Network Working Group
Request for Comments: 4448
Category: Standards Track

L. Martini, Ed.
E. Rosen
Cisco Systems, Inc.
N. El-Aawar
Level 3 Communications, LLC
G. Heron
Tellabs

April 2006

Encapsulation Methods for Transport of Ethernet over MPLS Networks

Status of This Memo

This document specifies an Internet standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "Internet Official Protocol Standards" (STD 1) for the standardization state and status of this protocol. Distribution of this memo is unlimited.

Copyright Notice

Abstract

An Ethernet pseudowire (PW) is used to carry Ethernet/802.3 Protocol Data Units (PDUs) over an MPLS network. This enables service providers to offer "emulated" Ethernet services over existing MPLS networks. This document specifies the encapsulation of Ethernet/802.3 PDUs within a pseudowire. It also specifies the procedures for using a PW to provide a "point-to-point Ethernet" service.

Table of Contents

1. Introduction

An Ethernet pseudowire (PW) allows Ethernet/802.3 [802.3] Protocol Data Units (PDUs) to be carried over a Multi-Protocol Label Switched [MPLS-ARCH] network. In addressing the issues associated with carrying an Ethernet PDU over a packet switched network (PSN), this document assumes that a pseudowire (PW) has been set up by using a control protocol such as the one as described in [PWE3-CTRL]. The design of Ethernet pseudowire described in this document conforms to the pseudowire architecture described in [RFC3985]. It is also assumed in the remainder of this document that the reader is familiar with RFC 3985.

The Pseudowire Emulation Edge-to-Edge (PWE3) Ethernet PDU consists of the Destination Address, Source Address, Length/Type, MAC Client Data, and padding extracted from a MAC frame as a concatenated octet sequence in their original order [PDU].

In addition to the Ethernet PDU format used within the pseudowire, this document discusses:

- Procedures for using a PW in order to provide a pair of Customer Edge (CE) routers with an emulated (point-to-point) Ethernet service, including the procedures for the processing of Provider Edge (PE)-bound and CE-bound Ethernet PDUs [RFC3985]

- Ethernet-specific quality of service (QoS) and security considerations

- Inter-domain transport considerations for Ethernet PW

The following two figures describe the reference models that are derived from [RFC3985] to support the Ethernet PW emulated services.

