**Final Assignment**

**Part 1**

**1. Identify and explain at least three security and privacy issues with metaverse technology.**

1. **Provide the detailed attacker model for each problem.**

* The metaverse is very closely related to the real world. Therefore, if the metaverse platform is hacked, real money can be stolen. Most of real metaverse platform uses virtual cryptocurrency, if attacker get user’s privacy, they can send their money to others.
* For using metaverse platform, many users use their own VR, AR device or IoT sensors to collect their privacy to service. So, if attacker success to attack, they can know about much more sensitive personal information of users.
* If the actual space is implemented with a digital-twin technology, as in the video, others can know about the structure of private places.

1. **Provide the possible defense/protection approach for each problem**

* Use NFT for protect digital assets. It can ensure their ownership.
* Encryption of various information like biometric information. Also, we should make new strict Metaverse Privacy Policy.
* As far as possible, we avoid building a virtual space that is exactly like the real world.

**2. Read the following article about the virtual Youtuber Rui:**

1. **Can you identify the area/features to proof that she is not a real human?**

If I hadn’t known that she is not a real human beforehand, I certainly wouldn't have noticed this. It is true that it is difficult to distinguish properly with the human eye.

1. **Explain the detailed approach to identify she is not a real human from the given video.**

The way to identify that she is not a real human is really similar with how to identify a fake in a deepfake technologies.

* Synchronization of voice and video(too difficult to know)
* Technical analyzing(using AI)

1. **What are three pros with the above technologies used for Rui? (provide the sufficient details)**

* It is easy to make personalized advertising. If a human being is customized according to the target of the advertisement, a very large advertisement effect can be seen.
* In virtual world, virtual human has same effect with real human. But they don’t eat, don't age, don’t sleep. So, it is easy to manage that human for economic activities.
* These technologies can remove discrimination in appearance. There are many people who get amazing talent to something, but someone can’t express their ability because of low self-confidence in their appearance.

1. **What are three cons with the above technologies used for Rui? (provide the sufficient details)**

* lookism may intensify. If everyone can make avatar that is virtual human with perfect appearance, People’s thought about appearance can be changed.
* Because they are not real human, sexual harassment against them can be exacerbated.(there are not legal protection for them)
* Virtual humans could replace real human jobs.

**Part 2**

**Flow.gz**

텍스트이(가) 표시된 사진

자동 생성된 설명

First, I make essential information lists by network traffic.

**1. Calculate the total number of TCP/80 connections.**

텍스트이(가) 표시된 사진

자동 생성된 설명

By using dport and prtc list, I can find the total number of TCP/80 connections.

Total count : 559998

**2. Calculate the average of the total number of TCP/80 connections per second.**

The start time is 2011-03-07 09:03:24 and the end time is 2011-03-09 02:14:27.

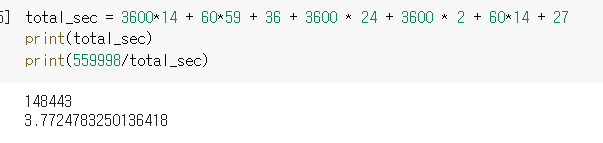
So the total seconds is as follows.

14h 59m 36s 3600\*14 + 60\*59 + 36

24h 3600 \* 24

2h 14m 27s 3600 \* 2 + 60\*14 + 27

* 3600\*14 + 60\*59 + 36 + 3600 \* 24 + 3600 \* 2 + 60\*14 + 27 = 148443



The average of the total number of TCP/80 per sec : 3.7724783250136418

**3. How many different destination IPs (dip) are there?**

텍스트이(가) 표시된 사진

자동 생성된 설명

The number of different destination : 140

**4. How many different dbyte are there?**

텍스트이(가) 표시된 사진

자동 생성된 설명

The number of different dbyte : 244

**5. Calculate the average, median, min, max, and std of packet sizes (dbyte) used in this network traffic.**

텍스트이(가) 표시된 사진

자동 생성된 설명

**6. Calculate the average, median, min, max, and std of packet sizes (sbyte) used in this network traffic.**

텍스트이(가) 표시된 사진

자동 생성된 설명

**http.gz**

텍스트이(가) 표시된 사진

자동 생성된 설명

First, I make essential information lists by network traffic.

