Network Working Group
Internet-Draft
Category: Informational
Expires: December 20, 2014

F. Jounay, Ed.
Orange CH
Y. Kamite, Ed.
NTT Communications
G. Heron
Cisco Systems
M. Bocci
Alcatel-Lucent

June 20, 2014

Requirements and Framework for Point-to-Multipoint Pseudowires over MPLS Packet Switched Networks

draft-ietf-pwe3-p2mp-pw-requirements-10.txt

Abstract

This document presents a set of requirements and a framework for providing a Point-to-Multipoint Pseudowire (PW) over MPLS Packet Switched Networks. The requirements identified in this document are related to architecture, signaling and maintenance aspects of Point-to-Multipoint PW operation. They are proposed as guidelines for the standardization of such mechanisms. Among other potential applications, Point-to-Multipoint PWs can be used to optimize the support of multicast layer 2 services (Virtual Private LAN Service and Virtual Private Multicast Service).

Status of This Memo

Table of Contents

1. Introduction

1.1. Problem Statement

As defined in the pseudowire architecture [RFC3985], a Pseudowire (PW) is a mechanism that emulates the essential attributes of a telecommunications service (such as a T1 leased line or Frame Relay) over an IP or MPLS Packet Switched Network. It provides a single service which is perceived by its user as an unshared link or circuit of the chosen service. A Pseudowire is used to transport layer 1 or layer 2 traffic (e.g. Ethernet, TDM, ATM, and FR) over a layer 3 PSN. Pseudowire Emulation Edge-to-Edge (PWE3) operates "edge to edge" to provide the required connectivity between the two endpoints of the PW.

The Point-to-Multipoint (P2MP) topology described in [I-D.ietf-l2vpn-vpms-frmwk-requirements] and required to provide P2MP Layer2 VPN service can be achieved using one or more P2MP PWs.
The use of PW encapsulation enables P2MP services transporting layer1 or layer2 data. This could be achieved using a set of point to point PWs, with traffic replication on the Provider Edge (PE), but at the cost of bandwidth efficiency, as duplicate traffic would be carried multiple times on shared links.

This document defines the requirements for a Point-to-Multipoint PW (P2MP PW). A P2MP PW is a mechanism that emulates the essential attributes of a P2MP telecommunications service such as a P2MP ATM VC over a Packet Switch Networks.
The required functions of P2MP PWs include encapsulating service-specific Protocol data Units (PDU) arriving at an ingress Attachment Circuit (AC), and carrying them across a tunnel to one or more egress ACs, managing their timing and order, and any other operations required to emulate the behavior and characteristics of the service as faithfully as possible.

1.2. Scope of this document

The document describes the general architecture of P2MP PW with reference model, mentions the notion of data encapsulation, and outlines specific requirements for the setup and maintenance of a P2MP PW. In this document, the requirements focus on the Single-Segment PW model. It is for further study how it should be realized in Multi-Segment PW model. For other aspects of P2MP PW implementation, such as packet processing (section 4) and Faithfulness of Emulated Services (section 7), the document refers to [RFC3916].

1.3. Conventions used in this document

Although this is a requirements specification not a protocol specification, the key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted to apply to protocol solutions designed to meet these requirements as described in [RFC2119] .

2. Definition

2.1. Acronyms

P2P: Point-to-Point
P2MP: Point-to-Multipoint
PW: Pseudowire
PSN: Packet Switched Network
SS-PW: Single-Segment Pseudowire
MS-PW: Multi-Segment Pseudowire

2.2. Terminology

This document uses terminology described in [RFC5659]. It also introduces additional terms needed in the context of P2MP PW.

P2MP PW, (also referred as PW Tree):
Point-to-Multipoint Pseudowire. A PW attached to a source Customer Edge (CE) used to distribute Layer1 or Layer2 traffic to a set of one or more receiver CEs. The P2MP PW is unidirectional (i.e., carrying traffic from Root PE to Leaf PEs), and optionally supports a return path.
P2MP SS-PW:

Point-to-Multipoint Single-Segment Pseudowire. A single segment P2MP PW set up between the Root PE attached to the source CE and the Leaf PEs attached to the receiver CEs. The P2MP SS-PW uses P2MP Label Switched Paths (LSP) as PSN tunnels. The requirements in this document is targeted for SS-PW model. Application of MS-PW (Multi-segment PW) model [RFC5254] is out of scope and left for future work.

