UNIVERSITY OF DUBLIN

TRINITY COLLEGE

FACULTY OF ENGINEERING &SYSTEMS SCIENCE

DEPARTMENT OF COMPUTER SCIENCES

B.A.(Mod.) Computer Science SS Examination

Trinity Term 1999

4BA2 - Systems Modelling

Wednesday 26th May

Goldsmith Hall

14.00 - 17.00

Professor F. Neelamkavil, Dr. T. Redmond, Mr. C. McGoldrick

Attempt FIVE questions. At least one from each section Please use separate answer books for each section (Queuing Tables are available from the invigilators)

SECTION A

- 1. Define the terms *Verification* and *Validation* in Modelling and Simulation. Explain the construction of a simple population dynamics model with emphasis on the verification and validation of the model.
- 2. The ISS (Information Systems Services) of a small university provide Internet access, word processing, scanning and printing facilities at one of its laboratories (Cyber-Lab) reserved for undergraduate students. The director of the ISS who has a limited budget is very keen to carry out a simulation study of the Cyber-Lab with a view to identifying the best possible physical layout and equipment configuration strategy that is likely to maximise student satisfaction and minimise the equipment budget. Assuming that you are in charge of this project, discuss the important features of the modelling and simulation study of the Cyber-Lab. Comment on the input/output data and the likely format of your final report to the Director of the ISS. You are free to make reasonable assumptions wherever necessary.

SECTION B

- 3. You have been retained by an Irish University to provide advice on cryptography and encryption systems. In the future, the University hopes to derive significant income from the provision of "distance-learning" undergraduate and postgraduate courses. Distance (or remote) learning courses involve the students studying from home and only attending college for tutorials or exams at weekends. In the first instance, the University wants to implement an infrastructure that allows secure and non-repudiable electronic submission of coursework by students taking these distance learning courses. The system should be as flexible as possible and facilitate secure delivery using web technologies, E-mail and FTP. Later, it is hoped to extend the implementation to allow examinations to be taken on-line. The University have been offered commercial products that use single DES or PGP or X.509 Digital Certificates but are unsure as to the differences between them. Taking account of the above requirements, advise your client as to:
- a) The issues they can expect to encounter in implementing their cryptographic infrastructure;
- b) The advisability of adopting any of the commercial offerings;
- c) Your recommendations for their infrastructure and future direction. Where appropriate, you may illustrate your answer with diagrams.
- 4. Hosts on the Internet are currently identified using a 32 bit address. Discuss the structure of this address and explain how and why subnet masking has evolved as it has. Describe briefly the procedure involved in the delivery of a datagram to the correct destination host. Comment on the limitations of the IPv4 addressing scheme and explain how the IPv6 addressing scheme aims to remedy these limitations.

Section C

- 5. a. Specify Little's Relation giving the meaning of each term. What is its importance and where can it be used?
 - b. Suggest how you would approach the problem of using a queueing theory model, or a series of models of increasing complexity, for estimating the performance of a computer system. Give examples.
 - c. Suggest three rules of thumb useful for using queueing theory in computer system design.
 - d. Discuss what is meant by the incremental improvement of performance by the successive removal of bottlenecks, and give an example.

The formula for the normalised response time of an interactive computer system using the machine repairman model is as follows with the usual notation:

$$\mu W = \frac{N}{\left(1 - p_o\right)} + \frac{\mu}{\alpha}$$

(W, the average response time, p0 (the probability the CPU-I/O system is idle), ρ the CPU-I/O system utilisation and λ the system throughput) and

$$p_o = \frac{1}{\sum_{n=0}^{N} \frac{N!}{(N-n)!} \left(\frac{\alpha}{\mu}\right)^n}$$

p0 may be calculated using the above equation where there are N active terminals in the system.

- a. Give in a diagram the plot of time in system versus number of active users. Sketch how the diagram will change for a CPU of twice the speed. Sketch how the diagram will change for a "Think Time" twice as fast.
- **b.** What is the significance of a "Normalised" versus a non-Normalised response and why is a Normalised Response preferred for theoretical purposes?
- c. What are the practical difficulties in the use of above formula.?
- d. Give the formula for the Kleinrock saturation point i.e. the number of active terminals n* at which saturation begins. How useful is this model for computing the number of terminals n* at which saturation begins? Why is this not very useful?
- e. Would this be a useful model ogf the College Internet Server why or why not? What is a better model and why?
- 7. The characteristics of an interactive computer system, consisting essentially of a CPU, Drum, 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.

