MPLS Introduction

Terminology

- LSR – Label Switch Router
- LER – Label Edge Router
- FEC – Forwarding Equivalent Class
- LSP – Label Switched Path
- FIB – Forwarding Information Base
- LIB – Label Information Base
- LFIB – Label Forwarding Information Base
- TIB – Tag Information Base
- PHP – Penultimate Hop Popping
- LDP – Label Distribution Protocol
- TDP – Tag Distribution Protocol
- RSVP – Resource Reservation Protocol
- CR-LDP – Constrained Routing LDP

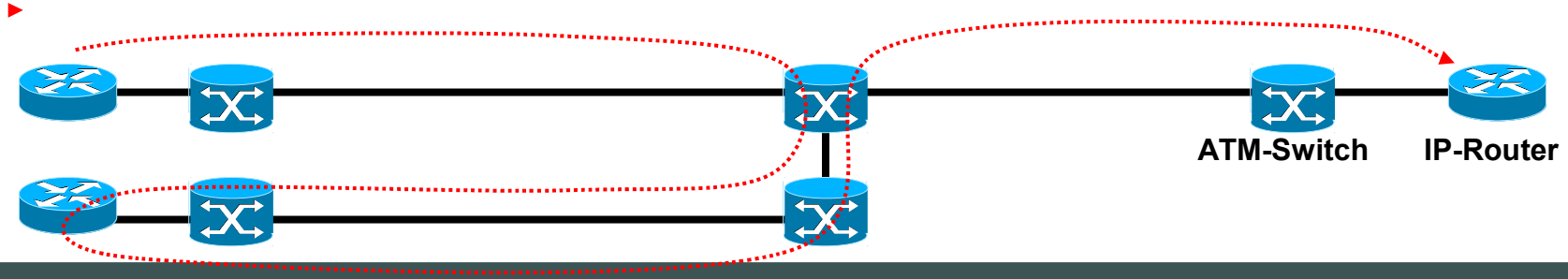
Why MPLS?

Once upon a time…

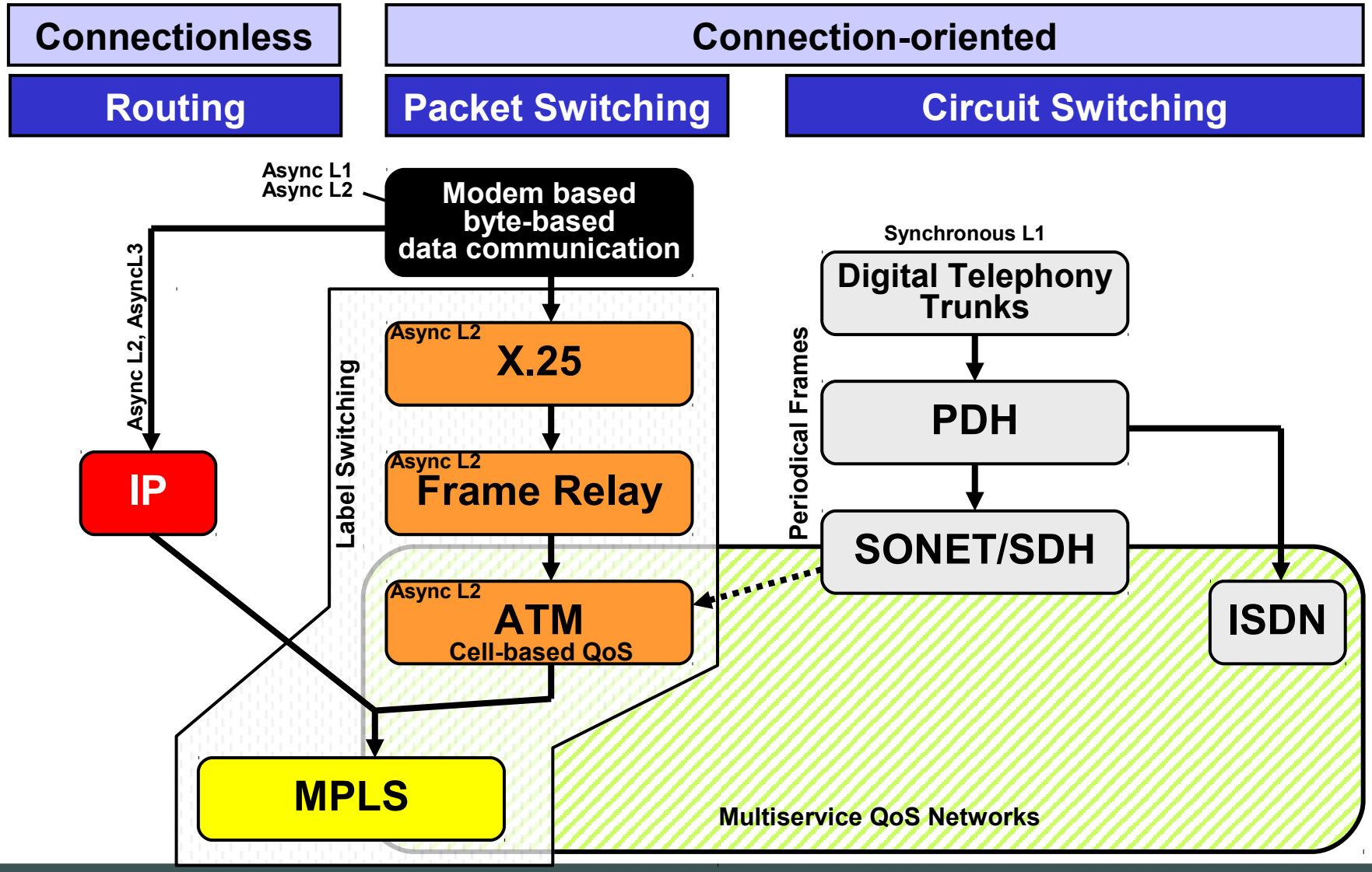
Drawbacks of IP Networks

- IP uses *structured* addresses for both:
  - Routing
  - Forwarding
- In other words: The "IP Routing Paradigm "
  - Hop-by-hop routing (slow)
  - Destination based routing (Large routing tables)
  - Least cost routing (no load balancing)
- ATM: Layer 2 and 3 topologies often different (hub & spoke)
  - Manual VC establishment necessary

TE?
QoS?
VPN?
Transport?

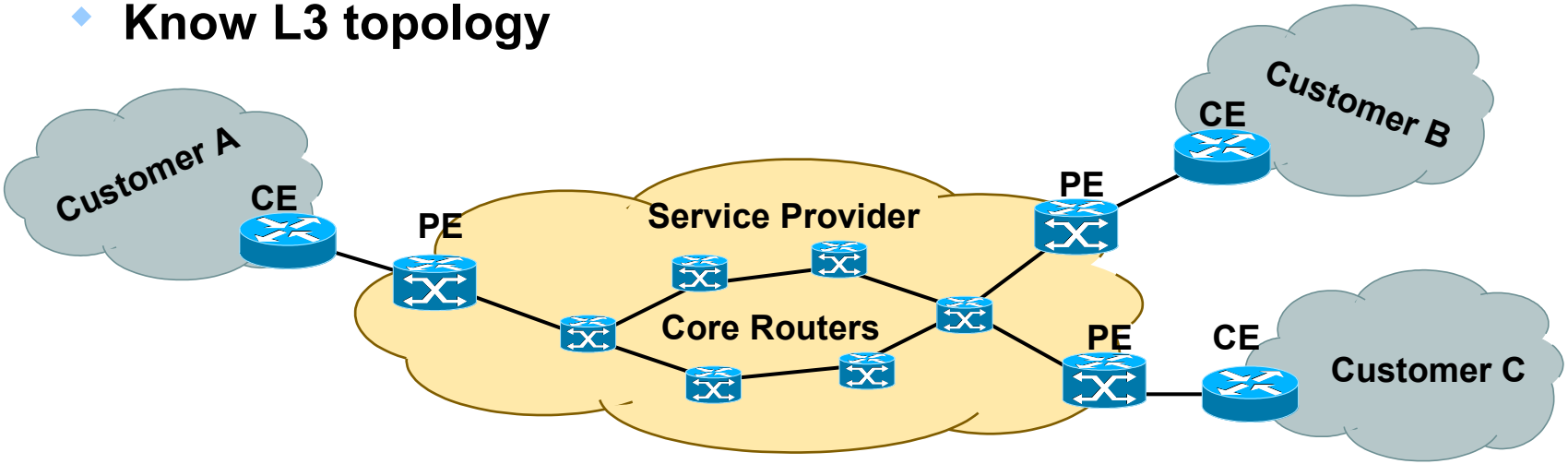


Networking Evolution

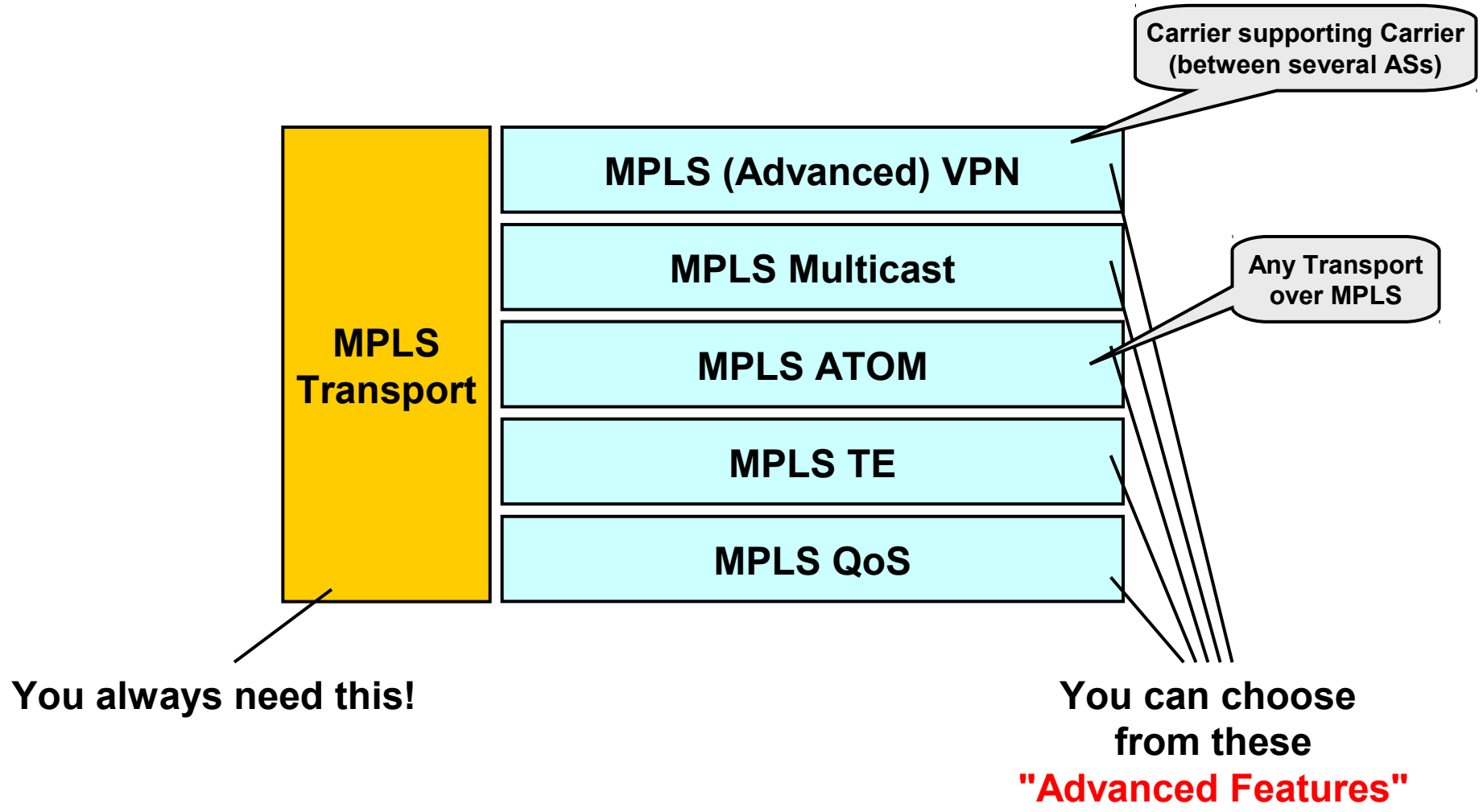


MPLS Idea

- MPLS is a provider technology
  - Application: Transport network!
- Inside versus border versus outside domains:
  - Core routers
  - Provider Edge routers (PE-routers)
  - Customer Edge routers (CE-routers)
- Also ATM switches can run MPLS
  - Know L3 topology



