CSE 333
Section 9
HW4, HTTP, and Boost
Logistics

- Exercise 11
  - Due Friday (03/03) @ 11 am
- Exercise 12
  - Released Friday (03/03)
  - Due next Wednesday (03/08) at 11 am
- Homework 4
  - Due next Thursday (03/09) @ 11:59 pm
HTTP Review
HTTP Review

1. What does HTTP stand for?

   HyperText Transfer Protocol

2. What layer does HTTP reside in?

   Application Layer

3. What does HTTP define?

   HTTP defines how we should send information between a client and a server
<table>
<thead>
<tr>
<th>Method</th>
<th>URI</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>GET</td>
<td>/courses/cse333/22wi/</td>
<td>HTTP/1.1</td>
</tr>
<tr>
<td>Host</td>
<td>courses.cs.washington.edu</td>
<td></td>
</tr>
<tr>
<td>Connection</td>
<td>keep-alive</td>
<td></td>
</tr>
<tr>
<td>sec-ch-ua:</td>
<td>&quot; Not A;Brand&quot;;v=&quot;99&quot;, &quot;Chromium&quot;;v=&quot;98&quot;, &quot;Google Chrome&quot;;v=&quot;98&quot;</td>
<td></td>
</tr>
<tr>
<td>sec-ch-ua-mobile:</td>
<td>?0</td>
<td></td>
</tr>
<tr>
<td>sec-ch-ua-platform:</td>
<td>&quot;nacOS&quot;</td>
<td></td>
</tr>
<tr>
<td>Upgrade-Insecure-Requests:</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>User-Agent:</td>
<td>Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/98.0.4758.109 Safari/537.36</td>
<td></td>
</tr>
<tr>
<td>Accept:</td>
<td>text/html, application/xhtml+xml, application/xml; q=0.9, image/avif, image/webp, image/apng, <em>/</em>; q=0.8, application/signed-exchange; v=b3;q=0.9</td>
<td></td>
</tr>
<tr>
<td>Sec-Fetch-Site:</td>
<td>none</td>
<td></td>
</tr>
<tr>
<td>Sec-Fetch-Mode:</td>
<td>navigate</td>
<td></td>
</tr>
<tr>
<td>Sec-Fetch-User:</td>
<td>?1</td>
<td></td>
</tr>
<tr>
<td>Sec-Fetch-Dest:</td>
<td>document</td>
<td></td>
</tr>
<tr>
<td>Accept-Encoding:</td>
<td>gzip, deflate, br</td>
<td></td>
</tr>
<tr>
<td>Accept-Language:</td>
<td>en-US, en;q=0.9, es;q=0.8, it;q=0.7, zh-CN;q=0.5, zh;q=0.5</td>
<td></td>
</tr>
<tr>
<td>Cookie:</td>
<td>rL_page_init_referer=RudderEncrypted%3AU2Fs GVkX1%2BpIj%2B1vZv50b9rEBtZr07gXF7fY40%3D; rL_page_init_referring_domain=RudderEncrypted%3AU2Fs GVkX1%2B3vk4w%2BxdaST8KA3fAqQe%2FlamkREM%3D; rL_anonymous_id=RudderEncrypted%3AU2Fs GVkX1%2Fm3x35zo6GyUCta1cDcjv1FSc0b0cibrq1I0NPgc1L1ZFvM8eqsJ0L19Lqzn3C86JQTre2ga9QurrQ%3D; rL_group_id=RudderEncrypted%3AU2Fs GVkX1%2B5ET%2BAl0e1WPU6E61450fQyBXKn08GsLw%3D; rL_group_trait=RudderEncrypted%3AU2Fs GVkX1%2B0tjYuogrYGTwycK0p4F7cmU3X%2Bylq1%3D; rL_user_id=RudderEncrypted%3AU2Fs GVkX1%2BnFbEzeBuVC905SVr212oVtvpPpBrna0P2Hn5ns0TKfVCvnFNLIK; rL_trait=RudderEncrypted%3AU2Fs GVkX1950AzIw7sf8308y6yr58r0TbbqA%2FMEY%3D</td>
<td></td>
</tr>
</tbody>
</table>
HTTP Request Format

[METHOD] [request-uri] HTTP/[version]\r\n
[headerfield1]: [fieldvalue1]\r\n
[headerfield2]: [fieldvalue2]\r\n
[...]

