

Network Working Group  
Internet-Draft  
Intended status: Informational  
Expires: August 19, 2012

P. Muley  
M. Aissaoui  
M. Bocci  
Alcatel-Lucent  
February 16, 2012

**Pseudowire Redundancy**  
**draft-ietf-pwe3-redundancy-06**

**Abstract**

This document describes a framework comprised of a number of scenarios and associated requirements for pseudowire (PW) redundancy. A set of redundant PWs is configured between provider edge (PE) nodes in single segment PW applications, or between Terminating PE nodes in Multi-Segment PW applications. In order for the PE/T-PE nodes to indicate the preferred PW to use for forwarding PW packets to one another, a new PW status is required to indicate the preferential forwarding status of active or standby for each PW in the redundancy set.

**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 [RFC 2119](#) [[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 August 19, 2012.

**Copyright Notice**

Copyright (c) 2012 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

<a href="#">1.</a>	Introduction . . . . .	<a href="#">3</a>
<a href="#">2.</a>	Terminology . . . . .	<a href="#">3</a>
<a href="#">3.</a>	Reference Models . . . . .	<a href="#">5</a>
<a href="#">3.1.</a>	PE Architecture . . . . .	<a href="#">5</a>
<a href="#">3.2.</a>	PW Redundancy Network Reference Scenarios . . . . .	<a href="#">5</a>
<a href="#">3.2.1.</a>	Single Multi-Homed CE . . . . .	<a href="#">5</a>
<a href="#">3.2.2.</a>	Multiple Multi-Homed CEs . . . . .	<a href="#">7</a>
<a href="#">3.2.3.</a>	Single-Homed CE With MS-PW Redundancy . . . . .	<a href="#">8</a>
<a href="#">3.2.4.</a>	PW Redundancy Between MTU-s in H-VPLS . . . . .	<a href="#">9</a>
3.2.5.	PW Redundancy Between VPLS Network Facing PEs (n-PEs) . . . . .	<a href="#">10</a>
<a href="#">3.2.6.</a>	Redundancy in a VPLS Bridge Module Model . . . . .	<a href="#">11</a>
<a href="#">4.</a>	Generic PW Redundancy Requirements . . . . .	<a href="#">12</a>
<a href="#">4.1.</a>	Protection Switching Requirements . . . . .	<a href="#">12</a>
<a href="#">4.2.</a>	Operational Requirements . . . . .	<a href="#">13</a>
<a href="#">5.</a>	Security Considerations . . . . .	<a href="#">13</a>
<a href="#">6.</a>	IANA Considerations . . . . .	<a href="#">13</a>
<a href="#">7.</a>	Major Contributing Authors . . . . .	<a href="#">13</a>
<a href="#">8.</a>	Acknowledgements . . . . .	<a href="#">14</a>
<a href="#">9.</a>	References . . . . .	<a href="#">14</a>
<a href="#">9.1.</a>	Normative References . . . . .	<a href="#">14</a>
<a href="#">9.2.</a>	Informative References . . . . .	<a href="#">15</a>
	Authors' Addresses . . . . .	<a href="#">15</a>



## 1. Introduction

The objective of PW redundancy is to provide sparing of attachment circuits (ACs), Provider Edge nodes (PEs), and Pseudowires (PWs) to eliminate single points of failure, while ensuring that only one active path between a pair of Customer Edge nodes (CEs).

In single-segment PW (SS-PW) applications, protection for the PW is provided by the PSN layer. This may be a Resource Reservation Protocol with Traffic Engineering (RSVP-TE) labeled switched path (LSP) with a fast-Reroute (FRR) backup or an end-to-end backup LSP. PSN protection mechanisms cannot protect against failure of the a PE node or the failure of the remote AC. Typically, this is supported by dual-homing a Customer Edge (CE) node to different PE nodes which provide a pseudowire emulated service across the PSN. A set of PW mechanisms is therefore required that enables a primary and one or more backup PWs to terminate on different PE nodes.

In multi-segment PW (MS-PW) applications, PSN protection mechanisms cannot protect against the failure of a switching PE (S-PE). A set of mechanisms that support the operation of a primary and one or more backup PWs via a different set of S-PEs is therefore required. The paths of these PWs are diverse in the sense that they are switched at different S-PE nodes.

