Network Working Group
Internet-Draft
Expires: March 9, 2008

M. Jork
Reef Point
A. Atlas
Google
L. Fang
Cisco Systems, Inc.

September 6, 2007

LDP IGP Synchronization
draft-ietf-mpls-igp-sync-00

Status of this Memo

Copyright Notice

Abstract

In networks depending on edge-to-edge establishment of MPLS forwarding paths via LDP, blackholing of traffic can occur in situations where the IGP is operational on a link and thus the link is used for IP forwarding but LDP is not operational on that link for whatever reason. This document describes a mechanism to avoid traffic loss due to this condition without introducing any protocol changes.

1. Introduction

LDP [RFC3036] establishes MPLS LSPs along the shortest path to a destination as determined by IP forwarding. In a common network design, LDP is used to provide label switched paths throughout the complete network domain covered by an IGP such as OSPF [RFC2328] or IS-IS [ISO.10589.1992], i.e. all links in the domain have IGP as well as LDP adjacencies.

A variety of services a network provider may want to deploy over an LDP enabled network depend on the availability of edge to edge label switched paths. In a L2 or L3 VPN scenario for example, a given PE router relies on the availability of a complete MPLS forwarding path to the other PE routers for the VPNs it serves. This means that along the IP shortest path from one PE router to the other, all the links need to have operational LDP sessions and the necessary label binding must have been exchanged over those sessions. If only one link along the IP shortest path is not covered by an LDP session, a blackhole exists and services depending on MPLS forwarding will fail. This might be a transient or a persistent error condition. Some of the reasons for it could be

- a configuration error,

- an implementation bug,

- the link has just come up and has an IGP adjacency but LDP has either not yet established an adjacency or session or distributed all the label bindings.

The LDP protocol itself has currently no means to indicate to a service depending on it whether there is an uninterrupted label switched path available to the desired destination or not.

2. Proposed Solution

The problem described above exists because LDP is tied to IP forwarding decisions but no coupling between the IGP and LDP operational state on a given link exists. If IGP is operational on a link but LDP is not, a potential network problem exists. So the solution described by this document is to prevent a link from being used for IP forwarding as long as LDP is not fully operational. This has some similarity to the mechanism specified in [RFC3137] which allows an OSPF router to advertise that it should not be used as a transit router. One difference is that [RFC3137] raises the link costs on all (stub) router links, while the mechanism described in here applies on a per-link basis.

In detail: when LDP is not "fully operational" (see below) on a given link, the IGP will advertise the link with maximum cost to avoid any transit traffic over it if possible. In the case of OSPF this cost is LSInfinity (16-bit value 0xFFFF) as proposed in [RFC3137]. Note that the link is not just simply removed from the topology because LDP depends on the IP reachability to establish its adjacency and session. Also, if there is no other link in the network to reach a particular destination, no additional harm is done by making this link available for IP forwarding at maximum cost.

LDP is considered fully operational on a link when an LDP hello adjacency exists on it, a suitable associated LDP session (matching the LDP Identifier of the hello adjacency) is established to the peer at the other end of the link and all label bindings have been exchanged over the session. The latter condition can not generally be verified by a router and some heuristics may have to be used. A simple implementation strategy is to wait some time after LDP session establishment before declaring LDP fully operational in order to allow for the exchange of label bindings. This is typically sufficient to deal with the link when it is being brought up. LDP protocol extensions to indicate the complete transmission of all currently available label bindings after a session has come up are conceivable but not addressed in this document.

The mechanism described in this document does not entail any protocol changes and is a local implementation issue. However, it is recommended that both sides of a link implement this mechanism to be effective and to avoid asymmetric link costs which could cause problems with IP multicast forwarding.

The problem space and solution specified in this document have also been discussed in an IEEE Communications Magazine paper [LDP-Fail].

3. Applicability

Example network scenarios that benefit from the mechanism described in here are MPLS VPNs and BGP-free core network designs where traffic can only be forwarded through the core when LDP forwarding state is available throughout.

In general, the proposed procedure is applicable in networks where the availability of LDP signaled MPLS LSPs and avoidance of blackholes for MPLS traffic is more important than always choosing an optimal path for IP forwarded traffic. Note however that non-optimal IP forwarding only occurs for a short time after a link comes up or when there is a genuine problem on a link. In the latter case an implementation should issue network management alerts to report the error condition and enable the operator to address it.

The usefulness of this mechanism also depends on the availability of alternate paths with sufficient bandwidth in the network should one link get costed out due to unavailability of LDP service over it.

On broadcast links with more than one IGP/LDP peer, the cost-out procedure can only be applied to the link as a whole and not an individual peer. So a policy decision has to be made whether the unavailability of LDP service to one peer should result in the traffic being diverted away from all the peers on the link.

4. Interaction With TE Tunnels

In some networks, LDP is used in conjunction with RSVP-TE which sets up traffic-engineered tunnels. The path computation for the TE tunnels is based on the TE link cost which is flooded by the IGP in addition to the regular IP link cost. The mechanism described in this document should only be applied to the IP link cost to prevent any unnecessary TE tunnel reroutes.

In order to establish LDP LSPs across a TE tunnel, a targeted LDP session between the tunnel endpoints needs to exist. This presents a problem very similar to the case of a regular LDP session over a link (the case discussed so far): when the TE tunnel is used for IP forwarding, the targeted LDP session needs to be operational to avoid LDP connectivity problems. Again, raising the IP cost of the tunnel while there is no operational LDP session will solve the problem. When there is no IGP adjacency over the tunnel and the tunnel is not advertised as link into the IGP, this becomes a local issue of the tunnel headend router.

5. Security Considerations

A DoS attack that brings down LDP service on a link or prevents it from becoming operational on a link will now additionally cause non-optimal IP forwarding within the network. However, as discussed above this is considered beneficial as it prevents MPLS traffic from being dropped.

6. IANA Considerations

This document has no actions for IANA.

7. References

[RFC3036] Andersson, L., Doolan, P., Feldman, N., Fredette, A., and B. Thomas, "LDP Specification", RFC 3036, January 2001.

[RFC2328] Moy, J., "OSPF Version 2", STD 54, RFC 2328, April 1998.

[RFC3137] Retana, A., Nguyen, L., White, R., Zinin, A., and D. McPherson, "OSPF Stub Router Advertisement", RFC 3137, June 2001.

[ISO.10589.1992] International Organization for Standardization, "Intermediate system to intermediate system intra-domain-routing routine information exchange protocol for use in conjunction with the protocol for providing the connectionless-mode Network Service (ISO 8473)", ISO Standard 10589, 1992.

[LDP-Fail] Fang, L., Atlas, A., Chiussi, F., Kompella, K., and G. Swallow, "LDP Failure Detection and Recovery", IEEE Communications Vol.42 No.10, October 2004.

Authors' Addresses

Markus Jork
Reef Point Systems
8 New England Executive Park
Burlington, MA 01803
US

Phone: +1 781 359 5071
Email: mjork@reefpoint.com

Alia Atlas
Google, Inc.
One Broadway, 7th Floor
Cambridge, MA 02142
US

Email: akatlas@google.com

Luyuan Fang
Cisco Systems, Inc.
300 Beaver Brook Road
Boxborough, MA 01719
US

Email: lufang@cisco.com

Full Copyright Statement

Intellectual Property

specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

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

Acknowledgment

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA).