Computer Networks

World Wide Web

Jianping Pan
Fall 2022

SIR TIM BERNERS-LEE
Inventor of the World Wide Web, the first web browser, and the protocols and algorithms allowing the Web to scale

A.M. TURING AWARD 2016

Sir Tim Berners-Lee 2016 ACM A.M. Turing Lecture "What is the World Wide Web & what is its future?"
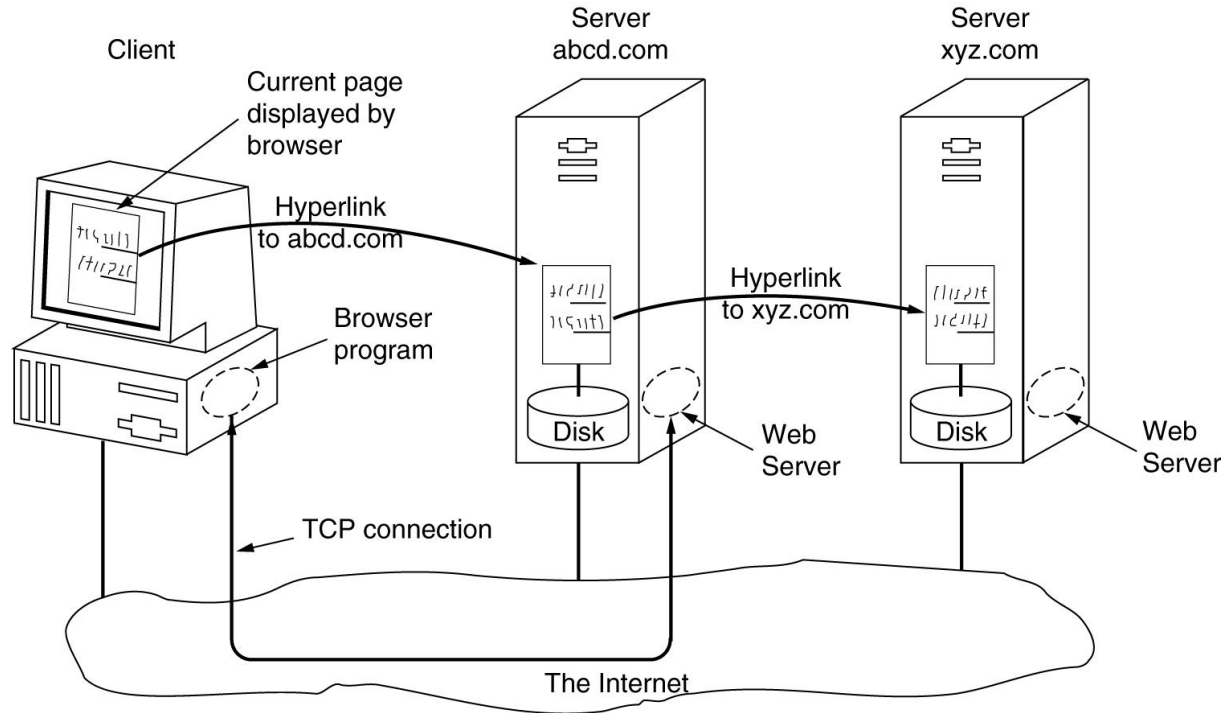
First things first

- Experienced almost all parts of CSc361?
  - Lectures, **T**utorials and **L**abs
  - Home **A**ssignment, **W**ritten and **P**rogramming
- 2nd lab session (L2) started on HTTP
- W1 posted on brightspace already
  - W1 (written assignment) due on Sep 23, 5pm
  - Will be marked and returned to you ASAP
  - Before the first **M**idterm (M1) for your prep

* I still need course reps for B03 and B06. Please drop me an email if you can volunteer, or find other ways to receive full 5%

World Wide Web

- ## Sir Tim Berners-Lee
  - 1989*, CERN; ... W3C; ... W3F!

- ## Hypertext and hypermedia
  - linked documents (existed earlier)

- ## Marc Andreessen
  - 1993, Mosaic, NCSA@UIUC

- ## Netscape Comm
  - Netscape navigator vs MS Internet explorer

* 30+ years by now

Q: Mozilla, IE, Firefox, Safari, Opera, Chrome?

Web overview

* page rank?

Uniform Resource Locator

- **http://**user:pass@**host**:port/**path/file**?input#anchor

Name	Used for	Example
http	Hypertext (HTML)	http://www.cs.vu.nl/~ast/
ftp	FTP	ftp://ftp.cs.vu.nl/pub/minix/README
file	Local file	file:///usr/suzanne/prog.c
news	Newsgroup	news:comp.os.minix
news	News article	news:AA0134223112@cs.utah.edu
gopher	Gopher	gopher://gopher.tc.umn.edu/11/Libraries
mailto	Sending e-mail	mailto:JohnUser@acm.org
telnet	Remote login	telnet://www.w3.org:80

* CSc360 file system?

