CS321: Computer Networks

WWW, HTTP

Dr. Manas Khatua
Assistant Professor
Dept. of CSE
IIT Jodhpur
E-mail: manaskhatua@iitj.ac.in

Standard Applications

- We would discuss six standard client-server application programs
  - World Wide Web (WWW), Hyper Text Transfer Protocol (HTTP)
  - File Transfer Protocol (FTP)
  - Electronic mail: SMTP, POP
  - TELNET
  - Secure Shell (ssh)
  - Domain Name System (DNS)

  - Dynamic Host Configuration Protocol (DHCP)
  - Simple Network Management Protocol (SNMP)

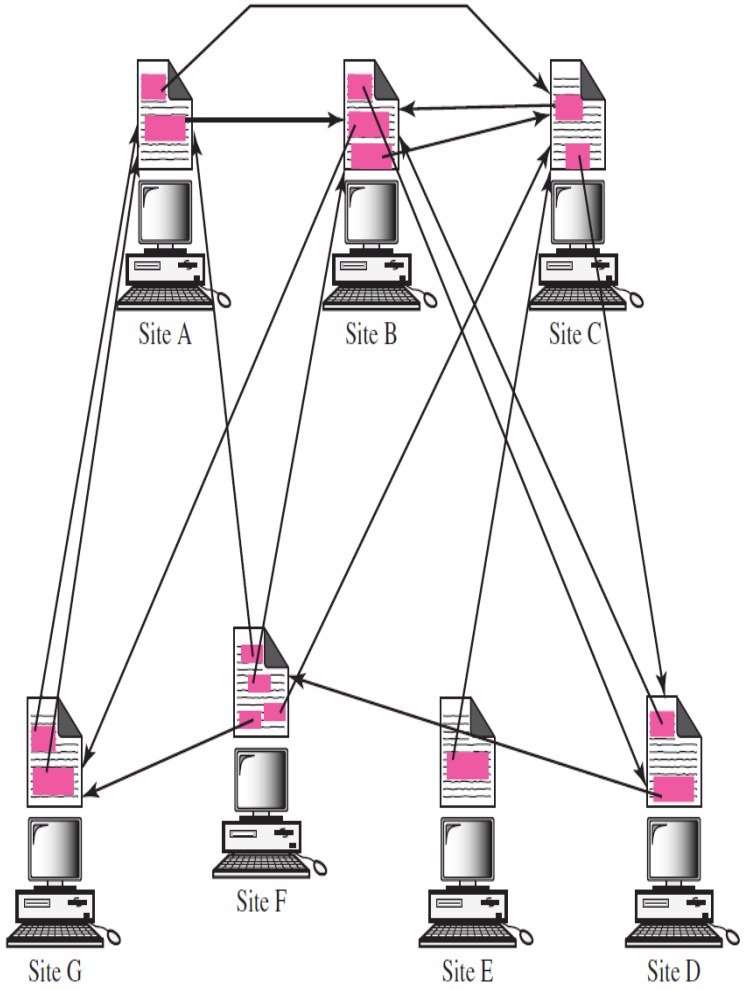
WWW

- The idea of the Web was first proposed by Tim Berners-Lee in 1989 at *CERN* to allow several researchers at different locations to access each others' researches.
- The commercial Web started in the early 1990s.
- The Web today is a repository of information in which the documents, called *web pages*, are distributed all over the world and related documents are linked together.
- The linking of web pages was achieved using a concept called *hypertext*.
- Today, the term *hypertext* has been changed to *hypermedia* (to indicate text, image, audio, video)

