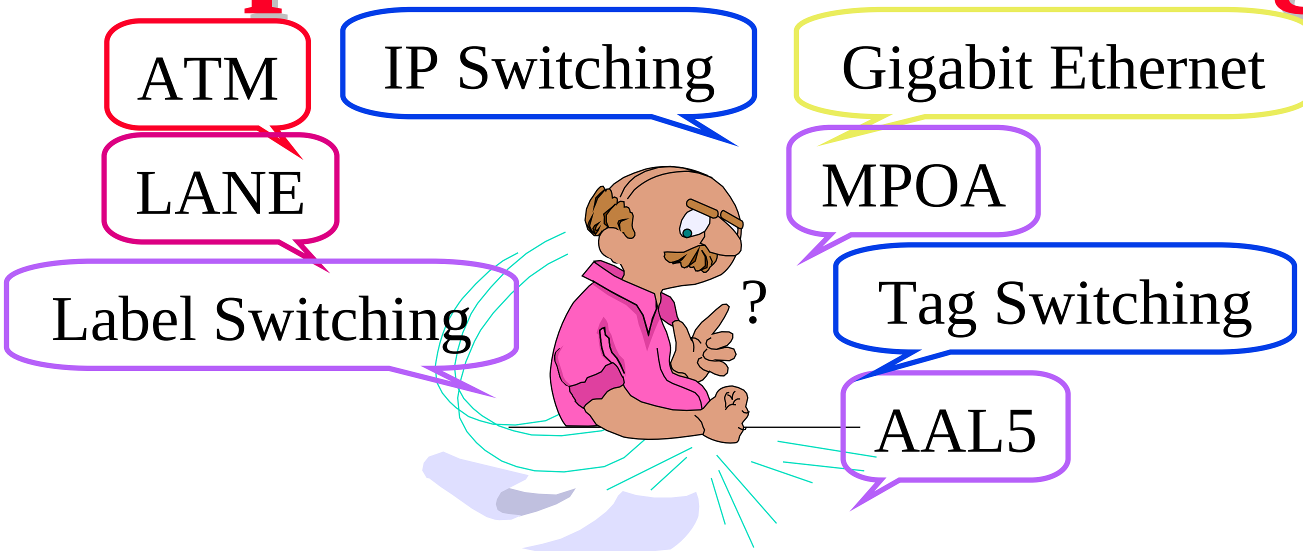
Recent Advances in Computer Networking



Raj Jain

Profe ... Sciences

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

N ... n/

Overview

- Networking Trends
- ATM Networks
- Legacy protocols over ATM
- IP/Tag/Label Switching
- Gigabit Ethernet

Networking Trends

- Networking Trends
- Impact of Networking
- Current Research Topics

ATM Networks

- ATM vs Phone Networks and Data Networks
- ATM Protocol Layers
- Cell Header Format
- AALs
- Physical Media
- Traffic Management: ABR vs UBR

ATM: Issues and Challenges

- Requirements for Success
- Economy of Scale
- High Performance
- Simplicity

LANE and IP over ATM

- LAN Emulation (LANE)
- IP over ATM (IPOA)
- Multicast Address Resolution Server (MARS)
- Next Hop Resolution Protocol (NHRP)
- Multiprotocol over ATM (MPOA)

IP Switching

- IP Switch
- Cell Switched Router
- Tag Switching (CISCO)
- ARIS (IBM)
- Multi-protocol label switching

Gigabit Ethernet

- Distance-Bandwidth Principle
- 10 Mbps to 100 Mbps
- Gigabit PHY Issues
- Gigabit MAC Issues
- Status
- ATM vs Gigabit Ethernet

Schedule (Tentative)

- 9:00-9:30 Course Introduction
- 9:30-10:30 Trends
- 10:30-10:45 *Coffee Break*
- 10:45-11:45 ATM Networks
- 11:45-12:00 ATM: Issues and Challenges
- 12:00-1:00 *Lunch Break*
- 1:00-2:00 LANE and IP over ATM
- 2:00-3:00 IP Switching and Alternatives
- 3:00-3:15 *Coffee Break*
- 3:15-5:00 Gigabit Ethernet

References

- You can get to all on-line references via: http://www.cis.ohio-state.edu/~jain/refs/hot_refs.htm

Disclaimer

- The technologies are currently evolving.
  $\Rightarrow$ Many statements are subject to change.
- Features not in a technology may be implemented later in that technology.
- Problems claimed to be in a technology may later not be a problem.

Networking Trends and Their Impact

Raj Jain
The Ohio State University
Columbus, OH 43210
Jain@CIS.Ohio-State.Edu

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

- Networking Trends
- Impact of Networking
- Current Research Topics

Trends

- Communication is more critical than computing
  - Greeting cards contain more computing power than all computers before 1950.
  - Genesis's game has more processing than 1976 Cray supercomputer.
- Internet: 0.3 M hosts in Jan 91 to 9.5 M by Jan 96 $\Rightarrow$ More than 5 billion (world population) in 2003

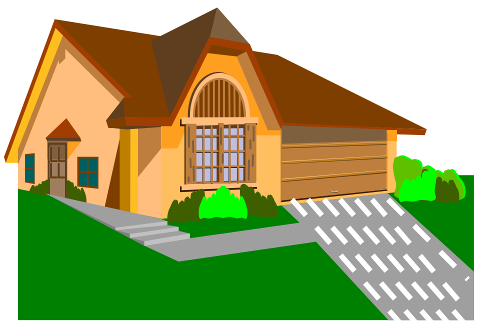
Stone Age to Networking Age

- Microwave ovens, stereo, VCRs, had some effect. But, Stone, iron, …, automotive, electricity, telephone, jet plane,…, networks caused a fundamental change in our life style
- In 1994, 9% of households with PC had Internet link. By 1997, 26%. Soon 98% … like TV and telephone.
- URL is more important than a company's phone number. (54 URLs in first 20 pages of March’97 Good Housekeeping.)
- Email is faster than telegrams

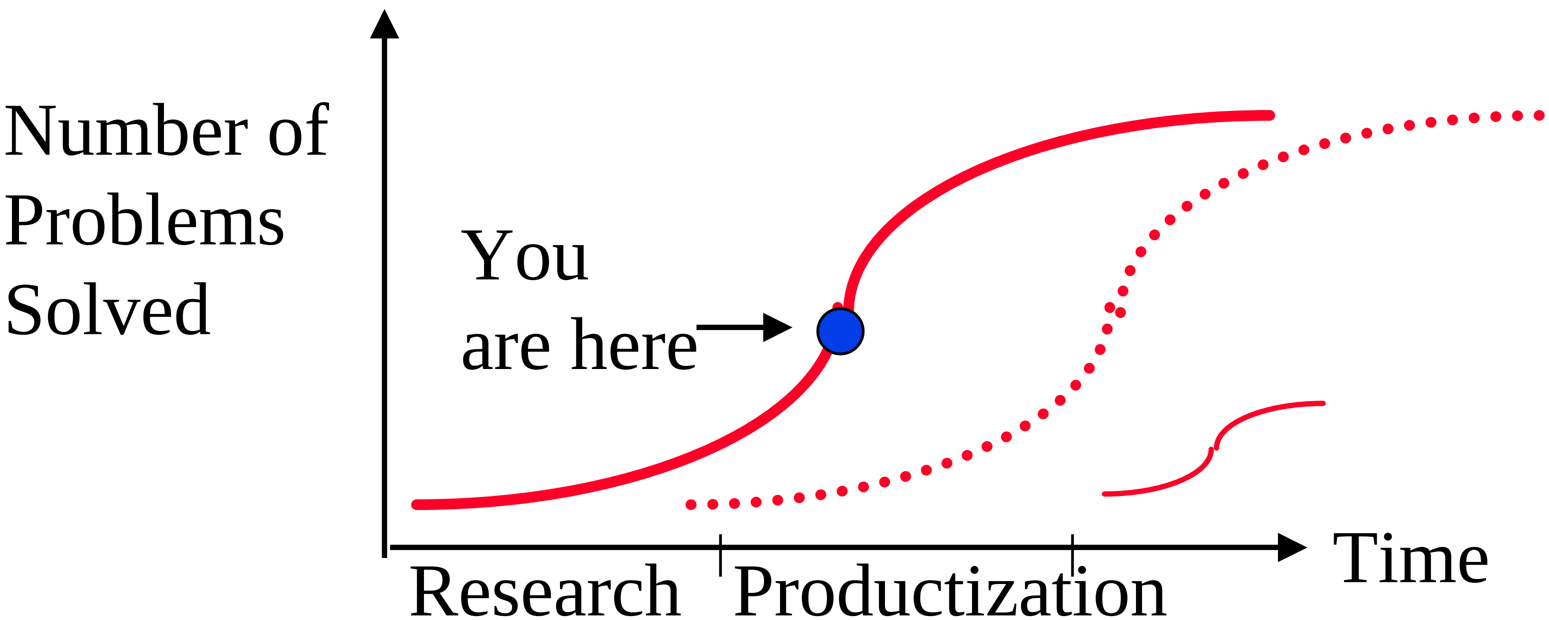
Raj Jain

Garden Path to I-Way

- Plain Old Telephone System (POTS) = 64 kbps = 3 ft garden path
- ISDN = 128 kbps = 6 ft sidewalk
- T1 Links to Businesses = 1.544 Mbps = 72 ft = 4 Lane roadway
- Cable Modem Service to Homes: = 10 Mbps = 470 ft = 26 Lane Driveway
- OC3 = 155 Mbps = 1 Mile wide superhighway
- OC48 = 2.4 Gbps = 16 Mile wide superhighway

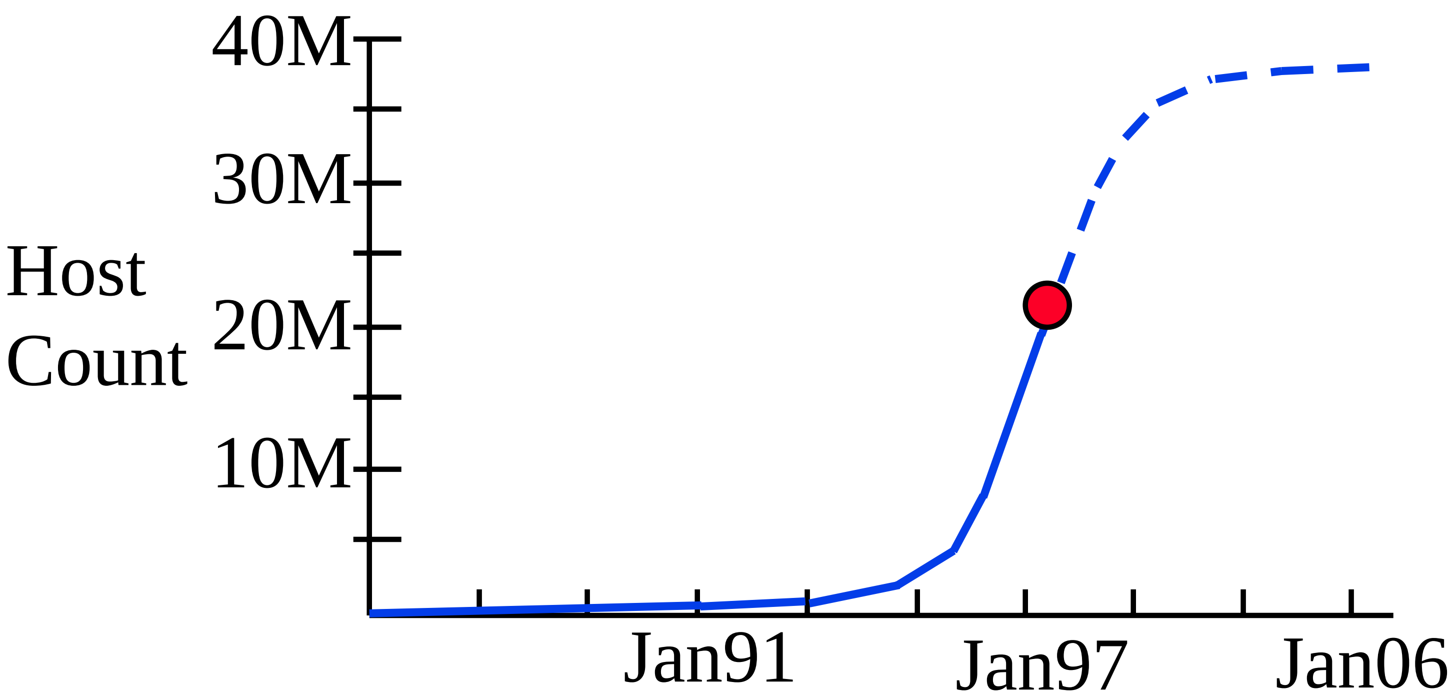


Life Cycles of Technologies



Raj Jain

Internet Technology



- **New Challenges**: Exponential growth in number of users. Exponential growth in bandwidth per user. Traffic management, Security, Usability, ...

Impact on R&D

- Too much growth in one year
  ⇒ Can't plan too much into long term
- Long term = 1_2 year or 10_2 years at most
- Products have life span of 1 year, 1 month, …
- Short product development cycles.
  Chrysler reduced new car design time
  from 6 years to 2.
- Distance between research and products has narrowed
  ⇒ Collaboration between researchers and developers
  ⇒ Academics need to participate in industry consortia

Impact on Education

- Technology is changing faster than our ability to learn
  ⇒ Your value (salary) decreases with experience (years out of college)
- Recent graduates know C++, HTML, Java, …
- A handheld device will have storage enough to carry a small library
- Computers have bigger memory than humans
  ⇒ Knowing where to find the information is more important than the information
- Human memory is pointer cache

New Challenges

- Networking is moving from specialists to masses ⇒ Usability (plug & play), security
- Exponential growth in number of users + Exponential growth in bandwidth per user ⇒ Traffic management
- Standards based networking for reduced cost
  ⇒ Important to participate in standardization forums
  ATM Forum, Frame Relay Forum, …
  Internet Engineering Task Force (IETF),
  Institute of Electrical and Electronic Engineers (IEEE)
  International Telecommunications Union (ITU), …

Raj Jain

Recent Trends

- Copper is still in.
  6-27 Mbps on phone wire.
  Fiber is being postponed.
- Shared LANs to Switched LANs
- Routing to Switching. Distinction is disappearing
- LANs and PBX's to Integrated LANs
- Bandwidth requirements are doubling every 4 months

Research Topics

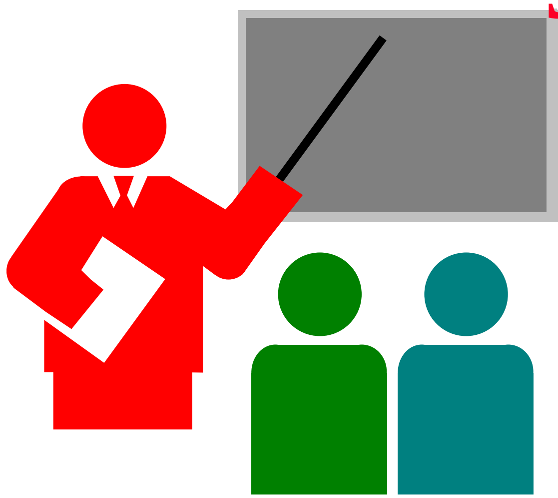
- Terabit networking: Wavelength division multiplexing, all-optical switching
- High-speed access from home
  $\Rightarrow$ Robust and high-bandwidth encoding techniques
- High-speed Wireless = More than 10 bit/Hz
  28.8 kbps on 30 kHz cellular $\Rightarrow$ 1 bit/Hz
- Traffic management, quality of service, multicasting:
  - Ethernet LANs, IP networks, ATM Networks
- Mobility
- Large network management Issues.

