Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: November 14, 2013

T. Beckhaus
Deutsche Telekom AG
B. Decraene
France Telecom
K. Tiruveedhula
Juniper Networks
M. Konstantynowicz
L. Martini
Cisco Systems, Inc.
May 13, 2013

LDP Downstream-on-Demand in Seamless MPLS
draft-ietf-mpls-ldp-dod-07

Abstract

Seamless MPLS design enables a single IP/MPLS network to scale over core, metro and access parts of a large packet network infrastructure using standardized IP/MPLS protocols. One of the key goals of Seamless MPLS is to meet requirements specific to access, including high number of devices, their position in network topology and their compute and memory constraints that limit the amount of state access devices can hold.This can be achieved with LDP Downstream-on-Demand (LDP DoD) label advertisement. This document describes LDP DoD use cases and lists required LDP DoD procedures in the context of Seamless MPLS design.

In addition, a new optional TLV type in the LDP Label Request message is defined for fast-up convergence.

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

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Copyright Notice

Table of Contents

1. Introduction

Seamless MPLS design [I-D.ietf-mpls-seamless-mpls] enables a single IP/MPLS network to scale over core, metro and access parts of a large packet network infrastructure using standardized IP/MPLS protocols. One of the key goals of Seamless MPLS is to meet requirements specific to access, including high number of devices, their position in network topology and their compute and memory constraints that limit the amount of state access devices can hold.

In general MPLS Label Switching Routers implement either LDP or RSVP for MPLS label distribution.

The focus of this document is on LDP, as Seamless MPLS design does not include a requirement for general purpose explicit traffic engineering and bandwidth reservation. Document concentrates on the unicast connectivity only. Multicast connectivity is subject for further study.

In Seamless MPLS design [I-D.ietf-mpls-seamless-mpls], IP/MPLS protocol optimization is possible due to a relatively simple access network topologies. Examples of such topologies involving access nodes (AN) and aggregation nodes (AGN) include:

a. A single AN homed to a single AGN.

b. A single AN dual-homed to two AGNs.

c. Multiple ANs daisy-chained via a hub-AN to a single AGN.

d. Multiple ANs daisy-chained via a hub-AN to two AGNs.

e. Two ANs dual-homed to two AGNs.

f. Multiple ANs chained in a ring and dual-homed to two AGNs.

The amount of IP RIB and FIB state on ANs can be easily controlled in the listed access topologies by using simple IP routing configuration with either static routes or dedicated access IGP. Note that in all of the above topologies AGNs act as the access border routers (access ABRs) connecting the access topology to the rest of the network. Hence in many cases it is sufficient for ANs to have a default route pointing towards AGNs in order to achieve complete network connectivity from ANs to the network.

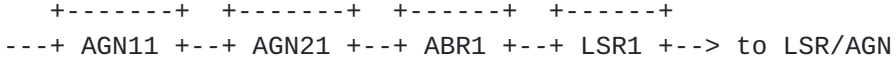
The amount of MPLS forwarding state however requires additional consideration. In general MPLS routers implement LDP Downstream Unsolicited (LDP DU) label advertisement [RFC5036] and advertise MPLS labels for all valid routes in their RIB. This is seen as an inadequate approach for ANs, which requires a small subset of the total routes (and associated labels) based on the required connectivity for the provisioned services. And although filters can be applied to those LDP DU labels advertisements, it is not seen as a suitable tool to facilitate any-to-any AN-driven connectivity between access and the rest of the MPLS network.

This document describes an access node driven "subscription model" for label distribution in the access. The approach relies on the standard LDP Downstream-on-Demand (LDP DoD) label advertisements as specified in [RFC5036]. LDP DoD enables on-demand label distribution ensuring that only required labels are requested, provided and installed. Procedures described in this document are equally applicable to LDP IPv4 and IPv6 address families. For simplicity the document provides examples based on LDP IPv4 address family.

The following sections describe a set of reference access topologies considered for LDP DoD usage and their associated IP routing configurations, followed by LDP DoD use cases and LDP DoD procedures in the context of Seamless MPLS design.

2. Reference Topologies

LDP DoD use cases are described in the context of a generic reference end-to-end network topology based on Seamless MPLS design [I-D.ietf-mpls-seamless-mpls] shown in Figure 1



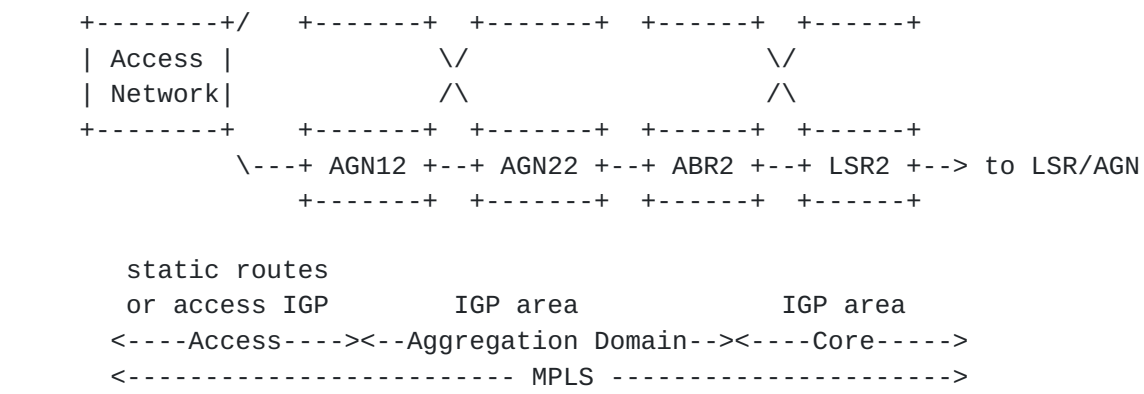


Figure 1: Seamless MPLS end-to-end reference network topology.

The access network is either single or dual homed to AGN1x, with either a single or multiple parallel links to AGN1x.

Seamless MPLS access network topologies can range from a single- or dual-homed access node to a chain or ring of access nodes, and use either static routing or access IGP (ISIS or OSPF). The following sections describe reference access topologies in more detail.

2.1. Access Topologies with Static Routing

In most cases access nodes connect to the rest of the network using very simple topologies. Here static routing is sufficient to provide the required IP connectivity. The following topologies are considered for use with static routing and LDP DoD:

a. [I1] topology - a single AN homed to a single AGN.

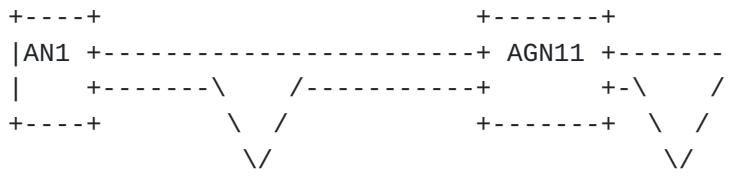
b. [I] topology - multiple ANs daisy-chained to a single AGN.

