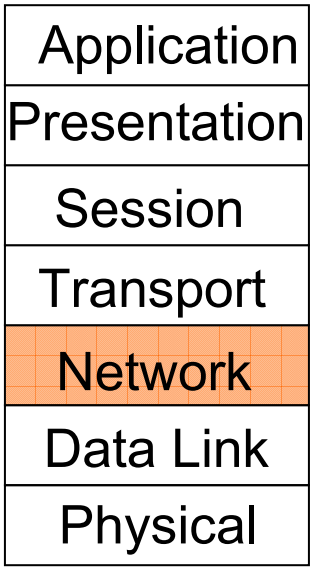
Internetworks – networks of networks

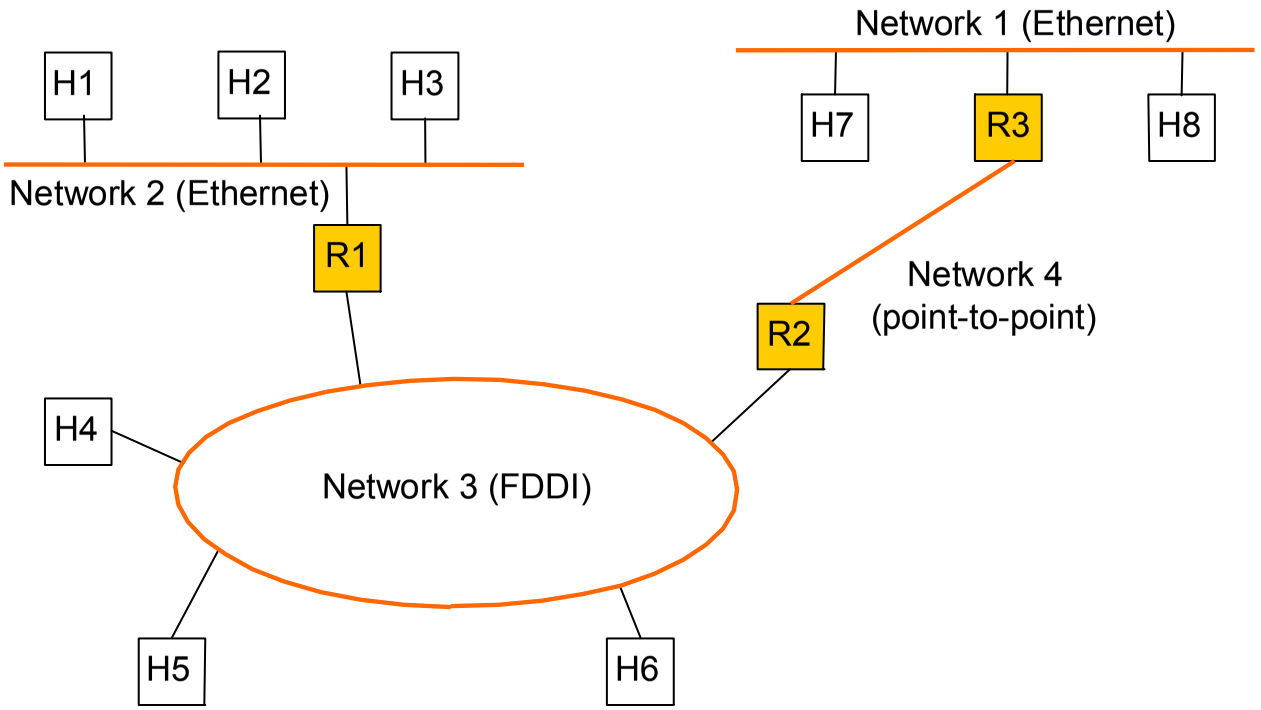
- Focus:
  - How do we build large networks?

- Start of the Network layer
  - Internetworks
  - Service models
  - IP, ICMP
  - We've covered some routing …



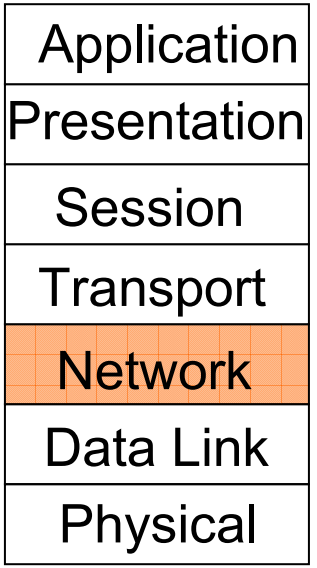
Internetworks

- Set of interconnected networks, e.g., the Internet
  - ***Key issues are scale and heterogeneity***



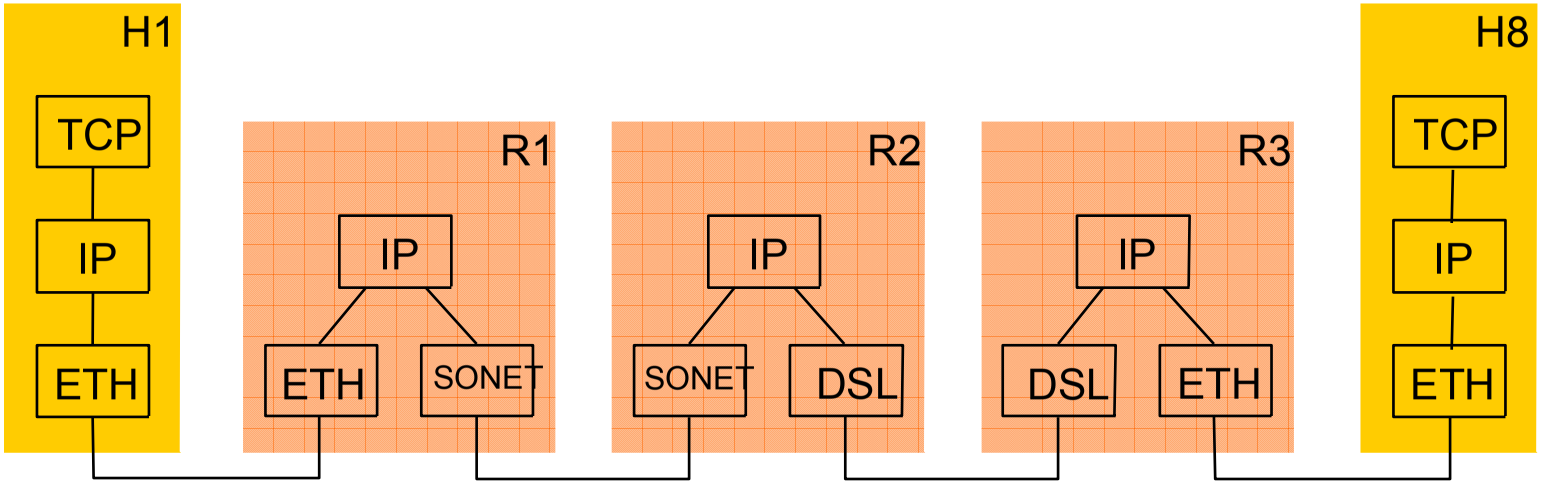
The Network Layer

- Job is to provide end-to-end data delivery between hosts on an internetwork
- Provides a higher layer of addressing



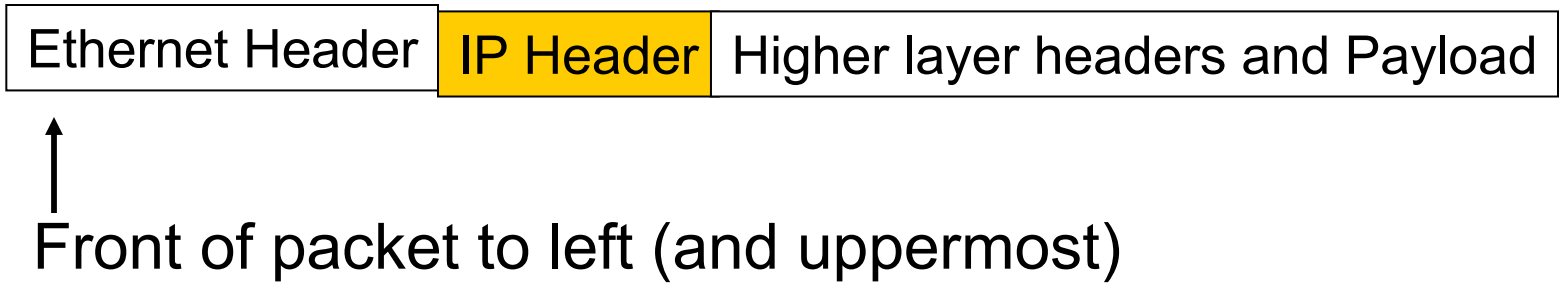
In terms of protocol stacks

- IP is the network layer protocol used in the Internet
- Routers are network level gateways
- Packet is the term for network layer PDUs



