An Analysis of the Modified Offered Load Approximation for the Nonstationary Erlang Loss Model

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

February 20, 1994

Abstract

A fundamental problem that led to the development of queueing theory is the probabilistic modelling of the number of busy lines in telephone trunk groups. Based on the behavior of real telephone systems, a natural model to use would be the $M_t/G/s/0$ queue, which has s servers, no extra waiting space and a nonhomogeneous Poisson arrival process (M_t). Unfortunately, so far queueing theory has provided an exact analysis for only the $M/G/s/0$ queue in steady state, which yields the Erlang blocking formula, and the $M_t/G/\infty$ queue, which treats nonstationary arrivals at the expense of having infinitely many servers. However, these results can be synthesized to create a *modified offered load* (MOL) approximation for the $M_t/G/s/0$ queue: the distribution of the number of busy servers in the $M_t/G/s/0$ queue at time t is approximated by the steady-state distribution of the stationary $M/G/s/0$ queue with an offered load (arrival rate times mean service time) equal to the mean number of busy servers in the $M_t/G/\infty$ queue at time t. In addition to being a simple effective approximation scheme, the MOL approximation makes all of the exact results for infinite server queues relevant to the analysis of nonstationary loss systems.

In this paper, we provide a rigorous mathematical basis for the MOL approximation. We find an expression for the difference between the $M_t/G/s/0$ queue length distribution and its MOL approximation. From this expression we extract bounds on the error and deduce when one distribution stochastically dominates the other.

Keywords: Performance Analysis, Traffic Theory, Nonstationary Queues, Nonstationary Erlang Loss Model, Erlang Blocking Formula, Infinite Server Queues, Phase-Type Service, Time Ordered Exponentials, Stochastic Dominance.

1 Introduction

The probabilistic modelling of the number of busy lines in telephone trunk groups is one of the fundamental problems that led to the development of queueing theory. It was first formulated as an $M/M/s/0$ queue by Erlang [7]. He gave an exact solution for the steady state distribution, which gave rise to the well known *Erlang blocking formula*. This formula states that if $Q_s(t)$ is the random queue length at time t for the $M/M/s/0$ system (queueing here means "waiting" for service completion), then

$$\lim_{t\to\infty} \mathsf{P}(Q_s(t) = s) = \beta_s(\lambda/\mu) \equiv \frac{(\lambda/\mu)^s}{s!} \bigg/ \sum_{k=0}^{s} \frac{(\lambda/\mu)^k}{k!} \tag{1.1}$$

where λ is the Poisson arrival rate, $1/\mu \equiv \mathsf{E}[S]$ is the mean of the exponentially distributed random service time S, and s equals the total number of servers (trunk lines). Since Poisson arrivals see time averages, $\beta_s(\lambda/\mu)$ is also the long-run proportion of arrivals that are lost.

The Erlang blocking formula also applies to the $M/G/s/0$ queue with a general service time distribution, having the same Poisson arrival rate and mean service time. This *insensitivity property* means that the assumption of exponential service is superfluous, which expands the model's range of applicability. Moreover, limit theorems for general point processes show that modelling the arrival process as Poisson is not too restrictive, e.g. see page 281 of Daley and Vere-Jones [2].

In fact, the most restrictive assumption in the $M/M/s/0$ model is having a constant arrival rate. Significant steps were made to solve this problem starting in the 1930's, see Palm [16], Khintchine [11], and Prékopa [17]. They found the exact solution for the time dependent distribution in the $M_t/G/\infty$ model. This infinite server queue captures the effect of a time varying mean arrival rate and general service times, but at the expense of letting the total number of servers be infinite. If $Q_\infty(t)$ equals the queue length at time t in the $M_t/G/\infty$ model, and $Q_\infty(t_0) = 0$ for some $t_0 < t$, then

$$\mathsf{P}(Q_\infty(t) = k) = e^{-m_\infty(t)} \frac{m_\infty(t)^k}{k!} \tag{1.2}$$

for all non-negative integers k, where

$$m_\infty(t) = \mathsf{E}\left[\int_{t-S}^{t} \lambda(\tau)\, d\tau\right], \tag{1.3}$$

with $\lambda(t) = 0$ for all $t < t_0$. A simple direct approximation for the blocking probability $\mathsf{P}(Q_s(t) = s)$ in the $M_t/G/s/0$ model is the tail probability $\mathsf{P}(Q_\infty(t) \geq s)$.

These exact solutions to the $M/G/s/0$ and $M_t/G/\infty$ models led to a better technique for approximating the time-dependent queue length distribution in the $M_t/G/s/0$ model. It is called the the *modified-offered-load approximation* (MOL), see Jagerman [8]. Since the Erlang blocking formula is a function of λ/μ, and λ/μ is the mean queue length in the steady-state stationary $M/G/\infty$ queue, we should obtain a reasonable approximation for the time-dependent blocking probability in the $M_t/G/s/0$ queue if we substitute $m_\infty(t)$ for λ/μ in the Erlang blocking formula. Thus the MOL approximation is

$$\mathsf{P}(Q_s(t) = s) \approx \beta_s\big(m_\infty(t)\big) = \mathsf{P}\left(Q_\infty(t) = s \mid Q_\infty(t) \leq s\right). \tag{1.4}$$

where β is given by (1.1) and $m_\infty(t)$ is given by (1.3).

From (1.4), we see that MOL enables us to apply the exact results for the $M_t/G/\infty$ model to the analysis of the $M_t/G/s/0$ model. For example, we applied the MOL approximation to help understand the impact of the service-time distribution in an $M_t/G/s/0$ queue in Davis et al. [3]. The MOL approximation was also a major motivation for the papers by Eick et al. [5], [6] on the $M_t/G/\infty$ model. Moreover, since a solution exists for the transient distribution of the $M_t/G_t/\infty$ queue, see Brown and Ross [1] and Massey and Whitt [15], we can apply the MOL approximation to the $M_t/G_t/s/0$ queue as well.

