CCAMP Working Group
Internet Draft

Zafar Ali
Jean-Philippe Vasseur
Anca Zamfir
Cisco Systems, Inc.
Jonathan Newton
Cable and Wireless

draft-ietf-ccamp-mpls-graceful-shutdown-07.txt

Graceful Shutdown in MPLS and Generalized MPLS
Traffic Engineering Networks

Status of this Memo

Copyright Notice

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/eliminate traffic disruption in the event of a planned
   shutdown of a network resource. These operations are equally
   applicable to both MPLS and its Generalized MPLS (GMPLS)
   extensions.

Table of Contents

1. Introduction

   When outages in a network are planned (e.g. for maintenance
   purpose), 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. Hence, 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 and its
Generalized MPLS (GMPLS) extensions.

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 removing the resource for
forwarding. The node initiating the graceful shutdown condition
is expected to introduce a delay between disabling the resource
in the control plane and removing the resource for forwarding.
This is to allow the control plane to gracefully divert the
traffic away from TE resource being gracefully shutdown. 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.

This document describes the mechanisms that can be used to
gracefully shutdown MPLS/ GMPLS Traffic Engineering on a
resource. As mentioned earlier, the graceful shutdown of the
Traffic Engineering capability on a resource could be
incorporated in the shutdown operation of an interface, but it is
a separate step that is taken before the IGP on the link is
brought down and before the interface is brought down at
different layers. This document only addresses TE nodes and TE
resources.

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.

LSP - An MPLS-TE/ GMPLS-TE Label Switched Path.

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

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

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

TE Link - The term TE link refers to single or a bundle of
physical link(s) or FA-LSP(s) on which traffic engineering is
enabled [RFC4206], [RFC4201].

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 Traffic Engineering.

- 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 resource(s) 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 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 LSP setup 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 shutdown is a last resort, it can be
used. Time or decision for removal of the resource being shutdown
is based on a local decision at the node initiating the graceful
shutdown procedure.

- It is required to give the ingress node the opportunity to take
actions in order to reduce/eliminate traffic disruption on the
LSP(s) that are using the network resources which are about to be
shutdown.

- Graceful shutdown mechanisms are equally applicable to intra-
domain and TE LSPs spanning multiple domains. Here, a domain is
defined as either an IGP area or an Autonomous System [RFC4726].

- Graceful shutdown is equally applicable to GMPLS-TE, as well as
packet-based (MPLS) TE LSPs.

- In order to make rerouting effective, it is required that when
  a node initiates the graceful shutdown of a resource, it
  identifies to all other network nodes the TE resource under
  graceful shutdown.
- Depending on switching technology, it may be possible to

shutdown 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], [RFC3784], [RFC4203] and [RFC4205] are not applicable when dealing with Inter-area and Inter-AS traffic engineering, as IGP LSA/LSP flooding is restricted to IGP areas/levels. Consequently, RSVP based mechanisms are required to cope with TE LSPs spanning multiple domains. At the same time, 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. Furthermore, graceful shutdown notification via IGP flooding is required to discourage a node from establishing new LSPs through the resources being shutdown. In the following sections the complementary mechanisms for RSVP-TE and IGP for Graceful Shutdown are described.

A node where a link or the whole node is being shutdown may first trigger the IGP updates as described in Section 4.1, introduce a delay to allow network convergence and only then use the signaling mechanism described in Section 4.2.

4.1 OSPF/ ISIS Mechanisms for graceful shutdown

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

OSPF and ISIS procedure for graceful shutdown of TE link(s) is similar to graceful restart of OSPF and ISIS as described in [RFC4203] and [RFC4205], respectively. Specifically, the node where graceful-shutdown of a link is desired originates the TE LSA/LSP containing Link TLV for the link under graceful shutdown with Traffic Engineering metric set to 0xffffffff, 0 as unreserved bandwidth, and if the link has LSC or FSC as its Switching Capability then also with 0 as Max LSP Bandwidth. A node may also specify a value for Minimum LSP bandwidth which is greater than the available bandwidth. This would discourage new LSP establishment through the link under graceful shutdown.

