Network Working Group
Internet Draft
Expires: August 2004

Thomas D. Nadeau
Cisco Systems, Inc.

Rahul Aggarwal
Juniper Networks
Editors

February 2004

Pseudo Wire (PW) Virtual Circuit Connection Verification (VCCV)
draft-ietf-pwe3-vccv-02.txt

Status of this Memo

Abstract

This document describes Virtual Circuit Connection Verification (VCCV) procedures for use with pseudowire connections. VCCV supports connection verification applications for pseudowire VCs regardless of the underlying MPLS or IP tunnel technology. VCCV makes use of IP based protocols such as Ping and MPLS LSP Ping to perform operations and maintenance functions. This is accomplished by providing an IP control channel associated with each pseudowire. A network operator may use the VCCV procedures to test the network's forwarding plane liveliness.

Contents

1. Contributors

Thomas D. Nadeau
Cisco Systems, Inc.
300 Beaver Brook Road
Boxborough, MA 01719
Email: tnadeau@cisco.com

Rahul Aggarwal
Juniper Networks
1194 North Mathilda Ave.
Sunnyvale, CA 94089
Email: rahul@juniper.net

George Swallow
Cisco Systems, Inc.
300 Beaver Brook Road
Boxborough, MA 01719
Email: swallow@cisco.com

Monique Morrow
Cisco Systems, Inc.
Glatt-com
CH-8301 Glattzentrum
Switzerland
Email: mmorrow@cisco.com

Yuichi Ikejiri
NTT Communication Corporation

Kenji Kumaki
KDDI Corporation

1-1-6, Uchisaiwai-cho, Chiyoda-ku
Tokyo 100-8019
Shinjuku-ku, JAPAN
Email: y.ikejiri@ntt.com

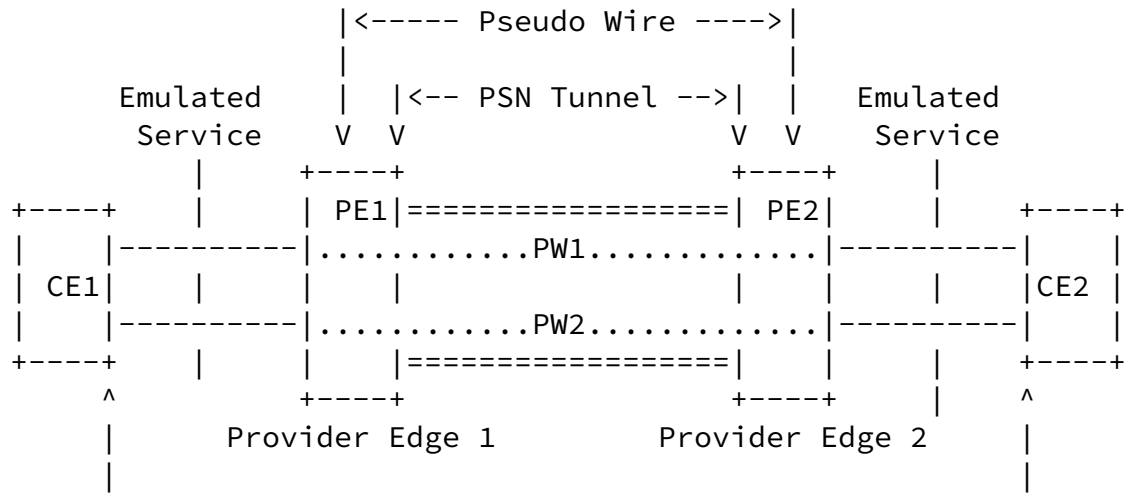
KDDI Bldg. 2-3-2,
Nishishinjuku,
Tokyo 163-8003,
JAPAN
E-mail: ke-kumaki@kddi.com

Peter B. Busschbach
Lucent Technologies
67 Whippany Road
Whippany, NJ, 07981
E-mail: busschbach@lucent.com

Vasile Radoaca
Nortel Networks
Billerica, MA, 01803
Email: vasile@nortelnetworks.com
Voice: 978-288-6097

2. Introduction

As network operators deploy pseudowire services, fault detection and diagnostic mechanisms particularly for the PSN portion of the network are pivotal. Specifically, the ability to provide end-to-end fault detection and diagnostics for an emulated pseudowire service is critical for the network operator. Operators have indicated in [MPLSOAMREQS] that such a tool is required for pseudowire deployments. This document describes procedures for PSN-agnostic fault detection and diagnostics called virtual circuit connection verification (VCCV).



