CSC358 Tutorial 2

KyoKeun Park

January 24, 2020

University of Toronto Mississauga

Question 1: Concept Review

(a) While writing software at the application layer, do we need worry about the transport-layer protocols? Do we need to worry about the network-layer protocols? Why or why not?

(b) What do we mean by "HTTP is stateless"?

(c) What do we mean by "Cookies enable states"?

(d) What's the difference between persistent and non-persistent HTTP?

(e) What are the pros and cons of web cache (proxy)?

For the answers, review the lectures, books, go to office hours, and use the discussion board!

Question 2: Web Cache Response Time

Consider the network in the picture below. There is an institutional network connected to the public Internet through a 15 Mbps access link. Suppose that the average object size is 850 Kbits and that the average request rate from the institution's browser to the origin server is 16 requests per second. We model the total average response time as the sum of the *average access delay* (i.e. the delay from Internet router to institution router) and the *average Internet delay*. For the average access delay, use the following formula:

$$\text{average access delay} = \frac{\alpha}{1 - \alpha \cdot \beta}$$

where α is the average time required to send an object over the access link and β is the arrival rate of objects to the access link. The *Internet delay* is the amount of time it take from when the router on the Internet side of the access link forwards an HTTP request until it receives the response. We assume the Internet delay is 3 seconds on average.

(a) Find the total average response time given the above configuration (without a web cache)

(b) Now suppose a web cache is installed in the institutional LAN. Suppose the cache hit rate is 0.6. Find the total average response time again.

Question 2: Web Cache Response Time

(a) Find the total average response time given the above configuration (without web cache)

From the question, we know that average object size is 850 Kbits, and the access link is rated 15 Mbps. So for α we get:

Question 2: Web Cache Response Time

From the question, we know that average object size is 850 Kbits, and the access link is rated 15 Mbps. So for α we get:

$$\alpha = \frac{850 \text{ Kbits}}{15 \text{ Mbps}} = \frac{850000 \text{ bits}}{15000000 \text{ bits/sec}} = 0.0567 \text{ sec}$$

4

Question 2: Web Cache Response Time

(a) Find the total average response time given the above configuration (without web cache)

From the question, we know that average object size is 850 Kbits, and the access link is rated 15 Mbps. So for α we get:

$$\alpha = \frac{850 \text{ Kbits}}{15 \text{ Mbps}} = \frac{850000 \text{ bits}}{15000000 \text{ bits/sec}} = 0.0567 \text{ sec}$$

Furthermore, we know that average request rate is 16 requests per second. Hence $\beta = 16$ req/sec. We can calculate the traffic intensity of access link by:

$$\alpha \cdot \beta = 0.0567 \text{ sec/req} \cdot 16 \text{ req/sec} = 0.907$$

Question 2: Web Cache Response Time

(a) Find the total average response time given the above configuration (without web cache)

From the question, we know that average object size is 850 Kbits, and the access link is rated 15 Mbps. So for α we get:

$$\alpha = \frac{850 \text{ Kbits}}{15 \text{ Mbps}} = \frac{850000 \text{ bits}}{15000000 \text{ bits/sec}} = 0.0567 \text{ sec}$$

Furthermore, we know that average request rate is 16 requests per second. Hence $\beta = 16$ req/sec. We can calculate the traffic intensity of access link by:

$$\alpha \cdot \beta = 0.0567 \text{ sec/req} \cdot 16 \text{ req/sec} = 0.907$$

Thus, the average access delay becomes:

$$\text{average access delay} = \frac{\alpha}{1 - \alpha \cdot \beta}$$

Question 2: Web Cache Response Time

(a) Find the total average response time given the above configuration (without web cache)

From the question, we know that average object size is 850 Kbits, and the access link is rated 15 Mbps. So for α we get:

$$\alpha = \frac{850 \text{ Kbits}}{15 \text{ Mbps}} = \frac{850000 \text{ bits}}{15000000 \text{ bits/sec}} = 0.0567 \text{ sec}$$

Furthermore, we know that average request rate is 16 requests per second. Hence $\beta = 16$ req/sec. We can calculate the traffic intensity of access link by:

$$\alpha \cdot \beta = 0.0567 \text{ sec/req} \cdot 16 \text{ req/sec} = 0.907$$

Thus, the average access delay becomes:

$$\text{average access delay} = \frac{\alpha}{1 - \alpha \cdot \beta} = \frac{0.0567 \text{ sec}}{1 - 0.907} = 0.6 \text{ sec}$$

