



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	李劲光		院系	计算机科学与技术学院		
班级	1903201		学号	L190202102		
任课教师	聂兰顺		指导教师	聂兰顺		
实验地点			实验时间			
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：

(注：实验报告模板中的各项内容仅供参考，可依照实际实验情况进行修改。)

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：

概述本次实验的主要内容，包含的实验项等。

- (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：

服务器端：

- 1) 首先是创建 `socket.socket(family,type[protocol])`。
- 2) 向服务器发出连接请求（hostname， port）如果连接有出错就返回`socket.error` 。
- 3) 进行过滤，用户和访问外部网站是否在被封名单里。
- 4) 进行通信发送数据， 将string类发送到连套字，等收到返回byte之后sendall将string中的数据发送到连接的套接字给用户。
- 5) 关闭套接字 `socket.close()`。

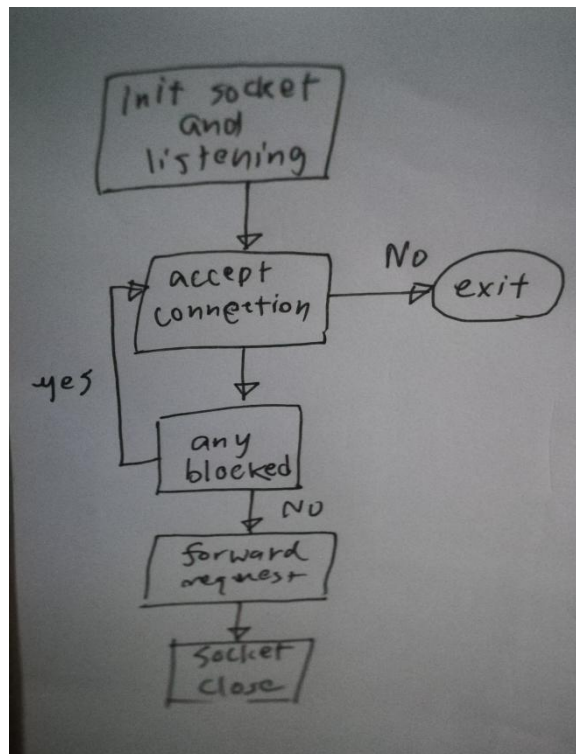
用户端:

- 1) 首先是创建 `socket.socket(family,type[protocol])`。
- 2) 请求到来后, 接受连接请求得到新的对应连接的套接字。
- 3) 向服务器发出连接请求 (`hostname`, `port`) 如果连接有出错就返回`socket.error`, 将
- 4) 进行过滤, 用户和访问外部网站是否在被封名单里。
- 5) 与服务器进行通信发送数据, 将`string`类发送到连套字, 等收到返回`byte`之后`send`将`string`中的数据发送到连接的套接字给用户。
- 6) 数据以字符串形式返回 `socket.recv(buffer)`, `buffer`大小要接收的最大数据量。
- 7) 关闭套接字 `socket.close()`。

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。

HTTP 代理服务器的基本原理

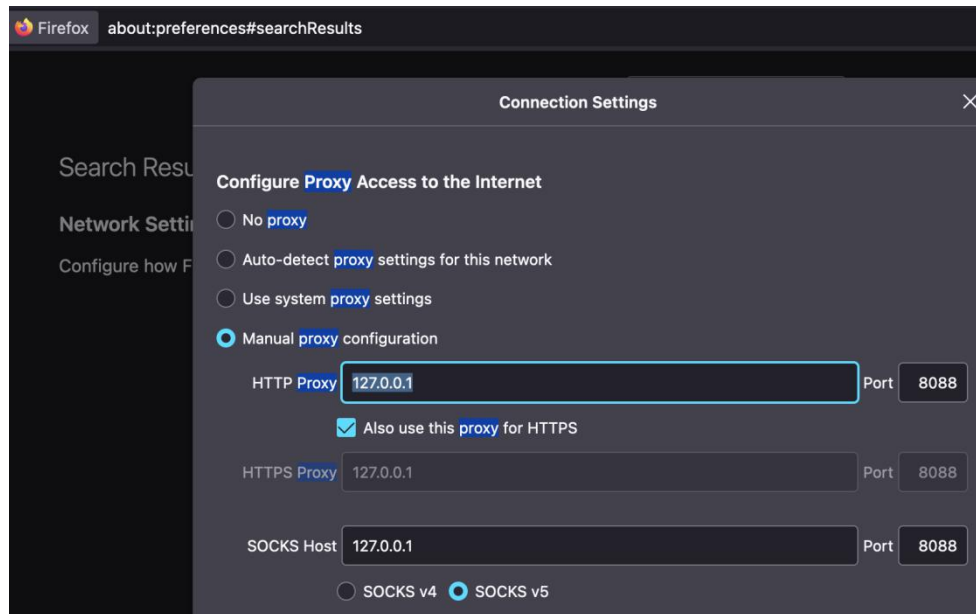
HTTP 代理服务器是网络的中间实体, 既是 Web 服务器又是 Web 客户端。代理位于 Web 客户端和 Web 服务器之间, 充当“中间人”。对于客户端, 代理扮演服务器的角色, 接收请求并返回响应; 对于服务器, 代理扮演客户端的角色, 发送请求和接收响应。它接受客户端的请求并根据客户端的请求连接到远程服务器, 并将请求发送给服务器, 然后将接收到的服务器的响应发送回客户端, 完成其基本功能。代理服务器的扩展功能可以通过对客户端/服务器已经接受/将要发送的消息进行处理来实现。