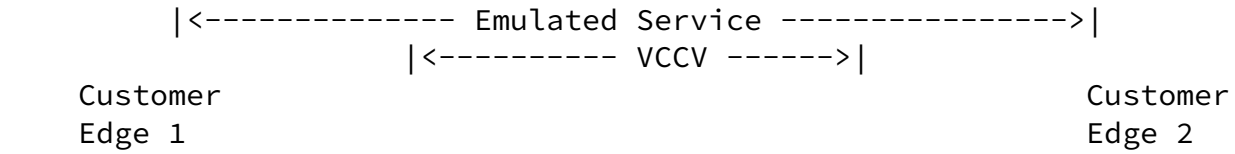
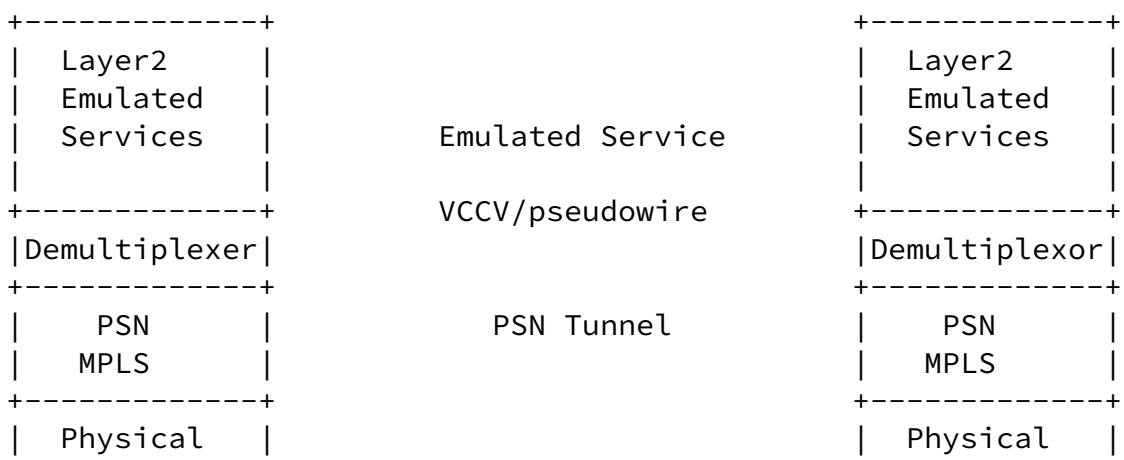


Figure 1: PWE3 VCCV Operation Reference Model

Figure 1 depicts the basic functionality of VCCV. VCCV provides several means of creating a control channel between PEs that attaches the VC under test.



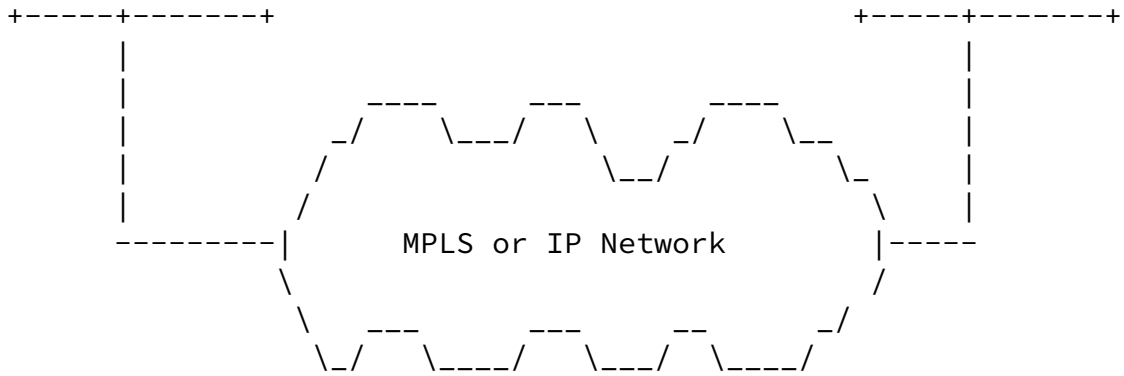


Figure 2: PWE3 Protocol Stack Reference Model including the VCCV control channel.

