CSE160: Computer Networks

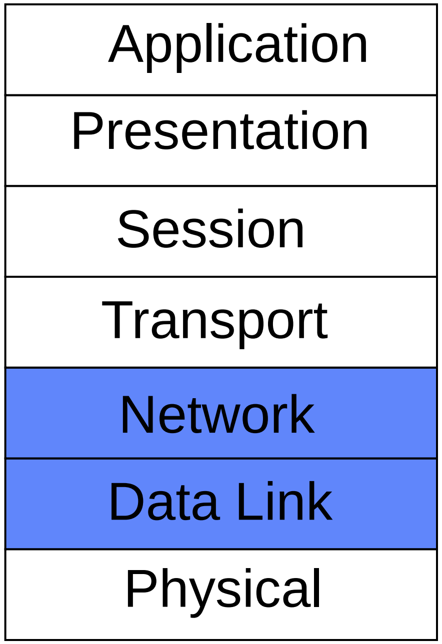
Lecture #08 – IP/ICMP and the Network Layer

2020-09-21

Professor
Alberto E. Cerpa

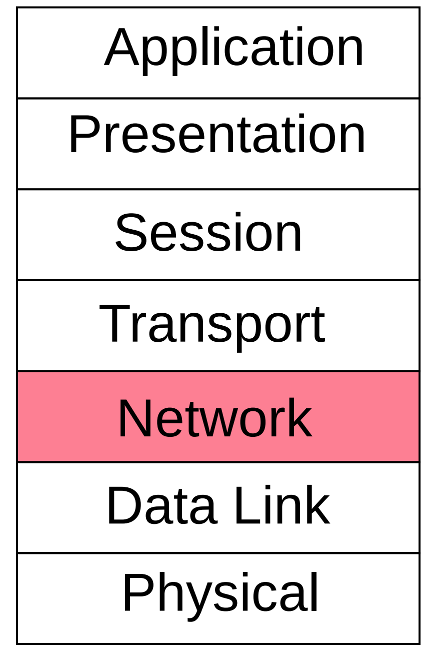
Last Time

- Focus:
  - What to do when one shared LAN isn't big enough?
- Interconnecting LANs
  - Bridges and LAN switches
  - But there are limits …



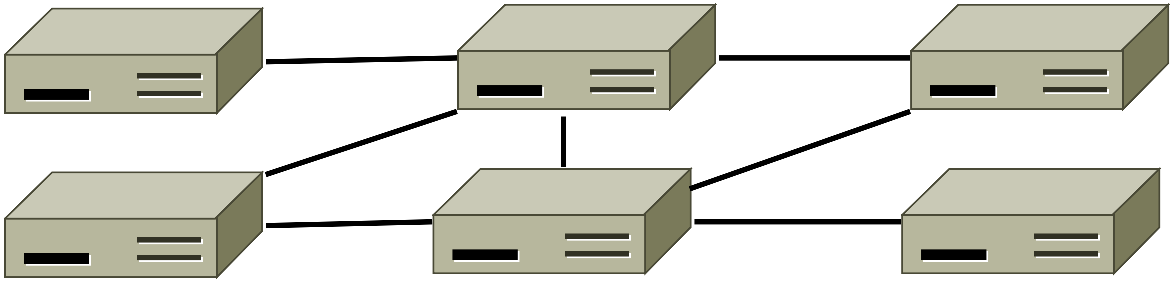
This Lecture

- Focus:
  - How do we build large networks?
- Introduction to the Network layer
  - Service models
  - Internetworks
  - IP
  - Packet Fragmentation and Path Discovery
  - ICMP



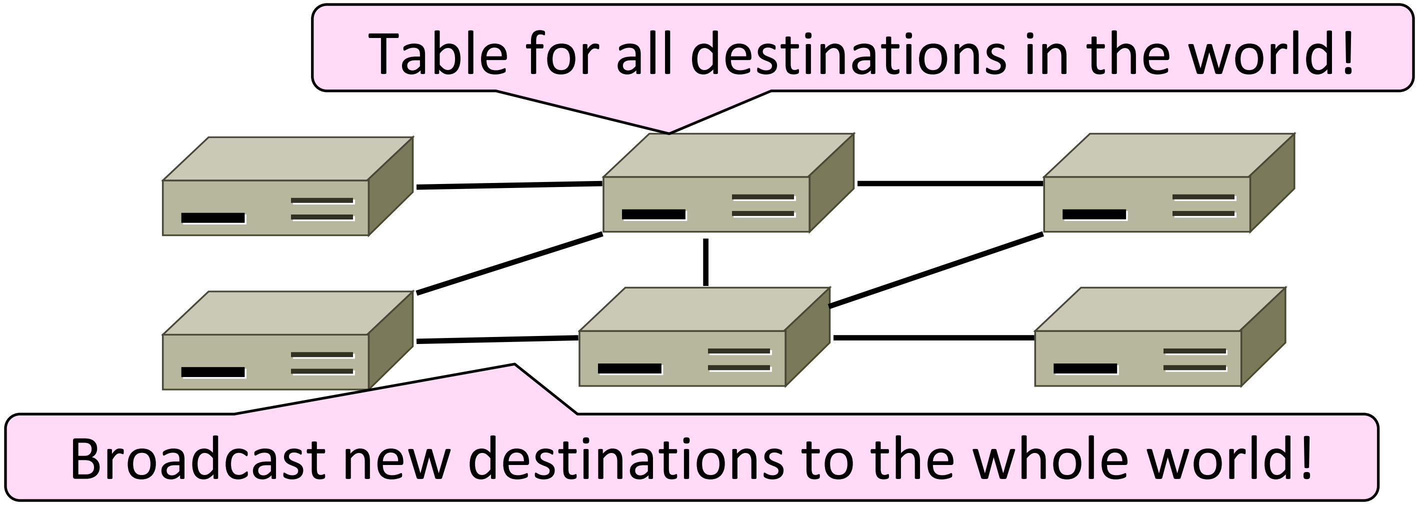
Why do we need a Network Layer?

- We can already build networks with links and switches and send frames between hosts …



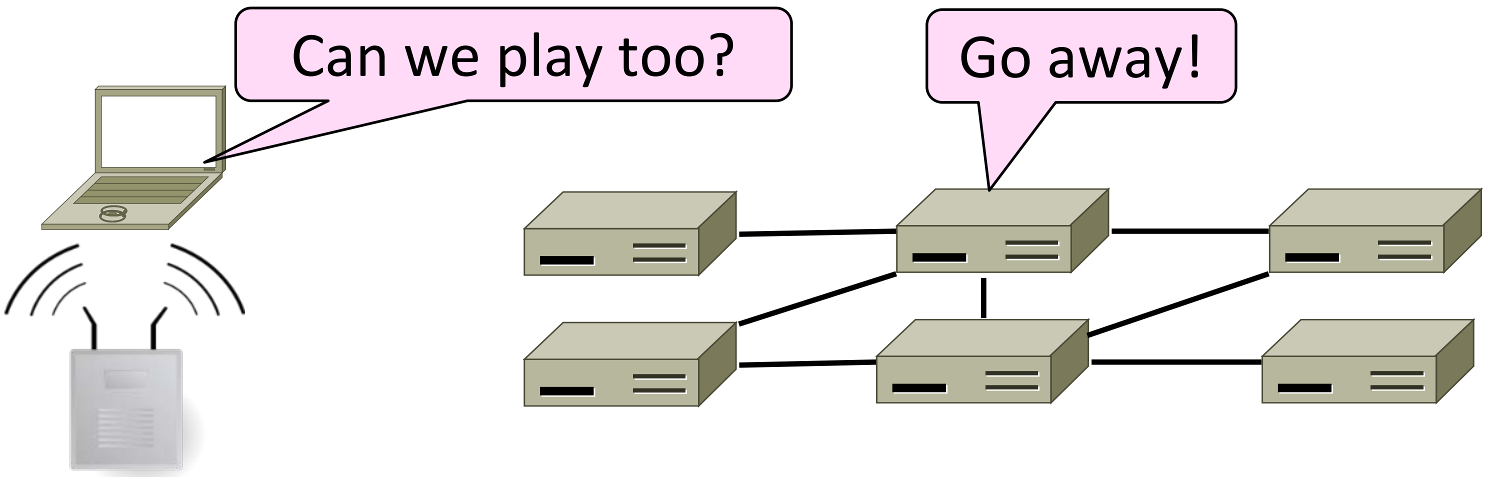
Shortcomings of Switches

1. Don’t scale to large networks
   - Blow up routing table
   - Broadcast over the Internet!
   - Convergence time of SPT ~ to network diameter



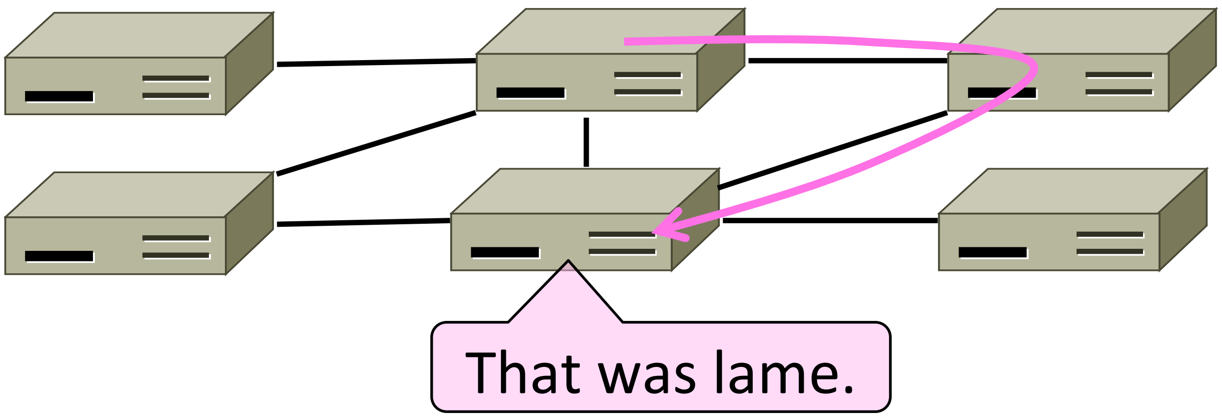
Shortcomings of Switches

2. Don't work across more than one link layer technology
   – Hosts on Ethernet + 4G + 802.11



Shortcomings of Switches

3. Don't give much traffic control
   – Want to plan routes/bandwidth

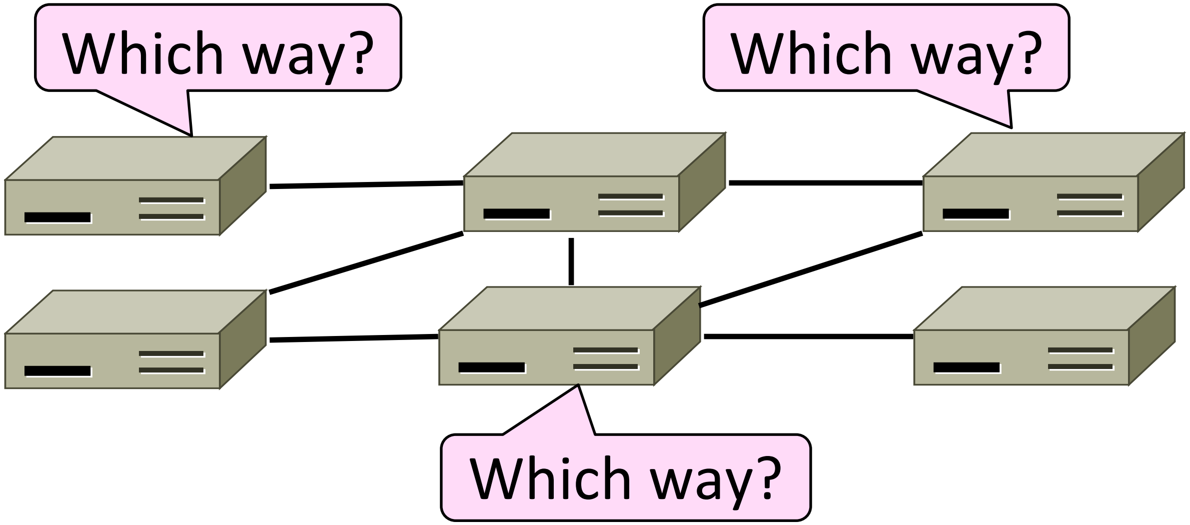


Network Layer Approach

- Scaling:
  - Hierarchy, in the form of hierarchical routing
  - Hierarchy, in the form of IP prefixes
- Heterogeneity:
  - IP for internetworking
- Bandwidth Control:
  - Lowest-cost routing
  - Later QOS (Quality of Service)

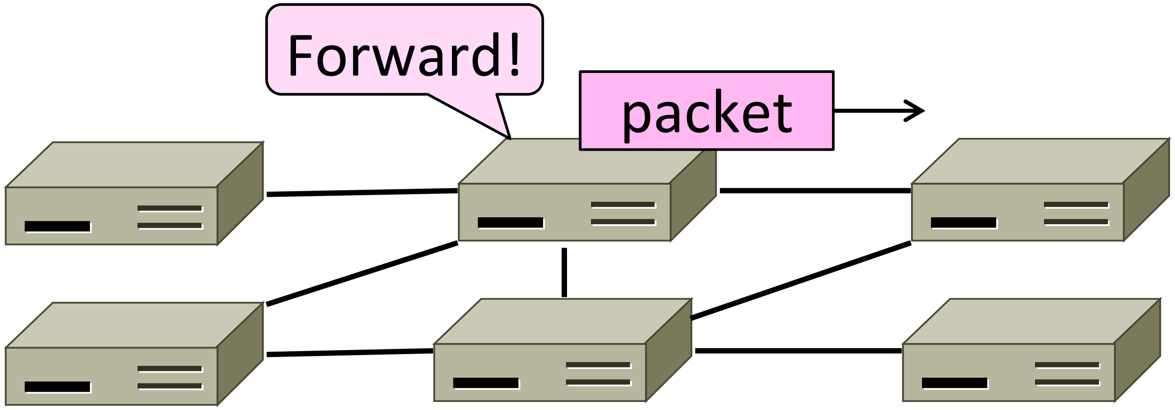
Routing vs Forwarding

- Routing is the process of deciding in which direction to send traffic
  - Network wide (global) and expensive



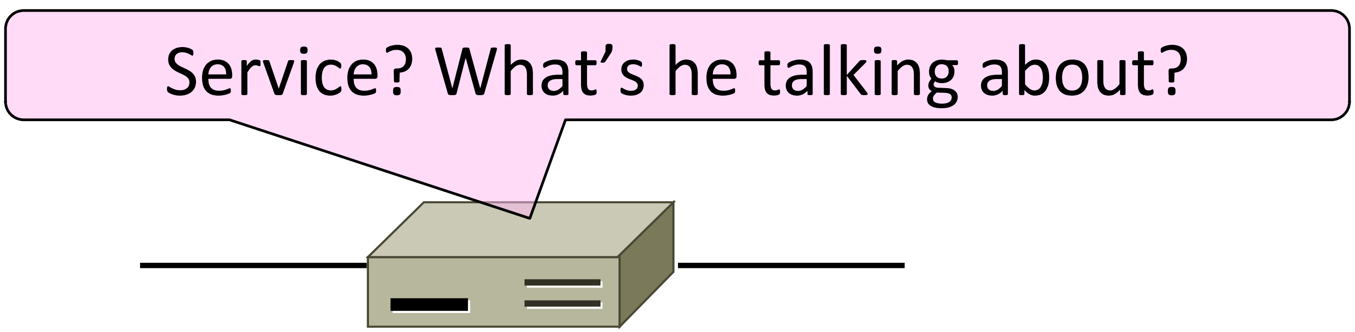
Routing vs Forwarding (2)

- Forwarding is the process of sending a packet on its way
  - Node process (local) and fast



Networking Services

- What kind of service does the Network Layer provide to the Transport Layer?
  - How is it implemented at routers?



