Network Working Group
Internet Draft
Expiration Date: Dec 2007

S. Bryant
M. Shand
Cisco Systems

June 2007

A Framework for Loop-free Convergence
<draft-ietf-rtgwg-lf-conv-frmwk-01.txt>

Abstract
This draft describes 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.

Table of Contents

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 occurs in both IP networks and MPLS networks that use LDP [LDP] 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 are emerging which are more sensitive to the packet disruption that occurs during a transition. To make the transition transparent to their users, these services 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 TE fast-reroute mechanism for MPLS [MPLS-TE]. Alternative mechanisms that might be deployed in an MPLS network and mechanisms that may be used in an IP network are work in progress in the IETF [IPFRR]. Any 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.

There is, however, little point in introducing new mechanisms into an IP network to provide fast re-route, without also deploying mechanisms that prevent the disruptive effects of micro-loops which may starve the repair or cause congestion loss as a result of looping packets.

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 mechanisms that have been proposed to address the micro-loop issue.

1. The Nature of Micro-loops

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 routers has the next router in the cycle as a next hop for a prefix.

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

1. Asymmetric link costs.
2. The existence of an equal cost path between a pair of routers which make different decisions regarding which path to use for forwarding a particular destination. Note that even routers which 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 TTL expires. This transiently increases the traffic over a link by as much as 128 times, and may cause the link to congest. 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 re-route repair, the inconsistency in the FIBs prevents some traffic from reaching the failure and hence being repaired. The repair may thus become starved of traffic and hence become ineffective. Thus in addition to the congestive damage, the repair is rendered ineffective by the micro-loop. Similarly, if the topology change is the result of management action the link could have been retained in service throughout the transition (i.e. the link acts as its own repair path), however, if micro-loops form, they prevent productive forwarding during the transition.

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 may significantly extend the duration of the packet disruption caused by the micro-loops. In some cases a packet may be subject to disruption from micro-loops which occur sequentially at links along the path, thus further extending the period of disruption beyond that required to resolve a single loop.

2. Applicability

Loop free convergence techniques are applicable [APPL] 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 failure.

In each case, a component may be a link or a router.

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.

3. Micro-loop Control Strategies.

Micro-loop control strategies fall into three basic classes:

1. Micro-loop mitigation

2. Micro-loop prevention

3. Micro-loop suppression

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 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 done by micro-loops to other traffic. This may be achieved by, for example, using a packet monitoring method, which detects that a packet is looping and drops it. Such schemes make no attempt to productively forward the packet throughout the network transition.

Note that all known micro-loop mitigation and micro-loop prevention mechanisms extend the duration of the re-convergence process. When the failed component is protected by a fast re-route 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 which 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.

4. Loop mitigation

The only known loop mitigation approach is the Path Locking with safe-neighbors (PLSN) method described in [PLSN]. 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 which 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 Type 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 "covert" announcement mechanism used to trigger the tunnel based schemes (see sections 5.2 to 5.4) could be used to cause the Type A and Type 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 both "bad-news" events, "good-news" events and SRLG failure.

5. 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

5.1. 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 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 which 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 which 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 [PF] has shown that there are a number of optimizations which can be applied to the problem in order to minimize the number of increments required.

The incremental cost change approach has the advantage over all other currently known loop prevention scheme that it requires no change to the routing protocol. It will work in any network because

it does not require any co-operation from the other routers in the network.

Where large metrics are used and no optimization is performed, the 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 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.

The number of increments required, and hence the time taken to fully converge, is significant 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 become unreachable for an extended period.

Where the micro-loop prevention mechanism was being used to support a fast re-route repair the network may be vulnerable to a second failure for the duration of the controlled re-convergence.

Where the micro-loop prevention mechanism was being used to support a reconfiguration of the network the extended time is less of an issue. In this case, because the real forwarding path is available throughout the whole transition, there is no conflict between concurrent change actions throughout the network.

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 members of the SRLG. This may be achieved by completing the transition of one link before starting the next, or by interleaving the changes. This can be achieved without the need for any protocol extensions, by for example, using existing identifiers to establish the ordering and the arrival of LSP/LSAs to trigger the generation of the next increment.

5.2. Nearside Tunneling

This mechanism works by creating an overlay network using tunnels whose path is not affected by the topology change and 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 ("covert") 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, the failure is then 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.

This method completes in bounded time, and is 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 co-operation 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 "covert" 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 first member of the SRLG that it will traverse on its way to the destination, and to tunnel that traffic to the router that is closest to that SRLG member. This will require multiple tunnel destinations, in the limiting case, one per SRLG member.

5.3. 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 a 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 mechanism, in which case no further computation is needed.

The mode of operation is otherwise identical to the nearside tunneling loop prevention method (Section 5.2).

5.4. 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 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 G-space [TUNNEL] using routers that are in F-space. Because packets are decapsulated on entry to G-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 [TUNNEL].

The mode of operation is otherwise identical to the nearside tunneling loop prevention method (Section 5.2).