In terms of packet formats

- View of a packet on the wire on network 1 or 2
- Routers work with IP header, not higher
  - Higher would be a "layer violation"
- Routers strip and add link layer headers



Network Service Models

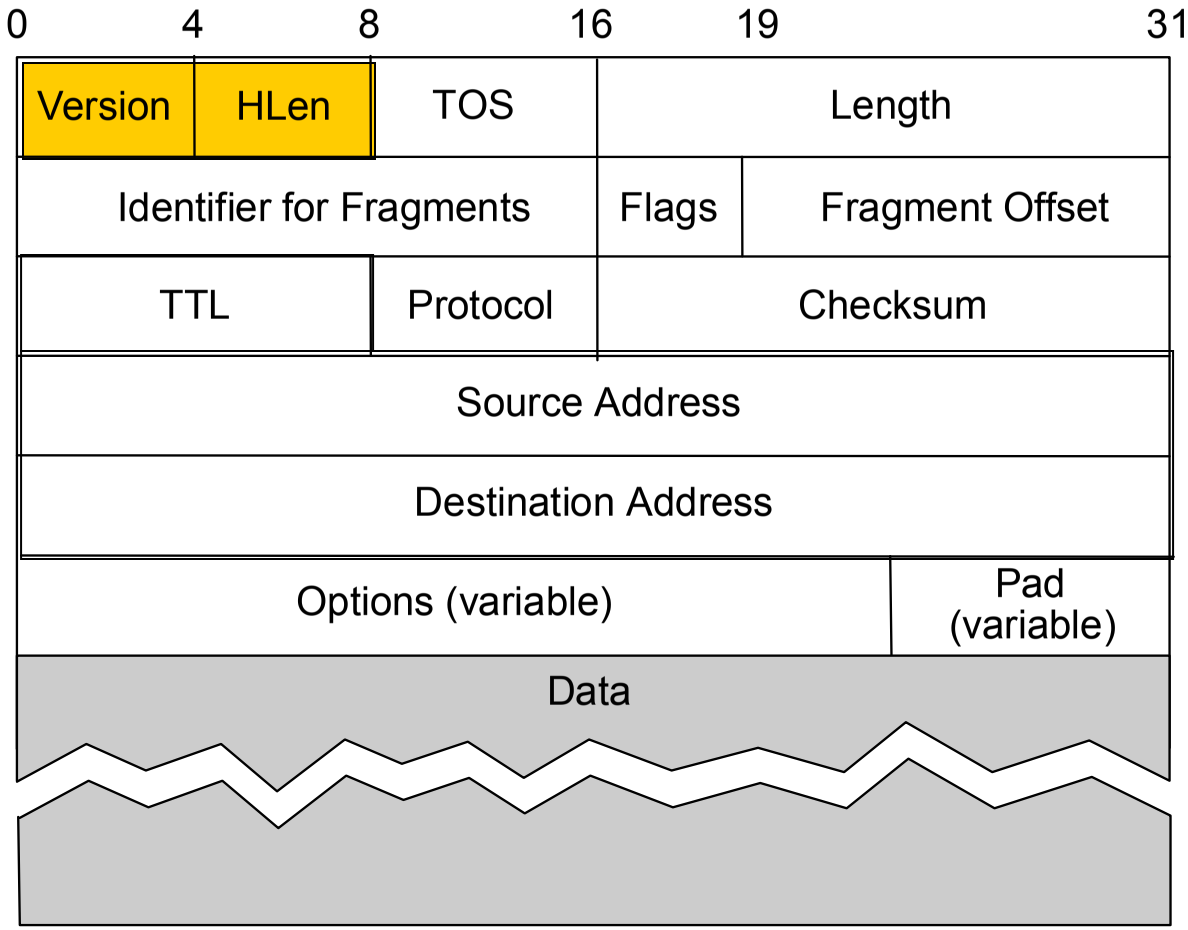
- Datagram delivery: postal service
  - Also connectionless, best-effort or unreliable service
  - Network can't guarantee delivery of the packet
  - Each packet from a host is routed independently
  - Example: IP
- Virtual circuit models: telephone
  - Also connection-oriented service
  - Signaling: connection establishment, data transfer, teardown
  - All packets from a host are routed the same way (router state)
  - Example: ATM, Frame Relay, X.25
- Pros and Cons?
  - Simplicity/robustness versus stronger resource allocation
  - These issues are at the heart of Internet evolution and QOS

Internet Protocol (IP)

- IP (RFC791) defines a "best effort" service
  - May be loss, reordering, duplication, and errors!
  - Currently IPv4 (IP version 4), IPv6 on the way
- Routers forward packets using predetermined routes
  - Routing protocols (RIP, OSPF, BGP) run between routers to maintain routes (routing table, forwarding information base)
- Global, hierarchical addresses, not flat addresses
  - 32 bits in IPv4 address; 128 bits in IPv6 address
  - ARP (Address Resolution Protocol) maps IP to MAC addresses
- We peek at the IPv4 formats to see what's involved …

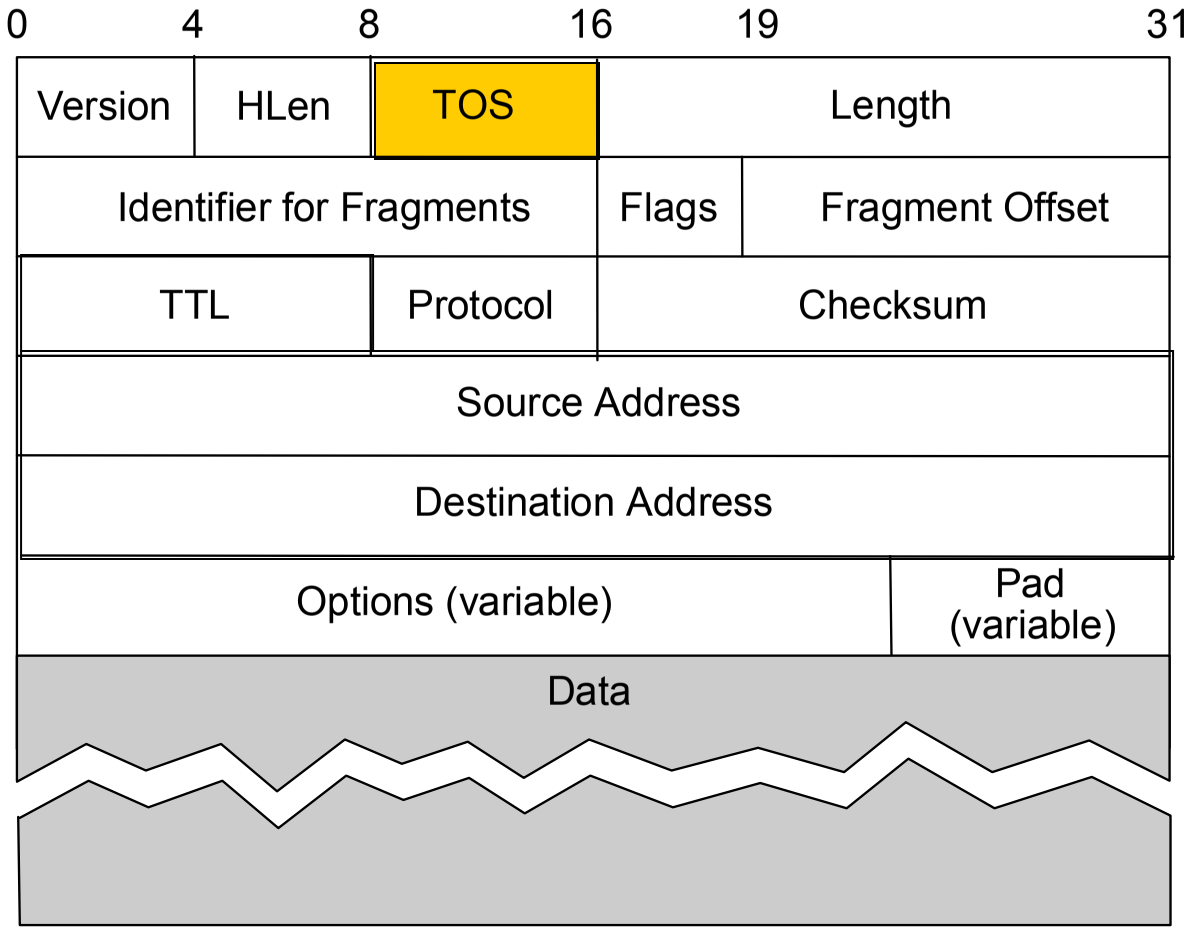
IPv4 Packet Format

- Version is 4
- Header length is number of 32 bit words
- Limits size of options



