REPRINTED FROM:

QUEUEING SYSTEMS

Queueing Systems 13 (1993) 183–250 183

Networks of infinite-server queues with nonstationary Poisson input

William A. Massey and Ward Whitt
AT&T Bell Laboratories, Murray Hill, NJ 07974-0636, USA

J.C. BALTZER AG, SCIENCE PUBLISHERS, BASEL, SWITZERLAND

Networks of infinite-server queues with nonstationary Poisson input

William A. Massey and Ward Whitt
AT&T Bell Laboratories, Murray Hill, NJ 07974-0636, USA

Received 10 February 1992; revised 5 October 1992

In this paper we focus on networks of infinite-server queues with nonhomogeneous Poisson arrival processes. We start by introducing a more general Poisson-arrival-location model (PALM) in which arrivals move independently through a general state space according to a location stochastic process after arriving according to a nonhomogeneous Poisson process. The usual open network of infinite-server queues, which is also known as a linear population process or a linear stochastic compartmental model, arises in the special case of a finite state space. The mathematical foundation is a Poisson-random-measure representation, which can be obtained by stochastic integration. It implies a time-dependent product-form result: For appropriate initial conditions, the queue lengths (numbers of customers in disjoint subsets of the state space) at any time are independent Poisson random variables. Even though there is no dependence among the queue lengths at each time, there is important dependence among the queue lengths at different times. We show that the joint distribution is multivariate Poisson, and calculate the covariances. A unified framework for constructing stochastic processes of interest is provided by stochastically integrating various functionals of the location process with respect to the Poisson arrival process. We use this approach to study the flows in the queueing network; e.g., we show that the aggregate arrival and departure processes at a given queue (to and from other queues as well as outside the network) are generalized Poisson processes (without necessarily having a rate or unit jumps) if and only if no customer can visit that queue more than once. We also characterize the aggregate arrival and departure processes when customers can visit the queues more frequently. In addition to obtaining structural results, we use the stochastic integrals to obtain explicit expressions for time-dependent means and covariances. We do this in two ways. First, we decompose the entire network into a superposition of independent networks with fixed deterministic routes. Second, we make Markov assumptions, initially for the evolution of the routes and finally for the entire location process. For Markov routing among the queues, the aggregate arrival rates are obtained as the solution to a system of input equations, which have a unique solution under appropriate qualifications, but not in general. Linear ordinary differential equations characterize the time-dependent means and covariances in the totally Markovian case.

Keywords: Nonstationary queues, time-dependent arrival rates, infinite-server queues, queueing networks, Poisson random measures, stochastic compartmental models, Poisson-arrival-location model, dependence among queue lengths at different times.

1. Introduction and summary

Since real queueing systems typically have arrival rates that vary significantly over time, it is important to consider queueing models with time-dependent arrival rates. There have been some notable advances, as can be seen from Asmussen and Rolski [1], Bambos and Walrand [2], Duda [20], Green et al. [29], Hall [30], Lemoine [42], Newell [52], Ong and Taaffe [54], Rolski [61] and references cited in these sources, but certainly more needs to be done.

INFINITE-SERVER MODELS

To contribute, we have been reexamining what seems to be the easiest case – infinite-server models with nonhomogeneous Poisson arrival processes. In Eick et al. [21,22] we studied a single $M_t/GI/\infty$ queue; here we study an open network of infinite-server queues with a nonhomogeneous Poisson external arrival process. We discuss previous work on infinite-server models below.

Infinite-server models represent the (usually) highly idealized situation in which different customers do not interfere with each other. Infinite-server models are obviously not appropriate to describe systems in which customers spend more time waiting to begin service than being served. Nevertheless, networks of infinite-server queues are of interest both in their own right and as approximations for networks of light-to-moderately loaded multiserver queues, possibly with finite waiting space. The nice theory for infinite-server models with time-dependent arrival rates is a useful frame of reference for examining more difficult finite-server models with time-dependent arrival rates. We can fruitfully view the time-dependent behavior of the finite-server model in relation to the more tractable analytical descriptions of its infinite-server couterpart; for further discussion, see Eick et al. [23] and section 10.

Networks of infinite-server queues arise in many different contexts under different names; e.g., see Kelly [37] and Whittle [72]. In "biological" applications, networks of infinite-server queues form special classes of *population processes*, see Kingman [40], Chiang [13], Kurtz [41] and Ethier and Kurtz [24], and *stochastic compartmental models*, see Sandberg [63], Faddy [25], Purdue [59], Brown [11], Jacques [33], Garzia and Lockhart [27], Matis [46] and Matis and Wehrly [47]. Another example is the number of active colonies over time in a branching process with immigration, see Pakes and Kaplan [55].

A motivating application for us is wireless (or mobile cellular) telecommunications systems; e.g., see Lee [43]. A highly idealized model, which initially ignores resource constraints, is the open network of infinite-server queues we study here. The different queues represent cells. Call originations are modeled as a nonhomogeneous Poisson process, with the nonhomogeneity capturing the important time-of-day effect. The customer's movement from cell to cell is represented by his route through the network. The service times at each queue are the time intervals when the customer is in that cell while the call is in progress. The number of busy servers at each queue represents the number of channels in use in that cell. We believe that descriptions of this idealized network model can provide useful information about the resources required to meet service demand.

The relevant theory for infinite-server models is quite well established in the theory of stochastic point processes and random measures, as in chapters 5–7 plus appendix 2 of Daley and Vere-Jones [17], Brémaud [8] and Serfozo [64]. Hence, a significant part of this paper should be regarded as a review, as was intended for the special issue. It is appropriate to draw on this theory, not only because it is relevant, but because queueing problems played a significant role in its development, through the groundbreaking work of Palm [56] and Khintchine [39].

THE $M_t/GI/\infty$ QUEUE

A single $M_t/GI/\infty$ queue is characterized by a deterministic arrival-rate function $\alpha \equiv \{\alpha(s)\}$, a generic service-time random variable S with cumulative distribution function (cdf) G and the initial conditions. The arrival process is understood to be a *nonhomogeneous Poisson process*. This means: (1) α is nonnegative and integrable over every finite interval, (2) the number of arrivals in the interval $(a, b]$ is Poisson with mean $\int_a^b \alpha(s)\mathrm{d}s$ and (3) the numbers of arrivals in disjoint intervals are independent. For simplicity, throughout this paper, we will also assume that *the arrival-rate function α is actually integrable over the entire real line $\mathbb{R}$.*

Since we focus on the history up to a fixed time t, it actually suffices to have α be integrable over the semi-infinite interval $(-\alpha, t]$. Moreover, most results only require that α be locally integrable, i.e., integrable over each bounded interval. The extra integrability assumption ensures that all expectations and integrals are finite and that the difference of two expectations (e.g., as in a covariance) are well defined. This extra assumption is not a serious restriction for applications, because it suffices to consider finite time intervals. Indeed, the integrability assumption helps us change our thinking from the familiar steady-state setting to the time-dependent transient setting considered here. (The stationary model can be considered as the limit of models with a constant arrival rate over a finite interval $(a, b]$ as $a \to -\infty$.) The integrability assumption simplifies the theory, because it means that we are considering a finite point process; see chapter 5 of Daley and Vere-Jones [17].

Time-dependent service times can be treated, as we indicate later, but the GI without a subscript means that we are now assuming that successive service times are i.i.d. and independent of the arrival process. We assume that $E[S] < \infty$. An important role is played by a random variable S_e with the associated *stationary – excess* (or stationary-age) *cdf*

$$G_e(t) \equiv P(S_e \leqslant t) = \frac{1}{\mathrm{E}[S]} \int_0^t G^c(u)\mathrm{d}u, \quad t \geqslant 0, \tag{1.1}$$

where $G^c(t) = 1 - G(t)$.

General initial conditions can be treated by separating the new arrivals after time 0 from the customers in the system at time 0, since they do not interact. The behavior of the customers in the system at time 0 depends on their residual service-time distributions. To avoid this complication, we assume that the system started empty in the infinite past. Of course, this includes the case of starting out empty at time 0 (which is equivalent to focusing only on the new arrivals) as a special case, but it includes more; see Thorisson [67] for additional discussion.

In this context, the fundamental result is:

THEOREM 1.1

In the $M_t/GI/\infty$ model above, for each t, the queue length $Q(t)$ (number of busy servers) at time t has a Poisson distribution with finite mean

$$m(t) \equiv \mathbf{E}[Q(t)] = \mathbf{E}\left[\int_{t-S}^{t} \alpha(s)\mathrm{d}s\right] = \mathbf{E}[\alpha(t - S_e)]\mathbf{E}[S], \tag{1.2}$$

and $Q(t)$ is independent of the departure process before time t. The departure process is a Poisson process with integrable time-dependent rate function

$$\delta(t) = \mathbf{E}[\alpha(t - S)], \quad t \in \mathbb{R}. \tag{1.3}$$

We use the integrability of α to imply that $m(t)$ in (1.2) is finite and that δ in (1.3) is integrable. As reviewed in Eick et al. [21], theorem 1.1 can be traced to Palm [56], Bartlett [4], Doob [19], Khintchine [39] and Prékopa [58], all before 1958. Prékopa [58] provided an especially appealing "sample-path" proof based on Poisson random measures; we use this same approach. Other relevant references are Mirasol [48], Newell [50,51], Renyi [60], Brown [9], Brown and Ross [10], Jagerman [34], Daley [16], Collins and Stoneman [15], Jackson and Aspeden [32], Foley [26], Blanc [6] and Carrillo [12].

In Eick et al. [21,22] we showed that additional insight can be gained by further examining formulas (1.2) and (1.3). First, since the mapping of α into m in (1.2) is *linear*, we can invoke *linear system theory*; e.g., this is one explanation for simple explicit expressions for m when α has special structure. For example, if α is quadratic before time t (ignoring difficulties caused by α being not integrable and sometimes negative, i.e., taking (1.2) as the definition), then $m(t)$ is also quadratic; moreover, $m(t)$ coincides with the *pointwise stationary approximation* (PSA) $\alpha(t)\mathbf{E}[S]$ except for a *time lag* and a *space shift*; i.e. if $\alpha(s) = a + bs + cs^2$ for $s \leqslant t$, then

$$m(t) = \alpha(t - \mathbf{E}[S_e])\mathbf{E}[S] + c\mathbf{Var}(S_e)\mathbf{E}[S], \tag{1.4}$$

where $\mathbf{Var}[S_e]$ is the variance of S_e (which depends on the first three moments of S); see theorem 3.1 of [21]. The PSA $\alpha(t)\mathbf{E}[S]$ would be the mean queue length in a stationary model with constant arrival rate $\alpha(t)$. Formula (1.4) can be a basis for approximations for infinite-server and finite-server models; see [21–23].

A significant feature of nonstationarity is that the celebrated insensitivity of the stationary $M/GI/\infty$ model is lost. Another example from [21] describes the effect of service-time variability, expressed via convex stochastic order, see Stoyan [66]. If S_1 and S_2 are two candidate service-time distributions (associated with $M_t/GI/\infty$ systems which are otherwise identical) with $\mathbf{E}[f(S_1)] \leqslant \mathbf{E}[f(S_2)]$ for all convex real-valued functions for which the expectations are well defined, i.e., if S_1 is less variable than S_2 in the *convex stochastic order*, denoted by $S_1 \leqslant_c S_2$ (which implies that $\mathbf{E}[S_1] = \mathbf{E}[S_2]$), then $m_1(t) \leqslant m_2(t)$ when α is *decreasing* before time t, but $m_1(t) \geqslant m_2(t)$ when α is *increasing* before time t; see theorem 2.1 of [21]. (For the case in which α is decreasing before t, we would allow α not to be integrable or require S_1 and S_2 to have finite combined support.)

OPEN NETWORKS

It is significant that similar results hold for open networks of infinite-server queues with nonhomogeneous Poisson arrival processes. From Kelly [37] and Whittle [72], we know that a very nice theory exists for the stationary version of the models we consider; i.e., the steady-state distributions of the vector queue-length process in a $\cdot/GI/\infty$ network with a homogeneous Poisson arrival process has a product-form. It is significant that the *time-dependent* vector queue-length distributions in the nonstationary models also have product form. Unfortunately, however, this result depends strongly on the infinite-server property. (This can be seen by considering two $\cdot/M/1$ queues in series with a Poisson arrival process, starting empty; see the appendix.)

A natural nonstationary network model with all nonstationarity in the external arrival process has N queues (or nodes) with independent stationary Markovian routing according to a substochastic matrix $\mathbf{P} \equiv \{p_{ij}\}$; here p_{ij} is the probability of going next to queue j immediately after completing service from queue i. We assume that $\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty$ for all i and j, so that all arrivals eventually leave, although this property is actually not needed for many results. As before, let the external arrival process be nonhomogeneous Poisson. Let successive arrivals be initially assigned to queue i with probability π_i, with successive assignments being mutually independent. (Since independently thinned Poisson processes are Poisson, see p. 31 of Daley and Vere-Jones [17], this is equivalent to independent nonhomogeneous Poisson arrivals to each queue.) This model is more general than it might appear, because the N "queues" do not actually have to be queues as we usually think of them; they can be classes which include other customer attributes as well as location. (This is a familiar device to extend the modeling power with stationary product-form queueing network models.)

In this basic model, the service times are mutually independent and independent of the arrival process, with the service times at queue i being distributed according to the random variable S_i with cdf G_i. Consistent with the Kendall notation, we call this model $(M_t/GI/\infty)^N/M$; the final M indicates independent stationary Markov routing. Moreover, the arrival process, service times and routings are understood to be mutually independent.

Let $\alpha \equiv \{\alpha(s)\}$ be the deterministic *external-arrival-rate function* (which we assume is nonnegative and integrable over $\mathbb{R}$). We use the familiar symbol λ to describe the aggregate rates at each queue (to and from other queues as well as outside the network). In particular, λ_i^+ denotes the *aggregate-arrival-rate function to queue i*, and λ_i^- denotes the *aggregate-departure-rate function from queue i*. Since we do not have stationarity, we need *not* necessarily have $\lambda_i^+(t) = \lambda_i^-(t)$.

Again we assume that the network started empty in the infinite past. Other initial conditions can be treated by analyzing the customers in the system at time 0 and the new arrivals separately. Considering only the customers initially in the system makes the model a closed network of infinite-server queues, which we will not discuss here. We remark that the $(M_t/GI/\infty)^N/M$ network becomes a special case of the semi-Markov compartmental model of Weiner and Purdue [69] and Purdue [59] when their arrival processes are allowed to be nonhomogeneous. The semi-Markov property allows transitions to depend on the completed service time as well as the queue index. This extension is covered by our generalizations of theorem 1.2 later.

In this context, here is the main result:

THEOREM 1.2

In the $(M_t/GI/\infty)^N/M$ model above, for each t, the queue lengths $Q_i(t)$ at time t, $1 \leqslant i \leqslant N$, are independent Poisson variables random variables with finite means

$$m_i(t) \equiv \mathbf{E}[Q_i(t)] = \mathbf{E}\left[\int_{t-S_i}^{t} \lambda_i^+(u)\mathrm{d}u\right] = \mathbf{E}[\lambda_i^+(t - S_{ie})]\mathbf{E}[S_i], \qquad (1.5)$$

where λ_i^+ is the aggregate-arrival-rate function to queue i, defined as the minimal nonnegative solution (or, equivalently, the unique integrable solution) to the system of input equations

$$\lambda_i^+(t) = \alpha(t)\pi_i + \sum_{j=1}^{N} \mathbf{E}[\lambda_j^+(t - S_j)]p_{ji}, \quad 1 \leqslant i \leqslant N. \qquad (1.6)$$

In addition, for each t, the vector $(Q_1(t), \ldots, Q_N(t))$ is independent of the external departure processes (from the network from each queue) before time t. The external departure processes are independent Poisson processes with integrable time-dependent rate functions

$$\delta_i(t) = \mathbf{E}[\lambda_i^+(t - S_i)]\left(1 - \sum_{j=1}^{N} p_{ij}\right), \quad 1 \leqslant i \leqslant N. \qquad (1.7)$$

Moreover, the aggregate arrival process to queue i (counting arrivals from other queues as well as from outside the network) and the aggregate departure process from queue i (counting flows to other queues as well as to outside the network) are Poisson processes if and only if no customer can visit queue i more than once.

From theorems 1.1 and 1.2, we see that, for each t, $Q_i(t)$ in theorem 1.2 has the same distribution as if queue i were an isolated $M_t/GI/\infty$ queue with Poisson external arrival process with arrival-rate function λ_i^+ determined by (1.6). However, as in the steady-state product-form theory, see Disney and Kiessler [18] and Walrand [68], the aggregate arrival process to queue i need not be a Poisson process. We expand on this point later.

The product-form result in theorem 1.2 follows so quickly from theorem 1.1 that it should perhaps be considered an immediate corollary, but of course it is important. In particular, as pointed out by Keilson and Servi [36], to prove the product-form property in theorem 1.2 it suffices to do the route assignments in advance, so that a probability is attached to each of countably many finite deterministic routes. We can think of the different route assignments as independent thinnings of the Poisson process. Since independent thinnings of a Poisson processes are independent Poisson processes, we can regard the model as a multiclass model with independent Poisson arrival processes and deterministic finite routes for each class. For each of these routes, we initially treat different visits to the same queue as different queues. Then we can apply theorem 1.1 inductively to the resulting tandem networks. Finally, we obtain the desired result because the sum of independent Poisson random variables is Poisson, with the independence among queues preserved.

Surprisingly, however, this time-dependent product-form result does not seem to be very well known. We are unaware of any textbook treatment. Specific references for parts of theorem 1.2 in addition to Keilson and Servi [36] are Bartlett [4], Kingman [40], Faddy [25], section 5 of Purdue [59], and Harrison and Lemoine [31]. We contribute to establishing theorem 1.2 by providing additional details; e.g., we construct an example (example 6.1) showing that the *input equations* (1.6) need *not* have a unique solution; however we prove the existence of a minimal nonnegative solution to (1.6). We also show that the input equations (1.6) have a unique solution among integrable functions. (The stationary analog of (1.6) is usually called the *throughput equation*, but a change of terminology is appropriate in the nonstationary case because the arrival rates typically do not equal the departure rates.) We also analyze the arrival and departure processes (both external and aggregate).

THE REST OF THIS PAPER

In this paper we prove theorem 1.2 (see remark 7.2), but we go considerably beyond that. We start in section 2 by introducing a more general model, called a *Poisson-arrival-location model* (PALM), and establish a product-form result for it (theorem 2.1). In a PALM a customer's location is specified by a continuous-time stochastic process with values in a general state space $\mathcal{S}$. We think of queues being associated with subsets of $\mathcal{S}$. We thus could have a countably or uncountably infinite collection of queues. Any finite partition of $\mathcal{S}$ can be identified with a finite network of queues. We primarily consider a network of N queues; then we focus on the special case in which the state space $\mathcal{S}$ is $\{1, \ldots, N, \Delta_*, \Delta^*\}$, where Δ_* and Δ^* represent outside the network, with Δ_* specifying not yet having arrived and Δ^* specifying having completed service. For this special case, we also stipulate that the sample paths of the location processes be piecewise-constant with only finitely many jumps.

The PALM is motivated by the wireless-telecommunication-system application; then, without reference to any system of cells, the customer's location within the system may be his spatial coordinates in Euclidean space $\mathbb{R}^3$. When we focus on a finite system of cells, the PALM reduces to an open queueing network. The more general PALM facilitates studying the performance of the system *as a function of the system of cells*. Even for networks of infinite-server queues, the PALM goes beyond the model in theorem 1.2 by allowing the routing (the sequence of queues visited) to be time-dependent and non-Markovian, and by allowing the service times to depend on the route and be time-dependent and stochastically dependent.

In section 2, we also examine the dependence among queues (disjoint subsets of $\mathcal{S}$). The product-form result implies that the queue lengths are independent, but this is only true at each fixed t. It is interesting to consider the dependence among the queue lengths *at different times*. We determine the joint distribution of the queue lengths at different times and calculate the covariances (theorem 2.2).

In the rest of the paper we focus on the open queueing network model. In section 3 we consider the general $(M_t/G_t/\infty)^N/G_t$ model. The symbol G denotes general stochastic dependence; the symbol GI for service times denotes i.i.d; the symbol M denotes Markovian; the symbol D denotes deterministic; and the subscript t denotes time-dependence. Stationarity is assumed when the subscript t is absent. *The symbols M, M_t and GI are also understood to mean that the model component* (arrival times, service times or routing) *is independent of other model components.* This is not assumed for G and G_t. Our notational conventions are summarized in table 1. The specific models we consider are displayed in fig. 1, with extra assumptions being added as we move down in the diagram. We obtain general structural results for the more general models and explicit formulas and algorithms for the less general models.

In section 3, we consider the flows in the $(M_t/G_t/\infty)^N/G_t$ model. Just as for stationary product-form networks, the aggregate arrival and departure processes at each queue in general do not have independent Poisson increments, even though they do in an acyclic network (theorems 3.4 and 3.5). In general, disjoint increments of the aggregate arrival and departure processes have what we call a *positive-linear Poisson distribution*, from which it follows that the increments are

Table 1
Modified Kendall notation for the open queueing network.

Queue characteristic	Symbol	Meaning
Arrival process	G_t	General nonstationary point process
	G	General stationary point process
	GI	Renewal process
	M_t	Nonstationary Poisson process
	M	Stationary Poisson process
Service times	G_t	General nonstationary sequences, route-dependent
	G	General stationary sequences, route-dependent
	GI	Independent i.i.d. sequences, general distributions
	M_t	Time-dependent exponential distributions
	M	Exponential distributions
	D	Deterministic
Routing discipline	G_t	General, time-dependent
	G	General, time-independent
	M_t	Nonstationary Markov chain
	M	Stationary Markov chain
	D	Deterministic

associated. We also calculate the first and second moments of the increments (theorem 3.6).

In section 4, we focus on rates and reversibility. We determine necessary and sufficient conditions for the arrival and departure processes to have well defined rates or intensity functions (theorems 4.1–4.2). We also discuss notions of *reversibility* for this nonstationary model. In particular, we show that a time-reversed $(M_t/G_t/\infty)^N/G_t$ network with departure rates is again an $(M_t/G_t/\infty)^N/G_t$ network, typically with a different external arrival-rate function (theorem 4.4).

In section 5, we focus on deterministic routes. By the argument used for theorem 1.2, the $(M_t/G_t/\infty)^N/G_t$ model can be reduced to an $(M_t/G_t/\infty)^N/D$ model with fixed deterministic routing. Thus many results follow from theorem 1.1 and associated results for a single $M_t/G_t/\infty$ queue (theorems 5.1 and 5.2).

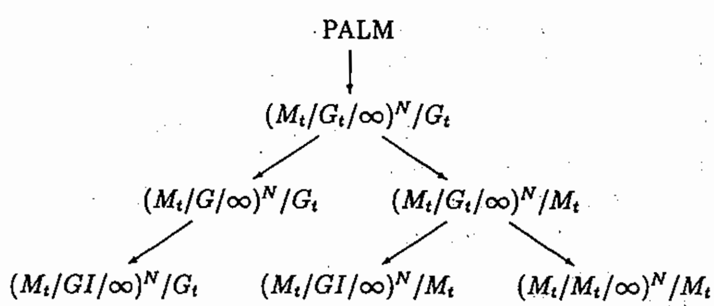

Fig. 1. Ordering the infinite server networks by generality.

In section 6, we focus on the $(M_t/GI/\infty)^N/G_t$ model, where the service times are mutually independent and i.i.d. at each queue. The extra structure here allows us to obtain nice explicit formulas. We show that if the external arrival-rate functions are polynomials or trigonometric polynomials, then so are the net arrival-rate functions to each queue. Hence, we obtain network generalizations of the explicit formulas in [21,22] (theorems 6.2 and 6.3). Polynomial arrival-rate functions may seem unrealistic because they usually converge to $\pm\infty$ as time approaches $-\infty$. However, for any time of interest, usually only the relatively recent history of the arrival-rate function is relevant. Hence, polynomial arrival-rate functions are in fact natural candidates as approximations.

In section 6 we note that many of the comparison results in section 2 of [21] hold for networks too. We also establish a new one (theorem 6.4). We conclude section 6 by examining the dependence between two queues in series. We identify the time lag which maximizes the covariance in the special cases of both exponential and both deterministic service-time distributions. We call these times *maximum-dependence times*. We believe that they will be useful in further analyses.

In section 7, we consider Markovian routing. We establish generalizations of the input equations (1.6) for the $(M_t/G_t/\infty)^N/M_t$ model when rates may not exist (theorem 7.1). In section 7 we also show that the input equations in (1.6) and in the more general $(M_t/GI/\infty)^N/M_t$ model have a minimal nonnegative solution (theorem 7.2). For the case of $(M_t/GI/\infty)^N/M$ model of theorem 1.2, we show that the input equations (1.6) have a unique solution among integrable functions (theorem 7.3). We also show how to solve the input equations in special cases (theorems 7.4 and 7.5).

