暨南大学本科实验报告专用纸

课程名称 计算机网络实验 成绩评定_____

实验项目名称 超文本传输协议（HTTP） 指导老师 某某某

实验项目编号 8 实验项目类型 综合类 实验地点 N117

学生姓名 某某 学号 xxxxxxxx

学院 网络空间安全学院 系____ 专业 网络空间安全

实验时间 2020 年 12 月 2 日 晚 上～ 2020 年 12 月 2 日 晚 上

（一）实验目的

1. 掌握 HTTP 的报文格式
2. 掌握 HTTP 的工作原理
3. 掌握 HTTP 常用方法

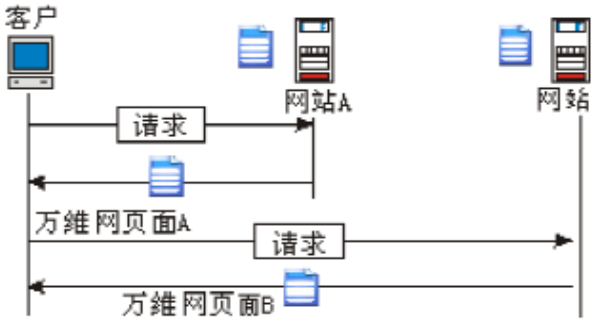
（二）实验环境

本次使用能访问广域网的主机

（三）实验原理

一、体系结构

万维网（WWW）服务是分布式的客户/服务器服务，客户使用浏览器能够得到服务器提供的服务，这种服务的提供是分布在许多网站中的，如下图所示::



每一个网站保存有一个或多个文档，叫做万维网页面。在图 13-1 所示的例子中，客户需要查看网站 A 的某些信息。浏览器用来读取万维网上的文档。它向网站 A 发送一个请求，这个请求包含了网站 A 和网站中万维网页面的地址，这个地址叫做统一资源定位符（URL）。网站 A 收到请求后，将指定的文档发送给这个客户。它用同样的方法与网站 B 通信。

1. **客户（浏览器）**

   许多厂商提供商用浏览器，可以解释和显示万维网页面。每一个浏览器通常由 3 个部分组成：控制程序、客户程序及解释程序。控制程序从键盘或鼠标接收输入，使用客户程序访问要浏览的文档。在文档找到后，控制程序就使用解释程序（可以是 HTML、Java 或 JavaScript 等）中的一个，把文档显示在屏幕上。

2. **服务器**

   万维网页面存储在服务器上。每当有客户请求到达时，对应的文档就发送给客户。为了提高效率，服务器通常在其高速缓存中存储被请求过的文档；对高速缓存进行访问要比磁盘快得多。通过多线程或多进程可使服务器的效率更加提高，服务器在同一时间可回答多个请求。

3. **统一资源定位符（URL）**

   客户要访问万维网页面就需要这个页面地址。为了方便地访问文档，HTTP 协议使用统一资源定位符（URL）。URL 是 Internet 上指定信息的标准。URL 由 4 部分组成：**协议、主机、端口和路径**。

   ● 协议：协议指定了用这个 URL 的客户/服务器程序。例如，HTTP 协议、FTP 协议和 TELNET 协议等。

   ● 主机：主机指明了信息所存放的地址，可以是逻辑地址也可以是相应的域名。存放万维网页面的计算机通常使用以字符“WWW”开始的域名，但这不是强制性的。

   ● 端口：端口指定了使用主机的某个端口，端口是可选的。如果包含了端口，那么端口就插入在主机和路径之间，和主机用冒号分隔开。

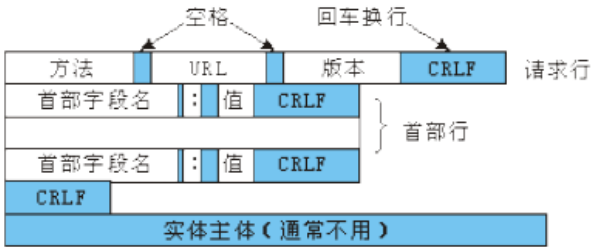
   ● 路径：路径指定了文件存放的位置。路径本身可以包含斜线，用于将目录与子目录和文件分隔开。

