

Project #3: HTTP Server with Authentication

CS 352 Internet Technology

Released: October 17, 2024; Due: October 31, 2024

Instructions

Please read these instructions carefully before you begin.

1. You must work on this project preferably in teams of 2.
2. You are free to discuss the project on Piazza or through other means with your peers and the instructors. You may refer to the course materials, textbook, and resources on the Internet for a deeper understanding of the topics. However, you cannot lift solutions from other students or from the web including GitHub, Stack Overflow, generative AI (e.g., ChatGPT), or other resources. Do not post this project or parts of it to question-answering services like Chegg, academic hosting sites such as CourseHero, or generative AI. All written and programmed solutions must be your team's original work. We run sophisticated software to detect plagiarism and carefully monitor student answers. If you are in doubt, please ask us.
3. You cannot post this project specification or your solutions to the project on your personal GitHub page or any other public-facing services.
4. For each question in the project report, please be clear and concise. Vague and rambling answers will receive zero credit.
5. For the report question on references and collaboration, please include anyone you discussed the project with, and also any resources you consulted to learn how to solve this project, including URLs of pages visited on the Internet. Use of generative AI is not permitted in this course. Please be specific about the aspect of the project that you got help with. You must be thorough and as complete as possible here. It is mandatory to answer this question.
6. We encourage you to *start early* and get the bulk of the work for this project done the week(s) before it is due, rather than holding back your most significant efforts until the submission date is too close.
7. There is a due date/time but no "time limit" on projects. That is, you may take as long as you need to work on the project, as long as you submit it on time. Further, you may make an unlimited number of submissions on Canvas.
8. If you have any questions or need clarifications on the project, please post them on Piazza or contact the course staff. We are here to help.

Overview

In project 3, you will implement an HTTP server that can parse and understand HTTP requests, and craft HTTP responses appropriate for an application that serves secret user data. Access to the secret user data will be authenticated through two mechanisms: (i) a username and password, transmitted by the client to the server using HTTP POST requests, and (ii) cookies assigned by the server and later presented by the client from a prior successful authentication. We will build simple HTML entity bodies to render browser-readable data for each case. You will only work with one program in this project, `server.py`.

Step 1: Take `server.py` for a spin

The project archive comes with starter code for `server.py`. This file already implements an HTTP server which serves a login page. Start the server program by either typing

```
python3 server.py
```

which spins up the server on port 8080 (default), or you can type a port number of your choice:

```
python server.py 45006
```

An amazing thing about the HTTP protocol is that you can interact with the server endpoint using your browser. Spin up the local browser and type `http://localhost:45006` (substitute 45006 by the port number you used when you started the server, which may be the default value 8080.)

You may see something that looks like this:

Please login



Name:

Password:

You can also interact with the server using a command line client like `curl`. You could type the command `curl http://localhost:45006/` (or `curl ilab1.cs.rutgers.edu:45006` if you were running the program on `ilab1`) to perform a transaction through the command line. You will see the HTML content of the page printed on the terminal. You can get `curl` to print more information using the `-v` flag. Here is an example of what you may see when you invoke `curl -v ilab1.cs.rutgers.edu:45006`

```
* Trying 128.6.13.2:45006...
* Connected to ilab1.cs.rutgers.edu (128.6.13.2) port 45006 (#0)
> GET / HTTP/1.1
> Host: ilab1.cs.rutgers.edu:45006
> User-Agent: curl/7.81.0
> Accept: */*
>
* Mark bundle as not supporting multiuse
```

```

< HTTP/1.1 200 OK
< Content-Type: text/html
* no chunk, no close, no size. Assume close to signal end
<
<h1>Please login</h1>
  <form action = "http://ilab1.cs.rutgers.edu:45006" method = "post">
    Name: <input type = "text" name = "username"> <br/>
    Password: <input type = "text" name = "password" /> <br/>
    <input type = "submit" value = "Submit" />
  </form>
* Closing connection 0

```

(You can ignore error messages you may receive at the beginning from IPv6 connection attempts if you use the `localhost` URL.) If you go back to the terminal running the server program, it should have printed some helpful messages regarding the request it received and the responses it sent out. Example:

Here is the headers

```

"""
GET / HTTP/1.1
Host: ilab1.cs.rutgers.edu:45006
User-Agent: curl/7.81.0
Accept: */*
"""

```

Here is the entity body

```

"""

```

```

"""

```

Here is the response

```

"""
HTTP/1.1 200 OK
Content-Type: text/html

<h1>Please login</h1>
  <form action = "http://ilab1.cs.rutgers.edu:45006" method = "post">
    Name: <input type = "text" name = "username"> <br/>
    Password: <input type = "text" name = "password" /> <br/>
    <input type = "submit" value = "Submit" />
  </form>

"""

```

Served one request/connection!

Step 2: Study `server.py`

In many ways, `server.py` is not that different from any of the TCP servers you implemented in project 1 or project 2. The key difference is that it receives HTTP protocol request messages and responds with HTTP protocol response messages.

It will be worthwhile understanding how `server.py` unpacks the data in the HTTP request to extract the headers and the entity body of the request. You will build on this code to extract specific headers and information from the entity body.

The `server.py` code also contains various strings that are helpful to return as entity bodies of the HTTP responses, such as `login_page`, `bad_creds_page`, and so on.

Pay specific attention to the places marked `TODO`; these are the locations you will insert the main application logic for this project.

Step 3: Build the databases

The server should read all of the username and passwords stored in the file `passwords.txt` (in plain text) provided along with this project archive. This file will contain multiple lines, one per user of the system, with their username and password (always one word) in that order, separated by a space. For example, in the sample passwords provided, one of the users of the system is named `bezos` with the password `amazon`. Your program should read in all this user data and maintain a data structure for lookup when the user attempts to authenticate. Lookups are case-sensitive: `BEZOS` is a different user from `bezos`.

The server should also read in the corresponding user secrets stored in plain text in the file `secrets.txt`, provided along with the project archive. This file will contain multiple lines, one per user of the system, with their username and secret data (always one word) in that order, separated by a space. For example, in the sample secrets provided, one of the users of the system is named `bezos` with the secret word `kaching`. Your program should read in all this user data and maintain a data structure to look up after successful authentication.

We may test up to 10 users. The line corresponding to each user in each file may be as large as 300 characters.

Step 4: Implement username-password authentication

The browser login form on `http://localhost:45006/` is set up to send POST requests when these details are typed into the respective fields. (Beware, the file is stored in plaintext, and any information you type in the form is sent in plain text, so please do not put any sensitive information in these!)

You can also send POST data from the command line using the `curl` tool. If you type the command

```
curl -d "username=bezos&password=amazon" http://127.0.0.1:45006/
```

it has the same effect as form data posted from a browser window. The `server.py` terminal output should show

Here is the entity body

```
""  
username=bezos&password=amazon  
""
```

in place of the (empty) entity body from before. Specifically, the command flag `-d` makes `curl` send a POST request with the provided data inserted into the entity body of the HTTP request.

Parse the entity body obtained by `server.py` to recover the username and password fields of the HTTP POST request. Compare these details with the information you read from `passwords.txt` in the previous step. The HTTP request may be valid (according to the HTTP spec) but the entity body may still miss the appropriate data for server-side processing. Your server code must be careful not to assume that both the necessary POSTed fields (`username/password`) will always appear in the entity body, or appear exactly in that order.

Case A: Username-password auth success. If the username and password fields do exist in the entity body, and they match with the username and password of a user in the passwords file, you can return the `success_page` to this user with the corresponding secret information that you read in the previous step. An example of this output on the browser looks like this, after `bezos` successfully authenticates:

Welcome!

[Click here to logout](#)

Your secret data is here:

kaching

To accomplish this in the code, you merely need to set the variable `html_content_to_send` to `success_page + secret`, where `secret` is the secret word of the corresponding user who just logged in.

Case B: Username-password auth failure. If exactly one among the username or password fields is absent in the entity body (i.e., exactly one field is present), or if both fields are present but the username is not in the passwords file, or the password did not match the corresponding username in the passwords file, then we ask the user to log in again. This is accomplished by setting `html_content_to_send` to `bad_creds_page`. You may see output like this:

Bad user/pass! Try again

Name:

Password:

You should be able to see HTML source code corresponding to these pages using appropriate `curl` commands as well, in case you want to test your program quickly on the terminal.

Step 5: Generate, send, and store a cookie

Now we get to the part where the server “remembers” prior successful authentications using cookies. Recall that cookies are a collaborative mechanism between the client and the server. The server assigns an opaque identifier to a successfully authenticated user, and sends it back in a `Set-Cookie` header in the HTTP response.

Step 5.1 In case A in the previous step (successful username-password authentication), generate a cookie value which is a random 64-bit value. You can accomplish this by invoking `random.getrandbits(64)` in your code. Capture this header-value pair in a string that contains the entire HTTP header line. If you assign `headers_to_send` to this string, it will be sent to the client along with the rest of the message. An example of doing this is:

```
rand_val = random.getrandbits(64)
headers_to_send = 'Set-Cookie: token=' + str(rand_val) + '\r\n'
```

(The `token=` is helpful to have `curl` record and reuse cookies, as we will see later.) An example output of the full HTTP response in this case from the terminal output of `server.py` looks like this:

```
Here is the response
"""
HTTP/1.1 200 OK
Set-Cookie: token=4338535727857038191
Content-Type: text/html
```

```
<h1>Welcome!</h1>
<form action="http://localhost:45006" method = "post">
<input type = "hidden" name = "action" value = "logout" />
<input type = "submit" value = "Click here to logout" />
</form>
<br/><br/>
<h1>Your secret data is here:</h1>
kaching
"""
```

Step 5.2 Store the cookie `rand_val` that you sent in step 5.2 into another data structure that can look up the cookie to obtain the corresponding pre-authenticated user name, if any.

Storing and replaying cookies using `curl`. The `curl` tool allows recording and presenting cookies through its “cookie-jar” flags, `-c` and `-b`. Type the command

```
curl -d "username=bezos&password=amazon" \  
  -c cookies.txt -b cookies.txt \  
  http://127.0.0.1:45006/
```

and see the cookie token stored in the file `cookies.txt`.

Subsequent browser requests should contain the cookie. Once you have authenticated successfully from a browser after implementing the functionality above, if you open a new browser tab pointing to `http://localhost:45006/`, you must be able to see the `Cookie:` header printed among the request headers on `server.py`’s output on the terminal. Testing the functionality from scratch on the browser will require you to delete cookies for the specific server domain (`localhost` or other hostname) from your browser settings. The specific way in which you achieve this will depend on the browser. For example, if you are using Google Chrome, searching for the word “cookie” after you visit `chrome://settings` will provide links to resources that help you find and delete cookies for the specific domain you are using for your server. As an aside, to avoid having to manually delete cookies on your browser when debugging/testing your server’s cookie functionality, it may be easier to use the `curl` client rather than a browser.

Step 6: Implement cookie-based authentication

We are now ready to implement cookie-based authentication for users accessing our server. Browsers, and in general any HTTP clients implementing cookies correctly (including `curl` with cookie jar), are expected to present cookies provided by the server in subsequent HTTP requests to that server.

Your task now is to validate these cookies on the server side. From the HTTP request headers, extract the `Cookie:` header, and check if it corresponds to any user whose cookie you already recorded in step 5 above.

Case C. Cookie validated. If there is a `Cookie` header in the request, and the cookie is one of the pre-recorded cookies, we assume it comes from a user who was successfully authenticated earlier. (In this project, we will not worry about cookie hijacks, where a malicious entity steals cookies from a legitimate user and presents on their behalf, or where a malicious client checks all possible cookie values by brute force.) In this case, the output is similar to that of case A (step 4): welcome the user and show their secret content.

Case D. Cookie invalid. If there is a `Cookie` header in the request, and the cookie’s value isn’t one of the pre-recorded ones, you must present an error to the user similar to case B (step 4).

If there is no `Cookie` header in the request, you must process the request as you did in Step 4, handling cases A and B.

Summary of application logic. After completing this step, the flow of application logic should look like:

1. (case C) cookie header present and valid. Show the user the secret page: send back `success_page` with the secret.
2. (case D) cookie header present but invalid cookie: send back `bad_creds_page`
3. (case A) cookie header absent, username-password present in entity body, user-pass combination successfully matches correct credentials from passwords file: send back `success_page` with the secret, *remember to set a new cookie!*
4. (case B) cookie header absent, either one of username/password fields missing or username-password both present in entity body, but user-pass combination does not match any correct credential from the passwords file: send back `bad_creds_page`
5. (basic case, already implemented in the starter code) cookie header absent, username-password absent in entity body. Send back `login_page`

Step 7: Implement logout function

Your browser will keep presenting the same cookie indefinitely if cookies aren't cleared or don't have an expiry date. The cookies we sent back in step 5 do not have an expiry date. Hence, a user, once logged in, must manually clear cookies each time they want to log out and avoid secret data being served by `server.py`. (As mentioned earlier, most browsers will allow cookies to be deleted on a per-domain basis through the browser settings.)

In this step, you will send back a cookie header from the server that explicitly corresponds to clearing the cookie from the client. This is accomplished by setting an expiry date for the cookie that is in the past. For example, if you set `headers_to_send` to the string

```
Set-Cookie: token=; expires=Thu, 01 Jan 1970 00:00:00 GMT\r\n
```

it instructs the HTTP client to clear its cookie named `token` for the server's domain, since the cookie is past its expiry date. If your server sends this header in the response, you can check on your browser's cookie list as well as `curl`'s cookie jar file (`cookies.txt` in the examples above) that the cookie has disappeared. (I have sometimes observed that `cookies.txt` itself disappears if there are no other cookies stored in it.)

Case E. Logout. Helpfully, the welcome HTML page that appears when a user is successfully authenticated contains a logout action button, that POSTs a hidden object called `action` with value `logout` to the server when the button is clicked. To implement a log-out function (clearing cookies in the process), all you need to do on the server side is to detect `action=logout` in the POST entity body, and send an expired cookie to the client when that happens. Further, send the HTML content corresponding to `logout_page` back to the user indicating they have logged out.

After implementing this step, the flow of application logic is enhanced by implementing log-out functionality at the beginning of the overall logic flow described in the summary at the end of the previous section. That is, the sequence of the logic in your program should be:

1. case E
2. case C
3. case D

4. case A
5. case B
6. basic case (already implemented in starter)

What you must submit and how we will test it

For your project submission on Canvas, please turn in `server.py` and your project report `report.pdf` in one Zip file. The command line of the program must be exactly the same as it was given to you (same as in step 1). The questions for the report are listed below. We will be running your code on the `ilab` machines with the default Python 3 version on those machines. **Please compress the files into a single Zip archive before uploading to Canvas.** Only one team member must submit.

Testing your program

We will test your program at least with the following cases and more. We may run the client on the same or a different machine from the one running your `server.py`. You may use this as a guide to check that your program implements all the required functionality correctly (we also mention the desired end result in parentheses). In all tests, all input files will be well formed, and we will only send well-formed HTTP packets crafted through a browser or `curl`. You should be able to craft requests corresponding to these cases through either your browser, or more likely, `curl`.

1. basic: no username or password posted, no cookies (login)
2. correct username and password posted, no cookies (success)
3. non-existent username posted with password, no cookies (bad credentials)
4. existing username posted with bad password, no cookies (bad credentials)
5. exactly one of username or password posted, other field missing, no cookies (bad credentials)
6. no username or password posted, valid cookie (success)
7. non-existent username or bad password for existing username, valid cookie (success)
8. correct username and password, valid cookie (success)
9. correct username and password, invalid cookie (bad credentials)
10. logout posted, valid cookie (logout)
11. logout posted, invalid cookie (logout)

Project report

Please answer the following questions for the project report.

1. Team details: Clearly state the names and netids of your team members (usually there are 2 of you).
2. Collaboration: Who did you collaborate with on this project? What resources and references did you consult? Please also specify on what aspect of the project you collaborated or consulted.
3. Is there any portion of your code that does not work as required in the description above? Please explain.

4. Did you encounter any difficulties? If so, explain.
 5. Please discuss **two** observations or facts you learned about HTTP and cookies in the process of working on this project. Please be specific and technical in your response.
- Contact the course staff on Piazza if you have any questions. Start early.