Networking

CS 4410
Operating Systems

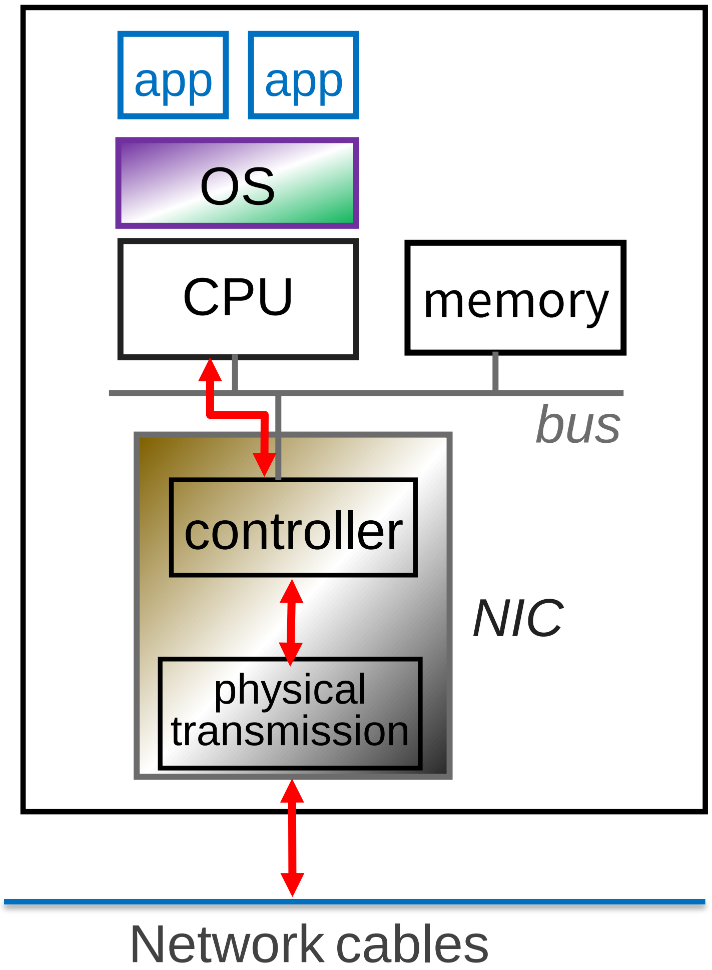
[R. Agarwal, L. Alvisi, A. Bracy, M. George, Kurose, Ross, E. Sirer, R. Van Renesse]

Application Layer
Transport Layer
Network Layer
Link Layer
Physical Layer

Link Layer: Local Area Networking (LAN) and Ethernet

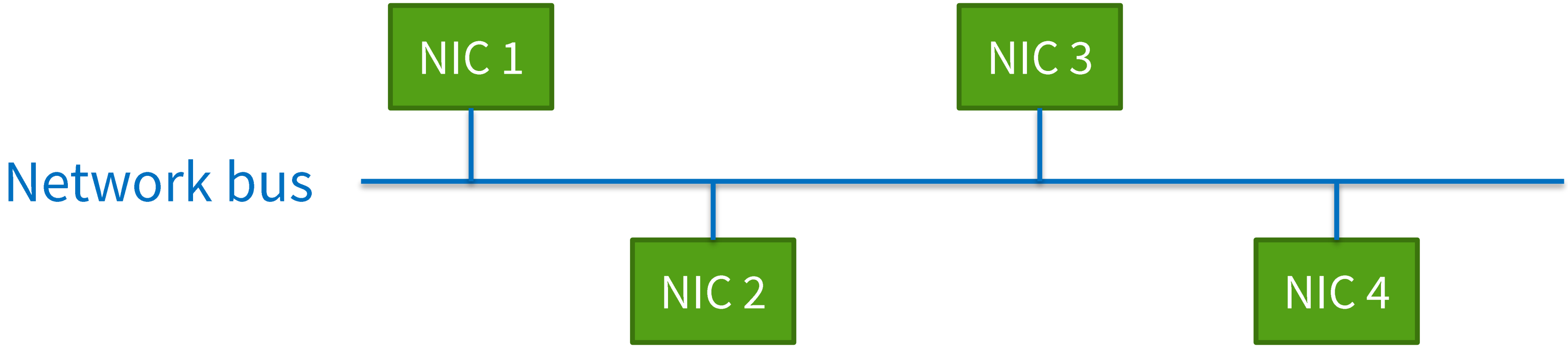
Link Layer

- Each host has one or more *NICs*
  - *Network Interface Cards*
    - Ethernet, 802.11, etc.
- Each NIC has a *MAC address*
  - *Media Access Control* address
  - Ethernet example: b8:e3:56:15:6a:72
  - Unique to network instance
    - often even globally unique
  - Does not change if NIC moves
- Messages are *packets* or *frames*