[headerfieldN]: [fieldvalueN]\r\n\r\n
[request body, if any]

Note: Double return indicates the end of the headers section
# HTTP Methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GET</td>
<td>The GET method requests a representation of the specified resource. Requests using GET should only retrieve data.</td>
</tr>
<tr>
<td>HEAD</td>
<td>The HEAD method asks for a response identical to that of a GET request, but without the response body.</td>
</tr>
<tr>
<td>POST</td>
<td>The POST method is used to submit an entity to the specified resource, often causing a change in state or side effects on the server.</td>
</tr>
<tr>
<td>PUT</td>
<td>The PUT method replaces all current representations of the target resource with the request payload.</td>
</tr>
<tr>
<td>DELETE</td>
<td>The DELETE method deletes the specified resource.</td>
</tr>
<tr>
<td>CONNECT</td>
<td>The CONNECT method establishes a tunnel to the server identified by the target resource.</td>
</tr>
<tr>
<td>OPTIONS</td>
<td>The OPTIONS method is used to describe the communication options for the target resource.</td>
</tr>
<tr>
<td>TRACE</td>
<td>The TRACE method performs a message loop-back test along the path to the target resource.</td>
</tr>
<tr>
<td>PATCH</td>
<td>The PATCH method is used to apply partial modifications to a resource.</td>
</tr>
</tbody>
</table>
HTTP/1.1 200 OK
Date: Mon, 21 May 2018 07:58:46 GMT
Server: Apache/2.2.32 (Unix) mod_ssl/2.2.32 OpenSSL/1.0.1e-fips
        mod_pubcookie/3.3.4a mod_uwa/3.2.1 Phusion_Passenger/3.0.11
Last-Modified: Mon, 21 May 2018 07:58:05 GMT
ETag: "2299elef-52-56cb2a9615625"
Accept-Ranges: bytes
Content-Length: 82
Vary: Accept-Encoding,User-Agent
Connection: close
Content-Type: text/html
Set-Cookie: 

<html><body>
<font color="chartreuse" size="18pt">Awesome!!</font>
</body></html>
HTTP Response Format

HTTP/[version]  [status code]  [reason]\r\n[headerfield1]:  [fieldvalue1]\r\n[headerfield2]:  [fieldvalue2]\r\n[...]
[headerfieldN]:  [fieldvalueN]\r\n\r\n[response body, if any]

Note: Double return indicates the end of the headers section
HTTP Response Status Codes

- **HTTP/1.1 200 OK**
  - The request succeeded and the requested object is sent

- **HTTP/1.1 404 Not Found**
  - The requested object was not found

- **HTTP/1.1 301 Moved Permanently**
  - The object exists, but its name has changed
  - The new URL is given as the “Location:” header value

- **HTTP/1.1 500 Server Error**
  - The server had some kind of unexpected error
HTTP Response Status Codes

- **HTTP/1.1 200 OK**
  - The request succeeded and the requested object is sent
  - HW4 Ex.: User requests a file that is successfully found by the server

- **HTTP/1.1 404 Not Found**
  - The requested object was not found
  - HW4 Ex.: User requests a file that has been deleted so server can’t find it

- **HTTP/1.1 301 Moved Permanently**
  - The object exists, but its name has changed
  - The new URL is given as the “Location:” header value
  - Ex.: washington.edu is redirected permanently to uowash.edu

- **HTTP/1.1 500 Server Error**
  - The server had some kind of unexpected error
  - Ex.: Corrupted browser cache
HW4 Overview
HW4: Web Server

1. Establish client connections
   a. Server socket set up in hw4/ServerSocket.cc

Socket API: Server TCP Connection

◦ Pretty similar to clients, but with additional steps:
  1) Figure out the IP address and port on which to listen
  2) Create a socket
  3) bind() the socket to the address(es) and port
  4) Tell the socket to listen() for incoming clients
  5) accept() a client connection

Helpful to refer to:
- Server-side networking lecture
- server_accept_rw_close.cc
- Exercise 11
HW4: Web Server

1. Establish client connections
   a. Server socket set up in hw4/ServerSocket.cc

2. Read client requests
   a. Parse incoming HTTP requests in hw4/HttpConnection.cc

Notes:
- Request may be split across multiple read()’s
- Parse by splitting strings – can have any number of headers
- Working just with GET (i.e., no body)
HW4: Web Server

1. Establish client connections
   a. Server socket set up
      in hw4/ServerSocket.cc

2. Read client requests
   a. Parse HTTP requests
      in hw4/HttpConnection.cc

3. Respond to requests
   a. Write HTTP responses
      in hw4/HttpServer.cc

Notes:
- Built up via string manipulation
- Must handle a variety of situations based on request
- Interact with index and static files
HW4: Web Server