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

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

$$= \frac{\mu_{1}\rho_{1}p_{i}}{\mu_{i}} \qquad i = 2,3, \dots M$$

$$\lambda_{i} = \mu_{1}\rho_{1}p_{1}$$

Buzen's algorithm:

$$x_1 = 1$$

 $x_i = \frac{\mu_1 p_i}{\mu_i}$ $i = 2,3, \dots M$
 $G(K) = g(K, M)$
 $g(k, m) = g(k, m - 1) + x_m g(k - 1, m)$ if $k > 0$ and $m > 1$
 $g(k, 1) = 1$ for $k = 0,1,2,\dots K$
 $g(0, m) = 1$ for $m = 1,2,\dots M$

- 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 original 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.

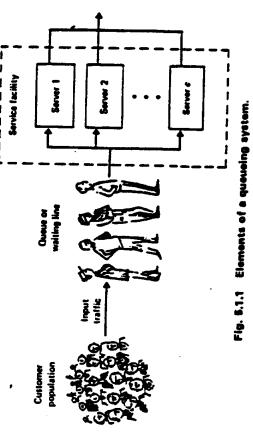
©University of Dublin 1999

しこう

Appendix C

QUEUEING THEORY DEFINITIONS AND FORMULAS

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



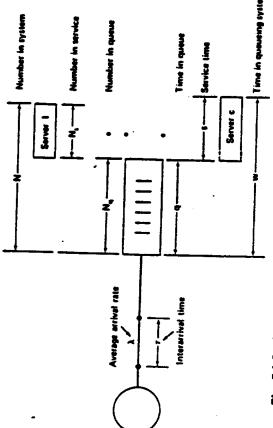


Fig. 5.1.2 Some random variables used in queueing theory models.

TABLE 1 Queueing Theory Notation and Definitions

A(t)	Distribution of interarrival time, $A(t) = P[\tau \le t]$.
B(c, u)	Erlang's B formula or the probability of $t = r[t \le t]$.
, ,	Erlang's B formula or the probability all c servers are busy in an M/M/c/c queueing system.
C(c, u)	Eriang's Commission the make the
	Erlang's C formula or the probability all c servers are busy in an M/M/c queueing system.
E	Janares 1/4/2000
D	Symbol for the number of servers in the service facility of a queueing system.
E[N]	The state of the s
-6-1	amposited (average of mean) number of distorners in the second many
$E[N_a]$	quedeling system. The letter L is also used for FTAN
_(. •1	Expected (average or mean) number of customers in the queue (waiting line)
E[N]	when the system is in the steady state. The symbol / is also used to prive t
-(· · · · · ·	
E[q]	system is in the steady state.
-[4]	Expected (average or mean) queueing time (does not include service time)
E[s]	" " " " " " " " " " " " " " " " " " "
حاد	The combat tree in the state of
Ff_1	and and tot 1/21.
E[au]	Expected (average or mean) interarrival time. $E[\tau] = 1/\lambda$, where λ is average
Ef1	— · · · · · · · · · · · · · · · · · · ·
E[w]	Expected (average or mean) waiting time in the system (this includes both
	quedang time and service time; when the system is in the steady steep. The
	recet in a size rice filt.
E _k	Symbol for Erlang-k distribution of interarrival or service time.
$E[N_{\mathfrak{q}} N_{\mathfrak{q}}>0]$	Expected (average or mean) queue length of nonempty guerres when the
Ef -1	system is in the steady state.
$E[q \mid q > 0]$	Expected (average or mean) waiting time in queue for customers delayed
Ecte	when the system is in the steady state. Same as in
FCFS	Symbol for. hist come, first served " queue discipline
FIFO	Symbol for "first in, first out," queue discipline which is identical mich
6	· C. J.
G	Symbol for general probability distribution of service time. Independence
61	
GI	Symbol for general independent interarrival time distribution.
K	waxunum number allowed in queueing system, including both these switches
	tot service and mose receiving service. Also size of constation in finite
•	
L	E[N], expected (average or mean) number in the queueing system when the
	system is in the ready state.
μ (·)	The natural logarithm function or the logarithm to the base e.
Lq	Elives expected (average or mean) number in the quere not including those
	m 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 I CES
À	Average (mean) arrival rate to queueing system. $\lambda = 1/E[\tau]$, where $E[\tau] = 2\sqrt{E[\tau]}$
	average interarrival time.
1	

