



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	卢兑琬		院系	计算机科学与技术学院		
班级	1803501		学号	L170300901		
任课教师			指导教师			
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						

计算学部

### 实验目的：

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。
- (3) 扩展 HTTP 代理服务器， 支持如下功能：
  - 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（钓鱼网站）

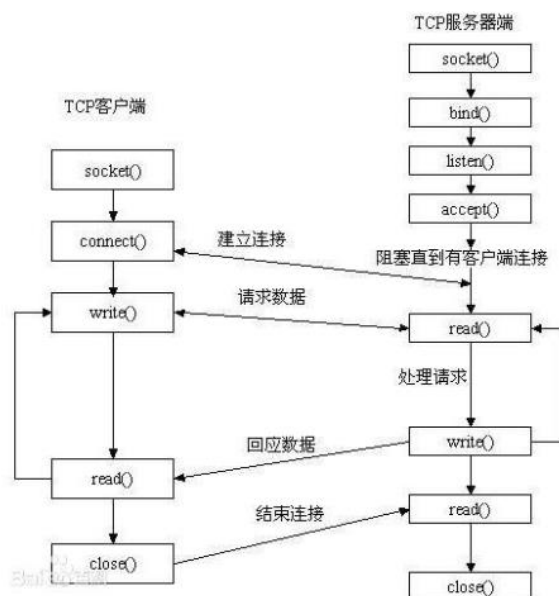
### 实验过程：

#### 1、 实验原理

##### (1) Socket 编程的客户端和服务端主要步骤

在 TCP/IP 网络应用中， 通信的两个进程之间相互作用的主要模式是客户/服务器（C/S 或 B/S） 模式， 即客户向服务器发出请求， 服务器接受到请求后， 提供相应的服务。

两者的工作步骤可以通过下面流程图直观地看到：



#### 服务器端：

其过程是首先服务器方要先启动， 并根据请求提供相应服务：

- (1) 打开一通信通道并告知本地主机， 它愿意在某一公认地址上的某端口接收客户请求； 对应的操作是申请一个 socket， 这时的 socket 称为“欢迎套接字”， 然后绑定（bind） 本地地址信息和欢迎套接字， 然后开放

监听 (listen)。

(2) 等待客户请求到达该端口；

(3) 接收到客户端的服务请求时，处理该请求并发送应答信号。在 TCP 实现过程中进行了三次握手操作，但是实际编写过程中通过 `accept` 函数即可实现上述操作，并建立连接，注意这个时候才真正建立起了与客户机传输数据的套接字。接收到并发服务请求，要激活一新进程来处理这个客户请求。新进程处理此客户请求，并不需要对其余请求作出应答。服务完成后，关闭此新进程与客户的通信链路，并终止。

(4) 返回第 (2) 步，等待另一客户请求。

(5) 关闭服务器，对应的也就是关闭服务器的欢迎套接字。

客户端：

(1) 打开一通信通道，即建立起要与服务器传输数据的套接字 `socket`，通过 `connect` 连接到服务器所在主机的特定端口；

(2) 向服务器发服务请求报文，等待并接收应答；继续提出请求.....

