Networks and Distributed Systems **Homework 6**
CS3700 Fall 2022

This homework is due at 11:59:59 PM on November 14, 2022 and is worth 3% of your grade.

Name: ____

NUID (with leading zeros): ____

Problem	**Possible**	**Score**
1	20	
2	10	
3	20	
4	30	
5	20	
Total	100	

1a. What are the two primary architectures of Distributed Systems? Explain how the differ. (10 pts)

1b. List the Pros and Cons of using TCP versus UDP as the transport protocol in a distributed system. Would it be smarter to make your own transport protocol? Justify your response. (10 pts)

2a. What is the main difference between HTTP and HTTPS? (5 pts)

2b. Can a web server use both HTTP and HTTPS? Explain your reasoning. (5 pts)

3. Many services "crawl" the web in order to provide useful services; the most common example of this is web search engines like Google. However, the operators of websites often wish to express what parts of their website should and should not be crawled. We're going to explore this functionality in this question.

3a. The robots.txt file is one way this can be accomplished. What is the format of this file, and where should it be placed on your website so that Google et al. can find it? (10 pts)

3b. Locate the robots.txt file for github.com. Give two examples of pages on Github's website that Google would be allowed to index, and two examples that Google would not be allowed to index. (10 pts)

4a. What is the `User-Agent` HTTP header? How is it used by web servers? (5 pts)

4b. Suppose that we built a custom web browser, and desired to only allow users who were running this particular web browser to visit our site (i.e., we did not want to allow users on Google Chrome to access it for security reasons). Would the `User-Agent` header be a good way of accomplishing this goal? Why or why not? (10 pts)

4c. Recall that the HTTP `Referer` header tells the server which web page "referred" it to the current request. However, this header field has raised a number of privacy concerns. Give an example of privacy issues on sites like Facebook or Twitter that is caused by the `Referer` header. (10 pts)

4d. If you were an operator of a site like Facebook, how might you ensure that the users who click on links to external sites from your site are not subject to these privacy issues? (5 pts)

5. For the next set of questions, you'll be using the developer tools in your browser. Note that for these questions you will need to disable any ad blocking or privacy-preserving browser extensions you have installed. Alternatively, use a browser that does not have any of these tools installed.

First, you should clear your browsers cache; typing CTRL-SHIFT-Delete opens the clear dialog in Firefox and Chrome. Next, in Firefox, go to Menu -> Web Developer -> Toggle Tools. In Chrome, go to Menu -> More Tools -> Developer Tools. Alternatively, in both browser you can type CTRL-SHIFT-I. Once the developer tools are open, go to the "Network" tab, then browse to `cnn.com`.

5a. Wait a few moments for CNN to finish loading. How many total requests did it take for the CNN website to load? Note: you don't need to scroll the page or interact with elements on the page, this will cause even more requests to be sent. (5 pts)

5b. How many requests were for HTML? How many for images? How many for JavaScript? (Hint: the filters in the "Network" tab will help.) (5 pts)

5c. Many of the resources that end up being included into the CNN homepage come from third-parties, i.e., companies that are not CNN. As representative examples, find some requests to *outbrain.com*, *googletagservices.com*, and *krxd.net* and click on them to inspect the HTTP requests and responses. What are these services doing? Why has CNN included them in their homepage (10 pts)