

Network Working Group
Internet Draft
Intended Status: Standards Track
Expires: March 2009

Praveen Muley
Mustapha Aissaoui
Matthew Bocci
Pranjal Kumar Dutta
Marc Lasserre
Alcatel-Lucent

Jonathan Newton
Cable & Wireless

Olen Stokes
Extreme Networks

Hamid Ould-Brahim
Nortel

Luca Martini
Cisco Systems Inc.

Giles Heron
Thomas Nadeau
BT

September 30, 2008

Preferential Forwarding Status bit definition
draft-ietf-pwe3-redundancy-bit-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 [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/lid-abstracts.html>

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on March 30, 2009.

Abstract

This document describes a mechanism for standby status signaling of redundant PWs between their termination points. 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 bit is needed to indicate a preferential forwarding status of active or standby for each PW in the redundancy set.

In addition, a second status bit is defined to allow peer PE/T-PE nodes to coordinate a switchover operation of the PW/MS-PW path.

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.....[3](#)
- [2](#). Motivation and Scope.....[4](#)
- [3](#). Terminology.....[7](#)
- [4](#). Modes of Operation.....[7](#)
 - [4.1](#). Independent Mode:.....[8](#)
 - [4.2](#). Master/Slave Mode:.....[9](#)
- [5](#). Signaling procedures of PW State Transition.....[10](#)
 - [5.1](#). PW Standby notification procedures in Independent mode...[10](#)
 - [5.2](#). PW Standby notification procedures in Master/Slave mode..11
 - [5.2.1](#). PW State Machine.....[12](#)

5.3. Coordination of PW Path Switchover.....	14
5.3.1. Procedures at the requesting endpoint.....	15
5.3.2. Procedures at the receiving endpoint.....	16
6. Applicability and Backward Compatibility.....	17
7. Security Considerations.....	17
8. IANA Considerations.....	17

8.1. Status Code for PW Preferential Forwarding Status.....	17
8.2. Status Code for PW Request Switchover Status.....	18
9. Acknowledgments.....	18
10. References.....	18
10.1. Normative References.....	18
10.2. Informative References.....	18
11. Appendix A - Applications of PW Redundancy Procedures.....	19
11.1. One Multi-homed CE with single SS-PW redundancy.....	19
11.2. Multiple Multi-homed CEs with single SS-PW redundancy...	21
11.3. Multi-homed CE with MS-PW redundancy.....	22
11.4. Single Homed CE with MS-PW redundancy.....	24
11.5. PW redundancy between H-VPLS MTU-s and PE-rs.....	25
Author's Addresses.....	27
Full Copyright Statement.....	28
Intellectual Property Statement.....	28
Acknowledgment.....	29

[1](#). Introduction

In single-segment PW (SS-PW) services such as VPWS and VPLS, protection for the PW is provided by the PSN layer. This may be an RSVP-TE LSP with a FRR backup and/or an end-to-end backup LSP. There are however applications where PSN protection is insufficient to fully protect the PWE3 service and pseudowire redundancy is required. These scenarios are described in the PW redundancy and framework document [\[5\]](#).

In a VPWS service, this is used to provide access AC redundancy to a CE device which is dual-homed to target PE nodes. In a HVPLS service, this is used to provide access PW redundancy to the MTU device which is dual-homed to two PE-r devices. PSN protection mechanisms cannot protect against failure of the target PE node or the failure of the remote AC. These scenarios rely on a set of two or more pseudowires to protect a given PWE3 service. Only one of these pseudowires is used by the PEs to forward user traffic on at any given time. This is the active PW. The other PWs in the set are considered standby and

are not used for forwarding unless they become active. In essence, this provides a 1:1 or N:1 PW protection with the possibility of multi-homing.

In order to support access AC or access PW redundancy, at least one of the PEs on which a PW terminates must be different from that on which the primary PW terminates, as described in [5]. Figure 1 illustrates an application of Active and Standby PWs.

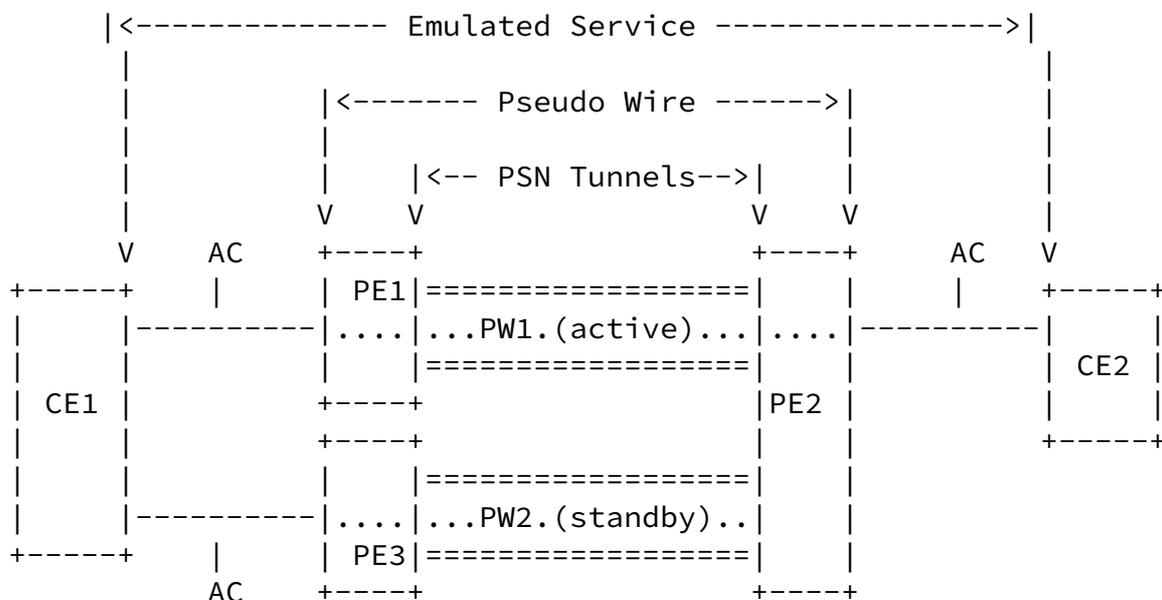


Figure 1: Reference Model for PW Redundancy

In multi-segment PW (MS-PW) applications, multiple MS-PWs are configured between a pair of T-PE nodes. The paths of these MS-PWs are diverse and are switched at different S-PE nodes. Only one of these MS-PWs is active at any given time. The others are put in standby. In these applications, 1:1 or N:1 PW redundancy is important to provide resilience in the event of failure of S-PE node since PSN protection mechanisms cannot, and the desire is that MS-PW based services have resiliency similar to that of SS-PW based services.

