Internet Engineering Task Force (IETF)
Request for Comments: 7032
Category: Standards Track
ISSN: 2070-1721

T. Beckhaus, Ed.
Deutsche Telekom AG
B. Decraene
Orange
K. Tiruveedhula
Juniper Networks
M. Konstantynowicz, Ed.
L. Martini
Cisco Systems, Inc.

October 2013

LDP Downstream-on-Demand in Seamless MPLS

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 networks including high number of devices, device position in network topology, and compute and memory constraints that limit the amount of state access devices can hold. This can be achieved with LDP Downstream-on-Demand (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.

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 5741.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at http://www.rfc-editor.org/info/rfc7032.

Table of Contents

1. Introduction

Seamless MPLS design [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, device position in network topology, and compute and memory constraints that limit the amount of state access devices can hold.

In general, MPLS Label Switching Routers (LSRs) 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. This document concentrates on the unicast connectivity only. Multicast connectivity is a subject for further study.

In Seamless MPLS design [SEAMLESS-MPLS], IP/MPLS protocol optimization is possible due to relatively simple access network topologies. Examples of such topologies involving access nodes (ANs) and aggregation nodes (AGNs) 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 Routing Information Base (RIB) and Forwarding Information Base (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 area 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.

However, the amount of MPLS forwarding state requires additional consideration. In general, MPLS routers implement LDP Downstream Unsolicited (LDP DU) label advertisements [RFC5036] and advertise MPLS labels for all valid routes in their RIB tables. This is seen as an inadequate approach for ANs, which require a small subset of the total routes (and associated labels) based on the required connectivity for the provisioned services. Although filters can be applied to those LDP DU label 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 AN-driven "subscription model" for label distribution in the access network. The approach relies on the standard 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 the 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.

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].

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 [SEAMLESS-MPLS] as shown in Figure 1.

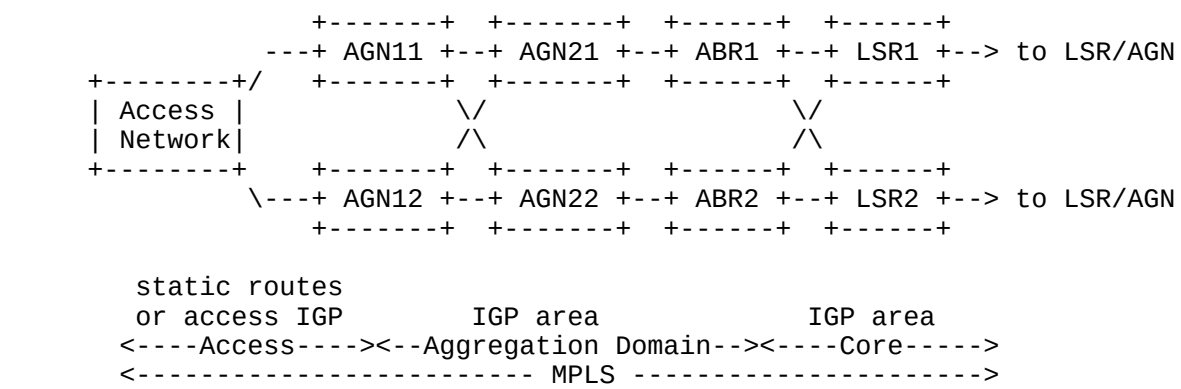


Figure 1: Seamless MPLS End-to-End Reference Network Topology

The access network is either single- or dual-homed to AGN1x, with either a single parallel link 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 it can use either static routing or access IGP (IS-IS 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.

