Types are advertised, the following precedence rules apply when choosing the single CC Type to use:

1. Type 1: PWE3 Control Word with 0001b as first nibble

2. Type 2: MPLS Router Alert Label

3. Type 3: MPLS PW Label with TTL == 1

For MPLS PWs, the CV Type of LSP Ping (0x02) is the default, and the CV Type of ICMP Ping (0x01) is optional.

8. IANA Considerations

8.1. VCCV Interface Parameters Sub-TLV

The VCCV Interface Parameters Sub-TLV codepoint is defined in [RFC4446]. IANA has created and will maintain registries for the CC Types and CV Types (bitmasks in the VCCV Parameter ID). The CC Type and CV Type new registries (see Sections 8.1.1 and 8.1.2, respectively) have been created in the Pseudo Wires Name Spaces, reachable from [IANA.pwe3-parameters]. The allocations must be done using the "IETF Consensus" policy defined in [RFC2434].

8.1.1. MPLS VCCV Control Channel (CC) Types

IANA has set up a registry of "MPLS VCCV Control Channel Types". These are 8 bitfields. CC Type values 0x01, 0x02, and 0x04 are specified in Section 4 of this document. The remaining bitfield values (0x08, 0x10, 0x20, 0x40, and 0x80) are to be assigned by IANA using the "IETF Consensus" policy defined in [RFC2434]. A VCCV

Control Channel Type description and a reference to an RFC approved by the IESG are required for any assignment from this registry.

MPLS Control Channel (CC) Types:

Bit (Value)	Description
Bit 0 (0x01)	Type 1: PWE3 Control Word with 0001b as first nibble (PW-ACH, see [RFC4385])
Bit 1 (0x02)	Type 2: MPLS Router Alert Label
Bit 2 (0x04)	Type 3: MPLS PW Label with TTL == 1
Bit 3 (0x08)	Reserved
Bit 4 (0x10)	Reserved
Bit 5 (0x20)	Reserved
Bit 6 (0x40)	Reserved
Bit 7 (0x80)	Reserved

The most significant (high order) bit is labeled Bit 7, and the least significant (low order) bit is labeled Bit 0, see parenthetical "Value".

8.1.2. MPLS VCCV Connectivity Verification (CV) Types

IANA has set up a registry of "MPLS VCCV Control Verification Types". These are 8 bitfields. CV Type values 0x01 and 0x02 are specified in Section 4 of this document. The remaining bitfield values (0x04, 0x08, 0x10, 0x20, 0x40, and 0x80) are to be assigned by IANA using the "IETF Consensus" policy defined in [RFC2434]. A VCCV Control Verification Type description and a reference to an RFC approved by the IESG are required for any assignment from this registry.

MPLS Connectivity Verification (CV) Types:

Bit (Value)	Description
Bit 0 (0x01)	ICMP Ping
Bit 1 (0x02)	LSP Ping
Bit 2 (0x04)	Reserved
Bit 3 (0x08)	Reserved
Bit 4 (0x10)	Reserved
Bit 5 (0x20)	Reserved
Bit 6 (0x40)	Reserved
Bit 7 (0x80)	Reserved

The most significant (high order) bit is labeled Bit 7, and the least significant (low order) bit is labeled Bit 0, see parenthetical "Value".

8.2. PW Associated Channel Type

The PW Associated Channel Types used by VCCV as defined in Sections 5.1.1 and 6.1 rely on previously allocated numbers from the Pseudowire Associated Channel Types Registry [RFC4385] in the Pseudo Wires Name Spaces reachable from [IANA.pwe3-parameters]. In particular, 0x21 (Internet Protocol version 4) MUST be used whenever an IPv4 payload follows the Pseudowire Associated Channel Header, or 0x57 MUST be used when an IPv6 payload follows the Pseudowire Associated Channel Header.

8.3. L2TPv3 Assignments

Section 8.3.1 through Section 8.3.3 are registrations of new L2TP values for registries already managed by IANA. Section 8.3.4 is a new registry that has been added to the existing L2TP name spaces, and will be maintained by IANA accordingly. The Layer Two Tunneling Protocol "L2TP" Name Spaces are reachable from [IANA.l2tp-parameters].

8.3.1. Control Message Attribute Value Pairs (AVPs)

An additional AVP Attribute is specified in Section 6.3.1. It was defined by IANA as described in Section 2.2 of [RFC3438].