**7. How many destination IP addresses are there?**

텍스트이(가) 표시된 사진

자동 생성된 설명

The number of different dips : 73

**8. How many different domains are there ?**

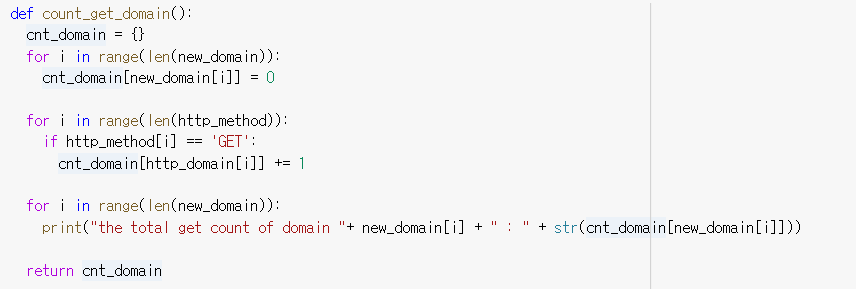
텍스트이(가) 표시된 사진

자동 생성된 설명

But there is ‘-‘ domain so if I except this, the total number is 31.

The number of different domains : 31(if except ‘-‘ domain)

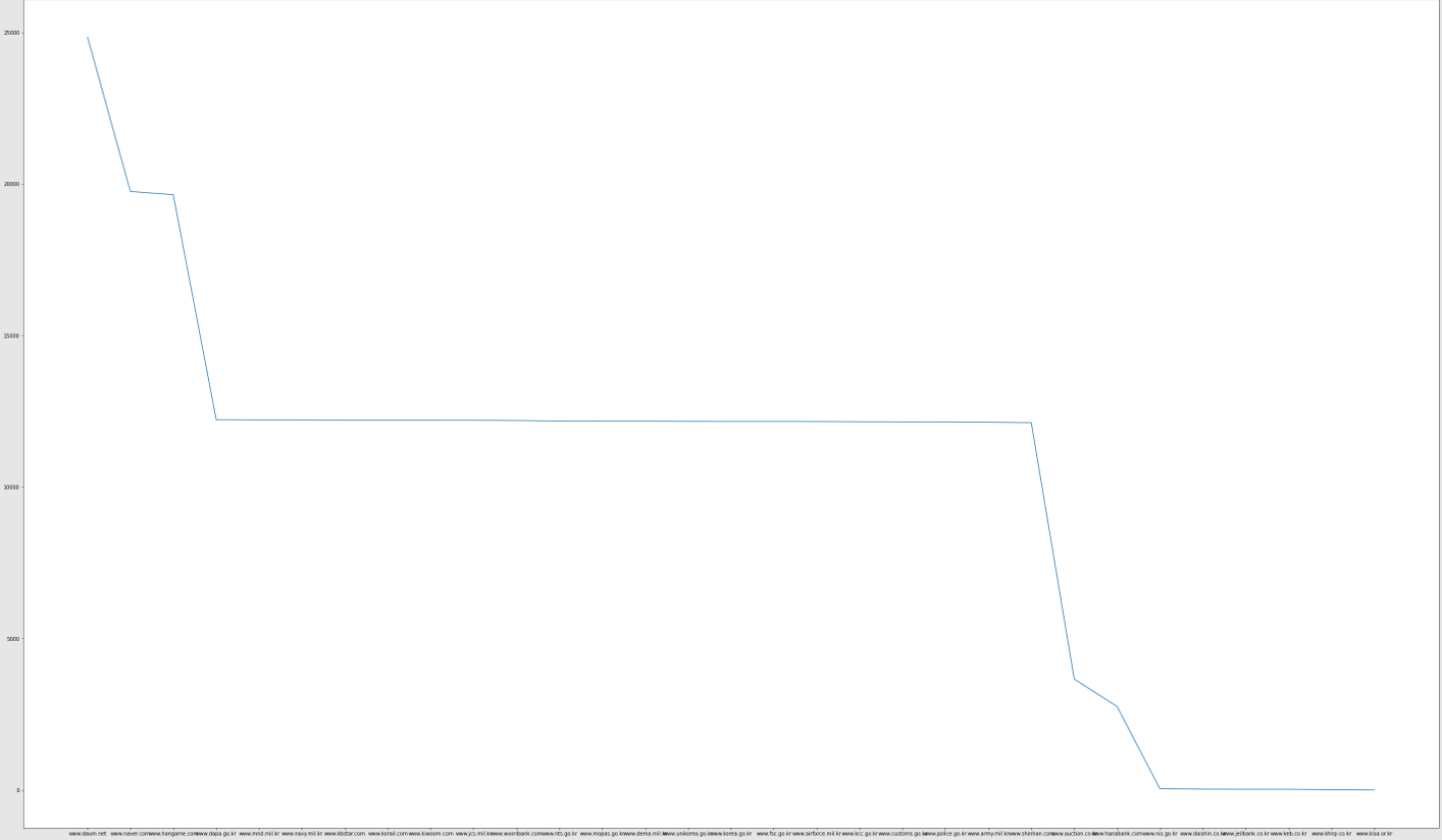
**9. Count the number of “GET” connections made to each domain.**



테이블이(가) 표시된 사진

자동 생성된 설명

**10. Plot the histogram of Q9 (above question) in a descending order, where the X-axis is the domain name, and the Y-axis is the count.**



**11. Calculate the average, median, min, max, variance, and standard deviation for the**

**number of “GET” connections made to each domain (Q9)?**

텍스트이(가) 표시된 사진

자동 생성된 설명

I except ‘-‘ domain in this code.