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. Similarly, If 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

shutdown, the routing procedure related to TE link removal is
used.

Neighbors of the node where graceful shutdown procedure is in
progress continues 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 uses LSP rerouting mechanism as defined in
[LSP-REROUTE]. 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 "Reroute" and an
error value of "TE link Graceful Shutdown required" for all
affected LSPs. Similarly, the node that is being gracefully
shutdown triggers a PathErr message with the error code "Reroute"
and an error value of "Node Graceful Shutdown required" for all
LSPs.

MPLS TE Link Bundling [RFC4201] requires that an 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 TE bundled link is being shutdown, the
remaining component links of the bundled TE link may still be
able to admit new LSPs. The node where graceful shutdown of a
component link is desired triggers a PathErr message with the
error code "Reroute" and the new error value of "Component link
Graceful Shutdown required" for all affected LSPs. The PathErr
message includes in the ERROR_SPEC the TE Link ID address. If the
last component link is being shutdown, procedure for gracefully
shutdown entire bundled TE link outlined above is be used,
instead.

If graceful shutdown of a label resource is desired, the node
initiating this action triggers a PathErr message with the error
code "Reroute" and the new error value of "Label resource
Graceful Shutdown required" for the affected LSP. The PathErr
message includes in the ERROR_SPEC the TE Link ID address.

The "Reroute" error code for the ERROR SPEC object is defined in

[LSP-REROUTE]. This document defines following four error value
for the "Reroute" error code [To Be Confirmed (TBC) by IANA upon
publication of this document]:

Error-value	Meaning	Reference
2 (TBC)	Node Graceful Shutdown required	This doc
3 (TBC)	TE link Graceful Shutdown required	This doc
4 (TBC)	Component link Graceful Shutdown required	This doc
5 (TBC)	Label resource Graceful Shutdown required	This doc

The PathErr message includes in the ERROR_SPEC the TE Link ID
address.

If unbundled TE link, component link of a bundled TE link, entire
bundled TE link, or label resource of a TE link is being
gracefully shutdown, the PathErr message includes the ERROR_SPEC
object containing IP address of the TE Link being gracefully
shutdown. If TE link is unnumbered, the PathErr message includes
the ERROR_SPEC object containing unnumbered ID and TE node ID for
the TE Link being gracefully shutdown. Similarly, if the TE node
is being gracefully shutdown, the PathErr message includes in the
ERROR_SPEC object the MPLS-TE node ID address.

When a head-end node, or border node receives a PathErr message
with "Reroute" error code and error value of "Node Graceful
Shutdown required" or "TE link Graceful Shutdown required", or
"Component link Graceful Shutdown required", or "Label resource
Graceful Shutdown required" it follows the procedures defined in
[LSP-REROUTE]. When performing path computation for the new LSP,
the head-end node, or border node avoids using the TE resources
identified by the IP address contained in the PathErr. If PCE is
used for path computation, head-end node or border node acts as
PCC to request the PCE via PCEP for path computation avoiding
resource being gracefully shutdown. The amount of time the head-
end node, or border node avoid using the TE resources identified
by the IP address contained in the PathErr is based on a local
decision at head-end node or border node.

If node initiating the graceful shutdown procedure received path
setup request for a new tunnel using resource being gracefully
shutdown, it sends a Path Error message with "Reroute" error code
in the ERROR SPEC object and an error value consistent with the
type of resource being gracefully shutdown. However, based on a
local decision, if node initiating the graceful shutdown
procedure received path setup request for an existing tunnel, it
may allow signaling for it. This is to allow resource being
gracefully shutdown as a "last resort". The node initiating the
graceful shutdown procedure can distinguish between new and
existing tunnels based on the tunnel ID in the SESSION object.

Time or decision for removal of the resource being shutdown from
forwarding is based on a local decision at the node initiating

the graceful shutdown procedure.

