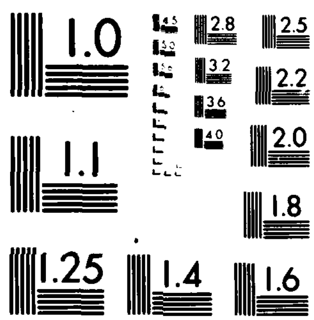
AD-A086 711 DELAWARE UNIV NEWARK APPLIED MATHEMATICS INST F/G 12/1
ALGORITHMIC ANALYSIS OF A MARKOVIAN MODEL FOR A SYSTEM WITH BAT--ETC(U)
MAR 80 J P COLARD, G LATOUCHE AFOSR-77-3236
UNCLASSIFIED TR-52B AFOSR-TR-80-0500 NL

1 of 1
AD
A086711

END
DATE
FILMED
8 80
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

APPLIED
MATHEMATICS INSTITUTE

University of Delaware
Newark, Delaware

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR-80-0500

2. GOVT ACCESSION NO.: AD-A086711

3. RECIPIENT'S CATALOG NUMBER:

4. TITLE (and Subtitle): ALGORITHMIC ANALYSIS OF A MARKOVIAN MODEL FOR A SYSTEM WITH BATCH AND INTERACTIVE JOBS.

5. TYPE OF REPORT & PERIOD COVERED: Interim

6. PERFORMING ORG. REPORT NUMBER:

7. AUTHOR(s): Jean Paul Colard, Guy Latouche

8. CONTRACT OR GRANT NUMBER(s): AFOSR-77-3236, NSF-ENG79-

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
University of Delaware
Applied Mathematics Institute
Newark, DE 19711

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F 2304/A5

11. CONTROLLING OFFICE NAME AND ADDRESS:
Air Force Office of Scientific Research/NM
Bolling AFB, Washington, DC 20332

12. REPORT DATE: Mar[redacted]80

13. NUMBER OF PAGES: 34

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): TR-52

15. SECURITY CLASS. (of this report): UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report):

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report):

18. SUPPLEMENTARY NOTES:

19. KEY WORDS (Continue on reverse side if necessary and identify by block number):

Queueing system, finite priority source, batch and interactive computer model, computational probability

20. ABSTRACT (Continue on reverse side if necessary and identify by block number):

A computer system serving both batch and interactive jobs is modeled as a single server queue, with an infinite source of ordinary customers and a finite source of priority customers.
The stability condition and the stationary probability distribution are determined. For a stable system, it is shown how the distribution and moments, of the waiting time, the sojourn time, the completion time, and the busy period, may be efficiently computed.

Algorithmic Analysis of a Markovian Model

for a System with Batch and Interactive Jobs

by

Jean Paul Colard and Guy Latouche

Laboratoire d'Informatique Theorique
Université Libre de Bruxelles
1050 Bruxelles BELGIUM

Applied Mathematics Institute
Technical Report Nr 52B
March 1980

This research was supported in part by the National Science Foundation under Grant Nr ENG-7908351 and by the Air Force Office of Scientific Research under Grant Nr AFOSR-77-3236.

Abstract

A computer system serving both batch and interactive jobs is modeled as a single server queue, with an infinite source of ordinary customers and a finite source of priority customers.

The stability condition and the stationary probability distribution are determined. For a stable system, it is shown how the distribution and moments, of the waiting time, the sojourn time, the completion time, and the busy period, may be efficiently computed.

Keywords

Queueing system, finite priority source, batch and interactive computer model, computational probability.

Introduction

The purpose of this paper is to present the algorithmic analysis of a single model for a computer system, serving batch and interactive jobs. The methodology used here has been introduced in recent years by Neuts [8-12].

The computer system under study is represented by a single server queue with infinite waiting room (see Figure 1). Batch jobs arrive according to a Poisson process with parameter λ_1. The time needed to process a batch job is exponential with parameter μ_1. Once a batch job is terminated, it leaves the system. Interactive jobs are submitted by a finite number, N, of interactive terminals. The time needed by a terminal to submit a job is exponential with parameter λ_2. Once the job is submitted, the terminal must wait until the job is processed, at which time the terminal initiates the submission of a new job. The time needed by the server to process an interactive job is exponential with parameter μ_2. The interactive jobs have preemptive priority over the batch jobs. It will readily be observed that such a system is a special case of the priority queueing model, with infinite ordinary source and finite priority source, which has already been studied by Avi-Itzhak and Naor [1], Jaiswal and Thiruvengadam [3], Thiruvengadam [13,14], and most comprehensively by Jaiswal [4]. The results in [4] are obtained via the analysis of the renewal process formed by successive busy periods of the queueing system. The results mainly consist of Laplace-Stieltjes transforms of distributions or of generating functions. These do not simplify very much in the single case considered here, where the service time distributions are exponential.

In contrast, the method used here will yield computationally efficient expressions. We shall determine the stability condition for the system and prove that the stationary probability vector has a matrix-geometric form.

Appropriately partitioning that vector $\underline{x}$ as $(\underline{x}_0, \underline{x}_1, \ldots)$, it is shown that

$$\underline{x}_i = \underline{\pi}(I-R)R^i, \qquad \text{for } i \geq 0,$$

where the positive vector $\underline{\pi}$, and the positive matrix R are explicitly determined. We show in Sections 2 to 4 how the distribution and moments of the waiting time, the sojourn time, the completion time, and the busy period are determined. A comprehensive discussion of stochastic models with embedded Markov chains having a matrix-geometric invariant probability vector may be found in Neuts [12].

To conclude this introduction, we remark that the simple network described in Figure 1 is not amenable to the approach of Baskett et.al. [2], since the two priority classes of jobs have different service rates.

1. The Stationary Distribution

Under the assumptions of Poisson arrivals and exponential services, the model may be described as a continuous-parameter Markov chain on the state space $\{(i,j),\ i \geq 0,\ 0 \leq j \leq N\}$, where i and j respectively denote the number of batch and interactive jobs in the queueing system.

The infinitesimal generator Q of the Markov chain is a block-tridiagonal matrix of the form

$$Q = \begin{bmatrix} A_1+A_2 & A_0 & 0 & 0 & \cdots \\ A_2 & A_1 & A_0 & 0 & \cdots \\ 0 & A_2 & A_1 & A_0 & \cdots \\ 0 & 0 & A_2 & A_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}, \tag{1}$$

where A_0, A_1 and A_2 are square matrices of order N+1, defined as follows. The matrix A_0 corresponds to transitions from states (i,j) to (i+1,j'), for $i \geq 0$, and is equal to $\lambda_1 I$. The matrix A_2 corresponds to transitions from states (i,j) to $(i-1,j')$, for $i \geq 1$; $(A_2)_{0,0}$ is equal to μ_1, all the other elements of A_2 are equal to zero. The matrix A_1 corresponds to transitions from states (i,j) to (i,j'), for $i \geq 1$, and is given by

$$A = \begin{bmatrix} -\lambda_1-N\lambda_2-\mu_1 & N\lambda_2 & 0 & \cdots & 0 \\ \mu_2 & -\lambda_1-(N-1)\lambda_2-\mu_2 & (N-1)\lambda_2 & \cdots & 0 \\ 0 & \mu_2 & -\lambda_1-(N-2)\lambda_2-\mu_2 & & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ & & & & \lambda_2 \\ 0 & 0 & \cdots & \mu_2 & -\lambda_1-\mu_2 \end{bmatrix}.$$

Let $\underline{x}$ denote the vector of stationary probabilities. It is the unique solution to the system $\underline{x}Q=\underline{0}$, $\underline{x}\ \underline{e}=1$, where $\underline{e}$ represents a vector with every component equal to 1. Let us partition $\underline{x}$ as $(\underline{x}_0,\underline{x}_1,\ldots)$, where $\underline{x}_i$ is an N+1-vector, and corresponds to the states $\{(i,j),\ 0 \leq j \leq N\}$. Furthermore, we define the matrix A by $A=A_0+A_1+A_2$. It is the infinitesimal generator of a finite, irreducible, continuous-parameter Markov chain. By $\underline{\pi}$ we denote its vector of stationary probabilities, i.e. $\underline{\pi}A=\underline{0}$, $\underline{\pi}\ \underline{e}=1$.

