Problem 1:

The fingerprinting website https://panopticlick.eff.org/ was visited on a Firefox browser as well as a Tor browser. In my instance, the Privacy Badger extension was enabled in the Firefox browser.

Here are the screenshots of the reports from panopticlick. The subparts of the problem are explained after the screenshots, with relevant information.

Firefox:

Test	Result
Is your browser blocking tracking ads?	✓ yes
Is your browser blocking invisible trackers?	✓ yes
Does your blocker stop trackers that are included in the so-called **"acceptable ads"** whitelist?	✓ yes
Does your browser unblock 3rd parties that promise to honor **Do Not Track**?	✗ no
Does your browser protect from **fingerprinting**?	✗ your browser has a unique fingerprint

Your browser fingerprint **appears to be unique** among the 235,794 tested in the past 45 days.

Currently, we estimate that your browser has a fingerprint that conveys **at least 17.85 bits of identifying information.**

Browser Characteristic	bits of identifying information	one in x browsers have this value	value
User Agent	16.85	117897.0	Mozilla/5.0 (Macintosh; Intel Mac OS X 10.15; rv:71.0) Gecko/2010010 1 Firefox/71.0
HTTP_ACCEPT Headers	1.96	3.9	text/html, */*; q=0.01 gzip, deflate, br en-US,en;q=0.5
Browser Plugin Details	1.27	2.42	undefined
Time Zone Offset	3.25	9.5	240
Time Zone	3.58	11.93	America/New_York
Screen Size and Color Depth	5.94	61.45	1680x1050x24
System Fonts	3.88	14.69	Andale Mono, Arial, Arial Black, Arial Hebrew, Arial Narrow, Arial Round ed MT Bold, Arial Unicode MS, Comic Sans MS, Courier, Courier New, Geneva, Georgia, Helvetica, Helvetica Neue, Impact, LUCIDA GRAND E, Microsoft Sans Serif, Monaco, Palatino, Tahoma, Times, Times New Roman, Trebuchet MS, Verdana, Wingdings, Wingdings 2, Wingdings 3 (via javascript)
Are Cookies Enabled?	0.18	1.13	Yes
Limited supercookie test	1.05	2.07	DOM localStorage: Yes, DOM sessionStorage: Yes, IE userData: No, o penDatabase: false, indexed db: true
Hash of canvas fingerprint	17.85	235794.0	2dabe85f0d9093e09fd4dab3918e0291
Hash of WebGL fingerprint	8.96	499.56	d1e597a0c8a4f515187a5cea1923a8ba
WebGL Vendor & Renderer	7.18	145.1	Intel Inc.~Intel(R) UHD Graphics 630
DNT Header Enabled?	1.04	2.06	True
Language	0.98	1.97	en-US
Platform	3.04	8.23	MacIntel
Touch Support	0.73	1.66	Max touchpoints: 0; TouchEvent supported: false; onTouchStart support ed: false
Ad Blocker Used	0.38	1.3	False
AudioContext fingerprint	2.74	6.66	35.7383295930922
CPU Class	0.15	1.11	N/A
Hardware Concurrency	5.56	47.09	16
Device Memory (GB)	0.74	1.68	N/A

Tor Browser:

Test	Result
Is your browser blocking tracking ads?	⚠ partial protection
Is your browser blocking invisible trackers?	⚠ partial protection
Does your blocker stop trackers that are included in the so-called **"acceptable ads"** whitelist?	no
Does your browser unblock 3rd parties that promise to honor **Do Not Track**?	no
Does your browser protect from **fingerprinting**?	⚠ partial protection

Within our dataset of several hundred thousand visitors tested in the past 45 days, only **one in 8732.56 browsers have the same fingerprint as yours.**

Currently, we estimate that your browser has a fingerprint that conveys **13.09 bits of identifying information.**

Browser Characteristic	bits of identifying information	one in x browsers have this value	value
User Agent	3.36	10.25	Mozilla/5.0 (Windows NT 10.0; rv:78.0) Gecko/20100101 Firefox/78.0
HTTP_ACCEPT Headers	1.96	3.9	text/html, */*; q=0.01 gzip, deflate, br en-US,en;q=0.5
Browser Plugin Details	1.28	2.42	undefined
Time Zone Offset	2.49	5.61	0
Time Zone	2.68	6.42	UTC
Screen Size and Color Depth	6.19	73.25	1000x900x24
System Fonts	7.75	215.13	Arial, Courier, Geneva, Georgia, Helvetica, Helvetica Neue, LUCIDA GRANDE, Monaco, Tahoma, Times, Times New Roman, Verdana (via javascript)
Are Cookies Enabled?	0.18	1.13	Yes
Limited supercookie test	1.05	2.07	DOM localStorage: Yes, DOM sessionStorage: Yes, IE userData: No, openDatabase: false, indexed db: true
Hash of canvas fingerprint	2.83	7.13	randomized
Hash of WebGL fingerprint	4.06	16.7	randomized
WebGL Vendor & Renderer	3.36	10.26	None
DNT Header Enabled?	0.96	1.94	False
Language	0.98	1.97	en-US
Platform	3.04	8.23	MacIntel
Touch Support	0.73	1.66	Max touchpoints: 0; TouchEvent supported: false; onTouchStart supported: false
Ad Blocker Used	0.38	1.3	False
AudioContext fingerprint	3.62	12.33	not available
CPU Class	0.15	1.11	N/A
Hardware Concurrency	2.39	5.23	2
Device Memory (GB)	0.74	1.68	N/A

a. The major differences I see from the two reports is that panopticlick states that the Firefox browser conveys 17.5 bits of identifying information. This is a higher value than that of what the Tor browser conveys (13.09 bits). This shows that using a tor browser increases source anonymity and makes it harder for the source to be uniquely identifiable.

   Another important difference that was observed from the reports is that the Firefox browser blocks tracking ads and invisible trackers. The tor browser only reports partial

protection from these. This is attributed to the Privacy Badger extension that was enabled on Firefox.

b. The key takeaway for the browsers is as explained below.
The Firefox browser blocks tracking ads, invisible trackers, as well as stops trackers that are included in the acceptable ads list. This is all attributed to the Privacy Badger Extension that is enabled as outlined at the start of the question. The report is shown below for Firefox.

Is your browser blocking tracking ads?	✓ yes
Is your browser blocking invisible trackers?	✓ yes
Does your blocker stop trackers that are included in the so-called **"acceptable ads"** whitelist?	✓ yes

At the same time, when you look at the corresponding reports for the Tor browser, it only provides partial protection from tracking ads and invisible trackers, with no protection to stop trackers that are included in the acceptable ads list. This is attributed to no extensions being enabled in the tor browser. The corresponding section of the report is shown below.

Is your browser blocking tracking ads?	⚠ partial protection
Is your browser blocking invisible trackers?	⚠ partial protection
Does your blocker stop trackers that are included in the so-called **"acceptable ads"** whitelist?	no

The second takeaway for the two browsers is that even though Firefox provides seemingly better protection from trackers, its browser fingerprint is much more unique than that of the Tor browser. The sections of the report that show this are displayed below.

For the Firefox browser:

Your browser fingerprint **appears to be unique** among the 235,794 tested in the past 45 days.

Currently, we estimate that your browser has a fingerprint that conveys **at least 17.85 bits of identifying information.**

For the Tor browser:

Within our dataset of several hundred thousand visitors tested in the past 45 days, only **one in 8732.56 browsers have the same fingerprint as yours.**

Currently, we estimate that your browser has a fingerprint that conveys **13.09 bits of identifying information.**

For context, consider the year 2007, when the population of the world was around 7 billion people. This means that the amount of entropy that is required to identify a human is about S = log2(1/6625000000) = 32.6 bits, which is approximately 33 bits of information.
Source: https://www.eff.org/deeplinks/2010/01/primer-information-theory-and-privacy

Clearly, we can see that the Mozilla browser is more uniquely identifiable than the Tor browser.

