

Routing Area Working Group  
Internet-Draft  
Intended status: Standards Track  
Expires: August 19, 2013

S. Litkowski  
B. Decraene  
Orange  
P. Francois  
IMDEA Networks/Cisco Systems  
February 15, 2013

**Microloop prevention by introducing a local convergence delay**  
**draft-litkowski-rtgwg-uloop-delay-00**

Abstract

This document describes a mechanism for link-state routing protocols to prevent a subset of transient loops during topology changes. It introduces a two-step convergence by introducing a delay between the network wide convergence and the node advertising the failure. As the network wide convergence is delayed in case of down events, this mechanism can be used for planned maintenance or for unplanned outage provided a fast reroute mechanism is used in conjunction to convert a failure into a less urgent topology change.

Simulation using real network topologies and costs, with pathological convergence behaviour, have been performed. While the proposed mechanism does not provide a complete solution, it is simple and it solves an interesting fraction of the transient loops in the analyzed SP topologies.

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 August 19, 2013.

#### Copyright Notice

Copyright (c) 2013 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.



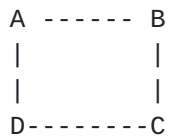
## Table of Contents

<a href="#">1.</a>	Introduction . . . . .	<a href="#">4</a>
<a href="#">2.</a>	Overview of the solution . . . . .	<a href="#">4</a>
<a href="#">3.</a>	Specification . . . . .	<a href="#">4</a>
<a href="#">3.1.</a>	Definitions . . . . .	<a href="#">4</a>
<a href="#">3.2.</a>	Current IGP reactions . . . . .	<a href="#">5</a>
<a href="#">3.3.</a>	Local delay . . . . .	<a href="#">5</a>
<a href="#">3.3.1.</a>	Link down event . . . . .	<a href="#">5</a>
<a href="#">3.3.2.</a>	Link up event . . . . .	<a href="#">6</a>
<a href="#">4.</a>	Use case . . . . .	<a href="#">6</a>
<a href="#">4.1.</a>	Applicable case . . . . .	<a href="#">6</a>
<a href="#">4.2.</a>	Non applicable case . . . . .	<a href="#">7</a>
<a href="#">5.</a>	Applicability . . . . .	<a href="#">7</a>
<a href="#">5.1.</a>	Topological applicability . . . . .	<a href="#">8</a>
<a href="#">6.</a>	Deployment considerations . . . . .	<a href="#">8</a>
<a href="#">7.</a>	Security Considerations . . . . .	<a href="#">9</a>
<a href="#">8.</a>	Acknowledgements . . . . .	<a href="#">9</a>
<a href="#">9.</a>	IANA Considerations . . . . .	<a href="#">9</a>
<a href="#">10.</a>	References . . . . .	<a href="#">10</a>
<a href="#">10.1.</a>	Normative References . . . . .	<a href="#">10</a>
<a href="#">10.2.</a>	Informative References . . . . .	<a href="#">10</a>
	Authors' Addresses . . . . .	<a href="#">10</a>



## 1. Introduction

In figure 1, upon link AD down event, for the destination A, if D updates its forwarding entry before C, a transient forwarding loop occurs between C and D. We have a similar loop for link up event, if C updates its forwarding entry A before D.



All the links have a metric of 1 except BC=5

Figure 1

## 2. Overview of the solution

This document defines a two-step convergence initiated by the router detecting the failure and advertising the topological changes in the IGP. This introduces a delay between the convergence of the local router compared to the network wide convergence. This delay is positive in case of "down" events and negative in case of "up" events.

This ordered convergence, is similar to the ordered FIB proposed defined in [[I-D.ietf-rtgwg-ordered-fib](#)], but limited to only one hop distance. The proposed mechanism reuses also some concept described in [[I-D.ietf-rtgwg-microloop-analysis](#)] with some limitation and improvements. As a consequence, it can only eliminate the loops between the node local to the event and its neighbors. In the SP topologies that were analyzed, this can avoid a high number of transient loops. On the other hand, as this mechanism is local to the router, it can be deployed incrementally with incremental benefit.

## 3. Specification

### 3.1. Definitions

This document will refer to the following existing IGP timers:

- o LSP\_GEN\_TIMER: to batch multiple local events in one single local LSP update. It is often associated with damping mechanism to slowdown reactions by incrementing the timer when multiple consecutive events are detected.



- o SPF\_TIMER: to batch multiple events in one single computation. It is often associated with damping mechanism to slowdown reactions by incrementing the timer when the IGP is instable.
- o IGP\_LDP\_SYNC\_TIMER: defined in [[RFC5443](#)] to give LDP some time to establish the session and learn the MPLS labels before the link is used.

