Link State Routing
Internet-Draft
Intended status: Standards Track
Expires: February 21, 2020

K. Talaulikar
P. Psenak
Cisco Systems, Inc.
H. Johnston
AT&T Labs

August 20, 2019

OSPF Reverse Metric
draft-ketant-lsr-ospf-reverse-metric-02

Abstract

This document specifies the extensions to OSPF that enables a router to signal to its neighbor the metric that the neighbor should use towards itself using link-local advertisement between them. The signalling of this reverse metric, to be used on link(s) towards itself, allows a router to influence the amount of traffic flowing towards itself and in certain use-cases enables routers to maintain symmetric metric on both sides of a link between them.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction 2
2. Use Cases 3
2.1. Symmetrical Metric Based on Reference Bandwidth 3
2.2. Adaptive Metric Signaling 4
3. Solution 5
4. LLS Reverse Metric TLV 6
5. LLS Reverse TE Metric TLV 6
6. Procedures 7
7. Backward Compatibility 9
8. IANA Considerations 9
9. Security Considerations 9
10. Contributors 10
11. Acknowledgements 10
12. References 10
12.1. Normative References 10
12.2. Informative References 11
Authors' Addresses 11

1. Introduction

Routers running the Open Shortest Path First (OSPFv2) [RFC2328] and OSPFv3 [RFC5340] routing protocols originate a Router-LSA (Link State Advertisement) that describes all its links to its neighbors and includes a metric which indicates its "cost" of reaching the neighbor over that link. Consider two routers R1 and R2 that are connected via a link. The metric for this link in direction R1->R2 is configured on R1 and in the direction R2->R1 is configured on R2. Thus the configuration on R1 influences the traffic that it forwards towards R2 but does not influence the traffic that it may receive from R2 on that same link.

This document describes certain use-cases where it is desirable for a router to be able to signal what we call as the "reverse metric" (RM) to its neighbor to adjust the routing metric on the inbound direction. When R1 signals its reverse metric on its link to R2, then R2 advertises this value as its metric to R1 in its Router-LSA instead of its locally configured value. Once this information is part of the topology then all other routers do their computation using this value which results in the desired change in traffic distribution that R1 wanted to achieve towards itself over the link from R2.

This document proposes an extension to OSPF link-local signaling (LLS) [RFC5613] for signalling the OSPF reverse metric using the LLS Reverse Metric TLV in Section 4, the reverse Traffic Engineering (TE) metric [RFC3630] using the LLS Reverse TE Metric TLV in Section 5 and describes the related procedures in section Section 6.

2. Use Cases

This section describes certain use-cases that OSPF reverse metric helps to address. The usage of OSPF reverse metric need not be limited to these cases and is intended to be a generic mechanism.

2.1. Symmetrical Metric Based on Reference Bandwidth

Certain OSPF implementations and deployments deduce the metric of links based on their bandwidth using a reference bandwidth. The OSPF MIB [RFC4750] has ospfReferenceBandwidth that is used by entries in the ospfIfMetricTable. This mechanism is leveraged in deployments where the link metrics get lowered or increased as bandwidth capacity is removed or added e.g. consider layer-2 links bundled as a layer-3 interface on which OSPF is enabled. In the situations where these layer-2 links are directly connected to the two routers, the link and bandwidth availability is detected and updated on both sides. This allows for schemes where the metric is maintained to be symmetric in both directions based on the bandwidth.

Now consider variation of the same deployment where the links between routers are not directly connected and instead are provisioned over a layer-2 network consisting of switches or other mechanisms for a layer-2 emulation. In such scenarios, as show in Figure 1, the router on one side of the link would not detect when the neighboring router has lost one of its layer-2 link and has reduced capacity to its layer-2 switch. Note that the number of links and their capacities on the router R0 may not be the same as those on R1, R2 and R3. The left hand side diagram shows the actual physical topology in terms of the layer-2 links while the right hand side diagram shows the logical layer-3 link topology between the routers.

```
                +--------+
                |   R0   |
                | Router |
                +--------+                       +--------+
    (a) Physical   ^ ^ ^           (b) Layer-3   |   R0   |
        Topology   | | |              Topology   +--------+
                   v v v                           ^ ^ ^
             +----------------+                    | | |
             | Layer 2 Switch |                    | | |
             |  (Aggregation) |                +---+ | +---+
             +----------------+                |     |     |
              ^^  ^ ^ ^ ^   ^                  v     |     v
              ||  | | | |   |              +------+  |  +------+
         +----+|  | | | |   |              |  R1  |  |  |  R3  |
         | +---+  | | | |   +----+         +------+  |  +------+
         v v      v v v v        v                   v
    +--------+  +--------+  +--------+           +--------+
    |   R1   |  |   R2   |  |   R3   |           |   R2   |
    | Router |  | Router |  | Router |           +--------+
    +-- -----+  +--------+  +--------+
```

