CS 484 Lecture 9: HTTP

Joshua Reynolds
Fall 2022

“As We May Think” Vannevar Bush, The Atlantic, 1945

“Consider a future device for individual use, which is a sort of mechanized private file and library.”

“On the top are slanting translucent screens…”

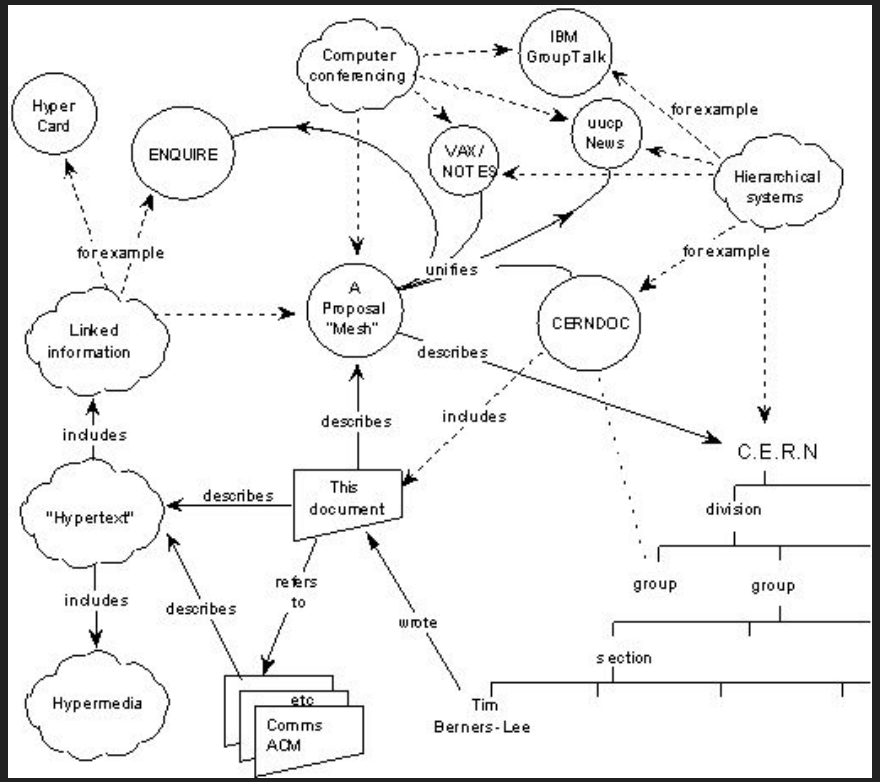
“...it would take him hundreds of years to fill the repository…”

“Business correspondence takes the same path.”

“If the user wishes to consult a certain book, he taps its code on the keyboard, and the title page of the book promptly appears before him”

“Wholly new forms of encyclopedias will appear”

Information Management: A Proposal, Tim Berners-Lee, 1989



https://www.w3.org/History/1989/proposal.html

Proposed linking information in a web instead of a hierarchical tree

← created on a 1989 computer!

HyperText: Non-Linear Text

HyperText was an existent system for linking files together within a filesystem.

Why not link files together from multiple filesystems?

Why only text files? What about images, videos, sound? ←HyperMedia

The oldest website in the world

https://www.w3.org/History/19921103-hypertext/hypertext/WWW/TheProject.html

The HyperText Transfer Protocol (HTTP)

A protocol to request and receive HyperText and HyperMedia

Like SMTP it is all ASCII and you could type it out yourself.

Like SMTP, there exist encrypted versions we will learn later

HTTP Demo

$ telnet www.google.com 80 # ← HTTP's standard port is 80

(telnet will automatically use CRLF line endings)

GET / HTTP/1.1

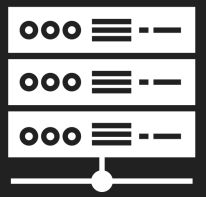
Host: www.google.com

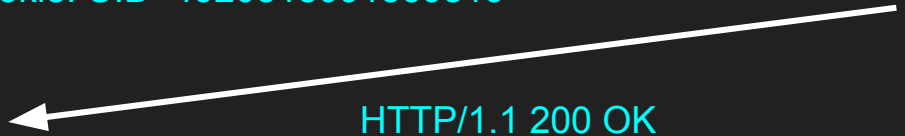
<CRLF>

<CRLF>

HTTP Client-Server Model

```
GET /index.html HTTP/1.1
Host: example.com
User-Agent: Mozilla/5.0 AppleWebKit (KHTML, like Gecko)
Accept-Encoding: gzip, deflate, br
Cookie: UID=4920318091369819
…
```





```
HTTP/1.1 200 OK
Date: Mon, 27 Jun 2029 11:48:44 GMT
Server: NGINX
Content-Length: 88
Content-Type: text/html

<html><body>
<h1>Hello, World!</h1>
…
```

