Fast Reroute

A high availability addition to MPLS

Shankar Rao
Sohel Ahmed
Qwest Communications

Richard Southern
Juniper Networks

Agenda

- Overview of MPLS FRR – what problem is this technology solving, and how does it work?
- Drivers for Qwest to implement FRR –alternative options evaluated, and why FRR?
- Real-world scenarios experienced on the Qwest network – did FRR help?
- Operational lessons learned, what can we do better?
- Conclusions

Fast Reroute

What is it?

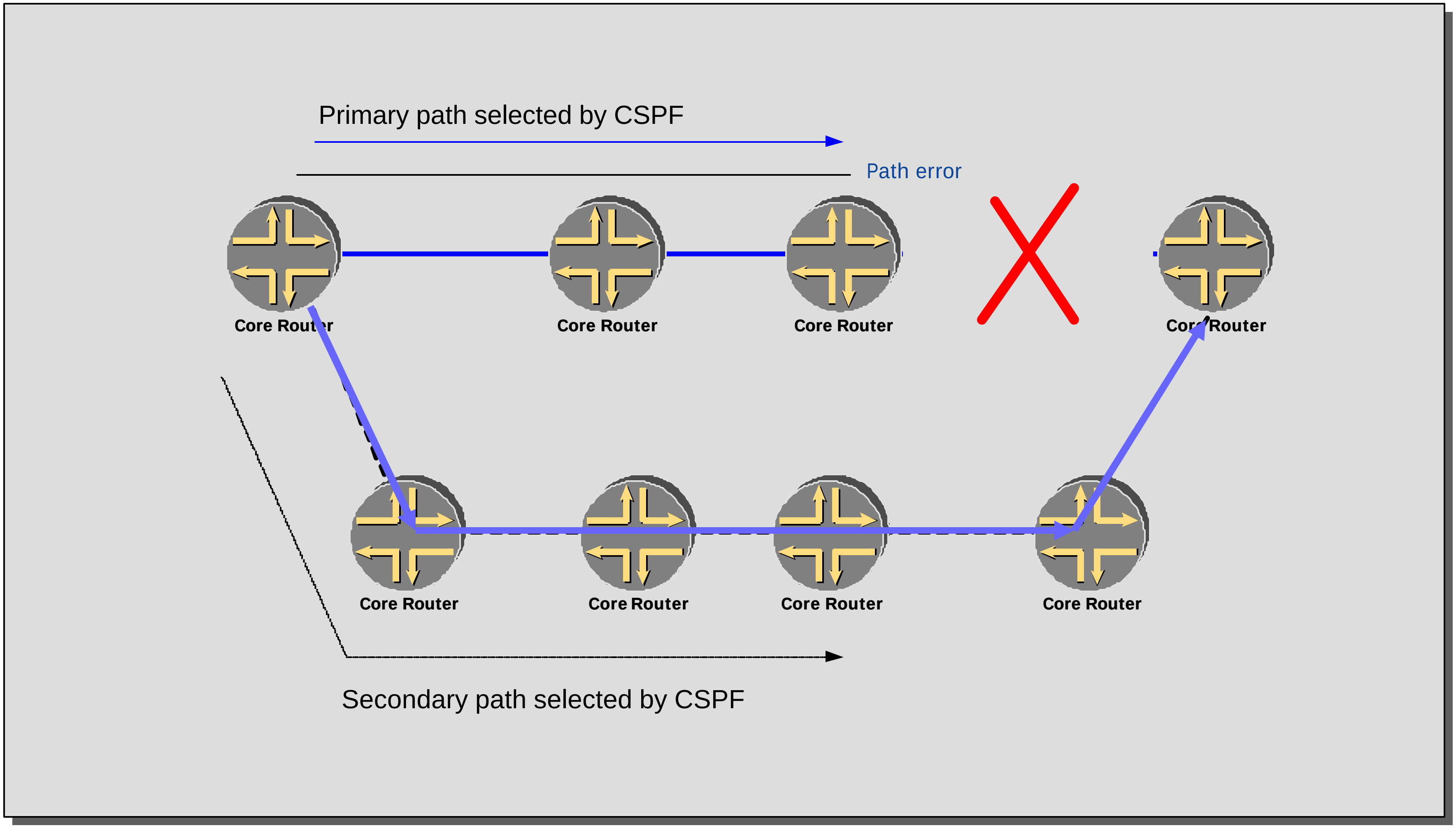
Planning for a failure (things to consider)

- Control plane failures
  - Graceful restart
    - Implemented in each protocol
- Data plane failures
  - L2 based solutions (for comparisons sake)
    - APS
    - Link bundling /aggregated sonet or Ethernet
  - **MPLS+RSVP Choices for protecting a LSP**
    - **Secondary LSP**
    - **Secondary Standby LSP**
    - **Fast reroute**

Recovery Speed (Secondary LSP)

- Secondary and Standby
  - Secondary: Ingress LSR needs to signal new LSP when primary LSP fails
    - Patherr and resvtear unicast to ingress LSR
  - IGP needs to change nexthop @ ingress LSR
    - May be additional built in delays to optimize SPF runs
  - Standby path is pre-computed
    - Saves CSPF run
  - Sum of delays is in 100^smS to 1S range
  - Packet loss may occur until LSP is redirected by ingress LSR

Example Secondary LSP



Recovery Speed (FRR)

- **Each node along the LSPs path takes care of protecting the LSP. Request is made by including** `detour` **and** `fast-reroute` **objects in RSVP** `PATH` **messages**
- **Can be used with other protection methods since it's a quasi-L2 solution, including secondary LSPs**

