Traffic Engineering (TE) & Multiprotocol Label Switching (MPLS)

Redes de Comunicações II

Licenciatura em Engenharia de Computadores e Informática

Prof. Amaro de Sousa (asou@ua.pt)

DETI-UA, 2024/2025

Traffic supported by a Network

- A **traffic flow** from a sender host A to a receiver host B is the sequence of IP packets sent by host A to host B during time
  - The **throughput** is the rate, in bits/second, of the traffic flow

---

- Example: host A downloads a file of 5 GBytes file from host B through TCP at an average download rate of 5000 IP packets/second.
  - Average download throughput of the file download:

    Assuming that each IP download packet has 20 Bytes (IP header) + 20 Bytes (TCP header) + 1460 Bytes (Data) = 1500 × 8 bits, then, the download throughput is 1500 × 8 × 5000 = 60000000 bps = 60 Mbps (Megabits/sec)
  - Average upload throughput of the file download:

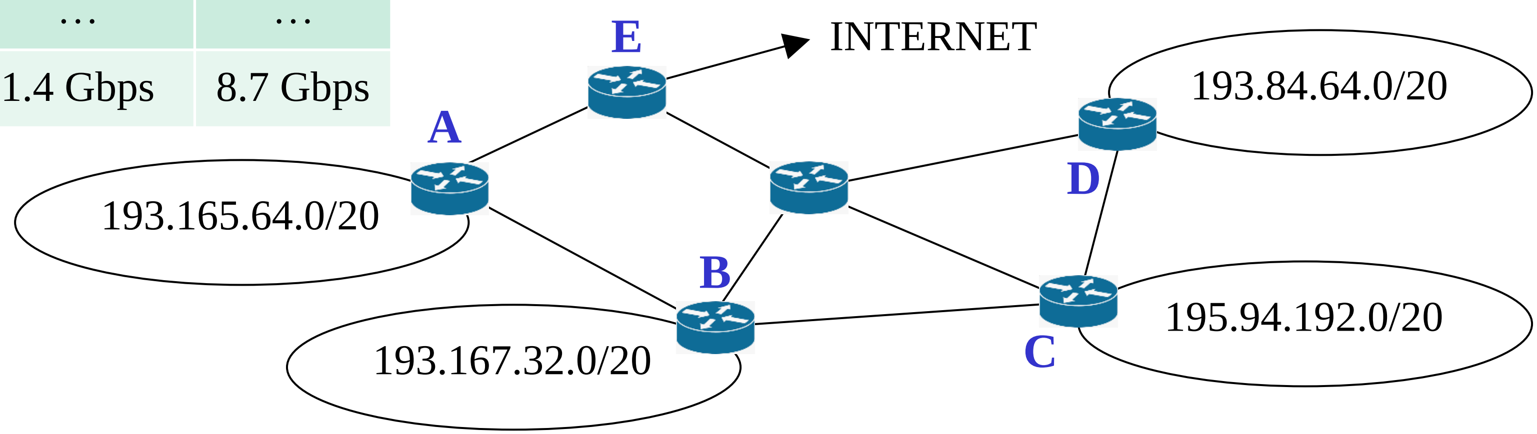
    Assuming that each 2 downloaded TCP segments generate one upload TCP segment (to acknowledge the received data) in an IP packet with 20 Bytes (IP header) + 20 Bytes (TCP header) = 40 × 8 bits, then, the upload throughput is 40 × 8 × 2500 = 800000 bps = 800 Kbps (Kilobits/sec)
  - File download time:

    The file size 5 GBytes is segmented in = $5 \times 1024^3 / 1460 \approx 3677$ TCP segments. So, the download time is around 3677 / 5000 = 0.74 seconds

Traffic supported by a Network

- An **aggregated flow** from one network element A to another element B is the set of traffic flows from A to B.
  - The **throughput** of an aggregated flow is the total rate, in bits/second, of all its traffic flows
- The **traffic throughput of a service** supported by a network is defined (in its simplest version) by the average throughput over time of the aggregated flows between all node pairs running the service.

N1	N2	N1→N2	N2→N1
A	B	2.5 Gbps	3.1 Gbps
A	C	1.3 Gbps	1.7 Gbps
…	…	…	…
D	E	1.4 Gbps	8.7 Gbps



Network Dimensioning vs. Traffic Engineering

- **Network Dimensioning** is the process of determining the required capacity on each element of the network to be able to carry the traffic throughput of all supported services
  - Network elements are links, routers, switches, etc…
  - Traffic throughputs are hard to predict!
- **Traffic Engineering** is the process selecting the routing paths to best fit the traffic of all supported services in the capacity resources of the network
  - Ensure the most desirable/appropriate routing paths for each aggregated flow of each service.
  - The aim is to use efficiently the capacity resources of the network.
    - For example, to prevent over-utilized (congested) links whilst other links are under-utilized.

Life Cycle of a Network

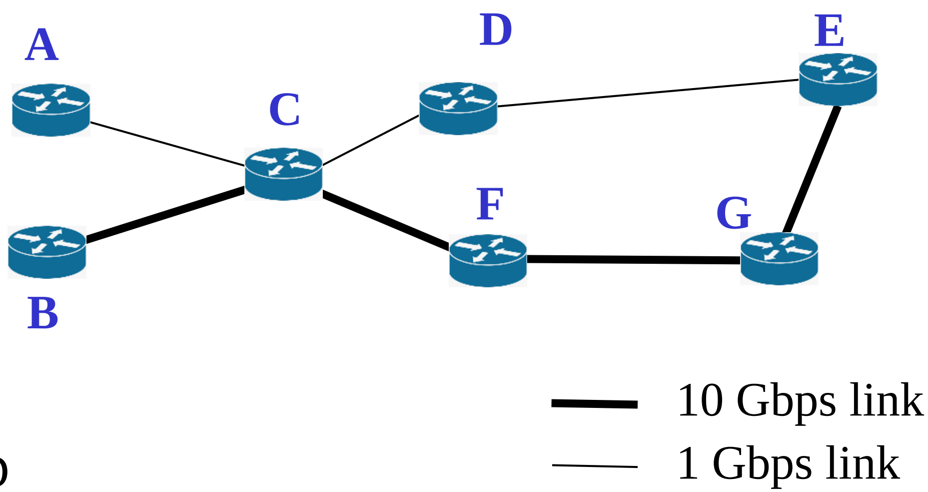
- A network life cycle is composed by two management phases:
  - A network dimensioning phase
    - The operator estimates the traffic throughput of the services the network must support in a given future time (for example, in the next year).
    - Then, it determines the required capacity of the network elements to carry the estimated traffic throughputs.
      - The aim is to minimize the investment cost.
    - Finally, the operator upgrades the network.
  - A traffic engineering phase
    - The operator changes the routing of the network in accordance with the traffic changes of the supported services.
    - This phase continues while network resources are enough to support all services
      - An efficient traffic engineering allows this phase to last for a longer time (minimizing the need for a new investment cost).
    - A new network dimensioning phase is required when the network capacity resources are no longer enough to support all services due to:
      - (i) throughput grow of the existing aggregated flows, (ii) new aggregated flows of the initial services and/or (iii) the introduction of new services.

Traffic Engineering with IGP Routing Protocols

- Traffic engineering is not ideal when using standard IGP protocols (as RIP or OSPF)

- With RIP, all flows are routed in C to next-hop D
  - There is a huge overflow (i.e., a drop) of 10 Gbps in link C-D
- With OSPF, and assigning appropriate OSPF port costs:
  - All flows can be routed in C through next-hop F
    - overflow of 1 Gbps in link C-F
  - All flows can be routed in C through next-hops D and C (with ECMP)
    - overflow of 4.5 Gbps in link C-D

N1	N2	N1→N2	N2→N1
A	E	1.0 Gbps	-
B	E	5.0 Gbps	-
C	E	5.0 Gbps	-



Traffic Engineering with IGP Routing Protocols

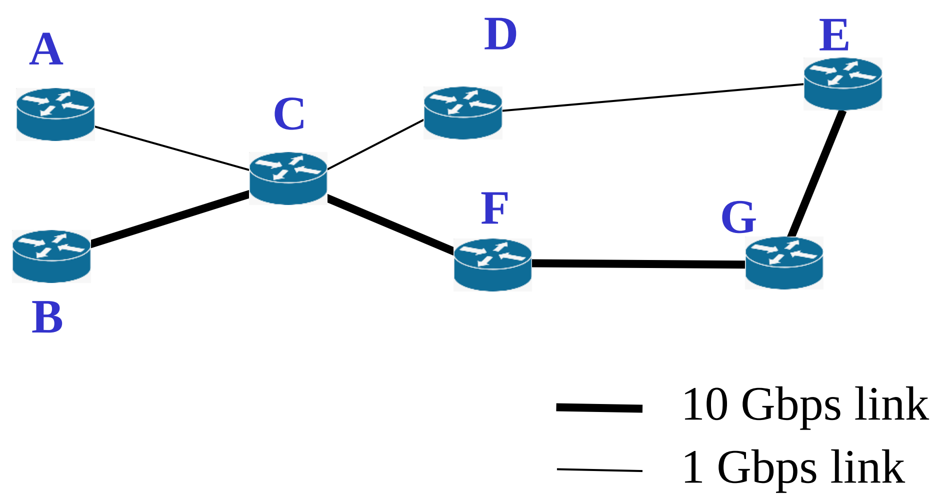
- In order to fit the traffic flow throughputs in the capacity resources of the network, we need to route each traffic flow through:

A – E: A → C → D → E

B – E: B → C → F → G → E

C – E: C → F → G → E

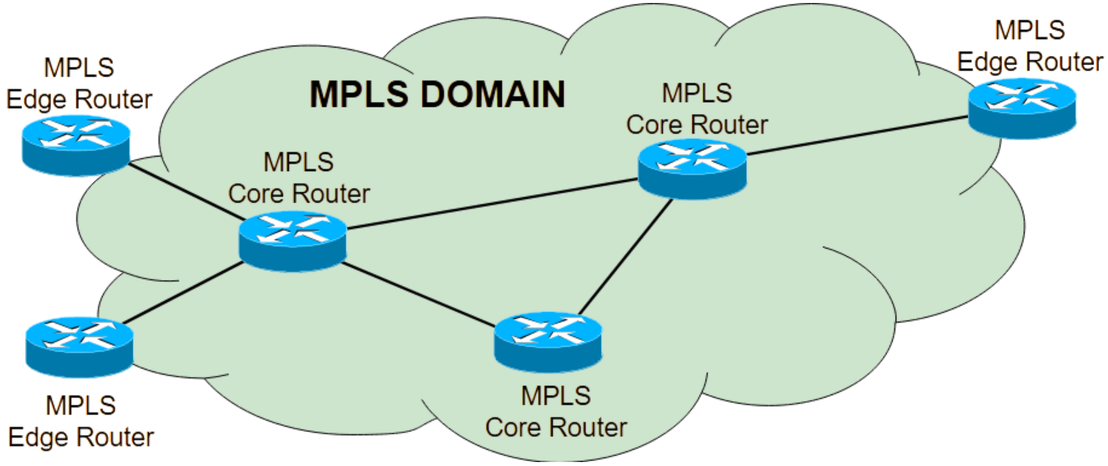
N1	N2	N1→N2	N2→N1
A	E	1.0 Gbps	-
B	E	5.0 Gbps	-
C	E	5.0 Gbps	-



- Note that:
  - this solution requires C to route flows based not only on the destination but also on the source
  - IGP protocols (as RIP or OSPF) are destination-based routing protocols

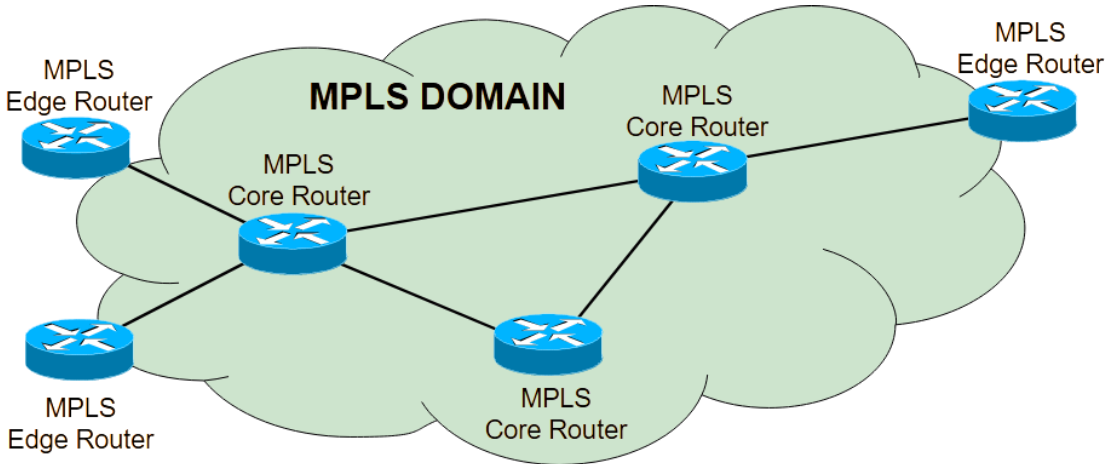
Multiprotocol Label Switching (MPLS)

- MPLS is a routing technique that forwards packets from one router to the next router based on labels rather than on network addresses.
  - While network addresses identify endpoints, labels identify established routing paths between endpoints.
  - MPLS can encapsulate packets of different protocols (IPv4 and IPv6), hence the ‘multiprotocol’ component of the name.
- In an **MPLS domain**, there are **Edge Routers** and **Core Routers**.



Multiprotocol Label Switching (MPLS)

- Packets are labelled at the entry Edge Router with a label of the first hop.
- Each Core Router forwards the packet based on its incoming label and swaps the label to the one of the next hop.
- The exit Edge Router removes the label.
- Advantages:
  - Simplification of the packet routing process on routers.
  - Full traffic engineering capability.
  - Simplification of the network management (a single protocol layer).

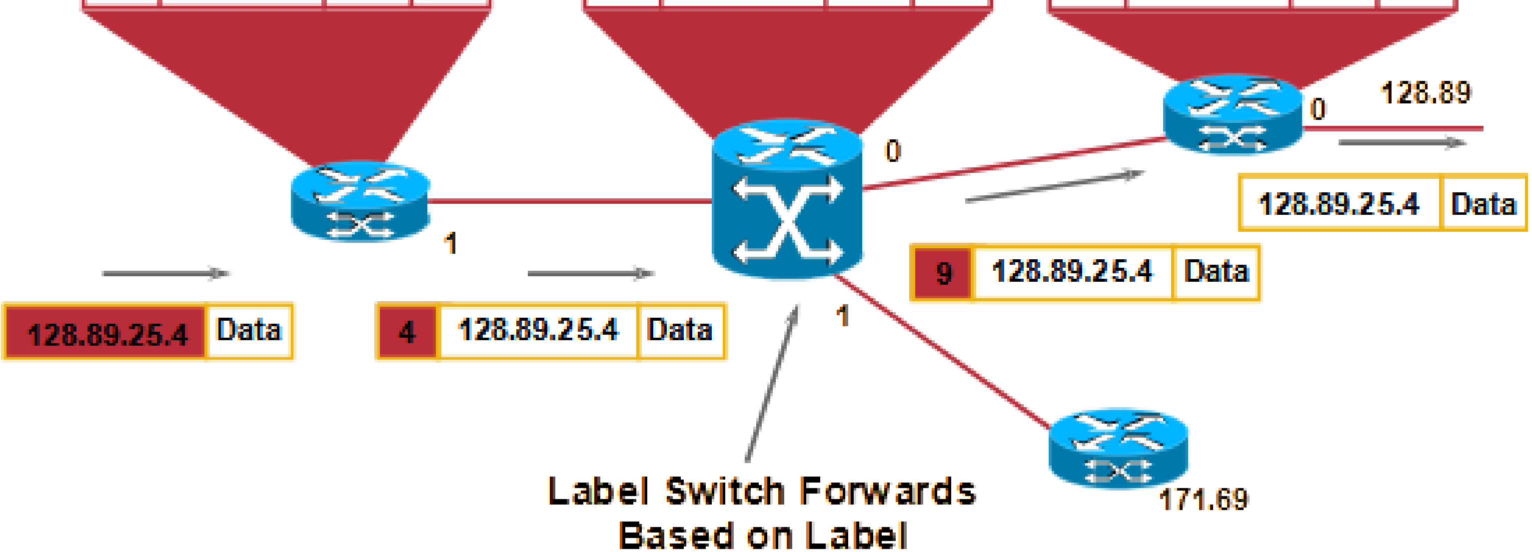


MPLS Switching Illustration

In Label	Address Prefix	Out I'face	Out Label
	128.89	1	4
	171.69	1	5
...	...	...	...

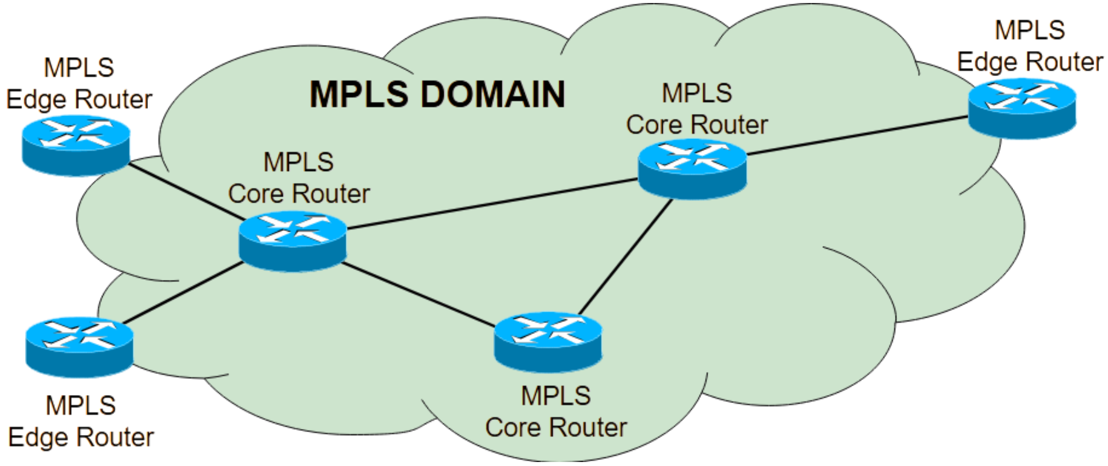
In Label	Address Prefix	Out I'face	Out Label
4	128.89	0	9
5	171.69	1	7
...	...	...	...

In Label	Address Prefix	Out I'face	Out Label
9	128.89	0	
...	...	...	...



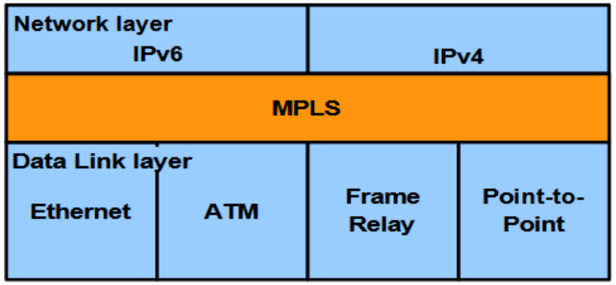
MPLS Fundamentals

- As a packet enters an MPLS domain, it is assigned a label based on its **Forwarding Equivalence Class (FEC)** defined at the entry **Edge Router**.
- FECs are groups of packet flows forwarded through the same **Label Switched Path (LSP)** by **Label Switching Routers (LSR)**.
  - Need a mechanism to assign and distribute labels to establish LSP paths.
- In **Core Routers**, routing is based on label-swapping and forwarding.
- MPLS works in two planes:
  - Control Plane - responsible for setup label swapping tables among LSRs.
  - Forwarding Plane - uses the label carried by each incoming packet and the label swapping table of the LSR to forward the packet.