Lemma 1

The vector $\underline{\pi}$ is given by

$$\pi_j = \frac{N!}{(N-j)!}\rho_2^j \left[\sum_{k=0}^{N} \frac{N!}{(N-k)!}\rho_2^k\right]^{-1}, \qquad \text{for } 0 \leq j \leq N,$$

where $\rho_2=\lambda_2/\mu$.

Proof. The proof is elementary.

Theorem 1

The system is stable if and only if

$$\rho = \lambda_1 \left[\sum_{k=0}^{N} \frac{N!}{(N-k)!} \rho_2^k \right] / \mu_1 < 1. \tag{2}$$

If the system is stable, the stationary probability vector $\underline{x}$ is given by $\underline{x}=(\underline{x}_0,\underline{x}_1,\ldots)$, where

$$\underline{x}_i = \underline{\pi}(I-R)R^i, \qquad \text{for } i \geq 0. \tag{3}$$

The matrix R is the unique nonnegative solution with maximal eigenvalue strictly less than one of the matrix quadratic equation

$$R^2A_2+RA_1+A_0 = 0. \tag{4}$$

The matrix R is equal to $\lim_{n\to\infty} R_n$, where the matrices R_n are defined as follows,

$$\begin{aligned} R_0 &= 0, \\ R_{n+1} &= -A_0A_1^{-1}-R_n^2A_2A_1^{-1}, \qquad \text{for } n\geq 0. \end{aligned} \tag{5}$$

Moreover, the matrix R is strictly positive, and its first column is equal to $\rho_1\underline{e}$, where $\rho_1=\lambda_1/\mu_1$.

Proof. This theorem is proved by repeating arguments given in [8-11]. We refer the reader to those papers for details and only indicate the main steps of the proof here.

Firstly, it results from [11], Theorem 1, that if the system is stable, then $\underline{x}_i=\underline{x}_0R^i$, for $i\geq 0$, where the matrix R satisfies the equation (4), the maximal eigenvalue of R is strictly less than one, the matrix R is the minimal nonnegative solution of (4).

Secondly, we show that the sequence of matrices R_n defined in (5), converges monotonically to the minimal nonnegative solution of (4). Also, we show that R_n is strictly positive, for $n\geq 1$.

Thirdly, we prove that the matrix quadratic equation (4) has a unique solution with maximal eigenvalue strictly less than one if and only if $\underline{\pi}A_2\underline{e} > \underline{\pi}A_0\underline{e}$. It results from Lemma 1 that the stability condition $\underline{\pi}A_2\underline{e} > \underline{\pi}A_0\underline{e}$ is equivalent to the inequality (2). It is easy to verify that if the system is stable, then $\underline{x}_0 = \underline{\pi}(I-R)$, which completes the proof of (3). Finally, we show that $RA_2\underline{e} = A_0\underline{e}$, which proves that the first column of R is equal to $\rho_1\underline{e}$.

In general, for systems which may be analyzed by the approach followed here, the matrix R must be numerically computed by using the recurrence relation (5). In our case however, it is possible to go further in the determination of the matrix R, as we show in the next theorem.

Theorem 2

The matrix R is explicitly given by

$$R = -\lambda_1 A_1^{-1} + \frac{\lambda_1^2}{1+\lambda_1 \underline{m}\,\underline{e}} A_1^{-1} M, \tag{6}$$

where $\underline{m}$ denotes the first row of the matrix A_1^{-1}, and the matrix M is equal to $\underline{e}\cdot\underline{m}$, i.e. each row of the matrix M is equal to $\underline{m}$.

Proof. The matrix R satisfies the equation

$$R = -A_0 A_1^{-1} - R^2 A_2 A_1^{-1}. \tag{7}$$

From the structure of A_0 and A_2, it results that (7) may be written as

$$R = -\lambda_1 A_1^{-1} - \mu_1 R^2 M^*, \tag{8}$$

where the first row of M^* is equal to $\underline{m}$, the remaining rows are equal to $\underline{0}$. Clearly, (8) is equivalent to

$$R = -\lambda_1 A_1^{-1} - \mu_1 R^* M^* = -\lambda_1 A_1^{-1} - \mu_1 \underline{r}^*\cdot\underline{m}, \tag{9}$$

where $\underline{r}^*$ denotes the first column of R^2, the first column of R^* is equal to $\underline{r}^*$, the remaining columns are equal to $\underline{0}$. Since the first column of R is equal to $\rho_1\underline{e}$ (by Theorem 1), then $\underline{r}^*=\rho_1 R\underline{e}$, and (9) becomes

$$R = -\lambda_1 A_1^{-1} - \lambda_1 R\underline{e}\cdot\underline{m}. \tag{10}$$

Postmultiplying both sides of (10) by $\underline{e}$, and replacing the obtained value for $R\underline{e}$ in (10), completes the proof of the theorem.

It is now a simple matter to determine special probabilities of interest. Let $\underline{p}$ and $\tilde{\underline{p}}$ respectively represent the marginal distribution of the number of interactive and batch jobs. Clearly, one has that $\underline{p} = \sum_{i=0}^{\infty} \underline{x}_i = \underline{\pi}$. The vector $\underline{\pi}$ is explicitly given in Lemma 1, and there is no difficulty in computing the moments for the marginal distribution of the interactive jobs. To determine the distribution and moments for the batch jobs is only slightly more difficult. We shall only state the following result. The proof is identical to that of [5], Lemma 1, and is not presented here.

<u>Lemma 2</u>

The marginal distribution of the number of batch jobs is given by

$$\tilde{p}_i = \underline{\pi}(I-R)R^i\underline{e}, \qquad \text{for } i\geq 0.$$

The ν^{th} moment β_ν, about the origin, is equal to $\beta_\nu=\underline{\pi}X_\nu\underline{e}$, for $\nu\geq 1$, where the matrices X_ν of order N+1 are recursively defined as follows.

$$X_0 = I,\ X_{n+1} = R(I-R)^{-1} \sum_{j=0}^{n} \binom{n+1}{j} X_j, \qquad \text{for } n\geq 0.$$

In particular, one has that

$$\beta_1 = \underline{\pi}R(I-R)^{-1}\underline{e},$$
$$\beta_2 = \underline{\pi}R(I+R)(I-R)^{-2}\underline{e},$$
$$\text{and} \quad \beta_3 = \underline{\pi}R(I+4R+R^2)(I-R)^{-3}\underline{e}.$$

Finally, we determine easily the following probabilities for the state of the server. The probability that the server is idle is equal to $x_{0,0}=\pi_0-\rho_1$. The probability that the server is busy and processing a batch job is equal to $\sum_{i=1}^{\infty} x_{i,0}=\rho_1$. The probability that the server is busy and processing an interactive job is equal to $1-p_0=1-\pi_0$.

2. The Waiting Time and the Sojourn Time for a Batch Job

It is obvious that the interactive jobs actually form an M/M/1 queue with finite source. This type of queue has been widely studied and therefore, we will only examine the batch jobs. This can be done by any of two methods. Firstly, by using the concept of phase type distributions - Neuts [6,7] - we obtain an explicit form for the probability distributions. The second approach provides expressions which are better suited to numerical computations.

