Internet Engineering Task Force (IETF)
Request for Comments: 5817
Category: Informational
ISSN: 2070-1721

Z. Ali
JP. Vasseur
A. Zamfir
Cisco Systems, Inc.
J. Newton
Cable and Wireless

April 2010

Graceful Shutdown in MPLS and Generalized MPLS Traffic Engineering Networks

Abstract

MPLS-TE Graceful Shutdown is a method for explicitly notifying the nodes in a Traffic Engineering (TE) enabled network that the TE capability on a link or on an entire Label Switching Router (LSR) is going to be disabled. MPLS-TE graceful shutdown mechanisms are tailored toward addressing planned outage in the network.

This document provides requirements and protocol mechanisms to reduce or eliminate traffic disruption in the event of a planned shutdown of a network resource. These operations are equally applicable to both MPLS-TE and its Generalized MPLS (GMPLS) extensions.

Status of This Memo

This document is not an Internet Standards Track specification; it is published for informational purposes.

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). Not all documents approved by the IESG are a candidate for any level of Internet Standard; see 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/rfc5817.

Table of Contents

1. Introduction

When outages in a network are planned (e.g., for maintenance purposes), some mechanisms can be used to avoid traffic disruption. This is in contrast with unplanned network element failure, where traffic disruption can be minimized thanks to recovery mechanisms, but may not be avoided. Therefore, a Service Provider may desire to gracefully (temporarily or indefinitely) remove a TE link, a group of TE links, or an entire node for administrative reasons such as link maintenance, software/hardware upgrade at a node, or significant TE configuration changes. In all these cases, the goal is to minimize the impact on the traffic carried over TE LSPs in the network by triggering notifications so as to gracefully reroute such flows before the administrative procedures are started.

These operations are equally applicable to both MPLS-TE [RFC3209] and its Generalized MPLS (GMPLS) extensions [RFC3471] [RFC3473].

This document describes the mechanisms that can be used to gracefully shut down MPLS-TE / GMPLS-TE on a resource such as a TE link, a component link within a bundled TE link, a label resource, or an entire TE node.

Graceful shutdown of a resource may require several steps. These steps can be broadly divided into two sets: disabling the resource in the control plane and disabling the resource in the data plane. The node initiating the graceful shutdown condition introduces a delay between the two sets to allow the control plane to gracefully divert the traffic away from the resource being gracefully shut down. The trigger for the graceful shutdown event is a local matter at the node initiating the graceful shutdown. Typically, graceful shutdown is triggered for administrative reasons, such as link maintenance or software/hardware upgrade.

2. Terminology

LSR: Label Switching Router. The terms node and LSR are used interchangeably in this document.

GMPLS: The term GMPLS is used in this document to refer to packet MPLS-TE, as well as GMPLS extensions to MPLS-TE.

TE Link: The term TE link refers to a single link or a bundle of physical links or FA-LSPs (see below) on which traffic engineering is enabled.

TE LSP: A Traffic Engineered Label Switched Path.

S-LSP: A segment of a TE LSP.

FA-LSP (Forwarding Adjacency LSP): An LSP that is announced as a TE link into the same instance of the GMPLS control plane as the one that was used to create the LSP [RFC4206].

ISIS-LSP: Link State Packet that is generated by IS-IS routers and that contains routing information.

LSA: Link State Advertisement that is generated by OSPF routers and that contains routing information.

TE LSA / TE-IS-IS-LSP: The traffic engineering extensions to OSPF / IS-IS.

Head-end node: Ingress LSR that initiated signaling for the Path.

Border node: Ingress LSR of a TE LSP segment (S-LSP).

PCE (Path Computation Element): An entity that computes the routes on behalf of its clients (PCC) [RFC4655].

Last-resort resource: If a path to a destination from a given head-end node cannot be found upon removal of a resource (e.g., TE link, TE node), the resource is called "last resort" to reach that destination from the given head-end node.

3. Requirements for Graceful Shutdown

This section lists the requirements for graceful shutdown in the context of GMPLS.

- Graceful shutdown is required to address graceful removal of one TE link, one component link within a bundled TE link, a set of TE links, a set of component links, label resources, or an entire node.

- Once an operator has initiated graceful shutdown of a network resource, no new TE LSPs may be set up that use the resource. Any signaling message for a new TE LSP that explicitly specifies the resource, or that would require the use of the resource due to local constraints, is required to be rejected as if the resource were unavailable.

- It is desirable for new TE LSP set-up attempts that would be rejected because of graceful shutdown of a resource (as described in the previous requirement) to avoid any attempt to use the resource by selecting an alternate route or other resources.

- If the resource being shut down is a last-resort resource, based on a local decision, the node initiating the graceful shutdown procedure can cancel the shutdown operation.

- It is required to give the ingress node the opportunity to take actions in order to reduce or eliminate traffic disruption on the TE LSPs that are using the network resources that are about to be shut down.

