

Enterprise-Scale Software Development (COMP5348)

Semester 1, 2009

Tutorial Week 6

Part A: Paper-based exercises

Q1: A disk is observed for 10 minutes, during which time 1200 tasks are completed by the whole system. The disk is observed to perform 2400 access, and the average time to service an access is 0.1s. What is the utilization of the disk? What is the service demand per task, on the disk? What is the average time between access completions, at the disk? *In 10min=600s, the disk performs 2400 accesses each taking 0.1s. Thus the disk is busy for 240s in 600s, so its utilization is $240/600=40\%$. The work done (240s) is performed for 1200 tasks, so each task has $240/1200 = 0.2$ s of service demand on the disk (alternative argument: each task needs 2 access, and each access takes 0.1s, so each task demands $2*0.1=0.2$ s of service from the disk). The disk completes 2400 accesses in 600s, so it completes 4 access per second, or equivalently the gap between access completions is 0.25s on average.*

Q2. Consider an open network consisting of two devices. Packets arrive at device 1 from the outside with a rate of 20 jobs/sec. Whenever a packet has been processed by device 1, it will leave the system with probability 0.5, and otherwise (probability 0.5) it moves to device 2. When a packet has been serviced at device 2, it leaves the system with probability 0.25 and moves to device 1 otherwise. What is the average number of visits to each device, for an arriving packet?

Let A be the average services per second at device 1, and let B be the average services per second at device 2. We see that $B = A/2$ (since the only source of work at device 2 is 50% of the services at device 1); similarly $A=(20+ 3B/4)$. Solving these equations gives $A=32$ and $B=16$. Thus every 20 jobs that arrive, give rise to 32 services at device 1, or an average of 1.6 visits per job. Similarly 20 arriving jobs give rise to 16 services at device 2, or 0.8 visits per job.

Q3. Consider a closed system with 20 terminals, each submitting jobs with a think time of 5s between jobs. Each job might be (this happens with 50% probability) of type A, which is a simple calculation that has average response time of 0.1s, or it might be of type B, which is a large data manipulation that has average response time of 3s. How many type A jobs are completed each second? How would the system behave if instead of having 20 terminals each submitting equal numbers of type A and type B jobs, we instead have 10 terminals that always submit type A, and 10 terminals that always submit type B?

*Let throughput for the whole system be X jobs/sec. The average response time over all jobs is $0.5*0.1s+0.5*3s=1.55s$. Thus the response time law says that $1.55 = N/X - Z$, where $N=20$ and $Z=5s$, giving $X=20/6.55$ or just over 3 (the significant figure is probably just 3). Since half of all jobs are type A, we have 1.5 jobs/sec completed of type A.*

If there are 10 terminals for type A, and 10 terminals for type B, then (because the type A jobs go very quickly, and the corresponding terminal gets to submit another job which will be the same type) there will be more type A jobs than type B jobs, in a given period of time. If we assume that response times for each type are unchanged in the different

system, we can apply the response time law separately to each class of terminals, and we get equations $0.1 = 10/XA - 5$, so XA (the throughput for type A jobs) is 1.9 jobs/s, while for type B, $3 = 10/XB - 5$, so the throughput for type B jobs is $XB = 1.25$ jobs/s

Part B: Work through some of the material from ch 5 of Menasce et al (handed out in class). The spreadsheets can be obtained from
<http://www.cs.gmu.edu/~menasce/perfbyd/efiles.html>