Hypertext Transfer Protocol (HTTP)

Raj Jain
Washington University in Saint Louis
Saint Louis, MO 63130
Jain@cse.wustl.edu

These slides are available on-line at:

http://www.cse.wustl.edu/~jain/cse473-05/

- Hypertext Transfer Protocol (HTTP)
- Hypertext Markup Language (HTML)
- Key HTTP Terms
- URI vs URL vs URN
- Intermediate HTTP Systems
- HTTP Message Structure
- HTTP 1.1 Features

Hypertext Transfer Protocol (HTTP)

- Hypertext: Documents contain pointers to other documents
- HTTP: Protocol used between web browsers and web servers
- Originally designed for hypertext. Not limited to text. Used for all types of media.
- Transaction oriented client/server protocol
- Uses TCP connections
- Stateless: New TCP connection for each transaction (Changed in HTTP v1.1)
- Presentation separate from content
  - Presentation controlled by the browser
  - Content provided by the server
- Hypertext markup language (HTML)
- Hypertext transfer protocol (HTTP)

Hypertext Markup Language (HTML)

- Header, Body, Anchors, Hyper References
- **Sample Code**:

```
<HTML>
<HEAD>
<TITLE>Hello</TITLE>
</HEAD>
<BODY>
<A
   HREF="http://www.cse.wustl.edu/~jain/greetings.html">Ho
   w are you?</A>
</BODY>
</HTML>
```

Output

Hello

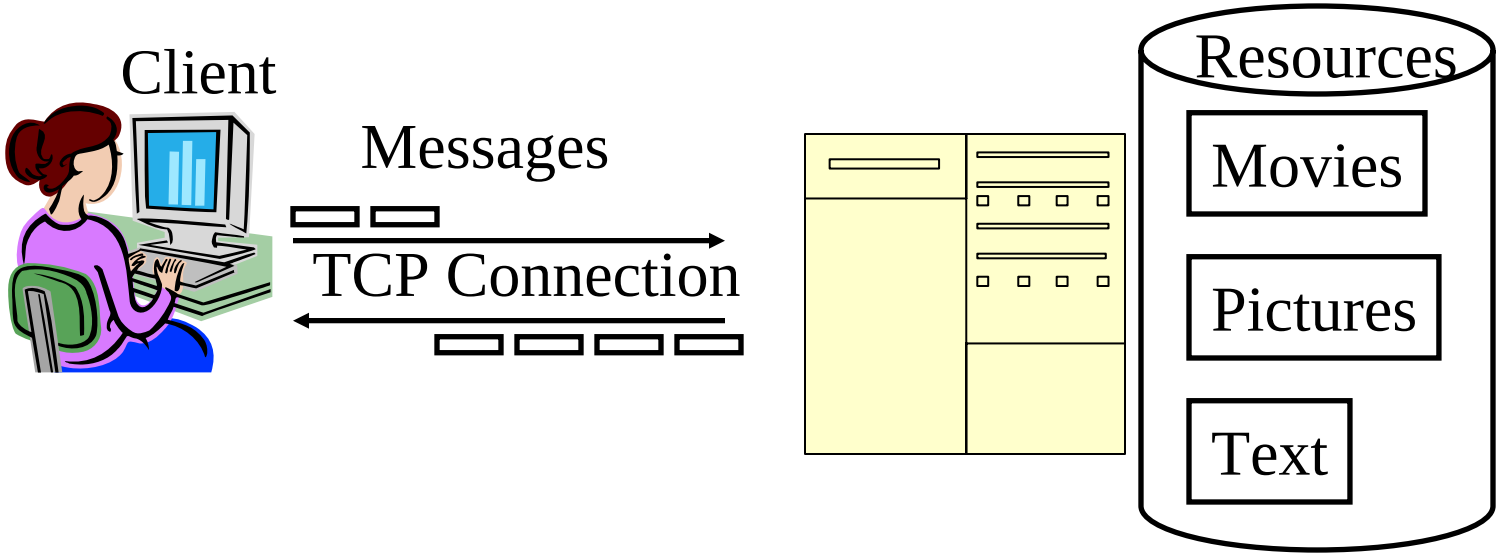
How are you?

A Sample List of HTML Tags

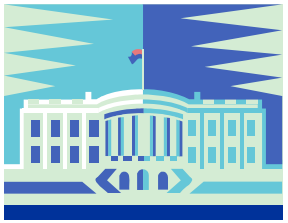
<A> </A>	Anchor (link or name)
<BODY> </BODY>	Contents
 	Break
<FORM> </FORM>	Input form
<H1> </H1>	Heading level 1
<HEAD> </HEAD>	Header of a document
<HR>	Horizontal Rule
<HTML> </HTML>	The doc type is HTML
<LI>	List Item
<OL> </OL>	Ordered List
	Paragraph break
<PRE> </PRE>	Preformatted text
<TITLE> </TITLE>	Document title
<UL>	Unnumbered list

Key HTTP Terms

- User agent: Browsers, spiders
- Client and Server systems
- Connection: TCP
- Message
- Resource: Object or service. Identified by URI/URN/URL
- Entity: Representation of a resource with a header and body



URI vs URL vs URN



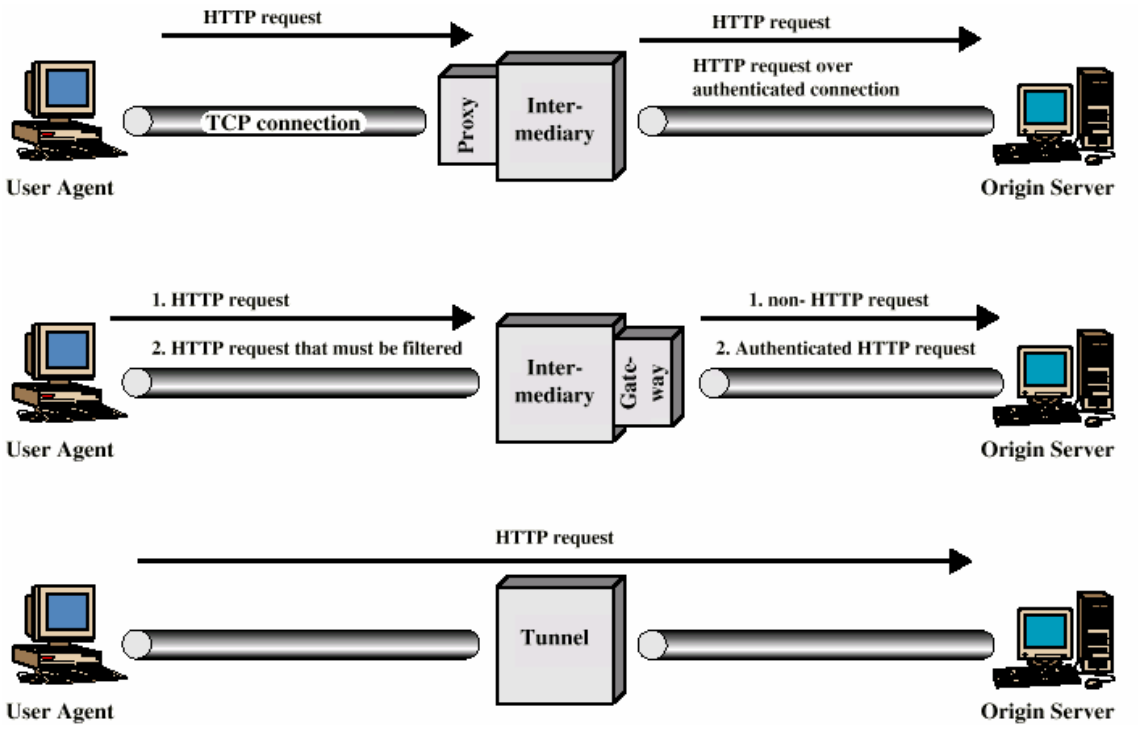
- Uniform Resource Name (URN): Host/path/object#name
  www.cse.wustl.edu/~jain/refs/wir_refs.htm#consortia
- Uniform Resource Identifier (URI):
  Name + Scheme (How to get it)
  Scheme://host/path/object#name

http://www.cse.wustl.edu/~jain/refs/wir_refs.htm#consortia

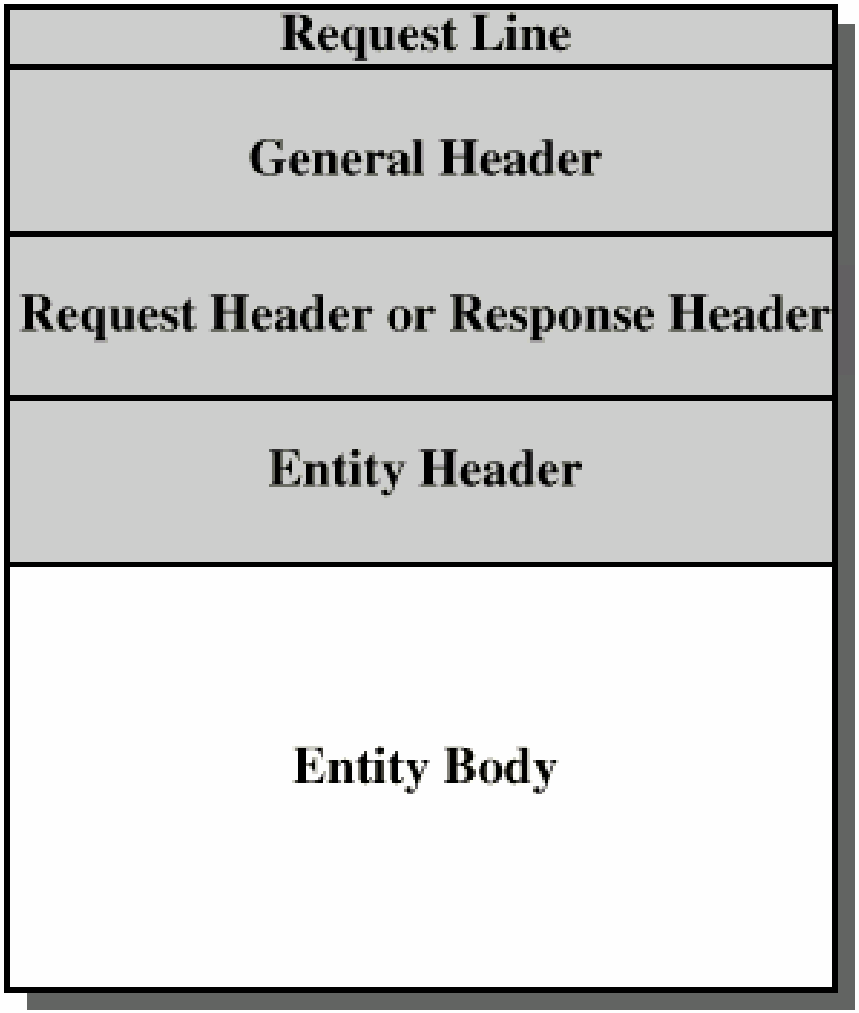
- Uniform Resource Locator (URL): URI based on location rather than any other attributes
- URI vs URL
  - IDs need not be location based ⇒ URLs are subset of URIs
  - Most URIs are based on location ⇒ URLs ≅ URIs
    In practice, URL and URI are used interchangeably.
  - URI is preferred in technical documentation.

Intermediate HTTP Systems

- Proxy: For performance enhancements. **Caches** pages.
- Gateway: For security. Used at firewall boundary.
- Tunnel: simple relay
- Origin server: Holds original content



HTTP Message Structure



Sample HTTP Exchange

telnet www.cse.wustl.edu http
Connected to hydra.cs.wustl.edu.
Escape character is '^]'.
GET /~jain/index.html HTTP/1.0 } Request Line
Accept: text/plain, text/html } Request Header

HTTP/1.0 200 OK } Response Line
Server: Netscape-Enterprise/2.0a
Date: Tue, 25 Feb 2005 05:04:11 GMT
Accept-ranges: bytes
Last-modified: Tue, 25 Feb 2005 05:03:07 GMT
Content-length: 84
Content-type: text/html
} Response Header

Sample HTTP Exchange (cont)

<HTML>
<HEAD>
<TITLE>Hello</TITLE>
</HEAD>
<BODY>
Hello! How are you?
</BODY>
</HTML>
Connection closed by foreign host.

Entity Header

Entity Body

HTTP Requests

- GET Return the contents
- HEAD Return the header
- POST Treat the document as a script and send some data to it
- PUT Replace the contents with some data
- DELETE Delete the indicated document

  etc.

HTTP Request Headers

Header	Description
From	Email address of user
User-Agent	Client software
Accept File	File types that client will accept
Accept-encoding	Compression methods
Accept-Language	Languages
Referrer	URL of the last document the client displayed
If-Modified-Since	Return document only if modified since specified
Content-length	Length (in bytes) of data to follow

HTTP Status Codes

Code	Text
2xx	Success
3xx	Redirection
301	Moved
302	Found
4xx	Client Errors
400	Bad Request
401	Unauthorized
404	Not found
5xx	Server Errors
500	Internal Error
502	Service overloaded

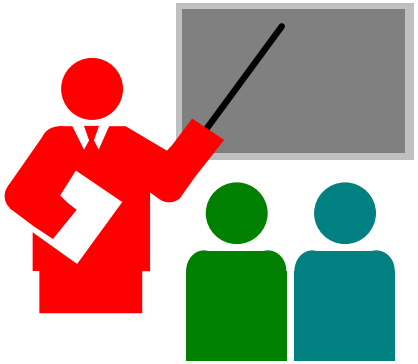
HTTP Response Headers

Header	Description
Server	Server software
Date	Current Date
Last-Modified	Modification date of document
Expires	Date at which document expires
Location	The location of the document in redirection responses
Pragma	A hint, e.g., no cache
MIME-version	
Link	URL of document's parent
Content-Length	Length in bytes
Allowed	Requests that user can issue, e.g., GET

HTTP 1.1 Features

- Persistent TCP Connections: Remain open for multiple requests
- Partial Document Transfers: Clients can specify start and stop positions
- Conditional Fetch: Several additional conditions
- Better content negotiation
- More flexible authentication

Summary



- Content separate from presentation
- Protocol messages in plain text
- HTML to define the media
- HTTP to transfer the media
- HTTP 1.1 allows for persistent connections

Reading Assignment

- Read Section 22.2 of Stallings' 7th edition