

This homework is due at the beginning of class on November 2, 2016 and is worth 1.5% of your grade.

Name: _____

CCIS Username: _____

Problem	Possible	Score
1	20	
2	25	
3	20	
4	30	
Total	95	

- 1a. Imagine a file is being shared via BitTorrent, and the file is divided into 8 BitTorrent pieces $\{A, B, C, D, E, F, G, H\}$. Now, suppose that you and two other leechers $\{l_1, l_2\}$ are in the system, and that everyone is connected to everyone else. The bitfields are as follows:

Leecher	Bitfield [ABCDEFGH]
you	[11100100]
l_1	[00000101]
l_2	[00011110]

Assume that each leecher can send and receive two pieces in each round. Given that you must upload something to each l_1 and l_2 in each round in order for them to upload something to you, which bitfield should you advertise to l_1 and l_2 in the next round in order to ensure that they stay interested in you for the longest amount of time? *Note that you can lie about your bitfield, and can reveal different bitfields to l_1 and l_2 .* (10 pts)

- 1b. Assuming ideal conditions for you, what is the minimum number of rounds until you have downloaded the entire file? (5 pts)

- 1c. Why does network address translation (NAT) make life difficult for peer-to-peer applications (such as BitTorrent), but does not generally affect client-server applications (such as Web servers)? (5 pts)

2. In this question, we're going to use the `telnet` program at the command line to manually make a HTTP request. It should be available on most modern machines, as well as all of the CCIS Linux machines. If you're not familiar with `telnet`, read the Linux manual page before beginning this question.

2a. Using `telnet`, connect to the machine `ssl-research.ccs.neu.edu` on the HTTP port. Issue a HTTP v1.0 request for the file `/`, providing no request headers. What is the response code, and what is the server telling you to do? Why might this be the case? (10 pts)

2b. The previous request did not work out the way we expected. Now, make a HTTP v1.1 request for `/` to the same site. What response code do you get now? How much content does the server give you? *Hint: There is one important header you need to be sure to include.* (10 pts)

2c. Now, repeat the previous request but tell the server that it is allowed to compress the content using `gzip`. How do you do this? How big is the content that is returned (on the wire)? (5 pts)

3. Many web sites “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 sites often wish to express what parts of their web pages 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 Northeastern’s main web site. Give an two example of different pages on Northeastern’s web site 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 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)