COMP519
Web Programming

Autumn 2015

Introductory Lectures

COMP519: Web Programming

- Lecturer: Russell Martin, Room 319 Ashton Building,
  email: `Russell.Martin(at)liverpool.ac.uk`

- Lectures: Monday (2pm), Wednesday (10am) in Brodie
  Tower, Room 702

  Practicals: Monday (4pm), Friday (10am) in George Holt,
  Lab 1

- Assessment: 4 programming assessments (75%) and
  multiple-choice final exam in January (25%)

- Course web page:
  `http://www.csc.liv.ac.uk/~martin/teaching/comp519/`

Module Outcomes

By the end of this module a student should

1. Understand some of the technology and protocols underlying the World Wide Web;

2. Become familiar with common tools and techniques for developing Web-based applications, both client-side and server-side; and

3. Develop a working knowledge of HTML, JavaScript, Python, and PHP as languages for developing Web applications.

Reasonable questions about the Internet and the Web

- ▶ What is the World Wide Web?

- ▶ Is it the same thing as the Internet?

- ▶ Who invented it?

- ▶ How old is it?

- ▶ How does it work?

- ▶ What kinds of things can it do (or what can we do using it)?

- ▶ What does it have to do with programming?

Web $\neq$ Internet

- ▶ Internet : a physical network connecting millions (or billions) of computers (and other devices) using the same protocols for sharing and transmitting information (TCP/IP).

  In reality, the Internet is a network of smaller networks.

- ▶ World Wide Web: a collection of interlinked multimedia documents that are stored on the Internet and accessed using a common protocol (HTTP).

Key distinction: Internet is hardware while the Web is software, along with data, documents, and other media.

Many other Internet-based applications exist, e.g. email, telnet, ftp, usenet, instant messaging services, file-sharing services, . . .

A (Very Brief) History of the Internet

The idea of a long-distance computer network traces back to early 1960s.

- ▶ Joseph Licklider at M.I.T. (a "time-sharing network of computers", the Intergalactic Computer Network).
- ▶ Paul Baran at RAND Corporation (tasked with designing a "survivable" communications system that could maintain communication between end points even after damage from a nuclear attack).
- ▶ Donald Davies at National Physics Laboratory in U.K. (developed packet-switching methodology to reduce costs associated with maintaining open transmissions).

In particular, the US Department of Defense was interested in the development of distributed, decentralized networks.

- ▶ survivability (i.e., network still functions despite a local attack)
- ▶ fault-tolerance (i.e., network still functions despite local failure)

A (Very Brief) History of the Internet (cont.)

The idea of a survivable and fault-tolerant network was very different from conventional highly-centralized services such as the phone networks and electrical power generation facilities.

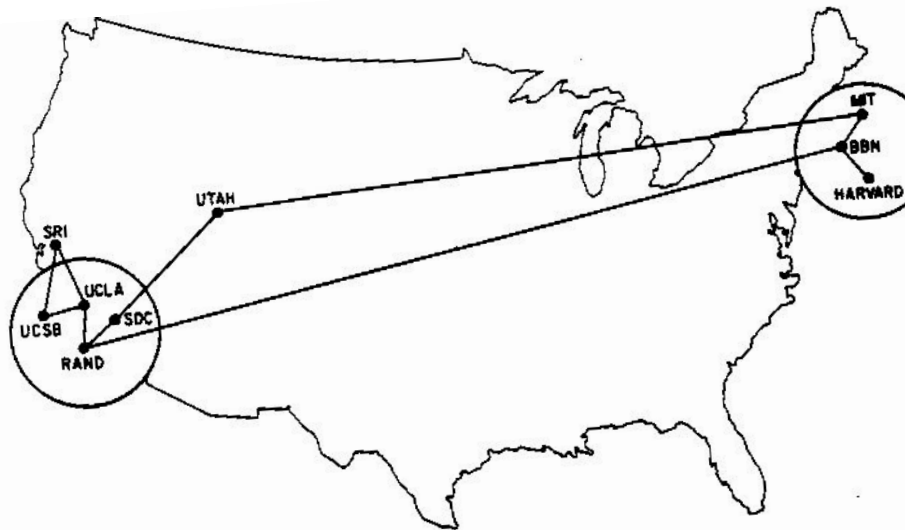In 1969, the Advanced Research Project Agency funded the ARPANET.

- ▶ Connected computers at UC Los Angeles, UC Santa Barbara, Stanford Research Institute, and University of Utah.
- ▶ Allowed researchers to share data, communicate 56Kb/sec communication lines (vs. 110 b/sec over phone lines).
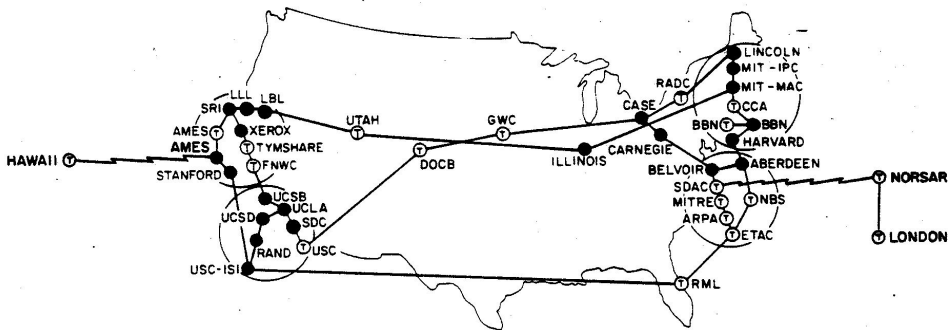
ARPANET (1969)
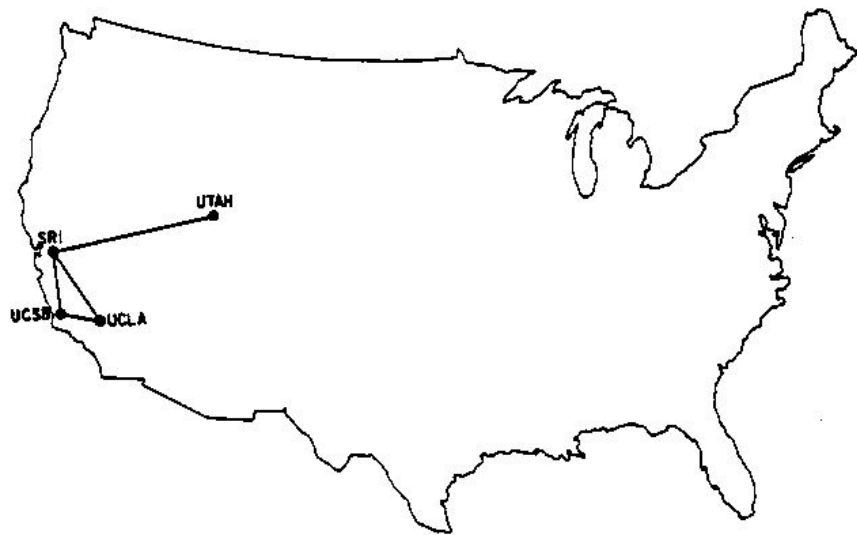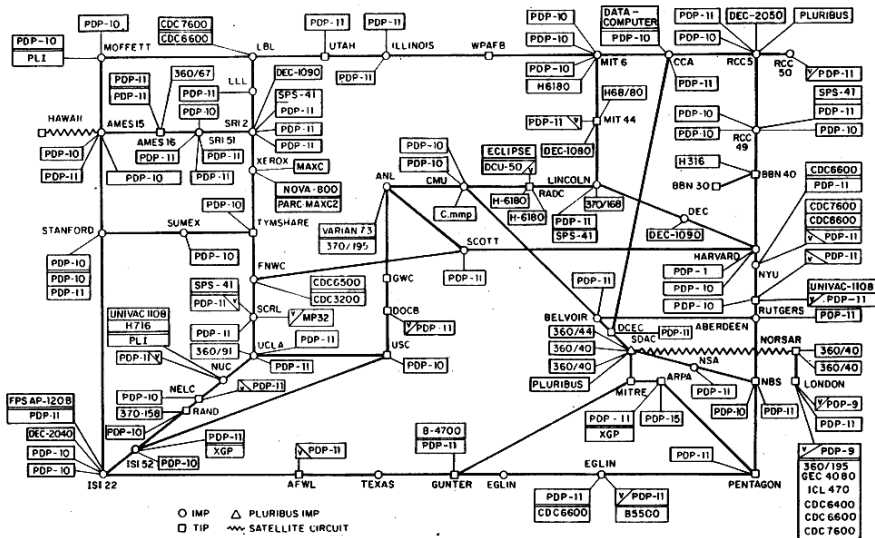
ARPANET (1970)

ARPANET (1973)

ARPANET (1977, from the Computer History Museum)

ARPANET LOGICAL MAP, MARCH 1977

(PLEASE NOTE THAT WHILE THIS MAP SHOWS THE HOST POPULATION OF THE NETWORK ACCORDING TO THE BEST INFORMATION OBTAINABLE, NO CLAIM CAN BE MADE FOR ITS ACCURACY)

NAMES SHOWN ARE IMP NAMES, NOT (NECESSARILY) HOST NAMES

○ IMP △ PLURIBUS IMP
□ TIP ∿ SATELLITE CIRCUIT

Internet Growth

Throughout the 1970s, the size of the ARPANET doubled every year.

