MPLS
Multiprotocol Label Switching

Prof. Rami Langar
LIGM/UPEM
Rami.Langar@u-pem.fr
http://perso.u-pem.fr/~langar

Outline

- Routing vs. switching
- Label Switching properties
- IPOA
- IP switching
- Multi Protocol Label Switching (MPLS)
- Label Distribution Protocol
- Traffic Engineering

Two forwarding paradigms

- Station breaks long message into packets
- Packets sent one at a time to the network
- Packets handled in two ways
  - Datagram (e.g., IP routing)
  - Virtual Circuit (e.g., ATM)

Datagram (e.g., IP routing)

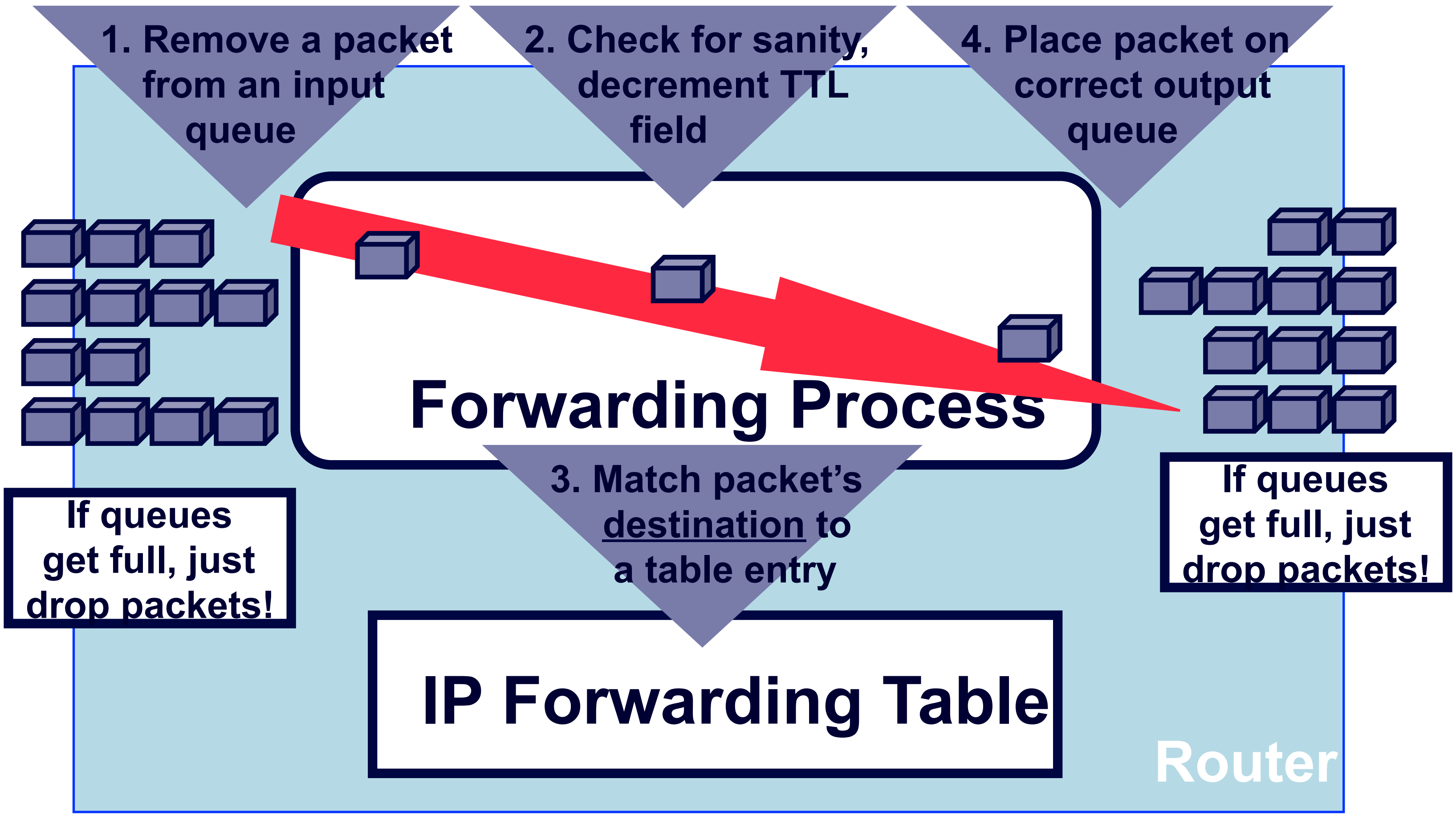
- Each packet treated independently
- Packets can take any practical route
- Packets may arrive out of order
- Packets may go missing
- Up to receiver to re-order packets and recover from missing packets

Conventional IP routing

- Router have to execute a forwarding decision for every received IP packet :
  - Routing algorithm execution (performed in advance),
  - Packet reception from the input queue,
  - Header analysis (Checksum, TTL, @IP, …),
  - Access control (optional),
  - Send the packet to the output queue.

=> Important queue congestion if all packets are routed in the same way.

Conventional IP routing



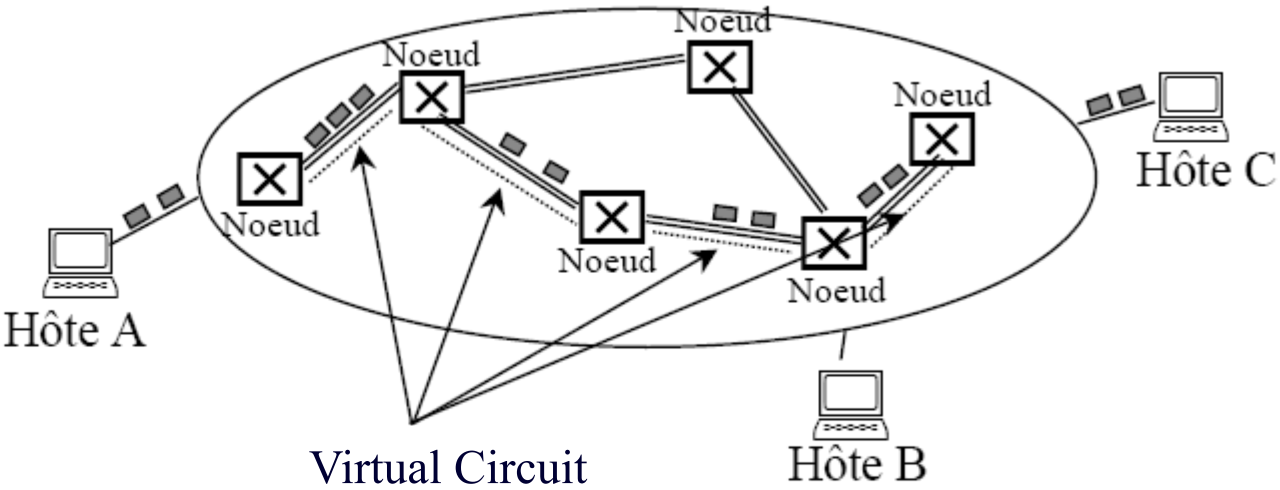
Conventional IP routing

- Advantages:
  - IP is the first defined and used protocol,
  - IP is the only protocol for global Internet working
    - Deployed in all machines since mid 90's
  - Connectionless: Simplicity, no need for signalization
  - Flexibility (adaptive routing),

- Disadvantages:
  - No QoS support,
  - High routing table size
  - Large IP Header (At least 20 bytes)
  - Poor performance (in early 2000's 10^5 packets/s VS 10^8 today)
  - Usually designed to obtain shortest path,
    - Do not take into account additional metrics.

Virtual Circuit (e.g., ATM)

- Preplanned route established before any packets sent
- Each packet contains a virtual circuit identifier instead of destination address
- No routing decisions required for each packet
- Three steps:
  - Establishment of the connection,
  - Data transfer,
  - Connection release.