In both of these applications, PW redundancy is important to maximise the resiliency of the emulated service.

This document describes framework for these applications and its associated operational requirements. The framework utilizes a new PW status, called the Preferential Forwarding Status of the PW. This is separate from the operational states defined in [RFC4447](#) [[RFC4447](#)]. The mechanisms for PW redundancy are modeled on general protection switching principles.

## 2. Terminology

- o Up PW: A PW which has been configured (label mapping exchanged between PEs) and is not in any of the PW defect states specified in [[RFC4447](#)]. Such a PW is available for forwarding traffic.
- o Down PW: A PW that has either not been fully configured, or has been configured and is in any one of the PW defect states specified in [[RFC4447](#)]. Such a PW is not available for forwarding traffic.



- o Active PW. An UP PW used for forwarding user, OAM and control plane traffic.
- o Standby PW. An UP PW that is not used for forwarding user traffic but may forward OAM and specific control plane traffic.
- o PW Endpoint: A PE where a PW terminates on a point where Native Service Processing is performed, e.g., A Single Segment PW (SS-PW) PE, a Multi-Segment Pseudowire (MS-PW) Terminating PE (T-PE), or a Hierarchical VPLS MTU-s or PE-rs.
- o Primary PW: the PW which a PW endpoint activates (i.e. uses for forwarding) in preference to any other PW when more than one PW qualifies for active state. When the primary PW comes back up after a failure and qualifies for the active state, the PW endpoint always reverts to it. The designation of Primary is performed by local configuration for the PW at the PE.
- o Secondary PW: when it qualifies for the active state, a Secondary PW is only selected if no Primary PW is configured or if the configured primary PW does not qualify for active state (e.g., is DOWN). By default, a PW in a redundancy PW set is considered secondary. There is no Revertive mechanism among secondary PWs.
- o Revertive protection switching. Traffic will be carried by the primary PW if it is UP and a wait-to-restore timer expires and the primary PW is made the Active PW.
- o Non-revertive protection switching. Traffic will be carried by the last PW selected as a result of previous active PW entering Operationally DOWN state.
- o Manual selection of PW. The ability for the operator to manually select the primary/secondary PWs.
- o MTU-s: A hierarchical virtual private LAN service Multi-Tenant Unit switch, as defined in [RFC4762](#) [[RFC4762](#)].
- o PE-rs: A hierarchical virtual private LAN service switch, as defined in [RFC4762](#).
- o n-PE: A network facing provider edge node, as defined in [RFC4026](#) [[RFC4026](#)].

This document uses the term 'PE' to be synonymous with both PEs as per [RFC3985](#) [[RFC3985](#)] and T-PEs as per [RFC5659](#) [[RFC5659](#)].

This document uses the term 'PW' to be synonymous with both PWs as



per [RFC3985](#) and SS-PWs, MS-PWs, and PW segments as per [RFC5659](#).

### 3. Reference Models

The following sections describe show the reference architecture of the PE for PW redundancy and its usage in different topologies and applications.

#### 3.1. PE Architecture

Figure 1 shows the PE architecture for PW redundancy, when more than one PW in a redundant set is associated with a single AC. This is based on the architecture in Figure 4b of [RFC3985](#) [[RFC3985](#)]. The forwarder selects which of the redundant PWs to use based on the criteria described in this document.

