Network Working Group                                          L. Jin
Internet-Draft                                          ZTE Corporation
Intended status: Standards Track                             F. Jounay
Expires: March 24, 2013                                  France Telecom
                                                          I. Wijnands
                                                     Cisco Systems, Inc
                                                           N. Leymann
                                                   Deutsche Telekom AG
                                                         Sep 20, 2012

        **LDP Extensions for Hub & Spoke Multipoint Label Switched Path**
                      **draft-ietf-mpls-mldp-hsmp-00.txt**

Abstract

   This draft introduces a hub & spoke multipoint LSP (short for HSMP
   LSP), which allows traffic both from root to leaf through P2MP LSP
   and also leaf to root along the co-routed reverse path.  That means
   traffic entering the HSMP LSP from application/customer at the root
   node travels downstream, exactly as if it was traveling downstream
   along a P2MP LSP to each leaf node, and traffic entering the HSMP LSP
   at any leaf node travels upstream along the tree to the root as if it
   is unicast to the root, except that it follows the path of the tree
   rather than ordinary unicast to the root.

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 [RFC2119].

Status of this Memo

Copyright Notice

Table of Contents

## 1.  Introduction

   The point-to-multipoint LSP defined in [RFC6388] allows traffic to
   transmit from root to several leaf nodes, and multipoint-to-
   multipoint LSP allows traffic from every node to transmit to every
   other node.  This draft introduces a hub & spoke multipoint LSP
   (short for HSMP LSP), which allows traffic both from root to leaf
   through P2MP LSP and also leaf to root along the co-routed reverse
   path.  That means traffic entering the HSMP LSP at the root node
   travels downstream, exactly as if it was traveling downstream along a
   P2MP LSP, and traffic entering the HSMP LSP at any other node travels
   upstream along the tree to the root.  A packet traveling upstream
   should be thought of as being unicast to the root, except that it
   follows the path of the tree rather than ordinary unicast to the
   root.

## 2.  Terminology

   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].

   This document uses some terms and acronyms as follows:

      mLDP: Multipoint extensions for LDP

      P2MP LSP: An LSP that has one Ingress LSR and one or more Egress
      LSRs.

      MP2MP LSP: An LSP that connects a set of nodes, such that traffic
      sent by any node in the LSP is delivered to all others.

      HSMP LSP: hub & spoke multipoint LSP.  An LSP allows traffic both
      from root to leaf through P2MP LSP and also leaf to root along the
      co-routed reverse path.

      PTP: The timing and synchronization protocol used by IEEE1588

## 3.  Applications

   In some cases, the P2MP LSP may not have a reply path for the OAM
   message (e.g, LSP Ping).  If P2MP LSP is provided by HSMP LSP, then
   the upstream path could be exactly used as the OAM message reply
   path.  This is especially useful in the case of P2MP LSP fault
   detection, performance measurement, root node redundancy and etc.
   There are several other applications that could take advantage of

   such kind of LDP based HSMP LSP as described below.

## 3.1.  Time synchronization scenario

   [IEEE1588] over MPLS is defined in [I-D.ietf-tictoc-1588overmpls].
   It is required that the LSP used to transport PTP event message
   between a Master Clock and a Slave Clock, and the LSP between the
   same Slave Clock and Master Clock must be co-routed.  By using point-
   to-multipoint technology to transmit PTP event messages from Master
   Clock at root side to Slave Clock at leaf side will greatly improve
   the bandwidth usage.  Unfortunately current point-to-multipoint LSP
   only provides unidirectional path from root to leaf, which cannot
   provide a co-routed reverse path for the PTP event messages.  LDP
   based HSMP LSP described in this draft provides unidirectional point-
   to-multipoint LSP from root to leaf and co-routed reverse path from
   leaf to root.

