Back to the roots of the M/D/s queue and the works of Erlang, Crommelin, and Pollaczek

A.J.E.M. Janssen
Philips Research. Digital Signal Processing Group
HTC-36, 5656 AE Eindhoven, The Netherlands
a.j.e.m.janssen@philips.com

J.S.H. van Leeuwaarden*
Eindhoven University of Technology and EURANDOM
P.O. Box 513 - 5600 MB Eindhoven, The Netherlands
j.s.h.v.leeuwaarden@tue.nl

On the occasion of the 10th anniversary of EURANDOM

Abstract

A.K. Erlang introduced the M/D/s queue in 1917, while F. Pollaczek and C.D. Crommelin formalized the theory using complex analysis and transforms. Let $D(s,\lambda)$ denote the stationary probability of experiencing no waiting time in the M/D/s queue with arrival rate λ and service requirement 1. We use $D(s,\lambda)$ as a vehicle to give an overview of some of the results we obtained over the last years, including explicit characterizations of the roots, the derivation of infinite series from expressions in terms of roots using Fourier sampling, and heavy-traffic limits obtained from square-root staffing. We propose to call $D(s,\lambda)$ the Erlang D formula, for which several new results are presented and compared to the results of Pollaczek.

Key Words and Phrases: M/D/s queue, Erlang D formula, queueing theory, roots, square-root staffing.

1 Introduction

The M/D/s queue, with Poisson arrivals, deterministic service times and s servers, has a deeply rooted place in queueing theory. It all started with the 1917 paper of A.K. Erlang, in which he introduced the M/D/s queue along with the M/M/s queue (Erlang delay model) and the M/M/s/s queue (Erlang loss model). All three models are of great historical interest, and the M/D/s queue in particular is illustrative for the difficulties occasioned by the absence of exponential service times. The pioneering work of Erlang, complemented by more formal works on the M/D/s queue by Pollaczek (1930a,b), and Crommelin (1932, 1934), laid the foundation of modern queueing theory and demonstrated the wealth of mathematical techniques that can be applied.

Erlang obtained expressions for the distribution of the stationary waiting time for up to three servers, while Crommelin (1932) obtained for any s the probability generation function (pgf) of the

*The research of JvL is supported by an NWO VENI grant.

stationary queue length distribution, expressed in terms of the s roots on and within the unit circle of $z^s = \exp(\lambda(z-1))$. From this pgf, Crommelin was able to obtain the stationary waiting time distribution. Two years earlier, Pollaczek (1930a,b) treated the M/D/s queue in two papers, deriving results in terms of infinite series. Pollaczek's work was difficult to read, since he relied on rather complicated analysis, so Crommelin (1934) gave an exposition of Pollaczek's theory for the M/D/s queue and found his own results in agreement with those of Pollaczek. Pollaczek too, found alternative expressions for his infinite series in terms of roots (see Pollaczek, 1930b, Eqns. (80) and (83)).

By now, queueing theory is a mature branch of applied mathematics of which Erlang is considered the founding father and Pollaczek the main pioneer of analytical methods. For Pollaczek, the M/D/s queue was merely a first step in his impressive collection of results, including investigations of the M/G/s, G/G/1 and G/G/s queue. Crommelin should be credited for introducing to queueing theory the generating function technique, which found numerous applications, including queues that permit analysis in terms of embedded processes, like bulk service queues, M/G/1 and G/M/1-type queues, and discrete or discrete-time queues. The generating function or transform technique typically leads to results in terms of implicitly defined, complex-valued roots of some equation.

We shall give an overview of some of the results we obtained over the last years. For a class of queues that contains the M/D/s queue, these results include explicit characterizations of roots, the derivation of infinite-series or Pollaczek-type results from expressions in terms of roots using Fourier sampling, and heavy-traffic results obtained from square-root staffing. In particular, we confine attention to the stationary probability that an arbitrary customer in the M/D/s queue experiences no waiting time (meets an empty queue). Henceforth, we denote this quantity by $D(s,\lambda)$, and refer to it as the Erlang-Pollaczek-Crommelin D formula, Erlang D formula in short.

Some of this work was done in, or has benefitted from, collaboration with colleagues including I.J.B.F. Adan, O.J. Boxma, D. Denteneer, J.A.C. Resing, E.M.M. Winands and B. Zwart.

2 The M/D/s queue

We consider the M/D/s queue and keep track of the number of customers waiting in the queue (without those in service) at the end of intervals equal to the constant service time (which we set to one). Customers arrive according to a Poisson process with rate λ and are served by at most s servers. Let Q_n denote the number of customers waiting in the queue at the end of interval n. The queue length process is then described by

$$Q_{n+1} = (Q_n + A_{\lambda,n} - s)^+, \quad n = 0, 1, \ldots, \tag{1}$$

where $x^+ = \max\{0, x\}$, and $A_{\lambda,n}$ denotes the number of customers that arrived at the queue during interval n. Obviously, the $A_{\lambda,n}$ are i.i.d. for all n, and copies of a Poisson random variable A_λ with mean λ. Note that due to the assumption of constant service times, the customers which are being served at the end of the considered interval should start within this interval, and for the same reason, the customers whose service is completed during this interval should start before its beginning.

Assume that $\mathbb{E}A_\lambda = \lambda < s$ and let Q denote the random variable that follows the stationary queue length distribution, i.e., Q is the weak limit of Q_n.

