

CSIS 4222

Ch 4: The Web and FTP

Application-Layer Protocols

- Defines the syntax and semantics of messages exchanged by network applications
- Specify two aspects of interaction

Aspect	Description
Data Representation	Syntax of data items that are exchanged, specific form used during transfer, translation of integers, characters, and files between computers
Data Transfer	Interaction between client and server, message syntax and semantics, valid and invalid exchange error handling, termination of interaction

Web Protocols

The World Wide Web is one of the most widely used services in the Internet

Standard	Purpose
HyperText Markup Language (HTML)	A representation standard used to specify the contents and layout of a web page
Uniform Resource Locator (URL)	A representation standard that specifies the format and meaning of web page identifiers
HyperText Transfer Protocol (HTTP)	A transfer protocol that specifies how a browser interacts with a web server to transfer data

Figure 4.2 Three key standards that the World Wide Web service uses.

Hypermedia/Hypertext

- *Hypermedia* system allows interactive access to a collections of documents
- Document can hold:
 - Text (*hypertext*)
 - Graphics
 - Sound
 - Animations
 - Video
- Documents are linked together
 - Distributed: documents stored on remote servers

Document Representation with HTML

- *HyperText Markup Language* (HTML) is the standard that specifies the syntax of web pages
- Pages are formatted with a *mark up language* (tags)
 - Allows browser to reformat to fit display
 - Different browsers could display the same document differently (e.g. PC vs. iPhone)

HTML Tags

Input

```
Hello<BR /><H1>This is a
Heading</H1><BR />Back to normal
```

Output

```
Hello
This is a Heading
Back to normal
```

Uniform Resource Locator (URL) to Identifying a Page

Syntax for URL:

```
protocol://computer_name:port/document_name%parameters
```

- Protocol to access page: can be http (default), ftp, file, mailto, etc.
- *computer_name* is DNS name of server where page is located
- *port* is TCP port to access page
- *document_name* is the file path on server to the page (resource)
- *parameters* for the page

Some of these components are optional or have default values

Links in Web Pages

- A document can contains *links* (pointers) to other documents
- Links are specified in HTML
- Links are *passive* - no action taken until selected
- HTML tags: `<a>` and ``
 - Linked document specified by parameter: `href="document URL"`
 - Content between HTML tags selects the link

```
<a href="http://www.stockton.edu/~olanm">My home page</a>
```

My home page

WWW Uses Client-Server Paradigm

Browser is the client:

- Application program
- User's interface to the Web
- Client to fetch information from Web server
- Displays information for user

Web server is the server:

- Has access to a set of web documents
- Responds to request from browser by sending a copy of document

HyperText Transport Protocol (HTTP) specifies commands for client-server interaction

Browser Architecture

Browser has many components:

- Display driver for painting screen
- HTML interpreter for HTML-formatted documents
- Other interpreters (e.g. Javascript, Flash)
- HTTP client to fetch HTML docs from WWW server
- Other clients for other protocols (e.g. ftp)
- Controller to accept input from user

Must be multi-threaded

Web Server Architecture

- Stores a set of Web documents or connects to an application server which generates documents dynamically
- Waits for incoming connections (usually listens on port 80)
- Accepts commands from connection
- Writes page to connection

Connecting to a Web Server

```
telnet www.stockton.edu 80 Request
Trying 134.210.1.200...
Connected to loki.stockton.edu.
Escape character is '^]'.
GET / HTTP/1.0
```

```
HTTP/1.1 200 OK Response
Date: Sun, 21 Mar 2010 20:59:31 GMT
Server: Apache
Last-Modified: Wed, 02 Mar 2005 17:05:41 GMT
ETag: "721e-164-4225f265"
Accept-Ranges: bytes
Content-Length: 356
Connection: close
Content-Type: text/html

<HTML>
<HEAD>
<title>The Richard Stockton College of New Jersey</title>
...
</HEAD></HTML>
Connection closed by foreign host.
```

Web Document Transfer with HTTP

Once it establishes a connection a browser sends an HTTP request to the server

Request	Description
GET	Requests a document; server responds by sending status information followed by a copy of the document
HEAD	Requests status information; server responds by sending status information, but does not send a copy of the document
POST	Sends data to a server; the server appends the data to a specified item (e.g., a message is appended to a list)
PUT	Sends data to a server; the server uses the data to completely replace the specified item (i.e., overwrites the previous data)

Figure 4.5 The four major HTTP request types.

Web Document Transfer with HTTP

The first line of a response header contains a status code

Status Code	Corresponding Status String
200	OK
400	Bad Request
404	Not Found

Caching in Browsers

- Important for optimization of web access
- Much of the content at a given site consists of large images
- Browser can reduce download times significantly by saving images in a cache on disk and using the cached copy

Caching in Browsers

- What if the document on the server changes after a browser stores a copy in its cache?
- Whenever a browser obtains a document from a web server, the header specifies the last time the document was changed
- A browser saves the Last-Modified date information along with the cached copy
 - A browser makes a HEAD request to the server and compares the Last-Modified date of the server's copy to the Last-Modified date in the cached
 - If the cached version is stale, the browser downloads the new version
- Browsers do not cache small items
 - The time to download the item with a GET request is approximately the same as the time to make a HEAD request and keeping many small items in a cache can increase cache lookup times