Juniper NETWORKS

Recovery Speed (FRR) -continued

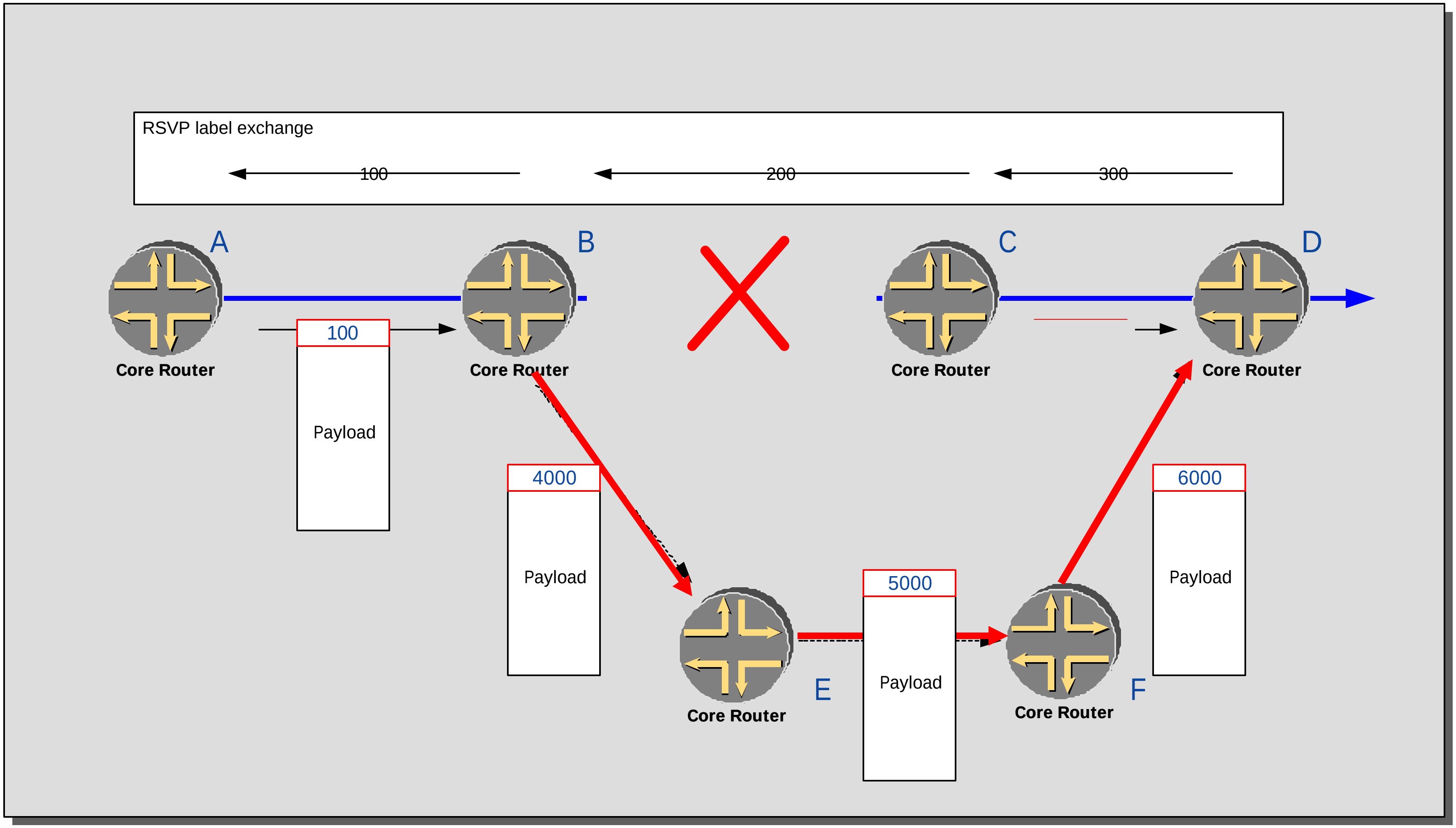
- **Delay in switching to FRR is limited by the failure detection delay and the propagation time to update the forwarding table of the change**
  - **Typically in the 10s to 100s of mS window (w/o F-FRR)**
    - **Sub 50mS numbers are possible if the other reroute labels are preloaded in the forwarding plane**
    - **50mS times are necessary for VoIP signal sync frames**
  - **Vendor specific implementation details may add extra time to the switch-over depending on IGP**
- **Packet loss is minimized to the 'unlucky few' that were transiting the link during the failure**

Juniper NETWORKS

FRR mechanisms (detour)

- **Detour (1:1)** (Juniper and Avici)
  - **Each LSP has its own detour LSP**
  - **Uses combined link and node protection**

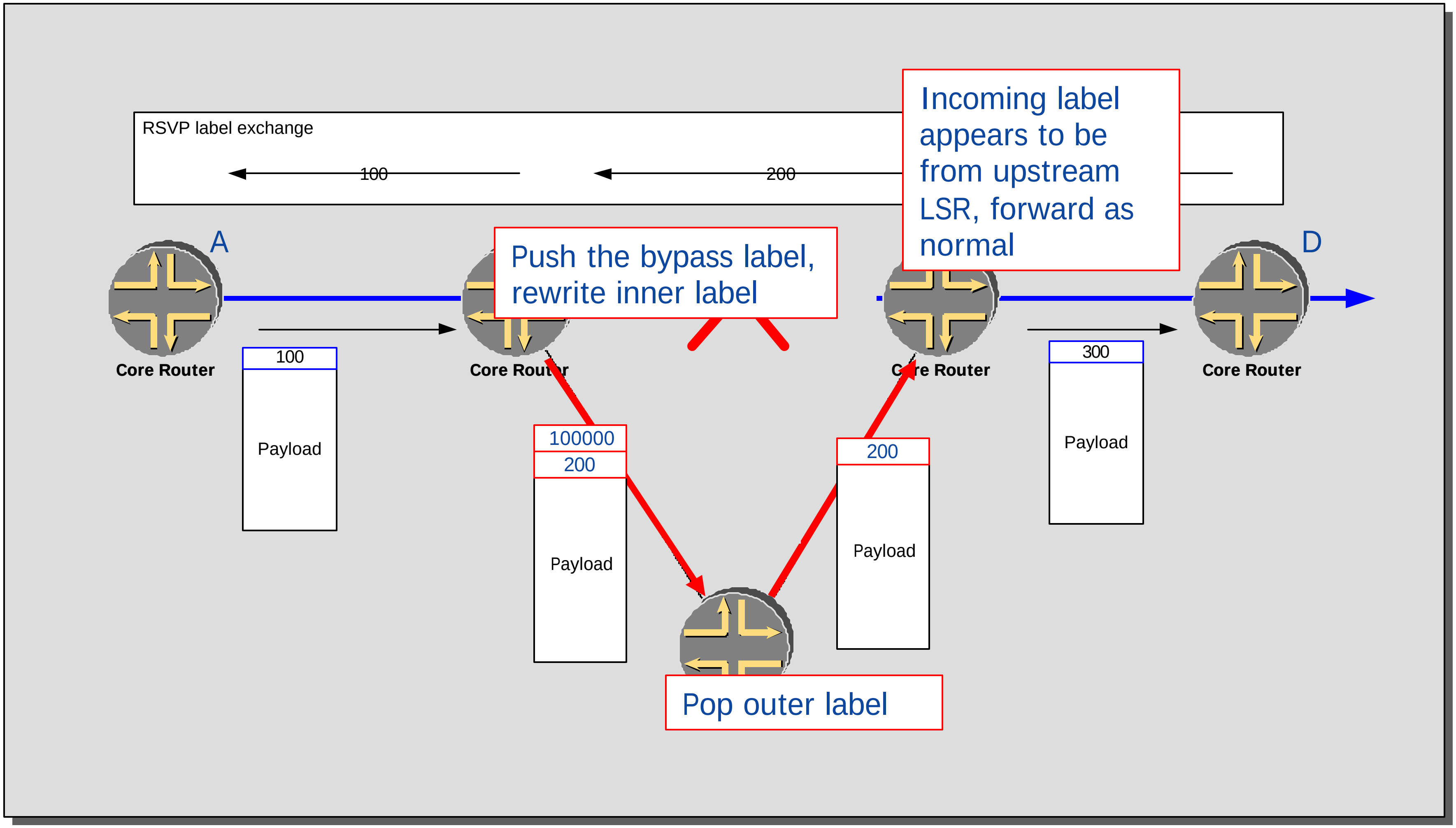
FRR Detour (link and node)



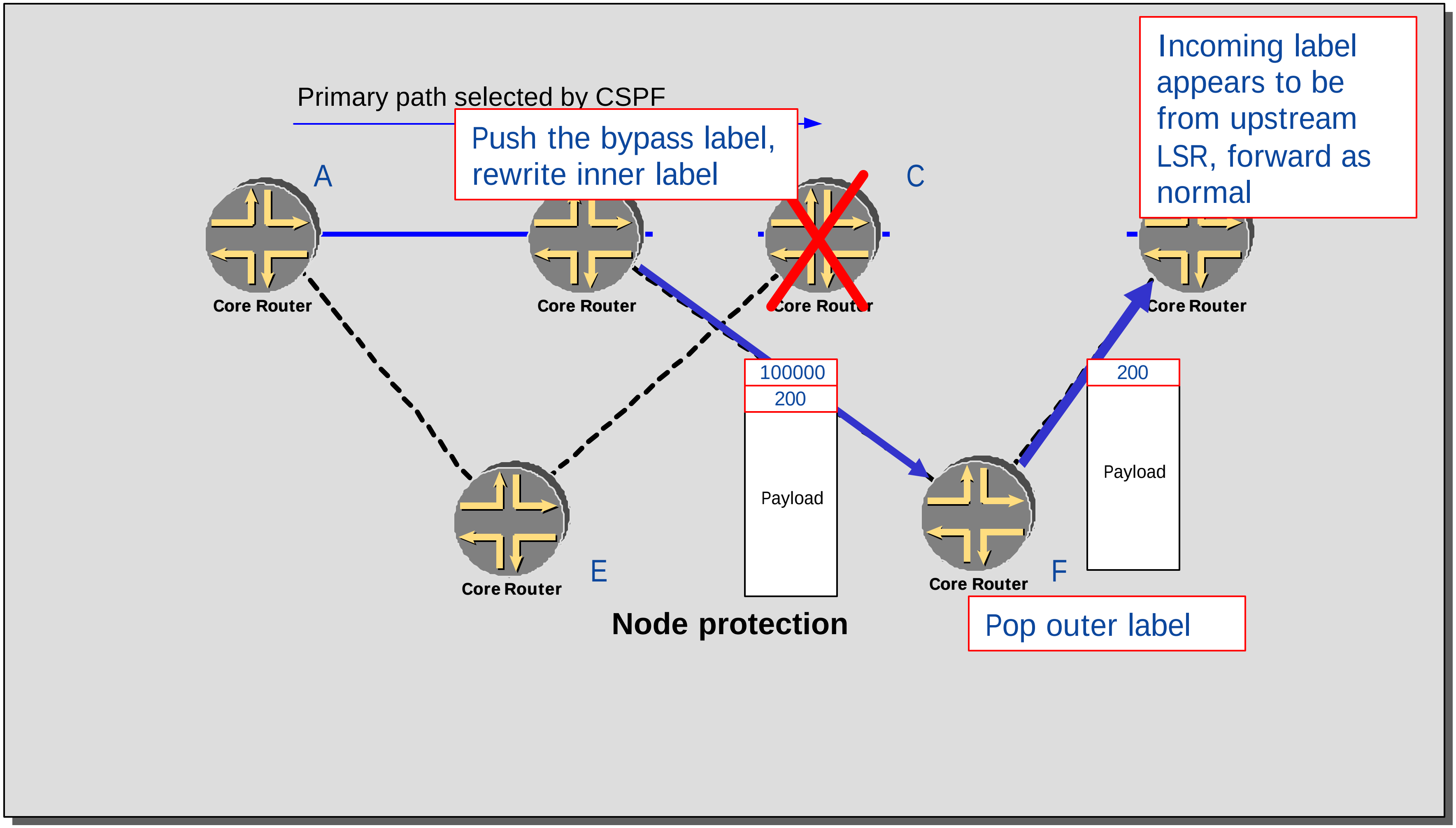
FRR mechanisms (facility backup)

- **Facility backup (N:1)**
  - **One or more LSPs share a common detour**
  - **Link protection (NHOP)** (Juniper, Cisco, Avici, others?)
    - **Merge Point (MP) is at the next hop, but on a different link**
    - **Protecting against multiple link outages**
    - **This is where most development time has been as ISPs have an immediate need to protect critical links**
  - **Node protection (NNHOP)** (Cisco)
    - **MP is at the next next-hop**
    - **This may be the next step, however graceful restart might work better here**

FRR link-protection bypass LSP



FRR using Node Protection



Complexity Comparison

- Secondary
  - Signaled by ingress LSR only, protects path
  - + additional constraints can be applied
  - + tries to stay away from primary path nodes and links
  - - additional management and planning
  - - switch is done at the ingress router only
- FRR
  - Each LSR along the path protects configured links
  - limited path constraints (can include BW, hold and setup priorities, links to avoid etc.)
  - + no additional path definitions configuration

Maintaining the protected LSP properties

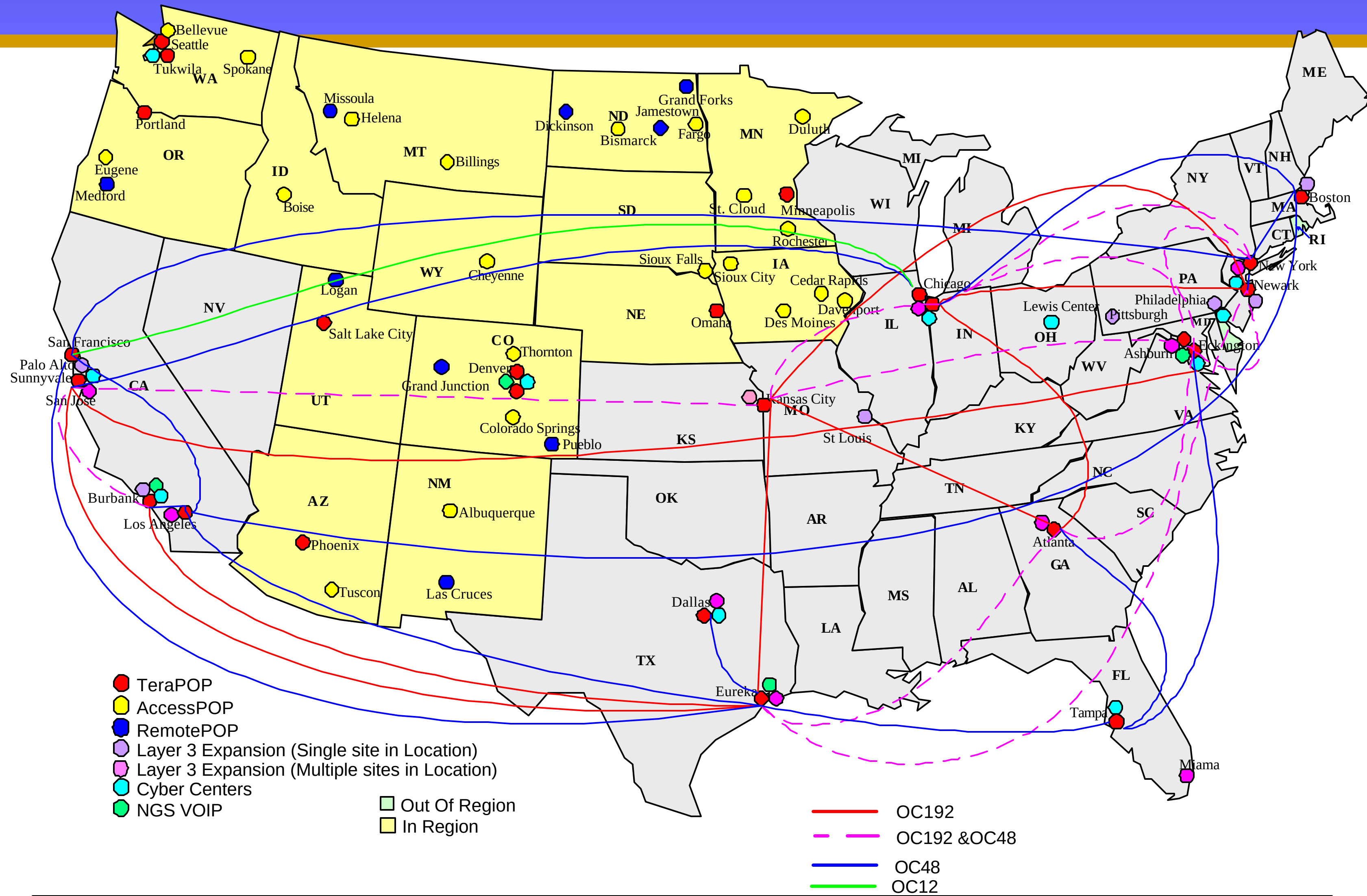
- **Secondary can…**
  - **make all the same BW requests as the primary**
  - **maintain CoS requirements**
  - **remain up even after primary path recovers**

- **FRR is intended as a short term fix**
  - **Builds on the existing LSPs properties since majority of the LSP will remain in place**
  - **BW may be shared in many-to-one (Facility) backup**
  - **Forward packets only until primary can handle the problem (may include a switch to secondary)**

Fast Reroute

Why on the Qwest network?

BB Logical



Last Update:
06/05/02
Gary Waldner

Recovery Requirements

Qwest. Spirit of Service

- OC48 SONET APS protected backbone circuits initially
- Partial mesh of OC192 unprotected wavelengths today, therefore need for protection at higher layer
- Higher layer protection must be comparable to SONET protection (~50ms)
- Voice/ATM traffic on IP network demanding stringent SLA's for network recovery (<100ms)
- Other SLA's - RTT < 100ms, Availability 99.999%, Packet Loss < 0.001%, Jitter < 5ms
- Need to protect (sub-second) against both link and node failures

Options

- IGP tweaks
  - http://www.nanog.org/mtg-0202/ppt/cengiz.pdf
  - Convergence times
    - Convergence as fast as today's technology allows, ~5secs.
    - Can be improved to sub second with enhancements to ISIS specification
- Graceful Restart Mechanisms (NSF)
  - Offers protection against RE/RP failures (by keeping such failures control plane transparent), but
  - Link failures/flapping links still a problem

Options

- Other HA mechanisms such as Stateful failover
  - RE/RP failures are transparent to peers/neighbors (sessions remain up)
  - Link failures/flapping links still a problem
  - Deployed/Field Tested implementations non-existent
- MPLS FRR
  - Both link and node protection possible
  - Promise of recovery times in order of 10's of ms, however,
  - Proprietary implementations
  - New technology, burden of operationalizing

Fast Reroute

Operational considerations

Failure Scenarios

Qwest. Spirit of Service

- Good (or not so good) feature – link fails, traffic is re-routed over backup, primary re-optimized, all happens transparently
- No special MPLS FRR monitoring needed (in theory) – rely on existing NMS to flag link/node failures, FRR keeps traffic moving
- MPLS control plane anomalies harder to detect – primary/backup paths setup transparently, backup paths only used for short periods of time
- Worst case – if FRR croaks, IGP always available as backup

Operational Concerns

Qwest. Spirit of Service

- Detecting LSP Primary/Detour Outages
  - SNMP/Syslog tools - proactive monitoring of data/control plane
- Traffic Management
  - Not actively used
- Trouble Shooting RSVP/LSP's
  - Additional control plane protocols to be learned and understood
  - Change in data plane forwarding
- Bug report/analysis and testing
  - Extensive testing to ensure that worst case does not get worse (i.e. IGP routing as fallback works)
- Training NOC/NMC/TAC
  - Keep changes transparent (to the extent possible) to existing troubleshooting methods

Fast Reroute

Real-world Scenarios

MPLS FRR Topology
LSP: chi-core-03 to jfk-core-01

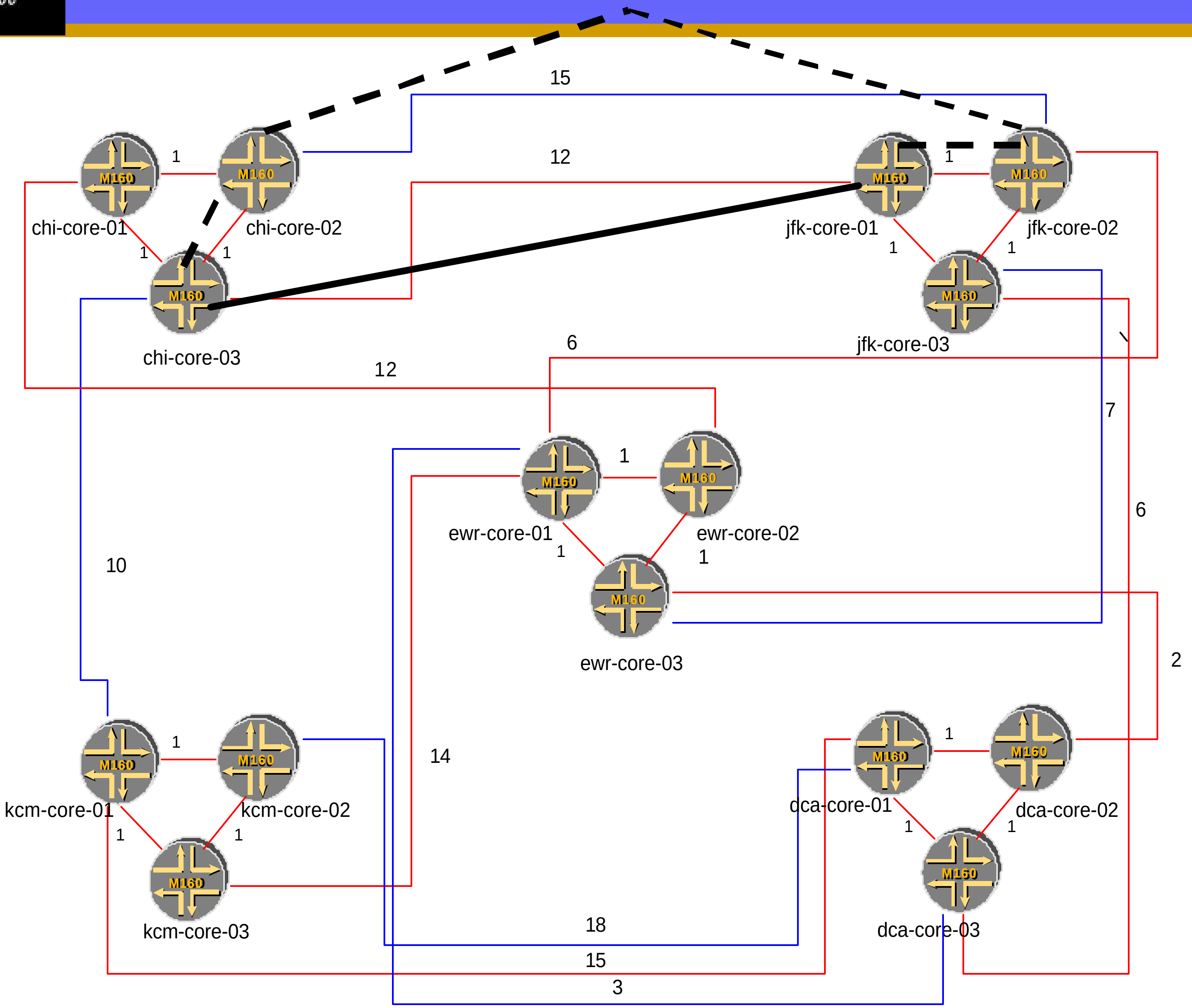
LSP Configuration

Primary LSP: chi-core-03 to jfk-core-01

ERO: chi-core-03, jfk-core-01

Detour LSP

ERO: chi-core-03, chi-core-02, jfk-core-02, jfk-core-01



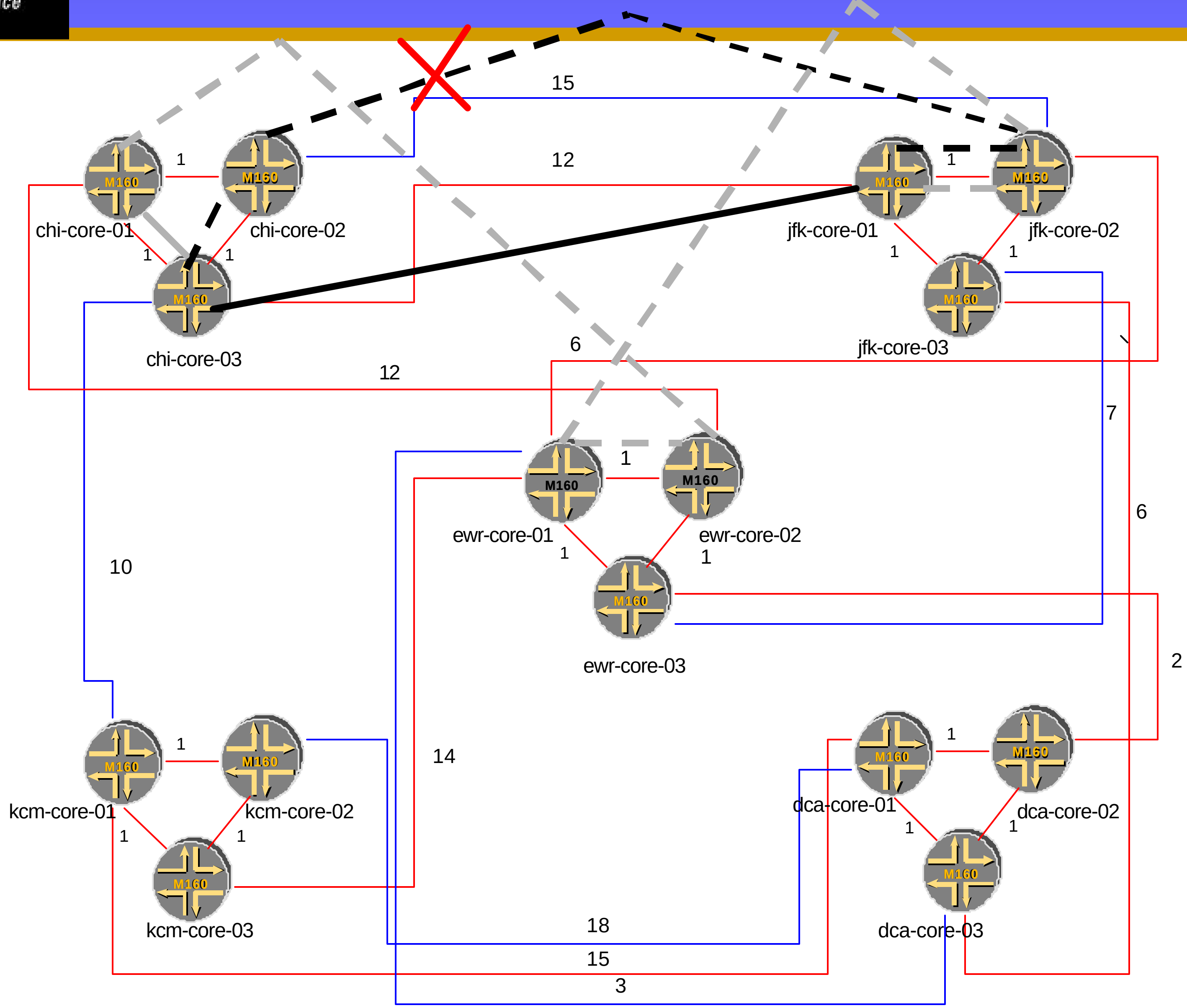
MPLS FRR Topology
LSP: chi-core-03 to jfk-core-01

Detour LSP goes down but Primary LSP remain active, a new Detour LSP establishes

Primary LSP: chi-core-03 to jfk-core-01

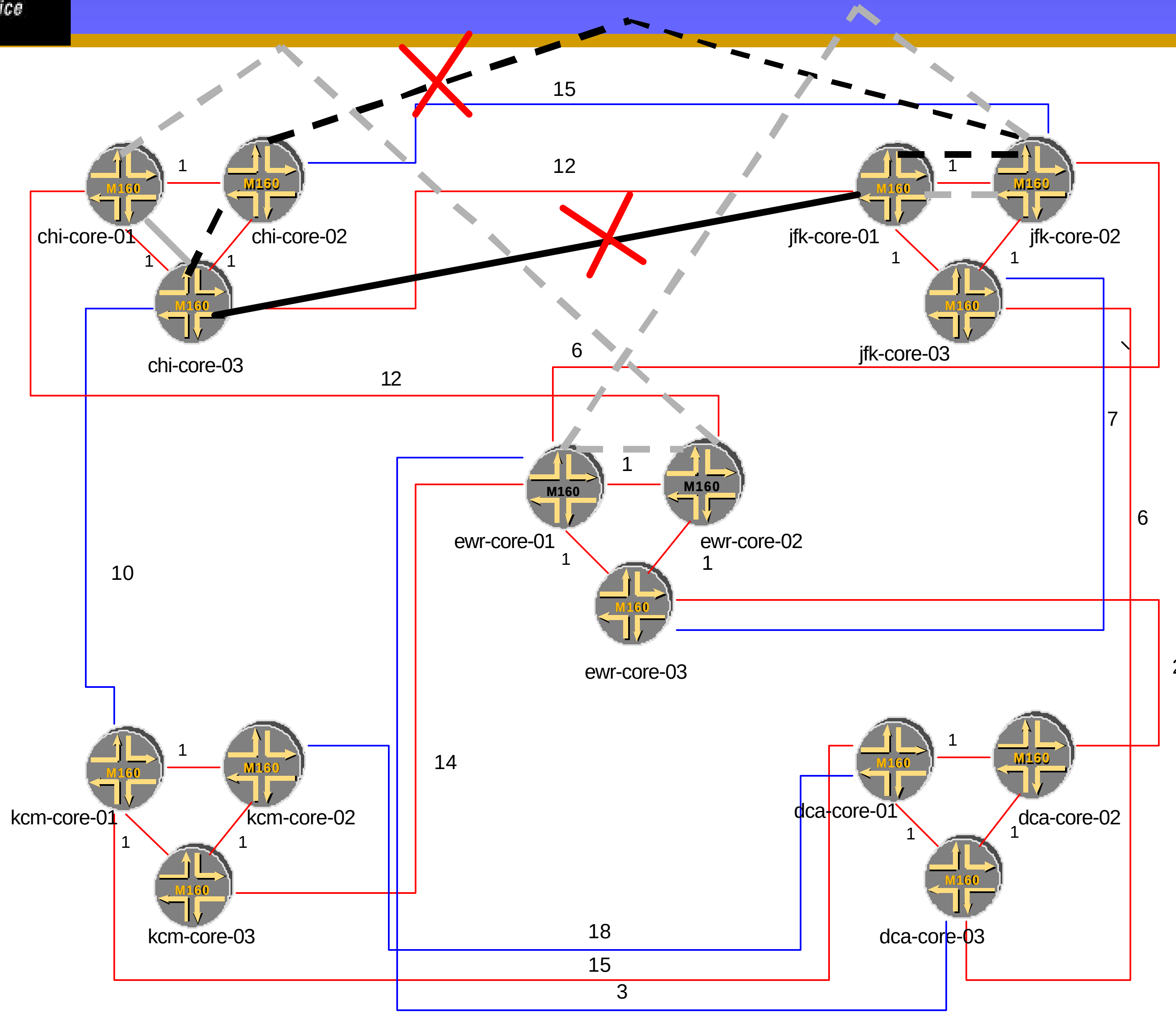
ERO: chi-core-03, jfk-core-01

New Detour LSP ERO: chi-core-03, chi-core-01, ewr-core-02, ewr-core-01, jfk-core-02, jfk-core-01



MPLS FRR Topology
LSP: chi-core-03 to jfk-core-01

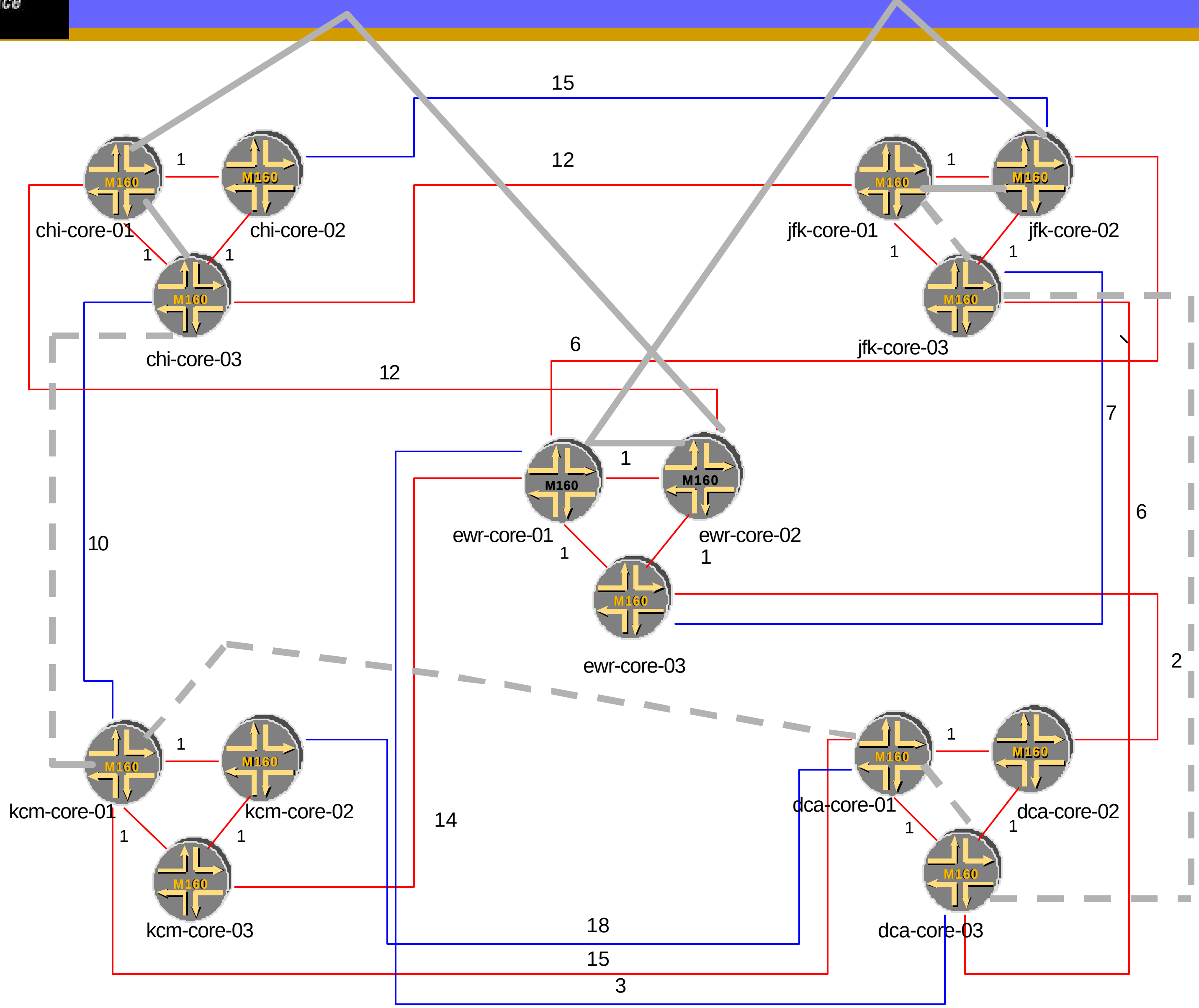
Primary LSP goes down as well, traffic switch to pre-configured detour immediately, after a CSPF calculation the existing detour becomes primary Primary LSP: chi-core-03 to jfk-core-01 ERO: chi-core-03, chi-core-01, ewr-core-02, ewr-core-01, jfk-core-02, jfk-core-01



