The number of overlapping customers

Onno Boxma*

December 8, 2023

Abstract

We consider the number of overlapping customers in several single- and multiserver queues, i.e., the number of customers whose visit to a service system at some point in time has an overlap with that of a tagged customer. Restricting ourself to the FCFS policy, we obtain the probability generating function and moments of the number of overlaps in the $M/G/1$, $M/G/1/N$, $M/M/c$ and $G/M/c$ queue and (as an approximation) the $M/G/c$ queue.

1 Introduction

The COVID-19 pandemic has triggered an interest in the number and lengths of contacts a customer experiences in a service facility. Long queues at, e.g., baggage checks and passport controls in airports, or in shops, give rise to many opportunities for transmission of infections. Several recent studies have tackled the problem how to organize a queue in order to increase the safety of customers; see, e.g., Perlman and Yechiali [8] and its references. Kang et al. [4] combine a spatial SIR (Susceptible-Infected-Recovered) model with queueing theory to compute the expected number of transmissions from an infected customer during its sojourn in a service facility, assuming all other customers are susceptible. Palomo and Pender [6] study the *overlap time*, viz., the amount of time two customers, k arrivals apart, spend in the service system together. They obtain its distribution for the $M/M/1$ queue, and present simulation results for the $G/G/1$ queue with lognormal interarrival and service time distributions. Pender and Palomo [7] investigate the overlapping time distribution in infinite server queues. Xu et al. [11] aim to design queueing topologies and flow control policies in order to

*Department of Mathematics and Computer Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands; o.j.boxma@tue.nl

reduce the risk of the spread of virus infections. They propose two metrics to measure the customers' infection risk in a service system, viz., the average overlapping time and the average number of overlaps (overlapped customers).

In the present paper we focus on *the distribution of the number of overlaps*, i.e., the number of customers whose visit to a service facility at some point in time has an overlap with that of a tagged customer. Our main finding is that one can obtain explicit exact expressions for the generating function of the number of overlaps in some of the most important single- and multiserver queues, as well as their moments. We successively obtain the probability generating function, and the mean and variance, of the number of overlaps in the $M/G/1$ and $M/G/1/N$ queue (Section 2), the $M/M/c$ queue (Section 3) and the $G/M/c$ queue (Section 4). We provide an approximation for the generating function of the number of overlaps in the $M/G/c$ queue in Section 5. In this paper we restrict ourself to the First-Come First-Served policy; in a future study, the impact of the service policy on the number of overlaps will be considered.

2 The $M/G/1$ queue

In this section we focus on the $M/G/1$ FCFS queue. We consider an arbitrary arriving customer C, the "tagged" customer, and we determine the steady-state distribution of the number of customers who overlap with C. First some notation: customers arrive according to a Poisson process with rate λ; a generic service time is denoted by B, with distribution $B(\cdot)$ and LST (Laplace-Stieltjes transform) $\beta(\cdot)$; and the load of the server, $\rho := \lambda \mathbb{E}B$, is assumed to be less than one. Under the FCFS discipline, C overlaps with all X customers seen upon its arrival and all Z customers arriving during its sojourn time S; the latter are also the customers left behind at C's departure. The (transform of the) steady-state joint distribution of X, S and Z was obtained in Theorem 3.1 of [2] (the mean of the residual service time seen upon arrival by a customer who meets k customers was already derived in [5], see also [9]):

$$\mathbb{E}[z^X \mathrm{e}^{-\omega S} r^Z] = (1-\rho)\beta(\omega+\lambda(1-r)) + Q(z\beta(\omega+\lambda(1-r)), \omega+\lambda(1-r)), \quad (1)$$

where

$$Q(z,\omega) := \frac{(1-\rho)z}{\beta(\lambda(1-z)) - z} \lambda(1-z) \frac{\beta(\omega) - \beta(\lambda(1-z))}{\lambda(1-z) - \omega}. \quad (2)$$

$Q(z, \omega)$ actually equals the joint transform of X and ζ with indicator function $I(X > 0)$, where ζ is the residual service time of the customer found in service by C.

