

Trường Đại học Khoa học Tự nhiên
Khoa Công Nghệ Thông Tin

BÁO CÁO ĐỒ ÁN



ĐỒ ÁN 1: SOCKET – PROXY SERVER
MÔN: MẠNG MÁY TÍNH

Thành viên:

Nguyễn Ngọc Minh Khánh	-	1712525
Phạm Đặng Đăng Khoa	-	1712533
Hoàng Anh Kiệt	-	1712542
Nguyễn Thành Ngôn	-	1712617

Mục lục

I. Phân công công việc:	3
II. Những hàm chức năng chính:.....	3
1. getWebServer:	3
2. getResponse:	4
3. getResponseFromWebServer:	4
4. requestThread:	5
5. runServer:	5
6. isForbidden:	5
7. getDateOfCache:	6
8. cacheIsTimeOut:.....	6
III. Cơ chế caching:	6
IV. Mức độ hoàn thành:	7
1. Các chức năng đã hoàn thành:	7
2. Các chức năng chưa làm được:	7
3. Mức độ hoàn thành đồ án:	7
V. Cách chạy chương trình và kết quả chạy được:.....	7
VI. Dùng Wireshark bắt gói tin:	11
VII. Tại sao lại cần dùng Proxy Server?	13
VIII. Tài liệu tham khảo:.....	14

I. Phân công công việc:

STT	MSSV	Họ và tên	Công việc
1	1712525	Nguyễn Ngọc Minh Khánh	<ul style="list-style-type: none">• Tìm hiểu giải pháp và các tài liệu hỗ trợ.• Thiết kế chương trình.• Cài đặt phần nhận gói tin response từ Web Server và trả về cho client.
2	1712533	Phạm Đăng Đăng Khoa	<ul style="list-style-type: none">• Tìm hiểu giải pháp và các tài liệu hỗ trợ• Thiết kế chương trình.• Cài đặt phần nhận và xử lý các thông tin từ gói tin request của client.
3	1712542	Hoàng Anh Kiệt	<ul style="list-style-type: none">• Tìm hiểu giải pháp và các tài liệu hỗ trợ• Thiết kế chương trình.• Cài đặt phần chạy đa luồng, caching và chạy Proxy Server (run Server).• Tổng hợp chương trình và viết báo cáo.
4	1712617	Nguyễn Thành Ngôn	<ul style="list-style-type: none">• Tìm hiểu giải pháp và các tài liệu hỗ trợ• Thiết kế chương trình.• Cài đặt phần chặn truy cập các trang web có tên miền thuộc file blacklist.conf.

II. Những hàm chức năng chính:

1. getWebServer:

- Hàm: void getWebServer(const string& request, string& host, string& port, string& url);
- Chức năng: Từ request nhận được từ Client gửi lên, hàm getWebServer sẽ phân tích từ chuỗi request ra các thông tin như tên miền (lưu vào host), port (lưu vào port) và url (lưu vào url).
- Các tham số:
 - + request: là chuỗi chứa request nhận được từ Client gửi lên.
 - + host: là biến dùng để chứa tên miền của Web Server sau khi phân tích từ chuỗi request.
 - + port: là biến dùng để chứa port của Web Server sau khi phân tích từ chuỗi request.
 - + url: là biến dùng để chứa url sau khi phân tích từ chuỗi request.

- Kết quả: Sau khi gọi hàm ta sẽ có các thông tin về Web Server mà Client gửi request như: tên miền (host), port và url.

2. **getResponse:**

- Hàm: void getResponse(const char* HostName, const char* request, const string& dir, SOCKET& Client_ProxyServer)
- Chức năng: nhận Response từ Web Server hoặc từ file Cache (nếu file cache đã tồn tại và chưa time out) để gửi về cho Client thông qua Socket Client_ProxyServer. Đồng thời tạo mới hoặc cập nhật lại (trường hợp bị time out) file cache.
- Các tham số:
 - + HostName: chuỗi tên miền của Web Server.
 - + request: chuỗi request mà client gửi lên.
 - + dir: đây là chuỗi url truyền vào để từ đó đặt tên cho file cache (1 url sẽ có 1 tên file riêng biệt, 1 file cache riêng).
 - + Client_ProxyServer: đây là Socket đã gửi request lên cho Proxy Server. Response (nhận được từ Web Server hoặc từ file Cache) sẽ được gửi trả về cho client thông qua Socket này.
- Kết quả: sau khi gọi hàm client sẽ nhận được gói tin phản hồi từ Web Server hoặc từ file Cache (nếu file cache này đã tồn tại và chưa time out), đồng thời tạo mới (nếu file cache chưa tồn tại) hoặc cập nhật (nếu file cache bị time out) file cache.