Root PE:
P2MP PW Root Provider Edge. The PE attached to the traffic source CE for the P2MP PW via an Attachment Circuit (AC).

Leaf PE:
P2MP PW Leaf Provider Edge. A PE attached to a set of one or more traffic receiver CEs, via ACs. The Leaf PE replicates traffic to the CEs based on its Forwarder function [RFC3985].

P2MP PSN Tunnel:
In the P2MP SS-PW topology, The PSN Tunnel is a general term

indicating a virtual P2MP connection between the Root PE and the Leaf PEs. A P2MP tunnel may potentially carry multiple P2MP PWs inside (aggregation). This document uses terminology from the document describing the MPLS multicast architecture [RFC5332] for MPLS PSN.

3. P2MP PW Requirements

3.1. Reference Model

As per the definition of [RFC3985], a pseudowire (PW) both originates and terminates on the edge of the same packet switched network (PSN). The PW label is unchanged between the originating and terminating Provider Edges (PEs). This is also known as a single-segment pseudowire (SS-PW), as the most fundamental network model of PWE3.

P2MP PW can be defined as Point-to-Multipoint connectivity from a Root PE connected to a traffic source CE to one or more Leaf PEs connected to traffic receiver CEs. It is considered to be an extended architecture of the existing unicast-based SS-PW technology.

Figure 1 describes the P2MP reference model which is derived from [RFC3985] to support P2MP emulated services.

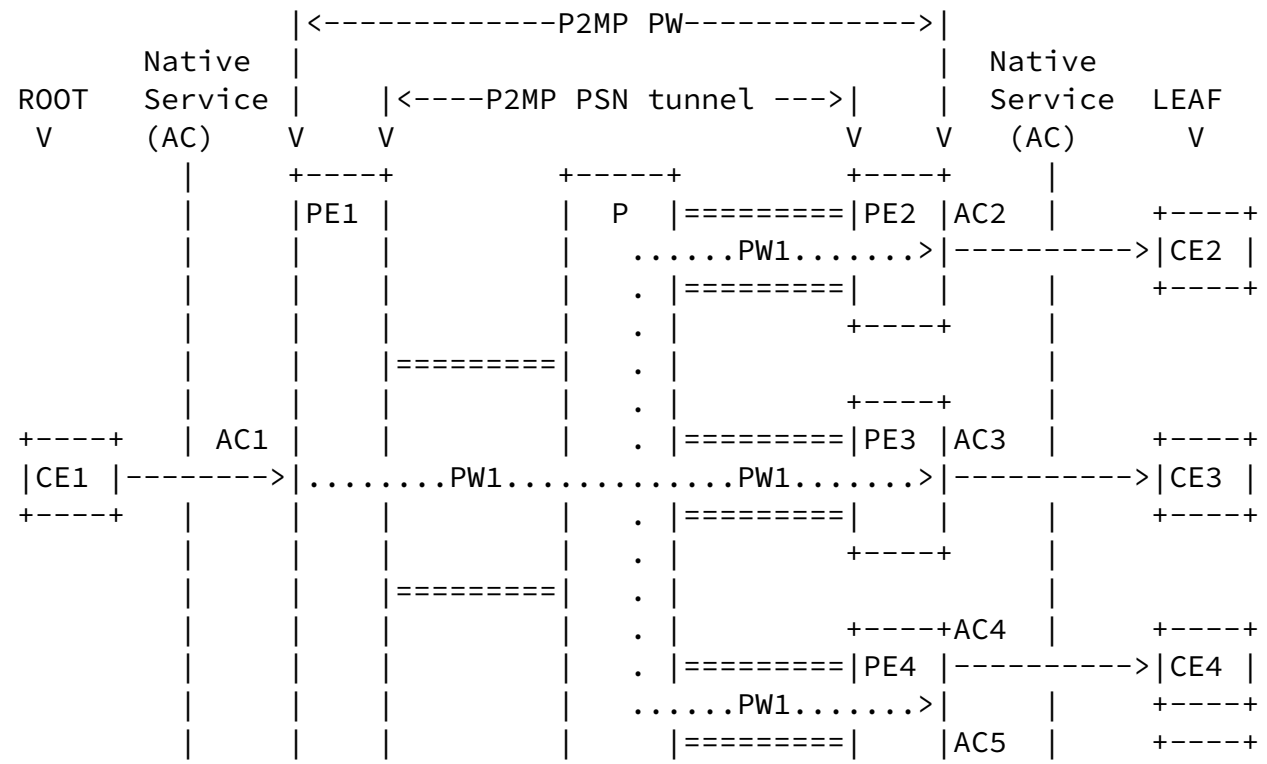
|<-------------P2MP PW------------->|
Native | | Native
ROOT Service | |<----P2MP PSN tunnel --->| | Service LEAF
V (AC) V V V V (AC) V
| +----+ +-----+ +----+ |
| |PE1 | | P |=========|PE2 |AC2 | +----+
| | | | ......PW1.......>|----------->|CE2 |
| | | | . |=========| | | +----+
			.	+----+
		=========	.	
			.	+----+
+----+	AC1			. |=========|PE3 |AC3	+----+
CE1	-------->	.........PW1.............PW1.......>	----------->	CE3
+----+				. |=========|		+----+
			.	+----+
		=========	.	
			.	+----+AC4	+----+
			.	=========	PE4	----------->	CE4
			......PW1.......>|	+----+
			|=========| |AC5	+----+

