THE DISTRIBUTIONAL LITTLE'S LAW AND ITS APPLICATIONS

DIMITRIS BERTSIMAS and DAISUKE NAKAZATO

Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received March 1991; revisions received January, June, October 1992; accepted March 1993)

This paper discusses the distributional Little's law and examines its applications in a variety of queueing systems. The distributional law relates the steady-state distributions of the number in the system (or in the queue) and the time spent in the system (or in the queue) in a queueing system under FIFO. We provide a new proof of the distributional law and in the process we generalize a well known theorem of Burke on the equality of pre-arrival and postdeparture probabilities. More importantly, we demonstrate that the distributional law has important algorithmic and structural applications and can be used to derive various performance characteristics of several queueing systems which admit distributional laws. As a result, we believe that the distributional law is a powerful tool for the derivation of performance measures in queueing systems and can lead to a certain unification of queueing theory.

One important trend in the queueing theory literature has been the development of laws that connect fundamental quantities of queueing systems. Probably the most well known result in this direction is that the steady-state number of customers in the system (or queue), denoted by L, and the waiting time W of a customer in the system (or queue) obey Little's law (1961): $E[L] = \lambda E[W]$, where λ is the arrival rate. Generalizations of Little's law include $E[H] = \lambda E[G]$ due to Stidham (1970) and Brumelle (1971) and the rate conservation laws of Miyazawa (1979, 1985). $E[H] = \lambda E[G]$ and the rate conservation laws have been shown to be essentially equivalent (see, for example, Sigman 1991). For a thorough survey of Little's law and its extensions the reader is referred to the paper by Whitt (1991).

Although the previous laws are important structural results, they do not address how the *distributions* of L and W are related, or how one can find $E[L]$, $E[W]$ or, even more ambitiously, the distributions of L and W. As a result, to find the distributions of L and W in specific queueing systems, researchers exploit the particular structure of the system.

It is well known that although $E[W]$ is invariant under a wide variety of service disciplines, the distribution of W is not. Therefore, to find a useful relation between the distributions of L and W, one has to specify the service discipline. Haji and Newell (1971) address the issue of relating the distributions of L and W, when the discipline of the queueing system is first in, first out (FIFO), and prove a general distributional law (see Theorem 1). Miyazawa (1979) and Franken et al. (1982, pp. 110–111) discuss these distributional laws further and offer additional insights. Keilson and Servi (1988) discuss these laws for the special case of a Poisson arrival process. Whitt (1991, subsection 8.4) contains a nice discussion of the distributional law.

Our goal in this paper is to understand the distributional laws further and demonstrate that although the distributional laws are more restrictive than Little's law, they can be used to derive performance measures for queueing systems that admit distributional laws. In this way, the derivation of performance measures in queueing systems that admit distributional laws can be done in a unified way. In particular, the contribution of the present paper can be summarized as follows:

1. We offer two proofs of the relation of the distributions of L and W in queueing systems, in which the arrival process is a general stationary process and the queueing discipline is FIFO. The first proof, which is a simple probabilistic proof from first principles, is similar to the one in Haji and Newell. Our second proof (Theorem 4), which is new, offers insight on the relations of pre-arrival, postdeparture, general time probabilities and the waiting time. Moreover, it is the natural matrix generalization of the proof technique of Keilson and Servi (1988).
2. Theorem 3 generalizes previous results of Burke (1956) (see also Franken et al., p. 112, Papaconstantinou and Bertsimas 1990, and Hebuterne 1988) on the relation of pretransition and postdeparture probabilities for stochastic processes with randomly distributed jumps. Although this generalization is of independent interest, it was the key for our second proof of the distributional law.
3. Most importantly, we attempt to demonstrate that the distributional law has important algorithmic and structural applications and can be used to derive various performance characteristics of several queueing systems which admit distributional laws. This is particularly important, because it shows that the distributional law is a powerful and unifying tool for the analysis of a wide variety of queueing systems.

Subject classification: Queues: approximations, algorithms, multichannel.
Area of review: STOCHASTIC PROCESSES AND THEIR APPLICATIONS.

0030-364X/95/4302-0298 $01.25

For the most part, our results were known in the literature. What is new is the unifying way of deriving them and providing new insights about them based on distributional laws.

a. We derive (Theorem 5) asymptotic closed-form relations in heavy traffic between $E[L^2]$ and $E[W^2]$ under FIFO, and between $E[L^+]$ (L^+ is the number of customers in the queue left behind from a departing customer) and $E[L]$ under FIFO for systems that obey distributional laws. In the context of the $GI/G/s$ queue, these results are special cases of more general results of Whitt (1971) based on heavy-traffic limit theorems.
b. We derive the well known asymptotic expressions (see, for example, Heyman and Sobel 1982, p. 483) in heavy traffic for $E[W]$ in a $GI/G/1$ queue (Theorem 6) and a $GI/D/s$ queue (Theorem 7) in terms of first and second moments of the interarrival and service time distributions using just the distributional laws.
c. We offer a new simple proof of the decomposition result (Doshi 1985) for the expected waiting time in vacation queues (Theorem 8) based on the distributional laws.
d. We show (Theorem 9) that in systems obeying distributional laws the queue length distribution is a mixture of geometric terms, provided that the waiting time distribution is a mixture of exponential terms. Our results are consistent with previous results of de Smit (1983) and Bertsimas (1990).
e. One major consequence of the distributional law is that one can find the queue length distribution from the waiting time distribution. Theorem 10 addresses the inverse problem. Using complex analysis techniques we show how to obtain the waiting time distribution from the queue length distribution.
f. In subsection 3.6 we derive closed-form formulas for the transforms of the queue length distributions in $GI/R/1$, $RI/G/1$, $GI/D/s$, $GI/D/\infty$ queues (R is the class of distributions with rational Laplace transforms). The waiting time distribution for such systems can be found using Hilbert factorization techniques (see, for example, Bertsimas et al. 1991). The distributional law enables one to find the transforms of the queue length distributions in such systems. For the $GI/R/1$ queue our expressions agree with the expressions of de Smit. We finally remark that just the knowledge of the distributional law is enough to fully characterize the waiting time distribution under FIFO for systems that have distributional laws for both the number in the queue and the number in the system.

The paper is structured as follows. Section 1 presents the first simple probabilistic proof of the distributional law and describes several queueing systems for which distributional laws hold. Section 2 presents our second proof for the case in which the interarrival distribution is mixed generalized Erlang. This section also includes a theorem on the relation of pretransition and postdeparture probabilities for stochastic processes with randomly distributed jumps. Section 3 contains algorithmic and structural consequences of the distributional law for a variety of queueing systems that admit distributional laws.

1. THE DISTRIBUTIONAL LAW

Consider a general queueing system, whose arrival process is stationary. We assume that customers arrive at and depart from the system one at a time. Let $N_a(t)$ be the number of customers up to time t for the ordinary process, where the time of the first interarrival time has the same distribution as the stationary interarrival time. Let $N_a^*(t)$ be the number of customers up to time t for the equilibrium process, where the time of the first interarrival time is distributed as the forward recurrence time of the arrival process.

The distributional law can be stated as follows.

Theorem 1. (Haji and Newell) *Let a given class C of customers have the following properties:*

1. *All arriving customers enter the system (or the queue) one at a time, remain in the system (or the queue) until served (there is no blocking, balking or reneging) and leave one at a time.*
2. *The customers leave the system (or the queue) in the order of arrival (FIFO).*
3. *New arriving class C customers do not affect the time in the system (or the queue) for previous class C customers.*

Then, given that they exist in steady state, the stationary waiting time W of the class C customers in the system (or the queue) and the stationary number L of the class C customers in the system (or queue) are related in distribution by:

$$L \stackrel{d}{=} N_a^*(W). \quad (1)$$

Proof. From some arbitrarily chosen time τ, which is chosen independently of the arrival process, we number the customers within the class C backward in time. In particular, the customer numbered 1 is the one who arrived most recently. The customer with the highest ordinal number n is the one getting served, or is at the head of the queue. According to this numbering, let τ_n be the arrival time of the nth class C customer in the system (or queue) and W_n be the waiting time in the system (or queue), i.e., τ_n, W_n are ordered in the reverse time direction. Assuming that the system has reached steady state, the distribution of W_n is the same as the stationary waiting time W.