The PGF (probability generating function) of O, the number of customers overlapping with C, is given by $\mathbb{E}[z^{X+Z}]$ or, alternatively, by $\mathbb{E}[z^X \mathrm{e}^{-\lambda(1-z)S}]$ (the last factor denoting the PGF of the number of arrivals during S). Either way, the result is obtained from (1) and (2) as follows, where we define $\gamma(z) := \lambda(1 - z\beta(\lambda(1-z)))$:

Theorem 1 *In the $M/G/1$ FCFS queue, the PGF of the number of overlaps O is given by*

$$\mathbb{E}[z^O] \;=\; \frac{(1-\rho)(1-z)\beta(\lambda(1-z))}{1-\beta(\lambda(1-z))}\frac{\beta(\lambda(1-z))[1-\beta(\gamma(z))]}{\beta(\gamma(z)) - z\beta(\lambda(1-z))}, \tag{3}$$

with mean

$$\mathbb{E}O = \frac{\lambda^2 \mathbb{E}B^2}{1-\rho} + 2\rho, \tag{4}$$

and variance

$$\begin{aligned}\mathrm{Var}[O] \;&=\; \frac{\lambda^3\mathbb{E}(B^3)(3+\rho)}{3(1-\rho)} + \frac{\lambda^4(\mathbb{E}(B^2))^2(1+\rho)}{2(1-\rho)^2} \\ &+\; \frac{2\lambda^2\mathbb{E}(B^2)}{1-\rho} + 2\rho + 2\lambda^2\mathrm{Var}[B]. \end{aligned}\tag{5}$$

Remark 1 *Of course, $\mathbb{E}O$ also follows by realizing that $\mathbb{E}O = \mathbb{E}[X+Z] = 2\mathbb{E}X$ while the latter expression is well-known (see, e.g., Chapter II.4 of [1]). For the overlap variance, we have made use of Expression (3.6) in [2] for the covariance of X and Z.*

Remark 2 *In the case of the $M/M/1$ queue, it readily follows from (3) (see also Remark 3.1 of [2]) that*

$$\mathbb{E}[z^O] \;=\; \frac{1-\rho}{1+\rho-2\rho z}, \tag{6}$$

and hence O is geometrically distributed:

$$\mathbb{P}(O = n) \;=\; \frac{1-\rho}{1+\rho}\left(\frac{2\rho}{1+\rho}\right)^n, \quad n = 0, 1, \ldots, \tag{7}$$

with mean

$$\mathbb{E}O = \frac{2\rho}{1-\rho}, \tag{8}$$

and variance

$$\mathrm{Var}[O] = \frac{2\rho + 2\rho^2}{(1-\rho)^2}. \tag{9}$$

In this particular case, a quick derivation of (6) may be based on the observation that $O = X + N(B_1^{res}) + N(B_2) + \cdots + N(B_X) + N(B)$, *where the residual service time* $B_1^{res} = \zeta$ *of the customer in service is exp(*μ*) distributed like all ordinary service times, and where the number of Poisson(*λ*) arrivals* $N(B_i)$ *during* B_i *is geom(*$\frac{\lambda}{\lambda+\mu}$*) distributed:*

$$\mathbb{E}[z^O] \;=\; \sum_{k=0}^{\infty}(1-\rho)\rho^k z^k (\frac{\mu}{\mu + \lambda(1-z)})^{k+1} \;=\; \frac{1-\rho}{1+\rho-2\rho z}. \tag{10}$$

While X *and the number of arrivals during* C*'s sojourn time are both geometrically distributed, their sum is not negative binomially distributed, because those two random variables are correlated. The fact that* O *is geometrically distributed, hence memoryless, may be intuitively explained by applying the memoryless property to* $O = \sum_{j=1}^{X}(1 + N(B_j)) + N(B)$.

2.1 The $M/G/1/N$ queue

In this subsection we briefly discuss the $M/G/1$ queue with a finite total capacity of N customers. To make the approach more transparent, we start with the $M/M/1/N$ queue. An arriving customer who meets N customers is lost. The arrival rate is λ, and the service rate is μ; $\rho := \lambda/\mu$. X again denotes the number of customers found by an admitted tagged customer C upon arrival, and Z is the number left behind by C upon departure; $O = X + Z$. However, Z no longer equals the number of customers who have arrived during C's sojourn time, because some of those arrivals may have been rejected.