Figure 1 P2MP PW Reference Model

This architecture applies to the case where a P2MP PSN tunnel extends between edge nodes of a single PSN domain to transport a unidirectional P2MP PW with endpoints at these edge nodes. In this model a single copy of each PW packet is sent over the PW on the P2MP PSN tunnel and is received by all Leaf PEs due to the P2MP nature of the PSN tunnel. The P2MP PW SHOULD be traffic optimized, i.e., only one copy of a P2MP PW packet or PSN tunnel (underlying layer) is sent on any single link along the P2MP path. P routers participate in P2MP PSN tunnel operation but not in the signaling of P2MP PWs.

The Reference Model outlines the basic pieces of a P2MP PW. However, several levels of replication needs to be considered when designing a P2MP PW solution:

- Ingress PE replication to CEs: traffic is replicated to a set of local receiver CEs
- P router replication in the core: traffic replicated by means of P2MP PSN tunnel (P2MP LSP)
- Egress PE replication to CEs: traffic replicated to local receiver CEs

Theoretically, it is also possible to consider Ingress PE replication in the core; that is, all traffic is replicated to a set of P2P PSN transport tunnels at ingress, not using P router replication at all.

However, this approach may easily lead to more than one-stream bandwidth consumption at a single link, particularly if the PSN tunnels logically go over the same physical link. Hence this approach is not preferred.

Specific operations that MUST be performed at the PE on the native data units are not described here since the required pre-processing (Forwarder (FWRD) and Native Service Processing (NSP)) defined in section 4.2 of [RFC3985] are also applicable to P2MP PW.

P2MP PWs are generally unidirectional, but a Root PE may need to receive unidirectional P2P return traffic from any Leaf PE. For that purpose the P2MP PW solution MAY support an optional return path from

each Leaf PE to Root PE.

3.2. P2MP PW and Underlying Layer

The definition of MPLS multicast encapsulation [RFC5332] specifies the procedure to carry MPLS packets that are to be replicated and a copy of the packet sent to each of the specified next hops. This notion is also applicable to P2MP PW (as a MPLS) packet carried by a P2MP PSN tunnel.

To be more precise, a P2MP PSN tunnel corresponds to a "point-to-multipoint data link or tunnel" described in [RFC5332] Section 3. Similarly, P2MP PW labels correspond to "the top labels (before applying the data link or tunnel encapsulation) of all MPLS packets that are transmitted on a particular point-to-multipoint data link or tunnel."

In P2MP PW architecture, PW label with PW-PDU [RFC3985] is replicated by underlying P2MP PSN tunnel layer in SS-PW network model. In other words, it is intended to utilize PSN technology designed for efficient multicast/broadcast transport. Note that PW label is unchanged and hidden in switching by transit P routers as long as the model of SS-PW is taken.

