TITOC Working Group   S. Davari
Internet Draft   A. Oren
Intended status: Standards   Broadcom
Expires: March 22, 2011   L. Martini
                                                               Cisco

                                                        Sep 22, 2010

              Transporting PTP messages (1588) over MPLS Networks
                    draft-davari-tictoc-1588overmpls-00.txt

Status of this Memo

Copyright Notice

Abstract

   This document defines the method for transporting PTP messages (PDUs)
   over an MPLS network to enable a proper handling of these
   packets (e.g. implementation of Transparent Clocks (TC)) in LSRs.

   The basic idea is to transport PTP messages inside dedicated MPLS
   LSPs. These LSPs only carry PTP messages and possibly Control and
   Management packets, but they do not carry customer traffic.

   Two methods for transporting 1588 over MPLS are defined. The first
   method is to transport PTP messages directly over the dedicated MPLS
   LSP via UDP/IP encapsulation, which is suitable for IP/MPLS networks.
   The second method is to transport PTP messages inside a PW via
   Ethernet encapsulation, which is more suitable for MPLS-TP networks.

Table of Contents

1. Introduction

The objective of Precision Time Protocol (PTP) is to synchronize
independent clocks running on separate nodes of a distributed system.
[IEEE1588] defines PTP messages for clock and time synchronization.
The PTP messages include PTP PDUs over UDP/IP (Annex D & E of
[IEEE1588]) and PTP PDUs over Ethernet (Annex F of [IEEE1588]). This

document defines mapping and transport of the PTP messages defined in
[IEEE1588] over MPLS networks.

PTP defines intermediate clock functions (called transparent clocks)
between the source of time (Master) and the Slave clocks. Boundary
Clocks (BC) form Master-Slave hierarchy with the Master clock as root.
The messages related to synchronization, establishing the Master-
Slave hierarchy, and signaling, terminate in the protocol engine of a
boundary clock and are not forwarded. Management messages however,
are forwarded to other ports on the boundary clock.

Transparent clocks modify a "correction field" (CF) within the
synchronization messages to compensate for residence and propagation
delays. Transparent clocks do not terminate synchronization, Master-
Slave hierarchy control messages or signaling messages.

There is a need to transport PTP messages over MPLS networks. The
MPLS network could be a transit network between 1588 Masters and
Slaves. The accuracy of the recovered clock improves and the Slave
logic simplifies when intermediate nodes (e.g. LSRs) properly handle
PTP messages (e.g. perform TC), otherise the jitter at the 1588
Slave may be excessive and therefore the Slave may not be able to
properly rcover the clock and time of day.

This document requires that MPLS nodes (LSRs) SHOULD be able to
support the Transparent Clock (TC) function, meaning that they should
be able to modify the CF of the proper PTP messages, via a 1-step or
2-step process. Such LSR is called "1588-aware LSR" in this document.

TC requires a 1588-aware LSR in the middle of an LSP to identify the
PTP messages and perform proper update of the CF.

More generally this document requires that an LSR SHOULD be
able to properly handle the PTP messages. For instance for those cases
when the TC function is not viable (e.g. due to layer violation) as
an alternative it should be possible to instead control the delay for

these messages on both directions across the node.

In the above cases it is beneficial that PTP packets can be easily identified when carried over MPLS.

This document provides two methods for transporting PTP messages over MPLS. The main objectives are for LSRs to be able to deterministically detect and identify the PTP messages.

2. 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 [RFC2119].

3. Terminology

1588:   The timing and synchronization as defined by IEEE 1588

PTP: The timing and synchronization protocol used by 1588

Master: The Source of 1588 Timing and clock

Slave: The Destination of 1588 Timing and clock that tries to
         follow the Master clock.

OC: Ordinary Clock

TC: Transparent Clocking, a time stamping method applied by
         intermediate nodes between Master and Slave

BC: Border Clock, is a node that recovers the Master clock via a
         Slave function and uses that clock as the Master for other
         Slaves.

PTP LSP: An LSP dedicated to carry PTP messages

PTP PW: A PW within a PTP LSP that MAY correspond to a Master/Slave

flow.

    CW: Pseudo wire Control Word

    HW: Hardware

    LAG: Link Aggregation

    ECMP: Equal Cost Multipath

    CF: Correction Field, a field inside certain PTP messages
                (message type 0-3)that holds the accumulative transit time
                inside intermediate switches

_____

[4](). Problem Statement

    When PTP messages are transported over MPLS networks, there is a need
    for intermediate LSRs to detect such messages and perform proper
    processing (e.g. Transparent Clock (TC)). Note the TC processing
    could be in the form of 1-Step or 2-Step time stamping.

