



哈尔滨工业大学  
Harbin Institute of Technology

## 计算机网络 课程实验报告

实验名称	HTTP 代理服务器的设计与实现					
姓名	熊峰		院系	计算机科学与技术		
班级	1903104		学号	1190200708		
任课教师	刘亚维		指导教师	刘亚维		
实验地点	格物 207		实验时间	2021/10/24		
实验课表现	出勤、表现得分(10)		实验报告 得分(40)		实验总分	
	操作结果得分(50)					
教师评语						

**实验目的：**

本次实验的主要目的。

熟悉并掌握 Socket 网络编程的过程与技术；深入理解 HTTP 协议，掌握 HTTP 代理服务器的基本工作原理；掌握 HTTP 代理服务器设计与编程实现的基本技能。

**实验内容：**

- (1) 设计并实现一个基本 HTTP 代理服务器。要求在指定端口（例如8080）接收来自客户的 HTTP 请求并且根据其中的 URL 地址访问该地址所指向的 HTTP 服务器（原服务器），接收 HTTP 服务器的响应报文，并将响应报文转发给对应的客户进行浏览。
- (2) 设计并实现一个支持 Cache 功能的 HTTP 代理服务器。要求能缓存原服务器响应的对象，并能够通过修改请求报文（添加 if-modified-since 头行），向原服务器确认缓存对象是否是最新版本。（选作内容，加分项目，可以当堂完成或课下完成）
- (3) 扩展 HTTP 代理服务器，支持如下功能：（选作内容，加分项目，可以当堂完成或课下完成）
  - a) 网站过滤：允许/不允许访问某些网站；
  - b) 用户过滤：支持/不支持某些用户访问外部网站；
  - c) 网站引导：将用户对某个网站的访问引导至一个模拟网站（钓鱼）。

**实验过程：**

以文字描述、实验结果截图等形式阐述实验过程，必要时可附相应的代码截图或以附件形式提交。

- (1) 实验使用C++ OOP完成，实验最初设计建立一个Server类、Client类，使用HttpProxyServer继承Server类和Client类提高代码整体的复用性，但由于子进程中并发的的问题，最终采用如下方法创建了一个Server类，作为HttpProxyServer类的父类，在子进程中实例化Client类，创建Client对象，以便和服务收发信息，由于采用C++ OOP特性，代码复用性大大增强。
- (2) 在子进程中实例化Server，可以将接收到的报文发送给http服务器，并将接受到的报文发送到http客户端，增加原服务器缓存的功能，同时可以将所收到的报文添加if-modified-since头行，可以检查保存的响应报文是否为最新版本。在实际使用时，代理服务器接收到http客户端的请求报文时，向http服务器发送请求报文，获取最新报文头，再检查本地cache，若已缓存，且响应报文头为304 Not Modified，则说明此时文件仍为最新版本，此时将本地缓存读入，并发送到http客户端，若此时已改变，则将响应报文写入，并向http客户端发送。若本地cache中不含客户所需文件，则将http服务器的响应内容写入本地文件，并向http客户端发送。
- (3) 扩展代理服务器功能：
  - a) 网站过滤：对接受到的报文分析，若报文分析出的url和host为所禁止的网站，则提示禁止访问信息，并不再给http服务器发送请求。
  - b) 用户过滤：由于设置SOCKADDR\_IN结构体，在接受http客户端的请求时，会保存客户端的ip和端口，因此可以根据所设置的要被过滤的用户，过滤访问外部网站的功能。
  - c) 网站引导：代理服务器接收到用户的请求报文时，分析用户的请求报文，若分析出的url和host符合被钓鱼的网站，则将其修改为钓鱼网站，并将请求报文的url和host修改为钓鱼网站，从而网站引导的效果。

**实验结果：**

- (1) 指定端口8080接受来自客户HTTP请求，并根据其中的 URL 地址访问该地址所指向的 HTTP 服务器（原服务器），接收 HTTP 服务器的响应报文，并将响应报文转发给对应的客户进行浏览。

```
HttpProxyServerRun
D:\Code\CLionProjects\ComputerNetworkExperiment\Lab1_Prox
Proxy加载WinSock成功!
代理服务正在运行, 当前代理端口为8080
HTTP客户端已连接!
当前线程ID=19252
Client Said:
GET http://cs.hit.edu.cn/ HTTP/1.1
Host: cs.hit.edu.cn
Proxy-Connection: keep-alive
Cache-Control: max-age=0
Upgrade-Insecure-Requests: 1
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) App
537.36
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,application/atom+xml;q=0.9,application/javascript;q=0.9,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,zh-TW;q=0.6
Cookie: _ga=GA1.3.944203100.1611215889; JSESSIONID=B12EC0
```



