COMP10001 Foundations of Computing
HTML and the Internet

Semester 2, 2016

Chris Leckie

September 12, 2016

Lecture Agenda

- Last lecture:
  - Recursion
- This lecture:
  - URLs and the Internet
  - HTML basics

Lecture Outline

1. The Internet
2. HTML

Addressing Machines: IPs I

- Each device on the Internet has a unique "IP address"
- In IPv4, IP addresses are represented as 4 "8-bit" integers, each in the range $[0, 255]$, e.g. 128.250.36.33 is Tim's main web server
- IP addresses can be allocated to a device either "statically" (an IP is reserved for a given device) or "dynamically" (an IP is allocated to a device dynamically when it connects to the Internet)

Addressing Machines: IPs II

- IP addresses can also be either "local" (addressable only on a private network) or "external" (directly addressable from anywhere on the internet)
  - in IPv4, a number of address ranges, including 192.168.x.x, are reserved for local IPs
- Devices with local IPs can often still access the open Internet through "proxy servers" or "gateways"
- There is increasing momentum to move to IPv6 (8×4-digit "hexadecimal" numbers) because we are rapidly running out of IP addresses

Addressing Machines: Hostnames

- Humans tend to find sequences of numbers hard to remember, so devices also tend to have "hostnames" such as `hum.csse.unimelb.edu.au` made up of (case-insensitive) letters and full stops
- Hostnames are structured hierarchically relative to a domain name (e.g. `unimelb.edu.au`) and end with a "top-level domain" (TLD, e.g. `au`)
- Hostnames resolve to IP addresses via the Domain Name System (DNS) using "name servers"

Quiz

- Can two hostnames share the same IP address?
- Can one hostname represent more than one IP address?

Structure of the Internet

- Devices address each other by their IP, but communicate via "Internet routers"

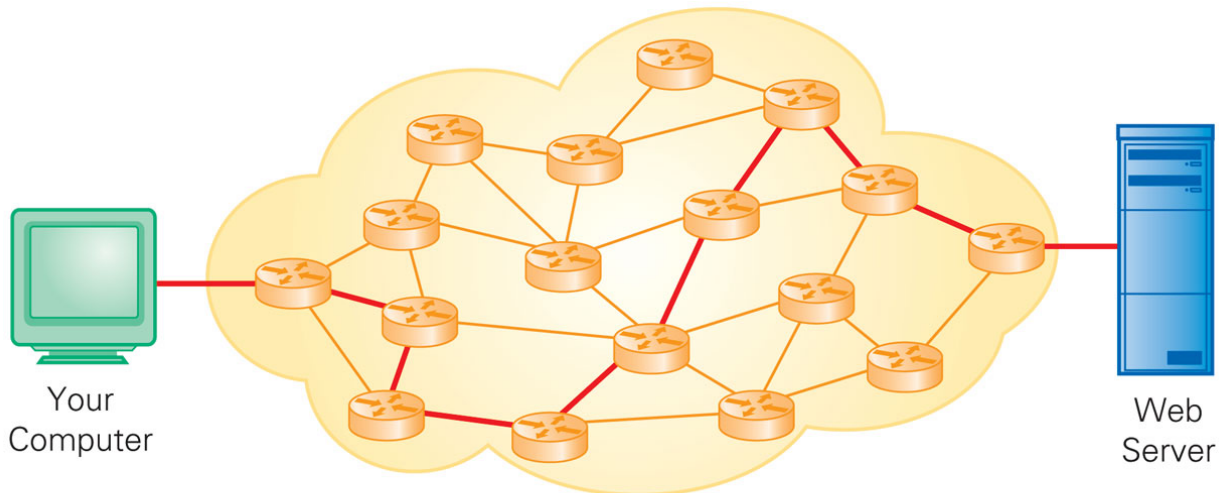
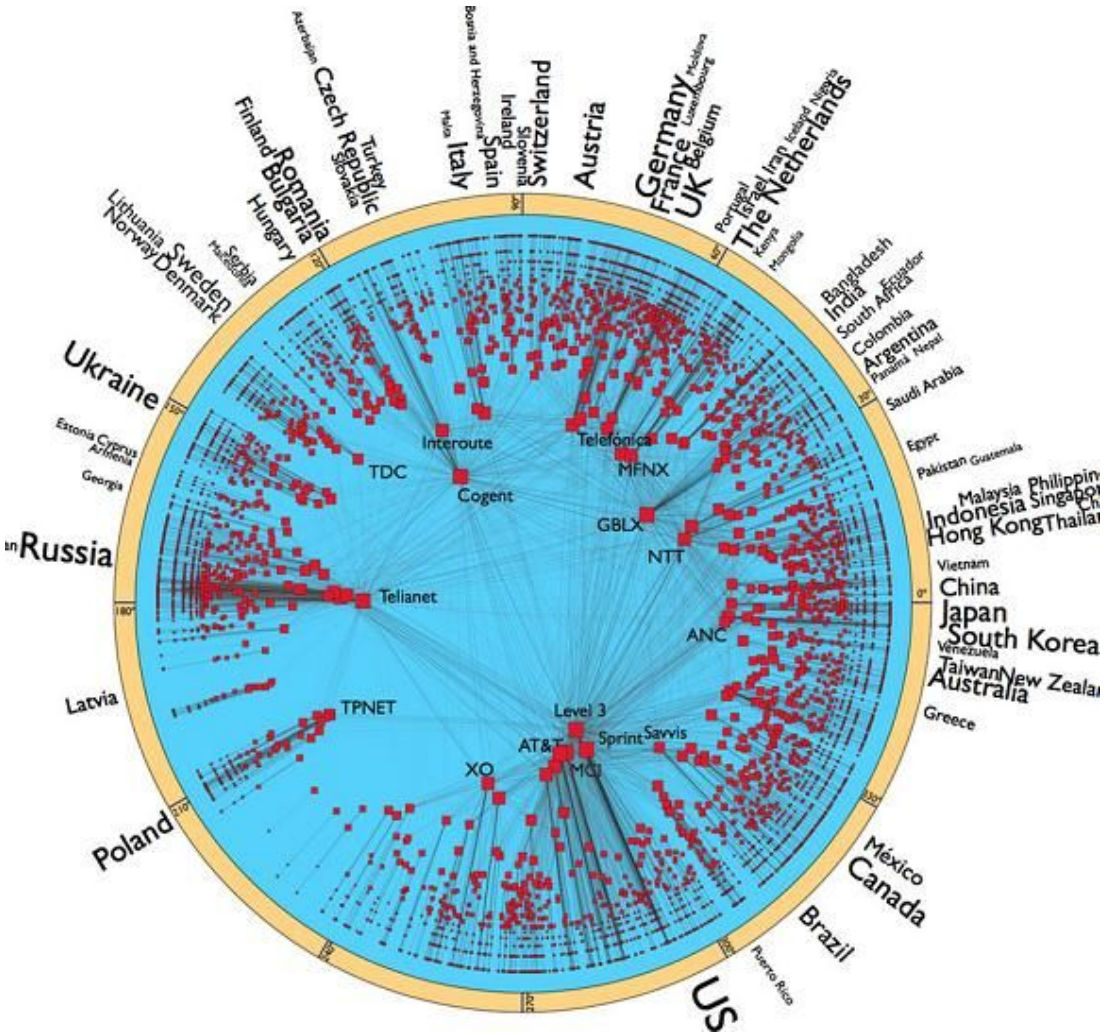


Figure 3.1. A diagram of the Internet.

- Communication is from a "client" to a "server"

Hyperbolic Map of the Internet



Source(s): `http://www.nature.com/ncomms/journal/v1/n6/full/ncomms1063.html`

Applications and Ports

- Multiple applications run over IP networks:
  - Email
  - World Wide Web (HTTP)
  - FTP
  - Chat/Instant Messaging
  - Video Streaming
- Machines communicate with other machines over the Internet via a collection of numbered "ports", differentiated by application:
  - 21 = FTP
  - 25 = outgoing email (SMTP)
  - 80 = HTTP
  - 110 = incoming email (POP3)

URLs

- Internet resources are "addressed" via URLs ("Uniform Resource Locators"):
  `http://nlp.stanford.edu:8080/parser/`
  `ftp://ftp.unimelb.edu.au/pub/www/ughb-book2007.tar.gz`
  `mms://www.microsoft.com/videos/a_streaming_video.wmv`
- URLs are made up of the following parts:

  scheme://hostname:port/path

  where:
  - scheme = the "protocol" for accessing the file
  - hostname = the device the file lives on
  - port = access port (optional)
  - path = where the file lives on the device

URL: Examples

- `http://nlp.stanford.edu:8080/parser/`
  - protocol = HTTP
  - hostname = `nlp.stanford.edu`
  - port = 8080
  - path = `parser`
- So what happens when I type `google.com` into my browser?
  - protocol = HTTP (client-side default)
  - hostname = `www.google.com` (client-side default)
  - port = 80 (by default from protocol)
  - path = `index.html` (or similar; server-side default)

Lecture Outline

HTML: Introduction

- The primary language used to "mark up" web documents is HTML ("Hypertext Markup Language"); we will focus on HTML5
- HTML is made up of "elements" ("tags" and "entities"), which are used to mark up "content"
- HTML tags are enclosed in "angle brackets" (e.g. `<tag>`), and take the form of: (1) "empty elements" (e.g. `<tag/>`; note backslash at *end* of tag), or (2) tag pairs (e.g. `<tag></tag>`; note backslash at *start* of closing tag)
- HTML tags may optionally contain "attributes"

HTML: Mark-up Basics

- Given some (textual) content:

```
How much wood could a woodchuck chuck
```

  we mark up regions with tag pairs, e.g.:

```
How much <b>wood</b> could a <i>wood</i>chuck chuck
```

  which renders as:

How much **wood** could a *wood*chuck chuck

- The basic textual tag pairs are:
  - `<i></i>`: *italics*
  - `<b></b>`: **bold**
  - `<u></u>`: <u>underline</u>

HTML: Stacking up Mark-up

- It is possible to stack up mark-up, but tags have to be closed in the reverse order of opening (a la a "stack"), i.e. must be nested within one another, e.g.

```
How much <u><b>wood</b></u> could a <i>wood</i>chuck chuck
```

which renders as:

How much **<u>wood</u>** could a *wood*chuck chuck

Lecture Summary

- What are IPs and hostnames, and what is their role?
- What are ports?
- What are URLs and how are they structured?
- What are HTML elements, tags, attributes and entities?