It is fairly straightforward to prove that $z^s = \mathrm{e}^{\lambda(z-1)}$ has s roots in $|z| \leq 1$ (using Rouché's theorem, see Sec. 3). Denote these roots by $z_0 = 1, z_1, \ldots, z_{s-1}$. For the pgf of Q, Crommelin derived the expression

$$Q(z) := \mathbb{E}(z^Q) = \frac{(z-1)(s-\lambda)}{z^s - \mathrm{e}^{\lambda(z-1)}} \prod_{k=1}^{s-1} \frac{z - z_k}{1 - z_k}, \tag{2}$$

from which it immediately follows that

$$\mathbb{P}(Q=0) = \mathrm{e}^{\lambda}(-1)^{s-1}(s-\lambda)\prod_{k=1}^{s-1}\frac{z_k}{1-z_k}. \tag{3}$$

Pollaczek, on the other hand, derived the expression

$$\mathbb{P}(Q=0) = \exp\Big\{-\sum_{l=1}^{\infty}\frac{1}{l}\sum_{m=1}^{\infty}\mathrm{e}^{-l\lambda}\frac{(l\lambda)^{ls+m}}{(ls+m)!}\Big\}. \tag{4}$$

In his derivation of (4), Pollaczek employed the following cyclic scheduling policy: Suppose the servers are numbered 1 to s and free servers are assigned in numerical order; then for first-come-first-served, if server i serves arrival j, it also serves arrivals $s+j,\ 2s+j,\ \ldots$. Since waiting times are independent of this assignment, the waiting time distribution is the same as that of a single server serving every sth arrival of a Poisson input, or rather, Erlang input. For the latter queueing system, the Laplace-Stieltjes transform of the waiting time is known to be (see e.g. Riordan, 1962, p. 118)

$$W(u) \ := \ \mathbb{E}(\mathrm{e}^{-uW}) \ = \ \frac{u(s-\lambda)\lambda^{s-1}}{\lambda^s\mathrm{e}^{-u}-(\lambda-u)^s}\prod_{k=1}^{s-1}\Big(1-\frac{u}{u_k}\Big), \tag{5}$$

where $u_k = \lambda(1-z_k)$. Distributional Little's law $W(\lambda(1-z)) = Q(z)$ and (2) give the same result. Cyclic scheduling policies play an important role in Franx (2001) and Jelenković et al. (2004). It immediately follows from (5) that the Erlang D formula is given by (Crommelin, 1932, Eq. (10) and Pollaczek, 1930b, Eq. (80))

$$D(s,\lambda) = \mathbb{P}(W=0) = \frac{s-\lambda}{\prod_{k=1}^{s-1}(1-z_k)}, \tag{6}$$

whereas Pollaczek's infinite-series counterpart (Pollaczek, 1930b, Eq. (83)) reads

$$D(s,\lambda) = \exp\Big\{-\sum_{l=1}^{\infty}\frac{1}{l}\sum_{m=0}^{\infty}\mathrm{e}^{-l\lambda}\frac{(l\lambda)^{ls+m}}{(ls+m)!}\Big\}. \tag{7}$$

In comparing (3), (4), (6) and (7), it should hold that

$$\prod_{k=1}^{s-1} z_k = (-1)^{s-1}\mathrm{e}^{-\lambda}\exp\Big\{\sum_{l=1}^{\infty}\frac{1}{l}\mathrm{e}^{-l\lambda}\frac{(l\lambda)^{ls}}{(ls)!}\Big\}, \tag{8}$$

and

$$\frac{s-\lambda}{\prod_{k=1}^{s-1}(1-z_k)} = \exp\Big\{-\sum_{l=1}^{\infty}\frac{1}{l}\sum_{m=0}^{\infty}\mathrm{e}^{-l\lambda}\frac{(l\lambda)^{ls+m}}{(ls+m)!}\Big\}. \tag{9}$$

Identities of this type were investigated by Crommelin, 1934, Appendix 1, in order to convince himself that his results agreed with those of Pollaczek. Direct proofs of (8) and (9) were also given in Janssen and van Leeuwaarden, 2005b, pp. 380-382 (for a larger class of queues). Hereto, we derived explicit expressions for the roots and treated the products occurring in the left-hand sides of (8) and (9) as Fourier aliasing series with terms given in analytic form, eventually leading to the right-hand sides of (8) and (9). The explicit expressions for the roots are discussed in the next section.

3 Roots in queueing theory

Crucial in applying the generating function technique is the investigation of the roots of some given equation, and usually these roots have no explicit representation. Crommelin (1932) and Pollaczek (1930b) used Rouché's theorem to prove the existence of such roots. We consider the roots in the unit disk of

$$z^s = A(z), \tag{10}$$

where $A(z)$ is the pgf of a discrete random variable A, with $A(0) > 0$, $\rho = A'(1)/s < 1$, and $s \in \mathbb{N}$.

In order to apply Rouché's theorem it is required that $A(z)$ has a radius of convergence larger than one, which is not true in general. For Crommelin this was obviously not an issue, since $A(z) = \exp(\lambda(z-1))$ is an entire function. A problem does occur when $A(z)$ is the pgf of a random variable with a heavy-tailed distribution (causing $A(z)$ to be non-analytic at $z = 1$), like the discrete Pareto distribution or Haight's zeta distribution, see Adan et al. (2006).

Define the period ν of a series $\sum_{-\infty}^{\infty} b_j z^j$ as the largest integer for which $b_j = 0$ whenever j is not divisible by ν. In Adan et al. (2006) the following result was established.

