UNIVERSITY OF DUBLIN

TRINITY COLLEGE

FACULTY OF ENGINEERING & SYSTEMS SCIENCES

DEPARTMENT OF COMPUTER SCIENCE

B.A. (Mod). in Computer Science
Senior Sophister Examination

Trinity Term 2001

4BA2 - Systems Modelling

Tuesday 22nd May *MANSION HOUSE* *14.00 - 17.00*

Professor Francis Neelamkavil, Dr. T. Redmond, Dr. Donal O'Mahony

Answer FIVE questions, at least one from each section
Please use separate answer books for each section
Queueing Tables are attached to paper.

SECTION A

1. Compare and contrast system simulation by using CA (Cellular Automata), GA (Genetic Algorithms) and Mathematical/Logical modelling.

   Comment on the major conclusions drawn from your CA simulation study of the *shark-fish* problem (class assignment).

2. Discuss any ***two*** of the following

   (a) Verification and validation of models.

   (b) Generation of random variates and their applications in simulation.

   (c) Systems concept.

   (d) Next-event time-advance approach to discrete system simulation.

SECTION B

3. Describe the fundamental label-swapping technique at the core of AsynchronousTransfer Mode (ATM) and show with the aid of a diagram how an ATM cell is passed through a sequence of ATM switches. How is the necessary state established to allow this to occur? What kind of protocol stack is used to allow both signalling and user information to be passed between an ATM terminal and the network.

4. Give an account of the purpose and realization of the X.500 directory system commenting on the data model and the way in which the system is queried. What future lies ahead for directory systems?

SECTION C

5. **a.** Suggest how you would approach the problem of using a queueing theory approach via a series of models of increasing complexity, for estimating the performance of a computer system. Give examples.

b. Suggest three rules of thumb useful for using queueing theory in computer system design.

c. Discuss what is meant by the incremental improvement of performance by the successive removal of bottlenecks, and give an example.

d. Specify Little's Relation giving the meaning of each term. What is its importance and where can it be used?

6. **a.** Discuss briefly what are the main characteristics of the mainframe. Why have not mainframes been replaced by PCs?

b. Discuss briefly the 3-layer hybrid architecture which it is claimed is replacing the traditional mainframe giving the advantages of each layer.

c. Comment briefly on why there has been a dramatic increase in mainframe sales in the last five years

7. The characteristics of an interactive computer system, consisting essentially of a CPU, Logical Drum, Logical Disk and Terminals, have been estimated as follows:

Mean CPU service time per interaction = 7 ms
Mean Drum service time per interaction = 10 ms
Mean Disk service time per interaction = 70 ms.

After an interaction with the CPU, the probability of a job finishing is 0.1, the probability of the job going to the Drum is 0.7, and the probability of going to the Disk is 0.2.

Continued/.........................

The Central Server model formulae (using the usual notation) are as follows:

$$\rho_i = \left[\frac{G(K-1)}{G(K)}\right] \qquad i = 1$$

$$= \frac{\mu_1 \rho_1 p_i}{u_i} \qquad i = 2, 3, \ldots\ldots\ldots M.$$

$$\lambda_t = \mu_1 p_1 \rho_1$$

Buzen's algorithm:

$$x_1 = 1$$

$$x_i = \frac{\mu_1 p_i}{\mu_i} \qquad i = 2, 3, \ldots\ldots\ M.$$

$$G(K) = g(K, M)$$

$$g(k, m) = g(k, m-1) + x_m\, g(k-1, m) \qquad \text{if } k > 0 \text{ and } m > 1$$

$$g(k, 1) = 1 \qquad \text{for } k = 0, 1, \ldots, K$$

$$g(0, m) = 1 \qquad \text{for } m = 1, 2, \ldots, M$$

a. Use the Central Server model to model the CPU/ IO inner subsystem of the computer system and use Little's relation for other quantities. Draw a sketch of your model. Estimate the resource utilisations at a multiprogramming level of 4 for all resources. Calculate the throughput, and W, the response time, assuming there are (N =) 40 terminals in use with Think Time = 3 seconds. State which is the bottleneck resource.

b. Replace the original bottleneck processor with three of the same speed (assume the bottleneck traffic is evenly spread over the 3 processors). Compute resource utilisation, throughput and response time. State which is the bottleneck resource and give your comments.

c. Replace the bottleneck processor with one which is twice as fast. Again compute the resource utilisations, throughput and response time. State the bottleneck resource and comment on your results.

d. Write a short note on extensions to the Central Server model.

