

**stichting  
mathematisch  
centrum**



---

AFDELING MATHEMATISCHE BESLISKUNDE  
(DEPARTMENT OF OPERATIONS RESEARCH)

BW 120/80

MAART

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

ALGORITHMS FOR THE STATE PROBABILITIES AND WAITING TIMES IN  
SINGLE SERVER QUEUEING SYSTEMS WITH RANDOM AND QUASIRANDOM  
INPUT AND PHASE-TYPE SERVICE TIMES

Preprint

---

**2e boerhaavestraat 49 amsterdam**

*Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, 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).*

Algorithms for the State Probabilities and Waiting Times in Single Server  
Queueing Systems with Random and Quasirandom Input and Phase-Type Service  
Times<sup>\*</sup>

by

H.C. Tijms<sup>\*\*</sup> & M.H. Van Hoorn<sup>\*\*\*</sup>

ABSTRACT. This paper presents a unifying algorithmic analysis for a general class of single server queueing systems with a state dependent Markovian input process and a phase-type service time distribution including single server queues with random and quasirandom input.

Using regenerative analysis we develop numerically stable and efficient recursion schemes to compute the state probabilities. The computation of the waiting times is based on the state probabilities.

KEY WORDS & PHRASES: *Single server queue, state-dependent Markovian input, phase-type service time, computational methods, state probabilities, waiting times.*

---

\* This paper will be submitted for publication elsewhere

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

\*\*\* Vrije Universiteit, Amsterdam



## 1. Introduction.

In recent years substantial contributions have been made to computational aspects of queueing analysis in particular in the field of teletraffic theory, e.g. cf. Bux and Herzog [1], Bux [2], Kampe and Kühn [5], Kobayashi [6], Kühn [7,8] and Neuts [10]. A powerful tool for the computational analysis of queueing systems is the use of phase-type distributions by which the queueing processes can be analysed by Markov processes.

In this paper we consider a single-server queueing system at which customers singly arrive according to a state-dependent Markovian input process (i.e. exponentially distributed interarrival times) with rate  $\lambda_j$  when  $j$  customers are in the system and each arriving customer joins the system. Customers are served in order of arrival where the service times of the customers are independent and distributed as the random variable  $S$  having probability distribution function  $F(t)$  with  $F(0) = 0$ . It is assumed that the sequence  $\{\lambda_j, j \geq 0\}$  is bounded and that conditions for statistical equilibrium are satisfied, e.g.  $\limsup_{n \rightarrow \infty} \lambda_n ES < 1$ . This single-server queueing system with state-dependent arrival rate covers a number of important queueing models with random or quasirandom input including

- (i) the M/G/1 queue with  $K \leq \infty$  waiting places,
- (ii) the cyclic queueing system as in figure 1 with a fixed number  $K$  of customers where service station 1 has  $c$  exponential servers and service station 2 has a single server with service time distribution  $F$ . Note that this cyclic queueing system has as special cases the finite capacity M/G/1 queue (take  $c = 1$ ) and the machine repair problem with a single repairman having a general repair time (take  $c = K$ ), cf. Kobayashi [6] and Lavenberg [9].

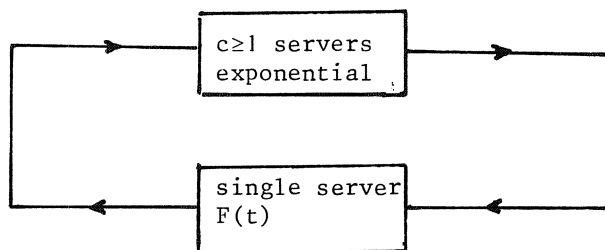


Figure 1.1. Closed cyclic queueing system.

We shall present a unifying algorithmic analysis for the state-probabilities and the waitingtimes in the above single-server queueing system where we assume that the service time distribution function  $F$  has a phase-type representation and is given by a hyperexponential distribution function or by a mixture of Erlang distribution functions with the same scale parameter. Observe that any probability distribution function concentrated on  $(0, \infty)$  can be approximated by a mixture of Erlang distribution functions with the same scale parameter, see Bux and Herzog [1] and Schassberger [13]. For the phase-type service time distributions a direct algorithmic approach is to solve the multi-dimensional equilibrium state equations for the resulting continuous-time Markov chains by iterative methods using overrelaxation or possibly by recursion using the method of macrostates given in Cooper [3] and Kühn [7]. However, this approach results in computational methods which may be computationally expensive or not numerically stable. In our algorithmic approach based on regenerative analysis and to some extent on aggregation of the equilibrium microstate equations the state probabilities are computed from numerically stable and efficient recursion schemes. The computation of the waiting times is based on the state probabilities.

In section 2 we present the regenerative analysis and derive the basic recurrence relation. In the sections 3 and 4 we specify the algorithms for the hyperexponential service time distribution and mixtures of Erlang service time distributions with the same scale parameter. Finally, in section 5 we give some numerical results.

