

浙江大学

本科实验报告

课程名称： 计算机网络基础

实验名称： 实现一个轻量级的 WEB 服务器

姓 名： 猜猜

学 院： 计算机学院

系： 数字媒体技术

专 业： 数字媒体技术

学 号： 猜猜

指导教师：

2019 年 12 月 23 日

浙江大学实验报告

实验名称：实现一个轻量级的 WEB 服务器 实验类型：编程实验

同组学生： 猜猜 实验地点： 计算机网络实验室

一、 实验目的

深入掌握 HTTP 协议规范，学习如何编写标准的互联网应用服务器。

二、 实验内容

- 服务程序能够正确解析 HTTP 协议，并传回所需的网页文件和图片文件
- 使用标准的浏览器，如 IE、Chrome 或者 Safari，输入服务程序的 URL 后，能够正常显示服务器上的网页文件和图片
- 服务端程序界面不做要求，使用命令行或最简单的窗体即可
- 功能要求如下：
 1. 服务程序运行后监听在 80 端口或者指定端口
 2. 接受浏览器的 TCP 连接（支持多个浏览器同时连接）
 3. 读取浏览器发送的数据，解析 HTTP 请求头部，找到感兴趣的部分
 4. 根据 HTTP 头部请求的文件路径，打开并读取服务器磁盘上的文件，以 HTTP 响应格式传回浏览器。要求按照文本、图片文件传送不同的 Content-Type，以便让浏览器能够正常显示。
 5. 分别使用单个纯文本、只包含文字的 HTML 文件、包含文字和图片的 HTML 文件进行测试，浏览器均能正常显示。
- 本实验可以在前一个 Socket 编程实验的基础上继续，也可以使用第三方封装好的 TCP 类进行网络数据的收发
- 本实验要求不使用任何封装 HTTP 接口的类库或组件，也不使用任何服务端脚本程序如 JSP、ASPX、PHP 等

三、 主要仪器设备

联网的 PC 机、Wireshark 软件、Visual Studio、gcc 或 Java 集成开发环境。

四、 操作方法与实验步骤

- 阅读 HTTP 协议相关标准文档，详细了解 HTTP 协议标准的细节，有必要的话使用 Wireshark 抓包，研究浏览器和 WEB 服务器之间的交互过程
- 创建一个文档目录，与服务器程序运行路径分开
- 准备一个纯文本文件，命名为 test.txt，存放在 txt 子目录下
- 准备好一个图片文件，命名为 logo.jpg，放在 img 子目录下
- 写一个 HTML 文件，命名为 test.html，放在 html 子目录下，主要内容为：

```

<html>
  <head><title>Test</title></head>
  <body>
    <h1>This is a test</h1>
    
    <form action="dopost" method="POST">
      Login:<input name="login">
      Pass:<input name="pass">
      <input type="submit" value="login">
    </form>
  </body>
</html>

```

- 将 test.html 复制为 noimg.html，并删除其中包含 img 的这一行。
- 服务端编写步骤（**需要采用多线程模式**）
 - a) 运行初始化，打开 Socket，监听在指定端口（**请使用学号的后 4 位作为服务器的监听端口**）
 - b) 主线程是一个循环，主要做的工作是等待客户端连接，如果有客户端连接成功，为该客户端创建处理子线程。该子线程的主要处理步骤是：
 1. 不断读取客户端发送过来的字节，并检查其中是否连续出现了 2 个回车换行符，如果未出现，继续接收；如果出现，按照 HTTP 格式解析第 1 行，分离出方法、文件和路径名，其他头部字段根据需要读取。

✧ 如果解析出来的方法是 GET

2. 根据解析出来的文件和路径名，读取响应的磁盘文件（该路径和服务端程序可能不在同一个目录下，需要转换成绝对路径）。如果文件不存在，第 3 步的响应消息的状态设置为 404，并且跳过第 5 步。
3. 准备好一个足够大的缓冲区，按照 HTTP 响应消息的格式先填入第 1 行（状态码=200），加上回车换行符。然后模仿 Wireshark 抓取的 HTTP 消息，填入必要的几行头部（需要哪些头部，请试验），其中不能缺少的 2 个头部是 Content-Type 和 Content-Length。Content-Type 的值要和文件类型相匹配（请通过抓包确定应该填什么），Content-Length 的值填写文件的字节大小。
4. 在头部行填完后，再填入 2 个回车换行
5. 将文件内容按顺序填入到缓冲区后面部分。