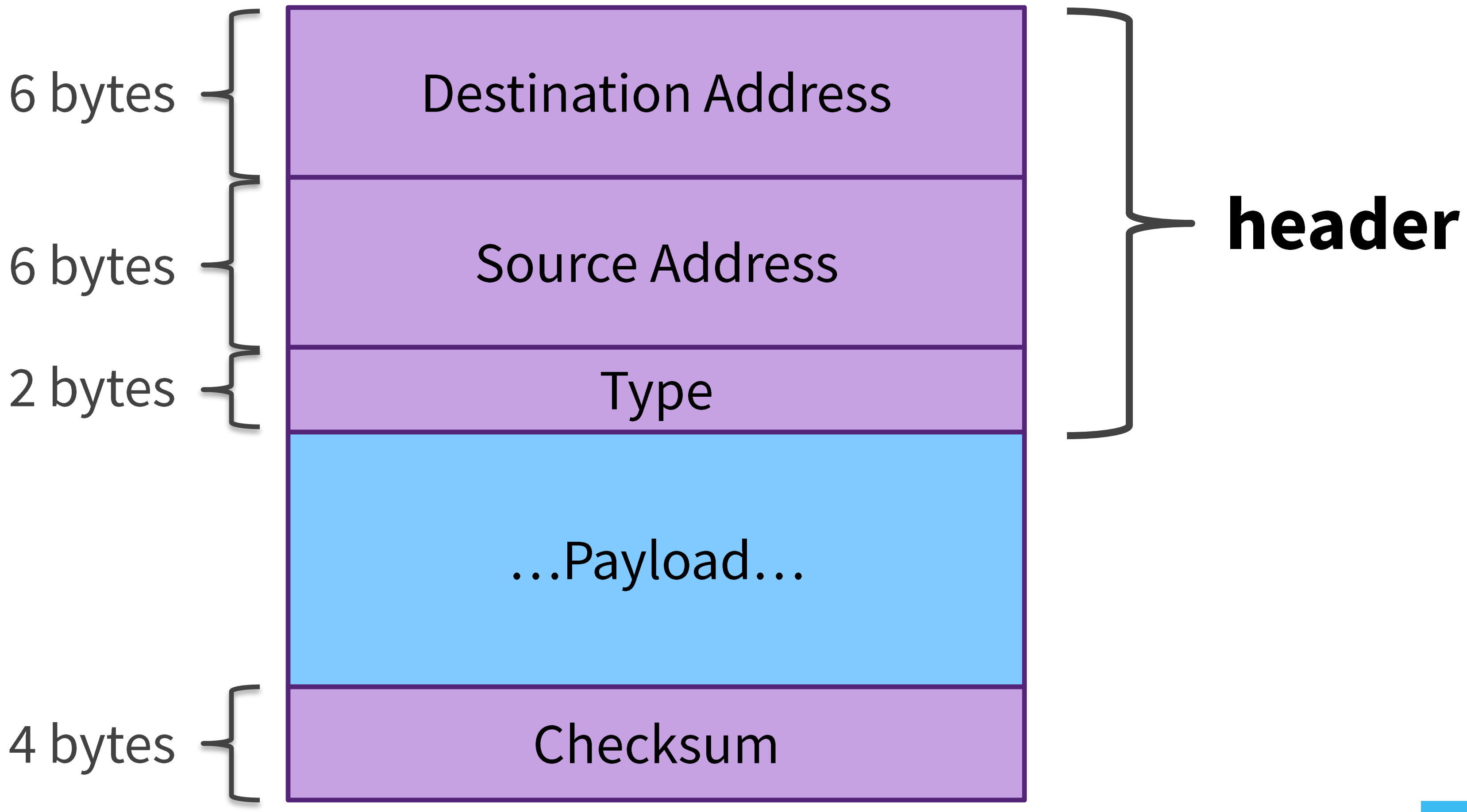
Example: Ethernet

- 1976, Metcalfe & Boggs at Xerox
  - Later at 3COM
- Based on the Aloha network in Hawaii
- Named after the "*luminiferous ether*"
- Centered around a broadcast bus
- Simple link-level protocol, scales pretty well
- Tremendously successful
- Still in widespread use
  - many orders of magnitude increase in bandwidth since early versions



Ethernet basics

An Ethernet packet



CRC Checksum

(Cyclic Redundancy Check)

- Basically a hash function on the packet
- Added to the end of a packet
- Used to detect malformed packets, e.g. electrical interference, noise

