IP Fast Reroute

Peter Psenak

IOS XR Engineering

ppsenak@cisco.com

Agenda

- Definition and Goals
- Terminology
- Overview
- Backup Path Computation
- Topology Coverage
- RIB, LDP and FIB Operation
- Availability
- Summary

IP Fast Reroute: Definition & Goal

- IP Fast Reroute is a mechanism that enables a router to rapidly switch traffic following an adjacent link and/or node failure, towards a pre-computed/pre-programmed loop-free alternative (LFA) path.

- The goal of LFA FRR is to reduce failure reaction time to 10s of milliseconds by using a pre-computed alternate next-hop, in the event that the currently selected primary next-hop fails, so that the alternate can be rapidly used when the failure is detected.

Terminology

IPFRR	IP Fast Reroute
LFA	Loop-free Alternate
Primary Path	Path that traffic takes under normal circumstances
Backup Path	Loop free path that traffic takes immediately after the failure
Post Conv. Path	Path that traffic takes after the network converges

IP/LDP LFA FRR: Overview

- **LFA Computation**:

  LFAs can be computed by IGP in two ways:

  Link-based (per-link)

  Prefix-based (per-prefix)

  Per-Link:

  In link-based LFAs, all prefixes reachable through the primary (protected) link share the same backup information. This means that the whole set of prefixes sharing the same primary also share the repair/FRR ability.

  Per-Prefix:

  Prefix-based LFAs allow computing backup information per prefix. Thus, repair/backup info computed for a given prefix using prefix-based LFA may be different than one computed by per-link LFA.

- The existence of a suitable LFA next-hop is dependent on the topology and the nature of the failure for which the alternate is calculated.

IP/LDP LFA FRR: Overview (2)

- **FRR triggers**

Once backup (repair) information has been pre-programmed in forwarding, a trigger is required to switch forwarding from primary information to using backup information.

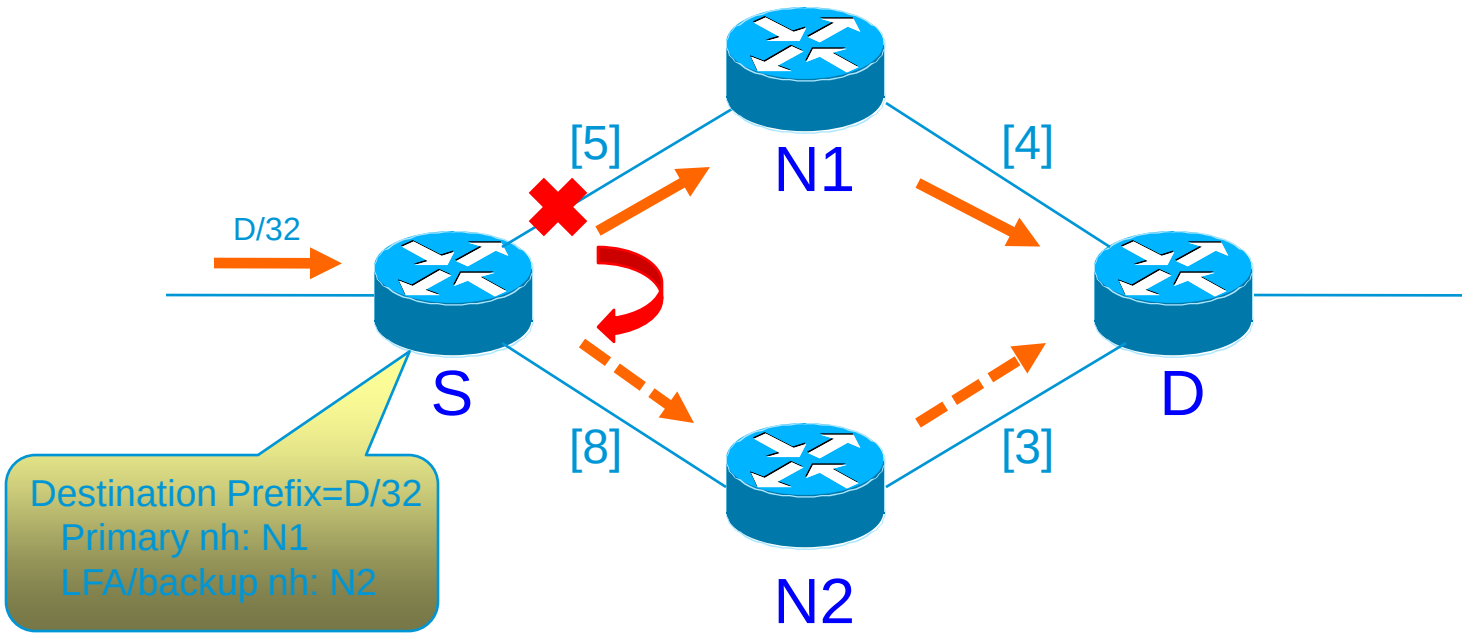
Supported FRR triggers are:

- Interface/link down
- BFD detection

- **Protection Speed**

sub-50 msec ,assuming that trigger/detection completes in < 10 msec.

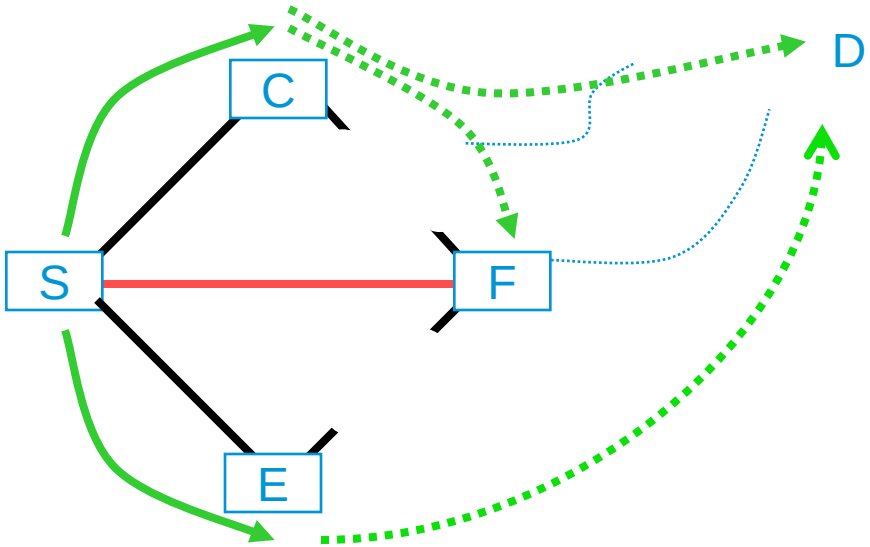
IP/LDP LFA FRR: An Example



NOTE:

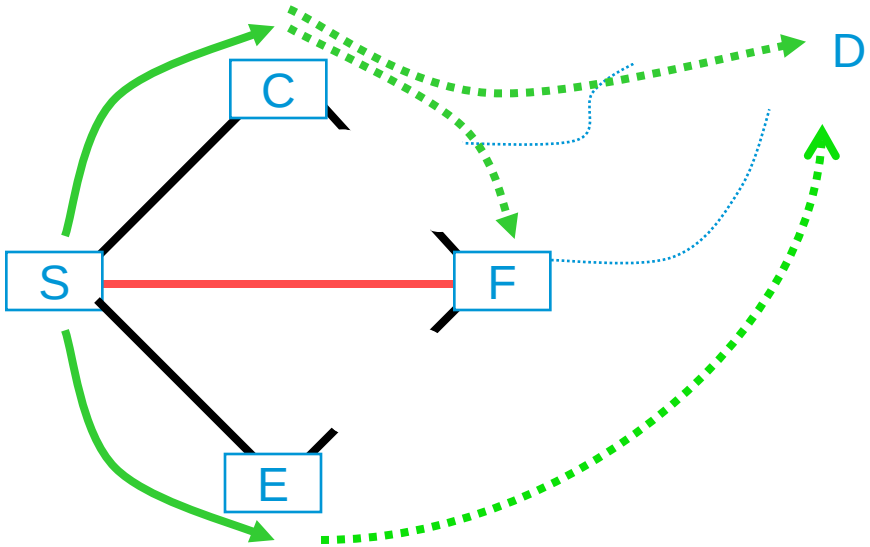
If cost of N2-D link >= 17 (N2-S + S-N1 + N1-D), then S can not elect N2 as LFAfor D upon S-N1 failure.

