Network Layer in Practice:
IP and ATM

Eytan Modiano
Massachusetts Institute of Technology
Laboratory for Information and Decision Systems

The TCP/IP Protocol Suite

- **Transmission Control Protocol / Internet Protocol**
- **Developed by DARPA to connect Universities and Research Labs**

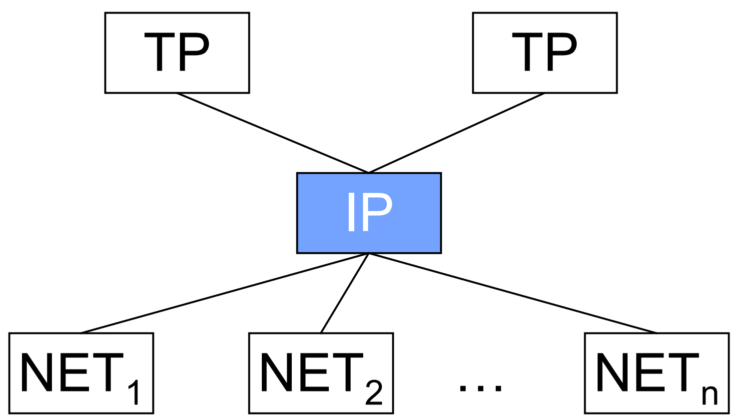
Four Layer model

Layer	Examples
Applications	Telnet, FTP, email, etc.
Transport	TCP, UDP
Network	IP, ICMP, IGMP
Link	Device drivers, interface cards

TCP - Transmission Control Protocol
UDP - User Datagram Protocol
IP - Internet Protocol

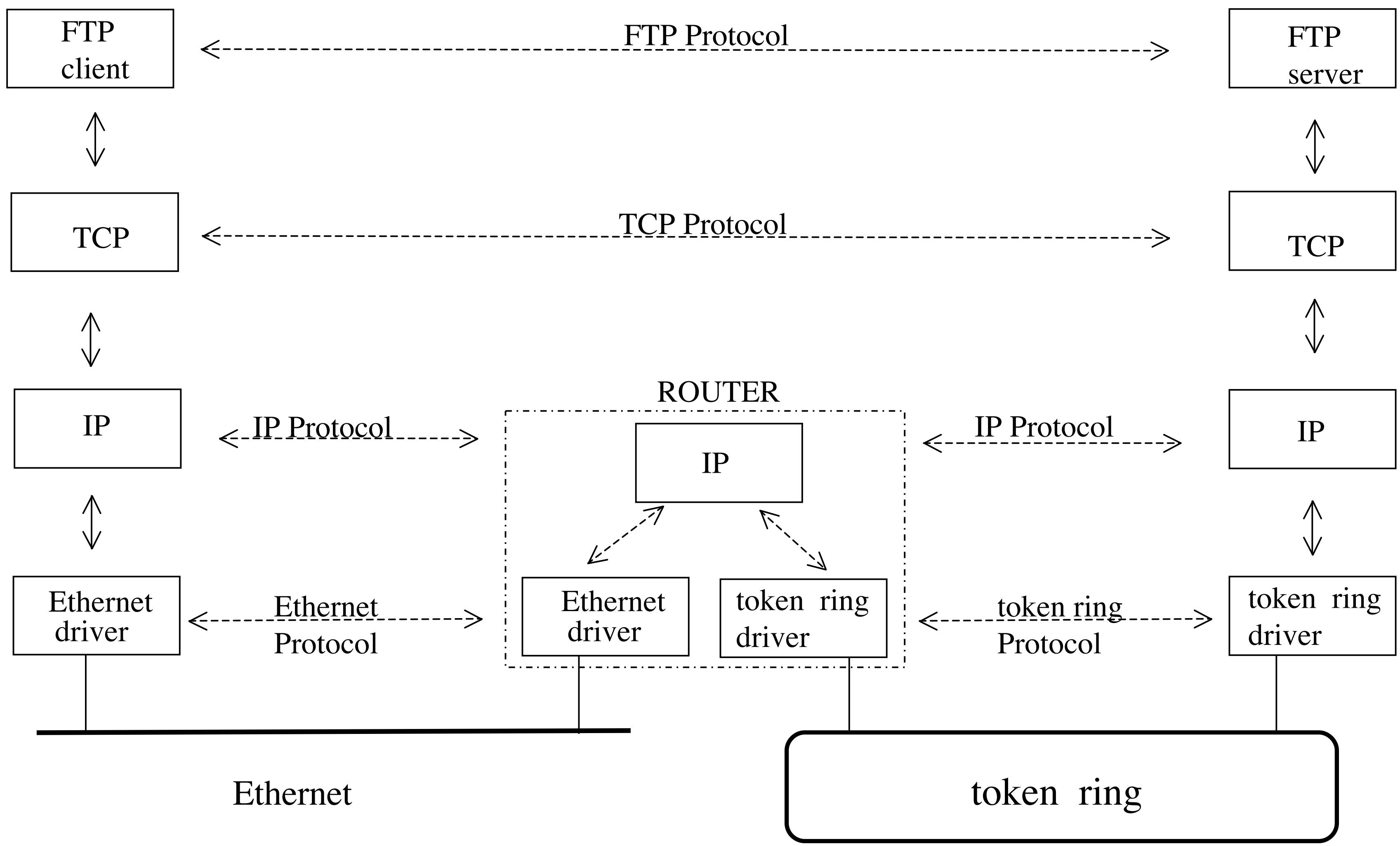
Internet Sub-layer

- **A sub-layer between the transport and network layers is required when various incompatible networks are joined together**
  - **This sub-layer is used at gateways between the different networks**
  - **In the internet this function is accomplished using the Internet Protocol (IP)**

