实验二

超文本传输协议

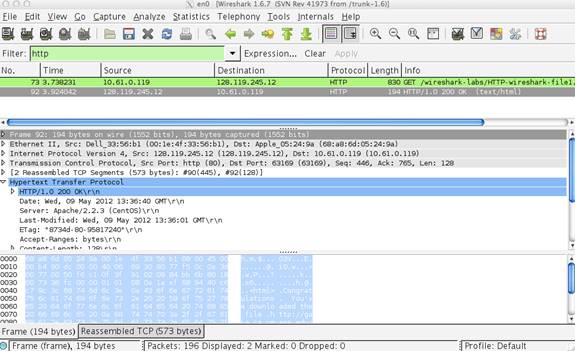
在这个实验室中，我们将使用Wireshark来研究HTTP协议，探索几个方面：基本的GET/response交互、HTTP消息格式、检索大型HTML文件、检索带有嵌入对象的HTML文件以及HTTP身份验证和安全性。在开始这些实验之前，请先复习一下课本的第2.2节[[1]](#footnote-1)。

一。基本的HTTP GET/response交互

让我们从下载一个非常简单的HTML文件开始我们对HTTP的探索——这个文件非常短，并且不包含嵌入的对象

1. 启动网络浏览器。
2. 启动Wireshark数据包嗅探器，如介绍性实验室中所述（但不要启动数据包捕获）。在“显示筛选器规范”窗口中输入“http”（仅输入字母，而不是引号），以便稍后在“数据包列表”窗口中仅显示捕获的HTTP消息对这里的HTTP协议感兴趣，并且不想看到所有捕获包的混乱
3. 稍等一分钟以上（我们很快就会知道原因），然后开始Wireshark数据包捕获。
4. 在浏览器http://gaia.cs.umass.edu/wireshark-labs/http-wireshark-file1.html中输入以下内容您的浏览器将显示非常简单的单行html文件。
5. 停止Wireshark数据包捕获。

您的Wireshark窗口应该类似于图1所示的窗口。如果您无法在实时网络连接上运行Wireshark，则可以下载在执行上述步骤时创建的数据包跟踪。[2]



**图1：**浏览器检索到http://gaia.cs.umass.edu/Wireshark-labs/http-Wireshark-file1.html后显示Wireshark

图1中的示例在“数据包列表”窗口中显示捕获了两条HTTP消息：GET消息（从浏览器到gaia.cs.umass.edu web服务器）和response消息（从服务器到浏览器）。“数据包内容”窗口显示所选消息的详细信息（在本例中为HTTP OK消息，这在“数据包列表”窗口中突出显示。）回想一下，由于HTTP消息是在TCP段中携带的，而TCP段是在IP数据包中携带的，而IP数据包是在以太网帧中携带的，因此Wireshark也会显示帧、以太网、IP和TCP数据包信息。我们希望最小化显示的非HTTP数据量（我们对这里的HTTP很感兴趣，我们将在稍后的实验室中研究这些其他协议），因此请确保帧、以太网、IP和TCP信息最左边的框有加号或右指三角形（这意味着存在隐藏的、未显示的信息），HTTP行有减号或下指三角形（这意味着将显示有关HTTP消息的所有信息）。

（注意：您应该忽略favicon.ico的任何HTTP GET和响应。如果您看到对该文件的引用，它是您的浏览器自动询问服务器（服务器）是否有一个小图标文件，该文件应该显示在您浏览器中显示的URL旁边。我们将忽略对这个讨厌的文件的引用。）。

通过查看HTTP GET和response消息中的信息，回答以下问题。——在回答以下问题时，您应该打印出GET和response消息（有关如何执行此操作的解释，请参阅Wireshark实验室介绍），并指出在消息中您在哪里找到了回答接下来的问题。当你交作业时，对输出进行注释，以便清楚你在输出中获得答案信息的位置（例如，对于我们的课程，我们要求学生用钢笔标记纸质副本，或者用彩色字体的文本对电子副本进行注释）。

1. 您的浏览器运行的是HTTP版本1.0还是1.1？服务器运行的是什么版本的HTTP？
2. 您的浏览器可以接受哪些语言（如果有的话）？
3. 你电脑的IP地址是多少？是gaia.cs.umass.edu服务器吗？
4. 从服务器返回到浏览器的状态代码是什么？
5. 您检索的HTML文件上次在服务器上修改是什么时候？
6. 有多少字节的内容被返回到您的浏览器？
7. 通过检查数据包内容窗口中的原始数据，您是否看到数据包列表窗口中未显示的任何数据头？如果是的话，说出一个。

在您对上述问题5的回答中，您可能会惊讶地发现，您刚刚检索到的文档是在您下载文档之前一分钟内最后一次修改的。因为（对于这个特定的文件），gaia.cs.umass.edu服务器正在将文件上次修改的时间设置为当前时间，并且每分钟执行一次。因此，如果您在两次访问之间等待一分钟，则该文件将显示为最近修改过的文件，因此您的浏览器将下载该文档的“查看”副本。

