Lecture 11: Internetworking

CSE 123: Computer Networks
Alex C. Snoeren

HW 2 due WEDNESDAY
Main challenge is heterogeneity of link layers:

- **Addressing**
  - Each network media has a different addressing scheme
- **Bandwidth**
  - Modems to terabits
- **Latency**
  - Seconds to nanoseconds
- **Frame size**
  - Dozens to thousands of bytes
- **Loss rates**
  - Differ by many orders of magnitude
- **Service guarantees**
  - “Send and pray” vs reserved bandwidth
Cerf & Kahn74, “A Protocol for Packet Network Intercommunication”
- Foundation for the modern Internet

- **Routers** forward **packets** from source to destination
  - May cross many separate networks along the way

- All packets use a common **Internet Protocol**
  - *Any* underlying data link protocol
  - *Any* higher layer transport protocol
IP Networking

CSE 123 – Lecture 11: Internetworking
Routers

- A router is a store-and-forward device
  - Routers are connected to multiple networks
  - On each network, looks just like another host
  - A lot like a switch, but supports multiple datalink layers and makes decisions at the network layer

- Must be explicitly addressed by incoming frames (L2)
  - Not at all like a switch, which is transparent
  - Removes link-layer header, parses IP header (L3)

- Looks up next hop, forwards on appropriate network
  - Each router need only get one step closer to destination
IP Philosophy

- Impose few demands on network
  - Make few assumptions about what network can do
  - No QoS, no reliability, no ordering, no large packets
  - No persistent state about communications; no connections

- Manage heterogeneity at hosts (not in network)
  - Adapt to underlying network heterogeneity
  - Re-order packets, detect errors, retransmit lost messages…
  - Persistent network state only kept in hosts (fate-sharing)

- Service model: **best effort**, a.k.a. *send and pray*
# IP Packet Header

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
<th>Offset</th>
</tr>
</thead>
<tbody>
<tr>
<td>ver</td>
<td>Protocol version</td>
<td>0-3</td>
</tr>
<tr>
<td>HL</td>
<td>Header length</td>
<td>4-7</td>
</tr>
<tr>
<td>TOS</td>
<td>Type of Service</td>
<td>8-11</td>
</tr>
<tr>
<td>length</td>
<td>Length of header</td>
<td>12-15</td>
</tr>
<tr>
<td>identification</td>
<td>Identification number of datagram</td>
<td>16-31</td>
</tr>
<tr>
<td>offset</td>
<td>Offset of first datagram data</td>
<td></td>
</tr>
<tr>
<td>TTL</td>
<td>Time to Live</td>
<td>32-34</td>
</tr>
<tr>
<td>protocol</td>
<td>Protocol type</td>
<td>35-36</td>
</tr>
<tr>
<td>header checksum</td>
<td>Header checksum</td>
<td>37-39</td>
</tr>
<tr>
<td>source address</td>
<td>Source address</td>
<td>40-43</td>
</tr>
<tr>
<td>destination address</td>
<td>Destination address</td>
<td>44-47</td>
</tr>
</tbody>
</table>

- **options (if any)**
- **data (if any)**

---

CSE 123 – Lecture 11: Internetworking
Version field

- Which version of IP is this?
  - Plan for change
  - Very important!

- Current versions
  - 4: most of Internet today
  - 6: new protocol with larger addresses
  - What happened to 5?
    Standards body politics.
Header length

- How big is IP header?
  - Counted in 32-bit words
  - Variable length
    » Options
  - Engineering consequences of variable length...

- Most IP packet headers are 20 bytes long
Type-of-Service

● How should this packet be treated?
  ❖ Care/don’t care for delay, throughput, reliability, cost
  ❖ How to interpret, how to apply on underlying net?
  ❖ Largely unused until 2000 (hijacked for new purposes, ECN & Diffserv)
Length

- How long is whole packet in bytes?
  - Includes header
  - Limits total packet to 64K
  - Redundant?

<table>
<thead>
<tr>
<th>ver</th>
<th>HL</th>
<th>TOS</th>
<th>offset</th>
</tr>
</thead>
<tbody>
<tr>
<td>identification</td>
<td>P</td>
<td>P</td>
<td>A</td>
</tr>
<tr>
<td>TTL</td>
<td>protocol</td>
<td>header checksum</td>
<td></td>
</tr>
<tr>
<td>source address</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>destination address</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>options (if any)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>data (if any)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TTL (Time-to-Live)

- How many more routers can this packet pass through?
  - Designed to limit packet from looping forever
- Each router decrements TTL field
- If TTL is 0 then router discards packet
Protocol

- Which transport protocol is the data using?
  - i.e. how should a host interpret the data

- TCP = 6
- UDP = 17
IP Checksum

- Header contains simple checksum
  - Validates content of header only

- Recalculated at each hop
  - Routers need to update TTL
  - Hence straightforward to modify

- Ensures correct destination receives packet
So what *does* IP do?

- **Addressing**
- **Fragmentation**
  - E.g. FDDI’s maximum packet is 4500 bytes while Ethernet is 1500 bytes, how to manage this?
- **Some** error detection
- **Routers only forward packets to next hop**
  - They do not:
    » Detect packet loss, packet duplication
    » Reassemble or retransmit packets

- Today we’ll talk about fragmentation
Fragmentation

- Different networks may have different maximum frame sizes
  - Maximum Transmission Unit (MTUs)
  - Ethernet 1.5K, FDDI 4.5K
- Router breaks up single IP packet into two or more smaller IP packets
  - Each fragment is labeled so it can be correctly reassembled
  - End host reassembles them into original packet
IP ID and Bitflags

- Source inserts unique value in identification field
  - Also known as the IPID
  - If packet is fragmented, the router copies this value into any fragments
- Offset field indicates position of current fragment (in bytes/8)
  - Zero for non-fragmented packet
- Bitflags provide additional information
  - More Fragments bit helps identify last fragment
  - Don’t Fragment bit prohibits (further) fragmentation
  - Note recursive fragmentation easily supported—just requires care with More Fragments bit
One large datagram becomes several smaller datagrams

(Offset actually encoded as bytes/8)
Costs of Fragmentation

- Interplay between fragmentation and retransmission
  - A single lost fragment may trigger retransmission
  - Any retransmission will be of entire packet (why?)

- Packet must be completely reassembled before it can be consumed on the receiving host
  - Takes up buffer space in the mean time
  - When can it be garbage collected?

- Why not reassemble at each router?
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 can set “don’t fragment” flag
  - Hosts discover limits, can size packets at source
    » ICMP protocol: special IP packet format for sending error msgs
  - Reassembly at destination as before
Aside: 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 Error Message Generation

Error during forwarding!
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
    - traceroute traces packet routes through Internet

- **Echo request/reply**
  - Used by the “ping” program
    - ping just tests for host liveness

- **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
  - Don’t waste valuable bandwidth sending tons of ICMP messages
For Next Time

- Read 3.2.5-6, 9.3.1 in P&D
- Homework 2 due WEDNESDAY
- Project 1 due FRIDAY!