

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN



BÁO CÁO ĐỒ ÁN MẠNG MÁY TÍNH

CHỦ ĐỀ: *LẬP TRÌNH SOCKET – PROXY SERVER*

Giảng viên hướng dẫn: Nguyễn Thanh Quân

TP. Hồ Chí Minh, ngày 10/06/2019

MỤC LỤC---

I.	THÔNG TIN THÀNH VIÊN NHÓM VÀ PHÂN CÔNG CÔNG VIỆC:.....	1
II.	BÁO CÁO ĐỒ ÁN:	1
1.	<i>Môi trường lập trình:</i>	1
2.	<i>Chức năng chính của chương trình:</i>	1
3.	<i>Các hàm và chức năng của từng hàm:</i>	1
4.	<i>Cách chạy chương trình và kiểm tra kết quả đạt được:</i>	3
5.	<i>Giao thức trao đổi giữa Web Client với Proxy Server và Proxy Server với Web Server:</i>	9
6.	<i>Sự cần thiết của Proxy Server trong thực tế:</i>	11
III.	ĐÁNH GIÁ MỨC ĐỘ HOÀN THÀNH ĐỒ ÁN:	12
IV.	NGUỒN THAM KHẢO:	13

I. THÔNG TIN THÀNH VIÊN NHÓM VÀ PHÂN CÔNG CÔNG VIỆC:

<i>STT</i>	<i>MSSV</i>	<i>Họ và tên</i>	<i>Phân công công việc</i>
1	1712173	Đặng Thái Gia Thuận	Tìm hiểu nội dung, kiểm thử lỗi, test chương trình, tổng hợp kết quả, viết báo cáo.
2	1712202	Nguyễn Trọng Văn	Tìm hiểu nội dung, xây dựng các chức năng chính của chương trình, tổng hợp kết quả.
3	1712292	Lý Quốc Bình	Tìm hiểu nội dung, kiểm thử lỗi, test chương trình, tổng hợp kết quả, viết báo cáo.

II. BÁO CÁO ĐỒ ÁN:

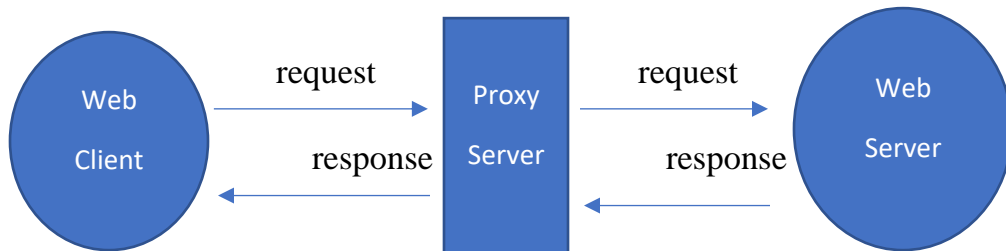
1. Môi trường lập trình:

_Công cụ hỗ trợ: Visual Studio 2013 – 2010.

_Ngôn ngữ lập trình: C/C++.

2. Chức năng chính của chương trình:

Chương trình hỗ trợ tải và hiển thị nội dung các trang web thông qua Proxy Server theo mô hình:



3. Các hàm và chức năng của từng hàm:

<i>STT</i>	<i>Tên hàm</i>	<i>Tham số truyền vào</i>	<i>Chức năng</i>	<i>Kết quả sau khi thực hiện</i>
1	<code>char *get_ip(char* host)</code>	Mảng kí tự là tên miền	Lấy địa chỉ ip từ domain name nhận được	Địa chỉ IP ứng với tên miền.

		của website		
2	<code>wchar_t *convertCharArrayToLPCWSTR(const char* charArray)</code>	Mảng ký tự.	Hàm chuyển mảng kí tự (char*) thành LPCWSTR.	
3	<code>char *get_host(char *request)</code>	Mảng kí tự là thông tin gói request mà Client nhận được từ truy cập trang website qua Web Brower.	Lấy host (domain name) từ gói request nhận được từ Client.	Domain name tương ứng với thông tin GET có được.
4	<code>std::string get_header(char* buff, std::string header)</code>	Mảng kí tự chứa thông tin gói request nhận được, header.	Trả về thông tin của header truyền vào (hàm tổng quát) với kiểu dữ liệu string.	Thông tin header. Ví dụ Domain name (header='Host',...)

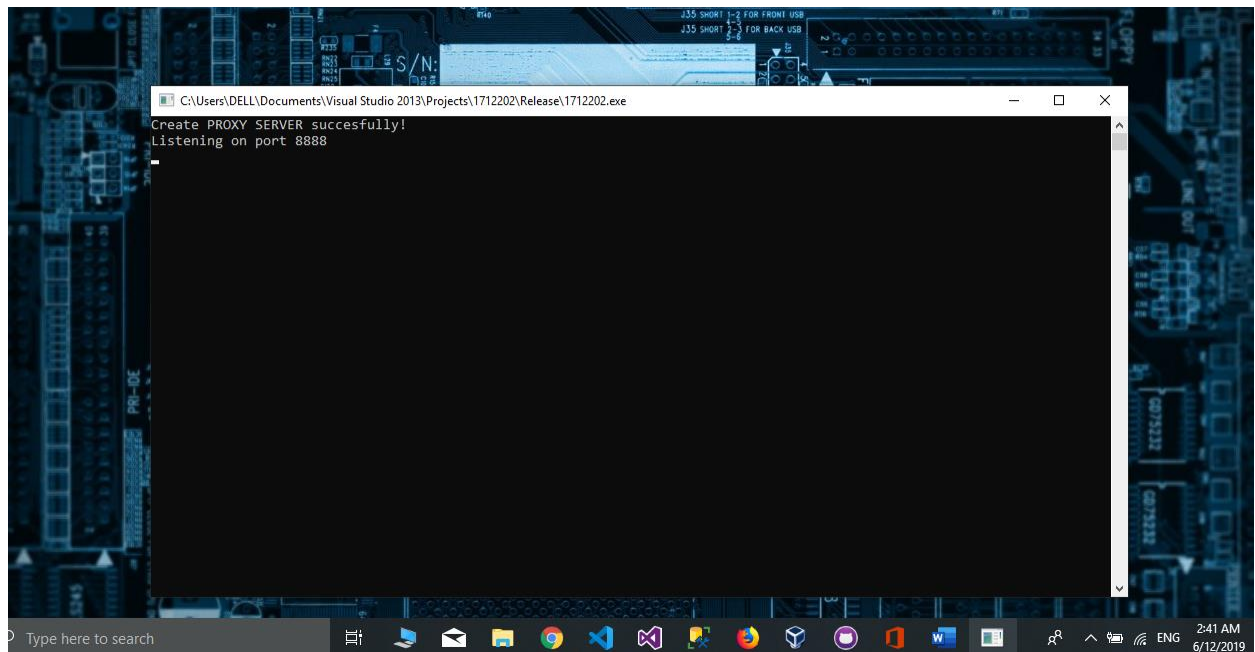
5	<code>DWORD socketSend(CSocket* socket, char* buffer, int size)</code>	Socket, mảng kí tự chứa thông tin gửi, kích thước nội dung thông tin.	Tương tự như cách phương thức <code>Send()</code> trong class <code>CSocket</code> thực hiện nhưng có thêm xử lý kiểm tra lượng thông tin mà đã được gửi có đủ hay chưa thông qua kích thước được truyền vào.	Gửi thông tin Socket nhận được đi.
6	<code>int isBlacklisted(std::string domain)</code>	Chuỗi kí tự là tên miền của website truy cập.	Kiểm tra xem tên miền có nằm trong file <code>blacklist.conf</code> (ngăn phép truy cập) hay không.	Trả về -1 nếu file không tồn tại, 0 nếu tên miền không nằm trong danh sách và 1 nếu có trong danh sách.