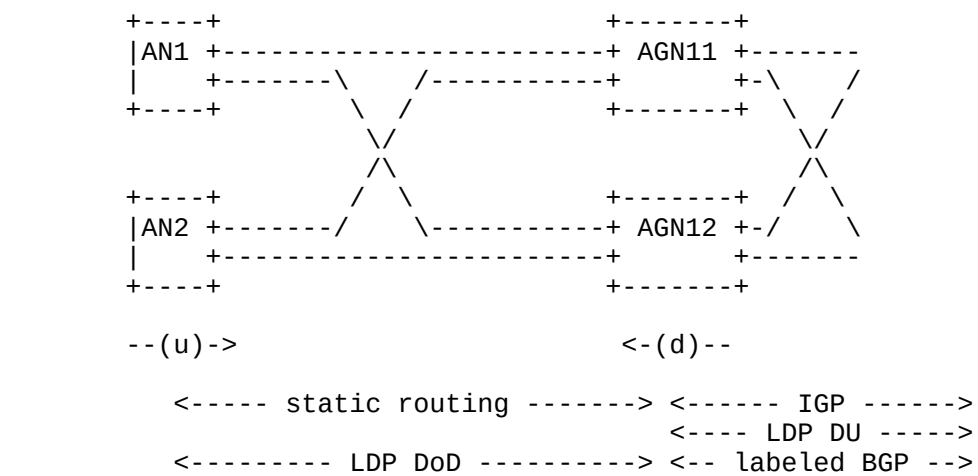
b. [I] topology - multiple ANs daisy-chained to a single AGN.

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 the [I1] topology.

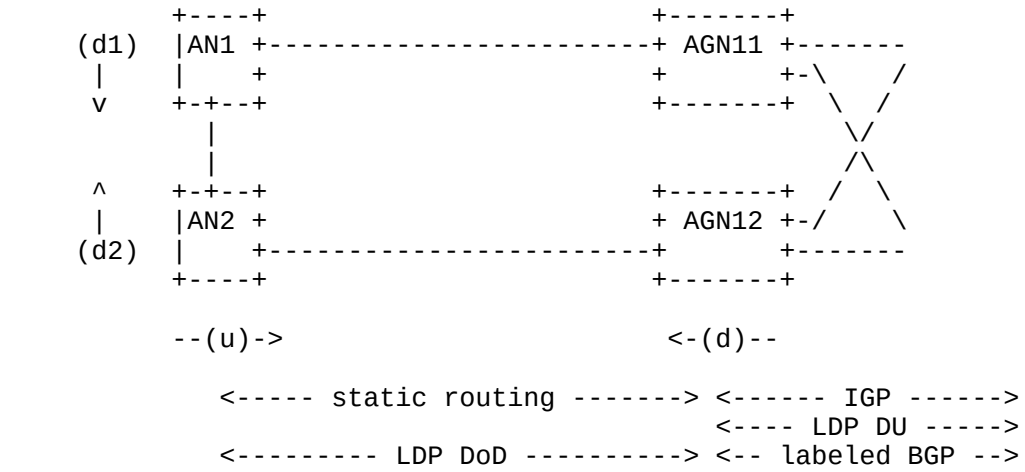


(u) static routes: 0/0 default, (optional) /32 routes
(d) static routes: AN loopbacks

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 IP routes [RFC3107].

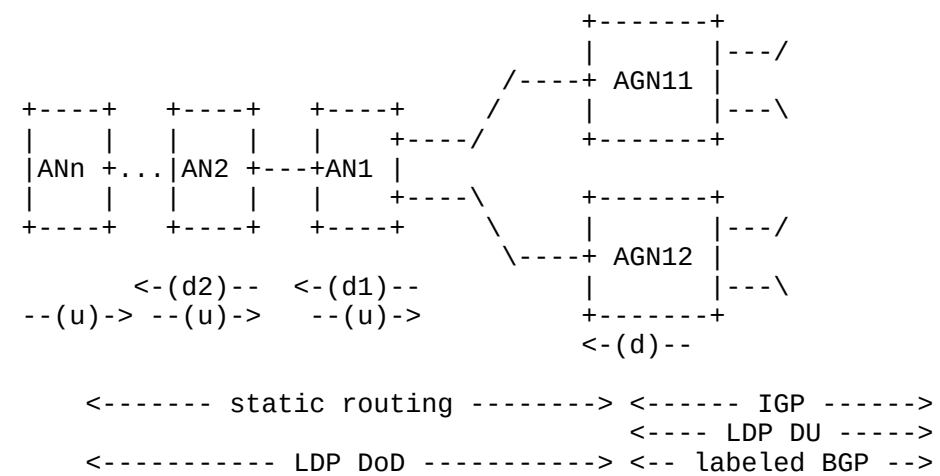
The reference static routing and LDP configuration for [U2] access topology is shown in Figure 3.



