Routing Area Working Group
Internet-Draft
Intended status: Standards Track
Expires: June 16, 2016

S. Litkowski
Orange Business Service
B. Decraene
Orange
M. Horneffer
Deutsche Telekom
December 14, 2015

Link State protocols SPF trigger and delay algorithm impact on IGP micro-loops
draft-ietf-rtgwg-spf-uloop-pb-statement-02

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.

In this document, we are trying to analyze the impact of using different Link State IGP implementations in a single network in regards of micro-loops. The analysis is focused on the SPF triggers and SPF delay algorithm.

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 [RFC2119].

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on June 16, 2016.

Copyright Notice

Copyright (c) 2015 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction

Link State IGP protocols are based on a topology database on which a SPF (Shortest Path First) algorithm like Dijkstra is implemented to find the optimal routing paths.

Specifications like IS-IS ([RFC1195]) propose some optimization of the route computation (See Appendix C.1) but not all the implementations are following those not mandatory optimizations.

We will call SPF trigger, the events that would lead to a new SPF computation based on the topology.

Link State IGP protocols, like OSPF ([RFC2328]) and IS-IS ([RFC1195]), are using plenty of timers to control the router

behavior in case of churn : SPF delay, PRC delay, LSP generation delay, LSP flooding delay, LSP retransmission interval ...

Some of those timers are standardized in protocol specification, some are not especially the SPF computation related timers.

For non standardized timers, implementations are free to implement it in any way. For some standardized timer, we can also see that rather than using static configurable values for such timer , implementations may offer dynamically adjusted timers to help controlling the churn.

We will call SPF delay, the timer that exists in most implementations that specifies the required delay before running SPF computation after a 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. We can observe that these micro-loops are formed when two routers do not update their Forwarding Information Base (FIB) for a certain prefix at the same time. The micro-loop phenomenon is described in [I-D.ietf-rtgwg-microloop-analysis].

Some micro-loop mitigation techniques have been defined by IETF (e.g. [RFC6976], [I-D.ietf-rtgwg-uloop-delay]) but are not implemented due to complexity or are not providing a complete mitigation.

In multi vendor networks, using different implementations of a link state protocol may favor micro-loops creation during convergence time due to discrepancies of timers. Service Providers are already aware to use similar timers for all the network as best practice, but sometimes it is not possible due to limitation of implementations.

This document will present why it sounds important for service provider to have consistent implementations of Link State protocols across vendors. We are particularly analyzing the impact of using different Link State IGP implementations in a single network in regards of micro-loops. The analysis is focused on the SPF triggers and SPF delay algorithm in a first step.

This document is only stating the problem, and defining some work items but its not intended to provide a solution.

2. Problem statement

```
                                  A ---- B
                                  |      |
                               10 |      | 10
                                  |      |
                                  C ---- D
                                  |  2   |
                                  Px     Px
```

Figure 1

In the figure above, A uses primarily the AC link to reach C. When the AC link fails, IGP convergence occurs. If A converges before B, A will forward traffic to C through B, but as B as not converged yet, B will loop back traffic to A, leading to a micro-loop.

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

Multiple factors (and combination of these factors) may increase the probability for a micro-loop to appear :

- o delay of failure notification : the more B is advised of the failure later than A, the more a micro-loop may appear.

- o SPF delay : most of the implementations supports a delay for the SPF computation to try to catch as many events as possible. If A uses a SPF delay timer of x msec and B uses a SPF delay timer of y msec and x < y, B would start converging after A leading to a potential micro-loop.

- o SPF computation time : mostly a matter of CPU power and optimizations like incremental SPF. If A computes SPF faster than B, there is a chance for a micro-loop to appear. CPUs are today faster enough to consider SPF computation time as negligeable (order of msec in a large network).

- o RIB and FIB prefix insertion speed or ordering : highly implementation dependant.

This document will focus on analysis SPF delay (and associated triggers).

3. SPF trigger strategies

Depending of the change advertised in LSP/LSA, the topology may be affected or not. An implementation can decide to not run SPF (and only run IP reachability) if the advertised change is not affecting topology.

Different strategies exists to trigger SPF :

1. Always run full SPF whatever the change to process.

2. Run only Full SPF when required : e.g. 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 change.

3. If topology does not change, only recompute reachability.

As pointed in Section 1, SPF optimization are not mandatory in specifications, leading to multiple strategies to be implemented.

4. SPF delay strategies

Implementations of link state routing protocols use different strategies to delay SPF :

1. Two steps.

2. Exponential backoff.

4.1. Two step SPF delay

The SPF delay is managed by four parameters :

- Rapid delay : amount of time to wait before running SPF.

- Rapid runs : amount of consecutive SPF runs that can run using rapid delay. When amount is exceeded router moves to slow delay.

- Slow delay : amount of time to wait before running SPF.

- Wait time : amount of time to wait without events before going back to rapid delay.