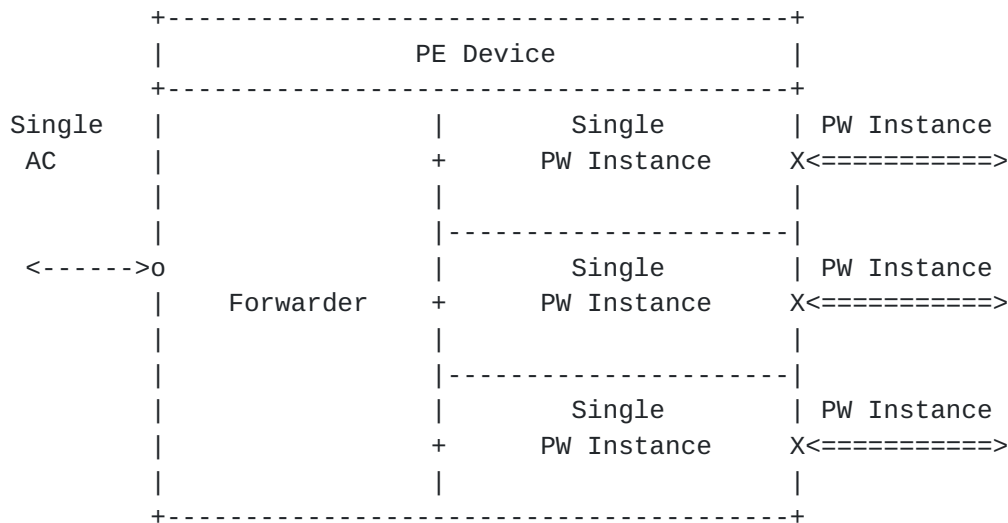


Figure 1: PE Architecture for PW redundancy

#### 3.2. PW Redundancy Network Reference Scenarios

This section presents a set of reference scenarios for PW redundancy.

##### 3.2.1. Single Multi-Homed CE

The following figure illustrates an application of single segment pseudowire redundancy. This scenario is designed to protect the emulated service against a failure of one of the PEs or ACs attached to the multi-homed CE. Protection against failures of the PSN tunnels is provided using PSN mechanisms such as MPLS Fast Reroute, so that these failures do not impact the PW.



CE1 is dual-homed to PE1 and PE3. A dual homing control protocol, the details of which are outside the scope of this document, selects which AC CE1 should use to forward towards the PSN, and which PE (PE1 or PE3) should forward towards CE1.

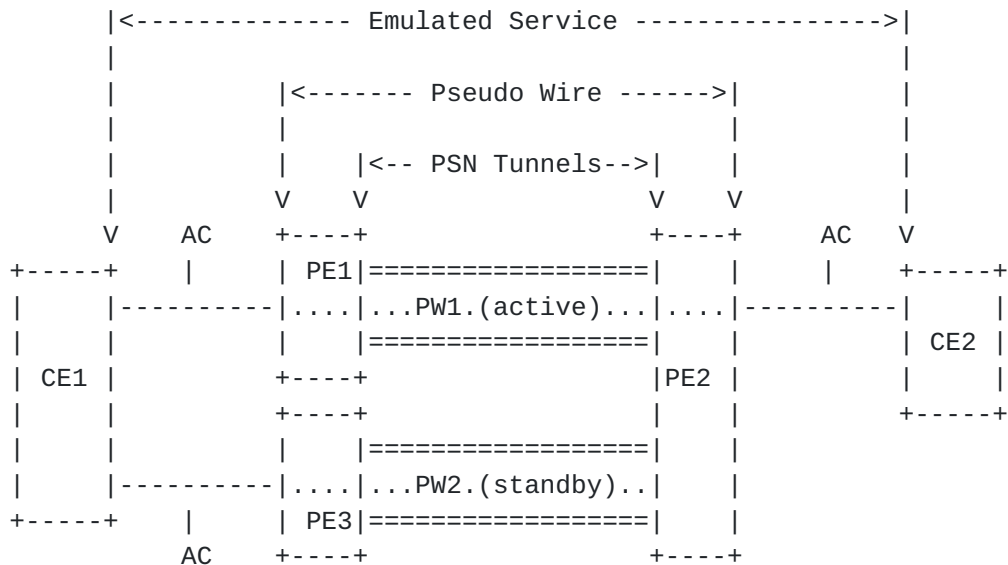


Figure 2: PW Redundancy with One Multi-Homed CE

In this scenario, only one of the PWs should be used for forwarding between PE1 / PE3, and PE2. PW redundancy determines which PW to make active based on the forwarding state of the ACs so that only one path is available from CE1 to CE2.

Consider the example where the AC from CE1 to PE1 is initially active and the AC from CE1 to PE3 is initially standby. PW1 is made active and PW2 is made standby in order to complete the path to CE2.