Theorem 3.1 *Let $z^s - A(z)$ have period ν. Then $z^s = A(z)$ has ν roots on the unit circle and exactly $s - \nu$ zeros in $|z| < 1$.*

The remainder of this section will deal with explicit expressions for the roots of (10). All results stem from Janssen and van Leeuwaarden (2005a,b). We first note that all roots of $z^s = A(z)$ lie on the curve

$$\mathcal{S}_{A,s} := \{z \in \mathbb{C} \mid |z| \le 1,\ |A(z)| = |z|^s\}.$$

Condition 3.2 *$\mathcal{S}_{A,s}$ is a Jordan curve with 0 in its interior, and $A(z)$ is zero-free on and inside $\mathcal{S}_{A,s}$.*

Let $C_{z^j}[f(z)]$ denote the coefficient of z^j in $f(z)$. We established the following result.

Theorem 3.3 *If Condition 3.2 is satisfied, we have the parametrization*

$$\mathcal{S}_{A,s} = \{\tilde{z}(\mathrm{e}^{\mathrm{i}\alpha}) \mid \alpha \in [0, 2\pi]\},$$

with

$$\tilde{z}(w) = \sum_{l=1}^{\infty} \frac{1}{l} C_{z^{l-1}}[A^{l/s}(z)] w^l \tag{11}$$

a power series with radius of convergence larger than one.

Theorem 3.3 thus implies that, under Condition 3.2, the roots of (10) in the unit disk have the explicit representations $z_k = \tilde{z}(w_k)$, $k = 0, 1, ..., s-1$ with $w_k = \exp(2\pi i k/s)$. Most functions $A(z)$ will satisfy Condition 3.2, although counterexamples can be constructed (see van Leeuwaarden, 2005, p. 57). For the Poisson case $A(z) = \exp(\lambda(z-1))$ with $0 \le \lambda < s$, Condition 3.2 is always satisfied, and hence (11) yields

$$z_k = \tilde{z}(w_k) = \sum_{l=1}^{\infty} \mathrm{e}^{-l\rho} \frac{(l\rho)^{l-1}}{l!} w_k^l, \quad k = 0, 1, ..., s-1. \tag{12}$$

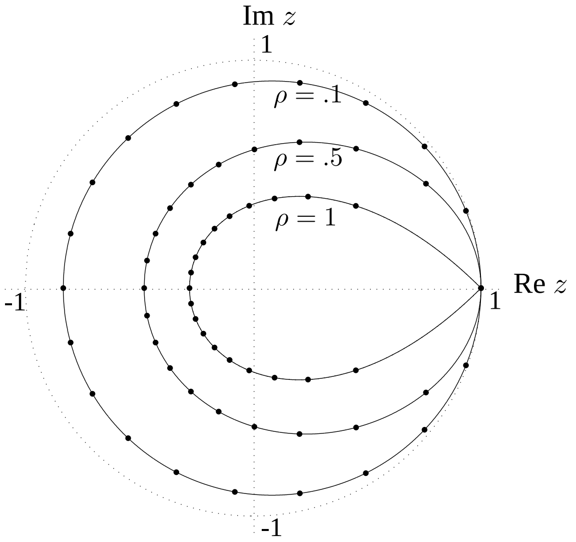

Figure 1: $\mathcal{S}_{A,s}$ for the Poisson case.

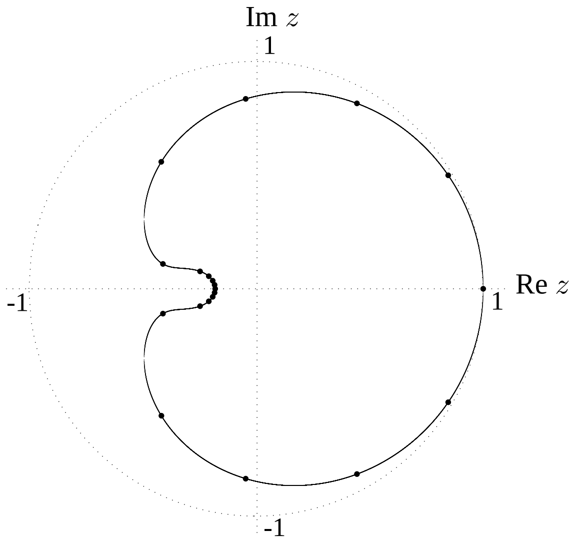

Figure 2: $\mathcal{S}_{A,s}$ for the binomial case.

When $A(z)$ is assumed to have no zeros in $|z| \leq 1$, we know that the s roots of $z^s = A(z)$ in $|z| \leq 1$ satisfy

$$z = wA^{1/s}(z), \quad w^s = 1. \tag{13}$$

For each feasible w, Eq. (13) can be shown to have one unique root in $|z| \leq 1$, and one could try to solve the equations by the successive substitutions

$$z_k^{(n+1)} = w_k A^{1/s}(z_k^{(n)}), \quad k = 0, 1, ..., s-1, \tag{14}$$

with starting values $z_k^{(0)} = 0$.

Lemma 3.4 *When for* $|z| \leq 1$, $A(z)$ *is zero-free and* $|\frac{\mathrm{d}}{\mathrm{d}z} A^{1/s}(z)| < 1$, *the fixed point equations* (14) *converge to the desired roots.*

A necessary condition for convergence of the scheme in (14) can be formulated as follows.