5.5. 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 could, for example, be achieved by allocating a Type of Service bit to the task [RFC791]. 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:

The packet marking bit may not available.

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.

5.6. MPLS New Labels

In an MPLS network that is using LDP [LDP] 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 5.2, the repairing routers issue a covert 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 which reaches the destination without traversing the failure (i.e. the next hop is in F-space with respect to the failure [TUNNEL]). 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 loop-free convergence with tunnels method (Section 5.2). 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 covert announcement, an overt announcement of the failure is issued. This announcement must not be issued until such time as all routers have carried out all of their covert announcement activities. On receipt of the overt 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 covert 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 time later, 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 method methods the new labels may be modified to provide loop prevention for "good news". There are also a number of optimizations of this method.

5.7. Ordered FIB Update

The Ordered FIB loop prevention method is described in [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 which is further away. By analyzing the reverse spanning tree over which traffic is directed

to the failed component in the old topology, it is possible to determine a strict ordering which 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 [TIMER]. The multiple is determined by the node's position in the reverse spanning tree, 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 was safe for the parent to instantiate its new routes.

The property of this approach is therefore that it imposes a delay which 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 co-operating nodes can give rise to loops for any traffic which traverses them (not just traffic which 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 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 which is based on pure signaling [OB]. Methods that use signaling as an optimization are safe because eventually they fall back on the established IGP mechanisms which 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 which must be proven to recover from any failure circumstance.

5.8. 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 [NTP] 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 and 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.

6. Using PLSN In Conjunction With Other Methods

All of the tunnel methods and packet marking can be combined with PLSN [PLSN] 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

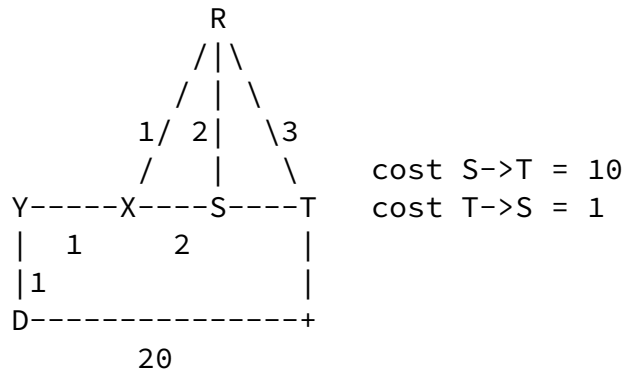
destination to be type C. The type C to type C traffic would be protected from micro-looping through the use of a loop prevention method.

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 which 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 region %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.

7. 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 that it had just come from. 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.

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

The problem with this class of micro-loop control strategies is that whilst they prevent collateral damage they do nothing to enhance the productive forwarding of packets during the network transition.

8. 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 will be complicated by interaction with the scheme previously deployed.

9. Comparison of Loop-free Convergence Methods

PLSN [PLSN] 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 is impractical as a general solution because it takes too long to complete. However, it is universally available, and hence may find use in certain 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 [NTP], 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 is high, bounded by the network diameter. The use of signaling as an accelerator will 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.

10. IANA considerations

There are no IANA considerations that arise from this draft.

11. Security Considerations

All micro-loop control mechanisms raise significant security issues which must be addressed in their detailed technical description.

12. Intellectual Property Statement

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such

rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

13. Disclaimer of Validity

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

14. Copyright Statement

15. Normative References

There are no normative references.

16. Informative References

Internet-drafts are works in progress available from <http://www.ietf.org/internet-drafts/>

[APPL] Bryant, S., Shand, M., "Applicability of Loop-free Convergence", <draft-bryant-shand-lf-applicability-03.txt>, June 2007, (work in progress).

[IPFRR] Bryant, S., Shand, M., "IP Fast-reroute Framework", <draft-ietf-rtgwg-ipfrr-framework-07.txt>, June 2007, (work in progress).

[LDP] Andersson, L., Doolan, P., Feldman, N., Fredette, A. and B. Thomas, "LDP Specification", RFC3036, January 2001.

[MPLS-TE] Ping Pan, et al, "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", RFC 4090, May 2005.

[NTP] RFC1305 Network Time Protocol (Version 3) Specification, Implementation and Analysis. D. Mills. March 1992.

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

[OFIB] Francois et. al., "Loop-free convergence using ordered FIB updates", <draft-ietf-rtgwg-ordered-fib-01.txt>, June 2007 (work in progress).

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

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

[RFC791] RFC791, "Internet Protocol Protocol" Specification, September 1981

[TIMER] S. Bryant, et. al. , "Synchronisation of Loop Free Timer Values", <draft-atlas-bryant-shand-lf-timers-02.txt>, October 2006 (work in progress)

[TUNNEL] Bryant, S., Shand, M., "IP Fast Reroute using tunnels", <draft-bryant-ipfrr-tunnels-02.txt>, Apr 2005 (work in progress).

17. 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