The third takeaway is from the detailed reports section of Panopticlick. If the user-agent field is observed from the Firefox browser and that of the Tor browser, we can see that the Firefox report highlights the exact machine that I was using (MacBook running OSX). However, the tor browser obfuscated this and showed that the browser was windows NT. See below that highlights this difference.

Firefox:

User Agent	16.85	117897.0	Mozilla/5.0 (Macintosh; Intel Mac OS X 10.15; rv:71.0) Gecko/20100101 Firefox/71.0

Tor:

User Agent	3.36	10.25	Mozilla/5.0 (Windows NT 10.0; rv:78.0) Gecko/20100101 Firefox/78.0

Finally, the fourth takeaway is that the Firefox browser was uniquely identifiable by using the canvas fingerprinting techniques. The GPU that was used to render this page was also uniquely identifiable. However, all these values were not available in the Tor report since it was randomized or obfuscated. Please see the corresponding reports below.

Firefox:

Hash of canvas fingerprint	17.85	235794.0	2dabe85f0d9093e09fd4dab3918e0291
Hash of WebGL fingerprint	8.96	499.56	d1e597a0c8a4f515187a5cea1923a8ba
WebGL Vendor & Renderer	7.18	145.1	Intel Inc.~Intel(R) UHD Graphics 630
DNT Header Enabled?	1.04	2.06	True

Tor:

Hash of canvas fingerprint	2.83	7.13	randomized
Hash of WebGL fingerprint	4.06	16.7	randomized
WebGL Vendor & Renderer	3.36	10.26	None
DNT Header Enabled?	0.96	1.94	False

Problem 2:

The .har files were parsed using a custom program for the domains cnn.com and macys.com and the questions that were asked were attempted. The results are described below.

a. The total number of third-party domains for file www.cnn.com.har is 134. The domains that are third-party are shown below as a screenshot that is taken directly from the program that generated these results.

   The total number of third-party domains for macys.com is provided after the first set of screenshots.

Domains
googletagservices.com
amazon-adsystem.com
criteo.net
cookielaw.org
outbrain.com
optimizely.com
doubleclick.net
jsdelivr.net
beemray.com
ugdturner.com
adsafeprotected.com
indexww.com
krxd.net
chartbeat.com
bing.com
bounceexchange.com
ads-twitter.com
tru.am
boomtrain.com
segment.com
demdex.net
scorecardresearch.com
tree.com
cloudflare.com
google.com
bleacherreport.net
bootstrapcdn.com
googletagmanager.com
google-analytics.com
lendingtree.com
rlcdn.com
rkdms.com
adsrvr.org
rubiconproject.com

```
|      rubiconproject.com      |
|          adnxs.com           |
|             t.co             |
|         usabilla.com         |
|       everesttech.net        |
|          segment.io          |
|         yieldmo.com          |
|       outbrainimg.com        |
|          3lift.com           |
|       casalemedia.com        |
|         facebook.net         |
|    googlesyndication.com     |
|       imrworldwide.com       |
|          turner.com          |
| d9t9vcvz5fqud.cloudfront.net |
|         facebook.com         |
|            cnn.io            |
|         onetrust.com         |
|         bluekai.com          |
|         zemanta.com          |
|         im-apps.net          |
|           agkn.com           |
|         mfadsrvr.com         |
|        bidswitch.net         |
|          criteo.com          |
|          navdmp.com          |
|         adition.com          |
|        powerlinks.com        |
|          eyeota.net          |
|         exelator.com         |
|          geistm.com          |
|          bttrack.com         |
|        crwdcntrl.net         |
|       creativecdn.com        |
|        quantserve.com        |
|          trustx.org          |
|          moatads.com         |
```

```
|          turn.com          |
|          2mdn.net          |
|        mathtag.com         |
|        sharedid.org        |
|         atdmt.com          |
|         yahoo.com          |
|         dyntrk.com         |
|           1rx.io           |
|         media.net          |
|     smartadserver.com      |
|    fonts.googleapis.com    |
|       sitescout.com        |
|        impdesk.com         |
|         socdm.com          |
|        gstatic.com         |
|      flashtalking.com      |
|         truste.com         |
|        bouncex.net         |
|        tidaltv.com         |
|         tapad.com          |
|         emxdgt.com         |
|        adswizz.com         |
|      advertising.com       |
|        addthis.com         |
|      spotxchange.com       |
|         fwmrm.net          |
|        videohub.tv         |
|        pubmatic.com        |
|        truoptik.com        |
|     yieldoptimizer.com     |
|        iasds01.com         |
|        twitter.com         |
|       chartbeat.net        |
|     summerhamster.com      |
|        trustarc.com        |
|         behave.com         |
|      adsymptotic.com       |
```