Condition 3.5 *Condition 3.2 should be satisfied and for all points* $z \in \mathcal{S}_{A,s}$ *there should hold that* $|\frac{\mathrm{d}}{\mathrm{d}z} A^{1/s}(z)| < 1$.

Condition 3.2 is much weaker than Condition 3.5, and hence (12) is applicable for a larger class of A than (14). Nevertheless, for $A(z) = \exp(\lambda(z-1))$, it is readily seen that $A(z) \neq 0$ and $|\frac{\mathrm{d}}{\mathrm{d}z} A^{1/s}(z)| < 1$ for $|z| \leq 1$, so that the scheme

$$z_k^{(n+1)} = w_k \exp\{\rho(z_k^{(n)} - 1)\}, \quad k = 0, 1, ..., s-1, \tag{15}$$

with starting values $z_k^{(0)} = 0$, converges to the desired roots. Figure 1 depicts $\mathcal{S}_{A,s}$ for $\rho = 0.1, 0.5, 1.0$. The dots on the curves indicate the roots z_k for the case $s = 20$ that can be determined from either (12) or (15).

While Condition 3.2 implies that $\mathcal{S}_{A,s}$ is a closed curve without double points, Condition 3.5 apparently does not hold for all such curves. Condition 3.5 can be compared with the notions convexity and starshapedness from the geometric theory of univalent functions.

Definition 3.6 *(i) A closed curve without double points is called starshaped with respect to a point in its interior if any ray from this point intersects the curve at exactly one point,*

(ii) A closed curve without double points is called convex when it is starshaped with respect to any point in its interior.

The following intriguing result then holds.

Theorem 3.7

$\mathcal{S}_{A,s}$ *is convex* $\Rightarrow$ *Condition* 3.5 *holds* $\Rightarrow$ $\mathcal{S}_{A,s}$ *is starshaped with respect to* 0.

3.1 An illustrative example

Consider the binomial case $A(z) = (p + qz)^n$ where $p,\, q \geq 0$, $p + q = 1$ and $A'(1) = nq < s$. We compute in this case

$$z_k = \sum_{l=1}^{\infty} \frac{1}{l}\, p^{l\sigma - l + 1}\, q^{l-1} \binom{l\sigma}{l-1} w_k^l, \quad k = 0, 1, ..., s-1, \tag{16}$$

where $\sigma := n/s$. In can be shown that the coefficients have exponential decay for $\sigma \geq 1$ and that, for $0 \leq \sigma < 1$, the coefficients have exponential decay if and only if

$$p^{\sigma-1} q (1-\sigma)^{1-\sigma} \sigma^{\sigma} < 1. \tag{17}$$

For $\sigma = 1/2$, $s = 20$, constraint (17) requires q to be less than $2(\sqrt{2} - 1)$. In Fig. 2 we plotted the $\mathcal{S}_{A,s}$ for $q = 0.82 < 2(\sqrt{2} - 1)$, and the dots indicate the roots z_k obtained by calculating the sum in (12) up to $l = 50$. When q is increased beyond $q > 2(\sqrt{2} - 1)$, $\mathcal{S}_{A,s}$ turns from a smooth Jordan curve containing zero into two separate closed curves, and (12) no longer holds.

Here is a demonstration of Thm. 3.7. From an inspection of $\mathcal{S}_{A,s}$ in Fig. 2 one sees that $\mathcal{S}_{A,s}$ is not starshaped with respect to 0, and one can thus immediately conclude that Condition 3.5 is not satisfied and hence the iteration (14) cannot be applied to determine the roots.

When $\sigma \geq 1$, we have that

$$\max_{z \in \mathcal{S}_{A,s}} |\tfrac{\mathrm{d}}{\mathrm{d}z} A^{1/s}(z)| = \max_{z \in \mathcal{S}_{A,s}} |\sigma q (p + qz)^{\sigma - 1}|$$

occurs at $z = 1$ and equals $\sigma q < 1$. For $\sigma \in (0, 1)$, it can be shown that Condition 3.5 holds if and only if

$$p^{\sigma-1} q (1+\sigma)^{1-\sigma} \sigma^{\sigma} < 1. \tag{18}$$

Now denote by $q_1(\sigma)$ and $q_2(\sigma)$ the suprema of q for which (17) and (18) hold, respectively. These values $q_1(\sigma)$ and $q_2(\sigma)$ are plotted in Fig. 3 as a function of $\sigma \in [0, 1]$. Observe that the set of values q for which (16) holds ($q < q_1(\sigma)$) is much larger than the set for which Condition 3.5 holds ($q < q_2(\sigma)$). To compare the condition of convexity and Condition 3.5 we just consider the case $\sigma = 1/2$: when $2 - \sqrt{2} < q < 2/3$ we have that $\mathcal{S}_{A,s}$ satisfies Condition 3.5 while $\mathcal{S}_{A,s}$ is not convex (since it is not starshaped with respect to points near the intersection point with the negative real axis).

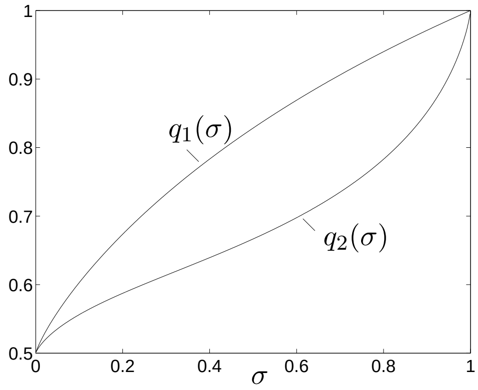

Figure 3: $q_1(\sigma)$ and $q_2(\sigma)$ for $\sigma \in (0,1)$.