On failure of the AC between CE1 and PE1, the forwarding state of the AC on PE3 transitions to Active. The preferential forwarding state of PW2 therefore needs to become active, and PW1 standby, in order to reestablish connectivity between CE1 and CE2. PE3 therefore uses PW2 to forward towards CE2, and PE2 uses PW2 instead of PW1 to forward towards CE1. PW redundancy in this scenario requires that the forwarding status of the ACs at PE1 and PE3 be signaled to PE2 so that PE2 can choose which PW to make active.

Changes occurring on the dual homed side of network due to a failure of the AC or PE are not propagated to the ACs on the other side of the network. Furthermore, failures in the PSN are not be propagated to the attached CEs.



### 3.2.2. Multiple Multi-Homed CEs

This scenario, illustrated in Figure 3, is also designed to protect the emulated service against failures of the ACs and failures of the PEs. Here, both CEs, CE1 and CE2, are dual-homed to their respective PEs, PE1 and PE2, and PE3 and PE4. The method used by the CEs to choose which AC to use to forward traffic towards the PSN is determined by a dual-homing control protocol. The details of this protocol are outside the scope of this document.

Note that the PSN tunnels are not shown in this figure for clarity. However, it can be assumed that each of the PWs shown is encapsulated in a separate PSN tunnel. Protection against failures of the PSN tunnels is provided using PSN mechanisms such as MPLS Fast Reroute, so that these failures do not impact the PW.

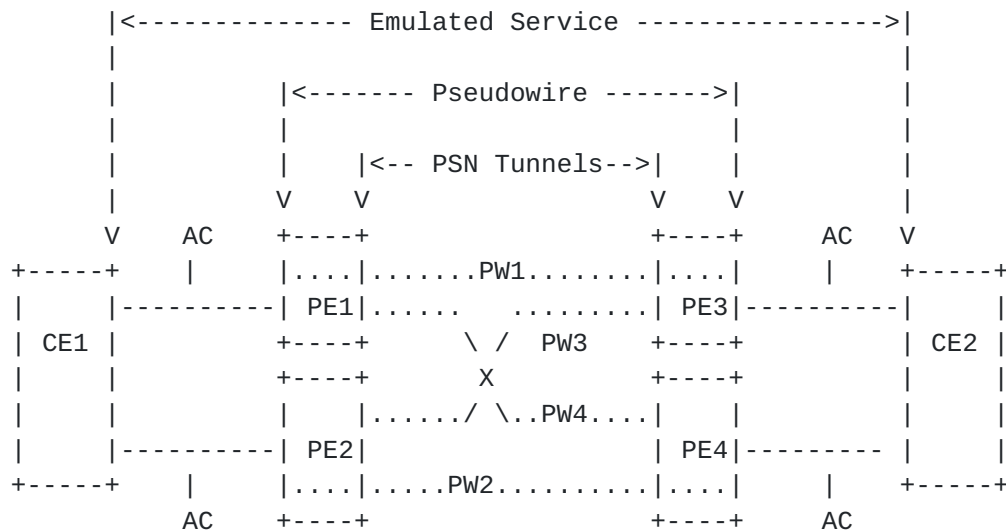


Figure 3: Multiple Multi-Homed CEs with PW Redundancy

PW1 and PW4 connect PE1 to PE3 and PE4, respectively. Similarly, PE2 has PW2 and PW3 connect PE2 to PE4 and PE3. PW1, PW2, PW3 and PW4 are all UP. In order to support N:1 or 1:1 protection, only one PW MUST be selected to forward traffic. This document defines an additional PW state that reflects this forwarding state, which is separate from the operational status of the PW. This is the 'Preferential Forwarding Status'.

If a PW has a preferential forwarding status of 'active', it can be used for forwarding traffic. The actual UP PW chosen by the combined set of PEs that interconnect the CEs is determined by considering the preferential forwarding status of each PW at each PE. The mechanisms for achieving this selection are outside the scope of this document. Only one PW is used for forwarding.



The following failure scenario illustrates the operation of PW redundancy in Figure 2. In the initial steady state, when there are no failures of the ACs, one of the PWs is chosen as the active PW, and all others are chosen as standby. The dual-homing protocol between CE1 and PE1/PE2 chooses to use the AC to PE2, while the protocol between CE2 and PE3/PE4 chooses to use the AC to PE4. Therefore the PW between PE2 and PE4 is chosen as the active PW to complete the path between CE1 and CE2.