Figure 1: Routers Interconnected over Layer-2 Network

In such a scenario, the amount of traffic that can be forwarded in bidirectional manner between say R0 and R1 is dictated by the lower of the link capacity of R0 and R1 to the layer-2 transport network. In this scenario, when one of the link from R1 to the switch goes down, it would increase its link metric to R0 from say 20 to 40. However, similarly R0 also needs to increase its link metric to R1 as well from 20 to 40 as otherwise, the traffic will hit congestion and get dropped.

When R1 has the ability to signal the OSPF reverse metric of 40 towards itself to R0, then R0 can also update its metric without any manual intervention to ensure the correct traffic distribution. Consider if some destinations were reachable from R0 via R1 previously and this automatic metric adjustment now makes some of those destinations reachable from R0 via R3. This allows some traffic load on the link R0 to R1 to now flow via R3 to these destinations.

2.2. Adaptive Metric Signaling

Now consider another deployment scenario where, as show in Figure 2, two routers AGGR1 and AGGR2 are connected to a bunch of routers R1 thru RN that are dual homed to them and aggregating the traffic from them towards a core network. At some point T, AGGR1 loses some of its capacity towards the core or is facing some congestion issue towards the core and it needs to reduce the traffic going through it and perhaps redirect some of that load via AGGR2 which is not facing a similar issue. Altering its own metric towards Rx routers would influence the traffic flowing through it in the direction from core to the Rx but not the other way around as desired.

```
              Core Network
          ^                ^
          |                |
          V                v
     +----------+    +----------+
     |  AGGR1   |    |  AGGR2   |
     +----------+    +----------+
       ^      ^        ^      ^
       |      |        |      |
       |      +-----------+   |
       |               |  |   |
       |      +--------+  |   |
       v      v           v   v
    +-----------+      +-----------+
    |    R1     |      |    RN     |
    |  Router   | ...  |  Router   |
    +-----------+      +-----------+
```

Figure 2: Adaptive Metric for Dual Gateways

In such a scenario, the AGGR1 router could signal an incremental value of OSPF reverse metric towards some or all of the Rx routers. When the Rx routers apply this signaled reverse metric offset value to the original metric on their links towards AGGR1 then the path via AGGR2 becomes a better path causing traffic towards the core getting diverted away from it. Note that the reverse metric mechanism allows such adaptive metric changes to be applied on the AGGR1 as opposed to being provisioning statically on the possibly large number of Rx routers.

3. Solution

To address the use-cases described earlier and to allow an OSPF router to indicate its reverse metric for a specific point-to-point or point-to-multipoint link to its neighbor, this document proposes to extend OSPF link-local signaling to advertise the Reverse Metric TLV in OSPF Hello packets. This ensures that the RM signaling is scoped ONLY to each specific link individually. The router continues to include the Reverse Metric TLV in its Hello packets on the link as long as it needs its neighbor to use that metric value towards itself. Further details of the procedures involve are specified in Section 6.

The RM signaling specified in this document is not required for broadcast or non-broadcast-multi-access (NBMA) links since the same objective is achieved there using the OSPF Two-Part Metric mechanism [RFC8042].

4. LLS Reverse Metric TLV

The Reverse Metric TLV is a new LLS TLV. It has following format:

```
    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            |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |     MTID      | Flags     |O|H|        Reverse Metric         |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

Type: TBD, suggested value 19

Length: 4 octet

MTID : the multi-topology identifier of the link ([RFC4915])

Flags: 1 octet, following are defined currently and the rest MUST be set to 0 and ignored on reception.

* H (0x1) : Indicates that neighbor should use value only if higher than its current metric value in use

* O (0x2) : Indicates that the reverse metric value provided is an offset that is to be added to the original metric

Reverse Metric: 2 octets, the value or offset of reverse metric to be used

5. LLS Reverse TE Metric TLV

The Reverse TE Metric TLV is a new LLS TLV. It has following format:

```
    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            |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |   Flags   |O|H|                 RESERVED                      |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                     Reverse TE Metric                         |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

Type: TBD, suggested value 20

Length: 4 octet

Flags: 1 octet, following are defined currently and the rest MUST be set to 0 and ignored on reception.