Two Network Service Models

- Datagram delivery: postal service
  - connectionless, best-effort or unreliable service
  - Network can’t guarantee delivery of the packet
  - Each packet from a host is routed independently
  - Example: IP

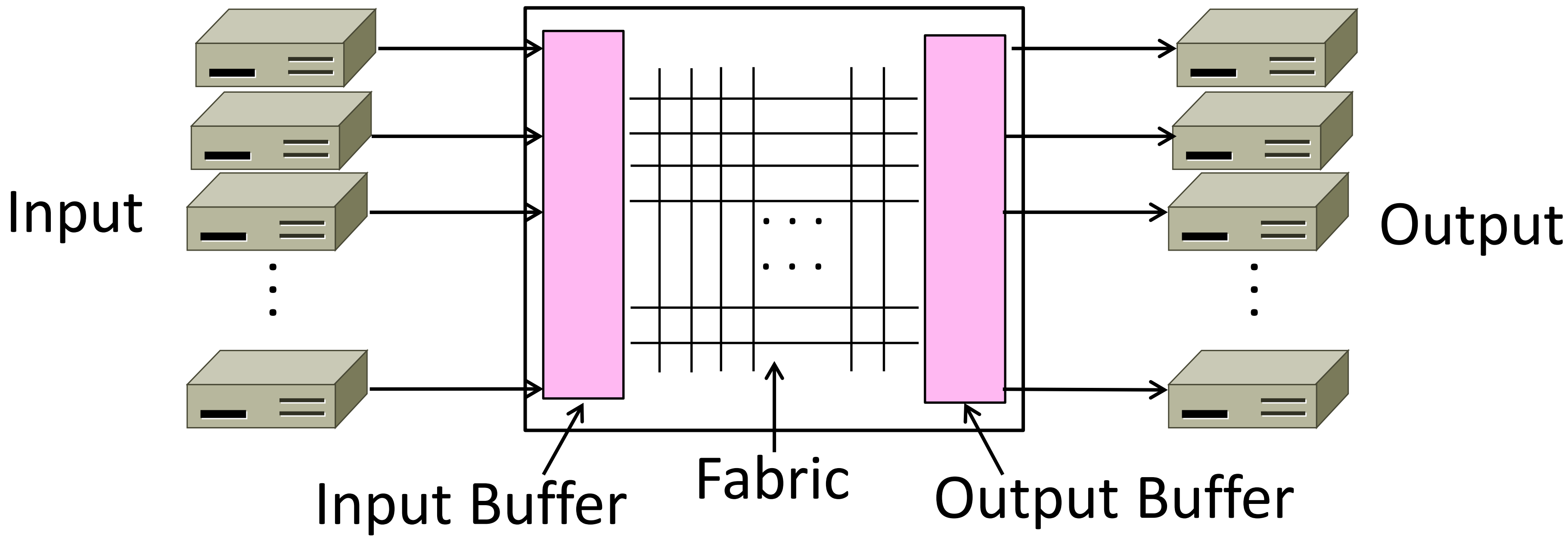
- Virtual circuit models: telephone
  - connection-oriented service
  - Signaling: connection establishment, data transfer, teardown
  - All packets from a host are routed the same way (router state)
  - Example: ATM, Frame Relay, X.25, SS7, MPLS

Store-and-Forward Packet Switching

- Both models are implemented with store-and-forward packet switching
  - Routers receive a complete packet, storing it temporarily if necessary before forwarding it onwards
  - We use statistical multiplexing to share link bandwidth over time

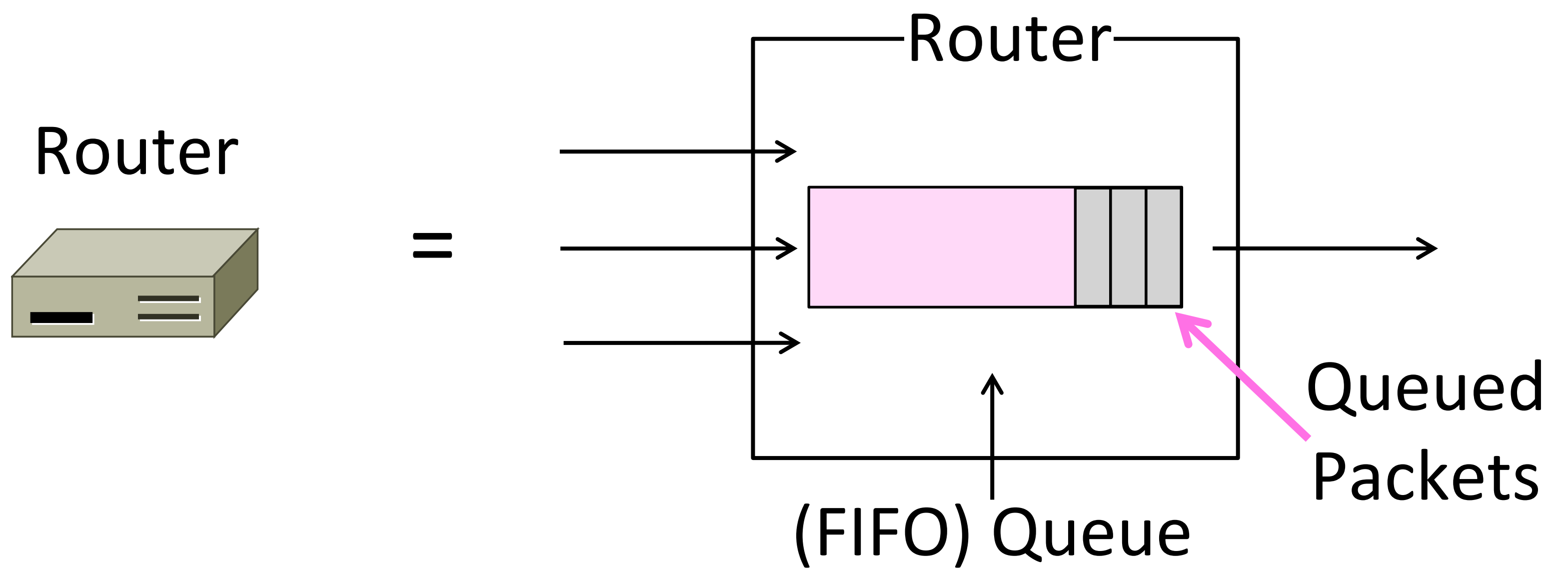
Store-and-Forward (2)

- Switching element has internal buffering for contention



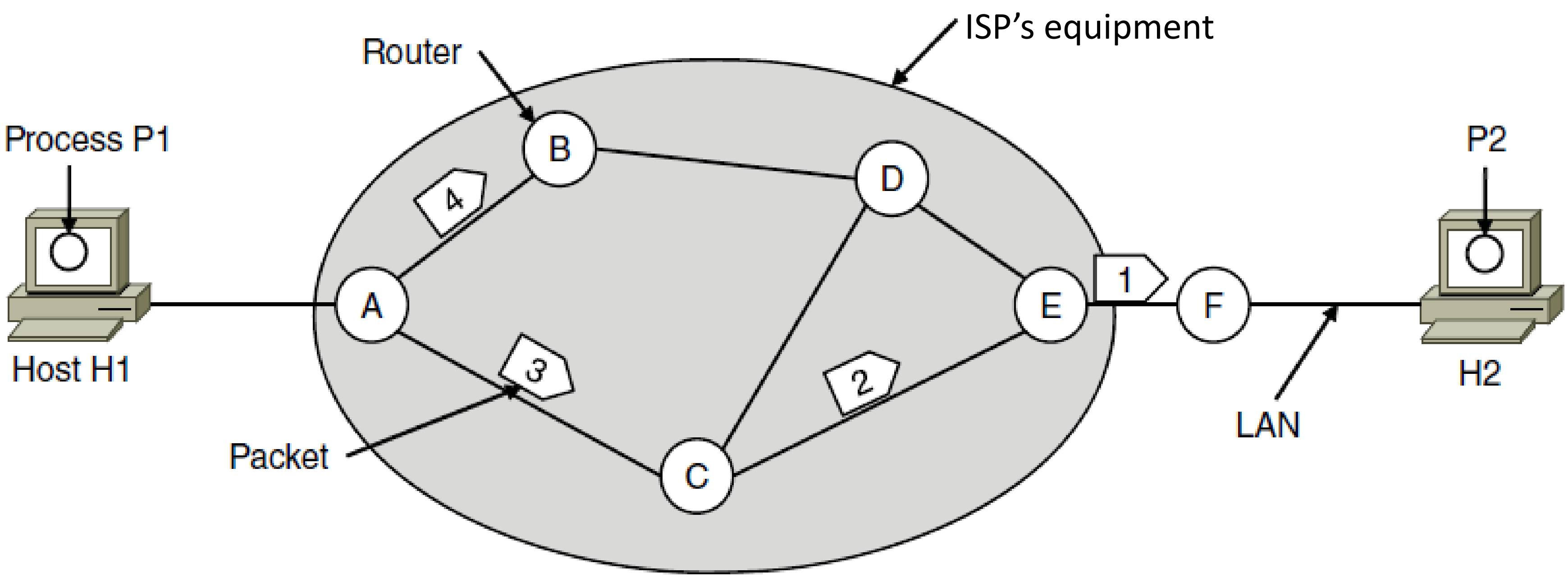
Store-and-Forward (3)

- Simplified view with per port output buffering
  - Buffer is typically a FIFO (First In First Out) queue
  - If full, packets are discarded (congestion, later)



Datagram Model

- Packets contain a destination address; each router uses it to forward packets, maybe on different paths



Datagram Model (2)

- Each router has a forwarding table keyed by address
  - Gives next hop for each destination address; may change

A’s table (initially)

A	
B	B
C	C
D	B
E	C
F	C

Dest. Line

A’s table (later)

A	
B	B
C	C
D	B
E	B
F	B

C’s Table

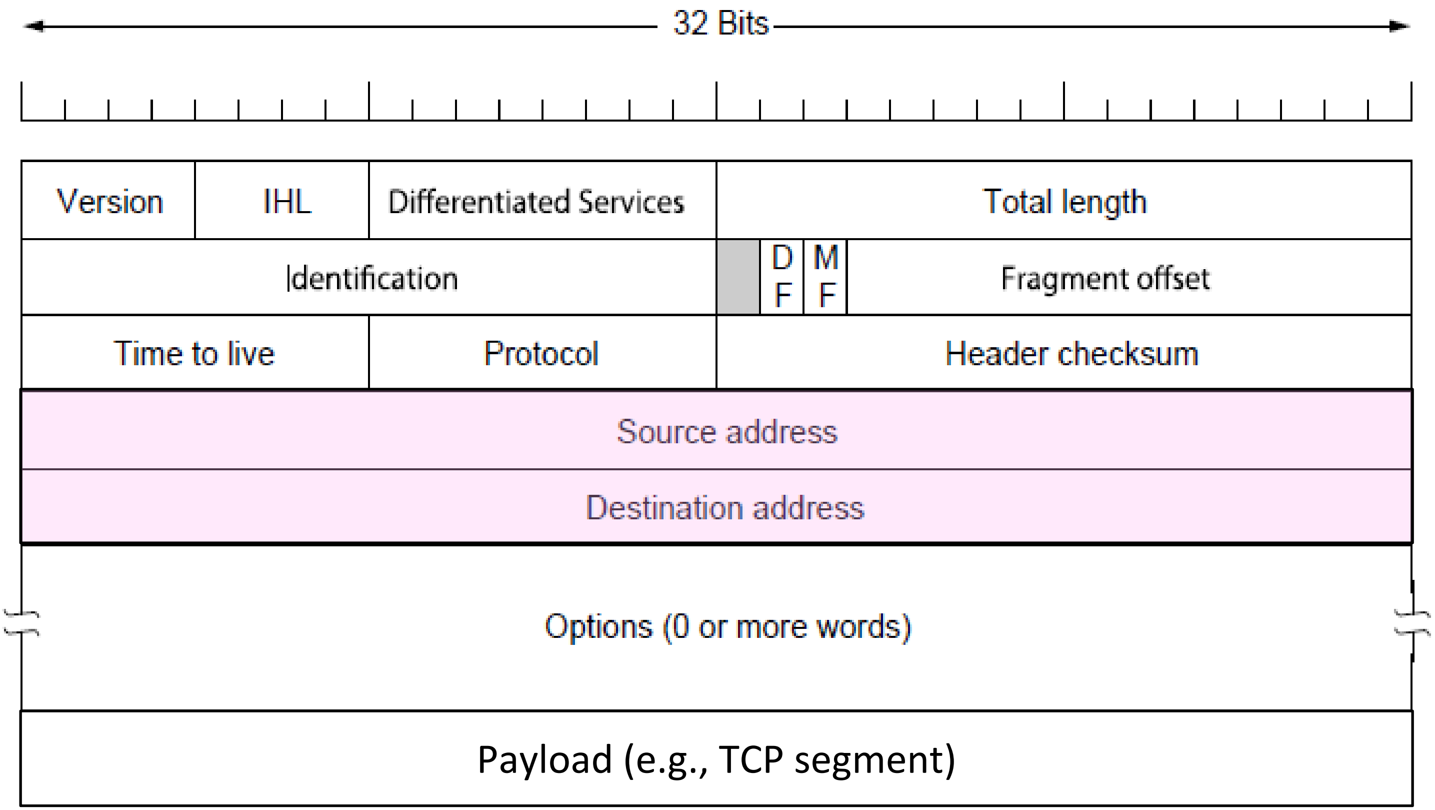
A	A
B	A
C	
D	E
E	E
F	E

E’s Table

A	C
B	D
C	C
D	D
E	
F	F

Internet Protocol (IP)

