

**ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG-HCM**  
**KHOA CÔNG NGHỆ THÔNG TIN**



**REPORT**  
**SOCKET**

**ĐỀ 2: CLIENT**

**MÔN: MẠNG MÁY TÍNH**

21120501 - Nguyễn Ngọc Gia Minh

2112016 - Nguyễn Thị Kiều Ngân

21120516 - Võ Bá Hoàng Nhất

**Lớp: 21CTT5**

# Mục lục

<b>I</b>	<b>Thông tin chung</b>	<b>2</b>
1	Thông tin nhóm . . . . .	2
2	Yêu cầu đồ án . . . . .	2
3	Bảng phân công nhiệm vụ và đánh giá mức độ hoàn thành .	3
<b>II</b>	<b>Kịch bản giao tiếp chương trình</b>	<b>3</b>
1	Tổng quan . . . . .	3
2	Kịch bản giao tiếp . . . . .	3
2.1	Kết nối . . . . .	3
2.2	Quản lý kết nối . . . . .	4
2.3	Tải và lưu file dạng "Content-Length" . . . . .	5
2.4	Tải và lưu thành file dạng "Transfer-Encoding: chunked" . . . . .	5
2.5	Tải các file trong folder . . . . .	6
2.6	Khả năng kết nối tới nhiều server . . . . .	6
<b>III</b>	<b>Hướng dẫn sử dụng chương trình</b>	<b>6</b>
<b>IV</b>	<b>Tài liệu tham khảo</b>	<b>8</b>

# I Thông tin chung

## 1 Thông tin nhóm

STT	MSSV	Họ tên
1	21120501	Nguyễn Ngọc Gia Minh
2	21120506	Nguyễn Thị Kiều Ngân
3	21120516	Võ Bá Hoàng Nhất

## 2 Yêu cầu đồ án

- Mục tiêu là viết một web client đơn giản. Web client sẽ gửi request đến web server, port 80 để tải nội dung của page và lưu vào file.

STT	Chức năng	Ý nghĩa
1	KẾT NỐI	Cho phép client kết nối đến server thông qua kết nối TCP
2	QUẢN LÝ KẾT NỐI	Khi client hoặc server mất kết nối đột ngột, không làm chương trình treo hay xảy ra lỗi
3	Tải và lưu thành file dạng "Content-Length"	Download file thành công
4	Tải và lưu thành file dạng "Transfer-Encoding: chunked"	Download file thành công
5	Tải các file trong folder	Lưu tất cả các file trong folder
6	Multiple requests	gửi nhiều requests trong một connection khi download file trong folder
7	Multiple connection	Concurrent, handle nhiều kết nối cùng lúc đến các web servers

### 3 Bảng phân công nhiệm vụ và đánh giá mức độ hoàn thành

STT	Người thực hiện	Nội dung	Mức độ hoàn thành
1	Nguyễn Ngọc Gia Minh	Tải và lưu thành file dạng "Transfer-Encoding: chunked" Tải các file trong folder	100% 100%
2	Nguyễn Thị Kiều Ngân	Multiple requests Multiple connection	100% 100%
3	Võ Bá Hoàng Nhất	Kết nối Quản lý kết nối Tải và lưu thành file dạng "Content-Length"	100% 100% 100%

## II Kịch bản giao tiếp chương trình

### 1 Tổng quan

- Xác định kiến trúc mạng: Client – Server
- Giao thức sử dụng tại tầng Transport: TCP
- Môi trường lập trình: Visual Studio Code
- Ngôn ngữ lập trình: Python
- Module sử dụng: threading, sys

### 2 Kịch bản giao tiếp

#### 2.1 Kết nối

- Client tạo một TCP socket, kết nối đến port 80 của web server. Sau đó gửi một yêu cầu tải file.
- Thủ tục bắt tay (handshaking):
  - + Client gửi 1 request yêu cầu kết nối dạng:

---

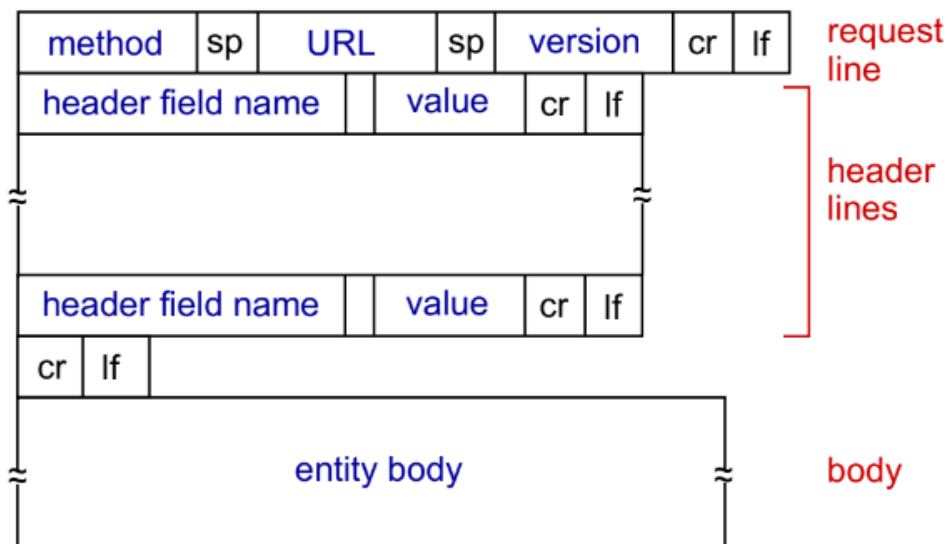
```
"GET /" + path + " HTTP/1.1\r\nHost: " +
    name_host + "\r\nConnection:
    keep-alive\r\n\r\n"
```

---

tới webserver trong đó Path là đường dẫn chứa file cần download, Host là tên máy chủ webserver.

Trong đó GET yêu cầu đối tượng được chỉ định trong đường dẫn (path) dẫn đến đối tượng, HTTP/1.1 chỉ định phiên bản giao thức HTTP đang được sử dụng là 1.1, Connection: keep-alive yêu cầu vẫn giữ kết nối sau khi server gửi xong response.

+ Server phản hồi lại với gói tin có dạng:



+ Lúc này đã hình thành kết nối TCP giữa client và server.

- Client sẽ tự động đóng kết nối sau khi ghi xong file data.

## 2.2 Quản lý kết nối

- Trong trường hợp thông điệp phản hồi không là 200 OK thì thông báo lỗi rồi đóng kết nối và kết thúc chương trình.
- Trong trường hợp client bị mất kết nối giữa chừng khi đang tải dữ liệu: Ở mỗi hàm nhận dữ liệu từ server thì có cài thời gian tối đa để tải dữ

liệu xuống. Khi web Server hay máy tính bị mất kết nối đột ngột, Vì mất kết nối nên quá trình tải dữ liệu xuống vượt quá thời gian nên thoát khỏi chương trình luôn mà không làm treo chương trình.

- Trong trường hợp đối tượng được yêu cầu đã bị chuyển đi nơi khác, thì lấy URL mới từ thông điệp phản hồi sau đó điều hướng truy cập lại vào địa chỉ mới này
- Trong trường hợp các lỗi do server như 404 hay 505 thì chỉ in ra lỗi là không tìm thấy hoặc là không hỗ trợ sau đó đóng kết nối và thoát chương trình.

### **2.3 Tải và lưu file dạng "Content-Length"**

- Với Content-Length: Ở phần header tồn tại một giá trị là kích thước phần body (content-length). Để tải dữ liệu xuống ở dạng này ta chỉ cần đọc dữ liệu lần lượt cho đến khi lượng data nhận về đúng bằng giá trị content-length thì kết thúc.

### **2.4 Tải và lưu thành file dạng "Transfer-Encoding: chunked"**

- Với Transfer- Encoding: chunked, luồng dữ liệu được chia thành các khối không chồng chéo lên nhau.

Các khối này sẽ được gửi và nhận độc lập. Trước mỗi khối dữ liệu là kích thước của nó(được tính bằng bytes). Quá trình gửi sẽ kết thúc khi receiver nhận được khối dữ liệu có kích thước bằng 0.

Kích thước của mỗi khối dữ liệu được truyền dưới dạng thập lục phân, theo sau là `\r\n` đóng vai trò phân cách dòng, tiếp sau là khối dữ liệu với kích thước trên.

