BÀI TẬP THỰC HÀNH SỐ 2
MÔN HỌC: NHẬP MÔN MẠNG MÁY TÍNH

GIAO THỨC HTTP

Trong bài thực hành này, chúng ta sẽ khám phá một vài khía cạnh của giao thức HTTP: thông điệp GET/response, cấu trúc của HTTP header, truy cập các file HTML dài, truy cập các file HTML có đính kèm các đối tượng, xác thực HTTP và bảo mật.

1 HTTP GET/response cơ bản

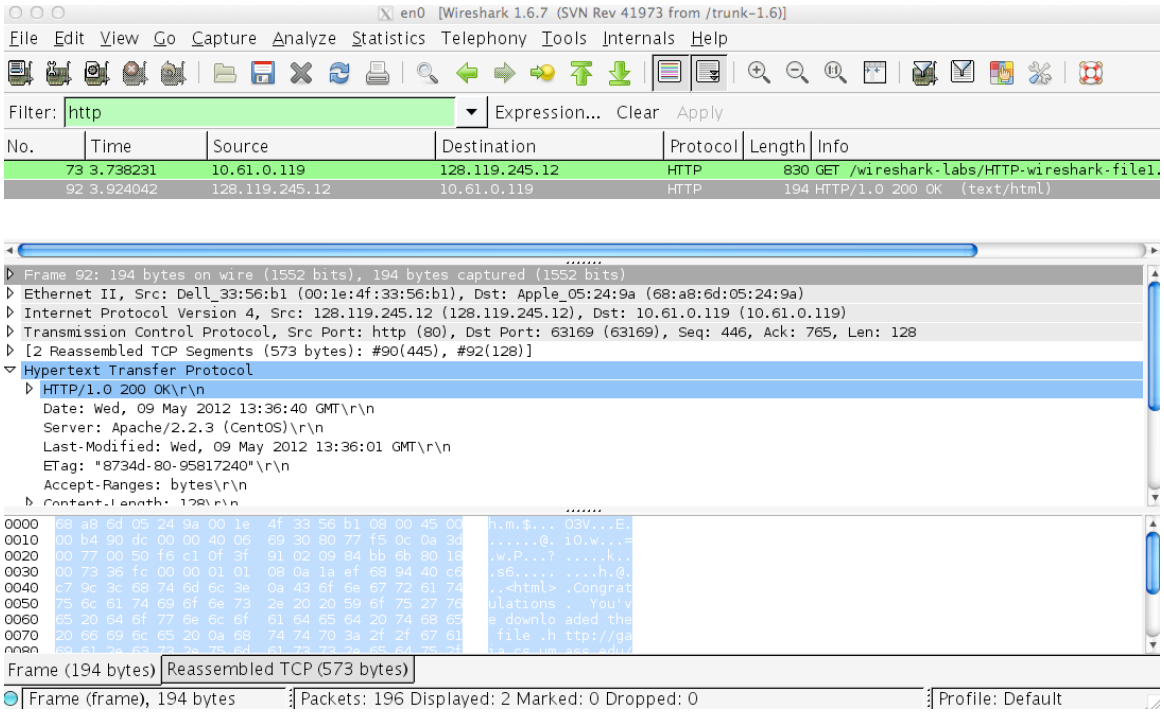
Chúng ta sẽ bắt đầu tìm hiểu HTTP bằng cách download một file HTML đơn giản.

Chú ý: nếu chúng ta không thể chạy Wireshark trên Internet thật sự thì có thể mở file http-ethereal-trace-1 có sẵn trong thư mục wireshark-traces.

Thực hiện các bước sau khi có kết nối Internet:

- Khởi động trình duyệt web
- Khởi động Wireshark và gõ "http" vào display-filter window để Wireshark chỉ hiển thị các thông điệp HTTP.
- Bắt đầu bắt gói tin.
- Gõ vào trình duyệt web: http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file1.html
- Dừng bắt gói tin.

Cửa sổ Wireshark lúc nào giống như trong hình 1.



Hình 1: Cửa sổ Wireshark sau khi trang web http://gaia.cs.umass.edu/wireshark-labs/ HTTP- wireshark-file1.html được hiển thị trên trình duyệt

Ví dụ trong hình 1 cho thấy packet-listing window chứa 2 thông điệp HTTP được bắt: thông điệp GET (từ trình duyệt gửi đến gaia.cs.umass.edu) và thông điệp response từ server đến trình duyệt. Packet-contents window hiển thị chi tiết của thông điệp được chọn (trong trường hợp này thông điệp HTTP OK đang được chọn). Tạm thời chúng ta chỉ quan tâm đến HTTP.

