Web server reconnaissance
Reconnaissance and fingerprinting

- Finding information about a target web server/web site
  - May be illegal to perform reconnaissance on a web server and web site without prior approval/permission.
  - Simulate via war games to demonstrate issues with trusting clients with URLs and filenames
- Fingerprinting information
  - Name and version of server
  - Database backend
  - Use of reverse proxy (nginx)
  - Programming language and web application server
1. Viewing HTTP headers

$ nc -C vulnerable 80
GET / HTTP/1.1
Host: vulnerable

HTTP/1.1 200 OK
Date: Sun, 03 Mar 2013 10:56:20 GMT
Server: Apache/2.2.16 (Debian)
X-Powered-By: PHP/5.3.3-7+squeeze14
Content-Length: 6988
Content-Type: text/html

$ nc -C oregonctf.org 80
HEAD / HTTP/1.1
Host: foobar

$ nc -C cs410.oregonctf.org 80
HEAD / HTTP/1.1
Host: cs410.oregonctf.org
2. Viewing source content

- Look for comments, links, or directory structure

- Wikipedia (en.wikipedia.org)
2. Viewing source content

- Hyatt.com

- GitHub
  - Hint: Asset cache busting pipeline

- Blackboard
2. Viewing source content

- External services
  - https://builtwith.com
  - https://wappalyzer.com
  - https://urlscan.io
3. Search engine signals

- Google, Yahoo, Bing, will crawl everything on your site unless you tell them otherwise.
  - Prevent via use of **robots.txt** file
  - Instructs search engine spiders how to interact with your content.
  - Can also reveal sensitive information
3. Search engine signals

- For hacker, robots.txt can contain interesting folders, files, and data to investigate.
  - Sometimes even passwords, usernames, ...
- Example
  - Specifies that no robots should visit any URL starting with "/cyberworld/map/" or "/tmp/", or /foo.html.

```text
# robots.txt for http://www.example.com/

User-agent: *
Disallow: /cyberworld/map/ # This is an infinite virtual URL space
Disallow: /tmp/ # these will soon disappear
Disallow: /foo.html
```
If the Joomla site is installed within a folder such as at e.g. www.example.com/joomla/ the robots.txt file MUST be moved to the site root at e.g. www.example.com/robots.txt AND the joomla folder name MUST be prefixed to the disallowed path, e.g. the Disallow rule for the /administrator/ folder MUST be changed to read Disallow: /joomla/administrator/

For more information about the robots.txt standard, see: http://www.robotstxt.org/orig.html

For syntax checking, see: http://tool.motoricerca.info/robots-checker.phtml

User-agent: *
Disallow: /administrator/
Disallow: /cache/
Disallow: /cli/
Disallow: /components/
Disallow: /images/
Disallow: /includes/
Disallow: /installation/
Disallow: /language/
Disallow: /libraries/
Disallow: /logs/
Disallow: /media/
Disallow: /modules/
Disallow: /plugins/
Disallow: /templates/
Disallow: /tmp/
4. Artifacts

- favicon.ico
  - Default icons indicate software package being used
  - Which package?

- Search-engine worms (Santy worm 2004)
  - phpBB
4. Artifacts

- Application-specific 404 error pages
  - Tomcat, Ruby on Rails
4. Artifacts

- Stack trace of web application
- Inject %00, %22, %27 to check for injection vulnerabilities
5. TLS transparency reports

- Rogue certificate authority can create valid certificates for sites it is not supposed to
  - Force all authorities to log every certificate (for HTTPS) issued to a central location
  - Browsers eventually will reject those that are not logged
  - But, exposes all of the names of machines an organization has generated certificates for
  - Potential targets for adversaries
    - [https://transparencyreport.google.com/https/certificates](https://transparencyreport.google.com/https/certificates)
    - [https://observatory.mozilla.org/](https://observatory.mozilla.org/) (TLS section)
- Demo: Lookup all oregonctf.org certificates and who issued them
6. Fuzzing

- Brute-force common directory names
  - (e.g. admin, config, conf, src)
  - Brute-force admin pages with default admin credentials
- `wfuzz` tool
  - Detect directories and pages on the web server using wordlists of common resource names.

```
$ wfuzz -c -z file,wordlist/general/common.txt --hc 404 http://vulnerable/FUZZ
```

- `nmap` tool
  - General tool supporting any number of scans
  - Can specifically be used to enumerate directories in web servers similar to `wfuzz`

```
nmap --script http-enum w.x.y.z
```
Lab: A0 Reconnaissance

- See handout