

PALS Working Group
Internet Draft
Intended status: Standard Track
Expires: April 2019

Italo Busi
Stewart Bryant
Andrew G. Malis
Jie Dong
Huawei

October 22, 2018

Pseudowire (PW) Control Word (CW) Stitching
draft-busi-pals-pw-cw-stitching-01.txt

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), 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 22, 2019.

Copyright Notice

Copyright (c) 2018 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.

Abstract

This document defines the behavior of a new type of Multi-Segment Pseudowire (MS-PW) Switching PE (S-PE) which enhances the S-PE functions defined in [RFC 6073], with the capability to switch an Ethernet pseudowire (PW) segment that uses the PW Control Word (CW) [RFC 4385] with an Ethernet PW segment that does not use the CW.

This new type of S-PE can be deployed in the network one hop away, at the MPLS layer, from a Terminating PE (T-PE) which does not support CW for Ethernet PW encapsulation [RFC 4448]. In this way, all the Ethernet PW packets sent through the MPLS network will have the CW and be protected against incorrect equal-cost-multi-path (ECMP) behavior as described in [I-D ETH-CW].

Table of Contents

1. Introduction.....	2
1.1. Assumptions.....	5
2. Terminology.....	5
2.1. Conventions Used in This Document.....	6
3. Control Word Stitching procedures.....	6
3.1. CW Stitching Signaling.....	7
4. VCCV Stitching Procedures.....	8
4.1. VCCV Stitching for CC Type 3.....	9
4.2. VCCV Stitching for CC Type 4.....	10
4.3. VCCV Stitching Signaling.....	12
5. Other Deployment Scenarios.....	13
6. Security Considerations.....	15
7. IANA Considerations.....	16
8. References.....	16
8.1. Normative References.....	16
8.2. Informative References.....	16
9. Acknowledgments.....	17

1. Introduction

In order to protect Ethernet pseudowire (PW) packets against incorrect equal-cost-multi-path (ECMP) behavior, which may cause out-of-order delivery of the payload Ethernet frames, the use of PW control word (CW) has been recommended in [I-D ETH-CW].

There are cases where service providers have existing deployments where the Provider Edge (PE) device is an old piece of equipment which does not support the CW for Ethernet PW encapsulation. In this case, the CW shall not be used as defined in [RFC 4448].

There are situations where replacing this PE with a new piece of equipment which supports CW for Ethernet PW is not acceptable because of economical or operational (e.g., service disruption time) reasons.

It may be beneficial to give operators an option to deploy (or re-use) another piece of equipment, located one hop away at the MPLS layer from this PE (typically physically co-located), which can add the CW to the Ethernet PW packets received from this PE, before sending them through an MPLS network.

This node should behave as a Switching PE (S-PE) as defined in [RFC 6073] and also be capable of switching an Ethernet pseudowire (PW) segment that uses the control word (CW) with an Ethernet PW segment that does not use the CW.

The reference network is shown in Figure 1 below.