Figure 2 depicts how the VCCV IP control channel is associated with the pseudowire. Ping and other IP messages are encapsulated

using the PWE3 encapsulation as described below in sections 5 and 6. These messages, referred to as VCCV messages, are exchanged only after the desire to exchange such traffic has been negotiated between the PEs (see section 8).

3. Overview of VCCV Modes of Operation

VCCV defines a set of messages that are exchanged between PEs to verify connectivity of the pseudowire. To make sure that pseudowire packets follow the same path as the data flow, they are encapsulated with the same labels. VCCV can operate in two modes:

1) as a diagnostic tool
2) as a fault detection tool

In the diagnostic mode, the operator triggers LSP-Ping, L2TPV3, or ICMP Ping modes depending on the underlying PSN. Since a pseudowire is bi-directional, it makes sense to require that the reply is send over the PSN tunnel that makes up the other half of the PW under test. For example, if the PSN is an MPLS LSP, the reply should be sent on the LSP representing the reverse path. If this fails, the operator can use other reply modes to determine what is wrong. The specific type of reply mode is indicated during PW circuit set-up (see section 6).

The fault detection mode provides a way to emulate fault detection mechanisms in other technologies, such as ATM for example. In the fault detection mode, the upstream PE sends BFD control messages periodically. [BFD-MPLS] describes procedures for using BFD to detect liveliness of MPLS LSPs. When the downstream PE doesn't receive these message for a defined period of time, it declares that direction of the PW down and it notifies the upstream PE. Based on the emulated

service, the PEs may send FDI and RDI indications over the related attachment circuits to notify the end points of the fault condition. This is described in more detail in [OAMMsgMap].

3.1 LSP Ping

When the PSN is MPLS, the LSP Ping header is used as described in [LSP-PING] for verifying the connectivity status of pseudo wires.

3.2 L2TPV3

When L2TPv3 is used as the underlying PSN, a VCCV mechanism is needed for the L2TPv3 session. The L2TPv3 control connection does employ a keepalive mechanism; however, this mechanism is not sufficent for fault detection and diagnostic of the L2TPv3 session i.e. data plane. In L2TPv3, a session is analogous to a PW. A L2TPv3 VCCV mechanism is needed in particular for verifying the session forwarding state at the egress router.

3.3 ICMP Ping

When IP is used as the PSN, ICMP ECHO packets can be used as the means by which connectivity verification is achieved using VCCV.

3.4 Bidirectional Forwarding Detection

When heart-beat indication is necessary for one or more pseudowires, the Bidirectional Forwarding Detection (BFD) [BFD] provides a light-weight means of continuous monitoring and propagation of forward and reverse defect indications. BFD can be used regardless of the underlying PSN technology.

4. MPLS as PSN

In order to apply IP monitoring tools to PWE3 circuits, VCCV creates a control channel between PWE3 PEs[PWEARCH]. Packets sent across this channel are IP packets, allowing maximum flexibility.

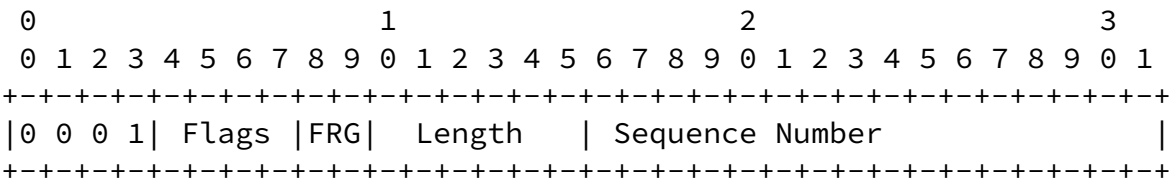
Ideally such a control channel would be completely in band. When a control word is present on virtual circuit, it is possible to indicate the control channel by setting a bit in the control header. This method is described in section 7.1 and is referred to as PWE3 inband VCCV.

However in order to address the case when the control header is not in use as well as to deal with a number of existent hardware devices, use of the MPLS Router Alert Label to indicate the IP control channel is also proposed. This is described in section 7.2.

The actual channel type is agreed through signaling as described in section 8.

