

CPSC 471 – 06 Assignment #1

Required Software

Wireshark version 3.4.8 or newer must be used for this assignment. A penalty of 10% will apply if using older version.

Submission

You are responsible for the content of your submitted file. You should double-check in Canvas after the submission to ensure the submitted file is not corrupted, is readable and contains all required files and the contents are what you intend to submit. Submitting a wrong file (from another class for example) is not a valid excuse.

Submit one zip file using naming such as: **firstnamelastname_p1.zip** in Canvas.

The zip file shall contain the following files:

1. Report file using naming such as **firstnamelastname_p1.pdf** (must be a pdf document, not Word). This report file contains the answers to all the questions in this assignment. You must use the same ordering number for your responses (do not change the question ordering number for your answers, doing so will result in 10% points penalty)
2. Five Wireshark capture files using naming such as **firstnamelastname_p1wsx.pcapng** (where x is number 1 to 5)

Managing Submission File Size

Manage your submission File Size by controlling your Wireshark captures and check the zip file size before submitting. Do not exceed 10 Megabytes. A penalty of 10% will be deducted for each 10 MB violation in excess from the allowed 10 MB. For example, penalty for a 10.1 MB to 20 MB submission is 10%, and for a 20.1 MB to 30 MB submission the penalty is 20%. Limit your Wireshark capture file size by starting and stopping the capture process accordingly. Do not let your Wireshark run continuously without stopping. This will create a very large capture file and unnecessary penalty.

Analyzing HTTP Messages using Wireshark

Use Wireshark to capture and analyze HTTP messages.

Perform the followings and provide answers for the 19 questions below.

Basic HTTP GET/response interaction

Let's begin our exploration of HTTP by downloading a very simple HTML file - one that is very short and contains no embedded objects. Do the following:

- Start up your web browser.
- Start up the Wireshark packet sniffer, as described in the Introductory lab (but don't yet begin packet capture). Enter "http" (just the letters, not the quotation marks) in the display-filter-specification window, so that only captured HTTP messages will be displayed later in the packet-listing window. (We're only interested in the HTTP protocol here, and don't want to see the clutter of all captured packets).
- Wait a bit more than one minute (we'll see why shortly), and then begin Wireshark packet capture.
- Enter the following to your browser
<http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file1.html>
Your browser should display the very simple, one-line HTML file.
- Stop Wireshark packet capture and save the Wireshark capture in file **first-namelastname_p1ws1.pcapng**.

Your Wireshark window should look like the window shown in Figure 1.

