Probability in the Engineering and Informational Sciences, **24**, 2010, 457–471.
doi:10.1017/S0269964810000100

LOSS PROBABILITIES FOR THE $M^X/G^Y/1/K+B$ BULK QUEUE

REMCO GERMS AND NICKY VAN FOREEST
Faculty of Economics and Business
University of Groningen
9700 AV Groningen, The Netherlands
E-mail: r.germs@rug.nl

In this article we analyze the $M^X/G^Y/1/K+B$ bulk queue. For this model, we consider three rejection policies: partial acceptance, complete rejection, and complete acceptance. For each of these policies, we are interested in the loss probability for an arriving group of customers and for individual customers within a group. To obtain these loss probabilities, we derive a numerically stable method to compute the limiting probabilities of the queue length process under all three rejection policies. At the end of the article we demonstrate our method by means of a numerical example.

1. INTRODUCTION

Consider the $M^X/G^Y/1/K+B$ queue with a buffer having capacity for K customers (excluding those in service) and a single server with maximal capacity B. Groups of customers arrive according to a Poisson process. Group sizes are independent and identically distributed (i.i.d.) random variables distributed according to some random variable X and are serviced in batches whose sizes are i.i.d. random variables distributed according to a random variable Y. An important practical problem is to find the customer and group loss probabilities for such systems; we provide one such application later. These probabilities clearly depend on the rejection policy. Common policies are the following: partial acceptance, in which the admitted part of a group is equal to the number of free places in the queue; complete rejection, in which the complete arriving group is rejected if its size exceeds the available buffer space; and complete acceptance, in which a group is completely accepted whenever part of it can be accepted.

The purpose of this article is to derive a numerically stable method to compute the (loss) probabilities of the $M^X/G^Y/1/K+B$ queue under all three rejection

© Cambridge University Press 2010 0269-9648/10 $25.00

https://doi.org/10.1017/S0269964810000100 Published online by Cambridge University Press

policies. We infer that it is easy to handle other rejection policies, although the three policies considered here appear to make the most practical sense, see also Perry and Asmussen [8].

We applied the bulk queue to find estimates for the loss probabilities in the sorting process of post parcels of a large international post delivery company. Post parcels enter an outlet (chute) at one end and are taken out at the other end by an employee. The arrival rate of parcels for the chute could temporarily exceed the working capacity of the employee so that the chute also serves as a buffer. When the chute overflows, parcels are directed to an overflow container and are sorted the next day. The target of the sorting center is to deliver at least a certain percentage of the parcels within one day. This process can be modeled as a queuing system with group arrivals, batch services, and complete rejection. The group sizes correspond to the sizes of the parcels (e.g., in centimeters), batch services to moving parcels from the chute, the sum of the sizes of the parcels in the chute form the queue length, and parcels are rejected (completely) when the queue length is such that the additional parcel does not fit into the chute. The problem is to determine the length of the chute such that the overflow (or loss) probability is not too large. Other applications and a detailed discussion of bulk queues can be found in Medhi [6] and references therein.

To obtain the loss probabilities, it is necessary to have expressions for the limiting distribution of the queue length process at (group) arrival epochs. A common approach for analyzing the queue length distribution for $M/G/1$ type of queues is to consider the Markov chain embedded at service completion epochs. To apply this method, we need to calculate the number $Z(s)$ of customers *accepted* during a service period of duration s. When the capacity of the buffer is infinite, $Z(s)$ is simply the number $N(s)$ of customers arriving during s. Since $N(s)$ is a compound Poisson random variable, its generating function has a simple form from which the required probability distribution can be obtained (see, e.g., Tijms [12, Theorem 1.2.1]). When the buffer size is bounded by K while the partial acceptance policy is in effect, the embedded Markov chain approach is also easy since now $Z(s) = \min\{N(s), k\}$, where k is the available space in the queue at the start of the service epoch. On the other hand, when the acceptance policy is either complete rejection or complete acceptance, there exists no simple form for $Z(s)$ (cf. Takagi [11, Sec. 5.6]) which might perhaps explain the lack of results in the literature for the complete acceptance and complete rejection policies. In case of complete acceptance, the queue length can exceed K, so that the state space becomes (much) larger than the state space with partial acceptance. Finally, when the policy is complete rejection, both the group size and the queue length at arrival epochs (which are generally not renewal moments of the queue length process) are required to decide to accept the group. Informally speaking, small groups might still fit, whereas large groups are blocked. To circumvent these problems, we use in our article an approach presented in Dudin, Shaban, and Klimenok [4] to analyze the $BMAP/G/1/K$ queue with complete rejection or complete acceptance.

Our work extends several models studied in the literature. The work by Nobel [7] was concerned with the $G^X/E_r/1/K$ queue with complete rejection but does not include batch services. Chang, Choi, and Kim [2] gave computational and analytical

https://doi.org/10.1017/S0269964810000100 Published online by Cambridge University Press

results for the $M^X/G^Y/1/K+B$ queue with partial acceptance and Sikdar and Gupta [10] extended this queuing model by including server vacations. Both articles, however, do not include complete rejection or complete acceptance. Gouweleeuw [5] provided an expression for the loss probability in case the group size is fixed, the service batch size is 1, and the system load is less than 1. None of these conditions is necessary for our approach. Tijms [12] provided numerous approximations for related queuing models such as the $M^X/G/c/K+c$ queue.

