Disclaimer

These extra solved exercises come from very old exams (dated between 5 and 15 years ago) when the program of the course was obviously different. They have been superficially checked and seem to be correct and compatible with the current course, but keep in mind that:

- Some important topics (i.e. computing visits from probabilities, exact solution of separable single class open models) are missing
- There could be some question using some formula / theoretical elements that are missing or it wre presented in a different way in the current course
- Some mistake in the solution might be present

Please use this as an additional study material, and if something in the solution do not convince you, you might be right and the problem could be due to one of the previous notes.

1. A coffee company requires its clients to first get a ticket with a progressive number, then numbers are called by one of three free counters, and finally costumers are offered a free coffee. We imagine that the three counters operators are all with identical service speed, and that only one customers out of two accept the free coffee, which is served by another single operator. The ticket machine can provide 12 tickets per minute, the mean service time for a costumer is 2 minutes, and the time required to prepare the free coffee is 30 seconds.

   (a) Draw a visual representation of the system.

   (b) Compute the visits and the demands of all the service centers.

   (c) Identify the bottleneck.

   (d) Which is the maximum number of costumers per minutes that the system is able to serve?

   (e) How many operators at counters and at the free coffee point should be present to make the ticket machine the bottleneck?

   (f) If costumers arrive at a rate of 1 per minute, the mean service time is 3 minutes at the counters. Which is the mean number of costumers in the shop in this case?

   SOLUTIONS:

   (a)

   (b) $D_{ticket} = \frac{1}{X_{ticket}}$; $D_{counter} = \frac{1}{3}120 = 40sec$; $D_{free} = \frac{1}{2}30 = 15sec$

   (c) Bottleneck is counter

   (d) $X_{max} = \frac{1}{D_{max}} = \frac{1}{40} = 0.025rec/sec = 1.5req/min$

   (e) Bisogna far scendere la D di ogni stazione sotto la D_{ticket}, aumentando il numero di stazioni. 6 counter e 4 free.

   (f) $N = XR$ con R somma dei tempi delle singole stazioni. $X_{max} = \frac{1}{60} = 0.01666$; $R = 60 + 60 + 15 = 135$; $N = 2.25$

1. A medium size company wants to evaluate the performances provided to their customers. The intranet of the company includes a web server, an application server and a database, called A, B and C, respectively. The intranet is represented as a closed model. In order to evaluate the network performance, a resources monitoring procedure was set-up and the results are the following:

2. For the system with all servers the following data were detected for an interval time of $T = 2h$. and $N = 10jobs$:
   C number of interactions completed by the system: $3600jobs$
   C_A number of operations completed by server A: $5400jobs$
   B_A busy time of server A: $7200s$
   B_C busy time of server C: $1800s$
   U_B utilization of server B: 0.625
   S_B service time of server B : $0.25s$
   S_C service time of server C : $0.1s$
   Z think time: $4s$

   For the complete system (i.e. with all servers) compute:

   (a) the service demand of each station.

   (b) the utilization of each server

   (c) the visits of each server

   (d) the average network response time.

   (e) Moreover, write the equations of the asymptotes of system throughput and compute for which value of N the system switches from light to heavy load.

   SOLUTIONS:

   (a) $D_k = V_k S_k = \frac{C_k}{C}\frac{B_k}{C_k} = \frac{B_k}{C}$. $D_A = 7200/3600 = 2\text{s}$, $D_C = 1800/3600 = 0.5\text{s}$. $B_B = U_B \cdot T = 0.625 \cdot 7200$. $D_B = 0.625 \cdot 7200/3600 = 1.25\text{s}$.
   (b) $U_k = B_k/T$. $U_A = 7200/7200 = 1$, $U_C = 1800/7200 = 0.25$, $U_B = 0.625$.
   (c) $v_A = 5400/3600 = 1.5, v_B = D_B/S_B = 1.25/0.25 = 5, v_C = 0.5/0.1 = 5$
   (d) $R = N/X - Z$. $X = C/T = 3600/7200 = 0.5\text{j/s}$. $Z = 10/0.5 - 4 = 16\text{s}$.
   (e) $\frac{N}{3.75 \cdot N+4} \leq X \leq \min\left(0.5, \frac{N}{3.75}\right)$. $N* = (D+Z)/D_{Max} = (2+1.25+0.5+4)/2 = 3.875$

2. A secure system is composed of a dispatcher (D), a token server T, and a computation server (P) elaborating critical process, and it is used by $N = 10$ users. The system throughput is $X = 0.366$ job/s. The utilization of the dispatcher is $U_D = 0.7686$, and its service time is $S_D = 0.6$ s. Each jobs is served two times by the token server to acquire and release the token ($V_T = 2$), and each operation takes $S_T = 1$ s. On the average, the critical process needs to be run only for one job every two ($V_P = 0.5$), but the computation server has the same utilization of the token server.

