WEB TECHNOLOGY

Chapter 1: WEB ESSENTIALS

Lecture Six

Compiled by:

Suwal@serij.com.np

Quick Lookback of Previous Class:

- Web 1.0
- Web 2.0
- Web 3.0
- & comparison among those topics

Topics for Today

- HTTP
- HTTP Request and Response Messages;
- & features

  - Remaining Slide referenced from: https://www.slideshare.net/ChuongMai/http-protocol-basic

  - Additional Information can be found on:
  - https://developer.mozilla.org/en-US/docs/Web/HTTP/Overview

VGU BIS 2012 – Computer Networks

HTTP Protocol

Mai Nam Chuong

Instructor: Prof. Dr. Friedbert Kaspar

Agenda

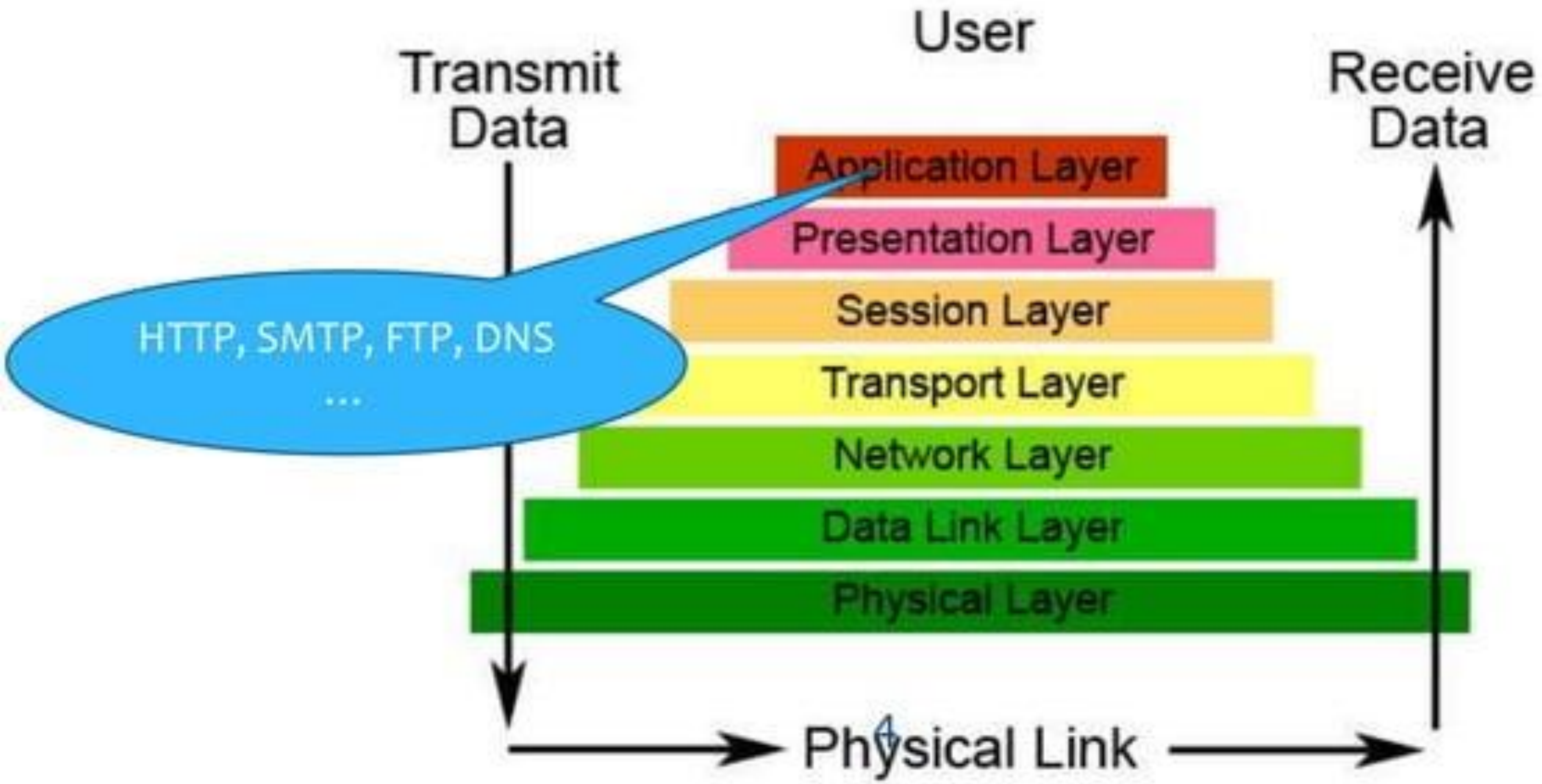
- Definitions
- HTTP Communication chain
- HTTP Transport binding
- HTTP Methods
- HTTP Messages
- HTTP Status Codes
- HTTP Caching
- Q&A

Communication Protocol?

- A set of rules and regulations that determine how data is transmitted.

