CS 455 Programming Homework 2
Design and implement a Proxy Server
Due Midnight Tuesday 9/22 on Angel

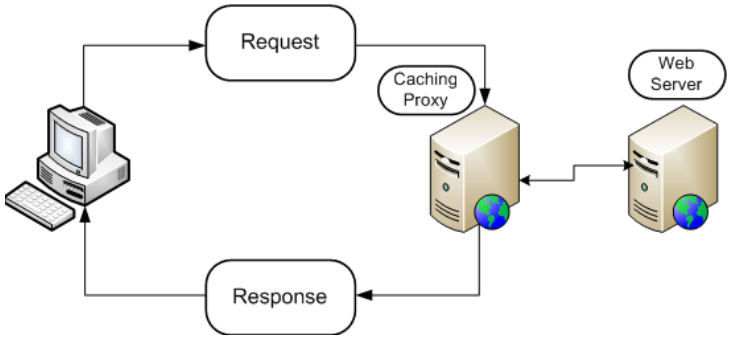
## 1. Introduction

In this programming homework, you will design and implement an HTTP proxy cache server in Python. *The main purpose is to understand how http and caching work.* Upon completion, you can configure your web browser to use your proxy to surf the web! A python template file is provided and can be used as a starting point. The next assignment will base on this assignment.

You may want to review the following materials if needed.

- Section 2.2 in the textbook
- http://tools.ietf.org/html/rfc2616 - detail description about the HTTP protocol. The proxy cache server will only support GET method in this assignment.

## 2. Programming Description




The above figure shows how proxy server works. Conceptually, the proxy sits between the client and the server. In the simplest case, instead of sending requests directly to the server the client sends all its requests to the proxy.

The proxy works as follows.

- The proxy listens for requests from clients
- When there is a request, the proxy spawns a new thread for handling the request
- The new thread checks if the request file is in the cache.
    a. If the file is in the cache, it sends an http request to the web server to check if the file is still updated. The web server will either responds with the newer copy of the file or a status code indicating that the file is updated. The proxy server then sends the file to the client and update the cache if necessary.
    b. If it is not, it contacts the web server for the file and then sends the file back to the client and update the cache.
- The proxy turns off the keep-alive for all http packets

**Important points**

- Please take time to design the program first before implementing it. The design should at least address data structure, concurrency issues, and error cases.
- While the proxy only handles GET requests from clients, it should handle all kinds of responses from origin web servers.
- For this assignment, you can assume that the proxy server can store all requests and responses before forwarding them.

## 3. Browser setup and evaluation

You will need to configure your web browser to use your proxy. How to do it depends on your specific browser. Most current browsers have their own local cache. Ask your best friend, Google ☺, how to disable local cache and set a proxy to your browser.

Use Wireshark to measure performance in terms of time with and without proxy (plot a graph) when requesting the following websites

- www.cnn.com
- www.vancouver.wsu.edu
- http://techcrunch.com

## 4. Submission guidelines and grading criteria

### a. Submission

Provide sufficient comments for your code. Write up a report that includes the design of the proxy server, the graph and comments in Section 3, and discussions on several factors that may slow down your proxy cache server how you can change the design to improve its performance.

**Soft copy (on Angel):** The report file, the code, and a README file that describes how to run your code. The code is expected to run on the lab machines.

### b. Grading criteria

| Criteria | Grade |
|---|---|
| Basic setup to relay Http GET requests and responses | 30 |
| Correct cache setup (e.g., data structure, access control…) | 30 |
| Support concurrent requests | 20 |
| Proper handling of all response codes from the web server | 10 |
| Comments, discussion, code quality, and documentation | 10 |