This document specifies a new PW status bit to indicate the preferential forwarding status of the PW for the purpose of notifying the remote PE of the active/standby state of each PW in the redundancy set. This status bit is different from the operational status bits already defined in the PWE3 control protocol [2]. In addition, a second status bit is defined to allow peer PE/T-PE nodes to coordinate a switchover operation of the PW/MS-PW path.

2. Motivation and Scope

The PWE3 control protocol [2] defines the following status codes in PW the status TLV to indicate the operational state for an AC and a PW:

0x00000000 - Pseudowire forwarding (clear all failures)

0x00000001 - Pseudowire Not Forwarding

0x00000002 - Local Attachment Circuit (ingress) Receive Fault

0x00000004 - Local Attachment Circuit (egress) Transmit Fault

0x00000008 - Local PSN-facing PW (ingress) Receive Fault

0x00000010 - Local PSN-facing PW (egress) Transmit Fault

The scenarios defined in [5] allow the provisioning of a primary PW and one or many secondary PWs in the same VPWS or VPLS service.

A PE node makes a selection of which PW to activate at any given time for the purpose of forwarding user packets. This selection takes into account the local operational state of the PW as well as the remote operational state of the PW as indicated in the status bits of the PW it received from the peer PE node.

In the absence of faults, all PWs are operationally UP both locally and remotely and a PE node needs to select a single PW to forward user packets to. This is referred to as the active PW. All other PWs will be in standby and must not be used to forward user packets.

In order for both ends of the service to select the same PW for forwarding user packets, it is proposed that a PE node communicates a new status bit to indicate the forwarding state of a PW to its peer

PE node.

In addition, a second status bit is defined to allow peer PE/T-PE nodes to coordinate a switchover operation of the PW/MS-PW path if required by the application.

Together, the mechanisms described in this document achieve the following PW protection capabilities:

- a. A mandatory 1:1 PW protection with a single active PW and one standby PW. An active PW can forward data traffic and control plane traffic, such as OAM packets. A standby PW does not carry data traffic.
- b. An optional N:1 PW protection scheme with a single active PW and N standby PWs.
- c. An independent mode of operation in which each PW endpoint node uses its own local rule to select which PW it intends to activate at any given time and advertises it to the remote endpoints. Only a PW which is operationally UP and which indicated Active status bit locally and remotely is in the

Active state and can be used to forward data packets. This is described in [Section 4.1](#). .

- d. An optional mechanism to allow PW endpoints to coordinate the switchover to a given PW path by using an explicit request/acknowledgment switchover procedure. This mechanism is complementary to the Independent mode of operation and is described in [Section 5.3](#). . This mechanism can be invoked manually by the user, effectively providing a manual switchover capability. It can also be invoked automatically to resolve a situation where the PW endpoints could not match the two directions of the PW.
- e. A Master/Slave mode in which one PW endpoint, the Master endpoint, selects and dictates to the other endpoint(s), the Slave endpoint(s), which PW to activate. This is described in [Section 4.2](#). .
- f. Multi-homing of a CE device to two or more PE/T-PE nodes.

- g. Multi-homing of a PE/T-PE node to two or more PE/T-PE nodes.
- h. Optionally, implementations can add support for precedence parameter to govern the selection of a PW when more than one PW qualify for the active state, as defined in sections [4.1.](#) and 4.2. The PW with the lowest precedence value has the highest priority. This is a local configuration parameter to the PW endpoint.
- i. Optionally, implementations can designate by configuration one PW in the 1:1 or N:1 set as a primary PW and the remaining as secondary PWs. If more than one PW qualify for the active state, as defined in sections [4.1.](#) and 4.2. , a PE/T-PE node selects the primary PW in preference to a secondary PW. In other words, the primary PW has implicitly the lowest precedence value. Furthermore, implementations can provide the option to revert to the primary PW immediately after it comes back up or after the expiration of a delay. The PE/T-PE node can use the PW precedence parameter to select a secondary PW among many that qualify for active state.

[Appendix A](#) shows how the mechanisms described in this document are used to achieve the desired protection behavior for the scenarios described in the PW redundancy and framework document [[5](#)].

[3.](#) Terminology

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 [[2](#)]. Such a PW is available for forwarding traffic.

DOWN PW: A PW that has either not been fully configured or has been and is in any of the PW defect states specified in [[2](#)]. Such a PW is not available for forwarding traffic.

Active PW: An UP PW used for forwarding user and OAM traffic.

Standby PW: An UP PW that is not used for forwarding user traffic, but may forward OAM traffic.

Primary PW: the PW which a PW endpoint activates in preference to any other PW when more than one PW qualify for active state. When the primary PW comes back up after a failure and qualifies for active state, the PW endpoint always reverts to it. The designation of Primary is performed by local configuration at the PW endpoint.

Secondary PW: when it qualifies for 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 operationally Down). By default, a PW in a redundancy PW set is considered secondary. There is no Revertive mechanism among secondary PWs.

PW Precedence: this is a configuration parameter that dictates the order in which a PW endpoint activates a PW among multiple PWs which qualify for active state. The lowest the precedence value, the highest is the priority of selecting the PW.

PW Endpoint: A PE where a PW terminates on an NSP e.g. A SS-PW PE, an MS-PW T-PE, or a VPLS MTU or PE-r.

4. Modes of Operation

There are two modes of operation for the use of the PW preferential forwarding status bits:

- o Independent mode
- o Master/Slave mode.

4.1. Independent Mode:

PW endpoint nodes independently select which PW they intend to make active and which PWs they intend to make standby. They advertise the corresponding Active/Standby forwarding state for each PW. Each PW endpoint compares local and remote status and uses the PW that is operationally UP at both endpoints and that shows Active states at both the local and remote endpoint.

If more than one PW qualify for the Active state, and the PW endpoint

implements the precedence parameter, it must use the PW with the lowest precedence value. The precedence parameter is optional.

