**ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN**

**KHOA CÔNG NGHỆ THÔNG TIN**

# **BÁO CÁO ĐỒ ÁN SỐ 01**

# **LẬP TRÌNH ỨNG DỤNG MẠNG BẰNG SOCKET**

## MẠNG MÁY TÍNH

## **GVHD:**

## **Lớp: 19CTT1**

## **Họ và tên những người thực hiện:**

* **BÙI LÊ TUẤN ANH**
* **NGÔ NHẬT TÂN**

***Thành phố Hồ Chí Minh, tháng 12 năm 2020***

**MỤC LỤC**

[**LỜI CAM KẾT** i](#_Toc55721589)

[**PHÂN CÔNG CÔNG VIỆC CHUNG** 1](#_Toc55721590)

[**GIỚI THIỆU ĐỒ ÁN** 3](#_Toc55721591)

[**HƯỚNG TIẾP CẬN VÀ CÁC BƯỚC THỰC HIỆN CHI TIẾT** 5](#_Toc55721592)

[**KIỂM THỬ ỨNG DỤNG VÀ ĐÁNH GIÁ CHUNG** 9](#_Toc55721593)

[**KẾT LUẬN** 15](#_Toc55721594)

[**TÀI LIỆU THAM KHẢO** 17](#_Toc55721595)

# **LỜI CAM KẾT**

Chúng tôi, Bùi Lê Tuấn Anh và Ngô Nhật Tân, cam kết:

* ***Mọi thành phần của đồ án này, bao gồm mã nguồn và cả báo cáo này, đều là sản phẩm do chúng tôi thực hiện.***
* ***Mọi tư liệu, mã nguồn tham khảo và các thành phần được lấy từ người khác sẽ được ghi nhận, tiếp thu ở mức độ phù hợp để đảm bảo quyền tác giả và không vi phạm các quy định của pháp luật cũng như môn học.***
* ***Mọi hành vi sao chép, làm giả hoặc chỉnh sửa mà không có sự cho phép theo quy định, chúng tôi sẽ chịu hoàn toàn trách nhiệm và sẽ chịu những hình thức xử lý của pháp luật, bộ môn, khoa và nhà trường.***

## **PHÂN CÔNG CÔNG VIỆC CHUNG**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **Họ tên** | **MSSV** | **Công việc** | **Hoàn tất** |
| ***1*** | ***Bùi Lê Tuấn Anh*** | ***19120163*** | * *Viết báo cáo* * *Sử dụng Github để theo dõi và chỉnh sửa các hàm* |  |
| ***2*** | ***Ngô Nhật Tân*** | ***19120128*** | * *Viết báo cáo* * *Sử dụng Github để theo dõi và chỉnh sửa các hàm* |  |
| **Cả nhóm** | | | * *Chỉnh sửa báo cáo và các tập tin mã nguồn.* |  |

**TRANG TRỐNG – BLANK PAGE**

## **GIỚI THIỆU ĐỒ ÁN**

*Đây là đồ án thứ nhất của học phần* ***“Mạng máy tính”****, giảng dạy tại* ***Khoa Công nghệ thông tin, Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Thành phố Hồ Chí Minh****. Đồ án này được thực hiện bởi hai thành viên:*

1. ***Bùi Lê Tuấn Anh – Mã số sinh viên: 19120163***
2. ***Ngô Nhật Tân – Mã số sinh viên: 19120128***

*Nội dung chính của bản báo cáo này tập trung phân tích những khía cạnh căn bản về lập trình ứng dụng mạng bằng* ***Socket*** *– một yếu tố vô cùng quan trọng trong giao thức* ***TCP/IP*** *giúp thiết lập và đảm bảo an toàn cho kết nối Internet của người dùng, với mô hình cơ bản nhất là máy chủ trung gian* ***(proxy)****. Thông qua đó, nhóm sẽ làm rõ hơn những mặt tích cực và hạn chế của kết nối thông qua giao thức* ***TCP/IP*** *sử dụng* ***socket*** *hoạt động trên máy chủ* ***proxy*** *được nhóm tự xây dựng và hoàn thiện.*

*Trong quá trình thực hiện đồ án này, nhóm chúng tôi không tránh khỏi những thiếu sót (do kiến thức còn hạn chế), nhưng chúng tôi đã và đang cố gắng hết sức để hoàn thiện nó tốt nhất trong khả năng của mình. Mong nhận được sự thông cảm từ mọi người.*

