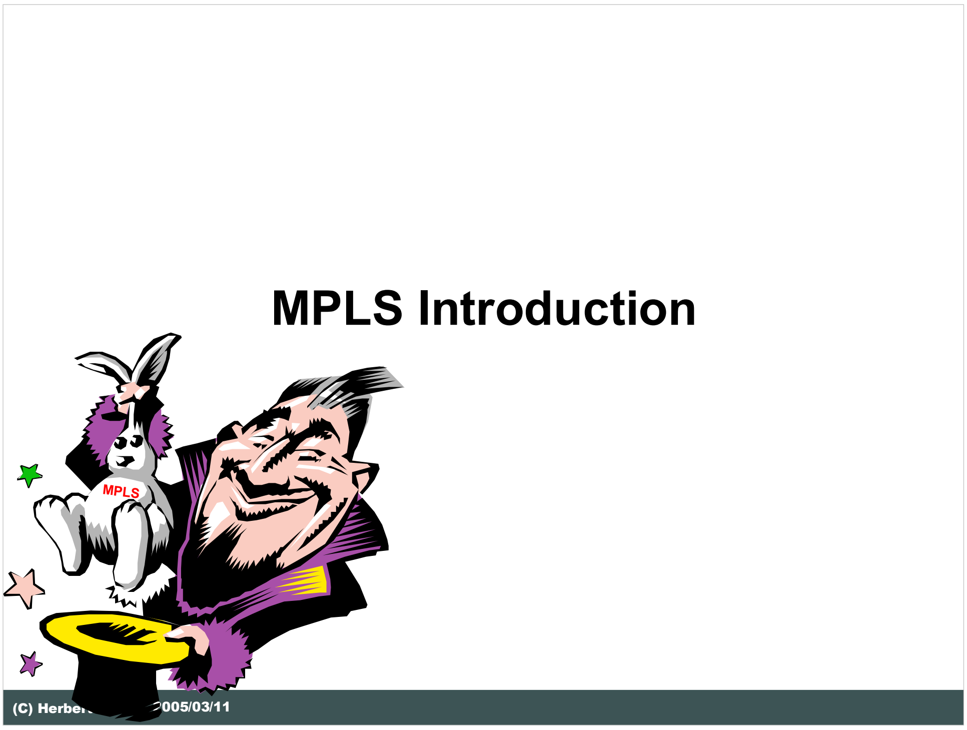
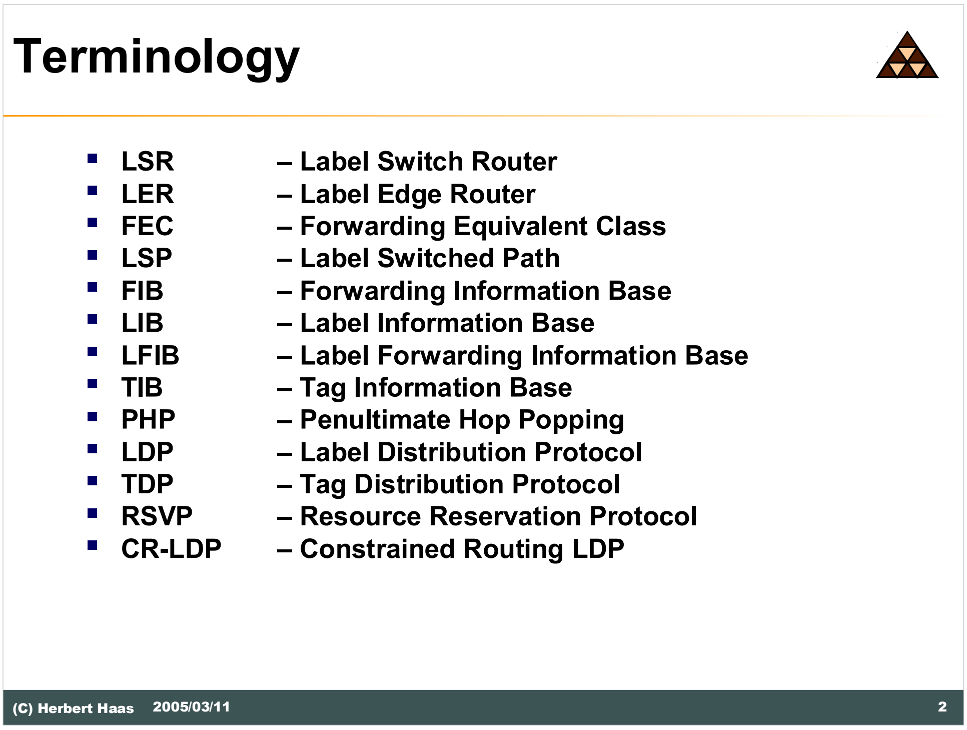


MPLS Introduction

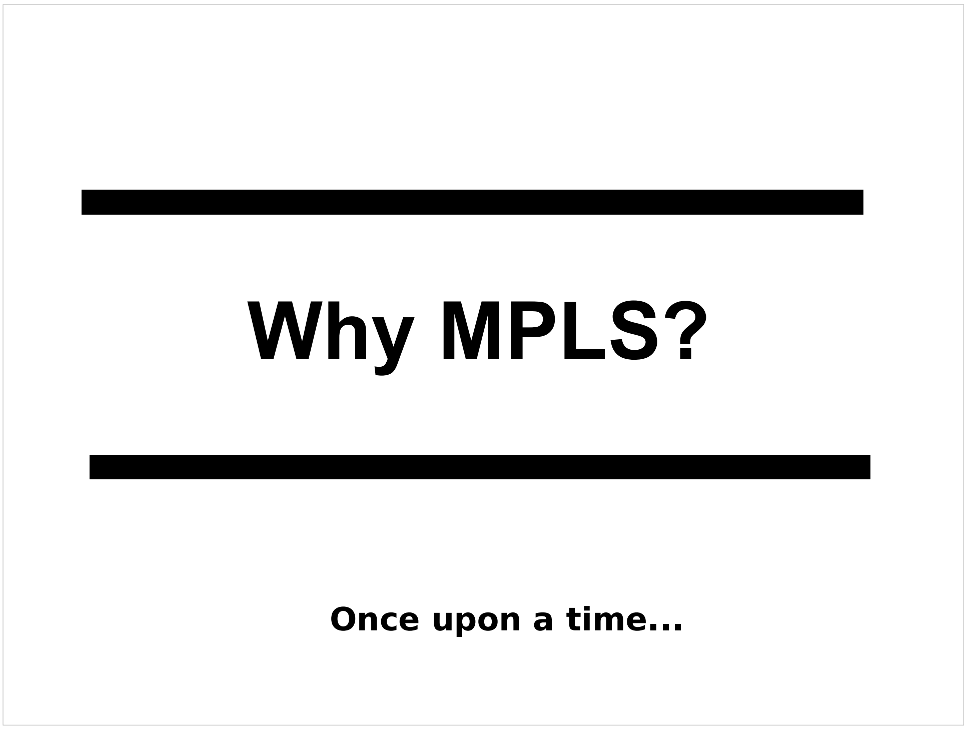
(C) Herber [illegible] 2005/03/11



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

This slide lists a few of the thousand important abbreviations.



Why MPLS?

Once upon a time...

Computer science:

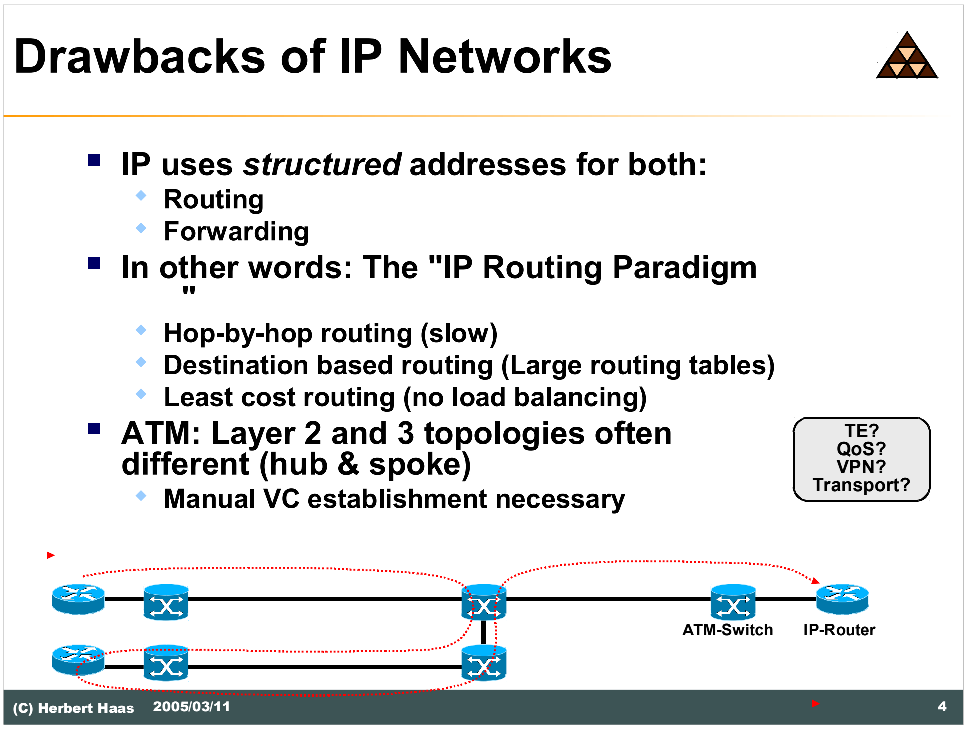
A study akin to numerology and astrology, but lacking the precision of the former and the success of the latter.

Networking science:

The costly enumeration of the obvious.

MPLS:

No science at all.



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

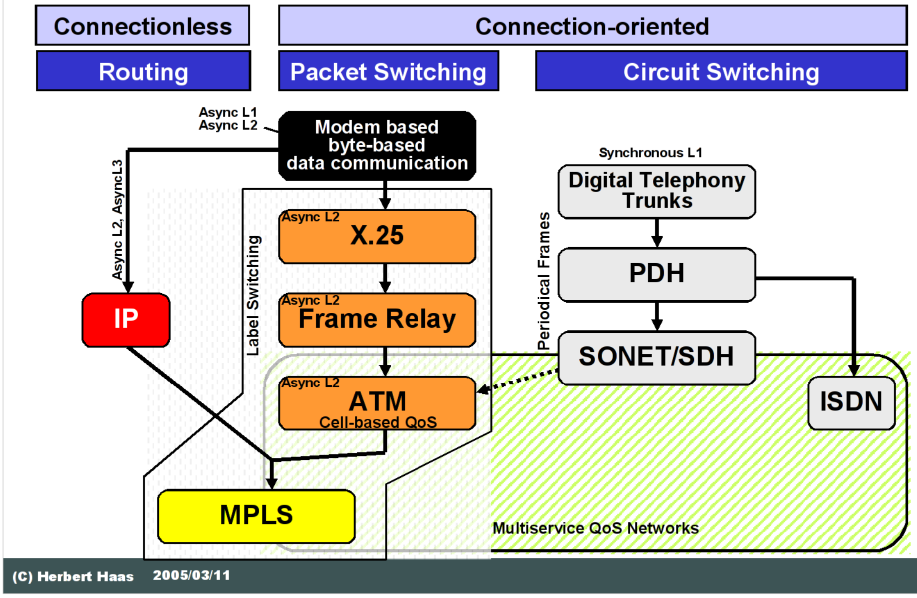
TE?
QoS?
VPN?
Transport?

ATM-Switch
IP-Router

(C) Herbert Haas 2005/03/11 4

Destination based least cost IP routing does not support load balancing. Although policy based routing is supported by most vendors this solution does not scale. Also there are no satisfying solutions available for Taffic Engineering (TE) and Quality of Service (QoS). Indeed there are some working IP VPN solutions (e. g. IPSec based) but it is still a scalability issue.

Networking Evolution



Connectionless

Connection-oriented

Routing

Packet Switching

Circuit Switching

Async L1
Async L2

Modem based
byte-based
data communication

Synchronous L1

Digital Telephony
Trunks

Async L2, AsyncL3

Label Switching

Periodical Frames

Async L2 X.25

Async L2 Frame Relay

Async L2 ATM
Cell-based QoS

IP

PDH

SONET/SDH

ISDN

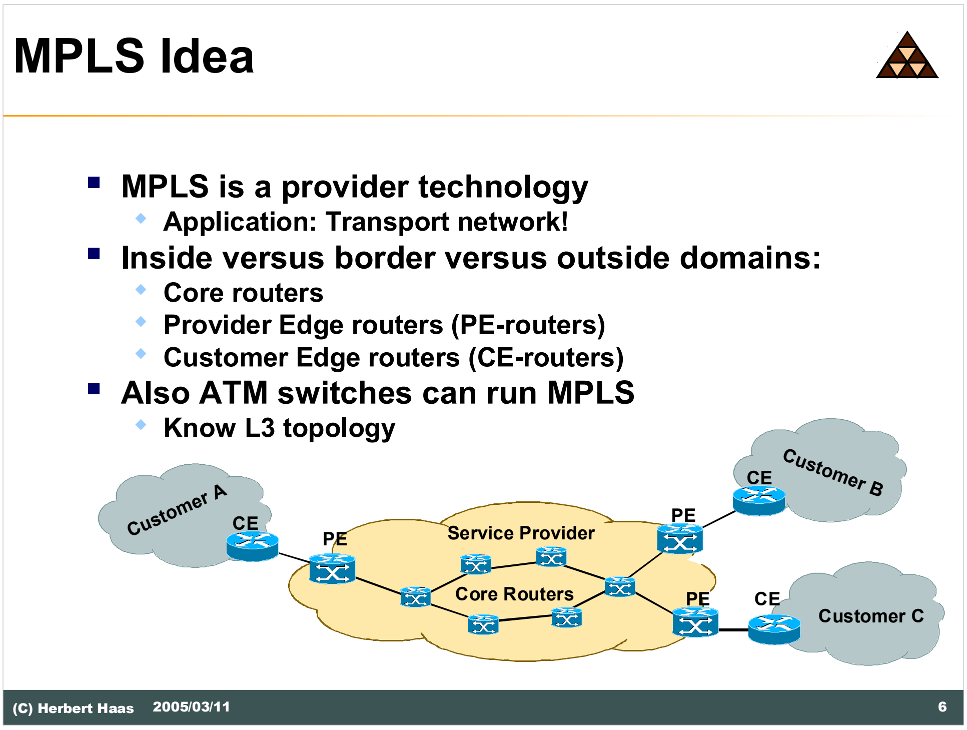
MPLS

Multiservice QoS Networks

(C) Herbert Haas 2005/03/11 5

The picture above summarizes the whole networking evolution. The most important concepts are mentioned such as connectionless and connection-oriented protocols, routed and packet-switched protocols, etc.

The only important message here is that MPLS aims to provide the best of all worlds ever known. This *must* be crazy...but is it crazy enough?



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

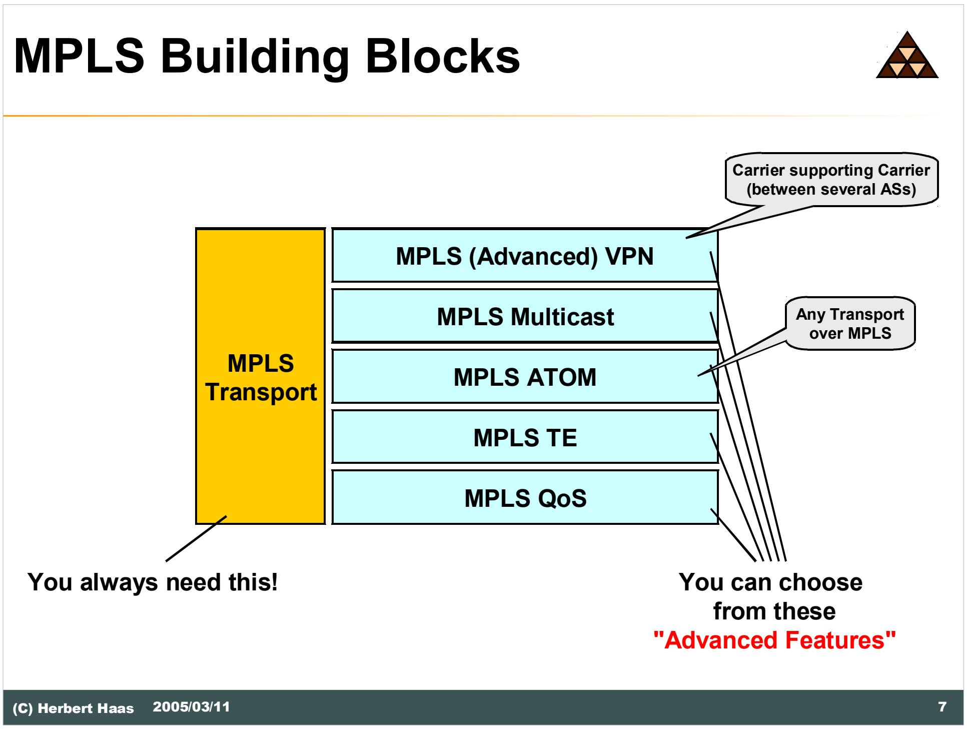
(C) Herbert Haas 2005/03/11 6

There is one unique basic concept with MPLS which is the idea of a border and a core network.

What if the border MPLS routers are somehow clever, determine where the packet has to go (perform the whole routing process) and add a simple but significant label on it (the packet), so that all subsequent MPLS routers (somehow) know what to do with it

Actually the whole principle had been stolen from the ATM world—but it has also been improved. ATM can only swap two labels (the VPI and the VCI, but mostly they are swapped together).

Wouldn't there be much greater flexibility if we had more labels per packet?



MPLS Building Blocks

Carrier supporting Carrier
(between several ASs)

MPLS Transport

MPLS (Advanced) VPN

MPLS Multicast

MPLS ATOM

MPLS TE

MPLS QoS

Any Transport
over MPLS

You always need this!

You can choose
from these
"Advanced Features"

The MPLS technology supports different types of so called MPLS Applications like the one shown in the graphic above.

•MPLS Transport is the base MPLS Application which needs to be configured if you want to use other MPLS Applications like MPLS VPN, MPLS TE etc. MPLS Transport can be used to replace pure layer 3 IP forwarding with Label switching.

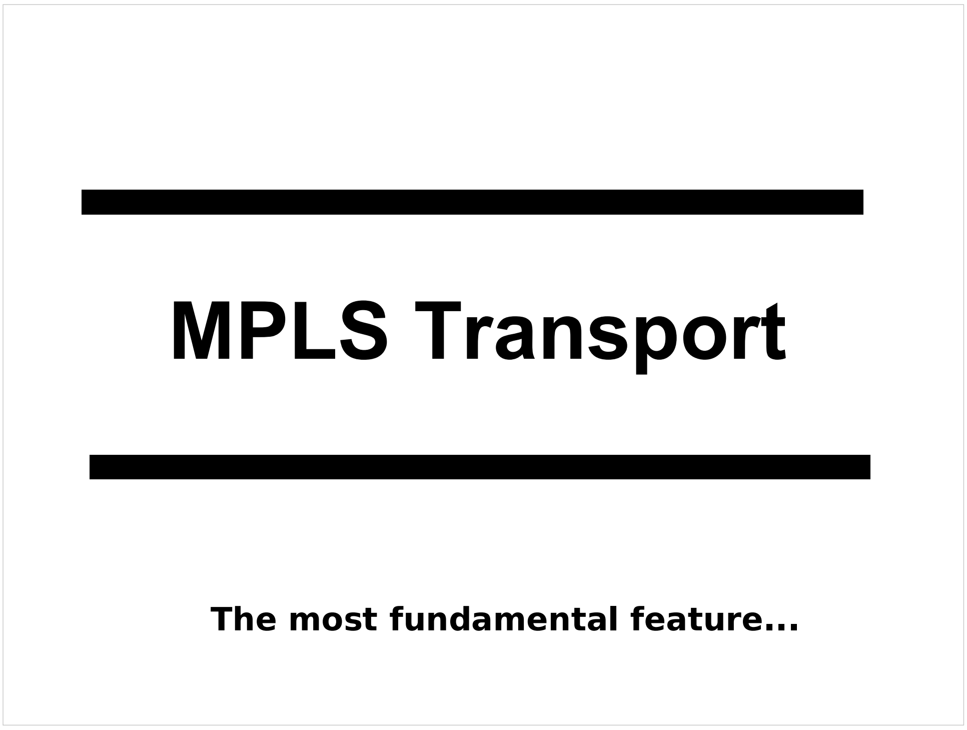
•MPLS VPN can be used to built closed user groups on top of the MPLS Transport system.