Example : Rapid delay = 50msec, Rapid runs = 3, Slow delay = 1sec, Wait time = 2sec

```
   SPF delay time
       ^
       |
       |
   SD- |             x xx x
       |
       |
       |
   RD- |   x  x   x                    x
       |
       +---------------------------------> Events
           |  |   |  | || |            |
                           < wait time >
```

4.2. Exponential backoff

The algorithm has two mode : fast mode and backoff mode. In backoff mode, the SPF delay is increasing exponentially at each run. The SPF delay is managed by four parameters :

- o First delay : amount of time to wait before running SPF. This delay is used only when SPF is in fast mode.

- o Incremental delay : amount of time to wait before running SPF. This delay is used only when SPF is in backoff mode and increments exponentially at each SPF run.

- o Maximum delay : maximum amount of time to wait before running SPF.

- o Wait time : amount of time to wait without events before going back to fast mode.

Example : First delay = 50msec, Incremental delay = 50msec, Maximum delay = 1sec, Wait time = 2sec

```
   SPF delay time
       ^
   MD- |               xx x
       |
       |
       |
       |
       |
       |             x
       |
       |
       |
       |          x
       |
   FD- |   x  x                        x
   ID  |
       +---------------------------------> Events
           |  |   |  | || |            |
                            < wait time >
          FM->BM -------------------->FM
```

5. Mixing strategies

```
                              S ---- E
                              |      |
                           10 |      | 10
                              |      |
                              D ---- A
                              |  2
                              Px
```

Figure 2

In the diagram above, we consider a flow of packet from S to D. We consider that S is using optimized SPF triggering (Full SPF is triggered only when necessary), and two steps SPF delay (rapid=150ms,rapid-runs=3, slow=1s). As implementation of S is optimized, Partial Reachability Computation (PRC) is available. We consider the same timers as SPF for delaying PRC. We consider that E is using a SPF trigger strategy that always compute Full SPF and exponential backoff strategy for SPF delay (start=150ms, inc=150ms, max=1s)

We also consider the following sequence of events (note : the time scale does not intend to represent a real router time scale where jitters are introduced to all timers) :

- o t0=0 ms : a prefix is declared down in the network. We consider this event to happen at time=0.
- o 200ms : the prefix is declared as up.
- o 400ms : a prefix is declared down in the network.
- o 1000ms : S-D link fails.

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

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

Route computation event time scale

In the table above, we can see that due to discrepancies in SPF management, after multiple events (different types of event), SPF delays are completely misaligned between nodes leading to long micro-loop creation.

The same issue can also appear with only single type of events as displayed below :

Time	Network Event	Router S events	Router E events
t0=0	Link DOWN		
10ms		Schedule SPF (in 150ms)	Schedule SPF (in 150ms)
160ms		SPF starts	SPF starts
161ms		SPF ends	
162ms		RIB/FIB starts	

```
   | 163ms  |                    |                  |     SPF ends     |
   | 164ms  |                    |                  |  RIB/FIB starts  |
   | 175ms  |                    |   RIB/FIB ends   |                  |
   | 178ms  |                    |                  |   RIB/FIB ends   |
   |        |                    |                  |                  |
   | 200ms  |     Link DOWN      |                  |                  |
   | 212ms  |                    | Schedule SPF (in |                  |
   |        |                    |      150ms)      |                  |
   | 214ms  |                    |                  | Schedule SPF (in |
   |        |                    |                  |      150ms)      |
   |        |                    |                  |                  |
   |        |                    |                  |                  |
   | 370ms  |                    |    SPF starts    |                  |
   | 372ms  |                    |     SPF ends     |                  |
   | 373ms  |                    |                  |    SPF starts    |
   | 373ms  |                    |  RIB/FIB starts  |                  |
   | 375ms  |                    |                  |     SPF ends     |
   | 376ms  |                    |                  |  RIB/FIB starts  |
   | 383ms  |                    |   RIB/FIB ends   |                  |
   | 385ms  |                    |                  |   RIB/FIB ends   |
   |        |                    |                  |                  |
   | 400ms  |     Link DOWN      |                  |                  |
   | 410ms  |                    | Schedule SPF (in | Schedule SPF (in |
   |        |                    |      150ms)      |      300ms)      |
   |        |                    |                  |                  |
   |        |                    |                  |                  |
   | 560ms  |                    |    SPF starts    |                  |
   | 561ms  |                    |     SPF ends     |                  |
   | 562ms  |   Micro-loop may   |  RIB/FIB starts  |                  |
   |        |  start from here   |                  |                  |
   | 568ms  |                    |   RIB/FIB ends   |                  |
   |        |                    |                  |                  |
   |        |                    |                  |                  |
   | 710ms  |                    |                  |    SPF starts    |
   | 713ms  |                    |                  |     SPF ends     |
   | 714ms  |                    |                  |  RIB/FIB starts  |
   | 716ms  |  Micro-loop ends   |                  |   RIB/FIB ends   |
   |        |                    |                  |                  |
   | 1000ms |     Link DOWN      |                  |                  |
   | 1010ms |                    | Schedule SPF (in | Schedule SPF (in |
   |        |                    |       1s)        |      600ms)      |
   |        |                    |                  |                  |
   |        |                    |                  |                  |
   |        |                    |                  |                  |
   |        |                    |                  |                  |
   | 1612ms |                    |                  |    SPF starts    |
   | 1615ms |                    |                  |     SPF ends     |
   | 1616ms |   Micro-loop may   |                  |  RIB/FIB starts  |
```

	start from here		
1626ms			RIB/FIB ends
2012ms		SPF starts	
2014ms		SPF ends	
2015ms		RIB/FIB starts	
2025ms	Micro-loop ends	RIB/FIB ends	

Route computation event time scale

6. Proposed work items

In order to enhance the current LinkState IGP behaviour, authors would encourage working on standardization of some behaviours.