If more than one PW qualify for the Active state, and the PW endpoint implements the primary/secondary procedures, it must use the primary PW in preference to all secondary PWs. If a primary PW is not available, it must use the secondary PW with the lowest precedence value. If the primary PW becomes available, a PW endpoint may revert to it immediately or after the expiration of a configurable delay. These primary/secondary procedures are optional.

In steady state with consistent configuration, a PE will always find an Active PW. However, it is possible that due to a misconfiguration, such a PW is not found. In the event that an active PW is not found, a management indication should be generated. If a management indication for failure to find an active PW was generated and an active PW is subsequently found, a management indication should be generated, effectively clearing the previous failure indication. Additionally, a PE may use the optional request switchover procedures described in [Section 5.3](#). to have both PE nodes switch to a common PW path.

There may also be transient conditions where endpoints do not share a common view of the active/standby state of the PWs. This could be caused by propagation delay of the T-LDP status messages between endpoints. In this case, the behavior of the receiving endpoint is outside the scope of this document.

Thus, in this mode of operation, the following definition of Active and Standby PW states apply:

- o Active State

A PW is considered to be in Active state when the PW labels are exchanged between its two endpoints in control plane, and the status bits exchanged between the endpoints indicate the PW is UP and Active

at both endpoints. In this state user traffic can flow over the PW in both directions.

- o Standby State

A PW is considered to be in Standby state when the PW labels are exchanged between its two endpoints in the control plane, but the status bits exchanged indicate the PW is in Standby state at one or both endpoints. In this state the endpoints MUST NOT forward data traffic over the PW but MAY allow PW OAM packets, e.g., VCCV, to be sent and received in order to test the liveness of standby PWs. If the PW is a spoke in H-VPLS, any MAC addresses learned via the PW should be flushed when it transitions to Standby state.

4.2. Master/Slave Mode:

One endpoint node of the redundant set of PWs is designated the Master and is responsible for selecting which PW both endpoints must use to forward user traffic.

The Master indicates the forwarding state in the Active/Standby status bit. The other endpoint node, the Slave, MUST follow the decision of the Master node based on the received status bits.

One endpoint of the PW, the Master, actively selects which PW to activate and uses it for forwarding user traffic. This status is indicated to the Slave node by setting the preferential forwarding status bit in the status bit TLV to Active. It does not forward user traffic to any other of the PW's in the redundancy set to the slave node and indicates this by setting the preferential forwarding status bit in the status bit TLV to Standby for those PWs. The master node MUST ignore any Active/Standby status bits received from the Slave nodes.

If more than one PW qualify for the Active state, and the PW endpoint implements the precedence parameter, it must use the PW with the lowest precedence value. The precedence parameter is optional.

If more than one PW qualify for the Active state, and the PW endpoint implements the primary/secondary procedures, it must use the primary PW in preference to all secondary PWs. If a primary PW is not available, it must use the secondary PW with the lowest precedence value. If the primary PW becomes available, a PW endpoint may revert to it immediately or after the expiration of a configurable delay. These primary/secondary procedures are optional. The Slave endpoint(s) are required to act on the status bits received from the Master. When the received status bit transitions from Active to

Standby, a Slave node MUST stop forwarding over the previously active PW. When the received status bit transitions from Standby to Active for a given PW, the Slave node MUST start forwarding user traffic over this PW.

There is a single PE/T-PE Master PW endpoint node and one or many PE/T-PE PW endpoint Slave nodes. The assignment of Master/Slave roles to the PW endpoints is performed by local configuration. Note that the above behavior assumes correct configuration of the Master and Slave endpoints. This document does not define a mechanism to detect errors in the configuration.

In this mode of operation, the following definition of Active and Standby PW states apply:

- o Active State

A PW is considered to be in Active state when the PW labels are exchanged between its two endpoints in control plane, and the status bits exchanged between the endpoints indicate the PW is UP at both endpoints, and the forwarding status sent by the Master endpoint indicates Active state. In this state user traffic can flow over the PW in both directions.

- o Standby State

A PW is considered to be in Standby state when the PW labels are exchanged between its two endpoints in the control plane, but the status bits sent by the Master endpoint indicate the PW is in Standby state. In this state the endpoints MUST NOT forward data traffic over the PW but MAY allow PW OAM packets, e.g., VCCV, to be sent and received. If the PW is a spoke in H-VPLS, any MAC addresses learned via the PW should be flushed when it transitions to Standby state.

5. Signaling procedures of PW State Transition

This section describes the extensions proposed and the processing rules for the extensions. It defines a new "PW preferential forwarding" bit in Status Code that is to be used with PW Status TLV proposed in [RFC 4447](#) [2]. The PW preferential forwarding bit when set is used to signal Standby state of PW by one terminating point to the other end.

5.1. PW Standby notification procedures in Independent mode

PW endpoint nodes independently select which PW they intend to use for forwarding, and which PWs they do not, based on the specific

application. They advertise the corresponding Active/Standby forwarding state for each PW. An active forwarding state is indicated by clearing the Active/Standby status bit in the PW status TLV. A standby forwarding state is indicated by setting the Active/Standby status bit in the PW status TLV. This advertisement occurs in both the initial label mapping message and in a subsequent notification message when the forwarding state transitions as a result of a state change in the specific application.

Each endpoint compares the updated local and remote status and effectively activates the PW which is operationally UP at both endpoints and which shows both local Active and remote Active states.

When a PW is in active state, the endpoints can forward both user packets and OAM packets.

When a PW is in standby state, the endpoints MUST NOT forward user packets over the PW but MAY forward PW OAM packets.

For MS-PWs, S-PEs MUST relay the PW status notification containing both the operational and preferential forwarding status bits between ingress and egress PWs.

5.2. PW Standby notification procedures in Master/Slave mode

Whenever the Master PW endpoint "actively" selects or deselects a PW for forwarding user traffic at its end, it explicitly notifies the event to the remote Slave endpoint. The slave endpoint carries out the corresponding action on receiving the PW state change notification.

If the PW preferential forwarding bit in PW Status TLV received by the slave is set, it indicates that the PW at the Master end is not used for forwarding and is thus kept in the Standby state, the PW MUST also not be used for forwarding at Slave endpoint. Clearance of the PW Preferential Forwarding bit in PW Status TLV indicates that the PW at the Master endpoint is used for forwarding and is in Active state, and the receiving Slave endpoint MUST activate the PW if it was previously not used for forwarding.