    PTP messages over Ethernet or IP can always be tunneled over MPLS.
    However the 1588 over MPLS mapping defined in this document is
    applicable whenever MPLS LSRs are 1588-aware and the intention is for
    those LSRs to perform proper processing on these packets.

    When 1588-awareness is needed PTP messages should NOT be transported
    over LSPs or PWs that are carrying customer traffic because LSRs
    perform Label switching based on the top label in the stack. To
    detect PTP messages inside such LSPs require special Hardware (HW) to
    do deep packet inspection at line rate. Even if one assumes a deep
    packet inspection HW at line rate exists, the payload can't be
    deterministically identified by LSRs because the payload type is a
    context of the PW label and the PW label and its context are only
    known to the Edge routers (PEs) and LSRs don't know what is a PW's
    payload (Ethernet, ATM, FR, CES, etc). Even if one assumes only
    Ethernet PWs are permitted in an LSP, the LSRs don't have the
    knowledge of whether PW Control Word (CW) is present or not and
    therefore can't deterministically identify the payload.

    Therefore a generic method is defined in this document that does not
    require deep packet inspection at line rate, and can

deterministically identify PTP messages. The defined method is
applicable to both MPLS and MPLS-TP networks.

5. Dedicated LSPs for PTP messages

   The method defined in this document can be used by LSRs to identify
   PTP messages in MPLS tunnels by using dedicated LSPs to carry PTP
   messages.

   Compliant implementations MUST use dedicated LSPs to carry PTP
   messages over MPLS. Let's call these LSPs as the "PTP LSPs" and the
   labels associated with these LSPs as "PTP labels". These LSPs could
   be P2P or P2MP LSPs. The PTP LSP between Master and Slaves MAY be
   P2MP or P2P LSP while the PTP LSP between each Slave and Master
   SHOULD be P2P LSP. The PTP LSP between a Master and a Slave and the
   PTP LSP between the same Slave and Master MUST be co-routed.
   Alternatively, a single bidirectional co-routed LSP can be used.
   The PTP LSP MAY be MPLS LSP or MPLS-TP LSP.

   The PTP LSPs could be configured or signaled via RSVP-TE/GMPLS. New
   RSVP-TE/GMPLS TLVs and objects are defined in this document to
   indicate that these LSPs are PTP LSPs.

   Note that the PTP LSPs MUST only carry PTP messages and MAY carry
   MPLS/MPLS-TP control and management messages such as BFD and LSP-Ping.

6. 1588 over MPLS Encapsulation

   This document defines two methods for carrying PTP messages over MPLS.
   The first method is carrying PTP messages over PTP LSPs and the
   second method is to carry PTP messages over dedicated Ethernet PWs
   (called PTP PWs) inside PTP LSPs.

6.1. 1588 over LSP Encapsulation

   The simplest method of transporting PTP messages over MPLS is to
   encapsulate PTP PDUs in UDP/IP and then encapsulate them in PTP LSP.
   The 1588 over LSP format is shown in Figure 1.

                    +----------------+
                    |PTP Tunnel Label|
                    +----------------+
                    | IPV4/V6 |

```
                    +---------------+
                    |     UDP       |
                    +---------------+
                    |   PTP PDU     |
                    +---------------+
```
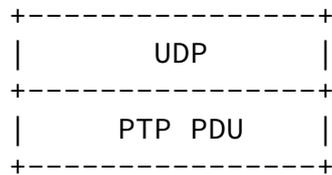
              Figure 1 - 1588 over LSP Encapsulation

    This encapsulation is very simple and is useful when the networks
    between 1588 Master and Slave are IP/MPLS networks.

    In order for an LSR to process PTP messages, the PTP Label MUST be
    the top label of the label stack.

    The UDP/IP encapsulation of PTP MUST follow Annex D and E of
    [IEEE1588].

6.2. 1588 over PW Encapsulation

    Another method of transporting 1588 over MPLS networks is by
    encapsulating PTP PDUs in Ethernet and then transporting them over
    Ethernet PW (PTP PW) as defined in [RFC4448] , which in turn is
    transported over PTP LSPs. Alternatively PTP PDUs MAY be encapsulated
    in UDP/IP/Ethernet and then transported over Ethernet PW.