*Chúng tôi cũng xin chân thành cảm ơn sự hỗ trợ tận tình từ các giảng viên* ***Khoa Công nghệ Thông tin*** *nói chung và bộ môn* ***Mạng máy tính*** *nói riêng. Chúng tôi cũng xin gửi lời cảm ơn và ghi nhận đối với những lập trình viên trên khắp thế giới đã không ngừng nghỉ làm việc để mang đến những chương trình hữu ích cho người sử dụng mà ở đó chúng tôi tham khảo và sử dụng mã nguồn một cách thận trọng nhất và tôn trọng nguyên tác nhất.*

*Toàn bộ mã nguồn (và cả báo cáo này) cũng sẽ được đăng tải trên* ***Github*** *tại* ***repository*** *(hiện tại đang ở chế độ riêng tư) mang tên* ***socketProject****. Các giảng viên có thể cung cấp địa chỉ email* ***đã đăng ký tài khoản Github*** *để nhóm cung cấp quyền truy cập, đảm bảo tính riêng tư cho đồ án này. Vui lòng gửi thông tin qua địa chỉ* [*19120163@student.hcmus.edu.vn*](mailto:19120163@student.hcmus.edu.vn)*. Sau đó các giảng viên có thể nhân bản (folk) repository này về và nghiên cứu thêm* ***(toàn bộ ở nhánh Master)****. Xin trân trọng cảm ơn.*

**TRANG TRỐNG – BLANK PAGE**

## **HƯỚNG TIẾP CẬN VÀ CÁC BƯỚC THỰC HIỆN CHI TIẾT**

1. **Thông tin về các cấu hình cho chương trình**:

|  |  |
| --- | --- |
| ***Chương trình lập trình*** | ***Trình duyệt*** |
| **Visual Studio Code** | Các trình duyệt thông dụng như ***Mozilla Firefox, Google Chrome, Microsoft Edge, Cốc Cốc…*** |
| ***Ngôn ngữ lập trình*** |
| **Python 3.9** |
| **Hệ điều hành máy tính** |
| **Windows 10**. |

