

实验 1：HTTP

班级_____ 学号 2021 姓名_____

一、实验内容

在入门实验中接触了 Wireshark 数据包嗅探器之后，现在我们将实验 Wireshark 来研究操作中的协议。在本实验中，我们将探讨 HTTP 协议的下面几个方面：

- 1、基本的 HTTP 获取/响应交互
- 2、HTTP 条件获取/响应交互
- 3、长文档的检索
- 4、带有内嵌对象的 HTTP

二、实验操作步骤及结果

1、基本的 HTTP 获取/响应交互

(1) 实验步骤

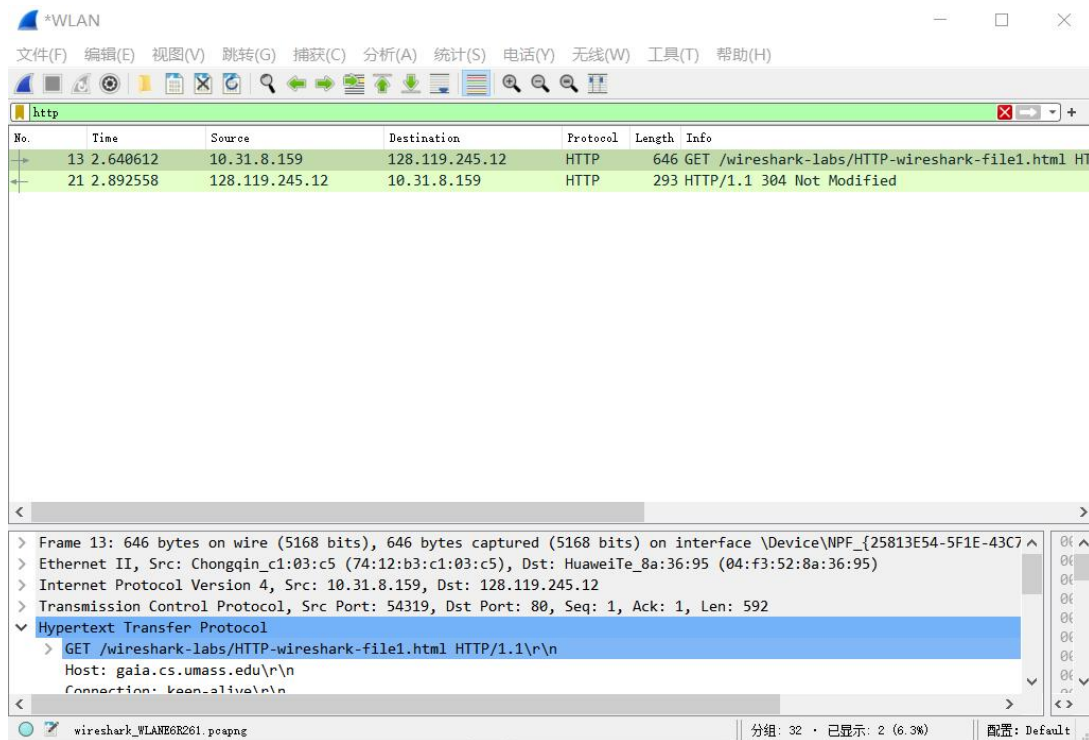
- ① 启动网络浏览器。
- ② 启动 Wireshark 数据包嗅探器，在显示筛选器规范窗口中输入“http”，使封包列表中只显示 http 信息。
- ③ 稍等一分钟以上，然后开始 Wireshark 数据包捕获。
- ④ 在浏览器中输入：

<http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file1.html>

浏览器显示一行非常简单的 HTML 文件。

Congratulations. You've downloaded the file <http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file1.html>!

