

THE $M^X/D/c$ BATCH ARRIVAL QUEUE

GEERT JAN FRANX

*Department of Mathematics
Vrije Universiteit Amsterdam
De Boelelaan 1081*

1081HV Amsterdam, The Netherlands

E-mail: franx@few.vu.nl

A surprisingly simple and explicit expression for the waiting time distribution of the $M^X/D/c$ batch arrival queue is derived by a full probabilistic analysis, requiring neither generating functions nor Laplace transforms. Unlike the solutions known so far, this expression presents no numerical complications, not even for high traffic intensities.

1. INTRODUCTION

In the $M^X/D/c$ queue, batches of customers arrive according to a Poisson process with rate λ . There are c identical servers, serving each customer on a first-come first-served basis during a constant time D . Batch sizes are independently and identically distributed. With probability b_i , a batch contains $i > 0$ customers ($\sum_{i=1}^{\infty} b_i = 1$). The average batch size $B = \sum_{i=1}^{\infty} ib_i$. In order to ensure irreducibility, we assume $b_i > 0$ for at least one such i that c and i have a greatest common divisor of 1. We assume the traffic intensity $\rho = \lambda BD/c < 1$.

By the use of generating functions, an explicit expression for the waiting time distribution of the $M^X/D/c$ queue was derived by Kuczura [1]. Due to alternating terms, which are, in general, much larger than their sum, Kuczura's result is not really practical for numerical purposes.

As a way to get around the problem of round-off errors, a recursion scheme is described in Tijms [2]. However, "for increasing c and ρ this recursion scheme (too) will ultimately be hampered by round off errors," for which case an asymptotic expansion is recommended.

This article presents an alternative probabilistic approach, leading to a simple formula for the waiting time distribution, which is numerically stable for all $\rho < 1$. The analysis does not require generating functions or Laplace transforms.

2. NUMBER OF CUSTOMERS PRESENT IN THE SYSTEM

Let $r_j(T)$ denote the probability that exactly j new customers arrive during an arbitrary time interval of length T . Given that in such a time interval, n batch arrivals take place, the probability $\Psi_j(n)$ of these n batches containing j new customers can be derived by recursion from the convolution formula:

$$\Psi_j(n) = \sum_{i=1}^{j-1} b_i \Psi_{j-i}(n-1),$$

$$\Psi_j(1) = b_j.$$

Since the number of batches arriving during an arbitrary time interval of length T is Poisson distributed, the following expression holds for $r_j(T)$:

$$r_j(T) = \sum_{n=1}^{\infty} \Psi_j(n) \frac{(\lambda T)^n}{n!} e^{-\lambda T}.$$

Our next objective will be to derive a set of equations for $p_i(t)$, the probability of the system holding i customers at time t . Observe that all customers in service at time t will have left the system at time $t + D$. Consequently, customers present at time $t + D$ either arrived during the time interval $(t, t + D]$ or were already waiting for service at time t . Hence, by conditioning on the number of customers present at time t , we find

$$p_i(t + D) = \sum_{j=0}^c p_j(t) r_i(D) + \sum_{j=c+1}^{i+c} p_j(t) r_{i+c-j}(D), \quad \text{for all } i \in \mathbb{N}_0. \quad (2.1)$$

The stationary distribution $p_i = \lim_{t \rightarrow \infty} p_i(t)$ is found by letting $t \rightarrow \infty$ in (2.1):

$$p_i = \sum_{j=0}^c p_j r_i(D) + \sum_{j=c+1}^{i+c} p_j r_{i+c-j}(D) \quad \text{for all } i \in \mathbb{N}_0. \quad (2.2)$$

Together with the normalization equation and the capacity utilization equation, (2.2) constitutes an infinite system of linear equations with a unique solution (due to the irreducibility of the system) that can be calculated in several ways. In Tijms [2], a fast Fourier transform method and a geometric tail approach to solving (2.2) are described, both of which are quite efficient.

3. THE WAITING TIME DISTRIBUTION

Customers arriving in $(m + 1)$ st position within their batch will not be served until the first m customers of their batch are being served. This type of customers will be

called customers of the m -class. Let W_m denote the waiting time of an arbitrary m -class customer in the stationary system. The corresponding m -class waiting time distribution is denoted by $F_m(x)$:

$$F_m(x) = P\{W_m \leq x\}.$$

In order to find an expression for $F_m(x)$, we assume that an arbitrary m -class customer enters the stationary system. We define $t = 0$ at the arrival of this customer, which will be called the “marked customer.” Let $u \leq D$ be some positive time lapse. Due to the memoryless property of the arrival process, the system receives, with probability $r_n(u)$, exactly n new customers during the time interval $(-u, 0)$, just prior to the arrival of our marked customer.

We will focus on the queuing position of our marked customer at $t = D - u \geq 0$, given that n new customers arrive during the time interval $(-u, 0)$. Together with the m customers of higher priority, from $t = -u$ onward there will be $m + n$ new customers that entered the system with higher priority than the marked customer. Another group of customers with higher priority are those who were already waiting for service at $t = -u$. Therefore, we introduce the following definitions:

$L_q(t) :=$ the queue length at time t

$$q_i(t) := P\{L_q(t) = i\}$$

$$q_i := \lim_{t \rightarrow \infty} q_i(t) = \lim_{t \rightarrow \infty} P\{L_q(t) = i\}.$$

Let $j = L_q(-u)$ denote the number of customers waiting for service at $t = -u$. All customers in service at $t = -u$ will have left the system at $t = D - u$. Consequently, at $t = D - u$, the system contains exactly $j + m + n$ customers with a higher priority than the marked customer. If $j + m + n < kc$, the service initiation of the marked customer will take place no later than $t = kD - u$, implying $W_m \leq kD - u$. On the other hand, if $j + m + n \geq kc$, the service start of the marked customer will take place after $t = kD - u$, implying $W_m > kD - u$. Thus, by conditioning on n , we find

$$P\{W_m \leq kD - u \mid n \text{ new customers arrive during } (-u, 0)\} = P\{L_q(-u) < kc - n - m\}. \tag{3.1}$$

According to PASTA, as proven by Wolff [3,4], it is possible to find the time-average probability of any state of the system by sampling the system at the jump times of some Poisson process $A \equiv \{A(t), t \geq 0\}$. The crux of Wolff’s central theorem is that the Poisson process need not necessarily be the arrival process. In fact, any Poisson process will do, provided that it satisfies the Lack of Anticipation Assumption (LAA), implying that its increments $A(t + \Delta t) - A(t)$ and the evolution of the system $\{U(s), 0 \leq s \leq t\}$ are independent for any fixed $t > 0$. Therefore, we can also apply PASTA to the Poisson process $A_m \equiv \{A_m(t), t \geq 0\}$, which is counting the arrivals of all batches containing more than m customers. Another Poisson process satisfying LAA is $A_m^u \equiv \{A_m(t + u), t \geq 0\}$ for any positive u . Here, all jumps are taking place exactly u time units before the corresponding jumps of A_m .

Thus, by applying PASTA to A_m^u in the $M^X/D/c$ system, we find for the right-hand side of equation (3.1),

$$\begin{aligned} P\{L_q(-u) < kc - n - m\} &= \lim_{t \rightarrow \infty} P\{L_q(t) < kc - n - m\} \\ &= \lim_{t \rightarrow \infty} \sum_{i=0}^{kc-n-m-1} q_i(t) = \sum_{i=0}^{kc-n-m-1} q_i. \end{aligned}$$

Now, by defining the cumulative probability $Q_j := \sum_{i=0}^j q_i = \sum_{i=0}^{j+c} p_i$, we can rewrite equation (3.1) as

$$P\{W_m \leq kD - u \mid n \text{ new customers arrive during } (-u, 0)\} = Q_{kc-n-m-1}.$$

Since negative queue lengths are impossible, we sum over all $n \leq kc - m - 1$ in order to obtain the (unconditional) waiting time distribution of m -class customers:

$$\begin{aligned} P\{W_m \leq kD - u\} &= \sum_{n=0}^{kc-m-1} Q_{kc-n-m-1} P\{n \text{ new customers arrive during } (-u, 0)\} \\ &= \sum_{n=0}^{kc-m-1} Q_{kc-n-m-1} r_n(u). \end{aligned}$$

Substituting $x = kD - u$, this result can be formulated in terms of x :

$$P\{W_m \leq x\} = \sum_{n=0}^{kc-m-1} Q_{kc-n-m-1} r_n(kD - x) \quad \text{for } (k-1)D \leq x < kD. \quad (3.2)$$

Since this expression contains only a finite number of positive terms and since $Q_{kc-n-m-1}$ and $r_n(kD - x)$ can be calculated as described in Section 2, (3.2) does not present any numerical complications, regardless of the traffic intensity ρ .

Our next objective will be to derive an expression for the general waiting time distribution $F(x) = P\{W \leq x\}$ for an arbitrary customer of unknown class. With probability $jb_j / \sum_{i=1}^{\infty} ib_i = jb_j/B$, the marked customer arrives in a batch containing j customers. Given that a customer arrives in a batch of j customers, there is a probability of $1/j$ that this customer is of the m -class, for $m = 0, \dots, j - 1$. (Here, $m = j$ is not possible since m is the number of customers with higher priority that arrived in the same batch as the marked customer.) By conditioning on the batch size j , we find the probability a_m of an arbitrary customer being m -class:

$$a_m = \sum_{j=m+1}^{\infty} \frac{1}{j} P\{\text{arrival takes place in batch of size } j\} = \frac{1}{B} \sum_{j=m+1}^{\infty} b_j.$$

Using these constants a_m , we can write the following expression for the general waiting time distribution:

$$P\{W \leq x\} = \sum_{m=0}^{\infty} a_m \sum_{n=0}^{kc-m-1} Q_{kc-n-m-1} r_n(kD - x) \quad \text{for } (k-1)D \leq x < kD. \quad (3.3)$$

References

1. Kuczura, A. (1973). Batch input to a multiserver queue with constant service times. *The Bell System Technical Journal* 52(1).
2. Tijms, H.C. (1994). *Stochastic models, an algorithmic approach*. New York: Wiley.
3. Wolff, R.W. (1982). Poisson arrivals see time averages. *Operations Research* 30: 223–231.
4. Wolff, R.W. (1989). *Stochastic modelling and the theory of queues*. Englewood Cliffs, NJ: Prentice-Hall.