Raj Jain

Research Topics (Cont)

- Information Glut ⇒ Intelligent agents for searching, digesting, summarizing information
- Scalable Voice/Video compression: 2400 bps to 1.5 Mbps video, 8 kbps voice
- Electronic commerce ⇒ Security, privacy, cybercash
- Active Networks ⇒ A "program" in place of addresses

Summary



- Networking is the key to productivity
- It is impacting all aspects of life $\Rightarrow$ Networking Age
- Profusion of Information
- Collaboration between researchers and developers
- Usability, security, traffic management

Key References

- See http://www.cis.ohio-state.edu/~jain/refs/ref_trnd.htm
- "The Next 50 years," Special issue of Communications of the ACM, Feb 1997.
- D. Tapscott, "The Digital Economy: Promise and Peril in the Age of Networked Intelligence," McGraw-Hill, 1995.
- T. Lewis, "The Next $10{,}000_2$ years," IEEE Computer, April/May 1996

ATM Networks: An Overview

Raj Jain

Professor of Computer and Information Sciences

The Ohio State University

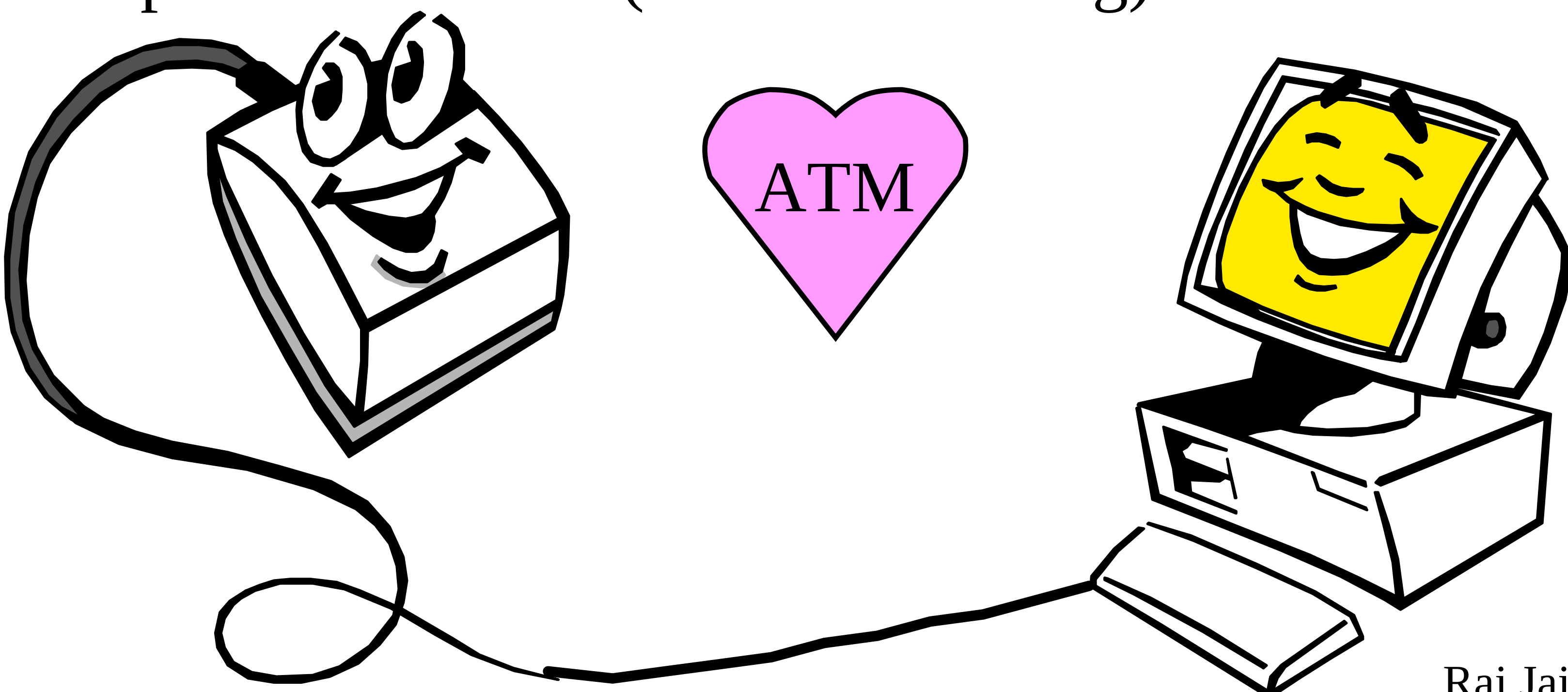
Columbus, OH 43210-1277

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

- ATM vs Phone Networks and Data Networks
- ATM Protocol Layers
- Cell Header Format
- AALs
- Physical Media
- Traffic Management: ABR vs UBR

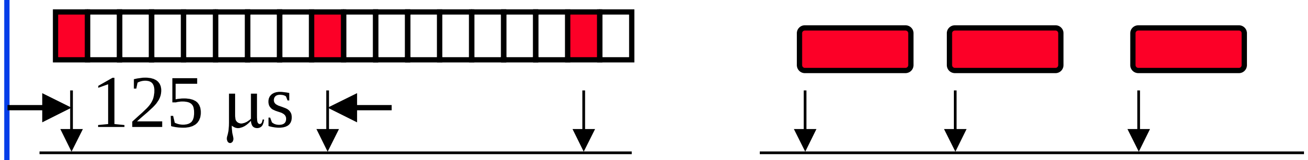
ATM

- ATM Net = Data Net + Phone Net
- Combination of Internet method of communication (packet switching) and phone companies' method (circuit switching)



Raj Jain

ATM vs Phone Networks



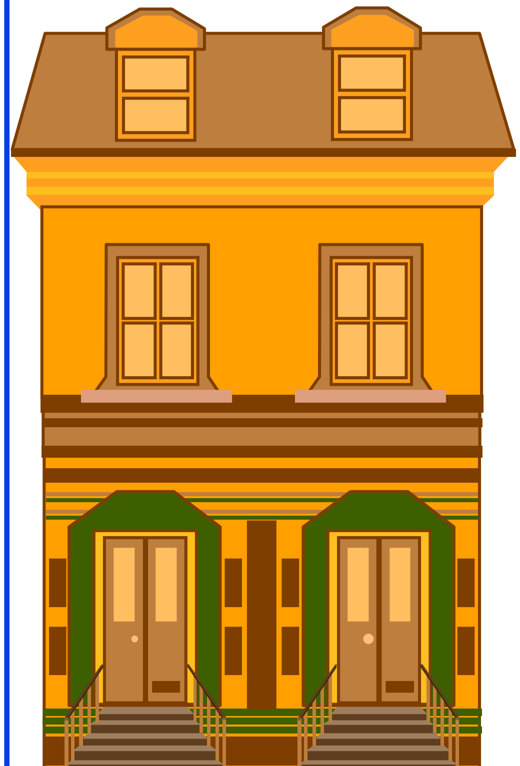
- Current phone networks are synchronous (periodic). ATM = Asynchronous Transfer Mode
- Phone networks use circuit switching. ATM networks use "Packet" Switching
- In phone networks, all rates are multiple of 8 kbps. With ATM service, you can get any rate. You can vary your rate with time.
- With current phone networks, all high speed circuits are manually setup. ATM allows dialing any speed.

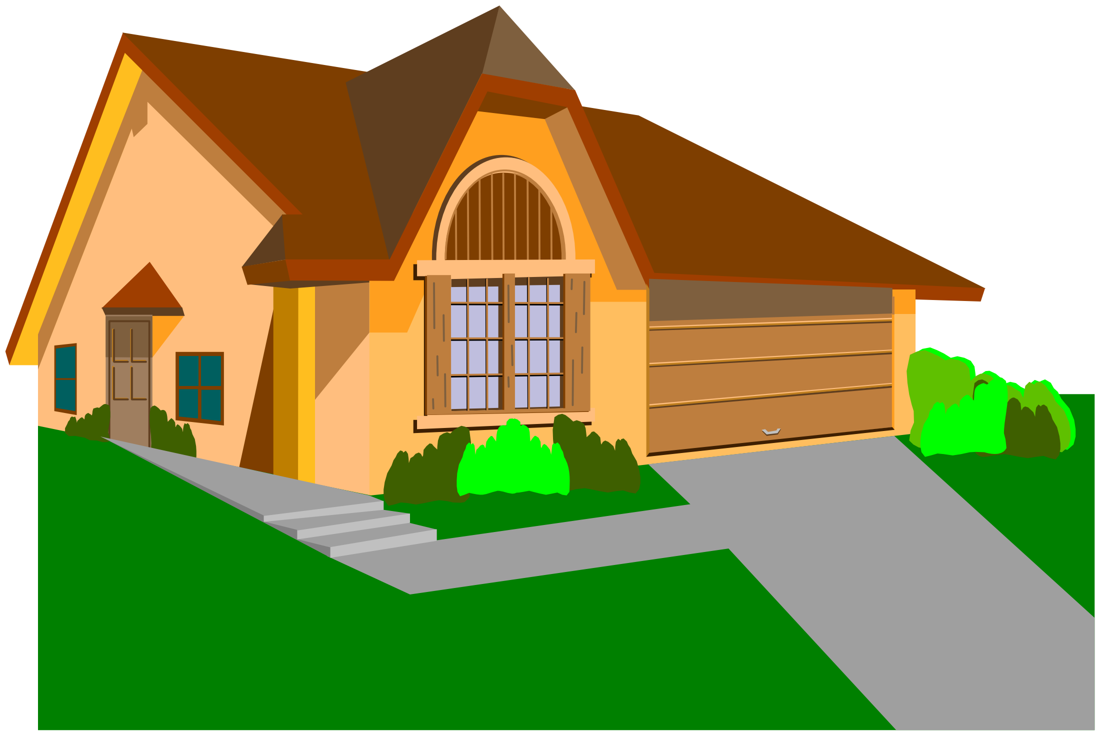
Raj Jain

ATM vs Data Networks

- Signaling: Internet Protocol (IP) is connectionless.
  You cannot reserve bandwidth in advance.
  ATM is connection-oriented.
  You declare your needs before using the network.
- PNNI: Path based on quality of service (QoS)
- Switching: In IP, each packet is addressed and processed individually.
- Traffic Management: Loss based in IP.
  ATM has 1996 traffic management technology.
  Required for high-speed and variable demands.
- Cells: Fixed size or small size is not important

Old House vs New House

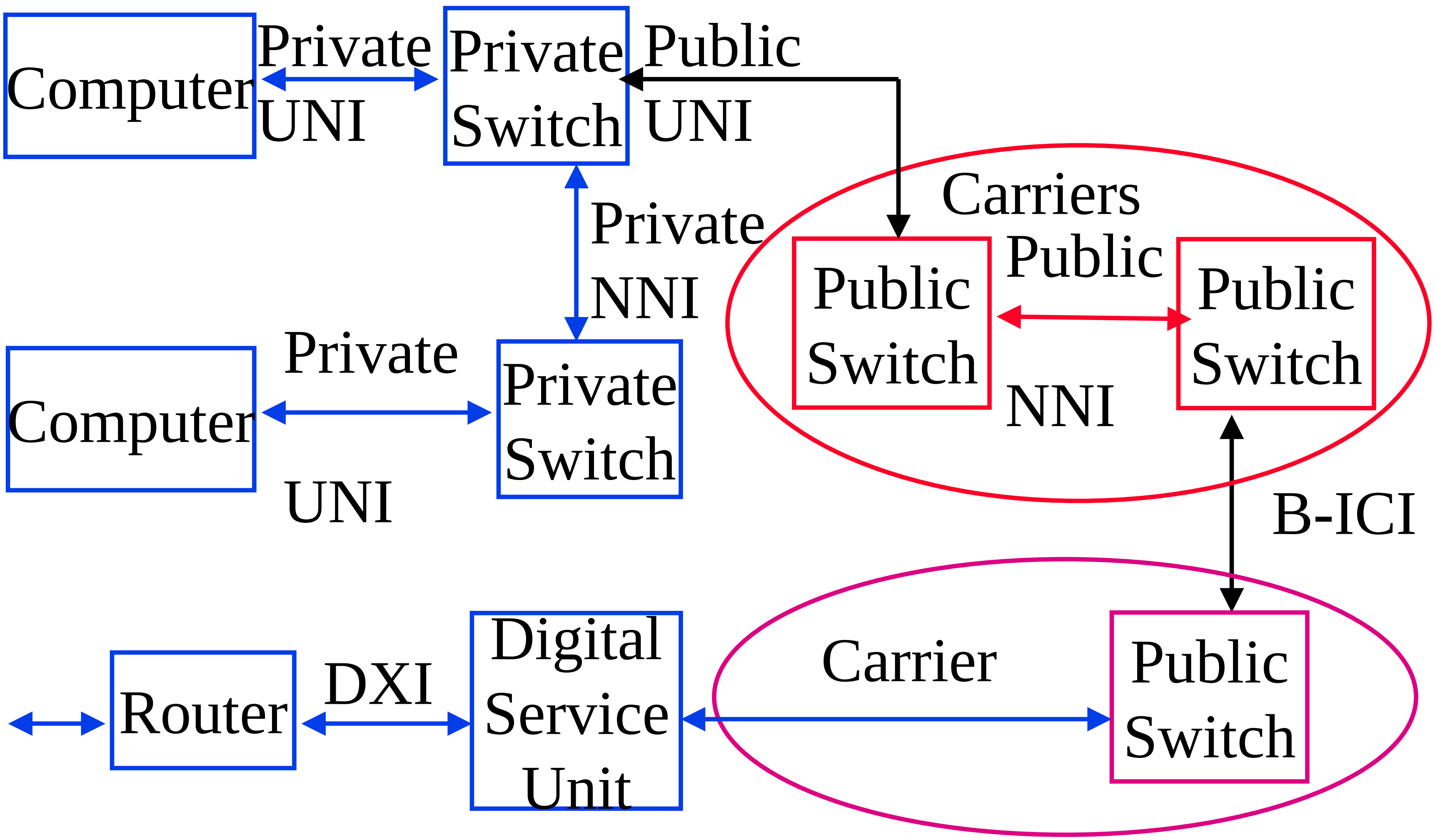




- New needs:
  Solution 1: Fix the old house (cheaper initially)
  Solution 2: Buy a new house (pays off over a long run)

Raj Jain

ATM Interfaces

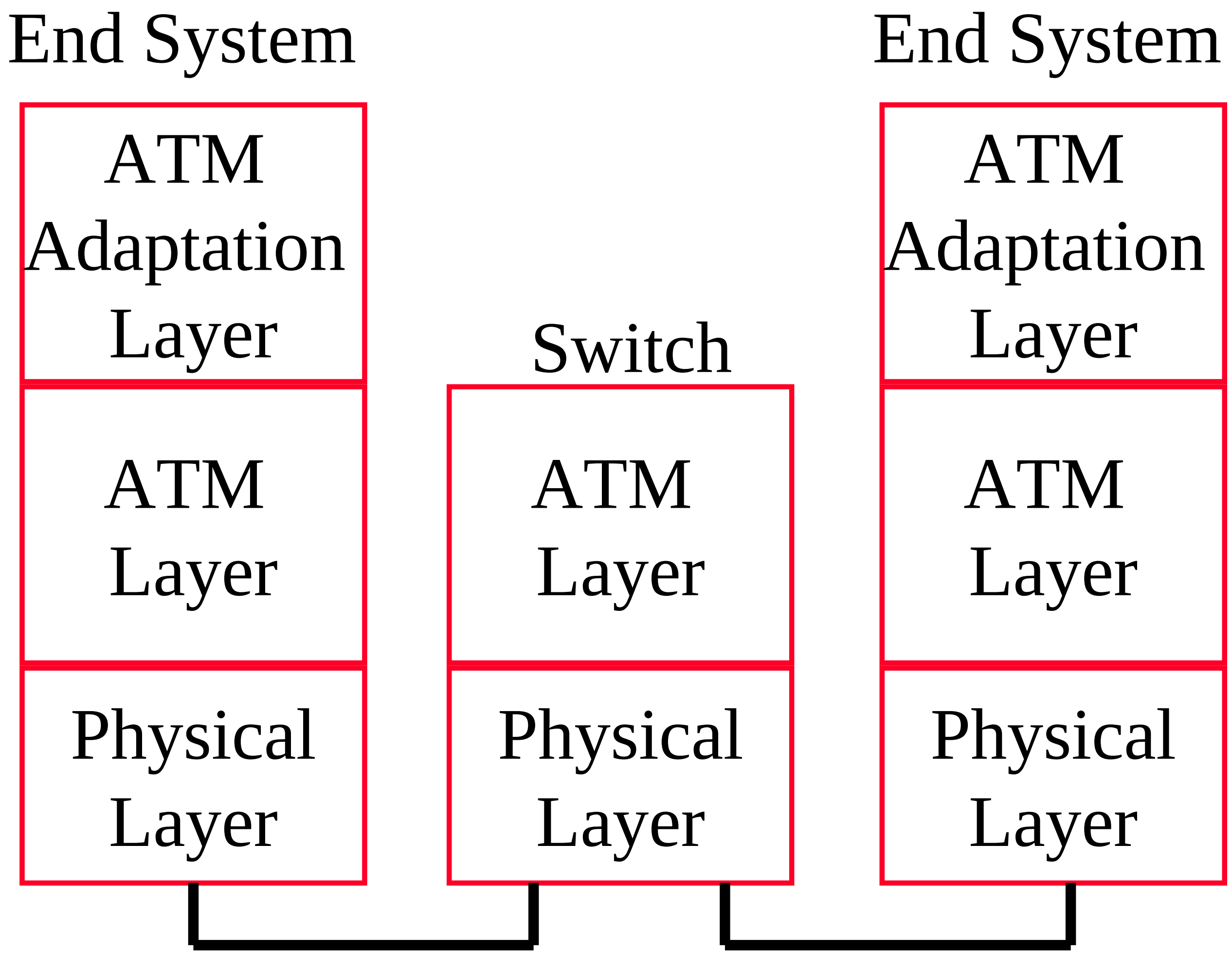


Raj Jain

ATM Interfaces

- User to Network Interface (UNI):
  Public UNI, Private UNI
- Network to Node Interface (NNI):
  - Private NNI (P-NNI)
  - Public NNI =Inter-Switching System Interface (ISSI)
    Intra-LATA ISSI (Regional Bell Operating Co)
  - Inter-LATA ISSI (Inter-exchange Carriers)
    ⇒ Broadband Inter-Carrier Interface (B-ICI)
- Data Exchange Interface (DXI)
  Between routers and ATM Digital Service Units (DSU)

Protocol Layers



Protocol Layers

- The ATM Adaptation Layer
  - How to break messages to cells
- The ATM Layer
  - Transmission/Switching/Reception
  - Congestion Control/Buffer management
  - Cell header generation/removal at source/destination
  - Cell address translation
  - Sequential delivery

