IP Fast Reroute Overview and Things we are struggling to solve

Alia K. Atlas

(mailto:aatlas@avici.com)

Gagan Choudhury

(mailto:gchoudhury@att.com)

David Ward

(mailto:dward@cisco.com)

Outline of Talk

- Motivations for IP Fast-Reroute
- Basic Framework & Loop-Free Alternates
- Advanced Mechanisms
  - U-turn Alternates
  - TE Tunnels
  - Tunnels with Directed Forwarding
  - Mechanism Comparisons
- Micro-Forwarding Loop Prevention
- Summary and Questions

What is the problem?

- Loss of connectivity
  - For which routes?
    - important IGP destinations and their recursive routes (IBGP/CBGP routes)
  - How Fast is required?:
    - Sub-Second: requirements for most IP network
    - sub-200ms: no app is sensitive to LoC <= 200ms
    - sub-50ms: *business requirement* for some fraction of IP networks

What is the problem.2 ?

- New Requirements emerging - Internet growing up
  - No longer “please don’t persistently oscillate my traffic” or dampen the effect of the control plane at all costs
  - VoIP
  - Video
  - Reduce impact of maintenance windows
  - Need determinism in IP networks

What’s in your network into the future?

- Sub-Second
  - conservatively met by current technology
  - deployment status:

  <u>*Discussed at previous NANOG and RIPE*</u>

- Sub-500ms
  - achievable goal, <u>*issue is determinism*</u>
- Sub-50ms
  - impossible

Why It Takes So Long

Detection of SONET layer failure

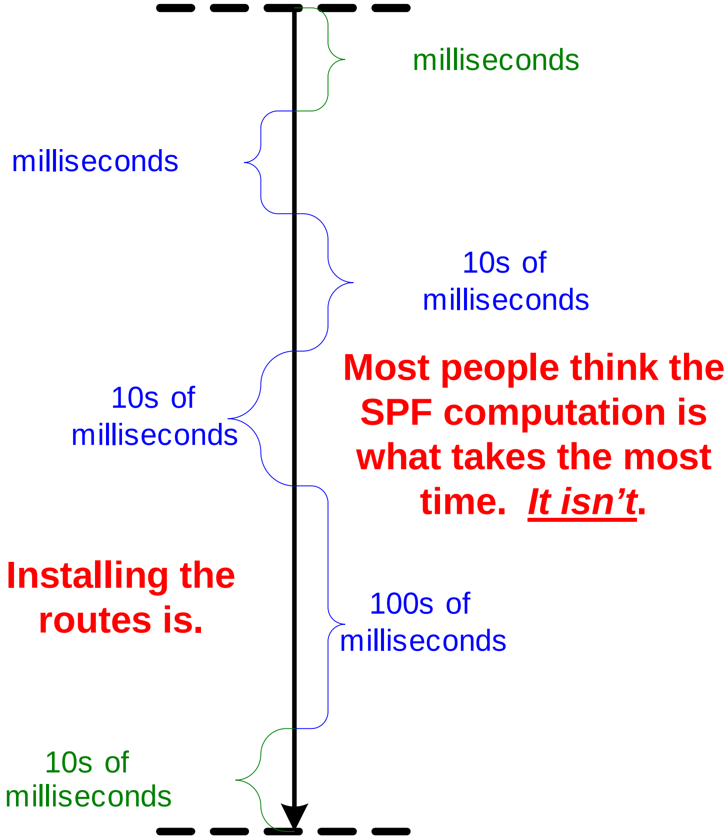
Report failure to Route Controller

Generate and flood an LSP

Trigger and Compute an SPF

Communicate new Next-Hops to linecards.

Install new Next-Hops into hardware path on each linecard.



milliseconds

milliseconds

10s of milliseconds

10s of milliseconds

Most people think the SPF computation is what takes the most time. *<u>It isn’t</u>*.

Installing the routes is.

100s of milliseconds

10s of milliseconds

What is in your network?
Fast IGP Convergence - a real example

- Convergence of the IGP <u>and its recursive routes</u>:

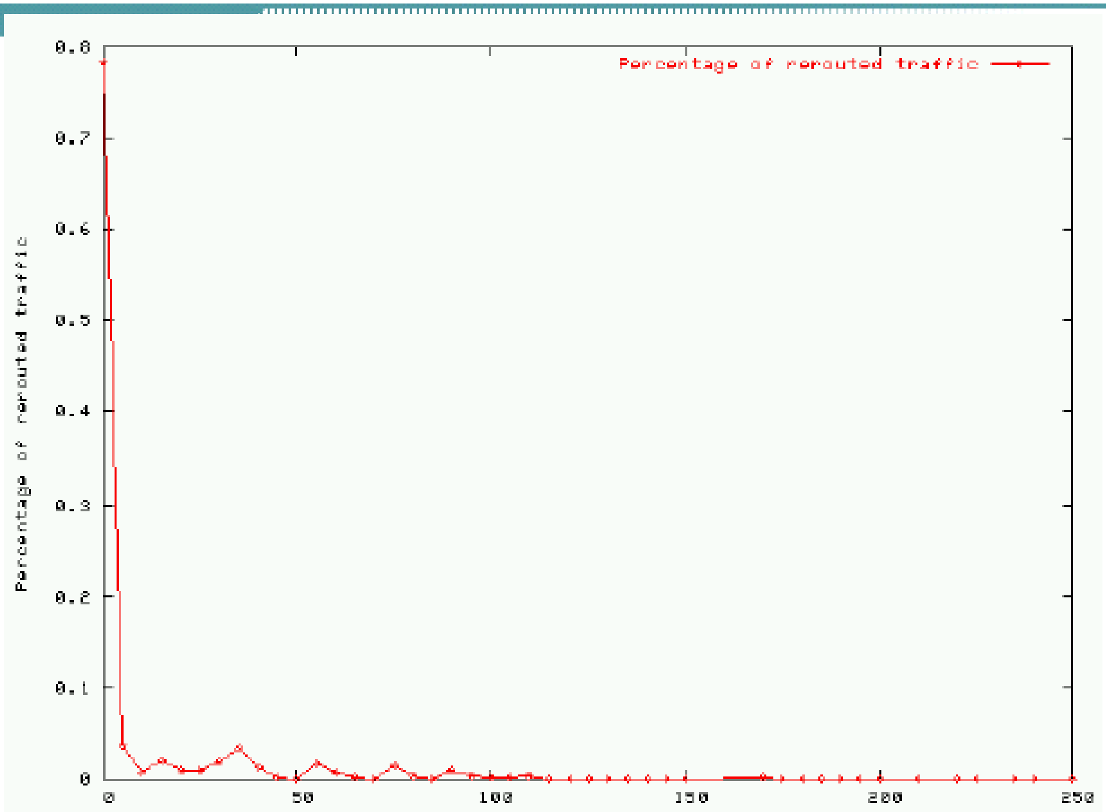
  - – Failure Detection (Sonet today, BFD emerging) < ~ 20ms
  - – Origination < ~ 10ms
  - – Queueing, Serialization, Propagation < 30ms
  - – Flooding < 5 * 2ms = 10ms
  - – SPF < n * 40us
  - – FIB update: p * 100us
  - – FIB Distribution Delay: 50ms
    - ~ 100ms + p * 0.1 ms
    - 500 important prefixes: ~ 150ms

- *<u>Worst-case</u>*: *<u>~ 280ms</u>*
  - –1500 nodes, 2500 prefixes

What’s not

- Propagation ~ 40r
- Detection (sonet) ·
- Infeasible to get tc 50ms

P: Propagation in ms (light speed)



Worldwide ISP with traffic matrix – summary for the failures of the 340 most loade links. Pessimistic definition of R

- MPLS FRR, SONET:
  - precomputation
  - local action (to avoid propagation/distribution)
  - tunneling (to avoid propagation/distribution)

Resilience for LDP Network

To support L3 and L2 VPNs, carriers deployed LDP to provide required MPLS path across network

Problem: Need Resiliency to failure for better SLAs

- Option 1: Use RSVP-TE Fast-Reroute
  - Deploy a new protocol (RSVP-TE).
  - Either create a new overlay network via a full-mesh of RSVP-TE tunnels
  - Or use 1-hop RSVP-TE tunnels with LDP to get only link protection
  - Or use 2-hop RSVP-TE tunnels with LDP to get link and node protection. This requires LDP sessions between a router and all its next-next-hops.
- Option 2: Use IP/LDP Fast-Reroute
- Option 3: Use combination in different parts of network.

Outline of Talk

- Motivations for IP Fast-Reroute
- Basic Framework & Loop-Free Alternates
- Advanced Mechanisms
  - U-turn Alternates
  - TE Tunnels
  - Tunnels with Directed Forwarding
  - Mechanism Comparisons
- Micro-Forwarding Loop Prevention
- Summary and Questions

IPFRR vs MPLS-FRR

- MPLS-FRR requires an MPLS infrastructure
- MPLS-FRR signals a source routed path around each protected failure.
- O(nk) repair paths must be set up in the network for link repair + O(nk^2) for node repair.

- IPFRR works on a pure IP network.
- No explicit routes.