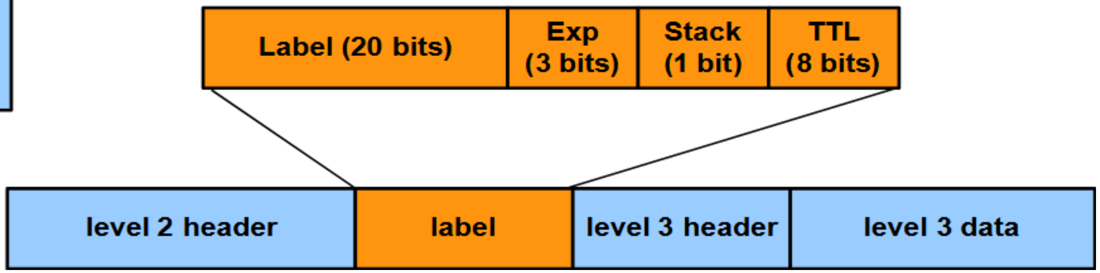


MPLS Labels

- Label is a 20-bit field that carries the value of the Label.
  - TTL field is IP independent – Similar purpose.
- On some Data Link (Layer 2) old technologies, label is given by the appropriate fields of their header.
  - ATM : VPI (Virtual Path ID) and VCI (Virtual Channel ID) fields.
  - Frame Relay: DLCI (Data Link Connection Identifier) field.
- On other Data Link technologies (Point-to-Point, Ethernet), the label is inserted between Layer 2 and Layer 3 headers.

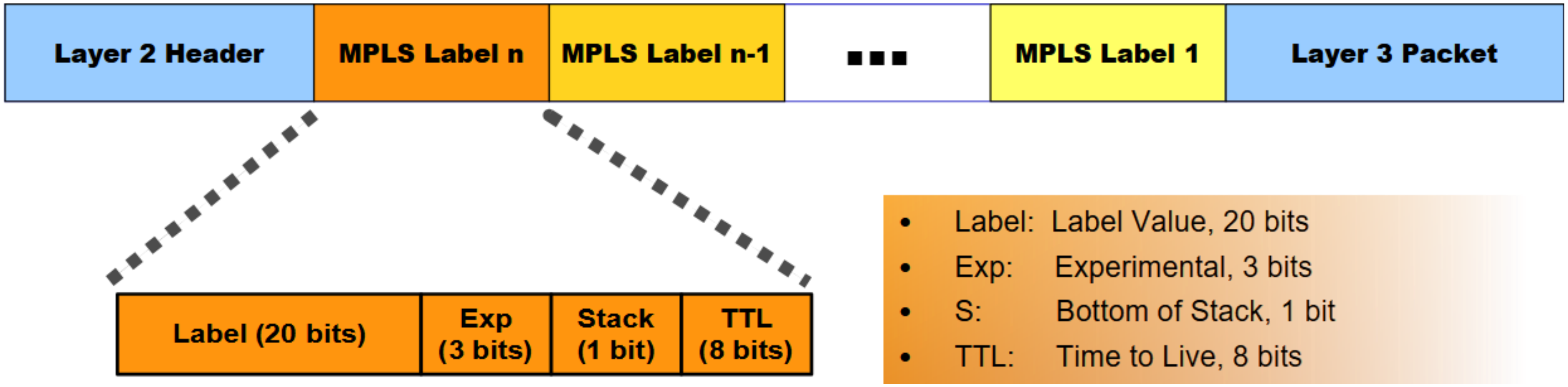


- Label: Label Value, 20 bits
- Exp: Experimental, 3 bits
- S: Bottom of Stack, 1 bit
- TTL: Time to Live, 8 bits



MPLS Label Stacking

- Labels can be used in stacks to support more efficiently multiple services:
  - Outer (last) label (first received) is used to switch the packets between Edge Routers in the MPLS core.
  - Inner labels are used to designate services, FECs, etc.
- S (Bottom of Stack) bit is set to one for the last entry in the label stack (i.e., for the bottom of the stack), and zero for all other labels.

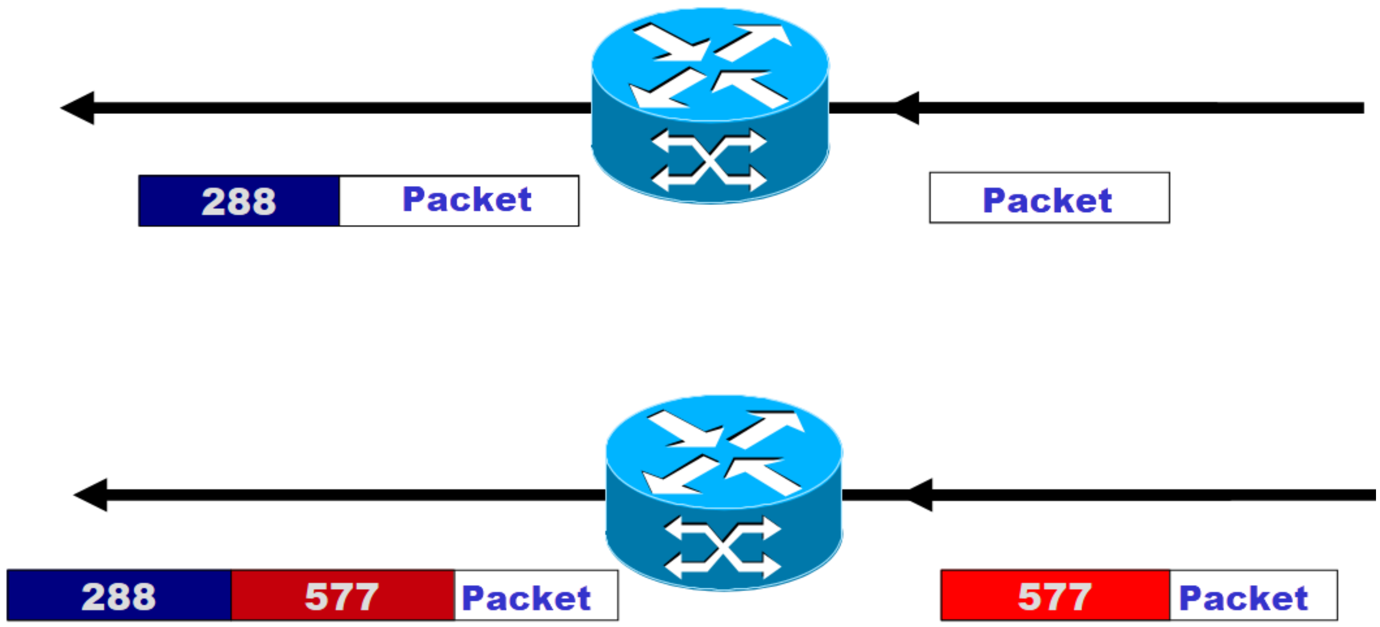


- Label: Label Value, 20 bits
- Exp: Experimental, 3 bits
- S: Bottom of Stack, 1 bit
- TTL: Time to Live, 8 bits

MPLS Label Swapping

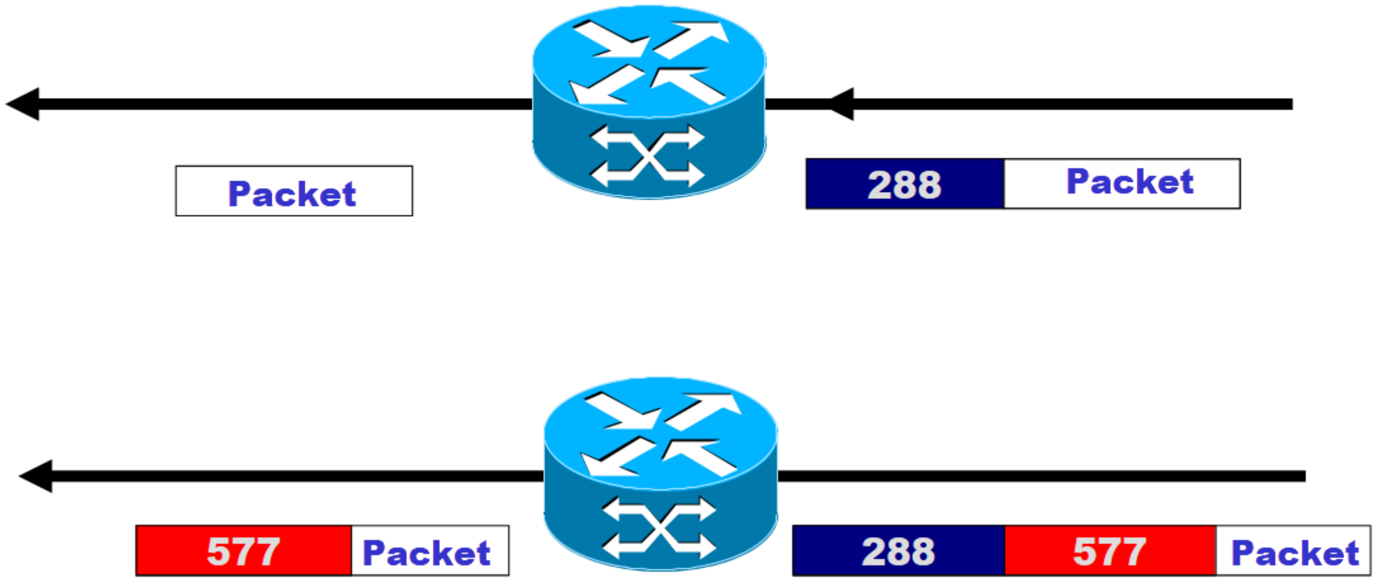
- Swapping is the process of changing the label of an incoming packet in the outgoing packet.
  - Labels are short and based on fixed-length fields.
  - Forwarding based on labels is much more efficient than forwarding based on IP routing tables.
  - Throughput capacity of MPLS routers is much higher than throughput capacity of IP routers.

