

**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 | 李卓凌 | | 院系 | 计算学部 | | |
| 班级 | 2103201 | | 学号 | 2021111000 | | |
| 任课教师 | 詹东阳 | | 指导教师 | 詹东阳 | | |
| 实验地点 | 格物213 | | 实验时间 | 2023.4.7 | | |
| 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。（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： |
| 1. **了解实验相关基础知识** 2. **代理服务器的概念**   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。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如下图所示：  C:\Users\midsummer\Desktop\捕获.JPG  代理服务器可以认为是TCP/IP网络应用的客户端和服务器端的结合。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。  针对本实验的特定要求，概括我们所要实现的代理服务器功能如下：  代理服务器在指定端口（本实验中所指定的是10240端口）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），代理服务器接收到浏览器对远程网站的浏览请求时，首先会查看浏览器来源的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（用户过滤功能）。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（网站过滤功能）；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（网站引导功能）。而对于Cache功能的实现，基本可以概括为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），若找到对象文件，则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<If-Modified-Since: 对象文件的最新被修改时间>，并向原 Web 服务器转发修改后的请求报文。若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。  除此之外，本实验要设计的服务器属于多用户代理服务器。首先，代理服务器创建 HTTP 代理服务的 TCP 主套接字，通过该主套接字监听等待客户端的连接请求。当客户端连接之后，创建一个子线程，由子线程执行上述一对一的代理过程，服务结束后子线程终止。与此同时，主线程继续接受下一个客户的代理服务。  HTTP代理服务器的流程图如下：     1. **TCP客户端与软件端的流程** 2. TCP客户端软件流程   1. 根据目标服务器IP地址与端口号创建套接字（socket），  2. 连接服务器（connect）：三次握手  3. 发送请求报文（send）  4. 接收返回报文（recv），返回3或者5  5. 关闭连接（closesocket）  根据上课的PPT，可以总结为：     1. TCP服务器端软件流程   1. 创建套接字（socket），绑定套接字的本地IP地址和端口号（bind），然后转到监听模式并设置连接请求队列大小（listen）。  2. 从连接请求队列中取出一个连接请求，并同意连接（accept）。在TCP连接过程中进行了三次握手。  3. 收到请求报文（recv）  4. 发送数据（send）返回3或者5  5. 关闭连接（closesocket）返回2  根据上课的PPT，可以总结为：     1. TCP软件端与服务器端流程图   http://c.hiphotos.baidu.com/baike/c0%3Dbaike80%2C5%2C5%2C80%2C26/sign=17baf4c7d739b60059c307e588395e4f/d000baa1cd11728b45647b06cafcc3cec3fd2c4c.jpg   1. **网络应用的Socket API(TCP)调用基本流程**      1. **常用C Socket编程函数及其功能**        1. **修改程序，完成基本功能**   对于代理服务器的实现，我采用基于C语言示例代码修改的方式完成。由于我使用的编程环境为Devcpp而非VS，程序初始提供的示例代码部分地方无法使用，因此，经过查询资料，发现有以下地方需要修改：   1. **主函数部分修改**   由于编程环境问题，不支持 int \_tmain(int argc, \_TCHAR\* argv[]) 的写法，因此，可以将其修改为 int main() 或 int main (int argc, char\* argv[])。   1. **goto语句后不能再定义变量**   在运行示例代码时，会出现错误，经过查阅相关资料，发现是在 goto 语句之后不能再定义新的变量，因此将相关代码注释即可：     1. **静态链接 Ws2\_32.lib 的问题**   示例代码修改无误后仍然无法运行，会出现很多引用无效的错误提示；  查阅相关资料，发现是这行代码的问题：    因为#pragma是VS的写法，而Devcpp内的C编译器不支持该表示格式，因此，想要运行该程序，我们需要使用g++对Ws2\_32.lib进行静态链接，相关语句如下：  g++ -Wno-unused-parameter 1.cpp -o server -lws2\_32  运行后，可以在相同文件夹下得到server.exe文件，执行该文件，即可得到我们代码的运行结果。   