•MPLS Multicast is needed if Multicast transport through an MPLS cloud is desired.

•MPLS Atom allows you to tunnel Ethernet, Frame-relay and ATM traffic through an MPLS domain.

•MPLS TE can be used to overcome load-balancing limitations of IP routing protocols by the use of traffic engineering tunnels.

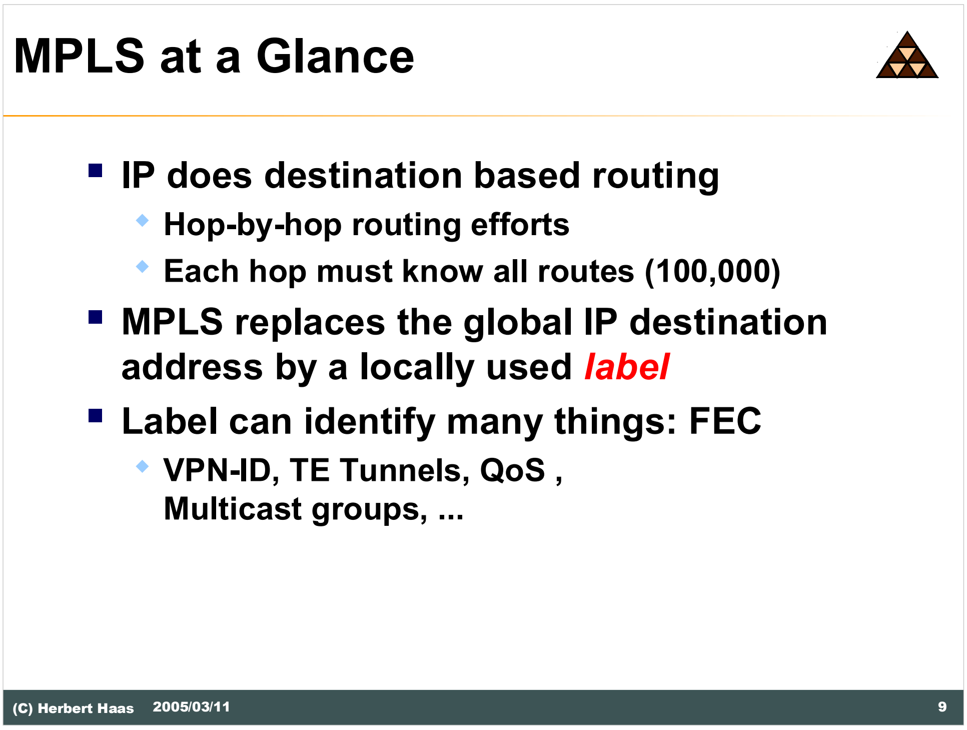
•MPLS QoS is used if you want to support different traffic classes inside your MPLS network.



MPLS Transport

The most fundamental feature...

If you understand MPLS Transport then you will follow the rest of it...



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, ...

(C) Herbert Haas 2005/03/11 9

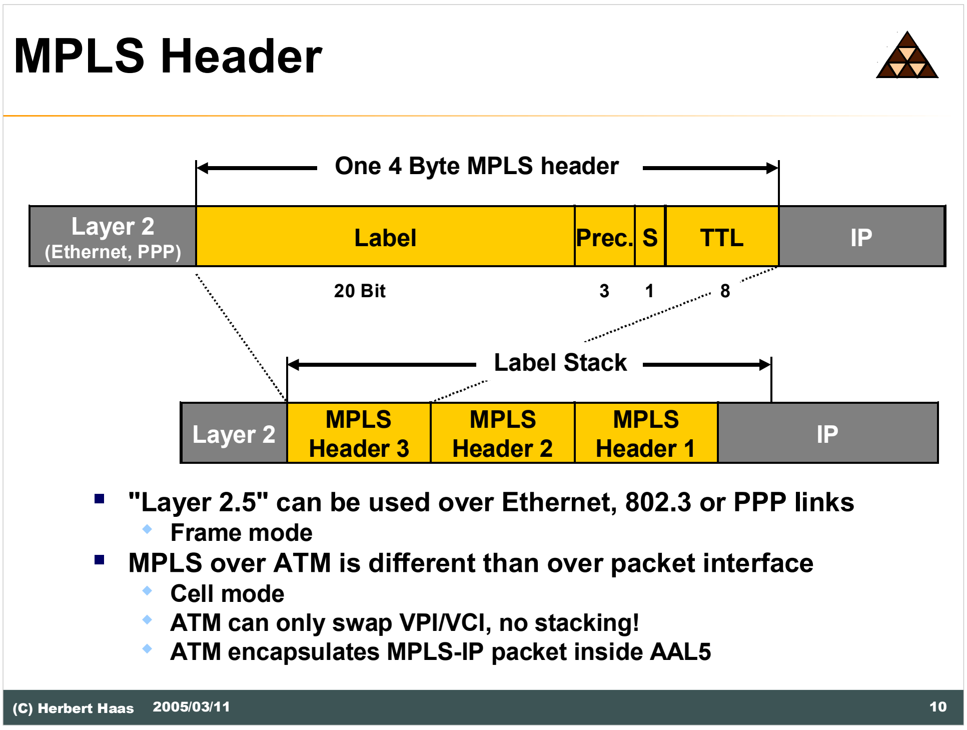
MPLS was formerly known as "tag switching" and was invented by Cisco. Today it is standardized as Multiprotocol Label Switching by the IETF.

The major difference between the IP and MPLS forwarding plane is:

An IP router uses the **longest match routing rule** when it scans through the IP forwarding table. This means the subnet mask information stored in the IP forwarding table determines how many bits of the incoming packets IP address must match with the IP entry in the forwarding table. In case of more than one match is found, the longest match wins.

The MPLS forwarding engine does not use the longest match routing rule. MPLS always requires an **exact match** between the incoming Label and the Label forwarding table.

A label in MPLS could also identify other things than only a destination, it could be used to “signal” a QoS group, Multicast group, MPLS TE tunnels etc. Therefore we assign a label to a **Forward Equivalent Class (FEC),** which has a common meaning. The FEC simply tells what the label stands for (e. g. a VPN, a next-hop, a QoS-class, ...).



MPLS Header

One 4 Byte MPLS header

Layer 2 (Ethernet, PPP)	Label	Prec.	S	TTL	IP
	20 Bit	3	1	8	

Label Stack

Layer 2	MPLS Header 3	MPLS Header 2	MPLS Header 1	IP

- "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

(C) Herbert Haas 2005/03/11 10

The MPLS Header is made up of four bytes and is located between the layer two header and the layer three header. The existence of an MPLS header is indicated by the layer two type field entry **0x8848**.

The MPLS header is made up of a:

•20 bit label field used for forwarding,

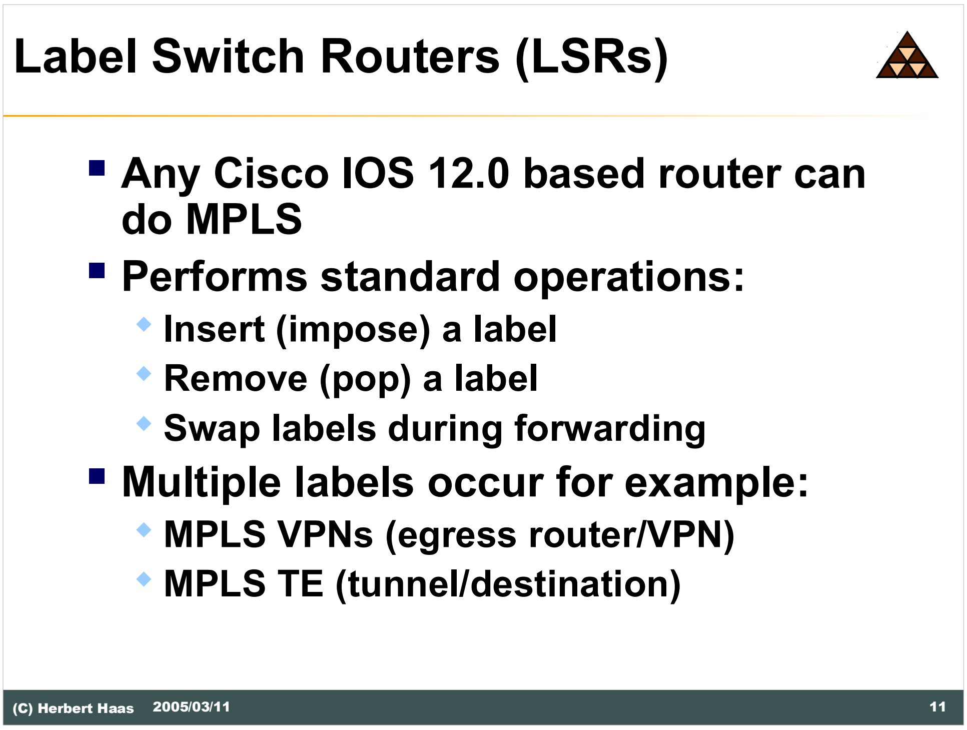
•3 Experimental bits typically used to carry IP Precedence settings,

•1 bit bottom of stack (0 indicates last label in the stack, 1 indicates there are some more labels on top of the bottom label)

•TTL field in which by default the IP TTL value is copied to when a Label is inserted.

If MPLS is used on top of ATM, the VPI/VCI field of the standard ATM cell header is used to carry the label information. There is no additional MPLS header involved because this would require hardware changes if you want to migrate existing ATM devices to support MPLS.

Note: The labels 0 to 15 are reserved. Therefore the lowest usable label number is 16 and the highest possible label is 1,048,575 (which is actually 2^20-1). Only four out of the 16 reserved labels have been defined by RFC 3032, which are: 0 "IPv4 Explicit Null Label", 1 "Router Alert Label", 2 "IPv6 Explicit Null Label", 3 "Implicit Null Label".



Label Switch Routers (LSRs)

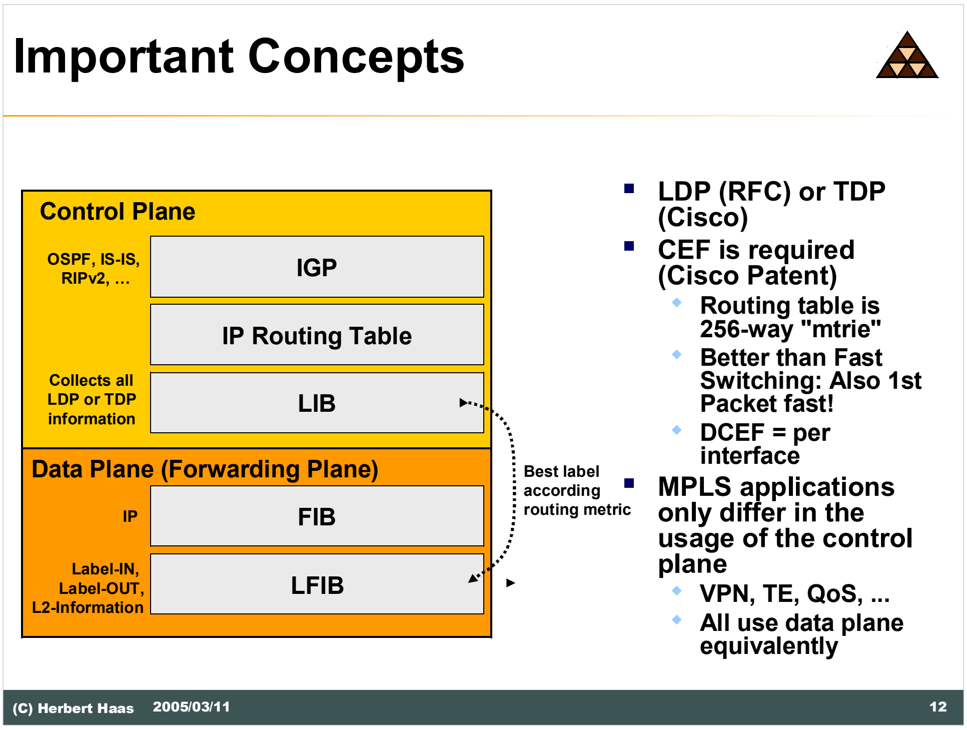
- Any Cisco IOS 12.0 based router can do MPLS
- Performs standard operations:
  - Insert (impose) a label
  - Remove (pop) a label
  - Swap labels during forwarding
- Multiple labels occur for example:
  - MPLS VPNs (egress router/VPN)
  - MPLS TE (tunnel/destination)

(C) Herbert Haas 2005/03/11 11

MPLS is basically a **software** solution. With Cisco IOS version 12.0, routers are able to perform CEF switching (explained soon in detail), which is the basis for MPLS. That is, nearly any Cisco router (except the smallest home office devices) are able to do MPLS.

MPLS routers are also called **"Label Switch Routers" (LSRs)** and must be able to perform the following basic operations: **Insert** (or "**impose**") a label (this is essential for edge routers), remove (or "**pop**") a label (this is essential for last hop routers), and **swap** labels (this is always done during packet forwarding).

Several reasons lead to a **label stack.** For example, with **MPLS VPNs**, the top label identifies the egress router while a second label identifies the VPN itself. Thus the egress router can (as soon as the packet arrived) pop the outermost label and forward the packet to the right interface according to the inner label. Another example is **MPLS Traffic Engineering (TE),** where the outer label points to the TE tunnel endpoint and the inner label to the final destination itself.



Important Concepts

Control Plane
OSPF, IS-IS, RIPv2, ...
IGP
IP Routing Table
Collects all LDP or TDP information
LIB
Data Plane (Forwarding Plane)
IP
FIB
Label-IN, Label-OUT, L2-Information
LFIB
Best label according routing metric

- 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

(C) Herbert Haas 2005/03/11 12

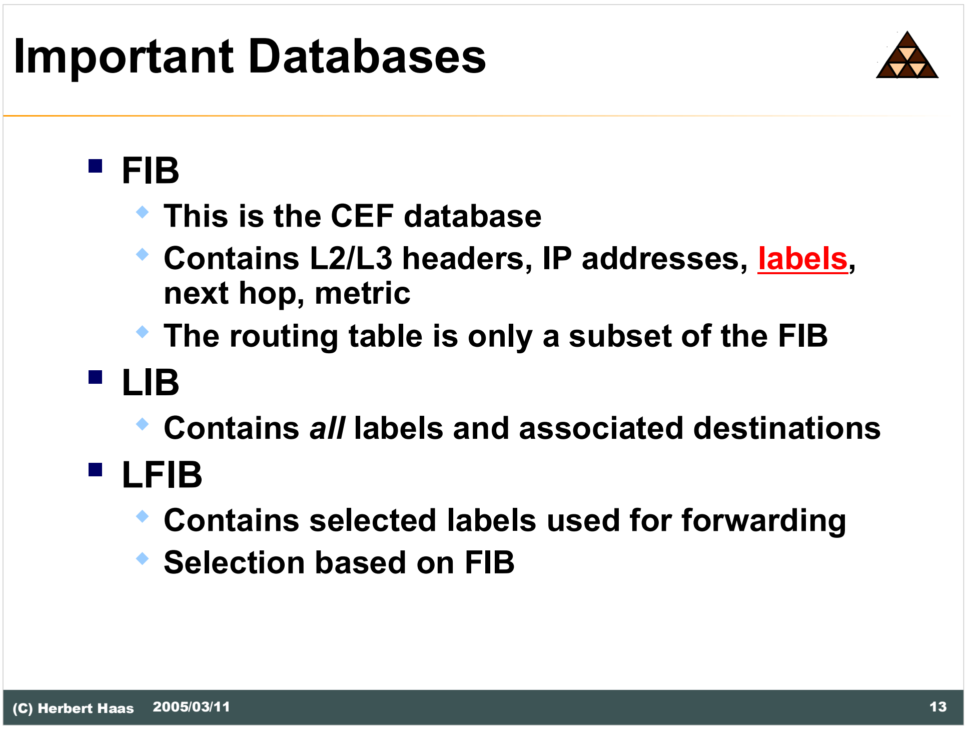
MPLS needs different types of tables which are interacting to provide MPLS forwarding functionality.

•The IP routing table is a common routing table which is built by the IGP in use.

•The FIB table is processed from the information held in the routing table plus all necessary layer 2 information and label Information needed for packet forwarding. All incoming IP packets are forwarded related to the information kept in the FIB table.

•The LIB table holds all the corresponding Label – IP Destination relationships. The LIB is built using either LDP or TDP updates. Both protocols distribute Label to IP prefix bindings. The LIB is a database of all possible labels.

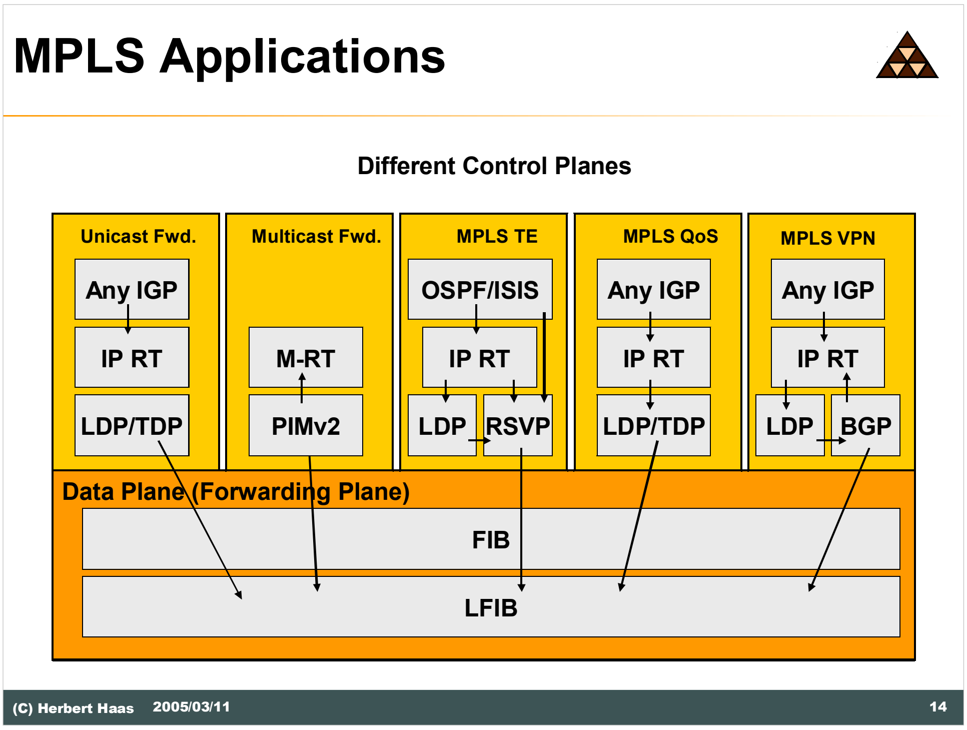
•The LFIB only holds the best Labels out of the LIB and is actually used to forward MPLS packets. What the best label in the LIB are is determined by the Next Hop information supplied by the local IGP.