4. Cách chạy chương trình và kiểm tra kết quả đạt được:

_Có thể chạy chương trình trực tiếp hoặc thông qua command line.

_Chạy trực tiếp:

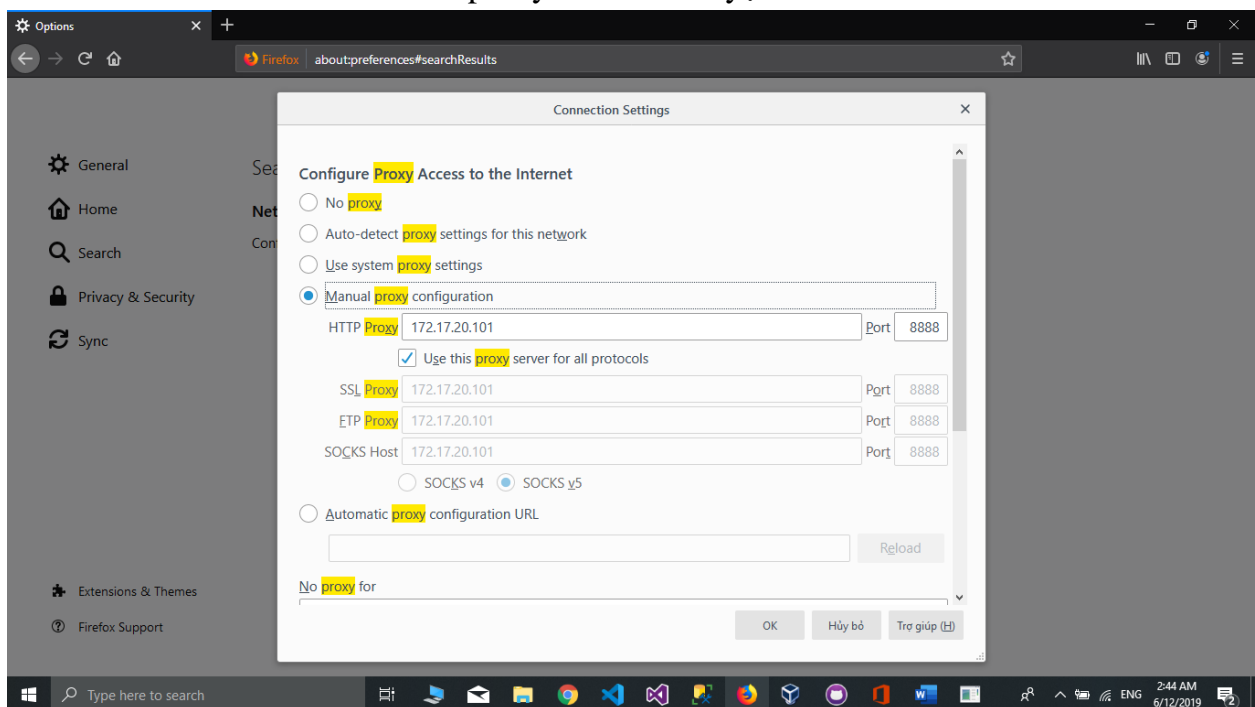
+Mở file *****.exe** trong thư mục Release. Nếu Proxy Server được tạo thành công và đang hoạt động trên Port 8888 ta sẽ thấy chương trình xuất ra thông báo:



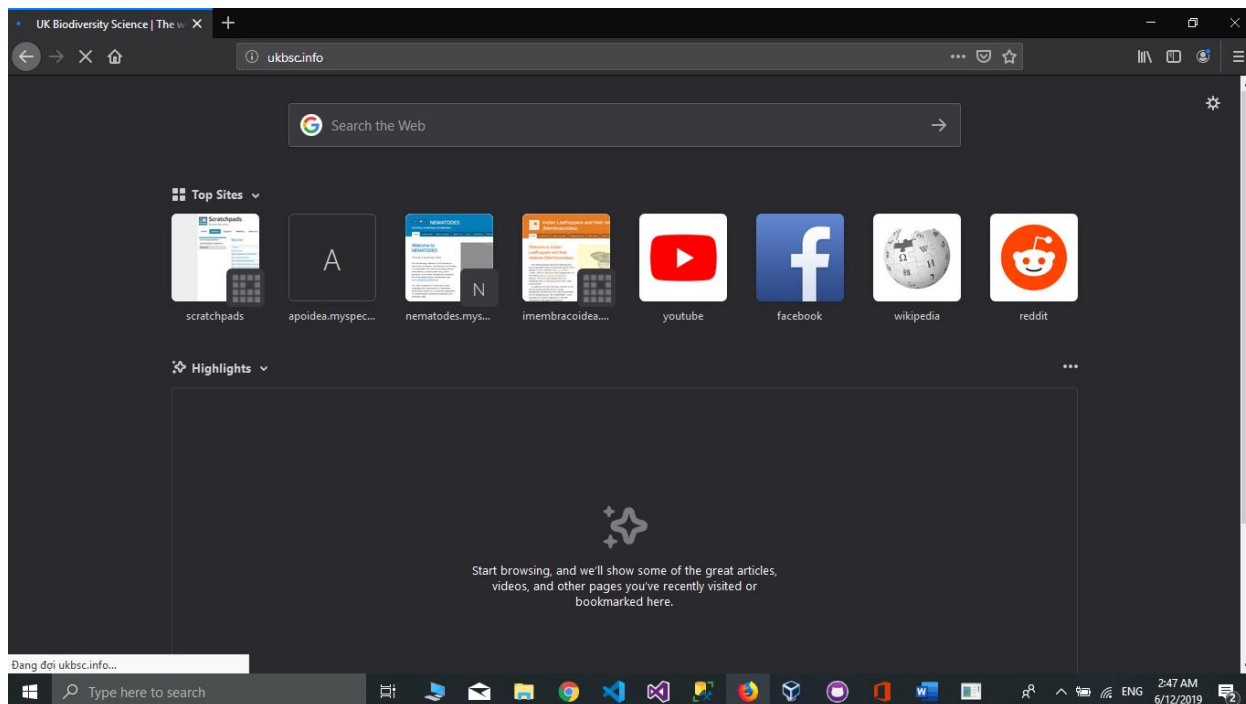
Nếu không chương trình sẽ báo lỗi và thoát.

+Lúc này ta có thể truy cập vào các website thông qua các trình duyệt Mozilla Firefox hoặc Chorme.

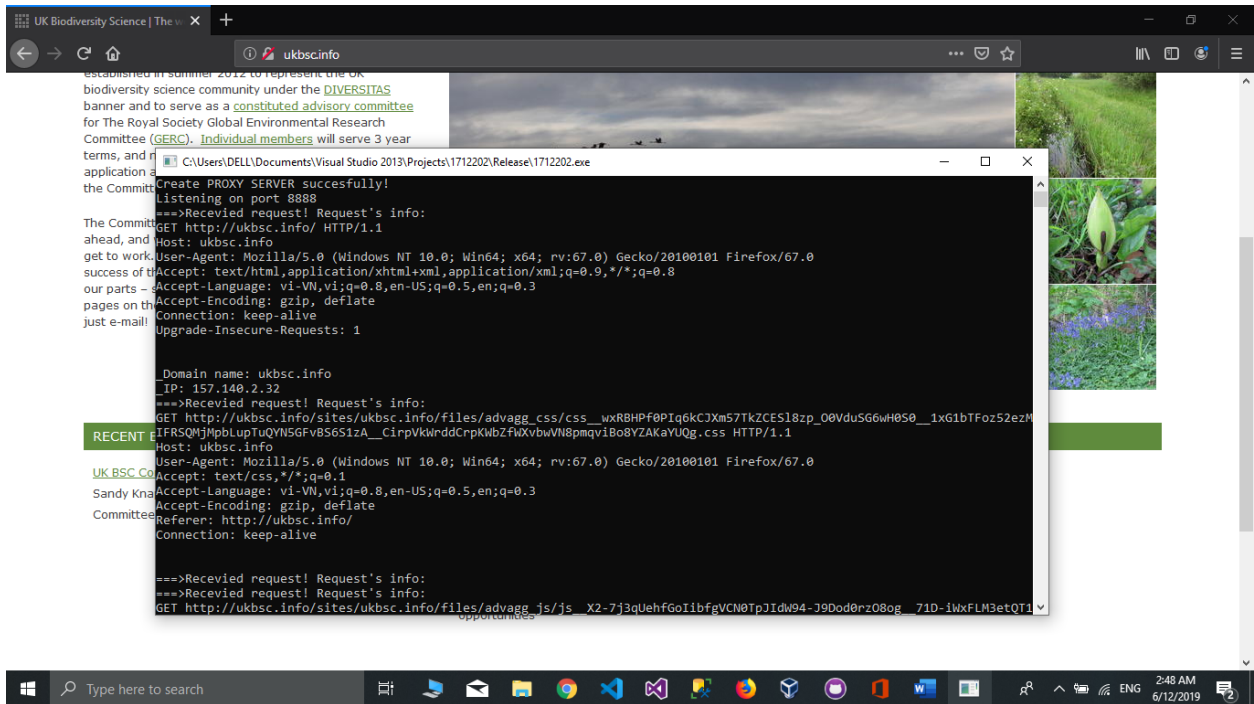
+Mở Firefox và vào *Options*. Trong thẻ tìm kiếm tìm kiếm từ khóa “*proxy*” để tiến hành cấu hình proxy cho trình duyệt.