MPLS Pushing Labels



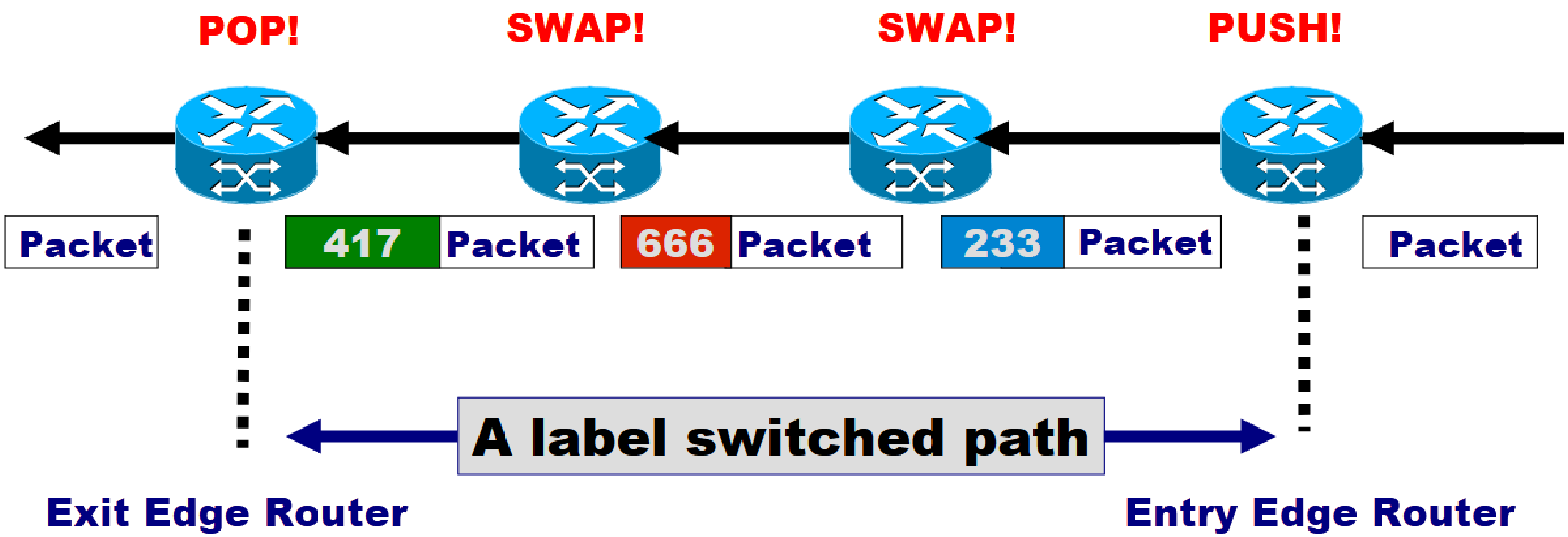
- Pushing is the process of adding a label to an incoming packet in the outgoing interface.

MPLS Popping Labels



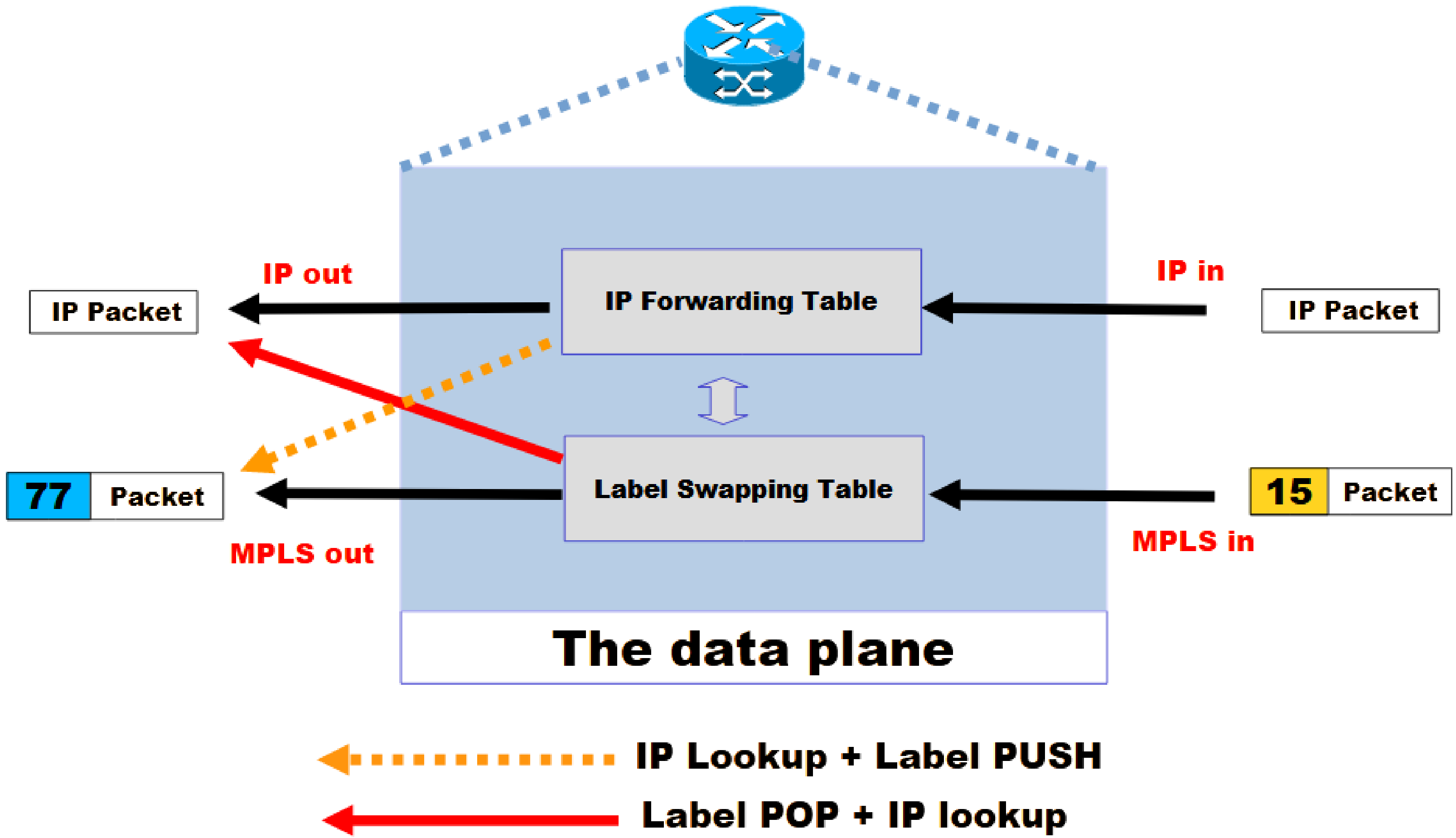
- Popping is the process of removing a label from an incoming packet in the outgoing interface.

A Label Switched Path (LSP)

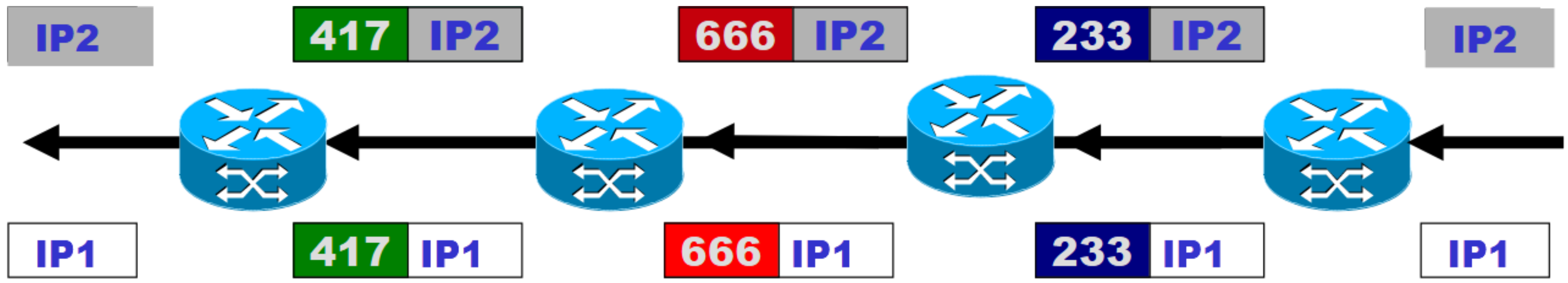


- An LSP is often referred to as a MPLS tunnel, as payload headers are not inspected inside of an LSP

Label Switched Router (LSR)



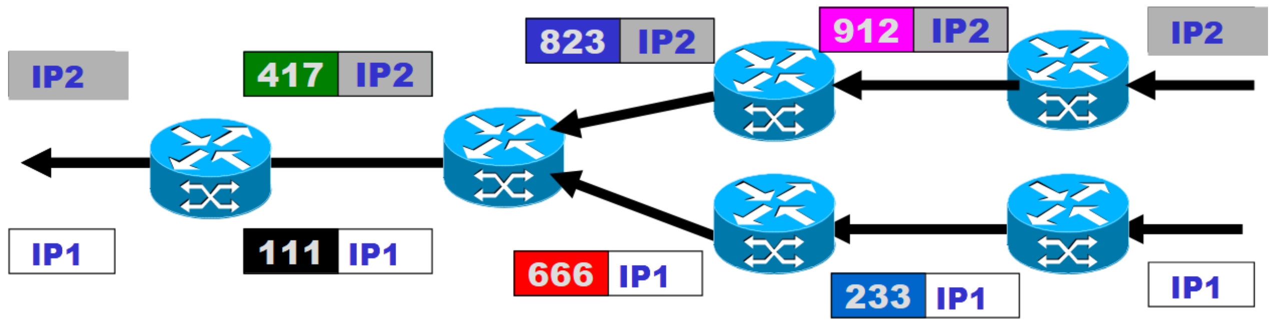
Forwarding Equivalence Class (FEC)



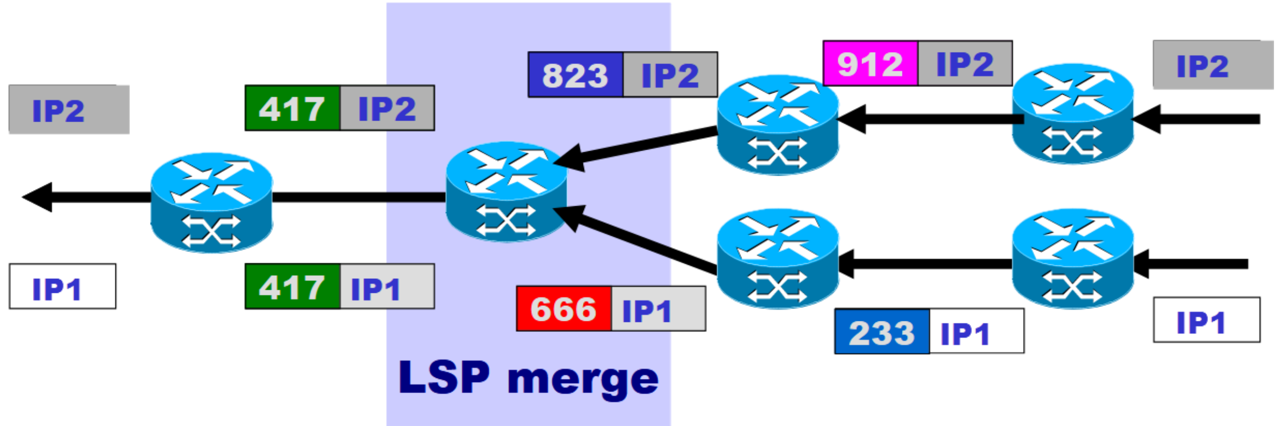
- Packets IP1 and IP2 are in the same FEC: they are forwarded in the same way.
  - IP layer headers are not inspected inside an MPLS LSP.
  - This means that inside of the tunnel, the LSRs do not need full IP forwarding tables.