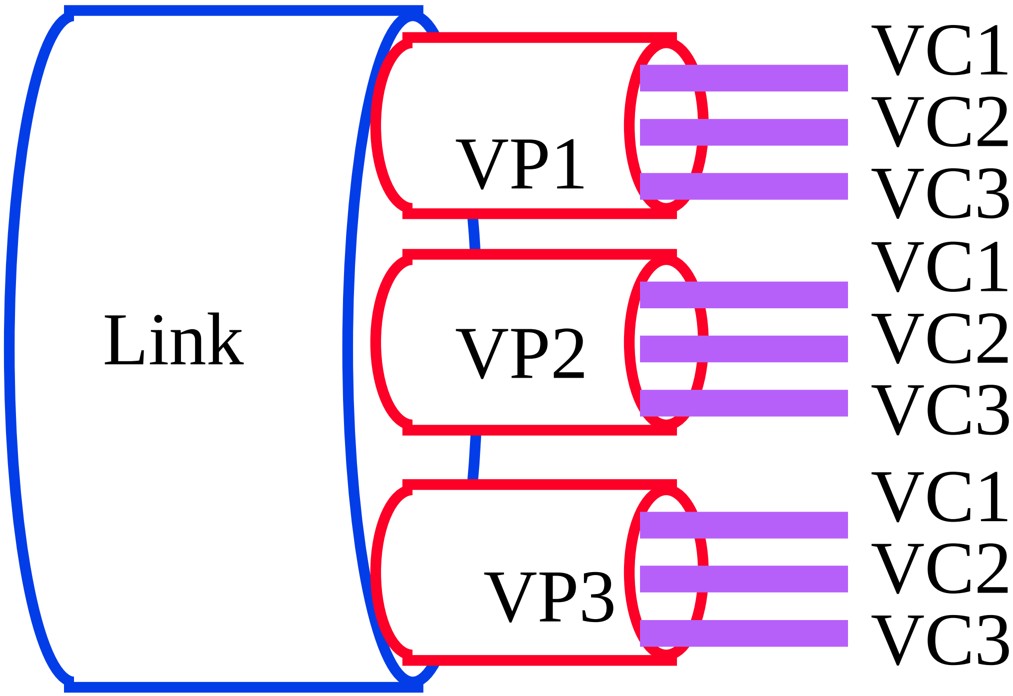
Cell Header Format

- GFC = Generic Flow Control
  - (Was used in UNI but not in NNI)
- VPI/VCI = 0/0 ⇒ Idle cell; 0/n ⇒ Signaling
- HEC: $1 + x + x^2 + x^8$

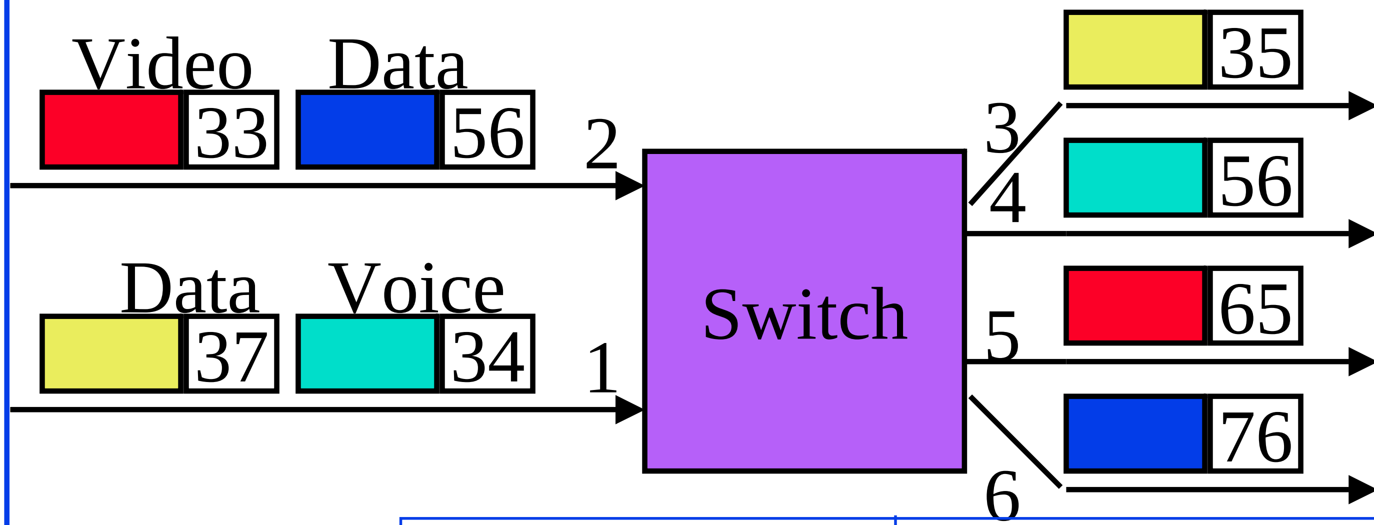
<table>
<tr><td>GFC/VPI</td><td colspan="2">VPI</td></tr>
<tr><td>VPI</td><td colspan="2">VCI</td></tr>
<tr><td colspan="3">VCI</td></tr>
<tr><td>VCI</td><td>PTI</td><td>CLP</td></tr>
<tr><td colspan="3">Header Error Check (HEC)</td></tr>
<tr><td colspan="3">Payload</td></tr>
</table>

Path vs Channels

- 24/28-bit connection identifier
  First 8/12 bits: Virtual Path,
  Last 16 bits: Virtual Circuit
- VP service allows new VC's w/o orders to carriers



VP/VC Assignment/Use



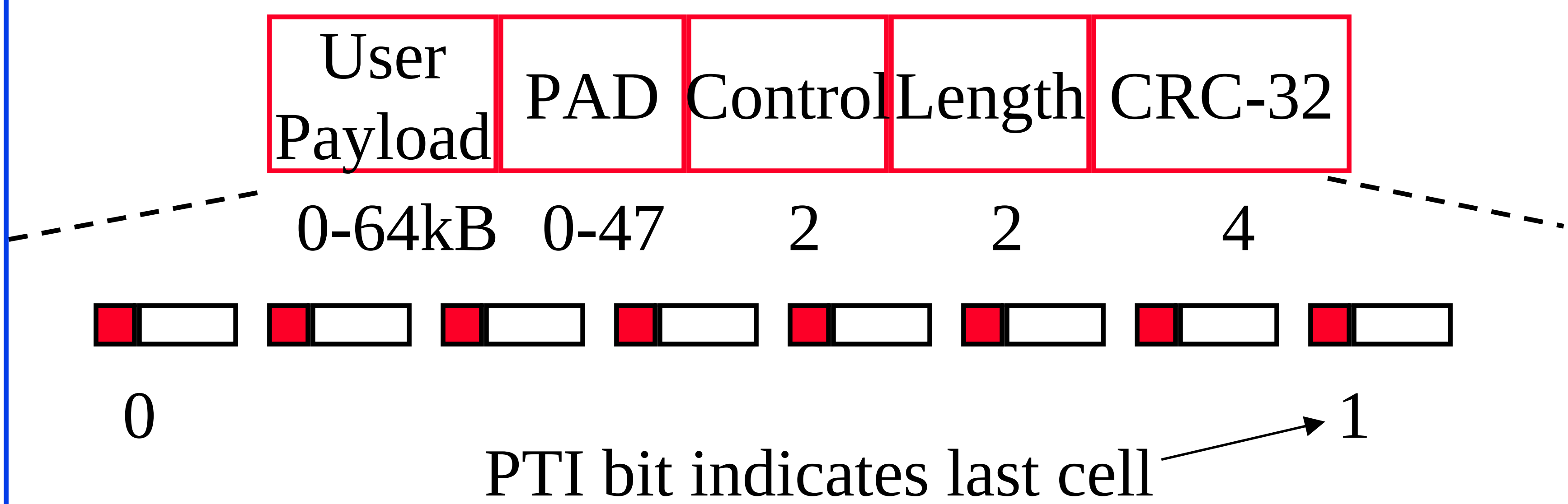
In		Out	
Port	VPI/VCI	Port	VPI/VCI
1	0/37	3	1/35
1	0/34	4	0/56
2	0/33	5	0/65
2	0/56	6	4/76

Original Classes of Traffic

	Class A	Class B	Class C	Class D
Time Sync	Yes	Yes	No	No
Bit Rate	Constant	Variable	Variable	Variable
Connection -Oriented	Yes	Yes	Yes	No
Examples	Circuit Emulation	Comp. Video	Frame Relay	SMDS
AAL	AAL1	AAL2	AAL3	AAL4

AAL 5

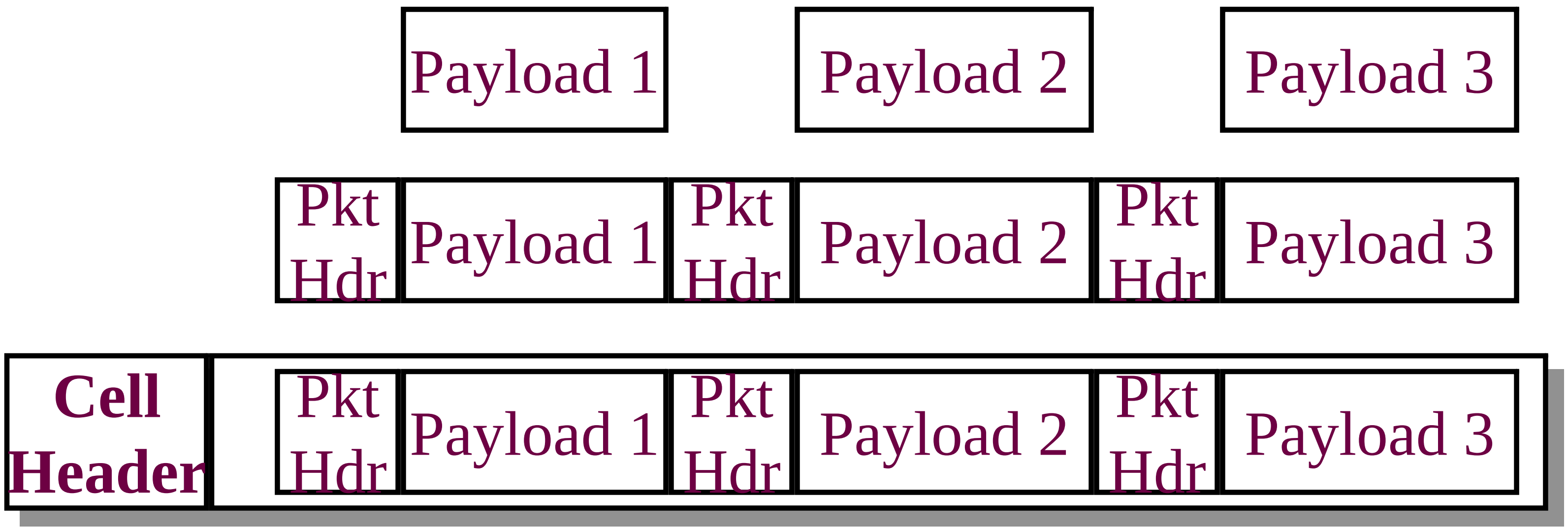
- Designed for data traffic
- Less overhead bits than AAL 3/4
  Simple and Efficient AAL (SEAL)
- No per cell length field, No per cell CRC



Raj Jain

AAL2

- Ideal for low bit rate voice
- Variable/constant rate voice
- Multiple users per VC
- Compression and Silence suppression
- Idle channel suppression



Raj Jain

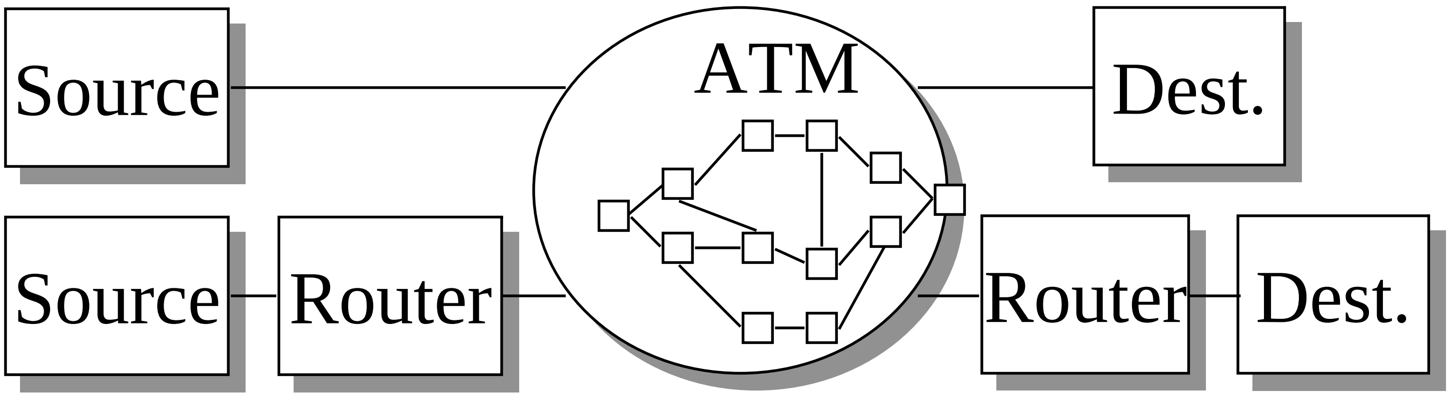
Physical Media Dependent Layers

- Multimode Fiber: 100 Mbps using 4b/5b, 155 Mbps SONET STS-3c, 155 Mbps 8b/10b
- Single-mode Fiber: 155 Mbps STS-3c, 622 Mbps
- Shielded Twisted Pair (STP): 155 Mbps 8b/10b
- Coax: 45 Mbps, DS3, 155 Mbps
- Unshielded Twisted Pair (UTP)
  - UTP-3 (phone wire) at 25.6 Mbps, 51.84 Mbps
  - UTP-5 (Data grade UTP) at 155 Mbps
- DS1, DS3, STS-3c, STM-1, E1, E3, J2, n × T1

Classes of Service

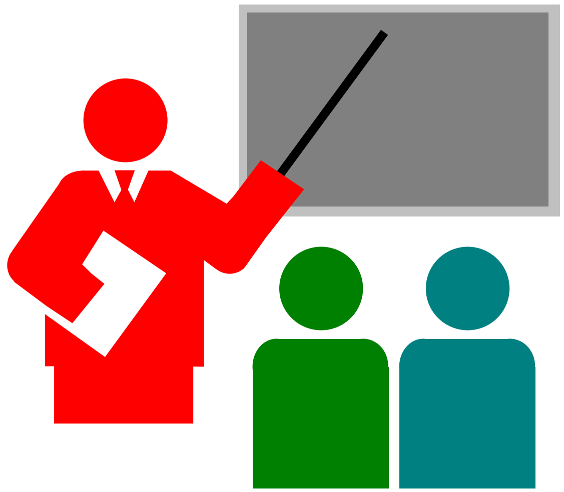
- ABR (Available bit rate):
  Source follows network feedback.
  Max throughput with minimum loss.
- UBR (Unspecified bit rate):
  User sends whenever it wants. No feedback. No guarantee. Cells may be dropped during congestion.
- CBR (Constant bit rate): User declares required rate. Throughput, delay and delay variation guaranteed.
- VBR (Variable bit rate): Declare avg and max rate.
  - rt-VBR (Real-time): Conferencing.
    Max delay guaranteed.
  - nrt-VBR (non-real time): Stored video.

ABR vs UBR



ABR	**UBR**
Queue in the source	Queue in the network
Pushes congestion to edges	No backpressure
Good if end-to-end ATM	Same end-to-end or backbone
Fair	Generally unfair
Good for the provider	Simple for user

Summary

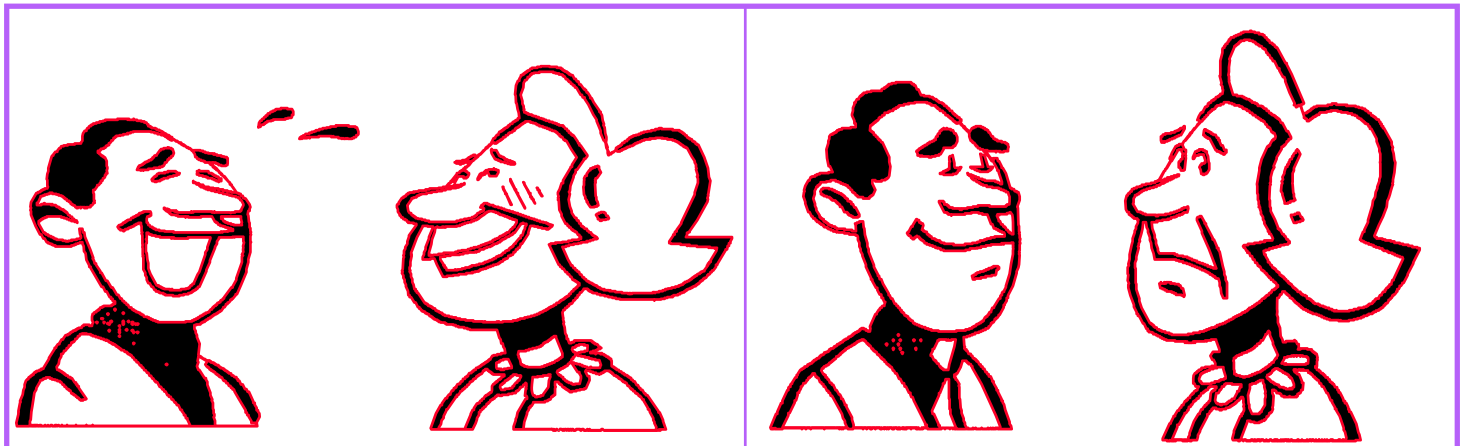


- ATM Overview: History, Why and What
- Protocol Layers: AAL, ATM, Physical layers, Cell format
- Interfaces: PNNI, NNI, B-ICI, DXI
- ABR, CBR, VBR, UBR

ATM : Key References

- See http://www.cis.ohio-state.edu/~jain/refs/atm_refs.htm
- G. Sackett and C. Y. Metz, "ATM and Multiprotocol Networking," McGraw-Hill, 1997 (Technical).
- S. Siu and R. Jain, "A brief overview of ATM: Protocol Layers, LAN Emulation and Traffic Management" Computer Communications Review (ACM SIGCOMM), April 1995. Available at http://www.cis.ohio-state.edu/~jain/
- ATM Forum specs are available at ftp://ftp.atmforum.com/pub/approved-specs/

