

CS118 Homework 2

Prithvi Kannan UID: 405110096

Problem 1

Amazon identifies users when shopping through cookies. Cookies are stored locally in the web browser as an ID that Amazon can use to personalize the shopping experience for that user/device/account. The web sever does this by creating a cookie within Amazon's database the first time the user visits the website. The cookie is sent back to the client in the response so that it can be stored in the client's browser. The next time the client visits Amazon, it will embed the cookie in the request, and Amazon can lookup the cookie and identify the user.

Problem 2

a. non-persistent HTTP

$$time_{IP} = RTT_1 + RTT_2 + \dots + RTT_n$$

$$time_{setup} = 2 * RTT_0$$

$$time_{objects} = 9 * 2 * RTT_0$$

$$time_{total} = RTT_1 + RTT_2 + \dots + RTT_n + 20 * RTT_0$$

b. non-persistent HTTP parallel up to 5

$$time_{IP} = RTT_1 + RTT_2 + \dots + RTT_n$$

$$time_{setup} = 2 * RTT_0$$

$$time_{objects} = 2 * 2 * RTT_0$$

$$time_{total} = RTT_1 + RTT_2 + \dots + RTT_n + 6 * RTT_0$$

c. persistent HTTP

$$time_{IP} = RTT_1 + RTT_2 + \dots + RTT_n$$

$$time_{setup} = 2 * RTT_0$$

$$time_{objects} = 9 * RTT_0$$

$$time_{total} = RTT_1 + RTT_2 + \dots + RTT_n + 11 * RTT_0$$

d. persistent HTTP arbitrary parallel

$$time_{IP} = RTT_1 + RTT_2 + \dots + RTT_n$$

$$time_{setup} = 2 * RTT_0$$

$$time_{objects} = RTT_0$$

$$time_{total} = RTT_1 + RTT_2 + \dots + RTT_n + 3 * RTT_0$$

Problem 3

SMTP uses "\r\n.\r\n." sequence, which represents a line containing only a period to indicate the end of a message body.

HTTP uses the **content-length** header field to indicate where the end of file is. There is no specific character for end of message body.

HTTP can't use the SMTP method because HTTP can transmit any data format, not just ASCII. Therefore, the SMTP ending sequence can actually be a message sent via HTTP, so HTTP must use the length to determine the end.

Problem 4

a. Yes, you can determine if the website was recently accessed by a computer in your department. Using the `dig` function, you can get the time the query took. If the time is very short (0 msec), then the website is in the DNS cache and was likely accessed recently.

b. As the administrator, you now have access to which website are being loaded into the DNS cache. A simple approach to determining the most popular web servers is to maintain a counter for each web server. For example, when a user visits `google.com`, increment the counter. By sorting this list by frequency, you can find the most popular websites accessed by your department.

Problem 5

$$F = 15Gbits, u_s = 30Mbps, d_{min} = d_i = 2Mbps$$

Client-Server:

$$D_{client-server} = \max\{NF/u_s, F/d_{min}\}$$

u\N	10	100
300 Kbps	7680	51200
2 Mbps	7680	51200

P2P:

$$D_{p2p} = \max\{F/u, F/d_{min}, \frac{NF}{u + \sum_{i=1}^N u_i}\}$$

u\N	10	100
300 Kbps	7680	25904
2 Mbps	7680	7680