Network Working Group
Internet Draft

Category: Informational
Expires: January 2006

J.-L. Le Roux
T. Morin
France Telecom
Vincent Parfait
Equant
Luyuan Fang
AT&T
Lei Wang
Telenor
Yuji Kamite
NTT Communications
Shane Amante
Level 3 Communications

July 2005

Requirements for multipoint extensions to the Label Distribution Protocol

draft-leroux-mpls-mp-ldp-reqs-00.txt

Status of this Memo

Abstract

This document lists a set of functional requirements for Label Distribution Protocol (LDP) extensions for setting up point-to-multipoint (P2MP) and potentially multipoint-to-multipoint (MP2MP) Label Switched Paths (LSP), in order to deliver point-to-multipoint applications over a Multi Protocol Label Switching (MPLS) infrastructure. It is intended that solutions that specify LDP procedures for setting up P2MP and MP2MP LSP satisfy these requirements.

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.

Table of Contents

1. Terminology

LSR: Label Switching Router

LSP: MPLS Label Switched Path

Ingress LSR: Router acting as a sender of an LSP

Egress LSR: Router acting as a receiver of an LSP

P2P LSP: A LSP that has one unique Ingress LSR and one unique Egress LSR

MP2P LSP: A LSP that has one or more Ingress LSRs and one unique Egress LSR

P2MP LSP: A LSP that has one unique Ingress LSR and one or more Egress LSRs

MP2MP LSP: A bidirectional LSP connecting a group of two or more LSRs acting equally as Ingress LSR or Egress LSR

Leaf LSR: Egress LSR of a P2MP LSP or Ingress/Egress LSR of a MP2MP LSP

MP LSP: P2MP LSP or MP2MP LSP

Transit LSR: A LSR of a MP LSP that has one or more downstream LSRs

Branch LSR: A LSR of a P2MP LSP that has more than one downstream LSRs

Hub LSR: A LSR of a MP2MP LSP that has two or more neighbour LSRs

Bud LSR: A LSR of a MP LSP that is an egress but also has one or more directly connected downstream LSRs

2. Introduction

Many operators have deployed LDP [LDP] for setting up point-to-point (P2P) and multipoint-to-point (MP2P) LSPs, in order to offer point-to -point services in MPLS backbones.

There are emerging requirements for supporting delivery of point-to-multipoint applications in MPLS backbones, such as those defined in [L3VPN-MCAST] and [L2VPN-MCAST].

An interesting and useful approach for operators who want to support point-to-multipoint traffic delivery on an MPLS backbone and have already deployed LDP for P2P traffic would be to rely on LDP extensions in order to setup point-to-multipoint (P2MP) LSPs and potentially multipoint-to-multipoint (MP2MP) LSPs. This would bring consistency with P2P MPLS applications and would ease the delivery of point-to-multipoint applications in an MPLS backbone.

This document lists a set of requirements for LDP extensions, for setting up P2MP LSPs and potentially MP2MP LSPs, so as to deliver P2MP traffic over a MPLS infrastructure.
It is intended that solutions that specify LDP procedures for P2MP and MP2MP LSP setup, satisfy these requirements.

Note that generic requirements for point-to-multipoint extensions to MPLS are out of the scope of this document. Rather this document describes solution specific requirements related to LDP extensions in order to set up P2MP and MP2MP LSPs.

Other mechanisms could be used for setting up P2MP and MP2MP LSPs, such as for instance PIM extensions, but these are out of the scope of this document. The objective is not to compare these mechanisms but rather to focus on the requirements for an LDP extension approach.

Section 3 points out the problem statement. Section 4 illustrates application scenarios. Finally section 5 addresses detailed requirements.

3. Problem Statement and Requirements Overview

3.1. Problem Statement

Many operators have deployed LDP [LDP] for setting up P2P and MP2P MPLS LSPs as PE-to-PE tunnels so as to carry point-to-point traffic essentially in Layer 3 and Layer 2 VPN networks.
There are emerging requirements for supporting multicast traffic delivery within these VPN infrastructures ([L3VPN-MCAST] and [L2VPN-MCAST]).
For various reasons, including consistency with P2P applications, and taking full advantages of MPLS network infrastructure, it would be highly desirable to use MPLS LSPs for the delivery of multicast traffic.
This could be implemented by setting up a group of P2P or MP2P LSPs,

but such an approach may be sub-optimal since it would result in data replication at the ingress LSR, and bandwidth inefficiency (duplicate data traffic within the network).
Hence new mechanisms are required that would allow traffic from an Ingress LSR to be efficiently delivered to a number of Egress LSRs in an MPLS backbone, avoiding duplicate copies of a packet on a given link.