RSVP-TE Fast-Reroute

PROs

- Provides Link, Node, and SRLG Protection
- Provides 100% Coverage except for Ingress & Egress Node Failures
- Understood and Deployed Technology
- Provides backup BW guarantees

Some CONs

- Overlay Network -> Scalability Concerns
- Operator Complexity?? – Many options & controls
- If No Need for TE, introduces new protocol just for resilience.
- Area & AS border routers are either single points of failure or make computation required unreasonable.
- Failure of Tunnel Ingress and Egress Cannot Be Protected Agains

Components of IPFRR solution

- Pre-computation of strategy
- Repair mechanism
- Reconvergence mechanism

IPFRR approach

- Node detects failure
- Node invokes pre-computed repair paths
- **Packet delivery restored (100%?)**
- Node generates and floods LSP describing failure
- All nodes recompute SPT and load new FIB using ***loop free*** convergence
- Remote (from failure) micro-forwarding loops can be separately solved via related micro-loop prevention techniques.
- Maximum disruption <50mS
  - Time to detect failure + a few mS

Precomputation of Strategy

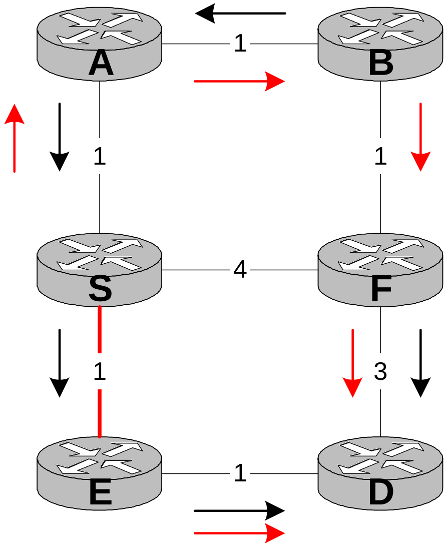
- There is insufficient time to compute the repair and install the repair paths when the failure is known.
- Strategy must be
  - Computed in advance
  - Consistent across the network
- Computation (for all methods) takes a significant time.

Micro-Loop Properties

- Independent decisions can cause micro-loops.
- Loops may occur between pairs of nodes or cycles of nodes.

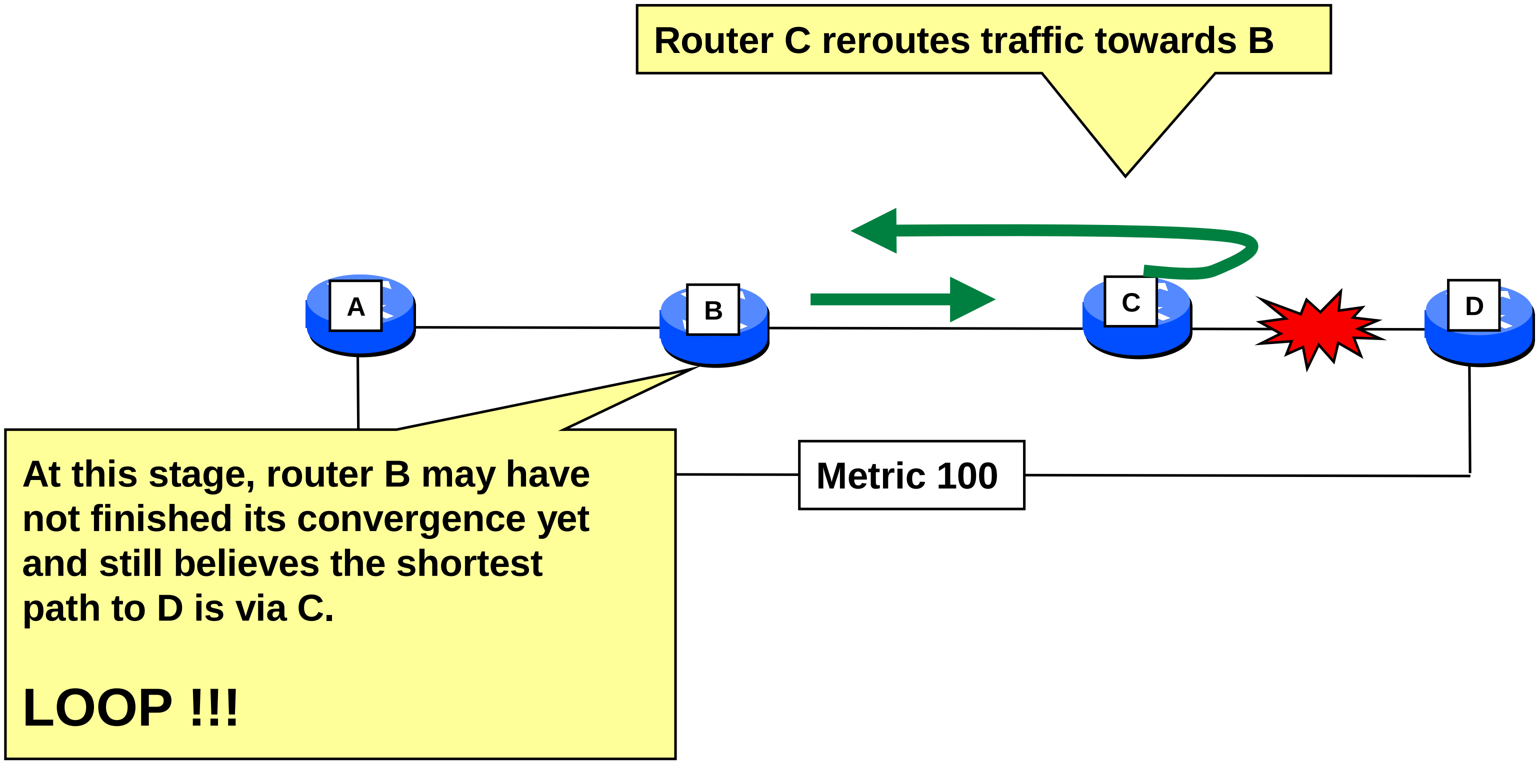
Duration depends on relative time to update FIBs.

- –Implementation differences
- –Number of affected destinations
- –Propagation time



Loss due to Loop duration may be longer (an order of magnitude) than Loss during the Fast Reroute failover.

uLoop - ***<u>Link Down</u>***



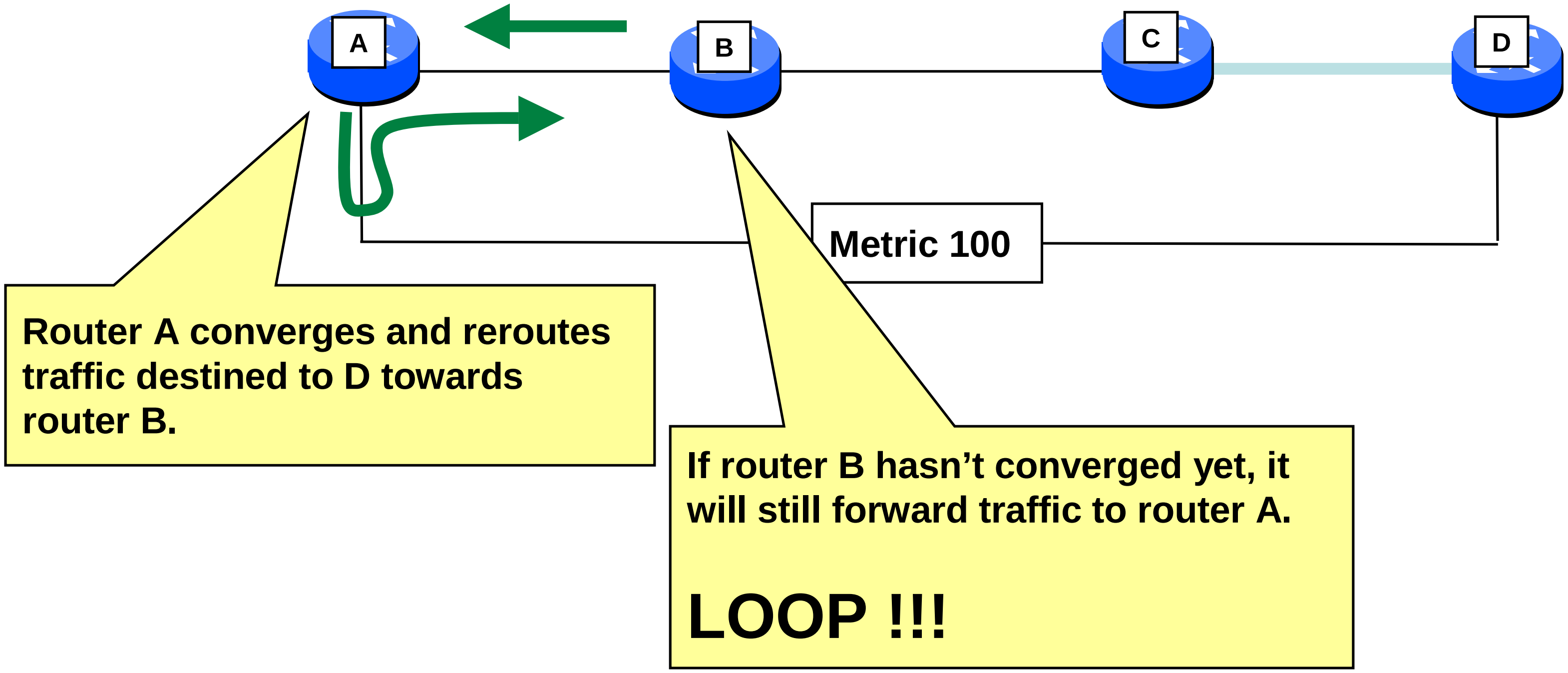
Router C reroutes traffic towards B

At this stage, router B may have not finished its convergence yet and still believes the shortest path to D is via C.

LOOP !!!

- After the protection takes effect, IGP convergence is desired and will occur… BUT
  - – this generally creates uloops
  - – all the routers do not converge at the same time

uLoop on ***Link Up***



Router A converges and reroutes traffic destined to D towards router B.

If router B hasn't converged yet, it will still forward traffic to router A.

LOOP !!!