✧ 如果解析出来的方法是 POST

6. 检查解析出来的文件和路径名，如果不是 dopost，则设置响应消息的状态为 404，然后跳到第 9 步。如果是 dopost，则设置响应消息的状态为 200，并继续下一步。
7. 读取 2 个回车换行后面的体部内容（长度根据头部的 Content-Length 字段的指示），并提取出登录名（login）和密码（pass）的值。**如果登录名是你的学号，密码是学号的后 4 位，则将响应消息设置为登录成功，否则将响应消息设置为登录失败。**
8. 将响应消息封装成 html 格式，如

<html><body>响应消息内容</body></html>

9. 准备好一个足够大的缓冲区，按照 HTTP 响应消息的格式先填入第 1 行（根据前面的情况设置好状态码），加上回车换行符。然后填入必要的几行头部，其中不能缺少的 2 个头部是 Content-Type 和 Content-Length。Content-Type 的值设置为 text/html，如果状态码=200，则 Content-Length 的值填写响应消息的字节大小，并将响应消息填入缓冲区的后面部分，否则填写为 0。
 10. 最后一次性将缓冲区内的字节发送给客户端。
 11. 发送完毕后，关闭 socket，退出子线程。
- c) 主线程还负责检测退出指令（如用户按退出键或者收到退出信号），检测到后即通知并等待各子线程退出。最后关闭 Socket，主程序退出。
- 编程结束后，将服务器部署在一台机器上（本机也可以）。在服务器上分别放置纯文本文件（.txt）、只包含文字的测试 HTML 文件（将测试 HTML 文件中的包含 img 那一行去掉）、包含文字和图片的测试 HTML 文件（以及图片文件）各一个。
 - 确定好各个文件的 URL 地址，然后使用浏览器访问这些 URL 地址，如 <http://x.x.x.x:port/dir/a.html>，其中 port 是服务器的监听端口，dir 是提供给外部访问的路径，请设置为与文件实际存放路径不同，通过服务器内部映射转换。
 - 检查浏览器是否正常显示页面，如果有问题，查找原因，并修改，直至满足要求
 - 使用多个浏览器同时访问这些 URL 地址，检查并发性

五、实验数据记录和处理

请将以下内容和本实验报告一起打包成一个压缩文件上传：

- 源代码：需要说明编译环境和编译方法，如果不能编译成功，将影响评分
- 可执行文件：可运行的.exe 文件或 Linux 可执行文件

以下实验记录均需结合屏幕截图（截取源代码或运行结果），进行文字标注（看完请删除本句）。

- 服务器的主线程循环关键代码截图（解释总体处理逻辑，省略细节部分）

```
1. while (true) {
2.     printf("\nwaiting for connection...\n");
3.
4.     //接受一个连接
5.     clientSocket = accept(serverSocket, (sockaddr*)&serverChannel, &len);
6.
7.     if (clientSocket < 0) {
8.         printf("accept failed\n");
9.     }
10.    else {
```

```

11.     printf("successfully connect\n");
12.     memset(buffer, 0, sizeof(buffer));
13.
14.     int ret;
15.     ret = recv(clientSocket, buffer, BUFFER_SIZE, 0);
16.     if (ret == SOCKET_ERROR) {
17.         printf("sorry receive failed\n");
18.     }
19.     else if (ret == 0) {
20.         printf("the client socket is closed\n");
21.     }
22.     else {
23.         printf("successfully receive\n");
24.         for (int i = 0; i < MAX; i++) {
25.             if (!isActive[i]) {
26.                 isActive[i] = true;
27.                 message msg(buffer, &isActive[i], clientSocket, i);
28.                 std::thread temp(&handleMSG, std::ref(msg));
29.                 t[i] = &temp;
30.                 t[i]->join();
31.                 break;
32.             }
33.         }
34.     }
35. }
36. }

