Slide supporting material

Lesson 13: MPLS Networks

Giovanni Giambene

Queuing Theory and Telecommunications: Networks and Applications
2nd edition, Springer

IP Over ATM

- Once defined IP networks, it became important to determine lower layer technologies able to efficiently **transport IP traffic**. We focus here on ATM, but this study is also suitable to other layer 2 technologies.
- The concept of **adjacency**: there is a layer 3 adjacency for two directly-connected IP routers; there is a layer 2 adjacency between to ATM nodes connected by virtual circuits; there is a layer 1 adjacency for interfaces connected to the same physical transmission medium.
- We speak about **interoperability** when two nodes work together, but at different OSI layers.

IP Over ATM (cont'd)

- The IETF RFC 1483 addressed the IP traffic over ATM and the problem of the **inefficient mapping of IP datagrams on ATM (short) cells**; the use of AAL5 was proposed. However, ATM AAL5 cannot multiplex different higher-layer traffic flows on the same virtual connection (different traffic flows from the same host need distinct ATM connections).
- Another problem was related to the **support of IP routing** in ATM networks as discussed in the next slide.

IP Over ATM

- Two models have been considered for IP traffic over ATM networks:
  - "**Overlay Model**" (i.e., the Classical IP over ATM, CIP, approach): IP/**AAL5**/ATM/PHY defined in RFC 1577 and RFC 2225. Here the problem is the duplication of routing and switching functions at layers 3 and 2, respectively (inefficiency). Permanent Virtual Connections (PVCs) are used.
  - "**Integrated Model**": IP+ATM(**AAL5**)/PHY. Here the problem is the complexity and the presence of many non-standard solutions.

