

HyperText Transfer Protocol (HTTP)

GURJOT SINGH

UG201113013

REVTI RAMAN SINGH

UG201110026



WHAT IS A PROTOCOL AND HTTP?

- Protocol: A Protocol is a standard procedure for defining and regulating communication. For example TCP, UDP, HTTP etc.
- HTTP is the foundation of data communication for the World Wide Web.
- The HTTP is the Web's application-layer protocol for transferring various forms of data between server and client like plaintext, hypertext, image, videos and sounds .

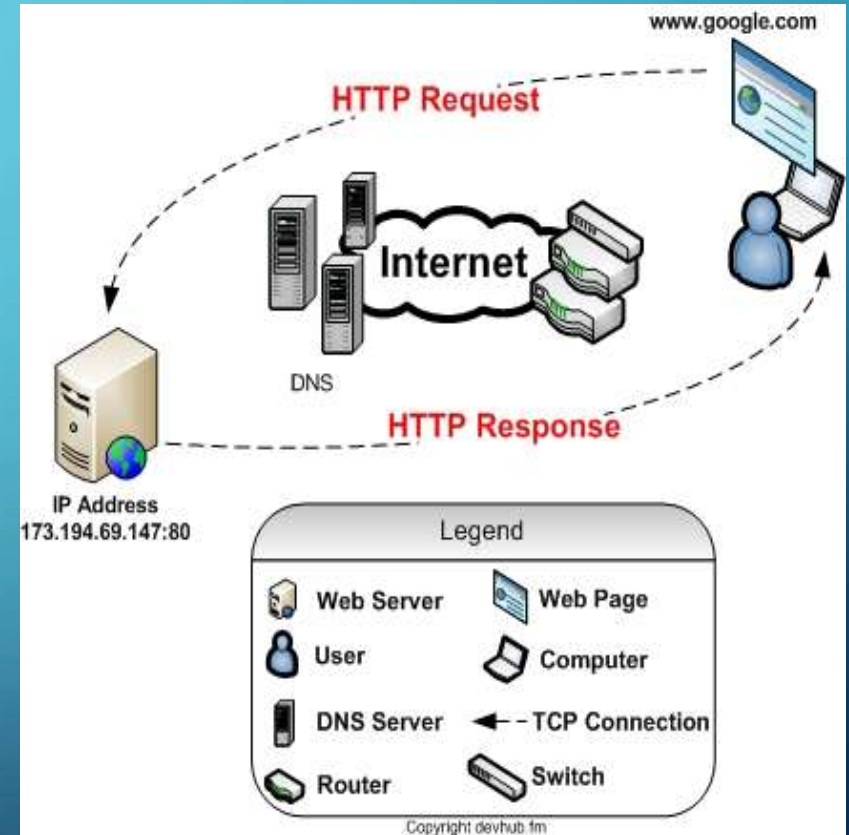


TERMINOLOGY

- IP Address: An Internet Protocol address (IP address) is a numerical label assigned to each device (e.g., computer, printer) participating in a computer network that uses the Internet Protocol for communication.
- TCP: Transmission Control Protocol (TCP) is one of the two original core protocols of the Internet Protocol Suite (IP), and is so common that the entire suite is often called *TCP/IP*. TCP provides reliable, ordered, error-checked delivery of a stream of octets between programs running on computers connected to an intranet or the public Internet.
- Port Number: A port number is a 16 bit number which when associated with IP address, completes the destination address for a communications session.
- Socket: A socket is nothing but a combination of IP address and port number. It is simply an end while communication.