- When C-D link comes up, both routers C and D will issue their new link-state packet with the new adjacency in it
- Routers close to the link change will likely converge prior to other nodes
- Routing loops seen in MPLS and IP networks

What you see today:

uloops do matter if x0ms is really the target

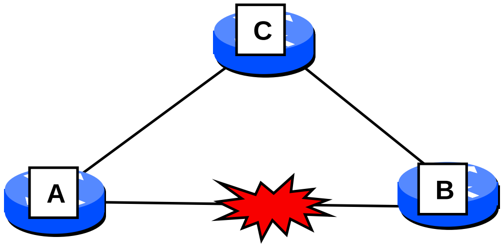
- *Dependent on the delta convergence time* between routers
  - The faster the IGP convergence, the smaller the deltas
  - The worst case is generally a few 100's of ms
- Solutions exist
  - Intuitively
    - Link Down: the closer to the failure, the later the FIB update should occur
    - Link Up: the closer to the repair, the sooner the FIB update should occur
- Can be leveraged to support interface graceful shutdown and no-shutdown

Repair Path Mechanisms

- ECMP
- Loop Free Alternates
  - A neighbor has a path to the destination which does not loop back to us
- Multi-hop repair paths
  - A node which is not a direct neighbor has a non-looping path to the destination

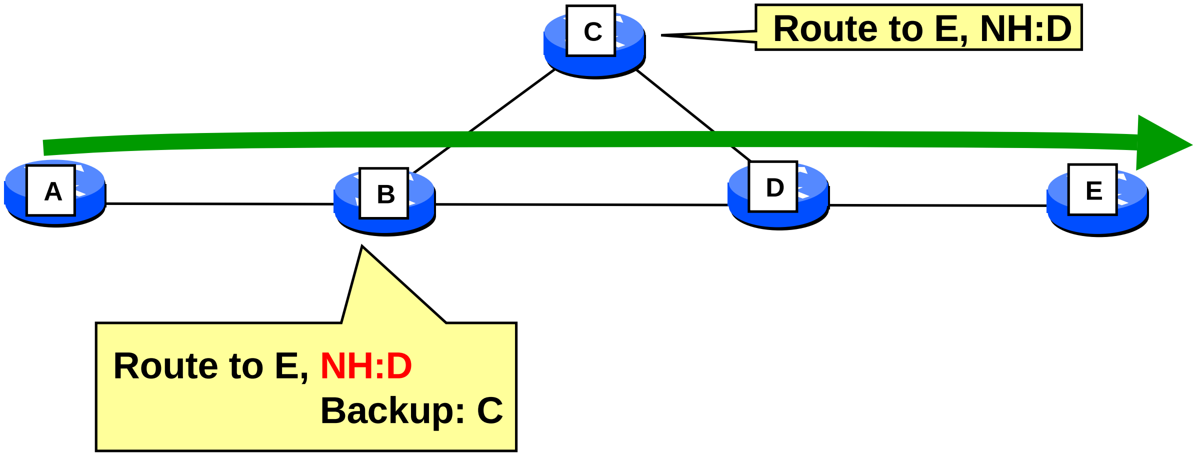
In all cases the path must work before, during and after convergence.

Foundation of Solutions: “Downstream Mode”



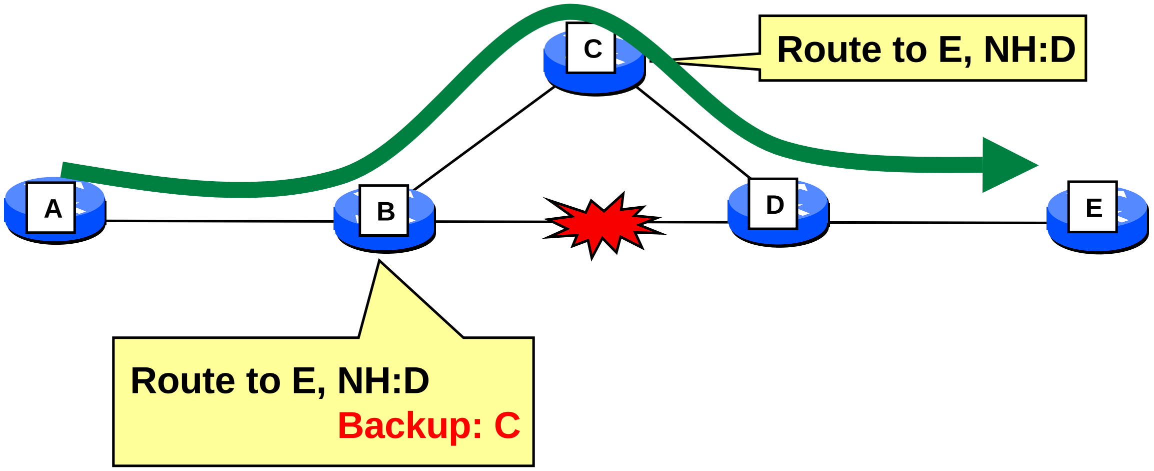
- When A-B fails, A, for sure, can locally reroute to C all its traffic normally sent onto link AB
- Obvious solution but still very applicable in practice
- The key is topologic shape and meshiness of network
- We have known about this algorithm for ~30 years
  - reduce complexity, add value, no extensions to protocols required

Downstream Routes.1



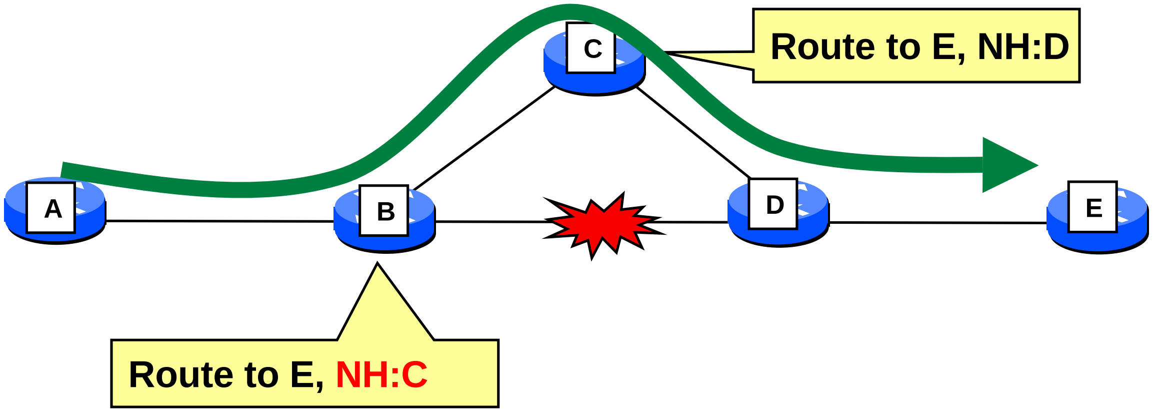
- Used when another neighbor can be safely used as an alternate next-hop for ALL protected traffic
- Upon B-D failure, router B can safely reroute to C all traffic it used to send to D
  - No loop, router C will forward to E and not back to B
- Precomputation without any new topology information:
  B just leverages its LSPDB/LSADB

Downstream Routes.2



- When link failure is detected, traffic is forwarded according to backup entry
- Local decision in the rerouting node
  - No need to signal anything
- Traffic is rerouted and meanwhile the IGP converges

Downstream Routes.3



Route to E, NH:D

Route to E, NH:C

- When IGP converges, nhop/oif of primary path is updated.
- Precomputation of backup’s is refreshed according to new topology
- Downstream routes do not work in all cases
  - Requires meshed topologies
  - Not always the case within core networks

Downstream Routes - Conclusion

- Downstream routes are ***easy to compute***
- RIB and FIB entries are populated with backup information
- Failure detection and traffic rerouting mechanism exactly the same as for MPLS-FRR
- Downstream routes require meshed topologies
  - The Triangle shape…
- ***Not always realistic*** in real backbones
- According to surveys, ***70 to 85 percent*** of the topology cases

Multi-hop repair mechanisms

- Aim is to fix the ***remaining 20%***
- Two classes of approaches
  - “Repair FIBs”
  - “Loose source routing” (or equivalent)

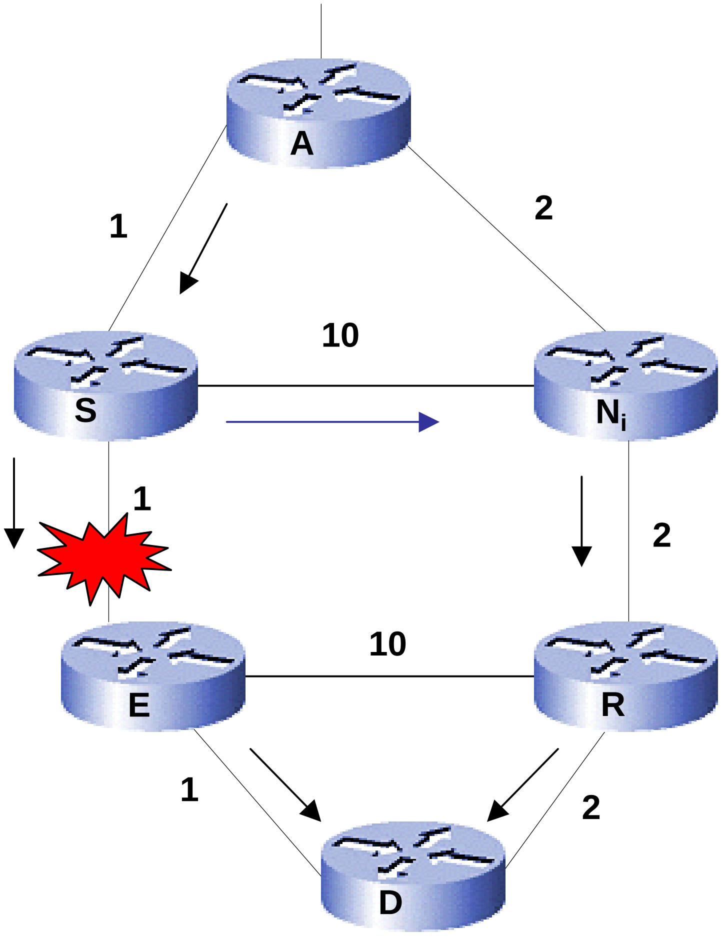
Summary: IP Fast-Reroute Framework

- Pre-Compute Alternates On Topology Change
- Use Alternate during a local failure
- Continue using Alternate until Network “converged” and new primary next-hops available.
- Alternate Next-Hop can be used for LDP as well as IGP/BGP to provide sub-second traffic re-direction.
- Remote (from failure) micro-forwarding loops can be separately solved via related micro-loop prevention techniques.