The article is organized as follows. After introducing the $M^X/G^Y/1/K+B$ queuing model with complete rejection in Section 2 we derive the limiting probabilities at departure epochs in Section 3. Section 4 is concerned with the loss probabilities as perceived by arriving groups and individual customers within the groups. Section 5 describes the algorithmic aspects of the numerical methods. Since now the overall structure is clear, we show in Section 6 how to deal with the other rejection policies. Finally, in Section 7 we present some numerical results.

2. MODEL

The $M^X/G^Y/1/K+B$ bulk queue can be described as follows. Groups of customers arrive according to a Poisson process with rate λ. The group sizes are i.i.d. integer random variables with distribution $P\{X=k\}=x_k, k\geq 1$, where X is a generic group size with finite mean $E(X)$. Customers are served in first come–first served (FCFS) order in service batches, which are also i.i.d. integer random variables distributed as the generic service batch size Y with probability mass function $P\{Y=l\}=y_l$, for $l=1,\ldots,B$, where B is the maximal server *capacity*. The buffer accepts groups according to the complete rejection policy; that is, if at an arrival moment, the queue has length q and the size of the group is X, the members of the group are only admitted if $q+X\leq K$, where $K<\infty$ is the size of the buffer. As any group larger than K cannot be accepted anyway, we assume for the moment that $x_k=0$ for all $k>K$; in Remark 2 we show how to relax this condition. Since the buffer is finite, the system cannot contain more than $K+B$ customers simultaneously. Pertaining to the distribution of Y, it is natural to assume that $B\leq K$; otherwise, set $y_K=\sum_{l=K}^{B} y_l$ and $y_k=0$ for $k=K+1,\ldots,B$. The batching policy is greedy in that if just before service starts, the queue contains less customers than the service capacity Y, the server takes all available customers into service. We assume that any arrival during a service joins the queue, if accepted, but cannot join a batch already in service. Service times of batches are assumed to be independent of the arrival process and the service batch size and form a set of i.i.d. random variables distributed according to the distribution function $G(s)=P(S\leq s)$ with finite mean $E(S)$.

Let $Q(t)$ denote the number of customers present in the queue at time $t\geq 0$. Let $Q(t)=\varnothing$ whenever the server is idle; otherwise, $Q(t)\in\{0,\ldots,K\}$. (The last element in this set is K, not $K+B$, since we are concerned with the queue length, not the number of customers in the system.) Thus, $E\subseteq\{\varnothing,0,1,\ldots,K\}$ is the state space of the process $\{Q(t)\}=\{Q(t),t\geq 0\}$. (Since service and arrivals are batchwise,

https://doi.org/10.1017/S0269964810000100 Published online by Cambridge University Press

E need not include all states in $\{\varnothing, 0, 1, \ldots, K\}$.) We assume the trajectories of this process to have left limits and to be right-continuous.

The server observes the queue length at service completion epochs and at arrival epochs of customers when the server is idle. Let $0 = T_0 < T_1 < T_2 < \cdots$ be the ordered sequence of these epochs and let $\{Q_n\} = \{Q_n, n \geq 0\}$ denote the (embedded) queue length process as observed by the server at these times, that is, we define

$$Q_n = \begin{cases} Q(T_n-) & \text{if } T_n \text{ is a service completion epoch} \\ Q(T_n-) + X & \text{if } T_n \text{ is an arrival epoch and the server is idle.} \end{cases}$$

Thus, Q_n is *either* the queue length just before service completion *or* the queue length just after the acceptance of the group of arriving customers. Observe that $Q_n \neq \varnothing$: the server cannot be idle just before service completion epochs, nor can $X = 0$. Thus, the state space of $\{Q_n\}$ is $E' = E \setminus \varnothing$.

It is clear (although it requires some technical arguments; see, e.g., Çinlar [3] or Asmussen [1]) that $\{Q_n, T_n\} = \{(Q_n, T_n); n \geq 0\}$ is a Markov renewal process embedded in $\{Q(t)\}$, so that $\{Q(t)\}$ is a semi-regenerative process. This means that for any n, the conditional distribution of $\{Q(t + T_0 + \cdots + T_n)\}_{t \geq 0}$, given $T_0, \ldots, T_n$ and $Q_0, \ldots, Q_n = i$, is the same as the conditional distribution of $\{Q(t)\}$, given $T_0 = 0$ and $Q_0 = i$. Hence, to characterize the conditional distribution of $\{Q(t)\}$, it suffices to specify the behavior of $\{Q(t)\}$ on the half-closed half-open interval (T_0, T_1). Let $T_0 = 0$, then $\{Q(t)\}$ must satisfy, on $[0, T_1)$,

$$Q(t) = \begin{cases} \varnothing & \text{if } Q_0 = 0 \\ [Q_0 - Y]_+ + Z(K - [Q_0 - Y]_+, t) & \text{if } Q_0 > 0, \end{cases} \tag{2.1}$$

where $[x]_+ = \max\{0, x\}$. Here, the random variable $Z(k, t)$ represents the number of *accepted* customers during $(0, t]$, given that k places are available in queue at time $T_0 = 0$ and $t \leq T_1$. Hence, $\{Z(k, t)\}$ is a pure birth process.