Question 2: Web Cache Response Time

(a) Find the total average response time given the above configuration (without web cache)

From the question, we know that average object size is 850 Kbits, and the access link is rated 15 Mbps. So for α we get:

$$\alpha = \frac{850 \text{ Kbits}}{15 \text{ Mbps}} = \frac{850000 \text{ bits}}{15000000 \text{ bits/sec}} = 0.0567 \text{ sec}$$

Furthermore, we know that average request rate is 16 requests per second. Hence $\beta = 16$ req/sec. We can calculate the traffic intensity of access link by:

$$\alpha \cdot \beta = 0.0567 \text{ sec/req} \cdot 16 \text{ req/sec} = 0.907$$

Thus, the average access delay becomes:

$$\text{average access delay} = \frac{\alpha}{1 - \alpha \cdot \beta} = \frac{0.0567 \text{ sec}}{1 - 0.907} = 0.6 \text{ sec}$$

Finally, we take consideration of the Internet delay of 3 seconds (from question):

$$3 \text{ sec} + 0.6 \text{ sec} = 3.6 \text{ sec}$$

(b) Now suppose a web cache is installed in the institutional LAN. Suppose the cache hit rate is 0.6. Find the total average response time again.

When web cache is installed in the institutional LAN, there are two possibilities upon request:

(b) Now suppose a web cache is installed in the institutional LAN. Suppose the cache hit rate is 0.6. Find the total average response time again.

When web cache is installed in the institutional LAN, there are two possibilities upon request:

1. Web cache misses and we need to forward the request to access link

(b) Now suppose a web cache is installed in the institutional LAN. Suppose the cache hit rate is 0.6. Find the total average response time again.

When web cache is installed in the institutional LAN, there are two possibilities upon request:

1. Web cache misses and we need to forward the request to access link

2. Web cache hits and the request can be resolved within LAN

Question 2: Web Cache Response Time

(b) Now suppose a web cache is installed in the institutional LAN. Suppose the cache hit rate is 0.6. Find the total average response time again.

Consider *cache miss case.* This will be similar to (a), as we would need to forward the request to access link and then wait for a response from Internet.

(b) Now suppose a web cache is installed in the institutional LAN. Suppose the cache hit rate is 0.6. Find the total average response time again.

Consider *cache miss case.* This will be similar to (a), as we would need to forward the request to access link and then wait for a response from Internet.

Note that the traffic intensity on the access link will now be reduced by 60%. (Why?)

Question 2: Web Cache Response Time

(b) Now suppose a web cache is installed in the institutional LAN. Suppose the cache hit rate is 0.6. Find the total average response time again.

Consider *cache miss case*. This will be similar to (a), as we would need to forward the request to access link and then wait for a response from Internet.

Note that the traffic intensity on the access link will now be reduced by 60%. (Why?)

Thus, average access delay becomes:

$$\frac{\alpha}{1 - 0.4 \cdot \alpha \cdot \beta} = \frac{0.0567 \text{ sec}}{1 - 0.4 \cdot 0.907} = 0.089 \text{ sec}$$

Question 2: Web Cache Response Time

(b) Now suppose a web cache is installed in the institutional LAN. Suppose the cache hit rate is 0.6. Find the total average response time again.

Consider *cache miss case.* This will be similar to (a), as we would need to forward the request to access link and then wait for a response from Internet.

Note that the traffic intensity on the access link will now be reduced by 60%. (Why?)

Thus, average access delay becomes:

$$\frac{\alpha}{1 - 0.4 \cdot \alpha \cdot \beta} = \frac{0.0567 \text{ sec}}{1 - 0.4 \cdot 0.907} = 0.089 \text{ sec}$$

Again, after considering Internet delay, we get:

$$3 \text{ sec} + 0.089 \text{ sec} = 3.089 \text{ sec}$$

(b) Now suppose a web cache is installed in the institutional LAN. Suppose the cache hit rate is 0.6. Find the total average response time again.

Now, let's consider *cache hit case.* Here, we would need to just communicate within institutional LAN:

(b) Now suppose a web cache is installed in the institutional LAN. Suppose the cache hit rate is 0.6. Find the total average response time again.

Now, let's consider *cache hit case.* Here, we would need to just communicate within institutional LAN:

$$\frac{850 \text{ Kbits}}{100 \text{Mbps}} = 8.5 \text{ ms}$$

(b) Now suppose a web cache is installed in the institutional LAN. Suppose the cache hit rate is 0.6. Find the total average response time again.

