4

CS 348 Computer Networks

Spring 2019

Instructor: Neha Karanjkar

The Web

- **WorldWideWeb (WWW)**: A set of **interlinked** web-pages.
- Invented by Sir Tim Berners-Lee, who created the first web-browser in 1990 at CERN.
- Client-server architecture:
  - A Client processe (web broswer) requests web-pages from a Server and displays them
- A Web Page can consists of several **objects**:
  - A base HTML file
  - Referenced objects such as images, audio file, Java applet etc.

The Web

- Key Components:
  1. A way to uniquely address each object that makes up a webpage: **URL** (Uniform Resource Locator)
  2. A Protocol for communication between the Client and Server Processes: **HTTP** (Hyper-Text Transfer Protocol)
  3. A standard for formatting and displaying the hyperlinked text to the end-user: **HTML** (Hyper-Text Markup Language)

Structure of a URL

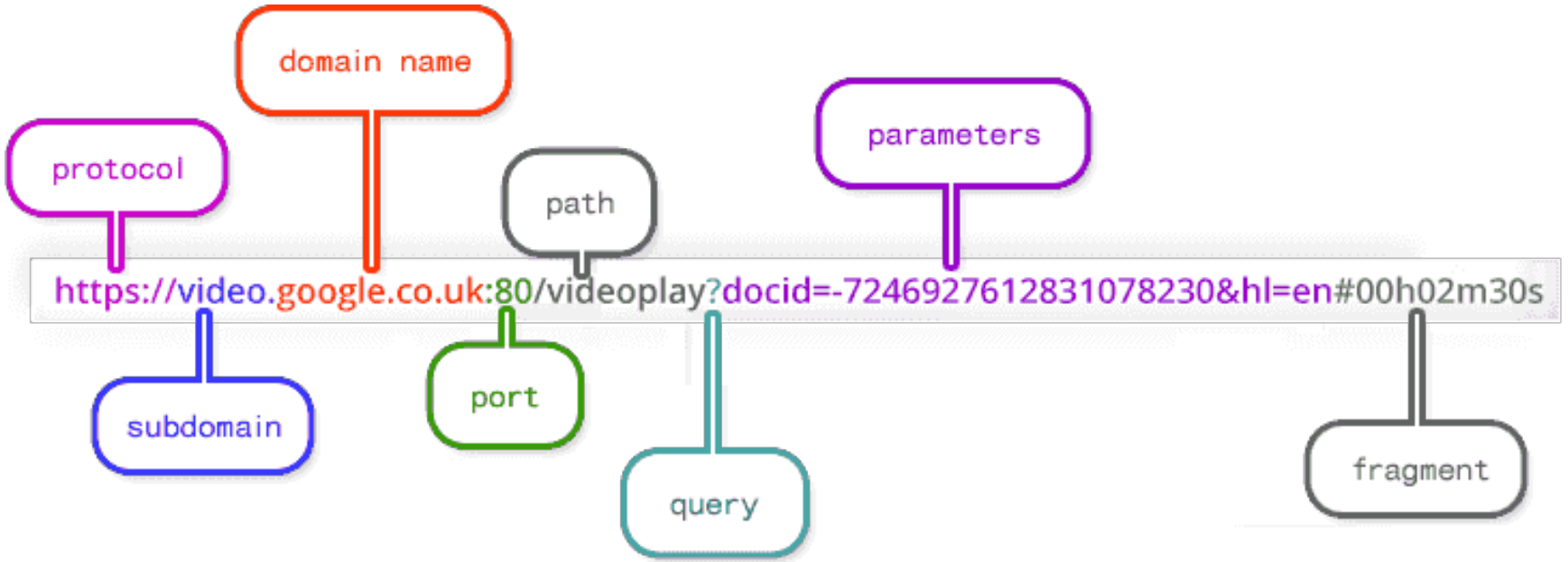


Image Source: https://doepud.co.uk/blog/anatomy-of-a-url

HTTP (Hyper-Text Transfer Protocol)

