Tim Duffy Homework #3 02/24/11

1) The 97th percentile confidence interval for the response time of a database transaction system is given by the interval [2-0.2, 2+0.2] in seconds. This confidence interval was computed using 48 samples. Answer the following questions:

a. Based on the above information, what do you think the standard deviation of the response time is?

Given: X= 2; E = 0.2; N = 48

$P(Z < (E/\sigma/\sqrt{N})) > (0.97 = 1-a/2)$

$a = 0.03 \Rightarrow a/2 = 0.015 \Rightarrow Z = 2.17$

Using the equation $E = (Z*\sigma) / \sqrt{N}$ we can determine that: $\sigma = (E * \sqrt{N}) / Z$

$\sigma = (0.2 * \sqrt{48}) / 2.17 =$ **0.6385**

b. Compute a 99th percentile confidence interval for this system.

Given: X=2; N=48; $\sigma = 0.6385$

$P(Z < (E/\sigma/\sqrt{N})) > (0.99 = 1-a/2)$

$a = 0.01 \Rightarrow a/2 = 0.005 \Rightarrow Z = 2.575$

$E = (Z*\sigma) / \sqrt{N} = (2.575 * 0.6385) / \sqrt{48} =$ **0.2373**

The 99th percentile confidence interval is [2-0.2373, 2+0.2373]

c. How many more samples do you need to make the 97th percentile confidence interval be [2-0.05, 2+0.05] ?

Given: X= 2; E = 0.05; $\sigma = 0.6385$

Z = (Value for z where area under curve > 0.97) = 1.89

$E = (Z*\sigma) / \sqrt{N}$ => $0.05 = (1.89 * 0.6385) / \sqrt{N}$

=> $\sqrt{N} = (1.89 * 0.6385) / 0.05 = 24.1353$

=> $N = (24.1353)^2 = 582.51 \approx 583$

You would need **535 more samples** to make the 97th percentile confidence interval be [2-0.05, 2+0.05].

d. What is the probability that the mean response time for 100 new transactions will be larger than 2.15 seconds?

2) The inter-arrival time of packets coming into a gateway is exponentially distributed, with an average incoming rate of 1,200,000 packets/min. Its outgoing transmission line has a rate of 24,000,000 bytes/sec. Packet lengths are distributed exponentially with the average packet size of 1,000 bytes.

a. What kind of queuing model do you think is an appropriate model for this gateway?

This would be best described as an M/M/1 system.

b. What is the mean time it takes a packet to be sent over the transmission line?

ρ = (1,200,0000 packets / min) / (24,000,000 bytes / sec) =
= (1,200,000 * 1,000 bytes) / 60 sec/min) / (24,000,000 bytes / sec)
= 20,000,000 / 24,000,000 = 0.8333

q = ρ / (1 - ρ) = 4.999
Tq = q /λ = 4.999 / 200,000 (packets/sec) = **0.025 msec**

c. What is the mean number of packets in the gateway?

q = ρ / (1 - ρ) = **4.999 packets**

d. What is the mean number of packets in the wait queue?

q = s + w => 4.999 = 1 + w => **w = 3.999 packets**

e. What is the average latency?

Tq = **0.025 msec**

f. What is the time that a packet spends in the wait queue?

Tw = w / λ = 3.999 / 200,000 = **0.02 msec**

g. What is the chance that a packet coming into the router will not have to wait in queue?

P(empty queue) = e^-λ = e^-200000 = 0

Due to the extremely high rate of incoming packets (200,000 / sec) there is no chance of finding an empty queue.

3) The buffer of a streaming media application can hold up to N packets. The processing time for each packet to be played out was found to be exponential with a mean of 30 msec. Degradation in the quality of the playout occurs as a result of buffer over-runs or as a result of a buffer under-runs. Buffer over-runs occur when a packet is dropped due to a filled buffer (causing a “blip” in the playout). Buffer under-runs occur when there are no packets to playout (causing a period of “plop” in the playout). It was measured that the packets arrive to the application as a Poisson process with a mean of 30 packets per second. Answer the following questions:

Given: λ = 30 packets / sec Ts = 30 msec = 0.03 sec

a. Assuming an infinite buffer, how many packets do you expect to find in the buffer (i.e. waiting to be played out)?

$\rho = \lambda * Ts = 30 * 0.03 = 0.9$
$q = \rho / (1- \rho) = 0.9 / 0.1 = 9$

Since each packet is processed individually (M/M/1), all but 1 packet in the system is in the queue.

w = packets in the queue = **8 packets**

b. Assuming an infinite buffer, what is the mean delay between receipt of a packet and its playout?