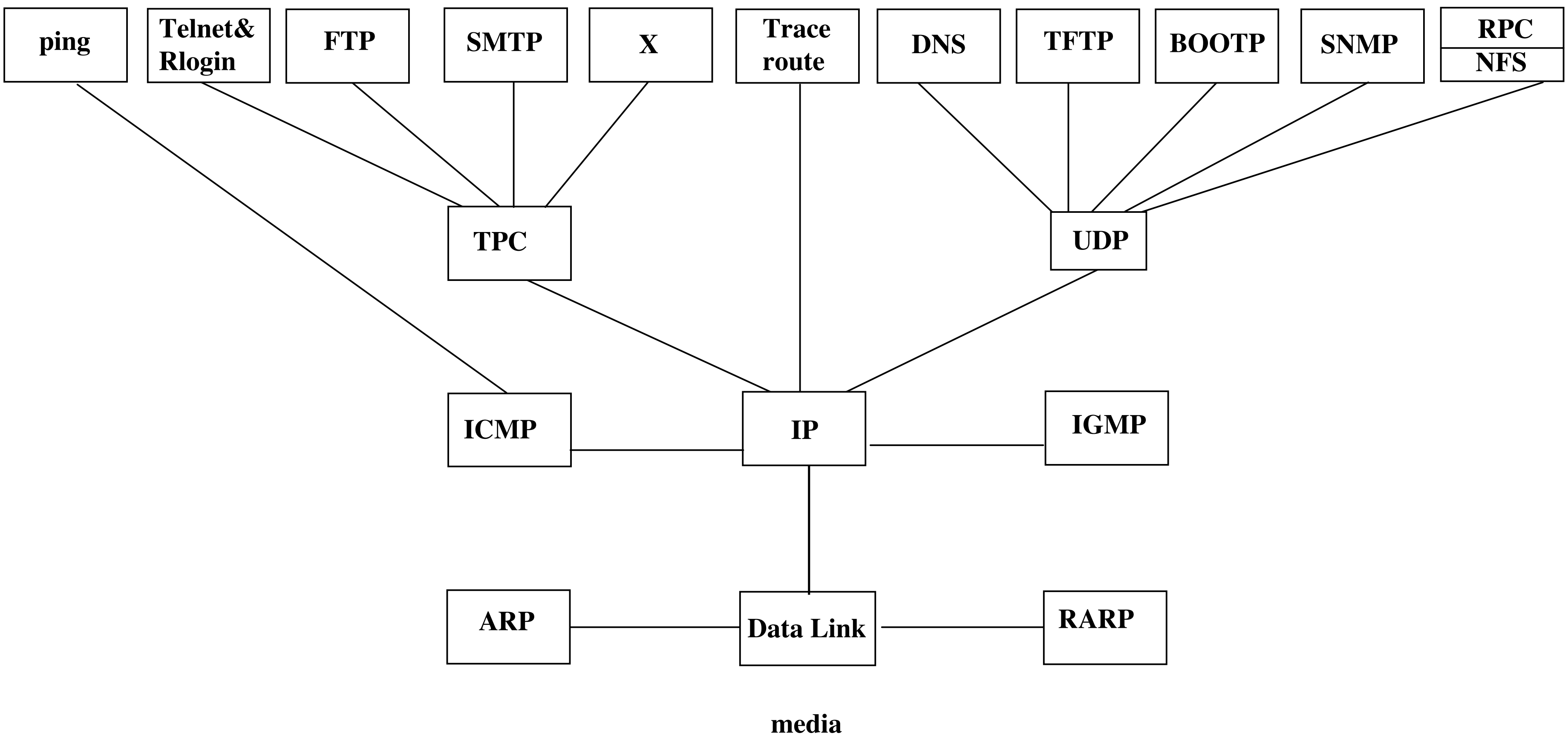


IP enables interoperability

Internetworking with TCP/IP

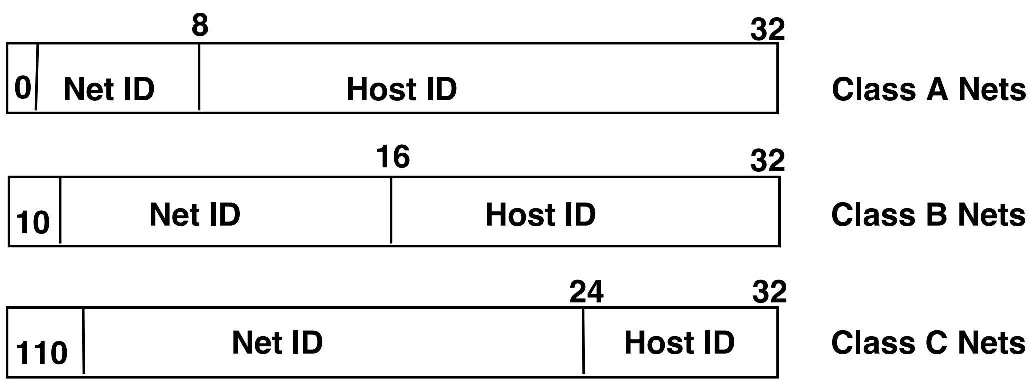


The TCP/IP Suite



IP addresses

- 32 bit address written as four decimal numbers
  – One per byte of address (e.g., 155.34.60.112)
- Hierarchical address structure
  – Network ID/ Host ID/ Port ID
  – Complete address called a socket
  – Network and host ID carried in IP Header
  – Port ID (sending process) carried in TCP header
- IP Address classes:



Class D is for multicast traffic

Host Names

- Each machine also has a unique name
- Domain name System: A distributed database that provides a mapping between IP addresses and Host names
- E.g., 155.34.50.112 => plymouth.ll.mit.edu

Internet Standards

- Internet Engineering Task Force (IETF)
  - Development on near term internet standards
  - Open body
  - Meets 3 times a year
- Request for Comments (RFCs)
  - Official internet standards
  - Available from IETF web page: http://www.ietf.org