We consider the queues in steady-state and we denote by w the time spent by a batch job in the incoming queue, by $W_q(x,i)$, the conditional distribution of w, given that upon arrival of the job, the system contained i batch jobs already and by $W_q(x)$ the distribution of w.

We will need the following matrices and vectors: S_0 is the square matrix of order N obtained by deleting the first row and the first column of A_1+A_0, S_i $(i>0)$ is the square matrix of order $i(N+1)$ given by:

$$\begin{bmatrix} C_1 & & & & \\ C_0 & C_1 & & 0 & \\ & C_0. & C_1 & . & \\ 0 & & \cdot & C_0 & \cdot C_1 \end{bmatrix},$$

where $C_1=A_1+A_0$ and $C_0=A_2$; $\underline{\alpha}_0$ is the N-vector obtained by deleting the first component of $\underline{x}_0$, and $\underline{\alpha}_i$ (i>0) is the i(N+1)-vector $(\underline{0},\underline{0},\ldots,\underline{0},\underline{x}_i)$.

Theorem 3

The distribution of w is given by:

$$W_q(x) = 1 - \sum_{i=0}^{\infty} \underline{\alpha}_i \exp(S_i x)\underline{e}, \qquad \text{for } x\geq 0.$$

The moments are given by:

$$E[w^\nu] = (-1)^\nu \cdot \nu! \sum_{i=0}^{\infty} \underline{\alpha}_i S_i^{-\nu} \underline{e}, \qquad \text{for } \nu\geq 0.$$

Proof. Given that the system contains i batch jobs already, w is equal to the time until absorption in a continuous parameter Markov chain, with transitions among transient states governed by S_i, and initial probability vector $(\underline{x}_i\underline{e})^{-1}\underline{\alpha}_i$. Therefore

$$W_q(x,i) = 1-(\underline{x}_i\underline{e})^{-1}\underline{\alpha}_i \exp(S_i x)\underline{e}, \qquad \text{for } x\geq 0,$$

and $$E[w^\nu|i] = (-1)^\nu\, \nu!(\underline{x}_i\underline{e})^{-1}\underline{\alpha}_i S_i^{-\nu}\underline{e}.$$

The proof is now immediate.

Remark. Although this theorem provides us with an explicit expression for $W_q(x)$, it is difficult to use in order to get numerical results, because it contains a double series $\sum_{i=0}^{\infty} \underline{\alpha}_i \left(\sum_{n=0}^{\infty} \frac{S_i^n x^n}{n!} \right) \underline{e}$.

These series are convergent, but involve large matrices. Nevertheless, the distribution of w may numerically be obtained by another method.

Clearly, w is the time until absorption in the Markov chain with infinitesimal generator Q_w and initial probability vector $\underline{x}=(\underline{x}_0,\underline{x}_1,\ldots,\underline{x}_k,\ldots)$, where

$$Q_w = \begin{bmatrix} C & 0 & 0 & 0 & \ldots \\ C_0 & C_1 & 0 & 0 & \ldots \\ 0 & C_0 & C_1 & 0 & \ldots \\ \vdots & 0 & C_0 & C_1 & 0 \\ & & \ddots & \ddots & \ddots \end{bmatrix},$$

and C is the matrix C_1 with a first row identically zero.

Let $y_{ij}(x)$ be the probability that this Markov chain is in the state (i,j) at time x. We have that $W_q(x)=y_{00}(x)$. The Kolmogorov equations for this chain are

$$\begin{aligned} \underline{y}_0'(x) &= \underline{y}_0(x)C+\underline{y}_1(x)C_0, \\ \underline{y}_i'(x) &= \underline{y}_i(x)C_1+\underline{y}_{i+1}(x)C_0, \qquad \text{for } i\geq 1, \end{aligned} \tag{11}$$

with the initial conditions $\underline{y}_i(0)=\underline{x}_i$.

The Markov chain Q_w can only move towards lower states. It is therefore obvious how to truncate the infinite system of differential equations (11). In order to lose a probability mass of at most ε in the tail of the distribution function $W_q(x)$, one truncates at the index K such that

$$\sum_{\nu=K+1}^{\infty} \underline{\pi}(I-R)R^{\nu}\underline{e} = \underline{\pi}R^{K+1}\underline{e} < \varepsilon$$

This approach does not lead to an explicit form for $W_q(x)$, but is easily implemented.

We define the sojourn time of a batch job in the system as the time between the moment the job enters the system and the moment it leaves the system. The sojourn time can be analyzed exactly like the waiting time.

3. The Completion Time of a Batch Job

Let $w_c(x)$ denote the distribution function of the completion time w_c of a batch job, i.e. the time between the moment when the processing of the job begins, and the moment when the job leaves the system. w_c is also the time until absorption into the absorbing state of the Markov chain with (N+2) states and infinitesimal generator Q_c given by

$$Q_c = \left[\begin{array}{c|ccc} 0 & 0 & \cdots & 0 \\ \hline \mu_1 & & & \\ 0 & & C_1 & \\ \vdots & & & \\ 0 & & & \end{array}\right],$$

and with the initial probability vector $\underline{v}=(0,\underline{e}_0)=(0,1,0,\ldots,0)$. Clearly now,

$$\begin{aligned} W_c(x) &= 1-\underline{e}_0 \exp(C_1 x)\underline{e}, \qquad \text{for } x\geq 0, \\ \text{and } E[w_c^\nu] &= (-1)^\nu \nu! \underline{e}_0 C_1^{-\nu} \underline{e}, \qquad \text{for } \nu\geq 1. \end{aligned} \tag{12}$$

Remarks.

a. $E[w_c]$ is equal to $\frac{\rho}{\lambda_1} = \frac{1}{\mu_1 \pi_0}$. (13)

Indeed, $E[w_c]$ is equal to minus the sum of the elements of the first row of C_1^{-1}; these elements are easily computed, since they are the solution of a system that differs only from the system for $\underline{\pi}$ in the first equation. It is now easy to prove (13).

b. Equation (12) may be written as

$$W_c(x) = -\sum_{k=1}^{\infty} c_k x^k, \tag{14}$$

where c_k is the sum of the elements of the first row of $\frac{1}{k!} C_1^k$. It is not necessary however to compute all the components of C_1^k in order to obtain c_k. If we denote by $\underline{c}^{(k)}$ the first row of $\frac{1}{k!} C_1^k$, one has that

$$\underline{c}^{(k)} = \frac{1}{k} \underline{c}^{(k-1)} C_1,$$

$$\text{and} \quad c_k = \underline{c}^{(k)} \underline{e} = - \frac{\mu_1}{k} \underline{c}^{(k-1)} \underline{e}_0.$$

c. The series (14) is unfortunately difficult to use for numerical purposes because its terms alternate in sign and have large absolute values (greater than 10^{30} in some cases). Since this series has a sum in the interval (0,1), it is clear that it is computationally unstable. Nevertheless, one easily obtains the distribution of w_c as the solution of a finite system of differential equations by using the same approach as for the waiting time.

4. The Busy Period of the System

The busy period is defined as the interval of time between the moment when the CPU becomes active (by the arrival of either an interactive or a batch job) and the first moment when the CPU becomes again inactive. We will have to distinguish two different types of busy periods, according as it begins with the arrival of an interactive (type 1) or a batch job (type 0).

