

ET 9

05348.-

1988

# SERIE RESEARCH MEMORANDA

ALGORITHMS AND APPROXIMATIONS FOR  
BATCH-ARRIVAL QUEUES

H.C. Tijms

Researchmemorandum 1988-9

Febr. '88



VRIJE UNIVERSITEIT  
FACULTEIT DER ECONOMISCHE WETENSCHAPPEN  
EN ECONOMETRIE  
AMSTERDAM

# ALGORITHMS AND APPROXIMATIONS FOR BATCH-ARRIVAL QUEUES

Henk C. Tijms

Operations Research Group  
Vrije Universiteit  
Amsterdam, The Netherlands

Batch-arrival queues are common in practice, but analytical tools for practical analysis are hardly known. The purpose of this paper is to present several computational methods that are easy to apply and give sufficiently accurate results.

## 1. INTRODUCTION

Queuing systems with batch arrivals rather than single arrivals have many applications in practice. A typical example is the analysis of message packetization in data communication systems, cf. Manfield and Tran-Gia [1]. Unlike single-arrival queueing systems for which many algorithms are available (cf. Tijms [2]), practically useful computational tools for batch-arrival models seem hardly known. In this paper, we present some algorithmic approaches that are generally applicable and sufficiently simple for practical use. We discuss both exact methods based on the concept of phase-type distributions and approximations based on interpolation of solutions for simpler models. In section 2 we consider the  $GI^X/G/1$  queue. Noting that any probability distribution can be arbitrarily closely approximated by a mixture of Erlangian distributions, we give a tractable exact method for computing the waiting-time probabilities when the service-time density is a mixture of Erlangians. Also, we present a simple approximation for the mean waiting time and the delay probability in the general  $GI^X/G/1$  queue. These approximations use the service time only through its first two moments but use the actual interarrival-time distribution. The latter is important because the characteristics of the input process have a much larger effect on the performance measures than the characteristics of the service process. Section 3 considers the finite capacity  $GI^X/G/1/N$  queue and gives a simple approximation method for the computation of the minimal buffer size for which the rejection probability of a customer is below a prespecified level.

## 2. THE $GI^X/G/1$ QUEUE

In subsection 2.1 we present an exact algorithm for the computation of the waiting-time probabilities when the service time is a mixture of Erlangians. This algorithm is based on the embedded Markov-chain method and exploits the fact that the state probabilities have a geometric tail. In subsection 2.2 we give approximations for the mean waiting time and the delay probability. These approximations use interpolation and easily computed exact results for the single-arrival  $GI/C_2/1$  queue with Coxian-2 service.

### 2.1. The $GI^X/E_{1,\dots,r}/1$ Queue

A fundamental and very useful result is that any probability distribution of a positive random variable can be arbitrarily closely approximated by a mixture of Erlangian distributions with density

$$(2.1) \quad b(x) = \sum_{i=1}^r q_i \mu^i \frac{x^{i-1}}{(i-1)!} e^{-\mu x}, \quad x \geq 0,$$

where  $q_i \geq 0$  and  $\sum_i q_i = 1$ . It is important that in this representation each Erlangian distribution has the same scale parameter.

Consider now the GI<sup>x</sup>/G/1 queue in which the service time of each customer has (2.1) as density. Batches of customers arrive according to a renewal process with a general interarrival-time density  $a(x)$ . The batch size has a general discrete probability distribution  $(\alpha_i, i \geq 1)$ . Denoting by the generic variables  $S$ ,  $A$  and  $B$  the service time of a customer, the interarrival time between batches and the batch size, we assume that the traffic intensity  $\rho = E(B)E(S)/E(A)$  is less than 1. Service is in order of arrival for customers belonging to different batches, while customers from a same batch are served according to their (random) positions in the batch. Denoting by  $W_n$  the delay in queue of the  $n^{\text{th}}$  served customer, we define

$$(2.2) \quad W_q(x) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n P(W_k \leq x), \quad x \geq 0,$$

as the long-run fraction of customers having a delay in queue of no more than  $x$ . The ordinary limit  $\lim_{n \rightarrow \infty} P(W_n \leq x)$  exists only if the batch size is aperiodic (i.e. there is no integer  $d \geq 2$  with  $\sum_n \alpha_{nd} = 1$ ). In general, the ordinary limit need not exist (e.g. take a constant batch size of 2, then the customers second in the batch have always to wait, while the other ones have a positive probability of no delay). In the sequel we use the following result from renewal theory. The long-run fraction of customers taking the  $j^{\text{th}}$  position in their batch is given by

$$(2.3) \quad \eta_j = \frac{1}{E(B)} \sum_{k=j}^{\infty} \alpha_k, \quad j=1,2,\dots$$