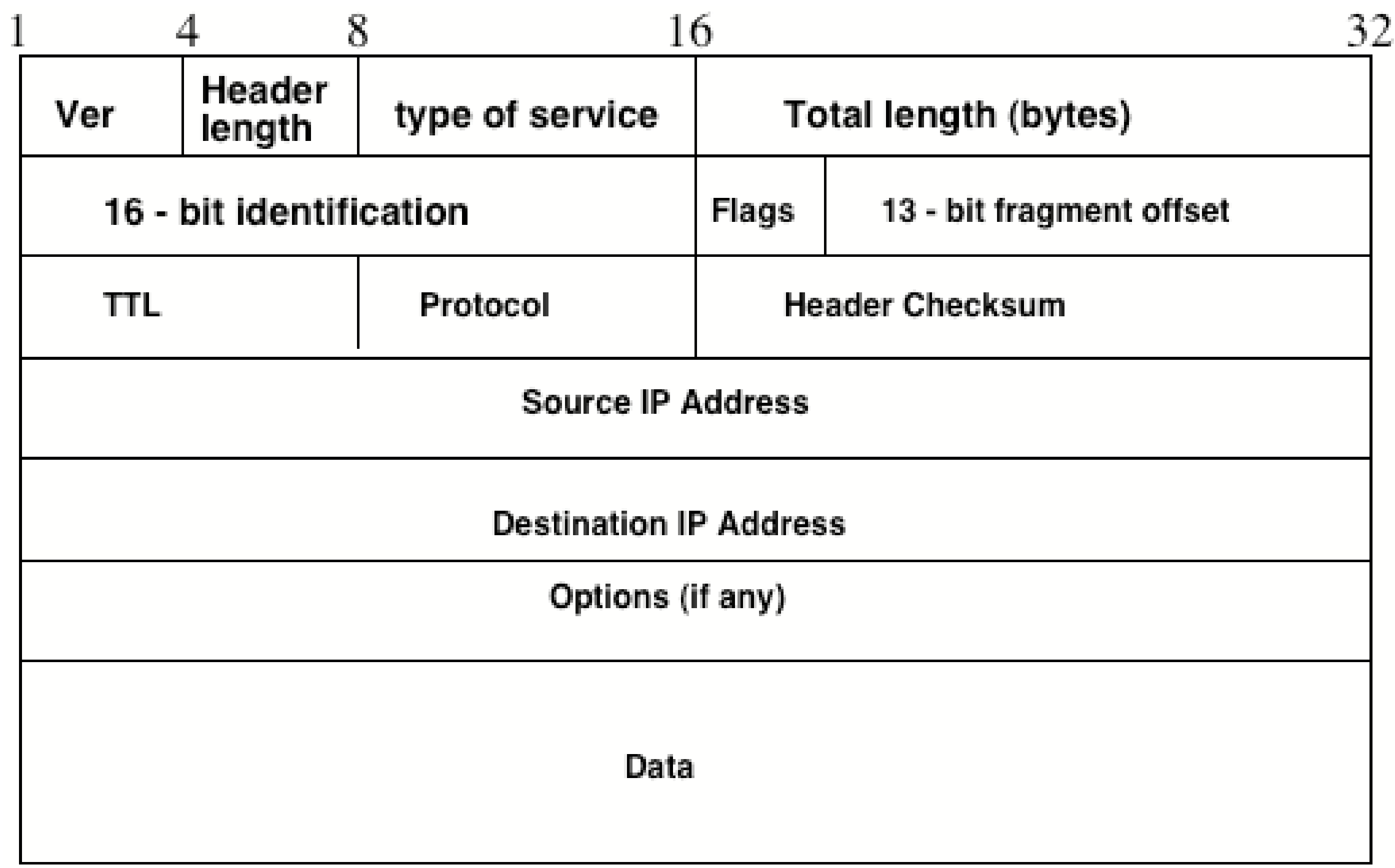
The Internet Protocol (IP)

- Routing of packet across the network
- Unreliable service
  - Best effort delivery
  - Recovery from lost packets must be done at higher layers
- Connectionless
  - Packets are delivered (routed) independently
  - Can be delivered out of order
  - Re-sequencing must be done at higher layers

- Current version V4

- Future V6
  - Add more addresses (40 byte header!)
  - Ability to provide QoS

Header Fields in IP

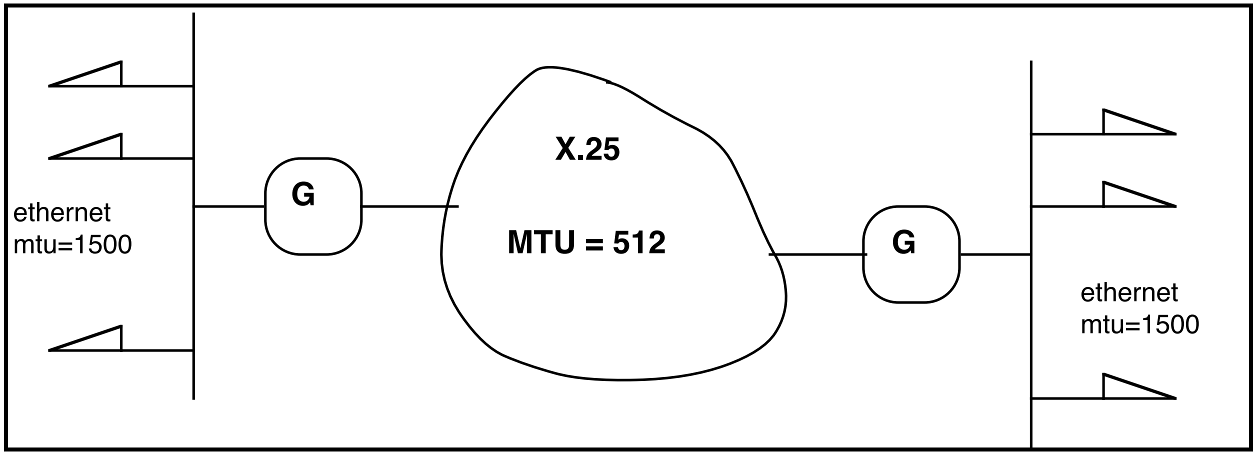


Note that the minimum size header is 20 bytes; TCP also has 20 byte header

IP HEADER FIELDS

- **Vers: Version # of IP (current version is 4)**
- **HL: Header Length in 32-bit words**
- **Service: Mostly Ignored**
- **Total length Length of IP datagram**
- **ID Unique datagram ID**
- **Flags: NoFrag, More**
- **FragOffset: Fragment offset in units of 8 Octets**
- **TTL: Time to Live in "seconds" or Hops**
- **Protocol: Higher Layer Protocol ID #**
- **HDR Cksum: 16 bit 1's complement checksum (on header only!)**
- **SA & DA: Network Addresses**
- **Options: Record Route,Source Route,TimeStamp**

FRAGMENTATION



- **A gateway fragments a datagram if length is too great for next network (fragmentation required because of unknown paths).**
- **Each fragment needs a unique identifier for datagram plus identifier for position within datagram**
- **In IP, the datagram ID is a 16 bit field counting datagram from given host**

POSITION OF FRAGMENT

- **Fragment offset field gives starting position of fragment within datagram in 8 byte increments (13 bit field)**
- **Length field in header gives the total length in bytes (16 bit field)**
  - **Maximum size of IP packet 64K bytes**
- **A flag bit denotes last fragment in datagram**
- **IP reassembles fragments at destination and throws them away if one or more is too late in arriving**