- ▶ First ARPANET e-mail sent in 1971.
- ▶ Decentralization made adding new computers easy.
- ▶ TCP/IP developed in the mid 1970s for more efficient packet routing.
- ▶ Migration of ARPANET to TCP/IP completed 1 January, 1983.
- ▶ ~1000 military & academic host computers connected by 1984.

In the 1980s, the U.S. government took a larger role in Internet development.

- ▶ Created NSFNET (National Science Foundation Net) for academic research in 1986.
- ▶ ARPANET was retained for military & government computers.

Internet Growth (cont.)

By the 1990s, the Internet connected virtually all colleges & universities.

Businesses and individuals also connecting as computing costs fell.

$\sim$1,000,000 computers by 1992.

In 1992, control of the Internet was transferred to a non-profit organization.

Internet Society: Internet Engineering Task Force
                    Internet Architecture Board
                    Internet Assigned Number Authority
                    World-Wide-Web Consortium (W3C)
                    . . .

Internet Growth (cont.)

Internet has exhibited exponential growth, doubling in size every 1-2 years (stats from Internet Software Consortium).

2,802,478,934 Internet users (approx. 39% of the world's population) (Source: Internet World Stats) (Dec, 2013)

Year	Computers on the Internet (at any one time?)
2011	~605,000,000
2006	439,286,364
2004	285,139,107
2002	162,128,493
1998	36,739,000
1996	12,881,000
1994	3,212,000
1992	992,000
1990	313,000
1988	56,000
1986	5,089
1984	1,024
1982	235

Internet Growth (cont.)

United Kingdom has 52.7 million users (approx. 89.8% of the population).

Partial map of the internet (circa Jan 2005, by the Opte Project (opte.org)

A (Very Brief) History of the World Wide Web

The idea of hypertext (cross-linked and inter-linked documents) traces back to Vannevar Bush in his 1945 *The Atlantic Monthly* article "As We May Think" (the "memex").

Online hypertext systems began to be developed in 1960s. E.g., Ted Nelson and Andy van Dam's Hypertext Editing System (HES), Doug Englebert's NLS (oN-Line System).

Aspen Movie Map (allowing virtual tour of Aspen, Colorado) was the first hypermedia application.

In 1987, Apple introduced HyperCard (a hypermedia system that predated the WWW).

History of the Web (cont.)

In 1989, Tim Berners-Lee at the European Particle Physics Laboratory (CERN) designed a hypertext system for linking documents over the Internet.

He designed a (Non-WYSIWYG) language for specifying document content which evolved into HyperText Markup Language (HTML).

Berners-Lee also designed a protocol for downloading documents and interpreting the content which evolved into HyperText Transfer Protocol (HTTP).

And he implemented the first browser – text-based, no embedded media.

And the Web was born!

History of the Web (cont.)

The Web was a generally obscure European research tool until 1993.

In 1993, Marc Andreessen and Eric Bina (at the National Center for Supercomputing Applications, a unit of the University of Illinois) developed Mosaic, one of the early graphical Web browsers that popularized the WWW for the general public (Erwise was the first one, ViolaWWW the second).

The intuitive, clickable interface helped make hypertext accessible to the masses made the integration of multimedia (images, video, sound, . . .) much easier.

Andreessen left NCSA to found Netscape in 1994. Netscape was (is?) a cheap/free browser and further popularized the Web (with a 75% market share in 1996).

History of the Web (cont.)

In 1995, Microsoft came out with Internet Explorer, version 1, based on licensing Spyglass Mosaic from Spyglass, Inc. In 2002–2003, Internet Explorer was estimated to have 95% market share.

Opera web browser released in 1996.

Netscape bought by AOL in 1998 for US$4.2 billion in stock.

Firefox web browser, version 1.0, released in November 2004. (As of Sept 2015, Firefox is on version 40.0.3).

Google Chrome released in 2008. In May 2012, Chrome overtook Internet Explorer as the most used browser worldwide.

In 2015, Chrome has an estimated 45% usage based on the last 15,000 hits on approximately 80,000 websites, with Safari second at 16.2%, Firefox third at 14.4%, and Internet Explorer fourth at 13.4% (source: W3Counter).

Web Growth

Today, the Web is the most visible aspect of the Internet.

Stats on right from Netcraft Web Server Survey.

Year	Computers on the Internet	Websites on the Internet
2015	?????	~893,000,000
2011	605,000,000	346,004,403
2008	?????	172,338,726
2006	439,286,364	85,507,314
2004	285,139,107	51,611,646
2002	162,128,493	38,760,373
2000	93,047,785	17,087,182
1998	36,739,000	2,410,067
1996	12,881,000	257,601
1994	3,212,000	2,738
1992	992,000	10

Difficulties caused by Internet and Web growth

Internet addresses are used to identify computers on the internet.

Internet Protocol version 4 (IPv4) was first defined in 1981 and is still in use today, but this uses a 32-bit number to specify addresses. (The IP addresses you are used to seeing like 138.253.187.203 or 141.101.120.15.)

IPv4 provides around 4.29 billion (i.e. 4.29×10^9) addresses.

For comparison, the estimated world population (July 2015) is about 7.3 billion (i.e. 7.3×10^9).

Difficulties caused by Internet and Web growth (cont.)

The Internet Assigned Numbers Authority exhausted their allocated addresses on February 3, 2011, and the Asia-Pacific Network Information Centre (regional internet registry) exhausted theirs on April 15, 2011.

Also, the Réseaux IP Européens Network Coordination Centre (regional internet registry for Europe, the Middle East, and parts of central Asia) exhausted their allotment on September 14, 2012.

LACNIC (the registry for Latin America and the Caribbean) depleted its address pool on June 10, 2014.

Other registries are expected to reach exhaustion within several years.

IPv6 was formally specified in 1999 and uses 128 bit addresses, but is also redesigned to allow more efficient routing, network aggregation, and ease of network reconfiguration. (And has about 3.4×10^{38} addresses.)

Difficulties caused by Internet and Web growth (cont.)

The main difficulty is that computers using IPv4 cannot "talk directly" to computers that use IPv6 (and vice-versa). Thus, various work-arounds and address-translation schemes are in use to mitigate communications between machines using different protocols.

All of these difficulties can result in delays in loading webpages and other communication problems.

The long-term solution is deployment of IPv6 to all machines. IPv6 has been in active deployment since June 2006, but as of 2014 more than 99% of all Internet traffic was carried with IPv4.

The changeover to IPv6 takes time and money...

Static vs. Dynamic Webpages

Many Web pages are still static in nature.

- ▶ Contents (text/links/images) are the same each time the webpage is accessed.
  e.g., online documents, many personal homepages

HyperText Markup Language (HTML) and Cascading Style Sheets (CSS) are used to specify text, image, and page format, along with styling the page for various effects (backgrounds, colors, table layout, page margins, etc.).

Static vs. Dynamic Webpages (cont.)

As the Web continues towards more and more online services and e-commerce continues to grow, Web pages must also provide dynamic content.

- ▶ Pages can be fluid, changeable (e.g., rotating banners, inclusion of "real-time" data, etc.).
- ▶ Must be able to react to the user's actions, request and process info, tailor services.
  e.g., amazon.com, YouTube, any e-commerce website, online email services, etc.

This course is about applying your programming skills to the development of dynamic webpages and web-based applications.

- ▶ Client-side programming
- ▶ Serve-side programming

Client-side Programming

Can download program with a webpage, execute the program on the client's machine.

Simple, generic, but sometimes insecure (e.g. cross-site scripting attacks).

JavaScript

- A scripting language for Web pages, developed by Netscape in 1995.
- Uses a C++/Java-like syntax, so familiar to many programmers, but simpler.

Client-side Programming (cont.)

JavaScript is good for adding dynamic features to Web page, controlling forms, and the GUI.

Requires users to have this technology enabled on their browsers.

JQuery, first released in 2006, is a JavaScript library to help with cross-browser compatibility issues.

JQuery is used by more than 65% of the 10,000,000 top websites (source: W^3Techs).

Java applets

- ▶ Can define small, special-purpose programs in Java called applets.
- ▶ Provides (almost) full expressive power of Java (but with more overhead).

Good for data-heavy tasks or more complex tasks such as graphics.

Server-side Programming

Can store a program on a web server, and supply a link from a webpage to execute that program. And you can also accept input from a user in terms of "filling in blanks" and/or file upload(s), etc.

The process of doing this can be more complex, requires server privileges, but can still be (mostly) secure with proper precautions.

Common Gateway Interface (CGI) programming

- ▶ Programs are written to conform to the CGI.
- ▶ When a webpage submits, the data from the page is sent as input to the CGI program. CGI program executes on the server and sends its results back to browser as a webpage.

Good if computation is large/complex or requires access to private data.

We will discuss CGI programming using Python, but other languages are possible as well (such as Perl, Ruby, etc.).

Server-side Programming (cont.)

Other server-side programming technologies include:

- ▶ Active Server Pages (ASP)
- ▶ Java Servlets
- ▶ PHP
- ▶ Server Side Includes
- ▶ Ajax (using JavaScript on the client side too)

Some of these are vendor-specific alternatives to CGI (such as Microsoft's ASP).
They often provide many of the same capabilities as CGI programs but use HTML-like tags (such as PHP).

Some of these technologies might require functionality to be enabled in the client's browser (e.g. Ajax generally requires the use of Javascript combined with PHP or some other server-based programming component).

An Exercise

Pick some of your favorite websites and try to identify

- the static components
- the dynamic components
    - Which sites are using JavaScript?
    - Which are using Java applets?
    - Which are using server-side elements such as CGI programs?