IPFRR – Per-link Computation



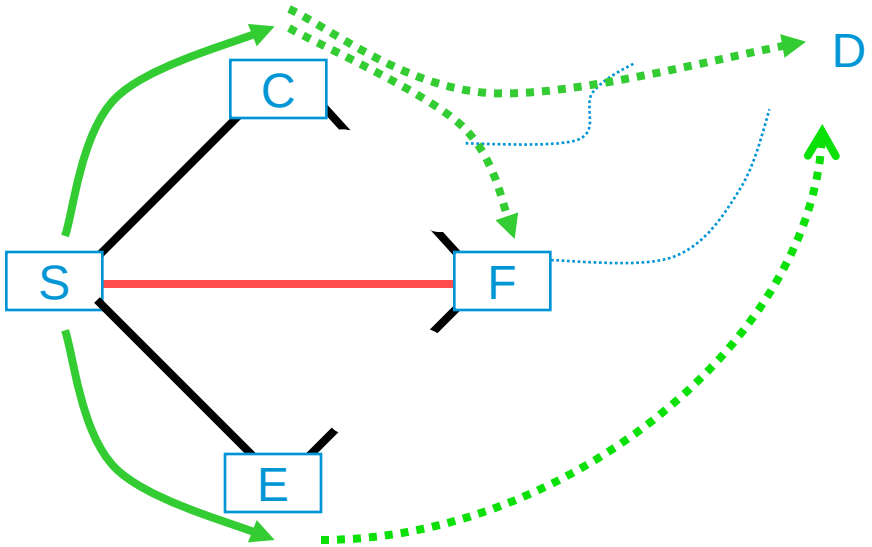
- Primary next-hop on S for destination D is F
- Can a packet that uses F as next-hop be sent toward F via any other direct neighbor of S without the neighbor sending it back to S?
- If such a neighbor is found, it can be used as backup next-hop

IPFRR – Per-link Computation (2)



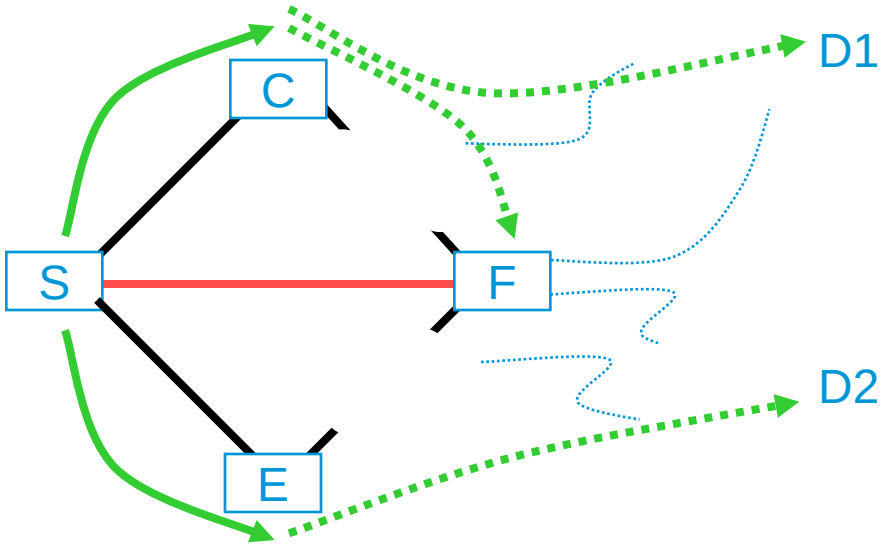
- This backup next-hop is valid for all prefixes that use F as primary next-hop
- During the calculation the assumption is made that the traffic will pas through F. In reality, traffic can be diverted to it's best path to D before it reaches F, completely avoiding F.