The goal of this paper is to create a mathematical theory supporting this heuristic approximation. In Section 4, we do so by constructing a formal solution to the error between the exact probability solution and the MOL approximation for the case of time dependent phase type service. From this main result, we derive simple, computable error bounds for MOL. For the $M_t/M/s/0$ queue, we will show that

$$\sup_{0\leq\tau\leq t}\Big|\mathsf{P}(Q_s(\tau)=s)-\beta_s\big(m_\infty(\tau)\big)\Big|\leq 2\int_0^t\beta_s\big(m_\infty(\tau)\big)\Big(1-\beta_s\big(m_\infty(\tau)\big)\Big)\left|\frac{dm_\infty}{dt}(\tau)\right|d\tau. \quad (1.5)$$

where we assume that the distribution of $Q_s(0)$ is the steady state $M/M/s/0$ distribution with parameter $m_\infty(0)$, which is a family of distributions that includes the point masses at 0 and s; see (4.6). For the more general $M_t/G/s/0$ system, we will also show that

$$\sup_{0\leq\tau\leq t}\Big|\mathsf{P}(Q_s(\tau)=s)-\beta_s\big(m_\infty(\tau)\big)\Big|\leq 2\cdot\int_0^t\beta_s\big(m_\infty(\tau)\big)\left|\frac{dm_\infty}{d\tau}(\tau)\right|d\tau; \quad (1.6)$$

see (4.7). These error bounds imply that the MOL approximation is asymptotically correct as either the derivative of $m_\infty(t)$ or the tail probability $\mathsf{P}(Q_\infty(t)\geq s)$ in the $M_t/G_t/\infty$ model approaches 0. In turn, these limits for the $M_t/G/\infty$ model hold as the derivative of $\lambda(t)$ approaches 0 and as $s\to\infty$. More generally, these bounds support the intuition that MOL should perform better when the arrival rate $\lambda(t)$ changes more slowly and when the blocking probability is lower.

We obtain alternative bounds for the $M_t/M/s/0$ system in Section 5 by using the ℓ_1-norm on cumulative distribution functions instead of the ℓ_1-norm on probability mass functions. For example, with the same initial conditions, if $\mu=1$ and λ is bounded with a bounded derivative λ' on $[0,\infty)$, then

$$\sup_{t\geq 0}\Big|\mathsf{E}\big[Q_s(t)\big]-m_\infty(t)\Big(1-\beta_s\big(m_\infty(t)\big)\Big)\Big|\leq|\lambda'|_\infty s\beta_s\big(|\lambda|_\infty\big), \quad (1.7)$$

where $|f|_\infty=\sup_{x\geq 0}|f(x)|$ for all bounded functions f on $[0,\infty)$. Note that (1.7) is uniform over all time.

In Section 6, we investigate in detail the special case of an $M/M/s/0$ model which experiences a change of parameters at time 0. Hence we are describing the transient behavior going from one stationary regime to another. Here we exploit the fact that the generator after time 0 is not time-dependent.

2 The $M_t/PH_t/s/0$ Queue

We define the $M_t/PH_t/s/0$ queueing system as follows. It has s independent servers, each with a common time-dependent phase-type service, and an arrival process that is non-homogeneous Poisson. The class of phase-type service-time distributions is quite general, because phase-type distributions are dense in the space of all distributions. This assumption enables us to construct an extended finite state space such that the queue length process is Markovian in continuous time. Let C equal the finite set of service phases (which we assume does not change with time). To obtain a general state description that makes our system Markovian, we count the number of customers in each phase of service. We define S_C to be the corresponding state space, allowing arbitrary numbers of customers. The states in S_C can be denoted by $\mathbf{k}$, where every $\mathbf{k} \in S_C$ is written as the formal sum

$$\mathbf{k} = \sum_{\alpha \in C} k_\alpha \mathbf{e}_\alpha, \tag{2.1}$$

such that $\mathbf{e}_\alpha$ is an independent basis vector, corresponding to the service phase α, and each k_α is a non-negative integer, representing the number of customers in service phase α. The set S_C is the state space for the case of $s = \infty$. In algebraic terms, S_C is referred to as the *free abelian semigroup* generated by the set C, in contrast to the free nonabelian semigroup structure used in Massey [13] for the state space of a multiclass single server queue. Finally, if we denote the *length* of $\mathbf{k}$ as $|\mathbf{k}|$, which equals $\sum_{\alpha \in C} k_\alpha$, then the state space for our queueing model $M_t/PH_t/s/0$, will be $S_C(s)$, where

$$S_C(s) = \{\, \mathbf{k} \,|\, \mathbf{k} \in S_C \;\text{ and }\; |\mathbf{k}| \le s \,\}. \tag{2.2}$$

Now let $\{\, \mathbf{Q}_s(t) \mid t \ge 0 \,\}$ be the Markovian queue length process with state space $S_C(s)$. Its infinitesimal generator will be constructed from the following parameters:

$\lambda_\alpha(t)$ = the external arrival rate at time t for a customer that initiates service in phase α.

$\mu_\alpha(t)$ = the service rate at time t for phase α.

$p_{\alpha\beta}(t)$ = the probability that phase β service is initiated at time t, given that phase α service has just terminated.

$q_\alpha(t)$ = the probability that the entire service has terminated at time t, given that phase α service has just terminated.