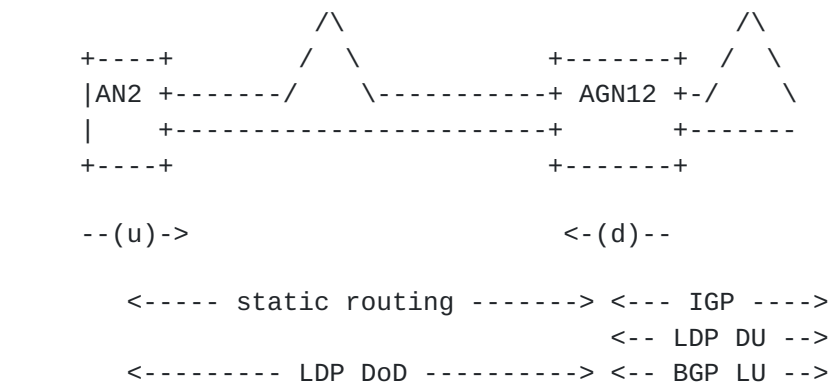
c. [V] topology - a single AN dual-homed to two AGNs.

d. [U2] topology - two ANs dual-homed to two AGNs.

e. [Y] topology - multiple ANs daisy-chained to two AGNs.

The reference static routing and LDP configuration for [V] access topology is shown in Figure 2. The same static routing and LDP configuration also applies to [I1] topology.



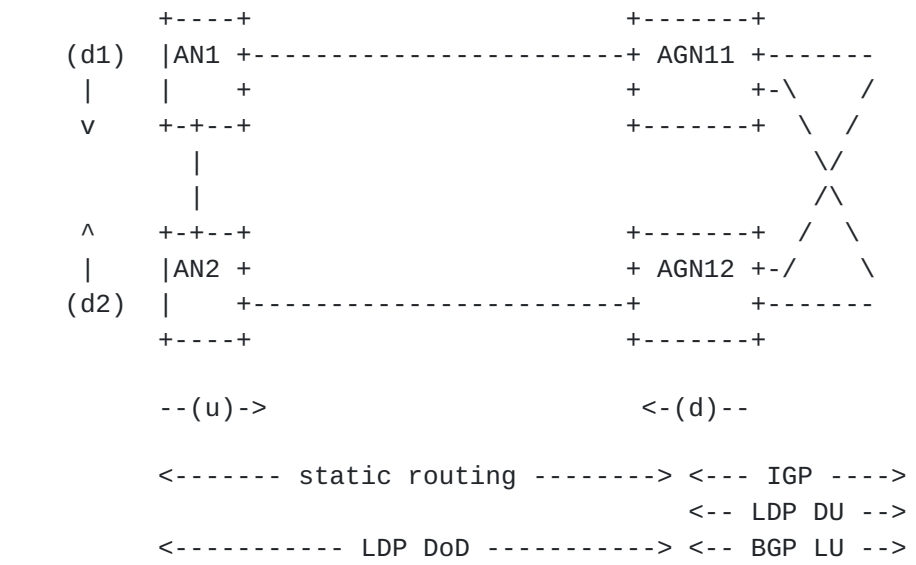


(u) static routes: 0/0 default, (optional) /32 routes
(d) static routes: AN loopbacks

Figure 2: [V] access topology with static routes.

In line with the Seamless MPLS design, static routes configured on AGN1x and pointing towards the access network are redistributed in either IGP or BGP labeled unicast (BGP-LU) [RFC3107].

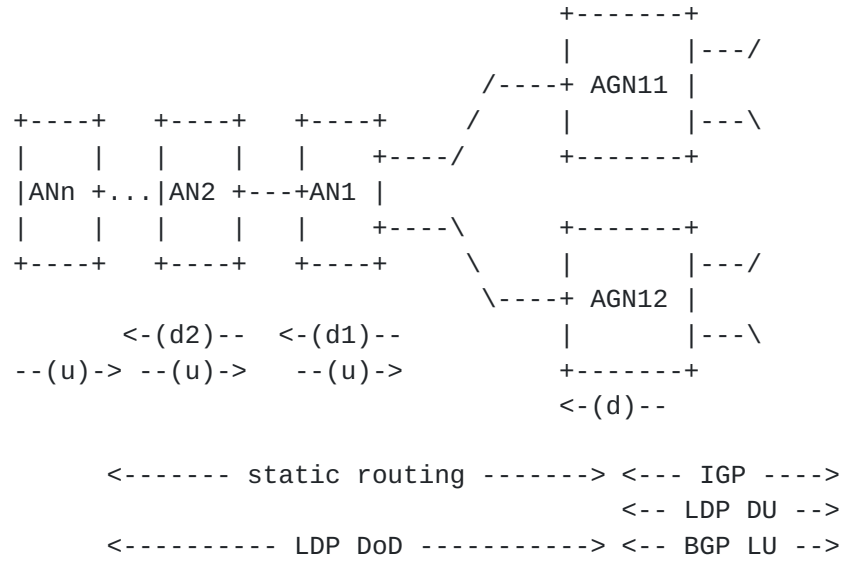
The reference static routing and LDP configuration for [U2] access topology is shown in Figure 3.



(u) static route 0/0 default, (optional) /32 routes
(d) static route for AN loopbacks
(d1) static route for AN2 loopback and 0/0 default with lower preference
(d2) static route for AN1 loopback and 0/0 default with lower preference

Figure 3: [U2] access topology with static routes.

The reference static routing and LDP configuration for [Y] access topology is shown in Figure 4. The same static routing and LDP configuration also applies to [I] topology.



(u) static routes: 0/0 default, (optional) /32 routes

(d) static routes: AN loopbacks [1..n]
(d1) static routes: AN loopbacks [2..n]
(d2) static routes: AN loopbacks [3..n]

Figure 4: [Y] access topology with static routes.

Note that in all of the above topologies parallel ECMP (or L2 LAG) links can be used between the nodes.

ANs support Inter-area LDP [RFC5283] in order to use the IP default route to match the LDP FEC advertised by AGN1x and other ANs.

2.2. Access Topologies with Access IGP

A dedicated access IGP instance is used in the access network to perform the internal routing between AGN1x and connected AN devices. Example of such IGP could be ISIS, OSPFv2&v3, RIPv2&RIPng. This access IGP instance is distinct from the IGP of the aggegation domain.

The following topologies are considered for use with access IGP routing and LDP DoD:

a. [U] topology - multiple ANs chained in an open ring and dual-homed to two AGNs.

b. [Y] topology - multiple ANs daisy-chained via a hub-AN to two AGNs.

The reference access IGP and LDP configuration for [U] access topology is shown in Figure 5.

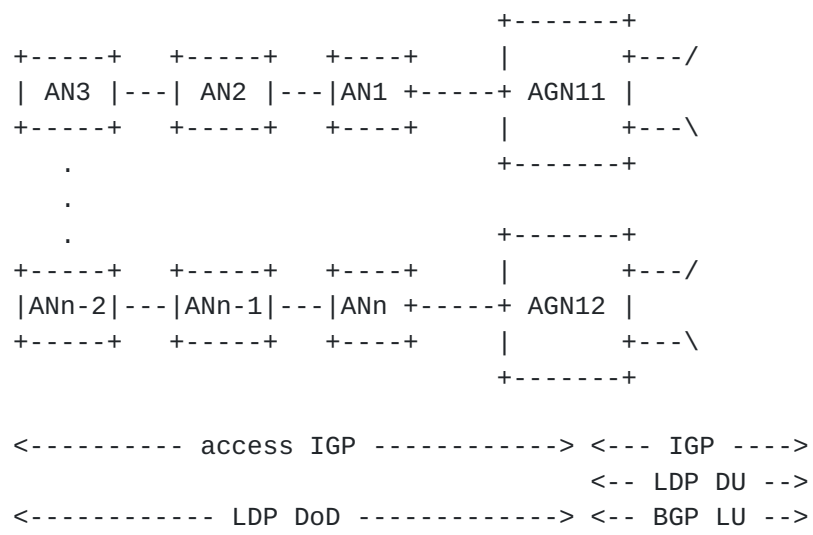


Figure 5: [U] access topology with access IGP.

