

Open Shortest Path First IGP
Internet-Draft
Intended status: Standards Track
Expires: January 8, 2017

S. Hegde
Juniper Networks, Inc.
P. Sarkar
H. Gredler
Individual
M. Nanduri
Microsoft Corporation
L. Jalil
Verizon
July 7, 2016

OSPF Link Overload
draft-ietf-ospf-link-overload-02

Abstract

When a link is being prepared to be taken out of service, the traffic needs to be diverted from both ends of the link. Increasing the metric to the highest metric on one side of the link is not sufficient to divert the traffic flowing in the other direction.

It is useful for routers in an OSPFv2 or OSPFv3 routing domain to be able to advertise a link being in an overload state to indicate impending maintenance activity on the link. This information can be used by the network devices to re-route the traffic effectively.

This document describes the protocol extensions to disseminate link overload information in OSPFv2 and OSPFv3.

Requirements Language

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 [RFC2119].

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any

time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on January 8, 2017.

Copyright Notice

Copyright (c) 2016 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	3
2. Motivation	3
3. Link overload sub-TLV	4
3.1. OSPF Link overload sub-TLV	4
4. Flooding Scope	4
4.1. Area scope flooding	4
4.2. Link scope flooding	5
5. Elements of procedure	5
5.1. Point-to-point links	5
5.2. Broadcast/NBMA links	6
5.3. Point-to-multipoint links	6
5.4. Unnumbered interfaces	6
6. Backward compatibility	7
7. Applications	7
7.1. Pseudowire Services	7
7.2. Controller based Traffic Engineering Deployments	8
8. Security Considerations	8
9. IANA Considerations	9
10. Acknowledgements	9
11. References	9
11.1. Normative References	9
11.2. Informative References	10
Authors' Addresses	10

1. Introduction

When a node is being prepared for a planned maintenance or upgrade, [RFC6987] provides mechanisms to advertise the node being in an overload state by setting all outgoing link costs to MAX-METRIC (0xffff). These procedures are specific to the maintenance activity on a node and cannot be used when a single link attached to the node requires maintenance.

In traffic-engineering deployments, LSPs need to be moved away from the link without disrupting the services. It is useful to be able to advertise the impending maintenance activity on the link and to have LSP re-routing policies at the ingress to route the LSPs away from the link.

Many OSPFv2 or OSPFv3 deployments run on overlay networks provisioned by means of pseudo-wires or L2-circuits. When the devices in the underlying network go for maintenance, it is useful to divert the traffic away from the node before the maintenance is actually scheduled. Since the nodes in the underlying network are not visible to OSPF, the existing stub router mechanism described in [RFC6987] cannot be used. Application specific to this use case is described in Section 7.1

This document provides mechanisms to advertise link overload state in the flexible encodings provided by RI LSA([RFC7770]) for OSPFv2 and OSPFv3. Throughout this document, OSPF is used when the text applies to both OSPFv2 and OSPFv3. OSPFv2 or OSPFv3 is used when the text is specific to one version of the OSPF protocol.

2. Motivation

The motivation of this document is to reduce manual intervention during maintenance activities. The following objectives help to accomplish this in a range of deployment scenarios.

1. Advertise impending maintenance activity so that the traffic from both directions can be diverted away from the link.
2. Allow the solution to be backward compatible so that nodes that do not understand the new advertisement do not cause routing loops.
3. Advertise the maintenance activity to other nodes in the network so that LSP ingress routers/controllers can learn the impending maintenance activity and apply specific policies to re-route the LSP for traffic-engineering based deployments.

For OSPFv3, Link-overload sub-tlv is carried in Link TLV as defined in [RFC5329] Link TLV is carried in the OSPFv3 area scoped RI-LSA [RFC7770]

4.2. Link scope flooding

The link local scope RI-LSA corresponds to the link on which the LSA arrives and there is no need to explicitly carry the link TLV. The Link overload sub-TLV is carried in the RI-LSA for both OSPFv2 and OSPFv3.

5. Elements of procedure

The Link Overload sub-TLV indicates that the link identified in which it is carried is overloaded. The node that has the link to be taken out of service SHOULD originate the Link Overload sub-TLV in the Link TLV in the RI LSA as defined in [RFC7770]. The link-overload information is carried as a property of the link and is flooded across the area. This information can be used by ingress routers or controllers to take special actions. Application specific to this use case is described in Section 7.2.

The precise action taken by the remote node at the other end of the link identified as overloaded depends on the link type.

5.1. Point-to-point links

The node that has the link to be taken out of service SHOULD set metric of the link to MAX-METRIC (0xffff) and re-originate the Router-LSA. The TE metric SHOULD be set to MAX-TE-METRIC-1 (0xffffffe) and the node SHOULD re-originate the TE Link Opaque LSAs. When a Link Overload sub-TLV is received for a point-to-point link either by link local or area scoped RI-LSA, the remote node SHOULD identify the local link which corresponds to the overloaded link and set the metric to MAX-METRIC (0xffff). The remote node MUST re-originate the router-LSA with the changed metric and flood into the OSPF area. The TE metric SHOULD be set to MAX-TE-METRIC-1 (0xffffffe) and the TE opaque LSA for the link MUST be re-originated with new value.