Outline of Talk

- Motivations for IP Fast-Reroute
- Basic Framework & Loop-Free Alternates
- Advanced Mechanisms
  - U-turn Alternates
  - TE Tunnels
  - Tunnels with Directed Forwarding
  - Mechanism Comparisons
- Micro-Forwarding Loop Prevention
- Summary and Questions

Loop-Free Alternates



The path from the neighbor N_i to the destination D must not go through the alternate-computing router S.

For Link-Protection: Avoid the pseudo-node (if any) on primary next-hop from S to E.

For Node Protection: Avoid the primary neighbor E.

For SRLG Protection: Avoid SRLGs on primary next-hop from S to E. Requires tracking SRLGs on shortest path from N_i to D.

Local Hold-down May Be Desirable

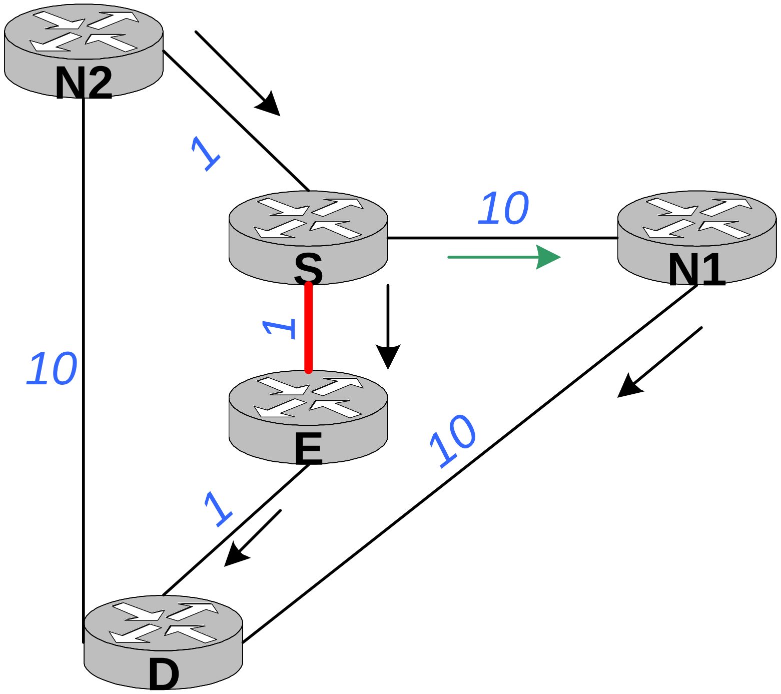
Local failure occurs on link from S to E.

S redirects traffic to alternate N1.

S reports into IGP about the local failure.

Network converges.

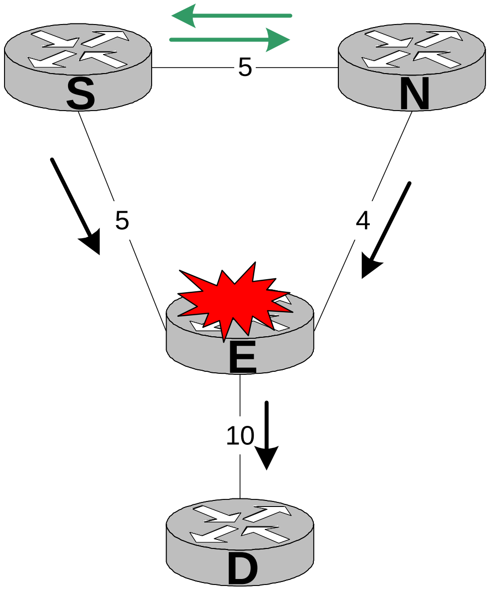
S directs traffic to new primary next-hop N2.



The hold-down at S ensures that traffic from S continues to reach the destination. Without this hold-down, S’s new primary next-hops may cause traffic to loop back to S if S’s new primary next-hops weren’t loop-free before the failure.

Use of Non-Downstream Paths

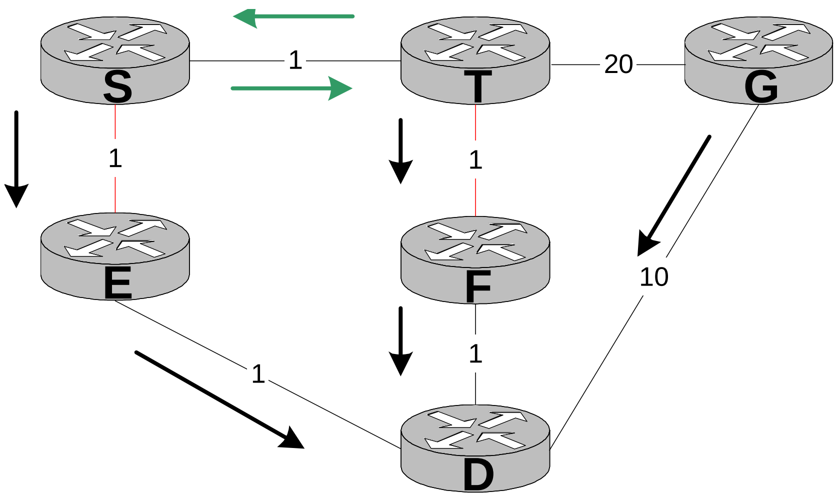
- Insufficiently Protected Failures can cause forwarding loops via the alternates.
  - Loop-free Link protecting alternates can loop when a node failure occurs



Two Solutions:

- Require that link-protecting alternates be downstream paths ($D_{opt}(N, D) < D_{opt}(S, D)$).
  - This avoids the potential for these loops – but may substantially reduce coverage.
- Use Other Node Protection Method – such as Non-Stop Routing™ or Graceful Restart.

Unprotected SRLG Failure



- If an SRLG fails and alternate doesn’t protect against that failure, then micro-looping can occur.
- This can happen whenever failure is more extensive than alternate protects against.
- This can be solved by using only downstream paths – but causes reduced coverage.

Interactions with Costed-Out Links

- Links can be given maximum cost because
  - BGP is not yet synchronized (RFC 3137)
  - LDP is not fully synchronized (draft-jork-ldp-igp-sync-00.txt)
  - Maintenance is being done on the link
  - Etc.
- IP/LDP Fast-Reroute must not use a costed-out link as an alternate next-hop.
  - This maintains the intention of making the link maximum cost.
  - Without rule, costed-out link would be likely to be used because it means that N’s path is very likely to be loop-free.

MIB Information

- Extension to IP Routing Table MIB
  (draft-atlas-rtgwg-ipfrr-ip-mib-00.txt)
  - Reports protection available per route per primary next-hop
  - Reports routes without protection and why
  - Provides summary counts of protected and unprotected routes
- Need Extension to LSR MIB
  - Report protection available per out-segment per in-segment.
  - Report in-segments without protection and why.
  - Provide summary counts of protected and unprotected routes.
- Need Extensions to IGP MIBs
  - Type of IP/LDP Fast-Reroute configured on MIB
  - Performance counters – such as times that single failure assumption was valid or violated.
  - Etc.

Basic Alternate Troubleshooting

- Show commands and MIB to report alternate next-hop(s) used for each route.
- Alternate next-hop is only used for brief period on a failure – but still want to verify functionality.
- Support ping and trace-route via alternate next-hop(s) to force packets onto alternate next-hop.

Network Coverage: Loop-Free Alternates

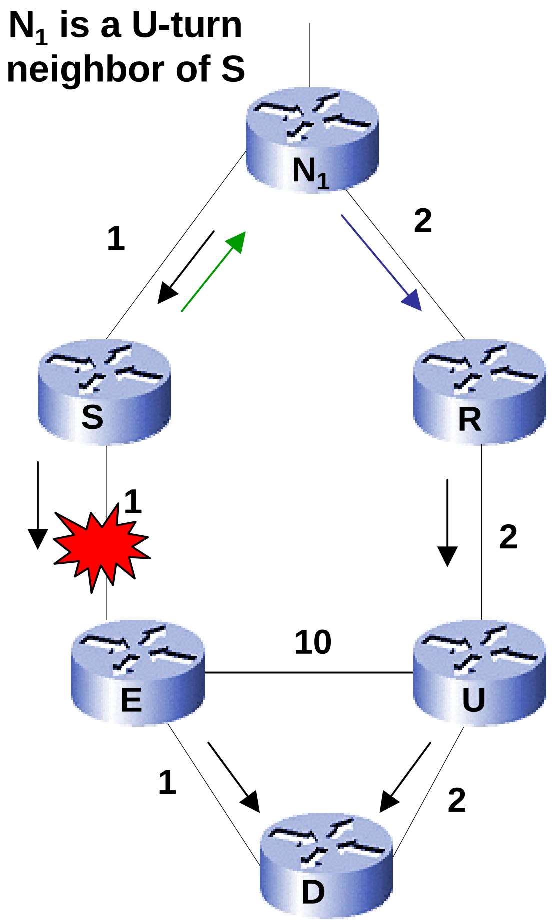
- Existence of Loop-Free Alternates depends strongly on network topology.
- Minor changes to network can lead to further improved coverage.
- Analysis below based on source/destination pairs, not % of traffic covered or % of link or node failures fully covered.