Such efficient traffic delivery requires setting up P2MP LSPs. A P2MP LSPs is an LSP starting at an Ingress LSR, and ending on a set of one or more Egress LSRs. Traffic sent by the Ingress LSR is replicated on one or more Branch LSRs down to Egress LSRs.

Requirements for setting up P2MP TE LSPs have been expressed in [P2MP-TE-REQ]. RSVP-TE extensions for setting up P2MP Traffic Engineered LSPs have been defined in [P2MP-TE-RSVP]. This approach is useful, in network environments where Traffic Engineering capabilities are required.
However, for operators that deployed LDP for setting up PE-to-PE unicast MPLS LSPs, and without the need of traffic engineering, an interesting approach would be using LDP extensions for setting up P2MP LSPs.

Note that there are other alternatives for setting up P2MP (e.g. PIM extensions defined in [PIM-MPLS]), that could be useful in various situations. These are out of the scope of this document.

This document focuses on the LDP approach for setting up P2MP LSPs. The following gives a set of guidelines that a specification of LDP extensions for setting up P2MP LSPs should follow.

3.2. Requirements overview

The multi-point (MP) LDP mechanism MUST support setting up P2MP LSPs, i.e. LSPs with one Ingress LSR and one or more egress LSRs, with traffic replication at some Branch LSRs.

For traffic delivery between a group of N LSRs which are acting indifferently as Ingress or Egress LSRs, it could be preferable to setup MP2MP LSP connecting all these LSRs, instead of having N P2MP LSPs. This would significantly reduce the amount of states that must be maintained on a given LSR.
The traffic sent by any Leaf LSRs of a MP2MP LSP is delivered to all other Leaf LSRs of the MP2MP LSP.

Hence the MP LDP mechanism SHOULD also allow setting up MP2MP LSPs, connecting a group of leaf LSRs.

Note that in the following we use the term MP LSP when referring to P2MP and MP2MP LSPs.

The MP LDP mechanism MUST allow the arbitrary addition or removal of leaves associated with a MP LSP.

The MP LDP mechanism MUST interoperate seamlessly with existing P2P and MP2P LDP mechanisms.
It is of paramount importance that MP LDP mechanisms MUST NOT impede the operation of existing P2P/MP2P LSPs.

Also the MP LDP mechanism SHOULD scale independently from the number of Leaf LSRs. For example, it SHOULD NOT create an extraordinary number of LFIB entries even as the number of leaves increases.

4. Application scenarios

To be completed in next revision

5. Detailed Requirements

5.1. MP LSPs

5.1.1. P2MP LSP

The MP LDP mechanism MUST support setting up P2MP LSPs.

A P2MP LSP has one Ingress LSR and one or more Egress LSRs. Traffic sent by the Ingress LSR is received by all Egress LSRs. The specific aspects related to P2MP LSP is the action required at a Branch LSR, where data replication occurs. Incoming labelled data is appropriately replicated to several outgoing interfaces which may use different labels. Only one copy of a packet MUST be sent on a given link of a P2MP LSP.

A P2MP LSP MUST be identified by a constant and unique identifier within the whole LDP domain, whatever the number of leaves, which may vary dynamically.
This identifier will be used so as to add/remove leaves to/from the P2MP tree.

5.1.2. MP2MP LSP

The MP LDP mechanism SHOULD allow setting up MP2MP LSPs.

A MP2MP LSP is a bidirectional LSP whose Leaf LSRs act indifferently as Ingress or Egress. Traffic sent by any Leaf LSRs is received by all other Leaf LSRs. Only one copy of a packet MUST be sent on a given link of a MP2MP LSP. The specific aspect related to a MP2MP LSP is the action required at Hub LSRs, where data replication occurs. A Hub LSR has more than two interfaces on the LSP. A Hub LSR replicates labeled packets received on any interface of the LSP to all other interfaces of the LSP, which may use different labels.
A Leaf LSR of a MP2MP LSP MUST NOT receive back a packet it had previously transmitted on the MP2MP LSP.

A MP2MP LSP MUST be identified by a constant and unique identifier within the whole LDP domain, whatever the number of leaves, which may vary dynamically.
This identifier will be used so as to add and remove leaves to and from the MP2MP tree.

5.1.3. MP LSP FEC

As with P2P MPLS technology [LDP], traffic MUST be classified into a FEC in this MP extension. All packets which belong to a particular P2MP or MP2MP FEC and which travel from a particular node MUST use the same MP LSP.