Authors are proposing the following work items :

- o Standardize SPF trigger strategy.
- o Standardize computation timer scope : single timer for all computation operations, separated timers ...
- o Standardize "slowdown" timer algorithm including its association to a particular timer : authors of this document does not presume that the same algorithm must be used for all timers.

Using the same event sequence as in figure 2, we may expect fewer and/or shorter micro-loops using standardized implementations.

Time	Network Event	Router S events	Router E events
t0=0	Prefix DOWN		
10ms		Schedule PRC (in 150ms)	Schedule SPF (in 150ms)
160ms		PRC starts	PRC starts
161ms		PRC ends	
162ms		RIB/FIB starts	PRC ends
163ms			RIB/FIB starts
175ms		RIB/FIB ends	

```
   | 176ms  |                    |                  |   RIB/FIB ends   |
   |        |                    |                  |                  |
   | 200ms  |     Prefix UP      |                  |                  |
   | 212ms  |                    | Schedule PRC (in |                  |
   |        |                    |      150ms)      |                  |
   | 213ms  |                    |                  | Schedule PRC (in |
   |        |                    |                  |      150ms)      |
   |        |                    |                  |                  |
   |        |                    |                  |                  |
   | 370ms  |                    |    PRC starts    |    PRC starts    |
   | 372ms  |                    |     PRC ends     |                  |
   | 373ms  |                    |  RIB/FIB starts  |     PRC ends     |
   | 374ms  |                    |                  |  RIB/FIB starts  |
   | 383ms  |                    |   RIB/FIB ends   |                  |
   | 384ms  |                    |                  |   RIB/FIB ends   |
   |        |                    |                  |                  |
   | 400ms  |    Prefix DOWN     |                  |                  |
   | 410ms  |                    | Schedule PRC (in | Schedule PRC (in |
   |        |                    |      300ms)      |      300ms)      |
   |        |                    |                  |                  |
   |        |                    |                  |                  |
   |        |                    |                  |                  |
   |        |                    |                  |                  |
   | 710ms  |                    |    PRC starts    |    PRC starts    |
   | 711ms  |                    |     PRC ends     |     PRC ends     |
   | 712ms  |                    |  RIB/FIB starts  |                  |
   | 713ms  |                    |                  |  RIB/FIB starts  |
   | 716ms  |                    |   RIB/FIB ends   |   RIB/FIB ends   |
   |        |                    |                  |                  |
   | 1000ms |   S-D link DOWN    |                  |                  |
   | 1010ms |                    | Schedule SPF (in | Schedule SPF (in |
   |        |                    |      150ms)      |      150ms)      |
   |        |                    |                  |                  |
   |        |                    |                  |                  |
   | 1160ms |                    |    SPF starts    |                  |
   | 1161ms |                    |     SPF ends     |    SPF starts    |
   | 1162ms |   Micro-loop may   |  RIB/FIB starts  |     SPF ends     |
   |        |  start from here   |                  |                  |
   | 1163ms |                    |                  |  RIB/FIB starts  |
   | 1175ms |                    |   RIB/FIB ends   |                  |
   | 1177ms |  Micro-loop ends   |                  |   RIB/FIB ends   |
   +--------+--------------------+------------------+------------------+
```

Route computation event time scale

As displayed above, there could be some other parameters like router computation power, flooding timers that may also influence micro-loops. In the figure 5, we consider E to be a bit slower than S,

leading to micro-loop creation. Despite of this, we expect that by aligning implementations at least on SPF trigger and SPF delay, service provider may reduce number or duration of micro-loops.

7. Security Considerations

This document does not introduce any security consideration.

8. Acknowledgements

Authors would like to thank Mike Shand for his useful comments.

9. IANA Considerations

This document has no action for IANA.

10. References

10.1. 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, <http://www.rfc-editor.org/info/rfc1195>.

[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>.

10.2. Informative References

[I-D.ietf-rtgwg-microloop-analysis] Zinin, A., "Analysis and Minimization of Microloops in Link-state Routing Protocols", draft-ietf-rtgwg-microloop-analysis-01 (work in progress), October 2005.

[I-D.ietf-rtgwg-uloop-delay] Litkowski, S., Decraene, B., Filsfils, C., and P. Francois, "Microloop prevention by introducing a local convergence delay", draft-ietf-rtgwg-uloop-delay-00 (work in progress), November 2015.

[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, <http://www.rfc-editor.org/info/rfc6976>.

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