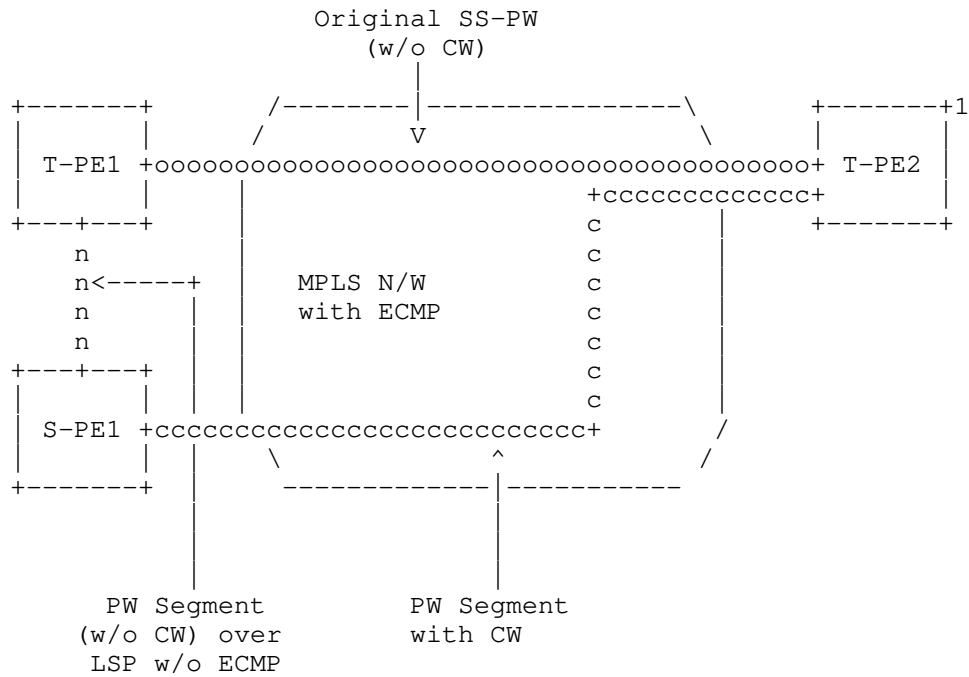


Figure 1 Reference Network

In Figure 1, T-PE1 is a device which is not capable of including a CW in Ethernet PW encapsulation, T-PE2 is a device which is capable to use the CW for Ethernet PW encapsulation, while S-PE1 is the new type of device defined in this document.

S-PE1 can be added to the network with minimum or no service disruption and PW redundancy [RFC 6718] or [RFC 7771] can be used to move the traffic from the old single-segment PW (SS-PW) without the CW to the new multi-segment PW (MS-PW) with the CW on the PW segment that passes through the MPLS network.

The deployment of the S-PE1, either as a new router or as an upgrade of an existing router, does not require any changes/upgrades to other nodes already installed within the network.

It is expected that in new deployments, all the Provider Edge (PE) devices are capable to insert the CW for Ethernet PW encapsulation and therefore the solution described in this document mainly applies to existing deployments where there are old pieces of equipment not being capable to support the CW for Ethernet PW encapsulation.

1.1. Assumptions

This document assumes that T-PE1 operates in the same way regardless of whether the PW is a SS-PW or a MS-PW, as defined in [RFC 6073]:

- o T-PE1 signals SS-PW with T-PE2 using T-LDP, as defined in [RFC 4447]
- o T-PE1 could be configured to signal a PW segment with S-PE1, as if it were T-PE2 using T-LDP, following the procedures defined in [RFC 6073].
- o T-PE1 is capable to set the PW-TTL value (i.e., the TTL value of the PW LSE) for Ethernet PW packets to a proper value that allows the Ethernet frames to be forwarded on the AC on T-PE2 (e.g., PW-TTL>2 in Figure 1): this could be done either via administrative configuration or through T-LDP information.

It is also assumed that if T-PE1 supports Pseudowire Virtual Circuit Connectivity Verification (VCCV), it can support at least CC Type 3 or CC Type 4. The underlying rationale for this assumption is that use of CC Type 2 for MS-PW is not allowed in [RFC 6073].

If T-PE1 supports CC Type 3, it is assumed that it is capable to set the PW-TTL value for the VCCV packets to a proper value that allows the VCCV packets to be recognized by T-PE2 by PW-TTL expiry (e.g., PW-TTL=2): this could be done either via administrative configuration or through T-LDP information.

It is assumed that S-PE1 is manually configured to switch between the two PW segments, following the procedure described in [RFC 6073].

If T-PE2 supports VCCV, it is configured to always advertise support for CC type 1. This would allow simplifying the VCCV switching process since CC type 1 is always used on the PW segment with CW.

2. Terminology

This document re-uses the terminology defined in [RFC 6073] for single-segment pseudo-wire (SS-PW), multi-segment PW (MS-PW), terminating provider edge (T-PE) and switching provider edge (S-PE).

This document uses the acronym PW-TTL to indicate the TTL value in the PW label stack entry (LSE).

2.1. Conventions Used in This Document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. Control Word Stitching procedures

The CW stitching procedure is performed by the S-PE1 on the Ethernet PW packets it is forwarding.