7 OSI layers

The Seven Layers of OSI



User

Transmit Data

Receive Data

Application Layer

Presentation Layer

Session Layer

Transport Layer

Network Layer

Data Link Layer

Physical Layer

Physical Link

HTTP, SMTP, FTP, DNS

…

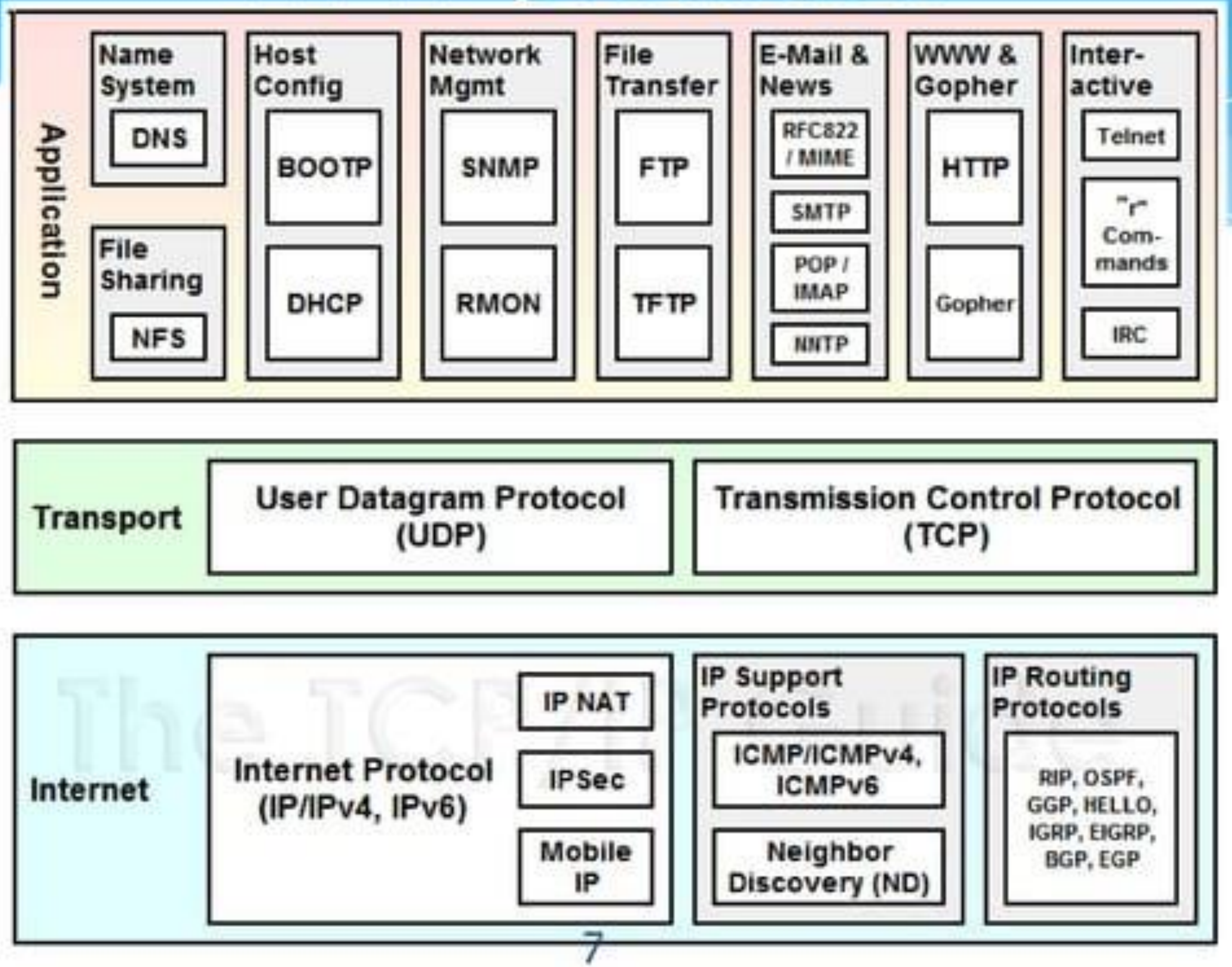
Application layer?

- Application vs. Application layer
  - Application layer provides services for user applications to use. Or,
  - User applications use protocols that reside at the application layer.
  - Ex: Internet Explorer – HTTP; Outlook – SMTP

HTTP protocol?

- Hypertext Transfer Protocol
  - an application-level protocol for distributed, collaborative, hypermedia information systems (RFC-2616)
  - Not only text but graphics and multimedia
  - A part of TCP/IP protocols suite

HTTP protocol?



Application

Name System
- DNS

File Sharing
- NFS

Host Config
- BOOTP
- DHCP

Network Mgmt
- SNMP
- RMON

File Transfer
- FTP
- TFTP

E-Mail & News
- RFC822 / MIME
- SMTP
- POP / IMAP
- NNTP

WWW & Gopher
- HTTP
- Gopher

Inter-active
- Telnet
- "r" Com-mands
- IRC

Transport

- User Datagram Protocol (UDP)
- Transmission Control Protocol (TCP)

Internet

Internet Protocol (IP/IPv4, IPv6)
- IP NAT
- IPSec
- Mobile IP

IP Support Protocols
- ICMP/ICMPv4, ICMPv6
- Neighbor Discovery (ND)

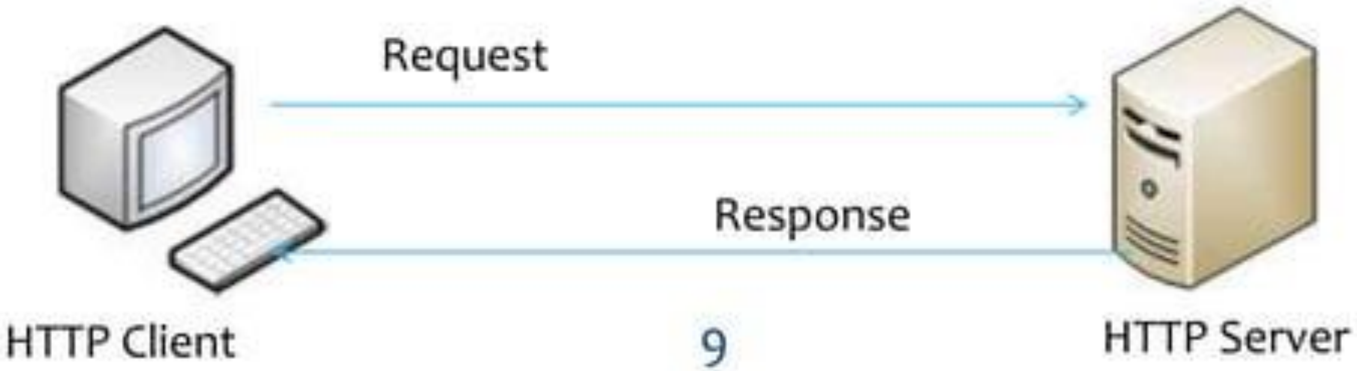
IP Routing Protocols
- RIP, OSPF, GGP, HELLO, IGRP, EIGRP, BGP, EGP

HTTP protocol?