4.1. PWE3 Inband VCCV

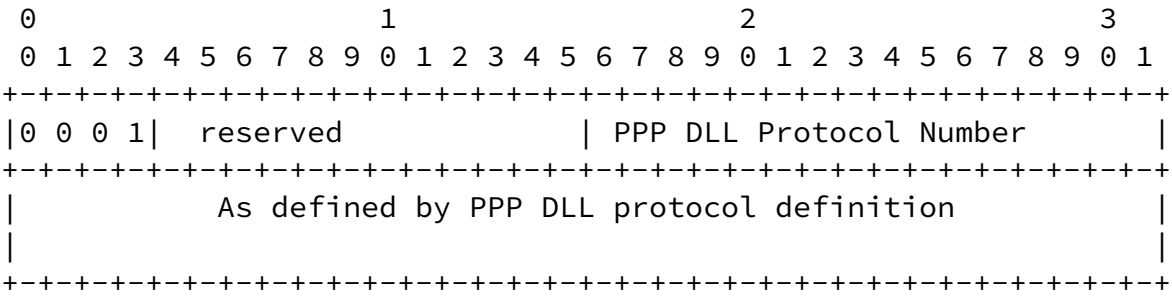
The PW set-up protocol determines whether a PW uses a control word. When a control word is used, it SHOULD have the following preferred form:



for the purpose of indicating VCCV control channel messages.

Note that for data, one uses the control word defined just above the MPLS payload [PWEARCH].

The PWE3 payload type identifier is defined as follows:



The first nibble 0000 indicates data. When the first nibble is=20 0001, the protocol of the frame is indicated by the Protocol Number. IP OAM flows are identified by either an IPv4 or IPv6 codepoint.

4.2. Router Alert Label Approach

When the control word is not used, or the receiving hardware

cannot divert control traffic based on information in the control word (i.e.: older hardware), an IP control channel can be created by including the MPLS router alert label immediately above the VC label. If the control word is in use on this VC

it is also included in the IP control flow.

5. IP Probe Traffic

For connectivity verification, both ICMP Ping and LSP-Ping packets may be used on the control channel. The type of packets used is indicated during signaling as described in section 6.

5.1. ICMP Ping

When ICMP packets are used, the source address should be set to the source address of the LDP session and the destination address to the destination of the LDP session. The Identifier and Sequence Number fields of the ICMP Echo Request/Echo Reply messages are used to track what VCs are being tested.

These fields are only interpreted by the sending PE. Specific use of these fields is an implementation matter.

5.2. MPLS Ping Packet

The LSP Ping header must be used as described [LSP-PING] and must also contain the sub-TLV of 8 for PW circuits. This sub-TLV must be sent containing the circuit to be verified as the "VC ID" field:

5.3 Bidirectional Forwarding Detection

When heart-beat indication is necessary for one or more pseudowires, the Bidirectional Forwarding Detection (BFD) [BFD] provides a light-weight means of continuous monitoring and propagation of forward and reverse defect indications.

In order to use BFD, both ends of the pseudowire connection must have signaled the existence of a control channel and the ability to run BFD. Once a node has both signaled and received signaling from its peer of these capabilities, it MUST begin sending BFD control packets. The packets MUST be sent on the control channel. The use of the control channel provides the context required to bind the BFD session to a particular pseudowire (FEC). Thus normal BFD initialization procedures are followed. BFD MUST be run in asynchronous mode.

It may be desirable to use LSP-Ping additionally for periodic diagnostics in addition to BFD for fault detection on the same PW. [BFD-MPLS] provides further details on how BFD can be used in

conjunction with LSP-Ping for detecting the liveliness of MPLS LSPs.

When one of the PEs (PE2) doesn't receive control messages from PE1 during the specified amount of time, or if it determines in another way that communication is lost,it declares that the PW in the direction from PE1 to PE2 is down. It stores the cause (e.g. control detection time expired) and sends a message to PE1 with H=3D0 (i.e. "I don't hear you"). In turn, PE1 declares the PW in the direction from PE1 to PE2 down and stores as cause: neighbor signaled session down. Depending on the emulated services, PE2 may send a FDI indication on its attachment circuits and PE1 may send an RDI indication on its attachment circuits.

BFD defines the following diagnostics:

0 -- No Diagnostic
1 -- Control Detection Time Expired
2 -- Echo Function Failed
3 -- Neighbor Signaled Session Down
4 -- Forwarding Plane Reset (Local equipment failure)
5 -- Path Down (Alarm Suppression)
6 -- Concatenated Path Down (Propagating access link alarm)
7 -- Administratively Down

Of these, 0 is used when the PW is up and 2 is not applicable to asynchronous mode.

