The Hypertext Transfer Protocol

Pat Morin
COMP 2405
Outline

• History of HTTP
• The HTTP protocol
  – HTTP Requests
  – HTTP replies
• Summary
History of HTTP

- Initial version (not standardized) by Tim Berners-Lee (CERN) and implemented in the World's first web browser/server (Dec. 1990)
- HTTP 0.9 defined in 1991
- HTTP 1.0 defined in 1996 as RFC 1945
- HTTP 1.1 defined in 1997 as RFC 2068
HTTP Protocol Summary

• A typical HTTP transaction:
  – Client (browser) open connection to server
  – Client sends request to server
  – Server processes request
  – Server replies to client
  – Server closes connection

• Variants
  – Connection stays open
  – A proxy sits between client and server
The HTTP Request

- Request includes a header and optional body, separated by a blank line

```
GET /~morin/index.html HTTP/1.1
Host: localhost:3128
User-Agent: Mozilla/5.0 (X11; U; Linux i686 (x86_64); en-US; rv:1.8)
Accept: text/xml,application/xml,application/xhtml+xml,text/html;q=0.9,text/plain;q=0.8,image/png,*/*;q=0.5
Accept-Language: en-us,en;q=0.5
Accept-Encoding: gzip,deflate
Accept-Charset: ISO-8859-1,utf-8;q=0.7,*;q=0.7
Keep-Alive: 300
Connection: keep-alive
```
Main HTTP Request Types

- **GET**
  - Get (retrieve) a document

- **HEAD**
  - Same as GET, but only retrieve header

- **PUT**
  - Put information on the server

- **POST**
  - Send information to the server

- **OPTIONS**
  - Get information about the server
The HTTP Response

- Response includes a header and optional body

```
HTTP/1.1 200 OK
Date: Wed, 03 Jan 2007 12:10:27 GMT
Server: Apache/2.0.54 (Fedora)
Last-Modified: Tue, 02 Jan 2007 16:01:12 GMT
ETag: "548268-1b2d-d6ce0200"
Accept-Ranges: bytes
Content-Length: 6957
Connection: close
Content-Type: text/html; charset=UTF-8

<!DOCTYPE html PUBLIC
"-//W3C//DTD HTML 4.01 Transitional//EN"
"http://www.w3.org/TR/html4/loose.dtd">
<html>
....
</html>
```
Common Response Codes

- 200 OK
  - Success

- 404 Not Found
  - The specified document does not exist

- 403 Forbidden
  - The specified document exists, but can not be accessed

- 301 & 302 Document Moved
  - The document is at the new (specified) location

- 4** Error
  - Any 4** message indicates an error has occurred
Summary

- The basic HTTP transaction is a simple request from the client followed by a reply from the server.
- Both requests and replies have a header followed by an optional body.
- 200 indicates a success.
- 4xx indicates an error.
- We will see more complicated transactions and header fields later on.