Open Shortest Path First IGP
Internet-Draft
Intended status: Standards Track
Expires: April 21, 2016

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

October 19, 2015

OSPF Link Overload
draft-ietf-ospf-link-overload-00

Abstract

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 [RFC3137] cannot be used.

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 in the underlying network devices. 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

Copyright Notice

Table of Contents

1. Introduction

When a node is being prepared for a planned maintenance or upgrade, [RFC3137] 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.

When a link is being prepared to be taken out of service, the traffic needs to be diverted from both ends of the link. Changing the metric on one side of the link is not sufficient to divert the traffic flowing in both directions. 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 rerouting policies at the ingress to route the LSPs away from the link.

It is useful for routers in 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 document provides mechanisms to advertise link overload state in the flexible encodings provided by OSPFv2 Prefix/Link Attribute Advertisement([I-D.ietf-ospf-prefix-link-attr]) and OSPFv3 Extended LSA ([I-D.ietf-ospf-ospfv3-lsa-extend]). 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.

4. Allow the link to be used as last resort link to prevent traffic disruption when alternate paths are not available.

3. Link overload sub-TLV

3.1. OSPFv2 Link overload sub-TLV

The Link Overload sub-TLV is carried as part of the Extended Link TLV as defined in [I-D.ietf-ospf-prefix-link-attr] for OSPFv2.

```
        0                   1                   2                   3
        0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |              Type             |             Length            |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |                     Remote IP address                         |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: Link Overload sub-TLV for OSPFv2

Type : TBA

Length: 4

Value: Remote IPv4 address. The remote IP4 address is used to identify the particular link that is in the overload state when there are multiple parallel links between two nodes.

3.2. OSPFv3 Link Overload sub-TLV

The Link Overload sub-TLV is carried in the Router-Link TLV as defined in the [I-D.ietf-ospf-ospfv3-lsa-extend] for OSPFv3. The Router-Link TLV contains the neighbor interface-id and can uniquely identify the link on the remote node.

```
        0                   1                   2                   3
        0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |              Type             |             Length            |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: Link Overload sub-TLV for OSPFv3

Type : TBA

Length: 0

4. Elements of procedure

The Link Overload sub-TLV indicates that the link identified in by the sub-TLV is overloaded. The node that has the link to be taken out of service sets metric of the link to MAX-METRIC (0xffff) and re-originates the Router-LSA. The TE metric is set to MAX-TE-METRIC-1 (0xfffffffe) and the node also re-originates the TE Link Opaque LSAs. The node SHOULD originate the Link Overload sub-TLV in the Extended Link TLV in the Extended Link Opaque LSA as defined in [I-D.ietf-ospf-prefix-link-attr] for OSPFv2 or in the E-Router-LSA as defined in [I-D.ietf-ospf-ospfv3-lsa-extend] for OSPFv3. This LSA should be flooded in the OSPF area. A node which supports this draft and is at the remote end of the link identified in the Link Overload sub-TLV MUST set the metric of the link in the reverse direction to MAX-METRIC. In addition, the TE metric MUST be changed to 0xfffffffe. The remote node MUST re-originate the Router-LSA and TE link opaque LSA with these updated metrics, and flood them into the area.

When the originator of the Link Overload sub-TLV purges the Extended Link Opaque 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.

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

4.1. Point-to-point links

When a Link Overload sub-TLV is received for a point-to-point link 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 (0xfffffffe) and the TE opaque LSA for the link MUST be re-originated with new value.

4.2. 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 outgoing link cannot be changed since all the neighbors are on same link. Setting the link cost to MAX-METRIC would impact paths going via all neighbors.

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 MT metric in the Network-to-Router Metric sub-TLV to MAX-METRIC 0xffff for OSPFv2 and OSPFv3. 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.

4.3. 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 neighbor 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.

4.4. Unnumbered interfaces

Unnumbered interface do not have a unique IP addresses and borrow address from other interfaces. [RFC2328] describes procedures to handle unnumbered interfaces. The link-data field in the Extended Link TLV carries the interface-id instead of the IP address. The Link Overload sub-TLV carries the remote interface-id in the Remote-ip-address field if the interface is unnumbered. Procedures to obtain interface-id of the remote side is defined in [RFC4203].

5. 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. .

6. Applications

6.1. Pseudowire Services

```
           ---------PE3----------------PE4----------
          |                                         |
          |                                         |
        CE1---------PE1----------------PE2---------CE2
          |                                         |
          |                                         |
           -----------------------------------------
                    Private VLAN
```

Figure 3: 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.

6.2. Controller based Traffic Engineering Deployments

```
                         _____________
                        |             |
           -------------| Controller  |--------------
          |             |_____________ |             |
          |                                          |
          |  ------- Primary Path ---------------    |
          PE1---------P1----------------P2---------PE2
                      |                  |
                      |                  |
                      |________P3________|

                         Alternate Path
```

Figure 4: 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 using 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 when P1->P2 link is being prepared for maintenance, the controller receives the link-overload information and sets up an alternate path via P1->P3->P2. Once the traffic is diverted, P1->P2 link can be taken out for maintenance/upgrade.

7. Security Considerations

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

8. IANA Considerations

This specification updates one OSPF registry:

OSPF Extended Link TLVs Registry

i) TBD - Link Overload sub TLV

OSPFV3 Router Link TLV Registry

i) TBD - Link Overload sub TLV

9. Acknowledgements

Thanks to Chris Bowers for valuable inputs and edits to the document.

10. References

10.1. Normative References

[I-D.ietf-ospf-ospfv3-lsa-extend]
Lindem, A., Mirtorabi, S., Roy, A., and F. Baker, "OSPFv3 LSA Extendibility", draft-ietf-ospf-ospfv3-lsa-extend-06 (work in progress), February 2015.

[I-D.ietf-ospf-prefix-link-attr]
Psenak, P., Gredler, H., Shakir, R., Henderickx, W., Tantsura, J., and A. Lindem, "OSPFv2 Prefix/Link Attribute Advertisement", draft-ietf-ospf-prefix-link-attr-03 (work in progress), February 2015.

[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.

10.2. 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>.

[RFC3137] Retana, A., Nguyen, L., White, R., Zinin, A., and D. McPherson, "OSPF Stub Router Advertisement", RFC 3137, DOI 10.17487/RFC3137, June 2001, <http://www.rfc-editor.org/info/rfc3137>.

[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>.

[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>.

Authors' Addresses

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

Email: shraddha@juniper.net

Pushpasis Sarkar
Juniper Networks, Inc.
Embassy Business Park
Bangalore, KA 560093
India

Email: psarkar@juniper.net

Hannes Gredler
Juniper Networks, Inc.
1194 N. Mathilda Ave.
Sunnyvale, CA 94089
US

Email: hannes@juniper.net

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

Email: mnanduri@microsoft.com

Luay Jalil
Verizon

Email: luay.jalil@verizon.com