

ICS 451: Today's plan

- DNS note
- World-Wide Web
- URI/URL
- HTML
- HTTP
 - protocol review
 - versions
 - methods
 - cookies

DNS note

- Domain names match in a case-independent manner
- resolving hawaii.edu gives the same result as resolving HAWAII.EDU or HaWali.EdU
- the path component of a URL may be case sensitive, but the domain name is not!

World-Wide Web

- Before the www, people on the Internet had three main applications:
 - ftp for file transfer
 - telnet for remote login
 - email for asynchronous communication
- the www was developed by physicists at CERN to exchange technical documents that contained *links* to other documents: **hypertext**

Components of the www

- a content Identification mechanism: URI/URL
- a content encoding: HTML
- a protocol to exchange content: HTTP

URI/URL

- Uniform Resource Identifier/Locator
- `scheme://authority path`
 - possibly add `?query` or `#fragment`
- `authority` is usually the domain name, but may:
 - begin with user-information `@`
 - end with `:port`
- `ftp://cnn.example.com&story=breaking_news@10.0.0.1/top_story.htm`
- `scheme` may be `http`, `ftp`, `mailto`, `file`, etc

HTML

- Inline (all ASCII) markup of text
- `<tag>content</tag>`
- `<tag option="value">content</tag>`
- `<` and `>` for literal `<` and `>`
- generally `<html><head>...</head>
<body>...</body></html>`
- ``

Web scripting

- HTML is designed to be general, but people always want more effects and better control over layout
- So web browsers interpret languages to provide such special effects
 - java and javascript
- The challenge is to prevent programs written in such scripts from harming the host or others
 - scripts are run in *sandboxes* that limit what the code can do

HTTP review

- Client-server exchange
- Each direction sends a header, possibly followed by a body
- Header ends with an empty line (like email)
- Header lines after the first have field: value
- Content-length: may tell us how much content to read
- Content-type: tells us what kind of content

HTTP versions

- 0.9: single request line, response is content
- 1.0: multi-line headers, response code
- 1.1: persistent connections
 - C->S Connection: keep-alive or close
 - S->C Keep-alive: timeout=3, max 5
 - Host: tells server which pages we want
important for virtual servers, i.e. several domain names mapping to the same IP

HTTP methods

- GET: give me this content
- HEAD: give me the header for this content
- POST: take this content

HTTP status codes

- 2xx: good
 - 200 OK
- 3xx: not here
 - 301 Moved Permanently (redirect)
 - 304 Not Modified (for If-Modified-Since)
- 4xx: client error
 - 404 Not Found
- 5xx: server error

HTTP status codes

- 2xx: good
 - 200 OK
- 3xx: not here
 - 301 Moved Permanently (redirect)
 - 304 Not Modified (for If-Modified-Since)
- 4xx: client error
 - 404 Not Found
- 5xx: server error

HTTP statelessness

- Each HTTP request is processed independently
 - the server does not consider history when serving content
 - this is very powerful, simplifies the server
- Sometimes history (state) is important
 - did this user authenticate?
 - did this user tell me what they prefer?

HTTP cookies

- S → C: Set-Cookie:
PREF=ID=4311228a5c422d51:FF=0:TM=1423
126138:LM=1423126139:S=20wOldft-
EOGtUHI; expires=Sat, 04-Feb-2017 08:48:59
GMT; path=/; domain=.google.com
- C → S: Cookie: ...
- Client stores cookie on disk until expiration
- Client sends cookies to server on every request to that domain

HTTP cookies and privacy

- The cookie mechanism is largely automated
- Users may not be aware of the information websites collect about them
- A 1-pixel image in a web page may contain a cookie from domain example.com
- If example.com places the same image in many web pages, they can track users as they visit the different web pages

HTTP cookies and privacy

- Users can tell browsers to discard cookies
- Some “cookies” (e.g. Flash cookies) are hard to discard
- Some websites might not work without cookies
 - cookies can provide better service!!!!
- Different jurisdictions have different laws
 - e.g. the EU vs. the US