stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCH BESLISKUNDE BW 135/81 JANUARI
(DEPARTMENT OF OPERATIONS RESEARCH)

H.C. TIJMS & M.H. VAN HOORN

COMPUTATIONAL METHODS FOR SINGLE-SERVER AND MULTI-SERVER QUEUES WITH MARKOVIAN INPUT AND GENERAL SERVICE TIMES

Preprint

kruislaan 413 1098 SJ amsterdam

Printed at the Mathematical Centre, 413 Kruislaan, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

1980 Mathematics subject classification: 60K25

Computational methods for single-server and multi-server queues with Markovian input and general service times*

by

H.C. Tijms** & M.H. van Hoorn***

ABSTRACT. We first consider a wide class of single-server queues with state dependent Markovian input including the finite capacity M/G/1 queue and the machine repair problem. We specify efficient and stable algorithms to compute the steady-state probabilities and the moments of the waiting time. Next we discuss the multi-server queue with Poisson input and general service times. We present for the steady-state probabilities good quality approximations to be computed from a stable recursive algorithm. As a by-product we obtain simple approximations for the delay probability and the moments of the waiting time. For the output process we derive tractable and good approximations for the moments of the interdeparture time. Also we discuss extensions to the finite capacity M/G/c queue and the machine repair problem with multiple repairmen having general repair times.

KEY WORDS & PHRASES: *Single-server queues, Markovian input, M/G/c queue, multiple-server machine repair problem, computational methods, approximations, state probabilities, waiting times, output process*

---

* This paper will be submitted for publication elsewhere.

** Advisor Mathematical Centre. Address: Vrije Universiteit, Amsterdam.

*** Vrije Universiteit, Amsterdam.

1. *Introduction.*

Until a few years ago the work in queueing analysis primarily concerned analytical results and little work was done on computationally tractable results and approximations. However, in recent years substantial contributions have been made to the field of algorithmic analysis of queues, cf. [3]-[4], [14]-[18] and [21]-[22] amongst others.

This paper presents a stable recursive method to compute the exact values of the steady-state probabilities in a wide class of single server queues with Markovian input and to compute approximate values of the steady-state probabilities in multi-server queues with random and quasirandom input and general service times.

In section 2 we give the algorithmic analysis for a class of single-server queues with state-dependent Markovian input, cf. also [32]. This class covers a number of important single-server queueing models with random and quasi-random input including the finite capacity M/G/1 queue (cf. [18] and [19]) and the machine servicing problem (cf. [15]). We specify stable and efficient algorithms to compute the steady-state probabilities and the waiting time moments in these useful queueing models for the performance analysis of computer systems.

In section 3 we first consider the M/G/c queue. For this multi-server queue exact methods to compute the steady-state probabilities are only available for deterministic service times (cf. [6] and [18]) and for phase-type service times (cf. [9]-[10], [12]-[13], [21] and [29]). However, in particular for phase-type service times the exact methods are only computationally feasible to a limited extent by the dimensionality of the equilibrium state equations for the continuous-time Markov chain representation. In general we may not expect computationally tractable methods can be developed so we have to resort to approximations. For the mean queue size approximations are given in [1], [23] and [30] amongst others. Approximations for the state probabilities are discussed in [8], [11] and [22].

We present for the steady-state probabilities different and improved approximations to be computed by a stable recursive algorithm. These good quality approximations yield as byproduct simple approximations for the delay probability and the moments of the actual waiting time. For the output process for which so far only complicated analytical results were known (cf. [24]) we derive as new result simple and good approximations for the moments of the interdeparture time. Further, for the case of deterministic services times we give somewhat different approximations that in particular result in a very accurate approximation for the delay probability improving the widely used Erlang delay probability approximation. Finally, we obtain for the first time tractable approximations for the finite capacity M/G/c queue and for the machine repair problem with multiple repairmen having general repair times.

2. *Computational methods for a general class of single server queues.*

Consider a single server queueing system where customers singly arrive according to a state-dependent Markovian input process with rate λ_j when j customers are in the system, i.e. the interarrival times are exponentially distributed with mean $1/\lambda_j$ when j customers are present. Each customer joins the system and the service times of the customers are independent random variables having a common probability distribution function F with $F(0) = 0$. To satisfy conditions for statistical equilibrium it is assumed that $\limsup_{n\to\infty}\lambda_n ES < 1$ where S denotes the service time of a customer. The server is never idle when customers are present.

This single server queueing model with state-dependent arrival rate covers a number of important systems including the finite capacity M/G/1 queue having only place for N customers (take $\lambda_j=\lambda$ for $0\le j<N$ and $\lambda_j=0$ for $j\ge N$) and the machine servicing problem with a single repairman and N identical machines having exponential running times with mean $1/\eta$ (take $\lambda_j=(N-j)\eta$ for $0\le j<N$ and $\lambda_j=0$ for $j\ge N$).

We first introduce some notation. Unless stated otherwise it is assumed that the system is empty at epoch 0. Define the following random variables

- T = the next epoch at which the system becomes empty,
- T_n = amount of time during which n customers are in the system in the busy cycle $(0,T]$, $n\ge 0$,
- N = number of customers served in the busy cycle $(0,T]$,
- N_n = number of service completion epochs at which the customer served leaves n other customers behind in the system in the busy cycle $(0,T]$, $n\ge 0$.

Further, define the state probabilities

$p_n = \lim_{t\to\infty} \Pr$ {at time t there are n customers in the system}, $n\geq 0$,

$\pi_n = \lim_{k\to\infty} \Pr$ {the k^{th} customer sees upon arrival n other customers in the system}, $n\geq 0$.

The above limits exist and the limiting distributions are probability distributions, cf. [27]. Note that in general $\pi_n \neq p_n$ except for Poisson input (i.e. $\lambda_j=\lambda$ for all $j\geq 0$). By the theory of regenerative processes (cf. [25] and [27]) and the up- and down crossing result that the long-run fraction of customers seeing upon arrival n other customers in the system equals the long-run fraction of customers leaving upon departure n other customers behind, we have

(2.1) $\quad p_n = ET_n/ET$ and $\pi_n = EN_n/EN$ for $n\geq 0$.

Further, by noting that the queueing system is equivalent to the queueing system in which customers arrive according to a Poisson process with rate $\lambda^* = \max_j \lambda_j$ and an arriving customer does not join the system with probability $1-\lambda_n/\lambda^*$ when n other customers are present and by using the property that Poisson arrivals see time averages (cf. [26]), it readily follows that (see [32] for details)

(2.2) $\quad \pi_n = \lambda_n p_n \Big/ \sum_{j=0}^{\infty} \lambda_j p_j$ for $n\geq 0$.

Also, noting that EN/ET equals the long-run average number of customers served per unit time and hence equals the long-run average number of customers joining the system per unit time, it follows from suitable versions of Little's formula that (cf. [32])

(2.3) $\quad EN/ET = \sum_{j=0}^{\infty} \lambda_j p_j$,

(2.4) $\quad (\sum_{j=0}^{\infty} \lambda_j p_j)ES$ = long-run average number of busy servers = $1 - p_0$.

Note that (2.1)-(2.3) and the first relation in (2.4) also apply to the multi-server case.

To derive a recurrence relation between the probabilities p_n and π_n, we define the following quantity. For $n\geq k\geq 0$, let

(2.5) A_{nk} = expected amount of time during which n customers are in the system until the next service completion epoch given that at epoch 0 a service is completed with k customers left behind in the system.

