IPv4

Christian Grothoff
christian@grothoff.org
http://grothoff.org/christian/

“As the builders say, the larger stones do not lie well without the lesser.” – Plato
The Network Layer

- Transports datagrams from sending to receiving host
- Network layer protocols are implemented on *every* host and router

The network layer is commonly referred to as **layer 3**.
The Internet Network Layer

application

transport layer: TCP, UDP

network

Routing protocols
- path selection
- RIP, OSPF, BGP

IP protocol
- addressing conventions
- datagram format
- packet handling conventions

forwarding table

ICMP protocol
- error reporting
- router "signaling"

data link layer

physical layer
Routing and Forwarding

**routing:** determine route taken by datagrams from source to destination

**forwarding:** move datagram from input device to output device
Datagram networks

- No call setup at network layer
- Routers keep no state about end-to-end connections
- Packets between the same source-destination pair may take different paths
IPv4 Address Format

- 32 bits
- 4 billion\(^1\) possible values
- Notation is dotted decimal, big-endian: 132.149.42.193
- Many values have special meanings

\(^1\)1 billion = 1,000,000,000
IP Addressing

- IP addresses identify interfaces
- Routers have multiple interfaces
- Hosts typically have only one network interface
- One interface can have multiple IPs
Devices in a **subnet** share the same subnet part (higher-order bits) of the IP address and can physically reach each other without the help of a router.
Subnets (2/3)

To determine the subnets, detach each interface from its host or router, creating islands of isolated networks. Those are the subnets.

Question: which devices count as routers?
Subnets (3/3)
Classless InterDomain Routing (CIDR)

- Subnet part of IP address are the higher-order bits
- Address format a.b.c.d/x specifies that subnet part has x bits
How are subnet identifiers assigned?

Administrators get a portion of their providers ISP’s address space:

- ISP has 200.23.16.0/20
- ISP gives Org 1 200.23.16.0/23: 11001000 00010111 00010000
- ISP gives Org 2 200.23.18.0/23: 11001000 00010111 00010010
- ISP gives Org 3 200.23.20.0/23: 11001000 00010111 00010100
- ...  
- ISP gives Org 7 200.23.30.0/23: 11001000 00010111 00011110
### Longest Prefix Matching

<table>
<thead>
<tr>
<th>Prefix match</th>
<th>Interface</th>
</tr>
</thead>
<tbody>
<tr>
<td>11001000 00010111 00010</td>
<td>0</td>
</tr>
<tr>
<td>11001000 00010111 00011000</td>
<td>1</td>
</tr>
<tr>
<td>11001000 00010111 00011</td>
<td>2</td>
</tr>
<tr>
<td>otherwise</td>
<td>3</td>
</tr>
</tbody>
</table>

Example: `11001000 00010111 00011000 10101010`?
Route aggregation

Organization 0
200.23.16.0/23

Organization 1
200.23.18.0/23

Organization 2
200.23.20.0/23

Organization 7
200.23.30.0/23

Fly-By-Night-ISP

"Send me anything with addresses beginning 200.23.16.0/20"

Internet

ISPs-R-Us

"Send me anything with addresses beginning 199.31.0.0/16"
Route aggregation

Organization 0
200.23.16.0/23

Organization 2
200.23.20.0/23

...  

Organization 7
200.23.30.0/23

Organization 1
200.23.18.0/23

Fly-By-Night-ISP

"Send me anything with addresses beginning 200.23.16.0/20"

Internet

ISP's-R-Us

"Send me anything with addresses beginning 199.31.0.0/16 or 200.23.18.0/23"
# The IP Header

<table>
<thead>
<tr>
<th>Offset</th>
<th>0-3</th>
<th>4-7</th>
<th>8-15</th>
<th>16-18</th>
<th>19-31</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Version</td>
<td>Hdr Len</td>
<td>TOS</td>
<td>Total Length</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>Identification</td>
<td></td>
<td>Flags</td>
<td>Frag. Offset</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>Time to Live</td>
<td>Protocol</td>
<td>Hdr. Checksum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>96</td>
<td>Source Address</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>128</td>
<td>Destination Address</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>160</td>
<td>Options (optional)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>160+</td>
<td>Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Where is the IP Header?

- Application layer
- Transport layer
- Network layer
- Data link layer
IP Fragmentation & Reassembly

- Network links have a Maximum Transfer Unit (MTU)
- IP datagrams that exceed the MTU are divided on the network
- Reassembly only happens at the final destination
- IP header bits are used to identify, order related fragments
Example

One large datagram becomes several smaller datagrams

- length = 4000
- ID = x
- fragflag = 0
- offset = 0

- length = 1500
- ID = x
- fragflag = 1
- offset = 0

- length = 1500
- ID = x
- fragflag = 1
- offset = 185

- length = 1040
- ID = x
- fragflag = 0
- offset = 370

1480 bytes in data field

offset = 1480/8
How does a host get an IP address?

- Hard-coded by system administrator in configuration
- Determined by network service: DHCP
- other options (⇒ IPv6)
The Dynamic Host Configuration Protocol

arriving client

DHCP discover
src: 0.0.0.0, 68
dest.: 255.255.255.255, 67
yiaddrr: 0.0.0.0
transaction ID 654

DHCP offer
src: 223.1.2.5, 67
dest.: 255.255.255.255, 68
yiaddr: 223.1.2.4
transaction ID: 654
Lifetime: 3600 secs

DHCP request
src: 0.0.0.0, 68
dest.: 255.255.255.255, 67
yiaddr: 223.1.2.4
transaction ID 655
Lifetime: 3600 secs

DHCP ACK
src: 223.1.2.5, 67
dest.: 255.255.255.255, 68
yiaddr: 223.1.2.4
transaction ID 655
Lifetime: 3600 secs

DHCP server: 223.1.2.5
The Internet Control Message Protocol

Communicate network-level information:

- Error reporting
- Diagnostics

<table>
<thead>
<tr>
<th>Type</th>
<th>Code</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0</td>
<td>dest net unreachable</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>dest host unreachable</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>dest prot unreachable</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>dest port unreachable</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>dest net unknown</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>dest host unknown</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>Echo reply</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>Echo request</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>source quench</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>router advertisement</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>router discovery</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>TTL expired</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>bad IP header</td>
</tr>
</tbody>
</table>
Traceroute

- Source send series of UDP segments to destination
- First has TTL=1, second TTL=2, etc.
- $n$-th router discards and sends ICMP TTL expired
- Ultimate destination returns ICMP PORT UNREACHABLE
- traceroute measures average RTTs for each step
How does an ISP get a block of addresses?

- Internet Corporation for Assigned Names and Numbers (ICANN)
How does an ISP get a block of addresses?

- ICANN (not really)
- Internet Assigned Numbers Authority (IANA)
How does an ISP get a block of addresses?

- ICANN (not really)
- IANA (not really)
- Regional Internet Registry (RIR)
How does an ISP get a block of addresses?

• ICANN (not really)
• IANA (not really)
• RIR (IPv4 only until 2011-2012)

“Their Internet usage is growing very rapidly, and even they can do the math: If everyone in China needed an IPv4 address — just one — this country would use up one third of the entire public IP address space.” – Vinton Cerf
Questions
Copyright

Copyright (C) 2010 Christian Grothoff

Verbatim copying and distribution of this entire article is permitted in any medium, provided this notice is preserved.