二. HTTP 协议简介

HTTP（超文本传输协议）主要用于访问万维网上的数据。协议以普通文本、超文本、音频、视频等格式传输数据。之所以称为超文本协议，原因是在应用环境中，它可以快速的在文档之间跳转。HTTP 在熟知端口 80 上使用 TCP 服务。

三. HTTP 报文格式

HTTP 报文有两种一般的类型：请求和响应。这两种类型的报文格式几乎是相同的。报文格式如下图所示：



四. HTTP 方法

HTTP 报文中的方法是客户端向服务器端发出的实际命令和请求。常用 HTTP 方法如下表所示：

方法	动作
GET	从服务器请求一个文档
HEAD	请求关于一个文档的信息，但不是这个文档本身
POST	从客户向服务器发送一些信息
PUT	从服务器向客户发送一些信息
TRACE	把到达的请求回送
CONNECT	用于使用代理连接
OPTION	询问关于可用的选项

五. HTTP 状态码

在响应报文中，请求行被替换为状态行，由 3 位数字组成，表示请求是否被理解或被满足。HTTP 协议的状态码如下表所示：

状态码	含义
200	OK。请求已成功。
302	Move temporarily请求的资源临时从不同的 URI响应请求。由于这样的重定向是临时的，客户端应当继续向原有地址发送以后的请求。
304	Not Modified文档的内容（自上次访问以来或者根据请求的条件）并没有改变。
400	Bad Request语义有误，当前请求无法被服务器理解。
401	Unauthorized当前请求需要用户验证。
403	Forbidden服务器收到请求，但是拒绝提供此服务。
404	Not Found请求资源不存在。
408	Request Timeout请求超时。
500	Internal Server Error服务器发生不可预知的错误。
503	Server Unavailable服务器当前不能处理客户端的请求，一段时间后可能恢复正常

六. 持久与非持久连接

在版本 1.1 以前的 HTTP 协议默认是非持久连接的，而在版本 1.1 中，持久连接是默认的连接。

1. 非持久连接

   在非持久连接中，对每一个请求/响应都要建立一次 TCP 连接。非持久连接的步骤如下：

   - 客户打开 TCP 连接，并发送请求。
   - 服务器发送响应，并关闭连接。
   - 客户读取数据，直到它遇到文件结束标记；然后它关闭连接。

   使用非持久连接时，对于在不同文件中的 N 个不同图片的请求，连接必须打开和关闭 N 次。非持久连接策略给服务器造成了很大的开销，因为服务器需要 N 个不同的缓存，而每次打开连接时都要使用慢开始过程。

2. 持久连接

   HTTP 版本 1.1 指明持久连接是默认的策略。在使用持久连接时，服务器在发送响应后，让连接继续为一些请求打开。服务器可以在客户发送关闭请求时等待或

关闭这个连接。发送端通常在每一个响应中都发送数据长度。但是，有时发送端并不知道数据的长度（例如动态文档或活动文档），这时服务器就把长度不知道这件事通知客户，并在发送数据后关闭这个连接，这样客户就知道数据结束的地方已经到了。

七. HTTP 代理服务器

HTTP 支持代理服务器（proxy server）。代理服务器保留对最近请求的响应的副本。在有代理服务器的情况下，HTTP 客户把请求发送给代理服务器。代理服务器检查它的高速缓存。如果在高速缓存中有这个响应，代理服务器就直接应答客户的请求；反之，如果在高速缓存中没有这个响应，代理服务器就把请求发送给相应的 HTTP 服务器，当 HTTP 服务器的响应到达代理服务器时，代理服务器将该响应转发给客户，同时将该响应存储到高速缓存中，以便以后为其它客户的请求使用。代理服务器减少了原始服务器的负荷，减小了通信量，也减小了延时。但是，由于使用了代理服务器，客户必须配置成接入到代理服务器而不是那个目标服务器。

（四）实验步骤

练习1 页面访问

各主机打开工具区的“拓扑验证工具”，选择相应的网络结构，配置网卡后，进行拓扑验证，如果通过拓扑验证，关闭工具继续进行实验，如果没有通过，请检查网络连接。本练习一人一组，现仅以主机 A 为例，其它主机参考主机 A 的操作。