   (a) Compute the visits V_D to the dispatcher and its demand D_D.

   (b) Compute the utilization U_T and the demand D_T of the token server. Determine also its expected busy time on an interval of $T = 30$ min.

   (c) Compute the demand D_P and the mean service time S_P of the computation server.

   (d) Knowing that the system response time is $R = 17.3224$ s, which is the think time of the clients? Which is the average number of clients in the system (that are not thinking)?

   (e) Which are minimum and maximum throughput that the system is expected to show for $N = 20$?

SOLUTIONS:

(a) $X_D = U_D/S_D = 0.768/0.6 = 1.28$ job/s; $V_D = X_D/X = 3.5$; $D_D = V_D S_D = 2.1s$
(b) $D_T = V_T S_T = 2s$; $U_T = X D_T = 0.732$; $B_T = TU_T = 30 * 0.732 = 21.96min = 1317.6s$
(c) $U_P = U_T$; $D_P = U_P/X = 0.732/0.366 = 2s$; $S_P = D_P/V_P = 2/0.5 = 4s$
(d) $R = N/X - Z$; $Z = N/X - R = 10/0.366 - 17.3224 = 10s$
$N_{elab} = XR = 0.366 * 17.3224 = 6.34$ job
(e) $N^* = (D + Z)/D_{max} = (10 + 2.1 + 2 + 2)/2.1 = 16.1/2.1 = 7.666$ job; we are in heavy-load regime.
$X_{min} = 20/(20 * 6.1 + 10) = 0.1515$ job/s;
$X_{max} = 1/D_{max} = 1/2.1 = 0.47619$ job/s

3. In an embedded system, it is known that the bottleneck is the disk, with a demand of $D_{disk} = 45ms$, while the total demand of the system is $D_{tot} = 140ms$. The system is run with an increasing number of jobs $N = 1$, $N = 5$, $N = 20$, and the throughput is measured as: $X(1) = 7.143job/s$, $X(5) = 17.18job/s$ and $X(20) = 21.798job/s$.

   (a) Compute the utilization of the disk for the three considered workloads

   (b) Compute the minimum response time that could be achieved in the three case, considering that it is a batch system (i.e. there are no terminal, and the think time can be considered $Z = 0$).

   (c) Compute the exact response time for the three cases.

   (d) Now consider the same system serving an open workload: which is the maximum arrival rate it can handle?

   SOLUTIONS
   a) $U(1) = 0.045 \cdot 7.143 = 0.32$, $U(5) = 0.045 \cdot 17.18 = 0.77$, $U(20) = 0.045 \cdot 21.798 = 0.98$.
   b) $N^* = \frac{D_{tot}}{D_{max}} = \frac{140}{45} = 3.11$
   $R_{min}(1) = D_{tot} = 0.14s$,
   $R_{min}(5) = 5 \cdot D_{max} = 5 \cdot 45ms = 0.225s$,
   $R_{min}(20) = 20 \cdot 45ms = 0.9s$
   c) $R(1) = 1/7.143 = 0.14s$, $R(5) = 5/17.18 = 0.291s$, $R(20) = 20/21.798 = 0.918s$.
   d) $\lambda_{max} = 1/0.045 = 22.2job/s$.

Question 1

The intranet of a company is accessed by $N = 100$ employees that have a mean think time $Z = 25sec$. The execution of a typical transaction requires 10 accesses to the web server ws, whose service time is $S_{ws} = 30ms$, utilized at 60%.

1. Compute the throughput X and the mean response time R of the intranet
2. How many accesses to the storage server *ss* are generated by a complete execution of a transaction knowing that its throughput is $X_{ss} = 20\, op/sec$?
3. It is known that the mean service time of the *storage server* S_{ss} is 45 ms. Compute its utilization and response time R_{ss}.

