Web Security I

But first… Web Technology extras

Anatomy of a Web Request

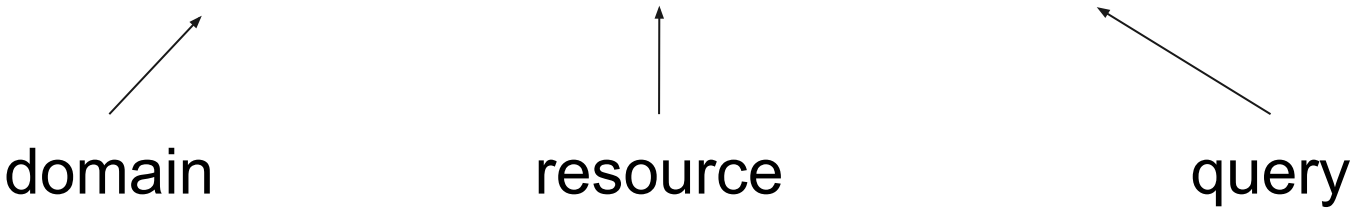
- User types `http://www.foo.com/about`
- Hits enter
- What happens?

Anatomy of a Web Request: URLs

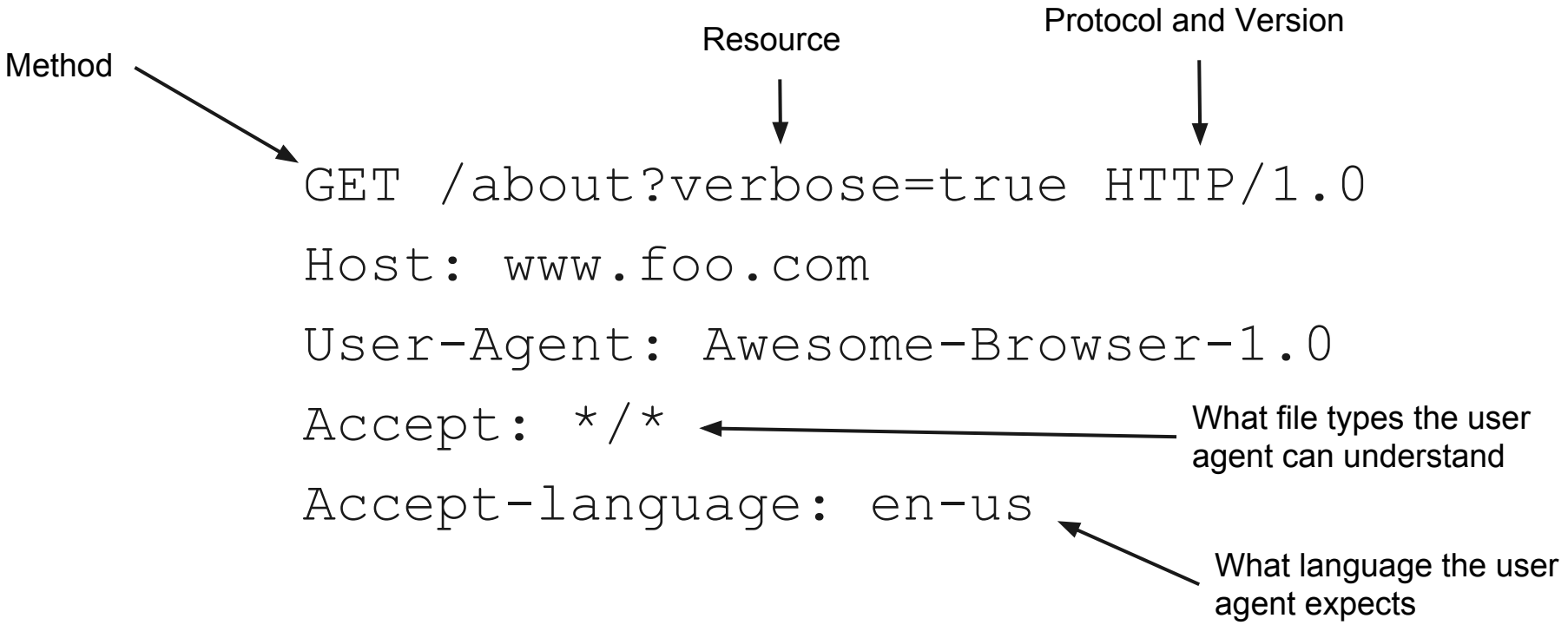
- How to parse a URL?
- First, parse the **protocol**
  - `protocol:<address>`
  - `http://www.foo.com/about`
  - `mailto:bernardo@cs.brown.edu`
  - `magnet:?xt=urn:sha1:YNCKHTQCWBTRNJIV4...`
- Browser might not speak protocol; open in external program (e.g., mail client)