6. OAM Capability Indication

To permit negotiation of the use and type of OAM for Connectivity Verification, a VCCV parameter is defined below. When a PE signals a PWE3 VC and desires OAM for that VC, it MUST indicate this during VC establishment using the messages defined below. Specifically for LDP it MUST include the VCCV parameter in the VC setup message.

As the overall method of PWE3 signaling is downstream, unsolicited, the decision of the type of IP control channel is left completely to the receiving control entity. When a PE sends a label for a PW, it uses the VCCV parameter to indicate the type of OAM channels it is willing to receive on that PW. The capablity of supporting a control channel MUST be signaled BEFORE the remote PE may send OAM messages, and then only on the type of control channel indicated.

If a PE receives OAM messages prior to sending a VCCV parameter, it MUST discard these messages and not reply

to them. In this case, the LSR SHOULD increment an error counter and optionally issues a system and/or SNMP notification to indicate to the system administrator that a mis-configuration exists.

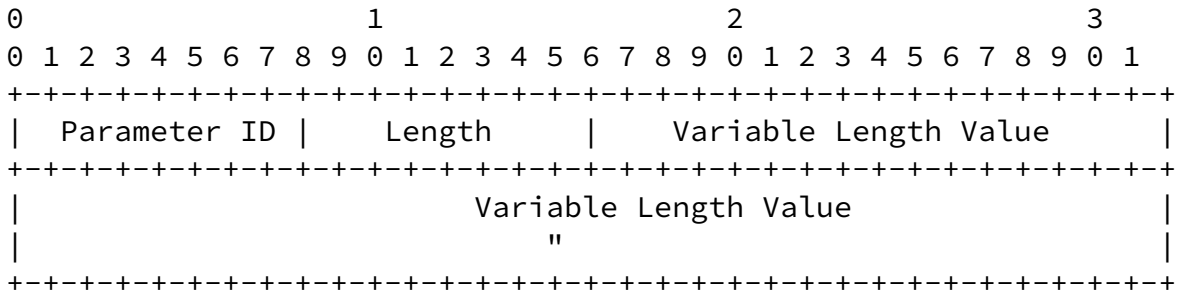
The requesting PE indicates its desire for the remote PE to support OAM capability by including the VCCV parameter with appropriate options set to indicate which methods of OAM are acceptable. The requesting PE MAY indicate multiple IP control channel options. The absence of the VCCV FEC TLV indicates that no OAM functions are supported or desired by the requesting PE. This last method MUST be supported by all PEs in order to handle backward-compatibility with older PEs. The receiving PE agrees to accept any of the indicated OAM types and options by virtue of establishing the VC. If it does not or cannot support at least one of the options specified, it MUST not establish the VC. If the requesting

PE wishes to continue, it may choose different options and try to signal the PW again.

6.1. Optional VCCV Parameter

[PWE3CONTROL] defines a VC FEC TLV for LDP. Parameters can be carried within that TLV to signal different capabilities for specific PWs. We propose an optional parameter to be used to indicate the desire to use a control channel for VCCV as follows.

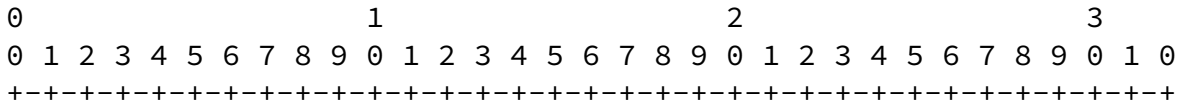
The TLV field structure is defined in [PWE3CONTROL] as follows:

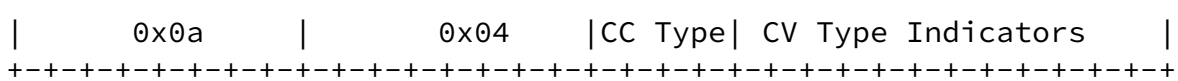


The VCCV parameter ID is defined as follows in [PWE3IANA]:

Parameter ID	Length	Description
0x0a	4	VCCV

The format of the VCCV parameter TLV is as follows:





The CC type field defines the type of IP control channel that will be used to receive. The defined values are:

0x01 PWE3 control word (0x0001 as first nibble of CW)
0x02 MPLS Router Alert Label

The CV Type Indicators field defines a bitmask used to indicate the specific type or types (i.e.: one or more) of IP control packets that may be sent on the control channel. The defined values are:

0x01 ICMP Ping
0x02 LSP Ping
0x04 BFD

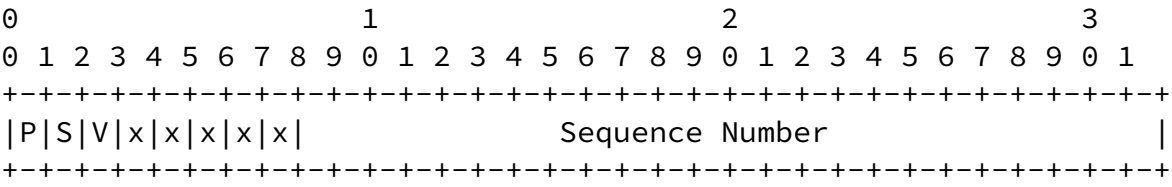
If none of the types above are supported, a CV Type Indicator of 0x00 SHOULD be transmitted to indicate this to the peer. However, if no capability is signaled, then the peer MUST assume that the peer has no VCCV capability.

7. L2TPV3 as PSN

When L2TPv3 is used as the underlying PSN, a VCCV mechanism is needed for the L2TPv3 session. The L2TPv3 control connection does employ a keepalive mechanism. However this mechanism is not sufficent for fault detection and diagnostic of the L2TPv3 session i.e. data plane. In L2TPv3 a session is analogous to a PW. A L2TPv3 VCCV mechanism is needed in particular for verifying the session forwarding state at the egress router.

When a PE verifies the connection status of a L2TPv3 session it must transmit a L2TPv3 VCCV message encoded in the L2TPv3 session packet.

The presence of a VCCV message in a L2TPv3 session packet can be indicated by reserving a bit in the default L2-specific sublayer format.



Default L2-Specific Sublayer Format with V bit.

The 'V' bit indicates that this is a VCCV session packet. If the PW has not been signaled to include a L2-specific sublayer format, other

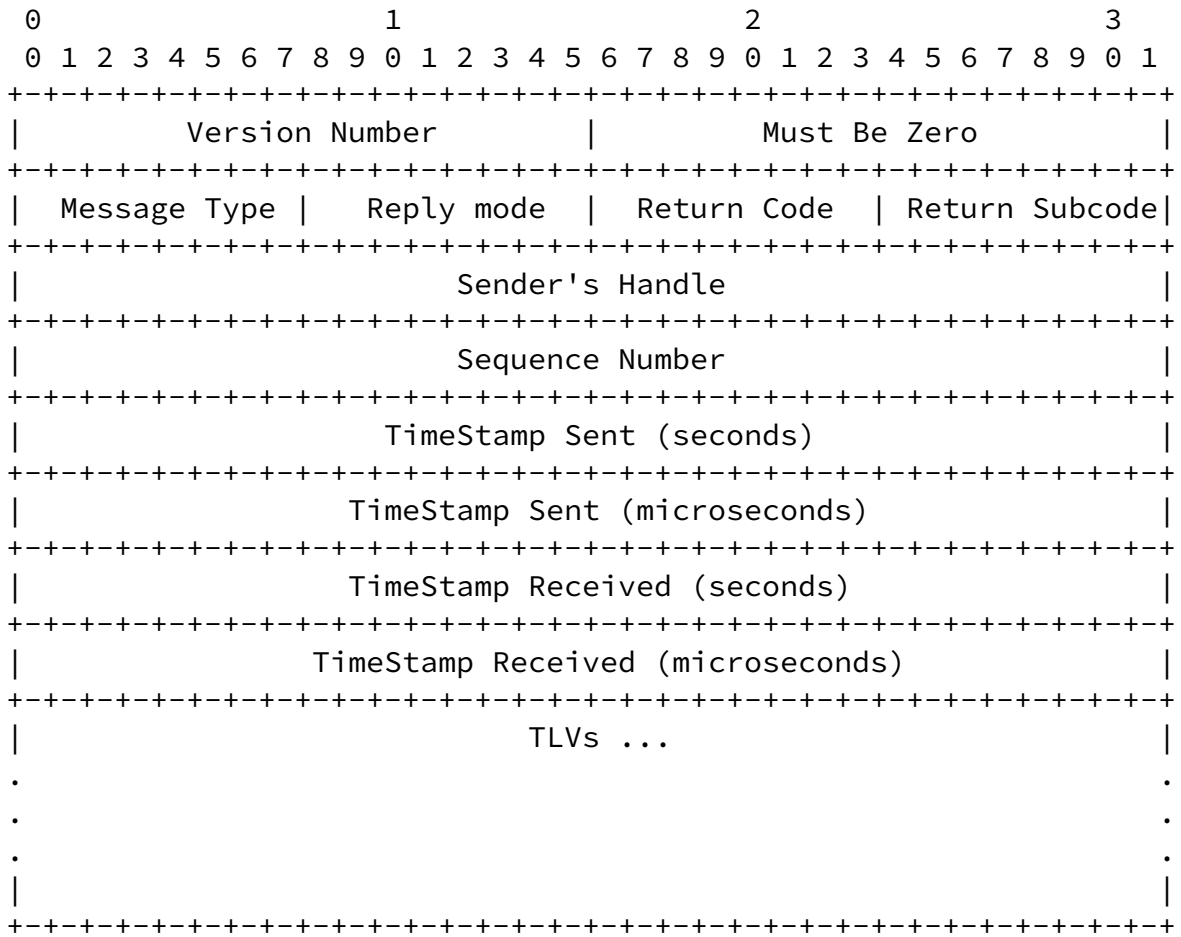
mechanisms are needed to indicate the VCCV message. Such mechanisms are for further study.