```

该循环通过阻塞调用 `accept` 进行监听客户端的消息，确认连接后开启子线程，接收客户端发送的消息，并对其进行处理。

- 服务器的客户端处理子线程关键代码截图（解释总体处理逻辑，省略细节部分）

```

1. void handleMSG(message msg)
2. {
3.     // 解析请求类型
4.     if (msg.data[0] == 'G') type = "GET";
5.     if (msg.data[0] == 'P') type = "POST";
6.     // 处理请求
7.     if (type == "POST") { //POST
8.         if (name == "2473" && passwd == "123") { //用户信息正确
9.             printf("Info: login succeeded!\n");
10.            int r = send(msg.clientSocket, msg.data, 10000, 0);
11.        }
12.        else { //用户信息错误

```

```

13.         int r = send(msg.clientSocket, msg.data, 10000, 0);
14.     }
15. }
16. else if (type == "GET" && data != "") {
17.     if (data.substr(0, 4) == "/hi/") {
18.         // 判断信息类型
19.         if (str == "txt") path = "catalog/txt/" + data.substr(4);
20.         else if (str == "html") path = "catalog/html/" + data.substr(4)
21.         ;
22.         sendMessage(path, msg);
23.     }
24.     else if (data.substr(0, 5) == "/img/") { //处理图片
25.         std::string path = "catalog/img/" + data.substr(5);
26.         sendMessage(path, msg);
27.     }
28. }

```

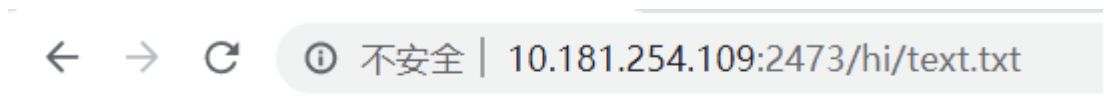
服务器通过解析客户端发来的信息，确定请求类型是 GET 请求还是 POST 请求，根据请求类型分情况（GET/POST）来处理信息。

- 服务器运行后，用 netstat -an 显示服务器的监听端口

TCP	0.0.0.0:1029	0.0.0.0:0	LISTENING
TCP	0.0.0.0:2473	0.0.0.0:0	LISTENING
TCP	0.0.0.0:5040	0.0.0.0:0	LISTENING
TCP	0.0.0.0:5357	0.0.0.0:0	LISTENING
TCP	0.0.0.0:6664	0.0.0.0:0	LISTENING

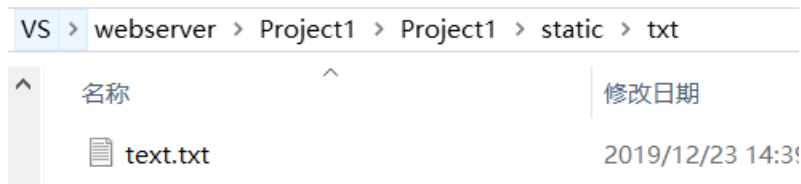
使用 netstat -an 命令后，可以看到服务器的监听端口（端口号为 2473），其目前处于 listen 状态。

- 浏览器访问纯文本文件（.txt）时，浏览器的 URL 地址和显示内容截图。



Hello world!

服务器上文件实际存放的路径：



服务器的相关代码片段：

```
1. if (str == "txt") {  
2.     path = "static/txt/" + data.substr(4);  
3. }
```