SOLUTIONS:

1) $D_{WS} = 10 \cdot 30ms$, $D_{WS} = 0,3sec$; $X = U_{WS}/D_{WS} = 2job/sec$; R = (N/X) - Z; R = (100 / 2)-25 = 25sec

2) $X_{ss} = V_{ss}X$; $20 = V_{ss} * 2$; $V_{ss} = 10op/sec$

3) $U_{ss} = X_{ss} * S_{ss}$; $U_{ss} = 20 * 0,045 = 0,9$ $R_{ss} = N_{ss}/X_{ss} = 100/20 = 5$

Question 1

A system is composed by terminals, a CPU (service time of 0.25 seconds, accessed 10 times per request) and a disk. With N users, the response time is 25 seconds, and the utilization of the CPU is 50%.
Un sistema è composto da terminali, una CPU (tempo di servizio di 0.25 *secondi, accesso in media* 10 *volte per richiesta) ed un disco. Con* N *utenti, il tempo di risposta è di* 25 *secondi, e l'utilizzazione della* CPU *è del* 50%.

1. Knowing that if we double N, the system has a response time of 55 sec, and the utilization of the CPU increases to 62.5%, compute both the think time and the number of users.
Sapendo che se si raddoppia N *il sistema ha un tempo di risposta pari* 55 *secondi, e l'utilizzazione della CPU cresce al* 62.5% *determinare il think time ed il numero di utenti.*

2. Knowing that the visits to the disk per request are 1 less than the visits to the CPU, and the response times are respectively $R_{CPU} = 0,75$ and $R_{disk} = 1$, compute the mean number of users in the CPU, in the disk and in the terminals.
Sapendo che le visite al disco per richiesta sono in media una in meno che quelle alla CPU, e che i tempi di risposta sono rispettivamente $R_{CPU} = 0,75$ *e* $R_{disk} = 1$, *calcolare il numero medio di utenti nella CPU, nel disco e nei terminali.*

SOLUTIONS:

a) $R = \frac{N}{X} - Z$. $X_{cpu}(N) = U_{cpu}(N)/S = 0.5/0.25 = 2$. $X_{cpu}(2N) = 0.625/0.25 = 2.5$. $X(N) = X_{cpu}(N)/V = 2/10 = 0.2$. $X(2N) = 2.5/10 = 0.25$.

$$\begin{cases} 25 = N/0.2 - Z \\ 55 = 2N/0.25 - Z \end{cases}$$

$$\begin{cases} 25 = 5N - Z \\ 55 = 8N - Z \end{cases}$$

$N = 10$, $Z = 25$.

b) $V_{disk} = V_{cpu} - 1 = 9$. $X_{disk} = 9 * X = 9 * 0.2 = 1.8$. $N = XR$. $N_{disk} = 1.8 * 1 = 1.8$. $N_{cpu} = 2 * 0.75 = 1.5$. $N_{term} = N - N_{disk} - N_{cpu} = 10 - 1.8 - 1.5 = 6.7$.

First Name - Nome: ....

Family Name - Cognome: ....

1. An intranet of a company consists of a web server WS, an application server AS and a storage server SS. The system has been measured for an interval of time of T sec. and the following data were detected:

   C number of interactions executed by the system: 3600 job
   C_{WEB} number of operations completed by the web server 5400 op
   C_{APP} number of operations completed by the application server 7200 op
   C_{SS} number of operations completed by the storage server 10800 op
   B_{WEB} busy time of web server 360 s
   B_{APP} busy time of application server 720 s
   B_{SS} busy time of storage server 1080 s
   Z think time 4 s
   The utilization of the local area network is negligible and thus should not be considered. Applying the operational analysis technique compute:

   (a) the service demands of the three servers D_{WEB}, D_{APP}, D_{SS}

   (b) knowing from the measurements that the utilization of the bottleneck is 0.6 , compute the throughput of the system X_0

   (c) compute the utilization of the two servers that are not bottleneck

   (d) write the equations of the asymptotes of system throughput and system response time (assume that it is a closed system)

   (e) if the bottleneck server will be substituted with a new one twice as fast, which will be the new bottleneck of the system? With this new configuration, would it be possible to have a system throughput of 4 j/s or greater?

   (f) in the original system would it be possible to have a response time of 0.7s with 16 users?

   (g) which is the duration of the observation interval T?

