

Network Working Group	S. Bryant, Ed.	
Internet-Draft	L. Martini	
Intended status: BCP	G. Swallow	
Expires: January 9, 2011	Cisco Systems	
	A. Malis	
	Verizon Communications	
	July 8, 2010	

[TOC](#)

Packet Pseudowire Encapsulation over an MPLS PSN draft-bryant-pwe3-packet-pw-04.txt

Abstract

This document describes a pseudowire mechanism that is used to transport a packet service over an MPLS PSN is the case where the client LSR and the server PE are co-resident in the same equipment. This pseudowire mechanism may be used to carry all of the required layer 2 and layer 3 protocols between the pair of client LSRs.

Requirements Language

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

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on January 9, 2011.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1.	Introduction
2.	Network Reference Model
3.	Client Network Layer Model
4.	Forwarding Model
5.	Design Considerations
6.	Encapsulation Approaches Considered
6.1.	A Protocol Identifier in the Control Word
6.2.	PID Label
6.3.	Parallel PWs
6.4.	Virtual Ethernet
6.5.	Recommended Encapsulation
7.	Packet PW Encapsulation
8.	Ethernet Functional Restrictions
9.	Congestion Considerations
10.	Security Considerations
11.	IANA Considerations
12.	Acknowledgements
13.	References
13.1.	Normative References
13.2.	Informative References
§	Authors' Addresses

1. Introduction

TOC

There is a need to provide a method of carrying a packet service over an MPLS PSN in a way that provides isolation between the two networks. The server MPLS network may be an MPLS network or a network conforming to the MPLS-TP [\[RFC5317\]](#) (Bryant, S. and L. Andersson, "Joint Working Team (JWT) Report on MPLS Architectural Considerations for a Transport

[Profile," February 2009.](#)). The client may also be either a MPLS network or a network conforming to the MPLS-TP. Considerations regarding the use of an MPLS network as a server for an MPLS-TP network are outside the scope of this document.

Where the client equipment is connected to the server equipment via physical interface, the same data-link type MUST be used to attach the clients to the Provider Edge equipments (PE)s, and a pseudowire (PW) of the same type as the data-link MUST be used [\[RFC3985\] \(Bryant, S. and P. Pate, "Pseudo Wire Emulation Edge-to-Edge \(PWE3\) Architecture," March 2005.\)](#). The reason that inter-working between different physical and data-link attachment types is specifically disallowed in the pseudowire architecture is because this is a complex task and not a simple bit-mapping exercise. The inter-working is not limited to the physical and data-link interfaces and the state-machines. It also requires a compatible approach to the formation of the adjacencies between attached client network equipment. As an example the reader should consider the differences between router adjacency formation on a point to point link compared to a multi-point to multi-point interface (e.g. Ethernet).

A further consideration is that two adjacent MPLS LSRs do not simply exchange MPLS packets. They exchange IP packets for adjacency formation, control, routing, label exchange, management and monitoring purposes. In addition they may exchange data-link packets as part of routing (e.g. IS-IS hellos and IS-IS LSPs) and for OAM purposes such as Link Layer Discovery protocol [IEEE standard 802.1AB-2009]. Thus the two clients require an attachment mechanism that can be used to multiplex a number of protocols. In addition it is essential to the correct operation of the network layer that all of these protocols fate share.

Where the client LSR and server PE is co-located in the same equipment, the data-link layer can be simplified to a simple protocol identifier (PID) that is used to multiplex the various data-link types onto a pseudowire. This is the method that described in this document.

2. Network Reference Model

[TOC](#)

The network reference model for the packet pseudowire operating in an MPLS network is shown in [Figure 1](#). This is an extension of Figure 3 "Pre-processing within the PWE3 Network Reference Model" from [\[RFC3985\] \(Bryant, S. and P. Pate, "Pseudo Wire Emulation Edge-to-Edge \(PWE3\) Architecture," March 2005.\)](#).

appropriate configuration of the embedded client network layer equipment (e.g. the embedded client LSR). Where the client was using the packet PW as part of a traffic engineered path, it is up to the operator of the client network to ensure that the server layer operator provides the necessary service level agreement.

4. Forwarding Model

[TOC](#)

The packet PW forwarding model is illustrated in [Figure 2 \(Packet PW Forwarding Model\)](#). The forwarding operation can be likened to a virtual private network (VPN), in which a forwarding decision is first taken at the client layer, an encapsulation is applied and then a second forwarding decision is taken at the server layer.

