CSC 591: Privacy (Fall 2019)
Home Assignment #3
Assigned: Tuesday, Oct. 1, 2019, Due: Tuesday, Oct. 15, 2019

Instruction: Completed homework should be typed (e.g., using LaTeX or word document) or hand-written clearly and scanned and uploaded into Moodle. You can discuss about how to use certain tools for data collection and analysis, but no collaboration is permitted to solve the problems.

1. Visit the following browser fingerprinting sites first from a Firefox Browser and then from a Tor Browser. What are the major differences that you see in each case? What is the basic takeaway from all the three comparisons? (provide screenshots of the tracking characteristics for the different visits highlighting the major differences). [points (5+5+5) +5]

    a. https://panopticlick.eff.org/
    b. https://amiunique.org/

2. Ghostery is a tracker blocking extension. Ghostery also has its own mobile browser that enables tracker blocking on mobile platforms. **Report** the number of trackers you see on desktop platform versus mobile platform (remember to go to setting to block all trackers, by default Ghostery doesn't block all trackers). **Identify** desktop-specific and mobile-specific trackers, if any (for each site visited). [points 5+5+5]

    Output should look like the following table:

Site	# of trackers found while using a desktop	# of trackers found while using a mobile	Desktop specific tracker (if any)	Mobile specific tracker (if any)

    Sites to visit –
    a. cnn.com
    b. alibaba.com
    c. espn.com
    visit: https://www.ghostery.com/products/?utm_source=ghostery.com&utm_campaign=products_menu

3. Visit the following sites and **list** the various third-party domains from which HTML objects are loaded. Then **construct** a connectivity graph showcasing the different **third-party domains** that are called while loading the following site (something like lecture 10 slide 53). [points 10+10]

    Sites to visit-
    a. macys.com
    b. cnn.com
    c. bankofamerica.com

    Hint: you can use the Google Developer tool (more tools->developer tool->network tab) to view the various http requests/responses generated while visiting a site. For Firefox you can use the web developer->inspector tool ->network tab). Remember to disable all privacy settings (like enable 3rd-party cookies, disable tracker blocking or any other extensions that you might be using; if you like you can launch a pristine VM for this task too). Then you can download the HAR files from the browser and analyze it using the following parser https://gist.github.com/tomatohater/8853161. To generate the graph, you can use any online tool (for example https://graphonline.ru/en/). You need to upload the HAR files and code used for generating the graph (with proper README).

CSC 591: Privacy (Fall 2019)
Home Assignment #3
Assigned: Tuesday, Oct. 1, 2019, Due: Tuesday, Oct. 15, 2019

4. AdBlock Plus uses the EasyList (https://easylist.to/easylist/easylist.txt) to filter unwanted web contents (each line is a filtering rule, so you can simply parse the file by extracting each line and putting it into an array). `adblockparser` is a package for working with AdBlock Plus filter rules (https://adblockplus.org/filter-cheatsheet). It can parse AdBlock Plus filters and match URLs against them (https://github.com/scrapinghub/adblockparser). Using the HAR files from question 3, write a code that will list the number of contents that should be blocked when visiting a site. Remember to pass the right options when checking if a URL should be blocked. For example, while visiting *example.com* if you see a HTTP request for *sample.com* then that request is a third-party request. Also, the HAR file contains the content type (like image or script) for each HTTP request. This information should also be passed as options when checking whether an URL should be blocked or not. Output should be in the following manner. [points 25]

Site	# of total HTTP requests	# of HTTP requests blocked	Third-party **domains (not URL)** blocked

   You need to upload the code used for generating the output (with proper README)

5. A list of Tor relays is available in the following link (https://torstatus.blutmagie.de/). Report the followings:
   a. List the top 5 countries hosting Tor relays [points 5]
   b. List the top 5 bandwidth-contributing relays [points 5]
   c. Venn diagram of the number of relays that act as guard, exit or middle relay. Also report the cumulative bandwidth for each category. (you can use any tool you like to generate the venn diagram; for example, https://python-graph-gallery.com/venn-diagram/) [points 10]

   You need to upload the code used for generating the output (with proper README)

Submission:

You have to submit the following files:

1. Merge all the written parts into a single pdf file <your unity id>_HW3.pdf (make sure you submit PDF file only)
2. Rename the program files (.c/.cpp/.java/.py) you used as <your unity id>_QX.extension. (for question no. X) [don't submit ipython script)

Zip all files into <your unity id>_HW3.zip and submit the zip file on Moodle.