3. **getResponseFromWebServer:**

- Hàm: void getResponseFromWebServer(const char* HostName, const char* request, SOCKET& Client_ProxyServer, fstream& fo)
- Chức năng: tạo Socket Proxy_WebServer kết nối đến Web Server thông qua tên miền (Host Name) để nhận gói tin phản hồi (response), sau đó gửi gói tin response về cho client thông qua Socket Client_ProxyServer đồng thời file lại file cache tương ứng (nếu thỏa điều kiện để cache). (Hàm này dùng để phục vụ cho hàm getResponse ở mục II.2)
- Các tham số:
 - + HostName: chuỗi tên miền của Web Server.

- + request: gói tin request mà client gửi lên bằng Socket Client_ProxyServer.
- + Client_ProxyServer: Socket mà nhận gói tin request từ client và trả về gói tin response cho client.
- + fo: đối tượng đang mở file cache, dùng để ghi lại file cache.
- Kết quả: sau khi gọi hàm sẽ nhận client sẽ nhận được gói tin response mà Proxy đã nhận từ Web Server đồng thời tạo mới hoặc cập nhật lại file cache.

4. requestThread:

- Hàm: unsigned __stdcall requestThread(void *)
- Chức năng: hàm điều khiển hoạt động của từng luồng (từng Thread), xử lý các kết nối đến Proxy Server từ hàng đợi. Mỗi luồng sẽ liên tục nhận request từ client và trả response về cho client (từ các kết nối đến Proxy Server), đồng thời ghi lại cache (nếu có). (Hàm viết theo cấu trúc để tạo luồng bằng hàm __beginthreadex).
- Các tham số: (hàm này không có tham số)
- Kết quả: hàm này sẽ luôn hoạt động (lặp vô hạn), liên tục nhận request và gửi response về cho client.

5. runServer:

- Hàm: void runServer()
- Chức năng: là hàm chính của chương trình giúp chạy Proxy Server. Khởi tạo Proxy Server (tạo Socket ProxyServer với port 8888), khởi tạo các luồng (các thread) chạy song song với nhau, lắng nghe và chấp nhận liên tục các kết nối từ client, push các kết nối vào hàng đợi để các thread xử lý các kết nối này.
- Các tham số: (hàm này không có tham số)
- Kết quả: hàm hoạt động liên tục (lặp vô hạn) giúp chạy Proxy Server.

6. isForbidden:

- Hàm: bool isForbidden(const string& host)
- Chức năng: kiểm tra tên miền (chuỗi host) có nằm trong file blacklist.conf (bị cấm truy cập) hay không ?

- Các tham số:
 - + host: tên miền cần kiểm tra
- Kết quả: nếu host nằm trong blacklist thì trả về true, ngược lại trả về false.

7. getDateOfCache:

- Hàm: bool getDateOfCache(fstream& fi, string& date)
- Chức năng: lấy chuỗi Date từ nội dung trong file cache lưu vào date.
- Các tham số:
 - + fi: đối tượng đang mở file cache để đọc nội dung.
 - + date: chuỗi date dùng để chứa kết quả Date được lấy từ nội dung file cache.
- Kết quả: Nếu thất bại (không lấy được chuỗi Date) thì trả về false. Ngược lại, trả về true và chuỗi date dùng để lưu kết quả.

8. cacheIsTimeOut:

- Hàm: bool cacheIsTimeOut(fstream& fi)
- Chức năng: Kiểm tra file cache đã Time Out hay chưa? (1 file cache là chưa Time Out nếu thời gian được lấy từ nội dung file cache và thời gian hệ thống phải cùng 1 ngày và thời gian chênh lệch không quá 1000 giây).
- Các tham số:
 - + fi: đối tượng đang mở file cache để đọc nội dung.
- Kết quả: Trả về true nếu file cache đã Time Out. Ngược lại trả về false.

III. Cơ chế caching:

- Từ mỗi URL của mỗi gói tin request, chương trình sẽ tạo ra 1 tên file riêng để lưu nội dung của gói tin response nhằm mục đích cache lại nội dung.
- Khi Proxy Server nhận request từ client, Proxy Server sẽ tìm xem file cache tương ứng có tồn tại hay đã bị Time Out chưa. Nếu file cache chưa tồn tại hoặc đã bị Time Out thì Proxy Server sẽ nhận gói tin response từ Web Server trả về cho client đồng thời cập nhật

lại file cache. Nếu file cache chưa time out, hợp lệ thì Proxy Server sẽ lấy nội dung từ file cache gửi về cho client.

- Một file cache chưa Time Out khi ngày giờ trong nội dung file cache và ngày giờ của hệ thống (máy tính) phải cùng 1 ngày và thời gian chênh lệch nhau không quá 1000 giây.

IV. Mức độ hoàn thành:

1. Các chức năng đã hoàn thành:

- Proxy Server hỗ trợ cho giao thức HTTP.
- Cho phép Client truy cập website thông qua Proxy Server (hỗ trợ 2 phương thức GET và POST).
- Xử lý được đồng thời các request từ Client.
- Chặn tất cả các truy cập đến các website có tên miền (host) nằm trong file blacklist.conf.
- Caching lại những nội dung của trang web để phục vụ cho các client khác nhanh hơn.