Wireshark 抓取的数据包截图（通过跟踪 TCP 流，只截取 HTTP 协议部分）：



No.	Time	Source	Destination	Protocol	Length	Info
8799	357.578130	117.144.244.59	10.181.254.109	HTTP	1300	HTTP/1.1 500 Internal Server Error (text/html)
8801	357.578136	10.181.171.165	10.181.254.109	HTTP	448	GET /img/logo.jpg HTTP/1.1
8803	357.584553	10.181.254.109	10.181.171.165	HTTP	3193	HTTP/1.1 200 OK (text/html)
8814	357.923785	10.181.171.165	10.181.254.109	HTTP	447	GET /favicon.ico HTTP/1.1
8973	377.821409	10.181.171.165	10.181.254.109	HTTP	512	GET /hi/text.txt HTTP/1.1
8975	377.832007	10.181.254.109	10.181.171.165	HTTP	66	HTTP/1.1 200 OK (text/html)
8988	378.317170	10.181.171.165	10.181.254.109	HTTP	446	GET /favicon.ico HTTP/1.1
9004	378.399913	10.181.171.165	10.181.254.109	HTTP	446	GET /favicon.ico HTTP/1.1
9042	382.008650	10.181.254.109	140.207.234.36	HTTP	671	GET /gchatpic_new/58666644D3B845E472B7ECFFB7ED159D691F06CD4423B
9141	382.152638	140.207.234.36	10.181.254.109	HTTP	159	HTTP/1.1 200 OK (image/jpeg)
9214	386.922015	10.181.254.109	140.207.234.21	HTTP	671	GET /gchatpic_new/D5FC9FAC6FB80A7DD2D1CDE8FE26082B74FEDD6FA8C5A
9265	387.015527	140.207.234.21	10.181.254.109	HTTP	85	HTTP/1.1 200 OK (image/png)

> Frame 8973: 512 bytes on wire (4096 bits), 512 bytes captured (4096 bits) on interface 0
> Ethernet II, Src: JuniperN 67:28:52 (88:e0:f3:67:28:52), Dst: IntelCor_40:36:98 (b4:6b:fc:40:36:98)
> Internet Protocol Version 4, Src: 10.181.171.165, Dst: 10.181.254.109
> Transmission Control Protocol, Src Port: 8159, Dst Port: 2473, Seq: 1, Ack: 1, Len: 458
▼ Hypertext Transfer Protocol
> GET /hi/text.txt HTTP/1.1\r\nHost: 10.181.254.109:2473\r\nConnection: keep-alive\r\nUpgrade-Insecure-Requests: 1\r\nUser-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/79.0.3945.88 Safari/537.36\r\nAccept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,image/apng,*/*;q=0.8,application/signed-exchange;v=b3;q=0.9\r\nAccept-Encoding: gzip, deflate\r\nAccept-Language: zh-CN,zh;q=0.9,en-US;q=0.8,en;q=0.7\r\n\r\n[Full request URI: http://10.181.254.109:2473/hi/text.txt]
[HTTP request 1/1]
[Response in frame: 8975]

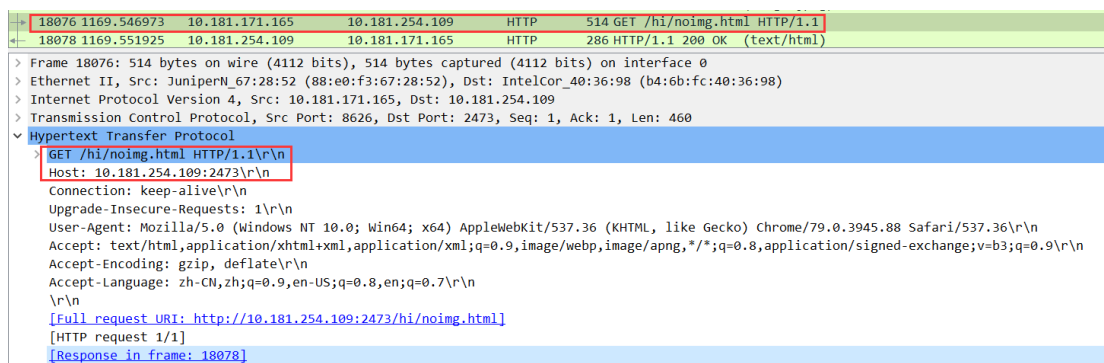
- 浏览器访问只包含文本的 HTML 文件时，浏览器的 URL 地址和显示内容截图。