IP Routing

- **Routing table at each node contains for each destination the next hop router to which the packet should be sent**
  - **Not all destination addresses are in the routing table**
    - **Look for net ID of the destination "Prefix match"**
    - **Use default router**
- **Routers do not compute the complete route to the destination but only the next hop router**
- **IP uses distributed routing algorithms: RIP, OSPF**
- **In a LAN, the "host" computer sends the packet to the default router which provides a gateway to the outside world**

Subnet addressing

- **Class A and B addresses allocate too many hosts to a given net**
- **Subnet addressing allows us to divide the host ID space into smaller "sub networks"**
  - **Simplify routing within an organization**
  - **Smaller routing tables**
  - **Potentially allows the allocation of the same class B address to more than one organization**
- **32 bit Subnet "Mask" is used to divide the host ID field into subnets**
  - **"1" denotes a network address field**
  - **"0" denotes a host ID field**

	16 bit net ID	16 bit host ID	
Class B Address	140.252	Subnet ID	Host ID
Mask	111111 111 1111111	11111111	00000000

Classless inter-domain routing (CIDR)

- Class A and B addresses allocate too many hosts to an organization while class C addresses don't allocate enough
  - This leads to inefficient assignment of address space
- Classless routing allows the allocation of addresses outside of class boundaries (within the class C pool of addresses)
  - Allocate a block of contiguous addresses

    E.g., 192.4.16.1 - 192.4.32.155

    Bundles 16 class C addresses

    The first 20 bits of the address field are the same and are essentially the network ID
  - Network numbers must now be described using their length and value (I.e., length of network prefix)

    E.g., 192.4.16.1/20
  - Routing table lookup using longest prefix match

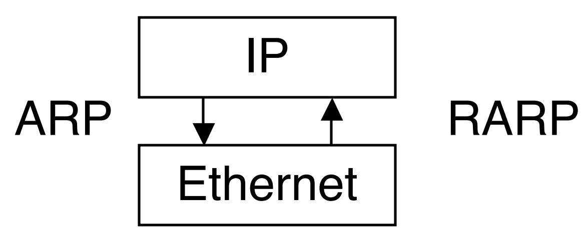
- Notice similarity to subnetting - "supernetting"

Dynamic Host Configuration (DHCP)

- Automated method for assigning network numbers
  - IP addresses, default routers
- Computers contact DHCP server at Boot-up time
- Server assigns IP address
- Allows sharing of address space
  - More efficient use of address space
  - Adds scalability
- Addresses are "least" for some time
  - Not permanently assigned

Address Resolution Protocol

- IP addresses only make sense within IP suite
- Local area networks, such as Ethernet, have their own addressing scheme
  - To talk to a node on LAN one must have its physical address (physical interface cards don't recognize their IP addresses)
- ARP provides a mapping between IP addresses and LAN addresses
- RARP provides mapping from LAN addresses to IP addresses
- This is accomplished by sending a "broadcast" packet requesting the owner of the IP address to respond with their physical address
  - All nodes on the LAN recognize the broadcast message
  - The owner of the IP address responds with its physical address
- An ARP cache is maintained at each node with recent mappings