Merging LSPs

- Two FECs to the same exit Edge Router with a common part in the two routing paths:

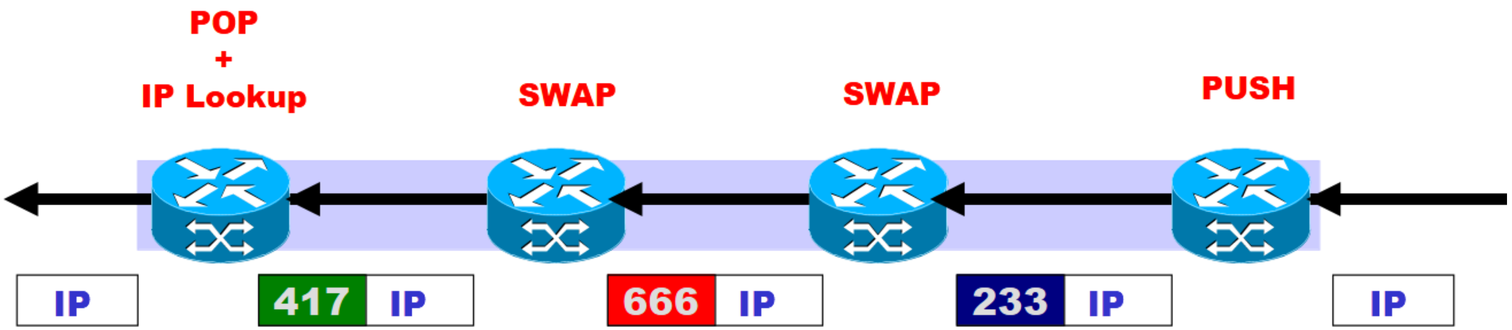


- The common part can be merged to reduce the size of the Label Swapping table:

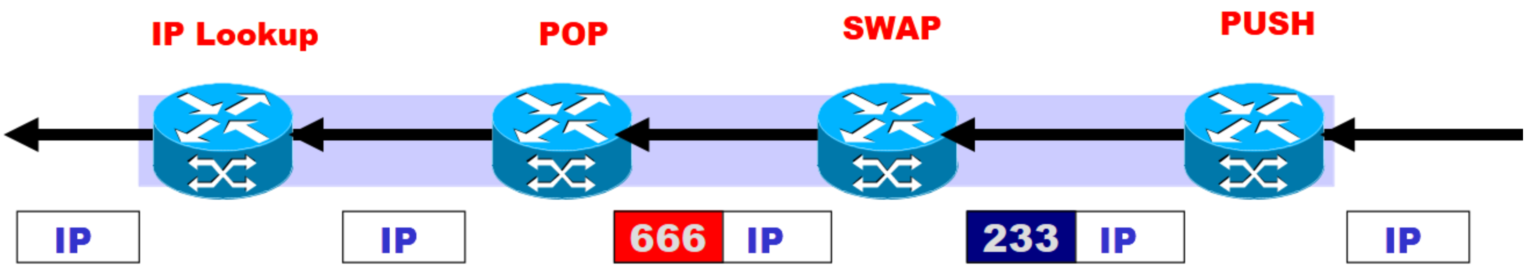


Penultimate Hop Popping (PHP)

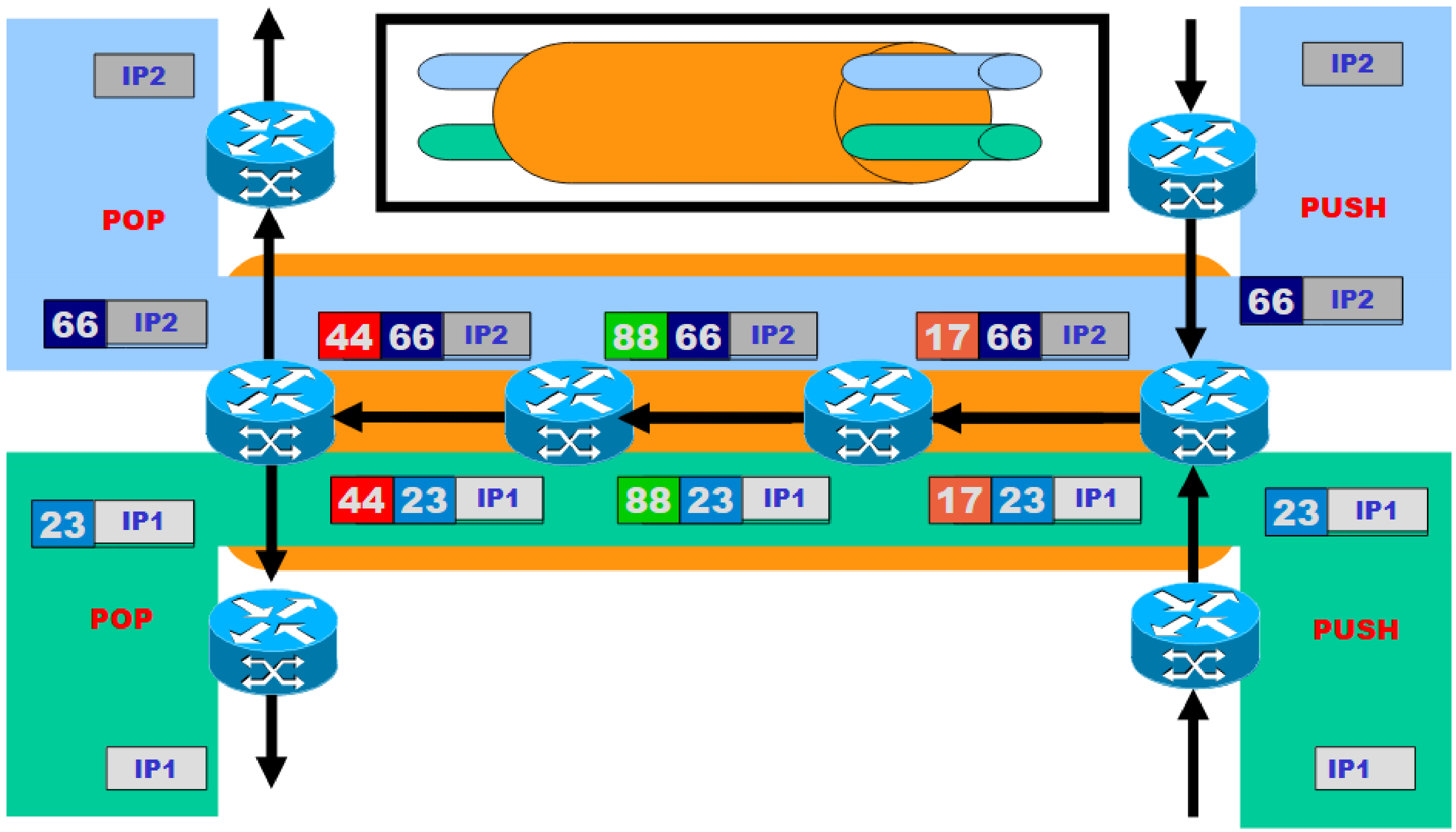
- An LSP from an entry Edge Router to an exit Edge Router:



- PHP is the process of label popping in the last Core Router before the packet reaches its exit Edge Router
  - it reduces the Edge Router load (usually, Edge Routers have more processing tasks than Core Routers)



LSP Hierarchy via Label Stacking



- The blue and the green LSPs are supported over the orange LSP in the central part of the MPLS network

MPLS Control Plane: establishment of LSPs

- LSPs for best-effort services
  - Path is chosen based on shortest paths of IGP protocol (OSPF, IS-IS).
  - Label Distribution Protocol (LDP).
- LSPs with resource reservation
  - Constrained by explicit path definition and/or performance requirements (e.g., available bandwidth).
  - Resource Reservation Protocol with Traffic Engineering (RSVP-TE).
    - Extension of RSVP to support traffic engineering and label distribution.
  - Constrained based Routing LDP (CR-LDP).
    - Extension of LDP to support constrained routing.
    - Deprecated!