With a reference to Figure 1, it performs the following operations, in the direction from T-PE1 to T-PE2:

1. It pops the MPLS label stack entry (LSE) of the LSP from T-PE1 to S-PE1, if not PHP-ed by the penultimate LSR
2. It swaps the PW label (and decrements the PW-TTL)
3. It adds the CW immediately following the bottom of the label stack
4. It pushes the MPLS LSE for the LSP to T-PE2, unless this LSP is a single-hop PHP-ed LSP.

It is worth noting that step 3 is the only addition to the S-PE forwarding rules defined in [RFC 6073].

In this step, the S-PE inserts also the sequence number field in the control word, following the rules defined in [RFC4448].

In the opposite direction, S-PE1 performs the following operations:

1. It pops the MPLS label stack entry (LSE) for the LSP from T-PE2 to S-PE1, if not PHP-ed by the penultimate LSR
2. It swaps the PW label (and decrements the PW-TTL)
3. It removes the CW, which is located immediately following the bottom of the label stack
4. It pushes the MPLS LSE for the LSP to T-PE2, unless this LSP is a single-hop PHP-ed LSP.

It is worth noting that step 3 is the only addition to the S-PE forwarding rules defined in [RFC 6073].

In this step, the S-PE MAY also process the sequence number field in the control word, following the rules defined in [RFC4448].

3.1. CW Stitching Signaling

S-PE1 negotiates CW capabilities with T-PE1 and T-PE2 following almost the same procedures defined in [RFC 4447] and [RFC 6073].

The only exception to the procedures defined in [RFC 6073] is that S-PE1, when signaling one PW segment, will always behave as if the CW is supported on the other PW segment.

This allows S-PE1 to negotiate different CW capabilities on different PW segments as well as to enable CW toward any T-PE that support CW insertion.

If the same CW capabilities are negotiated on both PW segments, then S-PE1 will behave as specified in [RFC 6073]. CW stitching, as defined in this document, is enabled if and only if different CW capabilities are negotiated on the two PW segments.

In case the S-PE considers the sequence number field in the control word, it SHALL follow the rules described in section 6.4 of [RFC4447].

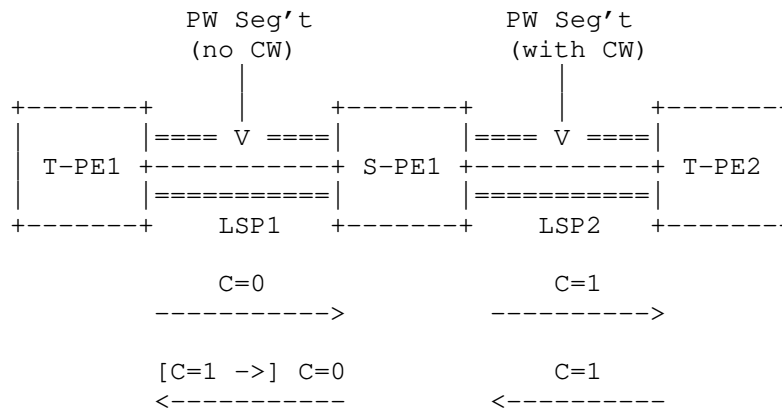


Figure 2 CW Stitching Signaling

Figure 2 shows an example of how CW capabilities are negotiated in the reference network scenario of Figure 1.

T-PE1 will send a T-LDP Label Mapping message with c=0 and T-PE2 will send a T-LDP Label Mapping message with C=1, following the procedures defined in section 6.2 of [RFC 4447] and amended by [I-D ETH-CW].

After S-PE1 receives the T-LDP Label Mapping message (with c=1) from T-PE2, it can send a T-LDP Label Mapping message back to T-PE2 (with c=1), following the procedures defined in section 6.2 of [RFC 4447], and a T-LDP Label Mapping messages to T-PE1 (with c=1), following the procedures of [RFC 6073].

After S-PE1 receives the T-LDP Label Mapping message (with c=0) from T-PE1, it can send a T-LDP Label Mapping message to T-PE2 (with c=1), as if it has received c=1 from T-PE1. It can also send a T-LDP Label Mapping message back to T-PE1 with c=0, following the procedures defined in section 6.2 of [RFC 4447].

If S-PE1 receives the T-LDP Label Mapping message (with c=0) from T-PE1 after having sent a T-LDP Label Mapping message with c=1 to T-PE1, a Label Withdraw message needs to be sent to T-PE1 before sending another Label Mapping message with c=1, as specified in section 6.2 of [RFC 4447].