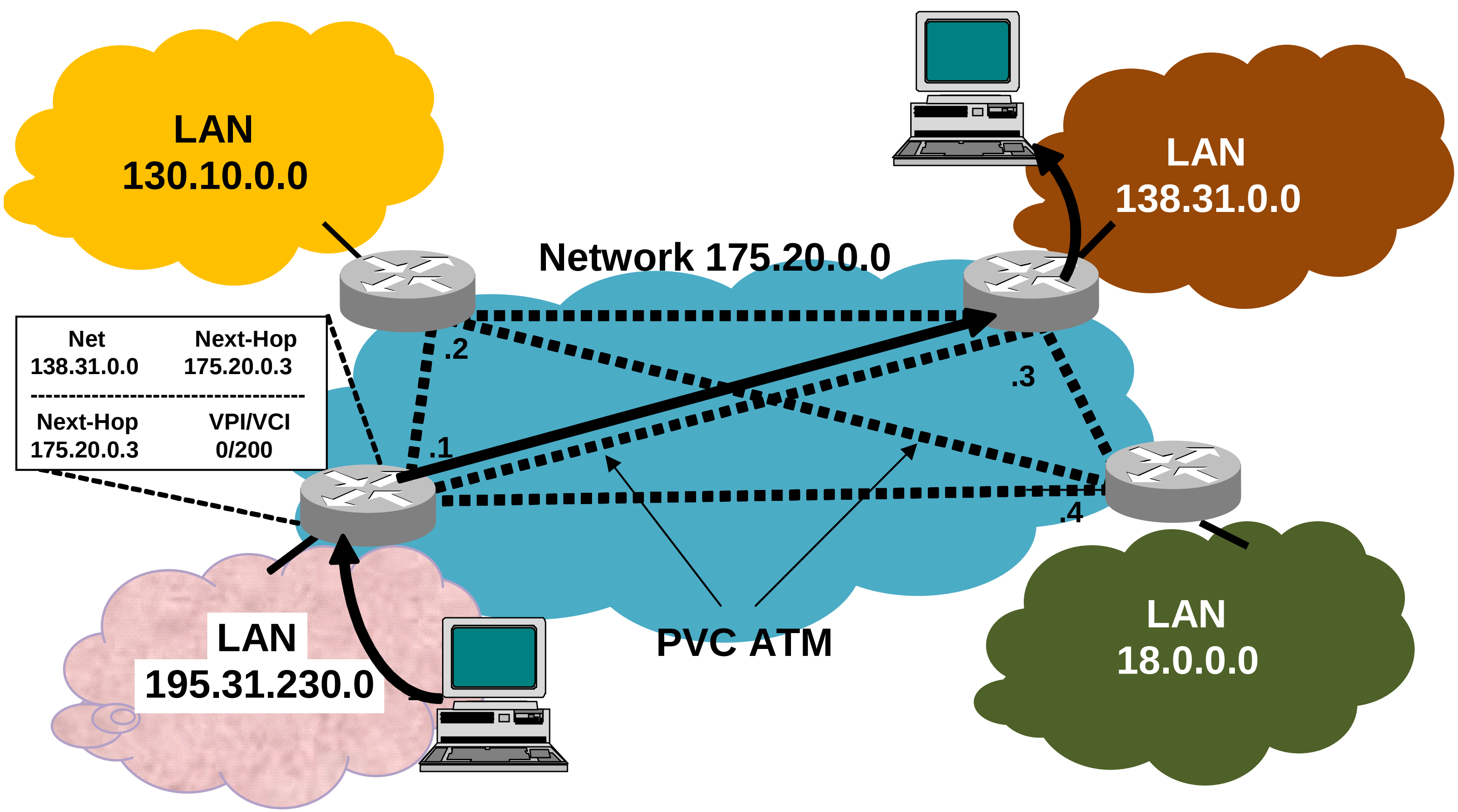
IP Over ATM: Overlay Model

- If the ATM network is not completely meshed, the **IP datagram is reassembled and segmented again at each crossed node (= router over ATM), by means of the SAR sub-layer of ATM**. This entails a waste of processing resources at each router that could cause congestion if a router is not appropriately designed.
- This is the reason why the overlay model tend to be implemented in a **full-mesh topology**, thus allowing layer 2 (ATM) adjacencies among routers.
- The use of a fully meshed ATM network is quite complicated due to the huge number of layer 2 adjacencies required with related PVCs.
  - Let us consider n routers connected through ATM; totally, **n(n–1)/2 bi-directional links [and n(n–1) layer 2 mono-directional ATM virtual circuits at layer 2]** are needed.

IP Over ATM: Overlay Model (cont'd)

