Network Working Group B Decraene
Internet Draft JL Le Roux
Document: draft-decraene-mpls-ldp-interarea-04.txt France Telecom
Expiration Date: September 2007
I Minei
Juniper Networks, Inc.

March 2007

LDP extension for Inter-Area LSP

Status of this Memo

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with Section 6 of BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

Abstract

To facilitate the establishment of Label Switched Paths (LSP) that would span multiple IGP areas in a given Autonomous System (AS), this document proposes a new optional label mapping procedure for the Label Distribution Protocol (LDP).

This procedure allows the use of a label if the Forwarding Equivalence Class (FEC) Element matches an entry in the routing table (RIB). Matching is defined by an IP longest match search and does not mandate an exact match.

1. 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.

2. Terminology

IGP Area: OSPF Area or IS-IS level

ABR: OSPF Area Border Router or IS-IS L1/L2 router

LSP: Label Switched Path

Intra-area LSP: LSP that does not traverse any IGP area boundary.

Inter-area LSP: LSP that traverses at least one IGP area boundary.

3. Introduction

Link state IGPs such as OSPF [OSPFv2] and IS-IS [IS-IS] allow the partition of an autonomous system into areas or levels so as to increase routing scalability within a routing domain.

However, [LDP] requires that the IP address of the FEC Element should *exactly* match an entry in the IP RIB: according to [LDP] section 3.5.7.1 (Label Mapping Messages Procedures) "An LSR receiving a Label Mapping message from a downstream LSR for a Prefix or Host Address FEC Element should not use the label for forwarding unless its routing table contains an entry that exactly matches the FEC Element".

Therefore, MPLS LSPs between LERs in different areas/levels are not setup unless the exact (/32 for IPv4) loopback addresses of all the LERs are redistributed across all areas.

The problem statement is discussed in section 3. Then, in section 4 we extend the Label Mapping Procedure defined in [LDP] so as to support the setup of contiguous inter-area LSPs while maintaining IP prefix aggregation on the ABRs. This basically consists of allowing for "Longest Match Based" Label Mapping.

4. Problem statement

Provider based MPLS VPN networks are expanding with the success of Layer 3 VPN ([L3-VPN]) and the new deployments of layer 2 VPNs ([VPLS-BGP], [VPLS-LDP]). Service Provider MPLS backbones are significantly growing both in terms of density with the addition of PEs to connect new customers and in terms of footprint as traditional

layer two aggregation networks are being replaced by IP/MPLS networks. As a consequence many providers need to introduce IGP areas. Inter-area LSPs, that is LSPs that traverse at least two IGP areas are required to ensure MPLS connectivity between PEs located in distinct IGP areas.

To set up the required MPLS LSPs between PEs in different IGP areas, services providers have currently three solutions: LDP with IGP route leaking, BGP [MPLS-BGP] over LDP with MPLS hierarchy, or also inter-area RSVP-TE [ID-RSVP-TE].

IGP route leaking consists in redistributing all /32 PE loopback addresses across area boundaries. As a result, LDP finds in the RIB an exact match for its FEC and sets up the LSP.
As a consequence, the potential benefits that a multi-area domain may yield are significantly diminished since a lot of addresses have to be redistributed by ABRs, and the number of IP entries in the LSDB and RIB maintained by every LSR of the domain (whatever the area/level it belongs to) cannot be minimized.

Service providers may also set up these inter-area LSPs by using MPLS hierarchy with BGP [MPLS-BGP] as a label distribution protocol between areas. The BGP next hop would typically be the ABRs and the BGP-created LSPs would be nested within intra-area LSPs setup by LDP between PEs and ABRs and between ABRs.
This solution is not adequate for Service Providers which don't want to run BGP on their P routers as it requires BGP on all ABRs. In addition, this scheme has an impact on the availability, as the recovery upon ABR failure relies on BGP convergence. Also MPLS hierarchy does not allow locally protecting the LSP against ABR failures (LDP Fast Reroute), and hence ensuring sub-50ms recovery upon ABR failure. The resulting convergence time may not be acceptable for stringent SLAs required for voice or mission critical applications. Finally, this solution requires a significant migration effort for Service Providers which started with LDP and IGP route

leaking to quickly set-up the fist inter-area LSPs.

Service providers may also setup these inter-area LSPs by using inter-area RSVP-TE [ID-RSVP-TE]. This is a relevant solution when RSVP-TE is already used for setting up intra-area LSPs, and inter-area traffic engineering features are required. In return this is not a desired solution when LDP is already used for setting up intra-area LSPs, and inter-area traffic engineering features are not required.

