Network Layer
Where we are in the Course

• Moving on up to the Network Layer!
Network Layer

- How to connect different link layer networks
  - Routing as the primary concern
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 of routing table, broadcast

Table for all destinations in the world!

Broadcast new destinations to the whole world!

CSE 461 University of Washington
Shortcomings of Switches (2)

2. Don’t work across more than one link layer technology
   - Hosts on Ethernet + 3G + 802.11 ...
Shortcomings of Switches (3)

3. Don’t give much traffic control
   • Want to plan routes / bandwidth

That was lame.
Network Layer Approach

• Scaling:
  • Hierarchy, in the form of prefixes

• Heterogeneity:
  • IP for internetworking

• Bandwidth Control:
  • Lowest-cost routing
  • Later QOS (Quality of Service)
Topics

• Network service models
  • Datagrams (packets), virtual circuits

• IP (Internet Protocol)
  • Internetworking
  • Forwarding (Longest Matching Prefix)
  • Helpers: ARP and DHCP
  • Fragmentation and MTU discovery
  • Errors: ICMP (traceroute!)
  • IPv6, scaling IP to the world
  • NAT, and “middleboxes”

• Routing Algorithms
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
  - Node process (local) and fast
Topic

• What kind of service does the Network layer provide to the Transport layer?
  • How is it implemented at routers?
Two Network Service Models

• Datagrams, or connectionless service
  • Like postal letters
  • (IP as an example)

• Virtual circuits, or connection-oriented service
  • Like a telephone call
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’s table (later)    C’s Table          E’s Table
  A
  B     B
  C     C
  D     B
  E     B
  F     B
  Dest. Line
```