$Tq = Tw + Ts = (w/ \lambda) + Ts$ = (8 packets / 30 packets / sec) + 0.03 sec = 0.267 + 0.03

Tq = 0.297 seconds = 297 msec

c. Assuming an infinite buffer and assuming that a streaming media object consists of 1,000 packets, how many “blips” and how many “plops” do you expect to hear for that object?

P(full buffer) = 0 With an infinite buffer, there will be **no blips.**

P(empty system) = $1 - \rho = 0.1$ Out of 1000 packet arrivals, there will be **100 plops**

d. Assuming N=5 packets, how many packets do you expect to find in the buffer (i.e. waiting to be played out)?

q = [ρ / (1- ρ)] - [(K+1)*ρ^(K+1) / (1 - ρ^(K+1))]
q = (0.9 / 0/1) - (6*0.9^6 / 1 - 0.9^6)
q = 9 - (3.188646 / 0.468559) = 2.19

w = q - 1 = 1.19 packets in the buffer.

e. Assuming N=5 packets, what is the mean delay between receipt of a packet and its playout?

$$Tq = Tw + Ts = (1.19/30 \text{ packets/sec})+0.03 \text{ sec} = \mathbf{0.06967\ seconds = 69.67\ msec}$$

f. Assuming N=5 packets and that a streaming media object consists of 1,000 packets, how many “blips” and how many “plops” do you expect to hear for that object?

$$P(\text{buffer is full}) = ((1-\rho)*p^K) / (1-\rho^{K+1}) = (0.1 * 0.9^6) / (1 - 0.9^6)$$
$$= 0.113420295$$

So, out of 1000 packets, there will be **about 114 blips**

$P(\text{empty system}) = 1 - \rho = 0.1$ => **100 plops**

4) A web site has a 2-tier architecture. The first "tier" acts as a firewall, which "rejects" requests that are deemed to be generated from hostile sources (e.g. requests that are identified as belonging to a source involved in a denial-of-service attack, for example). The second "tier" is the web server, which actually responds to the requests in earnest.

Assume that "legitimate" requests are Poisson distributed with rate L1 and that "hostile" requests are also Poisson distributed with rate L2. Also, assume that:

- The time it takes the firewall to either forward a request to the server or reject it is exponential with mean=5msec.
- The firewall rejects 95% of the "hostile" requests (i.e. it mis-classifies 5% of the hostile requests as legitimate).
- The time it takes the server to process a request (whether hostile or not) is 50msec.

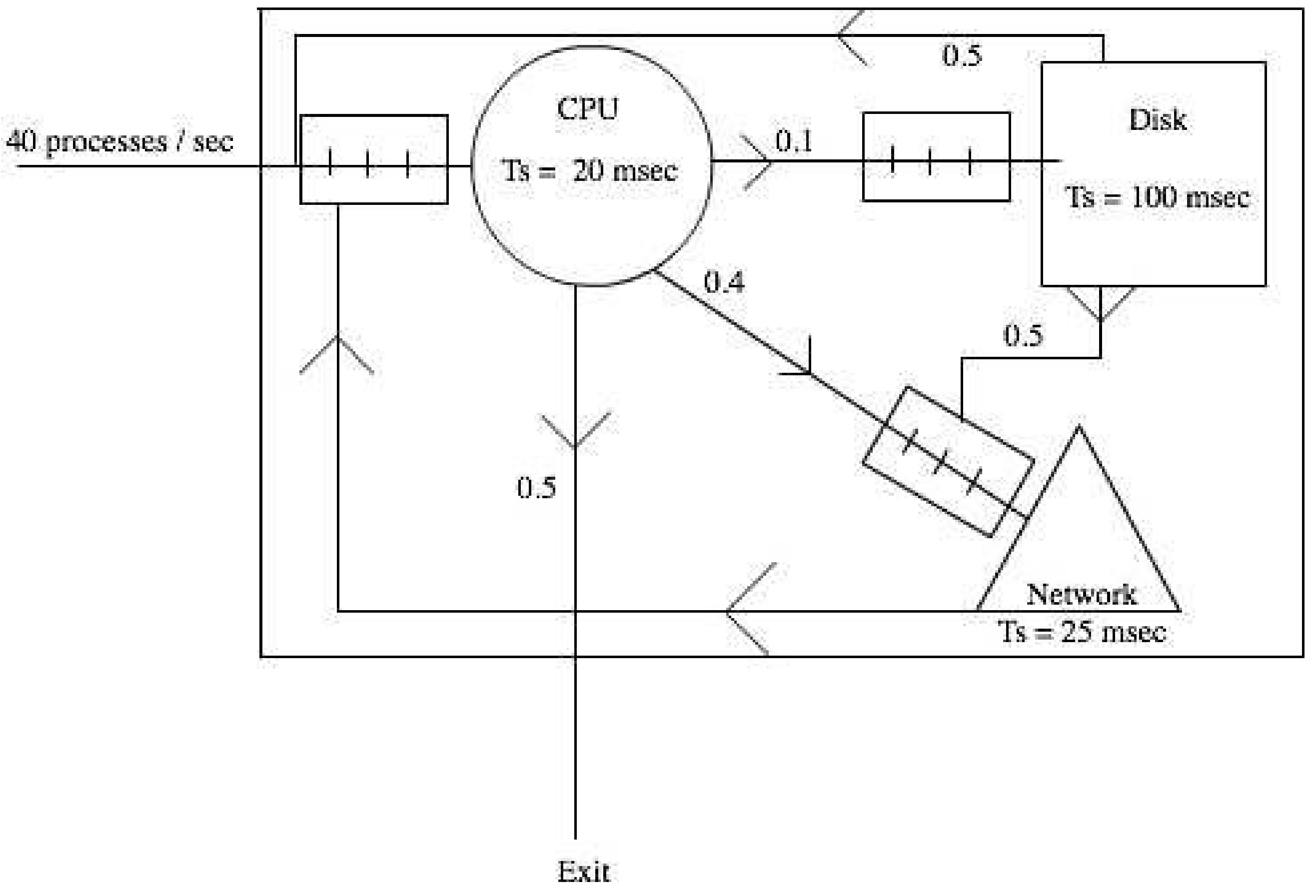
What conditions (spelled out in terms of L1 and L2) would this web site buckle under a denial of service attack?