MPLS FRR Topology
LSP: chi-core-03 to jfk-core-01

For the new primary LSP a new detour LSP establishes

ERO: chi-core-03, kcm-core-01, dca-core-01, dca-core-03, jfk-core-03, jfk-core-01



MPLS FRR Topology
LSP: chi-core-01 to ewr-core-01

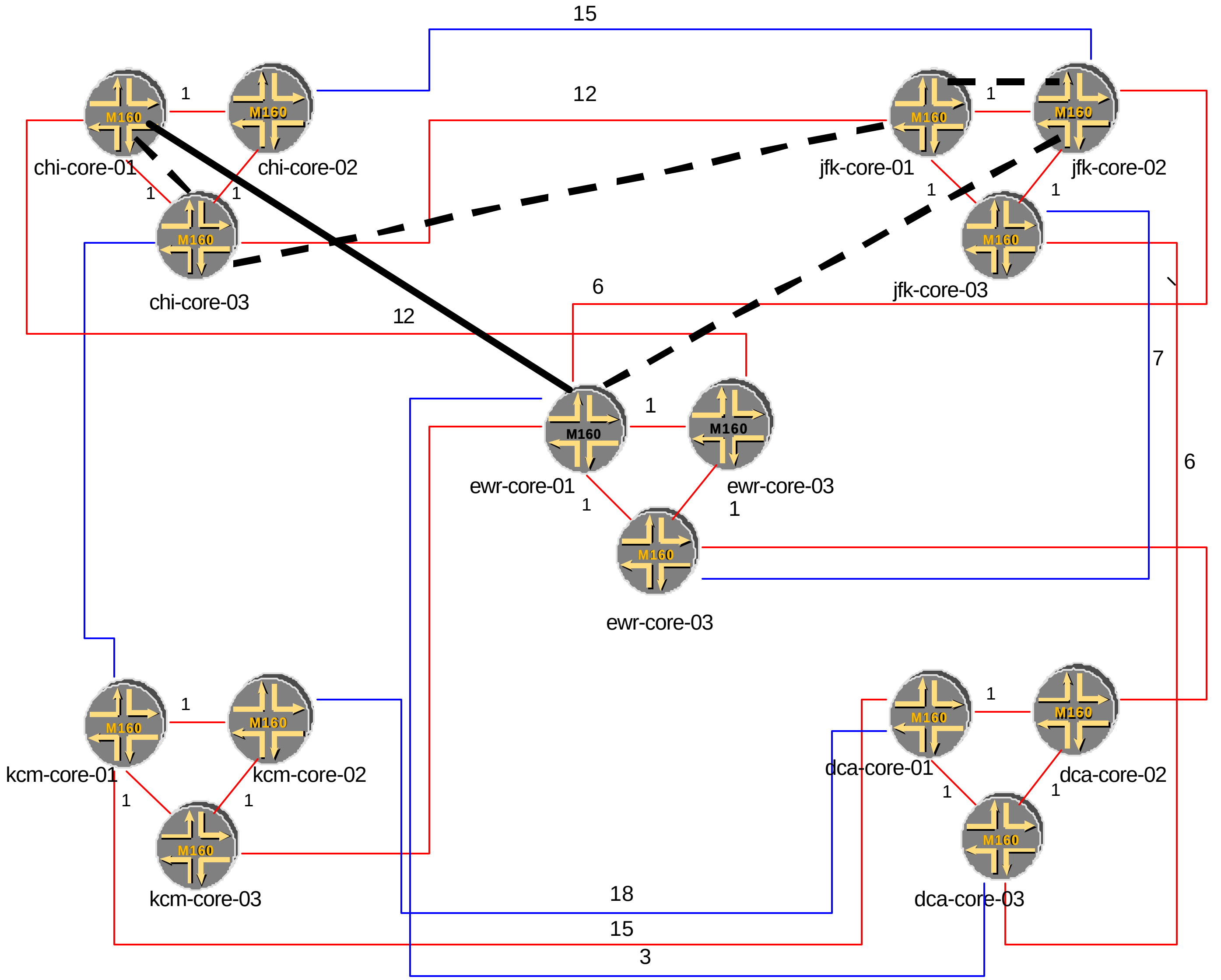
Both Primary and Detour LSPs are riding on same fiber path

LSP Configuration

Primary LSP: chi-core-01 to ewr-core-01

ERO: chi-core-01, ewr-core-02, ewr-core-01

Detour LSP ERO: chi-core-01, chi-core-03, jfk-core-01, jfk-core-02, ewr-core-01



MPLS FRR Topology
LSP: chi-core-01 to ewr-core-01

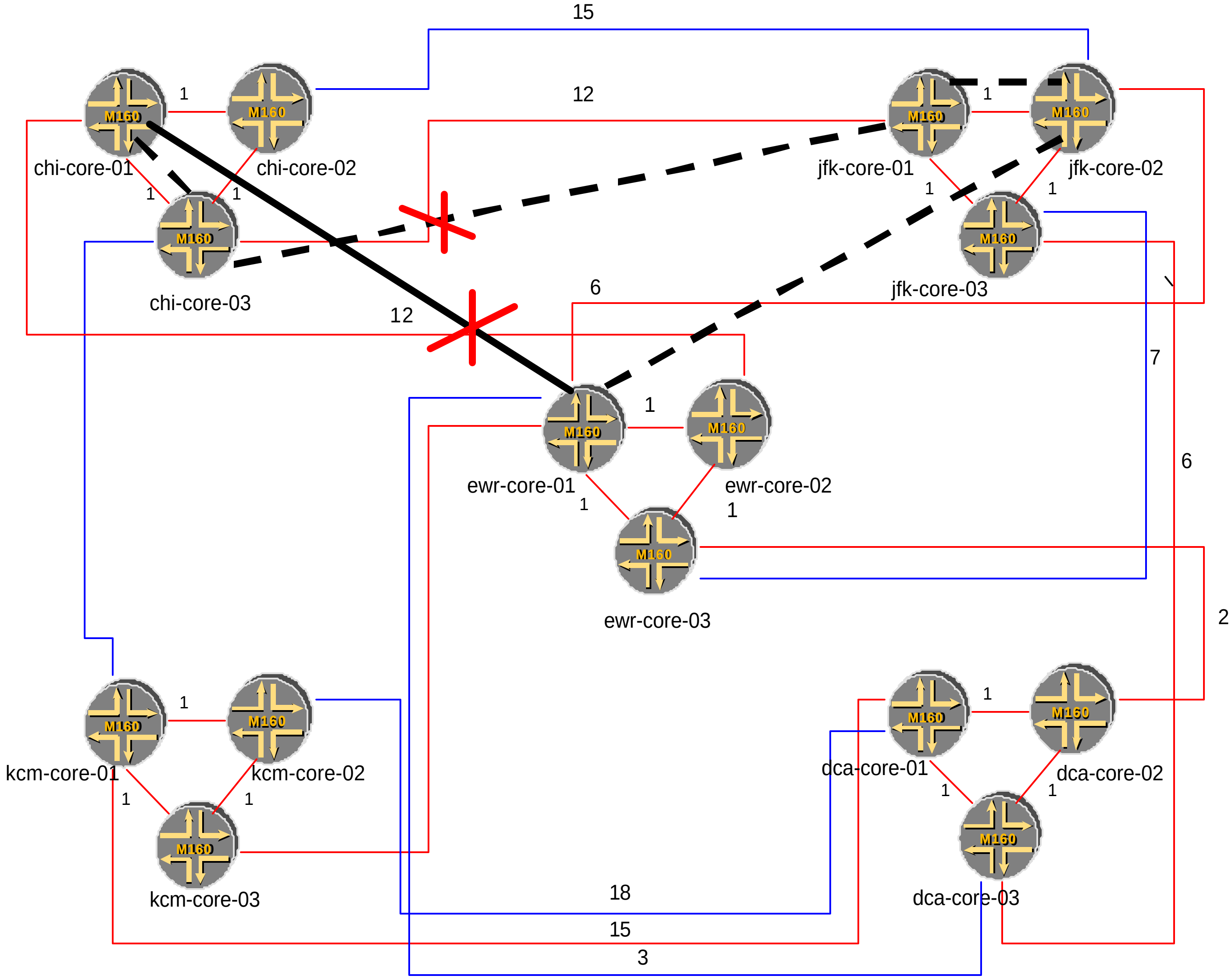
A BIG Fiber cut !