In section 8, we consider the totally Markovian $(M_t/M_t/\infty)^N/M_t$ model with a time-dependent service-rate function and Markovian routing. For this model, the vector of queue lengths is a time-dependent continuous-time Markov chain (CTMC). We show how this CTMC can be constructed from homogeneous Poisson processes by uniformization. We also show that the vector mean and covariance functions satisfy linear ordinary differential equations (ODEs) (theorems 8.2 and 8.3).

In section 9, we briefly discuss large-population approximations, the simplest being the deterministic fluid model obtained by considering only the means. Refined approximations are Gaussian-process approximations, as established for stationary infinite-server network models in Glynn and Whitt [28]. Our covariance formulas completely specify these approximating Gaussian processes. These deterministic and Gaussian approximations are useful, not so much for the infinite-server models themselves, but to provide insight into related approximation for more complicated non-linear models. Here we make contact with the limit theorems for population processes in Ethier and Kurtz [24] and references cited there and the deterministic compartmental models in Sandberg [63], Brown [11], Jacques [33] and Garzia and Lockhart [27].

We conclude in section 10 by briefly discussing how our results for networks of infinite-server queues with nonstationary Poisson input can be applied to approximately analyze finite-capacity networks of multi-server queues with nonstationary Poisson input. We briefly outline network extensions of the approximations in Eick et al. [23]. This concluding section presents promising directions for future work.

2. The Poisson-arrival-location model

We now introduce the *Poisson-arrival-location model* (PALM), which generalizes the open queueing network model. As in theorem 1.2, we assume that customers arrive exogenously according to a homogeneous Poisson process $A \equiv \{A(s) : s \in \mathbb{R}\}$ with integrable nonnegative deterministic *external-arrival-rate function* α; $A(s)$ represents the number of external arrivals in the interval $(-\infty, s]$. We think of each arrival moving through a general *state space* $\mathcal{S}$ according to some stochastic process. To account for arrival and departure, we let $\mathcal{S} = \mathcal{S}^0 \cup \{\Delta_*, \Delta^*\}$, with the *pre-arrival state* Δ_* representing that the customer has not yet entered the system, and the *post-service state* Δ^* representing that the customer did arrive but is currently outside the system. With the PALM, we focus on a generalization of the queue length, the number $Q_C(t)$ of customers in a subset C of $\mathcal{S}$ at time t.

An arrival at time s has an $\mathcal{S}$-valued *location process* $L_s \equiv \{L_s(t) : t \in \mathbb{R}\}$, with $L_s(t)$ representing the location of this customer at time t, as depicted in fig. 2. We typically use s to refer to an external arrival time and t to an "observation" time. Figure 2 depicts three arrivals before time t. The arrival epochs are indicated by the open circles appearing on the vertical time line and on the location-process sample paths. The observation time t is indicated by the dark circle on the time line and the location-process sample paths. In fig. 2 we see that the three arrivals are all still in the set $\mathcal{S}^0$ at time t, but only the second arrival is in the subset C. If this is the full history up to time t, then $Q_C(t) = 1$ for this realization.

For the wireless-telecommunication-system application, it is natural to consider customer location without reference to any specific cell system. To do this, we can let $\mathcal{S}^0$ be either $\mathbb{R}^2$ or $\mathbb{R}^3$ with the Euclidean metric. With $\mathcal{S}^0 = \mathbb{R}^3$, the location functions might be continuous $\mathbb{R}^3$-valued functions on $[s, T_s)$ with $L_s(t) = \Delta_*$ for $t < s$ and $L_s(t) = \Delta^*$ for $t \geqslant T_s$, for some s and T_s with $-\infty < s < T_s < \infty$. This makes the customer be in the system for a nonempty finite time interval. Even in this wireless-telecommunication-system application, the state space $\mathcal{S}^0$ might be more general, representing additional attributes of the customers.

For the queueing networks, we let $\mathcal{S}^0 = \{1, \ldots, N\}$ and we restrict attention to location functions with only finitely many discontinuities, and $L_s(t) = \Delta_*$ for all $t < s$ and $L_s(t) = \Delta^*$ for all $t \geqslant T_s$, for some s and T_s with $-\infty < s < T_s < \infty$. The lengths of the piecewise-constant segments are the service times in the queues. This formulation requires that the service times be strictly positive (which is without loss of generality in infinite-server systems).

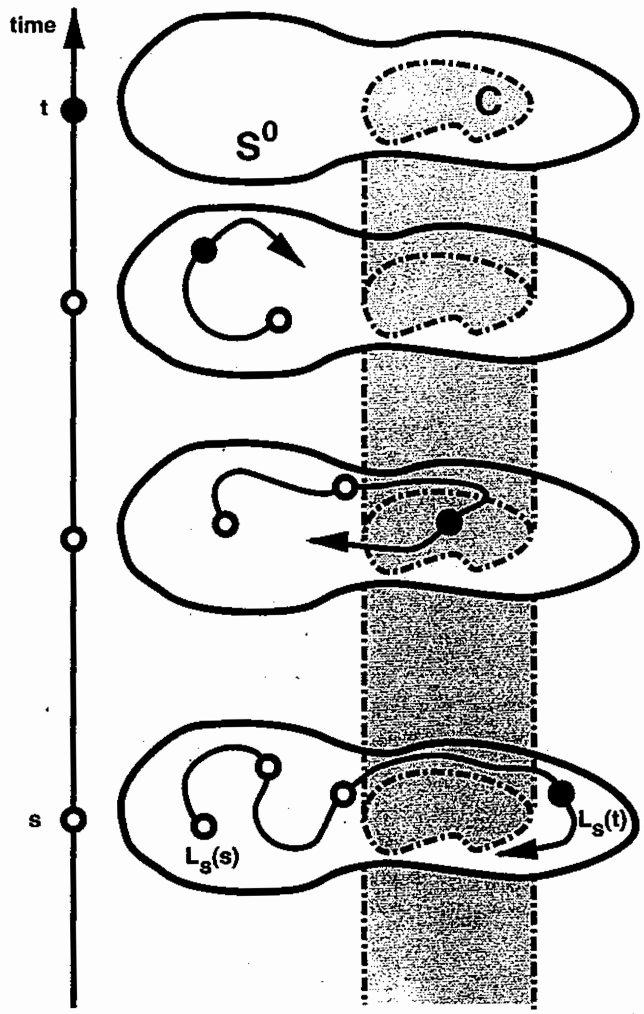

Fig. 2. A possible realization of a PALM.

THE POISSON-RANDOM-MEASURE REPRESENTATION

We represent a PALM as a *Poisson random measure* M on a product space $\mathbb{R} \times \Sigma$, where $\mathbb{R}$ is the real line and Σ is a general space, assumed to be a *Polish space*, i.e., a topological space metrizable as a complete separable metric space (CSMS); the specific metric does not matter; see chapters 5–7 plus appendix 2 in Daley and Vere-Jones [17], section IX.6 of Bourbaki [7] and Parthasarathy [57]. (The model also can be described as a *Poisson process* on $\mathbb{R} \times \Sigma$ or a *marked Poisson process* on $\mathbb{R}$ with *mark space* Σ.) An element (s, σ) of $\mathbb{R} \times \Sigma$ specifies a customer arrival time s and his location (as a function of time) σ.

A Poisson random measure $\mathbf{M}$ on $\mathbb{R} \times \Sigma$ is specified by its *intensity measure* $\boldsymbol{\mu}$, which here we assume is of the form

$$\boldsymbol{\mu}((a,b] \times \Gamma) = \int_a^b \alpha(s) P_s(\Gamma) \mathrm{d}s \tag{2.1}$$

for measurable rectangles $(a,b] \times \Gamma$ in $\mathbb{R} \times \Sigma$, where α is the external-arrival-rate functions and $P_s(\Gamma)$ is a measure on Σ for each s and a measurable function of s for each measurable subset Γ in Σ. The full distribution of $\mathbf{M}$ is characterized by extension; see pp. 177, 204–206, 607–610 of [17].

The Poisson random measure $\mathbf{M}$ has a distribution specified by

$$P(\mathbf{M}(B_1) = n_1, \ldots, \mathbf{M}(B_k) = n_k) = \prod_{i=1}^{k} \frac{\mathrm{e}^{-\gamma_i} \gamma_i^{n_i}}{n_i!}, \tag{2.2}$$

where

$$\gamma_i \equiv \mathbf{E}[\mathbf{M}(B_i)] = \int_{B_i} \mu(\mathrm{d}s, \mathrm{d}\sigma), \tag{2.3}$$

for any positive integer k, any k disjoint subsets $B_1, \ldots, B_k$ of $\mathbb{R} \times \Sigma$ and any k nonnegative integers $n_1, \ldots, n_k$; i.e., the numbers of points in disjoint subsets have independent Poisson distributions with means specified by (2.1) and (2.3).

THE SPACE OF LOCATION FUNCTIONS

The general space Σ above represents a set of possible sample paths of a location stochastic process. In particular, σ represents a *location function* depicting the location of a customer at time t. To make this function space Polish, we let Σ be the space $D(\mathbb{R}, \mathcal{S})$ of right-continuous $\mathcal{S}$-valued functions on $\mathbb{R}$ with limits from the left, where the state space $\mathcal{S}$ is Polish and $D(\mathbb{R}, \mathcal{S})$ is endowed with the usual Skorohod [65] J_1 topology, which makes $D(\mathbb{R}, \mathcal{S})$ Polish; see chapter 3 of Billingsley [5], section 2 of Whitt [70] and chapter 3 of Ethier and Kurtz [24].

As indicated above, to account for arrival and departure, we let $\mathcal{S} = \mathcal{S}^0 \cup \{\Delta_*, \Delta^*\}$. If m_0 is a metric on $\mathcal{S}^0$ inducing its topology and making it Polish, then a metric m on $\mathcal{S}$ making it Polish, for which the relative topology on $\mathcal{S}^0$ is the given topology, is determined by having $m(\Delta_*, \Delta^*) = m(\mathrm{s}, \Delta_*) = m(\mathrm{s}, \Delta^*) = 1$ for all $\mathrm{s} \in \mathcal{S}^0$ and

$$m(\mathrm{s}_1, \mathrm{s}_2) = m_0(\mathrm{s}_1, \mathrm{s}_2)/(1 + m_0(\mathrm{s}_1, \mathrm{s}_2)), \tag{2.4}$$

for $\mathrm{s}_1, \mathrm{s}_2 \in \mathcal{S}^0$.

The distribution of a PALM is specified via (2.1) by the external-arrival-rate function α and the probability distributions P_s on $\Sigma \equiv D(\mathbb{R}, \mathcal{S})$, which are the probability laws of the location processes L_s.

Remark 2.1

The specification above allows us to *simulate* **M** on a computer. In particular, we can generate a Poisson total number of arrivals with mean $\bar{\alpha} \equiv \int_{-\infty}^{\infty} \alpha(s)\mathrm{d}s$. Then these arrivals are assigned times on $\mathbb{R}$ according to i.i.d. random variables with density $\alpha(s)/\bar{\alpha}$. This produces a realization of the Poisson process A. Alternatively, assuming that α is bounded over a finite interval $[a, b]$, we can use a homogeneous Poisson process with arrival rate γ, where $\alpha(s) \leqslant \gamma, a \leqslant s \leqslant b$. We simulate points according to the homogeneous Poisson process and keep points at each time s by performing independent Bernoulli trials, keeping the point at time s with probability $\alpha(s)/\gamma$, i.e., we perform independent random thinnings; see Lewis and Shedler [44]. Then, for an arrival at time s, we simulate the additional location process according to the law $P_s(\Gamma)$. □

CONSTRUCTION VIA STOCHASTIC INTEGRATION

Another way to obtain the distribution of the Poisson random measure **M** is to directly construct the *random variables* $\mathbf{M}((a, b] \times \Gamma)$ via *stochastic integration* starting with the Poisson process A. In this construction, we use a sequence $\{Z_k : k \geqslant 1\}$ of i.i.d. random elements of a Polish space Σ' that are independent of A. We then define a family $\{L_s : s \in \mathbb{R}\}$ of random elements of Σ by letting

$$L_s = \psi(Z_{A(s)}, s), \quad -\infty < s < \infty, \tag{2.5}$$

where $\psi : \Sigma' \times \mathbb{R} \to \Sigma$ is jointly measurable. Using (2.5), we can define the random measure via the stochastic integral

$$\begin{aligned} \mathbf{M}(t, \Gamma) &\equiv \mathbf{M}((-\infty, t] \times \Gamma) \equiv \int_{-\infty}^{t} \mathbf{1}_{\{L_s \in \Gamma\}} \mathrm{d}A(s) \\ &\equiv \sum_{k=1}^{A(t)} \mathbf{1}_{\{L_{\hat{A}_k} \in \Gamma\}} = \sum_{k=1}^{\infty} \mathbf{1}_{\{\hat{A}_k \leqslant t, \psi(Z_k, \hat{A}_k) \in \Gamma\}} \end{aligned} \tag{2.6}$$

for any measurable set Γ in $D(\mathbb{R}, \mathbb{S})$, where $\hat{A}_k$ denotes the kth point of A and $\mathbf{1}_B$ is the *indicator function* of the set B. Note that the stochastic integral in (2.6) is just a countable sum. Thus no elaborate theory of stochastic integration is needed. This will be true throughout this paper.

Remark 2.2

In our model, $\Sigma = D(\mathbb{R}, \mathbb{S})$ and $L_s \equiv \{L_s(t) : t \in \mathbb{R}\}$ represents the location process of the customer arriving at time s. It is natural to think of $\{L_s : s \in \mathbb{R}\}$ as an uncountably infinite collection of stochastically independent random elements of Σ with time-dependent distributions. Such collections of uncountably infinite independent random elements are well defined, see section III.3 of Neveu [50], but they present measurability problems. Hence, we have instead defined $\{L_s : s \in \mathbb{R}\}$ in terms of the Poisson process and the (*countably infinite*) sequence of random elements $\{Z_k\}$ as in (2.5).

The random elements L_s in (2.5) are well defined for all s, but they are only relevant when A has a jump at s. The random elements $L_{s_1}, \ldots, L_{s_k}$ are typically *not* independent, but they are *conditionally independent* given that A has points at $s_1, \ldots, s_k$. □

It remains to specify the function ψ and the sequence $\{Z_k\}$ appearing in (2.5). One important special case is when the distribution of $\{L_s(t) : t \in \mathbb{R}\}$ is equal to the distribution of $\{L_0(t-s) : t \in \mathbb{R}\}$. Then it suffices to let $\Sigma' = \Sigma$ and let Z_k be distributed as L_0. Then we can have $\psi(\sigma, s)(t) = \sigma(s+t), t \in \mathbb{R}$, i.e., ψ can be the identity map modified by a time transition, which can be shown to be measurable on $D(\mathbb{R}, \mathcal{S}) \times \mathbb{R}$ (essentially because the projection maps are measurable; see p. 73 of Whitt [70] or p. 127 of Ethier and Kurtz [24]).

More generally, we can take $\{Z_k\}$ to be a sequence of i.i.d. uniforms on [0, 1], noting that any random element of a Polish space can be represented as $f(Z_k)$ for some measurable f; see section I.2 of Parthasarathy [57]. This means that given any Poisson random measure $\mathbf{M}$ with intensity measure $\boldsymbol{\mu}$ based on the pair (α, P) as in (2.3) there exists a representation of $\mathbf{M}$ of the form (2.5) and (2.6). However, the actual construction is not always easy. Of course, for a real-valued random variables with cdf G_s, it suffices to use the inverse cdf, i.e., to let Z_k be uniform on [0, 1] in (2.5) and let

$$\psi(\sigma', s) = G_s^{-1}(\sigma') \equiv \inf\{u : G_s(u) > \sigma'\}. \tag{2.7}$$

Explicit constructions can be done more generally by extending this idea, e.g., by using conditional cdf's, as we indicate below; see O'Brien [53]. (This specific construction applies directly to the representation of customer experience via finite queue sequences and service times, which we discuss in section 5.) Let $\stackrel{d}{=}$ denote equality in distribution.

LEMMA 2.1

Let $\{X_n : n \geqslant 1\}$ be a sequence of real-valued random variables and let $F(x_n | x_1, \ldots, x_{n-1}) = P(X_n \leqslant x_n | X_1 = x_1, \ldots, X_{n-1} = x_{n-1})$ be conditional probability distributions. Then $\{X_n : n \geqslant 1\} \stackrel{d}{=} \{\tilde{X}_n : n \geqslant 1\}$ where $\tilde{X}_n$ is defined recursively by

$$\tilde{X}_n = \inf\{s : F(s | \tilde{X}_1, \ldots, \tilde{X}_{n-1}) \geqslant U_n\}$$

and $\{U_n : n \geqslant 1\}$ is an i.i.d. sequence of uniform variables on [0, 1].

We can now apply the Watanabe Poisson-process-characterization theorem on p. 26 of Brémaud [8] to verify that the construction in (2.6) produces what we want. (As before, we can use extension to construct the full Poisson random measure $\mathbf{M}$ on $\mathbb{R} \times \Sigma$ from $\mathbf{M}(t, \Gamma)$ in (2.6).)

For this purpose, we now establish the required martingale property. We do this in greater generality than we need to characterize Poisson processes in order to cover other applications later. We give a direct proof of the martingale property, but it can also be obtained from the integration theorem on p. 27 of Brémaud [8]. (For this purpose, note that if s is a jump time of A, then

$$\phi(Z_{A(s)}, s) = \phi(Z_{A(s-)+1}, s),$$

so that the process is predictable with respect to the σ-fields $\mathcal{F}_s$ in lemma 2.2 below.)

LEMMA 2.2

Let $\{Z_k\}$ be an i.i.d. sequence of random elements of a Polish space Σ' distributed as Z. Let $\phi : \Sigma' \times \mathbb{R} \to \mathbb{R}$ be a measurable function such that

$$\int_{-\infty}^{t} \mathbf{E}[|\phi(Z,s)|]\alpha(s)\mathrm{d}s < \infty \quad \text{for all } t. \tag{2.8}$$

Then the process

$$M_\phi(t) = Z_\phi(t) - \mathbf{E}[Z_\phi(t)], \quad t \in \mathbb{R}, \tag{2.9}$$

where

$$Z_\phi(t) = \sum_{k=1}^{A(t)} \phi(Z_k, \hat{A}_k), \tag{2.10}$$

is a martingale with respect to the σ-fields $\mathcal{F}_t$ generated by $\{A(s) : s \leqslant t\}$ and $\{Z_k : k \leqslant A(t)\}$, and

$$\mathbf{E}[Z_\phi(t)] = \int_{-\infty}^{t} \mathbf{E}[\phi(Z,s)]\alpha(s)\mathrm{d}s. \tag{2.11}$$

Proof

It is easy to see that $Z_\phi(t)$ in (2.10) is adapted to $\mathcal{F}_t$. For all $s < t$, we have

$$\begin{aligned}\mathbf{E}[Z_\phi(t)|\mathcal{F}_s] &= Z_\phi(s) + \mathbf{E}\left[\sum_{k=1}^{A(t)-A(s)} \phi(Z_{k+A(s)}, \hat{A}_{k+A(s)}) \middle| \mathcal{F}_s\right] \\ &= Z_\phi(s) + \mathbf{E}[Z_\phi(t)] - \mathbf{E}[Z_\phi(s)].\end{aligned} \tag{2.12}$$

The second equality in (2.12) holds because the sum inside the conditional expectation is independent of $\mathcal{F}_s$. This follows from the fact that $\{\hat{A}_{k+A(s)} \leqslant t\} = \{A(t) - A(s) \geqslant k\}$. Consequently, the resulting conditional expectation is a constant. By taking expectations on both sides of the equation, we see that this constant must equal $\mathbf{E}[Z_\phi(t)] - \mathbf{E}[Z_\phi(s)]$.

Now we evaluate $\mathbf{E}[Z_\phi(t)]$. Since $\phi \to Z_\phi$ is a linear mapping, it is sufficient to let ϕ be a nonnegative function. We then have

$$\begin{aligned}
\mathbf{E}[Z_\phi(t)] &= \sum_{k=1}^{\infty} \mathbf{E}[\phi(Z_k, \hat{A}_k); \hat{A}_k \leqslant t] \\
&= \sum_{k=1}^{\infty} \int_{-\infty}^{t} \mathbf{E}[\phi(Z_k, s) | \hat{A}_k = s] P(\hat{A}_k \in \mathrm{d}s) \\
&= \sum_{k=1}^{\infty} \int_{-\infty}^{t} \mathbf{E}[\phi(Z, s)] \cdot \alpha(s) P(A(s) = k-1) \mathrm{d}s \\
&= \int_{-\infty}^{t} \mathbf{E}[\phi(Z, s)] \alpha(s) \mathrm{d}s \,.
\end{aligned} \tag{2.13}$$

The second to last step follows because $\{\hat{A}_k \leqslant s\} = \{A(s) \geqslant k\}$ and

$$\frac{\mathrm{d}}{\mathrm{d}s} P(A(s) \geqslant k) = \alpha(s) \cdot P(A(s) = k-1)\,, \tag{2.14}$$

which allows us to replace the measure $P(\hat{A}_k \in \mathrm{d}s)$ by $\alpha(s)P(A(s) = k-1)\mathrm{d}s$. Finally, from (2.13) and (2.8), we see that $\mathbf{E}[|Z_\phi(t)|] < \infty$ for all t. □

As in (2.6), our main application of lemma 2.2 is when ϕ is an indicator function. The following combines lemma 2.2 with the Watanabe theorem on p. 26 of Brémaud [8].

LEMMA 2.3

Let A be a nonhomogeneous Poisson process on $\mathbb{R}$ with integrable arrival-rate function α, let $\{Z_k : k \geqslant 1\}$ be an i.i.d. sequence of random elements of a Polish space Σ' and let $\Gamma_1, \ldots, \Gamma_k$ be disjoint subsets of another measurable space Σ. If L_s is defined by (2.5) for some jointly measurable function $\psi : \Sigma' \times \mathbb{R} \to \Sigma$, then $\{\mathbf{M}(t, \Gamma_1) : t \in \mathbb{R}\}, \ldots, \{\mathbf{M}(t, \Gamma_k) : t \in \mathbb{R}\}$ defined by (2.6) are independent Poisson processes determining the full random measure $\mathbf{M}$ on $\mathbb{R} \times \Sigma$. The arrival-rate function of $\mathbf{M}(t, \Gamma_i)$ is $\alpha(t)P(L_s \in \Gamma_i)$. Moreover, $\mathbf{M}(\infty, \Gamma_1), \ldots, \mathbf{M}(\infty, \Gamma_k)$ are independent Poisson random variables with mean

$$\mathbf{E}[\mathbf{M}(\infty, \Gamma_i)] = \int_{-\infty}^{\infty} P(L_s \in \Gamma_i) \alpha(s) \mathrm{d}s \,. \tag{2.15}$$

Below we apply lemma 2.3, not just for constructing the random measure $\mathbf{M}$ but for treating other stochastic integrals; then L_s in (2.5) need not be the location process.

We now apply lemma 2.3 to obtain a useful covariance formula. For random variables X_1 and X_2, let $\mathbf{Cov}[X_1, X_2]$ be the covariance of X_1 and X_2, i.e.,

$\mathbf{Cov}[X_1, X_2] = \mathbf{E}[X_1 X_2] - \mathbf{E}[X_1]\mathbf{E}[X_2]$. As with lemma 2.2, we give a direct proof using the Poisson property, but lemma 2.4 also follows from the product theorem for square integrable martingales; see pp. 79, 280 of Ethier and Kurtz [24].