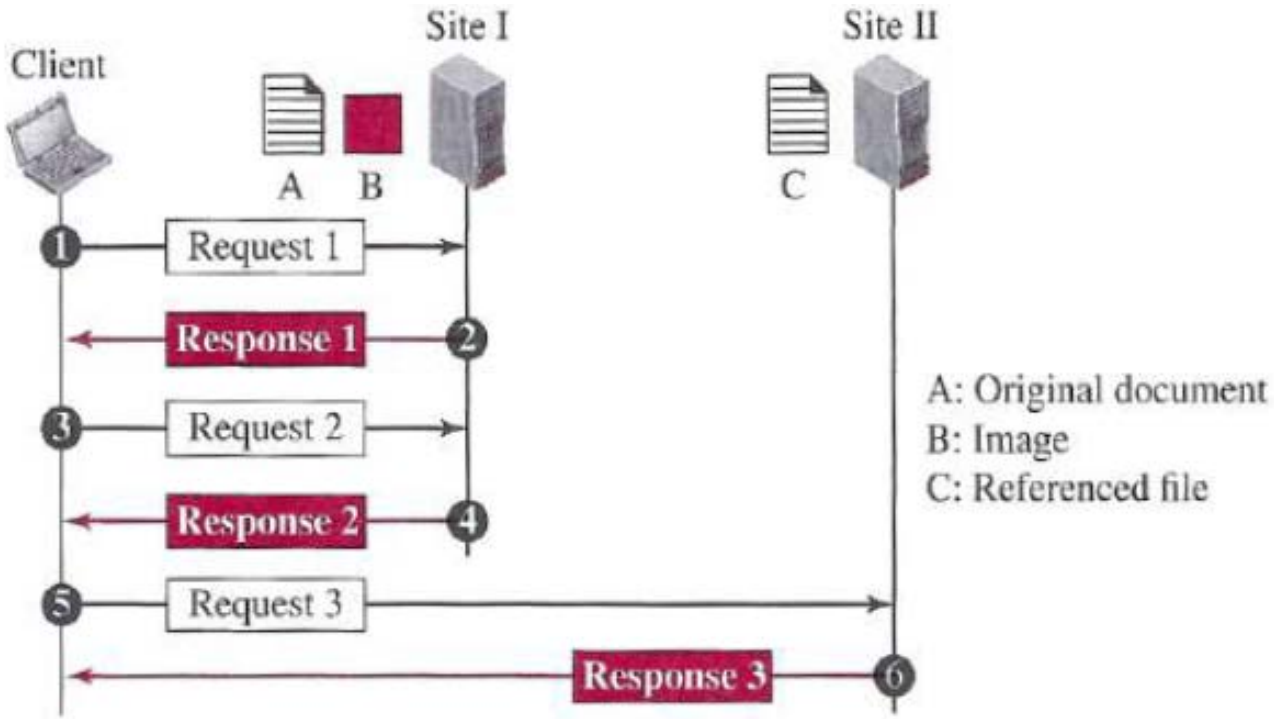
Architecture of WWW

- The WWW today is a distributed client-server service
- a client using a browser can access the services using a server

- The service provided is distributed over many locations called *sites*
- Each site holds one or more web pages
- Each web page can contain some links to other web pages in the same or other sites
- Each web page is a file with a name and address.



Cont…

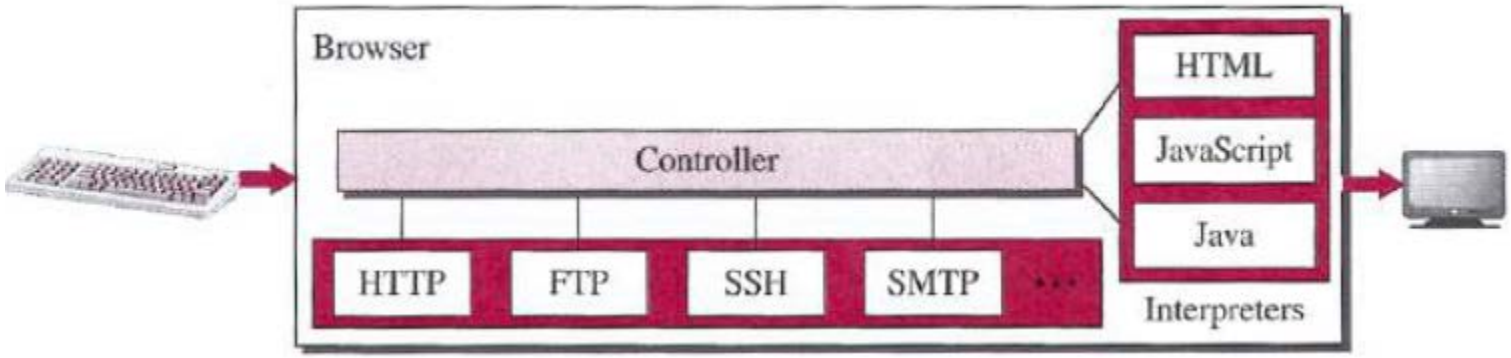


- In this example, we need three transactions if we want to see the whole document
  - retrieves a copy of the main document (file A),
  - retrieve a copy of the image (file B)
  - retrieving a copy of file C.

Web Client (Browser)

- Browser interprets and display a web page

- Each browser usually consists of three parts:
  - a controller
  - client protocols (used to access pages)
  - interpreters (used to display pages)



Web Server

- The web page is stored at the server.
- Each time a request arrives, the corresponding document is sent to the client.
- A server can become more efficient using the concept of cash memory, multithreading, multiprocessing.

- Popular Web Clients:
  - Internet Explorer
  - Netscape Navigator
  - Firefox
  - Google Chrome

- Popular Web Servers:
  - Apache
  - Microsoft Internet Information Server

Uniform Resource Locator (URL)

- A web page, as a file, needs to have a unique identifier to distinguish it from other web pages.

- To define a web page, we need three identifiers:
  - *host (*can be the IP-address/name of the server)
  - *port (*predefined for the client-server application)
  - *path* (The path identifies the location and the name of the file in the underlying OS. E.g., */top/next/last/myfile*)
- Further, we need to tell the browser what client-server application we want to use, which is called the *protocol*.

- To combine these four pieces together, the URL has been designed.

  protocol://host/path

  protocol://host:port/path

Web Documents

- The documents in the WWW can be grouped into three broad categories: Static, Dynamic, Active

- Static Documents:
  - fixed-content documents that are created and stored in a server.
  - In other words, the contents of the file are determined when the file is created, not when it is used.

  - Static documents are prepared using :
    HyperText Markup Language (HTML)
    Extensible Markup Language (XML)
    Extensible Style Language (XSL)
    Extensible Hypertext Markup Language (XHTML)

Cont…

- Dynamic Documents:
  - A dynamic document is created by a web server whenever a browser requests the document.
  - E.g,: retrieval of the time and date from a server
  - to retrieve a dynamic document we use:
    - *Java Server Pages (JSP)*
    - *Active Server Pages (ASP)*
    - *ColdFusion*

- Active Documents:
  - When we need a program or a script to be run at the client site
  - E.g.: a program that interacts with the user.
  - One way to create active document is to use Java applets, a program written in Java on the server. It is compiled and ready to be run.
  - Another way is to use JavaScripts, but download and run the script at the client site.

HyperText Transfer Protocol (HTTP)

- used to define how the client-server programs can be written to retrieve web pages from the Web.

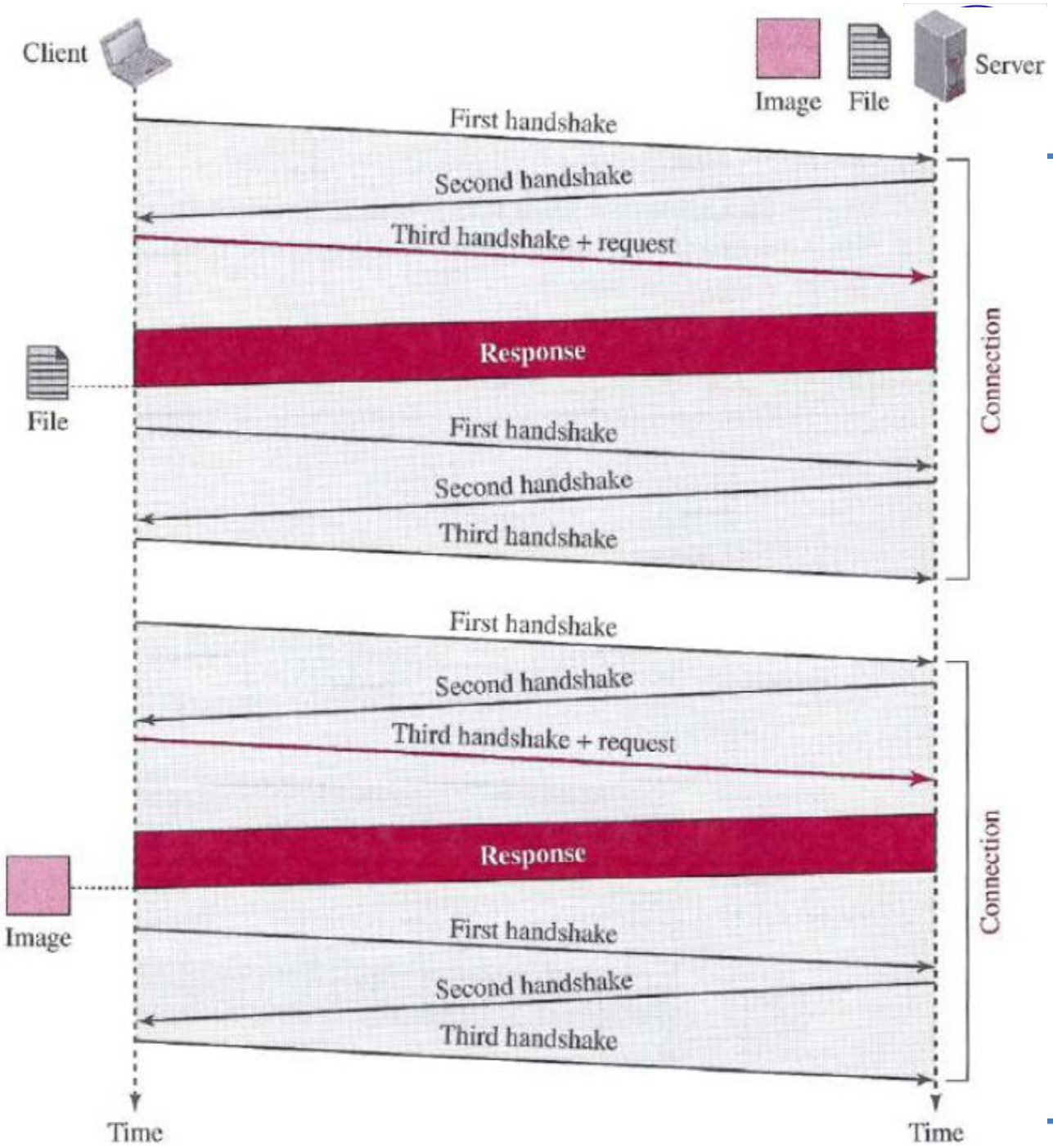
- An HTTP client sends a request; an HTTP server returns a response.
- HTTP uses the services of TCP

Cont…

- If the web pages are located on different servers
  - Create a new TCP connection for retrieving each object
- If some of the objects are located on the same server
  - *nonpersistent connection:* retrieve each object using a new TCP connection.
  - *persistent connection:* make a TCP connection and retrieve them all

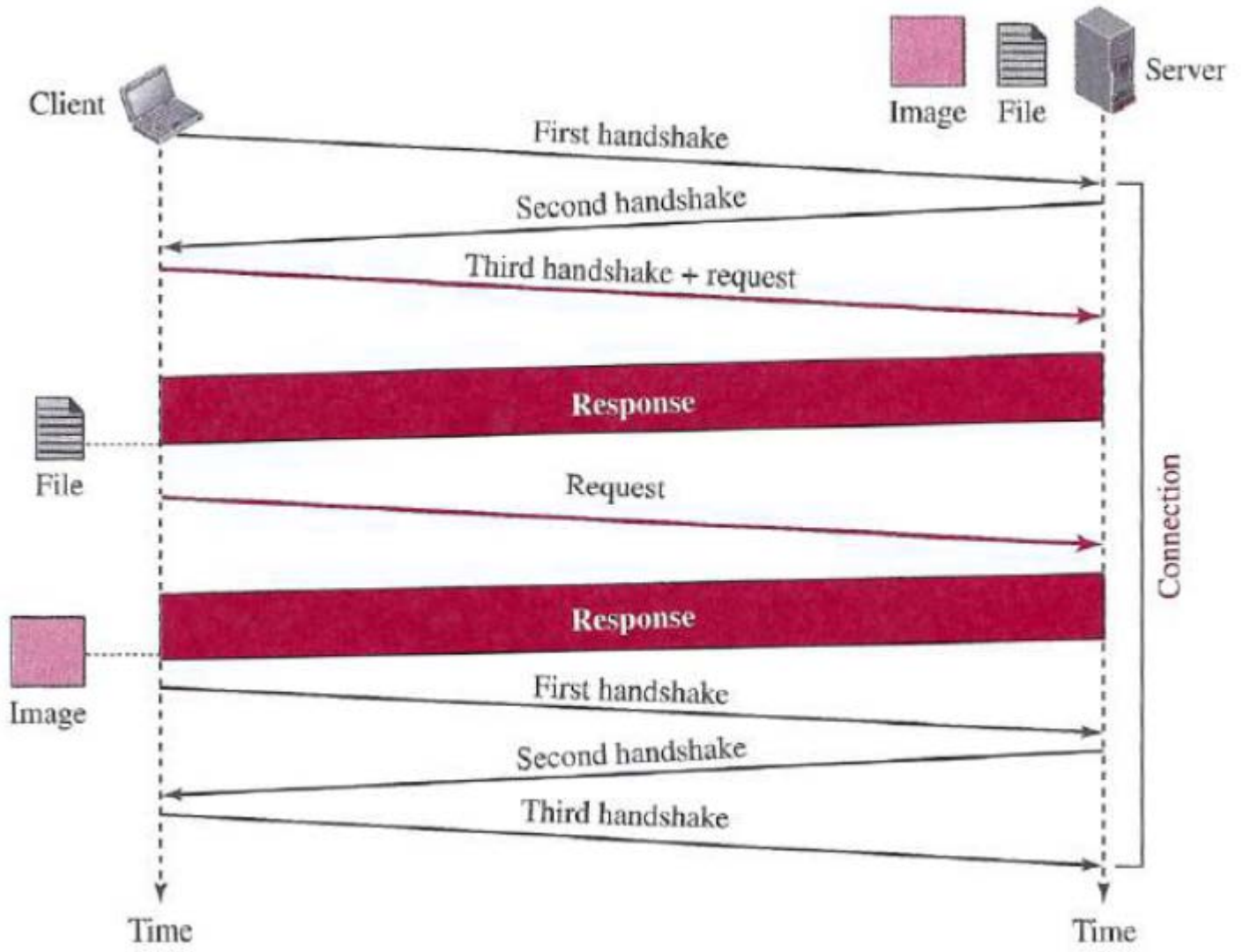
Cont…

Nonpersistent Connection

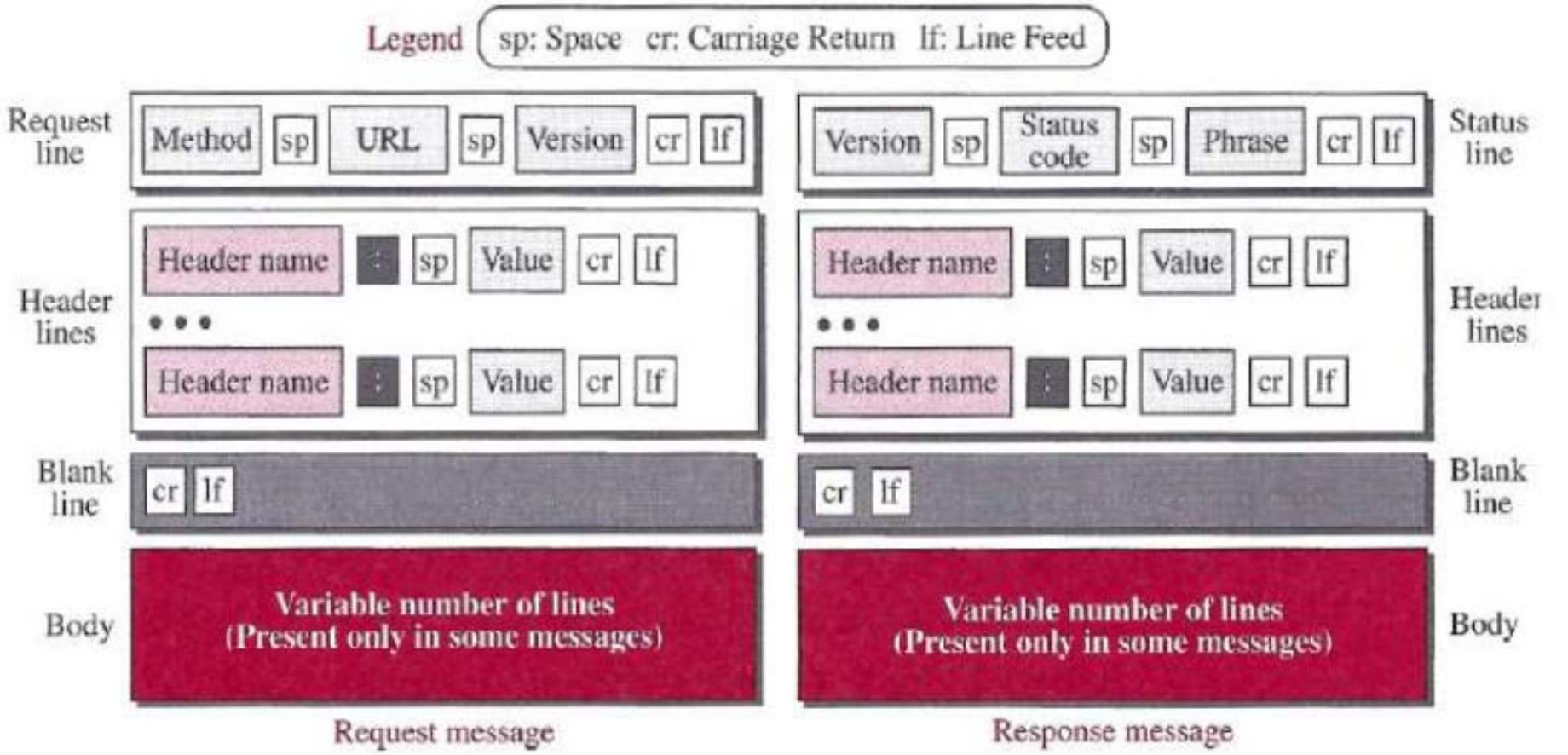


Cont…

Persistent Connection



Message Format in HTTP



Request Message

- There are three fields (*method*, *URL*, *version*) in request line separated by one space and terminated by two characters \n, \r.

- A line feed means moving one line forward. The code is \n. A carriage return means moving the cursor to the beginning of the line. The code is \r.
- Method: defines the request types. E.g., GET, PUT, POST, HEAD, …
  - GET: allows the client to send a request
  - PUT: allows the client to post a new web page on the server
  - HEAD: is used when the client needs only some information about the web page from the server
  - POST: is used to send some information to the server to be added to the web page or to modify the web page
  - TRACE: is used for debugging
  - DELETE: allows the client to delete a web page on the server
  - OPTIONS: allows the client to ask about the properties of a web page
  - CONNECT: a reserve method; it may be used by proxy servers

Cont…

- URL: defines the address and name of the corresponding web page
- Version: gives the version of the protocol

- After the request line, we can have zero or more *header* lines.
- Each header line sends additional information from the client to the server.
- Each header line has a header name, a colon, a space, and a header value.

- The body can be present in a request message. Usually, it contains the comment to be sent or the file to be published on the website when the method is PUT or POST.

Cont…

Table 26.2 *Request header names*

Header	*Description*
User-agent	Identifies the client program
Accept	Shows the media format the client can accept
Accept-charset	Shows the character set the client can handle
Accept-encoding	Shows the encoding scheme the client can handle
Accept-language	Shows the language the client can accept
Authorization	Shows what permissions the client has
Host	Shows the host and port number of the client
Date	Shows the current date
Upgrade	Specifies the preferred communication protocol
Cookie	Returns the cookie to the server (explained later)
If-Modified-Since	If the file is modified since a specific date

Response Message

- A response message consists of a status line, header lines, a blank line, and sometimes a body.

- There are three fields (version, status code, status phrase) in status line separated by spaces and terminated by a \n,\r
  - Version: defines the version of HTTP protocol
  - status code: defines the status of the request
  - status phrase: explains the status code in text form
- we can have zero or more response header lines
- Each header line sends additional information from the server to the client
- Each header line has a header name, a colon, a space, and a header value.

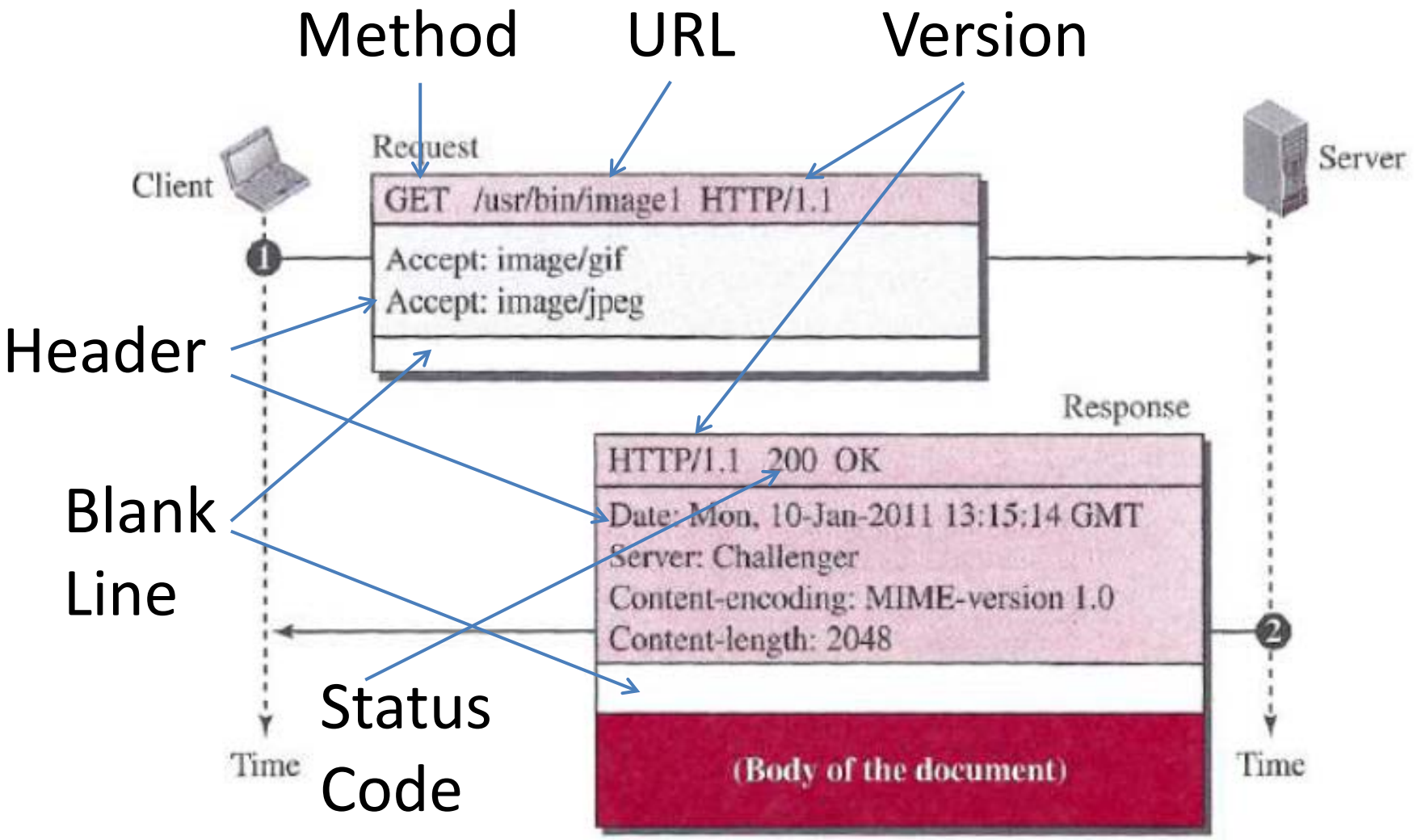
- The body contains the document to be sent from the server to the client.

