Draft document: Graceful Shutdown in MPLS Traffic Engineering Networks

Abstract
Graceful shutdown is a method for explicitly notifying a set of LSRs that either a Link or an entire node will remove itself from the network or the protocol is going to be disabled for a link or a node. Graceful shutdown mechanisms are tailored towards addressing the planned outage in the network.

This document provides protocol mechanisms so as to reduce/eliminate traffic disruption in the event of a planned shutdown of a network resource.

Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC-2119 [1].

Routing Area ID Summary

(This section to be removed before publication.)

SUMMARY

This document describes graceful shutdown mechanisms used in MPLS Traffic Engineering network.

WHERE DOES IT FIT IN THE PICTURE OF THE ROUTING AREA WORK?

This work requires protocol extension to signaling (RSVP-TE) and routing (OSPF/ISIS) protocols that are under IETF Routing Area.

WHY IS IT TARGETED AT THIS WG?

This work fits in the context of [RFC 3209], [RFC 3473] and extensions to these RFCs being defined in CCAMP.

RELATED REFERENCES

See the reference section.

Table of Contents

4.1. Graceful Shutdown of TE link(s).................................4
4.2. Graceful Shutdown of Component Link(s) in a Bundled TE Link....5
4.3. Graceful Shutdown of TE Node..................................5
4.4. Grace Period and Removal of Resource...........................5
5.1. Graceful Shutdown of TE link(s)...............................5
5.2. Graceful Shutdown of Component Link(s) in a Bundled TE Link....6
5.3. Graceful Shutdown of TE Node..................................6
9.1. Normative Reference...........................................7
9.2. Informative Reference.........................................8

1. Terminology
LSR - Label Switch Router

LSP - An MPLS Label Switched Path

Inter-AS MPLS TE LSP: TE LSP whose Head-end LSR and Tail-end LSR do not reside within the same Autonomous System (AS) or both Head-end LSR and Tail-end LSR are in the same AS but the TE tunnel transiting path may be across different ASes.

Inter-Area MPLS TE LSP: TE LSP whose Head-end LSR and Tail-end LSR do not reside within the same IGP area/level or both Head-end LSR and Tail-end LSR are in the same area/level but the TE tunnel transiting path may be across different areas/levels.

PCE: Path Computation Element whose function is to compute the path of a TE LSP it is not the head-end for. The PCE may be an LSR (e.g ABR or ASBR) in the context of some distributed PCE-based path computation scenario as defined in [INTER-AREA-AS] or a centralized Path Computation Element not forwarding packet.

2. Introduction

Some of the outages in a network are planned, in which case protocols extensions can be used so as to avoid traffic disruption by contrast with unplanned network element failure, where traffic disruption can be reduced but may not avoided. Hence, a Service Provider may desire to gracefully (temporarily or definitely) remove a TE Link, a group of TE Links or an entire node for administrative reasons such as link maintenance or LSR software/hardware upgrade. In all these cases, the goal is to minimize impact on the MPLS traffic engineered flows in the network.

In an MPLS Traffic Engineering (TE) enabled network, there are currently no defined mechanisms to allow for graceful shutdown of network resources (TE links or TE nodes). In this document, we describe graceful shutdown mechanisms for MPLS Traffic Engineering network. Specifically, this document proposes signaling and routing extensions to alert the head-end LSR of the graceful shutdown events.

Graceful shutdown mechanisms allow for the rerouting of the affected TE LSPs in a non-disruptive fashion using the so-called make-before-break technique. Furthermore, such mechanisms also prevent other network nodes to use network resources which are about to be shutdown, should new TE LSP be set up.

3. Applicability of IGP and RSVP-TE Mechanisms

An IGP based solution is not applicable when dealing with Inter-area and Inter-AS traffic engineering, as LSA or LSP flooding is restricted to IGP areas/levels. Consequently, RSVP based mechanisms are required to cope with TE LSPs spanning multiple domains, where a domain is defined as either an IGP area or an Autonomous System.
Nonetheless, RSVP mechanisms only convey the information for the transiting LSPs to the router along the upstream Path and not to all nodes in the network. Indication of graceful shutdown via IGP flooding is required to discourage a node from establishing new LSPs through the resources being shut-down.

The following sections specify OSPF/ISIS flooding and RSVP-TE signaling procedures for graceful removal of resources.

4. OSPF/ISIS Mechanisms for graceful shutdown

The procedures provided in this section are equally applicable to OSPF and ISIS.

4.1. Graceful Shutdown of TE link(s)

The link-attribute sub-TLV defined in [OSPF-LINK-ATTRI] and [ISIS-LINK-ATTRI] has been extended to allow a Service Provider to take a TE Link or a group of TE Links out of service for administrative reasons. Specifically, the node where graceful-shutdown of a link is desired MUST originate the TE LSA/LSP containing link-attribute sub-TLV with ‘local maintenance required’ bit set (see OSPF-LINK-ATTRI] and [ISIS-LINK-ATTRI] for encoding details).

Extension to link attribute sub-TLV is preferred over use of MAX-METRIC based solution, as links with MAX-METRIC bandwidth can be used as last resort links in path computation. Nonetheless, to deal with LSRs not compliant with this document, the node initiating graceful shutdown MAY also originate the TE LSA/LSP containing Link TLV with 0 unreserved bandwidth, Traffic Engineering metric set to 0xffffffff, and if the Link has LSC or FSC as its Switching Capability then also with 0 as Max LSP Bandwidth.

