PRESENTATION TITLE HERE

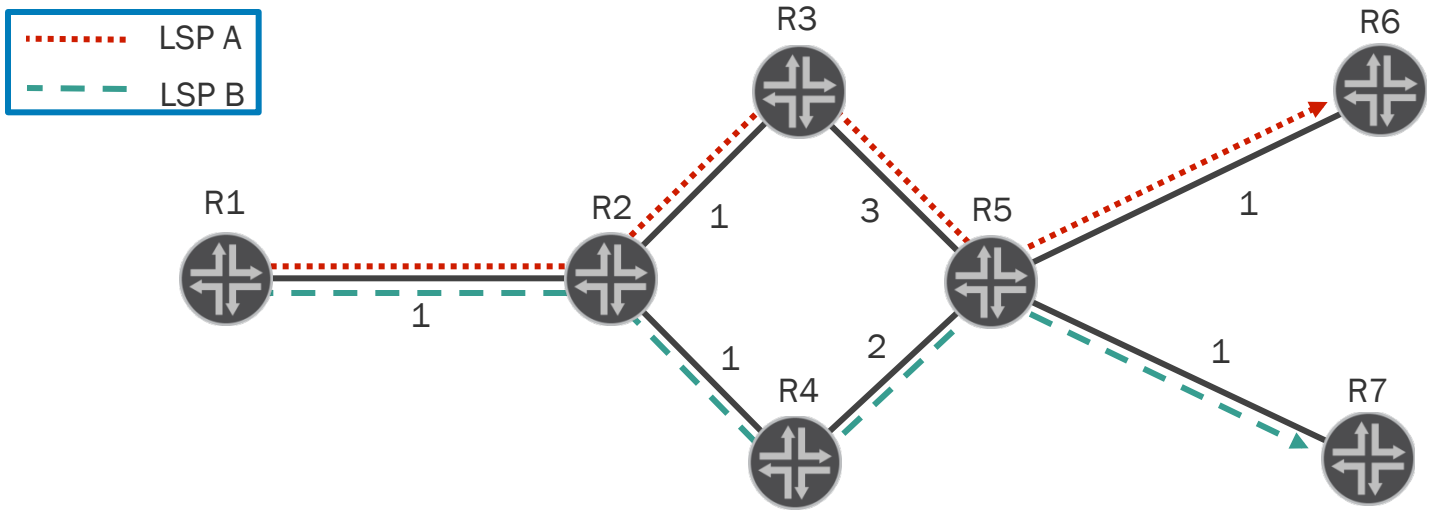
Presenter Name

Date

BENEFITS OF MPLS

Some benefits of MPLS include:

- Improved route lookup time by using labels to forward traffic
- Additional control over how traffic moves through the network using traffic engineering
- High Availability
- Increased scalability



HIGH AVAILABILITY

Ultimate goal – avoid business impact.

Business = working service – avoid service disruption or degradation.

Service disruption:

- Loss for connectivity = continues packet loss over period of time T
  - 1GE = 512B with 880Mp/h; 1% loss is 8Mp/h; is continuous - ~38 sec.
- Delay, jitter, random packet loss (non-continuous)
  - 1GE = 512B with 880Mp/h; 1% loss is 8Mp/h; is equally distributed – multiple breaks 4us each.

Challenges:

- Failures – loss of connectivity
- Congestions
  - Avoid: capacity planning and topology/link engineering, MPLS LSP path engineering (non-SPF routes).
  - Manage when happens: QoS tooklit.

MPLS LSP PATH ENGINEERING (NON-SPF ROUTES)

Technology is there

Very really used to fine tune path in order to optimize capacity utilization:

- OPEX intensive
- Multiple possible good states – no network banchmark
- Dynamic or not adaptive to changing traffic pattern
- Capacity is cheap, at least in Europe

Is used to control load balancing, SLRG, and as supportibe mechanism for other MPLS applications.

LABEL DISTRIBUTION PROTOCOLS

Overview of Label Distribution Protocols

- Often referred to as signaling protocols
- Dynamically establishes a LSP
  - Exchanges label information
- Examples of label distribution protocols.
  - Resource Reservation Protocol (RSVP)
  - Label Distribution Protocol (LDP)
  - Border Gateway Protocols (BGP)

HOW MPLS CAN HELP IN DEALING WITH FAILURES

Baseline

- Converged state = the optimal, stable state of routing as per routing policies and protocols algorithm.
- MPLS can't converge faster then IGP.
- LDP – after IGP converge (after last SPF run), active labels are selected from LIB.
- RSVP –
  - failure notification need to be delivered to head of each LSP (IGP flood, or PATH Err/ RESV Thear.) It take time and depend on network delay from point of failure.
  - Head end re-signal each LSP. It take time and depend on network diameter/ delay to and from tail of LSP (round-trip).

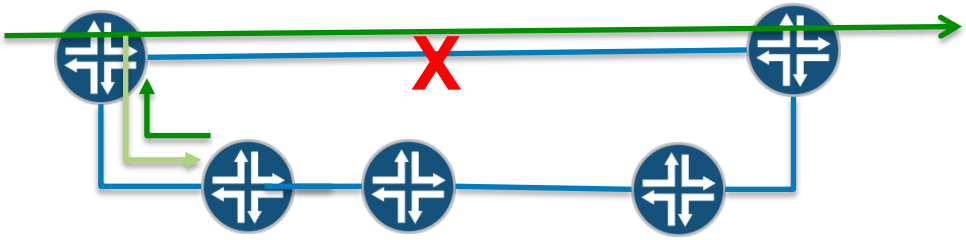
Pre-compute alternative path

- Local quick and dirty fix to keep packet delivery.
- Not converged state of network.
- LDP – follow IGP – IGP LFA (IPFRR)
- RSVP

LOOP FREE ALTERNATES AND LDP

LFA heavily depend on network topology and IGP costs

- Triangles are always good
- Squares and bigger makes problem



Requires RSVP-TE backup lsp to fix topology problem – escalate complexity.

Packet loss limited below 100ms in case of failure

Packet loss may appear at the end of convergence period of network – each node make SPF run ao it's own time.

Packet loss may appear what fault is fixed, because it is a topology change which triggers IGP convergence.

RSVP RECAP

PROPERTIES OF LABEL DISTRIBUTION PROTOCOLS

RSVP

- Is used for traffic engineering
- Internet standard for reserving resources
- Extended to support:
  - Explicit path configuration
  - Path numbering
  - Route recording
- Provides keepalive status for:
  - Visibility
  - Redundancy

BGP

- SCALABILITY !
- Inherit BGP path selection, NLRI attributes and policy
- Does not support engineered paths

RESOURCE RESERVATION PROTOCOL (RSVP)

RSVP:

- A generic quality of service (QoS) signaling protocol
- An Internet control protocol—uses IP as its network layer
- Designed originally for host-to-host usage
- Allows for bandwidth reservation
- *Not* a data transport protocol
- *Not* a routing protocol
  - Uses the IGP to determine paths

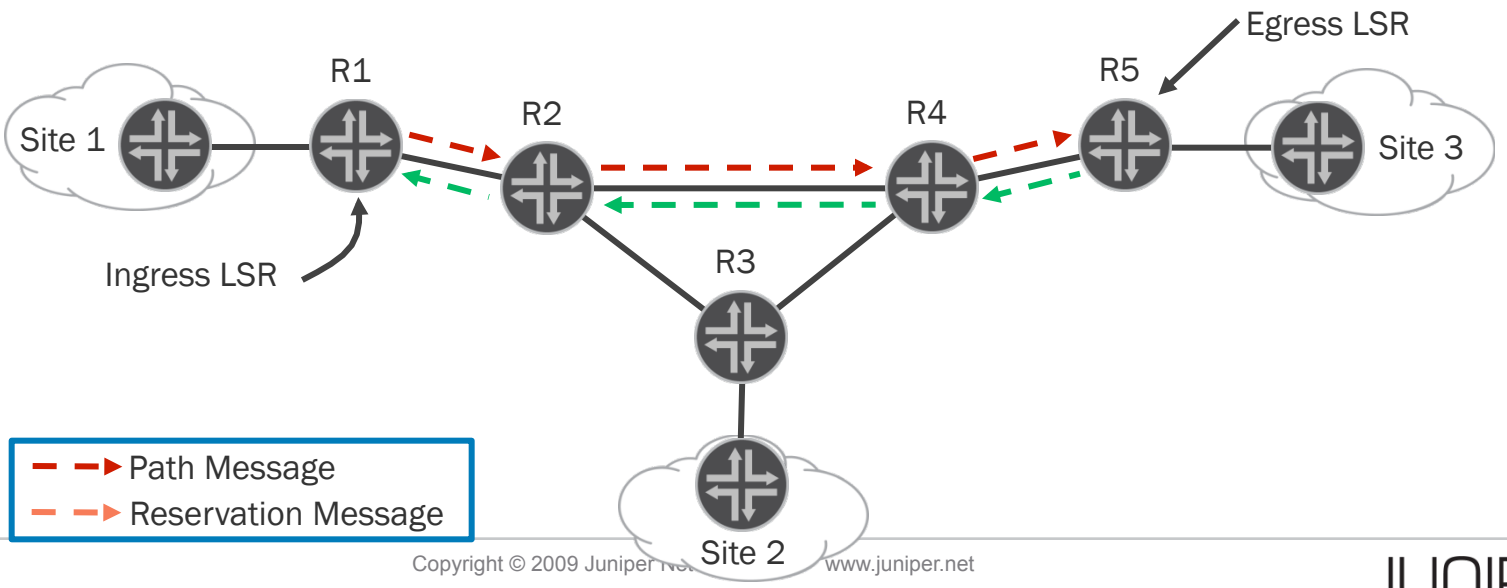
RSVP SESSIONS

Unidirectional data flows

- Ingress router initiates connection

Soft state

- Path and resources are maintained dynamically
- Can change during the life of the RSVP session
- Path message are sent downstream
- Reservation message are sent upstream



EXTENDED RSVP

Extensions added to support establishment and maintenance of LSPs

- Hello protocol
- Label distribution

RSVP is now used for router-to-router connectivity

TRAFFIC ENGINEERING EXTENSIONS: PATH

Path message extensions

- Mandatory:
  - SESSION Object: LSP_TUNNEL_IPv4
  - LABEL_REQUEST Object: Request LSRs to provide a label binding
- Optional:
  - EXPLICIT_ROUTE object (ERO): Specify predetermined path, independent of IGP path
    - Strict Hop: Informs the network of exact path to use
    - Loose Hop: LSP must be transited in order. IGP is used
  - RECORD_ROUTE object (RRO): Listing of the LSRs that the LSP tunnel traverses
  - SESSION_ATTRIBUTE object: Aids in session identification and also controls path setup priority, holding priority, and local-rerouting features
  - RSVP-HOP object: Contains the previous hop IP address

TRAFFIC ENGINEERING EXTENSIONS: RESV

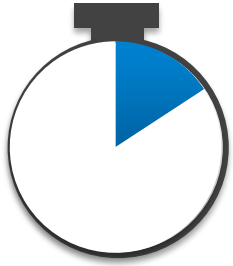
- Resv message extensions
  - Mandatory:
    - SESSION object: uniquely identifies the LSP being established
    - LABEL object: performs the upstream on demand label distribution process
    - STYLE object: specifies the reservation style (fixed-filter, wildcard-filter and shared-explicit)
  - Optional:
    - RECORD_ROUTE object: returns the LSPs path to the sender of the path message
    - HOP object: contains the previous hop IP address

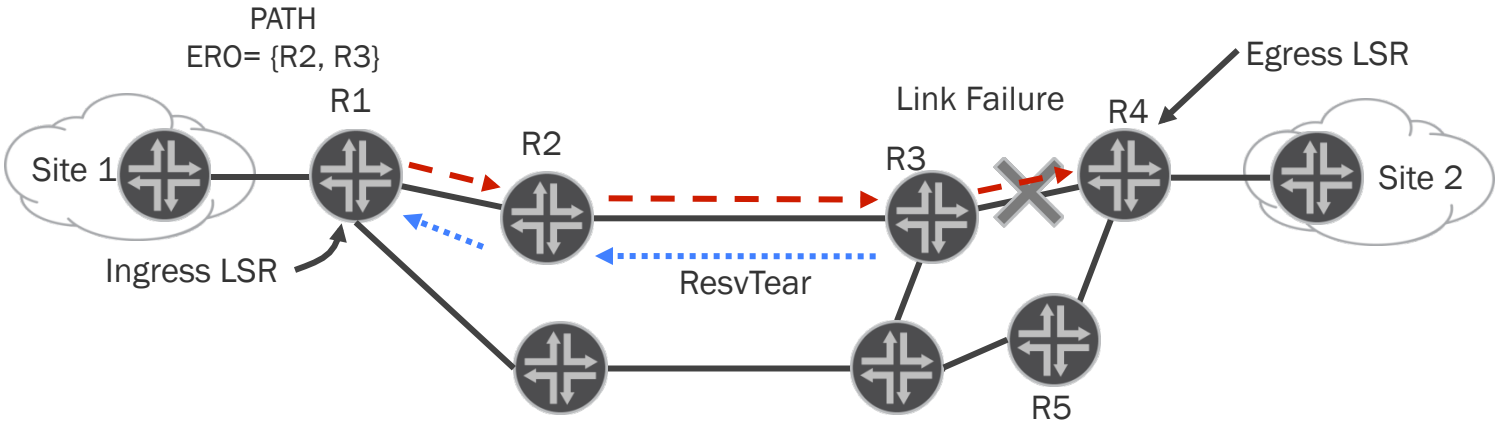
PROTECTION OF RSVP-TE LSP

TIME TO RECOVERY—NO PROTECTION (1)

R3 determines that there is a link failure between R3 and R4

- Packets traversing the LSP begin dropping
- R3 sends a ResvTear upstream towards the ingress router as notification of the failure

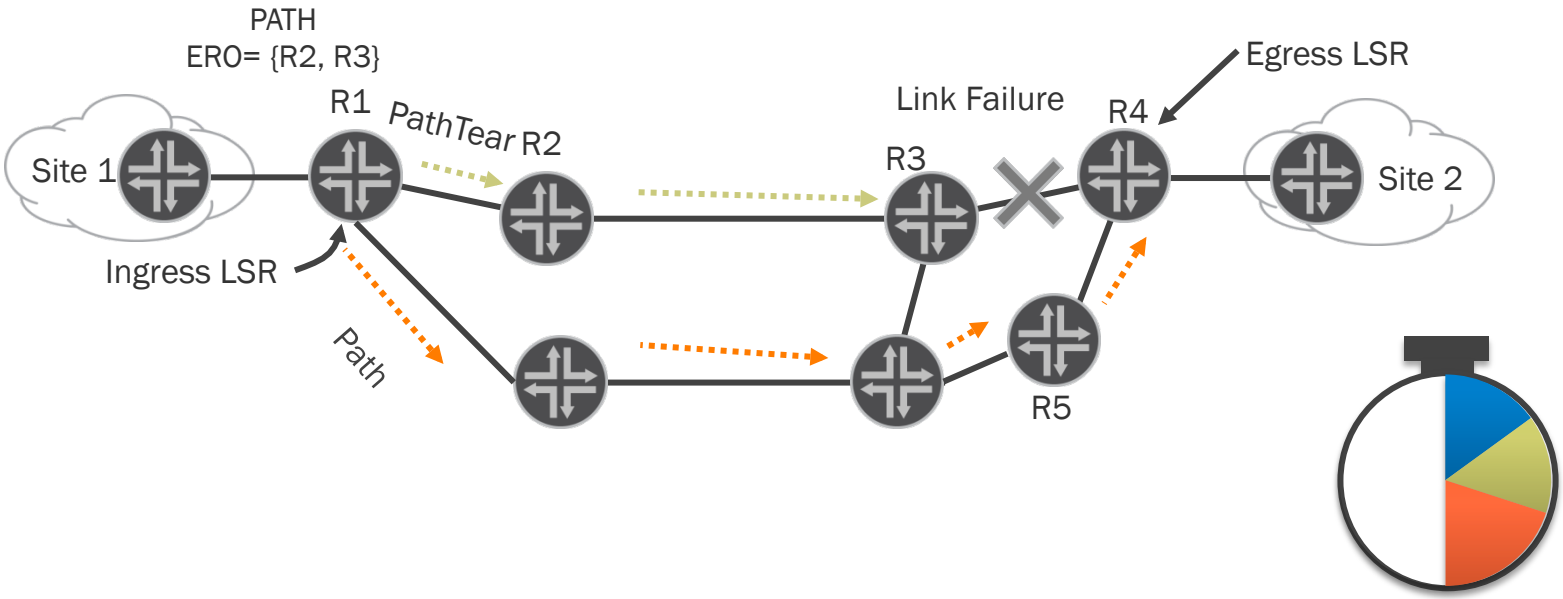




TIME TO RECOVERY—NO PROTECTION (2)