By (2.1), we have the useful representation that, with probability  $q_i$ , the service time of a customer is the sum of  $i$  independent phases each having an exponential distribution with the same mean  $1/\mu$ . Following Bux [3], we consider the embedded Markov chain  $(X_n)$  with  $X_n$  defined as the number of uncompleted service phases present just before the arrival of the  $n^{\text{th}}$  batch. This Markov chain is aperiodic and positive recurrent and hence has a limiting distribution  $(\pi_j, j \geq 0)$ . The steady-state probabilities  $\pi_j$  are the unique solution to the linear equations,

$$(2.4) \quad \pi_j = \sum_{i=0}^{\infty} p_{ij} \pi_i, \quad j=0,1,\dots$$

together with the normalization equation  $\sum_j \pi_j = 1$ , where the  $p_{ij}$ 's are the one-step transition probabilities of the Markov chain. To find  $W_q(x)$ , we need the quantity  $\pi_j^*$  defined as the long-run fraction of customers having  $j$  uncompleted service phases in front of them just after arrival of their batch,  $j \geq 0$ . Then,

$$(2.5) \quad 1 - W_q(x) = \sum_{k=0}^{\infty} e^{-\mu x} \frac{(\mu x)^k}{k!} \left[ 1 - \sum_{j=0}^k \pi_j^* \right], \quad x \geq 0,$$

using that  $\sum_{k=0}^{j-1} e^{-\mu x} (\mu x)^k / k!$  is the conditional probability that a customer having  $j$  phases in front of him must wait more than  $x$ . Clearly,

$$(2.6) \quad \pi_j^* = \sum_{k=0}^j \pi_k \sum_{m=1}^{j-k+1} \eta_m q_{j-k}^{(m-1)*}, \quad j=1,2,\dots$$

where  $(q_i^{n*})$  is the  $n$ -fold convolution of  $(q_j)$  with itself, i.e.  $q_i^{n*}$  is the probability that  $n$  customers represent a total of  $i$  phases. Also, define  $\gamma_\ell$  as the probability that one batch consists of a total of  $\ell$  phases. Then

$$(2.7) \quad \gamma_\ell = \sum_{k=1}^{\infty} \alpha_k q_\ell^{k*}, \quad \ell=1,2,\dots$$

The one-step transition probabilities  $p_{ij}$  of the Markov chain  $(X_n)$  are

$$(2.8) \quad p_{ij} = \sum_{\ell=1}^{\infty} \gamma_{\ell} a_{i+\ell-j} \quad \text{for } j \geq 1,$$

and  $p_{i0} = 1 - \sum_{j=0}^{\infty} p_{ij}$ , where  $a_k = \int_0^{\infty} e^{-\mu t} (1/k!) (\mu t)^k a(t) dt$ ,  $k \geq 0$ . From a computational point of view it is important to note that the probabilities  $p_{ij}$  for  $j \neq 0$  depend on  $i$  and  $j$  only through  $i-j$ . To solve numerically the infinite system of linear equations (2.4), the usual way is to reduce it to a finite system by a truncation integer  $M$  such that  $\sum_{i>M} \pi_i$  is sufficiently small. For the case of nonlight traffic, the resulting system may still be very large. However, for that case a much smaller system can be obtained by exploiting the fact the distribution  $\{\pi_j\}$  has a geometric tail. Therefore we need the mild assumption that the batch-size distribution  $\{\alpha_k\}$  has no extremely long tail. To be more precise, we assume that the power series  $A(x) = \sum_{k=1}^{\infty} \alpha_k x^k$  has a convergence radius  $R > 1$  and that  $A(x) \rightarrow \infty$  as  $x \rightarrow R$ . Then, by deep results from Markov chain theory,

$$(2.9) \quad \pi_j \sim \alpha \tau^j \quad \text{for } j \text{ large enough}$$