1. 主机 A 清空 IE 缓存。
2. 主机 A 启动协议分析器开始捕获数据，并设置过滤条件（提取 HTTP 协议）。
3. 主机 A 启动 IE 浏览器，在“地址”框中输入 http://服务器的 ip/experiment，并连接，服务器的 ip 默认为 172.23.65.63。
4. 主机 A 停止捕获数据，分析捕获到的数据，并回答以下问题：
   - 本练习使用 HTTP 协议的哪种方法？简述这种方法的作用。

     Get 方法。客户要从服务器读取文档时使用。
   - 根据本练习的报文内容，填写下表。

主机名	172.23.65.63
URL	http://172.23.65.63/experiment
服务器类型	Apache/2.3.46
传输文本类型	text/html
访问时间	Wed, 12 Dec 2020 18:34:07

   - 参考“会话分析”视图显示结果，绘制此次访问过程的报文交互图（包括 TCP 协议）。(在附页处)
   - 简述 TCP 协议和 HTTP 协议之间的关系。

     HTTP 是基于 TCP 的应用层协议，HTTP 的数据包的运输使用了 TCP 协议。

思考问题：

1. 一个主页是否只有一个连接？

   一个主页可能对应多个连接。

练习2 页面提交

本练习一人一组，现仅以主机 A 为例，其它主机参考主机 A 的操作。

1. 主机 B 启动协议分析器开始捕获数据，并设置过滤条件（提取 HTTP 协议）。
2. 主机 A 启动 IE 浏览器，在“地址”框中输入“http://服务器的 ip/experiment/post.html”，并连接，服务器的 ip 默认为 172.23.65.63。在返回页面中，填写“用户名”和“密码”，点击［确定］按钮。
3. 主机 A 停止捕获数据，分析捕获到的数据，并回答以下问题：
   - 本练习的提交过程使用 HTTP 协议的哪种方法？简述这种方法的作用。

     POST 方法。当客户要给服务器提供某些信息时使用。
   - 此次通信分几个阶段？每个阶段完成什么工作？

     两个阶段：页面的访问和页面的提交；或四个阶段：连接服务器，发送 POST 报文，服务器返回信息，断开连接。
   - 参考“会话分析”视图显示结果，绘制此次提交过程的报文交互图（包括 TCP 协议）(**在附页处**)。

练习3 获取页面信息

本练习一人一组，现仅以主机 A 为例，其它主机参考主机 A 的操作。

1. 主机 A 启动实验平台工具栏中的“TCP 工具”。
2. 主机 A 启动协议分析器开始捕获数据，并设置过滤条件（提取 HTTP 协议）。
3. 主机 A 在“TCP 工具”上，选中“客户端”单选框，设置“IP 地址”为服务器 IP（默认为 172.16.0.253）；设置“端口”为 80；单击［连接］按钮来和服务器建立连接。
4. 主机 A 在“TCP 工具”上，设置“发送数据（文本）”为以下内容：

   HEAD /experiment/ HTTP/1.1<CRLF>

   Host: 172.16.0.253<CRLF>

   <CRLF>

   点击［发送］按钮。(注：<CRLF> 是回车换行)

   点击［断开］按钮，断开 TCP 连接（由于不同 http 版本所遵循的规范不同，有些 HTTP 服务器不需要断开操作）。
5. 主机 A 在“TCP 工具”上的“显示数据（文本）”中察看服务器返回信息。
6. 主机 A 停止捕获数据，分析捕获到的数据。