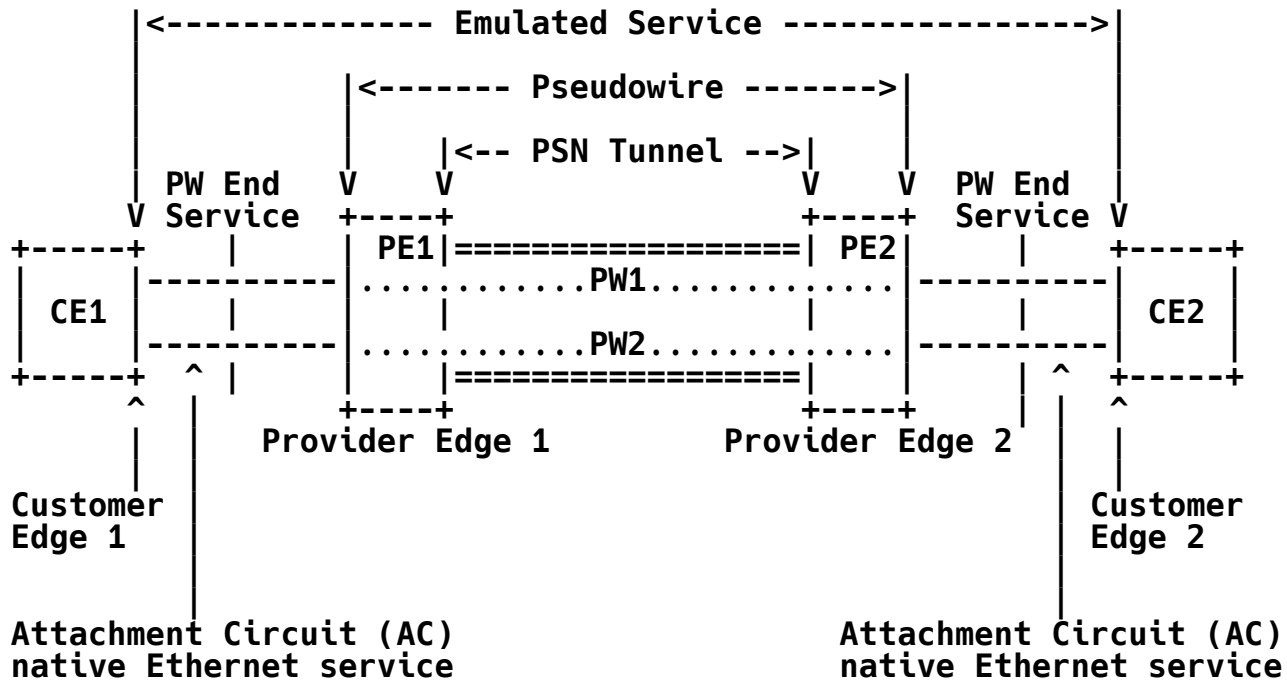


Figure 1: PWE3 Ethernet/VLAN Interface Reference Configuration

The "emulated service" shown in Figure 1 is, strictly speaking, a bridged LAN; the PEs have MAC interfaces, consume MAC control frames, etc. However, the procedures specified herein only support the case in which there are two CEs on the "emulated LAN". Hence we refer to this service as "emulated point-to-point Ethernet". Specification of the procedures for using pseudowires to emulate LANs with more than two CEs are out of the scope of the current document.

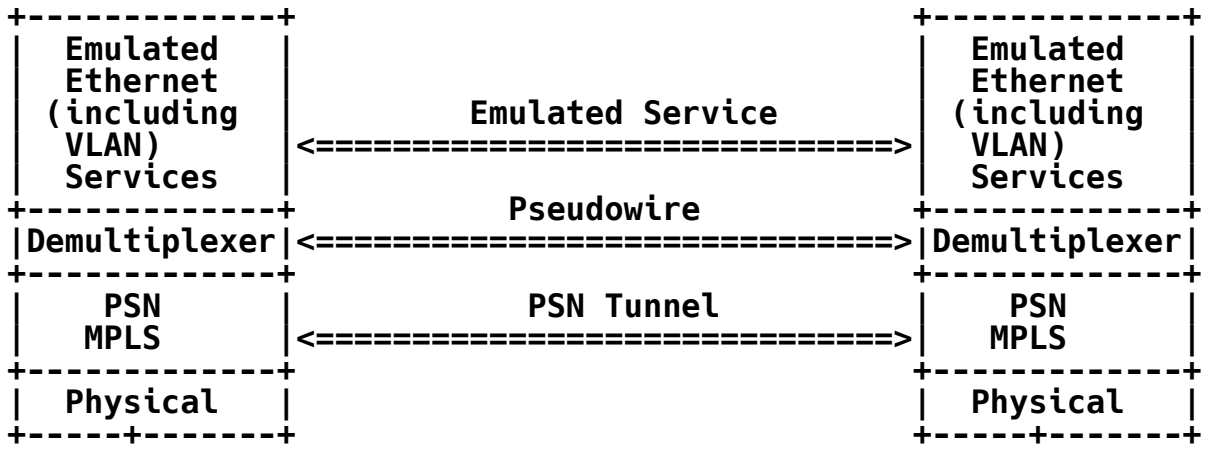


Figure 2: Ethernet PWE3 Protocol Stack Reference Model

For the purpose of this document, PE1 will be defined as the ingress router, and PE2 as the egress router. A layer 2 PDU will be received at PE1, encapsulated at PE1, transported, decapsulated at PE2, and transmitted out on the attachment circuit of PE2.

An Ethernet PW emulates a single Ethernet link between exactly two endpoints. The mechanisms described in this document are agnostic to that which is beneath the "Pseudowire" level in Figure 2, concerning itself only with the "Emulated Service" portion of the stack.

The following reference model describes the termination point of each end of the PW within the PE:

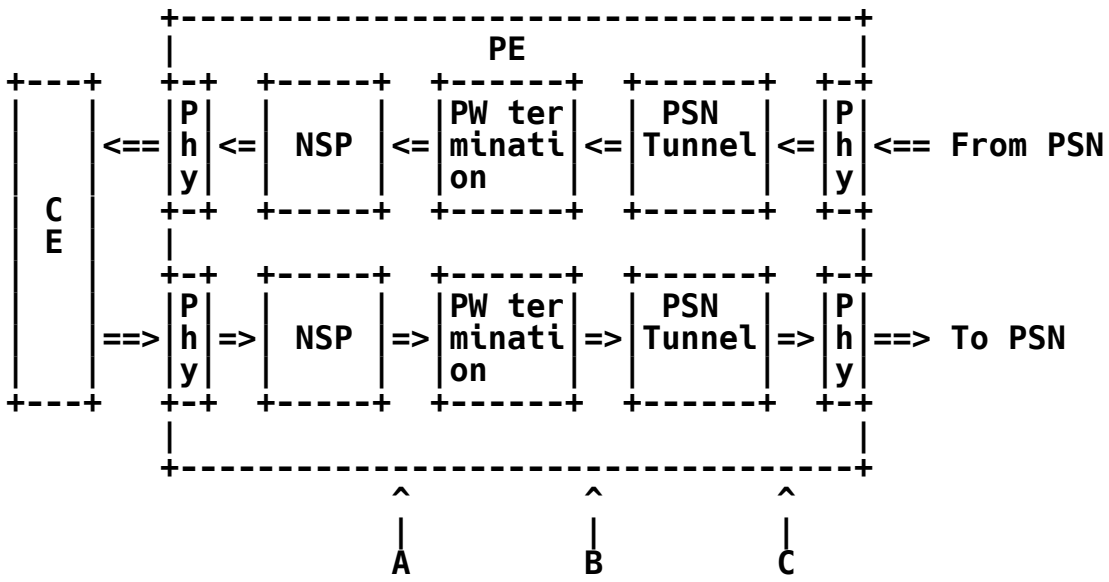


Figure 3: PW Reference Diagram

The PW terminates at a logical port within the PE, defined at point B in the above diagram. This port provides an Ethernet MAC service that will deliver each Ethernet frame that is received at point A, unaltered, to the point A in the corresponding PE at the other end of the PW.

The Native Service Processing (NSP) function includes frame processing that is required for the Ethernet frames that are forwarded to the PW termination point. Such functions may include stripping, overwriting or adding VLAN tags, physical port multiplexing and demultiplexing, PW-PW bridging, L2 encapsulation, shaping, policing, etc. These functions are specific to the Ethernet technology, and may not be required for the PW emulation service.

The points to the left of A, including the physical layer between the CE and PE, and any adaptation (NSP) functions between it and the PW terminations, are outside of the scope of PWE3 and are not defined here.

"PW Termination", between A and B, represents the operations for setting up and maintaining the PW, and for encapsulating and decapsulating the Ethernet frames as necessary to transmit them across the MPLS network.

An Ethernet PW operates in one of two modes: "raw mode" or "tagged mode". In tagged mode, each frame MUST contain at least one 802.1Q [802.1Q] VLAN tag, and the tag value is meaningful to the NSPs at the two PW termination points. That is, the two PW termination points must have some agreement (signaled or manually configured) on how to process the tag. On a raw mode PW, a frame MAY contain an 802.1Q VLAN tag, but if it does, the tag is not meaningful to the NSPs, and passes transparently through them.

Additional terminology relevant to pseudowires and Layer 2 Virtual Private Networking may be found in [RFC4026].

2. Specification of Requirements

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

3. Applicability Statement

The Ethernet PW emulation allows a service provider to offer a "port to port" Ethernet-based service across an MPLS packet switched network (PSN) while the Ethernet VLAN PW emulation allows an "Ethernet VLAN to VLAN" based service across an MPLS packet switched network (PSN).

The Ethernet or Ethernet VLAN PW has the following characteristics in relationship to the respective native service:

- An Ethernet PW connects two Ethernet ACs while an Ethernet VLAN PW connects two Ethernet VLAN ACs, supporting bidirectional transport of variable length Ethernet frames. The ingress Native Service Processing (NSP) function strips the preamble and frame check sequence (FCS) from the Ethernet frame and transports the frame in its entirety across the PW. This is done regardless of the presence of the 802.1Q tag in the frame. The egress NSP function receives the Ethernet frame from the PW and regenerates the preamble or FCS before forwarding the frame

to the attachment circuit. Since the FCS is not transported across either Ethernet or Ethernet VLAN PWs, payload integrity transparency may be lost. The OPTIONAL method described in [FCS] can be used to achieve payload integrity transparency on Ethernet or Ethernet VLAN PWs.

- For an Ethernet VLAN PW, VLAN tag rewrite can be achieved by NSP at the egress PE, which is outside the scope of this document.

- The Ethernet or Ethernet VLAN PW only supports homogeneous Ethernet frame type across the PW; both ends of the PW must be either tagged or untagged. Heterogeneous frame type support achieved with NSP functionality is outside the scope of this document.

- Ethernet port or Ethernet VLAN status notification is provided using the PW Status TLV in the Label Distribution Protocol (LDP) status notification message. Loss of connectivity between PEs can be detected by the LDP session closing, or by using [VCCV] mechanisms. The PE can convey these indications back to its attached Remote System.

