Shi Su
shis

Problem #1 : Using Tor

1.1 Preliminaries

1. What is your public IP address?
   Public IP address: **128.237.200.1**
   Result of whois:

```
NetRange:       128.237.0.0 - 128.237.255.255
CIDR:           128.237.0.0/16
NetName:        CMU-NET-128-237
NetHandle:      NET-128-237-0-0-1
Parent:         NET128 (NET-128-0-0-0-0)
NetType:        Direct Assignment
OriginAS:       AS9
Organization:   Carnegie Mellon University (CARNEG-Z)
RegDate:        1987-05-06
Updated:        2012-04-02
Ref:            http://whois.arin.net/rest/net/NET-128-237-0-0-1
```

Complete output available in **shis_hw4/whois.txt**

2. Display all Tor circuits your machine currently uses, as specified by the list of three Tor relays.
   Script available in **shis_hw4/circuits.py**
   It assumes that **tor** is installed and already launched tor with **ControlPort: 9151**

```
Circuit 1 (GENERAL)
 |- Nickname: IP, country, bandwidth
 |---------- fingerprint
 |- DanaScully: 213.246.56.62, fr, 10500
 |---------- 5158B6A386EECCCA74F1AFA11993F0E70B329D66
 |- bmwanon3: 176.9.1.211, de, 17100
 |---------- D123C0F8F562804693C47D68C61786023B295E98
 |- PsychoOnion3: 89.187.142.208, cz, 31400
 +---------- 64186650FFE4469EBBE52B644AE543864D32F43C

Circuit 2 (GENERAL)
 |- Nickname: IP, country, bandwidth
 |---------- fingerprint
 |- DanaScully: 213.246.56.62, fr, 10500
 |---------- 5158B6A386EECCCA74F1AFA11993F0E70B329D66
 |- 0011: 89.163.128.11, de, 12700
 |---------- FE757A32AE8B7CEB0AC45392411299166586C8C6
 |- CalyxInstitute12: 162.247.72.199, us, 6030
 +---------- B34CC9056250847D1980F08285B01CF0B718C0B6

Circuit 3 (GENERAL)
 |- Nickname: IP, country, bandwidth
 |---------- fingerprint
 |- DanaScully: 213.246.56.62, fr, 10500
 |---------- 5158B6A386EECCCA74F1AFA11993F0E70B329D66
 |- sputnik: 188.40.128.246, de, 16800
 |---------- AD19490C7DBB26D3A68EFC824F67E69B0A96E601
 |- epow0tornode: 146.0.32.144, de, 110000
 +---------- 35E8B344F661F4F2E68B17648F35798B44672D7E

Circuit 4 (GENERAL)
 |- Nickname: IP, country, bandwidth
 |---------- fingerprint
 |- DanaScully: 213.246.56.62, fr, 10500
 |---------- 5158B6A386EECCCA74F1AFA11993F0E70B329D66
 |- torfr01: 195.154.128.151, fr, 4680
 |---------- 4D5FAA22C650CE72CA43A078C40A91ED61461F36
 |- Unnamed: 192.42.116.161, nl, 63300
 +---------- C804BE8FB1C7C42D43C4A5E2039E77AA0FF3A8B4
```

Figure 1: Current Circuits

1.2 Measuring latencies

1. Measures the difference between response times with or without going over Tor.

Script available in **shis_hw4/measure_latency.py**
This script is used in 1.2.2 and 1.2.3
It assumes that **tor** is installed and already launched with **SocksPort: 9150, ControlPort: 9151**
It assumes that **PycURL** module is installed.

Below is the list of supported argument, when running without arguments, the script call url provided in the assignment, and retry 10 times.

Use "-u" to set the url to connect
Use "-r" to set number of retries

2. Give an estimate figure of how much overhead in latency Tor generates on a given connection. Run the above experiment 10 times, changing Tor identities (and thus, circuits) but connecting to the same website for each of the 10 runs, and provide a table summarizing the results.