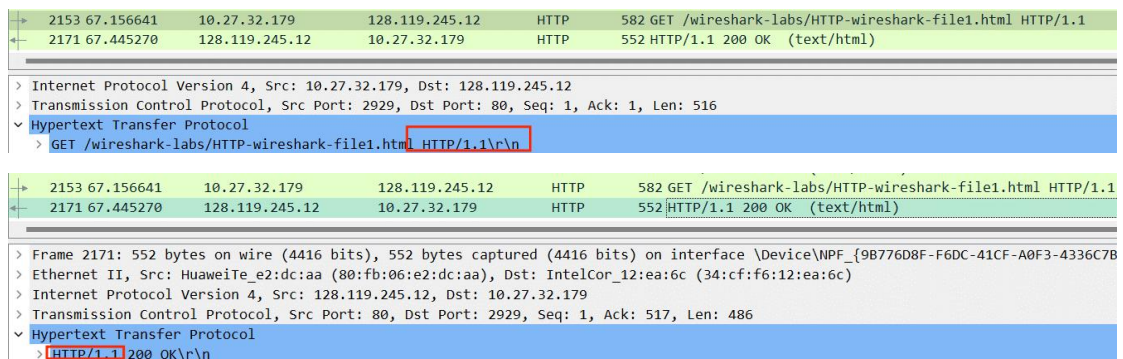
- ⑤ 返回 Wireshark 查看抓包情况。
- ⑥ 通过查看 HTTP GET 和响应消息中的信息，回答问题。



(2) 问题回答

1. Is your browser running HTTP version 1.0 or 1.1? What version of HTTP is the server running?

实验中使用的浏览器运行的是 HTTP/1.1，服务器运行的也是 HTTP/1.1。如图：



2. What languages (if any) does your browser indicate that it can accept to the server?

浏览器可以接受的语言：zh-CN; en; en-GB; en-US。如图：

Accept-Language: zh-CN, zh; q=0.9, en; q=0.8, en-GB; q=0.7, en-US; q=0.6\r\n

3. What is the IP address of your computer? Of the gaia.cs.umass.edu server?

客户机 IP 地址：10.31.8.159;

服务器地址:128.119.245.12。如下图所示：

```
Internet Protocol Version 4, Src: 10.31.8.159, Dst: 128.119.245.12
0100 .... = Version: 4
.... 0101 = Header Length: 20 bytes (5)
```

4. What is the status code returned from the server to your browser?

从服务器返回到浏览器的状态码是 200，如图：

```
▼ Hypertext Transfer Protocol
  ▼ HTTP/1.1 200 OK\r\n
```

5. When was the HTML file that you are retrieving last modified at the server?

检索的 HTML 文件上次在服务器修改的时间：世界时间 2023.06.29, 05:59:01, 星期四。如图：

```
Last-Modified: Thu, 29 Jun 2023 05:59:01 GMT\r\n
```

6. How many bytes of content are being returned to your browser?

包含 128 字节。如下图所示：

```
▼ Content-Length: 128\r\n
  [Content length: 128]
```

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.

并不是所有内容都能显示出来，数据包列表窗口中未显示的标题包括：Host 字段、Connection 字段、User-Agent 字段等：

```
▼ Hypertext Transfer Protocol
  > GET /wireshark-labs/HTTP-wireshark-file1.html HTTP/1.1\r\n
    Host: gaia.cs.umass.edu\r\n
    Connection: keep-alive\r\n
    Upgrade-Insecure-Requests: 1\r\n
    User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML,
    Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,image/ap
    Accept-Encoding: gzip, deflate\r\n
```