ATM Networking: Issues and Challenges Ahead



Raj Jain
Professor of Computer and Information Sciences
The Ohio State University
Columbus, OH 43210-1277
http://www.cis.ohio-state.edu/~jain/

Overview

- ❑ Requirements for Success
- ❑ Economy of Scale
- ❑ High Performance
- ❑ Simplicity

Ref: R. Jain, “ATM Networks: Issues and Challenges head,” NetWorld+Interop Engineering Conference, March 1995. Available on http://www.cis.ohio-state.edu/~jain/

Raj Jain

Networking: Failures vs Successes

- 1980: Broadband (vs baseband)
- 1981: PBX (vs Ethernet)
- 1984: ISDN (vs Modems)
- 1986: MAP/TOP (vs Ethernet)
- 1988: OSI (vs TCP/IP)
- 1991: DQDB
- 1992: XTP (vs TCP)
- 1994: CMIP (vs SNMP)

Requirements for Success

- Low Cost
- High Performance
- Killer Applications
- Timely completion
- Manageability
- Interoperability
- Coexistence with legacy LANs
  Existing infrastructure is more important than new technology

Raj Jain

Economy of Scale

- Technology is far ahead of the applications.
  Invention is becoming the mother of necessity.
  We have high speed fibers, but no video traffic.
- Low-cost is the primary motivator. Not necessity.
  Buyer's market (Like $99 airline tickets.)
  Why? vs Why not?
- Ten 100-MIPS computer cheaper than a 1000-MIPS
  Parallel computing, not supercomputing
- Ethernet was and is cheaper than 10 one-Mbps links.
- No FDDI if it is 10 times as expensive as Ethernet.
  10/100 Ethernet adapters = $50 over 10 Mbps

Challenge: Tariff

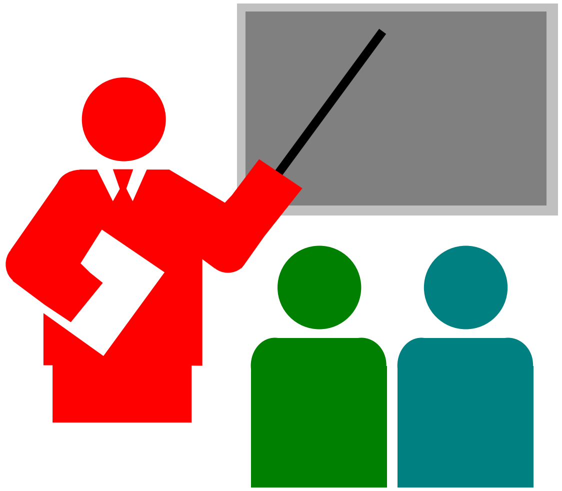
- Phone company's goal: How to keep the voice business and get into data too?
- Customer's goal: How to transmit the data cheaper?
- Tariff Today:
  - 64 kbps voice line = \$300/year
  - 45 Mbps line (\$45/mile/month)
    Coast to coast = \$180 k-240 k/year
    ⇒ 155 Mbps line = \$540 k - \$720 k/year
- Tomorrow: 155 Mbps = \$1k/month+ \$28/G cells
  ⇒ \$13k - \$45k/year

Raj Jain

Challenge: Simplicity

- One size fits all ⇒ Complexity
  Too many options too soon. Should work for
  - CBR and ABR
  - LAN and WAN
  - Private and Public
  - Low speed and High speed
- Switches have to do connection setup, route determination, address translation, anycasting, multicasting, flow control, congestion control, ...
- Many independent forums (ITU vs ATM Forum)
  ⇒ People energy divided

Summary



- High speed networking iff economy of scale
- Solving all problems can lead to complexity and failure.
- To succeed, ATM has to solve today's problem (data) well.

LANE and IP over ATM

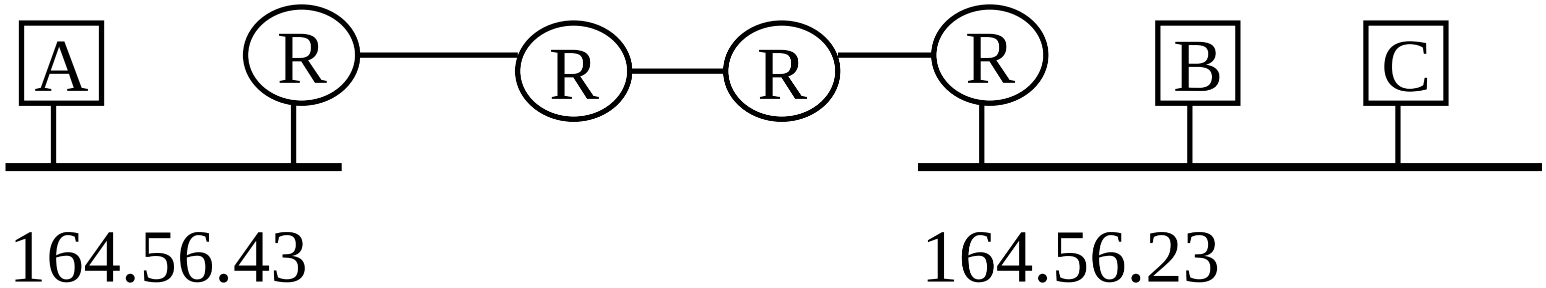
Raj Jain
Professor of Computer and Information Sciences
The Ohio State University

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

- LAN Emulation (LANE)
- IP over ATM (IPOA)
- Multicast Address Resolution Server (MARS)
- Next Hop Resolution Protocol (NHRP)
- Multiprotocol over ATM (MPOA)

IP Forwarding:Fundamentals

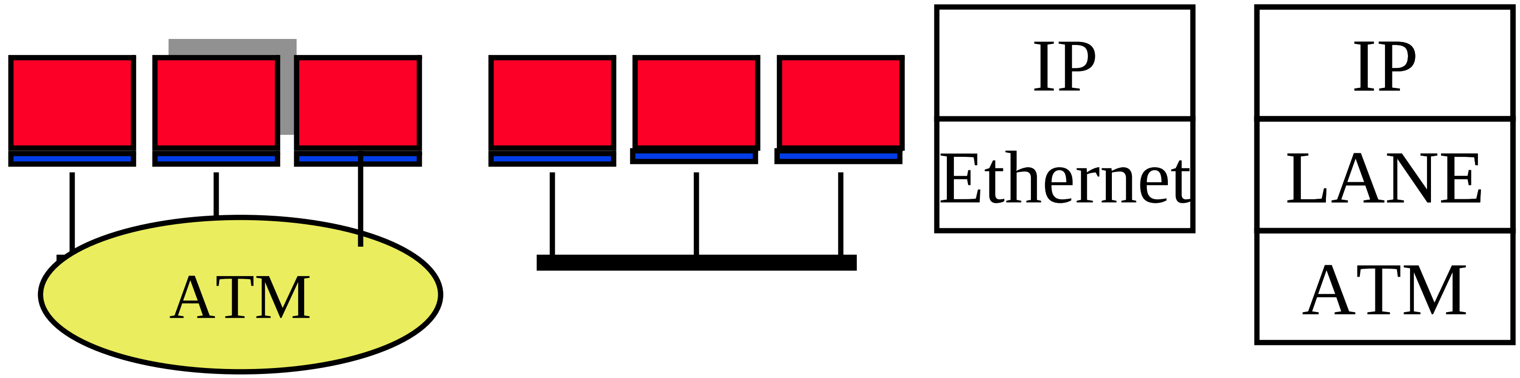
To: 164.56.23.34	From: 164.56.43.96	



- 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....

Raj Jain

LAN Emulation



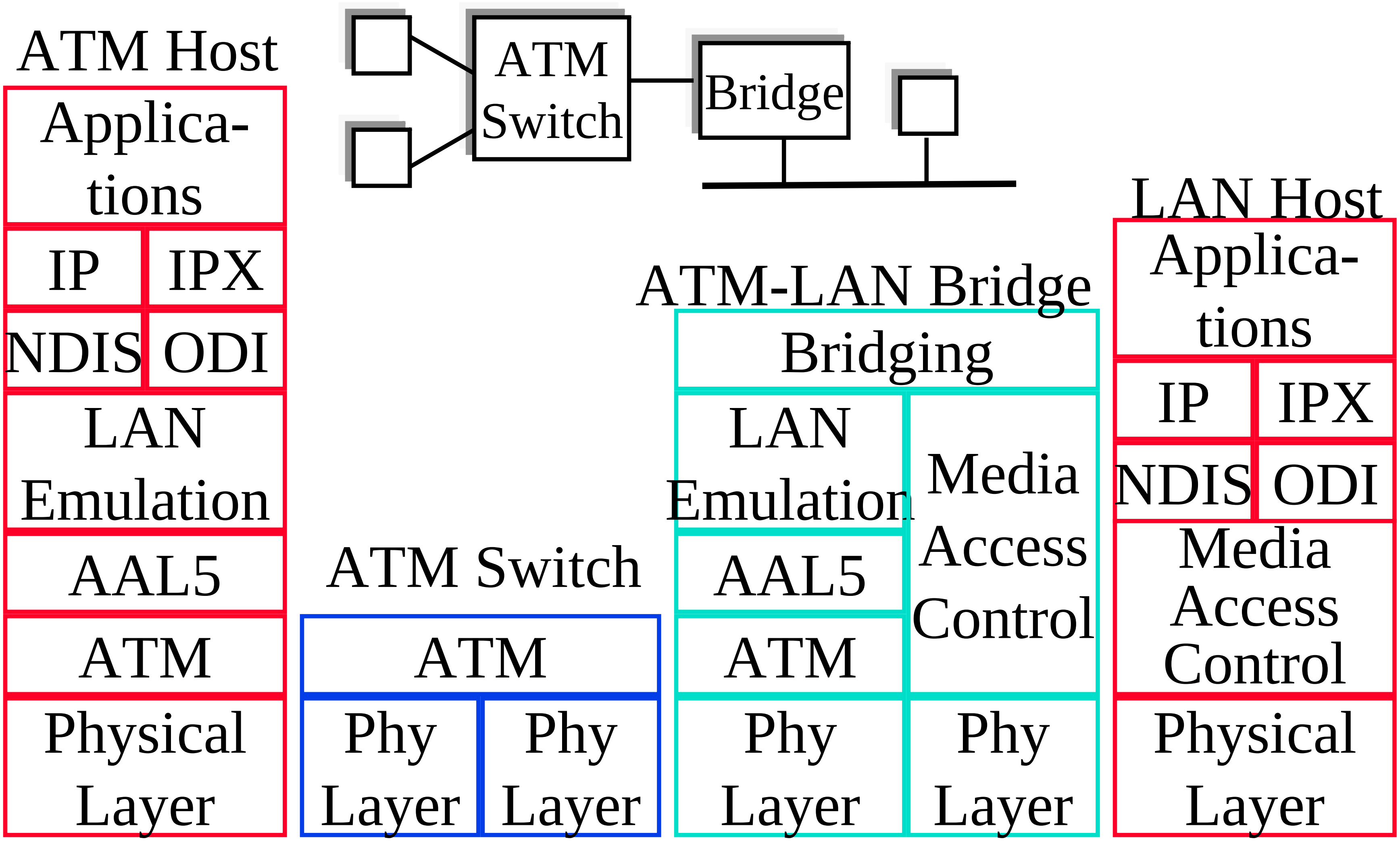
- Problem: Need new networking s/w for ATM
- Solution: Let ATM network appear as a virtual LAN
- LAN emulation implemented as a device driver below the network layer

Features

- One ATM LAN can be *n* virtual LANs
- Logical subnets interconnected via routers
- Need drivers in hosts to support each LAN
- Only IEEE 802.3 and IEEE 802.5 frame formats supported. (FDDI can be easily done.)
- Doesn't allow passive monitoring
- No token management (SMT), collisions, beacon frames.
- Allows larger frames.

LE Header (2 Bytes)	IEEE 802.3 or 802.5 Frame

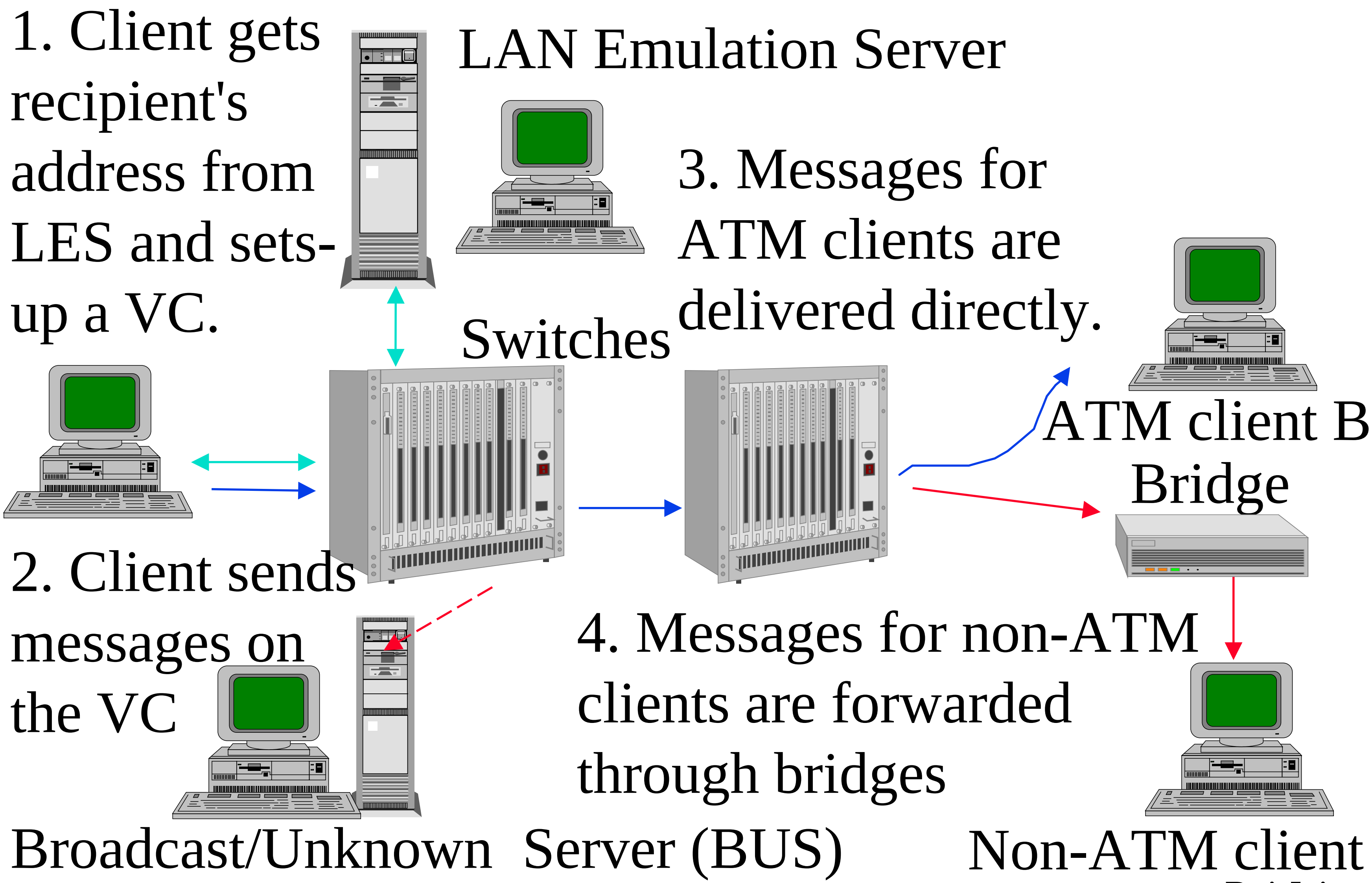
Protocol Layers



Protocol Layers (Cont)

- NDIS = Network Driver Interface Specification
- ODI = Open Datalink Interface
- IPX = NetWare Internetworking Protocol
- **LAN Emulation Software**:
    - LAN Emulation Clients in each host
    - LAN Emulation Servers
        - LAN Emulation Configuration server (LECS)
        - LAN Emulation Server (LES)
        - Broadcast and unknown server (BUS)

LAN Emulation



Operation

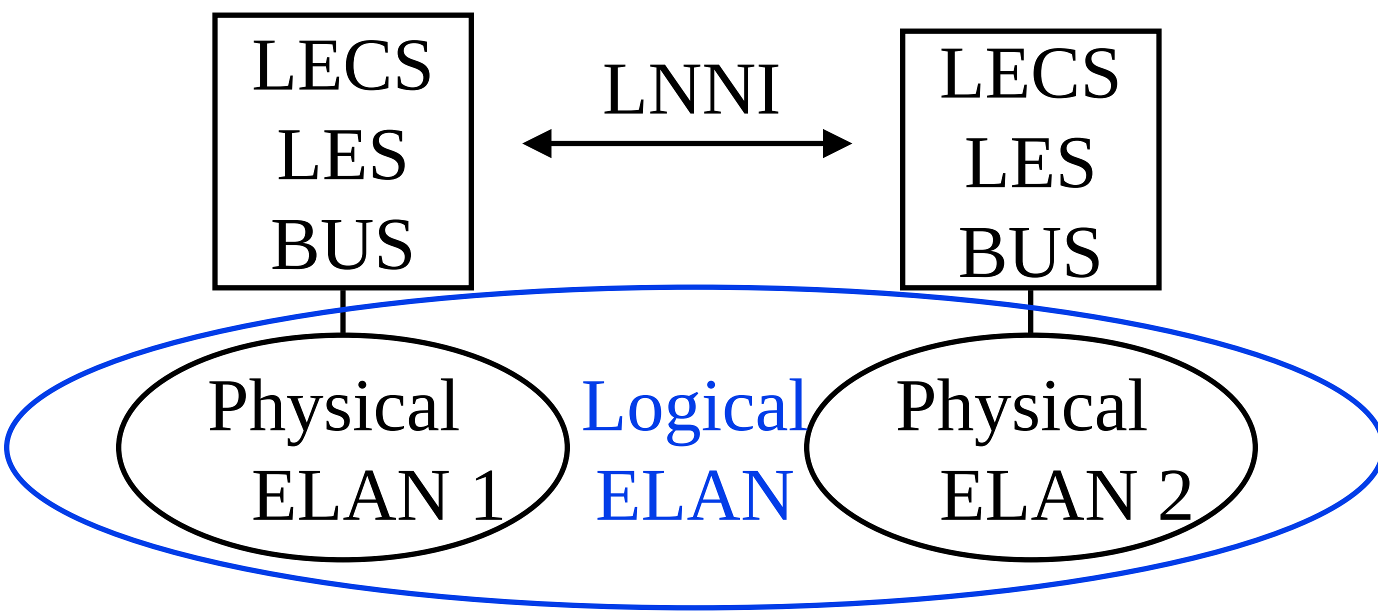
- Initialization:
  - Client gets address of LAN Emulation Configuration Server (LECS) from its switch, uses well-known LECS address, or well known LECS PVC
  - Client gets Server's address from LECS
- Registration:
  - Client sends a list of its MAC addresses to Server.
  - Declares whether it wants ARP requests.