If $p(\mathbf{k}, t) \equiv \mathsf{P}(\mathbf{Q}_s(t) = \mathbf{k})$, then for $|\mathbf{k}| < s$, $\mathbf{Q}_s(t)$ has the following set of forward equations:

$$\begin{aligned}\frac{d}{dt}p(\mathbf{k},t) = \sum_{\alpha\in C}\Big[\, &\lambda_\alpha(t)\mathrm{sgn}(k_\alpha)p(\mathbf{k}-\mathbf{e}_\alpha,t) + \mu_\alpha(t)(k_\alpha+1)q_\alpha(t)p(\mathbf{k}+\mathbf{e}_\alpha,t)\\ &+ \sum_{\beta\in C}\mu_\beta(t)(k_\beta+1)p_{\beta\alpha}(t)\mathrm{sgn}(k_\alpha)p(\mathbf{k}-\mathbf{e}_\alpha+\mathbf{e}_\beta,t) - (\lambda_\alpha(t)+\mu_\alpha(t)k_\alpha)p(\mathbf{k},t)\Big],\end{aligned}$$

where $\mathrm{sgn}(k)$ equals 0 if $k = 0$, and 1 if $k > 0$. When $|\mathbf{k}| = s$, we have

$$\begin{aligned}\frac{d}{dt}p(\mathbf{k},t) = \sum_{\alpha\in C}\Big[\, &\lambda_\alpha(t)\mathrm{sgn}(k_\alpha)p(\mathbf{k}-\mathbf{e}_\alpha,t)\\ &+ \sum_{\beta\in C}\mu_\beta(t)(k_\beta+1)p_{\beta\alpha}(t)\mathrm{sgn}(k_\alpha)p(\mathbf{k}-\mathbf{e}_\alpha+\mathbf{e}_\beta,t) - \mu_\alpha(t)k_\alpha p(\mathbf{k},t)\Big].\end{aligned}$$

Letting $\ell(S_C(s))$ be the vector space for real valued functions on $S_C(s)$, we can encode these equations as

$$\frac{d}{dt}\mathbf{p}(t) = \mathbf{p}(t)\mathbf{A}(t), \tag{2.3}$$

where

$$\mathbf{p}(t) = \sum_{\mathbf{k}\in S_C(s)} \mathsf{P}(\mathbf{Q}_s(t) = \mathbf{k})\mathbf{e}_\mathbf{k} \tag{2.4}$$

and $\mathbf{A}(t)$ is the corresponding infinitesimal generator that is a linear operator on $\ell(S_C(s))$ composed of the arrival and service rates for the queueing process. The $\mathbf{e}_\mathbf{k}$'s are the unit basis vectors for $\ell(S_C(s))$, where each $\mathbf{e}_\mathbf{k}$ corresponds to the indicator function for the singleton set $\{\mathbf{k}\}$. In general, $\mathbf{p}(t)$ is a *probability vector*, since it is a vector encoding of the probability distribution given by $p(\mathbf{k}, t)$. We will use the terms probability vector and probability distribution interchangeably. Formally, we can solve for $\mathbf{p}(t)$, and get

$$\mathbf{p}(t) = \mathbf{p}(0)\mathbf{E}_\mathbf{A}(t), \tag{2.5}$$

where $\mathbf{E}_\mathbf{A}(t)$ is the *time-ordered exponential* of the family of generators $\{\, \mathbf{A}(\tau) \,|\, 0 \le \tau \le t \,\}$. When $\mathbf{A}$ is a constant operator, then the corresponding time-ordered exponential is just $\exp(t\mathbf{A})$. In general, it is the unique operator solution to the equation

$$\frac{d}{dt}\mathbf{E}_\mathbf{A}(t) = \mathbf{E}_\mathbf{A}(t)\mathbf{A}(t), \tag{2.6}$$

where $\mathbf{E}_\mathbf{A}(0) = \mathbf{I}$, the identity operator. For all $0 \le \tau \le t$, it will also be useful to define

$$\mathbf{E}_\mathbf{A}(\tau, t) \equiv \mathbf{E}_\mathbf{A}(\tau)^{-1}\mathbf{E}_\mathbf{A}(t). \tag{2.7}$$

A thorough treatment of the issues of existence, uniqueness, and construction of time-ordered exponentials can be found in Dollard and Friedman [4].

3 The $M_t/PH_t/\infty$ Queue

Our approximate analysis of the $M_t/PH_t/s/0$ employs the exact solution for its infinite counterpart, the $M_t/PH_t/\infty$ queue. Let $\{\, \mathbf{Q}_\infty(t) \mid t \ge 0 \,\}$ be the $M_t/PH_t/\infty$ queue length process. Its marginal probabilities $q(\mathbf{k}, t) \equiv \mathsf{P}(\mathbf{Q}_\infty(t) = \mathbf{k})$ for all $\mathbf{k} \in S_C$, will then solve the following set of forward equations:

$$\begin{aligned}\frac{d}{dt}q(\mathbf{k}, t) = \sum_{\alpha\in C}\Big[\, &\lambda_\alpha(t)\mathrm{sgn}(k_\alpha)q(\mathbf{k} - \mathbf{e}_\alpha, t) + \mu_\alpha(t)(k_\alpha + 1)q_\alpha(t)q(\mathbf{k} + \mathbf{e}_\alpha, t) \\ &+ \sum_{\beta\in C} \mu_\beta(t)(k_\beta + 1)p_{\beta\alpha}(t)\mathrm{sgn}(k_\alpha)q(\mathbf{k} - \mathbf{e}_\alpha + \mathbf{e}_\beta, t) - (\lambda_\alpha(t) + \mu_\alpha(t)k_\alpha)q(\mathbf{k}, t)\Big].\end{aligned}$$