(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 the [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 Equal-Cost Multipath (ECMP) (or Layer 2 Link Aggregation Group (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 Forwarding Equivalence Class (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. Examples of such an IGP could be IS-IS, OSPFv2 and v3, or RIPv2 and RIPng. This access IGP instance is distinct from the IGP of the aggregation 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.

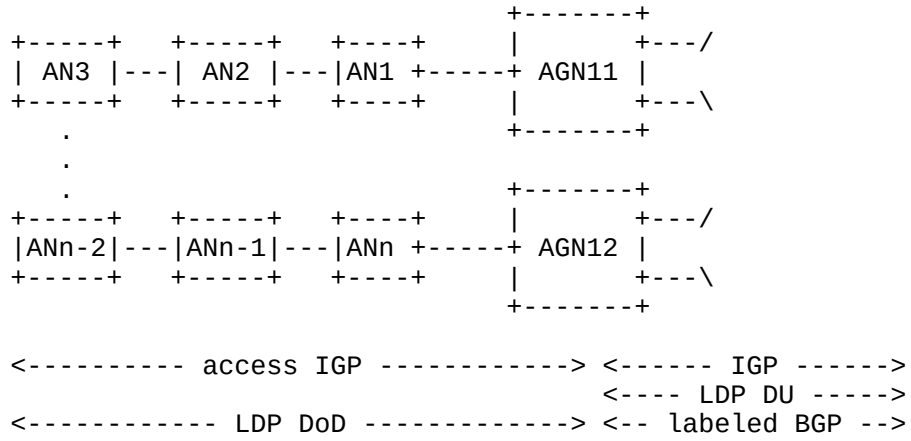


Figure 5: [U] Access Topology with Access IGP

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

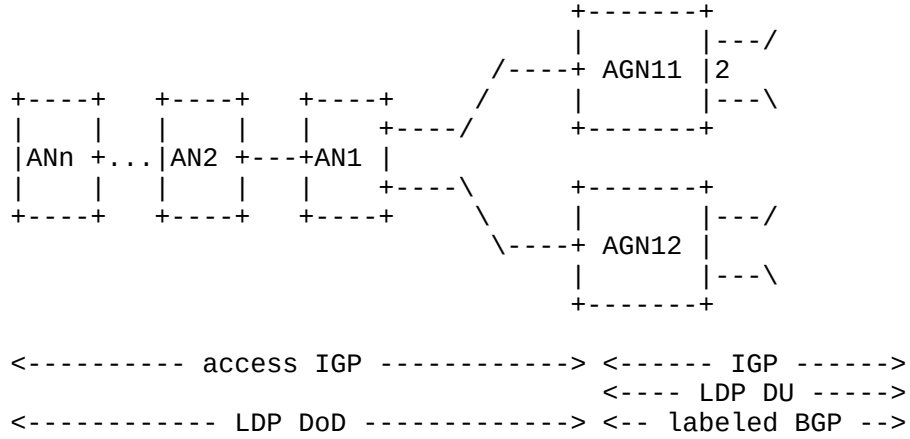


Figure 6: [Y] Access Topology with Access IGP

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 advertises 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 [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 Pseudowire Edge-to-Edge (PWE3) [RFC4447] and BGP/MPLS IP VPN [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 LSRs [RFC3031].

3.1. Initial Network Setup

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

The upstream AN/AGN1x requests labels over an LDP DoD session(s) from downstream AN/AGN1x for configured static routes if those static routes are configured with an 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 a policy on an AN/AGN1x.

The downstream AN/AGN1x responds 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 neighbor or if it is the egress node. In such a case, the downstream AN/AGN1x installs the advertised label as an incoming label in its label information base (LIB) and its label forwarding information base (LFIB). The upstream AN/AGN1x also installs the received label as an outgoing label in its 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 neighbor, it forwards the request to its downstream neighbor.