- Graceful shutdown mechanisms are equally applicable to intra-domain TE LSPs and those spanning multiple domains, as defined in [RFC4726]. Examples of such domains include IGP areas and Autonomous Systems.

- Graceful shutdown is equally applicable to packet and non-packet networks.

- In order to make rerouting effective, it is required that when a node initiates the graceful shutdown of a resource, it notifies all other network nodes about the TE resource under graceful shutdown.

- Depending on switching technology, it may be possible to shut down a label resource, e.g., shutting down a lambda in a Lambda Switch Capable (LSC) node.

4. Mechanisms for Graceful Shutdown

An IGP-only solution based on [RFC3630], [RFC5305], [RFC4203] and [RFC5307] is not applicable when dealing with inter-area and inter-AS traffic engineering, as IGP flooding is restricted to IGP areas/levels. An RSVP-based solution is proposed in this document to handle TE LSPs spanning multiple domains. In addition, in order to discourage nodes from establishing new TE LSPs through the resources being shut down, existing IGP mechanisms are used for the shutdown notification.

A node where a link or the whole node is being shut down first triggers the IGP updates as described in Section 4.1 and then, with some delay to allow network convergence, uses the signaling mechanism described in Section 4.2.

4.1. OSPF / IS-IS Mechanisms for Graceful Shutdown

This section describes the use of existing OSPF and IS-IS mechanisms for the graceful shutdown in GMPLS networks.

The OSPF and IS-IS procedures for graceful shutdown of TE links are similar to the graceful restart of OSPF and IS-IS as described in [RFC4203] and [RFC5307], respectively. Specifically, the node where graceful shutdown of a link is desired originates the TE LSA or IS-IS-LSP containing a Link TLV for the link under graceful shutdown with the Traffic Engineering metric set to 0xffffffff, 0 as unreserved bandwidth. If the TE link has LSC or FSC as its Switching Capability, then it also has 0 in the "Max LSP Bandwidth" field of the Interface Switching Capability Descriptor (ISCD) sub-TLV. A node may also specify a value that is greater than the available bandwidth in the "Minimum LSP bandwidth" field of the same ISCD sub-TLV. This would discourage new TE LSP establishment through the link under graceful shutdown.

If the 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. Similarly, if the graceful shutdown procedure is performed on a label resource within a TE link, the link attributes associated with the TE link are recomputed. If the removal of the component link or label resource results in a significant bandwidth change event, a new LSA is originated with the new traffic parameters. If the last component link is being shut down, the routing procedure related to TE link removal is used.

Neighbors of the node where graceful shutdown procedure is in progress continue to advertise the actual unreserved bandwidth of the TE links from the neighbors to that node, without any routing adjacency change.

When graceful shutdown at node level is desired, the node in question follows the procedure specified in the previous section for all TE links.

4.2 RSVP-TE Signaling Mechanisms for Graceful Shutdown

As discussed in Section 3, one of the requirements for the signaling mechanism for graceful shutdown is to carry information about the resource under graceful shutdown. For this purpose, the graceful shutdown procedure uses TE LSP rerouting mechanism as defined in [RFC5710].

Specifically, the node where graceful shutdown of an unbundled TE link or an entire bundled TE link is desired triggers a PathErr message with the error code "Notify" and error value "Local link maintenance required", for all affected TE LSPs. Similarly, the node that is being gracefully shut down triggers a PathErr message with the error code "Notify" and error value "Local node maintenance

required", for all TE LSPs. For graceful shutdown of a node, an unbundled TE link, or an entire bundled TE link, the PathErr message may contain either an [RFC2205] format ERROR_SPEC object or an IF_ID [RFC3473] format ERROR_SPEC object. In either case, it is the address and TLVs carried by the ERROR_SPEC object and not the error value that indicate the resource that is to be gracefully shut down.

MPLS-TE link bundling [RFC4201] requires that an TE LSP is pinned down to a component link. Consequently, 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 bundled TE link is being shut down, the remaining component links of the bundled TE link may still be able to admit new TE LSPs. The node where graceful shutdown of a component link is desired triggers a PathErr message with the error code "Notify" and error value of "Local link maintenance required". The rest of the ERROR_SPEC object is constructed using Component Reroute Request procedure defined in [RFC5710].

If graceful shutdown of a label resource is desired, the node initiating this action triggers a PathErr message with the error codes and error values of "Notify/Local link maintenance required". The rest of the ERROR_SPEC object is constructed using the Label Reroute Request procedure defined in [RFC5710].

When a head-end node, a transit node, or a border node receives a PathErr message with the error code "Notify" and error value "Local link maintenance required" or "Local node maintenance required", it follows the procedures defined in [RFC5710] to reroute the traffic around the resource being gracefully shut down. When performing path computation for the new TE LSP, the head-end node or border node avoids using the TE resources identified by the ERROR_SPEC object. If the PCE is used for path computation, the head-end (or border) node acting as PCC specifies in its requests to the PCE that path computation should avoid the resource being gracefully shut down. The amount of time the head-end node or border node avoids using the TE resources identified by the IP address contained in the PathErr is based on a local decision at that node.