LEMMA 2.4

If the assumptions of lemma 2.2 hold for two functions ϕ_1 and ϕ_2, and

$$\int_{-\infty}^{\infty} \mathbf{E}[|\phi_1(Z,s)\phi_2(Z,s)|]\alpha(s)\mathrm{d}s < \infty, \tag{2.16}$$

then

$$\mathbf{Cov}[Z_{\phi_1}(t), Z_{\phi_2}(t)] = \int_{-\infty}^{t} \mathbf{E}[\phi_1(Z,s)\phi_2(Z,s)]\alpha(s)\mathrm{d}s. \tag{2.17}$$

Proof

It is sufficient to prove (2.17) for nonnegative integer valued functions ϕ_1 and ϕ_2 because the covariance is bilinear and general $\phi_i(x)$ can be approximated by $\lfloor n\phi_i(x)\rfloor/n$, where $\lfloor x\rfloor$ is the greatest integer less than x. Given nonnegative-integer-valued ϕ_i, we express $Z_{\phi_i}(t)$ as linear functions of integrals of indicator functions, i.e.,

$$Z_{\phi_1}(t) = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} m \int_{-\infty}^{t} \mathbf{1}_{\{\phi_1(Z_{A(s)},s)=m,\phi_2(Z_{A(s)},s)=n\}}\mathrm{d}A(s) \tag{2.18}$$

and

$$Z_{\phi_2}(t) = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} n \int_{-\infty}^{t} \mathbf{1}_{\{\phi_1(Z_{A(s)},s)=m,\phi_2(Z_{A(s)},s)=n\}}\mathrm{d}A(s). \tag{2.19}$$

By lemma 2.3, the integrals are independent Poisson processes for different pairs (m,n). Hence

$$\begin{aligned}\mathbf{Cov}[Z_{\phi_1}(t), Z_{\phi_2}(t)] &= \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} mn\mathbf{Var}\left[\int_{-\infty}^{t} \mathbf{1}_{\{\phi_1(Z_{A(s)},s)=m,\phi_2(Z_{A(s)},s)=n\}}\mathrm{d}A(s)\right] \\ &= \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} mn\mathbf{E}\left[\int_{-\infty}^{t} \mathbf{1}_{\{\phi_1(Z_{A(s)},s)=m,\phi_2(Z_{A(s)},s)=n\}}\mathrm{d}A(s)\right] \\ &= \int_{-\infty}^{t} \mathbf{E}[\phi_1(Z_{A(s)},s)\phi_2(Z_{A(s)},s)]\alpha(s)\mathrm{d}s.\end{aligned} \tag{2.20}$$

□

THE PRODUCT-FORM RESULT

For any measurable subset C in $\mathbb{S}$, let $Q_C(t)$ be the random number of customers in the set C at time t. It is easy to see that

$$Q_C(t) = \int_{-\infty}^{\infty} \mathbf{1}_{\{L_s(t) \in C\}} \mathrm{d}A(s) = \mathbf{M}(q(t, C)), \tag{2.21}$$

where

$$q(t, C) = \{(s, l) \in \mathbb{R} \times D(\mathbb{R}, \mathbb{S}) : l(t) \in C\}. \tag{2.22}$$

We typically think of customers being in the system only *after their arrival*, but note that this is *not required* above.

We now state our product-form result.

THEOREM 2.1

Consider the PALM specified above (with integrable external-arrival-rate function α). For any positive integer k and any disjoint measurable subsets $C_1, \ldots, C_k$ of $\mathbb{S}$, the random variables $Q_{C_1}(t), \ldots, Q_{C_k}(t)$ are independent and Poisson-distributed for each t, with finite means

$$m_i(t) \equiv \mathbf{E}[Q_{C_i}(t)] = \mathbf{E}[\mathbf{M}(q(t, C_i))] = \int_{-\infty}^{\infty} P(L_s(t) \in C_i)\alpha(s)\mathrm{d}s. \tag{2.23}$$

Proof

Working with the Poisson random measure $\mathbf{M}$, we note that $q(t, C_1), \ldots, q(t, C_k)$ in (2.22) are disjoint subsets of $\mathbb{R} \times D(\mathbb{R}, \mathbb{S})$ when $C_1, \ldots, C_k$ are disjoint. Alternatively, working with the stochastic integrals, we apply the last statement in lemma 2.3. Note that, for any fixed t, the stochastic integral in (2.21) is a special case of (2.6); in (2.5) we use $\psi_t : \Sigma' \times R \to \mathbb{S}$ with

$$\psi_t(Z_{A(s)}, s) = \psi(Z_{A(s)}, s)(t) = L_s(t).$$

Finally, the means in (2.23) are necessarily finite because we have assumed that α is integrable. □

Remark 2.3

It is significant that the mean formulas (2.11), (2.15) and (2.23) do *not* depend on A being a Poisson process, as can be seen form (2.13). For the mean formula (2.13), we do use the fact that the conditional distribution of $\phi(Z_k, \hat{A}_k)$ given that $\hat{A}_k = s$ depends on s but not otherwise on k or $\hat{A}_k$, i.e.,

$$\mathbf{E}[\phi(Z_k, s) | \hat{A}_k = s] = \mathbf{E}[\phi(Z, s)].$$

We also use the fact that we have an external-arrival-rate function α,

$$\mathbf{E}[A(t)] = \int_{-\infty}^{t} \alpha(s)\mathrm{d}s,$$

as can be seen from (2.14), but this can be relaxed. We can replace the external-arrival-rate function $\alpha(s)$ by an external-arrival-intensity measure $\alpha(ds)$. In contrast, the martingale property in lemma 2.2, the covariance formula in lemma 2.4 and the distributional result in theorem 2.1 *do* depend on A being a Poisson process.

Remark 2.4

In many applications customers simultaneously use multiple resources, e.g., a call might use one circuit on each of several links connecting the source to destination; see Kelly [38]. The PALM encompasses this, because the space $\mathcal{S}^0$ could be a set of subsets of some other space, say Ω. Indeed, the PALM provides a network generalization of the infinite-server version of the model in [38], because customers may successively hold different subsets. In this framework, for each $\omega \in \Omega$, we can let $Q_{\{\omega\}}(t)$ be independent sum of $Q_{\{s\}}(t)$ for those $s \in \mathcal{S}^0$ for which $\omega \in s$. Hence, $Q_{\{\omega\}}(t)$ has a Poisson distribution for each ω and t as a consequence of theorem 2.1, but $Q_{\{\omega_1\}}(t)$ and $Q_{\{\omega_2\}}(t)$ are *not* necessarily independent. (This marginal Poisson property is noted by Keilson and Servi [36].) Indeed, the covariance is easily computed; $\mathbf{Cov}[Q_{\{\omega_1\}}(t), Q_{\{\omega_2\}}(t)]$ it is the sum of the means $\mathbf{E}[Q_{\{s\}}(t)]$ over all $s \in \mathcal{S}^0$ such that $\omega_1 \in s$ and $\omega_2 \in s$. A similar analysis applies to the covariance between queue lengths at different times in a PALM, as we show below.

DEPENDENCE AMONG THE QUEUE LENGTHS

By theorem 2.1, there is *no* dependence among the queue lengths at each time t, but there is interesting dependence at different times. (This is also true for the classical stationary product-form queueing networks.) We now determine the joint distribution of the queue lengths in a PALM at different times. For this purpose, we say that a k-dimensional probability distribution is *multivariate Poisson distribution* if it is the distribution of a random vector $(Y_1, \ldots, Y_k)$ where, for each i, Y_i is the sum of the variables in a finite subset of a collection $\{Z_1, \ldots, Z_n\}$ of mutually independent Poisson random variables; see pp. 298–300 of Johnson and Kotz [35] and p. 137 of Barlow and Proschan [3]. Since sums of independent Poisson variables are Poisson, it suffices for n to be at most $2^k - 1$.

The Poisson-random-measure representation of our PALM implies that the collection of random variables $\mathbf{M}(B_1), \ldots, \mathbf{M}(B_k)$ for subsets $B_1, \ldots, B_k$ of $\mathbb{R} \times \Sigma$ necessarily has a multivariate Poisson distribution, with dependence determined by nonempty intersections of the subsets $B_1, \ldots, B_k$. For example, given two random variables $\mathbf{M}(B_1)$ and $\mathbf{M}(B_2)$ for measurable subsets B_1 and B_2 in $R \times \Sigma$,

$$\mathbf{M}(B_1) = \mathbf{M}(B_1 \cap B_2) + \mathbf{M}(B_1 \cap B_2^c),$$

where B^c is the complement of B, and similarly for $\mathbf{M}(B_2)$, so that $\mathbf{M}(B_1)$ and $\mathbf{M}(B_2)$ each are sums of two independent Poisson random variables, with one in common. A corresponding representation as sums of subsets of $2^k - 1$ independent Poisson random variables holds for random variables $\mathbf{M}(B_1), \ldots, \mathbf{M}(B_k)$.

To express the consequences of this structure, recall that a bivariate probability mass function $p(j,k)$ is *totally positive of order 2* (TP_2) if for all integers $j_1 < j_2$ and $k_1 < k_2$

$$p(j_1,k_1)p(j_2,k_2) \geqslant p(j_1,k_2)p(j_2,k_1)\,;$$

see p. 92 of Barlow and Proschan [3]. Recall that two random variables X_1 and X_2 are *associated* if $\mathbf{Cov}[f(X_1), g(X_2)] \geqslant 0$ for all nondecreasing real-valued random variables f and g for which the expectations are well defined; see p. 29 of [3]. It is known that a multivariate distribution which is TP_2 in pairs is necessarily associated; see p. 149 of [3].

The following is the key lemma describing dependence between two random variables in our PALM; we omit the easy proof. Let $a \wedge b = \min\{a, b\}$.

LEMMA 2.5

For $i = 1, 2, 3$, let Y_i be independent nonnegative-integer-valued random variables with probability mass functions $p_i(k)$. Then $(Y_1 + Y_2, Y_1 + Y_3)$ has joint probability mass function

$$p(j,k) = \sum_{m=0}^{j \wedge k} p_1(m)p_2(j-m)p_3(k-m)\,, \tag{2.24}$$

which is TP_2 and thus associated. Moreover,

$$\mathbf{Cov}[Y_1 + Y_2,\, Y_1 + Y_3] = \mathbf{Var}[Y_1]\,. \tag{2.25}$$

If, in addition, p_i is Poisson for each i, then p in (2.24) is a bivariate Poisson distribution and

$$\mathbf{Cov}[Y_1 + Y_2,\, Y_1 + Y_3] = \mathbf{Var}[Y_1] = \mathbf{E}[Y_1]\,. \tag{2.26}$$

We now describe the joint distribution of $Q_{C_1}(t_1), \ldots, Q_{C_n}(t_n)$. The result for one queue ($\mathcal{S}^0 = \{1\}$) is theorem 1.2 of Eick et al. [21].

THEOREM 2.2

Consider the PALM and let $C_1, \ldots, C_n$ be (not necessarily disjoint) subsets of $\mathcal{S}$ and let $t_1, \ldots, t_n$ be time points. Then $(Q_{C_1}(t_1), \ldots, Q_{C_n}(t_n))$ has a multivariate Poisson distribution. Consequently, $Q_{C_1}(t_1), \ldots, Q_{C_n}(t_n)$ are associated random variables with

$$\begin{aligned}\mathbf{Cov}[Q_{C_1}(t_1), Q_{C_2}(t_2)] &= \mathbf{E}[\mathbf{M}(q(t_1, C_1) \cap q(t_2, C_2))] \\ &= \int_{-\infty}^{\infty} P(L_s(t_1) \in C_1, L_s(t_2) \in C_2)\alpha(s)\mathrm{d}s\,.\end{aligned} \tag{2.27}$$

Proof

Note that $Q_{C_i}(t_i) = \mathbf{M}(q(t_i, C_i)), 1 \leqslant i \leqslant n$. The multivariate Poisson property follows by the remarks before lemma 2.5. Then apply lemmas 2.4 and 2.5, noting that

$$\mathbf{M}(q(t_1, C_1) \cap q(t_2, C_2)) = \int_{-\infty}^{\infty} 1_{\{L_s(t_1) \in C_1, L_s(t_2) \in C_2\}} \mathrm{d}A(s) . \qquad \square$$

MODELING

We conclude this section with a brief discussion of modeling with PALMs. As should be clear from the above, to obtain useful PALMs for applications, the challenge is to obtain realistic location processes for which we can compute the time-dependent probability distributions, e.g., $P(L_s(t) \in C)$ in theorem 2.1.

A natural approach is *direct sampling from a function space*. This is most easily done starting with a finite or countably infinite set Σ_0 of location process sample paths $\{l^r(t) : t \in \mathbb{R}\}, r = 1, 2, \ldots$. We then stipulate that an arrival at time s has location function $\{l^r(t+s) : t \in \mathbb{R}\}$ with probability $p_s(r)$. The PALM is then specified by the set Σ_0 of sample paths, the assignment probabilities $p_s(r)$ and the external-arrival-rate function $\alpha(s)$.

Another natural approach is to treat the location processes as Markov. Then we can specify the location processes by their transition probabilities given the current state. This is illustrated by the $(M_t/M_t/\infty)^N/M_t$ model in section 8. The location processes are specified there by initial state probabilities $\pi_i(t)$, transition probabilities $p_{ij}(t)$ and individual service-rate functions $\mu_i(t)$. In applications these functions might be regarded as piecewise-constant functions of t. As indicated in section 8, the means and covariances can be found for this model by solving linear ODEs. Thus, this totally Markov $(M_t/M_t/\infty)^N/M_t$ model is relatively tractable.

We can also consider PALMs with Markov location processes, but without finite state spaces. For example, the location processes might be time-dependent Brownian motions in $\mathbb{R}^k$ or time-dependent reflected Brownian motions in $\mathbb{R}^k_+$.

Motivated by the wireless-telecommunication-system application, we introduce a new Markov location process for a PALM. This model has linear motion with random jumps; we call it a *Markov linear PALM*. A state s for each customer consists of his location $\mathbf{x}$ (a point in $\mathbb{R}^3$), *velocity* v a point in $[0, \infty)$) and *direction* $\mathbf{d}$ (a point in $\mathbb{R}^3$ with norm 1). Thus the state space $\mathcal{S}^0$ can be regarded as $\mathbb{R}^7$. In addition, there is a *jump-transition intensity function* $\eta(\mathbf{x}, v, \mathbf{d}, t)$ a collection of *initial-state probability measures* $\pi(t; \cdot)$ and a collection of *probability transition kernels* $p(t; (\mathbf{x}, v, \mathbf{d}), \cdot)$.

We assume that customers change their position continuously and linearly according to the current velocity v and direction $\mathbf{d}$, except at special jump transitions. Thus, if there are no jump transitions in the interval $[s, t]$ and if $\mathbf{x}(t)$ denotes the customer location at time t, then we have

$$\mathbf{x}(t) = \mathbf{x}(s) + v\mathbf{d}(t - s) , \tag{2.28}$$

where $\mathbf{d}$ is a direction vector with norm $\|\mathbf{d}\| = 1$. The jump transitions occur according to the jump-transition intensity function η.

As with any PALM, external arrivals occur according to the nonhomogeneous Poisson process A with external-arrival-rate function α. Upon arrival at time s, a customer is assigned an initial state $(\mathbf{x}, v, \mathbf{d})$ according to the probability measure $\pi(s; \cdot)$. (We assume that $\pi(s, C)$ is a probability measure on $\mathcal{S}^0$ for each s and a measurable function of s for each measurable subset C in $\mathcal{S}^0$.) Starting at time s in state $(\mathbf{x}, v, \mathbf{d})$, the customer moves from the initial location at constant velocity v in the direction $\mathbf{d}$. However, jump transitions occur at rate $\eta(\mathbf{x}, v, \mathbf{d}, t)$ in state $(\mathbf{x}, v, \mathbf{d})$ at time t. If a jump transition occurs in state $(\mathbf{x}_1, v_1, \mathbf{d}_1)$ at time t, then (as an approximation to the laws of physics) an instantaneous transition is made to state $(\mathbf{x}_2, v_2, \mathbf{d}_2)$ according to the transition probability kernel $p(t; (\mathbf{x}_1, v_1, \mathbf{d}_1), \cdot)$. (We assume that $p(t; (\mathbf{x}, v, \mathbf{d}), C)$ is a probability measure on $\mathcal{S}^0$ for each $(t; \mathbf{x}, v, \mathbf{d})$ and a measurable function on $\mathbb{R} \times \mathcal{S}^0$ for each measurable subset C on $\mathcal{S}^0$.) With probability $p(t, (\mathbf{x}_1, v_1, \mathbf{d}_1), \{\Delta^*\})$ the customer leaves the system at this jump transition point. This Markov linear PALM and variants seem to have the potential of realistically representing real systems, but it remains to effectively analyze these location processes.

For the open queueing network models considered below, the Markov property can be relaxed by assuming only that the successive transitions from queue to queue are Markov. A relatively simple model within this framework is the $(M_t/GI/\infty)^N/M_t$ model, which we discuss in section 7, in which the service times at the queues are mutually independent with general distributions depending only on the queue index. The model is then specified by the external-arrival-rate function α, the initial-state probabilities $\pi_i(t)$, the probability transition function $p_{ij}(t)$ and the service-time cdf's G_i, $1 \leqslant i \leqslant N$. The model specification is the same as for the $(M_t/M_t/\infty)^N/M_t$ model except that the service-time cdf's G_i replace the individual service-rate functions $\mu_i(t)$. The cdf's allow greater generality by getting away from the Markov property (the exponential special case), but they do not incorporate time dependence. As in theorem 2.1, the $(M_t/GI/\infty)^N/M_t$ model can be solved by iteratively solving the input equations, so that this model is also relatively tractable.

3. The $(M_t/G_t/\infty)^N/G_t$ open queueing network model

The $(M_t/G_t/\infty)^N/G_t$ open queueing network model is a special case of the PALM in section 2 in which $\mathcal{S}^0 = \{1, \ldots, N\}$, the location functions l are piecewise constant with only finitely many jumps and

$$l(t) = \Delta_* \text{ for } t < s \text{ and } l(t) = \Delta^* \text{ for } t \geqslant T_s \tag{3.1}$$

for some s and T_s with $-\infty < s < T_s < \infty$. Although it is not essential to do so, *we will assume that the time s in (3.1) is the external arrival time.* In this framework we write $Q_i(t)$ for $Q_{\{i\}}(t)$. Hence we have, by section 2,

$$Q_i(t) \equiv \int_{-\infty}^{\infty} 1_{\{L_s(t)=i\}} \mathrm{d}A(s)$$

and

$$m_i(t) \equiv \mathrm{E}[Q_i(t)] = \int_{-\infty}^{t} P(L_s(t) = i)\alpha(s)\mathrm{d}s\,,$$

with an analogous formula for $\mathbf{Cov}[Q_i(t_1), Q_j(t_2)]$.

A concrete example of the $(M_t/G_t/\infty)^N/G_t$ model that illustrates the generality is a nonstationary generalization of the semi-Markov compartmental models of Weiner and Purdue [69] and Purdue [59]. Then the location processes may be non-stationary semi-Markov processes. The transitions from queue to queue can depend on the length of the most recently completed service time as well as the queue index and the transition time.

With this additional structure, we can discuss the *external* arrival and departure processes, and the *aggregate* arrival and departure processes. We discuss these in turn. Our terminology is illustrated in fig. 3. The *external arrival process A* is the arrival process to the network from outside the network. The *external arrival process to queue i*, A_i, is the arrival process to queue i from outside the network; i.e., A_i counts the arrivals in A that visit queue i first. Similarly, the *external departure pro-*

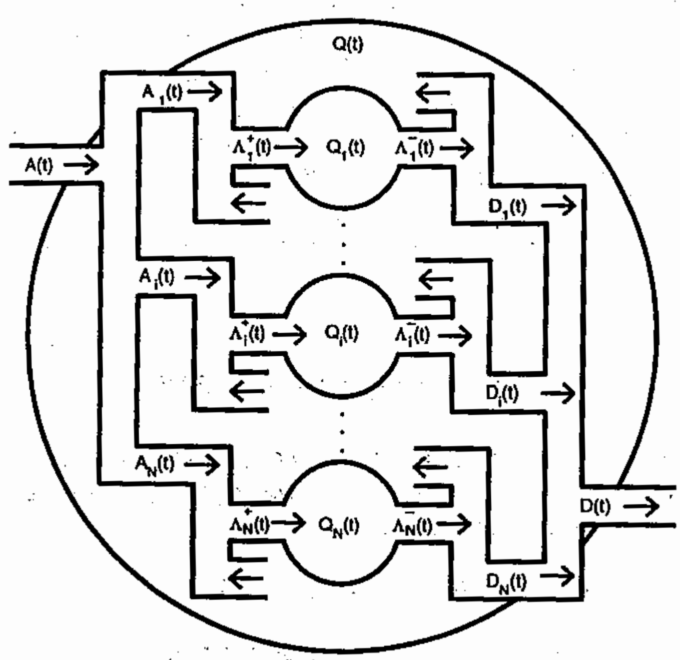

Fig. 3. Flow diagram for the processes related to infinite server networks.

cess D is the departure process from the network. The *external departure process from queue i*, D_i, is the departure process from the network from queue i. In other words, D_i counts those departures in D who visit queue i last.

In contrast, "aggregate" expresses the perspective of the individual queues. The *aggregate arrival process to queue i*, Λ_i^+, counts the arrivals to queue i from other queues as well as from outside the network. Similarly, the *aggregate departure process from queue i*, Λ_i^-, counts the departures from queue i that go to other queues as well as outside the network. Note that, in general, customers that are aggregate arrivals and departures may reappear again later as subsequent aggregate arrivals and departures, whereas this cannot happen for external arrivals and departures (under our model assumptions).

The processes $A, A_i, D, D_i, \Lambda_i^+$ and Λ_i^- are all understood to be counting processes; e.g., $\Lambda_i^+(t)$ counts the number of aggregate arrivals to queue i in $(-\infty, t]$. Since the external-arrival-rate function α is integrable, $A(\infty), A_i(\infty), D(\infty)$ and $D_i(\infty)$ are necessarily finite w.p.1, but we need to make extra assumptions to have $\Lambda_i^+(\infty)$ and $\Lambda_i^-(\infty)$ be finite.

Under our assumptions, the external arrival processes A and A_i have well defined rate or intensity functions α and α_i, but in general the other processes do not. When they do, the rate functions of D, D_i, Λ_i^+ and Λ_i^- will be denoted by $\delta, \delta_i, \lambda_i^+$ and λ_i^-, respectively.

For some of the results below, we use the following consequence of lemma 2.3. In the following lemma and its applications later, the counting processes fail to be Poisson processes only because the intensity measure need not be absolutely continuous; e.g., there can be multiple points at fixed times. The processes still have independent Poisson increments; we call these processes *generalized Poisson processes*. We give examples showing that the external departure process in an $M_t/G_t/\infty$ queue can be a generalized Poisson process without being a Poisson process after theorem 3.3. These examples apply to the following lemma too. In section 4 we give conditions for generalized Poisson processes to be Poisson processes.

LEMMA 3.1

Let $\{Z_k\}$ be an i.i.d. sequence of random elements of a Polish space Σ', let $\phi : \Sigma' \times \mathbb{R} \times \mathbb{R} \to \mathbb{S}$ be a measurable function with values in a Polish space $\mathbb{S}$ and let $X_s(t) = \phi(Z_{A(s)}, s, t)$. If $C_1, \ldots, C_k$ are disjoint subsets of $\mathbb{S}$ and if $1_{\{X_s(t) \in C_i\}}$ is nondecreasing in t for all s and i, then

$$\tilde{\mathbf{M}}(t, C_i) \equiv \int_{-\infty}^{\infty} 1_{\{X_s(t) \in C_i\}} \mathrm{d}A(s) \equiv \sum_{k=1}^{\infty} 1_{\{\phi(Z_k, \hat{A}_k, t) \in C_i\}}, \quad 1 \leqslant i \leqslant k, \tag{3.2}$$

are independent generalized Poisson processes with means

$$\mathrm{E}[\tilde{\mathbf{M}}(t, C_i)] = \int_{-\infty}^{\infty} P(X_s(t) \in C_i) \alpha(s) \mathrm{d}s. \tag{3.3}$$

Proof

Apply lemma 2.3 to show that the increments of $\tilde{\mathbf{M}}(t, C_i)$ for disjoint time intervals and different i are independent Poisson variables. In particular, for any positive integer m and time points $t_0 = -\infty < t_1 < \ldots < t_m < \infty$, consider

$$\tilde{\mathbf{M}}(t_j, C_i) - \tilde{\mathbf{M}}(t_{j-1}, C_i) = \int_{-\infty}^{\infty} \mathbf{1}_{\{(X_s(t_1),\ldots,X_s(t_m)) \in C_{ij}\}} \mathrm{d}A(s) , \tag{3.4}$$

where C_{ij} is the product set having the first $j-1$ components equal to C_i^c and the last $m-j+1$ components equal to C_i. Note that, by hypothesis, $\{X_s(t_{j-1}) \in C_i\} \subseteq \{X_s(t_j) \in C_i\}$ for all i and j. Hence, for each i, these m product sets are disjoint, and so the km integrals in (3.4) for $1 \leqslant i \leqslant k$ and $1 \leqslant j \leqslant m$ are independent Poisson random variables. Finally, (3.3) is a special case of (2.15). □

To immediately illustrate lemma 3.1, we give an extension of theorem 2.1. Let $Q_i^d(t,\tau)$ be the number of customers at queue i at time t that will *depart before time* $t+\tau$.

THEOREM 3.1

For each t, the processes $\{Q_i^d(t,\tau) : \tau \geqslant 0\}$ for $1 \leqslant i \leqslant N$ are independent generalized Poisson processes.

Proof

Note that

$$Q_i^d(t,\tau) = \int_{-\infty}^{\infty} \mathbf{1}_{\{T_s \leqslant t+\tau, L_s(t)=i\}} \mathrm{d}A(s)$$

for T_s in (3.1) and apply lemma 3.1. □

A similar result holds for $Q_i^a(t,\tau)$, the number of customers at queue i at time t that *arrived after time* $t-\tau$. In fact, the processes $\{Q_i^a(t;\tau) : \tau \geqslant 0\}$ are Poisson processes.