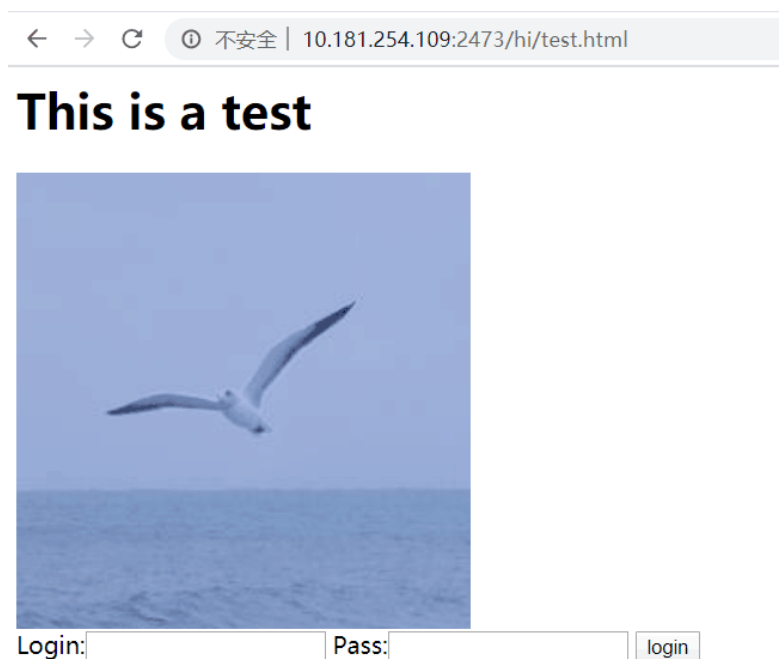
服务器文件实际存放的路径：

webserver > Project1 > Project1 > static > html	
名称	修改日期
 noimg.html	2018/11/14 15
 test.html	2019/12/23 17



Wireshark 抓取的数据包截图（只截取 HTTP 协议部分，包括 HTML 内容）：



- 浏览器访问包含文本、图片的 HTML 文件时，浏览器的 URL 地址和显示内容截图。



服务器上文件实际存放的路径：

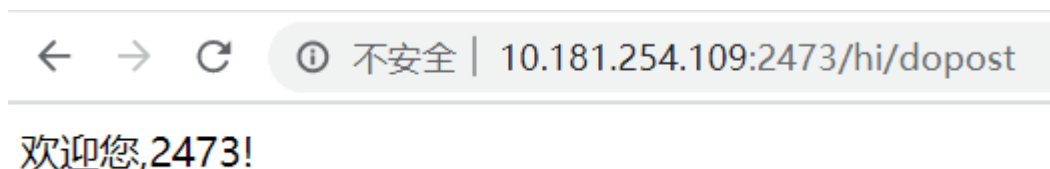
> webserver > Project1 > Project1 > static > html	
名称	修改日期
 noimg.html	2018/11/14 15
 test.html	2019/12/23 17

Wireshark 抓取的数据包截图（只截取 HTTP 协议部分，包括 HTML、图片文件的部分内容）：

18290	1228.602555	10.181.171.165	10.181.254.109	HTTP	513 GET /hi/test.html HTTP/1.1
18293	1228.606992	10.181.254.109	10.181.171.165	HTTP	333 HTTP/1.1 200 OK (text/html)
18296	1228.613992	10.181.254.109	10.181.171.165	HTTP	333 HTTP/1.1 200 OK (text/html)
18318	1229.737836	10.181.171.165	10.181.254.109	HTTP	448 GET /img/logo.jpg HTTP/1.1
18320	1229.741196	10.181.254.109	10.181.171.165	HTTP	3193 HTTP/1.1 200 OK (text/html)
18328	1229.810389	10.181.171.165	10.181.254.109	HTTP	447 GET /favicon.ico HTTP/1.1
18374	1242.154886	10.181.254.109	140.207.234.35	HTTP	671 GET /gchatpic_new/250031887357F3256E8A77E8923E61C1A630E5CCB23546/

> Frame 18290: 513 bytes on wire (4104 bits), 513 bytes captured (4104 bits) on interface 0
 > Ethernet II, Src: JuniperN_67:28:52 (88:e0:f3:67:28:52), Dst: IntelCor_40:36:98 (b4:6b:fc:40:36:98)
 > Internet Protocol Version 4, Src: 10.181.171.165, Dst: 10.181.254.109
 > Transmission Control Protocol, Src Port: 8678, Dst Port: 2473, Seq: 1, Ack: 1, Len: 459
 > Hypertext Transfer Protocol
 > GET /hi/test.html HTTP/1.1
 Host: 10.181.254.109:2473
 Connection: keep-alive
 Upgrade-Insecure-Requests: 1
 User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/79.0.3945.88 Safari/537.36
 Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,image/apng,*/*;q=0.8,application/signed-exchange;v=b3;q=0.9
 Accept-Encoding: gzip, deflate
 Accept-Language: zh-CN,zh;q=0.9,en-US;q=0.8,en;q=0.7
 [Full request URI: http://10.181.254.109:2473/hi/test.html]
 [HTTP request 1/1]
 [Response in frame: 18293]