2。HTTP条件GET/response交互

回想文本的第2.2.5节，大多数web浏览器执行对象缓存，因此在检索HTTP对象时执行条件GET。在执行以下步骤之前，请确保浏览器缓存为空。（要在Firefox下执行此操作，请选择“工具”->“清除最近的历史记录”并选中“缓存”框，或者对于Internet Explorer，请选择“工具”->“Internet选项”->“删除文件”；这些操作将从浏览器缓存中删除缓存的文件。）现在执行以下操作：

* 启动web浏览器，并确保浏览器缓存已清除，如上文所述。
* 启动Wireshark数据包嗅探器
* 在浏览器http://gaia.cs.umass.edu/wireshark-labs/http-wireshark-file2.html中输入以下URL您的浏览器应该会显示一个非常简单的五行html文件。
* 再次在浏览器中快速输入相同的URL（或只需选择浏览器上的“刷新”按钮）
* 停止Wireshark数据包捕获，并在“显示筛选器规范”窗口中输入“ttp”，以便稍后在“数据包列表”窗口中只显示捕获的HTTP消息
* （注意：如果您无法在实时网络连接上运行Wireshark，您可以使用http-ethereal-trace-2包跟踪来回答下面的问题；请参见脚注1。此跟踪文件是在一台author抯计算机上执行上述步骤时收集的。）

回答下列问题：

1. 检查从浏览器到服务器的第一个HTTP GET请求的内容。你在httpget中看到一行“F-MODIFIED-SINCE”吗？
2. 检查服务器响应的内容。服务器是否显式返回文件的内容？你怎么知道？
3. 现在检查从浏览器到服务器的第二个HTTP GET请求的内容。你在httpget中看到一行“F-MODIFIED-SINCE:”吗？如果是，那么“F-MODIFIED-SINCE:”标题后面有什么信息？
4. 服务器返回的响应第二个HTTP GET的HTTP状态代码和短语是什么？服务器是否显式返回文件的内容？牋解释。

三。检索长文档

到目前为止，在我们的示例中，检索到的文档是简单而简短的HTML文件。接下来让我们看看当我们下载一个很长的HTML文件时会发生什么

* 启动web浏览器，并确保浏览器缓存已清除，如上文所述。
* 启动Wireshark数据包嗅探器
* 在您的浏览器http://gaia.cs.umass.edu/wireshark-labs/http-wireshark-file3.html中输入以下URL您的浏览器应该会显示相当长的美国版权法案。
* 停止Wireshark包捕获，并在display filter specification窗口中输入“ttp”，以便只显示捕获的HTTP消息。
* （注意：如果您无法在实时网络连接上运行Wireshark，您可以使用http-ethereal-trace-3包跟踪来回答下面的问题；请参见脚注1。此跟踪文件是在一台author抯计算机上执行上述步骤时收集的。）

在“数据包列表”窗口中，您应该看到您的HTTP GET消息，然后是对httpget请求的多数据包TCP响应。——这个多数据包响应应该得到一点解释。——回想一下第2.2节（参见文本中的图2.9），HTTP响应消息由状态行、头行和头行组成在我们的HTTP GET中，响应中的实体体是整个请求的HTML文件。在我们的例子中，HTML文件很长，4500字节太大，不适合放在一个TCP包中。因此，单个HTTP响应消息被TCP分成几部分，每一部分都有包含在一个单独的TCP段中（参见文本中的图1.24）。在Wireshark的最新版本中，Wireshark将每个TCP段表示为一个单独的包，Wireshark显示的Info列中重新组合的PDU的CP段表示单个HTTP响应在多个TCP包中被分段。Wireshark的早期版本使用了continuation短语来表示HTTP消息的整个内容在多个TCP中被分段段……我们在此强调，在HTTP中没有“继续”消息！

回答下列问题：

1. 浏览器发送了多少HTTP GET请求消息？追踪中哪个包号包含票据或权利的GET消息？
2. 跟踪中的哪个数据包编号包含与对HTTP GET请求的响应相关联的状态代码和短语？
3. 回复中的状态码和短语是什么？
4. 承载单个HTTP响应和权利法案文本需要多少包含TCP段的数据？

四。包含嵌入对象的HTML文档

现在我们已经看到Wireshark如何显示捕获的大型HTML文件的数据包流量，我们可以看看当您的浏览器下载包含嵌入对象的文件时会发生什么情况，即，包含存储在另一个服务器上的其他对象（在下例中，图像文件）的文件。