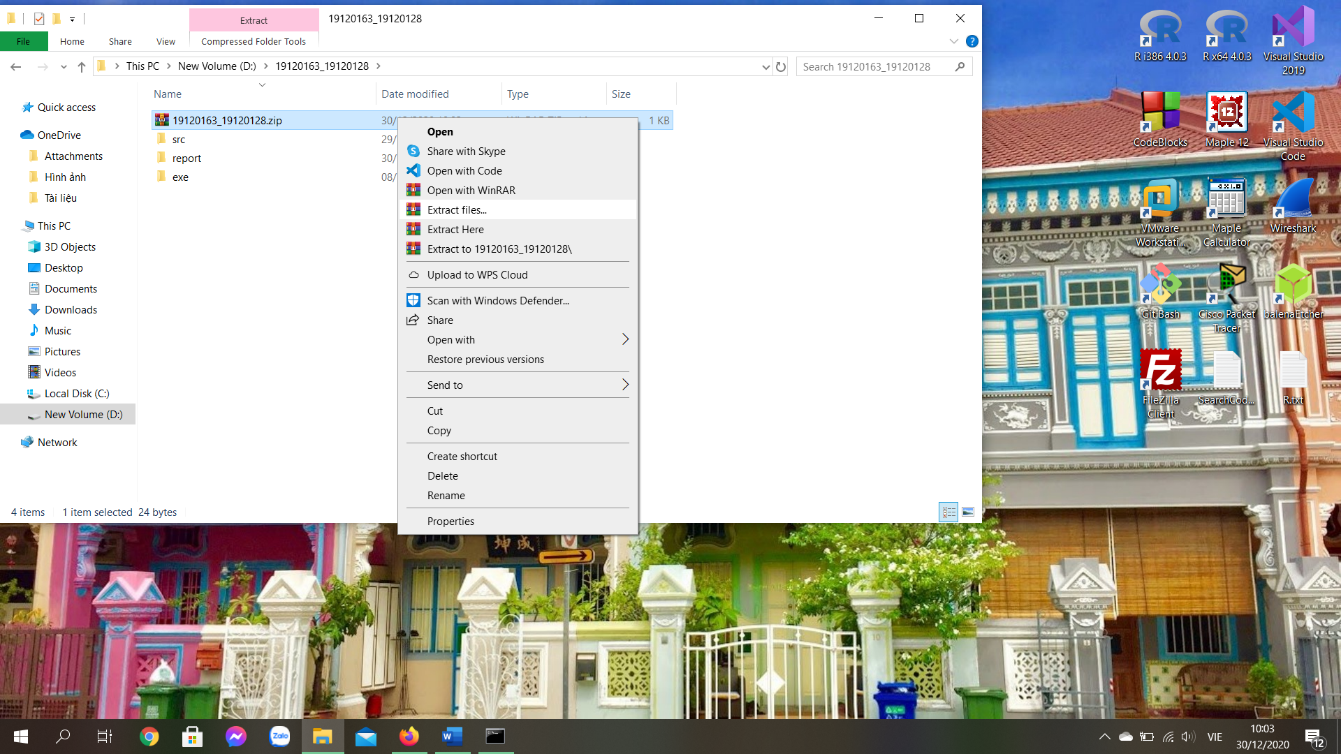
1. **Các bước thực hiện:**
   1. *Hàm xử lý file chặn các trang web*
      1. **Tham số truyền vào cho hàm**: Gồm 3 tham số
         1. ***cli\_socket***: ***Socket được khởi tạo từ máy khách***
         2. ***cli\_addr***: ***Địa chỉ của máy khách***, là một tuple (tương tự một vector hai chiều), có hai giá trị là **Địa chỉ IP** máy khách (“localhost”) và **Cổng truy cập** (mặc định là 8888)
         3. ***in4*:** Chứa ***toàn bộ thông tin*** về ***địa chỉ truy cập***, thời gian truy cập, ghi nhận việc lưu trữ dữ liệu (caching), thông tin của máy khách, ***phương thức truy cập (GET/POST)***…
      2. **Chức năng hàm**: *Giúp chặn các trang web nguy hại được lưu trữ trong file* ***“blacklist.conf”****, không cho người dùng truy cập vào để đảm bảo an toàn khi truy cập mạng Internet.*
   2. *Cụm hàm đọc và xử lý dữ liệu máy khách (Caching)* ***(bổ trợ cho các hàm chinh, không đi sâu phân tích, chỉ nêu chức năng)***
      1. Hàm: khóa địa chỉ tập tin dữ liệu.
      2. Hàm: mở khoa địa chỉ tập tin dữ liệu.
      3. Hàm: ghi dữ liệu vào log.
      4. Hàm: kiểm tra điều kiện có tiến hành lưu trữ dữ liệu (caching) hay không.
      5. Hàm: kiểm tra xem trang web đã được lưu trữ dữ liệu (caching) trước đó hay chưa.
      6. Hàm: lấy toàn bộ thông tin trong cache ra ngoài để sử dụng.
      7. Hàm: xử lý khi bộ nhớ lưu trữ bị đầy, xóa phần tử cache được sử dụng gần nhất.
      8. Hàm: thêm vào phần đầu gói tin các thông tin quan trọng.
      9. Hàm
         1. **Tham số truyền vào cho hàm**: Gồm hai tham số
            1. ***cli\_addr***: ***Địa chỉ của máy khách***, là một tuple (tương tự một vector hai chiều), có hai giá trị là **Địa chỉ IP** máy khách (“localhost”) và **Cổng truy cập** (mặc định là 8888)
            2. ***cli\_data***: ***Dữ liệu caching của máy khách***
         2. **Chức năng hàm**: Hỗ trợ đọc ***toàn bộ dữ liệu của máy khách***, bao gồm ***cổng và địa chỉ máy chủ***, ***dữ liệu từ máy khách***, ***phương thức kết nối (HTTP)***, ***truyền dữ liệu (GET/POST) …***
