Internet Engineering Task Force (IETF)
Request for Comments: 7823
Category: Informational
ISSN: 2070-1721

A. Atlas
J. Drake
Juniper Networks
S. Giacalone
Microsoft
S. Previdi
Cisco Systems

May 2016

Performance-Based Path Selection for Explicitly Routed Label Switched Paths (LSPs) Using TE Metric Extensions

Abstract

In certain networks, it is critical to consider network performance criteria when selecting the path for an explicitly routed RSVP-TE Label Switched Path (LSP). Such performance criteria can include latency, jitter, and loss or other indications such as the conformance to link performance objectives and non-RSVP TE traffic load. This specification describes how a path computation function may use network performance data, such as is advertised via the OSPF and IS-IS TE metric extensions (defined outside the scope of this document) to perform such path selections.

Status of This Memo

This document is not an Internet Standards Track specification; it is published for informational purposes.

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). Not all documents approved by the IESG are a candidate for any level of Internet Standard; see Section 2 of RFC 5741.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at http://www.rfc-editor.org/info/rfc7823.

Table of Contents

1. Introduction

In certain networks, such as financial information networks, network performance information is becoming as critical to data-path selection as other existing metrics. Network performance information can be obtained via either the TE Metric Extensions in OSPF [RFC7471] or IS-IS [RFC7810] or via a management system. As with other TE information flooded via OSPF or IS-IS, the TE metric extensions have a flooding scope limited to the local area or level. This document describes how a path computation function, whether in an ingress LSR or a PCE [RFC4655], can use that information for path selection for explicitly routed LSPs. The selected path may be signaled via RSVP-TE [RFC3209] [RFC3473] or simply used by the ingress with segment

routing [SEG-ROUTE-MPLS] to properly forward the packet. Methods of optimizing path selection for multiple parameters are generally computationally complex. However, there are good heuristics for the delay-constrained lowest-cost (DCLC) computation problem [k-Paths_DCLC] that can be applied to consider both path cost and a maximum delay bound. Some of the network performance information can also be used to prune links from a topology before computing the path.

The path selection mechanisms described in this document apply to paths that are fully computed by the head-end of the LSP and then signaled in an Explicit Route Object (ERO) where every sub-object is strict. This allows the head-end to consider IGP-distributed performance data without requiring the ability to signal the performance constraints in an object of the RSVP Path message.

When considering performance-based data, it is obvious that there are additional contributors to latency beyond just the links. Clearly end-to-end latency is a combination of router latency (e.g., latency from traversing a router without queueing delay), queuing latency, physical link latency, and other factors. While traversing a router can cause delay, that router latency can be included in the advertised link delay. As described in [RFC7471] and [RFC7810], queuing delay must not be included in the measurements advertised by OSPF or IS-IS.

Queuing latency is specifically excluded to insure freedom from oscillations and stability issues that have plagued prior attempts to use delay as a routing metric. If application traffic follows a path based upon latency constraints, the same traffic might be in an Expedited Forwarding Per-Hop Behavior (PHB) [RFC3246] with minimal queuing delay or another PHB with potentially very substantial per-hop queuing delay. Only traffic that experiences relatively low congestion, such as Expedited Forwarding traffic, will experience delays very close to the sum of the reported link delays.

This document does not specify how a router determines what values to advertise by the IGP; it does assume that the constraints specified in [RFC7471] and [RFC7810] are followed. Additionally, the end-to-end performance that is computed for an LSP path should be built from the individual link data. Any end-to-end characterization used to determine an LSP’s performance compliance should be fully reflected in the Traffic Engineering Database so that a path calculation can also determine whether a path under consideration would be in compliance.

1.1. Basic Requirements

The following are the requirements considered for a path computation function that uses network performance criteria.

1. Select a TE tunnel's path based upon a combination of existing constraints as well as on link-latency, packet loss, jitter, conformance with link performance objectives, and bandwidth consumed by non-RSVP-TE traffic.

2. Ability to define different end-to-end performance requirements for each TE tunnel regardless of common use of resources.

3. Ability to periodically verify with the TE Link State Database (LSDB) that a TE tunnel's current LSP complies with its configured end-to-end performance requirements.

4. Ability to move tunnels, using make-before-break, based upon computed end-to-end performance complying with constraints.

5. Ability to move tunnels away from any link that is violating an underlying link performance objective.

6. Ability to optionally avoid setting up tunnels using any link that is violating a link performance objective, regardless of whether end-to-end performance would still meet requirements.

7. Ability to revert back, using make-before-break, to the best path after a configurable period.

1.2. Oscillation and Stability Considerations

Past attempts to use unbounded delay or loss as a metric suffered from severe oscillations. The use of performance based data must be such that undamped oscillations are not possible and stability cannot be impacted.

The use of timers is often cited as a cure. Oscillation that is damped by timers is known as "slosh". If advertisement timers are very short relative to the jitter applied to RSVP-TE Constrained Shortest Path First (CSPF) timers, then a partial oscillation occurs. If RSVP-TE CSPF timers are short relative to advertisement timers, full oscillation (all traffic moving back and forth) can occur. Even a partial oscillation causes unnecessary reordering that is considered at least minimally disruptive.