MPLS Building Blocks



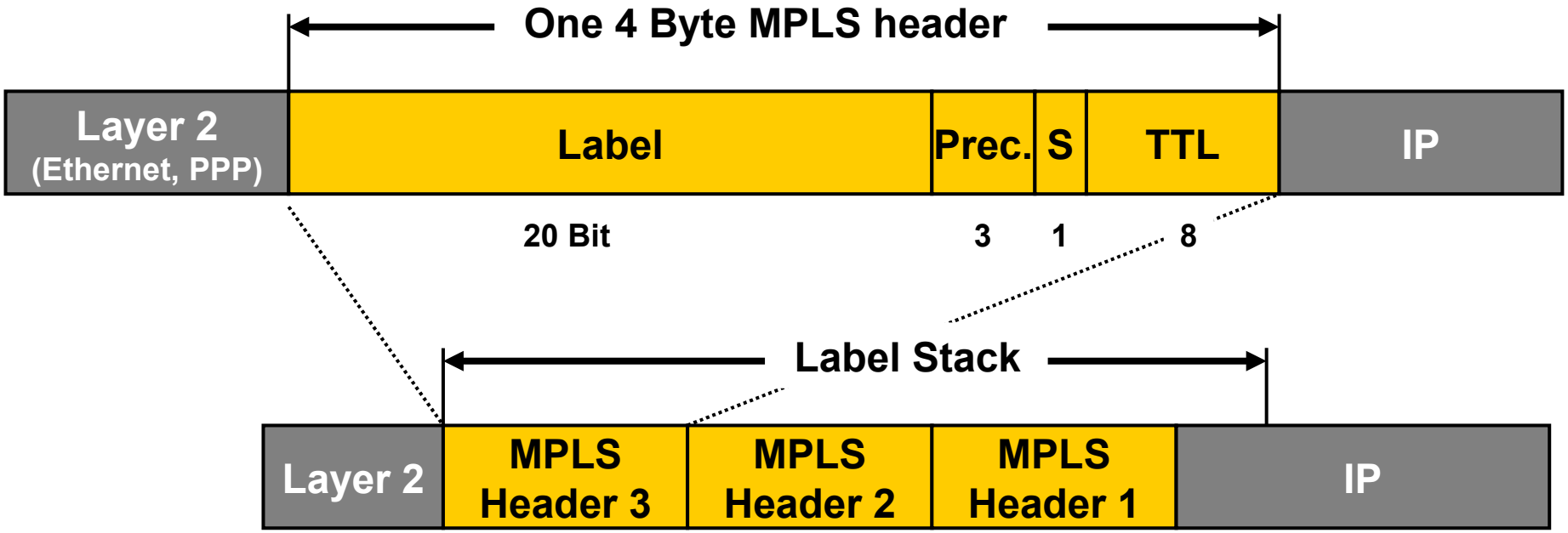
MPLS Transport

The most fundamental feature…

MPLS at a Glance

- IP does destination based routing
  - Hop-by-hop routing efforts
  - Each hop must know all routes (100,000)
- MPLS replaces the global IP destination address by a locally used *label*
- Label can identify many things: FEC
  - VPN-ID, TE Tunnels, QoS ,
    Multicast groups, ...

MPLS Header

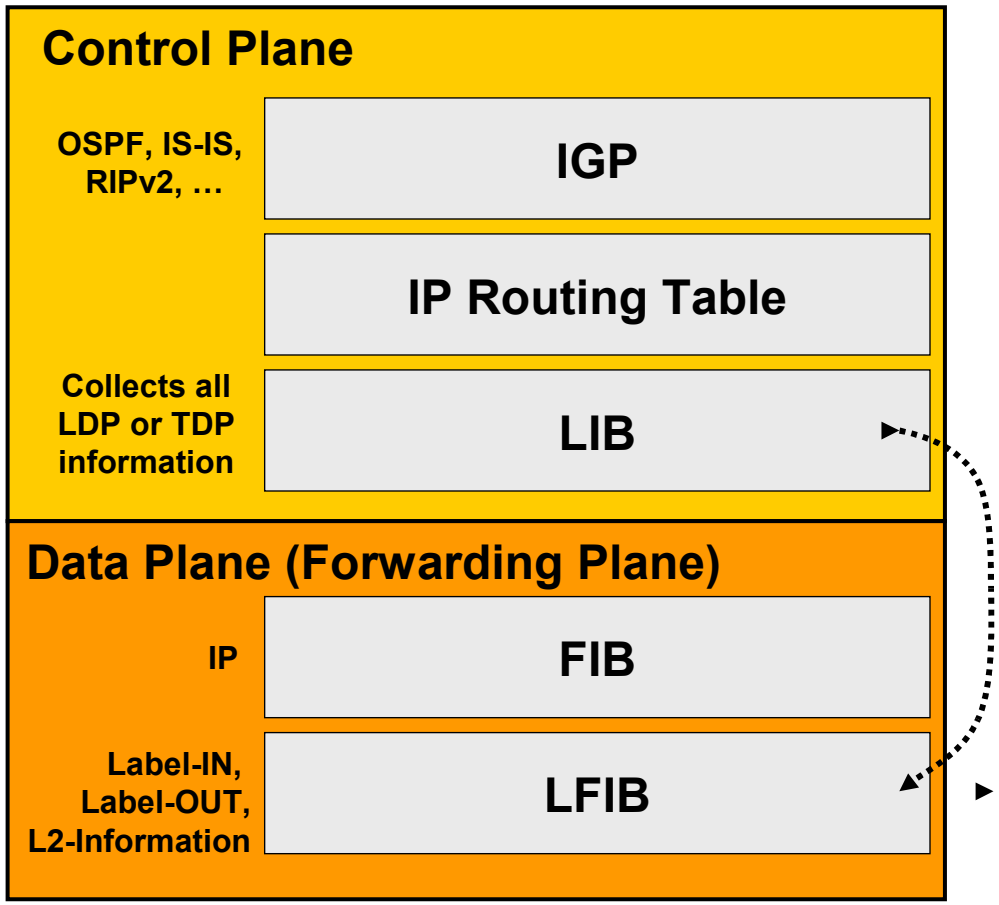


- "Layer 2.5" can be used over Ethernet, 802.3 or PPP links
  - Frame mode
- MPLS over ATM is different than over packet interface
  - Cell mode
  - ATM can only swap VPI/VCI, no stacking!
  - ATM encapsulates MPLS-IP packet inside AAL5

Label Switch Routers (LSRs)

- Any Cisco IOS 12.0 based router can do MPLS
- Performs standard operations:
  - Insert (impose) a label
  - Remove (pop) a label
  - Swap labels during forwarding
- Multiple labels occur for example:
  - MPLS VPNs (egress router/VPN)
  - MPLS TE (tunnel/destination)

Important Concepts

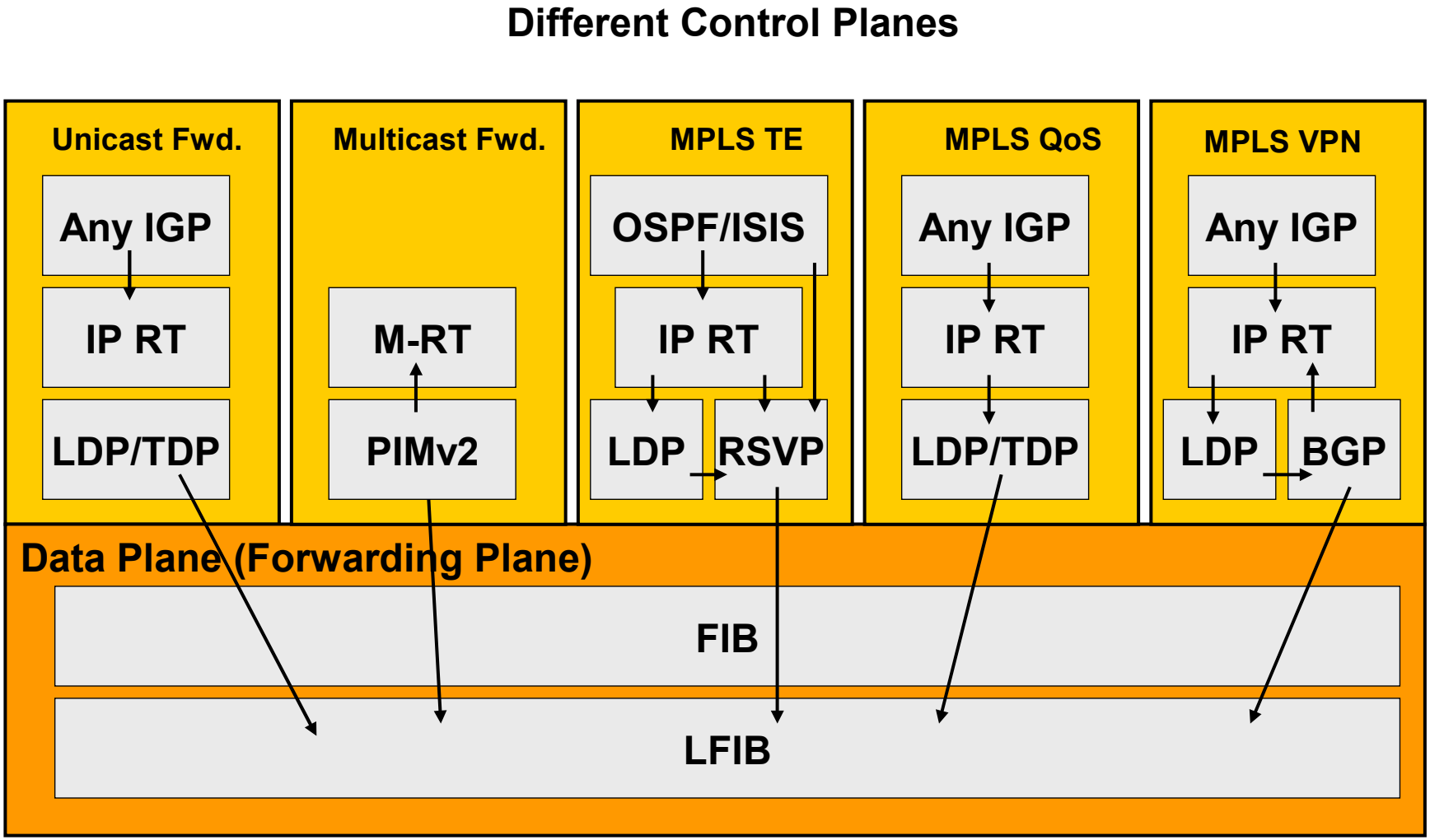


- LDP (RFC) or TDP (Cisco)
- CEF is required (Cisco Patent)
  - Routing table is 256-way "mtrie"
  - Better than Fast Switching: Also 1st Packet fast!
  - DCEF = per interface
- MPLS applications only differ in the usage of the control plane
  - VPN, TE, QoS, ...
  - All use data plane equivalently