Now for any $\mathbf{x}$ in $\ell(C)$, the vector space of real-valued functions on C, and any state $\mathbf{k} \in S_C$, define the following useful operations:

$$\mathbf{x}^\mathbf{k} \equiv \prod_{\alpha\in C} x_\alpha^{k_\alpha}, \quad \mathbf{k}! \equiv \prod_{\alpha\in C} k_\alpha!, \quad \text{and} \quad |\mathbf{x}| \equiv \sum_{\alpha\in C} |x_\alpha|, \tag{3.1}$$

where $x_\alpha = \mathbf{x}(\alpha)$. We will also represent $\mathbf{x}$ by the formal sum $\sum_{\alpha\in C} x_\alpha \mathbf{e}_\alpha$. Hence $|\mathbf{x}|$ is the ℓ_1-norm applied to $\mathbf{x}$. In this notation, the multinomial theorem is transformed into

$$\sum_{|\mathbf{k}|=s} \frac{\mathbf{x}^{\mathbf{k}}}{\mathbf{k}!} = \frac{|\mathbf{x}|^s}{s!}. \tag{3.2}$$

Theorem 8.2 of Massey and Whitt [15] gives the exact solution for the $M_t/PH_t/\infty$ queue with appropriate initial distributions, as

$$q(\mathbf{k},t) = \frac{e^{-m_\infty(t)}\mathbf{m}_\infty(t)^{\mathbf{k}}}{\mathbf{k}!}, \tag{3.3}$$

where $\mathbf{m}_\infty(t) = \sum_{\alpha\in C} m_\infty^\alpha(t)\mathbf{e}_\alpha$ and $m_\infty(t) = |\mathbf{m}_\infty(t)| = \sum_{\alpha\in C} m_\infty^\alpha(t)$, such that the $m_\infty^\alpha(t)$'s solve the set of differential equations

$$\frac{d}{dt} m_\infty^\alpha(t) = \lambda_\alpha(t) + \sum_{\beta\in C} \mu_\beta(t) m_\infty^\beta(t) p_{\beta\alpha}(t) - \mu_\alpha(t) m_\infty^\alpha(t) \tag{3.4}$$

for all $\alpha \in C$, with arbitrary $\mathbf{m}_\infty(0)$. The solution (3.3) is valid provided that the initial distribution $p(\mathbf{k},0)$ is also of the same form depending on the initial mean vector $\mathbf{m}_\infty(0)$.

4 The Fundamental Identity and Bounds for MOL

The MOL approximation is defined to be $p^*(\mathbf{k},t)$ for S_C, where

$$\mathsf{P}(\mathbf{Q}_s(t) = \mathbf{k}) \approx p^*(\mathbf{k},t) \equiv \frac{\mathbf{m}_\infty(t)^{\mathbf{k}}}{\mathbf{k}!} \Bigg/ \sum_{j=0}^{s} \frac{m_\infty(t)^j}{j!} = \mathsf{P}\left(\mathbf{Q}_\infty(t) = \mathbf{k} \mid |\mathbf{Q}_\infty(t)| \leq s\right), \tag{4.1}$$

where the components of the vector $\mathbf{m}_\infty(t) = \sum_{\alpha\in C} m_\infty^\alpha(t)\mathbf{e}_\alpha$ solve the differential equations given by (3.4), with arbitrary initial vector $\mathbf{m}_\infty(0)$. We now present our main result, which we prove in Section 7.

Theorem 4.1 *Let $\{\, \mathbf{Q}_s(t) \mid t \geq 0 \,\}$ be the Markovian queueing process for $M_t/\mathrm{PH}_t/s/0$ with the family of infinitesimal generators $\{\, \mathbf{A}(t) \mid t \geq 0 \,\}$. Let $\mathbf{p}(t)$ be the probability vector for the distribution of $\mathbf{Q}_s(t)$, with an initial distribution $\mathbf{p}(0) = \mathbf{p}^*(0)$, which is of the form (3.3) for arbitrary $\mathbf{m}_\infty(0)$. Let $\mathbf{p}^*(t)$ be the probability vector for the modified offered load approximation, then*

$$\mathbf{p}^*(t) - \mathbf{p}(t) = \sum_{|\mathbf{k}|=s} \int_0^t p^*(\mathbf{k},\tau)\cdot(\mathbf{p}^*(\tau) - \mathbf{e}_{\mathbf{k}})\mathbf{E}_{\mathbf{A}}(\tau,t)\, dm_\infty(\tau) \tag{4.2}$$

where $\mathbf{E}_{\mathbf{A}}(\tau,t)$ is given by (2.7), the signed *measure $dm_\infty(\tau)$ is formally the derivative of m_∞ times $d\tau$, and*

$$dm_\infty(\tau) = \left(\sum_{\alpha\in C} \lambda_\alpha(\tau) - \mu_\alpha(\tau) m_\infty^\alpha(\tau) q_\alpha(\tau)\right) d\tau. \tag{4.3}$$

We now apply Theorem 4.1 to obtain bounds and inequalities. First, we obtain bounds by simply bounding the time-ordered exponential $\mathbf{E_A}(\tau, t)$ in (4.2) by 1. It may be possible to obtain more refined relations by more carefully examining the time ordered exponential, as we illustrate by example in Section 6 below. Recall that $|\mathbf{x}|$ is the ℓ_1-norm, defined in (3.1). Let $|dm_\infty|(\tau)$ be the measure

$$|dm_\infty|(\tau) \equiv \left|\frac{dm_\infty}{d\tau}(\tau)\right| d\tau. \tag{4.4}$$

Corollary 4.2 *In the setting of Theorem 4.2, We have the following bounds for the error due to the modified offered load approximation:*

$$\begin{aligned}
\sup_{0\le\tau\le t} |\mathbf{p}^*(\tau) - \mathbf{p}(\tau)| &\le 2 \cdot \sum_{|\mathbf{k}|=s} \int_0^t p^*(\mathbf{k}, \tau)(1 - p^*(\mathbf{k}, \tau))\, |dm_\infty|(\tau) && (4.5)\\
&\le 2 \cdot \int_0^t \beta_s(m_\infty(\tau)) \left(1 - \frac{\beta_s(m_\infty(\tau))}{\binom{s+|C|-1}{s}}\right) |dm_\infty|(\tau) && (4.6)\\
&\le 2 \cdot \int_0^t \beta_s(m_\infty(\tau))\, |dm_\infty|(\tau), && (4.7)
\end{aligned}$$

where $\beta_s(m_\infty(t))$ is given in (1.4), i.e.

