Stream:	Internet Engineering Task Force (IETF)		
RFC:	9339		
Category:	Standards Track		
Published:	December 2022		
ISSN:	2070-1721		
Authors:	K. Talaulikar, Ed.	P. Psenak	H. Johnston
	Cisco Systems, Inc.	*Cisco Systems, Inc.*	*AT&T Labs*

RFC 9339
OSPF Reverse Metric

Abstract

This document specifies the extensions to OSPF that enable a router to use Link-Local Signaling (LLS) to signal the metric that receiving OSPF neighbor(s) should use for a link to the signaling router. When used on the link to the signaling router, the signaling of this reverse metric (RM) allows a router to influence the amount of traffic flowing towards itself. In certain use cases, this enables routers to maintain symmetric metrics on both sides of a link between them.

Status of This Memo

This is an Internet Standards Track document.

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). Further information on Internet Standards is available in Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at https://www.rfc-editor.org/info/rfc9339.

Table of Contents

1. Introduction

A router running the OSPFv2 [RFC2328] or OSPFv3 [RFC5340] routing protocols originates a Router-LSA (Link State Advertisement) that describes all its links to its neighbors and includes a metric that indicates its "cost" to reach the neighbor over that link. Consider two routers, R1 and R2, that are connected via a link. In the direction R1->R2, the metric for this link is configured on R1. In the direction R2->R1, the metric for this link 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 a router is required to signal what we call the "reverse metric" (RM) to its neighbor to adjust the routing metric in the inbound direction. When R1 signals its RM on its link to R2, 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, all other routers do their computation using this value. This may result in the desired change in the traffic distribution that R1 wanted to achieve towards itself over the link from R2.

This document describes extensions to OSPF LLS [RFC5613] to signal OSPF RMs. Section 4 specifies the LLS Reverse Metric TLV and Section 5 specifies the LLS Reverse TE Metric TLV. The related procedures are specified in Section 6.

1.1. 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.

2. Use Cases

This section describes certain use cases that are addressed by the OSPF RM. The usage of the OSPF RM need not be limited to these cases; it is intended to be a generic mechanism.

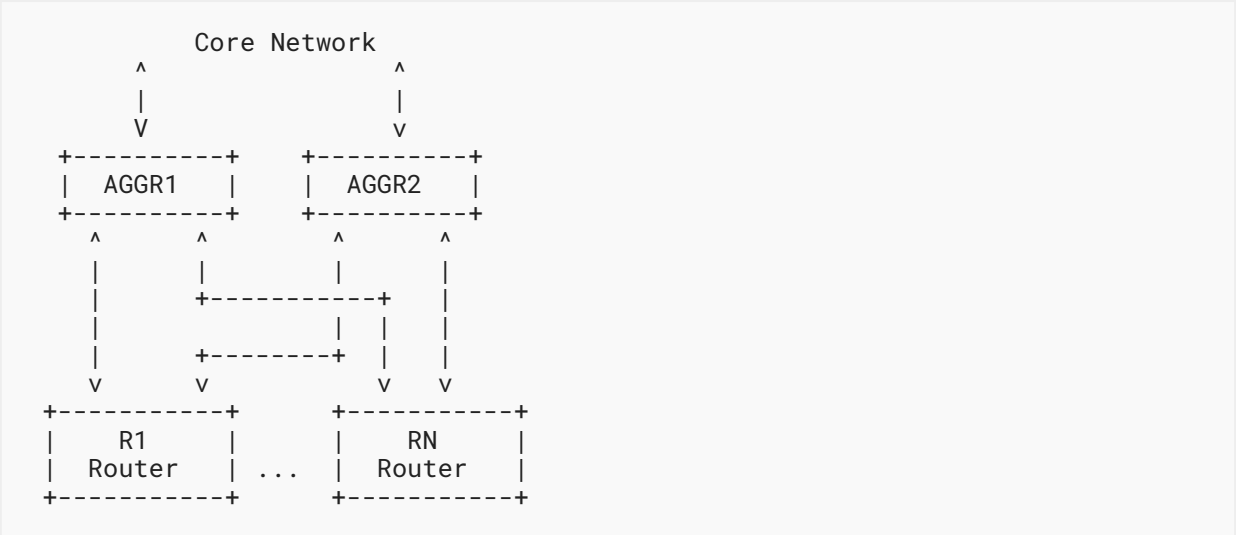

Figure 1: Reference Dual Hub-and-Spoke Topology

Consider a deployment scenario, such as the one shown in Figure 1, where routers R1 through RN are dual-home connected to AGGR1 and AGGR2 that are aggregating their traffic towards a core network.