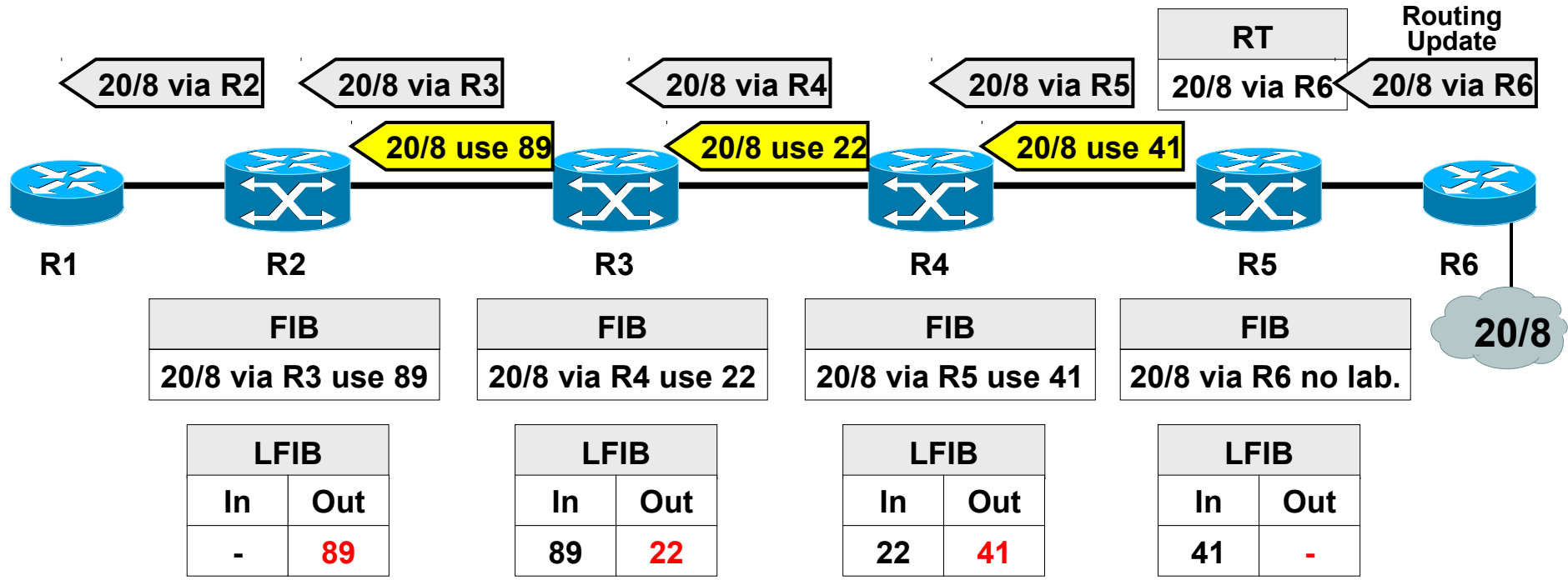
Important Databases

- FIB
  - This is the CEF database
  - Contains L2/L3 headers, IP addresses, labels, next hop, metric
  - The routing table is only a subset of the FIB
- LIB
  - Contains *all* labels and associated destinations
- LFIB
  - Contains selected labels used for forwarding
  - Selection based on FIB

MPLS Applications

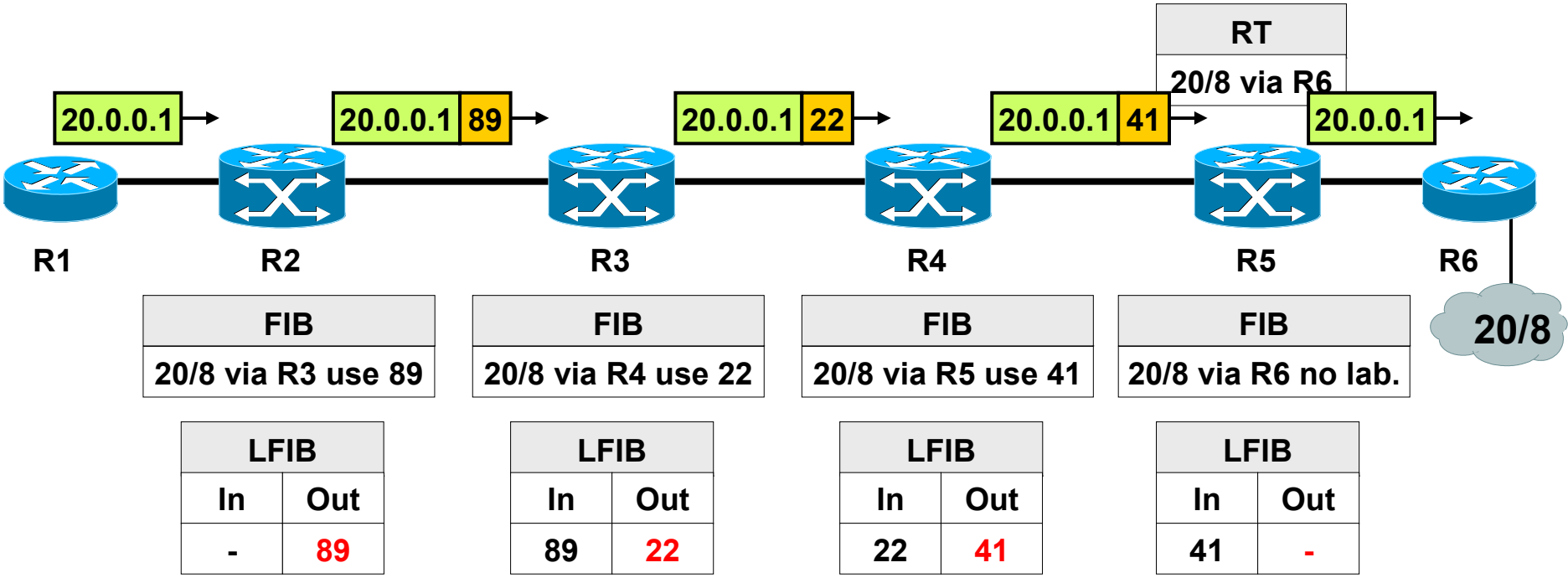


Label Switching (1)



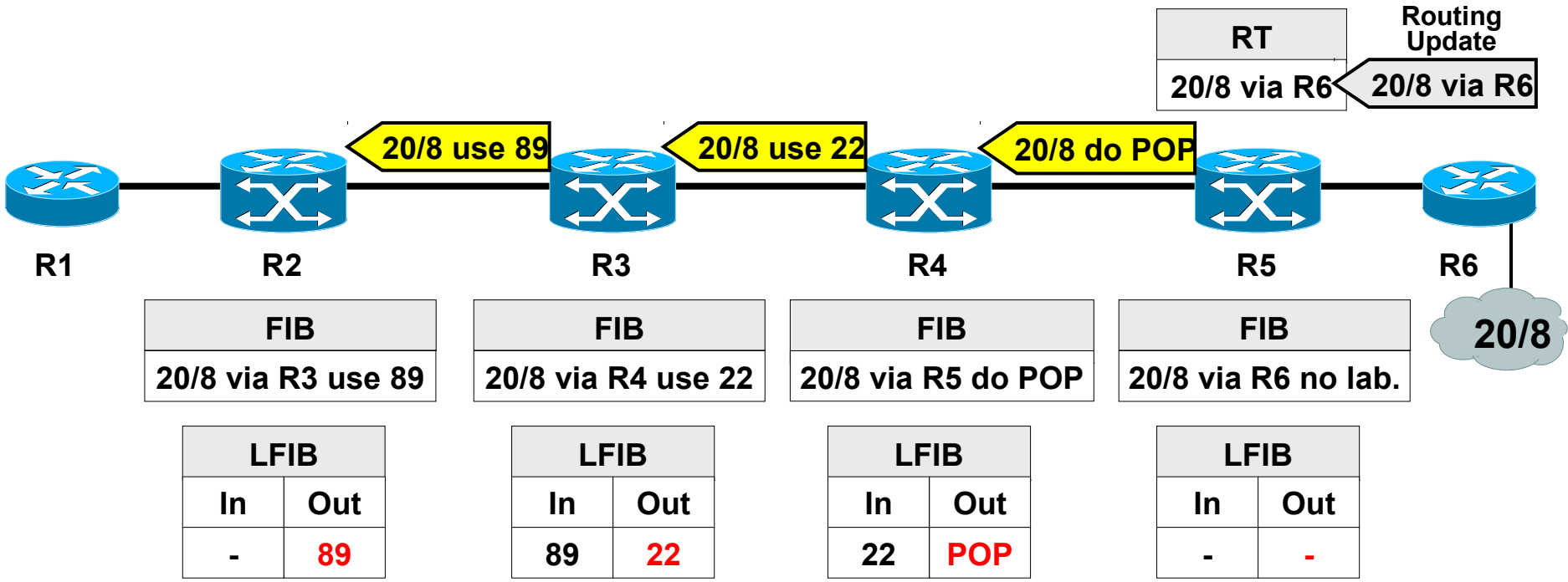
- Both routing updates and LDP/TDP distribute reachability information

Label Switching (2)



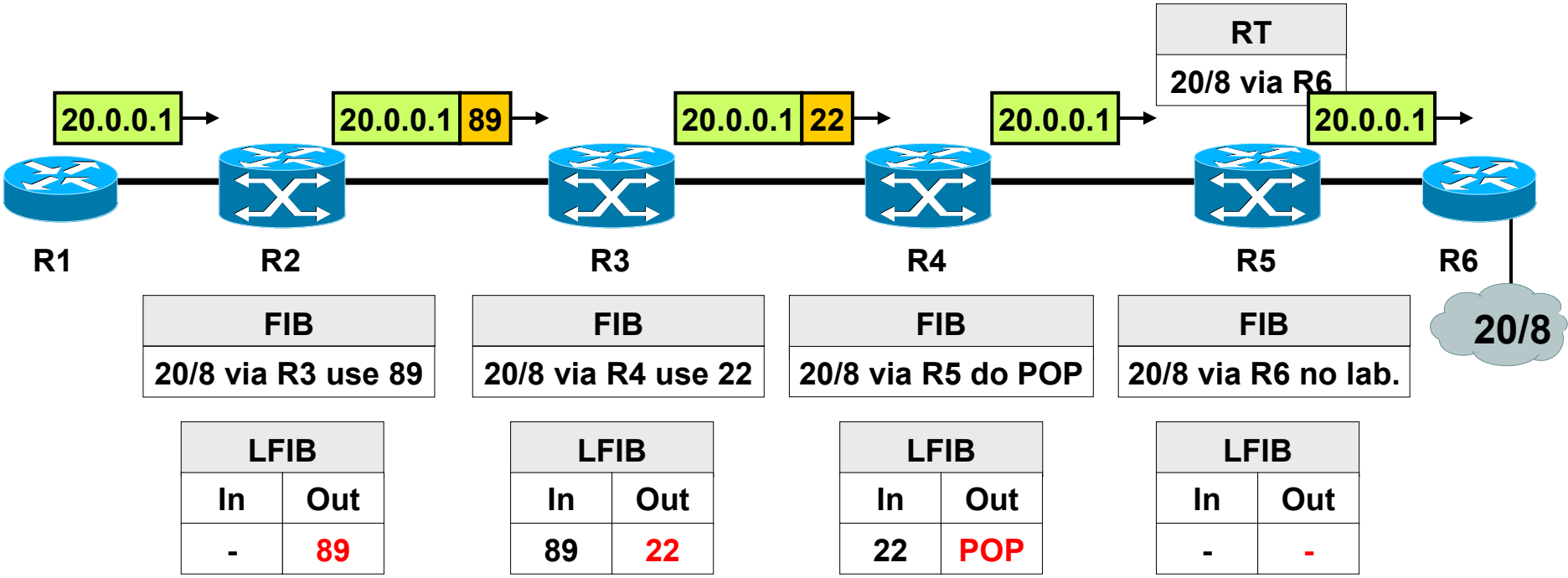
- R5 must perform double lookup:
  - LFIB tells "remove the label"
  - FIB tells "use next hop R6"
- Label should be removed one hop earlier (by R4) !!!!

