



TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN TP.HCM
KHOA CÔNG NGHỆ THÔNG TIN
HỆ CHÍNH QUI – LỚP CỬ NHÂN TÀI NĂNG 2014
MÔN: **LẬP TRÌNH MẠNG**

BÁO CÁO BÀI TẬP CÁ NHÂN 01

Giáo viên hướng dẫn: **Nguyễn Thanh Quân**

Sinh viên thực hiện

Nguyễn Quốc Bảo 1412661

TP.HCM, ngày 08 tháng 03 năm 2017

1. Ý nghĩa các hàm quan trọng

1.1. `parseByRegex()`

Lấy ra các matching group từ chuỗi đưa vào bằng Regular Expression. Dữ liệu trả về là một danh sách liên kết đơn chứa các matching group (kể cả full matching group).

1.2. `checkLink()`

Gửi truy vấn HTTP/HEAD để nhận thông tin về đường link được nhập vào như HTTP code, Content Type, Content Length,...

1.3. `socket()`

Đăng kí mở một socket với hệ điều hành.

1.4. `connect()`

Khởi tạo kết nối socket với server bên ngoài

1.5. `send()`

Gửi dữ liệu thông qua socket đã đăng kí và khởi tạo

1.6. `recv()`

Nhận thông tin được server gửi trả về máy qua socket. `recv()` chủ yếu có hai chế độ nhận là blocking và non-blocking