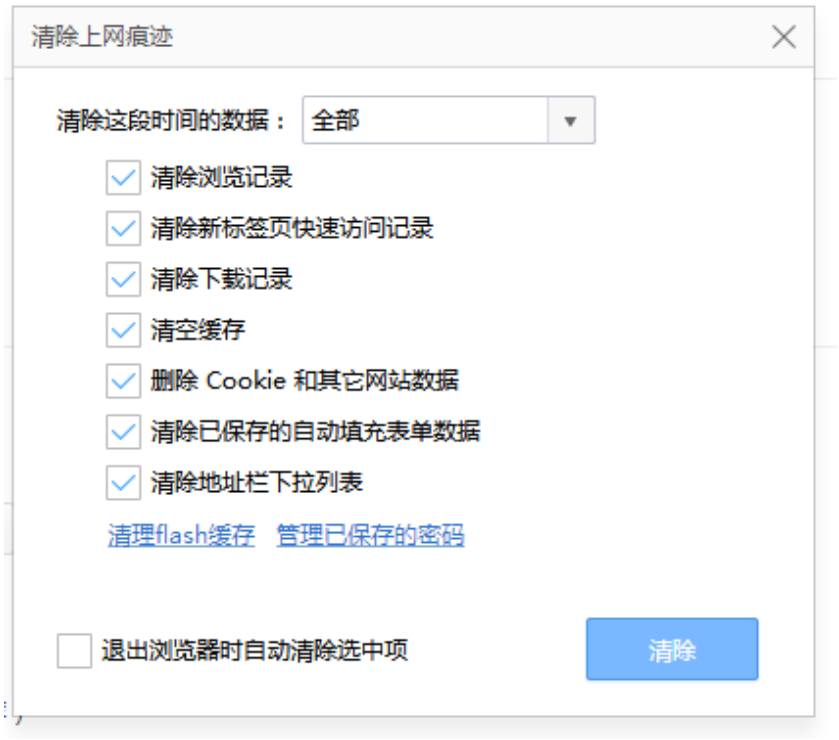
思考问题：

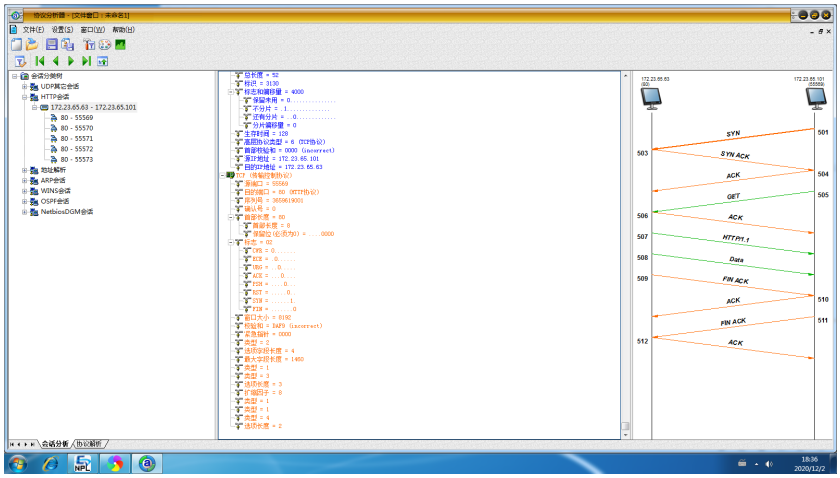
1. 同时打开多个浏览器窗口并访问一个 WEB 站点的不同页面时，系统是根据什么把返回的页面正确地显示到相应窗口的？

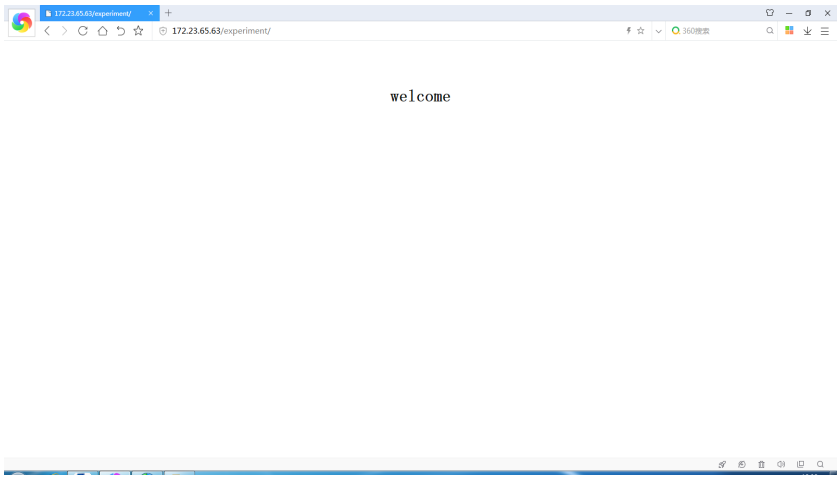
使用多个浏览器窗口访问一个 WEB 站点的不同页面时，每一个浏览器窗口可能对应一个或多个连接，每一个连接和数据报中的一个端口相对应，系统是根据这种对应关系把返回的页面正确地显示到相应窗口中。