Let $T_1^* = \tau - \tau_1$, i.e., T_1^* is distributed as the forward recurrence time of the arrival renewal process. For $n \geq 2$, let $T_n = \tau_{n-1} - \tau_n$.

The key observation for the proof is as follows. When an observer coming to the system at a random moment τ

sees at least n customers from class C, the nth most recently arrived customer among the class C is still waiting at that moment τ of the observation (see also Figure 1), i.e., for $n \geq 1$, $L \geq n$ if and only if $W_n > \tau - \tau_n$. Note that we have used Assumptions 1 and 2. Therefore

$$\Pr[L \geq n] = \Pr[W_n > \tau - \tau_n].$$

Now, because of Assumptions 2 and 3, W_n and $\tau - \tau_n = T_1^* + \sum_{i=2}^n T_i$ are independent. Indeed, every person arriving after time τ_n joins the queue *after* the nth customer and, therefore, each of these arrivals does not affect the waiting time W_n of that nth customer under Assumptions 2 and 3. Since the distribution of W_n is the same as the stationary waiting time W, we obtain

$$\Pr[L \geq n] = \int_0^\infty \Pr\left[T_1^* + \sum_{i=2}^n T_i < t\right] dF_W(t), \qquad (2)$$

which proves the theorem.

The distributional form of Little's law (1) has the following intuitive interpretation. In steady state, the number of class C customers in the system (queue) has the same distribution as the number of class C arrivals, arriving according to the equilibrium process, during the time spent in the system (queue).

Assumption 1 in the previous theorem does not allow for blocking, balking, or reneging. Assumption 2 does not allow overtaking, while Assumption 3 is typically not satisfied if there are dependencies among the interarrival times. Haji and Newell also include the case of batch arrivals, which we excluded in the previous theorem. The reason is that for Assumption 3 to be satisfied we need to assume that the distribution of the batch size is geometric, which is somewhat restrictive.

To develop a useful calculus so that we can exploit the distributional law, we will assume in the remainder of the paper that the arrival process is renewal. Let $\alpha(s)$ be the Laplace transform of the interarrival distribution, with arrival rate $\lambda = -1/\dot{\alpha}(0)$. Let $N_a(t)$ be the number of renewals up to time t for the ordinary renewal process and $N_a^*(t)$ be the number of renewals up to time t for the equilibrium renewal process. We will also use the notion of a complex integral $\oint_\Gamma f(s)\, ds$, for some complex function $f(s)$, in which we integrate the function $f(s)$ over a closed, simply connected curve Γ. In evaluating this type of integral, it is important that we specify whether the

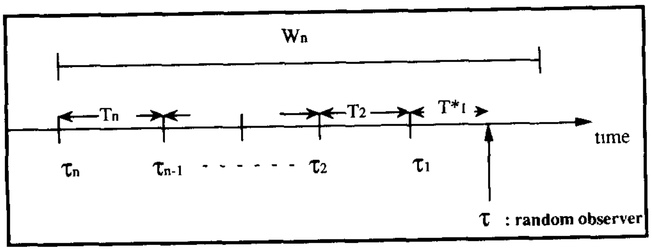

Figure 1. An illustration of Little's law.

curve Γ contains the singularities of $f(s)$ (see, for example, Henrici 1988, p. 243).

We now turn our attention to the relations among the generating function of L and the Laplace transform of W when the input process is a renewal process.

Theorem 2. *If arrivals of class C form a renewal process whose interarrival time has a transform $\alpha(s)$ and under the assumptions of Theorem 1, the cdf $F_W(t)$ of W and the generating function $G_L(z)$ of L satisfy the relation:*

$$G_L(z) = \int_0^\infty K(z, t)\, dF_W(t), \qquad (3)$$

where the kernel is the generating function of the equilibrium renewal process, i.e.,

$$K(z, t) = \sum_{n=0}^\infty z^n \Pr[N_a^*(t) = n]. \qquad (4)$$

The Laplace transform of the renewal generating function $K(z, t)$ is given by

$$K^*(z, s) = \int_0^\infty e^{-st} K(z, t)\, dt \qquad (5)$$

$$= \frac{1}{s} - \lambda \frac{(1 - z)(1 - \alpha(s))}{s^2(1 - z\alpha(s))}.$$

If the Laplace transform pdf $\phi_W(s)$ of the waiting time exists and is well defined, then

$$G_L(z) = \frac{1}{2\pi i} \oint_{\Gamma(z)} K^*(z, s) \phi_W(-s)\, ds, \qquad (6)$$

where in the contour integral we are integrating over a closed, simply connected curve $\Gamma(z)$ (depending on z) containing all singularities of $K^*(z, s)$, but not $\phi_W(-s)$.

Proof. From (1) we obtain

$$\Pr\{L = n\} = \int_0^\infty \Pr[N_a^*(t) = n]\, dF_W(t).$$

Multiplying both sides by z^n and adding over n leads to

$$G_L(z) = \sum_{n=0}^\infty z^n \Pr[L = n] = \int_0^\infty K(z, t)\, dF_W(t), \qquad (7)$$

where

$$K(z, t) = \sum_{n=0}^\infty z^n \Pr[N_a^*(t) = n].$$

It is well known (see Cox 1962, p. 37) that the Laplace transform of $K(z, t)$ is

$$\int_0^\infty e^{-st} K(z, t)\, dt = \frac{1}{s} - \lambda \frac{(1 - z)(1 - \alpha(s))}{s^2(1 - z\alpha(s))}.$$

Finally, we use the inverse Laplace transform formula for the kernel (see Kleinrock 1975, pp. 336, 352–353),

$$K(z, t) = \frac{1}{2\pi i} \oint e^{st} K^*(z, s)\, ds,$$

where the contour contains all singularities of $K^*(z, s)$ for a fixed z. We then have

$$G_L(z) = \frac{1}{2\pi i} \int_0^\infty \oint e^{st} K^*(z, s)\, ds\, dF_W(t).$$

Given that $\phi_W(s) = \int_0^\infty e^{-st}\, dF_W(t)$ exists and is well defined, we interchange the integrals and obtain that

$$G_L(z) = \frac{1}{2\pi i} \oint K^*(z, s) \phi_W(-s)\, ds.$$

Remark

The goal of this remark is to find the relation of prearrival, postdeparture probabilities and waiting time. The derivation here uses the argument of Keilson and Servi. Let L^-, L^+ be the number in the system (or the queue) just before an arrival or just after a departure, respectively, for a system that satisfies the assumptions of Theorem 2. The number of customers left behind in the system (or in the queue) by a departing customer a, is exactly the number of customers that arrived during the time customer a spent in the system (or in the queue), because the queue discipline is FIFO. As a result, $L^+ \stackrel{d}{=} N_a(w)$ and since the number of customers changes by one $L^+ \stackrel{d}{=} L^-$. As a result,

$$G_{L^+}(z) = \int_0^\infty K_o(z, t)\, dF_W(t), \tag{8}$$

where the kernel is the generating function of the ordinary renewal process, i.e.,

$$K_o(z, t) = \sum_{n=0}^{\infty} z^n \Pr[N_a(t) = n]. \tag{9}$$

It is well known (see Cox 1962) that the Laplace transform of the renewal generating function $K_o(z, t)$ is given by

$$K_o^*(z, s) = \int_0^\infty e^{-st} K_o(z, t)\, dt = \frac{1 - \alpha(s)}{s(1 - z\alpha(s))}. \tag{10}$$

If, in addition, the arrival process is Poisson and under Assumption 3 of Theorem 1, then

$$L^+ \stackrel{d}{=} L^- \stackrel{d}{=} L \stackrel{d}{=} N_a(W),$$

(note that in the case of Poisson arrivals $N_a^*(t) = N_a(t)$). In this case it is easy to prove that

$$K(z, t) = K_o(z, t) = e^{-\lambda(1-z)t}.$$

Substituting into (3) we obtain Keilson and Servi's (1988) result:

$$G_L(z) = \phi_W(\lambda - \lambda z).$$

Systems for Which the Distributional Law Holds

We emphasize that the distributional law holds for a variety of settings; these include the time and the number *in the queue* of a heterogeneous service priority $GI/G/s$ system for each priority class (within each class the discipline is FIFO). It holds even if successive service times are *dependent*. It does not hold, however, for the time and the number *in the system*, because Assumption 2 is violated (overtaking can occur). It also holds for both the number in the system and the number in the queue of a $GI/G/1$ system and a $GI/D/s$ system with priority classes. Moreover, it holds for the total sojourn time and the total number of a certain class in a queueing network without overtaking (such as tandem queues). Other applications include $GI/D/\infty$, $GI/G/1$ with vacations and exhaustive service.