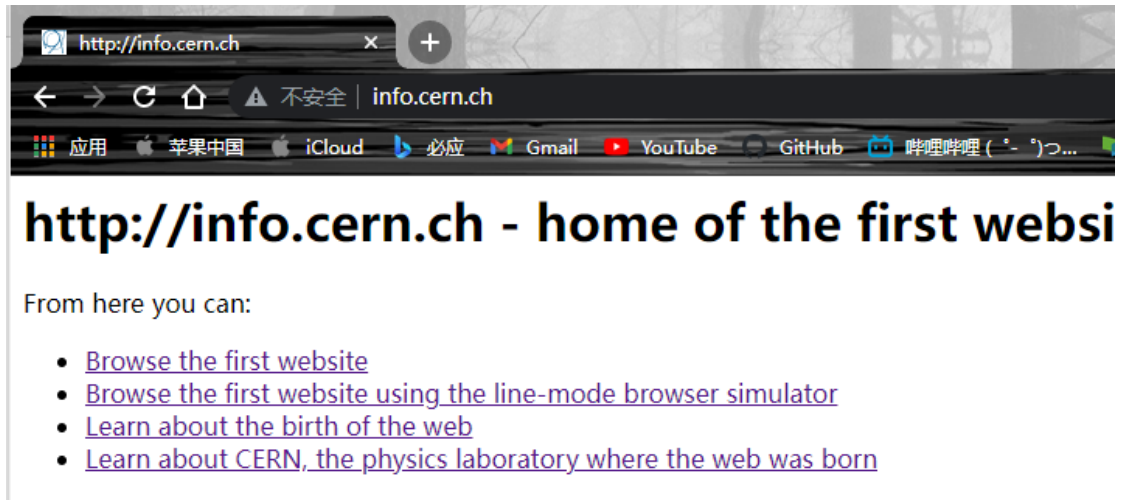
(2) 缓存原服务器响应的对象，并能够通过修改请求报文（添加 if-modified-since 头行），向原服务器确认缓存对象是否是最新版本。

将所访问的文件，写入本地cache，并对有Last-Modified字段的响应报文，将其写入cache首行，再将响应报文写入文件中。

```
1 Last-Modified: Wed, 05 Feb 2014 16:00:31 GMT
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16 HTTP/1.1 200 OK
17
18 Date: Sat, 30 Oct 2021 04:37:21 GMT
19
```

若响应报文回复Not Modified，则直接将本地cache发送。

```
From client to proxy.发送完成!
当前访问的url是:http://info.cern.ch/
From httpServer to proxy.正在接收!
check Last-Modified! Result: not modified!
向client发送cache缓存!
```



若未缓存，则写入文件并发送。

```
当前访问的url是:http://info.cern.ch/
From httpServer to proxy.正在接收!
写入cache并向client发送cache缓存!
From proxy to client.发送成功!
```



### (3) 扩展 HTTP 代理服务器

a) 网站过滤：允许/不允许访问某些网站；

过滤掉<http://www.gov.cn>

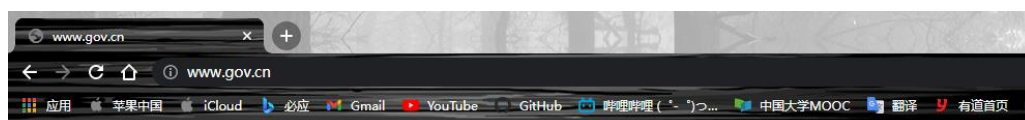
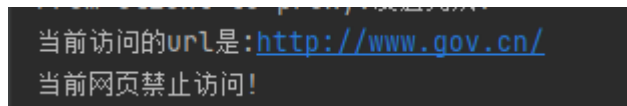
过滤前：