## 3.2.  VPMS scenario

   Point to multipoint PW described in [I-D.ietf-pwe3-p2mp-pw] requires
   to setup reverse path from leaf node (referred as egress PE) to root
   node (referred as ingress PE), if HSMP LSP is used to multiplex P2MP
   PW, the reverse path can also be multiplexed to HSMP upstream path to
   avoid setup independent reverse path.  In that case, the operational
   cost will be reduced for maintaining only one HSMP LSP, instead of
   P2MP LSP and n (number of leaf nodes) P2P reverse LSPs.

   The VPMS defined in [I-D.ietf-l2vpn-vpms-frmwk-requirements] requires
   reverse path from leaf to root node.  The P2MP PW multiplexed to HSMP
   LSP can provide VPMS with reverse path, without introducing
   independent reverse path from each leaf to root.

## 3.3.  IPTV scenario

   The mLDP based HSMP LSP can also be applied in a typical IPTV
   scenario.  There is usually only one location with senders but there
   are many receiver locations.  If IGMP is used for signaling between
   senders as IGMP querier and receivers, the IGMP messages from the
   receivers are travelling only from the leaves to the root (and from
   root towards leaves) but not from leaf to leaf.  In addition traffic
   from the root is only replicated towards the leaves.  Then leaf node
   receiving IGMP message (for SSM case) will join HSMP LSP, and then
   send IGMP message upstream to root along HSMP LSP.  Note that in
   above case, there is no node redundancy for IGMP querier.  And the
   node redundancy for IGMP querier could be provided by two independent
   VPMS instances with HSMP applied.

**4**.  **Setting up HSMP LSP with LDP**

   HSMP LSP is similar with MP2MP LSP described in [RFC6388], with the
   difference that the leaf LSRs can only send traffic to root node
   along the same path of traffic from root node to leaf node.

   HSMP LSP consists of a downstream path and upstream path.  The
   downstream path is same as MP2MP LSP, while the upstream path is only
   from leaf to root node, without communication between leaf and leaf
   nodes.  The transmission of packets from the root node of a HSMP LSP
   to the receivers is identical to that of a P2MP LSP.  Traffic from a
   leaf node follows the upstream path toward the root node, along the
   identical path of downstream path.

   For setting up the upstream path of a HSMP LSP, ordered mode MUST be
   used which is same as MP2MP.  Ordered mode can guarantee a leaf to
   start sending packets to root immediately after the upstream path is
   installed, without being dropped due to an incomplete LSP.

   Due to much of same behavior between HSMP LSP and MP2MP LSP, the
   following sections only describe the difference between the two
   entities.

**4.1**.  **Support for HSMP LSP setup with LDP**

   HSMP LSP also needs the LDP capabilities [RFC5561] to indicate the
   supporting for the setup of HSMP LSPs.  An implementation supporting
   the HSMP LSP procedures specified in this document MUST implement the
   procedures for Capability Parameters in Initialization Messages.
   Advertisement of the HSMP LSP Capability indicates support of the
   procedures for HSMP LSP setup.

   A new Capability Parameter TLV is defined, the HSMP LSP Capability.
   Following is the format of the HSMP LSP Capability Parameter.


```
    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
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |1|0|   HSMP LSP Cap(TBD IANA)    |         Length (= 1)        |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |1|  Reserved   |
   +-+-+-+-+-+-+-+-+
   Figure 1. HSMP LSP Capability Parameter encoding
```


   The HSMP LSP capability type is to be assigned by IANA.

## 4.2.  HSMP FEC Elements

Similar as MP2MP LSP, we define two new protocol entities, the HSMP
downstream FEC and upstream FEC Element.  If a FEC TLV contains an
HSMP FEC Element, the HSMP FEC Element MUST be the only FEC Element
in the FEC TLV.  The structure, encoding and error handling for the
HSMP downstream and upstream FEC Elements are the same as for the
MP2MP FEC Element described in [RFC6388] Section 4.2.  The difference
is that two additional new FEC types are used: HSMP downstream type
(TBD, IANA) and HSMP upstream type (TBD, IANA).

## 4.3.  Using the HSMP FEC Elements

In order to describe the message processing clearly, following
defines the processing of the HSMP FEC Elements, which is inherited
from [RFC6388] section 4.3.

