Internet Engineering Task Force (IETF)
Request for Comments: 5715
Category: Informational
ISSN: 2070-1721

M. Shand
S. Bryant
Cisco Systems

January 2010

# A Framework for Loop-Free Convergence

## 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 framework provides a summary of the causes and consequences of micro-loops and enables the reader to form a judgement on whether micro-looping is an issue that needs to be addressed in specific networks. It also provides a survey of the currently proposed mechanisms that may be used to prevent or to suppress the formation of micro-loops when an IP or MPLS network undergoes topology change due to failure, repair, or management action. When sufficiently fast convergence is not available and the topology is susceptible to micro-loops, use of one or more of these mechanisms may be desirable.

## Status of This Memo

Table of Contents

## 1. Introduction

When there is a change to the network topology (due to the failure or restoration of a link or router, or as a result of management action), the routers need to converge on a common view of the new topology and the paths to be used for forwarding traffic to each destination. During this process, referred to as a routing transition, packet delivery between certain source/destination pairs may be disrupted. This occurs due to the time it takes for the topology change to be propagated around the network together with the time it takes each individual router to determine and then update the forwarding information base (FIB) for the affected destinations. During this transition, packets may be lost due to the continuing attempts to use the failed component and due to forwarding loops. Forwarding loops arise due to the inconsistent FIBs that occur as a result of the difference in time taken by routers to execute the transition process. This is a problem that may occur in both IP networks and MPLS networks that use the label distribution protocol (LDP) [RFC5036] as the label switched path (LSP) signaling protocol.

The service failures caused by routing transitions are largely hidden by higher-level protocols that retransmit the lost data. However, new Internet services could emerge that are more sensitive to the packet disruption that occurs during a transition. To make the transition transparent to their users, these services would require a short routing transition. Ideally, routing transitions would be completed in zero time with no packet loss.

Regardless of how optimally the mechanisms involved have been designed and implemented, it is inevitable that a routing transition will take some minimum interval that is greater than zero. This has led to the development of a traffic engineering (TE) fast-reroute mechanism for MPLS [RFC4090]. Alternative mechanisms that might be deployed in an MPLS network or an IP network are current work items in the IETF [RFC5714]. The repair mechanism may, however, be disrupted by the formation of micro-loops during the period between the time when the failure is announced and the time when all FIBs have been updated to reflect the new topology.

One method of mitigating the effects of micro-loops is to ensure that the network reconverges in a sufficiently short time that these effects are inconsequential. Another method is to design the network topology to minimise or even eliminate the possibility of micro-loops.

The propensity to form micro-loops is highly topology dependent, and algorithms are available to identify which links in a network are subject to micro-looping. In topologies that are critically

susceptible to the formation of micro-loops, there is little point in introducing new mechanisms to provide fast reroute without also deploying mechanisms that prevent the disruptive effects of micro-loops. Unless micro-loop prevention is used in these topologies, packets may not reach the repair and micro-looping packets may cause congestion, resulting in further packet loss.

The disruptive effect of micro-loops is not confined to periods when there is a component failure. Micro-loops can, for example, form when a component is put back into service following repair. Micro-loops can also form as a result of a network-maintenance action such as adding a new network component, removing a network component, or modifying a link cost.

This framework provides a summary of the causes and consequences of micro-loops and enables the reader to form a judgement on whether micro-looping is an issue that needs to be addressed in specific networks. It also provides a survey of the currently proposed micro-loop mitigation mechanisms. When sufficiently fast convergence is not available and the topology is susceptible to micro-loops, use of one or more of these mechanisms may be desirable.

## 2. The Nature of Micro-Loops

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.

Micro-loops may form during the periods when a network is re-converging following ANY topology change and are caused by inconsistent FIBs in the routers. During the transition, micro-loops may occur over a single link between a pair of routers that temporarily use each other as the next hop for a prefix. Micro-loops may also form when each router in a cycle of three or more routers has the next router in the cycle as a next hop for a given prefix.

Cyclic loops may occur if one or more of the following conditions are met:

1. Asymmetric link costs.

2. An equal-cost path exists between a pair of routers, each of which makes a different decision regarding which path to use for forwarding to a particular destination. Note that even routers that do not implement equal-cost, multi-path (ECMP) forwarding must make a choice between the available equal-cost paths, and unless they make the same choice, the condition for cyclic loops will be fulfilled.