2.1. Link Maintenance

Before network maintenance events are performed on individual links, operators substantially increase (to maximum value) the OSPF metric simultaneously on both routers attached to the same link. In doing so, the routers generate new Router LSAs that are flooded throughout the network and cause all routers to shift traffic onto alternate paths (where available) with limited disruption (depending on the network topology) to in-flight communications by applications or end users. When performed successfully, this allows the operator to perform disruptive augmentation, fault diagnosis, or repairs on a link in a production network.

In deployments such as a hub-and-spoke topology (as shown in Figure 1), it is quite common to have routers with several hundred interfaces and individual interfaces that move anywhere from several hundred gigabits to terabits per second of traffic. The challenge in such conditions is that the operator must accurately identify the same point-to-point (P2P) link on two separate devices to increase (and afterward decrease) the OSPF metric appropriately and to do so in a coordinated manner. When considering maintenance for PE-CE links when many Customer Edge (CE) routers connect to a Provider Edge (PE) router, an additional challenge related to coordinating access to the CE routers may arise when the CEs are not managed by the provider.

The OSPF RM mechanism helps address these challenges. The operator can set the link on one of the routers (generally the hub, like AGGR1 or a PE) to a "maintenance mode". This causes the router to advertise the maximum metric for that link and to signal its neighbor on the same link to advertise maximum metric via the reverse metric signaling mechanism. Once the link maintenance is completed and the "maintenance mode" is turned off, the router returns to using its provisioned metric for the link and stops the signaling of RM on that link, resulting in its neighbor also reverting to its provisioned metric for that link.

2.2. Adaptive Metric Signaling

In Figure 1, consider that at some point in time (T), AGGR1 loses some of its capacity towards the core. This may result in a congestion issue towards the core on AGGR1 that it needs to mitigate by redirecting some of its traffic load to transit via AGGR2, which is not experiencing a similar issue. Altering its link metric towards the R1-RN routers would influence the traffic from the core towards R1-RN, but not the other way around as desired.

In such a scenario, the AGGR1 router could signal an incremental OSPF RM to some or all the R1-RN routers. When the R1-RN routers add this signaled RM offset to the provisioned metric on their links towards AGGR1, the path via AGGR2 becomes a better path. This causes traffic towards the core to be diverted away from AGGR1. Note that the RM mechanism allows such adaptive metric changes to be applied on the AGGR1 as opposed to being provisioned on a possibly large number of R1-RN routers.

The RM mechanism may be similarly applied between spine and leaf nodes in a Clos network [CLOS] topology deployment.

3. Solution

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

The RM mechanism specified in this document applies only to P2P, Point-to-Multipoint (P2MP), and hybrid-broadcast-P2MP ([RFC6845]) links. It is not applicable for broadcast or Non-Broadcast Multi-Access (NBMA) links since the same objective is achieved there using the OSPF Two-Part Metric mechanism [RFC8042] for OSPFv2. The OSPFv3 solution for broadcast or NBMA links is outside the scope of this document.

4. LLS Reverse Metric TLV

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

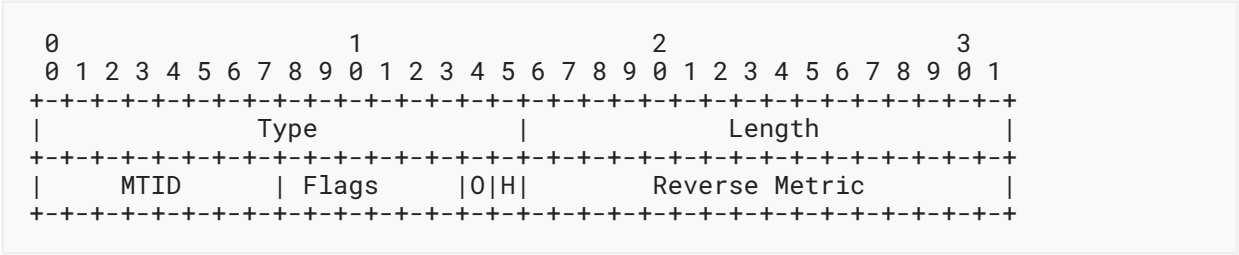

Figure 2: Reverse Metric TLV

where:

Type: 19

Length: 4 octets

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

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

H (0x1): Indicates that the neighbor should use the value only if it is higher than its provisioned metric value for the link.

O (0x2): Indicates that the RM value provided is an offset that is to be added to the provisioned metric.

Reverse Metric: Unsigned integer of 2 octets that carries the value or offset of RM to replace or be added to the provisioned link metric.

