

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN



BÁO CÁO ĐỒ ÁN MÔN HỌC
LẬP TRÌNH SOCKET
Môn: Mạng Máy Tính

Lớp: 19CTT3

Thành viên nhóm:

Lê Văn Định 19120477

Trần Hữu Đức 19120484

Nguyễn Nhật Duy 19120495

February 10, 2021

TP. Hồ Chí Minh, Tháng 12 Năm 2020

Mục lục

I Giới thiệu	3
1. Thành viên nhóm:	3
2. Đề tài: Lập trình Socket - Proxy Server	3
II Các hàm chức năng	3
1. Các biến toàn cục:	3
2. Các hàm chức năng:	4
3. Cách chạy chương trình:	5
III Quy trình chạy chương trình	6
IV Bắt gói tin HTTP, mô tả quá trình dữ liệu giữa Client – Proxy Server và Proxy Server – Web Server	10
1. Phân tích gói tin khi truy cập trang caoanxoa.azweb.vn:	10
2. Phân tích gói tin khi truy cập vào trang web trong blacklist.conf:	11
V Tổng kết <i>Giải thích tại sao lại cần Proxy Server trong thực tế</i>	12
1. Máy chủ Proxy trong thực tế:	12
2. Chức năng của chương trình Proxy:	12

I. Giới thiệu

1. Thành viên nhóm:

Họ và Tên	Nhiệm vụ	Mức độ hoàn thành
Lê Văn Định	Tổng hợp thông tin, thiết kế trình bày báo cáo đề án.	100%
Trần Hữu Đức	Quay và chỉnh sửa video mô tả quá trình chạy thử chương trình proxy.	100%
Nguyễn Nhật Duy	Lập trình chương trình proxy server.	100%

2. Đề tài: Lập trình Socket - Proxy Server

- Ngôn ngữ lập trình: Python.
- Chương trình Proxy Server sử dụng cho giao thức HTTP thỏa mãn những yêu cầu sau:
 - Proxy Server cho phép Client truy cập website thông qua các method: GET, POST.
 - Proxy Server chỉ hỗ trợ HTTP, không cần hỗ trợ HTTPS.
 - Proxy Server phải xử lý đồng thời được các request từ client.
 - Proxy Server sẽ lắng nghe tại port 8888, chờ kết nối từ client.
 - Proxy Server cấu hình 01 file **blacklist.conf**, mỗi dòng chứa các domain website cấm Client truy cập. Proxy Server phải chặn tất cả các truy cập đến các domain có trong danh sách này.

Ví dụ cho file blacklist.conf:

```
abc.com.vn
www.thongtin.net
www.lucky88.com
```

- Mức độ hoàn thành đề án: 100%.

II. Các hàm chức năng

1. Các biến toàn cục:

- LISTEN_PORT: cổng proxy phục vụ các kết nối.
- BUFFER_SIZE: kích thước vùng dữ liệu tiếp nhận và gửi đi của proxy.

- MAX_CONNECT: số kết nối tối đa mà proxy có thể phục vụ.
- BLACKLINKS: các địa chỉ không được truy cập.

2. Các hàm chức năng:

❖ connectServer(webServer, port, connect, address, request)

- Tham số:
 - webServer: máy chủ mà trình duyệt yêu cầu được kết nối.
 - port: cổng giao tiếp.
 - connect: kết nối TCP/IP giữa proxy và trình duyệt đã được thiết lập từ trước.
 - address: địa chỉ IP và port của trình duyệt.
 - request: gói tin yêu cầu của trình duyệt đã gửi cho proxy, tiếp tục được truyền đến cho máy chủ được yêu cầu.
- Chức năng: mở một kết nối socket đến server mà trình duyệt yêu cầu, gửi yêu cầu hợp lệ của người dùng đến server này và nhận lại các phản hồi, tiếp tục trả về các phản hồi này cho người dùng.
- Kết quả: Proxy trả về tất cả phản hồi của máy chủ được yêu cầu về lại cho client.