3. Topology changes affecting multiple links, including single node and line card failures.

Micro-loops have two undesirable side effects: congestion and repair starvation.

- A looping packet consumes bandwidth until it either escapes as a result of the re-synchronization of the FIBs or its time to live (TTL) expires. This transiently increases the traffic over a link by as much as 128 times, and may cause the link to become congested. This congestion reduces the bandwidth available to other traffic (which is not otherwise affected by the topology change). As a result, the "innocent" traffic using the link experiences increased latency and is liable to congestive packet loss.

- In cases where the link or node failure has been protected by a fast-reroute repair, an inconsistency in the FIBs may prevent some traffic from reaching the failure, and hence being repaired. The repair may thus become starved of traffic and thereby rendered ineffective.

Although micro-loops are usually considered in the context of a failure, similar problems of congestive packet loss and starvation may also occur if the topology change is the result of management action. For example, consider the case where a link is to be taken out of service by management action. The link can be retained in service throughout the transition, thus avoiding the need for any repair. However, if micro-loops form, they may cause congestion loss and may also prevent traffic from reaching the link.

Unless otherwise controlled, micro-loops may form in any part of the network that forwards (or in the case of a new link, will forward) packets over a path that includes the affected topology change. The time taken to propagate the topology change through the network, and the non-uniform time taken by each router to calculate the new shortest path tree (SPT) and update its FIB, contribute to the duration of the packet disruption caused by the micro-loops. In some cases, a packet may be subject to disruption from micro-loops that occur sequentially at links along the path, thus further extending the period of disruption beyond that required to resolve a single loop.

## 3. Applicability

Loop-free convergence techniques are applicable to any situation in which micro-loops may form, for example, the convergence of a network following:

1. Component failure
2. Component repair
3. Management withdrawal of a component
4. Management insertion or a component
5. Management change of link cost (either positive or negative)
6. External cost change, for example, change of external gateway as a result of a BGP change
7. A Shared Risk Link Group (SRLG) failure

In each case, a component may be a link, a set of links, or an entire router. Throughout this document, we use the term SRLG when describing the procedure to be followed when multiple failures have occurred, whether or not they are members of an explicit SRLG. In the case of multiple independent failures, the loop-prevention method described for SRLG may be used, provided it is known that all of these failures have been repaired.

Loop-free convergence techniques are applicable to both IP networks and MPLS-enabled networks that use LDP, including LDP networks that use the single-hop tunnel fast-reroute mechanism.

An assessment of whether loop-free convergence techniques are required should take into account whether or not the interior gateway protocol (IGP) convergence is sufficiently fast that any micro-loops are of such short duration that they are not disruptive, and whether or not the topology is such that micro-loops are likely to form.

## 4. Micro-Loop Control Strategies

Micro-loop control strategies fall into four basic classes:

1. Micro-loop mitigation
2. Micro-loop prevention
3. Micro-loop suppression
4. Network design to minimise micro-loops

A micro-loop-mitigation scheme works by re-converging the network in such a way that it reduces, but does not eliminate, the formation of micro-loops. Such schemes cannot guarantee the productive forwarding of packets during the transition.

A micro-loop-prevention mechanism controls the re-convergence of the network in such a way that no micro-loops form. Such a micro-loop-prevention mechanism allows the continued use of any fast repair method until the network has converged on its new topology and prevents the collateral damage that occurs to other traffic for the duration of each micro-loop.

A micro-loop-suppression mechanism attempts to eliminate the collateral damage caused by micro-loops to other traffic. This may be achieved by, for example, using a packet-monitoring method that detects that a packet is looping and drops it. Such schemes make no attempt to productively forward the packet throughout the network transition.

Highly meshed topologies are less susceptible to micro-loops, thus networks may be designed to minimise the occurrence of micro-loops by appropriate link placement and metric settings. However, this approach may conflict with other design requirements, such as cost and traffic planning, and may not accurately track the evolution of the network or temporary changes due to outages.

Note that all known micro-loop-prevention mechanisms and most micro-loop-mitigation mechanisms extend the duration of the re-convergence process. When the failed component is protected by a fast-reroute repair, this implies that the converging network requires the repair to remain in place for longer than would otherwise be the case. The extended convergence time means any traffic that is not repaired by an imperfect repair experiences a significantly longer outage than it would experience with conventional convergence.