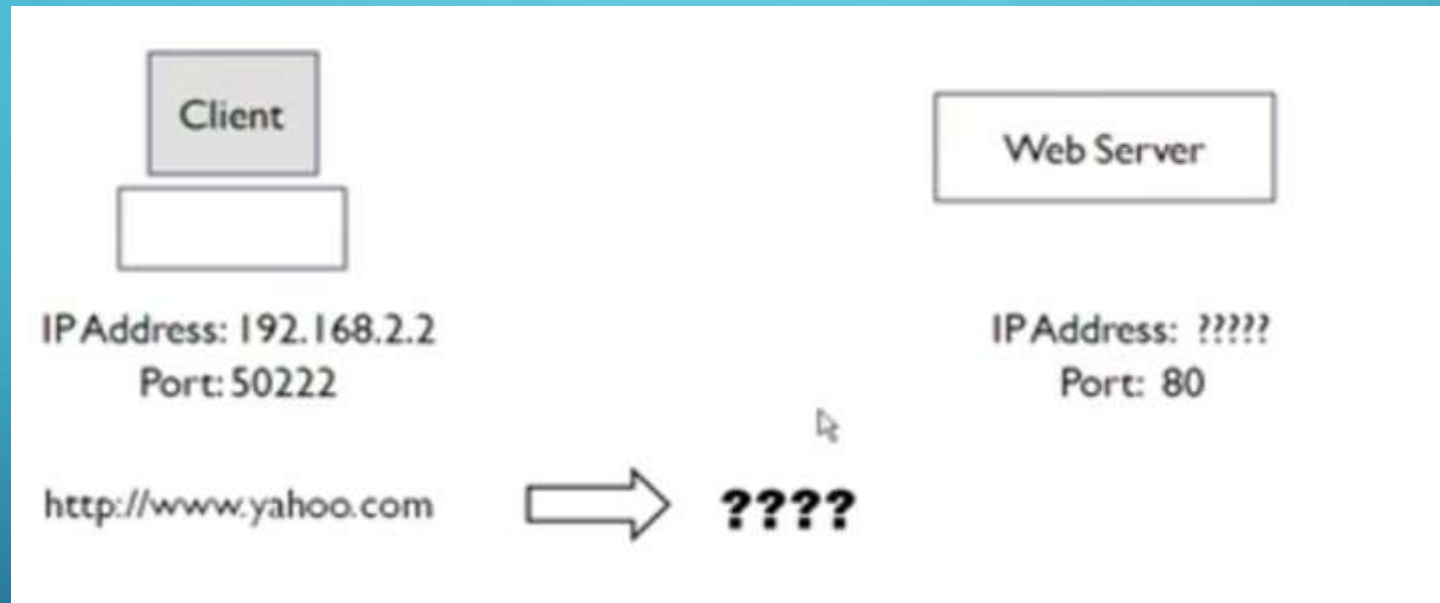
HOW HTTP WORKS?

- HTTP is implemented in two programs: a client program and a server program, executing on different end systems, talk to each other by exchanging HTTP messages.
- The HTTP client first initiates a TCP connection with the server. Once the connection is established, the browser and the server processes access TCP through their socket interfaces.



HOW HTTP WORKS?