- 浏览器输入正确的登录名或密码，点击登录按钮（login）后的显示截图。



服务器相关处理代码片段：

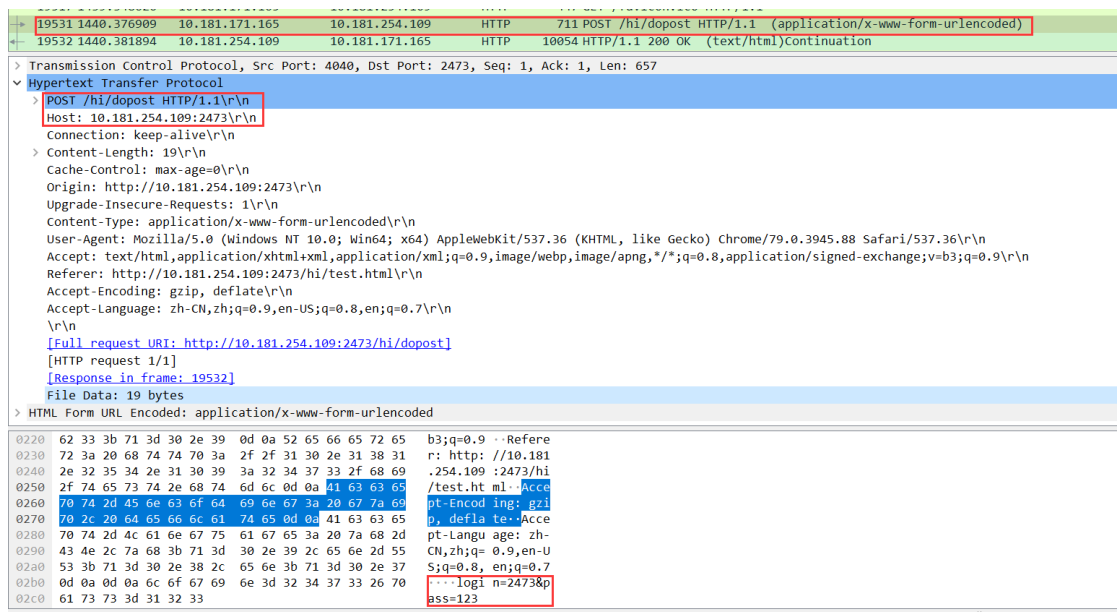
```
1. if (name == "2473" && passwd == "123") {
2.     char response[200];
3.     strcpy(response, "<html><body>欢迎您,");
4.     strcat(response, name.c_str());
5.     strcat(response, "!</body></html>\n");
6.     int len = strlen(response);
7.     char length[20];
8.     sprintf(length, "%d", len);
```

```

9.     strcpy(msg.data, "HTTP/1.1 200 OK\n");
10.    strcat(msg.data, "Content-Type: text/html;charset=gb2312\nContent-Length
    : ");
11.    strcat(msg.data, length);
12.    strcat(msg.data, "\n\n");
13.    strcat(msg.data, response);
14.    printf("Info: login succeeded!\n");
15.    int r = send(msg.clientSocket, msg.data, 10000, 0);
16.    if (r == SOCKET_ERROR) {
17.        printf("send failed\n");
18.        *msg.isActive = false;
19.        return;
20.    }
21.    printf("send succeeded\n");
22.    *msg.isActive = false;
23.    return;
24. }