Attribute Type	Description
96	VCCV Capability AVP

8.3.2. Default L2-Specific Sublayer Bits

The Default L2-Specific Sublayer contains 8 bits in the low-order portion of the header. This document defines one reserved bit in the Default L2-Specific Sublayer in Section 6.1, which was assigned by IANA following IETF Consensus [RFC2434].

Default L2-Specific Sublayer bits - per [RFC3931]
Bit 0 - V (VCCV) bit

8.3.3. ATM-Specific Sublayer Bits

The ATM-Specific Sublayer contains 8 bits in the low-order portion of the header. This document defines one reserved bit in the ATM-Specific Sublayer in Section 6.1, which was assigned by IANA following IETF Consensus [RFC2434].

ATM-Specific Sublayer bits - per [RFC4454]

Bit 0 - V (VCCV) bit

8.3.4. VCCV Capability AVP Values

This is a new registry that IANA maintains in the L2TP Name Spaces.

IANA created and maintains a registry for the CC Types and CV Types bitmasks in the VCCV Capability AVP, defined in Section 6.3.1. The allocations must be done using the "IETF Consensus" policy defined in [RFC2434]. A VCCV CC or CV Type description and a reference to an RFC approved by the IESG are required for any assignment from this registry.

IANA has reserved the following bits in this registry:

VCCV Capability AVP (Attribute Type 96) Values

L2TPv3 Control Channel (CC) Types:

Bit (Value)	Description
Bit 0 (0x01)	L2-Specific Sublayer with V-bit set
Bit 1 (0x02)	Reserved
Bit 2 (0x04)	Reserved
Bit 3 (0x08)	Reserved
Bit 4 (0x10)	Reserved
Bit 5 (0x20)	Reserved
Bit 6 (0x40)	Reserved
Bit 7 (0x80)	Reserved

L2TPv3 Connectivity Verification (CV) Types:

Bit (Value)	Description
Bit 0 (0x01)	ICMP Ping
Bit 1 (0x02)	Reserved
Bit 2 (0x04)	Reserved
Bit 3 (0x08)	Reserved
Bit 4 (0x10)	Reserved
Bit 5 (0x20)	Reserved
Bit 6 (0x40)	Reserved
Bit 7 (0x80)	Reserved

The most significant (high order) bit is labeled Bit 7, and the least significant (low order) bit is labeled Bit 0, see parenthetical "Value".

9. Congestion Considerations

The bandwidth resources used by VCCV are recommended to be minimal compared to those of the associated PW. The bandwidth required for the VCCV channel is taken outside any allocation for PW data traffic, and can be configurable. When doing resource reservation or network planning, the bandwidth requirements for both PW data and VCCV traffic need to be taken into account.

VCCV applications (i.e., Connectivity Verification (CV) Types) MUST consider congestion and bandwidth usage implications and provide details on bandwidth or packet frequency management. VCCV applications can have built-in bandwidth management in their protocols. Other VCCV applications can have their bandwidth configuration-limited, and rate-limiting them can be harmful as it could translate to incorrectly declaring connectivity failures. For all other VCCV applications, outgoing VCCV messages SHOULD be rate-limited to prevent aggressive connectivity verification consuming excessive bandwidth, causing congestion, becoming denial-of-service attacks, or generating an excessive packet rate at the CE-bound PE.

If these conditions cannot be followed, an adaptive loss-based scheme SHOULD be applied to congestion-control outgoing VCCV traffic, so that it competes fairly with TCP within an order of magnitude. One method of determining an acceptable bandwidth for VCCV is described in [RFC3448] (TFRC); other methods exist. For example, bandwidth or packet frequency management can include any of the following: a negotiation of transmission interval/rate, a throttled transmission rate on "congestion detected" situations, a slow-start after shutdown due to congestion and until basic connectivity is verified, and other mechanisms.

The ICMP and MPLS LSP PING applications SHOULD be rate-limited to below 5% of the bit-rate of the associated PW. For this purpose, the considered bit-rate of a pseudowire is dependent on the PW type. For pseudowires that carry constant bit-rate traffic (e.g., TDM PWs) the full bit-rate of the PW is used. For pseudowires that carry variable bit-rate traffic (e.g., Ethernet PWs), the mean or sustained bit-rate of the PW is used.

As described in Section 10, incoming VCCV messages can be rate-limited as a protection against denial-of-service attacks. This throttling or policing of incoming VCCV messages should not be more stringent than the bandwidth allocated to the VCCV channel to prevent false indications of connectivity failure.