Network	Alternate Type	% Node Protected	% Link Protected	%Link Protected where Node is Unavoidable	% Loop-Free	% Protected	% Unprotected
Average	Loop-Free only	55.50%	19.07%	4.97%	79.54%	79.54%	20.46%

IETF Standardization

- Drafts in Routing Area Working Group
- IP Fast-Reroute framework (draft-ietf-ipfrr-framework-02)
- Loop-Free Alternates (draft-ietf-ipfrr-base-spec-01)
- Advanced Mechanisms needed to improve coverage.
  - Draft-atlas-ip-local-protect-uturn-01
  - Draft-bryant-ipfrr-tunnels-01
  - TE tunnels as Alternates (draft-shen-nhop-fastreroute-00)

U-Turn Alternates



- N_1 can provide a U-turn alternate to S because:
  - N_1 itself has a loop-free node-protecting alternate path to reach D
  - N_1 can break the loop
  - N_1 is a U-Turn neighbor of S
- So S could use N_1 as an alternate, if N_1 were capable of breaking the loop when a failure happens.
- U-turn traffic can be explicitly marked or implicitly detected.

If N_1 receives U-turn traffic from its primary neighbor S, instead of forwarding that traffic back to S, N_1 forwards the traffic to its alternate R.

U-Turn Alternates

- Mechanism allows S to direct traffic to join the shortest-path tree at S’s neighbor’s neighbor.
- Can substantially increase coverage in real topologies.

Network	% Node Protected	% Link Protected	%Link Protected where Node is Unavoidable	% Loop-Free	% U-Turn	% Protected	% Unprotected
Average	72.00%	19.17%	7.28%	75.41%	23.04%	98.45%	1.55%

- Requires signaling of U-turn recipient capability.
- Allows protection of LDP traffic without additional LDP sessions or extensions.
- Same additional computational complexity as for loop-free alternates.
- U-turn alternates can be cascaded for still better coverage.

Basic Configuration Options

Per IGP Area, Enable IP/LDP Fast-Reroute

- Supports Loop-Free Alternates and U-turn Alternates
- Can disable use of U-turn Alternates for stand-alone solution
  - Will also disable signaling extensions
- Specify Local Hold-down Timer
  - Continues use of alternate until new primary is safe
- Configure use of non-downstream paths for better protection.

Per Interface

- Enable/Disable use of interface as an alternate next-hop
- Enable/Disable interface’s U-turn recipient ability

“Loose source routing”

- True LSR would be nice, but no hardware supports it (IPv6?)
- Can approximate in pure IP using tunnels
- One intermediate destination (way point) is all we need for most repairs
- Sometimes need to get packet to neighbor of way point.
  - “Directed forwarding”

Draft-bryant-ipfrr-tunnels-01.txt

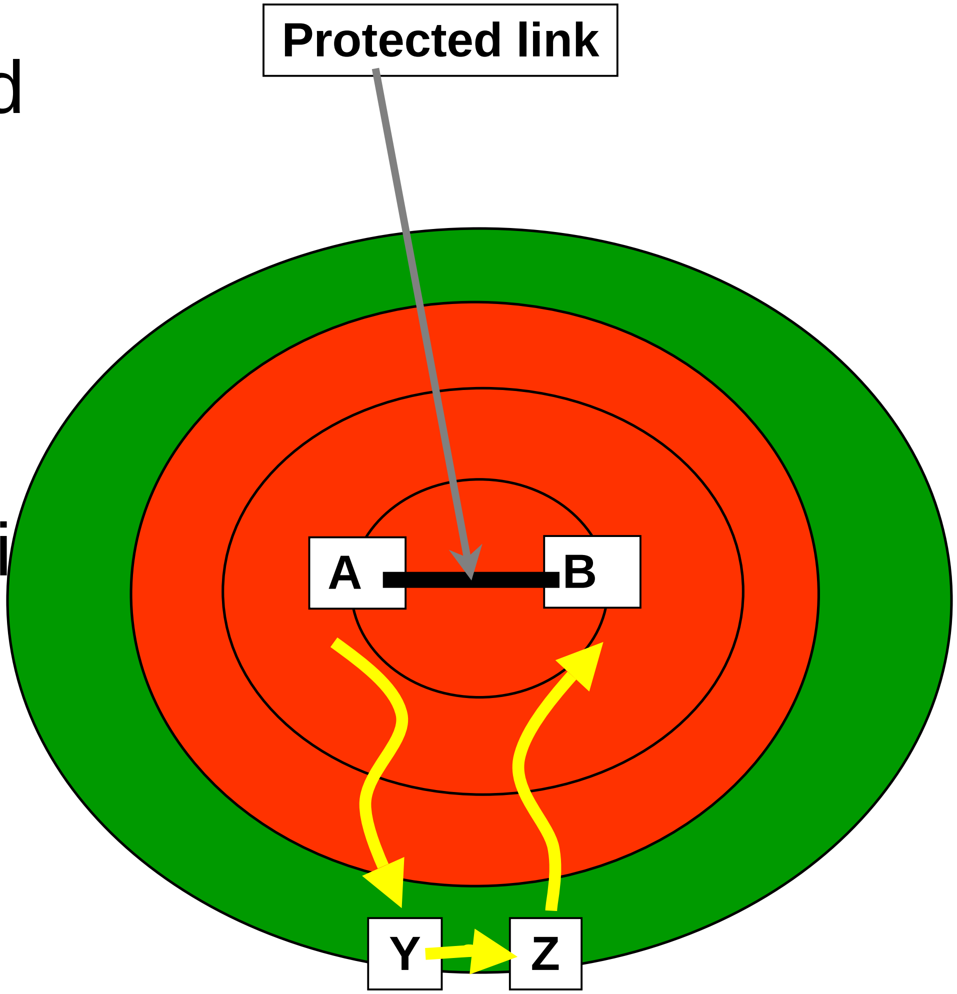
- Goal is ***100% coverage*** without asymmetric link costs.
- Uses ***IP tunnels*** combined with *directed forwarding*.
  - Any form of IP tunneling can be used: IPnIP, GRE, L2TP, etc
- *Directed forwarding* lets a router specify to the tunnel egress where to forward traffic. This could be done with MPLS labels
- Link protection has 100% coverage.
- Tunnels provide most of the additional coverage.
- Directed Forwarding adds the last couple percent of coverage (for link protection).
- To get 100% Node protection can require secondary repairs; this is a very small percentage of total.
- SRLGs to be addressed in next version of draft.

Precomputed Tunnel Solution

1. Y is reachable from A without using the protected link

2. Traffic A would have sent to B, would be forwarded from Z to the destination vi existing FIB

3. Tunnel can be any IP encapsulation L2TPv3, GRE, IPnIP or Tunnel can be MPLS Label



TE Tunnels as Alternates

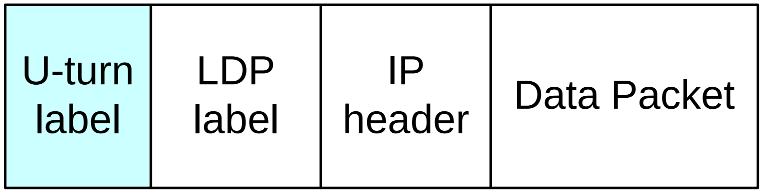
- Can use TE tunnels to enhance the topology and provide a direct adjacency to a router that could give loop-free alternates.
- Draft-shen-nhop-fastreroute-00.txt
  - Suggests using MPLS-TE tunnels to reach beyond immediate neighbors and gives details for computation.
  - Creation and Management of TE tunnels isn’t addressed.
  - Requires targeted LDP sessions or extensions to protect LDP traffic.
- Given TE support, provides a stand-alone solution for a router to improve coverage beyond loop-free alternates.
- If the network has an alternate path, an explicitly routed TE tunnel can use to always provide an alternate.

IP/LDP Fast-Reroute: Forwarding Implications

- Use a single longest-match tree of prefixes to determine forwarding result.
- Each forwarding result stores up to N primary next-hops plus 1 alternate next-hop.
- Pick one from the N primary next-hops
  - If selected next-hop is down, select among the N+1 for where to send it.
- For Scaling and Fast Repair:
  - Use indirection so prefixes can use same forwarding result.
  - Store and check interface state in forwarding path

U-Turn Alternates: Forwarding Implications

- Packets sent to alternate next-hop may require adding a single U-turn label – that is ALWAYS removed at the next router. U-turn label identifies packet as potential U-turn packet. Forwarding is done based on MPLS label or IP header underneath.



U-turn label

LDP label

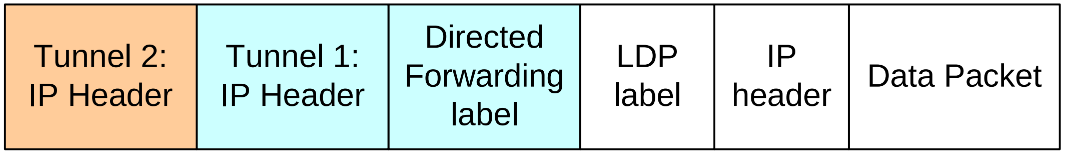
IP header

Data Packet

- If a potential U-turn packet would be sent to same neighbor as the packet was received from, select among the N+1 next-hops for where to send it.

IP Tunnels with Directed Forwarding: Forwarding Implications

- Multiple encapsulations may (infrequently) be required.
  - Add a directed forwarding label
  - Add an IP header to reach final waypoint
  - Add a second IP header to get to middle waypoint.
  - As SRLG and node protection is desired, more waypoints (with associated tunnels) may be required…
  - Multi-homed prefixes may require an extra tunnel.



Tunnel 2: IP Header

Tunnel 1: IP Header

Directed Forwarding label

LDP label

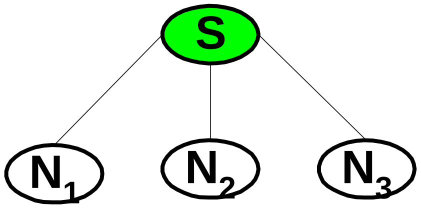
IP header

Data Packet

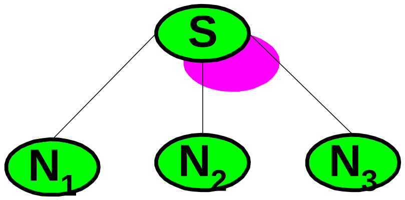
- Requires ability to remove IP headers and perform two lookups.
- Tunnel set-up time (may be router-internal) adds period of vulnerability to failures.