Then, by partioning the busy cycle (0,T] by means of the service completion epochs and using Wald's theorem (cf. [25]), it follows that

(2.6) $$ET_n = \sum_{k=0}^{n} A_{nk}\, EN_k \text{ for } n\geq 0.$$

Noting that $p_0ET = 1/\lambda_0$ by (2.1) and using the relations (2.1)-(2.3) and (2.6) we find the following recursive scheme

(2.7) $$p_0ET = 1/\lambda_0,\ p_nET = \sum_{k=0}^{n} \lambda_k A_{nk} p_k ET \text{ for } n\geq 1.$$

By the numerically stable algorithm (2.7) we can recursively compute the numbers p_0ET, p_1ET,... once we have evaluated the quantities A_{nk} to be discussed below. Next, using (2.2)-(2.4), the state probabilities p_n and π_n for $n\geq 0$ can be obtained in any desired accuracy by normalization. The above approach based on regenerative analysis is a fertile approach which can be applied to many variants of the M/G/1 queue, cf. [7] and [33]

Alternatively a recursive relation for the state probabilities π_n can be obtained as follows. Note that for $n\geq 1$ the long-run fraction of services having the property that at the completion exactly n customers are left behind equals the long-run fraction of services having the property that at its beginning at most n customers are present and during its execution the number of customers present exceeds the level n (cf. also [5] for a similar up- and downcrossings argument for the virtual waiting time process). Hence, using the second part of (2.1),

(2.8) $$\pi_n = \pi_0 b_{n1} + \sum_{k=1}^{n} \pi_k b_{n-k+1,k} \text{ for } n\geq 1,$$

where b_{nk} is the probability that at least n customers arrive during a service for which at the beginning k customers are present. The approaches (2.7) and (2.8) are of an equal simplicity but the regenerative approach (2.7) seems better suited for such variants as single-server queues with group arrivals, cf. [33].

We next discuss two wide cases in which efficient methods to compute the numbers A_{nk}, $n\geq k\geq 1$ can be given. Note that $A_{n0} = A_{n1}$ for $n\geq 1$.

Case 1. A phase-type service time.

Suppose that the service time distribution function F is a finite mixture of Erlang distribution functions, i.e.

$$(2.9)\qquad F(t) = \sum_{\ell=1}^{r} q_\ell E_{m_\ell,\mu_\ell}(t) \quad \text{with} \quad E_{m,\mu}(t) = 1 - \sum_{j=0}^{m-1} e^{-\mu t}\frac{(\mu t)^j}{j!}$$

Hence we can imagine that with probability q_ℓ the service of a customer requires the completion of m_ℓ independent phases having each an exponential distribution with mean $1/\mu_\ell$. It is known hat each probability distribution function concentrated on $(0,\infty)$ can be approximated with any prescribed accuracy by a phase-type distribution function as (2.9), cf. also [3]. Define $A_{nk}^{(\ell)}(i)$ as the expected amount of time during which n customers are present in a (remaining) service time that starts with k customers present and consists of i independent phases having each an exponential distribution with mean $1/\mu_\ell$. Considering such a remaining service time and using the memoryless property of the exponential distribution and the property that with probability $\mu_\ell/(\mu_\ell+\lambda_k)$ the current phase is completed before a customer arrives, we have for any $n \geq 1$, $1 \leq \ell \leq r$ and $1 \leq i \leq m_\ell$,

$$(2.10)\qquad A_{nk}^{(\ell)}(i) = \begin{cases} (\mu_\ell+\lambda_n)^{-1} + \mu_\ell(\mu_\ell+\lambda_n)^{-1}A_{nn}^{(\ell)}(i-1), & k = n \\ (\mu_\ell+\lambda_k)^{-1}\{\lambda_k A_{n,k+1}^{(\ell)}(i) + \mu_\ell A_{nk}^{(\ell)}(i-1)\}, & 1 \leq k \leq n-1 \end{cases}$$

where $A_{nk}^{(\ell)}(0) = 0$. Hence for any fixed n and ℓ we can recursively compute by a stable algorithm the numbers $A_{nk}^{(\ell)}(i)$ for $i = 1,\ldots,m_\ell$ and $k = n,\ldots,1$. Next we find for any $n \geq 1$

$$A_{nk} = \sum_{\ell=1}^{r} q_\ell A_{nk}^{(\ell)}(m_\ell), \quad 1 \leq k \leq n.$$

Define L_q as the steady-state queue size (excluding any customer in service) and let W_q be the steady-state queueing time of an arbitrary customer (excluding his service time). The moments of L_q follow from the state probabilities p_n. By Little's formula and (2.3)-(2.4), we generally have $EW_q = ESEL_q/(1-p_0)$. For a phase-type service time as (2.9), we can give an efficient algorithm to compute the higher moments of W_q. Therefore, define for $n \geq 1$ and $1 \leq \ell \leq r$,

$p_{ni}^{(\ell)} = \lim_{t\to\infty} \Pr\{$at time t there are n customers present and the residual service time consists of i independent exponential phases having each mean $1/\mu_\ell\}$, $1 \le i \le m_\ell$,

$\pi_{ni}^{(\ell)} = \lim_{k\to\infty} \Pr\{$at the arrival epoch of the k-th customer there are n other customers present and the residual service time consists of i independent exponential phases having each mean $1/\mu_\ell\}$, $1 \le i \le m_\ell$.

By the same arguments as used to prove (2.2), we have

$$\pi_{ni}^{(\ell)} = \lambda_n p_{ni}^{(\ell)} / \sum_{j=0}^{\infty} \lambda_j p_j = \lambda_n ES p_{ni}^{(\ell)}/(1-p_0) \quad \text{for all } n,\ell,i,$$

The moments of W_q follow from the probabilities $\pi_{ni}^{(\ell)}$, e.g.

$$EW_q^2 = \sum_{n,\ell,i} \pi_{ni}^{(\ell)} [\frac{i(i+1)}{\mu_\ell^2} + (n-1)\{\frac{2i}{\mu_\ell} ES + ES^2 + (n-2)(ES)^2\}],$$

where we assume service in order of arrival. To compute the $p_{ni}^{(\ell)}$, put for abbreviation

$$p_{0i}^{(\ell)} = q_\ell \delta(i,m_\ell) p_0 \qquad \text{where} \qquad \delta(i,j) = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{if } j \ne i \end{cases}$$

and define $X_1(t)$ = the number of customers present at time t and let $X_2(t) = (\ell,i)$ if at time t a service is in progress and has still i uncompleted phases having each mean $1/\mu_\ell$ and $X_2(t) = (0,0)$ otherwise. For the continuous-time Markov chain $(X_1(t),X_2(t))$ we have for any $n \ge 1$ and $1 \le \ell \le r$,

$$(2.11) \quad (\lambda_n + \mu_\ell) p_{ni}^{(\ell)} = \lambda_{n-1} p_{n-1,i}^{(\ell)} + \mu_\ell p_{n,i+1}^{(\ell)} + q_\ell \delta(i,m_\ell) \lambda_n p_n, \quad 1 \le i \le m_\ell,$$

where $p_{ni}^{(\ell)} = 0$ for $i = m_\ell + 1$. This relation is obtained by equating the rate at which the system leaves the microstate (n,ℓ,i) to the rate at which the system enters this state and inserting into this equation the local balance relation $\lambda_n p_n = \Sigma_h \mu_h p_{n+1,1}^{(h)}$. This latter relation in its turn follows by aggregating the microstates 0, ..., n into a macrostate and equating the rate at which the system leaves this macrostate to the rate at which the system enters this macrostate. For fixed $1 \le \ell \le r$ the state probabilities $p_{ni}^{(\ell)}$ can be computed by the stable recursive scheme (2.11) for $i = m_\ell, \ldots, 1$ and $n = 1,2,\ldots$ It is important to note that by first computing the probabilities p_n from (2.7) we can compute the $p_{ni}^{(\ell)}$ from a stable and efficient recursive scheme instead of by solving the difficult system of linear equilibrium equations.