As such, a solution MUST specify a FEC that is suitable for P2MP/MP2MP forwarding. Such P2MP/MP2MP FEC MUST be distinguished clearly from the exiting P2P/MP2P FEC.

5.2. Setting up, tearing down and modifying MP LSPs

The MP LDP mechanism MUST support the establishment, maintenance and teardown of MP LSPs in a scalable manner. This MUST include both the existence of a large amount of MP LSPs within a single network and a large amount of leaf LSRs for a single MP LSP.

In order to scale well with a large number of leaves it is RECOMMENDED to follow a leaf-initiated MP LSP setup approach. For that purpose, leaves will have to be aware of the MP LSP identifier. The way a Leaf LSR discovers MP LSPs identifiers SHOULD not be part of MP LDP extensions. Instead this SHOULD be part of the applications that will use MP LSPs, and it is out of the scope of this document.

The MP LDP mechanism MUST allow the dynamic addition and removal of leaves to and from a MP LSP. It is RECOMMENDED that these operations be leaf-initiated.
It is RECOMMENDED that these operations do not cause any additional processing except on the path from the Branch (or possibly Hub) LSR to the added or removed leaf LSR.

5.3. Label Advertisement

The MP LDP mechanism SHOULD support downstream unsolicited label advertisement mode. This is well suited to a leaf-initiated approach and is consistent with P2P/MP2P LDP operations.

In order to follow a leaf initiated LSP setup approach, MP LDP mechanism SHOULD support the Ordered label distribution control mode. Note that the Independent control mode is not relevant in a MP context, because the upstream LSRs cannot distribute labels independently like P2P/MP2P LDP, they must wait for label distribution from downstream LSRs.

Upstream label allocation ([MPLS-UPSTREAM]) may be particularly useful to avoid packet replication on LAN interfaces of a MP LSP, or when encapsulating the MP LSP into a P2MP TE tunnel.

Hence the MP LDP mechanism SHOULD also support upstream solicited label advertisement mode, where the solicitation is made by the downstream LSR, but the label is assigned by the upstream LSR.

Note that the existing base LDP specification [RFC3036] does not specify upstream solicited label advertisement. Hence specific extensions SHOULD be defined.

5.4. Data Duplication

Data duplication refers to the receipt of multiple copies of a packet by any leaf. Although this may be a marginal situation, it may also be detrimental for certain applications. Hence, data duplication SHOULD be avoided as much as possible, and limited to (hopefully rare) transitory conditions.

Note, in particular, that data duplication might occur if MP LSP rerouting is being performed (See also section 5.6).

5.5. Avoiding loops

The MP LDP mechanism SHOULD have a mechanism to avoid routing loops even during transient events. Furthermore, the MP LDP mechanism MUST avoid routing loops that may trigger unexpected non-localized exponential growth of traffic. Note that any loop-avoidance mechanism MUST respect scalability requirements, and particularly SHOULD scale independently from the number of Leaf LSRs.

5.6. MP LSP routing

As with P2P and MP2P LDP LSPs, the MP LDP mechanism MUST support hop-by-hop LSP routing. MP LSP LDP-based routing SHOULD rely upon the information maintained in LSR Routing Information Bases (RIB). For

instance, P2MP LSP routing could rely upon a shortest path to the Ingress LSR, and MP2MP LSP routing could rely upon a shortest path to one or more specific Hub LSRs. Note that unlike P2P/MP2P LDP routing, Equal Cost Multi Path (ECMP) MUST be avoided with MP LDP routing.

5.7. MP LSP Re-routing

The MP LDP mechanism MUST support the rerouting of a MP LSP in the following cases:

- -A better path exists (e.g. new link, netric change)
- -Network failure (link or node)
- -Planned maintenance

5.7.1. Rerouting on a Better Path

The MP LDP mechanism MUST allow for rerouting of a MP LSP in case a better path is created in the network, for instance as a result of a metric change, or the addition of links or nodes.
Traffic disruption MUST be minimized during such rerouting. It is RECOMMENDED that devices perform make-before-break for traffic on MP LSPÆs to minimize traffic disruption.
It SHOULD be feasible to avoid packet loss during such rerouting.

Unnecessary data duplication during such rerouting MUST also be minimized.

5.7.2. Rerouting due to a Network Failure

The MP LDP mechanism MUST allow for rerouting of a MP LSP in case of link or node failure(s). The rerouting time SHOULD be minimized as much as possible so as to reduce traffic disruption.

A mechanism MUST be defined to prevent constant MP LSP teardown and rebuild which may be caused by the instability of a specific link/node in the network.

