Fast Convergence

The Need for Fast Convergence

- Failure of a link/node results in disruption of network traffic
- These disruptions can last several seconds before the network reconverges
- Emergence of low latency applications such as IPTV, VoIP, business critical
- The distributed nature of the network places a limit on the minimum reconvergence time.
- How Fast ?
  - Sub Second: Achievable typical requirement for most IP networks
  - Sub 500ms: Required by Low Latency Applications like VoIP
  - Sub 50ms : Limited possibilities (achievable in a limited way using RSVP-TE traffic engineering)

Convergence Times Structure

- Detection of SDH/SONET layer failure
  - few milliseconds
- Report failure to router controller
  - Few milliseconds
- Change the LSAs affected by the failure and floods them
  - 10s of milliseconds
- Run SPF algorithm and computes the new OSPF shortest path tree
  - 10s of milliseconds
- Communicate and install new Next-Hops to line cards
  - **100s of milliseconds**

Loop-Free Alternate(LFA) or IP Fast Reroute (IP FRR) principles

- There's no reason Link State IGPs couldn't react faster
  - Every single router knows the whole topology of all attached areas and can thus easily calculate which of its neighbors could be feasible successors
- How?
  - OSPF or IS-IS routing process runs SPF, computes its own best paths, and installs them in the IP routing table (RIB)
  - After the network has converged, OSPF runs SPF algorithm from the perspective of its neighbors. If a neighbor's SPF tree doesn't use current router as the next hop for a specific destination, it's safe to use that neighbor as a feasible successor
  - The feasible successor information calculated by OSPF is downloaded in RIB and FIB, where it can be used immediately after the link failure

Fast Convergence not Always Possible
Loop Protection Rules

- Link loop protection
  - Dist (AlternSrc, Dest) < Dist (AlternSrc, Src) + Dist (Src,Dest)
- Node Loop Protection
  - Dist (AlternSrc, Dest) < Dist (AlternSrc, ProtNode) + Dist (ProtNode, Dest)
- Typical coverage without TE tunnels is around 65 – 85%

Examples:

A->B alternate path:

A-B link failure and A-B link cost = 10

A->B alternate path:

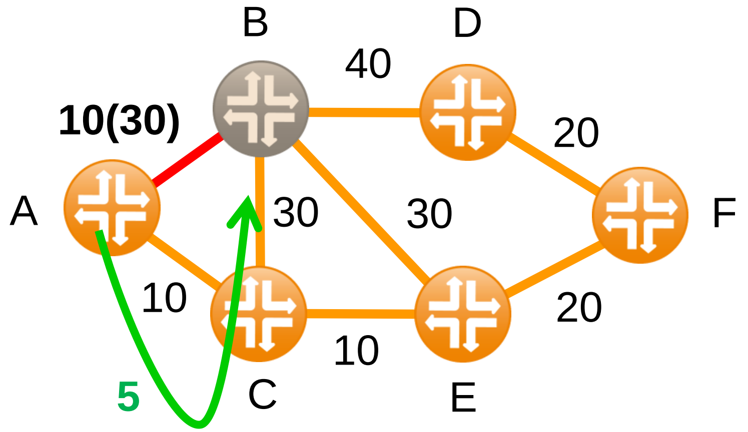
A-B link failure and A-B link cost = 30

A->D alternate path:

B node failure and A-B link cost = 10

A->D alternate path:

A-B link failure and A-B link cost = 10 & backup TE tunnel

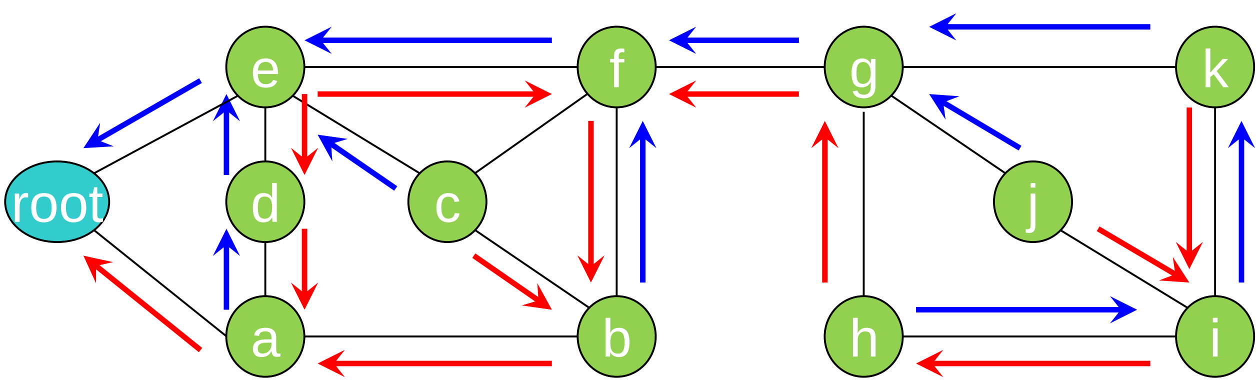


Maximally Redundant Trees (MRT)

- LFA/ IPFRR is topology-sensitive so generally lacks 100% protection for link & node failures
- Can augment LFA RSVP-TE tunnels
  - Manageability overhead

Maximally Redundant Trees

- A pair of directed spanning trees
- The common root is reachable along both of them
- The two paths along the two trees are maximally redundant



MRT Forwarding

- Fast-Reroute with MRT uses 3 forwarding topologies:
  - current default topology – next-hops computed by SPF
  - Blue MRT topology - MRTs computes next-hops
  - Red MRT - MRTs computes next-hops
- Each Router has 3 sets of next-hops (default, blue MRT, red MRT) to reach every other router.
- MPLS label distributed by LDP to indicate FEC and MT-ID (Multi Topology ID)
  - No label stacking needed
  - Just works with existing MPLS HW

