Application Layer homework

使用教材：English version: Computer Networking A Top-Down Approach

P1. True or false?

a. A user requests a Web page that consists of some text and three images. For this page, the client will send one request message and receive four response messages.

b. Two distinct Web pages (for example, `www.mit.edu/research.html` and `www.mit.edu/students.html`) can be sent over the same persistent connection.

c. With nonpersistent connections between browser and origin server, it is possible for a single TCP segment to carry two distinct HTTP request messages.

d. The `Date:` header in the HTTP response message indicates when the object in the response was last modified.

e. HTTP response messages never have an empty message body.

a. False. One request to get one response，非持久化连接

b. True. 持久化连接确实可以，一次发去两个 html 文件

c. False. 与 a 同理，一个 TCP segment 肯定只能一个 HTTP 请求

d. False. Date 是装的一个 request 被创建的时间，而不是对象上次被更改的时间

e. False. 比如 204 304 信息，肯定没有 message body

P4. Consider the following string of ASCII characters that were captured by Wireshark when the browser sent an HTTP GET message (i.e., this is the actual content of an HTTP GET message). The characters *<cr><lf>* are carriage return and line-feed characters (that is, the italized character string *<cr>* in the text below represents the single carriage-return character that was contained at that point in the HTTP header). Answer the following questions, indicating where in the HTTP GET message below you find the answer.

```
GET /cs453/index.html HTTP/1.1<cr><lf>Host: gai
a.cs.umass.edu<cr><lf>User-Agent: Mozilla/5.0 (
Windows;U; Windows NT 5.1; en-US; rv:1.7.2) Gec
ko/20040804 Netscape/7.2 (ax) <cr><lf>Accept:ex
t/xml, application/xml, application/xhtml+xml, text
/html;q=0.9, text/plain;q=0.8,image/png,*/*;q=0.5
```

```
<cr><lf>Accept-Language: en-us,en;q=0.5<cr><lf>Accept-
Encoding: zip,deflate<cr><lf>Accept-Charset: ISO
-8859-1,utf-8;q=0.7,*;q=0.7<cr><lf>Keep-Alive: 300<cr>
<lf>Connection:keep-alive<cr><lf><cr><lf>
```

a. What is the URL of the document requested by the browser?
b. What version of HTTP is the browser running?
c. Does the browser request a non-persistent or a persistent connection?
d. What is the IP address of the host on which the browser is running?
e. What type of browser initiates this message? Why is the browser type needed in an HTTP request message?

a. http 请求，host 是 gaia.cs.umass.edu，get 内容是/cs453/index.html。所以 url 是: http://gaia.cs.umass.edu/cs453/index.html
b. HTTP/1.1 所以 version 是 1.1
c. Connection: keep-alive 所以是 persistent connection
d. 信息未提及，无法得知
e. Mozilla/5.0 服务器需要这个信息，不同版本的 browser 只能得到特定 version 的对象信息，服务器需要确定是哪个类型的 browser 来发送对应版本的 message

P5. The text below shows the reply sent from the server in response to the HTTP GET message in the question above. Answer the following questions, indicating where in the message below you find the answer.

```
HTTP/1.1 200 OK<cr><lf>Date: Tue, 07 Mar 2008
12:39:45GMT<cr><lf>Server: Apache/2.0.52 (Fedora)
<cr><lf>Last-Modified: Sat, 10 Dec2005 18:27:46
GMT<cr><lf>ETag: "526c3-f22-a88a4c80"<cr><lf>Accept-
Ranges: bytes<cr><lf>Content-Length: 3874<cr><lf>
Keep-Alive: timeout=max=100<cr><lf>Connection:
Keep-Alive<cr><lf>Content-Type: text/html; charset=
ISO-8859-1<cr><lf><cr><lf><!doctype html public "-
//w3c//dtd html 4.0transitional//en"><lf><html><lf>
<head><lf> <meta http-equiv="Content-Type"
content="text/html; charset=iso-8859-1"><lf> <meta
name="GENERATOR" content="Mozilla/4.79 [en] (Windows NT
5.0; U) Netscape]"><lf> <title>CMPSCI 453 / 591 /
NTU-ST550ASpring 2005 homepage</title><lf></head><lf>
<much more document text following here (not shown)>
```

a. Was the server able to successfully find the document or not? What time was the document reply provided?
b. When was the document last modified?
c. How many bytes are there in the document being returned?
d. What are the first 5 bytes of the document being returned? Did the server agree to a persistent connection?

a. 成功找到了 状态信息是 200 时间是：Tue, 07, Mar, 2008, 12:39:45GMT

b. Last-modified time: Sat, 10, Dec, 2005, 18:27:46GMT
c. Content-length: 总共 3874bytes
d. 前五个 bytes 是 <!doc keep alive 允许了持久化连接

P7. Suppose within your Web browser, you click on a link to obtain a Web page. The IP address for the associated URL is not cached in your local host, so a DNS lookup is necessary to obtain the IP address. Suppose that n DNS servers are visited before your host receives the IP address from DNS; the successive visits incur an RTT of $RTT_1, \ldots, RTT_n$. Further suppose that the Web page associated with the link contains exactly one object, consisting of a large amount of HTML text. Let RTT_0 denote the RTT between the local host and the server containing the object. Assuming transmission duration of $0.002 \times RTT_0$ of the object, how much time elapses from when the client clicks on the link until the client receives the object?

