Piotr Gajowniczek
Instytut Telekomunikacji PW

Data Transmission

MPLS – Multiprotocol Label Switching

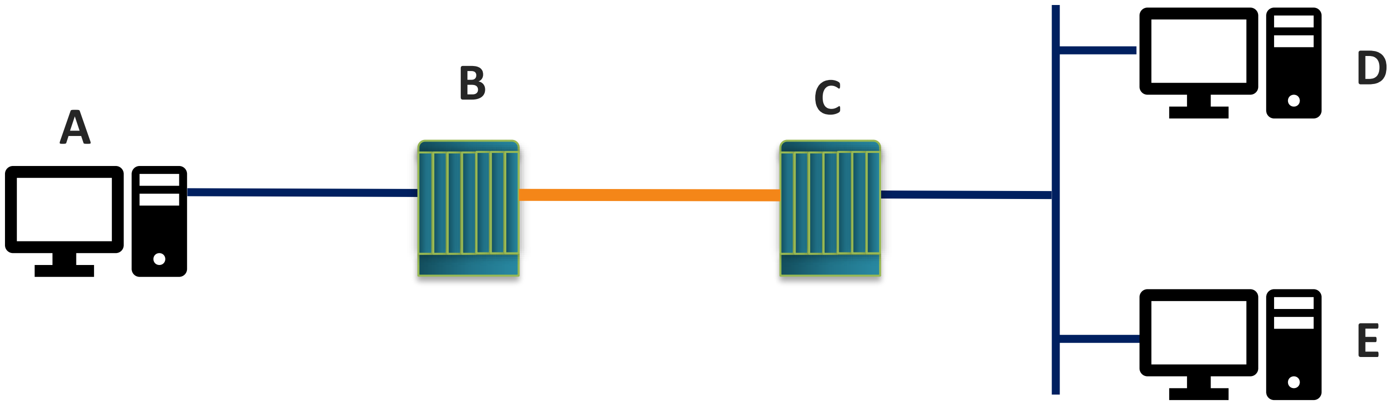
fundamentals and applications

MPLS – introduction

- *RFC 3031, RFC 3032 – MPLS architecture*
  - initially invented in the '90s to improve the forwarding speed of IP routers and introduce additional functionality to IP control plane
- *Nowadays, MPLS is very popular in the ISP core networks*
  - network virtualization (VPNs)
  - resource and service management
  - traffic engineering, QOS
  - network resiliency
- *Extension beyond the ISP core applications also exist*
  - GMPLS (Generalized MPLS) – single control plane extended to optical transport technologies
  - Seamless MPLS – extension of MPLS towards the access network

Layered model / switching vs routing

- *Switching (L2)*
  - findin the destination MAC address in MAC table (*forwarding table, bridge table*)
  - establish the outbound interface
  - send packet from input queue to output queue
  - packet is not modified in any way
- *Routing (L3)*
  - remove L2 header
  - lookup for *next hop* (*forwarding table; routing table*)
  - determine the correct destination L2 (MAC) address
  - *rewrite MAC header*
  - send packet to the output queue

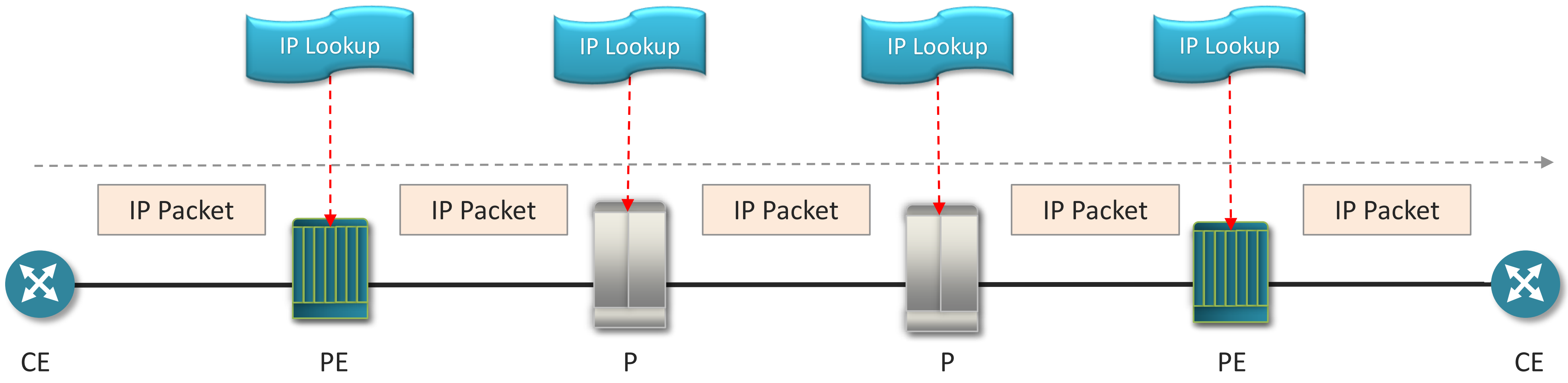


- L3 switch ?

IP routing ...

IP routing:

- *data link layer frame validation*
- *network-layer protocol demultiplexing*
- *IP packet validation*
- *forwarding decision — longest prefix match*
- *data link frame construction*

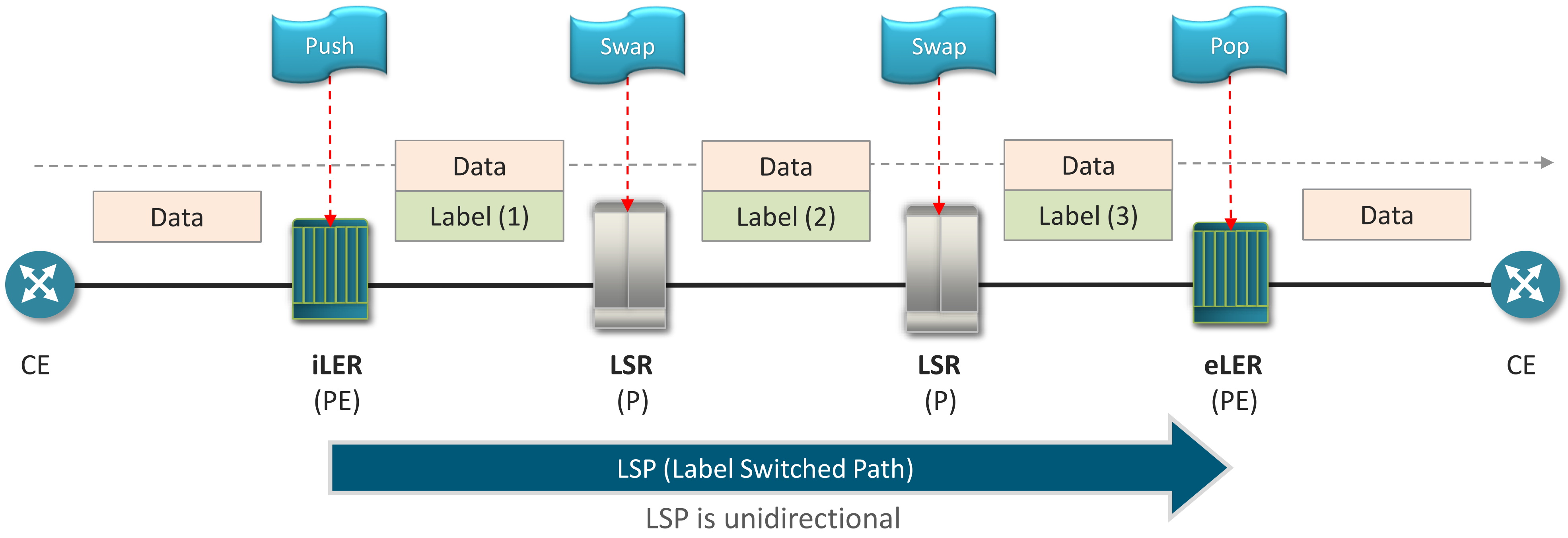


… vs MPLS

IP routing:

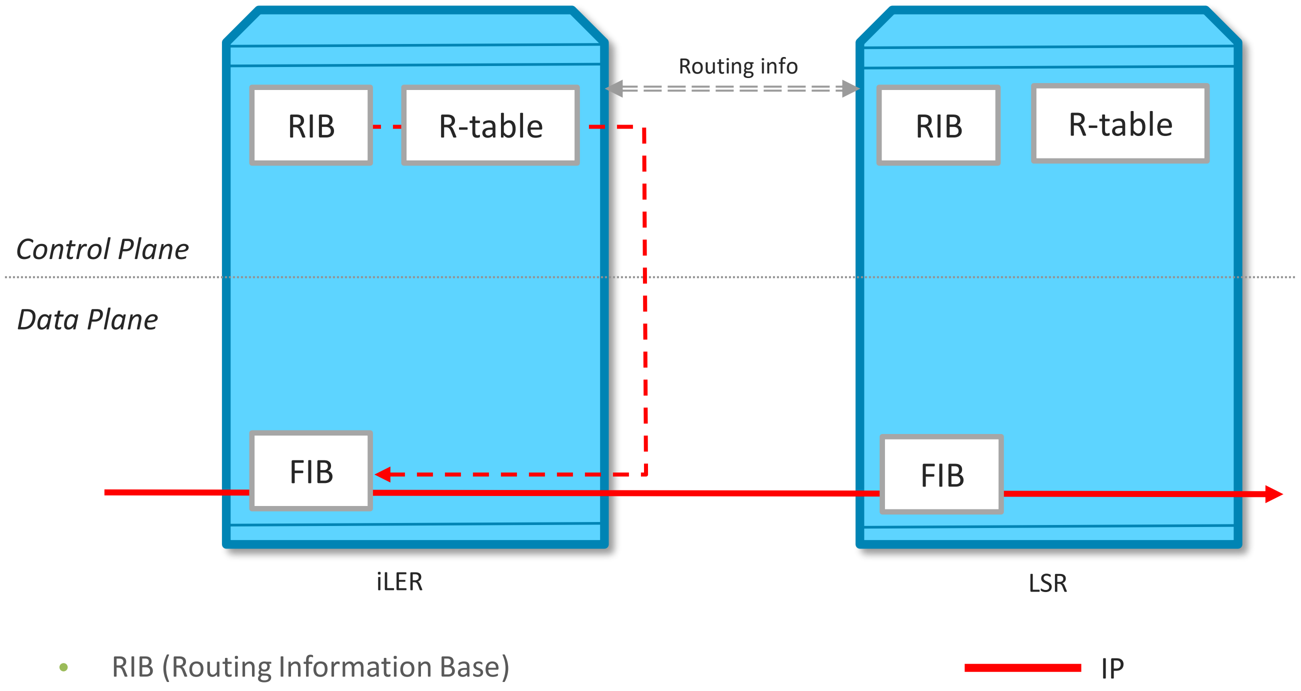
- *data link layer frame validation*
- *network-layer protocol demultiplexing*
- *IP packet validation*
- *forwarding decision — longest prefix match*
- *data link frame construction*

MPLS:
Label *Push*, *Swap* & *Pop*



IP control plane …

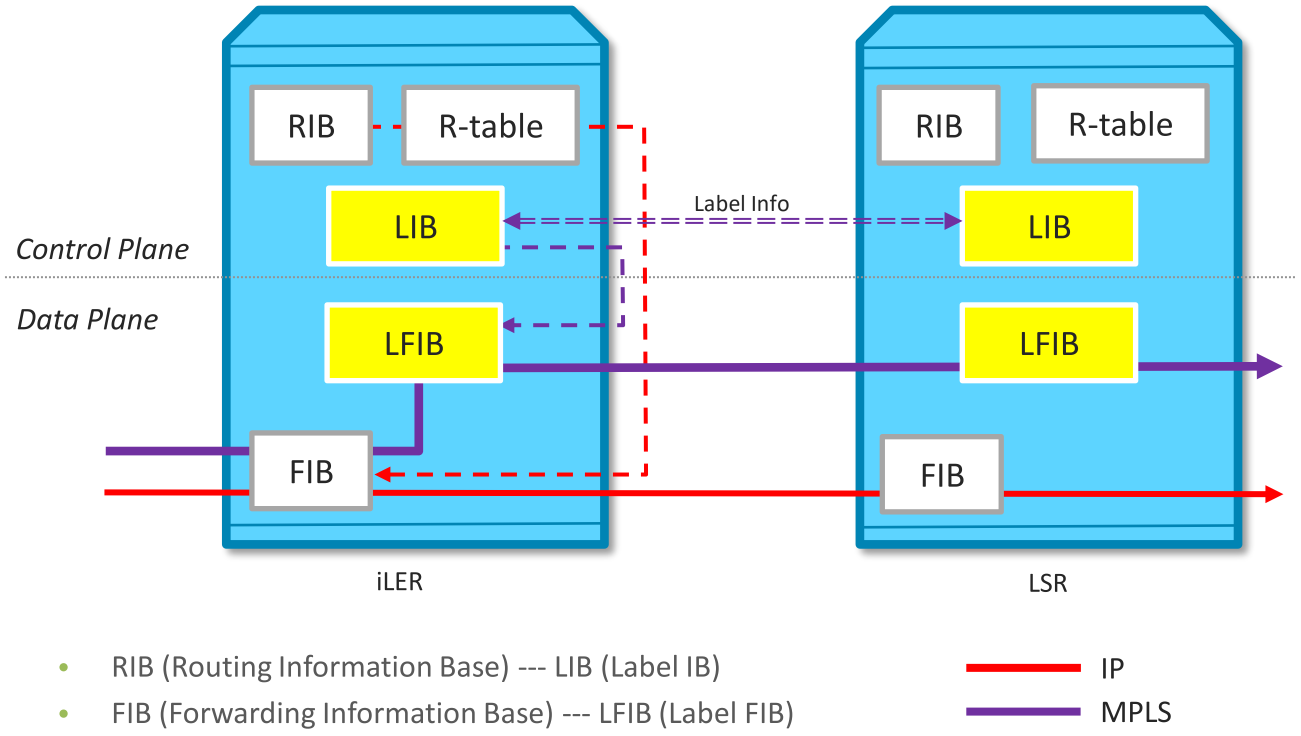
- *FEC (Forwarding Equivalence Class)*
  - IP routing – FEC = IP Prefix; FEC lookup done at each hop
  - MPLS – other FEC criteria possible, FEC lookup only at an iLER



- RIB (Routing Information Base)
- FIB (Forwarding Information Base)

… vs MPLS control plane

- *FEC (Forwarding Equivalence Class)*
  - IP routing – FEC = IP Prefix; FEC lookup done at each hop
  - MPLS – other FEC criteria possible, FEC lookup only at an iLER



- RIB (Routing Information Base) --- LIB (Label IB)
- FIB (Forwarding Information Base) --- LFIB (Label FIB)

- *Label retention*
  - Liberal
  - Conservative

MPLS – labels and tunnels

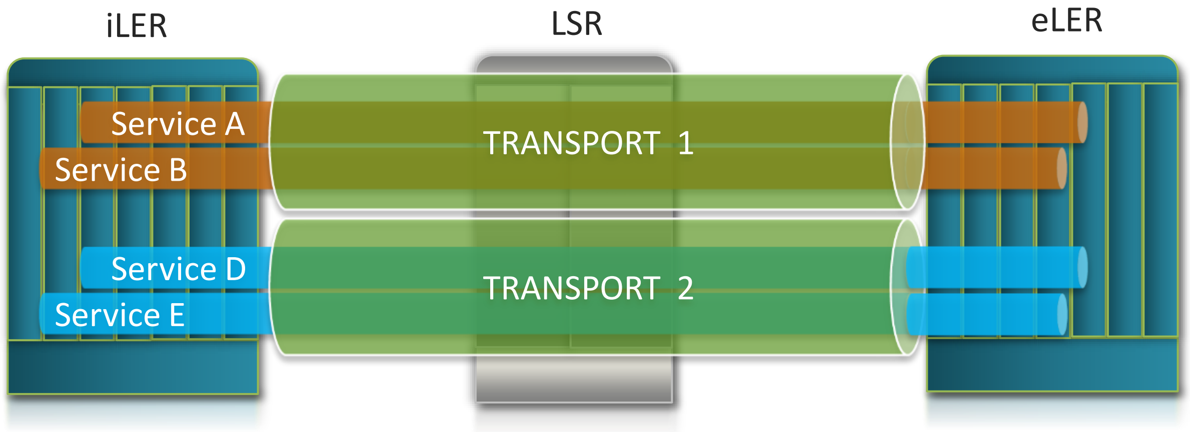
Data (MPLS Payload)
MPLS Header (top)
…
MPLS Header (bottom)
L2 Header

20 bit	3 bit	1 bit	8 bit
Label	EXP	S	TTL

Traffic Class (RFC 5462)

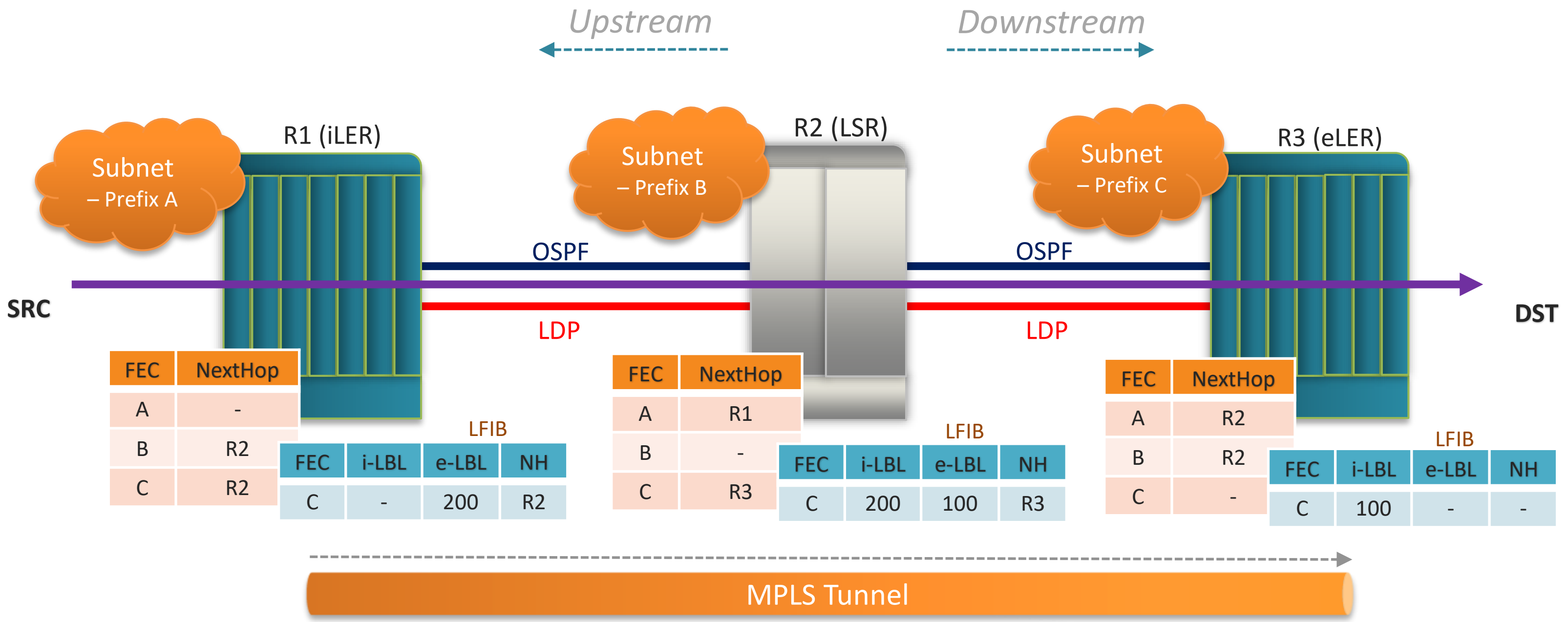
- Important only for external (top) label
- „Pipe Mode” and „Uniform Mode”

- Label stack (Frame Mode – label between L2 and the encapsulated data)
- Label has only **local significance** – the binding between a FEC and a label is not unique in the network



LSR handles only transport tunnels

MPLS – tunel set-up



- *MPLS tunnel set-up requires:*
  - FEC (IP prefixes) reachability = OSPF
  - distribution of label mappings between routers

- *Label distribution protocols*
  - LDP (Label Distribution Protocol) – „Downstream Unsolicited”
  - RSVP-TE (Resource Reservation Protocol) – „Downstream on Demand”

- *R1 – Request(FEC C)*
- *R3 – Response(FEC C, 100)*
- *R2 – Response(FEC C, 200)*

MPLS – label distribution protocols

transport tunnels

- *LDP (Label Distribution Protocol)*
  - TCP based
  - tunnels built based on IGP (full-mesh)
  - simple configuration
  - automatic creation of tunnels
  - no traffic engineering
  - convergence time depends on IGP
  - label distribution in „*downstream unsolicited*” approach
- *RSVP TE (Resource Reservation Protocol with Traffic Engineering)*
  - explicit tunnels (not following IGP paths)
  - additional constraints (administrative and TE-related) for advanced path calculation
  - bandwidth reservation for LSP (CAC)
  - advanced LSP protection against failures
  - label distribution in „*downstream on demand*” approach

service tunnels

- *T-LDP (Targeted LDP) – RFC 4447*
  - multi-hop LDP
  - for L2 services
  - for end-to-end tunnels between PE routers
- *MP-BGP (Multi-Protocol BGP) – RFC 4364*
  - for L3 services

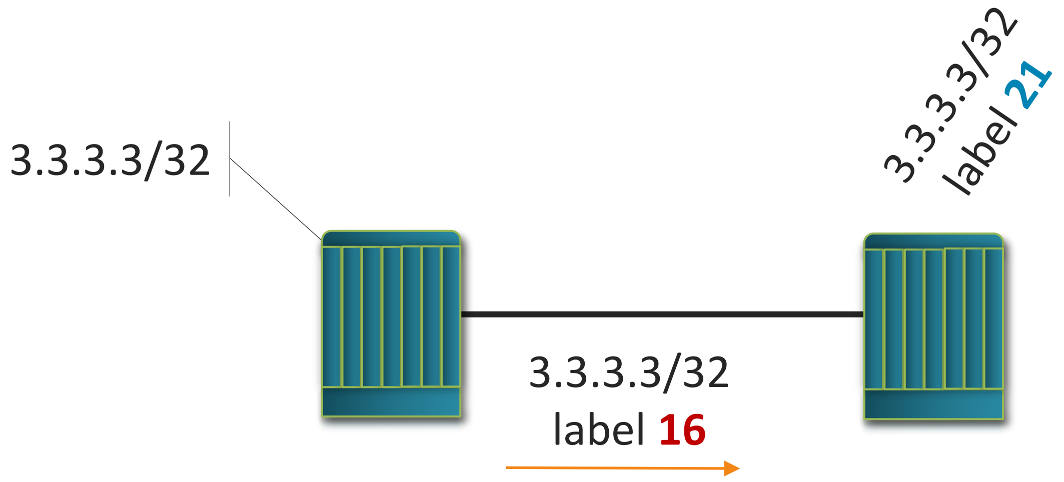
LDP (Label Distribution Protocol)

- *To establish LDP adjacency:*
  - router sends UDP multicast (224.0.0.2:646) hello packets to discover other neighbors
  - next, builds the neighbor adjacency (on loopback addresses) using a TCP connection for exchanging label information (each router has a unique ID called the LSR ID; much like in OSPF).
- *LDP creates a label binding for each prefix by default and sends them to neighbors (downstream unsolicited) – Label Mapping Advertisement, Label Withdraw*
- *LDP configuration*

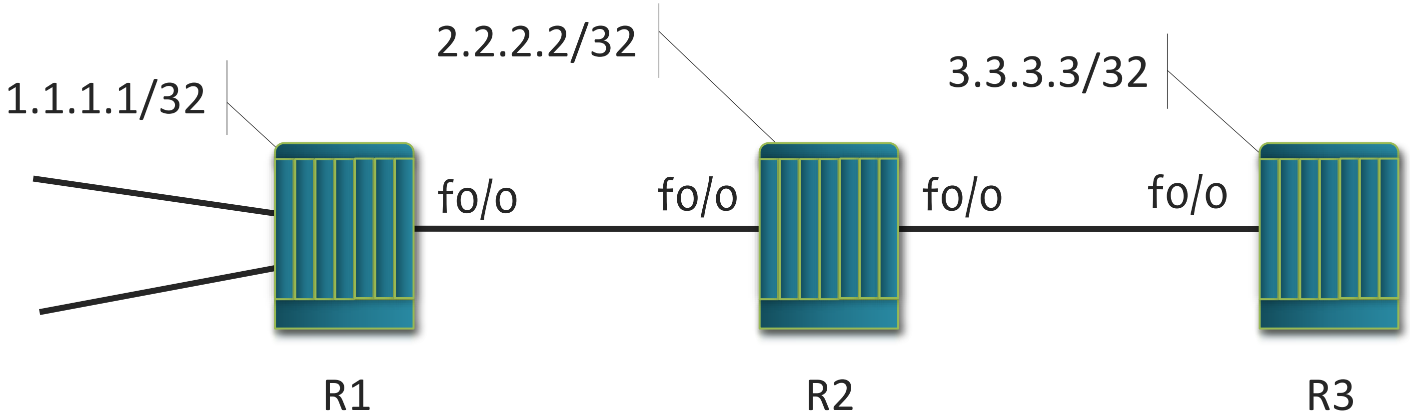
```
R1(config)#interface FastEthernet 0/0
R1(config-if)#mpls ip
```

```
R1#show mpls interfaces
Interface IP Tunnel BGP Static Operational
FastEthernet0/0 Yes (ldp) No No No Yes
FastEthernet0/1 Yes (ldp) No No No Yes
```

```
R1#show mpls ldp neighbor fe0/0
```



LDP



- *LIB (Label Information Base)*

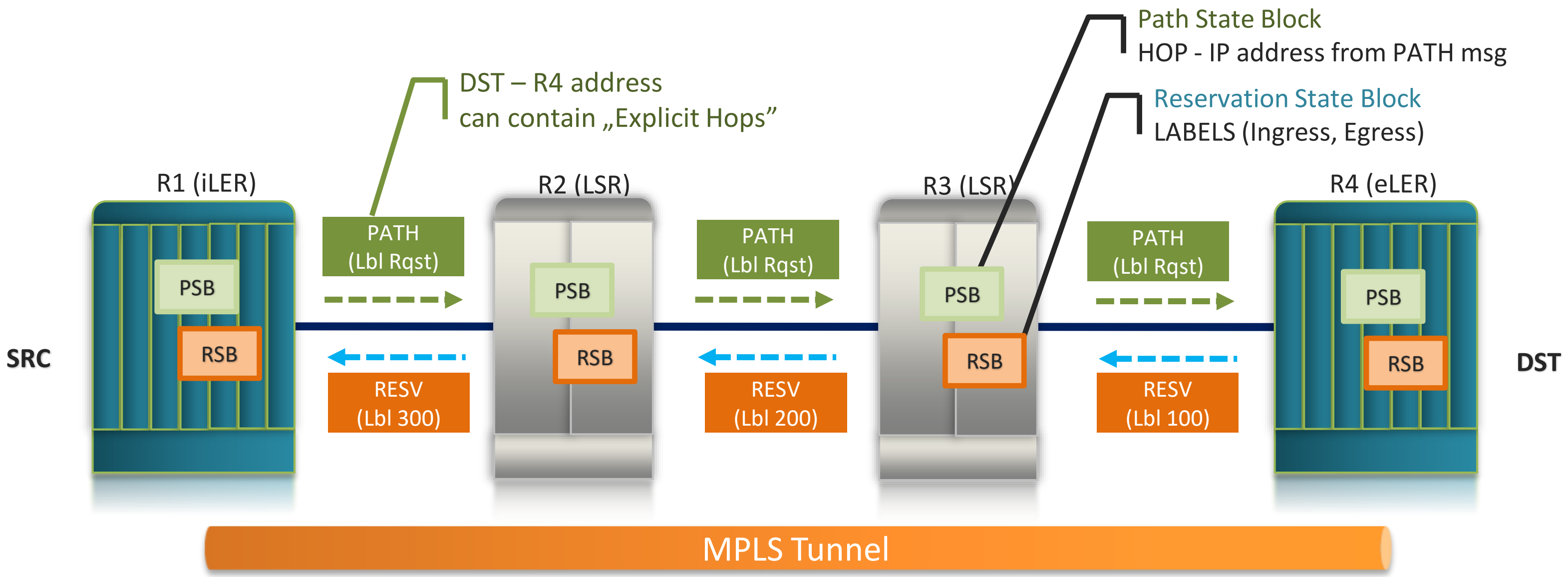
```
R1#show mpls ldp bindings
lib entry: 2.2.2.2/32, rev 7
  local binding: label: 16
  remote binding: lsr: 2.2.2.2:0, label: imp-null
lib entry: 3.3.3.3/32, rev 9
  local binding: label: 17
  remote binding: lsr: 2.2.2.2:0, label: 21
```

LFIB (Label Forwarding Information Base)

```
R1#show mpls forwarding-table
Local Outgoing  Prefix      Bytes Label  Outgoing    Next Hop
Label Label      or Tunnel   Switched     interface
                 Id
16     Pop Label  2.2.2.2/32     0           F0/0         10.1.1.1
17     21         3.3.3.3/32     0           F0/0         10.1.1.1
```

- *implicit null label – penultimate hop popping (PHP)*

MPLS – the RSVP protocol



- *RFC 3209 – RSVP as LDP*
- *RSVP TE:*
  - LSP definition
  - path calculation „outside" IGP metrics („link colors", bandwidth etc.)
  - tunel protection (Secondary Paths, Fast Reroute)
  - resource reservation (CAC)

- *MPLS tunnel can be composed of many paths (LSP-Paths).*
  - one „primary" path and seven „secondary"
  - only one active at a time

- *Other message types:*
  - PATH Tear: (downstream), RESV Tear: (upstream)
  - PATH Error, RESV Error:
  - Hello (RSVP heartbeat)
  - Summary Refresh (for less signalling)

RSVP Traffic Engineering: path calculation

- *Source routing*
- *Path Option - Explicit*
  - manual configuration at source router
  - high signalling overhead
- *Path Option – Dynamic = APC (Advanced Path Calculation)*
  - CSPF (Constrained Shortest Path First)
  - additional criteria
    - bandwidth reservation state
    - administrative criteria (link colors)
    - hop limit
    - TE metric
    - Explicit route („strict hops", „loose hops")
    - Shared Link Groups

reservations are made in the Control Plane

- actual bandwidth usage in the Data Plane is not considered
- requires relevant QoS solutions in the Data Plane

resiliency

- allows automated creation of backup paths and detours (Fast Reroute) that are disjoint with the primary path

RSVP-TE – how to force tunner route?

- *Signalling*
  - information about the route is conveyed in the RSVP PATH message in an ERO (Explicit Route Object)
  - ERO is updated in each intermediate router

- *Bandwidth reservation*
  - CSPF algorithm calculates a path with the required amount of unreserved bandwidth using data from TED database at source router
  - downstream:
    - reservation request is signaled in RSVP PATH message
    - each router checks bandwidth availability on outgoing link (CAC)
  - upstream:
    - bandwidth is reserved in each router on path (RSVP RESV message)
    - *Unreserved Bandwidth* – updated and advertised

- *Least-Fill Bandwidth Reservation rule*
  - if CSPF has found multiple paths with the same metric

- *relevant QOS policies in the Data Plane are required*

RSVP – TE variables

the need for additional constraints and link state data has to be reflected in routing protocol

↓

- *OSPF-TE (OSPF Traffic Engineering)*
  - RFC 2370: The OSPF Opaque LSA Option
  - Opaque LSAs – deneric mechanism for OSPF extensions

OSPF TE

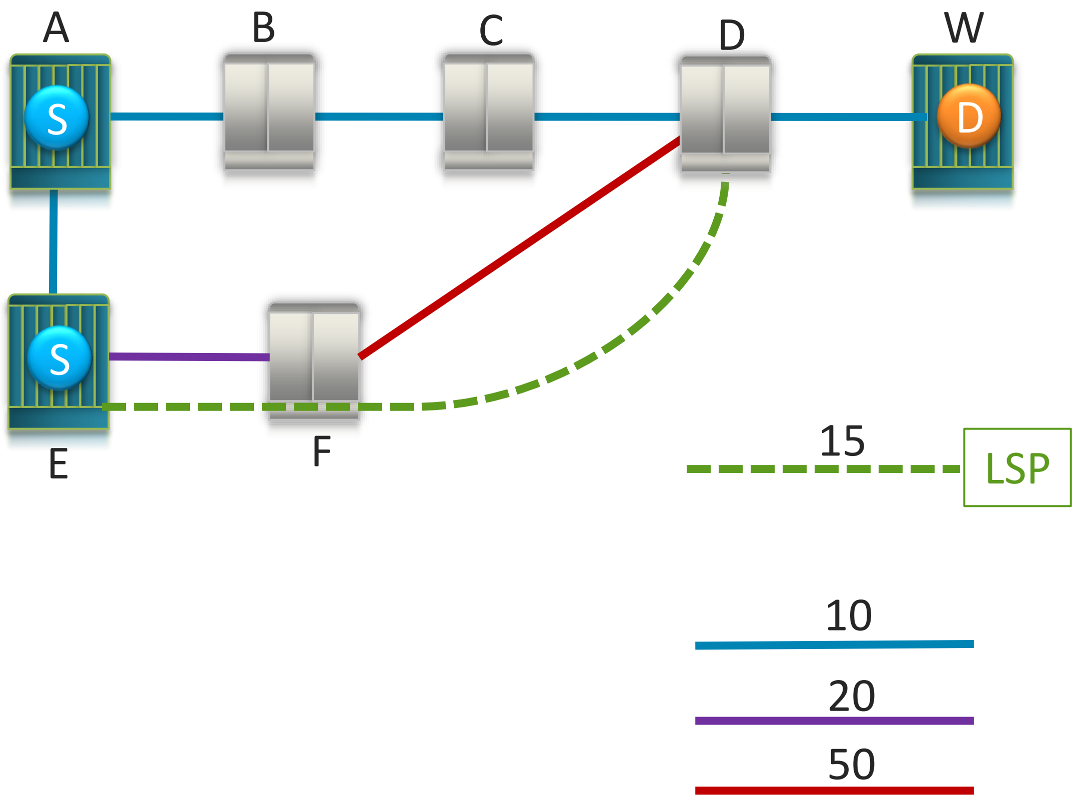
- *routers create additional database – TED (Traffic Engineering Database) for storing additional link attributes distributed by Opaque LSAs (Type 10)*
- *Opaque LSA Flooding – activated when:*
  - link state (up/down), link configuration of bandwidth reservation state changes
  - periodically (as in IGP)
- *Opaque LSA Type 10 contains Link TLV object, used to advertise information about links handled by RSVP-enabled routers:*
  - link type, link ID
  - IP addresses of interfaces on both sides of the link
  - TE metrics
  - maximum bandwidth
  - maximum reservable bandwidth (per LSP priority)
  - unreserved bandwidth (100 = 100%, overbooking possible)
  - administrative group
  - Shared Risk Link Group (SRLG)

MPLS – priorities and preemption

- *LSP Soft Preemption*
  - higher priority LSPs can preempt lower priority paths
  - priorities work in conjunction with knowledge of the Unreserved Bandwidth parameter – current values are advertised by OSPF TE for each priority level
- *Setup and Hold priorities (0 to 7, lower value = higher priority)*
  - LSP A can preempt LSP B if Setup Priority(A) < Hold Priority(B)
  - LSP priorities are signaled in RSVP PATH message, in SESSION_ATTRIBUTE object
- *RSVP Preemption-Timer & LSP Retry-Timer*
  - preemption by MBB (Make Before Break)
  - CSPF tries to find another route for preempted LSP
    - periodically (*Retry-Timer*)
    - preemption (status = down) after time defined in *Preemption-Timer* (unless a new route was found earlier)

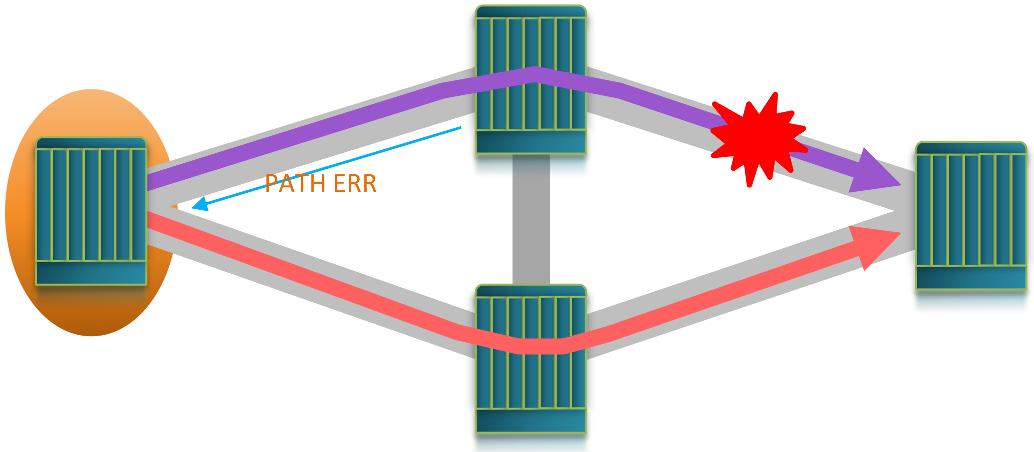
Using MPLS tunnels in IP routing

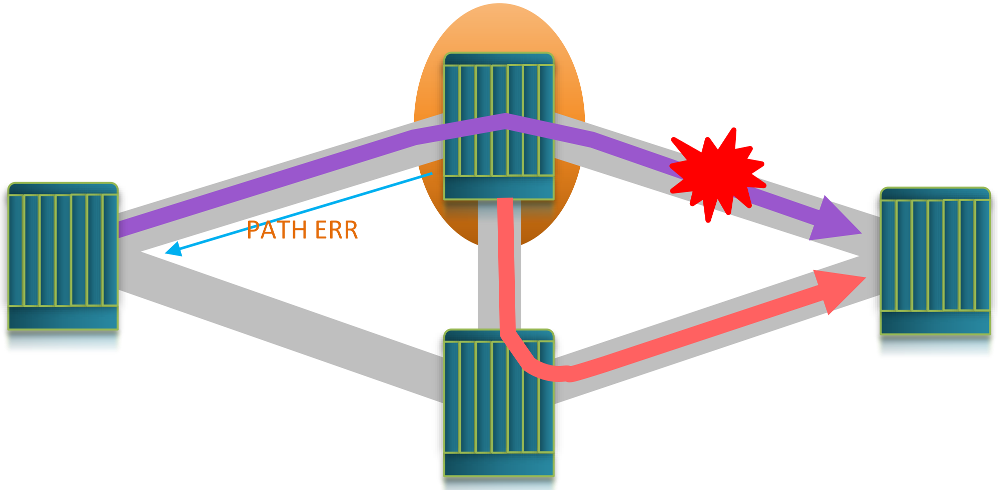
- *Static mapping*
  - LSR is configured to send packets through the tunnel
  - scalability issues
- *Dynamic mapping (done by IGP)*
  - LSP is treated as an interface (tunnel) with an associated metric
  - metric can influence routing decisions (choice between CSPF and SPF)
  - two cases:
    - IGP can use tunnel for SPF calculation only in an iLER (*autoroute announce*)
    - tunnel may be signaled in LSA for use by other routers to calculate shortest paths (*forwarding adjacency*)



MPLS – failure resiliency

- *factors influencing quality of protection*
  - avg. failure detection speed
    - OSPF Hello (30 s)
    - RSVP Hello (9 s)
    - Bidirectional Forwarding Detection (<1s)
      - » „IP level heartbeat”
  - speed of failure advertising
  - service restoration time (switchover speed)





- *Secondary LSP*
  - switchover at source router
  - switchover time depends on PATH ERR message delivery to source router
  - max. 7 standby (Secondary) paths
    - Hot Standby
    - Non-Standby

- *Fast Reroute*
  - local switchover (<50 ms)
  - PATH ERR conveys only the information about failure
  - detour paths are calculated automatically (CSPF)
  - protects against node and link failures
  - protection types:
    - One-to-One Backup (Detour)
    - Facility Backup (Bypass Tunnel)