2. RELATION OF PRETRANSITION AND POSTDEPARTURE PROBABILITIES AND ITS APPLICATION TO THE DISTRIBUTIONAL LAW

This section gives a different proof of the distributional law for a renewal, mixed, generalized Erlang arrival process by generalizing previous results of Burke; Papaconstantinou and Bertsimas; and Hebuterne on the relation of pretransition and postdeparture probabilities for stochastic processes with randomly distributed jumps. The advantage of this proof technique is that it leads to a closed-form expression for the kernel $K(z, t)$ in (4). This proof technique is a natural extension of Keilson and Servi's (1988) proof technique to arrival processes with rational Laplace transforms.

Our strategy for proving the distributional law is as follows.

1. We introduce the class of mixed generalized Erlang distributions.
2. We observe that when the interarrival time belongs to the class of mixed generalized Erlang distributions, the arrival process is a special case of a phase renewal process. Using the uniformization technique, a phase renewal process is interpreted as an embedded Markov chain at Poisson transition epochs.
3. In Theorem 3 we derive the relation of pretransition and postdeparture probabilities for stochastic processes with randomly distributed jumps. Since, by PASTA, an observer at a Poisson transition epoch sees time averages, we are thus able to relate the postdeparture and the general time probabilities.
4. We relate the postdeparture probabilities to the waiting time distribution.
5. We combine the last two relations and thus prove the distributional law.

2.1. The Coxian Distribution

This section will consider systems with mixed generalized Erlang (MGE) arrival processes (see Cox 1995) which can approximate any renewal arrival process arbitrarily closely. The stage representation of the MGE distribution is presented in Figure 2, i.e., we conceive the arrival process as an arrival timing channel (ATC) consisting of M consecutive exponential stages with rates $\lambda_1, \lambda_2, \ldots, \lambda_M$ and with probabilities $p_1, p_2, \ldots, p_M$ ($p_M = 1$) of entering the system after the completion of the 1st, 2nd, ..., Mth stage.

Let $a_k(t)$ be the pdf of the remaining interarrival time if the customer in the ATC is in stage $k = 1, \ldots, M$. Therefore, $a(t) = a_1(t)$ is the pdf of the interarrival time. For notational convenience we will drop the subscript for $k = 1$. Also, $1/\lambda$ denotes the mean interarrival time. Let $\alpha_k(s)$ be the Laplace transform of $a_k(t)$. Let $a_i^j(t)$ be the probability to move from stage $i \leq j$ of the ATC to stage j during the interval $[0, t)$ without having any new arrival.

We will also use the notation:

$$\vec{a}_1(t) = (a_1^1(t), \ldots, a_1^M(t))',$$

$$\vec{a}_k(t) = (0, \ldots, a_k^k(t), \ldots, a_k^M(t))'.$$

Here $\vec{\alpha}_k(s)$ denotes the Laplace transforms of $\vec{a}_k(t)$.

$$\vec{e}_j = (0, \ldots, 1, \ldots, 0)', \quad \vec{1} = (1, \ldots, 1, \ldots, 1)'.$$

By introducing the following upper semidiagonal matrix A_0 and the dyadic matrix A_1:

$$A_0 = \begin{bmatrix} \lambda_1 & -(1-p_1)\lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & -(1-p_2)\lambda_2 & \ddots & \vdots \\ \vdots & \ddots & & \ddots & \vdots \\ \vdots & & & \lambda_{M-1} & -(1-p_{M-1})\lambda_{M-1} \\ 0 & \cdots & \cdots & 0 & \lambda_M \end{bmatrix},$$

$$A_1 = \begin{bmatrix} -p_1\lambda_1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ -p_M\lambda_M & 0 & \cdots & 0 \end{bmatrix},$$

We can compactly express the transforms defined above as:

$$\vec{\alpha}'_k(s) = \vec{e}'_k(Is + A_0)^{-1},$$

$$\alpha_k(s) = -\vec{e}'_k(Is + A_0)^{-1}A_1\vec{e}_1 = \sum_{r=k}^{M} p_r\lambda_r\alpha_k^r(s)$$

$$= \sum_{r=k}^{M} p_r\lambda_r \frac{\prod_{i=k}^{r-1}(1-p_i)\lambda_i}{\prod_{i=k}^{r}(s+\lambda_i)},$$

$$\alpha(s) = -\text{trace}((Is + A_0)^{-1}A_1),$$

thus the interarrival pdf becomes

$$a(t) = -\text{trace}(e^{-A_0 t}A_1).$$

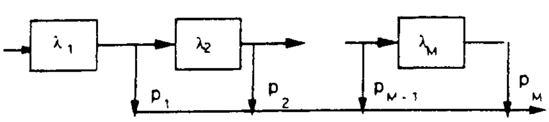

Figure 2. The Coxian class of distributions.

2.2. Uniformization of the Input Process

We will consider queueing systems with an input process that forms a renewal process with the interarrival time distribution being mixed generalized Erlang, which is a phase renewal process. We first observe that

$$(Is + A_0 + zA_1)^{-1} = (Is + A_0)^{-1} + \frac{z}{1 - z\alpha_1(s)} \begin{bmatrix} \alpha_1(s)\vec{\alpha}'_1(s) \\ \vdots \\ \alpha_M(s)\vec{\alpha}'_1(s) \end{bmatrix},$$

because for every pair of matrices C of full rank and D of rank 1,

$$(C + D)^{-1} = C^{-1} - \frac{C^{-1}DC^{-1}}{1 + \text{trace}(C^{-1}D)}.$$

By expressing this in real time we obtain

$$e^{-(A_0+zA_1)t} = \begin{bmatrix} a_1^1(t) & \cdots & a_1^M(t) \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_M^M(t) \end{bmatrix} + \sum_{n=1}^{\infty} z^n \begin{pmatrix} a_1(t) \\ \vdots \\ a_M(t) \end{pmatrix} * a_1^{(n-1)}(t) * (a_1^1(t) \cdots a_1^M(t)). \tag{11}$$

By interpreting this expression directly, it is clear that this is the phase renewal (generating) function of the arrival process.

We apply the uniformization technique (see, for example, Ross 1985, p. 261) to the phase renewal function:

$$e^{-(A_0+zA_1)t} = e^{-\nu t}e^{\nu t(I - 1/\nu A_0 - z/\nu A_1)},$$

where we choose $\nu = \max_{i=1,\ldots,M} \lambda_i < \infty$.

Let $P_0 = I - 1/\nu A_0$ and $P_1 = -1/\nu A_1$. Then

$$e^{-(A_0+zA_1)t} = \sum_{n=0}^{\infty} e^{-\nu t}\frac{(\nu t)^n}{n!}[P_0 + zP_1]^n.$$

The interpretation of this formula is that a transition occurs in a Poisson manner at rate ν with ATC phase transition probability P_0 and the *effective* arrival probability P_1. Note that a transition is either an arrival or a shift to the next exponential stage. We will use this interpretation in the next two subsections.

2.3. The Relation Between the Pretransition and Postdeparture Probabilities

Our goal in this subsection is to find a relation between the probabilities at pretransition Poisson epochs of the arrival process and postdeparture probabilities. In Theorem 3 we generalize the previous results of Burke; Papaconstantinou and Bertsimas; and Hebuterne on the relation of pretransition and postdeparture probabilities for stochastic processes with randomly distributed jumps.