In a solution, a P2MP PW MUST be supported over a single P2MP PSN tunnel as underlying layer of traffic distribution. Figure 2 gives an example of P2MP PW topology relying on a single P2MP LSP. The PW tree is composed of one Root PE (i1) and several Leaf PEs (e1, e2, e3, e4).
The mechanisms for establishing the PSN tunnel are outside the scope of this document, as long as they enable the essential attributes of the service to be emulated.

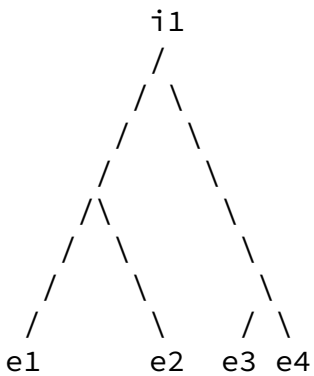

Figure 2 Example of P2MP Underlying Layer for P2MP PW

A single P2MP PSN tunnel MUST be able to serve more than one P2MP PW traffic in an aggregated way, i.e., multiplexing.

A P2MP PW solution MAY support different P2MP PSN tunneling technology (e.g., MPLS over GRE [RFC4023], or P2MP MPLS LSP) or different setup protocols. (e.g., MLDP [RFC6388], and P2MP RSVP-TE [RFC4875]).

The P2MP LSP associated to the P2MP PW can be selected either by user configuration or by dynamically using a multiplexing/demultiplexing mechanism.

The P2MP PW multiplexing SHOULD be used based on the overlap rate between P2MP LSP and P2MP PW. As an example, an existing P2MP LSP may attach more leaves than the ones defined as Leaf PEs for a given P2MP PW. It may be attractive to reuse it to minimize new configuration, but using this P2MP LSP would imply non-Leaf PEs (i.e. not part of the P2MP PW) to receive unwanted traffic.

Note: no special configuration is needed for non-Leaf PEs to drop those unwanted traffic because they do not have forwarding information entry unless they process corresponding P2MP PWs set-up operation (e.g. signaling).

The operator SHOULD determine whether the P2MP PW can accept partially multiplexing with P2MP LSP, and a minimum congruency rate may be defined. The Root PE can determine whether P2MP PW can multiplex to a P2MP LSP according to the congruency rate. The congruency rate SHOULD take into account several items, such as:

- the amount of overlap between the number of Leaf PEs of P2MP PW and existing egress PE routers of a P2MP LSP. If there is a complete overlap, the congruency is perfect and the rate is 100%.
- at the expense of the additional traffic (e.g. other VPNs) supported over the P2MP LSP.

With this procedure a P2MP PW is nested within a P2MP LSP. This allows multiplexing several PWs over a common P2MP LSP. Prior to the P2MP PW signaling phase, the Root PE determines which P2MP LSP will be used for this P2MP PW. The PSN Tunnel can be an existing PSN tunnel or the Root PE can create a new P2MP PSN tunnel. In addition, if ideal congruency rate is desired, if the P2MP PW has one or more extra leaf nodes that are not covered by the existing P2MP LSP, the

P2MP LSP SHOULD be modified or re-created to cover them.

3.3. P2MP PW Construction

[RFC5332] introduces two approaches to assign MPLS label (meaning PW label in P2MP PW context): Upstream-Assigned[RFC5331] and Downstream-Assigned. However, it is out of scope of this document which one should be used in PW construction. It is left to the specification of the solution work.

The following requirements apply to the establishment of P2MP PWs:

- PE nodes MUST be configurable with the P2MP PW identifiers and ACs.
- A discovery mechanism SHOULD allow the Root PE to discover the Leaf PEs, or vice versa.
- Solutions SHOULD allow single-sided operation at the Root PE for the selection of some AC(s) at the Leaf PE(s) to be attached to the PW tree so that the Root PE controls the Leaf attachment.

The Root PE SHOULD support a method to be informed about whether a Leaf PE has successfully attached to the PW tree.

3.4. P2MP PW Signaling Requirements

3.4.1. P2MP PW Identifier

The P2MP PW MUST be uniquely identified. This unique P2MP PW identifier MUST be used for all signaling procedures related to this PW (PW setup, Monitoring, etc).

3.4.2. PW type mismatch

The Root PE and Leaf PEs of a P2MP PW MUST be configured with the same PW type as defined in [RFC4446] for P2P PW. In case of a different type, a PE SHOULD abort attempts to attach the Leaf PE to the P2MP PW.