R1 reacts to the reception of a ResvTear for the LSP

- Path and Resv state blocks for LSP are removed
- R1 attempts to build a new LSP by sending a path message downstream
- Packets continue to drop

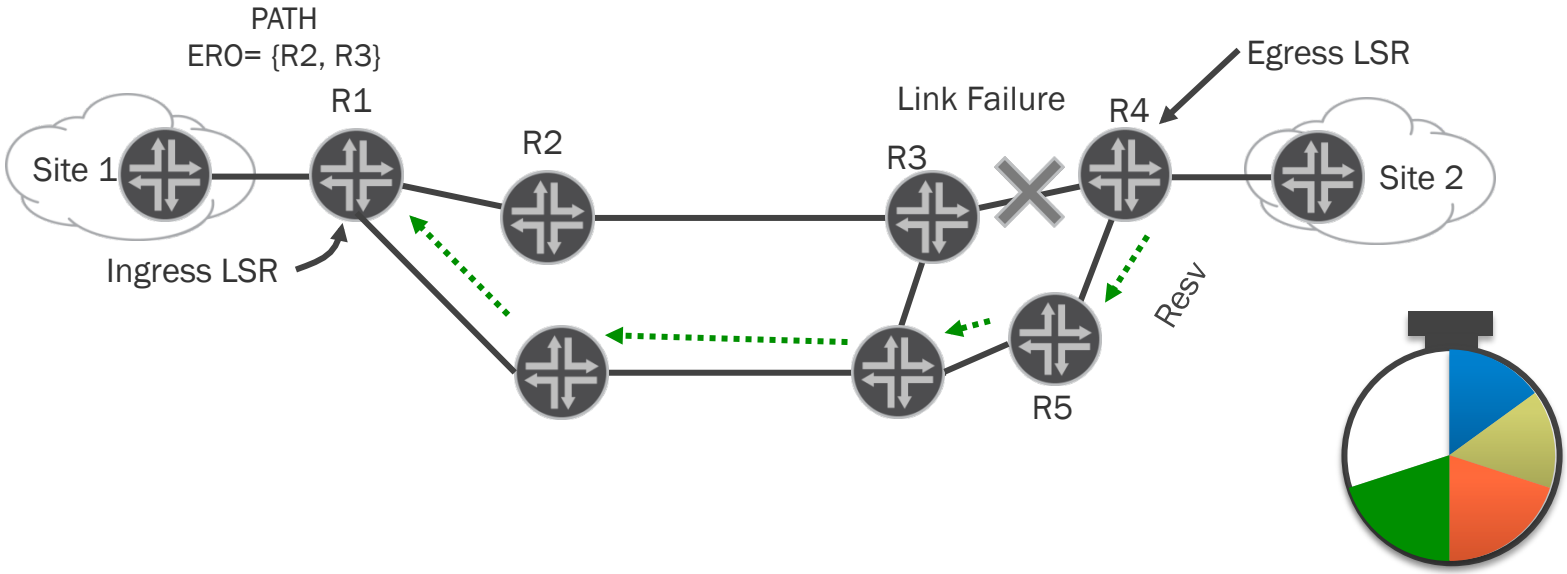


TIME TO RECOVERY—NO PROTECTION (3)

R4 reacts to the reception of a new Path for the LSP

- Answers by Resv message
- Packets continue to drop

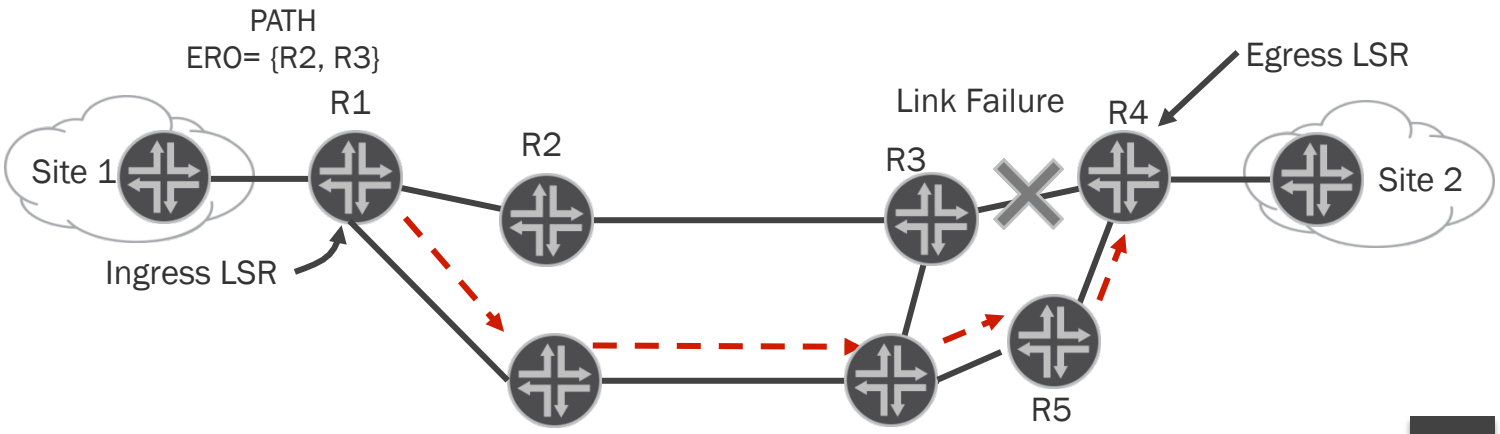
R1 receives Resv message, new LSP is established

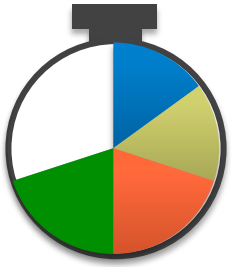


TIME TO RECOVERY—NO PROTECTION (4)

A new LSP is established around the failed link

- Packets are no longer dropped





PRIMARY AND SECONDARY LSP

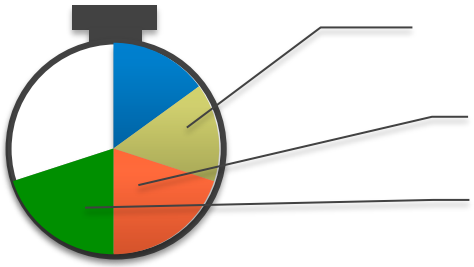
End-to-End protection

Secondary LSP takes 2nd best best form ingress to egress.

Secondary LSP can have different constrains (afinity group, BW, ERO) then primary.

Secondary LSP can be:

- Computed after failure – previous example, no CSPF.
- Pre-computed on ingress LSR, but not signaled – previous example w/ CSPF.
- Pre-computed and pre-signaled
  - Preestablishes and maintains secondary path
  - Eliminates LSP signaling delays when active path fails
  - Additional state information must be maintained – eacj LSP has 2 RSVP sessions.