```

Wireshark 抓取的数据包截图（HTTP 协议部分）

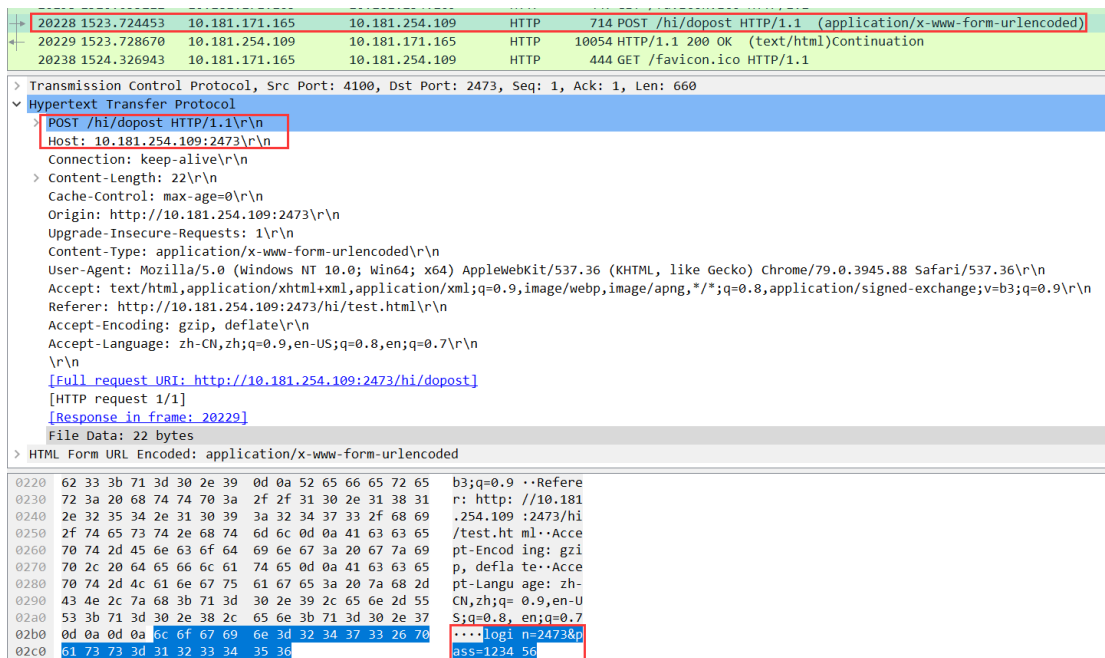


- 浏览器输入错误的登录名或密码，点击登录按钮（login）后的显示截图。

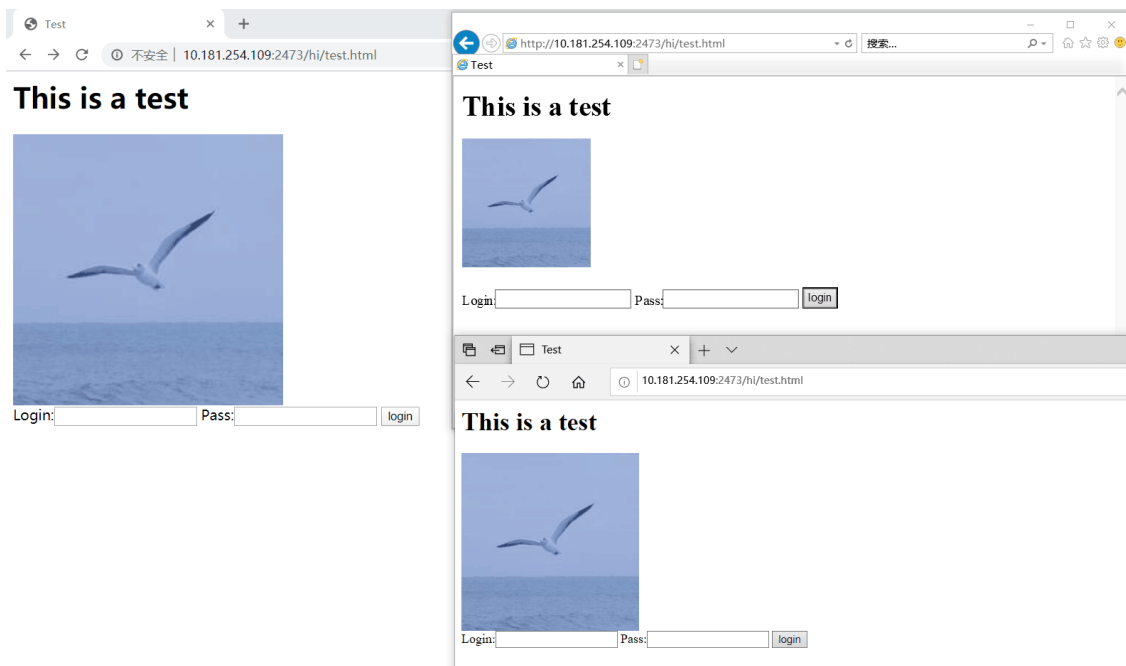


登录失败

- Wireshark 抓取的数据包截图（HTTP 协议部分）



- 多个浏览器同时访问包含图片的 HTML 文件时，浏览器的显示内容截图（将浏览器窗口缩小并列）



- 多个浏览器同时访问包含图片的 HTML 文件时,使用 `netstat -an` 显示服务器的 TCP 连接（截取与服务器监听端口相关的）

TCP	10.181.254.109:2473	10.181.171.165:4000	CLOSE_WAIT
TCP	10.181.254.109:2473	10.181.171.165:4201	TIME_WAIT
TCP	10.181.254.109:2473	10.181.171.165:4202	TIME_WAIT
TCP	10.181.254.109:2473	10.181.171.165:4250	TIME_WAIT
TCP	10.181.254.109:2473	10.181.171.165:4251	TIME_WAIT
TCP	10.181.254.109:2473	10.181.171.165:4252	TIME_WAIT
TCP	10.181.254.109:2473	10.181.171.165:4259	TIME_WAIT
TCP	10.181.254.109:2473	10.181.171.165:4260	TIME_WAIT
TCP	10.181.254.109:2473	10.181.171.165:4261	TIME_WAIT
TCP	10.181.254.109:2473	10.181.171.165:4266	TIME_WAIT
TCP	10.181.254.109:2473	10.181.171.165:4267	TIME_WAIT
TCP	10.181.254.109:2473	10.181.171.165:4269	TIME_WAIT