REMARK 2.1 For the special case where F is an Erlang distribution function $E_{r,\mu}$, the state probabilities p_n and the moments of W_q can be computed by another algorithm that is simpler than the above one. Therefore note that the number of uncompleted phases in the system uniquely determines the number of customers present. Defining f_j as the steady-state probability that at an arbitrary epoch there are j uncompleted phases in the system, we have

$$(2.11)\quad p_0 = f_0 \text{ and } p_n = \sum_{j=(n-1)r+1}^{nr} f_j \text{ for } n\geq 1$$

Put for abbreviation $\gamma_j=\lambda_{[j/r]}$ if j/r is an integer and $\gamma_j=\lambda_{[j/r]+1}$ otherwise, where [x] is the largest integer less than or equal to x. Then

$$(2.12)\quad \mu f_n = \sum_{k=n-r}^{n-1} \gamma_k f_k \text{ for } n\geq 1$$

where $f_j=0$ for $j<0$. This recursive relation follows by aggregating the microstates n,n+1,.. of the continuous-time Markov chain describing the number of uncompleted phases into a macrostate and equating the rate at which the system leaves this macrostate to the rate at which the system enters this macrostate. Finally, noting that $\gamma_k f_k/\Sigma_j \lambda_j p_j = \gamma_k f_k ES/(1-p_0)$ is the steady-state probability that an arriving customer finds k uncompleted phases present, the moments of W_q follows from the state probabilities f_j.

Case 2. The finite capacity M/G/1 queue and the machine repair problem.

For a general service time distribution function we can easily evaluate by numerical integration the numbers A_{nk} for both the finite capacity M/G/1 queue and the machine servicing problem. We only discuss here the latter model. Consider N identical machines with a single repairman where the running times of the machines are independent and exponentially distributed with mean $1/\eta$ and the repair time has the general distribution function F. This problem is a special case of the single-server model having state-dependent input with $\lambda_j=(N-j)\eta$ for $j<N$ and $\lambda_j=0$ for $j\geq N$ by taking the number of inoperative machines as state for the system. We have for $1\leq k\leq n\leq N$,

Table 2.1 The machine repair problem with a single repairman

$1/\rho$	c_S^2	N=5 ER	N=5 c_R	N=10 ER	N=10 c_R	N=20 ER	N=20 c_R	N=40 ER	N=40 c_R
5	1/10	1.70	0.47	5.03	0.34	15.0	0.13	35.0	0.07
	1/3	1.79	0.54	5.07	0.41	15.0	0.19	35.0	0.11
	1/2	1.84	0.59	5.09	0.45	15.0	0.22	35.0	0.13
	1/2*	1.85	0.58	5.10	0.45	15.0	0.22	35.0	0.13
	1	1.99	0.71	5.19	0.55	15.0	0.29	35.0	0.18
	2	2.17	0.96	5.31	0.73	15.0	0.40	35.0	0.25
	2*	2.13	1.02	5.24	0.75	15.0	0.40	35.0	0.25
	5	2.49	1.46	5.57	1.09	15.0	0.63	35.0	0.39
	25	2.97	3.14	6.06	2.37	15.0	1.43	35.0	0.88
	100	3.16	6.21	6.27	4.72	15.0	2.89	35.0	1.78
10	1/10	1.28	0.39	2.25	0.55	10.0	0.25	30.0	0.10
	1/3	1.33	0.48	2.39	0.61	10.0	0.31	30.0	0.13
	1/2	1.37	0.53	2.48	0.66	10.0	0.34	30.0	0.16
	1/2*	1.37	0.52	2.49	0.65	10.0	0.34	30.0	0.16
	1	1.47	0.68	2.73	0.76	10.0	0.43	30.0	0.21
	2	1.62	0.97	3.06	0.96	10.1	0.57	30.0	0.29
	2*	1.60	1.02	3.00	1.01	10.1	0.57	30.0	0.29
	5	1.93	1.54	3.68	1.37	10.3	0.85	30.0	0.45
	25	2.60	3.34	4.90	2.73	11.1	1.84	30.0	1.01
	100	2.94	6.51	5.52	5.25	11.6	3.67	30.0	2.06
20	1/10	1.13	0.29	1.38	0.46	3.02	0.61	20.0	0.18
	1/3	1.15	0.37	1.45	0.55	3.24	0.66	20.0	0.22
	1/2	1.17	0.42	1.50	0.60	3.39	0.69	20.0	0.25
	1/2*	1.17	0.41	1.50	0.59	3.40	0.69	20.0	0.25
	1	1.22	0.55	1.64	0.75	3.78	0.77	20.0	0.31
	2	1.31	0.85	1.88	1.04	4.33	0.93	20.0	0.41
	2*	1.30	0.90	1.86	1.09	4.25	0.97	20.0	0.42
	5	1.53	1.49	2.41	1.57	5.41	1.25	20.1	0.63
	25	2.21	3.47	3.89	3.04	7.98	2.28	20.8	1.37
	100	2.73	6.72	4.89	5.58	9.78	4.18	21.8	2.75
40	1/10	1.06	0.21	1.15	0.32	1.45	0.50	4.11	0.65
	1/3	1.07	0.27	1.18	0.40	1.54	0.59	4.45	0.69
	1/2	1.08	0.31	1.20	0.46	1.60	0.65	4.67	0.72
	1/2*	1.08	0.30	1.20	0.45	1.60	0.64	4.68	0.71
	1	1.10	0.42	1.27	0.60	1.78	0.79	5.26	0.78
	2	1.15	0.68	1.39	0.92	2.10	1.08	6.12	0.90
	2*	1.15	0.72	1.39	0.98	2.08	1.14	6.04	0.93
	5	1.28	1.31	1.71	1.57	2.88	1.56	7.90	1.14
	25	1.83	3.50	2.95	3.34	5.56	2.72	12.7	1.89
	100	2.47	6.89	4.35	5.91	8.31	4.56	17.1	3.26

$$(2.13)\quad A_{nk} = \int_0^\infty (1-F(t)) \binom{N-k}{n-k} (1-e^{-\eta t})^{n-k} e^{-\eta t(N-n)} dt,$$

as follows by noting that $A_{nk} = {}_0\int^\infty E\chi_t dt$ where $\chi_t=1$ if at time t the first repair is still in progress and n machines are broken down given that at epoch 0 a repair starts with k machines broken down, and $\chi_t=0$ otherwise.

We conclude this section by presenting some numerical results for the above machine repair problem. For the repair time S we consider the following cases with a phase-type distribution (2.9),

(i) $r=1$, $m_1=k$ $(c_S^2=1/k)$ for $k=1,2,3$ and 10,

(ii) $r=2$, $m_1=m_2=1$, $q_1=(1/2)[1+\{(c_S^2-1)/(c_S^2+1)\}^{1/2}]$, $q_1/\mu_1=q_2/\mu_2$ for $c_S^2=2$, 5, 25 and 100,

(iii) $r=2$, $m_1=1$, $m_2=3$, $q_1=(4-\surd 7)/6$, $\mu_1=\mu_2$ $(c_S^2=1/2)$,

(iv) $r=2$, $m_1=1$, $m_2=3$, $q_1=(19-\surd 145)/36$, $q_1/\mu_1=3q_2/\mu_2$ $(c_S^2=2)$,

where $c_S = \{ES^2/(ES)^2-1\}^{1/2}$ denotes the coefficient of variation of S. The latter two cases are denoted by * in table 2.1. In alle cases we take ES=1. For various values of $\rho = \eta ES$ and N we give in table 2.1 the mean ER and the coefficient of variation c_R of the response time $R = W_q + S$. Note that, by Little's formula $\lambda'(ER+1/\eta) = N$ with throughput $\lambda' = (1-p_0)/ES$, the server utilization $1-p_0$ determines ER.

3. *Approximations for the* M/G/c *queue and the machine repair problem with multiple repairmen.*

For clarity of presentation we first consider the infinite capacity M/G/c queue with $c>1$ servers where customers arrive in accordance with a Poisson process with rate λ and the service time S of a customer has a general probability distribution function F with $F(0)=0$. It is assumed that $\rho=\lambda ES/c<1$. An arriving customer joins the queue if he finds all c servers occupied or else he is served immediately by one of the free servers. A server will never remain idle if customers are waiting in the queue.