HTTP Client Request Format

```
GET /index.html HTTP/1.1
Host: example.com
User-Agent: Mozilla/5.0 AppleWebKit (KHTML, like Gecko)
Accept-Encoding: gzip, deflate, br
Cookie: UID=4920318091369819
Content-Length: 56

{
	“Hi” : “There”,
	“Server” : “!”
}
```

HTTP Request Methods

GET*

POST

PUT

HEAD*

OPTIONS*

DELETE

PATCH

***Idempotent**

HTTP Request Headers

Host: example.com

User-Agent:

Referer: https://developer.mozilla.org/testpage.html

Connection: keep-alive

HTTP Response Codes

100-199 Informational

200-299 Varieties of success

300-399 Look Elsewhere

400-499 Requester Error

500-599 Server Errors

600+ Custom messages

101 Switching Protocols

200 Success

204 Success, no content

301 Moved Permanently

304 Not Modified

400 Bad Request

404 Not Found

500 Server Error

503 Service Unavailable

HTTP Response Headers

HTTP/1.1 200 OK

cache-control: public, max-age=31536000

Strict-Transport-Security: max-age=31536000; includeSubDomains; preload

content-length: 6292

content-type: text/javascript

<html><body>

…

HTTP Response Body

Encoding, Compression, Length

HTTP MIME Types (Multipurpose Internet Mail Extensions)

- text/javascript
- image/jpeg
- video/mp4
- font/woff
- application/zip
- audio/ogg

HTTP Pipelining

A strategy to be more efficient in using a TLS connection

Send all your HTTP requests at once, then wait for all the answers to come.

Don’t wait for your answer before making the next request

Displaying HyperText and HyperMedia to Users

Historical Web Browsers:

- Line Mode Browser (1991)
- Lynx (1992) ← still a thing! Text-based
- NCSA Mosaic (1993) images and text on the same screen!
- Windows Explorer (1995)
- Firefox (2002)
- Safari (2003) ← About ⅛ market share
- Chrome (2008) ← About ⅔ market share

The Browser Wars

Netscape vs Microsoft – Get everyone to use your browser

Famous anti-trust litigation against Microsoft in 1998-2001:

- Bill Gates' deposition
- How to say nothing while answering every question
- https://www.youtube.com/watch?v=gRelVFm7iJE 10:30+

HyperText Markup Language

Like XML, it is a tree with opening and closing tags for everything

```
<html>
    <head></head>
    <body></body>
</html>
```

Gets parsed by the browser for rendering graphically

The tree of parsed objects is the Document Object Model (DOM)

Styling can be exported to separate Cascading Style Sheets (CSS)

Javascript

Automated Memory Management

Script interpreted, not compiled

TypeScript is a type-enforcing variant

STDERR written to browser development console

Instead of syscalls, uses Browser API’s

Single threaded (except for)

Event driven (onclick, onkeydown, onload, etc)

Can access rendered HTML in the Document Object Model (DOM)

WebAssembly

When performance really matters, on supporting browsers you can run something like assembly

Can be integrated with JavaScript

Still being decided how this will work

Not supported in all browsers

Discussion: Who can even make a new browser with so many requirements?

Embedding with iFrames

The browser keeps every Web Origin (scheme, hostname, port) separate from each other.