7.1. L2TPv3 VCCV Message

The VCCV message MUST contain a VCCV AVP. It does not contain a messageheader. A new AVP, called the VCCV AVP is defined. The usage of the L2TPv3 AVP format leaves room for adding further AVPs to this message in the future as needed.

7.1.1. L2TPv3 VCCV AVP

This AVP encodes the LSP Ping header as defined in [LSP-PING]. M and H bits must not be set. The attribute type is TBD. The LSP Ping header is not encapsulated in UDP. The modifications to the semantics of the fields of this header are specified here. Unless otherwise specified the semantics of the fields as explained in [LSP-PING] are to be followed. For reference the format of the LSP Ping header is shown below.



The version number is currently 1. The message type is one of the following:

1 - L2TPv3 VCCV Echo Request
2 - L2TPv3 VCCV Echo Reply

The Reply Mode is:

1 - Do not reply
2 - Reply using the L2TPv3 session

As explained in [LSP-PING] a reply mode of "do not reply" can be used

for one way connectivity tests. The VCCV message will normally contain a reply mode of "reply using the L2TPv3 session".

The return code can be set to the following by the receiver:

1 - Malformed echo request received
2 - One or more of the TLVs was not understood
3 - Replying router has a session mapping for the verified pseudowire
4 - Replying router does not have a mapping for the verified pseudowire

The LSP Ping header must contain the L2 Circuit ID TLV as defined in section 8.2. This TLV identifies the pseudo wire associated with the session, that is being verified. For L2TPv3 the remote PE address is the address of the session's remote end. A new PWID type is defined for L2TPv3, in addition to the ones defined in section 8.2:

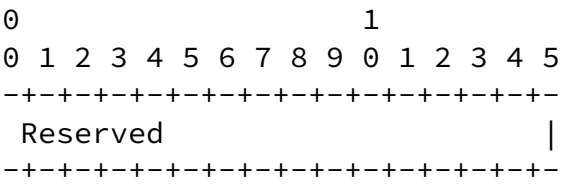
3. L2TPv3 Remote End Identifier AVP

7.2. L2TPv3 VCCV Capability Negotiation

A LCCE or a LAC should be able to indicate whether the session is capable of processing VCCV packets. This is done by including the optional VCCV capability AVP in an ICRQ, ICRP, OCRQ or OCRP.

7.2.1. L2TPv3 VCCV Capability AVP

This AVP specifies the VCCV capability. Its attribute type is TBD. The value field has the following format:



7.3. L2TPv3 VCCV Operation

A PE sends VCCV echo requests on a L2TPv3 signaled pseudo wire for fault detection and diagnostic of the L2TPv3 session. The destination

IP address in the echo request is set to the remote PE's IP address, while the source IP address is set to the local PE's IP address. The egress of the L2TPv3 session verifies the signaling and forwarding state of the pseudo wire, on reception of the VCCV message. Any faults detected can be signaled in the VCCV echo response. Its to be noted that the VCCV mechanism for L2TPv3 is primarily targeted at verifying the pseudo wire forwarding and signaling state at the egress PE. It also helps when L2TPv3 control and session paths are not identical.