Cont…

Table 26.3 *Response header names*

Header	*Description*
Date	Shows the current date
Upgrade	Specifies the preferred communication protocol
Server	Gives information about the server
Set-Cookie	The server asks the client to save a cookie
Content-Encoding	Specifies the encoding scheme
Content-Language	Specifies the language
Content-Length	Shows the length of the document
Content-Type	Specifies the media type
Location	To ask the client to send the request to another site
Accept-Ranges	The server will accept the requested byte-ranges
Last-modified	Gives the date and time of the last change

Example


Method
URL
Version
Client
Request
GET /usr/bin/image1 HTTP/1.1
Accept: image/gif
Accept: image/jpeg
Server
Header
Blank Line
Response
HTTP/1.1 200 OK
Date: Mon, 10-Jan-2011 13:15:14 GMT
Server: Challenger
Content-encoding: MIME-version 1.0
Content-length: 2048
(Body of the document)
Status Code
Time
Time

Conditional Request

- A client can add a condition in its request.

- In this case, the server will send the requested web page if the condition is met or inform the client otherwise.

Cont…

- The following shows how a client imposes the modification data and time condition on a request.

GET http://www.commonServer.com/information/file1 HTTP/1.1 — Request line

If-Modified-Since: Thu, Sept 04 00:00:00 GMT — Header line

Blank line

(Empty Body) — Empty body

- The status line in the response shows the file was not modified after the defined point in time. The body of the response message is also empty.

HTTP/1.1 304 Not Modified — Status line

Date: Sat, Sept 06 08 16:22:46GMT — First header line

Server: commonServer.com — Second header line

Blank line

(Empty Body) — Empty body

Cookies

- The World Wide Web was originally designed as a stateless entity.
  - A client sends a request; a server responds. Their relationship is over.

- Today the Web has other functions that need to remember some information about the clients.
  - Websites are being used as *electronic stores* that allow users to browse through the store, select wanted items, put them in an electronic cart, and pay at the end with a credit card.
  - Some websites need to allow access to *registered clients* only.
  - Some websites are used as *portals:* the user selects the web pages he wants to see.
  - Some websites are just *advertising* agencies.

- For these purposes, the cookie mechanism was devised.

Creating and Storing Cookies

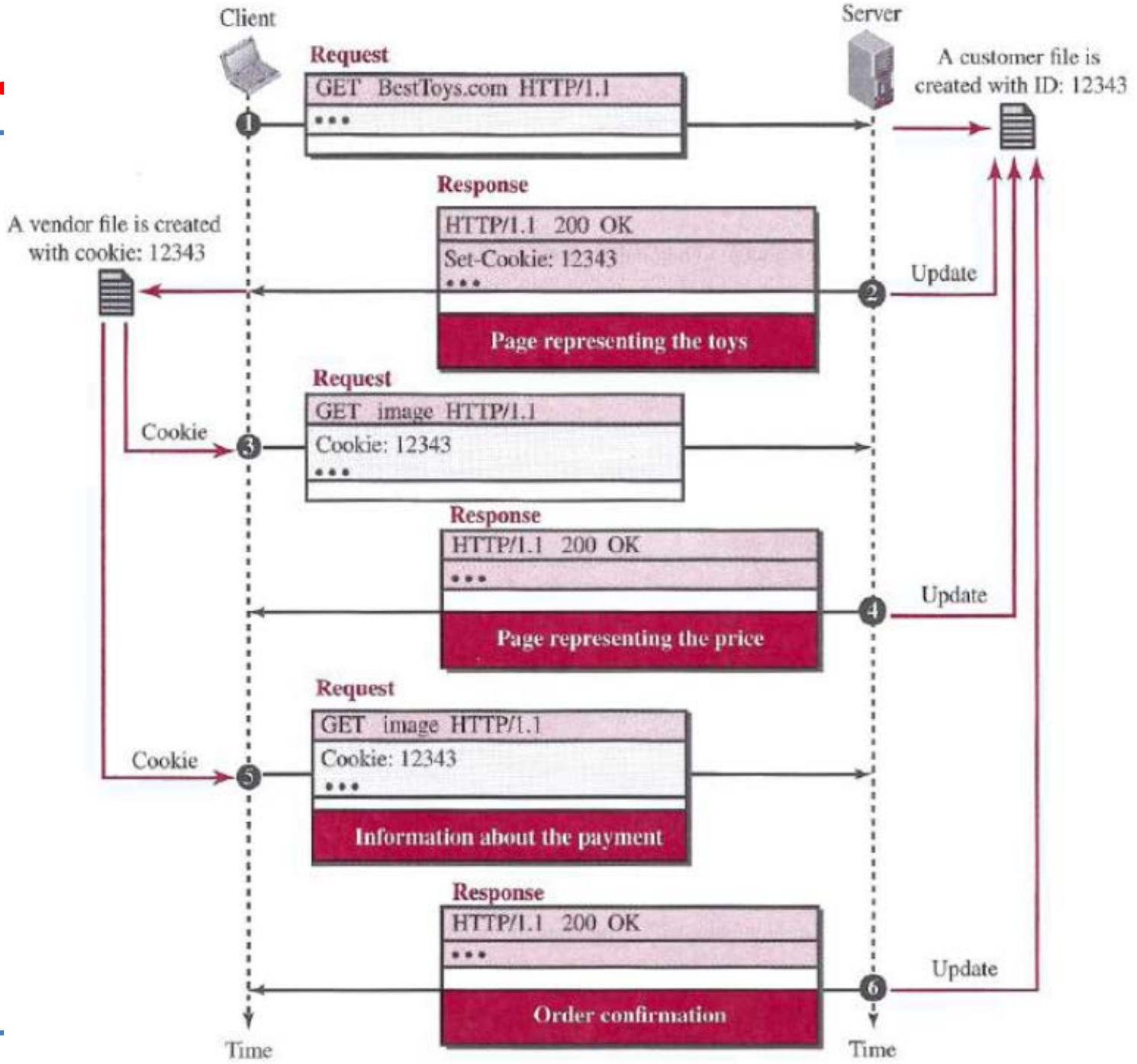
Three steps:

- When a server receives a request from a client, it stores information about the client in a file or a string.
  - The information may include the domain name of the client, the contents of the cookie (information about the client such as name, registration number, and so on), a timestamp, and other information depending on the implementation.
- The server includes the cookie in the response that it sends to the client.
- When the client receives the response, the browser stores the cookie in the cookie directory, which is sorted by the server domain name.

Using Cookies

- When a client sends a request to a server, the browser looks in the cookie directory to see if it can find a cookie sent by that server.
- If found, the cookie is included in the request.
- When the server receives the request, it knows that this is an old client, not a new one.
- Note that the contents of the cookie are never read by the browser or disclosed to the user.
- It is a cookie *made* by the server and *eaten* by the server.

Cont..


Client
Server
Request
GET BestToys.com HTTP/1.1
A customer file is created with ID: 12343
Response
A vendor file is created with cookie: 12343
HTTP/1.1 200 OK
Set-Cookie: 12343
Update
Page representing the toys
Request
Cookie
GET image HTTP/1.1
Cookie: 12343
Response
HTTP/1.1 200 OK
Update
Page representing the price
Request
Cookie
GET image HTTP/1.1
Cookie: 12343
Information about the payment
Response
HTTP/1.1 200 OK
Update
Order confirmation
Time
Time

Example

- *electronic store* (e-commerce): When a client selects an item and inserts it in a cart, a cookie that contains information about the item, such as its number and unit price, is sent to the browser. If the client selects a second item, the cookie is updated with the new selection information, and so on.

- *registered clients:* sends a cookie to the client when the client registers for the first time. For any repeated access, only those clients that send the appropriate cookie are allowed.

- *web portal:* When a user selects her favourite pages, a cookie is made and sent. If the site is accessed again, the cookie is sent to the server to show what the client is looking for.

- *advertising agencies*: When a user visits the main website and clicks the icon of a corporation, a request is sent to the advertising agency. The advertising agency sends the requested banner, but it also includes a cookie with the ID of the user.

Web Caching: Proxy Servers

- HTTP supports proxy servers.
- A proxy server is a computer that keeps copies of responses to recent requests.
  - The HTTP client sends a request to the proxy server.
  - The proxy server checks its cache.
  - If the response is not stored in the cache, the proxy server sends the request to the corresponding server.
  - Incoming responses are sent to the proxy server and stored for future requests from other clients.

- Advantages: reduces the load on the original server, decreases traffic, improves latency
- To use the proxy server, the client must be configured to access the proxy instead of the target server.
- The proxy servers are normally located at the client site.

Cache Update

- How long a response should remain in the proxy server before being deleted and replaced?

- Many Solutions (depending on requirement):
  - store the list of sites whose information remains the same for a while. E.g., news agency
  - add some headers to show the last modification time of the information.

HTTP Security

- HTTP per se does not provide security.

- But, HTTP can be run over the Secure Socket Layer (SSL). In this case, HTTP is referred to as HTTPS.

- HTTPS provides confidentiality, client and server authentication, and data integrity.

Thanks!