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**

This slide summarized the three important databases which had been introduced with MPLS.



MPLS Applications

Different Control Planes

Unicast Fwd.
Any IGP
IP RT
LDP/TDP

Multicast Fwd.
M-RT
PIMv2

MPLS TE
OSPF/ISIS
IP RT
LDP
RSVP

MPLS QoS
Any IGP
IP RT
LDP/TDP

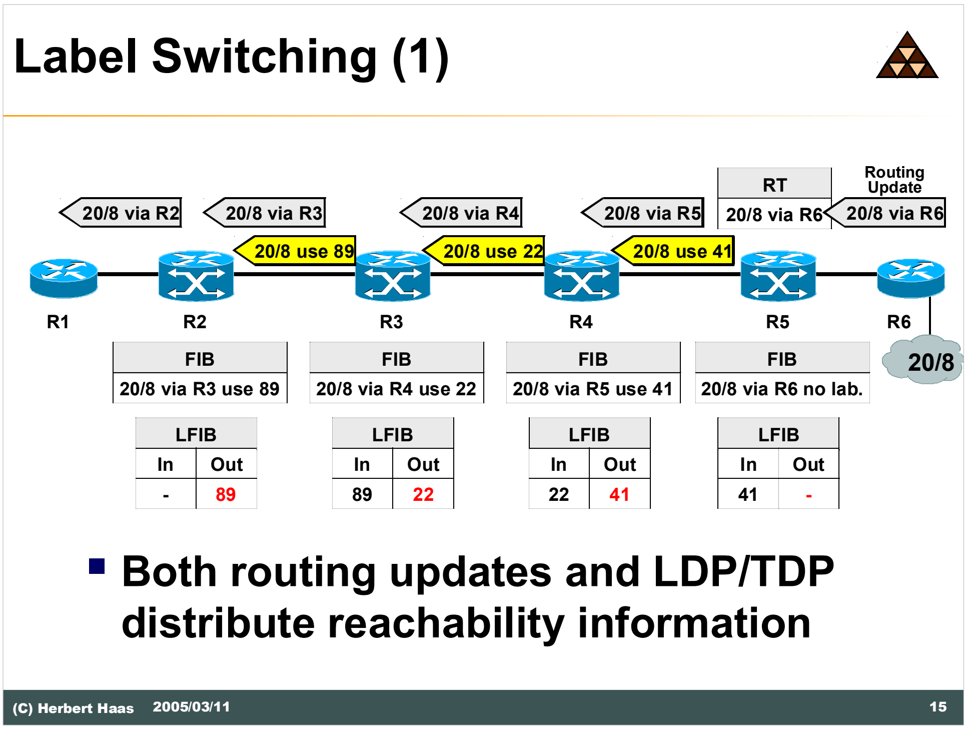
MPLS VPN
Any IGP
IP RT
LDP
BGP

Data Plane (Forwarding Plane)
FIB
LFIB

The diagram above illustrates how different MPLS applications use a different control plane. It is in fact the control plane which determines the FECs—in other words, what label-based forwarding is good for.

But all applications use the same (primitive) data plane.

Note that there are different types of MPLS-based Multicast. MPLS Multicast is discussed in another chapter, soon...



Label Switching (1)

FIB
20/8 via R3 use 89

FIB
20/8 via R4 use 22

FIB
20/8 via R5 use 41

FIB
20/8 via R6 no lab.

LFIB	
In	Out
-	89

LFIB	
In	Out
89	22

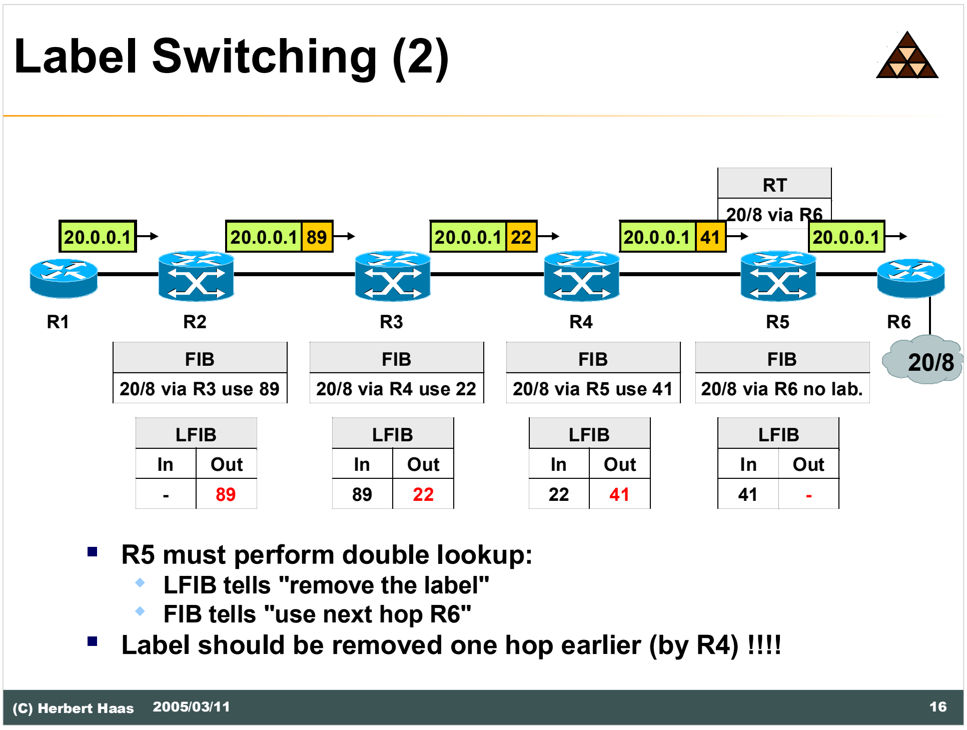
LFIB	
In	Out
22	41

LFIB	
In	Out
41	-

- Both routing updates and LDP/TDP distribute reachability information

(C) Herbert Haas 2005/03/11 15

The picture above shows how a label-switched path is established from left (near the “destination network” 20/8) to the right. Both routing updates and label distribution protocol (LDP or TDP) distribute reachability information for this destination network.



Label Switching (2)

FIB
20/8 via R3 use 89

FIB
20/8 via R4 use 22

FIB
20/8 via R5 use 41

FIB
20/8 via R6 no lab.

LFIB	
In	Out
-	89

LFIB	
In	Out
89	22

LFIB	
In	Out
22	41

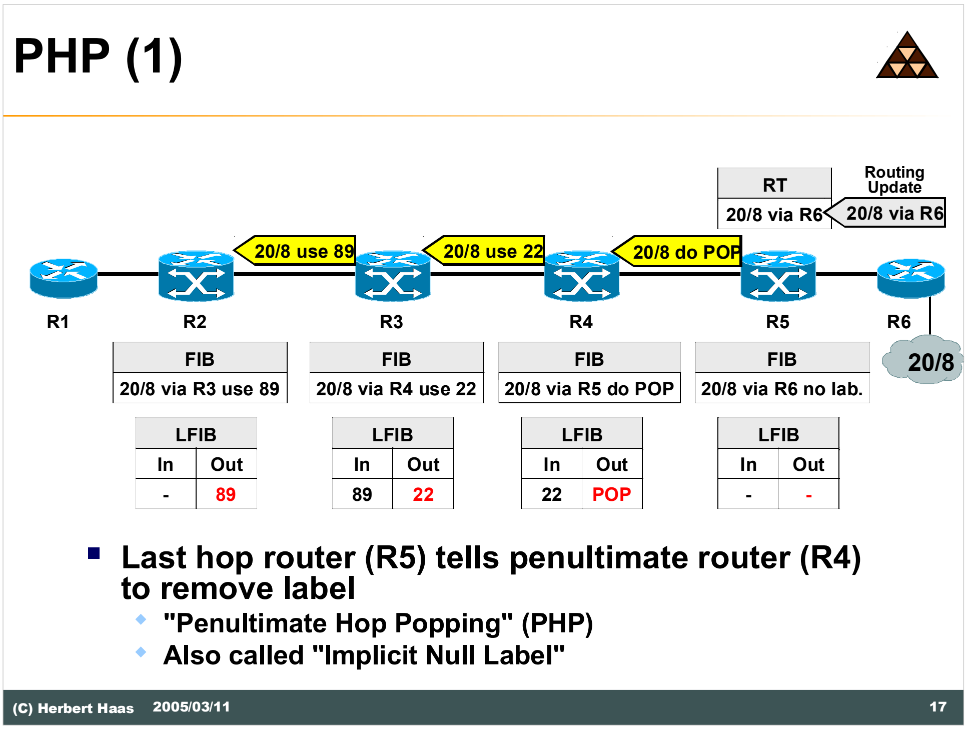
LFIB	
In	Out
41	-

- **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) !!!!**

(C) Herbert Haas 2005/03/11 16

The picture above shows how packets can now be sent using a MPLS header. Label switching is performed on each hop (LSR) inside the provider domain (R2, R3, R4, R5). The LFIB tables are used to perform a fast lookup.

But R5 cannot find any outgoing label in its LFIB. After this unsuccessful lookup, R5 looks into the FIB and determines the next hop. Note that this double lookup would be done for every packet! Therefore it would be reasonable to remove the label even one hop earlier (the penultimate hop, R4) in order to leave R5's LFIB empty.



PHP (1)

RT
20/8 via R6

FIB
20/8 via R3 use 89

FIB
20/8 via R4 use 22

FIB
20/8 via R5 do POP

FIB
20/8 via R6 no lab.

LFIB	
In	Out
-	89

LFIB	
In	Out
89	22

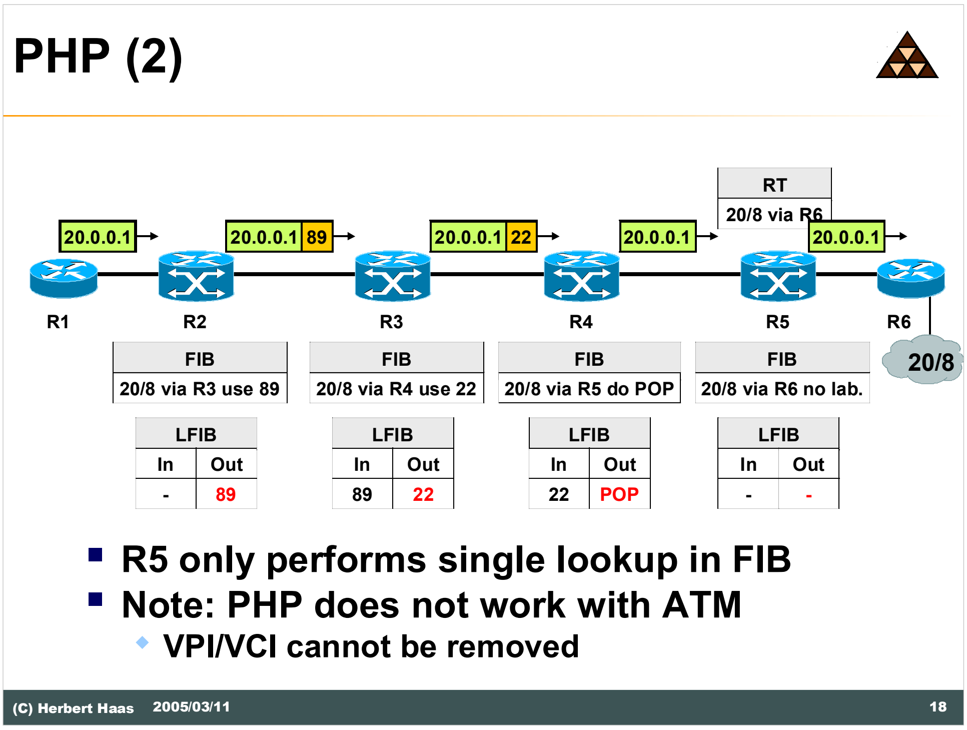
LFIB	
In	Out
22	POP

LFIB	
In	Out
-	-

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

(C) Herbert Haas 2005/03/11 17

In this scenario "Penultimate Hop Popping" (PHP) is illustrated. Now R5 does not allocate an incoming label for this destination but rather announces to R4 to use an "implicit null" label. It is also said, that R4 should perform the "POP" operation. The label number "3" had been reserved to represent the "do POP" command.



PHP (2)

RT

20/8 via R6

20.0.0.1 → R1 → R2

20.0.0.1 | 89 → R2 → R3

20.0.0.1 | 22 → R3 → R4

20.0.0.1 → R4 → R5

20.0.0.1 → R5 → R6

R1 R2 R3 R4 R5 R6

20/8

FIB
20/8 via R3 use 89

FIB
20/8 via R4 use 22

FIB
20/8 via R5 do POP

FIB
20/8 via R6 no lab.

LFIB	
In	Out
-	89

LFIB	
In	Out
89	22

LFIB	
In	Out
22	POP

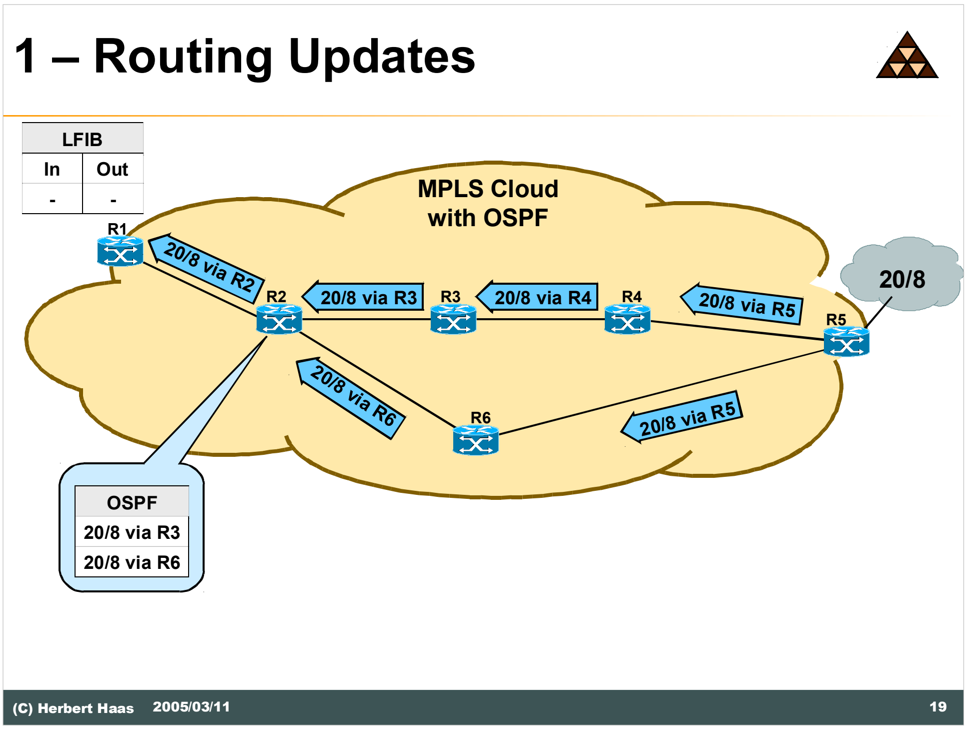
LFIB	
In	Out
-	-

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

(C) Herbert Haas 2005/03/11 18

Note that some router in between (e.g. R3) can be configured as **aggregation point**. That is, this router may aggregate several prefixes using a shorter prefix (e. g. 20/6) and a dedicated label. In this case the label-switched path is broken into two segments. The penultimate router (just before the aggregation router) already performs "POP" and the aggregation router therefore must perform a routing table lookup (this is necessary especially when the destination is more specific than the announced aggregate—there might be different downstream paths from the aggregation point).

Note: ATM LSRs must not aggregate because they cannot forward IP packets. Also aggregation must not be used in applications where an end-to-end tunnel is required, such as as in MPLS VPNs.



1 – Routing Updates

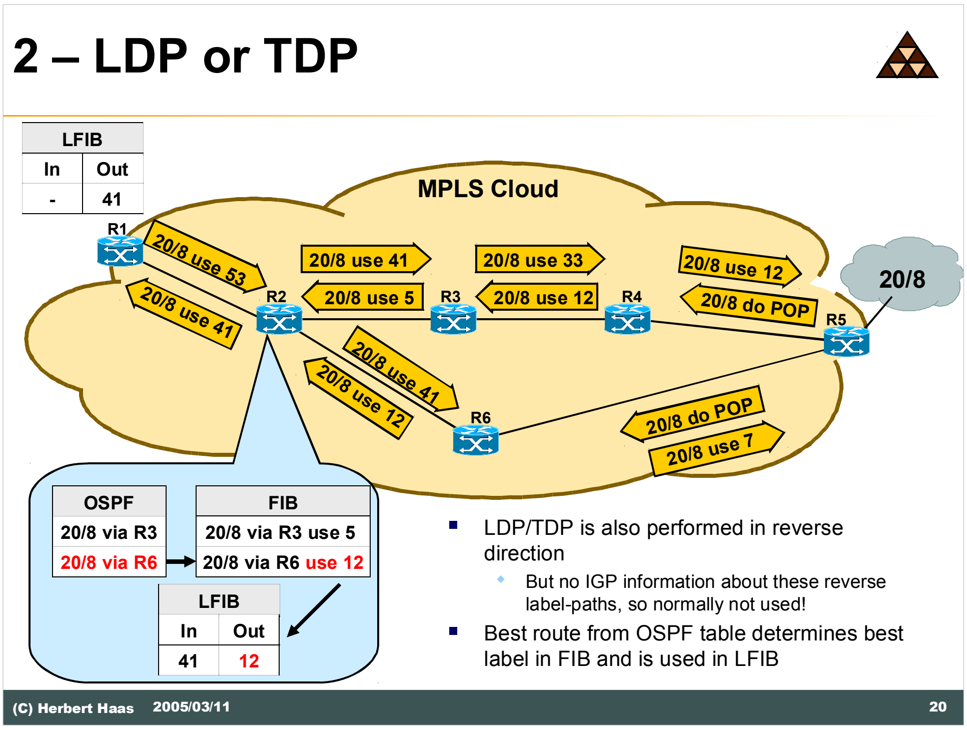
LFIB	
In	Out
-	-

OSPF
20/8 via R3
20/8 via R6

(C) Herbert Haas 2005/03/11 19

The first table that needs to be available is the routing table, which is build up in this example with the help of the OSPF link state routing protocol.

If only the MPLS Transport system is in use any IGP can be used. Only MPLS Transport in combination with MPLS TE requires a link state routing protocol like OSPF or ISIS.