1. Establish client connections
   a. Server socket set up
      in hw4/ServerSocket.cc

2. Read client requests
   a. Parse HTTP requests
      in hw4/HttpConnection.cc

3. Respond to requests
   a. Write HTTP responses
      in hw4/HttpServer.cc

4. Fix security vulnerabilities
   a. Escape characters in hw4/Utils.cc

Cross-site scripting flaw:
Clients can supply code for execution!

```
hello <script>alert("Boo!");</script>
```

Directory traversal attack:
Clients can gain unauthorized access to files!

```
/static/..//hw4/http333d.cc
```
HW4: Web Server

1. Establish client connections
   a. Server socket set up
      in hw4/ServerSocket.cc

2. Read client requests
   a. Parse HTTP requests
      in hw4/HttpConnection.cc

3. Respond to requests
   a. Write HTTP responses
      in hw4/HttpServer.cc

4. Fix security vulnerabilities
   a. Escape characters in hw4/Utils.cc

Steps 2, 3, and 4 involve a lot of string manipulation which can be tedious!
HW4 Tools Demo
(Telnet and Browser)

(Helpful for debugging!!!)
Executables

● Running make in hw4/ will produce **test_suite** and **http333d**
  ○ test_suite is always a good place to start to start debugging using the usual tools (program output and GDB)
  ○ You can test with http333d (even if it’s not complete) by sending it requests via **telnet** or a **web browser**

● You also have access to **solution_binaries/http333d** and **solution_binaries/http333d_withflaws**
  ○ Do NOT leave the version with flaws running on an attu machine!

● Launching the server:
  ○ `.http333d <port> ../projdocs/unit_test_indices/*`
  ○ Or replace `.http333d` with `.solution_binaries/http333d`
Writing an HTTP Request

● Generic HTTP request layout can be easily found in HttpRequest.h

● HW4 request types:
  ○ Example file request:
    ■ GET /static/test_tree/books/artofwar.txt HTTP/1.1
  ○ Example query request:
    ■ GET /query?terms=books+of+war HTTP/1.1

● Headers:
  ○ For HW4, can get away with no headers
  ○ Can see others in browser-generated requests or looking at the HTTP lecture slides
Sending a Request (and Viewing Response)

**Telnet:**

1. `telnet <HostName> <port>`
2. Manually type out HTTP request.
3. Finish request by hitting `[Enter]` twice.
4. Exit using `Ctrl+]` then `Ctrl+d` or entering in `quit`.

**Browser:**

1. `http://<HostName>:<port>`
2. Append URIs from last slide to URL to send different types of request.
3. Open the Network tab in Developer Tools to see request and response.
   a. `<F12>` or `Ctrl+Shift+I` (Chrome, Firefox)
   b. Warning: the UX here can be a little confusing
4. Close browser tab when done.
Debugging Your Server’s Responses

- Generic HTTP response layout can be easily found in HttpResponse.h
  - Responses will have a body!

- Copy what solution_binaries/http333d does
  - Can simply copy the HTML if you’re not familiar

- Content-Type is really important!
  - Tells client how to handle/interpret the response body
  - Your server should work with a variety of file types
    (in particular, see projdocs/bikeapalooza_2011/)
Using Telnet with HW4

1. Launch the server
   
   ./http333d <port> ../projdocs/ unit_test_indices/*

2. Connect with telnet
   
   telnet <HostName> <port>

3. Write an HTTP request and send it

4. To exit telnet:
   
   ○ Ctrl+]] then Ctrl+d
Debugging for HW4

- Telnet
  - Form requests to your server
  - Look at the HTTP Responses
- Browser Developer Tools (see lecture for a [brief] demo)
  - Can compare (between provided solution and your implementation):
    - The HTML
    - HTTP Requests Sent
    - HTTP Replies
- Other strategies you have been using all quarter :) 
  - Examining intermediate values (parsing/building replies)
Possible Security Vulnerabilities in HW4

● Cross-Site Scripting
  ○ **Problem**: Clients can supply code for execution
  ○ **Occurs when**: Code is not escaped
  ○ **Solution**: Escape untrusted user input

● Directory Traversal Attack
  ○ **Problem**: Clients can gain unauthorized access to files
  ○ **Occurs when**: Insufficient sanitization of user supplied files paths
  ○ **Solution**: Verify a file is in subdirectory, before completing a request.
Boost

Boost is a free C++ library that provides support for various tasks in C++

- **Note:** Boost does NOT follow the Google style guide!!!

