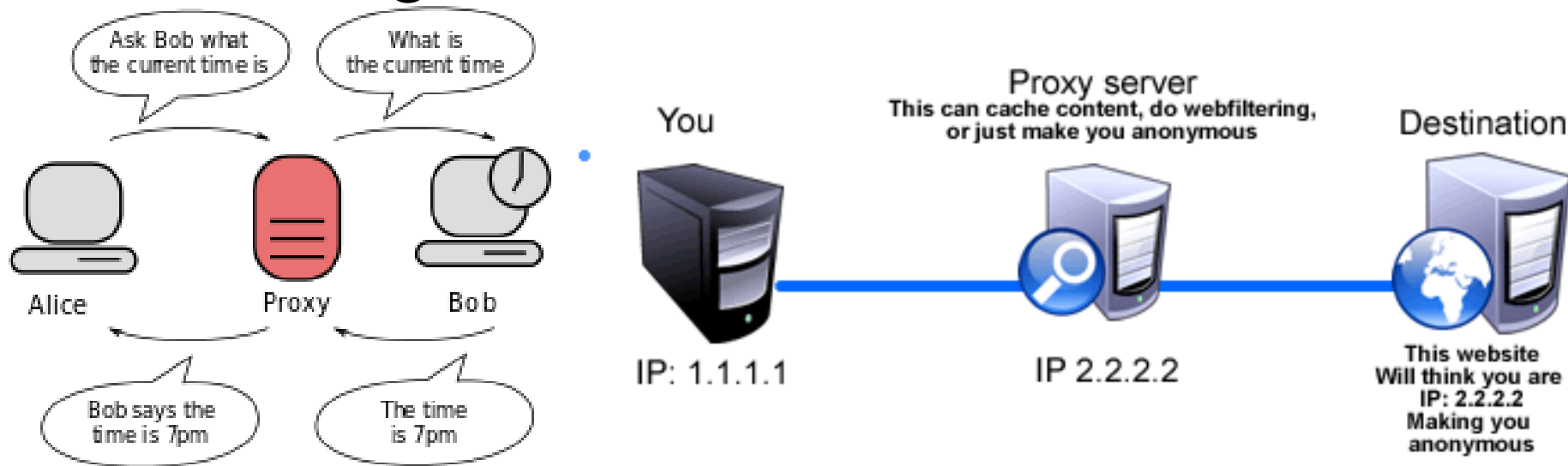# Computer Network Project 3
Simple Proxy Server
(Difficulty ★★★☆☆)

CSI4106-01

Fall, 2023

What is a proxy server?