1.

CS4BA21

Appendix C

QUEUEING THEORY DEFINITIONS AND FORMULAS

In Figs. 5.1.1 and 5.1.2, reproduced from Chapter 5, we indicate the elements and random variables used in queueing theory models. Table 1 is a compendium of the queueing theory definitions and notation used in this book. The remainder of Appendix C consists of tables of queueing theory formulas for the most useful models and figures to help with the calculations. APL functions are displayed in Appendix B to implement the formulas for most of the queueing models.

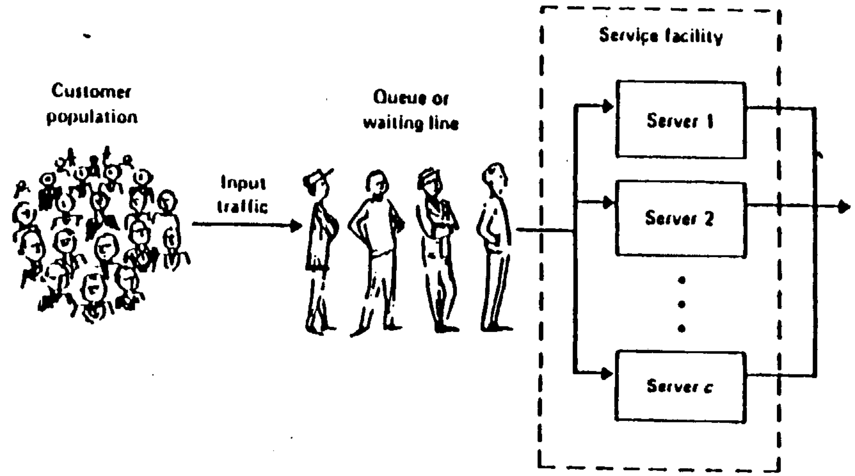

Fig. 5.1.1 Elements of a queueing system.

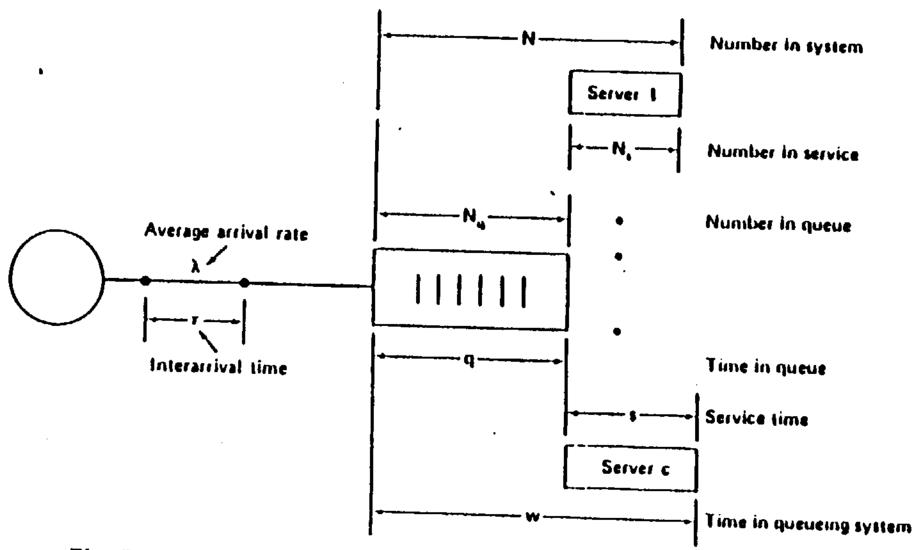

Fig. 5.1.2 Some random variables used in queueing theory models.