This document introduces the following two new timers :

- o ULOOP\_DELAY\_DOWN\_TIMER: slowdown the network wide IGP convergence in case of link down events.
- o ULOOP\_DELAY\_UP\_TIMER: slowdown the local node convergence in case of link up events.

### **[3.2.](#) Current IGP reactions**

Upon a change of status on an adjacency/link, the existing behavior of the router advertising the event is the following:

1. UP/Down event is notified to IGP.
2. IGP processes the notification and postpones the reaction in LSP\_GEN\_TIMER msec.
3. Upon LSP\_GEN\_TIMER expiration, IGP updates its LSP/LSA and floods it.
4. SPF is scheduled in SPF\_TIMER msec.
5. Upon SPF\_TIMER expiration, SPF is computed and RIB/FIB are updated.

### **[3.3.](#) Local delay**

#### **[3.3.1.](#) Link down event**

Upon an adjacency/link down event, this document introduces a change in step 5 in order to delay the local convergence compared to the network wide convergence: the node SHOULD delay the forwarding entry updates by ULOOP\_DELAY\_DOWN\_TIMER. Such delay SHOULD only be introduced if all the LSDB modifications processed are only reporting down local events. Note that determining that all topological change are only local down events requires analyzing all modified LSP/LSA as a local link or node failure will typically be notified by multiple nodes. If a subsequent LSP/LSA is received/updated and a new SPF computation is triggered before the expiration of



ULOOP\_DELAY\_DOWN\_TIMER, then the same evaluation SHOULD be performed.

As as result of this addition, routers local to the failure will converge slower than remote routers.

### **3.3.2. Link up event**

Upon an adjacency/link up event, this document introduces the following change in step 3 where the node SHOULD:

- o Firstly build a LSP/LSA with the new adjacency but setting the metric to MAX\_METRIC. It SHOULD flood it but not compute the SPF at this time.
- o Then build the LSP/LSA with the target metric but SHOULD delay the flooding of this LSP/LSA by SPF\_TIMER + ULOOP\_DELAY\_UP\_TIMER. MAX\_METRIC is equal to LSInfinity (0xFFFF) for OSPF and  $2^{24}-2$  (0xFFFFFE) for IS-IS.
- o Then continue with next steps (SPF computation) without waiting for the expiration of the above timer. In other word, only the flooding of the LSA/LSP is delayed, not the local SPF computation.

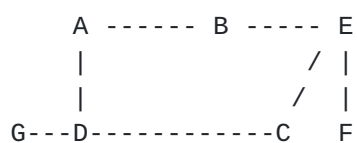
As as result of this addition, routers local to the failure will converge faster than remote routers.

If this mechanism is used in cooperation with "LDP IGP Synchronization" as defined in [RFC5443] then the mechanism defined in RFC 5443 is applied first, followed by the mechanism defined in this document. More precisely, the procedure defined in this document is applied once the LDP session is considered "fully operational" as per [RFC5443].

## **4. Use case**

As previously stated, the mechanism only avoids the forwarding loops on the links between the node local to the failure and its neighbor. Forwarding loops may still occur on other links.

### **4.1. Applicable case**



All the links have a metric of 1



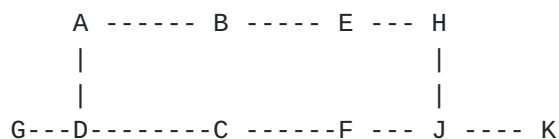
Figure 2

Let us consider the traffic from G to F. The primary path is G->D->C->E->F. When link CE fails, if C updates its forwarding entry for F before D, a transient loop occurs.

By implementing the mechanism defined in this document on C, when the CE link fails, C delays the update of his forwarding entry to F, in order to let some time for D to converge. When the timer expires on C, forwarding entry to F is updated. There is no transient forwarding loop on the link CD.

Note that C should implement a protection mechanism during the convergence delay in order to not increase the loss of traffic.

#### 4.2. Non applicable case



All the links have a metric of 1 except BE=15

Figure 3

Let us consider the traffic from G to K. The primary path is G->D->C->F->J->K. When the CF link fails, if C updates its forwarding entry to K before D, a transient loop occurs between C and D.

