CSE/EE 461
IP/ICMP and the Network Layer

Last Time

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

Application
Presentation
Session
Transport
Network
Data Link
Physical

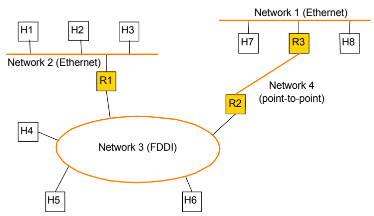
This Lecture

- Focus:
  - How do we build large networks?
- Introduction to the Network layer
  - Internetworks
  - Service models
  - IP, ICMP

Application
Presentation
Session
Transport
Network
Data Link
Physical

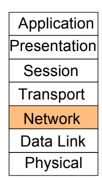
Internetworks

- Set of interconnected networks, e.g., the Internet
  - Scale and heterogeneity



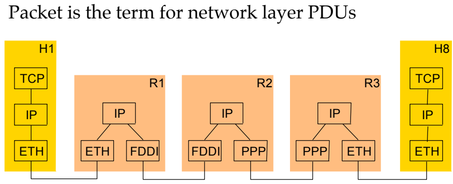
The Network Layer

- 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

Ethernet Header	IP Header	Higher layer headers and Payload

Front of packet to left (and uppermost)

Network Service Models

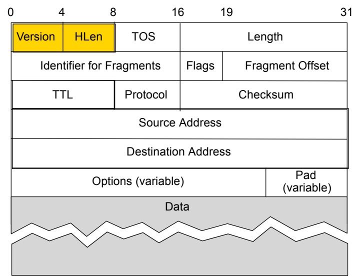
- Datagram delivery: postal service
  - connectionless, best-effort or unreliable service
  - Network can't guarantee delivery of the packet
  - Each packet from a host is routed independently
  - Example: IP
- Virtual circuit models: telephone
  - connection-oriented service
  - Signaling: connection establishment, data transfer, teardown
  - All packets from a host are routed the same way (router state)
  - Example: ATM, Frame Relay, X.25

Internet Protocol (IP)

- IP (RFC791) defines a datagram "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

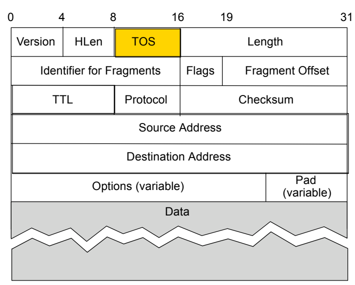
IPv4 Packet Format

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



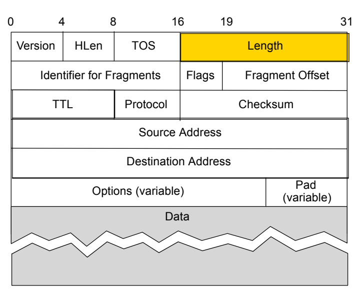
IPv4 Header Fields …

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



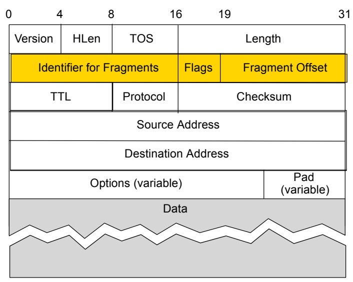
IPv4 Header Fields …

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



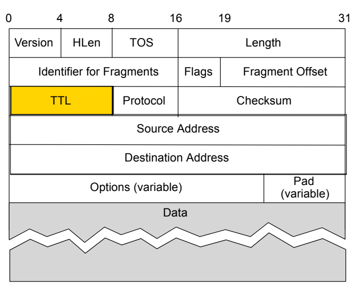
IPv4 Header Fields …

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



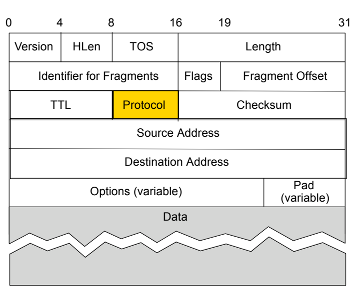
IPv4 Header Fields …

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



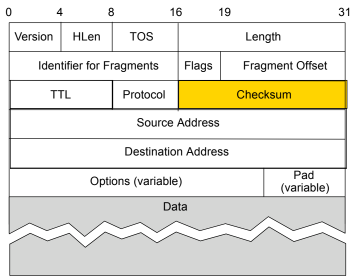
IPv4 Header Fields …

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



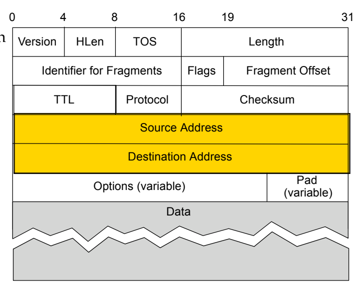
IPv4 Header Fields …

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



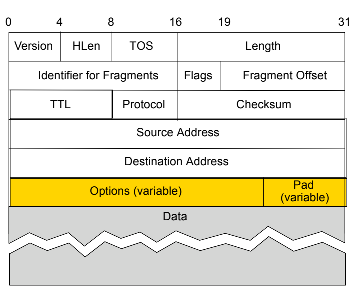
IPv4 Header Fields …

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



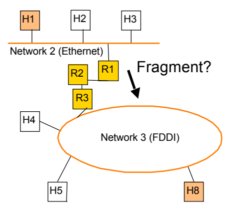
IPv4 Header Fields …

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



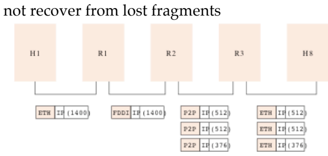
Fragmentation Issue

- 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



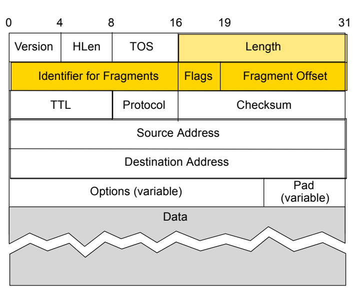
Fragmentation and Reassembly

- 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

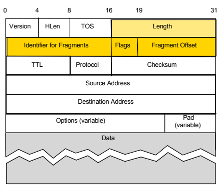


Fragment Fields

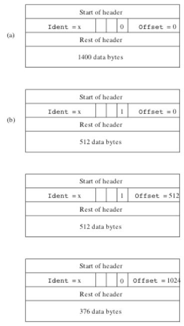
- Fragments of one packet identified by (source, dest, frag id) triple
  - Make unique
- Offset gives start, length changed
- Flags are More Fragments (MF) Don't Fragment (DF)



Fragmenting a Packet



Packet Format



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
- Higher level protocols must re-xmit entire datagram
  - on lossy network links, hard for packet to survive
- Efficient reassembly is hard
  - Lots of special cases
  - (think linked lists)

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

- Path MTU is the smallest MTU along path
  - Packets less than this size don't get fragmented
- Fragmentation is a burden for routers
  - We already avoid reassembling at routers
  - Avoid fragmentation too by having hosts learn path MTUs
- Hosts send packets, routers return error if too large
  - Hosts discover limits, can fragment at source
  - Reassembly at destination as before
- Learned lesson from IPv4, streamlined in IPv6

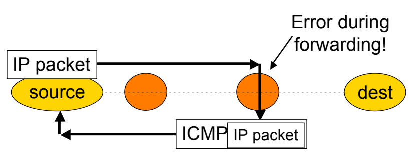
IP Addresses and IP Datagram Forwarding

- IP datagram (packet) contains destination address
- How the source gets the packet to the destination:
  - if source is on same network (LAN) as destination, source sends packet directly to destination host
  - else source sends data to a router on the same network as the source
  - router will forward packet to a router on the next network over
  - and so on…
  - until packet arrives at router on same network as destination; then, router sends packet directly to destination host
- Requirements
  - every host needs to know IP address of the router on its LAN
  - every router needs a routing table to tell it which neighboring network to forward a given packet on

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
- 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

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