Problem 1:

a) During a period of 1 second, 850000 × 16 bits are to be sent from Internet to Institutional router. Hence,

Δ=850000 / 100 × 10^6 = 0.0085 second

average access delay = $\frac{\Delta}{(1-\Delta\times\beta)} = \frac{0.0085}{(1-0.0085\times16)}$ = 0.00984 second

Using the formula:

total average response time = *average access delay* + *average Internet delay*

= 3.00984 seconds

b) *miss rate* = 0.4

Let *X* be the response time if the required file is not found in the institutional cache. Let *Y* be the response time if the required file is found in the institutional cache. Hence, we have:

total response time = 0.4*X* + 0.6*Y*

Now compute *X*, *Y* separately

$$X = \frac{\Delta}{(1-\Delta\times\beta)} + 3 = \frac{850000/100\times10^6}{(1-\frac{850000}{10^8}\times16)} + 3 = 3.00984 seconds$$

$$Y = \frac{850000}{1000} = 0.00085 seconds$$

Hence, we have

total response time = 0.4 *X* + 0.6 *Y* = 1.204446 seconds

Problem 2:

a) $T_{tp} = 2T_p$

b) $T_{tf} = B_f / R$

c) $T_{ti} = B_i / R$

d) $1T_{tp} = 2T_p$

e) We need the following times:

- $1T_{tp} = 2T_p$ for the connection setup; this will be repeated (*m*+1) times,
- $1T_{tp} + T_{tf} = 2T_p + B_f / R$ for requesting and receiving the base html file,

- $1T_{tp} + T_{ti} = 2T_p + B_i / R$ for requesting and receiving each of the m embedded images,
- Then, the total delay will be $(m + 1)\ T_{tp} + (1T_{tp} + T_{tf}) + m\ (1T_{tp} + T_{ti})$.

Consequently, the response time is $4\ (m + 1)\ T_p + B_f / R + mB_i / R$

f) We need the following times:

- $1T_{tp} = 2T_p$ for the connection setup,
- $1T_{tp} + T_{tf} = 2T_p + B_f / R$ for requesting and receiving the base html file,
- $1T_{tp} + T_{ti} = 2T_p + B_i / R$ for requesting and receiving each of the m embedded images,
- Then, the total delay will be $1\ T_{tp} + (1T_{tp} + T_{tf}) + m\ (1T_{tp} + T_{ti})$.

Consequently, the response time is $2\ (m + 2)\ T_p + B_f / R + mB_i / R$

Problem 3:

a) First consider parallel downloads using non-persistent connections. Parallel downloads would allow 10 connections to share the 200 bps bandwidth, giving each just 20 bps. Thus, recalling that $2 \times t_p$ is the time required to open a connection, t_p is the time to send a request and $(t_p + t_f)$ (where t_f is the transmission delay) is the time to download an object, the total time needed to receive all objects is given by:

$$T = 2 \times t_p + t_p + (t_p + t_f) + 2 \times t_p + t_p + (t_p + t_f)$$

$$= 2 \times t_p + t_p + t_p + \frac{300{,}000}{200} + 2 \times t_p + t_p + t_p + \frac{300{,}000}{\frac{200}{10}} = 8 \times t_p + 16500\text{s}$$

b) Now consider a persistent HTTP connection. The total time needed is given by:

$$T = 2 \times t_p + t_p + (t_p + t_f) + 10 \times (t_p + t_p + t_f)$$

$$= 2 \times t_p + t_p + t_p + \frac{300{,}000}{200} + 10 \times (t_p + t_p + \frac{300{,}000}{200}) = 24 \times t_p + 16500\text{s}$$

Assuming the speed of light is $300 * 10^6$ m/sec, then $t_p = 10 / (300 \times 10^6) = 3.33 \times 10^{-8}$s.

t_p is therefore negligible compared with the transmission delay. Thus, we see that persistent HTTP achieves the same download time than the non-persistent HTTP with parallel download.

Problem 4:

a) There is one trip to the local DNS server and then there are trips from the local DNS server to the Root, TLD server and two authoritative name servers. So the total would be 2+10+10+10+10 = 42 msec.

b) As there is no need to visit the Root and TLD servers (because of caching), then there would be 1 visit to the local DNS server and two visits to the utoronto.ca and ece.utoronto.ca authoritative name servers so 2+10+10 = 22 msec.

Problem 5:

a) For the client-server file distribution:
First, the server needs to send 150 Gbits to all 1 million peers, thus

$$time1 = \frac{150\ G \times 10^6}{100Gbps} = 1.5 \times 10^6 seconds$$

Second, each peer needs to download the file, thus

$$time2 = \frac{150\ G}{2Mbps} = 7.5 \times 10^4 seconds$$

The minimal distribution time for the client-server approach is:

$$D_{cs} = max\{time1,\ time2\} = 1.5 \times 10^6\ seconds$$

b) For the P2P file distribution:

First, the server at least needs to send one copy of the file, thus

$$time1 = \frac{150\ G}{100Gbps} = 1.5 seconds$$

Second, each peer needs to download the file, thus

$$time2 = \frac{150\ G}{2Mbps} = 7.5 \times 10^4 seconds$$

Third, a total of 1 million copies of the file need to be uploaded, thus

$$time3 = \frac{150\ G \times 10^6}{100Gbps + 500Kbps \times 10^6} = 2.5 \times 10^5 seconds$$

The minimal distribution time for the P2P approach is:

$$D_{P2P} = max\{time1, time2, time3\} = 2.5 \times 10^5\ seconds$$