2. CS4BA21

TABLE 1
Queueing Theory Notation and Definitions

$A(t)$	Distribution of interarrival time. $A(t) = P[\tau \leq t]$.
$B(c, u)$	Erlang's B formula or the probability all c servers are busy in an $M/M/c/c$ queueing system.
$C(c, u)$	Erlang's C formula or the probability all c servers are busy in an $M/M/c$ queueing system.
c	Symbol for the number of servers in the service facility of a queueing system.
D	Symbol for constant (deterministic) interarrival or service time distribution.
$E[N]$	Expected (average or mean) number of customers in the steady state queueing system. The letter L is also used for $E[N]$.
$E[N_q]$	Expected (average or mean) number of customers in the queue (waiting line) when the system is in the steady state. The symbol L_q is also used for $E[N_q]$.
$E[N_s]$	Expected (average or mean) number of customers receiving service when the system is in the steady state.
$E[q]$	Expected (average or mean) queueing time (does not include service time) when the system is in the steady state. The symbol W_q is also used for $E[q]$.
$E[s]$	Expected (average or mean) service time for one customer. The symbol W_s is also used for $E[s]$.
$E[\tau]$	Expected (average or mean) interarrival time. $E[\tau] = 1/\lambda$, where λ is average arrival rate.
$E[w]$	Expected (average or mean) waiting time in the system (this includes both queueing time and service time) when the system is in the steady state. The letter W is also used for $E[w]$.
E_k	Symbol for Erlang-k distribution of interarrival or service time.
$E[N_q \mid N_q > 0]$	Expected (average or mean) queue length of nonempty queues when the system is in the steady state.
$E[q \mid q > 0]$	Expected (average or mean) waiting time in queue for customers delayed when the system is in the steady state. Same as $W_{q\mid q>0}$.
FCFS	Symbol for "first come, first served." queue discipline.
FIFO	Symbol for "first in, first out." queue discipline which is identical with FCFS.
G	Symbol for general probability distribution of service time. Independence usually assumed.
GI	Symbol for general independent interarrival time distribution.
K	Maximum number allowed in queueing system, including both those waiting for service and those receiving service. Also size of population in finite population models.
L	$E[N]$, expected (average or mean) number in the queueing system when the system is in the steady state.
$\ln(\cdot)$	The natural logarithm function or the logarithm to the base e.
L_q	$E[N_q]$, expected (average or mean) number in the queue, not including those in service, for steady state system.
LCFS	Symbol for "last come, first served," queue discipline.
LIFO	Symbol for "last in, first out." queue discipline which is identical to LCFS.
λ	Average (mean) arrival rate to queueing system. $\lambda = 1/E[\tau]$, where $E[\tau]$ = average interarrival time.

3. CS4BA21

TABLE 1 *(Continued)*

λ_T	Average throughput of a computer system measured in jobs or interactions per unit time.
M	Symbol for exponential interarrival or service time distribution.
μ	Average (mean) service rate per server. Average service rate $\mu = 1/E[s]$, where $E[s]$ is the average (mean) service time.
N	Random variable describing number in queueing system when system is in the steady state.
N_q	Random variable describing number of customers in the steady state queue.
N_s	Random variable describing number of customers receiving service when the system is in the steady state.
O	Operating time of a machine in the machine repair queueing model (Sections 5.2.6 and 5.2.7). O is the time a machine remains in operation after repair before repair again is necessary.
$p_n(t)$	Probability that there are n customers in the queueing system at time t.
p_n	Steady state probability that there are n customers in the queueing system.
PRI	Symbol for priority queueing discipline.
PS	Abbreviation for "processor-sharing queue discipline." See Section 6.2.1.
$\pi_q(r)$	Symbol for rth percentile queueing time; that is, the queueing time that r percent of the customers do not exceed.
$\pi_w(r)$	Symbol for rth percentile waiting time in the system; that is, the time in the system (queueing time plus service time) that r percent of the customers do not exceed.
q	Random variable describing the time a customer spends in the queue (waiting line) before receiving service.
RSS	Symbol for queue discipline with "random selection for service."
ρ	Server utilization = traffic intensity/$c = \lambda E[s]/c = (\lambda/\mu)/c$. The probability that any particular server is busy.
s	Random variable describing service time for one customer.
SIRO	Symbol for queue discipline, "service in random order" which is identical with RSS. It means that each waiting customer has the same probability of being served next.
τ	Random variable describing interarrival time.
u	Traffic intensity = $E[s]/E[\tau] = \lambda E[s] = \lambda/\mu$. Unit of measure is the erlang.
w	Random variable describing the total time a customer spends in the queueing system, including both service time and time spent queueing for service.
$W(t)$	Distribution function for w. $W(t) = P[w \leq t]$.
W	$E[w]$, expected (average or mean) time in the steady state system.
$W_q(t)$	Distribution function for time in the queue. $W_q(t) = P[q \leq t]$.
W_q	$E[q]$, expected (average or mean) time in the queue (waiting line), excluding service time, for steady state system.
$W_{q\vert q>0}$	Expected (average or mean) queueing time for those who must queue. Same as $E[q \vert q > 0]$.
$W_s(t)$	Distribution function for service time. $W_s(t) = P[s \leq t]$.
W_s	$E[s]$, expected (average or mean) service time, $1/\mu$.