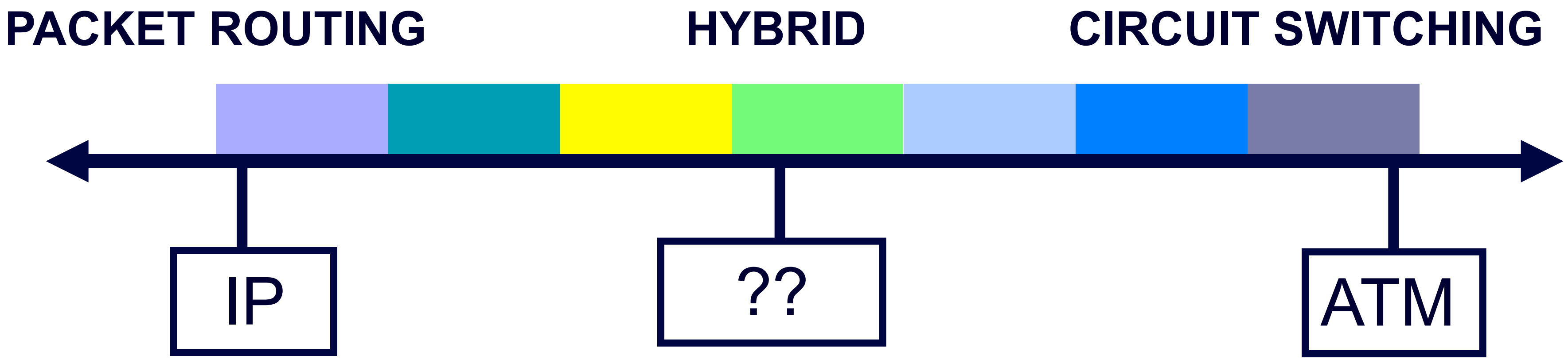
ATM

- Advantages :
  - High performance – 20 millions of cells/s,
  - Connection mode – Supports QoS,
  - Fast packet switching with fixed length packets (cells)
  - Integration of different traffic types (voice, data, video)

- Disadvantages:
  - Complex (More complex than routing),
  - Expensive
  - Need for signaling (Q.2931 for UNI and PNNI within ATM core networks)
  - Additional delay to establish the virtual path
  - Not widely adopted.

Best of both worlds

- Idea: Combine the forwarding algorithm used in ATM with IP.
  - The aim of IP switching is to combine L3 functionalities (simplicity, flexibility, …) with the high performance of L2 (i.e., switching is faster than routing).
  - The gain is considerable: in switching mode, the delay is 10 microsec ; 10 to 15 higher in routing mode. Instead of having a rate of 300 000 packets/s with a classic router, we obtain 3 to 15 millions of packets/s with a switch.

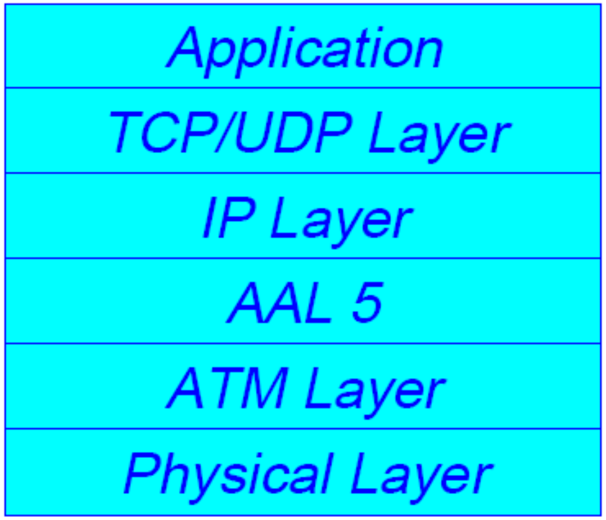


Solutions

- First approach : IP over ATM (the oldest solution)
  - Supports TCP/IP applications, which are highly used before the development of ATM.
  - Allows interoperability between ATM networks and existing networks that use IP.
    - LAN Emulation (ATM Forum)
    - MPOA (ATM Forum)
    - IPOA (IETF)
- Second approach : Label switching
    - IP switching
    - TAG switching
    - MPLS (IETF)

IPOA

- Classical IP over ATM (IETF RFC 1577 et 1483)
  - Interconnect several IP LANs using ATM links : conversion of IP @ to ATM @ and vice-versa,
  - LAN stations form Logical IP subnets (LIS),
  - Use of AAL5 to encapsulate IP packets within the ATM layer,
  - Use of ATMARP servers for the translation between IP @ and ATM @,
  - Then, establishment of VC between stations,
  - No possible broadcast within a LIS.



Pre-MPLS

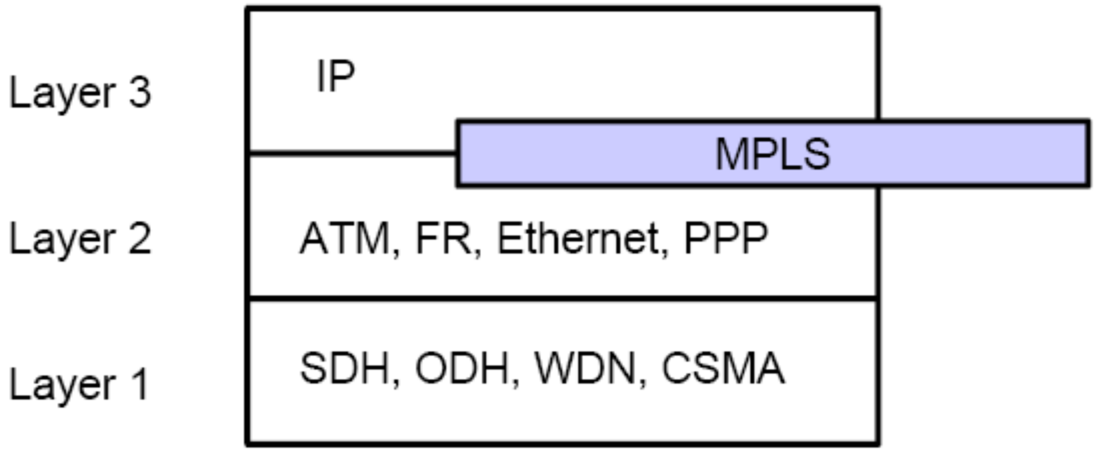
- IP-switching:
  - Proposed by IPSILON (Nokia) in 1994 and based on ATM,
  - Route determined by IP flow,
  - IP-switch nodes replace IP routers: first packet routed normally to determine the VP. Then, packets within the same flow are switched,
  - IP switch is an hybrid IP router/ATM switch.
- TAG-switching:
  - Proposed by CISCO in 1997,
  - Support other L2 technologies: ATM, Ethernet, PPP, …
  - Support resource reservation using RSVP
  - Basis for MPLS: Label switching.

MPLS

- Multi-Protocol Label Switching : IETF Working Group established in April 1997.
- Objective: develop a standard for IP switching and working on different types of networks such as ATM, Frame Relay, Ethernet, etc.
- Integrate the proposals of different companies:
  - TAG switching : CISCO (Ethernet),
  - IP switching : IPSILON (IP over ATM),
  - ARIS : IBM (IP over ATM, FR)
- Independent from routing protocols (RIP, OSPF, …),
- Can be used with IPv4 and IPv6.

MPLS : Bases

- Multi Protocol Label Switching is arranged between Layer 2 and Layer 3

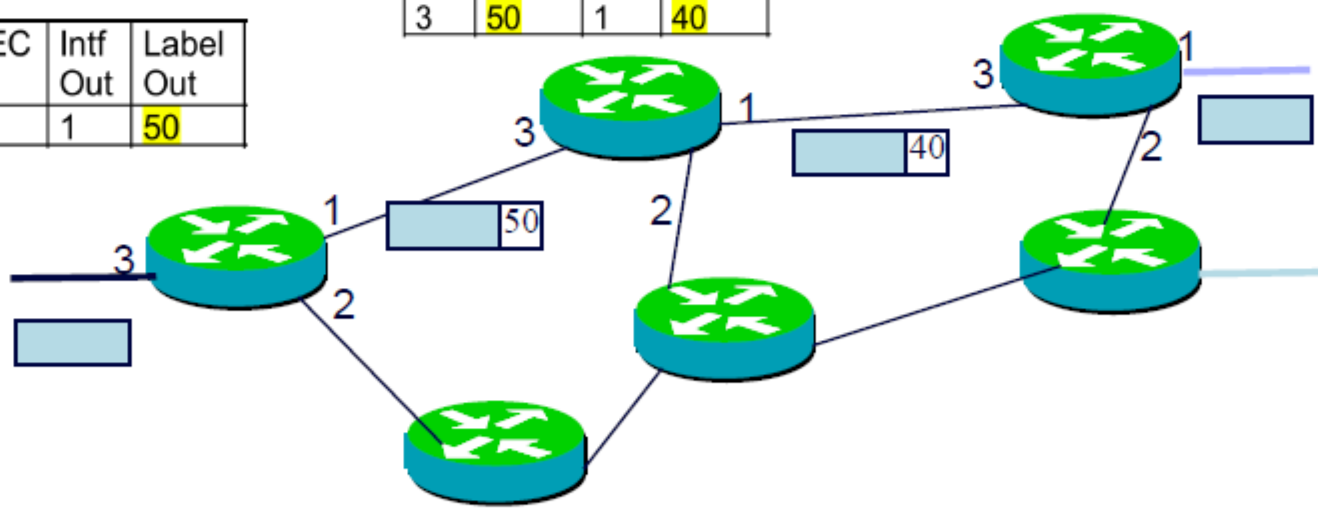


MPLS: Principle

- IP packet receives a LABEL upon the entrance in the network,
- The packet is then switched based on the LABEL value,
- The LABEL is changed at every hop,
- And so on and so forth until the arrival at the boundary of the MPLS domain.
- The label IP 'popped' and the IP packet is forwarded thanks to the destination IP address.