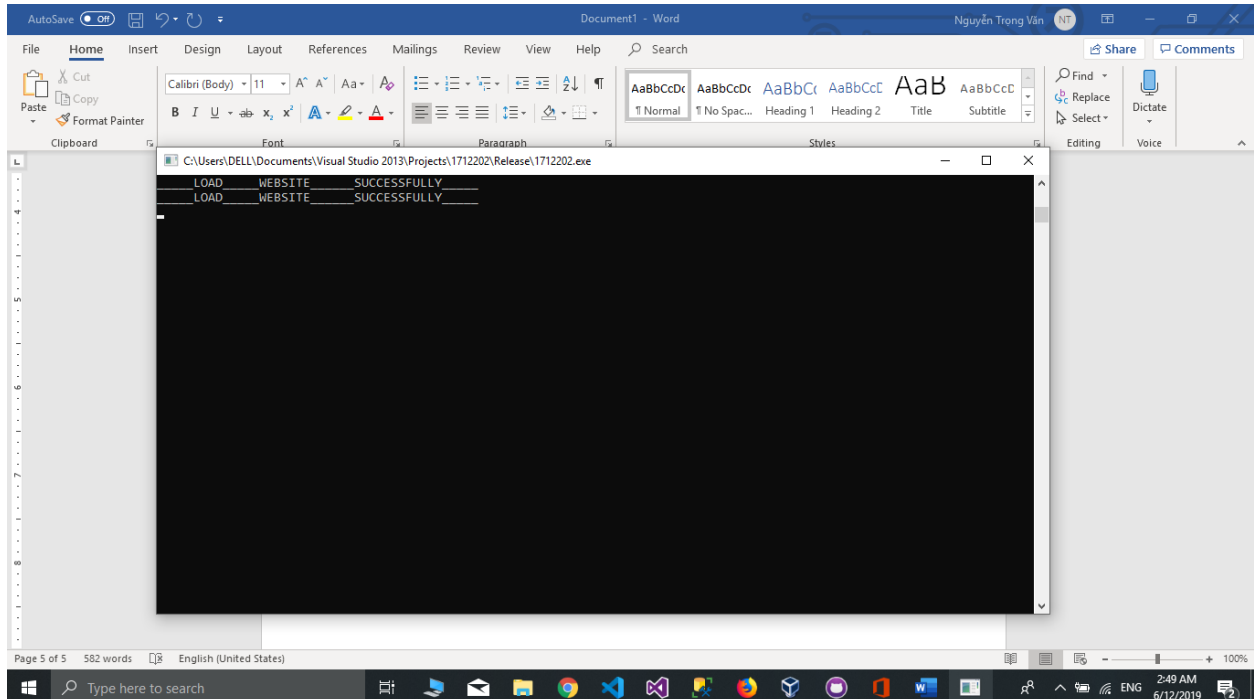
- +Trong thư mục cài đặt hiện ra ta chọn Port proxy là **8888**. HTTP Proxy là địa chỉ IP đường mạng mà máy tính bạn đang sử dụng.
- +Như vậy là đã xong. Cuối cùng dùng url hoặc gõ tên miền web muốn truy cập như các duyệt web bình thường. Ví dụ: <http://ukbsc.info/>



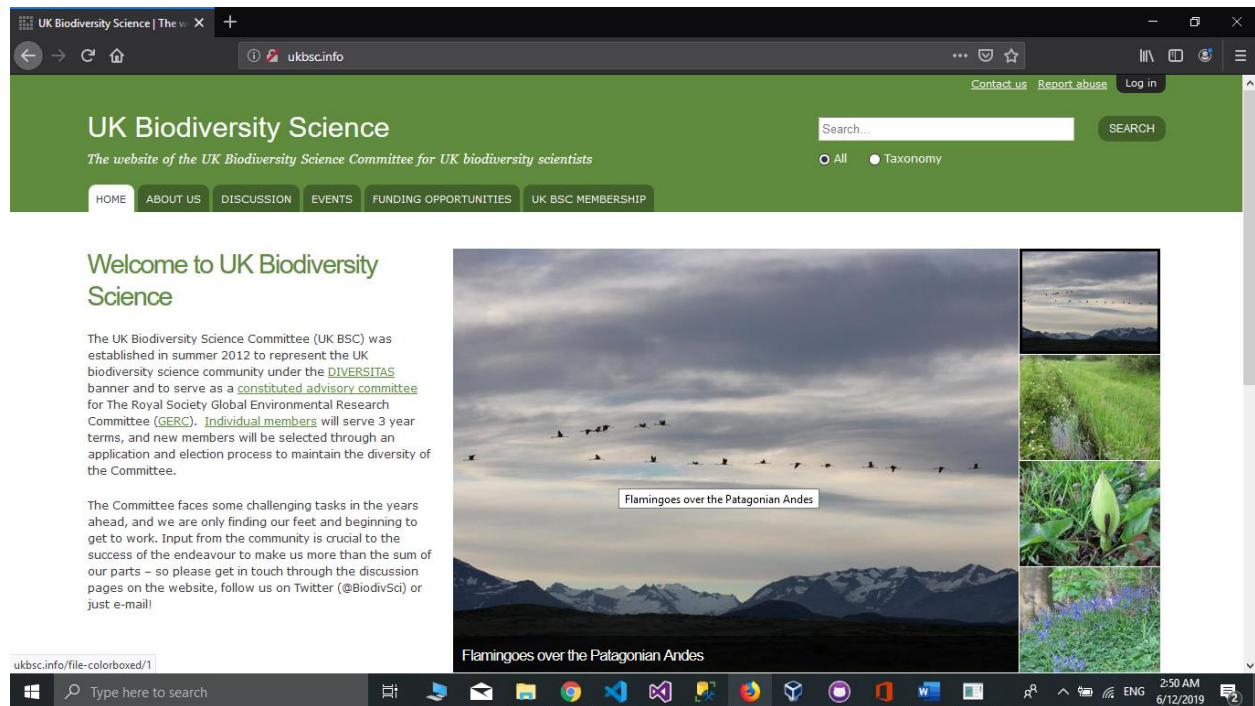
+Chương trình sẽ hiện lên thông tin mà gói tin nhận được cùng với domain name và địa chỉ IP ứng với tên miền mà ta đã truy cập.



+Khi load website thành công xuất hiện thông báo:



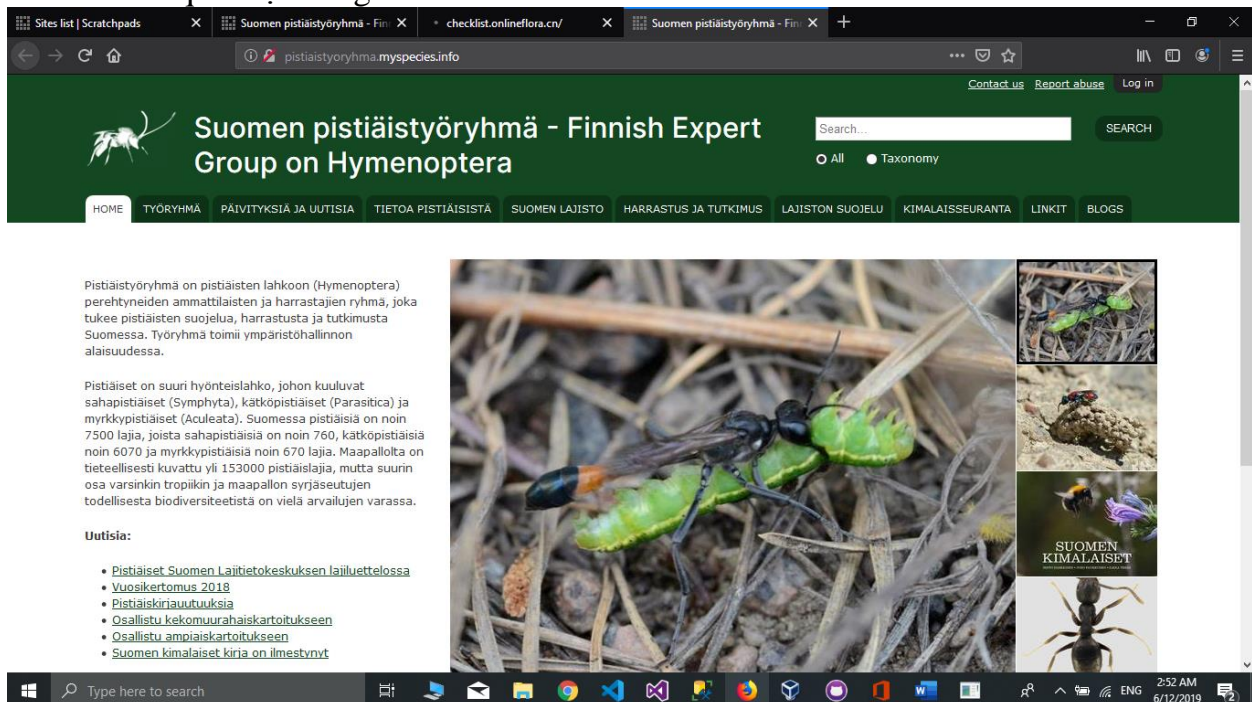
+Kết quả thu được:



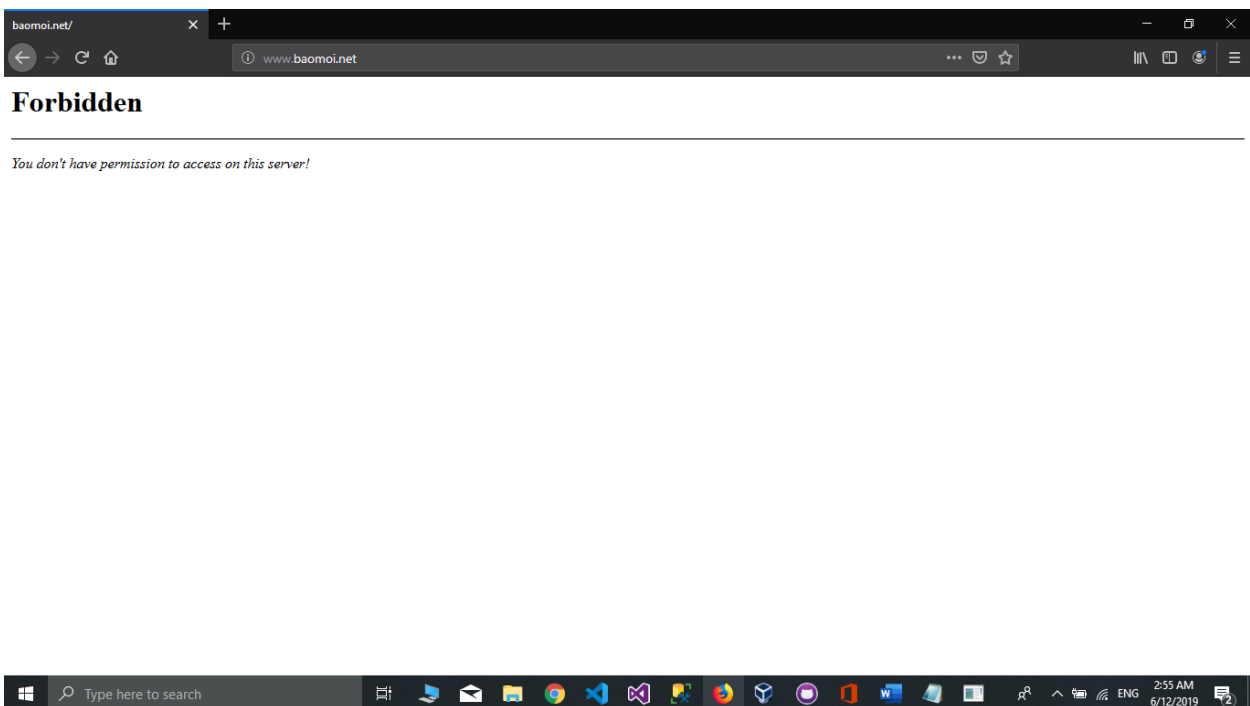
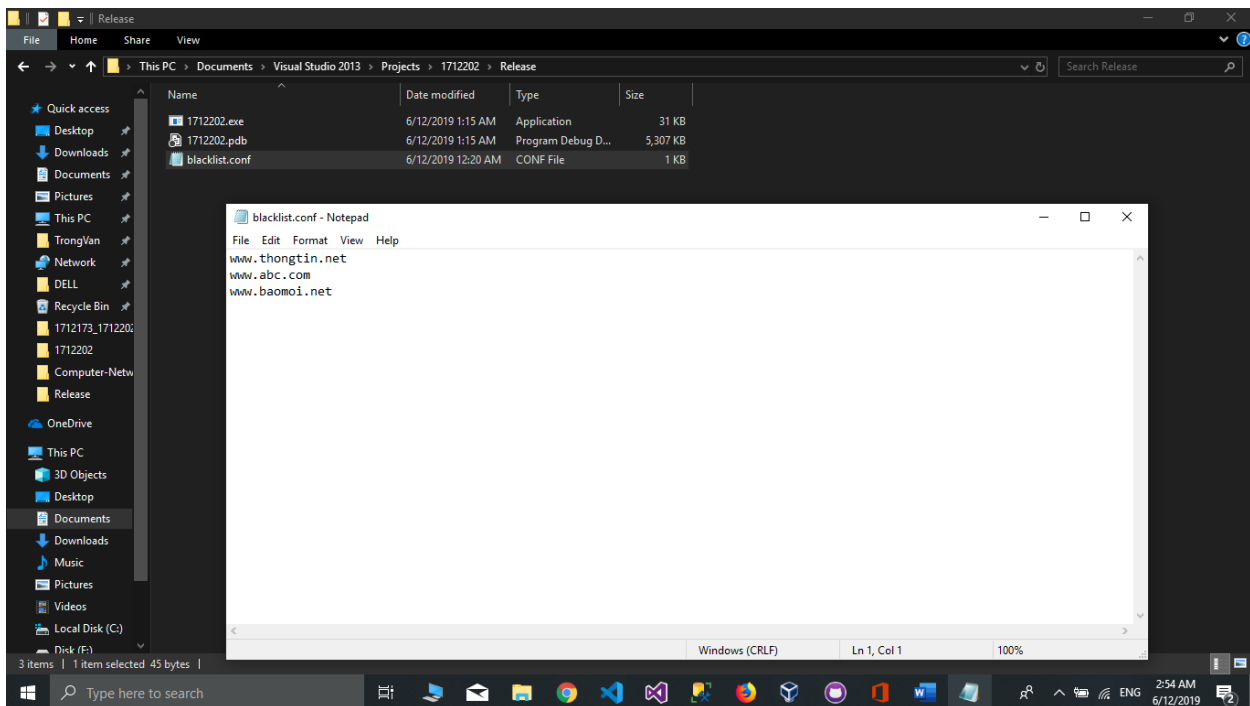
--Luru ý:

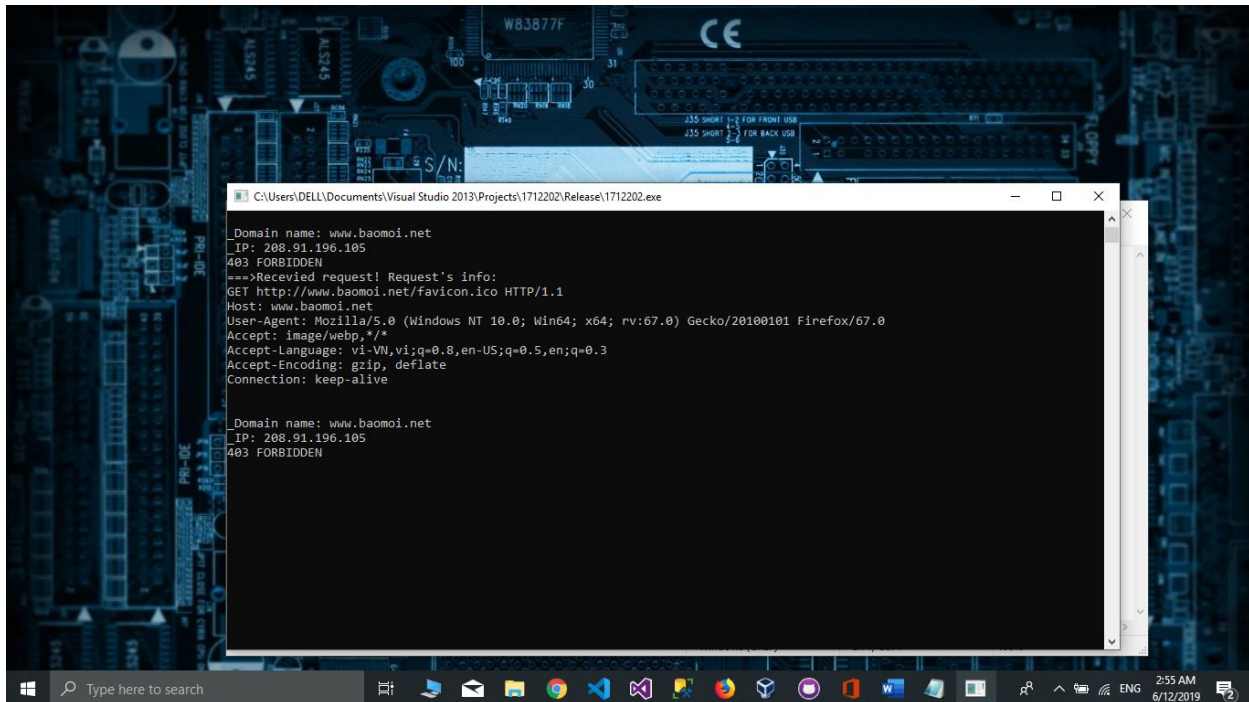
+Theo yêu cầu đề bài, chương trình chỉ hỗ trợ những trang web HTTP.

+Chương trình có thể xử lý đồng thời được nhiều request cùng lúc (đa luồng) qua việc dùng thread:



+Khi truy cập vào các domain được cấu hình trong file blacklist, sẽ trả về kết quả **403 FORBIDDEN**. Ví dụ với www.baomoi.net





5. Giao thức trao đổi giữa Web Client với Proxy Server và Proxy Server với Web Server:

— Khi thực hiện bắt gói tin tại WireShark trên mục http ta sẽ thấy được các trao đổi cụ thể giữa Client và Server:

No.	Time	Source	Destination	Protocol	Length	Info
2419	15.270359	172.17.20.101	157.140.2.32	HTTP	412	GET http://ukbcs.info/ HTTP/1.1
2429	15.559564	157.140.2.32	172.17.20.101	HTTP	766	HTTP/1.1 200 OK (text/html)
2438	15.993667	172.17.20.101	157.140.2.32	HTTP	524	GET http://ukbcs.info/sites/ukbcs.info/files/advagg_js/js_RvAe6WwL9ugGwyngroYoEoq3Gi267f4G3VP6rZAwLlM_eDc...
2441	15.996801	172.17.20.101	157.140.2.32	HTTP	542	GET http://ukbcs.info/sites/ukbcs.info/files/advagg_css/css_voxRBHPf0PIq6kC7Xm57kZCESl8zp_00VduS66wH050_1...
2444	16.004446	172.17.20.101	157.140.2.32	HTTP	524	GET http://ukbcs.info/sites/ukbcs.info/files/advagg_js/js_X2-7j3qUehfG0IibfgVCN07pIdw94-79D0d0rZ08og_71D...
2447	16.048732	172.17.20.101	157.140.2.32	HTTP	542	GET http://ukbcs.info/sites/ukbcs.info/files/advagg_css/css_Slsbo-Urelp3ChBwfv37Ick8XS0l1Iw5xavZDaATc_9...
2450	16.053871	172.17.20.101	157.140.2.32	HTTP	428	GET http://ukbcs.info/sites/all/libraries/mediaelement/build/mediaelement-and-player.min.js?v=2.1.6 HTTP/1...
2476	16.285752	157.140.2.32	172.17.20.101	HTTP	645	HTTP/1.1 200 OK (text/css)
2494	16.323979	157.140.2.32	172.17.20.101	HTTP	256	HTTP/1.1 200 OK (text/css)
2563	16.602845	157.140.2.32	172.17.20.101	HTTP	1166	HTTP/1.1 200 OK (application/javascript)
2583	16.772242	157.140.2.32	172.17.20.101	HTTP	184	HTTP/1.1 200 OK (application/javascript)