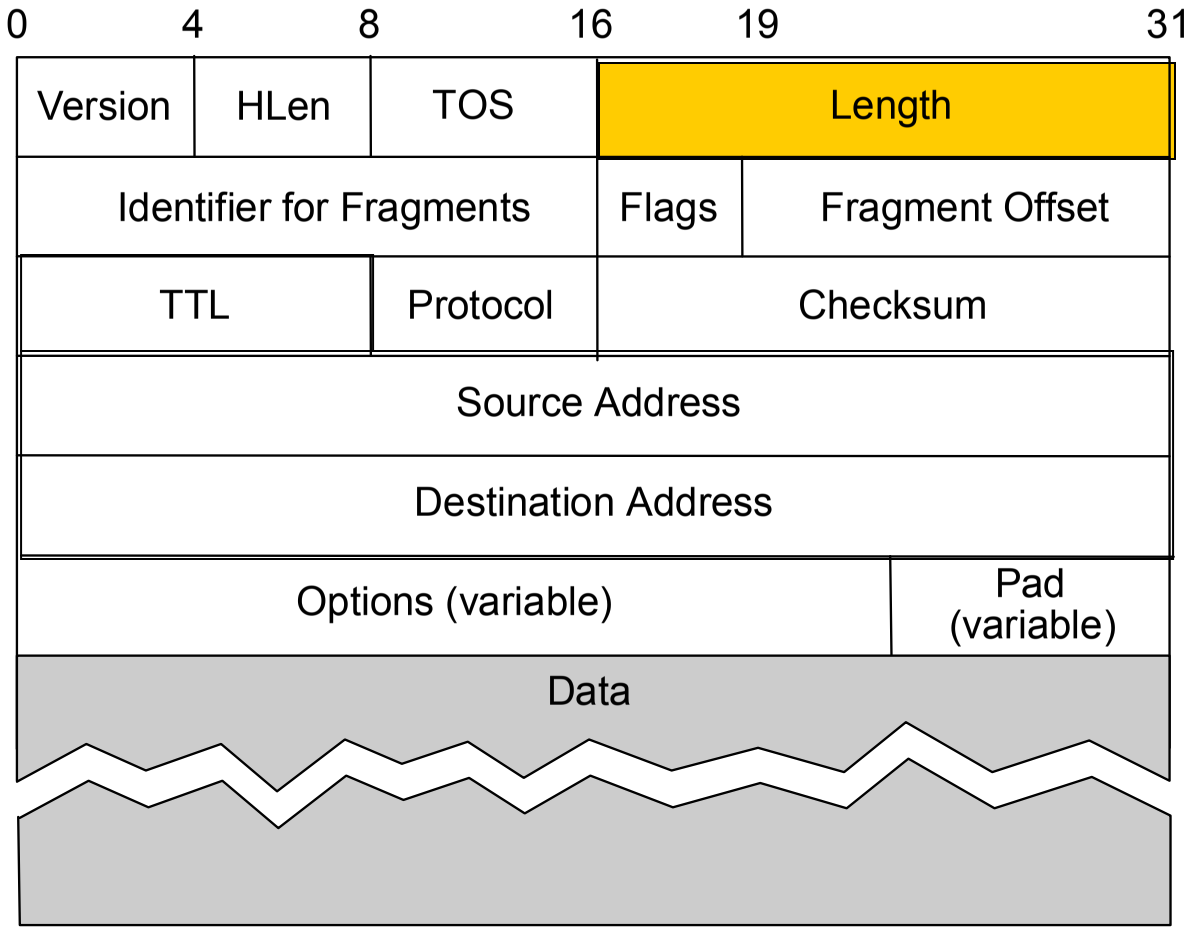
IPv4 Header Fields …

- Type of Service
- Abstract notion, never really worked out
  - Routers ignored
- But now being redefined for Diffserv



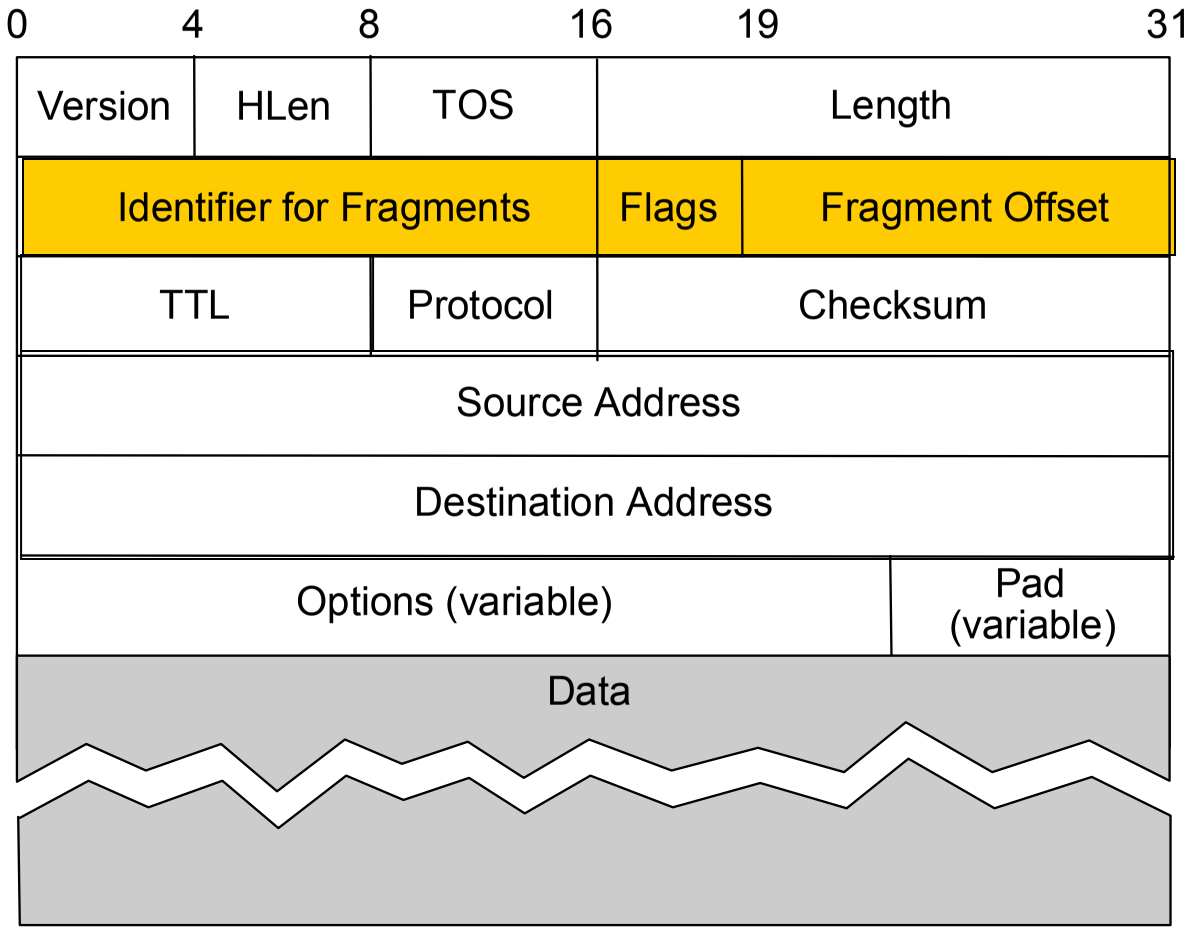
IPv4 Header Fields …

- Length of packet
- Min 20 bytes, max 65K bytes (limit to packet size)