Intf In	Label In	Intf Out	Label Out
3	50	1	40

FEC	Intf Out	Label Out
a	1	50



MPLS terminology (1)

- LABEL: short and fixed-length value to identify a FEC, usually of local significance.
- FEC: Forwarding Equivalence Class represents a group of packets that share the same requirements for their transport.
  - The assignment of a particular packet to a particular FEC is done just once (when the packet enters the network).
- Label Stack: successive and ordered LABELs.
- Label Switching Router (LSR): router with MPLS functionalities.
- Label Edge Router (LER): resides at the edge of an MPLS network and assigns and removes the labels from the packets.
- Label Switched Path (LSP): equiv. Virtual Circuit. Path through one or more LSRs at one level of hierarchy followed by packets in a particular FEC.
- FTN – FEC To NHLFE Map: used for forwarding unlabeled packets.
- ILM - Incoming Label Map: maps label to a set of NHLFE entries.
- NHLFE - Next Hop Label Forwarding Entry: contains the information necessary to forward a packet for which a label has already been assigned: the packet's next hop address, operation (whether the MPLS header of the packet must be swapped or popped), and the label.

MPLS terminology (2)

FEC-to-NHLFE map

FEC	NHLFE
132.12.17.0/25	(3)
123.1.4.192/26	(5)
129.175.32.0/24	(7)
129.175.23.0/25	(8)
147.193.160.0/19	(17)
0.0.0.0/0	(20)

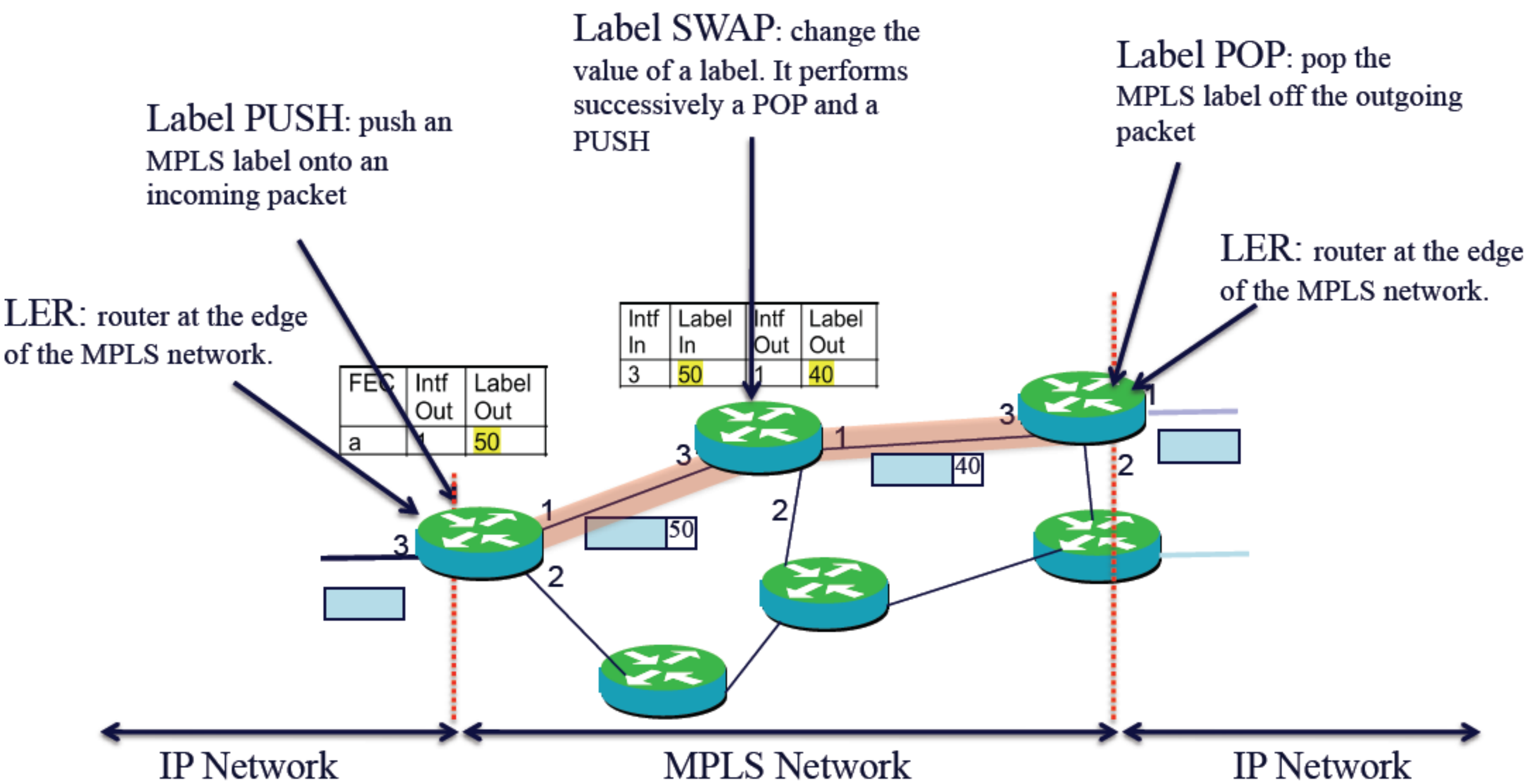
Incoming Label Map (ILM)

Label	NHLFE
15	(1)
22	(4)
145	(6)
234	(12)
456	(4)
989	(19)
1087	(2)

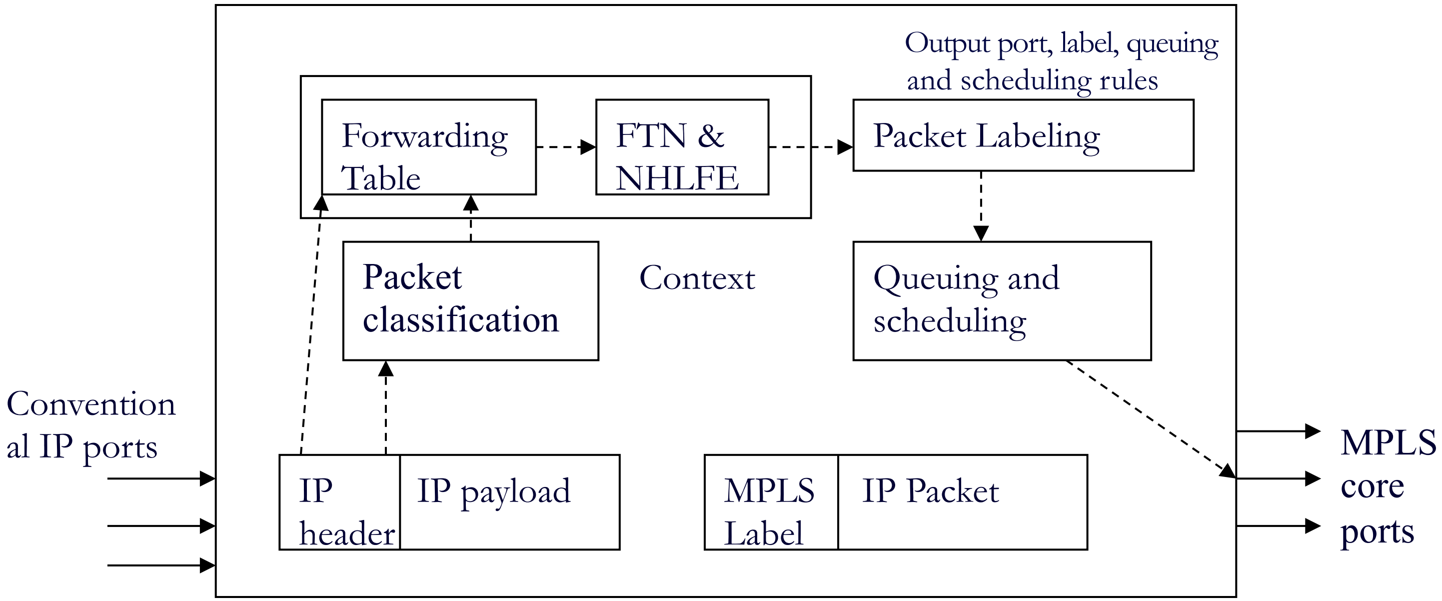
Next Hop Label Forwarding Entry (NHLFE)