1.  HSMP downstream LSP <X, Y> (or simply downstream <X, Y>): a HSMP
LSP downstream path with root node address X and opaque value Y.

2.  HSMP upstream LSP <X, Y> (or simply upstream <X, Y>): a HSMP LSP
upstream path for root node address X and opaque value Y which will
be used by any of downstream node to send traffic upstream to root
node.

3.  HSMP downstream FEC Element <X, Y>: a FEC Element with root node
address X and opaque value Y used for a downstream HSMP LSP.

4.  HSMP upstream FEC Element <X, Y>: a FEC Element with root node
address X and opaque value Y used for an upstream HSMP LSP.

5.  HSMP-D Label Map <X, Y, L>: A Label Map message with a single
HSMP downstream FEC Element <X, Y> and label TLV with label L. Label
L MUST be allocated from the per-platform label space of the LSR
sending the Label Map Message.

6.  HSMP-U Label Map <X, Y, Lu>: A Label Map message with a single
HSMP upstream FEC Element <X, Y> and label TLV with label Lu.  Label
Lu MUST be allocated from the per-platform label space of the LSR
sending the Label Map Message.

### 4.3.1.  HSMP LSP Label Map

This section specifies the procedures for originating HSMP Label Map
messages and processing received HSMP label map messages for a
particular HSMP LSP.  The procedure of downstream HSMP LSP is same as
that of downstream MP2MP LSP described in [RFC6388].  Under the
operation of ordered mode, the upstream LSP will be setup by sending

HSMP LSP mapping message with label which is allocated by upstream
LSR to its downstream LSR one by one from root to leaf node,
installing the upstream forwarding table by every node along the LSP.
Detail procedure of upstream HSMP LSP is different with that of
upstream MP2MP LSP, and is specified in below section.

All labels discussed here are downstream-assigned [RFC5332] except
those which are assigned using the procedures described in section 5.

Determining the upstream LSR for a HSMP LSP <X, Y> follows the
procedure for a MP2MP LSP described in [RFC6388] Section 4.3.1.1.

Determining one's downstream HSMP LSR procedure is much same as
defined in [RFC6388] section 4.3.1.2.  A LDP peer U which receives a
HSMP-D Label Map from a LDP peer D will treat D as downstream HSMP
LSR.

Determining the forwarding interface to an LSR has same procedure as
defined in [RFC6388] section 2.4.1.2.

### 4.3.1.1.  HSMP LSP leaf node operation

The leaf node operation is same as the operation of MP2MP LSP defined
in [RFC6388] section 4.3.1.4, only with different FEC element
processing and specified below.

A leaf node Z will send a HSMP-D Label Map <X, Y, L> to U, instead of
MP2MP-D Label Map <X, Y, L>. and expects a HSMP-U Label Map <X, Y,
Lu> from node U and checks whether it already has forwarding state
for upstream <X, Y>.  The created forwarding state on leaf node Z is
same as the leaf node of MP2MP LSP.  Z will push label Lu onto the
traffic that Z wants to forward over the HSMP LSP.

### 4.3.1.2.  HSMP LSP transit node operation

Suppose node Z receives a HSMP-D Label Map <X, Y, L> from LSR D, the
procedure is same as processing MP2MP-D Label Mapping message defined
in [RFC6388] section 4.3.1.5, and the processing protocol entity is
HSMP-D label mapping message.  The different procedure is specified
below.

Node Z checks if upstream LSR U already assigned a label Lu to
upstream <X, Y>.  If not, transit node Z waits until it receives a
HSMP-U Label Map <X, Y, Lu> from LSR U. Once the HSMP-U Label Map is
received from LSR U, node Z checks whether it already has forwarding
state upstream <X, Y> with incoming label Lu' and outgoing label Lu.
If it does, Z sends a HSMP-U Label Map <X, Y, Lu'> to downstream
node.  If it does not, it allocates a label Lu' and creates a new

label swap for Lu' with Label Lu over interface Iu.  Interface Iu is
determined via the procedures in Section 4.3.1.  Node Z determines
the downstream HSMP LSR as per Section 4.3.1, and sends a HSMP-U
Label Map <X, Y, Lu'> to node D.

Since a packet from any downstream node is forwarded only to the
upstream node, the same label (representing the upstream path) can be
distributed to all downstream nodes.  This differs from the
procedures for MPMP LSPs [RFC6388], where a distinct label must be
distributed to each downstream node.  The forwarding state upstream
<X, Y> on node Z will be like this {<Lu'>, <Iu Lu>}.  Iu means the
upstream interface over which Z receives HSMP-U Label Map <X, Y, Lu>
from LSR U. Packets from any downstream interface over which Z send
HSMP-U Label Map <X, Y, Lu'> with label Lu' will be forwarded to Iu
with label Lu' swap to Lu.

### 4.3.1.3.  HSMP LSP root node operation

Suppose root node Z receives a HSMP-D Label Map <X, Y, L> from node
D, the procedure is much same as processing MP2MP-D Label Mapping
message defined in [RFC6388] section 4.3.1.6, and the processing
protocol entity is HSMP-D label mapping message.  The different
procedure is specified below.

Node Z checks if it has forwarding state for upstream <X, Y>.  If
not, Z creates a forwarding state for incoming label Lu' that
indicates that Z is the LSP egress.  E.g., the forwarding state might
specify that the label stack is popped and the packet passed to some
specific application.  Node Z determines the downstream HSMP LSR as
per section 4.3.1, and sends a HSMP-U Label Map <X, Y, Lu'> to node
D.

Since Z is the root of the tree, Z will not send a HSMP-D Label Map
and will not receive a HSMP-U Label Map.

### 4.3.2.  HSMP LSP Label Withdraw

The HSMP Label Withdraw procedure is much same as MP2MP leaf
operation defined in [RFC6388] section 4.3.2, and the processing
protocol entities are HSMP FECs.  The only difference is process of
HSMP-U label release message, which is specified below.

When a transit node Z receives a HSMP-U label release message from
downstream node D, Z should check if there are any incoming interface
in forwarding state upstream <X, Y>.  If all downstream nodes are
released and there is no incoming interface, Z should delete the
forwarding state upstream <X, Y> and send HSMP-U label release
message to its upstream node.

**4.3.3**.  **HSMP LSP upstream LSR change**

   The procedure for changing the upstream LSR is the same as defined in
   [RFC6388] section 4.3.3, except it is applied to HSMP FECs.

**5**.  **HSMP LSP on a LAN**

   The procedure to process P2MP LSP on a LAN has been described in
   [RFC6388].  When the LSR forwards a packet downstream on one of those
   LSPs, the packet's top label must be the "upstream LSR label", and
   the packet's second label is "LSP label".

   When establishing the downstream path of a HSMP LSP, as defined in
   [RFC6389], a label request for a LSP label is send to the upstream
   LSR.  The upstream LSR should send label mapping that contains the
   LSP label for the downstream HSMP FEC and the upstream LSR context
   label.  At the same time, it must also send label mapping for
   upstream HSMP FEC to downstream node.  Packets sent by the upstream
   router can be forwarded downstream using this forwarding state based
   on a two label lookup.  Packets traveling upstream need to be
   forwarded in the direction of the root by using the label allocated
   by upstream LSR.

**6**.  **Redundancy considerations**

   In some scenario, it is necessary to provide two root nodes for
   redundancy purpose.  One way to implement this is to use two
   independent HSMP LSPs acting as active/standby.  At one time, only
   one HSMP LSP will be active, and the other will be standby.  After
   detecting the failure of active HSMP LSP, the root and leaf nodes
   will switch the traffic to the new active HSMP LSP which is switched
   from former standby LSP.  The detail of redundancy mechanism will be
   for future study.

**7**.  **Co-routed path exceptions**

   There are some exceptional cases that mLDP based HSMP LSP could not
   achieve co-routed path.  One possible case is using static routing
   between LDP neighbors; another possible case is IGP cost asymmetric
   generated by physical link cost asymmetric, or TE-Tunnels used
   between LDP neighbors.  The LSR/LER in HSMP LSP could detect if the
   path is co-routed or not, if not co-routed, an indication could be
   generated to the management system.

## 8.  Security Considerations

The same security considerations apply as for the MP2MP LSP described
in [RFC6388].

## 9.  IANA Considerations

This document requires allocation of two new LDP FEC Element types
from the "Label Distribution Protocol (LDP) Parameters registry" the
"Forwarding Equivalence Class (FEC) Type Name Space":

1. the HSMP-upstream FEC type - requested value TBD

2. the HSMP-downstream FEC type - requested value TBD

This document requires allocation of one new code points for the HSMP
LSP capability TLV from "Label Distribution Protocol (LDP) Parameters
registry" the "TLV Type Name Space":

HSMP LSP Capability Parameter - requested value TBD

## 10.  Acknowledgement

The author would like to thank Eric Rosen, Sebastien Jobert, Fei Su,
Edward, Mach Chen, Thomas Morin for their valuable comments.

## 11.  References

## 11.1.  Normative references

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

[RFC5332]   Eckert, T., Rosen, E., Aggarwal, R., and Y. Rekhter, "MPLS
            Multicast Encapsulations", RFC 5332, August 2008.

[RFC5561]   Thomas, B., Raza, K., Aggarwal, S., Aggarwal, R., and JL.
            Le Roux, "LDP Capabilities", RFC 5561, July 2009.

[RFC6388]   Wijnands, IJ., Minei, I., Kompella, K., and B. Thomas,
            "Label Distribution Protocol Extensions for Point-to-
            Multipoint and Multipoint-to-Multipoint Label Switched
            Paths", RFC 6388, November 2011.

[RFC6389]   Aggarwal, R. and JL. Le Roux, "MPLS Upstream Label

                    Assignment for LDP", RFC 6389, November 2011.

**11.2.  Informative References**

   [I-D.ietf-l2vpn-vpms-frmwk-requirements]
              Kamite, Y., JOUNAY, F., Niven-Jenkins, B., Brungard, D.,
              and L. Jin, "Framework and Requirements for Virtual
              Private Multicast Service (VPMS)",
              draft-ietf-l2vpn-vpms-frmwk-requirements-04 (work in
              progress), July 2011.

   [I-D.ietf-pwe3-p2mp-pw]
              Sivabalan, S., Boutros, S., and L. Martini, "Signaling
              Root-Initiated Point-to-Multipoint Pseudowire using LDP",
              draft-ietf-pwe3-p2mp-pw-04 (work in progress), March 2012.

   [I-D.ietf-tictoc-1588overmpls]
              Davari, S., Oren, A., Bhatia, M., Roberts, P., and L.
              Montini, "Transporting PTP messages (1588) over MPLS
              Networks", draft-ietf-tictoc-1588overmpls-02 (work in
              progress), October 2011.

   [IEEE1588]
              "IEEE standard for a precision clock synchronization
              protocol for networked measurement and control systems",
              IEEE1588v2 , March 2008.

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

   [RFC6374]  Frost, D. and S. Bryant, "Packet Loss and Delay
              Measurement for MPLS Networks", RFC 6374, September 2011.


Authors' Addresses

   Lizhong Jin
   ZTE Corporation
   889, Bibo Road
   Shanghai, 201203, China

   Email: lizhong.jin@zte.com.cn

    Frederic Jounay
    France Telecom
    2, avenue Pierre-Marzin
    22307 Lannion Cedex, FRANCE

    Email: frederic.jounay@orange.ch


    IJsbrand Wijnands
    Cisco Systems, Inc
    De kleetlaan 6a
    Diegem  1831, Belgium

    Email: ice@cisco.com


    Nicolai Leymann
    Deutsche Telekom AG
    Winterfeldtstrasse 21
    Berlin  10781

    Email: N.Leymann@telekom.de