Internet Engineering Task Force (IETF)
Request for Comments: 8541
Category: Informational
ISSN: 2070-1721

S. Litkowski
Orange Business Service
B. Decraene
Orange
M. Horneffer
Deutsche Telekom

March 2019

Impact of Shortest Path First (SPF) Trigger and Delay Strategies on IGP Micro-loops

Abstract

A micro-loop is a packet-forwarding loop that may occur transiently among two or more routers in a hop-by-hop packet-forwarding paradigm.

This document analyzes the impact of using different link state IGP implementations in a single network with respect to micro-loops. The analysis is focused on the Shortest Path First (SPF) delay algorithm but also mentions the impact of SPF trigger strategies.

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 candidates for any level of Internet Standard; see 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/rfc8541.

Table of Contents

1. Introduction

Link state IGP protocols are based on a topology database on which the SPF algorithm is run to find a consistent set of non-looping routing paths.

Specifications like IS-IS [RFC1195] propose some optimizations of the route computation (see Appendix C.1 of [RFC1195]), but not all implementations follow those non-mandatory optimizations.

In this document, we refer to the events that lead to a new SPF computation based on the topology as "SPF triggers".

Link state IGP protocols, like OSPF [RFC2328] and IS-IS [RFC1195], use multiple timers to control the router behavior in case of churn: SPF delay, Partial Route Computation (PRC) delay, Link State Packet (LSP) generation delay, LSP flooding delay, and LSP retransmission interval.

Some of the values and behaviors of these timers are standardized in protocol specifications, and some are not. The SPF computation-related timers have generally remained unspecified.

Implementations are free to implement non-standardized timers in any way. For some standardized timers, implementations may offer dynamically adjusted timers to help control the churn rather than use static configurable values.

"SPF delay" refers to the timer in most implementations that specifies the required delay before running an SPF computation after an SPF trigger is received.

A micro-loop is a packet-forwarding loop that may occur transiently among two or more routers in a hop-by-hop packet-forwarding paradigm. These micro-loops are formed when two routers do not update their Forwarding Information Bases (FIBs) for a certain prefix at the same time. The micro-loop phenomenon is described in [MICROLOOP-LSRP].

Two micro-loop mitigation techniques have been defined by IETF. The mechanism in [RFC6976] has not been widely implemented, presumably due to the complexity of the technique. The mechanism in [RFC8333] has been implemented. However, it does not prevent all micro-loops that can occur for a given topology and failure scenario.

In multi-vendor networks, using different implementations of a link state protocol may favor micro-loop creation during the convergence process due to discrepancies in timers. Service providers already know to use timers with similar values and behaviors for all of the network as a best practice, but this is sometimes not possible due to the limitations of implementations.

This document presents reasons for service providers to have consistent implementation of link state protocols across vendors. In particular, this document analyzes the impact of using different link state IGP implementations in a single network with regard to micro-loops. The analysis focuses on the SPF delay algorithm.

[RFC8405] defines a solution that partially addresses this problem statement, and this document captures the reasoning of the provided solution.

2. Problem Statement

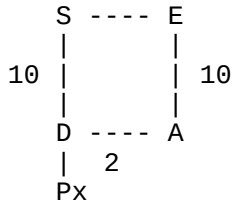

Figure 1: Network Topology Experiencing Micro-loops

Figure 1 represents a small network composed of four routers (S, D, E, and A). Router S primarily uses the SD link to reach the prefixes behind router D (named Px). When the SD link fails, the IGP convergence occurs. If S converges before E, S will forward the traffic to Px through E; however, because E has not converged yet, E will loop the traffic back to S, leading to a micro-loop.

The micro-loop appears due to the asynchronous convergence of nodes in a network when an event occurs.

Multiple factors (or a combination of factors) may increase the probability of a micro-loop appearing:

- Delay of failure notification: The greater the time gap between E and S being advised of the failure, the greater the chance that a micro-loop may appear.

