

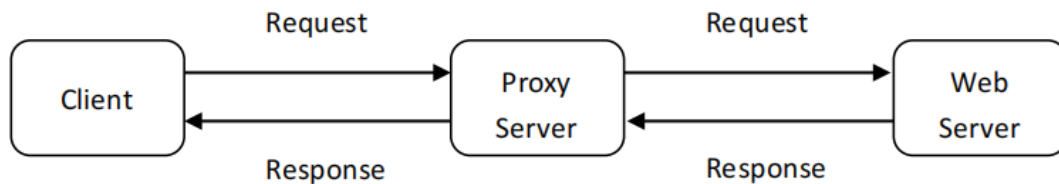
Lab 5:HTTP Web Proxy Server

191180048 黄奥成 通信工程

Attention: 此报告是由markdown编辑, 导出的pdf会有排版错误, 请尽量阅读markdown版本

一、实验原理与准备

在个编程作业中, 研发一个的Web代理服务器。当你的代理服务器从一个浏览器收到某对象的HTTP请求, 它生成对相同对象的一个新HTTP请求并向初始服务器发送。当该代理从初始服务器接收到具有该对象的HTTP响应时, 它生成一个包括该对象的新HTTP响应, 并发送给该客户。这个代理将是多线程的, 使其在相同时间能够处理多个请求。



由于之前科学上网时挂代理设置过, 因此在做实验前先备份设置, 然后设置实验环境

使用代理服务器

☒ 开

地址

127.0.0.1

端口

7890

请勿对以下列条目开头的地址使用代理服务器。若有多个条目，请使用英文分号 (;) 来分隔。

localhost;127.*;10.*;172.16.*;172.17.*;
172.18.*;172.19.*;172.20.*;172.21.*;
172.22.*;172.23.*;172.24.*;172.25.*

☒ 请勿将代理服务器用于本地(Intranet)地址

使用代理服务器

☒ 开

地址

127.0.0.1

端口

7890

请勿对以下列条目开头的地址使用代理服务器。若有多个条目，请使用英文分号 (;) 来分隔。

localhost

☐ 请勿将代理服务器用于本地(Intranet)地址

保存

二、实验过程与结果分析

1、输出

首先把实验结果挂出来，接下来结合代码与输出来分析

```
Output exceeds the size limit. Open the full output data in a text editor
Ready to serve...
Received a connection from: ('127.0.0.1', 4704)
b'GET http://gaia.cs.umass.edu/wireshark-labs/INTRO-wireshark-file1.html HTTP/1.1\r\nHost: gaia.cs.umass.edu\r\nProxy-Connection: keep-alive\r\nCache-Control: max-age=0\r\nUpgrade-Insecure-Requests: 1\r\nUser-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/99.0.4844.74 Safari/537.36\r\nAccept: text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,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\r\nIf-None-Match: "51-5daa0139d2242"\r\nIf-Modified-Since: Sun, 20 Mar 2022 05:59:02 GMT\r\n\r\n'
b'http://gaia.cs.umass.edu/wireshark-labs/INTRO-wireshark-file1.html'
filename = /gaia.cs.umass.edu/wireshark-labs/INTRO-wireshark-file1.html
filetouse = gaia.cs.umass.edu_wireshark-labs_INTRO-wireshark-file1.html
hostn = /gaia.cs.umass.edu/wireshark-labs/INTRO-wireshark-file1.html
('gaia.cs.umass.edu', 80)
Socket connected to port 80 of the host
Send the message successfully!
response = HTTP/1.1 304 Not Modified
Date: Sun, 20 Mar 2022 12:18:02 GMT
Server: Apache/2.4.6 (CentOS) OpenSSL/1.0.2k-fips PHP/7.4.27 mod_perl/2.0.11 Perl/v5.16.3
Etag: "51-5daa0139d2242"
open the file!
```

```
Ready to serve...
Received a connection from: ('127.0.0.1', 4705)
b'GET http://gaia.cs.umass.edu/wireshark-labs/INTRO-wireshark-file1.html HTTP/1.1\r\nHost: gaia.cs.umass.edu\r\nProxy-Connection: keep-alive\r\nCache-Control: max-age=0\r\nUpgrade-Insecure-Requests: 1\r\nUser-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/99.0.4844.74 Safari/537.36\r\nAccept: text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,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\r\nIf-None-Match: "51-5daa0139d2242"\r\nIf-Modified-Since: Sun, 20 Mar 2022 05:59:02 GMT\r\n\r\n'
b'http://gaia.cs.umass.edu/wireshark-labs/INTRO-wireshark-file1.html'
filename = /gaia.cs.umass.edu/wireshark-labs/INTRO-wireshark-file1.html
filetouse = gaia.cs.umass.edu_wireshark-labs_INTRO-wireshark-file1.html
fileexist = true
...
fileexist = true
Read from cache
Ready to serve...
Received a connection from: ('127.0.0.1', 9319)
```