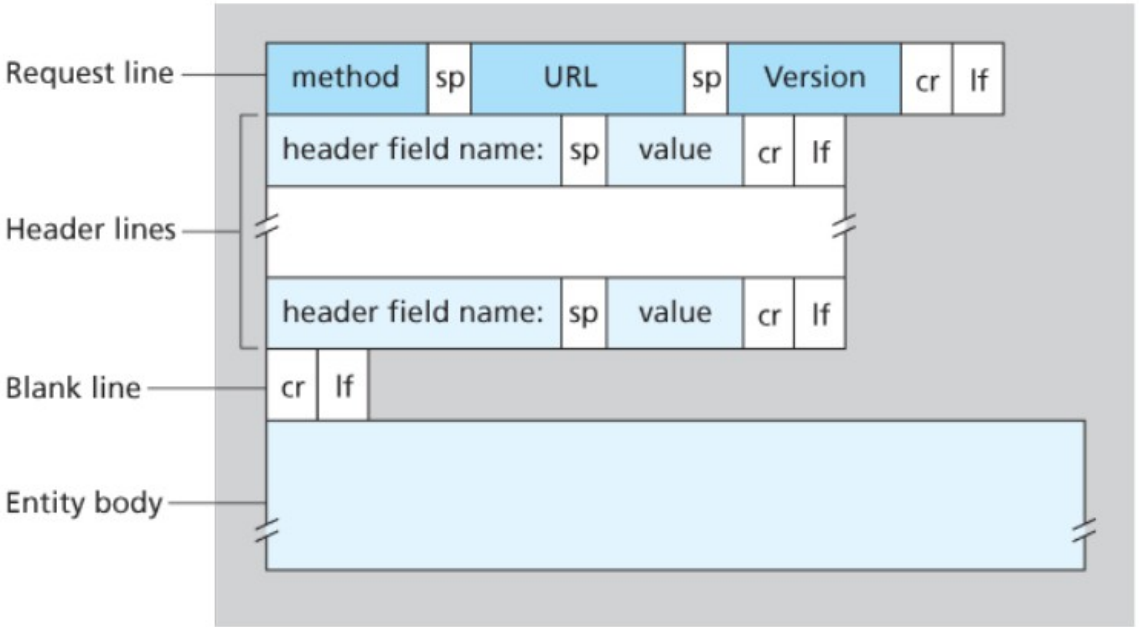
- Client sends a REQUEST and server sends back a RESPONSE.
- Both consist of simple, ASCII encoded text
- Uses TCP underneath (for reliable, in-order delivery)
- **An HTTP server is STATELESS**
  - Server treats each request independently, assuming no relation between successive requests
  - Server does not keep track of the Client's state

    Advantages of a stateless protocol...

HTTP (Hyper-Text Transfer Protocol)

- **Non-Persistent HTTP:**
  - Exactly one request-response pair per TCP connection. Then the connection is closed.
- **Persistent HTTP:**
  - Multiple requests sent over a single TCP connection. The connection is kept open by the server and client for a certain amount of time.
  - Reduces TCP connection overhead.
- Default behavior for most browsers:
  - Persistent HTTP
  - Use Pipelining and Parallelism in addition.

