

Network Working Group  
Internet Draft

M. Bocci  
M. Aissaoui  
Alcatel-Lucent

L. Martini  
Cisco

Expires: May 2008

November 12, 2007

Pseudowire Emulation MPLS PSN Congestion Status Bit  
draft-bocci-martini-pwe3-psn-congestion-bit-00.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 [Section 6 of 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/lid-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 April 12, 2008.

#### Copyright Notice

Copyright (C) The IETF Trust (2007).

#### Abstract

Internet-Draft

PWE3 Congestion Status Bit

November 2007

Draft-ietf-pwe3-congestion-frmwk-00.txt [2] describes requirements for a PE providing a PWE3 service to take action if congestion is detected in the underlying PSN. This draft provides a control plane mechanism to enable a downstream PE detecting congestion to signal that congestion state to an upstream PE so that it may take appropriate action on its PWs.

#### 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 [RFC-2119](#) [1].

#### Table of Contents

1. Introduction	2
2. Scope	3
3. Reference Model	3
4. PSN Congestion Detection Mechanisms	4
5. PSN Congestion Signaling Procedures	5
6. Prevention of Congestion State Oscillation.	5
7. Congestions Status Bit	6
8. IANA Considerations	6
9. Security Considerations	6
10. Acknowledgments	7
11. References	7

#### 1. Introduction

[2] provides requirements and a framework for congestion control for pseudowires. Many pseudowire types transport traffic, which does not behave in a manner similar to TCP when there is congestion in the underlying PSN. That is, they carry traffic such as TDM that does not automatically reduce its rate when congestion occurs. TCP/IP, on the other hand, will reduce its rate and sources will adjust to a sending rate that allows the control plane of the PSN to continue to function.

There is a requirement for pseudowires to support a mechanism by which the ingress PE to a PWE3 service can take action to prevent its pseudowires from congesting the PSN. However, although a number of methods to enable an egress PE to detect congestion exist (see

[2]), there is to-date no method for communicating that congestion state to the ingress PE.

This draft presents extensions to PW status signalling to achieve this. PW status signalling is used because it uses a reliable channel between the PEs; if the PSN is so congested that PW status messages are lost, then LDP hello messages will also be lost. This will cause the PE to declare the link down and the PWs to be released. A PE detecting congestion sends a PSN Congestion status notification to the ingress peer PE for the PW on which congestion is detected. The PE receiving this notification can then take relevant action on the offending PWs, such as releasing the PW or throttling the rate of the PW.

## 2. Scope

This draft describes a congestion notification mechanism where congestion is detected by an egress PE and congestion control actions on PWs are performed by an ingress PE. The draft assumes that each PW or PW segment is established using the PW control protocol [5].

## 3. Reference Model

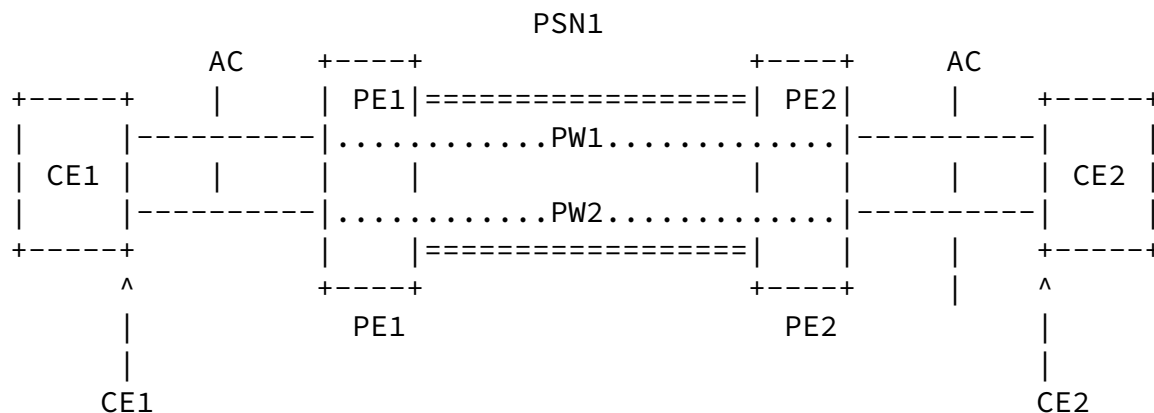


Figure 1 PWE3 reference Architecture

Figure 1 shows the PWE3 reference architecture, derived from [\[RFC3985\]](#).