2 – LDP or TDP

LFIB	
In	Out
-	41

OSPF
20/8 via R3
20/8 via R6

FIB
20/8 via R3 use 5
20/8 via R6 use 12

LFIB	
In	Out
41	12

- 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

(C) Herbert Haas 2005/03/11 20

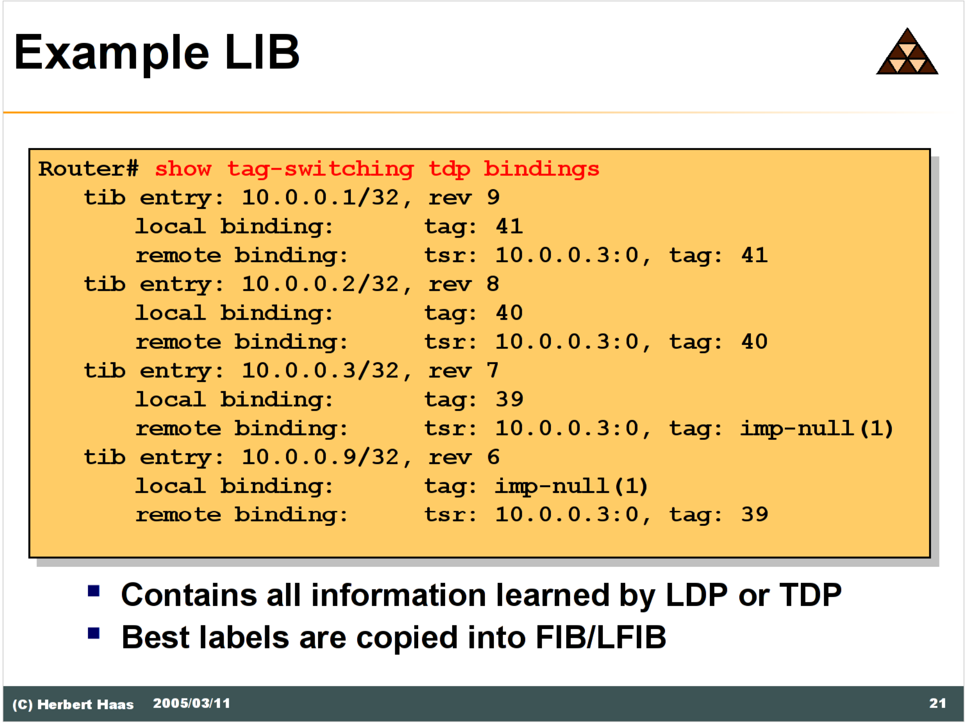
Allocated labels are advertised to all neighbor LSRs regardless of whether they are upstream or downstream.

Per-platform label allocation: typically an LFIB contains no incoming interface, so the same destination (next hop) can be associated with the same label for all interfaces. The LSR simply advertises the same label for the same destination through all interfaces. LSR announces label to adjacent LSRs only once even if there are parallel links between them. Advantage: Quicker label exchange, small LFIB. Drawback: Insecure: A third party router can send packets to the LSR even though the label was not announced to it.

Per-interface label space: LFIB contains incoming interface. Label can be reused per interface with different meanings.

POP (implicit null) removes outermost label.

PHP does not work on ATM because VPI/VCI cannot be removed. POP or "implicit null label" uses value 3 when being advertised to a neighbor.



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

This table shows the content of a LIB table.

tib entry ....specifies the destination network

local binding ....informs you about the locally generated Label

Remote binding ....specifies the Label binding information received from a neighbor router

tsr....indicates the neighbor routers LPD ID and its label space

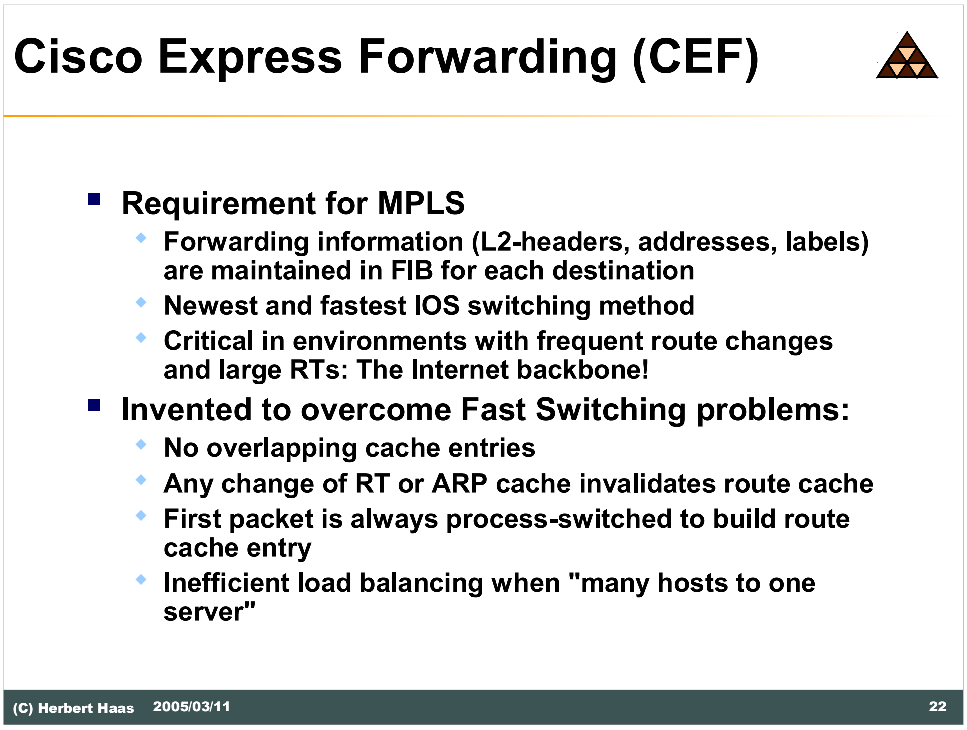
TIB = Tag information base (Cisco uses this terminology instead of LIB when TDP is used).

LPD/TDP ID:0 means per box label space (otherwise per interface label space)

Retention Modes

Liberal Retention Mode: All upstream/downstream labels from adjacent LSRs are stored in LIB. This improves convergence speed.

Conservative Retention Mode: Only next-hop labels are stored in LIB. Downstream on-demand distribution required during convergence phase.



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"

(C) Herbert Haas 2005/03/11 22

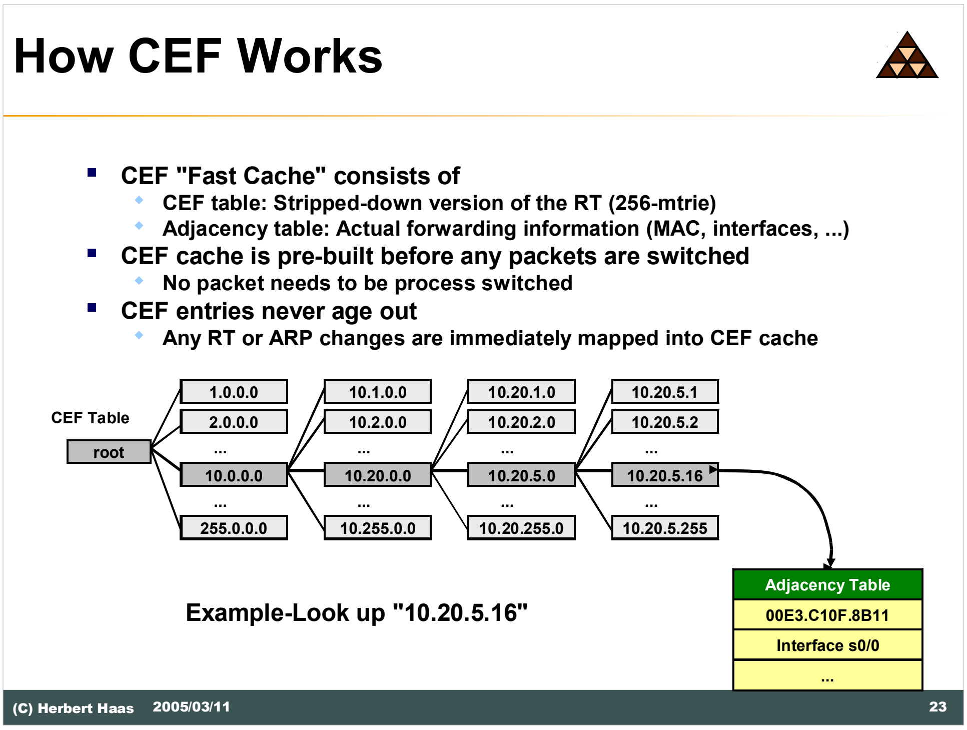
Many route changes occur in the Internet backbone, causing cache entries to be invalidated frequently. Therefore, a significant percentage of Internet traffic is process switched. First tests with IOS "ISP Geek images" under extreme conditions. Now CEF is the default switching mode in Cisco IOS Release 12.0 and the only switching mode on Cisco 12000 routers and Catalyst 8500.

Cisco IOS 12.0 knows several switching methods: Process Switching, Fast Switching, Autonomous Switching, Silicon Switching Engine (SSE) Switching, Optimum Switching, Distributed Fast Switching, CEF, Distributed CED (dCEF).

Process Switching was the first switching method implemented in IOS. It is simple (brute-force), slow, CPU demanding, non-optimized but at least platform independent.

Fast Switching: Cached subset of the routing table and MAC address tables. During Process Switching (which is still done for the first packet), the information learned is stored in a fast cache. This information contains route (next hop), interface and MAC header combinations. In order to avoid collisions in the fast cache, beginning with IOS 12.0, radix trees instead of hash tables are used.

Compared to process switching and fast switching technologies, CEF supports packet manipulation on the fly. This means the FIB table lookup also provides some additional information (e.g. precedence settings, Label information etc.) which are implemented in the outgoing data packet.



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"

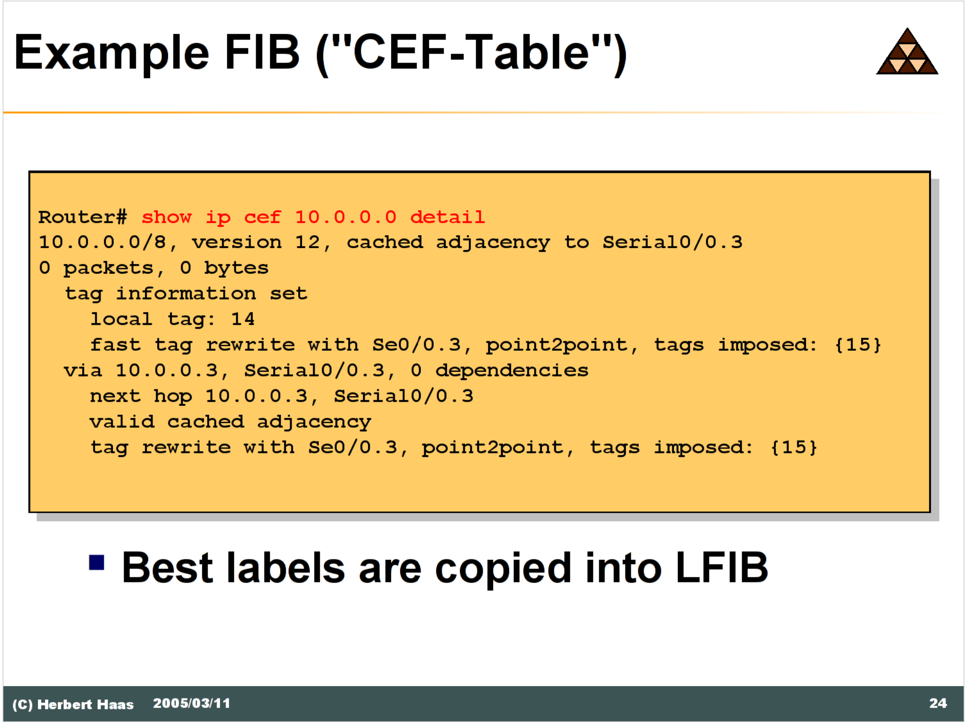
The CEF (FIB) table holds all the necessary information needed to rewrite the layer 2 and 3 header of an forwarded data packet. Changes in the routing table has to be reflected in the CEF table immediately.

mtree: tree of pointers; data is stored elsewhere.

Display CEF table information using **show ip cef summary.**

Display Adjacency table information: **show adjacency**.

dCEF: Very high performance boost. Each interface holds its own CEF table and is able to forward packets autonomously. Available on GSR, Cisco 7500 router



