Institute of Telecommunications
Warsaw University of Technology
2019

internet technologies and standards

- Piotr Gajowniczek
- Andrzej Bąk
- Michał Jarociński

MPLS – Multiprotocol Label Switching

MPLS – introduction

- *RFC 3031 – MPLS architecture*
  - initially for improving IP packet switching by simplified lookup
- *Now, a popular ISP core network technology*
  - virtualization and management of network resources, network services
  - traffic engineering, QOS
  - high availability
  - consolidation of services on a single infrastructure
    - ability to support various services, applications and solutions over a converged network infrastructure

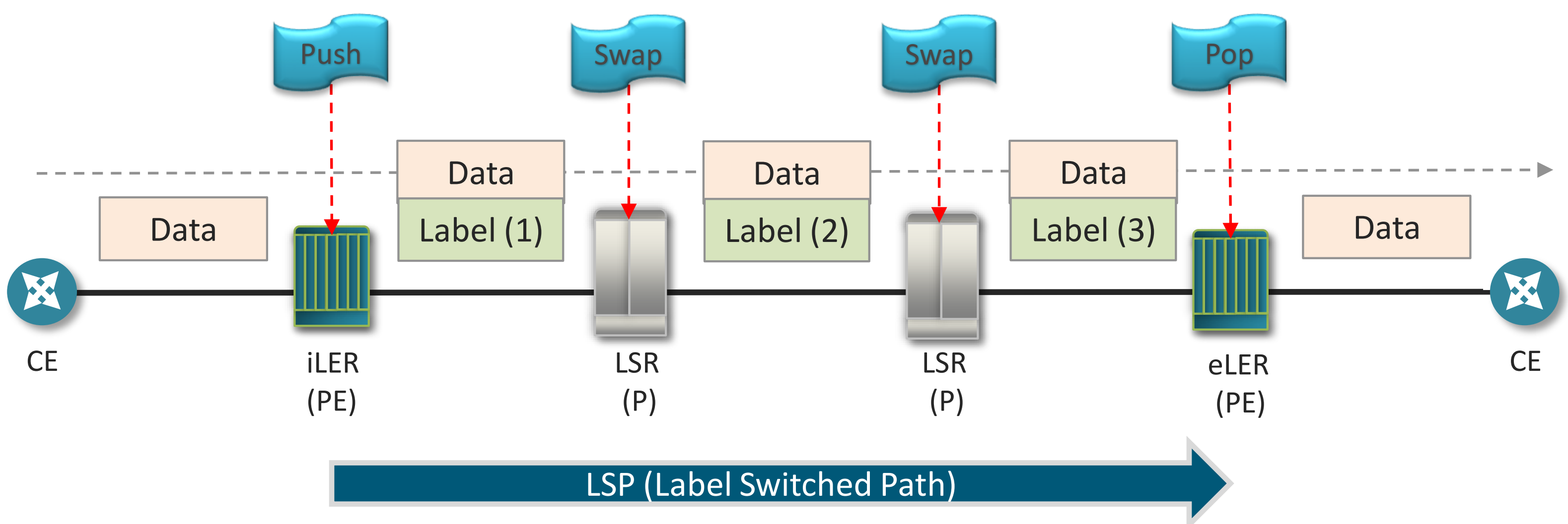
MPLS - introduction

IP routing:

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

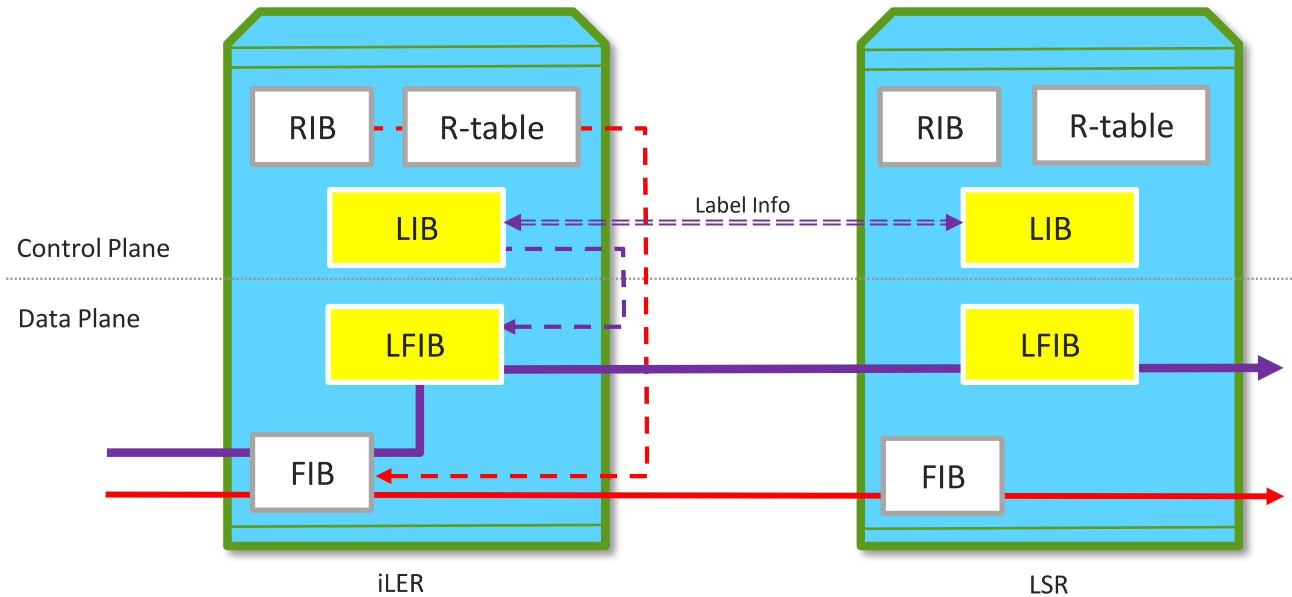
MPLS:

- Push, Swap & Pop



MPLS – 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
- *LIB (Label Information Base) and LFIB (Label Forwarding Inf. Base)*



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

IP

MPLS

MPLS – labels and tunnels

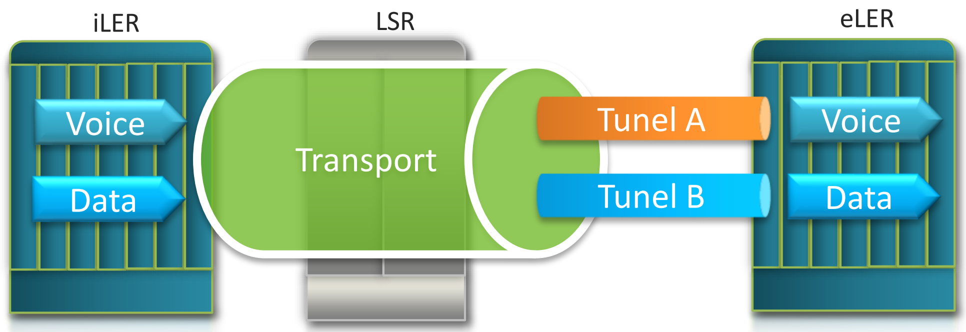
Data (MPLS Payload)
MPLS Header (top)
…
MPLS Header (bottom)
L2 Header

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

Traffic Class (RFC 5462)

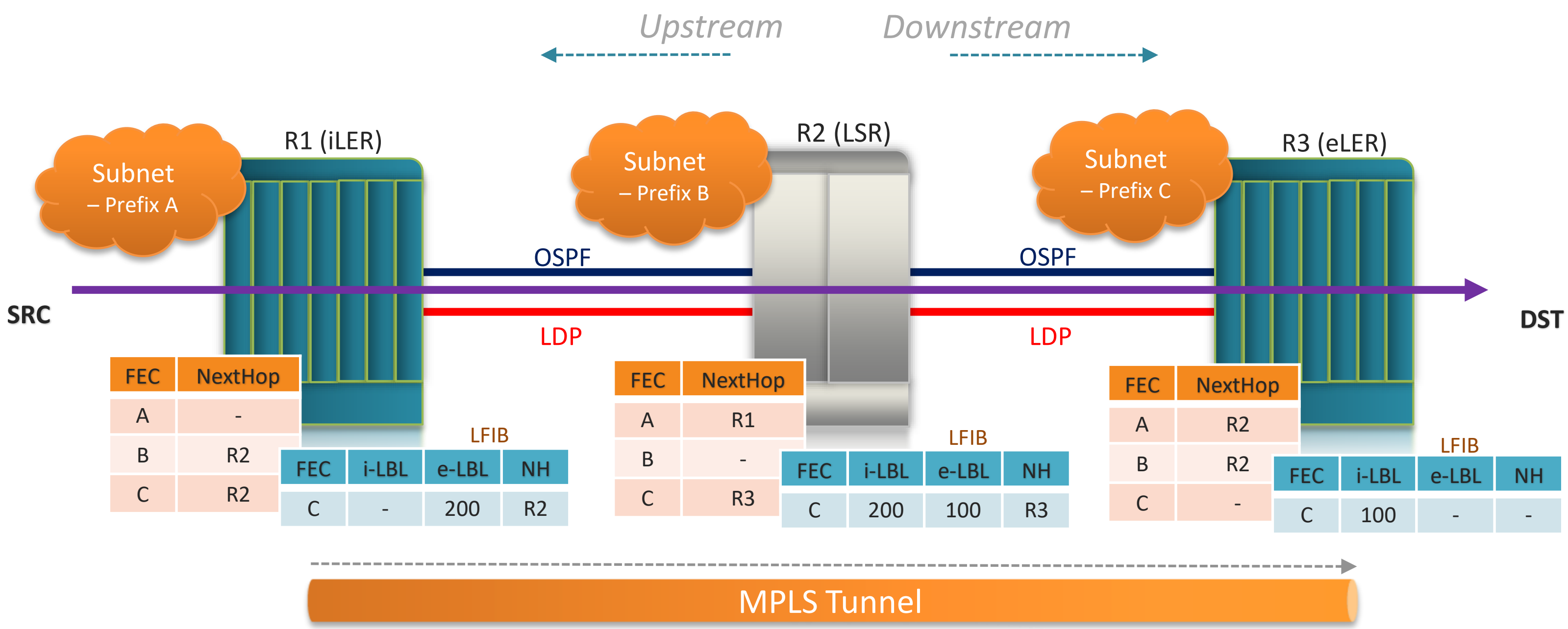
- Important only for external (top) label
- „Pipe Mode” and „Uniform Mode”

- Label stack (Frame Mode – label between L2 and the encapsulated data)



LSR handles only transport tunnels

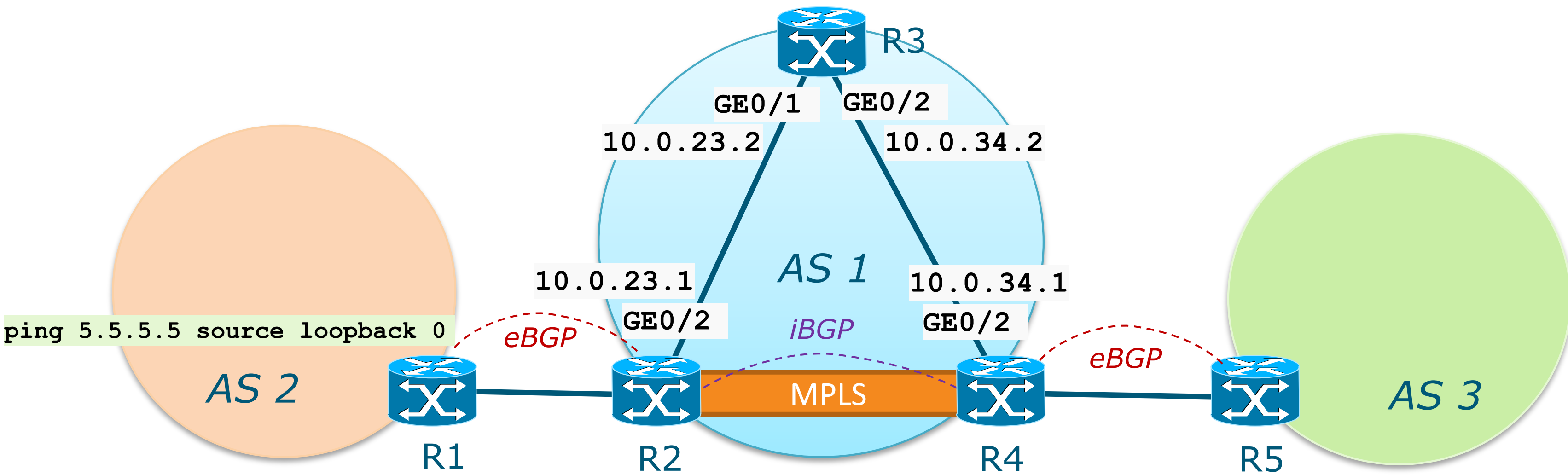
MPLS – tunel set-up



- *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)*

BGP free core

- *you can use MPLS tunnels to carry iBGP sessions between border routers*
  - the core routers do not need to handle IP addresses – no BGP necessary



- *starting MPLS (repeat on R2, R3, R4)*

```
R2(config)#interface GE0/2
R2(config-if)#mpls ip
```

```
R2#show mpls forwarding-table
Local  Outgoing    Prefix          Bytes Label  Outgoing    Next Hop
Label  Label       or Tunnel Id    Switched     interface
16     17          4.4.4.4/32      0            GE0/2       10.0.23.2
17     Pop Label   10.0.34.0/30    0            GE0/2       10.0.23.2
18     Pop Label   3.3.3.3/32      0            GE0/2       10.0.23.2
```

```
R3#show mpls forwarding-table
Local  Outgoing    Prefix          Bytes Label  Outgoing    Next Hop
Label  Label       or Tunnel Id    Switched     interface
16     Pop Label   2.2.2.2/32      0            GE0/1       10.0.23.1
17     Pop Label   4.4.4.4/32      0            GE0/2       10.0.34.1
```

LDP – Label Distribution Protocol

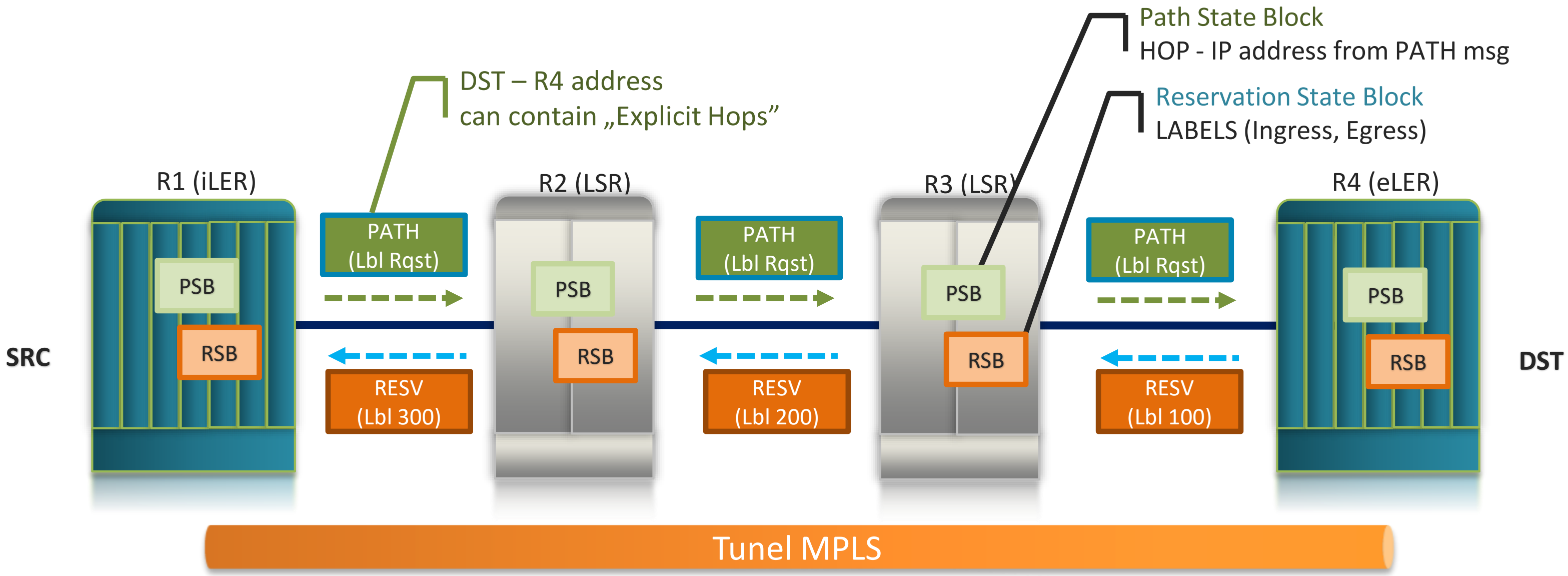
- LDP generates labels automatically and exchanges them between routers
  - first, the neighbor adjacency needs to be established (UDP multicast *hello* discovery protocol)
  - adjacency is built using TCP connection
  - routers have unique LSR IDs
  - labels are locally generated for each prefix in a routing table and added to the LIB
  - the LIB is a base for building LFIB
- the adjacent LDP neighbors exchange information stored in their LIBs
  - this distributes information on what labels should be used for different FECs

MPLS – label distribution protocols

- *LDP (Label Distribution Protocol)*
  - 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)*
  - tunnels can be defined administratively („outside" 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

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)
- *LSP = MPLS tunnel, can be composed of many paths (LSP-Paths).*
  - one „primary" path and seven „secondary"
  - only one active at a time
- *other RSVP messages:*
  - PATH Tear: (downstream), RESV Tear: (upstream)
  - PATH Error, RESV Error:
  - Hello (RSVP heartbeat)
  - Summary Refresh (for less signalling)

MPLS – Traffic Engineering – path calculation

- *Strict LSP*
  - manual configuration at source router
  - high signalling overhead

- *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

MPLS – 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 – CSPF algorithm

- *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*

MPLS – priorities and preemption

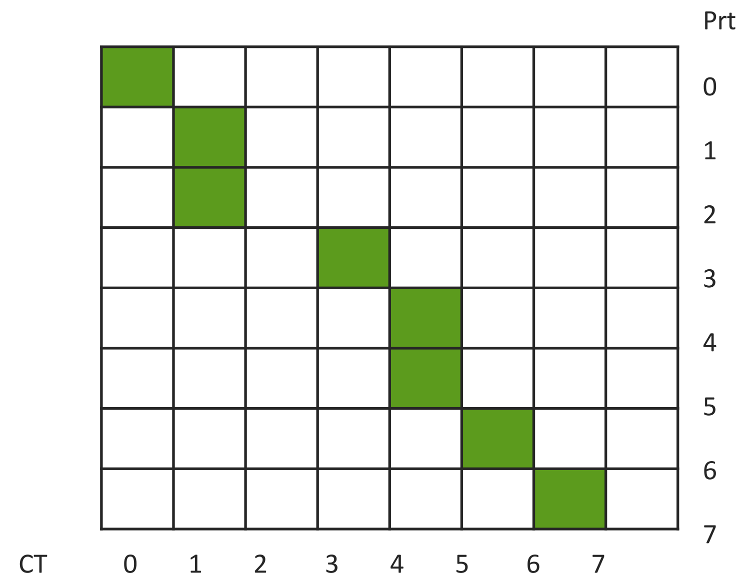
- *LSP Soft Preemption*
  - higher priority LSPs can preempt lower priority paths
  - priorities work in conjunction with knowledge of the Unreserved Bandwidth parameter – current values are advertised by OSPF TE for each priority level
- *Setup and Hold priorities (0 to 7, lower value = higher priority)*
  - 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)

MPLS – DiffServ Aware Traffic Engineering

Class Type	FC #	FC	
High Priority	7	NC (Network Control)	Signaling
	6	H1 (High-1)	Signaling or RT traffic
	5	EF (Expedited Forwarding)	RT services
	4	H2 (High-2)	
Assured	3	L1 (Low-1)	Low loss traffic, network management
	2	AF (Assured Forwarding)	Low loss traffic
Best Effort	1	L2 (Low-2)	„Best effort" data transfer
	0	BE (Best Effort)	

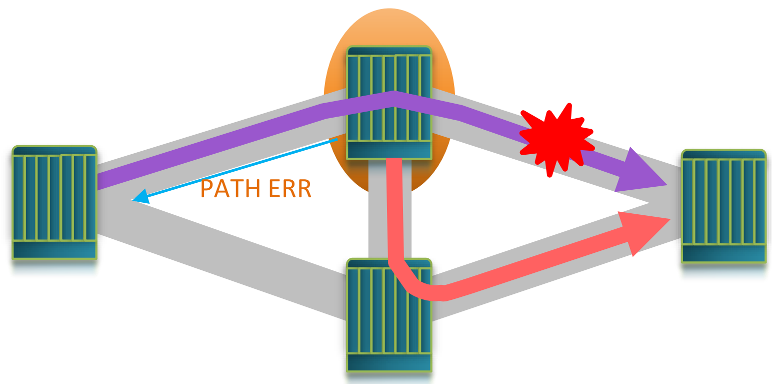
- *DiffServ-TE (RFC 3564) provides bandwidth reservations for traffic classes*
  - provides interworking between FC (Forwarding Class) QOS and MPLS TE
  - allows forwarding packets from different service classes to specific LSPs based on traffic class

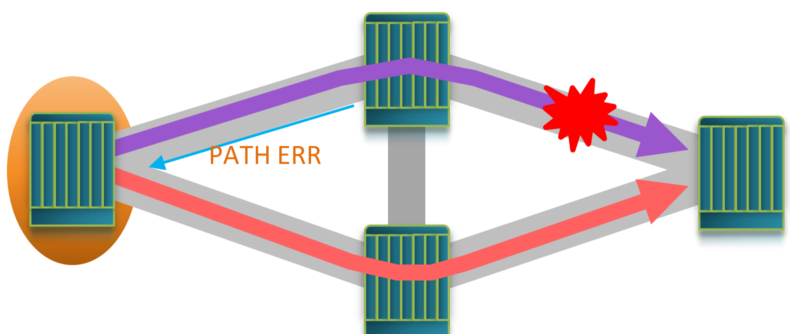
- *LSP may be assigned a traffic class (depending on service type)*
- *Using RSVP-TE the bandwidth can be reserved based on the TE-Class – a combination of traffic class and LSP priority*



MPLS – failure resiliency

- *factors influencing quality of protection*
  - 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)



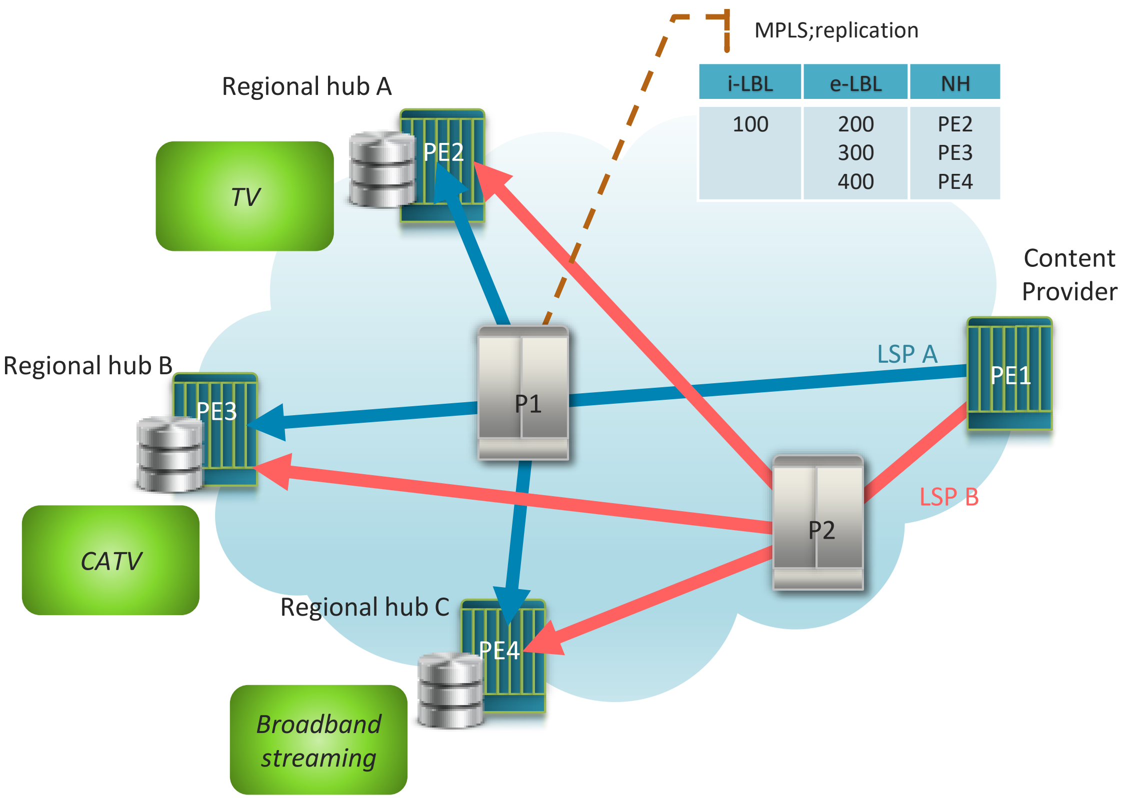


- *Fast Reroute*
  - local switchover (<50 ms)
  - PATH ERR conveys only the information about failure
  - detour paths are calculated automatically (CSPF)
  - protects against node and link failures
  - protection types:
    - One-to-One Backup (Detour)
    - Facility Backup (Backup Tunnel)

- *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

MPLS – Point-to-Multipoint LSPs

- *Point-to-Multipoint LSPs*
  - MPLS LSP with multiple endpoints
  - PE1 receives IP video stream and encapsulates it into an uni-directional P2MP LSP
  - P routers are responsible for stream replication
- *Advanced MPLS features can be used*
  - QOS/TE aware routing
  - RSVP CAC and bandwidth reservation
  - path resiliency
  - control over stream receivers
- *Additional protection level possible – two copies of the stream over disjoint LSP P2MP paths*
  - in video transport, target Head-End may perform „stream conditioning"



MPLS – P2MP LSP – RSVP signalling

- *P2MP LSP – a set of P2P LSPs (sub-LSPs)*
  - each sub-LSP is set up using separate PATH and RESV (ERO objects included in PATH messages targeted to different endpoints are different)
  - PATH and RESV – contain new object: Session Object
    - routers have to know the binding between sub-LSPs and P2MP path
    - required for proper replication in the data plane
- *LSP tree can be calculated by the source router or offline*
  - can be any tree (built under any criteria)
  - flexible solution