- Network layer of the Internet, uses datagrams
  - IPv4 carries 32 bit addresses on each packet (often 1.5 KB)



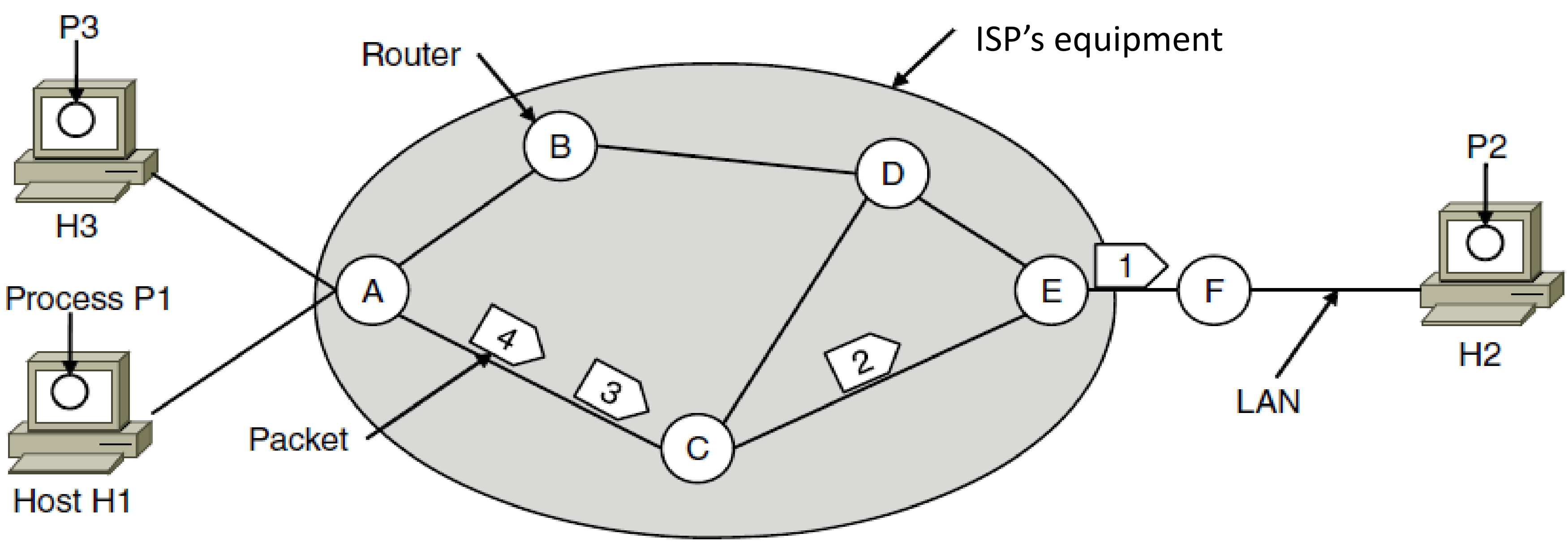
Virtual Circuit Model

- Three phases:
  1. Connection establishment, circuit is set up
     - Path is chosen, circuit information stored in routers
  2. Data transfer, circuit is used
     - Packets are forwarded along the path
  3. Connection teardown, circuit is deleted
     - Circuit information is removed from routers

- Just like a telephone circuit, but virtual in that no bandwidth need be reserved; statistical sharing of links

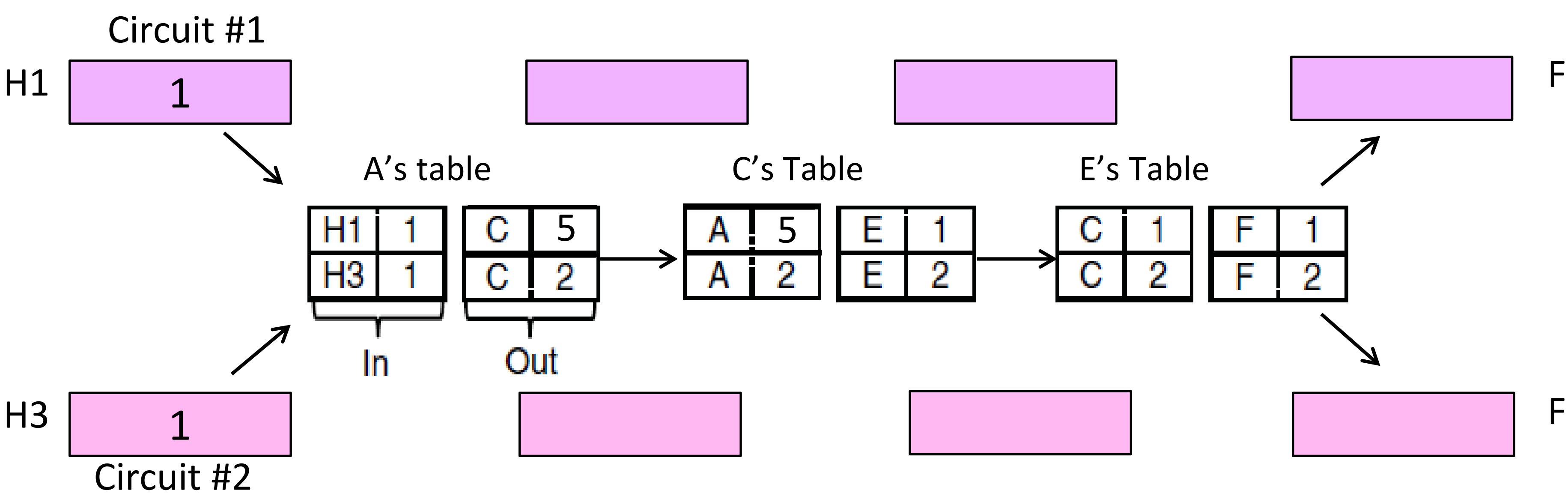
Virtual Circuits (2)

- Packets contain a <u>short label</u> to identify the circuit
  - Labels don't have global meaning, only unique for a link



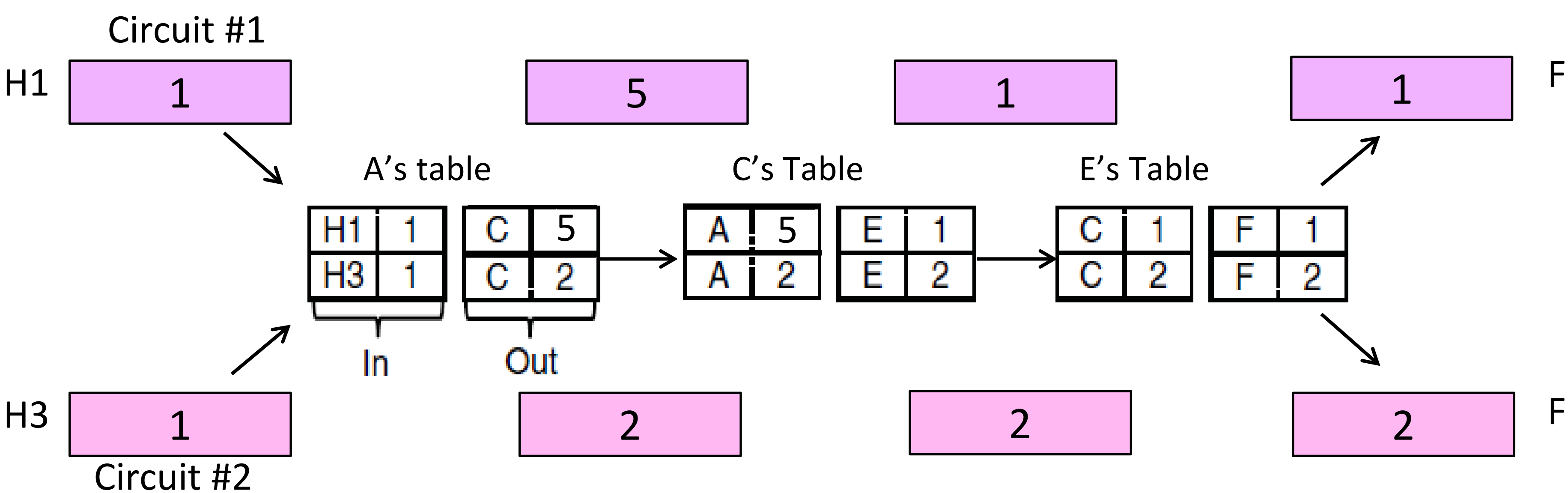
Virtual Circuits (3)

- Each router has a forwarding table keyed by circuit
  - Gives output line and next label to place on packet



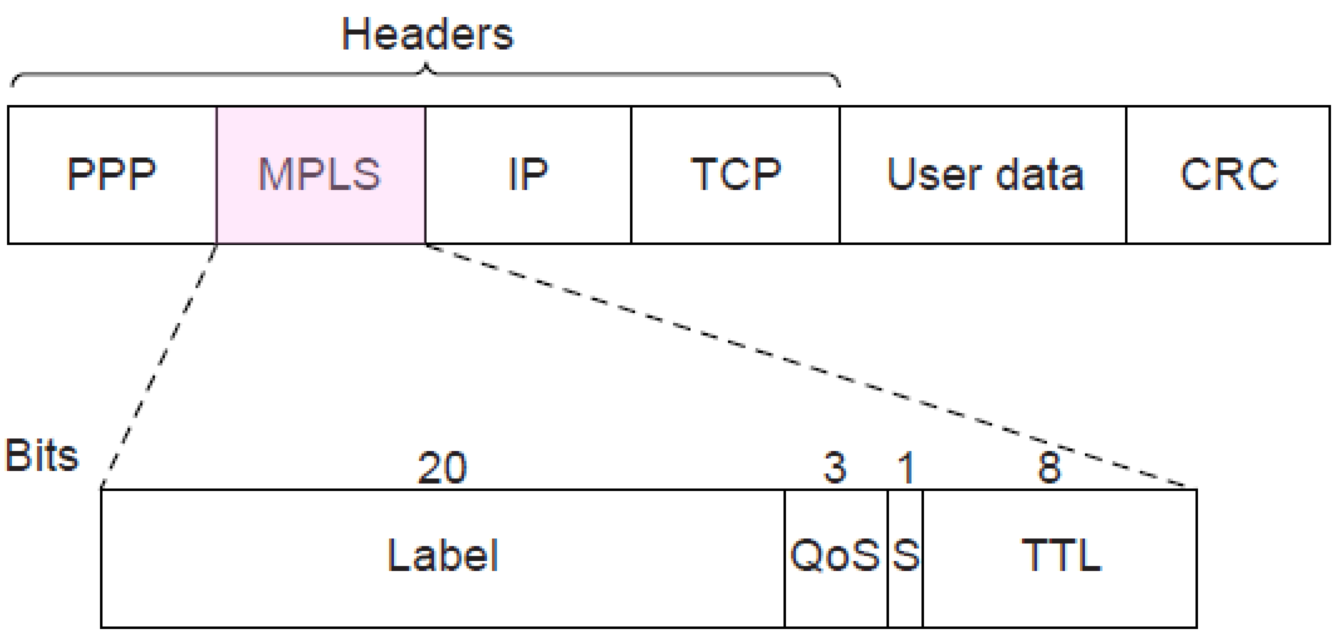
Virtual Circuits (4)

- Each router has a forwarding table keyed by circuit
  - Gives output line and next label to place on packet



Multi-Protocol Label Switching (MPLS)

- A virtual-circuit like technology widely used by ISPs
  - ISP sets up circuits inside their backbone ahead of time
  - ISP adds MPLS label to IP packet at ingress, undo at egress