- History
  - 1990 – HTTP/0.9
  - 1996 – HTTP/1.0
  - 1999 – HTTP/1.1
  - Next HTTP/2.0 (draft)

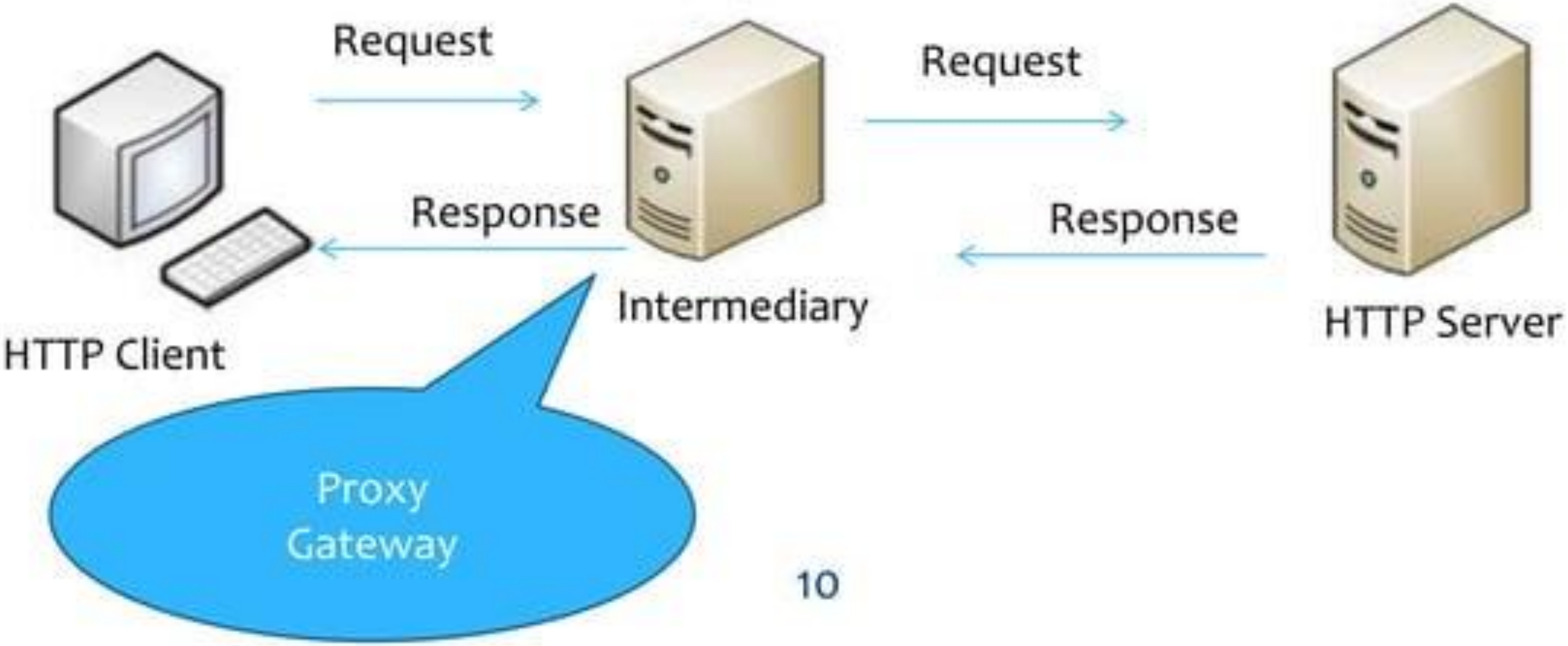
HTTP Communication chain

- The HTTP protocol is a request/response protocol
- A client sends a request to the server
- The server responds data



