

**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 | 贺宗磊 | | 院系 | 计算机科学与技术学院 | | |
| 班级 | 1703202 | | 学号 | 1170300421 | | |
| 任课教师 | 李全龙 | | 指导教师 | 李全龙 | | |
| 实验地点 |  | | 实验时间 | 2019.10.25 | | |
| 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： |
| 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。   1. 浏览器使用代理   为了使浏览器访问网址时通过代理服务器，必须进行相关设置，以 IE 浏览器设置为例：打开浏览器工具浏览器选项连接局域网设 置代理服务器，具体过程如图 1-1 所示。  （2）Socket 编程的客户端和服务器端主要步骤  客户端：  1. 根据目标服务器IP地址与端口号创建套接字，并连接服务器（三次握手）  2. 发送请求报文  3. 接收返回报文  4. 关闭连接  服务器端：  1.创建套接字，绑定套接字的本地IP地址和端口号，对端口进行监听。  2. 从连接请求队列中取出一个连接请求，并同意连接。在TCP连接过程中进行了三次握手。  3.接收客户端请求  4.对请求进行响应，发送响应数据  5.关闭连接  （3）HTTP 代理服务器的基本原理  主要功能：  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。  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。  原理：  代理服务器在指定端口（例如 8080）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，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<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。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。   1. HTTP 代理服务器的程序流程图     （5）HTTP 代理服务器实验验证过程以及实验结果；  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。  （1）实现一个代理服务器，在指定端口10240进行监听，接受来自客户端的 HTTP 请求，并返回来自服务器的响应。  1.首先初始化一个套接字，利用 blind() 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，这里就按照指导书上的要求设置为 “10240”。然后，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，为了提供效率，对每个请求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。  （2）实现一个支持 Cache 功能的 HTTP 代理服务器，利用外部文件存储来自服务器上的信息。  1.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。  2.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。  （3）扩展 HTTP 代理服务器，支持如下功能：  （a）网站过滤：  对请求过来的 HTTP 报文头部进行检测，提取出其中的访问地址 url ，检测其是否为要被屏蔽的网址，如果是，则直接跳转到代码中的 erro 部分，即关闭套接字，断开此次连接。代码片段如下：  if (strcmp (httpHeader->url, INVILID\_WEBSITE) == 0) {  printf("\n=====================================\n\n");  printf("-------------Sorry!!!该网站已被屏蔽----------------\n");  goto error;  }  （b）用户过滤：  更改套接字绑定的主机地址，这样的话，只要不是从该地址访问代理服务器的客户端，都会被该代理服务器屏蔽，部分代码如下：  ProxyServerAddr.sin\_addr.S\_un.S\_addr = inet\_addr("127.0.0.1");//仅本机用户可访问服务器  （c）网站引导：  检测请求过来的 HTTP 报文头部，如果发现访问的网址是要被钓鱼的网址，则将该网址引导到其他网站（钓鱼网址），通过更改 HTTP 头部字段的 url （访问网址）和 host 主机名来实现，部分代码如下：  case 'H'://Host  memcpy(httpHeader->host,&p[6],strlen(p) - 6);  //钓鱼  if(!strcmp("www.hit.edu.cn",httpHeader->host)) {  memcpy(httpHeader->host,&fish[7],strlen(fish) - 8);  }  printf("%s\n",httpHeader->host);  break;  if(p[0] == 'G') { //GET 方式  memcpy(httpHeader->method,"GET",3);  memcpy(httpHeader->url,&p[4],strlen(p) -13);  memcpy(url,httpHeader->url,sizeof(httpHeader->url));  //钓鱼  if(!strcmp("http://www.hit.edu.cn/",httpHeader->url)) {  memcpy(httpHeader->url,fish,1024);  } |
| 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 HTTP 代理服务器的实现，连接网址： “http://today.hit.edu.cn/” ，显示结果 2. 支持 Cache 功能的 HTTP 代理服务器实现，访问网址 http://today.hit.edu.cn/ 显示网址如下        1. 扩展 HTTP 代理服务器： 2. 网站过滤，不允许访问网址 http://jwts.hit.edu.cn/，访问结果如图      1. 用户过滤，屏蔽本机地址，结果如图：   3.网站引导（钓鱼）：将网址 http://www.hit.edu.cn/ 的访问引导到网址http://www.hrbnu.edu.cn/ ，运行结果如图：  正常情况下应为  钓鱼后为 |
| 问题讨论： |
| 1.最早根据指导书参考代码写代码时发现，http消息的拆分无法成功，后得知CodeBlocks 不支持 strtok\_s() 函数，完成相同功能需要调用 strtok() 函数。  2.在实验中钓鱼无法成功，后通过问同学得以解决，即钓鱼的处理需要在解析HTTP头部函数内进行  3.缓存时出现无法从缓存中连接网站，后分析得出是缓存出现乱码，通过对字符格式进行规范调整，后缓存数据正常，不在出现乱码。 |
| 心得体会： |
| 结合实验过程和结果给出实验的体会和收获。  对 socket 编程有了初步的了解，掌握了 HTTP 代理服务器的基本原理，对 HTTP 请求和响应原理有了更深的认识；同时，也对网站钓鱼、网站屏蔽等有了深刻的理解；对HTTP缓存印象深刻。。 |