5.7.3. Rerouting Due to Planned Maintenance

The MP LDP mechanism MUST support planned maintenance operations. It SHOULD be possible to reroute a MP-LSP before a link/node is deactivated for maintenance purposes. Traffic disruption MUST be minimized during such rerouting. It SHOULD be feasible to avoid packet loss during such rerouting.

Unnecessary traffic duplication during such rerouting MUST also be minimized.

5.8. Support for LAN interfaces

The MP LDP mechanism MUST provide a way for a Hub/Branch LSR to send a single copy of the data onto an Ethernet LAN interface and reach multiple adjacent downstream nodes. This requires that the same label be negotiated will all downstream LSRs for the LSP. In order to ease such negotiation an upstream label allocation approach may be used.

5.9. Support for encapsulation in P2P and P2MP TE tunnels

The MP LDP mechanism MUST support nesting MP LSPs into P2P and P2MP TE tunnels.
The MP LDP mechanism MUST provide a way for a Hub/Branch LSR of a MP LPS, which is also a Head End LSR of a P2MP TE tunnel, to send a single copy of the data onto the tunnel and reach all downstream LSRs on the MP LSP, which are also Egress LSRs of the tunnel. As with LAN interfaces, this requires that the same LDP label be negotiated with all downstream LSRs for the MP LDP LSP. In order to ease such negotiation, an upstream label allocation approach may be used.

5.10. Label spaces

Labels for MP LSPs and P2P/MP2P LSPs MAY be assigned from shared or dedicated label spaces.
MPLS Context Specific Label Spaces ([UPSTREAM-LABEL]) and particularly Upstream label spaces and Tunnel label spaces MAY be required to support upstream label allocation so as to avoid packet replication on LAN or P2MP TE Tunnel interfaces.

Note that dedicated label spaces will require the establishment of separate MP LDP sessions.

5.11. IPv4/IPv6 support

The MP LDP mechanism MUST be equally applicable to IPv4 and IPv6 traffic. Likewise, it SHOULD be possible to convey both kinds of traffic in a given MP LSP facility.

Also the MP LDP mechanism MUST support the establishment of LDP sessions over both IPv4 and IPv6 control planes.

5.12. Multi-Area LSPs

The MP LDP mechanism MUST support the establishment of multi-area MP LSPs, i.e. LSPs whose leaves do not all reside in the same IGP area. This SHOULD be possible without requiring the advertisement of Leaf LSRs' addresses across IGP areas.

5.13. OAM

LDP management tools ([LDP-MIB],...) MUST be enhanced to support MP LDP extensions. This may yield a new MIB module, which may possibly be inherited from the LDP MIB.

In order to facilitate correct management, MP LDP LSPs MUST have unique identifiers, otherwise it is impossible to determine which LSP is being managed.
OAM facilities will have special demands in MP MPLS environments especially within the context of tracing the paths and determining the connectivity of MP LSPs. Further and precise requirements and mechanisms for OAM purpose are out of the scope of this document. It is expected that a separate document will cover these requirements and mechanisms.

5.14. Graceful Restart and Fault Recovery

LDP Graceful Restart mechanisms [LDP-GR] and Fault Recovery [LDP-FT] mechanisms SHOULD be enhanced to support MP LDP LSPs.

Particularly [LDP-GR] applies only to downstream unsolicited label distribution. Hence new mechanisms are required to account for upstream label assignment, particularly in multi segment LANs.

5.15. Robustness

A solution SHOULD avoid whatever single points of failures or propose some technical solutions for a failover mechanism (e.g., redundancy/ failover of Hub LSRs).

5.16. Scalability

Scalability is a key requirement for the MP LDP mechanism. It MUST be designed to scale well with an increase in the number of any of the following:

- number of Leaf LSRs per MP LSP
- number of Branch/Hub LSRs per MP LSP
- number of MP LSPs per LSR

The size of a MP LSP state on a LSR SHOULD be independent of the number of leaves, and SHOULD only depend on the number of adjacent LSRs.

5.16.1. Orders of magnitude of the expected numbers of MP LSPs and leaves per LSP in operational networks

To be completed in next revision

5.17. Backward Compatibility

In order to allow for a smooth migration, the MP LDP mechanism SHOULD offer as much backward compatibility as possible. In particular, the solution SHOULD allow the setup of a MP LSP along non branch/hub transit LSRs that do not support MP LDP extensions.

Also, the MP LDP solution MUST interoperate seamlessly with current LDP mechanisms and inherit its capability sets from [LDP]. The MP LDP solution MUST not impede the operation of P2P/MP2P LSPs. A MP LSP solution MUST be designed in such a way that it allows P2P/MP2P and MP LSPs to be signalled on the same interface.

6. Evaluation criteria

6.1. Performances

The solution will be evaluated with respect to the following criteria:

(1) Time (in msec) to add or remove a Leaf LSR
(2) Time (in msec) to repair a MP LSP in case of link or node failure
(3) Scalability (state size, number of messages, message size).

Particularly, the MP LDP mechanism SHOULD be designed so that convergence times in case of link or node failure are minimized, in order to limit traffic disruption.

6.2. Complexity and Risks

The proposed solution SHOULD not introduce complexity to the current LDP operations to such a degree that it would affect the stability and diminish the benefits of deploying such MP LDP solution.

7. Security Considerations

This document does not introduce any new security issues beyond those inherent to LDP [LDP] and a MP LDP solution may use the same mechanisms.

8. Acknowledgment

We would like to thank Christian Jacquenet (France Telecom) and Hitoshi Fukuda (NTT Communications) for their highly useful comments and suggestions.

We would also like to thank authors of [P2MP-TE-REQ] from which some text of this document has been inspired.

9. References

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

[RFC3667] Bradner, S., "IETF Rights in Contributions", BCP 78, RFC 3667, February 2004.

[BCP79] Bradner, S., "Intellectual Property Rights in IETF Technology", RFC 3979, March 2005.

[LDP] L. Andersson et al. "LDP Specification", RFC 3036, January 2001

[L3VPN-MCAST] T. Morin, Ed., "Requirements for Multicast in L3 Provider-Provisioned VPNs", draft-ietf-l3vpn-ppvpn-mcast-reqts-01.txt, work in progress.

[L2VPN-MCAST] Y. Kamite et al. " Requirements for Multicast Support in Virtual Private LAN Services", draft-kamite-l2vpn-vpls-mcast-reqts-00.txt, work in progress

[P2MP-TE-REQ] S. Yasukawa, et. al., "Requirements for Point-to-Multipoint capability extension to MPLS", draft-ietf-mpls-p2mp-sig-requirement-03.txt, work in progress.

[P2MP-TE-RSVP] R. Aggarwal, et. al., "Extensions to RSVP-TE for Point to Multipoint TE LSPs", draft-ietf-mpls-rsvp-te-p2mp-02.txt, work in progress.

[PIM-MPLS] D. Farinacci, Y. Rekhter, E. Rosen, T. Qian, " Using PIM to Distribute MPLS Labels for Multicast Routes", draft-farinacci-mpls-multicast-03.txt.

[MPLS-UPSTREAM-LABEL] R. Aggarwal, Y. Rekhter, E. Rosen, "MPLS Upstream Label Assignment and Context Specific Label Space", draft-raggarwa-mpls-upstream-label-00.txt, work in progress.

[LDP-MIB] J. Cuchiarra et al. " Definitions of Managed Objects for the Multiprotocol Label Switching (MPLS), Label Distribution Protocol (LDP)", RFC3815, June 2004.

[LDP-GR] M. Leelanivas, Y. Rekhter, R. Aggarwal, " Graceful Restart Mechanism for Label Distribution Protocol" RFC3478, February 2003.

[LDP-FT] A. Farrel, " Fault Tolerance for the Label Distribution Protocol (LDP)", RFC3479, February 2003.

10. Authors' Addresses:

Jean-Louis Le Roux
France Telecom
2, avenue Pierre-Marzin
22307 Lannion Cedex
FRANCE
Email: jeanlouis.leroux@francetelecom.com

Thomas Morin
France Telecom
2, avenue Pierre-Marzin
22307 Lannion Cedex
FRANCE
Email: thomas.morin@francetelecom.com

Vincent Parfait
EQUANT
1041 Route des Dolines
Sophia Antipolis
06560 Valbonne
FRANCE
Email: vincent.parfait@equant.com

Luyuan Fang
AT&T
200 Laurel Avenue
Middletown, NJ 07748
USA
Email: luyuanfang@att.com

Lei Wang
Telenor
Snaroyveien 30
Fornebu 1331
NORWAY
Email: lei.wang@telenor.com

Yuji Kamite
NTT Communications Corporation
Tokyo Opera City Tower
3-20-2 Nishi Shinjuku, Shinjuku-ku,

Tokyo 163-1421,
JAPAN
Email: y.kamite@ntt.com

Shane Amante
Level 3 Communications, LLC
1025 Eldorado Blvd
Broomfield, CO 80021
USA
Email: shane@level3.net

11. Intellectual Property Statement

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