$$\beta_s\Big(m_\infty(t)\Big) = \sum_{|\mathbf{k}|=s} p^*(\mathbf{k}, t). \tag{4.8}$$

Proof: The first bound follows from Theorem 4.1, the identity

$$|\mathbf{p}^*(t) - \mathbf{e_k}| = 2\,(1 - p^*(\mathbf{k}, t)) \tag{4.9}$$

for all $\mathbf{k} \in S_C$, and fact that $|\mathbf{E_A}(t)| = 1$, where $|\cdot|$ is an operator norm induced by the ℓ_1 norm on row vectors.

For the second inequality, we observe that $x(1-x)$ is a concave function of x, and $\binom{s+|C|-1}{s}$ equals the number of states $\mathbf{k}$ with $|\mathbf{k}| = s$. Now apply Jensen's inequality to the first bound. ∎

If $\mathbf{Q}_s(0)$ has a distribution that is *not* of the form (4.1), then we can construct a process $\tilde{\mathbf{Q}}_s$ that has the same infinitesimal generator, but an initial distribution of the proper form. We then have

$$\sup_{0\le\tau\le t} |\mathbf{p}^*(\tau) - \mathbf{p}(\tau)| \le |\mathbf{p}(0) - \tilde{\mathbf{p}}(0)| + \sup_{0\le\tau\le t} |\mathbf{p}^*(\tau) - \tilde{\mathbf{p}}(\tau)|, \tag{4.10}$$

where $\tilde{\mathbf{p}}$ is the probability vector for $\tilde{\mathbf{Q}}_s$, and now the above corollary applies.

5 MOL Bounds for the $M_t/M_t/s/0$ Queue

Now we restrict ourselves to one class or $|C| = 1$, which gives us the $M_t/M_t/s/0$ queue. It follows that $S_C(s) = \{0, 1, \ldots, s\}$, which is a totally ordered set. Moreover, (4.2) simplifies to

$$\mathbf{p}^*(t) - \mathbf{p}(t) = \int_0^t \beta_s(m_\infty(\tau)) \cdot (\mathbf{p}^*(\tau) - \mathbf{e}_s)\mathbf{E_A}(\tau, t)\, dm_\infty(\tau). \tag{5.1}$$

The next proposition establishes a stochastic comparison between the $M_t/M_t/s/0$ queue and its MOL approximation. (All proofs appear at the end of the section.) We say that a probability vector $\mathbf{p}_1$ is *stochastically dominated* by $\mathbf{p}_2$, and write $\mathbf{p}_1 \leq_{st} \mathbf{p}_2$, if

$$\sum_{j=k}^{s} p_1(j) \leq \sum_{j=k}^{s} p_2(j) \quad \text{for all } k = 0, 1, \ldots, s. \tag{5.2}$$

In terms of operators and componentwise ordering of vectors, $\mathbf{p}_1 \leq_{st} \mathbf{p}_2$ is equivalent to $\mathbf{p}_1\mathbf{K} \leq \mathbf{p}_2\mathbf{K}$, where $\mathbf{K} = (\mathbf{I} - \mathbf{L})^{-1}$ with $\mathbf{L}$ equalling the left shift operator on row vectors, or

$$\mathbf{K} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & 1 & \cdots & 1 & 0 \\ 1 & 1 & 1 & \cdots & 1 & 1 \end{bmatrix}. \tag{5.3}$$

Theorem 5.1 *For the $M_t/M_t/s/0$ system, if $m_\infty(0) \leq \lambda(0)/\mu(0)$ and λ/μ is an increasing function on $[0,t]$, then the modified offered load distribution is stochastically dominated by the exact distribution for Q_s on $[0,t]$, or*

$$\mathbf{p}^*(\tau) \leq_{st} \mathbf{p}(\tau), \quad \text{for all } \tau \in [0,t]. \tag{5.4}$$

In particular, $\beta_s(m_\infty)$ underestimates the actual blocking probability on $[0,t]$. Conversely, if $m_\infty(0) \geq \lambda(0)/\mu(0)$ and λ/μ is an decreasing function on $[0,t]$, then the exact distribution for $Q_s(t)$ is stochastically dominated by the the modified offered load distribution at time t and $\beta_s(m_\infty)$ overestimates the actual blocking probability on $[0,t]$.

In order to obtain better bounds on the error of MOL in the blocking probability, we focus on the ℓ_1-norm on cumulative distribution functions (cdf's) or, equivalently complementary cdf's, instead of probability mass functions. For any vector $\mathbf{x}$ on $\{0,1,\ldots,s\}$ we define $|\mathbf{x}|_{\mathbf{K}}$ to equal $|\mathbf{xK}|$, i.e., the ℓ_1-norm applied to tail sums. Recall that $\mathsf{E}[X] = \sum_{k=0}^{\infty} \mathsf{P}(X > k)$ if X in a non-negative integer valued random variable. Thus, is $\mathbf{p_1}$ and $\mathbf{p_2}$ are two probability vectors corresponding to $\{0,1,\ldots,s\}$-valued random variables X_1 and X_2, we have

$$\max\left(\left|\mathsf{P}\left(X_1 = s\right) - \mathsf{P}\left(X_2 = s\right)\right|,\ \left|\mathsf{E}\left[X_1\right] - \mathsf{E}\left[X_2\right]\right|\right) \leq |\mathbf{p_1} - \mathbf{p_2}|_{\mathbf{K}}. \tag{5.5}$$

Theorem 5.2 *For all $t \geq 0$ with $\mathbf{p}(0) = \mathbf{p}^*(0)$, we have*

$$\begin{aligned} &|\mathbf{p}^*(t) - \mathbf{p}(t)|_{\mathbf{K}} \\ &\quad \leq \int_0^t \beta_s\Big(m_\infty(\tau)\Big)\left(s - m_\infty(\tau)\Big(1 - \beta_s\Big(m_\infty(\tau)\Big)\Big)\right) \exp\left(-\int_\tau^t \mu(r)dr\right) |dm_\infty|(\tau). \end{aligned} \tag{5.6}$$

We now apply Theorem 5.2 to obtain bounds that hold for all time. Note that $s\beta_s(x) \to 0$ as $s \to \infty$.