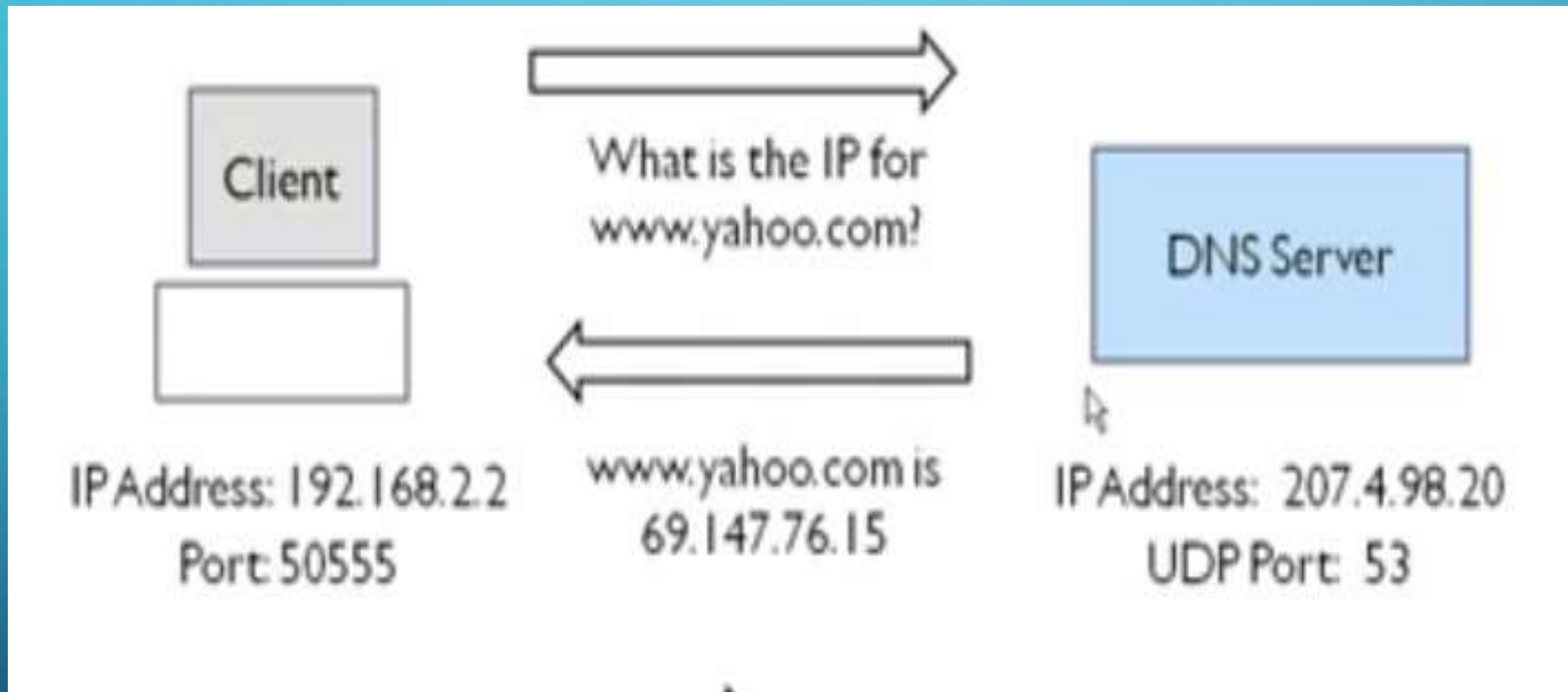
Suppose client wants to visit www.yahoo.com



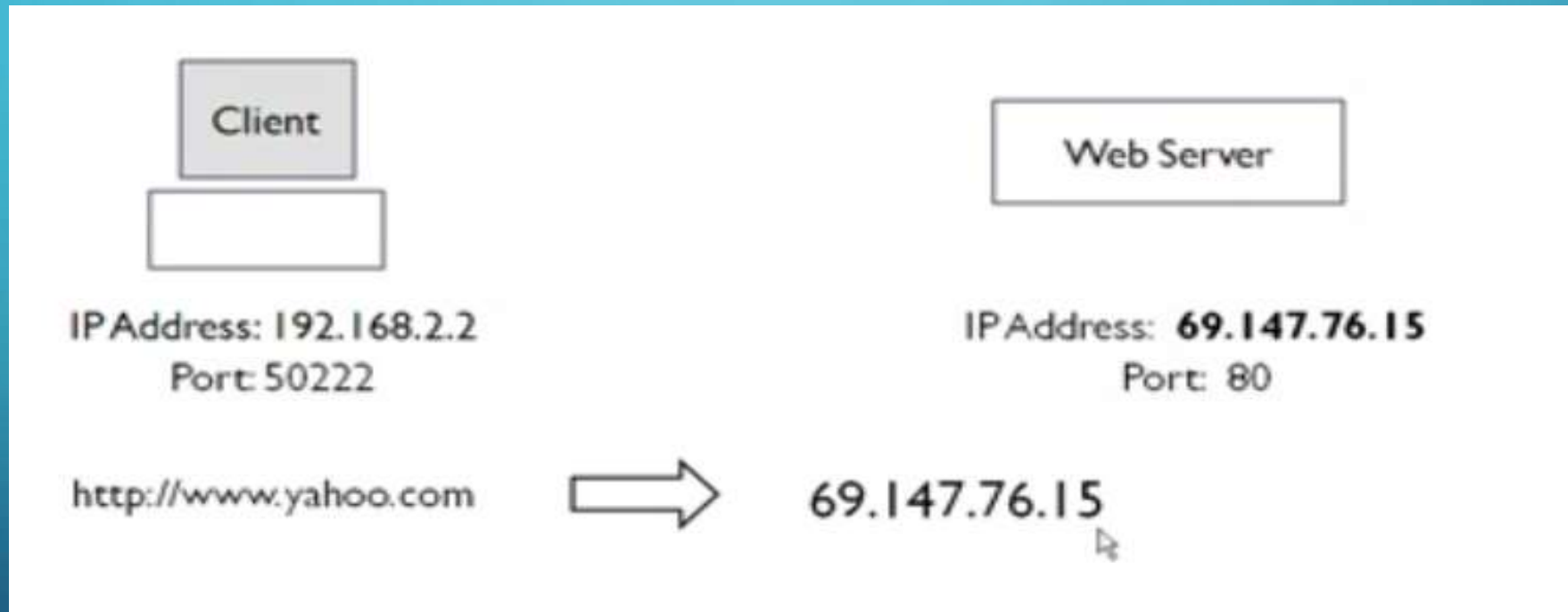
HOW HTTP WORKS? (CONT.)



