CSCI 3330 Cybersecurity

Word of the day

GROK vt. [from the novel "Stranger in a Strange Land", metaphorically meaning 'to be one with']

1. To understand, usually in a global sense. Connotes intimate and exhaustive knowledge.

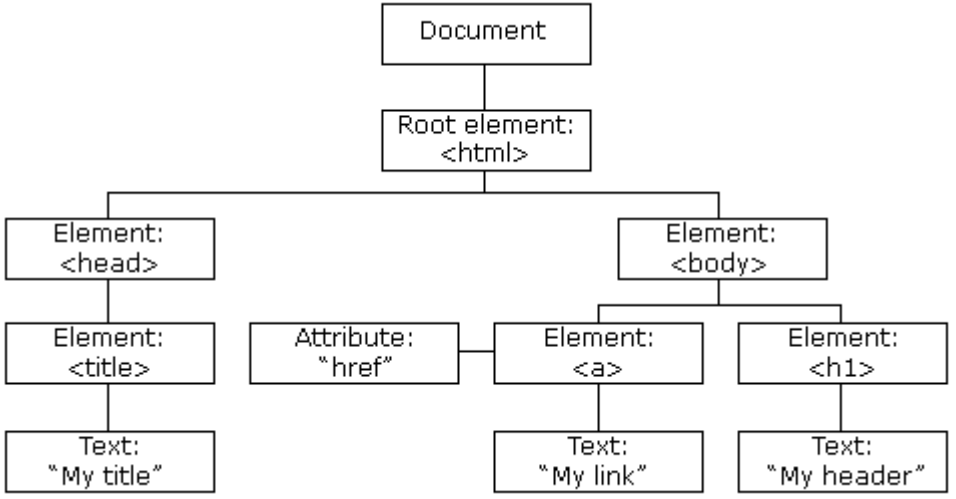
2. Used of programs, may connote merely sufficient understanding. "Almost all C compilers grok the `void' type these days."

HTML

- HyperText Markup Language

HTML

- HyperText Markup Language



Cookies

Cookies

- Allow sites to remember / track you
- Permits, e.g., logging in
- Session cookies
- Persistent cookies
- Key-value pairs

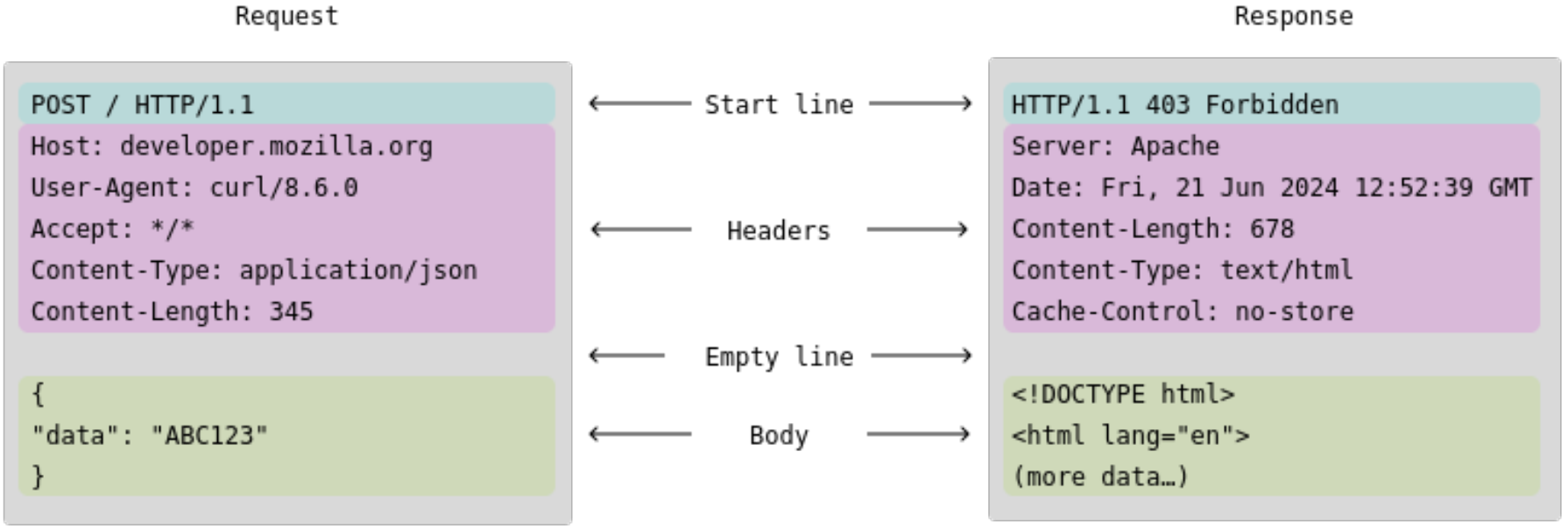
HTTP

- HyperText Transfer Protocol
- 1.1 most popular, 2.0 and 3.0 also exist
- Differences in serialization, transport
- Most primary features are equivalent

HTTP

- Different verbs
- GET: not make changes
- POST: make changes
- PUT, DELETE, etc.