On failure of the AC between the dual-homed CE1 and PE2, the preferential forwarding status of the PWs at PE1, PE2, PE3 and PE4 needs to change so as to re-establish a path from CE1 to CE2. Different mechanisms can be used to achieve this and these are beyond the scope of this document. After the change in status the algorithm for selection of PW needs to reevaluate and select PW to forward traffic. In this application each dual-homing algorithm, i.e., {CE1, PE1, PE2} and {CE2, PE3, PE4}, selects the active AC independently. There is therefore a need to signal the active status of each AC such that the PEs can select a common active PW for forwarding between CE1 and CE2.

Changes occurring on one side of network due to a failure of the AC or PE are not propagated to the ACs on the other side of the network. Furthermore, failures in the PSN are not be propagated to the attached CEs. Note that End-to-end native service protection switching can also be used to protect the emulated service in this scenario. In this case, PW3 and PW4 are not necessary.

If the CEs do not perform native service protection switching, they may instead may use load balancing across the paths between the CEs.

### **3.2.3. Single-Homed CE With MS-PW Redundancy**

This application is shown in Figure 4. The main objective is to protect the emulated service against failures of the S-PEs.



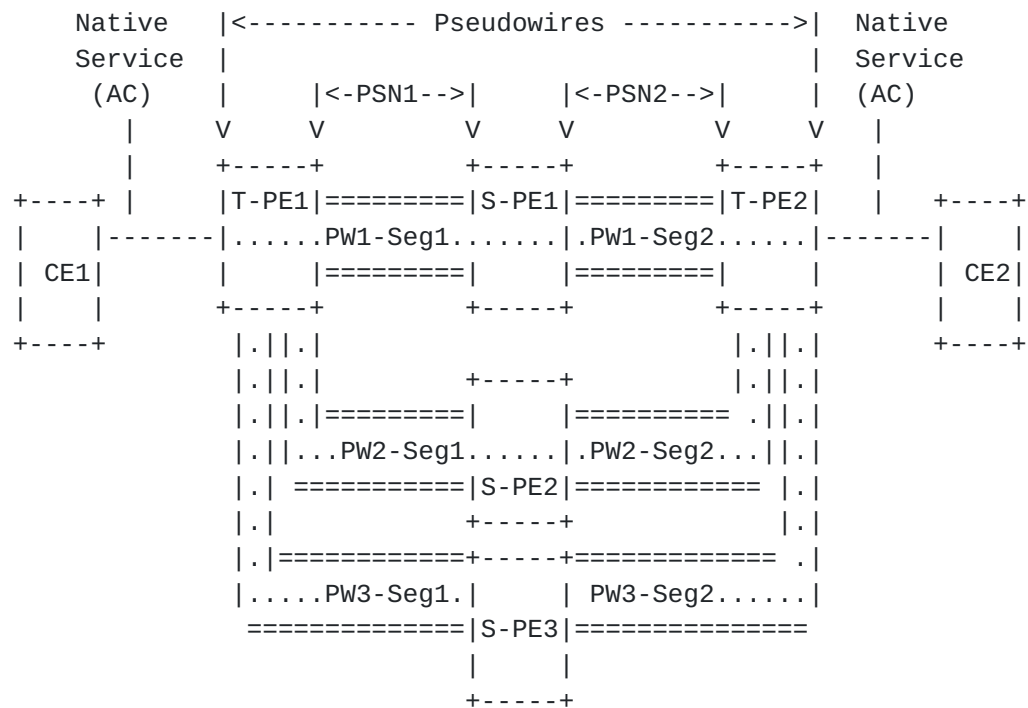


Figure 4: Single-Homed CE with MS-PW Redundancy

CE1 is connected to PE1 and CE2 to PE2, respectively. There are three multi-segment PWs. PW1 is switched at S-PE1, PW2 is switched at S-PE2, and PW3 is switched at S-PE3.

Since there is no multi-homing running on the ACs, the T-PE nodes would advertise 'Active' for the forwarding status based on a priority for the PW. Priorities associate meaning of 'primary PW' and 'secondary PW'. These priorities MUST be used in revertive mode as well and paths must be switched accordingly. The priority can be configuration or derivation from the PWid. Lower the PWid higher the priority. However, this does not guarantee selection of same PW by the T-PEs because, for example, mismatch of the configuration of the PW priority in each T-PE. The intent of this application is to have T-PE1 and T-PE2 synchronize the transmit and receive paths of the PW over the network. In other words, both T-PE nodes are required to transmit over the PW segment which is switched by the same S-PE. This is desirable for ease of operation and troubleshooting.