Source-Based Computation - Scales

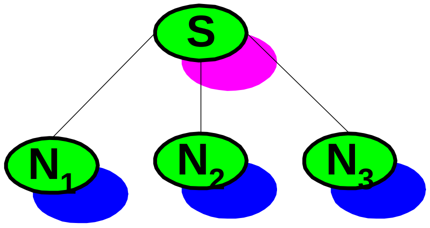
- IP Routing to compute primary paths - 1 SPF rooted at S



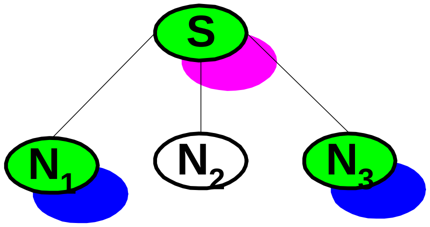
- Loop-Free Alternates – 1 SPF rooted at each neighbor of S plus 1 reverse SPF rooted at S



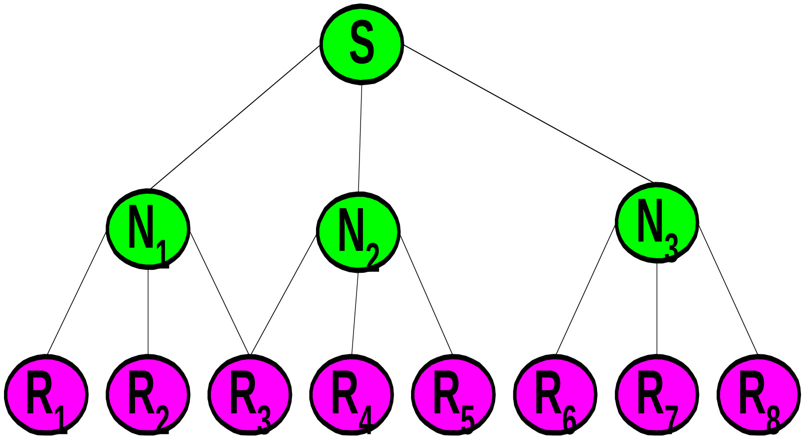
- U-Turn Alternates – 1 SPF rooted at each neighbor of S



- Number of SPFs required can be reduced by not allowing a neighbor to be used as an alternate.



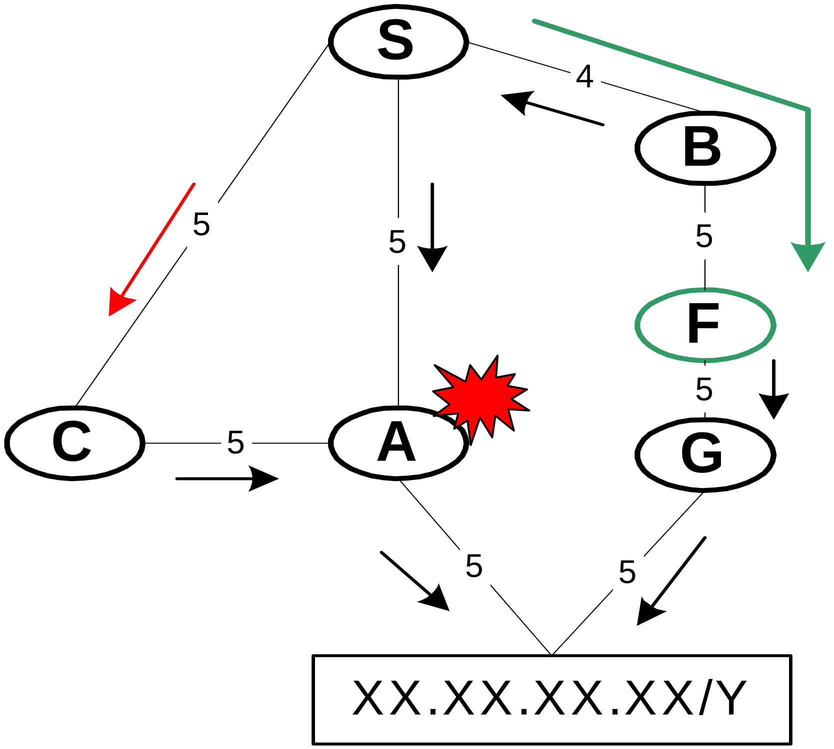
Destination-Based Computation



- One SPF per destination doesn't scale – Use Proxies to reduce number of SPFs.
- For node-protection, each next-next-hop is a proxy.
- IP Tunnels (plus Directed Forwarding) – 1 reverse SPF per proxy for simple single tunnel case
- RSVP-TE Tunnels – 1 CSPF per proxy

No configuration can reduce number of SPF computations required.

Proxy Problems: Multi-Homed Prefixes



- Using A as a proxy for the prefix gives an alternate that doesn’t protect against node failure.
- No acceptable proxy – must do an SPF per group of multi-homed prefixes
- Every link with a subnet may give a different multi-homed prefix.

Protecting Area/Level Border Routers requires even more computation.

Comparison Summary

Method	# of SPFs	Encapsulation Overhead	Average Coverage	Complexity
Loop-Free Alternates	2 + #neighbors	None	~80% Topology dependent	Low
U-turn Alternates + Loop-Free	2 + 2*#neighbors	None or 1 U-turn label	~98% Topology dependent	Medium
IP Tunnels + directed forwarding + Loop-Free	1 + #neighbors' neighbors + #multi-homed prefix groups	1 to 3+ IP headers plus 1 directed forwarding label	100% (with enough levels of tunnels)	High

Outline of Talk

- Motivations for IP Fast-Reroute
- Basic Framework & Loop-Free Alternates
- Advanced Mechanisms
  - U-turn Alternates
  - TE Tunnels
  - Tunnels with Directed Forwarding
  - Mechanism Comparisons
- Micro-Forwarding Loop Prevention
- Coverage & Capacity Needs for Example Core Networks

Microlooping is undesirable

- We have shown that there are a number of mechanisms that can prevent packet loss by fast re-route.
- BUT packets can still be lost due to micro-looping.
- Micro-loops form when the FIBs are inconsistent.
- Controlled convergence allows us to reduce of eliminate micro-looping.

Controlled convergence

- Made feasible for failure case by fast reroute
  - Traffic is not lost so can afford to take time
  - Can use common method for both failure and management change events
- Traditional convergence optimized for failure case without fast-reroute.

We can do better…

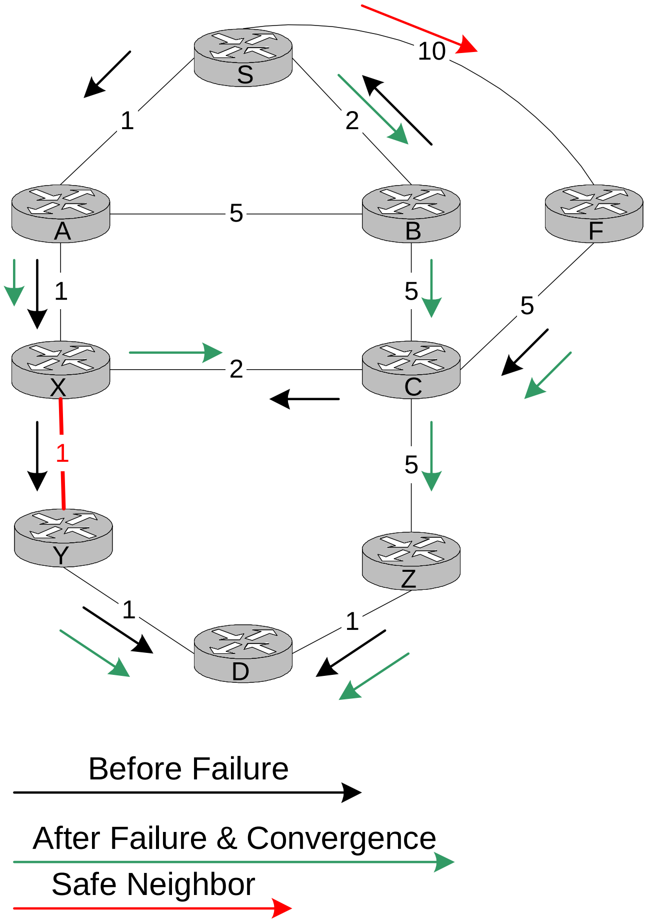
(but keep traditional as safe fall-back for single failure assumption violation.)

Path Locking Via Safe Neighbors

- Find a safe neighbor to provide transitional path
  - Loop-Free Neighbor before topology change and
  - Downstream Neighbor after topology change
- Three step approach:
  - Interval 1: Install safe neighbor(s) as primary next-hops.
  - Interval 2: If no safe neighbor, install new primary next-hops.
  - Interval 3: Install new primary next-hops
- Fixed Convergence Time regardless of Network Size

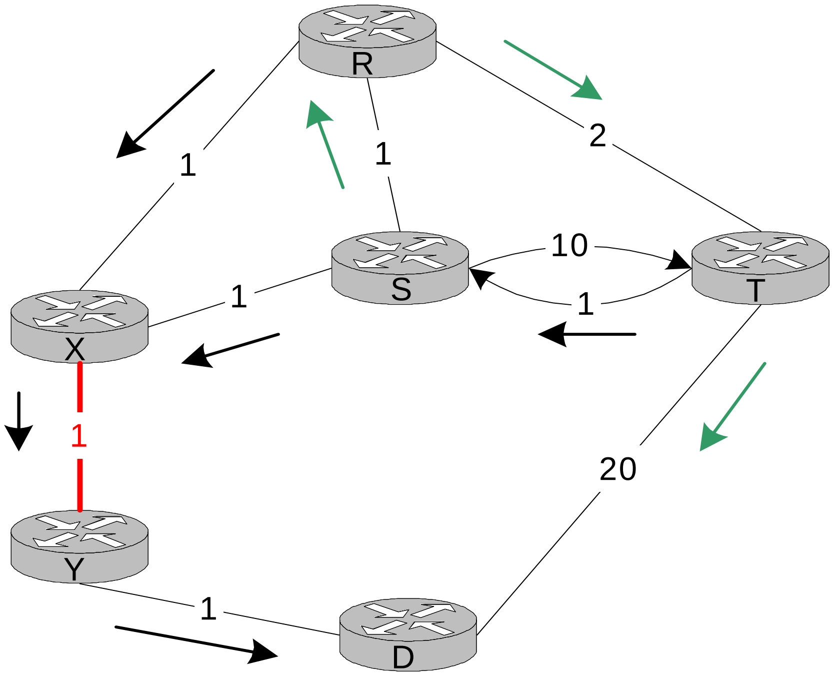
See draft-zinin-microloop-analysis-00

Example Prevention



- X->Y fails
- Interval 1:
  - S sends traffic to F
  - B sends traffic to C
  - C sends traffic to Z
- Interval 2:
  - X sends traffic to C
- Interval 3:
  - S sends traffic to B
- Avoids micro-loops between S & B and between X & C

Loops with Asymmetric Metrics



- R has a “safe neighbor” T – but micro-loop can form between T, S and R.
- Could avoid if a “safe neighbor” had to be a downstream path before the failure – but this reduces coverage.