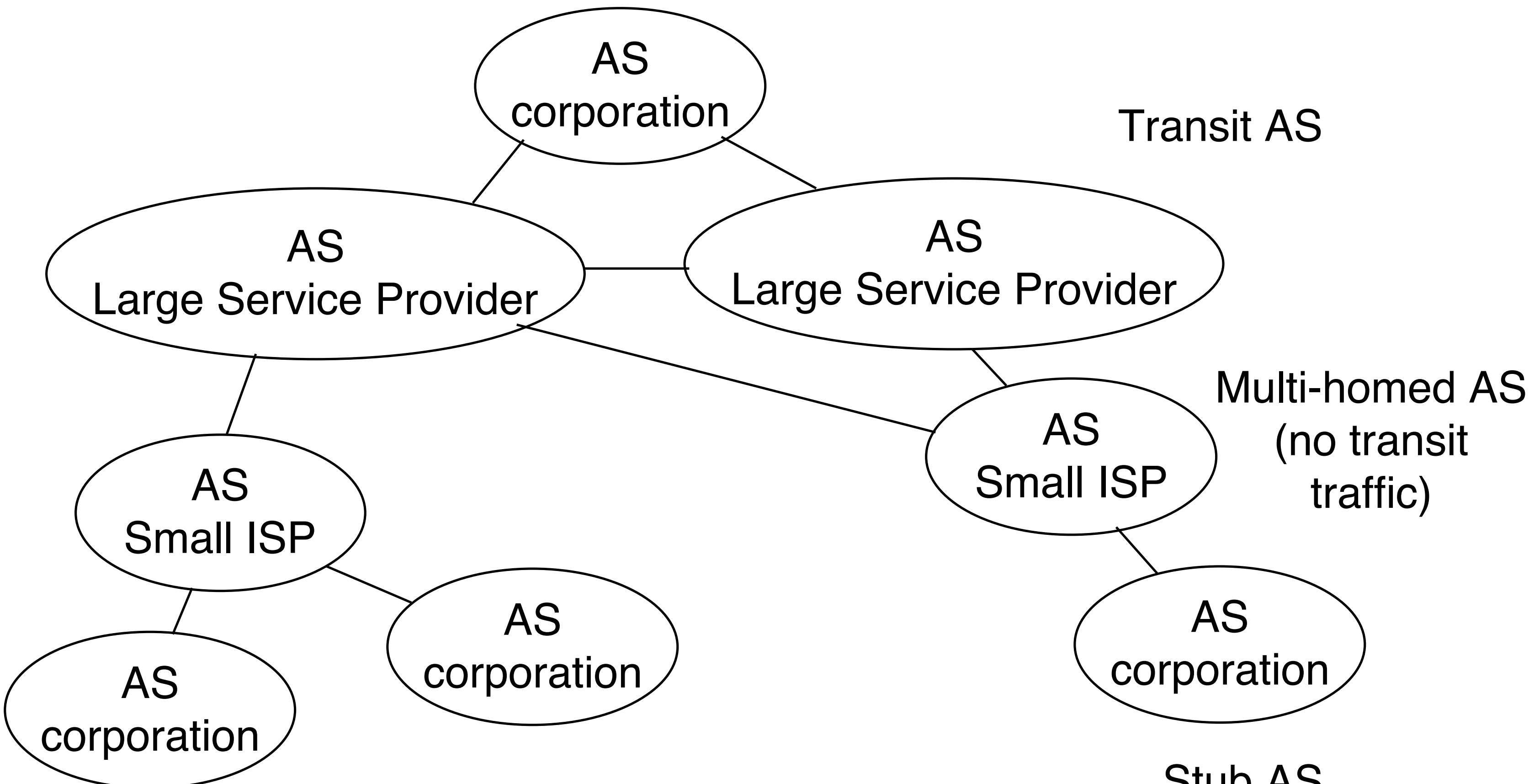


Routing in the Internet

- The internet is divided into sub-networks, each under the control of a single authority known as an Autonomous System (AS)
- Routing algorithms are divided into two categories:
  - Interior protocols (within an AS)
  - Exterior protocols (between AS's)
- Interior Protocols use shortest path algorithms
  - Distance vector protocols based on Bellman-ford algorithm
    - Nodes exchange routing tables with each other
    - E.g., Routing Information Protocol (RIP)
  - Link state protocols based on Dijkstra's algorithm
    - Nodes monitor the state of their links (e.g., delay)
    - Nodes broadcast this information to all of the network
    - E.g., Open Shortest Path First (OSPF)
- Exterior protocols route packets across AS's
  - Issues: no single cost metric, policy routing, etc..
  - Hierarchical routing based on "peering" agreements
  - Example protocols: Exterior Gateway protocol (EGP) and Border Gateway protocol (BGP)

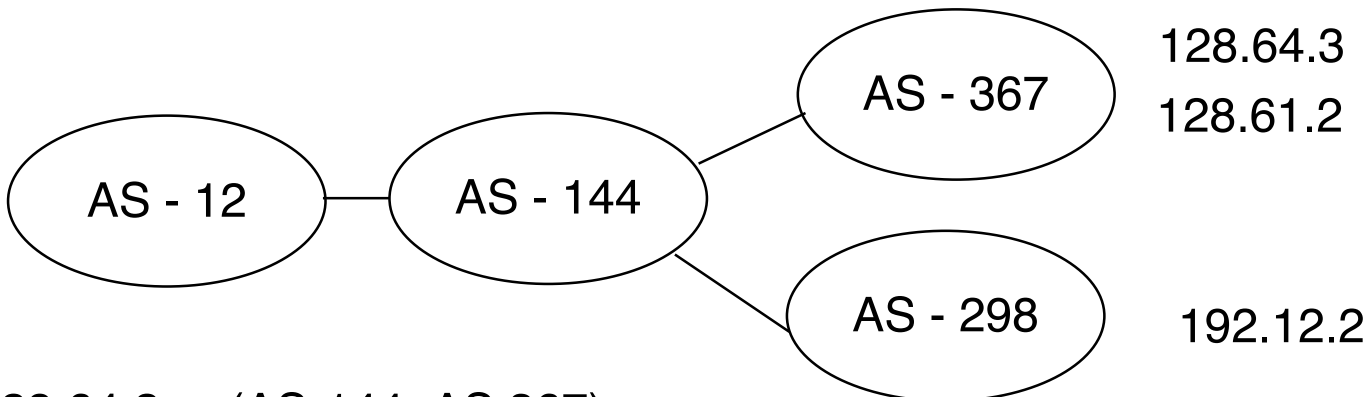
Border Gateway Protocol (BGP)

- **Routing between Autonomous systems**
  - **Find a path (no optimality) to destination (AS)**
    - **Path must satisfy policy criteria**



BGP overview

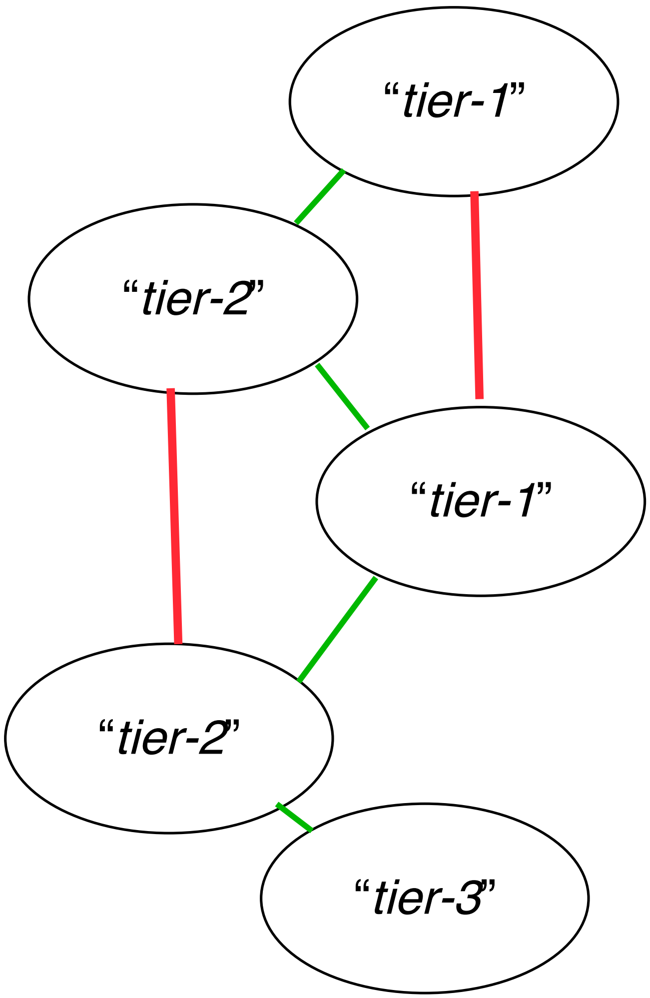
- **BGP speaker - one per AS**
  - **Establishes (TCP) session with other "speakers" to exchange reachability information**
- **Border "gateways" - routers that interface between AS's**
- **BGP advertises complete paths to destination AS**
  - **Avoid looping problems**
  - **Enable policy decisions (e.g., avoid certain AS's)**
  - **AS numbers - centrally assigned 16 bit numbers**
    - **Stub AS's don't need a number**



Path to 128.64.2 = (AS-144, AS 367)