Q: DocumentRoot and mapping?

HTML tags

- Anchor Links
  - \<a href="..."\>...\</a\>
- Image Objects
  - \<img src="..."\>

Tag	Description
\<html\> ... \</html\>	Declares the Web page to be written in HTML
\<head\> ... \</head\>	Delimits the page's head
\<title\> ... \</title\>	Defines the title (not displayed on the page)
\<body\> ... \</body\>	Delimits the page's body
\<h n \> ... \</h n \>	Delimits a level n heading
\<b\> ... \</b\>	Set ... in boldface
\<i\> ... \</i\>	Set ... in italics
\<center\> ... \</center\>	Center ... on the page horizontally
\<ul\> ... \</ul\>	Brackets an unordered (bulleted) list
\<ol\> ... \</ol\>	Brackets a numbered list
\<li\>	Starts a list item (there is no \</li\>)
\<br\>	Forces a line break here
\<p\>	Starts a paragraph
\<hr\>	Inserts a Horizontal rule
\<img src="..."\>	Displays an image here
\<a href="..."\> ... \</a\>	Defines a hyperlink

Q: linked vs embedded objects?

HTTP

- Hyper text transfer protocol
  - application layer protocol, ASCII format
    - HTTP/1.0: RFC1945 (1996); 1.1: RFC2068 (1997*)
  - typical client-server model: request-reply
    - client (browser): Mozilla, Opera, IE, Chrome, etc
    - server (web server)
      - Apache, Microsoft Internet information server (IIS)
  - normally uses service offered by TCP
    - http: 80; https: 443 (HTTP over SSL over TCP)

* 20+ years by now

* HTTP/2 (SPDY)? HTTP/3 (QUIC)

HTTP requests

- Request methods

Method	Description
GET	Request to read a Web page
HEAD	Request to read a Web page's header
PUT	Request to store a Web page
POST	Append to a named resource (e.g., a Web page)
DELETE	Remove the Web page
TRACE	Echo the incoming request
CONNECT	Reserved for future use
OPTIONS	Query certain options

- Request parameters (control headers)

Connection: keep-alive

HTTP responses

- ## Response codes

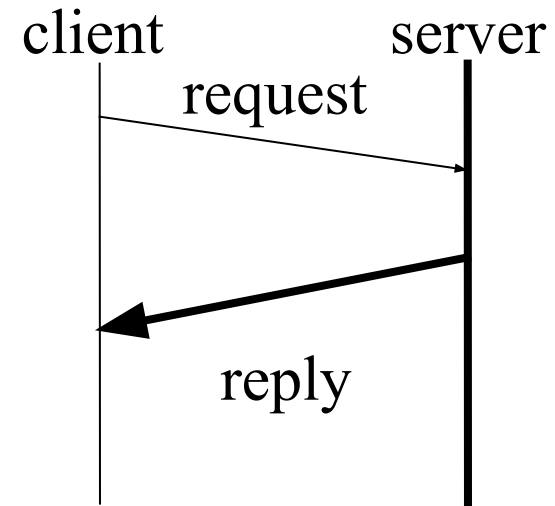| Code	Meaning	Examples
1xx	Information	100 = server agrees to handle client's request
2xx	Success	200 = request succeeded; 204 = no content present
3xx	Redirection	301 = page moved; 304 = cached page still valid
4xx	Client error	403 = forbidden page; 404 = page not found
5xx	Server error	500 = internal server error; 503 = try again later