❖ accessLog(webServer, port, time, code)

- Tham số:
 - webServer: máy chủ mà trình duyệt yêu cầu được kết nối.
 - port: cổng giao tiếp.
 - time: Thời điểm hiện tại.
 - code: Mã máy chủ proxy trả về cho trình duyệt.
- Chức năng: Lưu trữ lại thông tin các lần truy cập của trình duyệt vào file access.log. Thông tin lưu trữ bao gồm: máy chủ truy cập, cổng giao tiếp, thời gian truy cập, mã trả về (tượng trưng cho truy cập thành công hay không).
- Kết quả: Thông tin truy cập được lưu trong file access.log

❖ handleRequest(connect, address, request)

- Tham số:
 - connect: kết nối TCP/IP được thiết lập giữa trình duyệt và máy chủ proxy.
 - address: địa chỉ IP và port của trình duyệt.
 - request: gói tin yêu cầu của trình duyệt.
- Chức năng: Xử lý gói tin yêu cầu của trình duyệt, trích xuất địa chỉ máy chủ muốn kết nối và cổng giao tiếp. Nếu địa chỉ máy chủ hợp lệ (không xuất hiện trong file blacklist.conf), gọi hàm connectServer tiếp

tục xử lí. Nếu không, máy chủ proxy trả về mã 403, không kết nối địa chỉ máy chủ đó.

- Kết quả: Phân tích gói tin yêu cầu của trình duyệt, gọi hàm `connectServer` để xử lí tiếp. Mỗi lần xử lí các yêu cầu của trình duyệt, gọi hàm `accessLog(...)` để lưu trữ lại thông tin lần truy cập.
- ❖ `ProxyServer()`: Hàm chính.
 - Tham số: không yêu cầu tham số đầu vào.
 - Chức năng: Khởi tạo máy chủ proxy và chờ tiếp nhận các kết nối từ trình duyệt tới.
 - Kết quả: Tiếp nhận kết nối từ trình duyệt, gọi hàm `handleRequest` để xử lí các yêu cầu này.

3. Cách chạy chương trình:

- Luồng chương trình hoạt động là:
`ProxyServer` -> `handleRequest` -> `connectServer` hoặc trả về 403 Forbidden (nếu trang web bị cấm).
- Cấu hình proxy trong trình duyệt firefox: 127.0.0.1 – port 8888.
- Chạy chương trình `proxy.py` trong cmd: `$ python proxy.py`
- Truy cập trang web bất kì bằng trình duyệt Firefox để dùng thử máy chủ proxy đang chạy.

III. Quy trình chạy chương trình

Sau đây nhóm sẽ trình bày phần demo chạy chương trình.

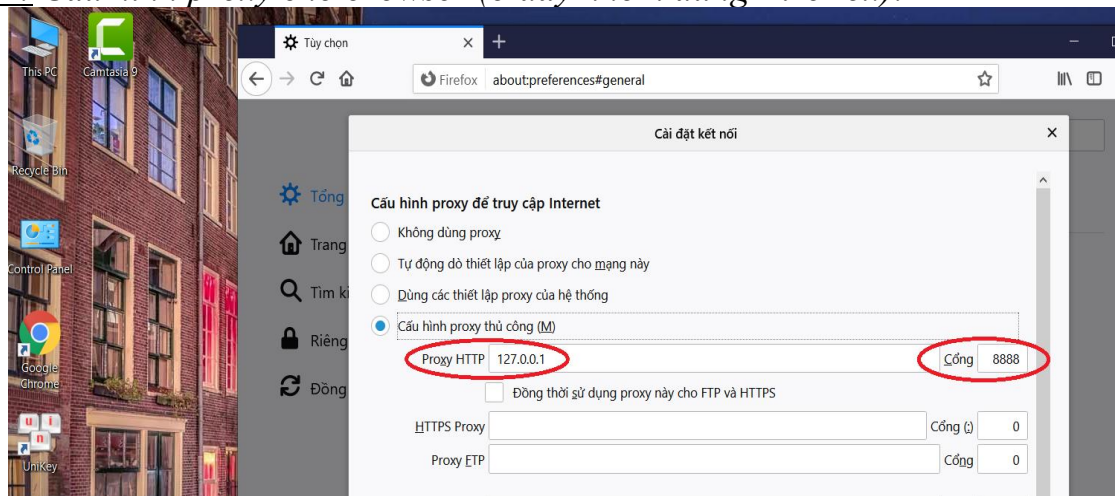
Để biết thêm chi tiết, truy cập đường link sau để xem kết quả chạy thử proxy server và bắt các gói tin bằng Wireshark:

<https://drive.google.com/drive/u/1/folders/1IR5oS31dd2nt5L-x4JT-0-exwBYoXX9Y>

(Nhóm làm link ở chế độ chỉ có người có link mới xem được để đảm bảo vấn đề không bị sao chép copy)

Các bước thực hiện và minh họa:

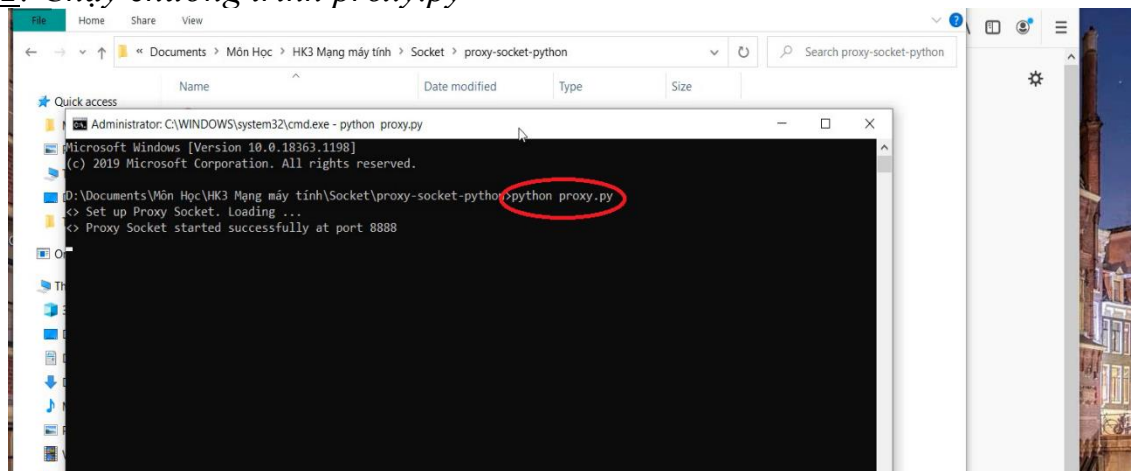
Bước 1: Cấu hình proxy cho browser (ở đây nhóm dùng Firefox):



Cấu hình proxy trong Firefox

- Tại mục Proxy HTTP ta nhập địa chỉ IP 127.0.0.1 và Cổng (hay Port) là 8888.
- Proxy Server bắt đầu lắng nghe tại port 8888 chờ kết nối từ Client.

Bước 2: Chạy chương trình proxy.py



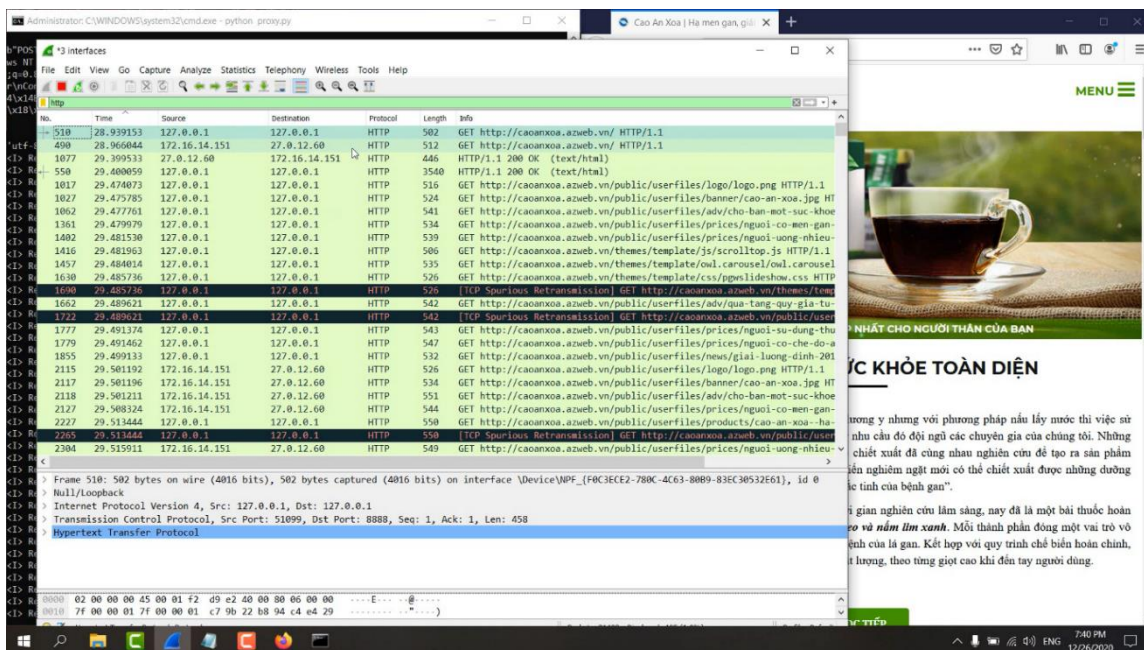
Chạy chương trình proxy server

- Khởi động chương trình proxy server trong CMD.