EXTERNAL ARRIVAL AND DEPARTURE PROCESSES

We first consider the queue where the arrival enters the network. Let A_i be the external arrival process to queue i in the $(M_t/G_t/\infty)^N/G_t$ model, which can be defined by

$$A_i(t) \equiv \int_{-\infty}^{t} \mathbf{1}_{\{L_s(s)=i\}} \mathrm{d}A(s) . \tag{3.5}$$

The following is an easy application of lemma 2.3; it just shows that independent thinnings of a Poisson process yields independent Poisson processes.

THEOREM 3.2

The external arrival processes $A_1, \dots, A_n$ are independent Poisson processes with external-arrival-rate functions

$$\alpha_i(s) = P(L_s(s) = i)\alpha(s). \tag{3.6}$$

The *external departure process* D can be defined as

$$D(t) \equiv \mathbf{M}(q(t, \{\Delta^*\})) = \int_{-\infty}^{\infty} \mathbf{1}_{\{L_s(t)=\Delta^*\}} \mathrm{d}A(s) = \int_{-\infty}^{\infty} \mathbf{1}_{\{T_s \leqslant t\}} \mathrm{d}A(s). \tag{3.7}$$

Similarly, the *external departure process from queue i*, D_i, can be defined by

$$D_i(t) \equiv \int_{-\infty}^{\infty} \mathbf{1}_{\{L_s(T_s-)=i, T_s \leqslant t\}} \mathrm{d}A(s). \tag{3.8}$$

The following is an extension of theorems 1.1, 2.1 and 2.2, which can be proved by applying lemmas 2.3 and 3.1 to (3.7), (3.8) and (2.21).

THEOREM 3.3

In the $(M_t/G_t/\infty)^N/G_t$ model, the vector external departure process before time t, $\{(D_1(s), \dots, D_N(s)) : s \leqslant t\}$, is independent of the vector $(Q_1(t), \dots, Q_N(t))$ for each t. Moreover, $Q_i(t), 1 \leqslant i \leqslant N$, are independent Poisson random variables and $D_i, 1 \leqslant i \leqslant N$, are independent generalized Poisson processes with means

$$\mathbf{E}[D_i(t)] = \int_{-\infty}^{\infty} P(L_s(T_s-) = i, T_s \leqslant t)\alpha(s)\mathrm{d}s \tag{3.9}$$

for T_s in (3.1).

EXAMPLE 3.1

The external departure processes D and D_i need *not* be Poisson processes, even when there is only one queue, because their intensity measures need not be absolutely continuous, so that they need not be either simple or orderly (see p. 28 of Daley and Vere-Jones [17]). To quickly see that the process D could have multiple points at a fixed time, consider a single queue with $\alpha(s) = 1$ for $0 \leqslant s \leqslant 1$ and $\alpha(s) = 0$ for $s < 0$. Let there be time-dependent deterministic service times, so that $T_s = 1, 0 \leqslant s \leqslant 1$. Then $D(1) - D(1-)$, the number of departures at time 1, is Poisson with mean $\int_0^1 \alpha(s)\mathrm{d}s$. Think of an infinite-capacity train that departs at time 1. □

EXAMPLE 3.2

In fact, the intensity measure of the departure process from an $M_t/G/\infty$ queue can be even more complicated. It need not be either absolutely continuous or discrete; it can be singular but continuous; see chapter 1 of Chung [14]. To see this, with α as above, let the departure times be $T_s = 1 + F^{-1}(s)$ for any cdf F on $[0, \infty)$, where $F^{-1}(s) = \inf\{t \geqslant 0 : F(t) > s\}, 0 \leqslant s \leqslant 1$. As in (2.7), $F^{-1}(U)$ has cdf F when

U is uniformly distributed on [0, 1]. Since the unordered arrival times in [0, 1] are distributed as independent uniform random variables over [0, 1] conditional on $A(1)$,

$$\mathrm{E}[D(1+t)|A(1)=k] = kF(t), \qquad t \geqslant 0,$$

so that

$$\mathrm{E}[D(1+t)] = \mathrm{E}[A(1)]F(t) = F(t), \quad t \geqslant 0.$$

Thus, $\mathrm{E}[D(1+t)]$ may have a singular continuous component as well as discrete and absolutely continuous components. □

THE AGGREGATE ARRIVAL AND DEPARTURE PROCESSES

We now describe the aggregate arrival and departure processes Λ_i^+ and Λ_i^-. Henceforth, where the results are similar, we only state results for Λ_i^+. Let $V_i^+(s,t)$ and $V_i^-(s,t)$, respectively, be the number of visits to queue i (arrivals to queue i) and completed visits to queue i (departures from queue i) up to time t by an arrival at time s. These processes are formally defined as in lemma 2.2. The processes Λ_i^+ and Λ_i^- are then defined as

$$\Lambda_i^+(t) \equiv \int_{-\infty}^{\infty} V_i^+(s,t)\mathrm{d}A(s) \quad \text{and} \quad \Lambda_i^-(t) \equiv \int_{-\infty}^{\infty} V_i^-(s,t)\mathrm{d}A(s). \tag{3.10}$$

Let $\Lambda_i^n(t)$ and $\tilde{\Lambda}_i^n(t)$ count the number of customers who have come to queue i exactly n times and at least n times, respectively, by time t. Let $\oplus$ denote the sum of mutually independent random elements; we use the familiar summation sign $\sum$ when there need not be independence of the summands. Thus we could reexpress results in theorems 2.1, 3.2 and 3.3 as

$$A = \bigoplus_{i=1}^{N} A_i, D = \bigoplus_{i=1}^{N} D_i \text{ and } Q(t) = \bigoplus_{i=1}^{N} Q_i(t) \quad \text{for each } t.$$

When the argument t is not included, we mean that the stochastic processes being added are mutually independent.

THEOREM 3.4

In the $(M_t/G_t/\infty)^N/G_t$ model,

$$\Lambda_i^+(t) = \sum_{n=1}^{\infty} \tilde{\Lambda}_i^n(t) = \bigoplus_{n=1}^{\infty} n\Lambda_i^n(t), \tag{3.11}$$

where

$$\Lambda_i^n(t) = \int_{-\infty}^{\infty} \mathbf{1}_{\{V_i^+(s,t)=n\}}\mathrm{d}A(s) \quad \text{and} \quad \tilde{\Lambda}_i^n(t) = \int_{-\infty}^{\infty} \mathbf{1}_{\{V_i^+(s,t)\geqslant n\}}\mathrm{d}A(s). \tag{3.12}$$

Moreover, the processes $\tilde{A}_i^n$ are generalized Poisson processes, with means

$$\mathbf{E}[\tilde{A}_i^n(t)] = \int_{-\infty}^{\infty} P(V_i^+(s,t) \geqslant n)\alpha(s)\mathrm{d}s < \infty, \tag{3.13}$$

and, for each t and for $n \geqslant 1$, $A_i^n(t)$ are mutually independent Poisson random variables with means

$$\mathbf{E}[A_i^n(t)] = \int_{-\infty}^{t} P(V_i^+(s,t) = n)\alpha(s)\mathrm{d}s < \infty. \tag{3.14}$$

Proof

To treat $\tilde{A}_i^n$, note that $\mathbf{1}_{\{V_i^+(s,t)\geqslant n\}}$ is nondecreasing in t for all s, i and n; then apply lemma 3.1. To treat $A_i^n(t)$ for each fixed t, apply lemma 2.3. □

We now state the consequences of theorem 3.4 for A_i^+ for A_i^-. It is important to note that so far we have not made assumptions strong enough to imply that $A_i^+(t)$ and $A_i^-(t)$ are finite. For this purpose, and to have finite variances, we assume that

$$\int_{-\infty}^{\infty} \mathbf{E}[V_i^+(s,\infty)^2]\alpha(s)\mathrm{d}s < \infty. \tag{3.15}$$

Note that $V_i^-(s,t) \leqslant V_i^+(s,t)$, so that no separate assumption is needed for $V_i^-(s,t)$.

THEOREM 3.5

Assume that (3.15) holds. The following eight statements are equivalent:

(i) $P(V_i^+(s,\infty) \leqslant 1) = 1$ for almost all s;
(ii) $P(V_i^-(s,\infty) \leqslant 1) = 1$ for almost all s;
(iii) A_i^+ is a generalized Poisson process;
(iv) A_i^- is a generalized Poisson process;
(v) $\mathbf{Var}[A_i^+(t)] = \mathbf{E}[A_i^+(t)]$ for all t;
(vi) $\mathbf{Var}[A_i^-(t)] = \mathbf{E}[A_i^-(t)]$ for all t;
(vii) $\mathbf{Cov}[Q_i(t), A_i^+(\infty) - A_i^+(t)] = 0$ for all t;
(viii) $\mathbf{Cov}[Q_i(t), A_i^-(t)] = 0$ for all t.

Proof

First, (i) and (ii) are equivalent, because $V_i^+(s,\infty) = V_i^-(s,\infty)$ for all s, by virtue of (3.1). Properties (i) and (ii) in turn imply (iii) and (iv) by theorem 3.4, because then $A_i^+ = \tilde{A}_i^1$. Obviously (iii) implies (v) and (iv) implies (vi). Next, to show that (v) implies (i), note that, by (2.11) and (2.16), (v) is equivalent to

$$\begin{aligned}\mathbf{E}\left[A_i^+(t)\right] &= \int_{-\infty}^{\infty} \mathbf{E}\left[V_i^+(s,t)\right]\alpha(s)\mathrm{d}s \\ &= \int_{-\infty}^{\infty} \mathbf{E}\left[V_i^+(s,t)^2\right]\alpha(s)\mathrm{d}s = \mathbf{Var}\left[A_i^+(t)\right].\end{aligned} \tag{3.16}$$

Since $V_i^+(s,t)$ is integer valued, we must have $V_i^+(s,t)$ equal to 0 or 1 for all t and almost all s with respect to the measure $\alpha(s)\mathrm{d}s$. This implies that $V_i^+(s,\infty)\leqslant 1$ for almost all s, and so (v) implies (i). Similarly, (vi) implies (ii). Finally, we consider (vii) and (viii). Note that (iii) implies (vii) and (viii), because under (iii), queue i behaves like an $M_t/G/\infty$ queue in isolation, except that the arrival process need not have an intensity. Thus, we can apply theorem 3.3 extended to this case. Now, (viii) implies (ii) because

$$\mathbf{Cov}[Q_i(t), A_i^-(t)] = \int_{-\infty}^{\infty} \mathbf{E}[V_i^-(s,t); L_s(t) = i]\alpha(s)\mathrm{d}s$$

for all t, by lemma 2.4. If $\mathbf{Cov}[Q_i(t), A_i^-(t)] = 0$ for all t, then

$$\mathbf{E}[V_i^-(s,t); L_s(t) = i] = 0$$

for all t almost everywhere with respect to $\alpha(s)\mathrm{d}s$. This implies for almost all s with respect to $\alpha(s)\mathrm{d}s$ that $V_i^-(s,t) = 0$ on $\{L_s(t) = i\}$, i.e., there are no completed visits from i by a customer at queue i, which implies (ii). To see that (vii) implies (i), we use lemma 2.4 to obtain

$$\mathbf{Cov}[Q_i(t), A_i^+(\infty) - A_i^+(t)] = \int_{-\infty}^{\infty} \mathbf{E}[V_i^+(s,\infty) - V_i^+(s,t); L_s(t) = i]\alpha(s)\mathrm{d}s.$$

Since $V_i^+(s,\infty)\geqslant V_i^+(s,t)$ for all s and t, (vii) is equivalent to

$$\mathbf{E}[V_i^+(s,\infty) - V_i^+(s,t); L_s(t) = i] = 0$$

for almost all s with respect to the measure $\alpha(s)\mathrm{d}s$. This says that any arrival that visits node i at any time t, never revisits node i. This is then equivalent to (i), which completes the proof. □

We now describe the aggregate arrival processes A_i^+ when customers *can* visit the queues more than once. A similar result holds for the aggregate departure processes A_i^-. For this purpose, we introduce a new class of multivariate distributions. We say that a k-dimensional discrete distribution is a *positive-linear multivariate Poisson distribution* if it is distributed as a vector $(Y_1,\dots,Y_k)$, where

$$Y_i = \sum_{j=1}^{\infty} m_{ij}Z_j, \quad 1\leqslant i\leqslant k, \tag{3.17}$$

$\{Z_j : j\geqslant 1\}$ is a sequence of mutually independent Poisson random variables and m_{ij} is a nonnegative integer for each i and j. The multivariate Poisson distribution in theorem 2.1 is the special case in which m_{ij} is equal to 0 or 1 for all i and j.

Let $\mathbb{Z}_+$ be the set of nonnegative integers.

LEMMA 3.2

If $\phi_1,\dots,\phi_n$ are $\mathbb{Z}_+$-valued, measurable functions in the setting of lemma 2.2, then $(Z_{\phi_1}(\infty), \ldots, Z_{\phi_n}(\infty))$ has a positive-linear multivariate Poisson distribution.

Proof

Let $\mathbf{m} = (m_1, \ldots, m_n)$ and $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n)$. Then

$$Z_{\phi_i}(\infty) = \sum_{\mathbf{m} \in \mathbf{Z}_+^n} m_i \cdot \int_{-\infty}^{\infty} \mathbf{1}_{\{\boldsymbol{\phi}(Z_{A(s)},s)=\mathbf{m}\}} \mathrm{d}A(s) \tag{3.18}$$

for all i, $1 \leqslant i \leqslant n$. To complete the proof apply lemma 2.3. □

THEOREM 3.6

Consider the $(M_t/G_t/\infty)^N/G_t$ model and suppose that (3.15) holds. Then the increments of A_i^+ for $1 \leqslant i \leqslant N$ have a positive-linear multivariate Poisson distribution. Consequently, the increments are associated with

$$\mathbf{E}[A_i^+(t_2) - A_i^+(t_1)] = \int_{-\infty}^{\infty} \mathbf{E}[V_i^+(s, t_2) - V_i^+(s, t_1)]\alpha(s)\mathrm{d}s < \infty,$$

$$\mathbf{Var}[A_i^+(t_2) - A_i^+(t_1)] = \int_{-\infty}^{\infty} \mathbf{E}[(V_i^+(s, t_2) - V_i^+(s, t_1))^2]\alpha(s)\mathrm{d}s < \infty,$$

and

$$\mathbf{Cov}[A_i^+(t_2) - A_i^+(t_1), A_j^+(t_4) - A_j^+(t_3)]$$
$$= \int_{-\infty}^{\infty} \mathbf{E}[(V_i^+(s, t_2) - V_i^+(s, t_1))(V_j^+(s, t_4) - V_j^+(s, t_3))]\alpha(s)\mathrm{d}s < \infty$$

for $t_1 < t_2$ and $t_3 < t_4$.

Proof

Apply lemma 3.2, noting that $A_i^+(t_{l+1}) - A_i^+(t_l) = Z_{\phi_{il}}(\infty)$, where

$$\phi_{il}(Z_{A(s)}, s) = V_i^+(s, t_{l+1}) - V_i^+(s, t_l)$$

for $1 \leqslant i \leqslant N$ and $1 \leqslant l \leqslant m$, where $t_1 \leqslant \ldots \leqslant t_{m+1}$. Since independent random variables are associated and increasing functions of associated random variables are associated, the increments are associated; see p. 30 of [3]. The formulas now follow from lemmas 2.2 and 2.4 or by direct calculation using (3.18). □

4. Rates and reversibility

We say that a counting process (point process) has a rate if its intensity measure is absolutely continuous with respect to Lebesgue measure; e.g., we say that the external departure process D has a rate δ if

$$\mathbf{E}[D(t)] = \int_{-\infty}^{t} \delta(s)\mathrm{d}s, \quad t \in \mathbb{R}, \tag{4.1}$$

for an integrable function δ; then we call δ the *external-departure-rate function*. In examples 3.1 and 3.2 we saw that the external departure process D need not have a rate, even in a single $M_t/G_t/\infty$ queue. In this section we determine necessary and sufficient conditions for the external departure process and other processes to have rates. Then we show that the $(M_t/G_t/\infty)^N/G_t$ model has a dynamic reversibility property when the external departure process has a rate.

RATES

By (3.7) and lemma 2.3, we can express the mean number of departures up to time t as

$$\mathbf{E}[D(t)] = \int_{-\infty}^{\infty} P(t_s \leqslant t)\alpha(s)\mathrm{d}s. \tag{4.2}$$

The following elementary result expresses standard definitions in the context of (4.2); e.g., see p. 104 of Royden [62].

THEOREM 4.1

(a) The intensity measure associated with $\mathbf{E}[D(t)]$ has an atom at t if and only if

$$\mathbf{E}[D(t) - D(t-)] = \int_{-\infty}^{\infty} P(t_s = t)\alpha(s)\mathrm{d}s > 0.$$

(b) The intensity measure associated with $\mathbf{E}[D(t)]$ is absolutely continuous, so that (4.1) holds if and only if for all a, b with $0 < a < b < \infty$ and $\epsilon > 0$ there is a $\delta > 0$ such that

$$\sum_{k=1}^{n} \mathbf{E}[D(t'_k) - D(t_k)] = \int_{-\infty}^{\infty} \sum_{k=1}^{n} P(t_k < T_s \leqslant t'_k)\alpha(s)\mathrm{d}s < \epsilon$$

for every finite collection $\{(t_k, t'_k) : 1 \leqslant k \leqslant n\}$ of nonoverlapping subintervals of $[a, b]$ with $\sum_{k=1}^{n}(t'_k - t_k) < \delta$.

Recall that not having an atom is necessary but not sufficient for absolute continuity. As illustrated by example 3.2, to have absolute continuity we also need to rule out the singular continuous case. The following expresses convenient sufficient conditions for absolute continuity. Condition (i) is used in the totally Markov case in section 8, while condition (ii) is used to treat $(M_t/G/\infty)^N/G_t$ models since they can be reduced to $(M_t/G/\infty)^N/D$ models.

THEOREM 4.2

The following are each sufficient conditions for the external departure process D to have a rate:

(i) If there exists a jointly measurable function $f_s(t)$ such that

$$P(T_s \leqslant t) = \int_{-\infty}^{t} f_s(u)\mathrm{d}u \tag{4.3}$$

for almost all s, then D has an integrable rate function

$$\delta(t) = \int_{-\infty}^{\infty} f_s(t)\alpha(s)\mathrm{d}s, \quad t \in \mathbb{R}.$$

(ii) If the customer sojourn-time distribution $P(T_s - s \leqslant t), t \in \mathbb{R}$, is independent of the customer arrival time s, then D has an integrable rate function

$$\delta(t) = \mathbf{E}[\alpha(t - T_0)].$$

Proof

For (i), combine (4.2) and (4.3) to get

$$\mathbf{E}[D(t)] = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{t} f_s(u)\mathrm{d}u \right] \alpha(s)\mathrm{d}s,$$

which is finite because α is integrable, and then apply Tonelli, p. 270 of Royden [62], to interchange the order of the integrals. For (ii), apply the proof of theorem 1.1, noting that to treat the external departure process D we can regard the entire network as a single $M_t/GI/\infty$ queue with service times distributed as T_0. Once again, the integrability of α implies that δ is integrable. □

We have seen that the existence of a rate for the external arrival process D can be determined by considering a single $M_t/G_t/\infty$ queue with service times $T_s - s$, where T_s is the departure time. In the same way we can treat the aggregate arrival and departure processes Λ_i^+ and Λ_i^-. We only discuss Λ_i^+; a similar result holds for Λ_i^-. The role of T_s is now played by

$$T_s^+(i,n) = \inf\{t : V_i^+(s,t) \geqslant n\}, \tag{4.4}$$

with the infimum equal $+\infty$ when it is not attained. Since the analysis is the same, we do not present further details.

We state some consequences for $(M_t/G/\infty)^N/G$ models. We apply theorems 3.3 and 3.4 with the analysis above.

THEOREM 4.3

For the $(M_t/G/\infty)^N/G$ model, the external departure processes D_i and the counting processes $\tilde{\Lambda}_i^n$ all are Poisson processes. Their (integrable) intensity functions are, respectively,

$$\delta_i(t) = \mathbf{E}[\alpha(t - T_0)]P(L_0(T_0-) = i) \tag{4.5}$$

and

$$\tilde{\lambda}_i^n(t) = \mathbf{E}[\alpha(t - T_0^+(i,n))], \tag{4.6}$$

where $\alpha(-\infty) = 0$ and $T_0^+(i,n)$ is defined in (4.4).

We can combine theorems 3.3 and 4.3 to conclude that the aggregate arrival processes Λ_i^+ have well defined intensity functions.

COROLLARY

In the $(M_t/G/\infty)^N/G$ model, if $\mathbf{E}[V_i^+(0,\infty)] < \infty$, then

$$\mathbf{E}[\Lambda_i^+(t)] = \int_{-\infty}^{t} \lambda_i^+(s)\mathrm{d}s < \infty ,$$

where

$$\lambda_i^+(t) = \sum_{n=1}^{\infty} \tilde{\lambda}_i^n(t)$$

for $\tilde{\lambda}_i^n$ in (4.6). Moreover, the processes Λ_i^+ are Poisson processes if and only if $P(V_i^+(0,\infty) \leqslant 1) = 1$.

Proof

From theorems 3.4 and 4.3, we have

$$\mathbf{E}[\Lambda_i^+(t)] = \sum_{n=1}^{\infty} \mathbf{E}[\tilde{\Lambda}_i^n(t)] = \sum_{n=1}^{\infty} \int_{-\infty}^{t} \tilde{\lambda}_i^n(s)\mathrm{d}s ,$$

which yields the result after interchanging the sums and integrals. The moments are finite because

$$\mathbf{E}[\Lambda_i^+(t)] \leqslant \mathbf{E}[V_i^+(0,\infty)] \int_{-\infty}^{\infty} \alpha(s)\mathrm{d}s < \infty .$$

Finally, theorem 3.5 implies that Λ_i^+ is a generalized Poisson process if and only if $\mu(\{s : V_i^+(s,\infty) \leqslant 1\}) = 1$. Here with i.i.d. location processes this condition is equivalent to $P(V_i(0,\infty) \leqslant 1) = 1$. The absolute continuity provides the remaining property to have a Poisson process. □

Remark 4.1

Even though Λ_i^+ and Λ_i^- in the $(M_t/G/\infty)^N/G$ queue are in general not Poisson processes, they are well studied processes, in particular *Poisson cluster processes*; see p. 236 of Daley and Vere-Jones [17]. A customer's first visit to queue i determines the cluster-center process, while subsequent visits to queue i provide the remaining points in the cluster. □

REVERSIBILITY

For stationary queueing network models, reversibility notions play a prominent role, so it is natural to look for nonstationary analogs here. First, note that theorems 2.1, 3.2 and 3.3 imply a form of *network dynamic-quasi-reversibility* for $(M_t/G_t/\infty)^N/G_t$ models, see p. 90 of Walrand [68], in that the *future arrivals, past departures and current state are mutually independent*; i.e., for each t, $\{A_1(s) - A_1(t) : s > t\}, \ldots, \{A_N(s) - A_N(t) : s > t\}, Q_1(t), \ldots, Q_N(t)$ and $\{D_1(s) : s \leqslant t\}, \ldots, \{D_N(s) : s \leqslant t\}$ are mutually independent (without any stationarity). However, in general this property does not hold for the aggregate flows at each queue; i.e., in general,

$$\{\Lambda_i^+(s) - \Lambda_i^+(t) : s > t\}, \{\Lambda_i^-(s) : (s) \leqslant t\} \text{ and } Q_i(t)$$

are *not* mutually independent. This is easy to see, because of the possibility of multiple visits. If a queue can be visited only once by each customer, then this dynamic quasi-reversibility property holds for that queue too. This is essentially just theorem 1.1 again.

Given an $(M_t/G_t/\infty)^N/G_t$ model for which the departure process D has a rate, form an associated *reverse-time* model with external arrival process A^* and external-arrival-rate function α^* defined in terms of the original external departure process D and its external-departure-rate function δ by setting

$$A^*(s) = D(\infty) - D((-s)-), \quad t \in \mathbb{R}. \tag{4.7}$$

Then A^* is a Poisson process with intensity $\alpha^*(s) = \delta(-s)$.

We construct the rest of the reverse-time system by continuing to reverse time. Given the arrival times $s_1, \ldots, s_k$ and the departure times $T_{s_1}, \ldots, T_{s_k}$ associated with the original system with external arrival process A, we have let the arrival times associated with A^* be $-T_{s_1}, \ldots, -T_{s_k}$. We now let the respective departure times be $-s_1, \ldots, -s_k$. Then, for $1 \leqslant m \leqslant k$, we let the reverse-time location processes be

$$L^*_{-T_{s_m}}(-t) = L_{s_m}(t), \quad t \in \mathbb{R}. \tag{4.8}$$

However, note that the arrival times A_m^* associated with A^* are the times $-T_{s_1}, \ldots, -T_{s_k}$ arranged in increasing order. From the above construction, we have the following result.