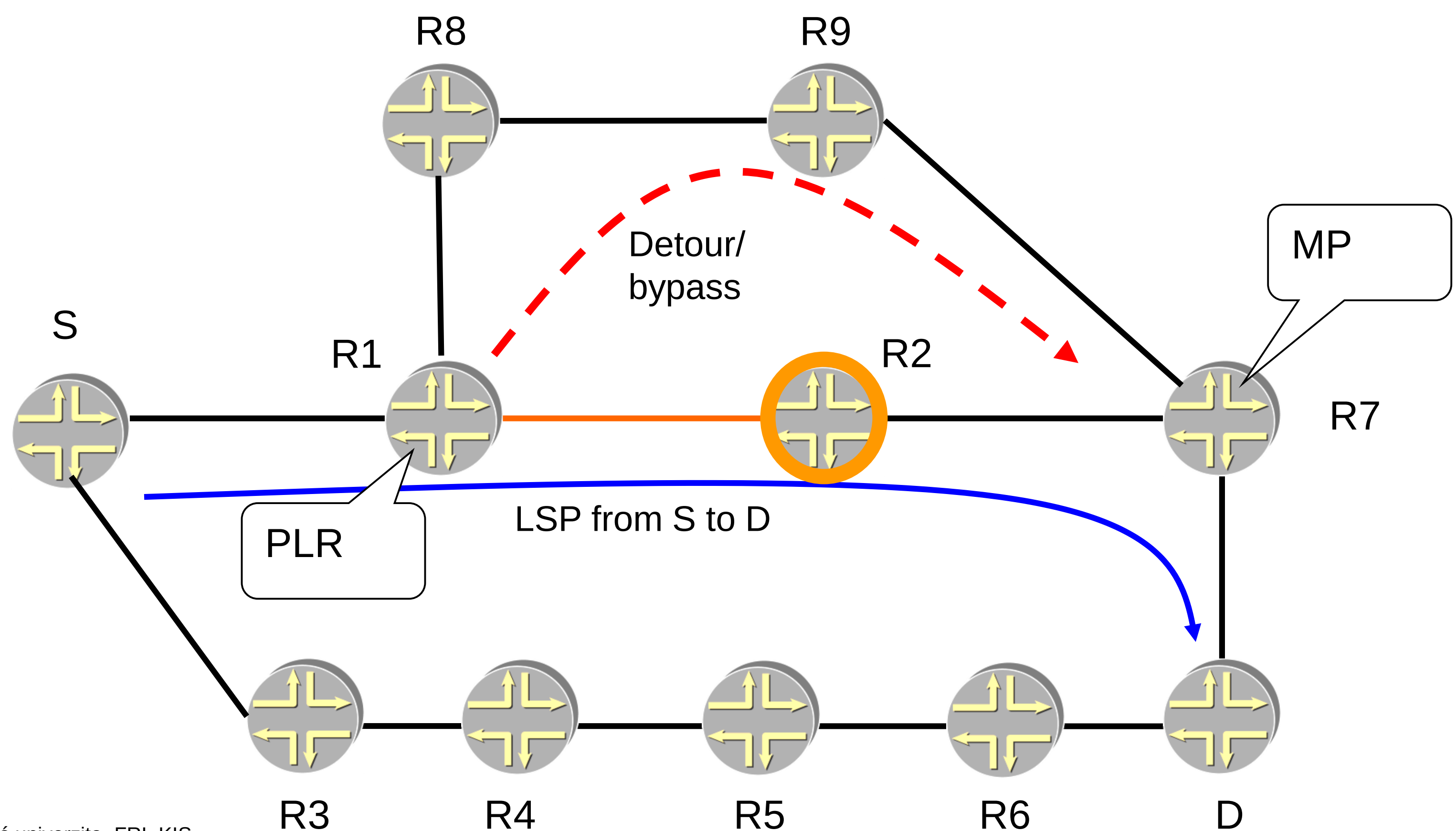
Traffic Engineering with RSVP-TE

- In MPLS, traffic engineering is inherently provided using explicitly routed paths.
  - Not IGP calculated shortest paths
- The LSPs are created independently, specifying different paths that are based on user-defined policies (bandwidth, delay, hop count, QoS, etc.). However, this may require extensive operator intervention.
- RSVP-TE (Resource Reservation Protocol) and CR-LDP (Constraint-based Routing LDP) are two possible approaches to supply dynamic traffic engineering and QoS in MPLS.

- RSVP-TE allows to carriers to provide edge-to-edge tunnels across their core networks
  - Full mesh required – might be a scalability issue
  - Management overhead
  - But allows fast traffic recovery (10s of msec) by introducing backup paths (FRR - Fast Reroute resiliency mechanism)

FRR General principles

- “Local repair” scheme
  - Upstream Router from protected link/node pre-signals protection path and pre-installs it in forwarding table.
  - If failure is detected, traffic is moved onto protection LSP.
  - PLR (Point of Local Repair), MP (Merge Point)



Protection Schemes

Path protection -> end-to-end backup path:

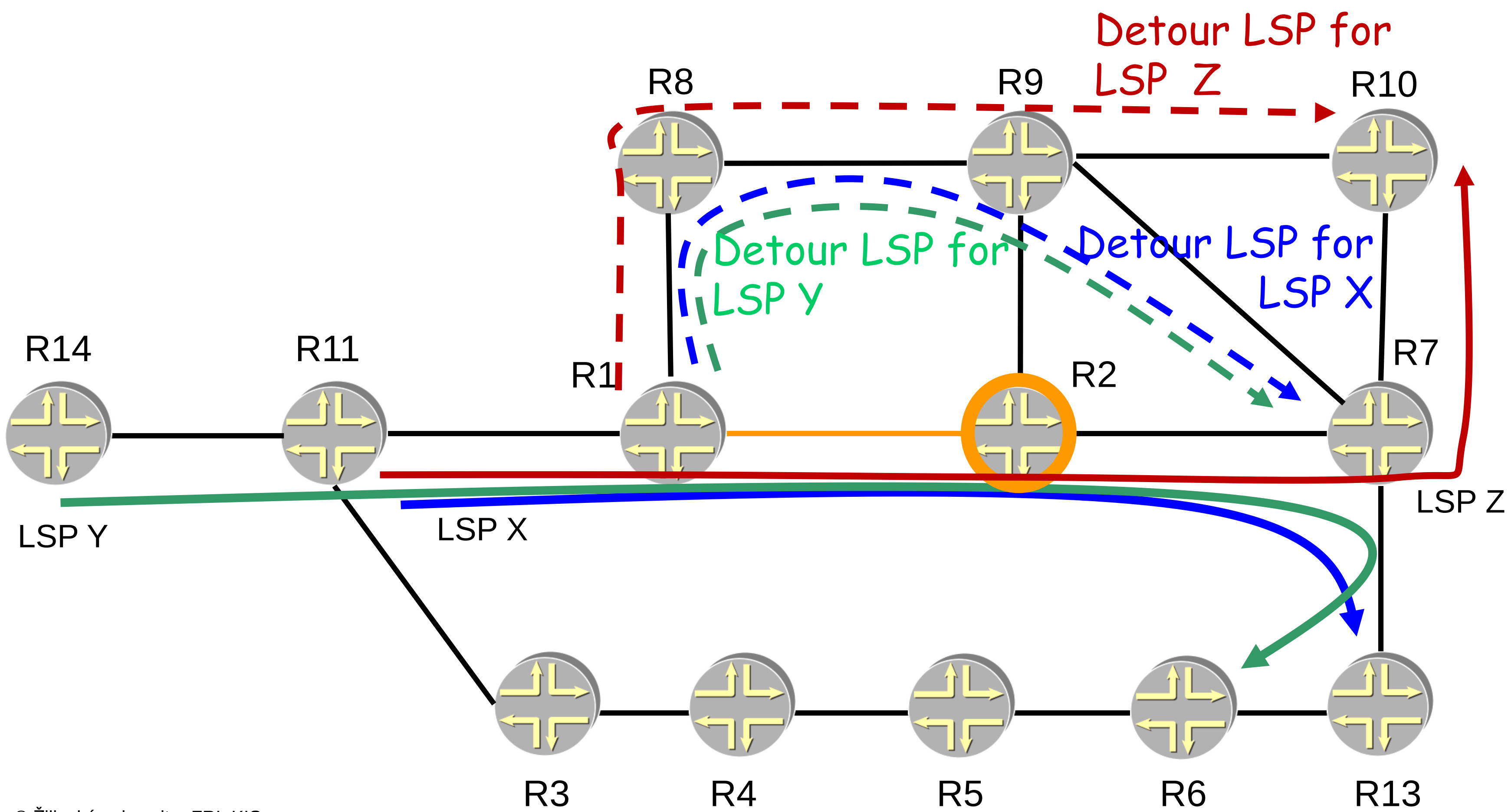
- Secondary LSP can be presignaled and make ready to take over the traffic
- RSVP error message is propagated to the LSP head end
- provides exact knowledge of where the traffic will flow following the failure
- Double-booking of resources

Local Repair Fast Reroute (FRR):

- One-to-one backup
  - Separate backup LSP (called Detour LSP) for each LSP that requires protection
- Facility backup
  - Bypass Tunnel created to protect a given facility (a link or a node). Multiple LSPs can share the same bypass tunnel
  - There are two variants:
    - Link protection
    - Node protection