Operation (Cont)

- Address Resolution:
  - Client sends ARP request to Server.
  - Unresolved requests sent to clients, bridges.
  - Server, Clients, Bridges answer ARP
  - Client setups a direct connection
- Broadcast/Unknown Server (BUS):
  - Forwards multicast traffic to all members
  - Clients can also send unicast frames for unknown addresses

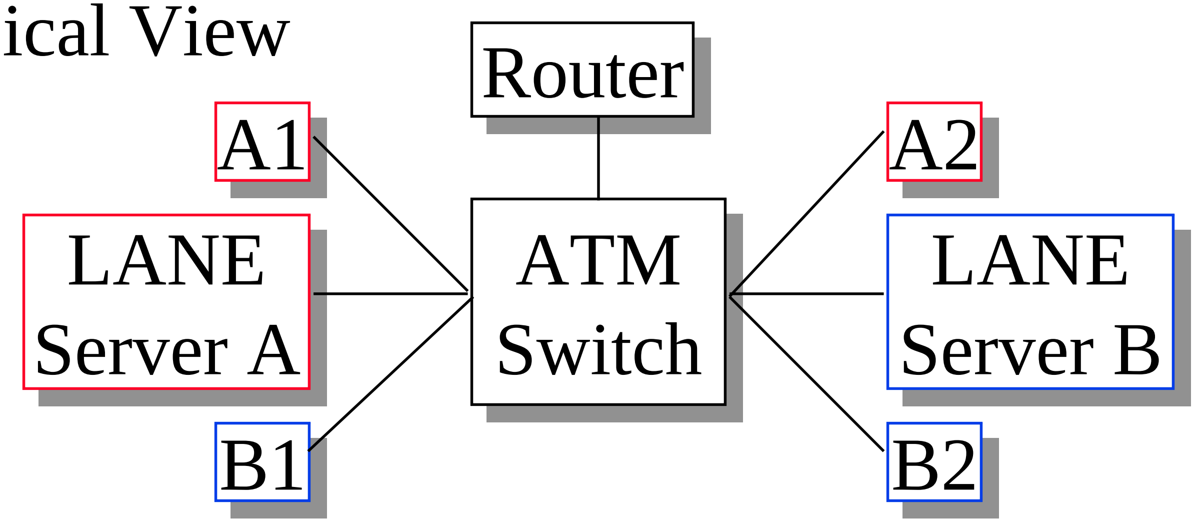
LANE v2.0

- Allows multiple LE Servers:
  LES, BUS, and LECS on a single ELAN
- LAN Emulation network-to-network interface (LNNI): Specifies interfaces for communication between the LE server entities.

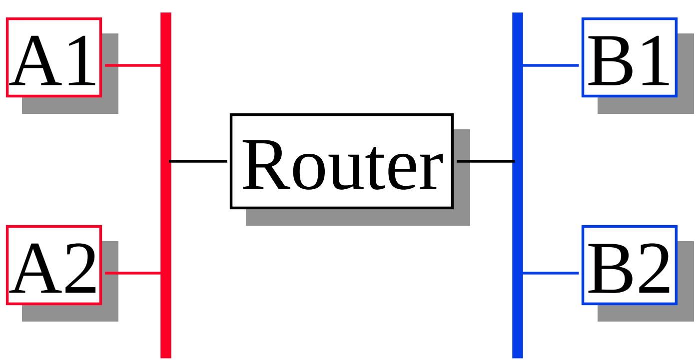


ATM Virtual LANs

- Physical View

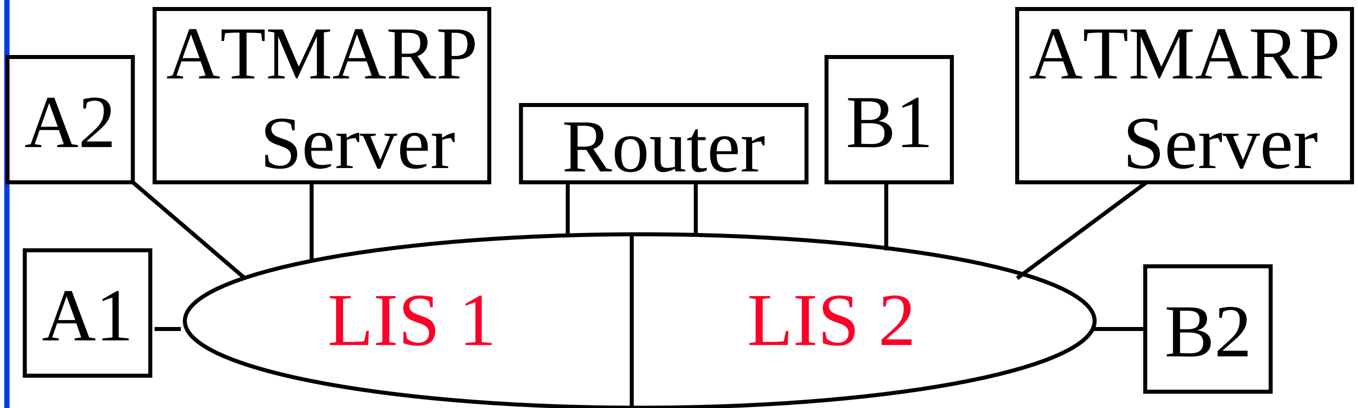


- Logical View



Raj Jain

Classical IP Over ATM



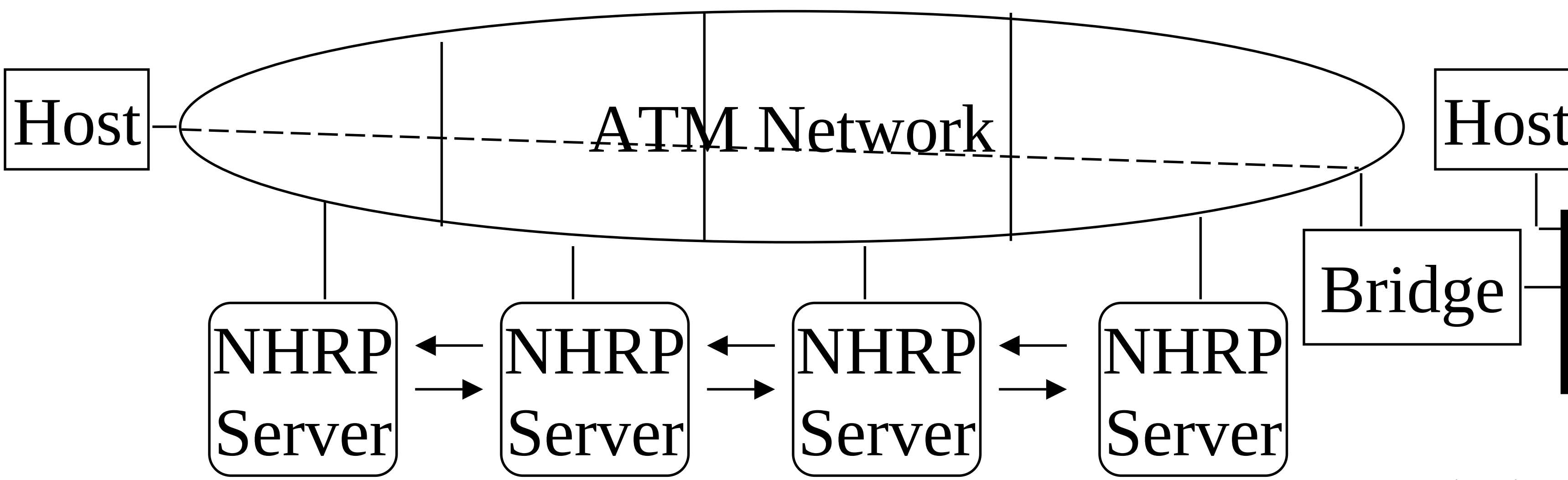
- ATM stations are divided in to Logical IP Subnets (LIS)
- ATMARP server translates IP addresses to ATM addresses.
- Each LIS has an ATMARP server for resolution
- IP stations set up a direct VC with the destination or the router and exchange packets.

IP Multicast over ATM

- Multicast Address Resolution Servers (MARS)
- Internet Group Multicast Protocol (IGMP)
- Multicast group members send IGMP join/leave messages to MARS
- Hosts wishing to send a multicast send a resolution request to MARS
- MARS returns the list of addresses
- MARS distributes membership update information to all cluster members

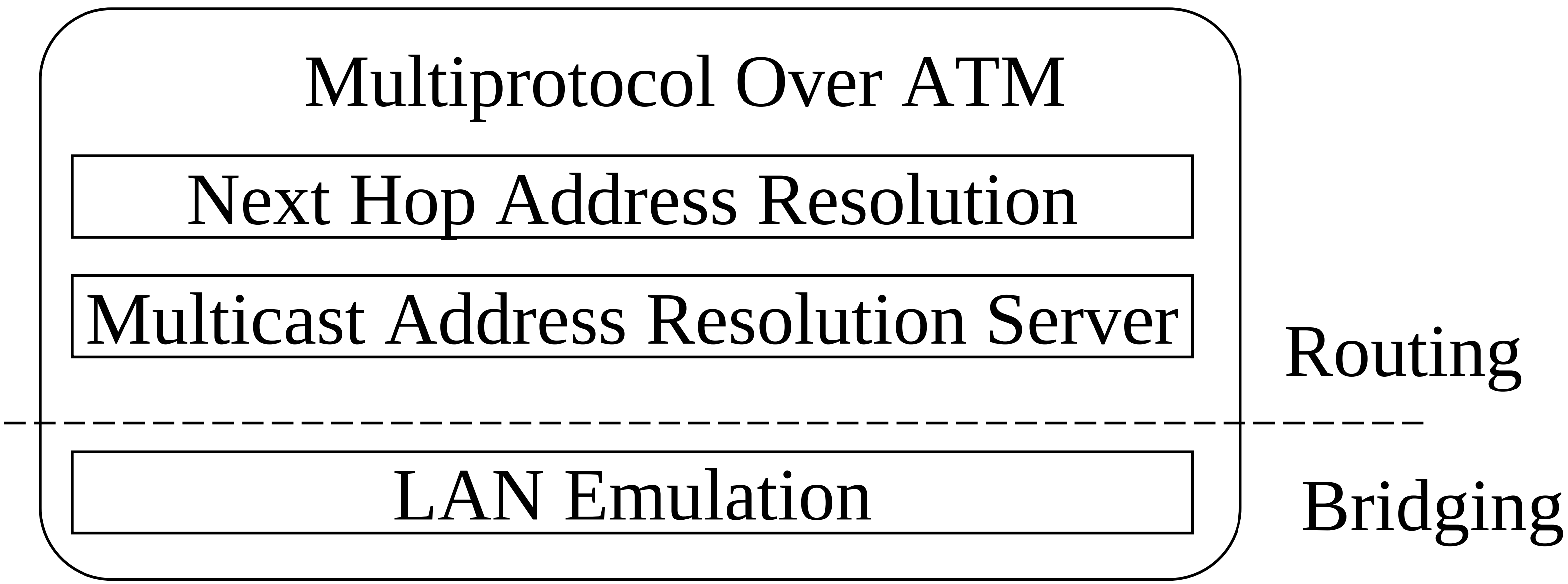
Next Hop Resolution Protocol

- Routers assemble packets ⇒ 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.



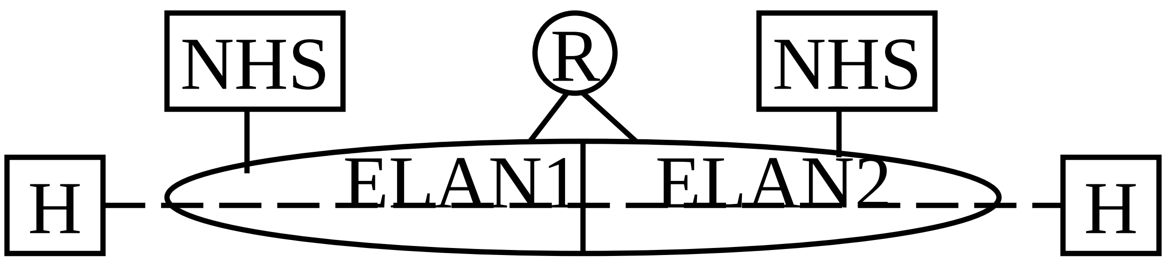
Multiprotocol Over ATM

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



MPOA (Cont)

- LANE operates at layer 2
- RFC 1577 operates at layer 3
- MPOA operates at both layer 2 and layer 3
  - MPOA can handle non-routable as well as routable protocols
- Layer 3 protocol runs directly over ATM
  - Can use ATM QoS
- MPOA uses LANE for its layer 2 forwarding



Key References

- See **http://www.cis.ohio-state.edu/~jain/refs/atm_refs.htm**
- "NBMA Next Hop Resolution Protocol (NHRP)", http://www.internic.net/internet-drafts/draft-ietf-rolc-nhrp-14.txt, 2/6/98.
- RFC 1577, "Classical IP and ARP over ATM," 1/20/94, http://ds.internic.net/rfc/rfc1577.txt
- LAN Emulation over ATM v1.0 Specification (Jan 1995), ftp://ftp.atmforum.com/pub/approved-specs/af-lane-0021.000.ps

IP Switching and Alternatives

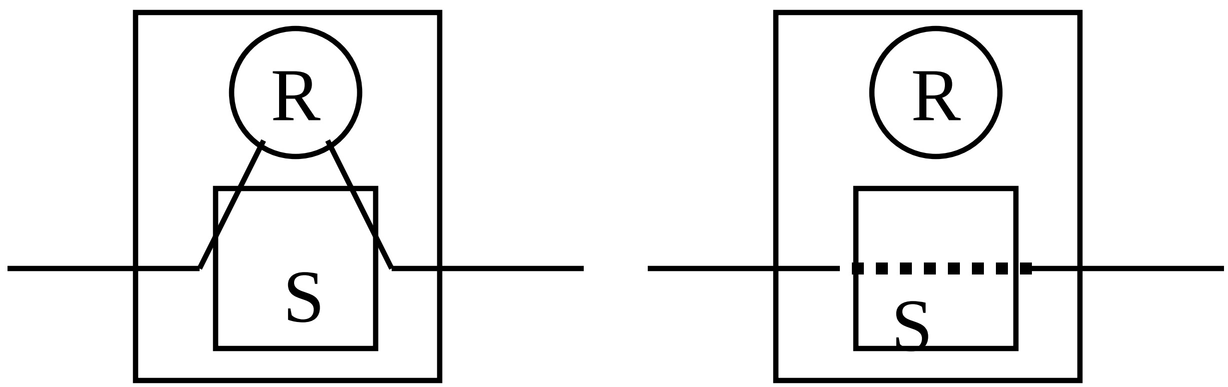
Raj Jain
Professor of Computer and Information Sciences
The Ohio State University

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

- IP Switch
- Cell Switched Router
- Tag Switching (CISCO)
- ARIS (IBM)
- Multi-protocol label switching