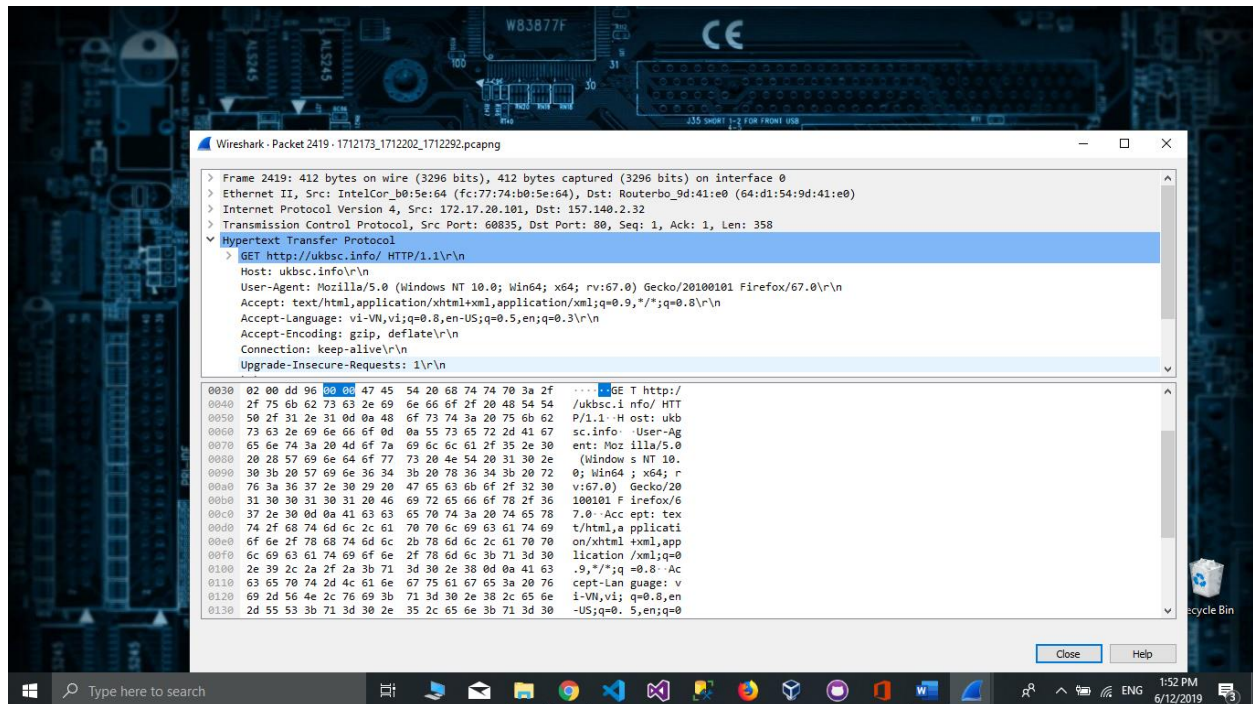
> Frame 2419: 412 bytes on wire (3296 bits), 412 bytes captured (3296 bits) on interface 0
 > Ethernet II, Src: IntelCor_b0:5e:64 (fc:77:74:b0:5e:64), Dst: Routerbo_9d:41:e0 (64:d1:54:9d:41:e0)
 > Internet Protocol Version 4, Src: 172.17.20.101, Dst: 157.140.2.32
 > Transmission Control Protocol, Src Port: 60835, Dst Port: 80, Seq: 1, Ack: 1, Len: 358
 > Hypertext Transfer Protocol

00f0 6c 69 63 61 74 69 6f 6e 2f 78 6d 6c 3b 71 3d 30 lication /xml;q=0
 0100 2e 39 2c 2a 2f 2a 3b 71 3d 30 2e 38 0d 0a 41 63 .9,/*;q =0.8:Ac
 0110 63 65 70 74 2d 4c 61 6e 67 75 61 67 65 3a 20 76 cept-lan guage: v
 0120 69 2d 56 4e 2c 76 69 3b 71 3d 30 2e 38 2c 65 6e i-VN;vi; q=0.0,en
 0130 2d 55 53 3b 71 3d 30 2e 35 2c 65 6e 3b 71 3d 30 -US;q=0.5,en;q=0
 0140 2e 33 0d 0a 41 63 63 65 70 74 2d 45 6e 63 6f 64 .3-Ac ce pt-Encod
 0150 69 6e 67 3a 20 67 7a 69 70 2c 20 64 65 66 6c 61 ling: gzi p, defla
 0160 74 65 0d 0a 43 6f 6e 6e 65 63 74 69 6f 6e 3a 20 te-Conn ection:
 0170 6b 65 65 70 2d 61 6c 69 76 65 0d 0a 55 70 67 72 keep-all ve- Upgr
 0180 61 64 65 2d 49 6e 73 65 63 75 72 65 2d 52 65 71 ade-Inse cure-Req
 0190 75 65 73 74 73 3a 20 31 0d 0a 0d 0a uests: 1 ----

_Cụ thể ta sẽ thấy được địa chỉ IP nguồn và đích trong quá trình trao đổi cũng như các port của Client (60835) và port mặc định của WEB Server (80).

_Quá trình trao đổi dữ liệu:

- Khi truy cập web thông qua web browser, sẽ gửi đến Web Client 1 gói truy vấn (request) bao gồm các thông tin như GET, giao thức HTTP, Host name, User-agent,...