   3. *Cụm hàm GET, POST*
      1. **Hàm GET**: 
      2. **Hàm POST**: 
      3. **Tham số truyền vào cho cả hai hàm**: Gồm 3 tham số
         1. ***cli\_socket***: ***Socket được khởi tạo từ máy khách***
         2. ***cli\_addr***: ***Địa chỉ của máy khách***, là một tuple (tương tự một vector hai chiều), có hai giá trị là **Địa chỉ IP** máy khách (“localhost”) và **Cổng truy cập** (mặc định là 8888)
         3. ***in4***: Chứa ***toàn bộ thông tin*** về ***địa chỉ truy cập***, thời gian truy cập, ghi nhận việc lưu trữ dữ liệu (caching), thông tin của máy khách, ***phương thức truy cập (GET/POST)***…
      4. **Chức năng hai hàm**:
         1. **Hàm GET** được dùng để *gửi yêu cầu lấy dữ liệu từ nguồn tài nguyên xác định*, ở đây là ***lấy toàn bộ thông tin trang web từ máy chủ***.
         2. **Hàm POST** được dùng để *gửi dữ liệu tới máy chủ* nhằm ***khởi tạo hoặc cập nhật nguồn tài nguyên***.
   4. *Cụm hàm cấu hình máy chủ proxy*
      1. *****Hàm khởi động chương trình***:
         1. **Tham số truyền vào cho hàm**: **Không có tham số**
         2. **Chức năng hàm**: *Khởi tạo socket, lắng nghe và chấp nhận kết nối từ người dùng. Khởi động máy chủ Proxy*.
      2. ***Hàm xử lý phương thức truyền thông tin và dữ liệu người dùng***: 
         1. **Tham số truyền vào cho hàm**:
            1. ***cli\_socket***: ***Socket được khởi tạo từ máy khách***
            2. ***cli\_addr***: ***Địa chỉ của máy khách***, là một tuple (tương tự một vector hai chiều), có hai giá trị là **Địa chỉ IP** máy khách (“localhost”) và **Cổng truy cập** (mặc định là 8888)
            3. ***cli\_data***: ***Dữ liệu caching của máy khách***
         2. **Chức năng hàm**: *Giúp định hướng kết nối, lựa chọn phương thức truyền dữ liệu và quản lý kết nối Proxy* (kiểm tra trang web đang truy cập có nằm trong **“danh sách đen”** hay không, có tiến hành lưu trữ dữ liệu trung gian **(caching)** hay không, hướng truy cập là **gửi thông tin lên** hay **lấy thông tin từ máy chủ trang web về**…)