Traffic from CE1 to CE2 enters the PSN via PE1 (the ingress PE). For the sake of the description in this draft, we assume that congestion occurs in PSN1 as a result of traffic on one or more PWs between PE1 and PE2. PE2 (the egress PE) detects congestion in PSN1 and signals

this congestion state to PE1 (the ingress PE). PE1 can then take actions on the PWs to mitigate congestion, as described in [\[2\]](#).

The reference architecture for multi-segment PWs is shown in Figure 2.

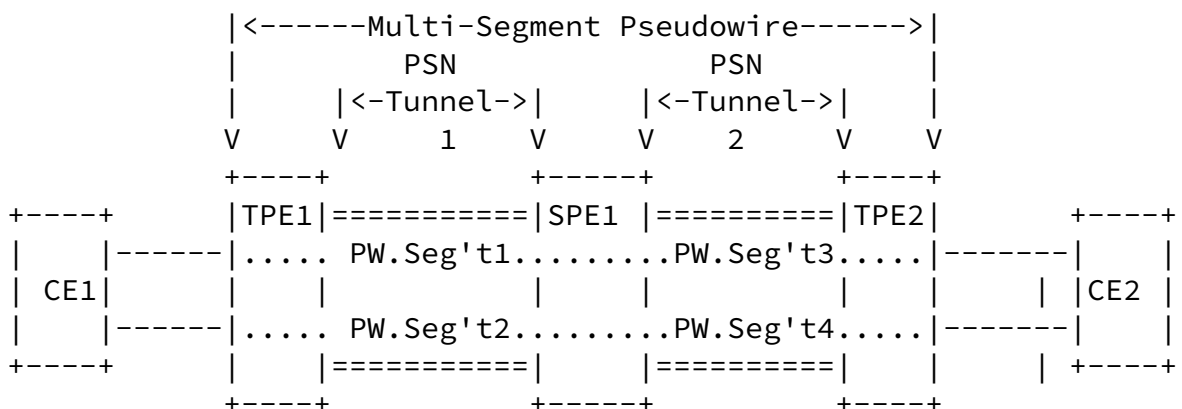


Figure 2 Reference Model for MS-PWs

Congestion occurring in PSN1 for traffic from CE1 to CE2 will be detected by SPE1, while congestion occurring in PSN2 will be detected by TPE2. In either case, the detected congestion state needs to be communicated to the ingress T-PE so that appropriate actions can be taken.

#### 4. PSN Congestion Detection Mechanisms

The protocol described in this draft relies on mechanisms for detecting PSN congestion specified in [\[2\]](#). At least one of these mechanisms MUST be used.

## [5.](#) PSN Congestion Signaling Procedures

Consider the case where congestion occurs in PSN1 for packets traveling from PE1 to PE2. This is detected by PE2 using one of the mechanisms described in [\[2\]](#). At the onset of this congestion, or when PE2 determines that PSN congestion is imminent, PE2 MUST send a PW status message with the PSN Congestion bit set. The status message MAY be sent as a wildcard notification message to each ingress PE for which the egress PE has detected at least one PW in congestion state.

On receipt of the status message, PE1 (the ingress PE) MUST implement congestion control on PWs that are destined for PE2. The precise details of the mechanisms used are outside the scope of this draft.

When PE1 detects that congestion in PSN1 has ceased, it MUST send a PW status message to PE1 with the PSN Congestion bit cleared. On receipt of a PW status message with the PSN congestion bit cleared, PE1 MAY cease applying congestion control to PWs destined for PE2. However, there may be some benefit to introducing a random delay to this cessation in order to prevent the PWs immediately re-congesting PSN1. This mechanism is described in [Section 6](#).

