Web-Based Information Systems

Fall 2004

CMPUT 410: Protocols

Dr. Osmar R. Zaïane

University of Alberta

Course Content

- Introduction
- Internet and WWW
- **Protocols**
- HTML and beyond
- Animation & WWW
- CGI & HTML Forms
- Javascript
- Databases & WWW
- Dynamic Pages
- Perl & Cookies
- SGML / XML
- CORBA & SOAP
- Web Services
- Search Engines
- Recommender Syst.
- Web Mining
- Security Issues
- Selected Topics

Preliminaries

Objectives of Lecture 3

Protocols

- Get acquainted with some communication protocols relevant to the Internet and the World-Wide Web

Outline of Lecture 3

- Protocols
- TCP/IP
- File Transfer Protocol
- HTTP
- URL Syntax

Communication Essentials

- For two parties to communicate efficiently, there is a need for a set of rules to exchange information.
- A protocol is a format or set of rules for communication, either over a network or between applications.
- Two machines communicating on the Internet use defined protocols to "understand" each other.
- There is a need to standardize protocols.

Internet Governance

- Internet Society:
- Non-profit, non-governmental, membership-based, body
- Two standards bodies:
  - IETF (Internet Engineering Task Force): e.g. internet security standards.
  - IAB (Internet Architecture Board): e.g. Internet addresses.

Web Governance

- World Wide Web Consortium (W3C) created in 1993.
- Organizational Hosts: MIT (USA), Keio University (Japan), and INRIA (France)
- W3C is responsible for web-related protocols (e.g. HTTP, HTML, and XML).
- W3C organizes work groups and ratifies WWW related standards.

Outline of Lecture 3

- Protocols
- TCP/IP
- File Transfer Protocol
- HTTP
- URL Syntax

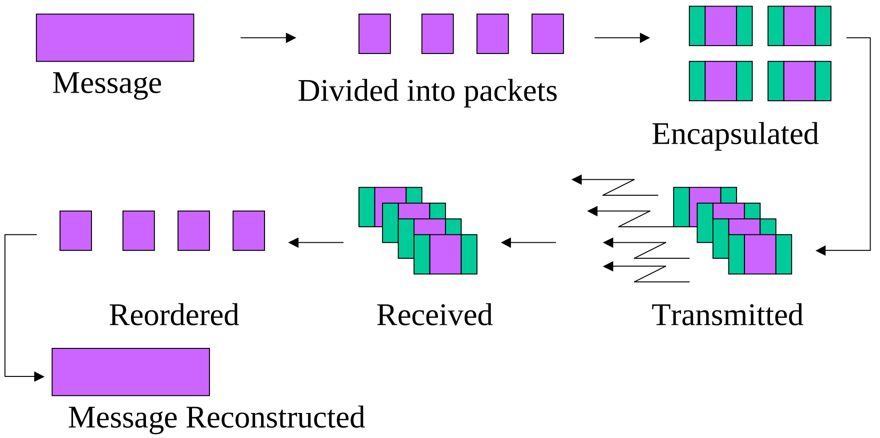
TCP/IP

- TCP/IP: Transmission Control Protocol/Internet Protocol.
- First defined by Vint Cerf and Bob Kahn in 1973, the protocol made the Internet possible and has become the default network protocol around the world.
- Every computer on the Internet supports TCP/IP.

TCP/IP

- Protocols on the Internet are layered protocols.
- Data transferred across the network with TCP/IP are transmitted in *packets*.
- TCP/IP is actually two protocols
- IP (the lowest layer) is responsible for moving the messages from one location to another. It does not guarantee the delivery.
- TCP guarantees the delivery of data by serializing the packets.

Packet Switching

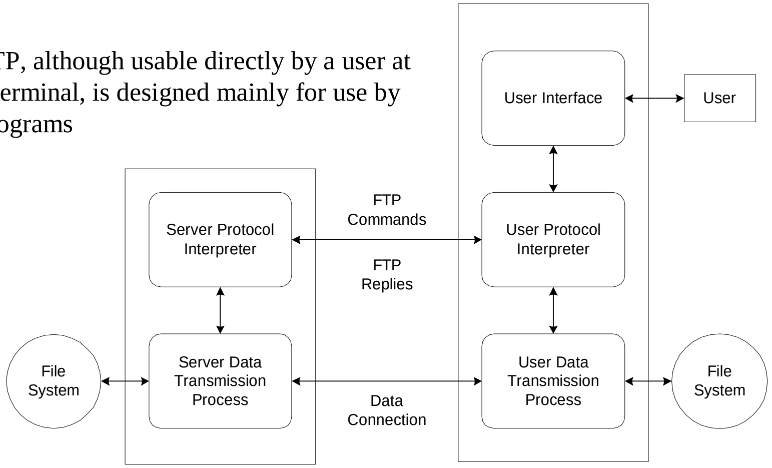


Outline of Lecture 3

- Protocols
- TCP/IP
- File Transfer Protocol
- HTTP
- URL Syntax

File Transfer Protocol (FTP)

FTP, although usable directly by a user at a terminal, is designed mainly for use by programs


User Interface
User
Server Protocol Interpreter
FTP Commands
User Protocol Interpreter
FTP Replies
File System
Server Data Transmission Process
User Data Transmission Process
File System
Data Connection

Reference RFC 959

- Control functions (commands) and reply codes are transferred over the control connection.
- All data transfer takes place over the data connection.
- The control connection must be "up" while data transfer takes place.
- The control connection uses the TELNET protocol.

Access Control Commands
USER specify user
PASS specify password
CWD change directory
CDUP change directory to parent
QUIT logout

Transfer Parameter Commands
PORT publish local data port
PASV server should listen
TYPE establish data representation
MODE establish transfer mode
STRU establish file structure

Service Commands
RETR retrieve file
STOR send file
STOU send file and save as unique
APPE send file and append
ABOR abort previous service command
PWD print working directory
LIST transfer list of files over data link

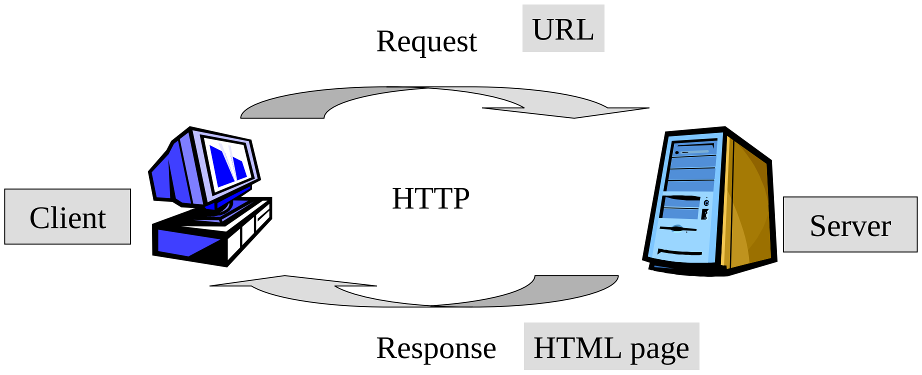
Outline of Lecture 3

- Protocols
- TCP/IP
- File Transfer Protocol
- HTTP
- URL Syntax

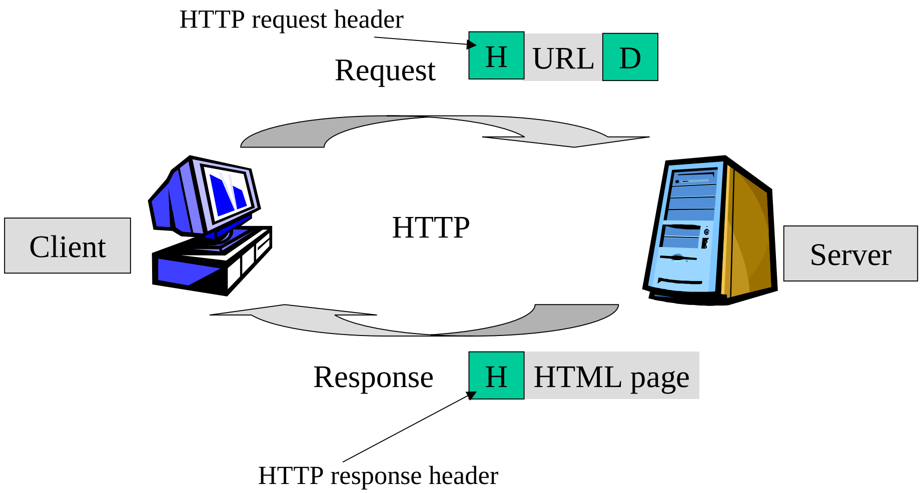
HyperText Transfer Protocol

- Is the underlying protocol of the Web.
- It runs on top of TCP/IP and determines the communication between the client (a browser) and a server (web server).

Client-Server Architecture



Client-Server Architecture



HTTP Request Header Example

```
Get /index.html HTTP/1.0
conncection: Keep-Alive
user-agent: Mozilla/3.0 [en] (sunos/4.006)
accept: image/gif, image/jpeg, image/png, */*
accept-language: en
authorization: alpha dG8tBYt6dG9t
pragma: no-cache
if-modified-since: 7 March 1996
```

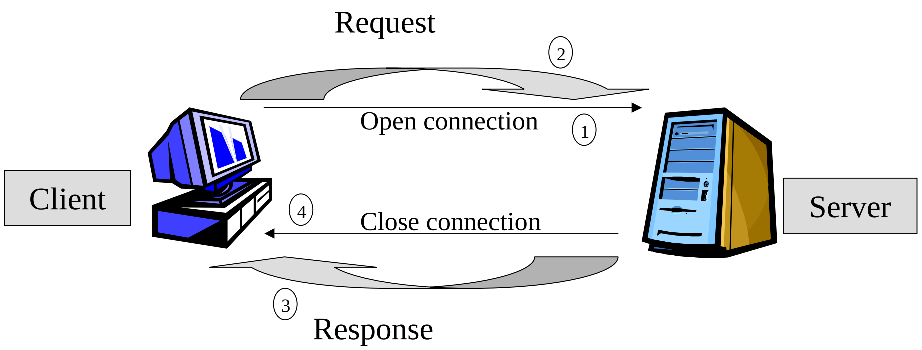
HTTP Response Header Example

```
HTTP/1.0 200 Document follows
MIME-version: 1.0
Server: CERN/3.6
Date: Tuesday, 12 March 1996  13:30 GMT
Content-Type : text/html
Content-Length : 2124
Last-Modified: Sunday, 10 March 1996  15:53 GMT

… message body …
```