2、服务端socket连接与接收报文

创建一个服务端socket用于接收报文

```
1 tcpSerSock = socket(AF_INET, SOCK_STREAM)
2 # Fill in start.
3 serverPort = 7890
4 tcpSerSock.bind(('', serverPort))
5 tcpSerSock.listen(1)
6 # Fill in end.
7 while 1:
8     # Start receiving data from the client
9     print('Ready to serve...')
10    tcpCliSock, addr = tcpSerSock.accept()
11    print('Received a connection from:', addr)
12    message = tcpCliSock.recv(4096)
13    print(message)
14
```

```
Received a connection from: ('127.0.0.1', 4704)
b'GET http://gaia.cs.umass.edu/wireshark-labs/INTRO-wireshark-file1.html HTTP/1.1\r\nHost: gaia.cs.umass.edu\r\nProxy-Connection: keep-alive\r\nCache-Control: max-age=0\r\nUpgrade-Insecure-Requests: 1\r\nUser-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/99.0.4844.74 Safari/537.36\r\nAccept: text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,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\r\nIf-None-Match: "51-5daa0139d2242"\r\nIf-Modified-Since: Sun, 20 Mar 2022 05:59:02 GMT\r\n\r\n'
```

3、将报文中的地址提取并处理

由于windows文件格式中不能出现'/'符号，因此将其替换为'_'。也即 `filetouse`

```
1 # Extract the filename from the given message
2 print(message.split()[1])
3 filename = message.split()[1].decode().partition("/")[2]
4 filetouse = message.split()[1].decode().partition("//")[2].replace('/',
5 '_')
6 print('filename = ' + filename)
7 fileExist = "false"
8 print('filetouse = ' + filetouse)
```

```
b'http://gaia.cs.umass.edu/wireshark-labs/INTRO-wireshark-file1.html'
filename = /gaia.cs.umass.edu/wireshark-labs/INTRO-wireshark-file1.html
filetouse = gaia.cs.umass.edu_wireshark-labs_INTRO-wireshark-file1.html
hostn = /gaia.cs.umass.edu/wireshark-labs/INTRO-wireshark-file1.html
```

4、尝试打开缓存

若open成功，则具有此网页的缓存。若失败，跳转到 `IOError`

```
1  try:# Check whether the file exist in the cache
2      f = open(filetouse[0:], "r")
3      outputdata = f.readlines()
4      print("fileexist = true")
5      fileExist = "true"
6      # ProxyServer finds a cache hit and generates a response message
7      #tcpCliSock.send("HTTP/1.0 200 OK\r\n")
8      #tcpCliSock.send("Content-Type:text/html\r\n")
9      # Fill in start
10     for i in range(0 , len(outputdata)):
11         tcpCliSock.send(outputdata[i].encode())
12     tcpCliSock.send('\r\n'.encode())
13     tcpCliSock.close()
14
15     # Fill in end.
16     print('Read from cache')
17     # Error handling for file not found in cache
```

```
Ready to serve...
Received a connection from: ('127.0.0.1', 4705)
b'GET http://gaia.cs.umass.edu/wireshark-labs/INTRO-wireshark-file1.html HTTP/1.1\r\nHost: gaia.cs.umass.edu\r\nProxy-Connection: keep-alive\r\nCache-Control: max-age=0\r\nUpgrade-Insecure-Requests: 1\r\nUser-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/99.0.4844.74 Safari/537.36\r\nAccept: text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,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\r\nIf-None-Match: "51-5daa0139d2242"\r\nIf-Modified-Since: Sun, 20 Mar 2022 05:59:02 GMT\r\n\r\n'
b'http://gaia.cs.umass.edu/wireshark-labs/INTRO-wireshark-file1.html'
filename = /gaia.cs.umass.edu/wireshark-labs/INTRO-wireshark-file1.html
filetouse = gaia.cs.umass.edu_wireshark-labs_INTRO-wireshark-file1.html
fileexist = true
...
fileexist = true
Read from cache
Ready to serve...
Received a connection from: ('127.0.0.1', 9319)
```