In multi-topology deployments [RFC4915], the Link overload Sub-TLV carried in an RI LSA corresponds to all the topologies the link belongs to. The receiver node SHOULD change the metric in the reverse direction corresponding to all the topologies to which the reverse link belongs.

When the originator of the Link Overload sub-TLV purges the RI-LSA or re-originates it without the Link Overload sub-TLV, the remote node

must re-originate the appropriate LSAs with the metric and TE metric values set to their original values.

5.2. Broadcast/NBMA links

Broadcast or NBMA networks in OSPF are represented by a star topology where the Designated Router (DR) is the central point to which all other routers on the broadcast or NBMA network connect logically. As a result, routers on the broadcast or NBMA network advertise only their adjacency to the DR. Routers that do not act as DR do not form or advertise adjacencies with each other. For the Broadcast links, the MAX-METRIC on the remote link cannot be changed since all the neighbours are on same link. Setting the link cost to MAX-METRIC would impact paths going via all neighbours.

The node that has the link to be taken out of service SHOULD set metric of the link to MAX-METRIC (0xffff) and re-originate the Router-LSA. The TE metric SHOULD be set to MAX-TE-METRIC-1(0xffffffe) and the node SHOULD re-originate the TE Link Opaque LSAs. For a broadcast link, the two part metric as described in [I-D.ietf-ospf-two-part-metric] is used. The node originating the Link Overload sub-TLV MUST set the metric in the Network-to-Router Metric sub-TLV to MAX-METRIC 0xffff for OSPFv2 and OSPFv3 and re-originate the LSAs the TLV is carried-in.

The nodes that receive the two part metric should follow the procedures described in [I-D.ietf-ospf-two-part-metric]. The backward compatibility procedures described in [I-D.ietf-ospf-two-part-metric] should be followed to ensure loop free routing.

5.3. Point-to-multipoint links

Operation for the point-to-multipoint links is similar to the point-to-point links. When a Link Overload sub-TLV is received for a point-to-multipoint link the remote node SHOULD identify the link which corresponds to the overloaded link and set the metric to MAX-METRIC (0xffff). The remote node MUST re-originate the Router-LSA with the changed metric and flood into the OSPF area.

5.4. Unnumbered interfaces

Unnumbered interface do not have a unique IP addresses and borrow address from other interfaces. The Link TLV carries the local and remote interface ids to uniquely identify the link when there are more than one parallel links between the nodes. Procedures to obtain interface-id of the remote side is defined in [RFC4203] and are

applicable to the Link TLV added in the RI LSA for the purpose of carrying the Link overload sub-tlv.

6. Backward compatibility

The mechanism described in the document is fully backward compatible. It is required that the originator of the Link Overload sub-TLV as well as the node at the remote end of the link identified as overloaded understand the extensions defined in this document. In the case of broadcast links, the backward compatibility procedures as described in [I-D.ietf-ospf-two-part-metric] are applicable. .

7. Applications

7.1. Pseudowire Services

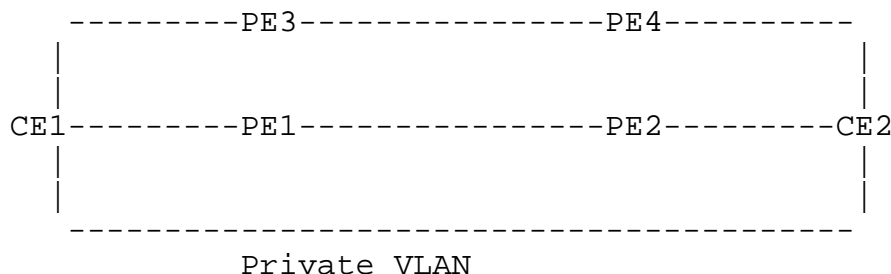


Figure 2: Pseudowire Services

Many service providers offer pseudo-wire services to customers using L2 circuits. The IGP protocol that runs in the customer network would also run over the pseudo-wire to create seamless private network for the customer. Service providers want to offer overload kind of functionality when the PE device is taken-out for maintenance. The provider should guarantee that the PE is taken out for maintenance only after the service is successfully diverted on an alternate path. There can be large number of customers attached to a PE node and the remote end-points for these pseudo-wires are spread across the service provider's network. It is a tedious and error-prone process to change the metric for all pseudo-wires in both directions. The link overload feature simplifies the process by increasing the metric on the link in the reverse direction as well so that traffic in both directions is diverted away from the PE undergoing maintenance. The link-overload feature allows the link to be used as a last resort link so that traffic is not disrupted when alternative paths are not available.

