Chapter 6: The Multi-Protocol Label Switching Architecture

TOPICS

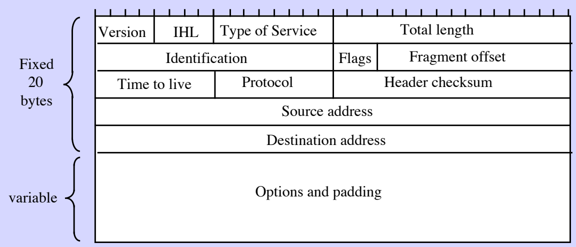
- IP: A primer
- The MPLS architecture
  - Label allocation schemes
  - Explicit routing
  - Label stack
  - Schemes for setting up an LSP
- MPLS over ATM

The Internet Protocol: A primer

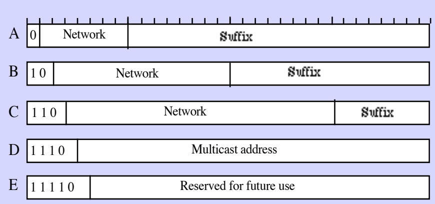
Features:

- IP is part of the TCP/IP suite of protocols.
- It provides a connectionless service.
- It does not guarantee delivery of IP packets.
- IP does not check the payload of an IP datagram for errors, but it only checks its IP header. IP will drop an IP datagram, if it finds that its header is in error. Lost or erroneous data is recovered by the destination's TCP.

The IPv4 header



IP addresses



Dotted decimal notation:
Each byte of the address is written in decimal, ranging from 0 to 255.
Example: the IP address 00000111 00000010 00000000 00000010 is written as 7.2.0.2.

Range of addresses..

- Class A: from 1.0.0.0 to 127.255.255.255,
- Class B: from 128.0.0.0 to 191.255.255.255,
- Class C: from 192.0.0.0 to 233.255.255.255.

Classless Inter-domain routing (CIDR)

Allocate blocks of class C network addresses to each ISP. Organizations using the ISP are sub-allocated a block of 2^n contiguous addresses.

This permits all these addresses to be advertised outside the ISP's network in an aggregate manner using part of the address known as the *prefix*.

ARP, RARP, and ICMP

- The TCP/IP protocol suite includes other protocols such as the
  - *address resolution protocol* (ARP)
  - *reverse address resolution protocol* (RARP)
  - *internet control message protocol* (ICMP)

The Multi-Protocol Label Switching (MPLS) Architecture

- MPLS is an IETF standard.
- It is based on *tag switching*, which was proposed by CISCO. (Tag switching was motivated by *IP switching*, a technique for switching IP packets over ATM)
- MPLS was developed exclusively for IP networks, despite the “multi-protocol” description in its title.

The IP router

- In order to understand the basic concepts behind MPLS, we need to take a look at the structure of an IP router.
- An IP router has a
  - *control* component, and a
  - *forwarding* component.

The control component

- It consists of routing protocols, such as OSPF, BGP, and PIM, which are used to construct routes and exchange routing information among routers.
- This information is used by the routers to construct a forwarding table (routing table), known as the *forwarding information base* (FIB).

The forwarding component

- It consists of procedures for forwarding an IP packet.
- The IP router uses the destination IP address to find an entry in the FIB, using the longest match algorithm. From this, it obtains an interface number, which is the output port connecting the IP router to the next-hop router, to which the IP packet should be sent.

Forwarding equivalent class (FEC)

- A FEC (pronounced *fec)* is the set of all forwarding addresses that have the same prefix.
- Thus, addresses in a router can be grouped into a number of disjoint FECs.
- IP packets belonging to the same FEC have the same output interface.
- In MPLS, an FEC is associated with a *label***.**

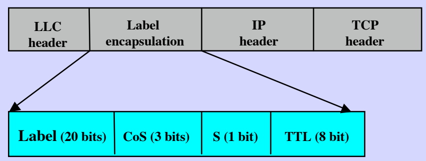
Labels

- A label is a short, fixed-length identifier that has local significance (i.e. it is valid on a single hop interconnecting two routers).
- A label in a packet represents the FEC to which the packet has been assigned.
- The label assigned to a packet is not an encoding of its destination address.

Where is the label carried?

- No space in IPv4 packet for a label.
- If the IP network is running on top of an ATM network/ Frame Relay, the label is carried in the VPI/VCI field/DLCI field.
- For Ethernet, and point-to-point connections running a link layer protocol, such as PPP, the label is encapsulated and inserted between the LLC header and the IP header (*shim header*).

Label encapsulation (Shim header)



CoS (class of service)
S - stack of labels
TTL (time to live)

Label stack

Label (20 bits)	CoS (3 bits)	S =0	TTL (8 bits)
Label (20 bits)	CoS (3 bits)	S =0	TTL (8 bits)
⋮			
Label (20 bits)	CoS (3 bits)	S = 1	TTL (8 bits)

Label Switching Routers (LSR)

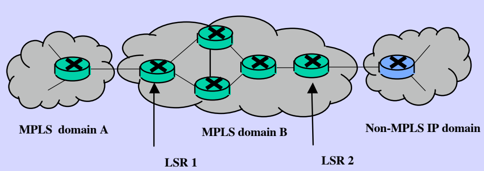
- An *LSR* is an IP router that runs MPLS. It is aware of MPLS control protocols and it operates one or more layer 3 routing protocols
  - It binds labels to FECs,
  - forwards packets based on their labels, and
  - it carries out the customary IP forwarding decision based on prefixes.

MPLS nodes

- An *MPLS node* is an LSR but it may not necessarily forward IP packets based on prefixes.

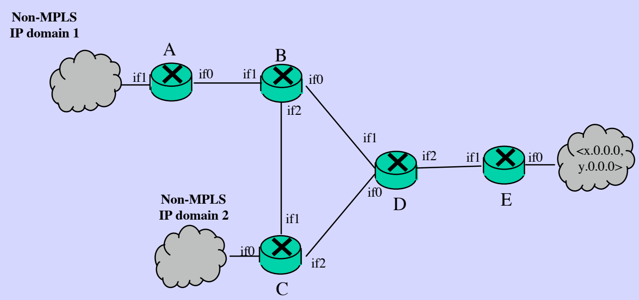
- *For simplicity, we will refer to an LSR or an MPLS node as an LSR*

An MPLS domain



- An MPLS domain is a contiguous set of MPLS nodes which are in the same routing or administrative domain.
- Within an MPLS domain, IP packets are switched using their labels.
- An MPLS domain maybe connected to other MPLS or non-MPLS domains.

An example of label switching



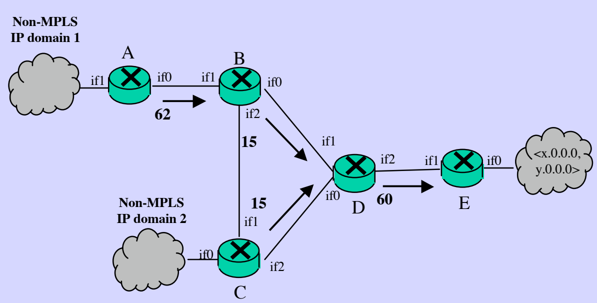
- When an LSR identifies this new FEC with the prefix <x.0.0.0, y.0.0.0>, it selects a label from a pool of free labels and it makes an entry into a table known as the *label forward information base* (LFIB). It also saves the label in its FIB in the entry associated with the FEC.
- The other LSRs also do the same.
- The entry in the LFIB of each LSR are:

LSR	*Incoming label*	*Outgoing label*	*Next hop*	*Outgoing interface*
A			LSR B	if 0
B	62		LSR D	if 0
C			LSR D	if 2
D	15		LSR E	if 2
E	60		LSR E	if 0

- LSRs distribute their local binding information to adjacent LSRs.
- LSR B sends its information to A, D, and C.
- A recognizes that it is upstream from B, and it uses the information to update the entry in its LFIB.
- Likewise for all the other LSRs.
- As a result, each entry in the LFIB of each LSR will be modified as follows:

LFIB	*Incoming label*	*Outgoing label*	*Next hop*	*Outgoing interface*
LSR A		62	LSR B	if 0
LSR B	62	15	LSR D	if 0
LSR C		15	LSR D	if 2
LSR D	15	60	LSR E	if 2
LSR E	60		LSR E	if 0

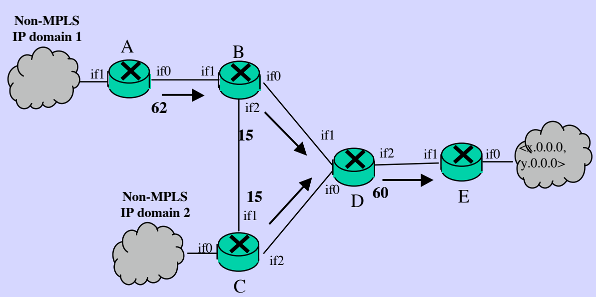
The labels allocated by the LSRs are as follows:



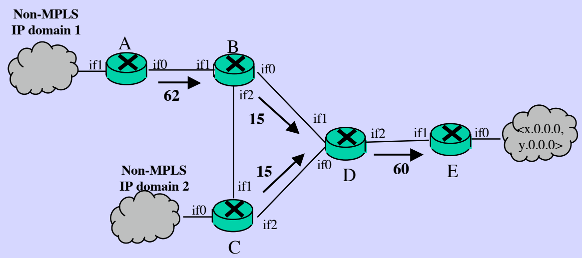
- Now, once the labels have been distributed and the entries have been updated, the forwarding of a packet belonging to this particular FEC is done using solely the labels in the LFIBs.
- Let us assume that A receives a packet with a label 100. A uses this label in its LFIB to locate the new outgoing label and interface. The old label is swapped with the new one and the packet is forwarded to interface 1.

Label switched path (LSP)

The sequence of routers for a particular packet P, i.e., A,B, D, E, or C, D, E is known as an LSP.



LSP ingress LSR. This is LSR A for the path A, B, D, E. It pushes the label 62 into the packet's label stack.
LSP egress LSR: This is LSR E for the same path. It forward the packet either using a lower-level label or the ordinary IP forwarding procedure.



Label allocation

- Two label allocation schemes are used in MPLS:
  - *Unsolicited downstream*
  - *Downstream on demand label distribution*

Unsolicited downstream

- An LSR distributes its binding of a label to a FEC to all its immediate neighbours.
- An *upstream* LSR (*upstream of the link as the traffic flows*) uses this label into its LFIB.
- A non-upstream LSR may either ignore it or store it for future use (for instance, a change in the topology may make it an upstream LSR).

Downstream-on-demand

- An upstream LSR requests from its downstream LSR for a label binding to a FEC.
- In this scheme the LSR that does the binding is the one on the downstream end (as far as data flows) of the link. However, it does not distribute the label until it is explicitly requested by an upstream LSR.

The next hop label forwarding entry (NHLFE)

- This the entry in the LFIB that binds an incoming label to an outgoing label. It provides the following information:
  - The packet's next hop
  - What operation to perform on the packet's label stack

NHLFE Operations:

The following three operations can be performed on the packet's label stack:

- *Replace the label at the top of the label stack with a specified new label.*
- *Pop the label stack.*
- *Replace the label at the top of the label stack with a specified new label, and then push one or more specified new labels onto the label stack.*

Incoming label map (ILM)

- This is used to map an incoming label to a set of NHLFEs.
- One of these entries is used.
- The ILM permits to introduce multi-paths for load balancing, protection, etc

Route selection

- In general, there are two methods for selecting an LSP for a particular FEC
  - *hop-by-hop routing*
  - *explicit routing.*

Hop-by-hop routing

- Each node chooses independently the next-hop for a FEC, as in the existing IP networks.
- The information for the next hop is typically provided by a routing protocol such as OSPF, BGP, etc

Explicit routing

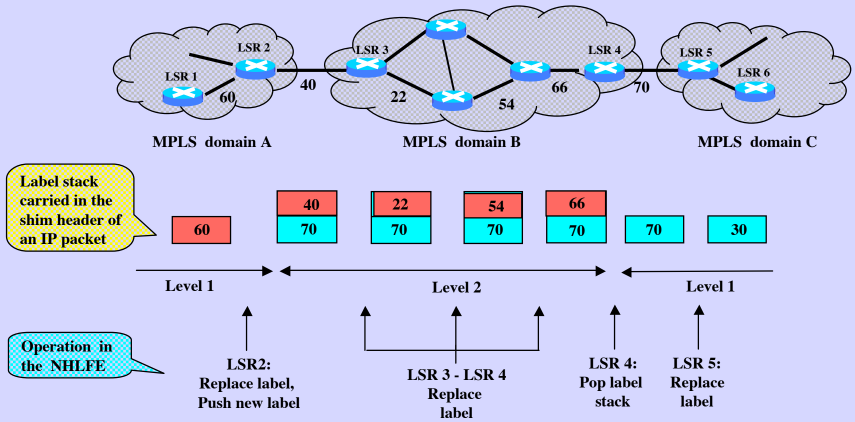
- An explicit route is a pre-defined path through the network, and it is known as *CR-LSP*.
- The route may be different to those advertised by the routing protocols.
- An LSR determines its next hop for the FEC based on the explicit route.

- An explicit route may be:
  - *Strictly explicitly routed*
    - That is, the entire LSP is specified
  - *Loosely explicitly routed*
    - Only part of the LSP is specified
- Explicit routing is used for a number of purposes such as *policy-based routing* or *traffic engineering*.

The label stack

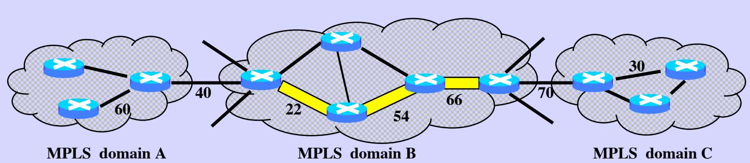
- MPLS allows a packet to carry a set of labels organized as a *stack*.
- When the packet is forwarded within a domain, it contains two labels. The label at the top of the stack is used for label switching within the interior LSRs. The label in the next level is used by the egress LSP LSR to forward the packet to the next ingress LSP LSR.

An example of label stack



Uses of the label stack: LSP tunnels

The labels in MPLS B domain form a tunnel. At the end of the tunnel, the LSR may not know where to forward the packet. This can be easily resolved using a label stack.



Schemes for setting-up an LSP

- We have assumed that FECs correspond to address prefixes, which are distributed via a dynamic routing algorithm.
- The setup of an LSP for these FECs can be done in one of the two following ways;
  - *Independent LSP control*
  - *Ordered LSP control*

Independent LSP control

- Each LSR when it recognizes a new FEC binds a label to it, and advertises it to its neighbours (as in the example given earlier on).
- This is similar to the way that conventional IP datagram routing works.

Ordered LSP control

- The allocation of labels proceeds backwards starting from the *egress LSP LSR*. That is:
  - An LSR only binds a label to a FEC if it is the egress LSR for that FEC, or
  - If it has already received a label binding for that FEC from its next hop LSR for that FEC.
- Ordered LSP control may be initiated by the ingress LSR as well.

Aggregation

- For each FEC an LSP is set up.
- Within an MPLS domain, it is possible that IP packets belonging to two or more FECs follow the same route. This can happen if these FECs have the same egress node.
- Within the same MPLS domain, the union of these FECs is also a FEC.

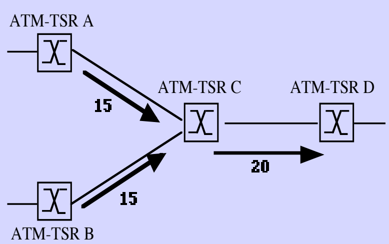
- Given a set of FECs that can be aggregated, it is possible to
  - Aggregate them to a single FEC
  - Aggregate them to several FECs
  - Not aggregate them.
- In ordered control an LSR can choose any of the above solutions.
- In independent control, the situation is more complex

MPLS over ATM

- An ATM switch that supports MPLS is referred to as the *ATM-LSR*.
- The user plane of the ATM switch remains intact, however the control plane is replaced by protocols such as OSPF, BGP, PIM and RSVP.
- The downstream label allocation on demand scheme is used.

- SVC encoding:
  - The label information is carried in the VPI/VCI field.
  - Each LSP is nothing else but an ATM SVC, and the label distribution protocol replaces the ATM signalling protocol.
  - No label stacks
- SVP encoding
  - The VPI field is used to encode the label, and the VCI field is used to encode the second label on the stack.

VC merging in ATM



ATM-LSR D will not be able to identify distinct AAL 5 frames, since cells from the upstream nodes will be interleaved!