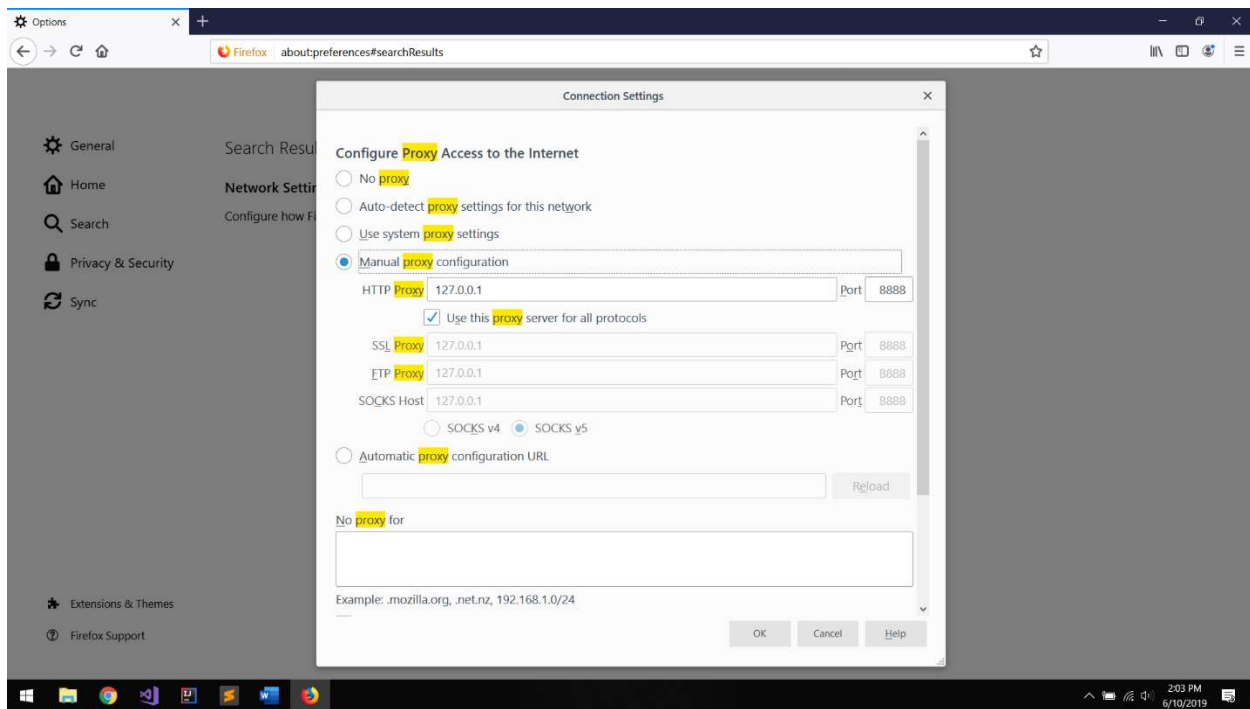
2. Các chức năng chưa làm được: (không có).

3. Mức độ hoàn thành đồ án:

Đã hoàn thành 100% yêu cầu của đồ án.

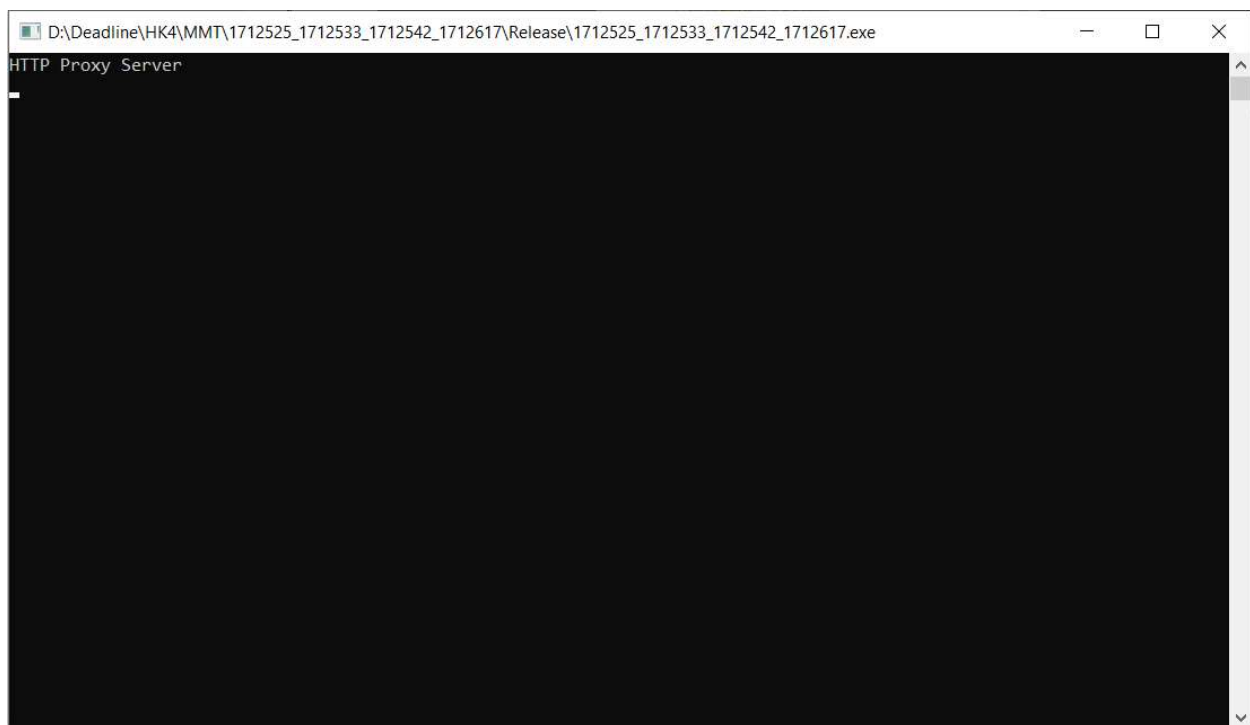
V. Cách chạy chương trình và kết quả chạy được:

- Mở trình duyệt lên và kết nối đến Proxy Server (IP: 127.0.0.1, Port: 8888):

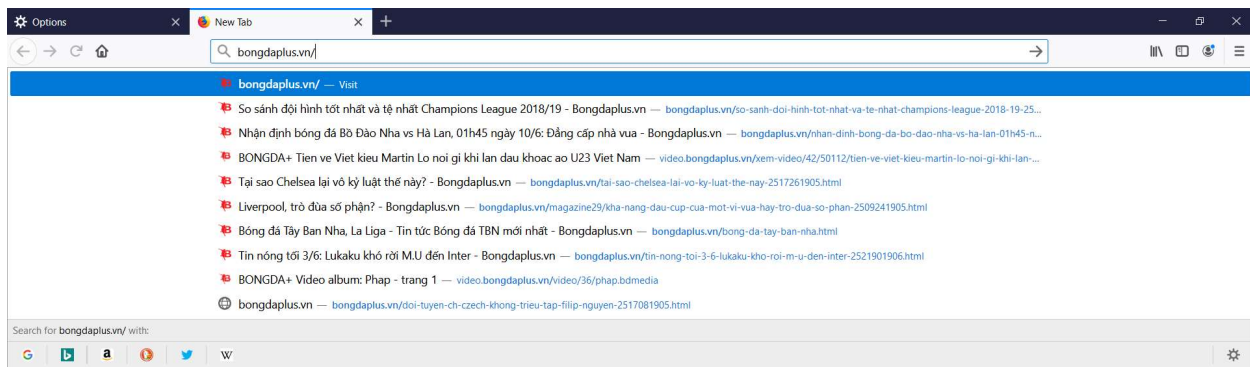


- Chạy chương trình Proxy Server
(1712525_1712533_1712542_1712617.exe):

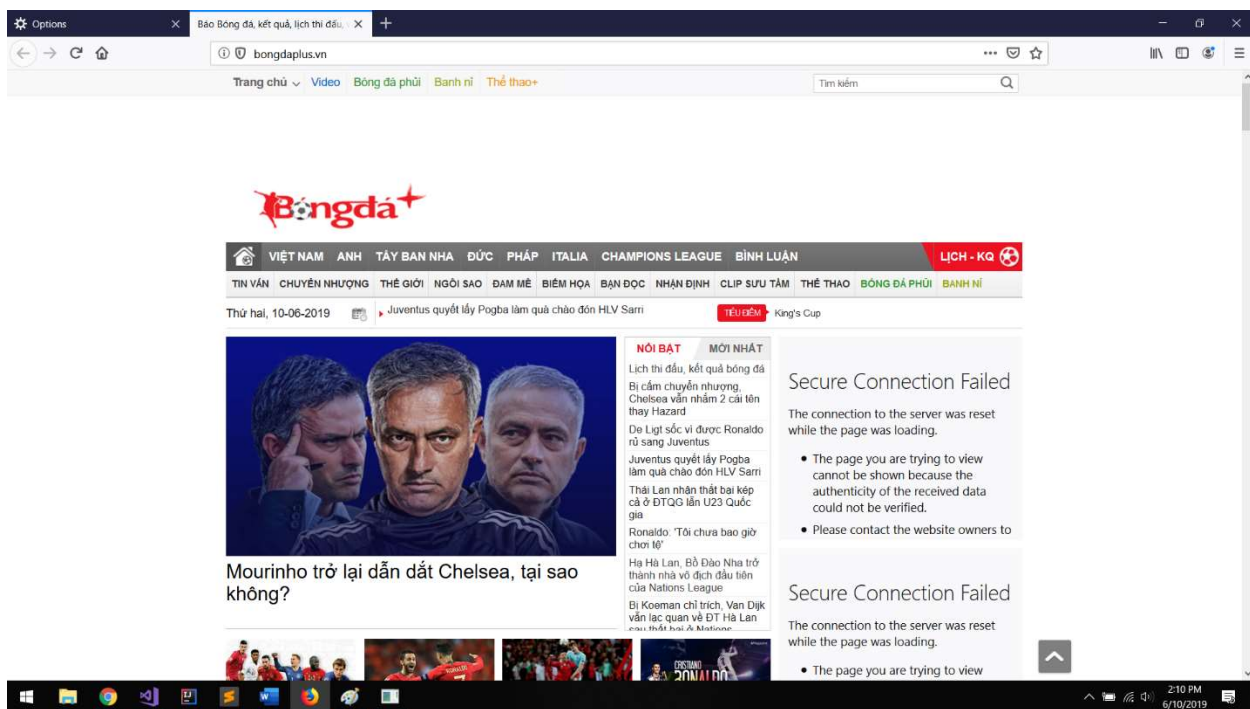
1712525_1712533_1712542_1712617	6/10/2019 11:21 A...	Application	45 KB
---------------------------------	----------------------	-------------	-------