4 New expressions for the Erlang D formula

We now give some new expressions for the Erlang D formula. All proofs are relegated to the appendix.

Theorem 4.1

$$D(s,\lambda) = \exp\Big\{ -\sum_{l=1}^{\infty} \frac{p(ls)}{l} \frac{\sqrt{ls}\mathrm{e}^{-l\alpha^2/2}}{\sqrt{2\pi}} \sum_{m=0}^{\infty} \frac{m^m \mathrm{e}^{-m}}{m!} \frac{\mathrm{e}^{-m\alpha^2/2s}}{m+ls} \Big\}, \tag{19}$$

with

$$\alpha \;=\; (-2s(1-\rho+\ln\rho))^{1/2}, \quad \alpha > 0, \tag{20}$$

and $p(n) = n^n \mathrm{e}^{-n}\sqrt{2\pi n}/n!$.

At first glance, (19) and (7) seem very different. Several basic manipulations of (7) lead to the following expression for $D(s,\lambda)$ that shows some more resemblance with (19):

$$D(s,\lambda) = \exp\Big\{ -\sum_{l=1}^{\infty} \frac{1}{l} \mathrm{e}^{-\frac{1}{2}\alpha^2 l} \sum_{m=0}^{\infty} \frac{p(ls+m)}{\sqrt{2\pi(ls+m)}} \Big(\Big(1-\frac{m}{m+ls}\Big)\mathrm{e}^{\frac{m}{m+ls}}\Big)^{m+ls} \Big(\frac{\lambda}{s}\Big)^m \Big\}. \tag{21}$$

From (21) we see that the series over m has a convergence rate $\frac{\lambda}{s}$. We can also bring (19) in a form that shows more resemblance with (7).

Theorem 4.2

$$D(s,\lambda) = \exp\Big\{ -\sum_{l=1}^{\infty} \frac{1}{l} \sum_{m=0}^{\infty} \mathrm{e}^{-l\lambda} \frac{(l\lambda)^{m+ls}}{(m+ls)!} \psi(l,m) \Big\}, \tag{22}$$

where

$$\psi(l,m) = \frac{(m+ls)!}{(ls)!m!} \frac{m^m}{(ls)^m} \frac{ls}{m+ls} \mathrm{e}^{-\lambda m/s}.$$

Indeed, comparing the new expression (22) with Pollaczek's result (7) shows close formal agreement, the difference beween the two being embodied by the factor ψ.

5 Square-root staffing

Being aware of the rather heavy analysis involved in their solutions to the M/D/s queue, Erlang, Crommelin and Pollaczek each developed some approximations. After all, it was about the modeling of real telephone exchanges, and both the implicitly defined complex-valued roots and the infinitely many convolutions were–at that time–not easy to get grip on. Since the computational burden increased with the number of servers, approximations would be particularly useful for large systems. Pollaczek, 1930b, Eqn. (84a), obtained for large s,

$$D(s,\lambda) = 1 - \frac{1}{1-\rho}\frac{(\rho \mathrm{e}^{1-\rho})^s}{\sqrt{2\pi s}}\left(1 + \mathcal{O}(s^{-1})\right). \tag{23}$$

The resulting approximation is useful in scaling the number of servers while ρ is kept constant. Pollaczek, 1946, p. 28 comments on (23) as follows:

> Cette formule approximative devient inutilisable dans le cas le plus important où, le nombre s des lignes parallèles étant grand, le coefficient de rendement ρ tend vers l'unité, c'est-à-dire où, pour un grand faisceau de lignes, l'on tend à approcher de l'état idéal d'une utilisation parfaite.

Pollaczek then proceeds to propose to scale the system such that $\rho = 1 - \gamma/\sqrt{s}$, with γ kept constant, for which he proves that

$$D(s,\lambda) = \frac{1}{2\pi \mathrm{i}}\oint_C \log\left(1 - e^{z^2/2+\gamma z}\right)\frac{\mathrm{d}z}{z} + \mathcal{O}(s^{-1}), \tag{24}$$

where C is a contour to the left of and parallel to the imaginary axis.

Pollaczek is, like on many occasions, ahead of his time. The scaling $\rho = 1 - \gamma/\sqrt{s}$, $s \to \infty$, is nowadays known as square-root staffing. It got immensely popular due to its application to call centers (see Borst et al. (2004); Janssen et al. (2008)), the modern counterparts of telephone exchanges. An equivalent scaling that we will use in the sequel is obtained from setting $s = \lambda + \beta\sqrt{\lambda}$. The parameter α, β and γ are closely related, as can be seen from

$$\tfrac{1}{2}\alpha^2 = s\sum_{n=2}^{\infty}\frac{(1-\rho)^n}{n}, \quad \beta = \frac{s-\lambda}{\sqrt{\lambda}}, \quad \gamma = \frac{s-\lambda}{\sqrt{s}} = \beta\rho^{\frac{1}{2}}.$$

Note that $\alpha \approx \sqrt{s}(1-\rho) = \gamma \approx \beta$ for large values of s.

