Network Working Group Internet Draft Intended Status: Informational Expires: March 2009 Praveen Muley Mustapha Aissaoui Matthew Bocci Pranjal Kumar Dutta Marc Lasserre Alcatel

> Jonathan Newton Cable & Wireless

> Olen Stokes Extreme Networks

Hamid Ould-Brahim Nortel

> Dave Mcdysan Verizon

Giles Heron Thomas Nadeau British Telecom

September 29, 2008

Pseudowire (PW) Redundancy draft-ietf-pwe3-redundancy-01.txt

Status of this Memo

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with <u>Section 6 of</u> BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

- The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt
- The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html

This Internet-Draft will expire on March 29, 2009.

Abstract

This document describes a framework comprised of few scenarios and associated requirements where PW redundancy is needed. A set of redundant PWs is configured between PE nodes in SS-PW applications, or between T-PE nodes in MS-PW applications. In order for the PE/T-PE nodes to indicate the preferred PW path to forward to one another, a new status is needed to indicate the preferential forwarding status of active or standby for each PW in the redundancy set.

Conventions used in this document

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 <u>RFC-2119</u> [<u>1</u>].

Table of Contents

<u>1</u> .	Terminology <u>3</u>
<u>2</u> .	Introduction
<u>3</u> .	Reference Model
	<u>3.1</u> . Multiple Multi-homed <u>5</u>
	<u>3.2</u> . Single Homed CE with MS-PW redundancy <u>6</u>
	3.3. PW redundancy between MTU-s8
	3.4. PW redundancy between n-PEs9
	3.5. PW redundancy in Bridge Module Model
<u>4</u> .	Generic PW redundancy requirements11
	4.1. Protection switching requirements11
	<u>4.2</u> . Operational requirements <u>11</u>
<u>5</u> .	Security Considerations <u>12</u>
<u>6</u> .	Acknowledgments <u>12</u>
<u>7</u> .	IANA considerations <u>12</u>
<u>8</u> .	References
	8.1. Normative References12
	8.2. Informative References
Aut	thor's Addresses
Int	tellectual Property Statement <u>15</u>

Muley et al. Expires March 29, 2009 [Page 2]

Disclaimer of Validity.....<u>15</u> Acknowledgment......<u>16</u>

1. Terminology

- o PW Terminating Provider Edge (T-PE). A PE where the customerfacing attachment circuits (ACs) are bound to a PW forwarder. A Terminating PE is present in the first and last segments of a MS-PW. This incorporates the functionality of a PE as defined in RFC3985 [3].
- o Single-Segment Pseudo Wire (SS-PW). A PW setup directly between two T-PE devices. Each PW in one direction of a SS-PW traverses one PSN tunnel that connects the two T-PEs.
- o Multi-Segment Pseudo Wire (MS-PW). A static or dynamically configured set of two or more contiguous PW segments that behave and function as a single point-to-point PW. Each end of a MS-PW by definition MUST terminate on a T-PE.
- o PW Segment. A part of a single-segment or multi-segment PW, which is set up between two PE devices, T-PEs and/or S-PEs.
- o PW Switching Provider Edge (S-PE). A PE capable of switching the control and data planes of the preceding and succeeding PW segments in a MS-PW. The S-PE terminates the PSN tunnels of the preceding and succeeding segments of the MS-PW.
- o PW switching point for a MS-PW. A PW Switching Point is never the S-PE and the T-PE for the same MS-PW. A PW switching point runs necessary protocols to setup and manage PW segments with other PW switching points and terminating PEs
- o Active PW. A PW whose preferential status is set to Active and Operational status is UP.
- o Standby PW. A PW whose preferential status is set to Standby and Operational status is UP.
- o Primary Path. The configured path which is preferred when revertive protection switching is used.
- o Secondary Path. One or more configured paths that are used by protection switching when current active PW path enters Operational DOWN state.