We first introduce some notation. Define the random variables T, T_n, N and N_n and the steady-state probabilities p_n and π_n as in section 2. Then (cf. (2.1)-(2.3)),

$$(3.1)\quad p_n = ET_n/ET,\ \pi_n = EN_n/EN \text{ for } n\geq 0,$$

$$(3.2)\quad p_n = \pi_n \text{ for } n\geq 0,\ EN/ET = \lambda.$$

Further, define the delay probability P_W, the mean queue size EL_q and the constant Ω by

$$(3.3)\qquad P_W = \sum_{n=c}^{\infty} p_n, \quad EL_q = \sum_{n=c}^{\infty} (n-c)p_n, \quad \Omega = \left\{ \sum_{k=0}^{c-1} \frac{(\lambda ES)^k}{k!} + \frac{(\lambda ES)^c}{c!(1-\rho)} \right\}^{-1}.$$

We write $p_n = p_n(\exp)$, $P_W = P_W(\exp)$ and $EL_q = EL_q(\exp)$ when the service time is exponentially distributed and we have the explicit results

$$(3.4)\qquad p_n(\exp) = \frac{(\lambda ES)^n}{n!}\Omega \text{ for } 0 \le n \le c-1, \quad p_n(\exp) = \frac{(\lambda ES)^n}{c!c^{n-c}}\Omega \text{ for } n \ge c,$$

$$(3.5)\qquad P_W(\exp) = \frac{(\lambda ES)^c}{c!(1-\rho)}\Omega, \quad EL_q(\exp) = \frac{(\lambda ES)^c \rho}{c!(1-\rho)^2}\Omega.$$

The quantity $P_W(\exp)$ is called the Erlang delay probability and is known to be a good approximation for P_W when the service time has a general distribution.

In general no explicit expression for p_n can be given. However, we can try to set up a recursive scheme as (2.7) to compute the state probabilities. In doing so, we encounter the difficulty that for a service completion epoch at which $j \ge 1$ customers are left behind the distribution function of the time until the next service completion epoch depends on the information of how long the remaining services are already in progress. To overcome this difficulty, we make an approximation by aggregating this required information and using distribution functions depending only the number of customers left behind. The specification of these distribution functions will determine our approximations. More precisely, we make the following approximation assumption.

APPROXIMATION ASSUMPTION. *For any* $1 \le j \le c-1$, *the random variables defined as the smallest of the remaining service times of the* j *services in progress at those service completion epochs at which* j *customers are left behind in the system are independent and have common probability distribution function* F_j^*. *For the service completion epochs at which* $j \ge c$ *customers are left behind in the system, the times until the next service completion epoch are independent random variables with common probability distribution function* F^* *where* F^* *is the same for all* $j \ge c$.

Define the quantities A_{nk} for $n \ge k \ge 0$ as in (2.5) with for $k \ge 1$ the stipulation that the approximation assumption applies. Under this assumption, we have

$$(3.6)\qquad ET_n \simeq \sum_{j=0}^{n} EN_j A_{nj} \text{ for } n = 0,1,\ldots,$$

and so, by (3.1)-(3.2),

$$(3.7)\qquad p_n ET \simeq \sum_{j=0}^{n} \lambda p_j ETA_{nj} \text{ for } n = 0,1,\ldots.$$

Together (3.7) and the relation $p_0 ET = 1/\lambda$ suggest to define $\{q_n, n\geq 0\}$ by

$$(3.8)\qquad q_0 = 1/\lambda \text{ and } q_n = \sum_{j=0}^{n} \lambda q_j A_{nj} \text{ for } n = 1,2,\ldots$$

from which stable recursive scheme we can successively compute the numbers $q_0, q_1, \ldots$ once we have evaluated the quantities A_{nk}. Next the state probabilities p_n, $n\geq 0$ can be approximated by

$$(3.9)\qquad p_n(\text{appr}) = q_n / \sum_{j=0}^{\infty} q_j \text{ for } n = 0,1,\ldots .$$

To evaluate the quantities A_{nk} we make in the approximation assumption the following *specification*

$$(3.10)\qquad 1-F_j^*(t) = (1-F_e(t))^j,\ 1\leq j\leq c \text{ and } F^*(t) = F(ct)$$

where F_e is the equilibrium distribution of F and is given by

$$(3.11)\qquad F_e(t) = (1/ES)\int_0^t (1-F(x)dx,\ t\geq 0.$$

To motivate this specification, note that if not all c servers are busy the M/G/c queue can be treated as an M/G/∞ queue for which we have the renewal-theoretic result that at an arbitrary epoch the remaining service times of services in progress (if any) are independent random variables with common probability distribution function F_e, cf. p. 161 in [28]. If all c c servers are busy we can treat the M/G/c queue with service time S as an M/G/1 queue with service time S/c, cf. also [22]. Note that the approximation assumption is satisfied for the M/M/c queue. In evaluating the quantities A_{nk} we next encounter the computational difficulty that, except for deterministic service times, the closed-form expression for A_{nk} with k<c involves an (n-k+1)-dimensional integral because of the phenomenon

that a newly started service may be completed before services in progress. Fortunately, by the specified form of F_j^* for $1 \le j \le c-1$, we can establish by induction a very simple expression for q_j for $j \le c-1$ through which we succeeded in eliminating the multi-dimensional integrals so that the ultimate recursive scheme involves only one-dimensional integrals. The following results have been proved in [31] (see the appendix for a simplified and more generally applicable proof).

THEOREM 3.1. Under the approximation assumption with specification (3.10),

$$(3.12) \quad p_n(\text{appr}) = \frac{(\lambda ES)^n}{n!} p_0(\text{appr}), \quad 0 \le n \le c-1,$$

$$(3.13) \quad p_n(\text{appr}) = \lambda p_{c-1}(\text{appr})\, \alpha_{n-c} + \lambda \sum_{j=c}^{n} p_j(\text{appr})\, \beta_{n-j}, \quad n \ge c,$$

with $p_0(\text{appr}) = \Omega$ and

$$(3.14) \quad \alpha_k = \int_0^\infty (1-F_e(t))^{c-1} (1-F(t)) e^{-\lambda t} \frac{(\lambda t)^k}{k!}, \quad k \ge 0$$

$$(3.15) \quad \beta_k = \int_0^\infty (1-F(ct)) e^{-\lambda t} \frac{(\lambda t)^k}{k!}, \quad k \ge 0.$$

Hence we can compute the approximations for the state probabilities by a stable recursive scheme where in general any recursion step requires the evaluation of two one-dimensional integrals which are easy to handle by numerical integration. Note that $p_j(\text{appr}) = p_j(\text{exp})$ for $0 \le j \le c-1$ and hence

$$(3.16) \quad P_W(\text{appr}) = P_W(\text{exp}),$$

so that as approximation for the delay probability we find the widely used Erlang delay probability which is in general a good approximation. We note that the approximations given in [11] for the state probabilities p_j are also equal to $p_j(\text{exp})$ for $j \le c-1$ but differ from our approximations for $j > c$. In [11] the approximations are given in a form inconvenient for computational purposes, however it was shown in [31] that these approximations can also be computed by a stable recursive scheme which is obtained from (3.12)-(3.13) by replacing $(1-F_e(t))^{c-1}(1-F(t))$ by $1-F(ct)$ in the integral in (3.14). We note that the latter approximations yield for the mean queue size the same approximation as found in [23].

Denote by L_q the queue size at an arbitrary epoch in the steady-state (excluding the customers in service, if any). Using generating functions we obtain after some algebra from (3.12)-(3.13),

$$(3.17)\quad EL_q(\text{appr}) = \tfrac{1}{2}(1+c_S^2)EL_q(\exp)\{1+(1-\rho)(2c\gamma_1 ES/ES^2-1)\}.$$

$$(3.18)\quad EL_q^2(\text{appr}) = \lambda P_W(\text{appr})\left[\{\gamma_1 + \frac{\lambda ES^2}{2c^2(1-\rho)}\}\{1+\frac{\lambda^2 ES^2}{c^2(1-\rho)}\} + \lambda\gamma_2 + \frac{\lambda^2 ES^3}{3c^3(1-\rho)}\right],$$

where $c_S = \{ES^2/(ES)^2 - 1\}^{\frac{1}{2}}$ denotes the coefficient of variation of the service time S and γ_k is defined by

$$(3.19)\quad \gamma_k = k\int_0^\infty t^{k-1}(1-F_e(t))^c dt,\ k\geq 1.$$

Note that when F is a NBUE-distribution function we have $ES/(c+1)\leq\gamma_1\leq ES/c$ by $1-F_e(t)\leq 1-F(t)\leq 1$. Similarly, the higher moments of L_q may be derived. From the moments of the queue size L_q we get the moments of the steady-state queueing time W_q of an arbitrary customer (excluding his service time). Under the assumption of service in order of arrival we have (see [20]),

$$(3.20)\quad EL_q(L_q-1)\ldots(L_q-k+1) = \lambda^k EW_q^k \text{ for } k\geq 1.$$

We note that the distribution function of W_q may be approximated by matching of moments, e.g. following [17] we may approximate the waiting time distribution $1-\Pr\{W_q>t|W_q>0\}$ for the delayed customers by a Weibull distribution function $1-\exp((at)^b)$ by matching the first two moments.