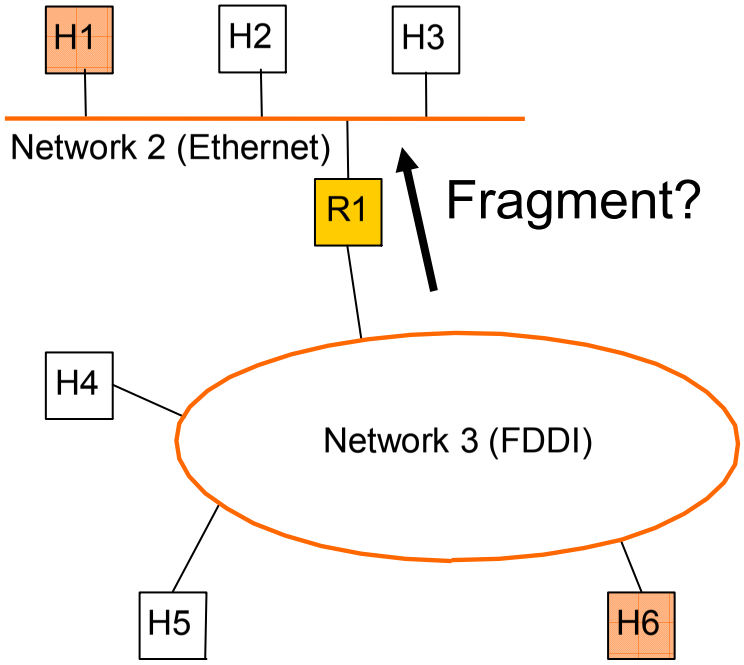
IPv4 Header Fields …

- Fragment fields
- Different LANs have different frame size limits
- May need to break large packet into smaller fragments



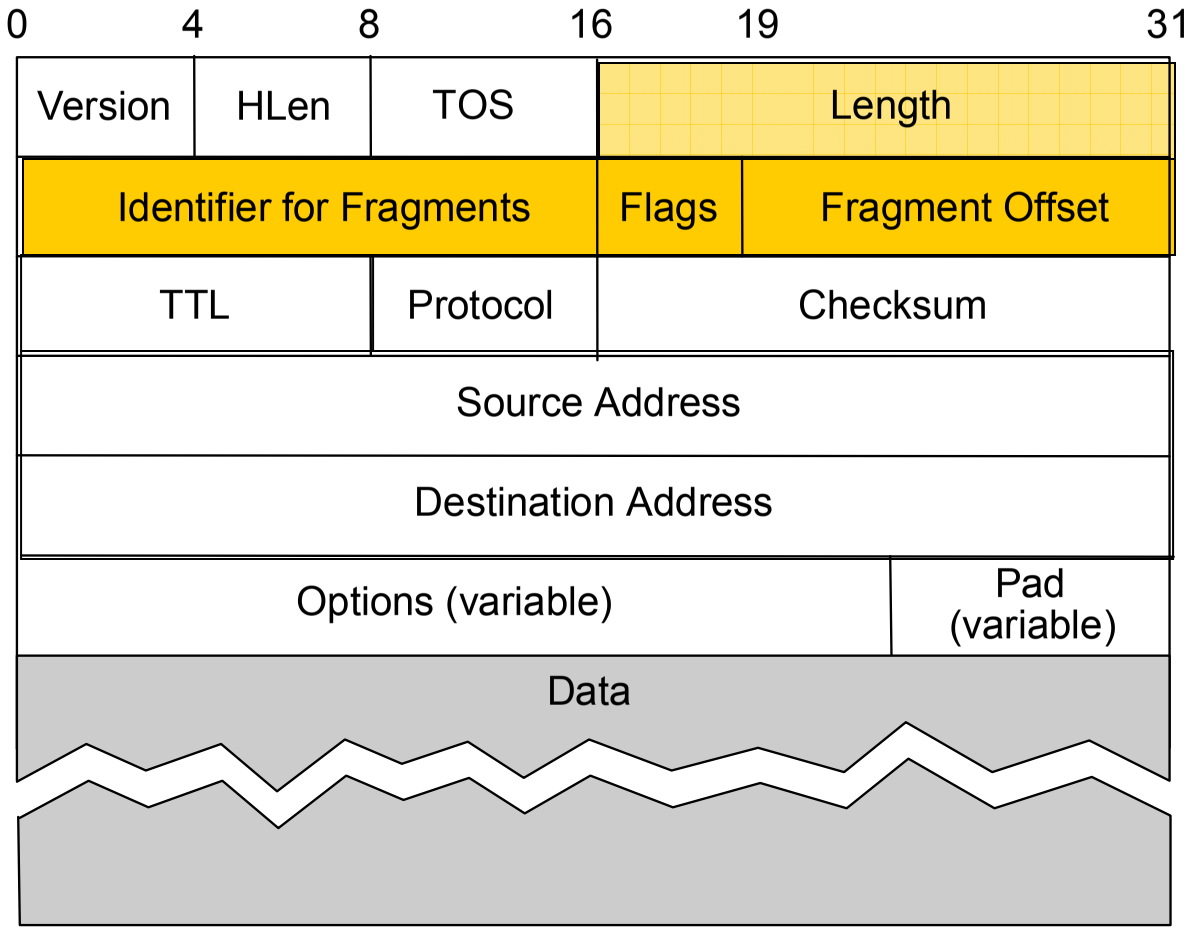
Fragmentation Issue

- Different networks may have different frame limits (MTUs)
  - Ethernet 1.5K, FDDI 4.5K
- Don't know if packet will be too big for path beforehand
  - IPv4: fragment on demand and reassemble at destination
  - IPv6: network returns error message so host can learn limit

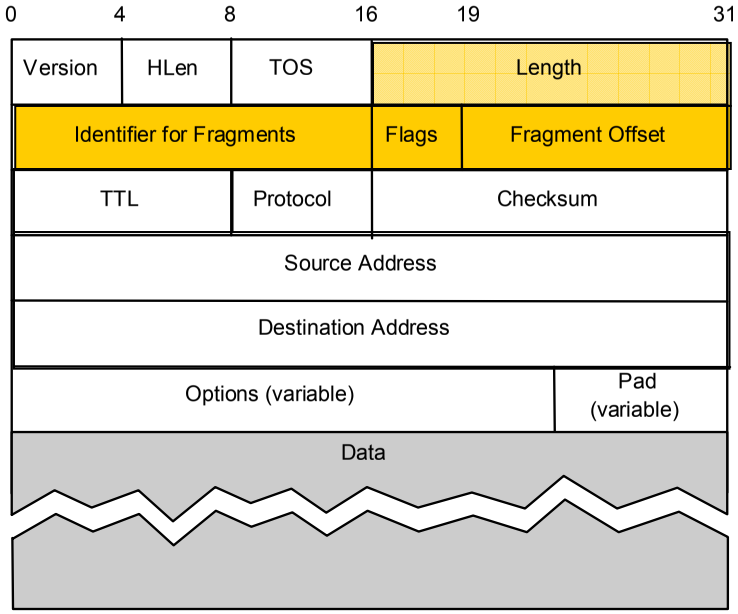


Fragment Fields

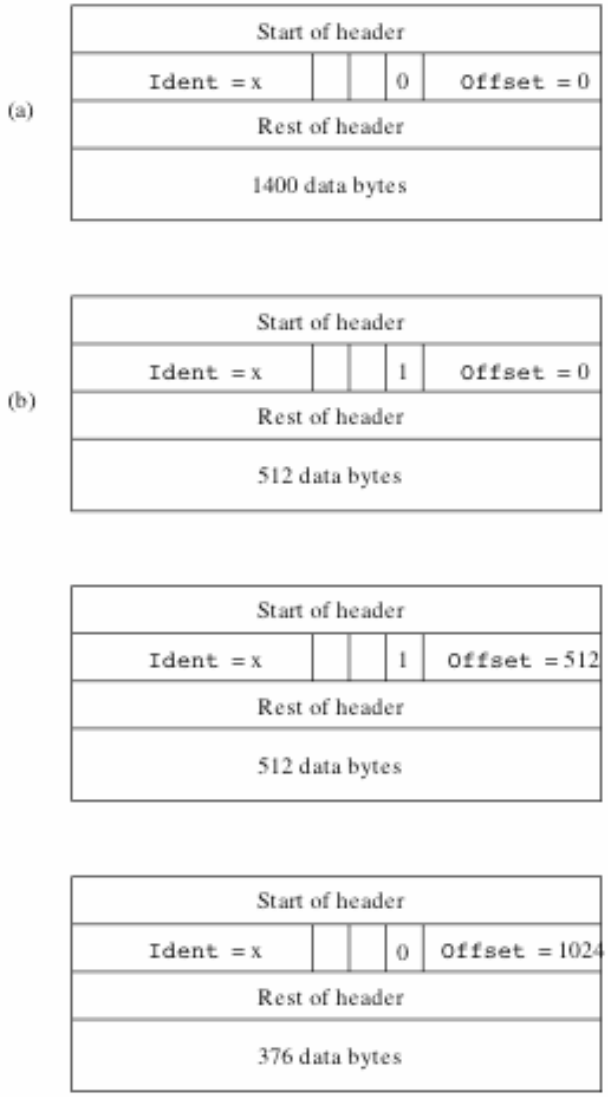
- Fragments of one packet identified by (source, dest, frag id) triple
  - Make unique
- Offset gives start, length changed
- Flags are More Fragments (MF) Don't Fragment (DF)