Denote by Q_λ the random variable Q for the M/D/s queue with $s = \lambda + \beta\sqrt{\lambda}$ and $\beta > 0$ fixed. From (1) we know that Q_λ satisfies

$$Q_\lambda \stackrel{d}{=} (Q_\lambda + A_\lambda - s)^+,$$

where A_λ again denotes a Poisson random variable with mean λ. In can be shown that $Q_\lambda/\sqrt{\lambda}$, as $\lambda \to \infty$, converges in distribution to the maximum M_β of the Gaussian random walk with drift $-\beta$ (see Janssen et al. (2007); Jelenković et al. (2004)). The Gaussian random walk is defined by the process $\{S_n : n \geq 0\}$ with $S_0 = 0$, $S_n = X_1 + \ldots + X_n$ and $X_1, X_2, \ldots$ independent and normally distributed random variables with mean $-\beta < 0$ and variance 1. Among other things, this implies the following limiting result for the Erlang D formula (see Jelenković et al. (2004); Janssen et al. (2007)).

Corollary 5.1

$$\lim_{\lambda\to\infty} D(\lambda+\beta\sqrt{\lambda},\lambda) = \mathbb{P}(M_\beta = 0), \quad \beta > 0. \tag{25}$$

Note that this also follows from (24) if one recognizes the integral on the right-hand side of (24) as $\mathbb{P}(M_\gamma = 0)$. Pollaczek (1946), right after introducing (24), remarks:

> ...et la dernière intégrale, qui se rapporte à la théorie de la fonction ζ de Riemann, peut aisément être développée de diverses manières en série suivant de simples fonctions de γ. C'est ce paramètre γ qui, pour les grandes valeurs de s, évalue dans quelle mesure l'utilisation d'un groupe de guichets, ou d'un faisceau de lignes mises en parallèle, s'approche du cas idéal $\rho = 1$.

Indeed, Pollaczek, 1931, Eqn. (45) gives the expression

$$\mathbb{P}(M_\beta = 0) = \sqrt{2}\beta \exp\Big\{\sum_{r=0}^{\infty} \frac{(2r)!(2\pi)^{-r-\frac{1}{2}}\sin(\frac{r\pi}{2}+\frac{\pi}{4})\zeta(\frac{1}{2}+r)}{2^{2r-1}(r!)^2(2r+1)}\Big(\frac{\beta}{\sqrt{2}}\Big)^{2r+1}\Big\},$$

valid for $\beta < 2\sqrt{\pi}$. Riemann's relation for ζ and some rewriting yields

$$\mathbb{P}(M_\beta = 0) = \sqrt{2}\beta \exp\Big\{\frac{\beta}{\sqrt{2\pi}}\sum_{r=0}^{\infty}\frac{\zeta(\frac{1}{2}-r)}{r!(2r+1)}\Big(\frac{-\beta^2}{2}\Big)^r\Big\}, \quad \beta < 2\sqrt{\pi}, \tag{26}$$

the form in which the result was derived by Chang and Peres, 1997, Thm. 1.1 on p. 788. A third proof of (26) was presented by Janssen and van Leeuwaarden (2007b,a), along with the derivation of similar expressions for the cumulants of M_β and analytic continuations that hold for all $\beta > 0$. Expression (26) is particularly useful for small values of β, in which case it can be complemented by the bounds

$$\begin{aligned}
\mathbb{P}(M_\beta = 0) &\leq 2\sqrt{1-\mathrm{e}^{-\beta^2/2}}\ \exp\Big\{-\frac{\beta}{\sqrt{\pi}}+\frac{1}{8}\beta^2\Big\},\\
\mathbb{P}(M_\beta = 0) &\geq 2\sqrt{1-\mathrm{e}^{-\beta^2/2}}\ \exp\Big\{-\frac{3\beta}{2\sqrt{2\pi}}+\frac{1}{8}\beta^2-\frac{\beta^3}{9\sqrt{2\pi}}\Big\},
\end{aligned}$$

that are valid for $0 < \beta \leq \sqrt{2/\pi}$. Also, for small β and large systems, we have $D(s,\lambda) \approx \sqrt{2}\beta$ and hence the number of servers for which $D(s,\lambda) = q \in (0,1)$ is approximately

$$s \approx \lambda + \frac{q}{\sqrt{2}}\sqrt{\lambda}.$$

Because the Gaussian random walk is the limiting process of the M/D/s queue length process, $\mathbb{P}(M_\beta = 0)$ serves as an approximation to $D(s,\lambda)$ in heavy-traffic (when $s = \lambda+\beta\sqrt{\lambda}$ and $\lambda, s \to \infty$). Real systems, though, have finite λ and s, and so it is interesting to find refinements to the heavy-traffic limit $\mathbb{P}(M_\beta = 0)$ in order to construct sharper approximations for $D(s,\lambda)$. These refinements follow from a result obtained in Janssen et al. (2007).

Theorem 5.2

$$-\ln \mathbb{P}(Q_\lambda = 0) \sim \frac{1}{\sqrt{2\pi}}\sum_{k=0}^{\infty} p_k s^{-k+1/2} G_{-(k+1)}(\alpha/\sqrt{s}), \tag{27}$$

where

$$G_k(a) = \sum_{l=1}^{\infty} l^{k+1/2} \int_a^{\infty} e^{-\frac{1}{2}lsx^2} y'(x)\mathrm{d}x,$$

$$p(n) = \frac{n^n \mathrm{e}^{-n}\sqrt{2\pi n}}{n!} \sim 1 - \frac{1}{12n} + \frac{1}{288n^2} + \ldots = \sum_{k=0}^{\infty} \frac{p_k}{n^k}, \quad n \to \infty,$$

and y defined implicitly as the solution to $-y - \ln(1-y) = \frac{1}{2}x^2$.