Make this delay not impacting traffic switchover

Avoid this delay

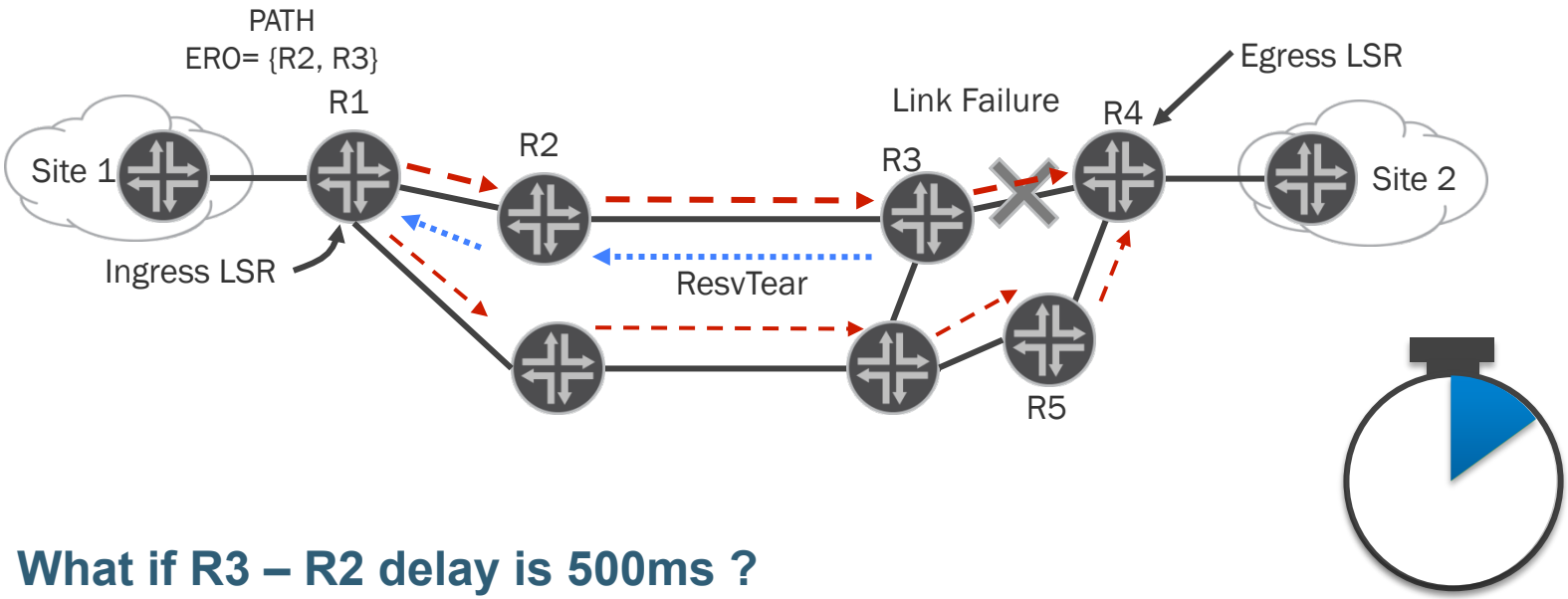
Avoid this delay

TIME TO RECOVERY— SECONDARY STANDBY

R3 determines that there is a link failure between R3 and R4

- Packets traversing the LSP begin dropping
- R3 sends a ResvTear upstream towards the ingress router as notification of the failure

R1 switch traffic to secondary LSP



What if R3 – R2 delay is 500ms ?

MOTIVATIONS FOR FAST REROUTE

Ask yourself these questions:

- Is there a way to get quicker failover in the event of primary LSP failure?
- How can I reduce packet loss when I lose my primary LSP?

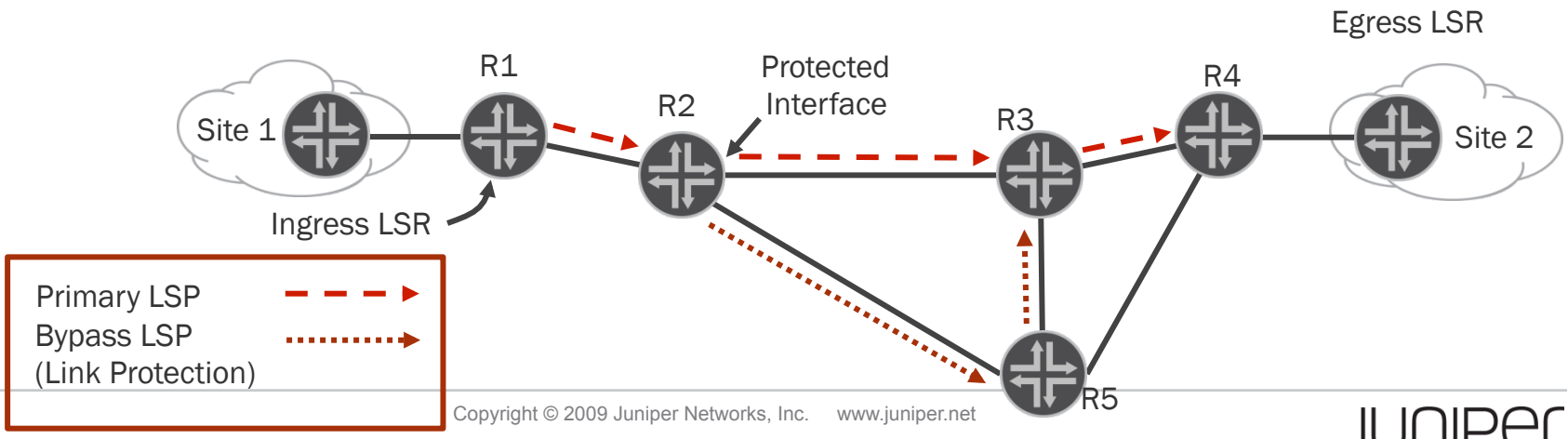
FAST REROUTE CHARACTERISTICS (1)

Short-term solution to reduce packet loss

- Implement the one-to-one backup method defined in RFC 4090 (use LSP backup)
- Implement the many-to-one backup method defined in RFC 4090 (use facility backup)
- When node or link fails, upstream node:
  - Immediately switch traffic to detour/bypass LSP
  - Signals failure to ingress LSR

Ingress LSP will re-signal LSP in order to get best possible path end-to-end

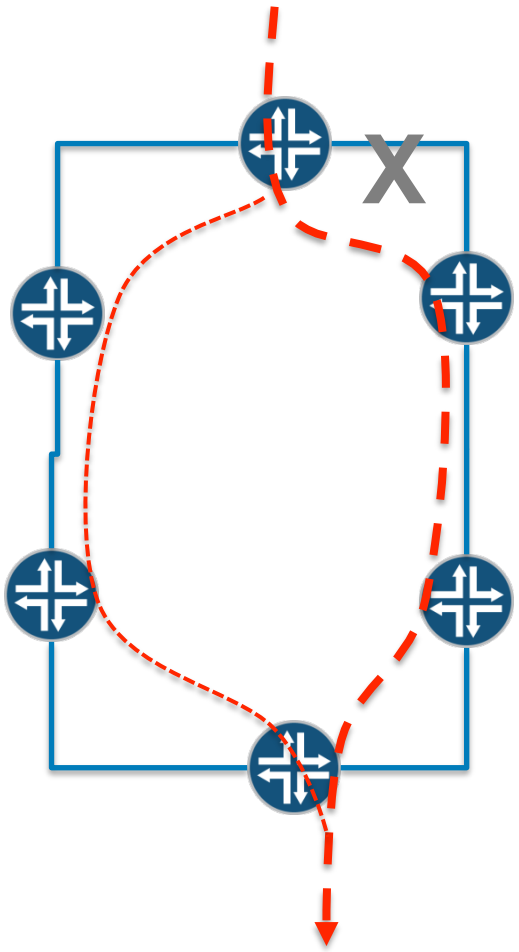
R1 is Point of Local Repair (PLR); R3 is Merging Point (MP)



ONE TO ONE FRR CHARACTERISTIC

Implement the one-to-one backup method defined in RFC 4090 (use LSP backup)

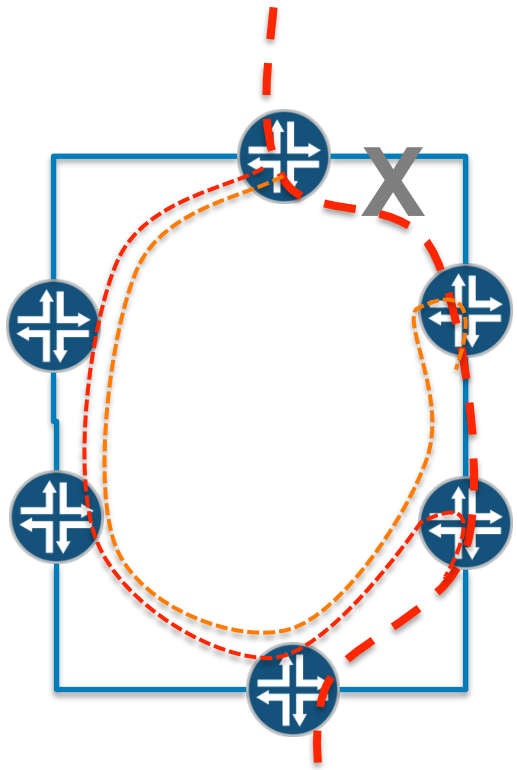
- Scalability issue – O(N*M) where
  - N is # of protected nodes
  - M is # of LSP
- Detour can inherit all constrains of primary LSP
- Ring friendly
- Label swap
- Use best path form PLR to egress LSP
- MP can be any node – first common node for detour and primary LSP



FACILITY BACKUP CHARACTERISTICS

Implement the many-to-one backup method defined in RFC 4090 (use facility backup)

- Good Scalability– O(N) where N is either
  - # of protected interfaces (link protection)
  - # of nodes behind protected node (node-link protection)
- Tunnels primary LSP over BYPASS LSP around failure - Label swap'n'push
- MP can be either
  - Node connected by faulty link (link protection)
  - Node immediately behind faulty node (node-link protection)
- Use best path form PLR to MP, but not egress LSR.



