COMP 431 — INTERNET SERVICES & PROTOCOLS

Spring 2020

Homework 4, February 21

Due: 10:30 am, February 28

Note: Late solutions for this assignment ***cannot*** be accepted.

Note: Please typeset your answers and submit on Gradescope. Please include detailed steps and explanations.

1. [5 pts] Suppose Alice want to send an email to a long-lost school friend, for whom she has just obtained a contact email address. Alice composes and send an email to that address. Discuss and list in detail the steps that would be involved starting from Alice's action above till her friend reads her email – in particular, list the series of all application-layer and transport layer protocols that are used in each step.

2. [6+6+2= 14 points] Consider a web page for a hypothetical company my*dog.com*. The HTML file for their main web page, *www.mydog.com*, is 20,000 bits and contains the following embedded URLs (*i.e.*, the following URLs are contained in the HTML source for *www.mydog.com*):

http://www.mybird.com/chirp-add.jpg	(10,000 bits)
http://www.mycat.com/meow-add.jpg	(10,000 bits)
http://www.mydog.com/dog-banner.jpg	(15,000 bits)
http://www.mycat.com/puss-chow-add.jpg	(10,000 bits)
http://www.mydog.com/dog-chow-add.jpg	(10,000 bits)
http://www2.mydog.com/cookie.crumb	(5,000 bits)

Consider the operation of the HTTP protocol to download the base page (*www.mydog.com*) and the embedded objects. Assume:

i) All network links connecting the browser to each server operate at 100 Mbps.

ii) The browser has an internal (empty) browser cache and that there is no proxy cache on the network.

iii) The round trip time from the browser to all servers in the *mydog.com* domain is 50 ms, the round trip time from the browser to *www.mycat.com* is 10 ms, and the round trip time from the browser to *www.mybird.com* is 20 ms.

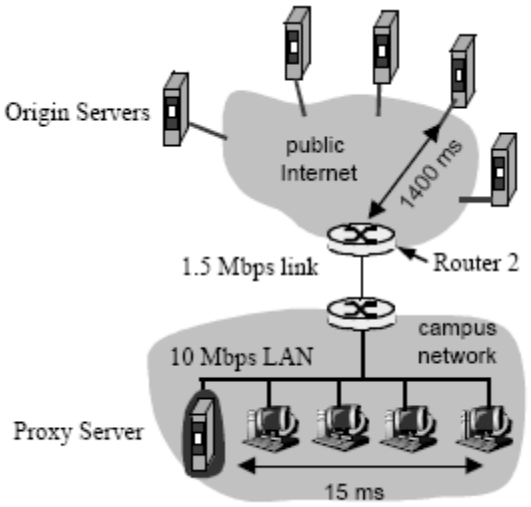
How long would it take to download the dog.com home page and all the embedded objects if the browser used:

a) Up to 4 parallel, non-persistent connections.

b) Persistent, non-parallel connections.

In answering these questions you should start with the time required for the initial TCP handshake (and assume here that the transmission time for the handshake is negligible). You may also ignore the overhead of DNS name resolution, assume there is no queuing or routing delay in the network, and assume that all objects are transferred in one network-layer frame.

3. [8+2 =10 points] Consider the network configuration for a small corporate campus shown in the figure below. The only traffic on the corporate network is generated by browsers accessing web objects stored by origin servers on the public Internet. A web proxy server with a cache has been installed on the corporate network and all the browsers configured to send HTTP/1.0 requests to the proxy server when the requested object cannot be found in the browser's local cache.



Measurements on this network indicate that the company browsers in aggregate send 100 HTTP requests per second to the proxy server and the mean size of all HTTP responses is 40,000 bits (the size of HTTP requests is negligible relative to the size of the responses).

If the requested object is found in the proxy's cache, the mean request/response latency for accessing the object is 15 milliseconds (including TCP connection overhead). If the requested object is not found in the proxy's cache, the proxy server makes an HTTP/1.0 request to the appropriate origin server and receives the object which is then added to the cache before the response is returned to the browser.

a. Assume that the mean HTTP request/response latency between Router 2 and all origin servers is 1,400 milliseconds (including TCP connection delays). Suppose we want to achieve an overall mean request/response latency of 450 milliseconds or less for all requests sent by the browsers to the proxy server. What is the smallest cache hit ratio at the proxy cache that will yield 450 milliseconds of latency for all requests? You may assume queuing and processing delays in routers 1 and 2 and all TCP connection establishment delays are negligible.

b. Given the hit ratio obtained in part 1, is the assumption of negligible queuing delays at router 2 valid or invalid? Give a clear and explicit analysis and justification for your answer.