When a component is returned to service, or when a network management action has taken place, this additional delay does not cause traffic disruption because there is no repair involved. However, the extended delay is undesirable because it increases the time that the network takes to be ready for another failure, and hence leaves it vulnerable to multiple failures.

## 5. Loop Mitigation

There are two approaches to loop mitigation.

- Fast convergence
- A purpose-designed, loop-mitigation mechanism

### 5.1. Fast Convergence

The duration of micro-loops is dependent on the speed of convergence. Improving the speed of convergence may therefore be seen as a loop-mitigation technique.

### 5.2. PLSN

The only known purpose-designed, loop-mitigation approach is the Path Locking with Safe-Neighbors (PLSN) method described in PLSN [ANALYSIS]. In this method, a micro-loop-free next-hop safety condition is defined as follows:

In a symmetric-cost network, it is safe for router X to change to the use of neighbor Y as its next hop for a specific destination if the path through Y to that destination satisfies both of the following criteria:

1. X considers Y as its loop-free neighbor based on the topology before the change, AND
2. X considers Y as its downstream neighbor based on the topology after the change.

In an asymmetric-cost network, a stricter safety condition is needed, and the criterion is that:

> X considers Y as its downstream neighbor based on the topology both before and after the change.

Based on these criteria, destinations are classified by each router into three classes:

- Type A destinations: Destinations unaffected by the change (type A1) and also destinations whose next hop after the change satisfies the safety criteria (type A2).

- Type B destinations: Destinations that cannot be sent via the new, primary next hop because the safety criteria are not satisfied, but that can be sent via another next hop that does satisfy the safety criteria.

- Type C destinations: All other destinations.

Following a topology change, type A destinations are immediately changed to go via the new topology. Type B destinations are immediately changed to go via the next hop that satisfies the safety criteria, even though this is not the shortest path. Type B destinations continue to go via this path until all routers have changed their type C destinations over to the new next hop. Routers must not change their type C destinations until all routers have changed their type A2 and B destinations to the new or intermediate (safe) next hop.

Simulations indicate that this approach produces a significant reduction in the number of links that are subject to micro-looping. However, unlike all of the micro-loop-prevention methods, it is only a partial solution. In particular, micro-loops may form on any link joining a pair of type C routers.

Because routers delay updating their type C destination FIB entries, they will continue to route towards the failure during the time when the routers are changing their type A and B destinations, and hence will continue to productively forward packets, provided that viable repair paths exist.

A backwards-compatibility issue arises with PLSN. If a router is not capable of micro-loop control, it will not correctly delay its FIB update. If all such routers had only type A destinations, this loop-mitigation mechanism would work as it was designed. Alternatively, if all such incapable routers had only type C destinations, the "loop-prevention" announcement mechanism used to trigger the tunnel-based schemes (see Sections 6.2 to 6.4) could be used to cause the type A and B destinations to be changed, with the incapable routers and routers having type C destinations delaying until they received the "real" announcement. Unfortunately, these two approaches are mutually incompatible.

Note that simulations indicate that in most topologies treating type B destinations as type C results in only a small degradation in loop prevention. Also note that simulation results indicate that in production networks where some, but not all, links have asymmetric costs, using the stricter asymmetric-cost criterion actually reduces the number of loop-free destinations because fewer destinations can be classified as type A or B.

This mechanism operates identically for:

- events that degrade the topology (e.g., link failure),
- events that improve the topology (e.g., link restoration), and
- shared risk link group (SRLG) failure.

## 6. Micro-Loop Prevention

Eight micro-loop-prevention methods have been proposed:

1. Incremental cost advertisement
2. Nearside tunneling
3. Farside tunneling
4. Distributed tunnels
5. Packet marking
6. New MPLS labels
7. Ordered FIB update
8. Synchronized FIB update

### 6.1. Incremental Cost Advertisement

When a link fails, the cost of the link is normally changed from its assigned metric to "infinity" in one step. However, it can be proved [OPT] that no micro-loops will form if the link cost is increased in suitable increments, and the network is allowed to stabilize before the next cost increment is advertised. Once the link cost has been increased to a value greater than that of the lowest alternative cost around the link, the link may be disabled without causing a micro-loop.