TCP	10.181.254.109:2473	10.181.171.165:8286	CLOSE_WAIT
TCP	10.181.254.109:2473	10.181.171.165:8289	CLOSE_WAIT
TCP	10.181.254.109:2473	10.181.171.165:8727	CLOSE_WAIT
TCP	10.181.254.109:2473	10.181.254.109:4976	CLOSE_WAIT

六、 实验结果与分析

根据你编写的程序运行效果，分别解答以下问题（看完请删除本句）：

- HTTP 协议是怎样对头部和体部进行分隔的？

通过一个空行来进行分隔。HTTP 报文的头部后紧跟一个空行，发送回车符和换行符，通知服务器接下来的是体部信息。

- 浏览器是根据文件的扩展名还是根据头部的哪个字段判断文件类型的？

根据头部的 Content-Type 字段内容来判断文件类型，比如“text/html”表示 html 文件。

- HTTP 协议的头部是不是一定是文本格式？体部呢？

是的。HTTP 消息是由普通 ASCII 文本组成的，包括头部和体部两个部分，头部是文本格式，体部可能不是文本格式。

- POST 方法传递的数据是放在头部还是体部？两个字段是用什么符号连接起来的？

体部。两个字段用&符号连接。

七、 讨论、心得

在实验过程中，最棘手的部分就是对 HTTP 报文的处理。解析报文的思路很简单，但落实到代码上却需要注意很多细节，尤其是 C++ 的字符串不支持 `split` 函数，导致想要截取需要的信息变得非常复杂和麻烦。

在读取服务器端的文件内容，打包发给浏览器端的过程中，遇到了一个困难：`txt` 文件和 `HTML` 文件都能正常显示，但图像却无法正确显示。打开浏览器的控制台开发工具，检测到在读取 `img` 时有一条报错信息 `net::ERR_CONTENT_LENGTH_MISMATCH`，也就是 HTTP 头部的 `Content-type` 字段和实际上的体部数据长度不匹配。在反复 `Debug` 后发现有 `strlen` 函数出现了莫名其妙的错误，没有解析出 `buffer` 的数据长度，修正后可以正常显示图像。