Corollary 5.3 *In the setting of Theorem 5.2, if λ is a bounded function on $[0, \infty)$, and μ is a constant function, then*

$$\sup_{t \geq 0} |\mathbf{p}^*(t) - \mathbf{p}(t)|_{\mathbf{K}} \leq \frac{s}{\mu} \max\Big(2|\lambda|_\infty, |\lambda|_\infty + \mu m_\infty(0)\Big) \beta_s \Big(\max\Big(\frac{|\lambda|_\infty}{\mu}, m_\infty(0)\Big)\Big). \tag{5.7}$$

If in addition, λ is differentiable and its derivative λ' is bounded on $[0, \infty)$, then

$$\sup_{t \geq 0} |\mathbf{p}^*(t) - \mathbf{p}(t)|_{\mathbf{K}} \leq \frac{s}{\mu} \max\Big(|\lambda(0) - \mu m_\infty(0)|, \frac{|\lambda'|_\infty}{\mu}\Big) \beta_s \Big(\max\Big(\frac{|\lambda|_\infty}{\mu}, m_\infty(0)\Big)\Big). \tag{5.8}$$

We remark that Corollary 5.3 is not good for the blocking probabilities, because we can use stochastic comparisons to directly deduce with proper initial conditions the sharper bound $\mathsf{P}(Q_s(t) = s) \leq \beta_s(|\lambda|_\infty/\mu)$; e.g., by Theorem 10 of Whitt [18]. However Corollary 5.3 yields useful bounds for the mean, as stated in (1.7).

Proof of Theorem 5.1: The basis vector $\mathbf{e}_s$ is a probability vector for the point mass distribution of being in state s, which is the maximum probability distribution, with respect to stochastic dominance, on $\{0, 1, \ldots, s\}$. It follows that the probability vector $\mathbf{p}^*(t)$ is always stochastically dominated by $\mathbf{e}_s$. Now $\mathbf{A}(t)$ for fixed t, is the generator for a birth-death process, which is stochastically monotone. Using Theorem 7.5 of Massey [14], it follows that the probability vector $\mathbf{p}^*(\tau)\mathbf{E}_{\mathbf{A}}(\tau, t)$ is always stochastically dominated by $\mathbf{e}_s\mathbf{E}_{\mathbf{A}}(\tau, t)$ for all $0 \leq \tau \leq t$. After combining this result with (4.2), we will be done once we show that the derivative of m_∞ is non-negative (or non-positive) on $[0, t]$. This will follow from the lemma below. ∎

Lemma 5.4 *If $m_\infty(0) \leq \lambda(0)/\mu(0)$ and λ/μ is a right-continuous, increasing function on $[0, t]$ then m_∞ is increasing on $[0, t]$ also. Conversely, if $m_\infty(0) \geq \lambda(0)/\mu(0)$ and λ/μ is a right-continuous, function on $[0, t]$ then m_∞ is decreasing on $[0, t]$.*

Proof: Since $|C| = 1$, (3.4) becomes

$$\frac{d}{dt} m_\infty(t) = \lambda(t) - \mu(t)\, m_\infty(t), \tag{5.9}$$

and so

$$m_\infty(t) = m_\infty(0) \exp\left(-\int_0^t \mu(\tau) d\tau\right) + \int_0^t \lambda(\tau) \exp\left(-\int_\tau^t \mu(v) dv\right) d\tau. \tag{5.10}$$

Now let $\rho \equiv \lambda/\mu$. Since by hypothesis, ρ is right-continuous and of bounded variation, we can apply the integration by parts formula (see page 104 of Daley and Vere-Jones [2]), and get

$$\frac{1}{\mu(t)} \frac{d}{dt} m_\infty(t) = (\rho(0) - m_\infty(0)) \exp\left(-\int_0^t \mu(\tau) d\tau\right) + \int_0^t \exp\left(-\int_\tau^t \mu(v) dv\right) d\rho(\tau). \tag{5.11}$$

We now observe that the hypothesis gives precisely the conditions that makes the two summands above non-negative or non-positive on $[0, t]$. ∎

Proof of Theorem 5.2 and Corollary 5.3: Since μ is constant, we can write (5.10) as

$$\begin{aligned} m_\infty(t) &= m_\infty(0)e^{-\mu t} + \int_0^t \lambda(t-\tau)e^{-\mu\tau}\,d\tau && (5.12)\\ &= \int_t^\infty \mu\, m_\infty(0)e^{-\mu\tau}\,d\tau + \int_0^t \lambda(t-\tau)e^{-\mu\tau}\,d\tau && (5.13)\\ &\le \int_0^\infty \max(|\lambda|_\infty, \mu\, m_\infty(0))e^{-\mu\tau}\,d\tau && (5.14)\\ &\le \max(\frac{|\lambda|_\infty}{\mu}, m_\infty(0)). && (5.15) \end{aligned}$$

Combining this with (5.9) gives us

$$\left|\frac{dm_\infty}{dt}\right|_\infty \le |\lambda|_\infty + \mu\,|m_\infty|_\infty \le |\lambda|_\infty + \max(|\lambda|_\infty, \mu\, m_\infty(0)) = \max(2|\lambda|_\infty, |\lambda|_\infty + \mu\, m_\infty(0)). \tag{5.16}$$

When λ' exists and is bounded, we have by (5.11)

$$\begin{aligned} \left|\frac{dm_\infty}{dt}\right|_\infty &= \left|(\lambda(0) - \mu\, m_\infty(0))e^{-\mu t} + \int_0^t e^{-\mu\tau}\lambda'(t-\tau)\,d\tau\right| && (5.17)\\ &\le |\lambda(0) - \mu\, m_\infty(0)|e^{-\mu t} + \frac{|\lambda'|_\infty}{\mu}(1-e^{-\mu t}) && (5.18)\\ &\le \max(|\lambda(0) - \mu\, m_\infty(0)|, \frac{|\lambda'|_\infty}{\mu}). && (5.19) \end{aligned}$$

Finally, we apply the lemma below:

Lemma 5.5 *If $\mathbf{E_A}(t)$ is the transition probability matrix for an $M_t/M_t/s/0$ queue at time t, then for any two probability vectors $\mathbf{p}$ and $\mathbf{q}$ we have*

$$|(\mathbf{p}-\mathbf{q})\mathbf{E_A}(t)\mathbf{K}| \le |(\mathbf{p}-\mathbf{q})\mathbf{K}| \exp\left(-\int_0^t \mu(\tau)\,d\tau\right). \tag{5.20}$$

Proof: If $\mathbf{A}$ is the generator for an $M/M/s/0$ queue, then it has the form

$$\mathbf{A} = \begin{bmatrix} -\lambda & \lambda & 0 & \cdots & 0 & 0 \\ \mu & -(\lambda+\mu) & \lambda & \cdots & 0 & 0 \\ 0 & 2\mu & -(\lambda+2\mu) & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & -(\lambda+(s-1)\mu) & \lambda \\ 0 & 0 & 0 & \cdots & s\mu & -s\mu \end{bmatrix}. \tag{5.21}$$

Using right and left shift operators, we have

$$\mathbf{A} = \lambda\mathbf{R} + \mu\boldsymbol{\Delta}\mathbf{L} - \lambda\mathbf{R}\mathbf{L} - \mu\boldsymbol{\Delta} = (\lambda\mathbf{R} - \mu\boldsymbol{\Delta})(\mathbf{I}-\mathbf{L}). \tag{5.22}$$

Since $\mathbf{K} = (\mathbf{I}-\mathbf{L})^{-1}$, then

$$\mathbf{K}^{-1}\mathbf{A}\mathbf{K} = (\mathbf{I}-\mathbf{L})(\lambda\mathbf{R} - \mu\boldsymbol{\Delta}) = \lambda\mathbf{R} + \mu\mathbf{L}\boldsymbol{\Delta} - \lambda\mathbf{L}\mathbf{R} - \mu\boldsymbol{\Delta}. \tag{5.23}$$

Translating back into matrix form, we get

$$\mathbf{K}^{-1}\mathbf{A}\mathbf{K} = \begin{bmatrix} 0 & \lambda & 0 & \cdots & 0 & 0 \\ 0 & -(\lambda+\mu) & \lambda & \cdots & 0 & 0 \\ 0 & \mu & -(\lambda+2\mu) & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & -(\lambda+(s-1)\mu) & \lambda \\ 0 & 0 & 0 & \cdots & (s-1)\mu & -(\lambda+s\mu) \end{bmatrix} \tag{5.24}$$

Let $\mathbf{A}^*$ be the lower righthand $s \times s$ submatrix of $\mathbf{K}^{-1}\mathbf{A}\mathbf{K}$, namely

$$\mathbf{A}^* = \begin{bmatrix} -(\lambda+\mu) & \lambda & \cdots & 0 & 0 \\ \mu & -(\lambda+2\mu) & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & -(\lambda+(s-1)\mu) & \lambda \\ 0 & 0 & \cdots & (s-1)\mu & -(\lambda+s\mu) \end{bmatrix}. \tag{5.25}$$

The off-diagonal terms of $\mathbf{A}^*$ are non-negative, which makes $\exp(t\mathbf{A}^*)$ a non-negative matrix. Moreover, the row sums of $\mathbf{A}^*$ are all non-negative and so $\exp(t\mathbf{A}^*)$ is substochastic meaning its ℓ_1 operator norm is less than or equal to 1. Since $\mathbf{A}^* + \mu\mathbf{I}$ has these same properties, we get

$$|\exp(t\mathbf{A}^*)| \leq \exp(-\mu t). \tag{5.26}$$

Now observe that if $\mathbf{p}$ and $\mathbf{q}$ are probability vectors then the first (or zero-th) entry of the row vector $(\mathbf{p}-\mathbf{q})\mathbf{K}$ is 0. Hence $\mathbf{A}$ acts on $(\mathbf{p}-\mathbf{q})\mathbf{K}$ the same way that $\mathbf{A}^*$ acts on the non-zero, righthanded, s-dimensional subvector of $(\mathbf{p}-\mathbf{q})\mathbf{K}$. Taking norms, we get

$$\begin{aligned} |(\mathbf{p}-\mathbf{q})\exp(t\mathbf{A})\mathbf{K}| &\leq \left|(\mathbf{p}-\mathbf{q})\mathbf{K}\exp(t\mathbf{K}^{-1}\mathbf{A}\mathbf{K})\right| \\ &\leq |(\mathbf{p}-\mathbf{q})\mathbf{K}|\,|\exp(t\mathbf{A}^*)| \\ &\leq |(\mathbf{p}-\mathbf{q})\mathbf{K}|\exp(-\mu t). \end{aligned} \tag{5.27}$$