Anatomy of a Web Request: URLs

- How to parse a URL?
- Second, parse the **address**
- Every protocol has its own address format
- For HTTP: domain, resource, query
  - `http://www.foo.com/about?verbose=true`



Anatomy of a Request: Header



Anatomy of a Request: DNS

- How do we know where `www.foo.com` is?
- DNS (Domain Name System)
  - Maps human-readable names…
    - `www.foo.com`
  - To IP addresses…
    - `184.73.167.182`
  - Hierarchical (`foo.com` controlls `www.foo.com`)
- That's all for now; more in the networks unit

Anatomy of a Request: Connection

- Browser looks up `www.foo.com`, gets `184.73.167.182`
- Makes a TCP connection to `184.73.167.182`
- Sends HTTP request over TCP connection
  - Important to specify host
  - Could be multiple web sites at the same IP!
- Waits for HTTP response

Anatomy of a Request: Serving

- Web server gets request for `/about`
- How does it know what to respond with?

Anatomy of a Request: Serving

- Web server gets request for `/about`
- How does it know what to respond with?
- Here's one popular way to do it
  - Not the only way
  - Not necessarily the best way
  - One of the oldest ways
  - *One of the most insecure ways :)

**Not actually insecure by design, but harder to make secure*

Anatomy of a Request: Serving

- Web server gets request for `/about`
- Server configured with **web root**
  - For us it's `/course/cs166/www`
  - Often `/var/www` or similar
- Resources path matches file in web root
  - e.g., `/var/www/about`

Anatomy of a Request: Serving

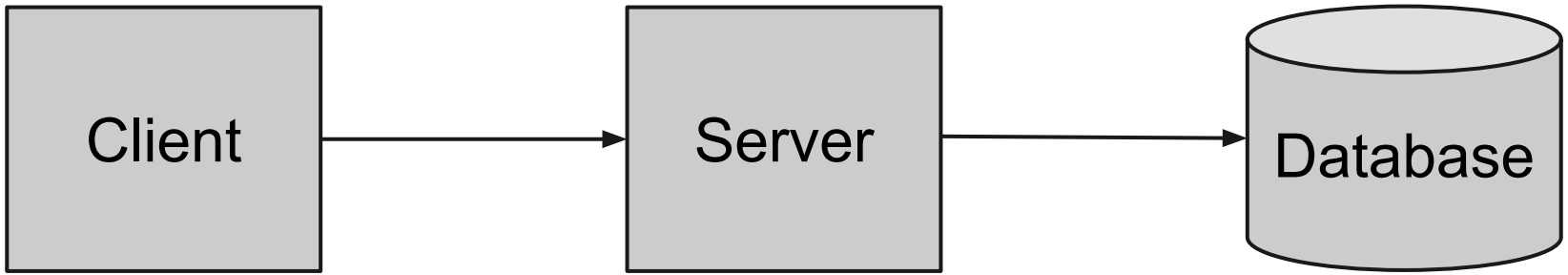
- Web server gets request for `/about`
- Server configured with **web root**
  - For us it's `/course/cs166/www`
  - Often `/var/www` or similar
- Resources path matches file in web root
  - e.g., `/var/www/about`
- ...but what about dynamic content?

Anatomy of a Request: Serving

- Paths specify *scripts* in web root
- Each web request runs code to generate response
  - e.g., `/index.php`
  - `/var/www/index.php`

Anatomy of a Request: Serving

- Often sites interact with databases
  - Save user data (usernames, passwords, etc)
  - Posts, comments, etc



Anatomy of a Request: Serving

- What about AJAX?
- HTTP requests aren’t for HTML, but just data
- Encoded as JSON, XML, etc