（五）实验结果与分析

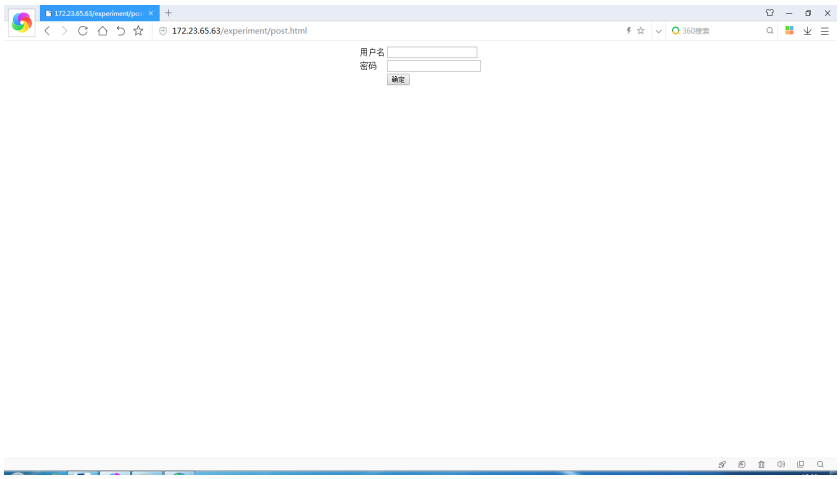
练习1 页面访问

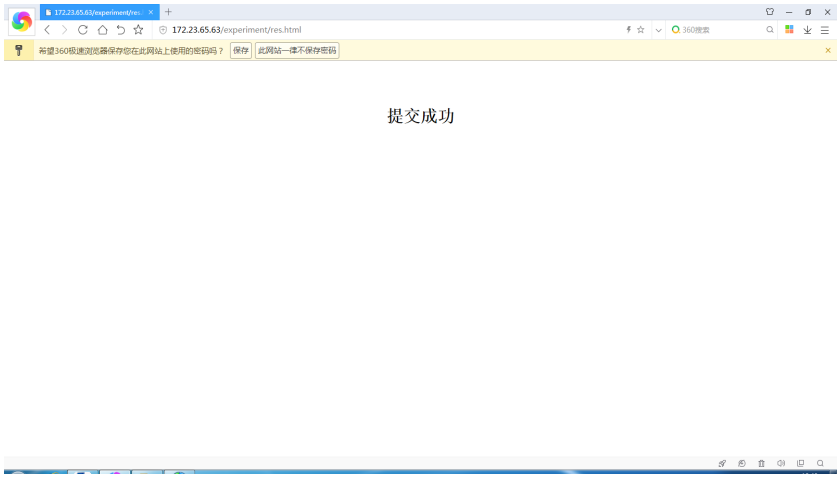


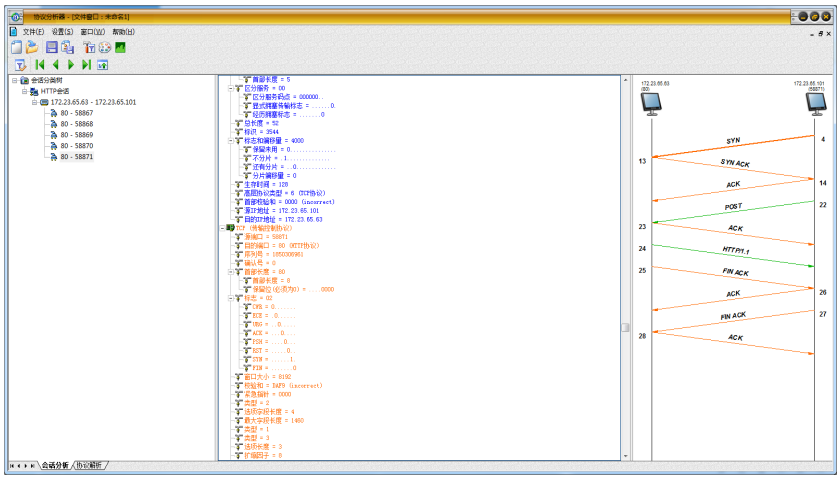




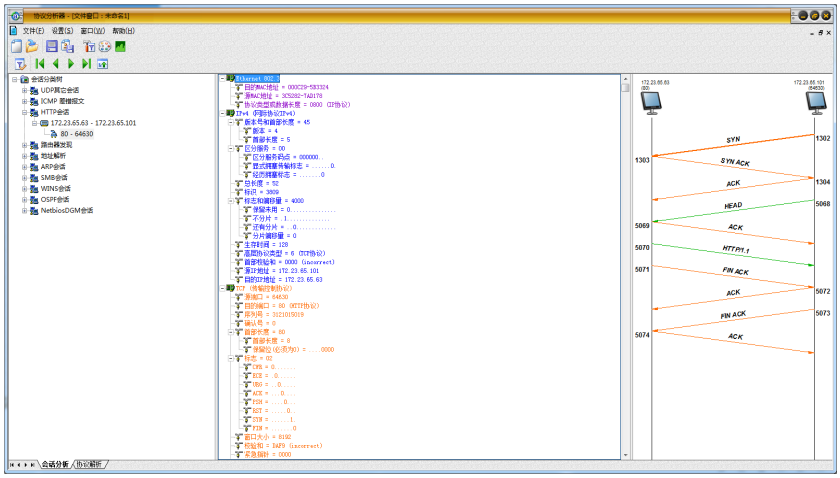
分析：清空浏览器缓存后访问网站，捕获到 HTTP 报文，从图中可以看出主机与服务器进行了 **“三次握手” 和 “四次挥手”**，HTTP 请求方法为 **GET**，HTTP 协议为 **HTTP/1.1**。

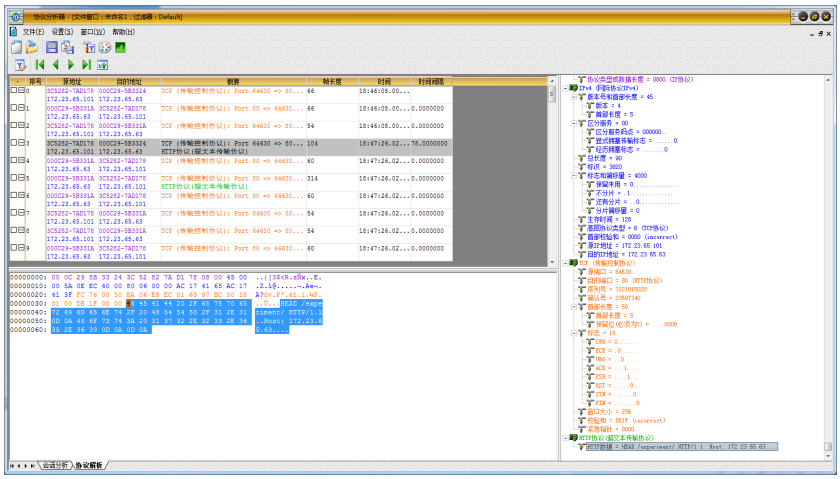






分析：再次清空浏览器缓存后访问网站，捕获到 HTTP 报文，从图中可以看出主机与服务器也是进行了 **“三次握手”** 和 **“四次挥手”**，HTTP 请求方法为 **POST**，HTTP 协议为 **HTTP/1.1**。