## 2. *The regenerative analysis and the recursion scheme.*

We first introduce some notation. Throughout this paper, we assume that the system is empty at epoch 0 unless stated otherwise. 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  $(0, T]$ ,  $n \geq 0$ ,
- $N$  = 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  $(0, T]$ ,  $n \geq 0$ .

Further, define the state probabilities

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

and

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

Also, define the actual waiting time distribution function

$$W(t) = \lim_{k \rightarrow \infty} \Pr \{ \text{the } k^{\text{th}} \text{ customer has to wait at most an amount of time } t \text{ before his service starts} \}, t \geq 0.$$

Denoting by the random variable  $W$  the steady-state waiting time of an arbitrary customer before being served, the complementary waiting time distribution function with respect to the delayed customers is given by

$$\Pr \{ W > t | W > 0 \} = \frac{1-W(t)}{1-W(0)}, t \geq 0$$

By the assumption of statistical equilibrium the above limits exist and the limiting distributions are probability distributions (cf. Stidham [14]). Note that in general  $p_n \neq \pi_n$  except for random input. We have the following results.

THEOREM 2.1.

$$(2.1) \quad p_n = ET_n/ET, \quad \pi_n = EN_n/EN \text{ for all } n \geq 0,$$

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

$$(2.3) \quad EN/ET = \sum_{j=0}^{\infty} \lambda_j p_j, \quad \sum_{j=0}^{\infty} \lambda_j p_j = (1-p_0)/ES.$$

PROOF. By the theory of regenerative processes (cf. Ross [12] and Stidham [14]) the first part of (2.1) follows and  $EN_n/EN$  equals the long-run expected fraction of customers who leave upon service completion  $n$  other customers behind in the system. Since customers both arrive and are served one by one, this fraction equals the long-run expected fraction of customers who see upon arrival  $n$  other customers in the system and this proves the second part of (2.1).

A rigorous and simple way to prove (2.2)-(2.3) is as follows. Choose a finite number  $\lambda^*$  such that  $\lambda^* \geq \lambda_j$  for all  $j \geq 0$  and define  $\phi_j = \lambda_j / \lambda^*$  for  $j \geq 0$ . Consider now the modified single-server queueing system at which customers arrive according to a Poisson process (random input) with rate  $\lambda^*$  and any customer seeing upon arrival  $j$  other customers in the system enters the system with probability  $\phi_j$  and leaves without joining the system with probability  $1 - \phi_j$ . The service time of each entering customer has the same distribution as in the original system. Clearly, in the modified queueing system the steady-state probability that an *entering* customer sees upon arrival  $n$  other customers in the system equals  $\pi_n$  and the steady-state probability that at an arbitrary epoch there are  $n$  customers in the system equals  $p_n$  for all  $n \geq 0$ . Moreover, in the modified queueing system arriving customers see time averages because of the random input, more precisely the steady-state probability that an arbitrary customer sees upon arrival  $n$  other customers in the system equals  $p_n$  (cf. Theorem 3 in Stidham [14]). Hence in the modified system the steady-state probability  $\pi_n$  that an entering customer sees upon arrival  $n$  other customers in the system is given by

$$p_n \phi_n / \sum_{j=0}^{\infty} p_j \phi_j = \lambda_n p_n / \sum_{j=0}^{\infty} \lambda_j p_j$$

since  $\sum_{j=0}^{\infty} p_j \phi_j$  is the steady-state probability that an arriving customer enters the system. To prove (2.3), note that by the theory of regenerative processes the long-run expected average number of customers served per unit time is equal to  $EN/ET$ . Hence in the modified system the long-run expected average number of customers entering the system per unit equals  $EN/ET$ . Since in the modified system the steady-state probability that an arriving customer enters the system is given by  $\sum_{j=0}^{\infty} p_j \phi_j = (1/\lambda^*) \sum_{j=0}^{\infty} \lambda_j p_j$  and the expected number of arrivals per unit time is  $\lambda^*$ , we get the first part of (2.3). Noting that the long-run expected fraction of time that the server is busy equals  $1 - p_0$ , we get by Little's formula (cf. Stidham [14,15]) that  $1 - p_0 = (EN/ET)ES$  which ends the proof.

To derive a recurrence relation between the probabilities  $p_n$  and  $\pi_n$ , we define the following quantity. For  $n \geq k \geq 1$ , let

$A_{n,k}$  = expected amount of time during which  $n$  customers are in the system in a service time  $S$  given that at the start of this service  $k$  customers are present.

Then, by partitioning the busy cycle  $(0, T]$  by means of the service completion



epochs and using Wald's theorem (cf. Ross [13]), it readily follows that

$$(2.4) \quad ET_0 = 1/\lambda_0$$

$$(2.5) \quad ET_n = A_{n,1} + \sum_{k=1}^n A_{n,k} EN_k \quad \text{for } n \geq 1.$$

Together relations (2.1)-(2.5) imply the following result.