Now, let's consider *cache hit case.* Here, we would need to just communicate within institutional LAN:
$$\frac{850 \text{ Kbits}}{100 \text{Mbps}} = 8.5 \text{ ms}$$

We know that *cache miss case* occurs 40% of the time and *cache hit case* occurs 60% of the time. So the average response time is:

Question 2: Web Cache Response Time

(b) Now suppose a web cache is installed in the institutional LAN. Suppose the cache hit rate is 0.6. Find the total average response time again.

Now, let's consider *cache hit case.* Here, we would need to just communicate within institutional LAN:

$$\frac{850 \text{ Kbits}}{100 \text{Mbps}} = 8.5 \text{ ms}$$

We know that *cache miss case* occurs 40% of the time and *cache hit case* occurs 60% of the time. So the average response time is:

$$0.6 \cdot 0.0085 \text{ sec} + 0.4 \cdot 3.089 \text{ sec} = 1.24 \text{ sec}$$

Consider a link over which a host can transmit at a rate of 150 bps in both directions. Suppose packets containing only *control* (e.g., HTTP request header, handshaking for setting up the connection) are 200 bits long, and packets containing *data* are 100Kbits long (including both header and body of the HTTP response). Assume that if a host uses N parallel connections then each gets $\frac{1}{N}$ of the link bandwidth. Now consider the HTTP protocol, and suppose that each downloaded object is 100 Kbits long (including both header and body of the HTTP response), and that the initial downloaded object contains 10 referenced objects from the same host.

(a) Using non-persistent HTTP without parallel connections, what's the total time taken to download all objects?

(b) How about using persistent HTTP without parallel connections?

(c) How about using non-persistent HTTP with 10 parallel connections?

(d) Comparing the strategies in (b) and (c), which one allows faster download of all objects? Is the difference significant?

(e) Suppose that the link is shared by Alice with four other users. Alice uses parallel instances of non-persistent HTTP, and the other four users use non-persistent HTTP without parallel downloads. Do Alice's parallel connections help her get web pages more quickly? Explain your answer.

(f) If the other four users now start opening parallel instances of non-persistent HTTP, then would Alice's parallel connections still be beneficial? Explain your answer.

(g) Now take the *propagation delay* into account and redo the analysis in Part (d). Assume that the propagation speed is 2×10^8 m/sec. Do some calculations and discuss.

(h) Suppose you are developing a web browser, should you implement parallel connections or not?

First, let us consider the RTT for both connection setup and HTTP request/response with the above setup.

$RTT_{connsetup}$:

RTT_{http}:

First, let us consider the RTT for both connection setup and HTTP request/response with the above setup.

$RTT_{connsetup}$: Need 2 *control* packets (why?)

$$\frac{2 \cdot 200 \text{ bits}}{150 \text{ bps}}$$

RTT_{http}:

First, let us consider the RTT for both connection setup and HTTP request/response with the above setup.

$RTT_{connsetup}$: Need 2 *control* packets (why?)

$$\frac{2 \cdot 200 \text{ bits}}{150 \text{ bps}}$$

RTT_{http}: Need a *control* packet and *data* packet (why?)

$$\frac{(200 + 100000) \text{ bits}}{150 \text{ bps}}$$

(a) Using non-persistent HTTP without parallel connections, what's the total time taken to download all objects?

(a) Using non-persistent HTTP without parallel connections, what's the total time taken to download all objects?

$$11 \cdot (RTT_{connsetup} + RTT_{http})$$ (why are there 11 connection setups?)

(a) Using non-persistent HTTP without parallel connections, what's the total time taken to download all objects?

$$11 \cdot (RTT_{connsetup} + RTT_{http}) \quad \text{(why are there 11 connection setups?)}$$
$$=11 \cdot \frac{(3 \cdot 200 + 100000) \text{ bits}}{150 \text{ bps}}$$

(a) Using non-persistent HTTP without parallel connections, what's the total time taken to download all objects?

$$11 \cdot (RTT_{connsetup} + RTT_{http}) \quad \text{(why are there 11 connection setups?)}$$
$$= 11 \cdot \frac{(3 \cdot 200 + 100000) \text{ bits}}{150 \text{ bps}}$$
$$= 7773.33 \text{ sec}$$

(b) How about using persistent HTTP without parallel connections?

