# 实验内容

- 1.基本HTTP GET/response交互
- 2.HTTP条件Get/response交互
- 3.检索长文件
- 4.具有嵌入对象的HTML文档
- 5.HTTP认证

# 实验步骤

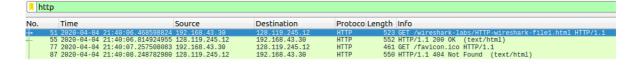
1.基本HTTP GET/response交互

先开浏览器稍等一下(原因后面会解释),然后打开 wireshark ,点击开始捕获之后,在浏览器输入 URL

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

捕获停止后的界面如下

这是数据包51号--请求消息



绿色部分的 Internet Protocol Version 4 其实就是 IPV4 的缩写,接着的 Transmission Control Protocol 就是 TCP 的全称,而且在这一行还显示了源地址跟目的地址的端口号等信息,最后的 Hypertext Transfer Protocol 就是 HTTP 的全称。

这是数据包55号--http响应消息

#### 回答问题

1.您的浏览器是否运行HTTP版本1.0或1.1?服务器运行什么版本的HTTP?

浏览器跟服务器都是运行的HTTP1.1版本

2.您的浏览器会从接服务器接受哪种语言(如果有的话)?

在请求消息图中的 Accept-Language 显示的是 zh-CN,也就是中文啦

3.您的计算机的IP地址是什么?gaia.cs.umass.edu服务器地址呢?

计算机的IP地址: 192.168.43.0

gaia.cs.umass.edu服务器地址: 128.119.245.12

4.服务器返回到浏览器的状态代码是什么?

200 ,代表OK

5.服务器上HTML文件的最近一次修改是什么时候?

在响应消息的图中显示的是

Last-Modified: Sat, 04 Apr 2020 05:59:02 GMT\r\n

6.服务器返回多少字节的内容到您的浏览器?

在响应消息的图中显示的是128字节

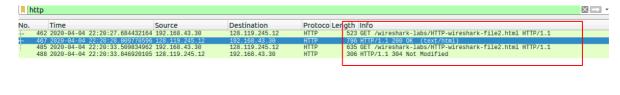
Accept-Ranges: bytes
Content-Length: 128\r\n

7.通过检查数据包内容窗口中的原始数据,你是否看到有协议头在数据包列表窗口中未显示?如果是,请举一个例子。

TCP

# 2.HTTP条件Get/response交互

先清除浏览器缓存,打开浏览器后,开始捕获,输入上次的URL,成功显示后刷新或者重新进入上次的 网址,停止捕获后的界面如下。



#### 回答问题

8.检查第一个从您浏览器到服务器的HTTP GET请求的内容。您在HTTP GET中看到了"IF-MODIFIED-SINCE"行吗?

# 

9.检查服务器响应的内容,服务器是否显式返回文件的内容?你是怎么知道的?

10.现在,检查第二个HTTP GET请求的内容。您在HTTP GET中看到了"IF-MODIFIED-SINCE:"行吗?如果是,"IF-MODIFIED-SINCE:"头后面包含哪些信息?

If-Modified-Since: Sat, 04 Apr 2020 05:59:02 GMT\r\n 包含的是上次服务器更新的时间。(因为gaia.cs.umass.edu服务器将这个特定文件的最后修改时间设置为当前时间,并且每分钟执行一次)

```
Frame 485: 635 bytes on wire (5080 bits), 635 bytes captured (5080 bits) on interface 0

Ethernet II, Src: IntelCor_11:2c:59 (9c:dai3e:11:2c:59), Dst: f6:b7:b3:8f:c3:92 (f6:b7:b3:8f:c3:92)

Internet Protocol Version 4, Src: 192.168.43.30, Dst: 128.119.245.12

Transmission Control Protocol, Src Port: 37282, Dst Port: 80, Seq: 1, Ack: 1, Len: 569

Hypertext Transfer Protocol

| GET Wireshark-labs/HTTP-wireshark-file2.html HTTP/1.l\r\n
| Host: gaia.cs.umass.edu\r\n
| Connection: keep-alive\r\n
| Cache-Control: max-age=0\r\n
| Upgrade-Insecure-Requests: 1\r\n
| User-Agent: Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/73.0.3683.86 Safari/537.36\r\n
| Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,image/apng,*/*;q=0.8,application/signed-exchange;v=b3\r\n
| Accept-Encoding: gzip, deflate\r\n
| Accept-Language: zh-CN, zh;q=0.9, en;q=0.8\r\n
| If-Modified-Since: Sat, 04 Apr 2020 05:59:02 GMT\r\n
| \r\n
| Full request URI: http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file2.html]
| HTTP request URI: http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file2.html]
| Response in frame: 488]
```

