Introducing HTTP and Webservers

CS174

Chris Pollett

Aug. 25, 2008.

Outline

- The Internet
- Web Browsers
- Web Servers

The Internet

- The internet began as ARPANET in 1969.
- The internet uses **TCP/IP** (Transmission Control Protocol/ Internet Protocol) to allow one computer to communicate with another.
- An IP address of a machine is a 32 bit number. (Four bytes).
- Routing computers use each of these bytes to figure out where to send message packets.
- Organizations are assigned blocks of IP addresses which they in turn use to assign an individual IP to a computer.
- For example, a small organization might have the addresses from 192.57.126.0 to 192.57.126.255.
- IP concerns routing packets, TCP uses packets to implement an end-to-end connection/channel between sender and receiver.
- Several protocols run on top of TCP/IP such as telnet (allows you to log in to a machine over the internet) and FTP (allows you to transfer files over the internet).

Domain Names

- To try to make computer addresses more meaningful, computers can be given text names such as:

  peano.mynetwork.org

- The whole name above is called a **fully qualified domain name**.
- The conversion from domain names to IP addresses is done by having your computer contact a domain name server (DNS) which has a conversion table.

The World Wide Web

- For this semester, the protocol on top of TCP/IP we will be interest in is called HTTP (Hypertext Transfer Protocol).
- It is used to retrieve from servers hypertext documents (one's with links and images. i.e., HTML) over the internet.
- It was first proposed in 1989 by a group including Tim Berners Lee at CERN.
- The World Wide Web is the collection of servers and client that use this protocol.

Web Browsers

- Are clients used to request hypertext documents from HTTP servers.
- The first browser developed by Berners-Lee was called WWW.
- The first widely used browser was NCSA Mosaic.
- It served as the basis for the first commercial browser Netscape.
- The most common modern browsers are: Internet Explorer, Firefox, Safari, and Opera.
- The site evolt.org has an executables for almost every browsers that has ever existed.
- Today, even most cell phones have browsers which support W3C compliant HTML and Javascript/ECMAscript.
- As a web programmer you should code to standards not any particular browser. In fact, large organizations (such as SJSU) might be required to produce only pages which conform to standards such as accessibility standards.

Web Servers

- The two most common servers of web documents are Apache (~70%) and IIS (~20%) (Microsoft Internet Information Server.)
- As Apache is free, I expect you to have it downloaded and installed by the time the first homework is due.
- There are two main variants of Apache: 1.3.x and 2.x. It should not matter which you use for this course.

Web Server Operation

- In a typical exchange between a browser and a web server, the browser telnets into port 80 of the web server and makes a GET request for a particular document.
- A web server has two sets of directories, the root of one of these is called the *document root* of the server. The web server looks in this directory to try to find the file to service the clients request.
- The root of the other directory the server uses is called the *server root*. Under it is stored the server and its support software.
- To initiate an exchange as above, the user typically supplies the browser with the protocol, say http; the machine, say peano.mynetwork.org; and the path to the file /index.html in one string http://peano.mynetwork.org/index.html called a **URL**.