- An intermediary for requests from clients seeking resources from other servers.
- Resides anywhere, between YOU and destination.
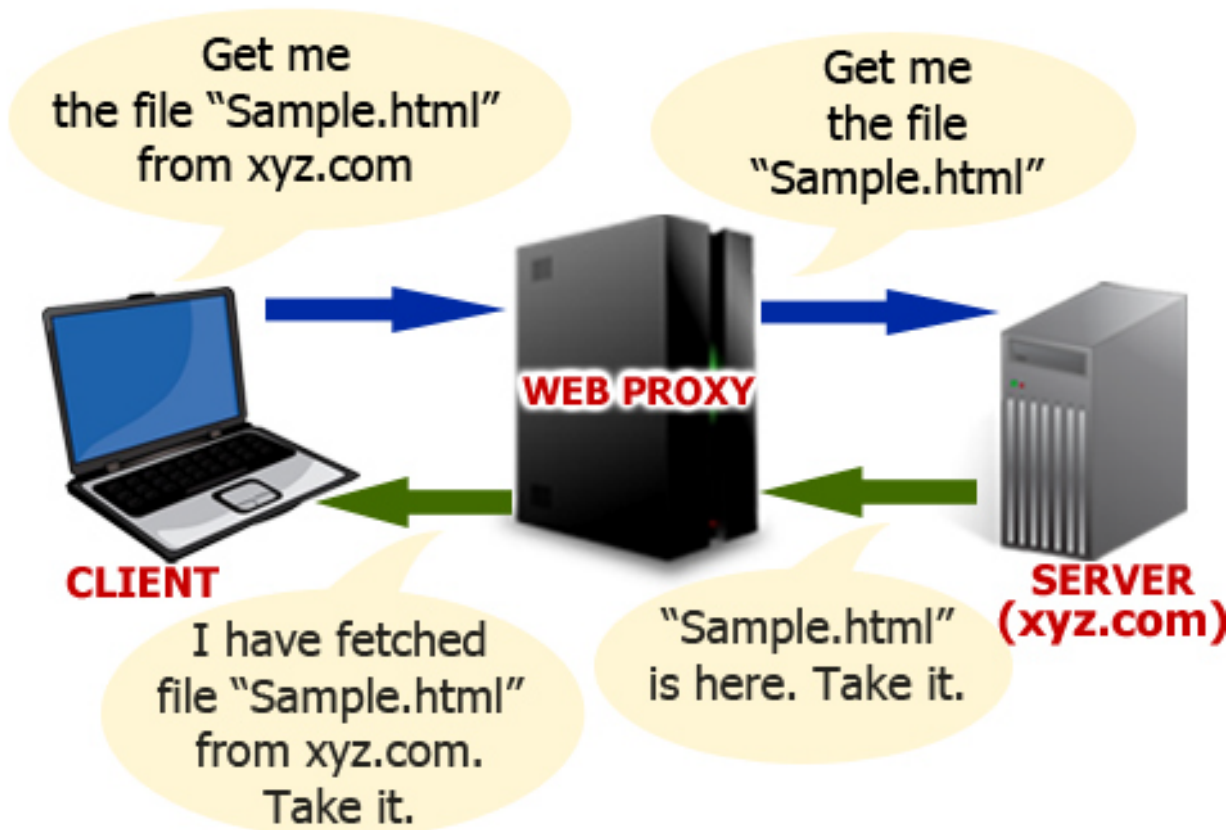- The fundamental function: intercepting and forwarding traffic.

The uses of proxy servers

- Filtering specific data (to block or redirect them)
- Logging or eavesdropping for monitoring traffic
- Caching data
- Etc.

- *(e.g.) SSL/DNS/CGI/HTTP proxy*
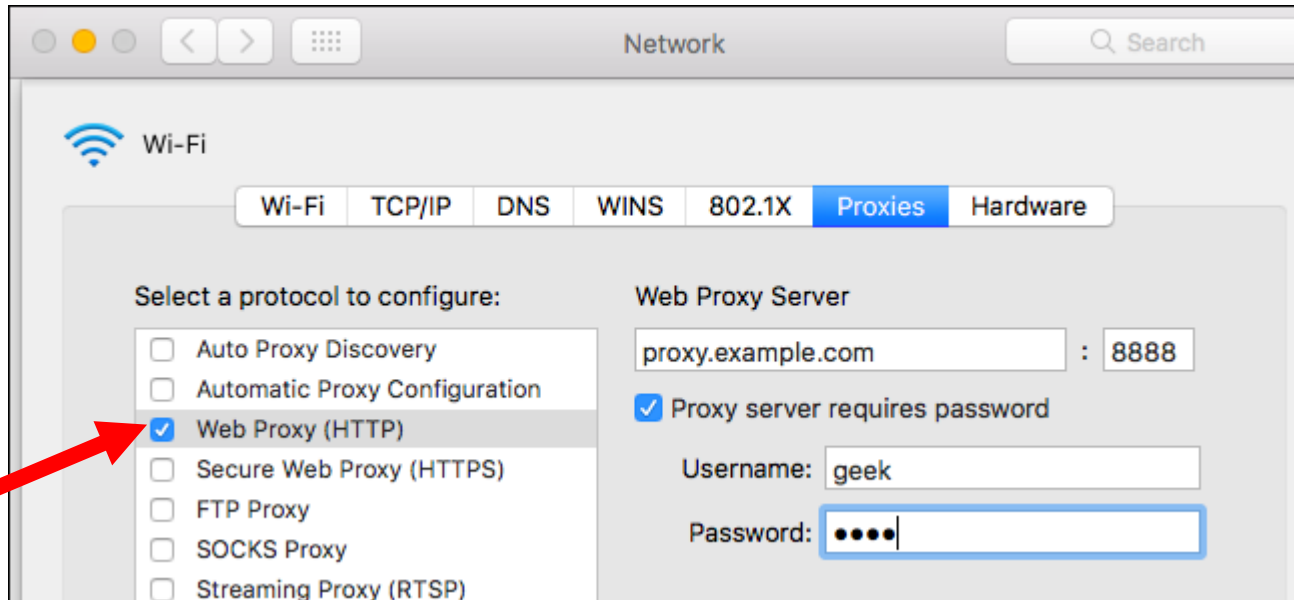- *(further…) Reverse, Anonymous, Tunneling…*