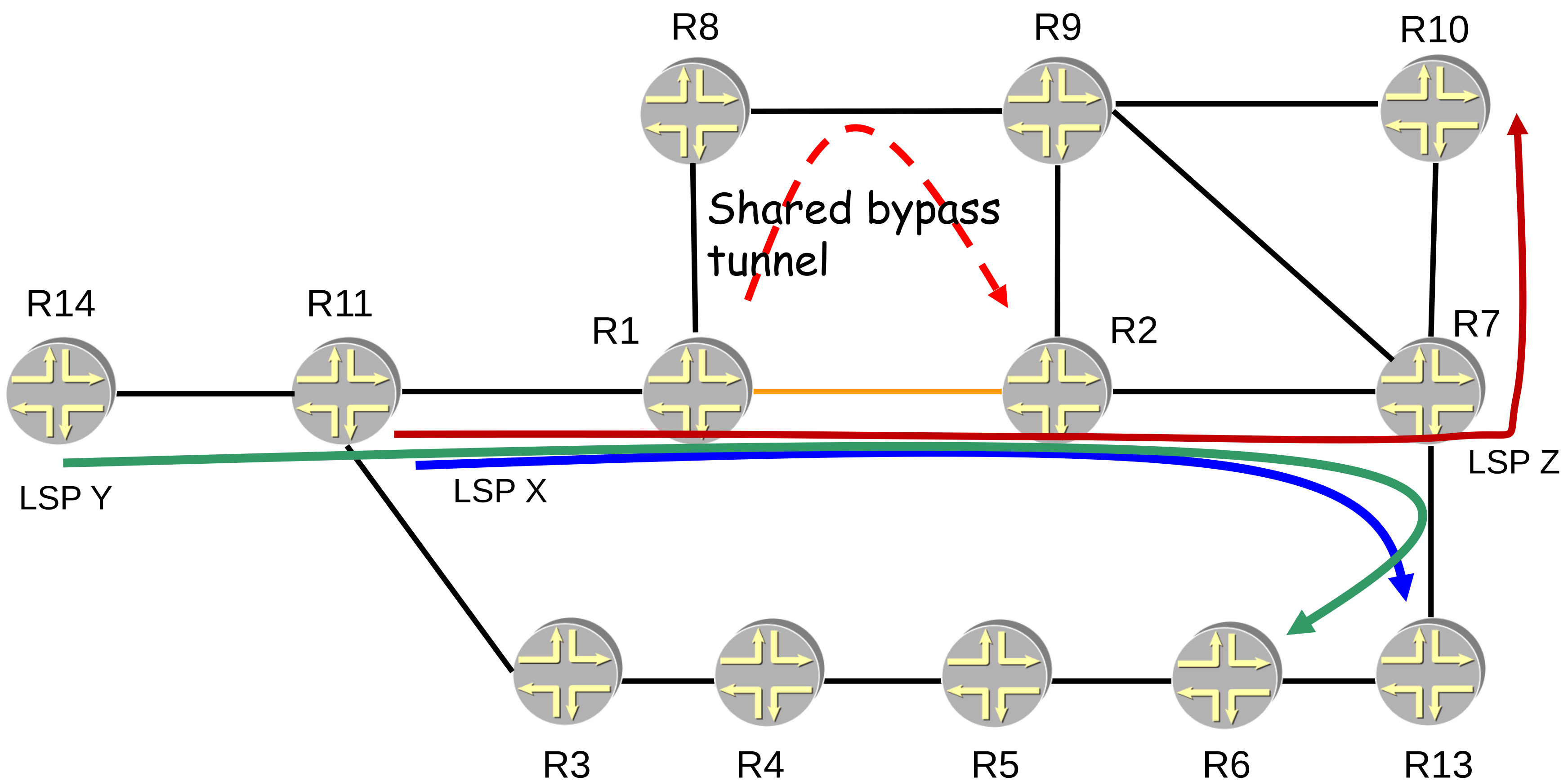
1:1 protection

Separate detour LSP for each protected LSP. 1:1 protection always protects downstream link and downstream node.



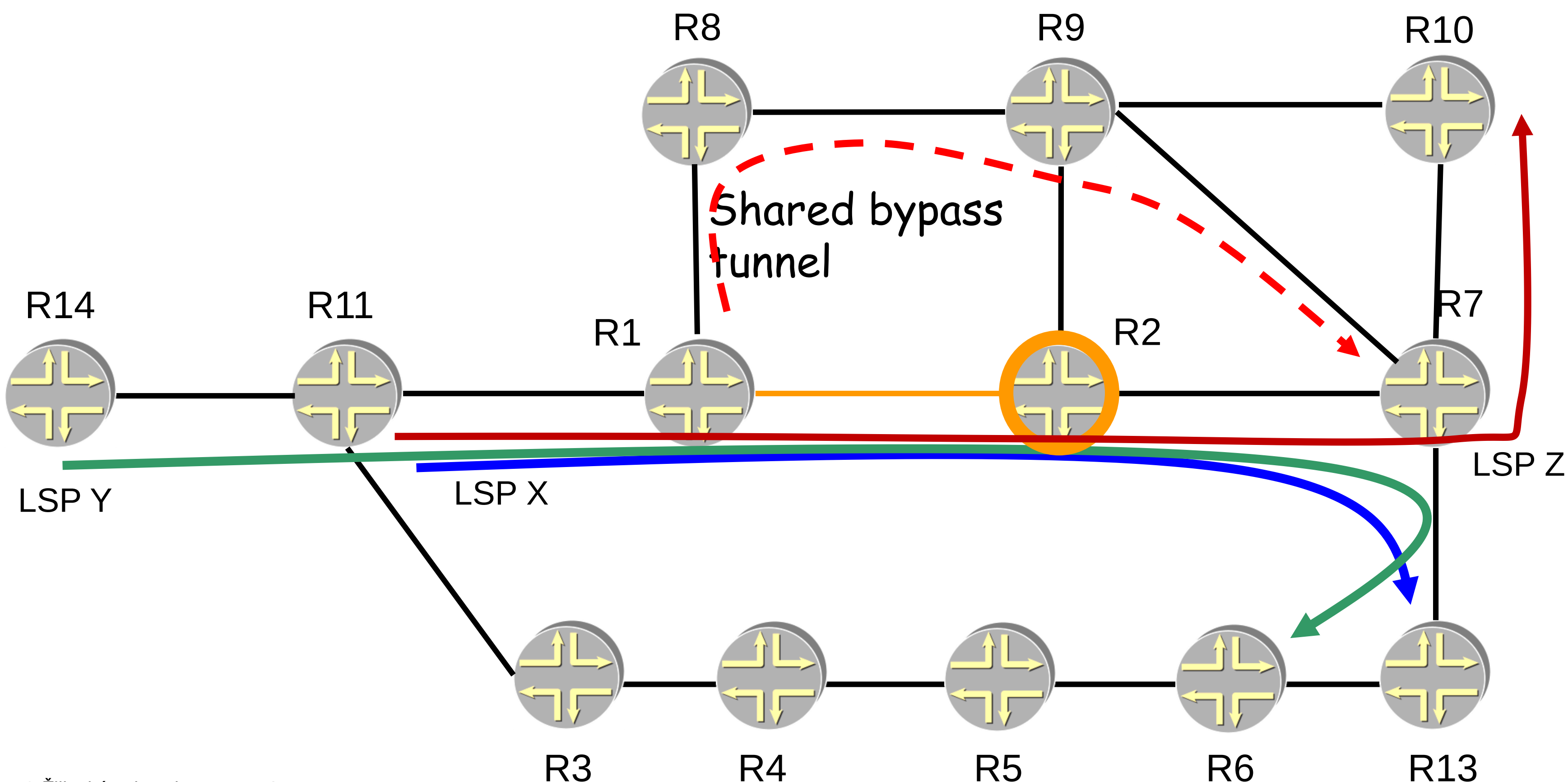
Facility protection, Link protection variant

R1 computes and signals a single bypass tunnel to the next-hop node, R2, that avoids the R1-R2 link. The bypass can be shared between all LSPs using R1->R2 link, if desired. Merge point is R2. Label stacking involved.



Facility protection, Node protection variant

R1 computes and signals bypass tunnel(s) to the next-next-hop node(s) which avoids R1-R2 link and R2 itself. Same bypass tunnel can be shared between all LSPs that pass from R1 to R2 and have the same next-next-hop. In the example, the 3 LSPs have the same next-next-hop so only one bypass was needed to protect R2



Ďakujem za pozornosť

roman dot kaloc at gmail dot com