- Các gói tin truy vấn sẽ được Web Client gửi tới Proxy Server.
- Tại Proxy Server sẽ tách lọc thông tin các gói tin nhận được như lấy domain name từ header host. Sau đó thực hiện lấy IP từ domain name. Kiểm tra tính hợp lệ của domain name (ngăn chặn nếu có truy cập trái phép).
- Proxy Server kết nối với Web Server thông qua port server 80 và địa chỉ IP vừa lấy được. Sau đó, gửi gói tin mà nó đã nhận được tới Web Server.
- Web Server sau khi nhận được thông tin từ Proxy Server sẽ tiến hành trả lại các gói tin response cho Proxy (bao gồm các nội dung trang web như: văn bản, hình ảnh,...nên kích thước các gói tin response khá lớn và có thể có nhiều gói được gửi).

The screenshot displays a Wireshark capture of network traffic. The top pane shows a list of captured packets, with several HTTP GET requests highlighted. The middle pane shows the details of the selected packet, including the HTTP request structure. The bottom pane shows the raw bytes of the packet.

- Proxy Server sẽ tiến hành nhận các gói tin từ Server Web nếu vẫn còn nhận được. Đồng thời gửi gói response đó về cho Web Client.
- Web Client sẽ nhận được response từ Proxy Server và trả kết quả về cho người dùng cuối.
- Nếu có đồng thời nhiều request gửi tới, cơ chế đa luồng sẽ phân luồng để xử lý từng request với chu trình như trên và trả kết quả về cho người dùng.

6. Sự cần thiết của Proxy Server trong thực tế:

Proxy Server đóng vai trò như cầu nối trong việc tiếp nhận và truyền tải cũng như trả về thông tin cho người dùng giữa Client và Server. Chúng nhận được yêu cầu từ Client và quyết định có chấp nhận thực hiện hay không, nếu chấp nhận thì tiếp đến Server và trả kết quả cuối cùng về. Vì vậy Proxy Server có vai trò quan trọng. Cụ thể:

- Quản lý được tất cả các thông tin truy xuất thông qua nó, ngăn chặn những truy xuất trái phép và nguy hiểm.
- Người dùng có thể truy cập Internet thông qua hệ thống riêng (Proxy Server) mà không phải các gói tin được truyền trực tiếp giữa hệ thống riêng và Internet. Điều này đảm bảo về tính an toàn và bảo mật các thông tin cá nhân riêng.
- Cơ chế caching lưu trữ lại nội dung các website đã được truy cập hoặc truy cập nhiều trong bộ nhớ đệm (cache). Từ đó việc truy xuất thông tin sẽ diễn ra nhanh hơn và thuận tiện hơn.

- Trong trường hợp băng thông dễ bị nghẽn do lượng truy cập lớn, Proxy Server có thể khắc phục được điều này, việc sử dụng băng thông sẽ hiệu quả hơn do có thể quản lý được các hoạt động của người dùng.
- Proxy Server hiểu được các giao thức cơ bản vì vậy, sử dụng nó sẽ có lợi trong việc logging.
- Đảm bảo an toàn cho hệ thống cục bộ nhờ việc sử dụng địa chỉ ẩn danh.

III. ĐÁNH GIÁ MỨC ĐỘ HOÀN THÀNH ĐỒ ÁN:

Cơ bản nhóm đã hoàn thành được các yêu cầu của đồ án đưa ra, tuy nhiên vẫn có nội dung chưa thực hiện được. Cụ thể:

<i>STT</i>	<i>Yêu cầu đồ án</i>	<i>Thực hiện được</i>	<i>Chưa thực hiện được</i>
1	Hỗ trợ HTTP 1.0 và 1.1	X	
2	Cho phép truy cập website thông qua Proxy Server.	X	
3	Hỗ trợ giao thức HTTP.	X	
4	Proxy Server chạy trên port 8888.	X	
5	Xử lý đồng thời được các request từ Client.	X	
6	Cấu hình file blacklist.conf và ngăn chặn các truy cập trái phép (khi tên miền có trong file trả về kết quả 403 FORBIDDEN).	X	
7	Thực hiện caching lại nội dung website cho các lần truy cập tới.		X

_Về cơ chế caching: nhóm đã suy nghĩ ra 1 cơ chế caching cho chương trình.

- Khi nhận được request từ Client, lấy tự domain name từ Host, sẽ tiến hành chuẩn hóa domain name thành 1 tên file có dạng domainname.txt.
- Kiểm tra file có tồn tại hay không? Nếu có nghĩa là lần truy cập vào tên miền lần trước đã có caching lại nội dung của website ứng với tên tiền và được lưu trong file. Đọc nội dung trong file ra vào 1 BUFFER và gửi nội dung đó về cho người dùng (đọc cho đến khi hết file).
- Ngược lại, nếu chưa tồn tại tên file thì tiến hành tạo 1 file với tên như đã chuẩn hóa.

- Khi Proxy Server nhận response từ Web Server và trả về cho Client sẽ tiến hành mở file và ghi nội dung gói tin response vào file như bộ nhớ cache và hoàn thành công việc caching.
 --Tuy nhiên, với ý tưởng này nhóm vẫn chưa thể thực hiện được (còn nhiều lỗi chưa được xử lý).
 _Đánh giá trên tổng thể án nhóm đã hoàn thành được 6/7 yêu cầu.
 _Đánh giá theo tỷ lệ phần trăm: 70-80 %.

IV. NGUỒN THAM KHẢO:

_Slide bài và tài liệu tham khảo, chương trình demo do giảng viên cung cấp.

<https://www.geeksforgeeks.org/socket-programming-cc/>

<https://developer.mozilla.org/en-US/docs/Web/HTTP/Overview>

<https://github.com/sameer2800/HTTP-PROXY>

<https://developer.mozilla.org/en-US/docs/Web/HTTP/Status/403>

<https://httpstatuses.com/403>

<https://docs.microsoft.com/en-us/windows/desktop/winsock/windows-sockets-error-codes-2>

<https://tech.vccloud.vn/proxy-la-gi-20181029114707565.htm>

<http://servergiarenhat.com/proxy-server-la-gi-loi-ich-khi-su-dung-proxy-server/>