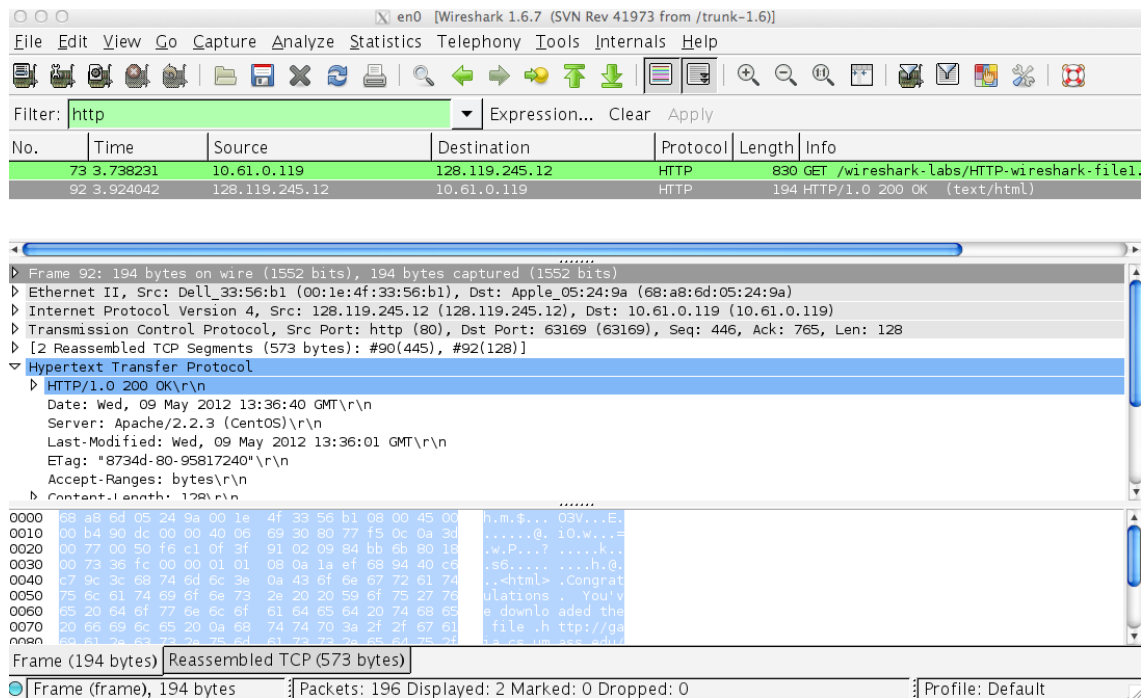


Figure 1: Wireshark Display after <http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file1.html> has been retrieved by your browser.

The example in Figure 1 shows in the packet-listing window that two HTTP messages were captured: the GET message (from your browser to the gaia.cs.umass.edu web server) and the response message from the server to your browser. The packet-contents window shows details of the selected message (in this case the HTTP OK message, which is highlighted in the packet-listing window). Recall that since the HTTP message was carried inside a TCP segment, which was carried inside an IP datagram, which was carried within an Ethernet frame, Wireshark displays the Frame, Ethernet, IP, and TCP packet information as well. We want to minimize the amount of non-HTTP data displayed (we're interested in HTTP here, and will be investigating these other protocols in later labs), so make sure the boxes at the far left of the Frame, Ethernet, IP and TCP information have a plus sign or a right-pointing triangle (which means there is hidden, not displayed information), and the HTTP line has a minus sign or a down-pointing triangle (which means that all information about the HTTP message is displayed).

(Note: You should ignore any HTTP GET and response for `favicon.ico`. If you see a reference to this file, it is your browser automatically asking the server if it (the server) has a small icon file that should be displayed next to the displayed URL in your browser. We'll ignore references to this pesky file in this lab.)

Answer the following questions

By looking at the information in the HTTP GET and response messages, answer the following questions. **When answering the following questions, you should include the screen capture of the GET and response messages and indicate where in the message you've found the information that answers the following questions.**

1. Is your browser running HTTP version 1.0 or 1.1? What version of HTTP is the server running?
2. What languages (if any) does your browser indicate that it can accept to the server?
3. What is the IP address of your computer? Of the gaia.cs.umass.edu server?
4. What is the status code returned from the server to your browser?
5. When was the HTML file that you are retrieving last modified at the server?
6. How many bytes of content are being returned to your browser?
7. By inspecting the raw data in the packet content window, do you see any headers within the data that are not displayed in the packet-listing window? If so, name one.

In your answer to question 5 above, you might have been surprised to find that the document you just retrieved was last modified within a minute before you downloaded the document. That's because (for this particular file), the gaia.cs.umass.edu server is setting the file's last-modified time to be the current time and is doing so once per minute. Thus, if you wait a minute between accesses, the file will appear to have been recently modified, and hence your browser will download a "new" copy of the document.

The HTTP CONDITIONAL GET/response interaction

Recall from Section 2.2.5 of the text, that most web browsers perform object caching and thus perform a conditional GET when retrieving an HTTP object. Before performing the steps below, make sure your browser's cache is empty. (To do this under Firefox, select *Tools->Clear Recent History* and check the Cache box, or for Internet Explorer, select *Tools->Internet Options->Delete File*; these actions will remove cached files from your browser's cache.) Now do the following:

- Start up your web browser, and make sure your browser's cache is cleared, as discussed above.
- Start up the Wireshark packet sniffer
- Enter the following URL into your browser
<http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file2.html>
Your browser should display a very simple five-line HTML file.
- Quickly enter the same URL into your browser again (or simply select the refresh button on your browser)
- Stop Wireshark packet capture and enter "http" in the display-filter-specification window, so that only captured HTTP messages will be displayed later in the packet-listing window. Save the Wireshark capture in file **firstnamelastname_p1ws2.pcapng**.