4. CS4BA21

TABLE 2

Relationships Between Random Variables of Queueing Theory Models

$u = E[s]/E[\tau] = \lambda E[s] = \lambda/\mu$	Traffic intensity in erlangs.
$\rho = u/c = \lambda E[s]/c = \lambda/c\mu$	Server utilization. The probability any particular server is busy.
$w = q + s$	Total waiting time in the system, including waiting in queue and service time.
$W = E[w] = E[q] + E[s] = W_q + W_s$	Average total waiting time in the steady state system.
$N = N_q + N_s$	Number of customers in the steady state system.
$L = E[N] = E[N_q] + E[N_s] = \lambda E[w] = \lambda W$	Average number of customers in the steady state system. $L = \lambda W$ is known as "Little's formula."
$L_q = E[N_q] = \lambda E[q] = \lambda W_q$	Average number in the queue for service for steady state system. $L_q = \lambda W_q$ is also called "Little's formula."

5.

CS4BA21

TABLE 3

Steady State Formulas for M/M/1 Queueing System

$$p_n = P[N = n] = (1-\rho)\rho^n, \qquad n = 0, 1, 2, \ldots$$

$$P[N \geq n] = \sum_{k=n}^{\infty} p_k = \rho^n, \qquad n = 0, 1, 2, \ldots$$

$$L = E[N] = \rho/(1-\rho), \qquad \sigma_N{}^2 = \rho/(1-\rho)^2.$$

$$L_q = E[N_q] = \rho^2/(1-\rho), \qquad \sigma^2_{N_q} = \rho^2(1+\rho-\rho^2)/(1-\rho)^2.$$

$$E[N_q \mid N_q > 0] = 1/(1-\rho), \qquad \mathrm{Var}[N_q \mid N_q > 0] = \rho/(1-\rho)^2.$$

$$W(t) = P[w \leq t] = 1 - e^{-t/W}, \qquad P[w > t] = e^{-t/W}.$$

$$W = E[w] = E[s]/(1-\rho), \qquad \sigma_w = W.$$

$$\pi_w(90) = W \ln 10 \approx 2.3W, \qquad \pi_w(95) = W \ln 20 \approx 3W.$$

$$\pi_w(r) = W \ln [100/(100-r)].$$

$$W_q(t) = P[q \leq t] = 1 - \rho e^{-t/W}, \qquad P[q > t] = \rho e^{-t/W}.$$

$$W_q = E[q] = \rho E[s]/(1-\rho).$$

$$\sigma_q{}^2 = (2-\rho)\rho E[s]^2/(1-\rho)^2.$$

$$E[q \mid q > 0] = W, \qquad \mathrm{Var}[q \mid q > 0] = W^2.$$

$$\pi_q(90) = W \ln(10\rho), \qquad \pi_q(95) = W \ln(20\rho).$$

$$\pi_q(r) = W \ln\left(\frac{100\rho}{100-r}\right).$$

All percentile formulas for q will yield negative values when ρ is small: all negative values should be replaced by zero. For example, if ρ is 0.02, then 98 percent of all customers do not have to queue for service so the 95th percentile value of q is zero: so are the 90th and 95th percentile values.

6. CS 4 BA21

TABLE 4
Steady State Formulas for M/M/1/K Queueing System

$(K \geq 1$ and $N \leq K)$

$$p_n = P[N = n] = \begin{cases} \dfrac{(1-u)u^n}{1-u^{K+1}} & \text{if } \lambda \neq \mu \text{ and } n = 0, 1, \ldots, K \\ \dfrac{1}{K+1} & \text{if } \lambda = \mu \text{ and } n = 0, 1, \ldots, K. \end{cases}$$

$p_K = P[N = K]$. Probability an arriving customer is lost.