Format of an HTTP Request

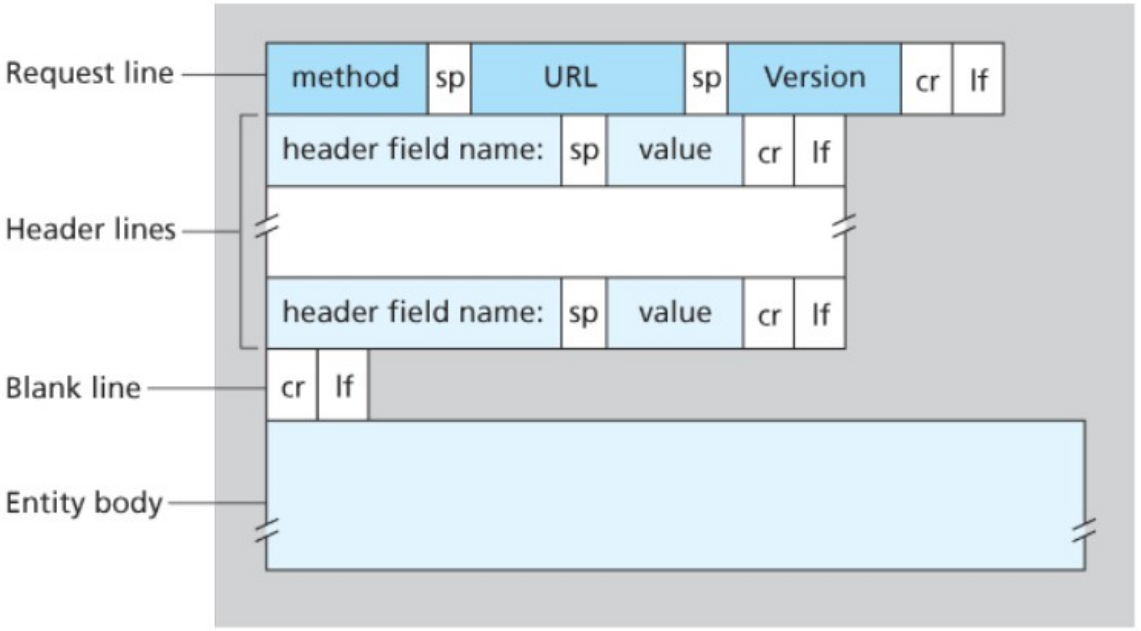


- **Example:**

```
GET /index.html HTTP/1.1\r\n
Host: www-net.cs.umass.edu\r\n
User-Agent: Firefox/3.6.10\r\n
Accept: text/html\r\n
Accept-Language: en-us,en;q=0.5\r\n
Keep-Alive: 115\r\n
Connection: keep-alive\r\n
\r\n
```

Image Source: Kurose and Ross 7th edition

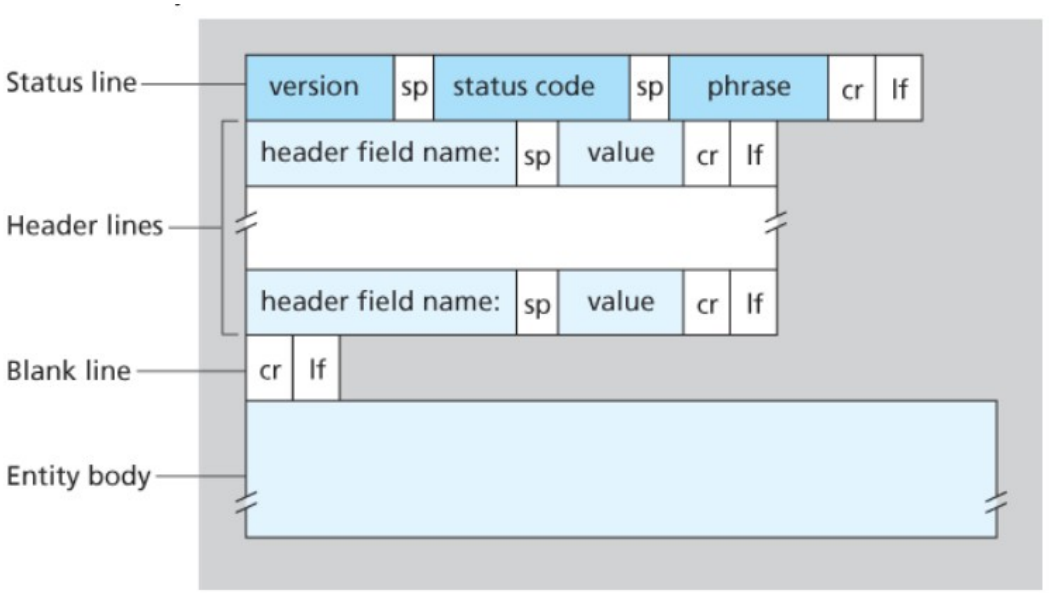
Format of an HTTP Request



- **Types of Requests:**
  - **GET**
  - **HEAD**
  - **POST**
  - **PUT**
  - **DELETE**

    ...