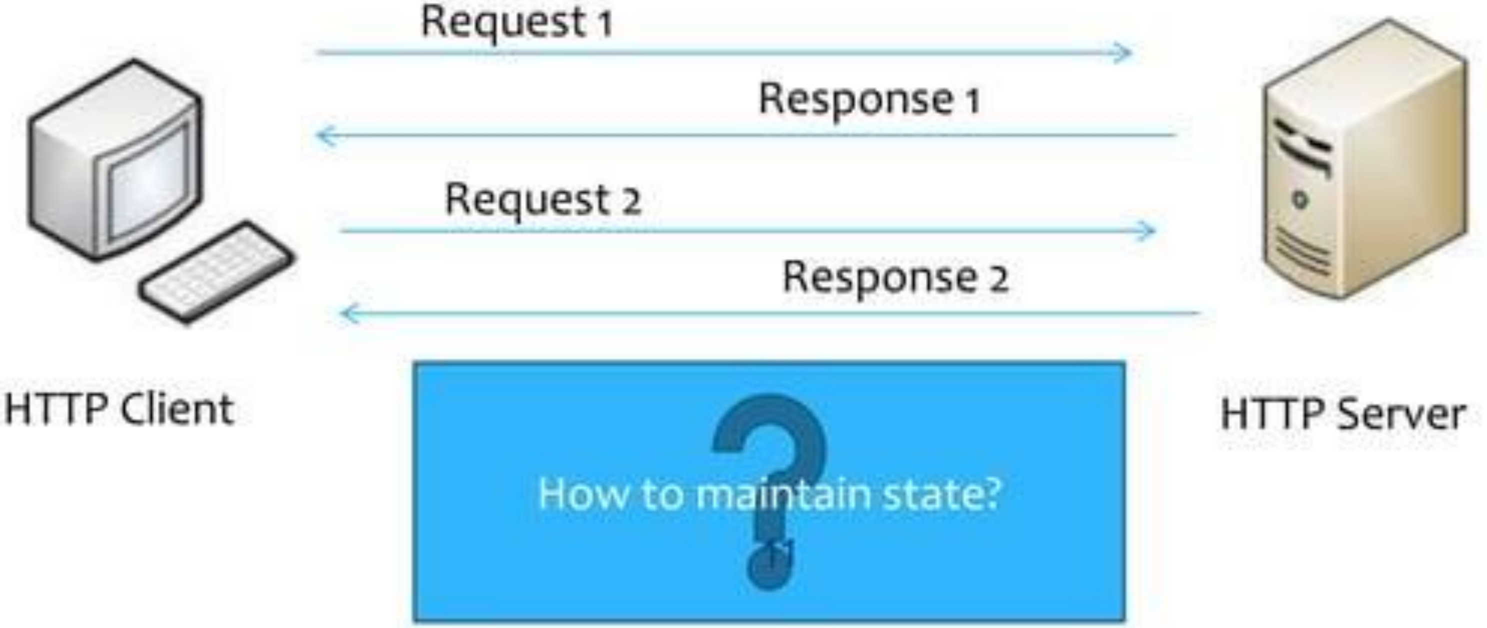
HTTP Communication chain

- HTTP Request/Response Chain with Intermediaries



HTTP stateless protocol

- Do not remember previous request/response chain



Request 1

Response 1

Request 2

Response 2

HTTP Client

HTTP Server

How to maintain state?

HTTP Transport binding

- Transitory connection
  - One connection is created for each request/response
- Persistent connection
  - The connection is kept open after each request/response set until the client is done requesting all the documents.

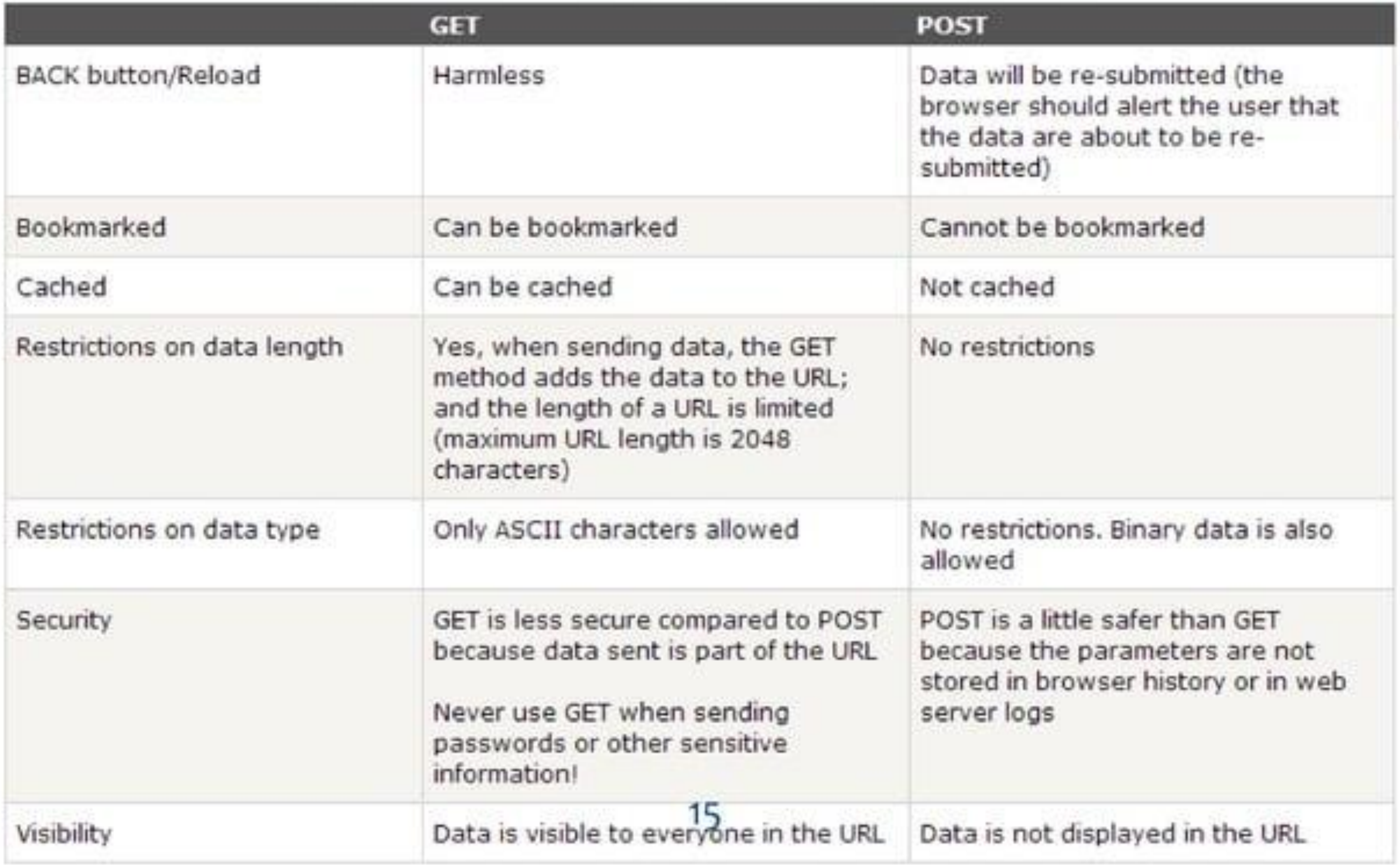
HTTP Methods

- **GET (*)**
- **HEAD (*)**
- **POST (*)**
- **OPTIONS**
- **PUT**
- **DELETE**
- **TRACE**
- **CONNECT**

HTTP GET vs. POST

- **GET** - Requests data from a specified resource
- **POST** - Submits data to be processed to a specified resource