Entry	Operation	Label	Next Hop	Interface
(1)	Swap	311	131.1.2.1	Eth0
(2)	Swap+push	786 555	131.1.2.5	Eth1
(3)	Push	561	131.1.2.1	Eth0
(4)	Pop	-		Eth0

MPLS terminology (3)

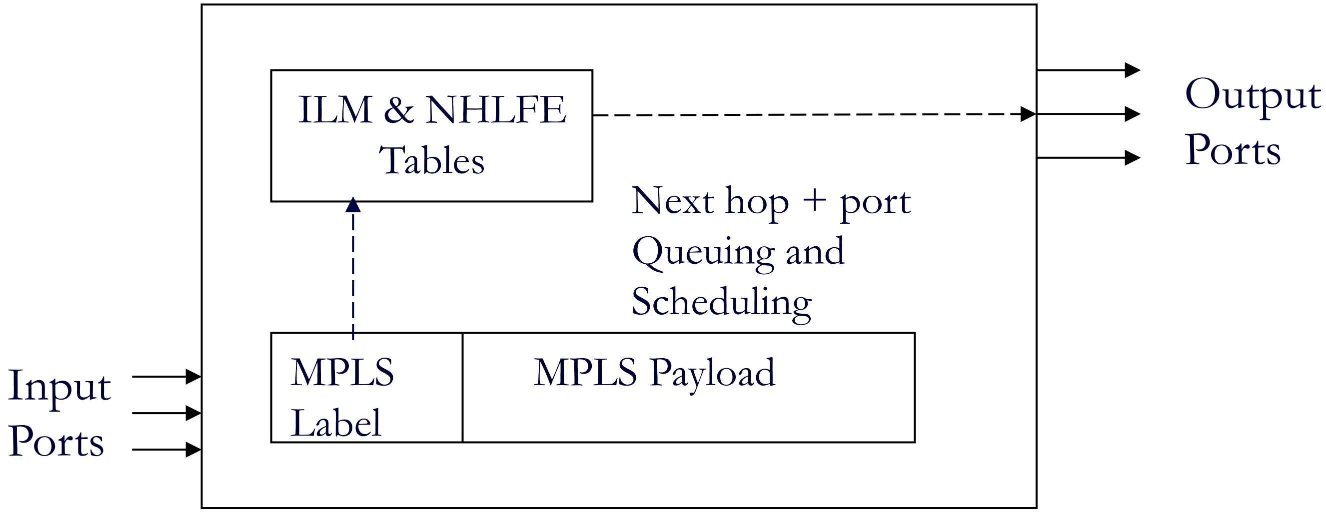


Ingress LER



- LER terminates and originates LSP's and performs both label based forwarding and conventional IP forwarding.
- **Ingress LER** - labels unlabeled packets and creates an initial MPLS frame by pushing one or more MPLS label entries.
- **Egress LER** - terminates LSP by popping the top MPLS stack entry.

LSR



- At each LSR, forwarding is done by the single index lookup into the switching table using the packet's MPLS label.
- The switching table is loaded a priori with a unique next-hop label, output port and queuing and scheduling rules.
- The establishment of mapping information is responsibility of control part - done using Label Distribution Protocols

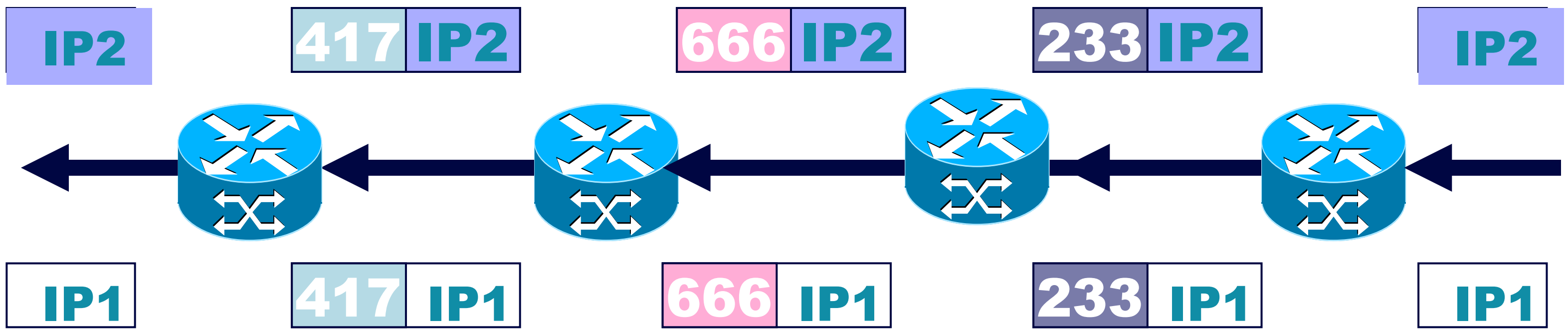
LSP Setup

- **“mpls-linux”** is an open source project

http://sourceforge.net/projects/mpls-linux/

- Scripts are designed to assist increating the entries in MPLS datastructures.
  - ./ingress.sh<label><interface><fwd_ip_add>
    - Updates NHLFE
  - ./lsr.sh<i/c_label><i/c_interface><o/g_label><o/g_interface> <fwd_ip_add>
    - Updates NHLFE + ILM
  - ./egress.sh<i/c_label><i/c_interface>
    - Updates ILM

Forwarding Equivalent Class (FEC)



Packets IP1 and IP2 are forwarded in the same way --- they are in the same FEC.

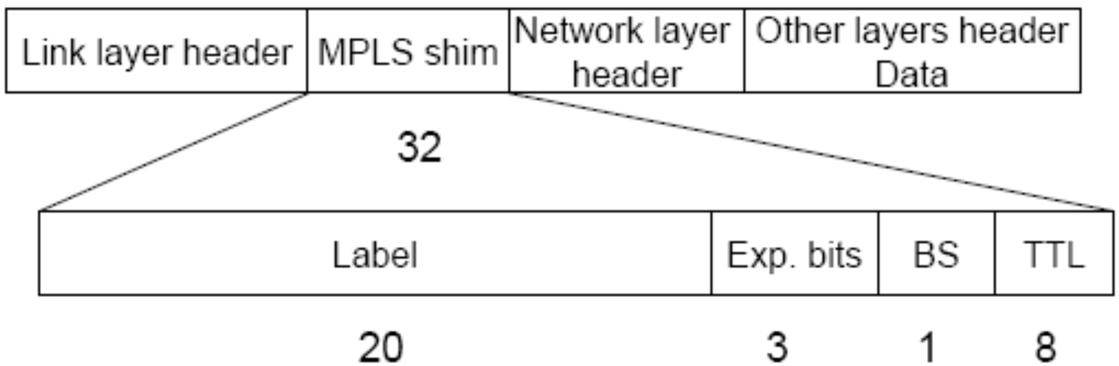
- FEC is a more general concept than virtual paths/channels used in ATM. All packets can be divided into subsets called FECs based on IP src @, IP dest @, IP protocol, src/dest port, etc.
- FECs can be
    - coarse grained consisting of all the packets with same destination address
        - Allow the overall system of be scalable where it is useful to handle large bundle of flows as a single class of traffic
        - Help in rerouting in event of occurrence of a fault
    - fine grained as in packets belonging to a particular application running between two hosts.
        - Help in providing different QoS to different flows.

LABEL definition

- LABEL is determined by the first router based on different parameters (TCP fields, IP address, etc.),
- LABEL identifies a FEC : Forwarding Equivalent Class (FEC = “a subset of packets that are all treated the same way by a router”),
- In conventional routing, a packet is assigned to a FEC at each hop (i.e. L3 lookup), in MPLS it is only done once at the network ingress nodes (at the edge of the network).
- LABEL will be used by the internal nodes as an index in their tables to determine the next hop and the value of the next LABEL,
- The concept of FECs provides for a great deal of flexibility and scalability.