(b) How about using persistent HTTP without parallel connections?

$$RTT_{connsetup} + 11 \cdot RTT_{http}$$

(b) How about using persistent HTTP without parallel connections?

$$RTT_{connsetup} + 11 \cdot RTT_{http}$$
$$= \frac{(2 \cdot 200 + 11 \cdot (200 + 100000)) \text{ bits}}{150 \text{ bps}}$$

(b) How about using persistent HTTP without parallel connections?

$$RTT_{connsetup} + 11 \cdot RTT_{http}$$
$$= \frac{(2 \cdot 200 + 11 \cdot (200 + 100000)) \text{ bits}}{150 \text{ bps}}$$
$$= 7350.67 \text{ sec}$$

(c) How about using non-persistent HTTP with 10 parallel connections?

For this case, we must first receive the HTTP file in order to initiate parallel connections for the referenced objects:

(c) How about using non-persistent HTTP with 10 parallel connections?

For this case, we must first receive the HTTP file in order to initiate parallel connections for the referenced objects:

$$\frac{(3 \cdot 200 + 100000) \text{ bits}}{150 \text{ bps}} + \frac{(3 \cdot 200 + 100000) \text{ bits}}{\frac{150}{10} \text{ bps}}$$

(why divide the bandwidth by 10 here?)

(c) How about using non-persistent HTTP with 10 parallel connections?

For this case, we must first receive the HTTP file in order to initiate parallel connections for the referenced objects:

$$\frac{(3 \cdot 200 + 100000) \text{ bits}}{150 \text{ bps}} + \frac{(3 \cdot 200 + 100000) \text{ bits}}{\frac{150}{10} \text{ bps}}$$

(why divide the bandwidth by 10 here?)

$$=7377.33 \text{ sec}$$

(c) How about using non-persistent HTTP with 10 parallel connections?

For this case, we must first receive the HTTP file in order to initiate parallel connections for the referenced objects:

$$\frac{(3 \cdot 200 + 100000) \text{ bits}}{150 \text{ bps}} + \frac{(3 \cdot 200 + 100000) \text{ bits}}{\frac{150}{10} \text{ bps}}$$

(why divide the bandwidth by 10 here?)

$$=7377.33 \text{ sec}$$

(d) Comparing the strategies in (b) and (c), which one allows faster download of all objects? Is the difference significant?

(e) Suppose that the link is shared by Alice with four other users. Alice uses parallel instances of non-persistent HTTP, and the other four users use non-persistent HTTP without parallel downloads. Do Alice's parallel connections help her get web pages more quickly? Explain your answer.

(f) If the other four users now start opening parallel instances of non-persistent HTTP, then would Alice's parallel connections still be beneficial? Explain your answer.

(e) Suppose that the link is shared by Alice with four other users. Alice uses parallel instances of non-persistent HTTP, and the other four users use non-persistent HTTP without parallel downloads. Do Alice's parallel connections help her get web pages more quickly? Explain your answer.

Yes. Alice has more connections. Since the bandwidth is equally shared across all connections, Alice will therefore have more share of the bandwidth.

(f) If the other four users now start opening parallel instances of non-persistent HTTP, then would Alice's parallel connections still be beneficial? Explain your answer.

(e) Suppose that the link is shared by Alice with four other users. Alice uses parallel instances of non-persistent HTTP, and the other four users use non-persistent HTTP without parallel downloads. Do Alice's parallel connections help her get web pages more quickly? Explain your answer.

Yes. Alice has more connections. Since the bandwidth is equally shared across all connections, Alice will therefore have more share of the bandwidth.

(f) If the other four users now start opening parallel instances of non-persistent HTTP, then would Alice's parallel connections still be beneficial? Explain your answer.

Yes. Without the parallel connections, Alice would be at a disadvantage in terms of bandwidth share.

(g) Now take the propagation delay into account and redo the analysis in Part (d). Assume that the propagation speed is 2×10^8 m/sec. Do some calculations and discuss.

Let the wire length be D, and let $t_{prop} = \frac{D}{2 \cdot 10^8}$ m/sec be the one-way propagation delay.

(g) Now take the propagation delay into account and redo the analysis in Part (d). Assume that the propagation speed is 2×10^8 m/sec. Do some calculations and discuss.

Let the wire length be D, and let $t_{prop} = \frac{D}{2 \cdot 10^8}$ m/sec be the one-way propagation delay. We need to think about total distance travelled in each case. Parallel trips counts as one.

(g) Now take the propagation delay into account and redo the analysis in Part (d). Assume that the propagation speed is 2×10^8 m/sec. Do some calculations and discuss.