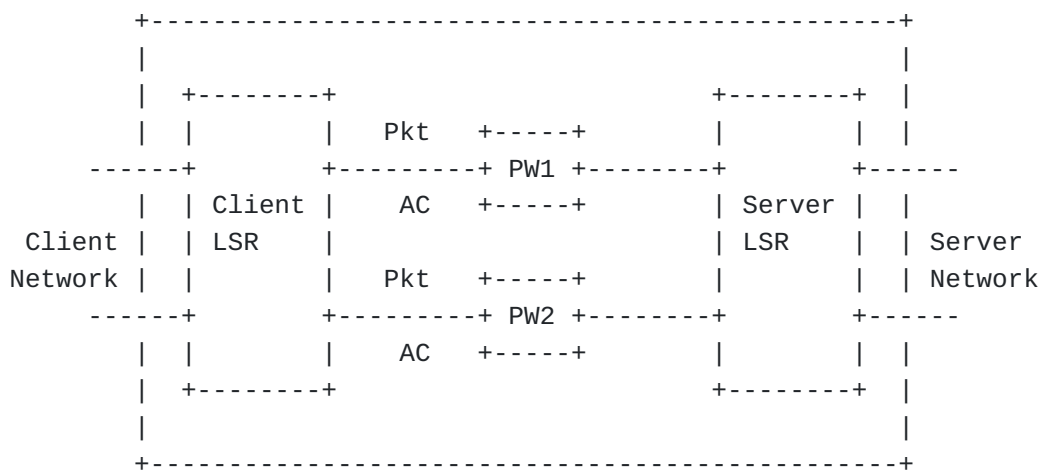


Figure 2: Packet PW Forwarding Model

A packet PW PE comprises three components, the client LSR, PW processor and a server LSR. Note that [\[RFC3985\] \(Bryant, S. and P. Pate, "Pseudo Wire Emulation Edge-to-Edge \(PWE3\) Architecture," March 2005.\)](#) does not formally indicate the presence of the server LSR because it does not concern itself with the server layer. However it is useful in this document to recognise that the server LSR exists.

It may be useful to first recall the operation of a layer two PW such as an Ethernet PW [\[RFC4448\] \(Martini, L., Rosen, E., El-Aawar, N., and G. Heron, "Encapsulation Methods for Transport of Ethernet over MPLS Networks," April 2006.\)](#) within this model. The client LSR is not present and packets arrive directly on the attachment circuit (AC)

which is part of the client network. The PW undertakes any header processing, if configured to do so, it then pushes the PW control word (CW), and finally pushes the PW label. The PW function then passes the packet to the LSR function which pushes the label needed to reach the egress PE and forwards the packet to the next hop in the server network. At the egress PE, the packet typically arrives with the PW label at top of stack, the packet is thus directed to the correct PW instance. The PW instance performs any required reconstruction using, if necessary, the CW and the packet is sent directly to the attachment circuit.

Now let us consider the case client layer MPLS traffic being carried over a packet PW. An LSR belonging to the client layer is embedded within the PE equipment. This is a type of native service processing element [\[RFC3985\] \(Bryant, S. and P. Pate, "Pseudo Wire Emulation Edge-to-Edge \(PWE3\) Architecture," March 2005.\)](#). This LSR determines the next hop in the client layer, and pushes the label needed by the next hop in the client layer. It then passes the packet to the correct PW instance indicating the packet protocol type. If the PW is configured to require a CW this is pushed. The PW instance then examines the protocol type and pushes a label that identifies the protocol type to the egress PE. The PW instance then proceeds as it would for a layer two PW, by pushing the PW label and then handing the packet to the server layer LSR for delivery. At egress, the packet again arrives with the PW label at the top of stack which causes the packet to be passed to the correct PW instance. This PW instance knows that the PW type is a packet PW, and hence that it needs to interpret the next label as a protocol type identifier. If necessary the CW is then popped and processed. The packet is then passed to the egress client LSR together with information that identifies the packet protocol type. The egress client LSR then forwards the packet in the normal manner for a protocol of that type.

Note that although the description above is written in terms of the behaviour of an MPLS LSR, the processing model would be similar for an IP packet, or indeed any other protocol type.

Note that the semantics of the PW between the client LSRs is a point to point link.