IP Switching

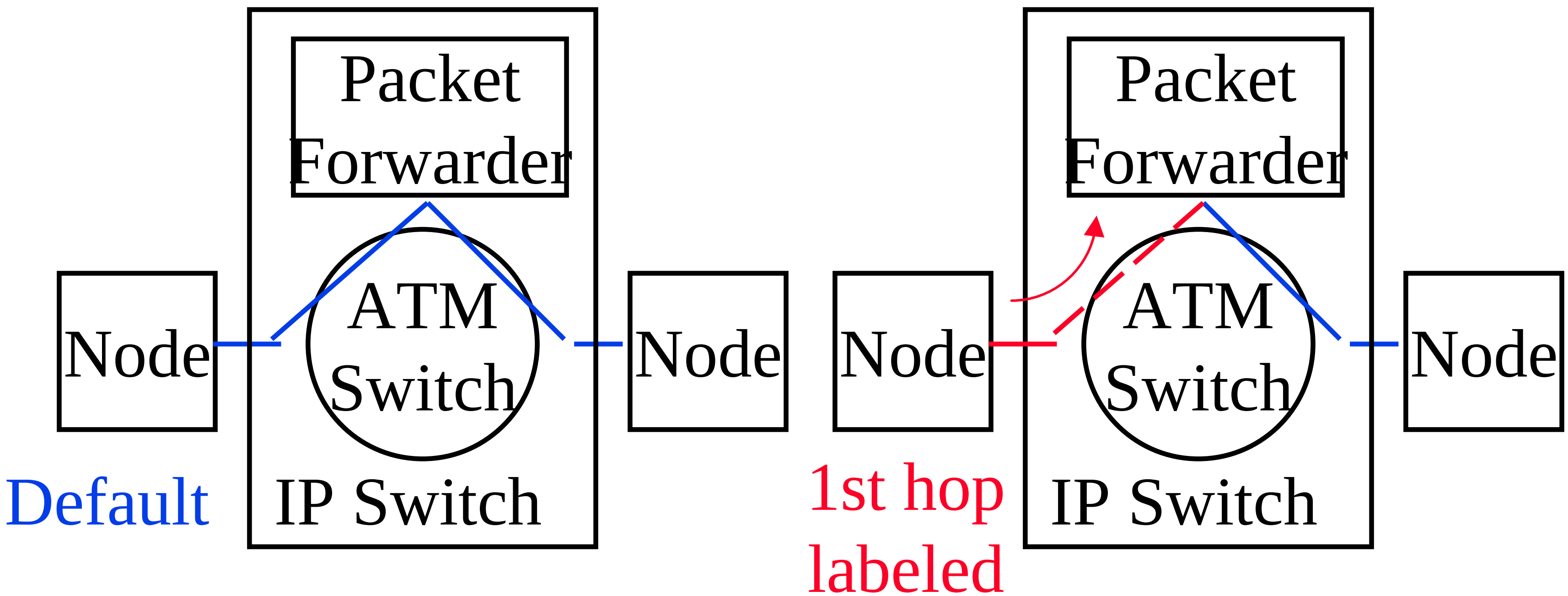
- Developed by Ipsilon
- Routing software in every ATM switch in the network
- Initially, packets are reassembled by the routing software and forwarded to the next hop
- Long term flows are transferred to separate VCs. Mapping of VCIs in the switch ⇒ No reassembly



Raj Jain

IP Switching: Steps 1-2

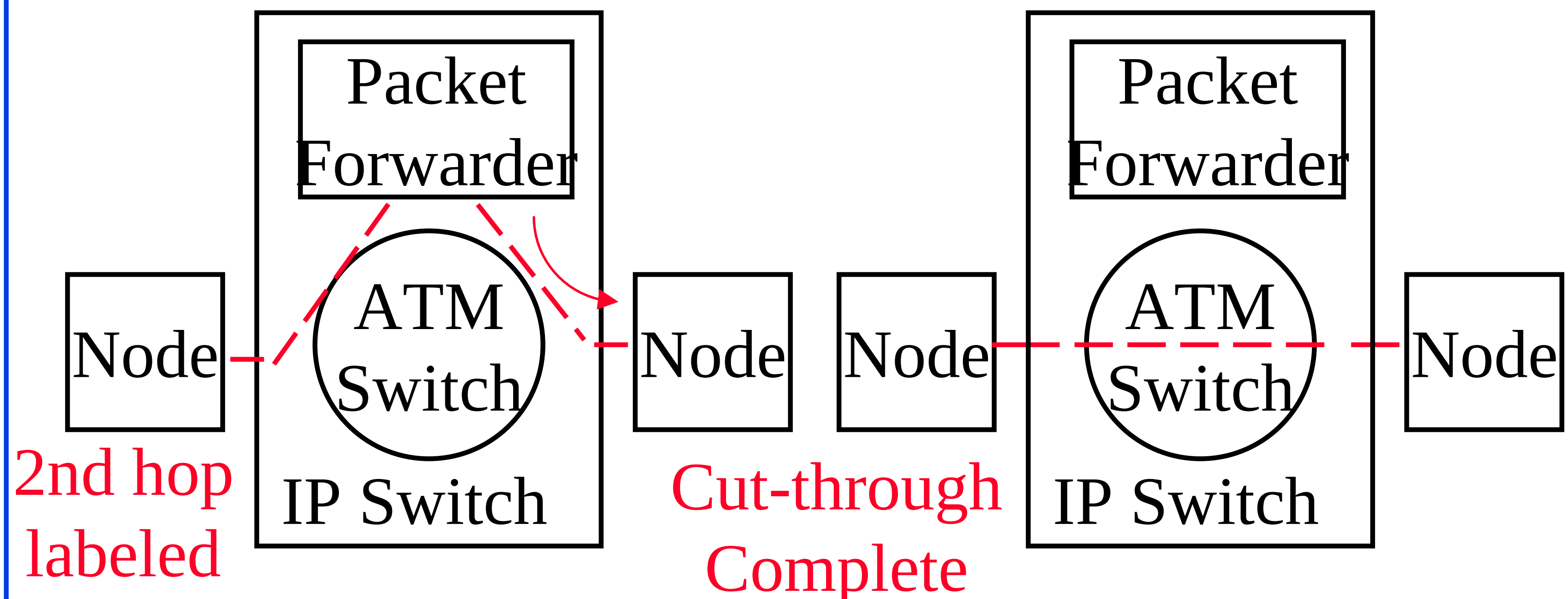
- If a flow is deemed to be "flow oriented", the node asks the upstream node to set up a separate VC.
- Downstream nodes may also ask for a new VC.



Raj Jain

IP Switching: Steps 3, 4

- After both sides of a flow have separate VCs, the router tells the switch to register the mapping for cut-through



Raj Jain

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 ⇒ pt-mpt VC per source
- Routing bypassed ⇒ Firewall bypassed
  - Solution: IP fields are deleted before segmentation and added after assembly ⇒ First packet has to go through firewall.
- Initially IP only. IPX supported via tunneling in IP.

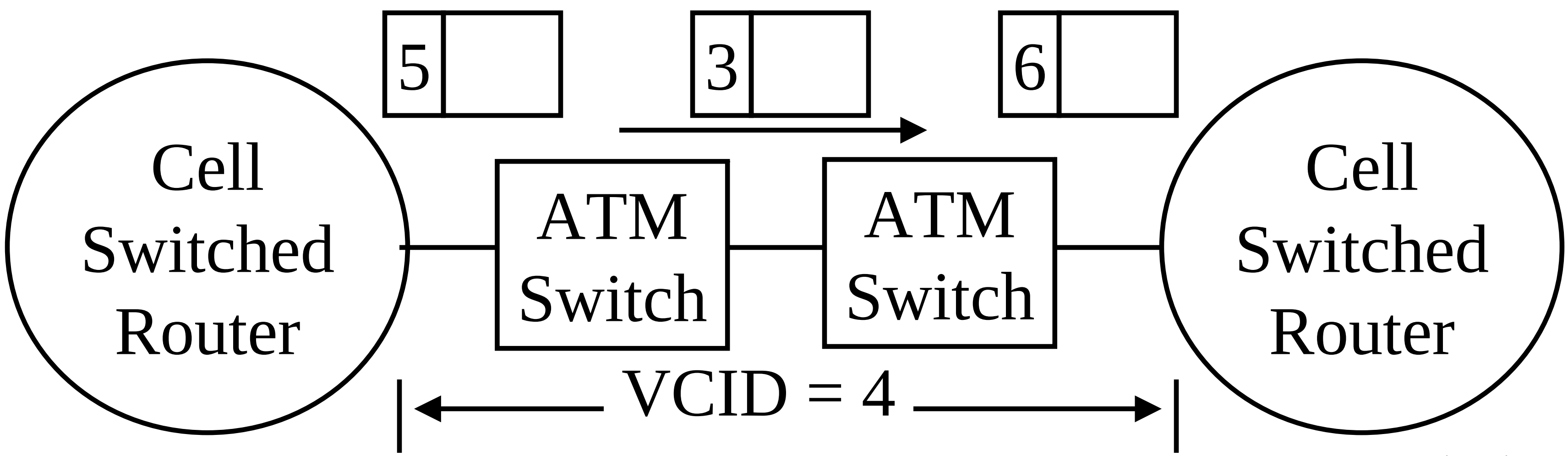
Raj Jain

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 ⇒ Switches between CSRs
- Upstream assigns a VCID and sends downstream



Raj Jain

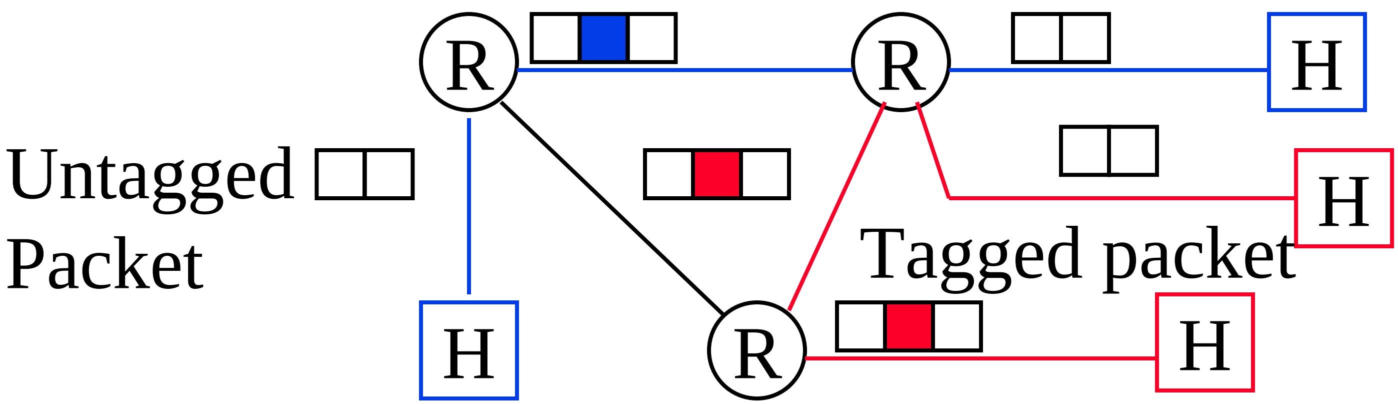
CSR (Cont)

- 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.



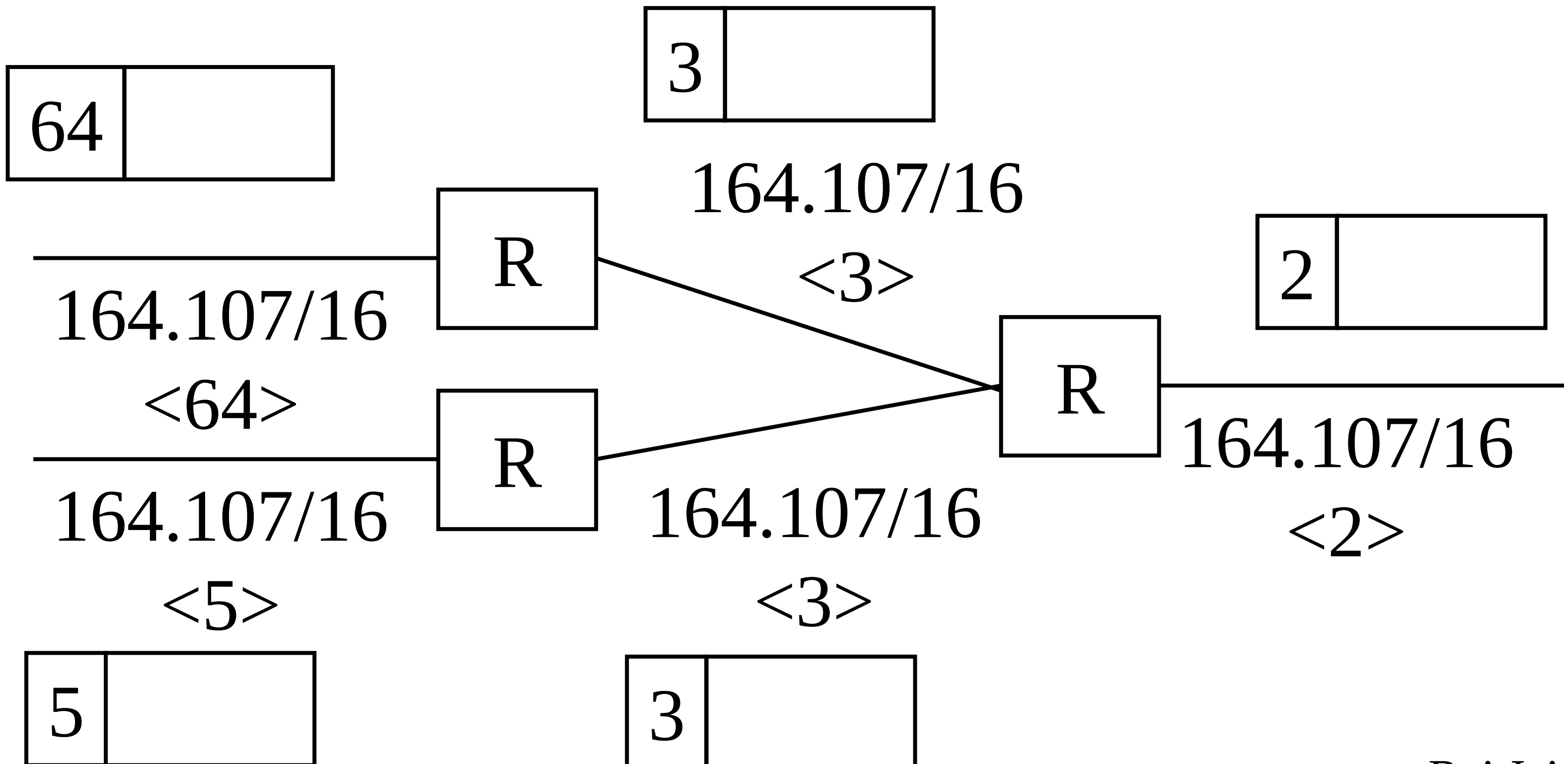
Raj Jain

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)

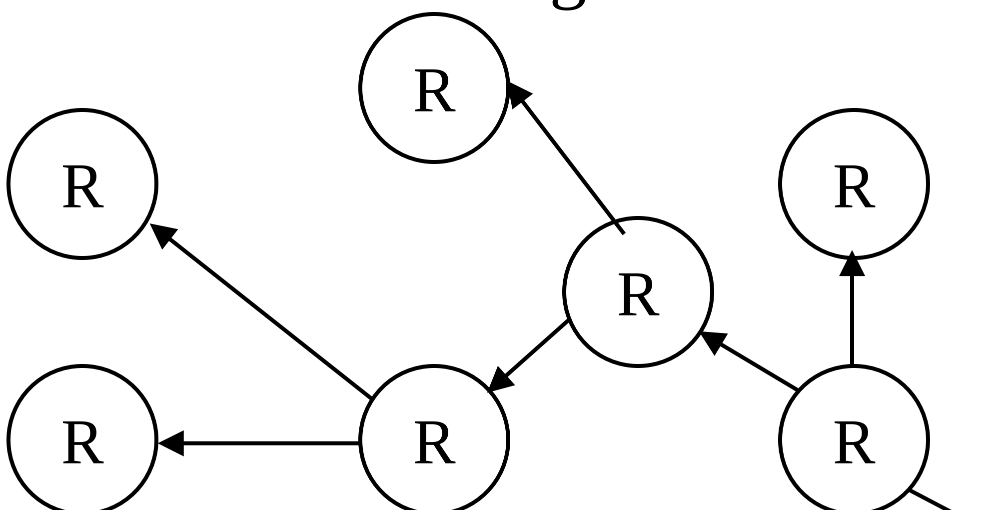
- One VC per routing table entry



Raj Jain

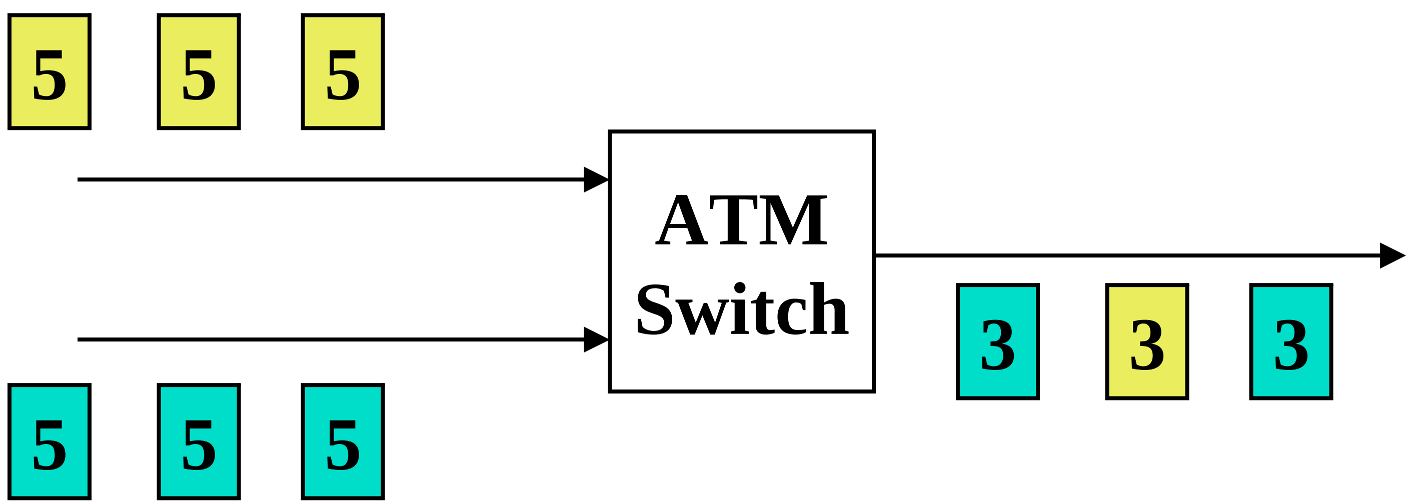
ARIS

- 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

Issue	IP Switch	CSR	Tag	ARIS
Datalink	ATM	ATM, FR	ATM, FR, Ethernet	ATM, FR Ethernet
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

Raj Jain

MPLS

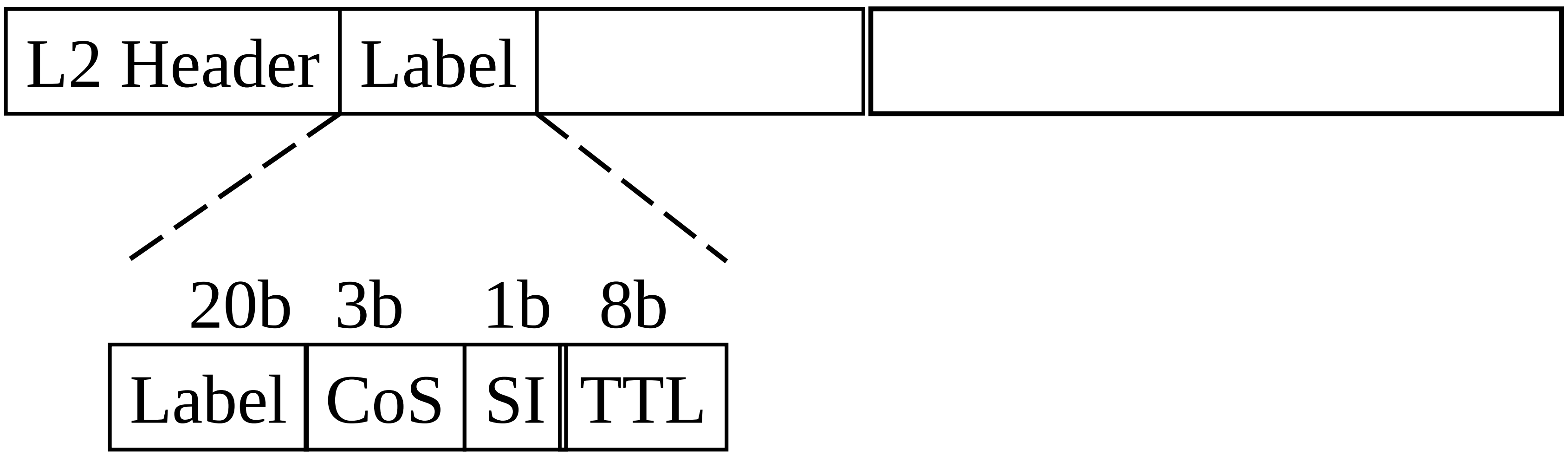
- Multiprotocol Label Switching
- IETF working group to develop switched IP forwarding
- Initially focused on IPv4 and IPv6. Technology extendible to other L3 protocols.
- Not specific to ATM. ATM or LAN.
- Not specific to a routing protocol (OSPF, RIP, ...)
- Optimization only. Labels do not affect the path. Only speed. Networks continue to work w/o labels

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
  $\Rightarrow$ 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



Raj Jain

Label Stacks

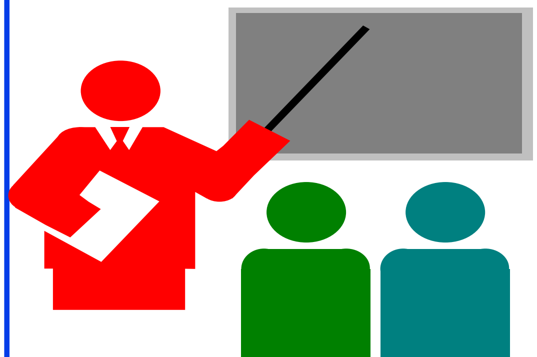
- Labels are pushed/popped as they enter/leave MPLS domain
- Routers in the interior will use Interior Gateway Protocol (IGP) labels. Border gateway protocol (BGP) labels outside.

L2 Header	Label 1	Label 2	…	Label n	

Raj Jain

MPLS: Issues

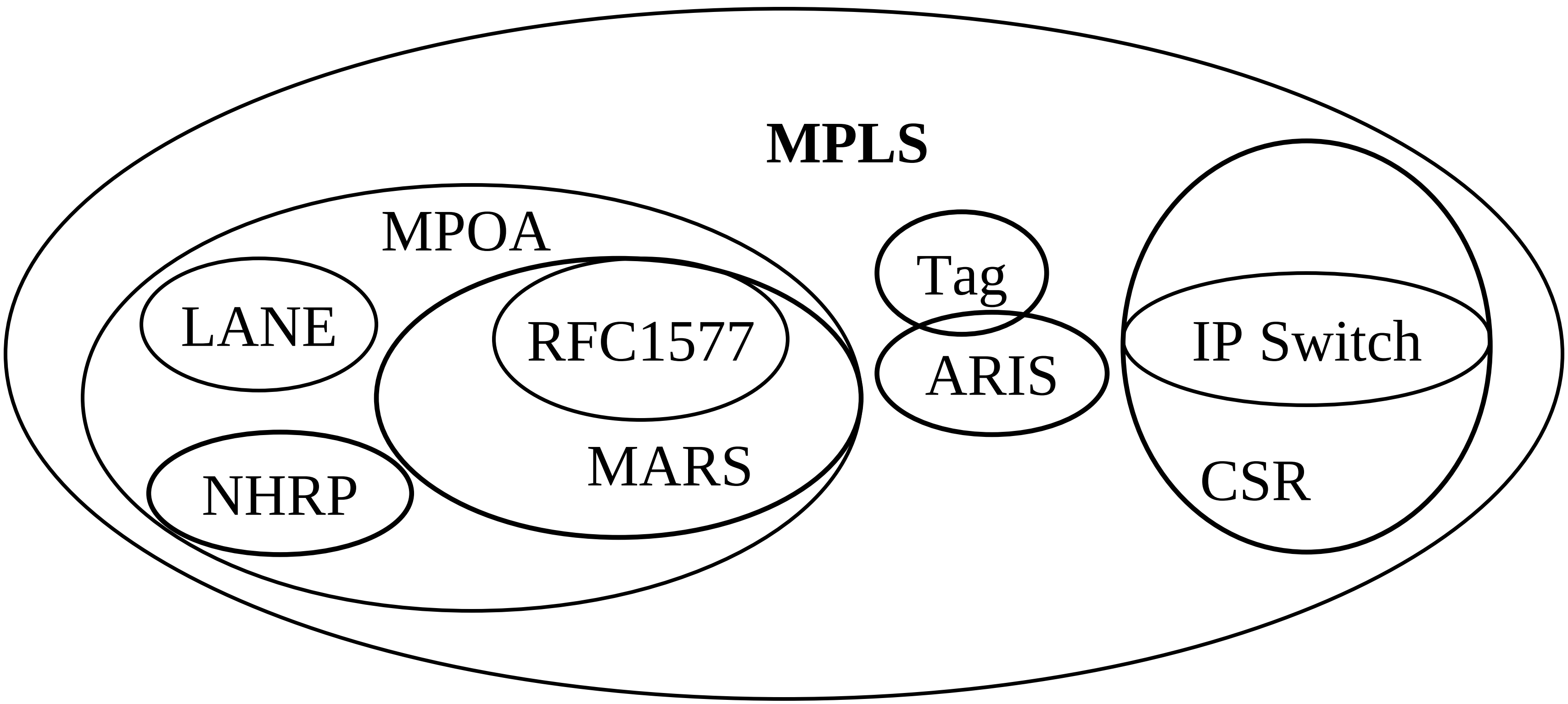
- Loop prevention, detection, survival
- Multicast:
  Multiple entries in label information base
- Multipath: Streams going to the same destination but different sources/port # may be assigned separate labels.
- Host involvement: Label-enabled hosts will avoid first hop reassembly
- Security: Label swapping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



Raj Jain

Key References

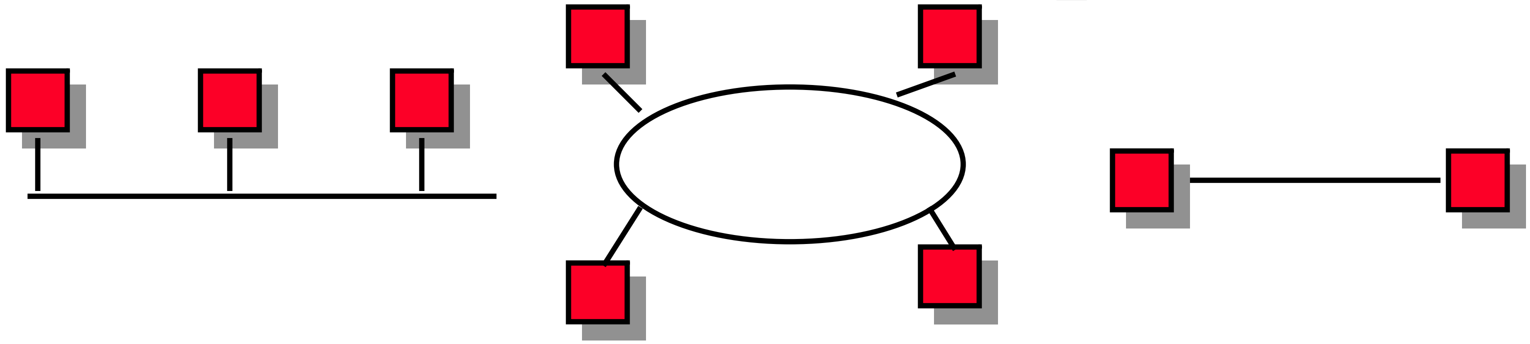
- See **http://www.cis.ohio-state.edu/~jain/refs/atm_refs.htm**
- "A Framework for Multiprotocol Label Switching", 11/26/1997, http://www.internic.net/internet-drafts/draft-ietf-mpls-framework-02.txt
- Multiprotocol Label Switching (mpls) working group at IETF. Email: mpls-request@cisco.com
- ATM Forum, "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

Gigabit Ethernet

Raj Jain
Professor of Computer and Information Sciences
The Ohio State University
Columbus, OH 43210
http://www.cis.ohio-state.edu/~jain/

- Distance-Bandwidth Principle
- 10 Mbps to 100 Mbps
- Gigabit PHY Issues
- Gigabit MAC Issues
- Status
- ATM vs Gigabit Ethernet

Distance-B/W Principle

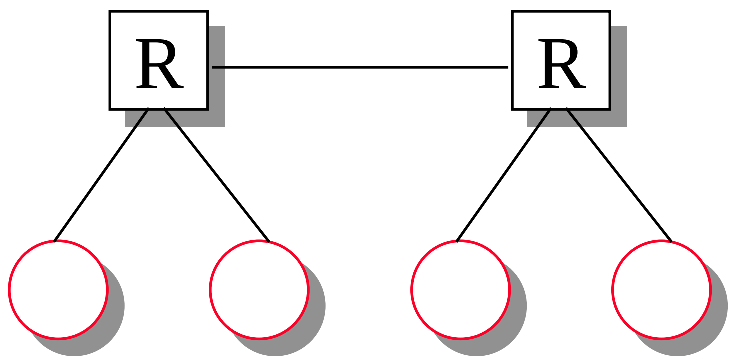


- Efficiency = Max throughput/Media bandwidth
- Efficiency is a decreasing function of α
  α = Propagation delay /Transmission time
  = (Distance/Speed of light)/(Transmission size/Bits/sec)
  = Distance×Bits/sec/(Speed of light)(Transmission size)
- Bit rate-distance-transmission size tradeoff.
- 100 Mb/s ⇒ Change distance or frame size

Raj Jain

Ethernet vs Fast Ethernet

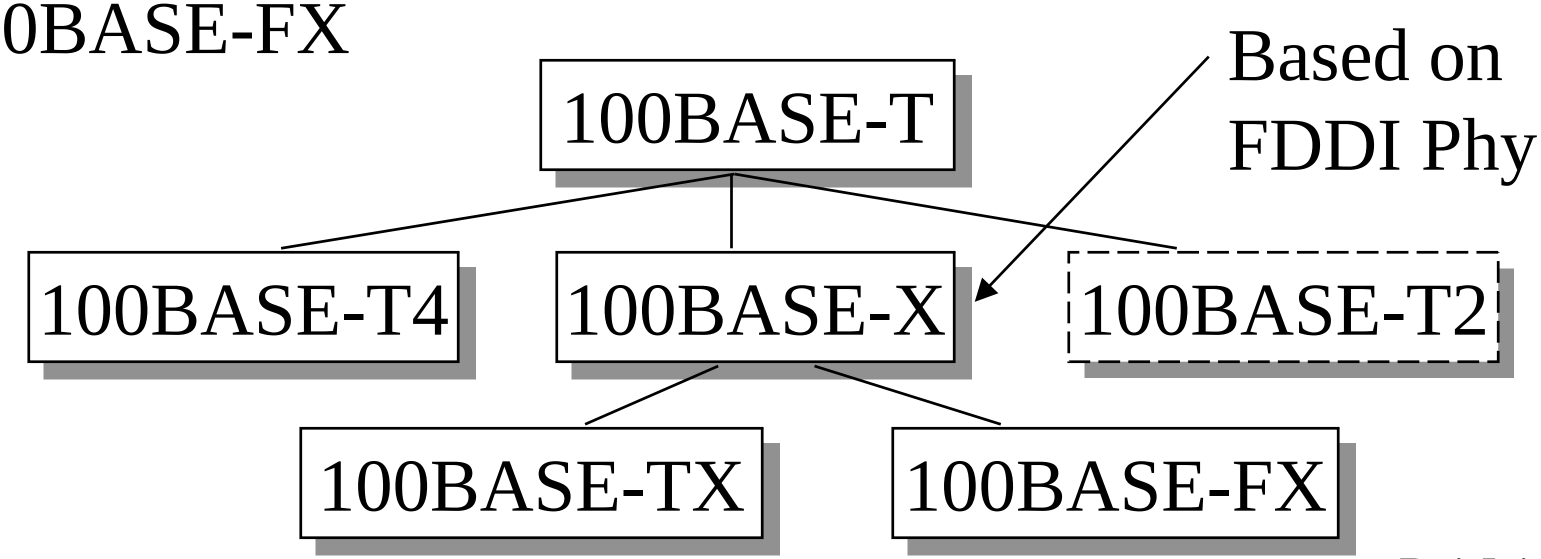
	Ethernet	Fast Ethernet
Speed	10 Mbps	100 Mbps
MAC	CSMA/CD	CSMA/CD
Network diameter	2.5 km	205 m
Topology	Bus, star	Star
Cable	Coax, UTP, Fiber	UTP, Fiber
Standard	802.3	802.3u
Cost	X	2X



Raj Jain

Fast Ethernet Standards

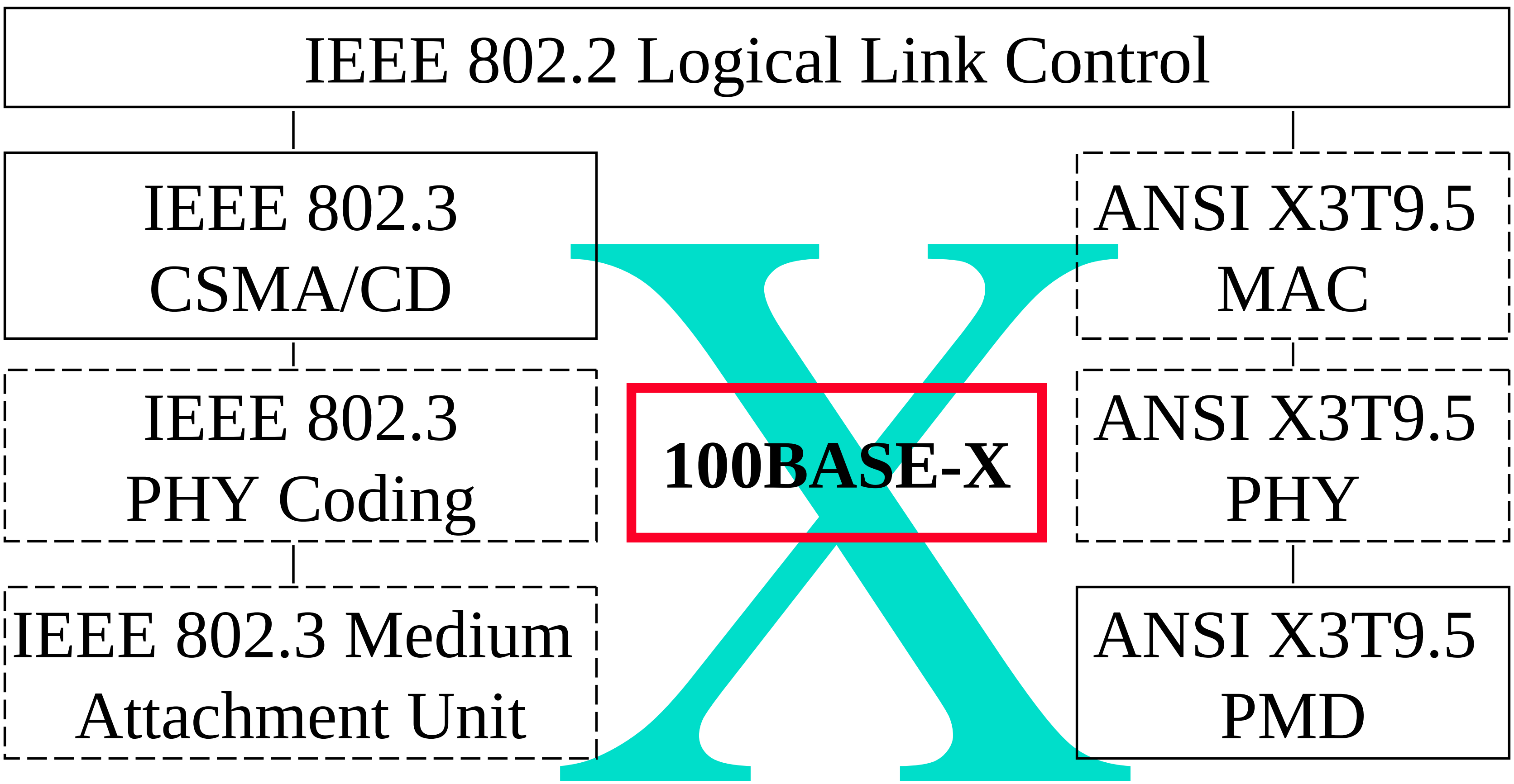
- **100BASE-T4:** 100 Mb/s over 4 pairs of CAT-3, 4, 5
- **100BASE-TX:** 100 Mb/s over 2 pairs of CAT-5, STP
- **100BASE-FX:** 100 Mbps CSMA/CD over 2 fibers
- **100BASE-X:** 100BASE-TX or 100BASE-FX
- **100BASE-T:** 100BASE-T4, 100BASE-TX, or 100BASE-FX



Raj Jain

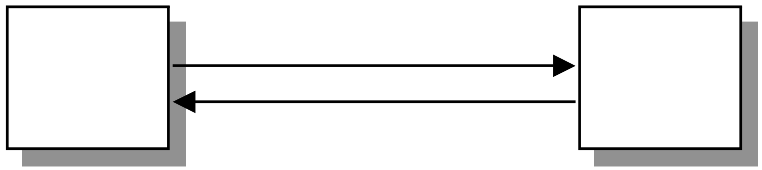
100 BASE-X

- X = Cross between IEEE 802.3 and ANSI X3T9.5



Raj Jain

Full-Duplex Ethernet



- Uses point-to-point links between TWO nodes
- Full-duplex bi-directional transmission
- Transmit any time
- Many vendors are shipping switch/bridge/NICs with full duplex
- No collisions $\Rightarrow$ 50+ Km on fiber.
- Between servers and switches or between switches

Gigabit Ethernet

- Being standardized by 802.3z
- Project approved by IEEE in June 1996
- 802.3 meets every three months ⇒ Too slow
  ⇒ Gigabit Ethernet Alliance (GEA) formed.
  It meets every two weeks.
- Decisions made at GEA are formalized at 802.3 High-Speed Study Group (HSSG)
- Based on Fiber Channel PHY
- Shared (half-duplex) and full-duplex version
- Gigabit 802.12 and 802.3 to have the same PHY

How Much is a Gbps?