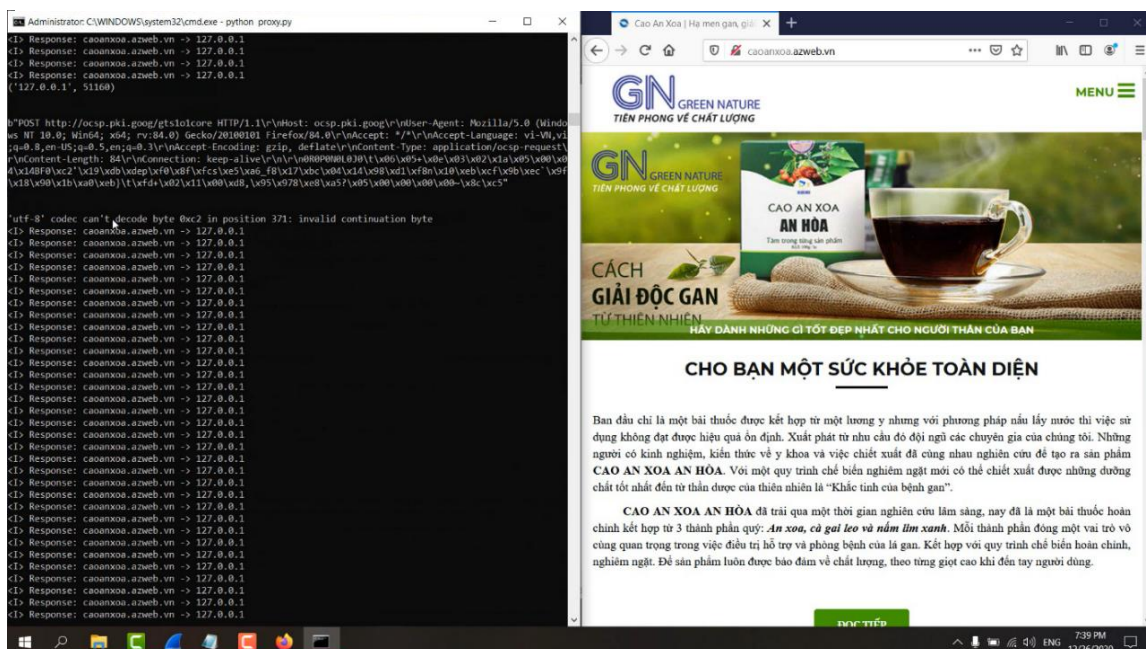
Bước 3: Dùng WireShark để bắt các gói tin HTTP.

Bước 4: Nhập các link trang Web hỗ trợ giao thức HTTP

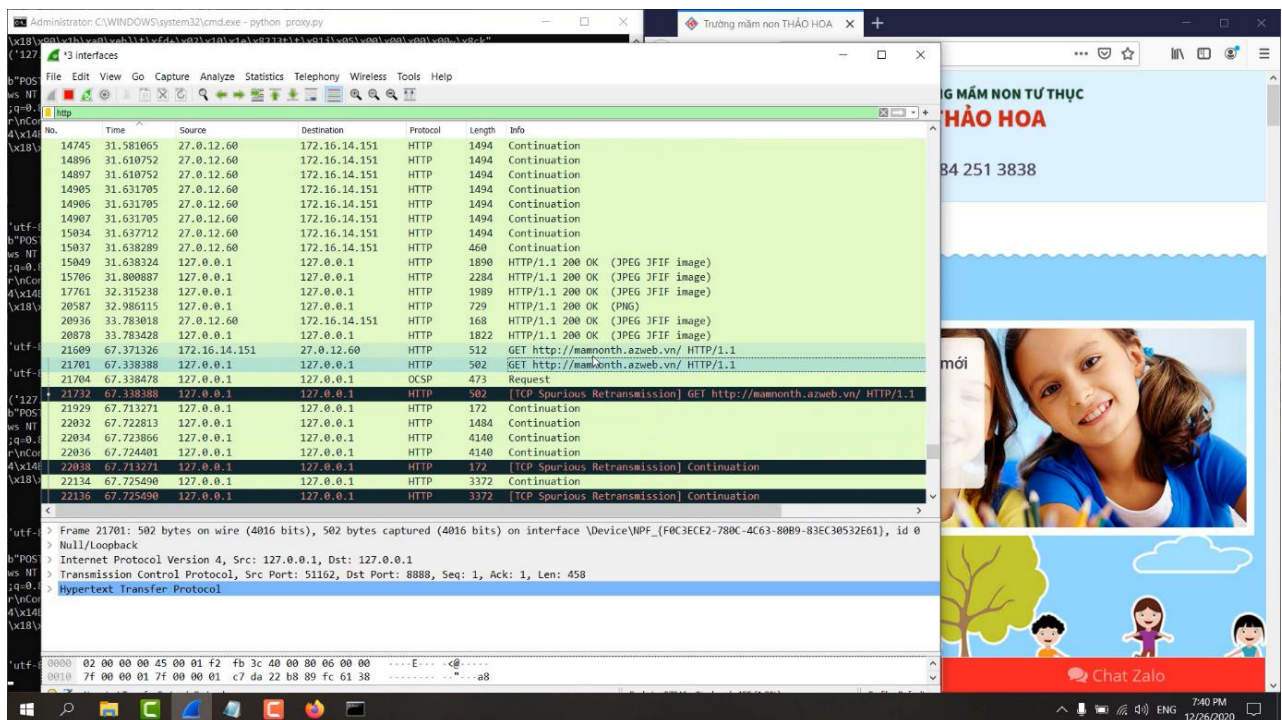
- Ví dụ như: caoanxoa.azweb.vn, mamnonth.azweb.vn, ...
- Sau đây là một số kết quả chạy được:



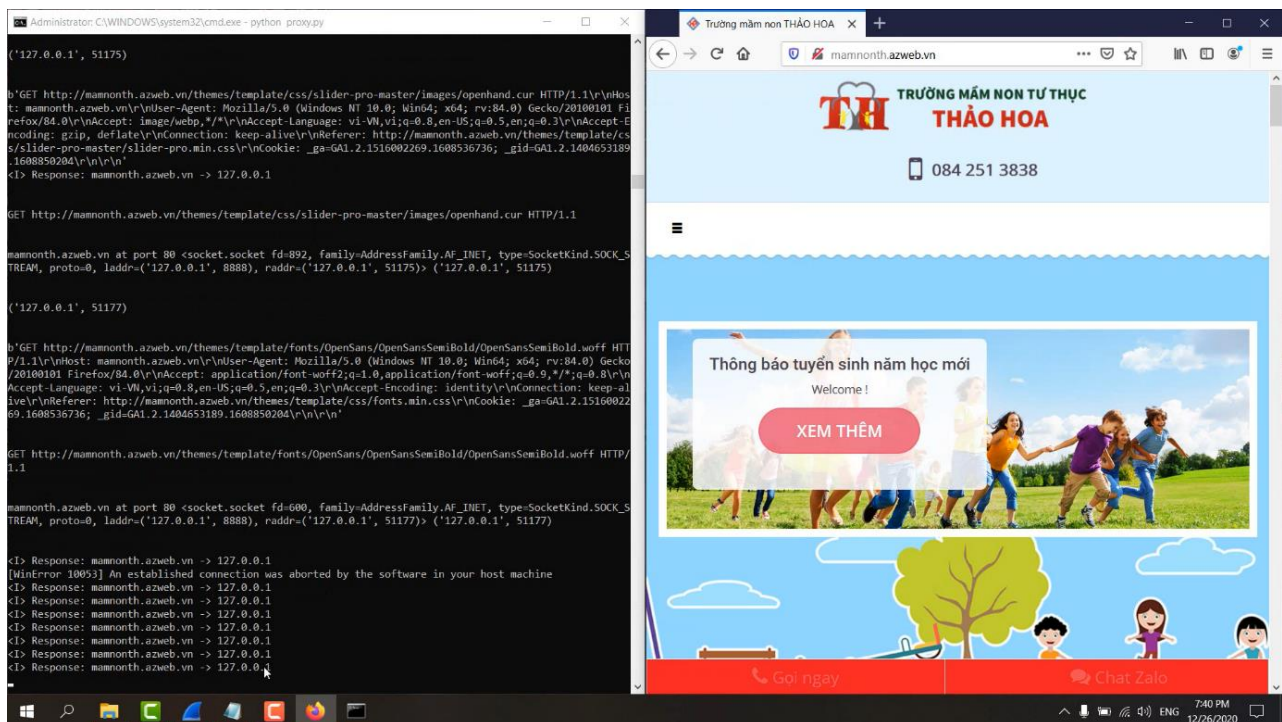
Truy cập trang caoanxoa.azweb.vn



Trang web được trả về hoàn toàn cho client



Truy cập trang mamnonth.azweb.vn



Bước 5: Nhập các link trang web bị cấm trong BlackList

- Ví dụ như tạo một file Blacklist.conf chứa các link như: abc.com.vn, www.thongtin.net, www.lucky88.com.
- Lần lượt nhập các link trên vào trình duyệt kết quả là trình duyệt sẽ trả về trang “403 Forbidden”.