Muley et al. Expires March 29, 2009 [Page 3]

- o Revertive protection switching. Traffic will be carried by primary path if it is Operationally UP and the wait-to-restore timer expires and primary path is made the Active PW.
- o Non-revertive protection switching. Traffic will be carried by the last PW path selected as a result of previous active path entering Operationally DOWN state.
- o Manual selection of PW path. Ability for the operator to manually select the primary/secondary paths.

2. Introduction

In single-segment PW (SS-PW) applications, protection for the PW is provided by the PSN layer. This may be an RSVP LSP with a FRR backup and/or an end-to-end backup LSP. There are applications however where the backup PW terminates on a different target PE node. PSN protection mechanisms cannot protect against failure of the target PE node or the failure of the remote AC.

In multi-segment PW (MS-PW) applications, a primary and one or more secondary PWs in standby mode are configured in the network. The paths of these PWs are diverse in the sense that they are switched at different S-PE nodes. In these applications, PW redundancy is important for the service resilience.

In some deployments, it is important for operators that particular PW is preferred if it is available. For example, PW path with least latency may be preferred.

This document describes framework for these applications and its associated operational requirements. The framework comprises of new required status called preferential status to PW apart from the operational status already defined in the PWE3 control protocol [2]. The definition and operation of the preferential status is covered in ref.[7]

3. Reference Model

Following figures shows the reference model for the PW redundancy and its usage in different topologies and applications.

Muley et al. Expires March 29, 2009 [Page 4]

3.1. Multiple Multi-homed



Figure 1 Multiple Multi-homed CEs with single SS-PW redundancy

In the Figure 1 illustrated above both CEs, CE1 and CE2 are dualhomed with PEs, PE1, PE2 and PE3, PE4 respectively. The method for dual-homing and the used protocols such as Multi-chassis Link Aggregation Group (MC-LAG) 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.

PE1 has PW1 and PW4 service connecting PE3 and PE4 respectively. Similarly PE2 has PW2 and Pw3 pseudo wire service connecting PE4 and PE3 respectively. PW1, PW2, PW3 and PW4 are all operationally UP. In order to support N:1 or 1:1 only one PW is required to be selected to forward the traffic. Thus the PW needs to reflect his new status apart from the operational status. We call this as preferential forwarding status with state representing 'active' the one carrying traffic while the other 'standby' which is operationally UP but not forwarding traffic. The method of deriving Active/Standby status of the AC is outside the scope of this document. In case of MC-LAG it is derived by the Link Aggregation Control protocol (LACP) negotiation.

A new algorithm needs to be developed using the preferential forwarding state of PW and select only one PW to forward.

On failure of AC between the dual homed CE1 in this case lets say PE1 the preferential status on PE2 needs to be changed.

Muley et al. Expires March 29, 2009 [Page 5]

Internet-Draft

Pseudowire (PW) Redundancy) September 2008

Different mechanisms/protocols can be used to achieve this and these are beyond the scope of this document. For example the MC-LAG control protocol changes the link status on PE2 to active. After the change in status the algorithm for selection of PW needs to revaluate and select PW to forward the traffic.

In this application, because each dual-homing algorithm running on the two node sets, i.e., {CE1, PE1, PE2} and {CE2, PE3, PE4}, selects the active AC independently, there is a need to signal the active status of the AC such that the PE nodes can select a common active PW path for end-to-end forwarding between CE1 and CE2. This helps in restricting the changes occurring on one side of network due to failure to the other side of the network. Note this method also protects against any single PE failure or some dual PE failures.

One Multi-homed CE with single SS-PW redundancy application is a subset of above. Only PW1 and PW3 exist in this case. This helps against AC failure and PE failure of dual homed AC. Similar requirements applies in usage MS-PW redundancy as well. An additional requirement applicable to MS-PW is forwarding of status notification through S-PE. In general from customer view, SS-PW and MS-PW has similar resiliency requirement.

There is also a 1:1 protection switching case that is a subset of the above where PW3 and PW4 are not present and the CEs do not perform native service protection switching, but instead may use load balancing. This protects against AC failures and can use the native service to indicate active/failed state.

If each CE homes to different PEs, then the CEs can implement native service protection switching, without any PW redundancy functions. All that the PW needs to do is detect AC, PE, or PSN tunnel failures and convey that information to both PEs at the end of the PW. This is applicable to MS-PW as well.