Relationships between AS's

- **ISP "tiers"**
  - Tier-1 ISP's - provide global reachability
  - Tier-2 ISP's - regional/country
  - Tier-3 ISP's - local
- **Provider-customer relationship (transit)**
  - Smaller AS's purchase internet access from larger ones
    - MIT purchases access from BBN (a regional provider)
    - BBN purchases access from AT&T a global provider
- **Peering**
  - ISP's of similar size are "peers" and forward each other's traffic at no charge
    - E..g, AT&T - MCI; MCI-BT
  - Paid peering: a small ISP may "purchase" the right to peer with a larger provider (different than a provider-customer relationship)
    - E.g., Israel Telecom - BT
- **Policy issues**
  - Which routes would an ISP advertise?
    - E.g., to a transit customer because they pay



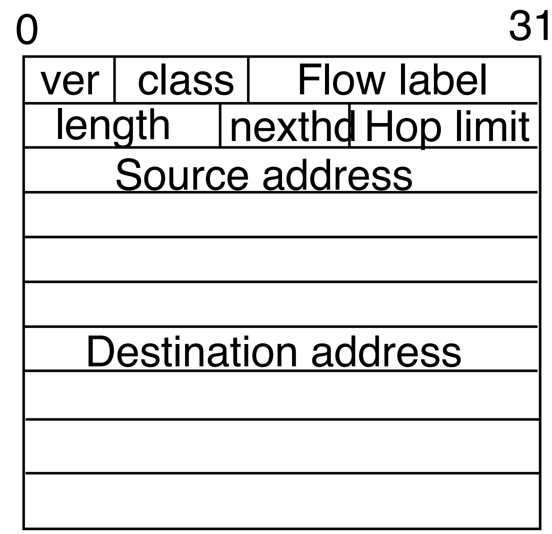
IPv6

- Effort started in 1991 as IPng
- Motivation
  - Need to increase IP address space
  - Support for real time application - “QoS”
  - Security, Mobility, Auto-configuration

- Major changes
  - Increased address space (16 bytes)
    - 1500 IP addresses per sq. ft. of earth!
    - Address partition similar to CIDR
  - Support for QoS via Flow Label field
  - Simplified header

- Is IPv6 really needed?
  - Most of the reasons for IPv6 have been taken care of in IPv4
  - Transition is complex

- Transition to IPv6
  - Cannot be done at once; must support joint operation
  - Dual-stack: routers run both IPv4 and IPv6
  - Tunneling: IPv6 packets carried in payload of IPv4 packets



Resource Reservation (RSVP)

- Service classes (defined by IETF)
  - Best effort
  - Guaranteed service
    - Max packet delay
  - Controlled load
    - emulate lightly loaded network via priority queueing mechanism (e.g., WFQ)
- Need to reserve resources at routers along the path
- RSVP mechanism
  - Packet classification
    - Associate packets with sessions (use flow field in IPv6)
  - Receiver initiated reservations to support multicast
  - "soft state" - temporary reservation that expires after 30 seconds
    - Simplify the management of connections
    - Requires refresh messages
  - Packet scheduling to guarantee service
    - Proprietary mechanisms (e.g., Weighted fair queueing)
- Scalability Issues
  - Each router needs to keep track of large number of flows that grows with the size (capacity) of the router

Differentiated Services (Diffserv)

- Unlike RSVP Diffserv does not need to keep track of individual flows
  - Allocate resources to a small number of classes of traffic
    - Queue packets of the same class together
  - E.g., two classes of traffic - premium and regular
    - Use one bit to differential between premium and regular packets
  - Issues
    - Who sets the premium bit?
      - "Edge routers"; ISP's; applications?
    - How is premium service different from regular?
      - "per-hop-behavior" (PHB): defines how each router treats a particular class of service
- IETF propose to use TOS field in IP header to identify traffic classes
  - Diffserv "code points" DSCP: 6-bit value that identifies a class of service
- Explicit forwarding (EF) PHB: forward with minimal delay
  - Total rate of EF traffic must be less than link rate
  - Give EF traffic strict priority over other traffic
  - Alternatively, use WFQ with high weight for EF traffic
- Assured forwarding PHB: packets are marked as being "in" or "out" of the customer's "traffic profile" , and treated accordingly
  - Profile represents a service agreement with the customer
  - Rarely drop packets within "profile"

Asynchronous Transfer Mode (ATM)

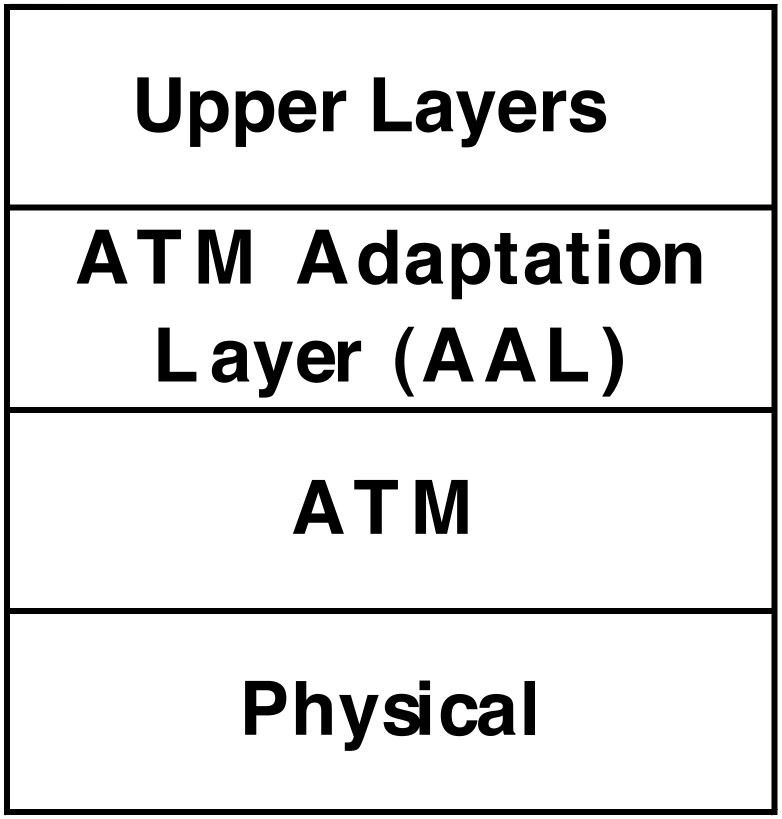
- 1980's effort by the phone companies to develop an integrated network standard (BISDN) that can support voice, data, video, etc.