In the previous subsection we introduced the matrices: $P_0 = I - 1/\nu A_0$ with elements $P_{0,i,j}\{i, j = 1, \ldots, M\}$, and $P_1 = -1/\nu\, A_1$ with elements $P_{1,i,j}\{i, j = 1, \dots, M\}$. We now introduce the notation we will use:

L^+ = The number of customers in the system (or queue) immediately after a departure epoch.

L^- = The number of customers in the system (or queue) just before a *transition* epoch of the arrival process. A transition includes both arrivals in the system and shifts to the next exponential stage of the ATC. We emphasize that L^- in this and the following subsection is *not* the number of customers before an *arrival* epoch. The motivation for considering L^- is that, because of the uniformization, the epochs of transition of the arrival process (n, i) are Poisson distributed with rate ν and thus we can apply PASTA.

R^+ = The ATC stage immediately after a departure epoch.

R^- = The ATC stage just before a transition epoch of the arrival process.

$\vec{P}_n^+ = \{\Pr[L^+ = n \cap R^+ = i]\}_{i=1}^{i=M}$ and $\vec{P}_n^- = \{\Pr[L^- = n \cap R^- = i]\}_{i=1}^{i=M}$.

$\vec{P}^+(z) = \sum_{n=0}^{\infty} z^n \vec{P}_n^+$ and $\vec{P}^-(z) = \sum_{n=0}^{\infty} z^n \vec{P}_n^-$.

We observe the system in the interval $(0, T)$ and define

$u(n, i)$ = The number of upward jumps of the number of customers n during the period $(0, T)$ such that $L^- = n$ and $R^- = i$.

$u^0(n, i)$ = The number of shifts from the ATC stage i to $i + 1$ during the period $(0, T)$ such that $L^- = n$ and $R^- = i$.

$d(n, i)$ = The number of downward jumps of the number of customers n during the period $(0, T)$ such that $L^+ = n$ and $R^+ = i$.

U = The total number of transitions (upward jumps and shifts) of the arrival process during the period $(0, T)$.

D = The total number of departures (downward jumps) during the period $(0, T)$.

We now prove the theorem on the relation of the pretransition and postdeparture probabilities.

Theorem 3. *Let $\vec{P}^-(z)$, $\vec{P}^+(z)$ be the generating functions for the pretransition and the postdeparture probabilities as defined above. Then*

$$\vec{P}^-(z) = \lambda(1 - z)\vec{P}^+(z)(A_0 + zA_1)^{-1}. \qquad (12)$$

Under Assumption 3 of Theorem 1, the generating function of the number of customers in the system L is

$$G_L(z) = \vec{P}^-(z)\vec{1}. \qquad (13)$$

Proof. We follow a method used by Papaconstantinou and Bertsimas and Hebuterne to establish the relation between pre-arrival and postdeparture probabilities in stochastic processes with random upward and downward jumps. We first write the flow balance equations; the left-hand sides correspond to flow out and the right sides correspond to flow in.

$$d(n-1, i) + u^0(n, i) + u(n, i) = d(n, i) + u^0(n, i-1) \qquad \{i > 1,\ n > 0\}$$

$$d(n-1, 1) + u^0(n, 1) + u(n, 1) = d(n, 1) + \textstyle\sum_{i=1}^{M} u(n-1, i) \qquad \{n > 0\}$$

$$u^0(0, i) + u(0, i) = d(0, i) + u^0(0, i-1) \qquad \{i > 1\}$$

$$u^0(0, 1) + u(0, 1) = d(0, 1). \qquad (14)$$

We divide all equations by U and we then take the limit as $T \to \infty$. Note that

$$\frac{D}{U} \to \frac{\lambda}{\nu}$$

$$\frac{1}{D} d(n, i) \to \Pr[L^+ = n \cap R^+ = i]$$

$$\frac{1}{U} u^0(n, i) \to \Pr[L^- = n \cap R^- = i] P_{0,i,i+1}$$

$$\frac{1}{U} u(n, i) \to \Pr[L^- = n \cap R^- = i] P_{1,i,1}$$

$$\frac{1}{U}(u^0(n, i) + u(n, i)) \to \Pr[L^- = n \cap R^- = i](1 - P_{0,i,i}).$$

Then (14) becomes in matrix form

$$\vec{P}_n^-(I - P_0) - \vec{P}_{n-1}^- P_1 = \frac{\lambda}{\nu}\vec{P}_n^+ - \frac{\lambda}{\nu}\vec{P}_{n-1}^+ \quad \{n > 0\}$$

$$\vec{P}_0^-(I - P_0) = \frac{\lambda}{\nu}\vec{P}_0^+.$$

Computing the generating functions $\vec{P}^-(z)$, $\vec{P}^+(z)$, we obtain

$$\vec{P}^-(z)[I - P_0 - zP_1] = \frac{\lambda(1-z)}{\nu}\vec{P}^+(z),$$

which leads to

$$\vec{P}^-(z) = \lambda(1 - z)\vec{P}^+(z)(A_0 + zA_1)^{-1}.$$

The epochs of transition of the arrival process (n, i) are Poisson distributed with rate ν. Therefore, by PASTA (see Wolff 1982) and using Assumption 3 of Theorem 1, the number of customers in the system L is equal in distribution to the pretransition number in the system L^-, and hence

$$G_L(z) = \vec{P}^-(z)\vec{1}.$$

2.4. The Relation of the Waiting Time and the Postdeparture Probabilities

Let $\phi_W(s)$ be the transform pdf of the waiting time of a class C customer and $F_W(t)$ be the cdf of the waiting time. In this subsection we relate the waiting time and the postdeparture probabilities.

Proposition 1. *Under assumptions 1 and 2 of Theorem 1 and for mixed, generalized Erlang interarrival times, the postdeparture probability generating function $\vec{P}^+(z)$ is represented as*

$$\vec{P}^+(z) = \vec{e}_1'\Phi_W(A_0 + zA_1),$$

where for any matrix D we symbolically define:

$$\Phi_W(D) \triangleq \int_0^\infty e^{-Dt}\, dF_W(t).$$

Proof. Because of Assumptions 1 and 2 in Theorem 1, the number of customers left behind by a customer departing from the system (or queue) is precisely the same as the number of customers that arrived during this customer's waiting time in the system (or queue). Therefore,

$$\Pr[L^+ = n \cap R^+ = i] = \int_0^\infty a^{(n)}(t) * a_1^i(t)\, dF_W(t).$$

Taking generating functions and using (11) we find

$$\vec{P}^+(z) = E[\vec{e}_1' e^{-(A_0+zA_1)W}],$$

and thus the result holds. □

2.5. A Matrix View of the Distributional Law

We now have all the necessary ingredients to give the second proof of the distributional law.

Theorem 4. *Under the assumptions of Theorem* 1 *and for mixed, generalized Erlang interarrival times characterized by the matrices* A_0, A_1, *the generating function* $G_L(z)$ *and the cdf* $F_W(t)$ *are related by*:

$$G_L(z) = \int_0^\infty K(z, t)\, dF_W(t),$$

where

$$K(z, t) = \lambda(1 - z)\vec{e}_1' e^{-(A_0+zA_1)t}(A_0 + zA_1)^{-1}\vec{1},$$

which leads to

$$G_L(z) = \lambda(1 - z)\vec{e}_1'\Phi_W(A_0 + zA_1) \cdot (A_0 + zA_1)^{-1}\vec{1}. \quad (15)$$

Proof. Combining (13) and (12) we obtain

$$G_L(z) = \lambda(1 - z)\vec{P}^+(z)(A_0 + zA_1)^{-1}\vec{1}.$$

Then by Proposition 1,

$$G_L(z) = \lambda(1 - z)\vec{e}_1'\Phi_W(A_0 + zA_1)(A_0 + zA_1)^{-1}\vec{1}.$$

Therefore,

$$G_L(z) = \int_0^\infty \lambda(1 - z)\vec{e}_1' e^{-(A_0+zA_1)t} \cdot (A_0 + zA_1)^{-1}\vec{1}\, dF_W(t),$$

i.e.,

$$K(z, t) = \lambda(1 - z)\vec{e}_1' e^{-(A_0+zA_1)t}(A_0 + zA_1)^{-1}\vec{1}.$$

Remarks

1. Note that (15) is a matrix generalization of Keilson and Servi's (1988) result that in the case of Poisson arrivals

$$G_L(z) = \phi_W(\lambda - \lambda z).$$

2. The transform of $K(z, t)$ is given by

$$K^*(z, s) = \lambda(1 - z)\vec{e}_1'(Is + A_0 + zA_1)^{-1} \cdot (A_0 + zA_1)^{-1}\vec{1}$$
$$= \frac{\lambda(1 - z)}{s}\vec{e}_1'\{(A_0 + zA_1)^{-1} - (Is + A_0 + zA_1)^{-1}\}\vec{1}.$$