Fragmenting a Packet



Packet Format



Fragment Considerations

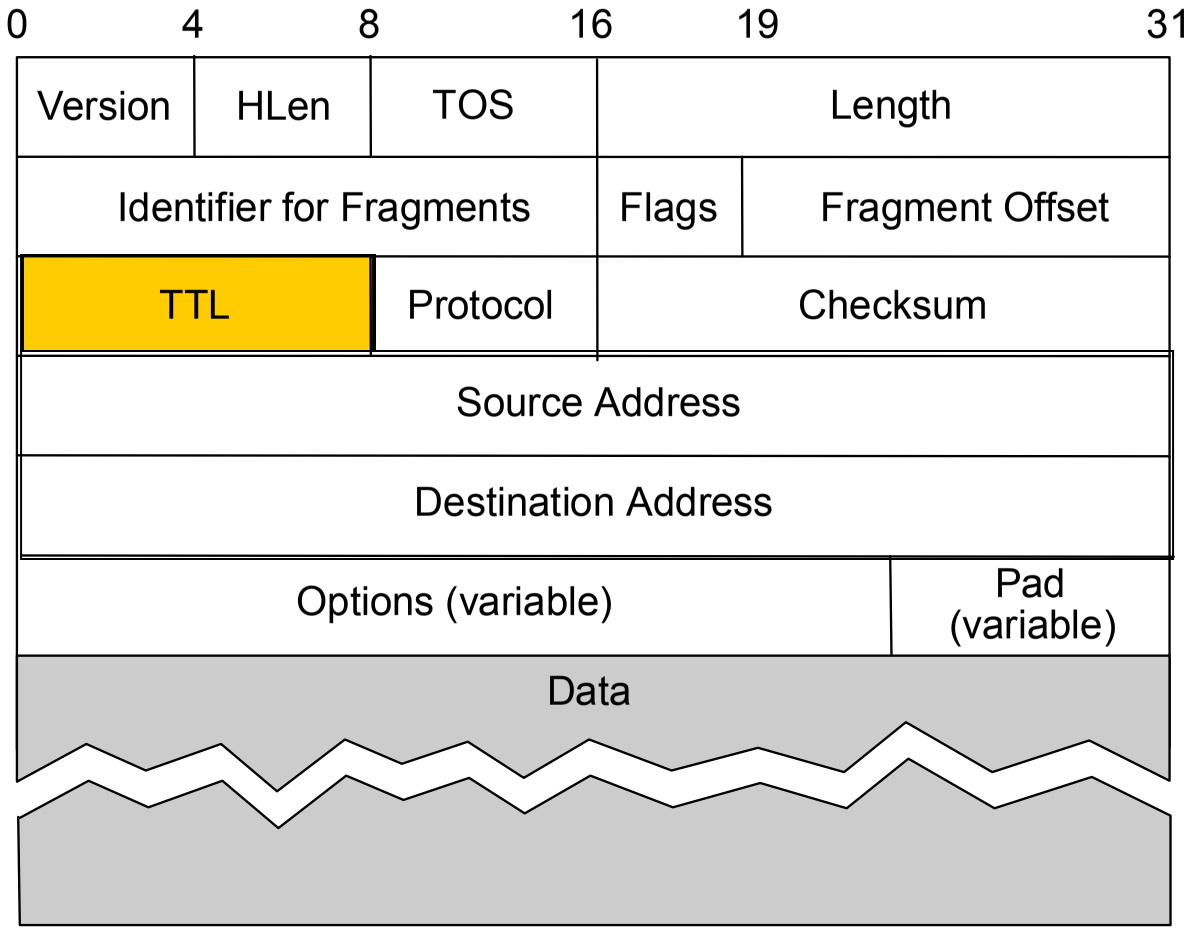
- Making fragments be datagrams provides:
  - Tolerance of reordering and duplication
  - Ability to fragment fragments
- Reassembly done at the endpoint
  - Puts pressure on the receiver, not network interior
- Consequences of fragmentation:
  - Loss of any fragments causes loss of entire packet
  - Need to time-out reassembly when any fragments lost

Avoiding Fragments with Path MTU Discovery

- Path MTU is the smallest MTU along path
  - Packets less than this size don't get fragmented
  - Idea: Avoid fragmentation too by having hosts learn path MTUs
- Non-option: send very small datagrams
  - Overly conservative, lots of header overhead
- Hosts send packets, routers return error if too large
  - Use DF flag
  - Hosts discover limits, can fragment at source
  - Reassembly at destination as before
- Learned lesson from IPv4, streamlined in IPv6

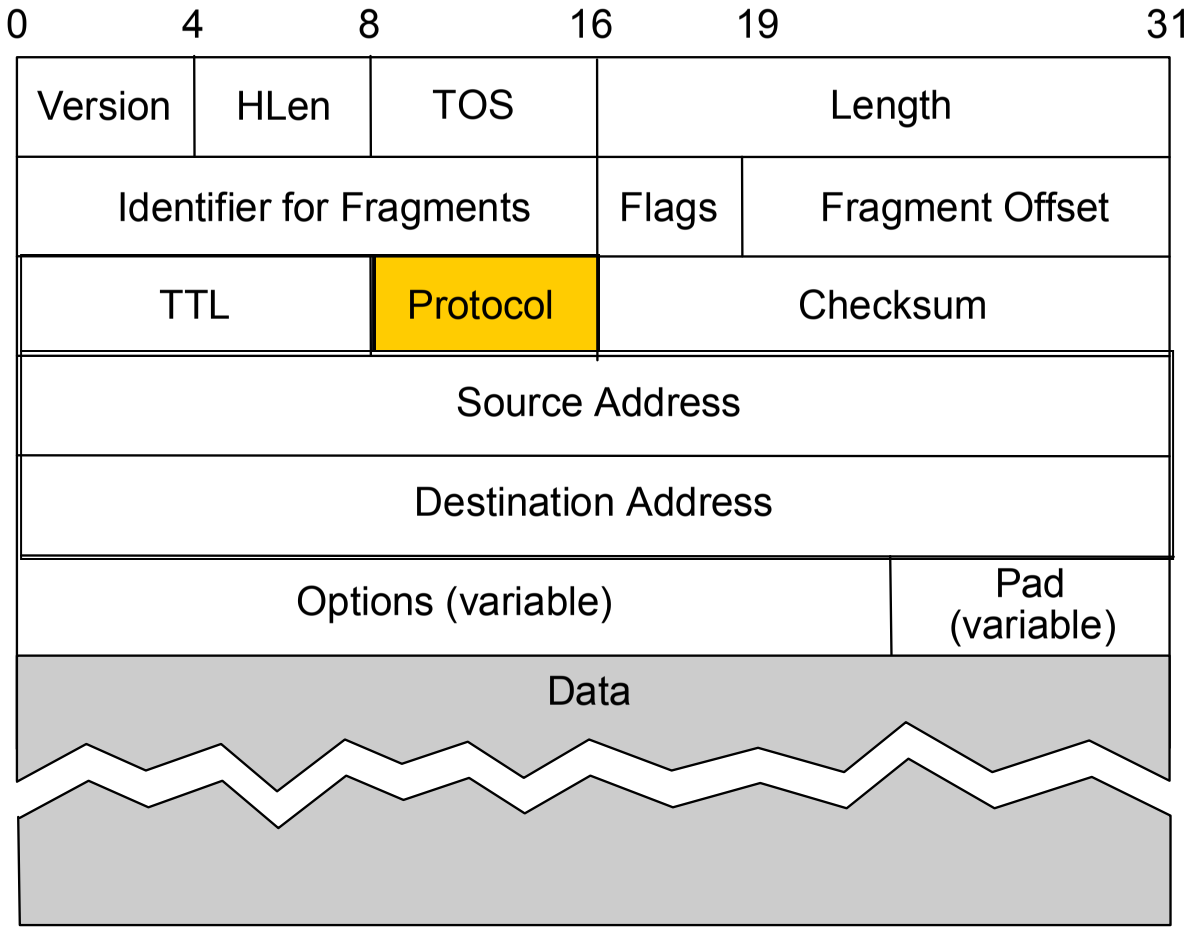
IPv4 Header Fields …

- Time To Live
- Decremented by router and packet discarded if = 0
- Prevents immortal packets