将网站过滤设置为<http://www.gov.cn>

过滤后：

浏览器无法访问<http://www.gov.cn>



#### 未连接到互联网

代理服务器出现问题，或者地址有误。

请试试以下办法：

- 联系系统管理员
- 检查代理服务器地址
- 运行 Windows 网络诊断

ERR\_PROXY\_CONNECTION\_FAILED

详情

b) 用户过滤：支持/不支持某些用户访问外部网站；

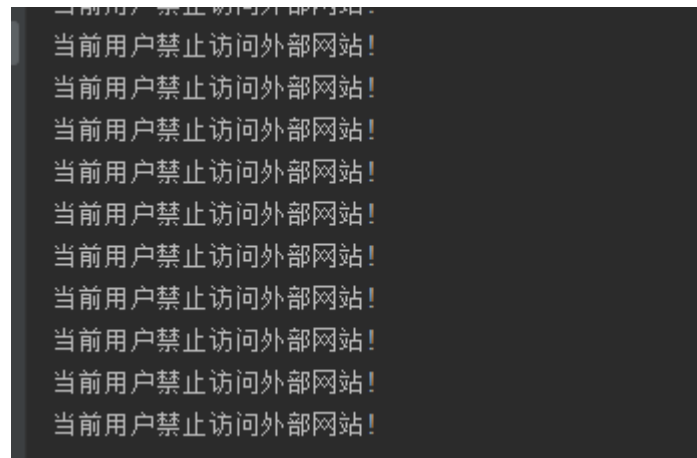
例如对本机限制，将过滤的ip设置为127.0.0.1.

过滤前：

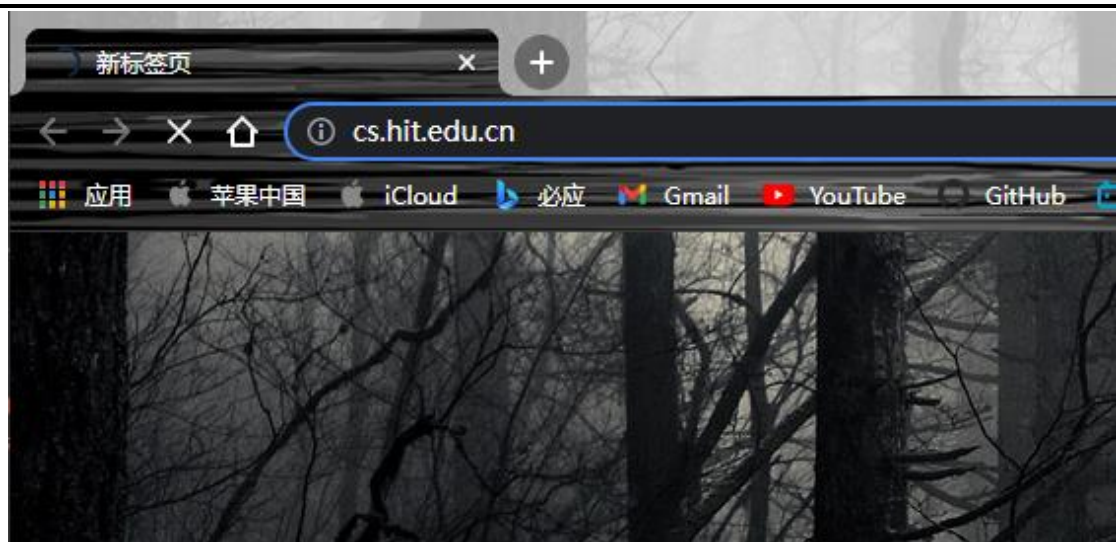
过滤前，网页正常打开。



过滤后：网站加载失败。







c) 网站引导：将用户对某个网站的访问引导至一个模拟网站（钓鱼）。  
将http://www.gov.cn跳转到http://jwts.hit.edu.cn。



问题讨论：

实验过程中，出现图片加载不全的情况。来发现是响应报文尚未被完全接收，实验应该将所有的报文都保存并发送到http客户端。

心得体会：

1. 对HTTP协议有了更深刻的了解；
2. 对HTTP报文头理解更加深刻；
3. 了解了代理服务器的工作原理；
4. 掌握socket编程的技能。