Similar procedures apply to MS-PWs. An egress T-PE or S-PE detecting PSN congestion sends a PW status message with the PSN congestion bit set to its peer S-PE or T-PE in the upstream direction. This T-PE or S-PE MAY also insert a PW switching TLV [\[4\]](#) with the prefix set to

indicate to an upstream T-PE or S-PE the location of the congestion. An S-PE receiving a PW status message relays it to the upstream segment irrespective of the state of the PSN Congestion bit, as described in [segment-pw]. The S-PE MAY also apply congestion control to PWs locally where it represents a policy control point between PSNs. A T-PE receiving a PW status message with the PSN Congestion bit set MUST apply congestion control to the affected PWs, as described as for SS-PWs described above.

## [6.](#) Prevention of Congestion State Oscillation.

The application of PW congestion control may enable the PSN to return to the un-congested state, causing the egress PE to signal no congestion to the ingress PE. However, the PSN may rapidly re-congest if the ingress PE immediately returns all of the PWs to their pre-

congestion sending rate, or immediately re-establishes all PWs which were released in order to prevent congestion.

In order to prevent such congestion state oscillations occurring, an ingress PE should introduce a per-PW random delay between receiving the PW status message with the PSN Congestion bit cleared, and returning each PW to its pre-congestion state (or allowing a PW that was released to be re-established as in the case of constant bit rate PW types). However it is highly desirable that the PE use a network bandwidth control method similar to the method used by TCP which gradually increases the window size until it experiences a dropped packet. In the case of a PW type that can be policed, or shaped to a specific network utilization bandwidth, the PW SHOULD, whenever possible be shaped to a much smaller network bandwidth utilization. When the congestion status bit is cleared, the PE can gradually increase the PW network bandwidth utilization until either it returns to the required full speed, or it starts to experience congestion again.

More details of this procedure will be explained in a subsequent version of this document.

## [7.](#) Congestions Status Bit

The PE/T-PE/S-PE nodes indicate to each other the congestion state of the intervening PSN using this bit.

0x00000TBD When the bit is set, it represents "PSN Congestion"

When the bit is cleared, it represents "No PSN Congestion"

## [8.](#) IANA Considerations

IANA needs to allocate the bit "0x00000TBD" to mean "PSN Congestion Status" in the PW status registry.

## [9.](#) Security Considerations

This draft introduces no new security considerations above those in [\[RFC3985\]](#) and [\[MS-ARCH\]](#).

Bocci & Martini

Expires May 12, 2008

[Page 6]

---

Internet-Draft

PWE3 Congestion Status Bit

November 2007

## [10.](#) Acknowledgments

The authors gratefully acknowledge the input of Dimitri Papadimitriou.

## [11.](#) References

- [1] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [2] Bryant et al; "Pseudowire Congestion Control Framework"; [draft-ietf-pwe3-congestion-frmwk-00.txt](#) ; February 2007
- [3] Bryant, S. and Pate, P. (Editors), "Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture", [RFC 3985](#), March 2005
- [4] Martini et al, "Segmented Pseudo Wire", Internet Draft, [draft-](#)

[ietf-pwe3-segmented-pw-05.txt](#), July 2007

- [5] Martini et al; Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP); [RFC 4447](#); April 2006

#### Author's Addresses

Matthew Bocci  
Alcatel-Lucent  
Email: [matthew.bocci@Alcatel-lucent.co.uk](mailto:matthew.bocci@Alcatel-lucent.co.uk)

Mustapha Aissaoui  
Alcatel-Lucent  
Email: [Mustapha.aissaoui@Alcatel-lucent.com](mailto:Mustapha.aissaoui@Alcatel-lucent.com)

Luca Martini  
Cisco  
Email: [lmartini@cisco.com](mailto:lmartini@cisco.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

Bocci & Martini

Expires May 12, 2008

[Page 7]

---

Internet-Draft

PWE3 Congestion Status Bit

November 2007

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 [BCP 78](#) and [BCP 79](#).

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](mailto: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 (2007).

This document is subject to the rights, licenses and restrictions contained in [BCP 78](#), 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.