3.4.3. Interface Parameters sub-TLV

Some interface parameters [RFC4446] related to the AC capability have

been defined according to the PW type and are signaled during the PW setup.

Where applicable, a solution is REQUIRED to ascertain whether the AC at the Leaf PE is capable of supporting traffic coming from the AC at the Root PE.

In case of a mismatch, the passive PE (Root or Leaf PE, depending on the signaling process) SHOULD support mechanisms to reject attempts to attach the Leaf PE to the P2MP PW.

3.4.4. Leaf Grafting/Pruning

Once the PW tree is established, the solution MUST allow the addition or removal of a Leaf PE, or a subset of leaves to/from the existing tree, without any impact on the PW tree (data and control planes) for the remaining Leaf PEs.

The addition or removal of a Leaf PE MUST also allow the P2MP PSN tunnel to be updated accordingly. This may cause the P2MP PSN tunnel to add or remove the corresponding Leaf PE.

3.4.5. Failure Detection and Reporting

Since the underlying layer has an End-to-End P2MP topology between the Root PE and the Leaf PEs, the failure reporting and processing procedures are implemented only on the edge nodes.

Failure events may cause one or more Leaf PEs to become detached from the PW tree. These events MUST be reported to the Root PE, using appropriate out-of-band or inband Operations, Administration, and Maintenance (OAM) messages for monitoring.
It MUST be possible for the operator to choose the out-of-band or inband Monitoring tools or both to monitor the Leaf PE status.
The solution SHOULD allow the Root PE to be informed of Leaf PEs failure for management purposes.

Based on these failure notifications, solutions MUST allow the Root PE to update the remaining leaves of the PW tree.

- A solution MUST support in-band status notification mechanism to detect failures:
  unidirectional point-to-multipoint traffic failure. This MUST be realized by enhancing existing unicast PW methods, such as VCCV

for seamless and familiar operation defined in [RFC5085].
- In case of failure, it MUST correctly report which Leaf PEs are affected. This MUST be realized by enhancing existing PW methods, such as LDP Status Notification. The notification message SHOULD include the type of fault (P2MP PW, AC or PSN tunnel).
- A Leaf PE MAY be notified of the status of the Root PE's AC.
- A solution MUST support OAM message mapping [RFC6310] at the Root PE and Leaf PE if a failure is detected on the source CE.

3.4.6. Protection and Restoration

It is assumed that if recovery procedures are required, the P2MP PSN tunnel will support standard MPLS-based recovery techniques (typically based on RSVP-TE). In that case a mechanism SHOULD be implemented to avoid race conditions between recovery at the PSN level and recovery at the PW level.

An alternative protection scheme MAY rely on the PW layer.

Leaf PEs MAY be protected via a P2MP PW redundancy mechanism. In the example depicted below, a standby P2MP PW is used to protect the active P2MP PW. In that protection scheme the AC at the Root PE MUST serve both P2MP PWs. In this scenario, the condition when to do the switchover SHOULD be implemented, e.g. one or all Leaf failure of active P2MP PW will trigger the whole P2MP PW's switchover.

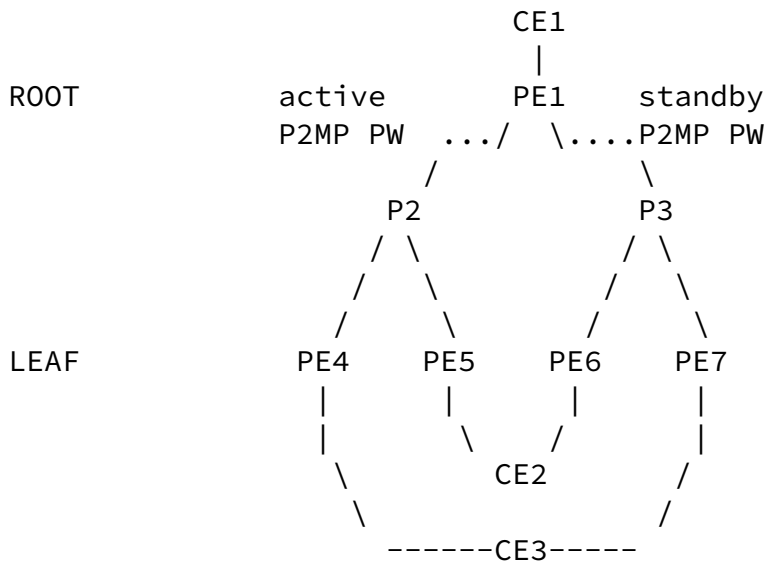

Figure 3: Example of P2MP PW redundancy for protecting Leaf PEs

Note that some of the nodes/links in this figure can be physically shared, which depends on the service provider policy of network redundancy.