#### 3.2.4. PW Redundancy Between MTU-s in H-VPLS

The following figure illustrates the application of use of PW redundancy to Hierarchical VPLS (H-VPLS). Here, a Multi-Tenant Unit switch (MTU-s) is dual-homed to two PE router switches (PE-rs).



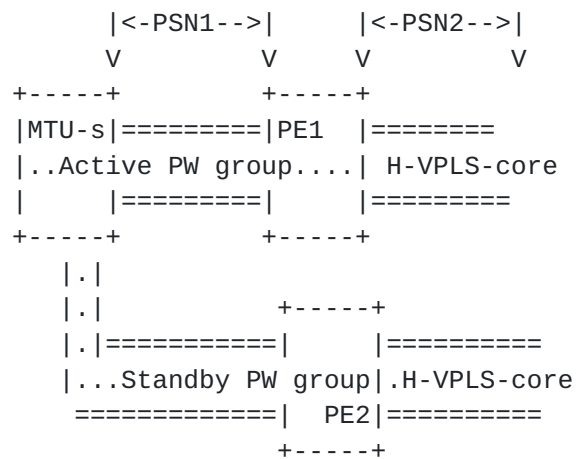


Figure 5: MTU-s Dual Homing in H-VPLS Core

In Figure 5, the MTU-s is dual homed to PE1 and PE2 and has spoke PWs to each of them. The MTU-s needs to choose only one of the spoke PWs ( the active PW) to one of the PE to forward the traffic and the other to standby status. The MTU-s can derive the status of the PWs based on local policy configuration. PE1 and PE2 are connected to the H-VPLS core on the other side of network. The MTU-s communicates the status of its member PWs for a set of VSIs having common status of Active or Standby. Here the MTU-s controls the selection of PWs to forward the traffic. Signaling using PW grouping with a common group-id in Pwid FEC Element or Grouping TLV in Generalized Pwid FEC Element as defined in [[RFC4447](#)] to PE1 and PE2 respectively, is recommended to scale better.

Whenever MTU-s performs a switchover, it needs to communicate to PE2 for the Standby PW group the changed status of active.

In this scenario, PE devices are aware of switchovers at MTU-s and could generate MAC Withdraw Messages to trigger MAC flushing within the H-VPLS full mesh. By default, MTU-s devices should still trigger MAC Withdraw messages as currently defined in [[RFC4762](#)] to prevent two copies of MAC withdraws to be sent (one by MTU-s and another one by PEs). Mechanisms to disable MAC Withdraw trigger in certain devices is out of the scope of this document.

### **3.2.5. PW Redundancy Between VPLS Network Facing PEs (n-PEs)**

Following figure illustrates the application of use of PW redundancy for dual homed connectivity between PE devices in a ring topology.



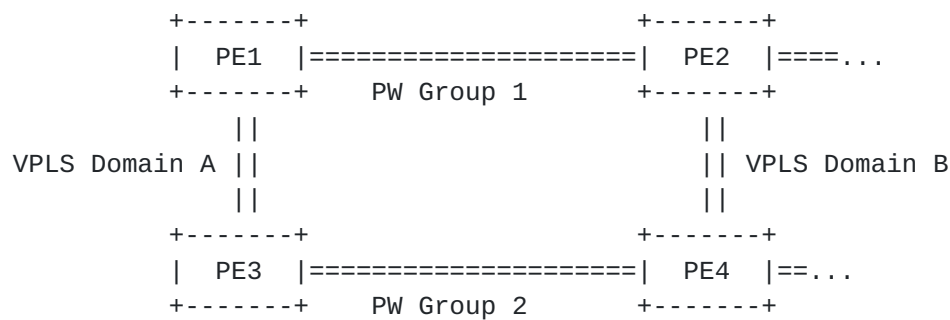


Figure 6: Redundancy in a Ring Topology

In Figure 6, PE1 and PE3 from VPLS domain A are connected to PE2 and PE4 in VPLS domain B via PW group 1 and group 2. Each of the PEs in the respective domains is connected to each other as well as forming the ring topology. Such scenarios may arise in inter-domain H-VPLS deployments where rapid spanning tree (RSTP) or other mechanisms may be used to maintain loop free connectivity of PW groups.