3.2. Single Homed CE with MS-PW redundancy

This is the main application of interest and the network setup is shown in Figure 2

Muley et al. Expires March 29, 2009 [Page 6]

Native	<		Pseudo	Wire		>	Nativ	'e
Service	1					1	Servi	ce
(AC)	Ì	<-PSN	1>	<-PSI	N2>	Í	(AC)	
I	V	V	V	V	V	V		
	+	+	+	+	+	+		
++	T - PE	E1 =====	===== S-PE	1 =====	===== T-	PE2		+ +
	-	PW1-S	Seg1	.PW1	-Seg2	-		
CE1	1	=====	=====	=====	=====			CE2
	+	+	+	+	+	+		
++		++
	1.1	.	+	+	.	.		
	1.1	. =====	=====	=====		.		
. PW2-Seg1 .PW2-Seg2 .								
. ======== S-PE2 ====================================								
	1.1		+	+		.		
	. =		====+	- +====	========	= .		
		PW3-S	Seg1.	PW3	-Seg2			
	===		===== S-PE	3 ====	========	===		
			+	- +				

Figure 2 Single homed CE with multi-segment pseudo-wire redundancy

In Figure 2, CE1 is connected to PE1 in provider Edge 1 and CE2 to PE2 in provider edge 2 respectively. There are three segmented PWs. A PW1, is switched at S-PE1, PW2, which is switched at S-PE2 and PW3, is switched at S-PE3.

Since there is no multi-homing running on the AC, the T-PE nodes would advertise 'Active' for the forwarding status based on the priority. 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 that paths of the PW are synchronized because for example of 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.

Muley et al. Expires March 29, 2009 [Page 7]

Internet-Draft

3.3. PW redundancy between MTU-s

Following figure illustrates the application of use of PW redundancy in spoke PW by dual homed MTU-s to PEs.

|<-PSN1-->| |<-PSN2-->| V V V V +---+ +---+ |MTU-s|======|PE1 |======= |...Active PW group....| H-VPLS-core | |=======| |======== +---+ +---+ 1.1 +---+ 1.1 |.|========| |=========== |...Standby PW group|.H-VPLS-core ========= PE2|======== +---+

Figure 3 Multi-homed MTU-s in H-VPLS core

In Figure 3, MTU-s is dual homed to PE1 and PE2 and has spoke PWs to each of them. MTU-s needs to choose only one of the spoke PW (active PW) to one of the PE to forward the traffic and the other to standby status. MTU-s can derive the status of the PWs based on local policy configuration. PE1 and PE2 are connected to H-VPLS core on the other side of network. MTU-s communicates the status of its member PWs for a set of VSIs having common status Active/Standby. Here MTU-s controls the selection of PWs to forward the traffic. Signaling using PW grouping with common group-id in PWid FEC Element or Grouping TLV in Generalized PWid FEC Element as defined in [2] to PE1 and PE2 respectively, is encouraged to scale better.

Whenever MTU-s performs a switchover, it needs to communicate to PE2-rs 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 [5] 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.

Muley et al. Expires March 29, 2009 [Page 8]

3.4. PW redundancy between n-PEs

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

+---+ +---+ +----+ PW Group 1 +----+ VPLS Domain A || || VPLS Domain B +---+ +---+ +----+ PW Group 2 +----+

In Figure 4, 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 PE in respective domain is connected to each other as well to form the ring topology. Such scenarios may arise in inter-domain H-VPLS deployments where RSTP or other mechanisms may be used to maintain loop free connectivity of PW groups.

Ref.[5] outlines about multi-domain VPLS service without specifying how redundant border PEs per domain per VPLS instance can be supported. In the example above, PW group1 may be blocked at PE1 by RSTP and it is desirable to block the group at PE2 by virtue of exchanging the PW preferential status as Standby. How the PW grouping should be done here is again deployment specific and is out of scope of the solution.

3.5. PW redundancy in Bridge Module Model

Figure 4 Redundancy in Ring topology



Figure 5 Bridge Module Model

In Figure 5, 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 respectively wasted.