- SPF delay: Most implementations support a delay for the SPF computation to catch as many events as possible. If S uses an SPF delay timer of x ms, E uses an SPF delay timer of y ms, and x < y, E would start converging after S, leading to a potential micro-loop.

- SPF computation time: This is mostly a matter of CPU power and optimizations like incremental SPF. If S computes its SPF faster than E, there is a chance for a micro-loop to appear. Today, CPUs are fast enough to consider the SPF computation time as negligible (on the order of milliseconds in a large network).

- SPF computation ordering: An SPF trigger can be common to multiple IGP areas or levels (e.g., IS-IS Level 1 and Level 2) or to multiple address families with multi-topologies. There is no specified order for SPF computation today, and it is implementation dependent. In such scenarios, if the order of SPF computation done in S and E for each area, level, topology, or SPF algorithm is different, there is a possibility for a micro-loop to appear.

- RIB and FIB prefix insertion speed or ordering: This is highly dependent on the implementation.

Even if all of these factors increase the probability of a micro-loop appearing, the SPF delay plays a significant role, especially in case of churn. As the number of IGP events increases, the delta between the SPF delay values used by routers becomes significant; in fact, it becomes the dominating factor (especially when one router increases its timer exponentially while another one increases it in a smoother way). Another important factor is the time to update the FIB. As of today, the total FIB update time is the major factor for IGP convergence. However, for micro-loops, what matters is not the total time but the difference in installing the same prefix between nodes. The time to update the FIB may be the main part for the first iteration but not for subsequent IGP events. In addition, the time to update the FIB is very implementation specific and difficult or impossible to standardize, while the SPF delay algorithm may be standardized.

As a consequence, this document will focus on an analysis of SPF delay behavior and associated triggers.

3. SPF Trigger Strategies

Depending on the change advertised in the LSP or LSA (Link State Advertisement), the topology may or may not be affected. An implementation may avoid running the SPF computation (and may only run an IP reachability computation instead) if the advertised change does not affect the topology.

Different strategies can trigger the SPF computation:

1. An implementation may always run a full SPF for any type of change.

2. An implementation may run a full SPF only when required. For example, if a link fails, a local node will run an SPF for its local LSP update. If the LSP from the neighbor (describing the same failure) is received after SPF has started, the local node can decide that a new full SPF is not required as the topology has not changed.

3. If the topology does not change, an implementation may only recompute the IP reachability.

As noted in Section 1, SPF optimizations are not mandatory in specifications. This has led to the implementation of different strategies.

4. SPF Delay Strategies

Implementations of link state routing protocols use different strategies to delay SPF computation. The two most common SPF delay behaviors are the following:

1. Two-step SPF delay

2. Exponential back-off delay

These behaviors are explained in the following sections.

4.1. Two-Step SPF Delay

The SPF delay is managed by four parameters:

- rapid delay: the amount of time to wait before running SPF after the initial SPF trigger event.
- rapid runs: the number of consecutive SPF runs that can use the rapid delay. When the number is exceeded, the delay moves to the slow delay value.
- slow delay: the amount of time to wait before running an SPF.
- wait time: the amount of time to wait without detecting SPF trigger events before going back to the rapid delay.

Figure 2 displays the evolution of the SPF delay timer (based on a two-step delay algorithm) upon the reception of multiple events. Figure 2 considers the following parameters for the algorithm: rapid delay (RD) = 50 ms, rapid runs = 3, slow delay (SD) = 1 s, wait time = 2 s.