PHP (1)



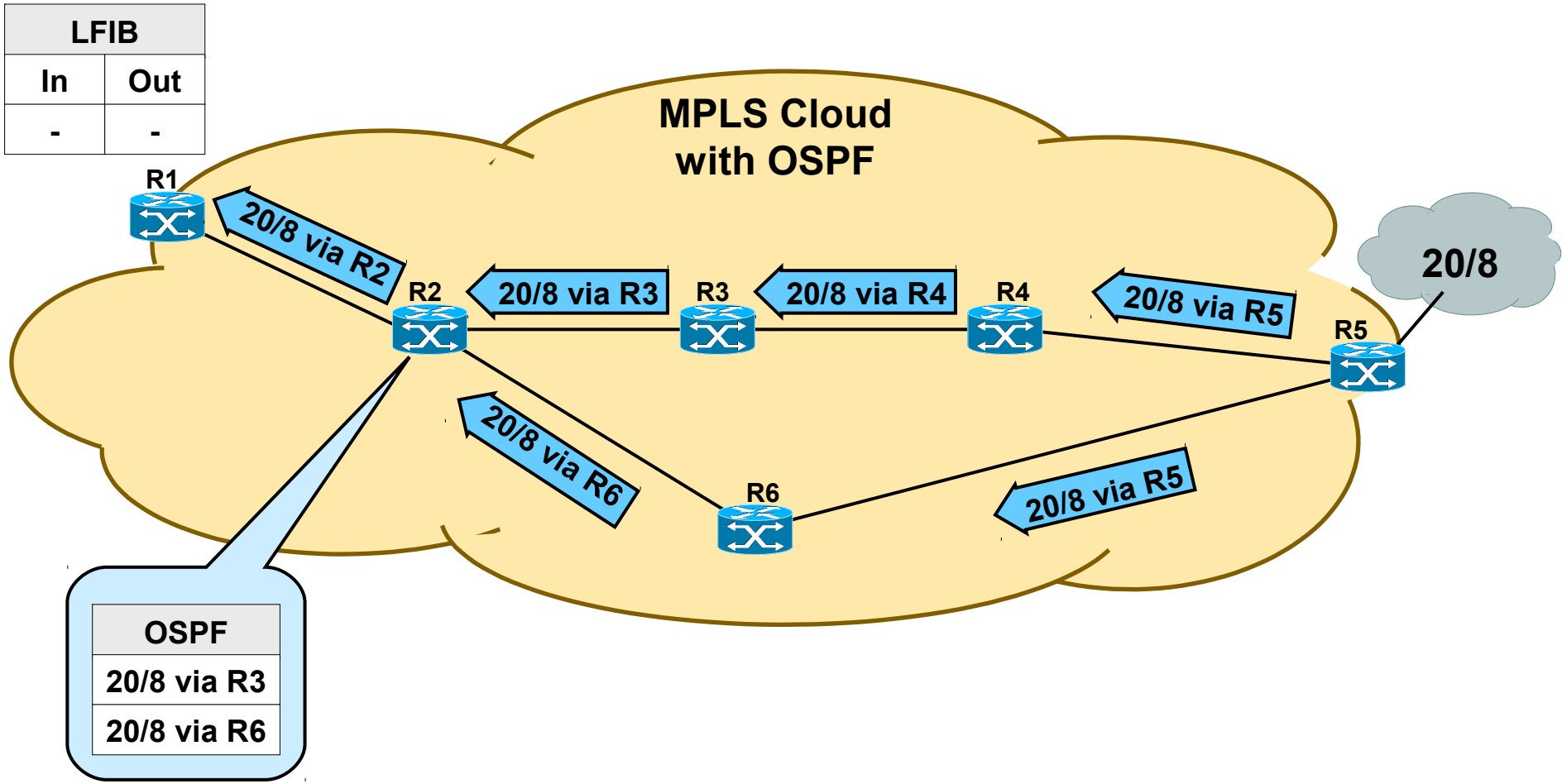
- Last hop router (R5) tells penultimate router (R4) to remove label
  - "Penultimate Hop Popping" (PHP)
  - Also called "Implicit Null Label"

PHP (2)

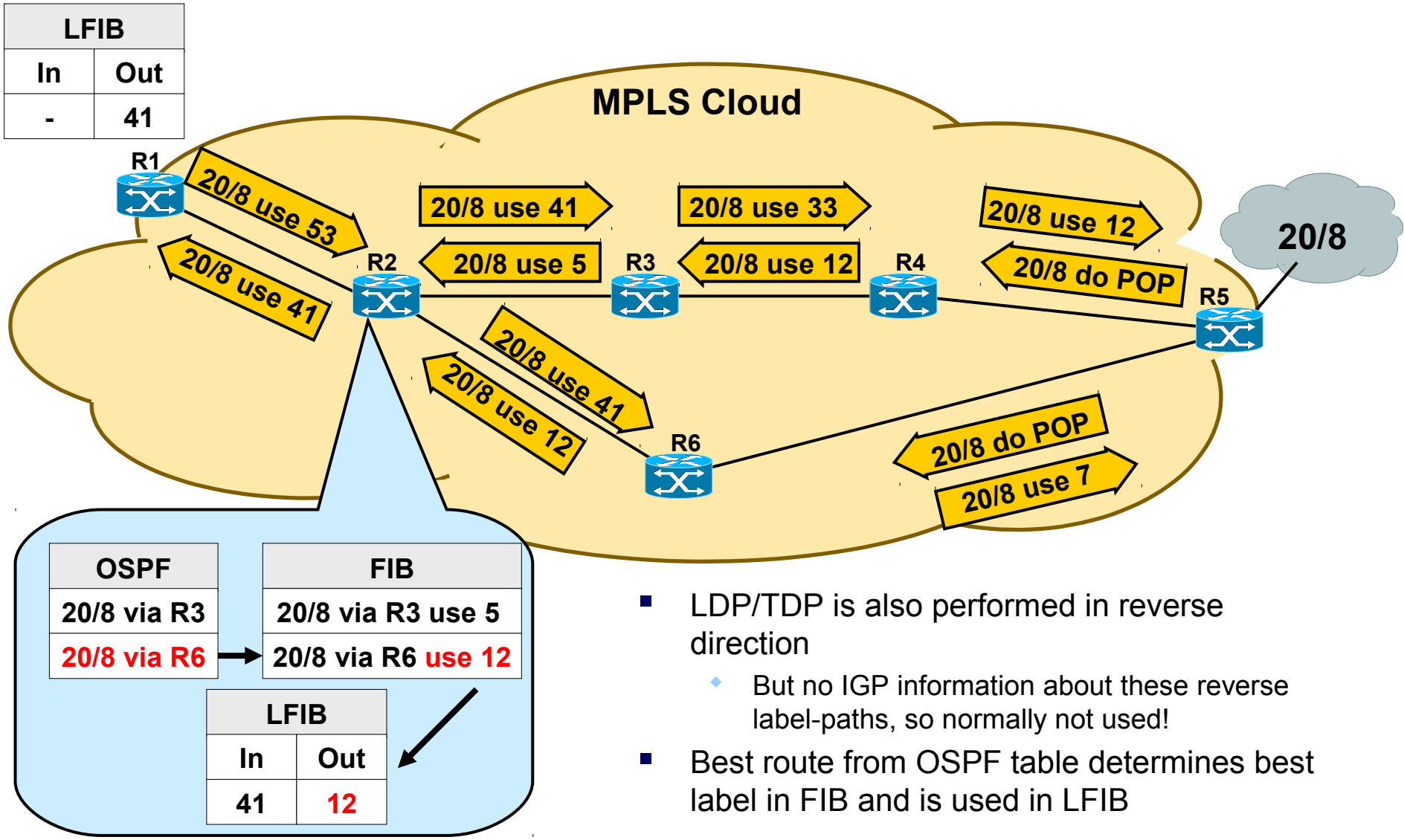


- R5 only performs single lookup in FIB
- Note: PHP does not work with ATM
  - VPI/VCI cannot be removed

1 – Routing Updates



2 – LDP or TDP



- LDP/TDP is also performed in reverse direction
  - But no IGP information about these reverse label-paths, so normally not used!
- Best route from OSPF table determines best label in FIB and is used in LFIB

Example LIB

```
Router# show tag-switching tdp bindings
    tib entry: 10.0.0.1/32, rev 9
        local binding:       tag: 41
        remote binding:      tsr: 10.0.0.3:0, tag: 41
    tib entry: 10.0.0.2/32, rev 8
        local binding:       tag: 40
        remote binding:      tsr: 10.0.0.3:0, tag: 40
    tib entry: 10.0.0.3/32, rev 7
        local binding:       tag: 39
        remote binding:      tsr: 10.0.0.3:0, tag: imp-null(1)
    tib entry: 10.0.0.9/32, rev 6
        local binding:       tag: imp-null(1)
        remote binding:      tsr: 10.0.0.3:0, tag: 39
```

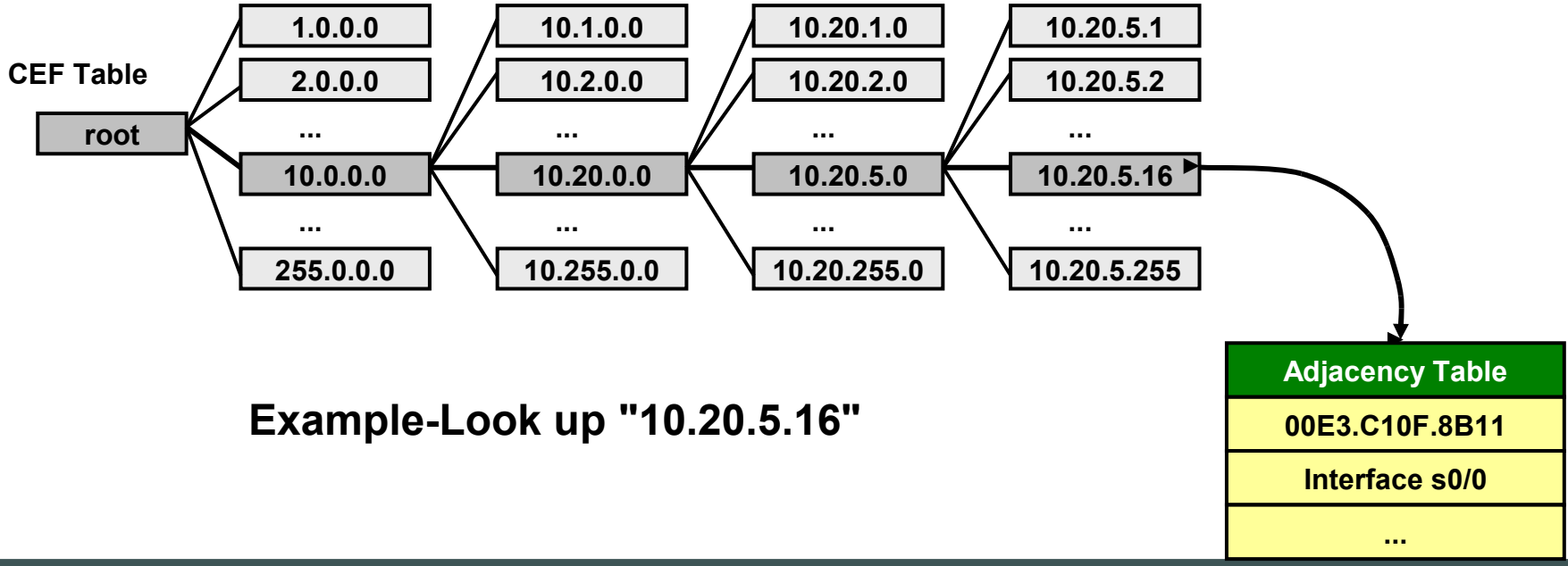
- Contains all information learned by LDP or TDP
- Best labels are copied into FIB/LFIB

Cisco Express Forwarding (CEF)

- Requirement for MPLS
  - Forwarding information (L2-headers, addresses, labels) are maintained in FIB for each destination
  - Newest and fastest IOS switching method
  - Critical in environments with frequent route changes and large RTs: The Internet backbone!
- Invented to overcome Fast Switching problems:
  - No overlapping cache entries
  - Any change of RT or ARP cache invalidates route cache
  - First packet is always process-switched to build route cache entry
  - Inefficient load balancing when "many hosts to one server"

How CEF Works

- CEF "Fast Cache" consists of
  - CEF table: Stripped-down version of the RT (256-mtrie)
  - Adjacency table: Actual forwarding information (MAC, interfaces, ...)
- CEF cache is pre-built before any packets are switched
  - No packet needs to be process switched
- CEF entries never age out
  - Any RT or ARP changes are immediately mapped into CEF cache



Example-Look up "10.20.5.16"

Example FIB ("CEF-Table")

```
Router# show ip cef 10.0.0.0 detail
10.0.0.0/8, version 12, cached adjacency to Serial0/0.3
0 packets, 0 bytes
  tag information set
    local tag: 14
    fast tag rewrite with Se0/0.3, point2point, tags imposed: {15}
  via 10.0.0.3, Serial0/0.3, 0 dependencies
    next hop 10.0.0.3, Serial0/0.3
    valid cached adjacency
    tag rewrite with Se0/0.3, point2point, tags imposed: {15}
```

- Best labels are copied into LFIB

Example LFIB

```
Router# show tag-switching forwarding-table detail
Local    Outgoing    Prefix          Bytes tag   Outgoing      Next Hop
tag      tag or VC   or Tunnel Id    switched    interface
35       Untagged    10.0.0.5/32     0           Se0/0.2       point2point
     MAC/Encaps=0/0, MTU=1500, Tag Stack{}
36       Pop tag     10.0.0.6/32     0           Se1/0.3       point2point
     MAC/Encaps=4/4, MTU=1500, Tag Stack{}
     1A31F422
37       39          10.0.0.7/32     0           Se1/0.1       point2point
     MAC/Encaps=4/8, MTU=1504, Tag Stack{39}
     80F1C300 00027000
```