This mechanism is RECOMMENDED to be used with PWs signaled in groups with common group-id in Pwid FEC Element or Grouping TLV in Generalized Pwid FEC Element defined in [2]. When PWs are provisioned with such grouping a termination point sends a single "wildcard" Notification message using a PW FEC TLV with only the group ID set, to denote this change in status for all affected PW connections. This status message contains either the PW FEC TLV with only the Group ID

Internet-Draft Preferential Forwarding Status Bit September 2008

set, or else it contains the PW Generalized FEC TLV with only the PW Grouping ID TLV. As mentioned in [2], the Group ID field of the PwID FEC Element, or the PW Grouping TLV used with the Generalized ID FEC Element, can be used to send status notification for all arbitrary set of PWs. For example, Group-ID in PwID may be used to send wildcard status notification message for a group of PWs when PwID FEC element is used for PW state signaling. When Generalized PwID FEC Element defined is used in PW state signaling, PW Grouping TLV may be used for wildcard status notification for a group of PWs.

For MS-PWs, S-PEs MUST relay the PW status notification containing both the operational and preferential forwarding status bits between ingress and egress PW segments.

[5.2.1.](#) PW State Machine

It is convenient to describe the PW state change behavior in terms of a state machine. The PW state machine is explained in detail in the two defined states and the behavior is presented as a state transition table. The same state machine is seamlessly applicable to PW Groups.

PW State Transition State Table

STATE	EVENT	NEW STATE
ACTIVE	PW put in Standby (master)	STANDBY
	Action: Transmit PW preferential forwarding bit set	
	Receive PW preferential forwarding bit set	STANDBY

Action: Stop forwarding over PW

Receive PW preferential forwarding ACTIVE

Muley et al.

Expires March 30, 2009

[Page 12]

Internet-Draft

Preferential Forwarding Status Bit

September 2008

bit set but bit not supported

Action: None

Receive PW preferential forwarding ACTIVE

bit clear

Action: None.

STANDBY

PW activated (master)

ACTIVE

Action: Transmit PW preferential forwarding bit clear

Receive PW preferential forwarding ACTIVE

bit clear

Action: Activate PW

Receive PW preferential forwarding STANDBY

bit clear but bit not supported

Action: None

Receive PW preferential forwarding STANDBY

bit set

Action: No action

Muley et al.

Expires March 30, 2009

[Page 13]

Internet-Draft

Preferential Forwarding Status Bit

September 2008

[5.3](#). Coordination of PW Path Switchover

There are PW redundancy applications which require that PE/T-PE nodes coordinate the switchover to a PW/MS-PW path such that both endpoints will be forwarding over the same path at any given time. One such application of redundant MS-PW paths is identified in [\[5\]](#). Multiple MS-PWs are configured between a pair of T-PE nodes. The paths of these MS-PWs are diverse and are switched at different S-PE nodes. Only one of these MS-PWs is active at any given time. The others are put in standby. The endpoints follow the Independent Mode procedures to activate the PW which is UP and advertised Active 'preferential forwarding' status bit by both endpoints.

The trigger for sending a request to switchover of the path of the MS-PW by one endpoint can be due to an operational event, example a failure, which caused the endpoints to not be able to match the Active 'preferential forwarding' status bit. The other trigger is the execution of an administrative maintenance operation by the network operator in order to move the traffic away from the node/links to be serviced.

Unlike the case of a Master/Slave mode of operation, the endpoint requesting the switchover requires explicit acknowledgement from the peer endpoint that the request is honored before it switches the path of the PW. Furthermore, any of the endpoints can make the request to switchover.

A new status bit is proposed to have a PE/T-PE node request the switchover to its peer. This bit will be referred to as 'request PW switchover' status bit. The 'preferential forwarding' status bit continues to be used by each endpoint to indicate its current local settings of the active/standby state of each PW in the redundancy

set. In other words, like in the Independent mode, it indicates to the far-end which of the PWs is being used to forward packets and which is being put in standby. It can thus be used as a way for the far-end to acknowledge the requested switchover operation.

The following procedures must be followed by both endpoints of a PW/MS-PW to coordinate the switchover of the PW/MS-PW path. These procedures are enabled only when the user configured the use of the 'request switchover' status bit at both endpoints.

S-PEs nodes MUST relay the PW status notification containing the operational status bits, as well as the 'preferential forwarding' and 'request switchover' status bits between ingress and egress PW segments.

[5.3.1](#). Procedures at the requesting endpoint

- a. The requesting endpoint sends a Status TLV in the LDP notification message with the 'request switchover' bit set on the PW it desires to switch to.
- b. The endpoint does not activate forwarding on that PW/MS-PW at this point in time. It may however enable receiving on that PW/MS-PW. Thus the 'preferential forwarding' status bit still reflects the currently used PW path.
- c. The requesting endpoint starts a timer while waiting the remote endpoint to acknowledge the request.
- d. If while waiting for the acknowledgment, the requesting endpoint receives a request from its peer to switchover to the same or a different PW path, it must perform the following:
 - i. If its system IP address is higher than that of the peer, this endpoint ignores the request and continues to wait for the acknowledgement from its peer.
 - ii. If its system IP address is lower than that of its peer, it aborts the timer and immediately starts the procedures of the receiving endpoint in [Section 5.3.2](#).
- e. If while waiting for the acknowledgment, the requesting

endpoint receives a status notification message from its peer with the 'preferential forwarding' status bit cleared in the requested PW and set in all other PWs, it must treat this as an explicit acknowledgment of the request and must perform the following:

- i. Abort the timer.
 - ii. Activate the PW path.
 - iii. Send an update status notification message with the 'preferential forwarding' status bit clear on the newly active PW and set in all other PWs in the redundancy set, and the 'request switchover' bit reset in all PWs in the redundancy set..
- f. If while waiting for the acknowledgment, the requesting endpoint detects that the requested PW went into operational Down state locally, and could use an alternate PW which is operationally UP, it must perform the following:

- i. Abort the timer.
 - ii. Issue a new request to switchover to the alternate PW.
 - iii. Re-start the timer.
- g. If while waiting for the acknowledgment, the requesting endpoint detects that the requested PW went into operational Down state locally, and could not use an alternate PW which is operationally UP, it must perform the following:
- i. Abort the timer.
 - ii. Send an update status notification message with the 'preferential forwarding' status bit unchanged and the 'request switchover' bit reset in all PWs in the redundancy set.
- h. If while waiting for the acknowledgment, the timer expired, the requesting endpoint assumes the request is rejected and will either issue a new request or do nothing.