When the MS-PW is completely setup:

- o T-PE1 is configured not to insert CW
- o T-PE2 is configures to insert CW
- o S-PE1 is configured to stitch the CW between the two PW segments

4. VCCV Stitching Procedures

When CW stitching is enabled, VCCV packets sent on the two PW segments would have different formats. In order to enable end-to-end OAM, S-PE1 needs to be capable to perform VCCV stitching.

Since support of CC Type 1 is REQUIRED by [RFC 5085] for PWs that support the CW, within this document it is RECOMMENDED that its use is always enabled at the T-PEs supporting the CW (e.g., T-PE2) such that, following the rules defined in [RFC 5085], when VCCV is in use, CC Type 1 is always used on the PW segment that support the CW.

Since [RFC 5085] does not define any mandatory CC Types for the PWs that do not support CW, different VCCV stitching procedures need to be defined depending on the CC Type supported by the T-PE not supporting the CW (e.g., T-PE1).

The VCCV stitching procedure is performed by S-PE1 on the VCCV packets it is forwarding.

In the traffic direction from T-PE2 and T-PE1 CC Type 1 is used: S-PE1 can distinguish VCCV and Ethernet PW packets by looking at the first nibble immediately following the bottom of the label stack which identifies either an ACH or a CW:

- o Ethernet PW packets are received with the CW: these packets need to be forwarded following the rules defined in section 3.
- o VCCV packets targeted at S-PE1 are received with the ACH and the PW-TTL=1: these packets should be processed by S-PE1 and not forwarded.
- o Other VCCV packets are received with the ACH and with a PW-TTL value greater than 1: these packets need to be forwarded following the rules defined in the following sections.

In the traffic direction from T-PE1 and T-PE2, the rules used to distinguish VCCV packets from Ethernet PW packets depends from the CC Type used on the PW segment without the CW.

4.1. VCCV Stitching for CC Type 3

In case CC Type 3 is used on the PW segment not using the CW, VCCV stitching needs to translate between CC Type 3 (without the CW) and CC Type 1. It is worth noting that when CC Type 3 is used on PW segments not using the CW, only IP-based CV types can be supported.

In the traffic direction from T-PE1 and T-PE2, S-PE1 can distinguish VCCV and Ethernet PW packets by looking at the PW-TTL value:

- o Ethernet PW packets are received with a PW-TTL value exceeding the PW-TTL distance from S-PE1 to T-PE2 (e.g., TTL>2): these packets need to be forwarded following the rules defined in section 3.
- o VCCV packets targeted at S-PE1 are received with PW-TTL=1: these packets should be processed by S-PE1 and not forwarded.
- o Other VCCV packets are received with a PW-TTL value greater than 1 and not exceeding the PW-TTL distance to T-PE2 (e.g., TTL=2): these packets need to be forwarded following the rules defined in this section.

With a reference to Figure 1, S-PE1 performs the following operations, in the direction from T-PE1 to T-PE2:

1. It pops the MPLS label stack entry (LSE) of the LSP from T-PE1 to S-PE1, if not PHP-ed by the penultimate LSR
2. It swaps the PW label (and decrements the PW-TTL)
3. It adds the ACH immediately following the bottom of the label stack (setting the ACH Channel Type based on the IP version field of the encapsulated IP packet)
4. It pushes the MPLS LSE for the LSP to T-PE2, unless this LSP is a single-hop PHP-ed LSP.

It is worth noting that step 3 is the only addition to the S-PE forwarding rules defined in [RFC 6073]: it is also the only step where the forwarding rules of VCCV packets are different from the forwarding rules defined for Ethernet PW packets in section 3.

S-PE1 can understand the IP version field of the encapsulated IP packet by looking at the first nibble immediately following the bottom of the label stack of the received packet.

In the opposite direction, S-PE1 performs the following operations:

1. It pops the MPLS label stack entry (LSE) for the LSP from T-PE2 to S-PE1, if not PHP-ed by the penultimate LSR
2. It swaps the PW label (and decrements the PW-TTL)
3. It removes the ACH, which is located immediately following the bottom of the label stack
4. It pushes the MPLS LSE for the LSP to T-PE2, unless this LSP is a single-hop PHP-ed LSP.