Next we derive for the output process approximations for the moments of the interdeparture time T_D between two consecutive service completion epochs in the steady-state.

THEOREM 3.2. Under (3.10),

$$(3.21)\quad ET_D^m(\text{appr}) = \frac{m!}{\lambda^m}\left[1-P_W(\text{appr})\{\rho-\frac{\lambda^m ES^m}{m!c^m} - (1-\rho)\sum_{i=1}^{m-1}\frac{\lambda^i}{i!}\gamma_i\}\right],\ m\geq 1,$$

where γ_i, $i\geq 1$ is given by (3.19). In particular, $ET_D(\text{appr}) = 1/\lambda$ is exact and $ET_D^2(\text{appr}) = (2/\lambda^2)\{1 - \rho P_W(\text{appr}) + (1-\rho)EL_q(\text{appr})\}$.

PROOF. Under the condition that the system is empty at epoch 0, define $M_k(t)$ as the probability that the service completions of the first k customers all occur beyond time t, $1 \le k \le c$. Also, define Q(t) as the steady-state probability that the time between two consecutive service completion epochs exceeds t. The steady-state probability that at a service completion epoch there are left i customers behind equals $\pi_i = p_i$ (cf. (3.1)-(3.2)) and so, under the approximation assumption with (3.10),

$$(3.22)\quad Q(t) = \sum_{i=0}^{c-1} p_i(\text{appr})(1-F_e(t))^i M_{c-i}(t) + (1-F(ct)) \sum_{i=c}^{\infty} p_i(\text{appr}),\ t>0,$$

By considering what may happen in $(0,\Delta t)$ for Δt small, it follows that for $1 \le k \le c-1$,

$$M_k(t+\Delta t) = (1-\lambda\Delta t)M_k(t) + \lambda\Delta t(1-F(t))M_{k-1}(t) + 0(\Delta t),\ t>0$$

and so, for $1 \le k \le c-1$,

$$(3.23)\quad \frac{dM_k(t)}{dt} = \lambda(1-F(t))M_{k-1}(t) - \lambda M_k(t),\ t>0,$$

where $M_0(t)=1$ for all t. Put for abbreviation

$$(3.24)\quad Q_1(t) = \sum_{i=0}^{c-1} p_i(\text{appr})(1-F_e(t))^i M_{c-i}(t),\ t>0.$$

Using (3.23) and the relations

$$(3.25)\quad \frac{d}{dt}(1-F_e(t))^{i+1} = -\frac{(i+1)}{ES}(1-F(t))(1-F_e(t))^i,\ i \ge 0,$$

$$(3.26)\quad p_{i+1}(\text{appr}) = \frac{\lambda ES}{i+1} p_i(\text{appr}),\ 0 \le i \le c-2,$$

we find after some algebra

$$\frac{dQ_1(t)}{dt} = -\lambda Q_1(t) - \frac{\lambda ES}{c} p_{c-1}(\text{appr}) \frac{d}{dt}(1-F_e(t))^c,\ t>0.$$

From this first-order differential equation, we get

$$(3.27)\quad Q_1(t) = \sum_{i=0}^{c-1} p_i(\text{appr})e^{-\lambda t} - \rho p_{c-1}(\text{appr}) \int_0^t e^{-\lambda(t-u)} d(1-F_e(u))^c,\ t>0.$$

Using the relation (cf.(3.12) and (3.4)-(3.5))

$$(3.28)\quad \rho p_{c-1}(\text{appr}) = (1-\rho)P_W(\text{appr})$$

and the relation $EX^k = k\int_0^\infty x^{k-1}\Pr\{X>x\}dx$, $k\geq 1$ for any nonnegative random variable X, we obtain (3.21) from (3.22), (3.24) and (3.27)-(3.28) after some algebra.

Denoting by c_D the coefficient of variation of the interdeparture time T_D, we have by Theorem 3.2 that $c_D^2(\text{appr})=1-2\rho P_W(\text{appr})+2(1-\rho)EL_q(\text{appr})$. In [26] it was empirically established that

$$(3.29)\quad L_q(\exp) \simeq (1-\rho)^{-1}\rho^{\sqrt{2(c+1)}} \text{ for } \rho \text{ close to } 1,$$

and so by (3.17) and $P_W(\exp) = (1-\rho)\rho^{-1}EL_q(\exp)$,

$$(3.30)\quad c_D^2(\text{appr}) \simeq 1-\rho^{\sqrt{2(c+1)}} + c_S^2\rho^{\sqrt{2(c+1)}} \text{ for } \rho \text{ close to } 1.$$

The above approximations are good quality approximations. This is supported by the findings that $P_W(\text{appr})$ is given by the good quality Erlang delay probability approximation and that $EL_q(\text{appr})$ is competitive to the extremely accurate approximations for the mean queue size specially developed in [1] and [30], cf. [31] for extensive numerical comparisons. Further, by [2] and [16], the following light- and heavy-traffic approximations agree with exact results,

$$\lim_{\rho\to 0}\frac{EL_q(\text{appr})}{EL_q(\exp)} = \frac{c\gamma_1}{ES} \text{ and } \lim_{\rho\to 1}\frac{EL_q(\text{appr})}{EL_q(\exp)} = \frac{ES^2}{2(ES)^2}.$$

The finite capacity M/G/c *queue.*

Consider the finite capacity M/G/c queueing system having only place for $N\geq c$ customers, i.e. any customer who finds upon arrival N other customers present does not enter and has no effect on the system. No restriction is imposed on $\rho=\lambda ES/c$. The above analysis and results require only obvious modifications. Define now π_n as the steady-state probability that an *entering* customer finds upon arrival n other customers present, $0\leq n\leq N-1$. We now have (cf. (2.1)-(2.3) with $\lambda_j=\lambda$ for $j\leq N-1$ and $\lambda_j=0$ for $j\geq N$),

$$(3.31)\quad p_n = \frac{ET_n}{ET},\ \pi_n = \frac{EN_n}{EN},\ \pi_n = \frac{p_n}{1-p_N} \text{ for } n\geq 0 \text{ and } \frac{EN}{ET} = \lambda(1-p_N).$$

Next a minor modification of the proof of Theorem 3.1 shows that under the approximation assumption with specification (3.10) the approximation $p_n(\text{appr})$ for the finite capacity M/G/c queue is given by (3.12) for $0\le n\le c-1$ and by (3.13) for $c\le n\le N-1$ whereas

$$(3.32)\quad p_N(\text{appr}) = \lambda p_{c-1}(\text{appr}) \sum_{k=N}^{\infty} \alpha_{k-c} + \lambda \sum_{j=c}^{N-1} p_j(\text{appr}) \sum_{k=N}^{\infty} \beta_{k-j}$$

where α_k, β_k are defined by (3.14)-(3.15). The relation (3.32) can be after some algebra reduced to

$$(3.33)\quad p_N(\text{appr}) = \rho p_{c-1}(\text{appr}) - (1-\rho) \sum_{k=c}^{N-1} p_k(\text{appr}).$$

