Lucrarea de laborator nr.2.
Protocolul HTTP

În acest laborator, vor fi explorate câteva aspecte ale protocolului HTTP – interacţiunea de bază GET/response, formate de mesaje HTTP, preluarea de fişiere HTML mari, preluare de fişiere HTML cu obiecte incluse precum şi autentificarea şi securitatea HTTP.

Interacţiunea de bază HTTP GET/răspuns

Vom descărca un fişier simplu HTML, scurt, care nu conţine obiecte inserate.

1. Deschideţi un browser web.
2. Porniţi Wireshark, dar nu porniţi deocamdată capturarea de pachete. La filtru, scrieţi "http" astfel încât doar mesajele HTTP capturate vor fi afişate. Nu ne interesează încă alte tipuri de pachete.
3. Aşteptaţi mai mult de un minut (o sa vedeţi imediat de ce) şi apoi începeţi captura Wireshark.
4. Scrieţi adresa următoare în browser:
   http://vlab.unitbv.ro/rc/HTTP_1.html

5. După încărcarea paginii opriţi captura Wireshark.

Fereastra Wireshark ar trebui să arate asemănător cu Figura 6 de mai jos. Dacă nu va merge Wireshark, puteţi descărca de pe CD o altă copie de pachete (un alt *packet-trace*). Dezarhivaţi fişierul *http-trace-1*, şi apoi încărcaţi-l în Wireshark. Puteţi vedea copia acestui şir de pachete folosind submeniul "File", alegând "Open" şi selectând fişierul *http-trace-1*.

Exemplul arată că două mesaje HTTP au fost capturate: mesajul GET (din browserul dvs. către web serverul *vlab.unitbv.ro*) şi mesajul răspuns de la server la browser. Fereastra care arată conţinutul pachetelor ("packet-contents") prezintă detaliile mesajului selectat (în cazul de faţă, ale mesajului HTTP GET). Reţineţi că, datorită faptului că mesajul HTTP a fost transportat într-un segment TCP, care a fost inclus într-o datagramă IP, care a fost încapsulată într-un cadru Ethernet, Wireshark va arăta informaţii despre cadrul Ethernet, IP şi TCP.

Dorim să minimizăm informaţiile referitoare la alte protocoale şi ne concentrăm pe HTTP; deci checkbox-urile din stânga, din dreptul Frame, Ethernet, IP si TCP au un semn "+", iar linia HTTP va avea un semn "-" (adică este afişată informaţia despre HTTP).

Răspundeţi la următoarele întrebări, ale căror răspunsuri le veţi găsi în mesajele HTTP GET şi "response":

1. Browser-ul este versiunea 1.0 sau 1.1? Ce versiune de HTTP rulează pe server?
2. Care limbi (dacă există) indică browser-ul că pot fi acceptate de către server?
3. Care este adresa IP a calculatorului dvs.? Dar a serverului *vlab.unitbv.ro*?
4. Care este codul "status" întors de la server către browser?
5. Când a fost modificat ultima oară pe server, fişierul HTML pe care îl descărcaţi?