Neighbors of the node under graceful shutdown procedure (either at the link, or a group of links) SHOULD continue advertise the actual unreserved bandwidth on the TE links from the neighbors to that node, without any routing adjacency change.

4.2. Graceful Shutdown of Component Link(s) in a Bundled TE Link

If graceful shutdown procedure is performed for a component link within a TE Link bundle and it is not the last component link available within the TE link, the link attributes associated with the TE link are recomputed. If the removal of the component link results in a significant change event, the TE link is re-flooded with the new traffic parameters. If the last component link is being shut-down, the
routing procedure outlined in Section 4.1 is used.

4.3. Graceful Shutdown of TE Node

In the event of Graceful Shutdown of the entire node, procedure outlined in Section 4.1 is applied to all the links advertised by the node under shutdown. Neighbors of the node under graceful shutdown procedure SHOULD continue advertise the actual unreserved bandwidth on the TE links from the neighbors to that node, without any routing adjacency change.

4.4. Grace Period and Removal of Resource

The node initiating the graceful shutdown condition SHOULD delay the removal of resources in question for some period determined by local policy. This is to allow other LSRs in the network to gracefully reroute their TE LSP away from the resources being removed.

5. RSVP-TE Signaling Mechanism for graceful shutdown

5.1. Graceful Shutdown of TE link(s)

The "local TE link maintenance required" error code as defined in [PATH-REOPT] is used to explicitly signal graceful shutdown of a link to the head-end LSR for triggering the reroute of an affected TE LSP. Specifically, the node where graceful-shutdown of a link or a set of links is desired MUST trigger a Path Error message with "local link maintenance required" sub-code for all affected LSPs. However, when a GS operation is performed along the path of a protected LSP, the PLR MAY trigger Fast Reroute [FRR] for the appropriate set of affected TE LSPs and forward a Path Error with "Tunnel locally repaired" sub-code, as per the procedures specified in [MPLS-FRR].

When a head-end node or an intermediate node (in the case of loose hops path computation) or a PCE receives Path Error notify message with sub-code "Local Maintenance on TE Link required Flag", it SHOULD immediately perform a make-before-break to avoid traffic loss. A head-end router or an intermediate node (in the case of loose hops path computation) or a PCE SHOULD avoid the IP address contained in the PathErr in performing path computation for rerouting the LSP.

5.2. Graceful Shutdown of Component Link(s) in a Bundled TE Link

MPLS TE Link Bundling draft [BUNDLE] requires that an LSP is pinned down to component link(s). Hence, when a component link is shut-down, the LSPs affected by such maintenance action needs to be re-signaled.

Graceful shutdown of a component link in a bundled TE link differs from graceful shutdown of unbundled TE link or entire bundled TE link. Specifically, in the former case, when only a subset of component links and not the entire TE bundled link is being shutdown, the remaining component links of the TE links may still be able to admit new LSPs. Consequently a new error sub-code for Path Error - Notify Error is
needed:

9 (TBA)  Local component link maintenance required

Error Sub-code for "Local component link maintenance required" is to be assigned by IANA.

If the last component link is being shut-down, the procedure outlined in Section 5.1 is used.

When a head-end node or an intermediate node (in the case of loose hops path computation) or a PCE receives an RSVP Path Error notification with sub-code "local component link maintenance required" Flag set, it SHOULD immediately perform a make-before-break to avoid traffic loss. The head-end router or an intermediate node (in the case of loose hops path computation) or a PCE MAY NOT avoid the IP address contained in the PathErr in performing path computation for rerouting the LSP. This is because, as mentioned earlier, this address is an IP address of the component link and the flag is an implicit indication that the TE link may still have capacity to admit new LSPs. However, if the ERO is to be computed such that it also provides details of the component link selection(s) along the Path, the component link selection with IP address contained in the PathErr SHOULD be avoided.

As in the case of singling for bundled TE link, the choice of the component link to use is always made by the sender of the Path/REQUEST message, a node receiving the Path Err with "Maintenance on component link required" Flag set SHOULD mark the component link blocked for any future selection.

5.3.  Graceful Shutdown of TE Node

When graceful shutdown at node level is desired, the node in question follows procedure specified in this section for all LSPs.

6.  Security Considerations

This document does not introduce new security issues. The security considerations pertaining to the original RSVP protocol [RSVP] remain relevant.

7.  Intellectual Property Considerations

Cisco Systems may have intellectual property rights claimed in regard to some of the specification contained in this document.

8.  Acknowledgments

The authors would like to acknowledge useful comments from their colleagues David Ward and Sami Boutros.

9.  Reference
9.1. Normative Reference


9.2. Informative Reference

[INTER-AREA-AS] Adrian Farrel, Jean-Philippe Vasseur, Arthi Ayyangar, "A Framework for Inter-Domain MPLS Traffic Engineering", draft-farrel-

Authors’ Address:

Zafar Ali
Cisco Systems, Inc.
100 South Main St. #200
Ann Arbor, MI 48104
USA
zali@cisco.com

Jean Philippe Vasseur
Cisco Systems, Inc.
300 Beaver Brook Road
Boxborough, MA - 01719
USA
Email: jpv@cisco.com

Anca Zamfir
Cisco Systems, Inc.
2000 Innovation Drive
Kanata, Ontario, K2K 3E8
Canada
ancaz@cisco.com


draft-ali-ccamp-mpls-graceful-shutdown-00.txt June 2004