- i. If the requesting node receives the acknowledgment after the request expired, it will treat it as if the remote endpoint unilaterally switched the path of the PW without issuing a request. In that case, it may issue a new request and follow the requesting endpoint procedures to synchronize transmit and receive paths of the PW.

5.3.2. Procedures at the receiving endpoint

- a. Upon receiving a status notification message with the 'request switchover' bit set on a PW different from the currently active one, and the requested PW is operationally UP, the receiving endpoint must perform the following:
 - i. Activate the PW.
 - ii. Send an update status notification message with the 'preferential forwarding' status bit clear on the newly active PW and set in all other PWs in the redundancy set, and the 'request switchover' bit reset in all PWs in the redundancy set.
- b. Upon receiving a status notification message with the 'request switchover' bit set on a PW different from the currently active

one, and the requested PW is operationally Down, the receiving endpoint must perform the following:

- i. Ignore the request and do nothing.

6. Applicability and Backward Compatibility

The mechanism defined in this document is OPTIONAL and is applicable to PWE3 applications where standby state signaling of PW or PW group is required.

A PE implementation that uses the mechanisms described in this document MUST negotiate the use of PW status TLV between its T-LDP peers as per [RFC 4447](#) [2]. If PW Status TLV is found to be not supported by either of its endpoint after status negotiation procedures, then the mechanisms specified in this document cannot be used.

A PE implementation compliant to [RFC 4447 \[2\]](#), and which does not support the generation or processing of the 'preferential forwarding' status bit or of the 'request switchover' status bit, will not set these bits in the status bits transmitted to a peer PE and will not examine them in the received status bits from a peer PE. The mechanisms specified in this document cannot be used.

[7.](#) Security Considerations

This document uses the LDP extensions that are needed for protecting pseudo-wires. It will have the same security properties as in the PWE3 control protocol [\[2\]](#).

[8.](#) IANA Considerations

We have defined the following codes for the pseudo-wire redundancy application.

[8.1.](#) Status Code for PW Preferential Forwarding Status

0x00000020 When the bit is set, it indicates "PW forwarding standby".

When the bit is cleared, it indicates "PW forwarding

Muley et al.

Expires March 30, 2009

[Page 17]

Internet-Draft

Preferential Forwarding Status Bit

September 2008

active".

[8.2.](#) Status Code for PW Request Switchover Status

0x00000040 When the bit is set, it represents "Request switchover to this PW".

When the bit is cleared, it represents no specific action.

[9.](#) Acknowledgments

The authors would like to thank Vach Kompella, Kendall Harvey, Tiberiu Grigoriu, John Rigby, Prashanth Ishwar, Neil Hart, Kajal Saha, Florin Balus, Philippe Niger, Dave McDysan, and Roman Krzanowski for their valuable comments and suggestions.

10. References

10.1. Normative References

- [1] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [2] Martini, L., et al., "Pseudowire Setup and Maintenance using LDP", [RFC 4447](#), April 2006.
- [3] Kompella, V., Lasserrre, M. , et al., "Virtual Private LAN Service (VPLS) Using LDP Signalling", [RFC 4762](#), January 2007.

10.2. Informative References

- [4] Martini, L., et al., "Segmented Pseudo Wire", [draft-ietf-pwe3-segmented-pw-09.txt](#), January 2009.
- [5] Muley, P., et al., "Pseudowire (PW) Redundancy", [draft-ietf-pwe3-redundancy-01.txt](#)", March 2009.
- [6] Bryant, S., et al., " Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture", [RFC 3985](#), March 2005

11. [Appendix A](#) - Applications of PW Redundancy Procedures

This section shows how the mechanisms described in this document are used to achieve the desired protection behavior for the scenarios described in the PW redundancy requirements and framework document [5].

11.1. One Multi-homed CE with single SS-PW redundancy

The following figure illustrates an application of single segment pseudo-wire redundancy.

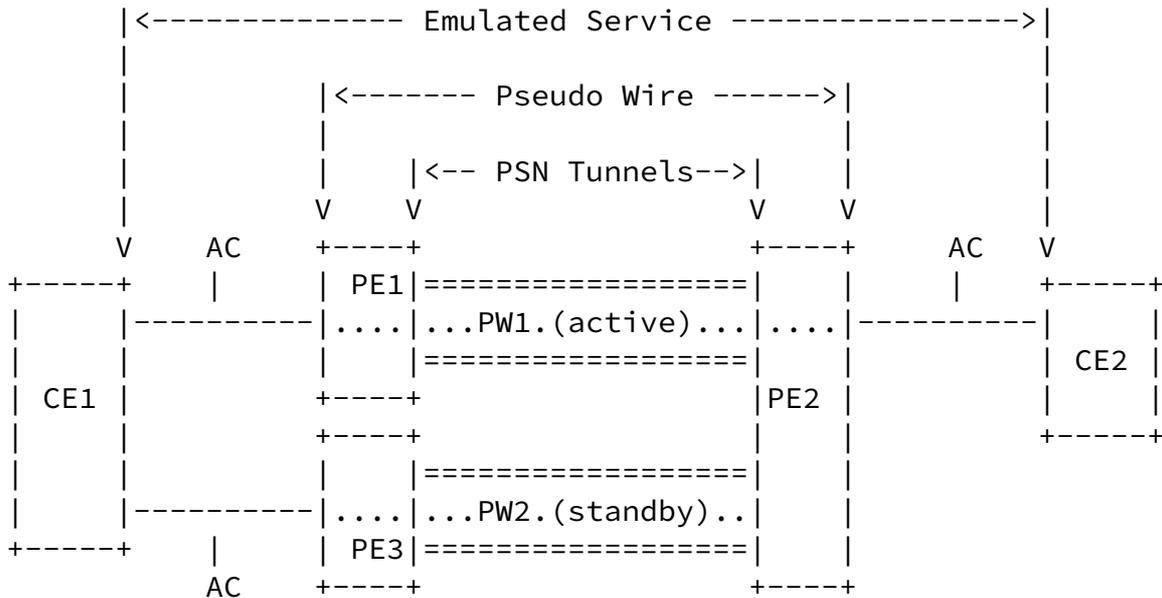


Figure 2 Multi-homed CE with single SS-PW redundancy

The application in Figure 2 makes use of the Independent mode of operation.

CE1 is dual homed to PE1 and to PE3 by attachment circuits. The method for dual-homing of CE1 to PE1 and to PE3 nodes, and the used protocols such as Multi-chassis Link Aggregation Group (MC-LAG), is outside the scope of this document.

In normal operation, PE1 and PE3 will advertise "Active" and "Standby" 'preferential forwarding' status bit respectively to PE2. This status reflects the forwarding state of the two AC's to CE1 as determined by the AC dual-homing protocol used. PE2 advertises 'preferential forwarding' status bit of "Active" on both PW1 and PW2 since the AC to CE2 is single homed. As both the local and remote operational and preferential forwarding status for PW1 are UP and Active, traffic is forwarded over PW1 in both directions.

On failure of the AC between CE1 and PE1, the forwarding state of the AC on PE3 transitions to Active. For example the MC-LAG control

protocol changes the link state on PE3 to active. PE3 then announces the newly changed 'preferential forwarding' status bit of "active" to PE2. PE1 will advertise a PW status notification message indicating that the AC between CE1 and PE1 is operationally down. PE2 matches the local and remote preferential forwarding status of "active" and operational status "Up" and select PW2 as the new active pseudo-wire to send traffic to.

On failure of PE1 node, PE3 will detect it and will transition the forwarding state of its AC to Active. The method by which PE3 detects that PE1 is down is outside the scope of this document. PE3 then announces the newly changed 'preferential forwarding' status bit of "active" to PE2. PE3 and PE2 match the local and remote preferential forwarding status of "active" and operational status "Up" and select PW2 as the new active pseudo-wire to send traffic to. Note that PE2 may have detected that the PW to PE1 went down via T-LDP Hello timeout or via other means. However, it will not be able to forward user traffic until it received the updated status bit from PE3.

Note in this example, the receipt of the operational status of the AC on the CE1-PE1 link is normally sufficient to have PE2 switch the path to PW2. However, the operator may want to trigger the switchover of the path of the PW for administrative reasons, i.e., maintenance, and thus the use of the 'preferential forwarding' status bit is required to notify PE2 to trigger the switchover.

Note that the primary/secondary procedures do not apply in this case as the PW Active/Standby status is driven by the AC forwarding state as determined by the AC dual-homing protocol used.

11.2. Multiple Multi-homed CEs with single SS-PW redundancy

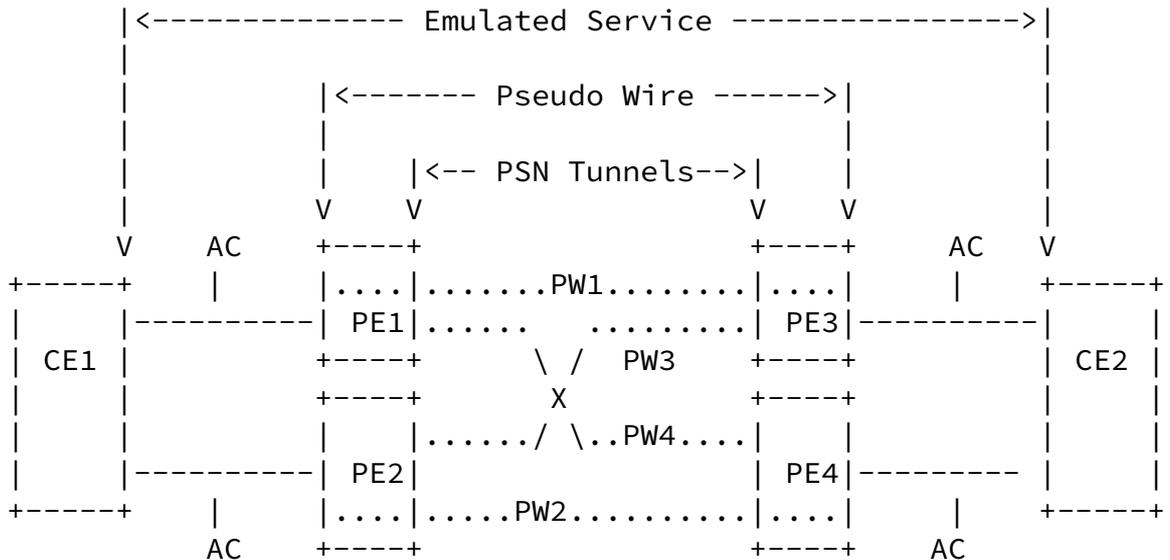


Figure 3 Multiple Multi-homed CEs with single SS-PW redundancy

The application in Figure 3 makes use of the Independent mode of operation.

CE1 is dual-homed to PE1 and PE2. CE2 is dual-homed PE3 and PE4. 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 advertises the preferential status "active" and operational status "UP" for pseudo-wires PW1 and PW4 connected to PE3 and PE4. This status reflects the forwarding state of the AC attached to PE1. PE2 advertises preferential status "standby" where as operational status "UP" for pseudo-wires PW2 and PW3 to PE3 and PE4. PE3 advertises preferential status "standby" where as operational status "UP" for pseudo-wires PW1 and PW3 to PE1 and PE2. PE4 advertise the preferential status "active" and operational status "UP" for pseudo-wires PW2 and PW4 to PE2 and PE1 respectively. 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. Thus by matching the local and remote preferential forwarding status of "active" and operational status "Up" of pseudo-wire, the PE nodes determine which PW should be in the Active state. In this case it is PW4 that will be selected.

On failure of the AC between CE1 and PE1, the forwarding state of the AC on PE2 is changed to Active. For example the MC-LAG control protocol changes the link state on PE2 to active. PE2 then announces the newly changed 'preferential forwarding' status bit of "active" to PE3 and PE4. PE1 will advertise a PW status notification message indicating that the AC between CE1 and PE1 is operationally down. PE2 and PE4 match the local and remote preferential forwarding status of "active" and operational status "Up" and select PW2 as the new active pseudo-wire to send traffic to.