for constants  $\alpha > 0$  and  $0 < \tau < 1$ . For nonlight traffic this asymptotic expansion turns out to apply already for relatively small values of  $j$ . Thus, by choosing an appropriate integer  $N$ , we reduce the infinite system (2.4) to a finite one by replacing  $\pi_i$  by  $\pi_N \tau^{i-N}$  for  $i > N$ . The constant  $\tau$  can be computed on beforehand. Substituting (2.9) into (2.4) yields that  $\tau$  satisfies the equation

$$(2.10) \quad G\left(\frac{1}{\tau}\right) a^*(\mu(1-\tau)) = 1,$$

where

$$(2.11) \quad G(z) = \sum_{\ell=1}^{\infty} \gamma_{\ell} z^{\ell} \quad \text{and} \quad a^*(s) = \int_0^{\infty} e^{-st} a(t) dt.$$

By (2.7),  $G(z) = \sum_{k=1}^{\infty} \alpha_k [Q(z)]^k$  with  $Q(z) = \sum_{i=1}^{\infty} q_i z^i$ . Note that  $G(z)$  has a convergence radius  $R' = R^{1/\tau}$ . The equation (2.10) has a unique solution on  $(1/R', 1)$ . This follows since  $F(1) = 1$ ,  $F(x) \rightarrow \infty$  as  $x \rightarrow 1/R'$ ,  $F'(1) < 0$  and  $F(x)$  is convex on  $(1/R', 1)$  with  $F(x) = G(1/x) a^*(\mu(1-x))$ . In fact  $F(x)$  is the generating function of the probabilities in the right hand side of (2.8). We remark that, by (2.6) and (2.9),  $\pi_j \sim \alpha_0 \tau^j$  for  $j$  large for some constant  $\alpha_0 > 0$ . Substituting this expansion into (2.5) yields

$$(2.12) \quad 1 - W_q(x) \sim \delta e^{-\mu(1-\tau)x} \quad \text{for } x \text{ large enough}$$

for some constant  $\delta > 0$ , in agreement with results in Van Ommeren [4]. To conclude this subsection, we note that for the particular case of Poisson arrivals of the batches the probabilities  $\pi_j$  can be recursively computed. Then, by the property Poisson arrivals see time averages,  $\pi_j$  equals the time-average probability of having  $j$  uncompleted service phases present. By equating the rate of which the system enters the macrostate of having at least  $j$  phases present to the rate at which the system enters that macrostate, we get the recursion scheme

$$(2.13) \quad \mu \pi_j = \sum_{i=0}^{j-1} \pi_i \lambda \cdot \sum_{\ell \geq j-i} \gamma_{\ell}, \quad j=1, 2, \dots,$$

starting with  $\pi_0 = 1 - \rho$ , where  $\lambda$  is the average arrival rate of batches. For the  $M^X/G/1$  queue with general service times, explicit expressions can be given for the average waiting time per customer ( $E(W_q)$ ) and the fraction of customers that is delayed ( $\Pi_W$ ),

$$(2.14) \quad E(W_q) = \frac{1}{2} (1 + c_S^2) \frac{\rho E(S)}{1 - \rho} + \frac{E(S)}{2(1 - \rho)} \left( \frac{E(B^2)}{E(B)} - 1 \right), \quad \Pi_W = 1 - \frac{(1 - \rho)}{E(B)},$$

where  $c_S^2 = \sigma^2(S)/E^2(S)$  denotes the squared coefficient of variation of the ser-

vice, time  $S$  of a customer. Also, for the  $M^X/G/1$  queue the coefficients  $\varphi$  and  $\theta$  of the asymptotic expansion  $1-W_q(x) \sim \varphi e^{-\theta x}$  for  $x$  large can be rather easily computed. Moreover, using this result, the result (2.14) and explicit results for the second moment of  $W_q(x)$  and the derivative of  $W_q(x)$  at  $x=0$ , the waiting-time distribution function  $1-W_q(x)$  may be approximated by a sum of three exponential functions for all  $x \geq 0$ , see Van Ommeren [4,5]. In many practical situations this approximation exists and gives very accurate results. An alternative approximation to  $1-W_q(x)$  for larger values of  $x$  is provided by approximating the higher percentiles of the waiting time by a linear interpolation of the asymptotic waiting-time percentiles for the special cases of deterministic and exponential services. The interpolation is based on the squared coefficient of variation of the service time. In the more general context of the multi-server  $M^X/G/c$  queue this useful approach is discussed in Eikeboom and Tijms [6]. This reference gives the details of the computation of the coefficients  $\varphi$  and  $\theta$  of the asymptotic expansion  $1-W_q(x) \sim \varphi e^{-\theta x}$  for  $x$  large for the particular cases of  $M^X/M/c$  queue and the  $M^X/D/c$  queue.

## 2.2. Approximations To The Average Delay And The Delay Probability

Before we can state the approximations, we need some preparatory results. A very useful probability distribution in queueing analysis is the Coxian-2 ( $C_2$ ) distribution. A positive random variable  $U$  is said to be  $C_2$ -distributed when  $U$  can be represented as

$$(2.15) \quad U = \begin{cases} U_1 + U_2 & \text{with probability } b \\ U_1 & \text{with probability } 1-b, \end{cases}$$

where  $U_1$  and  $U_2$  are two independent exponentials with respective means  $1/\mu_1$  and  $1/\mu_2$ . Equivalently, a probability density of a positive random variable is a  $C_2$ -density when its Laplace transform is the ratio of a polynomial of at most degree 1 to a polynomial of degree 2. A  $C_2$ -distributed random variable  $U$  has always a squared coefficient of variation  $c_u^2 \geq \frac{1}{2}$ . It is often convenient to fit a  $C_2$ -distribution to a positive random variable by matching its first two or its first three moments. Let  $X$  be a positive random variable with  $c_x^2 \geq \frac{1}{2}$  and  $m_i = E(X^i)$  denoting the  $i^{\text{th}}$  moment of  $X$ . If a three-moment fit to  $X$  by a  $C_2$ -distribution exists, the three parameters of this fit are given by

$$(2.16) \quad \mu_{1,2} = \frac{1}{2}(a_1 \pm \sqrt{a_1^2 - 4a_2}), \quad b = \frac{\mu_2}{\mu_1} (\mu_1 m_1 - 1),$$

where  $a_1 = 1/m_1 + \frac{1}{2}m_2/m_1$  and  $a_2 = (6m_1 - 3m_2/m_1) / [(3m_2^2/2m_1) - m_3]$ , see Van der Heijden [7]. An infinite number of two-moment fits to  $X$  by a  $C_2$ -distribution are always possible. A particularly useful two-moment fit is the one with the parameters

$$(2.17) \quad \mu_{1,2} = \frac{2}{m_1} \left( 1 \pm \sqrt{\frac{c_x^2 - \frac{1}{2}}{c_x^2 + 1}} \right), \quad b = \frac{\mu_2}{\mu_1} (\mu_1 m_1 - 1).$$

This  $C_2$ -distribution has the same first three moments as a gamma distribution. For the single-arrival  $GI/C_2/1$  queue the waiting-time distribution allows for a tractable analytical solution. Denote by  $E(W_{q1})$  and  $\Pi_{w1}$  the mean delay in queue and the probability of wait of a customer. Then, by general results in Cohen [8],

$$(2.18) \quad E(W_{q1}) = -\frac{(\mu_1 + \mu_2)}{\mu_1 \mu_2} + \frac{1}{\theta_1} + \frac{1}{\theta_2}, \quad \Pi_{w1} = 1 - \frac{\theta_1 \theta_2}{\mu_1 \mu_2},$$

where  $0 < \theta_1 < \min(\mu_1, \mu_2) \leq \theta_2$  are the two real roots of the equation

$$(2.19) \quad x^2 - (\mu_1 + \mu_2)x + \mu_1\mu_2 - (\mu_1\mu_2 - (1-b)\mu_1x) \int_0^{\infty} e^{-xt} a(t) dt = 0.$$

Here  $b$ ,  $\mu_1$  and  $\mu_2$  are the parameters of the Coxian-2 service. We now return to the batch-arrival  $GI^X/G/1$  queue with a general distribution for the service time  $S$  of a customer. Denote by  $E(W_q)$  and  $\Pi_w$  the long-run average delay per customer and the long-run fraction of customers that are delayed. The delay experienced by a customer consists of two independent components. The first component is the delay until the first member of his batch is served and the second component is the delay due to the service times of the members of his batch served before him. By (2.3), the average value of the second component equals  $\frac{1}{2}(E(B^2)/E(B)-1)E(S)$ . To find the average value of the first component, take the whole batch as a "super-customer" and consider the resulting single-arrival  $GI/G_x/1$  queue in which the interarrival times are the same as the times between the arrivals of a batch and the service time is the total time  $T$  to serve a whole batch. Denote by  $E(W_{q1})$  and  $\Pi_{w1}$  the mean delay and the delay probability for the latter queueing system. Then (cf. also Burke [9]),

$$(2.20) \quad E(W_q) = E(W_{q1}) + \frac{1}{2} \left( \frac{E(B^2)}{E(B)} - 1 \right) E(S), \quad \Pi_w = 1 - \frac{(1 - \Pi_{w1})}{E(B)}.$$

It remains to compute (approximately)  $E(W_{q1})$  and  $\Pi_{w1}$ . Therefore we use interpolation and the easily computed exact results (2.18) for the  $GI/C_2/1$  queue. The first three moments of the total time  $T$  needed to serve all customers from one batch are given by

$$E(T) = E(B)r_1, \quad E(T^2) = E(B)\bar{r}_2 + E(B^2)r_1^2, \quad E(T^3) = E(B)\bar{r}_3 + 3E(B^2)\bar{r}_2r_1 + E(B^3)r_1^3.$$

where  $r_i$  is the  $i^{\text{th}}$  moment of the service time  $S$  of a customer,  $\bar{r}_2 = r_2 - r_1^2$  and  $\bar{r}_3 = r_3 - 3r_1r_2 + 2r_1^3$ . In particular,  $c_T^2 = c_B^2 + c_S^2/E(B)$ . It is interesting to note that  $T$  has a  $C_2$ -distribution when the service time  $S$  of a customer has a  $C_2$ -distribution and the batch size  $B$  has a geometric distribution (this result follows by considering  $E(e^{-sT})$ ). To compute approximately  $E(W_{q1})$  and  $\Pi_{w1}$ , we suggest to proceed as follows. For the case of  $c_T^2 \geq \frac{1}{2}$ , fit a  $C_2$ -distribution to  $T$  by using the three-moment fit (2.16) or the two-moment fit (2.17) and next compute  $E(W_{q1})$  and  $\Pi_{w1}$  from (2.18). For the case of  $c_T^2 < \frac{1}{2}$ , we use (2.18) to compute the mean delay and the delay probability for several  $GI/C_2/1$  queues each having as interarrival times the times between the arrivals of batches and each having the same mean service time  $E(T)$ . Next we apply extrapolation based on the squared coefficient of variation of the service time to get approximations to  $E(W_{q1})$  and  $\Pi_{w1}$ . Denoting by  $P_i$  the value of the performance measure in the  $GI/C_2/1$  queue with squared coefficient of variation  $c_i^2$  for the service time and supposing that  $P_i$  is known for  $n$  different values of  $c_i^2$ , the value of the performance measure  $P$  for the  $GI/G/1$  queue with squared coefficient of variation  $c^2 = c_T^2$  for the service time can be approximated by

$$(2.21) \quad P(c^2) = \sum_{i=1}^n P_i \prod_{k=1}^n \frac{c^2 - c_k^2}{c_i^2 - c_k^2}.$$

The above approximations have the feature that the actual interarrival time distribution is used and thus should in general perform much better than the approximations in Krämer and Langenbach-Belz [10] using only the first two moments. In general it is hazardous to use the interarrival time only through its first two moments, see also Tijms [2]. In table 2.1 we give some numerical results for several  $D^X/E_k/1$  queues, where the batch size is constant or geometrically distributed with mean 2 and the mean service time of a customer is normalized as 1. The approximate results use the two-moment fit (2.17) when  $c_T^2 \geq \frac{1}{2}$  and use (2.21) with  $n=3$  and  $c_i^2 = 1/2, 3/4, 1$  when  $c_T^2 < \frac{1}{2}$ .

Table 2.1. Numerical results for the  $D^X/E_k/1$  queue.