2、HTTP 条件获取/响应交互

(1) 实验步骤

- ① 清理浏览器缓存。
- ② 打开 Wireshark，开始数据包捕获
- ③ 在浏览器中输入：

<http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file2.html>

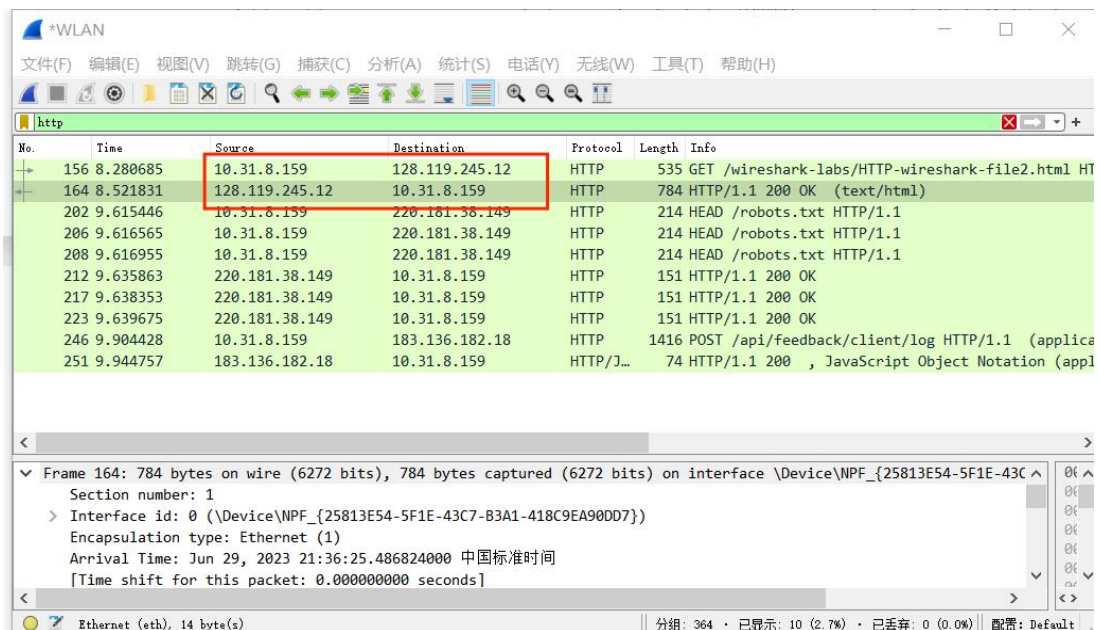
- ④ 刷新网页，停止数据包捕获，选择 http 过滤。

浏览器显示 5 行文字，内容如下：

Congratulations again! Now you've downloaded the file lab2-2.html.
This file's last modification date will not change.

Thus if you download this multiple times on your browser, a complete copy
will only be sent once by the server due to the inclusion of the IN-MODIFIED-SINCE
field in your browser's HTTP GET request to the server.

Wireshark 捕获情况如下：



(2) 问题回答

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?

资料：在 HTTP 中，通过使用 If-Modified-Since 请求头，可以避免服务器传输已经存在于客户端缓存中且未改变的响应数据。这有助于减少网络流量并提高性能，尤其是对频繁访问静态或缓慢变化的资源。如果文件在服务器上的最后修改时间与客户端发送的 If-Modified-Since 时间相同，服务器会返回一个 HTTP 304 Not Modified 响应，而不是文件内容，这样客户端就可以直接从本地缓存中获取文件，而不需要再次下载。

检查第一个 HTTP GET 请求的内容，没有找到 “If-Modified-Since”。说明客户端本地没有缓存最新修改的文件（确实在做这一步的时候我是第一次打开链接）。返回状态码为 200，说明返回了 HTML 文件检查服务器返回的相应的内容，可以查找到服务器明确返回的文件的内容。如下图所示：

```
Line-based text data: text/html (10 lines)
\n
<html>\n
\n
Congratulations again! Now you've downloaded the file lab2-2.html. <br>\n
This file's last modification date will not change. <p>\n
Thus if you download this multiple times on your browser, a complete copy <br>\n
will only be sent once by the server due to the inclusion of the IN-MODIFIED-SINCE<br>\n
field in your browser's HTTP GET request to the server.\n
\n
</html>\n
```


9. Inspect the contents of the server response. Did the server explicitly return the contents of the file? How can you tell?