The criterion for a link cost change to be safe is that any link that is subjected to a cost change of x can only cause loops in a part of the network that has a cyclic cost less than or equal to x. Because there may exist links that have a cost of one in each direction, resulting in a cyclic cost of two, this can result in the link cost having to be raised in increments of one. However, the increment can be larger where the minimum cost permits. Recent work [OPT] has

shown that there are a number of optimizations that can be applied to the problem in order to determine the exact set of cost values required, and hence minimise the number of increments.

It will be appreciated that when a link is returned to service, its cost is reduced in small steps from "infinity" to its final cost, thereby providing similar micro-loop prevention during a "good-news" event. Note that the link cost may be decreased from "infinity" to any value greater than that of the lowest alternative cost around the link in one step without causing a micro-loop.

When the failure is an SRLG, the link cost increments must be coordinated across all failing members of the SRLG. This may be achieved by completing the transition of one link before starting the next or by interleaving the changes.

The incremental cost change approach has the advantage over all other currently known loop-prevention schemes in that it requires no change to the routing protocol. It will work in any network because it does not require any cooperation from the other routers in the network.

Where the micro-loop-prevention mechanism is being used to support a planned reconfiguration of the network, the extended total reconvergence time resulting from the multiple increments is of limited consequence, particularly where the number of increments have been optimized. This, together with the ability to implement this technique in isolation, makes this method a good candidate for use with such management-initiated changes.

Where the micro-loop-prevention mechanism is being used to support failure recovery, the number of increments required, and hence the time taken to fully converge, is significant even for small numbers of increments. This is because, for the duration of the transition, some parts of the network continue to use the old forwarding path, and hence use any repair mechanism for an extended period. In the case of a failure that cannot be fully repaired, some destinations may therefore become unreachable for an extended period. In addition, the network may be vulnerable to a second failure for the duration of the controlled re-convergence.

Where large metrics are used and no optimization (such as that described above) is performed, the incremental cost method can be extremely slow. However, in cases where the per-link metric is small, either because small values have been assigned by the network designers or because of restrictions implicit in the routing protocol (e.g., RIP restricts the metric, and BGP using the autonomous system

(AS) path length frequently uses an effective metric of one or a very small integer for each inter AS hop), the number of required increments can be acceptably small even without optimizations.

### 6.2. Nearside Tunneling

This mechanism works by creating an overlay network using tunnels whose path is not affected by the topology change and then carrying the traffic affected by the change in that new network. When all the traffic is in the new, tunnel-based network, the real network is allowed to converge on the new topology. Because all the traffic that would be affected by the change is carried in the overlay network, no micro-loops form.

When a failure is detected (or a link is withdrawn from service), the router adjacent to the failure issues a new "loop-prevention" routing message announcing the topology change. This message is propagated through the network by all routers but is only understood by routers capable of using one of the tunnel-based, micro-loop-prevention mechanisms.

Each of the micro-loop-preventing routers builds a tunnel to the closest router adjacent to the failure. They then determine which of their traffic would transit the failure and place that traffic in the tunnel. When all of these tunnels are in place (determined, for example, by waiting a suitable interval), the failure is announced as normal. Because these tunnels will be unaffected by the transition and because the routers protecting the link will continue the repair (or forward across the link being withdrawn), no traffic will be disrupted by the failure. When the network has converged, these tunnels are withdrawn, allowing traffic to be forwarded along its new, "natural" path. The order of tunnel insertion and withdrawal is not important, provided that the tunnels are all in place before the normal announcement is issued and that the repair remains in place until normal convergence has completed.

This method completes in bounded time and is generally much faster than the incremental cost method. Depending on the exact design, it completes in two or three flood-SPF-FIB update cycles.

At the time at which the failure is announced as normal, micro-loops may form within isolated islands of non-micro-loop-preventing routers. However, only traffic entering the network via such routers can micro-loop. All traffic entering the network via a micro-loop-preventing router will be tunneled correctly to the nearest repairing router -- including, if necessary, being tunneled via a non-micro-loop-preventing router -- and will not micro-loop.

Where there is no requirement to prevent the formation of micro-loops involving non-micro-loop-preventing routers, a single, "normal" announcement may be made and a local timer used to determine the time at which transition from tunneled forwarding to normal forwarding over the new topology may commence.