- Label-to-interface mapping
- Synonym with: `show mpls forwarding-table`

IOS Standard Behavior

- Routers with packet interfaces
  - Per-platform label space !!!
  - Unsolicited label distribution
  - Liberal label retention !
  - Independent control
- Routers with ATM interfaces
  - Per-interface label space
  - On-demand label distribution
  - Conservative or liberal label retention
  - Independent control
- ATM switches
  - Per-interface label space
  - On-demand distribution
  - Conservative label retention
  - Ordered control

TDP Key Facts

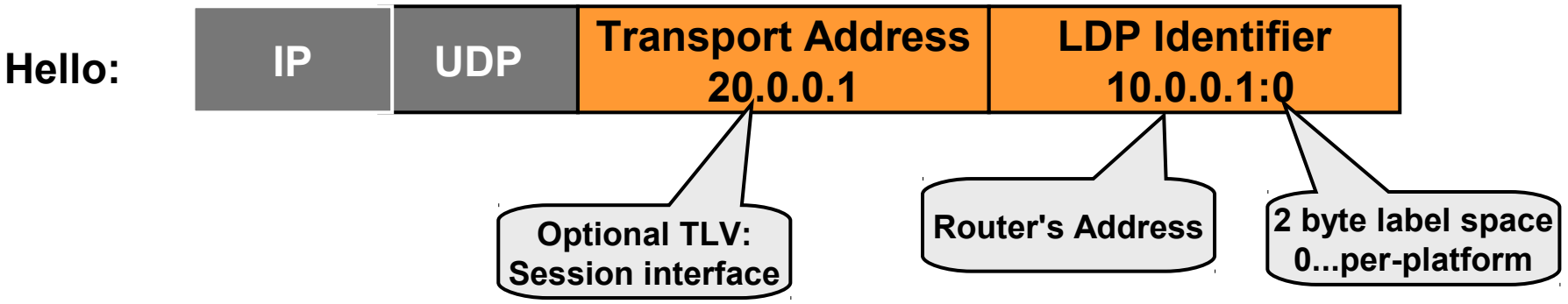
- Tag Distribution Protocol (TDP) invented by Cisco for distributing <label, prefix> bindings
  - Enabled by default
- Session establishment: UDP/TCP port 711
  - Hello messages via UDP, destination 224.0.0.2 (all subnet routers)
  - Session via TCP, incremental updates
- Not compatible with LDP
  - But can co-exist as long as two peers use same protocol

LDP Key Facts

- IETF standard, descendent of Cisco's proprietary TDP
- Same concept but port 646
  - Also to destination 224.0.0.2
- 6-byte TLV ("LDP-ID") identifies
  - Router (4 bytes)
  - Label space (2 bytes)
    - Per-platform label space is set to zero

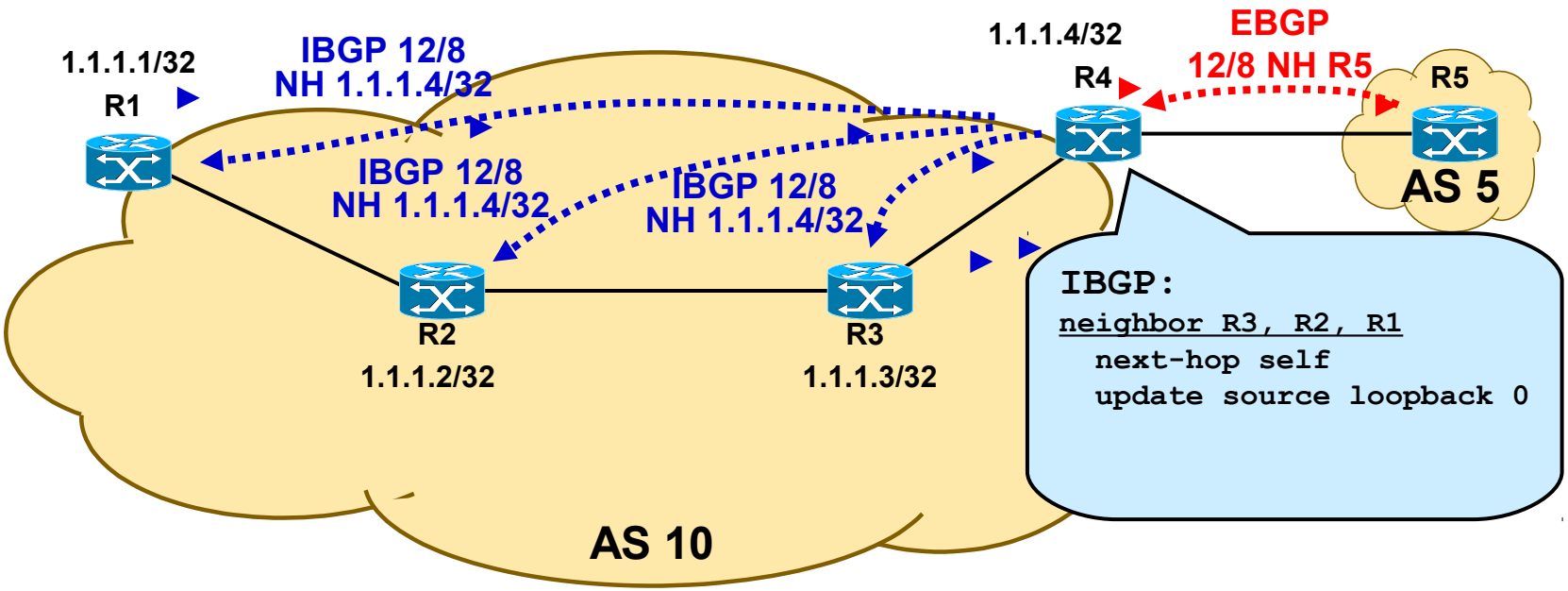
LDP Details

- One session per LDP identifier
  - Per-platform label space: 1 identifier for all links



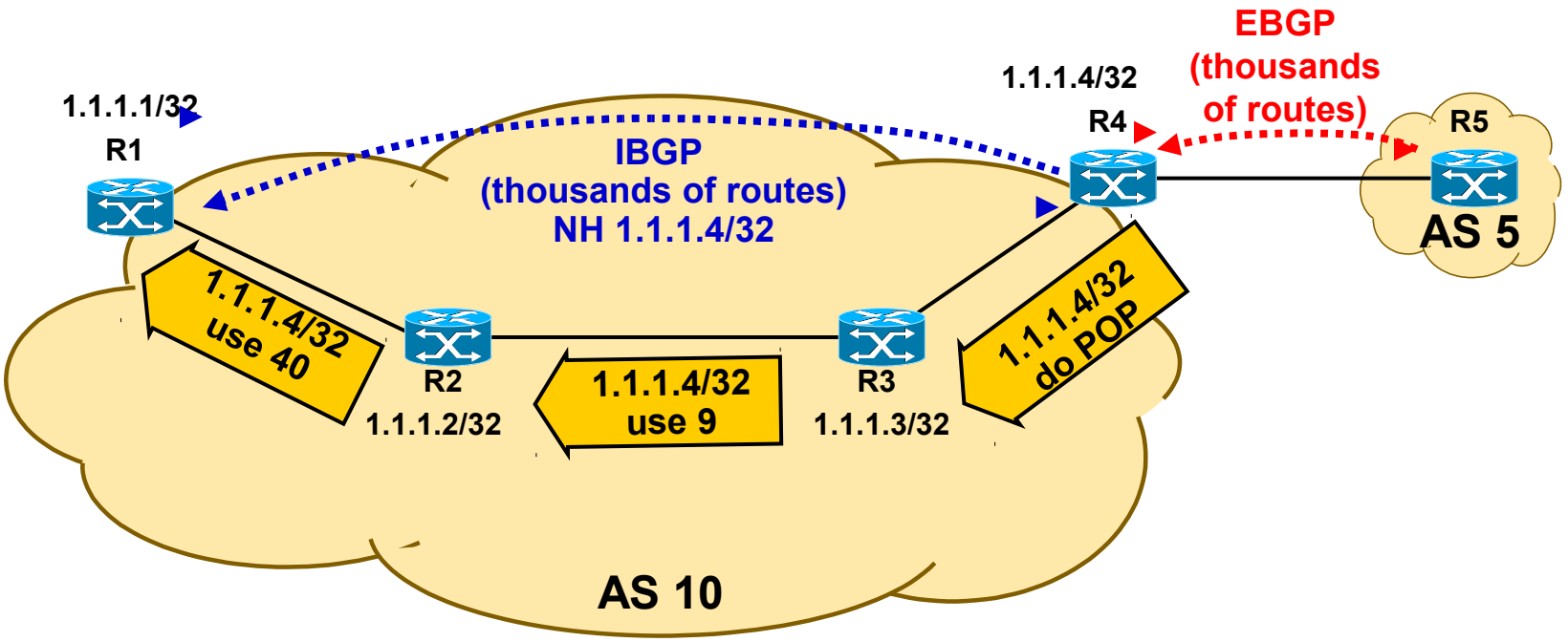
- TCP session initiated from router with highest address

BGP Standard Behavior



- Good style: Use loopback addresses and next hop self
  - BUT: Full mesh IBGP !!!
  - BUT: Each router has full routing table !!!
- IGP is used to propagate loopback addresses
  - 1.1.1.1/32, 1.1.1.2/32, 1.1.1.3/32, and 1.1.1.4/32
- Note: Sync Off
  - Otherwise IBGP routes would never be copied into the routing table
  - IBGP updates would only be propagated by PE-router if this network is reachable via IGP

MPLS and BGP



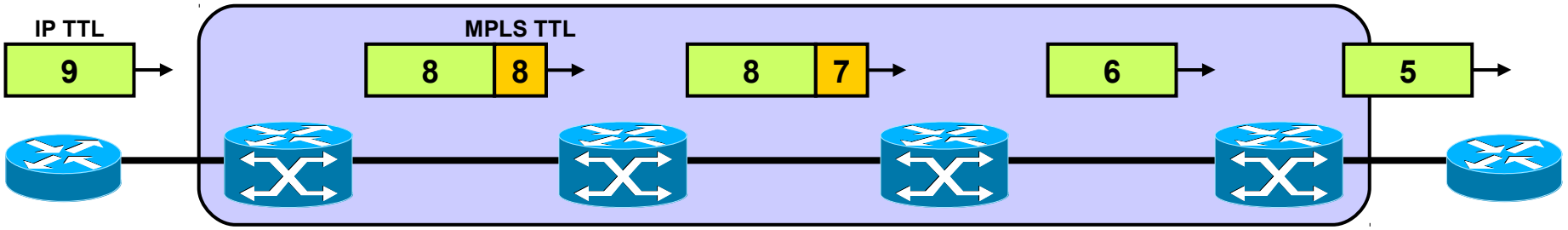
- FEC = Next Hop
  - Only PE routers must learn all external routes
  - Only the PE routers must be powerful
- IBGP sessions only between PE-routers

Note

- LSRs announce only one label (per destination) to adjacent LSRs
  - Even if there are parallel links between them
  - Insecure: Any neighbor can abuse label!
- After a link failure
  - All labels (and related information) are removed from the FIB/LFIB/LIB
  - After routing convergence FIB (RT) knows another path
  - New label is provided by LIB
- When broken link comes back again
  - LIB had already lost the label
  - Path broken!