HTTP GET vs. POST

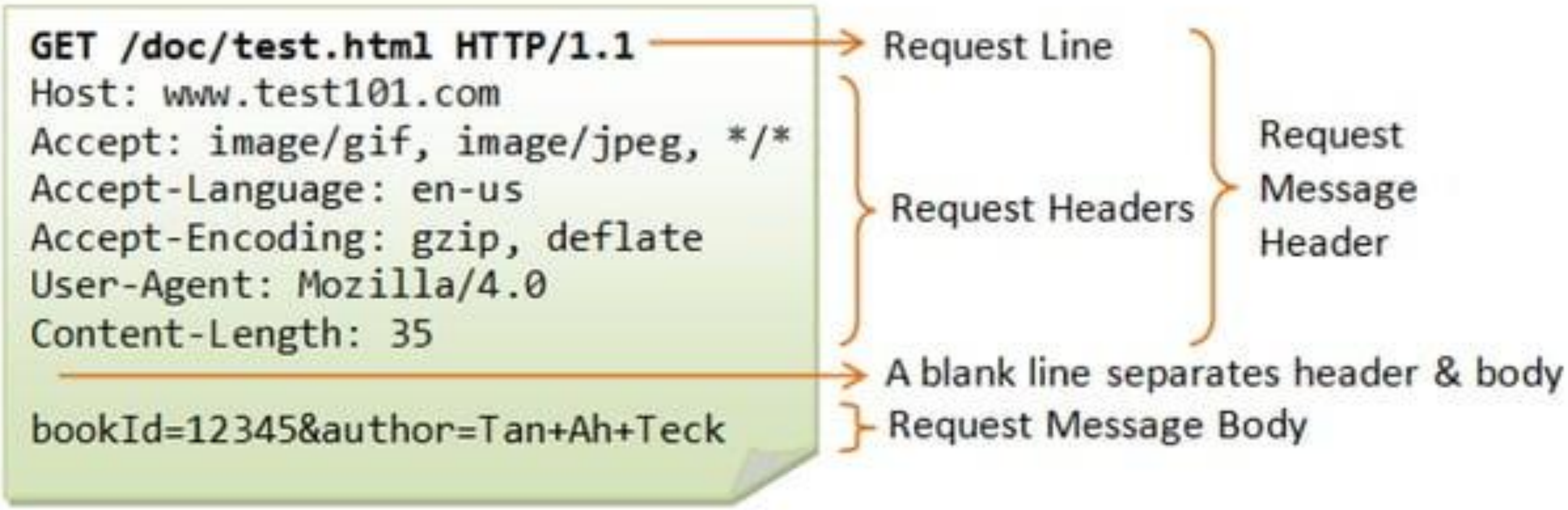


	GET	POST
BACK button/Reload	Harmless	Data will be re-submitted (the browser should alert the user that the data are about to be re-submitted)
Bookmarked	Can be bookmarked	Cannot be bookmarked
Cached	Can be cached	Not cached
Restrictions on data length	Yes, when sending data, the GET method adds the data to the URL; and the length of a URL is limited (maximum URL length is 2048 characters)	No restrictions
Restrictions on data type	Only ASCII characters allowed	No restrictions. Binary data is also allowed
Security	GET is less secure compared to POST because data sent is part of the URL Never use GET when sending passwords or other sensitive information!	POST is a little safer than GET because the parameters are not stored in browser history or in web server logs
Visibility	Data is visible to everyone in the URL	Data is not displayed in the URL

HTTP Message

- Communication between devices using the HTTP takes place via *HTTP messages*
- Two types: *requests* and *responses*.
- Text-based message format.

HTTP Request Message



```
GET /doc/test.html HTTP/1.1
Host: www.test101.com
Accept: image/gif, image/jpeg, */*
Accept-Language: en-us
Accept-Encoding: gzip, deflate
User-Agent: Mozilla/4.0
Content-Length: 35

bookId=12345&author=Tan+Ah+Teck
```