On failure of PE1 node, PE2 will detect it and will transition the forwarding state of its AC to Active. The method by which PE2 detects that PE1 is down is outside the scope of this document. PE2 then announces the newly changed 'preferential forwarding' status bit of "active" to PE3 and PE4. PE2 and PE4 match the local and remote preferential forwarding status of "active" and operational status "Up" and select PW2 as the new active pseudo-wire to send traffic to. Note that PE3 and PE4 may have detected that the PW to PE1 went down via T-LDP Hello timeout or via other means. However, they will not be able to forward user traffic until they received the updated status bit from PE2.

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 as per the procedures in the independent mode.

Note that the primary/secondary procedures do not apply in this use case as the Active/Standby status is driven by the AC forwarding state as determined by the AC dual-homing protocol used.

[11.3](#). Multi-homed CE with MS-PW redundancy

The following figure illustrates an application of multi-segment pseudo-wire redundancy.

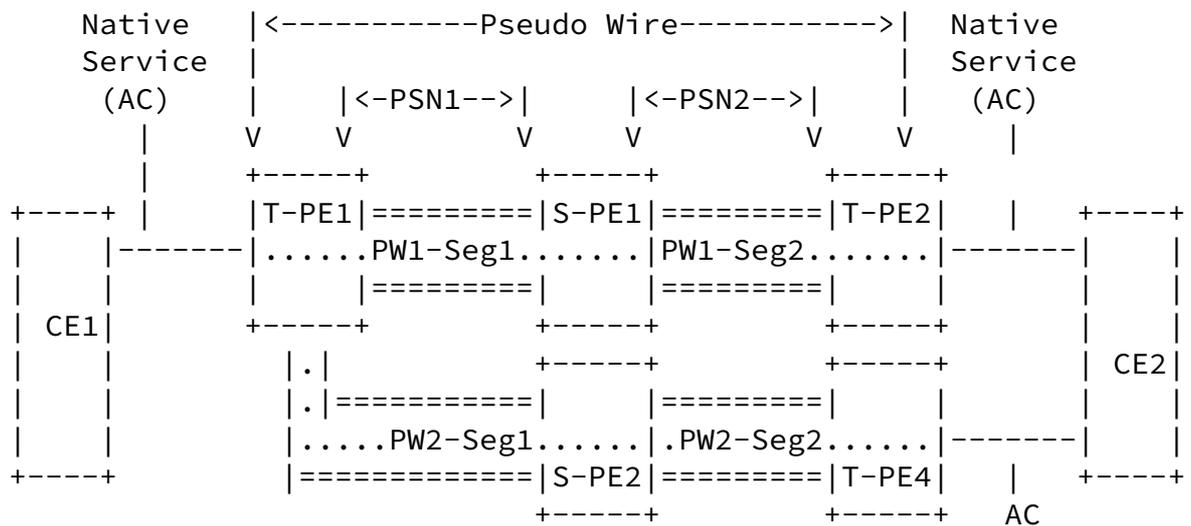


Figure 4 Multi-homed CE with MS-PW redundancy

The application in Figure 4 makes use of the Independent mode of operation.

CE2 is dual-homed to T-PE2 and T-PE4. PW1 and PW2 are used to extend the resilient connectivity all the way to T-PE1. PW1 has two segments and is active pseudo-wire while PW2 has two segments and is a standby pseudo-wire. This application requires support for MS-PW with segments of the same type as described in [4].

The operation in this case is the same as in the case of SS-PW as described in Section 11.1. . The only difference is that the S-PE nodes need to relay the PW status notification containing both the operational and forwarding status to the T-PE nodes.

11.4. Single Homed CE with MS-PW redundancy

The following is an application of the independent mode of operation along with the optional request switchover procedures in order to provide N:1 PW protection. A revertive behavior to a primary PW is shown as an example of configuring and using the primary/secondary procedures.

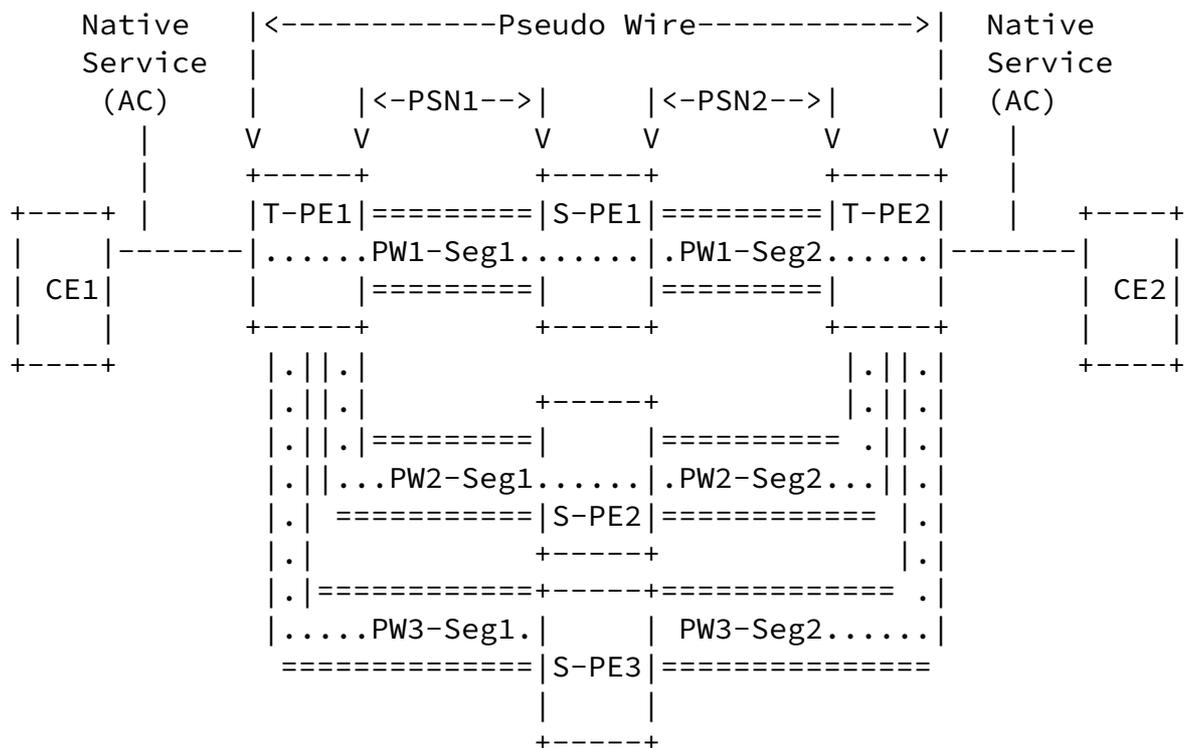


Figure 5 Single homed CE with multi-segment pseudo-wire redundancy
 CE1 is connected to PE1 in provider Edge 1 and CE2 to PE2 in provider