5. Design Considerations

[TOC](#)

A number of approaches to the design of a packet pseudowire (PW) have been investigated and have been described at the IETF. This section discusses the approaches that were analysed and the technical issues that the authors took into consideration in arriving at the proposed approach.

In a typical network there are usually no more than four network layer protocols that need to be supported: IPv4, IPv6, MPLS and CLNS although

any solution needs to be scalable to a larger number of protocols. The approaches considered in this document all satisfy this minimum requirement, but vary in their ability to support larger numbers of network layer protocols.

Additionally it is beneficial if the complete set of protocols carried over the network between in support of a set of CE peers fate share. It is additionally beneficial if a single OAM session can be used to monitor the behaviour of this complete set. During the investigation various views were expressed as to where on the scale from absolutely required to "nice to have" these benefits lay, but in the end they were not a factor in reaching our conclusion.

6. Encapsulation Approaches Considered

[TOC](#)

There are four candidate approaches that have been analysed:

1. A protocol identifier (PID) in the PW Control Word (CW)
 2. A PID label
 3. Parallel PWs - one per protocol.
 4. Virtual Ethernet
-

6.1. A Protocol Identifier in the Control Word

[TOC](#)

This is the approach that we proposed in draft 0 of this document . The proposal was that a Protocol Identifier (PID) would included in the PW control word (CW), by appending it to the generic control word [\[RFC5385\] \(Touch, J., "Version 2.0 Microsoft Word Template for Creating Internet Drafts and RFCs," February 2010.\)](#) to make a 6 byte CW (the version 0 draft actually included two reserved bytes to provide 32bit alignment, but let us assume that was optimized out). A variant of this is just to use a 2 byte PID without a control word.

This is a simple approach, and is basically a virtual PPP interface without the PPP control protocol. This has a smaller MTU than for example a virtual Ethernet would need, however in forwarding terms it is not as simple as the PID label or multiple PW approaches described next, and may not be deployable on a number of existing hardware platforms.

6.2. PID Label

[TOC](#)

This is the approach that we described in Version 2 of this document. The in this mechanism the PID is indicated by including a label after the PW label that indicates the protocol type as shown in [Figure 3 \(Encapsulation of a pseudowire with a pseudowire load balancing label\)](#).

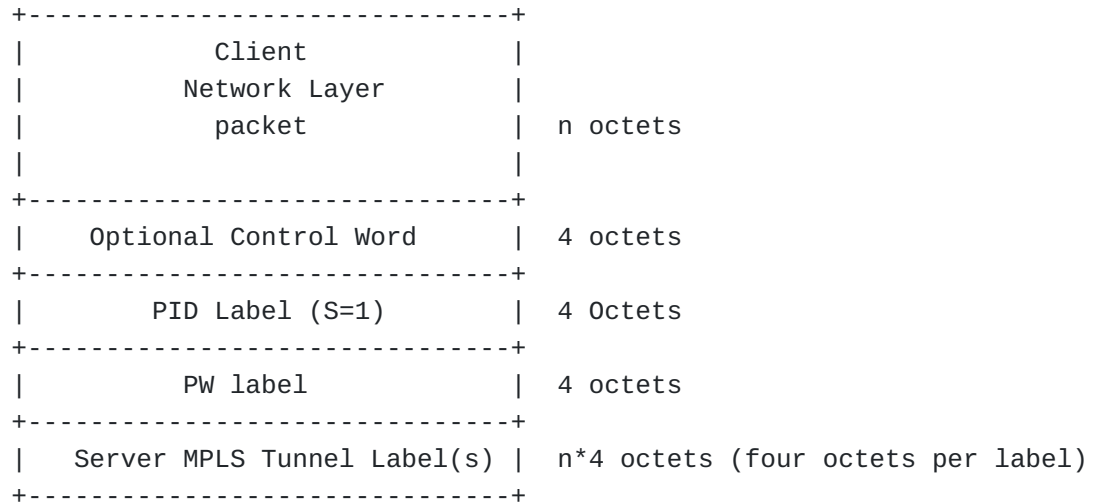


Figure 3: Encapsulation of a pseudowire with a pseudowire load balancing label

In the PID Label approach a new Label Distribution protocol (LDP) Forwarding Equivalence Class (FEC) element is used to signal the mapping between protocol type and the PID label. This approach complies with RFC3031.