But

$$\vec{e}_1'(Is + A_0 + zA_1)^{-1}\vec{1}$$
$$= \vec{e}_1'\left\{(Is + A_0)^{-1} - \frac{z}{1 - z\alpha(s)}(Is + A_0)^{-1}A_1 (Is + A_0)^{-1}\right\}\vec{1}$$
$$= \left\{1 + \frac{z}{1 - z\alpha(s)}\alpha(s)\right\}\vec{e}_1'(Is + A_0)^{-1}\vec{1}$$
$$= \frac{1}{1 - z\alpha(s)}\frac{1 - \alpha(s)}{s}$$
$$= \frac{1 - \alpha(s)}{s(1 - z\alpha(s))},$$

and similarly by taking $s \to 0$,

$$\vec{e}_1'(A_0 + zA_1)^{-1}\vec{1} = \frac{1}{\lambda(1 - z)}. \quad (16)$$

Therefore

$$K^*(z, s) = \frac{\lambda(1 - z)}{s}\left\{\frac{1}{\lambda(1 - z)} - \frac{1 - \alpha(s)}{s(1 - z\alpha(s))}\right\}$$
$$= \frac{1}{s} - \lambda\frac{(1 - z)(1 - \alpha(s))}{s^2(1 - z\alpha(s))},$$

thus giving an alternative proof of (5).

3. Theorem 4 is a special case of Theorem 1 for the case of a mixed generalized Erlang arrival process. The advantage of the proof technique that led to Theorem 4 is that it also provides a closed-form expression for the kernel $K(z, t)$. This has some interesting applications, as we show in the next section.

3. APPLICATIONS OF THE DISTRIBUTIONAL LAW

In this section we investigate some important structural and algorithmic consequences of the distributional law.

3.1. Relations Among the Second Moments

A useful application of the distributional law is a relation of the first two moments of the queue length and the waiting time. Formulas that relate higher moments of queue length and the waiting time were obtained in Brumelle (1972) (see also Miyazawa 1979). Our formula (18) is different, but equivalent with the one in Brumelle, while our asymptotic formulas (19) and (20) are, to the best of our knowledge, new. Although they are asymptotically correct only in heavy traffic, they have the important advantage of relating the second moment of the queue length distribution with just the first and second moments of the waiting time. Consider a queueing system which satisfies a distributional law. Let λ be the arrival rate and let μ be the total service rate (for example, in the case of the $GI/G/s$ queue with heterogeneous servers with rates $\mu_1, \ldots, \mu_s$, $\mu = \sum_{i=1}^{s} \mu_i$). Let $\rho = \lambda/\mu$ be the traffic intensity of the system. For the system to be stable we require that $\rho < 1$.

Theorem 5. *Under the assumptions of Theorem 1 with renewal arrivals, Little's law for the first and second moments is*

$$E[L] = \lambda E[W]. \tag{17}$$

$$E[L^2] = \lambda\left(E[W] + 2E\left[\int_0^W E[N_a(\tau)]\, d\tau\right]\right), \tag{18}$$

where $E[N_a(t)]$ is the renewal function whose Laplace transform is given by

$$\int_0^\infty e^{-st} E[N_a(t)]\, dt = \frac{\alpha(s)}{s(1-\alpha(s))}.$$

Asymptotically, as $\rho \to 1$

$$E[L^2] = \lambda^2 E[W^2] + \lambda c_a^2 E[W] - \frac{\lambda^3 E[A^3]}{3} + \frac{(c_a^2+1)^2}{2} + o(1), \tag{19}$$

$$E[L^+] = \lambda E[W] + \frac{c_a^2 - 1}{2} + o(1), \tag{20}$$

where c_a^2 is the square coefficient of variation of the interarrival distribution and $E[A^3]$ is the third moment of the interarrival time.

Proof. We first expand $K(z, t)$ as a Taylor series in terms of $\log(z)$:

$$K(z, t) = 1 + \lambda t \log(z) + \lambda\left(t + 2\int_0^t E[N_a(\tau)]\, d\tau\right) \cdot \frac{(\log(z))^2}{2} + o((\log(z))^2).$$

To see this, we substitute $z = e^u$ in (5) first, expand the expression in terms of u, and then perform the inverse Laplace transform term by term. Now if we compare it with

$$G_L(z) = \sum_{r=0}^{\infty} E[L^r] \frac{(\log(z))^r}{r!},$$

we obtain (17) and (18).

Let A be the interarrival time. As $\rho \to 1$, $W \to \infty$ and, thus, the asymptotically important contribution in (18) is from the terms $E[N_a(t)]$ for large t. As a result, we need to investigate the asymptotic behavior of $E[N_a(t)]$ as $t \to \infty$, or, equivalently, we need to investigate the asymptotic behavior of the Laplace transform of $E[N_a(t)]$ as $s \to 0$. To achieve this we expand the Laplace transform of $E[N_a(t)]$ in powers of s and obtain

$$\frac{\alpha(s)}{s(1-\alpha(s))} = \frac{\lambda}{s^2} + \frac{c_a^2 - 1}{2s} - \frac{\lambda^2 E[A^3]}{6} + \frac{\lambda^3 E[A^2]^2}{4} + o(1).$$

Inverting term by term, we obtain that

$$E[N_a(t)] = \lambda t + \frac{c_a^2 - 1}{2} + \left[\frac{\lambda^3 E[A^2]^2}{4} - \frac{\lambda^2 E[A^3]}{6}\right]\delta(t) + o(\delta(t)),$$

where $\delta(t)$ is Dirac's function. Substituting into (18) and performing the integration term by term we obtain (19). Starting with (8) and (10) and using the same technique as before we obtain (20).

We believe that the asymptotic behavior of the renewal function as $t \to \infty$, which is the key idea in the previous theorem, is a powerful tool leading to insightful results. Indeed, it plays an important role in approximations in queues (see Whitt 1982 and Fendick and Whitt 1989).

To better understand the previous theorem let us consider some examples.

1. For the Poisson case ($c_a^2 = 1$) the asymptotic expressions (19) and (20) are exact. Moreover, in this case the distributional law leads to an easy expression between the factorial moments. Since

$$E[z^L] = E[e^{-\lambda(1-z)W}]$$

successive differentiation leads to:

$$E[L(L-1)\ldots(L-r+1)] = \lambda^r E[W^r], \quad r = 1, 2, \ldots.$$

These expressions were derived in Brumelle (1972) and discussed in Franken (1976) and Miyazawa (1985).

2. In general, $E[L^2]$ and $E[L^+]$ depend on the entire distribution of W and not only on the first two moments. In heavy traffic, however, only the first moments are asymptotically important. To see this better, let us consider a queueing system with Erlang E_2 interarrival times, i.e., $\alpha(s) = (2\lambda/s + 2\lambda)^2$. Then

$$\frac{\alpha(s)}{s(1-\alpha(s))} = \frac{4\lambda^2}{s^2(s+4\lambda)},$$

which leads to

$$E[N_a(t)] = \lambda t - 1/4 + 1/4\, e^{-4\lambda t}.$$

Therefore,

$$E[L^2] = \lambda^2 E[W^2] + \frac{\lambda}{2} E[W] + 1/8 - 1/8\, E[e^{-4\lambda W}],$$

which is exactly expression (19), in which the $o(1)$ term is $-1/8\, E[e^{-4\lambda W}]$, which obviously depends on the entire distribution of W. In heavy traffic, however, $W \to \infty$ and, thus, the $o(1)$ term goes to zero exponentially fast.