		$E(W_q)$	$\Pi_w$	$E(W_q)$	$\Pi_w$	$E(W_q)$	$\Pi_w$
constant		$k=3, c_T^2=1/6$		$k=2, c_T^2=1/4$		$k=1, c_T^2=1/2$	
$\rho=0.5$	exa	0.511	0.511	0.535	0.524	0.657	0.558
	app	0.479	0.517	0.519	0.528	0.657	0.558
$\rho=0.8$	exa	0.870	0.689	1.137	0.723	2.014	0.774
	app	0.836	0.725	1.125	0.739	2.014	0.774
geometric		$k=3, c_T^2=2/3$		$k=2, c_T^2=3/4$		$k=1, c_T^2=1$	
$\rho=0.5$	exa	1.287	0.575	1.337	0.582	1.510	0.602
	app	1.263	0.575	1.321	0.583	1.510	0.602
$\rho=0.8$	exa	3.162	0.786	3.460	0.795	4.385	0.814
	app	3.122	0.792	3.434	0.799	4.385	0.814

### 3. MINIMAL BUFFER SIZE FOR THE $GI^X/G/1/N$ QUEUE

This section considers the  $GI^X/G/1/N$  queueing system with a finite buffer having only capacity for  $N$  customers (including any customer in service). If an arriving batch contains more customers than the remaining capacity of the buffer, the batch is rejected in its whole. The goal is to find the minimal buffer size such that the long-run fraction of customers that are rejected is below a prespecified level. An exact analysis for this design problem is possible when the service time of a customer has a pure Erlangian distribution. Then the analysis of section 2.1 can be modified to obtain the limiting distribution of the number of customers in the buffer. Only for pure Erlangian service the number of uncompleted service phases present determines unambiguously the number of customers present. The exact results for the  $GI^X/E_k/1/N$  queue can be used to approximate the minimal buffer size for the  $GI^X/G/1/N$  queue. As in section 2.2, this is done by interpolation with respect to the squared coefficient of variation of the service time. To be specific, define for the  $GI^X/G/1/N$  queue with a generally distributed service time  $S$  of a customer,  $\nu(\beta)$  as the smallest value of  $N$  such that the long-run fraction of customers that are rejected does not exceed a prespecified level  $\beta$ . Here  $\beta$  is typically a very small number. Denote by  $\nu_k(\beta)$  the minimal buffer size for the same queueing system except that the service time of a customer has now an Erlang- $k$  distribution with the same mean  $E(S)$  as before. Suppose that the exact value of  $\nu_k(\beta)$  has been calculated for several values of  $k$  (say,  $k \in K$ ). Then,  $\nu(\beta)$  can be approximated by the first integer larger or equal to

$$(3.1) \quad \nu_{\text{app}}(\beta) = \sum_{i \in K} \nu_i(\beta) \prod_{\substack{k \in K \\ k \neq i}} \frac{c_S^2 - 1/k}{1/i - 1/k}$$

provided  $c_S^2$  is not too large. The usefulness of this approximation approach has been demonstrated in Tijms [2] for a variety of finite-buffer queueing models. Next we show how to calculate  $\nu_k(\beta)$ . This will be done only for the case of Poisson arrivals of batches, see Nobel [11], for the general case. Consider the  $M^X/E_r/1/N$  queue. Define  $(\pi_j, j \geq 0)$  as the limiting distribution of the number of uncompleted service phases present just prior to the arrival of a batch. Then, the limiting probability of having  $i$  customers present just before the arrival of a batch is given by

$$(3.2) \quad \bar{\pi}_i = \frac{ir}{\sum_{j=(i-1)r+1}^i \pi_j}, \quad i=0,1,\dots,N.$$

Thus, using that the long-run fraction of customers belonging to a batch of size  $i$  equals  $i\alpha_i/E(B)$  (cf. [9]), it follows that the long-run fraction of customers that are rejected is given by

$$(3.3) \quad \pi_{rej}(N) = \sum_{k=0}^N \bar{\pi}_k \left(1 - \frac{1}{E(B)} \sum_{i=1}^{N-k} i\alpha_i\right).$$

