Network Working Group L. Jin
Internet-Draft
Intended status: Standards Track F. Jounay
Expires: April 19, 2014 France Telecom
I. Wijnands
Cisco Systems
N. Leymann
Deutsche Telekom
October 16, 2013

LDP Extensions for Hub & Spoke Multipoint Label Switched Path
draft-ietf-mpls-mldp-hsmp-03.txt

Abstract

This draft introduces a hub & spoke multipoint LSP (or HSMP LSP for short), 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 to each leaf node, exactly as if it is travelling downstream along a P2MP LSP to each leaf node. Upstream 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, and strictly follows the downstream path of the tree rather than routing protocol based unicast path 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].

This Internet-Draft will expire on April 19, 2014.

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 (or HSMP LSP for short), 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 is travelling 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 travelling upstream should be thought of as being unicast to the root, except that it follows the path of the tree rather than routing protocol based unicast path to the root. The combination of upstream LSPs initiated from all leaf nodes forms a multipoint-to-point LSP.

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:

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.

mLDP: Multipoint extensions for LDP

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.

PTP: The timing and synchronization protocol used by IEEE1588

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

3. Applications

In some cases, the P2MP LSP may not have a reply path for OAM messages (e.g, LSP Ping Echo Request). If P2MP LSP is provided by HSMP LSP instead, then the upstream path could be 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 a 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. 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 provides 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 LSP from leaf to root.

3.2. Virtual Private Multicast Service Scenario

Point to multipoint PW described in [I-D.ietf-pwe3-p2mp-pw] requires to set up 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 signalling between senders as IGMP querier [RFC3376] 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 report message (for source specific multicast case) will join HSMP LSP(use similar mechanism in [RFC6826] section 2), and then send IGMP report message upstream to root along HSMP upstream LSP. Note that in above case, there is no node redundancy for IGMP querier. And the node redundancy for IGMP querier[RFC3376] 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 an 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 a path that traverse the same nodes as the downstream node, but in reverse order.

For setting up the upstream path of an 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 similar behaviors 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 requires the LDP capabilities [RFC5561] for nodes to indicate that they support 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 Parameter. 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              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|S|  Reserved   |
+-+-+-+-+-+-+-+-+
Figure 1. HSMP LSP Capability Parameter encoding
```

The length SHOULD be 1, and the S bit and reserved bits are defined in [RFC5561] section 3.

The HSMP LSP Capability Parameter 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 Element and Upstream FEC Element. If a FEC TLV contains one of the HSMP FEC Elements, the HSMP FEC Element MUST be the only FEC Element in the FEC TLV. The structure, encoding and error handling for the HSMP Downstream FEC Element and Upstream FEC Element are the same as for the MP2MP FEC Element described in [RFC6388] Section 3.2. The difference is that two additional new FEC types are defined: HSMP Downstream FEC (TBD, IANA) and HSMP Upstream FEC(TBD, IANA).

4.3. Using the HSMP FEC Elements

In order to describe the message processing clearly, the entries in the list below define the processing of the HSMP FEC Elements. Additionally, the entries defined in [RFC6388] section 3.3 are also reused in the following sections.

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

2. HSMP upstream LSP <X, Y> (or simply upstream <X, Y>): an 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 Mapping <X, Y, L>: A Label Mapping 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 Mapping Message.

6. HSMP-U Label Mapping <X, Y, Lu>: A Label Mapping 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 Mapping Message.

4.3.1. HSMP LSP Label Map

This section specifies the procedures for originating HSMP Label Mapping messages and processing received HSMP Label Mapping messages for a particular HSMP LSP. The procedure of downstream HSMP LSP is same as that of downstream MP2MP LSP described in [RFC6388]. When LDP operates in Ordered Label Distribution Control mode [RFC5036], the upstream LSP will be set up by sending HSMP LSP LDP Label Mapping message with a label which is allocated by upstream LSR to its downstream LSR hop by hop from root to leaf node, installing the upstream forwarding table by every node along the LSP. The detail procedure of setting up 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 the HSMP LSP <X, Y> follows the procedure for a MP2MP LSP described in [RFC6388] Section 3.3.1.1.

Determining one's HSMP downstream LSR follows the procedure defined in [RFC6388] section 3.3.1.2. That is, an upstream LDP peer which receives a Label Mapping with HSMP downstream FEC Element from an LDP peer D will treat D as HSMP downstream LDP peer.

Determining the forwarding interface to an LSR follows the 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 3.3.1.4. The only difference is the FEC elements as specified below.

A leaf node Z will send an HSMP-D Label Mapping <X, Y, L> to U, instead of MP2MP-D Label Mapping <X, Y, L>, and expects an HSMP-U Label Mapping <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 an HSMP-D Label Mapping <X, Y, L> from LSR D, the procedure is much the same as processing MP2MP-D Label Mapping message defined in [RFC6388] section 3.3.1.5, and the processing protocol entity is HSMP-D Label Mapping message. The only difference is specified below.

Node Z checks if upstream LSR U already has assigned a label Lu to upstream <X, Y>. If not, transit node Z waits until it receives an HSMP-U Label Mapping <X, Y, Lu> from LSR U. Once the HSMP-U Label Mapping 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 an HSMP-U Label Mapping <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 an HSMP-U Label Mapping <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) SHOULD be distributed to all downstream nodes. This differs from the procedures for MP2MP 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 sends 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 an HSMP-D Label Mapping <X, Y, L> from node D, the procedure is much the same as processing MP2MP-D Label Mapping message defined in [RFC6388] section 3.3.1.6, and the processing protocol entity is HSMP-D Label Mapping message. The only difference 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 an HSMP-U Label Map <X, Y, Lu'> to node D.

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

4.3.2. HSMP LSP Label Withdraw

The HSMP Label Withdraw procedure is much the same as MP2MP leaf operation defined in [RFC6388] section 3.3.2, and the processing FEC Elements are HSMP FEC Elements. The only difference is the process of HSMP-U Label Release message, which is specified below.

When a transit node Z receives an 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. Otherwise, no HSMP-U Label Release message will be sent to the 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 3.3.3, only with different processing FEC Element, the HSMP FEC Element.

5. HSMP LSP on a LAN

The procedure to process the downstream HSMP LSP on a LAN is much the same as downstream MP2MP LSP described in [RFC6388] section 6.1.1.

When establishing the downstream path of an HSMP LSP, as defined in [RFC6389], a Label Request message for an LSP label is sent to the upstream LSR. The upstream LSR should send Label Mapping message that contains the LSP label for the downstream HSMP FEC and the upstream LSR context label defined in [RFC5331]. When the LSR forwards a packet downstream on one of those LSPs, the packet's top label must be the "upstream LSR context label", and the packet's second label is "LSP label". The HSMP downstream path will be installed in the context-specific forwarding table corresponding to the upstream LSR label. Packets sent by the upstream LSR can be forwarded downstream using this forwarding state based on a two-label lookup.

The upstream path of an HSMP LSP on a LAN is the same as the one on other kind of links. That is, the upstream LSR must send Label

Mapping message that contains the LSP label for upstream HSMP FEC to downstream node. Packets travelling upstream need to be forwarded in the direction of the root by using the label allocated for upstream HSMP FEC.

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 standby HSMP LSP which takes on the role as active HSMP LSP. The detail of redundancy mechanism is out of the scope.

7. Co-routed Path Exceptions

There are some exceptional cases when 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 should detect if the path is co-routed or not. If not co-routed, an alarm indication should be generated to the management system.

8. Failure Detection of HSMP LSP

The idea of LSP ping for HSMP LSPs could be expressed as an intention to test the LSP Ping Echo Request packets that enter at the root along a particular downstream path of HSMP LSP, and end their MPLS path on the leaf. The leaf node then sends the LSP Ping Echo Reply along the co-routed upstream path of HSMP LSP, and end on the root that are the (intended) root node.

New sub-TLVs are required to be assigned by IANA in Target FEC Stack TLV to define the corresponding HSMP-upstream FEC type and HSMP-downstream FEC type. In order to ensure the leaf node to send the LSP Ping Echo Reply along the HSMP upstream path, the R bit (Validate Reverse Path) in Global Flags Field defined in [RFC6426] is reused here.

The node processing mechanism of LSP Ping Echo Request and Echo Reply for HSMP LSP is inherited from [RFC6425] and [RFC6426] section 3.4, except the following:

1. The root node sending LSP Ping Echo Request message for HSMP LSP MUST attach Target FEC Stack with HSMP downstream FEC, and set R bit to '1' in Global Flags Field.

2. When the leaf node receiving the LSP Ping Echo Request, it MUST send the LSP Ping Echo Reply to the associated HSMP upstream path. The Reverse-path Target FEC Stack TLV attached by leaf node in Echo Reply message SHOULD contain the sub-TLV of associated HSMP upstream FEC.

9. Security Considerations

The same security considerations apply as for the MP2MP LSP described in [RFC6388] and [RFC6425].

Although this document introduces new FEC Elements and corresponding procedures, the protocol does not bring any new security issues compared to [RFC6388] and [RFC6425].

10. IANA Considerations

10.1. New LDP FEC Element types

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

The values should be allocated using the lowest free values from the "IETF Consensus"-range (0-127).

10.2. HSMP LSP capability TLV

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

The value should be allocated from the range 0x0901-0x3DFF (IETF Consensus) using the first free value within this range.

10.3. New sub-TLVs for the Target Stack TLV

This document requires allocation of two new sub-TLV types for inclusion within the LSP ping [RFC4379] Target FEC Stack TLV (TLV type 1).

1. the HSMP-upstream LDP FEC Stack - requested value TBD

2. the HSMP-downstream LDP FEC Stack - requested value TBD

The value should be allocated from the IETF Standards Action range (0-16383) that is used for mandatory and optional sub-TLVs that requires a response if not understood. The value should be allocated using the lowest free value within this range.

11. Acknowledgement

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

12. References

12.1. Normative references

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

[RFC5331] Aggarwal, R., Rekhter, Y., and E. Rosen, "MPLS Upstream Label Assignment and Context-Specific Label Space", RFC 5331, August 2008.

[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.

[RFC6425] Saxena, S., Swallow, G., Ali, Z., Farrel, A., Yasukawa, S., and T. Nadeau, "Detecting Data-Plane Failures in Point-to-Multipoint MPLS - Extensions to LSP Ping", RFC 6425, November 2011.

[RFC6426] Gray, E., Bahadur, N., Boutros, S., and R. Aggarwal, "MPLS On-Demand Connectivity Verification and Route Tracing", RFC 6426, November 2011.

12.2. 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-05 (work in progress), October 2012.

[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 Timing messages over MPLS Networks", draft-ietf-tictoc-1588overmpls-05 (work in progress), June 2013.

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

[RFC3376] Cain, B., Deering, S., Kouvelas, I., Fenner, B., and A. Thyagarajan, "Internet Group Management Protocol, Version 3", RFC 3376, October 2002.

[RFC4379] Kompella, K. and G. Swallow, "Detecting Multi-Protocol Label Switched (MPLS) Data Plane Failures", RFC 4379, February 2006.

[RFC5036] Andersson, L., Minei, I., and B. Thomas, "LDP Specification", RFC 5036, October 2007.

[RFC6826] Wijnands, IJ., Eckert, T., Leymann, N., and M. Napierala, "Multipoint LDP In-Band Signaling for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths",

RFC 6826, January 2013.

Authors' Addresses

Lizhong Jin
Shanghai, China

Email: lizho.jin@gmail.com

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