显式的显示了内容，因为有 Line-based text data。具体内容如下图所示：

```
Line-based text data: text/html (10 lines)
\n
<html>\n
\n
Congratulations again! Now you've downloaded the file lab2-2.html. <br>\n
This file's last modification date will not change. <p>\n
Thus if you download this multiple times on your browser, a complete copy <br>\n
will only be sent once by the server due to the inclusion of the IN-MODIFIED-SINCE<br>\n
field in your browser's HTTP GET request to the server.\n
\n
</html>\n
```

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?

有 IF-MODIFIED-SINCE。最后修改时间是 Thu, 29 Jun 2023 05:59:01。

```
Accept-Encoding: gzip, deflate\r\n
Accept-Language: zh-CN,zh;q=0.9,en;q=0.8\r\n
If-None-Match: "173-5ff3e65a210e3"\r\n
If-Modified-Since: Thu, 29 Jun 2023 05:59:01 GMT\r\n
```

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.

未显示的包含文件内容，因为近期请求过相同的文件，客户端中有缓存。该响应中无 line-based text data，并且状态码为 304，表明缓存中已有文件内容。

```
> [SEQ/ACK analysis]
TCP payload (240 bytes)
▼ Hypertext Transfer Protocol
> HTTP/1.1 304 Not Modified\r\n
Date: Thu, 29 Jun 2023 13:57:10 GMT\r\n
Server: Apache/2.4.6 (CentOS) OpenSSL/1.0.2k-fips PHP/7.4.33 mod_perl/2.0.11 Perl/v5.16.3\r\n
Connection: Keep-Alive\r\n
Keep-Alive: timeout=5, max=100\r\n
ETag: "173-5ff3e65a210e3"\r\n
\r\n
[HTTP response 1/1]
[Time since request: 0.240923000 seconds]
[Request in frame: 25]
```