- The maximum frame size that can be supported is limited by the PSN MTU minus the MPLS header size, unless fragmentation and reassembly are used [FRAG].

- The packet switched network may reorder, duplicate, or silently drop packets. Sequencing MAY be enabled in the Ethernet or Ethernet VLAN PW to detect lost, duplicate, or out-of-order packets on a per-PW basis.

- The faithfulness of an Ethernet or Ethernet VLAN PW may be increased by leveraging Quality of Service features of the PEs and the underlying PSN. (See Section 4.7, "QoS Considerations".)

4. Details Specific to Particular Emulated Services

4.1. Ethernet Tagged Mode

The Ethernet frame will be encapsulated according to the procedures defined later in this document for tagged mode. It should be noted that if the VLAN identifier is modified by the egress PE, the Ethernet spanning tree protocol might fail to work properly. If this issue is of significance, the VLAN identifier MUST be selected in such a way that it matches on the attachment circuits at both ends of the PW.

If the PE detects a failure on the Ethernet physical port, or the port is administratively disabled, it MUST send a PW status notification message for all PWs associated with the port.

This mode uses service-delimiting tags to map input Ethernet frames to respective PWs and corresponds to PW type 0x0004 "Ethernet Tagged Mode" [IANA].

4.2. Ethernet Raw Mode

The Ethernet frame will be encapsulated according to the procedures defined later in this document for raw mode. If the PE detects a failure on the Ethernet input port, or the port is administratively disabled, the PE MUST send an appropriate PW status notification message to the corresponding remote PE.

In this mode, all Ethernet frames received on the attachment circuit of PE1 will be transmitted to PE2 on a single PW. This service corresponds to PW type 0x0005 "Ethernet" [IANA].

4.3. Ethernet-Specific Interface Parameter LDP Sub-TLV

This LDP sub-Type Length Value [LDP] specifies interface-specific parameters. When applicable, it MUST be used to validate that the PEs, and the ingress and egress ports at the edges of the circuit, have the necessary capabilities to interoperate with each other. The Interface parameter TLV is defined in [PWE3-CTRL], the IANA registry with initial values for interface parameter sub-TLV types is defined in [IANA], but the Ethernet-specific interface parameters are specified as follows:

- 0x06 Requested VLAN ID Sub-TLV

  An Optional 16-bit value indicating the requested VLAN ID. This parameter MUST be used by a PE that is incapable of rewriting the 802.1Q Ethernet VLAN tag on output. If the ingress PE receives this request, it MUST rewrite the VLAN ID contained inside the VLAN Tag at the input to match the requested VLAN ID. If this is not possible, and the VLAN ID does not already match the configured ingress VLAN ID, the PW MUST not be enabled. This parameter is applicable only to PW type 0x0004.

4.4. Generic Procedures

When the NSP/Forwarder hands a frame to the PW termination function:

- The preamble (if any) and FCS are stripped off.
- The control word as defined in Section 4.6, "The Control Word", is, if necessary, prepended to the resulting frame. The conditions under which the control word is or is not used are specified below.
- The proper pseudowire demultiplexer (PW Label) is prepended to the resulting packet.
- The proper tunnel encapsulation is prepended to the resulting packet.
- The packet is transmitted.

The way in which the proper tunnel encapsulation and pseudowire demultiplexer is chosen depends on the procedures that were used to set up the pseudowire.

The tunnel encapsulation depends on how the MPLS PSN is set up. This can include no label, one label, or multiple labels. The proper pseudowire demultiplexer is an MPLS label whose value is determined by the PW setup and maintenance protocols.

When a packet arrives over a PW, the tunnel encapsulation and PW demultiplexer are stripped off. If the control word is present, it is processed and stripped off. The resulting frame is then handed to the Forwarder/NSP. Regeneration of the FCS is considered to be an NSP responsibility.

4.4.1. Raw Mode vs. Tagged Mode

When the PE receives an Ethernet frame, and the frame has a VLAN tag, we can distinguish two cases:

1. The tag is service-delimiting. This means that the tag was placed on the frame by some piece of service provider-operated equipment, and the tag is used by the service provider to distinguish the traffic. For example, LANs from different customers might be attached to the same service provider switch, which applies VLAN tags to distinguish one customer's traffic from another's, and then forwards the frames to the PE.

2. The tag is not service-delimiting. This means that the tag was placed in the frame by a piece of customer equipment, and is not meaningful to the PE.

Whether or not the tag is service-delimiting is determined by local configuration on the PE.

If an Ethernet PW is operating in raw mode, service-delimiting tags are NEVER sent over the PW. If a service-delimiting tag is present when the frame is received from the attachment circuit by the PE, it MUST be stripped (by the NSP) from the frame before the frame is sent to the PW.