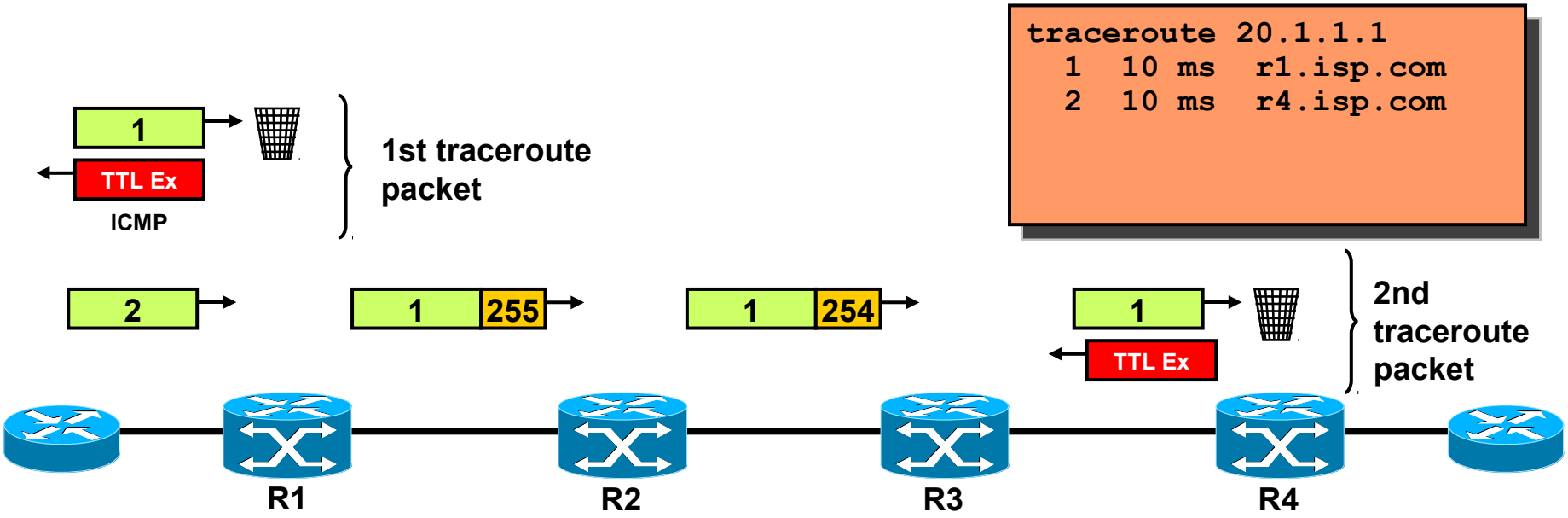
Normal TTL Usage

- Loop detection
  - LDP and TDP basically rely on IGP loop detection
  - Additionally a TTL field in the MPLS header prevents endless routing
- TTL Propagation: IP TTL is copied into MPLS header
  - Enabled by default on Cisco routers



Disable TTL Propagation

- No TTL copying between IP and MPLS header
- Ingress router assigns MPLS TTL 255
- Core routers are hidden
  - E. g. traceroute fails to show them



MPLS VPN

Where the complexity begins...

Two Major VPN Paradigms

- Overlay VPNs: Transparent P2P links
  - Well-known technology
  - Provider does not care about customer routing
  - Best customer isolation
- Peer VPNs: Participation in C-routing
  - Optimum routing
  - Simple provision of additional VPN
  - Problems with address space

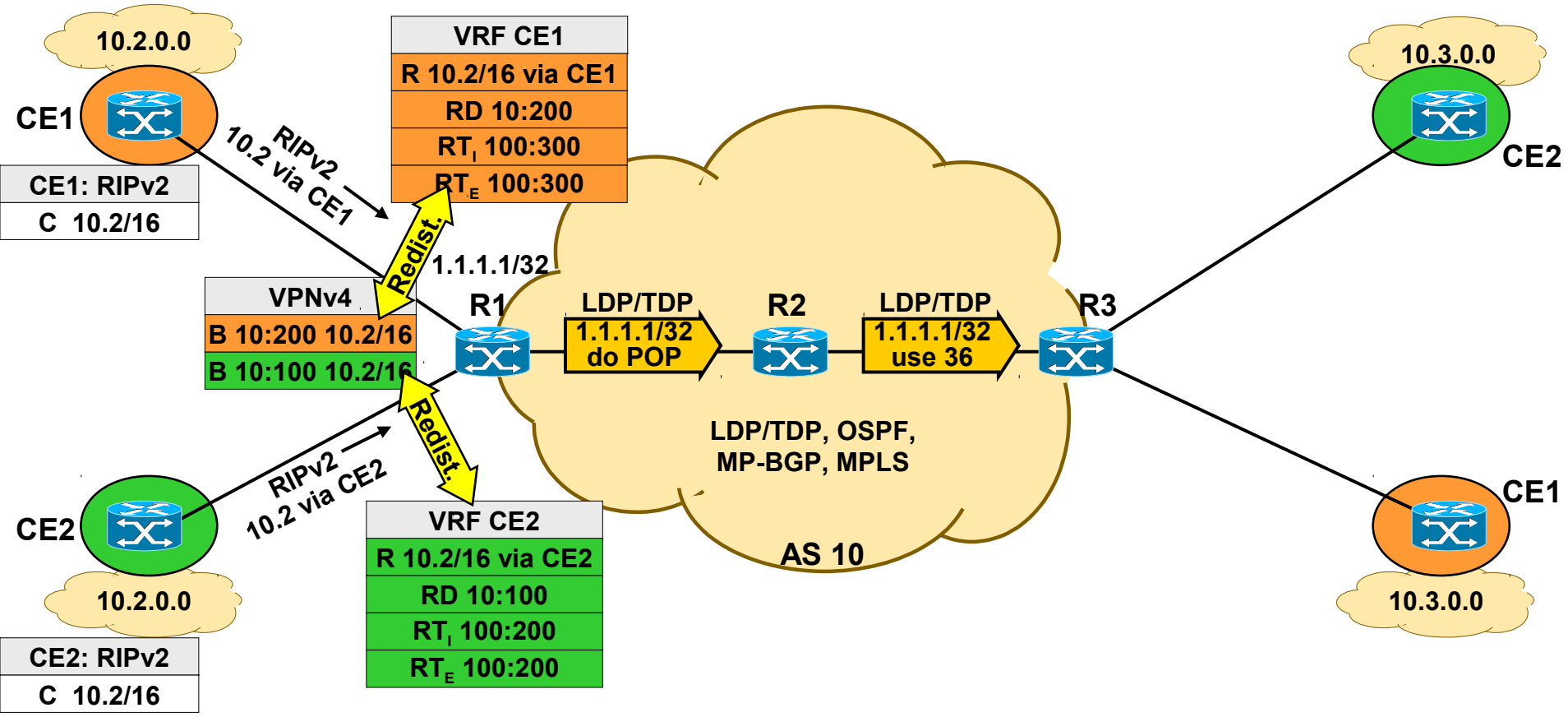
MPLS VPN – Best of Both Worlds

- PE routers participate in C-routing
  - Hence optimum routing between sites
  - Easy provisioning (sites only)
- PE routers allow route isolation
  - By using Virtual Routing and Forwarding Tables (VRF)
  - Allows overlapping address spaces
- Overlapping VPNs possible
  - By a simple (?) attribute syntax

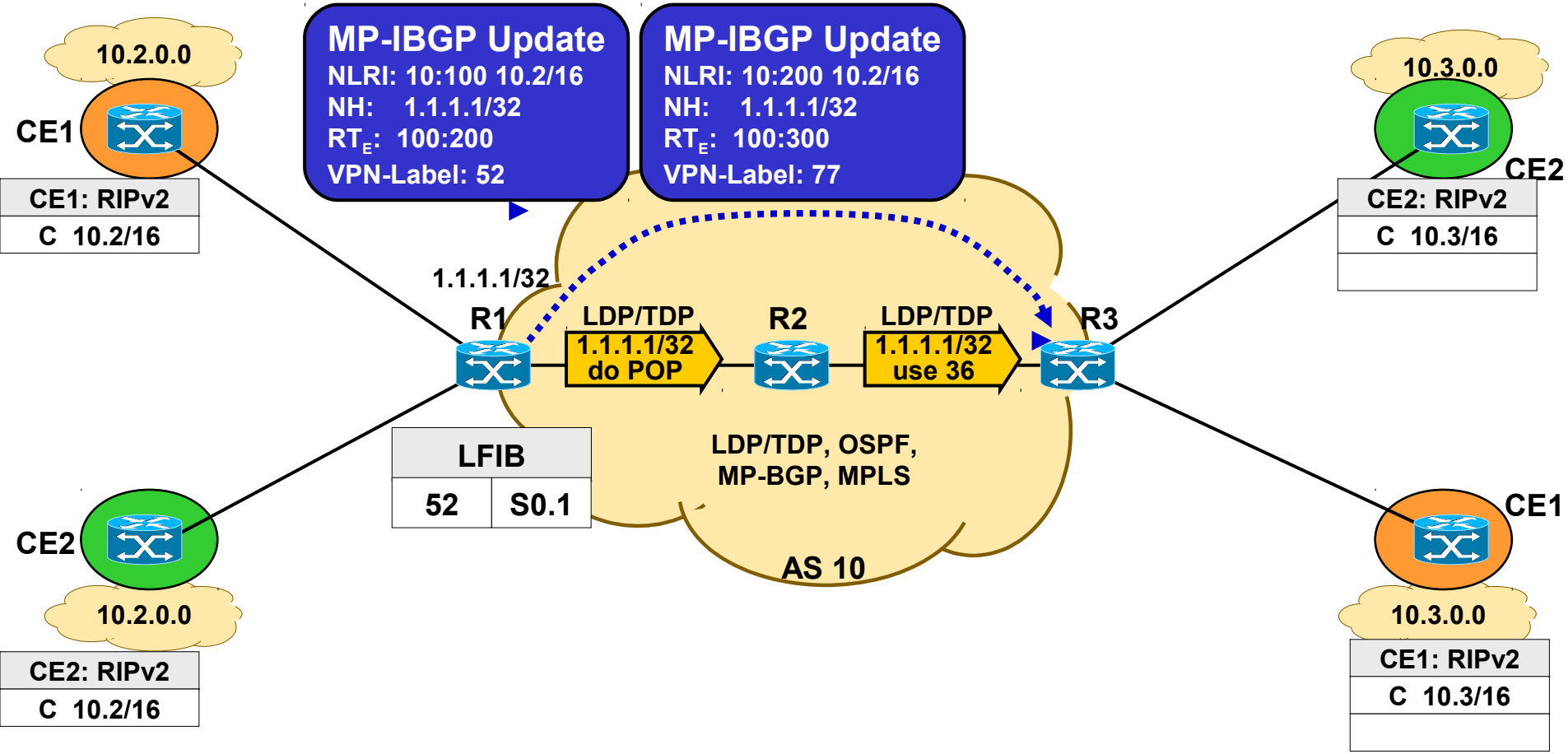
MPLS VPN – Principles

- Requires MPLS Transport
- Requires MP-BGP
  - Supports IPv4/v6, VPNv4, multicast
  - Default behavior: BGP-4
- VPNv4 uses 96 bit addresses
  - 64 bit Route Distinguisher (RD)
  - 32 bit IP address
- Every router uses one VRF for each VPN
  - Virtual Routing and Forwarding Table (VRF)