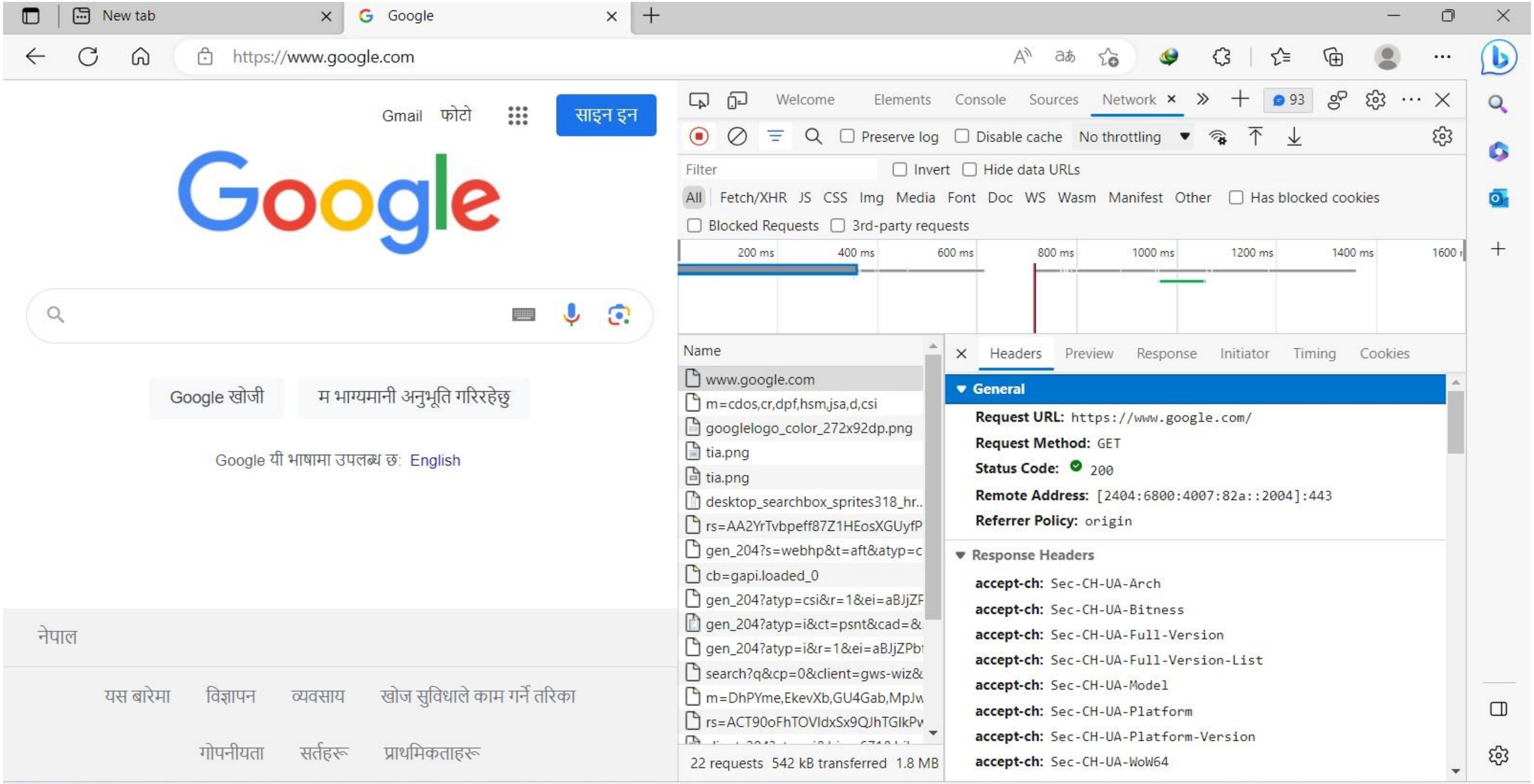
Request Line

Request Headers

Request Message Header

A blank line separates header & body

Request Message Body



www.google.com
m=cdos,cr,dpf,hsm,jsa,d,csi
googlelogo_color_272x92dp.png
tia.png
tia.png
desktop_searchbox_sprites318_hr..
rs=AA2YrTvbpeff87Z1HEosXGUyfP
gen_204?s=webhp&t=aft&atyp=c
cb=gapi.loaded_0
gen_204?atyp=csi&r=1&ei=aBJjZF
gen_204?atyp=i&ct=psnt&cad=&
gen_204?atyp=i&r=1&ei=aBJjZPbi
search?q&cp=0&client=gws-wiz&
m=DhPYme,EkevXb,GU4Gab,MpJw
rs=ACT90oFhTOVIdxSx9QJhTGIkPv

22 requests 542 kB transferred 1.8 MB

▼ General

Request URL: `https://www.google.com/`

Request Method: GET

Status Code: ✓ 200

Remote Address: [2404:6800:4007:82a::2004]:443

Referrer Policy: origin

▼ Response Headers

```
accept-ch: Sec-CH-UA-Arch
accept-ch: Sec-CH-UA-Bitness
accept-ch: Sec-CH-UA-Full-Version
accept-ch: Sec-CH-UA-Full-Version-List
accept-ch: Sec-CH-UA-Model
accept-ch: Sec-CH-UA-Platform
accept-ch: Sec-CH-UA-Platform-Version
accept-ch: Sec-CH-UA-WoW64
```