Let us introduce, for $k, i \geq 0$, $k + i \leq N - 1$,

$$\gamma_{k,i}(z) \;:=\; \mathbb{E}[z^Z|\text{C has k customers in front and i behind at a service start}]. \tag{11}$$

Observe that we have $\mathbb{E}[z^O]$ once $\gamma_{k,0}(z)$ is known:

$$\mathbb{E}[z^O] \;=\; \sum_{k=0}^{N-1}\mathbb{P}(X=k)z^k\gamma_{k,0}(z) \;=\; \sum_{k=0}^{N-1}\frac{1-\rho}{1-\rho^N}\rho^k z^k \gamma_{k,0}(z). \tag{12}$$

We shall derive a recursion that expresses $\gamma_{k,i}(z)$ into $\gamma_{k-1,\cdot}(z)$, for $k = 1, \ldots, N-1$, along with an expression for $\gamma_{0,i}(z)$. Thus it is possible to

obtain all $\gamma_{k,i}(z)$, and in particular $\gamma_{k,0}(z)$. Even though C may arrive during a service time, the residual service time still is $\exp(\mu)$ and hence we may pretend that a service starts right now.
By looking one service ahead, we can write, for $k = 1, 2, \ldots,\ k + i \le N - 1$:

$$\begin{aligned}\gamma_{k,i}(z) \quad &= \quad \sum_{j=0}^{N-k-i-1} (\frac{\lambda}{\lambda+\mu})^j \frac{\mu}{\lambda+\mu} \gamma_{k-1,i+j}(z) \\ &+ \quad (\frac{\lambda}{\lambda+\mu})^{N-k-i} \gamma_{k-1,N-k}(z); \end{aligned} \tag{13}$$

while for $i = 0, 1, \ldots, N - 1$,

$$\gamma_{0,i}(z) \ = \ \sum_{j=0}^{N-i-1} z^{i+j} (\frac{\lambda}{\lambda+\mu})^j \frac{\mu}{\lambda+\mu} \ + \ z^{N-1} (\frac{\lambda}{\lambda+\mu})^{N-i}. \tag{14}$$

In the $M/G/1/N$ case, a very similar recursive approach can be used, with one exception: when C arrives, the residual service time ζ of the customer in service has a different distribution than an arbitrary full service time, and it moreover depends on X. Theorem 4.1 of [2] presents the joint transform of X and the residual service time of the customer in service when C arrives, for the $M/G/1/N$ queue. This allows us to look ahead from C's arrival epoch to the start of the next service. From there, we have the same recursive scheme as in (13) and (14), except that terms like $(\frac{\lambda}{\lambda+\mu})^j \frac{\mu}{\lambda+\mu}$ have to be replaced by $\int_0^\infty \mathrm{e}^{-\lambda t} \frac{(\lambda t)^j}{j!} \, \mathrm{d}\mathbb{P}(B < t)$ (which again gives the probability of j arrivals during one service time):

$$\begin{aligned}\gamma_{k,i}(z) \quad &= \quad \sum_{j=0}^{N-k-i-1} \int_0^\infty \mathrm{e}^{-\lambda t} \frac{(\lambda t)^j}{j!} \mathrm{d}B(t) \ \gamma_{k-1,i+j}(z) \\ &+ \quad \sum_{j=N-k-i}^{\infty} \int_0^\infty \mathrm{e}^{-\lambda t} \frac{(\lambda t)^j}{j!} \mathrm{d}B(t) \ \gamma_{k-1,N-k}(z); \end{aligned} \tag{15}$$

while for $i = 0, 1, \ldots, N - 1$,

$$\gamma_{0,i}(z) \ = \ \sum_{j=0}^{N-i-1} z^{i+j} \int_0^\infty \mathrm{e}^{-\lambda t} \frac{(\lambda t)^j}{j!} \mathrm{d}B(t) + z^{N-1} \sum_{j=N-i}^{\infty} \int_0^\infty \mathrm{e}^{-\lambda t} \frac{(\lambda t)^j}{j!} \mathrm{d}B(t). \tag{16}$$

3 The $M/M/c$ queue

We now consider the $M/M/c$ queue with service in order of arrival. Assume that the service times are exponentially distributed with mean $1/\mu$. We assume that $\lambda < c\mu$ to ensure that steady-state distributions of the key performance measures exist. Just like for $M/G/1$ we have

$$\mathbb{E}[z^O] \;=\; \mathbb{E}[z^X \mathrm{e}^{-\lambda(1-z)S}], \tag{17}$$

and using Little's formula and PASTA it follows once more that $\mathbb{E}O = \mathbb{E}X + \lambda\mathbb{E}S = 2\mathbb{E}X$. The probabilities p_k of C seeing k customers upon arrival (and, by PASTA, also at an arbitrary epoch), are given by (cf. Section 4.5 of [10]),

$$\begin{aligned} p_k &= p_0 \frac{(\lambda/\mu)^k}{k!}, \quad k = 0, 1, \ldots, c, \\ p_k &= p_c (\frac{\lambda}{c\mu})^{k-c}, \quad k = c, c+1, \ldots, \end{aligned} \tag{18}$$

where p_0 easily follows by normalization:

$$p_0 = [\sum_{k=0}^{c-1} \frac{(\lambda/\mu)^k}{k!} + \frac{(\lambda/\mu)^c}{c!} \frac{c\mu}{c\mu - \lambda}]^{-1}. \tag{19}$$