- ATM uses small (53 Bytes) fixed size packets called "cells"
  - Why cells?
    - Cell switching has properties of both packet and circuit switching
    - Easier to implement high speed switches
  - Why 53 bytes?
  - Small cells are good for voice traffic (limit sampling delays)
    - For 64Kbps voice it takes 6 ms to fill a cell with data

- ATM networks are connection oriented
  - Virtual circuits

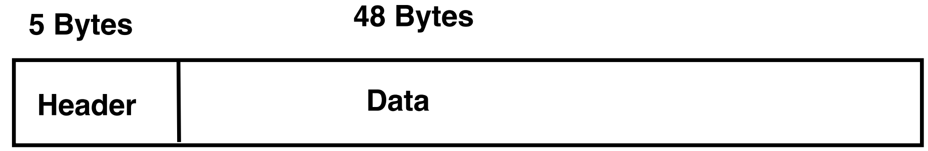
ATM Reference Architecture

- Upper layers
  - Applications
  - TCP/IP
- ATM adaptation layer
  - Similar to transport layer
  - Provides interface between upper layers and ATM
    - Break messages into cells and reassemble
- ATM layer
  - Cell switching
  - Congestion control
- Physical layer
  - ATM designed for SONET
    - Synchronous optical network
    - TDMA transmission scheme with 125 μs frames

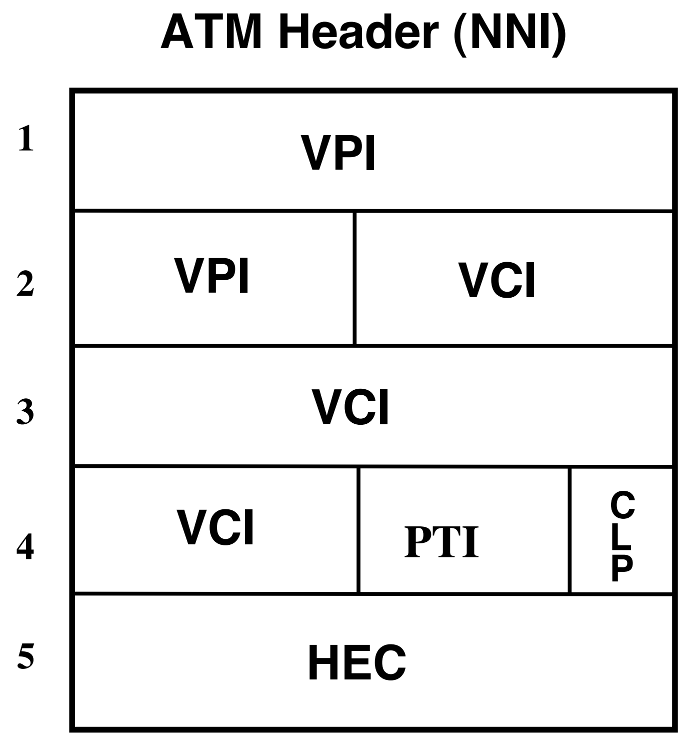


ATM Cell format

ATM Cell

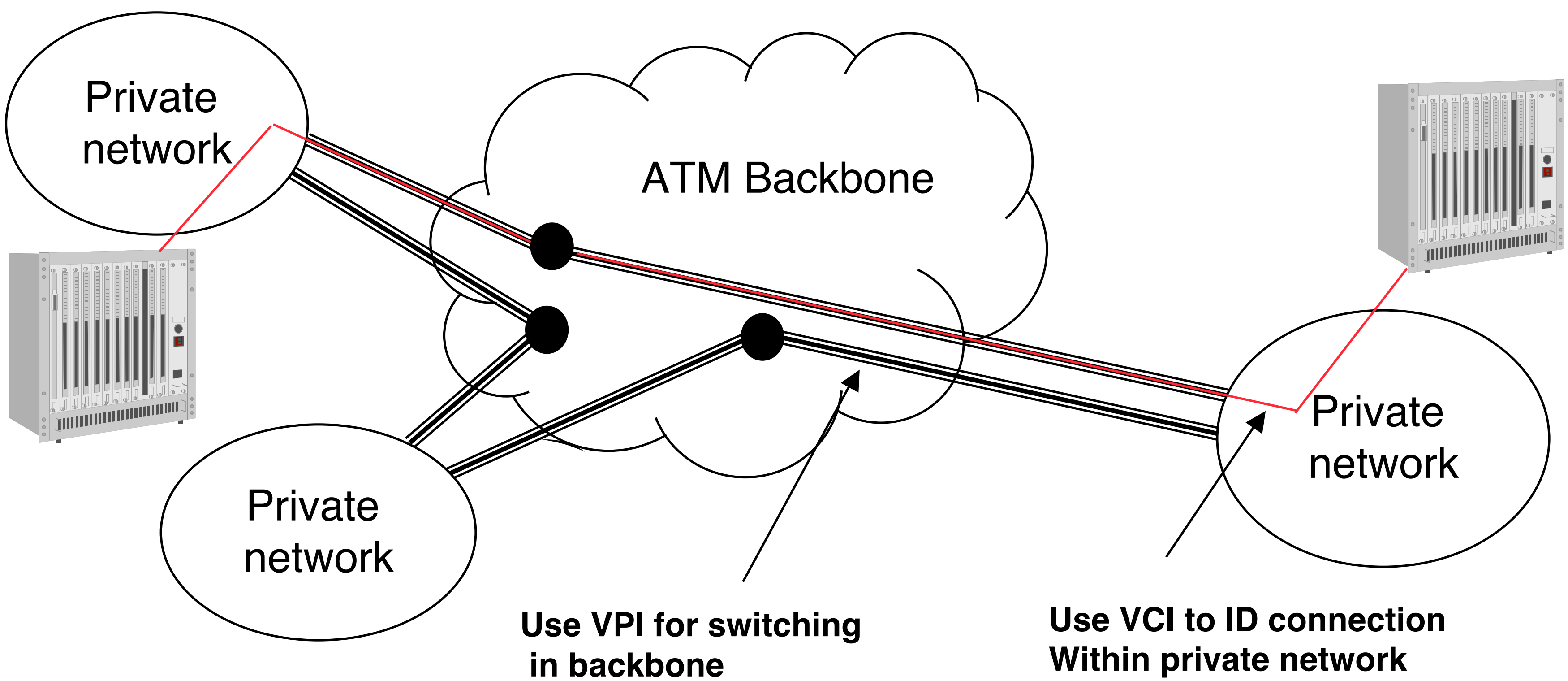


- Virtual circuit numbers (notice relatively small address space!)
  - Virtual channel ID
  - Virtual path ID
- PTI - payload type
- CLP - cell loss priority (1 bit!)
  - Mark cells that can be dropped
- HEC - CRC on header