Another interesting observation is that $E[L^+]$ depends in general on the queueing discipline. In heavy traffic, however, we found that (20) leads to

$$E[L^+] = E[L] + \frac{c_a^2 - 1}{2} + o(1).$$

Since $E[L]$ is independent of the queue discipline, we conclude that in heavy traffic $E[L^+]$ is independent of the queue discipline.

The above expressions not only offer structural insight linking together fundamental properties in queues, but they lead, as we see next, to closed-form asymptotic formulas for the expected waiting time for systems that have a distributional law for both the number in the system and in the queue.

3.2. Closed-Form Approximations for Systems With No Overtaking

Consider a queueing system in which the distributional law holds for both the number in the system and the number in the queue. Examples in this category include the *GI/G/1*, *GI/D/s* queues with priorities (in each priority the service discipline is FIFO), as well as the *GI/G/1* with vacations under FIFO. We will show in this subsection that the formulas we gave in the previous subsection lead to a closed-form formula for the expected waiting time.

The *GI/G/1* Queue

Let L, Q be the number in the system and queue, respectively, and S and W be the time spent in the system and queue. Let $1/\lambda$, $E[X]$, c_a^2, c_x^2 be the means and the square coefficients of variation for the interarrival and service time distributions. Let $E[A^3]$ be the third moment of the interarrival distribution. We will develop formula (21) for the expected waiting time as a function of these parameters using just the distributional laws. Then (21) is similar but not identical to the diffusion approximation for the expected waiting time in a *GI/G/1* queue (see, for example, Heyman and Sobel, p. 483), and is well known to be tight. We derive it just from the distributional laws.

Theorem 6. *For a GI/G/1 queue under FIFO, the expected waiting time in the queue is asymptotically* ($\rho \to 1$) *given by*

$$E[W_{GI/G/1}] = \frac{\rho^2 c_x^2 + \rho^2 + \rho c_a^2 - \rho + o(1)}{2\lambda(1-\rho)}. \quad (21)$$

Proof. For the *GI/G/1* queue the distributional law holds for both the number in the system and the number in the queue. Applying (17) and (19) we obtain

$$E[L] = \lambda E[S], \quad (22)$$

$$E[Q] = \lambda E[W], \quad (23)$$

$$E[L^2] = \lambda^2 E[S^2] + \lambda c_a^2 E[S] - \frac{\lambda^3 E[A^3]}{3} + \frac{(c_a^2+1)^2}{2} + o(1), \quad (24)$$

$$E[Q^2] = \lambda^2 E[W^2] + \lambda c_a^2 E[W] - \frac{\lambda^3 E[A^3]}{3} + \frac{(c_a^2+1)^2}{2} + o(1). \quad (25)$$

But $S = W + X$, where X is the service time and W, X are independent. Thus

$$E[S] = E[W] + E[X]. \quad (26)$$

$$E[S^2] = E[W^2] + E[X^2] + 2E[W]E[X]. \quad (27)$$

Finally, it is straightforward to verify that the transforms of $G_L(z) = E[z^L]$, and $G_Q(z) = E[z^Q]$, are related as:

$$G_L(z) = (1-z)(1-\rho) + zG_Q(z),$$

which leads by successive differentiation to

$$E[L] = E[Q] + \rho, \quad (28)$$

$$E[L^2] = 2E[Q] + E[Q^2] + \rho. \quad (29)$$

Substituting (26), (27), and (29) to (24) we obtain

$$E[Q^2] + 2E[Q] + \rho = \lambda^2(E[W^2] + E[X^2] + 2E[W]E[X]) + \lambda c_a^2(E[W] + E[X]) - \frac{\lambda^3 E[A^3]}{3} + \frac{(c_a^2+1)^2}{2} + o(1).$$

Substituting (23) and (25) to the above equation and solving for $E[W]$ we obtain

$$E[W] = \frac{\lambda^2 E[X^2] + \rho c_a^2 - \rho + o(1)}{2\lambda(1-\rho)},$$

which leads to (21), since $\lambda^2 E[X^2] = \rho^2(c_x^2 + 1)$.

Remarks

1. Although relations (22), (23), and (28) are known to hold under a variety of service disciplines they are not sufficient to derive $E[W]$. We used them in the more restrictive context of FIFO to illustrate that the distributional laws are sufficient to derive $E[W]$.
2. Note that the formula is exact for $c_a^2 = 1$ (the Pollaczek–Kintchine formula). In addition, the formula addresses the dependence of the expected waiting time on just the first two moments of the

interarrival and service time distributions. The $o(1)$ terms, which we have neglected, include the dependence of the expected waiting time on higher-order moments.

The *GI/D/s* Queue

The *GI/D/s* queue satisfies distributional laws for both the number in the system and the number in the queue. Using the well known result that the waiting time in the queue in the *GI/D/s* queue is the same as in the $GI^{(s)}/D/1$ queue, where $GI^{(s)}$ is the s convolution of the interarrival distribution we obtain the following theorem.

Theorem 7. *For the GI/D/s queue the expected waiting time is asymptotically given*

$$E[W_{GI/D/s}] = \frac{\rho c_a^2 + s(\rho^2 - \rho) + o(1)}{2\lambda(1-\rho)}, \tag{30}$$

where $\rho = \lambda E[X]/s$.

3.3. Stochastic Decomposition in Vacation Queues

In this subsection we consider a $GI/G/1$ queue with vacations as follows: When the system becomes empty, the server becomes inactive ("on vacation") for a duration V, independently of the arrival process, having Laplace transform $v(s)$. At the end of the vacation period, another vacation period begins if the system is empty. Otherwise, the system is again served exhaustively. We offer a new, simple proof of the well known decomposition result (Doshi) for the expected waiting time in vacation queues based on the distributional law.

Theorem 8. (Doshi) *For the $GI/G/1$ with vacations V under FIFO, the expected waiting time is the sum of the expected waiting time of a $GI/G/1$ and the forward recurrence time of the vacation V.*

Proof. Let L_v, Q_v, W_v, and S_v be the number of customers in the system, the number of customers in the queue, and the time spent in the queue and in the system *with vacations*, respectively. Let B be the number of customers in queue given that the server is on vacation. Let V^* be the forward recurrence time of the vacation period.

Applying the distributional law to this system we will get (22)–(27), but for L_v, Q_V, W_v, and S_v. By conditioning on whether the server is on vacation we obtain

$$G_{L_v}(z) = zG_{Q_v}(z) + (1-\rho)(1-z)G_B(z).$$

Differentiating twice we obtain

$$E[L_v] = E[Q_v] + \rho, \tag{31}$$

$$E[L^2] = 2E[Q_v] + E[Q_v^2] + \rho - 2(1-\rho)E[B]. \tag{32}$$

Moreover, applying the Little's law for the random variables B and V^* we obtain

$$E[B] = \lambda E[V^*]. \tag{33}$$

Solving the system of equations (22)–(27) and (31)–(33) we obtain

$$E[W_v] = E[V^*] + \frac{\rho^2 c_x^2 + \rho^2 + \rho c_a^2 - \rho + o(1)}{2\lambda(1-\rho)},$$

i.e., from (21)

$$E[W_v] = E[V^*] + E[W_{GI/G/1}].$$

We close this subsection by emphasizing that the same method leads to similar expressions for every system that has a distributional law for both the number in the system and the number in the queue. For example, a $GI/G/1$ queue with priorities can also be analyzed using the same techniques.

3.4. Structural Implications

In this subsection we present another interesting consequence of the distributional law, which offers structural insight into the class of distributions that can arise in queueing systems.

Theorem 9. *Under the assumptions of Theorem* 1, *if the waiting time distribution $F_W(t)$ is a mixture of exponential distributions, i.e.,*

$$F_W(t) = 1 - \sum_u A_u e^{-x_u t} \tag{34}$$

(x_u can be a complex number), the queue length distribution is a mixture of geometric terms. Namely

$$\Pr[L=n] = \lambda \sum_u A_u \frac{(1-w_u)^2}{x_u w_u} w_u^n, \tag{35}$$

$$G_L(z) = 1 - \lambda \sum_u A_u \frac{(1-z)(1-w_u)}{x_u(1-zw_u)}, \tag{36}$$

$$G_{L^+}(z) = \sum_u A_u \frac{1-w_u}{(1-zw_u)}, \tag{37}$$

$$w_u = \alpha(x_u).$$

Proof. From (3) we obtain

$$\begin{aligned} G_L(z) &= \int_0^\infty K(z,t)\left\{\sum_u A_u x_u e^{-x_u t}\right\} dt \\ &= \sum_u A_u \int_0^\infty x_u e^{-x_u t} K(z,t)\, dt \\ &= \sum_u A_u x_u K^*(z, x_u). \end{aligned}$$

Using (5) and then McLaurin expanding $G_L(z)$ in terms of z, we obtain (35). In a similar way we obtain (37).