Answer the questions

8. Inspect the contents of the first HTTP GET request from your browser to the server. Do you see an “IF-MODIFIED-SINCE” line in the HTTP GET?
9. Inspect the contents of the server response. Did the server explicitly return the contents of the file? How can you tell?
10. Now inspect the contents of the second HTTP GET request from your browser to the server. Do you see an “IF-MODIFIED-SINCE:” line in the HTTP GET? If so, what information follows the “IF-MODIFIED-SINCE:” header?
11. What is the HTTP status code and phrase returned from the server in response to this second HTTP GET? Did the server explicitly return the contents of the file? Explain.

Retrieving Long Documents

In our examples thus far, the documents retrieved have been simple and short HTML files. Let’s next see what happens when we download a long HTML file. Do the following:

- Start up your web browser, and make sure your browser’s cache is cleared, as discussed above.
- Start up the Wireshark packet sniffer
- Enter the following URL into your browser
<http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file3.html>
Your browser should display the rather lengthy US Bill of Rights.
- Stop Wireshark packet capture and enter “http” in the display-filter-specification window, so that only captured HTTP messages will be displayed. Save the capture in file **firstnamelastname_p1ws3.pcapng**

In the packet-listing window, you should see your HTTP GET message, followed by a multiple-packet TCP response to your HTTP GET request. This multiple-packet response deserves a bit of explanation. Recall from Section 2.2 (see Figure 2.9 in the text) that the HTTP response message consists of a status line, followed by header lines, followed by a blank line, followed by the entity body. In the case of our HTTP GET, the entity body in the response is the *entire* requested HTML file. In our case here, the HTML file is rather long, and at 4500 bytes is too large to fit in one TCP packet. The single HTTP response message is thus broken into several pieces by TCP, with each piece being contained within a separate TCP segment (see Figure 1.24 in the text). In recent versions of Wireshark, Wireshark indicates each TCP segment as a separate packet, and the fact that the single HTTP response was fragmented across multiple TCP packets is indicated by the “TCP segment of a reassembled PDU” in the Info column of the Wireshark display. Earlier versions of Wireshark used the “Continuation” phrase to indicate that the entire content of an

HTTP message was broken across multiple TCP segments. We stress here that there is no “Continuation” message in HTTP!

Answer the questions

12. How many HTTP GET request messages did your browser send? Which packet number in the trace contains the GET message for the Bill of Rights?
13. Which packet number in the trace contains the status code and phrase associated with the response to the HTTP GET request?
14. What is the status code and phrase in the response?
15. How many data-containing TCP segments were needed to carry the single HTTP response and the text of the Bill of Rights?

HTML Documents with Embedded Objects

Now that we’ve seen how Wireshark displays the captured packet traffic for large HTML files, we can look at what happens when your browser downloads a file with embedded objects, i.e., a file that includes other objects (in the example below, image files) that are stored on another server(s).

Do the following:

- Start up your web browser, and make sure your browser’s cache is cleared, as discussed above.
- Start up the Wireshark packet sniffer
- Enter the following URL into your browser
<http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file4.html>
Your browser should display a short HTML file with two images. These two images are referenced in the base HTML file. That is, the images themselves are not contained in the HTML; instead, the URLs for the images are contained in the downloaded HTML file. As discussed in the textbook, your browser will have to retrieve these logos from the indicated web sites. Our publisher’s logo is retrieved from the gaia.cs.umass.edu web site. The image of the cover for our 5th edition (one of our favorite covers) is stored at the caite.cs.umass.edu server. (These are two different web servers inside cs.umass.edu).
- Stop Wireshark packet capture and enter “http” in the display-filter-specification window, so that only captured HTTP messages will be displayed. Save the capture in file **firstnamelastname_p1ws4.pcapng**

Answer the questions

16. How many HTTP GET request messages did your browser send? To which Internet addresses were these GET requests sent?
17. Can you tell whether your browser downloaded the two images serially, or whether they were downloaded from the two web sites in parallel? Explain.