* H (0x1) : Indicates that neighbor should use value only if higher than its current TE metric value in use

* O (0x2) : Indicates that the reverse TE metric value provided is an offset that is to be added to the original TE metric

RESERVED: 24-bit field. SHOULD be set to 0 on transmission and MUST be ignored on receipt.

Reverse TE Metric: 4 octets, the value or offset of reverse traffic engineering metric to be used

6. Procedures

When a router needs to signal a RM value that its neigbhor(s) should use towards itself, it includes the Reverse Metric TLV in the LLS block of its hello messages sent on the link and continues to include this TLV for as long as it needs it's neighbor to use this value. The mechanisms used to determine the value to be used for the RM is specific to the implementation and use-case and is outside the scope of this document. e.g. in the use-case related to symmetric metric described in Section 2.1, the RM value may be derived based on the router's link's bandwidth with respect to the reference bandwidth.

A router receiving a hello packet from its neighbor that contains the Reverse Metric TLV on its link SHOULD use the RM value to derive the metric for the link in its Router-LSA to the advertising router.

When the O flag is set, the value in the TLV needs to be added to the existing original metric provisioned on the link to derive the new metric value to be used. When the O flag is clear, the value in the TLV should be directly used as the metric to be used. When H flag is set and O flag is clear, this is done only when the RM value signaled is higher than the provisioned metric value being used already. This mechanism applies only for point-to-point, point-to-multipoint and hybrid broadcast point-to-multipoint ([RFC6845]) links. For broadcast and NBMA links the OSPF Two-Part Metric mechanism [RFC8042] should be used in similar use-cases.

Implementations SHOULD provide a configuration option to enable the signaling of RM from a router to its neighbors and MAY provide a configuration option to disable the acceptance of the RM from its neighbors.

A router stops including the Reverse Metric TLV in its hello messages when it needs its neighbors to go back to using their own provisioned metric values. When that happens, a router which had modified its metric in response to receiving a Reverse Metric TLV from its neighbor should revert back to using its original provisioned metric value.

In certain scenarios, it is possible that two or more routers start the RM signaling on the same link. This could create collision scenarios. The following rules MUST be adopted by routers to ensure that there is no instability in the network due to churn in their metric due to signaling of RM:

- The RM value that is signaled by a router to its neighbor MUST NOT be derived from the reverse metric being signaled by any of its neighbor on any of its links.

- The RM value that is signaled by a router MUST NOT be derived from its own metric which has been modified on account of a RM signaled from any of its neighbors on any of its links. RM signaling from other routers can affect the router's own metric advertised in its Router-LSA. When deriving the RM values that a router signals to its neighbors, it should use its "original" local metric values not influenced by any RM signaling.

Based on these rules, a router MUST never start or stop or change its RM metric signaling based on the RM metric signaling initiated by some other router. Based on the local configuration policy, each router would end up accepting the RM value signaled by its neighbor and there would be no churn of metrics on the link or the network on account of RM signaling.

In certain use-case as described in Section 2.1 when symmetrical metrics are desired, the RM signaling can be enabled on routers on either ends of a link. In other use-cases as described in Section 2.2 RM signaling may need to be enabled on only router at one end of a link.

When using multi-topology routing with OSPF [RFC4915] a router MAY include multiple instances of the Reverse Metric TLV in the LLS block of its hello message - one for each of the topology for which it desires to signal the reserve metric for.

In certain scenarios, the OSPF router may also require the modification of the TE metric being advertised by its neighbor router towards itself in the inbound direction. The Reverse TE Metric TLV, using similar procedures as described above, MAY be used to signal the reverse TE metric by a router. The neighbor SHOULD use the reverse TE metric value to derive the TE metric to be used in the TE Metric sub-TLV of the Link TLV in its TE Opaque LSA [RFC3630].