- Vào trình duyệt truy cập đến trang web "bongdaplus.vn":



- Kết quả hiện thị sau khi tải xong trang web:



- Chương trình hiện thị:

```
D:\Deadline\HK4\MMT\1712525_1712533_1712542_1712617\Release\1712525_1712533_1712542_1712617.exe
Accept-Language: en-US,en;q=0.5
Accept-Encoding: gzip, deflate
Referer: http://bongdaplus.vn/ads/AdTopShare.htm?new=111002
Connection: keep-alive

Close socket
Close socket
Close socket
Close socket
Close socket
Close socket
Accept
GET http://apac-jp-sync.bidswitch.net/ul_cb/sync?ssp=ambient&bsw_own_uid= HTTP/1.1
Host: apac-jp-sync.bidswitch.net
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:67.0) Gecko/20100101 Firefox/67.0
Accept: image/webp,*/*
Accept-Language: en-US,en;q=0.5
Accept-Encoding: gzip, deflate
Referer: http://bongdaplus.vn/
Connection: keep-alive

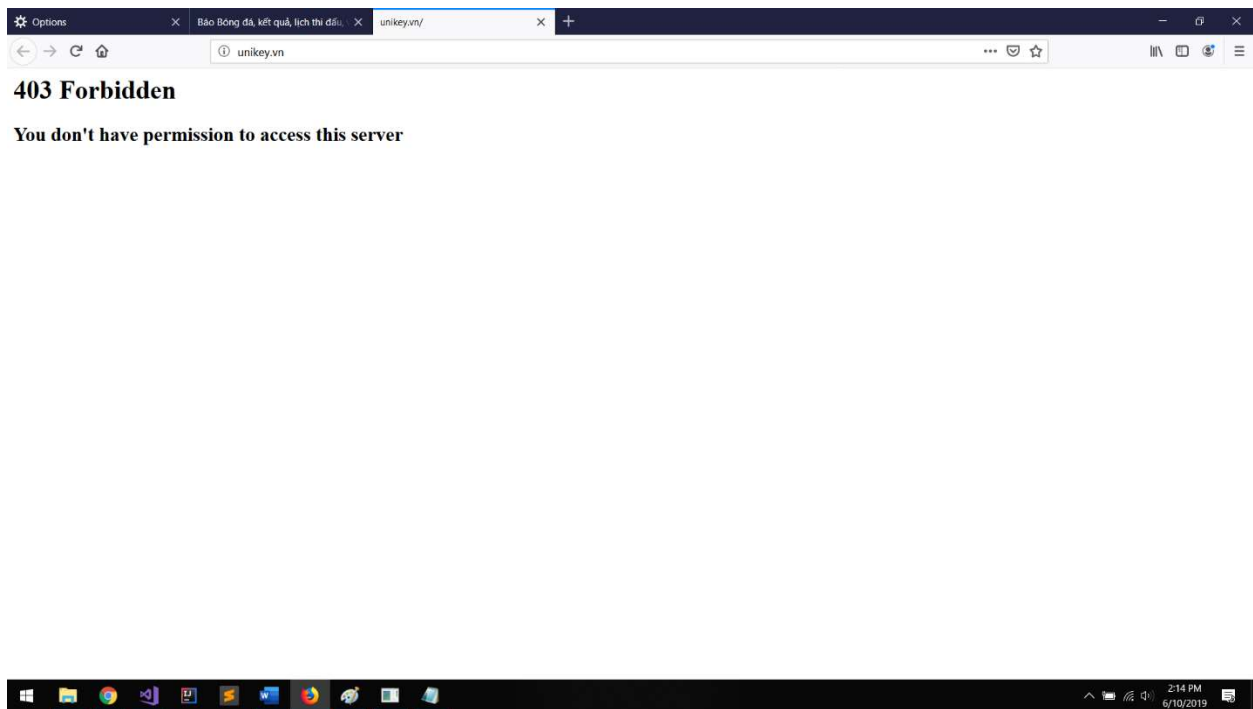
Close socket
Close socket
Accept
GET http://ss.yomedia.vn/sites/0.js HTTP/1.1
Host: ss.yomedia.vn
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:67.0) Gecko/20100101 Firefox/67.0
Accept: */*
Accept-Language: en-US,en;q=0.5
Accept-Encoding: gzip, deflate
Referer: http://bongdaplus.vn/ads/AdTopShare.htm?new=111002
Connection: keep-alive

Close socket
Close socket
Accept
GET http://analytics.yomedia.vn/aud?sw=1536&sh=864&scd=24&l=en-US&ie=j-0&ref=http%3A%2F%2Fbongdaplus.vn%2Fads%2FAdTopShare.htm%3Fnew%3D111002&uref=http%3A%2F%2Fbongdaplus.vn%2F&fgp=4511b5aea46587c37ba180b336b37b6
&wver=1.0.22&coId=1&wId=510&dt=be44f5ac2e0c297a0e9352aa8effc7c2ae55cfa742cb55949aaac09e01e8b1c3ec2cf93b1f3e95a525e5640c93fca1f5b816852dc12be8d9b65461fb404862f77e56b35270e704435b3c7180-9404457c4f4f270d009cfe7b6
fdd0e3774e5ff28016fa6d761224595b85fa03be65d8f3a5f8f4abbc02ac26f67399e4&kord=1560150599046 HTTP/1.1
Host: analytics.yomedia.vn
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:67.0) Gecko/20100101 Firefox/67.0
Accept: image/webp,*/*
Accept-Language: en-US,en;q=0.5
Accept-Encoding: gzip, deflate
Referer: http://bongdaplus.vn/ads/AdTopShare.htm?new=111002
Connection: keep-alive
```