实验结果：

采用演示截图、文字说明等方式，给出本次实验的实验结果。

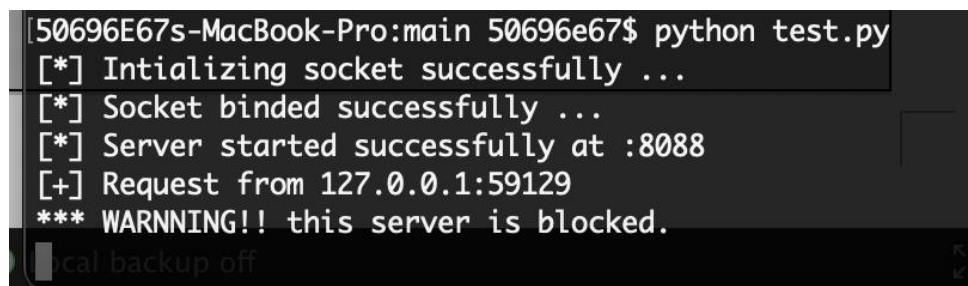
1) 设置代理



2) 访问今 80 的端口好号



如果访问到被封的网站地址：

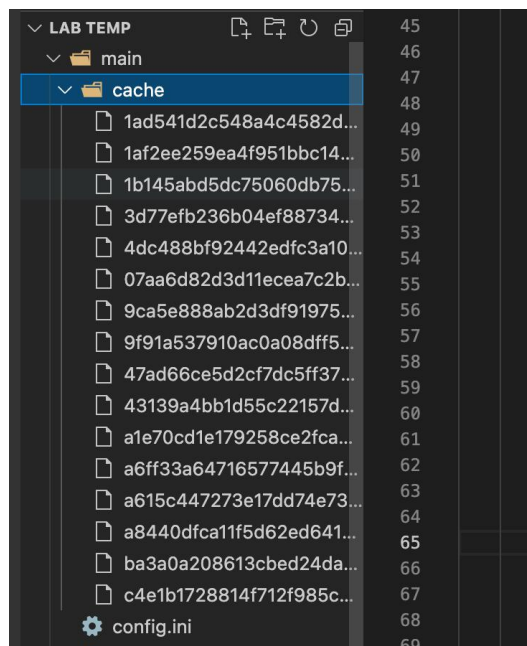


- 3) 当被封用户来访问某个网站，就直接拒绝请求而在终端显示。

```
[50696E67s-MacBook-Pro:main 50696e67$ python test.py
[*] Intializing socket successfully...
[*] Socket binded successfully...
[*] Server started successfully at :8088
[+] Request from 127.0.0.1:59589
*** WARNING!! this user is blocked.
```

- 4) 缓存功能：Cache 文件产生

```
[50696E67s-MacBook-Pro:main 50696e67$ python test.py
[*] Intializing socket successfully ...
[*] Socket binded successfully ...
[*] Server started successfully at :8088
[+] Request from 127.0.0.1:59323
[+] webserver : www.cri.cn:80
Cache miss: www.cri.cn/zhaopin/about
```



- 5) 在config文件里过滤功能，基本setting:

```
main > config.ini
1 [CONFIGURATION]
2 host=127.0.0.1
3 port=8088
4 buffer=65507
5 cache=1024
6 max=3
7
8 [BLOCK]
9 host=["http://httpbin.org/get", "www.portquiz.net"]
10 user=["192.168.2.2", "127.0.0.1"]
11
```

问题讨论：

对实验过程中的思考问题进行讨论或回答。

主要是cookie与cache的区别。cookie存储在客户端和存储一些字符串，但是缓存一般存储在服务器端的图片和文件等，缓存信息中的状态码304指的是没有更新缓存而返回给客户。使用缓存代理服务器，需要保存请求报文对应的响应报文，顺序不能错和信息不能丢失或者缺漏。

心得体会：

结合实验过程和结果给出实验的体会和收获。

理解了 HTTP 协议，HTTP 代理服务器的基本工作过程和原理，学习更多了关于 socket 网络编程与相关的技术。