**Machine Learning**

**12. First, read the train.csv and test.csv, and provide the dimension of each file.**

텍스트이(가) 표시된 사진

자동 생성된 설명

I used dictionary to store the data.

**13. If any value is ‘–‘, ‘#close”, null, or empty, then assign it to 0 in train.csv and test.csv.**

테이블이(가) 표시된 사진

자동 생성된 설명

This is same part with above lists(Q12)

**14. Write a python code to drop two columns: [' id.orig\_h ',' id.resp\_h '], which we will not consider as features for classification in train.csv.**

텍스트이(가) 표시된 사진

자동 생성된 설명

**15. Make each column in train.csv to be the following type:**

테이블이(가) 표시된 사진

자동 생성된 설명

In question 13, when the column “ts” is “#close”, there are empty value in original file so the example has 19733 Non-Null count. But in here I changed all empty value to 0, so all Non-Null count is same, 19735.

**16. Make each column in test.csv to be the following type:**

테이블이(가) 표시된 사진

자동 생성된 설명

I used same Dtype with above table. ( a little different with example)

**17. Apply “LabelEncoder()” when needed.**

I used labelEncoder after Q18~Q20

테이블이(가) 표시된 사진

자동 생성된 설명

**18. Write a python code to drop three columns:**

테이블이(가) 표시된 사진

자동 생성된 설명

**19. Store your final output to: train\_yourname.csv and test\_yourname.csv (save as csv file)**



텍스트이(가) 표시된 사진

자동 생성된 설명

**20. Execute the following code**

텍스트이(가) 표시된 사진

자동 생성된 설명

**21. Execute the remining code and report the final precision, recall, and F1.**텍스트이(가) 표시된 사진

자동 생성된 설명

**Part 3**

**1. Please describe gTLDs in your own words. How are they different from the top-level domain?**

TLD is a top level domain and it is every domain after final dot. TLD consists of a general top-level domain and country code top-level domain. In other words, gTLD(general top-level domain) is included in TLD. gTLD is top level domain that indicates specific purpose of usage.(companies, individuals ..)

**2. How many gTLDs are there as of 2022? List all of approved gTLDs.**

: There are 1248 gTLDs (iana.org)

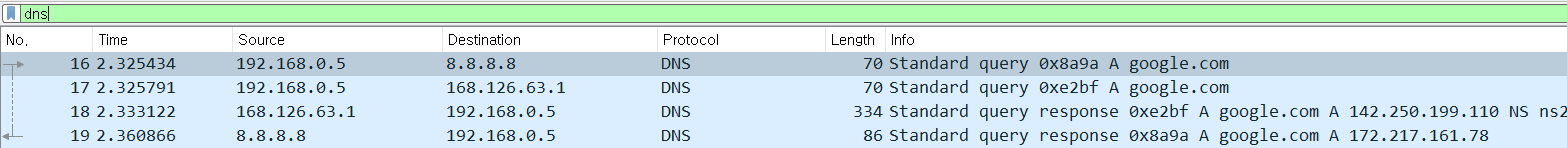
I’ll attach lists to separate file

**3. Use Chrome browser and change the default search engine to below (remove all other search engines):**

텍스트이(가) 표시된 사진

자동 생성된 설명

**4. Turn on the Wireshark and type “simon” in the browser, and capture the network traffic.**



Destination IP :