A PE must send VCCV packets on a L2TPv3 session only if it has signaled VCCV capability to the remote end and received VCCV capability from the remote end. If a PE receives VCCV packets and its not VCCV capable or it has not received VCCV capability indication from the remote end, it must discard these messages. In addition if a PE receives VCCV messages and it has not received VCCV capability from the remote end, it should increment an error counter. In this case the PE can optionally issue a system and/or SNMP notification.

8. Acknowledgments

The authors would like to thank Hari Rakotoranto, Michel Khouderchah, Bertrand Duvivier, Vanson Lim, Chris Metz, W.Mark Townsley, Eric Rosen, Dan Tappan,and Danny McPherson for their valuable comments and suggestions.

9. References

9.1 Normative References

[BFD] Katz, D., Ward, D., Bidirectional Forwarding Indication, draft-katz-ward-bfd-01.txt, August 2003, work in progress.

[PWREQ] Xiao, X., McPherson, D., Pate, P., "Requirements for Pseudo Wire Emulation Edge to-Edge (PWE3)", < draft-ietf-pwe3-requirements-08.txt>, December 2003

[PWE3FW] Prayson Pate, et al., Internet draft, Framework for Pseudo Wire Emulation Edge-to-Edge (PWE3), draft-ietf-pwe3-framework-01.txt, work in progress.

[PWEARCH] Bryant, S., Pate, P., "PWE3 Architecture", Internet Draft, < draft-ietf-pwe3-arch-06.txt>, October 2003

[PWE3IANA] Martini, L., Townsley, M., "IANA Allocations fo pseudo Wire Edge to Edge Emulation (PWE3)", draft-ietf-pwe3-iana-allocation-01.txt, June 2003, work in progress.

[L2SIG] Rosen, E., LDP-based Signaling for L2VPNs, Internet Draft <draft-rosen-ppvpn-l2-signaling-02.txt>, September 2002.

[LSPPING] Kompella, K., Pan, P., Sheth, N., Cooper, D., Swallow, G., Wadhwa, S., Bonica, R., " Detecting MPLS Data Plane Failures", Internet Draft < draft-ietf-mpls-lsp-ping-04.txt>, October 2003

[MARTINISIG] "Transport of Layer 2 Frames Over MPLS", Martini et. al., draft-martini-l2circuit-trans-mpls-10.txt, August 2002

9.2 Informative References

[ICMP] Postel, J. "Internet Control Message Protocol, RFC 792

[PWEATM] Martini, L., et al., "Encapsulation Methods for Transport of ATM Cells/Frame Over IP and MPLS=20 Networks", Internet Draft <draft-ietf-pwe3-atm-encap-00.txt>, October 2002

[MPLSOAMREQS] Nadeau, T., et al,"OAM Requirements for MPLS Networks, Internet Draft <draft-ietf-oam-requirements-02.txt>, June 2003.

[OAMMsgMap] Nadeau, T., et al, " Pseudo Wire (PW) OAM Message Mapping, Internet Draft < draft-nadeau-pwe3-oam-msg-map-04.txt>, January, 2004.

[PWE3CONTROL] L.Martini et al., "Transport of Layer 2 Frames over MPLS, Internet Draft, <draft-ietf-pwe3-control-protocol-01.txt>, May 2003

[PPVPNFW] Callon, R., Suzuki, M., Gleeson, B., Malis, A., Muthukrishnan, K., Rosen, E., Sargor, C., and J. Yu,=20 "A Framework for Provider Provisioned Virtual=20 Private Networks", Internet Draft <draft-ietf-ppvpn-framework-01.txt>, July 2001.

[RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, January 2001.

[BFD-MPLS] Aggarwal, R. Kompella, K. "BFD for MPLS LSPs", Internet Draft <draft-raggarwa-mpls-bfd-00.txt>, October 2003.

10. Security Considerations

TBD.

11. Intellectual Property Rights Notices

The IETF takes no position regarding the validity or scope of any intellectual property or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; neither does it represent that it has made any effort to identify any such rights. Information on=20 the IETF's procedures with respect to rights in standards-track and standards-related documentation can be found in BCP-11. Copies of claims of rights made available for publication and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification=20 can be obtained from the IETF Secretariat.
The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights which may cover technology that may be required to practice this standard. Please address the information to the IETF=20 Executive Director.

11. Full Copyright Statement