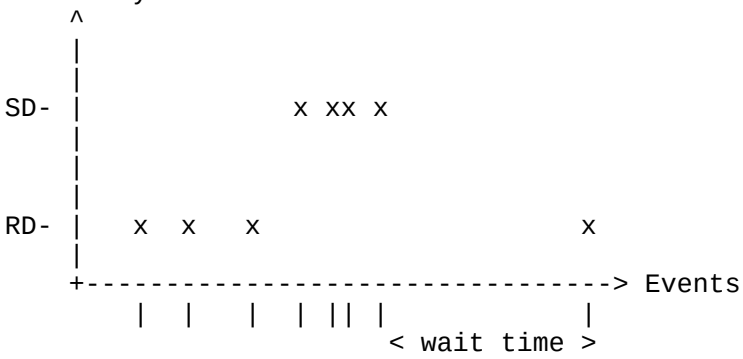

Figure 2: Two-Step SPF Delay Algorithm

4.2. Exponential Back-Off Delay

The algorithm has two modes: fast mode and back-off mode. In fast mode, the SPF delay is usually delayed by a very small amount of time (fast reaction). When an SPF computation is run in fast mode, the algorithm automatically moves to back-off mode (a single SPF run is authorized in fast mode). In back-off mode, the SPF delay increases exponentially in each run. When the network becomes stable, the algorithm moves back to fast mode. The SPF delay is managed by four parameters:

- first delay: amount of time to wait before running SPF. This delay is used only when SPF is in fast mode.
- incremental delay: amount of time to wait before running SPF. This delay is used only when SPF is in back-off mode and increments exponentially at each SPF run.
- maximum delay: maximum amount of time to wait before running SPF.
- wait time: amount of time to wait without events before going back to fast mode.

Figure 3 displays the evolution of the SPF delay timer (based on an exponential back-off delay algorithm) upon the reception of multiple events. Figure 3 considers the following parameters for the algorithm: first delay (FD) = 50 ms, incremental delay (ID) = 50 ms, maximum delay (MD) = 1 s, wait time = 2 s

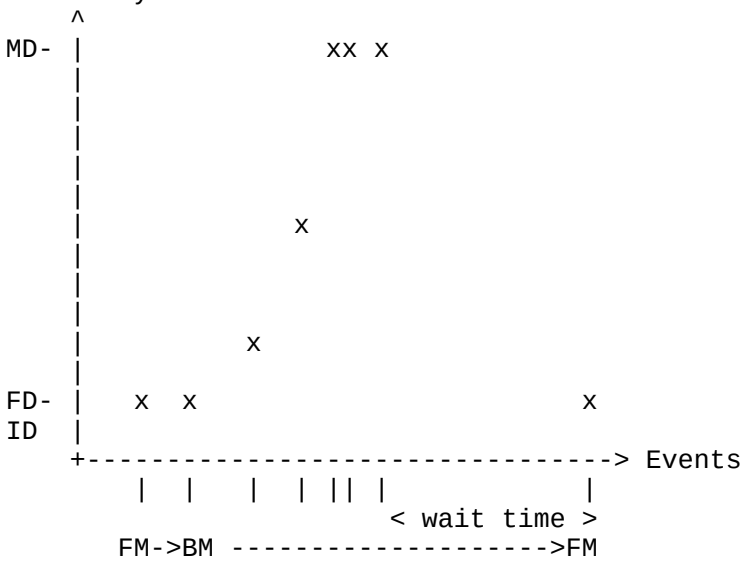

Figure 3: Exponential Back-Off Delay Algorithm

5. Mixing Strategies

Figure 1 illustrates a flow of packets from S to D. S uses optimized SPF triggering (full SPF is triggered only when necessary) and two-step SPF delay (rapid delay = 150 ms, rapid runs = 3, slow delay = 1 s). As the implementation of S is optimized, PRC is available. For PRC delay, we consider the same timers as for SPF delay. E uses an SPF trigger strategy that always computes a full SPF for any change and uses the exponential back-off strategy for SPF delay (first delay = 150 ms, incremental delay = 150 ms, maximum delay = 1 s).

Consider the following sequence of events:

- t0=0 ms: A prefix is declared down in the network. This event happens at time=0.
- 200 ms: The prefix is declared up.
- 400 ms: The prefix is declared down in the network.
- 1000 ms: S-D link fails.