SOLUTIONS
Score=7 , one for each question
a) $D_{APP} = B_{APP}/C = 720/3600 = 0.2s, D_{WEB} = 360/3600 = 0.1s, D_{SS} = 1080/3600 = 0.3s$
b) bottleneck = Storage Server $U_{SS} = X_0 D_{SS} = 0.6 \quad X_0 = 0.6/0.3 = 2j/s$
c) $U_{WEB} = 2 * 0.1 = 0.2 \;\; U_{APP} = 2 * 0.2 = 0.4$
d) $N^* = (D + Z)/D_{max} = 15.3333 users$
e) $new \; D_{SS} = 0.15s \quad new \; bottleneck = APP \; Server \; D_{APP} = 0.2s \; , \; X_0^{max} = 1/0.2 = 5j/s$ YES
f) NO, 16 users is greater than N*, thus $R \geq N D_{max} - Z = 16 * 0.3 - 4 = 0.8s$
g) if $X_0 = 2j/s$ then it will be $T = C/X_0 = 3600/2 = 1800s \;\; 30min$

1. A cloud SaaS application has a fixed number $N = 100$ of registered users, having a mean think time $Z = 15sec$. During its execution, the application utilizes a web-server (WS), having a mean service time $S_{WS} = 0.1sec$, a database (DB) with mean service time $S_{DB} = 0.05sec$ and a storage server (SS) with $S_{SS} = 0.01sec$. The complete execution of a transaction requires one access to the WS, 3 accesses to the DB, and 6 accesses to the SS. The utilization of the DB has been measured, its value is $U_{DB} = 0.9$. Compute:

   (a) the throughput of WS, of DB, of SS and of the system

   (b) the utilizations of the WS and of the SS

   (c) the system response time R

   (d) the system response time with $Z = 5$ sec

   SOLUTION: (a) $U_i = X_i S_i \quad X_{DB} = U_{DB}/S_{DB} = 0.9/0.05 = 18r/s \quad X_0 = X_i/V_i = 18/3 = 6trans$, $X_{WS} = X_0 V_{WS} = 6 * 1 = 6req$, $X_{SS} = X_0 V_{SS} = 6 * 6 = 36req$
   (b) $U_{WS} = X_{WS} S_{WS} = 6 * 0.1 = 0.6$, $U_{SS} = X_{SS} S_{SS} = 36 * 0.01 = 0.36$
   (c) $R = (N/X_0) - Z = (100/6) - 15 = 16.66666 - 15 = 1.66666sec$
   (d) impossible to compute, the data are not enough to compute it

1. An enterprise digital infrastructure consists of a web server (WS), an application server (AS), and a storage server (SS).

   The service demands D_i (i.e., the global service time required by a complete execution of a request to each of the components) are: $D_{WS} = 10ms$, $D_{AS} = 20ms$, $D_{SS} = 30ms$. The LAN and other components of the intranet are very *lightly* loaded and are *not* considered in the study.

   (a) In this intranet will be possible to have a throughput of $X = 40tr/sec$? (*show the computations*)

   (b) The management decide to use the storage of a cloud infrastructure. Half of the data stored in the local storage server will be allocated in this new cloud storage. At the end of the migration it will be: $D_{SS,\,int} = D_{SS,\,cloud} = 15ms$. Which will be the maximum throughput of this *new intranet*? Which is the bottleneck resource?

   (c) With a workload consisting of a constant number of users $N = 500users$ with think time $Z = 5sec$ the throughput of the intranet is $X = 25tr/sec$. Compute the response time R of the intranet and the utilization of the servers in the configuration with the cloud storage (*new intranet*).

   (d) Compute the minimum (theoretical) response time R of the new intranet with $N = 600users$ with think time $Z = 5sec$

   SOLUTION:
   a) $NO \;\; X_{max} = 1/D_{max} = 1/0.030 = 33.33tr/sec \; but \; 1/0.040 = 25ms$
   $(asD_{max} \; and \; this \; is \; not \; true- > D_{max} = 30ms$
   b) $X_{max} = 1/D_{max} = 1/0.020 = 50tr/sec$ The new bottleneck is AS (20ms)
   c) $R = (N/X) - Z = (500/25) - 5 = 15sec \quad U_{WS} = XD_{WS} = 25 * 0.010 = 0.25$
   $U_{AS} = XD_{AS} = 25 * 0.020 = 0.5 \quad U_{SS} = XD_{SS} = 25 * 0.015 = 0.375$