edge 2 respectively. There are three segmented PWs. A primary PW, PW1, is switched at S-PE1. A primary PW has the highest precedence. A secondary PW, PW2, which is switched at S-PE2 and has a precedence of 1. Finally, another secondary PW, PW3, is switched at S-PE3 and has a precedence of 2. The precedence is locally configured at the endpoints of the PW, i.e., T-PE1 and T-PE2.

T-PE1 and T-PE2 will select the PW they intend to activate based on their local and remote operational state as well as the local primary/precedence configuration. In this case, they will both advertise preferential forwarding' status bit of "active" on PW1 and of "standby" on PW2 and PW3. Assuming all PWs are operationally UP, T-PE1 and T-PE2 will use PW1 to forward user packets.

If PW1 fails, then the T-PE detecting the failure will send a status notification to the remote T-PE with a "PW Down" bit set as well as the 'preferential forwarding' status bit set on PW1. It will also clear 'preferential forwarding' status bit on PW2 as it is the next it has the next lowest precedence value. T-PE2 will also perform the same steps as soon as it is informed of the failure of PW1. Both T-PE nodes will perform a match on the 'preferential forwarding' status of "active" and operational status of "Up" and will use PW2 to forward user packets.

However this does not guarantee that paths of the PW are synchronized at all times. This may be due to a mismatch of the configuration of the PW priority in each T-PE. This may also be due to a failure which caused the endpoints to not be able to match the Active 'preferential forwarding' status bit and operational status bits. In this case, T-PE1 and/or T-PE2 can invoke the optional request switchover/acknowledgement procedures to synchronize the PW paths in both directions.

The trigger for sending a request to switchover can also be the execution of an administrative maintenance operation by the network operator in order to move the traffic away from the T-PE/S-PE nodes /links to be serviced.

In case the request switchover is sent by both endpoints simultaneously, both T-PEs send status notification with the newly selected PW with 'request switchover' bit set, waiting for response

from the other endpoint. In such situation, the T-PE with greater system address request is given precedence. This helps in synchronizing paths in event of mismatch of priority configuration as well.

11.5. PW redundancy between H-VPLS MTU-s and PE-rs

Following figure illustrates the application of use of PW redundancy in H-VPLS for the purpose of dual-homing an MTU-s node to PE nodes using PW spokes. This application makes use of the Master/Slave mode of operation.

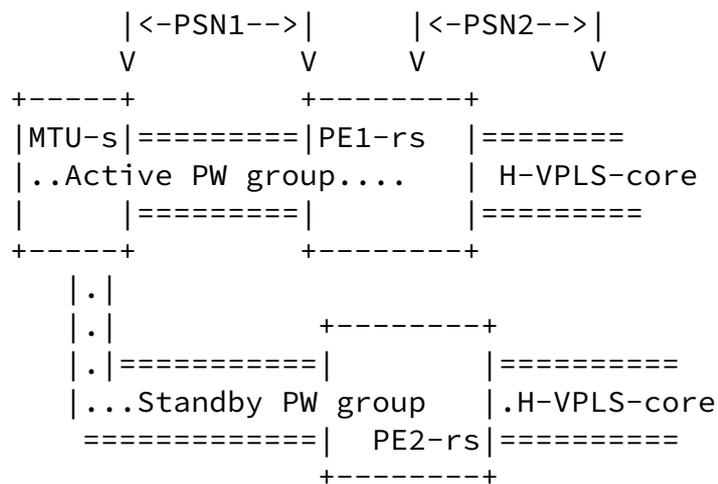


Figure 6 Multi-homed MTU-s in H-VPLS core

MTU-s is dual homed to PE1-rs and PE2-rs. The primary spoke PWs from MTU-s are connected to PE1-rs while the secondary PWs are connected to PE2. PE1-rs and PE2-rs are connected to H-VPLS core on the other side of network. MTU-s communicates to PE1-rs and PE2-rs the forwarding status of its member PWs for a set of VSIs having common

status Active/Standby. It may be signaled using PW grouping with common group-id in Pwid FEC Element or Grouping TLV in Generalized Pwid FEC Element as defined in [2] to scale better. MTU-s derives the status of the PWs based on local policy configuration. In this example, the primary/secondary procedures are used but this can be based on any other policy.

Whenever MTU-s performs a switchover, it sends a wildcard Notification Message to PE2-rs for the Standby PW group containing PW Status TLV with PW Standby bit cleared. On receiving the notification PE-2rs unblocks all member PWs identified by the PW group and state of PW group changes from Standby to Active. All procedures described in [Section 5.2](#). are applicable.

The use of the 'preferential forwarding' status bit in Master/Slave mode is similar to Topology Change Notification in RSTP controlled IEEE Ethernet Bridges but is restricted over a single hop. When these procedures are implemented, PE-rs 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 [6] to prevent two copies of MAC withdraws to be sent, one by MTU-s and another one by PE-rs nodes. Mechanisms to disable MAC Withdraw trigger in certain devices is out of the scope of this document

Author's Addresses

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

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

Matthew Bocci
Alcatel-Lucent
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@cw.com

Olen Stokes
Extreme Networks
Email: ostokes@extremenetworks.com

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

Luca Martini
Cisco Systems, Inc.
9155 East Nichols Avenue, Suite 400
Englewood, CO, 80112
Email: lmartini@cisco.com

Thomas Nadeau

Muley et al.

Expires March 30, 2009

[Page 27]

Internet-Draft

Preferential Forwarding Status Bit

September 2008

BT
tnadeau@lucidvision.com

Giles Heron
BT
giles.heron@gmail.com

Full Copyright Statement

Copyright (C) The IETF Trust (2008).

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.

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.

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 [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.

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