IPFRR – Per-link Computation (3)



- If C is chosen as a backup NH, C send traffic to D avoiding F. In this case the backup NH provides node protection as well
- If E is chosen a a backup NH, E sends traffic to D via F. In this case the backup NH provides link protection only.
- Per link IP FRR guarantees link protection only. Based on the topology, it may provide node protection too.

IPFRR – Per-prefix Computation



- For IGP route D1, S's primary path is link SF.
- S checks for each neighbor N (<>F) whether ND1 < NS + SD1

  "does the path from the neighbor to D1 avoid me?"

  If so, it is a loop-free alternate (LFA) to my primary path to D1

  Computation is done for each prefix independently

Backup Path Selection

- Single backup path selected for each primary path from the set of backup paths
- Per-link computation
  - First backup path selected – limited computation
  - ECMP paths to the protected neighbor are used as last resort
- Per-prefix computation
  - All possible backups are evaluated
  - Tie-breakers are used to select the best backup among them
  - CLI available to weight tie-breakers and influence selection
- Links can be excluded from becoming backup via CLI
  - Available with both per-link and per-prefix computation

Per-prefix LFA Backup Path Selection

- Default tie-breakers
  1. Prefer node protection
  2. Prefer line-card disjoint
  3. Prefer lowest backup metric
  4. Prefer backup path from the set of primary ECMP paths

- CLI to customize the tie-break policy

  Default is recommended. Simplicity.

Per-link versus Per-prefix computation

- Per-link computation
  - Simpler computation – single rSPF per protected neighbor
  - No prefix processing
  - Less CPU/memory intensive
  - Smaller coverage
- Per-prefix computation
  - More complex computation – full Dijkstra per neighbor
  - Per prefix backup computation
  - More CPU/memory intensive
  - Better Coverage
  - Better controlled backup selection from the available set of backups
- Selection should be based on the topology and required coverage

IGP Computation

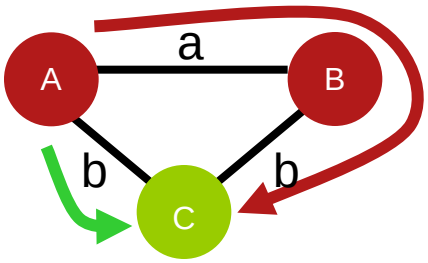
- Backup paths computation is done independently of the primary SPF and only after primary SPF is finished
  - there is a 500ms wait time after the last SPF finished, before the backup computation starts
- Backup paths computation is done in pieces, during the periods when no primary SPF is needed
- When an event that requires primary SPF arrives and SPF is scheduled, ongoing backup paths computation is aborted
- The objective is to not impact primary path convergence

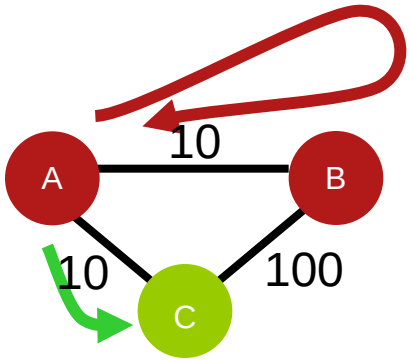
Backup path availability - Coverage

- All depends on the topology AND metric assignments
- Per-prefix computation has better coverage
- In some topologies backup may not be available

Backup path availability – Coverage (2)

- Triangles are usually good for both per-link and per-prefix computation
- Same coverage for both per-link and per-prefix
- All depends on metric assignment!!!