**TRANG TRỐNG – BLANK PAGE**

## **KIỂM THỬ ỨNG DỤNG VÀ ĐÁNH GIÁ CHUNG**

1. **Hướng dẫn sử dụng**

* Giải nén tập tin bằng lệnh **Extract files..**., chọn thư mục tương ứng rồi bấm **Extract/OK**

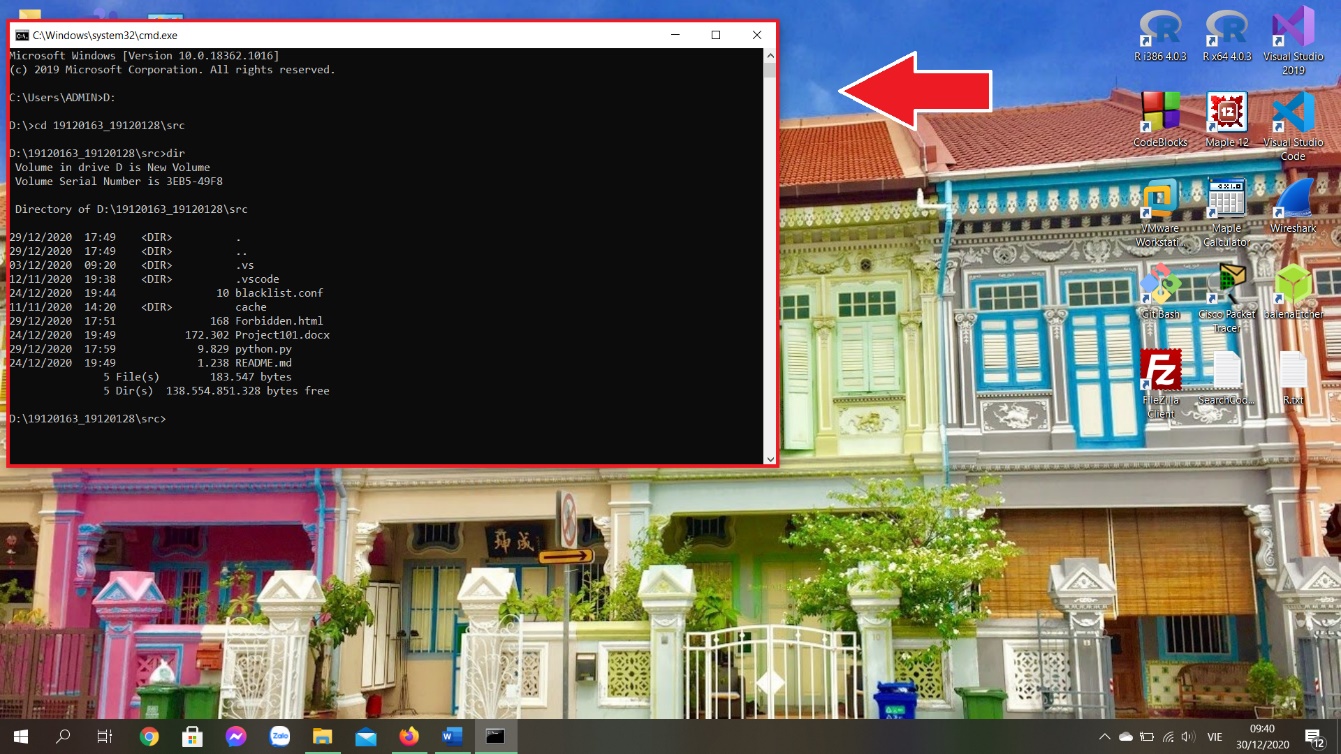
****

* Mở ứng dụng Command Prompt (Bằng cách gõ vào tổ hợp phím **Windows + R**, gõ “***cmd”*** và bấm **Enter** – hoặc bấm **Start** và tìm **Command Prompt**)

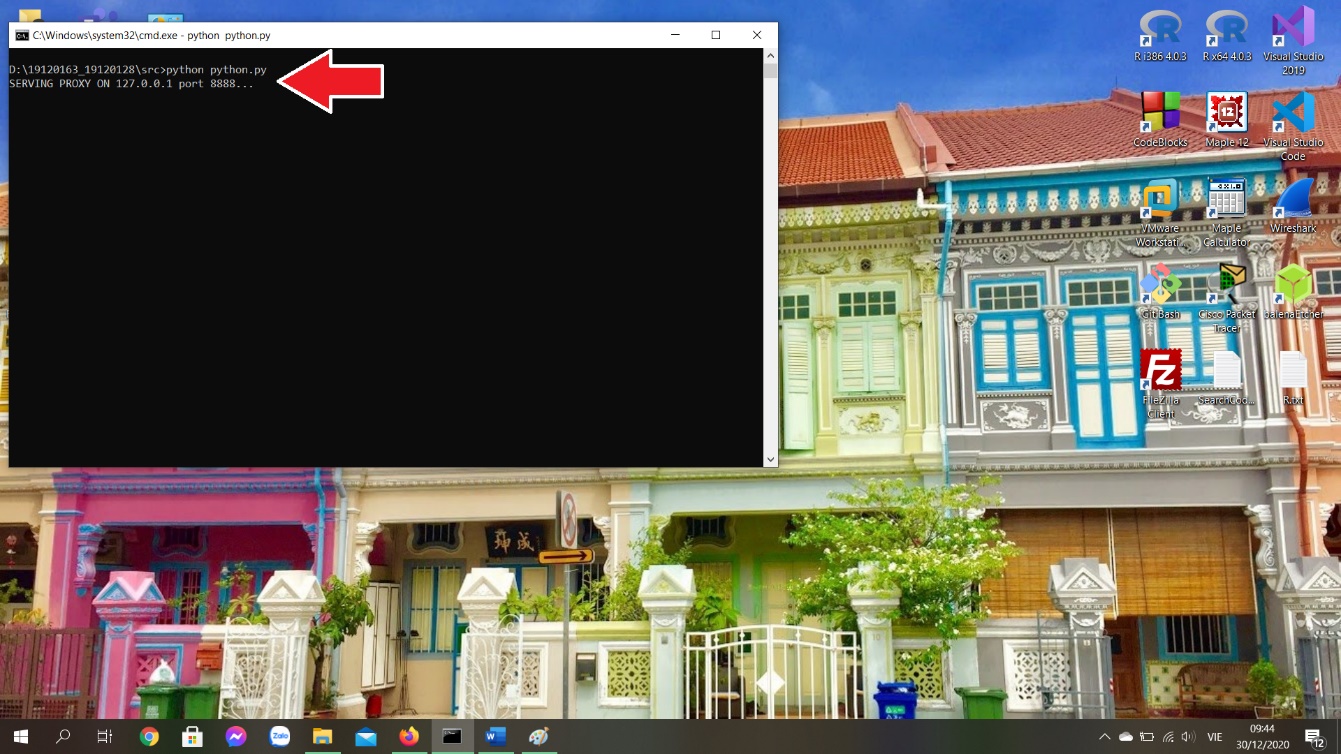
**Ảnh có chứa văn bản, tòa nhà

Mô tả được tạo tự động**

* Trỏ về thư mục chứa tập tin đã giải nén, xuất thông tin của thư mục bằng lệnh ***“dir”*** (hiện tại trên màn hình, do tất cả các tập tin đang cùng nằm trong 1 thư mục đưa lên **git/Github** nên bao gồm cả tập tin báo cáo và các thư mục khác):