ÅT	Average throughput of a computer system measured in jobs or interactions per unit time.
М	•
	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.	Random variable describing number of customers in the steady state queue.
N,	Random variable describing number of customers receiving service when the system is in the steady state.
0	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.
$p_{\mathbf{z}}(t)$	Probability that there are n customers in the queueing system at time t.
	Steady state probability that there are n customers in the queueing system.
<i>p</i> ≈ PRI	Symbol for priority queueing discipline.
PS	Abbreviation for "processor-sharing queue discipline." See Section 6.2.1.
$\pi_{\mathbf{q}}(r)$	Symbol for rth percentile queueing time; that is, the queueing time that r
.467	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.
4	Random variable describing the time a customer spends in the queue (wait-
_	ing 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.
5	Random variable describing service time for one customer.
STRO	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. $W(t) = P[w \le t]$.
W	E[w], expected (average or mean) time in the steady state system.
$W_{a}(t)$	Distribution function for time in the queue, $W_{i}(t) = P[q \le t]$.
w,	E[q], expected (average or mean) time in the queue (waiting line), excluding
4	service time, for steady state system.
W _{qiq>0}	Expected (average or mean) queueing time for those who must queue. Same as $E[q \mid q > 0]$.
$W_{\epsilon}(t)$	Distribution function for service time, $W_t(t) = P(s \le t)$.
W.	$E[s]$, expected (average or mean) service time. $1/\mu$.
•	ments and an analysis of the

TABLE 2
Relationships Between Random Variables of Queueing Theory Models

$$u = E[s]/E[\tau] = \lambda E[s] = \lambda/\mu$$

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

$$L = E[N] = \lambda E[q] = \lambda E[q] = \lambda W_q$$
Average number in the queue for service for steady state system. $L = \lambda W_q$ is also called "Little's formula."

$$p_{n} = P[N = n] = (1 - \rho)\rho^{n}, \qquad n = 0, 1, 2,$$

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

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

$$L_{q} = E[N_{q}] = \rho^{2}/(1 - \rho), \qquad \sigma_{N_{q}}^{2} = \rho^{2}(1 + \rho - \rho^{2})/(1 - \rho)^{2}.$$

$$E[N_{q}|N_{q} > 0] = 1/(1 - \rho), \qquad Var[N_{q}|N_{q} > 0] = \rho/(1 - \rho)^{2}.$$

$$W[t] = P[w \le t] = 1 - e^{-\nu w}, \qquad P[w > t] = e^{-\nu w}.$$

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

$$\pi_{w}(90) = W \text{ in } 10 \approx 2.3W, \qquad \pi_{w}(95) = W \text{ in } 20 \approx 3W.$$

$$\pi_{w}(r) = W \text{ in } [100/(100 - r)].$$

$$W_{q}(t) = P[q \le t] = 1 - \rho e^{-\nu w}, \qquad P[q > t] = \rho e^{-\nu w}.$$

$$W_{q} = E[q] = \rho E[s]/(1 - \rho).$$

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

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

$$\pi_{q}(90) = W \text{ in } (10\rho), \qquad \pi_{q}(95) = W \text{ in } (20\rho).$$

$$\pi_{q}(r) = W \text{ in } \left(\frac{100\rho}{100 - r}\right).$$

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.

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

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

 $\lambda_z = (1 - p_z)\lambda$. λ_z is the actual arrival rate at which customers enter the system.

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

$$L_a = E[N_a] = L - (1 - p_0)$$

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

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

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

$$W = E[w] = L/\lambda_1$$
.