   d) $R \geq ND_{max} - Z = (600 * 0.020) - 5 = 7sec$

1. An enterprise digital infrastructure consists of a web server (WS), an application server (AS), and a storage server (SS).
   The service demands D_i (i.e., the global service time required by a complete execution of a request) of the web server and of the application server are: $D_{WS} = 100ms$, $D_{AS} = 150ms$. The LAN and other components of the intranet are very *lightly* loaded and are *not* considered in the study.
   The service time of the Storage Server is $S_{SS} = 2ms$ and its throughput is $X_{SS} = 0.4op/ms$. The *measured* throughput of the network is $X_0 = 4tr/sec$.
   (a) How many operations (visits) a request execute on the Storage Server during a complete execution? Compute the service demand D_{SS} of the Storage Server
   (b) Compute the utilization of the three servers with the measured throughput.
   (c) Which server is the bottleneck of the network, why? Compute the maximum throughput of the network.
   (d) To support the business objective, a throughput of the network of $X_0 = 8tr/sec$ is required. Which are the servers that must be replicated? Compute the number of replicas for each server that must be used.
   (e) With the measured throughput of the network of $4tr/sec$ the number of customers in the network is $N = 6.1666 jobs$. Which is the network response time R? (consider Z=0)

   SOLUTION:
   a) $V_{SS} = X_{SS}/X_0 = 400/4 = 100 visits$ $D_{SS} = 100 * 2 =$ **200ms**
   b) $U_{WS} = X_0 * D_{WS} = 4 * 0.1 =$ **0.4** $U_{AS} = X_0 * D_{AS} = 4 * 0.15 =$ **0.6** $U_{SS} = X_0 * D_{SS} = 4 * 0.2 =$ **0.8**
   c) bottleneck is SS, max X is $X_{max} = 1/D_{max} = 1/0.2 =$ **5tr/sec**
   d) il Dmax necessario per avere $X_0 = 8tr/sec$ e' $D_{max} = 1/X_0 = 1/8 = 0.125sec$ quindi sono necessari **2 AS e 2 SS**
   e) $N = XR$ $R = 6.1666/4 = 1.541s$

1. The intranet of a company is accessed by $N = 100$ employees that have a mean think time $Z = 20sec$. The execution of a typical transaction requires 10 accesses to the web server ws, whose service time is $S_{ws} = 30ms$, utilized at 60%.

   (a) Compute the throughput X and the mean response time R of the intranet

   (b) How many accesses to the storage server ss are generated by a complete execution of a transaction knowing that its throughput is $X_{ss} = 20\, op/sec$?

   (c) It is known that the mean service time of the *storage server* S_{ss} is 45 ms. Compute its utilization U_{ss}.

   (d) Compute the maximum throughput that a system consisting of the web server and the storage server may obtain when the number of customers grows to infinite.

SOLUTIONS:

1) $D_{WS} = 10 \cdot 30ms$, $D_{WS} = 0,3sec$; $X = U_{WS}/D_{WS} = 2job/sec$; R = (N/X) - Z; R = (100 / 2)-20 = 30sec
2) $X_{ss} = V_{ss}X$; $20 = V_{ss} * 2$; $V_{ss} = 10op/sec$
3) $U_{ss} = X_{ss} * S_{ss}$; $U_{ss} = 20 * 0,045 = 0,9$
4) $D_{ss} = 10x0.045 = 0.45s$, $X_{max} = 1/Dmax = 1/0.45 = 2.2222$

Exercises – Performance Evaluation

10 / 5 / 2013

Exercise 1

The storage server of an intranet consists of two groups of disks, A and B, each having exponential distributed service times, with means S_A = 5ms and S_B = 3ms. The mean number of visits for the two components are V_A= 20 and V_B = 30. The throughput of A is 150 op./sec. The above data were collected when the system is processing a workload generated by 300 users with think time Z = 15 sec.

I) Compute the System Throughput X_0 and the Utilization of B. Which one of the two groups is the bottleneck?
II) Compute the system response time.
III) If the number of users increases to 400, which will be the new response time?