$\lambda_a = (1 - p_K)\lambda$. λ_a is the actual arrival rate at which customers enter the system.

$$L = E[N] = \begin{cases} \dfrac{u[1-(K+1)u^K + Ku^{K+1}]}{(1-u)(1-u^{K+1})} & \text{if } \lambda \neq \mu \\ \dfrac{K}{2} & \text{if } \lambda = \mu. \end{cases}$$

$$L_q = E[N_q] = L - (1 - p_0)$$

$$q_n = \frac{p_n}{1 - p_K}, \quad n = 0, 1, 2, \ldots, K-1.$$

q_n is the probability that there are n customers in the system just before a customer enters.

$$W(t) = P[w \leq t] = 1 - \sum_{n=0}^{K-1} q_n \sum_{k=0}^{n} e^{-\mu t} \frac{(\mu t)^k}{k!}.$$

$$W = E[w] = L/\lambda_a.$$

$$W_q(t) = P[q \leq t] = 1 - \sum_{n=0}^{K-2} q_{n+1} \sum_{k=0}^{n} e^{-\mu t} \frac{(\mu t)^k}{k!}.$$

$$W_q = E[q] = L_q/\lambda_a.$$

$$E[q \mid q > 0] = W_q/(1 - p_0).$$

$$\rho = (1 - p_K)u.$$

ρ is the true server utilization (fraction of time the server is busy).

TABLE 5
Steady State Formulas for M/M/c Queueing System

$u = \lambda/\mu = \lambda E[s]$, $\quad \rho = u/c$.

$$p_0 = P[N = 0] = \left[\sum_{n=0}^{c-1} \frac{u^n}{n!} + \frac{u^c}{c!\,(1-\rho)}\right]^{-1} = c!\,(1-\rho)C(c,u)/u^c.$$

$$p_n = \begin{cases} \dfrac{u^n}{n!}\,p_0 & \text{if } n = 0, 1, \ldots, c \\[2ex] \dfrac{u^n p_0}{c!\,c^{n-c}} & \text{if } n \ge c. \end{cases}$$

$$L_q = E[N_q] = \lambda W_q = \frac{uC(c,u)}{c(1-\rho)}, \qquad \sigma^2_{N_q} = \frac{\rho C(c,u)[1+\rho-\rho C(c,u)]}{(1-\rho)^2},$$

where $C(c,u) = P[N \ge c]$ = probability all c servers are busy is called Erlang's C formula.

$$C(c,u) = \frac{u^c}{c!}\Bigg/\left[\frac{u^c}{c!} + (1-\rho)\sum_{n=0}^{c-1}\frac{u^n}{n!}\right].$$

$$L = E[N] = L_q + u = \lambda W.$$

$$W_q(0) = P[q = 0] = 1 - \frac{p_c}{1-\rho} = 1 - C(c,u).$$

$$W_q(t) = P[q \le t] = 1 - \frac{p_c}{1-\rho}\,e^{-c(1-\rho)t/E[s]} = 1 - C(c,u)e^{-c(1-\rho)t/E[s]}.$$

$$W_q = E[q] = \frac{C(c,u)E[s]}{c(1-\rho)}, \qquad E[q \mid q > 0] = \frac{E[s]}{c(1-\rho)}.$$

$$\sigma_q^2 = \frac{[2 - C(c,u)]C(c,u)E[s]^2}{c^2(1-\rho)^2}. \qquad \pi_q(r) = \frac{E[s]}{c(1-\rho)}\ln\left(\frac{100C(c,u)}{100-r}\right).^{*}$$

$$\pi_q(90) = \frac{E[s]}{c(1-\rho)}\ln(10C(c,u)), \qquad \pi_q(95) = \frac{E[s]}{c(1-\rho)}\ln(20C(c,u)).^{*}$$

$$W(t) = P[w \le t] = \begin{cases} 1 + C_1 e^{-\mu t} + C_2 e^{-c\mu(1-\rho)t} & \text{if } u \ne c-1 \\ 1 - [1 + C(c,u)\mu t]e^{-\mu t} & \text{if } u = c-1. \end{cases}$$

where $C_1 = \dfrac{u - c + W_q(0)}{c - 1 - u}$ and $C_2 = \dfrac{C(c,u)}{c - 1 - u}$.

$$W = E[q] + E[s].$$

$$E[w^2] = \begin{cases} \dfrac{2C(c,u)E[s]^2}{u + 1 - c}\left\{\dfrac{1 - c^2(1-\rho)^2}{c^2(1-\rho)^2}\right\} + 2E[s]^2, & u \ne c-1. \\[2ex] 4C(c,u)E[s]^2 + 2E[s]^2, & u = c-1. \end{cases}$$

$$\sigma_w^2 = E[w^2] - E[w]^2$$

$$\left.\begin{aligned} \pi_w(90) &\approx W + 1.3\sigma_w \\ \pi_w(95) &\approx W + 2\sigma_w \end{aligned}\right\} \text{Martin's estimates}$$

* All percentile formulas for q yield negative values for low server utilization: all should be replaced by zero.