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}
```

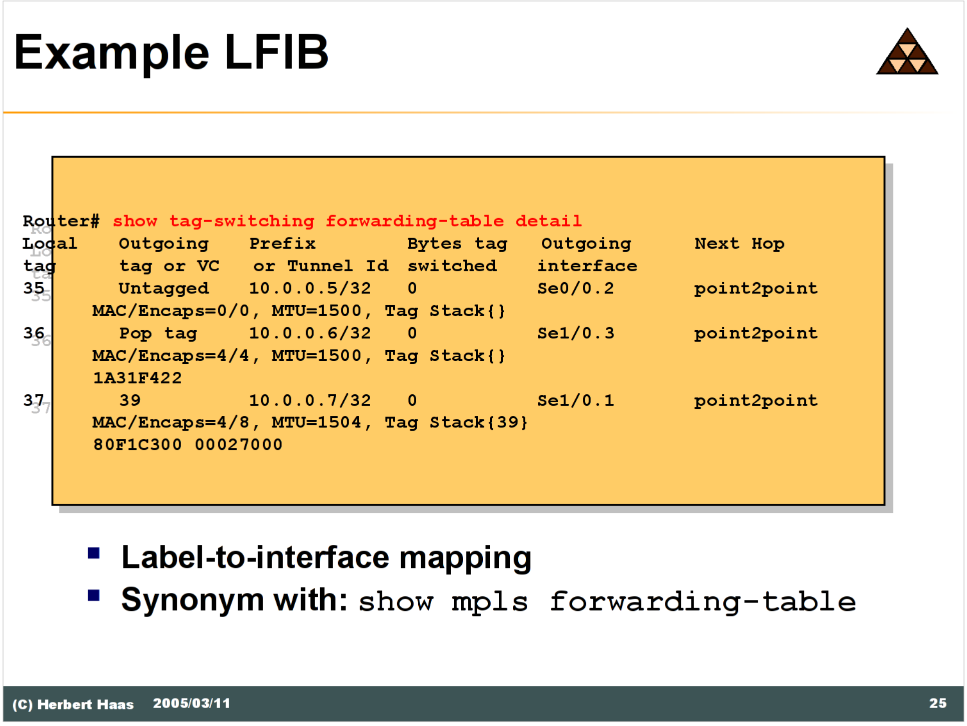
- **Best labels are copied into LFIB**

The listing above shows the content of a CEF (FIB) table.

The IP Destination 10.0.0.0/8 is reachable via the next hop router 10.0.0.3 and the interface that needs to be used is the interface S 0/0.3.

The Label that will be imposed is 15 learned from the next hop neighbor router.

The locally generated Label for this destination is 14. The local binding is propagated to all neighbors using the TDP/LDP protocol.



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`

(C) Herbert Haas 2005/03/11 25

This graphic displays the content of the LFIB table actually used for MPLS packet forwarding.

Local tag Locally generated Label,
represents incoming Label number

Outgoing tag Label received from the next hop neighbor

Prefix or Tunnel-id address or tunnel to which the packets with this tag are going.

MTU 1504 means that 1 label used.

MAC/Encaps Length in bytes of L2 header and length in bytes of packet encapsulation

This command is shown above with the additional keyword "**detail**".

Other keywords are:

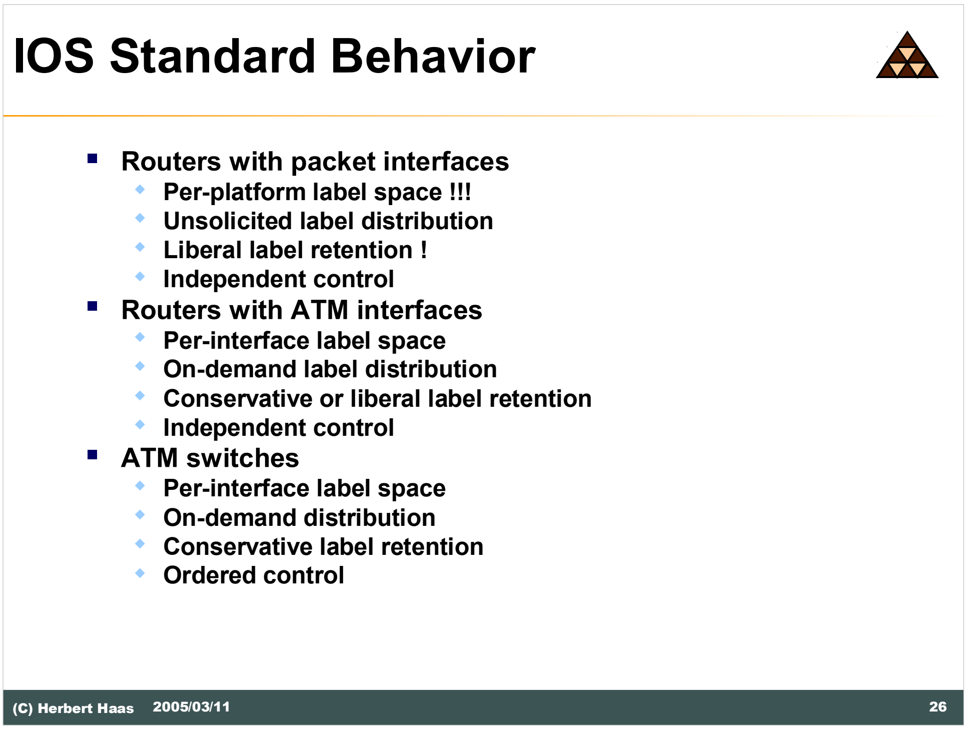
A.B.C.D (destination prefix)

interface (match outgoing interface)

next-hop (match next hop neighbor)

tags (match tag values)

tsp-tunnel (TSP tunnel id)



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

(C) Herbert Haas 2005/03/11 26

This slide summarized the main differences.

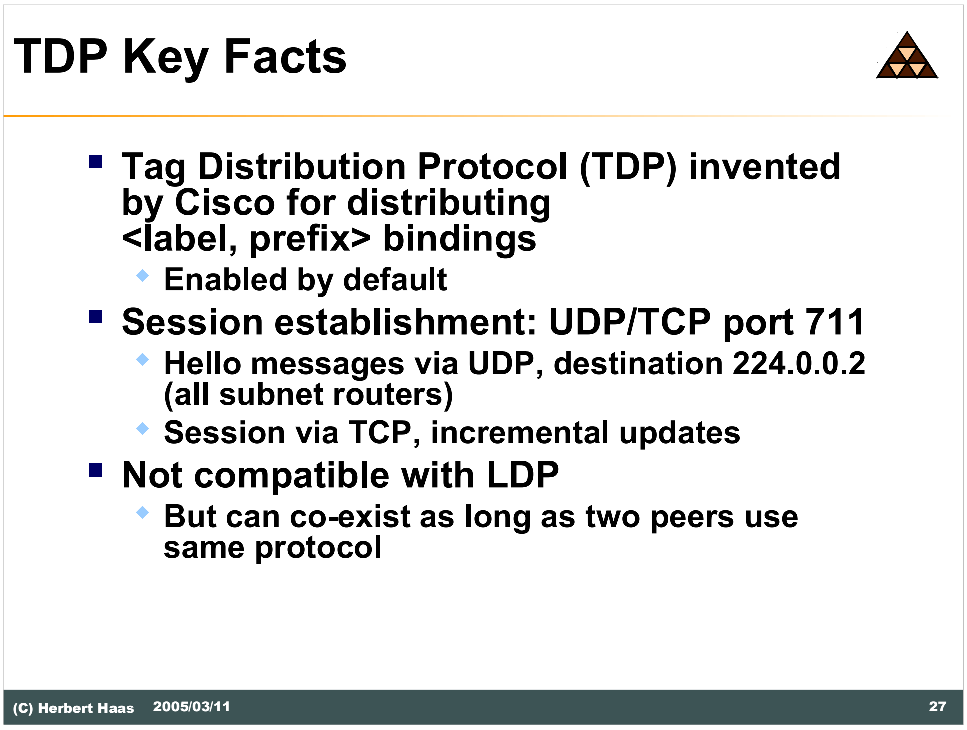
Note that routers performs a **per-platform label allocation**. That is, the LFIB does not contain any incoming interface, so the label must be unqiue on the entire router for a given destination. In other words, the same label can be used for a packet on any interface and will be forwarded to the same destination—this is the positive version.

Which label distribution and retention behavior is used depends on the interface type in use.

Unsolicited label distribution means that labels are advertised automatically without being asked...

Liberal label retention: All advertised labels are accepted, even from LSRs which are not next hop to the destination.

Conservative label retention: Advertised labels are only accepted from LSRs which are next hop LSRs for a given destination.



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

(C) Herbert Haas 2005/03/11 27

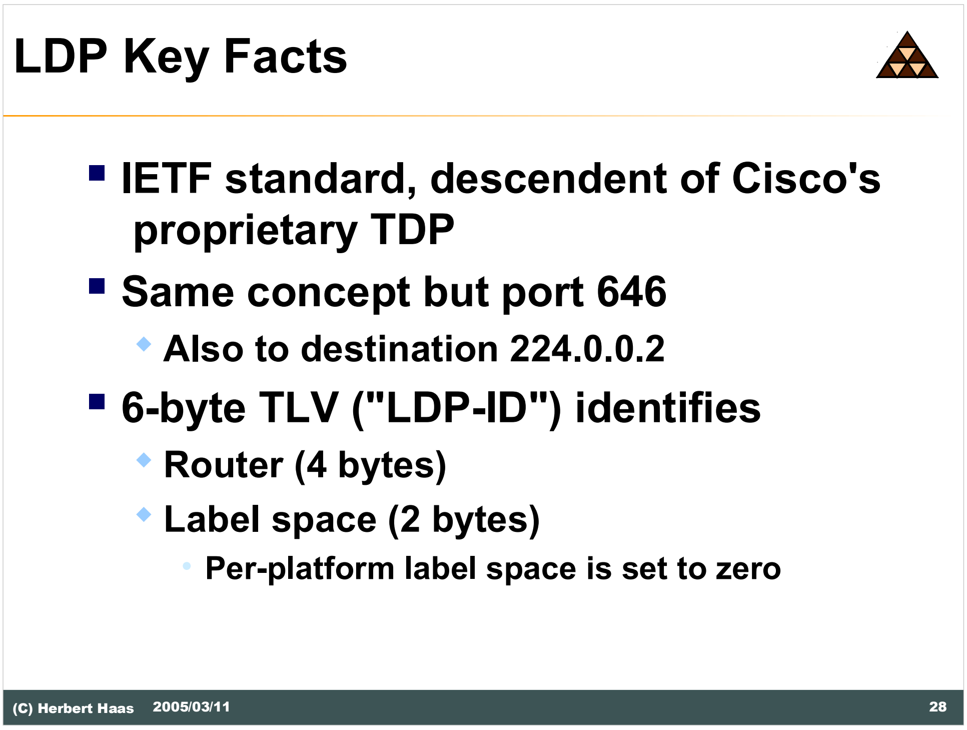
The TDP protocol was developed by Cisco and is used to distribute Lable-Prefix bindings between adjacent LSRs. Only in the case of MPLS TE TDP updates are also exchanged between not adjacent LSRs through so called Tunnel interfaces.

The TDP protocol is using both UDP and TCP at the transport layer. The TDP server process is addressed by the port number 711 and the updates are sent using the well known all routers Multicast address 224.0.0.2.

UDP is used in combination with a Hello procedure to detect neighboring LSRs.

The TCP protocol is used to reliable transport label binding information.

TDP is incompatible with LDP so neighboring LSRs need to use the same Protocol to allow a TDP/LDP session to come up.



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

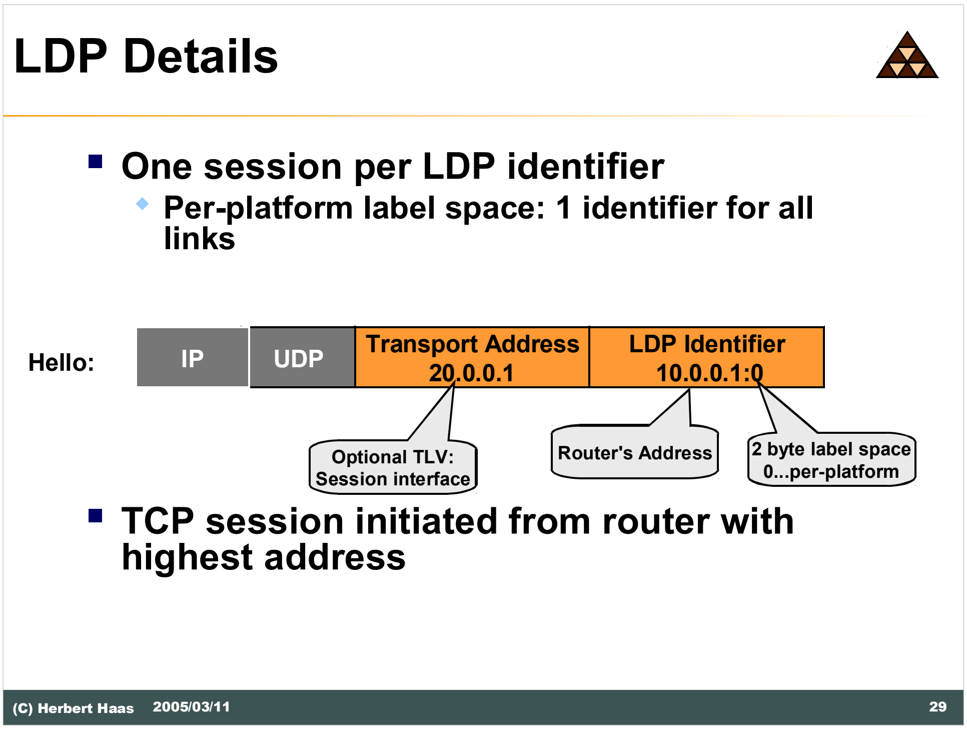
The LDP protocol is the standard protocol specified by the IETF. It works the same way like TDP does but they are incompatible as you can see just by the port numbers in use.

Reference: draft-ietf-mpls-ldp-07.txt

Combination of frame-mode and cell-mode (or multiple cell-mode) links result in multiple LDP sessions.

An LDP session is established by the router with the higher IP address.

Non-adjacent neighbors are discovered by unicast messages.



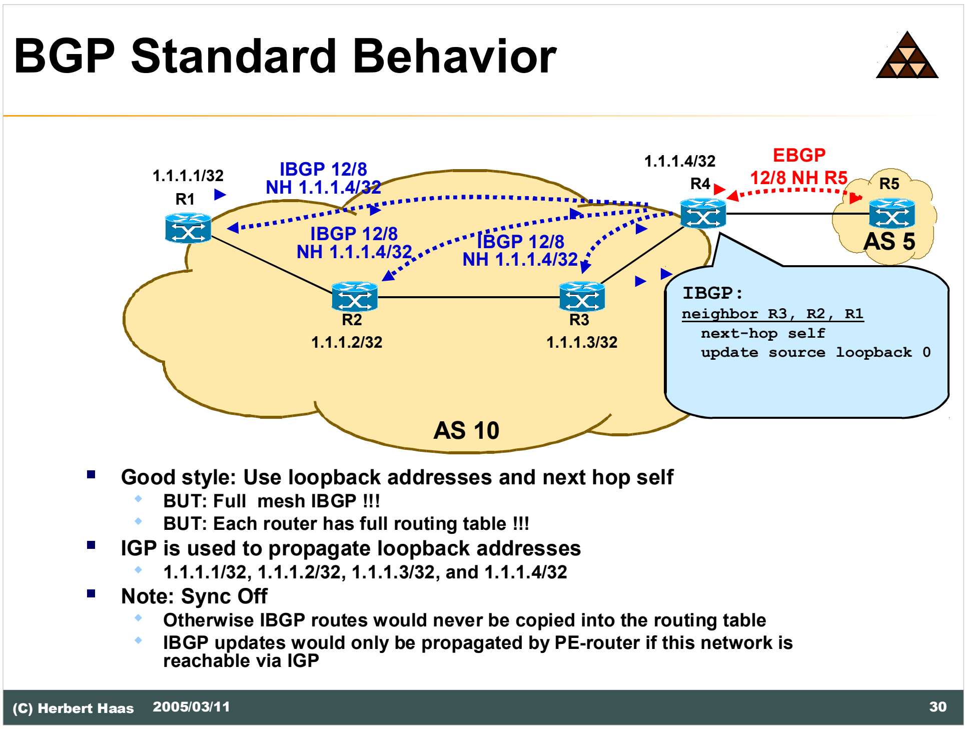
LDP Details

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

- TCP session initiated from router with highest address

Also **non-adjacent** LDP or TDP sessions can be established. In this case unicast addresses are used instead of multicast (for the hello packets).

Note: MPLS is enabled per interface. TDP is used by default on Cisco routers. If the router works in a mixed environment, enable both LDP and TDP for best interoperability.



BGP Standard Behavior

1.1.1.1/32

R1

IBGP 12/8
NH 1.1.1.4/32

IBGP 12/8
NH 1.1.1.4/32

IBGP 12/8
NH 1.1.1.4/32

1.1.1.4/32

R4

EBGP
12/8 NH R5

R5

AS 5

R2

1.1.1.2/32

R3

1.1.1.3/32

AS 10