Generic label format

- Shim Label

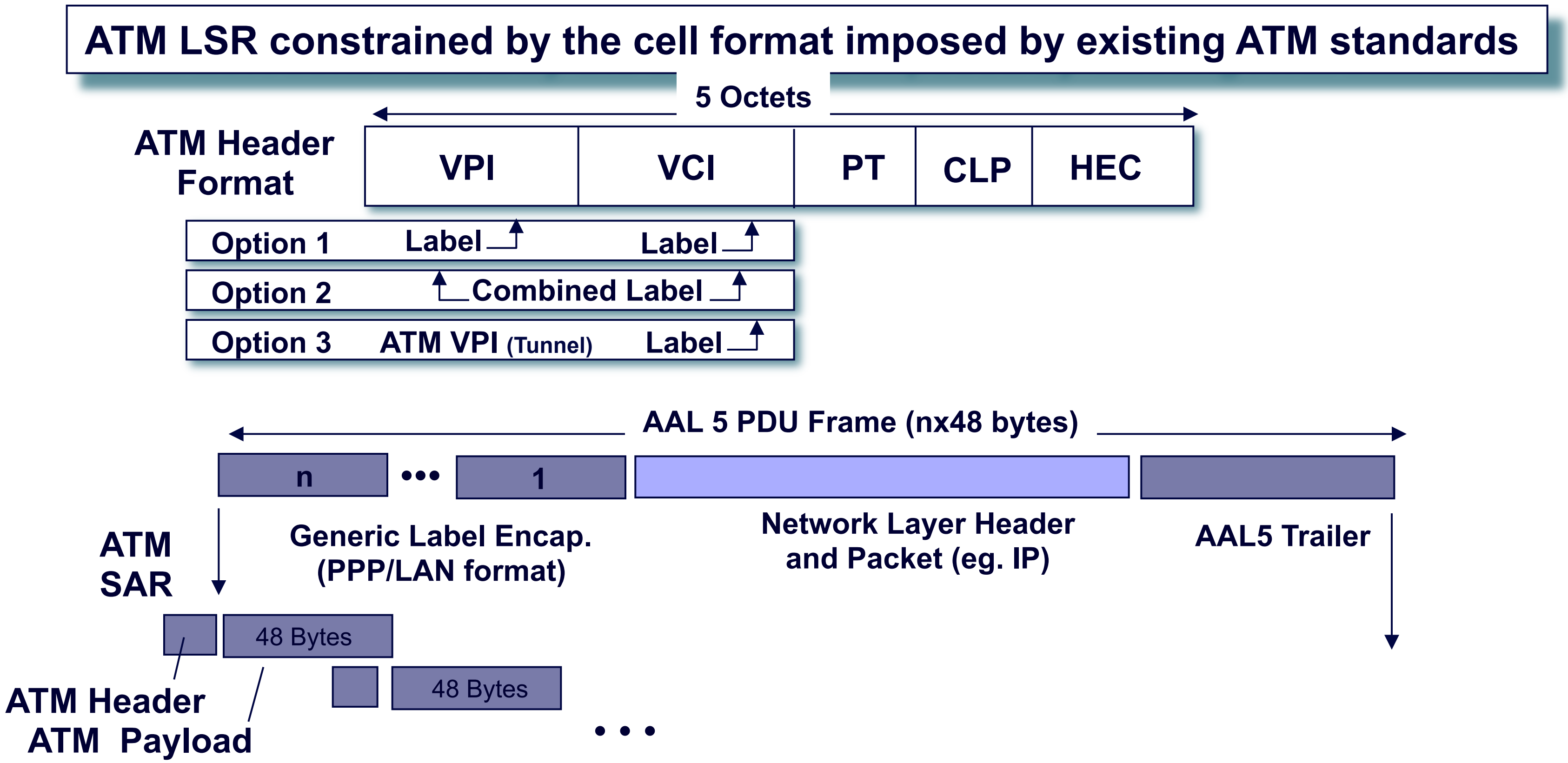


Exp.bits:	Experimental Bits, often used for Class of Service
BS:	Bottom of Stack bit, is set if no label follows
TTL:	Time To Leave, used in the same way like in IP

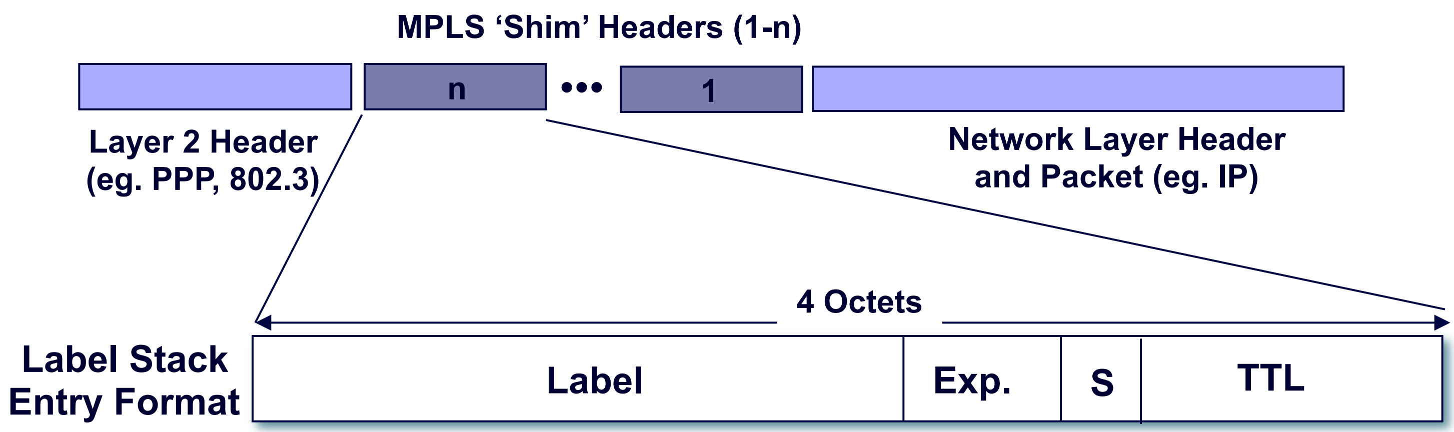
MPLS Link Layers

- MPLS is intended to run over multiple link layers
- Specifications for the following link layers currently exist:
  - ATM: label contained in VCI/VPI field of ATM header
  - Frame Relay: label contained in DLCI field in FR header
  - PPP/LAN: uses ‘shim’ header inserted between L2 and L3 headers
- Translation between link layers types must be supported

MPLS Encapsulation - ATM



MPLS Encapsulation - PPP & LAN Data Links



Label: Label Value, 20 bits (0-16 reserved)
Exp.: Experimental, 3 bits (was Class of Service)
S: Bottom of Stack, 1 bit (1 = last entry in label stack)
TTL: Time to Live, 8 bits

- TTL must be set to the value of the IP TTL field when packet is first labelled
- When last label is popped off stack, MPLS TTL to be copied to IP TTL field