In order to facilitate ECMP and IP Fast Reroute (IPFRR) Loop-Free Alternate (LFA) local-repair [RFC5286], the upstream AN/AGN1x also sends LDP DoD Label Requests to alternate next hops per its RIB, and installs received labels as alternate entries in its LIB and LFIB.

The AGN1x on the network side can use BGP labeled IP routes [RFC3107] in line with the Seamless MPLS design [SEAMLESS-MPLS]. In such a case, AGN1x will redistribute its static routes pointing to local ANs into BGP labeled IP routes to facilitate network-to-access traffic flows. Likewise, to facilitate access-to-network traffic flows, AGN1x will respond to access-originated LDP DoD Label Requests with label mappings based on its BGP labeled IP routes reachability for requested FECs.

3.1.2. AN with Access IGP

If access IGP is used, an AN(s) advertises its loopbacks over the access IGP with configured metrics. The AGN1x advertises a default route over the access IGP.

Routers request labels over LDP DoD session(s) according to their needs for MPLS connectivity (via Label Switching Paths (LSPs)). In particular, if AGNs, as per Seamless MPLS design [SEAMLESS-MPLS], redistribute routes from the IGP into BGP labeled IP routes [RFC3107], they request labels over LDP DoD session(s) for those routes.

Identical to the static route case, the downstream AN/AGN1x responds 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 neighbor), and installs the advertised label as an incoming label in its LIB and LFIB. The upstream AN/AGN1x also installs the received label as an outgoing label in its LIB and LFIB.

Identical to the static route case, in order to facilitate ECMP and IPFRR LFA local-repair, the upstream AN/AGN1x also sends LDP DoD Label Requests to alternate next hops per its RIB, and it installs received labels as alternate entries in its LIB and LFIB.

The AGN1x on the network side can use labeled BGP [RFC3107] in line with Seamless MPLS design [SEAMLESS-MPLS]. In such a case, AGN1x will redistribute routes received over the access IGP (and pointing to local ANs), into BGP labeled IP routes to facilitate network-to-access traffic flows. Likewise, to facilitate access-to-network traffic flows, the AGN1x will respond to access-originated LDP DoD Label Requests with label mappings based on its BGP labeled IP routes 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 the 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 (tLDP) 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. A tLDP session over an associated TCP/IP connection needs to be established to the PWE3 destination Provider Edge (PE). This is triggered either by an explicit tLDP 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 tLDP 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/non-MPLS connectivity operations that are 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 the 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 Network Layer Reachability Information (NLRI) needs to be exchanged between AN* and its BGP peers.

c. Based on configured BGP policies, VPNv4 routes need to be installed in the AN* VPN Routing and Forwarding (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 are 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 and 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* sends an LDP DoD label mapping request. A 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* accepts this label based on the exact route match of the 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 installs 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's are configured as equal-cost paths, AN* sends LDP DoD Label Requests to both AGN1x devices and installs all received labels in its LIB and LFIB.

In order for AN* to implement IPFRR LFA local-repair, AN* also sends LDP DoD Label Requests to alternate next hops per its RIB, and installs 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* implements ECMP functionality. This involves AN* using a 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.

The 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.

The AGN1x on the network side can use labeled BGP [RFC3107] in line with Seamless MPLS design [SEAMLESS-MPLS]. In such a case, the AGN1x will redistribute its static routes (or routes received from the access IGP) pointing to local ANs into BGP labeled IP routes to facilitate network-to-access traffic flows. Likewise, to facilitate access-to-network traffic flows, the AGN1x will respond to access-originated LDP DoD Label Requests with label mappings based on its BGP labeled IP routes reachability for requested FECs.

3.3. Service Changes and Decommissioning

Whenever the AN* service gets decommissioned or changed and connectivity to a specific destination is no longer required, the associated MPLS LSP label resources are to 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 tLDP session is no longer needed and is to 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 via local or remote reconfiguration can result in a specific BGP next hop(s) no longer being needed. The MPLS LSP(s) to that destination is no longer needed either.

In all of the above cases, the following operations related to LDP DoD 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* deletes the label from its LIB and LFIB, and releases it from the 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 is retained by AN*.