We denote by $h_0(k,\ell,x)$, $k \geq 0$, $\ell \geq 0$, the probability that starting in the state $(i+1,0)$, $i \geq 0$, at time 0, in the Markov chain with infinitesimal generator Q, the first visit to the state $(i,0)$ occurs no later than time x, and that exactly k batch and ℓ interactive jobs are processed during that first passage time. Similarly, we define $h_j(k,\ell,x)$, $j=1,2,\ldots,N$, for the first passage time from the state (i,j), $i \geq 0$, to the state $(i,0)$. Furthermore, we introduce the transforms

$$h_j^*(z,y,s) = \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} z^k y^\ell \int_0^\infty e^{-sx} dh_j(k,\ell,x).$$

$H^*(z,y,s)$ is a square matrix of order $(N+1)$ such that $H_{j0}^*(z,y,s)=h_j^*(z,y,s)$ for $j=0,1,\ldots,N$, the remaining elements are all zero and $\bar{H}=H^*(1,1,0)$. Finally, A_2' is a square matrix of order $(N+1)$ with only one non-zero element: $(A_2')_{10} = \mu_2$.

By following the argument in [9], we prove that the matrix $\bar{H}$ is stochastic if the system is stable. Then each row of $\bar{H}$ is equal to $\underline{e}_0$.

Let us denote by $\underline{m}$, $\underline{n}_B$, $\underline{n}_I$ the following vectors

$$\underline{m} = -\left[\frac{\partial}{\partial s} H^*(z,y,s)\underline{e}\right]_{(z=1,y=1,s=0)},$$

$$\underline{n}_B = \left[\frac{\partial}{\partial z} H^*(z,y,s)\underline{e}\right]_{(z=1,y=1,s=0)},$$

$$\underline{n}_I = \left[\frac{\partial}{\partial y} H^*(z,y,s)\underline{e}\right]_{(z=1,y=1,s=0)}.$$

Their components m_p, n_{Bp}, n_{Ip}, for $p=0,1$, are respectively equal to the mean busy period and the mean numbers of batch and interactive jobs processed during a busy period of type p.

<u>Theorem 4</u>

If the queueing system is stable ($\rho<1$), the vectors $\underline{m}$, $\underline{n}_B$, and $\underline{n}_I$ are the solutions of

$$M\underline{m} = -\underline{e} \tag{15}$$

$$M\underline{n}_B = -\underline{\mu}_1 \tag{16}$$

$$M\underline{n}_I = -\underline{\mu}_2 \tag{17}$$

where M is a non-singular matrix given by

$$M = A_1+A_0+A_0\bar{H}-A_2' + \frac{N\lambda_2}{\mu_1} A_2$$

and $\underline{\mu}_1 = {}^t(\mu_1,0,0,\ldots,0)$

$$\underline{\mu}_2 = {}^t(0,\mu_2,0,\ldots,0)$$

One obtains the following explicit forms:

$$m_0 = \frac{1}{\mu_1\pi_0(1-\rho)}, \qquad m_1 = \frac{1-\pi_0}{\pi_0 N\lambda_2(1-\rho)}, \tag{18}$$

$$n_{B0} = \frac{1}{(1-\rho)}, \qquad n_{B1} = \frac{\lambda_1(1-\pi_0)}{\pi_0 N\lambda_2(1-\rho)}, \tag{19}$$

$$n_{I0} = \frac{N\lambda_2}{\mu_1(1-\rho)}, \qquad n_{I1} = \frac{1-\rho\pi_0}{1-\rho}. \tag{20}$$

Proof. The proof of this theorem is purely technical and may be found in the appendix.

5. Other Models

In this section we shall briefly examine three queueing systems closely related to the system defined in the introduction.

Firstly, we shall consider the case where interactive jobs have non-preemptive priority over the batch jobs. If an interactive job is submitted while a batch job is being processed, the interactive jobs in the queue must wait until that job leaves the system. In this case, the state space may be represented by $\{(i,j),\ i=0,\ 0\le j\le N;\ i>0,\ -N\le j\le N\}$, where $j>0$ indicates that a batch job is being processed, while $j<0$ indicates that a batch job is being processed, and $|j|$ interactive jobs are waiting. The infinitesimal generator Q is now given by

$$Q = \begin{bmatrix} B_1 & B_0 & 0 & \cdots \\ B_2 & A_1 & A_0 & \cdots \\ 0 & A_2 & A_1 & A_0 \\ \vdots & \vdots & \vdots & & \ddots \end{bmatrix}, \qquad (21)$$

where the square matrices A_0, A_1, and A_2 are of order 2N+1, the square matrix B_1 is of order N+1, the matrices B_0 and B_2 are rectangular, with the appropriate dimensions. Each of these matrices is very sparse, we do not indicate their structure, since it is cumbersome and would serve no purpose here. It is clear that this model may be solved by using exactly the same approach as in Sections 1 through 4.

Secondly, we may consider the preemptive-delayed priority discipline. If an interactive job is submitted while a batch job is being processed, the latter is allocated a quantum of time, say D, during which it may terminate its processing. If it does not leave the system before D, it is sent back in the waiting queue and the interactive jobs are processed. If we assume that D has an exponential distribution, then the system is very similar to the non-preemptive priority models. Only the matrix A_1 in (21) is modified.

Thirdly, let us assume that the system contains $c \geq 1$ servers. Then some modifications occur, and we briefly present the most important ones. For each of the three queueing disciplines, the structure of the matrix Q is slightly modified. In addition, for the preemptive-delayed and the non-preemptive disciplines, a new variable must be introduced to describe the mixture of batch and interactive jobs being processed. This adds to the dimensionality of the problem since the elements themselves of the matrices A and B in (21) become matrices. For the preemptive discipline, no new variable is necessary and the stationary distribution, the waiting time, and the busy period may be analyzed in much the same way as in Sections 1, 2, and 4. However, the analysis of the completion time of a batch job will be different. It will be necessary to introduce a rule for selecting which batch job is to be interrupted. Different rules will yield different distributions for the completion time. It is our intention to discuss this point elsewhere.

6. Numerical Results

The parameters used in this section have been obtained after the analysis of some statistics from the Computer Center at our University; we obtained the following values:

$$\lambda_1 = 0.030 \text{ sec}^{-1}$$
$$\lambda_2 = 0.056 \text{ sec}^{-1}$$
$$\mu_1 = 0.5 \quad \text{sec}^{-1}$$
$$\mu_2 = 2.4 \quad \text{sec}^{-1}$$
$$N = 43 \text{ (mean number of active terminals).}$$

We used four of these values and let the fifth one vary. We will show here some results obtained by varying the number of terminals or the interactive service rate.

We show in Figure 2 how the traffic coefficient $\rho(n)$ depends on the number of terminals. It may be observed that 43 active terminals is a number near the critical region; a few supplementary terminals will bring the system to saturation.

Table I gives, for different values of N, the following information:

- the traffic coefficient ρ defined in Theorem 1;
- the mean number E[j] of interactive jobs and the percentiles .5, .9, .99, and .999, of the distribution $\underline{p}$;
- the mean number $E[i]_{\sim}$ of batch jobs and the same percentiles of the distribution $\underline{p}$;

- the mean value $E[w_I]$ and the percentiles .5, .9, .99 and .999, for the waiting time distribution of interactive jobs;
- the expected waiting time $E[w]$ and completion time $E[w_c]$ for batch jobs, the mean number m_{Ic} of interactive jobs processed during the completion time of a batch job;
- the expected length m_0 and m_1 of a busy period of type 0 and type 1.

On Figures 3 and 4 are represented the waiting time and completion time distributions for batch jobs for the same values of N as in Table I.

It appears clearly that increasing the number of terminals does not influence much the purely interactive components of the system, but disturbs much more the batch components, especially in the critical region N=35 to 47 (note that the system is unstable for N≥48).