24	0.011165	127.0.0.1	HTTP	391 GET http://abc.com.vn/ HTTP/1.1
34	0.026813	127.0.0.1	HTTP	44 HTTP/1.1 403 Forbidden (text/html)

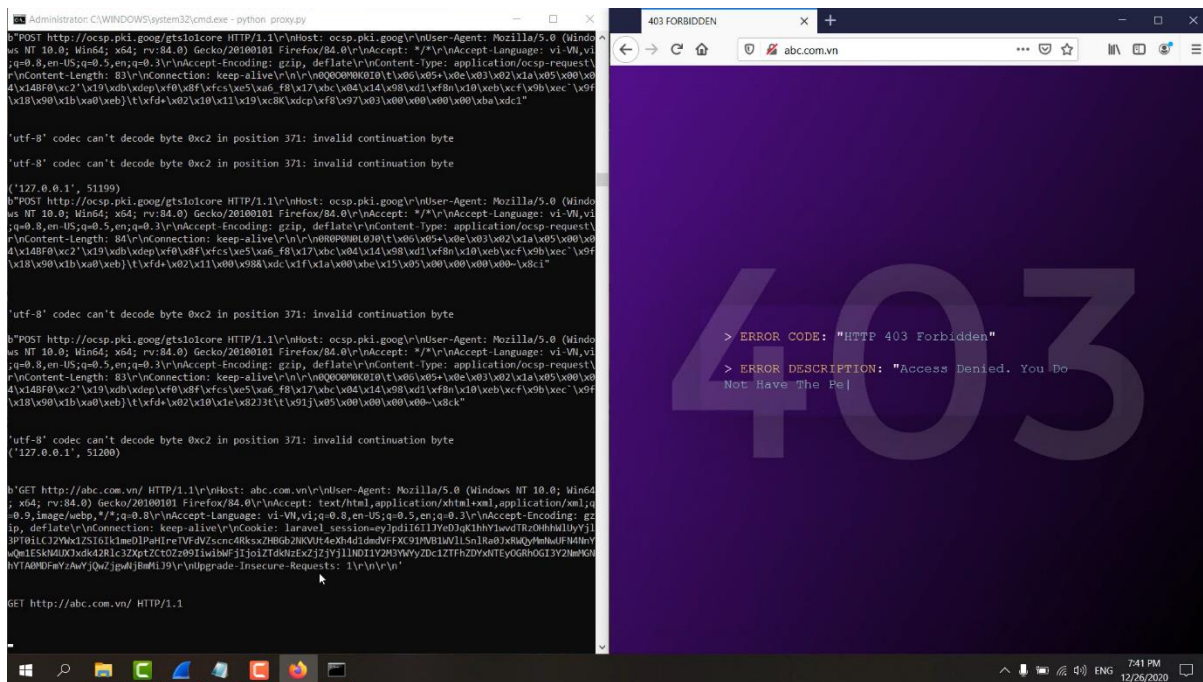
Truy cập trang abc.com.vn

378	18.903993	127.0.0.1	HTTP	403 GET http://www.thongtin.net/ HTTP/1.1
389	18.911284	127.0.0.1	HTTP	44 HTTP/1.1 403 Forbidden (text/html)

Truy cập trang www.thongtin.net

3948	85.098847	127.0.0.1	HTTP	401 GET http://www.lucky88.com/ HTTP/1.1
3959	85.115151	127.0.0.1	HTTP	44 HTTP/1.1 403 Forbidden (text/html)

Truy cập trang www.lucky88.com



Kết quả khi truy cập các trang web bị cấm

IV. Bắt gói tin HTTP, mô tả quá trình dữ liệu giữa Client – Proxy Server và Proxy Server – Web Server

1. Phân tích gói tin khi truy cập trang caoanxoa.azweb.vn:

No.	Time	Source	Destination	Protocol	Length	Info
510	28.939153	127.0.0.1	127.0.0.1	HTTP	502	GET http://caoanxoa.azwe
490	28.966044	172.16.14.151	27.0.12.60	HTTP	512	GET http://caoanxoa.azwe
1077	29.399533	27.0.12.60	172.16.14.151	HTTP	446	HTTP/1.1 200 OK (text/h
550	29.400059	127.0.0.1	127.0.0.1	HTTP	3540	HTTP/1.1 200 OK (text/h