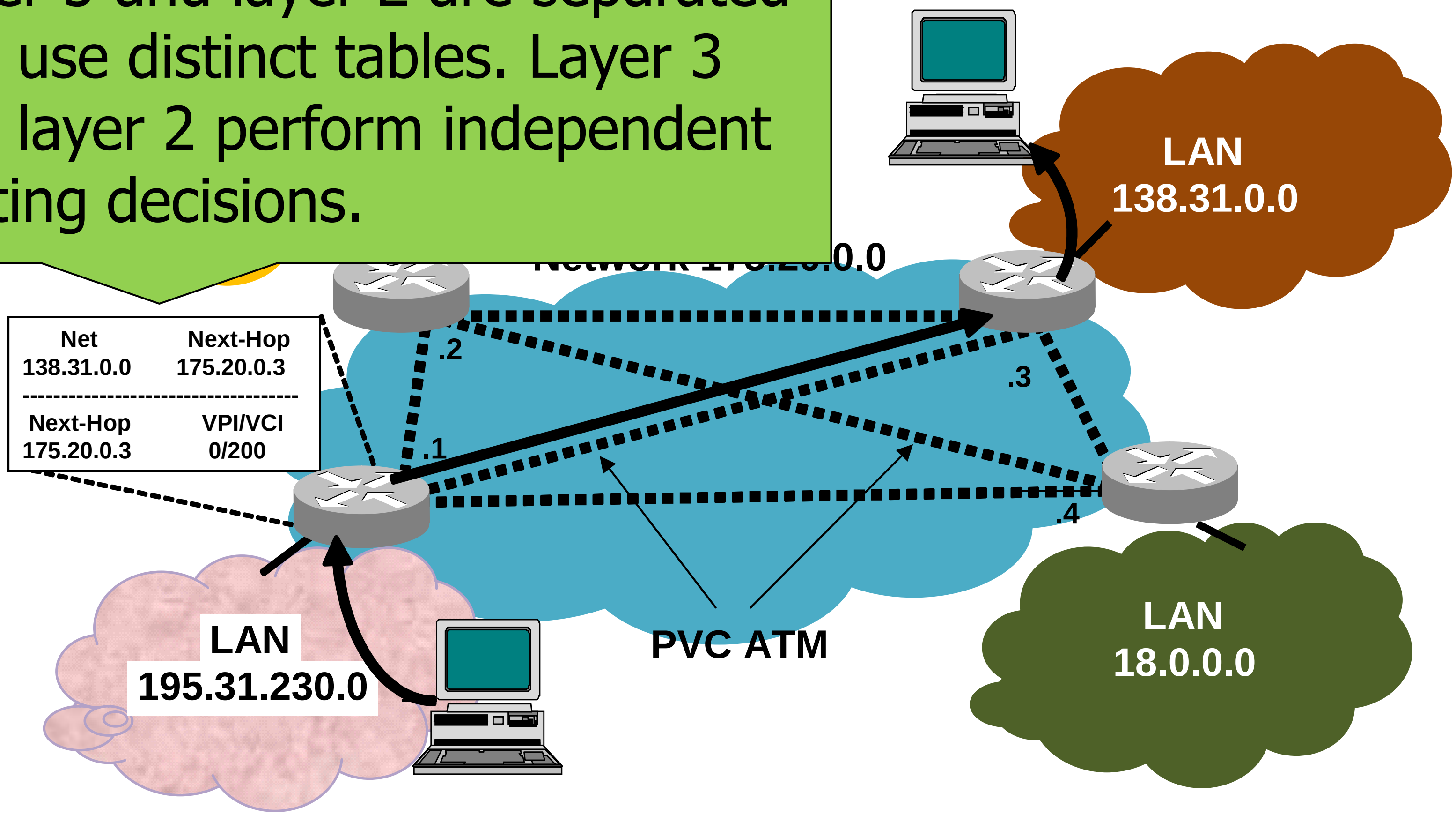
- If the ATM network (Logical IP Subnet, LIS) is configured to use only PVCs, **L = n(n–1) mono-directional virtual circuits** have to be pre-configured at layer 2 for n routers for the **full-mesh topology**. ATM level complexity is $O(n^2)$.
  - There is a limit in the number of layer 2 adjacencies that can be managed by an ATM switch.
  - The IP routing protocol of the **OSPF type needs to exchange $O(n \times L) = O(n^3)$ signaling messages to configure the routing tables** of a full-mesh LIS.
  - **When an ATM physical link fails**, all PVCs using that VC fail and many routers have to update their routes at the same time. This entails from **$O(n^3)$ to $O(n^4)$ routing messages** to be exchanged by OSPF among routers to reconfigure the paths. This is what is called **routing storm**, which may cause routing to become unstable after a single link failure event.