11.针对第二个HTTP GET,从服务器响应的HTTP状态码和短语是什么?服务器是否明确地返回文件的内容?请解释。

状态码短语:306 HTTP/1.1 304 Not Modified 服务并没有明确返回文件内容

No.		Time	Source	Destination	Protoco Lei	ngth Info
1	462	2020-04-04 22:20:27.684432164	192.168.43.30	128.119.245.12	HTTP	523 GET /wireshark-labs/HTTP-wireshark-file2.html HTTP/1.1
1	467	2020-04-04 22:20:28.009770596	128.119.245.12	192.168.43.30	HTTP	796 HTTP/1.1 200 OK (text/html)
-	485	2020-04-04 22:20:33.509834962	192.168.43.30	128.119.245.12	HTTP	635 GET /wireshark-labs/HTTP-wireshark-file2.html HTTP/1.1
+	488	2020-04-04 22:20:33.846920105	128.119.245.12	192.168.43.30	HTTP	306 HTTP/1.1 304 Not Modified

```
| Frame 488: 306 bytes on wire (2448 bits), 306 bytes captured (2448 bits) on interface 0
| Ethernet II, Src: f6:b7:b3:8f:c3:92 (f6:b7:b3:8f:c3:92), Dst: IntelCor_11:2c:59 (9c:da:3e:11:2c:59)
| Internet Protocol Version 4, Src: 128.119.245.12, Dst: 192.168.43.30
| Transmission Control Protocol, Src Port: 80, Dst Port: 37282, Seq: 1, Ack: 570, Len: 240
| Hypertext Transfer Protocol
| HTTP/1.1 304 Not Modified\r\n
| Date: Sat, 04 Apr 2020 14:2e:33 GMT\r\n
| Server: Apache/2.4.6 (CentOS) OpenSSL/1.0.2k-fips PHP/5.4.16 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: "173-5a270bb15aea1"\r\n
| \r\n
| [HTP response 1/1]
| [Time since request: 0.337085143 seconds]
| Request In frame: 485]
| [Request URI: http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file2.htm1]
```

## 3.检索长文件

这次的URL

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

## 捕获后的截图如下

```
Frame 628: 4927 bytes on wire (39416 bits), 4927 bytes captured (39416 bits) on interface 0

Ethernet II, Src: f6:b7:b3:8f:c3:92 (f6:b7:b3:8f:c3:92), Dst: IntelCor_11:2c:59 (9c:da:3e:11:2c:59)

Internet Protocol Version 4, Src: 128.119.245.12, Dst: 192.168.43.30

Transmission Control Protocol, Src Port: 80, Dst Port: 37584, Seq: 1, Ack: 458, Len: 4861

Hypertext Transfer Protocol

Intelligence of the data: Fext/Ntml (98 lines)
```

不知道为什么我检索的长文件,并没有显示TCP数据包之间分段的信息,在TCP的头消息中包的数量为 1......,本来按照书上的说法,这次的文件比较大,应该是有多个TCP数据包的,后来重复几次之后,发 现还是1个TCP数据包

在我们的例子中,HTML文件相当长,4500字节太大,一个TCP数据包不能容纳。因此,单个HTTP响应消息由TCP分成几个部分,每个部分包含在单独的TCP报文段中。在Wireshark的最新版本中,Wireshark将每个TCP报文段指定为独立的数据包,并且单个HTTP响应在多个TCP数据包之间分段的事实由Wireshark显示的Info列的"重组PDU的TCP段"指示。

#### 回答问题

12.您的浏览器发送多少HTTP GET请求消息?哪个数据包包含了美国权利法案的消息?

发送了1个HTTP GET请求消息,数据包620包含了美国权利法案的消息



13.哪个数据包包含响应HTTP GET请求的状态码和短语?

620

14.响应中的状态码和短语是什么?