3.4. Service Failure

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

In general, unless the service failure event modifies required MPLS connectivity, there is 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 Sections 3.2 and 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 goes into a failure state. Upon recovery from network failure, the service is to be re-established automatically.

The following additional LDP-related functions need to be supported to comply with Seamless MPLS [SEAMLESS-MPLS] fast service restoration requirements:

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

b. LDP session protection: LDP sessions are 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 the LDP session stays up.

c. IGP-LDP synchronization: If access IGP is used, LDP sessions between ANs, and between ANs and AGN1x, are 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 best downstream next hop.

3.5.2. AN Failure

If the AN fails, adjacent AN/AGN1x nodes remove all routes pointing to the failed node from their RIB tables (including /32 loopback belonging to the failed AN and any other routes reachable via the failed AN). In turn, this triggers the removal of associated outgoing /32 FEC labels from their LIB and LFIB tables.

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

If a specific /32 FEC(s) is no longer reachable from those ANs/AGN1x's, they also send LDP Label Withdraw messages to their upstream LSRs to notify them about the failure, and remove the associated incoming label(s) from their LIB and LFIB tables. Upstream LSRs, upon receiving a Label Withdraw, remove the signaled labels from their LIB/LFIB tables, and propagate LDP Label Withdraws across their upstream LDP DoD sessions.

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

AGN1x is notified about the AN failure via access IGP (if used) and/or cascaded LDP DoD Label Withdraw(s). AGN1x implements all relevant global-repair IP/MPLS procedures to propagate the AN failure towards the core network. This involves removing associated routes (in the access IGP case) and labels from its LIB and LFIB tables, and

propagating the failure on the network side using labeled BGP and/or core IGP/LDP DU procedures.

Upon the 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 is 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 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 remove routes pointing to the failed link immediately from the RIB, remove associated labels from their LIB and LFIB tables, and 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. The AN/AGN1x node stops using the failed link and immediately switches over (local-repair) to the remaining ECMP path or alternate path. The AN/AGN1x removes affected next hops and labels. If there is an AGN1x terminating the failed link, it immediately removes routes pointing to the failed link from the RIB, removes any associated labels from the LIB and LFIB tables, and propagates the failure on the network side using labeled BGP 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 to downstream AN/AGN1x nodes.

3.5.4. AGN Failure

If an AGN1x fails adjacent access then, depending on the access topology, the following cases apply to the network operation (topology references from Section 2 are shown in square brackets):

a. [I1, I] - ANs are isolated from the network - An AN adjacent to the failure immediately removes routes pointing to the failed AGN1x from the RIB, removes associated labels from the LIB and LFIB tables, and sends LDP Label Withdraw message(s) to its upstream neighbors. If access IGP is used, an IGP link update is sent.

b. [U2, U, V, Y] - at least one ECMP or alternate path remains. AN adjacent to failed AGN1x stops using the failed link and immediately switches over (local-repair) to the remaining ECMP path or alternate path by following LDP [RFC5036] procedures. (Appendix A.1.7 "Detect Change in FEC Next Hop")

Network-side procedures for handling AGN1x failure have been described in Seamless MPLS [SEAMLESS-MPLS].

3.5.5. AGN Network-Side Reachability Failure

If AGN1x loses network reachability to a specific destination or set of network-side destinations, AGN1x sends LDP Label Withdraw messages to its upstream ANs, withdrawing labels for all affected /32 FECs. Upon receiving those messages, ANs remove those labels from their LIB and LFIB tables, and use alternative LSPs instead (if available) as part of global-repair.

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

4. LDP DoD Procedures

All LDP Downstream-on-Demand implementations follow the Label Distribution Protocol as specified in [RFC5036]. 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 the upstream LSR to the 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.