7. Backward Compatibility

The signaling specified in this document happens at a link-local level between routers on that link. A router which does not support this specification would ignore the Reverse Metric and Reverse TE Metric LLS TLVs and take no actions to updates its metric in the other LSAs. As a result, the behavior would be the same as before this specification. Therefore, there are no backward compatibility related issues or considerations that need to be taken care of when implementing this specification.

8. IANA Considerations

This specification updates Link Local Signalling TLV Identifiers registry.

Following values are requested for allocation:

o TBD (Suggested value 19) - Reverse Metric TLV

o TBD (Suggested value 20) - Reverse TE Metric TLV

9. Security Considerations

The security considerations for "OSPF Link-Local Signaling" [RFC5613] also apply to the extension described in this document. The usage of the reverse metric TLVs is to alter the metrics used by routers on the link and influence the flow and routing of traffic over the network. Hence, modification of the Reverse Metric and Reverse TE

Metric TLVs may result in misrouting of traffic. If authentication is being used in the OSPF routing domain [RFC5709][RFC7474], then the Cryptographic Authentication TLV [RFC5613] SHOULD also be used to protect the contents of the LLS block.

Receiving a malformed LLS Reverse Metric or Reverse TE Metric TLVs MUST NOT result in a hard router or OSPF process failure. The reception of malformed LLS TLVs or sub-TLVs SHOULD be logged, but such logging MUST be rate- limited to prevent denial-of-service (DoS) attacks.

10. Contributors

Thanks to Jay Karthik for his contributions on the use-cases related to symmetric metric and the review of the solution.

11. Acknowledgements

12. References

12.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC2328] Moy, J., "OSPF Version 2", STD 54, RFC 2328, DOI 10.17487/RFC2328, April 1998, <https://www.rfc-editor.org/info/rfc2328>.

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

[RFC5340] Coltun, R., Ferguson, D., Moy, J., and A. Lindem, "OSPF for IPv6", RFC 5340, DOI 10.17487/RFC5340, July 2008, <https://www.rfc-editor.org/info/rfc5340>.

[RFC5613] Zinin, A., Roy, A., Nguyen, L., Friedman, B., and D. Yeung, "OSPF Link-Local Signaling", RFC 5613, DOI 10.17487/RFC5613, August 2009, <https://www.rfc-editor.org/info/rfc5613>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

12.2. Informative References

[RFC4750] Joyal, D., Ed., Galecki, P., Ed., Giacalone, S., Ed., Coltun, R., and F. Baker, "OSPF Version 2 Management Information Base", RFC 4750, DOI 10.17487/RFC4750, December 2006, <https://www.rfc-editor.org/info/rfc4750>.

[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, <https://www.rfc-editor.org/info/rfc4915>.

[RFC5709] Bhatia, M., Manral, V., Fanto, M., White, R., Barnes, M., Li, T., and R. Atkinson, "OSPFv2 HMAC-SHA Cryptographic Authentication", RFC 5709, DOI 10.17487/RFC5709, October 2009, <https://www.rfc-editor.org/info/rfc5709>.

[RFC6845] Sheth, N., Wang, L., and J. Zhang, "OSPF Hybrid Broadcast and Point-to-Multipoint Interface Type", RFC 6845, DOI 10.17487/RFC6845, January 2013, <https://www.rfc-editor.org/info/rfc6845>.

[RFC7474] Bhatia, M., Hartman, S., Zhang, D., and A. Lindem, Ed., "Security Extension for OSPFv2 When Using Manual Key Management", RFC 7474, DOI 10.17487/RFC7474, April 2015, <https://www.rfc-editor.org/info/rfc7474>.

[RFC8042] Zhang, Z., Wang, L., and A. Lindem, "OSPF Two-Part Metric", RFC 8042, DOI 10.17487/RFC8042, December 2016, <https://www.rfc-editor.org/info/rfc8042>.

Authors' Addresses

Ketan Talaulikar
Cisco Systems, Inc.
India

Email: ketant@cisco.com

Peter Psenak
Cisco Systems, Inc.
Apollo Business Center
Mlynske nivy 43
Bratislava 821 09
Slovakia

Email: ppsenak@cisco.com

Hugh Johnston
AT&T Labs
USA

Email: hugh_johnston@labs.att.com