3、长文档的检索

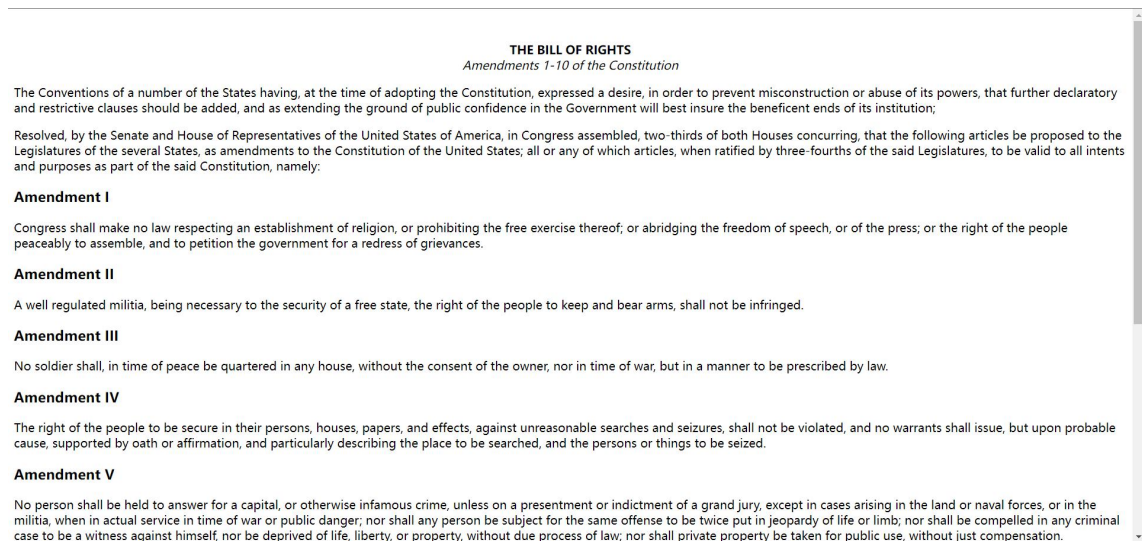
(1) 实验步骤

- ① 清理浏览器缓存。
- ② 打开 Wireshark，开始数据包捕获
- ③ 在浏览器中输入：

<http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file3.html>

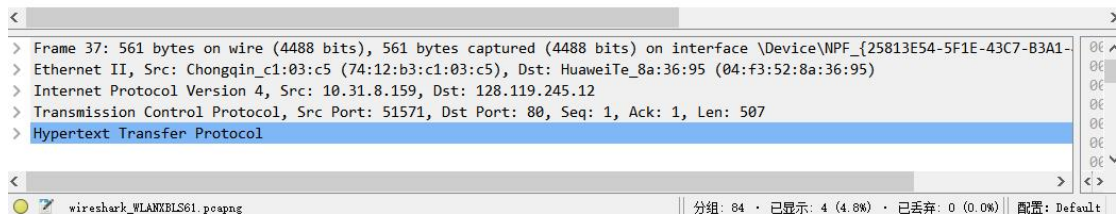
- ④ 刷新网页，停止数据包捕获，选择 http 过滤。

浏览器显示内容如下：



Wireshark 捕获情况如下：

No.	Time	Source	Destination	Protocol	Length	Info
37	8.120540	10.31.8.159	128.119.245.12	HTTP	561	GET /wireshark-labs/HTTP-wireshark-file3.html HTTP/1.1
44	8.365200	128.119.245.12	10.31.8.159	HTTP	775	HTTP/1.1 200 OK (text/html)
60	10.561835	10.31.8.159	119.147.19.225	HTTP	782	POST /mmtls/00007740 HTTP/1.1
67	10.737929	119.147.19.225	10.31.8.159	HTTP	963	HTTP/1.1 200 OK



(2) 问题回答

12. How many HTTP GET request messages did your browser send? Which packet number in the trace contains the GET message for the Bill or Rights?

共一条 HTTP GET 请求消息（No. 37），如下图所示：

No.	Time	Source	Destination	Protocol	Length	Info
37	8.120540	10.31.8.159	128.119.245.12	HTTP	561	GET /wireshark-labs/HTTP-wireshark-file3.html HTTP/1.1
44	8.365200	128.119.245.12	10.31.8.159	HTTP	775	HTTP/1.1 200 OK (text/html)
60	10.561835	10.31.8.159	119.147.19.225	HTTP	782	POST /mmtls/00007740 HTTP/1.1
67	10.737929	119.147.19.225	10.31.8.159	HTTP	963	HTTP/1.1 200 OK

HTTP GET 请求的下面一条（No. 44）包中含有 Bill 信息，如下图所示：

No.	Time	Source	Destination	Protocol	Length	Info
37	8.120540	10.31.8.159	128.119.245.12	HTTP	561	GET /wireshark-labs/HTTP-wireshark-file3.html HTTP/1.1
44	8.365200	128.119.245.12	10.31.8.159	HTTP	775	HTTP/1.1 200 OK (text/html)
60	10.561835	10.31.8.159	119.147.19.225	HTTP	782	POST /mmtls/00007740 HTTP/1.1
67	10.737929	119.147.19.225	10.31.8.159	HTTP	963	HTTP/1.1 200 OK

```

<
>
> Hypertext Transfer Protocol
> Line-based text data: text/html (98 lines)
<html><head> \n
<title>Historical Documents:THE BILL OF RIGHTS</title></head>\n
\n
\n
<body bgcolor="#ffffff" link="#330000" vlink="#666633">\n
<p><br>\n
</p>\n
<p></p><center><b>THE BILL OF RIGHTS</b><br>\n
<em>Amendments 1-10 of the Constitution</em>\n
</center>\n
\n
<n>The Conventions of a number of the States having at the time of adopting\n

```

13. Which packet number in the trace contains the status code and phrase associated with the response to the HTTP GET request?