Format of an HTTP Response



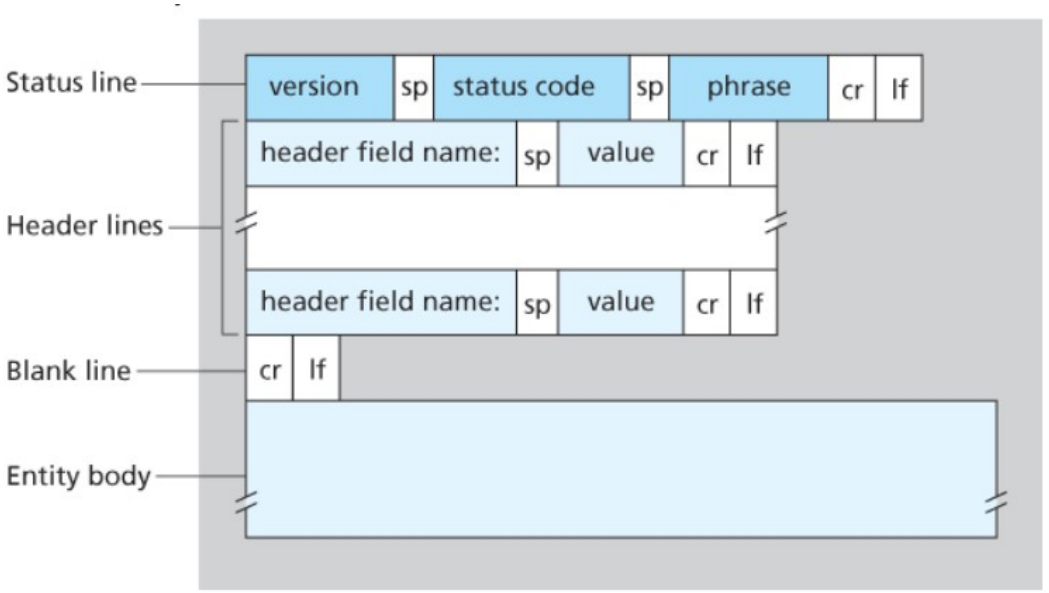
- **Example:**

```
HTTP/1.1 200 OK\r\n
Date: Sun, 26 Sep 2010 20:09:20 GMT\r\n
Server: Apache/2.0.52 (CentOS)\r\n
Last-Modified: Tue, 30 Oct 2007
   17:00:02 GMT\r\n
Content-Length: 2652\r\n
Keep-Alive: timeout=10, max=100\r\n
Connection: Keep-Alive\r\n
Content-Type: text/html\r\n
\r\n

data data data data data ...
```

Image Source: Kurose and Ross 7th edition

Format of an HTTP Response



Some Common Response Codes:

- **`200 OK`**
  request succeeded, requested object later in this msg
- **`301 Moved Permanently`**
  requested object moved, new location specified later in this msg (Location:)
- **`400 Bad Request`**
  request msg not understood by server
- **`404 Not Found`**
  requested document not found on this server
- **`505 HTTP Version Not Supported`**

Image Source: Kurose and Ross 7th edition

A Sample Browser Session

https://nehakaranjkar.github.io/348/samplewebpage/page1.html

Cookies

- A mechanism to identify a session/user on top of a stateless HTTP server
- How this works:
  - There is a server-side database
  - Server sends a “Set-cookie:<cookie ID>” header to the Client upon first visit
  - Client stores all cookies until they expire
  - Client includes a header “Cookie:<cookie ID>” in all subsequent requests to the same server
  - Server identifies the user/session from this cookie ID
- Privacy and Security Issues ....

Web Cache/Proxy Server

- The Caching concept... Cache HIT and Cache MISS
- Web Cache: stores copies of recently accessed pages and returns them to the requester on behalf of the origin web server.
- Purpose:
  - Reduce response times for client requests
  - Reduce traffic on the access link
- What if the webpage is updated in the origin and the cached copy is stale?
  - Freshness: Cached copy invalidated after a certain amount of time
  - Validation: Conditional GET (with a header “If-modified-since:<time>) to check if the cached copy should be updated