4. [6+4 = 10 pts] Consider a web page for a mythical company ***catmyth.com*** that contains the following embedded URLs (*i.e.*, the following URLs are contained in the HTML source for the dogmyth.com home page):

http://www.catmyth.com/logo.gif

http:// www.ads.catmyth.com/dog-banner.jpg

http://www.ads.cat.biz/meow-add.jpg

http://www.ads.bird.biz/chirp-add.jpg

http://www.cat.biz/happy-cat.jpg

http://www.catmyth.com/happy-dog.jpg

http://www2.catmyth.com/cookie.crumb

Consider the DNS name resolution that is required for a browser to locate and request all components of this web page beginning with the home page URL *http://www.catmyth.com*. Assume there is no browser cache of web objects, no web proxy server, and the browser's resolver does not cache any DNS results. In answering the following questions you may assume:

- The cache for the local authoritative DNS server is initially empty and no other resolvers attempt to use that server until all these HTTP requests are completely satisfied.

- The local authoritative DNS server always sends iterated queries.
- For the domain names in the above URLs, there is a different authoritative DNS server for each unique domain in the name hierarchy. Domain names with a prefix of "www" are all assigned to host machines.

(a) Show the contents of the cache (list each RR separately) on the local authoritative DNS server used by a browser accessing the dogmyth.com home page after the page has been entirely fetched. For each RR entry in the cache, show only the values of the name and type fields for the entry.

(b) How many DNS query/reply protocol message exchanges with other DNS servers are processed by the local authoritative DNS server during the process of fetching all components of the dogmyth.com home page? Explain how you obtained your answer.

5. [4+4+2 = 10 pts] Suppose a server wants to distribute a file of F bits to N peers using a P2P architecture. Assume a fluid model. For simplicity, assume that d_{min} is very large, so that peer download bandwidth is never a bottleneck.

a) Suppose that $u_s \leq (u_s + u_1 + \ldots + u_N)/N$. Specify a distribution scheme that has a distribution time of F/u_s.

[Hint: What if that the server divides the file into N parts, with i^{th} part having size $(u_i/\Sigma_i u_i)F$, and transmits i^{th} part to peer i at a rate proportional to the part size. Also what if each peer forwards the bits it receives to each of the other peers at the same rate. Try showing that this scheme is feasible (it can be handled by all links), and compute the total download time with this scheme.]

b) Suppose that $u_s \geq (u_s + u_1 + \ldots + u_N)/N$. Specify a distribution scheme that has a distribution time of $NF/(u_s + u_1 + \ldots + u_N)$.

[Hint: What if that the server divides the file into N+1 equal parts, and sends i^{th} part to peer i at a rate $r_i = u_i/(N-1)$. Also what if each peer forwards the bits it receives to each of the other peers at the same rate. Additionally, the server sends the $(N+1)^{th}$ part at rate r_{N+1} to each of the N peers, who do not forward these bits. Try figuring out what r_{N+1} should be so that this scheme is feasible (it can be handled by all links), and leads to the smallest download time.]

c) Conclude that the minimum distribution time is in general given by $\max\{F/u_s, NF/(u_s + u_1 + \ldots + u_N)\}$.

6. [12 pts] Review the plot in the slide numbered 9 in the P2P/CDNs slide-set, which quantitatively compares how the minimum download time scales with N for a client-server vs a peer-to-peer design. Plot and discuss the minimum download time as a function of N for the following situations:

a) Client upload rate = u, $u_s = 5u$, F/u = 1 hour, $d_{min} >= u_s$.

b) Client upload rate = u, $u_s = u$, F/u = 1 hour, $d_{min} >= u_s$.

c) Client upload rate = u, $u_s = 5u$, F/u = 1 hour, $d_{min} = u_s/2$.

7. [2+2=4 pts] Consider a peer Bob that joins a BitTorrent torrent, but wants to do "free-riding" (does not want to upload any data to any other peers).

a) Is it possible for Bob to receive a complete copy of the file that is shared by the swarm? Why or why not?

b) Bob claims that he can use a collection of multiple computers (with distinct IP addresses) to make his free-riding more efficient. How can that happen?

A Reminder on the Honor Code

Students are *encouraged* to work together on this homework assignment. *Acceptable collaboration* includes:

- Discussing the assigned problems to understand their meaning,
- Discussing possible approaches to assigned problems,

In *all* cases you must *explicitly* acknowledge *any and all substantive help received from other individuals* during the course of the preparation of your homework solution. That is, if you collaborate with other individuals then you *must* include an explicit acknowledgment in your homework solution of the persons from whom you received aid. You should include the

acknowledgement with your Honor Code pledge. Acknowledging others, if done properly, will not adversely affect your grade.

Unacceptable collaboration on written homework includes:

- Copying (verbatim use) of physical papers or computer files,[1] and
- Submission of solutions that are jointly authored, or authored either wholly or in part by other individuals.

The general rule to be followed is that the strategy and approach of solutions may be developed jointly but *all* actual solutions (*i.e.*, the final solution) must be *constructed* and *written up* individually. Work done jointly should not be done in sufficient detail as to make it a final solution. For example, solutions may *sketched* out jointly, however each student must construct the final form of their solution individually and write-up their own solution. Should questions arise the course of working on a problem please feel free to immediately contact the instructor either by telephone, electronic mail, or by an office visit. In principle, if you work with others in good faith and are honest and generous with your attributions of credit you will have no problems.

[1] This includes computer files that are copied and then edited and/or reformatted.