Time	Network Event	Router S Events	Router E Events
t0=0	Prefix DOWN		
10 ms		Schedule PRC (in 150 ms)	Schedule SPF (in 150 ms)
160 ms		PRC starts	SPF starts
161 ms		PRC ends	
162 ms		RIB/FIB starts	
163 ms			SPF ends
164 ms			RIB/FIB starts
175 ms		RIB/FIB ends	
178 ms			RIB/FIB ends
200 ms	Prefix UP		
212 ms		Schedule PRC (in 150 ms)	
214 ms			Schedule SPF (in 150 ms)
370 ms		PRC starts	
372 ms		PRC ends	
373 ms			SPF starts

373 ms		RIB/FIB starts	
375 ms			SPF ends
376 ms			RIB/FIB starts
383 ms		RIB/FIB ends	
385 ms			RIB/FIB ends
400 ms	Prefix DOWN		
410 ms		Schedule PRC (in 300 ms)	Schedule SPF (in 300 ms)
710 ms		PRC starts	SPF starts
711 ms		PRC ends	
712 ms		RIB/FIB starts	
713 ms			SPF ends
714 ms			RIB/FIB starts
716 ms		RIB/FIB ends	RIB/FIB ends
1000 ms	S-D link DOWN		
1010 ms		Schedule SPF (in 150 ms)	Schedule SPF (in 600 ms)
1160 ms		SPF starts	
1161 ms		SPF ends	
1162 ms	Micro-loop may start from here	RIB/FIB starts	
1175 ms		RIB/FIB ends	
1612 ms			SPF starts
1615 ms			SPF ends
1616 ms			RIB/FIB starts
1626 ms	Micro-loop ends		RIB/FIB ends

Table 1: Route Computation When S and E Use Different Behaviors and Multiple Events Appear

In Table 1, due to discrepancies in the SPF management and after multiple events of different types, the values of the SPF delay are completely misaligned between node S and node E, leading to the creation of micro-loops.

The same issue can also appear with only a single type of event as shown below:

Time	Network Event	Router S Events	Router E Events
t0=0	Link DOWN		
10 ms		Schedule SPF (in 150 ms)	Schedule SPF (in 150 ms)
160 ms		SPF starts	SPF starts
161 ms		SPF ends	
162 ms		RIB/FIB starts	
163 ms			SPF ends
164 ms			RIB/FIB starts
175 ms		RIB/FIB ends	
178 ms			RIB/FIB ends
200 ms	Link DOWN		
212 ms		Schedule SPF (in 150 ms)	
214 ms			Schedule SPF (in 150 ms)
370 ms		SPF starts	
372 ms		SPF ends	
373 ms			SPF starts
373 ms		RIB/FIB starts	
375 ms			SPF ends
376 ms			RIB/FIB starts
383 ms		RIB/FIB ends	
385 ms			RIB/FIB ends
400 ms	Link DOWN		
410 ms		Schedule SPF (in 150 ms)	Schedule SPF (in 300 ms)
560 ms		SPF starts	
561 ms		SPF ends	

562 ms	Micro-loop may start from here	RIB/FIB starts	
568 ms		RIB/FIB ends	
710 ms			SPF starts
713 ms			SPF ends
714 ms			RIB/FIB starts
716 ms	Micro-loop ends		RIB/FIB ends
1000 ms	Link DOWN		
1010 ms		Schedule SPF (in 1 s)	Schedule SPF (in 600 ms)
1612 ms			SPF starts
1615 ms			SPF ends
1616 ms	Micro-loop may start from here		RIB/FIB starts
1626 ms			RIB/FIB ends
2012 ms		SPF starts	
2014 ms		SPF ends	
2015 ms		RIB/FIB starts	
2025 ms	Micro-loop ends	RIB/FIB ends	

Table 2: Route Computation upon Multiple Link Down Events When S and E Use Different Behaviors

6. Benefits of Standardized SPF Delay Behavior

Table 3 uses the same event sequence as Table 1. Fewer and/or shorter micro-loops are expected using a standardized SPF delay.