200 OK

15.需要多少包含数据的TCP段来执行单个HTTP响应和权利法案文本?

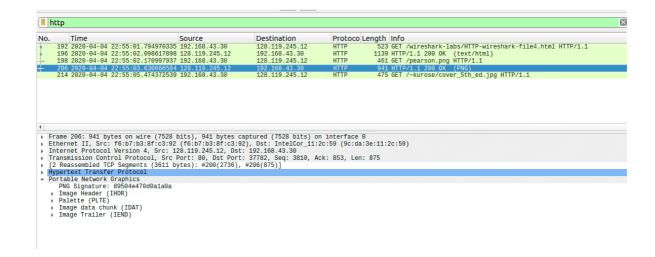
我这里包含了1个

# 4.具有嵌入对象的HTML文档

这次的URL

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

最终的实验结果

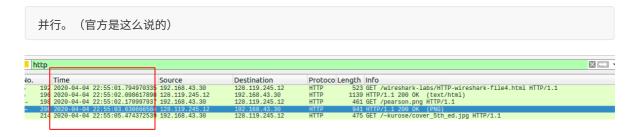


#### 回答问题

16.您的浏览器发送了几个HTTP GET请求消息?这些GET请求发送到哪个IP地址?

```
3个请求。发送到地址128.119.245.12。
```

17.浏览器从两个网站串行还是并行下载了两张图片?请说明。

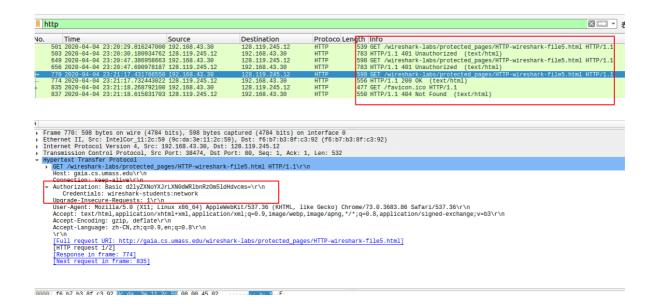


## 5.HTTP认证

这次的URL

http://gaia.cs.umass.edu/wireshark-labs/protected\_pages/HTTP-wireshark-file5.html

捕获截图



#### 回答问题

18.对于您的浏览器的初始HTTP GET消息,服务器响应(状态码和短语)是什么响应?

401 Unauthorized

19.当您的浏览器第二次发送HTTP GET消息时,HTTP GET消息中包含哪些新字段?

如图,多了一下这段

Authorization: Basic d2lyZXNoYXJrLXN0dWRlbnRzOm5ldHdvcms=\r\n

Credentials: wireshark-students:network

# 实验回顾

1.为什么要先开浏览器后稍等一会再开始捕获

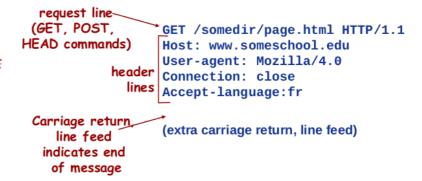
如果在打开浏览器之前开始捕获,那么会捕获到很多与此次实验无关的HTTP报文。

2.回顾HTTP请求消息跟响应消息的格式

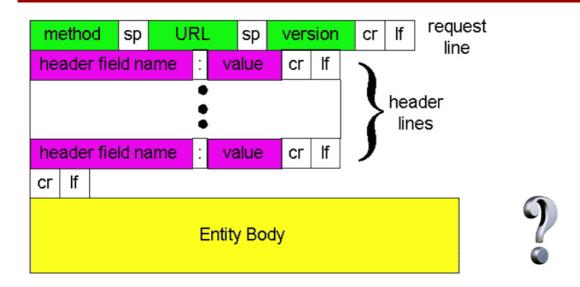
只有真的自己抓包看一看才印象比较深刻,有时候PPT看完了也就过

# HTTP请求消息

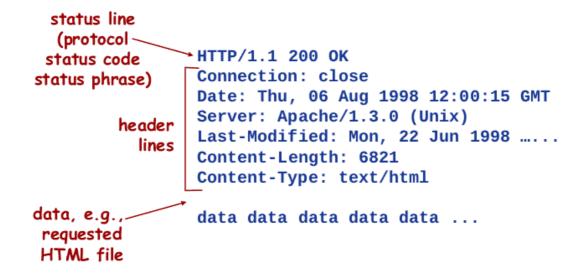
- ❖HTTP协议有两类消息
  - 请求消息(request)
  - 响应消息(response)
- ❖请求消息
  - ASCII: 人直接可读



