

IP Switching and Multiprotocol Label Switching

Raj Jain

Professor of Computer and Information Sciences

**Raj Jain is now at
Washington University in Saint Louis
Jain@cse.wustl.edu
<http://www.cse.wustl.edu/~jain/>**



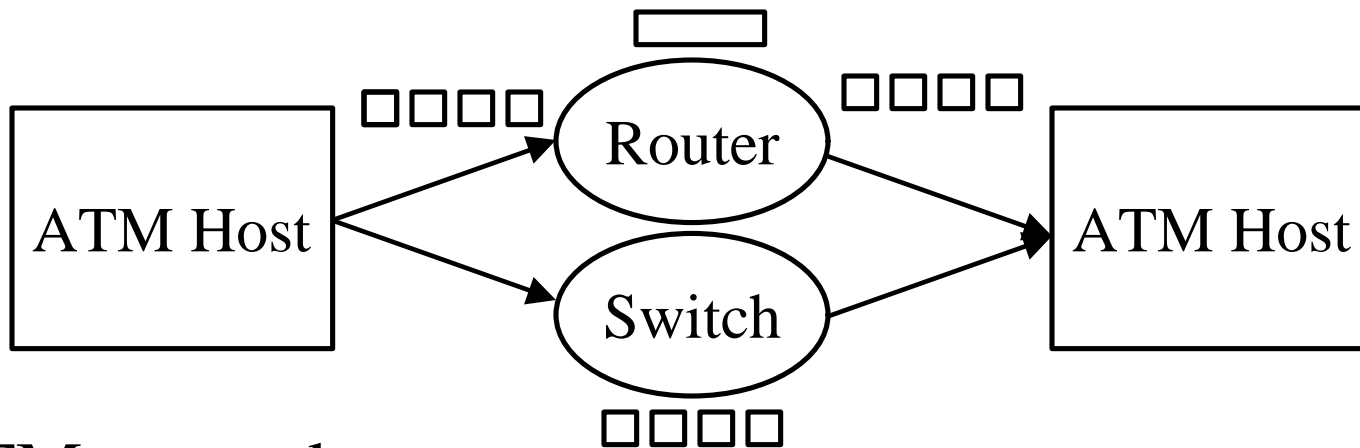
- ❑ IP Switching
- ❑ MPLS Overview
- ❑ Label Format
- ❑ Label Stacks
- ❑ Label Distribution Protocols

Routing vs Switching



- ❑ Routing: Based on address lookup. Max prefix match.
 - ⇒ Search Operation
 - ⇒ Complexity $\approx O(\log_2 n)$
- ❑ Switching: Based on circuit numbers
 - ⇒ Indexing operation
 - ⇒ Complexity $O(1)$
 - ⇒ Fast and Scalable for large networks and large address spaces
- ❑ These distinctions apply on all datalinks: ATM, Ethernet, SONET

Routing vs Switching over ATM

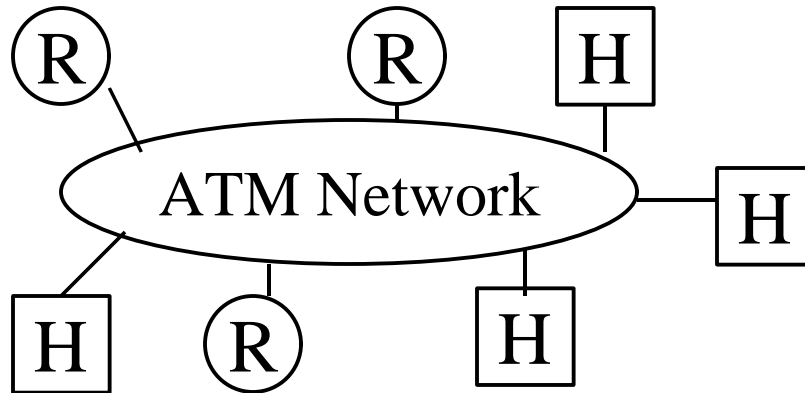


On ATM networks:

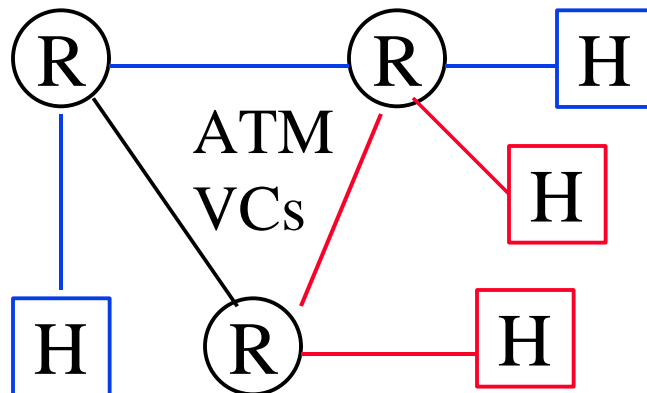
- ❑ IP routers use IP addresses
 - ⇒ Reassemble IP datagrams from cells
- ❑ IP Switches use ATM Virtual circuit numbers
 - ⇒ Switch cells
 - ⇒ Do not need to reassemble IP datagrams
 - ⇒ Fast

IP Switching

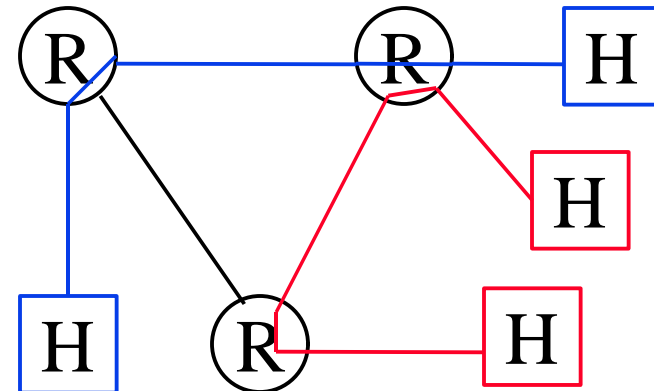
1. Original ATM Network



2. VCs at every hop



3. Short-circuit VCs



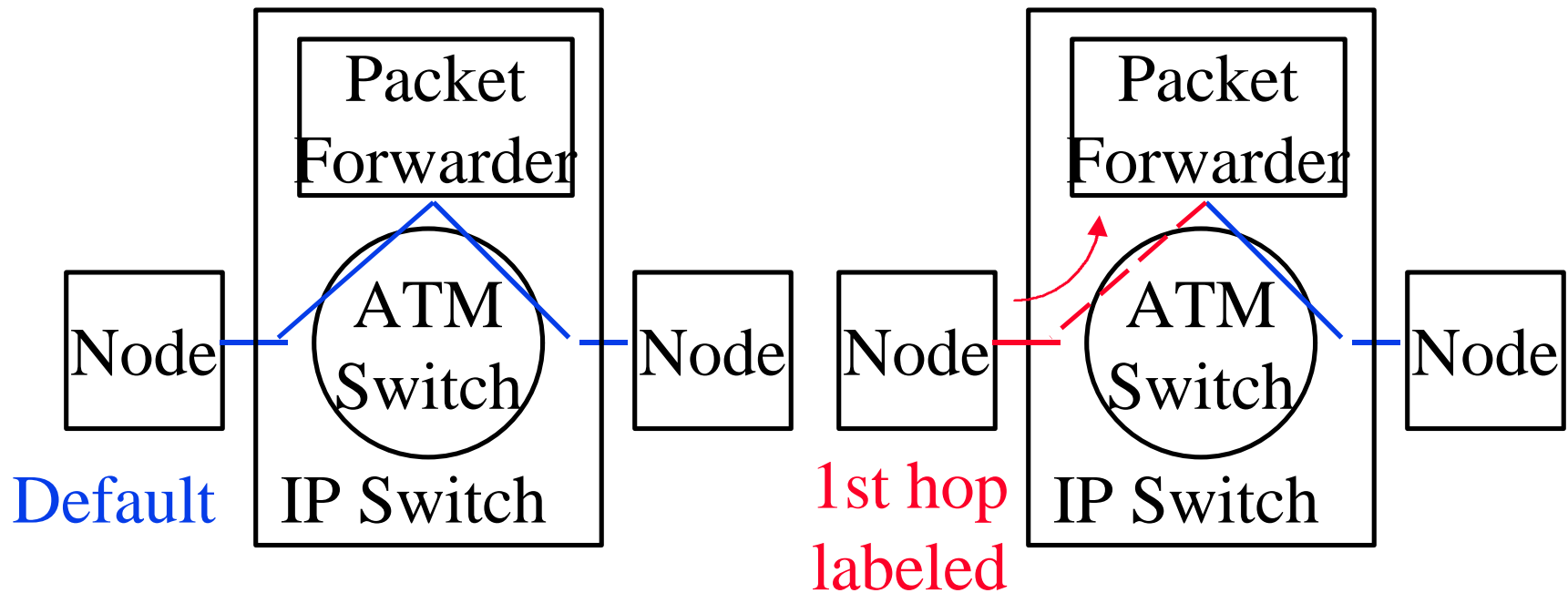
IP Switching

- ❑ Each ATM switch also has routing s/w
- ❑ Normally the packets are reassembled and forwarded in the router. Segmentation and reassembly in the forwarder.
- ❑ If a flow is deemed to be "flow oriented", previous node is told to set up a new VC. Forwarder uses cached info.
- ❑ Downstream nodes may also ask for a new VC. The switch then makes a mapping for cut-through
- ❑ Flow-oriented traffic: FTP, Telnet, HTTP, Multimedia

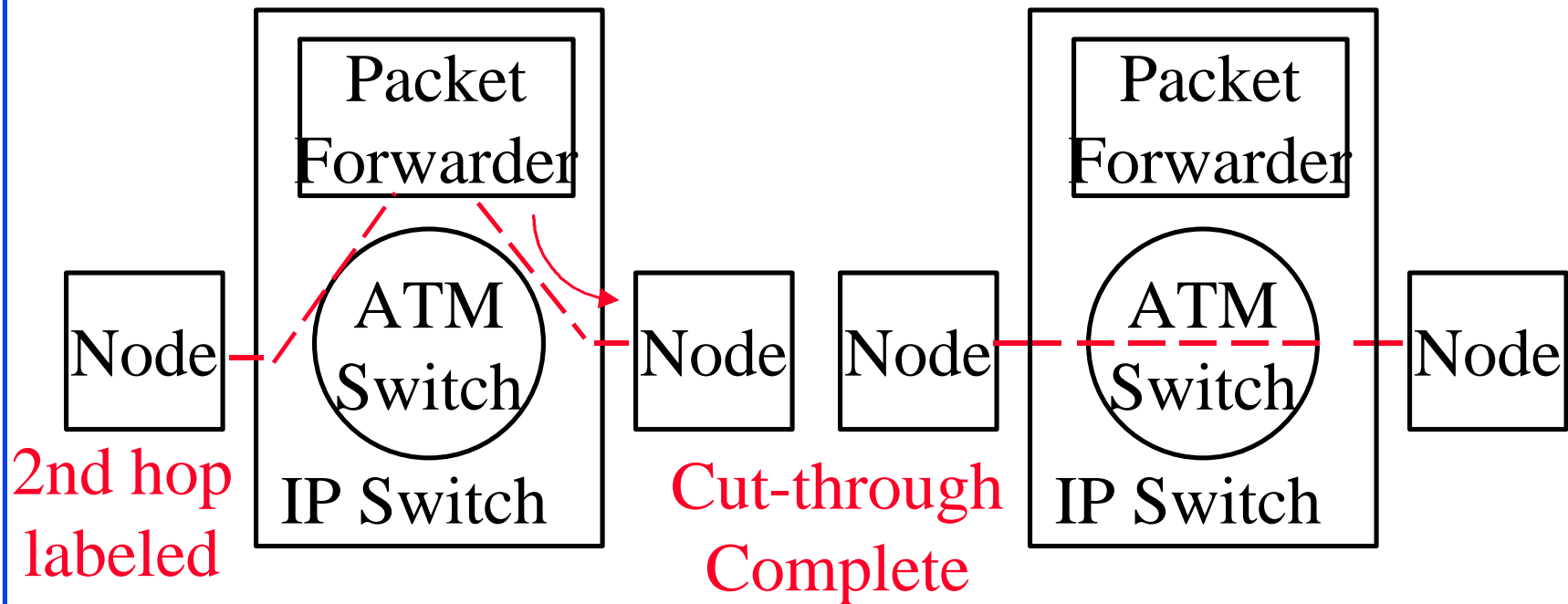
IP Switching (Cont)

- ❑ Short-lived Traffic: DNS query, SMTP, NTP, SNMP, request-response
- ❑ Ipsilon claims that 80% of packets and 90% of bytes are flow-oriented.
- ❑ Ipsilon Flow Management Protocol (IFMP)
- ❑ IP switching implemented as a s/w layer over an ATM switch
- ❑ Ipsilon claims their Generic Switch Management Protocol (GSMP) to be 2000 lines, and Ipsilon Flow Management Protocol (IFMP) to be only 10,000 lines of code

IP Switching: Steps 1-2



IP Switching: Steps 3, 4



Ipsilon's IP Switching: Issues

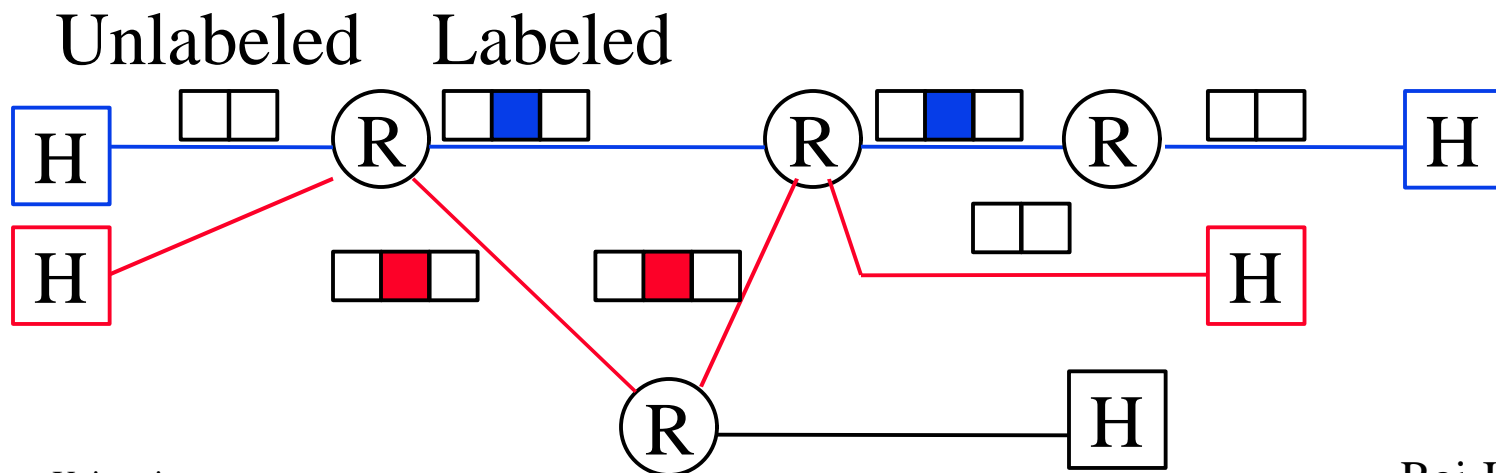
- ❑ VCI field is used as ID.
VPI/VCI change at switch
 - ⇒ Must run on **every** ATM switch
 - ⇒ non-IP switches not allowed between IP switches
 - ⇒ Subnets limited to one switch
- ❑ Cannot support VLANs
- ❑ Scalability: Number of VC \geq Number of flows.
 - ⇒ **VC Explosion** (1000 setups/sec.)
- ❑ Quality of service determined implicitly by the flow class or by RSVP
- ❑ ATM only

Other Competing Approaches

- ❑ Cisco: Tag Switching
 - ❑ IBM: Aggregate Route Based IP Switching (ARIS)
 - ❑ Toshiba: Cell-switched router
 - ❑ Cabletron: Secure Fast Virtual Network
 - ❑ 3Com: Fast IP
 - ❑ Cascade: IP Navigator
 - ❑ Bay Networks: Switch Node (packet-by-packet)
- ⇒ IETF: Multiprotocol label switching

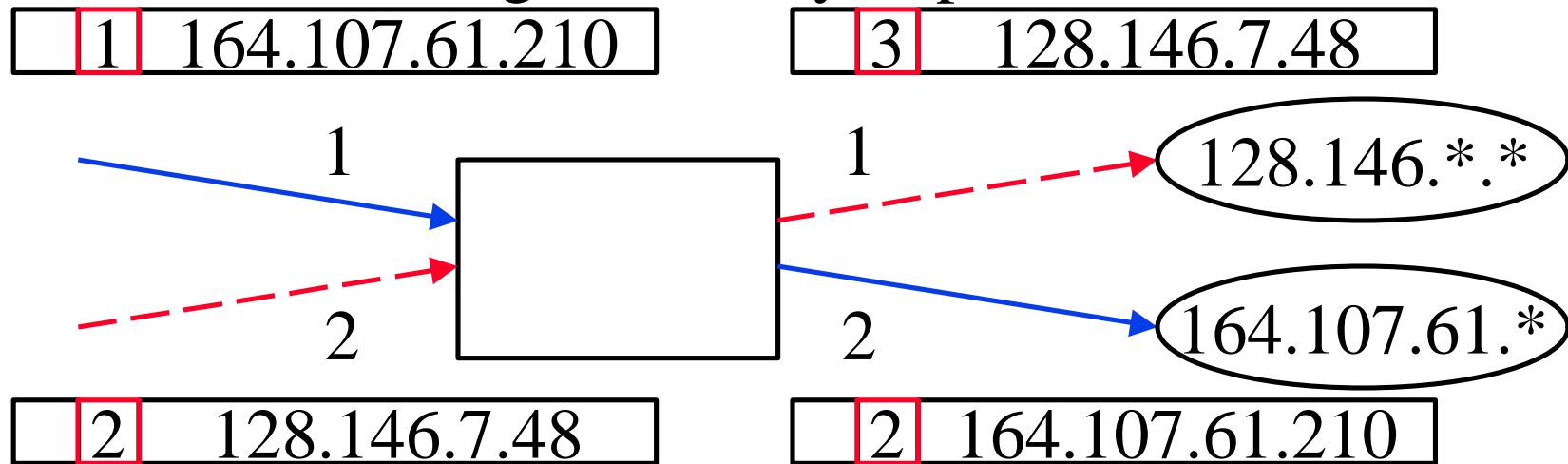
Label Switching

- ❑ Label = Circuit number = VC Id
- ❑ Ingress router/host puts a label. Exit router strips it off.
- ❑ Switches switch packets based on labels. Do not need to look inside \Rightarrow Fast.



Label Switching (Cont)

- Labels have local significance
- Labels are changed at every hop



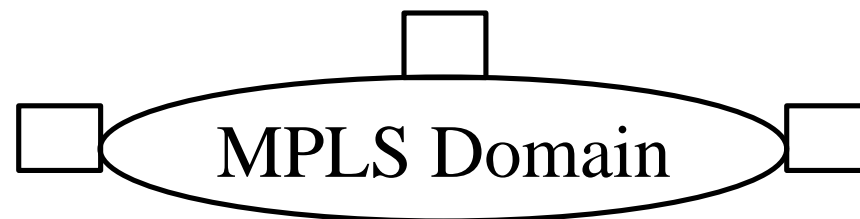
Input Port	Input Label	Adr Prefix	Output Port	Output Label
1	1	164.107.61.*	2	2
2	2	128.146.*.*	1	3

MPLS

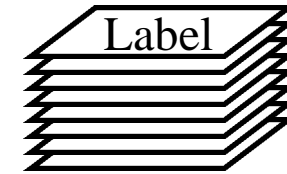
- ❑ Multiprotocol Label Switching
- ❑ IETF working group to develop switched IP forwarding
- ❑ Initially focused on IPv4 and IPv6.
Technology extendible to other L3 protocols.
- ❑ Not specific to ATM. ATM or LANs.
- ❑ Not specific to a routing protocol (OSPF, RIP, ...)

MPLS Terminology

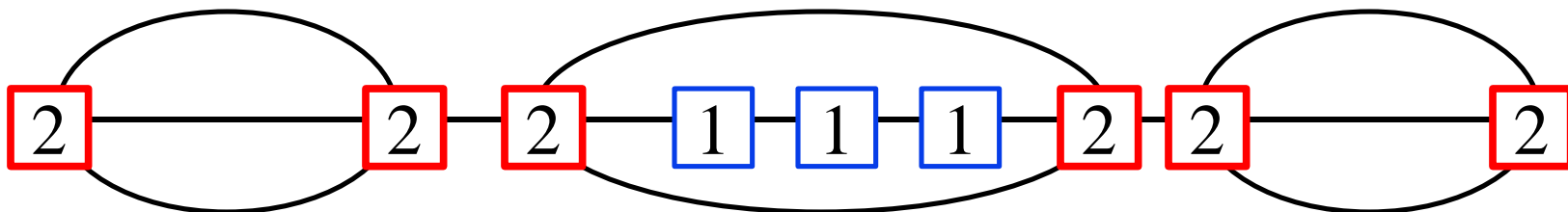
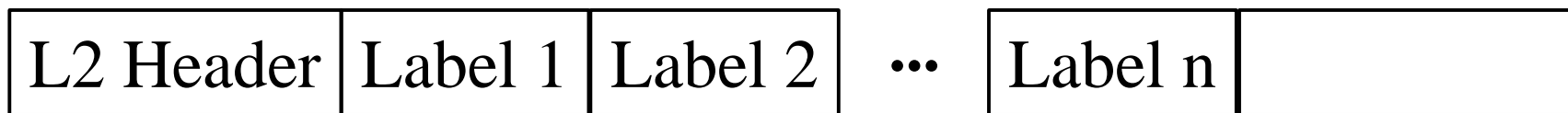
- ❑ Label = Short fixed length,
physically contiguous, locally significant
- ❑ Label Switching Router (LSR): Routers that use labels
- ❑ Forwarding Equivalence Class (FEC):
Same Path + treatment \Rightarrow Same Label
- ❑ MPLS Domain: Contiguous set of MPLS nodes in one
Administrative domain
- ❑ MPLS edge node = Egress or ingress node
- ❑ Label distribution protocol \cong Routing protocols



Label Stacks

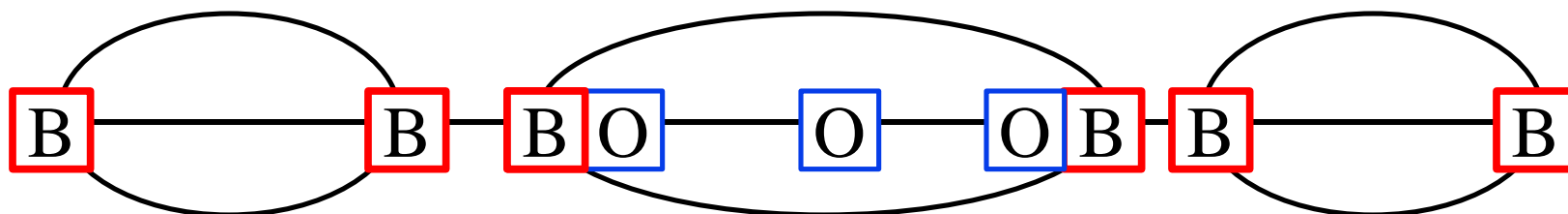


- ❑ A MPLS packet may have multiple labels
- ❑ Labels are pushed/popped as they enter/leave MPLS domain
- ❑ Stack allows hierarchy of MPLS domains
- ❑ Bottom label may indicate protocol (0=IPv4, 2=IPv6)

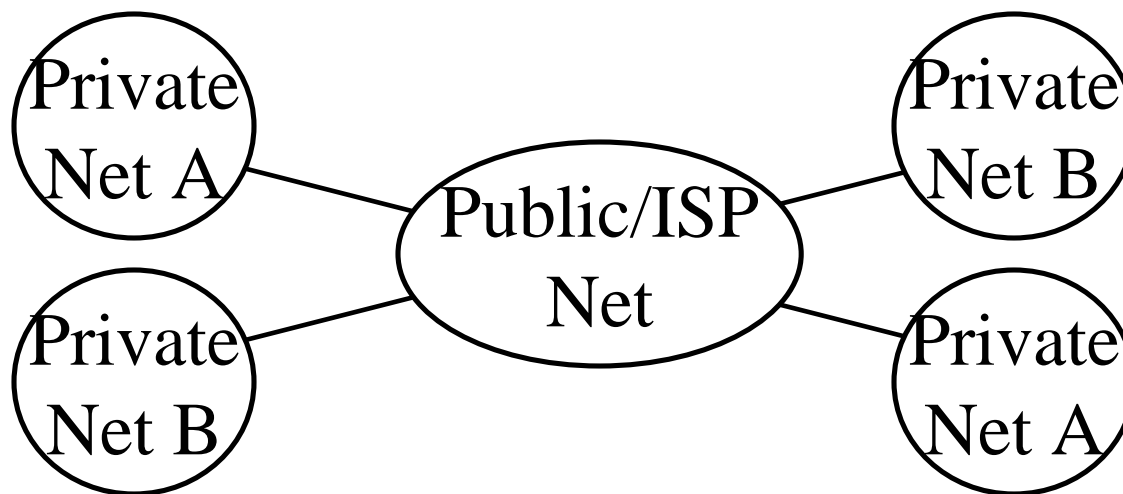


Label Stack Examples

1. BGP/OSPF Routing Hierarchy

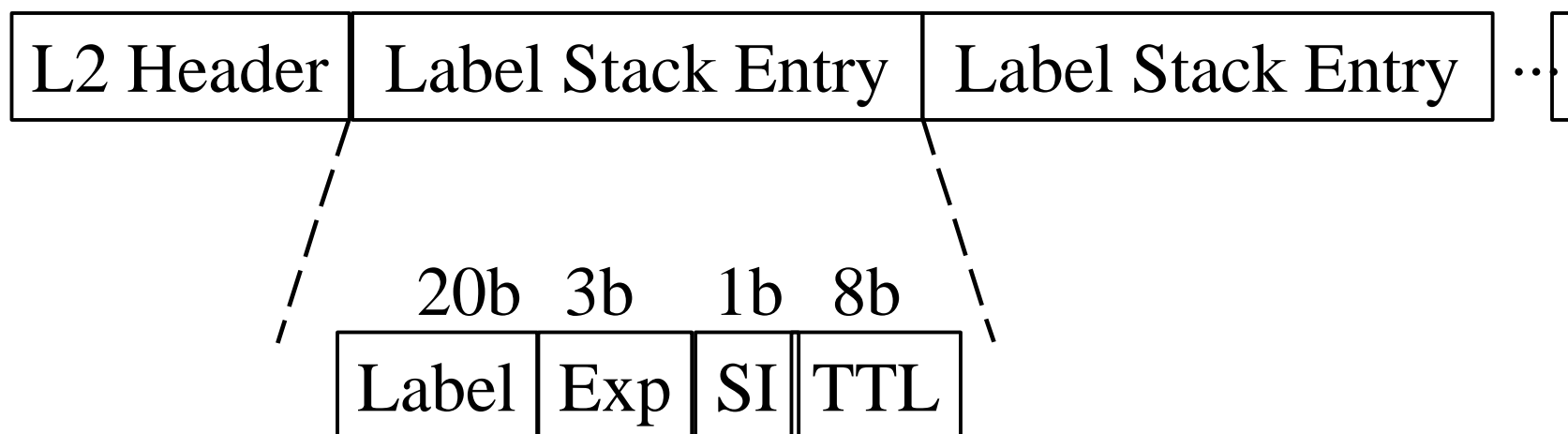


2. VPN: Top label used in public network.
Net A and B can use the same private addresses.



Label Stack Entry Format

- ❑ Labels = Explicit or implicit L2 header
- ❑ TTL = Time to live
- ❑ Exp = Experimental
- ❑ SI = Stack indicator, 1 \Rightarrow Bottom of Stack



Label Assignment

- ❑ Unsolicited: Topology driven \Rightarrow Routing protocols exchange labels with routing information.
Many existing routing protocols are being extended:
BGP, OSPF
- ❑ On-Demand:
 \Rightarrow Label assigned when requested,
e.g., when a packet arrives \Rightarrow latency
- ❑ A new Label Distribution Protocol called LDP is being defined.
- ❑ RSVP is being extended to allow label request and response

Label Distribution Protocol

- ❑ LDP peers: LSRs that exchange LDP messages. Using an LDP session.
- ❑ LDP messages:
 - Session establishment/termination messages
 - Discovery messages to announce LSRs (Hello)
 - Advertisement msgs to create/delete/change label
 - Notification messages for errors and advice
- ❑ Discovery messages are UDP based. All others TCP.
- ❑ Hello messages are sent on UDP port 646.
- ❑ Session establishment messages sent on TCP port 646.
- ❑ No multicast, multipath, or QoS in the first version.

LDP Messages

- Hello
- Initialization
- Label Request
- Label Mapping (Label Response)
- Label Withdraw (No longer recognized by downstream)
- Label Release (No longer needed by upstream)
- Label Abort Request
- KeepAlive
- Notification
- Address (advertise interface addresses)
- Address Withdraw
- Vendor-Private
- Experimental

LDP TLVs

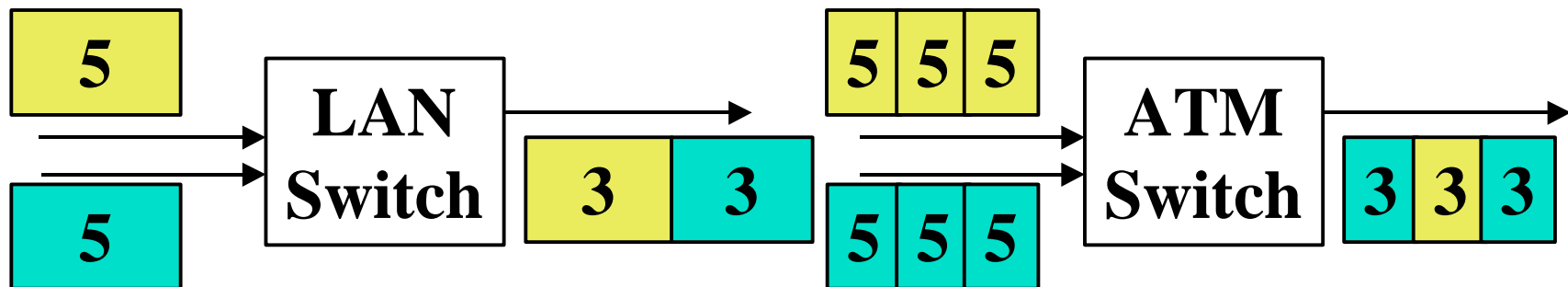
- ❑ FEC (Wild card, prefix, or host address)
- ❑ Address List
- ❑ Hop Count
- ❑ Path Vector
- ❑ Generic Label
- ❑ ATM Label
- ❑ Frame Relay Label
- ❑ Status
- ❑ Extended Status
- ❑ Returned PDU
- ❑ Returned Message
- ❑ Common Hello parameters

MPLS Over ATM

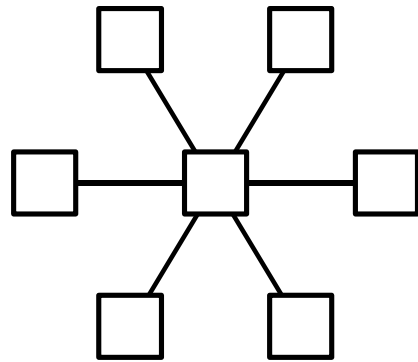
- ❑ With MPLS software, ATM switches can act as LSRs.
- ❑ VPI/VCI fields are used for labels.
- ❑ No Stack bit \Rightarrow Maximum two possible levels of hierarchy: VCI, VPI
All ATM switches should use the same encoding.
- ❑ No TTL field \Rightarrow Hops between ingress and egress can be computed during LSP setup.
Ingress router drops if $TTL < \text{hops to egress}$
- ❑ ATM LSRs need to participate in network layer routing protocols (OSPF, BGP)
- ❑ VPI/VCI space may be segmented for label switching and normal ATM switching

Stream Merging

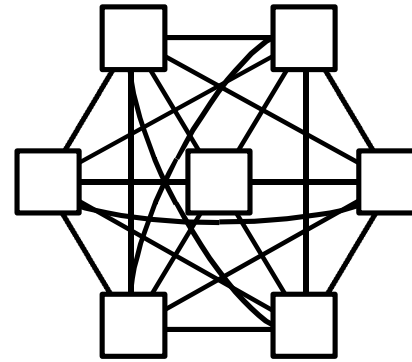
- ❑ Required for egress based labels. Helpful for mpt-to-pt streams.
- ❑ In ATM/AAL5, cells of frames on the same VC cannot be intermingled \Rightarrow VCs cannot be merged.
- ❑ VC-merge: Store all cells of a frame and forward together \Rightarrow Need more buffering. Delay.
- ❑ VP Merge: VPI = Labels, VCI = source



High-Speed Backbone Alternatives



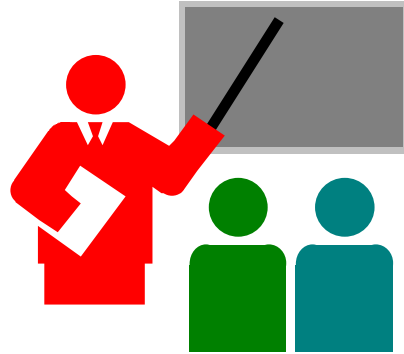
Physical Topology



Logical Topology

- ❑ High-speed (OC-3 and higher) ATM switches easily available. IP routers either not available or expensive.
- ❑ IP has no traffic engineering \Rightarrow Under/over-utilized links
- ❑ Logical \neq Physical \Rightarrow ATM has n^2 scaling problem
- ❑ MPLS takes the best of both IP and ATM networks
- ❑ Works on both ATM and non-ATM networks
 \Rightarrow Easier management

Summary



- ❑ IP Switching allows hop-by-hop switching of IP packets.
- ❑ MPLS combines the best of ATM and IP. Works on all media: ATM and non-ATM.
- ❑ Label is similar to circuit number or VC Id.
- ❑ Common routing protocols and RSVP are being extended to include label exchange. LDP is being defined.

Homework

- ❑ Read Section 20.6 of McDysan and Spohn
- ❑ IP Switching, http://www.cis.ohio-state.edu/~jain/cis788-97/ip_switching/index.htm