Time	Network Event	Router S Events	Router E Events
t0=0	Prefix DOWN		
10 ms		Schedule PRC (in 150 ms)	Schedule PRC (in 150 ms)
160 ms		PRC starts	PRC starts
161 ms		PRC ends	
162 ms		RIB/FIB starts	PRC ends
163 ms			RIB/FIB starts
175 ms		RIB/FIB ends	
176 ms			RIB/FIB ends
200 ms	Prefix UP		
212 ms		Schedule PRC (in 150 ms)	
213 ms			Schedule PRC (in 150 ms)
370 ms		PRC starts	PRC starts
372 ms		PRC ends	
373 ms		RIB/FIB starts	PRC ends
374 ms			RIB/FIB starts
383 ms		RIB/FIB ends	
384 ms			RIB/FIB ends
400 ms	Prefix DOWN		
410 ms		Schedule PRC (in 300 ms)	Schedule PRC (in 300 ms)
710 ms		PRC starts	PRC starts
711 ms		PRC ends	PRC ends
712 ms		RIB/FIB starts	
713 ms			RIB/FIB starts
716 ms		RIB/FIB ends	RIB/FIB ends
1000 ms	S-D link DOWN		

1010 ms		Schedule SPF (in 150 ms)	Schedule SPF (in 150 ms)
1160 ms		SPF starts	
1161 ms		SPF ends	SPF starts
1162 ms	Micro-loop may start from here	RIB/FIB starts	SPF ends
1163 ms			RIB/FIB starts
1175 ms		RIB/FIB ends	
1177 ms	Micro-loop ends		RIB/FIB ends

Table 3: Route Computation When S and E Use the Same Standardized Behavior

As displayed above, there can be other parameters, like router computation power and flooding timers, that may also influence micro-loops. In all the examples in this document comparing the SPF timer behavior of router S and router E, we have made router E a bit slower than router S. This can lead to micro-loops even when both S and E use a common standardized SPF behavior. However, by aligning implementations of the SPF delay, we expect that service providers may reduce the number and duration of micro-loops.

7. Security Considerations

This document does not introduce any security considerations.

8. IANA Considerations

This document has no actions for IANA.

9. References

9.1. Normative References

[RFC1195] Callon, R., "Use of OSI IS-IS for routing in TCP/IP and dual environments", RFC 1195, DOI 10.17487/RFC1195, December 1990, <https://www.rfc-editor.org/info/rfc1195>.

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

[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>.

9.2. Informative References

[MICROLOOP-LSRP] Zinin, A., "Analysis and Minimization of Microloops in Link-state Routing Protocols", Work in Progress, draft-ietf-rtgwg-microloop-analysis-01, October 2005.

[RFC6976] Shand, M., Bryant, S., Previdi, S., Filsfils, C., Francois, P., and O. Bonaventure, "Framework for Loop-Free Convergence Using the Ordered Forwarding Information Base (oFIB) Approach", RFC 6976, DOI 10.17487/RFC6976, July 2013, <https://www.rfc-editor.org/info/rfc6976>.

[RFC8333] Litkowski, S., Decraene, B., Filsfils, C., and P. Francois, "Micro-loop Prevention by Introducing a Local Convergence Delay", RFC 8333, DOI 10.17487/RFC8333, March 2018, <https://www.rfc-editor.org/info/rfc8333>.

Acknowledgements

The authors would like to thank Mike Shand and Chris Bowers for their useful comments.

Authors' Addresses

Stephane Litkowski
Orange Business Service

Email: stephane.litkowski@orange.com

Bruno Decraene
Orange

Email: bruno.decraene@orange.com

Martin Horneffer
Deutsche Telekom

Email: martin.horneffer@telekom.de