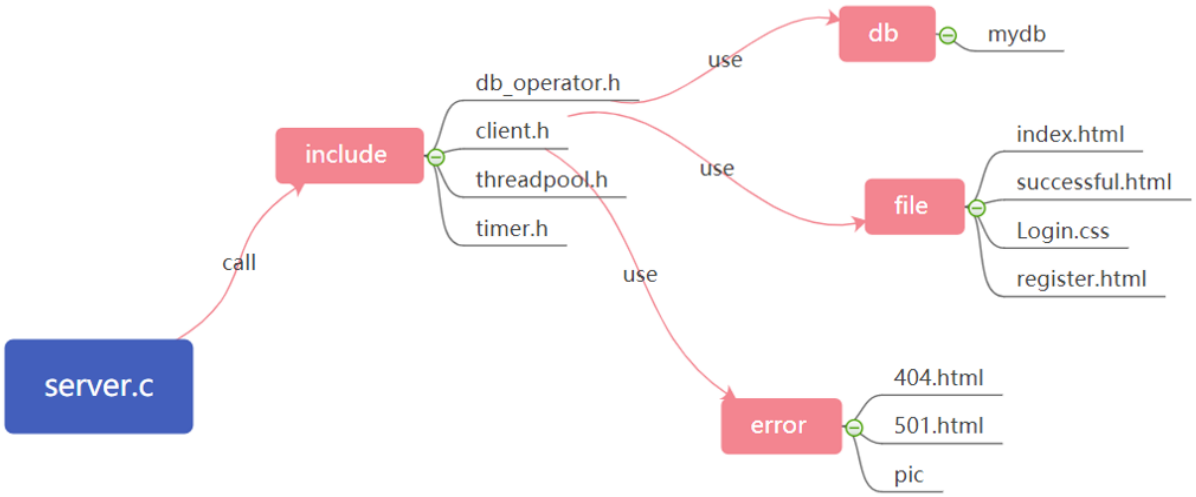




分析：用 TCP 工具先进行连接，再发送 HTTP 报文，响应的协议分析器捕捉到了 HTTP 报文，查看详细报文解析可以找到先前发送的报文。在 HTTP 报文中，空格和换行的规则必须样饿遵守，写网络通讯程序时空格一般用 “\r” 代替，换行一般用 “\n” 代替。

（六）附录

写一个 HTTP 服务器。其实 HTTP 服务器逻辑并不难，无非就是 HTTP 报文的解析以及响应。之前在准备项目的时候想写一个基于协程的服务器，为的是减少线程间加锁以及上下文切换造成的开销，不过还没写好。在此之前，倒是写过一个基于线程池的 HTTP 服务器，可处理 GET/POST 请求，egi 功能还没加进去。其大致框架如下：



主页面是一个登陆（GET 请求）界面，可注册（POST 请求），后端使用 sqlite3 实现用户数据的存储。使用 epoll ET 模式处理高并发连接，使用线程池减少线程切换开销。使用 Reactor 模式，主线程只接受连接，当有其他新的 IO 请求时分派一个新线程处理事务。