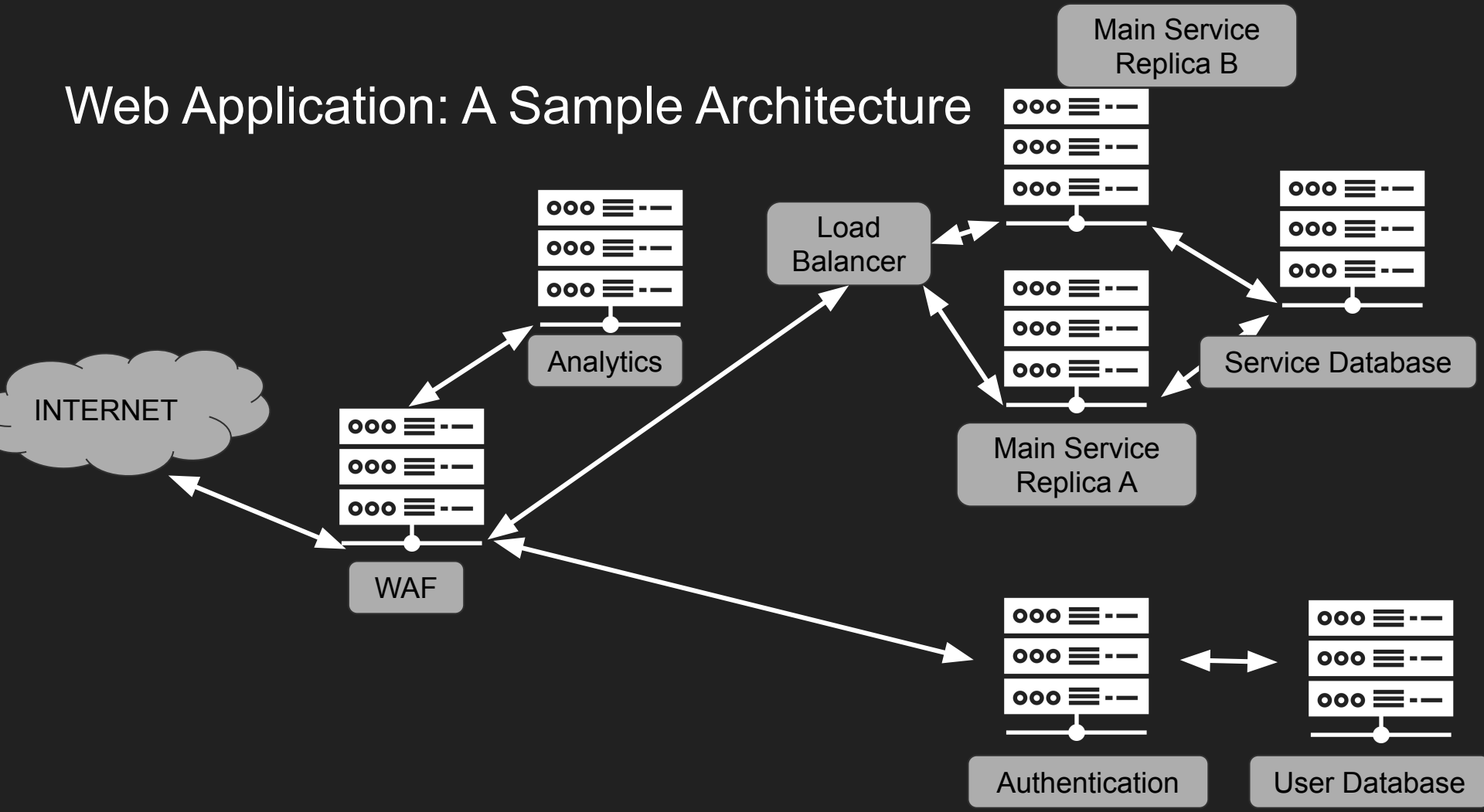
This is the Same Origin Policy - Prevents some XSS and prote

You cannot load content from elsewhere unless they specifically allow it.

Exception - other origins can be embedded in iframes unless they specifically include an HTTP response header to block it.

Browser Development Console Demo

Web Application: A Sample Architecture



Websockets – Not to be confused with network sockets!