No. 44 的报文包含了响应 HTTP GET 请求的状态码和短语。如下图所示:

No.	Time	Source	Destination	Protocol	Length	Info
37	8.120540	10.31.8.159	128.119.245.12	HTTP	561	GET /wireshark-labs/HTTP-wireshark-file3.html HTTP/1.1
44	8.365200	128.119.245.12	10.31.8.159	HTTP	775	HTTP/1.1 200 OK (text/html)
60	10.561835	10.31.8.159	119.147.19.225	HTTP	782	POST /mmtls/00007740 HTTP/1.1
67	10.737929	119.147.19.225	10.31.8.159	HTTP	963	HTTP/1.1 200 OK

```

<
>
> Frame 44: 775 bytes on wire (6200 bits), 775 bytes captured (6200 bits) on interface \Device\NPF_{25813E54-5F1E-43C7-B34}
> Ethernet II, Src: NewH3CTe_62:f5:50 (9c:54:c2:62:f5:50), Dst: Chongqin_c1:03:c5 (74:12:b3:c1:03:c5)
> Internet Protocol Version 4, Src: 128.119.245.12, Dst: 10.31.8.159
> Transmission Control Protocol, Src Port: 80, Dst Port: 51571, Seq: 4141, Ack: 508, Len: 721
> [4 Reassembled TCP Segments (4861 bytes): #41(1380), #42(1380), #43(1380), #44(721)]
> Hypertext Transfer Protocol
> Line-based text data: text/html (98 lines)
<html><head> \n
<title>Historical Documents:THE BILL OF RIGHTS</title></head>\n
\n
\n
<body bgcolor="#ffffff" link="#330000" vlink="#666633">\n
<p><br>\n
</p>\n
</n>\n

```

14. What is the status code and phrase in the response?

如上图所示: 状态码为 200, 响应短语为 OK。

15. How many data-containing TCP segments were needed to carry the single HTTP response and the text of the Bill of Rights?

需要 4 个 TCP 段。如下图所示:

No.	Time	Source	Destination	Protocol	Length	Info
37	8.120540	10.31.8.159	128.119.245.12	HTTP	561	GET /wireshark-labs/HTTP-wireshark-file3.html HTTP/1.1
44	8.365200	128.119.245.12	10.31.8.159	HTTP	775	HTTP/1.1 200 OK (text/html)
60	10.561835	10.31.8.159	119.147.19.225	HTTP	782	POST /mmtls/00007740 HTTP/1.1
67	10.737929	119.147.19.225	10.31.8.159	HTTP	963	HTTP/1.1 200 OK

```

<
>
> Frame 44: 775 bytes on wire (6200 bits), 775 bytes captured (6200 bits) on interface \Device\NPF_{25813E54-5F1E-43C7-B34}
> Ethernet II, Src: NewH3CTe_62:f5:50 (9c:54:c2:62:f5:50), Dst: Chongqin_c1:03:c5 (74:12:b3:c1:03:c5)
> Internet Protocol Version 4, Src: 128.119.245.12, Dst: 10.31.8.159
> Transmission Control Protocol, Src Port: 80, Dst Port: 51571, Seq: 4141, Ack: 508, Len: 721
> [4 Reassembled TCP Segments (4861 bytes): #41(1380), #42(1380), #43(1380), #44(721)]
> [Frame: 41, payload: 0-1379 (1380 bytes)]
> [Frame: 42, payload: 1380-2759 (1380 bytes)]
> [Frame: 43, payload: 2760-4139 (1380 bytes)]
> [Frame: 44, payload: 4140-4860 (721 bytes)]
> [Segment count: 4]
> [Reassembled TCP length: 4861]
> [Reassembled TCP Data: 485454502f312e3120323030204f4b0d0a446174653a205468752c203239204a756e2032...]
> Hypertext Transfer Protocol