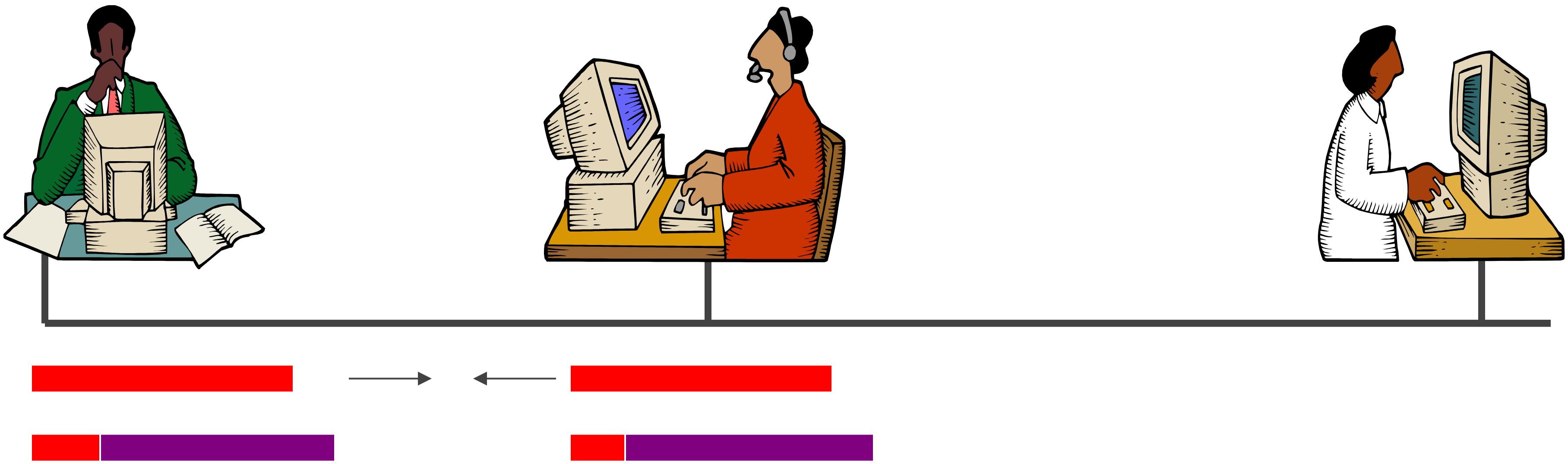
“CSMA/CD”

- Carrier sense
  - Listen before you speak
- Multiple access
  - Multiple hosts can access the network
- Collision detect
  - Detect and respond to cases where two hosts collide

Sending packets

- Carrier sense, broadcast if ether is available

Collisions



- What happens if two people decide to transmit simultaneously?

Collision Detection & Retransmission

- Detect collision by measuring incoming & outgoing signal strength
  - Shouldn't be receiving any incoming signal while transmitting
- Hosts involved in the collision promptly stop data transmission, sleep for a while, and attempt to retransmit
- How long they sleep is determined by how many collisions have occurred before (exponential backoff + random noise)
- Abort after 16 retries, hence no guarantee that a packet will get to its destination

Ethernet Features

- Completely distributed
  - No central arbiter

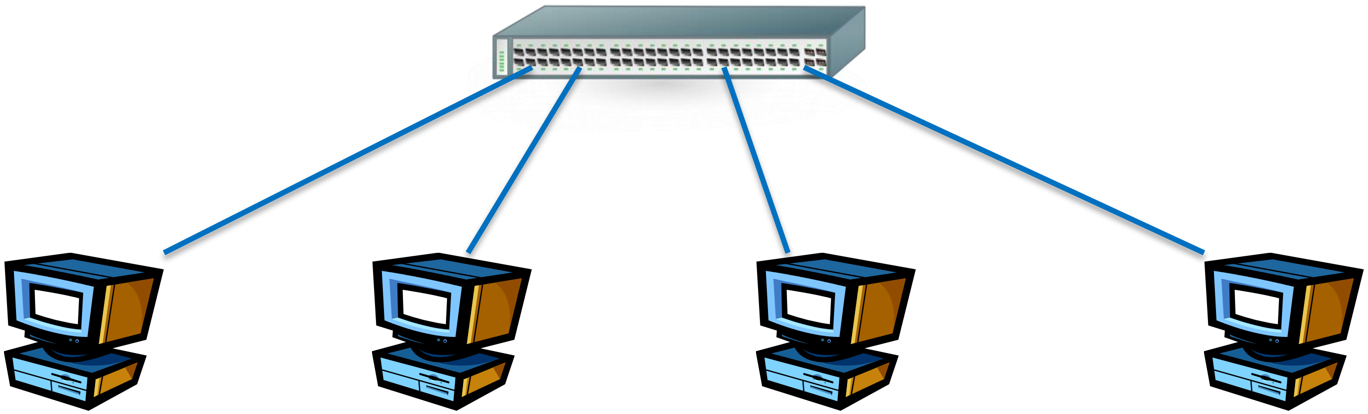
- Inexpensive
  - No state in the network
  - No extra hardware for arbitration
  - Cheap physical links (twisted pair of wires)

Ethernet Problems

- The endpoints are trusted to follow the collision-detect and retransmit protocol
  - Certification process tries to assure compliance
  - Not everyone always backs off exponentially
- Hosts are trusted to only listen to packets destined for them
  - But the data is available for all to see
    - All packets are broadcast on the wire
    - Can place Ethernet card in promiscuous mode and listen