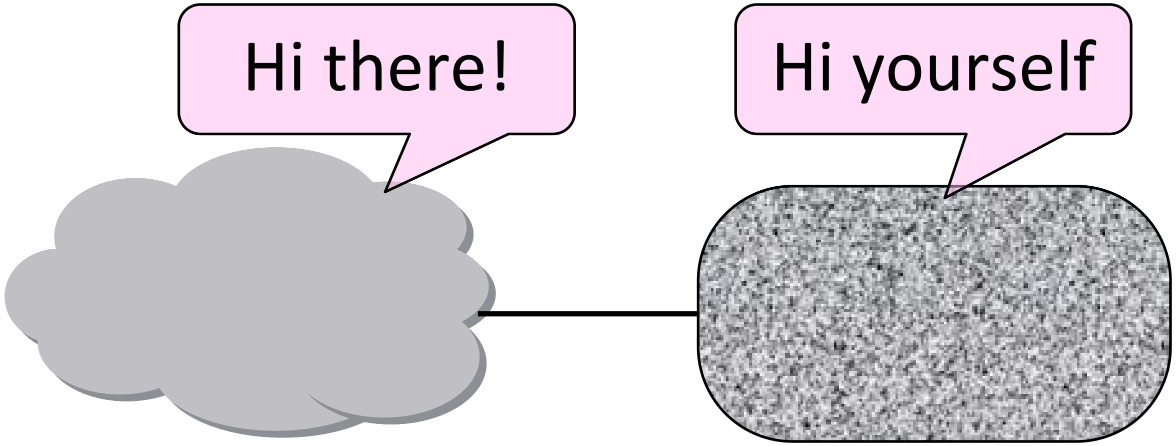
Datagrams vs Virtual Circuits

- Complementary strengths

Issue	Datagrams	Virtual Circuits
Setup phase	Not needed	Required
Router state	Per destination	Per connection
Addresses	Packet carries full address	Packet carries short label
Routing	Per packet	Per circuit
Failures	Easier to mask	Difficult to mask
Quality of service	Difficult to add	Easier to add

Internetworking

- How do we connect different networks together?
  - This is called internetworking
  - We'll look at how IP does it



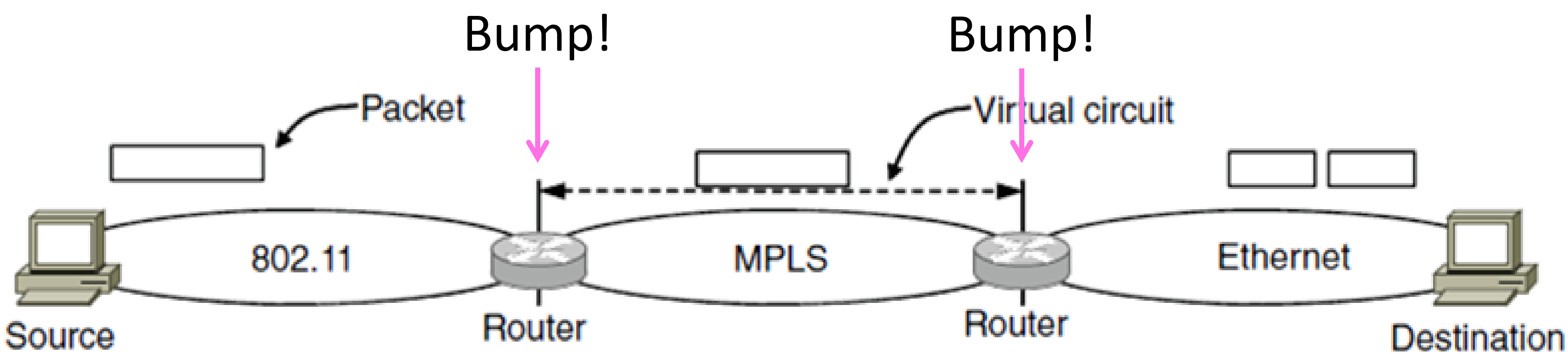
How Networks May Differ?

- Basically, in a lot of ways:
  - Service model (datagrams, VCs)
  - Addressing (what kind)
  - QOS (priorities, no priorities)
  - Packet sizes
  - Security (whether encrypted)

- Internetworking hides the differences with a common protocol (Uh oh!)

Connecting Datagram and VC networks

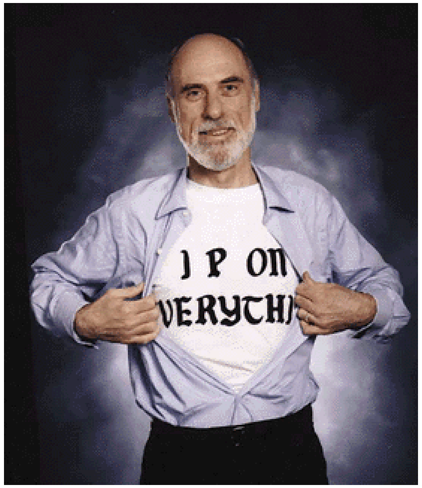
- An example to show that it's not so easy
  - Need to map destination address to a VC and vice-versa
  - A bit of a "road bump", e.g., might have to set up a VC



Internetworking – Cerf and Kahn

- Pioneers: Cerf and Kahn
  - "Fathers of the Internet"
  - In 1974, later led to TCP/IP
- Tackled the problems of interconnecting networks
  - Instead of mandating a single style networking technology

Vint Cerf



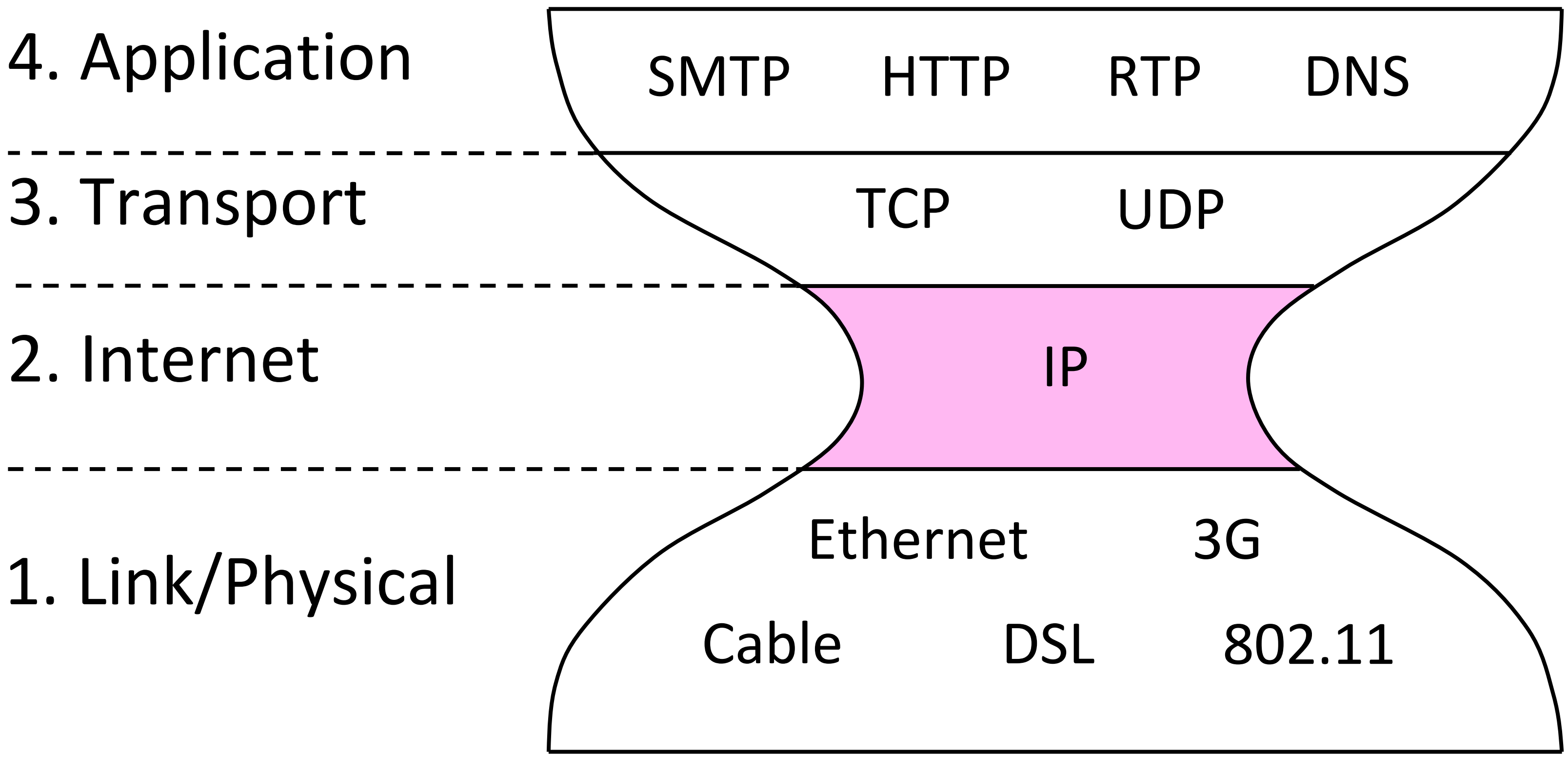
© 1996-2019 PCMag Digital Group

Bob Kahn

IEEE © 2019

Internet Reference Model

- Internet Protocol (IP) is the "narrow waist" of the Internet
  - Supports many different links below and apps above



IP as a Lowest Common Denominator

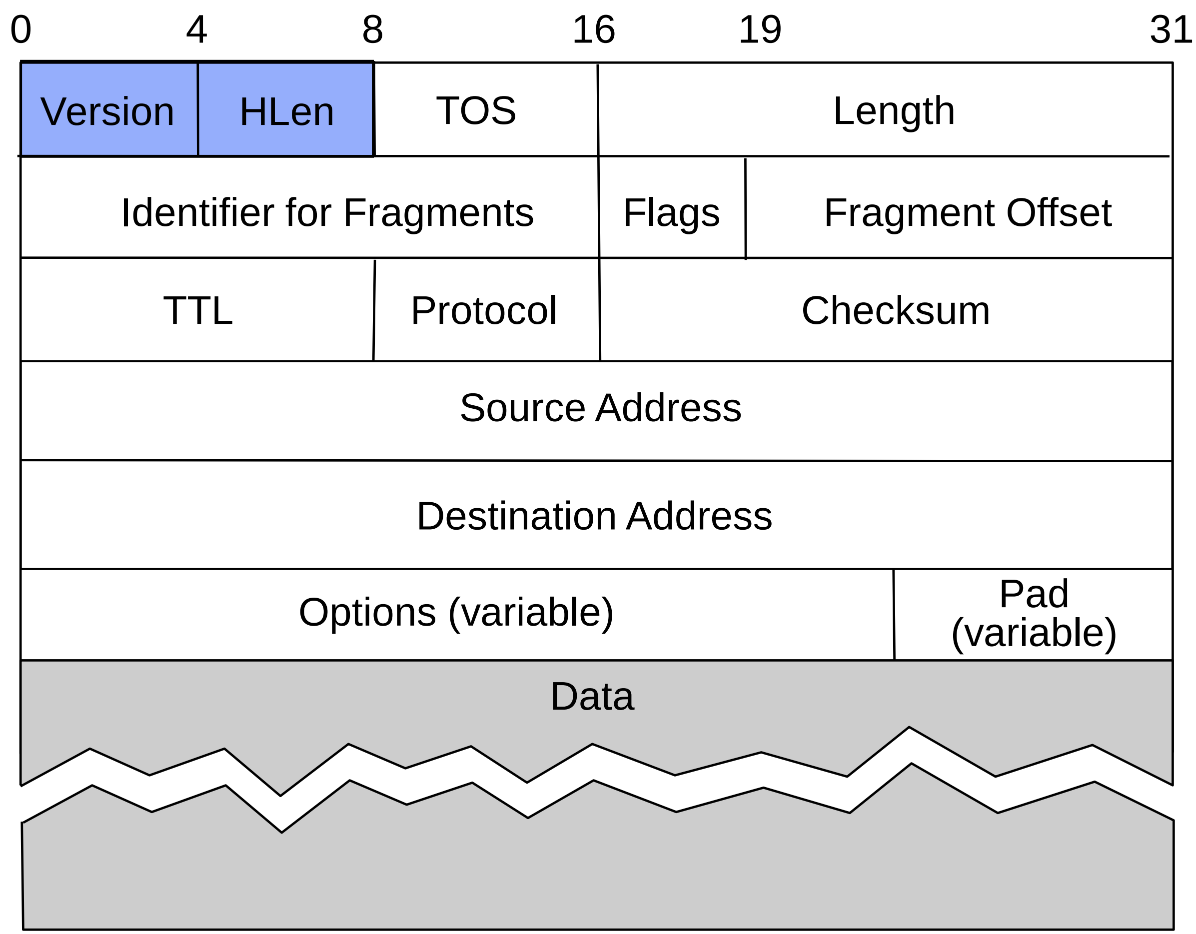
- Suppose only some networks support QOS or security etc.
  - Difficult for internetwork to support
- Pushes IP to be a “lowest common denominator”
  - Asks little of lower-layer networks
  - Gives little as a higher layer service

Internet Protocol (IP)

- IP (RFC791) defines a datagram “best effort” service
  - May have losses, reordering, duplication, and errors!
  - Currently IPv4 (IP version 4), IPv6 being adopted
- 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