```


4、带有内嵌对象的 HTTP

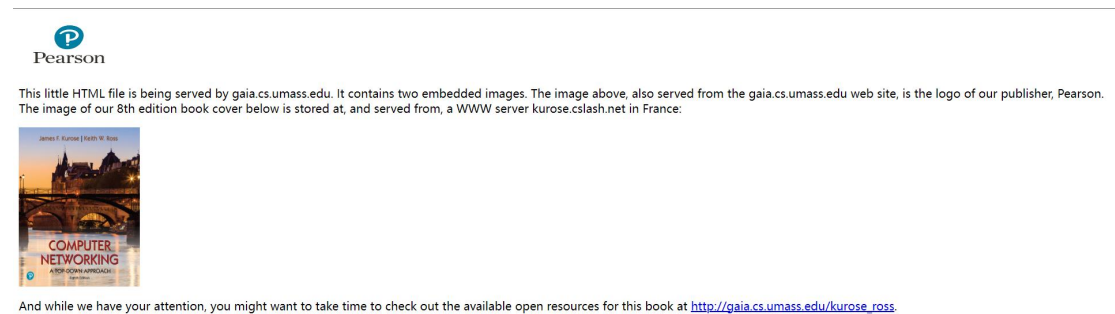
(1) 实验步骤

- ⑤ 清理浏览器缓存。
- ⑥ 打开 Wireshark，开始数据包捕获
- ⑦ 在浏览器中输入：

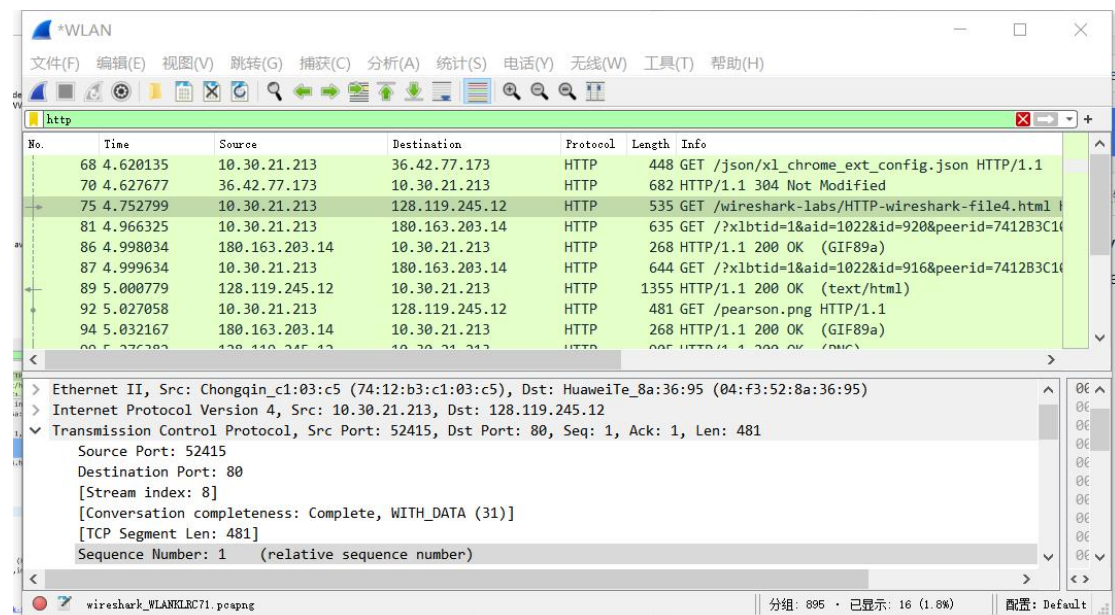
<http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file4.html>