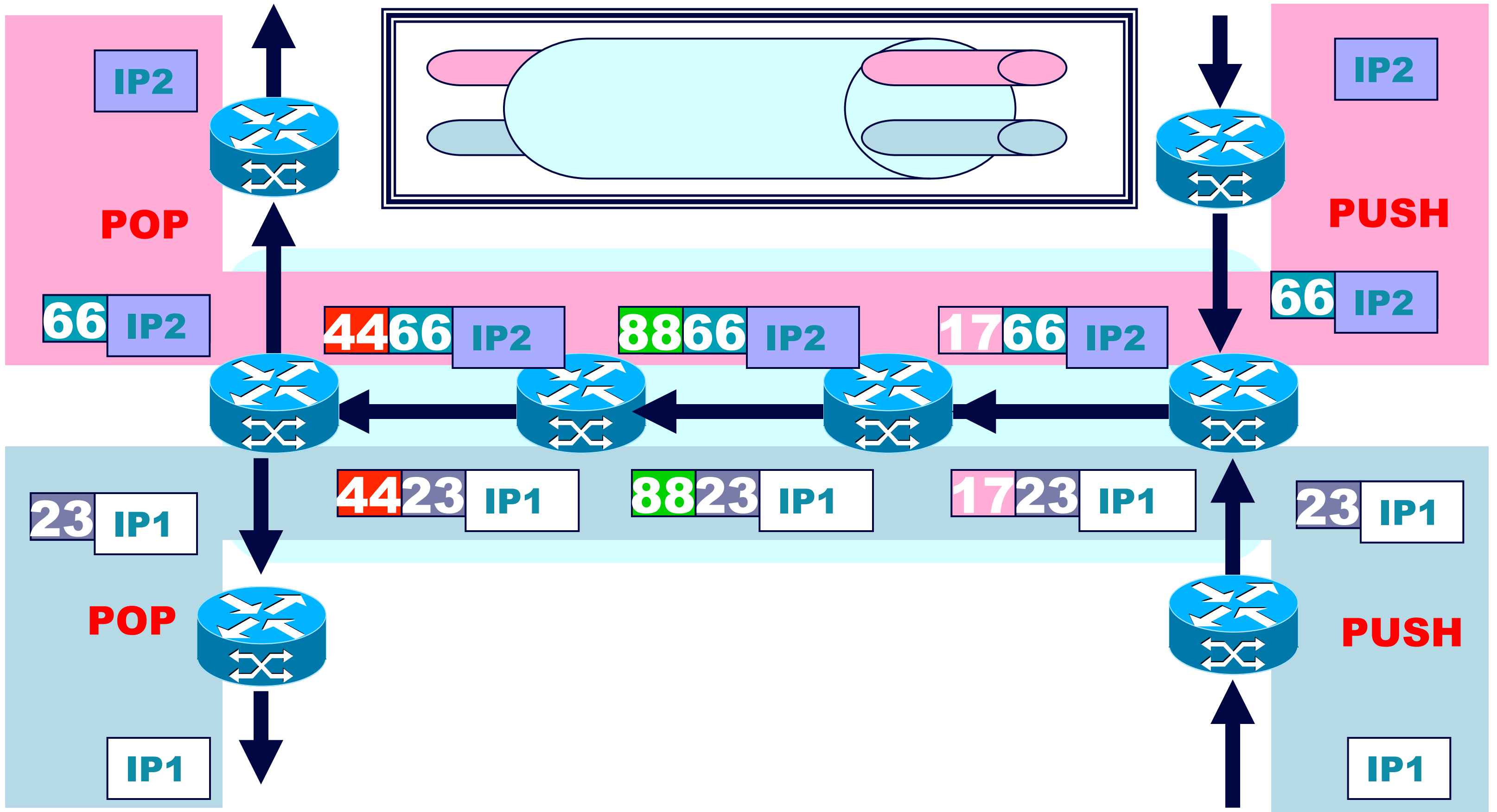
MPLS on PPP links and LANs uses 'Shim' Header Inserted Between Layer 2 and Layer 3 Headers

Label stack

Label (20 bits)	Exp (3 bits)	S = 0	TTL (8 bits)
Label (20 bits)	Exp (3 bits)	S = 0	TTL (8 bits)
• • •			
Label (20 bits)	Exp (3 bits)	S = 1	TTL (8 bits)

- MPLS allows hierarchical labels supported as LIFO stack.
- A packet is always processed based on the top label regardless of other labels that may be below it.
- Each label stack entry is 32 bits.
  - 20 bits for label
  - 3 bits for experimentation
  - 8 bits for TTL and 1 stack bit.

LSP hierarchy via label stacking



Basic MPLS Control plane

- Consists of
  - Network Layer routing protocols to distribute routing information between LSRs.
  - Label binding procedures to convert this routing information into the forwarding tables needed for label switching
- QoS routing requires additional information about availability of resources in the network and QoS requirements of each flow.
- A signaling protocol is also needed for reserving needed resources along the a selected route

  e.g. CR-LDP, RSVP-TE

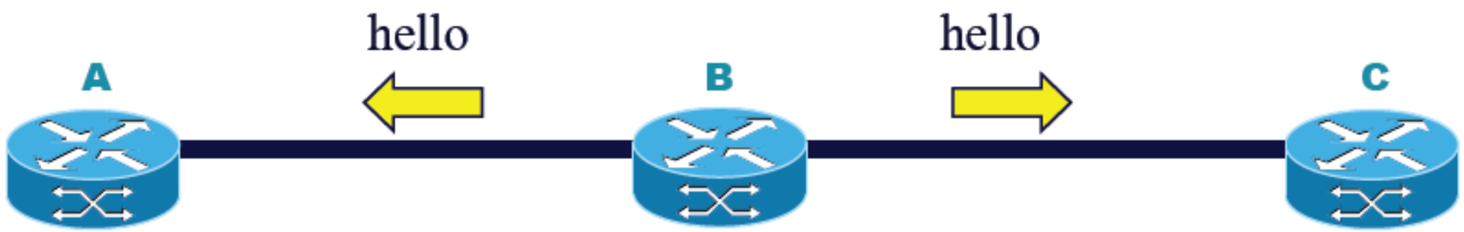
Label Distribution Protocol - LDP

- Dynamic distribution of label binding information
- It is used to map FECs to labels, which, in turn, create LSPs.
- Two LSRs must establish a bidirectional LDP session to exchange label binding information.
- Labels are always "downstream assigned".
  - Label information flows in the direction opposite to that of data packets
- Two modes of label distribution
  - **Unsolicited Downstream** - each LSR distributes label bindings to its upstream LSRs (even if they haven't explicity requested them).
  - **Downstream-on-demand** - each LSR allows LSR upstream to explicitly request the label binding for a particular FEC.
- LDP sessions are established between LDP peers in the MPLS network (not necessarily adjacent).
- Supports only hop-by-hop routing
- Allows LSR discovery.

LDP message categories

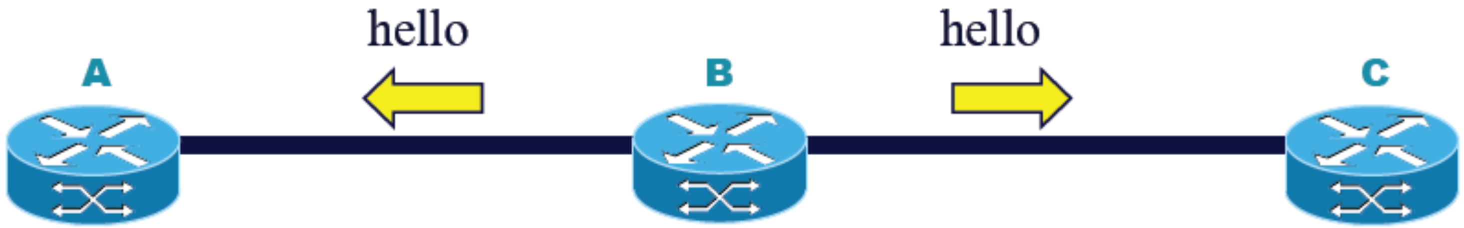
- LDP message types:
  - **discovery messages**—used to announce and maintain the presence of an LSR in a network (HELLO messages sent periodically using UDP/multicast)
  - **session messages**—used to establish, maintain, and terminate sessions between LDP peers (use TCP)
  - **advertisement messages**—used to create, change, and delete label mappings for FECs (use TCP)
  - **notification messages**—used to provide advisory information and signal error information (use TCP)

Discovery messages



- LDP discovery is a mechanism that enables an LSR to discover potential LDP peers:
- An LSR periodically sends LDP Link Hellos out the interface.
- Hello messages contain range of labels.
- Receipt of an LDP Link Hello on an interface identifies a "Hello adjacency" with a potential LDP peer reachable at the link level on the interface.
- An LSR maintains a hold timer with each Hello adjacency. If the timer expires without receipt of a matching Hello from the peer, the LSR deletes the Hello adjacency.