VPI/VCI

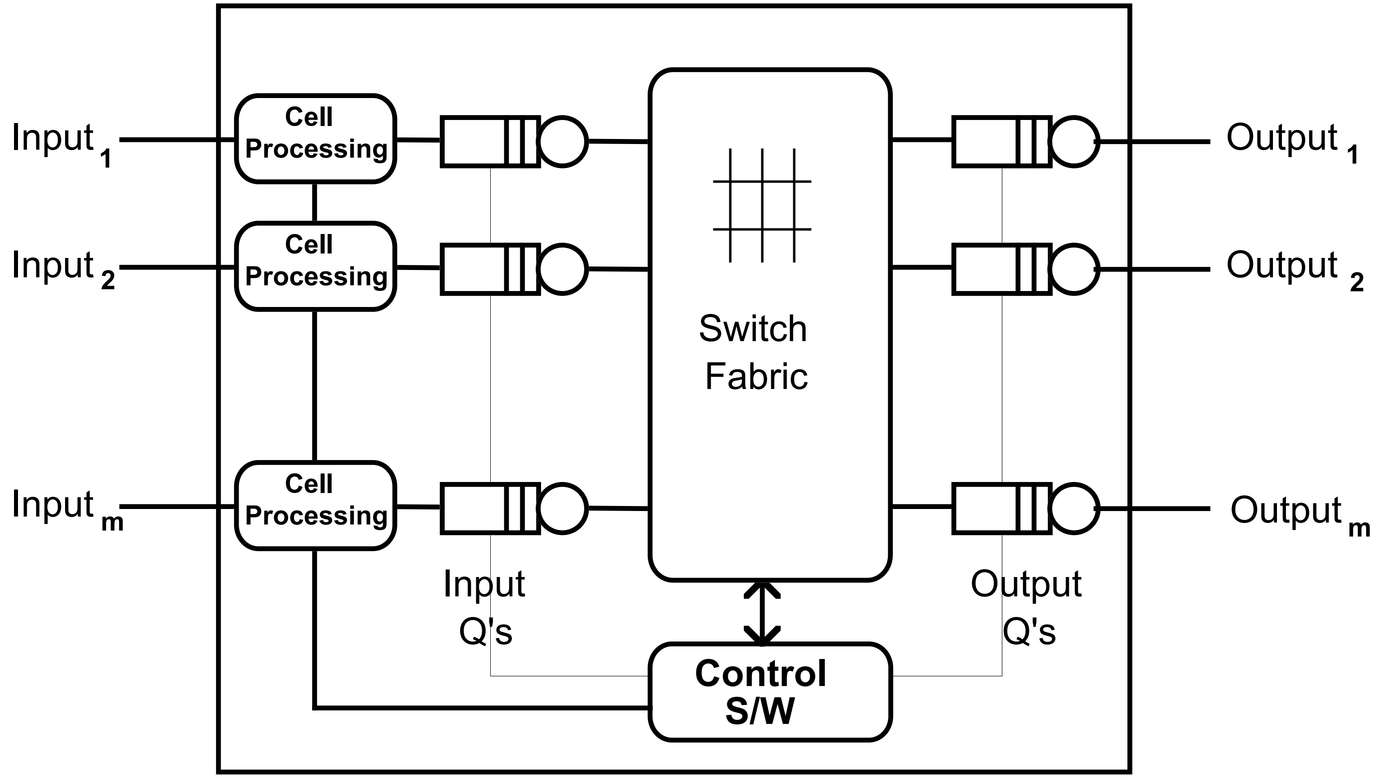
- **VPI identifies a physical path between the source and destination**
- **VCI identifies a logical connection (session) within that path**
  - **Approach allows for smaller routing tables and simplifies route computation**



ATM Service Categories

- **Constant Bit Rate (CBR) - e.g. uncompressed voice**
  - **Circuit emulation**
- **Variable Bit Rate (rt-VBR) - e.g. compressed video**
  - **Real-time and non-real-time**
- **Available Bit Rate (ABR) - e.g. LAN interconnect**
  - **For bursty traffic with limited BW guarantees and congestion control**
- **Unspecified Bit Rate (UBR) - e.g. Internet**
  - **ABR without BW guarantees and congestion control**

ATM cell switches



- **Design issues**
  - **Input vs. output queueing**
  - **Head of line blocking**
  - **Fabric speed**

ATM summary

- **ATM is mostly used as a “core” network technology**

- **ATM Advantages**
  - Ability to provide QoS
  - Ability to do traffic management
  - Fast cell switching using relatively short VC numbers

- **ATM disadvantages**
  - It not IP - most everything was design for TCP/IP
  - It’s not naturally an end-to-end protocol
    - Does not work well in heterogeneous environment
    - Was not design to inter-operate with other protocols
    - Not a good match for certain physical media (e.g., wireless)
  - Many of the benefits of ATM can be “borrowed” by IP
    - Cell switching core routers
    - Label switching mechanisms

Label Switching and MPLS

- Router makers realize that in order to increase the speed and capacity they need to adopt a mechanism similar to ATM
  - Switch based on a simple tag not requiring complex routing table look-ups
  - Use virtual circuits to manage the traffic (QoS)
  - Use cell switching at the core of the router
- First attempt: IP-switching
  - Routers attempt to identify flows
    - Define a flow based on observing a number of packets between a given source and destination (e.g., 5 packets within a second)
  - Map IP source-destination pairs to ATM VC's
    - Distributed algorithm where each router makes its own decision
- Multi-protocol label switching (MPLS)
  - Also known as Tag switching
  - Does not depend on ATM
  - Add a tag to each packet to serve as a VC number
    - Tags can be assigned permanently to certain paths

Label switching can be used to create a virtual mesh with the core network

- **Routers at the edge of the core network can be connected to each other using labels**
- **Packets arriving at an edge router can be tagged with the label to the destination edge router**
  - **"Tunneling"**
  - **Significantly simplifies routing in the core**
  - **Interior routers need not remember all IP prefixes of outside world**
  - **Allows for traffic engineering**
    - **Assign capacity to labels based on demand**