Remark. To produce the curves of Figure 3, we have solved the system (11) for each value of N. The system was truncated at the percentile .999 in order to lose at most a probability mass .001 in the tail of $W_q(x)$. We have not computed $W_q(x)$ for N=47, because the resulting finite system of differential equations would have been of order 48×125 which was too large for our program to handle.

In Table II and Figures 5 and 6, we represent essentially the same information as in Table I and Figures 3 and 4 for different values of μ_2. The last row of Table II indicates the maximum number N_{max} of terminals that may be active without causing the system to become unstable.

It appears that improving the performances of the interactive components of the system has a profound effect on the whole system, especially on the batch activity. This is reflected on each characteristic, whether they be moments or percentiles of queue length or waiting time distributions (see also Table III).

Appendix

Firstly, we prove that matrix $H^*(z,y,s)$ is a solution of the following matrix equation:

$$(A_0+A_2'')H^{*2}+(A_1-sI-A_2'-A_2'')H^*+zA_2+yA_2' = 0 \qquad (A.1)$$

where A_2'' is a square matrix of order (N+1) with the unique non-zero element $(A_2'')_{01} = N\lambda_2$.

Proof. One easily proves the following equations:

$$\begin{aligned}
\frac{d}{dx} h_0(k,\ell,x) &= \lambda_1 h_0 * h_0(k,\ell,x) \\
&\quad + N\lambda_2 h_1 * h_0(k,\ell,x) \\
&\quad + \mu_1 \delta_{k1} \delta_\ell \\
&\quad - [\lambda_1+N\lambda_2+\mu_1] h_0(k,\ell,x), \\
\frac{d}{dx} h_1(k,\ell,x) &= \lambda_1 h_1 * h_0(k,\ell,x) \\
&\quad + (N-1)\lambda_2 h_2(k,\ell,x) \\
&\quad + \mu_2 \delta_{\ell 1} \delta_{k0} \\
&\quad - [\lambda_1+(N-1)\lambda_2+\mu_2] h_1(k,\ell,x), \\
\frac{d}{dx} h_j(k,\ell,x) &= \lambda_1 h_j * h_0(k,\ell,x) \\
&\quad + (N-j)\lambda_2 h_{j+1}(k,\ell,x) \\
&\quad + \mu_2 h_{j-1}(k,\ell-1,x) \\
&\quad - [\lambda_1+(N-j)\lambda_2+\mu_2] h_j(k,\ell,x), \text{ for } 2 \le j \le M.
\end{aligned}$$

Remark. We define $h_{N+1}(k,\ell,x) = 0$ for all (k,ℓ,x).

Taking the generating functions of the Laplace-Stieltjes transforms, we get (writing h_j^* for $h_j^*(z,y,s)$):

$$\lambda_1 h_0^{*2}+N\lambda_2 h_1^* h_0^*-[\lambda_1+N\lambda_2+\mu_1]h_0^*-sh_0^*+\mu_1 z = 0,$$

$$\lambda_1 h_1^* h_0^*-[\lambda_1+(N-1)\lambda_2+\mu_2]h_1^*+(N-1)\lambda_2 h_2^*-sh_1^*+\mu_2 y = 0,$$

$$\lambda_1 h_j^* h_0^*-[\lambda_1+(N-j)\lambda_2+\mu_2]h_j^*+(N-j)\lambda_2 h_{j+1}^*+\mu_2 h_{j-1}^*-sh_j^* = 0.$$

This last system becomes in matrix notation:

$$A_0 H^{*2}+A_2'' H^{*2}+A_1 H^*-A_2'' H^*-A_2' H^*-sH^*+zA_2+yA_2' = 0.$$

One now easily obtained Formulas (15) to (17) by deriving formula (A.1) respectively with respect to s, z, or y, by right multiplying by $\underline{e}$, and by taking the result for z=1, y=1, and s=0.

The non-singularity of the matrix M follows from the fact that det $M=-\mu_1(-\mu_2)^N(1-\rho) \neq 0$ for $\rho<1$.

In order to obtain Formulas (18) to (20), it is not necessary to inverse the whole matrix M, it is sufficient to determine the first two rows $\underline{r}_0$ and $\underline{r}_1$ of M^{-1}. They can be obtained by solving two linear systems very similar to the system providing the vector $\underline{\pi}$. Now Formulas (18) to (20) immeidately result from

$$m_0 = -\underline{r}_0\underline{e}, \qquad m_1 = -\underline{r}_1\underline{e},$$

$$m_{B0} = -\underline{r}_0\underline{\mu}_1, \qquad m_{B1} = -\underline{r}_1\underline{\mu}_1,$$

$$n_{I0} = -\underline{r}_0\underline{\mu}_2, \qquad n_{I1} = -\underline{r}_1\underline{\mu}_2.$$

References

[1] AVI-ITZHAK, B. and NAOR, P. (1961).
On a problem of preemptive priority queueing.
Operations Research, IX, 664-672.

[2] BASKETT, F., CHANDY, K. M., MUNTZ, R. R., and PALACIOS, F. G. (1975).
Open, closed, and mixed networks of queues with different classes of customers.
Journal of the ACM, XXII, 248-260.

[3] JAISWAL, N. K. and THIRUVENGADAM, K. (1967).
Finite-source priority queues.
SIAM Journal on Applied Mathematics, XV, 1278-1293.

[4] JAISWAL, N. K. (1968).
Priority Queues.
Academic Press, New York and London.

[5] LATOUCHE, G. (1978).
Algorithmic analysis of a multiprogramming-multiprocessor computer system.
Institute of Mathematical Sciences, Univ. of Delaware, Newark, Delaware
Department of Statistics and Computer Science, Technical Report No. 78/13.

[6] NEUTS, M. F. (1975).
Probability distributions of phase type.
in - Liber Amicorum Professor Emeritus H. Florin - pp. 173-206,
Department of Mathematics, University of Louvain, Belgium.

[7] NEUTS, M. F. (1978).
Renewal processes of phase type.
Naval Research Logistics Quarterly, XXV, 445-454.

[8] NEUTS, M. F. (1978).
Markov chains with applications in queueing theory, which have a matrix-geometric invariant probability vector.
Adv. Appl. Prob., X, 185-191.

[9] NEUTS, M. F. (1978).
The M/M/1 Queue with randomly varying arrival and service rates.
Opsearch, XV, 139-157.

[10] NEUTS, M. F. (1978).
Further results on the M/M/1 queue with randomly varying rates.
Opsearch, XV, 158-168.

[11] NEUTS, M. F. (1980).
The probabilistic significance of the rate-matrix in matrix-geometric invariant vectors.
J. Appl. Prob., (to appear).

[12] NEUTS, M. F.
Matrix-geometric Solutions in Stochastic Models - An Algorithmic Approach.
The Johns Hopkins University Press - to appear.

[13] THIRUVENGADAM, K. (1964).
A priority assignment in machine interference problem.
Opsearch, I, 197-216.

[14] THIRUVENGADAM, K. (1965).
Studies in waiting line problems.
Ph.D. Thesis, University of Delhi, Delhi, India.