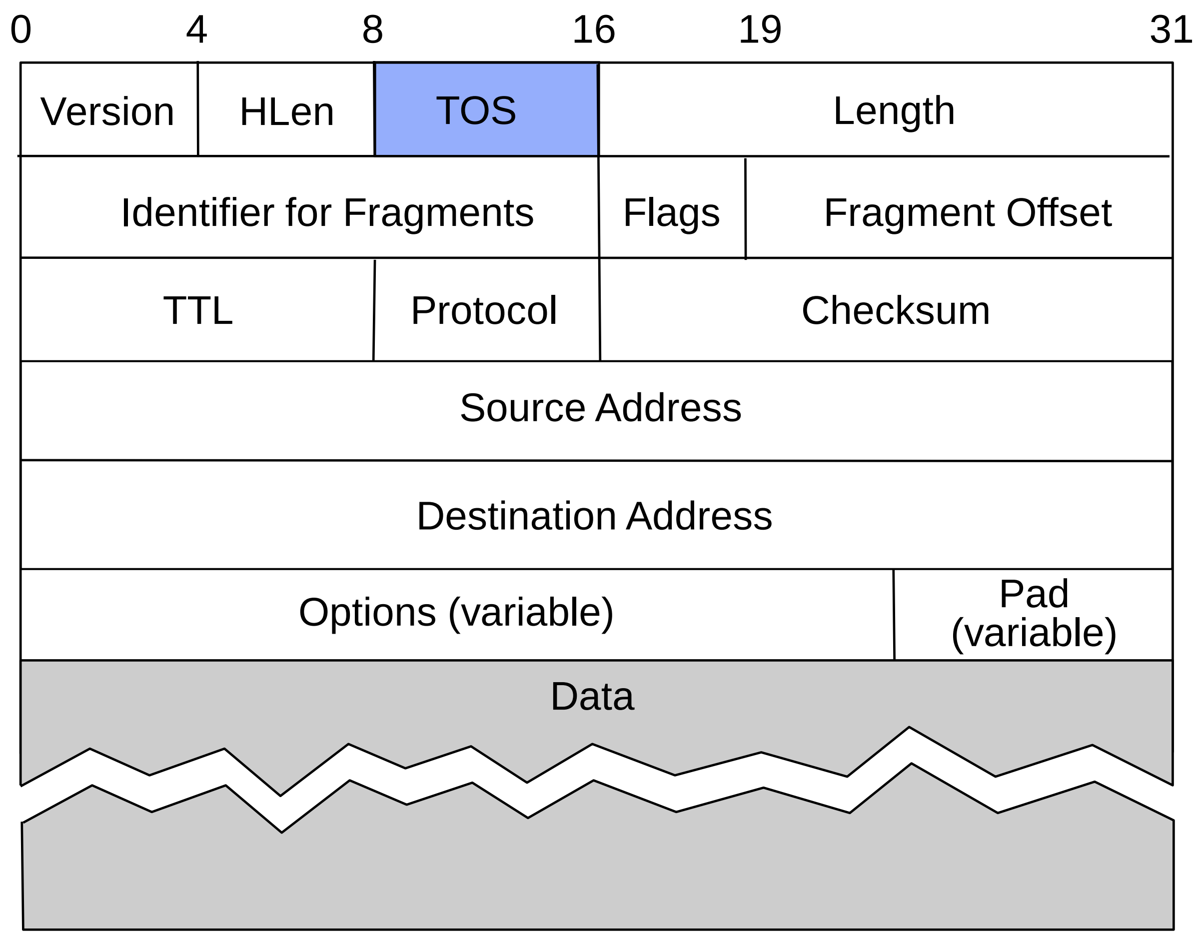
IPv4 Packet Format

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



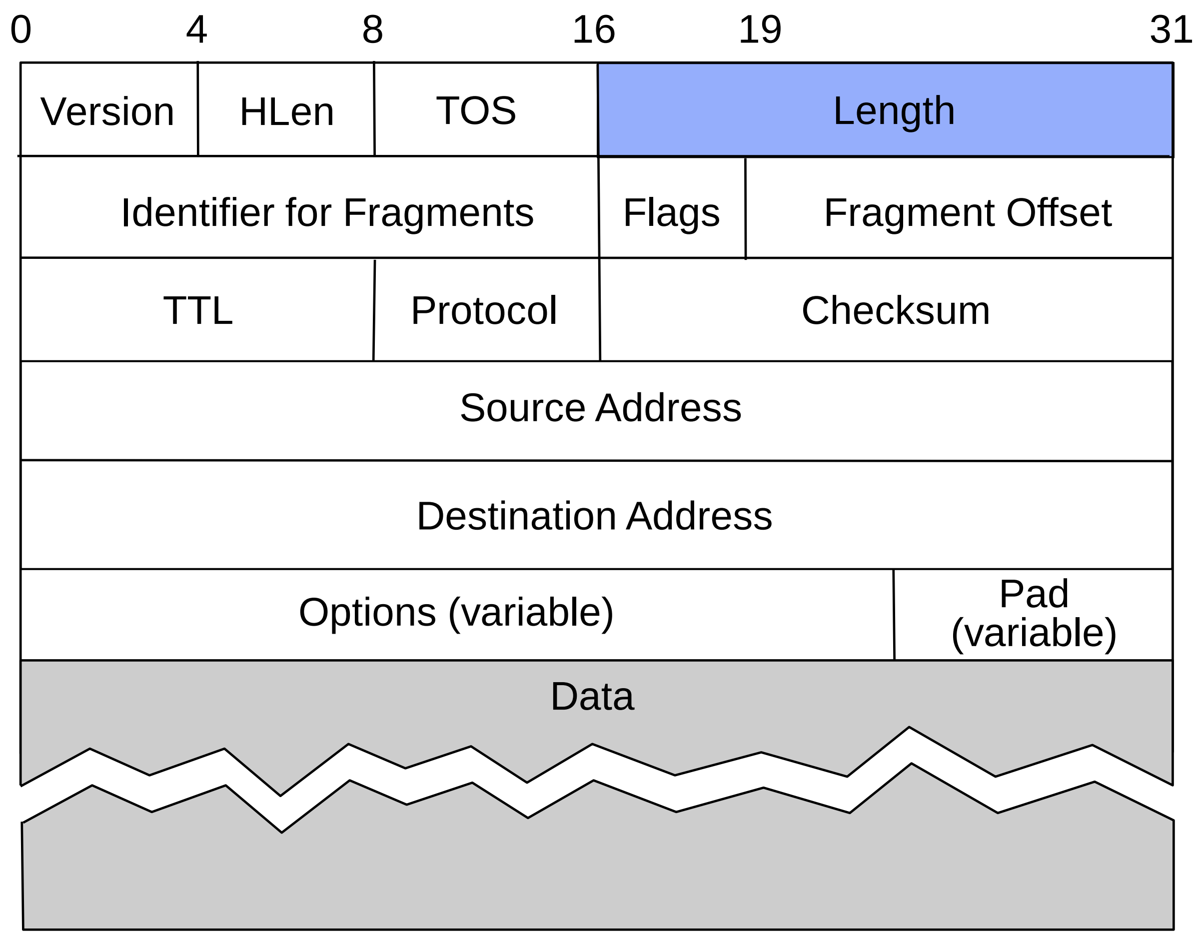
IPv4 Header Fields …

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



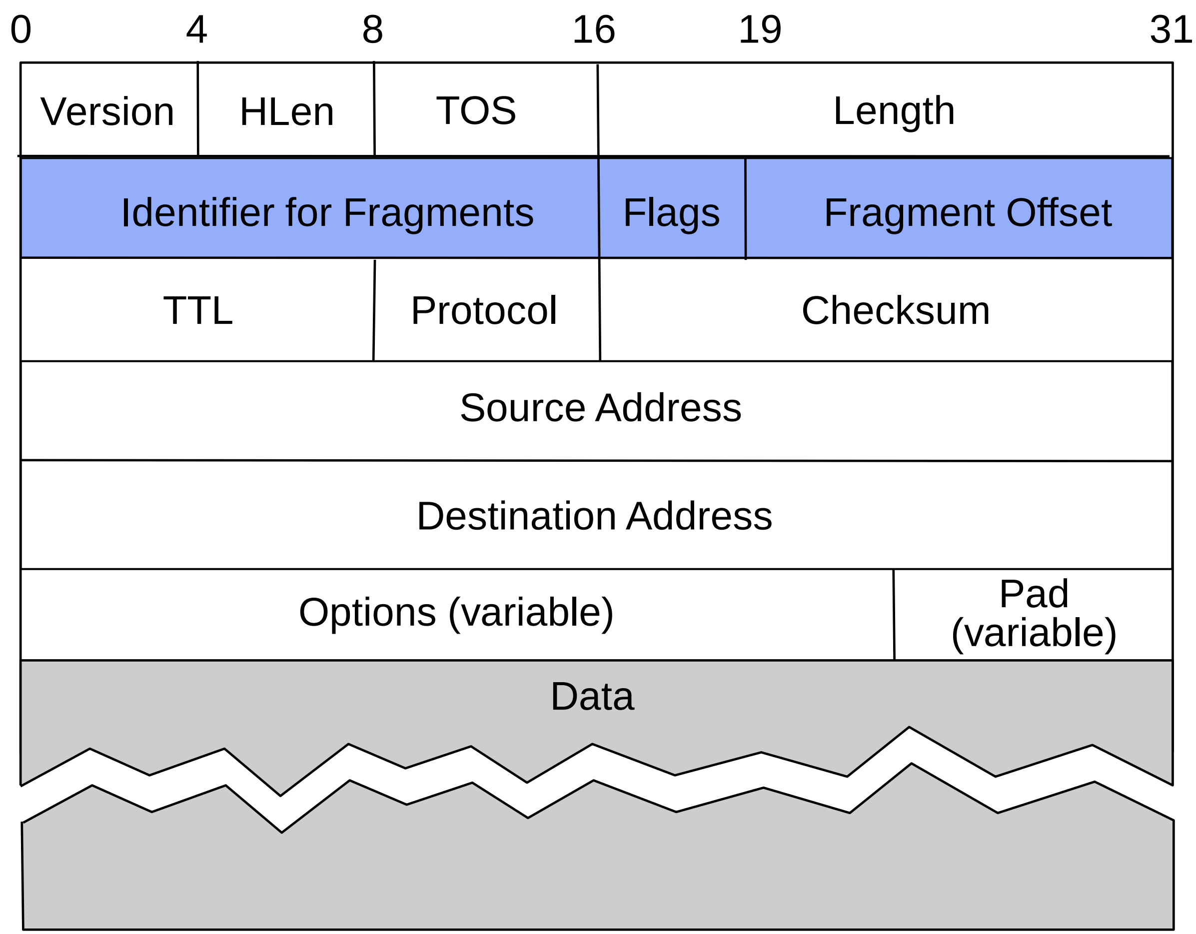
IPv4 Header Fields …

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



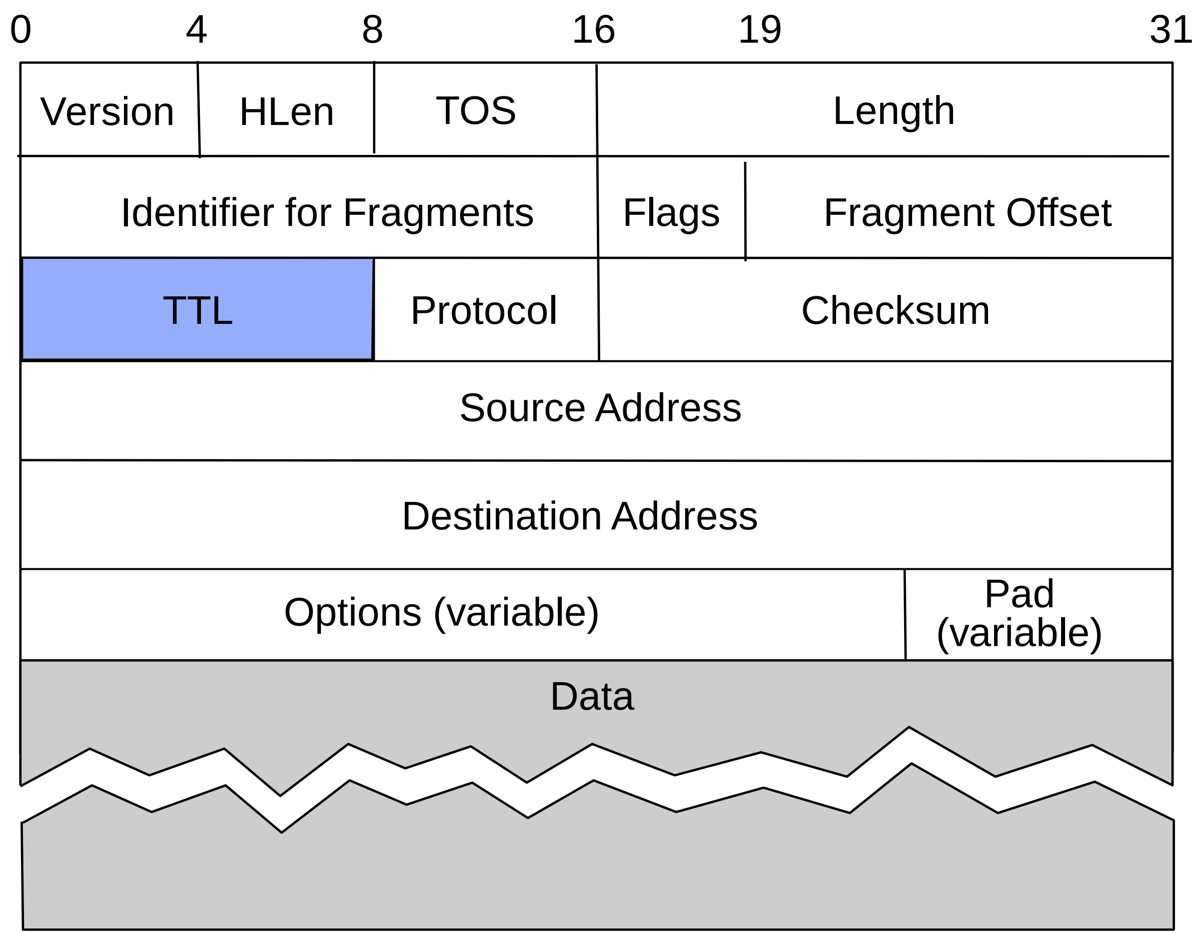
IPv4 Header Fields …

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



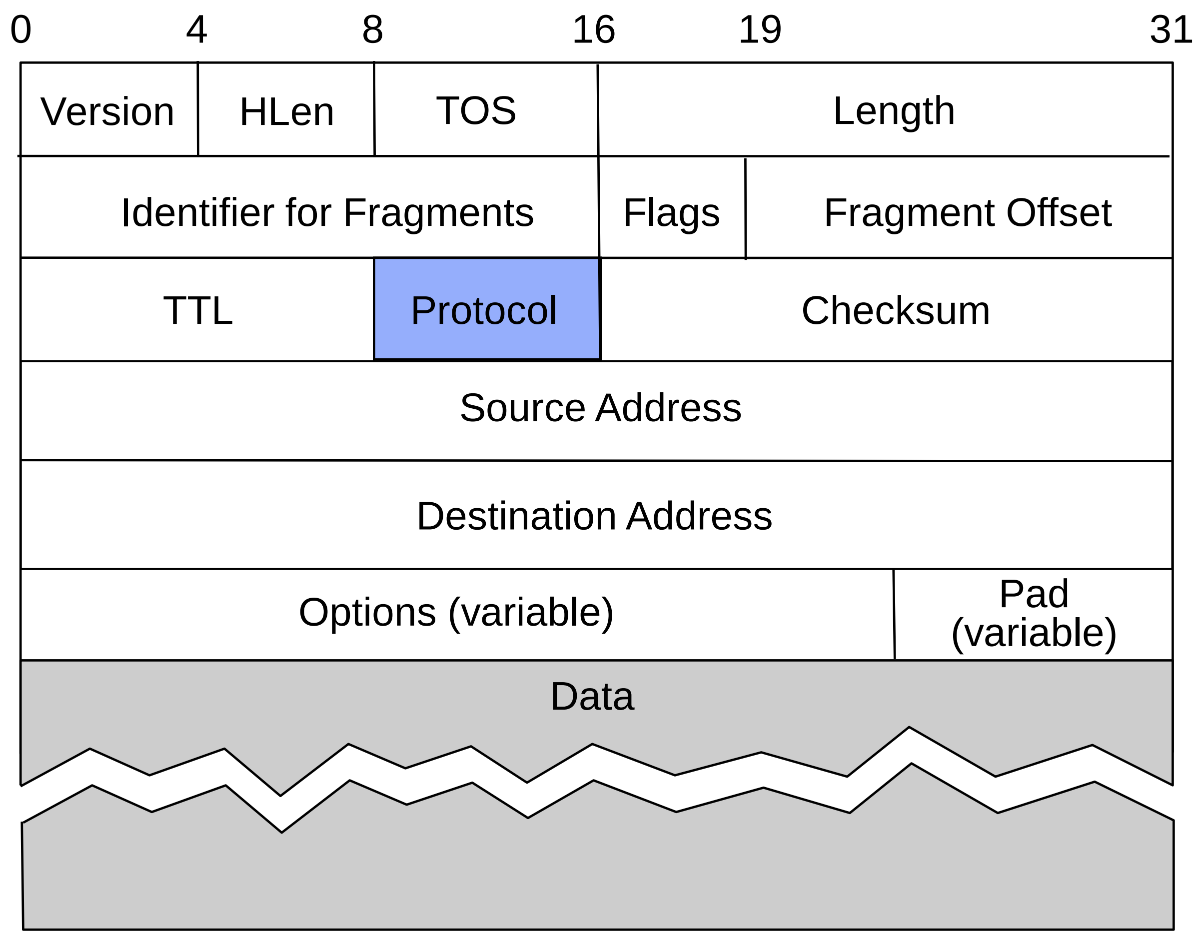
IPv4 Header Fields …