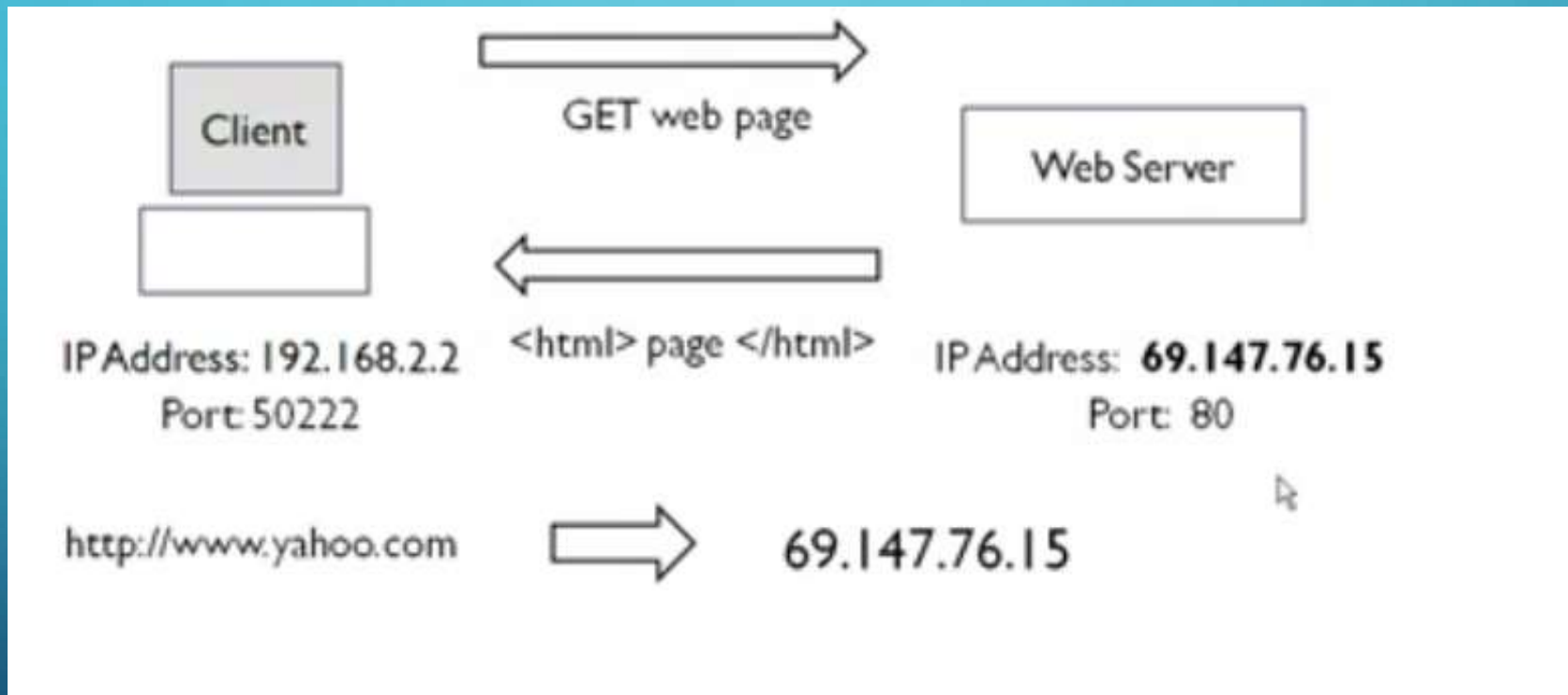
HOW HTTP WORKS? (CONT.)