# HTTP请求消息的通用格式



# HTTP响应消息



# 参考资料

- <u>Wireshark实验HTTP</u>
- <u>HTTP验证方案</u>
- <u>HTTP与HTTPS的区别</u>

# 附录

1.谷歌浏览器清除缓存

(记录一下防止以后忘了)

来到 更多工具 -> 清除浏览数据 。选择 过去一个小时 。



# 2.HTTP验证方案

参考资料第二点,是一个对HTTP验证方案的简单介绍

如果HTTP客户端(例如Web浏览器)请求作为受保护领域一部分的页面,则服务器将以401未经授权状态码进行响应,并在其响应中包含WWW-Authenticate标头字段。此标头字段必须至少包含一个适用于所请求页面的身份验证质询。

接下来,客户端发出另一个请求,这次包括一个Authentication标头字段,其中包含适用于服务器身份验证质询的客户端凭据。

如果服务器接受凭据,它将返回请求的页面。否则,它将返回另一个"401未经授权"响应,以通知客户端身份验证已失败。

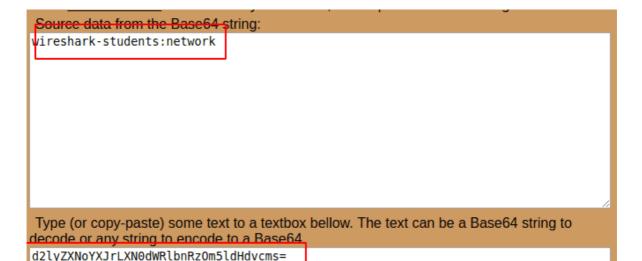
## 3.解码过程

您输入的用户名(wireshark-students)和密码(network)按照客户端HTTP GET消息中请求头的 "Authorization:Basic"的字符串(d2lyZXNoYXJrLXN0dWRlbnRzOm5ldHdvcms=)编码。

所谓用户名和密码可能"加密",只是以一种称为Base64格式的格式进行编码。打开在线<u>解码工具</u>,输入HTTP GET消息中请求头的 Authorization:Basic 的字符串,也就是 d2lyZXNoYXJrLXN0dWRlbnRzOm5ldHdvcms= 。选择解码 decode 。

Type (or copy-paste) some text to a textbox bellow. The text can be a Base64 string to decode or any string to encode to a Base64 d2lyZXNoYXJrLXN0dWRlbnRz0m5ldHdvcms= or select a file to convert to a Base64 string. Convert the source data 选择文件 未选择任何文件 What to do with the source data: encode the source data to a Base64 string (base64 encoding) - Maximum characters por line: 76 decode the data from a Base64 string (base64 decoding) Output data: output to a textbox (as a string) export to a binary file, filename: base64.bin Note: The source file is handled as a binary data. The textbox is handled as a string data, default character set for the textbox is 'iso-8859-1'. You can change the charset using form bellow.

最后转换出来的 ASCII 码,用户名字跟密码都有了。



## 4.HTTP与HTTPS

### 参考资料第三点,就截取一些片段吧

HTTPS协议的主要作用可以分为两种:一种是建立一个信息安全通道,来保证数据传输的安全;另一种就是确认网站的真实性。

简单来说,HTTPS协议是由SSL+HTTP协议构建的可进行加密传输、身份认证的网络协议,要比 http协议安全。

#### HTTPS和HTTP的区别主要如下:

- 1、https协议需要到ca申请证书,一般免费证书较少,因而需要一定费用。
- 2、http是超文本传输协议,信息是明文传输,https则是具有安全性的ssl加密传输协议。
- 3、http和https使用的是完全不同的连接方式,用的端口也不一样,前者是80,后者是443。
- 4、http的连接很简单,是无状态的;HTTPS协议是由SSL+HTTP协议构建的可进行加密传输、身份认证的网络协议,比http协议安全。

HTTP使用TCP三次握手建立连接,客户端和服务器需要交换3个包;HTTPS除了TCP的三个包,还要加上ssl握手需要的9个包,所以一共是12个包。