By implementing the mechanism defined in this document on C, when the link CF fails, C delays the update of his forwarding entry to K, letting time for D to converge. When the timer expires on C, forwarding entry to F is updated. There is no transient forwarding loop between C and D. However, a transient forwarding loop may still occur between D and A. In this scenario, this mechanism is not enough to address all the possible forwarding loops. However, it does not create additional traffic loss. Besides, in some cases -such as when the nodes update their FIB in the following order C, A, D, for example because the router A is quicker than D to converge- the mechanism may still avoid the forwarding loop that was occurring.

## 5. Applicability

Simulations have been run on multiple service provider topologies. So far, only link down event have been tested.



### 5.1. Topological applicability

Topology	Gain
T1	71%
T2	81%
T3	62%
T4	50%
T5	70%
T6	70%
T7	59%
T8	77%

Table 1: Number of Repair/Dst that may loop

We evaluated the efficiency of the mechanism on eight different service provider topologies (different network size, design). The benefit is displayed in the table above. The benefit is evaluated as follows:

- o We consider a tuple (link A-B, destination D, PLR S, backup nexthop N) as a loop if upon link A-B failure, the flow from a router S upstream from A (A could be considered as PLR also) to D may loop due to convergence time difference between S and one of his neighbor N.
- o We evaluate the number of potential loop tuples in normal conditions.
- o We evaluate the number of potential loop tuples using the same topological input but taking into account that S converges after N.
- o Gain is how much loops we succeed to suppress.

## 6. Deployment considerations

Transient forwarding loops have the following drawbacks :

- o Limit FRR efficiency : even if FRR is activated in 50msec, as soon as PLR has converged, traffic may be affected by a transient loop.
- o It may impact traffic not directly concerned by the failure (due to link congestion).



Our local delay proposal is a transient forwarding loop avoidance mechanism (like OFIB). Even if it does not prevent all transient loops to happen, the efficiency versus complexity comparison of the mechanism makes it a good solution.

Delaying convergence time is not an issue if we consider that the traffic is protected during the convergence. It would be up to the service provider to implement the local delay only for protected destinations or for all destinations. Considering that a service provider may implement the local delay for non protected destinations, it must be aware that delaying convergence will increase the loss duration on the affected link but at the same time, will prevent some other link to be congested. As a best practice, we recommend to activate the local delay only for protected destinations.

## **7. Security Considerations**

This document does not introduce change in term of IGP security. The operation is internal to the router. The local delay does not increase the attack vector as an attacker could only trigger this mechanism if he already has the ability to disable or enable an IGP link. The local delay does not increase the negative consequences as if an attacker has the ability to disable or enable an IGP link, it can already harm the network by creating instability and harm the traffic by creating forwarding packet loss and forwarding loss for the traffic crossing that link.

## **8. Acknowledgements**

We wish to thank the authors of [[I-D.ietf-rtgwg-ordered-fib](#)] for introducing the concept of ordered convergence: Mike Shand, Stewart Bryant, Stefano Previdi, Clarence Filsfils, Pierre Francois, and Olivier Bonaventure.

## **9. IANA Considerations**

This document has no actions for IANA.

## **10. References**



### **10.1. Normative References**

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [RFC5443] Jork, M., Atlas, A., and L. Fang, "LDP IGP Synchronization", [RFC 5443](#), March 2009.
- [RFC5715] Shand, M. and S. Bryant, "A Framework for Loop-Free Convergence", [RFC 5715](#), January 2010.

### **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-ordered-fib]  
Shand, M., Bryant, S., Previdi, S., Filsfils, C., Francois, P., and O. Bonaventure, "Framework for Loop-free convergence using oFIB", [draft-ietf-rtgwg-ordered-fib-09](#) (work in progress), January 2013.
- [I-D.ietf-rtgwg-remote-lfa]  
Bryant, S., Filsfils, C., Previdi, S., Shand, M., and S. Ning, "Remote LFA FRR", [draft-ietf-rtgwg-remote-lfa-01](#) (work in progress), December 2012.
- [RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", [RFC 3630](#), September 2003.
- [RFC6571] Filsfils, C., Francois, P., Shand, M., Decraene, B., Uttaro, J., Leymann, N., and M. Horneffer, "Loop-Free Alternate (LFA) Applicability in Service Provider (SP) Networks", [RFC 6571](#), June 2012.

### **Authors' Addresses**

Stephane Litkowski  
Orange

Email: [stephane.litkowski@orange.com](mailto:stephane.litkowski@orange.com)



Bruno Decraene  
Orange

Email: [bruno.decraene@orange.com](mailto:bruno.decraene@orange.com)

Pierre Francois  
IMDEA Networks/Cisco Systems

Email: [pierre.francois@imdea.org](mailto:pierre.francois@imdea.org)