THEOREM 4.4

The reverse-time system associated with an $(M_t/G_t/\infty)^N/G_t$ system with a departure rate is another $(M_t/G/\infty)^N/G$ system with a departure rate for which $A^*(\infty) = A(\infty)$. The number of busy servers in the two systems is related by $Q_i^*(t) = Q_i(-t)$ for all i and t. The reverse-time system associated with the reverse-time system is the original system.

Remark 4.2

We have assumed that the departure process D has a rate in theorem 4.4 only so that the reverse-time arrival process A^* has a rate. We could have reversibility without this condition if we allowed the external arrival process to be a generalized Poisson process. The theory can be extended to this case. □

Remark 4.3

The reverse-time construction above makes

$$\{Q_i^*(-t), 1 \leqslant i \leqslant N, t \in \mathbb{R}\} \stackrel{\mathrm{d}}{=} \{Q_i(t) : 1 \leqslant i \leqslant N, t \in \mathbb{R}\}, \tag{4.9}$$

but the arrival processes A^* and A typically have *different external-arrival-rate functions*. Indeed, in the $(M_t/G/\infty)^N/G$ model,

$$\alpha^*(s) = \delta(-s) = \mathbf{E}[\alpha(-s - T_0)]. \tag{4.10}$$

However, if T_0 is deterministic and α is symmetric about 0 and periodic with period T, then

$$\alpha^*(s) = \mathbf{E}[\alpha(-s - T_0)] = \alpha(-s - T_0) = \alpha(-s) = \alpha(s). \tag{4.11}$$

An example of such an arrival-rate function is $\alpha(s) = a + b\cos(\gamma s)$.

Not only are the arrival-rate functions α^* and α typically different, but so are the distributions of the location processes L_s^* and L_s. However, it is possible for L_s^* and L_s to have the same distribution. For example, consider the $(M_t/GI/\infty)^N/D$ model with routing by a deterministic finite queue sequence. If the reversed queue sequence is the same as the forward queue sequence, then $L_s^* = L_s$. Consistent with previous terminology, we can say the $(M_t/G_t/\infty)^N/G_t$ model with a departure rate is *fully reversible* if $\alpha^* = \alpha$ and $L_s^* \stackrel{\mathrm{d}}{=} L_s$ for all s, but the weaker notion in theorem 4.4 seems worthwhile. □

5. Decomposition into deterministic routes

We now discuss the reduction of the $(M_t/G_t/\infty)^N/G_t$ model to the $(M_t/G_t/\infty)^N/D$ model by focusing on customers following different routes (queue sequences) separately. As discussed after theorem 1.2, this is a convenient device for proofs, but it also can be exploited in modeling. It is natural to think of the overall model as a multiclass model in which each class follows its own deterministic route. We might specify the model by identifying the relevant routes and indicating the time-dependent external-arrival-rate function for each.

LOCATION PROCESSES VIA ROUTES AND SERVICE TIMES

Since it is natural to specify the location functions $\{l(t) : t \geqslant 0\}$ in the $(M_t/G_t/\infty)^N/G_t$ model via the finite sequence of queues visited and the service times at these queues, we could let Σ be the set of $(2k)$-tuples $(n_1, \ldots, n_k; v_1, \ldots, v_k)$, where n_i is the index of the ith queue visited and v_i is the service time there. We also let Δ_* and let Δ^* be elements of Σ. We then topologize Σ by making it the countable topological sum over the positive integers k of the product spaces $\{1,\ldots,N\}^k \times \mathbb{R}^k$ with their usual topologies and $\{\Delta_*, \Delta^*\}$. By propositions 1 and 2 on pp. 195, 196 of Bourbaki [7], this makes Σ Polish. Indeed, Σ defined this way is equivalent to the location-function representation. This is made precise by the following result; we omit the easy proof.

LEMMA 5.1

The mapping of piecewise-constant location functions l in $D(\mathbb{R}, \mathbb{S})$ with finitely many jumps and $l(t) = \Delta_*$ for $t < s$ and $l(t) = \Delta^*$ for $t \geqslant T_s$ for some s and T_s with $-\infty < s < T_s < \infty$, where D is endowed with the Skorohod [65] J_1 topology into the space of queue sequences and service times with the topological-sum topology above is a homeomorphism.

Moreover, we can decompose the arrival process A into a sequence of independent Poisson processes each with its own finite deterministic route (sequence of queues visited). To do this, we can index the finite queue sequences $(n_1, \ldots, n_k)$ by a unique route r and let A^r be the arrival process for route r, $P(\langle L_s \rangle = r)$ be the conditional probability of route r being initiated at time s given that there is an arrival at time s, and let $Q_i^r(t)$ be the number of customers following route r that are at queue i at time t. Let R be the (at most countably infinite) set of all routes. For each route r, we can use (2.5) and lemma 2.1 to construct the service times. Using lemma 2.3 once again, we have the following result.

THEOREM 5.1

The $(M_t/G_t/\infty)^N/G_t$ model can be expressed as the superposition of independent $(M_t/G_t/\infty)^N/D$ models, i.e., the arrival processes A^r associated with different routes r are independent Poisson processes with arrival-rate functions

$$\alpha^r(s) = \alpha(s) P(\langle L_s \rangle = r) . \tag{5.1}$$

Moreover, for $1 \leqslant i \leqslant N$ and $r \in R$, the queue lengths $Q_i^r(t)$ are independent Poisson random variables, and the external departure processes D_i^r are mutually independent generalized Poisson processes.

Remark 5.1

It should be noted that the fixed routing in the $(M_t/G_t/\infty)^N/D$ model is *not* a special case of Markov routing, because a fixed route may visit a given queue more than once and go to a different queue after each visit. However, the analysis can be *reduced* to the $(M_t/G_t/\infty)^N/M$ model with homogeneous Markov routing, because we can always regard different visits to the same queue as different queues by convention (as discussed after theorem 1.2). Thus, the $(M_t/G_t/\infty)^N/G_t$ model actually goes beyond the model of theorem 1.2 only by allowing stochastic dependence and time dependence in the service times. □

To efficiently express the results in this multi-route framework, we introduce some additional notation. First, let $|r|$ be the length of route r. At the same time, let r be a mapping from the set $\{1, \ldots, |r|\}$ of *site indices* to the set $\{1, \ldots, N\}$ of *queue indices*, so that $r(k)$ is the index of the queue at the kth site on route r, and $r^{-1}(i)$ is the set of sites on route r that are queue i, e.g., if the sequence of queues visited on route r is $(7, 1, 4, 1)$, then $|r| = 4, r(1) = 7$ and $r^{-1}(-1) = \{2, 4\}$.

For a given route r, let $\Lambda_k^r \equiv \{\Lambda_k^r(t) : t \in \mathbb{R}\}$ be the departure process from the kth site of route r, with Λ_0^r the external arrival process for route r. Let $\theta_i(R)$ be the set of all routes that *start* at queue i, and let $\theta^i(R)$ be the set of all routes that *terminate* at queue i; i.e.,

$$\theta_i(R) = \{r \in R : r(1) = 1\} \text{ and } \theta^i(R) = \{r \in R : r(|r|) = i\} . \tag{5.2}$$

Once again, let $\oplus$ and $\sum$ denote sums for which independence is, and is not, assumed for the summands, respectively.

THEOREM 5.2

The counting processes Λ_k^r are generalized Poisson processes and the following decompositions hold for the external processes:

$$A_i = \bigoplus_{r \in \theta_i(R)} \Lambda_0^r \quad \text{and} \quad D_i = \bigoplus_{r \in \theta^i(R)} \Lambda_{|r|}^r .$$

Moreover, the aggregate arrival and departure processes decompose as

$$\Lambda_i^+ = \bigoplus_{r \in R} \left(\sum_{k \in r^{-1}(i)} \Lambda_{k-1}^r \right) \text{ and } \Lambda_i^- = \bigoplus_{r \in R} \left(\sum_{k \in r^{-1}(i)} \Lambda_k^r \right) .$$

Proof

First, observe that

$$\Lambda_0^r(t) = \int_{-\infty}^{t} \mathbf{1}_{\{\langle L_s \rangle = r\}} \mathrm{d}A(s)$$

and

$$\Lambda_{|r|}^r(t) = \int_{-\infty}^{\infty} \mathbf{1}_{\{\langle L_s \rangle = r, T_s \leqslant t\}} \mathrm{d}A(s) .$$

Now we have

$$A_i(t) = \int_{-\infty}^{t} \mathbf{1}_{\{L_s(s) = i\}} \mathrm{d}A(s) = \bigoplus_{r \in R} \left(\int_{-\infty}^{t} \mathbf{1}_{\{L_s(s) = i, \langle L_s \rangle = r\}} \mathrm{d}A(s) \right)$$

$$= \bigoplus_{r \in \theta_i(R)} \left(\int_{-\infty}^{t} \mathbf{1}_{\{\langle L_s \rangle = r\}} \mathrm{d}A(s) \right) = \bigoplus_{r \in \theta_i(R)} \Lambda_0^r(t) ,$$

and

$$D_i(t) = \int_{-\infty}^{\infty} \mathbf{1}_{\{L_s(T_s-)=i, T_s \leqslant t\}} \mathrm{d}A(s)$$
$$= \bigoplus_{r\in R} \left(\int_{-\infty}^{\infty} \mathbf{1}_{\{L_s(T_s-)=i, \langle L_s\rangle = r, T_s \leqslant t\}} \mathrm{d}A(s) \right)$$
$$= \bigoplus_{r\in \theta^i(R)} \left(\int_{-\infty}^{\infty} \mathbf{1}_{\{\langle L_s\rangle = r, T_s \leqslant t\}} \mathrm{d}A(s) \right) = \bigoplus_{r\in \theta^i(R)} \Lambda^r_{|r|}(t) .$$

If $V^r_k(s,t)$ equals one when the arrival at time s has left its kth site before time t and zero otherwise, we then have

$$\Lambda^k_r(t) = \int_{-\infty}^{\infty} V^r_k(s,t) \mathrm{d}\Lambda^r_0(s) = \int_{-\infty}^{\infty} V^r_k(s,t) \mathbf{1}_{\{\langle L_s\rangle = r\}} \mathrm{d}A(s) ,$$

with

$$V^+_i(s,t) = \bigoplus_{r\in R, k\in r^{-1}(i)} \sum V^r_{k-1}(s,t) \mathbf{1}_{\{\langle L_s\rangle = r\}}$$

and

$$V^-_i(s,t) = \bigoplus_{r\in R} \sum_{k\in r^{-1}(i)} V^r_k(s,t) \mathbf{1}_{\{\langle L_s\rangle = r\}} .$$

Here, it is important to observe that $V^k_r(s,t)$ can be written as an indicator function so it is immediate that Λ^r_k is a generalized Poisson process. The final two decomposition results hold because

$$\Lambda^+_i(t) = \int_{-\infty}^{\infty} V^+_i(s,t) \mathrm{d}A(s) = \bigoplus_{r\in R} \left(\int_{-\infty}^{\infty} V^+_i(s,t) \mathbf{1}_{\{\langle L_s\rangle = r\}} \mathrm{d}A(s) \right)$$
$$= \bigoplus_{r\in R} \left(\sum_{k\in r^{-1}(i)} \int_{-\infty}^{\infty} V^r_{k-1}(s,t) \mathrm{d}\Lambda^r_0(s) \right) = \bigoplus_{r\in R} \left(\sum_{k\in r^{-1}(i)} \Lambda^r_{k-1}(t) \right)$$

and

$$\Lambda^-_i(t) = \int_{-\infty}^{\infty} V^-_i(s,t) \mathrm{d}A(s) = \bigoplus_{r\in R} \left(\int_{-\infty}^{\infty} V^-_i(s,t) \mathbf{1}_{\{\langle L_s\rangle = r\}} \mathrm{d}A(s) \right)$$
$$= \bigoplus_{r\in R} \left(\sum_{k\in r^{-1}(i)} \int_{-\infty}^{\infty} V^r_k(s,t) \mathrm{d}\Lambda^r_0(s) \right) = \bigoplus_{r\in R} \left(\sum_{k\in r^{-1}(i)} \Lambda^r_k(t) \right) . \qquad \square$$

We now see what is gained with additional structure. In particular, we now assume that there is no time dependence in the service times. In this framework, let

$\bar{S}^r_k$ be the sojourn time (sum of service times) at the first k sites of route r and let $k^r = \{k, k+1, \ldots, |r|\}$. Let $x^+ = \max\{x, 0\}$.

THEOREM 5.3

For the $(M_t/G/\infty)^N/G_t$ model, the arrival process at the kth site of route r, A^r_k, is a Poisson process for each r and k. The aggregate arrival and departure rate functions are

$$\lambda^+_i(t) = \sum_{r \in R} \sum_{k \in r^{-1}(i)} \mathbf{E}[\alpha^r(t - \bar{S}^r_{k-1})]$$

and

$$\lambda^-_i(t) = \sum_{r \in R} \sum_{k \in r^{-1}(i)} \mathbf{E}[\alpha^r(t - \bar{S}^r_k)]\,. \tag{5.3}$$

The queue lengths have means

$$m_i(t) = \sum_{r \in R} \sum_{k \in r^{-1}(i)} \mathbf{E}\left[\int_{t-\bar{S}^r_k}^{t-\bar{S}^r_{k-1}} \alpha^r(s)\mathrm{d}s\right] \tag{5.4}$$

and covariances

$$\mathbf{Cov}[Q_i(t), Q_j(u)] = \sum_{r \in R} \sum_{k \in r^{-1}(i)} \sum_{l \in r^{-1}(j) \cap k^r} \mathbf{E}\left[\int_{(t-\bar{S}^r_k) \vee (u-\bar{S}^r_l)}^{(t-\bar{S}^r_{k-1}) \wedge (u-\bar{S}^r_{l-1})} \alpha^r(s)\mathrm{d}s\right]^+ . \tag{5.5}$$

Proof

The formulas in (5.3) follow from theorems 1.1 and 5.2. Formulas (5.4) and (5.5) hold because

$$\begin{aligned}
\mathbf{E}[Q_i(t)] &= \int_{-\infty}^{t} P(L_s(t) = i)\alpha(s)\mathrm{d}s \\
&= \sum_{r \in R} \int_{-\infty}^{t} P(L_s(t) = i | \langle L_s \rangle = r)\alpha^r(s)\mathrm{d}s \\
&= \sum_{r \in R} \sum_{k \in r^{-1}(i)} \int_{-\infty}^{t} P(\bar{S}^r_{k-1} < t - s \leqslant \bar{S}^r_k)\alpha^r(s)\mathrm{d}s \\
&= \sum_{r \in R} \sum_{k \in r^{-1}(i)} \mathbf{E}\left[\int_{t-\bar{S}^r_k}^{t-\bar{S}^r_{k-1}} \alpha^r(s)\mathrm{d}s\right]
\end{aligned}$$

and

$$\begin{aligned}\mathrm{Cov}[Q_i(t), Q_j(u)] &= \int_{-\infty}^{t} P(L_s(t) = i, L_s(u) = j)\alpha(s)\mathrm{d}s \\ &= \sum_{r\in R}\int_{-\infty}^{t} P(L_s(t) = i, L_s(u) = j \mid \langle L_s\rangle = r)\alpha^r(s)\mathrm{d}s \\ &= \sum_{r\in R}\sum_{k\in r^{-1}(i)}\sum_{l\in r^{-1}(j)\cap k^r}\int_{-\infty}^{t} P(\bar{S}^r_{k-1} < t-s \leqslant \bar{S}^r_k, \bar{S}^r_{l-1} < u-s \leqslant \bar{S}^r_l) \\ &\quad \times \alpha^r(s)\mathrm{d}s \\ &= \sum_{r\in R}\sum_{k\in r^{-1}(i)}\sum_{l\in r^{-1}(j)\cap k^r}\mathbf{E}\left[\int_{(t-\bar{S}^r_k)\vee(u-\bar{S}^r_l)}^{(t-\bar{S}^r_{k-1})\wedge(u-\bar{S}^r_{l-1})}\alpha^r(s)\mathrm{d}s\right]^+ . \qquad \square\end{aligned}$$

UNIFORM ACCELERATION

We conclude this section by describing an asymptotic expansion for the $(M_t/G/\infty)^N/G$ network. This extends results for the $M_t/GI/\infty$ queue in section 3 of Eick et al. [21]; see remark 3.4. For additional background, see Massey [45]. In particular, we construct a family of systems indexed by positive ϵ and let $\epsilon \downarrow 0$. This speeds up the rates, so that the behaviour at time t is increasingly determined by the local behavior (near t) of the model data.

For each $\epsilon > 0$, let $(Q^\epsilon_1(t), \ldots, Q^\epsilon_N(t))$ be the vector queue-length process in an $(M_t/G/\infty)^N/G$ network, where the arrival rate is $\alpha(t)/\epsilon$ and, for a given route r, the sojourn time from the arrival to site 1 until the departure from site k is $\epsilon\bar{S}^r_k$. Since we are using G routing, the quantity $P(\langle L_s\rangle = r)$ is independent of the arrival time s, and so the "accelerated" version of the rate $\alpha^r(t)$ is simply $\alpha^r(t)/\epsilon$. Making use of the Taylor series expansion results from Eick et al. [21], we get the following expansion for the mean.

THEOREM 5.4

Consider an $(M_t/G/\infty)^N/G$ network, where α is an $(n+1)$-times continuously differentiable function of compact support on $(-\infty, t]$. If $\mathrm{E}[(\bar{S}^r_k)^{n+2}] < \infty$ for all $r \in R$ and $k = 1, \ldots, |r|$, then

$$m_i(t,\epsilon) \equiv \mathrm{E}[Q^\epsilon_i(t)] = \sum_{j=0}^{n}(-\epsilon)^j \cdot \sum_{r\in R}\sum_{k\in r^{-1}(i)}\frac{D^j\alpha^r(t)}{(j+1)!}\mathrm{E}[(\bar{S}^r_k)^{j+1} - (\bar{S}^r_{k-1})^{j+1}]$$
$$+ O(\epsilon^{n+1}) \quad \text{as } \epsilon \downarrow 0,$$

where $D^j\alpha^r$ is the jth derivative of α^r. In particular,

$$\lim_{\epsilon\downarrow 0} m_i(t,\epsilon) = \sum_{r\in R}\alpha^r(t) \cdot \sum_{k\in r^{-1}(i)}\mathrm{E}[S^r_k],$$

where $S^r_k \equiv \bar{S}^r_k - \bar{S}^r_{k-1}$ equals the amount of time spent at site k along route r.

6. The $(M_t/GI/\infty)^N/G_t$ model

In this section we make the additional assumption that the service times of each customer at the successive queues are mutually independent with a common distribution at each queue. A service time at queue i is S_i with cdf G_i. We assume that $\mathbf{E}[S_i] < \infty$ for all i, which makes the associated stationary excess variable S_{ie} in (1.1) well defined. By theorem 5.1, we can reduce this model to an $(M_t/GI/\infty)^N/D$ network. *The upshot is that the* $(M_t/GI/\infty)^N/G_t$ *model reduces to a single* $M_t/GI/\infty$ *queue*, essentially as described after theorem 1.2 for the special case of M routing.

Moreover, for phase-type service-time distributions the $(M_t/GI/\infty)^N/G_t$ model can be reduced to the totally Markovian $(M_t/M/\infty)^N/M$ model, with all time-dependence in the arrival process, for which additional results are given in sections 7 and 8 below. First, we focus on individual queue sequences, as indicated above, to obtain the $(M_t/GI/\infty)^N/D$ model. Then we regard the different visits to the same queue along the route as different queues. This produces a tandem network without repeated queues, which is a special case of Markovian routing without time-dependence, i.e., a special case of the $(M_t/GI/\infty)^N/M$ model. Finally, we make separate queues for each of the exponential service phases in the phase-type distribution at each queue (as described on p. 174 of Whitt [71]).

SIMPLE FORMULAS

We gain from restriction to the $(M_t/GI/\infty)^N/G_t$ model by obtaining simple formulas. In particular, the model reduction implies that the network can be described by a minor modification of theorem 1.1.

THEOREM 6.1

For the $(M_t/GI/\infty)^N/G_t$ model, in addition to the results of theorem 5.3, we have

$$\lambda_i^-(t) = \mathbf{E}[\lambda_i^+(t - S_i)], \tag{6.1}$$

$$m_i(t) = \mathbf{E}\left[\int_{t-S_i}^{t} \lambda_i^+(s)\mathrm{d}s\right] = \mathbf{E}[\lambda_i^+(t - S_{ie})]\mathbf{E}[S_i], \tag{6.2}$$

and

$$\mathbf{Cov}[Q_i(t), Q_j(u)] = \sum_{r\in R}\sum_{k\in r^{-1}(i)}\sum_{l\in r^{-1}(j)\cap k^r} \mathbf{E}\left[\int_{(t-S_i)\vee(u-S_i-T_{kl}^r-S_j)}^{t\wedge(u-S_i-T_{kl}^r)} \lambda_k^r(s)\mathrm{d}s\right]^+, \tag{6.3}$$

where $T_{kl}^r = \bar{S}_{l-1}^r - \bar{S}_k^r$.

From theorem 6.1, it is apparent that most of the structural results for a single $M_t/GI/\infty$ queue in Eick et al. [21,22] carry over immediately to the queues within an $(M_t/GI/\infty)^N/G_t$ network. From (5.1) and (5.3), we see that the role of the arrival-rate function λ in theorem 1.1 is replaced by the route-r arrival-rate function α^r with $\alpha^r(s) = \alpha(s)P(\langle L_s\rangle) = r)$. In particular, $\alpha^r(s)$ *depends on* $\alpha(s)$ *and* $P(\langle L_s\rangle = r)$ *through their product*. For example, α^r might be increasing without α increasing.

As in [21], for the structural results, we take (5.1), (5.3) and (6.1)–(6.3) as the definitions, without worrying about the arrival rate having to be negative sometimes, as occurs when we postulate that α^r is a polynomial. To consider what happens when the arrival-rate function is a polynomial, let $m_i^r(t)$ be the mean number of route-r customers at queue i at time t, and let $\delta^r(t)$ be the external route-r departure rate at time t. The following is the generalization of theorem 3.1 of [21]; it expresses $m_i^r(t)$ and $\delta^r(t)$ as the *PSA*s $\alpha^r(t)\mathbf{E}[S_i]$ and $\alpha^r(t)$ modified by both time shifts and a space shift.

THEOREM 6.2

In the $(M_t/GI/\infty)^N/G_t$ model, if α^r is a polynomial of order k before time t, then so are m_i^r and δ^r. In particular, if $\alpha^r(s) = a_r + b_r s + c_r s^2$ for $s \leqslant t$, then

$$m_i^r(t) = \sum_{k\in r^{-1}(i)} \alpha^r(t - \mathbf{E}[\bar{S}_{k-1}^r] - \mathbf{E}[S_{ie}])\mathbf{E}[S_i] + c_r\left(\mathbf{Var}(S_{ie}) + \sum_{k\in r^{-1}(i)} \mathbf{Var}(\bar{S}_{k-1}^r)\right)\mathbf{E}[S_i]$$

and

$$\delta^r(t) = \alpha^r(t - \mathbf{E}[\bar{S}_{|r|}^r]) + c_r\mathbf{Var}(\bar{S}_{|r|}^r)\,.$$

Note that the conditions on α^r in theorem 6.2 reduce to conditions on the external-arrival-rate function α in the $(M_t/GI/\infty)^N/G$ model (when $P(\langle L_s\rangle = r)$ is independent of s). We can then add over routes r containing queue i to obtain corresponding results for queue i.

COROLLARY

In the $(M_t/GI/\infty)^N/G$ model with $\alpha^r(s) = p_r\alpha(s)$, if $\alpha(s) = a + bs + cs^2$ for $s \leqslant t$, then

$$m_i(t) = \sum_{r\in R} m_i^r(t) = \sum_{r\in R}\sum_{k\in r^{-1}(i)} p_r\alpha(t - \mathbf{E}[\bar{S}_k^r] - \mathbf{E}[S_{ie}])\mathbf{E}[S_i] + c\left(\mathbf{Var}(S_{ie}) + \sum_{r\in R}\sum_{k\in r^{-1}(i)} p_r\mathbf{Var}(\bar{S}_k^r)\right)\mathbf{E}[S_i]$$

and

$$\delta(t) = \sum_{r\in R} \delta^r(t) = \sum_{r\in R} p_r \alpha(t - \mathbf{E}[\bar{S}^r_{|r|}]) + c \sum_{r\in R} p_r \mathbf{Var}(\bar{S}^r_{|r|}) .$$

An analog of theorem 6.2 holds for sinusoidal arrival rate functions. The following is the natural generalization of theorem 4.1 of Eick et al. [22], which is provided in the same way. Let N^r_i be the number of times queue i appears on route r.