Cách giải quyết yêu cầu đề bài :

Kiểm tra “chunked” có hiện diện trong Transfer- Encoding header?

- + nếu có -> ta sẽ tải file theo kiểu mã hóa truyền dữ liệu theo khối (Transfer - Encoding: chunked)

Ta sẽ đọc dữ liệu theo dòng :

+ Dòng đầu sẽ chứa thông tin kích thước của mỗi khối dữ liệu được truyền dưới dạng thập lục phân, theo sau là `\r\n`.

+ Chuyển về dạng thập phân và ta kiểm tra xem nó có khác 0 hay không

+ Nếu khác 0 tiến hành đọc khối dữ liệu

+ Sau mỗi khối dữ liệu với kích thước trên sẽ là `\r\n`

+ Lặp lại các bước trên đến khi kích thước của khối dữ liệu cần đọc là 0, và 1 dòng `\r\n` sau đó thì ta dừng vòng lặp.

## 2.5 Tải các file trong folder

- Khi download file trong subfolder, kiểm tra đường dẫn tới có chứa các file con hay không.
- Nếu đường dẫn đó chứa các file con thì client gửi tiếp 1 request tới webserver đó yêu cầu tải file con trong folder. Vì lúc này connection ban đầu vẫn chưa được đóng nên không cần phải khởi tạo lại connection mới mà chỉ cần gửi request ngay trên chính con đường này.
- Nếu đường dẫn đó không chứa các file con thì chỉ tải file.

## 2.6 Khả năng kết nối tới nhiều server

- Trong trường hợp kết nối tới nhiều server cùng lúc thì khởi tạo các thread và cho kết nối song song độc lập nhau.

# III Hướng dẫn sử dụng chương trình

- Mở cửa sổ Command Prompt trên hệ điều hành của bạn  
Hướng dẫn sử dụng chương trình trên hệ điều hành Windows:
- Truy cập tới thư mục chứa chương trình client của chúng tôi bằng lệnh  
cd + địa chỉ của thư mục
- Chạy chương trình của chúng tôi theo dạng như sau

---

```
client.exe <URLs of pages which need downloading>
```

---

Trong trường hợp chỉ cần tải 1 page hay 1 folder

Ví dụ:

```
client.exe http://www.fit.hcmus.edu.vn/
```

Trong trường hợp cần tải nhiều page hoặc folder cùng lúc

Ví dụ

```
client.exe http://www.fit.hcmus.edu.vn/ http://www.google.com
```

Lưu ý: Các URL sẽ cách nhau bởi khoảng trắng.

- Sau khi tải file hoặc folder từ bất kỳ 1 URL nào thì trên màn hình sẽ hiển thị thông báo "Download successfully"

```
C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.22000.1219]
(c) Microsoft Corporation. All rights reserved.

E:\nam hai\mmt\final_socket>python client.py http://example.com/
Download successfully

E:\nam hai\mmt\final_socket>
```

- Trong trường hợp trang web truy cập tới không được tìm thấy thì trên màn hình sẽ in ra là "Not Found"

```
C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.22000.1219]
(c) Microsoft Corporation. All rights reserved.

E:\nam hai\mmt\final_socket>python client.py http://example.com/abc
Not Found!!

E:\nam hai\mmt\final_socket>
```

- Trong trường hợp server không hỗ trợ phiên bản giao thức HTTP1.1 thì trên màn hình sẽ in ra là "Not support"
- Trong trường hợp trong lúc tải bị mất kết nối, sau khoảng 30s thì cứ mỗi file bị thiếu sẽ là 1 lần in dòng "A connection attempt failed!!!" từ



đó ta có thể biết được còn bao nhiêu file còn thiếu nữa và bị mất kết nối khi tải file bao nhiêu, còn file nào chưa được tải xuống.

[illegible]

## IV Tài liệu tham khảo

1. Giáo trình Mạng máy tính - ĐH KHTN TP.HCM, 2020, chương 2
2. Computer Networking: A Top-Down Approach 8th edition Jim Kurose, Keith Ross Pearson, 2020, chapter 2
3. Chunked transfer encoding, wikipedia, URL: [https://en.wikipedia.org/wiki/Chunked\\_transfer\\_encoding](https://en.wikipedia.org/wiki/Chunked_transfer_encoding)