IP Over ATM: Overlay Model with Fully Mesh Topology



IP Over ATM: Overlay Model with Fully Mesh Topology

Layer 3 and layer 2 are separated and use distinct tables. Layer 3 and layer 2 perform independent routing decisions.



IP Over ATM: Integrated Model

- This is an evolution of the CIP approach to reduce the functional redundancies between IP and ATM for what concerns routing.
- The routing of IP packets is made by **enriching the routing table at each router with the correspondence between the "next hop" and the VPI/VCI identifiers** of the ATM PVC that connects to the next router.
    - **New types of routers are needed** to support routing integrated with layer 2 information.
    - **IP datagrams are conveyed by ATM cells on virtual circuits determined by an IP routing protocol (e.g., OSPF)**.

Multi-Protocol Label Switching (MPLS)

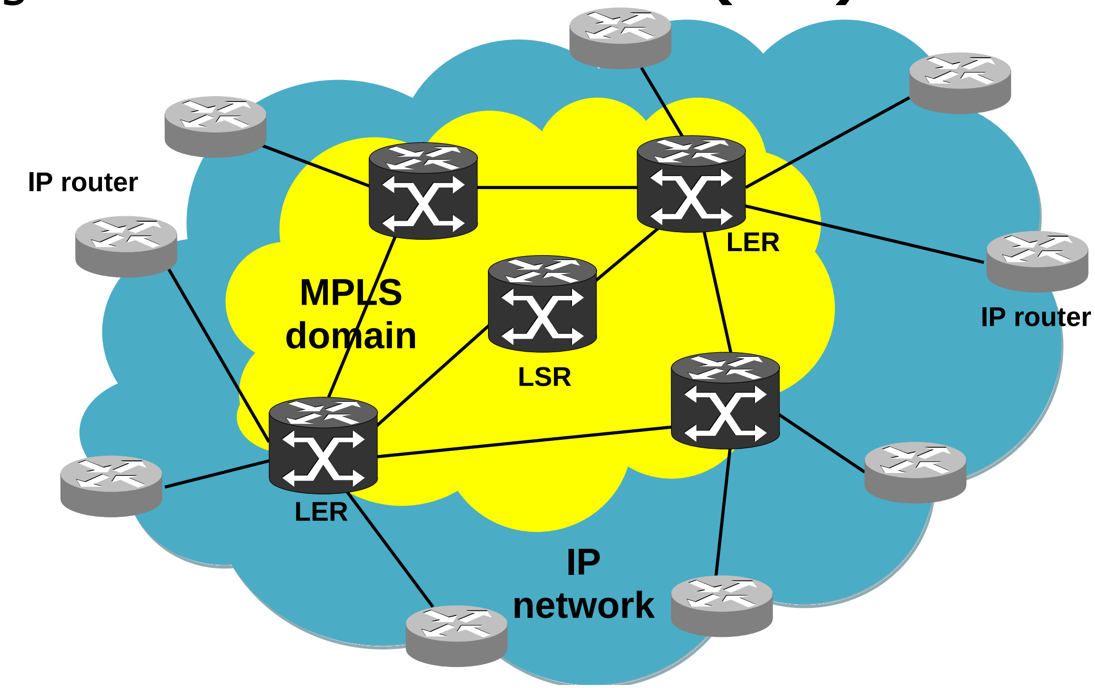
- Starting from the integrated approach, MPLS (RFC 3031) was standardized for:
  - An **efficient integration** of network layer traffic with different layer 2 technologies, not only ATM;
  - **A speed increase in forwarding IP traffic** at the nodes;
  - Enriching the IP routing with new functionalities (e.g., **traffic engineering** aspects);
  - A greater scalability in IP networks in managing huge traffic loads and to provide services like Virtual Private Networks (VPN);
  - Introducing **mechanisms for QoS support** in IP networks that typically provide best-effort services.

Multi-Protocol Label Switching (MPLS)

- MPLS (1997) is a **connection-oriented protocol** employed to route IP traffic over different layer 2 technologies such as ATM, Frame Relay, and Ethernet.
- The fundamental elements of an MPLS network are:
  - **Label Edge Routers (LERs)**: these are high-speed routers placed at the boundary of the MPLS network (domain) and are used to determine the associations between hops and labels (i.e., the label-switched path).
  - **Label Switch Routers (LSRs)**: these high-speed (core) routers are used to switch data units on the basis of the labels they convey.

Multi-Protocol Label Switching (MPLS), cont'd

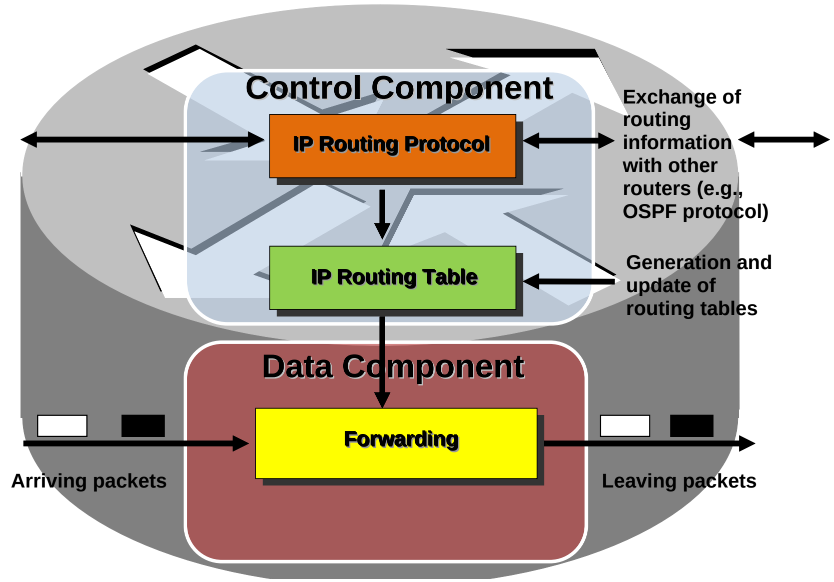
- An MPLS network if physically inserted in an IP network, but its operation is ideally distinguished. LERs (at the borders) receive IP datagrams; LERs label these datagrams on the basis of their destination and forward them through the MPLS domain along a given **Label-Switched Path (LSP) as in a tunnel**.



The LSP is build with a **set-up phase** where all the routers along the path are instructed on how to switch the labels.