```
A
  B     B
  C     C
  D     B
  E     B
  F     B
  Dest. Line
```

```
A
  A
  B     A
  C
  D     E
  E     E
  F     E
```

```
A
  A
  B     D
  C
  D     D
  E
  F     F
```
IP (Internet Protocol)

- Network layer of the Internet, uses datagrams (next)
- 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 short label to identify the circuit
- Labels don’t have global meaning, only unique for a link
Virtual Circuits (4)

• Each router has a forwarding table keyed by circuit
  • Gives output line and next label to place on packet
MPLS (Multi-Protocol Label Switching, §5.6.5)

• 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

<table>
<thead>
<tr>
<th>Issue</th>
<th>Datagrams</th>
<th>Virtual Circuits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setup phase</td>
<td>Not needed</td>
<td>Required</td>
</tr>
<tr>
<td>Router state</td>
<td>Per destination</td>
<td>Per connection</td>
</tr>
<tr>
<td>Addresses</td>
<td>Packet carries full address</td>
<td>Packet carries short label</td>
</tr>
<tr>
<td>Routing</td>
<td>Per packet</td>
<td>Per circuit</td>
</tr>
<tr>
<td>Failures</td>
<td>Easier to mask</td>
<td>Difficult to mask</td>
</tr>
<tr>
<td>Quality of service</td>
<td>Difficult to add</td>
<td>Easier to add</td>
</tr>
</tbody>
</table>
Internetworking (IP)
Topic

• 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 network technology
Internet Reference Model

• Internet Protocol (IP) is the “narrow waist”
  • Supports many different links below and apps above

1. Link
   - Ethernet
   - Cable
   - DSL
   - 802.11

2. Internet
   - IP

3. Transport
   - TCP
   - UDP

4. Application
   - SMTP
   - HTTP
   - RTP
   - DNS
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
IPv4 (Internet Protocol)

- Various fields to meet straightforward needs
  - Version, Header (IHL), Total length, Protocol, and Header Checksum

<table>
<thead>
<tr>
<th>Field</th>
<th>32 Bits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version</td>
<td></td>
</tr>
<tr>
<td>IHL</td>
<td></td>
</tr>
<tr>
<td>Differentiated Services</td>
<td></td>
</tr>
<tr>
<td>Total length</td>
<td></td>
</tr>
<tr>
<td>Identification</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>M</td>
</tr>
<tr>
<td>Time to live</td>
<td>Protocol</td>
</tr>
<tr>
<td>Source address</td>
<td></td>
</tr>
<tr>
<td>Destination address</td>
<td></td>
</tr>
<tr>
<td>Options (0 or more words)</td>
<td></td>
</tr>
<tr>
<td>Payload (e.g., TCP segment)</td>
<td></td>
</tr>
</tbody>
</table>
### IPv4 (2)

- Some fields to handle packet size differences (later)
  - Identification, Fragment offset, Fragment control bits

![IPv4 Header Diagram](image)

<table>
<thead>
<tr>
<th>Field</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version</td>
<td>4 bits</td>
</tr>
<tr>
<td>IHL</td>
<td>4 bits</td>
</tr>
<tr>
<td>Differentiated Services</td>
<td>8 bits</td>
</tr>
<tr>
<td>Total length</td>
<td>16 bits</td>
</tr>
<tr>
<td>Identification</td>
<td>13 bits</td>
</tr>
<tr>
<td>Fragment offset</td>
<td>13 bits</td>
</tr>
<tr>
<td>Time to live</td>
<td>8 bits</td>
</tr>
<tr>
<td>Protocol</td>
<td>8 bits</td>
</tr>
<tr>
<td>Header checksum</td>
<td>16 bits</td>
</tr>
<tr>
<td>Source address</td>
<td>32 bits</td>
</tr>
<tr>
<td>Destination address</td>
<td>32 bits</td>
</tr>
<tr>
<td>Options (0 or more words)</td>
<td>32 bits</td>
</tr>
<tr>
<td>Payload (e.g., TCP segment)</td>
<td>64 bits</td>
</tr>
</tbody>
</table>
IPv4 (3)

- Other fields to meet other needs (later, later)
  - Differentiated Services, Time to live (TTL)

Later, with QOS

Later, with ICMP

Payload (e.g., TCP segment)
IPv4 (4)

- Network layer of the Internet, uses datagrams
  - Provides a layer of addressing above link addresses (next)

<table>
<thead>
<tr>
<th>Version</th>
<th>IHL</th>
<th>Differentiated Services</th>
<th>Total length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>D</td>
<td>M</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Identification</td>
<td></td>
</tr>
<tr>
<td>Time to live</td>
<td>Protocol</td>
<td>Header checksum</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source address</th>
</tr>
</thead>
<tbody>
<tr>
<td>Destination address</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Options (0 or more words)</th>
</tr>
</thead>
</table>

Payload (e.g., TCP segment)
IP Addresses