If an Ethernet PW is operating in tagged mode, every frame sent on the PW MUST have a service-delimiting VLAN tag. If the frame as received by the PE from the attachment circuit does not have a service-delimiting VLAN tag, the PE must prepend the frame with a dummy VLAN tag before sending the frame on the PW. This is the default operating mode. This is the only REQUIRED mode.

In both modes, non-service-delimiting tags are passed transparently across the PW as part of the payload. It should be noted that a single Ethernet packet may contain more than one tag. At most, one of these tags may be service-delimiting. In any case, the NSP function may only inspect the outermost tag for the purpose of adapting the Ethernet frame to the pseudowire.

In both modes, the service-delimiting tag values have only local significance, i.e., are meaningful only at a particular PE-CE interface. When tagged mode is used, the PE that receives a frame from the PW may rewrite the tag value, or may strip the tag entirely, or may leave the tag unchanged, depending on its configuration. When raw mode is used, the PE that receives a frame may or may not need to add a service-delimiting tag before transmitting the frame on the attachment circuit; however, it MUST not rewrite or remove any tags that are already present.

The following table illustrates the operations that might be performed at input from the attachment circuit:

Tag->	service delimiting	non service delimiting
Raw Mode	1st VLAN Tag Removed	no operation performed
Tagged Mode	NO OP or Tag Added	Tag Added

4.4.2. MTU Management on the PE/CE Links

The Ethernet PW MUST NOT be enabled unless it is known that the MTUs of the CE-PE links are the same at both ends of the PW. If an egress router receives an encapsulated layer 2 PDU whose payload length (i.e., the length of the PDU itself without any of the encapsulation headers) exceeds the MTU of the destination layer 2 interface, the PDU MUST be dropped.

4.4.3. Frame Ordering

In general, applications running over Ethernet do not require strict frame ordering. However, the IEEE definition of 802.3 [802.3] requires that frames from the same conversation in the context of link aggregation (clause 43) are delivered in sequence. Moreover, the PSN cannot (in the general case) be assumed to provide or to guarantee frame ordering. An Ethernet PW can, through use of the control word, provide strict frame ordering. If this option is enabled, any frames that get misordered by the PSN will be dropped or reordered by the receiving PW endpoint. If strict frame ordering is a requirement for a particular PW, this option MUST be enabled.

4.4.4. Frame Error Processing

An encapsulated Ethernet frame traversing a pseudowire may be dropped, corrupted, or delivered out-of-order. As described in [PWE3-REQ], frame loss, corruption, and out-of-order delivery are considered to be a "generalized bit error" of the pseudowire. PW frames that are corrupted will be detected at the PSN layer and dropped.

At the ingress of the PW, the native Ethernet frame error processing mechanisms MUST be enabled. Therefore, if a PE device receives an Ethernet frame containing hardware-level Cyclic Redundancy Check (CRC) errors, framing errors, or a runt condition, the frame MUST be discarded on input. Note that defining this processing is part of the NSP function and is outside the scope of this document.

4.4.5. IEEE 802.3x Flow Control Interworking

In a standard Ethernet network, the flow control mechanism is optional and typically configured between the two nodes on a point-to-point link (e.g., between the CE and the PE). IEEE 802.3x PAUSE frames MUST NOT be carried across the PW. See Appendix A for notes on CE-PE flow control.

4.5. Management

The Ethernet PW management model follows the general PW management model defined in [RFC3985] and [PWE3-MIB]. Many common PW management facilities are provided here, with no additional Ethernet specifics necessary. Ethernet-specific parameters are defined in an additional MIB module, [PW-MIB].

4.6. The Control Word

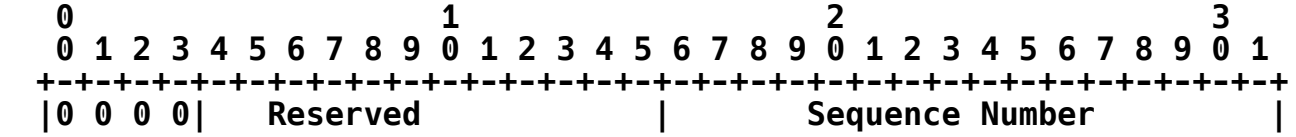
The control word defined in this section is based on the Generic PW MPLS Control Word as defined in [PWE3-CW]. It provides the ability to sequence individual frames on the PW, avoidance of equal-cost multiple-path load-balancing (ECMP) [RFC2992], and Operations and Management (OAM) mechanisms including VCCV [VCCV].