> Frame 510: 502 bytes on wire (4016 bits), 502 bytes captured (4016 bits) on interface \Device\NPF_{F0C3EC}

> Null/Loopback

> Internet Protocol Version 4, Src: 127.0.0.1, Dst: 127.0.0.1

> Transmission Control Protocol, Src Port: 51099, Dst Port: 8888, Seq: 1, Ack: 1, Len: 458

> Hypertext Transfer Protocol

- Gói tin đầu tiên: Web Browser gửi gói tin yêu cầu tới Proxy Server để lấy trang Web **caoanxoa.azweb.vn** thông qua phương thức GET. Nhận thấy **caoanxoa.azweb.vn** không nằm trong blacklist.conf nên Proxy Server cho thông qua và không chặn trang này. Để thấy dst port: 8888 là port của proxy server giống như đã cấu hình.

Src Port: 51099 - Dst Port: 8888

No.	Time	Source	Destination	Protocol	Length	Info
510	28.939153	127.0.0.1	127.0.0.1	HTTP	502	GET http://caoanxoa.azwe
490	28.966044	172.16.14.151	27.0.12.60	HTTP	512	GET http://caoanxoa.azwe
1077	29.399533	27.0.12.60	172.16.14.151	HTTP	446	HTTP/1.1 200 OK (text/h
550	29.400059	127.0.0.1	127.0.0.1	HTTP	3540	HTTP/1.1 200 OK (text/h

> Frame 490: 512 bytes on wire (4096 bits), 512 bytes captured (4096 bits) on interface \Device\NPF_{2BB5B5}

> Ethernet II, Src: LiteonTe_3f:62:13 (00:f4:8d:3f:62:13), Dst: Routerbo_e6:74:7d (b8:69:f4:e6:74:7d)

> Internet Protocol Version 4, Src: 172.16.14.151, Dst: 27.0.12.60

> Transmission Control Protocol, Src Port: 51100, Dst Port: 80, Seq: 1, Ack: 1, Len: 458

> Hypertext Transfer Protocol

- Gói tin thứ hai: Proxy Server gửi gói tin yêu cầu tới Web Server để yêu cầu lấy trang **caoanxoa.azweb.vn** từ Web Server, thông qua phương thức GET. IP của **caoanxoa.azweb.vn** là 27.0.12.60, cũng là IP của Web Server.

Src Port: 51100 - Dst Port: 80

No.	Time	Source	Destination	Protocol	Length	Info
510	28.939153	127.0.0.1	127.0.0.1	HTTP	502	GET http://caoanxoa.azweb.vn/ HTTP/1.1
490	28.966044	172.16.14.151	27.0.12.60	HTTP	512	GET http://caoanxoa.azweb.vn/ HTTP/1.1
1077	29.399533	27.0.12.60	172.16.14.151	HTTP	446	HTTP/1.1 200 OK (text/html)
550	29.400059	127.0.0.1	127.0.0.1	HTTP	3540	HTTP/1.1 200 OK (text/html)

> Frame 1077: 446 bytes on wire (3568 bits), 446 bytes captured (3568 bits) on interface \Device\NPF_{2BB5B5E7-9D93-4C5C-B...
 > Ethernet II, Src: Routerbo_e6:74:7d (b8:69:f4:e6:74:7d), Dst: LiteonTe_3f:62:13 (00:f4:8d:3f:62:13)
 > Internet Protocol Version 4, Src: 27.0.12.60, Dst: 172.16.14.151
 > Transmission Control Protocol, Src Port: 80, Dst Port: 51100, Seq: 7201, Ack: 459, Len: 392
 > [6 Reassembled TCP Segments (7592 bytes): #1072(1440), #1073(1440), #1074(1440), #1075(1440), #1076(1440), #1077(392)]
 > Hypertext Transfer Protocol
 > Line-based text data: text/html (474 lines)

- Gói tin thứ ba: Web Server phản hồi yêu cầu của Proxy Server, trả về mã code 200 tức là OK thành công.

