Network Working Group   S. Sivabalan
Internet-Draft   S. Boutros
Intended status: Standards Track   Cisco Systems, Inc.
Expires: May 03, 2016   H. Shah
                                                          Ciena Corp.
                                                           S. Aldrin
                                                         Google Inc.
                                                       M. Venkatesan
                                                            Comcast.
                                                   November 03, 2015

MAC Address Withdrawal over Static Pseudowire
draft-ietf-pals-mpls-tp-mac-wd-03.txt

Abstract

   This document specifies a mechanism to signal MAC address withdrawal
   notification using PW Associated Channel (ACH). Such notification is
   useful when statically provisioned PWs are deployed in VPLS/H-VPLS
   environment.

Requirements Language

   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 [RFC2119].

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). Note that other groups may also distribute
   working documents as Internet-Drafts. The list of current Internet-
   Drafts is at http://datatracker.ietf.org/drafts/current/.

   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."

   This Internet-Draft will expire on May 3, 2016.

Copyright Notice

Table of Contents

1. Introduction

   An LDP-based MAC Address Withdrawal Mechanism is specified in
   [RFC4762] to remove dynamically learned MAC addresses when the source
   of those addresses can no longer forward traffic. This is
   accomplished by sending an LDP Address Withdraw Message with a MAC
   List TLV containing the MAC addresses to be removed, to all other PEs
   over the LDP sessions. [RFC7361] describes an optimized MAC
   withdrawal mechanism which can be used to remove only the set of
   MAC addresses that need to be re-learned in H-VPLS networks.
   [RFC7361] also describes optimized MAC Withdrawal operations
   in PBB-VPLS networks.

A PW can be signaled via the LDP or can be statically provisioned. In the case of static PW, LDP based MAC withdrawal mechanism cannot be used. This is analogous to the problem and solution described in [RFC6478] where PW OAM message has been introduced to carry PW status TLV using in-band PW Associated Channel. In this document, we propose to use PW OAM message to withdraw MAC address(es) learned via static PW.

Thus, MAC withdraw signaling for static PW re-uses concepts of

- in-band signaling mechanisms used by static PW status signaling and

- MAC withdrawal mechanisms described by [RFC4762] and [RFC7361]

The MAC withdraw signaling is a best effort scheme. It is an attempt to optimize the network convergence by reducing blackholes caused by PW failover for protected PWs. The protocol defined in this document addresses possible loss of MAC withdraw signal due to network congestion, but do not assure the guarenteed delivery, as is the case for the LDP based MAC withdraw signaling. In the event that MAC withdraw signaling does not reach the intended target, the fallback to MAC re-learning due to bi-directional traffic or as a last resort to user configured MAC entries age out, will resume the traffic via new PW path. Such fallbacks would cause temporary blackout but does not render network permanently unusable.

2. Terminology

The following terminologies are used in this document:

ACK: Acknowledgement for MAC withdraw message.

LDP: Label Distribution Protocol.

MAC: Media Access Control.

PE: Provide Edge Node.

MPLS: Multi Protocol Label Switching.

PW: PseudoWire.

PW OAM: PW Operations, Administration and Maintenance.

TLV: Type, Length, and Value.

VPLS: Virtual Private LAN Services.

3. MAC Withdraw OAM Message

LDP provides a reliable packet transport for control plackets for
dynamic PWs. This can be contrasted with static PWs which rely on
re-transmission and acknowledgments (ACK) for reliable OAM packet
delivery as described in [RFC6478]. The proposed solution for MAC
withdrawal over static PW also relies on re-transmissions and ACKs.
However, ACK is mandatory. A given MAC withdrawal notification is
sent as a PW OAM message, and the sender re-transmits the message for
a configured number of times in the absence of an ACK response for
the sequence numbered message. The receiver removes the MAC
address(es) for a given sequence number MAC withdraw signaling and
sends the ACK response. The receipt of same or lower sequence number
message is responded with ACK but does not cause removal of MAC
addresses. A new TLV to carry the sequence number has been defined.