Switched Ethernet

- Today's Ethernet deployments are much faster
- In wired settings, *Switched Ethernet* is now the norm
  - All hosts connect to a switch, which forwards packets
  - Each p2p connection is a mini Ethernet set-up
  - More secure (no snooping), no possibility of collisions
  - Switches organize into a spanning tree
- Not to be confused with Ethernet *Hub*
  - A hub simply connects the wires into one big shared wire



Wireless

- 802.11 protocols inherit many of the Ethernet concepts
- Full compatibility with Ethernet interface
    - Same address and packet formats

Application Layer
Transport Layer
Network Layer
Link Layer
Physical Layer

Network Layer

Network Layer

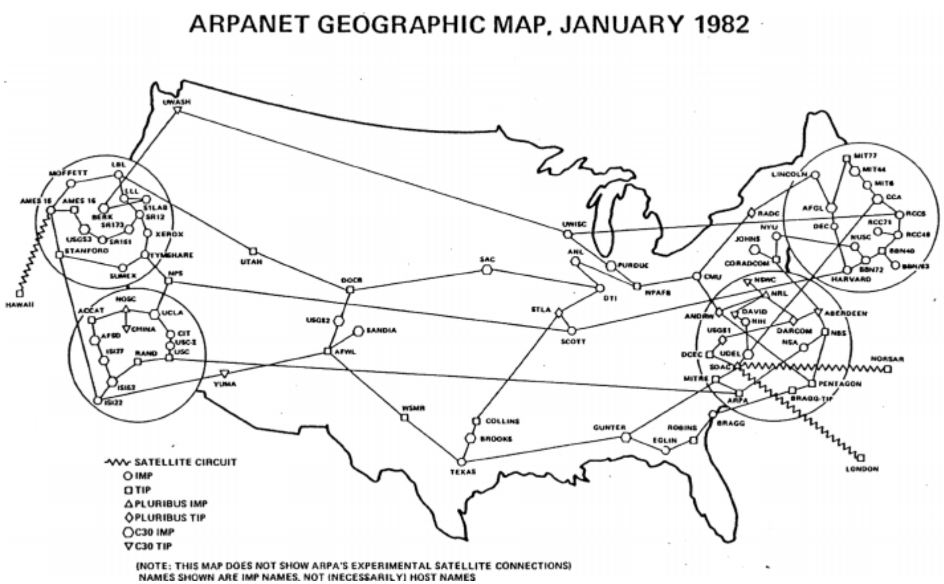
- There are lots of Local Area Networks
  - each with their own
    - address format and allocation scheme
    - packet format
    - maximum packet size
    - LAN-level protocols, reliability guarantees
- Wouldn't it be nice to tie them all together?
  - Nodes with multiple NICs can provide the glue!
  - Standardize address and packet formats
- This gives rise to an "Internetwork"
  - aka WAN (wide-area network)

Internetworking Origins

- Expensive supercomputers scattered throughout the US
- Researchers scattered differently throughout the US
- Needed a way to connect researchers to expensive machinery

Internetworking Origins

- Department of Defense initiated studies on how to build a resilient global network (60s, 70s)
  - How do you coordinate a nuclear attack?
- Interoperability and dynamic routing are a must
  - Along with a lot of other properties
- Result: *Internet (orig. ARPAnet, then NSFnet)*
- A **complex** system with **simple** components



Internet Overview

- Every host is assigned, and identified by, an <u>IP address</u>
- Messages are called <u>datagrams</u>
  - the term *packet* is probably more common though…
- Each datagram contains a <u>header</u> that specifies the destination address
- The network <u>routes</u> datagrams from the source to the destination

IP

- Internetworking protocol
  - Network layer
- Common address format
- Common packet format for the Internet
  - Specifies what packets look like
  - *Fragments* long packets into shorter packets
  - *Reassembles* fragments into original shape
- IPv4 vs IPv6
  - IPv4 is what most people use
  - IPv6 more scalable and clears up some of the messy parts