[PWE3-CW] states, "If a PW is sensitive to packet misordering and is being carried over an MPLS PSN that uses the contents of the MPLS payload to select the ECMP path, it MUST employ a mechanism which prevents packet misordering." This is necessary because ECMP implementations may examine the first nibble after the MPLS label stack to determine whether the labelled packet is IP or not. Thus, if the source MAC address of an Ethernet frame carried over the PW without a control word present begins with 0x4 or 0x6, it could be mistaken for an IPv4 or IPv6 packet. This could, depending on the configuration and topology of the MPLS network, lead to a situation where all packets for a given PW do not follow the same path. This may increase out-of-order frames on a given PW, or cause OAM packets to follow a different path than actual traffic (see Section 4.4.3, "Frame Ordering").

The features that the control word provides may not be needed for a given Ethernet PW. For example, ECMP may not be present or active on a given MPLS network, strict frame sequencing may not be required, etc. If this is the case, the control word provides little value and is therefore optional. Early Ethernet PW implementations have been deployed that do not include a control word or the ability to process one if present. To aid in backwards compatibility, future implementations MUST be able to send and receive frames without the control word present.

In all cases, the egress PE MUST be aware of whether the ingress PE will send a control word over a specific PW. This may be achieved by configuration of the PEs, or by signaling, as defined in [PWE3-CTRL].

The control word is defined as follows:



In the above diagram, the first 4 bits MUST be set to 0 to indicate PW data. The rest of the first 16 bits are reserved for future use. They MUST be set to 0 when transmitting, and MUST be ignored upon receipt.

The next 16 bits provide a sequence number that can be used to guarantee ordered frame delivery. The processing of the sequence number field is OPTIONAL.

The sequence number space is a 16-bit, unsigned circular space. The sequence number value 0 is used to indicate that the sequence number check algorithm is not used. The sequence number processing algorithm is found in [PWE3-CW].

4.7. QoS Considerations

The ingress PE MAY consider the user priority (PRI) field [802.1Q] of the VLAN tag header when determining the value to be placed in a QoS field of the encapsulating protocol (e.g., the EXP fields of the MPLS label stack). In a similar way, the egress PE MAY consider the QoS field of the encapsulating protocol (e.g., the EXP fields of the MPLS label stack) when queuing the frame for transmission towards the CE.

A PE MUST support the ability to carry the Ethernet PW as a best-effort service over the MPLS PSN. PRI bits are kept transparent between PE devices, regardless of the QoS support of the PSN.

If an 802.1Q VLAN field is added at the PE, a default PRI setting of zero MUST be supported, a configured default value is recommended, or the value may be mapped from the QoS field of the PSN, as referred to above.

A PE may support additional QoS support by means of one or more of the following methods:

i. One class of service (CoS) per PW End Service (PWES), mapped to a single CoS PW at the PSN.
ii. Multiple CoS per PWES mapped to a single PW with multiple CoS at the PSN.
iii. Multiple CoS per PWES mapped to multiple PWs at the PSN.

Examples of the cases above and details of the service mapping considerations are described in Appendix B.

The PW guaranteed rate at the MPLS PSN level is PW service provider policy based on agreement with the customer, and may be different from the Ethernet physical port rate.

5. Security Considerations

The Ethernet pseudowire type is subject to all of the general security considerations discussed in [RFC3985] and [PWE3-CTRL].

The Ethernet pseudowire is transported on an MPLS PSN; therefore, the security of the pseudowire itself will only be as good as the security of the MPLS PSN. The MPLS PSN can be secured by various methods, as described in [MPLS-ARCH].

Security achieved by access control of MAC addresses is out of the scope of this document. Additional security requirements related to the use of PW in a switching (virtual bridging) environment are not discussed here as they are not within the scope of this document.

6. PSN MTU Requirements

The MPLS PSN MUST be configured with an MTU that is large enough to transport a maximum-sized Ethernet frame that has been encapsulated with a control word, a pseudowire demultiplexer, and a tunnel encapsulation. With MPLS used as the tunneling protocol, for example, this is likely to be 8 or more bytes greater than the largest frame size. The methodology described in [FRAG] MAY be used to fragment encapsulated frames that exceed the PSN MTU. However, if [FRAG] is not used and if the ingress router determines that an encapsulated layer 2 PDU exceeds the MTU of the PSN tunnel through which it must be sent, the PDU MUST be dropped.