Boost adds many string algorithms that you may have seen in Java

- Include with `#include <boost/algorithm/string.hpp>`

We are showcasing a few we think could be useful for HW4, but more can be found here:

- [https://www.boost.org/doc/libs/1_60_0/doc/html/string_algo.html](https://www.boost.org/doc/libs/1_60_0/doc/html/string_algo.html)
void boost::trim(string& input);

- Removes all leading and trailing whitespace from the string
- input is an input and output parameter (non-const reference)

string s("   HI    ");
boost::algorithm::trim(s);

// results in s == "HI"
**replace_all**

```cpp
void boost::replace_all(string& input, const string& search, const string& format);
```

- Replaces all instances of `search` inside `input` with `format`

```cpp
string s("ynrnrt");
boost::algorithm::replace_all(s, "nr", "e");
// results in s == "yeet"
```
replace_all

```cpp
void boost::replace_all(string& input, const string& search, const string& format);
```

- Replaces all instances of `search` inside `input` with `format`

```cpp
string s("queue?");
boost::algorithm::replace_all(s, "que", "q");
// results in s == "que?"
```

`replace_all()` guarantees that ‘format’ will be in the final result if-and-only-if ‘search’ existed.

`replace_all()` makes a single pass over `input`. 
split

```cpp
void boost::split(vector<string>& output,
      const string& input,
      boost::PredicateT match_on,
      boost::token_compress_mode_type compress);
```

- Split the string by the characters in `match_on`
- Note: leading and trailing delimiter is ignored (treated as "")

```cpp
boost::PredicateT boost::is_any_of(const string& tokens);
```

- Returns predicate that matches on any of the characters in `tokens`
split Examples

vector<string> tokens;
string s("I-am--split");

boost::split(tokens, s, boost::is_any_of("-")
    boost::token_compress_on);
// results in tokens == ["I", "am", "split"]

boost::split(tokens, s, boost::is_any_of("-")
    boost::token_compress_off);
// results in tokens == ["I", "am", "", "split"]
**Exercise 1**

Write a function called `ExtractRequestLine` that takes in a well-formatted HTTP request as a string and returns a map with the keys as method, uri, version and the values from the corresponding request.

**Example Input:**

```
"GET /index.html HTTP/1.1\r\nHost: www.mywebsite.com\r\nConnection: keep-alive\r\nUpgrade-Insecure-Requests: 1\r\n"
```

**Map Returned:**

```
{
    "method" : "GET",
    "uri" : "/index.html",
    "version" : "HTTP/1.1"
}
```
Exercise 1

Write a function called `ExtractRequestLine` that takes in a well-formatted HTTP request as a string and returns a map with the keys as `method`, `uri`, `version` and the values from the corresponding request.

**Example Input:**

```
"GET /index.html HTTP/1.1\r\nHost: www.mywebsite.com\r\nConnection: keep-alive\r\nUpgrade-Insecure-Requests: 1\r\n"
```

**Map Returned:**

```json
{
  "method" : "GET",
  "uri"   : "/index.html",
  "version": "HTTP/1.1"
}
```
Exercise 1

```cpp
map<string,string> ExtractRequestLine(const string& request) {
    vector<string> lines;
    boost::split(lines, request, boost::is_any_of("\r\n"),
                 boost::token_compress_on);
    vector<string> components;
    string firstLine = lines[0];
    boost::split(components, firstLine, boost::is_any_of(" "),
                 boost::token_compress_on);
    map<string, string> res;
    res["method"] = components[0];
    res["uri"] = components[1];
    res["version"] = components[2];
    return res;
}
```
Exercise 2

Write a function RemoveDuplicates that takes in a string that contains words separated by whitespace and returns a vector that contains all of the words in that string, in the same order as they show up, but with no duplicates. Ignore all leading and trailing whitespace in the input string.

Example:
RemoveDuplicates(" Hi I'm sorry jon sorry hi\nhihi hi hi ")
should return vector:
["Hi", "I'm", "sorry", "jon", "hi", "hihi"]
Exercise 2

Write a function `RemoveDuplicates` that takes in a string that contains words separated by whitespace and returns a vector that contains all of the words in that string, in the same order as they show up, but with no duplicates. Ignore all leading and trailing whitespace in the input string.

**Example:**
`RemoveDuplicates(" Hi I'm sorry jon sorry hi hihi hi hi ")` should return vector:
`["Hi", "I'm", "sorry", "jon", "hi", "hihi"]`
```cpp
vector<string> RemoveDuplicates(const string& input){
    string copy(input);
    boost::algorithm::trim(copy);
    std::vector<string> components;
    boost::split(components, copy, boost::is_any_of(" 	\n"),
                   boost::token_compress_on);

    std::vector<string> result;
    std::set<string> unique_components;
    for (const auto& comp : components) {
        if (unique_components.find(comp) == unique_components.end()) {
            result.push_back(comp);
            unique_components.insert(comp);
        }
    }
    return result;
}
```