Multi-Protocol Label Switching (MPLS), cont'd

- In the IP world, routing can be functionally decomposed into the data plane and the control plane:
  - The **data component** is in charge of the actual forwarding of IP packets from input to output across a switch or router. The **forwarding table** maintained by a router and the information carried in the packet header are used to forward the packets to the next hop.
  - The **control component** is responsible for the construction and maintenance of the forwarding table. It consists of one or more **routing protocols** that support the exchange of information among routers and the procedures used by a router to update its forwarding table.



FEC in MPLS

- Choosing the next hop is the composition of two functions:
  - The first function partitions all IP packets into a **set of Forwarding Equivalence Classes (FECs)**;
  - The second **maps each FEC to a next hop**.
- The assignment of a packet to a particular FEC is done just once, when the **packet reaches the ingress LER**.
  - A FEC corresponds to a path in the MPLS domain with suitable characteristics in terms of available bandwidth, priority, etc.
  - A group of packets can use the same FEC (sharing the same transport requirements).
  - The definition of a FEC depends on several aspects such as: the address of the destination network, the precedence level, the existence of a source-destination reserved path and traffic engineering considerations.

FEC in MPLS (cont'd)

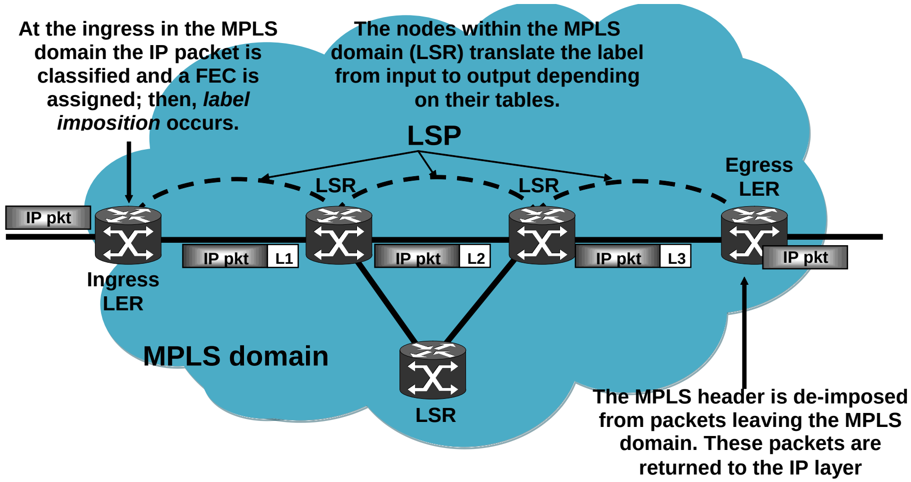
- When a packet reaches an LSR internal to the MPLS domain, its label is examined to decide the output interface where to forward the packet and the new label to be used. **It is therefore not necessary to scan the entire IP routing table; the forwarding procedure is immediate.**
  - Within the MPLS domain, the IP packet header is not used.
- A **FEC corresponds to a local label for each hop along the path in the MPLS domain.**
- When a packet is forwarded to its next hop, **the local label is sent along with it**.

Use of Labels in MPLS

- A **label binding is the association of a local label and a FEC** for the switching operations at an LSR .
- An **LSP is the set of label bindings** for the different hops from input to output of an MPLS network.
- It is the **responsibility of LER to recognize the FEC corresponding to a given packet**, then to assure that the packet is forwarded in the related LSP by imposing the appropriate label on top of the packet. The label-to-FEC correspondence has to be unique.
- Each LSR builds a **table** to specify how a packet must be forwarded.
- **MPLS handles labels just like all other virtual circuit identifiers are handled in other virtual circuit switching technologies.**