The reference access IGP and LDP configuration for [Y] access topology is shown in Figure 6.

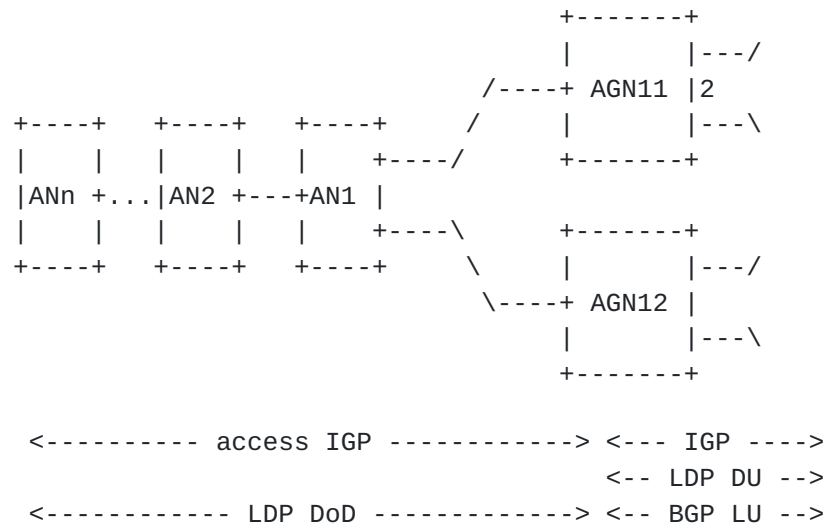


Figure 6: [Y] access topology with access IGP.

Note that in all of the above topologies parallel ECMP (or L2 LAG) links can be used between the nodes.

In both of the above topologies, ANs (ANn ... AN1) and AGN1x share the access IGP and advertise their IPv4 and IPv6 loopbacks and link addresses. AGN1x advertise a default route into the access IGP.

ANs support Inter-area LDP [RFC5283] in order to use the IP default route for matching the LDP FECs advertised by AGN1x or other ANs.

3. LDP DoD Use Cases

LDP DoD use cases described in this document are based on the Seamless MPLS scenarios listed in Seamless MPLS design [I-D.ietf-mpls-seamless-mpls]. This section illustrates these use cases focusing on services provisioned on the access nodes and clarifies expected LDP DoD operation on the AN and AGN1x devices. Two representative service types are used to illustrate the service use cases: MPLS PWE3 [RFC4447] and BGP/MPLS IPVPN [RFC4364].

Described LDP DoD operations apply equally to all reference access topologies described in Section 2. Operations that are specific to certain access topologies are called out explicitly.

References to upstream and downstream nodes are made in line with the definition of upstream and downstream LSR [RFC3031].

LDP DoD procedures follow the LDP specification [RFC5036], and are equally applicable to LDP IPv4 and IPv6 address families. For simplicity examples are provided for LDP IPv4 address family only.

3.1. Initial Network Setup

An access node is commissioned without any services provisioned on it. The AN may request labels for loopback addresses of any AN, AGN or other nodes within Seamless MPLS network for operational and management purposes. It is assumed that AGN1x has required IP/MPLS configuration for network-side connectivity in line with Seamless MPLS design [I-D.ietf-mpls-seamless-mpls].

LDP sessions are configured between adjacent ANs and AGN1x using their respective loopback addresses.

3.1.1. AN with Static Routing

If access static routing is used, ANs are provisioned with the following static IP routing entries (topology references from Section 2 are listed in square brackets):

a. [I1, V, U2] - Static default route 0/0 pointing to links connected to AGN1x. Requires support for Inter-area LDP [RFC5283].

b. [U2] - Static /32 routes pointing to the other AN. Lower preference static default route 0/0 pointing to links connected to the other AN. Requires support for Inter-area LDP [RFC5283].

c. [I, Y] - Static default route 0/0 pointing to links leading towards AGN1x. Requires support for Inter-area LDP [RFC5283].

d. [I, Y] - Static /32 routes to all ANs in the daisy-chain pointing to links towards those ANs.

e. [I1, V, U2] - Optional - Static /32 routes for specific nodes within Seamless MPLS network, pointing to links connected to AGN1x.

f. [I, Y] - Optional - Static /32 routes for specific nodes within the Seamless MPLS network, pointing to links leading towards AGN1x.

Upstream AN/AGN1x should request labels over LDP DoD session(s) from downstream AN/AGN1x for configured static routes if those static routes are configured with LDP DoD request policy and if they are pointing to a next-hop selected by routing. It is expected that all configured /32 static routes to be used for LDP DoD are configured with such policy on AN/AGN1x.

Downstream AN/AGN1x should respond to the Label Request from the upstream AN/AGN1x with a Label Mapping if requested route is present in its RIB, and there is a valid label binding from its downstream or it is the egress node. In such case downstream AN/AGN1x must install the advertised label as an incoming label in its label table (LIB) and its forwarding table (LFIB). Upstream AN/AGN1x must also install the received label as an outgoing label in their LIB and LFIB. If the downstream AN/AGN1x does have the route present in its RIB, but does not have a valid label binding from its downstream, it should forward the request to its downstream.

In order to facilitate ECMP and IPFRR LFA local-repair, the upstream AN/AGN1x must also send LDP DoD label requests to alternate next-hops per its RIB, and install received labels as alternate entries in its LIB and LFIB.

AGN1x node on the network side may use BGP labeled unicast [RFC3107] in line with the Seamless MPLS design [I-D.ietf-mpls-seamless-mpls]. In such a case AGN1x will be redistributing its static routes pointing to local ANs into BGP labeled unicast to facilitate network-to-access traffic flows. Likewise, to facilitate access-to-network traffic flows, AGN1x will be responding to access-originated LDP DoD label requests with label mappings based on its BGP labeled unicast reachability for requested FECs.

3.1.2. AN with Access IGP

If access IGP is used, AN(s) advertise their loopbacks over the access IGP with configured metrics. AGN1x advertise a default route over the access IGP.

Routers request labels over LDP DoD session(s) according to their needs for MPLS connectivity (LSPs). In particular if AGNs, as per Seamless MPLS design [I-D.ietf-mpls-seamless-mpls], redistribute routes from the IGP into BGP labeled unicast [RFC3107], they should request labels over LDP DoD session(s) for those routes.

Identically to the static route case, downstream AN/AGN1x should respond to the Label Request from the upstream AN/AGN1x with a Label Mapping (if the requested route is present in its RIB, and there is a valid label binding from its downstream), and must install the advertised label as an incoming label in its LIB and LFIB. Upstream AN/AGN1x must also install the received label as an outgoing label in their LIB and LFIB.

Identically to the static route case, in order to facilitate ECMP and IPFRR LFA local-repair, upstream AN/AGN1x must also send LDP DoD label requests to alternate next-hops per its RIB, and install received labels as alternate entries in its LIB and LFIB.

AGN1x node on the network side may use BGP labeled unicast [RFC3107] in line with Seamless MPLS design [I-D.ietf-mpls-seamless-mpls]. In such case AGN1x will be redistributing routes received over the access IGP (and pointing to local ANs), into BGP labeled unicast to facilitate network-to-access traffic flows. Likewise, to facilitate access-to-network traffic flows AGN1x will be responding to access originated LDP DoD label requests with label mappings based on its BGP labeled unicast reachability for requested FECs.