HTTP Authentication

Finally, let's try visiting a web site that is password-protected and examine the sequence of HTTP message exchanged for such a site. The URL

http://gaia.cs.umass.edu/wireshark-labs/protected_pages/HTTP-wireshark-file5.html is password protected. The username is "**wireshark-students**" (without the quotes), and the password is "**network**" (again, without the quotes). So, let's access this "secure" password-protected site. Do the following:

- Make sure your browser's cache is cleared, as discussed above, and close your browser. Then, start up your browser
- Start up the Wireshark packet sniffer
- Enter the following URL into your browser
http://gaia.cs.umass.edu/wireshark-labs/protected_pages/HTTP-wireshark-file5.html
Type the requested username and password into the pop up box.
- Stop Wireshark packet capture and enter "http" in the display-filter-specification window, so that only captured HTTP messages will be displayed later in the packet-listing window. Save the capture in file **firstnamelastname_p1ws5.pcapng**

Now let's examine the Wireshark output. You might want to first read up on HTTP authentication by reviewing the easy-to-read material on "HTTP Access Authentication Framework" at [http://frontier.userland.com/stories/storyReader\\$2159](http://frontier.userland.com/stories/storyReader$2159)

Answer the questions

18. What is the server's response (status code and phrase) in response to the initial HTTP GET message from your browser?
19. When your browser's sends the HTTP GET message for the second time, what new field is included in the HTTP GET message?

The username (wireshark-students) and password (network) that you entered are encoded in the string of characters (d2lyZXNoYXJrLXN0dWRlbnRzOm5ldHdvcm91) following the "Authorization: Basic" header in the client's HTTP GET message. While it may appear that your username and password are encrypted, they are simply encoded in a format known as Base64

format. The username and password are *not* encrypted! To see this, go to <http://www.motobit.com/util/base64-decoder-encoder.asp> and enter the base64-encoded string d2lyZXNoYXJrLXN0dWRIbnRz and decode. *Voila!* You have translated from Base64 encoding to ASCII encoding, and thus should see your username! To view the password, enter the remainder of the string Om5ldHdvcms= and press decode. Since anyone can download a tool like Wireshark and sniff packets (not just their own) passing by their network adaptor, and anyone can translate from Base64 to ASCII (you just did it!), it should be clear to you that simple passwords on WWW sites are not secure unless additional measures are taken.

Fear not! As we will see in Chapter 8, there are ways to make WWW access more secure. However, we'll clearly need something that goes beyond the basic HTTP authentication framework!

Example Screen Capture and Highlights

When answering each question, you should include a Wireshark screen capture and highlight the area(s) to indicate where you get the answer from.

The captures must contain sufficient surrounding detail information relevant to the question.

The screen captures must be clear and readable to get full credit.

The following is an example:

```
Frame 135 (488 bytes on wire, 488 bytes captured)
Ethernet II, Src: Cisco-Li 48:1f:1b:00:22:6b, Dst: IntelCor dc:36:d0 (00:22:fa:dc:12:35:42), Dst Port: 55428 (55428), Seq: 1, Ack: 43
Internet Protocol, Src: 128.128.128.128, Dst: 192.168.1.101 (192.168.1.101)
Transmission Control Protocol, Seq: 1, Ack: 43
Hypertext Transfer Protocol
  HTTP/1.1 200 OK\r\n
    (Expert Info (Chat/Sequence): HTTP/1.1 200 OK\r\n)
    Request Version: HTTP/1.1
    Response Code: 200
    Date: Wed, 27 Oct 2010 11:26:58 GMT\r\n
    Server: Apache/2.2.52 (CentOS)\r\n
    Last-Modified: Wed, 27 Oct 2010 11:26:01 GMT\r\n
    ETag: "8734d-80-7d/4e440"\r\n
    Accept-Ranges: bytes\r\n
    Content-Length: 128\r\n
    [Content length: 128]
    Keep-Alive: timeout=10, max=100\r\n
    Connection: Keep-Alive\r\n
    Content-Type: text/html; charset=ISO-8859-1\r\n
    \r\n
Line-based text data: text/html
<html>\n
  Congratulations. You've downloaded the file \n
  http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file1.html!\n
</html>\n
```

Return status: 200 OK

server running http

document last modified

content: 128

End of Assignment 1