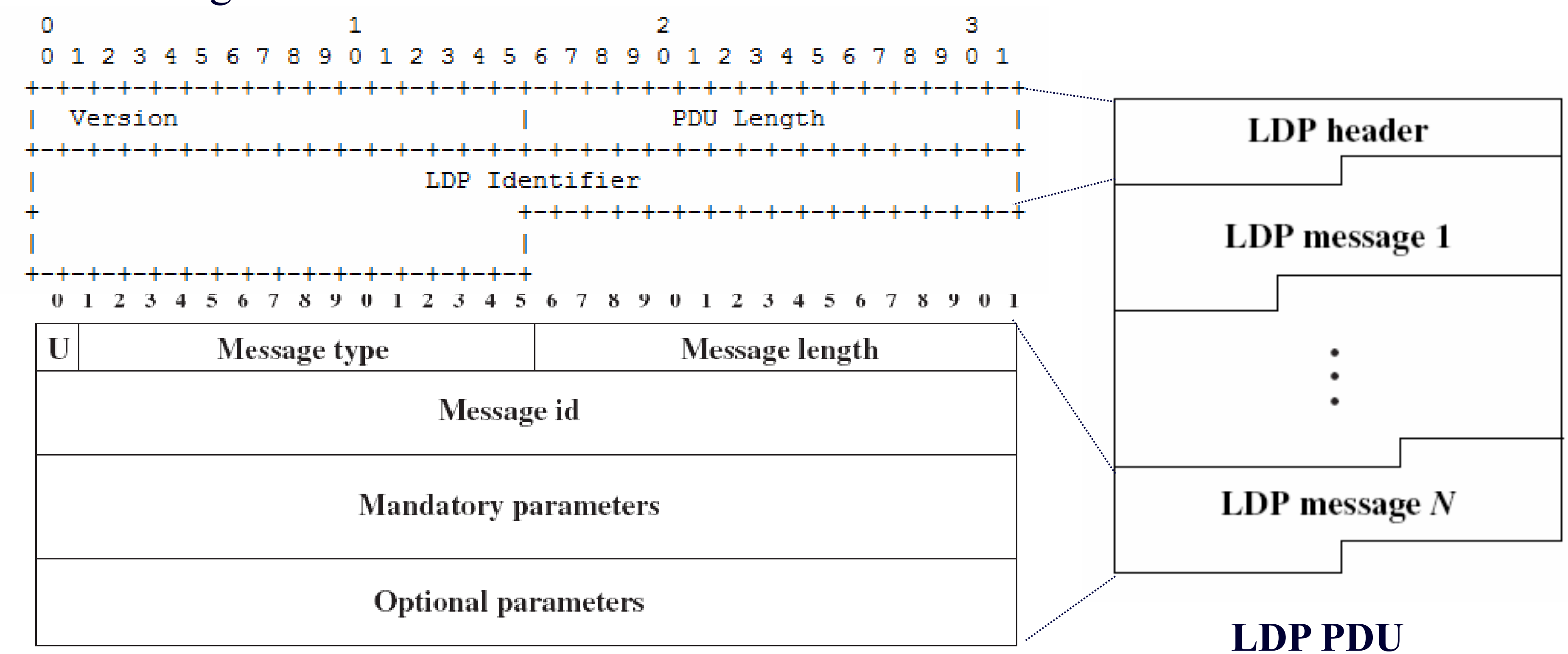
LDP Sessions



- Session establishment is a two step process:
  - Transport Connection Establishment: the exchange of Hellos results in the creation of a Hello adjacency at LSR1 that serves to bind the link and the label spaces between both routers.
  - Session Initialization: negotiate session parameters by exchanging LDP Initialization messages. (LDP protocol version, timer values, label distribution method…)
  - The session is closed if no LDP message (KeepAlive message, LSP establishment…) is received during the KeepAlive Timer.

LDP PDU format

- LDP messages have a common structure that uses a Type-Length-Value (TLV) encoding scheme.
- Messages are sent as LDP PDUs. Each PDU can contain more than one LDP message. Each LDP PDU is an LDP header followed by one or more LDP message