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



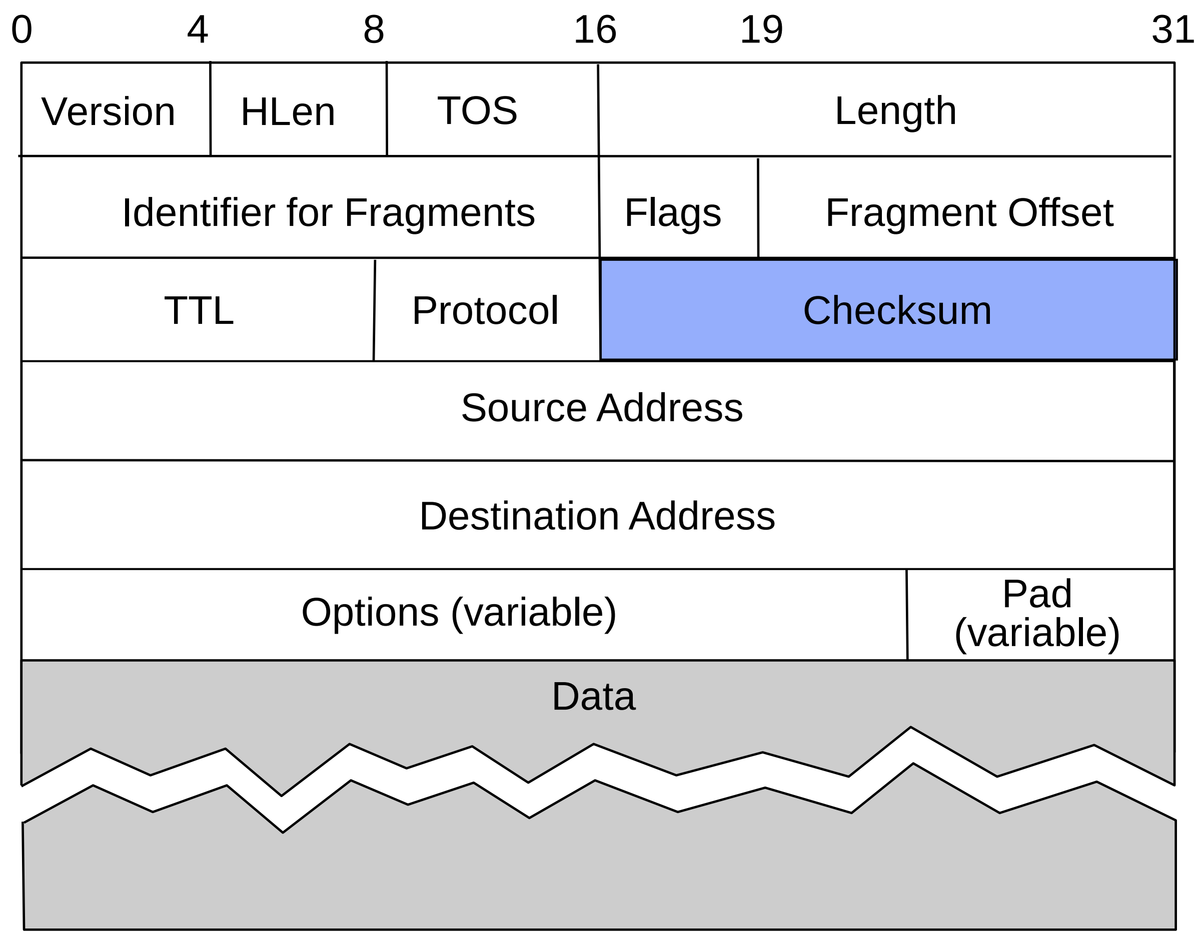
IPv4 Header Fields …

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



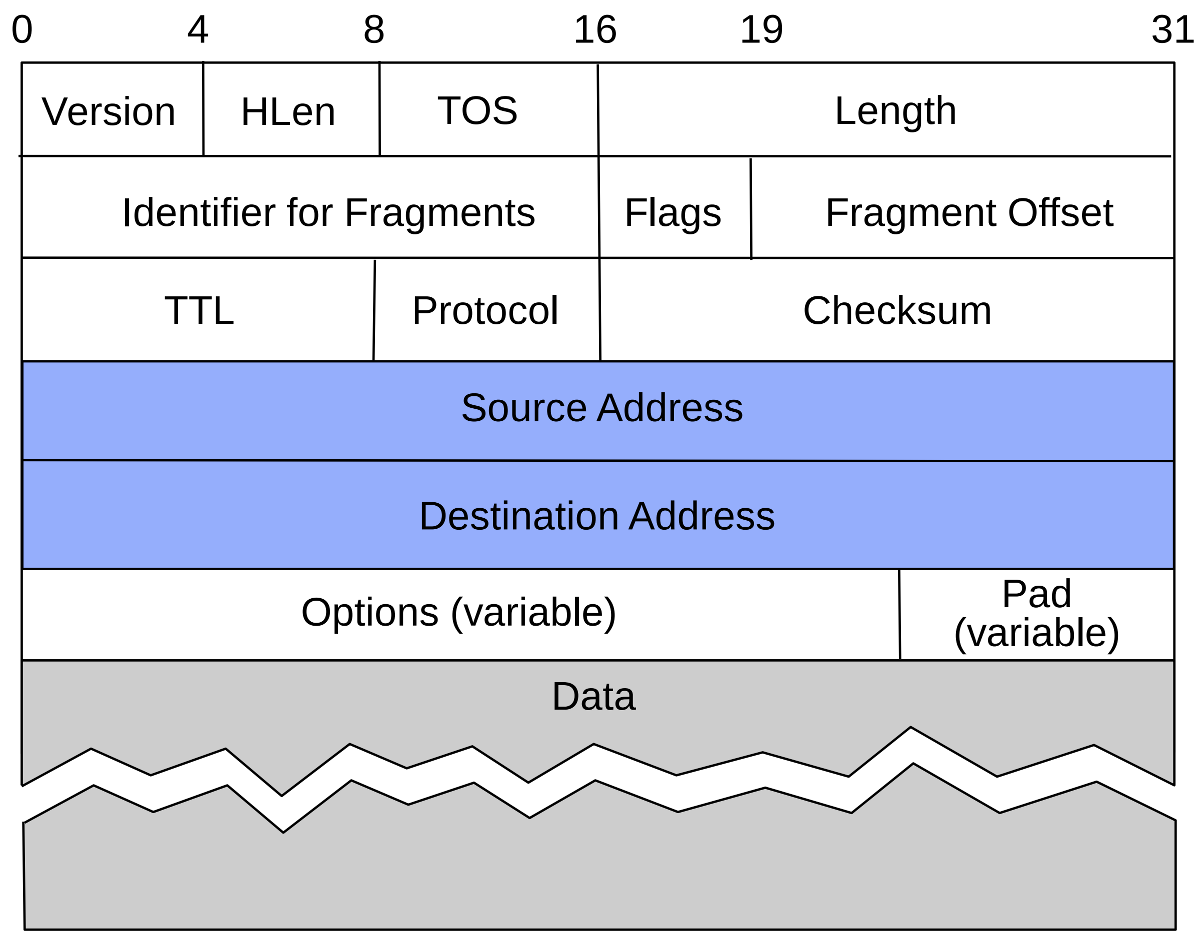
IPv4 Header Fields …

- Header checksum
- Recalculated by routers. Why?
  - (TTL drops)
- Doesn’t cover data
- Disappears for IPv6



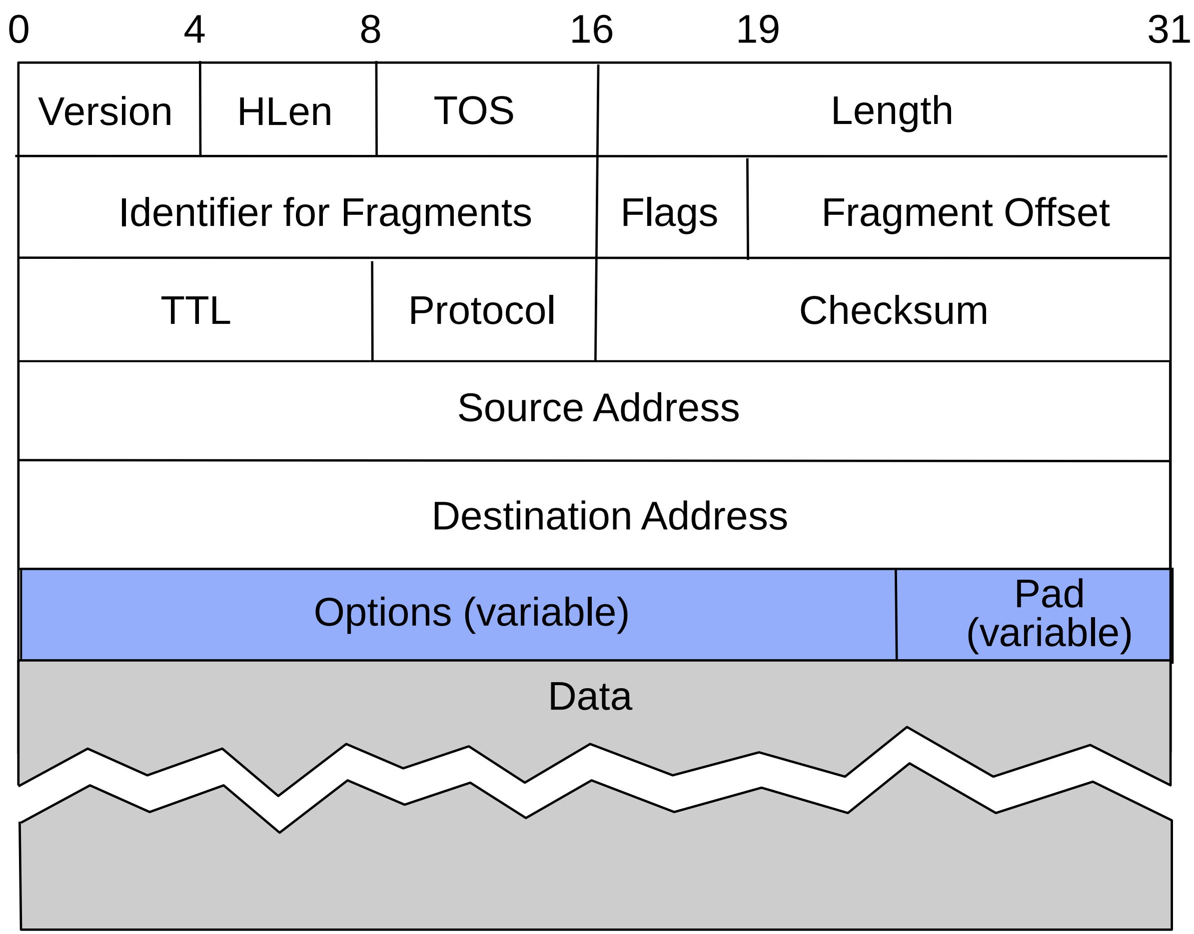
IPv4 Header Fields …

- Source & destination IP addresses
  - Not Ethernet
- Unchanged by routers
- Not authenticated by default