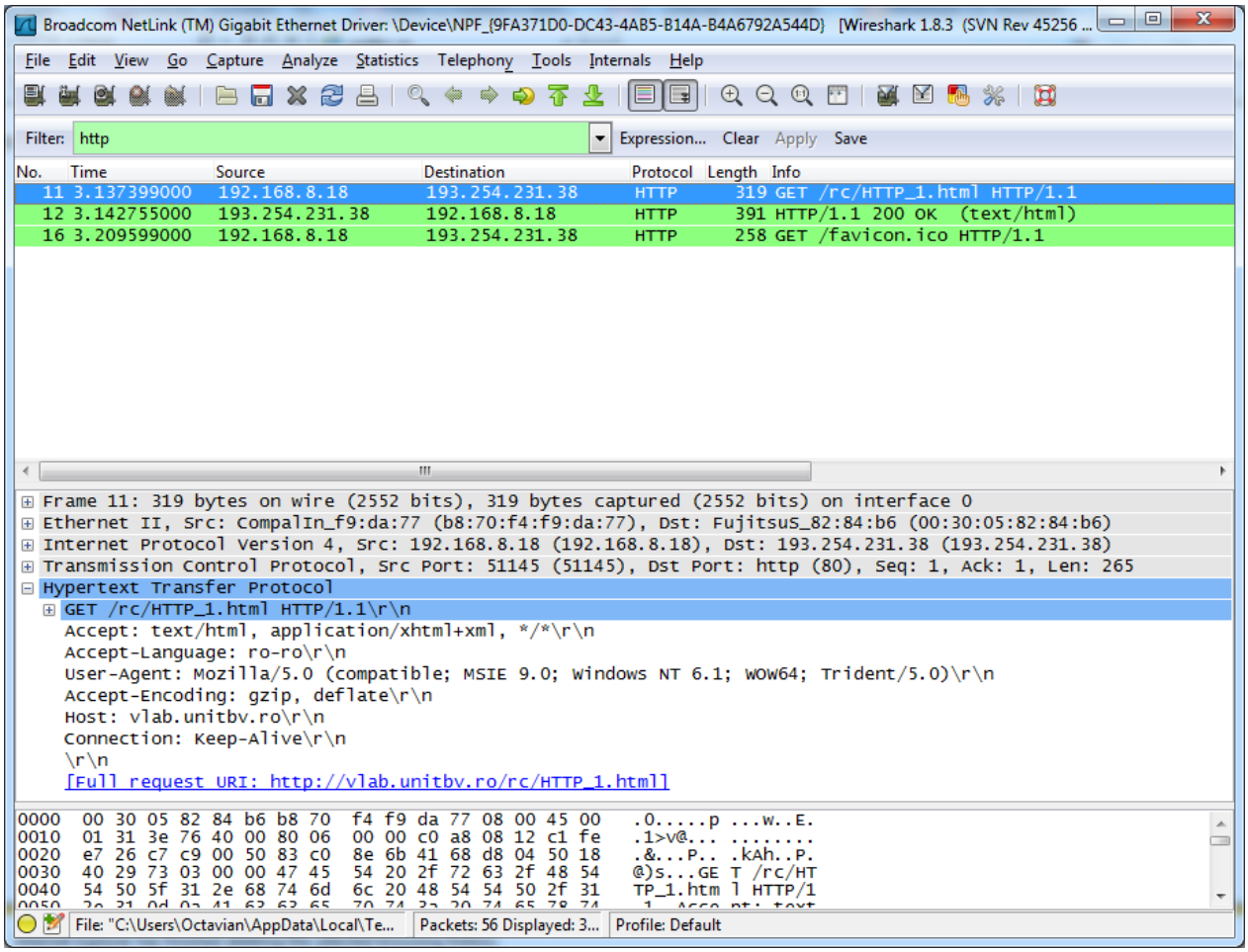

Figura 6 - Fereastra Wireshark după descărcarea fişierului

6. Câţi octeţi de payload sunt returnaţi către browser?
7. Priviţi fereastra "packet content". Puteţi vedea vreun header în datele care nu sunt afişate în fereastra packet-listing? Daca da, numiţi unul.

Referitor la întrebarea nr.5, se poate să fiţi surprinşi de faptul că documentul abia descărcat a fost modificat în minutul dinaintea descărcării lui. Acest lucru se datorează faptului că (pentru acest fişier), serverul *vlab.unitbv.ro* setează "last-modified time" ca să fie "current time", şi face asta în fiecare minut. Deci, dacă aşteptaţi un minut între accesări, fişierul va apărea ca fiind modificat recent, iar browser-ul va descărca o nouă copie a documentului.

Interacţiunea HTTP CONDITIONAL GET/response

Ştim că majoritatea browser-elor ascund obiectele (fac "object caching"), deci fac un GET condiţional atunci când descarcă un obiect HTTP. Înainte de a continua, ștergeți cache-ul din browser. Pentru Firefox avem *Tools* → *Clear Private Data* sau pentru Internet Explorer, *Tools* → *Internet Options* → *Delete File*. Acum să continuăm:

- Deschideţi browser-ul cu cache-ul gol.
- Porniţi Wireshark.
- Scrieţi următorul URL în browser:

http://vlab.unitbv.ro/rc/HTTP_2.html

Ar trebui să vedeţi un fişier HTML simplu, format din 5 linii.

- Acum daţi refresh la pagină.
- Opriţi Wireshark, şi la filtru scrieţi "http", astfel încât să vedem doar mesajele HTTP capturate.

Dacă nu va merge Wireshark, puteţi descărca de pe CD fișierul *http-trace-2.*

Răspundeţi la următoarele întrebări:

8. Priviţi prima cerere HTTP-GET de la browser către server. Vedeţi o linie "IF-MODIFIED-SINCE" ?
9. Priviţi răspunsul serverului. A întors serverul în mod explicit conţinutul fişierului? Justificaţi.
10. Acum priviţi a doua cerere HTTP-GET de la browser către server. Vedeţi o linie "IF-MODIFIED-SINCE" ? Dacă da, ce informaţie urmează după header-ul lui "IF-MODIFIED-SINCE" ?
11. Priviţi al doilea răspuns al serverului. A întors serverul în mod explicit conţinutul fişierului? Justificaţi.

Descărcarea de documente lungi

În exemplele de până acum, documentele descărcate au fost scurte. Vom face următoarele acţiuni pentru a analiza un document mai lung:

- Deschideţi browser-ul cu cache-ul gol.
- Porniţi Wireshark.
- Scrieţi următorul URL în browser:

http://vlab.unitbv.ro/rc/HTTP_3.html

Browserul ar trebui să vă arate un fişier lung.