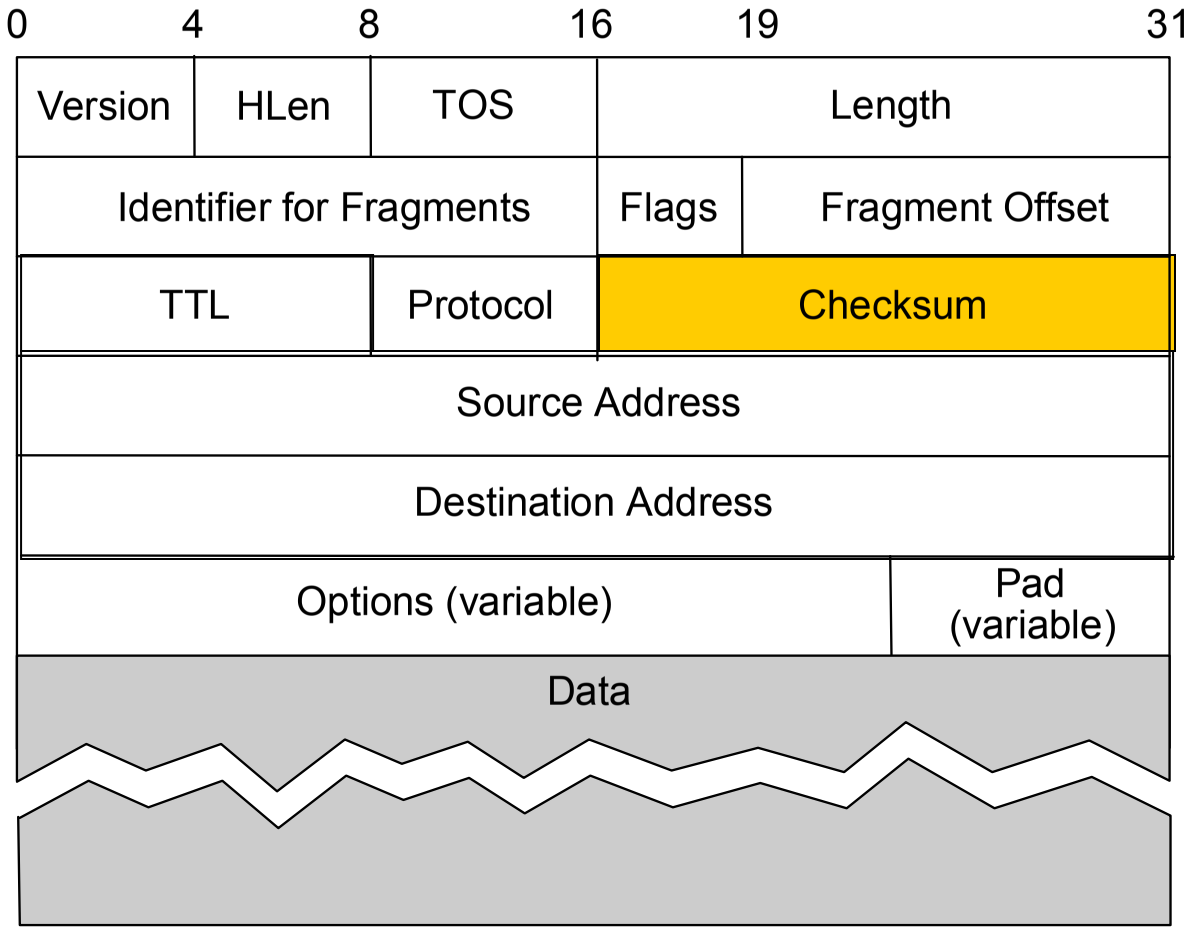
IPv4 Header Fields …

- Identifies higher layer protocol
  - E.g., TCP, UDP



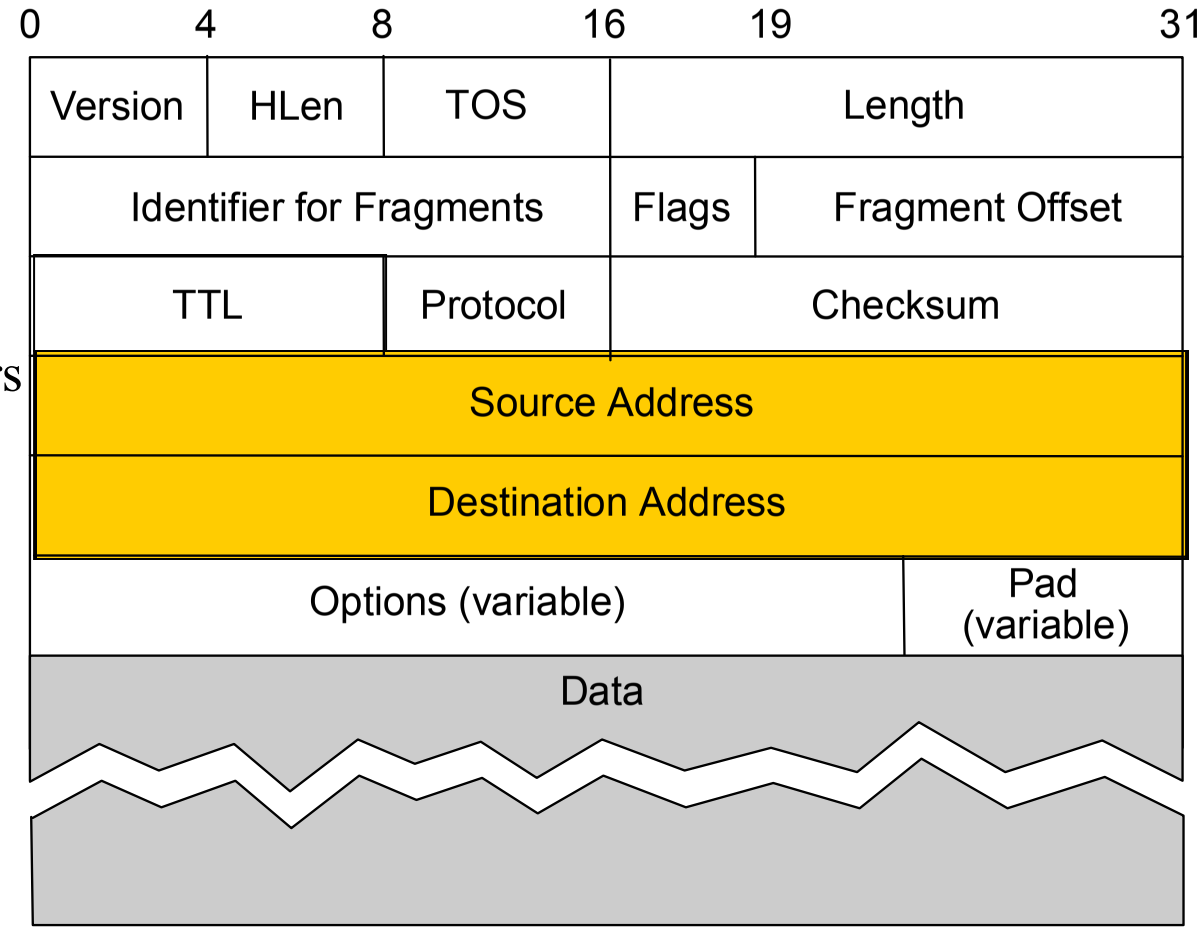
IPv4 Header Fields …

- Header checksum
- Recalculated by routers (TTL drops)
- Doesn't cover data
- Disappears for IPv6



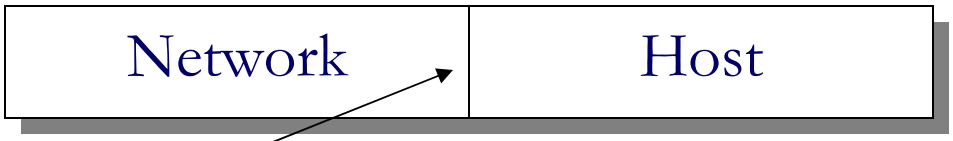
IPv4 Header Fields …

- Source/destination addresses
  - Not Ethernet
- Unchanged by routers
- Not authenticated by default



IP Addresses, e.g., 192.168.1.1

- 32 bits, hierarchical, conceptually split into 2 parts:

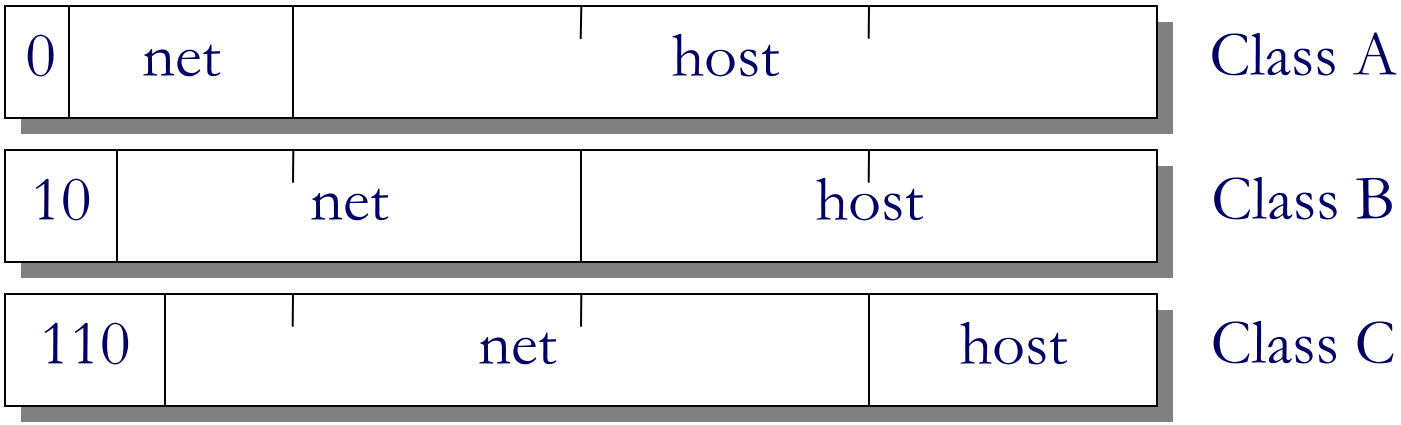


  - Flexible boundary

11111111	11111111	11000000	00000000	mask
11101000	01010101	01000000	00000000	address

- Host must learn its address, usually via DHCP
  - Unlike Ethernet addresses, which typically are burned into ROM

Old-style IP Address Classes



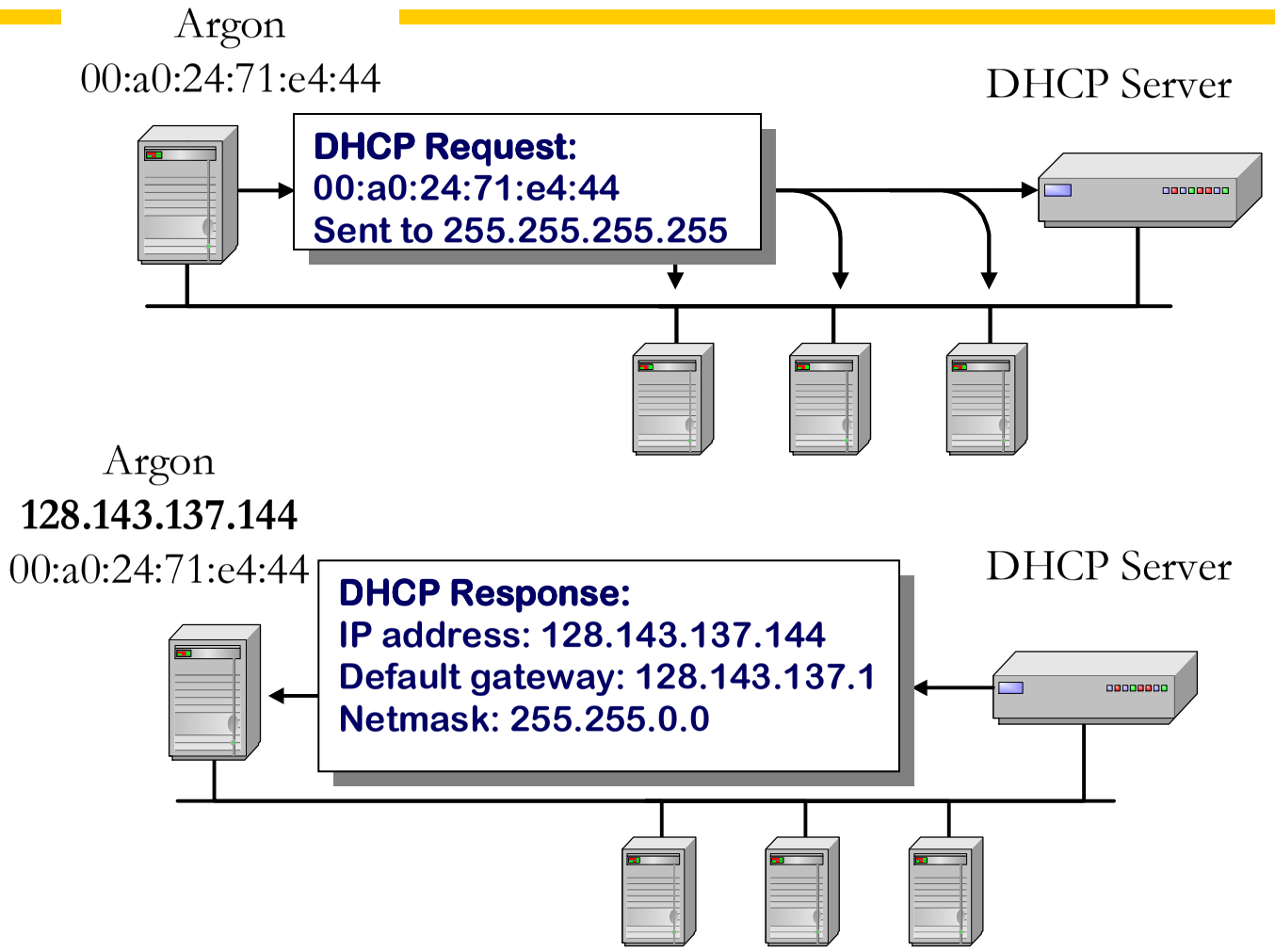
- Network mask defined as part of the address
  - Three sizes, class A, B and C, for different size networks.
- Now Classless Interdomain Routing (CIDR) [previous]
  - Adds flexibility; network mask is carried in the routing protocol
  - Internet routes are to variable length "prefixes," e.g., MIT = 18.0.0.0/8
- Also notion of subnet
  - Mask used to decide if host is on a network

Dynamic Host Configuration Protocol (DHCP)

- Q: How does a host get an IP address?
- A: DHCP, designed in 1993

- DHCP is widespread for the dynamic assignment of IP addresses, e.g., CSE, your cable company, …
- Host broadcasts a request; DHCP server responds with an IP
- Extensions:
  - Supports temporary allocation (“leases”) of IP addresses
  - DHCP client can acquire all IP configuration parameters

DHCP Interaction (simplified)



IP Forwarding

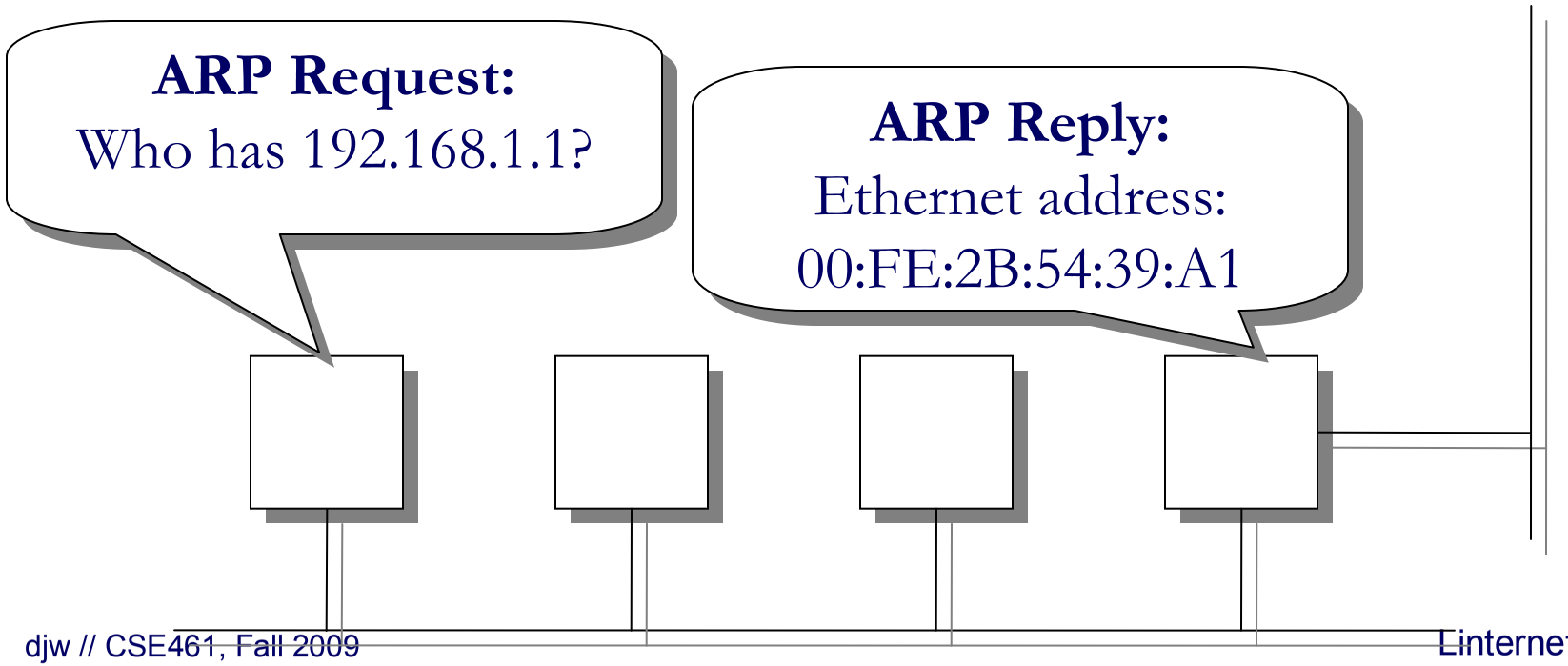
Destination	Gateway
Default (0/0)	192.168.1.1
192.168.1.0/24	Link #4