Test URL: https://stem.torproject.org/tutorials/to_russia_with_love.html

n	direct	tor	difference	abs diff	rank
1	1.03	0.28	0.75	0.75	6
2	0.92	0.18	0.75	0.75	6
3	0.94	0.26	0.69	0.69	2.5
4	0.95	0.26	0.69	0.69	2.5
5	1.01	0.26	0.74	0.74	4
6	0.94	0.18	0.75	0.75	6
7	1.06	0.19	0.87	0.87	8
8	0.90	0.27	0.63	0.63	1
9	1.47	0.26	1.21	1.21	9
10	1.46	0.18	1.28	1.28	10

Using **Wilcoxon Signed-Rank Test** with the data above, we got test statistic value:
$T_- = 0 \quad T_+ = |1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9| = 55$

And we're trying to prove hypothesis H_1 over H_0:
H_0: Distribution of response time of Tor(D_T) and direct connection(D_D) are identical
H_1, D_T is the right shifted of the D_D, i.e. latency of Tor is higner than direct connection
H_1, D_T is the left shifted of the D_D, i.e. latency of Tor is lower than direct connection

Because we're testing D_T is the right shifted of the D_D, so we will use test statistic $T_- = 0$
Critical Values $T_0 = 5$ in Wilcoxon Signed Rank one-tailed test when **n=10,** α **= 0.01** [1]
$T_- <= T_0$, T_- falls in the rejection region of T_0
So the data provide sufficient evidence to conclude that D_T is the right shifted of D_D, at α = 0.01. i.e. The latency for Tor connection is higher than direct connection.

3. Measure the latency of connecting to Tor hidden service. Compare it to the latency to its "clearnet" version using the Tor network, and bypassing the Tor network. Explain why the hidden service address does not use https

Compare between hidden service and clearnet version over Tor:

n	hidden	clear w/ Tor	difference	abs diff	rank
1	5.59	11.86	-6.17	6.17	-10
2	1.68	1.43	0.25	0.25	4
3	1.61	1.52	0.09	0.09	1
4	1.81	1.29	0.52	0.52	6
5	3.63	1.26	2.37	2.37	9
6	0.74	1.27	0.75	0.75	8
7	0.95	1.43	-0.53	0.53	-7
8	0.76	1.27	-0.51	0.51	-5
9	0.78	1.00	-0.22	0.22	-3
10	0.75	0.87	-0.12	0.12	-2

Using Wilcoxon Signed-Rank Test with the data above, we got test statistic value:
$T_- = |-10 + -7 + -5 + -3 + -2| = 27, \quad T_+ = |4 + 1 + 6 + 9 + 8| = 28, \quad T_0 = 5$
Both T_+ and T_- is larger than critical value, thus fall out of H_0's rejection region.

So there's no sufficient evidence to conclude that the distribution of latency of hidden service is different from the latency of the clearnet version over Tor.

Compare between hidden service and clearnet version **NOT** going through Tor:

n	hidden	clear w/o Tor	difference	abs diff	rank
1	5.59	0.15	5.44	5.44	10
2	1.68	0.11	1.57	1.57	7
3	1.61	0.11	1.50	1.50	6
4	1.81	0.11	1.70	1.70	8
5	3.63	0.10	3.53	3.53	9
6	0.74	0.11	0.63	0.63	2
7	0.95	0.10	0.85	0.85	5
8	0.76	0.14	0.62	0.62	1
9	0.78	0.11	0.67	0.67	4
10	0.75	0.10	0.65	0.65	3

$T_- = 0 \quad T_+ = |1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9| = 55$
Critical Values $T_0 = 5$ in Wilcoxon Signed Rank one-tailed test when n=10, α = 0.01 [1]
$T_- <= T_0$, T_- falls in the rejection region of T_0
So the data provide sufficient evidence to conclude that D_H is the right shifted of D_C, at α = 0.01. i.e. The latency for hidden service is higher than the clearnet version not going over Tor.

Connections with the Tor hidden services are going through Tor circuits, which are already all encrypted, so there's no need for https to reenforce confidentiality. Also the address of a hidden service came from the hash of the hidden service's public key, thus Tor client can verify

the address using the public key got from the directory servers, so the CA to sign certifications in https model is not required.