5. Security Considerations

This document introduces two new error values for "Reroute" error code of the ERROR SPEC object defined in [LSP-REROUTE]. This document also introduces ways to make resources unavailable for the control plane. It is therefore recommended that procedures in [RFC2747], which provides mechanisms to protect against external agents compromising the RSVP signaling state in an RSVP agent, be used. Specifically, [RFC2747] mechanisms provide some degree of protection to the head-end node or border node RSVP agent against making resources unavailable for control plan from an external agent sending Path Error messages with existing or new error code and error values. In summary, existing security considerations specified in [LSP-REROUTE], [RFC2747], [RFC2205], [RFC3209], [RFC4736], [RFC3471], [RFC3473] and [MPLS-GMPLS-SECURITY] remain relevant and suffice.

This document relies on existing procedures for advertisement of TE LSA/LSP containing Link TLV. Tampering with TE LSAs may have an effect on traffic engineering computations, and it is suggested that any mechanisms used for securing the transmission of normal OSPF LSAs/ ISIS LSPs be applied equally to all Opaque LSAs/ LSPs this document uses. In summary, existing security considerations specified in [RFC3630], [RFC3784], [RFC4203], [RFC4205] and [MPLS-GMPLS-SECURITY] remain relevant and suffice.

6. IANA Considerations

The "Reroute" error code for the ERROR SPEC object is defined in [LSP-REROUTE]. This document defines following four error value for the "Reroute" error code [To Be Confirmed (TBC) by IANA upon publication of this document]:

Error-value	Meaning	Reference
2 (TBC)	Node Graceful Shutdown required	This doc
3 (TBC)	TE link Graceful Shutdown required	This doc
4 (TBC)	Component link Graceful Shutdown required	This doc
5 (TBC)	Label resource Graceful Shutdown required	This doc

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. Reference

8.1 Normative Reference

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

[RFC4736] Jean-Philippe Vasseur, et al "Reoptimization of MPLS Traffic Engineering loosely routed LSP paths", RFC 4736, November 2006.

[LSP-REROUTE] Berger, L., Papadimitriou, D., and J. Vasseur, "PathErr Message Triggered MPLS and GMPLS LSP Reroute", draft-ietf-mpls-gmpls-lsp-reroute-01 (work in progress) September 2008.

8.2 Informative Reference

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

[RFC3784] Smit, H. and T. Li, "Intermediate System to Intermediate System (IS-IS) Extensions for Traffic Engineering (TE)", RFC 3784, June 2004.

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

[RFC4205] Kompella, K., Ed., and Y. Rekhter, Ed., "Intermediate System to Intermediate System (IS-IS) Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4205, October 2005.

[RFC2205] Braden, R. Ed. et al, "Resource ReSerVation Protocol (RSVP) Version 1, Functional Specification", RFC 2205, December 1997.

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

[RFC3473] Berger, L., "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., Ayyangar A., "A Framework for Inter-Domain MPLS Traffic Engineering", RFC 4726, November 2006.

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

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

[RFC2747] Baker, F., Lindell, B., and M. Talwar, "RSVP
Cryptographic Authentication", RFC 2747, January 2000.
[MPLS-GMPLS-SECURITY] Fang, L. et al, "Security Framework for
MPLS and GMPLS Networks", draft-fang-mpls-gmpls-security-
framework-01.txt, work in progress.

9. Authors' Address:

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
jonathan.newton@cw.com

10. Intellectual Property Considerations

The IETF takes no position regarding the validity or scope of any
Intellectual Property Rights or other rights that might be
claimed to pertain to the implementation or use of the technology
described in this document or the extent to which any license
under such rights might or might not be available; nor does it
represent that it has made any independent effort to identify any
such rights. Information on the procedures with respect to
rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any
assurances of licenses to be made available, or the result of an
attempt made to obtain a general license or permission for the

use of such proprietary rights by implementers or users of this
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.

11. Disclaimer of Validity

This document and the information contained herein are provided
on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE
REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE
IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL
WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY
WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE
ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR
FITNESS FOR A PARTICULAR PURPOSE.

12. Copyright Statement