10. Security Considerations

Routers that implement VCCV create a Control Channel (CC) associated with a pseudowire. This control channel can be signaled (e.g., using LDP or L2TPv3 depending on the PWE3) or statically configured. Over this control channel, VCCV Connectivity Verification (CV) messages are sent. Therefore, three different areas are of concern from a security standpoint.

The first area of concern relates to control plane parameter and status message attacks, that is, attacks that concern the signaling of VCCV capabilities. MPLS PW Control Plane security is discussed in Section 8.2 of [RFC4447]. L2TPv3 PW Control Plane security is discussed in Section 8.1 of [RFC3931]. The addition of the connectivity verification negotiation extensions does not change the security aspects of Section 8.2 of [RFC4447], or Section 8.1 of [RFC3931]. Implementation of IP source address filters may also aid in deterring these types of attacks.

A second area of concern centers on data-plane attacks, that is, attacks on the associated channel itself. Routers that implement the VCCV mechanisms are subject to additional data-plane denial-of-service attacks as follows:

An intruder could intercept or inject VCCV packets effectively providing false positives or false negatives.

An intruder could deliberately flood a peer router with VCCV messages to deny services to others.

A misconfigured or misbehaving device could inadvertently flood a peer router with VCCV messages which could result in denial of services. In particular, if a router has either implicitly or explicitly indicated that it cannot support one or all of the types of VCCV, but is sent those messages in sufficient quantity, it could result in a denial of service.

To protect against these potential (deliberate or unintentional) attacks, multiple mitigation techniques can be employed:

VCCV message throttling mechanisms can be used, especially in distributed implementations which have a centralized control-plane

processor with various line cards attached by some control-plane data path. In these architectures, VCCV messages may be processed on the central processor after being forwarded there by the receiving line card. In this case, the path between the line card and the control processor may become saturated if appropriate VCCV traffic throttling is not employed, which could lead to a complete denial of service to users of the particular line card. Such filtering is also useful for preventing the processing of unwanted VCCV messages, such as those which are sent on unwanted (and perhaps unadvertised) control channel types or VCCV types.

Section 8.1 of [RFC4447] discusses methods to protect the data plane of MPLS PWs from data-plane attacks. However the implementation of the connectivity verification protocol expands the range of possible data-plane attacks. For this reason implementations MUST provide a method to secure the data plane. This can be in the form of encryption of the data by running IPsec on MPLS packets encapsulated according to [RFC4023], or by providing the ability to architect the MPLS network in such a way that no external MPLS packets can be injected (private MPLS network).

For L2TPv3, data packet spoofing considerations are outlined in Section 8.2 of [RFC3931]. While the L2TPv3 Session ID provides traffic separation, the optional Cookie field provides additional protection to thwart spoofing attacks. To maximize protection against a variety of data-plane attacks, a 64-bit Cookie can be used. L2TPv3 can also be run over IPsec as detailed in Section 4.1.3 of [RFC3931].

A third and last area of concern relates to the processing of the actual contents of VCCV messages, i.e., LSP Ping and ICMP messages. Therefore, the corresponding security considerations for these protocols (LSP Ping [RFC4379], ICMPv4 Ping [RFC0792], and ICMPv6 Ping [RFC4443]) apply as well.

11. Acknowledgements

The authors would like to thank Hari Rakotoranto, Michel Khouderchah, Bertrand Duvivier, Vanson Lim, Chris Metz, W. Mark Townsley, Eric Rosen, Dan Tappan, Danny McPherson, Luca Martini, Don O'Connor, Neil Harrison, Danny Prairie, Mustapha Aissaoui, and Vasile Radoaca for their valuable comments and suggestions.

12. References

12.1. Normative References

[RFC0792] Postel, J., "Internet Control Message Protocol", STD 5, RFC 792, September 1981.

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

[RFC3032] Rosen, E., Tappan, D., Fedorkow, G., Rekhter, Y., Farinacci, D., Li, T., and A. Conta, "MPLS Label Stack Encoding", RFC 3032, January 2001.

[RFC3931] Lau, J., Townsley, M., and I. Goyret, "Layer Two Tunneling Protocol - Version 3 (L2TPv3)", RFC 3931, March 2005.

[RFC4379] Kompella, K. and G. Swallow, "Detecting Multi-Protocol Label Switched (MPLS) Data Plane Failures", RFC 4379, February 2006.

[RFC4385] Bryant, S., Swallow, G., Martini, L., and D. McPherson, "Pseudowire Emulation Edge-to-Edge (PWE3) Control Word for Use over an MPLS PSN", RFC 4385, February 2006.