HOW HTTP WORKS? (CONT.)



HOW HTTP WORKS? (CONT.)

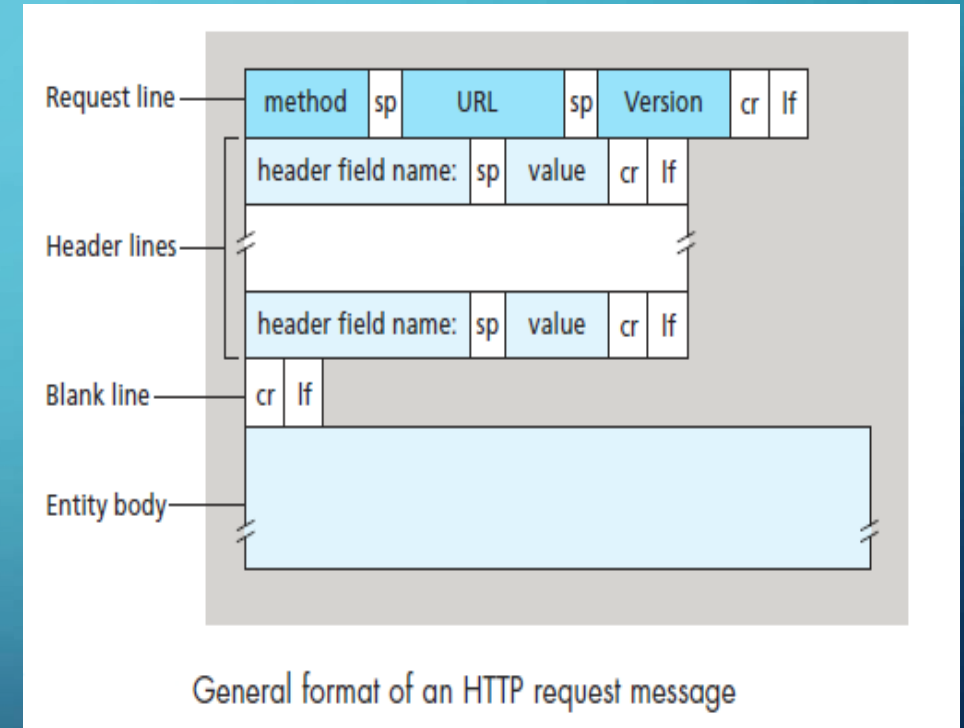


PERSISTENT AND NON-PERSISTENT CONNECTIONS

- In non-persistent connection each request/response pair are sent over a separate TCP connection.
- In persistent connections all of the requests and their corresponding responses are sent over the same TCP connection.

HTTP REQUEST MESSAGE

- The first line of an HTTP request message is called the **request line**; the subsequent lines are called the **header lines**. The request line has three fields: the method field, the URL field, and the HTTP version field. The method field can take on several different values, including GET, POST, HEAD, PUT, and DELETE etc. The great majority of HTTP request messages use the GET method. The GET method is used when the browser requests an object, with the requested object identified in the URL field.

