Networking Fundamentals:
IP, DNS, URL, MIME

Lecture 10
Internet Protocol (IP) Addresses

- A unique 32-bit number
  - Assigned to device connected to internet
  - An address for delivery of packets

- Written in “dotted-decimal” notation
  - Divided into 4 fields separated by “.”
  - Each field is 8 bits, ie 0-255 decimal
    - 10100100011010110111101100000110
    - 10100100.01101011.01111011.00000110
    - 164.107.123.6

- Some are reserved: eg, 127.0.0.1
Abstract Value vs Encoding

- Abstraction: 32-bit integer value
- Encodings
  - Dotted decimal
  - Dotted hex
  - Dotted octal
  - Hexadecimal
  - Decimal
  - Binary
  - Etc...
- Recall: abstraction, representation, correspondence relation
Address Space

- Organizations are allocated blocks of contiguous address to use
- 32 bits means 4 billion addresses
  - Population of the earth: 7 billion
  - Not enough addresses to go around!
- The end is predictable
  - Techniques like NAT developed to help
- In fact, the end has come!
  - Feb 2011: Last block was allocated
IPv6

- 128 bits
  - ~10^40 addresses; we’re good for a while
  - A small (but growing) fraction of IP traffic
    www.worldipv6launch.org/measurements/

- Recommended format (canonical):
  - Divide into 8 fields separated by "::"
  - Each field 16 bits, ie 4 hex digits (0-FFFF)
  - Omit leading 0’s in a field
  - If there are consecutive fields with value 0, compress them as "::"
  - Compress at most one such set of 0’s
    - Otherwise encoding could be ambiguous
    - Compress the longest sequence
Canonical Format: Uniqueness

2001:0db8:0000:0000:0000:ff00:0042:8329
2001:0db8:0000:0000:0000:ff00:0042:8329

2001:db8:0:0:0:ff00:42:8329
2001:db8:0:0:0:ff00:42:8329

2001:db8::ff00:42:8329
2001:db8::ff00:42:8329
Domain Names

- String corresponds to an IP address
  - `cse.ohio-state.edu` is easier than `164.107.123.6`
  - See host, whois

- Case insensitive: Lower-case standard

- A partial map (almost)
  - DNS maps lower-case strings → IP addresses
  - Multiple strings can map to same address!
  - Some strings map to multiple addresses (unusual)!
Domain Name Hierarchy

- Separated by .’s
  - Don’t confuse with dotted decimal!

- Right-to-left hierarchy
  - Top-level domain is right-most field
    - edu, com, net, gov, countries (ca, it, ...)
  - Second-level domain to its left
  - Then third, fourth, etc, no limit
    - www.sos.state.oh.us

- Hostname + Domain Name = Fully Qualified Domain Name (FQDN)
  - stdlinux.cse.ohio-state.edu
Name Servers

- Act as a phonebook for lookup

Client view:
- Given a FQDN, return IP address
- Partial map: FQDNs → IP addresses

Implementation view:
- Hierarchical by domain
- Local caching for recently retrieved items
Protocols

- Systematic ordering of messages
  - Phone rings
  - Callee answers by saying “Hello”
  - Caller answers by saying “Hello”

- Different protocols use different messages, different sequencing, etc
  - In Italy, callee answers by saying “Pronto”
Network Layering: Abstraction

- One protocol is built on top of another
  - Application level: FTP, HTTP, SSH, SMTP, TELNET
  - Transport: TPC, UDP
  - Internet: IP

- Each protocol assumes certain behavior from layer below
  - IP routes packets to destination (unreliable)
  - TCP creates a reliable, in-order channel
  - HTTP delivers web pages
Network Ports

- A single host has many ports
- Application-level protocols have default port
  - ftp -> 20
  - http -> 80
  - imap -> 143
  - ssh -> 22
  - smtp -> 25
  - telnet -> 23

- A “web server” is a running program, waiting, listening for a call (on port 80)
  - See telnet
URL

- Uniform Resource Locator
  
  `scheme://FQDN:port/path?query#fragment`

- Schemes include http, ftp, mailto, file...
  - Case insensitive, prefer lower case

- Port is optional (uses default)
  - 80 for http

- Variety of formats, depending on scheme
  
  `ftp://doe@ftp.cse.ohio-state.edu`
  `mailto://brutus.1@osu.edu`

- FQDN is case insensitive, prefer lower case
Abstract Value and Encoding

- Concrete invariant (convention)
  - No space, ;, :, & in encoding
  - To represent these characters in abstract value, use \%hh instead (hh is ASCII code in hex)
    - \%20 for space
  - Q: What about \% in abstract value?

- Recall: correspondence relation
Document Root

- Web server configured to serve documents from a location in file system
  - “document root”: /class/3901
  - File: /class/3901/labs/lab2.html
  - URL: http://www.cse.osu.edu/labs/lab2.html

- Slashes in path should be for server’s OS (but forward slashes are common)
- Virtual servers: multiple doc roots
- Proxy servers: remote doc roots
MIME

- Multipurpose Internet Mail Extensions
  - Used to be for mail attachments
- How to interpret a file
  - File is a blob of bits (encoding)
  - How to map this blob into (abstract) value? Colors, sounds, characters, etc?
  - Recall: correspondence relation
- Syntax: type/subtype
  - plain/text, plain/html
  - image/gif, image/jpg
  - video/mpeg, video/quicktime
Determining MIME Type

- The sender (web server) determines MIME type of document being sent
  - Rules map file extensions to MIME types
- If file arrives without MIME info, receiver has to guess
  - File extension may suggest the MIME type
- Some types handled by browser itself
- Others require plugin or application
- Experimental MIME subtypes: x-
  - application/x-gzip
Summary

- IP address are unique on network
  - IPv4 vs IPv6
- DNS maps strings to IP addresses
  - Domains nested hierarchically
- URLs identify resources on network
  - Scheme, host, path
- MIME type defines a file’s encoding
  - Correspondence