How a http proxy server works?

- **Transparent Proxy on HTTP**

How to configure a HTTP proxy?

- Goggle "how to set up proxy server on windows / mac / linux"

Assignment

- Write a code of http <u>transparent</u> proxy server (50pt)
  - HTTP 1.0/1.1 only (Do not consider HTTPS)
- Implement key features (30pt)
  - URL Filtering
  - Image Filtering
- Write a report (20pt)

Hint (1) Managing Two Sockets!

- One socket for receiving Client Request!
- Another socket for fetching Response of the client request
- You need to forward Requests/Responses.
- Hint: **Where should requests/responses go?**

Hint (2) Persistent Connection

- You do not have to implement persistent connection, but your proxy server may not work properly if you access a web server supporting the persistent connection.

- You have to <u>modify</u> HTTP header (Keep-Alive or Connection) so that your proxy server <u>does not</u> communicate with the client or server using the persistent connection.

- Hint: **Disable persistent connection**

Hint (3) Headers

- In addition to Keep-Alive header, some headers should not be retransmitted by the proxy server.
- On the other hand, some headers should be relayed by the proxy server without modification.
- Refer to https://developer.mozilla.org/en-US/docs/Web/HTTP/Headers
- Hint: **Remove or modify Hop-by-hop headers**

Key Feature (1)

- **Goal: Make a domain-level redirection**
  - Redirect any url containing "korea" to http://mnet.yonsei.ac.kr/
- **(Warning) This is not HTTP/1.1 302 Moved**
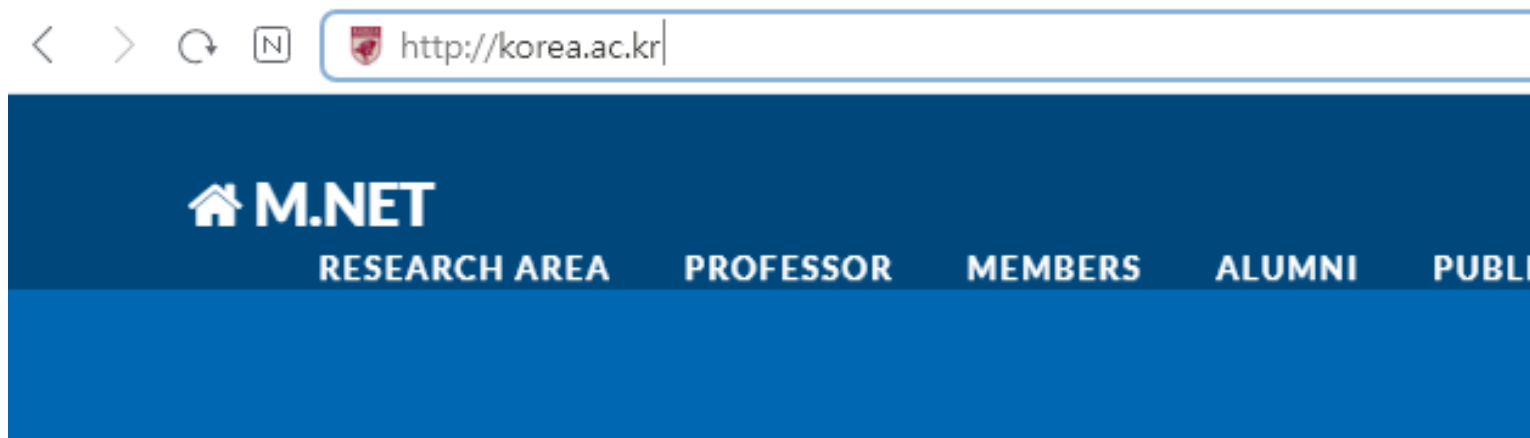
Key Feature (2)

- http://[URL]?**image_off**
  - From the request with this query string, all requests for <u>image</u> will be dropped. (use a flag)
- To stop it: http://[URL]?**image_on**

Sample Web Page

A random photo.