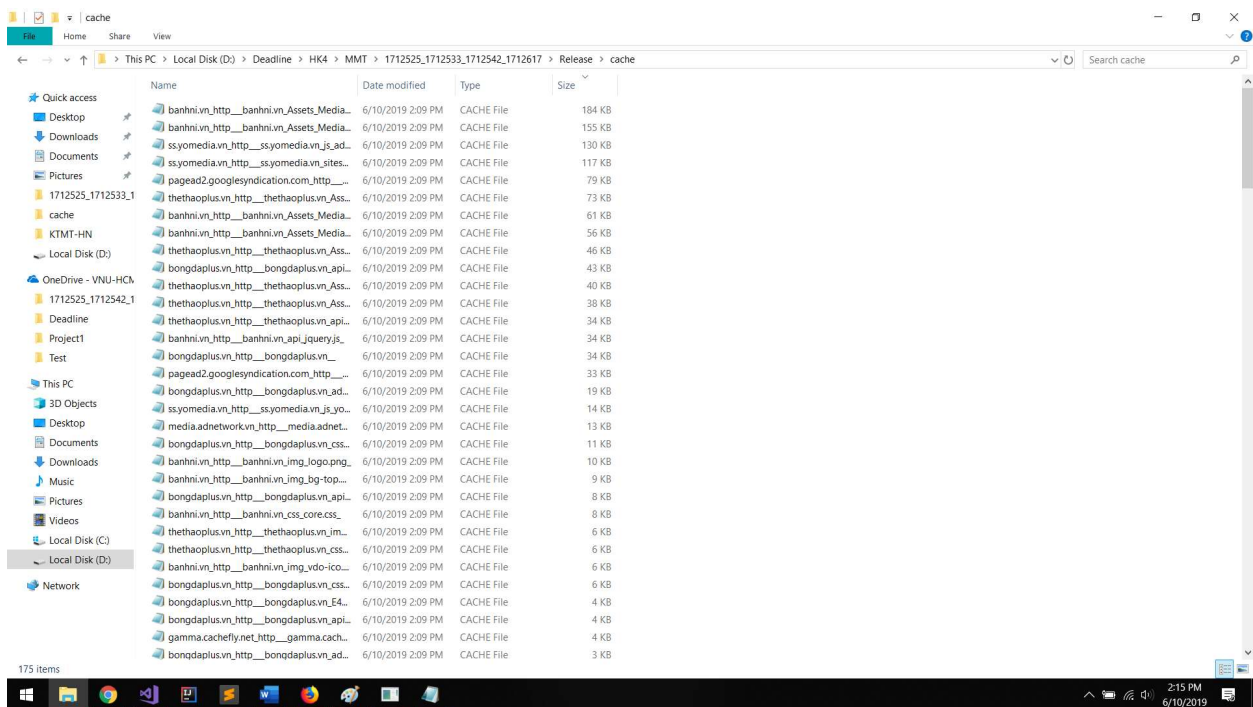
- Thư mục blacklist.conf:

```
blacklist - Notepad
File Edit Format View Help
unikey.vn
www.cplusplus.com
```

- Truy cập trang web "unikey.vn":