[RFC4762] outlines multi-domain VPLS services without specifying how multiple redundant border PEs per domain per VPLS instance can be supported. In the example above, PW group 1 may be blocked at PE1 by RSTP and it is desirable to block the group at PE2 by virtue of exchanging the PW preferential forwarding status of Standby. How the PW grouping should be done here is again deployment specific and is out of scope of the solution.

### 3.2.6. Redundancy in a VPLS Bridge Module Model

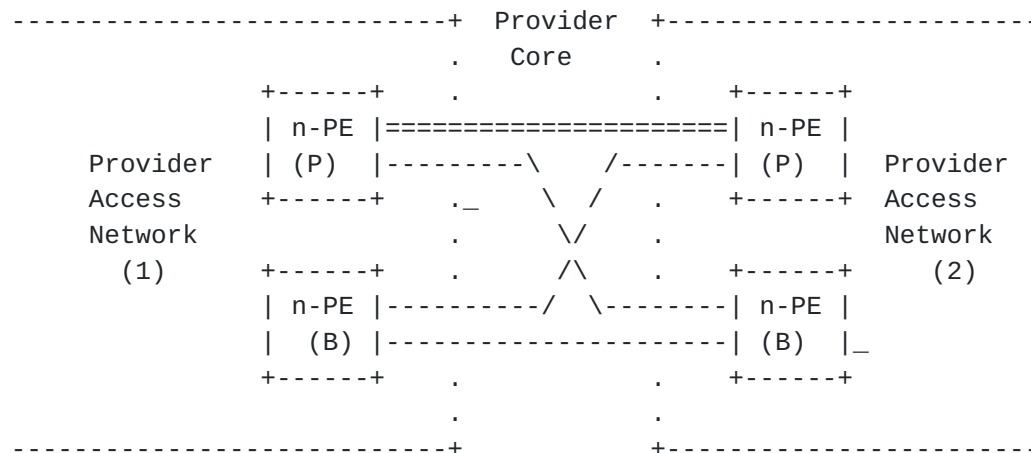


Figure 7: Bridge Module Model

In Figure 7, two provider access networks, each having two n-PEs, where the n-PEs are connected via a full mesh of PWs for a given VPLS instance. As shown in the figure, only one n-PE in each access



network is serving as a Primary PE (P) for that VPLS instance and the other n-PE is serving as the backup PE (B). In this figure, each primary PE has two active PWs originating from it. Therefore, when a multicast, broadcast, and unknown unicast frame arrives at the primary n-PE from the access network side, the n-PE replicates the frame over both PWs in the core even though it only needs to send the frames over a single PW (shown with == in the figure) to the primary n-PE on the other side. This is an unnecessary replication of the customer frames that consumes core-network bandwidth (half of the frames get discarded at the receiving n-PE). This issue gets aggravated when there is three or more n-PEs per provider, access network. For example if there are three n-PEs or four n-PEs per access network, then 67% or 75% of core-BW for multicast, broadcast and unknown unicast are wasted, respectively.

In this scenario, n-PEs can disseminate the status of PWs active/standby among them and furthermore to have it tied up with the redundancy mechanism such that per VPLS instance the status of active/backup n-PE gets reflected on the corresponding PWs emanating from that n-PE.

#### **4. Generic PW Redundancy Requirements**

##### **4.1. Protection Switching Requirements**

- o Protection architectures such as N:1, 1:1 or 1+1 are possible. 1:1 protection MUST be supported. The N:1 protection case is less efficient in terms of the resources that must be allocated and hence this SHOULD be supported. 1+1 protection architecture MAY be supported, but its definition is for further study.
- o Non-revertive behavior MUST be supported, while revertive behavior is OPTIONAL.
- o Protection switchover can be triggered by the operator e.g. using a Manual lockout/force switchover, or it may be triggered by a signal failure i.e. a defect in the PW or PSN. Both methods MUST be supported and signal failure triggers MUST be treated with a higher priority than any local or far-end operator-initiated trigger.
- o Note that a PE MAY be able to forward packets received from a standby status PW in order to avoid black holing of in-flight packets during switchover. However, in the case of use of VPLS, all VPLS application packets received from standby PWs MUST be dropped, except for OAM packets.