Frank da Cruz
Sat Jan 17 12:07:32 2004
Updated 26 May 2019 for HTML5.

CONTENTS

1. Creating a Web Page
2. HTML Syntax
3. Special Characters
4. Converting Plain Text to HTML
5. Effects
6. Lists
7. Links
8. Tables
9. Viewing your Web page
10. Installing your Web Page on the Internet
11. Where to go from here

http://[URL]

Sample Web Page

A random photo.

Frank da Cruz
Sat Jan 17 12:07:32 2004
Updated 26 May 2019 for HTML5.

CONTENTS

1. Creating a Web Page
2. HTML Syntax
3. Special Characters
4. Converting Plain Text to HTML
5. Effects
6. Lists
7. Links
8. Tables
9. Viewing your Web page
10. Installing your Web Page on the Internet
11. Where to go from here

You can create a Web page on your desktop computer but nobody can see it but you. If your want other people to be able to see your Web pages, you need an account a computer that has a Web server. Nowadays most people have their own computers on their desks, but normally they don't have Web servers and anyway you don't want the whole world coming into your desktop computer to see your web page because (a) it's not designed for that, and (b) who

http://[URL]?image_off

Requirements

- HTTP 1.0/1.1 only (not HTTPS)
- Do not care performance or race condition issues
  - Don't have to use mutex or lock
- Run with custom port
  - python ./prx.py 9001
  - ./prx 9001 (in case of C/C++)
- Terminated by CTRL + C
  - Make sure the sockets are closed.

Requirements (cont'd)

- OS: Ubuntu 18.04 or later
- Language: Python3 or C/C++ (gcc)
  - Provide 'build.sh' file if you use C/C++
  - Do not forget library options (e.g., -lpthread)
- You should <u>not</u> use 3rd party libraries
  - Use standard python/C/C++ libraries
  - But you can use GNU C library

Requirements (cont'd)

- When a request from a client is arrived, your proxy server should print out the precise logs as follows:

```
Starting proxy server on port 9001
----------------------------------------------
1 [O] Redirected [X] Image filter
[CLI connected to 127.0.0.1:4994]
[CLI ==> PRX --- SRV]
  > GET korea.ac.kr/
  > Mozilla/5.0 (Windows NT 6.1; WOW64)
[SRV connected to mnet.yonsei.ac.kr:80]
[CLI --- PRX ==> SRV]
  > GET mnet.yonsei.ac.kr
  > Mozilla/5.0 (Windows NT 6.1; WOW64)
[CLI --- PRX <== SRV]
  > 200 OK
  > text/html; charset=utf-8 4238bytes
[CLI <== PRX --- SRV]
  > 200 OK
  > text/html; charset=utf-8 4238bytes
[CLI disconnected]
[SRV disconnected]
----------------------------------------------
2 [X] Redirected [X] Image filtered …
```

```
Starting proxy server on port PORT_NUMBER
----------------------------------------------
NO. [?] URL redirection [?] Image filter
// ? is O or X (mode indicators)
[CLI connected to CLI_IP:CLI_PORT]
[CLI ==> PRX --- SRV]
  > Request_1st_Line_from_Client
  > User-Agent_1st_Line_from_Client
[SRV connected to DST_DOMAIN:DST_PORT]
[CLI --- PRX ==> SRV]
  > Request_1st_Line_to_Server
  > User-Agent_1st_Line_to_Server
[CLI --- PRX <== SRV]
  > Response_Status_Code_from_Server
  > Response_MIME_Type Response_Sizebytes
[CLI <== PRX --- SRV]
  > Response_Status_Code_to_Client
  > MIME_Type_to_Client Size_to_Clientbytes
[CLI disconnected]
[SRV disconnected]
```

Requirements (cont'd)