COMMON PRACTICE

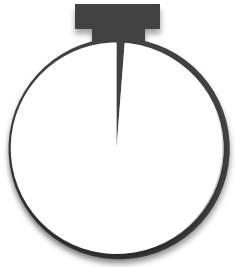
In most cases facility backup is in use

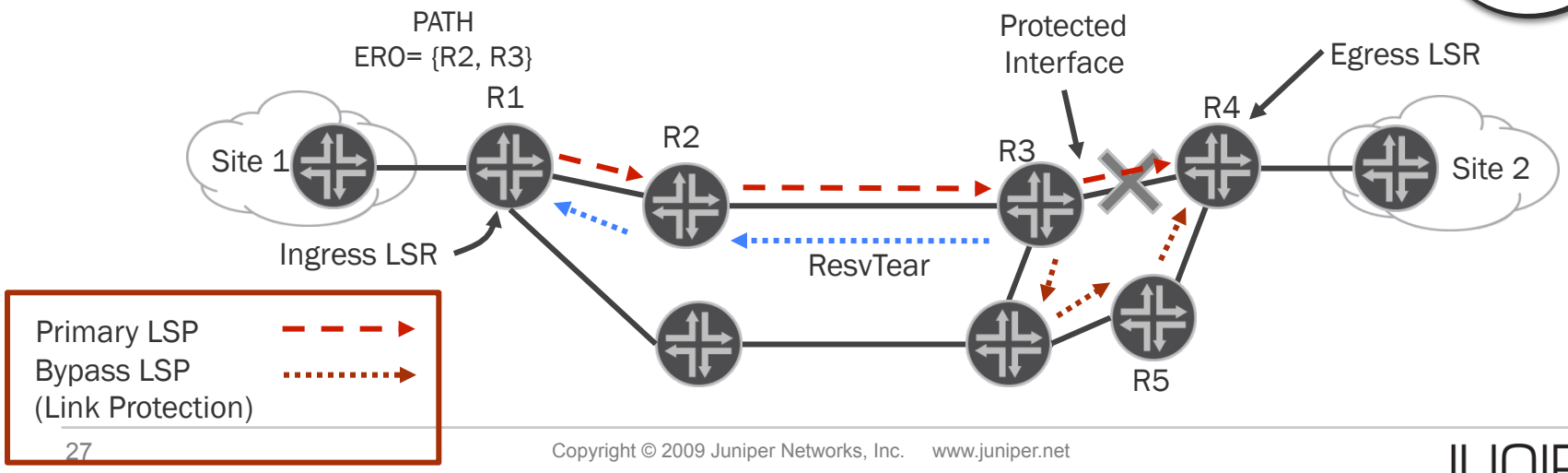
Mostly because of scalability.

LINK PROTECTION OVERVIEW

Protects interfaces instead of entire LSP path

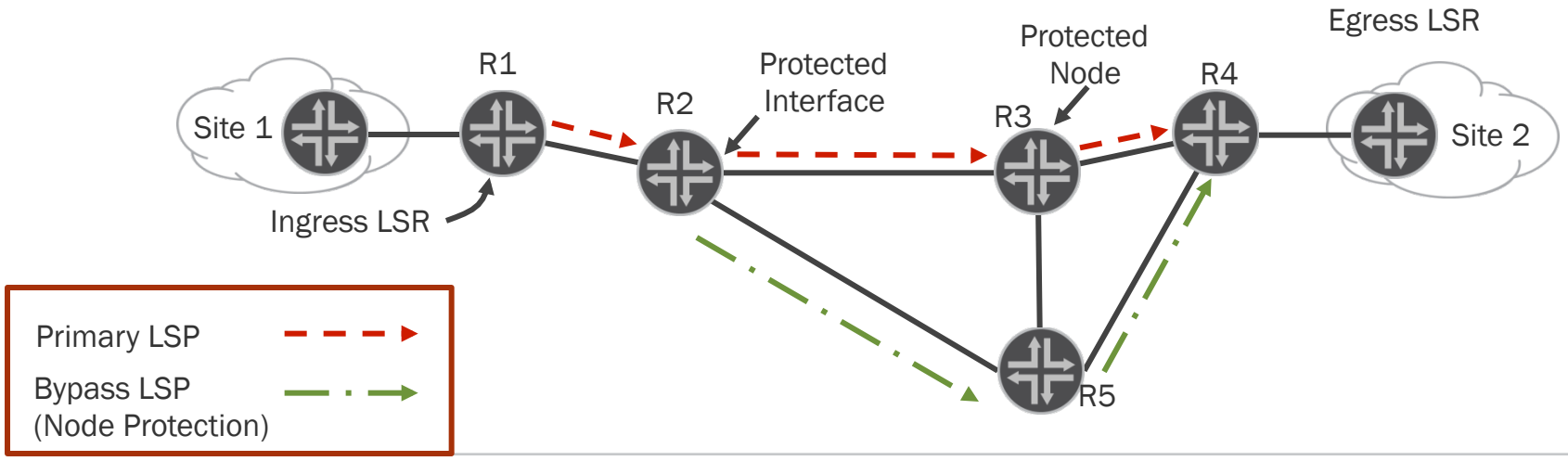
- Implements the facility backup method defined in RFC 4090
- LSPs must be flagged to make use of a bypass LSP
- Bypass LSP established around protected interface to adjacent node
  - Established BEFORE any failure, in normal state of network.
  - Uses CSPF to calculate bypass LSP
  - Can add ERO to influence CSPF routing of bypass LSP





NODE PROTECTION OVERVIEW

- Protects against failure of downstream node
  - Uses similar mechanisms to link protection
  - Relies on RSVP hello timers to determine node failure
  - LSPs must be flagged to make use of a bypass LSP
  - One bypass LSP established around downstream node



FACILITY BACKUP AND RING

Some capacity is wasted, as traffic U-turns

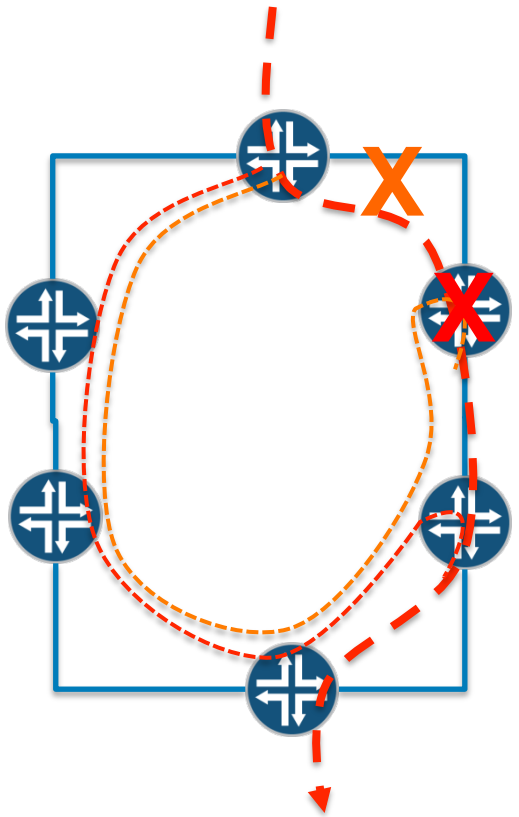
This is only short-term, until network converge, and lead-end re-signal LSP.