This technique has the disadvantage that it requires traffic to be tunneled during the transition. This is an issue in IP networks because not all router designs are capable of high-performance IP tunneling. It is also an issue in MPLS networks because the encapsulating router has to know the label set that the decapsulating router is distributing.

A further disadvantage of this method is that it requires cooperation from all the routers within the routing domain to fully protect the network against micro-loops.

When a new link is added, the mechanism is run in "reverse". When the loop-prevention announcement is heard, routers determine which traffic they will send over the new link and tunnel that traffic to the router on the near side of that link. This path will not be affected by the presence of the new link. When the "normal" announcement is heard, they then update their FIB to send the traffic normally, according to the new topology. Any traffic encountering a router that has not yet updated its FIB will be tunneled to the near side of the link, and will therefore not loop.

When a management change to the topology is required, again exactly the same mechanism protects against micro-looping of packets by the micro-loop-preventing routers.

When the failure is an SRLG, the required strategy is to classify traffic according the furthest failing member of the SRLG that it will traverse on its way to the destination, and to tunnel that traffic to the repairing router for that SRLG member. This will require multiple tunnel destinations -- in the limiting case, one per SRLG member.

### 6.3. Farside Tunnels

Farside tunneling loop prevention requires the loop-preventing routers to place all of the traffic that would traverse the failure in one or more tunnels terminating at the router (or, in the case of node failure, routers) at the far side of the failure. The properties of this method are a more uniform distribution of repair traffic than is achieved using the nearside tunnel method and, in the case of node failure, a reduction in the decapsulation load on any single router.

Unlike the nearside tunnel method (which uses normal routing to the repairing router), this method requires the use of a repair path to the farside router. This may be provided by the not-via [NOT-VIA] mechanism, in which case no further computation is needed.

The mode of operation is otherwise identical to the nearside tunneling loop-prevention method (Section 6.2).

### 6.4. Distributed Tunnels

In the distributed tunnels loop-prevention method, each router calculates its own repair and forwards traffic affected by the failure using that repair. Unlike the fast reroute (FRR) case, the actual failure is known at the time of the calculation. The objective of the loop-preventing routers is to get the packets that would have gone via the failure into Q-space [FRR-TUNN] using routers that are in P-space. Because packets are decapsulated on entry to Q-space, rather than being forced to go to the farside of the failure, more optimum routing may be achieved. This method is subject to the same reachability constraints described in [FRR-TUNN].

The mode of operation is otherwise identical to the nearside tunneling loop-prevention method (Section 6.2).

An alternative distributed tunnel mechanism is for all routers to tunnel to the not-via address [NOT-VIA] associated with the failure.

### 6.5. Packet Marking

If packets could be marked in some way, this information could be used to assign them to one of:

- the new topology,
- the old topology, or
- a transition topology.

They would then be correctly forwarded during the transition. This mechanism works identically for both "bad-news" and "good-news" events. It also works identically for SRLG failure. There are three problems with this solution:

- A packet-marking bit may not be available, for example, a network supporting both the differentiated services architecture [RFC2475] and explicit congestion notification [RFC3168] uses all eight bits of the IPv4 Type of Service field.

- The mechanism would introduce a non-standard forwarding procedure.
- Packet marking using either the old or the new topology would double the size of the FIB; however, some optimizations may be possible.

### 6.6. MPLS New Labels

In an MPLS network that is using [RFC5036] for label distribution, loop-free convergence can be achieved through the use of new labels when the path that a prefix will take through the network changes.

As described in Section 6.2, the repairing routers issue a loop-prevention announcement to start the loop-free convergence process. All loop-preventing routers calculate the new topology and determine whether their FIB needs to be changed. If there is no change in the FIB, they take no part in the following process.

The routers that need to make a change to their FIB consider each change and check the new next hop to determine whether it will use a path in the OLD topology that reaches the destination without traversing the failure (i.e., the next hop is in P-space with respect to the failure [FRR-TUNN]). If so, the FIB entry can be immediately updated. For all of the remaining FIB entries, the router issues a new label to each of its neighbors. This new label is used to lock the path during the transition in a similar manner to the previously described method for loop-free convergence with tunnels (Section 6.2). Routers receiving a new label install it in their FIB for MPLS label translation, but do not yet remove the old label and do not yet use this new label to forward IP packets, i.e., they prepare to forward using the new label on the new path but do not use it yet. Any packets received continue to be forwarded the old way, using the old labels, towards the repair.