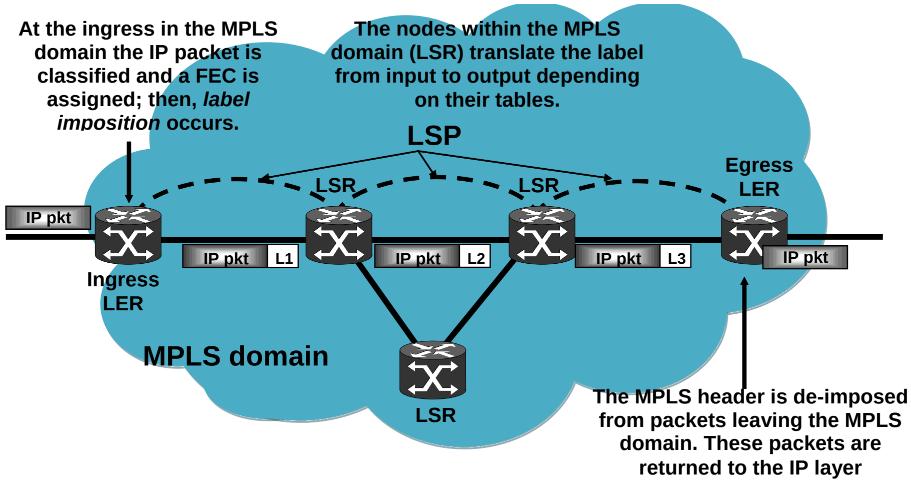
Forwarding in an MPLS Domain

- Let us consider an IP packet entering an MPLS domain. When a packet arrives at the first router of the MPLS domain, **ingress LER**, the IP packet header is analyzed.
- Let us assume that the data in the IP packet header match an already-defined FEC and related LSP in a LER table. Then, the ingress LER inserts (i.e., pushes) the **MPLS header** (with label L1) on top of the IP packet.



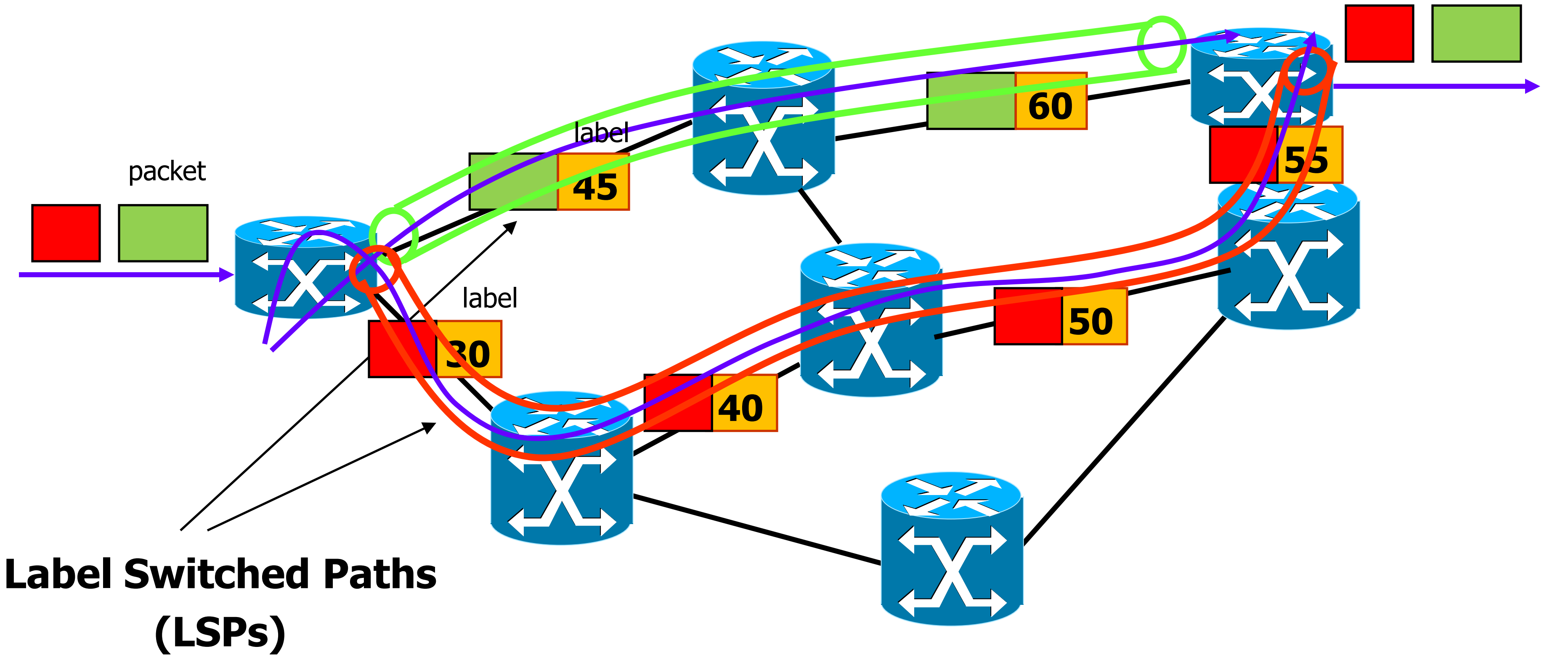
Forwarding in an MPLS Domain (cont'd)

- Subsequent LSRs along the LSP in the MPLS domain update the MPLS header by **swapping the label** (L1 against L2, L2 against L3) according to the instructions in the LSR tables.
- Finally, the last router of the LSP, called **egress LER**, removes (i.e., pops) the MPLS header (i.e., L3), so that the packet can be handled by subsequent MPLS-unaware IP routers.