No cache 对这 n 个 DNS servers 连续访问，则获取到 IP 地址的路程上的总时间是：
RTT1 + RTT2 + …… + RTTn
RTT0 是 local host 和 server 之间的 RTT，需要一个 RTT0 来建立两者之间的 TCP 连接，需要一个 RTT0 来接收对象。所以最终总时间是：
2*RTT0 + RTT1 + RTT2 + …… + RTTn

P8. Consider Problem P7 again and assume $RTT_0 = RTT_1 = RTT_2 = \ldots RTT_n = RTT$, Furthermore, assume a new HTML file, small enough to have negligible transmission time, which references nine equally small objects on the same server. How much time elapses with

a. non-persistent HTTP with no parallel TCP connections?
b. non-persistent HTTP with the browser configured for 6 parallel connections?
c. persistent HTTP?

a. 由 P7，原本是 (n+2)RTT 时间，现在总共由 9 个同等的 objects，接下来其余八个只用花 2RTT 时间即可。所以总时间：(n+2) RTT+16RTT = (n+18) RTT
b. 6 个并行，总共 9 个，两次额外时间即可 (n+2) RTT + 4RTT = (n+6) RTT
c. 持久化连接相比于非持久化连接也就是每次 transmission 减少了一个 RTT（建立连接的时间），所以是：(n+2) RTT + 8RTT = (n+10) RTT

P10. Consider a 30-meter link, over which a sender can transmit at a rate of 300 bits/sec in both directions. Suppose that packets containing data are 100,000 bits long, and packets containing only control (e.g., ACK or

handshaking) are 200 bits long. Assume that N parallel connections each get $1/N$ of the link bandwidth. Now, consider the HTTP protocol and suppose that each downloaded object is 100 Kbits long, and that the initial downloaded object contains 10 referenced objects from the same sender. Would parallel downloads via parallel instances of non-persistent HTTP make sense in this case? Now consider persistent HTTP. Do you expect significant gains over the non-persistent case? Justify and explain your answer.

30m link　300bits/s　包大小 100000bits　10 个 referenced objects
假设没有 propagation delay。
Parallel download

1. 非持久化连接：300bits/s 分带宽给 N 个 connection，一个收到 300/N bits/s
   200/300 * 3 + 100000/300 + 200/30 * 3 + 100000/30 = 3689 s
2. 持久化连接：200/300 * 3 +100000/300 + 10*(200/300 + 100000/300) = 3675 s

持久化连接更快，省去了很多建立连接的时间，但没有特别明显的优势。

P16. How does SMTP mark the end of a message body? How about HTTP? Can HTTP use the same method as SMTP to mark the end of a message body? Explain.

SMTP 在单独的行上添加一个单独的点. 来标记结尾
HTTP 协议中并没有明确的结尾标记来标识一个 HTTP 报文的结束。但它有内容长度标记，“Content-Length header field” 可以表示 message body 的长度，可以计算得到结尾地方。
SMTP 是 ascii 码编码形式才可以使用点来表示结尾，HTTP 的 message 是二进制形式。

P21. Suppose that your department has a local DNS server for all computers in the department. You are an ordinary user (i.e., not a network/system administrator). Can you determine if an external Web site was likely accessed from a computer in your department a couple of seconds ago? Explain.

直觉上来想普通（非特权）用户应该是无法获取到同地位的其它用户的访问信息的，包括时间。但是通过查询资料，有个工具叫 dig，可以查询本地 DNS 服务器的 website。
当 dig 一个网站时，首先将会返回找到这个网站的查询时间。如果该网站前几秒刚刚被访问过，那它的 entry 应该还被本地 DNS cache 保存着，而同一个机构同样的 DNS server 和 cache，因此当我此时访问时是直接从本地 cache 中拿到信息，响应时间应该是 0 ms，这样就说明前几秒刚有人访问了。

P25. Consider an overlay network with N active peers, with each pair of peers having an active TCP connection. Additionally, suppose that the TCP connections pass through a total of M routers. How many nodes and edges are there in the corresponding overlay network?

N 个 active peers 肯定是 N 个 nodes
每一对 peers 都有一个 active connection，所以全连接，N(N-1)/2 个 edges

P26. Suppose Bob joins a BitTorrent torrent, but he does not want to upload any data to any other peers (he wants to be a so-called free-rider).

a. Alice who has been using BitTorrent tells Bob that he cannot receive a complete copy of the file that is shared by the swarm. Is Alice correct or not? Why?

b. Charlie claims that Alice is wrong and that he has even been using a collection of multiple computers (with distinct IP addresses) in the computer lab in his department to make his downloads faster, using some additional coordination scripting. What could his script have done?

a. incorrect. P2P 并不会强制要求用户必须上传东西才能下载东西。如果有人上传这个 file 且不要求 Bob 上传啥东西，Bob 就可以直接下载得到。

b. 并行化下载加速，分配这个 file 的不同部分给不同的主机下载。可能抢占 department 中其它人的带宽。