A similar approach to PID label is described in Section 3.4.5 of [\[I-D.ietf-mpls-tp-framework\] \(Bocci, M., Bryant, S., Frost, D., Levrau, L., and L. Berger, "A Framework for MPLS in Transport Networks," May 2010.\)](#). In this case when the client is a network layer packet service such as IP or MPLS, a service label and demultiplexer label (which may be combined) is used to provide the necessary identifications needed to carry this traffic over an LSP.

The authors surveyed the hardware designs produced by a number of companies across the industry and concluded that whilst the approach complies with the MPLS architecture, it may conflict with a number of designer's interpretation of the existing MPLS architecture. This led to concerns that the approach may result in unexpected difficulties in the future. Specifically there is an assumption in many designs that a forwarding decision should be made on the basis of a single label. Whilst the approach is attractive, it cannot be supported by many

commodity chip sets and this would require new hardware which would increase the cost of deployment and delay the introduction of a packet PW service.

6.3. Parallel PWs

[TOC](#)

In this approach one PW is constructed for each protocol type that must be carried between the PEs. Thus a complete packet PW would therefore consist of a bundle of PWs. This model would be very simple and efficient from a forwarding point of view. The number of parallel PWs required would normally be relatively small. In a typical network there are usually no more than four network layer protocols that need to be supported: IPv4, IPv6, MPLS and CLNS although any solution needs to be scalable to a larger number of protocols.

There are a number of serious downsides with this approach:

1. From an operational point of view the lack of fate sharing between the protocol types can lead to complex faults which are difficult to diagnose.
2. There is an undesirable trade off in the OAM related to the first point. Either we would have to run an OAM on each PW and bind them together which lead to significant protocol and software complexity and does not scale well. Alternatively we would need to run a single OAM session on one of the PWs as a proxy for the others and then diagnose any more complex failure on a case by case basis. To some extent the issue of fate sharing between protocols in the bundle (for example the assumed fate sharing between CLNS and IP in IS-IS) can be mitigated through the use of BFD.
3. The need to configure, manage and synchronize the behaviour of a group of PWs as if they were a single PW leads to an increase in control plane complexity.

The Parallel PW mechanism is therefore an approach which simplifies the forwarding plane, but only at a cost of a considerable increase in other aspects of the design and in particular operation of the PW.

6.4. Virtual Ethernet

[TOC](#)

Using a virtual Ethernet to provide a packet PW would require PEs to include a virtual (internal) Ethernet interface and then to use an Ethernet PW [\[RFC4448\]](#) (Martini, L., Rosen, E., El-Aawar, N., and G.

[Heron, "Encapsulation Methods for Transport of Ethernet over MPLS Networks," April 2006.](#)) to carry the user traffic. This is conceptually simple and can be implemented today without any further standards action, although there are a number of applicability considerations that it is useful to draw to the attention of the community. Conceptually this is a simple approach and some deployed equipments can already do this. However the requirement to run a complete Ethernet adjacency lead us to conclude that there was a need to identify a simpler approach. The packets encapsulated in an Ethernet header have a larger MTU than the other approaches, although this is not considered to be an issue on the networks needing to carry packet PWs. The virtual Ethernet mechanism was the first approach that the authors considered, before the merits of the other approaches appeared to make them more attractive. As we shall see below however, the other approaches were not without issues and it appears that the virtual Ethernet is preferred approach to providing a packet PW.

6.5. Recommended Encapsulation

[TOC](#)

The operational complexity and the breaking of fate sharing assumptions associated with the parallel PW approach would suggest that this is not an approach that should be further pursued.

The PID Label approach gives rise to the concerns that it will break implicit behavioural and label stack size assumptions in many implementations. Whilst those assumptions may be addressed with new hardware this would delay the introduction of the technology to the point where it was unlikely to gain acceptance in competition with an approach that needed no new protocol design and is already supportable on many existing hardware platforms.

The PID in the CW leads to the most compact protocol stack, is simple and requires minimal protocol work. However it is a new forwarding design, and apart from the issue of the larger packet header and the simpler adjacency formation offers no advantage over the virtual Ethernet.

The above considerations bring us back to the virtual Ethernet, which is a well known protocol stack, with a well known (internal) client interface. It is already implemented in many hardware platforms and is therefore readily deployable. The authors conclude that having considered a number of initially promising alternatives, the simplicity and existing hardware make the virtual Ethernet approach to the packet PW the most attractive solution.