```
IBGP:
neighbor R3, R2, R1
  next-hop self
  update source loopback 0
```

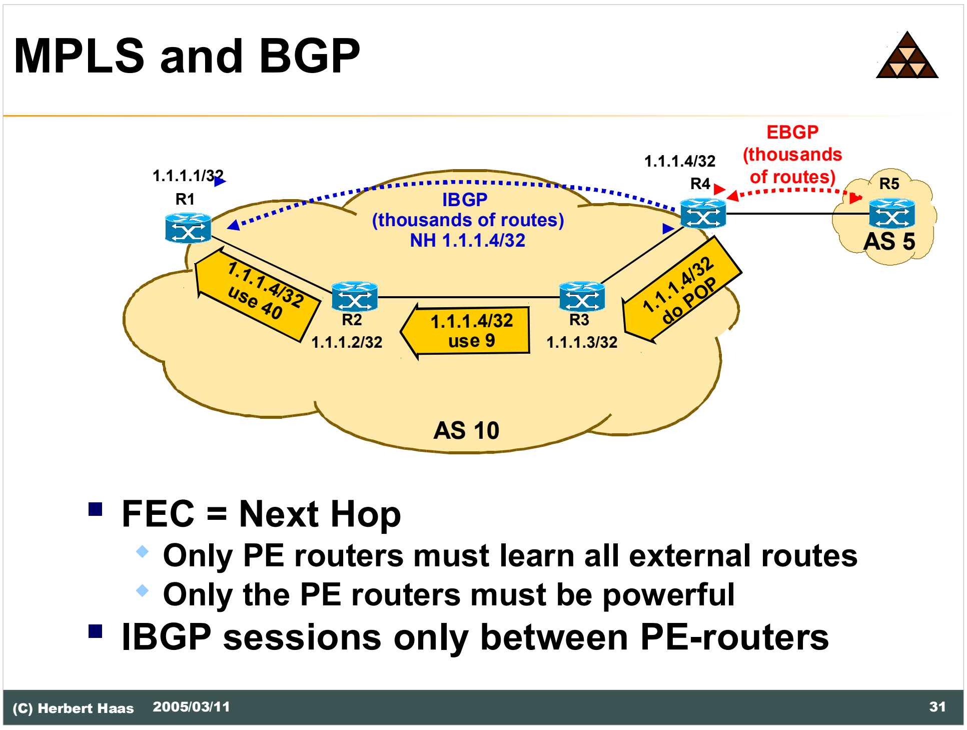
- Good style: Use loopback addresses and next hop self
  - BUT: Full mesh IBGP !!!
  - BUT: Each router has full routing table !!!
- IGP is used to propagate loopback addresses
  - 1.1.1.1/32, 1.1.1.2/32, 1.1.1.3/32, and 1.1.1.4/32
- Note: Sync Off
  - Otherwise IBGP routes would never be copied into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

Note: Sync is on by default (Cisco). "Update source loopback" makes IBGP updates using the loopback address as source address of update messages.

Note: The loopback addresses are specified as neighbor addresses.

Note: Next-hop self is necessary for the PE-routers because BGP otherwise assumes R5 to be the next hop AND there is no label to R5 if the IGP was not started on the external link.

Do not summarize PE loopback addresses as it would break the label-switching path. Therefore it is a good practice to use host-route loopback addresses with subnet masks of 32 bits. Equivalently do not use next-hop-self on confederation boundaries as it would also break the label-switching path.



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

(C) Herbert Haas 2005/03/11 31

For IGP derived routes a FEC represents an IP destination network.

For BGP derived routes a FEC represents the BGP Next Hop attribute.

This means that all routes which are imported by an EBGP Peer into an autonomous system are reachable via one and the same Label which points towards the EBGP Peers loopback address in the case NEXT HOP SELF is used on the EBGP Peer.

Therefore P routers don t need to run BGP because they are able to forward packets for external locations using the Label information derived from the EBGP Peers loopback address.

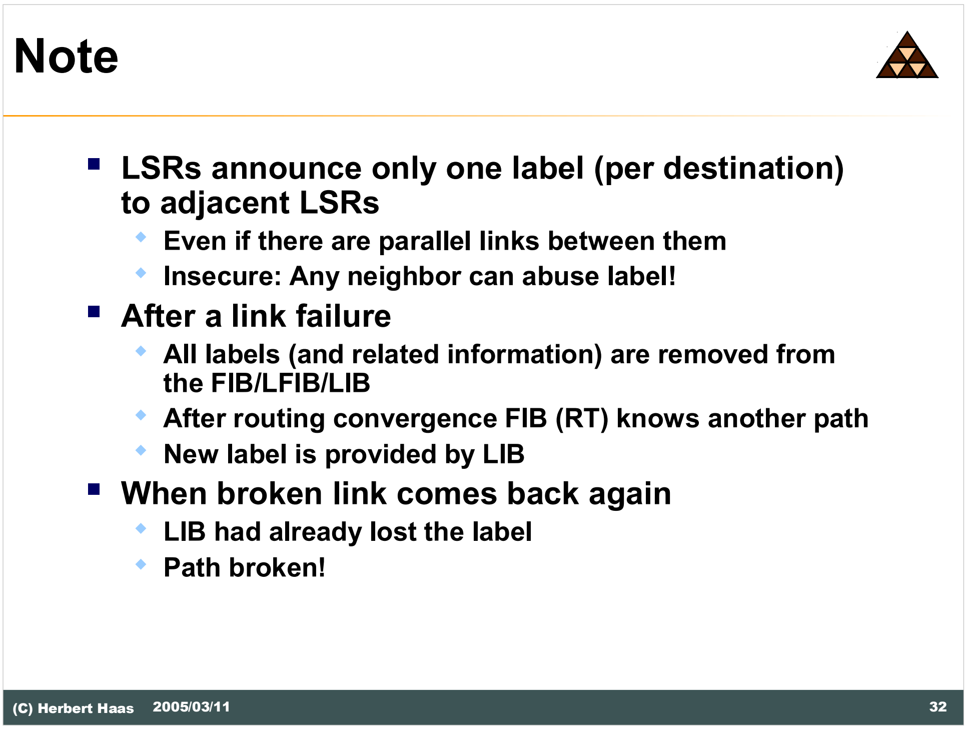
Advantages summary:

The BGP *topology* has been much *simplified*—only the AS edge routers need to run BGP with full Internet routing.

Core routers do not require much *memory*. The Internet routing table (by 2002) comprises about 100,000 routes which may require more than *50 MB* of memory for the *BGP table, IP routing table, and CEF's FIB table and distributed FIB tables*).

Changes in the Internet do not impact core routers!

Private (RFC 1918) addresses can be used inside the core. Note that in this case the *TTL propagation* must be *disabled*—otherwise a traceroute would show private addresses.



Note

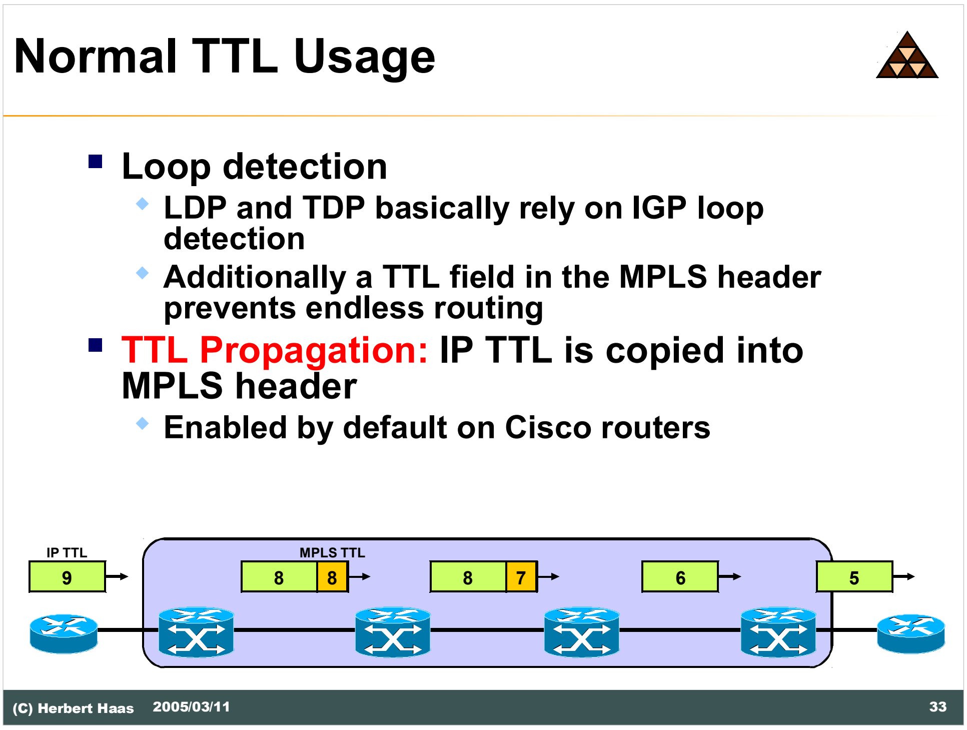
- LSRs announce only one label (per destination) to adjacent LSRs
  - Even if there are parallel links between them
  - Insecure: Any neighbor can abuse label!
- After a link failure
  - All labels (and related information) are removed from the FIB/LFIB/LIB
  - After routing convergence FIB (RT) knows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

(C) Herbert Haas 2005/03/11 32

LDP/TDP sessions are between routers not between interfaces—that's why label announcements are only sent once, even if there are parallel links between them. Therefore the LFIB is smaller and forwarding quicker.

But on the other hand, as the label is not bound to any interface, any neighbor can abuse the label and send a packet with this label to the router. The router does not (can not) check whether the packet had been received on the right interface for the given label.

Note that the label for a given destination is lost when a link is broken and comes back again. MPLS TE provides some measures against this.



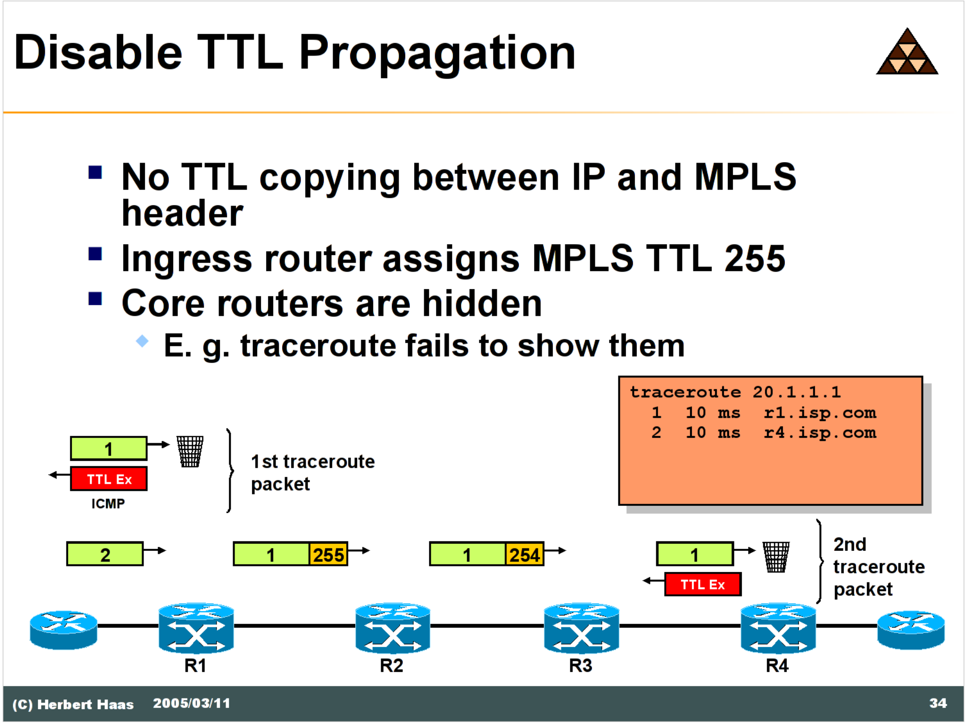
Normal TTL Usage

- 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

IGP protocols typically provide strong mechanisms to avoid routing loops. Nevertheless, the MPLS header carries a TTL field which provides additional protection against endless looping—for example caused by misconfigured static routes.

TTL Propagation: This mechanism is enabled by default (at least on Cisco routers) and ensures that the IP TTL value is also processed inside the MPLS domain. Actually, the IP TTL value is copied into the MPLS header. Within the MPLS domain only the MPLS TTL value is decremented.

Upon ingress, the IP TTL is copied to the MPLS header, upon egress the MPLS TTL is copied back to the IP header.



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

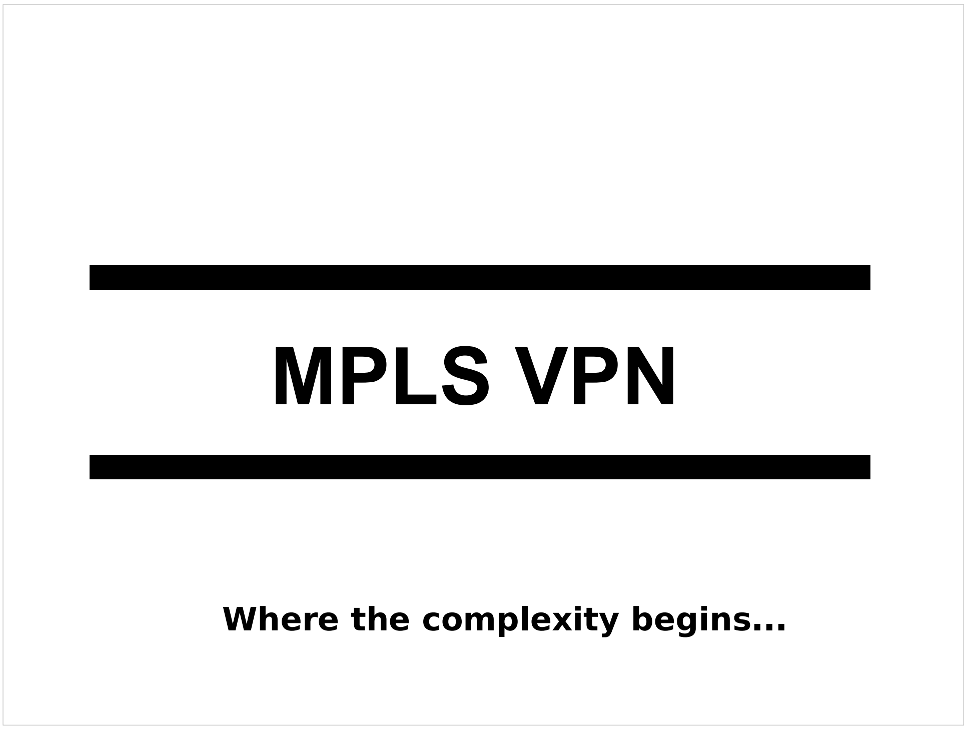
```
traceroute 20.1.1.1
  1  10 ms  r1.isp.com
  2  10 ms  r4.isp.com
```

(C) Herbert Haas 2005/03/11 34

As the example above shows, only the **ingress** and the **egress** LSRs are seen by traceroute.

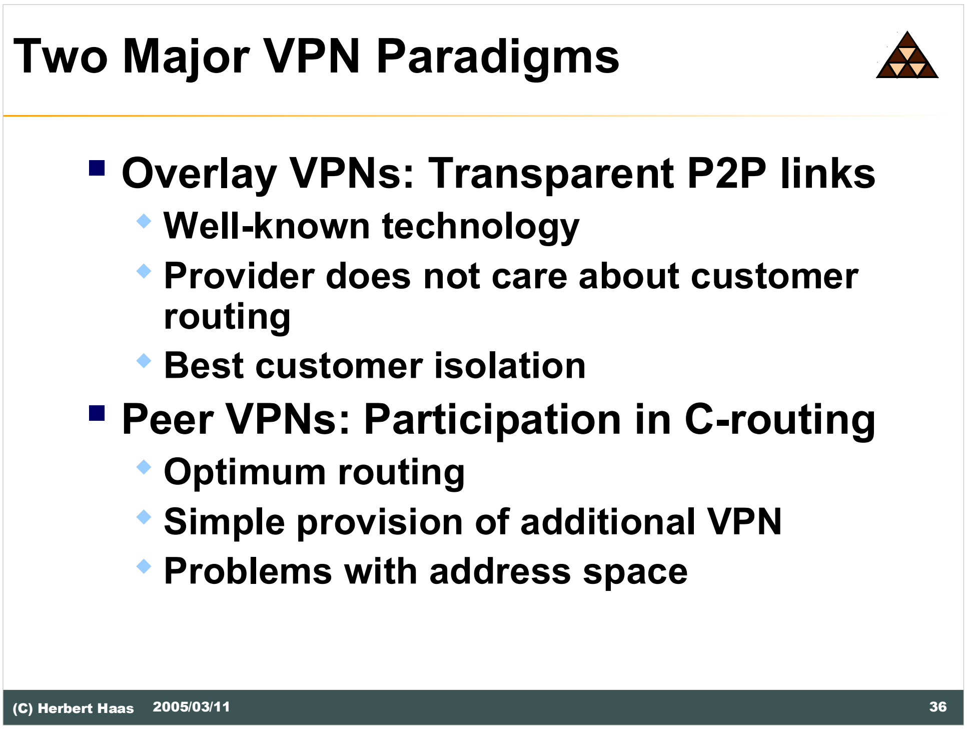
Note: If a traceroute would be started from any LSR (e. g. R1) every downstream router would be visible in the traceroute output. This is because TTL propagation can only be disabled for forwarded traffic. Traceroute from LSRs does not use the initial TTL value of 255.

Note: When TTL propagation should be disabled, it has to be disabled on **all** LSRs in the core! Frequently, ISPs forget to disable TTL propagation on some core routers. This typically lead to wrong traceroute results.



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

VPN services can be offered based on two major paradigms:

Overlay VPNs requires service providers to provide virtual point-to-point links between customer sites. The service provider does not see customer routes and is responsible only for providing point-to-point transport of customer data. All routing protocols run directly between customer routers.

Layer 1 solutions: Classical TDM technologies such as E1, ISDN, SONET/SDH.

Layer 2 solutions: FR, ATM, X.25.

Layer 3 solutions: IPsec, GRE whereas access (dialup) environments use L2TP, PPTP or L2F.

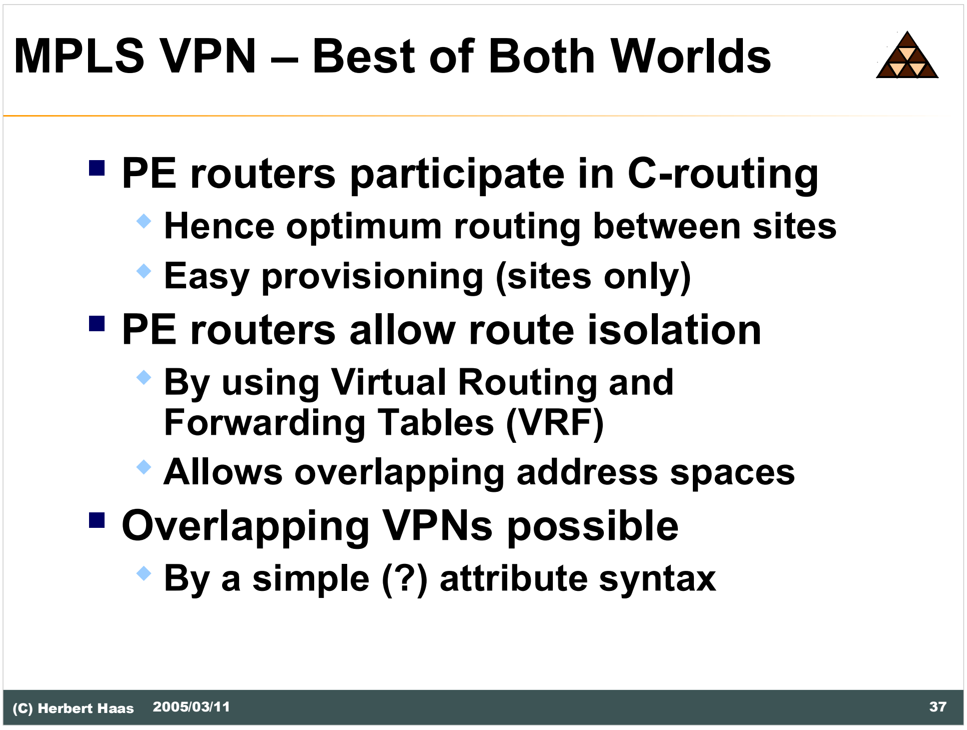
Peer-to-Peer VPNs requires service providers to participate in customer routing.

The isolation of the customers is realized via **packet filters** on PE routers at the PE-CE interfaces.

Another alternative is to implement **controlled route distribution** where each customer has a dedicated PE router which only knows about this customer's routes.

Peer VPNs allow a much **simpler provision** of additional VPNs because only the sites are provisioned, not the links between them.

Note: All customers **share** the same (provider-assigned or public) **address space.**



MPLS VPN – Best of Both Worlds

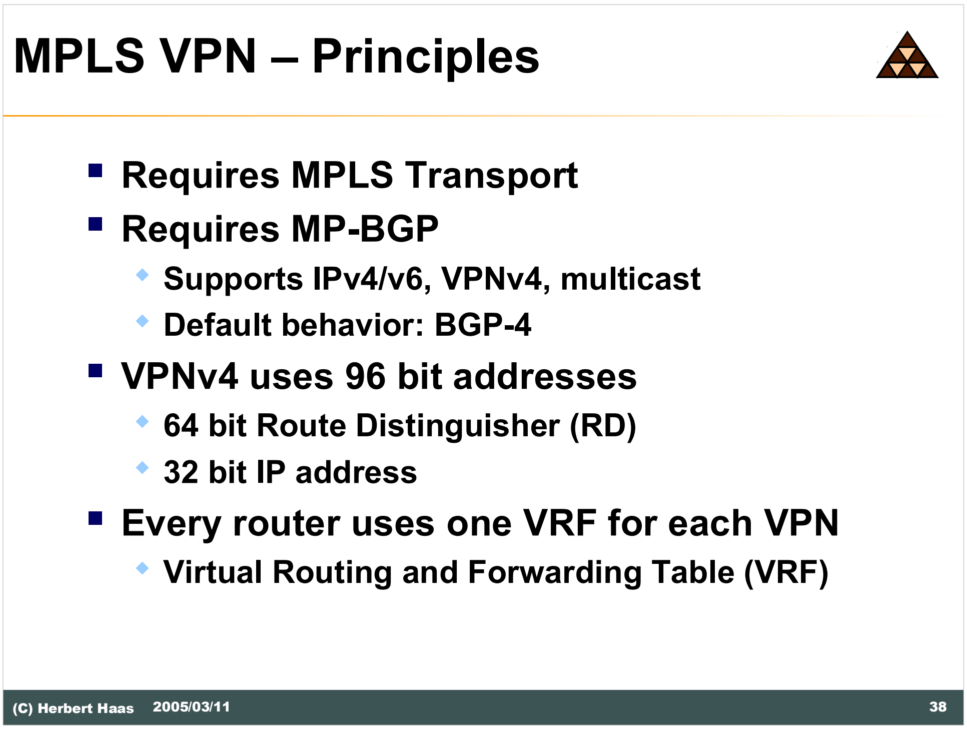
- PE routers participate in C-routing
  - Hence optimum routing between sites
  - Easy provisioning (sites only)
- PE routers allow route isolation
  - By using Virtual Routing and Forwarding Tables (VRF)
  - Allows overlapping address spaces
- Overlapping VPNs possible
  - By a simple (?) attribute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

The MPLS VPN solution combines the **best of both worlds** (overlapping and peer VPN).

Here the PE routers participate in C-routing which allows for **easy provisioning** and **optimum site-connections**. But the core routers do not need to carry much routing information. Only the PE routers must have some power.

Site isolation is provided by **Virtual Routing and Forwarding Tables (VRFs)** which are explained soon. This method allows for overlapping address space or overlapping VPNs (but not both together).

The main task is to specify which routes should be imported into which VRF. This is accomplished by special attributes during the configuration. The principle is easy (as you will see) but the attribute-syntax looks...strange (as you will see).



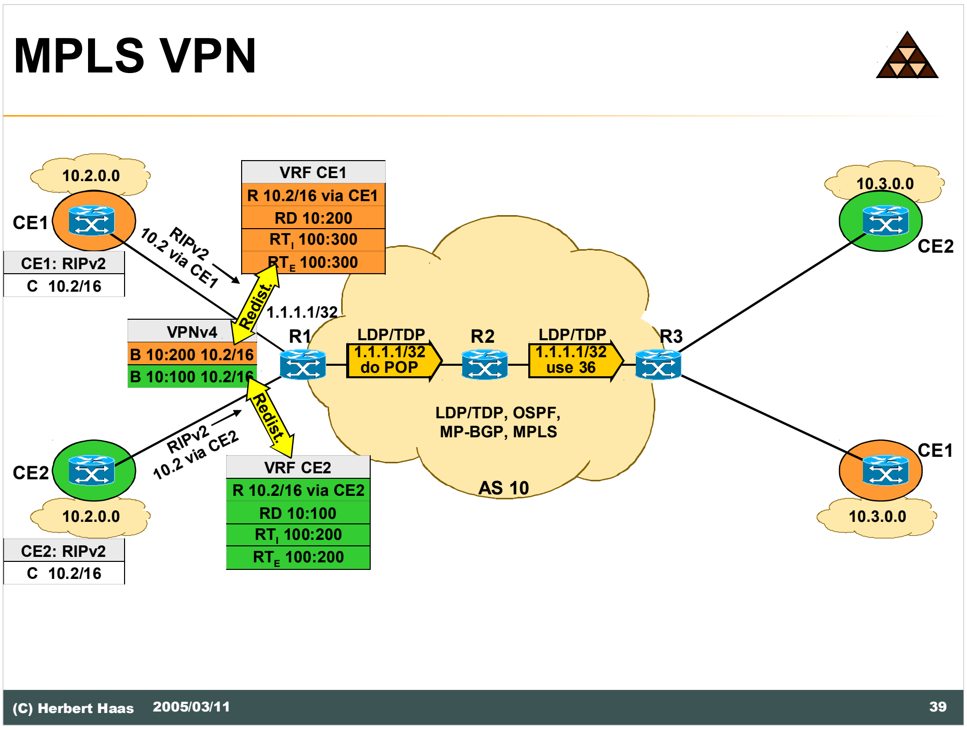
MPLS VPN – Principles

- 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)

(C) Herbert Haas 2005/03/11 38

For MPLS VPN services its **mandatory** to have an properly working MPLS Transport system already in place. Furthermore MP-BGP needs to be set up to allow the exchange of VPNV4 updates and VPN Label information.

A VPNV4 address is made up of a **64 bit Route Distinguisher (RD)** and a 32 bit IPV4 address. This VPNV4 address is needed to allow overlapping address spaces inside different VPNs. Every PE router holds different VRFs which holds address information for one or more VPNs, depending whether simple VPNs or overlapping VPNs are in use.



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

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

10.3.0.0

CE2

CE1

10.3.0.0

(C) Herbert Haas 2005/03/11 39

Each interface is exclusively member of the global routing process OR one VRF. The RD, RTi, and RTe are manually configured by the administrator. Each VRF has configured exactly one RD , but can have one or more RTi and RTE. The RD identifies each VPN (unless overlapping VPNs are configured). Routes for a VPNs are learned via an standard routing process running between the PE and the CE router such as RIPv2, OSPF, EIGRP and EBGP.

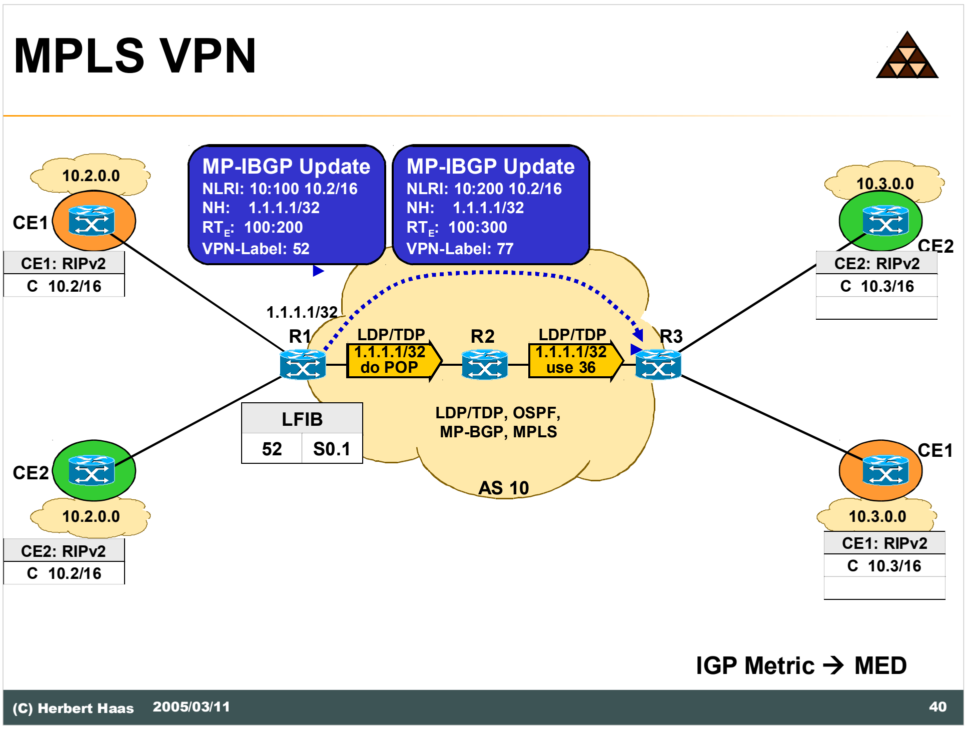
RIPv2, EIGRP or EBGP are good choices because a link state protocol such as OSPF would be limited to approx. 28 processes (theoretically a total of 32 routing processes). RIPv2, EIGRP and EBGP on the other hand can maintain many sub-processes, consuming only one process-number.

Bidirectional redistribution needs to be configured between MP-BGP and OSPF, RIPv2 and EIGRP, which copies the IGP information into the MP-BGP VPNv4 table and vice versa. Redistribution is not needed when EBGP is used as the PE-CE routing protocol.

Learned routes and the preconfigured RD is redistributed from the VRF tables into the MP-BGP VPNv4 table and since BGP makes triggered updates, this information is sent to the peers.

Note: the MP-BGP VPNv4 Table does not show the RTe, but the RTe is copied into to the BGP-database during the redistribution process.

MPLS VPN



MP-IBGP Update
NLRI: 10:100 10.2/16
NH: 1.1.1.1/32
RT_E: 100:200
VPN-Label: 52

MP-IBGP Update
NLRI: 10:200 10.2/16
NH: 1.1.1.1/32
RT_E: 100:300
VPN-Label: 77

CE1: RIPv2
C 10.2/16

CE2: RIPv2
C 10.2/16

CE2: RIPv2
C 10.3/16

CE1: RIPv2
C 10.3/16

LFIB	
52	S0.1

IGP Metric → MED

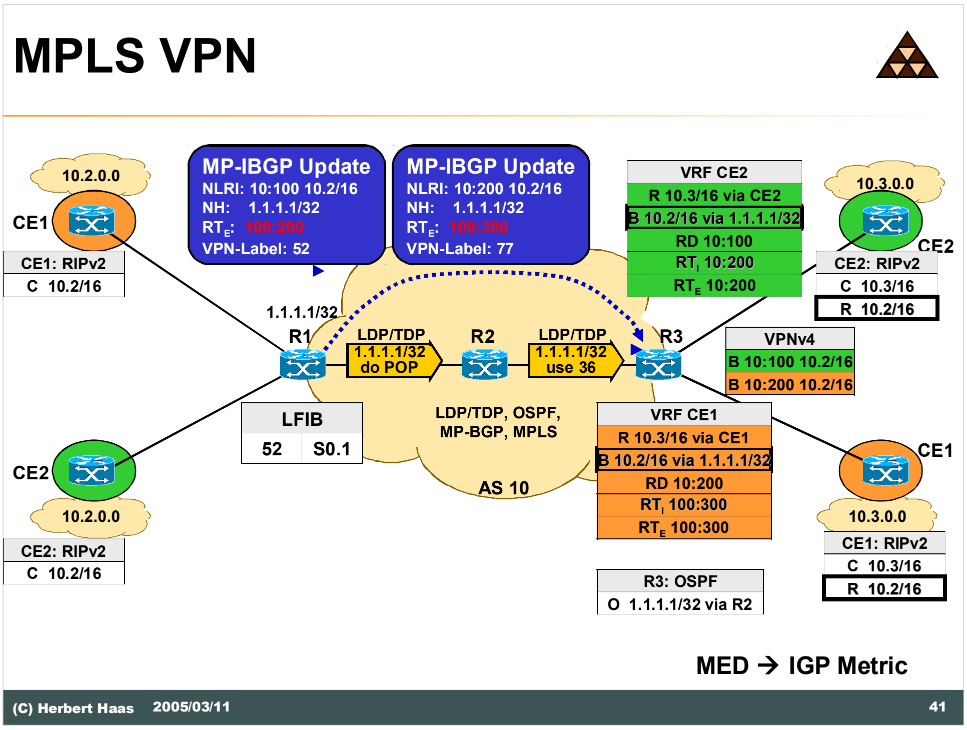
(C) Herbert Haas 2005/03/11 40

The RD together with the IPv4 address makes up the VPNv4 address which is propagated via MP-BGP updates. These VPNv4 addresses are now used in the NLRI fields of the BGP update instead of traditional IPv4 addresses. Also the RTe is carried with this update using extended community attributes as well as the VPN Label information.

