CS F303 Computer Networks: 2020-21 semester II

Assignment 1

This assignment is to be done individually. The deadline for this assignment is 25th February, 2021, 23:59 hours. Submissions have to be made on CMS. **Your code will be passed through standard code similarity checkers. If a high match is found between the submissions of two or more students, all involved students will summarily be awarded a grade penalty. No request for concessions will be entertained in this matter by anyone.**

The assignment will be graded on 30.

Assignment Statement:
We often access the web through a proxy. Squid is a caching and forwarding HTTP web proxy. In this assignment, you have to write a program using sockets in the C language, without using any high level libraries, to download the **page** and **image** from http://info.in2p3.fr/ through a squid proxy, which has been set up by the instructors. The proxy configuration is as follows:
Server: 182.75.45.22
Port: 13128
Login: csf303
Password: csf303

If you receive a HTTP 30x response you have to handle that and follow the redirection. You have to figure out the HTTP protocol for proxy access on your own. It will help to set your browser to the above proxy, access http://info.in2p3.fr/ from the browser and look at Wireshark traces from your browser while the request is being made and responded to. Also refer to the HTTP 1.1 RFC. For simplicity, just write a single program http_proxy_download.c, which should run as follows:

```
$ gcc http_proxy_download.c -o http_proxy_download.out
$ ./http_proxy_downoad.out info.in2p3.fr 182.75.45.22 13128 csf303 csf303 index.html logo.gif
```

where, as would be obvious, the command line arguments are respectively the URL, proxy IP, proxy port, login, password, filename to save html as, filename to save logo as.

Do not hardcode any parameter in your program! Your program should work even if the proxy IP changes, or the login/password changes, or image is changed, or the base URL is changed.

Please note:
1. Only static websites will be part of the test cases
2. Https will not need to be handled either directly or as redirects
3. Image download will only be tested for http://info.in2p3.fr/
4. A sample testing script has been provided on Piazza

You will submit just the single http_proxy_download.c file. **The program must compile and run on Ubuntu 18.04 or Ubuntu 20.04.** If you need access to these environments, please write on Piazza. The first line of the file should contain the BITS email and name of the student **exactly in the prescribed format below.** Also briefly describe your approach in comments following student details.

Test cases:
1. Correctly downloading the image from http://info.in2p3.fr/. This is the only image we will test for. For other websites image download will not be tested. You can hardcode the location and name of the image file in the base html for this test case. However, in the unlikely event that they change the image or the filename it should still work.
2. Correct download of a second website (image will not be checked)
3. Correction download of a third website (image will not be checked)
4. Correct handling of http redirects

We have allowed the following domains to be accessed through the proxy to enable you to test your code:

- .google.com
- .india.gov.in
- .bits-pilani.ac.in
- .in2p3.fr
- .go.com
- .jandarshan.cg.nic.in

We will use automated evaluations on Ubuntu systems and marking will be binary (either it works or doesn't), so it is very important to ensure that you adhere to the guidelines. If you do not adhere to the guidelines and our parser fails to extract your names you will not be marked for this assignment. We will also use heuristics in our parser to check that you have used the correct libraries, have not used higher level libraries, and have no hardcoded parameters. **There will be no manual rechecking for part marking if your program fails the tests.** We will release a test script which you can use to test that the names and IDs are being correctly parsed.

Adhere to the following format at the head of your program (numbers on the left are line numbers, not part of the program):

```
1. /* f20151234@hyderabad.bits-pilani.ac.in John Doe */
2.
3. /* Brief description of program...*/
4. /* ... */
5.
6. # include, etc.
```