The previous theorem is applicable in a variety of queueing systems. In Bertsimas and Nakazato (1990) we show that a general $GI/G/s$ system with heterogeneous servers satisfies the exponentiality assumption of the Theorem 9. Moreover, if the service time distributions have rational Laplace transform, then the number of exponential terms in (34) is finite.

As an illustration of the usefulness of Theorem 9, we apply it to find a closed-form expression for the queue length distribution of the $GI/R/1$ queue. Let $\beta(s) = \beta_N(s)/\beta_D(s)$ be the Laplace transform of the service time distribution, where $\beta_D(s)$, $\beta_N(s)$ are polynomials of degree m and less than m, respectively. Let $\alpha(s)$ be the Laplace transform of the interarrival distribution. Using the Hilbert factorization method, one can derive the transform of the waiting time distribution for the $GI/R/1$ queue (see, for example, Bertsimas et al.) as:

$$\phi_W(s) = \frac{\beta_D(s)}{\beta_D(0)} \prod_{r=1}^{m} \frac{x_r}{x_r + s},$$

where x_r, $r = 1, \ldots, m$ are the m roots of the equation

$$\alpha(z)\beta(-z) = 1, \quad Re(z) > 0.$$

Expanding $\phi_W(s)$ in partial fractions and inverting we find that

$$F_W(t) = 1 - \sum_{r=1}^{m} \frac{\beta_D(-x_r)}{\beta_D(0)} \prod_{i \neq r} \frac{x_i}{x_i - x_r} e^{-x_r t}.$$

Applying Theorem 9 we find that the queue length distribution is given by

$$\Pr\{Q = n\} = \lambda \sum_{r=1}^{m} \frac{\beta_D(-x_r)}{\beta_D(0)} \prod_{i \neq r} \frac{x_i}{x_i - x_r} \frac{(1 - w_r)^2}{x_r w_r} w_r^n, \tag{38}$$

where $w_r = \alpha(x_r)$.

3.5. The Inverse Problem

Equation 3 gives the generating function of the number in the system (or in the queue) as an integral transformation of the distribution of the time in the system (or in the queue). Therefore, once the waiting time is known it is easy to find the queue length distribution through (3). It is interesting, however, to find an inverse of this linear transformation and express W in terms of L. Our goal is to find a kernel $\bar{K}(z, t)$ so that

$$F_W(t) = \frac{1}{2\pi i} \oint \bar{K}(z, t) G_L(z)\, dz,$$

where the contour contains all the singularities of $G_L(z)$ but none of $\bar{K}(z, t)$.

Theorem 10. *Under the assumptions of Theorem 1, when the waiting time distribution $F_W(t)$ is a mixture of exponential distributions,*

$$F_W(t) = \frac{1}{2\pi i\sqrt{-1}} \oint \bar{K}(z, t) G_L(z)\, dz,$$

where

$$\bar{K}(z, t) = \frac{\alpha^{-1}\left(\frac{1}{z}\right)}{\lambda(1 - z)^2} (e^{-\alpha^{-1}(1/z)t} - 1).$$

Proof. Assuming that the waiting time distribution is a mixture of exponential distributions, we obtain from (36)

$$\begin{aligned} G_L(z) &= 1 - \lambda \sum_u A_u \frac{(1 - z)(1 - \alpha(x_u))}{x_u(1 - z\alpha(x_u))}, \\ &= 1 - \lambda \sum_u A_u \frac{(1 - \alpha(x_u))}{x_u \alpha(x_u)} \\ &\quad + \lambda \sum_u A_u \frac{(1 - \alpha(x_u))^2}{x_u \alpha(x_u)(1 - z\alpha(x_u))}. \end{aligned}$$

Since both the left- and the right-hand sides must have the same singularity structure, $G_L(z)$ must be singular at $z = 1/\alpha(x_u)$. Therefore, from the last term of the right-hand side, we obtain

$$\underset{z=1/\alpha(x_u)}{\text{Residual}}\; G_L(z) = -\frac{\lambda(1 - \alpha(x_u))^2}{x_u \alpha(x_u)^2} A_u.$$

Let z_0 be a singular point of $G_L(z)$, i.e., $x_u = \alpha^{-1}(1/z_0)$ (assuming that there exists a unique x_u such that $\Re x_u > 0$ for each given singular point $|z_0| > 1$). From (34), we have

$$F_W(t) = -\sum_{z_0} \frac{\alpha^{-1}\left(\frac{1}{z_0}\right)}{\lambda(1 - z_0)^2} (1 - e^{-\alpha^{-1}(1/z_0)t}) \cdot \{\underset{z=z_0}{\text{Residual}}\; G_L(z)\}.$$

Expressing the last expression in terms of a Cauchy integral (see Henrici, p. 243), we obtain

$$F_W(t) = \frac{1}{2\pi i} \oint -\frac{\alpha^{-1}\left(\frac{1}{z}\right)}{\lambda(1 - z)^2} \cdot (1 - e^{-\alpha^{-1}(1/z)t}) G_L(z)\, dz.$$

Therefore,

$$\bar{K}(z, t) = -\frac{\alpha^{-1}\left(\frac{1}{z}\right)}{\lambda(1 - z)^2} (1 - e^{-\alpha^{-1}(1/z)t}).$$

3.6. Algorithmic Applications

In this section we use the distributional law to derive the distribution for the number in the system.

The *RI/G*/1 and *GI/R*/1 Queues

In (38) we derived the queue length distribution of the $GI/R/1$ queue using the Laplace transform of the waiting time. We will now use the distributional law to find the queue length distribution of the $RI/G/1$ queue.

Let $\alpha(s) = \alpha_N(s)/\alpha_D(s)$ be the Laplace transform of the interarrival distribution, where $\alpha_D(s)$, $\alpha_N(s)$ are polynomials of degree m and less than m, respectively. Let $\beta(s)$ be the Laplace transform of the service time distribution. Using the Hilbert factorization method, one can derive the waiting time distribution for the $RI/G/1$ queue (see Bertsimas et al.) as:

$$\phi_W(s) = \frac{\alpha_D(0)}{\alpha_D(-s)} \frac{(1-\rho)s}{\lambda(1-\alpha(-s)\beta(s))} \prod_{r=1}^{k-1} \frac{x_r + s}{x_r},$$

where x_r, $r = 1, \ldots, k-1$ are the $k-1$ roots of the equation

$$\alpha(z)\beta(-z) = 1, \qquad Re(z) < 0.$$

Note that for $k = 1$ the product $\Pi_{r=1}^{k-1} x_r + s/x_r$ is defined to be 1. In addition, for $k = 1$ the formula reduces to the well known Pollaczek–Kintchine formula for the $M/G/1$ queue. Applying the distributional law (15) and diagonalizing the matrix $\Phi_W(A_0 + zA_1)$ we can find the queue length distribution as:

$$G_Q(z) = \lambda(1-z)\vec{e}'_1 S(z) \begin{bmatrix} \frac{\phi_W(\theta_1(z))}{\theta_1(z)} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\phi_W(\theta_k(z))}{\theta_k(z)} \end{bmatrix} S^{-1}(z)\vec{1}, \tag{39}$$

where $S(z)$ is a matrix with columns the eigenvectors of $A_0 + zA_1$ and $\theta_i(z)$, $i = 1, \ldots, k$ are the eigenvalues of $A_0 + zA_1$, which are calculated from the equation: $z\alpha(\theta_i(z)) = 1$.