Bằng cách quan sát HTTP GET và HTTP response, trả lời các câu hỏi sau:

1. Trình duyệt đang sử dụng phiên bản HTTP 1.0 hay 1.1? Phiên bản HTTP server đang sử dụng là bao nhiêu?
2. Trình duyệt hỗ trợ những ngôn ngữ nào?
3. Địa chỉ IP của máy tính chúng ta là bao nhiêu? Của gaia.cs.umass.edu server là bao nhiêu?
4. Mã trạng thái (status code) trả về từ server là gì?
5. Thời điểm file HTML được thay đổi lần cuối tại server là lúc nào?
6. Server đã trả về cho trình duyệt bao nhiêu bytes nội dung?

2 HTTP GET/response có điều kiện

Hầu hết các web browsers đều hỗ trợ caching và thực hiện HTTP GET có điều kiện. Trước khi thực hiện các bước sau, xóa cache của trình duyệt (đối với Firefox, chọn *Tools->Clear Recent History* và chọn *Cache box* hoặc đối với Internet Explorer thì chọn *Tools->Internet Options->Delete File*).

Chú ý: nếu chúng ta không thể chạy Wireshark trên Internet thật sự thì có thể mở file http-ethereal-trace-2 trong thư mục wireshark-traces.

Thực hiện các bước sau khi có kết nối Internet:

- Khởi động trình duyệt và cần đảm bảo cache của trình duyệt đã được xóa.
- Khởi động Wireshark và bắt đầu bắt gói tin
- Từ trình duyệt, truy cập đến địa chỉ sau http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file2.html .

Trình duyệt sẽ hiển thị một file HTML đơn giản gồm có 5 dòng.

- Nhanh chóng nhập URL đó và truy cập đến một lần nữa (hoặc chọn refresh button trên trình duyệt).
- Dừng bắt gói tin và nhập "http" vào cửa sổ display-filter để hiển thị các thông điệp HTTP.

Trả lời các câu hỏi sau:

7. Xem xét nội dung của HTTP GET đầu tiên. Chúng ta có thấy dòng "IF-MODIFIED-SINCE" hay không?

8. Xem xét nội dung phản hồi từ server. Server có thật sự trả về nội dung của file HTML hay không? Tại sao?

9. Xem xét nội dung của HTTP GET thứ 2. Chúng ta có thấy dòng "IF-MODIFIED-SINCE" hay không? Nếu có, giá trị của IF-MODIFIED-SINCE là gì?

3 Truy cập các trang dài

Trong các ví dụ của chúng ta, trang được truy cập là những files HTML ngắn và đơn giản. Chúng ta sẽ xem xét điều gì xảy ra khi download một file HTML dài.

Chú ý: nếu chúng ta không thể chạy Wireshark trên Internet thật sự thì có thể mở file http-ethereal-trace-3 trong thư mục wireshark-traces.

Thực hiện các bước sau khi có kết nối Internet:

- Khởi động web browser và đảm bảo cache được xóa.
- Khởi động Wireshark và bắt đầu bắt gói tin.
- Từ trình duyệt, truy cập đến địa chỉ sau: http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file3.html
- Dừng bắt gói tin và nhập "http" vào display-filter window để hiển thị các thông điệp HTTP.

Trong packet-listing window, chúng ta sẽ thấy theo sau HTTP GET là nhiều gói tin TCP phản hồi. Ở trường hợp của chúng ta, file HTML có nội dung dài, 4500 bytes là quá lớn để có thể chứa trong một gói tin TCP. Chính vì thế HTTP response được TCP tách ra thành nhiều gói nhỏ, mỗi gói chứa trong một TCP segment. Trong các phiên bản Wireshark gần đây, Wireshark xác định mỗt TCP segment là một gói tin riêng biệt và thông điệp HTTP response được phân rã ra thành nhiều gói tin TCP được xác định bởi dòng "TCP segment of reassembled PDU" trong cột Info. Các phiên bản Wireshark cũ hơn thì sử dụng "Continuation".

Trả lời các câu hỏi sau

10. Trình duyệt đã gửi bao nhiêu HTTP GET? Dòng "**THE BILL OF RIGHTS**" được chứa trong gói tin phản hồi thứ mấy?
11. Gói tin phản hồi thứ mấy chứa mã trạng thái và ý nghĩa của nó?
12. Mã trạng thái và ý nghĩa của HTTP response là gì?
13. Cần bao nhiêu TCP segments để chứa hết HTTP response và nội dung của

The Bill of Rights?