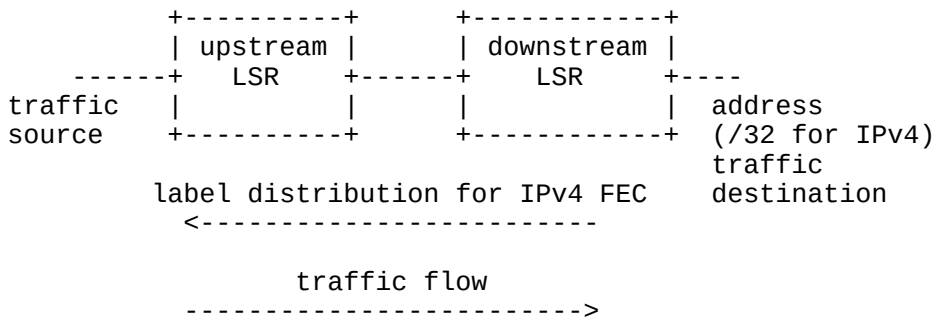


Figure 7: LDP Label Assignment Direction

4.1. LDP Label Distribution Control and Retention Modes

The LDP specification [RFC5036] defines two modes for label distribution control, following the definitions in the 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 an upstream LSR to be notified about the downstream failure and for an application using the LSP to switch over to an alternate path, even if such a path exists.

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

- Conservative label retention mode: If operating in DoD mode, an LSR will request label mappings only from the next-hop LSR according to routing. The main advantage of the conservative label retention 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 label retention 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 label retention mode: When operating in DoD mode with liberal label retention mode, 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 label retention 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. In turn, this 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 label 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 label 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 be requested only after the access IGP convergence.

Adhering to the overall design goals of Seamless 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 mode: enables pre-programming of alternate next hops with associated FEC labels.

In Seamless MPLS [SEAMLESS-MPLS], an AGN1x acts as an access ABR connecting access and metro domains. To enable failure propagation between those domains, the access ABR implements ordered label distribution control when redistributing routes/FECs between the

access side (using LDP DoD and static or access IGP) and the network side (using labeled BGP [RFC3107] or core IGP with LDP Downstream Unsolicited label advertisements).

4.2. LDP DoD Session Negotiation

An access LSR/ABR proposes the DoD label advertisement by setting the "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 the LDP specification [RFC5036].

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

If an access LSR is acting in an active role, it re-attempts the LDP session immediately. If the access LSR receives the same Downstream Unsolicited mode again, it follows the exponential backoff algorithm as defined in the LDP specification [RFC5036] with a 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 is used for PWE3 FECs. Even if the LDP DoD label advertisement has been negotiated for IPv4 and IPv6 LDP FECs as described earlier, the LDP session uses a 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

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

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

b. An access LSR/ABR is configured with a service in line with service use cases described in Sections 3.2 and 3.3.

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

d. Configuration with access IGP: An access LSR/ABR link to an adjacent node comes up, and an LDP DoD session is established. In this case, the access LSR sends 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 are sent to any next-hop LSRs and alternate LSRs.

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

a. Downstream access LSR/ABR: The requested FEC is an IGP or static route, and there is an LDP label already learned 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, the downstream LSR sends a Label Request message for the same FEC to the next-next-hop downstream LSR. In such a case, the 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: The requested FEC is a BGP labeled IP routes [RFC3107], and this BGP route is the best selected for this FEC.

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

4.3.2. Label Request Retry

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

If there is no response to the Label Request message sent, the LDP specification [RFC5036] (Section A.1.1) states that the LSR does not send another request for the same label to the peer and mandates that a duplicate Label Request be considered a protocol error and 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 does not send a duplicate Label Request message.

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 sends a Label Abort message to the downstream LSR.

To address the case of slower convergence resulting from described LDP behavior in line with the 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.

The downstream access LSR/ABR withdraws a label for a specific FEC in the following cases:

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

b. If an 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 the downstream LDP session is lost.

c. If an 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 an LDP DoD ingress label is associated with an outgoing label assigned by LDP (DoD or DU), the next hop in the route has changed, and

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

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

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

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

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

The downstream access LSR/ABR withdraws a label if the local route configuration (e.g., /32 loopback) is deleted.

Note: For any events inducing next-hop change, a downstream access LSR/ABR attempts 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 the LDP DoD session, then the downstream access LSR/ABR sends a Label Request on the new next-hop session. If the downstream access LSR/ABR doesn't get a label mapping for some duration, then and only then does the downstream access LSR/ABR withdraw the upstream label.

4.5. Label Release

If an access LSR/ABR no longer needs 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 an upstream access LSR/ABR receives an unsolicited label mapping on a DoD session, it releases the label by sending a Label Release message.

The access LSR/ABR sends 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 next hop in the route has 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, the access LSR/ABR requests 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, the 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 [RLFA]. If the access LSR/ABR doesn't already know those labels, it requests them.

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

5. LDP Extension for LDP DoD Fast-Up Convergence

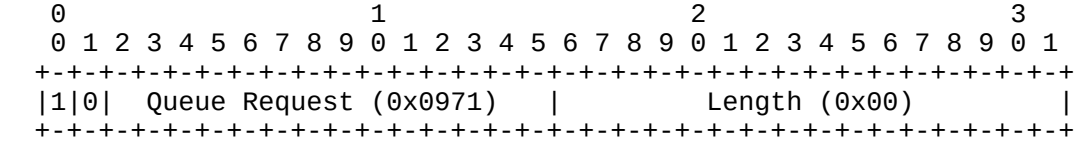
In some conditions, the exponential backoff algorithm usage described in Section 4.3.2 can result in a wait time that is longer than desired to get a successful LDP label-to-route mapping. An example is when a specific route is unavailable on the downstream LSR when the label mapping request from the upstream is received, but later comes back. In such a case, using the exponential backoff algorithm can 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 is to be treated as a normative reference. The downstream and upstream LSRs SHOULD implement this extension if fast-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 the 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 the requested route is not available, the downstream LSR MUST ignore this unknown TLV and send a "No Route" notification back. This 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 the F-bit being clear (=0), the unknown TLV is not forwarded with the message.

Type = 0x0971
Queue Request TLV (allocated by IANA).

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

The 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 a route is available; if so, it replies with a 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 remove the original request from the queue and send back a Label Request Aborted notification [RFC5036].

If the downstream LSR does not support the Queue Request TLV, and the 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 the LDP specification [RFC5036].

This procedure ensures backward compatibility.

6. IANA Considerations

6.1. LDP TLV Type

This document uses a new Optional Parameter, Queue Request TLV, in the Label Request message defined in Section 5. IANA already maintains a registry of LDP parameters called the "TLV Type Name Space" registry, as defined by RFC 5036. The following assignment has been made:

TLV type	Description
0x0971	Queue Request TLV

7. Security Considerations

MPLS LDP DoD deployment in the access network is subject to the same security threats as any MPLS LDP deployment. It is recommended that baseline security measures be 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 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 Very high bit-rate Digital Subscriber Line (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, route advertisements).

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

7.1. LDP DoD Native Security Properties

MPLS LDP DoD operation is request driven, and unsolicited label mappings are not accepted by upstream LSRs by design. This inherently limits the potential of an unauthorized third party injecting unsolicited label mappings on the wire.

This native security property enables an ABR LSR to act as a gateway to the MPLS network and to control the requests coming from any access LSR and prevent cases when the access LSR attempts to get access to an unauthorized FEC or remote LSR after being compromised.

In the event that an access LSR gets compromised and manages to advertise a FEC belonging to another LSR (e.g., in order to 'steal' third-party data flows, or breach the privacy of a VPN), such an access LSR would also have to influence the routing decision for affected FECs on the ABR LSR to attract the flows. The following measures need to be considered on an ABR LSR to prevent such an event from occurring:

a. Access with static routes: An access LSR cannot influence ABR LSR routing decisions due to the static nature of routing configuration, a native property of the design.

b. Access with IGP - access FEC "stealing": If the compromised access LSR is a leaf in the access topology (leaf node in topologies I1, I, V, Y described earlier), this will not have any adverse effect, due to the leaf IGP 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 only possible for this access LSR to attract traffic destined to the nodes upstream from it. Such a 'man-in-the-middle attack' can quickly be detected by upstream access LSRs not receiving traffic and by the LDP TCP session being lost.

c. Access with IGP - network FEC "stealing": The compromised access LSR can use IGP to advertise a "stolen" FEC prefix belonging to the network side. This case can be prevented by giving a better administrative preference to the BGP labeled IP routes versus access IGP routes.