(3) 请求结束后关闭通信通道并终止。

从上面所描述过程可知：

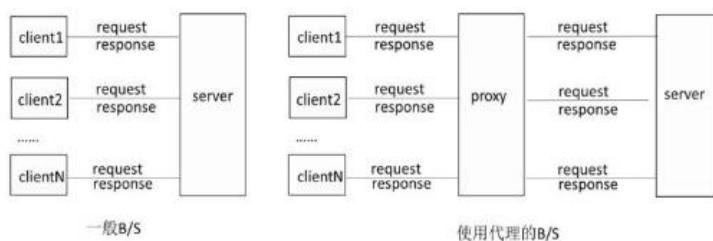
(1) 客户与服务器进程的作用是非对称的，因此代码不同。

(2) 服务器进程一般是先启动的。只要系统运行，该服务进程一直存在，直到正常或强迫终止。

## (2) HTTP 代理服务器的基本原理与流程图

代理服务器，俗称“翻墙软件”，允许一个网络终端（一般为客户端），通过这个服务与另一个网络终端（一般为服务器）进行非直接的连接。

如图 1-1 所示，为普通 Web 应用通信方式与采用代理服务器的通信方式的对比。



代理服务器可以认为是 TCP/IP 网络应用的客户端和服务端的结合。

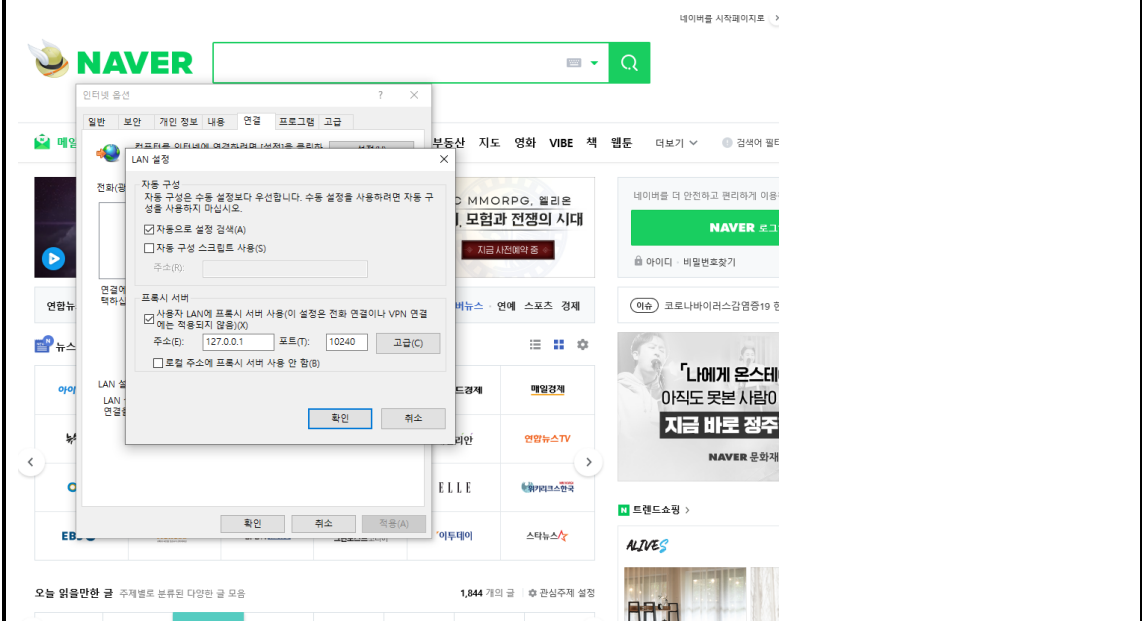
一方面，它是浏览器客户端的服务器端，另一方面，它也是目标服务器的客户端。浏览器将请求报文发送给代理服务器，代理服务器经过一些处理或者不经过处理，将请求报文转发给目标服务器；目标服务器相应请求报文发出响应报文，代理服务器接受到响应报文之后直接将响应报文转发给浏览器客户端。

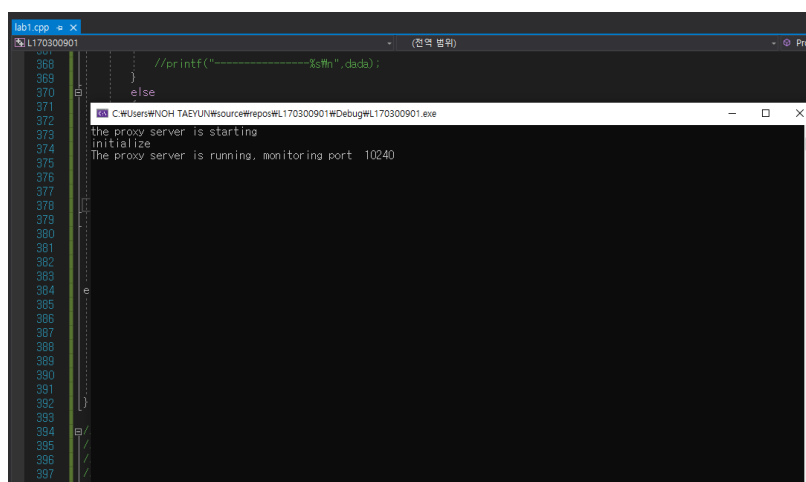
代理服务器在指定端口（例如 10240）监听浏览器的访问请求（需要在客户端浏览器进行相应的设置），接收到浏览器对远程网站的浏览请求时，首先查看浏览器来源的 ip 地址，如果属于被限制的用户，则认为没有接受到访问请求。否则，查看其请求的 host 主机，如果属于不允许访问的主机，则默认不向目标服务器发送请求；如果属于被引导的网站，则对该网站的请求报文中的 host 主机地址和 url 进行更改，代理服务器开始在代理服务器的缓存中检索 URL 对应的对象（网页、图像等对象），找到对象文件后，提取该对象文件的最新被修改时间；代理服务器程序在客户的请

