

CSC 498: Web Programming

Haidar Harmanani

Department of Computer Science and Mathematics
Lebanese American University
Byblos, 1401 2010 Lebanon

1



Wednesday, October 15, 2008

1

Internet History

- The first node of ARPAnet was established at UCLA in 1969.
- In the late 1970s and 1980s, many networks were developed (BITNET, CSNET, NSFnet) locally.
- Internet is a huge collection of computers connected in a communications network.
- The Internet cannot continue to grow indefinitely. So... when will it stop?

2

9/28/08

CSC 498: Web Programming



Wednesday, October 15, 2008

2

Estimation of Computers connected to the Internet

Year	Number of Computers on the Internet
1983	562
1984	1,024
1990	290,000
1991	500,000
1995	4,852,000
1996	9,472,000
1999	43,230,000
2000	73,000,000
2001	...

3

9/28/08

CSC 498: Web Programming



Wednesday, October 15, 2008

3

Network Protocols

- It is a special set of rules that end points in a telecommunication connection use when they communicate.
- Transmission Control Protocol (TCP) – To exchange messages with other Internet points at the information packet level.
- Internet Protocol (IP) – To send and receive messages at the Internet Address level.
- Hypertext Transfer Protocol (HTTP) – To deliver HTML, sound, audio files on the World Wide Web.

4

9/28/08

CSC 498: Web Programming



Wednesday, October 15, 2008

4

Internet Protocol Address

- The Internet devices are identified by
 - Names for people
 - Numeric addresses for computers
- The IP address is a unique 32-bit number
 - Divided into 4 parts (1 byte each)
 - 191.57.126.0
- Next generation : IPv6 – 128-bit



Domain Name & DNS

- People have difficulty dealing with and remembering numbers, machines on the Internet also have textual names.
 - www.csc.byblos.lau.edu.lb
- There are name servers on the Internet that provide Domain Name Services (DNS).
 - Returns the name or IP address of any machine in the domain.



The World Wide Web

- Affectionately called “The Web”
- It is a collection of information stored on the networked computers over the world.
- The WWW was proposed in 1989 by Tim Berners-Lee at CERN.



How does the Web work?

- The web information is stored in Web pages.
- The web pages are stored in computers that are called “Web servers.”
- The computer reading the pages is called a web client
 - Most commonly Internet Explorer or Firefox.
- The web server waits for the request from the web clients over the Internet.
 - Internet Information Server (IIS) or Apache.



Web or Internet?

- They are not the same thing.
- The Internet is a collection of computers or networking devices connected together.
 - They have communication between each other.
 - Decentralized design that there is no centralized body controls how the Internet functions.
- The Web is a collection of documents that are interconnected by hyper-links.
 - These documents are accessed by web browsers and provided by web servers.

9 9/28/08

CSC 498: Web Programming



Wednesday, October 15, 2008

9

Internet Terminology

- Client
 - Any computer on the network that requests services from another computer on the network.
- Server
 - Any computer that receives requests from client computers, processes and sends the output.
- Web Page
 - Any page that is hosted on the Internet.
- Web Development
 - The process of creating, modifying web pages.

10 9/28/08

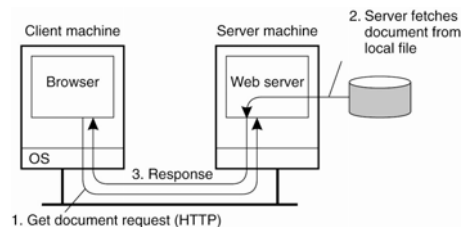
CSC 498: Web Programming



Wednesday, October 15, 2008

10

Traditional Web-Based Systems



- Essence: The WWW is a huge client-server system with millions of servers; each server hosting thousands of hyperlinked documents:
 - Documents are generally represented in text (plain text, HTML, XML)
 - Alternative types: images, audio, video, but also applications (PDF, PS)
 - Documents may contain scripts that are executed by the client-side software

11 9/28/08

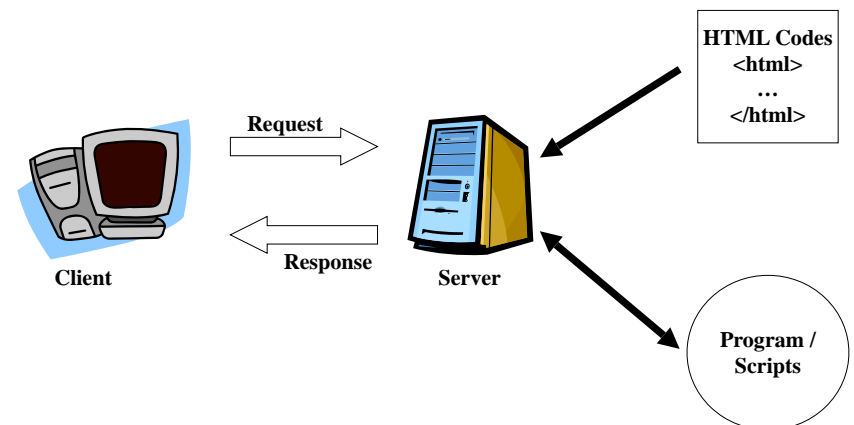
CSC 498: Web Programming



Wednesday, October 15, 2008

11

The HTTP Request/Response Model



12 9/28/08

CSC 498: Web Programming



Wednesday, October 15, 2008

12

Valid HTTP Request/Response mesg

- Provides additional information

```
GET /index.html HTTP/1.0
Host: www.anyhost.com
User-Agent : Mozilla/4.5 [en] (WinNT; I)
Accept : image/gif, image/jpeg, */*
Accept-language : en
Accept-charset : iso-8859-1, *, utf-8

HTTP/1.0 200 OK
Last-Modified: Mon, 20 Dec 1999 ...
Date: Tue, 11 Jan 2002 ...
Status: 200
Content-Type: text/html
Servlet-Engine: Tomcat Web Server
Content-Length: 59

<html>
...
</html>
```

13 9/28/08

CSC 498: Web Programming



Wednesday, October 15, 2008

13

Web Browser (Web Client)

- It is a program that retrieves information from the Web.
 - Microsoft Internet Explorer
 - Most commonly used browsers
 - Runs only on PCs
 - Firefox
 - Runs on many different computing platforms
 - Safari
 - From Apple
 - Opera
 - The fastest browser on Earth
 - Supports Bittorrent transfer
 - Lynx
 - Text based web client
 - Chrome
 - Recent web browser from Google

14 9/28/08

CSC 498: Web Programming



Wednesday, October 15, 2008

14

Some Statistics

- Until July 02

– Internet Explorer 6.x	39%		
– Internet Explorer 5.x	51%		
– Internet Explorer 4.x	2%		
– Netscape	3%		
– Others	1%		
– Win98/ME	64%	, Win 95	4%
– WinNT	5%	, Win2000	20%
– MAC	1%	, www.thecounter.com	

15 9/28/08

CSC 498: Web Programming



Wednesday, October 15, 2008

15

Internet Explorer

- Version 1.0 - August 1995
- Version 2.0 - November 1995
- Version 3.0 - August 1996
- Version 4.0 - October 1997
 - Support CSS & DOM, but no XML
- Version 5.0 - March 1999
- Version 5.5 - July 2000
- Version 6.0 - August 2001
- Version 7.0 - October 18, 2006
- Version 8.0 - August 2008 [Beta 2]

16 9/28/08

CSC 498: Web Programming



Wednesday, October 15, 2008

16

Netscape Navigator

- Netscape 1.0 - December 1994
- Netscape 2.0 - March 1996
- Netscape 3.0 - August 1996
- Netscape 4.0
- Mozilla
- Netscape 5.0 - Skipped
- Netscape 6.0 - November 2000
- Netscape 6.1 - August 2001
- Netscape 6.2 - November 2001
- Netscape 7.0 - August 2002



Mozilla Firefox

- In 1998, the Mozilla Organization was formed
 - Largely funded by Netscape
 - Coordinated the development of Netscape 5 based on the communicator source code
 - Too difficult to work with, it was decided to scrap Netscape 5 and re-write the source code
 - The re-written source code was in the form of the Mozilla web browser, which, with a few additions, Netscape 6 was based on.
 - Netscape discontinued its browser on 28 December 2007 after releases 7 and 8 due to low market share



Mozilla Firefox

- The Firefox project went through many versions
 - Version 1.0 was released on November 9, 2004.
 - Version 1.5 was released on November 29, 2005
 - Version 2 was released on October 24, 2006
 - Included updates to the tabbed browsing environment, the extensions manager, the GUI, and the find, search and software update engines; a new session restore feature; inline spell checking; and an anti-phishing feature which was implemented by Google as an extension, and later merged into the program itself.
 - In December 2007, Firefox Live Chat was launched
 - Mozilla Firefox 3 was released on June 17, 2008
 - Now mostly funded by Google!



Other Browsers

- Safari
 - From Apple
 - Windows version released
- Opera
 - Supposedly one of the fastest
- Chrome
 - Open Source
 - Released on September 2, 2008 by Google
 - Has good security support



Recent Statistics [2008]

- Internet Explorer 73.81%
- Firefox 18.43%
- Opera 0.71%
- Safari 6.14%
- Netscape 0.62%
- Other 0.29%
- Chrome ?

Source: Net Applications



Web Server

- It is a program that waits for requests from the web browser.
- It provides four major functions
 - Serving web pages
 - Running gateway programs (CGI) and returning output
 - Controlling access to the server
 - Monitoring and logging all access
- E.g. Apache, IIS, Netscape Web server, ...



Microsoft Internet Information Services (IIS)

- An enterprise-level Web server that is included with Windows
- FTP Site
 - Used for transferring large files across the Internet
- HTTP Site
 - Used most frequently to request documents from Web servers
- SMTP Virtual Server
 - Sends and receives electronic mail
- Web Site Content Directory
 - Directory containing the documents that clients will view



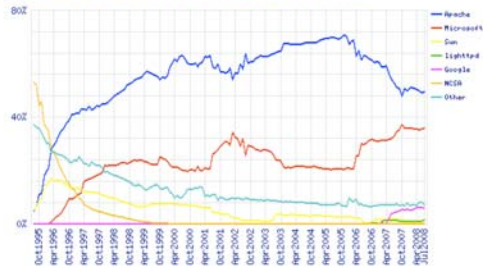
Apache Web Server

- Currently the most popular Web server
- Stability
- Efficiency
- Portability
- Open-source



Web Server Statistics

Market Share for Top Servers Across All Domains August 1995 - July 2008



Top Developers

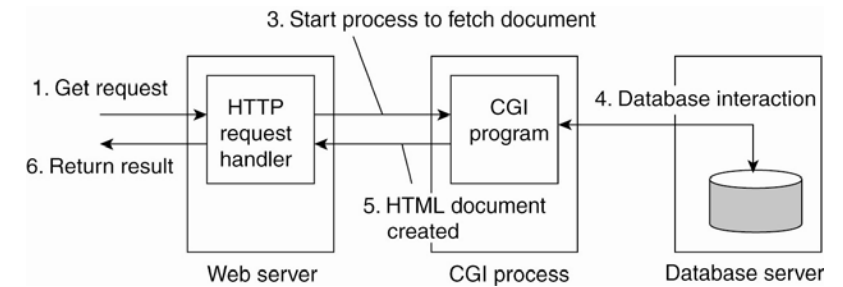
Developer	June 2008	Percent	July 2008	Percent	Change
Apache	84,647,780	49.12%	86,845,154	49.49%	0.37
Microsoft	60,995,528	35.39%	62,411,537	35.57%	0.17
Google	10,468,720	6.07%	10,001,763	5.70%	-0.37
lighttpd	1,532,952	0.89%	2,942,469	1.68%	0.79

• http://news.netcraft.com/archives/web_server_survey.html

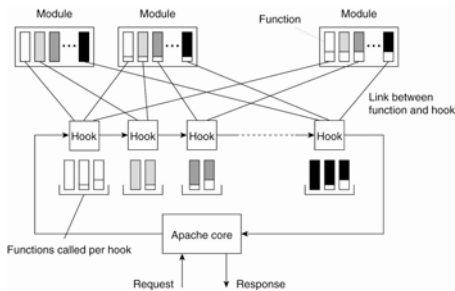


Multitiered Architectures

- Web sites were organized into three tiers:



The Apache Web Server

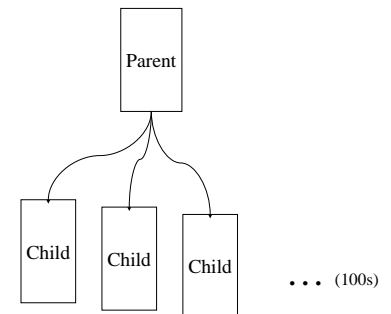


- Around 50% of all Web sites are based on Apache.
- The server is internally organized more or less according to the steps needed to process an HTTP request



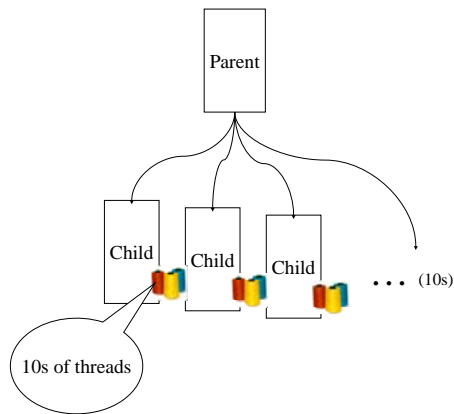
Classic "Prefork" Model

- Apache 1.3, and
- Apache 2.0 Prefork
- Many Children
- Each child handles one connection at a time.



Multithreaded "Worker" Model

- Apache 2.0 Worker
- Few Children
- Each child handles many concurrent connections.



9/28/08

CSC 498: Web Programming

29



Wednesday, October 15, 2008

29

Dynamic Content: Modules

- Extensive API
- Pluggable Interface
- Dynamic or Static Linkage

30

9/28/08

CSC 498: Web Programming



Wednesday, October 15, 2008

30

In-process Modules

- Run from inside the httpd process
 - CGI (mod_cgi)
 - mod_perl
 - mod_php
 - mod_python
 - mod_tcl

31

9/28/08

CSC 498: Web Programming

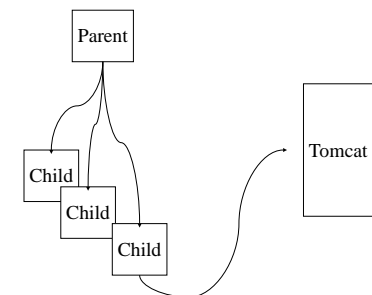


Wednesday, October 15, 2008

31

Out-of-process Modules

- Processing happens outside of httpd (eg. Application Server)
- Tomcat
 - mod_jk/jk2, mod_jserv
- mod_proxy
- mod_jrun



9/28/08

CSC 498: Web Programming

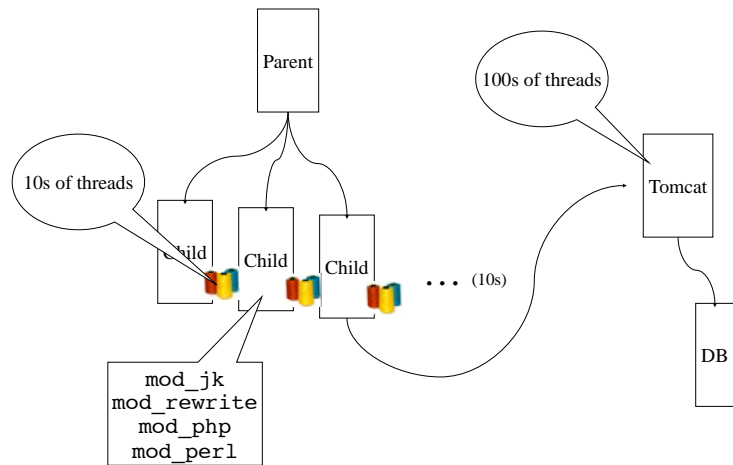
32



Wednesday, October 15, 2008

32

Architecture: The Big Picture



33 9/28/08

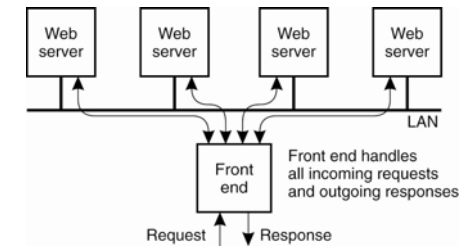
CSC 498: Web Programming



Wednesday, October 15, 2008

33

Web Server Clusters (1)



- To improve performance and availability, WWW servers are often clustered in a way that is transparent to clients
- Problem: The front end may easily get overloaded, so that special measures need to be taken.
 - Transport-layer switching: Front end simply passes the TCP request to one of the servers, taking some performance metric into account.
 - Content-aware distribution: Front end reads the content of the HTTP request and then selects the best server.

34 9/28/08

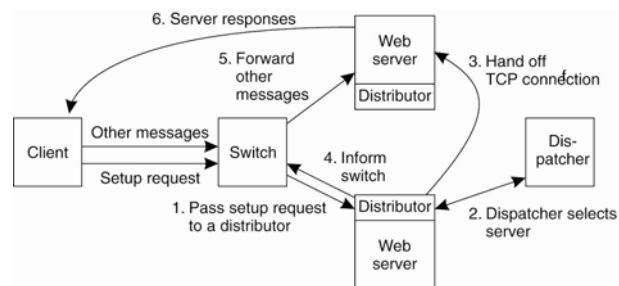
CSC 498: Web Programming



Wednesday, October 15, 2008

34

Web Server Clusters (2)



35 9/28/08

CSC 498: Web Programming



Wednesday, October 15, 2008

35

Naming (1)

Scheme	Host name	Pathname
http	:// www.cs.vu.nl	/home/steen/mbox

(a)

Scheme	Host name	Port	Pathname
http	:// www.cs.vu.nl	: 80	/home/steen/mbox

(b)

Scheme	Host name	Port	Pathname
http	:// 130.37.24.11	: 80	/home/steen/mbox

(c)

- Often-used structures for URLs. (a) Using only a DNS name. (b) Combining a DNS name with a port number. (c) Combining an IP address with a port number.

36 9/28/08

CSC 498: Web Programming



Wednesday, October 15, 2008

36

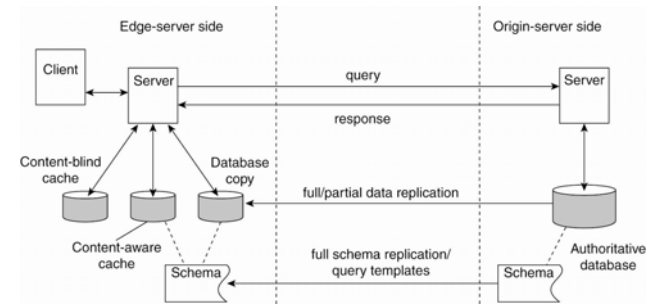
Naming (2)

Name	Used for	Example
http	HTTP	http://www.cs.vu.nl:80/globe
mailto	E-mail	mailto:steen@cs.vu.nl
ftp	FTP	ftp://ftp.cs.vu.nl/pub/minix/README
file	Local file	file:/edu/book/work/chp/11/11
data	Inline data	data:text/plain;charset=iso-8859-7,%e1%e2%e3
telnet	Remote login	telnet://flits.cs.vu.nl
tel	Telephone	tel:+31201234567
modem	Modem	modem:+31201234567;type=v32

- Examples of URIs.



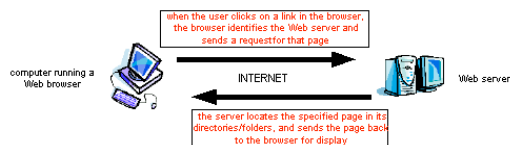
Replication of Web Applications



- Alternatives for caching and replication with Web applications.



World Wide Web



- The Web is the world's largest client/server system
 - communication occurs via message passing
 - within browser, select URL of desired page
 - browser requests page from server
 - server responds with message containing
 - type of page (HTML, gif, pdf, zip, ...)
 - page contents
 - browser uses type info to correctly display page
 - if page contains other items (images, applets, ...),
 - browser must request each separately



HTTP

- Hypertext Transfer Protocol (HTTP):
 - application-level protocol for distributed, collaborative, hypermedia information systems
 - generic, stateless, object-oriented
 - can be used for many tasks, such as name servers & distributed object management systems
 - underlying language of the Web
- HTTP/1.0 allowed only connectionless message passing
 - each request/response required a new connection
 - to download a page with images required multiple connections can overload the server, require lots of overhead
- HTTP/1.1 provides persistent connection by default
 - once client & server connect, remains open until told to close it (*or timeout*) reduces number of connections, saves overhead
 - client can send multiple requests without waiting for responses e.g., can request all images in a page at once



HTTP Methods

Operation	Description
Head	Request to return the header of a document
Get	Request to return a document to the client
Put	Request to store a document
Post	Provide data that are to be added to a document (collection)
Delete	Request to delete a document

- Operations supported by HTTP.



GET request

- most URL's (Uniform Resource Locator) have the form:

```
protocol://serverName URI
e.g., http://www.csc.liv.ac.uk/~avf/
index.html
```

- To retrieve a document via HTTP from the server, issue a GET request using the HTTP protocol

```
GET URI HTTP/1.1
Host: serverName
```

- Web server only knows the contents of the GET request message
 - automatically generated by browser when you select a URL
 - could also come from a link checker, a search engine robot, ...
 - can come directly from a telnet connection using port 80



GET example

- `bash-3.1$ telnet www.csc.liv.ac.uk 80`
- Trying 10.128.0.3...
- Connected to www.csc.liv.ac.uk (10.128.0.3).
- Escape character is '^]'.
GET /~martin/index.html HTTP/1.1
- Host: www.csc.liv.ac.uk

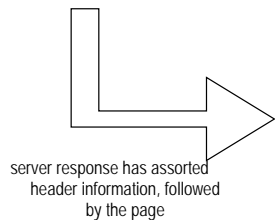
```
HTTP/1.1 200 OK
Date: Mon, 08 Oct 2007 10:01:15 GMT
Server: Apache/2.0.58 HP-UX_Apache-based_Web_Server (Unix)
 mod_perl/1.99_16 Perl/v5.8.7 DAV/2 PHP/5.0.4
Last-Modified: Mon, 01 Oct 2007 14:55:16 GMT
ETag: "ec3f-1122-9fd83d00"
Accept-Ranges: bytes
Content-Length: 4386
Content-Type: text/html

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en"
lang="en">

<head>
...
</head>

<body>
...
</body>
</html>

Connection closed by foreign host.
```



Response Header Fields

- the first line of the server's response contains a status code

- | | |
|-----------------------------|--|
| – 200 OK | request was processed successfully |
| – 301 Moved permanently | document has been moved |
| – 304 Not modified | if cached version is up-to-date |
| – 400 Bad request | syntax error in client's request |
| – 403 Forbidden | client is not allowed access (e.g., protected) |
| – 404 Not found | file could not be found |
| – 500 Internal server error | server failed |
| – 503 Service unavailable | server is overloaded |



Other response header fields

- in addition to the status code, the server's response may include
 - Date response time (in GMT)
 - Server identification info on the server
 - Last-modified time document was last changed (in GMT)
 - Content-length size of document, in bytes
 - Content-type file format (e.g., html, gif, pdf)
 - Expires prevents browser from caching beyond date

45 9/28/08

CSC 498: Web Programming



Wednesday, October 15, 2008

45

File not found

- `bash-3.1$ telnet www.csc.liv.ac.uk 80`
- Trying 10.128.0.3...
- Connected to www.csc.liv.ac.uk (10.128.0.3).
- Escape character is '^J'.
- GET /~/martin/foo.html HTTP/1.1
- Host: www.csc.liv.ac.uk

if file not found, response includes 404 status code and generic error page

```
HTTP/1.1 404 Not Found
Date: Mon, 08 Oct 2007 10:10:35 GMT
Server: Apache/2.0.58 HP-UX_Apache-based_Web_Server
      (Unix) mod_perl/1.99_16 Perl/v5.8.7 DAV/2 PHP/5.0.4
Accept-Ranges: bytes
Transfer-Encoding: chunked
Content-Type: text/html

a0
<html>
<head>
<title>404 Not Found</title>
<link rel="stylesheet" href="~/fonts.css" type="text/
css">
</head>
<body>
. . .
<td id="main_content">
  <div id="topmargin"><a name="skipnav"
  tabindex="1"></a></div>
  <!-- Beginning of main page content -->
. . .
  <h1>Page Not Found</h1>
  <p class="text">The requested URL
  was not found on this server.</p>
. . .
</html>
Connection closed by foreign host.
```

46 9/28/08

CSC 498: Web Programming



Wednesday, October 15, 2008

46

Directories as URI's

- Trying 10.128.0.3...
- Connected to www.csc.liv.ac.uk (10.128.0.3).
- Escape character is '^J'.
- GET /~/martin/ HTTP/1.1
- Host: www.csc.liv.ac.uk

if a directory is specified, will look for a file named index.html

```
HTTP/1.1 200 OK
Date: Mon, 08 Oct 2007 10:18:00 GMT
Server: Apache/2.0.58 HP-UX_Apache-based_Web_Server (Unix)
      mod_perl/1.99_16 Perl/v5.8.7 DAV/2 PHP/5.0.4
Last-Modified: Mon, 01 Oct 2007 14:55:16 GMT
ETag: "ec3f-1122-9fd83d00"
Accept-Ranges: bytes
Content-Length: 4386
Content-Type: text/html

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml11/DTD/xhtml11-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en"
lang="en">
<head>
<title> Russell Martin's Home Page </title>
. . .
</head>
<body>
. . .
</body>
</html>
Connection closed by foreign host.
```