3.2. Service Provisioning and Activation

Following the initial setup phase described in Section 3.1, a specific access node, referred to as AN*, is provisioned with a network service. AN* relies on LDP DoD to request the required MPLS LSP(s) label(s) from downstream AN/AGN1x node(s). Note that LDP DoD operations are service agnostic, that is, they are the same independently of the services provisioned on the AN*.

For illustration purposes two service types are described: MPLS PWE3 [RFC4447] service and BGP/MPLS IPVPN [RFC4364].

MPLS PWE3 service - for description simplicity it is assumed that a single segment pseudowire is signaled using targeted LDP FEC128 (0x80), and it is provisioned with the pseudowire ID and the loopback IPv4 address of the destination node. The following IP/MPLS

operations need to be completed on the AN* to successfully establish such PWE3 service:

a. LSP labels for destination /32 FEC (outgoing label) and the local /32 loopback (incoming label) need to be signaled using LDP DoD.

b. Targeted LDP session over an associated TCP/IP connection needs to be established to the PWE3 destination PE. This is triggered by either an explicit targeted LDP session configuration on the AN* or automatically at the time of provisioning the PWE3 instance.

c. Local and remote PWE3 labels for specific FEC128 PW ID need to be signaled using targeted LDP and PWE3 signaling procedures [RFC4447].

d. Upon successful completion of the above operations, AN* programs its RIB/LIB and LFIB tables, and activates the MPLS PWE3 service.

Note - only minimum operations applicable to service connectivity have been listed. Other non IP/MPLS connectivity operations that may be required for successful service provisioning and activation are out of scope in this document.

BGP/MPLS IPVPN service - for description simplicity it is assumed that AN* is provisioned with a unicast IPv4 IPVPN service (VPNv4 for short) [RFC4364]. The following IP/MPLS operations need to be completed on the AN* to successfully establish VPNv4 service:

a. BGP peering sessions with associated TCP/IP connections need to be established with the remote destination VPNv4 PEs or Route Reflectors.

b. Based on configured BGP policies, VPNv4 BGP NLRIs need to be exchanged between AN* and its BGP peers.

c. Based on configured BGP policies, VPNv4 routes need to be installed in the AN* VRF RIB and FIB, with corresponding BGP next-hops.

d. LSP labels for destination BGP next-hop /32 FEC (outgoing label) and the local /32 loopback (incoming label) need to be signaled using LDP DoD.

e. Upon successful completion of above operations, AN* programs its RIB/LIB and LFIB tables, and activates the BGP/MPLS IPVPN service.

Note - only minimum operations applicable to service connectivity have been listed. Other non IP/MPLS connectivity operations that may be required for successful service provisioning are out of scope in this document.

To establish an LSP for destination /32 FEC for any of the above services, AN* looks up its local routing table for a matching route, selects the best next-hop(s) and associated outgoing link(s).

If a label for this /32 FEC is not already installed based on the configured static route with LDP DoD request policy or access IGP RIB entry, AN* must send an LDP DoD Label Mapping request. Downstream AN /AGN1x LSR(s) checks its RIB for presence of the requested /32 and associated valid outgoing label binding, and if both are present, replies with its label for this FEC and installs this label as incoming in its LIB and LFIB. Upon receiving the Label Mapping the AN* must accept this label based on the exact route match of advertised FEC and route entry in its RIB or based on the longest match in line with Inter-area LDP [RFC5283]. If the AN* accepts the label it must install it as an outgoing label in its LIB and LFIB.

In access topologies [V] and [Y], if AN* is dual homed to two AGN1x and routing entries for these AGN1x are configured as equal cost paths, AN* must send LDP DoD label requests to both AGN1x devices and install all received labels in its LIB and LFIB.

In order for AN* to implement IPFRR LFA local-repair, AN* must also send LDP DoD label requests to alternate next-hops per its RIB, and install received labels as alternate entries in its LIB and LFIB.

When forwarding PWE3 or VPNv4 packets AN* chooses the LSP label based on the locally configured static /32 or default route, or default route signaled via access IGP. If a route is reachable via multiple interfaces to AGN1x nodes and the route has multiple equal cost paths, AN* must implement Equal Cost Multi-Path (ECMP) functionality. This involves AN* using hash-based load-balancing mechanism and sending the PWE3 or VPNv4 packets in a flow-aware manner with appropriate LSP labels via all equal cost links.

ECMP mechanism is applicable in an equal manner to parallel links between two network elements and multiple paths towards the destination. The traffic demand is distributed over the available paths.

AGN1x node on the network side may use BGP labeled unicast [RFC3107] in line with Seamless MPLS design [I-D.ietf-mpls-seamless-mpls]. In such case AGN1x will be redistributing its static routes (or routes received from the access IGP) pointing to local ANs into BGP labeled

unicast to facilitate network-to-access traffic flows. Likewise, to facilitate access-to-network traffic flows AGN1x will be responding to access originated LDP DoD label requests with label mappings based on its BGP labeled unicast reachability for requested FECs.

3.3. Service Changes and Decommissioning

Whenever AN* service gets decommissioned or changed and connectivity to specific destination is not longer required, the associated MPLS LSP label resources should be released on AN*.

MPLS PWE3 service - if the PWE3 service gets decommissioned and it is the last PWE3 to a specific destination node, the targeted LDP session is not longer needed and should be terminated (automatically or by configuration). The MPLS LSP(s) to that destination is no longer needed either.

BGP/MPLS IPVPN service - deletion of a specific VPNv4 (VRF) instance, local or remote re-configuration may result in specific BGP next-hop(s) being no longer needed. The MPLS LSP(s) to that destination is no longer needed either.

In all of the above cases the following LDP DoD related operations apply:

- If the /32 FEC label for the aforementioned destination node was originally requested based on either tLDP session configuration and default route or required BGP next-hop and default route, AN* should delete the label from its LIB and LFIB, and release it from downstream AN/AGN1x by using LDP DoD procedures.

- If the /32 FEC label was originally requested based on the static /32 route configuration with LDP DoD request policy, the label must be retained by AN*.

3.4. Service Failure

A service instance may stop being operational due to a local or remote service failure event.

In general, unless the service failure event modifies required MPLS connectivity, there should be no impact on the LDP DoD operation.

If the service failure event does modify the required MPLS connectivity, LDP DoD operations apply as described in Section 3.2 and Section 3.3.

3.5. Network Transport Failure

A number of different network events can impact services on AN*. The following sections describe network event types that impact LDP DoD operation on AN and AGN1x nodes.

3.5.1. General Notes

If service on any of the ANs is affected by any network failure and there is no network redundancy, the service must go into a failure state. When the network failure is recovered from, the service must be re-established automatically.

The following additional LDP-related functions should be supported to comply with Seamless MPLS [I-D.ietf-mpls-seamless-mpls] fast service restoration requirements as follows:

a. Local-repair - AN and AGN1x should support local-repair for adjacent link or node failure for access-to-network, network-to-access and access-to-access traffic flows. Local-repair should be implemented by using either IPFRR LDP LFA, simple ECMP or primary/backup switchover upon failure detection.

b. LDP session protection - LDP sessions should be configured with LDP session protection to avoid delay upon the recovery from link failure. LDP session protection ensures that FEC label binding is maintained in the control plane as long as LDP session stays up.

c. IGP-LDP synchronization - If access IGP is used, LDP sessions between ANs, and between ANs and AGN1x, should be configured with IGP-LDP synchronization to avoid unnecessary traffic loss in case the access IGP converged before LDP and there is no LDP label binding to the downstream best next-hop.

3.5.2. AN Node Failure

AN node fails and all links to adjacent nodes go down.