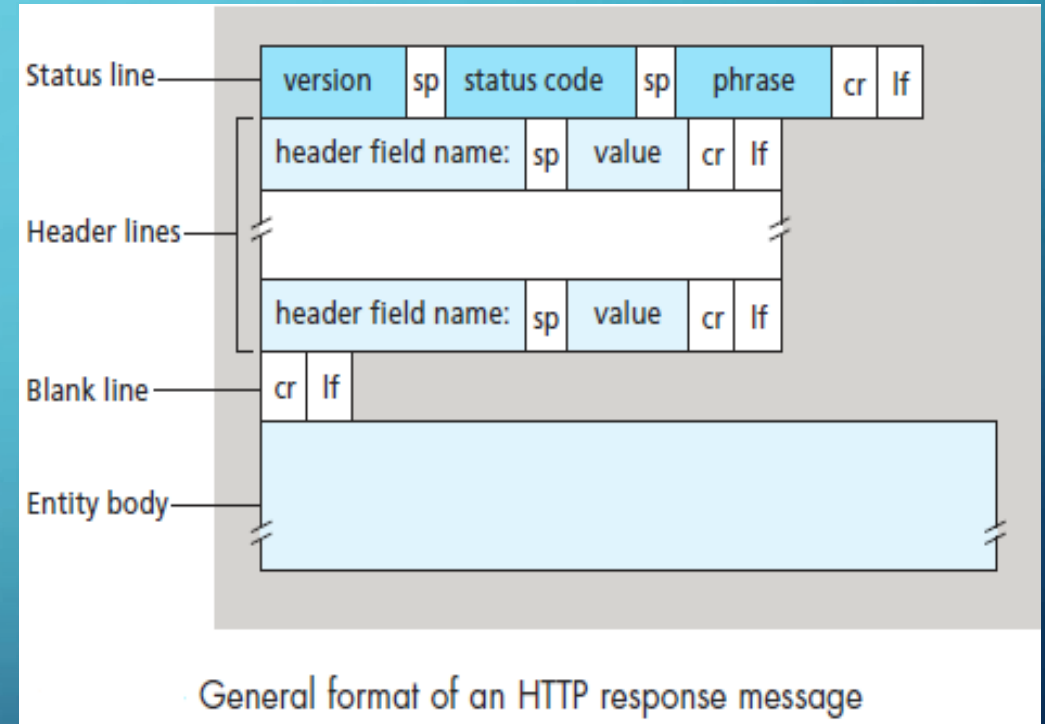


REQUEST METHODS

- GET: Retrieve Document identified in URL
- HEAD: Retrieve meta information about document identified in URL
- DELETE: Delete specified URL
- OPTIONS: Request information about available options
- PUT: Store document under specified URL
- POST: Give information to server
- TRACE: Loopback request message
- CONNECT: For use by Proxies

HTTP RESPONSE MESSAGES

- It has three sections: an initial status line, header lines, and then the entity body. The entity body contains the requested object itself. The status line has three fields: the protocol version field, a status code, and a corresponding status message.



SOME COMMON STATUS CODES AND ASSOCIATED PHRASES

- Some common status codes and associated phrases include:
 - 200 OK: Request succeeded and the information is returned in the response.
 - 301 Moved Permanently: Requested object has been permanently moved; the new URL is specified in Location: header of the response message. The client software will automatically retrieve the new URL.
 - 400 Bad Request: This is a generic error code indicating that the request could not be understood by the server.
 - 404 Not Found: The requested document does not exist on this server.
 - 505 HTTP Version Not Supported: The requested HTTP protocol version is not supported by the server.