Muley et al. Expires March 29, 2009 [Page 10]

In this scenario, Standby PW signaling defined in [7] can be used among n-PEs that can disseminate the status of PWs (active or blocked) among themselves 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.

<u>4</u>. Generic PW redundancy requirements

4.1. Protection switching requirements

- o Protection architecture such as N:1,1:1 or 1+1 can be used. N:1 protection case is somewhat inefficient in terms of capacity consumption hence implementations SHOULD support this method while 1:1 being subset and efficient MUST be supported. 1+1 protection architecture can be supported but is left for further study.
- o Non-revertive mode MUST be supported, while revertive mode is an optional one.
- o Protection switchover can be operator driven like Manual lockout/force switchover or due to signal failure. Both methods MUST be supported and signal failure MUST be given higher priority than any local or far end request.

<u>4.2</u>. 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 involved with protection switching MUST support the configuration of revertive or non-revertive protection switching mode.
- o (T-)PEs involved with protection switching SHOULD support the local invocation of protection switching.

Muley et al. Expires March 29, 2009 [Page 11]

- o (T-)PEs involved with protection switching SHOULD support the local invocation of a lockout of protection switching.
- o In standby status PW can still receive packets in order to avoid black holing of in-flight packets during switchover. However in case of use of VPLS application packets are dropped in standby status except for the OAM packets.

5. Security Considerations

This document expects extensions to LDP that are needed for protecting pseudo-wires. It will have the same security properties as in LDP $[\underline{4}]$ and the PW control protocol $[\underline{2}]$.

6. Acknowledgments

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.

7. IANA considerations

This document has no actions for IANA.

8. References

8.1. Normative References

- Bradner, S., "Key words for use in RFCs to Indicate [1] Requirement Levels", BCP 14, RFC 2119, March 1997.
- [2] Martini, L., et al., "Pseudowire Setup and Maintenance using LDP", RFC 4447, April 2006.
- [3] Bryant, S., et al., " Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture", March 2005
- [4] Andersson, L., Minei, I., and B. Thomas, "LDP Specification", RFC 5036, January 2001
- [5] Kompella, V., Lasserrre, M., et al., "Virtual Private LAN Service (VPLS) Using LDP Signalling", RFC 4762, January 2007

<u>8.2</u>. Informative References

- Martini, L., et al., "Segmented Pseudo Wire", <u>draft-ietf-pwe3-</u> [6] segmented-pw-09.txt, January 2009.
- Muley, P. et al., "Preferential forwarding status bit", draft-[7] ietf-pwe3-redundancy-bit-01.txt, March 2009.
- IEEE Std. 802.1D-2003-Media Access Control (MAC) Bridges. [8]

Author's Addresses

Praveen Muley Alcatel 701 E. Middlefiled Road Mountain View, CA, USA Email: Praveen.muley@alcatel-lucent.com

Mustapha Aissaoui Alcatel 600 March Rd Kanata, ON, Canada K2K 2E6 Email: mustapha.aissaoui@alcatel-lucent.com

Matthew Bocci Alcatel Voyager Place, Shoppenhangers Rd Maidenhead, Berks, UK SL6 2PJ Email: matthew.bocci@alcatel-lucent.co.uk

Pranjal Kumar Dutta Alcatel-Lucent Email: pdutta@alcatel-lucent.com

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

Jonathan Newton Cable & Wireless Email: Jonathan.Newton@cwmsg.cwplc.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 BT Email: giles.heron@gmail.com

Muley et al. Expires March 29, 2009 [Page 14]

Thomas Nadeau BT Email: tnadeau@lucidvision.com

Intellectual Property Statement

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in <u>BCP 78</u> and <u>BCP 79</u>.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Disclaimer of Validity

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Copyright Statement

Copyright (C) The IETF Trust (2008).

Muley et al. Expires March 29, 2009 [Page 15]

This document is subject to the rights, licenses and restrictions contained in <u>BCP 78</u>, and except as set forth therein, the authors retain all their rights.

Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.