Backup path availability – Coverage (3)

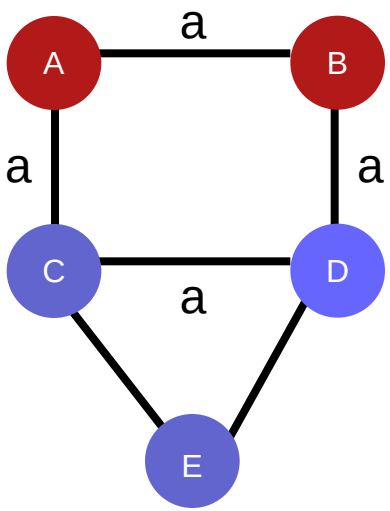
- Square topology

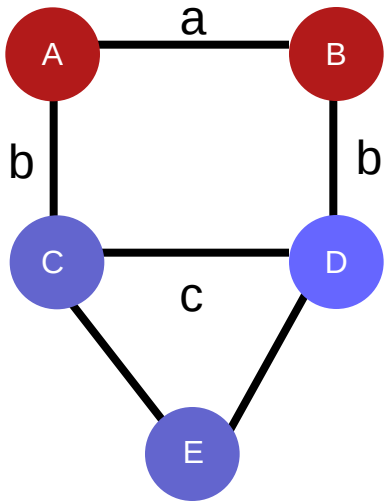
  Per-link computation:

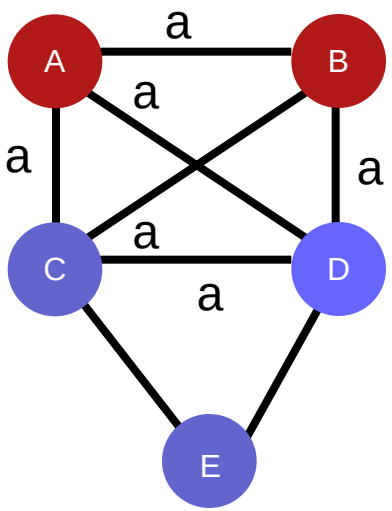
  If all links have same metric, no backup path is available for A-C link

  Tweaking metric may provide backup path for some nodes.
  If (b+c)<(a+b) then on A primary NH C has a backup NH B. There is no backup on C for NH A though anymore. (*It makes more sense to protect the traffic in the C->A direction as more prefixes are affected by the failure in this direction* =>c>a)

  Adding a cross links makes the backup available for all NH on A (triangle)







Backup path availability – Coverage (4)

- Square topology

  Per-prefix computation:

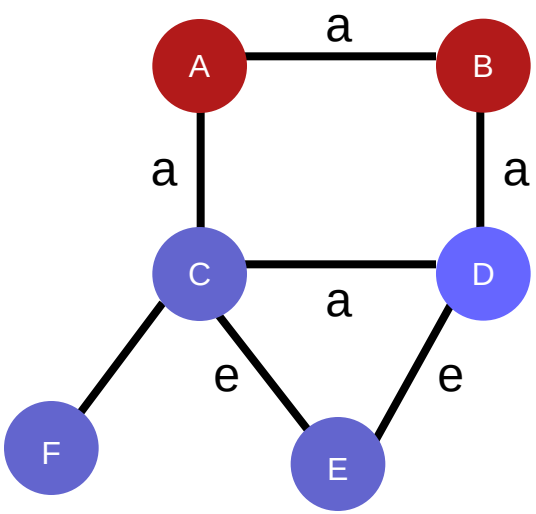
  Backup available to all prefixes behind E (assuming the E's uplinks have equal cost)