By (3.31) and (3.33) we have

$$(3.34)\quad \rho\pi_{c-1}(\text{appr}) = (1-\rho)\Pi_W(\text{appr}) + (1-p_N(\text{appr}))^{-1} p_N(\text{appr}),$$

where $\Pi_W = \Sigma_{k=c}^{N-1} \pi_k$ denotes the steady-state probability that an entering customer will be delayed. Noting that π_i gives the steady-state probability that at a service completion epoch i customers are left behind and using (3.34), an examination of the proof of Theorem 3.2 shows that the corresponding approximations for the moments of the interdeparture time T_D between two consecutive departures are given by

$$(3.35)\quad ET_D^m(\text{appr}) = \frac{m!}{\lambda^m} [(1-p_N(\text{appr}))^{-1} - \Pi_W(\text{appr})(\rho - \frac{\lambda^m ES^m}{m!c^m} - (1-\rho)\sum_{i=1}^{m-1} \frac{\lambda^i}{i!}\gamma_i) + (1-p_N(\text{appr}))^{-1} p_N(\text{appr}) \sum_{i=1}^{m-1} \frac{\lambda^i}{i!}\gamma_i],\ m\ge 1.$$

Note that for the special case of no waiting room (i.e. N=c) we have the remarkable result $p_n(\text{appr})=C^{-1}(\lambda ES)^n/n!$, $0\le n\le c$ with $C=\Sigma_{k=0}^{c}(\lambda ES)^k/k!$, that is the approximations are equal to the exact values having the famous insensitivity property of depending on the service time only through the first moment.

The machine repair problem with multiple repairmen.

Consider the machine repair problem with N identical machines and c repairmen where $1<c\le N$. The machines have independent running times with

a common exponential distribution with mean $1/\eta$ and the repair time S of a broken-down machine has a general probability distribution function F. Define the state of the system as the number of machines broken down and let $\lambda_j=(N-j)\eta$ for $0\le j\le N$. Under the approximation assumption with specification (3.10), a generalisation of the proof of Theorem 3.1 as given in the appendix yields

$$p_n(\mathrm{appr}) = \binom{N}{n}(\eta ES)^n p_0(\mathrm{appr}),\quad 0\le n\le c-1$$

$$p_n(\mathrm{appr}) = \lambda_{c-1}p_{c-1}(\mathrm{appr}) \int_0^\infty (1-F_e(t))^{c-1}(1-F(t))\phi_{nc}(t)dt + \\ + \sum_{j=c}^{n} \lambda_j\, p_j(\mathrm{appr}) \int_0^\infty (1-F(ct))\phi_{nj}(t)dt,\quad c\le n\le N$$

where

$$\phi_{nj}(t) = \binom{N-j}{n-j}\,(1-e^{-\eta t})^{n-j}\,e^{-\eta t\,(N-n)},\quad t>0$$

We refer to [34] for more details and numerical results.

The above approximations have been derived under the specification (3.10) in the approximation assumption. We now discuss a slightly different specification which in particular yields useful results for the case of deterministic service times. For the M/D/c queue with the service times equal to the constant D, consider the specification in which F_j^* for $1\le j\le c-2$ and F^* are the same as in (3.10) but the "boundary" distribution function F_{c-1}^* is chosen as

$$(3.36)\qquad F_{c-1}^*(t) = \begin{cases} 1, & t\ge D/c, \\ 0, & t<D/c. \end{cases}$$

Denoting the corresponding approximations by a bar, a minor modification of the proof of Theorem 3.1. gives the following approximations for the M/D/c queue

$$(3.37)\qquad p_n(\overline{\mathrm{appr}})= \frac{(\lambda D)^n}{n!}\, p_0(\mathrm{appr}),\quad 0\le n\le c-2,$$

$$(3.38)\qquad p_n(\overline{\mathrm{appr}})= \lambda p_{c-2}(\overline{\mathrm{appr}})\int_0^D \left(1-\frac{t}{D}\right)^{c-2} \frac{(\lambda t)^{n-c+1}}{(n-c+1)!}\, e^{-\lambda t}\, dt + \\ + \lambda \sum_{j=c-1}^{n} p_j(\overline{\mathrm{appr}}) \int_0^{D/c} \frac{(\lambda t)^{n-j}}{(n-j)!}\, e^{-\lambda t}\, dt,\quad n\ge c-1.$$

with $p_0(\overline{appr})=\Omega$. In particular we find

$$(3.39)\quad P_W(\overline{appr}) = P_W(\exp) - (\frac{\eta_1}{\eta_2}-1)\frac{(\lambda D)^{c-1}}{(c-1)!}\Omega,$$

where

$$\eta_1 = \frac{c-1}{D}\int_0^D (1-\frac{t}{D})^{c-2}e^{-\lambda t}dt \text{ and } \eta_2 = e^{-\lambda D/c}.$$

Note that $P_W(\overline{appr}) < P_W(\exp)$ since $\eta_1 > \eta_2$ as is readily verified. Numerical results show that $P_W(\overline{appr})$ is a very accurate approximation for the delay probability in the M/D/c queue and considerably improves in almost all cases the good Erlang delay probability approximation $P_W(\exp)$. Also we find after some algebra

$$(3.40)\quad EL_q(\overline{appr}) = \frac{(\lambda D)^{c-1}}{(c-1)!}\Omega\{\frac{\rho^2}{2(1-\rho)^2}+(\frac{\eta_1}{\eta_2}-1)\},$$

$$(3.41)\quad EL_q^2(\overline{appr}) = \frac{(\lambda D)^{c-1}}{(c-1)!}\Omega\{\frac{9\rho^2-11\rho^3+7\rho^4}{6(1-\rho)^3}-\frac{2\rho^2}{(1-\rho)(c+1)}-(\frac{\eta_1}{\eta_2}-1)\}.$$

Further $ET_D^m(\overline{appr})$ for $m\geq 1$ can be easily derived. Although $ET_D(\overline{appr})=1/\lambda$ is exact, our numerical results indicate that $ET_D^m(\overline{appr})$ is less good than $ET_D^m(appr)$ for $m\geq 2$.

We conclude this section by presenting some numerical results. In table 3.1 we consider the M/D/c queue for several values of ρ and c and we give the delay probability P_W, the mean queue size EL_q and the coefficient of variation cvL_q of the queue size. The top numbers in table 3.1 correspond to the exact values, the second top numbers correspond to the approximate values of (3.16)-(3.18) and the third top numbers correspond to the approximate values (3.39)-(3.41). The exact values were taken from [18]. In the tables 3.2-3.3 we deal with the following three phase-type densities represented by an Erlang density, a mixture of Erlang densities and a hyperexponential density,

Case i $\quad f(t) = \mu^2 te^{-\mu t}\ (c_S^2=0.5)$,

Case ii $\quad f(t) = p\mu e^{-\mu t}+(1/2)(1-p)\mu^3t^2e^{-\mu t}$ with $p=2/3-(1/6)\sqrt{7}\ (c_S^2=0.5)$

Case iii $\quad f(t) = p\mu_1 e^{-\mu_1 t}+(1-p)\mu_2 e^{-\mu_2 t}$ with $\frac{p}{\mu_1}=\frac{1-p}{\mu_2}$, $p=(1+1/\sqrt{5})/2\ (c_S^2=1.5)$.

For these phase-type service times we give in table 3.2 the delay probability P_W, the mean queue size EL_q and the coefficient of variation cvL_q of the queue size for several values of c with a traffic intensity $\rho=0.8$, where the top numbers in table 3.2 correspond to the exact values. The exact values were computed by using the decomposition method of [29] to solve equilibrium state equations. Finally, table 3.3 concerns the coefficient of variation c_D of the interdeparture time T_D for deterministic service times ($c_S^2=0$) and the above three phase-type service times for several values of c where the traffic intensity $\rho=0.8$. The top numbers in table 3.3 give the simulated actual values of c_D with a 95% percent confidence interval and the second top numbers give the approximate value of c_D corresponding to (3.21).

Appendix.

PROOF OF THEOREM 3.1. Under the condition that the system is empty at epoch 0, define $M_{nk}(t)$ as the joint probability that exactly n customers arrive in (0,t) and that the service completions of the first k customers all occur beyond time t, $1\leq k\leq c$ and $n\geq k$. In the same way as (3.23), we derive for $1\leq k\leq c$ and $n\geq k$,

$$\text{(A.1)}\qquad \frac{d}{dt} M_{nk}(t) = -\lambda M_{nk}(t) + \lambda(1-F(t))M_{n-1,k-1}(t)dt, \quad t>0,$$

where $M_{n0}(t)$ is defined by

$$\text{(A.2)}\qquad M_{n0}(t) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}, \quad n\geq 0, \ t\geq 0.$$

Using the approximation assumption with specification (3.10) and using the argument below (2.13), it follows that

$$\text{(A.3)}\qquad A_{nj} = \int_0^\infty (1-F_e(t))^j M_{n-j,n-j}(t)dt, \quad 0\leq j\leq n\leq c-1,$$

$$\text{(A.4)}\qquad A_{nj} = \int_0^\infty (1-F_e(t))^j M_{n-j,c-j}(t)dt, \quad 0\leq j\leq c-1, \ n\geq c,$$

$$\text{(A.5)}\qquad A_{nj} = \int_0^\infty (1-F(ct))e^{-\lambda t}\frac{(\lambda t)^{n-j}}{(n-j)!}dt, \quad c\leq j\leq n.$$

By (A.1) and (3.25) we can rewrite (A.3) for $0\leq j<n$ as

$$(A.6) \qquad A_{nj} = \int_0^\infty (1-F_e(t))^j (1-F(t)) M_{n-j-1,n-j-1}(t)dt + $$
$$- \frac{j}{\lambda ES} \int_0^\infty (1-F_e(t))^{j-1}(1-F(t))M_{n-j,n-j}(t)dt.$$

where for j=0 the second term in the right side of (A.6) vanishes. Now we first derive from (3.8) that

$$(A.7) \qquad q_k = \frac{1}{\lambda} \frac{(\lambda ES)^k}{k!}, \; 0 \le k \le c-1.$$

By (A.7) and (3.9) we get (3.12). We prove (A.7) by induction on n. Clearly, by (3.8), (A.7) holds for k=0. Fix $1 \le n \le c-1$. Assume that (A.7) holds for $k=0,\dots,n-1$. Using this induction assumption and (A.6), it is readily verified from (3.8) that

$$(A.8) \qquad (1-\lambda A_{nn})q_n = \frac{(\lambda ES)^{n-1}}{(n-1)!} \int_0^\infty (1-F_e(t))^{n-1}(1-F(t))M_{00}(t)dt.$$

Further, by partial integration, we get from (A.3) that

$$(A.9) \qquad 1-\lambda A_{nn} = \frac{n}{ES} \int_0^\infty (1-F_e(t))^{n-1}(1-F(t))e^{-\lambda t}dt.$$

By (A.8)-(A.9), we get (A.7) for k=n. Next, we verify (3.13). By rewriting (A.4) in the same way as (A.3) and using (A.7), we find for $n \ge c$

$$(A.10) \qquad \sum_{j=0}^{c-1} \lambda q_j A_{nj} = \lambda q_{c-1} \int_0^\infty (1-F_e(t))^{c-1}(1-F(t))M_{n-c,0}(t)dt.$$

By (3.8)-(3.9), (A.5) and (A.10) we get (3.13). Finally, by (3.12)-(3.13) and $\Sigma_0^\infty p_n(\text{appr})=1$, we verify $p_0(\text{appr})=\Omega$.

Acknowledgements

We would like to thank R. Corver and H. Groenevelt for their help in writing computer programs.

Table 3.1 *Delay probability and queue size for the M/D/C queue (exact, appr.* $\overline{\textit{appr.}}$*)*

	ρ=0.5			ρ=0.8			ρ=0.9			ρ=0.95		
	P_W	EL_q	cvL_q	P_W	EL_q	cvL_q	P_W	EL_q	cvL_q	P_W	EL_q	cvL_q
c=2	.3233	.1767	3.110	.7091	1.445	1.515	.8471	3.865	1.231	.9227	8.824	1.110
	.3333	.1944	2.986	.7111	1.517	1.471	.8526	3.965	1.207	.9256	8.940	1.097
	.3193	.1807	3.102	.6915	1.442	1.523	.8393	3.850	1.236	.9180	8.801	1.113
c=3	.2253	.1308	3.717	.6325	1.329	1.614	.8077	3.721	1.271	.9018	8.665	1.129
	.2368	.1480	3.531	.6472	1.424	1.551	.8171	3.861	1.237	.9070	8.832	1.110
	.2227	.1326	3.730	.6226	1.319	1.630	.7996	3.694	1.280	.8968	8.627	1.133
c=5	.1213	.0766	5.051	.5336	1.156	1.787	.7478	3.495	1.339	.8692	8.411	1.159
	.1304	.0869	4.778	.5541	1.256	1.706	.7625	3.660	1.296	.8778	8.617	1.136
	.1206	.0750	5.124	.5283	1.134	1.814	.7427	3.451	1.353	.8658	8.351	1.167
c=10	.0331	.0237	9.612	.3847	.8786	2.163	.6469	3.101	1.474	.8116	7.952	1.218
	.0361	.0254	9.152	.4092	.9523	2.057	.6687	3.256	1.423	.8256	8.164	1.192
	.0330	.0212	9.929	.3875	.8400	2.200	.6491	3.029	1.494	.8128	7.856	1.230
c=15	.0104	.0080	17.08	.2955	.7012	2.510	.5771	2.820	1.587	.7695	7.610	1.266
	.0113	.0081	16.40	.3192	.7501	2.384	.6026	2.949	1.532	.7870	7.803	1.239
	.0104	.0067	17.87	.3016	.6559	2.550	.5842	2.730	1.609	.7743	7.489	1.279
c=25				.1900	.4774	3.196	.4793	2.412	1.782	.7063	7.088	1.344
				.2091	.4954	3.029	.5079	2.497	1.721	.7284	7.239	1.317
				.1973	.4301	3.236	.4920	2.302	1.806	.7164	6.931	1.359
c=50				.0776	.2142	5.123	.3355	1.779	2.203	.6012	6.195	1.497
				.0870	.2073	4.840	.3639	1.795	2.125	.6291	6.264	1.470
				.0819	.1789	5.161	.3522	1.649	2.225	.6185	5.987	1.516
c=100				.0176	.0541	10.94	.1953	1.110	2.989	.4751	5.077	1.740
				.0196	.0470	10.34	.2169	1.072	2.871	.5065	5.047	1.710
				.0185	.0404	11.02	.2099	.9833	2.998	.4979	4.820	1.760
c=200				.0013	.0043	41.14	.0837	.5194	4.707	.3351	3.766	2.149
				.0014	.0033	39.24	.0945	.4672	4.497	.3653	3.642	2.107
				.0013	.0028	41.79	.0914	.4282	4.683	.3590	3.476	2.163

Table 3.2 *Delay prob. and queue size for phase-type services (exact, appr.)*

	case i (c_S^2=0.5)			case ii (c_S^2=0.5)			case iii (c_S^2=1.5)		
	P_W	EL_q	cvL_q	P_W	EL_q	cvL_q	P_W	EL_q	cvL_q
c=2	.7087	2.148	1.485	.7083	2.151	1.478	.7131	3.522	1.490
	.7111	2.169	1.475	.7111	2.132	1.487	.7111	3.484	1.501
c=3	.6432	1.964	1.585	.6426	1.969	1.577	.6503	3.183	1.595
	.6472	1.992	1.570	.6472	1.951	1.585	.6472	3.148	1.610
c=4	.5914	1.816	1.675	.5907	1.823	1.667	.6003	2.917	1.689
	.5964	1.847	1.657	.5964	1.804	1.674	.5964	2.890	1.705
c=5	.5484	1.693	1.758	.5477	1.700	1.750	.5584	2.697	1.775
	.5541	1.723	1.737	.5541	1.679	1.758	.5541	2.680	1.792
c=6	.5116	1.586	1.837	.5108	1.594	1.829	.5224	2.510	1.855
	.5178	1.615	1.814	.5178	1.571	1.836	.5178	2.501	1.874
c=7	.4794	1.493	1.913	.4786	1.501	1.904	.4908	2.346	1.933
	.4859	1.520	1.888	.4859	1.476	1.912	.4859	2.345	1.952
c=8	.4508	1.409	1.986	.4501	1.417	1.977	.4627	2.202	2.007
	.4576	1.434	1.959	.4576	1.391	1.986	.4576	2.207	2.028
c=9	.4253	1.333	2.057	.4245	1.342	2.049	.4373	2.073	2.079
	.4322	1.357	2.029	.4322	1.314	2.057	.4322	2.083	2.101
c=10	.4021	1.265	2.127	.4014	1.274	2.119	.4143	1.956	2.150
	.4092	1.286	2.098	.4092	1.245	2.127	.4092	1.971	2.173
c=15	.3122	.9955	2.466	.3116	1.004	2.457	.3241	1.507	2.490
	.3192	1.008	2.430	.3192	.9719	2.466	.3192	1.536	2.518