求报文首部插入<code>&lt;If-Modified-Since:</code> 对象文件的最新被修改时间<code>&gt;</code>，并向原 Web 服务器转发修改后的请求报文。如果代理服务器没有该对象的缓存，则会直接向原服务器转发请求报文，并将原服务器返回的响应直接转发给客户端，同时将对象缓存到代理服务器中。代理服务器程序会根据缓存的时间、大小和提取记录等对缓存进行清理。

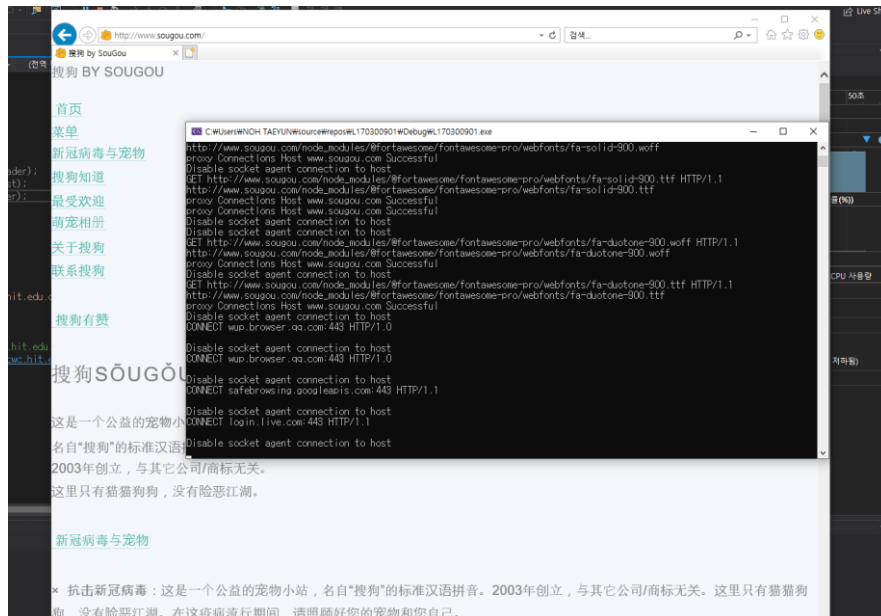
```
graph TD
    subgraph Browser [浏览器]
        direction TB
        B1[ ]
        B2[ ]
        B3[ ]
        B4[ ]
    end
    subgraph ProxyServer [代理服务器]
        direction TB
        P1[Proxysocket()] --> P2[Bind()]
        P2 --> P3[Listen()]
        P3 --> P4[Accept()]
        P4 --> P5["Recvfrom() // 从客户端"]
        P5 --> P6[Serversocket()]
        P6 --> P7[Connect()]
        P7 --> P8["Sento() // 到服务器"]
        P8 --> P9["Recvfrom() // 从服务器"]
        P9 --> P10["Sento() // 到客户端"]
        P10 --> P11[close() // 关闭所有的]
    end
    subgraph OriginServer [原服务器]
        direction TB
        O1[ ]
        O2[ ]
        O3[ ]
    end

    P4 -- "建立连接" --> B1
    B2 -- "请求报文" --> P4
    P7 -- "建立连接" --> O1
    O2 -- "请求报文" --> P8
    P9 -- "响应报文" --> O3
    P10 -- "响应报文" --> B3
    P11 -- "关闭连接" --> B4
    P11 -- "关闭连接" --> O3
```





3. 在 IE 浏览器输入 [www.sougou.com](http://www.sougou.com), 在程序运行窗口发现了请求报文, 并在浏览器端接收到了网页的数据, 说明代理服务器基本功能实现。



4. （拓展功能 1-支持 Cache 功能）我们在加入缓冲区的代码中设置一个断点，如图

```

for (i = length2; i < recvSize; i++) {
    Buffer2[length++] = Buffer[i];
}
delete CacheBuffer;

ret = send(((ProxyParam*)lpParameter)->serverSocket, Buffer2, strlen(Buffer2)
+ 1, 0);
recvSize = recv(((ProxyParam
*)lpParameter)->serverSocket, Buffer2, MAXSIZE, 0);

if (recvSize <= 0) {
    goto error;
}

const char* blank = " ";
const char* Modd = "304";
if (!memcmp(&Buffer2[9], Modd, strlen(Modd)))
{
    ret = send(((ProxyParam
*)lpParameter)->clientSocket, cache[find].buffer, strlen(cache[find].buffer) + 1, 0);
    goto error;
}
}