- ⑧ 刷新网页，停止数据包捕获，选择 http 过滤。

浏览器显示内容如下：



Wireshark 捕获情况如下：



(2) 问题回答

16. How many HTTP GET request messages did your browser send? To which Internet addresses were these GET requests sent?

浏览器发送了 3 条 HTTP 获取请求消息：

主页发往：128.119.245.12
图片 1 发往：128.119.245.12
图片 2 发往：178.79.137.164
如下图所示：

No.	Time	Source	Destination	Protocol	Length	Info
6	0.243938	10.30.21.213	128.119.245.12	HTTP	561	GET /wireshark-labs/HTTP-wireshark-file4.html HTTP/1.1
10	0.490227	128.119.245.12	10.30.21.213	HTTP	1355	HTTP/1.1 200 OK (text/html)
13	0.502132	10.30.21.213	128.119.245.12	HTTP	481	GET /pearson.png HTTP/1.1
17	0.744939	10.30.21.213	178.79.137.164	HTTP	448	GET /8E_cover_small.jpg HTTP/1.1
20	0.747690	128.119.245.12	10.30.21.213	HTTP	905	HTTP/1.1 200 OK (PNG)
23	0.983301	178.79.137.164	10.30.21.213	HTTP	225	HTTP/1.1 301 Moved Permanently
94	2.361702	10.30.21.213	110.249.194.71	HTTP/1.1	963	POST / HTTP/1.1, JavaScript Object Notation (application/json)
98	2.408283	110.249.194.71	10.30.21.213	HTTP	60	HTTP/1.1 200 OK
698	17.981564	10.30.21.213	14.119.104.254	HTTP	214	HEAD /robots.txt HTTP/1.1
700	18.018167	14.119.104.254	10.30.21.213	HTTP	151	HTTP/1.1 200 OK
714	19.199036	10.30.21.213	113.219.132.72	HTTP	341	GET /msdownload/update/v3/static/trustedr/en/disallowlist.dat HTTP/1.1
715	19.226895	113.219.132.72	10.30.21.213	HTTP	448	HTTP/1.1 304 Not Modified
722	19.658046	10.30.21.213	110.249.194.71	HTTP/1.1	963	POST / HTTP/1.1, JavaScript Object Notation (application/json)
726	19.697617	110.249.194.71	10.30.21.213	HTTP	60	HTTP/1.1 200 OK

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.

浏览器下载两张图片存在先后顺序，从上图中的 Time 一栏中可见，两个请求的时间不同。如下图所示：

No.	Time	Source	Destination	Protocol	Length	Info
123	16.375545	10.63.30.60	128.119.245.12	HTTP	557	GET /wireshark-labs/HTTP-wireshark-file4.html HTTP/1.1
125	16.619983	128.119.245.12	10.63.30.60	HTTP	1375	HTTP/1.1 200 OK (text/html)
126	16.641248	10.63.30.60	128.119.245.12	HTTP	503	GET /pearson.png HTTP/1.1
134	16.886347	128.119.245.12	10.63.30.60	HTTP	949	HTTP/1.1 200 OK (PNG)
139	16.929325	10.63.30.60	178.79.137.164	HTTP	470	GET /8E_cover_small.jpg HTTP/1.1
144	17.188634	178.79.137.164	10.63.30.60	HTTP	245	HTTP/1.1 301 Moved Permanently

三、实验中存在问题分析

本次实验刚开始是就遇到了问题，我捕获不到所需的 HTTP 内容，起初以为是选择的捕获接口有误的原因，更换了接口，但是问题依然没有解决。在网上查了各种方法都没有解决，花了一下午的时间。查阅文档发现可以用压缩包里的文件进行实验，但是我觉得还是应该自己试着捕获一下，毕竟实验课主打一个体验。某天我重启电脑，再次尝试，居然就解决了这个问题。（这个故事告诉我们解决不了的问题可以先放一放）

后续实验中，有时候会出现状态码不符合实验所需条件的情况，后来发现，清理浏览器缓存重新尝试后问题得到解决，我也慢慢熟悉了 Wireshark 的操作。此外，由于实验文档是纯英文的，有时候不太能理解题目的含义，需要仔细阅读文本理解含义。在借助翻译工具和查阅相关资料后问题也得到了解决。

总之，遇到问题后应该积极尝试解决问题的方法，多查阅资料，又或者是寻求他人的帮助都是良策，这是我这次实验的收获。