



Hypertext Transfer Protocol

HTTP

Dr. Sam Hsu
Computer Science & Engineering
Florida Atlantic University



HTTP

- Request-Response Type Protocol
- Two Simple Properties
- Message Types
- Byte Range Operations
- Persistent Connections
- Pipelining
- Cache
- Chunked Encoding
- Cookies
- Formats of Dates and Times
- Efficiency
- URLs



Introduction (1/2)

- The "native" WWW protocol for data retrieval
 - Was designed for distributed, collaborative, hypermedia information systems
 - Original intent was for transferring hypertext documents.
 - Has been in use by WWW since 1990.
- An application-level, client/server protocol
 - Files of any data type can be transferred.

3



Introduction (2/2)

- Currently used almost exclusively with TCP
 - Port number 80.
 - One independent TCP connection per HTTP request
 - For older versions: 0.9, 1.0
 - One persistent TCP connection for multiple HTTP requests
 - For newer version: 1.1

4

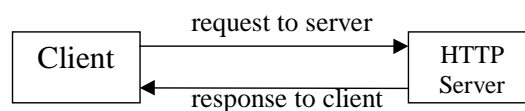
A Request/Response Type (1/4)

- A request/response protocol between a client and a server
 - The requesting client is known as the *user agent*.
 - Examples include a Web browser, spider, or other end-user tool.
 - The responding server is referred to as the *origin server*.
- There may be some intermediaries between a client and its server.
 - Such as proxies, gateways, and tunnels.

5

A Request/Response Type (2/4)

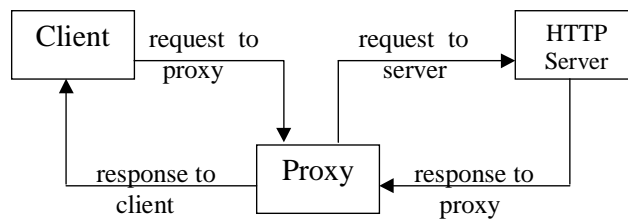
Direct



6

A Request/Response Type (3/4)

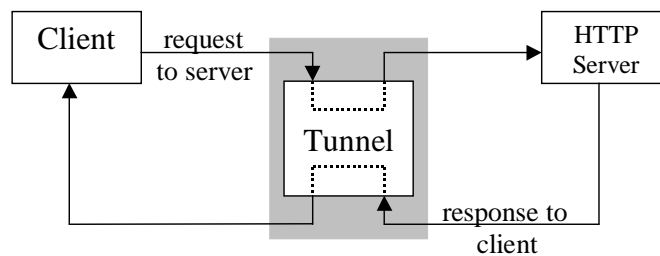
Through a proxy agent



7

A Request/Response Type (4/4)

Through a tunnel agent



8



Two Simple Properties

- Stateless
 - No memory of what files were transferred before.
- Pseudo-anonymous clients
 - Server only sees the IP address of a client, not individual users.

9



HTTP Is Negotiable

- Contents of negotiation:
 - Media type
 - Character set
 - Language
- Negotiation can be:
 - Agent-driven: from client.
 - Server-driven: from server.
 - Transparent: from an intermediate cache on behalf of the server.

10



HTTP Message Types (1/2)

- Request

request-line

header (0 or more)

<blank line>

body

where format of *request-line* is

method request-URI HTTP-version

- A *method* may be one of **HEAD, GET, POST, PUT, DELETE, TRACE, or OPTIONS.**

11



HTTP Message Types (2/2)

- Response

status-line

header (0 or more)

<blank line>

body

Where format of *status-line* is

HTTP-version response-code response-phrase

12



Some HTTP Header Names

<i>Header Name</i>	<i>Request</i>	<i>Response</i>	<i>Body</i>
Allow			•
Authorization	•		
Content-Encoding			•
Content-Length			•
Content-Type			•
Date	•	•	
Expires			
From	•		
If-Modified-Since	•		
Last-Modified			•
Location		•	
MIME-Version	•	•	
Pragma	•	•	
Referer	•		
Server		•	
User-Agent	•		
WWW-Authenticate		•	

13



Some HTTP 3-digit Response Codes

<i>Response</i>	<i>Description</i>
	Success
200	OK, request succeeded.
201	OK, new resource created (POST command).
202	Request accepted but processing not completed.
204	OK, but no content to return.
	Redirection; further action need be taken by user agent.
301	Requested resource has been assigned a new permanent URL.
302	Requested resource resides temporarily under a different URL
304	Document has not been modified (conditional GET).
	Client error.
400	Bad request.
401	Unauthorized; request requires user authentication.
403	Forbidden for unspecified reason.
404	Not found.
	Server error.
500	Internal server error.
501	Not implemented.
502	Bad gateway; invalid response from gateway or upstream server.
503	Service temporarily unavailable.