TABLE 8

Steady State Formulas for M/M/2 Queueing System

$\rho = \lambda E[s]/2 = u/2.$

$p_0 = P[N = 0] = (1 - \rho)/(1 + \rho).$

$p_n = P[N = n] = 2p_0\rho^n = \dfrac{2(1-\rho)\rho^n}{(1+\rho)}, \qquad n = 1, 2, 3, \ldots.$

$\left.\begin{array}{l}\pi_w(90) \approx W + 1.3\sigma_w \\ \pi_w(95) \approx W + 2\sigma_w\end{array}\right\}$ Martin's estimate.

$L_q = E[N_q] = \dfrac{2\rho^3}{1-\rho^2}, \qquad \sigma^2_{N_q} = \dfrac{2\rho^3[(\rho+1)^2 - 2\rho^3]}{(1-\rho^2)^2}$

$C(2, u) = P[\text{both servers busy}] = 2\rho^2/(1+\rho).$

$L = E[N] = L_q + u = 2\rho/(1-\rho^2).$

$W_q(0) = P[q = 0] = (1 + \rho - 2\rho^2)/(1+\rho).$

$W_q(t) = P[q \le t] = 1 - [(2\rho^2)/(1+\rho)]e^{-2\mu t(1-\rho)}.$

$W_q = E[q] = \rho^2 E[s]/(1-\rho^2). \qquad E[q \mid q > 0] = E[s]/2(1-\rho).$

$\sigma_q^2 = \rho^2(1 + \rho - \rho^2)E[s]^2/(1-\rho^2)^2.$

$\pi_q(r) = \dfrac{E[s]}{2(1-\rho)} \ln\left(\dfrac{200\rho^2}{(100-r)(1+\rho)}\right).$

$\pi_q(90) = \dfrac{E[s]}{2(1-\rho)} \ln\left(\dfrac{20\rho^2}{1+\rho}\right). \qquad \pi_q(95) = \dfrac{E[s]}{2(1-\rho)} \ln\left(\dfrac{40\rho^2}{1+\rho}\right).$[a]

$$W(t) = P[w \le t] = \begin{cases} 1 - \dfrac{(1-\rho)}{1-\rho-2\rho^2}e^{-\mu t} + \dfrac{2\rho^2}{1-\rho-2\rho^2}e^{-2\mu t(1-\rho)} & \text{if } u \ne 1 \\ 1 - \left[1 + \dfrac{\mu t}{3}\right]e^{-\mu t} & \text{if } u = 1. \end{cases}$$

$W = E[s]/(1-\rho^2).$

$$E[w^2] = \begin{cases} \dfrac{\rho^2 E[s]^2[1 - 4(1-\rho)^2]}{(2\rho-1)(1-\rho)(1-\rho^2)} + 2E[s]^2. & u \ne 1. \\ \frac{19}{9}E[s]^2. & u = 1. \end{cases}$$

$\sigma_w^2 = E[w^2] - E[w]^2.$

[a] All percentile formulas for q yield negative values for low server utilization: all such should be replaced by zero.

CS4BA21

TABLE 24

Steady State Formulas for the Finite Population Queueing Model of Interactive Computing With Processor-Sharing