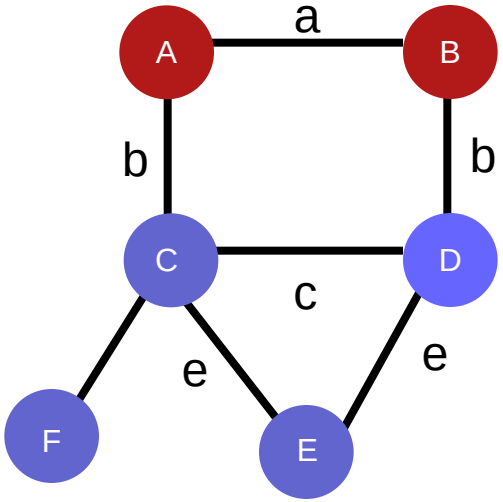
  Backup available on C for NH A for prefixes behind B

  If all links have same metric, no backup path is available for prefixes connected to C only – e.g. F

  Metric assignments is important

  Better coverage compared to per-link computation





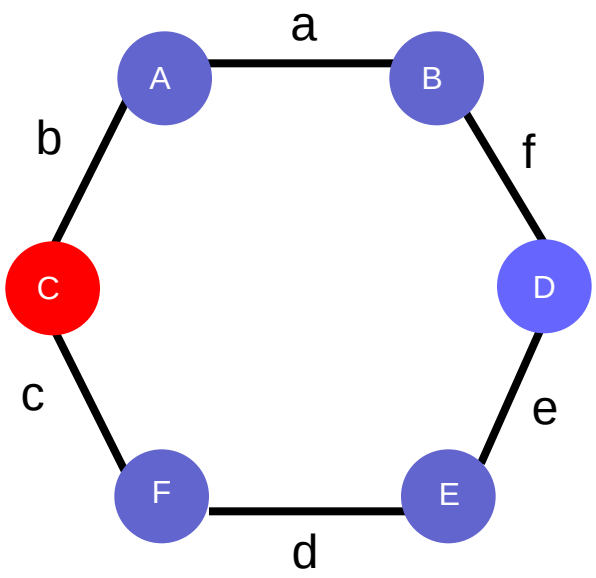
Backup path availability – Coverage (5)

	Per-link LFA Backup	Per-link LFA Node Protection	Per-prefix LFA Backup	Per-prefix LFA Node Protection
Triangle	Y	Y	Y	Y
Full-Mesh	Y	Y	Y	Y
Square	Limited to One Direction (1) applies	Limited to One Direction (1) applies	Y (2) applies	Y (2) applies

(1): Backup available in A->C or C->A direction only, not both.
(2): Backup not available for prefixes connected to or single homed to C. This can be fixed in one direction by metric tweaking

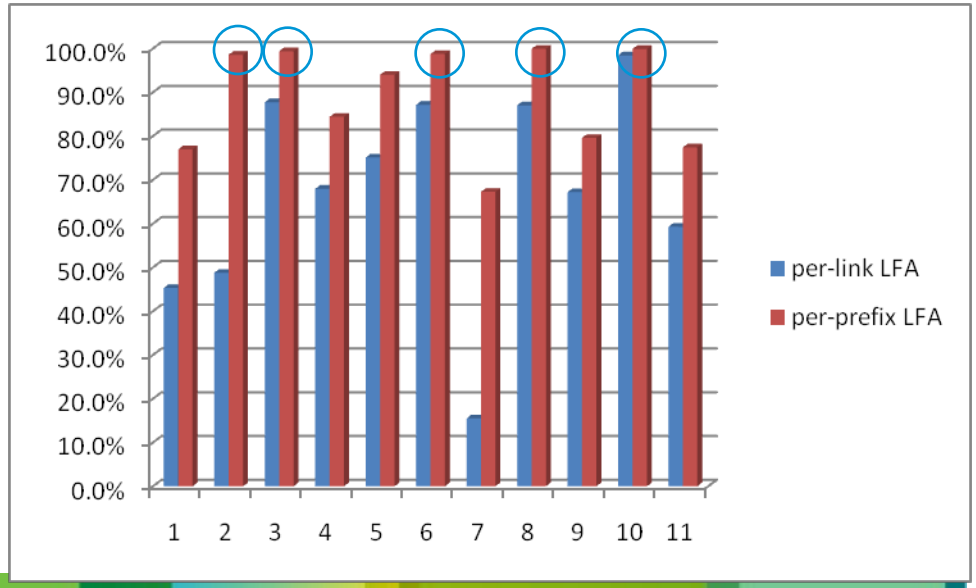
Backup path availability – Coverage (6)