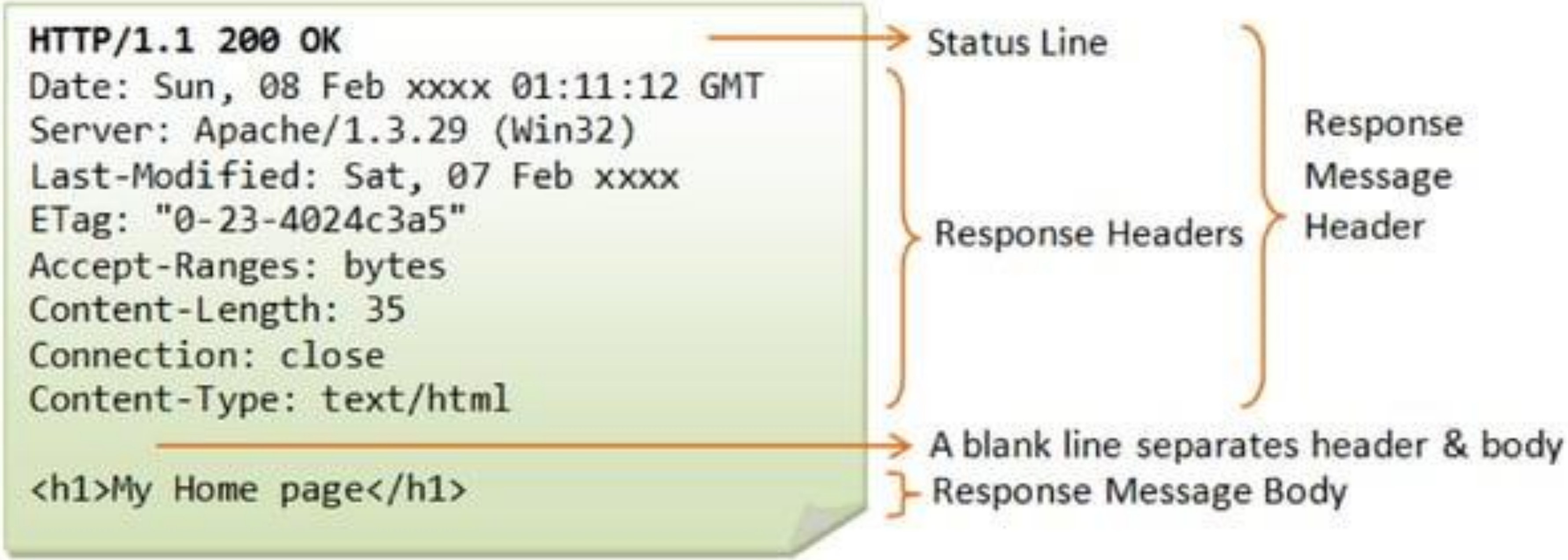
HTTP Request Headers

- **Accept:**
  - tell the server what Internet media types it is willing to accept in a response
  - Accept: type/subtype; text/plain; image/jpg; */*
- **Accept-Charset**
  - specifies what character sets the client is willing to accept in a response
  - Accept-Charset: iso-8859-5, UTF-8
- **Accept-Encoding**
  - specifies what content encodings the client is willing to accept
  - Accept-Encoding: compress, gzip
- **Accept-Language**
  - indicate what languages the client supports
  - Accept-Language: en-gb; vi-vn

HTTP Request Headers (cont.)

- **Host:**
  - Specifies the Internet host
  - Host: www.vgu.edu.vn:8080
- **User-Agent**
  - Specifies what agent (application) is sending the request
  - User-Agent: Mozilla/5.0
- **Cookie**
  - Cookie data which previously set by the server
  - Cookie: user=1; category=pc;
- **Referer**
  - the URL of the resource from which the URL of the current request was obtained
  - Referer: http://google.com

HTTP Response Message



```
HTTP/1.1 200 OK
Date: Sun, 08 Feb xxxx 01:11:12 GMT
Server: Apache/1.3.29 (Win32)
Last-Modified: Sat, 07 Feb xxxx
ETag: "0-23-4024c3a5"
Accept-Ranges: bytes
Content-Length: 35
Connection: close
Content-Type: text/html

<h1>My Home page</h1>
```

HTTP Response Headers

- **Server**
  - Server info that serves the request
  - Server: Apache/2.4.1 (Unix)
- **Expires**
  - gives the date/time after which the response is considered
  - Expires: Thu, 20 Jul 2013 16:00:00 GMT
- **Refresh**
  - Used in redirection, or when a new resource has been created.
  - Refresh: 5; url=http://www.vgu.edu.vn

HTTP Status codes

- **1xx (Informational):** Request received, server is continuing the process.
- **2xx (Success)**: The request was successfully received, understood, accepted and serviced.
- **3xx (Redirection)**: Further action must be taken in order to complete the request.
- **4xx (Client Error)**: The request contains bad syntax or cannot be understood.
- **5xx (Server Error)**: The server failed to fulfill an apparently valid request.

Common HTTP Status codes

- **200 OK:** The request is fulfilled.
- **301 Move Permanently:** The resource requested for has been permanently moved to a new location
- **304 Not Modified:** Server notifies that the resource requested has not been modified.

Common HTTP Status codes

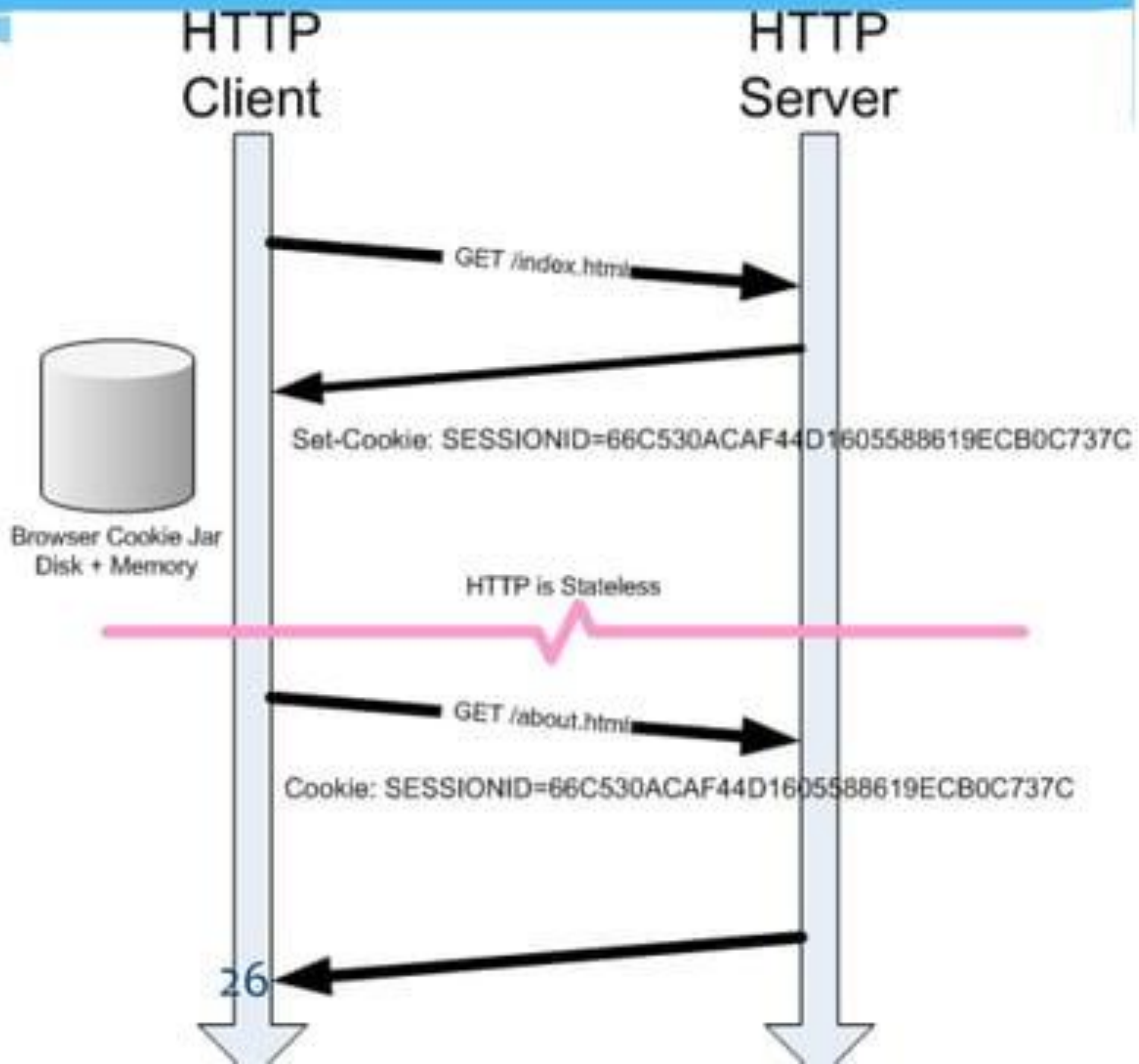
- **400 Bad Request**: Server could not understand the request (syntax error).
- **401 Authentication Required**: The requested resource is protected, and require client's credential.
- **403 Forbidden**: Server refuses to supply the resource, regardless of identity of client.
- **404 Not Found**: The requested resource cannot be found in the server.