The expansion in (27) provides sharp approximations, and in most cases, one term suffices to get accurate results, so that

$$\mathbb{P}(Q_\lambda = 0) \approx \exp\Big\{ -\frac{\sqrt{s}}{\sqrt{2\pi}} G_{-1}(\alpha/\sqrt{s}) \Big\}. \tag{28}$$

In Janssen et al. (2007) we showed that y admits the power series representation $y(x) = \sum_{n=1}^{\infty} a_n x^n$, $|x| < 2\sqrt{\pi}$, with

$$a_1 = 1,\ a_2 = -\frac{1}{3},\ a_3 = \frac{1}{36},\ a_4 = \frac{1}{270},\ a_5 = \frac{1}{4320}.$$

From (4) and (7) we see that

$$-\ln D(s,\lambda) = -\ln \mathbb{P}(Q_\lambda = 0) + \sum_{l=1}^{\infty} \frac{1}{l} \mathrm{e}^{-l\lambda} \frac{(l\lambda)^{ls}}{(ls)!}. \tag{29}$$

Combining (28) with $y'(x) \approx 1 - \frac{2}{3}x$ and (29) yields the approximation

$$D(s,\lambda) \approx \mathbb{P}(M_\alpha = 0) \cdot \exp\Big\{ \frac{-1}{3\sqrt{2\pi s}} \sum_{l=1}^{\infty} \frac{1}{l^{3/2}} e^{-\frac{1}{2}\alpha^2 l} \Big\}.$$

Many other results of this type were derived in Janssen et al. (2007), along with series expansions for the functions G_k with terms that involve the Riemann zeta function.

A Proofs of the results in Section 4

A.1 Proof of (19)

In Janssen et al. (2007), Theorem 1, it is proved that

$$-\ln \mathbb{P}(Q = 0) = \sum_{l=1}^{\infty} \frac{p(ls)}{l} \frac{1}{\sqrt{2\pi}} \int_{\alpha\sqrt{l}}^{\infty} \mathrm{e}^{-\frac{1}{2}x^2} y'(x/\sqrt{ls})\mathrm{d}x, \tag{30}$$

with α as in (20) and y defined implicitly as the solution to $-y - \ln(1-y) = \frac{1}{2}x^2, x \in \mathbb{C}$. For the sake of completeness, we shortly repeat the derivation of (30) in Appendix A.3. From Janssen et al. (2007), Lemma 16, we have

$$y(x) = 1 - \sum_{m=1}^{\infty} \frac{m^{m-1}}{m!\mathrm{e}^m} \mathrm{e}^{-\frac{1}{2}mx^2}, \quad |\arg(x)| \le \pi/4,$$

from which we find that

$$\frac{1}{\sqrt{2\pi}}\int_{\alpha\sqrt{l}}^{\infty}\mathrm{e}^{-\frac{1}{2}x^2}y'(x/\sqrt{ls})\mathrm{d}x = \frac{\sqrt{ls}}{\sqrt{2\pi}}\int_{\alpha/\sqrt{s}}^{\infty}\mathrm{e}^{-\frac{1}{2}lsx^2}y'(x)\mathrm{d}x$$
$$= \frac{\sqrt{ls}}{\sqrt{2\pi}}\sum_{m=1}^{\infty}\frac{m^m\mathrm{e}^{-\frac{1}{2}(m+ls)\alpha^2/s}}{m!\mathrm{e}^m(m+ls)} = \frac{\sqrt{ls}}{2\pi}\mathrm{e}^{-\frac{1}{2}\alpha^2 l}\sum_{m=1}^{\infty}\frac{p(m)}{\sqrt{m}}\frac{\mathrm{e}^{-m\alpha^2/2s}}{m+ls}. \quad (31)$$

Combining (31) with (29) and $\mathrm{e}^{-l\lambda}\frac{(l\lambda)^{ls}}{(ls)!} = \frac{p(ls)}{\sqrt{2\pi ls}}\mathrm{e}^{-l\alpha^2/2}$ completes the proof.

A.2 Proof of (22)

(19) can be written as

$$-\ln D(s,\lambda) = \frac{1}{2\pi}\sum_{l=1}^{\infty}\frac{1}{l}\sum_{m=0}^{\infty}p(ls)p(m)\sqrt{\frac{ls}{m}}\frac{\mathrm{e}^{-\frac{1}{2}(m+ls)\alpha^2/s}}{m+ls}. \quad (32)$$

Using

$$\frac{1}{2\pi}p(ls)p(m)\sqrt{\frac{ls}{m}} = \frac{(ls)^{ls+1}m^m\mathrm{e}^{-m-ls}}{(ls)!m!}$$

and

$$\mathrm{e}^{-\frac{1}{2}(m+ls)\alpha^2/s} = \rho^{m+ls}\mathrm{e}^{m+ls}\mathrm{e}^{-l\lambda}\mathrm{e}^{-m\rho},$$

the result follows from (32) upon some rearranging of terms.