Titles of Tables

Titles of Figures

Table I

Characteristics of the system behaviour for different values of N.
(λ_1=.030, λ_2=.056, μ_1=.5, μ_2=2.4)

N	20	25	30	35	40	45	47
ρ	.108	.134	.174	.243	.375	.677	.909
E[j]	.75	1.13	1.67	2.47	3.71	5.63	6.65
.5	0	1	1	2	3	5	6
.9	2	3	4	6	8	12	13
.99	5	6	8	10	13	17	19
.999	7	8	11	14	17	21	23
E[i]	.14	.19	.28	.46	.94	3.59	17.60
.5	0	0	0	0	0	2	12
.9	1	1	1	1	3	9	41
.99	2	2	3	4	6	18	83
.999	3	3	4	6	9	28	125
$E[w_I]$	.31	.47	.69	1.03	1.55	2.35	2.77
.5	.00	.09	.31	.62	1.12	1.93	2.38
.9	1.06	1.47	1.99	2.73	3.76	5.17	5.84
.99	2.63	3.29	4.13	5.23	6.65	8.42	9.22
.999	4.13	5.00	6.08	7.43	9.09	11.06	11.93
E[w]	1.04	1.81	3.38	7.14	18.91	97.08	556.44
$E[w_c]$	3.61	4.47	5.81	8.10	12.50	22.55	30.30
m_{Ic}	3.86	5.93	9.15	14.63	25.21	49.33	67.93
m_0	4.04	5.16	7.04	10.69	20.01	69.74	333.40
m_1	.81	1.03	1.38	2.07	3.78	12.71	59.63

Table II

Characteristics of the system behaviour for different values of μ_2.
(λ_1=.030, λ_2=.056, μ_1=.5, N=43)

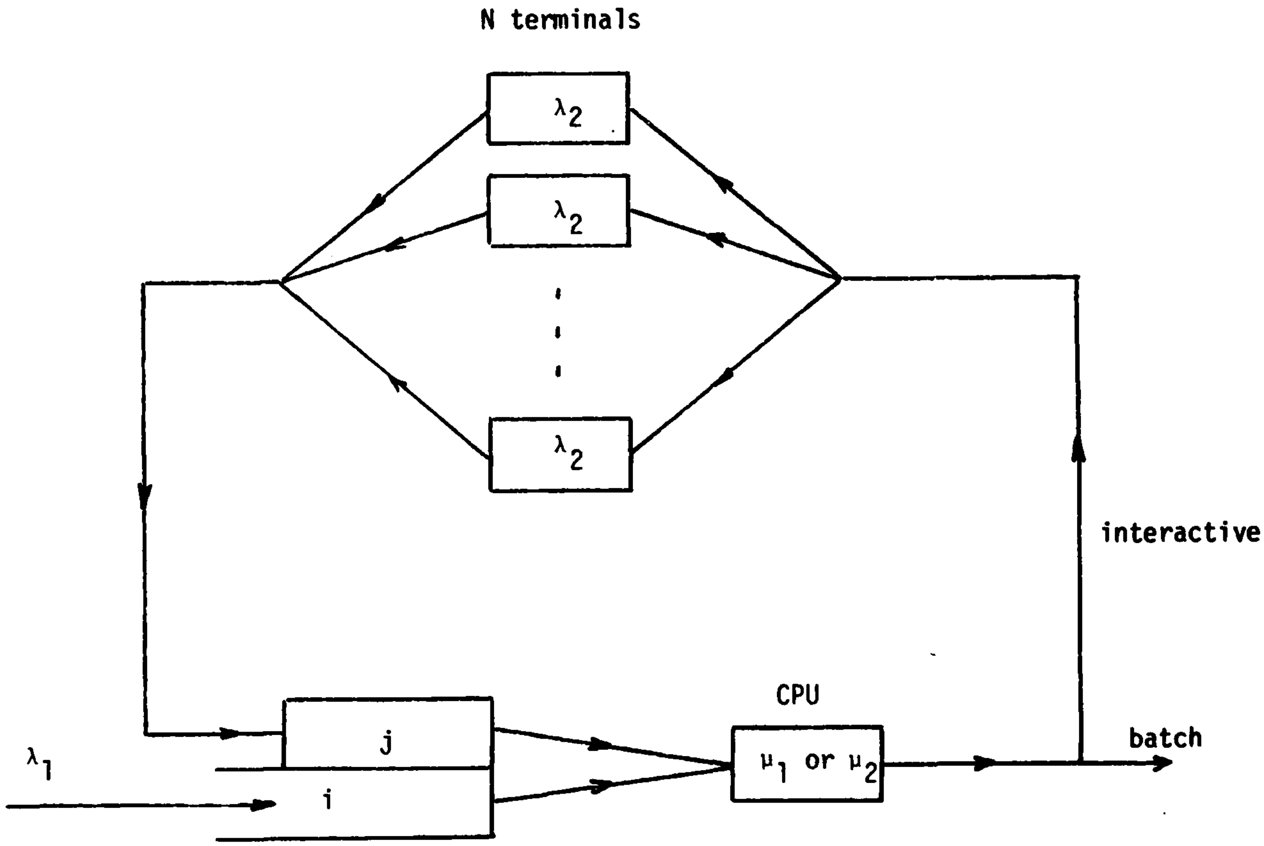
μ_2	2.55	2.50	2.45	2.40	2.35	2.30	2.25	2.20	2.15
ρ	.401	.435	.475	.522	.579	.649	.734	.840	.974
E[j]	3.96	4.21	4.47	4.76	5.08	5.43	5.81	6.23	6.68
.5	3	3	4	4	4	5	5	6	6
.9	9	9	10	10	11	11	12	12	13
.99	14	15	15	16	16	17	17	18	18
.999	18	19	19	20	20	21	21	22	22
E[i]	1.05	1.23	1.47	1.82	2.33	3.18	4.84	9.38	69.03
.5	1	1	1	1	1	2	3	6	48
.9	3	3	4	5	6	8	12	22	160
.99	7	7	9	10	13	17	24	45	320
.999	10	11	13	16	19	25	37	68	480
$E[w_I]$	1.55	1.68	1.83	1.98	2.16	2.36	2.58	2.83	3.11
.5	1.14	1.26	1.40	1.55	1.73	1.93	2.16	2.41	2.70
.9	3.74	3.99	4.26	4.55	4.88	5.23	5.61	6.02	6.47
.99	6.56	6.91	7.28	7.67	8.10	8.55	9.04	9.56	10.12
.999	8.92	9.33	9.77	10.24	10.73	11.26	11.82	12.42	13.06
E[w]	21.63	26.50	33.26	43.10	58.38	84.48	136.88	284.58	2268.57
$E[w_c]$	13.37	14.50	15.83	17.41	19.31	21.62	24.46	28.00	32.48
m_{Ic}	29.00	31.25	33.88	36.98	40.68	45.12	50.54	57.21	65.53
m_0	22.34	25.66	30.14	36.44	45.89	61.52	91.89	175.10	1269.67
m_1	3.98	4.63	5.51	6.75	8.61	11.68	17.66	34.03	249.38
N_{max}	50	49	48	47	46	45	44	44	43

Table III

Percentage in reduction of the mean interactive processing time, the mean interactive sojourn time, and the mean batch sojourn time, for increasing μ_2.