5. LLS Reverse TE Metric TLV

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

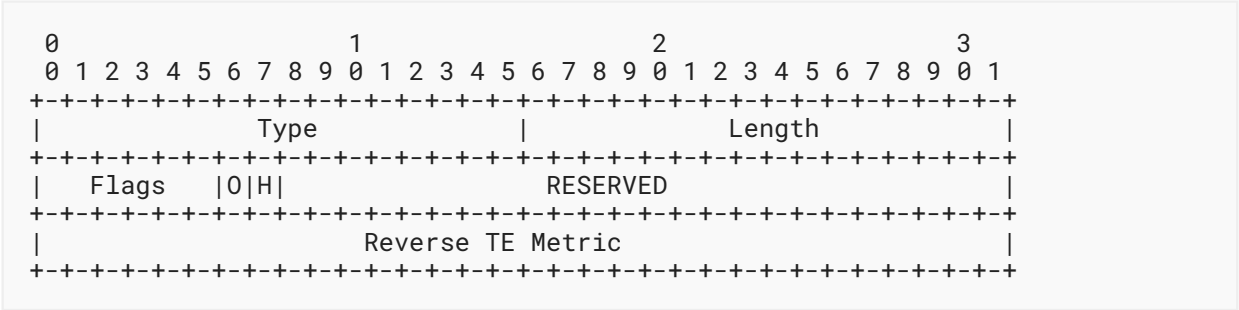

Figure 3: Reverse TE Metric TLV

where:

Type: 20

Length: 4 octets

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

H (0x1): Indicates that the neighbor should use the value only if it is higher than its provisioned TE metric value for the link.

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

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

Reverse TE Metric: Unsigned integer of 4 octets that carries the value or offset of reverse traffic engineering metric to replace or to be added to the provisioned TE metric of the link.

6. Procedures

When a router needs to signal an RM value that its neighbor(s) should use for a link towards the router, it includes the Reverse Metric TLV in the LLS block of its Hello packets sent on that link and continues to include this TLV for as long as the router needs its 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. For example, the RM value may be derived based on the router's link bandwidth with respect to a reference bandwidth.

A router receiving a Hello packet from its neighbor that contains the Reverse Metric TLV on a link **MUST** use the RM value to derive the metric for the link to the advertising router in its Router-LSA when the RM feature is enabled (refer to Section 7 for details on enablement of RM). When the O flag is set, the metric value to be advertised is derived by adding the value in the TLV to the provisioned metric for the link. The metric value 0xffff (maximum interface cost) is advertised when the sum exceeds the maximum interface cost. When the O flag is clear, the metric value to be advertised is copied directly from the value in the TLV. When the H flag is set and the O flag is clear, the metric value to be advertised is copied directly from the value in the TLV only when the RM value signaled is higher than the provisioned metric for the link. The H and O flags are mutually exclusive; the H flag is ignored when the O flag is set.

A router stops including the Reverse Metric TLV in its Hello packets when it needs its neighbors to go back to using their own provisioned metric values. When this happens, a router that has modified its metric in response to receiving a Reverse Metric TLV from its neighbor **MUST** revert to using its provisioned metric value.

In certain scenarios, two or more routers may start the RM signaling on the same link. This could create collision scenarios. The following guidelines are **RECOMMENDED** for adoption to ensure that there is no instability in the network due to churn in their metric caused by the signaling of RM:

- The RM value that is signaled by a router to its neighbor should not be derived from the RM being signaled by any of its neighbors on any of its links.
- The RM value that is signaled by a router to its neighbor should not be derived from the RM being signaled by any of its neighbors on any of its links. RM signaling from other routers can affect the router's metric advertised in its Router-LSA. When deriving the RM values that a router signals to its neighbors, it should use its provisioned local metric values not influenced by any RM signaling.

Based on these guidelines, a router would not start, stop, or change its RM signaling based on the RM signaling initiated by some other routers. 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 cases when symmetrical metrics are desired (e.g., when metrics are derived based on link bandwidth), the RM signaling can be enabled on routers on either end of a link. In other use cases (as described in Section 2.1), RM signaling may need to be enabled only on the 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 packet (one for each of the topologies for which it desires to signal the RM). A router **MUST NOT** include more than one instance of this TLV per MTID. If more than a single instance of this TLV per MTID is present, the receiving router **MUST** only use the value from the first instance and ignore the others.

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. Using similar procedures to those described above, the Reverse TE Metric TLV **MAY** be used to signal the reverse TE metric for router links. The neighbor **MUST** use the reverse TE metric value to derive the TE metric advertised in the TE Metric sub-TLV of the Link TLV in its TE Opaque LSA [RFC3630] when the reverse metric feature is enabled (refer Section 7 for details on enablement of RM). The rules for doing so are analogous to those given above for the Router-LSA.