Both Primary and Detour LSP goes down

MPLS fails to do the FRR.

Routing falls back to traditional IGP

Fate Sharing ?



MPLS FRR Topology
Unstable router

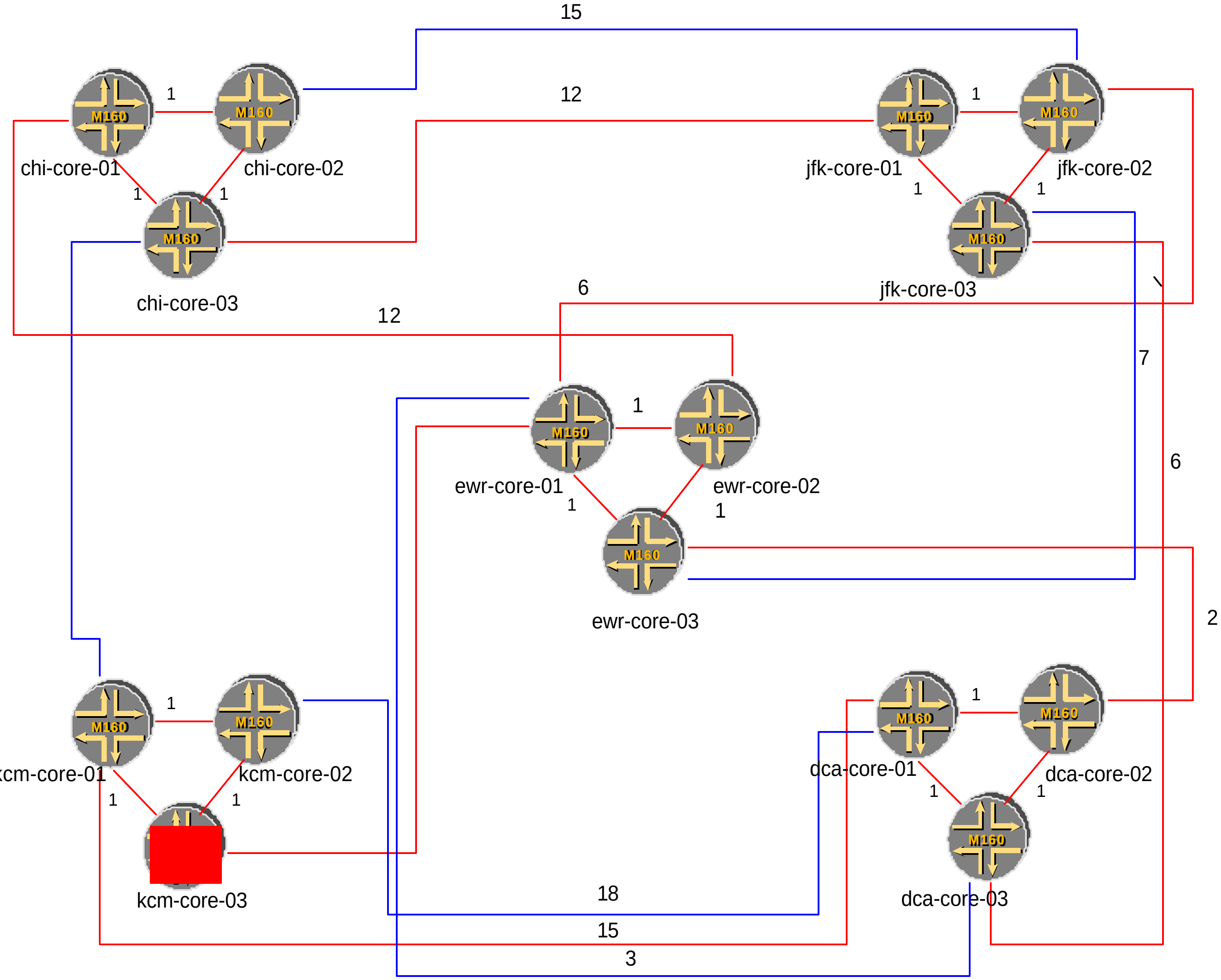
Unstable Router, kcm-core-03

All Links are constantly flapping in kcm-core-03

Both ISIS and IBGP flapping as well

MPLS comes to rescue. Traffic traversing over this router is forwarded over the detour paths avoiding kcm-core-03 node all together until router
Becomes stable.

Optimizer timer comes into play



Troubleshooting Tips

- Show mpls lsp name <lsp name>extensive
- Show rsvp neighbor
- Show rsvp session name <lsp name> detail
- Show rsvp interface detail
- Show mpls interface
- Show log <mpls-tracefile>
- Show log <rsvp-tracefile>

Fast Reroute

Lessons Learned / Conclusions

Conclusions

- Sub-second protection for both control/data plane failures necessary
- FRR provides sub-second recovery from data plane failures today
  - ISIS convergence can be improved, but best times (today) are in order of multiple seconds
  - FRR works, but
    - Requires implementation of a new technology
    - Lacks widely-deployed interoperable implementations
    - Can use enhancements such as detection of data plane "liveness", SRLG, QoS/TE conformance etc.
    - Manageability improvements (if doing lot of traffic management)

Conclusions

- HA mechanisms such as Graceful Restart / Stateful failover are interesting for control plane protection
- Keep it simple (as much as possible) - by relying on pure IGP metrics, no complex traffic management
- Use TE features only when needed – for example severe outage scenarios where real-time traffic must be protected (modeling required)
- IGP convergence timers must also be improved, since FRR protection is only on the core
- Tread carefully – control protocol scaling limits not completely known

Thank you!