MPLS VPN

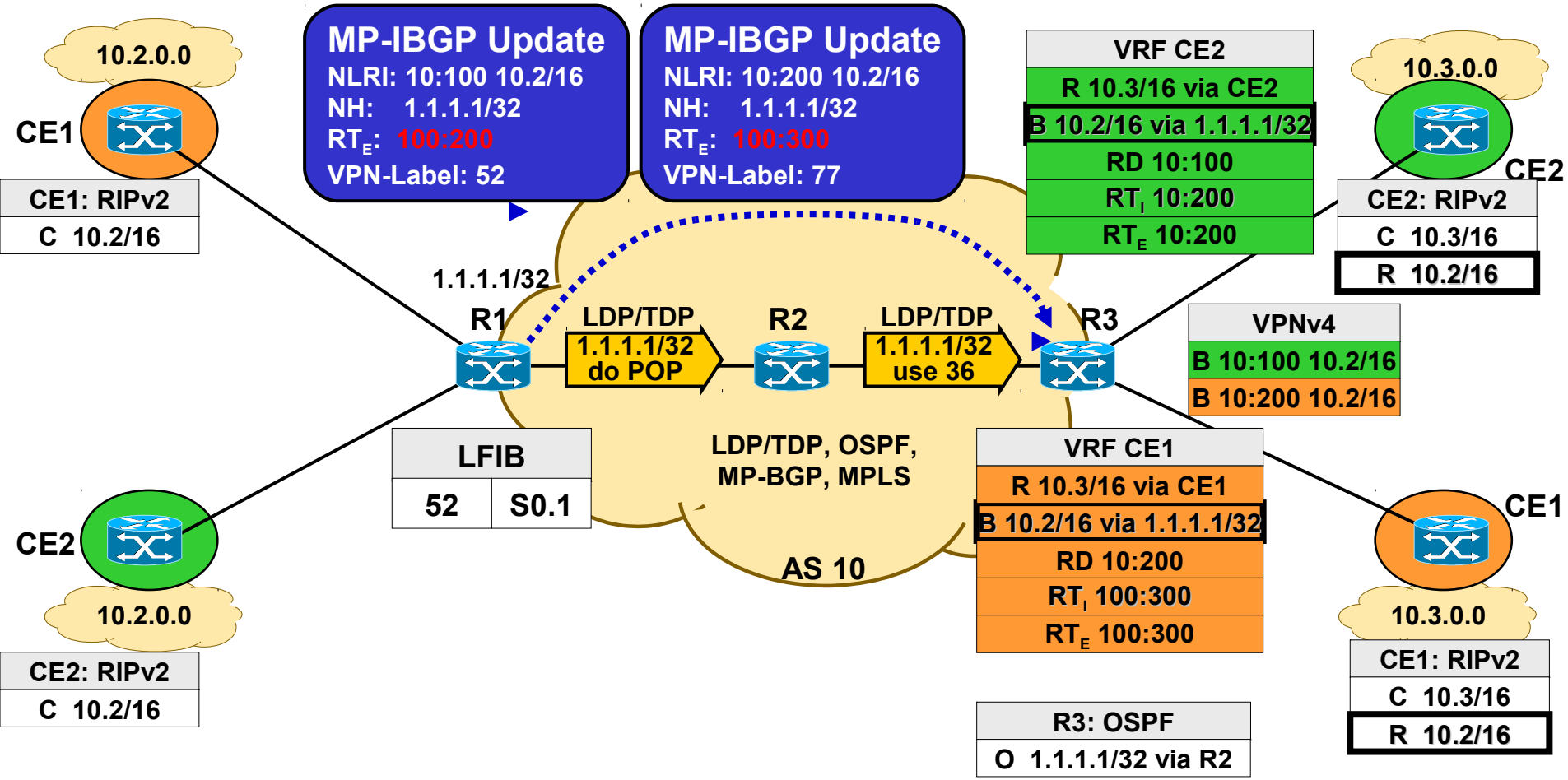

10.2.0.0
CE1
CE1: RIPv2
C 10.2/16
RIPv2
10.2 via CE1
VRF CE1
R 10.2/16 via CE1
RD 10:200
RT_I 100:300
RT_E 100:300
Redist.
1.1.1.1/32
VPNv4
B 10:200 10.2/16
B 10:100 10.2/16
R1
LDP/TDP
1.1.1.1/32
do POP
R2
LDP/TDP
1.1.1.1/32
use 36
R3
10.3.0.0
CE2
LDP/TDP, OSPF,
MP-BGP, MPLS
AS 10
Redist.
RIPv2
10.2 via CE2
CE2
10.2.0.0
CE2: RIPv2
C 10.2/16
VRF CE2
R 10.2/16 via CE2
RD 10:100
RT_I 100:200
RT_E 100:200
CE1
10.3.0.0

MPLS VPN



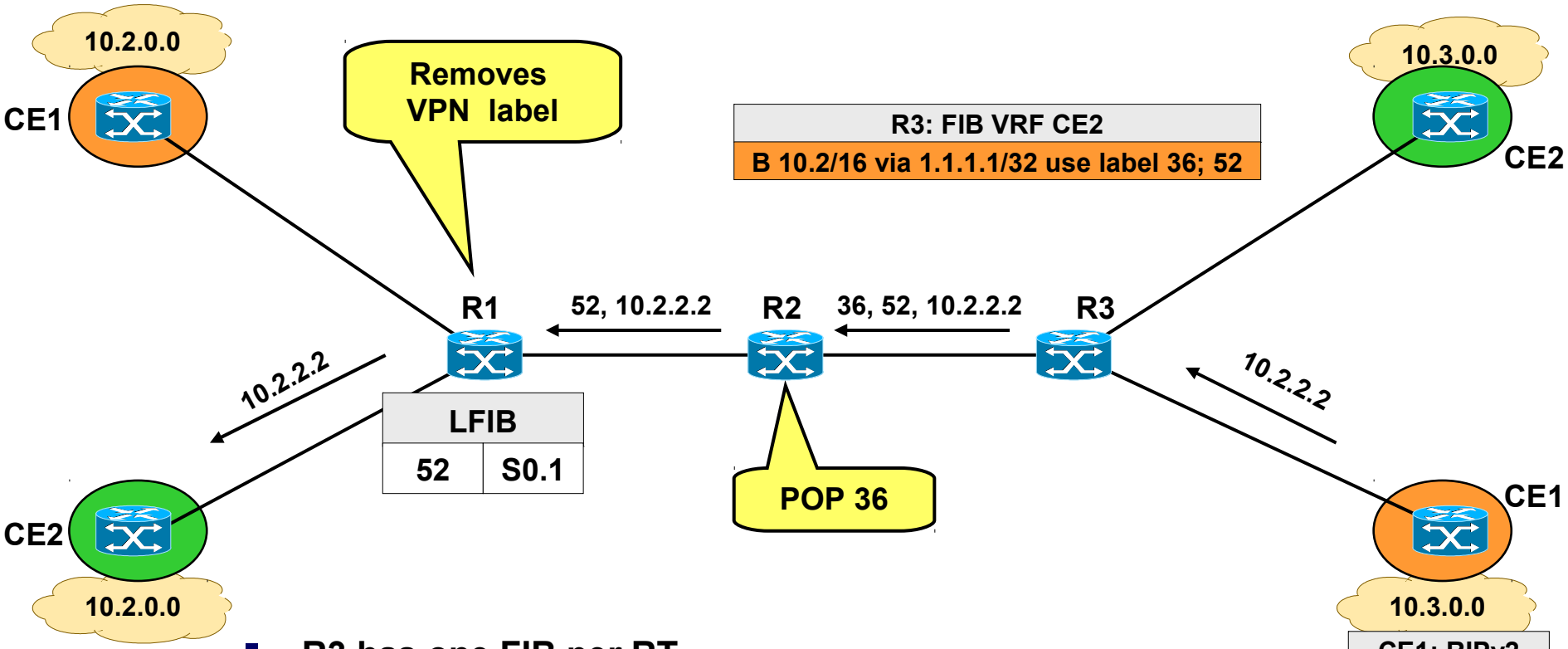
IGP Metric → MED

MPLS VPN



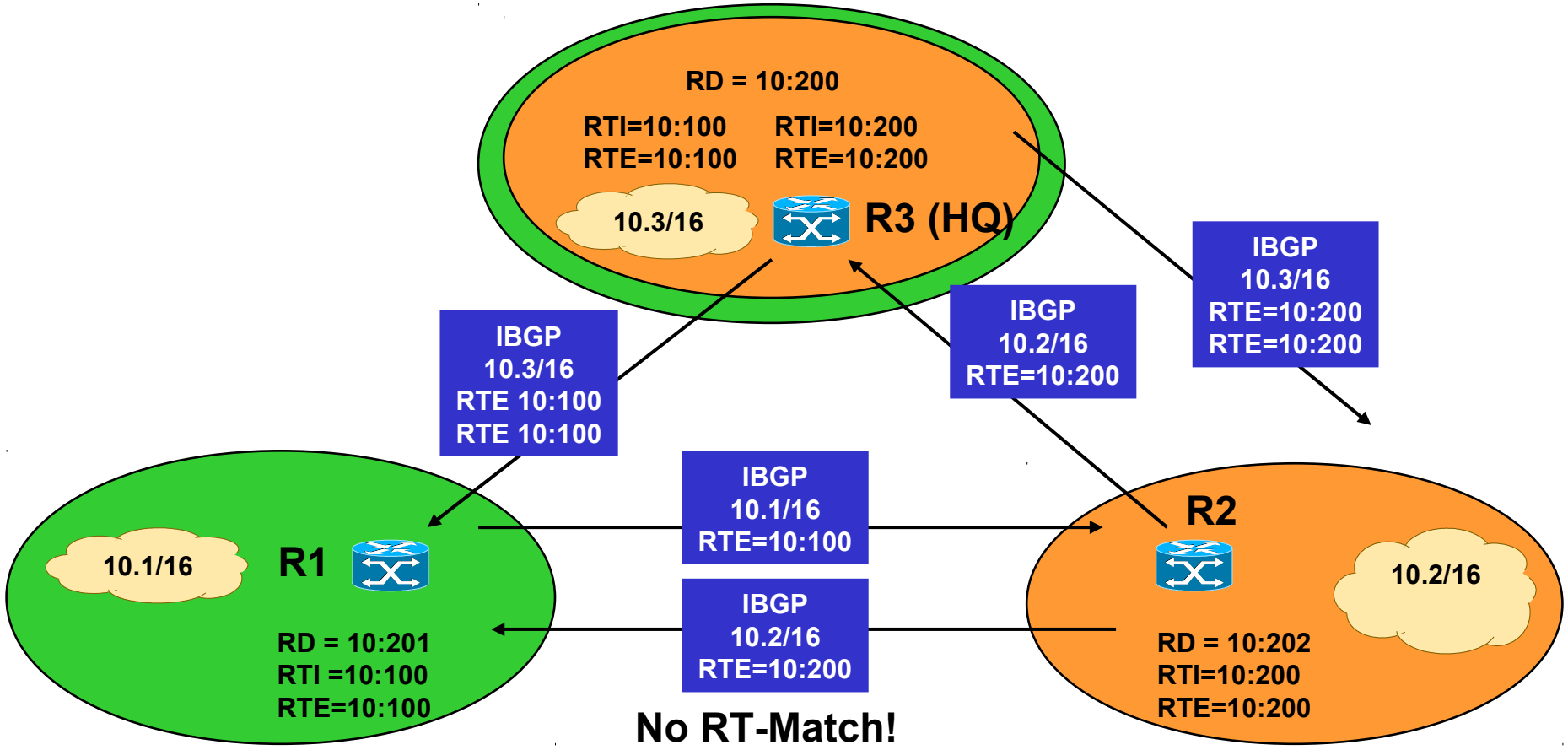
MED → IGP Metric

Transparent for IGP



- R3 has one FIB per RT
  - One FIB for global RT
  - One FIB for VRF CE1 RT
  - One FIB for VRF CE2 RT
- Each MPLS-Router has exactly one LFIB
  - PE routers must be connected with CE routers via (sub) interfaces

Overlapping VPNs



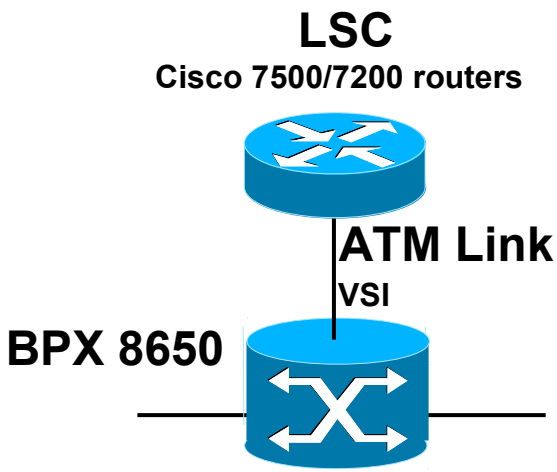
- IBGP Split Horizon Rule assures that R3 (HQ) does not forward routes learned by peers
- IP networks must be unique in overlapping situations!

Cell-based MPLS

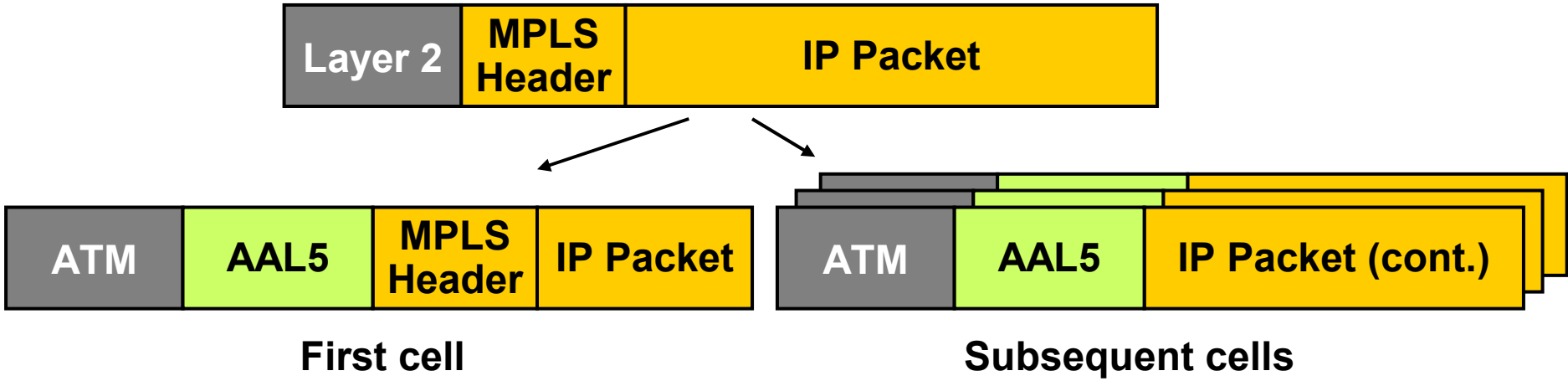
If you need this...