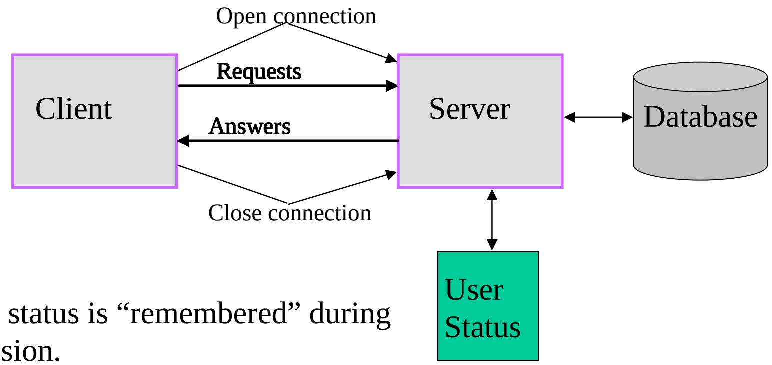
Blank line

Client-Server Architecture



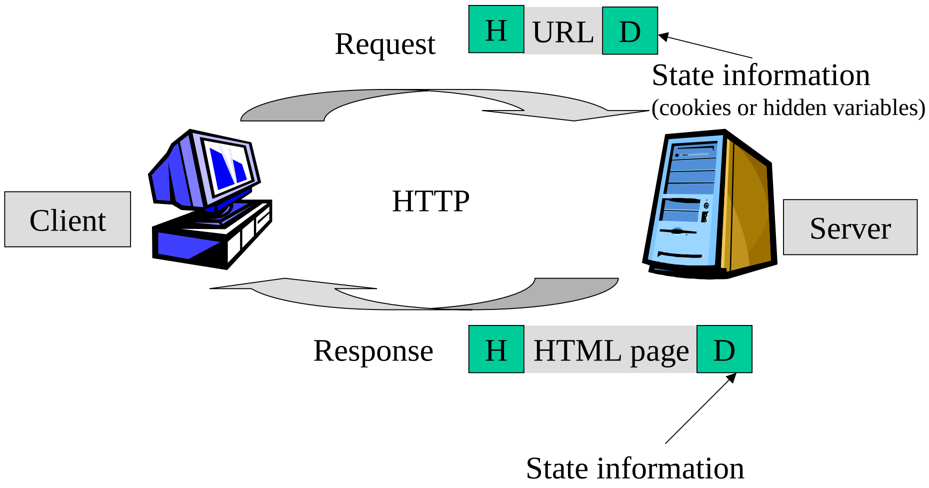
HTTP has stateless sessions

Database Client-Server Architecture



User status is “remembered” during a session.
Client opens and closes sessions.

Simulation of Status in Stateless Session



State Information for Web-Based Applications

- Statelessness of HTTP allows the server to handle a large number of clients & frees resources quickly.
- An application may need to keep persistent data.
- HTTP/1.0 added the specification for a *Keep-Alive* connection that is either closed by client or server.
- HTTP/1.1 specifies persistent connections by default that can be closed or timed-out (eliminates some overhead).

Common Status Codes

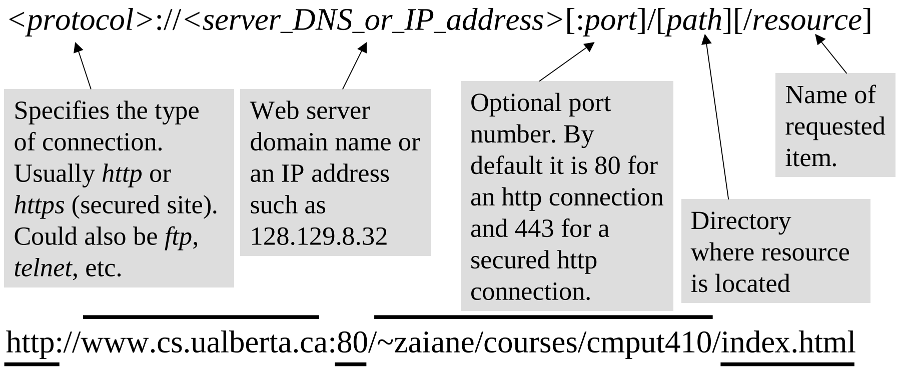
200	OK	Request successfully processed
302	Found	Client is redirected to new location
304	Not Modified	Client should use a local copy
400	Bad Request	Server could not interpret the request
401	Unauthorized	Client is not authorised to access the data
403	Forbidden	Access to the resource is forbidden
404	Not Found	Requested data was not found by server
500	Server Error	Error occurred while processing request

Outline of Lecture 3

- Protocols
- TCP/IP
- File Transfer Protocol
- HTTP
- URL Syntax

Uniform Resource Locator

A type of URI, Universal Resource Identifier



<protocol>://*<server_DNS_or_IP_address>*[:*port*]/[*path*][/*resource*]

Specifies the type of connection. Usually *http* or *https* (secured site). Could also be *ftp*, *telnet*, etc.

Web server domain name or an IP address such as 128.129.8.32

Optional port number. By default it is 80 for an http connection and 443 for a secured http connection.

Directory where resource is located

Name of requested item.

http://www.cs.ualberta.ca:80/~zaiane/courses/cmput410/index.html