[TOC](#)

7. Packet PW Encapsulation

The client network work layer packet encapsulation into a packet PW is shown in [Figure 4](#).

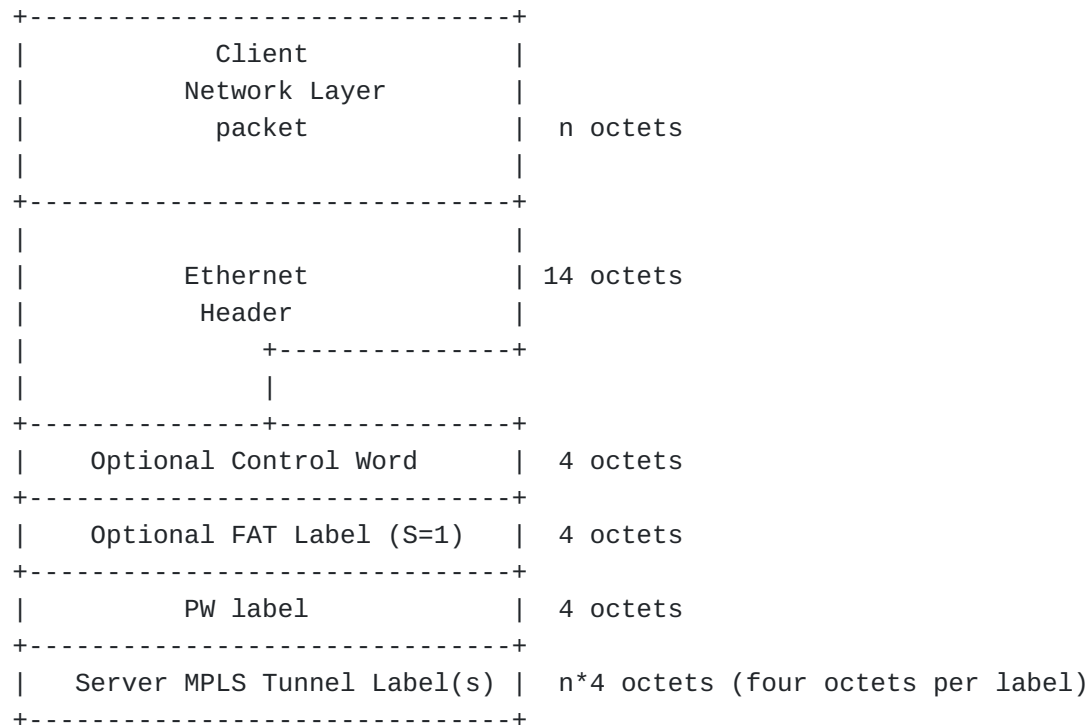


Figure 4

This conforms to the PW protocols stack as defined in [\[RFC4448\]](#) (Martini, L., Rosen, E., El-Aawar, N., and G. Heron, "Encapsulation Methods for Transport of Ethernet over MPLS Networks," April 2006.) and [\[I-D.ietf-pwe3-fat-pw\]](#) (Bryant, S., Filsfils, C., Drafz, U., Kompella, V., Regan, J., and S. Amante, "Flow Aware Transport of Pseudowires over an MPLS PSN," January 2010.).

This is unremarkable except to note that the stack does not retain 32 bit alignment between the virtual Ethernet header and the PW optional control word (or the PW label when the optional components are not present in the PW header). This loss of 32 bit of alignment is necessary to preserve backwards compatibility with the Ethernet PW design [\[RFC4448\]](#) (Martini, L., Rosen, E., El-Aawar, N., and G. Heron, "Encapsulation Methods for Transport of Ethernet over MPLS Networks," April 2006.)

Considerations concerning the allocation of a suitable Ethernet address the virtual Ethernet will be discussed in a future version of this document.

8. Ethernet Functional Restrictions

[TOC](#)

The use of Ethernet as the encapsulation mechanism for traffic between the server LSRs is a convenience based on the widespread availability of existing hardware. In this application there is no requirement for any Ethernet feature other than its protocol multiplexing capability. Thus, for example, the Ethernet OAM is NOT REQUIRED.

The use and applicability of Ethernet VLANs, 802.1p, and 802.1Q between PEs will be discussed in a future revision.