The Root PE MAY be protected via a P2MP PW redundancy mechanism. In the example depicted below, a standby P2MP PW is used to protect the active P2MP. A single AC at the Leaf PE MUST be used to attach the CE to the primary and the standby P2MP PW. The Leaf PE MUST support protection mechanisms in order to select the active P2MP PW.

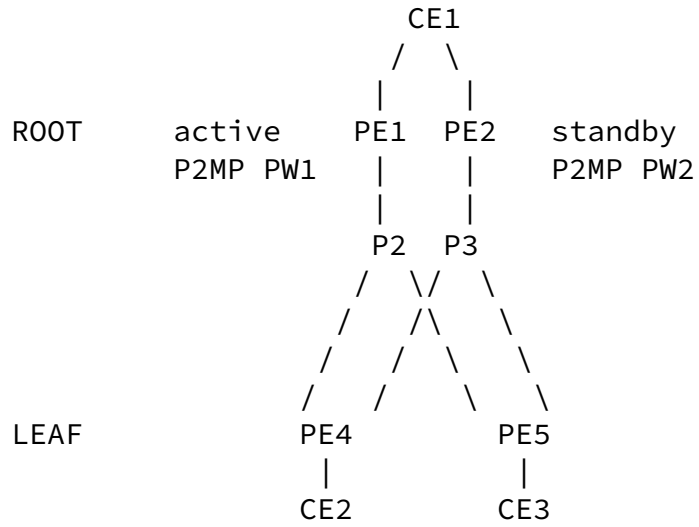

Figure 4: Example of P2MP PW redundancy for protecting Root PEs

3.4.7. Scalability

The solution SHOULD scale at worst linearly for message size, memory requirements, and processing requirements, with the number of Leaf PEs.
Increasing the number of P2MP PWs between a Root PE and a given set of Leaf PEs SHOULD NOT cause the P router to increase the number of entries in its forwarding table by the same or greater proportion. Multiplexing P2MP PWs to P2MP PSN Tunnels achieves this.

4. Backward Compatibility

Solutions MUST be backward compatible with current PW standards. Solutions SHOULD utilize existing capability advertisement and negotiation procedures for the PEs implementing P2MP PW endpoints.

The implementation of OAM mechanisms also implies the advertisement of PE capabilities to support specific OAM features.
The solution MAY allow advertising P2MP PW OAM capabiltities.
A solution MUST NOT allow a P2MP PW to be established to PEs that do not support P2MP PW functionality. It MUST have a mechanism to report an error for incompatible PEs.

In some cases, upstream traffic is needed from downstream CEs to upstream CEs. The P2MP PW solution SHOULD allow a return path (i.e. from the Leaf to the Root) that provides upstream connectivity.

In particular, the same ACs MAY be shared between downstream and upstream directions. For downstream, a CE receives traffic originated by the Root PE over its AC. For upstream, the CE MAY also send traffic destined to the same Root PE over the same AC.

5. Security Considerations

The security requirements common to PW are raised in Section 10 of [RFC3916]. P2MP PW is a variant of the initial P2P PW definition, and those requirements also apply to P2MP PW. The security considerations from [RFC5920], [RFC3985] and [RFC6941] also apply respectively to IP/MPLS and MPLS-TP deployment scenario.
Some issues specifically due to P2MP topology MUST be addressed in the definition of the solution:
- The solution SHOULD provide means to guarantee the traffic delivery to receivers (Integrity, Confidentially)
- The solution SHOULD support means to protect the P2MP PW as a whole against attacks that would lead to any kind of denial-of-service. Specifically, it would be desirable to consider safeguard mechanisms to avoid any negative impact on the whole PW Tree under the attack against its particular receiver(s). Considerations about both control plane and data plane are necessary.

6.IANA Considerations

This document does not require any IANA action.