Adjacent AN/AGN1x nodes remove all routes pointing to the failed link(s) from their RIB tables (including /32 loopback belonging to the failed AN and any other routes reachable via the failed AN). This in turn triggers the removal of associated outgoing /32 FEC labels from their LIB and LFIB tables.

If access IGP is used, the AN node failure will be propagated via IGP link updates across the access topology.

If a specific /32 FEC(s) is not reachable anymore from those AN/AGN1x, they must also send LDP Label Withdraw to their upstream LSRs to notify about the failure, and remove the associated incoming label(s) from their LIB and LFIB tables. Upstream LSRs upon receiving Label Withdraw should remove the signaled labels from their LIB/LFIB tables, and propagate LDP Label Withdraw across their upstream LDP DoD sessions.

In [U] topology there may be an alternative path to routes previously reachable via the failed AN node. In this case adjacent AN/AGN1x should invoke local-repair (IPFRR LFA, ECMP) and switchover to alternate next-hop to reach those routes.

AGN1x gets notified about the AN failure via either access IGP (if used) and/or cascaded LDP DoD label withdraw(s). AGN1x must implement all relevant global-repair IP/MPLS procedures to propagate the AN failure towards the core network. This should involve removing associated routes (in access IGP case) and labels from its LIB and LFIB tables, and propagating the failure on the network side using BGP-LU and/or core IGP/LDP-DU procedures.

Upon AN coming back up, adjacent AN/AGN1x nodes automatically add routes pointing to recovered links based on the configured static routes or access IGP adjacency and link state updates. This should be then followed by LDP DoD label signaling and subsequent binding and installation of labels in LIB and LFIB tables.

3.5.3. AN/AGN Link Failure

Depending on the access topology and the failed link location different cases apply to the network operation after AN link failure (topology references from Section 2 in square brackets):

a. [all] - link failed, but at least one ECMP parallel link remains - nodes on both sides of the failed link must stop using the failed link immediately (local-repair), and keep using the remaining ECMP parallel links.

b. [I1, I, Y] - link failed, and there are no ECMP or alternative links and paths - nodes on both sides of the failed link must remove routes pointing to the failed link immediately from the RIB, remove associated labels from their LIB and LFIB tabels, and must send LDP label withdraw(s) to their upstream LSRs.

c. [U2, U, V, Y] - link failed, but at least one ECMP or alternate path remains - AN/AGN1x node must stop using the failed link and immediately switchover (local-repair) to the remaining ECMP path or alternate path. AN/AGN1x must remove affected next-hops and

labels from its tables and invoke LDP Label Withdraw as per point (a) above. If there is an AGN1x node terminating the failed link, it must remove routes pointing to the failed link immediately from the RIB, remove associated labels from their LIB and LFIB tabels, and must propagate the failure on the network side using BGP-LU and/or core IGP procedures.

If access IGP is used AN/AGN1x link failure will be propagated via IGP link updates across the access topology.

LDP DoD will also propagate the link failure by sending label withdraws to upstream AN/AGN1x nodes, and Label Release messages downstream AN/AGN1x nodes.

3.5.4. AGN Node Failure

AGN1x fails and all links to adjacent access nodes go down.

Depending on the access topology, following cases apply to the network operation after AGN1x node failure (topology references from Section 2 in square brackets):

a. [I1, I] - ANs are isolated from the network - AN adjacent to the failure must remove routes pointing to the failed AGN1x node immediately from the RIB, remove associated labels from their LIB and LFIB tabels, and must send LDP label withdraw(s) to their upstream LSRs. If access IGP is used, an IGP link update should be sent.

b. [U2, U, V, Y] - at least one ECMP or alternate path remains - AN adjacent to failed AGN1x must stop using the failed link and immediately switchover (local-repair) to the remaining ECMP path or alternate path. AN must remove affected routes and labels from its tables and invoke LDP Label Withdraw as per point (a) above.

Network side procedures for handling AGN1x node failure have been described in Seamless MPLS [I-D.ietf-mpls-seamless-mpls].

3.5.5. AGN Network-side Reachability Failure

AGN1x loses network reachability to a specific destination or set of network-side destinations.

In such event AGN1x must send LDP Label Withdraw messages to its upstream ANs, withdrawing labels for all affected /32 FECs. Upon receiving those messages ANs must remove those labels from their LIB and LFIB tables, and use alternative LSPs instead if available as

part of global-repair. In turn ANs should also sent Label Withdraw messages for affected /32 FECs to their upstream ANs.

If access IGP is used, and AGN1x gets completely isolated from the core network, it should stop advertising the default route 0/0 into the access IGP.

4. LDP DoD Procedures

Label Distribution Protocol is specified in [RFC5036], and all LDP Downstream-on-Demand implementations follow [RFC5036] specification. This section does not update [RFC5036] procedures, but illustrates LDP DoD operations in the context of use cases identified in Section 3 in this document, for information only.

In the MPLS architecture [RFC3031], network traffic flows from upstream to downstream LSR. The use cases in this document rely on the downstream assignment of labels, where labels are assigned by the downstream LSR and signaled to the upstream LSR as shown in Figure 7.

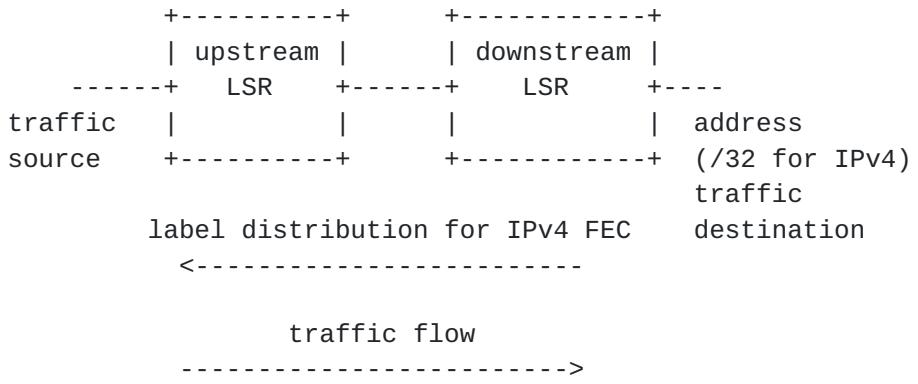


Figure 7: LDP label assignment direction

4.1. LDP Label Distribution Control and Retention Modes

LDP protocol specification [RFC5036] defines two modes for label distribution control, following the definitions in MPLS architecture [RFC3031]:

- Independent mode - an LSR recognizes a particular FEC and makes a decision to bind a label to the FEC independently from distributing that label binding to its label distribution peers. A new FEC is recognized whenever a new route becomes valid on the LSR.

- Ordered mode - an LSR needs to bind a label to a particular FEC if it knows how to forward packets for that FEC (i.e. it has a

route corresponding to that FEC) and if it has already received at least one Label Request message from an upstream LSR.

Using independent label distribution control with LDP DoD and access static routing would prevent the access LSRs from propagating label binding failure along the access topology, making it impossible for upstream LSR to be notified about the downstream failure and for an application using the LSP to switchover to an alternate path, even if such a path exists.

LDP protocol specification [RFC5036] defines two modes for label retention, following the definitions in MPLS architecture [RFC3031]:

- Conservative mode - If operating in Downstream on Demand mode, an LSR will request label mappings only from the next hop LSR according to routing. The main advantage of the conservative mode is that only the labels that are required for the forwarding of data are allocated and maintained. This is particularly important in LSRs where the label space is inherently limited, such as in an ATM switch. A disadvantage of the conservative mode is that if routing changes the next hop for a given destination, a new label must be obtained from the new next hop before labeled packets can be forwarded.