8.8.8.8

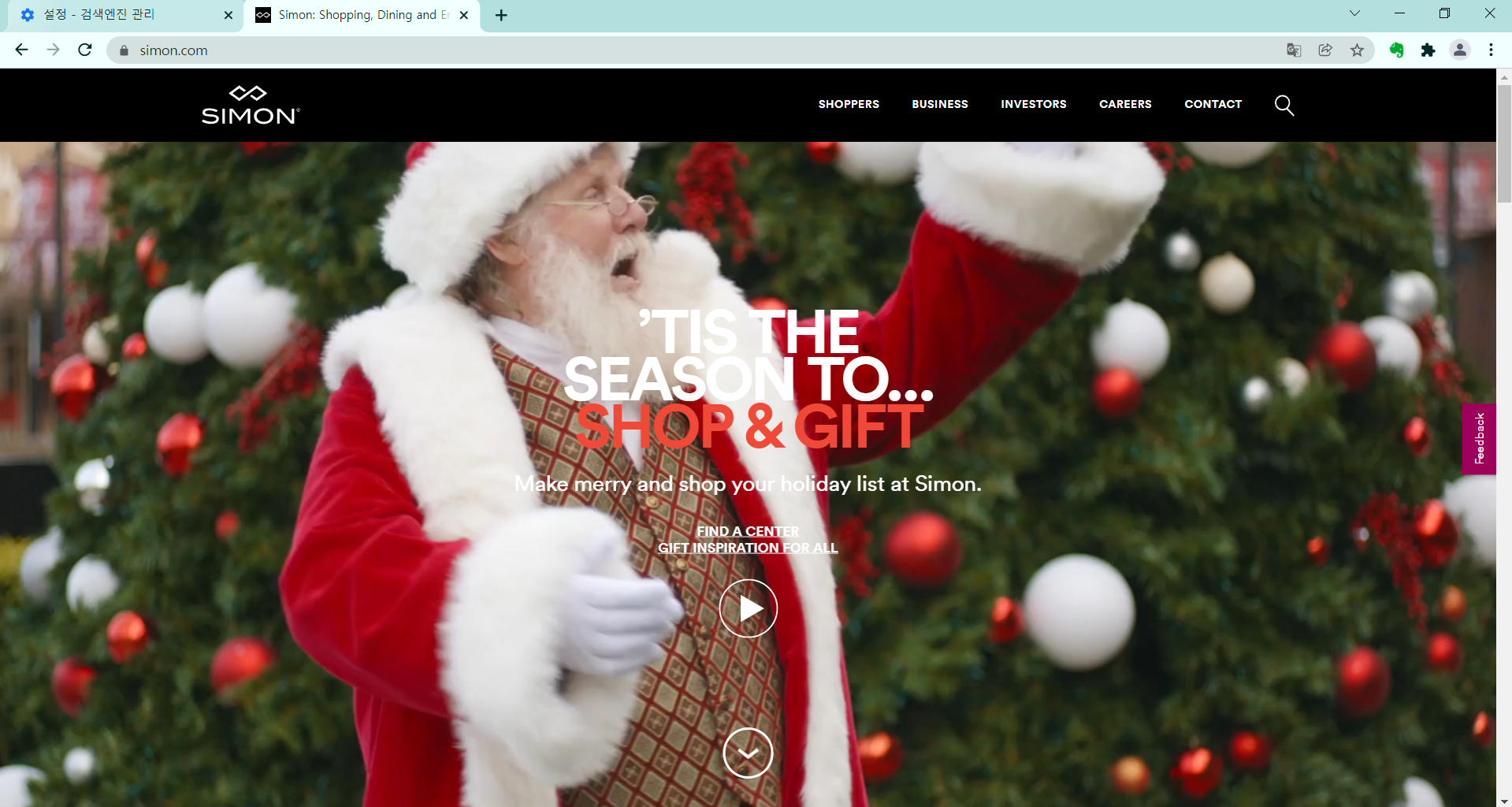
168.126.63.1

When I search word “simon” in that browser, it has to find IP address from domain. So that browser query to KT DNS server because it is set as default DNS server in my network. But KT DNS server can’t find IP address of that domain, so access to google DNS server to find.(8.8.8.8 is google’s DNS server ip address)

**5. Turn on the Wireshark and type “simon.com” in the browser, and capture the network traffic.**

테이블이(가) 표시된 사진

자동 생성된 설명

I

Destination IP :

168.126.63.1

When the browser query to KT DNS, it can find IP address of the domain “simon.com”. but for getting IP address of total domain, KT DNS server has to query recursively to other DNS server. So there are many connection to KT DNS server.

**6. Describe the network traffic difference between Q4 and Q5.**

In Q5, There is gTLD “.com” in full domain name. but Q4 is not. So in Q4, Local DNS server can’t access to com DNS server by using tld.

**7. If gTLD is “simon”, how the browser should resolve gTLD = “simon” compared to “simon” as a search query? For example, following is the search query or gTLD?**

It doesn’t have “.” Character to identify tld, so it is just the search query.

**8. What are the 3 pros and 3 cons of having above gTLD?**

Pros:

* It is easy to know what the brand of that domain is.
* Can make domain more shortly.(if there are just few TLDs, domain names become long.
* easy to remember the domain name for consumers

Cons:

* There are danger of cybersquatting.
* Famous gTLD competition intensifies to be exposed in search engines.
* misleading sites are likely to increase.

**9. Describe the potential security issue and vulnerability of using gTLD.**

By Attackers, it is important that they buy many domains before related organizations or individuals do for phishing or Cybersquatting. Today, there are too many gTLD than before, so it is easy to make similar domain name with attack target. It means that organizations and companies should pay for buying domain that match their names with different gTLDs. And It is easy for users to accidentally enter a phishing site.