```

图中 309 行的 Buffer2 是在缓冲区找到原网页，并且在原请求报文上加上了 If-Modified-Since: 段的请求报文，刷新网页，查看 Buffer2:

5. （拓展功能 2-a-网站过滤）在程序中设置了一个数组

以下四个网站主机被禁，接下来我们在浏览器输入被禁网站之一 <http://today.hit.edu.cn/>，看其反馈

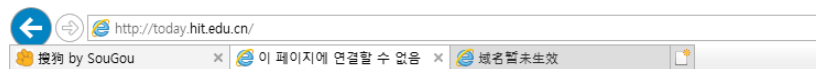
```

SOCKET ProxyServer;
sockaddr_in ProxyServerAddr;
const int ProxyPort = 10240;

const char* host[10] = { "today.hit.edu.cn", "jwts.hit.edu.cn", "www.hit.edu.cn" };
const int host_number = 3;

const char* host_to_another = "cwc.hit.edu.cn"; //cs.hit.edu.cn
const char* another[2] = { "cs.hit.edu.cn", "http://cwc.hit.edu.cn" }; //

```

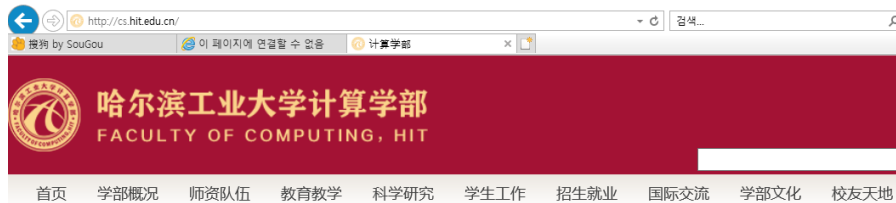


## 이 페이지에 연결할 수 없음

- 웹 주소 <http://today.hit.edu.cn> 이(가) 올바른지 확인하세요.
- [Bing에서 이 페이지 검색](#)
- [페이지 새로 고침](#)

▼ 추가 정보

연결 문제 진단



首页 学部概况 师资队伍 教育教学 科学研究 学生工作 招生就业 国际交流 学部文化 校友天地

### 学部新闻

【致敬百年 引领未来】哈工大计算学部致海内外校友捐款倡议...

亲爱的哈工大计算机校友：丁酉盛夏，百年芳华。2020年6月，我们共同的母年华诞。这是哈工大发展历史上重要的里程碑，更是哈工大广大校友隆重的纪念日。

【致敬百年 引领未来】哈工大计算学部致海内外校友捐款倡议...

计算学部成功举办哈尔滨工业大学首届四史教育知识竞赛

计算学部校友参观校航天馆、博物馆及学部部分研究中心

计算学部举办校友学术报告会

计算学部校友会召开理事会第一次会议

### 通知公告

MORE

- 计算学部关于开展2020年秋季学期“名师支团子”活动的通知 11-04
- 计算学部各本科团支部：为深入贯彻落实党的十九大精神和习近平新时代中国特色社会主义思想，全面... 10-30
- 关于“神经网络机器翻译核心技术及产业化”项目申报2020年度北... 10-30
- 北京百度网讯科技有限公司等申报的《神经网络机器翻译核心技术及产业化》项目，拟申请2020年北京市... 10-23
- 计算学部2020年度研究生国家奖学金候选人公示 10-23
- 根据《哈尔滨工业大学计算学部的2020年度研究生国家奖学金评审办法》的有关规定，计算学部本着公开、公平、公正的原则，经学部... 10-23

### 党建之窗

【研究生主题党日】计算学部网安博士生党支部开展“打造网安国之重器”主题党日

计算学部软件学院参加示范性软件学院领导干部“弘扬雷锋精神”党性教育

计算学部召开2020年秋季学生党建工作布置会

计算学部关于开展“学习四史 砥砺前行”主题党日活动的通知

中共哈尔滨工业大学计算学部党员代表大会顺利召开

问题讨论：

1. CodeBlocks 无法使用 `strtok_s()` 函数，这里需要切换成 `strtok()` 函数。
2. `goto` 语句后面，不能再新定义变量，否则，会出现一些问题

心得体会：

经过此次实验， 在实践过程中很清晰地学到了 TCP 协议在传输数据的流程和方式；熟悉了 Socket 网络编程的过程与技术； 同时也更清晰地掌握了 HTTP 代理服务器的基本工作原理； 掌握了 HTTP 代理服务器设计与编程实现的基本技能。 并且也在这基础上了解了浏览器在进行搜索网页过程中， 网络所起的作用以及具体的工作原理， 数据的传输方式等等， 受益匪浅。