But what if primary LSP has constrain which do not allow for re-signaling?

- Affinity group
- Strict/loose hop of faulty link/ node address

Use secondary path with other constrains, or

Traffic stay on bypass until failure get fixed.



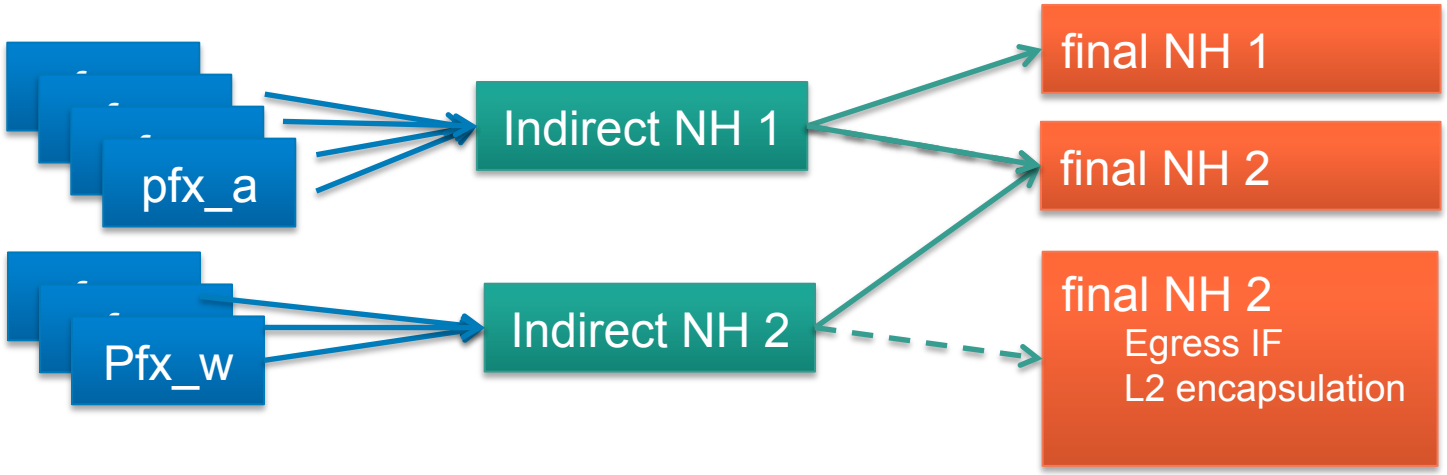
WHY FRR IS SO QUICK AND HOW QUICK?

No need to any routing computation. No control plane involved.

For each destination (yes even 100's of 1000's form BGP) backup egress interface and encapsulation descriptor are programmed in FIB.

After failure is discovered, ASIC need to change pointer from active be backup descriptor.

- ASIC FIB data structure is crucial here – indirection!



HOW TO FIND PATH FOR BYPASS

Requirements

- Best path from PLR to MP – both are on primary e2e LSP
- Best path is via faulty link/node
- BYPASS can’t be routed base on shortest path !

BYPASS need to be signaled with ERO to overcome protected link/node

- Manually configured – not scalable
- Computed by CSPF – to calculate BYPASS ERO, PLR use TED, “removes” protected link/node form it and runs SPF.
- Other constrains (affinity group, bandwitch) can be put on BYPASS ERO calculation

CONSTRAINT-BASED ROUTING OVERVIEW

Modified shortest-path-first algorithm

Integrates TED data

- IGP topology information, available bandwidth, and Administrative group
- Determines optimal path and setup order according to user-provided constraints
  - Maximum hop count (for fast reroute detours)
  - Bandwidth
  - Strict or loose routing (EROs)
  - Administrative groups
  - Priority

Prunes nonqualifying paths and performs SPF on remaining routes

- The result is either an ERO that is handed to RSVP for signaling, or a *no route to host* error message

TOPOLOGY INFORMATION DISTRIBUTION

IGP extensions propagate additional information

- IS-IS uses TLV tuples
- OSPF uses opaque LSA Type 10
- Information propagated within area or level only

Information propagated:

- Bandwidth available
- Administrative Groups (link colors)
- Router ID

IGP EXTENSION EXAMPLE: OSPF

```
user@R1> show ospf database opaque-area detail

    OSPF database, Area 0.0.0.0
OpaqArea*1.0.0.3          192.168.2.1      0x80000002     4   0x22 0xe2c
124
  Area-opaque TE LSA
  Link (2), length 100:
    Linktype (1), length 1:
      2
    LinkID (2), length 4:
      172.22.220.2
    LocIfAdr (3), length 4:
      172.22.220.1
    RemIfAdr (4), length 4:
      0.0.0.0
    TEMetric (5), length 4:
      1
    MaxBW (6), length 4:
      1000Mbps
    MaxRsvBW (7), length 4:
      1000Mbps
    UnRsvBW (8), length 32:
        Priority 0, 1000Mbps
        Priority 1, 1000Mbps
        Priority 2, 1000Mbps
        Priority 3, 1000Mbps
        Priority 4, 1000Mbps
        Priority 5, 1000Mbps
        Priority 6, 1000Mbps
        Priority 7, 1000Mbps
    Color (9), length 4:
      0
```

THE CSPF ALGORITHM

For BYPASS protecting given link/node:

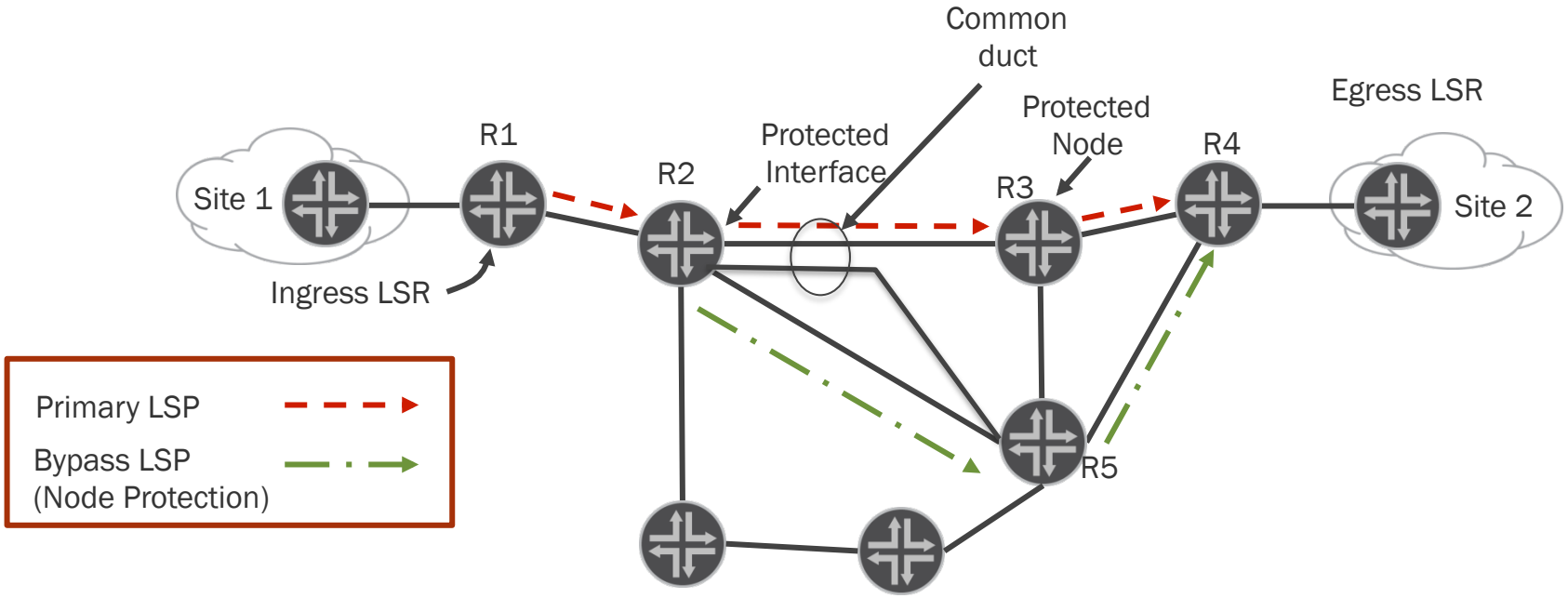
1. Prune protected link/links of protected node
2. Prune links with insufficient bandwidth (if BYPASS has BW constrain)
3. Prune links that do not contain an included color (if BYPASS has affinity group constrain)
4. Prune links that contain an excluded color (if BYPASS has affinity group constrain)
5. Calculate shortest path from ingress to egress consistent with ERO
6. If equal-cost paths exist, choose the path whose last hop address equals the LSP's destination
7. Select among equal-cost paths (least hop, then fill related criteria)
8. Pass explicit route (ERO) to RSVP

