

**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 | 王子奕 | | 院系 | 计算学部 | | |
| 班级 | 1937101 | | 学号 | 1190200121 | | |
| 任课教师 | 李全龙 | | 指导教师 | 李全龙 | | |
| 实验地点 | G207 | | 实验时间 | 2021.10.28 | | |
| 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： |
| 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。  3. 网站过滤:允许/不允许访问某些网站;  4. 用户过滤:支持/不支持某些用户访问外部网站;  5. 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： |
| 1. 浏览器使用代理   利用Chrom浏览器中的Proxy SwitchyOmega插件进行代理操作，通过切换不同的代理方案进行IP地址的转换和代理服务设置。     1. Socket 编程的客户端和服务器端的主要步骤   （1）首先浏览器上的Socket 客户端，要获取目的服务器的 IP 地址、端口号以及传输层协议(TCP 或 UDP)，创建Socket用于通信即可，消息传输完毕后断开 Socket 连接。  （2）其次Socket服务器端和客户端类似注意几种信息，并进行并行操作，创建多个线程。  （3）对于TCP的通信，服务器最开始创建一个Socket提前进行联系进行控制，每个线程中创建一个新的Socket用于与源主机进行通信。对于UDP的通信，无需提前建立联系，直接创建Socket进行通信即可。   1. HTTP 代理服务器的基本原理   基本原理：利用HTTP代理服务器转发源主机的HTTP对网页的访问请求报文到目的服务器，之后将目的服务器发回的报文再转发到源主机上。   1. HTTP代理服务器流程图      1. 技术重点及具体实施方案 2. 解析获取URL与IP地址     观察报文信息对URL和IP进行字符串提取     1. 初始化配置信息   将配置信息存入config.yaml中    不需要的内容注释即可   1. 缓存机制   若当前请求报文对应内容在本地有，则向目的主机发送加入If-modified-since的条件GET方法，若返回消息为304，即在这段时间内未更新过，则直接返回本地。反之为目标文件已经发生了变化或者缺少 Last-Modified 头部都需要重新转发请求报文至目的主机，并将返回报文重写到本地。   1. 权限机制   封目的主机：判断当前访问目标网址是否被封，判断当前目的主机的一部分是否在forbidden\_web名单中。封IP，判断当前user是否在forbidden\_user名单中。钓鱼，若当前网址为钓鱼网址，直接将返回报文信息修改为对应地fishing名单网址。   1. 转发机制   在代理服务器的实现中主要涉及了三种 Socket ，分别为：  其一为代理服务器用于处理 TCP 请求的 Socket。在本次实验中，将这个 Socket 的默认绑定在本地的 2001 端口；  其二为用于直接与源主机连接的 Socket，用于接受来自源主机的 HTTP 请求报文和从代理服务器将 HTTP 的响应报文转发至源主机。；  其三为代理服务器代替源主机与目的服务器进行连接的 Socket，主要负责将源主机的HTTP 请求转发发送至其目的服务器，并获取返回的HTTP响应报文。   1. 多线程机制   每次获得源主机的一个 TCP 请求，服务器会启动一个线程中并开启一个此种 Socket 用于处理该连接。 |
| 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. 基本功能   访问教务处网站可观察到缓存文件夹的变化，输出的报文信息如下图     1. 缓存机制   输出该报文来源为[CACHE]即本次请求报文在本地有缓存且这段时间未被修改过（返回GET消息为304）     1. 封号机制   当前主机（127.0.0.1）访问任何网站都不可以     1. 禁止访问某些网站     可以观察到访问该页面（wyydj1.top）为禁止访问，提示信息[REJECT HOSTNAME]   1. 钓鱼网站     访问钓鱼网址（示例：fishing.com）导向哈工大官网 |
| 问题讨论： |
| 在https协议上有复杂的加密信息，仅靠目前代码无法实现。此外封一些网站的操作需要利用re.match进行字符串匹配，当前访问网址有一部分被封就无法访问。 |
| 心得体会： |
| 通过本次实验我深刻理解了HTTP代理服务器的基本工作流程，Socket编程的基本语法与流程，当然目前实现的操作与真实的代理服务器还相差甚远，本次实验我收获颇丰。 |