Code	Type	Example Reasons
1xx	Informational	Request received, continuing process
2xx	Success	Action successfully received, understood, and accepted
3xx	Redirection	Further action must be taken to complete the request
4xx	Client error	Request contains bad syntax or cannot be fulfilled
5xx	Server error	Server failed to fulfill an apparently valid request
Five types of HTTP result codes.		


```
josh@blackbox:~$ telnet en.wikipedia.org 80
Trying 208.80.152.2...
Connected to rr.pmtpa.wikimedia.org.
Escape character is '^]'.
GET /wiki/Main_Page http/1.1
Host: en.wikipedia.org
```

```
HTTP/1.0 200 OK
Date: Thu, 03 Jul 2008 11:12:06 GMT
Server: Apache
X-Powered-By: PHP/5.2.5
Cache-Control: private, s-maxage=0, max-age=0, must-revalidate
Content-Language: en
Vary: Accept-Encoding, Cookie
X-Vary-Options: Accept-Encoding;list-contains=gzip, Cookie;string-contains=enwikiToken;string-contains=enwikiLoggedOut;string-contains=enwiki_session;
string-contains=centralauth Token;string-contains=centralauth_Session;string-contains=centralauth_LoggedOut
Last-Modified: Thu, 03 Jul 2008 10:44:34 GMT
Content-Length: 54218
Content-Type: text/html; charset=utf-8
X-Cache: HIT from sq39.wikimedia.org
X-Cache-Lookup: HIT from sq39.wikimedia.org:3128
Age: 3
X-Cache: HIT from sq38.wikimedia.org
X-Cache-Lookup: HIT from sq38.wikimedia.org:80
Via: 1.0 sq39.wikimedia.org:3128 (squid/2.6.STABLE18), 1.0 sq38.wikimedia.org:80 (squid/2.6.STABLE18)
Connection: close
```