7. Normative References

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

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

[PWE3-CTRL] Martini, L., El-Aawar, N., Heron, G., Rosen, E., Tappan, D., and T. Smith, "Pseudowire Setup and Maintenance using the Label Distribution Protocol (LDP)", RFC 4447, April 2006.

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

[802.3] IEEE802.3-2005, ISO/IEC 8802-3: 2000 (E), "IEEE Standard for Information technology -- Telecommunications and information exchange between systems -- Local and metropolitan
area networks -- Specific requirements -- Part 3: Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications", 2005.

[802.1Q] ANSI/IEEE Standard 802.1Q-2005, "IEEE Standards for Local and Metropolitan Area Networks: Virtual Bridged Local Area Networks", 2005.

[PDU] IEEE Std 802.3, 1998 Edition, "Part 3: Carrier sense multiple access with collision detection (CSMA/CD) access method and physical layer specifications" figure 3.1, 1998

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

8. Informative References

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

[PW-MIB] Zelig, D. and T. Nadeau, "Ethernet Pseudo Wire (PW) Management Information Base", Work in Progress, February 2006.

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

[PWE3-MIB] Zelig, D., Ed. and T. Nadeau, Ed., "Pseudo Wire (PW) Management Information Base", Work in Progress, February 2006.

[LDP] Andersson, L., Doolan, P., Feldman, N., Fredette, A., and B. Thomas, "LDP Specification", RFC 3036, January 2001.

[FRAG] Malis, A. and W. Townsley, "PWE3 Fragmentation and Reassembly", Work in Progress, February 2005.

[FCS] Malis, A., Allan, D., and N. Del Regno, "PWE3 Frame Check Sequence Retention", Work in Progress, September 2005.

[VCCV] Nadeau, T., Ed. and R. Aggarwal, Ed., "Pseudo Wire Virtual Circuit Connectivity Verification (VCCV)", Work in Progress, August 2005.

[RFC2992] Hopps, C., "Analysis of an Equal-Cost Multi-Path Algorithm", RFC 2992, November 2000.

[RFC4026] Andersson, L. and T. Madsen, "Provider Provisioned Virtual Private Network (VPN) Terminology", RFC 4026, March 2005.

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

9. Significant Contributors

Andrew G. Malis
Tellabs
90 Rio Robles Dr.
San Jose, CA 95134

EMail: Andy.Malis@tellabs.com

Dan Tappan
Cisco Systems, Inc.
1414 Massachusetts Avenue
Boxborough, MA 01719

EMail: tappan@cisco.com

Steve Vogelsang
ECI Telecom
Omega Corporate Center
1300 Omega Drive
Pittsburgh, PA 15205

EMail: stephen.vogelsang@ecitele.com

Vinai Sirkay
Reliance Infocomm
Dhirubai Ambani Knowledge City
Navi Mumbai 400 709
India

EMail: vinai@sirkay.com

Vasile Radoaca
Nortel Networks
600 Technology Park
Billerica MA 01821

EMail: vasile@nortelnetworks.com

Chris Liljenstolpe
Alcatel
11600 Sallie Mae Dr.
9th Floor
Reston, VA 20193

EMail: chris.liljenstolpe@alcatel.com

Kireeti Kompella
Juniper Networks
1194 N. Mathilda Ave
Sunnyvale, CA 94089

EMail: kireeti@juniper.net

Tricci So
Nortel Networks 3500 Carling Ave.,
Nepean, Ontario,
Canada, K2H 8E9.

EMail: tso@nortelnetworks.com

XiPeng Xiao
Riverstone Networks
5200 Great America Parkway
Santa Clara, CA 95054

EMail: xxiao@riverstonenet.com

Christopher O. Flores
T-Systems
10700 Parkridge Boulevard
Reston, VA 20191
USA

EMail: christopher.flores@usa.telekom.de

David Zelig
Corrigent Systems
126, Yigal Alon St.
Tel Aviv, ISRAEL

EMail: davidz@corrigent.com

Raj Sharma
Luminous Networks, Inc.
10460 Bubb Road
Cupertino, CA 95014

EMail: raj@luminous.com

Nick Tingle
TiMetra Networks
274 Ferguson Drive
Mountain View, CA 94043

EMail: nick@timetra.com

Sunil Khandekar
TiMetra Networks
274 Ferguson Drive
Mountain View, CA 94043

EMail: sunil@timetra.com

Loa Andersson
TLA-group

EMail: loa@pi.se

Appendix A. Interoperability Guidelines

A.1. Configuration Options

The following is a list of the configuration options for a point-to-point Ethernet PW based on the reference points of Figure 3:

Service and Encap on A	Encap on C	Operation at B ingress/egress	Remarks
1) Raw	Raw - Same as A		
2) Tag1	Tag2	Optional change of VLAN value	VLAN can be 0-4095 Change allowed in both directions
3) No Tag	Tag	Add/remove Tag field	Tag can be 0-4095 (note i)
4) Tag	No Tag	Remove/add Tag field	(note ii)

Figure 4: Configuration Options

Allowed combinations:

Raw and other services are not allowed on the same NSP virtual port (A). All other combinations are allowed, except that conflicting VLANs on (A) are not allowed. Note that in most point-to-point PW applications the NSP virtual port is the same entity as the physical port.