Obviously,  $\nu_r(\beta)$  is the smallest  $N$  with  $\pi_{rej}(N) \leq \beta$ . To find the  $\pi_j$ 's, we invoke the property that Poisson arrivals see time averages. That is, the limiting probability  $\pi_j$  is equal to the limiting probability of having  $j$  uncompleted phases present at an arbitrary time. By equating the rate at which the system leaves the macrostate of having at least  $j$  uncompleted phases present to the rate at which the system enters that state, we get

$$(3.4) \quad \mu\pi_j = \sum_{i=0}^{j-1} \pi_i \lambda \sum_{k \geq (j-i)/r}^{[(Nr-i)/r]} \alpha_k, \quad j=1, 2, \dots, Nr,$$

where  $\lambda$  is the average arrival rate of batches and  $\mu$  is the average service rate of a phase (i.e.  $r/\mu$  is the mean service time of a customer). The probabilities  $\pi_j$  can be recursively computed from (3.4) together with  $\sum_{j=0}^{Nr} \pi_j = 1$ . In table 3.1 we give some numerical results for the  $M^X/E_{10}/1/N$  queue, where the batch size is constant or geometrically distributed with  $E(B)=2$ . For several values of  $\beta$ , we give both the exact and approximate values of  $\nu(\beta)$ , where the approximate value is obtained by extrapolation of the exact results for  $E_1$  and  $E_2$  services. Also, we include in table 3.1 the exact values of  $\nu_1(\beta)$  for exponential service to show that in general  $\nu_1(\beta)$  cannot be used as first-order approximation to  $\nu(\beta)$ , cf. also [1]. Note from the results in table 3.1 that  $\nu(\beta)$  is linear in  $\ln(\beta)$  for  $\beta$  sufficiently small, or equivalently,  $\pi_{rej}(N) \sim \gamma e^{-\theta N}$  for constants  $\gamma, \theta > 0$  when  $N$  is sufficiently large. This empirical finding is useful for computational purposes since the computational burden grows quickly when  $\beta$  gets very small.

Table 3.1. Minimal buffer size  $\nu(\beta)$  for the  $M^X/E_{10}/1/N$  queue

		constant					geometric				
		-2	-3	-4	-5	-6	-2	-3	-4	-5	-6
$\beta$		10	10	10	10	10	10	10	10	10	10
$\rho=0.5$	exa	8	12	16	20	24	15	22	28	35	42
	app	9	14	17	20	24	15	23	29	35	43
	exp	10	15	20	25	31	16	24	32	40	48
$\rho=0.8$	exa	17	28	39	50	61	26	43	61	78	96
	app	17	29	39	50	61	26	44	62	79	97
	exp	22	37	53	69	84	31	52	74	96	118
$\rho=0.9$	exa	26	48	72	95	118	39	74	109	144	180
	app	27	50	72	96	120	40	74	109	146	180
	exp	35	67	100	133	166	48	92	137	181	226

#### REFERENCES

- [1] Manfield, D.R. and Tran-Gia, Ph., Analysis of a Finite Storage System with Batch Input Arising out of Message Packetization, IEEE Trans. Comm. 30 (1982), 456-463.
- [2] Tijms, H.C., Stochastic Modelling and Analysis (Wiley, New York, 1986).
- [3] Bux, W., Single Server Queues with General Interarrival and Phase Type Service Time Distributions, 9th ITC, Torremolinos 1982.



- [4] Van Ommeren, J.C.W., Exponential Expansion of the Tail of the Waiting Time Probability in the Single Server Queue with Batch Arrivals, to appear in Adv. Appl. Prob. 20 (1988)
- [5] Van Ommeren, J.C.W., Simple approximations for the Batch Arrival  $M^X/G/1$  Queue, Research Report 1-88, Institute for Econometrics, Vrije Universiteit, Amsterdam.
- [6] Eikeboom, A.M. and Tijms, H.C., Waiting-Time Percentiles in the Multi-Server  $M^X/G/c$  Queue with Batch Arrivals, Prob. in the Engr. and Inform. Sci. 1 (1987), 75-96.
- [7] Van der Heijden, M.C., On the Three-Moment Approximation of a General Distribution by a Coxian Distribution, to appear in Prob. in the Engr. and Inform. Sci. 2 (1988).
- [8] Cohen, J.W., The Single Server Queue, 2nd. ed. (North-Holland, Amsterdam, 1982).
- [9] Burke, P.J., Delays in Single-Server Queues with Batch-Input, Oper. Res. 23, (1975), 830-833.
- [10] Krämer, W. and Langenbach-Belz, M., Approximate Formulae for the Delay in the Queueing System GI/G/1, 8th ITC, Melbourne, 1976.
- [11] Nobel, R.D., Practical Approximations for Finite-Buffer Queueing Models with Batch Arrivals, to appear in Europ. J. Oper. Res. 27 (1988).