The received MP-IBGP update is then imported into all VRFs which hold a matching RTi and optionally redistributed towards the connected CE routers. During the import from the VPNV4 table to the VRF the RD is removed resulting in a standard IPV4 address.

The IGP Metric (i. e. the RIPv2 hop count) is copied into BGP MED attributes, in order to carry this information to the other side.

MPLS VPN



MED → IGP Metric

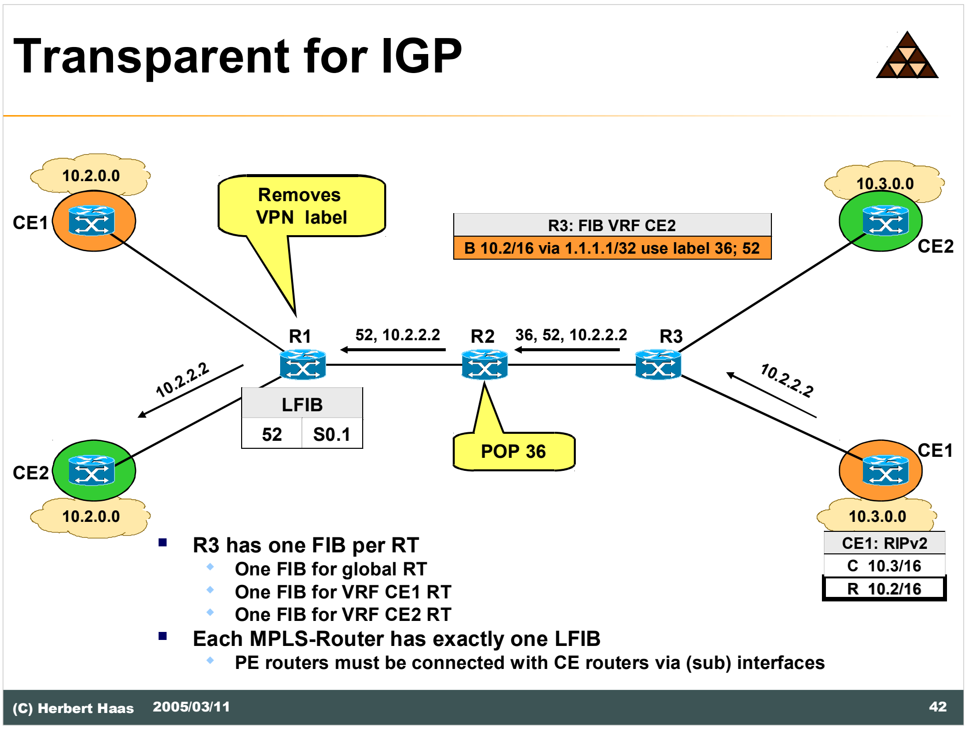
(C) Herbert Haas 2005/03/11 41

The RTi (import) is used locally by a VRF instance to determine which routes will be imported in the VRF-table and which not.

Routes are only copied into the VRF if the RTe matches the RTi. This route must be **redistributed** into the RIPv2 process.

Also a MPLS-label for this VPN is communicated via IBGP and is directly copied into the CEF table (FIB) of the peer PE router.

The MED attribute is copied into the hop-count field of the RIPv2 update. Thus, CE1 and CE2 on the right side learn about the metric which was specified on the other edge of the provider. The MPLS network is fully transparent to RIPv2 and only increases the IGP metric by one.



Transparent for IGP

R3: FIB VRF CE2
B 10.2/16 via 1.1.1.1/32 use label 36; 52

LFIB	
52	S0.1

CE1: RIPv2
C 10.3/16
R 10.2/16

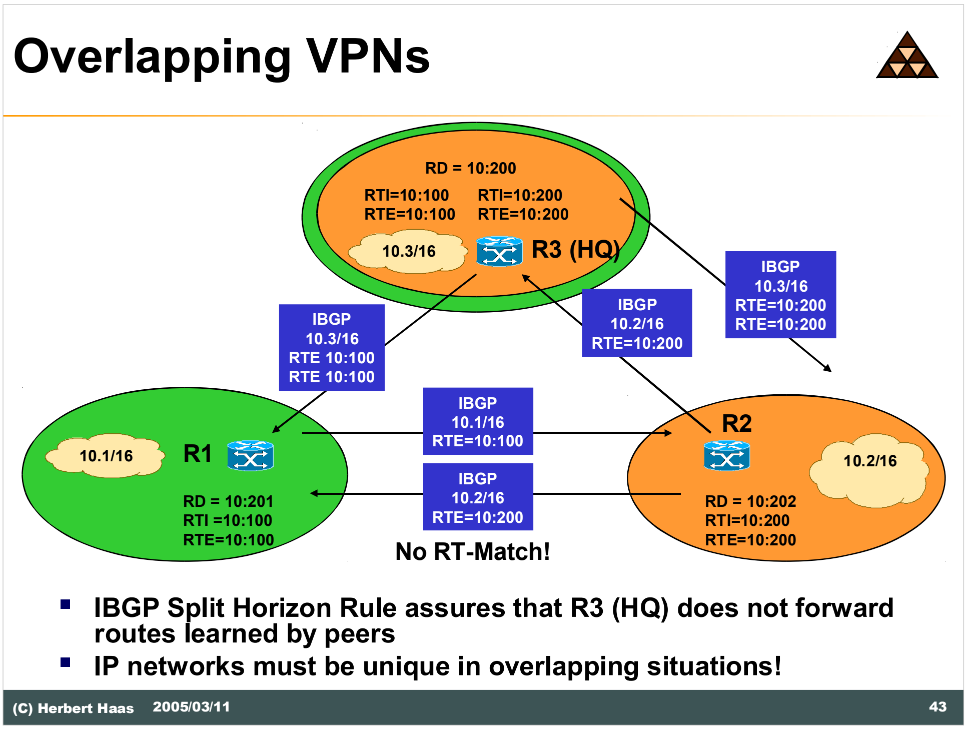
- R3 has one FIB per RT
  - One FIB for global RT
  - One FIB for VRF CE1 RT
  - One FIB for VRF CE2 RT
- Each MPLS-Router has exactly one LFIB
  - PE routers must be connected with CE routers via (sub) interfaces

(C) Herbert Haas 2005/03/11 42

Now IP packets can be forwarded between the VPNs. For example, IP packets to 10.2.2.2 are forwarded from the CE1 router (right side) to the next hop VRF-R3, which adds the labels {36; 52} into the MPLS header, according to its FIB.

R2 pops the MPLS-Transport header and R1 can quickly deliver the IP packet to the correct VPN according to the remaining VPN label {52} which is stored in the LFIB table at R1 pointing to the interface of the appropriate VPN.

R1 removes the MPLS-VPN label {52} before the IP packet is delivered to CE2 (left side). Thus, the VPNs do not recognize any MPLS network in-between; MPLS is completely transparent.



Overlapping VPNs

RD = 10:200

RTI=10:100 RTE=10:100

RTI=10:200 RTE=10:200

10.3/16

R3 (HQ)

IBGP 10.3/16 RTE 10:100 RTE 10:100

IBGP 10.2/16 RTE=10:200

IBGP 10.3/16 RTE=10:200 RTE=10:200

10.1/16

R1

RD = 10:201 RTI =10:100 RTE=10:100

IBGP 10.1/16 RTE=10:100

IBGP 10.2/16 RTE=10:200

No RT-Match!

R2

10.2/16

RD = 10:202 RTI=10:200 RTE=10:200

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

When using simple VPNs the RTi is equal to the RTe (keyword "both" when configuring) , but when overlapping VPNs are used, the Route Targets need to be different according to the desired communication behavior.

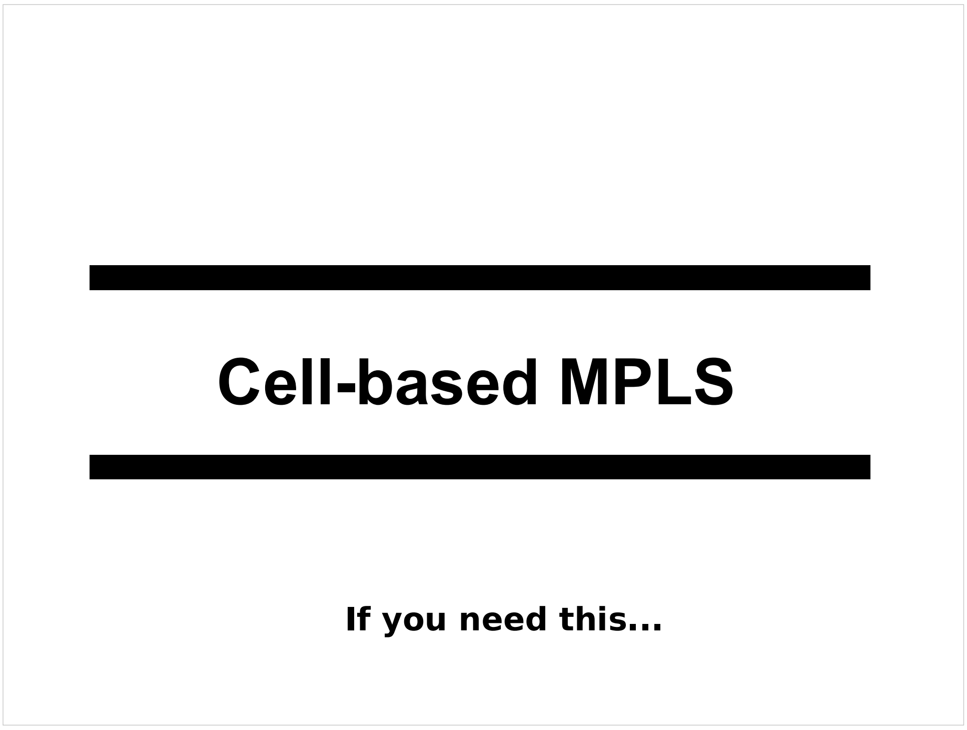
In our example all routes from the VPN-green and VPN-red are propagated to R3 (HQ) and copied into the VRF table due to the configured RTe and RTi values.

If R3 sends its update towards R1 and R3 all routes **(except routes learned from IBGP sessions)** out of R3s VRF are propagated to R1 and R2 with both RTEs attached. These routes are then imported by R1 and R2 into the appropriate VRF tables.

Due to the IBGP split horizon rule R3 does not propagate routes learned from R2 towards R3 and vice versa. So without the IBGP split horizon rule MPLS VPNs would not exist.

Note: Both RTi and RTe can be configured multiple times. For example one VRF on a router can have specified three different RTi values. Therefore, all IBGP updates whose RTe values match one of the specified RTi values can be imported.

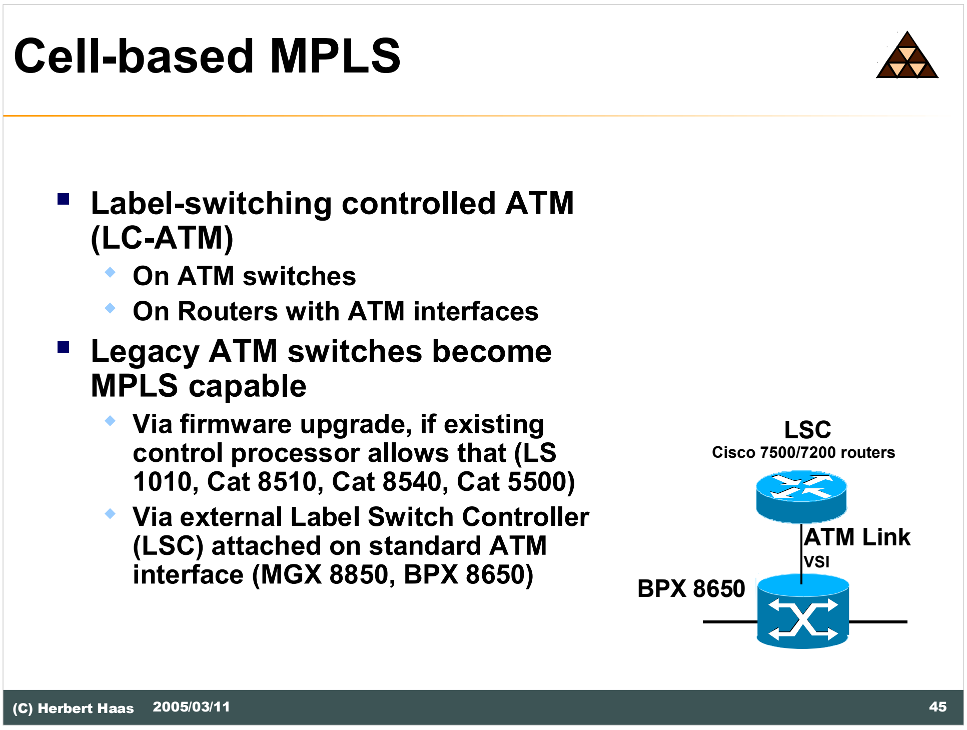
Note: some older IOS versions require that at least one RTi and one RTe are identical.



Cell-based MPLS

If you need this...

ATM? Try some, buy some...



Cell-based MPLS

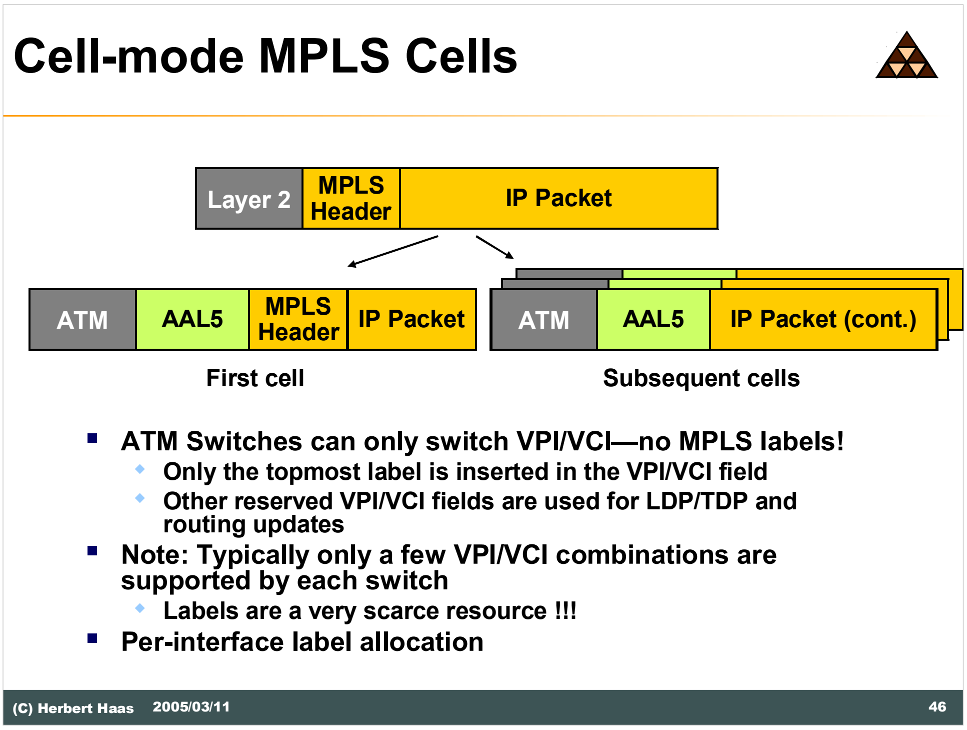
- Label-switching controlled ATM (LC-ATM)
  - On ATM switches
  - On Routers with ATM interfaces
- Legacy ATM switches become MPLS capable
  - Via firmware upgrade, if existing control processor allows that (LS 1010, Cat 8510, Cat 8540, Cat 5500)
  - Via external Label Switch Controller (LSC) attached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

LSC
Cisco 7500/7200 routers
ATM Link
VSI
BPX 8650

Enabling cell-based MPLS on Cisco IOS-based ATM switches is identical as enabling frame-based MPLS on IOS routers. When enabling cell-based ATM on IOS routers with ATM interfaces, the command **interface atm X/X/X tag-switching** must be used. The keyword tag-switching here reserves the VC 0/32 for control messages.

LSC is available for Cisco BPX switches. A special **Virtual Switch Interface (VSI)** protocol is used between the standard ATM interface and the LSC. The VSI basically only supports VC additions and deletions. All higher MPLS operations are performed by the LSC using VC 0/32.

One main advantage of Cell-mode ATM is to avoid NSAP addressing (and mapping) which is needed to run PNNI.



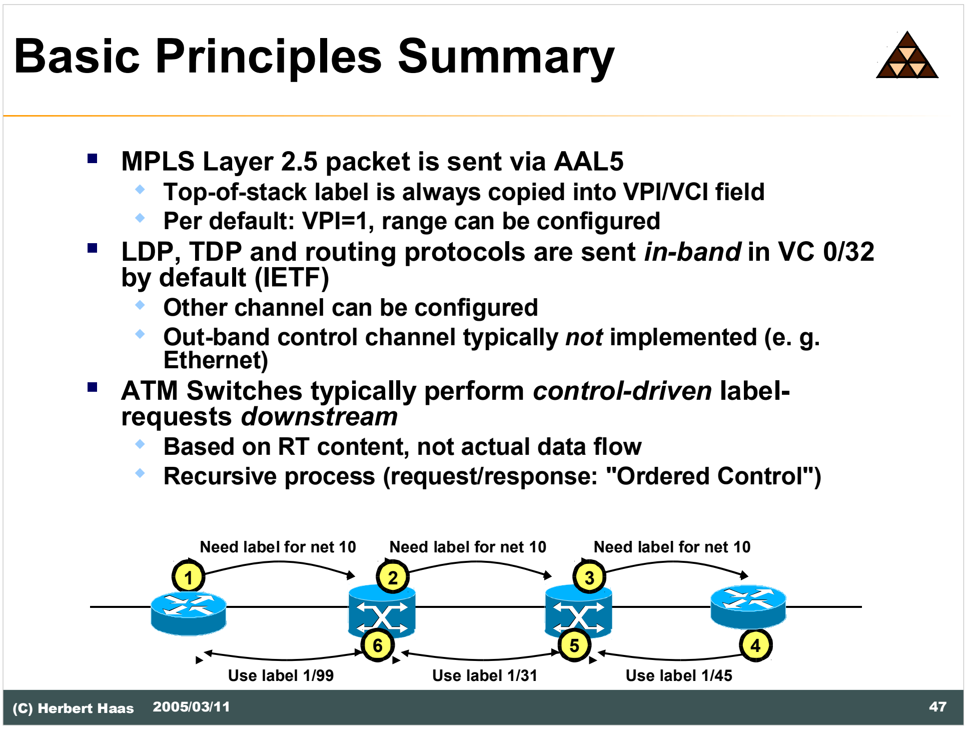
Cell-mode MPLS Cells

- 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

(C) Herbert Haas 2005/03/11 46

The top label is always copied into the (VPI/) VCI fields. LDP/TDP sessions are established via reserved VPI/VCI labels. Typically a ATM switch only provides a few VPI/VCI numbers, so it is difficult to adapt all MPLS labels used in a router network.

Note that LC-ATM provides a per-interface label allocation since the ATM switching matrix (= LFIB) always contains the incoming interface! That is, same labels can be reused on different interfaces on the same machine. This has a security advantage: Labeled packets are only accepted on that interfaces where the labels had been previously assigned.



Basic Principles Summary

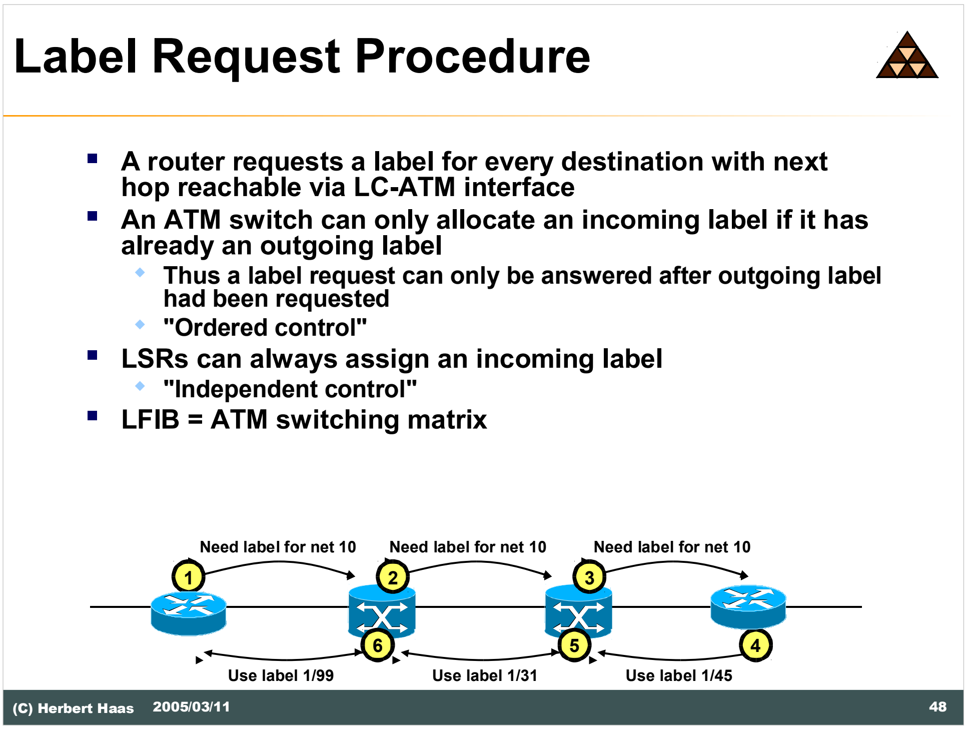
- MPLS Layer 2.5 packet is sent via AAL5
  - Top-of-stack label is always copied into VPI/VCI field
  - Per default: VPI=1, range can be configured
- LDP, TDP and routing protocols are sent *in-band* in VC 0/32 by default (IETF)
  - Other channel can be configured
  - Out-band control channel typically *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")

The main difference between frame-based MPLS in routers and cell-based MPLS is the following: Routers can handle both IP packets (LDP, TDP, routing updates) and labeled-packets (MPLS data packets on layer 2.5). But ATM switches can ONLY handle VPI/VCI-labeled packets.

As the top-of-stack MPLS label is now always used in the VPI/VCI field, there must be a dedicated VC for control packets such as LDP, TDP, and routing protocols.

Per default, only the 16-bit VCI value carries the label value. Note that VPI values are a scarce resource. Therefore the VPI value is set to 1 per default. Optionally, a VPI range can be specified.

The MPLS control VC is by default configured on VC 0/32 and must use LLC/SNAP encapsulation of IP packets as defined in RFC 1483. The corresponding IOS keyword is aal5snap.

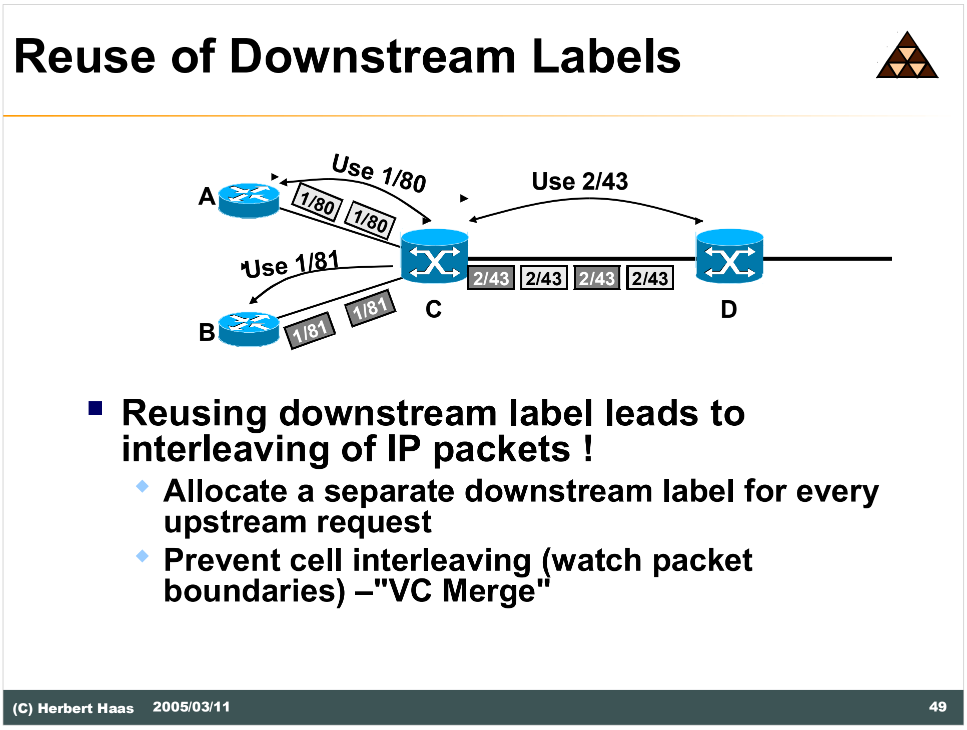


Label Request Procedure

- 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

Labels are requested via LDP/TDP as soon as an edge router (LSR) learns about a destination which is reachable via a next hop through a LC-ATM interface.

Each ATM-LSR can only allocate a label for this (requested) destination when it knows an outgoing label already. Therefore the response message must be delayed and another label request is sent downstream. Only when the last LSR on the right side, (or ATM-LSR which is the egress ATM LSR and needs L3 functionality) receives the request, it allocates a label and sends a response to the label request. Note that this last (egress) ATM LSR has no outgoing label as it is directly connected with the destination network. We assume that "net 10" is located at the right side next to the rightmost LSR.



Reuse of Downstream Labels

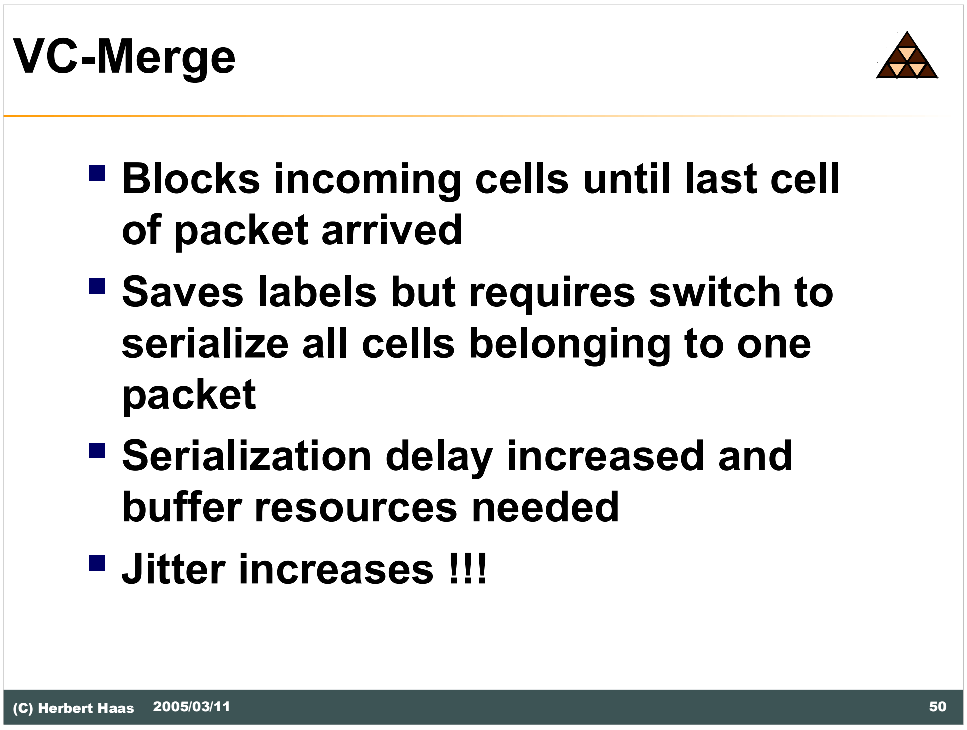
- **Reusing downstream label leads to interleaving of IP packets !**
  - **Allocate a separate downstream label for every upstream request**
  - **Prevent cell interleaving (watch packet boundaries) –"VC Merge"**

(C) Herbert Haas 2005/03/11 49

Note the difference to the old AAL5 problem: All cells belonging to one AAL5 IP packet are not interleaved with the cells of another IP packet received on the same interface—that is: from the same source (having the same VPI/VCI). But a switch may indeed interleave the cells of different VPI/VCIs. The only problem occurs if some cells are lost, especially the last cell which indicates packet boundaries.

The problem illustrates above involves two sources (A and B) whose cells are switched downstream with the same label. This is possible in a normal MPLS network which consists of routers only! But with LC-ATM the packets would be interleaved and cannot be reassembled correctly anymore.

Therefore, two solutions are implemented: Avoid cell interleaving (and assure packet interleaving only) or allocate separate downstream labels for every upstream request.



VC-Merge

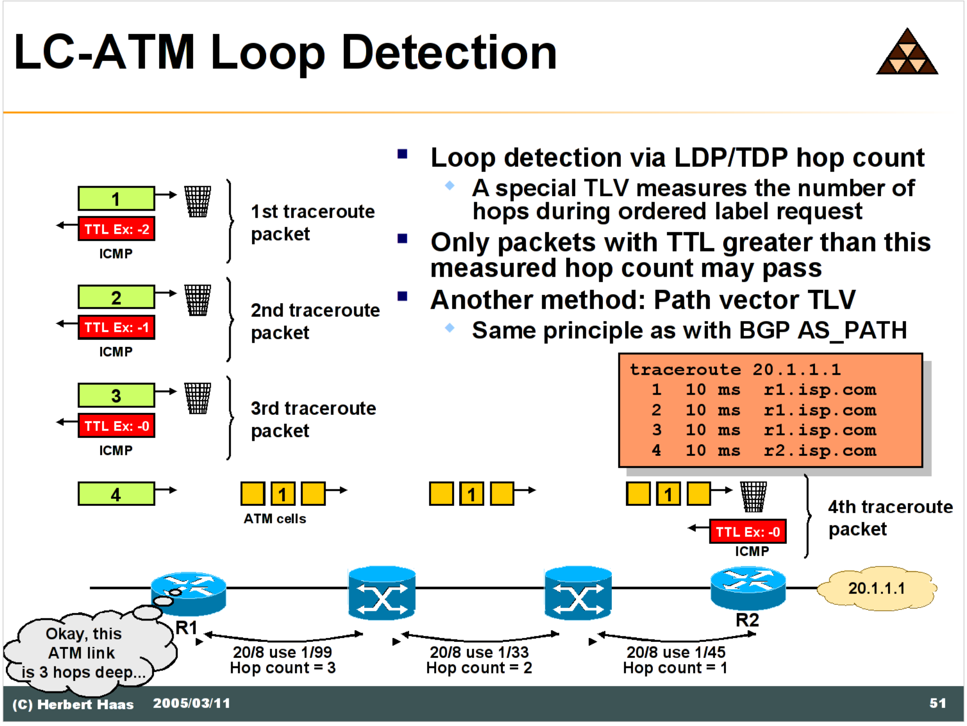
- Blocks incoming cells until last cell of packet arrived
- Saves labels but requires switch to serialize all cells belonging to one packet
- Serialization delay increased and buffer resources needed
- Jitter increases !!!

(C) Herbert Haas 2005/03/11 50

AAL5 only marks the end-cell of a IP-packet. Therefore it is not possible to aggregate several MPLS-VCs into one VC using a unique label because as cells are interleaved, subsequent switches cannot reassemble the IP packets.

If switches support "**VC Merge**" then they are capable to buffer all cells belonging to one IP packet and send them at once. That is, the switches avoid to interleave cells of different IP packets.

But most implementations **block** all other interfaces in the meanwhile! Then the forwarding delay of a complete packet depends on concurrent packets. **Jitter** occurs! This solution transforms the cell-based ATM network in a classical frame-based network!



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

