COMP9321 Web Application Engineering 12s1

Helen Hye-young Paik

Service Oriented Computing Group, CSE, UNSW

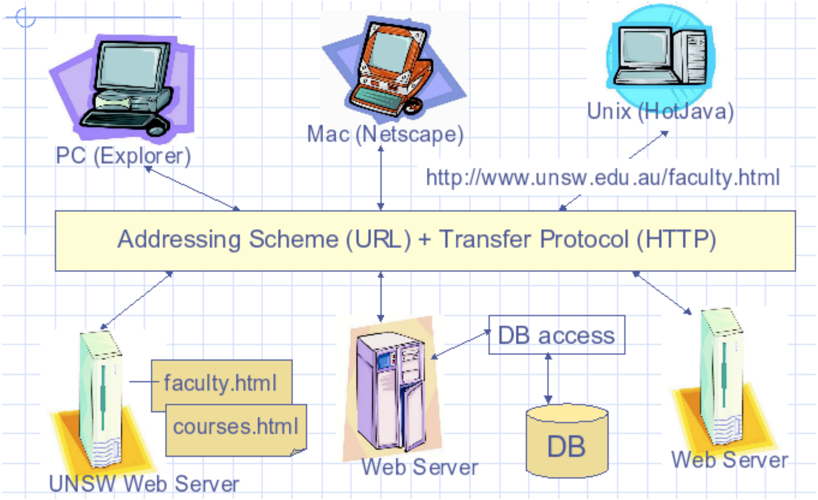
Week 1

Topic: Web Essentials - Clients, Servers and Communication

References used for the Lecture:

- Web Application Architecture - Shklar and Rosen

Basic Web Architecture: Universal Readership ...

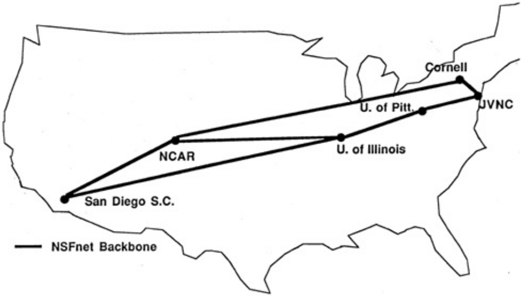


Web Essentials - Clients, Servers and Communication

Essential elements?

- Web browsers to surf the Web
- The server systems to supply information to the browsers
- *The computer networks supporting browser-server communication*

the start of the Internet backbone:-



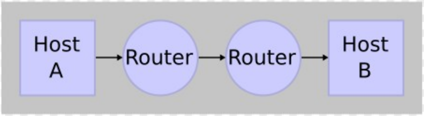
NSFnet Backbone Network

National Center For Atmospheric Research
March 19, 1986

Web Essentials - Clients, Servers and Communication

- 1990 - NSFNET (National Science Foundation - USA) - collection of computer networks connected via the public backbone and communicating across networks using TCP/IP.
- But the NSFNET terms of usage required purely commercial traffic was not allowed over the backbone. The purpose of the Internet was still to support research and education
- Restriction on commercial traffic was revoked in 1991 - laying the foundation of the Internet we see today
- NSF left the role of the Internet backbone operator, and private telco firms took over in 1995
- These firms have direct access to the backbone, and paid by Internet Service Providers (ISPs), who in turn are paid by their subscribers.

Web Essentials - Basic Internet Protocols: TCP/IP



IP (Internet Protocol)

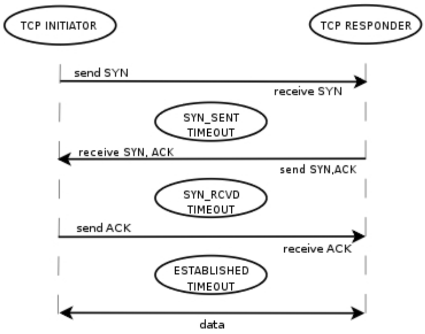
- Each device on the Internet is associated with an IP address (e.g., 192.0.34.16)
- When an application on a computer (*source*) wants to send data to another computer (*destination*), IP software creates a packet (header+data) and send it via the network.
- If destination is on another network, the IP software sends the packet to a *gateway* which selects a computer on one of the other networks and sends the packet to that computer - this is repeated until the packet reaches its destination
- IP software on the destination will receive the packet and pass its data up to an application that is waiting for the data

Web Essentials - Basic Internet Protocols: TCP/IP

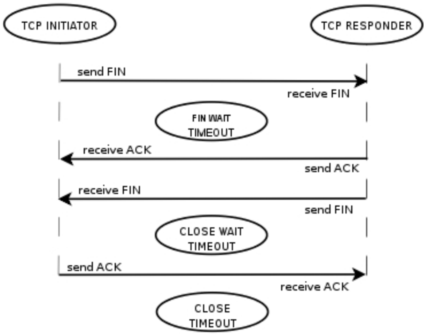
TCP (Transmission Control Protocol)

- TCP works with IP to add 'reliability' as IP-only communication can lose packets
- TCP adds the concept of 'connection' between a sender and a receiver
- a connection must be established before a packet can be sent

TCP connection establishment



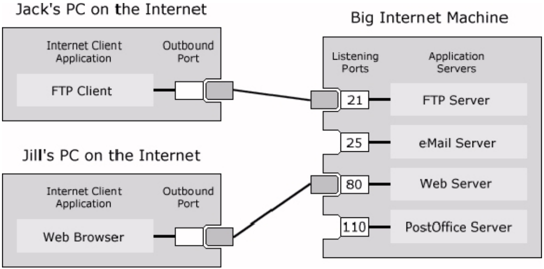
TCP connection termination



Web Essentials - Basic Internet Protocols: TCP/IP

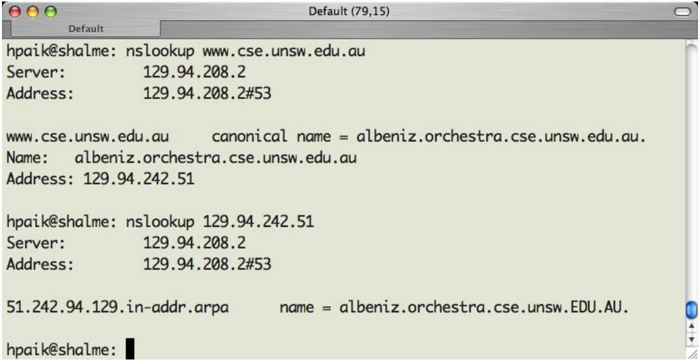
TCP (Transmission Control Protocol)

- An important feature that TCP adds to IP is the concept of 'port'
- Having ports allows TCP to be used to communicate with many different applications (e.g., mail server, file server, web server, remote login server. etc.)
- port numbers 0-1023 are reserved for a user with administration permissions, 1024-65535 can generally be used by the first application on a system that requests the port



Web Essentials - Basic Internet Protocols: Domain Names

Domain Name Service (DNS) : DNS provides a way to map back and forth IP addresses and host names (e.g., 129.94.242.51 - www.cse.unsw.edu.au)

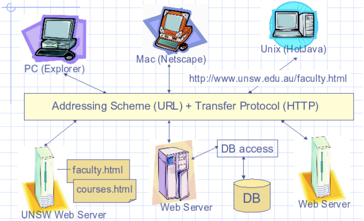


```
hpaik@shalme: nslookup www.cse.unsw.edu.au
Server:         129.94.208.2
Address:        129.94.208.2#53

www.cse.unsw.edu.au     canonical name = albeniz.orchestra.cse.unsw.edu.au.
Name:   albeniz.orchestra.cse.unsw.edu.au
Address: 129.94.242.51

hpaik@shalme: nslookup 129.94.242.51
Server:         129.94.208.2
Address:        129.94.208.2#53

51.242.94.129.in-addr.arpa      name = albeniz.orchestra.cse.unsw.EDU.AU.

hpaik@shalme: 
```

The first two lines - DNS detail

Web Essentials - Higher-Level Protocols

- TCP/IP and DNS enable the computers to communicate back and forth. The question is "what are you going to with it"?
- A variety of high level protocols are used to communicate once a TCP connection is established: SMTP, FTP and Telnet are good examples of widely used higher-level protocols.
- The most prevalent high-level protocol has to be HTTP (Hypertext Transport Protocol). The primary TCP-based protocol used for communication between Web servers and Web browsers.



Building Blocks of the Web (Shklar and Rosen, pg. 31)

Three basic components devised by Tim Berners-Lee:

- A Uniform Notation Scheme for addressing resources (URL)
- A protocol for transporting messages (HTTP)
- A markup language for formatting hypertext documents (HTML)

Web Essentials - URL

General notation associated with a URL

```
scheme://host[:port]/path/.../[;url-params][?query-string]
[#anchor]
```

- scheme - underlying protocol to be used
- host[:port] - hostname or IP address of the web server. For HTTP, the default port is 80.
- path - Path to the resource from the root directory of the web server.
- url-params - Optional name=value pairs - used mainly for JSESSIONID
- query-string - string of name=value pairs separated by ampersand (&) or semi-colons
- anchor - reference to a position marker in the requested document

An http-scheme URL

`http://www.example.org:56789/a/b/c.txt?t=win&s=chess#para5`

authority part:

- after 'http://' through to the next slash - www.example.org:56789
- It consists of either a domain name or an IP address
- optionally followed by a port number (if omitted, port 80 is implied)

path part:

- after the authority through to question mark (?)
- /a/b/c.txt (/ is part of the path, ? is not)
- much like a file path in file system ...

query string part:

- after the path up to a number sign #
- contains a set of name-value pairs, separated by & (e.g., t=win&s=chess)

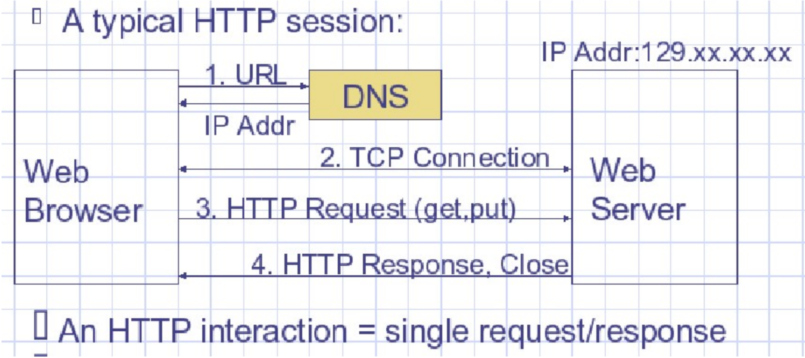
fragment identifier part:

- after the number sign #, not including #

Web Essentials - HTTP

- The basic structure of HTTP communication follows a "*request-response*" model.
- An HTTP interaction is initiated by a client sending a request message to the server; the server is then expected to generate a response message.
- The format of request and response messages is dictated by HTTP.
- HTTP is stateless - each request-response pair is a separate interaction
- There's no way to batch related requests

Web Essentials - HTTP



- A typical HTTP session:

- An HTTP interaction = single request/response

Web Essentials - HTTP

HTTP Request (from browser to server):

It is composed of `Request Line + Header + (additional data)`

Syntax for the Request Line:

```
Request-Method sp Request-URI sp HTTP-version CRLF
```

eg, GET `http://www.smh.com.au/index.html` HTTP/1.1

- There must be a newline (CRLF) between the header and the additional data part.
- Common Request methods: GET, POST, HEAD ...
- Request header: `User-Agent`, `Referer`, `Authorization`.
- Additional data (body): parameters (POST), block of data

On Request Header Fields

```
POST /servlet/EchoHttpRequest HTTP/1.1
host: www.example.org:8080
user-agent: Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.4)
   Gecko/20030624
accept: text/xml, application/xml, application/xhtml+xml,
   image/png, image/jpeg, image/gif;q=0.2, */*;q=0.1
accept-language: en-us, en:q=0.5
accept-encoding: gzip, deflate
accept-character: ISO-8859-1, utf-8;q=0.7
connection: keep-alive
keep-alive: 300
content-type: application/x-www-form-urlencoded
content-length: 13

request body starts ...
```

HTTP Request Methods (version 1.1)

Method	Description
GET	It the simplest, most used. It simply retrieves the data identified by the URL. If the URL refers to a script (CGI, servlet, and so on), it returns the data produced by the script.
HEAD	It only returns HTTP headers without the document body.
POST	It is like GET. Typically, POST is used in HTML forms. POST is used to transfer a block of data to the server.
OPTIONS	It is used to query a server about the capabilities it provides. Queries can be general or specific to a particular resource.
PUT	It stores the body at the location specified by the URI. It is similar to the PUT function in FTP.
DELETE	It is used to delete a document from the server. The document to be deleted is indicated in the URI section of the request.
TRACE	It is used to tract the path of a request through firewall and multiple proxy servers. TRACE is useful for debugging complex network problems and is similar to the traceroute tool.

HTTP/1.1 Status Code Classes and common codes[1]

Digit	Class	Standard Use
1	Informational	Provides information to client before request processing has been completed
2	Success	Request has been successfully processed
3	Redirection	Client needs to use a different resource to fulfill request
4	Client Error	Client's request is not valid
5	Server Error	An error occurred during server processing

Code	Reason Phrase	Usual Meaning
200	OK	Request processed normally
301	Moved Permanently	URI for the requested resource has changed
401	Unauthorised	The resource is password protected, and the user has not yet supplied a valid password
403	Forbidden	The resource is present on the server, but is read protected
404	Not Found	No resource corresponding the URI was found
500	Internal Server Error	Server software detected an internal failure

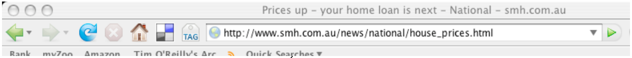
[1] p.19 in (WebTech)

Web Essentials - HTTP

HTTP Response (from server to browser):

- Composed of `Status Line + Header + Body`
- Status line: 200 OK, 404 Not Found, etc.
- Header:
    - Content-Type, Content-Language, Content-Length, Cache-control, etc.
- Body:
    - Body contains the requested data
    - Body is in specific MIME format (eg., text/HTML)
    - MIME (Multipurpose Internet Mail Extension): text (plain, HTML), multimedia data, applications such as PDF, PowerPoint, etc.

reproduced from: page 18, Head First Servlets & JSP, Basham, et. al., O'Reilly,



Prices up – your home loan is next – National – smh.com.au

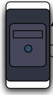
http://www.smh.com.au/news/national/house_prices.html

The user types a URL

The browser creates an HTTP GET request

```
GET /news/national/house_prices.html
Host: www.smh.com.au
User-Agent Mozilla/5.0 (Macintosh)
...
```

the HTTP GET is sent to the server



the server finds the page ...

house_prices.html

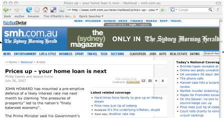
```
<html><body>
<h1 align=center>Sydney Morning Herald</h1>
<p>House prices are up. Following yesterday's
inflation figures ... </p>
```

And generates an HTTP response

```
HTTP/1.1 200 OK
Set-Cookie ...

<html><body>
<h1 align=center>Sydney Morning Herald</h1>
<p>House prices are up. Following yesterday's
inflation figures ... </p>
```

the HTTP response is sent to the browser

The browser renders the HTML



Prices up – your home loan is next – National – smh.com.au

http://www.smh.com.au/news/national/prices-up--you

Google
Rank
myZoo
Amazon
Tim O'Reilly's Arc...
Quick Searches ▾
FairfaxDigital
NEWS | MYCAREER | DOMAIN | DRIVE | FINANCE | MOBILE | RSVP | TRAVEL
member centre | login | register
smh.com.au
The Sydney Morning Herald
the (sydney) magazine
ONLY IN The Sydney Morning Herald
NEWS ENTERTAINMENT LIFE & STYLE BUSINESS SPORT TRAVEL TECH SECTIONS CLASSIFIEDS JOBS AUTO REAL ESTATE DATING
Search her
Home » National » Article

Prices up - your home loan is next

Phillip Coorey and Jessica Irvine
July 26, 2007

JOHN HOWARD has mounted a pre-emptive defence of a likely interest rate rise next month by claiming "the pressures of prosperity" led to the nation's "finely balanced economy".

The Prime Minister said his Government's

Latest related coverage

- Hard times force family to give up on lifelong dream
- Price rises just tip of iceberg
- Analysis It's the underlying inflation, stupid
- Your say: Another rate rise

Today's National Coverag

- Brimble tapes revealed at
- Online sex pests unwatch
- UK considers 56 days' det
- The phone calls
- Haneef case hits a surpris review
- Norfolk murder sickening,
- Hopes for Fromelles excav
- On the blower: no one to council keeps you up
- Price rises just tip of icebe
- Court tells charity to reins unjust sackings

The user reads the content

Improving Performance using HTTP features

HTTP 1.1 includes several methods for optimising delivery of Web content.

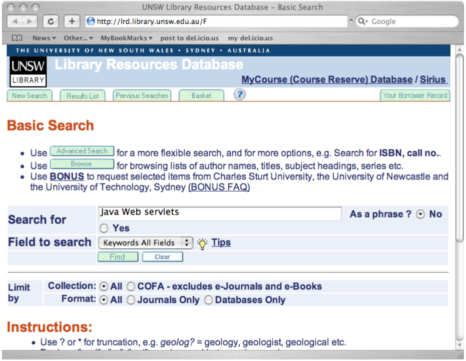
Includes information about the following:

- Connection management
- Caching Support
- Content negotiation

Using these features, you can tune your application to use caching effectively and tune your server for scalability and performance.

Moving Web Site to Web Applications

Typical HTML Interactions


UNSW Library Resources Database – Basic Search
http://lrd.library.unsw.edu.au/F
Google
News
Other...
MyBookMarks
post to del.icio.us
my del.icio.us
THE UNIVERSITY OF NEW SOUTH WALES • SYDNEY • AUSTRALIA
UNSW
LIBRARY
Library Resources Database
MyCourse (Course Reserve) Database / Sirius
New Search
Results List
Previous Searches
Basket
?
Your Borrower Record

Basic Search

- Use Advanced Search for a more flexible search, and for more options, e.g. Search for **ISBN, call no.**.
- Use Browse for browsing lists of author names, titles, subject headings, series etc.
- Use **BONUS** to request selected items from Charles Sturt University, the University of Newcastle and the University of Technology, Sydney (BONUS FAQ)

Search for Java Web servlets **As a phrase ?** ● **No** ○ **Yes**

Field to search Keywords All Fields Tips

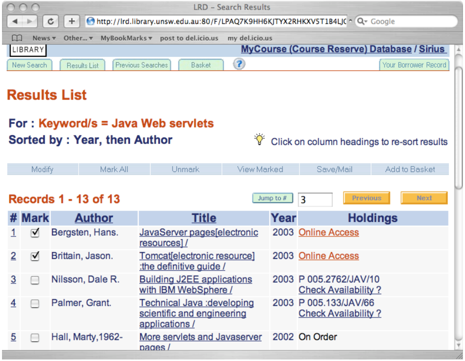
Find Clear

Limit by **Collection:** ● **All** ○ **COFA - excludes e-Journals and e-Books**

Format: ● **All** ○ **Journals Only** ○ **Databases Only**

Instructions:

- Use ? or * for truncation, e.g. *geolog?* = geology, geologist, geological etc.



Results List

For : Keyword/s = Java Web servlets

Sorted by : Year, then Author

Click on column headings to re-sort results

Modify | Mark All | Unmark | View Marked | Save/Mail | Add to Basket

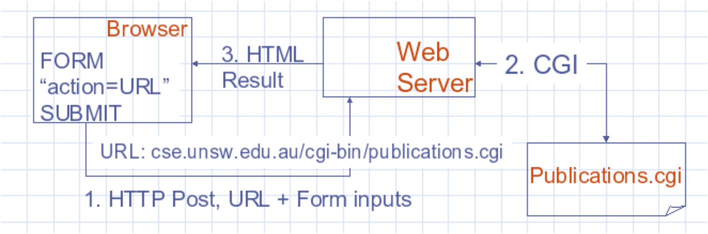
Records 1 - 13 of 13 Jump to # 3 Previous Next

#	Mark	Author	Title	Year	Holdings
1	☑	Bergsten, Hans.	JavaServer pages[electronic resources] /	2003	Online Access
2	☑	Brittain, Jason.	Tomcat[electronic resource] :the definitive guide /	2003	Online Access
3	□	Nilsson, Dale R.	Building J2EE applications with IBM WebSphere /	2003	P 005.2762/JAV/10 Check Availability ?
4	□	Palmer, Grant.	Technical Java :developing scientific and engineering applications /	2003	P 005.133/JAV/66 Check Availability ?
5	□	Hall, Marty,1962-	More servlets and Javaserver pages /	2002	On Order

- The responses are generated dynamically depending on your input.
- The response also include 'hooks' for further interactions

Common Gateway Interface (CGI)

- A Web server accepts certain kinds of URLs as a request for the execution of a specific program
- Basically, the CGI technology allows a browser to initiate a request to run a program. In turn, the Web server:
  - Identifies the program (a Web application) to be run,
  - Pass the user input to the program,
  - Executes the program
  - Pass the generated output (an HTML document) to the browser


Browser
FORM
“action=URL”
SUBMIT
3. HTML
Result
Web
Server
2. CGI
URL: cse.unsw.edu.au/cgi-bin/publications.cgi
Publications.cgi
1. HTTP Post, URL + Form inputs

CGI is replaced by other 'better' technologies

CGI was not a scalable solution for big applications (eg., e-commerce).

- each client request makes the Web server spawn a new process of the requested CGI program
- creating a new process is expensive
  - consumes a lot of CPU cycles, memory, etc.

Gradually, new and better technologies emerged:

- **PHP:**. An open-source technology. It comes with rich built-in functionality, such as file upload. It is becoming more and more popular.
- **Servlet and Java Server Pages (JSP):**. Introduced by Sun Microsystems in 1996. Java's answer to a better CGI technology.
- **Active Server Pages (ASP) and ASP.NET:**. MS's answer to a better CGI technology. It is part of Microsoft .NET initiative.

Collectively these (and a lot more others) are called 'Web application frameworks'.

n-Tier applications

Tier - A layer in your application through which the client request has to pass to be resolved. E.g. A 3-tier application consists of the interface (JSP) , business logic (servlets) and database (RDBMS). Applications can have many tiers (*n*-tier).

How many tiers should your application contain ?

Benefits

- Applications are modular and can re-use components - reduces development time and minimises errors (if designed right)
- Distributed processing - Components can run on multiple machines
- Overloaded components can be isolated for replication

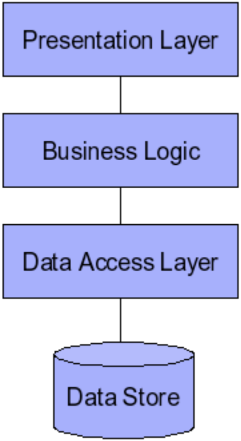
Disadvantages

- Message passing among components brings overheads
- Increased complexity of development
- Synchronization of distributed components

Different Layers in an Application

Different solutions for each layer

- Presentation Tier
  - Exposes Functionality to the Users, Communicates their commands to the lower tiers
- Business logic Tier
  - Processes user input, Performs logical evaluations & calculations, and Decides which view to expose to users
- Data Access Tier
  - Interfaces with the database, stores, updates and maintains persistence within the application



Web Essential - Web servers

Server features[2]:

- The server calls on TCP software and waits for connection requests to one or more ports
- When a connection request is received, the server dedicates a "subtask" to handling this connection
- The subtask establishes the TCP connection and receives an HTTP request
- The subtask examines the Host header field of the request to determine which "virtual host" should receive this request and invokes software for this host.
- The virtual host software maps the Request-URI field of the HTTP request start line to a resource on the server

[2] page 30 in (WebTech)