Label Distribution Protocol (LDP)

- LDP specification is defined in RFC 5036 (Oct 2007)
  - There is a LDPv6 version to support IPv6, defined in RFC 7552: Updates to LDP for IPv6 (June 2015)
- It includes automatic discovery of LSRs (based on Hello messages running over UDP).
- Label mappings are exchanged over TCP for reliability.
- Label mappings are assigned by LSRs in the opposite direction of the LSPs (illustrated in slide 29)
- Modes of behaviour are negotiated during session initialization:
  - Label distribution (independent of ordered)
  - Label advertisement (unsolicited or on-demand)
  - Label retention (liberal or conservative)

Modes of Behaviour of LDP

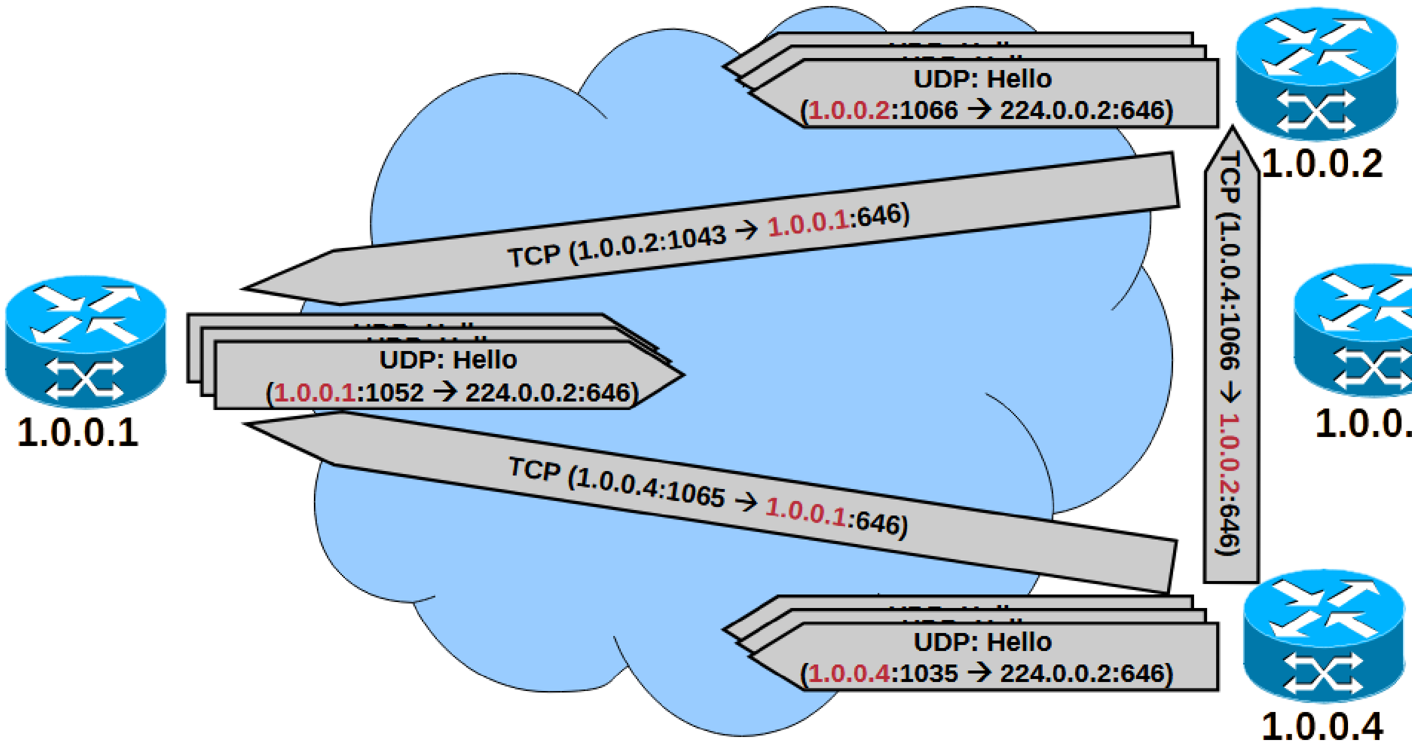
- Label distribution (independent of ordered)
  - Independent: each LSR may advertise label mappings to its neighbours at any time
  - Ordered: each LSR may advertise a label mapping only for a FEC for which it has a label mapping for the FEC next hop, or for which the LSR is the egress
- Label advertisement (downstream unsolicited or on-demand)
  - Downstream Unsolicited: label mapping advertisements for all routes may be received from all neighbour LSRs
  - Downstream on Demand: LSRs request label mappings from next-hop LSR (according to routing) and label mapping advertisements are sent upon requests
- Label retention (liberal or conservative)
  - Liberal: label mappings received from a neighbour LSR are retained regardless of whether the LSR is the next hop for the advertised mapping
  - Conservative: received label mappings are retained only if they are from a valid next-hop LSR (according to routing)

LDP Messages

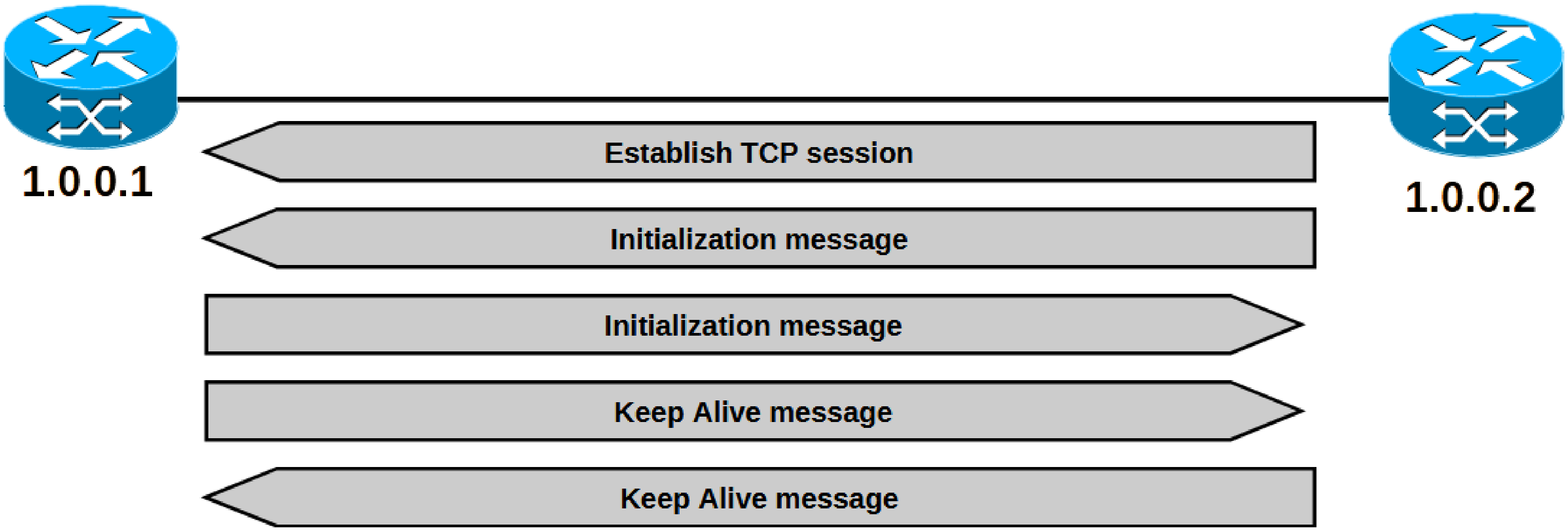
- Discovery messages
  - Announce and maintain the presence of an LSR in a network.
  - **Hello Messages** (UDP) sent to “all-routers” multicast address (224.0.0.2).
  - Once neighbour is discovered, a LDP session is established over TCP.
- Session messages
  - Establish (**Initialization Message**) and maintain (**Keep Alive Message**) sessions between LDP peers.
- Advertisement messages
  - When a new LDP session is initialized and before sending label information, an LSR advertises its interface addresses with one or more **Address Messages**.
  - An LSR withdraw previously advertised interface addresses with **Address Withdraw Messages**.
  - Create, change, and delete label mappings for FECs, with **Label Mapping**, **Label Request**, **Label Abort Request**, **Label Withdraw**, and **Label Release Messages**.
- Notification messages
  - Provide advisory information and/or error information.

LDP Neighbour Discovery

- Hello messages (UDP) are periodically sent on all interfaces enabled for MPLS to a “all-routers” multicast address (224.0.0.2).
- On each network, when a pair of MPLS enabled routers know each other (each one has received a Hello message from the other), the router with the highest IP address establishes of an LDP/TCP session between them.
- Both TCP and UDP messages use well-known LDP port number 646.

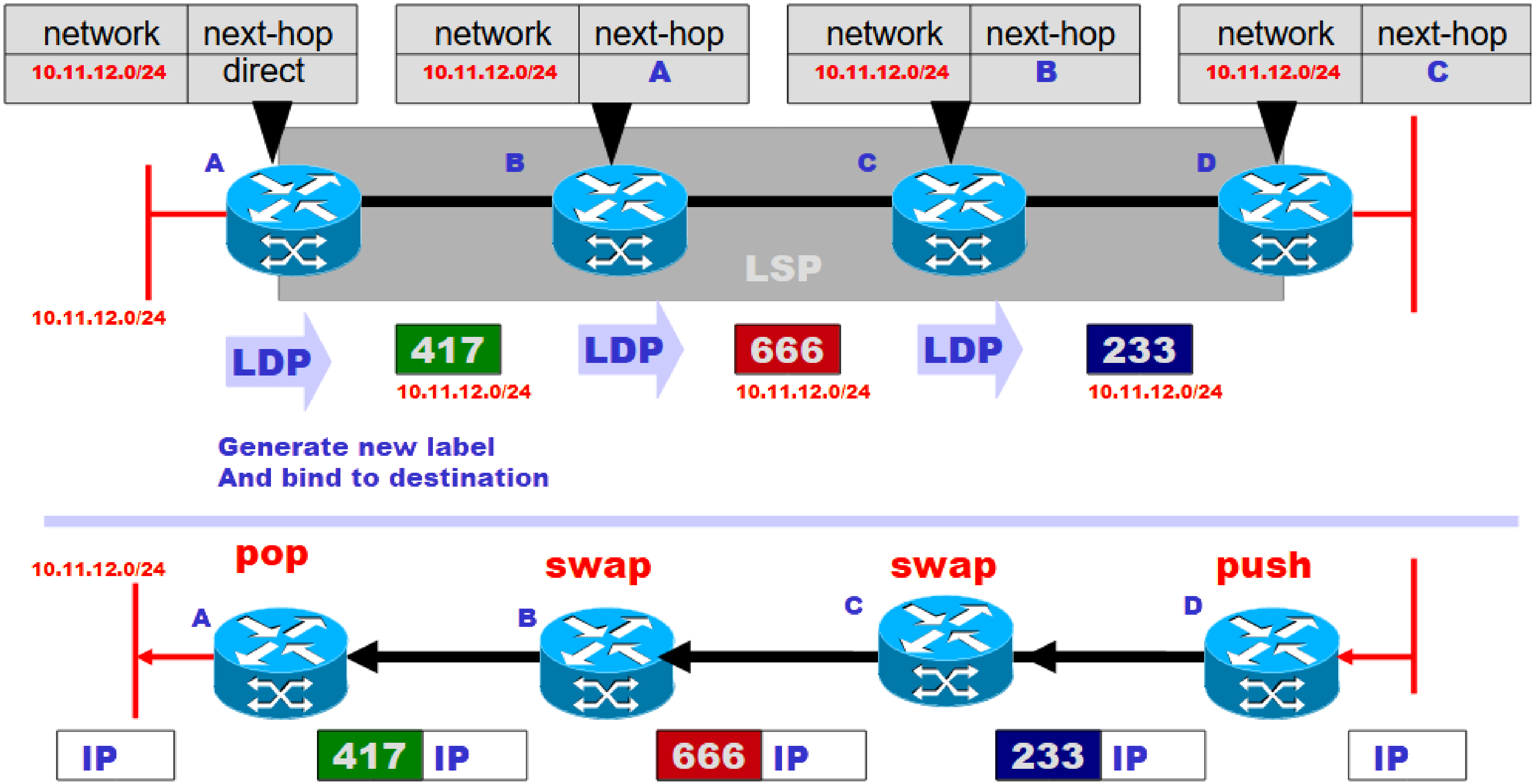


LDP Session Negotiation



- Peers first exchange Initialization messages.
- The session is ready to exchange label mappings after receiving the first Keep Alive message.
  - Keep Alive messages are resent periodically to maintain the LDP/TCP session active.