#### **4.2. Operational Requirements**

- o (T-)PEs involved in protecting a PW SHOULD automatically discover and attempt to resolve inconsistencies in the configuration of primary/secondary PW.
- o (T-)PEs involved in protecting a PW SHOULD automatically discover and attempt to resolve inconsistencies in the configuration of revertive/non-revertive protection switching mode.
- o (T-)PEs that do not automatically discover or resolve inconsistencies in the configuration of primary/secondary, revertive/non-revertive, or other parameters MUST generate an alarm upon detection of an inconsistent configuration.
- o (T-)PEs participating in PW redundancy MUST support the configuration of revertive or non-revertive protection switching modes.
- o (T-)PEs participating in PW redundancy SHOULD support the local invocation of protection switching.
- o (T-)PEs participating in PW redundancy SHOULD support the local invocation of a lockout of protection switching.
- o

#### **5. Security Considerations**

This document requires extensions to LDP that are needed for protecting pseudowires. These will inherit at least the same security properties as LDP [[RFC5036](#)] and the PW control protocol [[RFC4447](#)].

#### **6. IANA Considerations**

This document has no actions for IANA.

#### **7. Major Contributing Authors**

The editors would like to thank Pranjal Kumar Dutta, Marc Lasserre, Jonathan Newton, Hamid Ould-Brahim, Olen Stokes, Dave Mcdysan, Giles Heron and Thomas Nadeau who made a major contribution to the development of this document.



Pranjal Dutta  
Alcatel-Lucent  
Email: pranjal.dutta@alcatel-lucent.com

Marc Lasserre  
Alcatel-Lucent  
Email: marc.lasserre@alcatel-lucent.com

Jonathan Newton  
Cable & Wireless  
Email: Jonathan.Newton@cw.com

Olen Stokes  
Extreme Networks  
Email: ostokes@extremenetworks.com

Hamid Ould-Brahim  
Nortel  
Email: hbrahim@nortel.com

Dave McDysan  
Verizon  
Email: dave.mcdysan@verizon.com

Giles Heron  
Cisco Systems  
Email: giles.heron@gmail.com

Thomas Nadeau  
Computer Associates  
Email: tnadeau@lucidvision.com

## **8. Acknowledgements**

The authors would like to thank Vach Kompella, Kendall Harvey, Tiberiu Grigoriu, Neil Hart, Kajal Saha, Florin Balus and Philippe Niger for their valuable comments and suggestions.

## **9. References**

### **9.1. Normative References**

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [RFC3985] Bryant, S. and P. Pate, "Pseudo Wire Emulation Edge-to-



Edge (PWE3) Architecture", [RFC 3985](#), March 2005.

[RFC4026] Andersson, L. and T. Madsen, "Provider Provisioned Virtual Private Network (VPN) Terminology", [RFC 4026](#), 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)", [RFC 4447](#), April 2006.

[RFC4762] Lasserre, M. and V. Kompella, "Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling", [RFC 4762](#), January 2007.

[RFC5036] Andersson, L., Minei, I., and B. Thomas, "LDP Specification", [RFC 5036](#), October 2007.

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

## **[9.2.](#) Informative References**

[RFC6073] Martini, L., Metz, C., Nadeau, T., Bocci, M., and M. Aissaoui, "Segmented Pseudowire", [RFC 6073](#), January 2011.

### Authors' Addresses

Praveen Muley  
Alcatel-Lucent

Email: [praveen.muley@alcatel-lucent.com](mailto:praveen.muley@alcatel-lucent.com)

Mustapha Aissaoui  
Alcatel-Lucent

Email: [mustapha.aissaoui@alcatel-lucent.com](mailto:mustapha.aissaoui@alcatel-lucent.com)

Matthew Bocci  
Alcatel-Lucent

Email: [matthew.bocci@alcatel-lucent.com](mailto:matthew.bocci@alcatel-lucent.com)