At some time after the loop-prevention announcement, a normal routing announcement of the failure is issued. This announcement must not be issued until such time as all routers have carried out all of their activities that were triggered by the loop-prevention announcement. On receipt of the normal announcement, all routers that were delaying convergence move to their new path for both the new and the old labels. This involves changing the IP address entries to use the new labels AND changing the old labels to forward using the new labels.

Because the new label path was installed during the loop-prevention phase, packets reach their destinations as follows:

- If they do not go via any router using a new label, they go via the repairing router and the repair.

- If they meet any router that is using the new labels, they get marked with the new labels and reach their destination using the new path, back-tracking if necessary.

When all routers have changed to the new path, the network is converged. At some later time, when it can be assumed that all routers have moved to using the new path, the FIB can be cleaned up to remove the, now redundant, old labels.

As with other methods, the new labels may be modified to provide loop prevention for "good news". There are also a number of optimizations of this method.

### 6.7. Ordered FIB Update

The ordered FIB loop prevention method is described in "Loop-free convergence using oFIB" [oFIB]. Micro-loops occur following a failure or a cost increase, when a router closer to the failed component revises its routes to take account of the failure before a router that is further away. By analyzing the reverse shortest path tree (rSPT) over which traffic is directed to the failed component in the old topology, it is possible to determine a strict ordering that ensures that nodes closer to the root always process the failure after any nodes further away, and hence micro-loops are prevented.

When the failure has been announced, each router waits a multiple of the convergence timer [LF-TIMERS]. The multiple is determined by the node's position in the rSPT, and the delay value is chosen to guarantee that a node can complete its processing within this time. The convergence time may be reduced by employing a signaling mechanism to notify the parent when all the children have completed their processing, and hence when it is safe for the parent to instantiate its new routes.

The property of this approach is therefore that it imposes a delay that is bounded by the network diameter, although in many cases it will be much less.

When a link is returned to service, the convergence process above is reversed. A router first determines its distance (in hops) from the new link in the NEW topology. Before updating its FIB, it then waits a time equal to the value of that distance multiplied by the convergence timer.

It will be seen that network-management actions can similarly be undertaken by treating a cost increase in a manner similar to a failure and a cost decrease similar to a restoration.

The ordered FIB mechanism requires all nodes in the domain to operate according to these procedures, and the presence of non-cooperating nodes can give rise to loops for any traffic that traverses them (not just traffic that is originated through them). Without additional mechanisms, these loops could remain in place for a significant time.

It should be noted that this method requires per-router ordering but not per-prefix ordering. A router must wait its turn to update its FIB, but it should then update its entire FIB.

When an SRLG failure occurs, a router must classify traffic into the classes that pass over each member of the SRLG. Each router is then independently assigned a ranking with respect to each SRLG member for which they have a traffic class. These rankings may be different for each traffic class. The prefixes of each class are then changed in the FIB according to the ordering of their specific ranking. Again, as for the single failure case, signaling may be used to speed up the convergence process.

Note that the special SRLG case of a full or partial node failure can be dealt with without using per-prefix ordering by running a single reverse-SPF computation rooted at the failed node (or common point of the subset of failing links in the partial case).

There are two classes of signaling optimization that can be applied to the ordered FIB loop-prevention method:

- When the router makes NO change, it can signal immediately. This significantly reduces the time taken by the network to process long chains of routers that have no change to make to their FIB.

- When a router HAS changed, it can signal that it has completed. This is more problematic since this may be difficult to determine, particularly in a distributed architecture, and the optimization obtained is the difference between the actual time taken to make the FIB change and the worst-case timer value. This saving could be of the order of one second per hop.

There is another method of executing ordered FIB that is based on pure signaling [SIG]. Methods that use signaling as an optimization are safe because eventually they fall back on the established IGP mechanisms that ensure that networks converge under conditions of packet loss. However, a mechanism that relies on signaling in order to converge requires a reliable signaling mechanism that must be proven to recover from any failure circumstance.

### 6.8. Synchronised FIB Update

Micro-loops form because of the asynchronous nature of the FIB update process during a network transition. In many router architectures, it is the time taken to update the FIB itself that is the dominant term. One approach would be to have two FIBs and, in a synchronized action throughout the network, to switch from the old to the new. One way to achieve this synchronized change would be to signal or otherwise determine the wall clock time of the change and then execute the change at that time, using NTP [RFC1305] to synchronize the wall clocks in the routers.

This approach has a number of major issues. Firstly, two complete FIBs are needed, which may create a scaling issue; secondly, a suitable network-wide synchronization method is needed. However, neither of these are insurmountable problems.

Since the FIB change synchronization will not be perfect, there may be some interval during which micro-loops form. Whether this scheme is classified as a micro-loop-prevention mechanism or a micro-loop-mitigation mechanism within this taxonomy is therefore dependent on the degree of synchronization achieved.

This mechanism works identically for both "bad-news" and "good-news" events. It also works identically for SRLG failure. Further consideration needs to be given to interoperating with routers that do not support this mechanism. Without a suitable interoperating mechanism, loops may form for the duration of the synchronization delay.

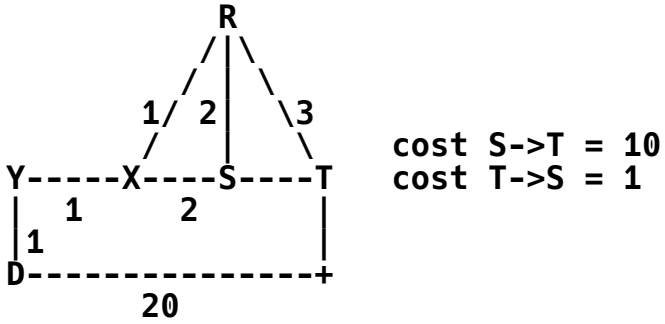
## 7. Using PLSN in Conjunction with Other Methods

All of the tunnel methods and packet marking can be combined with PLSN (see Section 5.2 of this document and [ANALYSIS]) to reduce the traffic that needs to be protected by the advanced method. Specifically, all traffic could use PLSN except traffic between a pair of routers, both of which consider the destination to be type C. The type-C-to-type-C traffic would be protected from micro-looping through the use of a loop-prevention method.

However, determining whether the new next-hop router considers a destination to be type C may be computationally intensive. An alternative approach would be to use a loop-prevention method for all local type C destinations. This would not require any additional computation, but would require the additional loop-prevention method to be used in cases that would not have generated loops (i.e., when the new next-hop router considered this to be a type A or B destination).

The amount of traffic that would use PLSN is highly dependent on the network topology and the specific change, but would be expected to be in the range of 70% to 90% in typical networks.

However, PLSN cannot be combined safely with ordered FIB. Consider the network fragment shown below:




On failure of link XY, according to PLSN, S will regard R as a safe neighbor for traffic to D. However, the ordered FIB rank of both R and T will be zero, and hence these can change their FIBs during the same time interval. If R changes before T, then a loop will form around R, T, and S. This can be prevented by using a stronger safety condition than PLSN currently specifies, at the cost of introducing more type C routers, and hence reducing the PLSN coverage.

## 8. Loop Suppression

A micro-loop-suppression mechanism recognizes that a packet is looping and drops it. One such approach would be for a router to recognize, by some means, that it had seen the same packet before. It is difficult to see how sufficiently reliable discrimination could be achieved without some form of per-router signature, such as route recording. A packet-recognizing approach therefore seems infeasible.

An alternative approach would be to recognize that a packet was looping by recognizing that it was being sent back to the place from which it had just come. This would work for the types of loop that form in symmetric-cost networks, but would not suppress the cyclic loops that form in asymmetric networks or as a result of multiple failures.

This mechanism operates identically for both "bad-news" events, "good-news" events, and SRLG failure.

## 9. Compatibility Issues

Deployment of any micro-loop-control mechanism is a major change to a network. Full consideration must be given to interoperation between routers that are capable of micro-loop control and those that are not. Additionally, there may be a desire to limit the complexity of micro-loop control by choosing a method based purely on its simplicity. Any such decision must take into account that if a more capable scheme is needed in the future, its deployment might be complicated by interaction with the scheme previously deployed.