Point to multipoint and multipoint to multipoint operation of the virtual Ethernet is not supported.

9. Congestion Considerations

[TOC](#)

A packet pseudowire is normally used to carry IP, MPLS and their associated support protocols over an MPLS network. There are no congestion considerations beyond those that ordinarily apply to an IP or MPLS network. Where the packet protocol being carried is not IP or MPLS and the traffic volumes are greater than that ordinarily associated with the support protocols in an IP or MPLS network, the congestion considerations being developed for PWs apply [\[RFC3985\]](#) (Bryant, S. and P. Pate, "Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture," March 2005.), [\[RFC5659\]](#) (Bocci, M. and S. Bryant, "An Architecture for Multi-Segment Pseudowire Emulation Edge-to-Edge," October 2009.).

10. Security Considerations

[TOC](#)

The virtual Ethernet approach to packet PW introduces no new security risks. A more detailed discussion of pseudowire security is given in [\[RFC3985\]](#) (Bryant, S. and P. Pate, "Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture," March 2005.), [\[RFC4447\]](#) (Martini, L., Rosen, E., El-Aawar, N., Smith, T., and G. Heron, "Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP)," April 2006.) and [\[RFC3916\]](#) (Xiao, X., McPherson, D., and P. Pate, "Requirements for Pseudo-Wire Emulation Edge-to-Edge (PWE3)," September 2004.).

[TOC](#)

11. IANA Considerations

This section may be removed on publication.

There are no IANA action required by the publication of this document.

12. Acknowledgements

[TOC](#)

The authors acknowledge the contribution make by Sami Boutros, Giles Herron, Siva Sivabalan and David Ward to this document.

13. References

[TOC](#)

13.1. Normative References

[TOC](#)

[RFC2119]	Bradner, S. , " Key words for use in RFCs to Indicate Requirement Levels ," BCP 14, RFC 2119, March 1997 (TXT , HTML , XML).
[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 (TXT).
[RFC4448]	Martini, L., Rosen, E., El-Aawar, N., and G. Heron, " Encapsulation Methods for Transport of Ethernet over MPLS Networks ," RFC 4448, April 2006 (TXT).

13.2. Informative References

[TOC](#)

[I-D.ietf-mpls-tp-framework]	Bocci, M., Bryant, S., Frost, D., Levrau, L., and L. Berger, " A Framework for MPLS in Transport Networks ," draft-ietf-mpls-tp-framework-12 (work in progress), May 2010 (TXT).
[I-D.ietf-pwe3-fat-pw]	Bryant, S., Filsfils, C., Drafz, U., Kompella, V., Regan, J., and S. Amante, " Flow Aware Transport of Pseudowires over an MPLS PSN ," draft-ietf-pwe3-fat-pw-03 (work in progress), January 2010 (TXT).
[RFC3916]	Xiao, X., McPherson, D., and P. Pate, " Requirements for Pseudo-Wire Emulation Edge-to-Edge (PWE3) ," RFC 3916, September 2004 (TXT).
[RFC3985]	

	Bryant, S. and P. Pate, " Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture ," RFC 3985, March 2005 (TXT).
[RFC5317]	Bryant, S. and L. Andersson, " Joint Working Team (JWT) Report on MPLS Architectural Considerations for a Transport Profile ," RFC 5317, February 2009 (TXT , PDF).
[RFC5385]	Touch, J., " Version 2.0 Microsoft Word Template for Creating Internet Drafts and RFCs ," RFC 5385, February 2010 (TXT).
[RFC5659]	Bocci, M. and S. Bryant, " An Architecture for Multi-Segment Pseudowire Emulation Edge-to-Edge ," RFC 5659, October 2009 (TXT).

Authors' Addresses

[TOC](#)

	Stewart Bryant (editor)
	Cisco Systems
	250, Longwater, Green Park,
	Reading, Berks RG2 6GB
	UK
Email:	stbryant@cisco.com
	Luca Martini
	Cisco Systems
	9155 East Nichols Avenue, Suite 400
	Englewood, CO 80112
	USA
Email:	lmartini@cisco.com
	George Swallow
	Cisco Systems
	1414 Massachusetts Ave
	Boxborough, MA 01719
	USA
Email:	swallow@cisco.com
URI:	
	Andy Malis
	Verizon Communications
	117 West St.
	Waltham, MA 02451
	USA
Email:	andrew.g.malis@verizon.com