  – 400: bad request

- ## Response parameters
- ## Response data

* ECS404? Computer Networks Research Lab ;-) BTW, ECS360: Computer Networks Teaching Lab
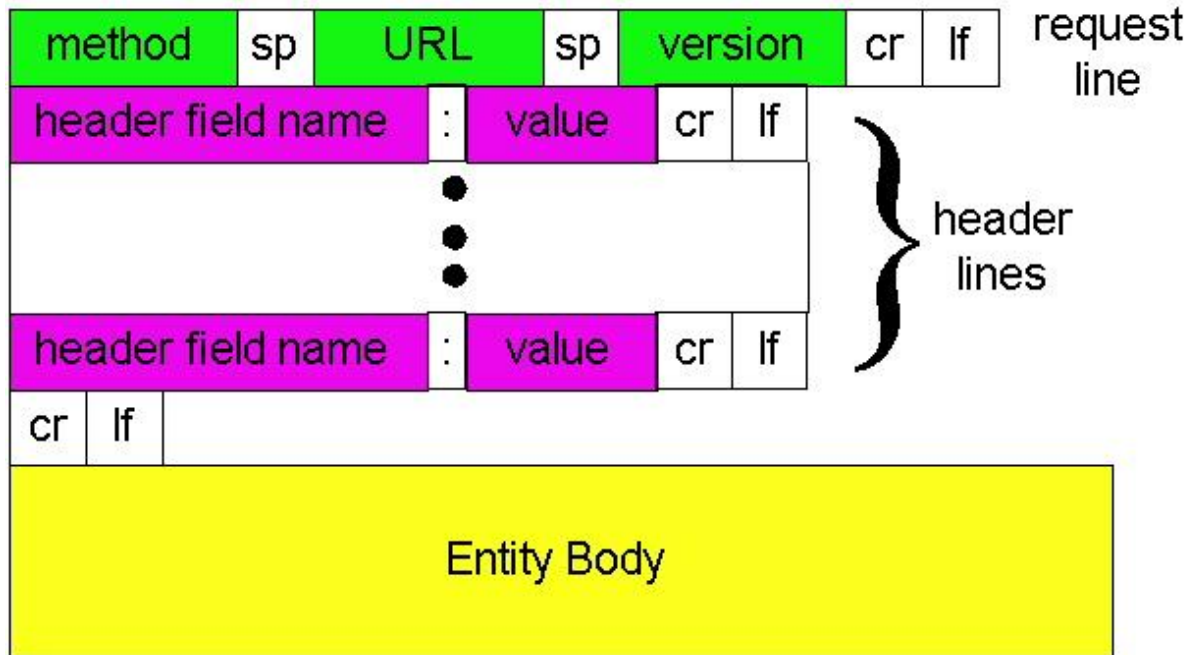
HTTP examples

- # wget -d www.google.com

    Connecting to www.google.com:80... Caching www.google.com <-> 66.102.7.104

    Created fd 3.

    connected!

    ---request begin---

    GET / HTTP/1.0

    User-Agent: Wget/1.7

    Host: www.google.com

    Accept: */*

    Connection: Keep-Alive

    ---request end---

    HTTP request sent, awaiting response...

    HTTP/1.0 302 Found

    Location: http://www.google.ca/

    Cache-Control: private

    Content-Type: text/html

    Server: GWS/2.1

    Content-Length: 218

client   server

request

reply

Q: syntax, semantics, sequence?

Message-level view

Web browsing examples

- http://www.a.com/index.html
  <html>
  <img src="http://www.a.com/x.gif">
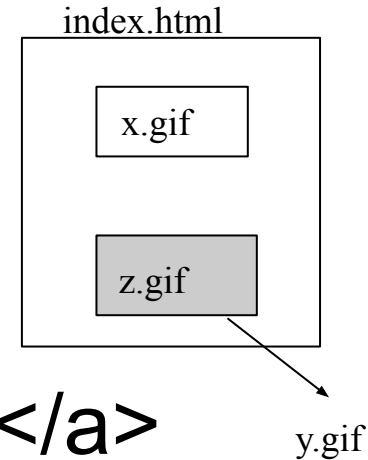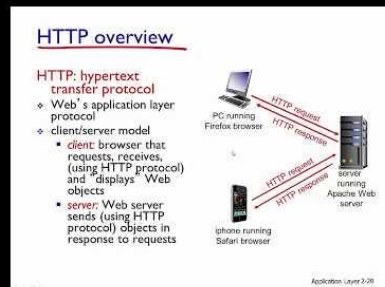
  <img src="http://www.a.com/z.gif">
  </html>

index.html

x.gif

z.gif

y.gif

- In your favorite web browser
  - URL: http://www.a.com
  - Q: how many HTTP requests?

discussion

This lecture

- HTTP basics
  - why "web"?
    - linked objects: links
    - embedded objects: e.g., images
  - URL
    - "http://user:pass@host:port/path/file?input#anchor"?!!
    - how to interpret?
  - HTTP service model
    - HTTP over TCP
      - connection-oriented, reliable data transfer service
      - Connection: keep-alive?
    - request/response
      - classic client-server model

https://www.youtube.com/watch?v=BaMa4u4Fio4

Next lecture

- HTTP intermediate
  - More on HTTP persistence
    - how many TCP connections do you need?
    - or, how fast can you get your web page?
      - PLT: page load time
      - a "metric" for browser wars
  - beyond the classic client-server model
    - web proxy, cache server, ... (you name it!)

- Read
  - KR4, Chapter 2