If the node initiating the graceful shutdown procedure receives a path setup request for a new tunnel-using resource being gracefully shut down, it sends a PathErr message with "Notify" error code in the ERROR SPEC object and an error value consistent with the type of resource being gracefully shut down. However, based on a local decision, if an existing tunnel continues to use the resource being gracefully shut down, the node initiating the graceful shutdown procedure may allow that resource being gracefully shut down to be

used as a "last resort". The node initiating the graceful shutdown procedure can distinguish between new and existing tunnels by inspecting the SENDER TEMPLATE and SESSION objects.

If the resource being shut down is a last-resort resource, it can be used; i.e., based on a local decision, the node initiating the graceful shutdown procedure can cancel the shutdown operation. Similarly, based on a local decision, the node initiating the graceful shutdown procedure can delay the actual removal of resource for forwarding. This is to give time to the network to move traffic from the resource being shut down. For this purpose, the node initiating graceful shutdown procedure follows the Reroute Request Timeout procedure defined in [RFC5710].

5. Manageability Considerations

When a TE link is being shut down, a linkDown trap as defined in [RFC2863] should be generated for the TE link. Similarly, if a bundled TE link is being shut down, a linkDown trap as defined in [RFC2863] should be generated for the bundled TE link, as well as for each of its component links. If a TE node is being shut down, a linkDown trap as defined in [RFC2863] should be generated for all TE links at the node.

6. Security Considerations

This document introduces no new security considerations as it describes usage of existing formats and mechanisms. This document relies on existing procedures for advertisement of TE LSA / IS-IS-LSPs containing Link TLVs. Tampering with TE LSAs / IS-IS-LSPs may have an effect on traffic engineering computations, and it is suggested that any mechanisms used for securing the transmission of normal LSAs / IS-IS-LSPs be applied equally to all Opaque LSAs / IS-IS-LSPs that this document uses. Existing security considerations specified in [RFC3630], [RFC5305], [RFC4203], [RFC5307], and [MPLS-GMPLS-SEC] remain relevant and suffice. Furthermore, the Security Considerations section in [RFC5710] and section 9 of [RFC4736] should be used for understanding the security considerations related to the formats and mechanisms used in this document.

7. Acknowledgments

The authors would like to thank Adrian Farrel for his detailed comments and suggestions. The authors would also like to acknowledge useful comments from David Ward, Sami Boutros, and Dimitri Papadimitriou.

8. References

8.1. Normative References

[RFC2205] Braden, R., Ed., Zhang, L., Berson, S., Herzog, S., and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, September 1997.

[RFC5710] Berger, L., Papadimitriou, D., and JP. Vasseur, "PathErr Message Triggered MPLS and GMPLS LSP Reroutes", RFC 5710, January 2010.

8.2. Informative References

[RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.

[RFC4736] Vasseur, JP., Ed., Ikejiri, Y., and R. Zhang, "Reoptimization of Multiprotocol Label Switching (MPLS) Traffic Engineering (TE) Loosely Routed Label Switched Path (LSP)", RFC 4736, November 2006.

[RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, September 2003.

[RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", RFC 5305, October 2008.

[RFC4203] Kompella, K., Ed., and Y. Rekhter, Ed., "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, October 2005.

[RFC5307] Kompella, K., Ed., and Y. Rekhter, Ed., "IS-IS Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 5307, October 2008.

[RFC3471] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, January 2003.

[RFC3473] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.

[RFC4726] Farrel, A., Vasseur, J.-P., and A. Ayyangar, "A Framework for Inter-Domain Multiprotocol Label Switching Traffic Engineering", RFC 4726, November 2006.

[RFC4201] Kompella, K., Rekhter, Y., and L. Berger, "Link Bundling in MPLS Traffic Engineering (TE)", RFC 4201, October 2005.

[RFC4206] Kompella, K. and Y. Rekhter, "Label Switched Paths (LSP) Hierarchy with Generalized Multi-Protocol Label Switching (GMPLS) Traffic Engineering (TE)", RFC 4206, October 2005.

[RFC4655] Farrel, A., Vasseur, J.-P., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.

[RFC2863] McCloghrie, K. and F. Kastenholz, "The Interfaces Group MIB", RFC 2863, June 2000.

[MPLS-GMPLS-SEC] Luyuan F., Ed., "Security Framework for PLS and GMPLS Networks", Work in Progress, March 2010.

Authors' Addresses

Zafar Ali
Cisco systems, Inc.,
2000 Innovation Drive
Kanata, Ontario, K2K 3E8
Canada
EMail: 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
EMail: ancaz@cisco.com

Jonathan Newton
Cable and Wireless
EMail: jonathan.newton@cw.com