- Liberal mode - When operating in Downstream on Demand mode with Liberal Label retention, an LSR might choose to request label mappings for all known prefixes from all peer LSRs. The main advantage of the Liberal Label retention mode is that reaction to routing changes can be quick because labels already exist. The main disadvantage of the liberal mode is that unneeded label mappings are distributed and maintained.

Note that the conservative label retention mode would prevent LSRs from requesting and maintaining label mappings for any backup routes that are not used for forwarding. This in turn would prevent the access LSRs (AN and AGN1x nodes) from implementing any local protection schemes that rely on using alternate next-hops in case of the primary next-hop failure. Such schemes include IPFRR LFA if access IGP is used, or a primary and backup static route configuration. Using LDP DoD in combination with liberal retention mode allows the LSR to request labels for the specific FEC from primary next-hop LSR(s) and the alternate next-hop LSR(s) for this FEC.

Note that even though LDP DoD operates in a liberal retention mode, if used with access IGP and if no LFA exists, the LDP DoD will introduce additional delay in traffic restoration as the labels for the new next-hop will get requested only after the access IGP convergence.

Adhering to the overall design goals of Seamless MPLS [I-D.ietf-mpls-seamless-mpls], specifically achieving a large network scale without compromising fast service restoration, all access LSRs (AN and AGN1x nodes) use LDP DoD advertisement mode with:

- Ordered label distribution control - enables propagation of label binding failure within the access topology.

- Liberal label retention - enables pre-programming of alternate next-hops with associated FEC labels.

In Seamless MPLS [I-D.ietf-mpls-seamless-mpls] AGN1x node acts as an access ABR connecting access and metro domains. To enable failure propagation between those domains, access ABR implements ordered label distribution control when redistributing routes/FEC between the access-side (using LDP DoD and static or access IGP) and the network-side (using BGP labeled unicast [RFC3107] or core IGP with LDP Downstream Unsolicited label advertisement.

4.2. LDP DoD Session Negotiation

Access LSR/ABR should propose the Downstream-on-Demand label advertisement by setting "A" value to 1 in the Common Session Parameters TLV of the Initialization message. The rules for negotiating the label advertisement mode are specified in LDP protocol specification [RFC5036].

To establish a Downstream-on-Demand session between the two access LSR/ABRs, both should propose the Downstream-on-Demand label advertisement mode in the Initialization message. If the access LSR only supports LDP DoD and the access ABR proposes Downstream Unsolicited mode, the access LSR should send a Notification message with status "Session Rejected/Parameters Advertisement Mode" and then close the LDP session as specified in LDP protocol specification [RFC5036].

If an access LSR is acting in an active role, it should re-attempt the LDP session immediately. If the access LSR receives the same Downstream Unsolicited mode again, it should follow the exponential backoff algorithm as defined in the LDP protocol specification [RFC5036] with delay of 15 seconds and subsequent delays growing to a maximum delay of 2 minutes.

In case a PWE3 service is required between the adjacent access LSR/ABR, and LDP DoD has been negotiated for IPv4 and IPv6 FECs, the same LDP session should be used for PWE3 FECs. Even if LDP DoD label advertisement has been negotiated for IPv4 and IPv6 LDP FECs as described earlier, LDP session should use Downstream Unsolicited label advertisement for PWE3 FECs as specified in PWE3 LDP [RFC4447].

4.3. Label Request Procedures

4.3.1. Access LSR/ABR Label Request

Upstream access LSR/ABR will request label bindings from adjacent downstream access LSR/ABR based on the following trigger events:

a. Access LSR/ABR is configured with /32 static route with LDP DoD Label Request policy in line with initial network setup use case described in Section 3.1.

b. Access LSR/ABR is configured with a service in line with service use cases described in Section 3.2 and Section 3.3.

c. Configuration with access static routes - Access LSR/ABR link to adjacent node comes up and LDP DoD session is established. In this case access LSR should send Label Request messages for all /32 static routes configured with LDP DoD policy and all /32 routes related to provisioned services that are covered by default route.

d. Configuration with access IGP - Access LSR/ABR link to adjacent node comes up and LDP DoD session is established. In this case access LSR should send Label Request messages for all /32 routes learned over the access IGP and all /32 routes related to provisioned services that are covered by access IGP routes.

e. In all above cases requests must be sent to next-hop LSR(s) and alternate LSR(s).

Downstream access LSR/ABR will respond with Label Mapping message with a non-null label if any of the below conditions are met:

a. Downstream access LSR/ABR - requested FEC is an IGP or static route and there is an LDP label already learnt from the next-next-hop downstream LSR (by LDP DoD or LDP DU). If there is no label for the requested FEC and there is an LDP DoD session to the next-next-hop downstream LSR, downstream LSR must send a Label Request message for the same FEC to the next-next-hop downstream LSR. In such case downstream LSR will respond back to the requesting upstream access LSR only after getting a label from the next-next-hop downstream LSR peer.

b. Downstream access ABR only - requested FEC is a BGP labelled unicast route [RFC3107] and this BGP route is the best selected for this FEC.

Downstream access LSR/ABR may respond with a Label Mapping with explicit-null or implicit-null label if it is acting as an egress for the requested FEC, or it may respond with "No Route" notification if no route exists.

4.3.2. Label Request Retry

Following LDP specification LDP specification [RFC5036], if an access LSR/ABR receives a "No route" Notification in response to its Label Request message, it should retry using an exponential backoff algorithm similar to the backoff algoritm mentioned in the LDP session negotiation described in Section 4.2.

If there is no response to the sent Label Request message, the LDP specification [RFC5036] (section A.1.1, page# 100) states that the LSR should not send another request for the same label to the peer and mandates that a duplicate Label Request is considered a protocol error and should be dropped by the receiving LSR by sending a Notification message.

Thus, if there is no response from the downstream peer, the access LSR/ABR should not send a duplicate Label Request message again.

If the static route corresponding to the FEC gets deleted or if the DoD request policy is modified to reject the FEC before receiving the Label Mapping message, then the access LSR/ABR should send a Label Abort message to the downstream LSR.

To address the case of slower convergence resulting from described LDP behavior in line with LDP specification [RFC5036], a new LDP TLV extension is proposed and described in Section 5.

4.4. Label Withdraw

If an MPLS label on the downstream access LSR/ABR is no longer valid, the downstream access LSR/ABR withdraws this FEC/label binding from the upstream access LSR/ABR with the Label Withdraw Message [RFC5036] with a specified label TLV or with an empty label TLV.

Downstream access LSR/ABR should withdraw a label for specific FEC in the following cases:

a. If LDP DoD ingress label is associated with an outgoing label assigned by BGP labelled unicast route, and this route is withdrawn.

b. If LDP DoD ingress label is associated with an outgoing label assigned by LDP (DoD or DU) and the IGP route is withdrawn from the RIB or downstream LDP session is lost.

c. If LDP DoD ingress label is associated with an outgoing label assigned by LDP (DoD or DU) and the outgoing label is withdrawn by the downstream LSR.

d. If LDP DoD ingress label is associated with an outgoing label assigned by LDP (DoD or DU), route next-hop changed and

   * there is no LDP session to the new next-hop. To minimize probability of this, the access LSR/ABR should implement LDP-IGP synchronization procedures as specified in [RFC5443].

   * there is an LDP session but no label from downstream LSR. See note below.

e. If access LSR/ABR is configured with a policy to reject exporting label mappings to upstream LSR.