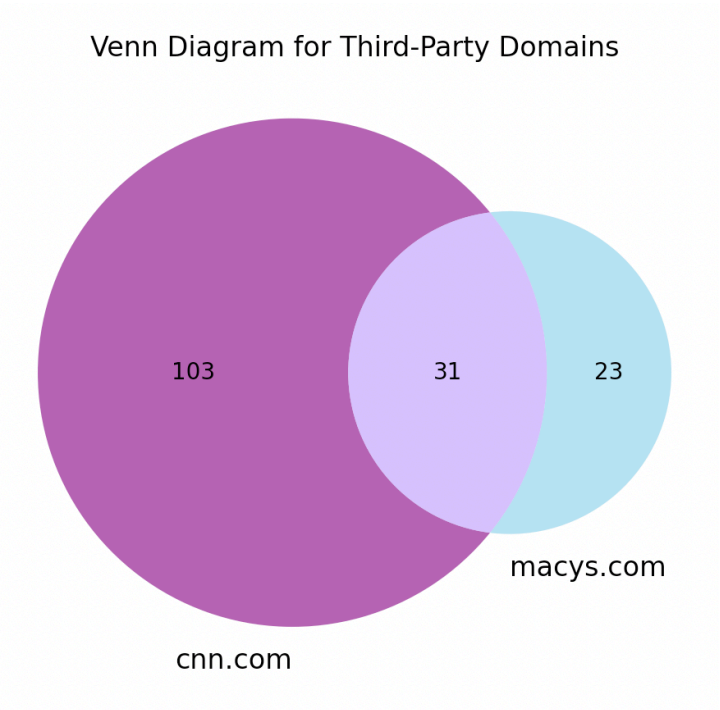
```
|          netmng.com          |
|          simpli.fi           |
|           bidr.io            |
|         adentifi.com         |
|       tribalfusion.com       |
|           w55c.net           |
|         owneriq.net          |
|       resetdigital.co        |
|         taboola.com          |
|        serverbid.com         |
|          adgrx.com           |
|         mxptint.net          |
|          dotomi.com          |
|       ipredictive.com        |
|          gumgum.com          |
|          adform.net          |
|          rundsp.com          |
|      acuityplatform.com      |
|         zorosrv.com          |
|     technoratimedia.com      |
|        eyereturn.com         |
|          appier.net          |
|        brealtime.com         |
|          adhigh.net          |
|        contextweb.com        |
|         gwallet.com          |
|        bidtheatre.com        |
+------------------------------+
```

The total number of third-party domains for file www.macys.com.har is 54. The domains that are third-party are shown below as a screenshot that is taken directly from the program that generated these results.

Domains
macysassets.com
tiqcdn.com
go-mpulse.net
demdex.net
omtrdc.net
everesttech.net
narrativ.com
doubleclick.net
owneriq.net
yimg.com
ads-twitter.com
hlserve.com
rlcdn.com
d1n00d49gkbray.cloudfront.net
pinimg.com
rmtag.com
agkn.com
facebook.com
medallia.com
facebook.net
bing.com
taboola.com
criteo.com
googletagmanager.com
akamaihd.net
mathtag.com
outbrain.com
bam-x.com
criteo.net
storetail.net
ibmcloud.com
yahoo.com
twitter.com
t.co
krxd.net

```
|          tapad.com          |
|          openx.net          |
|          adnxs.com          |
|        contextweb.com       |
|       casalemedia.com       |
|       linksynergy.com       |
|     google-analytics.com    |
|        pinterest.com        |
|         kampyle.com         |
|          akstat.io          |
|          google.com         |
|      rubiconproject.com     |
|         pubmatic.com        |
|          reson8.com         |
|         bluekai.com         |
|        bidswitch.net        |
|          digitru.st         |
|          adform.net         |
|         smarterhq.io        |
+-----------------------------+
```