- No LFA protection in the ring
  - A-C link failure on A
  - On B (a+b) < (f+e+d+c)
  - No backup path available
  - Solvable, but some way of tunneling required
  - Future work planned to cover ring topologies



Backbone Applicability

- Based on SP backbone topologies
  - Per-Link LFA: 70% of the links are protected
  - Per-Prefix LFA: 94% of the prefixes across all links are protected
  - implies a tight process to plan the topology



RIB, LDP and FIB operation

- Per-prefix FRR backup information is downloaded from IGP->RIB->FIB as part of the route update itself
- LDP associates local and output labels with each path (including backup paths) as usual
- LDP downloads labels as usual to FIB
- FIB programs backup information in hardware on LCs
- IGP backup path calculation algorithm (for e.g., per-prefix OR per-link LFA) is transparent to rest of components (RIB/LDP/FIB) that always see backup paths as per-prefix property
- On FRR trigger, traffic loss < 50msec is achieved by using the pre programmed info in HW to redirect the traffic on to backup paths.

LFA FRR Traffic Flows

- Traffic Flow types supported with LFA FRR (IOS-XR):

  - LFA FRR per-link and per-prefix LFA FRR supports IPV4, L2VPN, L3VPN, MPLS labeled unicast traffic.
  - LFA FRR does not support IPv6
  - mcast not supported

LFA FRR IOX Release 3.9

- OSPF per-link IPFRR is available from 3.9
- ISIS per-link IPFRR available from 3.5
- IPFRR forwarding is supported in 3.9 on following platforms:

  CRS

  Viking (ASR9k)

  - Only pure IPv4 unicast is supported, no MPLS
  - No Support for Bundle Interfaces
  - TE Tunnel is not available as a backup

LFA FRR IOX Release 4.0.1

- IP and MPLS traffic protection
- IPFRR supported on all supported IOS-XR platforms:
  - CRS-1, XR12000, ASR9000
- ISIS support for IPv4 per-link/per-prefix based LFA
- OSPF support for IPv4 per-link based LFA
  - Per-prefix based LFA is available in 4.2.0 (FCS 12/2011)
- IP/LDP LFA FRR feature is supported on “IPv4 unicast” family as follows:
  - IP (OSPF): Supported under both default and non-default VRFs
  - IP (ISIS): Supported only under default VRF
  - LDP: Supported only under default VRF

LFA FRR IOX Release 4.0.1

- LDP support for per-prefix LFA FRR (prefix-based and/or link-based IGP LFA)
- TE Tunnel used as an LFA
  - Must be manually specified on the candidate-list
- Bundle interface support
- Internal testing:
  - 15k IGP prefixes reprogrammed on FRR trigger in less then 50ms on all platforms

LFA FRR Feature Restrictions:

- A link can be either be configured for per-link LFA or per-prefix LFA. Not both on same link.
- A given interface should be configured for LFA FRR or TE FRR, not both on same link.
- GRE tunnel is not supported either as primary link or backup link.
- No LFA for primary paths whose output interface is via TE tunnel.

IOS Availability

- 15.1(2)S
  - •ISIS per-prefix LFA
  - •IP and MPLS/LDP support
  - •Platforms: 7600
- 15.1(3)S, XE3.4
  - •OSPF per-prefix LFA
  - •Platforms: ASR1k, 7600

Summary

- Simple
  - the router computes everything automatically
- < 50msec
  - pre-computed, pre-installed, enabled on link down event
- Deployment friendly
- no IETF protocol change, no interop testing, incremental deployment
- Requires topology planning
  - •backup availability depends on the topology
- LFA FRR is not meant to replace MPLS TE FRR
  - •replaces MPLS TE FRR when a simple solution exists

Thank you.