HTTP request / response



Useful tools

- netcat…
- wget…
- curl…

- Let’s look at developer tools:
- https://news.ycombinator.com/

Components of a URL

protocol://host:port/path?query#fragment

- protocol: http or https (encrypted)
- host: name, address (IPv4, IPv6, etc.)
- port: by default, 80 for HTTP, 443 for HTTPS
- path: path to resource
- query: key=value&pairs=ofstuff
- fragment: bookmark

python’s urllib

- Demo using urllib.parse.urlparse()

- Query string a.k.a. “GET” parameters
- https://www.google.com/search?q=test

- “POST” parameters
- BODY of the HTTP request
- Not visible in the URL
- Often the result of submitting a form

JavaScript

- Modify DOM
- document.getElementById()
- document.documentElement()
- element.style.camelCaseStyleName
- document.cookie

Cross-site scripting (XSS)

- Inject malicious Javascript
- Another in-band signal
- Common payload when testing: alert('xss');

XSS types

- Non-persistent
- https://www.google.com/search?q=<script>alert('xss');</script>

XSS types

- Persistent
- Submitting a form, storing it in a database
- Advantages:
  - Don’t need to send someone a URL
  - Don’t need to expose attack in URL

XSS types

- “DOM-based”
- Aren’t they all DOM-based?!?
- Client-side, JavaScript based
- element.innerHTML = userInput;

XSS cheat sheet

https://web.archive.org/web/20120808002214/http://ha.ckers.org/xss.html

XSS control characters

- "" + searchQuery + ""
- What are they?

How do we prevent?

- Escaping control characters
- Removing in-band signal
  - E.g., for DOM-based, innerText vs. innerHTML
- What if we want to allow *some* HTML?

Blacklisting vs. whitelisting

- Blacklisting/blocklisting: block all on a list
- Whitelisting/allowlisting: only allow all on a list

- Parse HTML into abstract syntax tree
- Filter abstract syntax tree
- Parse and filter CSS (style="", <style />, etc.)
- Unparse abstract syntax tree into HTML

- Whitelist: HTML tags
- Whitelist: HTML attributes
- Blacklist: CSS expression()
- https://bowdoin.instructure.com/courses/7468/pages/test-xss/edit

Regular expressions

- Fast, expressive string matching
- Power: compile regex → NFA → DFA
- A DFA matches a string of *n* length in O(*n*) time
  - No backtracking!

Matching an email

```
(?:[a-z0-9!#$%&'*+\x2f=?^_`\x7b-\x7d~\x2d]+(?:\.[a-z0-9!#$
%&'*+\x2f=?^_`\x7b-\x7d~\x2d]+)*|"(?:[\x01-\x08\x0b\x0c\
x0e-\x1f\x21\x23-\x5b\x5d-\x7f]|\\[\x01-\x09\x0b\x0c\x0e-\
x7f])*")@(?:(?:[a-z0-9](?:[a-z0-9\x2d]*[a-z0-9])?\.)+[a-z0-
9](?:[a-z0-9\x2d]*[a-z0-9])?|\[(?:(?:(2(5[0-5]|[0-4][0-9])|
1[0-9][0-9]|[1-9]?[0-9]))\.){3}(?:(2(5[0-5]|[0-4][0-9])|
1[0-9][0-9]|[1-9]?[0-9])|[a-z0-9\x2d]*[a-z0-9]:(?:[\x01-\
x08\x0b\x0c\x0e-\x1f\x21-\x5a\x53-\x7f]|\\[\x01-\x09\x0b\
x0c\x0e-\x7f])+)\])
```

Simple matches

- VISA/Mastercard

```
[0-9]{4}-?[0-9]{4}-?[0-9]{4}-?[0-9]{4}
```

- U.S. phone number

```
([0-9]{3}-?)?[0-9]{3}-?[0-9]{4}
```

Regular expression basics

- Concatenation `xy`
- Alternation `x|y`
- Kleene star `x*`
- Grouping/operator precedence `(x|y)*`

Quantifier shortcuts

- x* (zero or more)
- x? (zero or one)
- x+ (one or more x)
- x{n} (n-many x)

Other rules

- [0-9a-fA-F] (character ranges)
- [^0-9] (negative character ranges)
- . (wildcard)
- ^ (beginning of line)
- $ (end of line)
- \ (escape special symbol to get bare character)

Expressions to match…

- Hex: 0xdeadbeef, 0X123ABC
- A directory traversal attack (../../../etc/passwd)
- Parenthesis matching: (), (()), ((((()))))

- Limitations in expressiveness
  - Can only match “regular” languages
  - Cannot do parenthesis matching
  - Cannot do HTML tag matching (!?)

In-class exercise

- https://brainhammer.com/regexplayground/
- Minimum: 2
- Expected: 4
- Cool: 6

Cross-site request forgeries (CSRF)

- Same origin policy: scripts from one origin should not be able to read or manipulate content from another origin

Assignment 2

- Deadline: September 25, 2025, 10:00pm
- Setuid binaries, directory traversal, command line injection
  https://tildesites.bowdoin.edu/~j.knockel/csci3330f25/assignment2.pdf

Assignment 3

- Deadline: October 2, 2025, 10:00pm
- Setuid binaries, directory traversal, command line injection
  https://tildesites.bowdoin.edu/~j.knockel/csci3330f25/assignment3.pdf