BYPASS AND SHARED UNDERPLAYING RESOURCE

CSPF sees only L3 topology – R2-R3 and R2-R5 are independent

But they may share same FO, DWDM, or duct/pipe

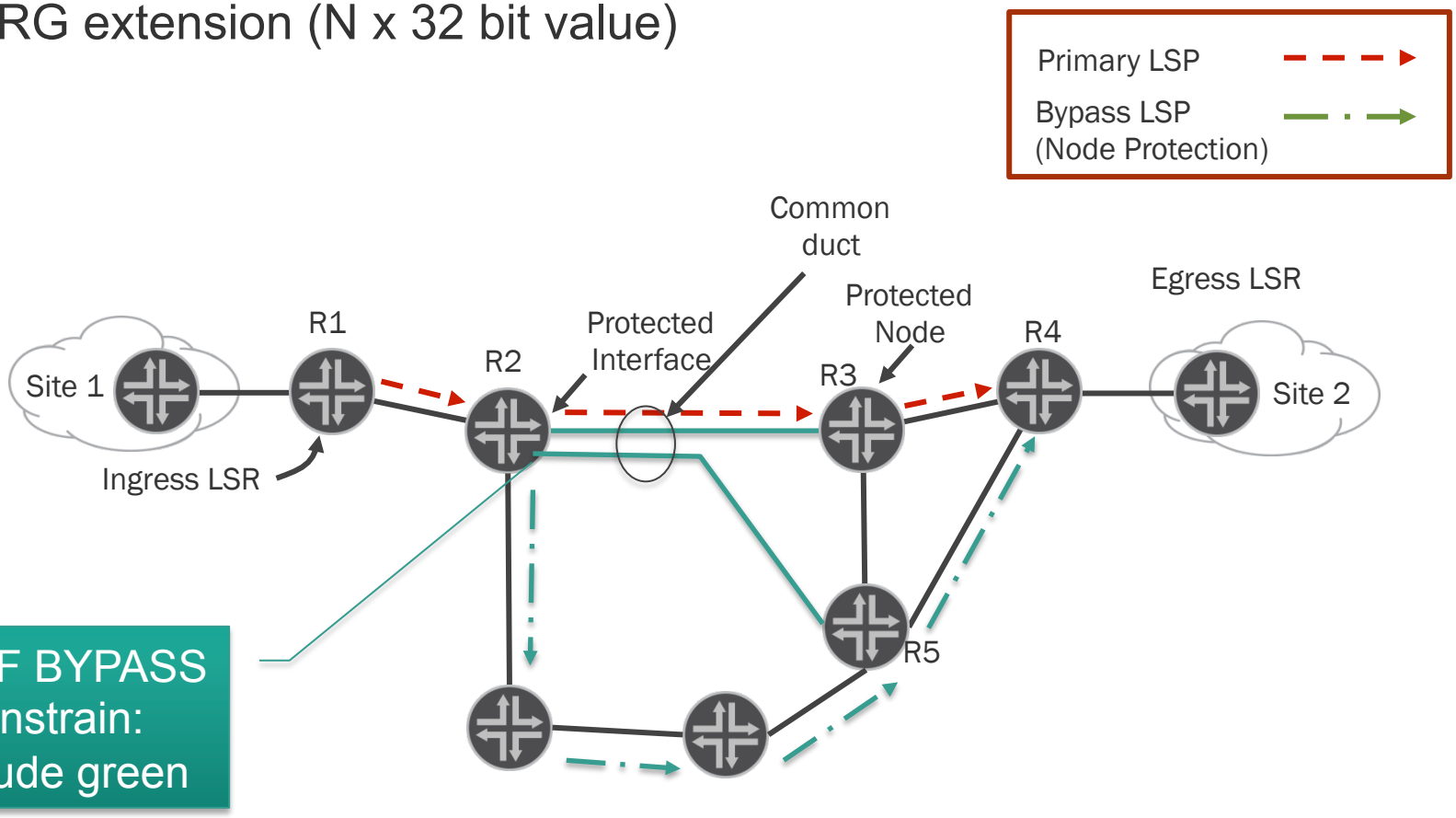
Need to feed CSPF with this information



SLRG INFORMATION FOR CSPF

Mark links sharing risk by same label

- Affinity group (only single 32 vector)
- SLRG extension (N x 32 bit value)



FAILURE EVENT SEQUENCE

When failure happen:

- We loose ~ 50 ms of traffic
- MPLS FRR locally patch the failure – traffic restored
- E2E LSP ingress router notified
- Ingress router re-signal new e2e LSP optimal for current topology – traffic still use old LSP and BYPASS
- Ingress router move traffic onto new LSP
  - Traffic use new e2e LSP
  - Old LSP exist and use BYPASS, but no traffic on it.
- After while, ingress router tear down old LSP

Switching form old LSP to now LSP is LOSS-LESS

- Thanks to Make-Before-Break

FAILURE FIXED

Traffic is on LSP, which is not optimal

- It was optimal during failure
- Fixed link/node change topology – better path possible.

LSP need to be configured for optimization (not necessary default)

Sequence of events

- Ingress router is notified about new link by IGP
- Ingress router signal new e2e LSP optimal for current topology – traffic still use old LSP
- After some time, ingress router move traffic onto new LSP
  - Traffic use new e2e LSP
  - Old LSP exist, but no traffic on it.
- After while, ingress router tear down old LSP

Switching form old LSP to now LSP is LOSS-LESS

- This is very different to IGP/LDP/LFA convergency

FRR IMPACT ON NETWORK SCALING

Challenges

- E2E LSP full mesh among PEs
- Number of BYPASSES
- MBB – doubling number of LSP for some time

Network grow

- Telco considers Next-Gen Architecture with 10.000's of MPLS nodes.
- Simple LSP full mesh do not scale
- Even IGP/LDP will not scale