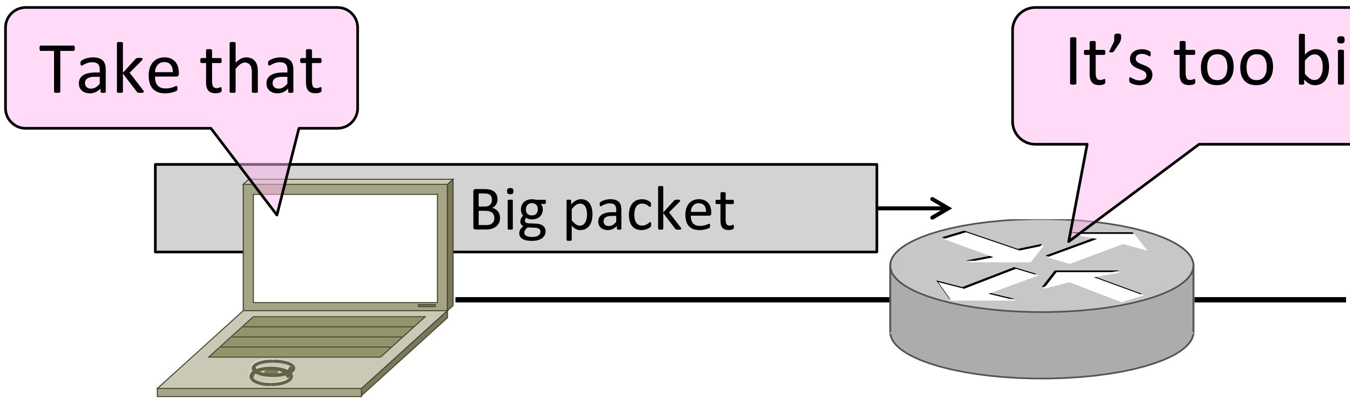
IPv4 Header Fields …

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



Packet Fragmentation

- How do we connect networks with different packet sizes?
  - Need to split up packets, or discover the largest size to use



Fragmentation Issue

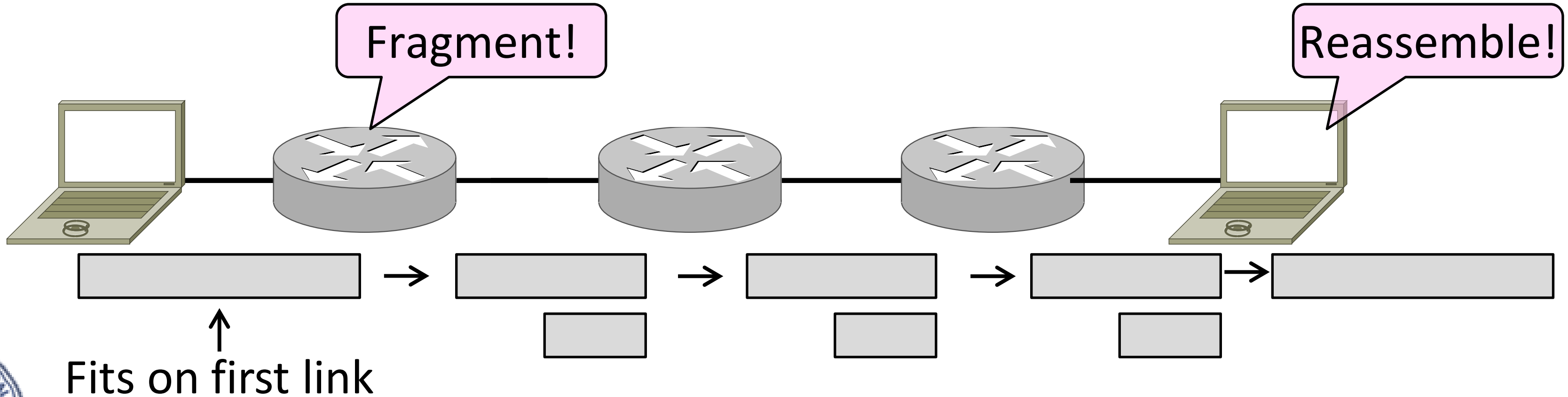
- Different networks may have different maximum packet sizes
  - Or Maximum Transmission Unit (MTU)
  - Ethernet 1.5K, FDDI 4.5K, WiFi 2.3K
- Prefer large packets for efficiency
  - But what size is too large?
  - Difficult because node does not know complete network path
- Don't know if packet will be too big for path beforehand
  - IPv4: fragment on demand and reassemble at dest.
  - IPv6: network returns error so host can learn limit

Packet Size Solutions

- Fragmentation (now)
  - Split up large packets in the network if they are too big to send
  - Classic method, dated
- Discovery (next)
  - Find the largest packet that fits on the network path and use it
  - IP uses today instead of fragmentation

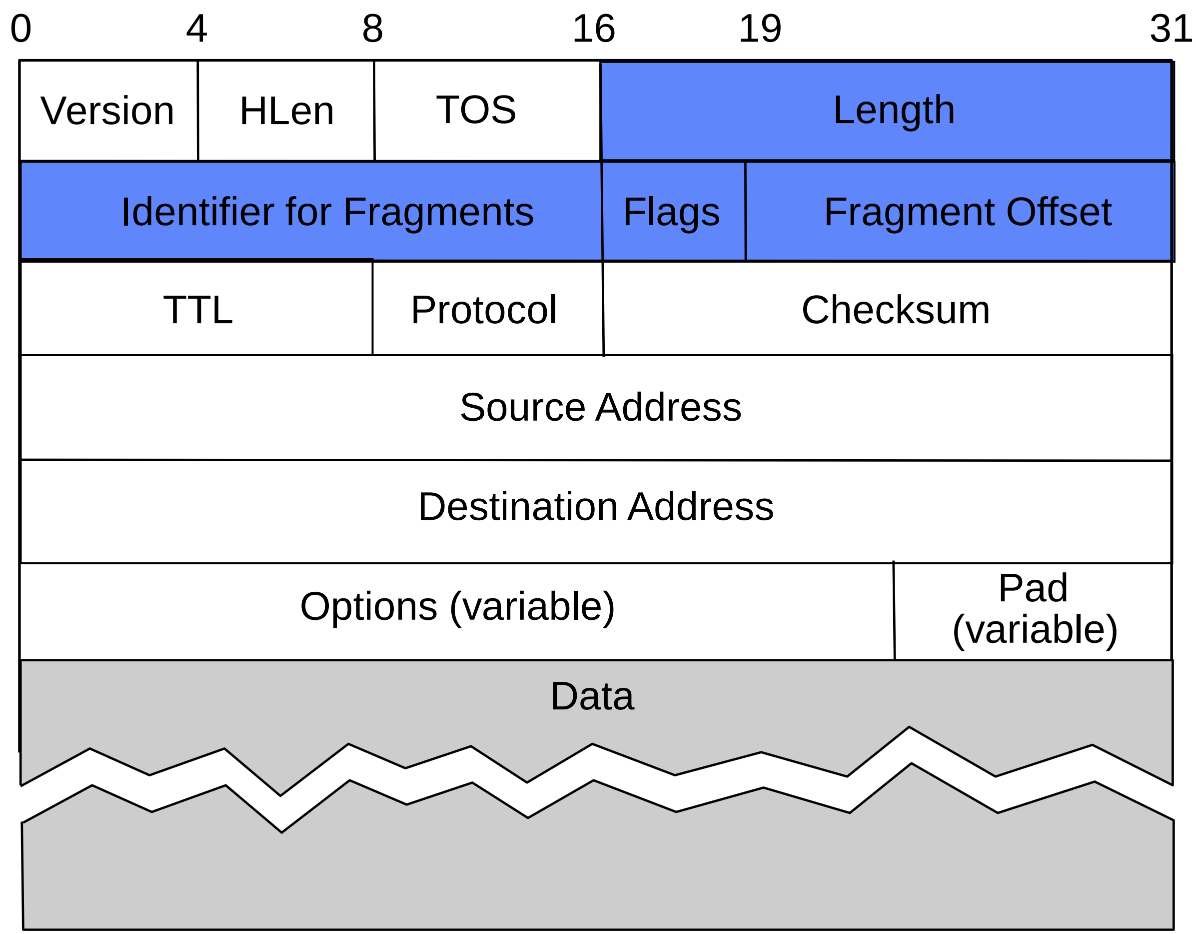
Fragmentation and Reassembly

- Strategy
  - fragment when necessary (MTU < Datagram size)
  - try to avoid fragmentation at source host
  - refragmentation is possible
  - fragments are self-contained IP datagrams
  - delay reassembly until destination host
  - do not recover from lost fragments



Fragment Fields

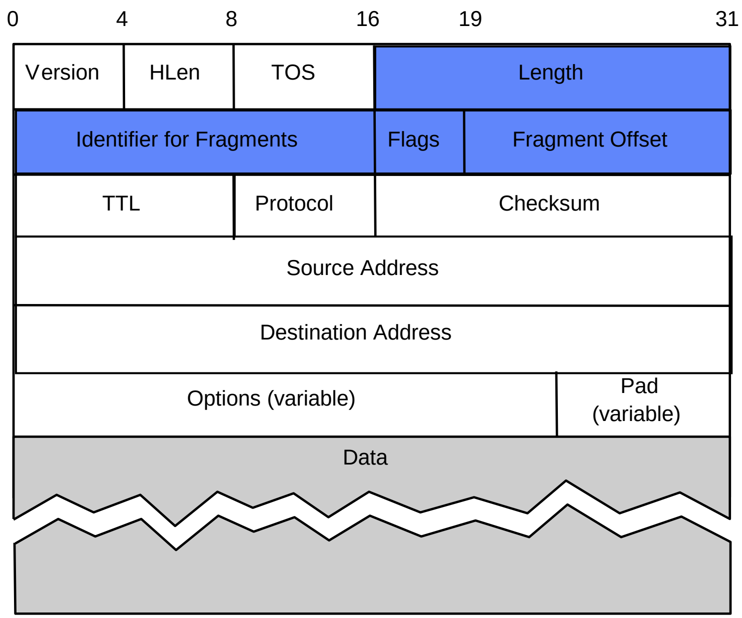
- Fragments of one packet identified by (source, dest, frag id) triple
  - Make unique
- Offset gives start, length changed
- Flags are More Fragments (MF) Don't Fragment (DF)



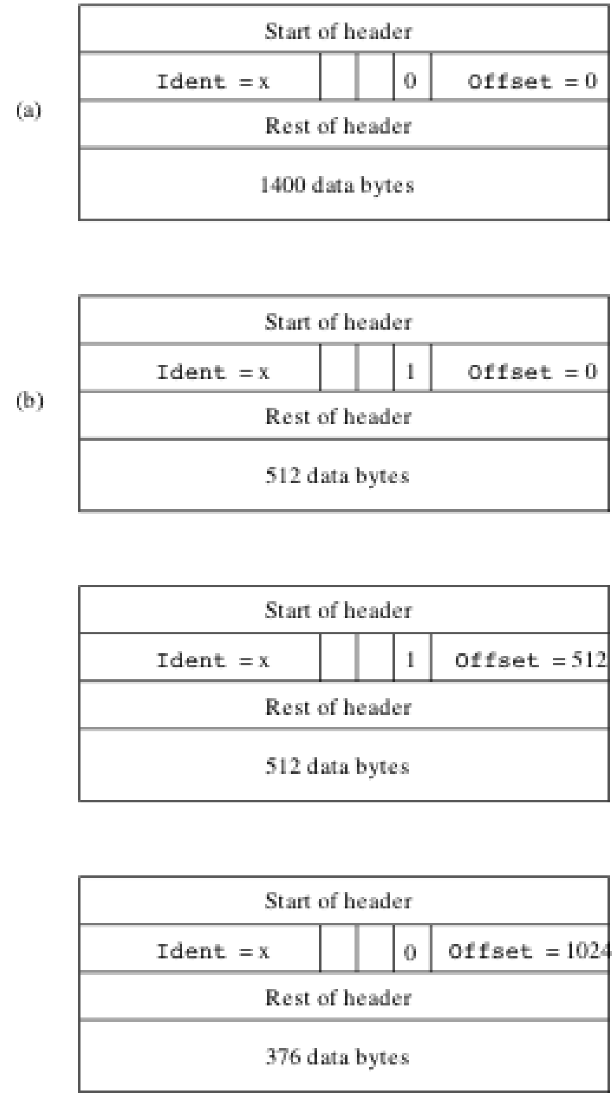
IPv4 Fragmentation Procedure

- Routers split a packet that is too large
  - Typically break into large pieces
  - Copy IP header to pieces
  - Adjust length on pieces
  - Set offset to indicate position
  - Set MF (More Fragments) on all pieces except last
- Receiving host reassembles the pieces:
  - Identification field links pieces together, MF tells receiver when it has all pieces

Fragmenting a Packet



Packet Format



- How do we differentiate a non-fragmented packet (a) from the last fragment in a fragmented packet (b).3?