Delay variation or jitter is affected by even small traffic levels. At even tiny traffic levels, the probability of a queue occupancy of one can produce a measured jitter proportional to or equal to the packet serialization delay. Very low levels of traffic can increase the probability of queue occupancies of two or three packets enough to further increase the measured jitter. Because jitter measurement is extremely sensitive to very low traffic levels, any use of jitter is likely to oscillate. However, there may be uses of a jitter measurement in path computation that can be considered free of oscillation.

Delay measurements that are not sensitive to traffic loads may be safely used in path computation. Delay measurements made at the link layer or measurements made at a queuing priority higher than any significant traffic (such as Differentiated Services Code Point (DSCP) CS7 or CS6 [RFC4594], but not CS2 if traffic levels at CS3 and higher or Expedited Forwarding and Assured Forwarding can affect the measurement). Making delay measurements at the same priority as the traffic on affected paths is likely to cause oscillations.

2. Using Performance Data Constraints

2.1. End-to-End Constraints

The per-link performance data available in the IGP [RFC7471] [RFC7810] includes: unidirectional link delay, unidirectional delay variation, and link loss. Each (or all) of these parameters can be used to create the path-level link-based parameter.

It is possible to compute a CSPF where the link latency values are used instead of TE metrics; this results in ignoring the TE metrics and causing LSPs to prefer the lowest-latency paths. In practical scenarios, latency constraints are typically a bound constraint rather than a minimization objective. An end-to-end latency upper bound merely requires that the path computed be no more than that bound and does not require that it be the minimum latency path. The latter is exactly the DCLC problem to which good heuristics have been proposed in the literature (e.g., [k-Paths_DCLC]).

An end-to-end bound on delay variation can be used similarly as a constraint in the path computation on what links to explore where the path’s delay variation is the sum of the used links’ delay variations.

For link loss, the path loss is not the sum of the used links’ losses. Instead, the path loss fraction is 1 - (1 - loss_L1)* (1 - loss_L2)*...*(1 - loss_Ln), where the links along the path are L1 to Ln with loss_Li in fractions. This computation is discussed in

more detail in Sections 5.1.4 and 5.1.5 in [RFC6049]. The end-to-end link loss bound, computed in this fashion, can also be used as a constraint in the path computation.

The heuristic algorithms for DCLC only address one constraint bound but having a CSPF that limits the paths explored (i.e., based on hop count) can be combined [hop-count_DCLC].

2.2. Link Constraints

In addition to selecting paths that conform to a bound on performance data, it is also useful to avoid using links that do not meet a necessary constraint. Naturally, if such a parameter were a known fixed value, then resource attribute flags could be used to express this behavior. However, when the parameter associated with a link may vary dynamically, there is not currently a configuration-time mechanism to enforce such behavior. An example of this is described in Section 2.3, where links may move in and out of conformance for link performance objectives with regards to latency, delay variation, and link loss.

When doing path selection for TE tunnels, it has not been possible to know how much actual bandwidth is available that includes the bandwidth used by non-RSVP-TE traffic. In [RFC7471] and [RFC7810], the Unidirectional Available Bandwidth is advertised as is the Residual Bandwidth. When computing the path for a TE tunnel, only links with at least a minimum amount of Unidirectional Available Bandwidth might be permitted.

Similarly, only links whose loss is under a configurable value might be acceptable. For these constraints, each link can be tested against the constraint and only explored in the path computation if the link passes. In essence, a link that fails the constraint test is treated as if it contained a resource attribute in the exclude-any filter.

2.3. Links out of Compliance with Link Performance Objectives

Link conformance to a link performance objective can change as a result of rerouting at lower layers. This could be due to optical regrooming or simply rerouting of an FA-LSP. When this occurs, there are two questions to be asked:

a. Should the link be trusted and used for the setup of new LSPs?

b. Should LSPs using this link automatically be moved to a secondary path?

2.3.1. Use of Anomalous Links for New Paths

If the answer to (a) is no for link latency performance objectives, then any link that has the Anomalous bit set in the Unidirectional Link Delay sub-TLV [RFC7471] [RFC7810] should be removed from the topology before a path calculation is used to compute a new path. In essence, the link should be treated exactly as if it fails the exclude-any resource attributes filter [RFC3209].

Similarly, if the answer to (a) is no for link loss performance objectives, then any link that has the Anomalous bit set in the Link Loss sub-TLV should be treated as if it fails the exclude-any resource attributes filter.

2.3.2. Links Entering the Anomalous State

When the Anomalous bit transitions from clear to set, this indicates that the associated link has entered the Anomalous state with respect to the associated parameter; similarly, a transition from set to clear indicates that the Anomalous state has been exited for that link and associated parameter.

When a link enters the Anomalous state with respect to a parameter, this is an indication that LSPs using that link might also no longer be in compliance with their performance bounds. It can also be considered an indication that something is changing that link and so it might no longer be trustworthy to carry performance-critical traffic. Naturally, which performance criteria are important for a particular LSP is dependent upon the LSP's configuration; thus, the compliance of a link with respect to a particular link performance objective is indicated per performance criterion.

At the ingress of a TE tunnel, a TE tunnel may be configured to be sensitive to the Anomalous state of links in reference to latency, delay variation, and/or loss. Additionally, such a TE tunnel may be configured to either verify continued compliance, to switch immediately to a standby LSP, or to move to a different path.

When a sub-TLV is received with the Anomalous bit set when previously it was clear, the list of interested TE tunnels must be scanned. Each such TE tunnel should have its continued compliance verified, be switched to a hot standby, or do a make-before-break to a secondary path.

It is not sufficient to just look at the Anomalous bit in order to determine when TE tunnels must have their compliance verified. When changing to set, the Anomalous bit merely provides a hint that

interested TE tunnels should have their continued compliance verified.

2.3.3. Links Leaving the Anomalous State

When a link leaves the Anomalous state with respect to a parameter, this can serve as an indication that those TE tunnels, whose LSPs were changed due to administrative policy when the link entered the Anomalous state, may want to reoptimize to a better path. The hint provided by the Anomalous state change may help optimize when to recompute for a better path.

3. Security Considerations

This document is not currently believed to introduce new security concerns.

4. References

4.1. Normative References

[RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, DOI 10.17487/RFC3209, December 2001, <http://www.rfc-editor.org/info/rfc3209>.

[RFC7471] Giacalone, S., Ward, D., Drake, J., Atlas, A., and S. Previdi, "OSPF Traffic Engineering (TE) Metric Extensions", RFC 7471, DOI 10.17487/RFC7471, March 2015, <http://www.rfc-editor.org/info/rfc7471>.

[RFC7810] Previdi, S., Ed., Giacalone, S., Ward, D., Drake, J., and Q. Wu, "IS-IS Traffic Engineering (TE) Metric Extensions", RFC 7810, DOI 10.17487/7810, May 2016, <http://www.rfc-editor.org/info/rfc7810>.

4.2. Informative References

[hop-count_DCLC] Agrawal, H., Grah, M., and M. Gregory, "Optimization of QoS Routing", 6th IEEE/AACIS International Conference on Computer and Information Science, DOI 10.1109/ICIS.2007.144, July 2007, <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=4276447>.

[k-Paths_DCLC]
Jia, Z. and P. Varaiya, "Heuristic methods for delay constrained least cost routing using k-shortest-paths", IEEE Transactions on Automatic Control, vol. 51, no. 4, April 2006, <http://dx.doi.org/10.1109/TAC.2006.872827>.

[RFC3246] Davie, B., Charny, A., Bennet, J., Benson, K., Le Boudec, J., Courtney, W., Davari, S., Firoiu, V., and D. Stiliadis, "An Expedited Forwarding PHB (Per-Hop Behavior)", RFC 3246, DOI 10.17487/RFC3246, March 2002, <http://www.rfc-editor.org/info/rfc3246>.

[RFC3473] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, DOI 10.17487/RFC3473, January 2003, <http://www.rfc-editor.org/info/rfc3473>.

[RFC4594] Babiarz, J., Chan, K., and F. Baker, "Configuration Guidelines for DiffServ Service Classes", RFC 4594, DOI 10.17487/RFC4594, August 2006, <http://www.rfc-editor.org/info/rfc4594>.

[RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, DOI 10.17487/RFC4655, August 2006, <http://www.rfc-editor.org/info/rfc4655>.

[RFC6049] Morton, A. and E. Stephan, "Spatial Composition of Metrics", RFC 6049, DOI 10.17487/RFC6049, January 2011, <http://www.rfc-editor.org/info/rfc6049>.

[SEG-ROUTE-MPLS]
Filsfils, C., Ed., Previdi, S., Ed., Bashandy, A., Decraene, B., Litkowski, S., Horneffer, M., Shakir, R., Tantsura, J., and E. Crabbe, "Segment Routing with MPLS data plane", Work in Progress, draft-ietf-spring-segment-routing-mpls-04, March 2016.

Acknowledgements

The authors would like to thank Curtis Villamizar for his extensive detailed comments and suggested text in Sections 1 and 1.2. The authors would like to thank Dhruv Dhody for his useful comments and his care and persistence in making sure that these important corrections weren’t missed. The authors would also like to thank Xiaohu Xu and Sriganesh Kini for their reviews.

Contributors

Dave Ward and Clarence Filsfils contributed to this document.

Authors' Addresses

Alia Atlas
Juniper Networks
10 Technology Park Drive
Westford, MA 01886
United States

Email: akatlas@juniper.net

John Drake
Juniper Networks
1194 N. Mathilda Ave.
Sunnyvale, CA 94089
United States

Email: jdrake@juniper.net

Spencer Giacalone
Microsoft

Email: spencer.giacalone@gmail.com

Stefano Previdi
Cisco Systems
Via Del Serafico 200
Rome 00142
Italy

Email: sprevidi@cisco.com