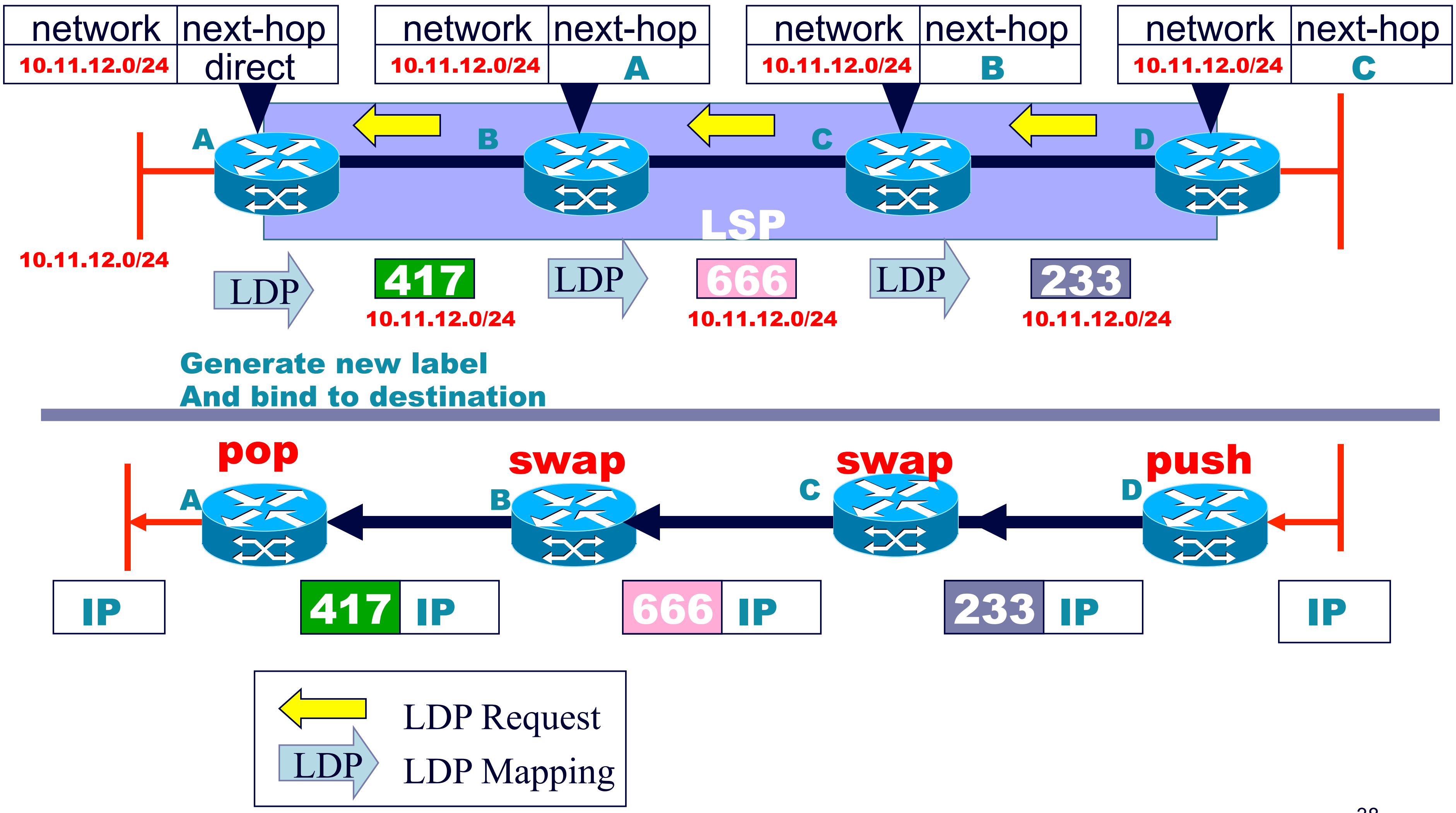
LSP setup

- MPLS provides two options to set up an LSP
  - hop-by-hop routing with LDP

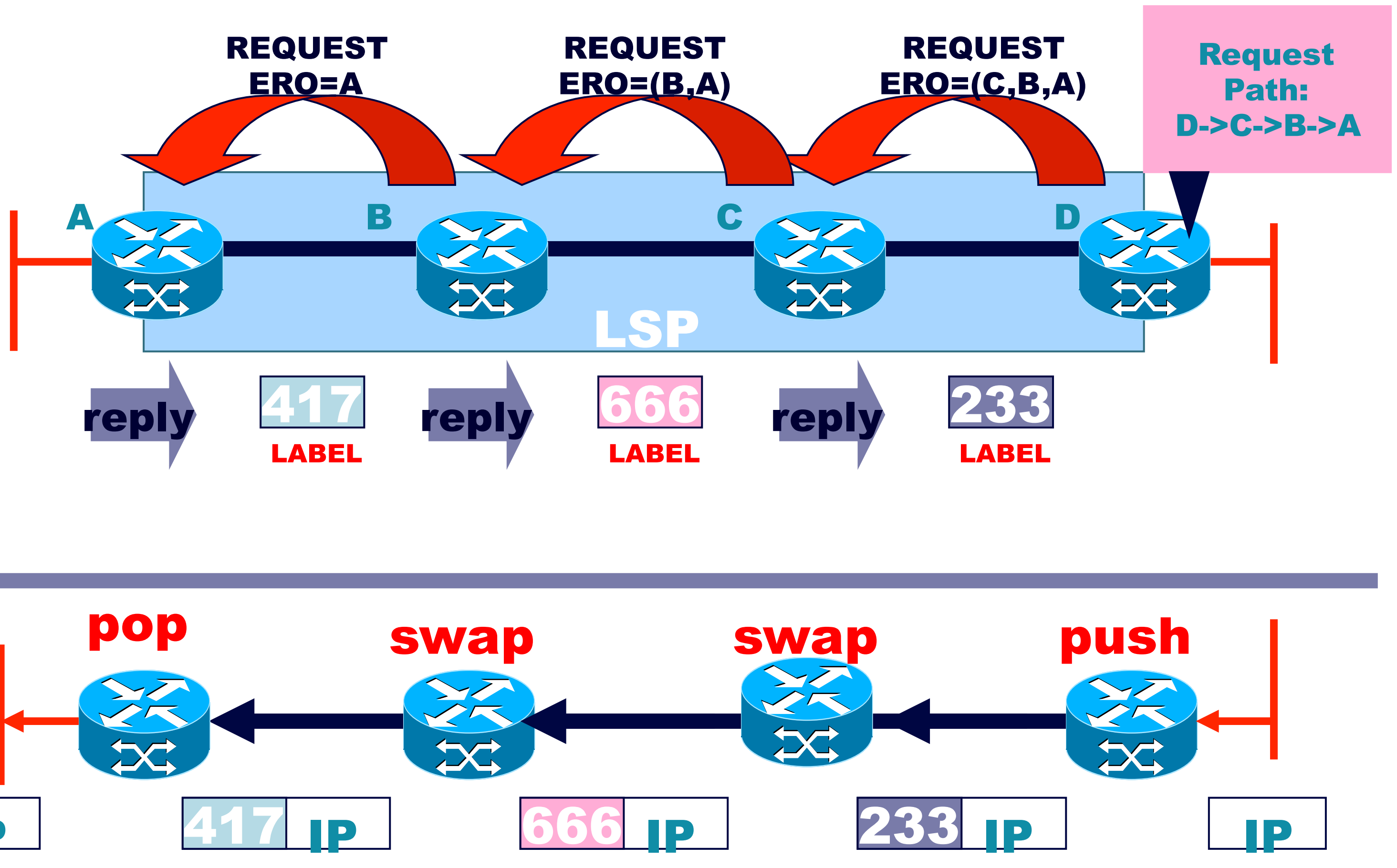
    Each LSR independently selects the next hop for a given FEC. LSRs support any available routing protocols (OSPF…).
  - Explicit routing

    Is similar to source routing. The ingress LSR specifies the list of nodes through which the packet traverses.
- The LSP setup for an FEC is unidirectional. The return traffic must take another LSP!

LDP and hop-by-hop routing



Explicit path setup



Hop-by-Hop vs. Explicit Routes

- Hop-by-Hop
  - Distributed control
  - Trees rooted at destination
  - Destination based forwarding
- Explicit Routing
  - Originates at source
  - Paths from sources to destinations
  - Traffic to path mapping based on what configuration commands your vendor(s) provide

Explicit routing shows great promise for traffic engineering

Traffic Engineering

- In MPLS, traffic engineering is inherently provided using explicitly routed paths.
- The LSPs are created independently, specifying different paths that are based on user-defined policies. However, this may require extensive operator intervention.
- RSVP-TE and CR-LDP are two possible approaches to supply dynamic traffic engineering and QoS in MPLS.

CR-LDP : Constraint-based Routing LDP

- CR-LDP: add explicit routes and resource reservation to a label distribution protocol.
  - Takes into account parameters, such as link characteristics (bandwidth, delay, etc.), hop count, and QoS.
- It is entirely possible that a longer (in terms of cost) but less loaded path is selected:
  - Load balancing
- Drawback: It adds more complexity to routing calculations.

RSVP-TE

- RSVP-TE: add label distribution and explicit routes to a resource reservation protocol:
  - Extension of RSVP for LSP tunnels.
  - Request bandwidth and traffic conditions on a defined path.
- Drawback:
  - Requires regular refreshes: Soft state periodically refreshed,
  - IntServ QoS model: scalability.

MPLS Operation

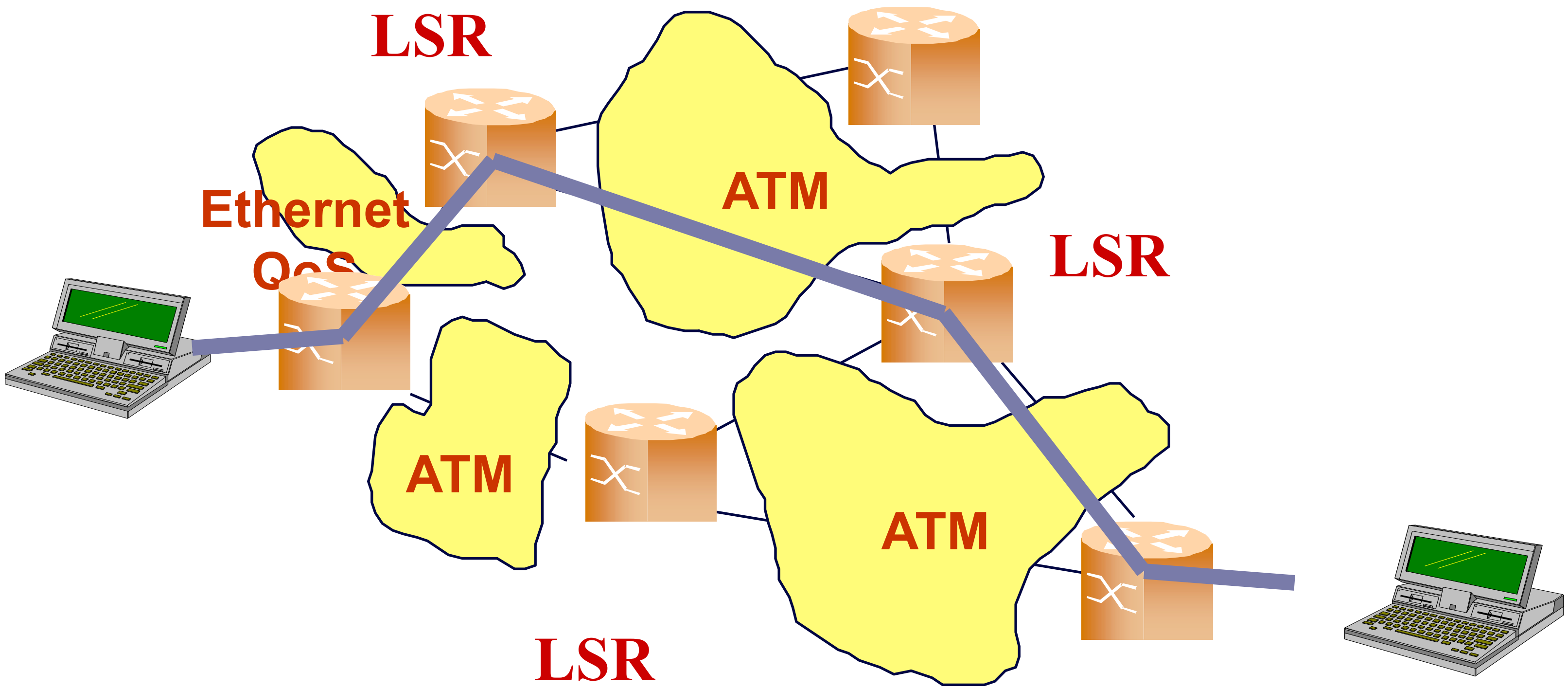
- The following steps must be taken for a data packet to travel through an MPLS domain.
  - label creation and distribution
  - table creation at each router
  - label-switched path creation
  - label insertion/table lookup
  - packet forwarding

MPLS Advantages/Disadvantages

- Advantages:
  - Improves packet-forwarding performance in the network
  - Supports QoS and CoS for service differentiation
  - Supports network scalability
  - Integrates IP and ATM in the network
  - Builds interoperable networks
- Disadvantages:
  - An additional layer is added
  - The router has to understand MPLS

Transition towards MPLS

- Assume a network interconnecting ATM and Ethernet sub-networks via LSRs. Is it an MPLS network?



Protocol stack in MPLS