7. Contributing Authors

Philippe Niger
France Telecom

2, avenue Pierre-Marzin
22307 Lannion Cedex
France

Email: philippe.niger@orange-ftgroup.com

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

EMail: lmartini@cisco.com

Lei Wang
Telenor
Snaroyveien 30
Fornebu 1331
Norway

Email: lei.wang@telenor.com

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

Email: rahul@juniper.net

Simon Delord
Telstra
380 Flinders lane. Melbourne

Email: simon.delord@gmail.com

Martin Vigoureux
Alcatel-Lucent France
Route de Villejust
91620 Nozay
France

Email: martin.vigoureux@alcatel-lucent.fr

Lizhong Jin
ZTE Corporation

889, Bibo Road
Shanghai, 201203, China

Email: lizho.jin@gmail.com

8. Acknowledgments

The authors thank the following people: the authors of [RFC4461] since the structure and content of this document were, for some sections, largely inspired by [RFC4461], JL Le Roux and A. Cauvin for the discussions, comments and support, Adrian Farrel for his Routing Area Director review, and IESG reviewers.

9. References

9.1. Normative References

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

[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.

[RFC5332] Eckert, T., Rosen, E., Aggarwal, R., and Y. Rekhter, "MPLS Multicast Encapsulations", RFC 5332, August 2008.

[RFC5659] Bocci, M. and S. Bryant, "An Architecture for Multi-Segment Pseudowire Emulation Edge-to-Edge", RFC 5659, October 2009.

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

[RFC6310] Aissaoui, M., Busschbach, P., Martini, L., Morrow, M., Nadeau, T., and Y(J). Stein, "Pseudowire (PW) Operations, Administration, and Maintenance (OAM) Message Mapping", RFC 6310, July 2011.

9.2. Informative References

[I-D.ietf-l2vpn-vpms-frmwk-requirements] Kamite, Y., Jounay, F., Niven-Jenkins, B., Brungard, D., and L. Jin, "Framework and Requirements for Virtual Private Multicast Service (VPMS)", draft-ietf-l2vpn-vpms-frmwk-requirements-05 (work in progress), October 2012.

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

[RFC4461] Yasukawa, S., "Signaling Requirements for Point-to-Multipoint Traffic-Engineered MPLS Label Switched Paths (LSPs)", RFC 4461, April 2006.

[RFC4875] Aggarwal, R., Papadimitriou, D., and S. Yasukawa, "Extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs)", RFC 4875, May 2007.

[RFC5085] Nadeau, T. and C. Pignataro, "Pseudowire Virtual Circuit Connectivity Verification (VCCV): A Control Channel for Pseudowires", RFC 5085, December 2007.

[RFC5254] Bitar, N., Bocci, M., and L. Martini, "Requirements for Multi-Segment Pseudowire Emulation Edge-to-Edge (PWE3)", RFC 5254, October 2008.

[RFC5331] Aggarwal, R., Rekhter, Y., and E. Rosen, "MPLS Upstream Label Assignment and Context-Specific Label Space", RFC 5331, August 2008.

[RFC6388] Wijnands, IJ., Minei, I., Kompella, K., and B. Thomas, "Label Distribution Protocol Extensions for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths", RFC 6388, November 2011.

[RFC5920] Fang, L., "Security Framework for MPLS and GMPLS Networks", RFC 5920, July 2010.

[RFC6941] Fang, L., Niven-Jenkins, B., Mansfield, S., Graveman, R., "MPLS Transport Profile (MPLS-TP) Security Framework", RFC 6941, April 2013.

Authors' Addresses

Frederic Jounay (editor)
Orange CH
4 rue caudray 1020 Renens
Switzerland

Email: frederic.jounay@orange.ch

Yuji Kamite (editor)
NTT Communications Corporation
Granpark Tower
3-4-1 Shibaura, Minato-ku
Tokyo 108-8118
Japan

Email: y.kamite@ntt.com

Giles Heron
Cisco Systems, Inc.
9 New Square
Bedfont Lakes
Feltham
Middlesex
TW14 8HA
United Kingdom

Email: giheron@cisco.com

Matthew Bocci
Alcatel-Lucent Telecom Ltd
Voyager Place
Shoppenhangers Road
Maidenhead
Berks

United Kingdom

Email: matthew.bocci@alcatel-lucent.co.uk