Solution

I) $D_A = V_A \cdot S_A = 20 \cdot 0.005 = 0.1s$;
$U_A = X_A \cdot S_A = 150 \cdot 0.005 = 0.75$;

$U_A = X_0 \cdot D_A \rightarrow X_0 = \frac{U_A}{D_A} = \frac{0.75}{0.1} = 7.5\ int/sec$

$U_B = X_0 \cdot D_B = X_0 \cdot (S_B \cdot V_B) = 7.5 \cdot 30 \cdot 0.003 = 0.675$

II) $R = \frac{N}{X_0} - z = \frac{300}{7.5} - 15 = 40 - 15 = 25\ s;$

III) The response time **R when the number of users increases to400cannot be computed** with the operational analysis equations since we cannot use the value of system throughput obtained when the N was 300. In this new case we do not know the new throughput.

Exercise 2

An intranet is composed of 5 web servers used in parallel, 3 application servers used in parallel, and one storage server. The other components on the intranet (e.g., switches, gateways, load balancers, firewalls, network) are not considered since their utilization is very low. The server connected in parallel are used in a balanced way. The complete execution of a transaction requires (service demands) 750 ms to the web server, 600 ms to the application server and 300 ms to the storage server.

Compute the maximum throughput of the intranet.

Solution

$D_{ws} = \frac{750}{5} = 150\ ms\ for\ each\ web\ server;\ D_{as} = \frac{600}{3} = 200\ ms\ for\ each\ application\ server;$

$D_{ss} = 300\ ms\ for\ the\ storage\ server;$

$$X_{max} = \frac{1}{D_{max}} = \frac{1}{300} = 0{,}003\ int/sec$$

Exercise 3

The throughput of a disk is 100 I/O operations per second. To complete a given request 20 visits to the disk are required. The number of users is 100 and the response time is 15 seconds.

Compute the users think time.

Solution

$$R = \frac{N}{X_0} - z \rightarrow Z = \frac{N}{X_0} - R = \frac{100}{X_0} - 15$$

$$X_D = V_D \cdot X_0 \rightarrow X = \frac{X_D}{V_D} = \frac{100}{20} = 5\ \text{int/sec}$$

$$Z = \frac{100}{5} - 15 = 20 - 15 = 5\ \text{sec}$$

Exercise 4

Let's consider an intranet that can be accessed by a large number of users. The execution of a single request pass through an application server (AS), which has a service time S = 300 ms, then through a database server (DS), which has a service time S = 250 ms, and then back through the application server. A request must pass through the system firewall before entering the intranet and before exiting from it. The firewall service time per visit is S = 10 ms.

I) Compute the maximum throughput of the system.

II) It is possible to have a Response Time R < 9 s? At which conditions?

Solution

By drawing the intranet model, the visits $V_{as} = 2;\ V_{ds} = 1;\ V_{fw} = 2;$ *can be obtained.*

Demands can then be computed. $D_{as} = 600;\ D_{ds} = 250;\ D_{fw} = 20;$

I) $X_{max} = \frac{1}{D_{max}} = \frac{1}{600};$

II) If we assume that the intranet is modeled by a **closed model** with a think time $Z = 0$, we have:

$N\,D_{max} - Z \le R$; $N\,D_{max} \le R$; $N\ 600 \le 9000$; $N \le \frac{9000}{600}$; $N \le 15$

If we assume that the intranet is modeled by an **open model,** even if:

$X_0 \le \frac{1}{D_{max}}; X_0 \le \frac{1}{600}$

thus, the system is not saturated, there is no pessimistic bound on the response time. For such reason, we can not provide conditions at which the constraint was satisfied, as done in the **closed model**.

Exercise 5

A web server of a company is connected to an intranet and is accessed by the employees that work internally in the company resulting in a population of fixed size: N=21 users. The average think time of the users is Z=20 sec. A complete execution of a request generate a load of V_s=20 operations to a specific storage device whose utilization is U_s= 0.30. The service time of the storage device per each visit is S_s =0.025 sec.

I) Determine the average system response time R

II) Compute the average throughput and system response time with N=40 users.

Solution

I) $U_s = X_s \cdot S_s \rightarrow X_s = \frac{U_s}{S_s} = \frac{0.3}{0.025} = 12 \text{ int/sec};$

$X_s = V_s \cdot X_0 \rightarrow 12 = 20 \cdot X \rightarrow X = \frac{12}{20} = 0.6 \text{ int/sec};$

$R = \frac{N}{X_0} - z = \frac{21}{0.6} - 20 = 35 - 20 = 15 \text{ sec};$

II) The response time **R when the number of users increases to 40 cannot be computed** with the operational analysis equations since we cannot use the value of system throughput obtained when the N was 21. Indeed, in this case we do not know the new throughput.