To avoid the above drawbacks, there is a need for an LDP based solution which allows setting up contiguous inter-area LSPs while avoiding leaking of /32 PE loopback addresses across area boundaries, and hence keeping all the benefits of IGP hierarchy.

In that context, this document defines a new LDP Label Mapping Procedure so as to support the setup of contiguous inter-area LSPs

while maintaining IP prefix aggregation on the ABRs. This procedure is similar to the one defined in [LDP] but performs a longest match when searching the FEC element in the RIB.

5. Label Mapping Procedure

This document defines a new label mapping procedure for LDP. It MUST be possible to activate/deactivate this procedure by configuration and it SHOULD be deactivated by default. It MAY be possible to activate it on a per prefix basis.

With this new longest match label mapping procedure, a LSR receiving a Label Mapping message from a neighbor LSR for a Prefix Address FEC Element SHOULD use the label for MPLS forwarding if its routing table contains an entry that matches the FEC Element and the advertising LSR is a next hop to reach the FEC. If so, it SHOULD advertise the FEC Element and a label to its LDP peers.

By "matching FEC Element", one should understand an IP longest match. That is, either the LDP FEC element exactly matches an entry in the IP RIB or the FEC element is a subset of an IP RIB entry. There is no match for other cases such as the FEC element is a superset of a RIB entry.

Note that with this longest match Label Mapping Procedure, each LSP established by LDP still strictly follows the shortest path(s) defined by the IGP.

FECs selected by this "Longest Match" label mapping procedure will be distributed in an ordered way. However this procedure is applicable to both independent and ordered distribution control mode.

As per RFC 3036, LDP has already some interactions with the RIB. In particular, it needs to be aware of the following events:

- prefix UP when a new IP prefix appears in the RIB
- prefix DOWN when an existing prefix disappears
- next-hop change when an existing prefix have new next hop following a routing change.

With the longest match procedure, multiple FECs may be concerned by a single RIB prefix change. The LSR must check all the FECs which are a subset of this RIB prefix. So some LDP reactions following a RIB event are changed:

- When a new prefix appears in the RIB, the LSR MUST check if this prefix is a better match for some existing FECs. E.g. the FEC elements 192.0.2.1/32 and 192.0.2.2/32 used the IP RIB entry 192.0.0/16 and a new more specific IP RIB entry 192.0.2/24 appears. This may result in changing the LSR used as next hop and hence the NHLFE for this FEC.
- When a prefix disappears in the RIB, the LSR MUST check all FEC elements which are using this RIB prefix as best match. For each

FEC, if another RIB prefix is found as best match, LDP MUST use it. This may result in changing the LSR used as next hop and hence the NHLFE for this FEC. Otherwise, the LSR MUST remove the FEC binding and send a label withdraw message.
- When the next-hop of a RIB prefix change, the LSR must change the NHLFE of all the FEC elements using this prefix.

6. Application examples

6.1. Inter-area LSPs

Consider the following example of an autonomous system with one backbone area and two edge areas:

```
                         Area "B"

                 Level 2 / Backbone area
```

```
                 +--------------------------+
        Area "A" |                          |  Area "C"
                 |                          |
        Level 1  |                          |  Level 1 / area
                 |        P1                |
      +----------+                          +-------------+
      |          |                 P2       |         PE1 | 192.0.2.1/32
      |          |                          |             |
      |PE4      ABR2                       ABR1       PE2 | 192.0.2.2/32
      |          |        P3                |             |
      |          |                          |         PE3 | 192.0.2.3/32
      +----------+                          +-------------+
                 |                          |
                 +--------------------------+
```

Figure 1: An IGP domain with two areas attached to the Backbone Area.

Note that this applies equally to IS-IS and OSPF. An ABR refers here either to an OSPF ABR or to an IS-IS L1/L2 node.

All routers are MPLS enabled and MPLS connectivity (ie an LSP) is required between all PE routers.

In the "egress" area "C", the records available are:

```
IGP RIB                          LDP FEC elements:
  192.0.2.1/32                      192.0.2.1/32
  192.0.2.2/32                      192.0.2.2/32
  192.0.2.3/32                      192.0.2.3/32
```


The area border router ABR1 advertises in the backbone area:

- the aggregated IP prefix 192.0.2/24 in the IGP
- all the individual IP FEC elements (/32) in LDP

In the "backbone" area "B", the records available are:

```
IGP RIB                          LDP FEC elements:
  192.0.2/24                       192.0.2.1/32
                                   192.0.2.2/32
                                   192.0.2.3/32
```

The area border router ABR2 advertises in the area "A":
- an aggregated IP prefix 192.0/16 in the IGP
- all the individual IP FEC elements (/32) in LDP

In the "ingress" area "A", the records available are:

```
IGP RIB                         LDP FEC elements:
  192.0/16                        192.0.2.1/32
                                  192.0.2.2/32
                                  192.0.2.3/32
```

In this situation, one LSP is established between the ingress PE4 and every egress PE of area C while maintaining IP prefix aggregation on the ASBRs.