THEOREM 2.2.

$$(2.6) \quad p_0 ET = 1/\lambda_0$$

$$(2.7) \quad p_n ET = A_{n,1} + \sum_{k=1}^n \lambda_k A_{n,k} p_k ET \quad \text{for } n \geq 1.$$

Once we have evaluated the quantities  $A_{n,k}$  we can recursively compute the numbers  $p_0 ET, p_1 ET, \dots$  from the numerically stable recursion scheme (2.6)-(2.7) and next obtain from (2.2) the state probabilities  $\pi_n, n \geq 0$  in any desired accuracy by normalization. The above regenerative approach introduced in Hordijk and Tijms [8] is a fertile approach which has been applied amongst others to derive approximations for the state probabilities in the M/G/c queue, see Tijms, Van Hoorn and Federgruen [17].

For hyperexponential service time distributions and mixtures of Erlang service time distributions we discuss efficient procedures to evaluate the quantities  $A_{n,k}$  and computational methods for the waiting times in the next sections 3 and 4.

### 3. Algorithms for the hyperexponential service time distribution.

In this section we assume a hyperexponential service time distribution function

$$F(t) = \sum_{i=1}^r q_i (1 - e^{-\mu_i t}),$$

i.e. with probability  $q_i$  the service time of a customer is exponentially distributed with mean  $1/\mu_i$  (service time of type  $i$ ). We first specify the evaluation of the numbers  $A_{n,k}$  needed in the recursive computation of the state probabilities  $p_n$  and next we discuss the computation of the waiting times.

Clearly we have for all  $n \geq k \geq 1$

$$(3.1) \quad A_{n,k} = \sum_{i=1}^r q_i A_{n,k,i}$$

where

$A_{n,k,i}$  = expected amount of time during which  $n$  customers are in the system in one service time of type  $i$  given that at the start of this service  $k$  customers are present.

Using the property that with probability  $\lambda_j / (\mu + \lambda_j)$  a customer arrives before the completion of an exponential service with mean  $1/\mu$  when  $j$  customers are present and using the memoryless property of the exponential distribution, it follows that for all  $n \geq k$  and  $1 \leq i \leq r$ ,

$$(3.2) \quad A_{n,k,i} = \left( \frac{1}{\mu_i + \lambda_n} \right) \prod_{j=k}^{n-1} \left( \frac{\lambda_j}{\mu_i + \lambda_j} \right)$$

This completes the determination of the state probabilities  $p_n$  and  $\pi_n$ .

To determine the waiting time distribution function  $W$ , we first observe that the stochastic process  $\{X(t), t \geq 0\}$  with  $X(t)$  describing jointly the number of customers present and the type of the service in progress at time  $t$  (if any) is a two-dimensional continuous-time Markov chain. To determine  $W$ , we define for  $n \geq 1$  and  $1 \leq i \leq r$  the state probabilities,

$$p_{ni} = \lim_{t \rightarrow \infty} \Pr\{\text{at time } t \text{ there are } n \text{ customers in the system and the service in progress is of type } i\}$$

$$\pi_{ni} = \lim_{k \rightarrow \infty} \Pr\{\text{the } k^{\text{th}} \text{ customer sees upon arrival } n \text{ other customers in the system and a service in progress of type } i\}.$$

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

$$(3.3) \quad \pi_{ni} = \lambda_n p_{ni} / \sum_{j=0}^{\infty} \lambda_j p_j \quad \text{for } n \geq 1, 1 \leq i \leq r$$

The (unconditional) waiting time distribution function  $W(t)$  is given by

$$(3.4) \quad W(t) = \pi_0 + \sum_{n=1}^{\infty} \sum_{i=1}^r \pi_{ni} W(t|n,i) \quad \text{for } t \geq 0$$

where the conditional waiting time function  $W(t|n,i)$  gives the probability that an arriving customer has to wait at most a time  $t$  before his service starts given that he finds upon arrival  $n$  other customers in the system and a service in progress of type  $i$ . We first consider the evaluation of the state probabilities  $\pi_{ni}$ . Therefore we define for  $n \geq 1$  and  $1 \leq i \leq r$  the random variable

$T_{ni}$  = amount of time during which  $n$  customers are in the system and a service of type  $i$  is in progress in the busy cycle  $(0, T]$ .

Then, by the theory of regenerative processes,

$$(3.5) \quad p_{ni} = ET_{ni}/ET \quad \text{for } n \geq 1, 1 \leq i \leq r$$

and, by the same arguments as used to derive (2.5),

$$(3.6) \quad ET_{ni} = q_i A_{n,1,i} + \sum_{k=1}^n q_i A_{n,k,i} EN_k \quad \text{for } n \geq 1, 1 \leq i \leq r.$$

Together (2.1)-(2.3) and (3.5)-(3.6) imply

$$(3.7) \quad p_{ni} ET = q_i A_{n,1,i} + q_i \sum_{k=1}^n \lambda_k A_{n,k,i} p_k ET \quad \text{for } n \geq 1, 1 \leq i \leq r.$$

Note from (2.5)-(2.6), (3.1) and (3.7) that the recursive computation of the numbers  $p_n ET$  and  $p_{ni} ET$  ( $1 \leq i \leq r$ ) can be simultaneously done for  $n \geq 1$ . We further note that alternatively the state probabilities  $p_{ni}$  and so  $p_n = \sum_{i=1}^r p_{ni}$  may be obtained by solving the equilibrium micro-state equations for the above continuous-time Markov chain  $\{X(t), t \geq 0\}$ . However, this will require the iterative solution of a two-dimensional system of linear equations which compares quite unfavourably from a computational point of view with our numerically stable recursion scheme.

We now turn to the computation of the waiting time function  $W$  given by (3.4). The main computational problem is the computation of the conditional waiting time distributions  $W(t|n,i)$ . Both for service in order of arrival, service in random order and service in reverse order of arrival, we can derive by standard arguments a system of first-order linear differential equations in these conditional waiting time functions, cf. Cooper [3], Kühn [7] and Riordan [11]. However, for large systems of linear differential

equations computational difficulties will arise. For service in order of arrival another computational procedure is suggested by the specific form of  $F(t)$ . Therefore denote  $F^{(n)}$  the  $n$ -fold convolution of  $F$  with itself and let  $E_{k,\mu}(t)$  be the probability distribution function of the sum of  $k$  independent exponentials with common mean  $1/\mu$ , i.e.

$$(3.8) \quad E_{k,\mu}(t) = 1 - \sum_{j=0}^{k-1} e^{-\mu t} \frac{(\mu t)^j}{j!}, \quad t \geq 0$$

with the usual convention that  $F^{(0)}$  and  $E_{0,\mu}$  are degenerate distributions concentrated at 0. Then, by the assumption of service in order of arrival,

$$(3.9) \quad W(t|n,i) = E_{1,\mu_i} * F^{(n-1)}(t), \quad n \geq 1, \quad 1 \leq i \leq r$$

where  $*$  denotes the convolution operator. Further, assuming for ease the most important case  $r=2$ , we can also write

$$(3.10) \quad W(t|n,1) = \sum_{k=0}^{n-1} \binom{n-1}{k} q_1^k q_2^{n-1-k} E_{k+1,\mu_1} * E_{n-1-k,\mu_2}(t)$$

with

$$(3.11) \quad E_{i,\mu_1} * E_{j,\mu_2}(t) = E_{i,\mu_1}(t) - \sum_{k=0}^{j-1} e^{-\mu_2 t} t^{k+i} \frac{\mu_2^k \mu_1^i}{k!(i-1)!} \alpha_{i,k}(t)$$

where

$$(3.12) \quad \alpha_{i,k}(t) = \int_0^1 e^{-(\mu_1 t - \mu_2 t)x} (1-x)^k x^{i-1} dx.$$

Note that very efficient numerical procedures are available to evaluate the cumulative Poisson probabilities  $1 - E_{i,\mu}(t)$  and that the integral for  $\alpha_{i,k}(t)$  is well-suited for evaluation by numerical integration. In fact  $\alpha_{i,k}(t)$  gives the Laplace transform of the Beta distribution. We can compute  $W(t|n,i)$  from (3.10)-(3.12) for smaller values of  $n$  and from (3.9) for larger values of  $n$  by approximating  $F^{(n-1)}$  by a normal distribution function.

We conclude by remarking that for the waiting time distribution several approximation procedures based on matching of moments have been suggested in the literature. An approximation procedure given in Riordan [11] and Kühn [7] is to approximate the complementary distribution function  $\Pr\{W>t|W>0\}$

respectively the complementary distribution functions  $1-W(t|n,i)$  by sums of exponentials by matching the first  $k$  moments (if possible). A simple approximation procedure suggested by Kühn [8] is to approximate  $1-P\{W>t|W>0\}$  by a Weibull distribution function  $1-\exp(-(at)^b)$  by matching the first two moments. Note that, by using Little's formula (cf. Stidham [14,15]) the first part of (2.3), (3.4) and (3.9),

$$(1-\pi_0)E(W|W>0) = (1 / \sum_{j=0}^{\infty} \lambda_j p_j) \sum_{n=1}^{\infty} (n-1) p_n,$$

$$(1-\pi_0)E(W^2|W>0) = \sum_{n=1}^{\infty} \sum_{i=1}^r \pi_{ni} \left\{ \frac{2}{\mu_i} + (n-1)ES^2 + \right.$$

$$\left. + (n-1)(n-2)(ES)^2 + 2(n-1) \frac{ES}{\mu_i} \right\},$$

where  $ES = \sum_{i=1}^r q_i / \mu_i$  and  $ES^2 = \sum_{i=1}^r 2q_i / \mu_i^2$ .

#### 4. Algorithms for mixtures of Erlang distribution functions.

In this section we assume that the service time probability distribution function  $F$  is a mixture of Erlang distribution functions with the same scale parameter, i.e.

$$(4.1) \quad F(t) = \sum_{k=1}^r q_k E_{k,\mu}(t), \quad t \geq 0$$

where the Erlang distribution function  $E_{k,\mu}(t)$  is given by (3.8). For any probability distribution function  $G$  concentrated on  $(0, \infty)$  there exists a sequence of phase-type distribution functions as (4.1) converging weakly to  $G$ , see Schassberger [13]. An algorithm to determine a phase-type distribution function as (4.1) fitting with prescribed accuracy a given probability distribution function has been given in Bux and Herzog [1].

We first specify the evaluation of the numbers  $A_{n,k}$  needed in the recursive computation of the state probabilities  $p_n$  and next we turn to the computation of the waiting time distribution function  $W$ . The following terminology will be useful. By (4.1), the service time of a customer can be imagined to consist of  $k$  independent *phases* with probability  $q_k$  where each phase requires an exponential service time with the same mean  $1/\mu$ .

A service time of a customer and a service in progress are said to be of type  $i$  if this service consists of  $i$  phases still to complete. We now define the quantities  $A_{n,k,i}$ , the state probabilities  $p_{ni}$  and  $\pi_{ni}$  and the conditional waiting time distribution functions  $W(t|n,i)$  in the same way as in section 3. We again have

$$(4.2) \quad A_{n,k} = \sum_{i=1}^r q_i A_{n,k,i}, \quad n \geq k \geq 1,$$

$$(4.3) \quad \pi_{ni} = \lambda_n p_{ni} / \sum_{j=0}^{\infty} \lambda_j p_j, \quad n \geq 1, \quad 1 \leq i \leq r,$$

$$(4.4) \quad W(t) = \pi_0 + \sum_{n=1}^{\infty} \sum_{i=1}^r \pi_{ni} W(t|n,i), \quad t \geq 0.$$

For any fixed  $n \geq 1$ , we have by the properties of the exponential distribution

$$(4.5) \quad A_{n,n,i} = 1 / (\mu + \lambda_n), \quad 1 \leq i \leq r,$$

$$(4.6) \quad A_{n,k,i} = \frac{\lambda_k}{\mu + \lambda_k} A_{n,k+1,i} + \frac{\mu}{\mu + \lambda_k} A_{n,k,i-1}, \quad 1 \leq k \leq n, \quad 1 \leq i \leq r$$

where  $A_{n,k,0} = 0$ . Thus for any fixed  $n \geq 1$  starting with  $A_{n,n,1}$  we can recursively compute from a stable and efficient scheme the numbers  $A_{n,k,i}$  for  $k=n, \dots, 1$  and  $i=1, \dots, r$  since we only need  $A_{n,k+1,i}$  and  $A_{n,k,i-1}$  to compute  $A_{n,k,i}$ .

REMARK 4.1. If

$$\lambda_n = \begin{cases} \lambda & \text{for } M \leq n < K \\ 0 & \text{for } n \geq K \end{cases}$$

for some  $M < K \leq \infty$ , then we can directly compute  $A_{n,k}$  for  $k \geq M$  from

$$A_{n,k} = \begin{cases} \int_0^{\infty} (1-F(t)) e^{-\lambda t} \frac{(\lambda t)^{n-k}}{(n-k)!} dt, & k \leq n < K \\ \int_0^{\infty} (1-F(t)) \sum_{j=K-k}^{\infty} e^{-\lambda t} \frac{(\lambda t)^j}{j!} dt, & n = K. \end{cases}$$

This relation follows by noting that  $A_{n,k} = \int_0^{\infty} E \chi_{n,k}(t) dt$  where, conditionally that at epoch 0 a new service starts with  $k$  customers present,  $\chi_{n,k}(t) = 1$  if at time  $t$  this service is still in progress and  $n$  customers are present and  $\chi_{n,k}(t) = 0$  otherwise.

We now turn to the computation of the state probabilities  $\pi_{ni}$  and the conditional waiting time distribution functions  $W(t|n,i)$ . The probabilities  $p_{ni}$  could be evaluated from a stable recursion relation similar to (3.7), but this would in general involve the recursive computation of a four-dimensional vector of numbers. However, a more efficient stable computational method for the state probabilities  $p_{ni}$  can be developed by combining the recursion scheme (2.6)-(2.7) and an aggregated system of equilibrium state equations for the continuous-time Markov chain  $\{X(t), t \geq 0\}$  where  $X(t)$  jointly gives the number of customers present and the type of the service in progress (if any) at time  $t$ . We have the following result.

THEOREM 4.1.

$$(4.7) \quad \mu p_{n+1,1} = \lambda_n p_n, \quad n \geq 0,$$

$$(4.8) \quad (\mu + \lambda_n) p_{ni} = \lambda_{n-1} \sum_{j=i}^r p_{n-1,j} \left( \frac{\mu}{\mu + \lambda_n} \right)^{j-i} + \\ + \mu p_{n+1,1} \sum_{j=i}^r q_j \left( \frac{\mu}{\mu + \lambda_n} \right)^{j-i}, \quad n \geq 1, 1 \leq i \leq r,$$

where  $p_{0j} = p_0$  for  $1 \leq j \leq r$ .

PROOF. The equilibrium microstate equations for the continuous-time Markov chain  $\{X(t)\}$  are given by

$$(4.9) \quad \lambda_0 p_0 = \mu p_{11}$$

$$(4.10) \quad (\mu + \lambda_1) p_{1i} = \mu p_{1,i+1} + \mu q_i p_{21} + \lambda_0 q_i p_0, \quad 1 \leq i < r,$$

$$(4.11) \quad (\mu + \lambda_1) p_{1r} = \mu q_r p_{21} + \lambda_0 q_r p_0,$$

$$(4.12) \quad (\mu + \lambda_n) p_{ni} = \mu p_{n,i+1} + \mu q_i p_{n+1,1} + \lambda_{n-1} p_{n-1,i}, \quad n \geq 2, 1 \leq i < r,$$

$$(4.13) \quad (\mu + \lambda_n) p_{nr} = \mu q_r p_{n+1,1} + \lambda_{n-1} p_{n-1,r}, \quad n \geq 2.$$

These equations equate for each microstate  $(n,i)$  the rate at which the system leaves this state to the rate of which the system enters this state. By summing the microstate equations over  $i$  for fixed  $n$  and using induction on  $n$ , we readily obtain (4.7). The relation (4.7) can be directly obtained by aggregating the states  $(n,1), \dots, (n,r)$  to a macrostate and equating the rate at which the system enters this macrostate to the rate at which the system leaves this macrostate, cf. also Cooper [3] and Kühn [9]. Finally, we get (4.8) from (4.10)-(4.11) and (4.12)-(4.13) by repeated substitution.

It now follows that once we have computed the probabilities  $p_n, n \geq 0$  from (2.6)-(2.7) we can compute the probabilities  $p_{ni}$  and so  $\pi_{ni}$  ( $1 \leq i \leq r$ ) for  $n=1,2,\dots$  as follows. We first compute  $p_{n+1,1}$  from (4.7) and next we compute the probabilities  $p_{ni}$  for  $1 \leq i \leq r$  from (4.8) using the previously computed values of  $p_{n-1,j}$  ( $1 \leq j \leq r$ ). Note that this computational scheme based on (2.6)-(2.7) and (4.7)-(4.8) is numerically stable as opposed to the alternative computational scheme computing recursively both the  $p_{ni}$  ( $1 \leq i \leq r$ ) and  $p_n$  directly from (4.7)-(4.8) with the boundary probability  $p_0$  as parameter but involving the evaluation of differences which may cause a loss of accuracy during the computations.

We next consider the computation of the conditional waiting times  $W(t|n,i)$ . By the phase representation (4.1) of the service time, we have under the assumption of service in order of arrival that for all  $n \geq 1$  and  $1 \leq i \leq r$

$$(4.14) \quad W(t|n,i) = \sum_{k=i+n-1}^{\infty} q_{k-i}^{(n-1)} E_{k,u}(t), \quad t \geq 0$$

where  $\{q_j^{(n)}\}$  denotes the  $n$ -fold convolution of  $\{q_j\}$  with itself, i.e.

$$(4.15) \quad q_j^{(n)} = \sum_{h=1}^r q_h q_{j-h}^{(n-1)}, \quad n \geq 1, j \geq n$$

with  $q_0^{(0)} = 1$  and  $q_j^{(0)} = 0, j \neq 0$ . Noting that the cumulative Poisson probabilities  $1 - E_{k,u}(t)$  can be very efficiently evaluated, we can compute the exact value of  $W(t)$  from (4.4) and (4.14)-(4.15). For the special case of the infinite capacity  $M/G/1$  queue (i.e.  $\lambda_j = \lambda$  for all  $j \geq 0$ ) we can obtain  $W(t)$  in a simpler way by directly determining the steady-state distribution of the number of uncompleted phases in the system, see Bux [2] and see also the relations (4.17)-(4.18) below (note that for the  $M/G/1$  queue another computational method for  $W(t)$  is to numerically solve the simple integro-differential equation of Takács [16]). Further we remark that a simple and accurate approximation for the waiting time distribution may be obtained



by fitting a Weibull distribution function to  $1 - \Pr\{W > t | W > 0\}$  by matching the first two moments which are easily evaluated from (4.4).

We conclude this section by giving a very simple computational method for the state probabilities and the waiting times for the special case where  $q_r = 1$ , i.e. the service time distribution function is the Erlang distribution function

$$F(t) = E_{r, \mu}(t), \quad t \geq 0.$$

Hence each customer represents  $r$  independent phases having each an exponential service time with mean  $1/\mu$ . Define now  $Y(t)$  as the number of uncompleted phases in the system at time  $t$  and let for the continuous-time Markov chain  $\{Y(t), t \geq 0\}$

$$f_j = \lim_{t \rightarrow \infty} \Pr\{Y(t) = j\}, \quad j \geq 0.$$

Then, since the number of uncompleted phases uniquely determines the number of customers present,

$$(4.16) \quad p_0 = f_0, \quad p_n = \frac{nr}{\sum_{j=(n-1)r+1}^{nr}} f_j \quad \text{for } n \geq 1.$$

Put for abbreviation

$$\gamma_j = \begin{cases} \lambda_{\lceil j/r \rceil} & \text{if } j/r \text{ is an integer} \\ \lambda_{\lceil j/r \rceil + 1} & \text{otherwise,} \end{cases}$$

where  $\lceil x \rceil$  is the largest integer less than or equal to  $x$ . We have

$$(4.17) \quad W(t) = \eta_0 + \sum_{n=1}^{\infty} \eta_n E_{n, \mu}(t), \quad t \geq 0 \quad \text{with } \eta_n = \gamma_n f_n / \sum_{j=0}^{\infty} \lambda_j p_j, \quad n \geq 0,$$

by noting that  $\eta_n$  is the steady-state probability that an arriving customer finds upon arrival  $n$  phases in the system. Finally, we find

$$(4.18) \quad \mu f_n = \sum_{k=n-r}^{n-1} \gamma_k f_k, \quad n \geq 1$$

where  $f_j = 0$  for  $j < 0$ , as may be seen by aggregating the set of states  $\{n, n+1, \dots\}$  of the continuous-time Markov chain  $\{Y(t)\}$  into a macrostate and equating the rate at which the system leaves this macrostate to the rate at which the system enters this macrostate. From (4.18) we can recursively compute the probabilities  $f_n$ ,  $n \geq 1$  by taking  $f_0$  as parameter.

5. *Some numerical results.*

As illustration consider the M/G/1 queue with finite capacity  $K$  (cf. also Kühn [7] and Lavenberg [9]) where customers arrive according to a Poisson process with rate  $\lambda$  and that customers who find upon arrival  $K$  other customers in the system do not enter and have no effect on the system. The finite capacity M/G/1 queue is a special case of the single-server queue having state dependent arrival rate  $\lambda_j$  with  $\lambda_j = \lambda$  for  $0 \leq j \leq K-1$  and  $\lambda_j = 0$  for  $j \geq K$ .

Denote by  $c = (ES^2 / (ES)^2 - 1)^{1/2}$  the coefficient of variation of the service time  $S$  and denote by  $\rho = \lambda ES$  the offered load. We consider the following two examples where the service time  $S$  has a phase-type distribution function as in (4.1),

$$K=15, q_1=(2/7)(3-\sqrt{2}), q_2=1-q_1, \mu=1+q_2 \quad (c^2=0.75)$$

and

$$K=15, q_2=5/6, q_5=1/6, \mu=2.5 \quad (c^2=3).$$

Both for  $\rho=0.8$  and  $\rho=1.5$  we give in table 5.1 for these two examples the values of the probability  $\sum_{j>n} \pi_j$  that an arbitrary *entering* customer sees upon arrival more than  $n$  other customers in the system and the values of the complementary waiting time distribution function  $\Pr\{W>tES|W>0\}$  for the delayed entering customers. The number between brackets gives the difference between the Weibull approximation of  $\Pr\{W>tES|W>0\}$  and the exact value. Note that the Weibull approximation is quite satisfactory. Also note from table 5.1 that the complementary waiting time probabilities are not necessarily increasing in the coefficient of variation of the service time. Finally, for the four cases considered in table 5.1 the values of the steady-state probability  $p_K$  that an arbitrary arriving customer finds upon arrival the system full are given by 0.00468, 0.00345, 0.33356 and 0.33344 respectively. The computing time per example was a few seconds on the CDC 7300.