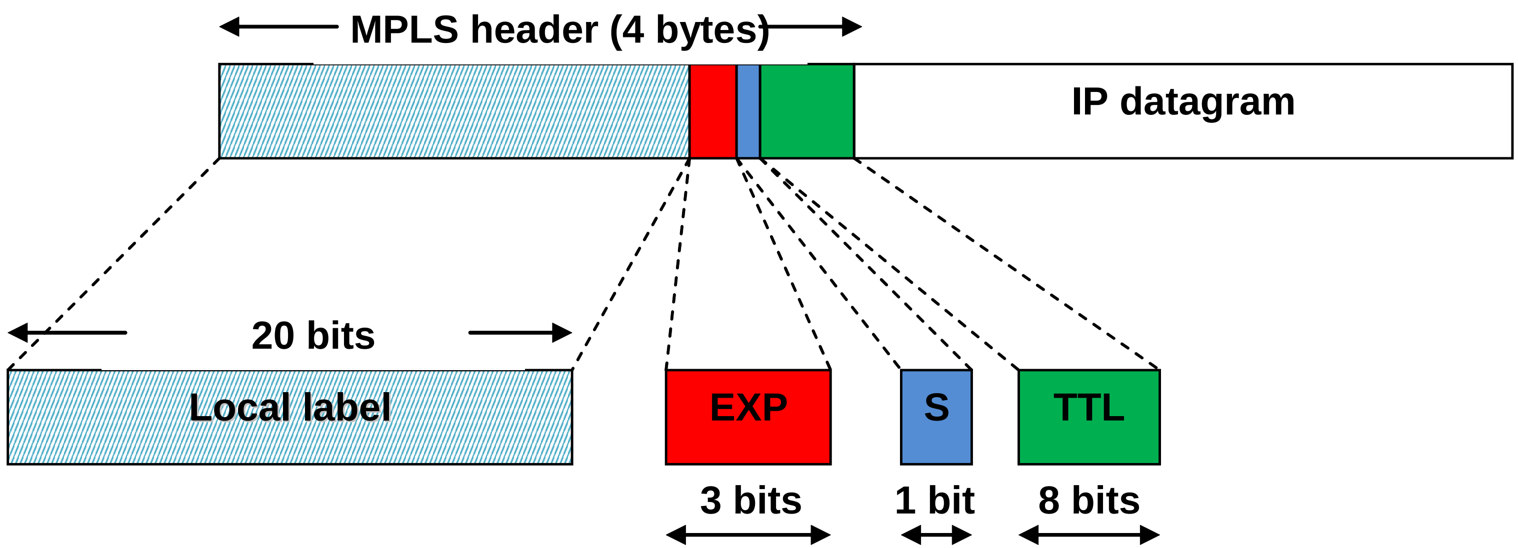
LSPs and Forwarding

- An LSP is analogous to a virtual path in connection-oriented networks.



MPLS Header

- **The shim MPLS header (4 bytes) is composed of four fields of fixed length: label field (20 bits), EXP field (3 bits), S flag (1 bit) and Time-To-Live, TTL, field (8 bits)**.
- The shim MPLS header is between the MAC header and the IP header.



MPLS Header (cont'd)

- **Label**: This 20-bit field carries the actual value of the local label. The characterization of this field depends on the protocol used to assign and to distribute the labels among LSRs.
- **EXP**: These 3 bits have an experimental use in order to identify traffic classes or network congestion.
  - MPLS can encode congestion by means of a single EXP bit (ECN bit). The 2 other EXP bits can be used for other scopes (e.g., PHB encoding with the DiffServ approach for QoS provision).
- **S**: This bit is used for label stack functions, that is when multiple **MPLS headers are stacked** (S = 1 denotes the last element in the stack: at the next hop the IP datagram leaves the current MPLS domain).
- **TTL**: This field is a counter decreased at each hop; it is used to reproduce at the MPLS level the same mechanism used at the IP level.

Thank you!

giovanni.giambene@gmail.com