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

$$W_a = E[q] = L_a/\lambda_a.$$

$$E[q|q>0] = W_a/(1-p_0).$$

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

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

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

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

$$p_n = \begin{pmatrix} u^n \\ n! \end{pmatrix} p_0 \quad \text{if } n = 0, 1, \dots, c$$

$$\left| \frac{u^n p_0}{c! c^{n-\epsilon}} \right| \quad \text{if } n \ge c.$$

$$L_{q} = E[N_{q}] = \lambda W_{q} = \frac{uC(c, u)}{c(1-\rho)}, \qquad \sigma_{N_{q}}^{2} = \frac{\rho C(c, u)[1+\rho-\rho C(c, u)]}{(1-\rho)^{2}}.$$

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

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

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

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

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

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

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

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

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

where
$$C_1 = \frac{u - c + W_4(0)}{c - 1 - u}$$
 and $C_2 = \frac{C(c, u)}{c - 1 - u}$

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

$$E[w^2] = \frac{|2C(c, u)E[s]^2|}{|u - 1 - c|} \frac{|1 - c^2(1 - \rho)^2|}{|c^2(1 - \rho)^2|} + 2E[s]^2, \quad u = c - 1.$$

$$|4C(c, u)E[s]^2 + 2E[s]^2. \quad u = c - 1.$$

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

$$\pi_{-}(90) = W + 1.3\sigma_{-}$$

 $\pi_{-}(95) = W + 2\sigma_{-}$ Martin's estimates

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

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

$$\rho_0 = P[N = 0] = (1 - \rho)/(1 + \rho).$$

$$\rho_n = P[N = n] = 2\rho_0 \rho^n = \frac{2(1 - \rho)\rho^n}{(1 + \rho)}. \qquad n = 1, 2, 3,$$

$$\pi_n(90) \approx W + 1.3\sigma_n | \text{Martin's estimate.}$$

$$\pi_n(95) \approx W + 2\sigma_n | \text{Martin's estimate.}$$

$$L_n = E[N_n] = \frac{2\rho^3}{1 - \rho^2}. \qquad \sigma_{N_n}^2 = \frac{2\rho^3[(\rho + 1)^2 - 2\rho^3]}{(1 - \rho^2)^3}$$

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

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

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

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

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

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

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

$$\pi_n(2) = \frac{E[s]}{2(1 - \rho)} \ln \left(\frac{200\rho^2}{(1 + \rho)}\right). \qquad \pi_n(2) = \frac{2\rho^2}{2(1 - \rho)} \ln \left(\frac{40\rho^2}{1 + \rho}\right).$$

$$W(t) = P[w \le t] = \frac{1 - (1 - \rho)}{1 - \rho - 2\rho^2} e^{-2\omega t(1 - \rho)} \text{ if } u = 1.$$

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

$$U = E[s]/(1 - \rho^$$

 $\sigma_{-}^{2} = E[w^{2}] - E[w]^{2}.$

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

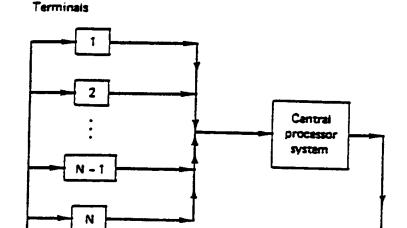


Fig. 6.3.1 Finite population queueing model of interactive computer system. Special case in which the central processor system consists of a single CPU with processor-sharing queue discipline.

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

$$\rho=1-\rho_0,$$

and the average throughput

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

The average response time

$$W = \frac{NE[s]}{1 - \rho_0} - E[t].$$

K Circulating markers (programs)

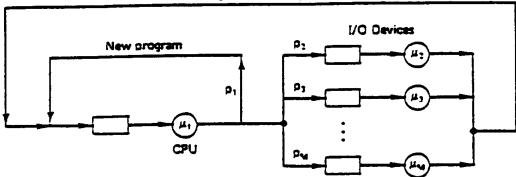


Fig. 6.3.4 Central server model of multiprogramming.

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

$$\rho_{i} = \begin{cases} G(K-1)/G(K) & i = 1\\ \frac{\mu_{1} \rho_{1} p_{i}}{\mu_{i}} & i = 2, 3, ..., \mathcal{M}. \end{cases}$$
 (6.3.25)

The average throughput it is given by

$$\lambda_{\mathsf{T}} = \mu_1 \rho_1 \rho_1. \tag{6.3.26}$$

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

$$W = \frac{N}{\lambda_T} - E[t] = \frac{N}{\mu_1 \rho_1 \rho_2} - E[t]. \tag{6.3.27}$$