    Both Raw and Tagged modes for Ethernet PW are permitted. The 1588
    over PW format is shown in Figure 2.
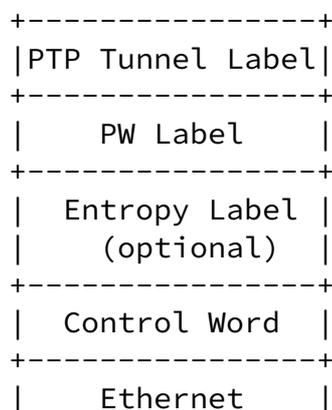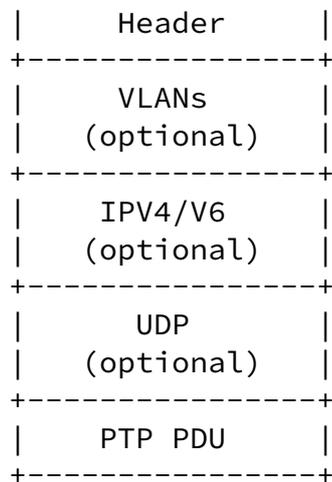
                    +---------------+
                    |PTP Tunnel Label|
                    +---------------+
                    | PW Label |
                    +---------------+
                    | Entropy Label |
                    | (optional) |
                    +---------------+
                    | Control Word |
                    +---------------+
                    | Ethernet |
```