Table 3.3 *The coefficient of variation of the output process (sim., appr.)*

	c_S^2 = 0	case i(c_S^2=0.5)	case ii(c_S^2=0.5)	case iii(c_S^2=1.5)
c=2	.7438 (± .0072)	.8836 (± .0058)	.8979 (± .0064)	1.065 (± .0077)
	.6849	.8543	.8455	1.121
c=3	.8074 (± .0073)	.9136 (± .0069)	.9294 (± .0064)	1.043 (± .0077)
	.7308	.8725	.8632	1.106
c=4	.8418 (± .0080)	.9321 (± .0063)	.9502 (± .0046)	1.030 (± .0072)
	.7617	.8856	.8760	1.096
c=5	.8644 (± .0068)	.9474 (± .0051)	.9635 (± .0068)	1.021 (± .0064)
	.7847	.8959	.8861	1.089
c=10	.9303 (± .0040)	.9734 (± .0029)	.9859 (± .0041)	1.007 (± .0038)
	.8522	.9273	.9182	1.065
c=15	.9527 (± .0054)	.9849 (± .0047)	.9921 (± .0041)	1.006 (± .0039)
	.8884	.9273	.9370	1.051

REFERENCES.

1. BOXMA, O.J., COHEN, J.W. & HUFFELS, N. (1979), Approximations of the mean waiting time in an M/G/c queueing system. *Oper. Res.* 27, 1115-1127.

2. BURMAN, D.Y. & SMITH, D.R. (1981), A light-traffic theorem for multi-server queues, Report, Bell Laboratories, Holmdel, New Jersey.

3. BUX, W. and HERZOG, U. (1977), The phase-concept, approximation of measured data and performance analysis, in: K.M. Chandy and M. Reiser (eds.), Proceedings of the International Symposium on Computer Performance, Modeling, Measurement and Evaluation, North-Holland, Amsterdam, pp. 23-38.

4. BUX, W. (1979), Single server queues with general interarrival and phase-type service time distributions-computational algorithms, Proceedings of the 9th International Teletraffic Congress (Torremolinos).

5. COHEN, J.W. (1977), On up- and downcrossings, *J. Appl. Prob.* 14, 405-410.

6. CROMMELIN, C.D. (1932), Delay probability formulae when the holding times are constant. *P.O. Elec. Engrs. J.* 25, 41-50.

7. FEDERGRUEN, A. & TIJMS, H.C. (1980), Computation of the stationary distribution of the queue size in an M/G/1 queueing system with variable service rate, *J. Appl. Prob.* 17, 515-522.

8. HALACHMI, B. & FRANTA, W.R. (1978), A diffusion process approximation to the multi-server queue, *Manag. Sci.* 24, 522-529.

9. HEFFER, J.C. (1969), Steady-state solution of the $M/E_k/c(\infty,FIFO)$ queueing system, *CORS J.* 7, 16-30.

10. HILLIER, F.S. & LO, F.D. (1971), Tables for multiple-server queueing systems involving Erlang distributions, Technical Report no. 31, Dept. of Operations Research, Stanford Univ., Stanford.

11. HOKSTAD, P. (1978), Approximations for the M/G/m queue, *Oper. Res.* 26, 511-523.

12. HOKSTAD, P. (1980), The steady-state solution of the $M/K_2/m$ queue, *Adv. Appl. Prob.* 12, 799-823.

13. ISHIKAWA, A. (1979), On the equilibrium solution for the queueing system $GI/E_k/m$, *TRU Mathematics*, Vol. 15, No 1, 47-66.

14. KAMPE, G. & KÜHN, P.J. (1976), Graded delay systems with infinite or finite source traffic and exponential or constant holding time, Proceedings of the 8th International Teletraffic Congress (Melbourne), 256-1/10.

15. KOBAYASHI, H. (1976), *Modelling and Analysis: an Introduction to System Performance and Evaluation Methodology*, Addison-Wesley, Reading.

16. KÖLLERSTRÖM, J. (1974), Heavy traffic theory for queues with several servers, *J. Appl. Prob.* 11, 544-552.

17. KÜHN, P.J. (1972), On the Calculation of Waiting Times in Switching and Computer Systems, 15th Report on Studies in Congestion Theory, Institute of Switching and Data Technics, University of Stuttgart.

18. KÜHN, P.J. (1976), *Tables on Delay Systems*, Institute of Switching and Data Technics, University of Stuttgart.

19. LAVENBERG, S.S. (1975), The steady-state queueing time distribution for the M/G/1 finite capacity queue. Manag. Sci. 25, 501-506.

20. MARSHALL, K.T. & WOLFF, R.W. (1971), Customer average and the time average queue lengths and waiting times, *J. Appl. Prob.* 8, 535-542.

21. NEUTS, M.F. (1980), *Matrix-Geometric Solutions in Stochastic Models - An algorithmic Approach*, The John Hopkins University Press (to appear).

22. NEWELL, G. (1973), *Approximate Stochastic Behaviour of n-Server Service Systems with Large n*, Lecture Notes in Economics and Mathematical Systems, 87, Springer-Verlag, Berlin.

23. NOZAKI, S.A. & ROSS, S.M. (1978), Approximations in finite-capacity multi-server queues with Poisson arrivals, *J. Appl. Prob.* 15, 826-834.

24. PACK, C.D. (1978), The output of a multiserver queueing system, *Oper. Res.* 26, 492-509.

25. ROSS, S.M. (1970), *Applied Probability Models with Optimization Applications*, Holden-Day, San Francisco.

26. SAKASEGAWA, H. (1977), An approximation formula $L_q \simeq \alpha\rho^{\beta}/(1-\rho)$, *Ann. Inst. Statist. Math.* 29 Part A, 67-75.

27. STIDHAM, S., Jr. (1972), Regenerative processes in the theory of queues with applications to the alternating priority queue, *Adv. Appl. Prob.* 4, 542-557.

28. TAKACS, L. (1962), *Introduction to the Theory of Queues*, Oxford University Press.

29. TAKAHASHI, Y. & TAKAMI, Y. (1976), A numerical method for the steady-state probabilities of a GI/G/c queueing system in a general class, *J. Oper. Res. Soc. Japan* 19, 147-157.

30. TAKAHASHI, Y. (1977), An approximation formula for the mean waiting time of an M/G/c queue, *J. Oper. Res. Soc. Japan* 20, 150-163.

31. TIJMS, H.C., VAN HOORN, M.H. & FEDERGRUEN, A. (1979), Approximations for the steady-state probabilities in the M/G/c queue, Research Report 49, Vrije Universiteit, Amsterdam (to appear in *Adv. Appl. Prob.*).

32. TIJMS, H.C. & VAN HOORN, M.H. (1980), Algorithms for the state probabilities and waiting times in single server queueing systems with random and quasi-random input and phase-type service times, Research Report 51, Vrije Universiteit (to appear in *OR Spektrum)*

33. VAN HOORN, M.H. (1980), Algorithms for the state probabilities in a general class of single server queueing systems with group arrivals, Research Report 56, Vrije Universiteit, Amsterdam.

34. VAN HOORN, M.H. (1981), Approximations for the machine repair problem with multiple repairmen (forthcoming).