- IPv4 uses 32-bit addresses
  - Later we’ll see IPv6, which uses 128-bit addresses
- Written in “dotted quad” notation
  - Four 8-bit numbers separated by dots

\[
\begin{align*}
\text{8 bits} & \quad \text{8 bits} & \quad \text{8 bits} & \quad \text{8 bits} \\
aaaaaaaabbabbcccccdddddddd & \leftrightarrow & A.B.C.D \\
0001001000111100000000000000001 & \leftrightarrow &
\end{align*}
\]
IP Prefixes

• Addresses are allocated in blocks called prefixes
  • Addresses in an L-bit prefix have the same top L bits
  • There are $2^{32-L}$ addresses aligned on $2^{32-L}$ boundary
IP Prefixes (2)

• Written in “IP address/length” notation
  • Address is lowest address in the prefix, length is prefix bits
  • E.g., 128.13.0.0/16 is 128.13.0.0 to 128.13.255.255
  • So a /24 (“slash 24”) is 256 addresses, and a /32 is one address

00010010|00011111|00000000|xxxxxxxxx ↔ 128.13.0.0/16
Classful IP Addressing

- Originally, IP addresses came in fixed size blocks with the class/size encoded in the high-order bits.
- They still do, but the classes are now ignored.

<table>
<thead>
<tr>
<th>Class</th>
<th>Network Address Range</th>
<th>Number of Addresses</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0-15</td>
<td>$2^{24}$</td>
</tr>
<tr>
<td>B</td>
<td>16-63</td>
<td>$2^{16}$</td>
</tr>
<tr>
<td>C</td>
<td>64-127</td>
<td>$2^{8}$</td>
</tr>
</tbody>
</table>

Network portion | Host portion
Classful IP Addressing

- This is an ARPANet assignment.
IP Forwarding

- Addresses on one network belong to a unique prefix
- Node uses a table that lists the next hop for prefixes

<table>
<thead>
<tr>
<th>Prefix</th>
<th>Next Hop</th>
</tr>
</thead>
<tbody>
<tr>
<td>192.24.0.0/19</td>
<td>D</td>
</tr>
<tr>
<td>192.24.12.0/22</td>
<td>B</td>
</tr>
</tbody>
</table>
Longest Matching Prefix

- Prefixes in the table might overlap!
  - Combines hierarchy with flexibility

- **Longest matching prefix** forwarding rule:
  - For each packet, find the longest prefix that contains the destination address, i.e., the most specific entry
  - Forward the packet to the next hop router for that prefix
Longest Matching Prefix (2)

<table>
<thead>
<tr>
<th>Prefix</th>
<th>Next Hop</th>
</tr>
</thead>
<tbody>
<tr>
<td>192.24.0.0/19</td>
<td>D</td>
</tr>
<tr>
<td>192.24.12.0/22</td>
<td>B</td>
</tr>
</tbody>
</table>

192.24.6.0  \rightarrow
192.24.14.32 \rightarrow
192.24.54.0  \rightarrow

192.24.63.255 More specific
192.24.15.255 /22
192.24.12.0 /19
192.24.0.0 IP address
IP Address Work Slide:

• Route to D = 192.00011x.x.x
• Route to B = 192.00011000.000011x.x
• 192.24.6.0 = 192.00011000.00001100.00000000
• 192.24.14.32 = 192.00011000.00001110.00010000
• 192.24.54.0 = 192.00011000.00110110.00000000
Longest Matching Prefix (2)

<table>
<thead>
<tr>
<th>Prefix</th>
<th>Next Hop</th>
<th>IP address</th>
<th>More specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>192.24.0.0/19</td>
<td>D</td>
<td>192.24.0.0</td>
<td>/19</td>
</tr>
<tr>
<td>192.24.12.0/22</td>
<td>B</td>
<td>192.24.12.0</td>
<td>/22</td>
</tr>
</tbody>
</table>

- 192.24.6.0 → D
- 192.24.14.32 → B
- 192.24.54.0 → D
Host/Router Distinction

• In the Internet:
  • Routers do the routing, know way to all destinations
  • Hosts send remote traffic (out of prefix) to nearest router
Host Forwarding Table

• Give using longest matching prefix
  • 0.0.0.0/0 is a default route that catches all IP addresses

<table>
<thead>
<tr>
<th>Prefix</th>
<th>Next Hop</th>
</tr>
</thead>
<tbody>
<tr>
<td>My network prefix</td>
<td>Send to that IP</td>
</tr>
<tr>
<td>0.0.0.0/0</td>
<td>Send to my router</td>
</tr>
</tbody>
</table>
Flexibility of Longest Matching Prefix

• Can provide default behavior, with less specifics
  • Send traffic going outside an organization to a border router (gateway)

• Can special case behavior, with more specifics
  • For performance, economics, security, ...
Performance of Longest Matching Prefix

• Uses hierarchy for a compact table
  • Relies on use of large prefixes

• Lookup more complex than table
  • Used to be a concern for fast routers
  • Not an issue in practice these days
Issues?

• Where does this break down?
Issues?

• Where does this break down?

Bootstrapping (DHCP)
Finding Link nodes (ARP)
Really big packets (Fragmentation)
Errors in the network (ICMP)
Running out of addresses (IPv6, NAT)