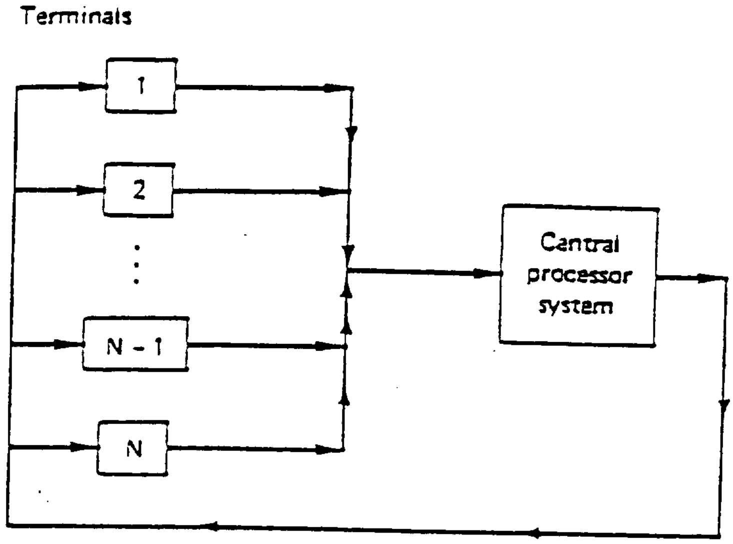

Fig. 6.3.1 Finite population queueing model of interactive computer system. Special case in which the central processor system consists of a single CPU with processor-sharing queue discipline.

The CPU operates with the processor-sharing queue discipline. CPU service time is general with the restriction that the Laplace-Stieltjes transform must be rational. The same restriction holds on think time. $E[t] = 1/\alpha$ is the average think time with $E[s] = 1/\mu$ the average CPU service time. Then

$$p_0 = \left[\sum_{n=0}^{N} \frac{N!}{(N-n)!}\left(\frac{E[s]}{E[t]}\right)^n\right]^{-1} = \left[\sum_{n=0}^{N} \frac{N!}{(N-n)!}\left(\frac{\alpha}{\mu}\right)^n\right]^{-1}.$$

The CPU utilization

$$\rho = 1 - p_0,$$

and the average throughput

$$\lambda_T = \frac{\rho}{E[s]} = \frac{1-p_0}{E[s]}.$$

The average response time

$$W = \frac{NE[s]}{1-p_0} - E[t].$$

CS4BA21.

10.

TABLE 27

Steady State Equations of Central Server Model of Multiprogramming

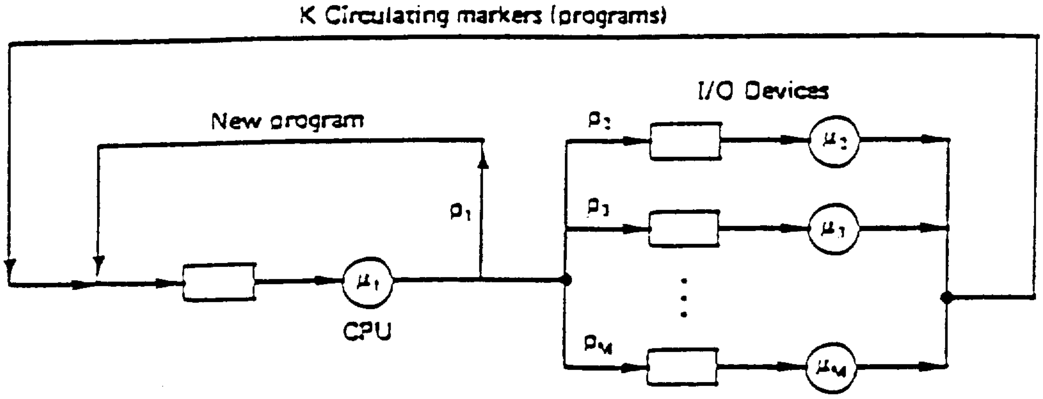

Fig. 6.3.4 Central server model of multiprogramming.

For the assumptions of the model see Section 6.3.4.
Calculate $G(0), G(1), \ldots, G(K)$ by Algorithm 6.3.1 (Buzen's Algorithm).
Then the server utilizations are given by

$$\rho_i = \begin{cases} G(K-1)/G(K) & i = 1 \\ \dfrac{\mu_1 \rho_1 p_i}{\mu_i} & i = 2, 3, \ldots, M. \end{cases} \tag{6.3.25}$$

The average throughput λ_T is given by

$$\lambda_T = \mu_1 \rho_1 p_1. \tag{6.3.26}$$

If the central server model is the central processor model for the interactive computing system of Fig. 6.3.1, then the average response time W is calculated by

$$W = \frac{N}{\lambda_T} - E[t] = \frac{N}{\mu_1 \rho_1 p_1} - E[t]. \tag{6.3.27}$$