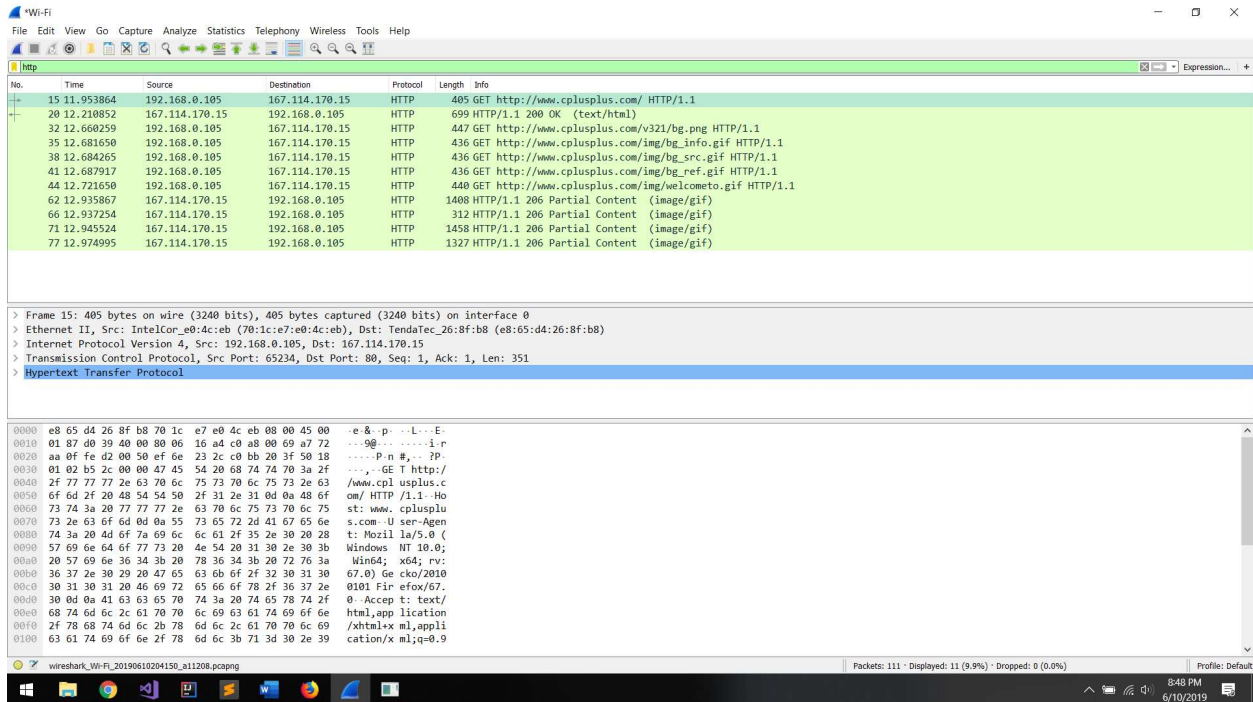


- Thư mục cache (chứa các file cache):

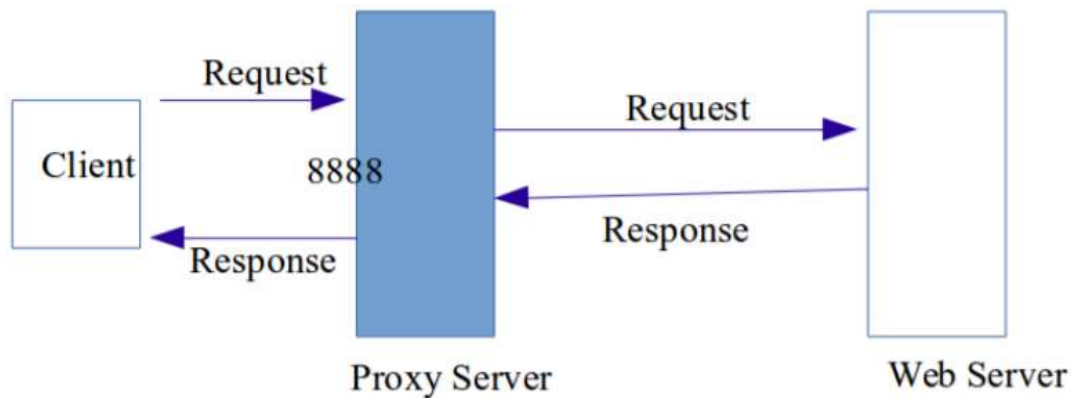


VI. Dùng Wireshark bắt gói tin:

- Các gói tin bắt được khi truy cập trang web "www.cplusplus.com":



- Quá trình gửi nhận dữ liệu giữa Client - Proxy Server và Proxy Server - Web Server:



- + Đầu tiên client sẽ gửi gói http request lên cho Proxy Server.
- + Từ gói tin http request nhận được từ client, Proxy Server sẽ lấy ra những thông tin cần thiết (như host, port, url). Proxy Server sẽ kiểm tra xem host có thuộc blacklist hay không? Nếu có sẽ gửi ngay gói tin http response 403 forbidden về client. Nếu không Proxy Server sẽ kiểm tra nội dung trong file cache còn hạn hay không? Nếu còn thì Proxy Server sẽ gửi nội dung file cache về cho client. Nếu đã time out

hoặc file cache không tồn tại, Proxy Server sẽ chuyển tiếp gói tin http request lên cho Web Server (với thông tin host, port sau khi đã phân tích từ gói tin http request):

```

+ 15 11.953864 192.168.0.105 167.114.170.15 HTTP 405 GET http://www.cplusplus.com/ HTTP/1.1
v Hypertext Transfer Protocol
> GET http://www.cplusplus.com/ HTTP/1.1\r\n
Host: www.cplusplus.com\r\n
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:67.0) Gecko/20100101 Firefox/67.0\r\n
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8\r\n
Accept-Language: en-US,en;q=0.5\r\n
Accept-Encoding: gzip, deflate\r\n
Connection: keep-alive\r\n
Upgrade-Insecure-Requests: 1\r\n
\r\n
[Full request URI: http://www.cplusplus.com/]
[HTTP request 1/1]
[Response in frame: 20]

```

+ Tiếp theo Proxy Server sẽ nhận gói tin http response từ Web Server, sau đó chuyển tiếp về cho Client, đồng thời cập nhật lại file cache.

```

+ 20 12.210852 167.114.170.15 192.168.0.105 HTTP 699 HTTP/1.1 200 OK (text/html)

```

+ Sau đó, tương tự các bước ở trên Client sẽ tiếp tục gửi các gói tin request và nhận các đối tượng đầy đủ cho 1 trang web:

```

+ 35 12.681650 192.168.0.105 167.114.170.15 HTTP 436 GET http://www.cplusplus.com/img/bg_info.gif HTTP/1.1
38 12.684265 192.168.0.105 167.114.170.15 HTTP 436 GET http://www.cplusplus.com/img/bg_src.gif HTTP/1.1
41 12.687917 192.168.0.105 167.114.170.15 HTTP 436 GET http://www.cplusplus.com/img/bg_ref.gif HTTP/1.1
44 12.721650 192.168.0.105 167.114.170.15 HTTP 440 GET http://www.cplusplus.com/img/welcome.gif HTTP/1.1
62 12.935867 167.114.170.15 192.168.0.105 HTTP 1408 HTTP/1.1 206 Partial Content (image/gif)
66 12.937254 167.114.170.15 192.168.0.105 HTTP 312 HTTP/1.1 206 Partial Content (image/gif)
71 12.945524 167.114.170.15 192.168.0.105 HTTP 1458 HTTP/1.1 206 Partial Content (image/gif)
77 12.974995 167.114.170.15 192.168.0.105 HTTP 1327 HTTP/1.1 206 Partial Content (image/gif)

```

VII. Tại sao lại cần dùng Proxy Server?

- **Giúp quản lý được mọi thông tin vào ra:** Do mọi thông tin truy xuất (các gói tin request và response) đều phải thông qua Proxy Server nên giúp ta có thể quản lý được các thông tin vào ra và kiểm soát việc sử dụng của client (Ví dụ: có thể biết được người dùng Proxy Server truy cập vào trang web nào).
- **Giúp tiết kiệm băng thông và tăng tốc độ truy cập:** do Proxy Server có chức năng caching nên giúp hỗ trợ nhiều người dùng tốt hơn, tiết kiệm băng thông và tăng tốc độ tải trang.

- **Có thể cấm client truy cập vào một số trang web:** Proxy Server có thể ngăn client truy cập vào 1 trang web nằm trong danh sách cấm.
- **Giúp bảo mật tốt hơn, truy cập an toàn hơn:** Do mọi gói tin đều phải thông qua Proxy Server, mọi tin request chỉ được gửi đi bởi Proxy Server nên giúp bảo mật tốt hơn, truy cập ẩn danh (Web Server không biết thông tin, địa chỉ IP của client; chỉ biết thông tin của Proxy Server). Ngoài ra, người dùng còn có thể truy cập vào các trang web mà mình bị chặn IP (do thực hiện thông qua Proxy Server).

VIII. Tài liệu tham khảo:

1. Bài giảng Lập Trình Mạng - Lương Ánh Hoàng:

https://www.academia.edu/9813735/B%C3%A0i_gi%E1%BA%A3ng_L%E1%BA%ACP_TR%C3%8CNH_M%E1%BA%A0NG_Ng%C6%B0%E1%BB%9Di_so%E1%BA%A1n_L%C6%B0%C6%A1ng_%C3%81nh_Ho%C3%A0ng_M%E1%BB%A5c_l%E1%BB%A5c_b%C3%A0i_gi%E1%BA%A3ng

2. Thread for Win32:

<https://www.bogotobogo.com/cplusplus/multithreaded2A.php>

3. Hypertext Transfer Protocol -- HTTP/1.1:

<https://www.w3.org/Protocols/rfc2616/rfc2616.txt>

4. Tài liệu Socket: (do giảng viên cung cấp)

--- HẾT ---