6.2. Use of static routes

Consider the following example where a LER is dual-connected to two LSRs:

```
 +--------LSR1----
 |         |
LER        |
 |         |
 +--------LSR2----
```

Figure 2: LER dual-connected to two LSRs.

In some situations, especially on the edge of the network, it is valid to use static IP routes between the LER and the two LSRs. If necessary, the BFD protocol can be used to quickly detect loss of connectivity.

The LDP specification defined in [LDP] would require on the ingress LER the configuration and the maintenance of one IP route per egress LER and per outgoing interface.

The longest match Label Mapping Procedure described in this document only requires one IP route per outgoing interface.

7. Caveats for deployment

7.1. Deployment consideration

LSRs compliant with this document are backward compatible with LSRs that comply with [LDP].

For the successful establishment of end-to-end MPLS LSPs whose FEC are aggregated in the RIB, this specification must be implemented on all LSRs in all areas where IP aggregation is used.

If all IP prefixes are leaked in the IGP backbone area and only stub areas use IP aggregation, LSRs in the backbone area don't need to be compliant with this document.

7.2. Impact on routing convergence time

In case of an egress LER failure, performing IP route aggregation on ABRs will change the routing convergence behavior. The IGP will not propagate the notification of the egress LER failure outside of the egress area and failure notification will rely on LDP signaling through the end-to-end propagation of the LDP withdraw message. This failure notification may be faster or slower depending on the implementations, the IGP timers used and the network topology (network diameter).

For failure of links and other nodes (Ps, ABRs), the failure notification and the convergence is unchanged. The convergence time may be improved because the RIB has fewer entries to update.

8. Security Considerations

The longest match Label Mapping procedure described in this document does not introduce any change as far as the Security Consideration section of [LDP] is concerned.

9. References

9.1. Normative References

[LDP] L. Andersson, P. Doolan, N. Feldman, A. Fredette, B. Thomas, "LDP Specification", RFC 3036, January 2001

[MPLS] E. Rosen, A. Viswanathan, R. Callon, " Multiprotocol Label Switching Architecture", RFC 3031, January 2001

9.2. Informative References

[MP-BGP]   Bates, T., Chandra, R., Katz, D. and Rekhter, Y, "Multiprotocol Extensions for BGP-4", RFC 4760, January 2007.

[L3-VPN]   Rosen, E., Rekhter, Y. ," BGP/MPLS IP Virtual Private Networks (VPNs) ", RFC 4374, February 2006

[MPLS-BGP] Rekhter, Y., Rosen, E., "Carrying Label Information in BGP-4", RFC 3107, May 2001

[OSPFv2]   Moy, J.,"OSPF Version 2", RFC 2328, April 1998

[IS-IS]   Callon, R., "Use of OSI IS-IS for Routing in TCP/IP and Dual Environments", RFC 1195, December 1990

[VPLS-BGP] Kompella, K., Rekhter, Y., "Virtual Private LAN Service (VPLS) Using BGP for Auto-discovery and Signaling", RFC 4761, January 2007.

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

[ID-RSVP-TE] Farrel, Ayyangar, Vasseur, " Inter domain MPLS and GMPLS Traffic Engineering - RSVP-TE extensions", draft-ietf-ccamp-inter-domain-rsvp-te, work in progress.

10. Acknowledgments

Authors would like to thank Yakov Rekhter, Stefano Previdi, Vach Kompella, Benoit Fondeviole, Gilles Bourdon and Christian Jacquenet for the useful discussions on this subject, their review and comments.

11. Author's Addresses

Bruno Decraene
France Telecom
38-40 rue du General Leclerc
92794 Issy Moulineaux cedex 9
France
bruno.decraene@orange-ftgroup.com

Jean-Louis Le Roux
France Telecom
2, avenue Pierre-Marzin

22307 Lannion Cedex
France
jeanlouis.leroux@orange-ftgroup.com

Ina Minei
Juniper Networks
1194 N. Mathilda Ave.
Sunnyvale, CA 94089
ina@juniper.net

Intellectual Property Considerations

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Copyright Statement

Copyright (C) The IETF Trust (2007).

This document is subject to the rights, licenses and restrictions contained in BCP 78, and except as set forth therein, the authors retain all their rights.

Disclaimer of Validity

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Acknowledgment

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