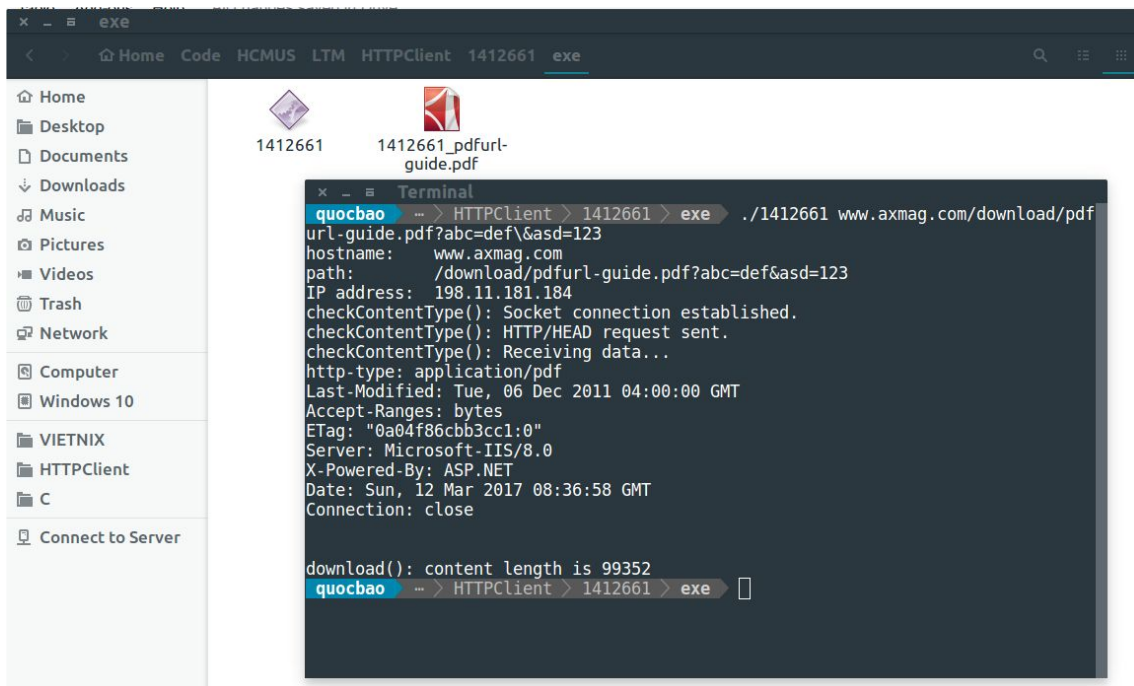
2. Cách chạy của chương trình

2.1. Tổng quan

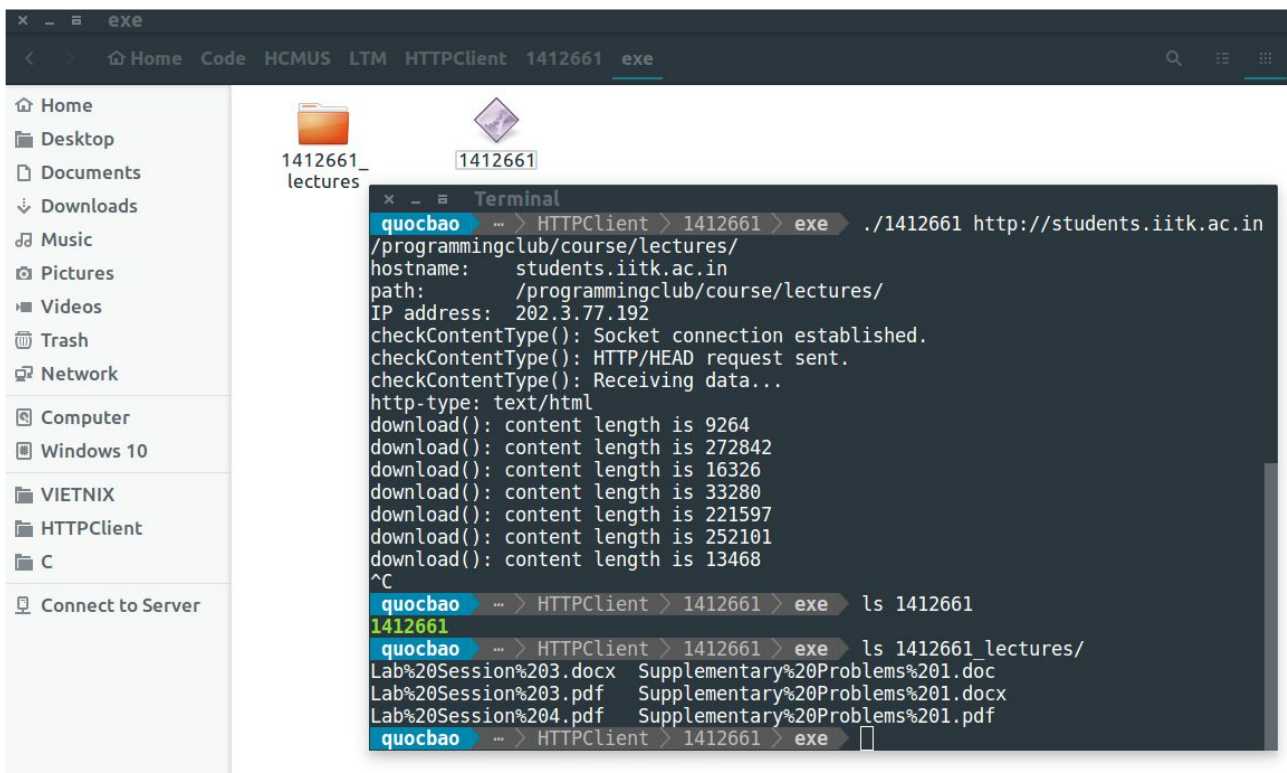
- **Bước 1:** Phân giải các thông tin như hostname, path, địa chỉ IP từ đường link nhập vào ban đầu
- **Bước 2:** Gửi truy vấn HTTP/HEAD để kiểm tra đường link nhập vào có hợp lệ hay không (có bị trả về code HTTP 4xx hay 5xx hay không?), Content Type là gì, Content Length là bao nhiêu,....
 - Nếu mã HTTP trả về là 4xx hoặc 5xx thì dừng chương trình.
- **Bước 3:**
 - Nếu Content Type trả về không phải là text/html thì tiến hành gửi truy vấn HTTP/GET để download file về.
 - Nếu Content Type là text/html thì tiến hành gửi truy vấn HTTP/GET để tải nội dung HTML về. Từ nội dung HTML trả về rút trích ra đường link các file và tải lần lượt từng file về.

2.2. Kết quả chạy chương trình

2.2.1. Download file đơn



2.2.2. Download nhiều file



3. Phân tích quá trình download một file

Xem file `capture.pcapng` đi kèm với báo cáo:

- Dòng 4,8,9 là quá trình bắt tay 3 bước của giao thức TCP để thiết lập kết nối
- Dòng 10 là truy vấn HTTP/HEAD được gửi lên server.
- Dòng 16, 17 là phản hồi của server cho truy vấn HTTP/HEAD
- Dòng 18-22 là truy vấn HTTP/GET gửi lên server để yêu cầu download file.
- Dòng 23-156 là các packet chứa file đang được tải về
- Dòng 157 là xác nhận hoàn tất phiên gửi dữ liệu
- Dòng 158-162 là thủ tục đóng kết nối TCP giữa server và client.

4. Ưu điểm của chương trình

- Được thực hiện bằng nhiều công cụ:
 - IDE: CodeBlocks
 - Memory tool: Valgrind (dùng để phát hiện lỗi memory leakage).
 - Debugger: gdb
- **Không có lỗi memory leakage**
- Quản lý bộ nhớ linh hoạt bằng việc cast qua lại giữa kiểu dữ liệu của pointer và số nguyên:
 - File `function.c` dòng 1 và 2:
 - `#pragma GCC diagnostic ignored "-Wpointer-to-int-cast"`
 - `#pragma GCC diagnostic ignored "-Wint-to-pointer-cast"`
- Sử dụng Regular Expression (POSIX Extended)
- Sử dụng danh sách liên kết đơn để quản lý bộ nhớ.
- Ít sử dụng và khai báo vùng nhớ stack (không dùng bộ nhớ stack để lưu các phần dữ liệu lớn như dữ liệu download về).

5. Nhược điểm của chương trình

- Không có comment (không kịp làm)
- Chưa sử dụng thread nên quá trình download rất chậm do phải diễn ra tuần tự.