7.2. Controller based Traffic Engineering Deployments

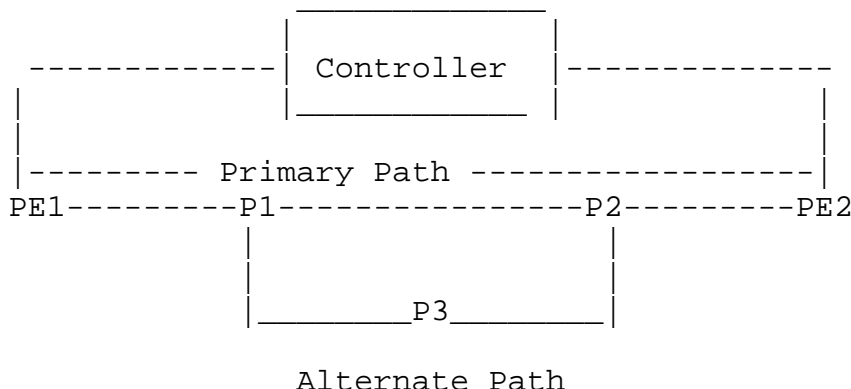


Figure 3: Controller based Traffic Engineering

In controller-based deployments where the controller participates in the IGP protocol, the controller can also receive the link-overload information as a warning that link maintenance is imminent. Using this information, the controller can find alternate paths for traffic which use the affected link. The controller can apply various policies and re-route the LSPs away from the link undergoing maintenance. If there are no alternate paths satisfying the traffic engineering constraints, the controller might temporarily relax those constraints and put the service on a different path.

In the above example, PE1->PE2 LSP is set-up which satisfies a constraint of 10 GB bandwidth on each link. The links P1->P3 and P3->P2 have only 1 GB capacity. and there is no alternate path satisfying the bandwidth constraint of 10GB. When P1->P2 link is being prepared for maintenance, the controller receives the link-overload information, as there is no alternate path available which satisfies the constraints, controller chooses a path that is less optimal and sets up an alternate path via P1->P3->P2 temporarily. Once the traffic is diverted, P1->P2 link can be taken out for maintenance/upgrade.

8. Security Considerations

This document does not introduce any further security issues other than those discussed in [RFC2328] and [RFC5340].

9. IANA Considerations

This specification updates one OSPF registry:

OSPF Link TLVs Registry

i) TBD - Link Overload sub TLV

OSPFV3 Link TLV Registry

i) TBD - Link Overload sub TLV

OSPF Router Information (RI)TLVs Registry

i) TBD - Link TLV

10. Acknowledgements

Thanks to Chris Bowers for valuable inputs and edits to the document.
Thanks to Jeffrey Zhang and Acee Lindem for inputs.

11. References

11.1. Normative References

[I-D.ietf-ospf-two-part-metric]

Wang, L., Lindem, A., DuBois, D., Julka, V., and T. McMillan, "OSPF Two-part Metric", draft-ietf-ospf-two-part-metric-01 (work in progress), July 2015.

[RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, DOI 10.17487/RFC3630, September 2003, <<http://www.rfc-editor.org/info/rfc3630>>.

[RFC5329] Ishiguro, K., Manral, V., Davey, A., and A. Lindem, Ed., "Traffic Engineering Extensions to OSPF Version 3", RFC 5329, DOI 10.17487/RFC5329, September 2008, <<http://www.rfc-editor.org/info/rfc5329>>.

[RFC7770] Lindem, A., Ed., Shen, N., Vasseur, JP., Aggarwal, R., and S. Shaffer, "Extensions to OSPF for Advertising Optional Router Capabilities", RFC 7770, DOI 10.17487/RFC7770, February 2016, <<http://www.rfc-editor.org/info/rfc7770>>.

11.2. Informative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<http://www.rfc-editor.org/info/rfc2119>>.
- [RFC2328] Moy, J., "OSPF Version 2", STD 54, RFC 2328, DOI 10.17487/RFC2328, April 1998, <<http://www.rfc-editor.org/info/rfc2328>>.
- [RFC4203] Kompella, K., Ed. and Y. Rekhter, Ed., "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, DOI 10.17487/RFC4203, October 2005, <<http://www.rfc-editor.org/info/rfc4203>>.
- [RFC4915] Psenak, P., Mirtorabi, S., Roy, A., Nguyen, L., and P. Pillay-Esnault, "Multi-Topology (MT) Routing in OSPF", RFC 4915, DOI 10.17487/RFC4915, June 2007, <<http://www.rfc-editor.org/info/rfc4915>>.
- [RFC5340] Coltun, R., Ferguson, D., Moy, J., and A. Lindem, "OSPF for IPv6", RFC 5340, DOI 10.17487/RFC5340, July 2008, <<http://www.rfc-editor.org/info/rfc5340>>.
- [RFC6987] Retana, A., Nguyen, L., Zinin, A., White, R., and D. McPherson, "OSPF Stub Router Advertisement", RFC 6987, DOI 10.17487/RFC6987, September 2013, <<http://www.rfc-editor.org/info/rfc6987>>.

Authors' Addresses

Shraddha Hegde
Juniper Networks, Inc.
Embassy Business Park
Bangalore, KA 560093
India

Email: shraddha@juniper.net

Pushpasis Sarkar
Individual

Email: pushpasis.ietf@gmail.com

Hannes Gredler
Individual

Email: hannes@gredler.at

Mohan Nanduri
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052
US

Email: mmanduri@microsoft.com

Luay Jalil
Verizon

Email: luay.jalil@verizon.com