****

* Gõ lệnh “***python python.py***”. Xuất hiện dòng lệnh “**SERVING 127.0.0.1 ON PORT 8888…**”

****

* Truy cập vào Mozilla Firefox, Google Chrome…, mở **Cài đặt (Settings)**, tìm thông tin về cài đặt Proxy, thay đổi thông tin như sau:
  + Địa chỉ máy chủ/Server Address: ***localhost*** hoặc ***127.0.0.1***
  + Cổng/Port: **8888**
    - Bấm **Lưu/Save** hoặc **OK**
* **Bước này có thể thực hiện đầu tiên, trước hoặc sau khi khởi động Proxy đều được. Hình ảnh này được chụp lại trước khi khởi động Proxy.**

**Ảnh có chứa văn bản

Mô tả được tạo tự động**

* Sau đó truy cập vào các trang web HTTP như trên màn hình. Sẽ có báo lỗi nếu web không truy cập được. Dưới đây là một trang web truy cập thành công:

**Ảnh có chứa văn bản, ảnh chụp màn hình, máy tính xách tay

Mô tả được tạo tự động**

1. **Đánh giá ứng dụng**
   1. Quá trình bắt gói tin bằng Wireshark
   2. Đánh giá chung về chương trình

***Chú thích:*** *Một chức năng được tính là* ***ĐẠT*** *nếu hoàn thành từ 50-65% yêu cầu trở lên. Mức độ* ***CHƯA ĐẠT*** *chỉ hoàn thành dưới 50% yêu cầu. Đánh giá chung là trung bình tỉ lệ của 5 chức năng (nếu chức năng chưa đạt thì tính chênh lệch so với mức 50%).*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **Chức năng** | **Nhận xét** | **Mức độ (%)** | |
| **Đạt** | **Chưa đạt** |
| 1 | Proxy Server cho phép Client truy cập website thông qua các method: GET, POST | ***Hoàn thành một phần, vẫn còn một số lỗi chưa khắc phục*** | **70%** |  |
| 2 | Proxy Server chỉ hỗ trợ HTTP, không cần hỗ trợ HTTPS | ***Đã cấp quyền hỗ trợ HTTP, đã truy cập được*** | **100%** |  |
| 3 | Proxy Server phải xử lý đồng thời được các request từ client | ***Hoàn thành ở mức chấp nhận được*** | **80%** |  |
| 4 | Proxy Server sẽ lắng nghe tại port 8888, chờ kết nối từ client | ***Đã hoàn thành việc cấu hình cổng cho máy chủ proxy, đảm bảo việc kết nối đúng cổng*** | **100%** |  |
| 5 | 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 | ***Hoàn thành việc chặn các trang web bằng mã lỗi 403*** | **100%** | **40%** |
| **ĐÁNH GIÁ CHUNG** | | | **90%** |  |

**TRANG TRỐNG – BLANK PAGE**

## **KẾT LUẬN**

Proxy mang lại nhiều tính năng quan trọng cho việc sử dụng mạng, bao gồm:

* ***Tường lửa và bộ lọc***: Chức năng này là do Proxy làm việc ở tầng Ứng dụng (Application), khi được cấu hình độc lập sẽ giúp thông tin cá nhân và lịch sử duyệt web được bảo mật. Đồng thời, với việc sử dụng Proxy có thể thay đổi địa chỉ IP, ta có thể truy nhập vào các nội dung bị chặn hoặc chặn các nội dung không mong muốn. Điều này đồng nghĩa với việc ta có thể kiểm soát việc sử dụng Internet của tất cả mọi người.
* ***Chia sẻ kết nối***: Proxy đóng vai trò là máy chủ trung gian, có thể gom kết nối từ các máy khác trong cùng một công ty hay một doanh nghiệp để mở rộng kết nối một cách hiệu quả mà tiết kiệm được tài nguyên.
* ***Lưu trữ thông tin trang web (Caching)***: Việc proxy đóng vai trò là 1 máy chủ trung gian lưu trữ dữ liệu trang web giúp tăng tốc độ truy cập và tiết kiệm băng thông tốt nhất, không cần phải truy cập thẳng tới máy chủ gốc.

**TRANG TRỐNG – BLANK PAGE**

## **TÀI LIỆU THAM KHẢO**