- Opriţi Wireshark, şi la filtru scrieţi 'http' astfel încât să vedem doar mesaje HTTP.

Dacă nu va merge Wireshark, puteţi descărca de pe CD fișierul *http-trace-3.*

Acum ar trebui să se vadă mesajul HTTP GET, urmat de un răspuns cu pachete multiple. Să explicăm acest răspuns. Ştim că mesajul HTTP de răspuns se compune dintr-o linie de status, apoi urmează linii de header, apoi o linie blank, urmată de restul corpului mesajului ("entity body"). În cazul lui HTTP GET al nostru, "entity body" al răspunsului este *întregul fişier HTML cerut*. Fişierul nostru este lung, iar 6255 de octeţi nu încap într-un pachet TCP. Rezultă că singurul răspuns HTTP va fi împărţit în mai multe bucăţi de către TCP, fiecare bucată fiind conţinută într-un segment TCP separat. Fiecare segment TCP este văzut de către Wireshark ca un pachet separat. Faptul că un singur răspuns a fost fragmentat în mai multe pachete TCP este indicat de cuvântul "Continuation".

ATENTIE!!! Nu există vreun mesaj "*Continuation*" in Wireshark !!!

Răspundeţi la următoarele întrebări:

12. Câte mesaje de cerere HTTP GET au fost trimise de către browser?
13. De câte segmente TCP care conţin date, a fost nevoie pentru a trimite singurul răspuns HTTP?
14. Care este codul status asociat răspunsului cererii HTTP GET?
15. Există vreo linie status HTTP în transmisiunea asociată cu cuvântul "Continuation" TCP?

Documente HTML cu obiecte incluse

Să vedem acum ce se întâmplă când descărcăm un fişier cu obiecte încorporate, de exemplu o poză care este stocată pe alt server.

- Deschideţi browser-ul cu cache-ul gol.
- Porniţi Wireshark.
- Scrieţi următorul URL în browser:
  http://vlab.unitbv.ro/rc/HTTP_4.html

Dacă nu va merge Wireshark, puteţi descărca de pe CD fişierul *http-trace-4*.

Ar trebui să vedeţi un fişier HTML scurt cu 2 imagini. Aceste imagini nu sunt conţinute în HTML, ci adresa (URL-ul) imaginilor este conţinută în fişierul descărcat. Browserul trebuie să retragă aceste logo-uri de la anumite web site-uri. Logo-ul din exemplul nostru este luat de pe site-ul http://webbut.unitbv.ro. Imaginea coperţii este stocată la serverul http://grouper.ieee.org.

- Opriţi Wireshark, şi la filtru scrieţi "http" astfel încât să vedem doar mesaje HTTP.

Răspundeţi la următoarele întrebări:

16. Câte mesaje de cerere HTTP GET au fost trimise de către browser? Către care adresă?
17. Puteţi spune dacă browser-ul a descărcat cele 2 imagini una după alta, sau au fost descărcate simultan de la cele două site-uri? Justificaţi.

Autentificarea HTTP

Acum vom accesa o pagina html protejată cu parolă. URL-ul http://vlab.unitbv.ro/rc/protejat/HTTP_5.html este parolat.
Username: *user*
Parola: *student*
- Goliţi cache-ul browser-ului. Închideţi browser-ul şi porniţi-l din nou.
- Porniţi Wireshark.
- Accesaţi http://vlab.unitbv.ro/rc/protejat/HTTP_5.html.Introduceţi datele de autentificare de mai sus
- Opriţi captura Wireshark şi la filtru scrieţi ”http” astfel încât să vedem doar mesaje HTTP.

Dacă nu va merge Wireshark, puteţi descărca de pe CD fișierul *http-trace-5*.

Examinaţi captura şi răspundeţi la întrebările de mai jos. Este recomandat să consultaţi şi scurtul material de la http://frontier.userland.com/stories/storyReader$2159.

18. Care este răspunsul server-ului la primul mesaj HTTP GET ?
19. Când browser-ul trimite mesajul HTTP GET a doua oară, ce câmp nou apare în mesaj?

User-ul şi parola pe care le-ați folosit sunt codate în string-ul Basic *dXNlcjpzdHVkZW50* care urmează după header-ul “*Authorization: Basic*” în mesajul HTTP GET al clientului. User-ul şi parola NU SUNT criptate, doar codate în formatul Base64.

Puteți verifica, accesând pagina http://www.base64decode.org/ şi introducând string-ul codat Base64. Ar trebui să vă vedeţi user-ul şi parola.

Datorită faptului că oricine poate folosi Wireshark şi traduce din Base64, este de la sine înţeles că parolele simple pe site-uri www nu sunt securizate decât dacă se iau măsuri adiţionale. Despre aceste metode de securizare vom vorbi mai târziu.

Cuprins