- 622,000,000 bps = OC-12
- 800,000,000 bps (100 MBps Fiber Channel)
- 1,000,000,000 bps
- 1,073,741,800 bps = 2^{30} bps (2^{10} = 1024 = 1k)
- 1,244,000,000 bps = OC-24
- 800 Mbps ⇒ Fiber Channel PHY
  ⇒ Shorter time to market
- Decision: 1,000,000,000 bps ⇒ 1.25 GBaud PHY
- Not multiple speed ⇒ Sub-gigabit Ethernet rejected
- 1000Base-X

Physical Media

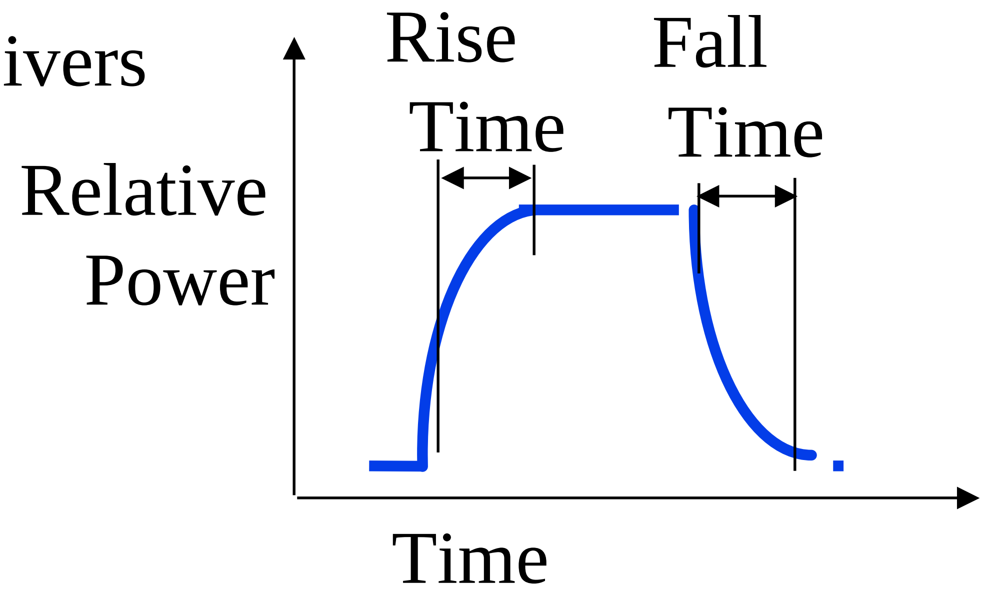
- Unshielded Twisted Pair (UTP-5): 4-pairs
- Shielded Twisted Pair (STP)
- Multimode Fiber: 50 μm and 62.5 μm
  - Use CD lasers
- Single-Mode Fiber
- Bit Error Rate better than 10^{-12}

How Far Should It Go?

- Full-Duplex:
  - Fiber Channel: 300 m on 62.5 μm at 800 Mbps ⇒ 230 m at 1000 Mbps
  - Decision: 500 m at 1000 Mbps ⇒ Minor changes to FC PHY
- Shared:
  - CSMA/CD without any changes ⇒ 20 m at 1 Gb/s (Too small)
  - Decision: 200 m shared ⇒ Minor changes to 802.3 MAC

PHY Issues

- Fiber Channel PHY:
  100 MBps = 800 Mbps
  ⇒ 1.063 GBaud using 8b10b
- Changes to get 500 m on 62.5-μm multimode fiber
  - Modest decrease in rise and fall times of the transceivers



Raj Jain

- Symbol Codes for Specific Signals: Jam, End-of-packet, beginning of packet
- PHY-based flow Control: No.
  Use the XON/XOFF flow control of 802.3x

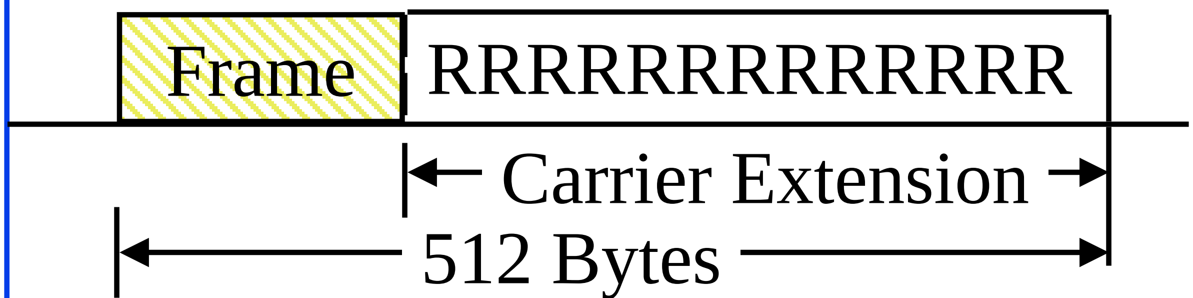
850 nm vs 1300 nm lasers

- 850 nm used in 10Base-F
  - Cannot go full distance with 62.5-μm fiber
  - 500 m with 50-μm fiber
  - 250 m with 62.5-μm fiber
- 1300 nm used in FDDI but more expensive
  - Higher eye safety limits
  - Better Reliability
  - Start with 550 m on 62.5-μm fiber
  - Could be improved to 2 km on 62.5-μm fiber
    $\Rightarrow$ Needed for campus backbone

Media Access Control Issues

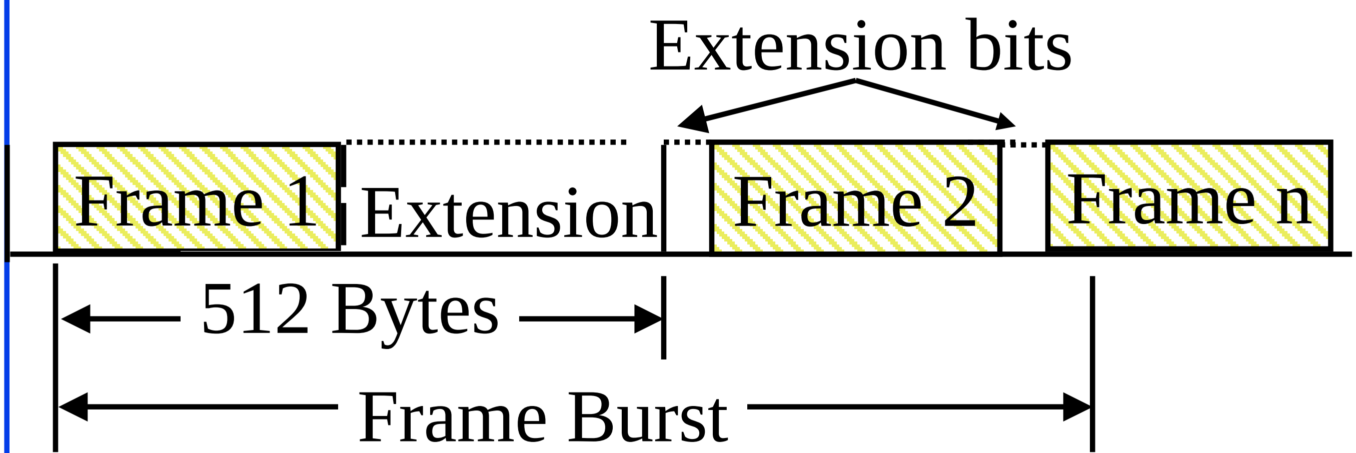
- Carrier Extension
- Frame Bursting
- Buffered Distributor

Carrier Extension



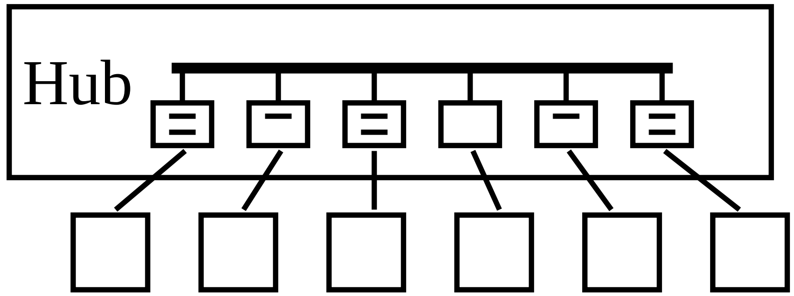
- 10 Mbps at 2.5 km ⇒ Slot time = 64 bytes
- 1 Gbps at 200 m ⇒ Slot time = 512 bytes
- Continue transmitting control symbols.
  Collision window includes the control symbols
- Control symbols are discarded at the destination
- Net throughput for small frames is only marginally better than 100 Mbps

Frame Bursting

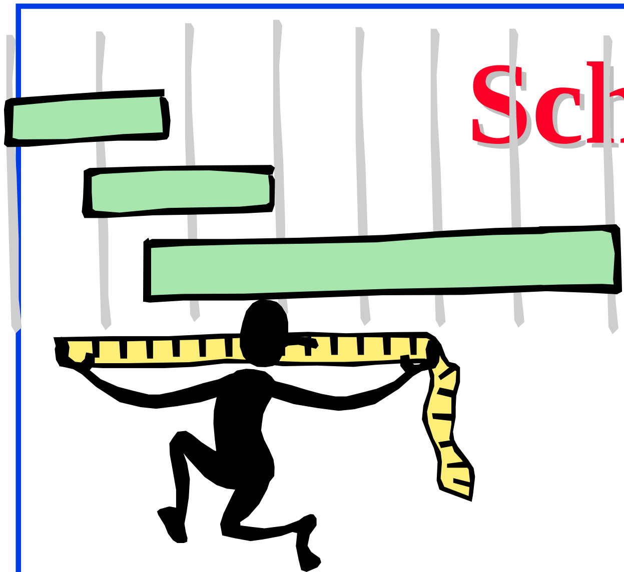


- Don't give up the channel after every frame
- After the slot time, continue transmitting additional frames (with minimum inter-frame gap)
- Interframe gaps are filled with extension bits
- No no new frame transmissions after 8192 bytes
- Three times more throughput for small frames

Buffered Distributor



- All incoming frames are buffered in FIFOs
- CSMA/CD arbitration inside the box to transfer frames from an incoming FIFO to all outgoing FIFOs
- Previous slides were half-duplex. With buffered distributor all links are full-duplex with frame-based flow control
- Link length limited by physical considerations only



Schedule

- November 1996: Proposal cutoff
- January 1997: First draft
- March 1997: Second draft
- July 1997: Working Group Ballot
- March 1998: Approval

Status

- On Schedule
- First draft reviewed in January 97
- Fourth draft was issued in December'97
- 1000Base-X: Gigabit Ethernet based on Fiber Channel Phy
- Phy modified for 1000 Mbps operation
- Phy modified for ISO 11801 standard for premises cabling $\Rightarrow$ 550 m intra-building backbone runs
  $\Rightarrow$ 1300-nm lasers on 62.5-μm multimode fiber
  850-nm lasers on 62.5-μm fiber ok for 300 m

1000Base-X

- 1000Base-LX: 1300-nm laser transceivers
  - 2 to 550 m on 62.5-μm or 50-μm multimode, 2 to 3000 m on 10-μm single-mode
- 1000Base-SX: 850-nm laser transceivers
  - 2 to 300 m on 62.5-μm, 2 to 550 m on 50-μm. Both multimode.
- 1000Base-CX: Short-haul copper jumpers
  - 25 m 2-pair shielded twinax cable in a single room or rack.

    Uses 8b/10b coding ⇒ 1.25 Gbps line rate

1000Base-T

- 100 m on 4-pair Cat-5 UTP
  - Network diameter of 200 m
  - Requires new coding schemes
  - Under development.
  - New PAR approved in March 1997
  - 802.3ab task force

Design Parameter Summary

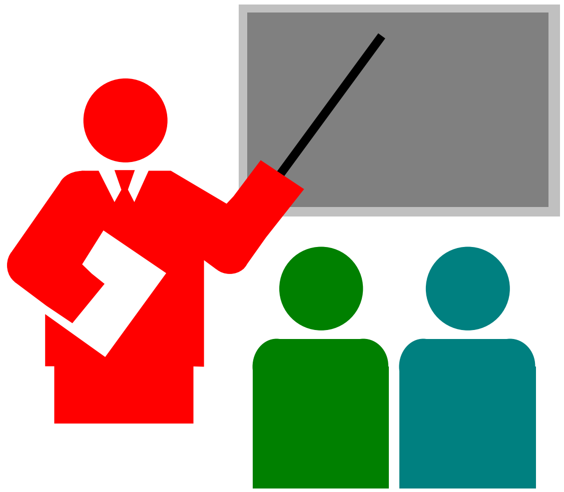
Parameter	10 Mbps	100 Mbps	1 Gbps
Slot time	512 bt	512 bt	4096 bt
Inter Frame Gap	9.6 μs	0.96 μs	0.096 μs
Jam Size	32 bits	32 bits	32 bits
Max Frame Size	1518 B	1518 B	1518 B
Min Frame Size	64 B	64 B	64 B
Burst Limit	N/A	N/A	8192 B

- bt = bit time

ATM vs Gb Ethernet

Issue	ATM	Gigabit Ethernet
Media	SM Fiber, MM Fiber, UTP5	Mostly fiber
Max Distance	Many miles using SONET	260-550 m
Data Applications	Need LANE, IPOA	No changes needed
Interoperability	Good	Limited
Ease of Mgmt	LANE	802.1Q VLANs
QoS	PNNI	802.1p (Priority)
Signaling	UNI	None/RSVP (?)
Traffic Mgmt	Sophisticated	802.3x Xon/Xoff

Summary



- Ethernet will run at 1000 Mbps
- Will compete with ATM for campus backbone and desktop
- Both shared and full-duplex links
- Fully compatible with current Ethernet

References

- For a detailed list of references, see http://www.cis.ohio-state.edu/~jain/refs/gbe_refs.htm

- "Media Access Control (MAC) Parameters, Physical Layer Repeater and Management Parameters for 1000 Mb/s Operation," IEEE Draft P802.3z/D4.2, March 17, 1998.

References (Cont)

- Email Reflector:
  stds-802-3-hssg@mail.ieee.org
  - To join send email to majordomo@mail.ieee.org
  - subscribe stds-802-3-hssg@mail.ieee.org <your email address>
- FTP Site:
  ftp://stdsbbs.ieee.org/pub/802_main/802.3/gigabit
- Gigabit Ethernet Consortium
  http://www.gigabit-ethernet.org

Final Review: 13 Hot Facts

1. Networking is critical and growing exponentially.
2. Networking is the key to productivity
3. LAN Emulation allows current LAN applications to run on ATM
4. Classical IP allows address resolution using LIS servers
5. NHRP allows shortcuts between ATM hosts
6. MARS allows multicast address resolution.

7. MPOA combines LANE, NHRP, and MARS and reduces the need for routers
8. IP switching allows some IP packets to go through an ATM network without reassembly at intermediate routers.
9. To succeed, ATM has to solve today's problem (data) at a price competitive to LANs.
10. 100 Mbps Ethernet limited to 200 m to desktop. Not limited in full-duplex mode.
11. Gigabit Ethernet will compete with ATM for campus backbone and desktop

12. Gigabit Ethernet will support both shared and full-duplex links

13. Most gigabit Ethernet links will be full-duplex