7. Operational Guidelines

The signaled RM does not alter the OSPF metric parameters stored in a receiving router's persistent provisioning database.

Routers that receive an RM advertisement **SHOULD** log an event to notify system administration. This will assist in rapidly identifying the node in the network that is advertising an OSPF metric or TE metric different from what is configured locally on the device.

When the link TE metric is raised to the maximum value, either due to the RM mechanism or by explicit user configuration, this **SHOULD** immediately trigger the CSPF (Constrained Shortest Path First) recalculation to move the TE traffic away from that link.

An implementation **MUST NOT** signal RM to neighbors by default and **MUST** provide a configuration option to enable the signaling of RM on specific links. An implementation **MUST NOT** accept the RM from its neighbors by default. An implementation **MAY** provide configuration to accept the RM globally on the device, or per area, but an implementation **MUST** support configuration to enable/disable acceptance of the RM from neighbors on specific links. This is to safeguard against inadvertent use of RM.

For the use case in Section 2.1, it is **RECOMMENDED** that the network operator limit the period of enablement of the reverse metric mechanism to be only the duration of a network maintenance window.

[RFC9129] specifies the base OSPF YANG data model. The required configuration and operational elements for this feature are expected to be introduced as an augmentation to this base OSPF YANG data model.

8. Backward Compatibility

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

9. IANA Considerations

IANA has registered code points from the "Link Local Signalling TLV Identifiers (LLS Types)" registry in the "Open Shortest Path First (OSPF) Link Local Signalling (LLS) - Type/Length/Value Identifiers (TLV)" registry group for the TLVs introduced in this document as follows:

- 19 - Reverse Metric TLV
- 20 - Reverse TE Metric TLV

10. Security Considerations

The security considerations for "OSPF Link-Local Signaling" [RFC5613] also apply to the extension described in this document. The purpose of using 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 traffic being misrouted. If authentication is being used in the OSPFv2 routing domain [RFC5709] [RFC7474], then the Cryptographic Authentication TLV [RFC5613] **MUST** also be used to protect the contents of the LLS block.

A router that is misbehaving or misconfigured may end up signaling varying values of RMs or toggle the state of RM. This can result in a neighbor router having to frequently update its Router LSA, causing network churn and instability despite existing OSPF protocol mechanisms (e.g., MinLSInterval, and [RFC8405]). It is **RECOMMENDED** that implementations support the detection of frequent changes in RM signaling and ignore the RM (i.e., revert to using their provisioned metric value) during such conditions.

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.

11. References

11.1. 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>.

11.2. Informative References

[CLOS] Clos, C., "A study of non-blocking switching networks", The Bell System Technical Journal, Vol. 32, Issue 2, DOI 10.1002/j.1538-7305.1953.tb01433.x, March 1953, <https://doi.org/10.1002/j.1538-7305.1953.tb01433.x>.

[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>.

[RFC8405] Decraene, B., Litkowski, S., Gredler, H., Lindem, A., Francois, P., and C. Bowers, "Shortest Path First (SPF) Back-Off Delay Algorithm for Link-State IGPs", RFC 8405, DOI 10.17487/RFC8405, June 2018, <https://www.rfc-editor.org/info/rfc8405>.

[RFC8500] Shen, N., Amante, S., and M. Abrahamsson, "IS-IS Routing with Reverse Metric", RFC 8500, DOI 10.17487/RFC8500, February 2019, <https://www.rfc-editor.org/info/rfc8500>.

[RFC9129] Yeung, D., Qu, Y., Zhang, Z., Chen, I., and A. Lindem, "YANG Data Model for the OSPF Protocol", RFC 9129, DOI 10.17487/RFC9129, October 2022, <https://www.rfc-editor.org/info/rfc9129>.

Acknowledgements

The authors would like to thank:

- Jay Karthik for his contributions to the use cases and the review of the solution.
- Les Ginsberg, Aijun Wang, Gyan Mishra, Matthew Bocci, Thomas Fossati, and Steve Hanna for their review and feedback.
- Acee Lindem for a detailed shepherd's review and comments.
- John Scudder for his detailed AD review and suggestions for improvement.

The document leverages the concept of RM for IS-IS, its related use cases, and applicability aspects from [RFC8500].

Authors' Addresses

Ketan Talaulikar (EDITOR)
Cisco Systems, Inc.
India
Email: ketant.ietf@gmail.com

Peter Psenak
Cisco Systems, Inc.
Apollo Business Center
Mlynske nivy 43
821 09 Bratislava
Slovakia
Email: ppsenak@cisco.com

Hugh Johnston
AT&T Labs
United States of America
Email: hugh.johnston@att.com