The *GI/D/s* Queue

As we observed in subsection 3.2, the waiting time in the queue for the $GI/D/s$ is exactly the same as in a $G^{(s)}/D/1$ queue, where $G^{(s)}$ is the s convolution of the interarrival distribution. Therefore, we can solve the $RI/D/s$ queue using the results of the previous paragraph for the $RI/D/1$, because the class R is closed under convolutions.

The *GI/D/∞* Queue

In this case, L is the number in the system and W is the time spent in the system. Because of the deterministic service with mean $1/\mu$, it is clearly a system with no overtaking. Moreover, because of the presence of infinite number of servers there is no waiting and thus $f_W(t) = \delta(t - 1/\mu)$. From the distributional law therefore

$$G_L(z) = K\left(z, \frac{1}{\mu}\right).$$

If, in addition, the arrival time is Poisson, i.e., $K(z, t) = e^{-\lambda(1-z)t}$, then we obtain the well known result that the number in the system has a Poisson distribution with parameter λ/μ.

Systems With No Overtaking

In subsection 3.2 we used the distributional laws to find closed-form approximations for the expected waiting time in systems for which the distributional law holds for both the number in the system and the number in the queue. We want to argue that for such systems the distributional law completely characterizes all the distributions of interest, i.e., just the knowledge of the distributional law has all the probabilistic information needed to solve for these distributions. Although the actual solution might need arguments from complex analysis, the distributional laws fully characterize such systems. This important idea was observed by Keilson and Servi (1990) for systems with Poisson arrival process. We generalize it here for systems with arbitrary i.i.d. interarrival time distributions.

Let L, Q be the number in the system and queue, respectively, and S and W is the time spent in the system and queue. From Theorem 1

$$G_L(z) = \frac{1}{2\pi i} \oint K^*(z, s)\phi_S(-s)\, ds, \tag{40}$$

$$G_Q(z) = \frac{1}{2\pi i} \oint K^*(z, s)\phi_W(-s)\, ds. \tag{41}$$

But if $\beta(s)$ is the transform of the service time distribution then

$$\phi_S(s) = \phi_W(s)\beta(s). \tag{42}$$

Finally, depending on the system being solved $G_L(z)$, $G_Q(z)$ are related. For example, for the $GI/G/1$ queue,

$$G_L(z) = (1-\rho)(1-z) + zG_Q(z). \tag{43}$$

Solving the system of equations (40), (41), (42), and (43) we can find an integral equation for the transform of the waiting time pdf

$$\frac{1}{2\pi i\sqrt{-1}} \oint K^*(z, s)\phi_W(-s)(\beta(-s) - z)\, ds = (1-\rho)(1-z). \tag{44}$$

For the special case of the $M/G/1$ queue it is easy to derive the Pollaczek–Khintchine formula. To solve (44) we need to use the calculus of residuals and regularity arguments from complex analysis. What is important is that just the knowledge of the distributional Little's law for systems with no overtaking in both the number in the queue and in the system is sufficient to fully characterize the queueing system.

ACKNOWLEDGMENT

We thank Ward Whitt for his comments on an earlier version of the paper and for pointing out the Haji and Newell reference. We thank the two careful reviewers of the paper for their insightful comments, which improved the paper significantly. The research of the first author was partially supported by NSF grants DDM-9014751, DDM-9010332, and by a Presidential Young Investigator award DDM-9158118 with matching funds from Draper Laboratory.

REFERENCES

BERTSIMAS, D. 1990. An Analytic Approach to a General Class of $G/G/s$ Queueing Systems. *Opns. Res.* **38,** 139–155.

BERTSIMAS, D., AND D. NAKAZATO. 1990. On the Characteristic Equation of General Queueing Systems. Operations Research Center Technical Report, MIT, Cambridge, Mass.

BERTSIMAS, D., J. KEILSON, D. NAKAZATO AND H. ZHANG. 1991. Transient and Busy Period Analysis of the $GI/G/1$ Queue as a Hilbert Factorization Problem. *J. Appl. Prob.* **28,** 873–885.

BRUMELLE, S. L. 1971. On the Relation Between Customer and Time Averages in Queues. *J. Appl. Prob.* **8,** 508–520.

BRUMELLE, S. L. 1972. A Generalization of $L = \lambda W$ to Moments of Queue Length and Waiting Times. *Opns. Res.* **20,** 1127–1136.

BURKE, P. J. 1956. The Output of Queueing Systems. *Opns. Res.* **6,** 699–704.

COX, D. R. 1955. A Use of Complex Probabilities in the Theory of Stochastic Processes. *Proc. Cambridge Phil. Soc.* **51,** 313–319.

COX, D. R. 1962. *Renewal Theory*. Chapman and Hall, New York.

DOSHI, B. 1985. A Note on Stochastic Decomposition in a $GI/G/1$ Queue With Vacations or Setup Times. *J. Appl. Prob.* **22,** 419–428.

FENDICK, K., AND W. WHITT. 1989. Measurements and Approximations to Describe the Offered Traffic and Predict the Average Workload in a Single-Server Queue. *Proc. IEEE*, 171–194.

FRANKEN, P. 1976. Some Applications of the Theory of Stochastic Point Processes in Queueing Theory. *Math. Nachr.* **70,** 303–319 (in German).

FRANKEN, P., D. KONIG, U. ARNDT AND V. SCHMIDT. 1982. *Queues and Point Processes*. John Wiley, New York.

HAJI, R., AND G. NEWELL. 1971. A Relation Between Stationary Queue and Waiting Time Distributions. *J. Appl. Prob.* **8,** 617–620.

HEBUTERNE, G. 1988. Relation Between States Observed by Arriving and Departing Customers in Bulk Systems. *Stoch. Proc. and Their Applic.* **27,** 279–289.

HENRICI, P. 1988. *Applied and Computational Complex Analysis, Vol. 1*, 2nd ed. John Wiley, New York.

HEYMAN, D., AND M. SOBEL. 1982. *Stochastic Models in Operations Research, Vol. 1*. Prentice Hall, New York.

KEILSON, J., AND L. SERVI. 1988. A Distributional Form of Little's Law. *O. R. Letts.* **7,** 223–227.

KEILSON, J., AND L. SERVI. 1990. The Distributional Form of Little's Law and the Fuhrmann-Cooper Decomposition. *O. R. Letts.* **9,** 239–247.

KLEINROCK, L. 1975. *Queueing Systems; Vol. 1: Theory*. John Wiley, New York.

LITTLE, J. 1961. A Proof of the Theorem $L = \lambda W$. *Opns. Res.* **9,** 383–387.

MIYAZAWA, M. 1979. A Formal Approach to Queueing Processes in the Steady State and Their Applications. *J. Appl. Prob.* **16,** 332–346.

MIYAZAWA, M. 1985. The Intensity Conservation Law for Queues With Randomly Changed Service Rate. *J. Appl. Prob.* **22,** 408–418.

PAPACONSTANTINOU, X., AND D. BERTSIMAS. 1990. Relation Between the Prearrival and Postdeparture State Probabilities and the FCFS Waiting Time Distribution in the $E_k/G/s$ Queue. *Naval Res. Logist.* **37,** 135–149.

ROSS, S. 1985. *Introduction to Probability Models, 3rd ed.* Academic Press, Florida.

DE SMIT, J. H. A. 1983. The Queue $GI/M/s$ With Customers of Different Types or the Queue $GI/H_m/s$. *Adv. Appl. Prob.* **15,** 392–419.

SIGMAN, K. 1991. A Note on a Sample-Path Rate Conservation Law and Its Relationship With $H = \lambda G$. *Adv. Appl. Prob.* **23,** 662–665.

STIDHAM, S. 1970. On the Optimality of the Single-Server Queueing System. *Opns. Res.* **18,** 708–732.

WHITT, W. 1971. Weak Convergence Theorems for Priority Queues: Preemptive Resume Priority Discipline. *J. Appl. Prob.*, 74–94.

WHITT, W. 1982. Approximating a Point Process by a Renewal Process, I: Two Basic Methods. *Opns. Res.* **30,** 125–147.

WHITT, W. 1991. A Review of $L = \lambda W$ and Extensions. *Queue. Syst.* **9,** 235–268.

WOLFF, R. 1982. Poisson Arrivals See Time Averages. *Opns. Res.* **30,** 223–231.