1. **程序基本功能描述**   在修改完成后，示例代码便可正常运行，以下简述示例代码的流程与功能：   1. 初始化一个套接字，利用 bind() 函数将该套接字与服务器 host 地址绑定，地址设为 “127.0.0.1”；同时也要绑定端口号，在示例代码中，端口号被设置为“10240”。在设置完毕后，利用 listen() 函数对该端口进行监听。 2. 通过设置 accept() 函数，对每个到来的请求进行接收和相应，为提高效率，对每个请求，代理服务器都创建一个新的线程来处理。 3. 利用 recv() 和 send() 函数，接收来自客户端的 HTTP 请求，并通过这个代理服务器将该请求转发给服务器；同时，服务器也将获得的响应发给代理服务器，然后代理服务器再将该响应发送给客户端。在这里，代理服务器相当于一个中介，提供一个代理的服务，所有的请求和响应都经过它。 4. 处理完成后，等待 200 ms 后，关闭该线程，并清理缓存，然后继续接收并处理下一个请求。对于客户端而言，它只要将正常发送的请求发给代理服务器，就可以接收到对应的响应。 5. **拓展HTTP代理服务器（网站过滤，用户过滤，网站引导）**   为便于后续实验验证，进行如下设定。  设置被屏蔽的网站为选课系统（http://jwts.hit.edu.cn/），设置钓鱼网站为（http://pku.edu.cn/）重定向至今日哈工大（http://today.hit.edu.cn/），相关示例网站可以更换.   1. **网站过滤**   在线程执行函数ProxyThread中，解析TCP报文中的HTTP头部，将HTTP头部中的URL与对应屏蔽网站URL BANNED\_WEB进行比较，如果相同，则打印相关语句并直接跳转到error，从而实现网站过滤功能。     1. **用户过滤**   修改ProxyServerAddr.sin\_addr.S\_un.S\_addr的值为inet\_addr("127.0.0.1")，表示只允许本机访问代理服务器，从而实现用户过滤功能。代码如下：     1. **网站引导**   网站引导功能同样在线程执行函数ProxyThread中实现。实现思路为：比较HTTP头部中的URL字段是否与PHISHING\_WEB\_SRC相同，若相同，则说明客户端想要访问的网址是重定向源网址，需要执行重定向操作。具体而言，在重定向操作中，我们分部分构造302报文，将其中的IP地址修改为钓鱼目的IP地址，在最后，将修改好的302报文通过send函数返回给客户端。  网站引导的核心函数如下图所示：     1. **实现Cache功能**   客户端第一次请求服务器中的数据时，代理服务器将该请求返回的响应缓存下来，存到本地的文件下。  当客户端第二次访问该数据时，代理服务器检查本地是否有该请求的响应，如果没有，则继续缓存；如果有，则向服务器发送一个请求，该请求需要增加 “If-Modified-Since” 字段，通过此字段，告知服务器缓存资源最后修改的时间（可以将 “Date” 字段进行解析），服务器通过对比最后修改时间来判断缓存是否过期，如果没过期，服务器返回状态码304，代理服务器直接将本地缓存发送给客户端；如果缓存过期，服务器返回状态码200，同时返回一个更新过的响应，代理服务器接收后，将该响应发回给客户端，并更新本地缓存. |
| 实验结果： |
| 1. **修改浏览器的代理设置**   按照实验指导书修改浏览器设置如下：     1. **通过代码实现静态链接，并运行程序，实现基础功能**   输入g++ -Wno-unused-parameter 1.cpp -o server -lws2\_32生成可执行文件server.exe，点击运行，然后打开浏览器，打开一个使用http协议的网站，如：http://today.hit.edu.cn/，可以看到如下结果：    这说明我们的代理服务器成功接收了来自客户的 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客户进行浏览。   1. **验证代理服务器的Cache功能**   再次在IE浏览器地址栏输入http://today.hit.edu.cn/进行访问，结果如下：     1. **验证代理服务器的HTTP拓展功能** 2. **网站过滤**   设定的屏蔽网站为[http://jwts.hit.edu.cn](http://today.hit.edu.cn)，输入网址尝试进行访问：     1. **用户过滤**   修改程序代码，将不允许访问的用户IP地址修改为127.0.0.1，如图所示：    可以看出，代理服务器不会对我们的访问做出相应，ip为127.0.0.1的用户被成功屏蔽。   1. **网站引导**   设定的钓鱼源网站为http://pku.edu.cn/，目的网站为[http://today.hit.edu.cn](http://jwts.hit.edu.cn)，在地址栏输入http://jwc.hit.edu.cn尝试进行访问：    网页被成功重定向。 |
| 问题讨论： |
| 这个实验因为有老师的参考代码，所以基础功能实现起来是非常简单的，对我而言不好的一点就是参考代码是在VS里可以正常运行，但我用的Devcpp运行起来会有几个报错，具体解决方案我在上面也说过了；然后就是额外功能的实现，由于上课对代码这方面的讲解少之甚少，只知道一些基本函数的功能，所以不得不去网上查阅资料，询问学长，所幸最后还是比较顺利的解决了。希望以后的培养方案可以把实验课理论部分单独抽出来，对学生的帮助会更大，理解会更好，现在这样子难免有些一知半解。 |
| 心得体会： |
| 结合实验过程和结果给出实验的体会和收获。   1. 对TCP协议传输数据的流程和方式有了更深的体会； 2. 对 socket 编程方法有了初步的了解，通过动手实践有了很大收获； 3. 了解了 HTTP 代理服务器的基本原理，掌握了 HTTP 代理服务器设计与编程实现的基本技能，对 HTTP 请求和响应原理有了更深的认识； 4. 对网站钓鱼、网站屏蔽、用户屏蔽的机制有了深刻的理解； 5. 对HTTP Cache缓存的作用有了直观印象 |