The upstream access LSR/ABR responds to the Label Withdraw Message with the Label Release Message [RFC5036].

After sending Label Release message to downstream access LSR/ABR, the upstream access LSR/ABR should resend Label Request message, assuming upstream access LSR/ABR still requires the label.

Downstream access LSR/ABR should withdraw a label if the local route configuration (e.g. /32 loopback) is deleted.

Note: For any events inducing next hop change, downstream access LSR/ABR should attempt to converge the LSP locally before withdrawing the label from an upstream access LSR/ABR. For example if the next-hop changes for a particular FEC and if the new next-hop allocates labels by LDP DoD session, then the downstream access LSR/ABR must send a

Label Request on the new next-hop session. If downstream access LSR/ABR doesn't get Label Mapping for some duration, then and only then downstream access LSR/ABR must withdraw the upstream label.

4.5. Label Release

If an access LSR/ABR does not need any longer a label for a FEC, it sends a Label Release Message [RFC5036] to the downstream access LSR/ABR with or without the label TLV.

If upstream access LSR/ABR receives an unsolicited Label Mapping on DoD session, they should release the label by sending Label Release message.

Access LSR/ABR should send a Label Release message to the downstream LSR in the following cases:

a. If it receives a Label Withdraw from the downstream access LSR/ABR.

b. If the /32 static route with LDP DoD Label Request policy is deleted.

c. If the service gets decommissioned and there is no corresponding /32 static route with LDP DoD Label Request policy configured.

d. If the route next-hop changed, and the label does not point to the best or alternate next-hop.

e. If it receives a Label Withdraw from a downstream DoD session.

4.6. Local Repair

To support local-repair with ECMP and IPFRR LFA, access LSR/ABR must request labels on both the best next-hop and the alternate next-hop LDP DoD sessions, as specified in the Label Request procedures in Section 4.3. If remote LFA is enabled, access LSR/ABR needs a label from its alternate next-hop toward the PQ node and needs a label from the remote PQ node toward its FEC/destination. If access LSR/ABR doesn't already know those labels, it must request them.

This will enable access LSR/ABR to pre-program the alternate forwarding path with the alternate label(s), and invoke IPFRR LFA switch-over procedure if the primary next-hop link fails.

5. LDP Extension for LDP DoD Fast-Up Convergence

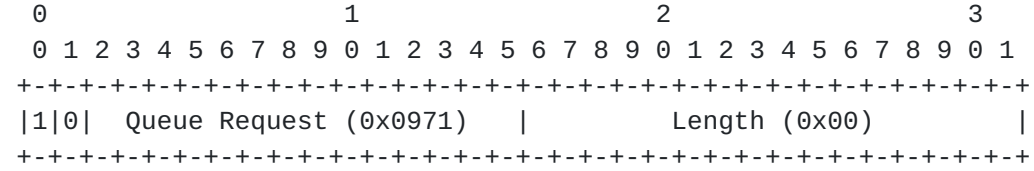
In some conditions, the exponential backoff algorithm usage described in Section 4.3.2 may result in a longer than desired wait time to get a successful LDP label to route mapping. An example is when a specific route is unavailable on the downstream LSR when the Label Mapping request from the upstream is received, but later comes back. In such case using the exponential backoff algorithm may result in a max delay wait time before the upstream LSR sends another LDP Label Request.

This section describes an extension to the LDP DoD procedure to address fast-up convergence, and as such should be treated as a normative reference. The downstream and upstream LSRs SHOULD implement this extension if the improvement in up convergence is desired.

The extension consists of the upstream LSR indicating to the downstream LSR that the Label Request SHOULD be queued on the downstream LSR until the requested route is available.

To implement this behavior, a new Optional Parameter is defined for use in the Label Request message:

Optional Parameter	Length	Value
Queue Request TLV	0	see below



U-bit = 1
Unknown TLV bit. Upon receipt of an unknown TLV, due to U-bit being set (=1), the unknown TLV MUST be silently ignored and the rest of the message processed as if the unknown TLV did not exist. In case requested route is not available, the downstream LSR MUST ignore this unknown TLV and send a "no route" notification back. Ensures backward compatibility.

F-bit = 0
Forward unknown TLV bit. This bit applies only when the U-bit is set and the LDP message containing the unknown TLV is to be forwarded. Due to F-bit being clear (=0), the unknown TLV is not forwarded with the containing message.

Type
Queue Request Type value to be allocated by IANA.

Length = 0x00
Specifies the length of the Value field in octets.

Specified operation is as follows.

To benefit from the fast-up convergence improvement, the upstream LSR sends a Label Request message with a Queue Request TLV.

If the downstream LSR supports the Queue Request TLV, it verifies if route is available and if so it replies with Label Mapping as per existing LDP procedures. If the route is not available, the downstream LSR queues the request and replies as soon as the route becomes available. In the meantime, it does not send a "no route" notification back. When sending a Label Request with the Queue Request TLV, the upstream LSR does not retry the Label Request message if it does not receive a reply from its downstream peer

If the upstream LSR wants to abort an outstanding Label Request while the Label Request is queued in the downstream LSR, the upstream LSR sends a Label Abort Request message, making the downstream LSR to remove the original request from the queue and send back a notification Label Request Aborted [RFC5036].

If the downstream LSR does not support the Queue Request TLV, and requested route is not available, it ignores this unknown TLV and sends a "no route" notification back in line with [RFC5036]. In this case the upstream LSR invokes the exponential backoff algorithm described in Section 4.3.2 following standard LDP specification LDP specification [RFC5036].

This described procedure ensures backward compatitibility.

6. IANA Considerations

6.1. LDP TLV TYPE

This document uses a new a new Optional Parameter Queue Request TLV in the Label Request message defined in Section 5. IANA already maintains a registry of name LDP "TLV TYPE NAME SPACE" defined by RFC5036. The following value is suggested for assignment:

TLV type	Description
0x0971	Queue Request TLV

7. Security Considerations

MPLS LDP Downstream on Demand deployment in the access network is subject to similar security threats as any MPLS LDP deployment. It is recommended that baseline security measures are considered as described in Security Framework for MPLS and GMPLS networks [RFC5920] and the LDP specification [RFC5036] including ensuring authenticity and integrity of LDP messages, as well as protection against spoofing and Denial of Service attacks.

Some deployments may require increased measures of network security if a subset of Access Nodes are placed in locations with lower levels of physical security e.g. street cabinets (common practice for VDSL access). In such cases it is the responsibility of the system designer to take into account the physical security measures (environmental design, mechanical or electronic access control, intrusion detection), as well as monitoring and auditing measures (configuration and Operating System changes, reloads, routes advertisements).

But even with all this in mind, the designer still should consider network security risks and adequate measures arising from the lower level of physical security of those locations.

7.1. Security and LDP DoD

7.1.1. Access to network packet flow direction

An important property of MPLS LDP Downstream on Demand operation is that the upstream LSR (requesting LSR) accepts only mappings it sent a request for (in other words the ones it is interested in), and does not accept any unsolicited label mappings by design.

This limits the potential of an unauthorized third party fiddling with label mappings operations on the wire. It also enables ABR LSR to monitor behaviour of any Access LSR in case the latter gets compromised and attempts to get access to an unauthorized FEC or remote LSR. Note that ABR LSR is effectively acting as a gateway to the MPLS network, and any Label Mapping requests made by any Access LSR are processed and can be monitored on this ABR LSR.

7.1.2. Network to access packet flow direction

Another important property of MPLS LDP DoD operation in the access is that the number of access nodes and associated MPLS FECs per ABR LSR is not large in number, and they are all known at the deployment time. Hence any changes of the access MPLS FECs can be easily controlled and monitored on the ABR LSR.