b. The Venn diagram that represent the number of third-party domains in both the sites along with the common number of domains is shown below.



This data was obtained from the program output that was compiled for this question. As we can see, cnn.com has a total of 134 third party domains, and macys.com has a total of 54 third-party domains. The two sites share 31 common third-party domains between each other.

c. The third-party domains that are common across the two sites are displayed below. The output was obtained from the code.

Serial	Domain
1	criteo.net
2	outbrain.com
3	doubleclick.net
4	krxd.net
5	bing.com
6	ads-twitter.com
7	demdex.net
8	google.com
9	googletagmanager.com
10	google-analytics.com
11	rlcdn.com
12	rubiconproject.com
13	adnxs.com
14	t.co
15	everesttech.net
16	casalemedia.com
17	facebook.net
18	facebook.com
19	bluekai.com
20	agkn.com
21	bidswitch.net
22	criteo.com
23	mathtag.com
24	yahoo.com
25	tapad.com
26	pubmatic.com
27	twitter.com
28	owneriq.net
29	taboola.com
30	adform.net
31	contextweb.com

Problem 3:

The adblocker plus python package was downloaded and a program was written in order to determine what requests would be blocked based on all the requests that are made from the .har file.

The question discusses that the right options should be provided to the AdBlock checking functions to ensure that the checks against the rules are done properly. On investigating the different options in the requests on both the .har files, we can see that the results belong to the following categories:

```
options = ('image', 'xhr', 'document', 'fetch', 'font', 'script', 'styles
heet', 'other')
```

These options that are displayed above were extracted from the "_resourceType" attribute that is present for each of the requests. On comparison of these options with the AdBlock plus documentation at https://help.eyeo.com/en/adblockplus/how-to-write-filters#element-hiding, xhr was represented as xmlhttprequest, and fetch was removed since that is not a valid option.

In each request contained from the har file, the following checks take place:

1. Identify if the URL is first-party or third-party.
2. Identify if a valid option is present.
3. Ensure that if the option is xhr, the translation is made to represent xhr as xmlhttprequest.
4. Count the number of unblocked vs blocked requests.
5. Increase the counter of the total number of requests.

The parsed report as per the questions requirements is shown below. This output was obtained from the code.

Site	# of total HTTP Requests	# of HTTP Requests Blocked
www.cnn.com	747	388
www.macys.com	202	18

Note: The code with the Readme is attached to the submissions file.

Problem 4:

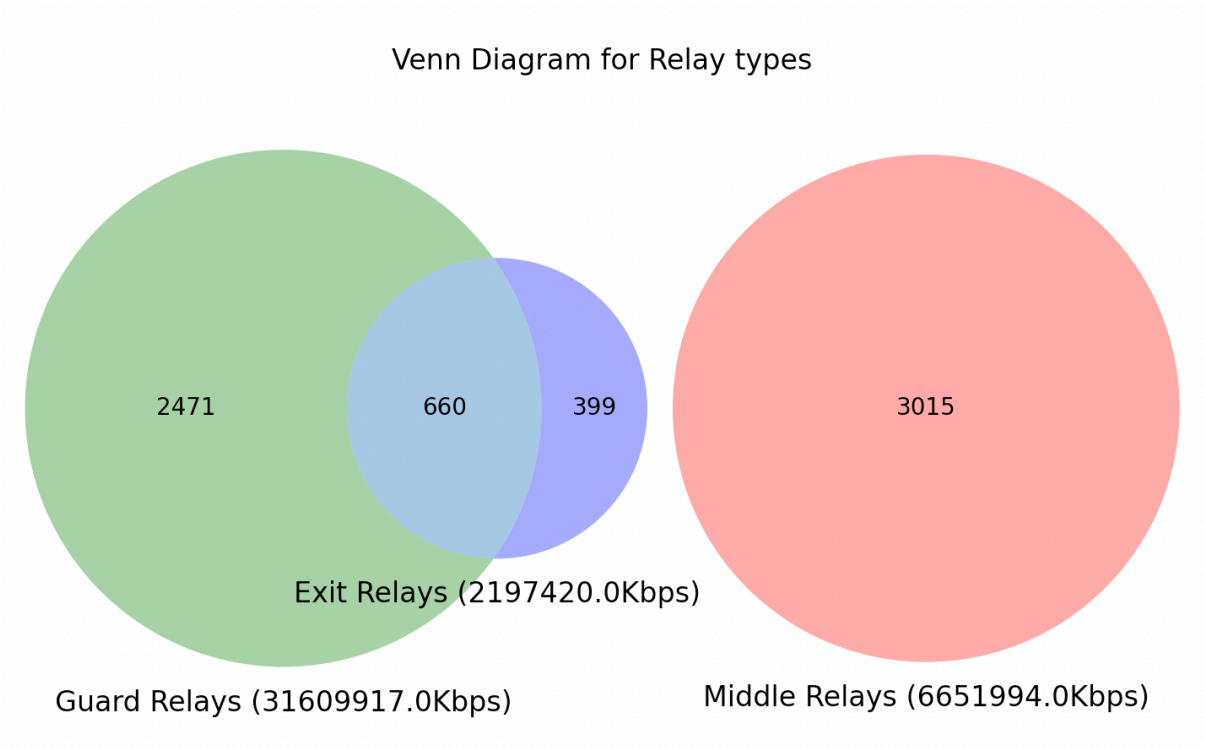
It was noticed that the excel sheet had entries were multiple tor relays either had the same IP address or the same hostname. In order to be able to uniquely identify each of these tor relays, an MD5 hash algorithm on the combination of router name, bandwidth, ip and the hostname was used to generate a unique identifier for each of the 6545 entries in the csv file that was provided for the assignment.

a. The top 5 countries that host the greatest number of Tor relays are:
['DE', 'US', 'FR', 'NL', 'CA']

b. The top 5 relays that contribute bandwidth are:

Router Name	Hostname	IP Address	Bandwidth
CalyxInstitute03	162.247.74.213	162.247.74.213	117100
PinkiePieParty	178-165-72-177-kh.maxnet.ua	178.165.72.177	100409
Unnamed	ns3082025.ip-145-239-66.eu	145.239.66.236	86850
Unnamed	ns340204.ip-5-39-69.eu	5.39.69.166	80597
Unnamed	ns3127631.ip-54-38-92.eu	54.38.92.43	78916

c. The Venn diagram for the distribution of the number relays that are acting as middle, exit and guard are shown below.



The distribution of the relays suggests that there is a total of 6545 relays which conforms to the total number of entries in the csv file (compared with the number to rows).
There is a total of 3015 middle relays, 399 relays that hold only the exit role, and 660 relays that hold the exit as well as the guard role. In addition, there are 2471 relays that act as pure guard nodes.
The table below was computed for the cumulative bandwidths of each of the above categories.

Cumulative Bandwidths Per group			
Guard Relay Only BW (kbps)	Middle Relay Only BW (kbps)	Exit Relay Only BW (kbps)	Exit + Guard Relay BW (kbps)
31609917.0	6651994.0	2197420.0	12270918.0

The data above suggests that there is 31.6 Gbps of bandwidth available in the guard relays only. Middle relays account for 6.65 Gbps of bandwidth, 12.27 Gbps of bandwidth for relays that are behaving as both exit and guard relays. However, there is only 2.197 Gbps of bandwidth available for exit only relays. This is a bottleneck in the performance of the Tor network.