1.3 Using exits to circumvent censorship

1. Try to connect to http://dogo.ece.cmu.edu/tor-homework/public/ and http://dogo.ece.cmu.edu/tor-homework/secret/ using a regular browser. How different are the results?

When opening secret url with regular browser, it returns **403 Forbidden**.
When opening public url with regular browser, it shows "**the public part of the server is working**"

2. Write a Stem script to constrain Tor to specific exits in different countries.

Script available in **shis_hw4/exit_node.py**
It assumes that **tor** is installed and added to PATH, and the **PycURL** module is installed.
It assumes port **9150, 9151** are not occupied, and will launch with **SocksPort: 9150, ControlPort: 9151**

Below is the list of supported argument, when running without arguments, the script call url provided in the assignment, and try to switch circuits 5 times in a country.

Use "-u" to set the url to connect
Use "-r" to set number of s to switch circuit for exist nodes in a country

3. Use your script to find as many countries as possible in which the website http://dogo.ece.cmu.edu/tor-homework/secret/ is not blocked. Provide a list of these countries, as well as the exact date/time at which you made each request verifying that each of these countries was not blocked.

Using the script in 1.3.2, I was able to connect to the secret website from the ExitNodes in following countries.

Find GB returns 200, 2015-12-02 03:16:32
Find RU returns 200, 2015-12-02 03:16:35
Find FR returns 200, 2015-12-02 03:17:19
Find AU returns 200, 2015-12-02 03:17:24
Find PL returns 200, 2015-12-02 03:17:28
Find CH returns 200, 2015-12-02 03:19:51
Find HU returns 200, 2015-12-02 03:34:21
Find GR returns 200, 2015-12-02 03:36:40
Find LU returns 200, 2015-12-02 03:47:58
Find RO returns 200, 2015-12-02 11:04:33
Find SE returns 200, 2015-12-02 11:14:24

Problem #2 : WikiLeaks and anonymity

1. Decide whether you support banning the design and use of anonymous networks or not, and prepare four arguments to support your position, and one mitigating factor

I oppose banning anonymous networks.

Anonymous network provides a cheaply accessible channel for people to communicate without the fear for the hurts or embarrassments that are likely to occur if their identities are reviewed. It protects people in some special positions like journalists and dissidents as well as common people who values their privacies.

As mentioned above, a lot of users with legitimate purposes benefits from anonymous networks, and they generally don?t attract public attention, while in the contrast, those bad incidents related to Tor, anonymous network or any other unfamiliar techniques will make an interesting news. Like the maintainer of Tor said "Often people don't have a good measure of how many polite Tor users are connecting to their service, you never notice them until there's an impolite one."[2] It?s far from advisable to ban a wildly desired channel simply for those over magnified issues.

Also it?s difficult to carry out and enforce the bill to ban the design and use of anonymous network, for even a complicated anonymous network like Tor are consist of basic cryptography primitives and network operations. It's impractical to ban those building blocks and hard to restrain their combination to limit the outcome. It will either leave ambiguity in the bill/law or make the bill/law easy to bypass by simple variations.

And back to the original issue about WikiLeaks, the information is submitted to WikiLeaks voluntarily means the root cause of sensitive information leaking probably lies in the poor access control and lack of audit trail. If the source is not controlled, banning one channel only makes information flow to other new and sneakier ones.

But still we cannot deny that anonymous networks have been used for illegal activities, and made them harder to trace. Which to certain degree, lowers the cost of some illegal action like copyright violation while making the cost to address those issue higher.

[1] Stat.ufl.edu,. (2015). Retrieved 4 Dec. 2015, from
http://www.stat.ufl.edu/ winner/tables/wilcox_signrank.pdf
[2] The Tor Project, I. (2015). Tor Project: Abuse FAQ. Torproject.org. Retrieved 4 December 2015, from https://www.torproject.org/docs/faq-abuse.html.en#Bans