- Image filter example

```
Starting proxy server on port 9001
-----------------------------------------------
1 [X] Redirected [O] Image filter
[CLI connected to 127.0.0.1:4994]
[CLI ==> PRX --- SRV]
  > GET www.any.url?image_off
  > Mozilla/5.0 (Windows NT 6.1; WOW64)
[SRV connected to www.any.url:80]
[CLI --- PRX ==> SRV]
  > GET www.any.url
  > Mozilla/5.0 (Windows NT 6.1; WOW64)
[CLI --- PRX <== SRV]
  > 200 OK
  > text/html; charset=utf-8 4238bytes
[CLI <== PRX --- SRV]
  > 200 OK
  > text/html; charset=utf-8 4238bytes
[CLI disconnected]
[SRV disconnected]
-----------------------------------------------
2 [X] Redirected [O] Image filter …
```

```
-----------------------------------------------
2 [X] Redirected [O] Image filter
[CLI connected to 127.0.0.1:4994]
[CLI ==> PRX --- SRV]
  > GET www.any.url/someimage.png
  > Mozilla/5.0 (Windows NT 6.1; WOW64)
[SRV connected to www.any.url:80]
[CLI --- PRX ==> SRV]
  > GET www.any.url/someimage.png
  > Mozilla/5.0 (Windows NT 6.1; WOW64)
[CLI --- PRX <== SRV]
  > 200 OK
  > image/png 5969 bytes
[CLI <== PRX --- SRV]
  > 404 Not Found
[CLI disconnected]
[SRV disconnected]
-----------------------------------------------
3 [X] Redirected [O] Image filter …
…
7 [X] Redirected [X] Image filter
[CLI connected to 127.0.0.1:4994]
[CLI ==> PRX --- SRV]
  > GET www.any.url?image_on
```

Your report must include

- **Introduction/Reference** (2pts)
  - Software environment, programming language you used, version, reference and so on.
- **Flow chart or Diagram** (8pts)
  - Must show the logic of your program
  - Focus on describing how your client and server work

Your report must include (cont'd)

- **Logical explanations block by block in detail.** (10pts)
  - It is different with brief comments in your source code!!!
  - In your report, write what the blocks do and why you implemented those functions.

Background

- TCP (socket programming)
- HTTP Packet (request and response)
- Basic idea of HTML tag with <style>
- https://en.wikipedia.org/wiki/Proxy_server
- HTTP Persistent Connection
  - https://www.safaribooksonline.com/library/view/http-the-definitive/1565925092/ch04s05.html

HTTP site sample

- As http is insecure, most of the websites redirect to https if you use http.
- You might have trouble finding the right websites to test.
- We provide URLs for testing
  - http://mnet.yonsei.ac.kr/hw/test.html
    - A sample page containing an image
  - http://mnet.yonsei.ac.kr/hw/test2.html
    - A sample page containing two images (jpg and svg)

Deliverable

- **Only one zip file of "YourID_3.zip"**
  - **If your ID is 2021147123, 2021147123_3.zip should be your deliverable file name.**
  - **Zip files, not the folder they are in**

- **In the zip file only the two files must be included without any folder**
  - **report.pdf**
  - **prx.py or prx.c(cpp)**
    - if you use C/C++ language, include build.sh as well
    - Files with different format/name will be a deduction [2pts each]

•DUE DATE

- •12/Dec/2023 23:59:59 KST
- •**No exception for exceeding deadline**

•Delay Policy

- •-33%pts for ~13/Dec 23:59:59
- •-66%pts for ~14/Dec 23:59:59
- •-100%pts for 15/Dec 00:00:00~

You will get 0 points if you…

- **Copy your friend's codes**

- **Use a 3rd-party API or codes.**

Score Policy
Maximum Score = 100pts

1	Not submitted / not working / missing files	**0 pts**
2	Overdue ➜ Delay	-33% pts/day
3	**The rules or directions whose scores are not specified/ are not followed**	-10 pts/rule
4	**A 3rd party framework is used**	**0 pts**
5	**Plagiarizing / Over-implementation (Any kinds of Suspicion of Code-copy)**	**0 pts**
6 7	**Impolite Report / Lack of Comments**	**0 pts/ -50% pts**