Path Locking Coverage

- Micro-Forwarding Loops still possible between nodes without safe neighbors
- Analysis of real topologies shows pretty good (>90%) coverage.
- Similar set of unprotected destinations as for loop-free alternates.
- Collateral damage possible for protected traffic if looping traffic across a common link.

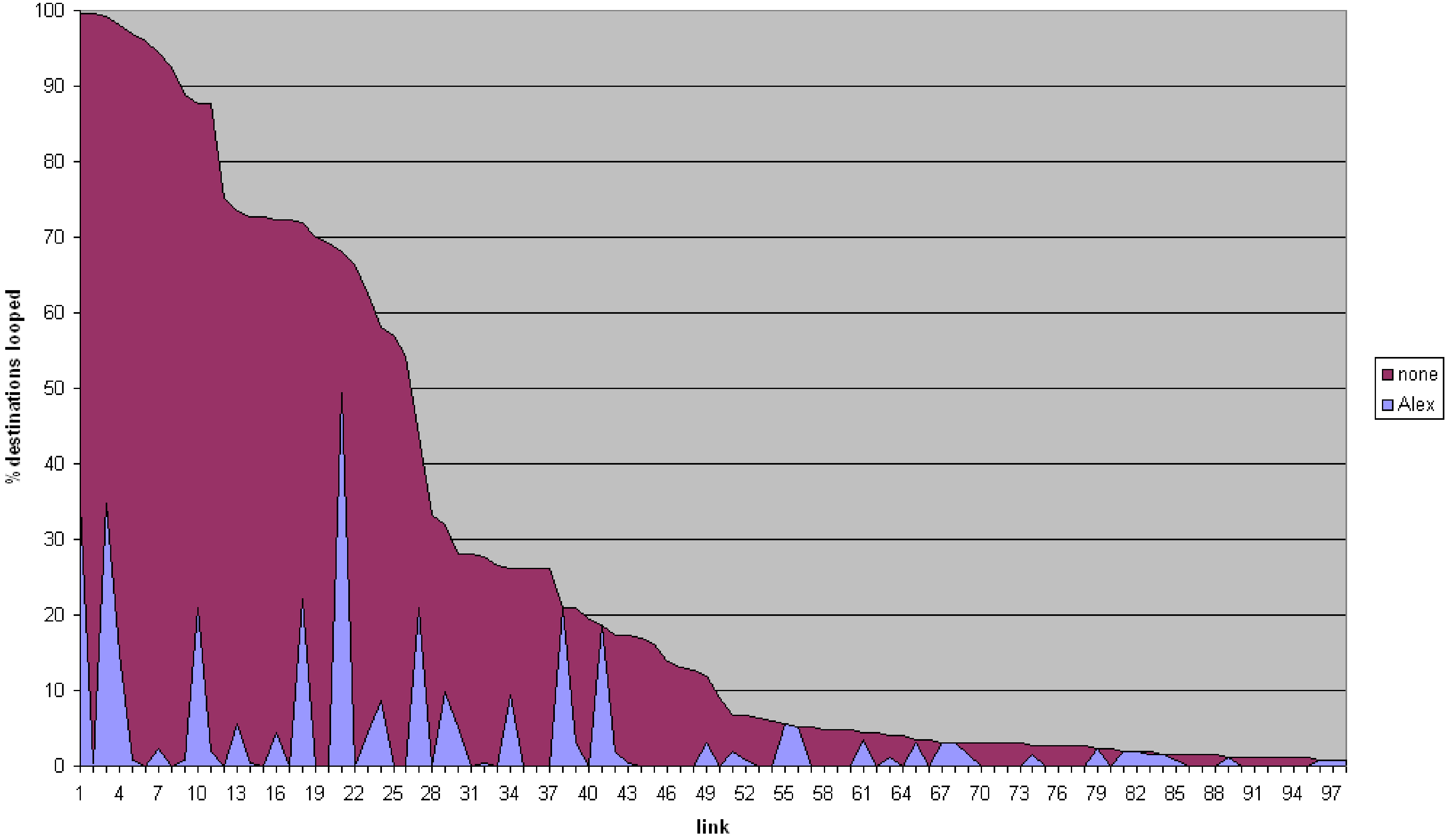
Path Locking Techniques

- Obtain a fixed convergence delay regardless of network.
- Avoid ordering issue by providing transitional paths.
- Handles SRLGs
- Different methods to
  - Determine/Create transitional paths
  - Direct traffic to use transitional paths

Standard trade-off of complexity versus coverage.

1. Tunnels for Transitional Paths
2. Safe Neighbors for Transitional Next-Hops
3. Marked Packets to Use Transitional Topology
4. U-turn Packets to Use New Topology

Typical Coverage

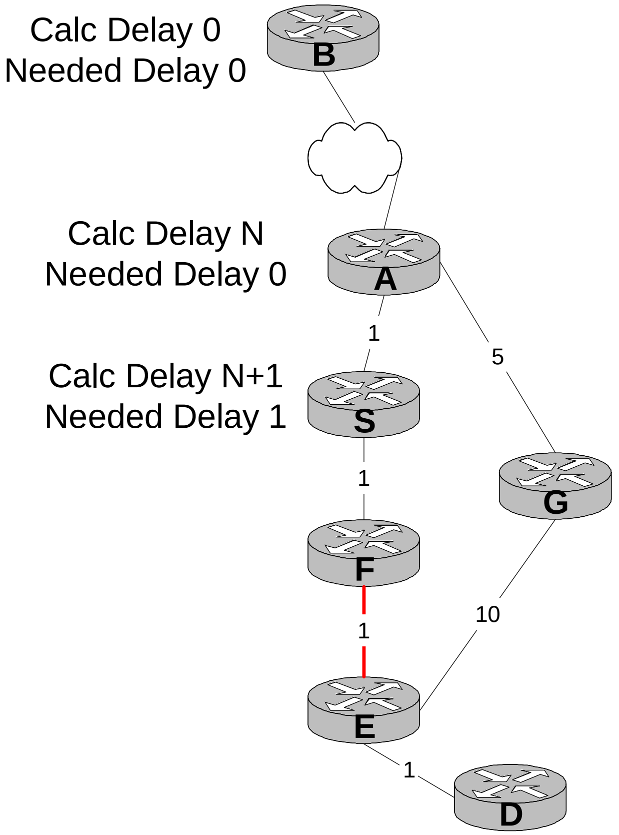


Ordered FIB Installation

- Determine “safe” ordering for FIB installation
  - bad news: update from edge to failure,
  - good news: update from change to edge
- Each router computes its “rank” with respect to the change.
- Delays for a number of worst-case FIB compute/install times proportional to its rank.

Delay Proportional to Network Diameter

- For Good News, rSPF gives necessary depth.
- For Bad News, rSPF is overly pessimistic for some topologies.
- Strategies to reduce unnecessary delay
  - Prune rSPF by only considering the branch across the failure – but still too pessimistic.
  - Run SPF rooted at edge nodes to correctly prune them – but doesn't scale.



Avoids all micro-loops and requires single FIB install per prefix

Delay dependent on network diameter so may be unacceptable.

Ordered FIB changes

- For any isolated link/node change
- Determine “safe” ordering for FIB installation
  - bad news: update from edge to failure,
  - good news: update from change to edge
- Each router computes its “rank” with respect to the change.
- Delays for a number of worst-case FIB compute/install times proportional to its rank.

Ordered FIB Installation Summary

- No forwarding changes required.
- No signalling required at time of change.
- Complete prevention of loops for isolated SRLG, node, or link changes.

- Requires cooperation from all routers
- May delay re-convergence for tens of seconds
- SRLGs require per destination delays and may delay re-convergence more.

See draft-bryant-shand-lf-conv-frmwk-00

Ordered SPF Summary

- No forwarding changes required.
- No signalling required at time of change.
- Complete prevention of loops for isolated node or link changes.

- Requires cooperation from all routers
- May delay re-convergence for tens of seconds (unless optional signalling used)
- SRLGs require per destination delays and may delay re-convergence more.

Method Comparison

Name	Pros	Cons	Delay (in FIB compute/install
Ordered SPFs	Control plane only.	SRLGs require destination-based decision.	High (Bounded by network diameter)
Path Locking	Deals with SRLGs & uncorrelated changes. Various depending on sub-method	Completeness depends on additional forwarding mechanisms.	Small (3)

Outline of Talk

- Motivations for IP Fast-Reroute
- Basic Framework & Loop-Free Alternates
- Advanced Mechanisms
  - U-turn Alternates
  - TE Tunnels
  - Tunnels with Directed Forwarding
  - Mechanism Comparisons
- Micro-Forwarding Loop Prevention
- Summary and Questions

Need for Requirements Draft or better, vocal operators

- There are X repair mechanism and Y convergence mechanisms on the table.
- Each mechanism has distinct advantages and disadvantages.
- Only objective way to chose between the mechanisms is to start from an understanding of the operator requirements.