```
traceroute 20.1.1.1
  1   10 ms   r1.isp.com
  2   10 ms   r1.isp.com
  3   10 ms   r1.isp.com
  4   10 ms   r2.isp.com
```

(C) Herbert Haas 2005/03/11 51

Loop detection relies primarily upon IGP mechanisms but additionally, there is a special TLV for LDP hop count.

When the "already explained control-driven ordered label request" is performed, the LDP protocol carries a **special TLV** which **counts** all hops between egress and ingress device.

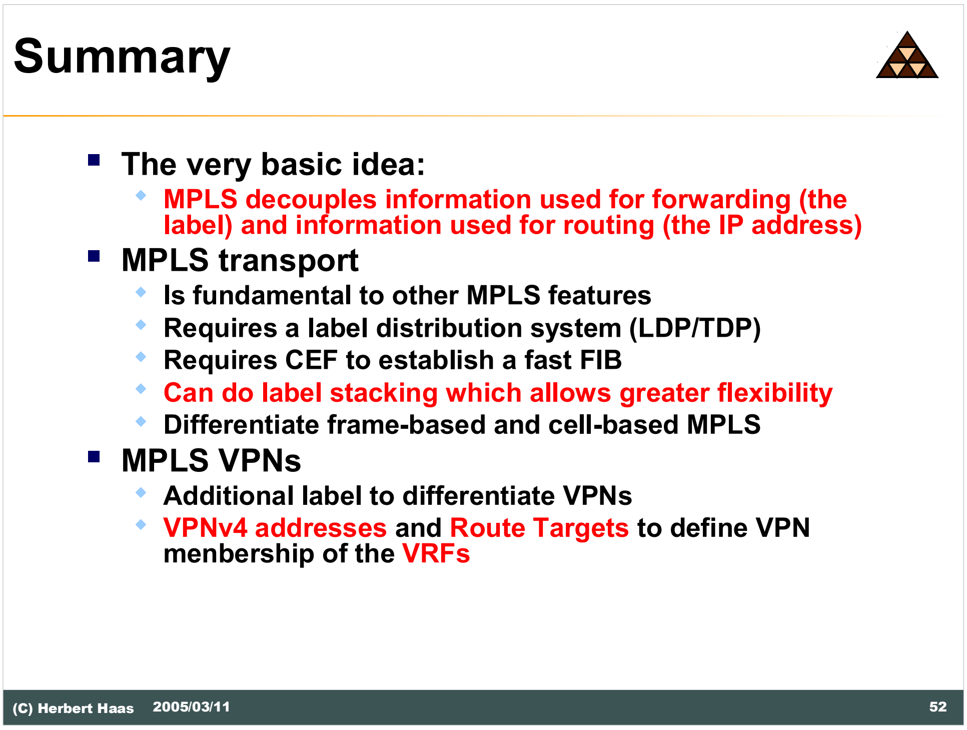
Now the ingress router R1 knows the number of hops that must be passed until the egress router R2 is reached.

When a traceroute packet arrives, R1 **subtracts** the measured hop count from the given IP TTL value of the packet. If the result is zero, or less the packet is discarded. Only if the result is one or greater, the packet is forwarded.

But note: if the result (of the TTL subtraction) is only one, the packet is indeed forwarded but only reaches the egress router R2 where the TTL is once more reduced and the packet is dropped. Obviously, in order to surpass the LC-ATM domain, the TTL must be greater than the measured hop count plus one.

Note: The maximum number of hops can also be specified for LDP.

The **Path vector TLV** is another LDP loop detection mechanism. This TLV contains all intermediate ATM LSRs IDs in the path. When the sending LSR detects its own ID within the vector it ignores the message. This is the same principle as with BGP's AS_PATH attribute.



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