## 10. Comparison of Loop-Free Convergence Methods

PLSN [ANALYSIS] is an efficient mechanism to prevent the formation of micro-loops but is only a partial solution. It is a useful adjunct to some of the complete solutions but may need modification.

Incremental cost advertisement in its simplest form is impractical as a general solution because it takes too long to complete. Optimized incremental cost advertisement, however, completes in much less time and requires no assistance from other routers in the network. It is therefore useful for network-reconfiguration operations.

Packet marking is probably impractical because of the need to find the marking bit and to change the forwarding behavior.

Of the remaining methods, distributed tunnels is significantly more complex than nearside or farside tunnels and should only be considered if there is a requirement to distribute the tunnel decapsulation load.

Synchronised FIBs is a fast method but has the issue that a suitable synchronization mechanism needs to be defined. One method would be to use NTP [RFC1305]; however, the coupling of routing convergence to a protocol that uses the network may be a problem. During the transition, there will be some micro-looping for a short interval because it is not possible to achieve complete synchronization of the FIB changeover.

The ordered FIB mechanism has the major advantage that it is a control-plane-only solution. However, SRLGs require a per-destination calculation and the convergence delay may be high, bounded by the network diameter. The use of signaling as an accelerator may reduce the number of destinations that experience the full delay, and hence reduce the total re-convergence time to an acceptable period.

The nearside and farside tunnel methods deal relatively easily with SRLGs and uncorrelated changes. The convergence delay would be small. However, these methods require the use of tunneled forwarding, which is not supported on all router hardware, and raises issues of forwarding performance. When used with PLSN, the amount of traffic that was tunneled would be significantly reduced, thus reducing the forwarding performance concerns. If the selected repair mechanism requires the use of tunnels, then a tunnel-based loop prevention scheme may be acceptable.

## 11. Security Considerations

This document analyzes the problem of micro-loops and summarizes a number of potential solutions that have been proposed. These solutions require only minor modifications to existing routing protocols and therefore do not add additional security risks. However, a full security analysis would need to be provided within the specification of a particular solution proposed for deployment.

## 12. Acknowledgments

The authors would like to acknowledge contributions to this document made by Clarence Filsfils.

## 13. Informative References

[ANALYSIS] Zinin, A., "Analysis and Minimization of Microloops in Link-state Routing Protocols", Work in Progress, October 2005.

[FRR-TUNN] Bryant, S., Filsfils, C., Previdi, S., and M. Shand, "IP Fast Reroute using tunnels", Work in Progress, November 2007.

[LF-TIMERS] Atlas, A., Bryant, S., and M. Shand, "Synchronisation of Loop Free Timer Values", Work in Progress, February 2008.

[NOT-VIA] Shand, M., Bryant, S., and S. Previdi, "IP Fast Reroute Using Not-via Addresses", Work in Progress, July 2009.

[OPT] Francois, P., Shand, M., and O. Bonaventure, "Disruption free topology reconfiguration in OSPF networks", IEEE INFOCOM May 2007, Anchorage.

[RFC1305] Mills, D., "Network Time Protocol (Version 3) Specification, Implementation", RFC 1305, March 1992.

[RFC2475] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., and W. Weiss, "An Architecture for Differentiated Services", RFC 2475, December 1998.

[RFC3168] Ramakrishnan, K., Floyd, S., and D. Black, "The Addition of Explicit Congestion Notification (ECN) to IP", RFC 3168, September 2001.

[RFC4090] Pan, P., Swallow, G., and A. Atlas, "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", RFC 4090, May 2005.

[RFC5036] Andersson, L., Minei, I., and B. Thomas, "LDP Specification", RFC 5036, October 2007.

[RFC5714] Shand, M. and S. Bryant, "IP Fast Reroute Framework", RFC 5714, January 2010.

[SIG] Francois, P. and O. Bonaventure, "Avoiding transient loops during IGP convergence", IEEE INFOCOM March 2005, Miami.

[oFIB] Francois, P., "Loop-free convergence using oFIB", Work in Progress, February 2008.

## Authors' Addresses

Mike Shand
Cisco Systems
250, Longwater Ave,
Green Park, Reading, RG2 6GB
United Kingdom

EMail: mshand@cisco.com

Stewart Bryant
Cisco Systems
250, Longwater Ave,
Green Park, Reading, RG2 6GB
United Kingdom

EMail: stbryant@cisco.com