LDP and Hop-by-Hop routing



MPLS Traffic Engineering

- Traffic Engineering is the process selecting the LSPs to best fit the traffic in the capacity resources of the network
  - LDP is appropriate to set up LSPs supporting best-effort services (services that work properly whatever resources are available)
  - LSPs supporting services requiring guaranteed QoS need to be established with an associated reservation of resources
- LSPs with resource reservation
  - Resource Reservation Protocol with Traffic Engineering (RSVP-TE)
    - Extension of RSVP to support traffic engineering and label distribution
  - Constrained based Routing LDP (CR-LDP): deprecated
- The routes of LSPs with resource reservation can be computed:
  - Offline: by a management system
  - Online: through Constrained based Routing
    - LSPs are established dynamically and automatically when needed and teared down when are not required anymore

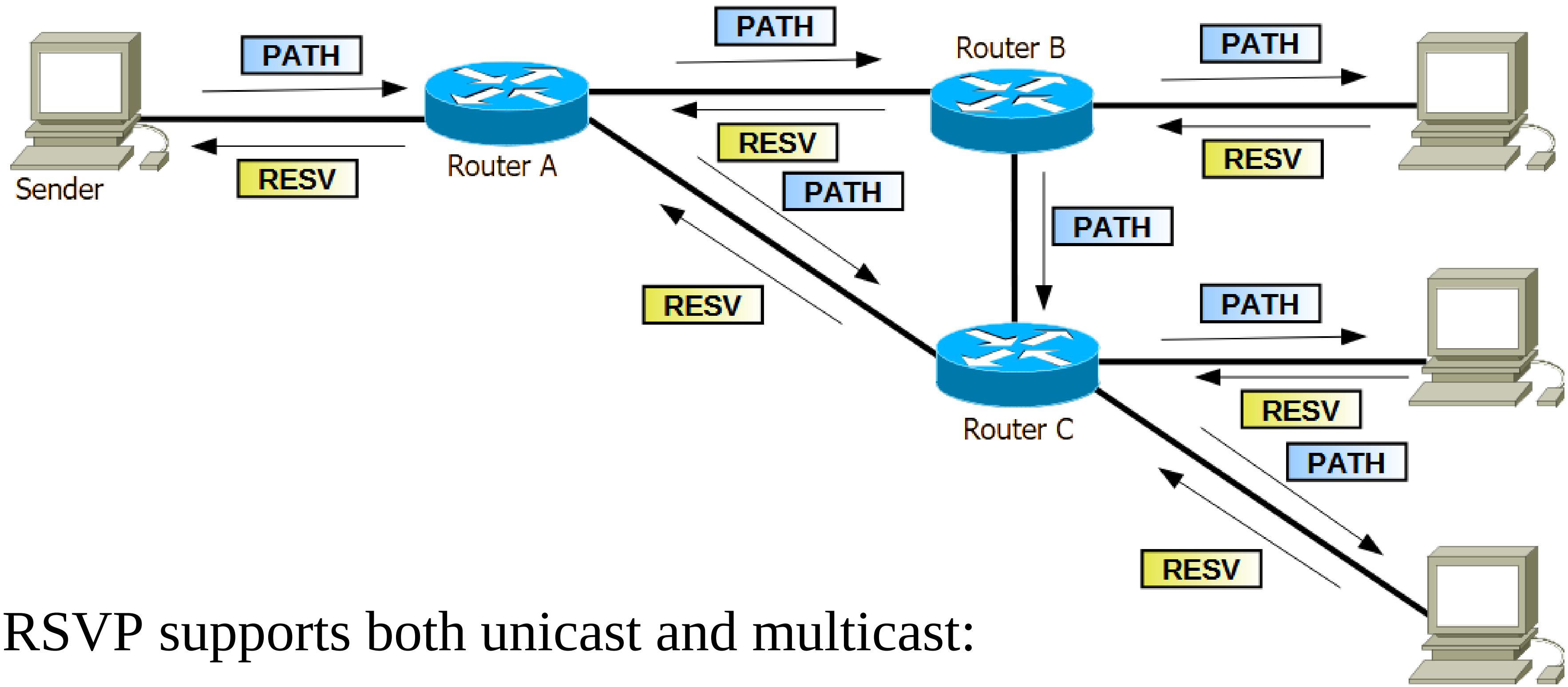
Resource Reservation Protocol (RSVP)

- RSVP is a signalling protocol allowing to request resource needs between hosts and network devices for IP flows:
  - The source describes the characteristics of the IP packets flow.
  - The destination(s) describe(s) the reservation he/they need.
  - The Routers in the routing path from source to destination(s) set up the requested reservation on their outgoing interfaces.
- Signalling is based on PATH and RESV messages:
  - PATH announces the IP flow characteristics by the sender.
  - RESV specifies the reservation request by the receiver(s).
  - If the reservation is not possible in a router, a RESV ERR message is sent by the router towards the receiver(s).
- RSVP runs over IP (IP protocol type = 46).
- PATH and REST message are sent periodically to maintain the routers' reservation states (soft states).
- RSVP defines a "Session" with a particular destination and transport-layer protocol to identify a data flow.

RSVP Messages

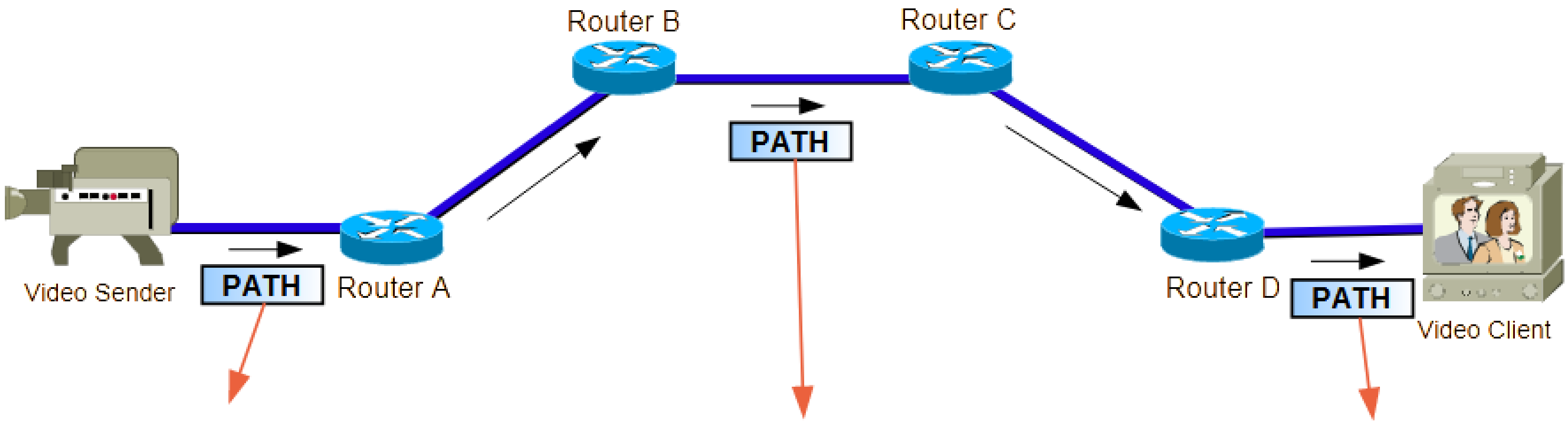
- PATH (Type = 0x01)
  - **Tspec** (“flow traffic specification”): contains the parameters that describe the traffic source based on the “Token Bucket” model
- RESV (Type = 0x02)
  - **Tspec**: the same that was received on the PATH message
  - **FilterSpec** (“filter specification”): contains the flow descriptor that enables routers to identify packets belonging to this reservation (source address, destination address, protocol type, source port number, destination port number, any combination of these parameters)
  - **Rspec** (“flow reservation specification”): contains the parameters describing the reservation that the receiver wants to become supported
    - Rspec is specified if the receiver wants a service of the “guaranteed service” type: it specifies the bandwidth to be reserved in the path
    - When Rspec is not specified, it means that the receiver wants a service of the “controlled load” type: it is up to the routers to decide the bandwidth to be reserved to that the data flow does not suffer congestion.