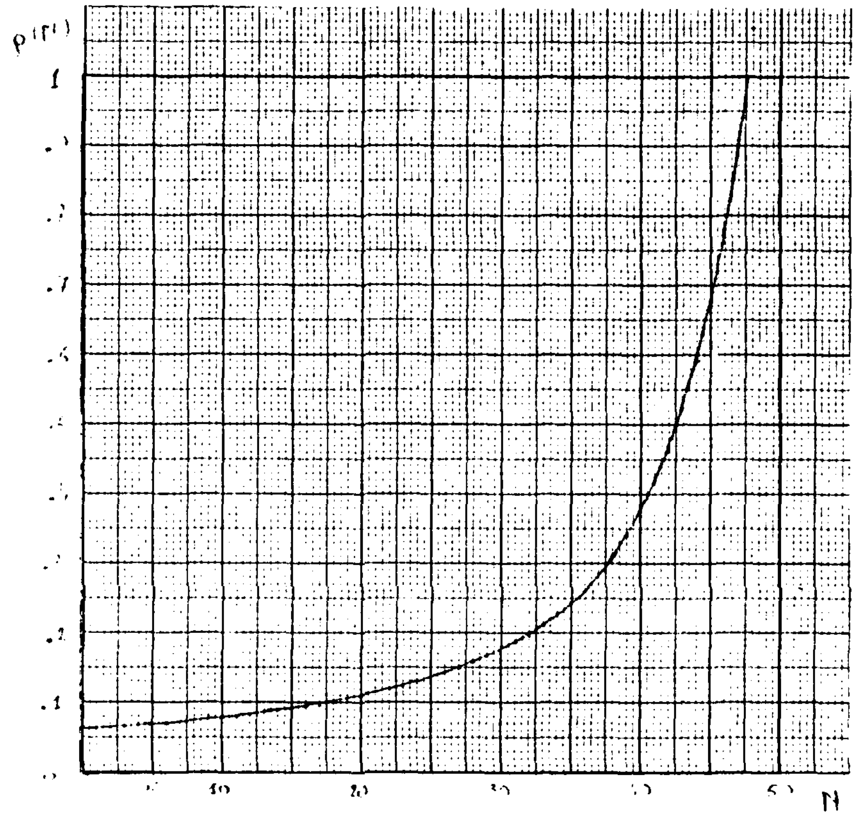
μ_2	2.40 → 2.55	2.15 → 2.30	2.15 → 2.55
mean int. proc. time	- 6%	-6,5%	-16%
mean int. soj. time	-19%	-22%	-46%
mean batch soj. time	-42%	-95%	-98.5%

Figure 1



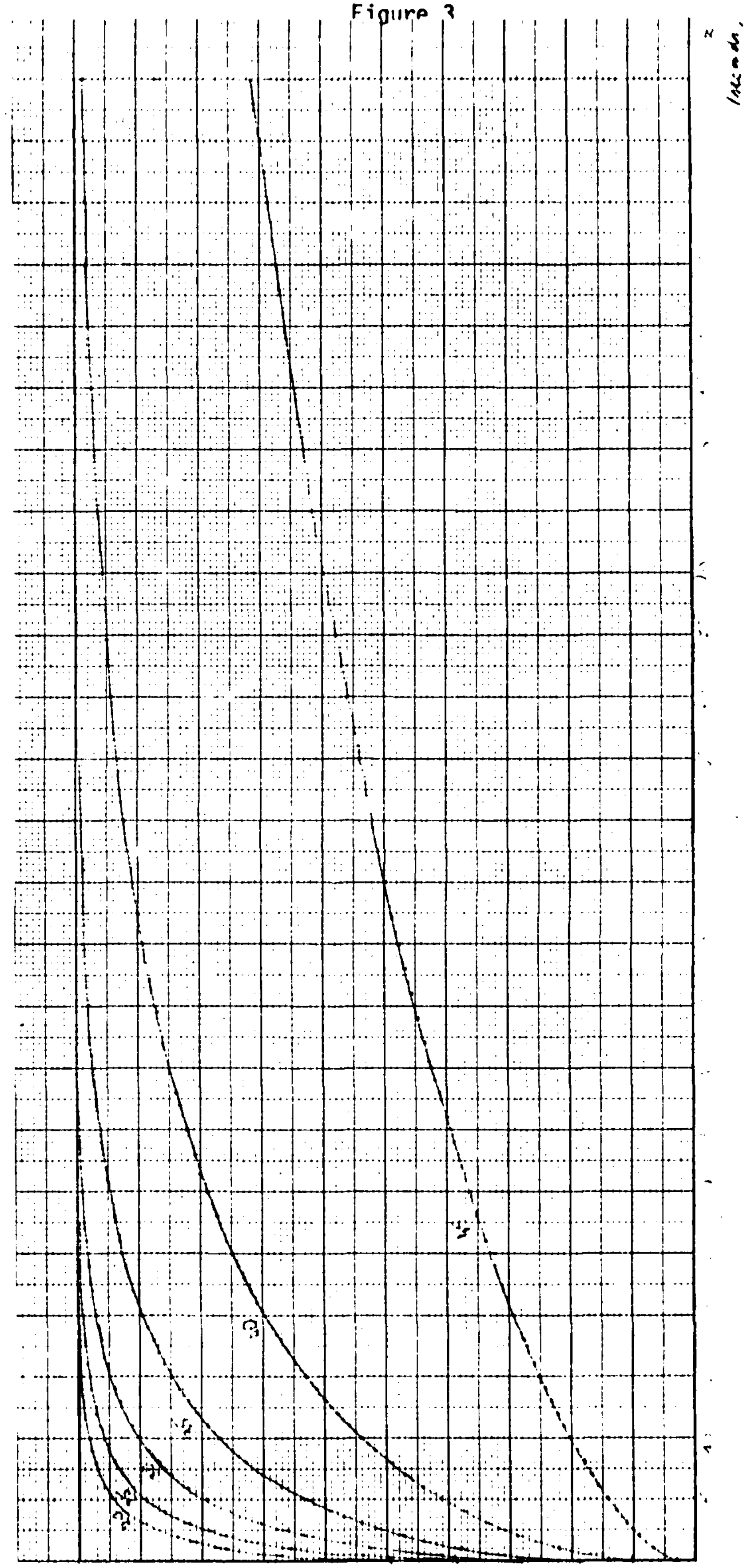
The model.

Figure 2



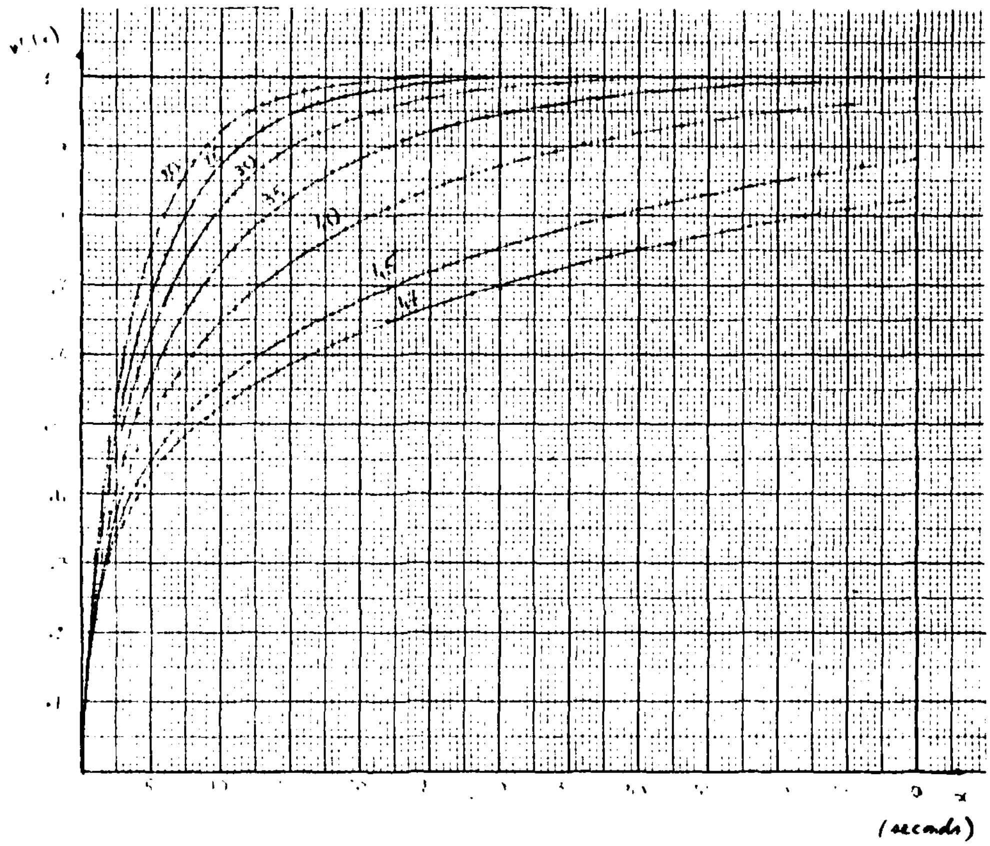
The traffic coefficient $\rho(N)$.