IP: Narrow Waist

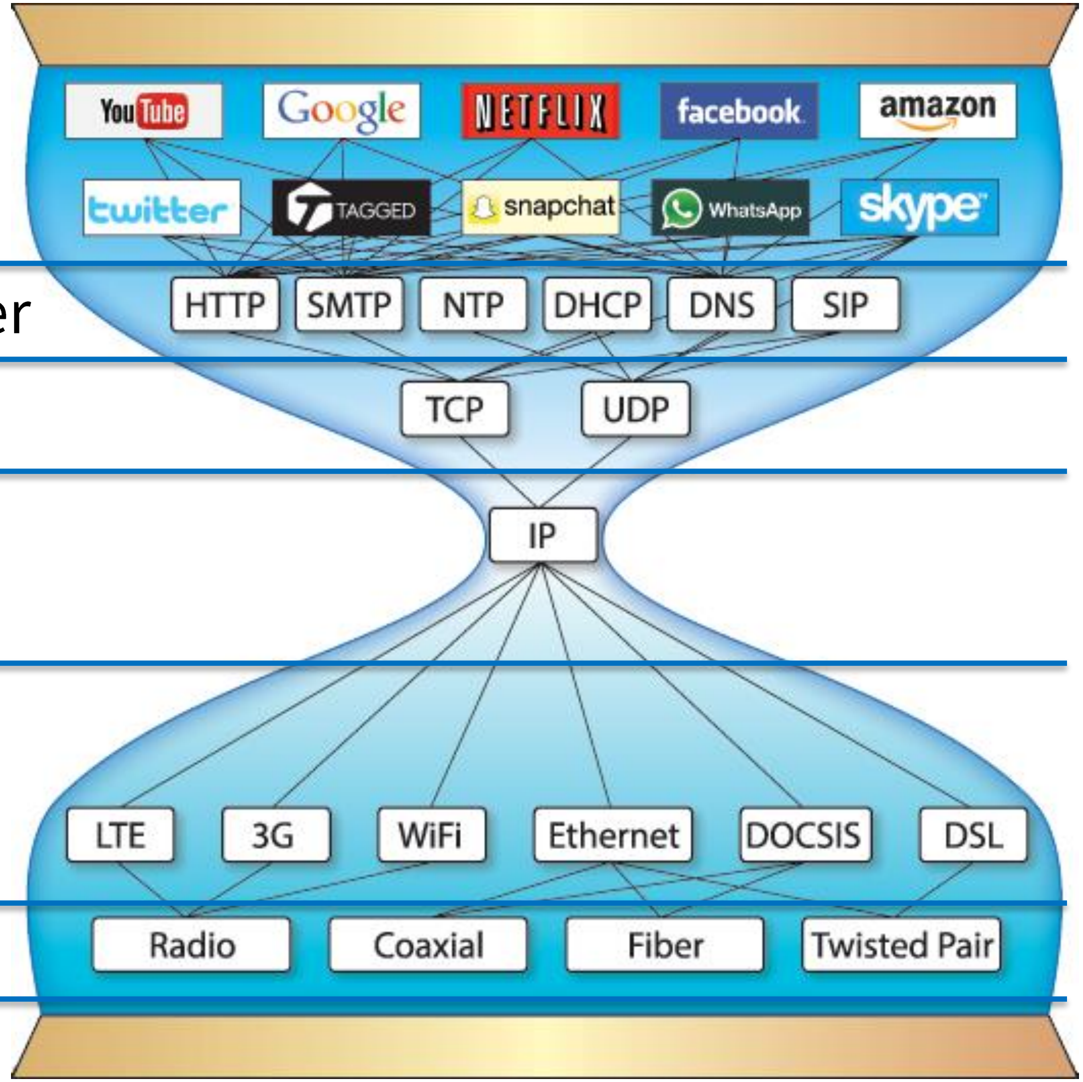
Application Layer

Transport Layer

Network Layer

Data Link Layer

Physical Layer



from: http://if-we.clients.labzero.com/code/posts/what-title-ii-means-for-tcp/

IP Addressing

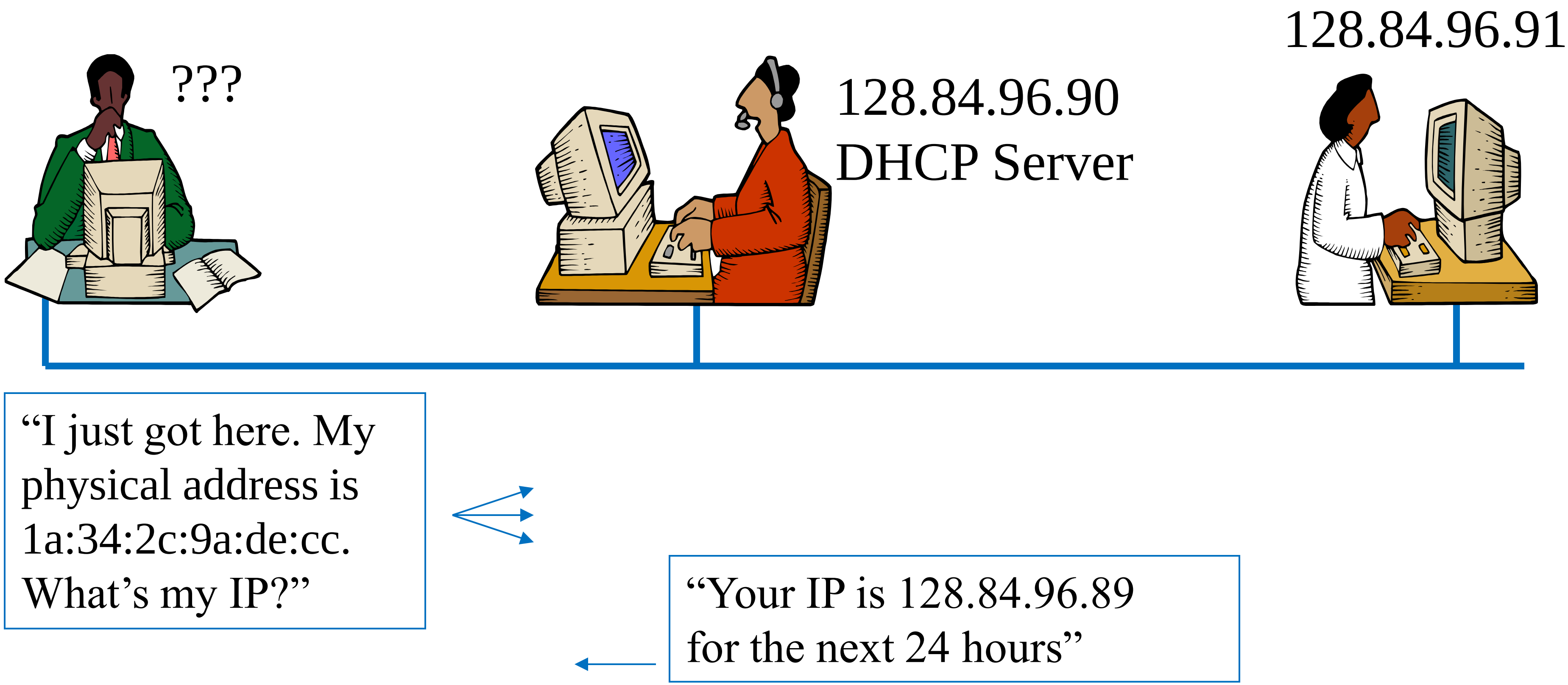
- Every (active) NIC has an IP address
  - IPv4: 32-bit descriptor, e.g. 128.84.12.43
  - IPv6: 128-bit descriptor (but only 64 bits "functional")
  - Will use IPv4 unless specified otherwise…
- Each Internet Service Provider (ISP) owns a set of IP addresses
- ISPs assign IP addresses to NICs
- IP addresses can be re-used
- Same NIC may have different IP addresses over time