My PC's routing table (netstat –r)

- Match on the longest prefix
  - Routers in the Internet may have 100s of 1000s of prefixes
- Example:
  - Send a packet to my printer (192.168.1.254)
  - Send a packet to cnn (157.166.224.25)

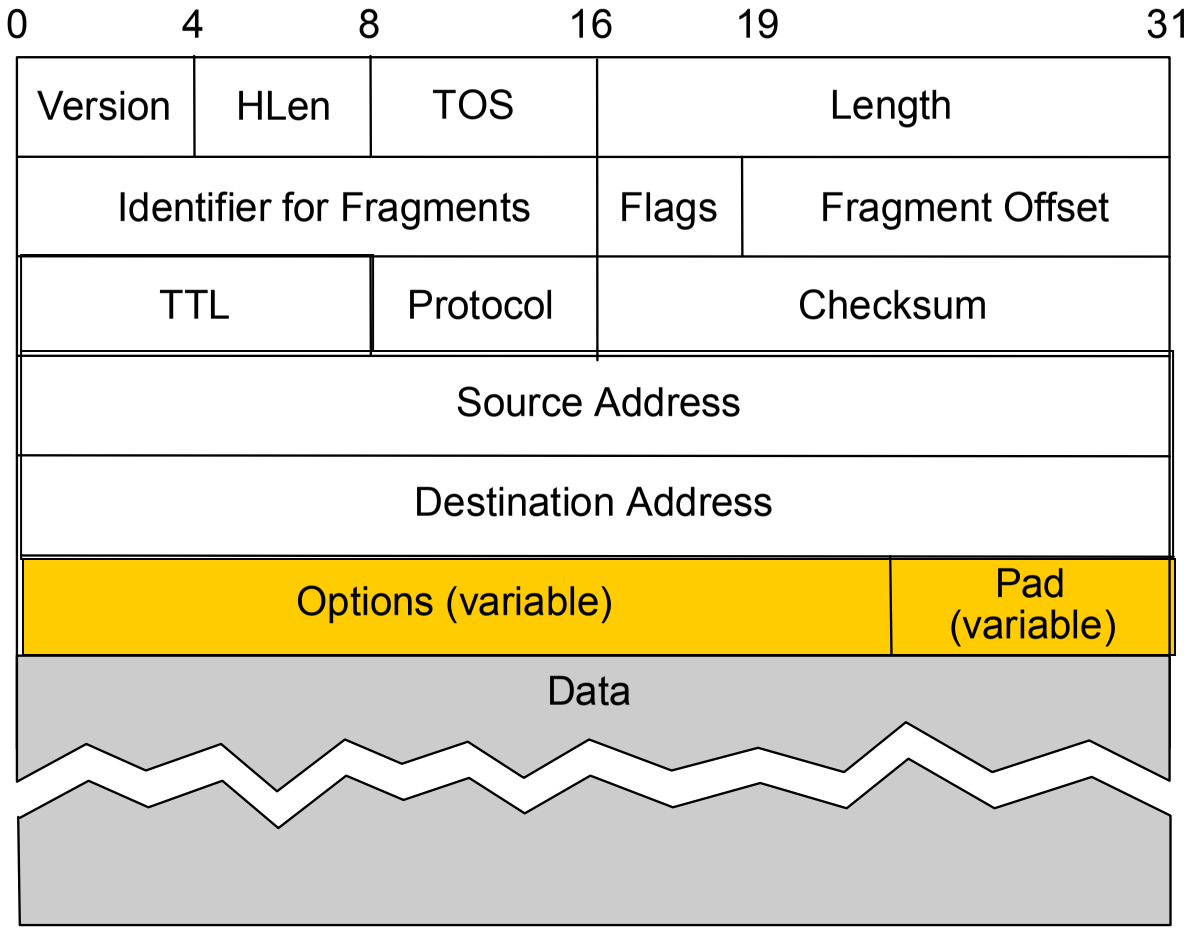
Address Resolution Protocol (ARP)

- Problem: We know a destination IP address, but how do we find the actual device on the LAN with that address?
- Solution: ARP maps local IP to local Ethernet addresses



IPv4 Header Fields …

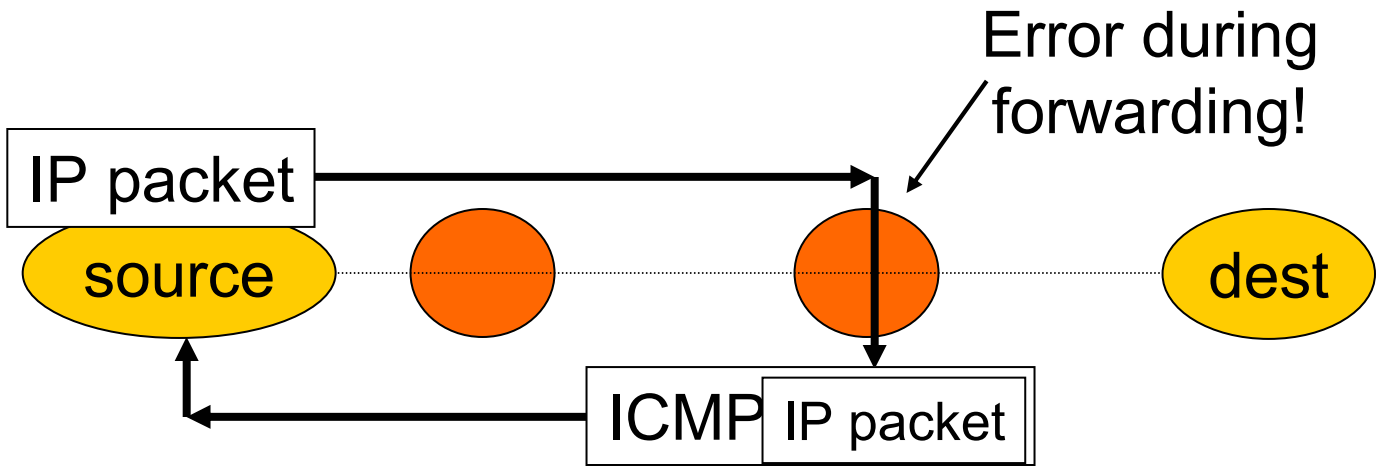
- IP options indicate special handling
  - Timestamps
  - "Source" routes
- Rarely used …



ICMP

- What happens when things go wrong?
  - Need a way to test/debug a large, widely distributed system
- ICMP = Internet Control Message Protocol (RFC792)
  - Companion to IP – required functionality
- Used for error and information reporting:
  - Errors that occur during IP forwarding
  - Queries about the status of the network

ICMP Generation



Common ICMP Messages

- Destination unreachable
  - "Destination" can be host, network, port or protocol
- Redirect
  - To shortcut circuitous routing
- TTL Expired
  - Used by the "traceroute" program
- Echo request/reply
  - Used by the "ping" program

- ICMP messages include portion of IP packet that triggered the error (if applicable) in their payload

ICMP Restrictions

- The generation of error messages is limited to avoid cascades … error causes error that causes error!

- Don't generate ICMP error in response to:
  - An ICMP error
  - Broadcast/multicast messages (link or IP level)
  - IP header that is corrupt or has bogus source address
  - Fragments, except the first

- ICMP messages are often rate-limited too.