And then, even in the event when Access LSR manages to advertise a FEC that belongs to another LSR (e.g. in order to 'steal' third party data flows, or breach a privacy of VPN), such Access LSR will have to influence the routing decision for affected FEC on the ABR LSR. Following measures should be considered to prevent such event from occurring:

a. ABR LSR - access side with static routes - this is not possible for Access LSR. Access LSR has no way to influence ABR LSR routing decisions due to static nature of routing configuration here.

b. ABR LSR - access side with IGP - this is still not possible if the compromised Access LSR is a leaf in the access topology (leaf node in topologies I1, I, V, Y described earlier in this document), due to the leaf metrics being configured on the ABR LSR. If the compromised Access LSR is a transit LSR in the access topology (transit node in topologies I, Y, U), it is possible for this Access LSR to attract to itself traffic destined to the nodes upstream from it. However elaborate such 'man in the middle attack' is possible, but can be quickly detected by upstream Access LSRs not receiving traffic, and legitimate traffic from them getting dropped.

c. ABR LSR - network side - designer should consider giving a higher administrative preference to the labeled unicast BGP routes vs. access IGP routes.

In summary MPLS in access design with LDP DoD has number of native properties that prevent number of security attacks and make their detection quick and straightforward.

Following two sections describe other security considerations applicable to general MPLS deployments in the access.

7.2. Data Plane Security

Data plane security risks applicable to the access MPLS network are listed below (a non-exhaustive list):

a. packets from a specific access node flow to an altered transport layer or service layer destination.

b. packets belonging to undefined services flow to and from the access network.

c. unlabelled packets destined to remote network nodes.

Following mechanisms should be considered to address listed data plane security risks:

1. addressing (a) - Access and ABR LSRs should NOT accept labeled packets over a particular data link, unless from the Access or ABR LSR perspective this data link is known to attach to a trusted system based on employed authentication mechanism(s), and the top label has been distributed to the upstream neighbour by the receiving Access or ABR LSR.

2. addressing (a) - ABR LSR MAY restrict network reachability for access devices to a subset of remote network LSR, based on authentication or other network security technologies employed towards Access LSRs. Restricted reachability can be enforced on the ABR LSR using local routing policies, and can be distributed towards the core MPLS network using routing policies associated with access MPLS FECs.

3. addressing (b) - labeled service routes (e.g. MPLS/VPN, tLDP) are not accepted from unreliable routing peers. Detection of unreliable routing peers is achieved by engaging routing protocol detection and alarm mechanisms, and is out of scope of this document.

4. addressing (a) and (b) - no successful attacks have been mounted on the control plane and has been detected.

5. addressing (c) - ABR LSR MAY restrict IP network reachability to and from the access LSR.

7.3. Control Plane Security

Similarly to Inter-AS MPLS/VPN deployments [RFC4364], the control plane security is prerequisite to the data plane security.

To ensure control plane security access LDP DoD sessions should only be established with LDP peers that are considered trusted from the local LSR perspective, meaning they are reachable over a data link that is known to attach to a trusted system based on employed authentication mechanism(s) on the local LSR.

The security of LDP sesions is analyzed in [I-D.ietf-karp-routing-tcp-analysis], and its reading is recommended. Specifically the TCP/IP MD5 authentication option [RFC5925] should be used with LDP as described in LDP specification [RFC5036]. If TCP/IP MD5 authentication is considered not secure enough, the designer may consider using a more elaborate and advanced TCP Authentication Option TCP-AO [RFC5925] for LDP session authentication.

Access IGP (if used) and any routing protocols used in access network for signaling service routes should also be secured in a similar manner. Refer to [I-D.ietf-karp-routing-tcp-analysis] and [RFC6863] for further analysis of security properties of IS-IS and OSPF IGP routing protocols.

For increased level of authentication in the control plane security for a subset of access locations with lower physical security, designer could also consider using:

- different crypto keys for use in authentication procedures for these locations.

- stricter network protection mechanisms including DoS protection, interface and session flap dampening.

7.4. Network Node Security

If a network node, especially an Access Node, is not located in a physically secured and controlled location, then this Access Node should implement some measures to provide a level of protection of the key(s) used to its authenticate to the network, so as to avoid an attacker to get those keys easily. Software tools should monitor and keep checking the integrity of the Access Node configuration and software version. Note that this is not specific to the node using LDP DoD. In the contrary, the use of LDP DoD will allow the upstream /network to check, log and possibly deny the FEC requests from the Access Node.

8. Acknowledgements

The authors would like to thank Nischal Sheth, Nitin Bahadur, Nicolai Leymann, George Swallow, Geraldine Calvignac, Ina Minei, Eric Gray and Lizhong Jin for their suggestions and review. Additional thanks go to Adrian Farrel for thorough pre-publication review and editing suggestions.

9. References

9.1. Normative References

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

[RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, January 2001.

[RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, February 2006.

[RFC4447] Martini, L., Rosen, E., El-Aawar, N., Smith, T., and G. Heron, "Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP)", RFC 4447, April 2006.

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

[RFC5283] Decraene, B., Le Roux, JL., and I. Minei, "LDP Extension for Inter-Area Label Switched Paths (LSPs)", RFC 5283, July 2008.

9.2. Informative References

[I-D.ietf-karp-routing-tcp-analysis]
Jethanandani, M., Patel, K., and L. Zheng, "Analysis of BGP, LDP, PCEP and MSDP Issues According to KARP Design Guide", draft-ietf-karp-routing-tcp-analysis-07 (work in progress), April 2013.

[I-D.ietf-mpls-seamless-mpls]
Leymann, N., Decraene, B., Filsfils, C., Konstantynowicz, M., and D. Steinberg, "Seamless MPLS Architecture", draft-ietf-mpls-seamless-mpls-02 (work in progress), October 2012.

[RFC3107] Rekhter, Y. and E. Rosen, "Carrying Label Information in BGP-4", RFC 3107, May 2001.

[RFC5443] Jork, M., Atlas, A., and L. Fang, "LDP IGP Synchronization", RFC 5443, March 2009.

[RFC5920] Fang, L., "Security Framework for MPLS and GMPLS Networks", RFC 5920, July 2010.

[RFC5925] Touch, J., Mankin, A., and R. Bonica, "The TCP Authentication Option", RFC 5925, June 2010.

[RFC6863] Hartman, S. and D. Zhang, "Analysis of OSPF Security According to the Keying and Authentication for Routing Protocols (KARP) Design Guide", RFC 6863, March 2013.

Authors' Addresses

Thomas Beckhaus
Deutsche Telekom AG
Heinrich-Hertz-Strasse 3-7
Darmstadt 64307
Germany

Phone: +49 6151 58 12825
Email: thomas.beckhaus@telekom.de

Bruno Decraene
France Telecom
38-40 rue du General Leclerc
Issy Moulineaux cedex 9 92794
France

Email: bruno.decraene@orange.com

Kishore Tiruveedhula
Juniper Networks
10 Technology Park Drive
Westford, Massachusetts 01886
USA

Phone: 1-(978)-589-8861
Email: kishoret@juniper.net

Maciek Konstantynowicz
Cisco Systems, Inc.
10 New Square Park, Bedfont Lakes
London
United Kingdom

Email: maciek@cisco.com

Luca Martini
Cisco Systems, Inc.
9155 East Nichols Avenue, Suite 400
Englewood, CO 80112
USA

Email: lmartini@cisco.com