Cell-based MPLS

- Label-switching controlled ATM (LC-ATM)
  - On ATM switches
  - On Routers with ATM interfaces
- Legacy ATM switches become MPLS capable
  - Via firmware upgrade, if existing control processor allows that (LS 1010, Cat 8510, Cat 8540, Cat 5500)
  - Via external Label Switch Controller (LSC) attached on standard ATM interface (MGX 8850, BPX 8650)



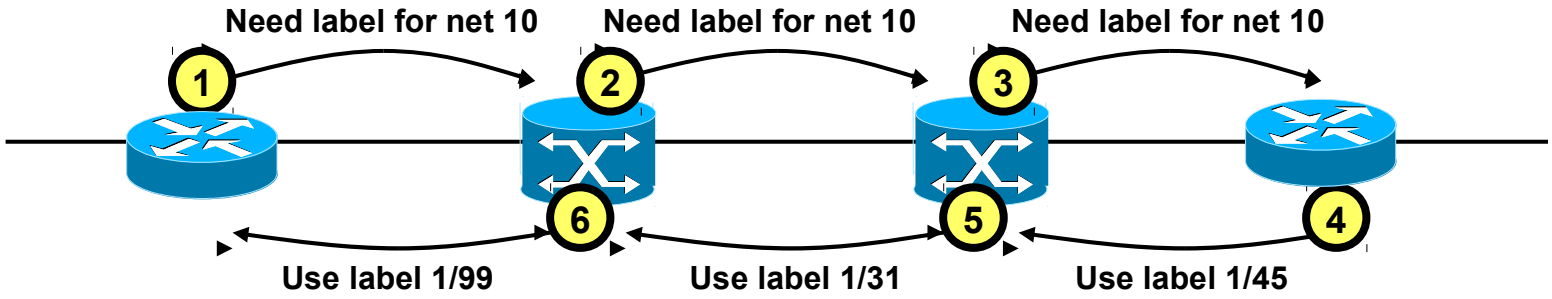
Cell-mode MPLS Cells



- ATM Switches can only switch VPI/VCI—no MPLS labels!
  - Only the topmost label is inserted in the VPI/VCI field
  - Other reserved VPI/VCI fields are used for LDP/TDP and routing updates
- Note: Typically only a few VPI/VCI combinations are supported by each switch
  - Labels are a very scarce resource !!!
- Per-interface label allocation

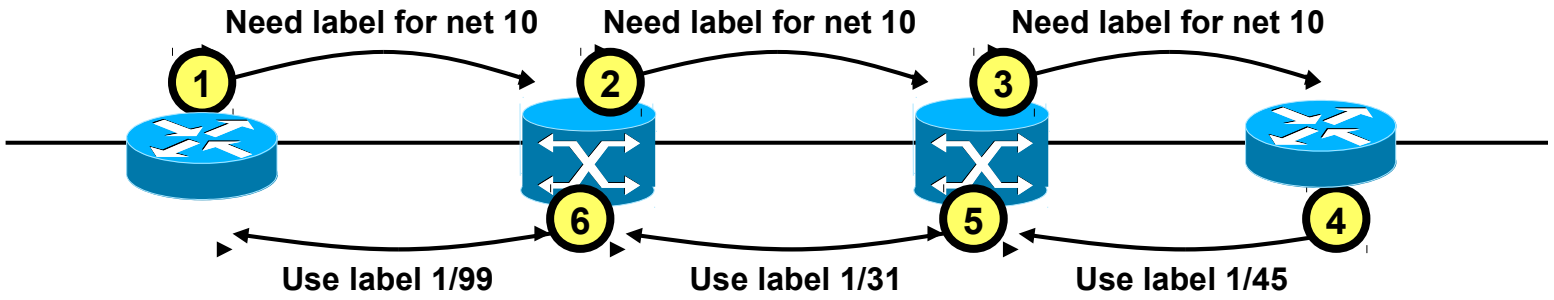
Basic Principles Summary

- MPLS Layer 2.5 packet is sent via AAL5
  - Top-of-stack label is always copied into VPI/VCI field
  - Per default: VPI=1, range can be configured
- LDP, TDP and routing protocols are sent *in-band* in VC 0/32 by default (IETF)
  - Other channel can be configured
  - Out-band control channel typically *not* implemented (e. g. Ethernet)
- ATM Switches typically perform *control-driven* label-requests *downstream*
  - Based on RT content, not actual data flow
  - Recursive process (request/response: "Ordered Control")

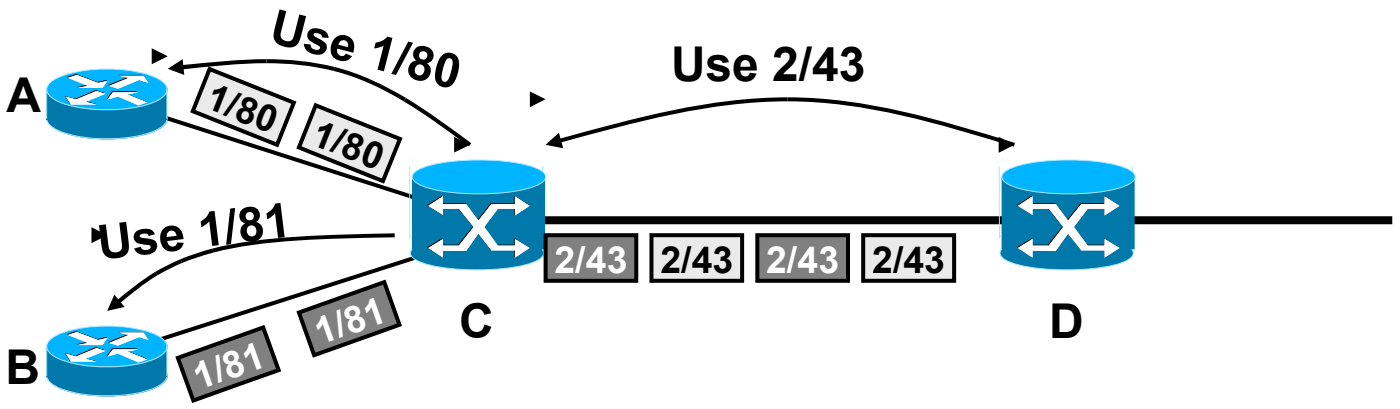


Label Request Procedure

- A router requests a label for every destination with next hop reachable via LC-ATM interface
- An ATM switch can only allocate an incoming label if it has already an outgoing label
  - Thus a label request can only be answered after outgoing label had been requested
  - "Ordered control"
- LSRs can always assign an incoming label
  - "Independent control"
- LFIB = ATM switching matrix



Reuse of Downstream Labels

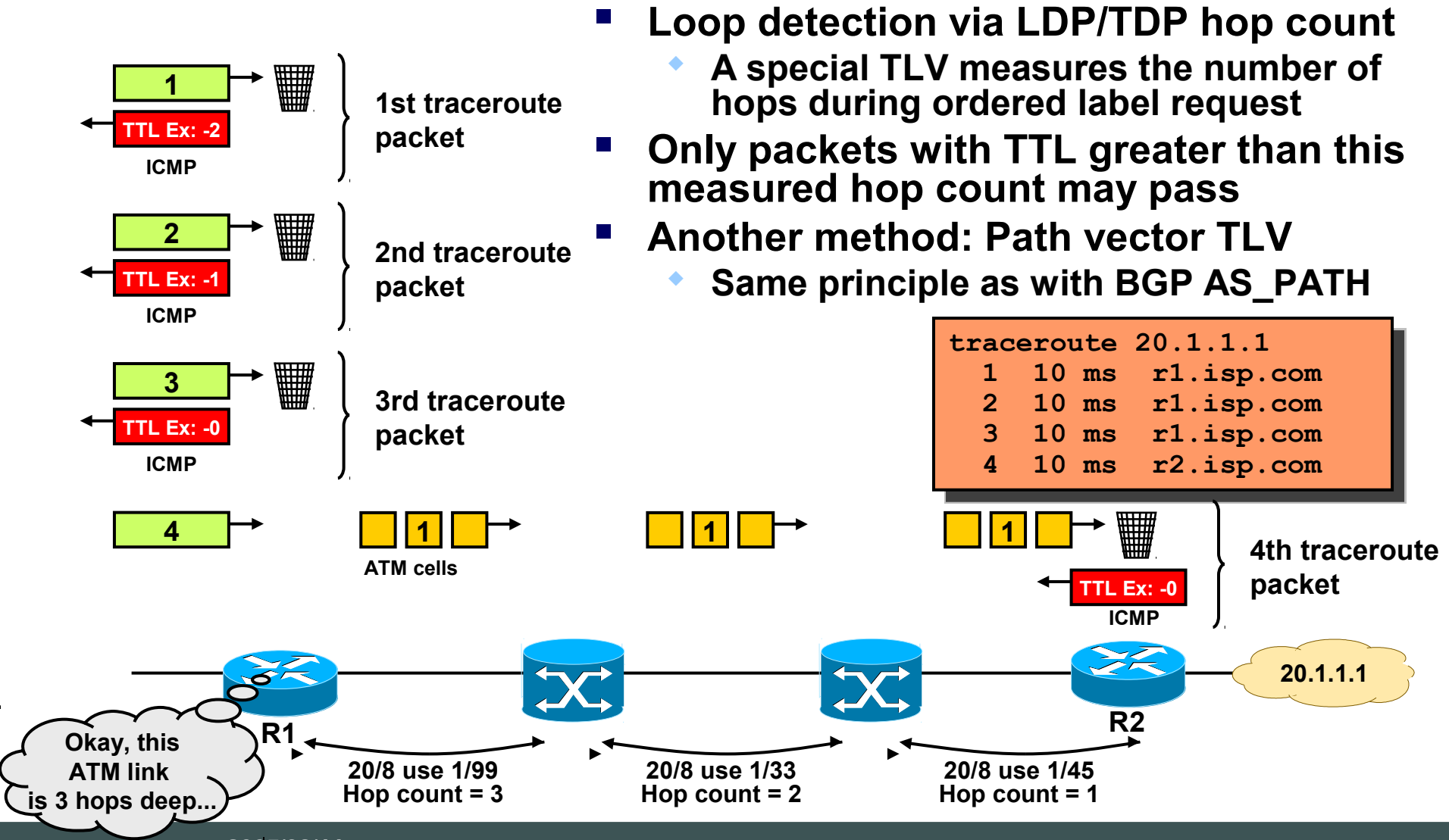


- Reusing downstream label leads to interleaving of IP packets !
  - Allocate a separate downstream label for every upstream request
  - Prevent cell interleaving (watch packet boundaries) –"VC Merge"

VC-Merge

- Blocks incoming cells until last cell of packet arrived
- Saves labels but requires switch to serialize all cells belonging to one packet
- Serialization delay increased and buffer resources needed
- Jitter increases !!!

LC-ATM Loop Detection



- Loop detection via LDP/TDP hop count
  - A special TLV measures the number of hops during ordered label request
- Only packets with TTL greater than this measured hop count may pass
- Another method: Path vector TLV
  - Same principle as with BGP AS_PATH

Summary

- The very basic idea:
  - MPLS decouples information used for forwarding (the label) and information used for routing (the IP address)
- MPLS transport
  - Is fundamental to other MPLS features
  - Requires a label distribution system (LDP/TDP)
  - Requires CEF to establish a fast FIB
  - Can do label stacking which allows greater flexibility
  - Differentiate frame-based and cell-based MPLS
- MPLS VPNs
  - Additional label to differentiate VPNs
  - VPNv4 addresses and Route Targets to define VPN menbership of the VRFs