Figure 3



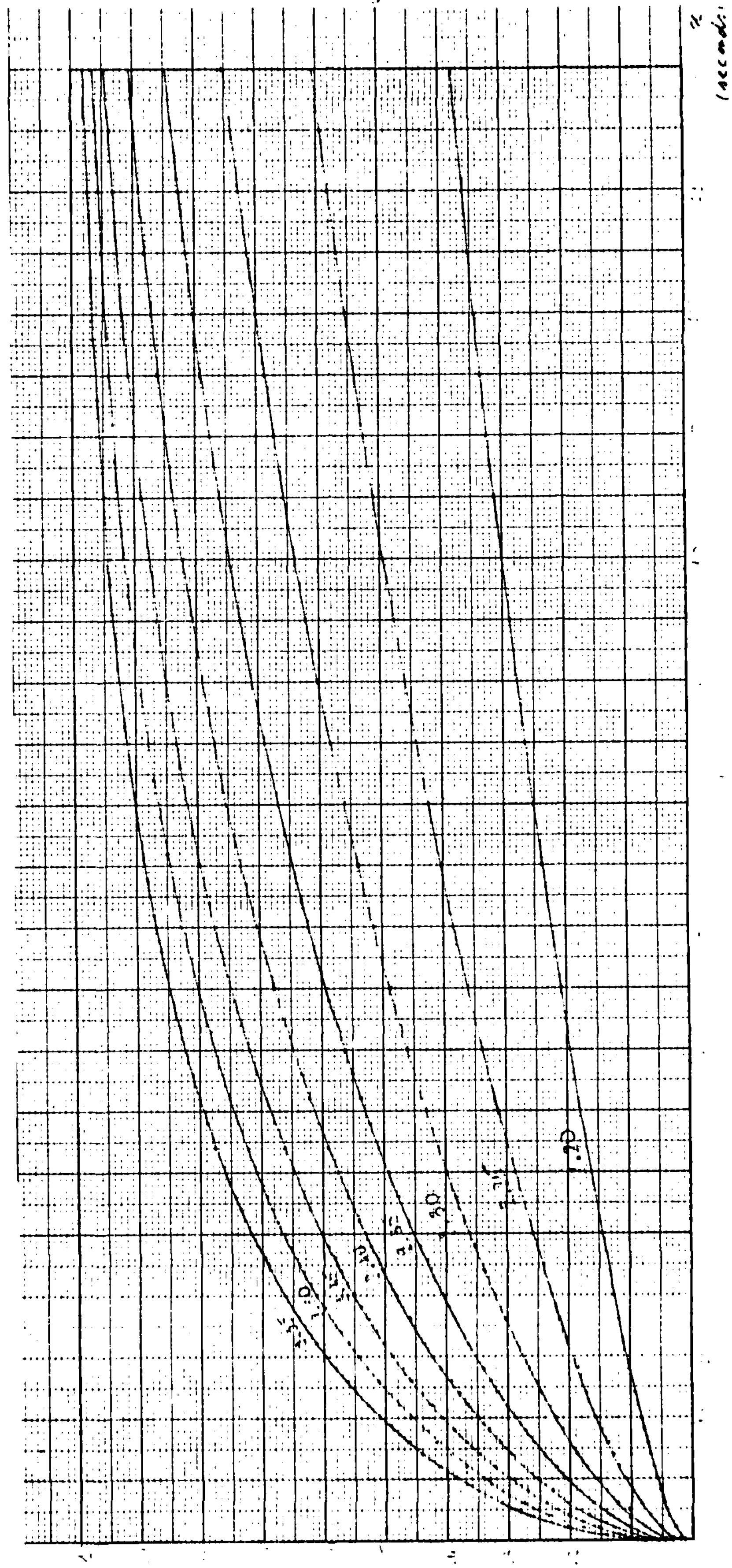
The waiting time distribution for batch jobs. Different values of N.

Figure 4



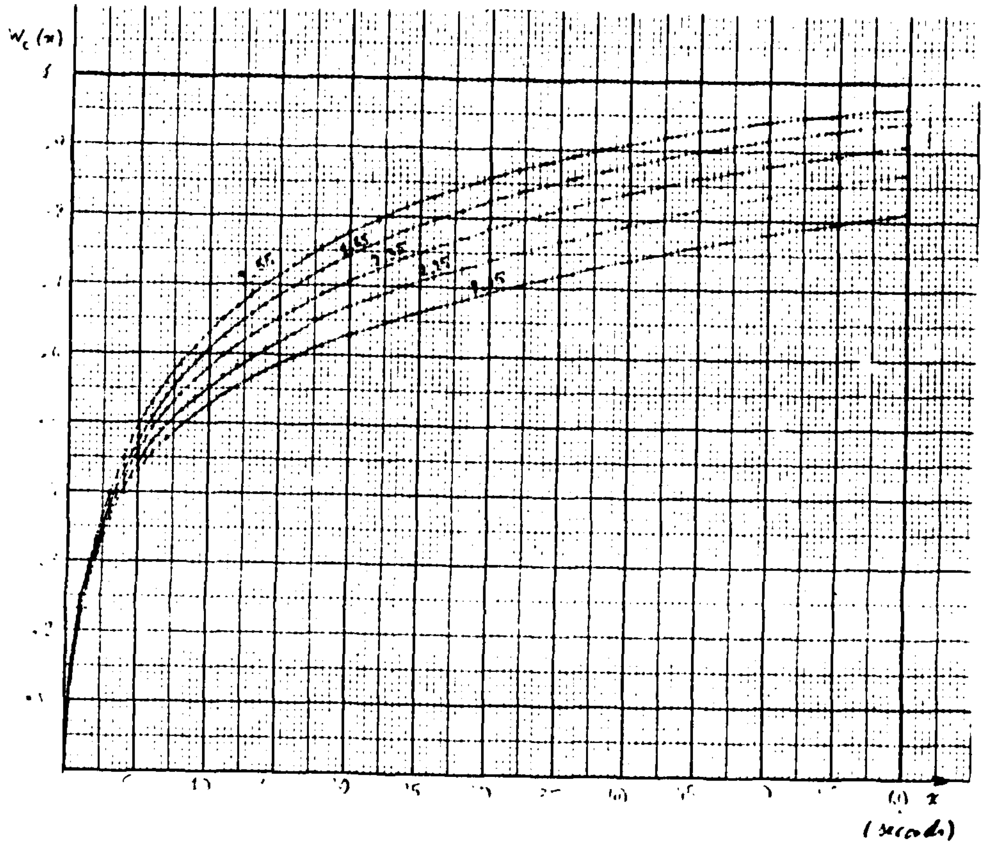
The completion time distribution for batch jobs.
Different values of N.

Figure 5



The waiting time distribution for batch jobs. Different values of μ_2.

Figure 6



The completion time distribution for batch jobs.
Different values of μ_2.

DATE

ILME

-8