What does it cost.1?

- Requires more spfs to be run on each node
  - More control plane work means CPU and OS busier
- More memory used to store backup paths
- Doesn’t require ubiquitous network deployment but, any network holes “don’t help”
  - No flag days required
- Some solutions require protocol extensions
- Some solutions require per packet marking
- Not all solutions protect 100% of the links

What does it cost.2?

- *Difficult to debug* during FRR event as it is defined to be transient
- Operational requirements (show commands, policy filters, additional data to be signaled, configured, ++) not complete
  - Still working on algorithms
- Not applicable for any Traffic Engineering - still need to lay out metrics correctly with failure models and traffic matrices

Multicast Challenges

- A number of tentative solutions have been proposed.
- Always needs at least one level of tunnel encapsulation in addition to unicast repair mechanism
- When to start and stop sending to or accepting from alternates

Multiple simultaneous failures

- Cannot pre-compute (don’t know what else might have failed)
- For two failures:-
  - Repairs may be completely independent (which works)
  - One may traverse the other failure (which works)
  - Each may traverse the other’s failure (which fails due to looping)
- Simplest approach is to fall back to conventional convergence when a second unrelated LSP is received.

Key Questions 1

1. Is 500ms (from conventional fast convergence) good enough?
2. Is this worthwhile, or should we rely on MPLS-FRR?
3. What services need to be protected (and how fast)?
4. Will you protect everything, or just strategic resources?
5. Do we need full or partial coverage of the attempted protection

   - By link,

   - By node,

   - By prefix
6. If partial – how partial?
7. How predictable does the coverage need to be?
8. Do we need to support - IGP, Multicast, BGP?

Key Questions - 2

9. Can the network be re-engineered to help?

   - change costs, add/rm links (change network design), etc.

10. Any allergies, phobias or religious problems?
11. How understandable does the repair process need to be?
12. What debug support is required?
12. Multiple failures – non SRLG?
13. Where in the network will this be used – core – edge – all?
14. Is this strictly an SP problem, or are Enterprises also interested?

Key Questions 3

15. Timing of cutover, post-failure convergence, re-computation of repair strategy? (second and third stages could be 10s of seconds)

- Type of encapsulation considerations?
- Is miss-ordering during transition to repair important?
- Is miss-ordering during convergence important?
- How do rate simplicity vs completeness?
- Should the IPFRR/micro-loop strategy be common with one that works with LDP?

16. What level of complexity do you want in the SW solution vs network design solution?

IPFRR Architecture Conclusions

- IPFRR provides a way of computing and using backup/repair paths on IP networks
  - There are many solutions to the problem and more emerging
  - ***Complete link, node and SRLG protection end to end***
- IPFRR uses SPF/reverse-SPF algorithm to compute repair paths (well known algorithms)
- IPFRR is NOT a replacement of MPLS-FRR
  - Same service on different networks
  - Can build networks with multiple, complementary repair techniques

Real World Analysis

Gagan Choudhury - ATT

Coverage & Capacity Needs for Example Core Networks

- We analyze the performance of IP Fast Reroute techniques in example realistic networks
- Techniques Considered:
  - Loop-Free Alternate and U-Turn Alternate
- Protection type
  - Link Failure Only, and Router + Link Failures
- Performance Measures
  - Coverage/Efficiency: Fraction of traffic loss that can be protected
  - Capacity Need: Amount of additional capacity needed in the network to account for fast reroute

Networks Under Study

- Network 1
  - An IP Backbone Network With 36 Routers and 49 Links
  - A Point-to-Point Traffic Matrix Between Every Pair of Routers
  - The Network and Traffic Matrix roughly similar to an existing Network
  - An Enhanced Version of the Network Also Considered with two Additional Links to Improve The Fast Reroute Performance
- Network 2
  - An IP Backbone + Access Network With 93 Routers and 180 Links
    - Backbone: 30 Routers and 54 Links
    - Access: 63 Routers and 126 Links
  - A Point-to-Point Traffic Matrix Between Every Pair of Access Routers
  - The Network and Traffic Matrix roughly similar to a Planned Future Network

IP Fast Reroute efficiency

- For Every Single Failure Scenario we define the quantity

  IP Fast Reroute Efficiency = (X - Y)/X

  - X is the amount of traffic that would be lost if no IP Fast Reroute were used
  - Y is the amount of traffic that would be lost in the presence of IP Fast Reroute
  - For destinations without an alternate next hop, traffic to that destination would be lost and contribute to Y.
- Show average value over all failure scenarios of a given type.
  - Each scenario of a given type considered equally likely
  - Types are single link failure and single router failure.

Some Link Failure Scenarios

Network Type	Fast Reroute Efficiency Averaged Over all Single Link Failure Scenarios	
	Only Loop-Free	Loop-Free + U-Turn
Network 1	33%	88%
Network 1 + Two Added Links to Improve Efficiency	43%	98%
Network 2	89%	100%

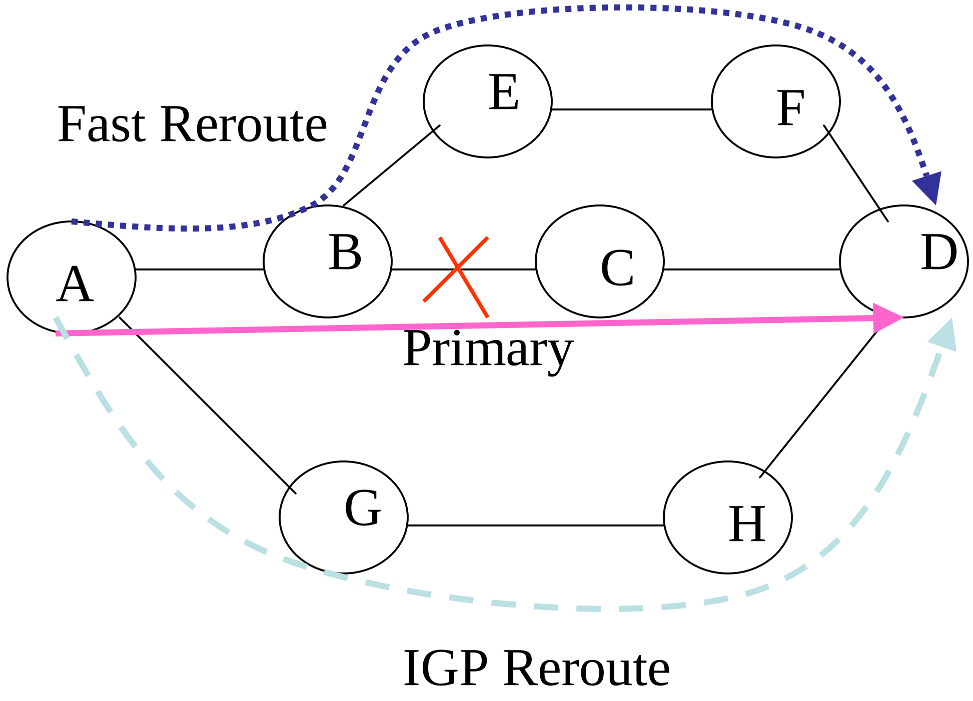
Link + Router Failure Scenarios and IP Fast Reroute End-to-End Versus Backbone Only

Loop-Free + U-Turn Alternate Applied to Network 2

Failure Scenario	Fast Reroute Algorithm Applied at	Fast Reroute Efficiency Averaged Over		
		All Backbone Link Failures Only	All Backbone + Access Link Failures	All Backbone Router Failures
Link Failures Only	Backbone Only	100%	72.5%	
	Backbone + Access	100%	100%	
Link + Router Failures (Avoids Router Even on Link Failures)	Backbone Only	86.2%	66.3%	53.3%
	Backbone + Access	86.2%	74.6%	64.7%

Extra Resource Needed With IP Fast Reroute

- A Failure Event Will First Trigger IP Fast Reroute and Then IGP Reroute
- The Path Taken By Fast Reroute and IGP Reroute Need Not Be The same and we have to keep enough capacity on both paths



Resource Needs for IP Fast Reroute

(Network 2, Router Avoiding Loop-Free + U-Turn Alternates)

- Network Designed To Have Enough Capacity Under
  - Every Single Link Failure Scenarios
  - Every Single Router Failure Scenarios
- 85% Max Link Utilization Allowed During IGP Reroute
- 85% or 110% Max Link Utilization Allowed During IP Fast Reroute

Backbone Resource Need	Only IGP Reroute	IGP + Fast Reroute With Max Link Utilization During Fast Reroute at	
		85%	110%
Total OC-48s	829	864	834
Total OC-48 Miles	574,992	616,571	583,232

Network Analysis Conclusions

- For Link Failure Scenarios, IP Fast Reroute Efficiency of
  - Loop-Free Alternate may vary wildly depending on the network topology
  - Loop-Free + U-Turn Alternates is usually very high (88-100%)
  - Strategically adding some links may greatly enhance efficiency (from 88% to 98% for Network 1 with Loop-Free + U-Turn Alternates)
- Using Router-Avoiding Alternate Next-hops avoids the Potential of Routing Loops Following a Router Failure but it may also reduce the fast reroute efficiency
- It is possible to have IP Fast Reroute only in the Backbone but with reduced efficiency
- Since IP Fast Reroute Path and IGP Reroute Path are not the same, it is necessary to have some additional capacity to support Fast Reroute
  - In a particular example, the additional backbone resource need is around 4% in terms of number of OC48s and around 7% in terms of OC48-Miles
  - The Additional Resource Need May Be significantly Reduced by allowing higher utilization during Fast Reroute (this would result in some lower-priority packet loss)