**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 site first from a Firefox Browser and then from a Tor Browser.
   1. What are the major differences that you see for the **two browsers**? [points 5]
   2. What is the basic takeaway for the two browsers? (compare screenshots for the two browsers **highlighting** the major differences for a given fingerprinting sites). Include the screenshots in your answer. [points 5]

<https://panopticlick.eff.org/>

1. You are given the HAR files (HTTP Archive format) for the following sites: *macys.com* and *cnn.com*.
   1. How many **third-party domains** that are loaded while visiting these sites? [points 15]
   2. Draw a Venn diagram showing overlap among the two sites? [points 10]
   3. List the third-party domains that are common across the two sites. [points 5]

You can analyze the HAR files using the following parser <https://gist.github.com/tomatohater/8853161>.

You compute third-party domain you can use *get\_fld* (<https://pypi.org/project/tld/>)

**You need to upload the code used for generating the result** (with proper README).

1. AdBlock Plus uses the EasyList (you are given the EasyList inside the assignment) 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 python package for working with AdBlock Plus filter rules. It can parse AdBlock Plus filters and match URLs against them (<https://github.com/scrapinghub/adblockparser>). Using the HAR files from question 2, write a code that will list the number of contents that should be blocked when visiting a site. Remember to pass **all** 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 this 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. Your report should be formatted in the following manner. [points 40]

|  |  |  |
| --- | --- | --- |
| Site | # of total HTTP requests | # of HTTP requests blocked |
|  |  |  |

Passing all options: rules.should\_block("http://ads.example.com/notbanner", **{'script': True, third-party: True}**)

**Script, image, domain, third-party options are most important**

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

1. A list of Tor relays is given to you. Report the followings:
   1. List the top 5 countries hosting Tor relays [points 5]
   2. List the top 5 bandwidth-contributing relays [points 5]
   3. Venn diagram of the **number** of relays that act as guard, exit or middle relay. Also report the **cumulative** **bandwidth (verified bandwidth)** for each category in the plot. (you can use any tool you like to generate the Venn diagram; for example, <https://python-graph-gallery.com/venn-diagram/>). A relay that is neither a guard nor an exit is assumed to be a middle relay. Comment on the distribution of relay types. [points 6+4]

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

**Submission:**

You have to submit three files:

1. Merge all the written parts into a single pdf file named <your unity id>\_HW3.**pdf**.

2. Rename the program file you used for as <your unity id>\_HW3\_QX.extension (e.g., .c/.cpp/.java/.py).

3. Add a README file regarding how to run your code.

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