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" lang="en" dir="ltr">
  <head>
    <meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
    <meta name="keywords" content="Main Page,1778,1844,1863,1938,1980 Summer Olympics,2008,2008 Guizhou riot,2008 Jerusal
    ...
    ... This content has been removed to save space ...
    "Non-profit organization">nonprofit</a> <a href="http://en.wikipedia.org/wiki/Charitable_organization" title="Charitable organization">charity</a>.<b
    r /></li>
    <li id="privacy"><a href="http://wikimediafoundation.org/wiki/Privacy_policy" title="wikimedia:Privacy policy">Privac
    y policy</a></li>
    <li id="about"><a href="/wiki/Wikipedia:About" title="Wikipedia:About">About Wikipedia</a></li>
    <li id="disclaimer"><a href="/wiki/Wikipedia:General_disclaimer" title="Wikipedia:General disclaimer">Disclaimers</a>
  </li>
  </ul>
</div>
  <script type="text/javascript">if (window.runOnLoadHook) runOnLoadHook();</script>
<!-- Served by srv93 in 0.050 secs. --></body></html>
Connection closed by foreign host.
josh@blackbox:~$
```

Escape character is '^]'.
Request

GET /wiki/Main_Page http/1.1
Host: en.wikipedia.org
Response headers

HTTP/1.0 200 OK
Response body

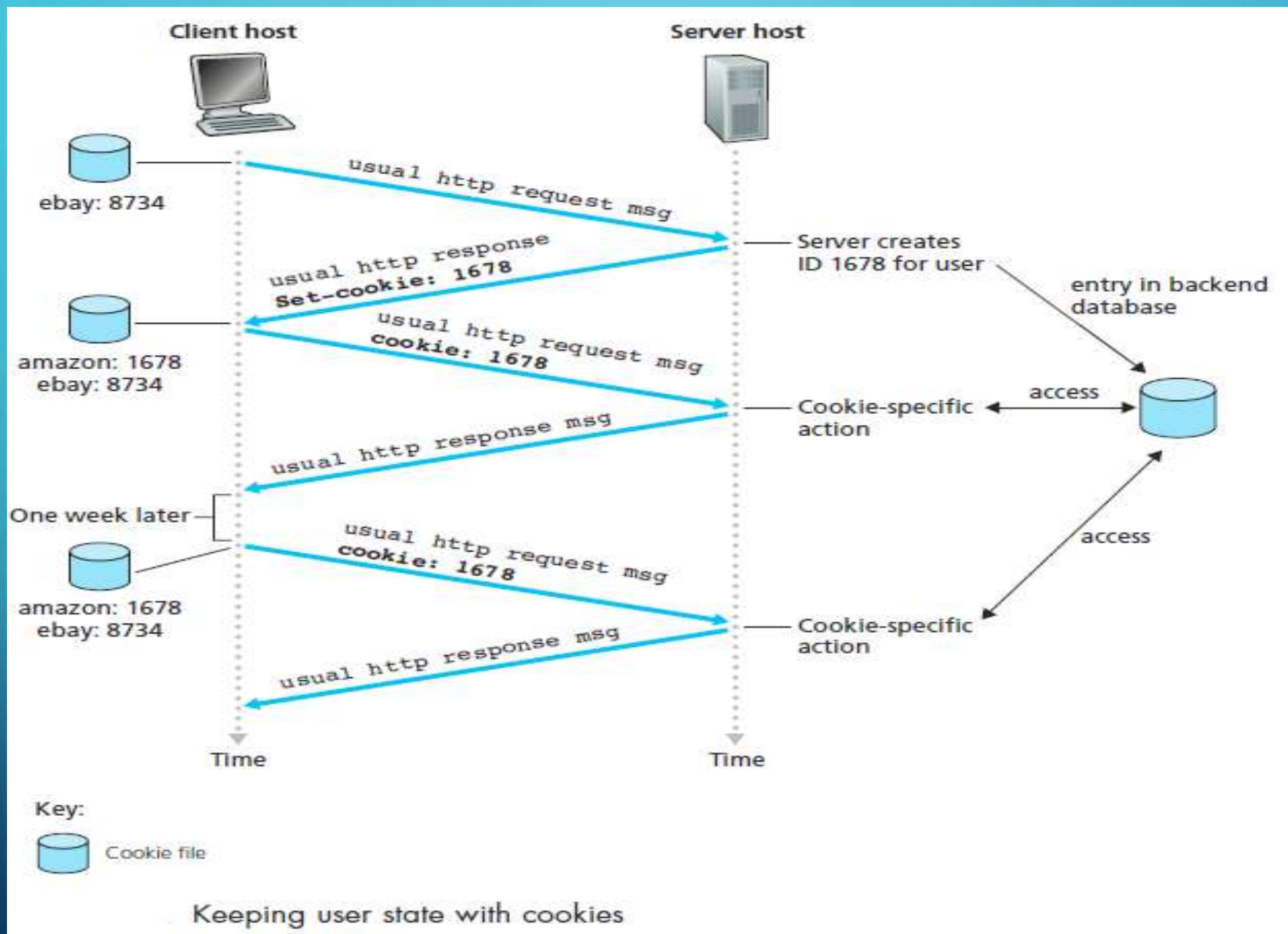
An HTTP request using telnet

HTTP: A STATELESS PROTOCOL

- HTTP is a stateless protocol, because an HTTP server maintains no information about the clients.
- If a particular client asks for the same object twice in a period of a few seconds, the server does not respond by saying that it just served the object to the client; instead, the server resends the object, as it has completely forgotten what it did earlier.

USER-SERVER INTERACTIONS: COOKIES

- HTTP server being stateless, simplifies server design and has permitted engineers to develop high-performance Web servers that can handle thousands of simultaneous TCP connections. However, it is often desirable for a Web site to identify users, either because the server wishes to restrict user access or because it wants to serve content as a function of the user identity. For these purposes, HTTP uses cookies. Cookies allow sites to keep track of users.
- The cookie technology has four components:
 - a cookie header line in the HTTP response message
 - a cookie header line in the HTTP request message
 - a cookie file kept on the user's end system and managed by the user's browser
 - a back-end database at the Web site



WEB CACHING

- Also called proxy server, is a network entity that satisfies HTTP requests on the behalf of an origin Web server.
- The Web cache has its own disk storage and keeps copies of recently requested objects in this storage.
- A Web cache can substantially reduce the response time for a client request, particularly if the bandwidth between the client and the origin server is much less than the bandwidth between the client and the cache.
- Web caches can substantially reduce traffic