Exercise 6

The Intranet of a medium scale company consists of three servers, namely A, B and C, which represent the web server of the clients, the application server and the database server, respectively. The number of users is constant, N=20 users. In order to evaluate the performance of the system a 10 minutes monitoring phase has been performed. The following data have been collected:

Server B number of completions, C_B= 150 op.
Server C number of completions, C_c= 300 op.
Network (intranet) completions, C= 100 op.
Server B busy time, B_B= 300s
Server C busy time, B_C= 100s

It is also known that the maximum throughput achievable by the intranet is 0.2 trans/sec. Compute:

I) the system throughput during the measurement phase

II) the service demands of all the servers and determine the server which should be upgraded to achieve the maximum gain of the network performance

III) the utilizations of all the servers

IV) the number of visits at server B

V) the system response time if the users have a mean think time of 30 sec.

Solution

I) $X_0 = \frac{C}{T} = \frac{100}{600} = \frac{1}{6} = 0.1666 \text{ int/sec}$

II) $D_B = \frac{B_B}{C} = \frac{300}{100} = 3\ sec; D_C = \frac{B_C}{C} = \frac{100}{100} = 1\ sec;$

Given that $X_{max} < \min\left\{\frac{1}{D_B}, \frac{1}{D_C}\right\}$ then $D_{max} = D_A = \frac{1}{X_{max}} = \frac{1}{0.2} = 5$ sec, A is the bottleneck.

III) $D_k = \frac{U_k T}{C} \rightarrow U_A = \frac{D_A C}{T} = \frac{5 \cdot 100}{600} = 0.833; U_B = \frac{D_B C}{T} = \frac{3 \cdot 100}{600} = \frac{3}{6} = {}^1\!/_2 = 0.5;$

$U_C = \frac{D_C C}{T} = \frac{1 \cdot 100}{600} = \frac{1}{6} = 0.1667$

IV) $V_B = \frac{C_B}{C} = \frac{150}{100} = 1.5 \text{ visits}$

V) $R = \frac{N}{X_0} - z = \frac{20}{\frac{1}{6}} - 30 = 120 - 30 = 90 \text{ sec}$

Exercise 7 - Performance of an IT infrastructure

Let's consider an IT infrastructure consisting of a Web Server (WS), an Application Server (AS) and a Storage Server (SS).After 1 hour measurement, during which N = 50 users were working continuously, the following data have been collected:

C_0 total number of jobs executed by the system: 5400 j

C_{WS}Number of WS completed operations: 54000 op

C_{AS}Number of AS completed operations: 32400 op

C_{SS} Number of SS completed operations: 10800 op

B_{WS} WS total *activity* time: 1800 sec

B_{AS} AS total *activity* time: 720 sec

B_{SS} SS total *activity* time: 900 sec

Z Mean think time 5 sec

Using Operational Analysis equations:

1. Compute the visits V_i to the three servers during a complete job execution, their global service requests D_i and determine the bottleneck resource of the IT infrastructure

2. Compute response time when N = 50 users are connected, as well as the maximum throughput when the number of users tends to infinity (asymptotic value)

3. Let's substitute the bottleneck resource determined at point 1) with another, two times (2x) more powerful. Does the bottleneck migrate to another resource? If so, which one? Compute the new value of the asymptotic throughput.

Solution:

1)$V_{WS} = C_{WS}/ C_0 = 54000 / 5400 = 10$; $V_{AS} = C_{AS}/C = 32400/5400 = 6$; $V_{SS} = 2$;
$U_{WS} = B_{WS}/T = 1800/3600 = 1/2$; $U_{AS} = B_{AS}/T = 1/5$; $U_{SS} = 1/4$;
$X = C/T = 5400/3600 = 3/2$ j/s;
$D_{WS} = U_{WS}/X_0 = (1/2)/(3/2) = 1/3$s; $D_{AS} = U_{AS}/X_0 = 2/15$s; $D_{SS} = U_{SS}/X_0 = 1/6$sec;
bottleneck = WebServerD_{max} = (1/3)s;

2)$R = (N/X) - Z = (50/3/2) - 5 = 85/3$s; $X_{max} = 1/D_{max} = 3$ j/s;

3)new D_{WS} = 1/6 sec hence Web Server and Storage Server are now both bottlenecks, new volume asymptote of throughput $X_{max} = 1/D_{max}$ = 6 j/s;