		Pr{W>tES   W>0}		$\sum_{j>n} \pi_j$		
$\rho$	t	$c^2=0.75$	$c^2=3$	n	$c^2=.75$	$c^2=3$
0.8	0	1.0000( 0 )	1.0000( 0 )	0	.79529	.79653
	.25	.94821(+0.0105)	.94572(+0.0083)	1	.61826	.60861
	.50	.89682(+0.0151)	.88985(+0.0131)	2	.47591	.45966
	.75	.84683(+0.0176)	.83570(+0.0159)	3	.36414	.34511
	1.0	.79875(+0.0188)	.78419(+0.0171)	4	.27710	.25775
	1.5	.70920(+0.0182)	.68944(+0.0163)	5	.20953	.19136
	2	.62868(+0.0152)	.60512(+0.0132)	6	.15712	.14096
	3	.49251(+0.0063)	.46443(+0.0050)	7	.11650	.10272
	4	.38437(-0.0023)	.35507(-0.0027)	8	.08500	.07372
	5	.29862(-0.0085)	.27030(-0.0081)	9	.06059	.05173
	8	.13406(-0.0119)	.11424(-0.0103)	10	.04167	.03504
	10	.07387(-0.0069)	.06024(-0.0055)	11	.02701	.02239
	12	.03753(-0.0014)	.02889(-0.0005)	12	.01564	.01280
	14	.01715(+0.0021)	.01216(+0.0024)	13	.00683	.00552
	15	.01108(+0.0029)	.00745(+0.0029)	14	0	0
1.5	0	1.0000( 0 )	1.0000( 0 )	0	.99948	.99975
	3	.99348(+0.0028)	.99630(+0.0020)	1	.99860	.99928
	5	.97702(-0.0014)	.98584(-0.0007)	2	.99719	.99846
	6	.95911(-0.0065)	.97349(-0.0051)	3	.99492	.99709
	7	.92945(-0.0120)	.95133(-0.0109)	4	.99132	.99478
	8	.88325(-0.0152)	.91334(-0.0154)	5	.98560	.99092
	9	.81691(-0.0137)	.85324(-0.0152)	6	.97649	.98445
	10	.73012(-0.0070)	.76768(-0.0081)	7	.96202	.97360
	11	.62689(+0.0032)	.65918(+0.0041)	8	.93901	.95541
	12	.51503(+0.0133)	.53666(+0.0167)	9	.90244	.92492
	13	.40399(+0.0196)	.41268(+0.0241)	10	.84430	.87380
	14	.30236(+0.0203)	.29932(+0.0234)	11	.75187	.78811
	15	.21604(+0.0158)	.20485(+0.0157)	12	.60493	.64447
	17	.09650(+0.0011)	.08113(-0.0036)	13	.37134	.40367
	20	.02125(-0.0071)	.01379(-0.0075)	14	0	0

Table 5.1 Numerical results

*References*

1. Bux, W. and Herzog, U., 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, 1977, pp. 23-38.
2. Bux, W., Single server queues with general interarrival and phase-type service time distributions-computational algorithms, Proceedings of the 9<sup>th</sup> International Teletraffic Congress (Torremolinos), 1979.
3. Cooper, R.B., Introduction to Queueing Theory, MacMillan, New York, 1972.
4. Hordijk, A. and Tijms, H.C., A simple proof of the equivalence of the limiting distributions of the continuous-time and the embedded process of the queue size in the M/G/1 queue, Statist. Neerlandica 30, 1976, pp. 97-100.
5. Kampe, G. and Kühn, P.J., Graded delay systems with infinite or finite source traffic and exponential or constant holding time, Proceedings of the 8<sup>th</sup> International Teletraffic Congress (Melbourne), 1976, pp. 256-1/10.
6. Kobayashi, H., Modeling and Analysis: An Introduction to System Performance and Evaluation Methodology, Addison-Wesley, Reading, 1976.
7. Kühn, P.J., On the Calculation of Waiting Times in Switching and Computer Systems, 15<sup>th</sup> Report on Studies in Congestion Theory, Institute of Switching and Data Technics, University of Stuttgart, 1972.
8. Kühn, P.J., Tables on Delay Systems, Institute of Switching and Data Technics, University of Stuttgart, 1976.
9. Lavenberg, S.S., The steady-state queueing time distribution for the M/G/1 finite capacity queue. Manag. Sci. 25, 1975, pp. 501-506.
10. Neuts, M.F., Algorithms for the waiting time distributions under various queue disciplines in the M/G/1 queue with service time distribution of phase-type, in: M.F. Neuts (ed.), Algorithmic Methods in Probability Theory, North-Holland, Amsterdam, 1977.
11. Riordan, J., Stochastic Service Systems, Wiley, New York, 1962.
12. Ross, S.M., Applied Probability Models with Optimization applications, Holden-Day, San Francisco, 1970.
13. Schassberger, R., Warteschlangen, Springer-Verlag, Berlin, 1973.
14. Stidham, S. Jr., Regenerative processes in the theory of queues with applications to the alternating priority queue, Adv. Appl. Prob. 4, 1972, pp. 542-577.

15. Stidham, S. Jr., A last word on  $L=\lambda W$ , Operations Res., 22, 1972, pp. 417-421.
16. Takács, L., An Introduction to the Theory of Queues, Oxford University Press, New York, 1962.
17. Tijms, H.C., Van Hoorn, M.H. and Federgruen, A., Approximations for the steady-state probabilities in the M/G/c queue, Research Report 49, Vrije Universiteit, Amsterdam (to appear in Adv. Appl. Prob.).