It is worth noting that step 3 is the only addition to the S-PE forwarding rules defined in [RFC 6073]: it is also the only step where the forwarding rules of VCCV packets are different from the forwarding rules defined for Ethernet PW packets in section 3.

4.2. VCCV Stitching for CC Type 4

In case CC Type 4 is used on the PW segment not using the CW, VCCV stitching needs to translate between CC Type 4 and CC Type 1. It is

worth noting that in this case both IP-based and ACH-based CV types can be supported.

In the traffic direction from T-PE1 and T-PE2, S-PE1 can distinguish VCCV and Ethernet PW packets by looking at GAL LSE right after the PW LSE:

- o Ethernet PW packets are received without a GAL LSE: these packets need to be forwarded following the rules defined in section 3.
- o VCCV packets targeted at S-PE1 are received with the GAL LSE and with the PW-TTL=1: these packets should be processed by S-PE1 and not forwarded.
- o Other VCCV packets are received with the GAL LSE and with a PW-TTL value greater than 1: these packets need to be forwarded following the rules defined in this section.

With a reference to Figure 1, S-PE1 performs the following operations, in the direction from T-PE1 to T-PE2:

1. It pops the MPLS label stack entry (LSE) of the LSP from T-PE1 to S-PE1, if not PHP-ed by the penultimate LSR
2. It swaps the PW label (and decrements the PW-TTL)
3. It removes the GAL LSE at the bottom of the label stack
4. It sets the S-bit of the PW LSE since the PW LSE becomes the new bottom of the label stack
5. It pushes the MPLS LSE for the LSP to T-PE2, unless this LSP is a single-hop PHP-ed LSP.

It is worth noting that steps 3 and 4 are the only additions to the S-PE forwarding rules defined in [RFC 6073]: they are also the only steps where the forwarding rules of VCCV packets are different from the forwarding rules defined for Ethernet PW packets in section 3.

In the opposite direction, S-PE1 performs the following operations:

1. It pops the MPLS label stack entry (LSE) for the LSP from T-PE2 to S-PE1, if not PHP-ed by the penultimate LSR
2. It swaps the PW label (and decrements the PW-TTL)
3. It inserts the GAL LSE at the bottom of the label stack

4. It clears the S-bit of the PW LSE since the PW LSE is no longer at the bottom of the label stack
5. It pushes the MPLS LSE for the LSP to T-PE2, unless this LSP is a single-hop PHP-ed LSP.

It is worth noting that steps 3 and 4 are the only additions to the S-PE forwarding rules defined in [RFC 6073]: they are also the only steps where the forwarding rules of VCCV packets are different from the forwarding rules defined for Ethernet PW packets in section 3.

4.3. VCCV Stitching Signaling

S-PE1 negotiates VCCV capabilities with T-PE1 and T-PE2 following almost the same procedures defined in [RFC 5085] and [RFC 6073].

If the same CW capabilities are negotiated on both PW segments, then S-PE1 will behave as specified in [RFC 6073]. VCCV stitching, as defined in this document, is enabled if and only if different CW capabilities are negotiated on the two PW segments.

If S-PE1 supports VCCV stitching for CC Type 3, and it knows the PW-TTL distance to both T-PE1 and T-PE2:

- o If T-PE1 advertises support for CC Type 3, S-PE1 advertises support for CC Type 1 to T-PE2
- o If T-PE2 advertises support for CC Type 1, S-PE1 behaves toward T-PE1 if it supports CC Type 3 and T-PE2 has advertised support for CC Type 3, following the procedure defined in [RFC 6073]

If S-PE1 supports VCCV stitching for CC Type 4:

- o If T-PE1 advertises support for CC Type 4, S-PE1 advertises support for CC Type 1 to T-PE2
- o If T-PE2 advertises support for CC Type 1, S-PE1 behaves toward T-PE1 as if it supports CC Type 4 and T-PE2 has advertised support for CC Type 4, following the procedure defined in [RFC 6073]

CV types are advertised based on S-PE1 capabilities as per [RFC 6073] with the following additional rule:

- o S-PE1 can advertise support for ACH-based CV types if and only if it supports VCCV stitching for CC Type 4

This rule ensures that only IP-based CV types are negotiated between T-PE1, T-PE2 and S-PE1 when VCCV stitching for CC Type 3 is used.

