

**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 | 王丙昊 | | 院系 | 计算机科学与技术 | | |
| 班级 | 1603108 | | 学号 | 1160300302 | | |
| 任课教师 | 聂兰顺 | | 指导教师 | 聂兰顺 | | |
| 实验地点 | 格物207 | | 实验时间 | 2018-10-31 | | |
| 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。）  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. 实现一个基本的代理服务器，在制定端口接收来自客户的ＨＴＴＰ请求并根据其中的ｕｒｌ地址访问服务器，接收HTTP服务器的响应报文并将其转发给客户进行浏览。 2. 实现一个支持Ｃａｃｈｅ的代理服务器。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首部行），向原服务器确认是否为最新版本。 3. 扩展功能，包括网站过滤、用户过滤和网络钓鱼 |
| 实验过程： |
| 1. **HTTP代理服务器实现的基本步骤** 2. 指定代理服务器的IP地址（使用回环地址１２７．０．０．１）以及端口号，在该端口创建一个套接字并初始化，等待客户的HTTP请求。 3. 每收到客户发来的HTTP请求，创建一个新的线程处理客户的请求。 4. 接收客户端的请求报文，首先判断该报文是否为空，如果为空，直接关闭与客户端的连接，并结束线程。否则，构造代理服务器的请求报文，并创建与原服务器进行通信的套接字与之进行通信。构造请求报文的过程中，可以通过修改请求报文，来实现网络过滤、用户过滤和网络钓鱼等功能；并可以根据是否已经缓存该ｕｒｌ的响应报文，在首部行中加入ｉｆ－ｍｏｄｉｆｉｅｄ－ｓｉｎｃｅ字段 5. 对于请求报文加入ｉｆ－ｍｏｄｉｆｉｅｄ－ｓｉｎｃｅ字段的情况，需要根据响应报文中的Last－modified字段与缓存文件中Last－modified相比较的情况，来决定是否要将新的报文发送给客户端（并缓存），还是将缓存的报文发送给客户端；对于请求报文中没有加入ｉｆ－ｍｏｄｉｆｉｅｄ－ｓｉｎｃｅ字段的情况，接收响应报文，直接缓存并发送给客户端即可 6. **实现HTTP代理服务器的关键技术及解决方案**    1. 创建HTTP代理服务器套接字      * 1. 网络过滤   通过判断ｕｒｌ．ｈｏｓｔｎａｍｅ是否在禁止访问的域名集合中       * 1. 用户过滤   通过判断客户端的ＩＰ地址是否在禁止用户的集合中。       * 1. 网络钓鱼   使用一个字典，如果收到的ｕｒｌ．ｈｏｓｔｎａｍｅ在该字典的ｋｅｙｓ集合中，则将ｈｏｓｔｎａｍｅ替换成它所对应的ｖａｌｕｅ值。     * 1. Ｃａｃｈｅ的实现   在本地维持维护一个缓存文件夹，每当从原服务器收到响应报文，如果没有缓存或者缓存的数据已经被修改过，则将新的响应报文缓存，缓存文件采用路径的md5码命名。如果HTTP代理服务器发送请求报文的过程中，发现本地有缓存，需要在原请求报文中加入ｉｆ－ｍｏｄｉｆｉｅｄ－ｓｉｎｃｅ字段，报文中的时间要采用ＧＭＴ格式    报文中时间与本次发送时间进行比较时，需要将两个ＧＭＴ格式的时间转化为ｄａｔｅｔｉｍｅ格式的时间，以便于进行比较 |
| 实验结果： |
| **（1）基本功能的测试：**  HTTP代理服务器在１２７.０.０.１：８０８０等待客户的HTTP请求，验证前要配置浏览器的代理服务器。      访问　ｊｗｔｓ．ｈｉｔ．ｅｄｕ．ｃｎ，可以访问    下面为访问ｊｗｔｓ时的请求报文    **（2）Cache的测试**：  这里使用BIT的官网进行测试(因为jwts.hit.edu.cn的返回报文中没有Last-Modified字段)，首先第一次访问[www.bit.edu.cn](http://www.bit.edu.cn)，可以访问    报文如下    可以在本地Cache文件夹里发现缓存文件：    可以查看缓存文件的内容：    当再一次访问[www.bit.edu.cn](http://www.bit.edu.cn)时，同样可以访问，查看报文    可以看到此时提示，该页面没有被更新过    **（3）网站过滤功能的测试：**    设today.hit.edu.cn为禁止访问网站，此时HTTP代理服务器会返回给客户端403 Forbidden    **（4）用户过滤功能的测试：**    测试时，将回环测试IP 127.0.0.1设置为禁止访问IP，则访问jwts.hit.edu.cn    此时会禁止访问  **（5）网络钓鱼功能的测试：**  测试时，设置为，如果访问hitgs.hit.edu.cn，则会跳转到jwts.hit.edu.cn，结果如下：      可以发现，输入的是hitgs.edu.cn，但是却被引导至了jwts.hit.edu.cn |
| 问题讨论： |
| 1. **HTTP代理服务器的基本原理**   HTTP代理服务器是能够代表初始Ｗｅｂ服务器来满足HTTP请求的网络实体，它有自己的磁盘存储空间。可以通过浏览器设置，将客户所有的ＨＴＴＰ请求首先指向ＨＴＴＰ代理服务器。HTTP代理服务器是服务器同时又是客户，当它接收浏览器的请求并发回响应时，它是一个服务器。当它向初始服务器发出请求并接收响应时，它是一个客户。   1. **HTTP代理服务器的程序流程图**      1. **实现HTTP代理服务器的关键技术及解决方案**   见**实验过程**中的网络过滤、用户过滤、Cache以及网络钓鱼等技术的实现   1. **源代码**   见proxy.py |
| 心得体会： |
| 1. 更加熟悉并掌握了HTTP代理服务器的基本原理 2. 对Python的Socket编程进行了实践，有了更深的理解 3. 掌握了HTTP代理服务器设计与实现的基本技能 |