Src Port: 80 - Dst Port: 51100

No.	Time	Source	Destination	Protocol	Length	Info
510	28.939153	127.0.0.1	127.0.0.1	HTTP	502	GET http://caoanxoa.azweb.vn/ HTTP/1.1
490	28.966044	172.16.14.151	27.0.12.60	HTTP	512	GET http://caoanxoa.azweb.vn/ HTTP/1.1
1077	29.399533	27.0.12.60	172.16.14.151	HTTP	446	HTTP/1.1 200 OK (text/html)
550	29.400059	127.0.0.1	127.0.0.1	HTTP	3540	HTTP/1.1 200 OK (text/html)

> Frame 550: 3540 bytes on wire (28320 bits), 3540 bytes captured (28320 bits) on interface \Device\NPF_{F0C3ECE2-780C-4C6...
 > Null/Loopback
 > Internet Protocol Version 4, Src: 127.0.0.1, Dst: 127.0.0.1
 > Transmission Control Protocol, Src Port: 8888, Dst Port: 51099, Seq: 4097, Ack: 459, Len: 3496
 > [2 Reassembled TCP Segments (7592 bytes): #547(4096), #550(3496)]
 > Hypertext Transfer Protocol
 > Line-based text data: text/html (474 lines)

- Gói tin thứ tư: Proxy Server trả lại toàn bộ phản hồi nhận được từ Web Server về lại cho Web Browser, trả về mã code 200 tức là OK, tải trang Web về trình duyệt thành công.

Src Port: 8888 - Dst Port: 51099

2. Phân tích gói tin khi truy cập vào trang web trong blacklist.conf:

Lấy ví dụ trang abc.com.vn (các trang khác trong blacklist.conf cũng tương tự):

No.	Time	Destination	Protocol	Length	Info
24	0.011165	127.0.0.1	HTTP	391	GET http://abc.com.vn/ HTTP/1.1
34	0.026813	127.0.0.1	HTTP	44	HTTP/1.1 403 Forbidden (text/html)

> Frame 3948: 401 bytes on wire (3208 bits), 401 bytes captured (3208 bits) on interface \Device\NPF_Loopback, id 0
 > Null/Loopback
 > Internet Protocol Version 4, Src: 127.0.0.1, Dst: 127.0.0.1
 > Transmission Control Protocol, Src Port: 64101, Dst Port: 8888, Seq: 1, Ack: 1, Len: 357
 > Hypertext Transfer Protocol

Truy cập trang abc.com.vn

- Gói tin đầu tiên Web Browser yêu cầu tới Proxy Server lấy trang web abc.com.vn thông qua phương thức GET.

Src Port: 51200 - Dst Port: 8888

No.	Time	Destination	Protocol	Length	Info
24	0.011165	127.0.0.1	HTTP	391	GET http://abc.com.vn/ HTTP/1.1
34	0.026813	127.0.0.1	HTTP	44	HTTP/1.1 403 Forbidden (text/html)

▶ Frame 3948: 401 bytes on wire (3208 bits), 401 bytes captured (3208 bits) on interface \Device\NPF_{Loopback}, id 0 ▶ Null/Loopback ▶ Internet Protocol Version 4, Src: 127.0.0.1, Dst: 127.0.0.1 ▶ Transmission Control Protocol, Src Port: 64101, Dst Port: 8888, Seq: 1, Ack: 1, Len: 357 ▶ Hypertext Transfer Protocol
--

Truy cập trang abc.com.vn

- Khi chương trình proxy.py chạy phát hiện ra trang web này nằm trong blacklist.conf nên Proxy Server không gửi yêu cầu tới abc.com.vn và trả về mã code 403 cho Client ở gói tin thứ 2.

Src Port: 8888/ Dst Port: 51200

V. Tổng kết

Giải thích tại sao lại cần Proxy Server trong thực tế

1. Máy chủ Proxy trong thực tế:

- Proxy Server (Máy chủ proxy) hoạt động như một cổng nối giữa người dùng và Internet. Đây là một server trung gian giữa người dùng cuối và trang web họ truy cập. Các máy chủ proxy cung cấp các chức năng, bảo mật và riêng tư khác nhau phụ thuộc vào nhu cầu của bạn hoặc chính sách công ty.
- Để kiểm soát việc sử dụng Internet:
 - Tiết kiệm băng thông và cải thiện tốc độ.
 - Bảo mật riêng tư.
 - Cải thiện bảo mật.
 - Truy cập vào các tài nguyên bị chặn.

2. Chức năng của chương trình Proxy:

- Chương trình hoàn thành chức năng cơ bản của máy chủ proxy: Tiếp nhận và xử lý các yêu cầu của client, nếu máy chủ client yêu cầu hợp lệ, tiếp tục kết nối tới máy chủ đó. Nếu không, ngăn chặn client tiếp tục kết nối tới máy chủ này.
- Ngoài ra, chương trình còn có chức năng ghi lại thông tin các lần truy cập của người dùng vào file “access.log”. Thông tin bao gồm: máy chủ yêu cầu, cổng kết nối, thời điểm truy cập và tình trạng kết nối.