If T-PE1 supports CC Type 4 and S-PE1 supports VCCV stitching for CC Type 4, then VCCV stitching for CC Type 4 is used and both IP-based and ACH-based CV capabilities can be negotiated depending on T-PE1, T-PE2 and S-PE1 CV capabilities.

If T-PE1 does not support CC Type 4, it will advertise support only for IP-based CV types and therefore only IP-based CV capabilities can be negotiated depending on T-PE1, T-PE2 and S-PE1 CV capabilities.

If S-PE1 does not support VCCV stitching for CC Type 4, it will advertise support only for IP-based CV types and therefore only IP-based CV capabilities can be negotiated depending on T-PE1, T-PE2 and S-PE1 CV capabilities.

5. Other Deployment Scenarios

The solution described in this document is quite generic and can be used in different deployment scenarios, in addition to the reference network outline in Figure 1, without requiring any change to the behavior of the S-PE, as defined in this document.

A possible deployment scenario is shown in Figure 3 where both T-PEs are not capable to insert the CW:

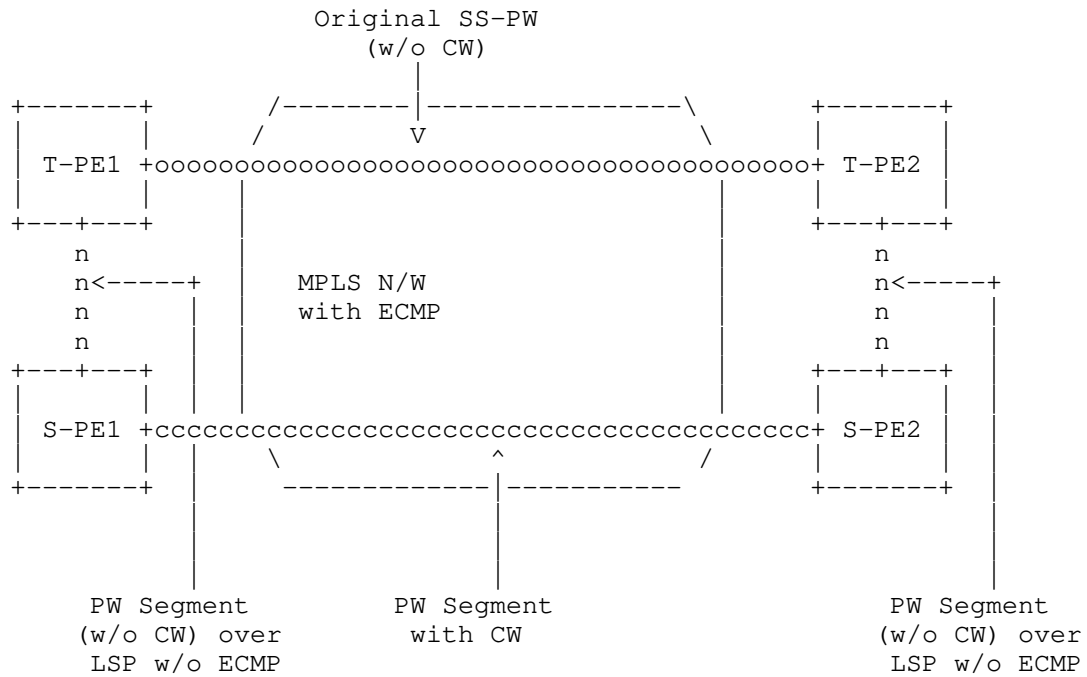


Figure 3 Reference network with two T-PEs not capable to insert CW

In this scenario, two S-PEs needs to be deployed: S-PE1 in front of T-PE1 and S-PE2 in front of T-PE2.

S-PE1 and S-PE2 operate as defined in this document: these operations are the same even if one or both the PW segments switched by one S-PE are terminated at a T-PE or at another S-PE.