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

The following two sections describe other security considerations applicable to general MPLS deployments in the access network.

7.2. Data-Plane Security

Data-plane security risks applicable to the access MPLS network include:

a. Labeled packets from a specific access LSR that are sent to an unauthorized destination.

b. Unlabeled packets that are sent by an access LSR to remote network nodes.

The following mechanisms apply to MPLS access design with LDP DoD that address listed data-plane security risks:

1. addressing (a): Access and ABR LSRs do 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 control-plane security as described in Section 7.3 and the top label has been distributed to the upstream neighbor by the receiving access or ABR LSR.

2. addressing (a) - The ABR LSR restricts network reachability for access devices to a subset of remote network LSRs, based on control-plane security as described in Section 7.3, FEC filters, and routing policy.

3. addressing (a): Control-plane authentication as described in Section 7.3 is used.

4. addressing (b): The ABR LSR restricts IP network reachability to and from the access LSR.

7.3. Control-Plane Security

Similar to Inter-AS MPLS/VPN deployments [RFC4364], control-plane security is a prerequisite for data-plane security.

To ensure control-plane security access, LDP DoD sessions are established only 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 sessions is analyzed in the LDP specification [RFC5036] and in [RFC6952] ("Analysis of BGP, LDP, PCEP, and MSDP Issues According to the Keying and Authentication for Routing Protocols (KARP) Design Guide"). Both documents state that LDP is subject to two different types of attacks: spoofing and denial-of-service attacks.

The threat of spoofed LDP Hello messages can be reduced by following guidelines listed in the LDP specification [RFC5036]: accepting Basic Hellos only on interfaces connected to trusted LSRs, ignoring Basic Hellos that are not addressed to all routers in this subnet multicast group, and using access lists. LDP Hello messages can also be secured using an optional Cryptographic Authentication TLV as specified in "LDP Hello Cryptographic Authentication" [CRYPTO-AUTH] that further reduces the threat of spoofing during the LDP discovery phase.

Spoofing during the LDP session communication phase can be prevented by using the TCP Authentication Option (TCP-AO) [RFC5925], which uses a stronger hashing algorithm, e.g., SHA1 as compared to the traditionally used MD5 authentication. TCP-AO is recommended as being more secure as compared to the TCP/IP MD5 authentication option [RFC5925].

The threat of a denial-of-service attack targeting a well-known UDP port for LDP discovery or a TCP port for LDP session establishment can be reduced by following the guidelines listed in [RFC5036] and in [RFC6952].

Access IGP (if used) and any routing protocols used in the access network for signaling service routes also need to be secured following best practices in routing protocol security. Refer to the KARP IS-IS security analysis document [KARP-ISIS] and to [RFC6863] ("Analysis of OSPF Security According to the Keying and Authentication for Routing Protocols (KARP) Design Guide") for further analysis of security properties of IS-IS and OSPF IGP routing protocols.

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 to Stephen Kent for review and guidance specifically for the security section.

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

[CRYPTO-AUTH] Zheng, L., Chen, M., and M. Bhatia, "LDP Hello Cryptographic Authentication", Work in Progress, August 2013.

[KARP-ISIS] Chunduri, U., Tian, A., and W. Lu, "KARP IS-IS security analysis", Work in Progress, March 2013.

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

[RFC5286] Atlas, A. and A. Zinin, "Basic Specification for IP Fast Reroute: Loop-Free Alternates", RFC 5286, September 2008.

[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.

[RFC6952] Jethanandani, M., Patel, K., and L. Zheng, "Analysis of BGP, LDP, PCEP, and MSDP Issues According to the Keying and Authentication for Routing Protocols (KARP) Design Guide", RFC 6952, May 2013.

[RLFA] Bryant, S., Filsfils, C., Previdi, S., Shand, M., and N. So, "Remote LFA FRR", Work in Progress, May 2013.

[SEAMLESS-MPLS] Leymann, N., Ed., Decraene, B., Filsfils, C., Konstantynowicz, M., Ed., and D. Steinberg, "Seamless MPLS Architecture", Work in Progress, July 2013.

Authors' Addresses

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

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

Bruno Decraene
Orange
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 (editor)
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