ρ (Tier 1) = (L1 + L2) * 5 msec
ρ (Tier 2) = (L1 + 0.5*L2) * 50 msec
ρ (System) = ρ (Tier 1) + ρ (Tier 2)

So the website would buckle under a DoS attack if $\rho > 1$, or:

1 < ((L1 + L2) * 5 msec) + ((L1 + 0.5*L2) * 50 msec) =>
1 < (5*L1 + 5*L2) + (50*L1 + 25*L2) =>
1 < (55*L1 + 30*L2)

5) a. Draw a queuing network representation of the system.



b. Find the average turnaround time (i.e. total response time) for processes submitted to the above system.

CPU: 40	processes/sec = 0.5X	=>	λcpu = 80 processes/sec
	Ts = 20 msec = 0.02 sec	=>	ρ = (80 * 0.02) / 2 cores = 0.8
	q = ρ / (1- ρ)	=>	q = 0.8/0.2 = 5
Disk:	λdisk = (0.1 * λcpu)	=>	λdisk = 8 processes/sec
	Ts = 100 msec = 0.1 sec	=>	ρ = (8 * 0.1) = 0.8
	q = ρ / (1- ρ)	=>	q = 0.8/0.2 = 5
Network:	λnetwork = (0.5 * λdisk) + (0.4 λcpu) = 4 + 32 = 36 processes/sec		
	Ts = 25 msec = 0.025 sec	=>	ρ = (36 * 0.025) = 0.9
	q = ρ / (1- ρ)	=>	q = 0.9/0.1 = 9

System: **Tq-system =** Tq-cpu + Tq-disk + Tq-network
= (λcpu * qcpu) + (λdisk * qdisk) + (λnetwork * qnetwork)
= (80 * 5) + (8 * 5) + (36 * 9) =
= 400 + 40 + 324 = **746 msec**

c. What service (CPU, Disk, or network) is the "bottleneck" in the above system? Why?

The CPU is the bottleneck in the system. This is interesting since it has the smallest service time. However, since it must serve a very large number of processes each second, the CPU causes the system to bottleneck.

d. What process arrival rate will render this system unstable, i.e., will make one of the queues grow infinitely long?

If the utilization of one process exceeds 100%, its queue will grow infinitely long. Since the CPU is the bottleneck of the system, it will be be most effected by an increased arrival rate.

Since the CPU has 2 cores, utilization must exceed 2. We can use the fact that $\rho = \lambda * Ts$ to determine the arrival rate at which the system fails.

$\rho = X * Ts \quad => \quad 2 = X * 0.02 \quad => \quad X = 2 / 0.02 = 100$

But remember, because some processes return to the CPU, this is not the actual arrival rate!

$0.5\ \lambda = X => \lambda = X/2 \quad => \lambda = 50$

Any arrival rate **greater than 50 processes/second** will cause the system to become unstable.