

# **IP Switching: Issues and Alternatives**

Raj Jain

Professor of Computer and Information Sciences  
The Ohio State University

<http://www.cis.ohio-state.edu/~jain/>



- ❑ LANE, IPOA, NHRP, MPOA
- ❑ IP Switch
- ❑ Cell Switched Router
- ❑ Tag Switching (CISCO)
- ❑ ARIS (IBM)
- ❑ Multi-protocol label switching

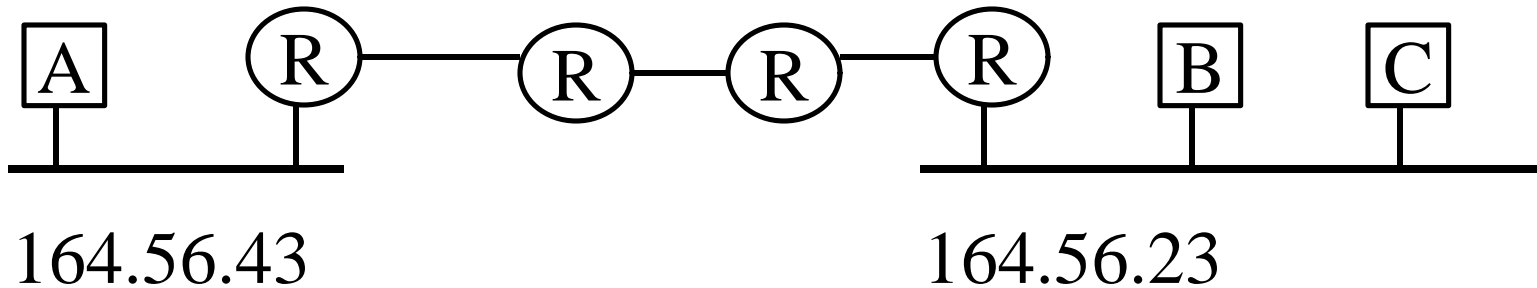
# L3 Switching

- Layer 3 forwarding at wire speeds
  - Switching based on Layer 3 (L3) header
  - Switched IP forwarding
  - Several million packets per second (Mpps)
  - 8 Mpps announced by ODS
- Layer 2 switching  $\Rightarrow$  Large flat networks
  - Problem: Broadcast, security
  - Solution: Virtual LANs (VLANs)
    - $\Rightarrow$  Need routing between VLANs

# IP Forwarding: Fundamentals

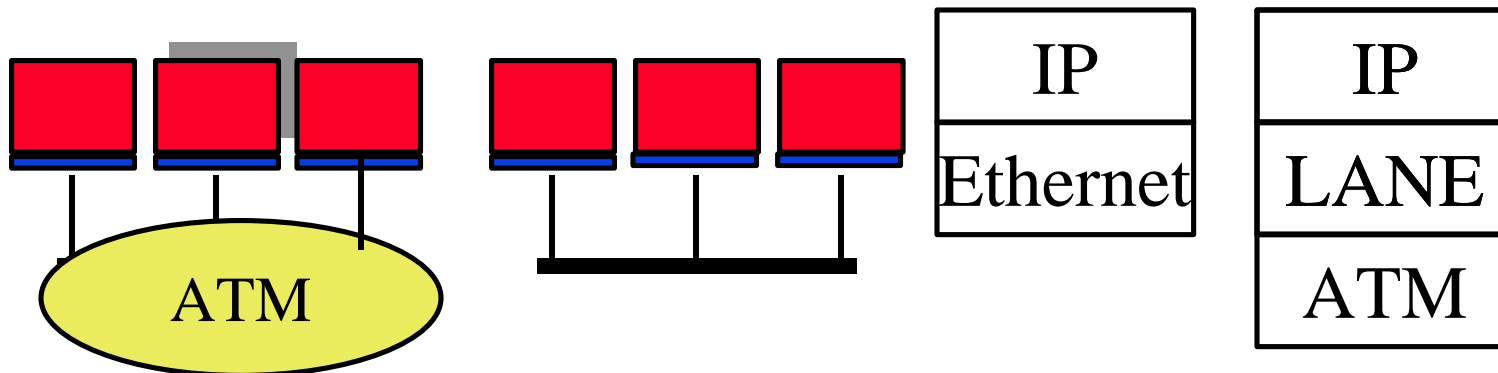
To: 164.56.23.34

From: 164.56.43.96



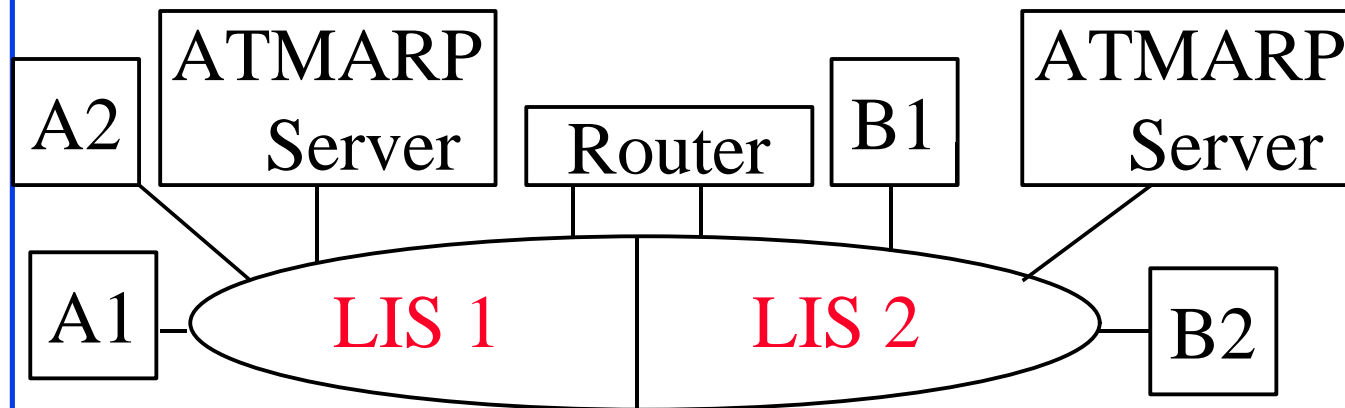
- ❑ IP routers forward the packets towards the destination subnet
- ❑ On the same subnet, routers are not required.
- ❑ IP Addresses: 164.56.23.34  
Ethernet Addresses: AA-23-56-34-C4-56  
ATM : 47.0000 1 614 999 2345.00.00.AA....

# LAN Emulation



- ❑ LAN Emulation driver replaces Ethernet driver and passes the networking layer packets to ATM driver.
- ❑ Each ATM host is assigned an Ethernet address.
- ❑ LAN Emulation Server translates Ethernet addresses to ATM addresses
- ❑ Hosts set up a VC and exchange packets
- ❑ All software that runs on Ethernet can run on LANE

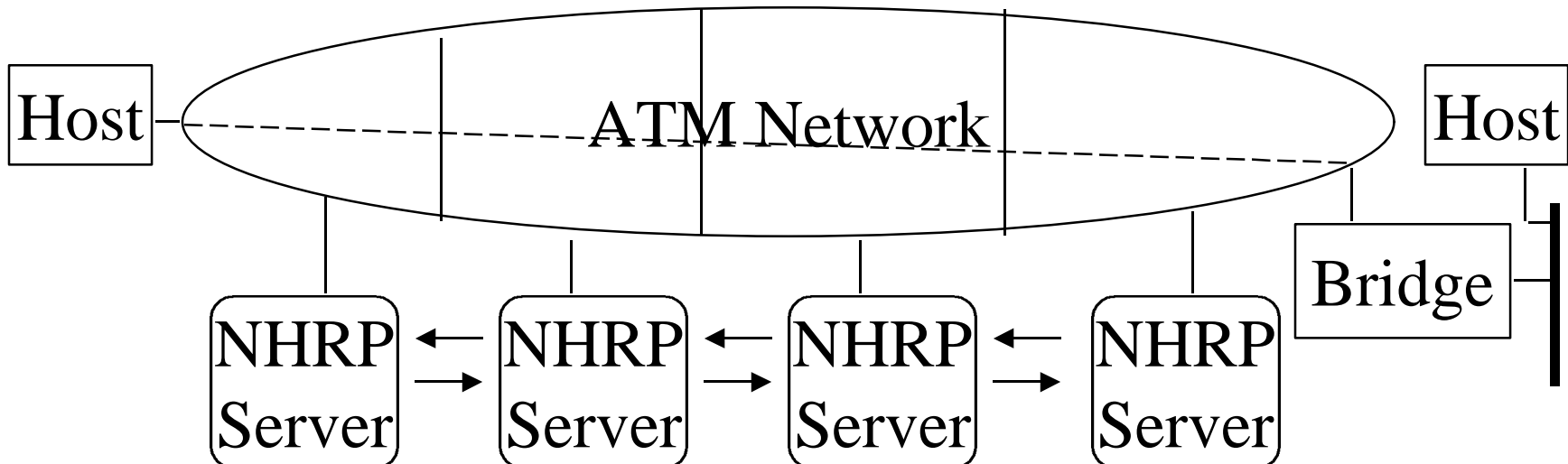
# Classical IP Over ATM



- ❑ ATM stations are divided into Logical IP Subnets (LIS)
- ❑ ATMARP server translates IP addresses to ATM addresses.
- ❑ Each LIS has an ATMARP server for resolution
- ❑ IP stations set up a direct VC with the destination or the router and exchange packets.

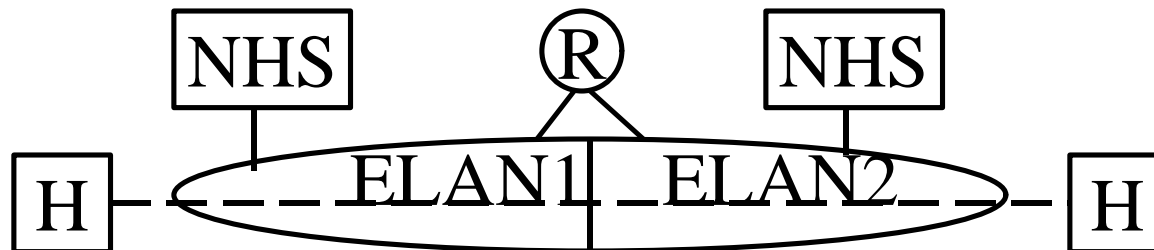
# Next Hop Resolution Protocol

- ❑ Routers assemble packets  $\Rightarrow$  Slow
- ❑ NHRP servers can provide ATM address for the edge device to any IP host
- ❑ Can avoid routers if both source and destination are on the same ATM network.



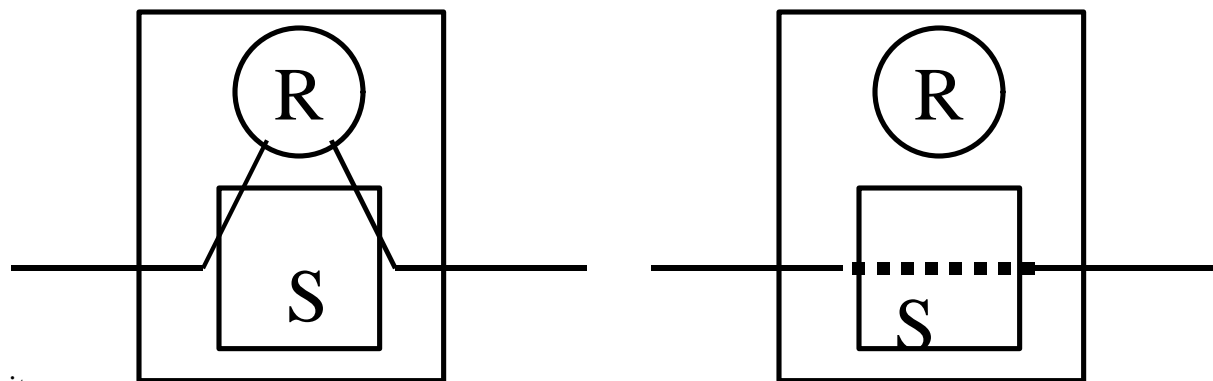
# Multiprotocol Over ATM

- ❑ MPOA= LANE + “NHRP+”
- ❑ Extension of LANE
- ❑ Uses NHRP to find the shortcut to the next hop
- ❑ No routing (reassembly) in the ATM network



# IP Switching

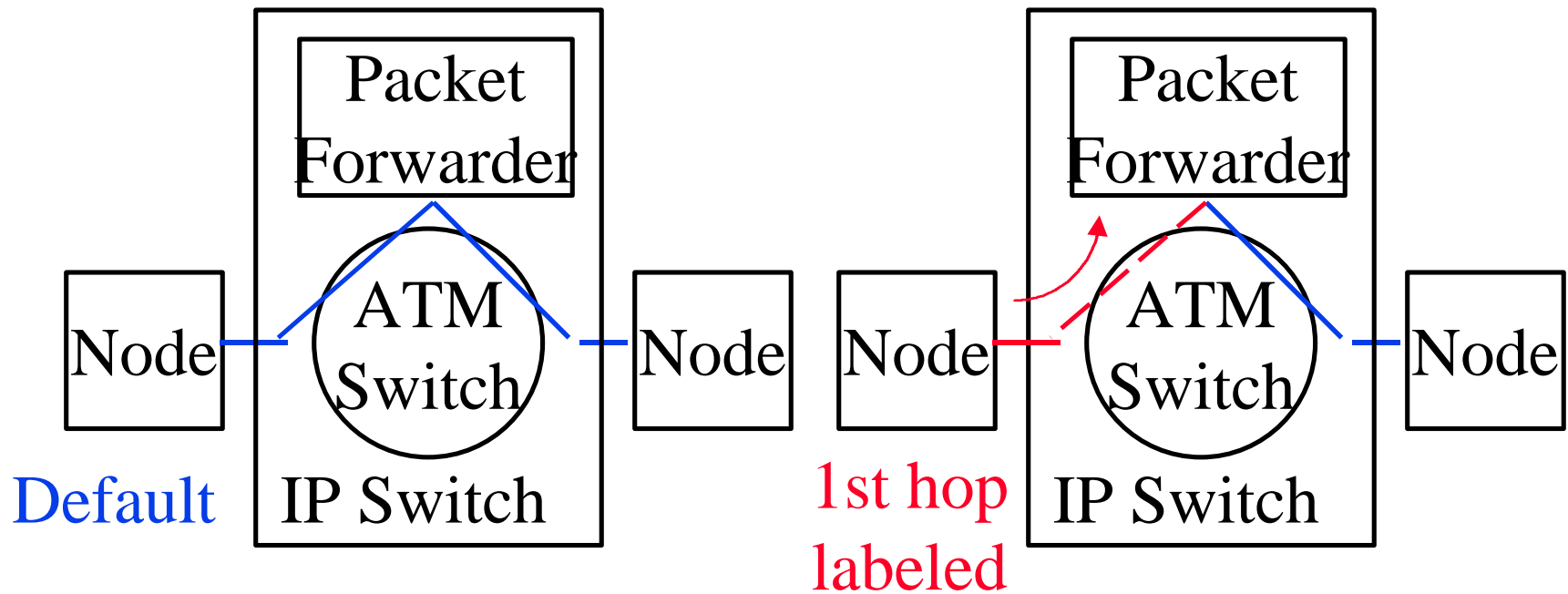
- ❑ Developed by Ipsilon
- ❑ Routing software in every ATM switch in the network
- ❑ Initially, packets are reassembled by the routing software and forwarded to the next hop
- ❑ Long term flows are transferred to separate VCs. Mapping of VCIs in the switch  $\Rightarrow$  No reassembly



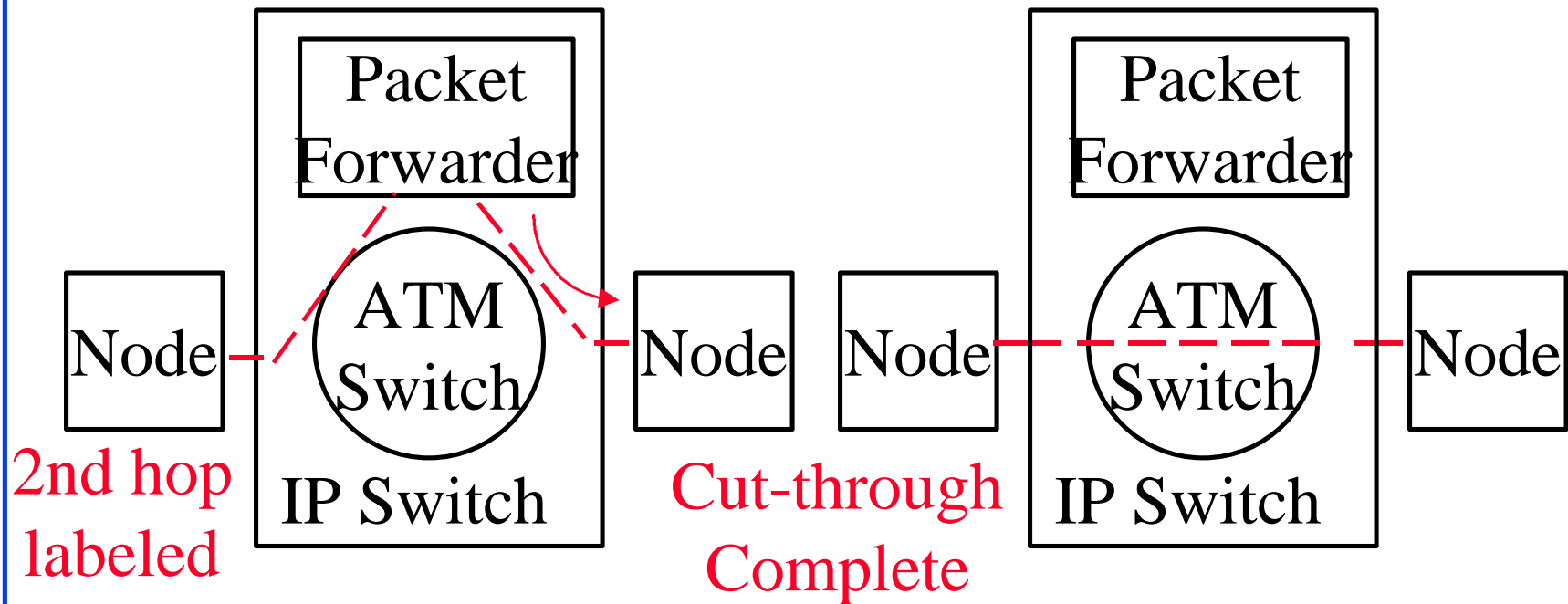
# IP Switching

- ❑ If a flow is deemed to be "flow oriented", the node asks the upstream node to set up a separate VC.
- ❑ Downstream nodes may also ask for a new VC.
- ❑ After both sides of a flow have separate VCs, the router tells the switch to register the mapping for cut-through

# IP Switching: Steps 1-2



# IP Switching: Steps 3, 4



# IP Switching (Cont)

- ❑ Flow-oriented traffic: FTP, Telnet, HTTP, Multimedia
- ❑ Short-lived Traffic: DNS query, SMTP, NTP, SNMP, request-response  
Ipsilon claims that 80% of packets and 90% of bytes are flow-oriented.
- ❑ 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

# Ipsilon's IP Switching: Features

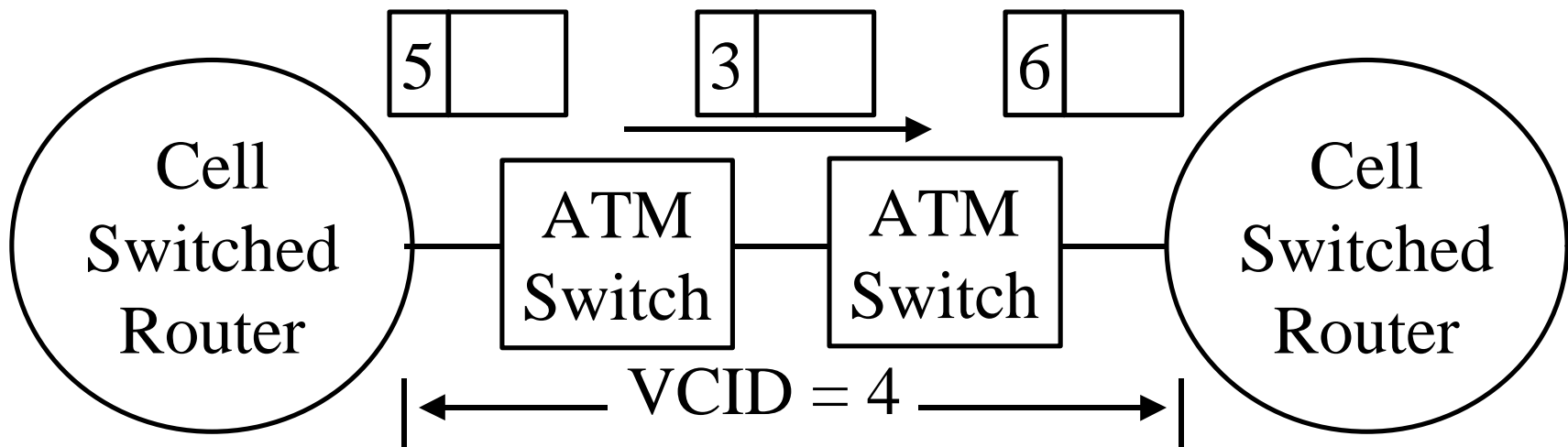
- ❑ Runs as added software on an ATM switch
- ❑ Implemented by several vendors
- ❑ Multicast flows  $\Rightarrow$  pt-mpt VC per source
- ❑ Routing bypassed  $\Rightarrow$  Firewall bypassed
  - Solution: IP fields are deleted before segmentation and added after assembly  $\Rightarrow$  First packet has to go through firewall.
- ❑ Initially IP only. IPX supported via tunneling in IP.

# 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

# Cell Switched Router (CSR)

- ❑ Proposed by Toshiba
- ❑ Flow driven (similar to Ipsilon)
- ❑ VCID separate from VCI  $\Rightarrow$  Switches between CSRs
- ❑ Upstream assigns a VCID and sends downstream



# CSR (Cont)

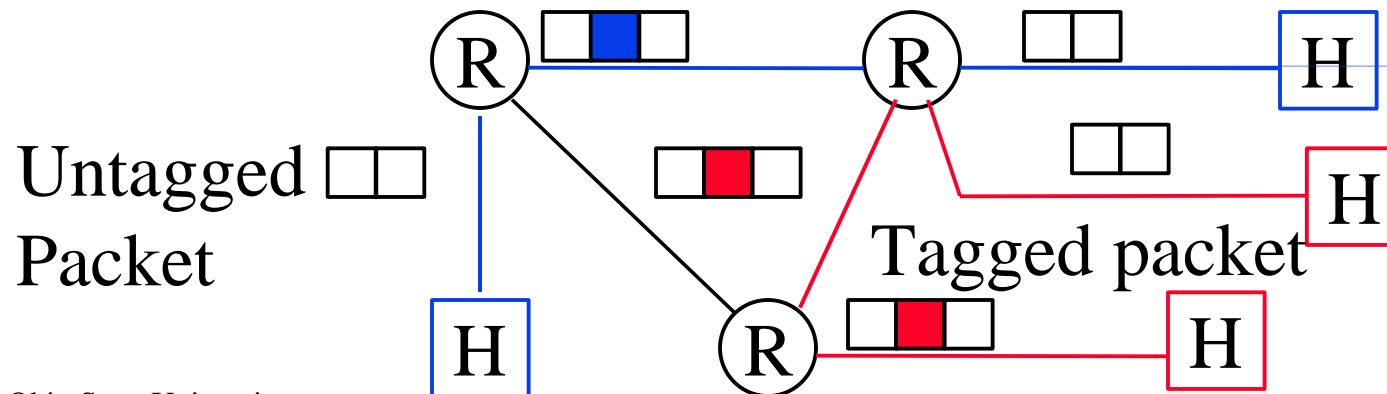
- ❑ VCs are set up in advance and are bounded as needed
- ❑ Classifies flows by IP source/destination address pair
- ❑ Soft connections  $\Rightarrow$  Periodically refreshed

# Tag Switching

- ❑ Proposed by CISCO
- ❑ Similar to VLAN tags
- ❑ Tags can be explicit or implicit L2 header



- ❑ Ingress router/host puts a tag. Exit router strips it off.

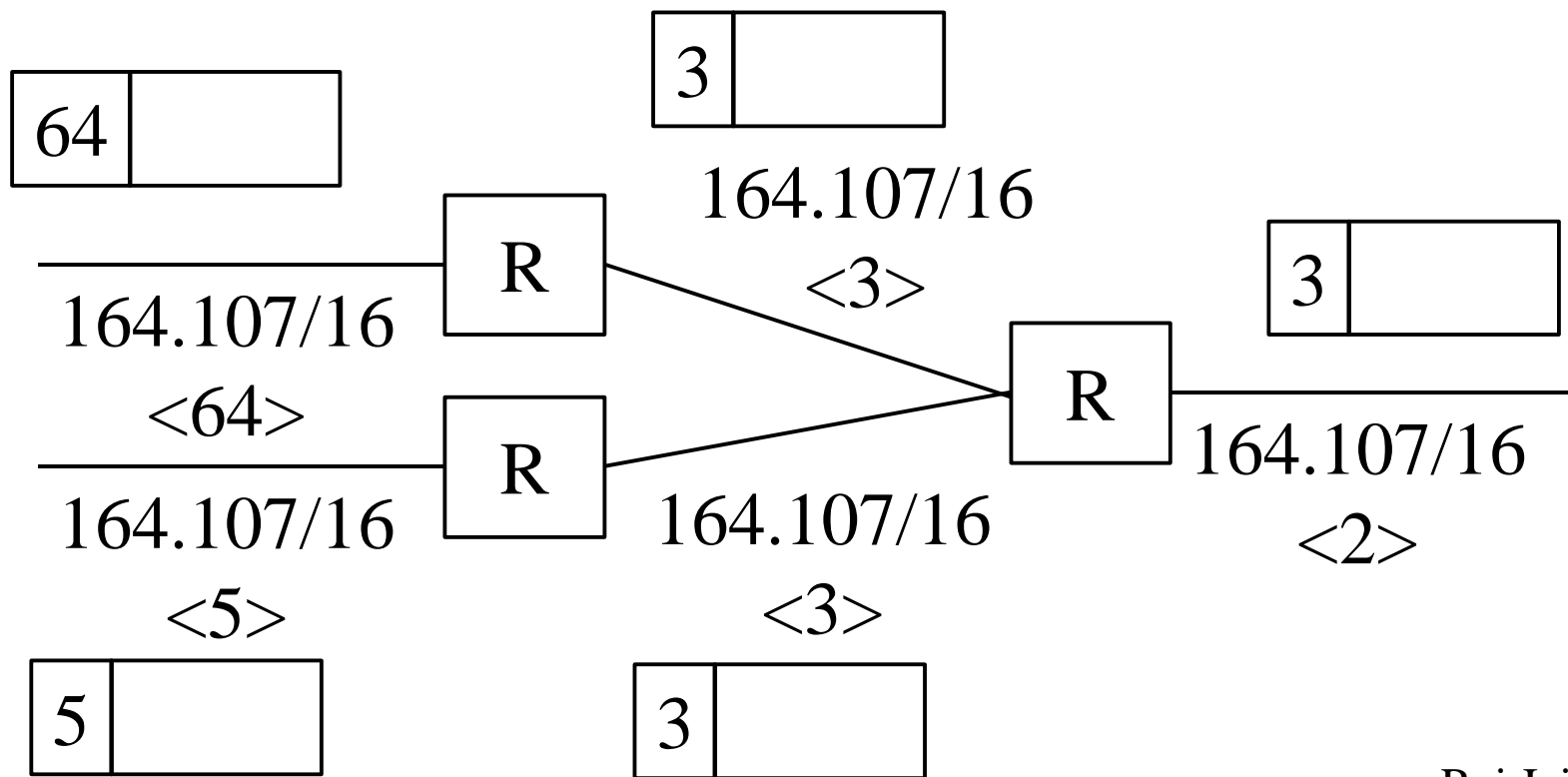


# Tag Switching (Cont)

- ❑ Switches switch packets based on labels.  
Do not need to look inside  $\Rightarrow$  Fast.
- ❑ One memory reference compared to 4-16  
in router
- ❑ Tags have local significance  
 $\Rightarrow$  Different tag at each hop (similar to VC #)

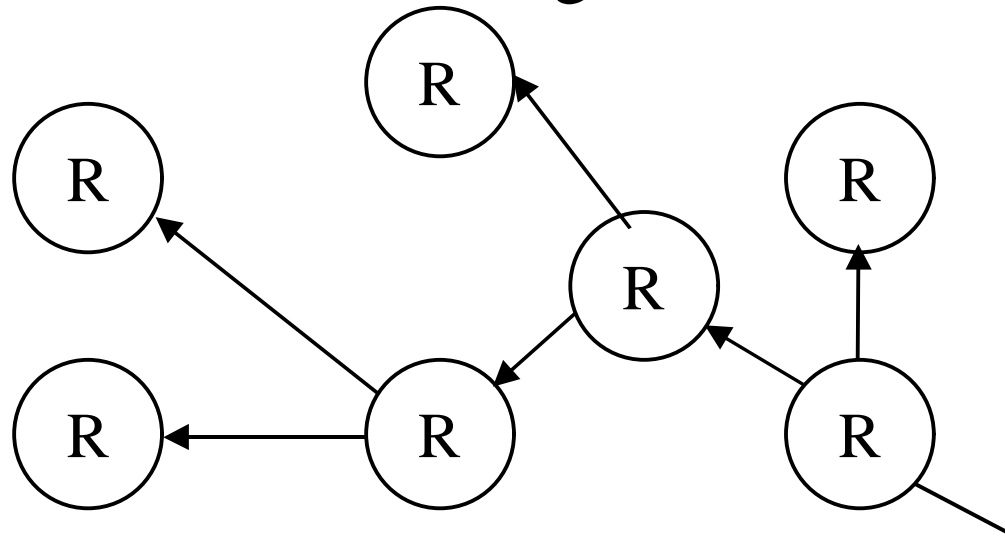
# Tag Switching (Cont)

- One VC per routing table entry



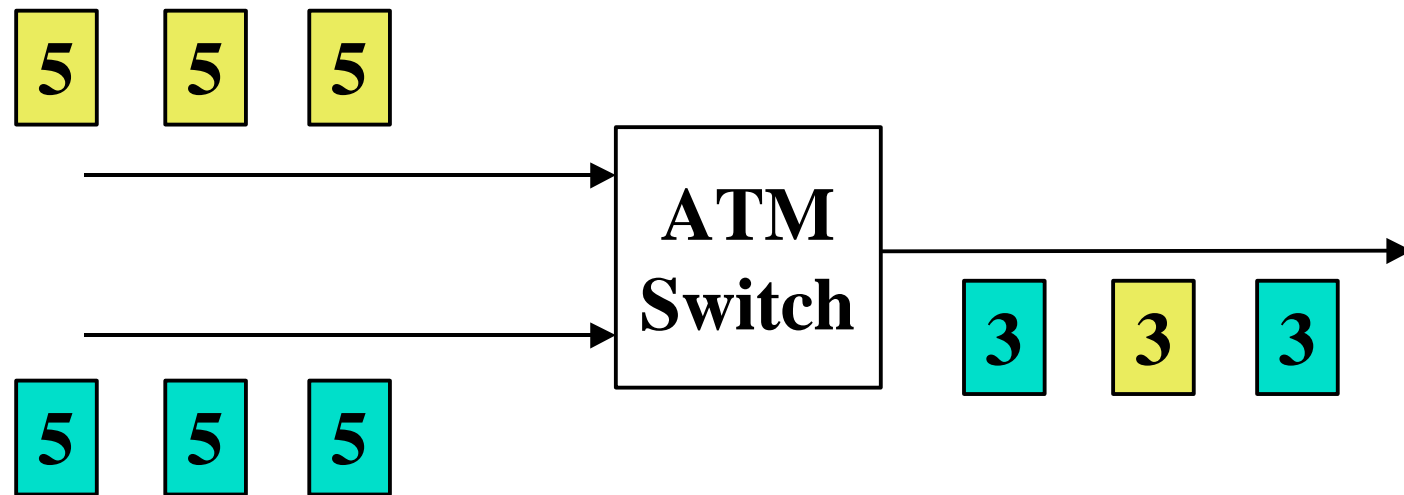
# ARIS

- ❑ Aggregate Route-Based IP Switch
- ❑ Proposed by IBM
- ❑ Topology based. One VC per egress router.
- ❑ Egress router initiates the setup of switched path
- ❑ Supports LAN media switching



# ARIS (Cont)

- mpt-to-pt VC  $\Rightarrow$  VC merge
- Integrated Switch Routers (ISRs)
- Globally unique labels  $\Rightarrow$  Each ISR has a VCI block



# Alphabet Soup

- ❑ CSR Cell Switched Router
- ❑ ISR Integrated Switch and Router
- ❑ LSR Label Switching Router
- ❑ TSR Tag Switching Router
- ❑ Multi layer switches, Swoters
- ❑ DirectIP
- ❑ FastIP
- ❑ PowerIP

# Switched IP Forwarding: Comparison

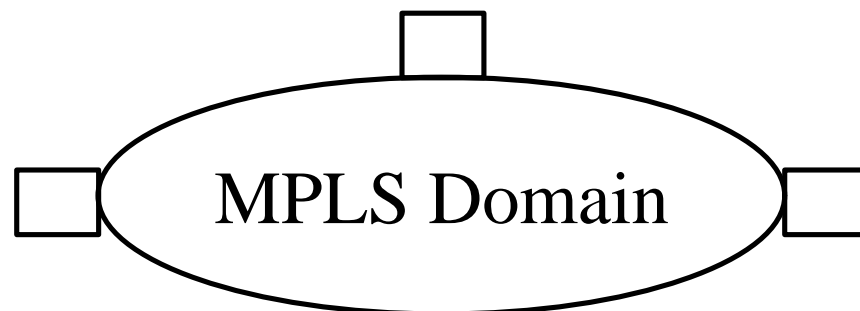
<b>Issue</b>	<b>IP Switch</b>	<b>CSR</b>	<b>Tag</b>	<b>ARIS</b>
Datalink	ATM	ATM, FR	ATM, FR, Ethernet	ATM, FR
Network Layer	IP	IP	IP, XNS, ...	IP
Initiator	Downstream	Both	Both	Egress
VC Setup Protocol	IFMP	FANP	TDP	ARIS
Mapping	Traffic	Traffic	Topology	Topology
# of VCs	# of L4 flows	# of L3 flows	# of routes	# of Egress routers

# 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 LAN.
- ❑ Not specific to a routing protocol (OSPF, RIP, ...)
- ❑ Optimization only. Labels do not affect the path.  
Only speed. Networks continue to work w/o labels
- ❑ Complete spec by the end of 1997

# Terminology

- ❑ Label = Short fixed length, physically contiguous, locally significant
- ❑ Stream =  $\Sigma$  flows = pt-pt, pt-mpt, mpt-pt, mpt-mpt
- ❑ Stream Merge  $\Rightarrow$  Stream =  $\Sigma$  streams
- ❑ Label information base (LIB)  $\cong$  Routing info base
- ❑ Label distribution protocol (LDP)  $\cong$  Routing protocols
- ❑ MPLS edge node = Egress or ingress node

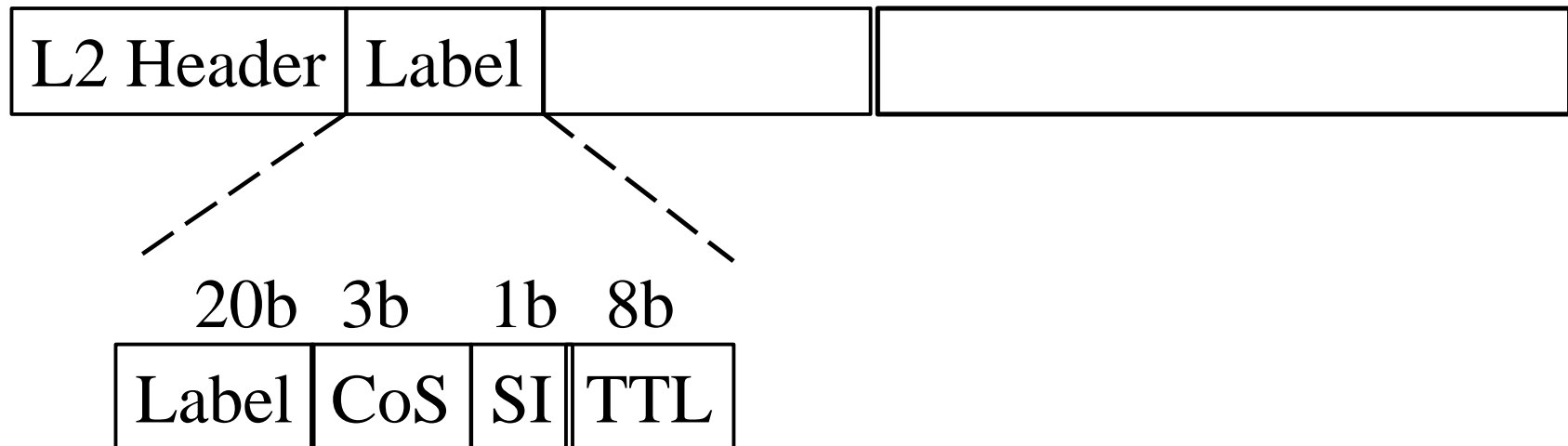


# Label Assignment

- ❑ Binding between a label and a route
- ❑ Traffic, topology, or reservation driven
- ❑ Traffic: Initiated by upstream/downstream/both
- ❑ Topology: One per route, one per MPLS egress node.
- ❑ Labels may be preassigned
  - ⇒ first packet can be switched immediately
- ❑ Reservations: Labels assigned when RSVP “RESV” messages sent/received.
- ❑ Unused labels are "garbage collected"
- ❑ Labels may be shared, e.g., in some multicasts

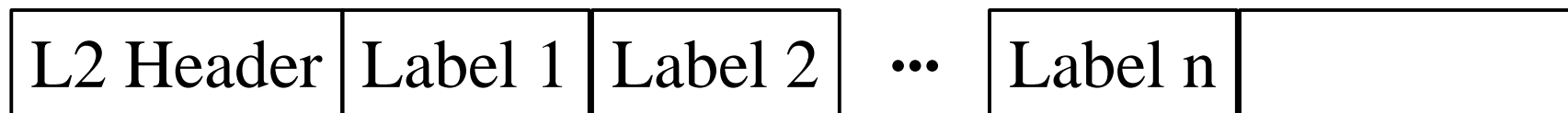
# Label Format

- ❑ Labels = Explicit or implicit L2 header
- ❑ TTL = Time to live
- ❑ CoS = Class of service
- ❑ SI = Stack indicator



# Label Stacks

- ❑ Labels are pushed/popped as they enter/leave MPLS domain
- ❑ Routers in the interior will use Interior Gateway Protocol (IGP) labels. Border gateway protocol (BGP) labels outside.

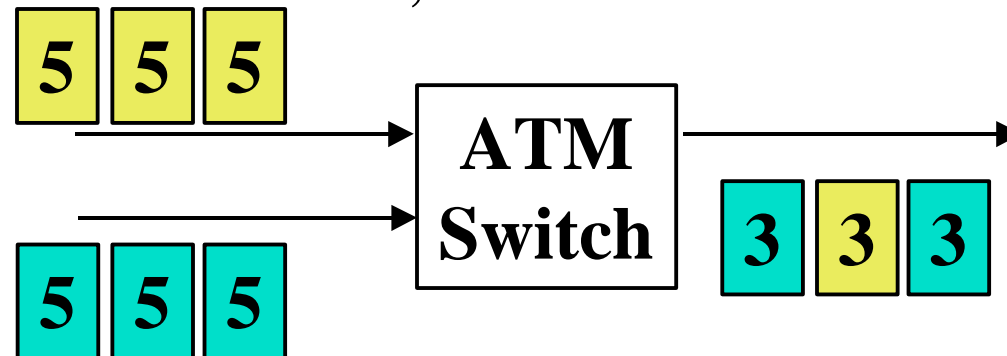


# Label Distribution

- ❑ Who assigns labels for communication between A and B?
  - A, B, or someone else?
  - Downstream, upstream, ...
- ❑ Where is the control for the entire path?  
A, B, ingress or egress LSR?
- ❑ Separate protocol or existing route distribution mechanisms?
  - Tag Distribution Protocol (TDP)
  - Flow Attribute Notification Protocol (FANP)

# 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



# MPLS on ATM: Issues

- ❑ VCI field is sufficient for one level tagging  
VPI may be used for the 2nd level
- ❑ LSR switches need to participate in network layer routing protocols (OSPF, BGP)
- ❑ Multiple tags per destination may be used to avoid frame merging
- ❑ VPI/VCI space may be segmented for label switching and normal ATM switching

# Other Issues

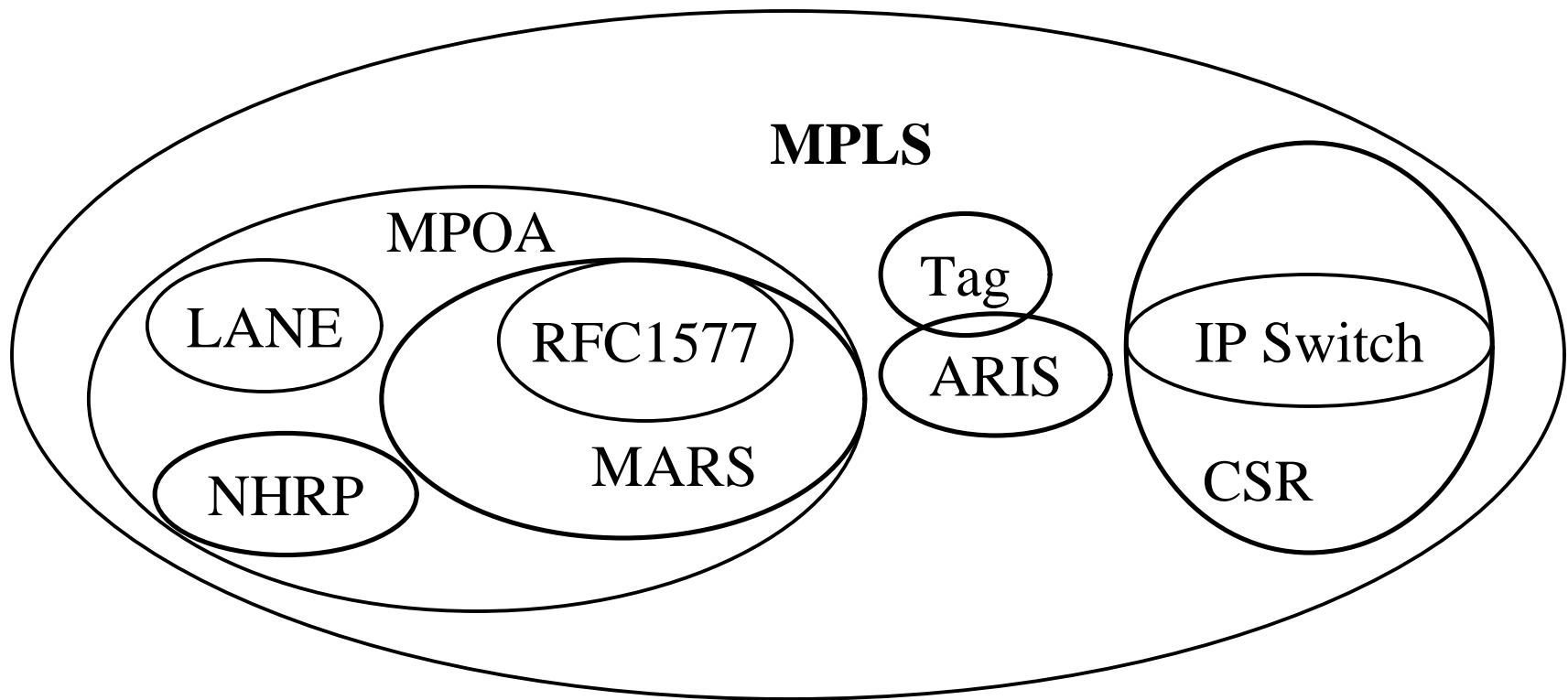
- ❑ Loop prevention, detection, survival
- ❑ Multicast:  
Multiple entries in label information base
- ❑ Multipath: Streams going to the same destination but different sources/port # may be assigned separate labels.
- ❑ Host involvement: Label-enabled hosts will avoid first hop reassembly
- ❑ Security: Label swapping may be terminated before firewall



# Summary

- ❑ IP Switching: Traffic-based, per-hop VCs, downstream originated
- ❑ CSR: Traffic-based, VCs (VCID), originated by downstream/upstream/both
- ❑ Tag switching: Topology based, one VC per route
- ❑ ARIS: Topology based, one VC per egress router
- ❑ MPLS combines various features of IP switching, CSR, Tag switching, ARIS

# Summary (Cont)



# Key References

- ❑ For a detailed list of references see [http://www.cis.ohio-state.edu/~jain/refs/atm\\_refs.htm](http://www.cis.ohio-state.edu/~jain/refs/atm_refs.htm)
- ❑ "A Framework for Multiprotocol Label Switching", 05/12/1997, <http://www.internic.net/internet-drafts/draft-ietf-mpls-framework-00.txt>
- ❑ RFC 2098, "Toshiba's Router Architecture Extensions for ATM : Overview", 02/04/1997, 18 pp., <http://ds.internic.net/rfc/rfc2098.txt>

## References (Cont)

- ❑ RFC 2105, "Cisco Systems' Tag Switching Architecture Overview", 02/06/1997, 13 pp., <http://ds.internic.net/rfc/rfc2105.txt>
- ❑ "ARIS: Aggregate Route-Based IP Switching", 03/26/1997, <http://www.internic.net/internet-drafts/draft-viswanathan-aris-overview-00.txt>
- ❑ Multiprotocol Label Switching (mpls) working group at IETF. Email: [mpls-request@cisco.com](mailto:mpls-request@cisco.com)
- ❑ RFC 1954, "Transmission of flow labeled IPv4 on ATM datalinks Ipsilon V1.0," 5/22/96.

## References (Cont)

- ❑ ATM Forum, "MPOA V1.0," Letter Ballot, June 1997, (available to ATM Forum members only)  
<http://www-mo.atmforum.com/ftp/atm/letter-ballot/af-mpoa-0087.000.ps>
- ❑ "NBMA Next Hop Resolution Protocol (NHRP)",  
<http://www.internic.net/internet-drafts/draft-ietf-rolc-nhrp-11.txt>, 3/5/97.
- ❑ RFC 1577, "Classical IP and ARP over ATM,"  
1/20/94, <http://ds.internic.net/rfc/rfc1577.txt>

# References (Cont)

- LAN Emulation over ATM v1.0 Specification (Jan 1995), <ftp://ftp.atmforum.com/pub/approved-specs/af-lane-0021.000.ps>