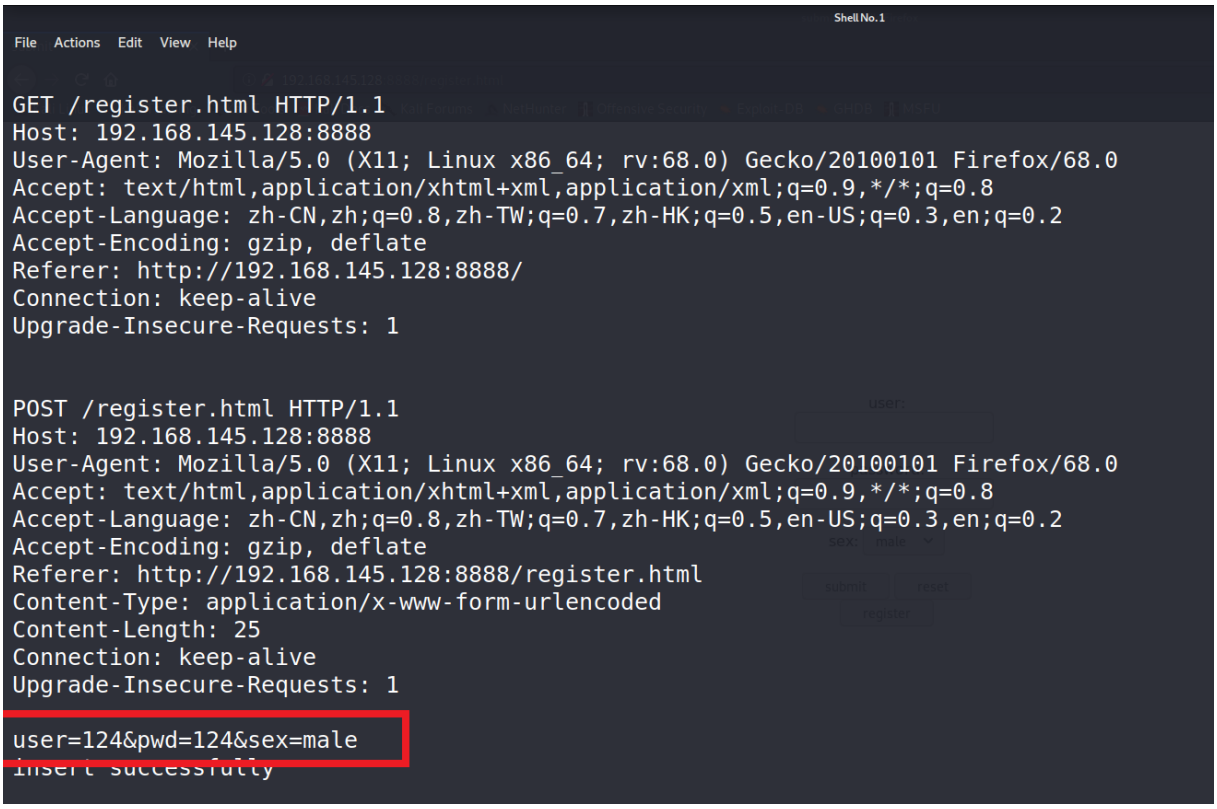


图 1: POST 请求会把数据写在实体主体

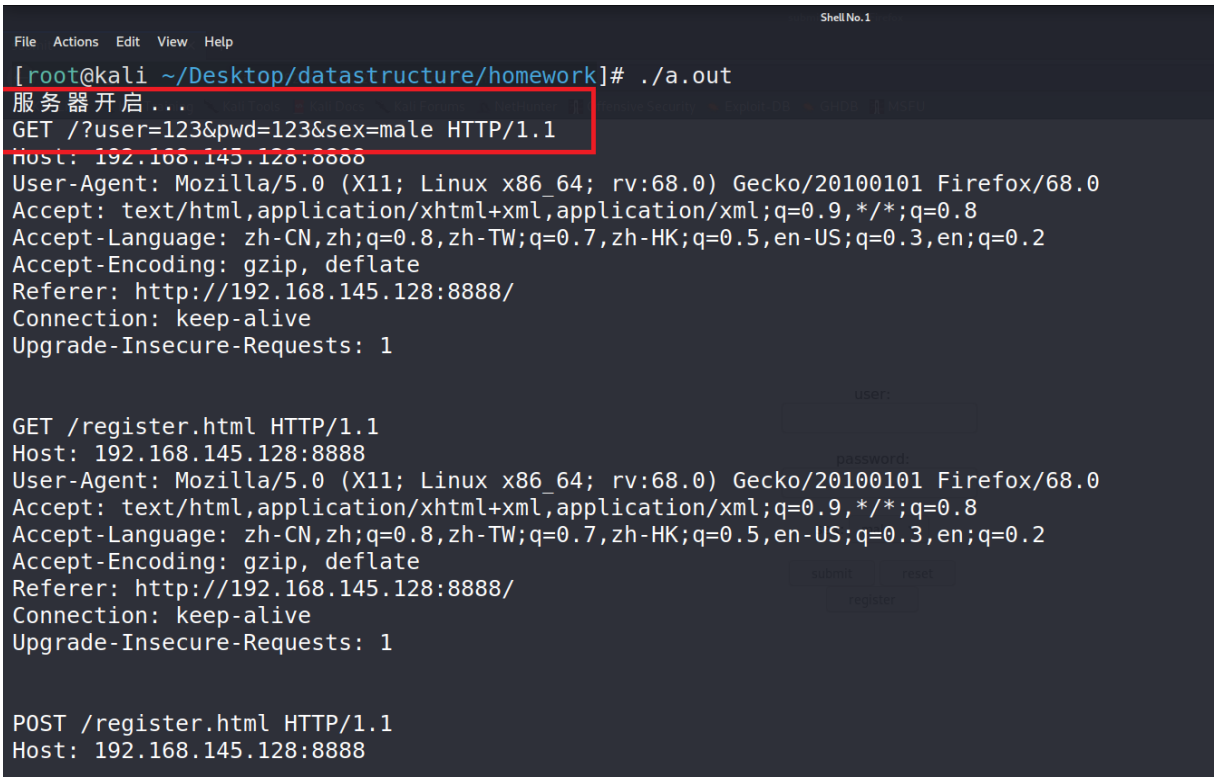


图 2: GET 请求会把数据写在请求报文首部的请求行