RSVP signalling



- RSVP supports both unicast and multicast:
  - Both messages are routed based on the IP routing protocol
    - Multicast routing protocols will be the next module of this course unit
  - PATH messages: paths are selected by the routers
  - RESV messages: routers set up the reservation on the selected paths
    - In multicast, routers make reservation aggregation

RSVP PATH (Example)

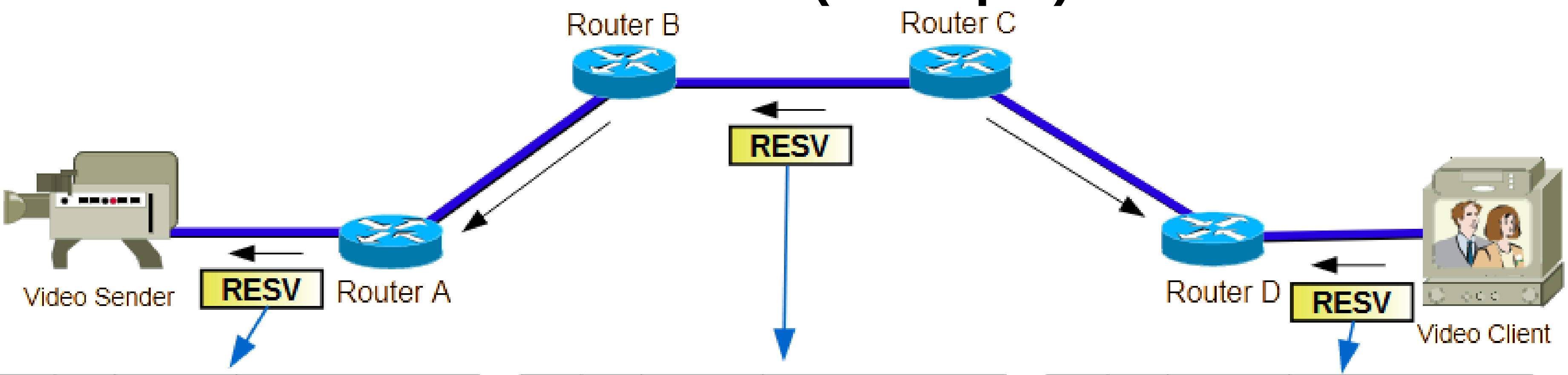


Vs.: 4	iHL: 5	Service	Total Length: 60	
Identification		Flg	Fragment Offset	
Time to Live	Protocol: 46	Header Checksum		
Source Address: Video Server				
Destination Address: Video Client				
1	0	Type: 1	Checksum	
Send_TTL	0	Message Length: 40		
SESSION Length.: 12		Class Nº: 1	Class Type: 1	
Destination Address: Video Client				
Protocol ID	Flags	Destination port		
RSVP_HOP Length. : 12		Class Nº: 3	Class Type: 1	
Last Hop Address: Video Server				
Logical Interface Handle of the last node (LIH)				
TIME_VALUES Length: 8		Class Nº: 5	Class Type: 1	
Update Period (ms)				

Vs.: 4	iHL: 5	Service	Total Length: 60	
Identification		Flg	Fragment Offset	
Time to Live	Protocol: 46	Header Checksum		
Source Address: Video Server				
Destination Address: Video Client				
1	0	Type: 1	Checksum	
Send_TTL	0	Message Length: 40		
SESSION Length: 12		Class Nº: 1	Class Type: 1	
Destination Address: Video Client				
Protocol ID	Flags	Destination Port		
RSVP_HOP Length: 12		Class Nº: 3	Class Type: 1	
Last Hop Address: Router B				
Logical Interface Handle of the last node (LIH)				
TIME_VALUES Length: 8		Class Nº: 5	Class Type: 1	
Update Period (ms)				

Vs.: 4	iHL: 5	Service	Total Length: 60	
Identification		Flg	Fragment Offset	
Time to Live	Protocol: 46	Header Checksum		
Source Address: Video Server				
Destination Address: Video Client				
1	0	Type: 1	Checksum	
Send_TTL	0	Message Length: 40		
SESSION Length: 12		Class Nº: 1	Class Type: 1	
Destination Address: Video Client				
Protocol ID	Flags	Destination Port		
RSVP_HOP Length: 12		Class Nº: 3	Class Type: 1	
Last Hop Address: Router D				
Logical Interface Handle of the last node (LIH)				
TIME_VALUES Length: 8		Class Nº: 5	Class Type: 1	
Update Period (ms)				

RSVP RESV (Example)



Vs.: 4	IHL: 5	Service	Total Length
Identification		Flg	Fragment Offset
Time to Live	Protocol: 46	Header Checksum	
Source Address: Router A			
Destination Address: Video Server			
1	0	Type: 2	Checksum
Send_TTL	0	Message Length	
SESSION Length: 12		Class Nº: 1	Class Type: 1
Destination Address: Video Client			
Protocol Id	Flags	Destination protocol port	
RSVP_HOP Length: 12		Class Nº: 3	Class Type: 1
Address of the last node: Router A			
Logical Interface Handle of the last node (LIH)			
TIME_VALUES Length: 8		Class Nº: 5	Class Type: 1
Update period (ms)			
STYLE Object Length : 8		Class Nº: 8	Class Type: 1
Flags	Style Option Vector: 0x00000A (FF)		
FLOWSPEC Length		Class Nº: 9	Class Type
FLOWSPEC object contents			
FILTER_SPEC Length: 12		Class Nº: 10	Class Type: 1
Source Address: Video Server			
Reserved	Reserved	Source protocol port	

Vs.: 4	IHL: 5	Service	Total Length
Identification		Flg	Fragment Offset
Time to Live	Protocol: 46	Header Checksum	
Source Address: Router C			
Destination Address: Router B			
1	0	Type: 2	Checksum
Send_TTL	0	Message Length	
SESSION Length: 12		Class Nº: 1	Class Type: 1
Destination Address: Video Client			
Protocol Id	Flags	Destination protocol port	
RSVP_HOP Length: 12		Class Nº: 3	Class Type: 1
Address of the last node: Router C			
Logical Interface Handle of the last node (LIH)			
TIME_VALUES Length: 8		Class Nº: 5	Class Type: 1
Update period (ms)			
STYLE Object Length : 8		Class Nº: 8	Class Type: 1
Flags	Style Option Vector: 0x00000A (FF)		
FLOWSPEC Length		Class Nº: 9	Class Type
FLOWSPEC object contents			
FILTER_SPEC Length: 12		Class Nº: 10	Class Type: 1
Source Address: Video Server			
Reserved	Reserved	Source protocol port	

Vs.: 4	IHL: 5	Service	Total Length
Identification		Flg	Fragment Offset
Time to Live	Protocol: 46	Header Checksum	
Source Address: Video Client			
Destination Address: Router D			
1	0	Type: 2	Checksum
Send_TTL	0	Message Length	
SESSION Length: 12		Class Nº: 1	Class Type: 1
Destination Address: Video Client			
Protocol Id	Flags	Destination protocol port	
RSVP_HOP Length: 12		Class Nº: 3	Class Type: 1
Address of the last node: Video Client			
Logical Interface Handle of the last node (LIH)			
TIME_VALUES Length: 8		Class Nº: 5	Class Type: 1
Update period (ms)			
STYLE Object Length : 8		Class Nº: 8	Class Type: 1
Flags	Style Option Vector: 0x00000A (FF)		
FLOWSPEC Length		Class Nº: 9	Class Type
FLOWSPEC object contents			
FILTER_SPEC Length: 12		Class Nº: 10	Class Type: 1
Source Address: Video Server			
Reserved	Reserved	Source protocol port	

Resource Reservation Protocol with Traffic Engineering (RSVP-TE)

- Defined in:
  - RFC 3209: RSVP-TE: Extensions to RSVP for LSP Tunnels (Dec 2001)
  - RFC 5151: Resource Reservation Protocol-Traffic Engineering (RSVP-TE) Extensions (Feb 2008)
- RSVP-TE is an extension of RSVP.
- It allows to map traffic flows onto the physical network topology through LSPs.
- It requires resource and constraint network information.
  - Provided by Extended Link State Protocols (IS-IS or OSPF with traffic engineering extensions).
    - RFC 3630: Traffic Engineering (TE) Extensions to OSPF Version 2 (Sep 2003)
    - RFC 5305: IS-IS Extensions for Traffic Engineering (Oct 2008)

Extensions to RSVP for LSP Tunnels

- The SENDER_TEMPLATE (or FilterSpec) object together with the Session object uniquely identifies an LSP tunnel (flow).
- LSP Tunnel related new objects
  - **Explicit Route**: carried in PATH and contains a sequence of variable-length items, that can be IPv4 prefixes, IPv6 prefixes or autonomous system numbers.
  - **Label Request**: carried in PATH requesting a label for a specific tunnel/flow.
  - **Label**: carried in RESV messages and contain a single label for a specific tunnel/flow.
  - Record Route: carried in PATH and RESV, used to collect detailed path information and useful for loop detection and diagnostics.
  - Session Attribute: carried in PATH, used to define the type and name of the session/tunnel/flow, also used to define priority values.
- LSP Tunnel related new object types
  - Session object new types: LSP_TUNNEL_IPv4 and LSP_TUNNEL_IPv6
  - Tspec object new types: LSP_TUNNEL_IPv4 and LSP_TUNNEL_IPv6
  - FilterSpec object new types: LSP_TUNNEL_IPv4 and LSP_TUNNEL_IPv6

Traffic Engineering with MPLS

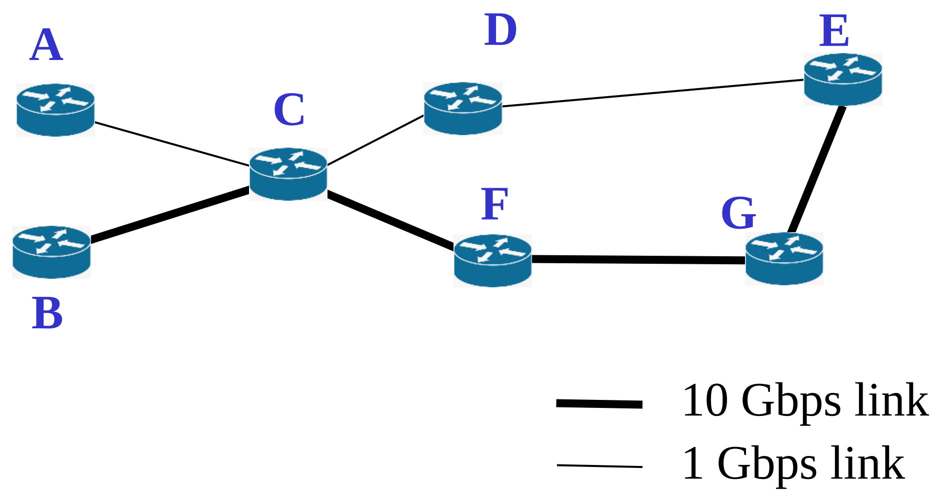
- In order to fit the traffic flow throughputs in the capacity resources of the network, we need to route each traffic flow through:

A – E: A → C → D → E

B – E: B → C → F → G → E

C – E: C → F → G → E

N1	N2	N1→N2	N2→N1
A	E	1.0 Gbps	-
B	E	5.0 Gbps	-
C	E	5.0 Gbps	-



- This solution can be implemented with MPLS:
  - setting routers A, B, C and E as edge LSRs
  - establishing the 3 LSPs with RSVP-TE