Fragment Considerations

- Making fragments be datagrams provides:
  - Tolerance of loss, 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

Fragmentation Issues Summary

- Causes inefficient use of resources within the network
  - BW, CPU at both the routers and the hosts
- Higher level protocols must re-xmit entire datagram
  - On lossy network links, hard for packet to survive
  - Tends to magnify loss rate
- Efficient reassembly is hard
  - Lots of special cases
  - (think linked lists)
- Security vulnerabilities too

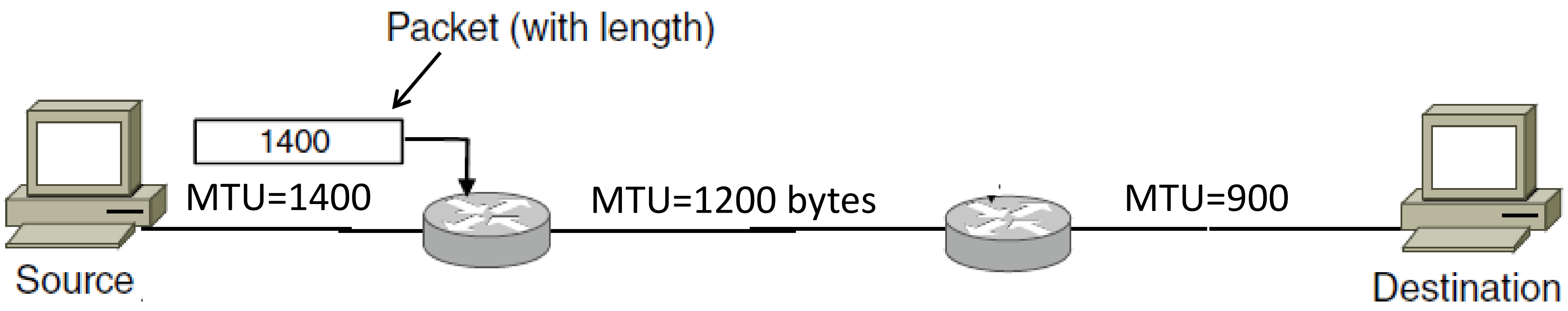
Avoiding Fragmentation

- Always send small datagrams
  - Might be too small
- “Guess” MTU of path
  - Use DF flag. May have large startup time
- Discover actual MTU of path
  - One RT delay w/help, much more w/o.
  - “Help” requires router support
- Guess or discover, but be willing to accept your mistakes

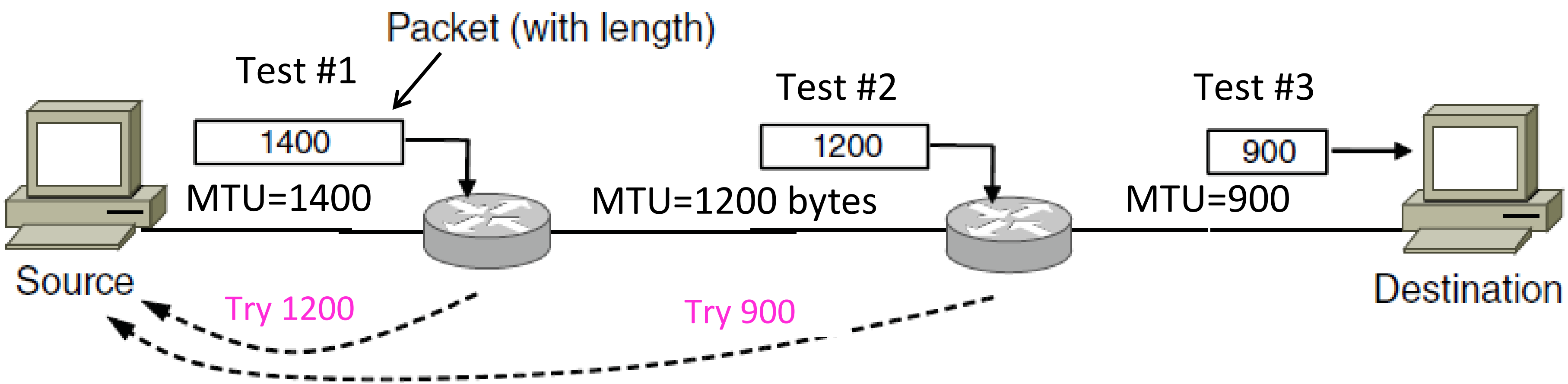
Path MTU Discovery

- Discover the MTU that will fit
  - So we can avoid fragmentation
  - The method in use today
- Host tests path with large packet
  - Routers provide feedback if too large; they tell host what size would have fit

Path MTU Discovery (2)



Path MTU Discovery (3)

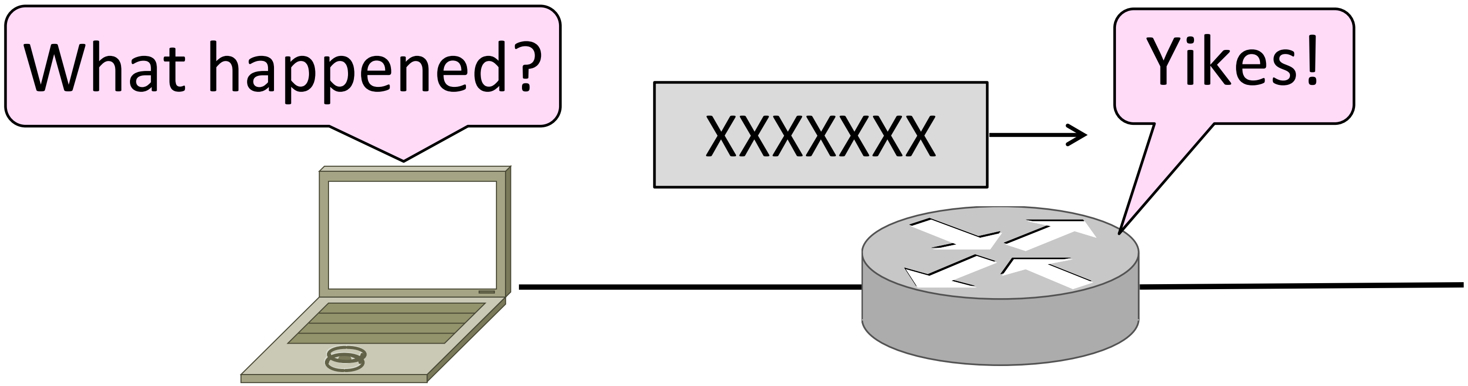


Path MTU Discovery (4)

- Process may seem involved
  - But usually quick to find right size
- Path MTU depends on the path and so can change over time
  - Search is ongoing
- Implemented with ICMP (next)
  - Set DF (Don't Fragment) bit in IP header to get feedback

Error Handling

- What happens when something goes wrong during forwarding?
  - Need to be able to find the problem
  - Need a way to test/debug a large, widely distributed system

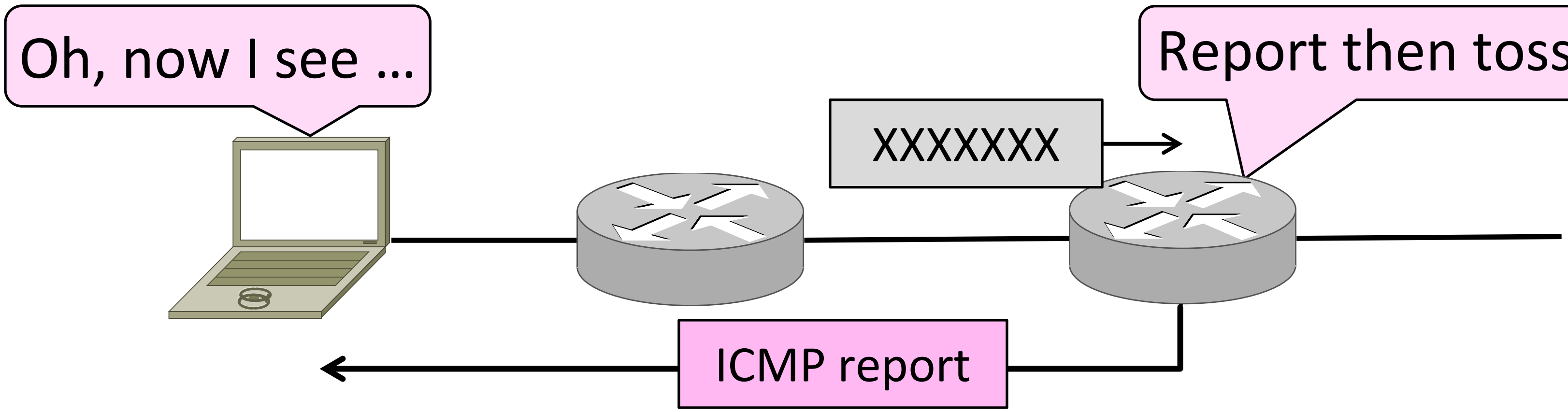


Internet Control Message Protocol (ICMP)

- ICMP = Internet Control Message Protocol (RFC792)
  - Companion to IP – required functionality
  - They are implemented together
  - Sits on top of IP, ICMP messages are carried by IP packets with IP Protocol field equal to 1
- Provides error report and testing:
  - Error is at router during IP forwarding
  - Also testing that hosts can use
    - e.g.: queries about the status of the network

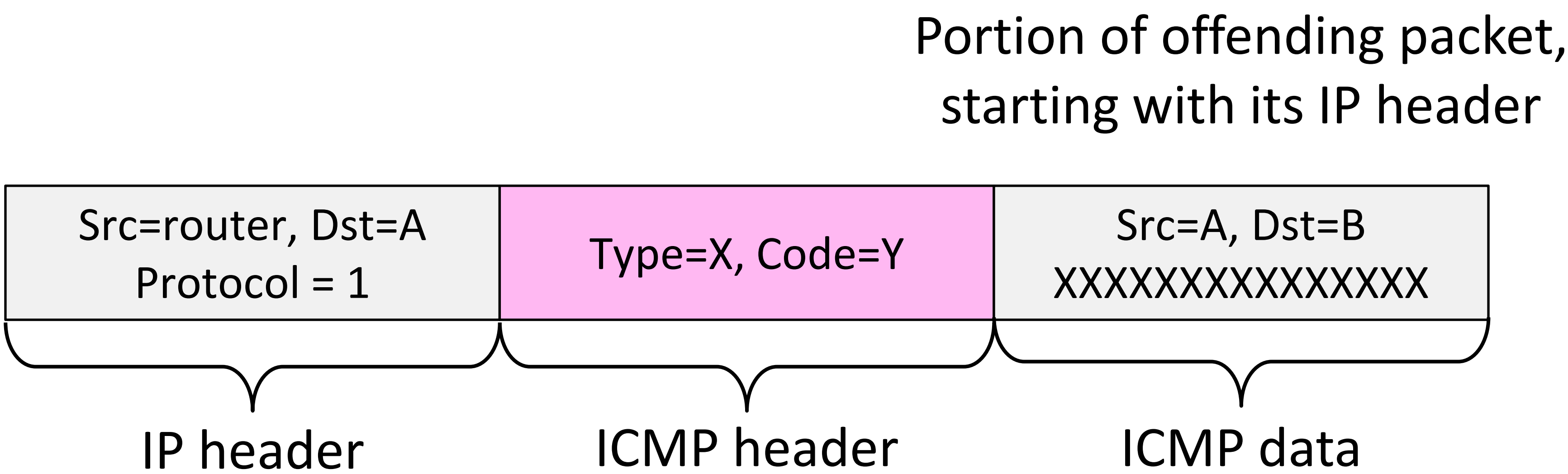
ICMP Errors

- When a router encounters an error while forwarding:
  - It sends an ICMP error report back to the IP source address
  - It discards the problematic packet; host needs to rectify



ICMP Message Format

- Each ICMP message has a Type, Code and Checksum
- Often carry the start of the offending packet as payload
- Each message is carried in an IP packet



Common ICMP Messages

- Destination unreachable
  - “Destination” can be host, network, port or protocol
- Packet needs fragmenting but DF is set
- Redirect
  - To shortcut circuitous routing
- TTL Expired
  - Used by the “traceroute” program
- Echo request/reply
  - Used by the “ping” program
- Cannot Fragment
- Busted Checksum

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

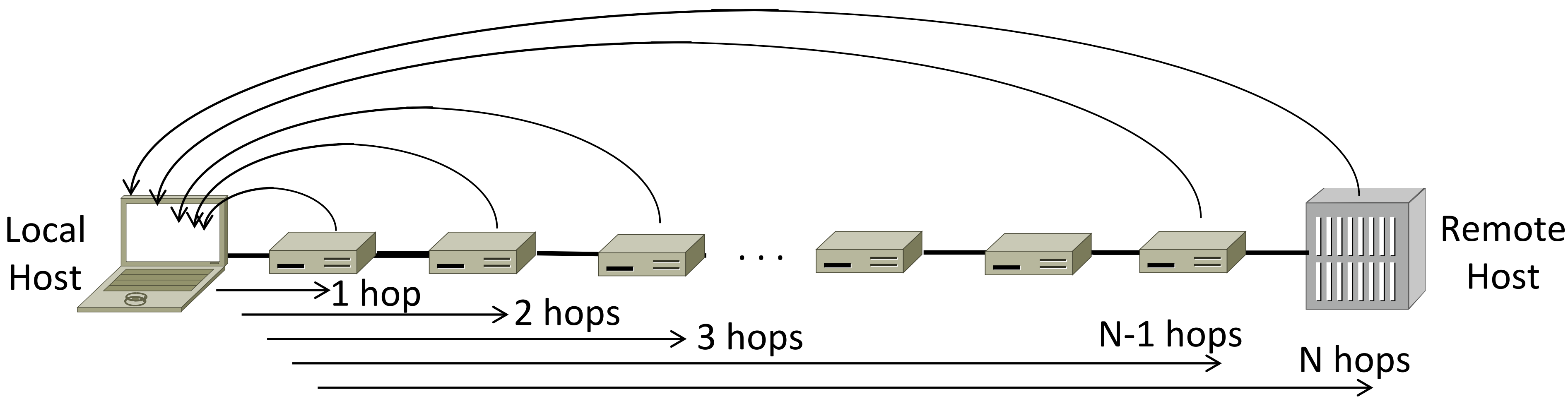
Traceroute

- IP header contains TTL (Time to live) field
  - Decremented every router hop, with ICMP error if it hits zero
  - Protects against forwarding loops

<table>
<tr><td>Version</td><td>IHL</td><td colspan="2">Differentiated Services</td><td colspan="5">Total length</td></tr>
<tr><td colspan="4">Identification</td><td></td><td>DF</td><td>MF</td><td colspan="2">Fragment offset</td></tr>
<tr><td colspan="2">Time to live</td><td colspan="2">Protocol</td><td colspan="5">Header checksum</td></tr>
<tr><td colspan="9">Source address</td></tr>
<tr><td colspan="9">Destination address</td></tr>
<tr><td colspan="9">Options (0 or more words)</td></tr>
</table>

Traceroute (2)

- Traceroute repurposes TTL and ICMP functionality
  - Sends probe packets increasing TTL starting from 1
  - ICMP errors identify routers on the path



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.

Key Concepts

- Network layer provides end-to-end data delivery across an internetwork, not just a LAN
  - Datagram and virtual circuit service models
  - IP/ICMP is the network layer protocol of the Internet
- Up next: More detailed look at routing and addressing