Notes:

i. Mode #3 MAY be limited to adding VLAN NULL only, since change of VLAN or association to specific VLAN can be done at the PW CE-bound side.

ii. Mode #4 exists in layer 2 switches, but is not recommended when operating with PW since it may not preserve the user's PRI bits. If there is a need to remove the VLAN tag (for TLS at the other end of the PW), it is recommended to use mode #2 with tag2=0 (NULL VLAN) on the PW and use mode #3 at the other end of the PW.

A.2. IEEE 802.3x Flow Control Considerations

If the receiving node becomes congested, it can send a special frame, called the PAUSE frame, to the source node at the opposite end of the connection. The implementation MUST provide a mechanism for terminating PAUSE frames locally (i.e., at the local PE). It MUST operate as follows: PAUSE frames received on a local Ethernet port SHOULD cause the PE device to buffer, or to discard, further Ethernet frames for that port until the PAUSE condition is cleared. Optionally, the PE MAY simply discard PAUSE frames.

If the PE device wishes to pause data received on a local Ethernet port (perhaps because its own buffers are filling up or because it has received notification of congestion within the PSN), then it MAY issue a PAUSE frame on the local Ethernet port, but MUST clear this condition when willing to receive more data.

Appendix B. QoS Details

Section 4.7, "QoS Considerations", describes various modes for supporting PW QOS over the PSN. Examples of the above for a point-to-point VLAN service are:

- The classification to the PW is based on VLAN field, but the user PRI bits are mapped to different CoS markings (and network behavior) at the PW level. An example of this is a PW mapped to an E-LSP in an MPLS network.

- The classification to the PW is based on VLAN field and the PRI bits, and frames with different PRI bits are mapped to different PWs. An example is to map a PWES to different L-LSPs in MPLS PSN in order to support multiple CoS over an L-LSP-capable network, or to map a PWES to multiple L2TPv3 sessions [L2TPv3].

  The specific value to be assigned at the PSN for various CoS is out of the scope of this document.

B.1. Adaptation of 802.1Q CoS to PSN CoS

It is not required that the PSN will have the same CoS definition of CoS as defined in [802.1Q], and the mapping of 802.1Q CoS to PSN CoS is application specific and depends on the agreement between the customer and the PW provider. However, the following principles adopted from 802.1Q, Table 8-2, MUST be met when applying the set of PSN CoS based on user's PRI bits.

	#of available classes of service							
User Priority	1	2	3	4	5	6	7	8
0 Best Effort (Default)	0	0	0	1	1	1	1	2
1 Background	0	0	0	0	0	0	0	0
2 Spare	0	0	0	0	0	0	0	1
3 Excellent Effort	0	0	0	1	1	2	2	3
4 Controlled Load	0	1	1	2	2	3	3	4
5 Interactive Multimedia	0	1	1	2	3	4	4	5
6 Interactive Voice	0	1	2	3	4	5	5	6
7 Network Control	0	1	2	3	4	5	6	7

Figure 5: IEEE 802.1Q CoS Mapping

B.2. Drop Precedence

The 802.1P standard does not support drop precedence; therefore, from the PW PE-bound point of view there is no mapping required. It is, however, possible to mark different drop precedence for different PW frames based on the operator policy and required network behavior. This functionality is not discussed further here.

PSN QoS support and signaling of QoS are out of the scope of this document.

Authors' Addresses

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

EMail: lmartini@cisco.com

Nasser El-Aawar
Level 3 Communications, LLC.
1025 Eldorado Blvd.
Broomfield, CO, 80021

EMail: nna@level3.net

Giles Heron
Tellabs
Abbey Place
24-28 Easton Street
High Wycombe
Bucks
HP11 1NT
UK

EMail: giles.heron@tellabs.com

Eric C. Rosen
Cisco Systems, Inc.
1414 Massachusetts Avenue
Boxborough, MA 01719

EMail: erosen@cisco.com

Full Copyright Statement

Intellectual Property

Acknowledgement

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA).