   **Micro-loop prevention by introducing a local convergence delay**
               **draft-ietf-rtgwg-uloop-delay-08**

Abstract

   This document describes a mechanism for link-state routing protocols
   to prevent local transient forwarding loops in case of link failure.
   This mechanism proposes a two-step convergence by introducing a delay
   between the convergence of the node adjacent to the topology change
   and the network wide convergence.

   As this mechanism delays the IGP convergence it may only be used for
   planned maintenance or when fast reroute protects the traffic between
   the link failure time and the IGP convergence.

   The proposed mechanism is limited to the link down event in order to
   keep the mechanism simple.

   Simulations using real network topologies have been performed and
   show that local loops are a significant portion (>50%) of the total
   forwarding loops.

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

Table of Contents

## 1.  Acronyms

      FIB: Forwarding Information Base

      FRR: Fast ReRoute

      IGP: Interior Gateway Protocol

      LFA: Loop Free Alternate

      LSA: Link State Advertisement

      LSP: Link State Packet

      MRT: Maximum Redundant Trees

      OFIB: Ordered FIB

      PLSN: Path Locking via Safe Neighbor

      RIB: Routing Information Base

      RLFA: Remote Loop Free Alternate

      SPF: Shortest Path First

      TTL: Time To Live

## 2.  Introduction

   Micro-forwarding loops and some potential solutions are well
   described in [RFC5715].  This document describes a simple targeted
   mechanism that prevents micro-loops that are local to the failure.
   Based on network analysis, local failures make up a significant
   portion of the micro-forwarding loops.  A simple and easily
   deployable solution for these local micro-loops is critical because
   these local loops cause some traffic loss after a fast-reroute
   alternate has been used (see Section 3.1).

Consider the case in Figure 1 where S does not have an LFA (Loop Free
Alternate) to protect its traffic to D when the S-D link fails.  That
means that all non-D neighbors of S on the topology will send to S
any traffic destined to D; if a neighbor did not, then that neighbor
would be loop-free.  Regardless of the advanced fast-reroute (FRR)
technique used, when S converges to the new topology, it will send
its traffic to a neighbor that was not loop-free and thus cause a
local micro-loop.  The deployment of advanced fast-reroute techniques
motivates this simple router-local mechanism to solve this targeted
problem.  This solution can work with the various techniques
described in [RFC5715].

```
     D ------ C
     |        |
     |        | 5
     |        |
     S ------ B
```
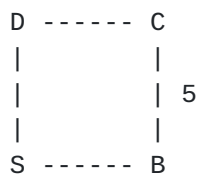
     Figure 1

In the Figure 1, all links have a metric of 1 except B-C which has a
metric of 5.  When S-D fails, a transient forwarding loop may appear
between S and B if S updates its forwarding entry to D before B does.

## 3.  Transient forwarding loops side effects

Even if they are very limited in duration, transient forwarding loops
may cause significant network damage.

## 3.1.  Fast reroute inefficiency

```
        D
     1 |
       |    1
       A ------ B
       |        |     ^
    10 |        | 5   | T
       |        |     |
       E--------C
       |    1
     1 |
       S
```
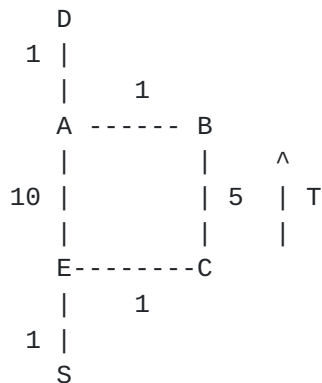
     Figure 2 - RSVP-TE FRR case

In the Figure 2, we consider an IP/LDP routed network.  An RSVP-TE
tunnel T, provisioned on C and terminating on B, is used to protect

the traffic against C-B link failure (the IGP shortcut feature,
defined in [RFC3906], is activated on C ).  The primary path of T is
C->B and FRR is activated on T providing an FRR bypass or detour
using path C->E->A->B.  On router C, the next hop to D is the tunnel
T thanks to the IGP shortcut.  When C-B link fails:

1.  C detects the failure, and updates the tunnel path using a
    preprogrammed FRR path.  The traffic path from S to D becomes:
    S->E->C->E->A->B->A->D.

2.  In parallel, on router C, both the IGP convergence and the TE
    tunnel convergence (tunnel path recomputation) are occurring:

    *  The Tunnel T path is recomputed and now uses C->E->A->B.

    *  The IGP path to D is recomputed and now uses C->E->A->D.

3.  On C, the tail-end of the TE tunnel (router B) is no longer on
    the shortest-path tree (SPT) to D, so C does not continue to
    encapsulate the traffic to D using the tunnel T and updates its
    forwarding entry to D using the nexthop E.

If C updates its forwarding entry to D before router E, there would
be a transient forwarding loop between C and E until E has converged.

The table 1 below describes a theoretical sequence of events
happening when the B-C link fails.  This theoretical sequence of
events should only be read as an example.

| Network condition | Time | Router C events | Router E events |
|-----------|------------|-----------------|-----------------------|
| S->D Traffic OK | | | |
| S->D Traffic lost | t0 | Link B-C fails | Link B-C fails |
| | t0+20msec | C detects the failure | |
| S->D Traffic OK | t0+40msec | C activates FRR | |

| | | t0+50msec | C updates its local LSP/LSA | |
|---|---|---|---|---|
| | | t0+60msec | C schedules SPF (100ms) | |
| | | t0+70msec | C floods its local updated LSP/LSA | |
| | | t0+87msec | | E receives LSP/LSA from C and schedules SPF (100ms) |
| | | t0+117msec | | E floods LSP/LSA from C |
| | | t0+160msec | C computes SPF | |
| | | t0+165msec | C starts updating its RIB/FIB | |
| | | t0+193msec | | E computes SPF |
| | | t0+199msec | | E starts updating its RIB/FIB |
| S->D Traffic lost | | t0+255msec | C updates its RIB/FIB for D | |
| | | t0+340msec | C convergence ends | |
| S->D Traffic OK | | t0+443msec | | E updates its RIB/FIB for D |
| | | t0+470msec | | E convergence ends |

Table 1 - Route computation event time scale

The issue described here is completely independent of the fast-reroute mechanism involved (TE FRR, LFA/rLFA, MRT ...) when the primary path uses hop-by-hop routing.  The protection enabled by fast-reroute is working perfectly, but ensures a protection, by

definition, only until the PLR has converged (as soon as the PLR has
converged, it replaces its FRR path by a new primary path).  When
implementing FRR, a service provider wants to guarantee a very
limited loss of connectivity time.  The previous example shows that
the benefit of FRR may be completely lost due to a transient
forwarding loop appearing when PLR has converged.  Delaying FIB
updates after the IGP convergence may allow to keep the fast-reroute
path until the neighbors have converged and preserves the customer
traffic.

## 3.2.  Network congestion

```
           1
      D ------ C
      |        |
    1 |        | 5
      |        |
 A -- S ------ B
   / |     1
  F   E
```
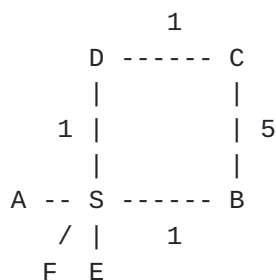
        Figure 3

In the figure above, as presented in Section 2, when the link S-D
fails, a transient forwarding loop may appear between S and B for
destination D.  The traffic on the S-B link will constantly increase
due to the looping traffic to D.  Depending on the TTL of the
packets, the traffic rate destined to D, and the bandwidth of the
link, the S-B link may become congested in a few hundreds of
milliseconds and will stay congested until the loop is eliminated.

The congestion introduced by transient forwarding loops is
problematic as it can affect traffic that is not directly affected by
the failing network component.  In the example, the congestion of the
S-B link will impact some customer traffic that is not directly
affected by the failure: e.g.  A to B, F to B, E to B.  Class of
service may mitigate the congestion for some traffic.  However, some
traffic not directly affected by the failure will still be dropped as
a router is not able to distinguish the looping traffic from the
normally forwarded traffic.

## 4.  Overview of the solution

This document defines a two-step convergence initiated by the router
detecting a failure and advertising the topological changes in the
IGP.  This introduces a delay between network-wide convergence and
the convergence of the local router.

The proposed solution is limited to local link down events in order
to keep the solution simple.

This ordered convergence is similar to the ordered FIB proposed
defined in [RFC6976], but it is limited to only a "one hop" distance.
As a consequence, it is more simple and becomes a local-only feature
that does not require interoperability.  This benefit comes with the
limitation of eliminating transient forwarding loops involving the
local router only.  The proposed mechanism also reuses some concepts
described in [I-D.ietf-rtgwg-microloop-analysis].

## 5.  Specification

### 5.1.  Definitions

This document will refer to the following existing IGP timers.  These
timers may be standardized or implemented as a vendor specific local
feature.

o  LSP_GEN_TIMER: The delay used to batch multiple local events in
   one single local LSP/LSA update.  In IS-IS, this timer is defined
   as minimumLSPGenerationInterval in [ISO10589].  In OSPF version 2,
   this timer is defined as MinLSInterval in [RFC2328].  It is often
   associated with a vendor specific damping mechanism to slow down
   reactions by incrementing the timer when multiple consecutive
   events are detected.

o  SPF_DELAY: The delay between the first IGP event triggering a new
   routing table computation and the start of that routing table
   computation.  It is often associated with a damping mechanism to
   slow down reactions by incrementing the timer when the IGP becomes
   unstable.  As an example, [I-D.ietf-rtgwg-backoff-algo] defines a
   standard SPF (Shortest Path First) delay algorithm.

This document introduces the following new timer:

o  ULOOP_DELAY_DOWN_TIMER: used to slow down the local node
   convergence in case of link down events.

### 5.2.  Regular IGP reaction

Upon a change of the status of an adjacency/link, the regular IGP
convergence behavior of the router advertising the event involves the
following main steps:

1.  IGP is notified of the Up/Down event.

2.  The IGP processes the notification and postpones the reaction for
    LSP_GEN_TIMER msec.

3.  Upon LSP_GEN_TIMER expiration, the IGP updates its LSP/LSA and
    floods it.

4.  The SPF computation is scheduled in SPF_DELAY msec.

5.  Upon SPF_DELAY timer expiration, the SPF is computed, then the
    RIB and FIB are updated.

## 5.3.  Local events

The mechanism described in this document assumes that there has been
a single link failure as seen by the IGP area/level.  If this
assumption is violated (e.g. multiple links or nodes failed), then
regular IP convergence must be applied (as described in Section 5.2).

To determine if the mechanism can be applicable or not, an
implementation SHOULD implement logic to correlate the protocol
messages (LSP/LSA) received during the SPF scheduling period in order
to determine the topology changes that occured.  This is necessary as
multiple protocol messages may describe the same topology change and
a single protocol message may describe multiple topology changes.  As
a consequence, determining a particular topology change MUST be
independent of the order of reception of those protocol messages.
How the logic works is left to the implementation.

Using this logic, if an implementation determines that the associated
topology change is a single local link failure, then the router MAY
use the mechanism described in this document, otherwise the regular
IP convergence MUST be used.

Example:

```
     +--- E ----+--------+
     |          |        |
A ---- B -------- C ------ D
```
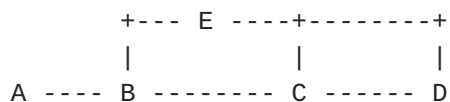
     Figure 4

Let router B be the computing router when the link B-C fails.  B
updates its local LSP/LSA describing the link B->C as down, C does
the same, and both start flooding their updated LSP/LSAs.  During the
SPF_DELAY period, B and C learn all the LSPs/LSAs to consider.  B
sees that C is flooding an advertisement that indicates that a link
is down, and B is the other end of that link.  B determines that B
and C are describing the same single event.  Since B receives no

other changes, B can determine that this is a local link failure and
may decide to activate the mechanism described in this document.

## 5.4.  Local delay for link down

Upon an adjacency/link down event, this document introduces a change
in step 5 (Section 5.2) in order to delay the local convergence
compared to the network wide convergence.  The new step 5 is
described below:

> 5.  Upon SPF_DELAY timer expiration, the SPF is computed.  If the
> condition of a single local link-down event has been met, then an
> update of the RIB and the FIB MUST be delayed for
> ULOOP_DELAY_DOWN_TIMER msecs.  Otherwise, the RIB and FIB SHOULD
> be updated immediately.

If a new convergence occurs while ULOOP_DELAY_DOWN_TIMER is running,
ULOOP_DELAY_DOWN_TIMER is stopped and the RIB/FIB SHOULD be updated
as part of the new convergence event.

As a result of this addition, routers local to the failure will
converge slower than remote routers.  Hence it SHOULD only be done
for a non-urgent convergence, such as for administrative de-
activation (maintenance) or when the traffic is protected by fast-
reroute.

## 6.  Applicability

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

## 6.1.  Applicable case: local loops
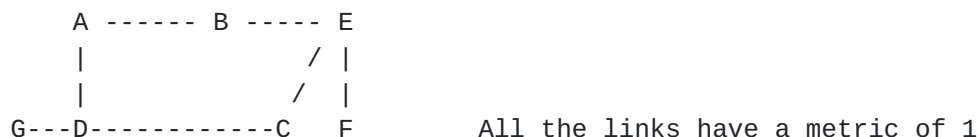
```
    A ------ B ----- E
    |             / |
    |            /  |
    |           /   |
  G---D-----------C   F        All the links have a metric of 1

      Figure 5
```

Let us consider the traffic from G to F.  The primary path is
G->D->C->E->F.  When link C-E fails, if C updates its forwarding
entry for F before D, a transient loop occurs.  This is sub-optimal
as C has FRR enabled and it breaks the FRR forwarding while all
upstream routers are still forwarding the traffic to itself.

By implementing the mechanism defined in this document on C, when the C-E link fails, C delays the update of its forwarding entry to F, in order to allow some time for D to converge.  FRR on C keeps protecting the traffic during this period.  When the timer expires on C, its forwarding entry to F is updated.  There is no transient forwarding loop on the link C-D.

## 6.2.  Non applicable case: remote loops

```
      A ------ B ----- E --- H
      |                   |
      |                   |
  G---D--------C ------F --- J ---- K
```
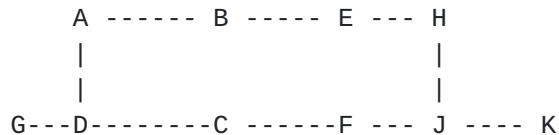
 All the links have a metric of 1 except BE=15

      Figure 6

Let us consider the traffic from G to K.  The primary path is G->D->C->F->J->K.  When the C-F 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 C-F fails, C delays the update of its forwarding entry to K, allowing time for D to converge.  When the timer expires on C, its 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 would have otherwise occurred.

## 7.  Simulations

Simulations have been run on multiple service provider topologies.

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

             Table 2 - 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 neighbors 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  The gain is how many loops (both remote and local) we succeed to
      suppress.

   On topology 1, 71% of the transient forwarding loops created by the
   failure of any link are prevented by implementing the local delay.
   The analysis shows that all local loops are prevented and only remote
   loops remain.

8.  Deployment considerations

   Transient forwarding loops have the following drawbacks:

   o  They limit FRR efficiency: even if FRR is activated within 50msec,
      as soon as PLR has converged, the traffic may be affected by a
      transient loop.

o   They may impact traffic not directly affected by the failure (due
    to link congestion).

This local delay proposal is a transient forwarding loop avoidance
mechanism (like OFIB).  Even if it only addresses local transient
loops, the efficiency versus complexity comparison of the mechanism
makes it a good solution.  It is also incrementally deployable with
incremental benefits, which makes it an attractive option both for
vendors to implement and service providers to deploy.  Delaying the
convergence time is not an issue if we consider that the traffic is
protected during the convergence.

The ULOOP_DELAY_DOWN_TIMER value should be set according to the
maximum IGP convergence time observed in the network (usually
observed in the slowest node).

The proposed mechanism is limited to link down events.  When a link
goes down, it eventually goes back up.  As a consequence, with the
proposed mechanism deployed, only the link down event will be
protected against transient forwarding loops while the link up event
will not.  If the operator wants to limit the impact of the transient
forwarding loops during the link up event, it should take care of
using specific procedures to bring the link back online.  As
examples, the operator can decide to put back the link online out of
business hours or it can use some incremental metric changes to
prevent loops (as proposed in [RFC5715]).

9.  Examples

We will consider the following figure for the associated examples :

```
      D
    1 |          F----X
      |     1    |
      A ------ B
      |          |
   10 |          | 5
      |          |
      E--------C
      |     1
    1 |
      S
```
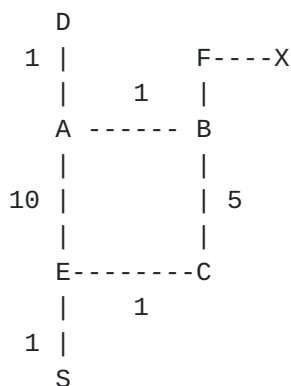
         Figure 7

The network above is considered to have a convergence time about 1
second, so ULOOP_DELAY_DOWN_TIMER will be adjusted to this value.  We
also consider that FRR is running on each node.

## 9.1.  Local link down

   The table 3 describes the events and associated timing that happen on
   router C and E when link B-C goes down.  It is based on a theoretical
   sequence of event that should only been read as an example.  As C
   detects a single local event corresponding to a link down (its LSP +
   LSP from B received), it applies the local delay down behavior and no
   microloop is formed.

```
+-----------+------------+-----------------+--------------------+
| Network   |    Time    | Router C events |  Router E events   |
| condition |            |                 |                    |
+-----------+------------+-----------------+--------------------+
|   S->D    |            |                 |                    |
|  Traffic  |            |                 |                    |
|    OK     |            |                 |                    |
|           |            |                 |                    |
|   S->D    |     t0     |  Link B-C fails |   Link B-C fails   |
|  Traffic  |            |                 |                    |
|   lost    |            |                 |                    |
|           |            |                 |                    |
|           | t0+20msec  |  C detects the  |                    |
|           |            |     failure     |                    |
|           |            |                 |                    |
|   S->D    | t0+40msec  |  C activates FRR|                    |
|  Traffic  |            |                 |                    |
|    OK     |            |                 |                    |
|           |            |                 |                    |
|           | t0+50msec  |   C updates its |                    |
|           |            |   local LSP/LSA |                    |
|           |            |                 |                    |
|           | t0+60msec  | C schedules SPF |                    |
|           |            |      (100ms)    |                    |
|           |            |                 |                    |
|           | t0+67msec  |    C receives   |                    |
|           |            |  LSP/LSA from B |                    |
|           |            |                 |                    |
|           | t0+70msec  |   C floods its  |                    |
|           |            |  local updated  |                    |
|           |            |     LSP/LSA     |                    |
|           |            |                 |                    |
|           | t0+87msec  |                 | E receives LSP/LSA |
|           |            |                 | from C and schedules|
|           |            |                 |    SPF (100ms)     |
|           |            |                 |                    |
|           | t0+117msec |                 |  E floods LSP/LSA  |
|           |            |                 |      from C        |
|           |            |                 |                    |
|           | t0+160msec |  C computes SPF |                    |
|           |            |                 |                    |
|           | t0+165msec |   C delays its  |                    |
|           |            |  RIB/FIB update |                    |
|           |            |     (1 sec)     |                    |
|           |            |                 |                    |
|           | t0+193msec |                 |   E computes SPF   |
|           |            |                 |                    |
|           | t0+199msec |                 |  E starts updating |
```

```
|         |            |                 |     its RIB/FIB     |
|         |            |                 |                     |
|         | t0+443msec |                 |    E updates its    |
|         |            |                 |    RIB/FIB for D     |
|         |            |                 |                     |
|         | t0+470msec |                 | E convergence ends  |
|         |            |                 |                     |
|         | t0+1165msec|    C starts     |                     |
|         |            |   updating its  |                     |
|         |            |     RIB/FIB     |                     |
|         |            |                 |                     |
|         | t0+1255msec|  C updates its  |                     |
|         |            |   RIB/FIB for D |                     |
|         |            |                 |                     |
|         | t0+1340msec|  C convergence  |                     |
|         |            |       ends      |                     |
+---------+------------+-----------------+---------------------+
```

Table 3 - Route computation event time scale

Similarly, upon B-C link down event, if LSP/LSA from B is received
before C detects the link failure, C will apply the route update
delay if the local detection is part of the same SPF run.  The table
4 describes the associated theoretical sequence of events.  It should
only been read as an example.

| Network condition | Time | Router C events | Router E events |
|-----------|------------|-----------------|---------------------|
| S->D Traffic OK | | | |
| S->D Traffic lost | t0 | Link B-C fails | Link B-C fails |
| | t0+32msec | C receives LSP/LSA from B | |
| | t0+33msec | C schedules SPF (100ms) | |
| | t0+50msec | C detects the failure | |
| S->D | t0+55msec | C activates FRR | |

```
|  Traffic   |            |                  |                      |
|    OK      |            |                  |                      |
|            |            |                  |                      |
|            | t0+55msec  |  C updates its   |                      |
|            |            |  local LSP/LSA   |                      |
|            |            |                  |                      |
|            | t0+70msec  |  C floods its    |                      |
|            |            |  local updated   |                      |
|            |            |     LSP/LSA      |                      |
|            |            |                  |                      |
|            | t0+87msec  |                  |  E receives LSP/LSA  |
|            |            |                  | from C and schedules |
|            |            |                  |    SPF (100ms)       |
|            |            |                  |                      |
|            | t0+117msec |                  |  E floods LSP/LSA    |
|            |            |                  |       from C         |
|            |            |                  |                      |
|            | t0+160msec |  C computes SPF  |                      |
|            |            |                  |                      |
|            | t0+165msec |   C delays its   |                      |
|            |            |  RIB/FIB update  |                      |
|            |            |     (1 sec)      |                      |
|            |            |                  |                      |
|            | t0+193msec |                  |    E computes SPF    |
|            |            |                  |                      |
|            | t0+199msec |                  |   E starts updating  |
|            |            |                  |      its RIB/FIB     |
|            |            |                  |                      |
|            | t0+443msec |                  |    E updates its     |
|            |            |                  |    RIB/FIB for D     |
|            |            |                  |                      |
|            | t0+470msec |                  | E convergence ends   |
|            |            |                  |                      |
|            | t0+1165msec|     C starts     |                      |
|            |            |   updating its   |                      |
|            |            |     RIB/FIB      |                      |
|            |            |                  |                      |
|            | t0+1255msec|  C updates its   |                      |
|            |            |  RIB/FIB for D   |                      |
|            |            |                  |                      |
|            | t0+1340msec|  C convergence   |                      |
|            |            |       ends       |                      |
+-----------+------------+------------------+----------------------+
```

             Table 4 - Route computation event time scale

9.2.  Local and remote event

   The table 5 describes the events and associated timing that happen on
   router C and E when link B-C goes down, in addition F-X link will
   fail in the same time window.  C will not apply the local delay
   because a non local topology change is also received.  The table 5 is
   based on a theoretical sequence of event that should only been read
   as an example.

```
+-----------+-----------+----------------+------------------------+
| Network   |   Time    | Router C events |    Router E events    |
| condition |           |                |                        |
+-----------+-----------+----------------+------------------------+
|   S->D    |           |                |                        |
|  Traffic  |           |                |                        |
|    OK     |           |                |                        |
|           |           |                |                        |
|   S->D    |    t0     | Link B-C fails |     Link B-C fails     |
|  Traffic  |           |                |                        |
|   lost    |           |                |                        |
|           |           |                |                        |
|           | t0+20msec |  C detects the |                        |
|           |           |     failure    |                        |
|           |           |                |                        |
|           | t0+36msec |  Link F-X fails |     Link F-X fails     |
|           |           |                |                        |
|   S->D    | t0+40msec | C activates FRR |                        |
|  Traffic  |           |                |                        |
|    OK     |           |                |                        |
|           |           |                |                        |
|           | t0+50msec |  C updates its |                        |
|           |           |  local LSP/LSA |                        |
|           |           |                |                        |
|           | t0+54msec |   C receives   |                        |
|           |           | LSP/LSA from F |                        |
|           |           | and floods it  |                        |
|           |           |                |                        |
|           | t0+60msec | C schedules SPF |                       |
|           |           |     (100ms)    |                        |
|           |           |                |                        |
|           | t0+67msec |   C receives   |                        |
|           |           | LSP/LSA from B |                        |
|           |           |                |                        |
|           | t0+69msec |                | E receives LSP/LSA     |
|           |           |                | from F, floods it and  |
|           |           |                | schedules SPF (100ms)  |
|           |           |                |                        |
|           | t0+70msec |   C floods its |                        |
```

| | | | |
|------|------|------|------|
| | | local updated LSP/LSA | |
| | t0+87msec | | E receives LSP/LSA from C |
| | t0+117msec | | E floods LSP/LSA from C |
| | t0+160msec | C computes SPF | |
| | t0+165msec | C starts updating its RIB/FIB (NO DELAY) | |
| | t0+170msec | | E computes SPF |
| | t0+173msec | | E starts updating its RIB/FIB |
| S->D Traffic lost | t0+365msec | C updates its RIB/FIB for D | |
| S->D Traffic OK | t0+443msec | | E updates its RIB/FIB for D |
| | t0+450msec | C convergence ends | |
| | t0+470msec | | E convergence ends |

Table 5 - Route computation event time scale

## 9.3. Aborting local delay

The table 6 describes the events and associated timing that happen on
router C and E when link B-C goes down.  In addition, we consider
what happens when F-X link fails during local delay of the FIB
update.  C will first apply the local delay, but when the new event
happens, it will fall back to the standard convergence mechanism
without further delaying route insertion.  In this example, we
consider a ULOOP_DELAY_DOWN_TIMER configured to 2 seconds.  The table

   6 is based on a theoretical sequence of event that should only been
   read as an example.

```
+-----------+------------+------------------+---------------------+
| Network   |    Time    | Router C events  |  Router E events    |
| condition |            |                  |                     |
+-----------+------------+------------------+---------------------+
|   S->D    |            |                  |                     |
|  Traffic  |            |                  |                     |
|    OK     |            |                  |                     |
|           |            |                  |                     |
|   S->D    |     t0     |   Link B-C fails |   Link B-C fails    |
|  Traffic  |            |                  |                     |
|    lost   |            |                  |                     |
|           |            |                  |                     |
|           | t0+20msec  |   C detects the  |                     |
|           |            |      failure     |                     |
|           |            |                  |                     |
|   S->D    | t0+40msec  |  C activates FRR |                     |
|  Traffic  |            |                  |                     |
|    OK     |            |                  |                     |
|           |            |                  |                     |
|           | t0+50msec  |   C updates its  |                     |
|           |            |   local LSP/LSA  |                     |
|           |            |                  |                     |
|           | t0+60msec  |  C schedules SPF |                     |
|           |            |      (100ms)     |                     |
|           |            |                  |                     |
|           | t0+67msec  |    C receives    |                     |
|           |            |  LSP/LSA from B  |                     |
|           |            |                  |                     |
|           | t0+70msec  |    C floods its  |                     |
|           |            |   local updated  |                     |
|           |            |      LSP/LSA     |                     |
|           |            |                  |                     |
|           | t0+87msec  |                  | E receives LSP/LSA  |
|           |            |                  | from C and schedules|
|           |            |                  |     SPF (100ms)     |
|           |            |                  |                     |
|           | t0+117msec |                  |   E floods LSP/LSA  |
|           |            |                  |       from C        |
|           |            |                  |                     |
|           | t0+160msec |   C computes SPF |                     |
|           |            |                  |                     |
|           | t0+165msec |   C delays its   |                     |
|           |            |  RIB/FIB update (2|                    |
|           |            |       sec)       |                     |
|           |            |                  |                     |
|           | t0+193msec |                  |    E computes SPF   |
|           |            |                  |                     |
|           | t0+199msec |                  |  E starts updating  |
```

| | | | | its RIB/FIB |
| | | | | |
| | t0+254msec | Link F-X fails | Link F-X fails |
| | | | | |
| | t0+300msec | C receives LSP/LSA from F and floods it | |
| | | | | |
| | t0+303msec | C schedules SPF (200ms) | |
| | | | | |
| | t0+312msec | E receives LSP/LSA from F and floods it | |
| | | | | |
| | t0+313msec | E schedules SPF (200ms) | |
| | | | | |
| | t0+502msec | C computes SPF | |
| | | | | |
| | t0+505msec | C starts updating its RIB/FIB (NO DELAY) | |
| | | | | |
| | t0+514msec | | E computes SPF |
| | | | | |
| | t0+519msec | | E starts updating its RIB/FIB |
| | | | | |
| S->D Traffic lost | t0+659msec | C updates its RIB/FIB for D | |
| | | | | |
| S->D Traffic OK | t0+778msec | | E updates its RIB/FIB for D |
| | | | | |
| | t0+781msec | C convergence ends | |
| | | | | |
| | t0+810msec | | E convergence ends |

Table 6 - Route computation event time scale

## 10.  Comparison with other solutions

As stated in Section 4, the proposed solution reuses some concepts
already introduced by other IETF proposals but tries to find a
tradeoff between efficiency and simplicity.  This section tries to
compare behaviors of the solutions.

### 10.1.  PLSN

PLSN ([I-D.ietf-rtgwg-microloop-analysis]) describes a mechanism
where each node in the network tries to avoid transient forwarding
loops upon a topology change by always keeping traffic on a loop-free
path for a defined duration (locked path to a safe neighbor).  The
locked path may be the new primary nexthop, another neighbor, or the
old primary nexthop depending how the safety condition is satisfied.

PLSN does not solve all transient forwarding loops (see
[I-D.ietf-rtgwg-microloop-analysis] Section 4 for more details).

Our solution reuses some concept of PLSN but in a more simple
fashion:

o  PLSN has three different behaviors: keep using old nexthop, use
   new primary nexthop if it is safe, or use another safe nexthop,
   while the proposed solution only has one: keep using the current
   nexthop (old primary, or already activated FRR path).

o  PLSN may cause some damage while using a safe nexthop which is not
   the new primary nexthop in case the new safe nexthop does not
   provide enough bandwidth (see [RFC7916]).  This solution may not
   experience this issue as the service provider may have control on
   the FRR path being used preventing network congestion.

o  PLSN applies to all nodes in a network (remote or local changes),
   while the proposed mechanism applies only on the nodes connected
   to the topology change.

### 10.2.  OFIB

OFIB ([RFC6976]) describes a mechanism where the convergence of the
network upon a topology change is ordered in order to prevent
transient forwarding loops.  Each router in the network must deduce
the failure type from the LSA/LSP received and computes/applies a
specific FIB update timer based on the failure type and its rank in
the network considering the failure point as root.