The format of the MAC address withdraw OAM message is shown in
Figure 1. The MAC withdraw PW OAM message follows same guidelines
used in [RFC6478], whereby first 4-bytes of OAM message header
is followed by message specific field and a set of TLVs relevant
for the message. Since the MAC withdrawal PW OAM message is not
refreshed forever, a MAC address withdraw OAM message MUST contain a
"Sequence Number TLV" otherwise the entire message is dropped. It
MAY contain MAC Flush Parameter TLVs defined in [RFC7361] when
static PWs are deployed in H-VPLS and PBB-VPLS scenarios. The first
2 bits of the sequence number TLV are reserved and MUST be set to 0
on transmit and ignored on receipt.

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |0 0 0 1|Version|   Reserved    |(TBD) MAC Withdraw OAM Message |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |           Reserved            |  TLV Length   |A|R| Flags     |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |Res| Sequence Number TLV (TBD) |  Sequence Number TLV Length   |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                        Sequence Number                        |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                                                               |
   |                       MAC List TLV                            |
   ~               MAC Flush Parameter TLV (optional)              ~
   |                                                               |
   |                                                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
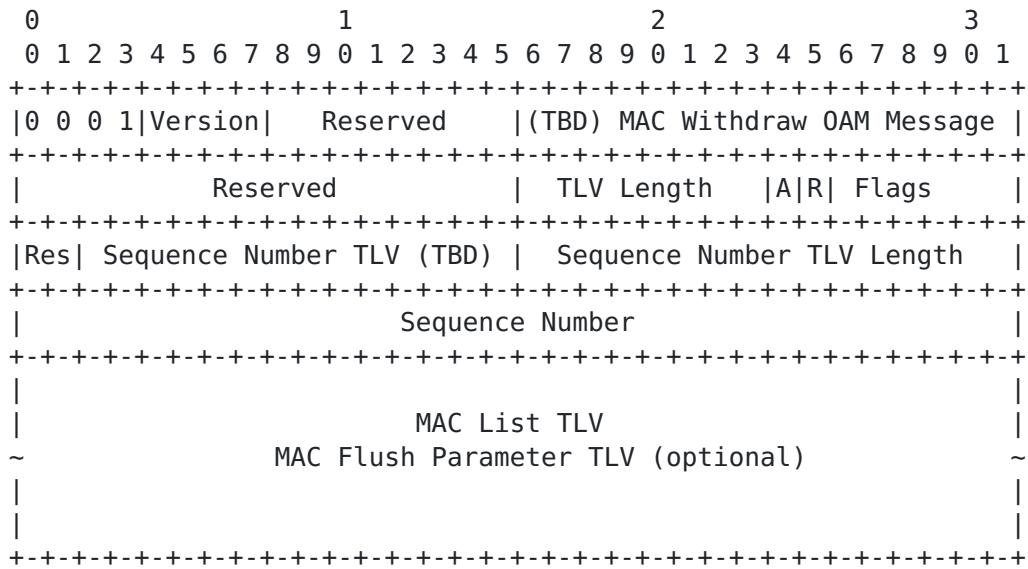
            Figure 1: MAC Address Withdraw PW OAM Packet Format

In this section, MAC List TLV and MAC Flush Parameter TLV are
collectively referred to as "MAC TLV(s)". The definition and
processing rules of MAC List TLV are described by [RFC4762], and the
corresponding rules of MAC Flush Parameter TLV are governed by
[RFC7361].

"TLV Length" is the total length of all TLVs in the message, and
"Sequence Number TLV Length" is the length of the sequence number
field.

A single bit (called A-bit) is set by a receiver to acknowledge
receipt and processing of a MAC Address Withdraw OAM message.
In the acknowledge message, with A-bit set, MAC TLV(s) is/are
excluded.

A single bit (called R-bit) is set to indicate if the sender is
requesting reset of the sequence numbers. The sender sets this bit
when the Pseudowire is restarted and has no local record of send and
expected receive sequence number.

The Sequence number TLV MUST be the first TLV in the message.

The lack of reliable transport protocol for the in-band OAM
necessitates a presence of sequencing and acknowledgement
scheme so that the receiver can recognize newer message from
retransmitted older messages. The [RFC4385] describes the details
of sequence number handling which includes overflow detection for
sequence number field of size 16-bits. This document leverages
the same scheme with the two exemptions
    - sequence number field is of size 32-bits
    - overflow detection is simplified such that sequence

number exceed 2,147,483,647 (0x7FFFFFFF) is considered
overflow and reset to 1.

4. Operation

   This section describes how the initial MAC withdraw OAM messages are
   sent and retransmitted, as well as how the messages are processed and
   retransmitted messages are identified.

4.1. Operation of Sender

   Each PW is associated with a counter to keep track of the sequence
   number of the transmitted MAC withdrawal messages. Whenever a node
   sends a new set of MAC TLVs, it increments the transmitted sequence
   number counter, and include the new sequence number in the message.
   The transmit sequence number is initialized to 1 at the onset, after
   the wrap and after the sequence number reset request receipt.
   Hence the transmit sequence number is set to 2 in the first MAC
   withdraw message sent after the sequence number is initialized
   to 1.

   The sender expects an ACK from the receiver within a time interval
   which we call "Retransmit Time" which can be either a default (1
   second) or configured value. If the ACK does not arrive within the
   Retransmit Time, the sender retransmits the message with the same
   sequence number as the original message. The retransmission MUST
   be ceased when an ACK is received. In order to avaoid continuous
   retransmissions in the absence of acknowledgements, a method of
   suppressing retransmission MUST be implemented. A simple and
   well used approach is to cease retransmission after a small
   number of transmissions. A one second retransmission with two
   retries in the absence of an ACK response is RECOMMENDED.
   However, both the interval and the number of retries are a local
   matter which present no interworking issues and thus the operator
   MAY configure different values. Alternatively, an increasing
   backoff delay with a larger number of retries MAY be implemented
   to improve scaling issues. Whilst there are no interworking issues
   with any of these methods, the implementer must be mindful of not
   introducing network congestion and must take into account of
   decaying value of delayed MAC withdraw signaling against possible
   relearning due to bidirectional traffic or MAC age timeout.

   During the period of retransmission, if a need to send a new MAC
   withdraw message with updated sequence number arises then
   retransmission of the older unacknowledged withdraw message MUST be
   suspended and retransmit time for the new sequence number MUST be
   initiated. In essence, sender engages in retransmission logic only
   for the latest send withdraw message for a given PW.

   In the event that a Pseudowire was deleted and re-added or the router
   is restarted with configuration, the local node may lose information
   about the send sequence number of previous incarnation. This becomes
   problematic for the remote peer as it will continue to ignore the

received MAC withdraw messages with lower sequence numbers. In such
cases, it is desirable to reset the sequence numbers at both ends of
the Pseudowire. The 'R' reset bit is set in the first MAC withdraw
to notify the remote peer to reset the send and receive sequence
numbers. The 'R' bit must be cleared in subsequent MAC withdraw
messages after the acknowledgement is received

4.2. Operation of Receiver

Each PW is associated with a register to keep track of the expected
sequence number of the MAC withdrawal message and is initialized to
1. Whenever a MAC withdrawal message is received, and if the
sequence number on the message is greater than the value in the
register, the MAC address(es) contained in the MAC TLV(s) is/are
removed, and the register is updated with the received sequence
number. The receiver sends an ACK whose sequence number is the
same as that in the received message.

If the sequence number in the received message is smaller than or
equal to the value in the register, the MAC TLV(s) is/are not
processed. However, an ACK with the received sequence number MUST be
sent as a response. The receiver processes the ACK message as an
acknowledgement for all the MAC withdraw messages sent up to the
sequence number present in the ACK message and terminates
retransmission.

The handling of the sequence number is described in section 3.

A MAC withdraw message with 'R' bit set MUST be processed by
resetting the send and receive sequence number first. The rest of
MAC withdraw message processing is performed as described above. The
acknowledgement is sent with 'R' bit cleared.

5. Security Consideration

The security measures described in [RFC4447], [RFC5085], and
[RFC6073] are adequate for the proposed mechanism.

6. IANA Considerations

6.1. MPLS G-Ach type

This document requests IANA to assign new channel type (requested
value 0x0028) from the registry named "MPLS Generalized Associated
Channel (G-ACh) Types (including Pseudwire Associated Channel
Types)". The description of the new channel type is "MAC Withdraw
OAM Message". [TO BE REMOVED: This registration should take place at
the following location: http://www.iana.org/assignments/g-ach-
parameters/g-ach-parameters.xhtml]. The channel type value of 0x0028

is requested as it is used in implementations that are deployed in
the field.

6.2. Sequence Number TLV

This document requests IANA to assign a new TLV Type
(requested value 0x0001) from the existing LDP "TLV Type Name
Space" registry. The description for the new TLV Type is
"Sequence Number TLV".

[Note to IANA TO BE REMOVED BY THE RFC EDITOR: This registration
should take place at the following location:
http://www.iana.org/assignments/ldp-namespaces/ldp-namespaces.xhtml].
In an earlier revision of this draft, we created a new sub-TLV
registry with one entry, a new "Sequence Number TLV" with the
value 0x0001. This has been implemented by several vendors. The
IESG proposed that rathen than create a new sub-TLV registry, we
just allocate a new code point from the existing LDP "TLV Type Name
Space" registry. In this registry, the value 0x0001 is available for
allocation by standards action, so we request this code point for
the new "Sequence Number" TLV type to avoid needing to change
the existing implementations.

7. References

7.1. Normative References

[RFC4385] Bryant, S., Swallow, G., Martini, L., and D. McPherson,
           "Pseudowire Emulation Edge-to-Edge (PWE3) Control Word for
           Use over an MPLS PSN", RFC 4385, February 2006.

[RFC4447] Martini, L., Rosen, E., El-Aawar, N., Smith, T., and G.
           Heron, "Pseudowire Setup and Maintenance Using the Label
           Distribution Protocol (LDP)", RFC 4447, April 2006.

[RFC4762] Lasserre, M. and V. Kompella, "Virtual Private LAN Service
           (VPLS) Using Label Distribution Protocol (LDP) Signaling",
           RFC 4762, January 2007.

[RFC5085] Nadeau, T. and C. Pignataro, "Pseudowire Virtual Circuit
           Connectivity Verification (VCCV): A Control Channel for
           Pseudowires", RFC 5085, December 2007.

[RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an
           IANA Considerations Section in RFCs", BCP 26, RFC 5226,
           May 2008.

   [RFC6073]   Martini, L., Metz, C., Nadeau, T., Bocci, M., and M.
               Aissaoui, "Segmented Pseudowire", RFC 6073, January 2011.

   [RFC6478]   Martini, L., Swallow, G., Heron, G., and M. Bocci,
               "Pseudowire Status for Static Pseudowires", RFC 6478, May
               2012.

   [RFC7361]   Dutta, P., Balus, F., Stokes, O., Calvignac, G., and D.
               Fedyk, "LDP Extensions for Optimized MAC Address
               Withdrawal in a Hierarchical Virtual Private LAN Service
               (H-VPLS)", RFC 7361, September 2014.

   [RFC2119]   Bradner, S., "Key words for use in RFCs to Indicate
               Requirement Levels", BCP 14, RFC 2119, March 1997.

7.2. Informative References

Authors' Addresses

   Siva Sivabalan
   Cisco Systems, Inc.
   2000 Innovation Drive
   Kanata, Ontario K2K 3E8
   Canada

   Email: msiva@cisco.com

   Sami Boutros
   Cisco Systems, Inc.
   170 West Tasman Dr.
   San Jose, CA 95134
   US

   Email: sboutros@cisco.com

   Himanshu Shah
   Ciena Corp.
   3939 North First Street
   San Jose, CA 95134
   US

   Email: hshah@ciena.com

Sam Aldrin
Google Inc.

Email: aldrin.ietf@gmail.com

Mannan Venkatesan
Comcast.
1800 Bishop Gate Blvd
Mount Laurel, NJ 08075
US

Email: mannan_venkatesan@cable.comcast.com