C's sojourn time equals its service time when $X < c$, while when $X = k \geq c$, it equals its service time plus a sum of $k - c + 1$ independent exponentially distributed random variables with mean $1/(c\mu)$ – the minimum of c (residual) service times. Hence

$$\mathbb{E}[z^X \mathrm{e}^{-\lambda(1-z)S}] \;=\; \frac{\mu}{\mu + \lambda(1-z)} [\sum_{k=0}^{c-1} p_k z^k + \sum_{k=c}^{\infty} p_k z^k (\frac{c\mu}{c\mu + \lambda(1-z)})^{k-c+1}]. \tag{20}$$

Combining (17), (18) and (20), we obtain the following result.

Theorem 2 *In the $M/M/c$ FCFS queue, the PGF of the number of overlaps O is given by*

$$\mathbb{E}[z^O] \;=\; p_0 \frac{\mu}{\mu + \lambda(1-z)} [\sum_{k=0}^{c-1} \frac{(\lambda/\mu)^k}{k!} z^k + \frac{(\lambda/\mu)^c}{c!} z^c \frac{c\mu}{c\mu + \lambda(1-2z)}]. \tag{21}$$

The mean number of overlaps equals

$$\mathbb{E}O = 2\frac{\lambda}{\mu} + 2\frac{(\lambda/\mu)^c}{c!} p_0 \frac{\lambda}{c\mu} \frac{1}{(1 - \frac{\lambda}{c\mu})^2}, \tag{22}$$

and its variance is

$$\begin{aligned}\mathrm{Var}(O) &= 2\frac{\lambda}{\mu}+(\frac{\lambda}{\mu})^2+p_0(\frac{\lambda}{\mu})^2\frac{(\frac{\lambda}{\mu})^c}{c!}\frac{1}{(1-\rho)^2}\times \\ &\times \left[\frac{3}{c}(\frac{1}{\rho}-1)+\frac{8}{c}\frac{1}{1-\rho}+\frac{2}{c^2}\frac{1}{\rho}-4p_0(\frac{\lambda}{\mu})^2\frac{(\frac{\lambda}{\mu})^c}{c!}\frac{1}{(1-\rho)^2}\right].\end{aligned} \tag{23}$$

In presenting $\mathbb{E}O$ we have used a result for $\mathbb{E}X$ from [10], p. 288. $\mathbb{E}O$ was first obtained in Corollary 2 of [11].

4 The $G/M/c$ queue

Consider the $G/M/c$ queue with service in order of arrival. A generic interarrival time is denoted by A, with distribution $A(\cdot)$ and LST $\alpha(\cdot)$. The mean service time is again denoted by $1/\mu$. Equation (17) no longer holds, but we now have, with $a(S)$ the number of arrivals during the sojourn time S of C:

$$\mathbb{E}[z^O] = \mathbb{E}[z^{X+a(S)}]. \tag{24}$$

We refer to Section 4.6.1 of Tijms [10] for a discussion of the probabilities p_k of C meeting $X=k$ customers upon arrival. He points out that

$$p_k = \eta^{k-c+1}p_{c-1}, \quad k \geq c-1, \tag{25}$$

where η is the unique solution in $[0,1]$ of the equation

$$x = \int_0^\infty \mathrm{e}^{-c\mu(1-x)t}\,\mathrm{d}A(t). \tag{26}$$

Tijms subsequently shows how the probabilities $p_0,\dots,p_{c-1}$ can be obtained in the cases of deterministic, generalized Erlang and Coxian-2 interarrival time distributions.

Let us now determine $\mathbb{E}[z^{a(S)}|X=k]$. If $k<c$ then $S \sim \exp(\mu)$, and it is readily seen that

$$\mathbb{E}[z^{a(S)}|X=k<c] = \sum_{j=0}^{\infty} z^j[\alpha^j(\mu)-\alpha^{j+1}(\mu)] = \frac{1-\alpha(\mu)}{1-\alpha(\mu)z}. \tag{27}$$

If $k \geq c$ then we can write $S = E_{k-c+1,c\mu}+B$, where $E_{k-c+1,c\mu}$ denotes an Erlang distributed random variable with $k-c+1$ phases that are exponentially distributed with rate $c\mu$ and B is an $\exp(\mu)$ service time; those two random variables are independent. We can hence write:

$$\begin{aligned}\mathbb{E}[z^{X+a(S)}|X=k\geq c] &= z^k\mathbb{E}[z^{a(E_{k-c+1,c\mu}+B)}] \qquad (28)\\ =\quad z^k\int_{u=0}^{\infty}\int_{t=0}^{u}&\mathbb{E}[z^{a(u)}]\mu\mathrm{e}^{-\mu(u-t)}c\mu\frac{(c\mu t)^{k-c}}{(k-c)!}\mathrm{e}^{-c\mu t}\mathrm{d}t\mathrm{d}u,\end{aligned}$$

yielding

$$\begin{aligned}&\sum_{k=c}^{\infty}p_k\mathbb{E}[z^{X+a(S)}|X=k] \qquad (29)\\ =\quad & z^c p_c\sum_{k=c}^{\infty}(\eta z)^{k-c}\int_{u=0}^{\infty}\int_{t=0}^{u}\mathbb{E}[z^{a(u)}]\mu\mathrm{e}^{-\mu(u-t)}c\mu\frac{(c\mu t)^{k-c}}{(k-c)!}\mathrm{e}^{-c\mu t}\mathrm{d}t\mathrm{d}u\\ =\quad & z^c p_c\int_{u=0}^{\infty}\int_{t=0}^{u}\mathbb{E}[z^{a(u)}]\mu\mathrm{e}^{-\mu(u-t)}c\mu\mathrm{e}^{-c\mu(1-\eta z)t}\mathrm{d}t\mathrm{d}u\\ =\quad & z^c p_c c\mu^2\int_{u=0}^{\infty}\mathbb{E}[z^{a(u)}]\frac{\mathrm{e}^{-\mu u}-\mathrm{e}^{-c\mu(1-\eta z)u}}{c\mu(1-\eta z)-\mu}\mathrm{d}u\\ =\quad & z^c p_c\frac{c\mu^2}{c\mu(1-\eta z)-\mu}\left[\frac{1}{\mu}\frac{1-\alpha(\mu)}{1-\alpha(\mu)z}-\frac{1}{c\mu(1-\eta z)}\frac{1-\alpha(c\mu(1-\eta z))}{1-\alpha(c\mu(1-\eta z))z}\right].\end{aligned}$$

The above results in the following theorem; the moments are obtained by straightforward but tedious differntiations.

Theorem 3 *In the $G/M/c$ FCFS queue, the PGF of the number of overlaps O is given by*

$$\begin{aligned}\mathbb{E}[z^O] \quad=\quad & \sum_{k=0}^{c-1}p_k z^k\frac{1-\alpha(\mu)}{1-\alpha(\mu)z} \qquad (30)\\ +\quad & z^c p_c\frac{c\mu^2}{c\mu(1-\eta z)-\mu}\left[\frac{1}{\mu}\frac{1-\alpha(\mu)}{1-\alpha(\mu)z}-\frac{1}{c\mu(1-\eta z)}\frac{1-\alpha(c\mu(1-\eta z))}{1-\alpha(c\mu(1-\eta z))z}\right].\end{aligned}$$

The mean number of overlaps equals

$$\begin{aligned}\mathbb{E}[O] \quad=\quad & \mathbb{E}[X]+\sum_{k=0}^{c-1}p_k\frac{\alpha(\mu)}{1-\alpha(\mu)} \qquad (31)\\ +\quad & p_c\frac{c\mu^2}{c\mu(1-\eta)-\mu}\left[\frac{1}{\mu}\frac{\alpha(\mu)}{1-\alpha(\mu)}-\frac{1}{c\mu(1-\eta)}\frac{\alpha(c\mu(1-\eta))}{1-\alpha(c\mu(1-\eta))}\right],\end{aligned}$$

where $\mathbb{E}[X]=\sum_{k=0}^{c-1}kp_k+p_c[\frac{c}{1-\eta}+\frac{\eta}{(1-\eta)^2}]$.