47 9/28/08

CSC 498: Web Programming



Wednesday, October 15, 2008

47

Redirection

- `bash-2.05b$ telnet www.csc.liv.ac.uk 80`
- Trying 10.0.64.233...
- Connected to www.csc.liv.ac.uk.
- Escape character is '^J'.
- GET /~/avf/ HTTP/1.1

since URI is missing / at end, browser must do 2 requests

```
HTTP/1.1 301 Moved Permanently
Date: Tue, 06 Sep 2005 13:49:15 GMT
Server: HP Apache-based Web Server/1.3.27 (Unix)
      mod_perl/1.27 PHP/4.2.2
Location: http://www.csc.liv.ac.uk/~avf/
Transfer-Encoding: chunked
Content-Type: text/html; charset=iso-8859-1

148
<!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML 2.0//EN">
<HTML><HEAD>
<TITLE>301 Moved Permanently</TITLE>
</HEAD><BODY>
<H1>Moved Permanently</H1>
The document has moved <A HREF="http://
www.csc.liv.ac.uk/~avf/">here</A>.<P>
<HR>
<ADDRESS>HP Apache-based Web Server/1.3.27 Server at
www.csc.liv.ac.uk Port 80</ADDRESS>
</BODY></HTML>

0
Connection closed by foreign host.
```

48 9/28/08

CSC 498: Web Programming



Wednesday, October 15, 2008

48

Request Header Fields

- the client can specify additional information in the request
 - User-Agent specifies the browser version
 - Referer tells server where the user came from
useful for logging and customer tracking
 - From contains email address of user
generally not used for privacy reasons
 - Authorization can send username & password
used with documents that require authorization
 - If-Modified-Since only send document if newer than specified date

49 9/28/08

CSC 498: Web Programming

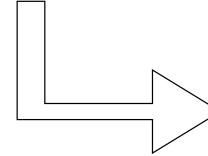


Wednesday, October 15, 2008

49

Conditional GET

- bash-2.05b\$ telnet www.csc.liv.ac.uk 80
- Trying 10.0.64.233...
- Connected to www.csc.liv.ac.uk.
- Escape character is '^J'.
- GET /~avf/ HTTP/1.1
- Host: www.csc.liv.ac.uk



```
HTTP/1.1 304 Not Modified
Date: Tue, 06 Sep 2005 14:08:58 GMT
Server: HP Apache-based Web Server/1.3.27
      (Unix) mod_perl/1.27 PHP/4.2.2
ETag: "10b3e-1000-431452ef"

Connection closed by foreign host.
```

since the document has not been modified since the specified date, the page is not sent by the server (status code 304)

50 9/28/08

CSC 498: Web Programming



Wednesday, October 15, 2008

50

Other Request Methods

- HEAD similar to GET, but requests header information only useful for checking to see if a document exists, how recent
- POST similar to GET, but encodes inputs differently useful for submitting form contents to a CGI program, and is also often used in PHP scripts for submitting information
- PUT upload a document to the server
- DELETE delete a document from the server

51 9/28/08

CSC 498: Web Programming

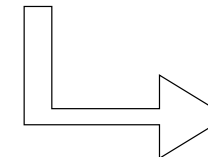


Wednesday, October 15, 2008

51

HEAD example

- bash-2.05b\$ telnet www.csc.liv.ac.uk 80
- Trying 10.0.64.233...
- Connected to www.csc.liv.ac.uk.
- Escape character is '^J'.
- HEAD /~avf/index.html HTTP/1.1
- Host: www.csc.liv.ac.uk



server does not send the page, only the header information

```
HTTP/1.1 200 OK
Date: Tue, 06 Sep 2005 14:23:24 GMT
Server: HP Apache-based Web Server/1.3.27 (Unix)
      mod_perl/1.27 PHP/4.2.2
Last-Modified: Tue, 30 Aug 2005 12:37:03 GMT
ETag: "10b3e-1000-431452ef"
Accept-Ranges: bytes
Content-Length: 4096
Content-Type: text/html

Connection closed by foreign host.
```

52 9/28/08

CSC 498: Web Programming



Wednesday, October 15, 2008

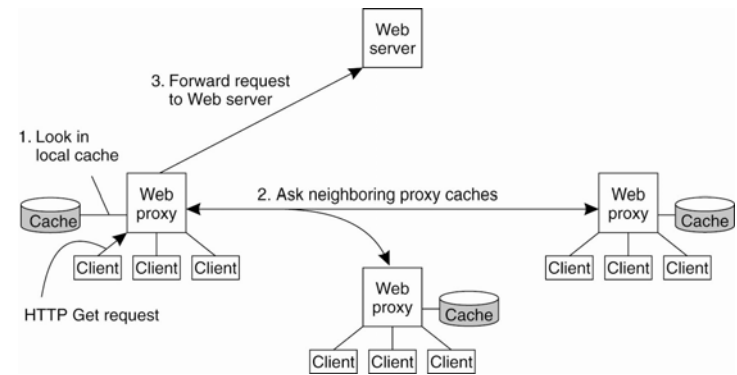
52

Caching

- Browsers cache pages to save downloading
 - maintain temporary storage (cache) for recent pages
 - when a page is requested, check to see if already in cache
 - if not in the cache, issue GET request
 - when response message arrives,
 - display page and store in cache (along with header info)
 - if already stored in the cache, send GET request with If-Modified-Since header set to the data of the cached page
 - when response message arrives,
 - if status code 200, then display and store in cache
 - if status code 304, then display cached version instead



Web Caching



- The principle of cooperative caching.



Cookies

HTTP message passing is transaction-based, stateless

- many e-commerce apps require persistent memory of customer interactions
 - e.g., *amazon.com*
 - remembers your name, credit card, past purchases, interests*
- Netscape's solution: cookies
 - a cookie is a collection of information about the user
 - server can download a cookie to the client's machine using the "Set-cookie" header in a response
 - Set-cookie: CUSTOMER=Dave_Reed; PATH=/; EXPIRES=Thursday, 29-Jan-04 12:00:00
 - when user returns to URL on the specified path, the browser returns the cookie data as part of its request
 - Cookie: CUSTOMER=Joe_Doe

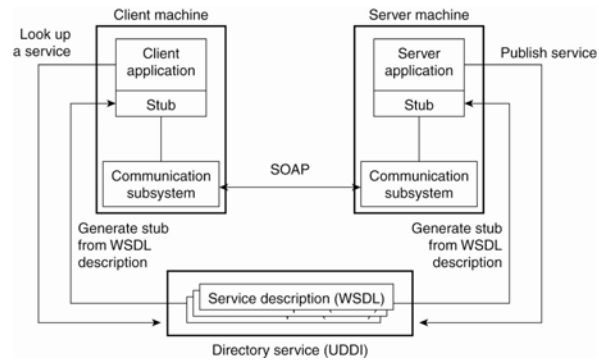


Web Services

- What are Web services?
 - They are a distributed computing architecture.
- Who is using Web services now?
 - Industry technologies
- Which approach should we use - .NET or J2EE?
- Requestor, Registry, Provider



Web Services Fundamentals



- The Web is more than just user-site interaction:
 - Sites could offer services to other sites
 - standardization is badly needed.

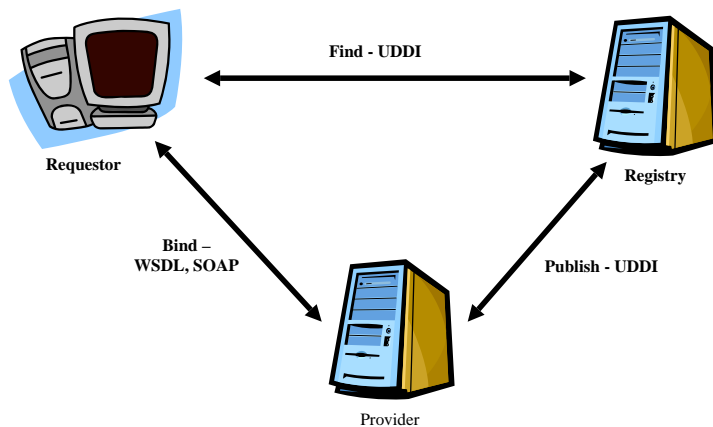


Critical Elements of a Basic Web Services Architecture

Format	XML (Format)	For presenting data and information	
Services	UDDI (Publish) A directory service	WSDL (Find) A protocol for applications to find a service	SOAP (Bind) A protocol that enables applications to agree the communication
Network	The Internet The Internet, using TCP/IP protocols...		



Service-Oriented Architecture



Developing Web services using J2EE & .NET

- Reference: [TheServerSide](#)