This mechanism allows to solve all the transient forwarding loop in a network at the price of introducing complexity in the convergence process that may require a strong monitoring by the service provider.

Our solution reuses the OFIB concept but limits it to the first hop that experiences the topology change.  As demonstrated, the mechanism proposed in this document allows to solve all the local transient forwarding loops that represents an high percentage of all the loops. Moreover limiting the mechanism to one hop allows to keep the network-wide convergence behavior.

## 11.  Implementation Status

At this time, there are three different implementations of this mechanism.

o  Implementation 1:

   *  Organization: Cisco

   *  Implementation name: Local Microloop Protection

   *  Operating system: IOS-XE

   *  Level of maturity: production release

   *  Coverage: all the specification is implemented

   *  Protocols supported: ISIS and OSPF

   *  Implementation experience: tested in lab and works as expected

   *  Comment: the feature gives the ability to choose to apply the
      delay to FRR protected entry only

   *  Report last update: 10-11-2017

o  Implementation 2:

   *  Organization: Cisco

   *  Implementation name: Local Microloop Protection

   *  Operating system: IOS-XR

   *  Level of maturity: deployed

   *  Coverage: all the specification is implemented

   *  Protocols supported: ISIS and OSPF

   *  Implementation experience: deployed and works as expected

   *  Comment: the feature gives the ability to choose to apply the
      delay to FRR protected entry only

   *  Report last update: 10-11-2017

   o  Implementation 3:

   *  Organization: Juniper Networks

   *  Implementation name: Microloop avoidance when IS-IS link fails

   *  Operating system: JUNOS

   *  Level of maturity: deployed (hidden command)

   *  Coverage: all the specification is implemented

   *  Protocols supported: ISIS only

   *  Implementation experience: deployed and works as expected

   *  Comment: the feature applies to all the ISIS routes

   *  Report last update: 10-11-2017

## [12].  Security Considerations

   This document does not introduce any change in term of IGP security.
   The operation is internal to the router.  The local delay does not
   increase the number of attack vectors as an attacker could only
   trigger this mechanism if he already has be ability to disable or
   enable an IGP link.  The local delay does not increase the negative
   consequences.  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.

## [13].  Acknowledgements

   We would like to thanks the authors of [RFC6976] for introducing the
   concept of ordered convergence: Mike Shand, Stewart Bryant, Stefano
   Previdi, and Olivier Bonaventure.

## 14.  IANA Considerations

   This document has no actions for IANA.

## 15.  References

### 15.1.  Normative References

   [ISO10589]
             "Intermediate System to Intermediate System intra-domain
             routeing information exchange protocol for use in
             conjunction with the protocol for providing the
             connectionless-mode network service (ISO 8473)",
             ISO 10589, 2002.

   [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>.

### 15.2.  Informative References

   [I-D.ietf-rtgwg-backoff-algo]
             Decraene, B., Litkowski, S., Gredler, H., Lindem, A.,
             Francois, P., and C. Bowers, "SPF Back-off algorithm for
             link state IGPs", draft-ietf-rtgwg-backoff-algo-05 (work
             in progress), May 2017.

   [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.

   [RFC3906]  Shen, N. and H. Smit, "Calculating Interior Gateway
             Protocol (IGP) Routes Over Traffic Engineering Tunnels",
             RFC 3906, DOI 10.17487/RFC3906, October 2004,
             <https://www.rfc-editor.org/info/rfc3906>.

   [RFC5715]  Shand, M. and S. Bryant, "A Framework for Loop-Free
             Convergence", RFC 5715, DOI 10.17487/RFC5715, January
             2010, <https://www.rfc-editor.org/info/rfc5715>.

   [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>.

   [RFC7916]  Litkowski, S., Ed., Decraene, B., Filsfils, C., Raza, K.,
              Horneffer, M., and P. Sarkar, "Operational Management of
              Loop-Free Alternates", RFC 7916, DOI 10.17487/RFC7916,
              July 2016, <https://www.rfc-editor.org/info/rfc7916>.

Authors' Addresses

   Stephane Litkowski
   Orange


   Email: stephane.litkowski@orange.com


   Bruno Decraene
   Orange

   Email: bruno.decraene@orange.com


   Clarence Filsfils
   Cisco Systems

   Email: cfilsfil@cisco.com


   Pierre Francois
   Individual

   Email: pfrpfr@gmail.com