14



Example 1

- polaris% **telnet www.cse.fau.edu 80**
- Trying 131.91.96.106...
- Connected to polaris.cse.fau.edu.
- Escape character is '^['.
- **HEAD / http/1.0**

- HTTP/1.1 200 OK
- Date: Sat, 10 Nov 2007 15:41:28 GMT
- Server: Apache/2.0.48 (Unix) mod_ssl/2.0.48 OpenSSL/0.9.7c DAV/2 PHP/4.3.4
- Last-Modified: Mon, 29 Oct 2007 19:05:07 GMT
- ETag: "450923-12ed-60ef3ec0"
- Accept-Ranges: bytes
- Content-Length: 4845
- Connection: close
- Content-Type: text/html; charset=ISO-8859-1

- Connection closed by foreign host.

15



Example 2

- polaris% **telnet www.cse.fau.edu 80**
- Trying 131.91.96.106...
- Connected to polaris.cse.fau.edu.
- Escape character is '^['.
- **GET / http/1.0**

- HTTP/1.1 200 OK
- Date: Sat, 10 Nov 2007 15:46:50 GMT
- Server: Apache/2.0.48 (Unix) mod_ssl/2.0.48 OpenSSL/0.9.7c DAV/2 PHP/4.3.4
- Last-Modified: Mon, 29 Oct 2007 19:05:07 GMT
- ETag: "450923-12ed-60ef3ec0"
- Accept-Ranges: bytes
- Content-Length: 4845
- Connection: close
- Content-Type: text/html; charset=ISO-8859-1

- <HTML>
- <HEAD>
- <STYLE TYPE='text/css'>
- ...

16



Example 3

- polaris% telnet www.cse.fau.edu 80
- Trying 131.91.96.106...
- Connected to polaris.cse.fau.edu.
- Escape character is '^']'.
- **GET /images/facstaff/fs_hsu.jpg http/1.0**
- **From: sam@cse.fau.edu**

- HTTP/1.1 200 OK
- Date: Sat, 10 Nov 2007 15:49:47 GMT
- Server: Apache/2.0.48 (Unix) mod_ssl/2.0.48 OpenSSL/0.9.7c DAV/2 PHP/4.3.4
- Last-Modified: Tue, 06 Jun 2006 13:23:24 GMT
- ETag: "2a8138-1519-27cb0b00"
- Accept-Ranges: bytes
- Content-Length: 5401
- Connection: close
- Content-Type: image/jpeg
- ...

17



Example 4

- polaris% telnet www.cse.fau.edu 80
- Trying 131.91.96.106...
- Connected to polaris.cse.fau.edu.
- Escape character is '^']'.
- **GET /images/facstaff/fs_hsu.jpg http/1.0**
- **If-Modified-Since: Sat, 01 Jan 2001 00:00:01 GMT**

- HTTP/1.1 200 OK
- Date: Sat, 10 Nov 2007 15:52:45 GMT
- Server: Apache/2.0.48 (Unix) mod_ssl/2.0.48 OpenSSL/0.9.7c DAV/2 PHP/4.3.4
- Last-Modified: Tue, 06 Jun 2006 13:23:24 GMT
- ETag: "2a8138-1519-27cb0b00"
- Accept-Ranges: bytes
- Content-Length: 5401
- Connection: close
- Content-Type: image/jpeg
- ...
- Connection closed by foreign host.

18



Example 5

- polaris% **telnet www.cse.fau.edu 80**
- Trying 131.91.96.106...
- Connected to polaris.cse.fau.edu.
- Escape character is '^']'.
- **OPTIONS * http/1.0**

- HTTP/1.1 200 OK
- Date: Sat, 10 Nov 2007 15:57:49 GMT
- Server: Apache/2.0.48 (Unix) mod_ssl/2.0.48 OpenSSL/0.9.7c DAV/2 PHP/4.3.4
- Allow: GET,HEAD,POST,OPTIONS,TRACE
- Content-Length: 0
- Connection: close
- Content-Type: text/plain; charset=ISO-8859-1

- Connection closed by foreign host.

19



Example 6

- polaris% **telnet www.cse.fau.edu 80**
- Trying 131.91.96.106...
- Connected to polaris.cse.fau.edu.
- Escape character is '^']'.
- **TRACE home.netscape.com http/1.0**

- HTTP/1.1 400 Bad Request
- Date: Sat, 10 Nov 2007 16:01:09 GMT
- Server: Apache/2.0.48 (Unix) mod_ssl/2.0.48 OpenSSL/0.9.7c DAV/2 PHP/4.3.4
- Vary: accept-language,accept-charset
- Accept-Ranges: bytes
- Connection: close
- Content-Type: text/html; charset=iso-8859-1
- Content-Language: en
- Expires: Sat, 10 Nov 2007 16:01:09 GMT

- <?xml version="1.0" encoding="ISO-8859-1"?>
- <!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
- ...
- Connection closed by foreign host.

20



Example 7

- D:\sam **tracert home.netscape.com**
- Tracing route to www-ld1.netscape.com [207.200.75.200]
- over a maximum of 30 hops:
- 1 10 ms 20 ms <10 ms segate.fau.edu [131.91.80.1]
- 2 20 ms 10 ms 20 ms s5-0-0-6-0.tamqfl1-cr1.bbnplanet.net [4.24.96.21]
- 3 10 ms * 20 ms fa1-0-0.tamqfl1-cr2.bbnplanet.net [4.24.4.166]
- 4 20 ms 30 ms 31 ms s4-1-0.atlanta1-br2.bbnplanet.net [4.24.98.226]
- 5 20 ms * 40 ms core4-hssi5-0-0.Atlanta.cw.net [204.70.10.169]
- 6 * 300 ms 110 ms corerouter1.SanFrancisco.cw.net [204.70.9.131]
- 7 80 ms 90 ms 100 ms bordercore2.SanFrancisco.cw.net [166.48.14.1]
- 8 100 ms * 101 ms netscape.SanFrancisco.cw.net [166.48.15.254]
- 9 100 ms 100 ms * h-207-200-69-74.netscape.com [207.200.69.74]
- 10 * * * Request timed out.
- 11 * * * Request timed out.
- . . .

21



Byte Range Operations

- Client can request portion of a resource.
- Examples:
 - Range: bytes=2–100
 - Range: bytes=500–700, 900–

22



Example 8

- polaris% telnet www.cse.fau.edu 80
- Trying 131.91.96.106...
- Connected to polaris.cse.fau.edu.
- Escape character is '^'].
- GET / http/1.0
- Range: bytes=100-120

- HTTP/1.1 206 Partial Content
- Date: Sat, 10 Nov 2007 16:14:23 GMT
- Server: Apache/2.0.48 (Unix) mod_ssl/2.0.48 OpenSSL/0.9.7c DAV/2 PHP/4.3.4
- Last-Modified: Mon, 29 Oct 2007 19:05:07 GMT
- ETag: "450923-12ed-60ef3ec0"
- Accept-Ranges: bytes
- Content-Length: 21
- Content-Range: bytes 100-120/4845
- Connection: close
- Content-Type: text/html; charset=ISO-8859-1
- ...

23



Persistent Connections

- In HTTP/1.1, when a connection is made, it remains open until it either gets timed out or receives a connection close request from either end.
 - Multiple HTTP requests can thus be made via a single TCP connection.
 - Similar to the *keep-alive* option supported in HTTP/1.0.
 - Resulting in fewer roundtrips for better performance.
- Persistent connections are default in HTTP/1.1.

24



Example 9

- polaris: **telnet www.fau.edu 80**
- Trying 131.91.128.83...
- Connected to www.fau.edu.
- Escape character is '^']'.
- **HEAD / http/1.1**
- **Host: www.fau.edu**

- HTTP/1.1 200 OK
- Date: Sat, 10 Nov 2007 16:18:13 GMT
- Server: Apache/2.0.48 (Unix) mod_ssl/2.0.48 OpenSSL/0.9.7c DAV/2 PHP/4.3.4
- Last-Modified: Mon, 29 Oct 2007 19:05:07 GMT
- ETag: "450923-12ed-60ef3ec0"
- Accept-Ranges: bytes
- Content-Length: 4845
- Content-Type: text/html; charset=ISO-8859-1

- <some time delay, nothing displayed>
- Connection closed by foreign host.

25



Example 10

- polaris% **telnet www.fau.edu 80**
- Trying 131.91.128.83...
- Connected to www.fau.edu.
- Escape character is '^']'.
- **OPTIONS * http/1.1**
- **Connection: close**
- **Host: www.fau.edu**

- HTTP/1.1 200 OK
- Date: Sat, 10 Nov 2007 16:22:49 GMT
- Server: Apache/1.3.29 (Unix) PHP/5.0.1
- Content-Length: 0
- Allow: GET, HEAD, OPTIONS, TRACE
- Connection: close

- Connection closed by foreign host.

26



Pipelining (1/2)

- A new feature in HTTP/1.1 to allow clients to send multiple requests at once.
 - Server can thus send multiple answers together in one connection.
 - An example would be to send several in-line images on a Web page to a requesting client browser at once.
- Is particularly useful for high latency connections, such as a satellite Internet connection.

27



Pipelining (2/2)

- Caution: One should not pipeline requests that are not "idempotent".
 - A sequence is idempotent if a single execution of the entire sequence always yields a result that is not changed by a re-execution of all, or part, of that sequence.
 - That is, multiple identical requests should have the same effect as a single request.
 - HEAD, GET, OPTIONS, and TRACE are considered to be idempotent.
 - Be cautious in using **GET** – It can be misused to cause a change in server state.

28



Example 11

```
polaris% telnet www.fau.edu 80
Trying 131.91.128.83...
Connected to www.fau.edu.
Escape character is '^]'.
OPTIONS * http/1.1
Host: www.fau.edu

HEAD / http/1.1
Host: www.fau.edu

HTTP/1.1 200 OK
Date: Sat, 10 Nov 2007 16:53:39 GMT
Server: Apache/1.3.29 (Unix) PHP/5.0.1
Content-Length: 0
Allow: GET, HEAD, OPTIONS, TRACE

HTTP/1.1 200 OK
Date: Sat, 10 Nov 2007 16:53:39 GMT
Server: Apache/1.3.29 (Unix) PHP/5.0.1
Last-Modified: Thu, 31 Aug 2006 17:54:42 GMT
ETag: "a36112-3a5-44f72262"
...
```

29



Example 12

```
polaris% telnet www.fau.edu 80
Trying 131.91.128.83...
Connected to www.fau.edu.
Escape character is '^]'.
OPTIONS * http/1.1
Host: www.fau.edu

HTTP/1.1 200 OK
Date: Sat, 10 Nov 2007 16:58:50 GMT
Server: Apache/2.0.48 (Unix) mod_ssl/2.0.48 OpenSSL/0.9.7c DAV/2 PHP/4.3.4
Allow: GET,HEAD,POST,OPTIONS,TRACE
Content-Length: 0
Content-Type: text/plain; charset=ISO-8859-1

HEAD / http/1.1
Connection: close
Host: www.fau.edu

HTTP/1.1 200 OK
Date: Sat, 10 Nov 2007 16:58:57 GMT
Server: Apache/2.0.48 (Unix) mod_ssl/2.0.48 OpenSSL/0.9.7c DAV/2 PHP/4.3.4
Last-Modified: Mon, 29 Oct 2007 19:05:07 GMT
ETag: "450923-12ed-60ef3ec0"
Accept-Ranges: bytes
Content-Length: 4845
Connection: close
Content-Type: text/html; charset=ISO-8859-1

Connection closed by foreign host.
```

30



Cache

- A local copy of response messages.
- Five categories of cache control directives:
 - Cachable
 - Not cachable
 - How old can it be
 - Don't serve anything past its age
 - Don't transform

31



Chunked Encoding

- Data transfer in chunks of known length.
- For sending data with an unknown initial length due to persistent connections.
- Encoding format:

Size	data chunk	...	Size	data chunk	0 size	footer
------	------------	-----	------	------------	--------	--------

32



Cookie

- A *cookie* is a small piece of data containing some user-specific information.
 - Embedded in the HTML documents flowing back and forth between the user's browser and the server.
- Cookies may be created/stored the following ways:
 - Generated by client to reside on client site.
 - Generated by server to reside on client site.
 - Generated by server to reside on server site.

33



Digest Access Authentication

- To avoid passing user's password in plain text across the network.
- Using the MD5 (Message Digest 5) algorithm to generate a *digest* on the concatenation of password and some other values.
- The *digest* is then sent over the network for authentication purposes.

34



Formats of Dates and Times

- Based on UNIX *asctime()*
 - Sat Nov 01 14:50:20 1997
- Based on RFC 850
 - Saturday, 01-Nov-97 14:50:20 GMT
- Based on RFC 1123 (required by HTTP/1.1)
 - Sat, 01 Nov 1997 14:50:20 GMT

35



Gateway

- A server acting as an intermediary for some other server.
- Unlike a proxy, a gateway receives requests as it were the original server.
- A requesting client may not be aware that it is communicating with a gateway.

36



Proxy

- An intermediary program which acts as both a server and a client for the purpose of making requests on behalf of other clients.
- Requests received by a proxy may be serviced internally or passed on, with possible translation, to other servers.

37



Tunnel

- An intermediary program acting as a blind relay between two connections.
- Not considered a party to the HTTP communication, although it may be initiated by an HTTP request.

38



Version

- Included in each request/response.
- Specifies the upper limit on the capabilities of the sender.

39



Efficiency

- Time consuming three-way handshaking process for establishing TCP connections
 - E.g. a typical page with several images which need a separate HTTP request for each image.
 - Each HTTP request requires one TCP connection in HTTP/0.9 and HTTP/1.0.
- Some suggestions
 - One TCP connection for several transfer requests.
 - Maintaining several TCP connections.

40



User Identification

- Adding states.
- Some suggestions
 - UID explicitly included.
 - Hidden UID.
 - Page retrieved should contain parameter-value pair (Netscape).

41



URLs (1/2)

- Used to designate objects within the WWW.
- Most widely used one within the family technically known as universal resource identifiers (URI)
 - URL: Universal Resource Locator.
 - URN: Universal Resource Name.
 - URC: Uniform Resource Characteristics.

42



URLs (2/2)

- Composed of three parts
 - Protocol
 - Host [[:port]
 - Pathname [#ref]
- Says nothing about the object to which it points
 - Only some informal hints by name.
- An open issue
 - How to keep it up-to-date?

43



MIME

- Multiple Internet Mail Extensions
 - A standard used to encode binary data as printable ASCII text for transmission across the Internet.
- Format of a MIME type:
 - *type/subtype*

44



MIME Types

<i>Type/Subtype</i>	<i>Description</i>
text/html	Data is HTML
text/plain	Data is plain text
text/richtext	Data is rich text format, meeting RFC 1523
text/tab-separated-values	Data is tab separated text
video/mpeg	Data is MPEG video
video/quicktime	Data is Quicktime movie
audio/x-wav	Data is Microsoft "wav" format audio file
image/gif	Data is GIF image
image/tiff	Data is TIFF image
application/msword	Data is Microsoft Word file
application/rtf	Data is Microsoft rich text format
application/zip	Data is compressed with PKZIP
message/rfc822	E-mail message, Internet format
multipart/mixed	contains multiple MIME types

45



For More Information

- <http://wdvl.com/Authoring/HTML/Tutorial/http.html>
- http://www.yahoo.com/Computers_and_Internet/Internet/World_Wide_Web/HTTP
- <http://www.ncsa.uiuc.edu/SDG/IT94/Proceedings/DDay/mogul/HTTPLatency.html>
- <http://www.w3.org/Addressing/URL/url-spec.html>
- RFC 2616 – R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, T. Berners-Lee, "Hypertext Transfer Protocol – HTTP/1.1", June 1999.
- <http://www.w3.org/Protocols/rfc2616/rfc2616.html>
- RFC 1945 – T. Berners-Lee, R. Fielding, & H. Frystyk, "Hypertext Transfer Protocol – HTTP/1.0", 05/07/1996
- David Whalen, "The Unofficial Netscape Cookie FAQ", <http://www.cookiecentral.com/faq/>

46