IP “subnetting”

- An IP address consists of a prefix of size *n* and a suffix of size 32 – *n*
  - Either specified by a number, e.g., 128.84.32.00/24
  - Or a “netmask”, e.g., 255.255.255.0 (in case n = 24)
- A “subnet” is identified by a prefix and has 2^{32-n} addresses
  - Suffix of “all zeroes” or “all ones” reserved for broadcast
  - Big subnets have a short prefix and a long suffix
  - Small subnets have a long prefix and a short suffix

Addressing & DHCP

How to get an IP address:



DHCP is used to discover IP addresses (and more)

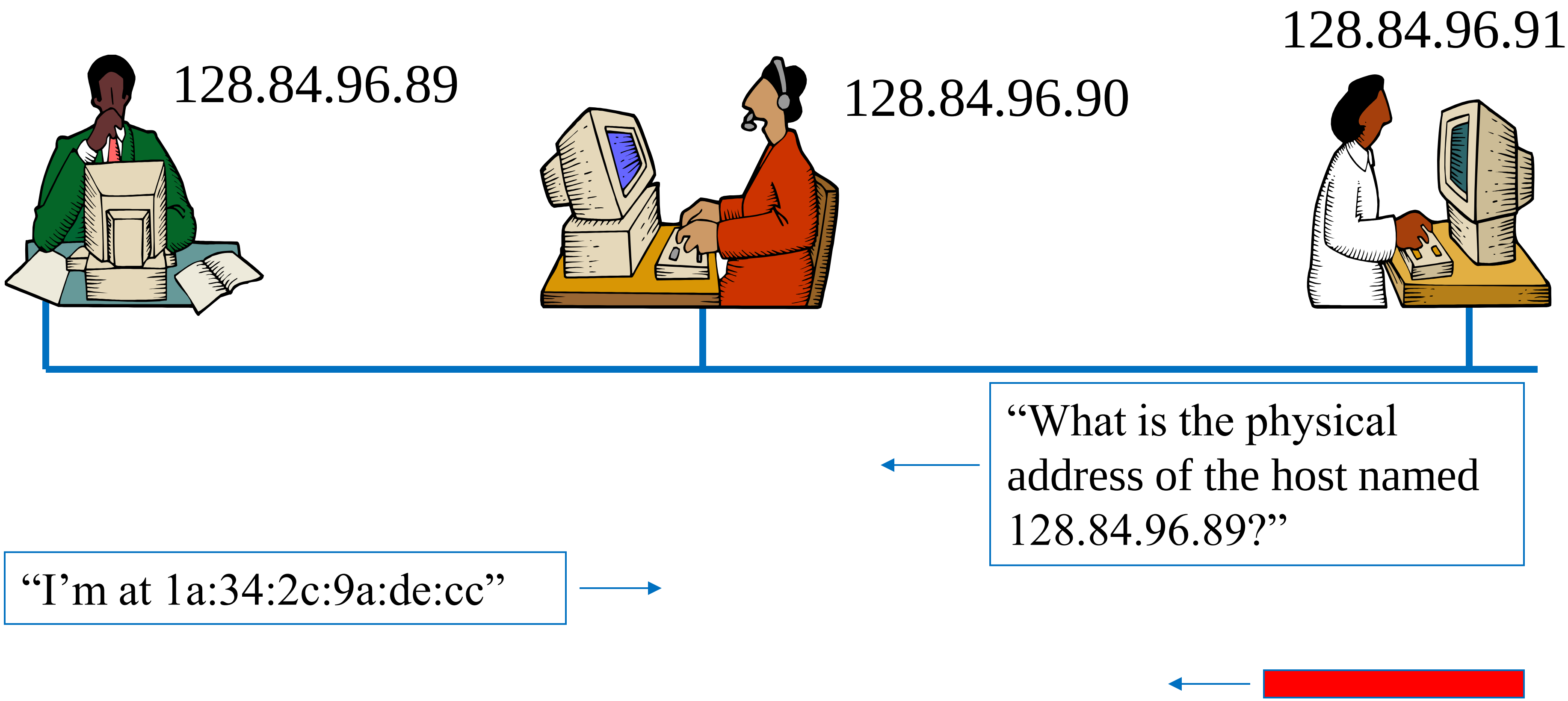
DHCP = Dynamic Host Configuration Protocol