The variance of the number of overlaps equals

$$
\begin{aligned}
& \mathrm{Var}[O] = \mathbb{E}[O(O-1)] + \mathbb{E}[O] - (\mathbb{E}[O])^2 && (32)\\
= \quad & \mathbb{E}[O] - (\mathbb{E}[O])^2 + \sum_{k=0}^{c-1} k(k-1)p_k \; + \; 2\sum_{k=0}^{c-1} kp_k \frac{\alpha(\mu)}{1-\alpha(\mu)} \; + \; \sum_{k=0}^{c-1} 2p_k (\frac{\alpha(\mu)}{1-\alpha(\mu)})^2 \\
+ \quad & c(c-1)p_c \frac{1}{1-\eta} \; + \; 2p_c \frac{c\mu\eta}{c\mu(1-\eta)-\mu} \frac{1}{1-\eta} \; + \; 2p_c (\frac{c\mu\eta}{c\mu(1-\eta)-\mu})^2 \frac{1}{1-\eta} \\
+ \quad & 2[cp_c \frac{c\mu^2}{c\mu(1-\eta)-\mu} + p_c \frac{c^2\mu^3\eta}{(c\mu(1-\eta)-\mu)^2}] \times \\
\times \quad & [\frac{1}{\mu} \frac{\alpha(\mu)}{1-\alpha(\mu)} \; - \; \frac{\eta}{c\mu(1-\eta)^2} \; - \; \frac{1}{c\mu(1-\eta)} \frac{\alpha(\xi)}{1-\alpha(\xi)}] \\
+ \quad & p_c \frac{c\mu^2}{c\mu(1-\eta)-\mu} [\frac{2}{\mu} (\frac{\alpha(\mu)}{1-\alpha(\mu)})^2 \; - \; \frac{4\eta^2}{c\mu(1-\eta)^3} \; - \; 2\frac{\eta}{c\mu(1-\eta)^2} \frac{\alpha(\xi)}{1-\alpha(\xi)} \\
+ \quad & \frac{1}{c\mu(1-\eta)} \frac{1}{(1-\alpha(\xi))^2} (-2\alpha^2(\xi) \; + \; c\mu\eta\alpha(\xi)\alpha'(\xi) \; - \; (c\mu\eta)^2(\alpha'(\xi))^2 \; + \; 2c\mu\eta\alpha'(\xi))];
\end{aligned}
$$

here $\alpha(\xi) := \alpha(c\mu(1-\eta))$ *and* $\alpha'(\xi) := \frac{\mathrm{d}}{\mathrm{d}\xi}\alpha(\xi)|_{\xi=c\mu(1-\eta)}$.

Remark 3 *In the case of the* $M/M/c$ *queue, we have* $\alpha(s) = \frac{\lambda}{\lambda+s}$ *and* $\eta = \lambda/(c\mu) = \rho$, *and (30) and (31) are seen to reduce to (21) and (22). This expression for* η *in* $M/M/c$ *is actually the reason why we did not simplify (32) slightly by dividing some numerators and denominators by* μ. *We have also refrained from regrouping terms in (32), to make it easier for the reader to verify the calculations.*
In the case of the $G/M/1$ *queue, i.e.,* $c = 1$, *the only unknown probability is* p_0, *which by normalization turns out to be* $1-\eta$. *Hence in this case,*

$$
\begin{aligned}
& \mathbb{E}[z^O] = (1-\eta)\frac{1-\alpha(\mu)}{1-\alpha(\mu)z} && (33)\\
+ \quad & z\eta(1-\eta)\frac{\mu^2}{\mu(1-\eta z)-\mu}\left[\frac{1}{\mu}\frac{1-\alpha(\mu)}{1-\alpha(\mu)z} - \frac{1}{\mu(1-\eta z)}\frac{1-\alpha(\mu(1-\eta z))}{1-\alpha(\mu(1-\eta z))z}\right].
\end{aligned}
$$

5 The $M/G/c$ queue

Consider the $M/G/c$ queue with service in order of arrival and with arrival rate λ. A generic service time is denoted by B, with distribution $B(\cdot)$ and LST $\beta(\cdot)$. Once more, Equation (17) holds. Denoting the steady-state *waiting time* by W, so that $S = W + B$, it follows from (17) that

$$\begin{aligned} \mathbb{E}[z^O] &= \mathbb{E}[z^X \mathrm{e}^{-\lambda(1-z)S}] \qquad (34)\\ &= \beta(\lambda(1-z)) \left[\sum_{k=0}^{c-1} p_k z^k + \sum_{k=c}^{\infty} p_k z^k \mathbb{E}[\mathrm{e}^{-\lambda(1-z)W} | X = k] \right]. \end{aligned}$$

Unfortunately, in the $M/G/c$ FCFS queue with $c \geq 2$, no exact expressions for the probabilities $p_k = \mathbb{P}(X = k)$ are known, and even the mean waiting time $\mathbb{E}[W]$ is unknown. Therefore we propose an approximation for $\mathbb{E}[z^O]$, based on Approximation Assumption 4.5.1 of [10]:
Suppose that right after a service completion epoch k customers are present. (i) if $k < c$, then the time until the next service completion is distributed as the minimum of k independent random variables $B_1^{res}, \ldots, B_k^{res}$, where $\mathbb{P}(B^{res} < x) = \int_0^x \frac{1-B(y)}{\mathbb{E}[B]}\, \mathrm{d}y$ (the residual service time distribution); (ii) If $k \geq c$, then the time until the next service completion is distributed as B/c.
The rationale behind this approximation is that, for $k < c$, the system behaves like an $M/G/\infty$ system (for which (i) holds), while for $k \geq c$ the system behavior resembles that of an $M/G/1$ queue with a c times as fast server (for which (ii) holds). It should be observed, though, that in reality X and the time until the next service completion are dependent, as one can already see in Equation (2). Anyway, based on the above Approximation Assumption, Tijms provides an approximation for the probabilities p_k, in Equation (4.5.1) of [10]: the p_k are assumed to satisfy

$$p_k = p_0 \frac{(\lambda/\mu)^k}{k!}, \quad k = 0, 1, \ldots, c-1, \qquad (35)$$

(like in $M/G/\infty$) and

$$p_k = \lambda a_{k-c} p_{c-1} + \lambda \sum_{m=c}^{k} b_{k-m} p_m, \quad k = c, c+1, \ldots, \qquad (36)$$

where

$$a_n = \int_0^\infty (\mathbb{P}(B^{res} > t))^{c-1} \mathbb{P}(B > t) \mathrm{e}^{-\lambda t} \frac{(\lambda t)^n}{n!}\, \mathrm{d}t, \quad n = 0, 1, \ldots, \qquad (37)$$

$$b_n = \int_0^\infty \mathbb{P}(B > ct) \mathrm{e}^{-\lambda t} \frac{(\lambda t)^n}{n!}\, \mathrm{d}t, \quad n = 0, 1, \ldots\ . \qquad (38)$$

This set of equations can be solved recursively; p_0 eventually follows by normalization.

According to the approximation assumption, we can write W, if positive, as the sum of B^{res}/c and $k-c$ independent random variables that are all distributed as B/c. Hence we propose the following approximation:

$$\mathbb{E}[z^O] \;\approx\; \beta(\lambda(1-z))\left[\sum_{k=0}^{c-1} p_k z^k + \sum_{k=c}^{\infty} p_k z^k \frac{1-\beta(\frac{\lambda}{c}(1-z))}{\mathbb{E}[B](\lambda/c)(1-z)}\beta^{k-c}(\frac{\lambda}{c}(1-z))\right], \tag{39}$$

where the p_k are determined by (35)-(38). This can be translated into a more compact expression in the following way. It follows from (36) that

$$\sum_{k=c}^{\infty} p_k z^k \;=\; \lambda p_{c-1} z^c \sum_{k=c}^{\infty} a_{k-c} z^{k-c} \;+\; \lambda \sum_{m=c}^{\infty} p_m z^m \sum_{k=m}^{\infty} b_{k-m} z^{k-m}. \tag{40}$$

Using (37) and (38), it readily follows from (40) that

$$\sum_{k=c}^{\infty} p_k z^k \;=\; \lambda p_{c-1} z^c \frac{\int_0^\infty (\mathbb{P}(B^{res} > t))^{c-1} \mathbb{P}(B > t) \mathrm{e}^{-\lambda(1-z)t}\, \mathrm{d}t}{1-\lambda\int_0^\infty \mathbb{P}(B > ct)\mathrm{e}^{-\lambda(1-z)t}\, \mathrm{d}t}. \tag{41}$$

Now observe that the second term in the righthand side of (39) contains an expression $\sum_{k=c}^{\infty} p_k z^k \beta^{k-c}(\frac{\lambda}{c}(1-z))$, and use (41) with z replaced by $z\beta(\frac{\lambda}{c}(1-z))$:

$$\begin{aligned} \mathbb{E}[z^O] \;\approx\; & \beta(\lambda(1-z))[\sum_{k=0}^{c-1} p_k z^k + \frac{1-\beta(\frac{\lambda}{c}(1-z))}{\mathbb{E}[B](\lambda/c)(1-z)} \lambda p_{c-1} z^c \times \\ \times \quad & \frac{\int_0^\infty (\mathbb{P}(B^{res} > t))^{c-1} \mathbb{P}(B > t)\mathrm{e}^{-\lambda(1-z\beta(\lambda(1-z)))t}\, \mathrm{d}t}{1-\lambda\int_0^\infty \mathbb{P}(B > ct)\mathrm{e}^{-\lambda(1-z\beta(\lambda(1-z)))t}\, \mathrm{d}t}]. \end{aligned} \tag{42}$$

Remark 4 *For $M/M/c$, the approximation for $\mathbb{E}[z^O]$ reduces to the exact expression for $\mathbb{E}[z^O]$ given in Theorem 2. Indeed, now the righthand side of (41) reduces to $\lambda p_{c-1} z^c/(\mu c - \lambda z)$, which yields that $p_k = p_{c-1}(\frac{\lambda}{c\mu})^{k-c+1}$, in agreement with (18).*

Remark 5 *An approximation for $\mathbb{E}[O]$ can be obtained by observing that (again) $\mathbb{E}[O] = \mathbb{E}[X] + \lambda\mathbb{E}[S] = 2\mathbb{E}[X]$. Several accurate approximations for $\mathbb{E}W$, and hence for $\mathbb{E}[X]$, are known; see, e.g., [3] and Section 4.5.3 of [10]. Alternatively, one could obtain an approximation for $\mathbb{E}[O]$ by differentiating the approximation for $\mathbb{E}[z^O]$ and taking $z = 1$.*

Remark 6 *A more refined approximation for $\mathbb{E}[z^O]$ might be obtained along the following lines. As we are assuming that, when $X \geq c$, the system* *behaves like an $M/G/1$ queue with generic service time B/c, we might use the $M/G/1$ result from Section 2 that gives the joint transform of X and residual service time of the customer in service, at an arrival epoch. That is likely to give a more accurate result then taking the p_k from (4.5.1) of [10] and assuming that the time until the first service completion is distributed as B^{res}/c* independent *of the value of X, apart from the given that $X \geq c$.*

Acknowledgment. This research is funded by the NWO Gravitation project NETWORKS, grant number 024.002.003.

References

[1] J.W. Cohen (1982). *The Single Server Queue.* North-Holland Publ. Cy., 2nd ed.

[2] O.J. Boxma (1984). Joint distribution of sojourn time and queue length in the $M/G/1$ queue with (in)finite capacity. *Eur. J. Oper. Res.* **16**, 246-256.

[3] O.J. Boxma, J.W. Cohen and N. Huffels (1979). Approximations of the mean waiting time in an $M/G/s$ queueing system. *Oper. Res.* **27**, 1115-1127.

[4] K. Kang, S. Doroudi, M. Delasay and A. Wickeham (2023). A queueing-theoretic framework for evaluating transmission risks in service facilities during a pandemic. *Production and Operations Management* **32**, 1453-1470.

[5] A. Mandelbaum and U. Yechiali (1979). The conditional residual service time in the $M/G/1$ queue. http://www.math.tau.ac.il/ uriy/publications (No. 30a).

[6] S. Palomo and J. Pender (2021). Measuring the overlap with other customers in the single server queue. *2021 Winter Simulation Conference (WSC)*, IEEE, 1-12.

[7] J. Pender and S. Palomo (2021). Overlap times in the infinite server queue. *arXiv:2104.14437.*

[8] Y. Perlman and U. Yechiali (2020). Reducing risk of infection – the COVID-19 queueing game. *Safety Science* **132**, 104987.

[9] K. Sigman and U. Yechiali (2007). Stationary remaining service time conditional on queue length. *Oper. Res. Letters* **35**, 581-583.

[10] H.C. Tijms (1994). *Stochastic Models – An Algorithmic Approach.* John Wiley & Sons.

[11] J. Xu, Y.M. Ko, M. Kong and J. Pender (2023). Queueing management for reducing the overlaps of customers in service systems.