Common HTTP Status codes

- **500 Internal Server Error:** Server is error in the server-side program responding to the request.
- **503 Service Unavailable:** Server cannot response due to overloading or maintenance. The client can try again later.
- **504 Gateway Timeout:** Proxy or Gateway indicates that it receives a timeout from an upstream server.

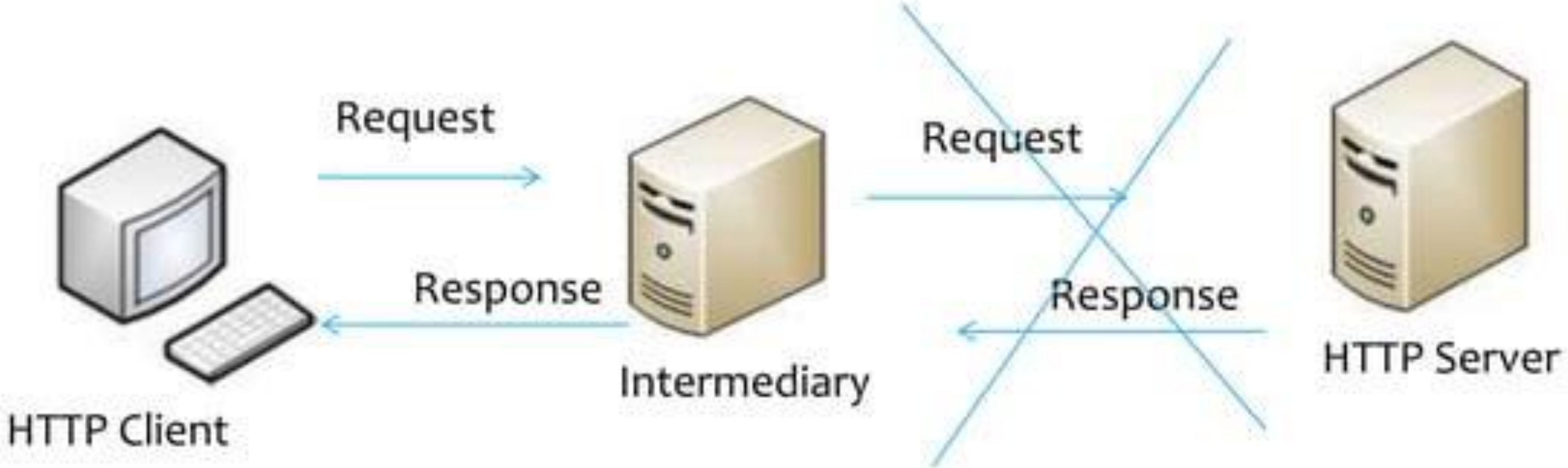
HTTP Cookie

- HTTP is a stateless protocol. Cookies provide a mechanism to "maintain state".


HTTP
Client
HTTP
Server
GET /index.html
Set-Cookie: SESSIONID=66C530ACAF44D1605588619ECB0C737C
Browser Cookie Jar
Disk + Memory
HTTP is Stateless
GET /about.html
Cookie: SESSIONID=66C530ACAF44D1605588619ECB0C737C

HTTP Caching

- Cache: local copy of a resource



HTTP Caching

- Expiration Model
  - How long a resource is valid?
  - Eliminate requests to server
  - Reduce network round-trip
  - Cache-Control: max-age=3600
  - Expires: Tue, 15 Aug 2013 07:19:00 GMT

HTTP Caching

- Validation Model
  - When the resource last changed?
  - Eliminate full response to client
  - Reduce network bandwidth
  - If-Modified-Since: Thu, 17 Aug 2006 17:53:51 GMT
  - ETag: "686897696a7c876b7e"

References

- RFC-2616: http://www.w3.org/Protocols/rfc2616/rfc2616.html
- TCP GUIDE: http://www.tcpipguide.com/free/t_TCPIPHypertextTransferProtocolHTTP.htm
- IBM Software Information Center - The HTTP protocol http://publib.boulder.ibm.com/infocenter/cicsts/v3r1/topic/com.ibm.cics.ts31.doc/dfhtl/topics/dfhtl29.htm

Q&A