执行以下操作：

* 启动web浏览器，并确保浏览器缓存已清除，如上文所述。
* 启动Wireshark数据包嗅探器
* 在浏览器http://gaia.cs.umass.edu/wireshark-labs/http-wireshark-file4.html中输入以下URL您的浏览器应显示一个包含两个图像的简短html文件。这两个图像在基本的HTML文件中被引用，也就是说，图像本身不包含在HTML中，而是图像的URL包含在下载的HTML文件中。如教科书中所述，您的浏览器必须从指定的网站检索这些徽标。我们的出版商徽标从gaia.cs.umass.edu网站检索。我们第5版（我们最喜爱的封面之一）的封面图像存储在caite.cs.umass.edu服务器上。（这是cs.umass.edu中的两个不同的web服务器）。
* 停止Wireshark包捕获，并在display filter specification窗口中输入“ttp”，以便只显示捕获的HTTP消息。
* （注意：如果您无法在实时网络连接上运行Wireshark，您可以使用http-ethereal-trace-4包跟踪来回答下面的问题；请参见脚注1。此跟踪文件是在一台author抯计算机上执行上述步骤时收集的。）

回答下列问题：

1. 浏览器发送了多少HTTP GET请求消息？-这些GET请求被发送到哪些Internet地址？
2. 你能告诉你的浏览器是连续下载这两张图片，还是同时从两个网站下载的吗？解释一下。

5 HTTP身份验证

最后，让我们试着访问一个受密码保护的网站，并检查为这样一个网站交换的HTTP消息的顺序

http://gaia.cs.umass.edu/wireshark-labs/protected\_pages/http-wireshark-file5.html受密码保护。用户名是搘ireshark students搘（不带引号），密码是搉etwork搉（同样，不带引号）。因此，让抯访问受密码保护的站点。请执行以下操作：

* 如上所述，确保浏览器缓存已清除，然后关闭浏览器。然后，启动浏览器
* 启动Wireshark数据包嗅探器
* 在浏览器中输入以下URL http://gaia.cs.umass.edu/wireshark-labs/protected\_pages/http-wireshark-file5.html在弹出框中键入请求的用户名和密码。
* 停止Wireshark数据包捕获，并在“显示筛选器规范”窗口中输入“ttp”，以便稍后在“数据包列表”窗口中只显示捕获的HTTP消息
* （注意：如果您无法在实时网络连接上运行Wireshark，您可以使用http-ethereal-trace-5包跟踪来回答下面的问题；请参见脚注2。此跟踪文件是在一台author抯计算机上执行上述步骤时收集的。）

现在让我们来检查Wireshark的输出。您可能想先阅读HTTP认证，方法是查看HTTP://frontier.userland.com/stories/storyReader$2159上关于TTP访问认证框架的易读材料

回答下列问题：

1. 服务器抯响应（状态代码和短语）对来自浏览器的初始HTTP GET消息的响应是什么？
2. 当您的浏览器第二次发送HTTP GET消息时，httpget消息中包含了什么新字段？

您输入的用户名（wireshark学生）和密码（网络）以字符串（d2lyzxnoyxjrlxn0dwrlbnzom5ldhdvcms=）编码，在客户端的HTTP GET消息的“Authorization:Basic”头之后。虽然您的用户名和密码可能看起来是加密的，它们只是以一种称为Base64格式的格式进行编码。用户名和密码未加密！要查看这个，请转到http://www.motobit.com/util/base64-decoder-encoder.asp，输入base64编码字符串d2lyzxnoyxjrlxn0dwrlbnz并解码！你已经从Base64编码转换成ASCII编码，因此应该看到你的用户名！要查看密码，请输入字符串Om5ldHdvcms=的其余部分，然后按decode键。因为任何人都可以下载Wireshark之类的工具，并嗅探通过其网络适配器的数据包（不只是他们自己的数据包），任何人都可以将Base64转换为ASCII（您刚刚做到了！），您应该清楚，除非采取其他措施，否则WWW站点上的简单密码是不安全的。

不要害怕！正如我们将在第8章中看到的，有一些方法可以使WWW访问更加安全。然而，我们显然需要一些超越基本HTTP身份验证框架的东西！

[[一]](" \l "_ftnref1" \o ")

[[二]](" \l "_ftnref2" \o ")下载zip文件http://gaia.cs.umass.edu/wireshark-labs/wireshark-traces.zip并解压缩文件http-ethereal-trace-1。这个zip文件中的跟踪是由Wireshark在一台作者的计算机上运行，同时执行Wireshark实验室中指定的步骤来收集的。下载跟踪后，您可以将其加载到Wireshark中，并使用“文件”下拉菜单查看跟踪，选择“打开”，然后选择http-ethereal-trace-1跟踪文件结果显示应该类似于图1。（Wireshark用户界面在不同的操作系统上和Wireshark的不同版本中显示的稍有不同）。

1. 参考图和章节编号均适用于课本 [↑](#footnote-ref-1)