THEOREM 6.3

In the $(M_t/GI/\infty)^N/G_t$ model, if the route r arrival-rate function α^r is sinusoidal, i.e., if $\alpha^r(s) = \bar{\alpha}_r + \beta_r \sin(\gamma_r s)$ for positive constants $\bar{\alpha}_r, \beta_r$ and γ_r and all $s \leqslant t$, then

$$m^r_i(t) = N^r_i \bar{\alpha}_r \mathbf{E}[S_i] + \beta_r \left(\sin(\gamma_r t) \sum_{k\in r^{-1}(i)} \mathbf{E}[\cos(\gamma_r[\bar{S}^r_k + S_{ie}])] \right.$$
$$\left. - \cos(\gamma_r t) \sum_{k\in r^{-1}(i)} \mathbf{E}[\sin(\gamma_r[\bar{S}^r_k + S_{ie}])] \right) \mathbf{E}[S_i]$$

and

$$\delta^r(t) = \bar{\alpha}_r + \beta_r \left(\sin(\gamma_r t) \mathbf{E}[\cos(\gamma_r \bar{S}^r_{|r|})] - \cos(\gamma_r t) \mathbf{E}[\sin(\gamma_r \bar{S}^r_{|r|})] \right) .$$

Remark 6.1

As in [22], we can apply the sine and cosine addition formulas to express $m^r_i(t)$ and $\delta^r(t)$ in theorem 6.3 in terms of sums of products of terms of the form $\mathbf{E}[f(\gamma S^r_k)]$ and $\mathbf{E}[f(\gamma S_{ie})]$ where f is the sine or cosine. □

COROLLARY

In the $(M_t/GI/\infty)^N/G$ model with $\alpha^r(s) = p_r \alpha(s)$, if $\alpha(s) = \bar{\alpha} + \beta \sin(\gamma s)$ for all $s \leqslant t$, then

$$m_i(t) = A + \beta \sin(\gamma t) - C \cos(\gamma t) ,$$

where

$$A = \sum_{r\in R} N^r_i \bar{\alpha}^r \mathbf{E}[S_i] = \bar{\alpha} \left(\sum_{r\in R} p_r N^r_i \right) \mathbf{E}[S_i] ,$$

$$B = \beta \sum_{r\in R} \sum_{k\in r^{-1}(i)} p_r \mathbf{E}[\cos(\gamma[\bar{S}^r_k + S_{ie}])] \mathbf{E}[S_i] ,$$

$$C = \beta \sum_{r\in R} \sum_{k\in r^{-1}(i)} p_r \mathbf{E}[\sin(\gamma[\bar{S}^r_k + S_{ie}])] \mathbf{E}[S_i] ,$$

and

$$\delta(t) = \bar{\alpha} + D\sin(\gamma t) - E\cos(\gamma t),$$

where

$$D = \beta \sum_{r\in R} p_r \mathbf{E}[\cos(\gamma \bar{S}^r_{|r|})] \text{ and } E = \beta \sum_{r\in R} p_r \mathbf{E}[\sin(\gamma \bar{S}^r_{|r|})].$$

STOCHASTIC COMPARISONS

As mentioned in section 1, theorem 2.1(b) of [21] describes the effect of service-time variability, as expressed via convex stochastic order. This result takes a slightly different form here, involving increasing convex stochastic order. We say S_1 is less than or equal to S_2 in the *increasing convex stochastic order*, and write $S_1 \leqslant_{ic} S_2$, if $\mathbf{E}[f(S_1)] \leqslant \mathbf{E}[f(S_2)]$ for all increasing convex real-valued functions. As in [21], it is notable that the effect of the service-time variability depends on the shape of the arrival-rate function, which here is α^r.

THEOREM 6.4

Consider two $(M_t/GI/\infty)^N/G_t$ models indexed by subscripts 1 and 2 which differ only in their service times at queue i, one being S_{i1} and the other S_{i2}, and suppose that $S_{i1} \leqslant_c S_{i2}$.

(a) If queue i appears only once on route r and α^r is increasing (decreasing) before time t, then $m^r_{i1}(t) \geqslant (\leqslant) m^r_{i2}(t)$.

(b) If queue i appears more than once on route r and α^r is increasing and concave (decreasing and convex) before time t, then $m^r_{i1}(t) \geqslant (\leqslant) m^r_{i2}(t)$.

(c) If α^r is convex (concave) before time t, then $\delta^r_1(t) \leqslant (\geqslant) \delta^r_2(t)$ and at queue $k, k \neq i, m^r_{k1}(t) \leqslant (\geqslant) m^r_{k2}(t)$.

Proof

For part (a) apply the argument in theorem 2.1(b) of Eick et al. [21] with (6.2), i.e., $S_{i1} \leqslant_c S_{i2}$ implies that $S_{i1e} \leqslant_{st} S_{i2e}$ (ordinary stochastic order), so that $\mathbf{E}[f(S_{i1e})] \leqslant \mathbf{E}[f(S_{i2e})]$ for all increasing real-valued functions f. By (5.3) and theorem 2.4(a) of Eick et al. [21], λ^+_i is increasing (decreasing) before time t. For part (b), note that $S_{i1} \leqslant_c S_{i2}$ implies that $\bar{S}^r_{k1} \leqslant_c \bar{S}^r_{k2}$ for all k and $S_{i1e} \leqslant_{st} S_{i2e}$. Since $\leqslant_c$ and $\leqslant_{st}$ both imply $\leqslant_{ic}$, we have the conclusion because $\alpha^r(t-x)$ as a function of x has the required structure. Finally, for part (c) simply apply the convex stochastic order. □

Remark 6.2

The stochastic comparisons in theorems 2.1a, 2.2 and 2.4 of Eick et al. [21] extend easily to the network setting.

DEPENDENCE IN TANDEM QUEUES

We have described the dependence among the queues at different time points in theorems 2.2, 5.3 and 6.1. We now examine the consequences of this general result by considering the special case of a two-queue tandem $(M_t/GI/\infty)^N/D$ network. Then

$$\mathbf{Cov}[Q_1(t), Q_2(t+u)] = \mathbf{E}\left[\int_{(S_1-u)^+}^{S_1-(u-S_2)^+} \alpha(t-x)\mathrm{d}x\right]$$
$$= \int_0^\infty \alpha(t-x)\int_x^{u+x} P(S_2>u+x-y)P(S_1\in \mathrm{d}y)\mathrm{d}x\,. \quad (6.4)$$

Of interest is the *maximum-dependence time* $u_{\max} \equiv u_{\max}(\alpha, t, G_1, G_2)$, which is the value (or values) of u maximizing (6.4). In general, $u_{\max}$ depends on the external-arrival-rate function α, the time t and the two service-time cdf's G_1 and G_2.

EXAMPLE 6.1

If S_1 and S_2 are *both exponentially distributed* with means μ_1^{-1} and μ_2^{-1}, where $\mu_1 \neq \mu_2$, then

$$\mathbf{Cov}[Q_1(t), Q_2(t+u)] = \mu_1\int_0^\infty \alpha(t-x)\int_x^{u+x} \mathrm{e}^{-\mu_2(u+x-y)}\cdot \mathrm{e}^{-\mu_1 y}\mathrm{d}y\mathrm{d}x$$
$$= \mu_1\int_0^\infty \alpha(t-x)\mathrm{e}^{-\mu_2(u+x)}\int_x^{u+x} \mathrm{e}^{(\mu_2-\mu_1)y}\mathrm{d}y\mathrm{d}x$$
$$= \mu_1\int_0^\infty \alpha(t-x)[\mathrm{e}^{-\mu_2(u+x)}\mathrm{e}^{(\mu_2-\mu_1)(u+x)} - \mathrm{e}^{(\mu_2-\mu_1)x})/(\mu_2-\mu_1)]\mathrm{d}x$$
$$= \mu_1\int_0^\infty \alpha(t-x)[(\mathrm{e}^{-\mu_1(u+x)} - \mathrm{e}^{-(\mu_2 u+\mu_1 x)})/(\mu_2-\mu_1)]\mathrm{d}x$$
$$= \left(\frac{\mathrm{e}^{-\mu_1 u}-\mathrm{e}^{-\mu_2 u}}{\mu_2-\mu_1}\right)\int_0^\infty \alpha(t-x)\mu_1\mathrm{e}^{-\mu_1 x}\mathrm{d}x$$
$$= \left(\frac{\mathrm{e}^{-\mu_1 u}-\mathrm{e}^{-\mu_2 u}}{\mu_2-\mu_1}\right)\lambda_1^-(t)\,,$$

where λ_1^- is the aggregate-departure-rate function from queue 1. The value t^* maximizing $\lambda_1^-(t)$ is characterized by theorem 2.6 and corollary 2.8 of [21].

For all external-arrival-rate functions α and for all times t,

$$u_{\max} = \frac{\log\mu_1 - \log\mu_2}{\mu_1-\mu_2}\,.$$

By the mean-value theorem, we have

$$\min\{\mathbf{E}[S_1], \mathbf{E}[S_2]\} \leqslant u_{\max} \leqslant \max\{\mathbf{E}[S_1], \mathbf{E}[S_2]\}\,.$$

If $\mu_1 = \mu_2 = \mu$, then by the same reasoning

$$\mathbf{Cov}[Q_1(t), Q_2(t+u)] = u e^{-\mu u} \lambda_1^+(t)$$

and $u_{\max} = \mu^{-1} = \mathbf{E}[S_1] = \mathbf{E}[S_2]$. □

EXAMPLE 6.2

If S_1 and S_2 are *both deterministic* with means μ_1^{-1} and μ_2^{-1}, then

$$\begin{aligned}\mathbf{Cov}[Q_1(t), Q_2(t+u)] &= \int_0^\infty \alpha(t-x) P(S_1 + S_2 > u + x > S_1 > x)\mathrm{d}x \\ &= \int_0^\infty \alpha(t-x) \int_x^{u+x} P(S_2 > u + x - y) P(S_1 \in \mathrm{d}y)\mathrm{d}x \\ &= \int_{(S_1-u)^+}^{S_1} \alpha(t-x) P(S_2 > u + x - S_1)\mathrm{d}x \\ &= \left(\int_{(S_1-u)^+}^{S_1-(u-S_2)^+} \alpha(t-x)\mathrm{d}x \right)^+ ,\end{aligned}$$

and the maximal value for the covariance will be

$$\sup\left\{ \left(\int_{(S_1-u)^+}^{S_1-(u-S_2)^+} \alpha(t-x)\mathrm{d}x \right)^+ : S_1 \wedge S_2 \leqslant u \leqslant S_1 \vee S_2 \right\}. \tag{6.5}$$

For the case of $S_1 \leqslant S_2$ the supremum in (6.5) will be $\int_0^{S_1} \alpha(t-x)\mathrm{d}x$, which can be attained by any value of u within the given range; this is easy to see in fig. 4. The arrival rate function α is integrated from $t - S_1 + (u - S_2)^+$ to $t - (S_1 - u)^+$ for fixed u. Graphing these two functions with respect to u, the region of integration for α is represented as a vertical line segment whose endpoints touch these piecewise linear curves. Here it is clear that some u such that $S_1 \leqslant u \leqslant S_2$ gives the maximal integral for α, and in this case, any such u generates the line segment. On the other hand, if $S_1 \geqslant S_2$, then by fig. 5 the choice of u that maximizes the integral will depend on the choice of function α. □

In summary, we note that the value of u which maximizes the covariance is independent of the external-arrival rate function α and the time t when the service distributions are both exponentially distributed, as well as when the service times are both deterministic with $S_1 \leqslant S_2$, but not when the service times are both deterministic with $S_1 > S_2$.

7. The $(M_t/G_t/\infty)^N/M_t$ network: Markovian routing

In this section we consider Markovian routing. Here we consider the special case of the $(M_t/G_t/\infty)^N/G_t$ model in section 3 with M_t (time-dependent Marko-

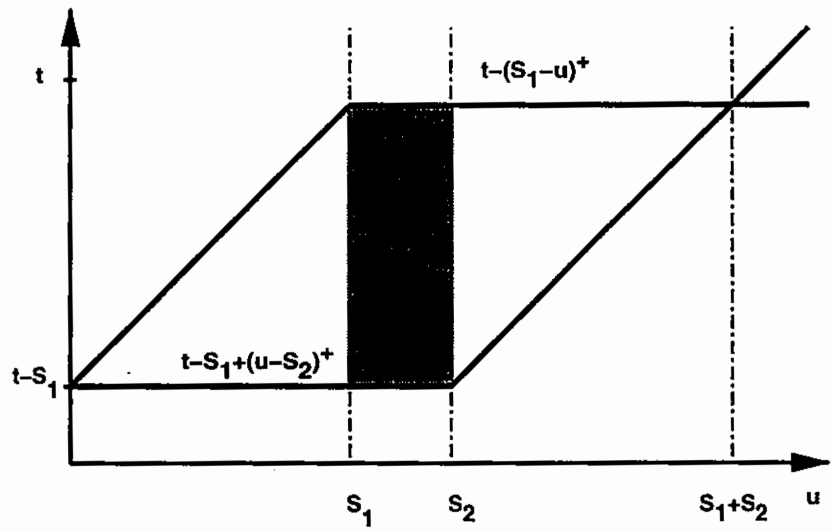

Fig. 4. Graph of integration intervals for tandem network with $S_1 < S_2$.

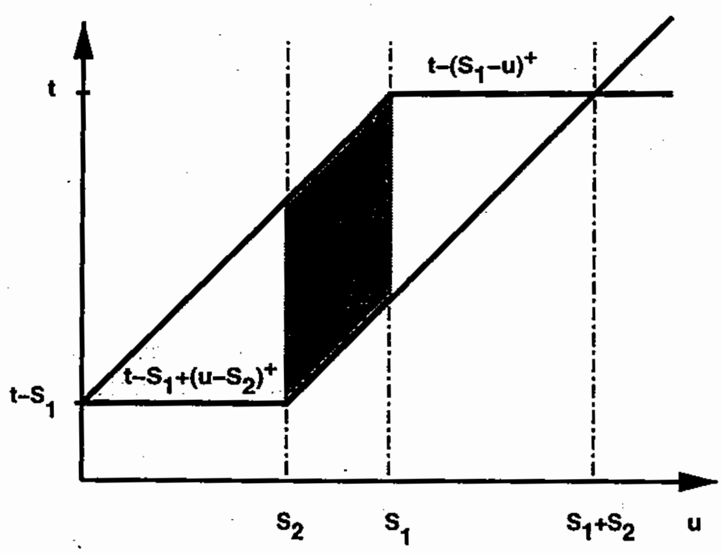

Fig. 5. Graph of integration intervals for tandem network with $S_1 > S_2$.

vian) routing. This means that the queue sequence is determined by initial-state probability $\pi_i(t)$ and the transition probabilities $p_{ij}(t)$, independently of the arrival process and service times. We first allow the service times to have general time dependence and stochastic dependence, but later we restrict attention to the GI special case.

GENERALIZED INPUT EQUATIONS WITHOUT RATES

As noted in section 3, the processes Λ_i^+ and Λ_i^- do not necessarily have rate functions, but since $\mathbf{E}[\Lambda_i^+(t)]$ and $\mathbf{E}[\Lambda_i^-(t)]$ are well defined by theorem 3.3 and nondecreasing in t, we can let λ_i^+ and λ_i^- be nonnegative measures such that

$$\mathbf{E}[\Lambda_i^+(t)] = \int_{-\infty}^{t} \lambda_i^+(\mathrm{d}s) \text{ and } \mathbf{E}[\Lambda_i^-(t)] = \int_{-\infty}^{t} \lambda_i^-(\mathrm{d}s)\,, \tag{7.1}$$

where these expectations may be infinite. In order for these integrals to be finite, we assume that (3.15) holds for all i, where $\mathbf{E}[V_i^+(s,\infty)]$ depends on the initial state probability $\pi_i(s)$ and the transition probability $p_{ij}(t)$ for $t \geqslant s$ in a rather complicated way.

We now determine generalized input equations for this model. Note that we do not obtain a simple expression for the departure measure λ_i^- in terms of the arrival measure λ_i^+ at this level of generality.

Our starting point is the obvious sample-path conservation relation

$$\Lambda_i^+(t) = A_i(t) + \sum_{j=1}^{N} \Lambda_{ij}^-(t)\,, \tag{7.2}$$

where $\Lambda_{ij}^-(t)$ counts the number of departures from queue i by time t that go next to queue j. For each $t \in \mathbb{R}$, and i, $1 \leqslant i \leqslant N$, let $\{\Pi_{ij}(t) : 1 \leqslant j \leqslant N\}$ be a collection of disjoint subintervals of $[0,1]$ for which $\Pi_{ij}(t)$ has Lebesgue measure $p_{ij}(t)$.

THEOREM 7.1

In the $(M_t/G_t/\infty)^N/M_t$ model with assumption (3.15) for $k=1$, the process Λ_{ij}^- can be represented as

$$\Lambda_{ij}^-(t) = \int_{-\infty}^{t} \left(\int_{s}^{t} V_i^-(s,\mathrm{d}u) \mathbf{1}_{\{U^i_{A(s),V_i^-(s,u)} \in \Pi_{ij}(u)\}} \right) \mathrm{d}A(s)\,, \tag{7.3}$$

where $\{U^i_{k,m} : 1 \leqslant i \leqslant N, k \geqslant 1, m \geqslant 1\}$ is a countable collection of i.i.d. uniform random variables on $[0,1]$ independent of A, so that

$$\mathbf{E}[\Lambda_{ij}^-(t)] = \int_{-\infty}^{t} \lambda_i^-(\mathrm{d}u) p_{ij}(u) \tag{7.4}$$

and we have the generalized input equations

$$\mathbf{E}[\Lambda_i^+(t)] \equiv \int_{-\infty}^t \lambda_i^+(\mathrm{d}s) = \int_{-\infty}^t \alpha_i(s)\mathrm{d}s + \sum_{j=1}^N \int_{-\infty}^t \lambda_j^-(\mathrm{d}s)p_{ji}(s)\,. \qquad (7.5)$$

Proof

The generalized input equations (7.5) follow easily from (7.2), (7.4) and theorems 3.2 and 3.4. The inner integral in (7.3) expresses the fact that the conditional probability of the kth transition from queue i being to j given that the kth transition takes place at time u has probability $p_{ij}(u)$ independent of the history up to that time. To justify this inner integral, let $\hat{\Lambda}_k^-$ be the kth aggregate departure epoch from queue i for an arrival at time s, with $\hat{\Lambda}_k^- = +\infty$ if $V_i(s,\infty) < k$. Then

$$\int_s^t V_i^-(s,\mathrm{d}u)\mathbf{1}_{\{U^i_{A(s),V_i^-(s,u)} \in \Pi_{ij}(u)\}} = \sum_{k=1}^\infty \mathbf{1}_{\{\hat{\Lambda}_k^- \leqslant t\}}\mathbf{1}_{\{U^i_{A(s),k} \in \Pi_{ij}(\hat{\Lambda}_k^-)\}}\,, \qquad (7.6)$$

where $\hat{\Lambda}_{k+1}^-$ in general depends on $\hat{\Lambda}_1,\ldots,\hat{\Lambda}_k$ and $U^i_{A(s),1},\ldots,U^i_{A(s),k}$. We then obtain (7.4) by taking expectations in (7.3). By lemma 2.2 and (3.15),

$$\mathbf{E}[\Lambda_{ij}^-(t)] = \int_{-\infty}^t \mathbf{E}\left[\int_s^t V_i^-(s,\mathrm{d}u)\mathbf{1}_{\{U^i_{A(s),V_i^-(s,u)} \in \Pi_{ij}(u)\}}\right]\alpha(s)\mathrm{d}s\,. \qquad (7.7)$$

By (7.7), (7.6) and Tonelli, plus conditioning and unconditioning on $\hat{\Lambda}_k^-$,

$$\begin{aligned}
\mathbf{E}[\Lambda_{ij}^-(t)] &= \int_{-\infty}^t \sum_{k=1}^\infty \mathbf{E}[\mathbf{1}_{\{\hat{\Lambda}_k^- \leqslant t\}}p_{ij}(\hat{\Lambda}_k^-)]\alpha(s)\mathrm{d}s \\
&= \int_{-\infty}^t \int_s^t \mathbf{E}[V_i^-(s,\mathrm{d}u)p_{ij}(u)]\alpha(s)\mathrm{d}s \\
&= \int_{-\infty}^t \left(\int_{-\infty}^u \mathbf{E}[V_i^-(s,\mathrm{d}u)]\alpha(s)\mathrm{d}s\right)p_{ij}(u) \\
&= \int_{-\infty}^t \lambda_i^-(\mathrm{d}u)p_{ij}(u)\,.
\end{aligned}$$

□

INPUT EQUATIONS FOR THE $(M_t/GI/\infty)^N/M_t$ MODEL

We now assume that the service times only depend on the queue index. This model is a special case of the $(M_t/GI/\infty)^N/G_t$ model studied in section 6, which can be reduced to the $(M_t/GI/\infty)^N/D$ model. Hence, the results of sections 2–6 all apply to it.

Paralleling theorem 3.4, let $\hat{\Lambda}_i^n$ be the aggregate arrival process at queue i associated with customers making their nth transition (visit to a queue); i.e., $\hat{\Lambda}_i^1$ is the external arrival process to queue i, A_i. Let $\hat{\lambda}_i^n$ be the arrival rate at queue i associated with customers making their nth transition.

THEOREM 7.2

In the $(M_t/GI/\infty)^N/M_t$ model with the assumptions above, the aggregate arrival rate functions are well defined and satisfy the input equations

$$\lambda_i^+(t) = \alpha_i(t) + \sum_{j=1}^{N} \mathrm{E}[\lambda_j^+(t - S_j)]p_{ji}(t), \quad \text{for } 1 \leqslant i \leqslant N. \tag{7.8}$$

The aggregate arrival-rate function λ_i^+ can be defined as

$$\lambda_i^+(t) = \sum_{n=1}^{\infty} \hat{\lambda}_i^n(t) < \infty \text{ a.e.}, \tag{7.9}$$

where $\{\hat{\lambda}_i^n\}$ can be defined recursively as $\hat{\lambda}_i^1(t) = \alpha_i(t)$, and, for $n \geqslant 1$,

$$\hat{\lambda}_i^{n+1}(t) = \sum_{j=1}^{N} \mathrm{E}[\hat{\lambda}_j^n(t - S_j)]p_{ji}(t). \tag{7.10}$$

The aggregate arrival-rate functions λ_i^+ can be characterized as the minimal nonnegative solution to (7.8). Moreover, the mean number of busy servers at queue i is given by (1.5).

Proof

By theorem 3.2, A_i is a Poisson process with intensity $\alpha_i(t)$. By (3.15) and the corollary to theorem 4.3, the intensities $\lambda_i^+(t)$ and $\lambda_i^-(t)$ are well defined with

$$\int_{-\infty}^{\infty} \lambda_i^-(t)\mathrm{d}t \leqslant \int_{-\infty}^{\infty} \lambda_i^+(t)\mathrm{d}t = \int_{-\infty}^{\infty} \mathrm{E}[V_i^+(s, \infty)]\alpha(s)\mathrm{d}s < \infty.$$

Formula (7.9) holds because $\Lambda_i^+ = \sum_{n=1}^{\infty} \hat{\Lambda}_i^n$. We obtain (7.10) by applying theorem 1.1 and Poisson thinning of the departure processes recursively. We then obtain (7.8) by appropriate summing (7.10), e.g.,

$$\lambda_j^-(t) = \sum_{n=1}^{\infty} \mathrm{E}[\hat{\lambda}_j^n(t - S_j)] = \mathrm{E}[\lambda_j^+(t - S_j)].$$

We now show that λ_i^+ is the minimal nonnegative solution to (7.8). For this purpose, we introduce the operator Φ mapping N-tuples of nonnegative measurable functions $\mathbf{x} \equiv (x_1, \ldots, x_N)$ into themselves by

$$\Phi_i(\mathbf{x}) = \alpha(t)\pi_i(t) + \sum_{j=1}^{N} \mathrm{E}[x_j(t - S_j)]p_{ji}(t), \quad 1 \leqslant i \leqslant N, \tag{7.11}$$

i.e., by substituting $\mathbf{x}$ for $\boldsymbol{\lambda}$ in the right side of (7.8). Note that Φ is monotone with the usual partial order: $\mathbf{x} \leqslant \mathbf{y}$ if $x_i(t) \leqslant y_i(t)$ for all i and t. Moreover, $\Phi^n(\boldsymbol{\theta})$ converges monotonically to the solution specified by (7.9)–(7.10) when $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_N)$ with $\theta_i(t) = 0, t \in \mathbb{R}$. Suppose that $\boldsymbol{\lambda}^*$ is a second nonnegative solution. Since $\boldsymbol{\theta} \leqslant \boldsymbol{\lambda}^*, \Phi^n(\boldsymbol{\theta}) \leqslant \Phi^n(\boldsymbol{\lambda}^*) = \boldsymbol{\lambda}^*$ for all n. Since $\Phi^n(\boldsymbol{\theta}) \to \boldsymbol{\lambda}$ as $n \to \infty, \boldsymbol{\lambda} \leqslant \boldsymbol{\lambda}^*$ too. Hence, $\boldsymbol{\lambda}$ is indeed the minimum nonnegative solution. Finally, we establish (1.5) by using the decomposition in (7.9)–(7.10) and theorem 1.1. In particular, let $\hat{m}_i^n(t)$ be the mean number of customers at queue i at time t who have made n transitions. Then

$$m_i(t) = \sum_{n=1}^{\infty} \hat{m}_i^n(t)$$

and theorem 1.1 applies to $\hat{m}_i^n(t)$ because $\hat{\Lambda}_i^n$ is a Poisson process. □

We now show that the input equations (1.6) as well as (7.8) do *not* necessarily have a unique solution.