A.3 Proof of (30)

For $n = 0, 1, \ldots$ we let

$$s_n(z) = \sum_{k=0}^{n}\frac{z^k}{k!}, \quad z \in \mathbb{C}.$$

With $\rho = \lambda/s$ and $n = ls$ (so that $\lambda l = n\rho$), and

$$q(\xi) = e^{1-\xi}\xi, \quad \xi \in \mathbb{C},$$

we have from Szegö, 1922, p. 50 (or Abramowitz and Stegun, 1970, 6.5.13 on p. 262),

$$\sum_{j=n+1}^{\infty}e^{-l\lambda}\frac{(l\lambda)^j}{j!} \;=\; 1 - e^{-\lambda l}s_n(\lambda l) \;=\; \frac{n^{n+1}e^{-n}}{n!}\int_0^{\rho}q^n(\xi)\mathrm{d}\xi.$$

Using this relation we can rewrite (4) as

$$-\ln\mathbb{P}(Q=0) \;=\; s^{1/2}\sum_{l=1}^{\infty}\frac{p(ls)}{\sqrt{2\pi l}}\int_0^{\rho}q^{ls}(\xi)\mathrm{d}\xi, \quad (33)$$

with $p(n)$ as defined in Thm. 4.1. We then consider the equation

$$f(y) := -\ln q(1-y) = \tfrac{1}{2}x^2, \quad x \in \mathbb{C}, \quad (34)$$

from which y is to be solved. We note that

$$f(y) = \tfrac{1}{2}y^2 + \tfrac{1}{3}y^3 + \tfrac{1}{4}y^4 + \ldots,$$

whence there is a solution $y(x)$ that is analytic around $x = 0$ and that satisfies $y(x) = x + \mathcal{O}(x^2)$ as $x \to 0$. Furthermore, since f increases from 0 to ∞ as y increases from 0 to 1, we have that $y(x)$ increases from 0 to ∞, and for any $x \geq 0$ there is a unique non-negative solution $y = y(x)$ of (34). Furthermore, with α as in (20), it holds that $q^{ls}(\rho) = e^{-\frac{1}{2}\alpha^2 l}$, and

$$\int_0^\rho q^{ls}(\xi)\mathrm{d}\xi = \frac{1}{\sqrt{ls}} \int_{\alpha\sqrt{l}}^\infty e^{-\frac{1}{2}x^2} y'(x/\sqrt{ls})\mathrm{d}x. \tag{35}$$

Substituting (35) into (33) yields (30).

References

M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions*. Dover, New York, 1970.

I.J.B.F. Adan, J.S.H. van Leeuwaarden, and E.M.M. Winands. On the application of Rouché's theorem in queueing theory. *Operations Research Letters*, 34:355–360, 2006.

S. Borst, A. Mandelbaum, and M. Reiman. Dimensioning large call centers. *Operations Research*, 52:17–34, 2004.

J.T. Chang and Y. Peres. Ladder heights, Gaussian random walks and the Riemann zeta function. *Annals of Probability*, 25:787–802, 1997.

C.D. Crommelin. Delay probability formulae when the holding times are constant. *Post Office Electrical Engineers Journal*, 25:41–50, 1932.

C.D. Crommelin. Delay probability formulae. *Post Office Electrical Engineers Journal*, 26:266–274, 1934.

G.J. Franx. A simple solution for the M/D/c waiting time distribution. *Operations Research Letters*, 29:221–229, 2001.

A.J.E.M. Janssen and J.S.H. van Leeuwaarden. Cumulants of the maximum of the Gaussian random walk. *Stochastic Processes and their Applications*, 117:1928–1959, 2007a.

A.J.E.M. Janssen and J.S.H. van Leeuwaarden. Analytic computation schemes for the discrete-time bulk service queue. *Queueing Systems*, 50:141–163, 2005a.

A.J.E.M. Janssen and J.S.H. van Leeuwaarden. A discrete queue, Fourier sampling on Szegö curves, and Spitzer formulas. *International Journal of Wavelets, Multiresolution and Information Processing*, 3:361–387, 2005b.

A.J.E.M. Janssen and J.S.H. van Leeuwaarden. On Lerch's transcendent and the Gaussian random walk. *Annals of Applied Probability*, 17:421–439, 2007b.

A.J.E.M. Janssen, J.S.H. van Leeuwaarden, and B. Zwart. Corrected asymptotics for a multi-server queue in the Halfin-Whitt regime. *Accepted for Queueing Systems*, 2007.

A.J.E.M. Janssen, J.S.H. van Leeuwaarden, and B. Zwart. Refining square root staffing by expanding Erlang C. *Accepted for Operations Research*, 2008.

P. Jelenković, A. Mandelbaum, and P. Momčilović. Heavy traffic limits for queues with many deterministic servers. *Queueing Systems*, 47:54–69, 2004.

F. Pollaczek. Über eine Aufgabe der Wahrscheinlichkeitstheorie. I. *Mathematische Zeitschrift*, 32: 64–100, 1930a.

F. Pollaczek. Über eine Aufgabe der Wahrscheinlichkeitstheorie. II. *Mathematische Zeitschrift*, 32: 729–750, 1930b.

F. Pollaczek. Sur l'application de la théorie des fonctions au calcul de certaines probabilités continues utilisées dans la théorie des réseaux téléphoniques. *Annales de l'Institut Henri Poincaré*, 10:1–55, 1946.

F. Pollaczek. Über zwei Formeln aus der Theorie des Wartens vor Schaltergruppen. *Elektrische Nachrichtentechnik*, 8:256–268, 1931.

J. Riordan. *Stochastic Service Systems*. Wiley, New York, 1962.

G. Szegö. Über eine Eigenschaft der Exponentialreihe. *Sitzungsberichte der Berliner Math. Gesellschaft*, 22:50–64, 1922.

J.S.H. van Leeuwaarden. *Queueing Models for Cable Access Networks*. Ph.D. thesis, Eindhoven University of Technology, The Netherlands, 2005.