We assume that the embedded Markov chain $\{Q_n\}$ with state space E' is irreducible and aperiodic and that the Markov renewal process $\{Q_n, T_n\}$ is aperiodic. Since E' is also finite, it follows that $\{Q_n\}$ is positive recurrent.

Assuming that all of these conditions are satisfied, the limiting distributions π of the embedded Markov chain $\{Q_n\}$ and p of the semi-regenerative process $\{Q(t)\}$ exist; that is, for $j \in E'$,

$$\begin{aligned} \pi_j &= \lim_{n \to \infty} P\{\text{at time } T_n, j \text{ customers wait in queue}\} \\ &= \lim_{n \to \infty} P\{Q_n = j\}, \\ p_\varnothing &= \lim_{t \to \infty} P\{\text{at time } t \text{ the server is idle}\} \\ &= \lim_{t \to \infty} P\{Q(t) = \varnothing\}, \end{aligned}$$

https://doi.org/10.1017/S0269964810000100 Published online by Cambridge University Press

$$p_j = \lim_{t\to\infty} P\{\text{at time } t \text{ the server is busy and } j \text{ customers wait in queue}\}$$
$$= \lim_{t\to\infty} P\{Q(t) = j\}.$$

Finally, for later purposes, we define here $\bar{x}_k = \sum_{l=k}^{\infty} x_l$, $\bar{y}_k = \sum_{l=k}^{\infty} y_l = \sum_{l=k}^{B} y_l$, and the elapsed (residual) service time distribution $G^e(t) = (E(S))^{-1} \int_0^t 1 - G(s)\, ds$.

3. THE STATIONARY DISTRIBUTION OF THE EMBEDDED CHAIN

In this section we derive the set of linear equations that the stationary distribution π of the embedded chain $\{Q_n\}$ has to satisfy. To this end, we specify the transition matrix of $\{Q_n\}$. We start with deriving a numerically stable procedure to compute the semi-Markov kernel corresponding to the Markov renewal process $\{Q_n, T_n\}$. Taking a simple limit then yields the transition matrix.

The semi-Markov kernel $H(t) = \{H_{ij}(t); i,j \in E', t \geq 0\}$, is defined by (cf., Çinlar [3] or Asmussen [1])

$$H_{ij}(t) = P\{Q_{n+1} = j, T_{n+1} - T_n \leq t \mid Q_n = i\}$$
$$= P\{Q_1 = j, T_1 \leq t \mid Q_0 = i\}.$$

One readily verifies the following representation for $H(t)$:

$$H_{ij}(t) = \begin{cases} \left(1 - e^{-\lambda t}\right) x_j & \text{if } i = 0, j \geq 1 \\ \int_0^t p_{ij}(s)\, dG(s) & \text{if } i \geq 1, j \geq [i - B]_+ \\ 0 & \text{else,} \end{cases} \tag{3.1}$$

where $p_{ij}(s) = P\{Q_1 = j \mid Q_0 = i, T_1 = s\}$, for $i \geq 1, j \geq [i - B]_+$, where T_1 is a here service completion epoch since $i > 0$.

As is clear from the previous expression, the next step is to rewrite $p_{ij}(s)$ in a more convenient form. Observe first that $Q_1 = Q(s-)$, so that $p_{ij}(s) = P\{Q(s-) = j \mid Q_0 = i, T_1 = s\}$. Therefore, we can use (2.1) to write

$$p_{ij}(s) = P\{[i - Y]_+ + Z(K - [i - Y]_+, s) = j\}. \tag{3.2}$$

Second, we can expand $p_{ij}(s)$ in terms of the probability to accept j customers during half-open half-closed $(0, s)$, given that the queue has k positions available at $T_0 = 0$ while no orders depart from the system—that is, in terms of

$$R(j, k, s) = P\{Z(k, s) = j\}, \qquad 0 \leq j \leq k.$$

https://doi.org/10.1017/S0269964810000100 Published online by Cambridge University Press

For notational ease we extend the definition of R such that $R(j,k,s) = 0$ whenever $j < 0$. It follows from (3.2) and the independence of Y and Z that

$$p_{ij}(s) = \sum_{l\geq 1} P\{Y = l\} R(j - [i-l]_+, K - [i-l]_+, s)$$

$$= \bar{y}_i R(j,K,s) + \sum_{l=1}^{i-1} y_l R(j-i+l, K-i+l, s). \tag{3.3}$$

We can find a suitable expression for $R(j,k,s)$ by conditioning on the number $N(s)$ of groups arriving in the half-open half-closed interval $(0,s)$. Noting that $N(s)$ is Poisson distributed, this yields, for $0 \leq j \leq k \leq K$,

$$R(j,k,s) = \sum_{i=0}^{\infty} P\{Z(k,s) = j \mid N(s) = i\} P\{N(s) = i\}.$$

Now observe that, analogous to the arguments used in the uniformization method (cf. Tijms [12]), the function $s \to P\{Z(k,s) = j \mid N(s) = i\}$ is constant on $s > 0$. Thus, letting

$$U(j,k,i) = P\{Z(k,s) = j \mid N(s) = i\},$$

the above becomes

$$R(j,k,s) \sum_{i=0}^{\infty} U(j,k,i) e^{-\lambda s} \frac{(\lambda s)^i}{i!}. \tag{3.4}$$

In the following lemma we use the Kolmogorov (forward) equations of $R(j,k,s)$ to derive an efficient recursion for $U(j,k,i)$.

LEMMA 1: *The probability $U(j,k,i)$, $i \geq 0$ and $0 \leq j \leq k \leq K$, satisfies the recursion*

$$U(j,k,i+1) = U(j,k,i)\bar{x}_{k+1-j} + \sum_{l=0}^{j-1} U(l,k,i) x_{j-l}, \tag{3.5}$$

with initial conditions

$$U(j,k,0) = \begin{cases} 1 & \text{if } j = 0, \\ 0 & \text{if } j > 0. \end{cases} \tag{3.6}$$

PROOF: Since the batch interarrival times are exponentially distributed, it follows for small $h > 0$ that

$$R(j,k,s+h) = (1 - \lambda h) R(j,k,s) + \lambda h R(j,k,s) \bar{x}_{k+1-j}$$

$$+ \lambda h \sum_{l=0}^{j-1} R(l,k,s) x_{j-l} + o(h).$$

[Note that when $j = k$, then $\bar{x}_{k+1-j} = 1$ and the above becomes $R(k,k,s+h) = R(k,k,s) + \lambda h \sum_{l=0}^{k-1} R(l,k,s) x_{k-l} + o(h)$, as expected.]

https://doi.org/10.1017/S0269964810000100 Published online by Cambridge University Press

After some algebra and taking the limit $h \to 0$, we arrive at

$$\frac{d}{ds}R(j,k,s) = \lambda(\bar{x}_{k+1-j} - 1)R(j,k,s) + \lambda \sum_{l=0}^{j-1} R(l,k,s)\, x_{j-l}.$$

After substituting (3.4) in the above and simplifying the result, we obtain

$$\sum_{i=0}^{\infty} \frac{(\lambda s)^i}{i!} \big[U(j,k,i+1) - U(j,k,i)\big]$$
$$= \sum_{i=0}^{\infty} \frac{(\lambda s)^i}{i!} \left[(\bar{x}_{k+1-j} - 1)U(j,k,i) + \sum_{l=0}^{j-1} U(l,k,i) x_{j-l} \right].$$

As this has to hold true for all $s \geq 0$, the probabilities $U(j,k,i)$ must satisfy

$$U(j,k,i+1) - U(j,k,i) = (\bar{x}_{k+1-j} - 1)U(j,k,i) + \sum_{l=0}^{j-1} U(l,k,i) x_{j-l}$$

for all $i \geq 0$. This immediately simplifies to (3.5) for $j < k$.

The initial conditions follow from observing in (3.4) that $R(j,k,0) = U(j,k,0)$ and that $R(j,k,0) = 1$ if $j = 0$, and 0 otherwise. ■

Since $\{Q_n\}$, with transition matrix $H(\infty)$, is ergodic by assumption, its stationary distribution $\pi = (\pi_0, \ldots, \pi_K)$ exists and satisfies $\pi = \pi H(\infty)$. Taking the limit $s \to \infty$ in (3.1), π is therefore the unique solution of

$$\pi_0 = \sum_{i \geq 1} \pi_i \int_0^{\infty} p_{i0}(s)\, dG(s),$$
$$\pi_j = \pi_0 x_j + \sum_{i \geq 1} \pi_i \int_0^{\infty} p_{ij}(s) dG(s), \qquad j \in E' \setminus 0, \tag{3.7}$$
$$\sum_{j \in E'} \pi_j = 1.$$

In Section 5 we address how to rewrite this into a form that is readily implementable in computer code.

Remark 1: The recursion for $U(j,k,i)$ can be given a nice probabilistic interpretation. This will also prove useful in Section 6, where we need similar expressions for the $U(j,k,i)$ suitable to capture the acceptance behavior of the partial and complete

https://doi.org/10.1017/S0269964810000100 Published online by Cambridge University Press

acceptance policies. Let u correspond to the arrival time of the last batch. Then

$$\begin{aligned} U(j,k,i+1) &= P\{Z(k,s)=j \mid N(s)=i+1\} \\ &= P\{Z(k,s)=j \mid N(u-)=i\}P\{X>k-j\} \\ &\quad + \sum_{l=0}^{j-1} P\{Z(k,s)=l \mid N(u-)=i\}P\{X=j-l\} \\ &= U(j,k,i)\,\bar{x}_{k+1-j} + \sum_{l=0}^{j-1} U(l,k,i)x_{j-l}. \end{aligned}$$

Remark 2: In Section 2 we assumed that the group sizes are not allowed to exceed K (i.e., $x_k = 0$ for all $k > K$). This can be relaxed quite easily, however, by finding a suitable expression in (3.1) for $H_{0j}(t) = P\{Q_1 = j, T_1 \le t \mid Q_0 = 0\}$. Suppose that the number of batch arrivals in $[0,t)$ is $n > 1$. Since the n-batch is the batch that starts the busy period, the other batches are rejected. Thus, the size of all arrivals but the last exceed K, so that the first line in (3.1) should be replaced by

$$\begin{aligned} P\{Q_1 = j, T_1 \le t \mid Q_0 = 0\} &= \sum_{n=1}^{\infty} x_j (1-\bar{x}_{K+1})^{n-1} e^{-\lambda t} \frac{(\lambda t)^n}{n!} \\ &= \frac{x_j}{1-\bar{x}_{K+1}} \left(e^{-\lambda \bar{x}_{K+1} t} - e^{-\lambda t}\right). \end{aligned}$$

If $\bar{x}_{K+1} = 0$ this reduces, of course, to the first line of (3.1).

4. LOSS PROBABILITIES

In this section we consider the loss probabilities for the arriving groups and the individual customers within the groups. Since these losses occur at *arrival* epochs—whereas we have by the previous section only the probabilistic information π at embedded points T_n, being either a service completion epoch or an arrival epoch of customers when the server is idle—we need to convert π into long runtime average probabilities p by using semi-regeneration. Then, from PASTA, we conclude that the arriving batches and customers also observe p.

The following theorem relates the limiting distributions p to π by using semi-regeneration. The embedded Markov chain is $\{Q_n\}$ with stationary distribution π, and the cycle C, being the length of the interval between two successive embedded Markov points T_n and T_{n+1}, is a service interval of length S when $Q_n \ge 1$ and an interarrival time when $Q_n = 0$. Hence,

$$E(C) = \frac{\pi_0}{\lambda} + \sum_{i \ge 1} \pi_i E(S).$$

https://doi.org/10.1017/S0269964810000100 Published online by Cambridge University Press

THEOREM 2: *The limiting distribution* $p = (p_\varnothing, p_0, \ldots, p_K)$ *satisfies*

$$p_\varnothing = \frac{1}{\lambda}\frac{\pi_0}{E(C)}, \tag{4.1}$$

$$p_j = \frac{E(S)}{E(C)}\sum_{i\geq 1}\pi_i \int_0^\infty p_{ij}(s)\, dG^e(s), \qquad j \in E'. \tag{4.2}$$

PROOF: To prove Theorem 2, we use Çinlar [3, Thm. 6.6.12], which states that for $j \in E$,

$$p_j = \frac{1}{E(C)}\sum_{i\in E'}\pi_i \int_0^\infty \psi_i(t,j)\, dt, \tag{4.3}$$

provided that $\{Q_n, T_n\}$ is an ergodic process, $E(C) < \infty$, and the function $t \to \psi_i(t,j) = P\{Q(t) = j,\ T_1 > t \mid Q_0 = i\}$ is directly Riemann integrable for every $i \in E'$ and $j \in E$.

To check these conditions, note that the first two conditions are true by the assumptions made in Section 2. From (3.1) and the fact that T_1 equals a service time S when $Q_0 > 0$ and an interarrival time when $Q_0 = 0$, it follows that

$$\psi_i(t,j) = \begin{cases} e^{-\lambda t} & \text{if } i = 0, j = \varnothing \\ p_{ij}(t)(1 - G(t)) & \text{if } i \geq 1, j \geq [i - B]_+ \\ 0 & \text{else.} \end{cases} \tag{4.4}$$

It is clear that $\psi_0(t,\varnothing) = e^{-\lambda t}$ is directly Riemann integrable. From (4.4), we have that $\psi_i(t,j) \leq 1 - G(t)$ for $i \geq 1$. Note that $1 - G(t)$ is directly Riemann integrable since it is nonincreasing and $\int_0^\infty 1 - G(t)\, dt = E(S) < \infty$. These facts show that $t \to \psi_i(t,j)$ is directly Riemann integrable for every $i \in E'$ and $j \in E$.

For $j = \varnothing$, (4.1) follows directly from (4.3) and (4.4). Let $j \in E'$. Then by (4.3), (4.4), and using $dG^e(t) = (1 - G(t))\, dt/E(S)$, (4.2) directly follows. ■

We next derive the group loss probability under the complete rejection policy. Let Γ correspond to the event that a group is lost. Since, by PASTA, arriving batches see the queue length distributed as p_i, $i \in E$, it must be that at an arrival epoch s,

$$\begin{aligned} P\{\Gamma\} &= P\{X > K - Q(s)\} \\ &= \sum_{j\in E} P\{X > K - j \mid Q(s) = j\}p_j \\ &= \sum_{j\in E} \bar{x}_{K-j+1}p_j \\ &= p_\varnothing \bar{x}_{K+1} + \sum_{j=0}^{K} p_j x_{K-j+1}. \end{aligned} \tag{4.5}$$

https://doi.org/10.1017/S0269964810000100 Published online by Cambridge University Press

To calculate the loss probability for an arbitrary customer, we use the following renewal-theoretic result:

$$q_k := P\{\text{an arbitrary customer belongs to a group of size } k\} = \frac{kx_k}{E(X)}. \tag{4.6}$$

Let γ correspond to the event that an arbitrary customer is lost and define $\bar{q}_k = \sum_{l=k}^{\infty} q_l$, then

$$\begin{aligned} P\{\gamma\} &= P\left\{\gamma \,\middle|\, \begin{matrix}\text{customer sees 0 customers}\\ \text{in queue upon arrival}\end{matrix}\right\} p_\varnothing \\ &\quad + \sum_{j=0}^{K} P\left\{\gamma \,\middle|\, \begin{matrix}\text{customer sees } j \text{ customers}\\ \text{in queue upon arrival}\end{matrix}\right\} p_j \\ &= P\left\{\begin{matrix}\text{customer belongs to a group}\\ \text{of size larger than } K\end{matrix}\right\} p_\varnothing \\ &\quad + \sum_{j=0}^{K} P\left\{\begin{matrix}\text{customer belongs to a group}\\ \text{of size larger than } K-j\end{matrix}\right\} p_j \\ &= p_\varnothing \bar{q}_{K+1} + \sum_{j=0}^{K} p_j \bar{q}_{K-j+1}. \end{aligned} \tag{4.7}$$

5. ALGORITHMIC ASPECTS

In the previous two sections we derived a numerical procedure to compute the limiting probabilities π and p. In this section we discuss the algorithmic aspects of the procedure in more detail and introduce some more notation to bring the relation between π and p to the fore. Moreover, we show how the precision of our numerical method can be specified in advance.

Combining (3.3), (3.4), (3.7), and (4.2) and interchanging the integral and the summations (which is allowed by the positivity of the summands and integrands), we can write

$$\pi_0 = \sum_{i\geq 1} \pi_i \sum_{m=0}^{\infty} A(i,0,m) a_m, \tag{5.1}$$

$$\pi_j = \pi_0 x_j + \sum_{i\geq 1} \pi_i \sum_{m=0}^{\infty} A(i,j,m) a_m, \qquad j \in E' \setminus 0, \tag{5.2}$$

$$p_j = \frac{E(S)}{E(C)} \sum_{i\in E'} \pi_i \sum_{m=0}^{\infty} A(i,j,m) a^e_m, \qquad j \in E', \tag{5.3}$$

https://doi.org/10.1017/S0269964810000100 Published online by Cambridge University Press

where

$$A(i,j,m) = \bar{y}_i U(j,K,m) + \sum_{l=1}^{i-1} y_l U(j-i+l, K-i+l, m), \tag{5.4}$$

$$a_m = \int_0^\infty e^{-\lambda s} \frac{(\lambda s)^m}{m!}\, dG(s), \tag{5.5}$$

$$a_m^e = \int_0^\infty e^{-\lambda s} \frac{(\lambda s)^m}{m!}\, dG^e(s). \tag{5.6}$$

(Note that $U(j,k,m) = 0$ if $j < 0$.)

The numerical method that we have developed in the previous sections leads to the following algorithm:

Step 1 Compute (by numerical integration or if possible explicitly) the mixed Poisson probabilities a_m and a_m^e from (5.5) and (5.6).

Step 2 Compute $A(i,j,m)$ for $i,j \in E'$ and $m \geq 0$ by means of the recursion in Lemma 1 and (5.4).

Step 3 Use standard numerical procedures to compute π from (5.1) and (5.2).

Step 4 Compute p using (4.1) and (5.3).

Note that the computations in our approach only involve additions and multiplications of positive and bounded numbers, thereby preventing a loss a significant digits. Observe also that explicit expressions for the a_m and a_m^e can be given for the cases of deterministic and phase-type services.

Following Sericola and Tuffin [9], the probabilities a_m and a_m^e can be used as a criterion to stop the computation of $V(i,j) := \sum_{m=0}^\infty A(i,j,m)\, a_m$ in (5.1). From (5.4), we have that $A(i,j,m) \leq 1$ for all $i \geq 1$ and $j, m \geq 0$. Now, define for a given tolerance $\epsilon > 0$,

$$N = \min\left\{ n; \sum_{i=0}^{n} a_i \geq 1 - \epsilon \right\}$$

$$V_N(i,j) = \sum_{m=0}^{N} A(i,j,m) a_m.$$

We then have

$$V(i,j) = V_N(i,j) + e_N,$$

where e_N satisfies

$$e_N = \sum_{m \geq N+1} A(i,j,m) a_m \leq \sum_{m \geq N+1} a_m$$

$$= 1 - \sum_{m=0}^{N} a_m \leq \epsilon.$$

Similar reasoning applies to the computation of $\sum_{m=0}^\infty A(i,j,m) a_m^e$.

https://doi.org/10.1017/S0269964810000100 Published online by Cambridge University Press

6. PARTIAL AND COMPLETE ACCEPTANCE

The structure of the algorithm described in the previous section enables us to also analyze other rejection policies. Specifically, to compute the limiting probabilities of the queue length at departure, arbitrary and group arrival epochs for the partial and complete acceptance policy we need to find the form of the matrixes $A(i,j,m)$ as in (5.4). For the sake of completeness, we present for both policies the (conditional) transition probabilities $p_{ij}(s)$ and $U(j,k,i)$, from which $A(i,j,m)$ then easily follows. Note that the assumption that $x_k = 0$ for $k > K$ is no longer required, as groups can now be partially or completely accepted.

We start with the partial acceptance policy. In this case, the semi-Markov kernel is given by

$$H_{ij}(t) = \begin{cases} \left(1 - e^{-\lambda t}\right) x_j & \text{if } i = 0, 1 \le j \le K-1 \\ \left(1 - e^{-\lambda t}\right) \bar{x}_K & \text{if } i = 0, j = K \\ \int_0^t p_{ij}(s)\, dG(s) & \text{if } i \ge 1, j \ge [i - B]_+ \\ 0 & \text{else.} \end{cases}$$

Expanding $p_{ij}(s)$ also leads to (3.3). Hence, we obtain the same expression for $A(i,j,m)$ as in (5.4) while now $U(j,k,i)$, $i \ge 0$, satisfies the recursion

$$U(j,k,i+1) = \sum_{l=0}^{j-1} U(l,k,i) x_{j-l}, \qquad 1 \le j < k \le K,$$

$$U(k,k,i+1) = U(k,k,i) + \sum_{l=0}^{k-1} U(l,k,i) \bar{x}_{k-l},$$

with the same initial conditions as (3.6).

Under the complete acceptance policy, the semi-Markov kernel is also given by (3.1), whereas for $i = 1, \ldots, K$ and $j \ge 0$, the probabilities $p_{ij}(s)$ are given by (3.3). However, the complete acceptance policy allows the queue length to exceed K so that we need to specify $p_{ij}(s)$ also for $i > K$. It can be checked that for $K + 1 \le i < K + B$ (recall that $R(j,k,s) = 0$ if $j < 0$ or $k < 0$),

$$p_{ij}(s) = \sum_{l=1}^{B} y_l R(j - i + l, K - i + l, s) \quad \text{if } j < K \text{ or } j \ge i,$$

$$p_{ij}(s) = y_{i-j} + \sum_{l=i-j+1}^{B} y_l R(j - i + l, K - i + l, s) \quad \text{if } K \le j < i,$$

and for $i \ge K + B$,

$$p_{ij}(s) = \begin{cases} y_{i-j} & \text{if } i - B \le j < i \\ 0 & \text{otherwise.} \end{cases}$$

https://doi.org/10.1017/S0269964810000100 Published online by Cambridge University Press

The $R(j,k,s)$ are given by (3.4), whereas the probabilities $U(j,k,i)$, $k = 0,\dots,K$, satisfy the recursions

$$U(j,k,i+1) = \sum_{l=0}^{j-1} U(l,k,i)x_{j-l} \quad \text{if } j = 1,\dots,k-1,$$

$$U(j,k,i+1) = U(j,k,i) + \sum_{l=0}^{k-1} U(l,k,i)x_{j-l} \quad \text{if } j \geq k,$$

with initial conditions (3.6).

We now consider the loss probabilities under the partial and complete acceptance policies. Recall that Γ and γ respectively correspond to the event that a group is lost and that an arbitrary customer is lost. Then it is not difficult to check that for the complete acceptance policy

$$P\{\Gamma\} = P\{\gamma\} = 1 - p_{\varnothing} - \sum_{j=0}^{K-1} p_j.$$

Under the partial acceptance policy, it is preferable to interpret Γ as the event that a group of customers *overflows*—that is, when an arriving group does not fit completely into the queue (see Nobel [7]). The reasoning leading to (4.5) can also be applied to the partial acceptance policy, so that $P\{\Gamma\}$ is also given by (4.5). To calculate $P\{\gamma\}$, we use (4.6) to obtain

$$\eta_k := P\{\text{an arbitrary customer occupies the } k\text{th position in the group}\} = \sum_{m \geq k} \frac{x_m}{E(X)}$$

for $k \geq 1$. Define $\bar{\eta}_k = \sum_{m=k}^{\infty} \eta_m$ to be the probability that an arbitrary customers occupies a position greater than or equal to k in his group. Then analogous to the derivation of (4.7), we obtain

$$P\{\gamma\} = \sum_{j=0}^{K} p_j \bar{\eta}_{K-j+1}.$$

7. NUMERICAL EVALUATION

To demonstrate the algorithms presented in this article, we implemented them in a computer code.[1] We verified our algorithms by testing our code on numerical experiments from Gouweleeuw [5], Tijms [12, Sec. 9.8.3], and Chang et al. [2]. The results match exactly, except in one case of Chang et al. [2], which we discuss below.

[1] The Python code can be obtained from the first author at request.

https://doi.org/10.1017/S0269964810000100 Published online by Cambridge University Press

We apply the algorithm of Section 5 to the $M^X/D^Y/1/10+10$ queue under the partial acceptance, complete rejection, and complete acceptance policies. This model derives from the numerical example in Chang et al. [2], who analyzed the queuing process with the partial acceptance policy. In this example, $x_1 = 0.25$, $x_3 = 0.5$, and $x_5 = 0.25$ so that $E(X) = 3$, and $y_2 = 0.5$ and $y_{10} = 0.5$ so that $E(Y) = 6$. Furthermore, service times are deterministic with $E(S) = 10$.

Table 1 presents the limiting distribution p of the $M^X/D^Y/1/10+10$ queue at an arbitrary epoch for different loadings of the system, where $\rho = \lambda E(X)E(S)/E(Y)$ is the offered load to the system. Note that the values in Table 1 for the partial acceptance policy for $\rho = 1.0$ and $\rho = 2.0$ do not match the values presented in Chang et al. [2]. This is due to a mistake made in the computation of the limiting probabilities at an arbitrary epoch from the limiting probability at an embedded epoch in the work of Chang et al. [2]. We have corrected this error in Table 1.

Table 1. Limiting Distribution of the $M^X/D^Y/1/10+10$ Queue at an Arbitrary Epoch

	Partial Acceptance			Complete Rejection			Complete Acceptance		
	$\rho = 0.5$	$\rho = 1.0$	$\rho = 2.0$	$\rho = 0.5$	$\rho = 1.0$	$\rho = 2.0$	$\rho = 0.5$	$\rho = 1.0$	$\rho = 2.0$
$p_\varnothing$	.1791	.0352	.0023	.1793	.0352	.0023	.1690	.0262	.0011
p_0	.3078	.2252	.1245	.3080	.2252	.1245	.2904	.1674	.0584
p_1	.0967	.0635	.0319	.0970	.0636	.0319	.1029	.0741	.0411
p_2	.0279	.0206	.0092	.0288	.0211	.0092	.0366	.0438	.0352
p_3	.1026	.0967	.0622	.1037	.0975	.0623	.1020	.0873	.0456
p_4	.0487	.0487	.0306	.0531	.0552	.0330	.0518	.0573	.0391
p_5	.0525	.0613	.0417	.0569	.0692	.0452	.0536	.0619	.0406
p_6	.0491	.0633	.0463	.0589	.0963	.0833	.0469	.0600	.0410
p_7	.0252	.0385	.0281	.0348	.0744	.0781	.0276	.0449	.0368
p_8	.0506	.1180	.1293	.0424	.1226	.2079	.0339	.0690	.0673
p_9	.0166	.0416	.0517	.0214	.0736	.1533	.0251	.0590	.0638
p_{10}	.0431	.1873	.4421	.0156	.0660	.1690	.0258	.0979	.1940
p_{11}	.0000	.0000	.0000	.0000	.0000	.0000	.0165	.0654	.1357
p_{12}	.0000	.0000	.0000	.0000	.0000	.0000	.0116	.0539	.1233
p_{13}	.0000	.0000	.0000	.0000	.0000	.0000	.0037	.0174	.0406
p_{14}	.0000	.0000	.0000	.0000	.0000	.0000	.0027	.0144	.0365
Sum	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Mean	2.3738	4.7370	6.9633	2.2480	4.3674	6.3744	2.5927	5.4926	8.1740

Table 2. Loss Probability for an Arriving Group of Customers and for an Arbitrary Customer Within a Group in the $M^X/D^Y/1/10+10$ Queue.

	Partial Acceptance			Complete Rejection			Complete Acceptance		
	$\rho = 0.5$	$\rho = 1.0$	$\rho = 2.0$	$\rho = 0.5$	$\rho = 1.0$	$\rho = 2.0$	$\rho = 0.5$	$\rho = 1.0$	$\rho = 2.0$
$P\{\Gamma\}$	0.1121	0.3324	0.5965	0.0869	0.2559	0.4802	0.0603	0.2490	0.5301
$P\{\gamma\}$	0.0835	0.2759	0.5390	0.1132	0.3170	0.5673	0.0603	0.2490	0.5301

https://doi.org/10.1017/S0269964810000100 Published online by Cambridge University Press

Table 2 shows the loss probability for an arriving group of customers $P\{\Gamma\}$ and for individual customers within a group $P\{\gamma\}$ in the $M^X/D^Y/1/10+10$ queue under the three rejection policies for the different loadings of the system. As can be seen from Table 2, the complete rejection policy always results in the highest loss probability for customers, followed by the partial acceptance policy, which is intuitively clear.

References

1. Asmussen, S. (2003). *Applied probability and queues*. Berlin: Springer-Verlag.
2. Chang, S., Choi, D., & Kim, T.-S. (2004). Performance analysis of a finite-buffer bulk-arrival bulk-service queue with variable server capacity. *Stochastic Analysis and Applications 22*(5), 1151–1173.
3. Çinlar, E. (1975). *Introduction to stochastic processes*. Englewood Cliffs, NJ: Prentice-Hall.
4. Dudin, A., Shaban, A., & Klimenok, V. (2005). Analysis of a queue in the $BMAP/G/1/N$ system. *International Journal of Simulation: Systems, Science and Technology 6*(1–2), 13–23.
5. Gouweleeuw, F. (1994). The loss probability in finite-buffer queues with batch arrivals and complete rejection. *Probability in the Engineering and Informational Sciences* 8, 221–227.
6. Medhi, J. (2003). *Stochastic models in queuing theory*. San Diego: Academic Press.
7. Nobel, R. (1989). Practical approximations for finite-buffer queuing models with batch arrivals. *European Journal of Operational Research 38*, 44–55.
8. Perry, D. & Asmussen, S. (1995). Rejection rules in the $M/G/1$ queue. *Queueing Systems 19*, 105–130.
9. Sericola, B. & Tuffin, B. (1999). A fluid queue driven by a Markovian queue. *Queueing Systems 31*, 253–264.
10. Sikdar, K. and Gupta, U. (2008). On the batch arrival batch service queue with finite buffer under server's vacation: $M^X/G^Y/1/N$ queue. *Computers and Mathematics with Applications 56*, 2861–2873.
11. Takagi, H. (1993). *Queuing analysis: Finite systems*. Amsterdam: North–Holland.
12. Tijms, H. (2003). *A first course in stochastic models*. Chichester, UK: Wiley.

https://doi.org/10.1017/S0269964810000100 Published online by Cambridge University Press