EXAMPLE 7.1

Here we show that the input equations (1.6) need not have a unique solution. In this case of homogeneous Markov routing, $\boldsymbol{\lambda} + \boldsymbol{\gamma}$ satisfies (1.6) when $\boldsymbol{\lambda}$ does if and only if

$$\gamma_i(t) = \sum_{j=1}^{N} \mathbf{E}[\gamma_j(t - S_j)]p_{ji}, \quad 1 \leqslant i \leqslant N. \tag{7.12}$$

To obtain a simple example, suppose that $N = 1, P(S_j = 1) = 1$ and $P_{11} = 1/2$. We now seek solutions to (7.12) of the form $\gamma_i(t) = \gamma_i(\lfloor t \rfloor)$ for all t, where $\lfloor t \rfloor$ is the greatest integer less than or equal to t. Hence, (7.12) reduces to the equation $\gamma_1(k) = \gamma_1(k-1)/2$ for sequences $\{\gamma_1(k) : k \in \mathbb{Z}\}$, where $\mathbb{Z}$ is the set of integers, which has the positive finite solution $\gamma_1(k) = 2^{-k}$ for all k. □

UNIQUENESS AMONG INTEGRABLE FUNCTIONS

We now return to the $(M_t/GI/\infty)^N/M$ model of theorem 1.2. We have Markovian routing with all time-dependence in the external arrival process. The routing is determined by the initial-state probabilities π_i and the transition probabilities p_{ij}. Now we assume that

$$\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty \quad \text{for all } i \text{ and } j, \tag{7.13}$$

which is a necessary and sufficient condition for (3.15) for homogeneous Markov routing.

We first show that the input equations (1.6) have a unique solution among integrable functions in this case. (Note that the second solution in example 7.1 is not integrable.) Then we show how to solve the input equations in special cases.

As in the proof of theorem 7.2, we use the right side of the input equations to define an operator. In particular, here let $\mathbf{F}$ be the space of N-tuples $\mathbf{x} \equiv (x_1, \ldots, x_N)$ of integrable real-valued functions on $\mathbb{R}$ with the L_1 norm

$$\|\mathbf{x}\| = \sum_{i=1}^{N} \int_{-\infty}^{\infty} |x_i(t)| \mathrm{d}t . \tag{7.14}$$

Let $\Phi : \mathbf{F} \to \mathbf{F}$ be the operator defined by

$$\Phi_i(\mathbf{x})(t) = \alpha_i(t) + \sum_{j=1}^{N} \mathbf{E}[x_j(t - S_j)] p_{ji} . \tag{7.15}$$

As before, the operator Φ is monotone with the usual partial order, i.e., $\mathbf{x} \leqslant \mathbf{y}$ if $x_i(t) \leqslant y_i(t)$ for all i and t. Since α is integrable and $\|\Phi(\mathbf{x})\| \leqslant \|\boldsymbol{\alpha}\| + \|\mathbf{x}\|$, Φ indeed maps $\mathbf{F}$ into $\mathbf{F}$. Note that the input equations (1.6) are simply the *fixed-point equation*

$$\mathbf{x} = \Phi(\mathbf{x}) . \tag{7.16}$$

We shall show that the fixed point equation (7.16) has a unique solution by showing that Φ is an N-stage contraction, and applying the Banach–Picard fixed point theorem. Let $\Phi^{(n)}$ be the n-fold iterate of Φ.

THEOREM 7.3

Consider the $(M_t/GI/\infty)^N/M$ model with assumption (7.13). For all $\mathbf{x}, \mathbf{y} \in \mathbf{F}$ and $n \geqslant 1$,

$$\|\Phi^{(n)}(\mathbf{x}) - \Phi^{(n)}(\mathbf{y})\| \leqslant \|\mathbf{x} - \mathbf{y}\| \tag{7.17}$$

and there exists γ, $0 < \gamma < 1$, such that

$$\|\Phi^{(N)}(\mathbf{x}) - \Phi^{(N)}(\mathbf{y})\| \leqslant \gamma \|\mathbf{x} - \mathbf{y}\| , \tag{7.18}$$

so that the fixed-point equation (7.16) (equivalently, the input equations (1.6)) has a unique solution $\boldsymbol{\lambda}$ in $\mathbf{F}$. Moreover,

$$\|\boldsymbol{\lambda}\| \leqslant \frac{N}{1-\gamma} \|\boldsymbol{\alpha}\| \tag{7.19}$$

and

$$\|\Phi^{(nN)}(\mathbf{x}) - \boldsymbol{\lambda}\| \leqslant \gamma^n \|\mathbf{x} - \boldsymbol{\lambda}\| . \tag{7.20}$$

Proof

We have noted that Φ maps $\mathbf{F}$ into itself. Note that

$$\int_{-\infty}^{\infty} |\Phi_i(\mathbf{x})(t) - \Phi_i(\mathbf{y})(t)| \mathrm{d}t \leqslant \sum_{j=1}^{N} \left(\int_{-\infty}^{\infty} |x_j(t) - y_j(t)| \mathrm{d}t \right) p_{ji} .$$

By induction on n,

$$\|\Phi^{(n)}(\mathbf{x}) - \Phi^{(n)}(\mathbf{y})\| \leqslant \sum_{i=1}^{N}\sum_{j=1}^{N}\left(\int_{-\infty}^{\infty} |x_j(t) - y_j(t)|\mathrm{d}t\right)p_{ji}^{(n)} \leqslant \gamma_n \|\mathbf{x} - \mathbf{y}\|,$$

where

$$\gamma_n = \max_{1 \leqslant j \leqslant N}\left\{\sum_{i=1}^{N} p_{ji}^{(n)}\right\}.$$

Since $\gamma_n \leqslant 1$ for all n and there exists γ such that $\gamma_N \leqslant \gamma < 1$ by (7.13), we have established (7.17) and (7.18). For (7.19), note that

$$\|\Phi^{(n)}(\mathbf{\theta})\| \leqslant \|\boldsymbol{\alpha}\| + \|\boldsymbol{\alpha}\mathbf{P}\| + \ldots + \|\boldsymbol{\alpha}\mathbf{P}^n\|,$$

so that

$$\|\boldsymbol{\lambda}\| = \lim_{n\to\infty} \|\Phi^{(nN)}(\mathbf{\theta})\| \leqslant \frac{(\|\boldsymbol{\alpha}\| + \ldots + \|\boldsymbol{\alpha}\mathbf{P}^{n-1}\|)}{1-\gamma} \leqslant \frac{N\|\boldsymbol{\alpha}\|}{1-\gamma}.$$

For (7.20), apply (7.18) with $\boldsymbol{\lambda} = \Phi^{(nN)}(\boldsymbol{\lambda})$. □

Remark 7.1

Theorem 7.3 above also applies to models with time-dependent routing if the time-dependent probabilities are appropriately dominated by fixed transition probabilities, i.e., if

$$p_{ij}(t) \leqslant p_{ij} \tag{7.21}$$

for all i, j and t, where $\{p_{ij}\}$ satisfies (7.13).

Remark 7.2

Theorem 1.2 is proved by combining theorems 2.1, 3.3, 3.5, 4.3, 7.2 and 7.3.

Remark 7.3

In theorem 7.3 we could work with functions that are only required to be integrable over $(-\infty, t]$. The bound in remark 7.1 also only need hold over $(-\infty, t]$.

SOLVING THE INPUT EQUATIONS IN SPECIAL CASES

We now indicate how to solve the input equations (1.6) in special cases. Our first result supplements theorem 6.2. As before, we ignore difficulties concerning α being sometimes nonnegative; we take (1.6) as the definition.

THEOREM 7.4

In the $(M_t/GI/\infty)^N/M$ model with assumption (7.13), if $\alpha(s) = a_0 + a_1 s + \ldots + a_m s^m$ for $s \leqslant t$, then there is a solution of the input equations (1.6) of the same form, i.e., $\lambda_i^+(t) = a_{i0} + a_{i1}t + \ldots + a_{im}t^m$, where the coefficients satisfy the system of linear equations

$$a_{ik} = a_k\pi_i + \sum_{j=1}^{N}\sum_{l=k}^{m} a_{jl}(-1)^{l-k}\binom{l}{k}\mathbf{E}[S_j^{l-k}]p_{ji}\,, \tag{7.22}$$

for $1 \leqslant i \leqslant N$ and $1 \leqslant k \leqslant m$. The equations (7.22) have a unique solution, which can be found by solving m systems of N linear equations recursively, going backwards in k.

Proof

Substituting the polynomial form into (1.6), we get

$$\sum_{k=o}^{m} a_{ik}t^k = \sum_{k=0}^{m} a_k\pi_i t^k + \sum_{j=1}^{N}\sum_{k=0}^{m} a_{jk}\mathbf{E}[(t-S_j)^k]p_{ji}\,. \tag{7.23}$$

Collecting coefficients of t^k in (7.23), we obtain (7.22). Considering these equations recursively backwards in k, starting with $k = m$, we see that they have a unique solution. In particular, the kth equation is of the form

$$a_{ik} = b_{ik} + \sum_{j=1}^{N} a_{jk}p_{ji}\,, \tag{7.24}$$

where

$$b_{ik} = a_k\pi_i + \sum_{l=k+1}^{m}\sum_{j=1}^{N} a_{jl}(-1)^{l-k}\binom{l}{k}\mathbf{E}[S_j^{l-k}]p_{ji}\,, \tag{7.25}$$

which is known when we consider the kth equation. In matrix notation, (7.24) is

$$\mathbf{A}_k = \mathbf{B}_k + \mathbf{A}_k\mathbf{P}\,, \tag{7.26}$$

so that

$$\mathbf{A}_k = \mathbf{B}_k(\mathbf{I}-\mathbf{P})^{-1} \tag{7.27}$$

with $(\mathbf{I}-\mathbf{P})^{-1}$ existing by (7.13). □

Remark 7.4

In matrix notation, we can express $\mathbf{B}_k$ in (7.26) as

$$\mathbf{B}_k = \mathbf{C}_k + \sum_{l=k+1}^{m}(-1)^{l-k}\binom{l}{k}(\mathbf{A}_l \circ \mathbf{E}[\mathbf{S}^{l-k}])\mathbf{P}\,, \tag{7.28}$$

where $(\mathbf{A}_l \circ \mathbf{E}[\mathbf{S}^{l-k}])_j = a_{jl}\mathbf{E}[\mathbf{S}_j^{l-k}]$. □

We now consider the case of sinusoidal external arrival rate functions, and thus supplement theorem 6.3. Here let $\mathrm{i} = \sqrt{-1}$.

THEOREM 7.5

In the $(M_t/G/\infty)^N/M$ model with assumption (7.13), suppose that we have a Fourier-series expansion for the external arrival-rate function α, i.e.,

$$\alpha(s) = \sum_{n=-\infty}^{\infty} \alpha_n \mathrm{e}^{2\pi n \mathrm{i} s} \quad \text{for } s \leqslant t. \tag{7.29}$$

Then there is a solution to (1.6) of the same form, i.e.,

$$\lambda_j^+(t) = \sum_{n=-\infty}^{\infty} \lambda_{jn} \mathrm{e}^{2\pi n \mathrm{i} t}, \quad 1 \leqslant j \leqslant N, \tag{7.30}$$

where the coefficients λ_{jn} in (7.30) are obtained for each n as the unique solution of the linear equations

$$\lambda_{jn} = \alpha_n \pi_j + \sum_{k=1}^{N} \lambda_{kn} \mathrm{E}[\mathrm{e}^{-2\pi n \mathrm{i} S_k}] p_{kj}, \quad 1 \leqslant j \leqslant N. \tag{7.31}$$

Proof

Substitute (7.30) into (1.6) to obtain

$$\sum_{n=-\infty}^{\infty} \lambda_{jn} \mathrm{e}^{2\pi \mathrm{i} n t} = \sum_{n=-\infty}^{\infty} \alpha_n \mathrm{e}^{2\pi n \mathrm{i} t} \pi_j + \sum_{k=1}^{N} \sum_{n=-\infty}^{\infty} \lambda_{kn} \mathrm{E}[\mathrm{e}^{2\pi \mathrm{i} n (t - S_k)}] p_{kj}. \tag{7.32}$$

Now collect coefficients of $\mathrm{e}^{2\pi \mathrm{i} n t}$ in (7.32) to obtain (7.31). By (7.13), (7.31) has a unique solution $(\lambda_{1n}, \ldots, \lambda_{Nn})$ for each n, just as in (7.27). □

8. The totally Markov $(M_t/M_t/\infty)^N/M_t$ model

In this section we consider the Markovian $(M_t/M_t/\infty)^N/M_t$ model in which service completions at queue i at time t take place at rate $\mu_i(t)$ for each busy server. This is the case in which the location processes $\{L_s(t) : t \in \mathbb{R}\}$ are Markov. Here the vector queue-length process $(Q_1, \ldots, Q_N) \equiv \{(Q_1(t), \ldots, Q_N(t)) : t \in \mathbb{R}\}$ is a nonstationary continuous-time Markov chain (CTMC) determined by the functions $\alpha(t), \pi_i(t), p_{ij}(t)$ and $\mu_i(t)$. There is substantial related literature for this special case under linear Markov population processes and linear Markov compartmental models; see Kingman [40], Faddy [25], section 5 of Purdue [59], Kurtz [41] and Ethier and Kurtz [24].

In this section, we first show how to construct this CTMC and then we show that the mean and covariance functions satisfy linear ordinary differential equations (ODEs). The linear ODE for the mean function is standard, but the linear ODE for the covariance function seems to be new.

CONSTRUCTION VIA UNIFORMIZATION

We indicate one way to construct the nonhomogeneous CTMC $(Q_1, \ldots, Q_N)$. As before, we assume that $\bar{\alpha} \equiv \int_{-\infty}^{\infty} \alpha(s)\mathrm{d}s < \infty$. Now we assume in addition that

$$\bar{\mu} \equiv \sup\{\mu_i(t) : t \in \mathbb{R}, 1 \leqslant i \leqslant N\} < \infty. \tag{8.1}$$

We start by generating a Poisson number of arrivals with mean $\bar{\alpha}$. Given this number, we let the times of these arrivals be i.i.d. with density $\alpha(s)/\bar{\alpha}$. This produces a realization of the external arrival process A. If there are k external arrivals at times $s_1, \ldots, s_k$, then we assign them to their initial queues independently, with the probability of queue i being selected at time s_j being $\pi_i(s_j)$.

Given that there is a total of k external arrivals, generate all other events using a homogeneous Poisson process $M \equiv \{M(t) : t \in \mathbb{R}\}$ with rate $k\bar{\mu}$ for $\bar{\mu}$ in (8.1). Let the k external arrivals be labeled. We start at the time of the first external arrival. At each point from M after this first external arrival, we generate our transition or no transition. We consider the successive external arrivals in turn. If the jth external arrival is in the system at queue i, then that customer has a transition with probability $\mu_i(t)/k\bar{\mu}$. We let the transitions for different external arrivals be mutually exclusive events. Thus, the remaining probability is the probability of no transition at that point of M. Given that a customer at queue i has a transition at time t, that transition is treated as an independent event, leading to queue j with probability $p_{ij}(t)$ and outside the network with probability $1 - \sum_{j=1}^{N} p_{ij}(t)$. Without making any assumptions on $p_{ij}(t)$, the number of customers in the network at any time is at most k, given that k external arrivals have been generated. As a consequence, we have

$$\sum_{i=1}^{N} m_i(t) \leqslant \bar{\alpha} \quad \text{for all } t. \tag{8.2}$$

Given that a customer arrives at queue i at time t, that customer has a service time S_{it} which is independent of the history of the network up to time t. Moreover, the distribution of S_{it} is a *time-dependent exponential*, i.e.,

$$P(S_{it} > x) = \mathrm{e}^{-\int_t^x \mu_i(u)\mathrm{d}u}, \quad x > 0. \tag{8.3}$$

Note that we have not made assumptions implying that $P(S_{it} < \infty) = 1$. From (8.3), it is apparent that $P(S_{it} < \infty) = 1$ for all i and t if and only if

$$\int_t^{\infty} \mu_i(u)\mathrm{d}u = \infty \tag{8.4}$$

for some t and all i.

Assumption (8.1) implies that

$$P(S_{it} > x) \geqslant e^{-\bar{\mu}x}, \quad x > 0, \tag{8.5}$$

i.e., S_{it} is stochastically greater than or equal to an exponential random variable with mean $\bar{\mu}^{-1}$.

We now show that the aggregate arrival-rate and departure-rate functions λ_i^+ and λ_i^- are well defined. For this purpose, let $f(n_1, \ldots, n_k; t_1, \ldots, t_{k+1})$ be the probability density of an arrival at time t_1, following route $(n_1, \ldots, n_k)$, being at queue n_i in the interval $[t_i, t_{i+1})$ and leaving the network at time t_{k+1}. From the construction above, we can characterize this joint density.

LEMMA 8.1

The density $f(n_1, \ldots, n_k; t_1, \ldots, t_{k+1})$ is well defined, being

$$f(n_1, \ldots, n_k; t_1, \ldots, t_{k+1}) = \alpha(t_1)\pi_{n_1}(t_1) \exp\left(-\int_{t_1}^{t_2} \mu_{n_1}(u)\mathrm{d}u\right)\mu_{n_1}(t_2)$$
$$\times p_{n_1,n_2}(t_2) \ldots \exp\left(-\int_{t_k}^{t_{k+1}} \mu_{n_k}(u)\mathrm{d}u\right)\mu_{n_k}(t_{k+1})\left(1 - \sum_{j=1}^{N} p_{n_k,j}(t_{k+1})\right). \tag{8.6}$$

THEOREM 8.1

The aggregate arrival-rate and departure-rate functions λ_i^+ and λ_i^- are well defined with

$$\lambda_i^+(t) = \alpha(t)\pi_i(t) + \sum_{j=1}^{N} \lambda_j^-(t)p_{ji}(t) \tag{8.7}$$

and

$$\lambda_i^-(t) = \mu_i(t) \int_{-\infty}^{t} \lambda_i^+(s)e^{-\int_s^t \mu_i(u)\mathrm{d}u}\mathrm{d}s. \tag{8.8}$$

Moreover, λ_i^+ and λ_i^- are integrable over finite intervals.

Proof

To show that λ_i^+ is well defined, it suffices to sum (8.6) over all routes $(n_1, \ldots, n_k)$ such that $n_l = i$ for some l and sum over all visits to queue i. If $n_l = i$, then we integrate over the set of $(t_1, \ldots, t_{k+1})$ such that $-\infty < t_1 < \ldots < t_{l-1} < t_l = t < t_{l+1} < \ldots < t_{k+1} < \infty$. This yields $\lambda_i^+(t)$. Similarly, we can construct μ_i and show it is well defined and equal to (8.8). Formula (8.7) is a consequence of theorem 7.1 with the extra properties here. Finally, since $\sum_{i=1}^{N} \lambda_i^-(t) \leqslant \bar{\alpha}\bar{\mu}$ and $\lambda_i^+(t) \leqslant \alpha(t) + \sum_{i=1}^{N} \lambda_j^-(t)$, λ_i^+ and λ_i^- are integrable over finite intervals. □

LINEAR ODEs

We now show that the mean and covariance functions satisfy linear ODEs. For this purpose, we first show that the mean function is differentiable.

LEMMA 8.2

For each i, $m_i(t)$ is differentiable in t almost everywhere and

$$\frac{\mathrm{d}m_i(t)}{\mathrm{d}t} = \lambda_i^+(t) - \lambda_i^-(t) \text{ a.e.} \tag{8.9}$$

Proof

Note that, for any $\epsilon > 0$,

$$Q_i(t+\epsilon) - Q_i(t) = A_i^+(t+\epsilon) - A_i^+(t) - A_i^-(t+\epsilon) + A_i^-(t)\,,$$

so that

$$\begin{aligned} m_i(t+\epsilon) - m_i(t) &\equiv \mathbf{E}[Q_i(t+\epsilon)] - \mathbf{E}[Q_i(t)] \\ &= \mathbf{E}[A_i^+(t+\epsilon) - A_i^+(t)] - \mathbf{E}[A_i^-(t+\epsilon) - A_i^-(t)] \\ &= \int_t^{t+\epsilon} \lambda_i^+(u)\mathrm{d}u - \int_t^{t+\epsilon} \lambda_i^-(u)\mathrm{d}u \end{aligned}$$

by theorem 8.1, so that indeed (8.9) holds. □

THEOREM 8.2

The mean functions m_i are the unique bounded solution of the linear ODE

$$\frac{\mathrm{d}m_i(t)}{\mathrm{d}t} = \alpha_i(t) + \sum_{j=1}^{N} m_j(t)\mu_j(t)p_{ji}(t) - m_i(t)\mu_i(t), \quad 1 \leqslant i \leqslant N\,, \tag{8.11}$$

which necessarily has the property that $m_i(t) \to 0$ as $t \to -\infty$.

Proof

To establish existence, note that

$$\lambda_i^-(t) = m_i(t)\mu_i(t)\,. \tag{8.12}$$

Then combine (8.7), (8.9) and (8.12). Uniqueness is immediate by direct construction if $\alpha(t) = 0$ for $t < t_0$ for some t_0. More generally, since $\int_{-\infty}^{t} \alpha(u)\mathrm{d}u \to 0$ as $t \to -\infty$, we must have $m_i(t) \to 0$ as $t \to -\infty$; i.e., the solution is the limit as $n \to \infty$ of the sequence of unique solutions associated with the sequence of external arrival-rate functions $\{\alpha_n : n \geqslant 1\}$, where $\alpha^{(n)}(t) = \alpha(t)$ for $t \geqslant -n$ and $\alpha^{(n)}(t) = 0$ for $t < -n$. As n increases, $\alpha^{(n)}$ increases and $m^{(n)}(t)$ increases. Since $m^{(n)}(t) \leqslant \bar{\alpha}$,

$m^{(n)}(t)$ converges to a proper limit as $n \to \infty$ for all i and t. This establishes the uniqueness. □

Let $\pi_{ij}(s,t)$ be the *nonstationary transition probabilities for each customer*, i.e., the probability of being at queue j at time t given that the customer was at queue i at time s, where $s<t$. The following is proved like lemma 8.2.

LEMMA 8.3

Under (8.1), $\{\pi_{ij}(s,t) : t \geqslant s\}$ is the unique solution to

$$\frac{\mathrm{d}\pi_{ij}(s,t)}{\mathrm{d}t} = \sum_{k=1}^{N} \pi_{ik}(s,t)\mu_k(t)p_{kj}(t) - \pi_{ij}(s,t)\mu_j(t) \tag{8.13}$$

for all s, i and j.

We now show that the covariances $c_{ij}(s,t) \equiv \mathbf{Cov}[Q_i(s), Q_i(t)]$ also satisfy a linear ODE.

THEOREM 8.3

The covariance functions can be expressed as

$$c_{ij}(s,t) = m_i(s)\pi_{ij}(s,t), \quad s<t, \tag{8.14}$$

for $\pi_{ij}(s,t)$ in (8.13), so that $c_{ij}(s,t)$ is the unique solution to the ODE

$$\frac{\mathrm{d}}{\mathrm{d}t} c_{ij}(s,t) = \sum_{k=1}^{N} c_{ik}(s,t)\mu_k(t)p_{k_j}(t) - c_{ij}(s,t)\mu_j(t)\,. \tag{8.15}$$

Proof

Formula (8.14) is an easy consequence of the Markov and infinite-server properties. Formula (8.15) is obtained by multiplying (8.13) by $m_i(s)$. □

We conclude this section with a uniform acceleration expansion in this Markovian setting. Paralleling the treatment in section 5, we let the arrival and departure rates in the model indexed by ϵ be $\alpha_i(t)/\epsilon$ and $\mu_i(t)/\epsilon$, but we keep the transition function $p_{ij}(t)$ fixed. As before, let $m_i(t,\epsilon) = \mathbf{E}[Q_i^{\epsilon}(t)]$. For a vector $\mathbf{x}$, let $\Delta(\mathbf{x})$ be the diagonal matrix with elements of $\mathbf{x}$ on the diagonal.

The following result holds by essentially the same argument as in Massey [45].

THEOREM 8.4

Consider an accelerated version of an $(M_t/M_t/\infty)^N/M_t$ network, where α_i, μ_i and p_{ij} are $(n+1)$-times differentiable functions of t. If $\mathbf{m}(t,\epsilon) = (m_1(t,\epsilon), \ldots, m_N(t,\epsilon))$, then as $\epsilon \downarrow 0$,

$$\mathbf{m}(t,\epsilon) = \sum_{j=0}^{n} \epsilon^j \mathbf{m}^{(j)}(t) + O(\epsilon^{n+1}),$$

where

$$\mathbf{m}^{(0)}(t) = \boldsymbol{\alpha}(t) \cdot (\mathbf{I} - \mathbf{P}(t))^{-1} \Delta(\boldsymbol{\mu}(t))^{-1}$$

and for $n \geqslant 0$,

$$\mathbf{m}^{(n+1)}(t) = -\left(\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{m}^{(n)}(t)\right)(\mathbf{I} - \mathbf{P}(t))^{-1} \Delta(\boldsymbol{\mu}(t))^{-1}.$$

9. Large-population and fluid approximations

Given that the number of busy servers in each queue at time t is Poisson, approximations are not greatly needed for this model, but the performance of approximations for this model can provide insight into the performance of similar approximations for more complicated models, because the approximations are so easy to understand in this setting.

Here by "large-population approximation" we refer to the natural approximations when the mean number of busy servers at queue i at time t, $m_i(t)$, is large. Since the distribution is Poisson, the variance equals the mean and the standard deviation is its square root $\sqrt{m_i(t)}$. Hence, when the mean is large, the full distribution tends to concentrate closely about the mean (in a relative sense). This provides support for the first-order *deterministic approximation*

$$Q_i(t) \approx m_i(t), \quad t \geqslant 0, \tag{9.1}$$

and indicates that the error in (9.1) is of order $O(\sqrt{m_i(t)})$. This deterministic approximation has wider applicability, because (as noted in remark 2.3) the formula for the mean $m_i(t)$ does *not* depend on the Poisson property; i.e., it is valid for $(G_t/G_t/\infty)^N/G_t$ networks.

It should be noted that the deterministic mean-value approximation (9.1) can also be obtained directly as a *deterministic fluid approximation*. Instead of discrete customers, we assume that a continuous fluid arrives according to the external-arrival-rate function α. The fluid flow within the system is then described by the external-departure-rate functions δ_i and aggregate-arrival-rate and aggregate-departure-rate functions λ_i^+ and λ_i^-, while the fluid contents at the queues are described by the mean functions m_i. It is significant that these deterministic fluid quantities actually coincide with their expected-value counterparts in the stochastic model (due to the linearity of the stochastic model). They are also asymptotically correct as the populations grow.

Moreover, it is well known that the Poisson distribution approaches the normal distribution as the mean grows. Hence, we also have the second-order *normal approximation*

$$Q_i(t) \approx m_i(t) + \sqrt{m_i(t)}N(0,1)\,, \tag{9.2}$$

where $N(0,1)$ is a standard (mean 0, variance 1) normal random variable.

We have also characterized the joint distribution of $\{Q_i(t) : 1 \leqslant i \leqslant N, t \in \mathbb{R}\}$ in theorem 2.2. From the multivariate central limit theorem, we obtain a full *Gaussian-process approximation* when the means are large, i.e.,

$$\{(Q_1(t), \ldots, Q_N(t)) : t \geqslant 0\} \approx \{(X_1(t), \ldots, X_N(t)) : t \geqslant 0\}\,, \tag{9.3}$$

where $\{(X_1(t), \ldots, X_N(t)) : t \geqslant 0\}$ is a Gaussian stochastic process with the exact means and covariances determined here; i.e., for all positive integers k and all k-tuples $(t_1, \ldots, t_k)$, the joint distribution of

$$(X_1(t_1), \ldots, X_N(t_1), X_1(t_2), \ldots, X_N(t_2), \ldots, X_1(t_k), \ldots, X_N(t_k)) \tag{9.4}$$

has a kN-dimensional Gaussian (normal) distribution, which is completely specified by its means

$$\mathbf{E}[X_i(t_j)] = \mathbf{E}[Q_i(t_j)] \tag{9.5}$$

and covariances

$$\mathbf{Cov}[X_i(t_k), X_j(t_m)] = \mathbf{Cov}[Q_i(t_k), Q_j(t_m)]\,, \tag{9.6}$$

as determined by theorems 2.1 and 2.2.

In summary, there are three significant points about the Gaussian-process approximation: First, it characterizes all the joint distributions (over multiple queues and multiple time points); second, it has the exact means and covariances; and third, it is asymptotically correct as the means grow.

For the Markovian $(M_t/M_t/\infty)^N/M_t$ model, we saw in section 8 that the time-dependent means are characterized by a linear ODE. In this context, the deterministic fluid approximation (9.1) thus reduces to the linear ODE. This corresponds to the linear compartmental models in Sandberg [63], Brown [11], Jacques [33], and Garzia and Lockhart [27]. (The connection between the deterministic linear compartmental models and their stochastic counterparts seems to be quite well known; e.g., see Purdue [59].) This is a natural framework for introducing other model features that make the model nonlinear.

The full Gaussian-process approximation can also be developed by different methods, e.g., as in Kurtz [41], Ethier and Kurtz [24] and Glynn and Whitt [28]. These alternative methods are important to treat models that cannot be solved exactly. For example, the approach in Glynn and Whitt [28] yields full Gaussian approximations for open networks of infinite-server queues with non-Markovian arrival processes. Glynn and Whitt [28] consider only stationary models, but their approach also applies to $(G_t/G_t/\infty)^N/G_t$ models. For the $(M_t/M_t/\infty)^N/M_t$ model, we obtain the covariances for the Gaussian-process approximation by solving a second linear ODE, as indicated in theorem 8.3.

Since $(M_t/GI/\infty)^N/G_t$ models with phase-type service-time distributions can be reduced to $(M_t/M/\infty)^N/M$ models, the totally Markovian model has wide applicability.

10. Approximations for loss networks

We have been studying networks of infinite-server queues with nonstationary Poisson input largely because we believe that the results will help us approximately analyze nonstationary networks of queues with finitely many servers. In this section we briefly describe how the two basic approximation methods for computing blocking probabilities in single-station loss models described in Eick et al. [23] (which have a long history; see Palm [56], Newell [52] and Jagerman [34]) can be extended to loss networks. These and other related approximations remain to be more carefully studied.

To be concrete, we consider the $(M_t/GI/s/0)^N/G_t$ loss network; i.e., now we assume that each queue has only finitely many servers without extra waiting space. We assume that customers who find all servers busy at any queue upon arrival immediately leave the network without affecting future arrivals. Moreover, we assume that the service times are mutually independent and i.i.d. at each queue. As before, we assume that the external arrival process is a nonhomogeneous Poisson process with integrable external-arrival-rate function. Our object is to approximately determine the time-dependent blocking probabilities at each queue in the $(M_t/GI/s/0)^N/G_t$ model.

Recall that in the stationary $M/GI/s/0$ model the *Erlang blocking formula* is

$$B(n,x) = (x^n/n!) \Big/ \sum_{k=0}^{n} x^k/k!\,, \tag{10.1}$$

where n is the *number of servers* and x is the *offered load*, i.e., $x = \lambda \mathrm{E}[S]$ where λ is the arrival rate and $\mathrm{E}[S]$ is the mean service time.

THE POINTWISE-STATIONARY APPROXIMATION

The idea of the pointwise-stationary approximation (PSA) is to use the steady-state Erlang blocking formula in (10.1) associated with the time-dependent instantaneous arrival rates. To implement this idea with $(M_t/GI/s/0)^N/G_t$ loss networks, we can apply one of the infinite-server methods here to determine the time-dependent aggregate-arrival-rate functions λ_i^+ at each queue in the associated infinite-server model; e.g., by theorem 5.3 or theorem 7.2. The simple direct approach is then to use the Erlang blocking formula (10.1) at each queue with these time-dependent aggregate infinite-server arrival rates.

A natural refinement is to use the computed time-dependent blocking probabilities to modify our estimate of the time-dependent aggregate arrival rates in the loss model. For example, we can introduce an extra time-dependent probability of leaving the network (going to Δ^*) at each queue equal to the blocking probability. We then iteratively calculated time-dependent aggregate arrival rates and time-dependent blocking probabilities, hopefully arriving at a stable solution, as in the Erlang-fixed-point and reduced-load approximations in Kelly [38].

THE MODIFIED-OFFERED-LOAD APPROXIMATION

Just as with PSA, the idea of the modified-offered-load (MOL) approximation is to use the steady-state Erlang loss formula associated with an instantaneous arrival rate but with MOL we use the time-dependent mean number of busy servers in the infinite-server model for the instantaneous offered load (arrival rate times mean service time). As with PSA, this is exact in the stationary $M/GI/s/0$ model; see Jagerman [34].

To implement this idea with loss networks, we can apply one of the infinite-server methods here to determine the time-dependent mean number of busy servers at queue i, $m_i(t)$, for each i. The simple direct approach is then to use the steady-state Erlang blocking formula in (10.1) at each queue with these time-dependent means $m_i(t)$ serving as the offered loads. As with PSA, it is natural to consider iterative refinements in which the estimated time-dependent blocking probabilities are used to modify our estimates of the time-dependent offered loads.

Remark 10.1

As noted in remark 2.4, the infinite-server results extend to cover models in which customers simultaneously use multiple resources, as in Kelly [38]. In that setting, a natural candidate for PSA is based on the time-dependent Erlang fixed-point approximation as described in [38] associated with instantaneous time-dependent arrival rates. If customers do not move within the network, then these instantaneous time-dependent arrival rates are the given external arrival rates. Otherwise, we can determine the aggregate infinite-server arrival rates and then apply the Erlang fixed-point approximation to them. A natural candidate for MOL is the time-dependent Erlang fixed-point approximation as described in [38] associated with the time-dependent offered loads, which are estimated by the time-dependent means $m_i(t)$ from the infinite-server model.

Remark 10.2

A principal contribution of Eick et al. [23] is to develop a refined hybrid approximation for $M_t/GI/s/0$ models that combines PSA and MOL. It is natural to look for such refinements in the network setting.

Remark 10.3

The infinite-server covariance formulas developed in this paper seem promising for developing refined approximations for loss networks.

Appendix: A product-form counterexample

The product-form results in theorems 1.2 and 2.1 depend strongly on the infinite-server property of our networks. To demonstrate this, we now show that the time-dependent queue-length distribution is in general *not* product-form even in a stationary Markovian open Jackson network model that starts empty. In particular, we consider an $(M/M/1)^2/M$ network containing two single-server queues in tandem; i.e., the routing probabilities are $p_{12} = 1$ and $p_{ij} = 0$ otherwise.

Let this model have external arrive-rate function $\alpha(t) = \lambda, t \geqslant 0$, with $\lambda < 1$, and exponential service times with means $\mu_1^{-1} = 1$ and $\mu_2^{-1} = \epsilon$ for small ϵ. Hence, it suffices to consider the service rate at the second queue as being 0, and we do in our calculations under this assumption. It is then not difficult to show that in general we do *not* have

$$P(Q_1(t) = 0)P(Q_2(t) = 0) = P(Q_1(t) = 0, Q_2(t) = 0)\,. \qquad \text{(A.1)}$$

To see this, we indicate how to calculate the three quantities in (A.1). First,

$$P(Q_1(t) = 0) = 1 - \lambda B(t)\,, \qquad \text{(A.2)}$$

where $B(t)$ is the $M/M/1$ busy-period cdf; see corollary 4.2.3 of Abate and Whitt [73]. Second,

$$\begin{aligned} P(Q_2(t) = 0) =& P(\text{no arrivals to queue1}) \\ &+ P(\text{at least one arrival to queue 1, who does not go to queue 2}) \\ =& \mathrm{e}^{-\lambda t} + \int_0^t \lambda \mathrm{e}^{-\lambda u} \mathrm{e}^{-(t-u)} \mathrm{d}u = \mathrm{e}^{-\lambda t} + \frac{\lambda \mathrm{e}^{-t}(1 - \mathrm{e}^{-(\lambda+1)t})}{\lambda + 1} \end{aligned} \qquad \text{(A.3)}$$

and, third,

$$P(Q_1(t) = Q_2(t) = 0) = \mathrm{e}^{-\lambda t}\,. \qquad \text{(A.4)}$$

Formulas (A.1)–(A.4) would imply that

$$\begin{aligned} B(t) &= \lambda^{-1}\left(1 - \frac{P(Q_1(t) = Q_2(t) = 0)}{P(Q_2(t) = 0)}\right) \\ &= \frac{1 - \mathrm{e}^{-\lambda t}}{\lambda + (\lambda + 1)\mathrm{e}^{(1-\lambda)t} - \lambda \mathrm{e}^{-(\lambda+1)t}}\,, \end{aligned} \qquad \text{(A.5)}$$

which does not hold. Hence, (A.1) does not hold.

One might object to this example because the long-run arrival rate at the second queue (greatly) exceeds the service rate, so that it is an unstable model. An alternative demonstration without unstable models is via light-traffic asymptotics, as in Whitt [74]. Then we let both service rates be 1, but we make the arrival rate be very very small, so that the probability of at least k external arrivals by time t is of order $O(\epsilon^k)$ as $\epsilon \to 0$. It suffices to consider only the cases of 0, 1 or 2 arrivals. Then

$$\begin{aligned} P(Q_1(t) = 0) &= 1 - a_1\epsilon - a_2\epsilon^2 + o(\epsilon^2), \\ P(Q_2(t) = 0) &= 1 - b_1\epsilon - b_2\epsilon^2 + o(\epsilon^2), \\ P(Q_1(t) = Q_2(t) = 0) &= 1 - c_1\epsilon - c_2\epsilon^2 + o(\epsilon^2). \end{aligned}$$

In order to have (A.1), we need

$$\begin{aligned} &(1 - a_1\epsilon - a_2\epsilon^2 + o(\epsilon^2))(1 - b_1\epsilon - b_2\epsilon^2 + o(\epsilon^2)) \\ &\quad = (1 - (a_1 + b_1)\epsilon - (a_2 + b_2 - a_1 b_1)\epsilon^2 + o(\epsilon^2) \\ &\quad = (1 - c_1\epsilon - c_2\epsilon^2 + o(\epsilon^2). \end{aligned} \tag{A.6}$$

Detailed calculations show that $c_1 = a_1 + b_1$, but we need not have $c_2 = a_2 + b_2 - a_1 b_1$.

References

[1] S. Asmussen and T. Rolski, Risk theory in a periodic environment: the Cramér–Lundberg approximation and Lundberg's inequality, Institute for Electronic Systems, University of Aaborg, Denmark (1991).
[2] N. Bambos and J. Walrand, On queues with periodic inputs, J. Appl. Prob. 26 (1989) 381–389.
[3] R.E. Barlow and F. Proschan, *Statistical Theory of Reliability and Life Testing* (Holt, Rinehart and Winston, New York, 1975).
[4] M.S. Bartlett, Some evolutionary stochastic processes, J. Roy. Statist. Soc. B 11 (1949) 211–229.
[5] P. Billingsley, *Convergence of Probability Measures* (Wiley, New York, 1968).
[6] J.P.C. Blanc, The transient behavior of networks of infinite server nodes, Centre for Mathematics and Computer Science, Amsterdam, Report OS-R8408 (1984).
[7] N. Bourbaki, *Elements of Mathematics, General Topology*, Part 2 (Addison–Wesley, Reading, MA, 1966).
[8] P. Brémaud, *Point Processes and Queues* (Springer, New York, 1981).
[9] M. Brown, An invariance property of a Poisson process, J. Appl. Prob. 6 (1969) 453–458.
[10] M. Brown and S.M. Ross, Some results for infinite server Poisson queues, J. Appl. Prob. 6 (1969) 604–611.
[11] R.F. Brown, Compartmental system analysis, IEEE Trans. Biomed. Eng. BME-27 (1980) 1–11.
[12] M.J. Carrillo, Extensions of Palm's theorem: A review, Manag. Sci. 37 (1991) 739–744.
[13] C.L. Chiang, *An Introduction to Stochastic Processes and their Applications* (Krieger, Huntington, NY, 1980).
[14] K.L. Chung, *A Course in Probability Theory*, 2nd ed. (Academic Press, New York, 1974).
[15] T. Collings and C. Stoneman, The $M/M/\infty$ queue with varying arrival and departure rates, Oper. Res. 24 (1976) 760–773.
[16] D.J. Daley, Queueing output processes, Adv. Appl. Prob. 8 (1976) 395-415.
[17] D.J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes* (Springer, New York, 1988).

[18] R.L. Disney and P.C. Kiessler, *Traffic Processes in Queueing Networks, A Markov Renewal Approach* (The Johns Hopkins University Press, Baltimore, 1987).
[19] J.L. Doob, *Stochastic Processes* (Wiley, New York, 1953).
[20] A. Duda, Diffusion approximations for time-dependent queueing systems, IEEE J. Sel. Areas Commun. SAC-4 (1986) 905–918.
[21] S.G. Eick, W.A. Massey and W. Whitt, The physics of the $M_t/G/\infty$ queue, Oper. Res. 41 (1993), to appear.
[22] S.G. Eick, W.A. Massey and W. Whitt, $M_t/G/\infty$ queues with sinusoidal arrival rates, Manag. Sci. 39 (1993) 241–252.
[23] S.G. Eick, W.A. Massey and W. Whitt, Infinite-server approximations for multi-server loss models with time-dependent arrival rates, AT&T Bell Laboratories, Murray Hill, NJ, in preparation.
[24] S.N. Ethier and T.G. Kurtz, *Markov Processes, Characterization and Convergence* (Wiley, New York, 1986).
[25] M.J. Faddy, Stochastic compartmental models as approximations to more general stochastic systems with the general stochastic epidemic as an example, Adv. Appl. Prob. 9 (1977) 448–461.
[26] R.D. Foley, The non-homogeneous $M/G/\infty$ queue, Opsearch 19 (1982) 40–48.
[27] M.R. Garzia and C.M. Lockhart, Nonhierarchical communications networks: an application of compartmental modeling, IEEE Trans. Commun. COM-37 (1989) 555–564.
[28] P.W. Glynn and W. Whitt, A new view of the heavy-traffic limit theorem for infinite-server queues, Adv. Appl. Prob. 23 (1991) 188–209.
[29] L. Green, P. Kolesar and A. Svoronos, Some effects of nonstationarity on multiserver Markovian queueing systems, Oper. Res. 39 (1991) 502–511.
[30] R.W. Hall, *Queueing Methods for Services and Manufacturing* (Prentice–Hall, Englewood Cliffs, NJ, 1991).
[31] J.M. Harrison and A.J. Lemoine, A note on networks of infinite-server queues, J. Appl. Prob. 18 (1981) 561–567.
[32] R.R.P. Jackson and P. Aspden, A transient solution to the multistage Poisson queueing system with infinite servers, Oper. Res. 28 (1980) 618–622.
[33] J.A. Jacques, *Compartmental Analysis in Biology and Medicine*, 2nd ed. (University of Michigan Press, Ann Arbor, 1985).
[34] D.L. Jagerman, Nonstationary blocking in telephone traffic, Bell Syst. Tech. J. 54 (1975) 625–661.
[35] N.L. Johnson and S. Kotz, *Discrete Distributions* (Wiley, New York, 1969).
[36] J. Keilson and L.D. Servi, Networks of non-homogeneous $M/G/\infty$ systems, GTE Laboratories, Waltham, MA (1989).
[37] F.P. Kelly, *Reversibility and Stochastic Networks* (Wiley, New York, 1979).
[38] F.P. Kelly, Loss networks, Ann. Appl. Prob. 1 (1991) 319–378.
[39] A.Y. Khintchine, *Mathematical Methods in the Theory of Queueing* (Trudy Mat. Inst. Steklov 49, in Russian (1955)). (English translation by Charles Griffin and Co., London, 1960).
[40] J.F.C. Kingman, Markov population processes, J. Appl. Prob. 6 (1969) 1–18.
[41] T.G. Kurtz, *Approximation of Population Processes*, CBMS-NSF Regional Conf. Series in Appl. Math. 36 (SIAM, Philadelphia, 1981).
[42] A.J. Lemoine, Waiting time and workload in queues with periodic Poisson input, J. Appl. Prob. 26 (1989) 390–397.
[43] W.C.Y. Lee, *Mobile Cellular Telecommunications Systems* (McGraw–Hill, New York, 1989).
[44] P.A.W. Lewis and G.S. Shedler, Simulation of nonhomogeneous Poisson processes by thinning, Naval Res. Logist Quart. 29 (1979) 403–413.
[45] W.A. Massey, Asymptotic analysis of the time dependent $M/M/1$ queue, Math. Oper. Res. 10 (1985) 305–327.

[46] J.H. Matis, An introduction to stochastic compartmental models in pharmacokinetics, in: *Pharmacokinetics*, eds. A. Percele and A. Rescigno (Plenum, New York, 1988) pp. 113–128.

[47] J.H. Matis and T.E. Wehrly, Generalized stochastic compartmental models with Erlang transit times, J. Pharmacokin. Biopharm. 18 (1990) 589–607.

[48] N.M. Mirasol, The output of an $M/G/\infty$ queueing system is Poisson, Oper. Res. 11 (1963) 282–284.

[49] J. Neveu, *Mathematical Foundations of the Calculus of Probability* (Holden–Day, San Francisco, 1965).

[50] G.F. Newell, The $M/G/\infty$ queue, SIAM J. Appl. Math. 14 (1966) 86–88.

[51] G.F. Newell, *Approximate Stochastic Behavior of n-Server Service Systems with Large n*, Lecture Notes in Econ. and Math. Systems 87 (Springer, Berlin, 1973).

[52] G.F. Newell, *Applications of Queueing Theory* (Chapman and Hall, London, 1982).

[53] G.L. O'Brien, The comparison method for stochastic processes, Ann. Probab. 3 (1975) 80–88.

[54] K.L. Ong and M.R. Taaffe, Nonstationary queues with interrupted Poisson arrivals and unreliable/repairable servers, Queueing Systems 4 (1989) 27–46.

[55] A.G. Pakes and N. Kaplan, On the subcritical Bellman-Harris process with immigration, J. Appl. Prob. 11 (1974) 652–668.

[56] C. Palm, Intensity variations in telephone traffic, Ericsson Technics 44 (1943) 1–189 (in German). (English translation by North-Holland, Amsterdam, 1988).

[57] K.R. Parthasarathy, *Probability Measures on Metric Spaces* (Academic, New York, 1967).

[58] A. Prékopa, On secondary processes generated by a random point distribution of Poisson type, Annales Univ. Sci. Budapest de Eötvös Nom. Sectio Math. 1 (1958) 153–170.

[59] P. Purdue, Stochastic compartmental models: a review of the mathematical theory with ecological applications, in: *Compartmental Analysis of Ecosystems Models*, eds. J.H. Matis, B.C. Patten and G.C. White (International Co-operative Publishing House, Fairland, MD, 1979) pp. 223–260.

[60] A. Renyi, Remarks on the Poisson process, Studia Sci. Math. Hungar. 2 (1967) 119–123.

[61] T. Rolski, Queues with nonstationary inputs, Queueing Systems 5 (1989) 113–130.

[62] H.L. Royden, *Real Analysis*, 2nd ed. (London, Macmillan, 1968).

[63] I.W. Sandberg, On the mathematical foundations of compartmental analysis in biology, medicine and ecology, IEEE Trans. Circuit Syst. CAS-25 (1978) 273–279.

[64] R.F. Serfozo, Point processes, in: *Stochastic Models*, eds. D.P. Heyman and M.J. Sobel (North-Holland, Amsterdam, 1990) pp. 1–93.

[65] A.V. Skorokhod, Limit theorems for stochastic processes, Theor. Prob. Appl. 1 (1956) 261–290.

[66] D. Stoyan, *Comparison Methods for Queues and Other Stochastic Models* (Wiley, New York, 1983).

[67] H. Thorisson, On regenerative and ergodic properties of the k-server queue with non-stationary Poisson arrivals, J. Appl. Prob. 22 (1985) 893–902.

[68] J. Walrand, *An Introduction to Queueing Networks* (Prentice–Hall, Englewood Cliffs, NJ, 1989).

[69] D. Weiner and P. Purdue, A semi-Markov approach to stochastic compartmental models, Commun. Statist. A6 (1977) 1231–1243.

[70] W. Whitt, Some useful functions for functional limit theorems, Math. Oper. Res. 5 (1980) 67–85.

[71] W. Whitt, On the heavy-traffic limit theorem for $GI/G/\infty$ queues, Adv. Appl. Prob. 14 (1982), 171–190.

[72] P. Whittle, *Systems in Stochastic Equilibrium* (Wiley, New York, 1986).

[73] J. Abate and W. Whitt, Transient behavior of the $M/M/1$ queue via Laplace transforms, Adv. Appl. Prob. 20 (1988) 145–178.

[74] W. Whitt, Deciding which queue to join: some counterexamples, Oper. Res. 34 (1986) 55–62.