GET ws://example.com:8181/ HTTP/1.1
Host: localhost:8181
Connection: Upgrade
Pragma: no-cache
Cache-Control: no-cache
Upgrade: websocket
Sec-WebSocket-Version: 13
Sec-WebSocket-Key: q4xkcO32u266gldTuKaSOw==

HTTP/1.1 101 Switching Protocols
Upgrade: websocket
Connection: Upgrade
Sec-WebSocket-Accept: fA9dggdnMPU79lJgAE3W4TRnyDM=

Web Sockets – Protocol includes frames w/ Op Codes

0x08 - End Connection

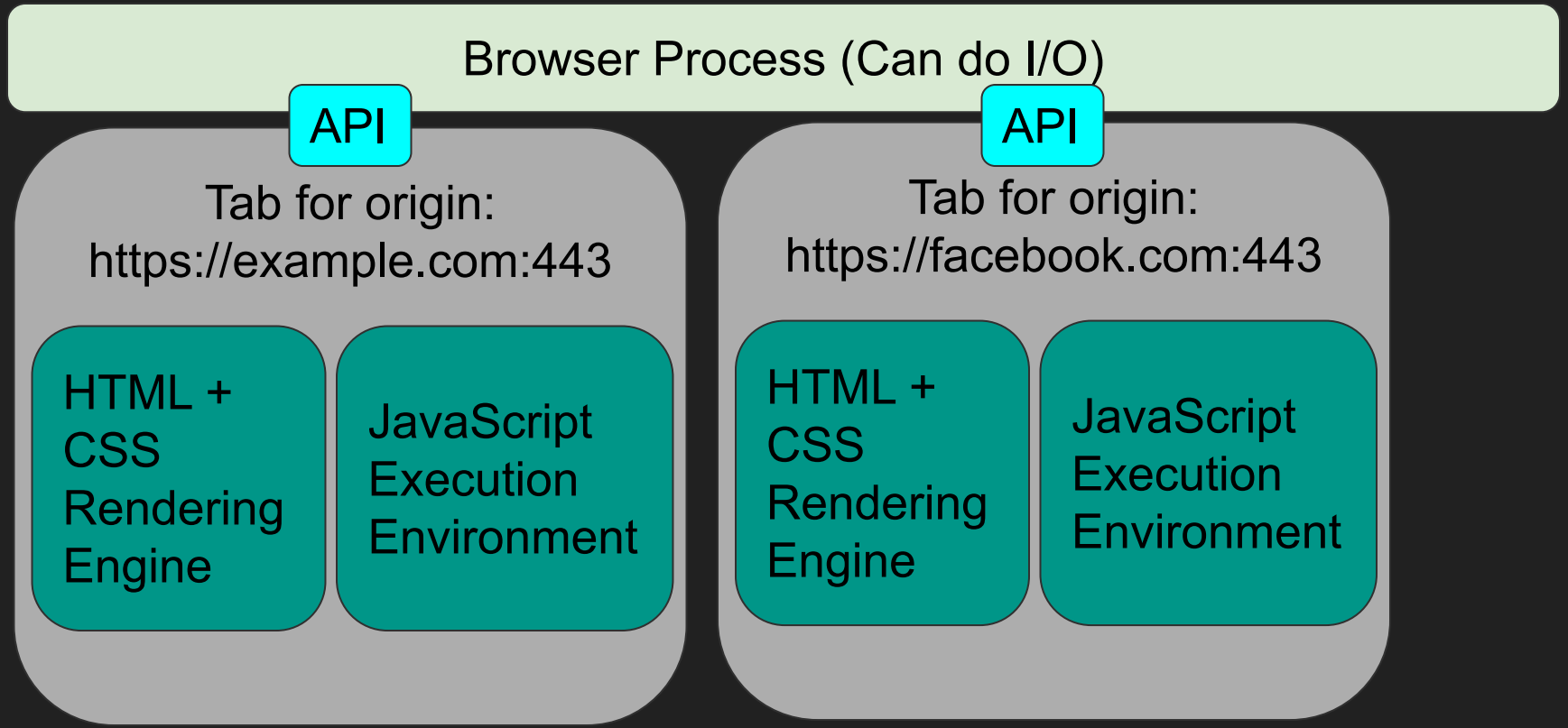
0x01 - UTF-8 Text

0x02 - Binary Data

0x09 - Ping (You still there?)

0x0A - Pong (Yup)

Browser Process Isolation (Sandbox)



Content Security Policies

Can be listed in <head> or HTTP response headers

Says where HTML, Javascript, CSS, and media can be loaded from on this Web Origin

Contains rules like:

Content-Security-Policy: default-src 'self' *.example.com; img-src *

FTP File Transfer Protocol (port 21)

Used to be supported in browsers until recently

A protocol for reading a filesystem directly from elsewhere

Still supported in Windows file explorer app

Image Attribution

Smartphone by IconHome from Noun Project 

Server by Creative Mania from Noun Project