```
                    |     Header     |
                    +----------------+
                    |     VLANs      |
                    |   (optional)   |
                    +----------------+
                    |    IPV4/V6     |
                    |   (optional)   |
                    +----------------+
                    |      UDP       |
                    |   (optional)   |
                    +----------------+
                    |    PTP PDU     |
                    +----------------+
```

             Figure 2 - 1588 over PW Encapsulation

   The Control Word (CW) as specified in [RFC4448] is SHOULD be used
   to ensure a more robust detection of PTP messages inside the MPLS
   packet. If CW is used, the use of Sequence number is optional.

   The use of VLAN and UDP/IP are optional. Note that 1 or 2 VLANs MAY
   exist in the PW payload.

   In order for an LSR to process PTP messages, the top label of the
   label stack (the Tunnel Label) MUST be from PTP label range. However
   in some applications the PW label may be the top label in the stack,
   such as cases where there is only one-hop between PEs. In such cases,
   the PW label SHOULD be chosen from the PTP Label range.

   An Entropy label [Fat PW] MAY be present at the bottom of stack.

   The Ethernet encapsulation of PTP MUST follow Annex F of [IEEE1588]
   and the UDP/IP encapsulation of PTP MUST follow Annex D and E of
   [1EEE1588].

7. 1588 Message Transport

   1588 protocol comprises of a number of message types. A subset of PTP
   messages that require TC processing are:

        SYNC

FOLLOW_UP

DELAY_REQ (Delay Request)

DELAY_RES (Delay Response)

PDELAY_REQ (Peer Delay Request)

PDELAY_RESP (Peer Delay Response)

PDELAY_RESP_FOLLOW_UP (Peer Delay Response Follow up)

SYNC, FOLLOW_UP, DELAY_REQ and DELAY_RESP are exchanged between
Master and Slave and MUST be transported over PTP LSPs.

PDELAY_REQ, PDELAY_RESP, and PDELAY_RESP_FOLLOW_UP are exchanged
between adjacent routers and MAY be transported over PTP LSPs.

For a given instance of 1588 protocol SYNC, FOLLOW_UP, and DELAY_RESP
MUST be transported over the same PTP LSP in the direction from
Master to Slave, while DELAY_REQ MUST be transported over another PTP
LSP in the reverse direction meaning in the direction from Slave to
Master. These PTP LSPs, which are in opposite directions MUST be
congruent and co-routed. Alternatively, a single bidirectional
co-routed LSP can be used.

Other PTP message types are end-to-end messages between Master and
Slave that don't need to be processed by intermediate routers. These
message types MAY be carried in PTP Tunnel LSPs or any other LSP.
When these PTP messages are carried in PTP LSPs there is no need to
distinguish between the PTP message types, since the CF of these
messages will be ignored by Slave clock.

8. Protection and Redundancy

In order to ensure continuous uninterrupted operation of 1588 Slaves,
usually as a general practice, Redundant Masters are tracked by each
Slave. It is the responsibility of the network operator to ensure
that physically disjoint PTP tunnels that don't share any link are
used between the redundant Masters and a Slave.

When redundant Masters are tracked by a Slave, any PTP LSP or PTP PW

failure will trigger the slave to switch to the Redundant Master.
However LSP/PW protection such as Linear Protection Switching (1:1,
1+1), Ring protection switching or MPLS Fast Reroute (FRR) SHOULD
still be used to ensure the LSP/PW is ready for a future failure.

Note that any protection or reroute mechanism that adds additional
label to the label stack, such as Facility Backup Fast Reroute, MUST
ensure that the pushed label is a PTP Label to ensure proper
processing of PTP messages by LSRs in the backup path.

9. ECMP and LAG

To ensure the proper operation of 1588 Slaves, the physical path for
PTP messages from Master to Slave and vice versa MUST be the same for
all PTP messages listed in section 7 and MUST not change even in
presence of ECMP and LAG in the MPLS network.

The network operator MUST either ensure that the ECMP or LAG hashing
algorithms keep the PTP messages described in section 7 and belonging
to the same 1588 flow on the same link and path, or MUST disable LAG
and/or ECMP for the PTP LSPs and/or PWs.

10. OAM, Control and Management

In order to manage PTP LSPs and PTP PWs, they MAY carry OAM, Control
and Management messages. These control and management messages can be
differentiated from PTP messages via already defined IETF methods.

In particular BFD [RFC5880], [RFC5884] and LSP-Ping [RFC4389] MAY run
over PTP LSPs via UDP/IP encapsulation or via GAL/G-ACH. These
Management protocols are easily identified by the UDP Destination
Port number or by GAL/ACH respectively.

Also BFD, LSP-Ping and other Management messages MAY run over PTP PW
via one of the defined VCCVs (Type 1, 2 or 3) [RFC5085]. In this case
G-ACH, Router Alert Label (RAL), or PW label (TTL=1) are used to
identify such Management messages.

11. FCS Recalculation

Ethernet FCS MUST be recalculated at every LSR that performs the TC
processing and FCS retention described in [RFC4720] MUST not be used.

12. RSVP-TE/GMPLS Extensions for support of 1588

   RSVP-TE/GMPLS signaling MAY be used to setup the PTP LSPs. A new object
   or TLV is required to signal that this is a PTP LSP. The OFFSET from
   bottom of label stack to the start of the PTP PDU MAY also be signaled.
   The LSRs that receive and process the RSVP-TE/GMPLS messages MAY use
   the OFFSET to locate the PTP 'correction field' (CF).

   Note that the new object/TLV Must be ignored by LSRs that are not
   compliant to this specification.

   The signaling details will be added in future versions of the draft.

13. Backward compatibility with non-1588-aware LSRs

   It is most beneficial that all LSRs in the path of a PTP LSP be
   1588-aware LSRs. This would ensure teh highest quality time and
   clock synchronization by 1588 Slaves. However, this specification
   does not mandate that all LSRs in path of a PTP LSP be 1588-aware.

   Non-1588-aware LSRs just switch the MPLS packets carrying 1588
   messages as data packets.

14. Other considerations

   The use of Explicit Null (Label= 0 or 2) is accepatble as long as
   either the Explicit Null label is the bottom of stack label
   (applicable only to UDP/IP encapsulation) or the label below the
   Explicit Null label is a PTP label.

   The use of Penultimate Hop Poping (PHP) is acceptable as long as
   either the PHP label is the bottom of stack label (applicable
   only to UDP/IP encapsulation) or the label below the PHP label
   is a PTP label.

15. Security Considerations

   MPLS PW security considerations in general are discussed in [RFC3985]
   and [RFC4447], and those considerations also apply to this document.

An experimental security protocol is defined in [1]. The PTP
security extension and protocol provide group source authentication,
message integrity, and replay attack protection for PTP messages.

16. IANA Considerations

A new TLV is required to signal that PTP LSPs. IANA needs to assign
the new TLV Type.

17. References

17.1. Normative References

   [IEEE1588] IEEE Standard for a Precision Clock Synchronization
              Protocol for Networked Measurement and Control Systems,
              IEEE 1588-2008

   [RFC4448] Martini, L., Rosen, E., El-Aawar, and G.Heron,
              "Encapsulation Methods for Transport of Ethernet over
              MPLS Networks", April 2006.

   [RFC4389] K. Kompella, G. Swallow, "Detecting Multi-Protocol Label
              Switched (MPLS) Data Plane Failures", February 2006

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

   [RFC5880]   D. Katz, D. Ward, "Bidirectional Forwarding Detection",
               June 2010

   [RFC3985]   Bryant, S. and P. Pate, "Pseudo Wire Emulation Edge-to-
               Edge (PWE3) Architecture", March 2005

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

   [RFC5884]   R. Aggarwal, K. Kompella, T. Nadeau, G. Swallow,
               "Bidirectional Forwarding Detection for MPLS",
               June 2010

   [RFC4720] A. Malis, D.Allan,N. Del Regno, "Pseudowire Emulation
              Edge-to-Edge (PWE3)Frame Check Sequence Retention",
              November 2006

17.2. Informative References

    [Fat PW]   S. Bryant, "Flow Aware Transport of Pseudowires over an
              MPLS PSN", January 2010

18. Acknowledgments

Authors' Addresses

   Shahram Davari
   Broadcom Corp.
   3151 Zanker Road
   San Jose, CA 95134

   Email: davari@ieee.org
          davari@boadcom.com

   Amit Oren
   Broadcom Corp.
   3151 Zanker Road
   San Jose, CA 95134

   Email: amito@broadcom.com

   Luca Martini
   Cisco Systems
   San Jose,CA

   Email: Lmartini@cisco.com