[RFC4443] Conta, A., Deering, S., and M. Gupta, "Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification", RFC 4443, March 2006.

[RFC4446] Martini, L., "IANA Allocations for Pseudowire Edge to Edge Emulation (PWE3)", BCP 116, RFC 4446, April 2006.

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

12.2. Informative References

[IANA.l2tp-parameters]
Internet Assigned Numbers Authority, "Layer Two Tunneling Protocol "L2TP"", April 2007, <http://www.iana.org/assignments/l2tp-parameters>.

[IANA.pwe3-parameters]
Internet Assigned Numbers Authority, "Pseudo Wires Name Spaces", June 2007, <http://www.iana.org/assignments/pwe3-parameters>.

[MSG-MAP] Nadeau, T., "Pseudo Wire (PW) OAM Message Mapping", Work in Progress, March 2007.

[RFC2434] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 2434, October 1998.

[RFC3438] Townsley, W., "Layer Two Tunneling Protocol (L2TP) Internet Assigned Numbers Authority (IANA) Considerations Update", BCP 68, RFC 3438, December 2002.

[RFC3448] Handley, M., Floyd, S., Padhye, J., and J. Widmer, "TCP Friendly Rate Control (TFRC): Protocol Specification", RFC 3448, January 2003.

[RFC3916] Xiao, X., McPherson, D., and P. Pate, "Requirements for Pseudo-Wire Emulation Edge-to-Edge (PWE3)", RFC 3916, September 2004.

[RFC3985] Bryant, S. and P. Pate, "Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture", RFC 3985, March 2005.

[RFC4023] Worster, T., Rekhter, Y., and E. Rosen, "Encapsulating MPLS in IP or Generic Routing Encapsulation (GRE)", RFC 4023, March 2005.

[RFC4377] Nadeau, T., Morrow, M., Swallow, G., Allan, D., and S. Matsushima, "Operations and Management (OAM) Requirements for Multi-Protocol Label Switched (MPLS) Networks", RFC 4377, February 2006.

[RFC4448] Martini, L., Rosen, E., El-Aawar, N., and G. Heron, "Encapsulation Methods for Transport of Ethernet over MPLS Networks", RFC 4448, April 2006.

[RFC4454] Singh, S., Townsley, M., and C. Pignataro, "Asynchronous Transfer Mode (ATM) over Layer 2 Tunneling Protocol Version 3 (L2TPv3)", RFC 4454, May 2006.

[RFC4928] Swallow, G., Bryant, S., and L. Andersson, "Avoiding Equal Cost Multipath Treatment in MPLS Networks", BCP 128, RFC 4928, June 2007.

Appendix A. Contributors' Addresses

George Swallow
Cisco Systems, Inc.
300 Beaver Brook Road
Boxborough, MA 01719
USA

EMail: swallow@cisco.com

Monique Morrow
Cisco Systems, Inc.
Glatt-com
CH-8301 Glattzentrum
Switzerland

EMail: mmorrow@cisco.com

Yuichi Ikejiri
NTT Communication Corporation
1-1-6, Uchisaiwai-cho, Chiyoda-ku
Tokyo 100-8019
Shinjuku-ku
JAPAN

EMail: y.ikejiri@ntt.com

Kenji Kumaki
KDDI Corporation
KDDI Bldg. 2-3-2
Nishishinjuku
Tokyo 163-8003
JAPAN

EMail: ke-kumaki@kddi.com

Peter B. Busschbach
Alcatel-Lucent
67 Whippany Road
Whippany, NJ, 07981
USA

EMail: busschbach@alcatel-lucent.com

Rahul Aggarwal
Juniper Networks
1194 North Mathilda Ave.
Sunnyvale, CA 94089
USA

EMail: rahul@juniper.net

Luca Martini
Cisco Systems, Inc.
9155 East Nichols Avenue, Suite 400
Englewood, CO, 80112
USA

EMail: lmartini@cisco.com

Authors' Addresses

Thomas D. Nadeau (editor)
Cisco Systems, Inc.
300 Beaver Brook Road
Boxborough, MA 01719
USA

EMail: tnadeau@lucidvision.com

Carlos Pignataro (editor)
Cisco Systems, Inc.
7200 Kit Creek Road
PO Box 14987
Research Triangle Park, NC 27709
USA

EMail: cpignata@cisco.com

Full Copyright Statement

Intellectual Property