Let the wire length be *D*, and let $t_{prop} = \frac{D}{2 \cdot 10^8}$ m/sec be the one-way propagation delay. We need to think about total distance travelled in each case. Parallel trips counts as one.

For persistent HTTP without parallel connections, 2 t_{prop} is required for setup connection, and then 2 for each 11 objects.

(g) Now take the propagation delay into account and redo the analysis in Part (d). Assume that the propagation speed is 2×10^8 m/sec. Do some calculations and discuss.

Let the wire length be *D*, and let $t_{prop} = \frac{D}{2 \cdot 10^8}$ m/sec be the one-way propagation delay. We need to think about total distance travelled in each case. Parallel trips counts as one.

For persistent HTTP without parallel connections, 2 t_{prop} is required for setup connection, and then 2 for each 11 objects. Hence, there will be 24 · t_{prop} total propagation delay.

(g) Now take the propagation delay into account and redo the analysis in Part (d). Assume that the propagation speed is 2×10^8 m/sec. Do some calculations and discuss.

Let the wire length be *D*, and let $t_{prop} = \frac{D}{2 \cdot 10^8}$ m/sec be the one-way propagation delay. We need to think about total distance travelled in each case. Parallel trips counts as one.

For persistent HTTP without parallel connections, 2 t_{prop} is required for setup connection, and then 2 for each 11 objects. Hence, there will be $24 \cdot t_{prop}$ total propagation delay.

For non-persistent HTTP with 10 parallel connections, we need 4 t_{prop} for the first object, and then 4 t_{prop} for the other 10 objects. Hence, there will be $8 \cdot t_{prop}$ total propagation delay.

(g) Now take the propagation delay into account and redo the analysis in Part (d). Assume that the propagation speed is 2×10^8 m/sec. Do some calculations and discuss.

Now, let the total delay of (b) equal to (c) and solve for critical t_{prop}:

$$7350.67 + 24 \cdot t_{prop} = 7377.33 + 8 \cdot t_{prop}$$

(g) Now take the propagation delay into account and redo the analysis in Part (d). Assume that the propagation speed is 2×10^8 m/sec. Do some calculations and discuss.

Now, let the total delay of (b) equal to (c) and solve for critical t_{prop}:

$$7350.67 + 24 \cdot t_{prop} = 7377.33 + 8 \cdot t_{prop} \implies t_{prop} = 1.67 \text{ sec}$$

(g) Now take the propagation delay into account and redo the analysis in Part (d). Assume that the propagation speed is 2×10^8 m/sec. Do some calculations and discuss.

Now, let the total delay of (b) equal to (c) and solve for critical t_{prop}:

$$7350.67 + 24 \cdot t_{prop} = 7377.33 + 8 \cdot t_{prop} \implies t_{prop} = 1.67 \text{ sec}$$

Using the $t_{prop} = \frac{D}{2 \cdot 10^8}$ formula, we can solve for critical D:

(g) Now take the propagation delay into account and redo the analysis in Part (d). Assume that the propagation speed is 2×10^8 m/sec. Do some calculations and discuss.

Now, let the total delay of (b) equal to (c) and solve for critical t_{prop}:

$$7350.67 + 24 \cdot t_{prop} = 7377.33 + 8 \cdot t_{prop} \implies t_{prop} = 1.67 \text{ sec}$$

Using the $t_{prop} = \frac{D}{2 \cdot 10^8}$ formula, we can solve for critical D:

$$D = 1.67 \cdot 2 \cdot 10^8 = 3.34 \cdot 10^8 \text{ metres}$$

(g) Now take the propagation delay into account and redo the analysis in Part (d). Assume that the propagation speed is 2×10^8 m/sec. Do some calculations and discuss.

Now, let the total delay of (b) equal to (c) and solve for critical t_{prop}:

$$7350.67 + 24 \cdot t_{prop} = 7377.33 + 8 \cdot t_{prop} \implies t_{prop} = 1.67 \text{ sec}$$

Using the $t_{prop} = \frac{D}{2 \cdot 10^8}$ formula, we can solve for critical D:

$$D = 1.67 \cdot 2 \cdot 10^8 = 3.34 \cdot 10^8 \text{ metres}$$

This means that if wire is shorter than 334,000,000 metres, (b) will have smaller delay. Otherwise, (c) will have smaller delay.

(h) Suppose you are developing a web browser, should you implement parallel connections or not? Have a discussion.

Hint: Consider our answers from (b) to (g). In what situation would parallel connections perform better?