5、缓存网页

创建一个 `clientsocket` 用于发送报文（转发报文）。连接80端口并且将之前接收的请求报文发送给"真正"的服务端。然后接收其返回的报文 `response`。将其写入以 `filetouse` 为文件名的文件中

```
1  except IOError:
2      if fileExist == "false":
3          # Create a socket on the proxyserver
4          tcpCacheSock = socket(AF_INET, SOCK_STREAM)
5          hostn = filename.replace("www.", "", 1)
6          print('hostn = ' + hostn)
7
8          try:
9              # Connect to the socket to port 80
10             tcpCacheSockPort = 80
11             serverName = hostn.split("/")[1]
12             print((serverName, tcpCacheSockPort))
13             tcpCacheSock.connect((serverName, tcpCacheSockPort))
14             print('Socket connected to port 80 of the host')
```

```

15         tcpCacheSock.send(message)
16         print('Send the message sucessfully!')
17         #获取服务器返回内容
18         response = b''
19         rec = tcpCacheSock.recv(4096)
20         while rec:
21             response += rec
22             rec = tcpCacheSock.recv(4096)
23
24         print('response = ' + response.decode())
25         tmpFile = open("./" + filetype, "w")
26         print('open the file!')
27         tmpFile.writelines(response.decode().replace('\r\n', '\n'))
28         tmpFile.close()
29         #sys.exit()
30
31     except:
32         print("Illegal request")
33     else:
34         # HTTP response message for file not found
35         print('ERROR!!')
36         # Close the client and the server sockets
37     tcpCliSock.close()
38     tcpSersSock.close()
39

```

三、option experience

1、404 NOT FOUND

若网页不存在，则返回的报文中使用 `split` 分离后的字节流为 `404`。

```

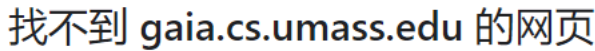
Socket connected to port 80 of the host
...
b'HTTP/1.1 404 Not Found\r\nDate: Mon, 21 Mar 2022 01:14:21 GMT\r\nServer: Apache/2.4.6 (CentOS) OpenSSL/1.0.2k-fips PHP/7.4.27 mod_perl/2.0.11 Perl/v5.16.3\r\nContent-Length: 238\r\nContent-Type: text/html; charset=iso-8859-1\r\n\r\n<!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML 2.0//EN">\n<html>\n<head>\n<title>404 Not Found</title>\n</head>\n<body>\n<h1>Not Found</h1>\n<p>The requested URL /wireshark-lab/INTRO-wireshark-file1.html was not found on this server.</p>\n</body>\n</html>\n'
404

```

```

1
2 response = b''
3 rec = tcpCacheSock.recv(4096)
4 while rec:
5     response += rec
6     rec = tcpCacheSock.recv(4096)
7
8 if response.split()[0] == b'404':
9     print('404')
10    tcpCliSock.send("HTTP/1.1 404 Not Found\r\n\r\n".encode())
11    tcpCliSock.close()
12    continue
13
14 tmpFile = open("./" + filetype, "w")
15 tmpFile.writelines(response.decode().replace('\r\n', '\n'))
16

```



HTTP ERROR 404

2、POST方法

四、Something interesting

由于我做实验的机器为台式机（后称PC）且其始终连接校园网（网线）为内网，因而很容易的在校内各处访问我的PC。这里使用我的笔记本尝试访问！

很不幸的是，我的笔记本与PC发送的报文的编码方式不同

由于时间原因（ddl貌似已经到了），下次实验报告会补全这部分内容

2、转发至公网

使用sakura frp 转发至公网。

```
2022/03/21 09:29:58 I Tunnel/J6CXFKXC [ef10f0db] 限速已更新: 10 Mibit/s
Tunnel/J6CXFKXC TCP 类型隧道启动成功
Tunnel/J6CXFKXC 使用 [cn-cd-dx-7.natfrp.cloud:40361] 来连接到你的隧道
Tunnel/J6CXFKXC 或使用 IP 地址连接 (不推荐) : [218.89.171.148:40361]
2022/03/21 09:29:59 I Tunnel/J6CXFKXC [ef10f0db] [vh1**3th.J6CXFKXC] 隧道启动成功
2022/03/21 09:30:15 I Service/RemoteManager 远程管理已连接
```

1. **TypeError:** a bytes-like object is required, not 'str' : <https://blog.csdn.net/u011675334/article/details/108768482>