BGP can scale - use it for label distribution

- No TE or FRR capability
- BGP free core

DIVIDE AND CONQUERED

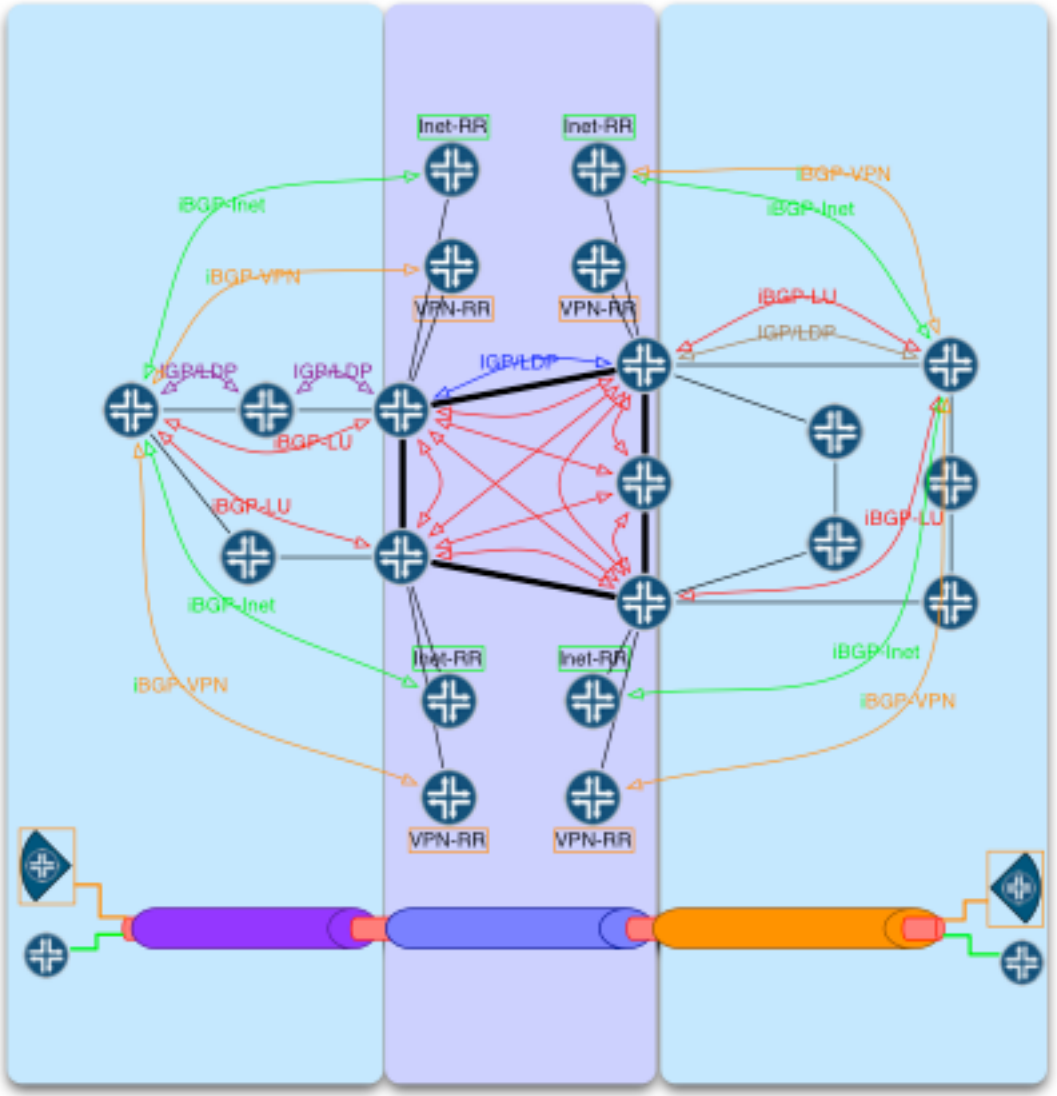
Seamless MPLS architecture

Regions runs independent IGP and LDP or RSVP.

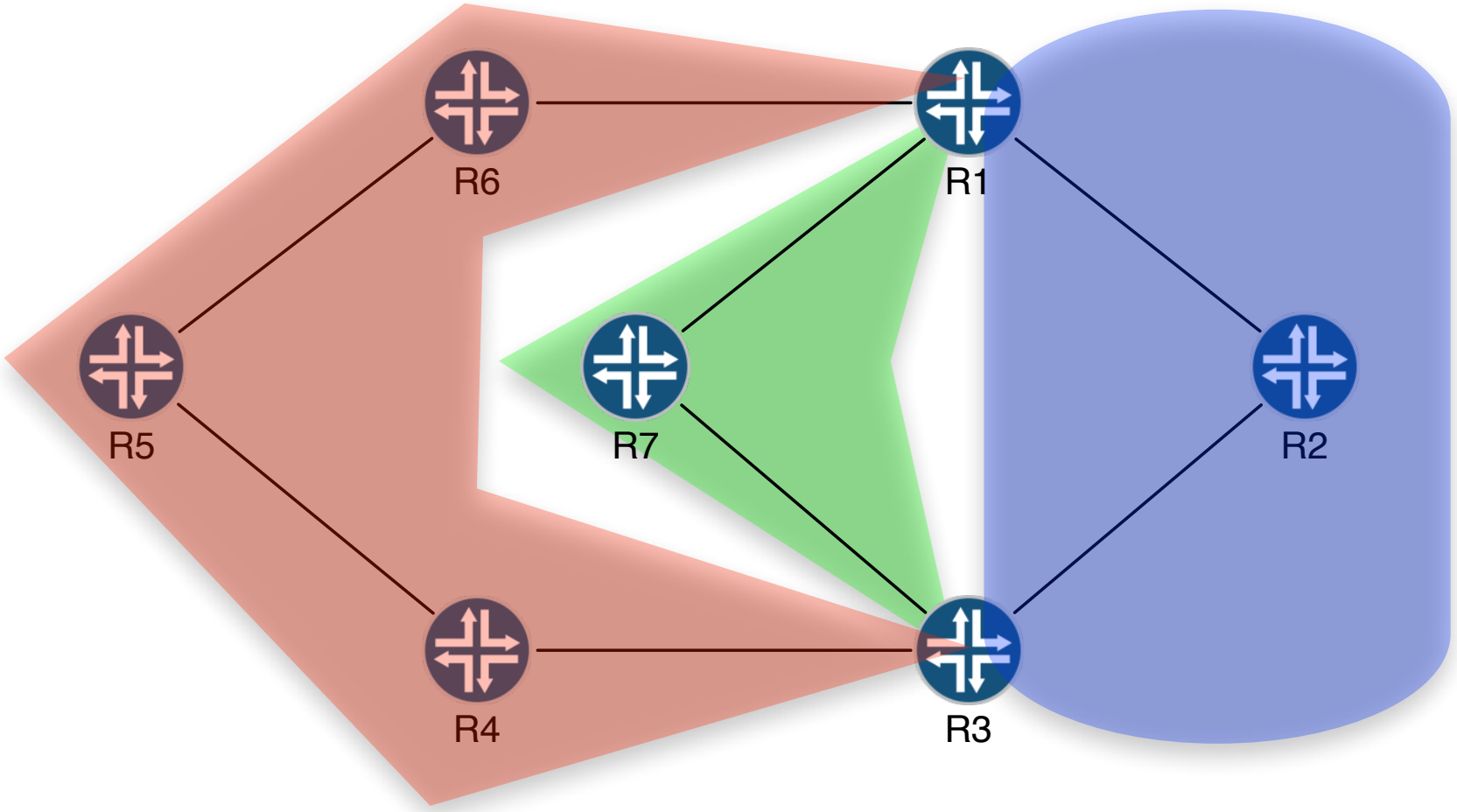
All PE in region runs iBGP-LU (RFC 3107)

Regions connected by either

- eBGP-LU
- RR
  - has IF in multiple regions
  - NHS for iBGP paths



SMALL EXAMPLE



JUNIPER ICONS

Generic Devices (cont.)

Blue versions of any generic icons represent general Juniper versions of each.


SBC Option 1


SBC Option 2

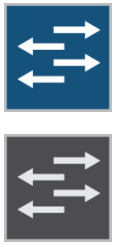

Secured Generic Router


Generic Router, GCSN

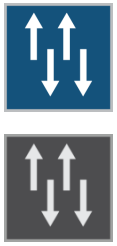

BSR-VSR


Logical

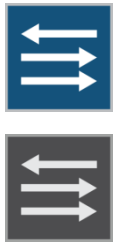

L2/L3 Switch


L2 Switch, L3 Switch


GGSN


SGSN


MSC


RNC, BSC


Lightning

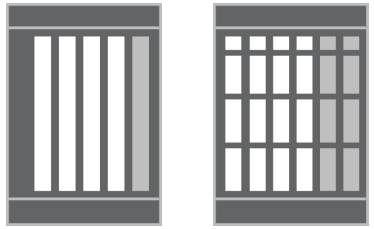

AirWaves


CPE Antenna


Generic Racks


1 Rack 1 Unit


1 Rack Multiple Units


Generic Core Routers

JUNIPER ICONS

Generic Devices (cont.)

Cable Head-End, Head-End, Edge QAM

OLT

Cable Modem Termination, M-CMTS

Switch, Gigabit Ethernet, MPLS, MSC, Class 5, Layer 2, RNC/BNC, SGSN, Ethernet LAN,

Multiservice Security, Routing Switching

Multiservice Security, VOIP, Routing Switching

Integrated

Media Gateway

Dial Access Aggregation, Optical Mix Multiplexer, ROADM Multiplexer

Optical OCX

ONT

SONET Switch, ATM Switch

Frame Relay Switch

Video BNG

Voice Gateway

Voice Home Gateway

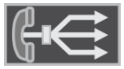
Voice Softswitch

VoIP Gateway

Wireless Access Point

DWDM Switch, WDM Switch

Continuous Systems



PBX

MSAN, Access Node, DSLAM

PSTN

JUNIPER ICONS

Location (cont.)

Network Cloud

Scenario/Location

Toolkits

NFP Web Services

Toolkit, VPN Toolkit, QoS Toolkit, Security Toolkit

Unlocked

Unlocked with Key

Locked with Key

everywhere