DHCP

- Each LAN (usually) runs a DHCP server
  - you probably run one at home inside your “router box”
- DHCP server maintains
  - the IP subnet that it owns (say, 128.84.245.00/24)
  - a map of IP address <-> MAC address
    - possibly with a timeout (called a “lease”)
- When a NIC comes up, it broadcasts a DHCPDISCOVER message
  - if MAC address in the map, respond with corresponding IP address
  - if not, but an IP address is unmapped and thus available, map that IP address and respond with that
- DHCP also returns the netmask
- Note: NICs can also be statically configured and don’t need DHCP

Addressing & ARP

Virtual to physical translation for IP addresses

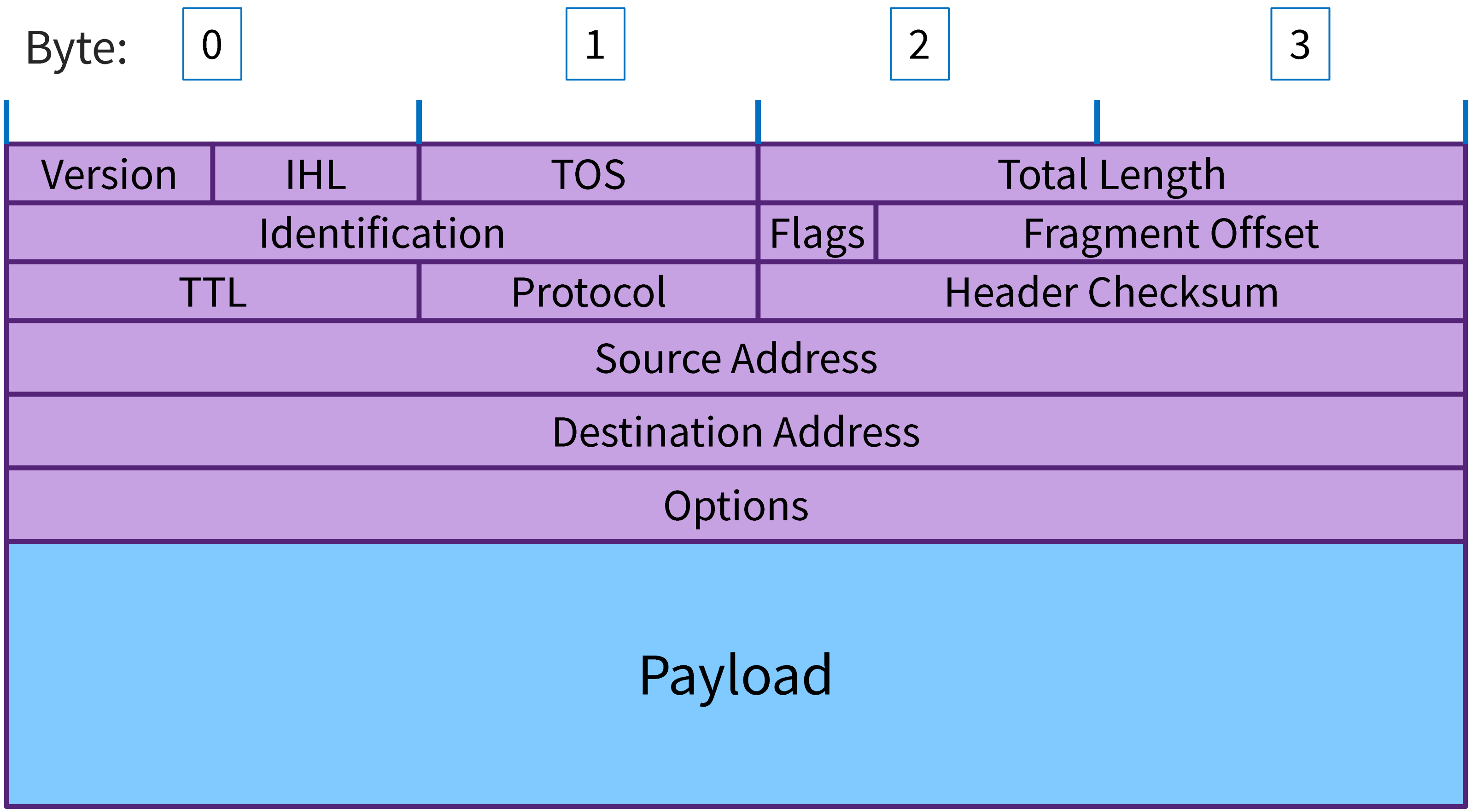


- ARP is used to discover MAC addresses *on same subnet*
  - ARP = Address Resolution Protocol

Scale?

- ARP and DHCP are broadcast protocols
- Only scale to single subnet
- Need more to scale to the Internet!
- Routing (next lecture)

IPv4 packet layout



IP Header Fields

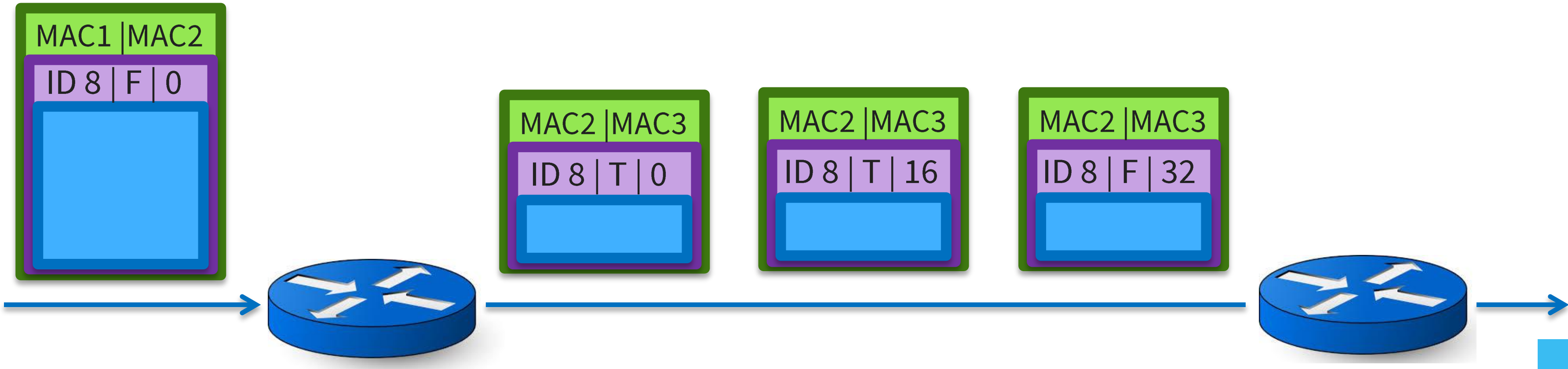
- Version (4 bits): 4 or 6
- IHL (4 bits): Internet Header Length in 32-bit words
  - usually 5 unless options are present
- TOS (1 byte): type of service (not used much)
- Total Length (2 bytes): length of packet in bytes
- Id (2 bytes), Flags (3 bits), Fragment Offset (13 bits)
  - used for fragmentation/reassembly. Stay tuned
- TTL (1 byte): Time To Live. Decremented at each hop
- Protocol (1 byte): TCP, UDP, ICMP, …
- Header Checksum (2 bytes): to detect corrupted headers

IP Fragmentation

- Networks have different maximum packet sizes
  - "MTU": Maximum Transmission Unit
    - Big packets are sometimes desirable – less overhead
    - Huge packets are not desirable – reduced response time for others
- High-level protocols could try to figure out the minimum MTU along the network path, but
    - Inefficient for links with large MTUs
    - The route can change underneath
- Consequently, IP can transparently fragment and reassemble packets

IP Fragmentation Mechanics

- Source assigns each datagram an “identification”
- At each hop, IP can divide a long datagram into N smaller datagrams
- Sets the More Fragments bit except on the last packet
- Receiving end puts the fragments together based on *Identification* and *More Fragments* and *Fragment Offset (times 8)*



IP Options (not well supported)

- Source Routing: The source specifies the set of hosts that the packet should traverse
- Record Route: If this option appears in a packet, every router along a path attaches its own IP address to the packet
- Timestamp: Every router along the route attaches a timestamp to the packet
- Security: Packets are marked with user info, and the security classification of the person on whose behalf they travel on the network
  - Most of these options pose security holes and are generally not implemented