An even more generic deployment scenario is shows in Figure 3:

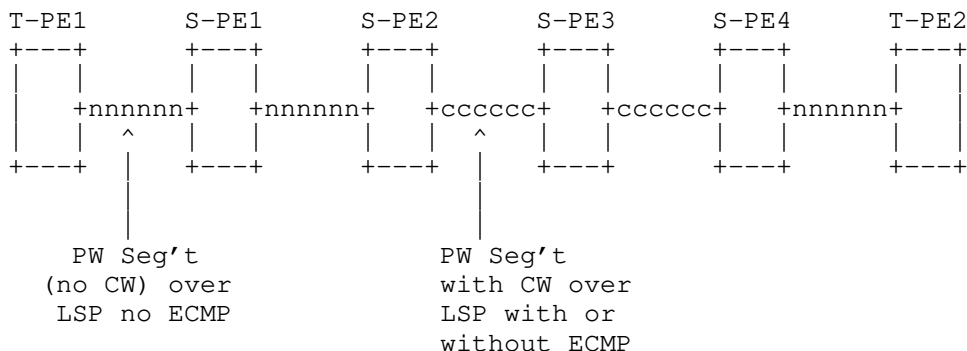


Figure 4 More generic Reference network

In this case a MS-PW can be setup with some PW segments using the CW and other not using the CW.

S-PE1 and S-PE3 operates as defined in [RFC 6073] while S-PE2 and S-PE4 operate as defined in this document: these operations are the same even if one or both the PW segments switched by one S-PE are terminated at a T-PE or at another S-PE operating as defined in [RFC 6073] or at another S-PE operating as defined in this document.

The operations are also the same if the PW segment not using the CW is setup over a link or over an MPLS network.

In order to achieve the desired behavior, i.e., to avoid the issues described in [I-D ETH-CW], care must be taken by the operator to make sure that no ECMP is used within the MPLS network carrying the PW segments without the CW.

The operations described in this document work also if static configuration is used instead of T-LDP to setup some or all the PW segments.

The operations described in this document work also if dynamic MS-PW signaling procedures, as defined in [RFC7267], are used instead of static configuration of the S-PEs.

6. Security Considerations

The method described in this document adds no security issues beyond those encountered in a network running multi-segment Ethernet pseudowires with the Control Word over MPLS, as previously discussed in [RFC4385], [RFC4448] and [RFC7267]. Such networks are normally private, well managed, highly controlled environments.

7. IANA Considerations

This document makes no IANA requests.

8. References

8.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC4385] Bryant, S. et al., "Pseudowire Emulation Edge-to-Edge (PWE3) Control Word for Use over an MPLS PSN", RFC 4385, February 2006.
- [RFC4447] Martini, L. et al., "Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP)", RFC 4447, April 2006.
- [RFC4448] Martini, L. et al., "Encapsulation Methods for Transport of Ethernet over MPLS Networks", RFC 4448, April 2006.
- [RFC5085] Nadeu, T., Pignataro, C., " Pseudowire Virtual Circuit Connectivity Verification (VCCV): A Control Channel for Pseudowires", RFC 5085, December 2007.
- [RFC6073] Martini, L. et al., "Segmented Pseudowire", RFC 6073, January 2011.
- [RFC7267] Martini, L. et al., "Dynamic Placement of Multi-Segment Pseudowires", RFC 7267, June 2014.
- [RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, May 2017.
- [I-D ETH-CW] Bryant, S. et al., "Use of Ethernet Control Word RECOMMENDED", draft-ietf-pals-ethernet-cw, work in progress.

8.2. Informative References

- [RFC6718] Muley, P. et al., "Pseudowire Redundancy", RFC 6718, August 2012.

[RFC7771] Malis, A. et al., "Switching Provider Edge (S-PE) Protection for MPLS and MPLS Transport Switching Provider Edge (S-PE) Protection for MPLS and MPLS Transport", RFC 7771, January 2016.

9. Acknowledgments

This document was prepared using 2-Word-v2.0.template.dot.

Authors' Addresses

Italo Busi
Huawei

Email: italo.busi@huawei.com

Stewart Bryant
Huawei

Email: stewart.bryant@gmail.com

Andrew G. Malis
Huawei

Email: agmalis@gmail.com

Jie Dong
Huawei

Email: jie.dong@huawei.com

