Network Working Group
Internet Draft
Expires: May 2008

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.

November 19, 2007

Preferential Forwarding Status bit definition
draft-muley-dutta-pwe3-redundancy-bit-02.txt

Status of this Memo

This Internet-Draft will expire on May 19, 2008.

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

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 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 [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 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.

```
        |<-------------- Emulated Service ---------------->|
         |                                                  |
         |          |<------- Pseudo Wire ------>|          |
```

```
         |          |                            |          |
         |          |    |<-- PSN Tunnels-->|    |          |
         |          V    V                  V    V          |
         V    AC    +----+                  +----+     AC   V
   +-----+    |     | PE1|==================|    |     |    +-----+
   |     |----------|....|...PW1.(active)...|....|----------|     |
   |     |          |    |==================|    |          | CE2 |
   | CE1 |          +----+                  |PE2 |          |     |
   |     |          +----+                  |    |          +-----+
   |     |          |    |==================|    |
   |     |----------|....|...PW2.(standby)..|    |
```



```
   +-----+    |     | PE3|==================|    |
              AC    +----+                  +----+
```

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, PW redundancy is important to provide resilience in the event of failure of S-PE node since PSN protection mechanisms cannot.

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

The PWE3 control protocol [2] defines the following status codes 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..

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 traffic.

Standby PW: An UP PW that is not used for forwarding user traffic.

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.

In steady state, a PE will always find an Active PW. However, it is possible that due to a misconfiguration, such a PW is not found. The behavior of a PE in this case is outside the scope of this document.

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 liveliness of standby PWs.

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 forward user

traffic to any other of the PW's in the redundant 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.

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.

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.

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. 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 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         STANDBY

          bit set

               Action: Stop forwarding over PW
```



```
               Receive PW preferential forwarding         ACTIVE
```

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

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 LDP status 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 set in the requested PW, 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 set to the newly active PW 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 set to the newly active PW 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

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 and Philippe Niger for their valuable comments and suggestions.

10. 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] Martini, L., et al., "Segmented Pseudo Wire", draft-ietf-pwe3-segmented-pw-05.txt, July 2007.

[4] Bryant, S., et al., " Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture", RFC 3985, March 2005

[5] Praveen, Pranjal et al., "draft-muley-pwe3-redundancy-01.txt", March 2007.

Author's Addresses

Praveen Muley
Alcatel-lucent
701 E. Middlefiled 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

Full Copyright Statement

Intellectual Property Statement

Acknowledgment

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