Now we consider the transition matrix $\prod_{i=1}^{n} \exp(t_i\mathbf{A}_i)$, where $\mathbf{A}_i$ is an $M/M/s/0$ generator for each i. By (5.27) and induction,

$$\left|(\mathbf{p}-\mathbf{q})\prod_{i=1}^{n}\exp(t_i\mathbf{A}_i)\,\mathbf{K}\right| \leq \left|(\mathbf{p}-\mathbf{q})\mathbf{K}\right| \exp\left(-\sum_{i=1}^{n}\mu_i t_i\right). \tag{5.28}$$

However, we can approximate $\mathbf{E}_{\mathbf{A}}(t)$ arbitrarily closely by $\prod_{i=1}^{n}\exp(t_i\mathbf{A}_i)$. This allows us to deduce (5.20). ∎

6 Example: Changing $M/M/s/0$ Rates in Midstream

Suppose we consider the case of $\lambda(t) = \lambda_+$ and $\mu(t) = \mu_+$ for all $t \geq 0$ and $\mathbf{p}(0) = \mathbf{p}^*(0)$ where $m_\infty(0) = \lambda_-/\mu_-$. The time-dependent behavior of Q_s is that of a stationary $M/M/s/0$ queue with rates λ_- and μ_- for all time $t < 0$, that suddenly switches to rates λ_+ and μ_+

for all time $t \geq 0$. We now want to compute an upper bound for the error between the transient distribution of Q_s and its MOL approximation. Now, in addition to Theorem 4.1, we exploit the fact that $\mathbf{E}_{\mathbf{A}}(t) = \exp(\mathbf{A}t)$, where $\mathbf{A}$ is the infinitesimal generator of the $M/M/s/0$ queue with parameters λ_+ and μ_+. In this case, (5.1) becomes

$$\mathbf{p}^*(t) - \mathbf{p}(t) = \int_0^t \beta_s\Big(m_\infty(\tau)\Big) \cdot (\mathbf{p}^*(\tau) - \mathbf{e}_s) \exp((t-\tau)\mathbf{A})\, dm_\infty(\tau). \tag{6.1}$$

If we let $\rho_+ \equiv \lambda_+/\mu_+$ and $\rho_- \equiv \lambda_-/\mu_-$, then

$$m_\infty(\tau) = \rho_- \exp(-\mu_+\tau) + \rho_+(1 - \exp(-\mu_+\tau)) \tag{6.2}$$

and

$$\frac{d}{d\tau} m_\infty(\tau) = \mu_+(\rho_+ - \rho_-)\exp(-\mu_+\tau). \tag{6.3}$$

Since Q_s is reversible, see page 32 of Keilson [10], the generator $\mathbf{A}$ is diagonally similar to a symmetric matrix. By the spectral decomposition theorem, we have

$$\exp(t\mathbf{A}) = \mathbf{1}^T\boldsymbol{\pi} + \sum_{j=1}^{s} \exp(-\sigma_j t)\boldsymbol{\Delta}(\sqrt{\boldsymbol{\pi}})^{-1}\mathbf{x}_j^T\mathbf{x}_j\boldsymbol{\Delta}(\sqrt{\boldsymbol{\pi}}), \tag{6.4}$$

where $\boldsymbol{\pi}$ is the steady state probability vector for $\mathbf{A}$ such that $\boldsymbol{\pi}\mathbf{A} = \mathbf{0}$, $\sqrt{\boldsymbol{\pi}}$ is the positive vector whose components are the square roots of the components of $\boldsymbol{\pi}$, and $\boldsymbol{\Delta}(\sqrt{\boldsymbol{\pi}})$ is the corresponding diagonal matrix. The negatives of the $s+1$ real numbers $0 < \sigma_1 < \cdots < \sigma_s$ are the eigenvalues for $\mathbf{A}$. Finally, $\{\sqrt{\boldsymbol{\pi}}, \mathbf{x}_1, \ldots, \mathbf{x}_s\}$ is the corresponding set of orthonormal eigenvectors for $\boldsymbol{\Delta}(\sqrt{\boldsymbol{\pi}})\mathbf{A}\boldsymbol{\Delta}(\sqrt{\boldsymbol{\pi}})^{-1}$. The eigenvalues and eigenvectors for this model are readily obtained, as shown by Lederman and Reuter [12] and Karlin and McGregor [9]. In this case the orthogonal polynomials are the Poisson-Charlier polynomials, also see Jagerman [8].

Since $\mathbf{p}^*(\tau)$ and $\mathbf{e}_s$ are probability vectors, we have

$$(\mathbf{p}^*(\tau) - \mathbf{e}_s)\mathbf{1}^T = 0. \tag{6.5}$$

Using the Cauchy-Schwartz inequality, we have

$$|\mathbf{x}_j\boldsymbol{\Delta}(\sqrt{\boldsymbol{\pi}})| \leq 1 \quad \text{and} \quad |\mathbf{x}_j\boldsymbol{\Delta}(\sqrt{\boldsymbol{\pi}})^{-1}| \leq \sqrt{\sum_{k=0}^{s}\frac{1}{\pi_k}}. \tag{6.6}$$

Hence, we get

$$|(\mathbf{p}^*(\tau) - \mathbf{e}_s)\cdot\exp((t-\tau)\mathbf{A})| \leq 2\cdot\sqrt{\sum_{k=0}^{s}\frac{1}{\pi_k}}\cdot\sum_{j=1}^{s}\exp(-\sigma_j(t-\tau)). \tag{6.7}$$

Combining all of these results, we obtain

$$|\mathbf{p}^*(t) - \mathbf{p}(t)| \leq 2\,\mu_+\,\beta_s(\overline{\rho})\,|\rho_+ - \rho_-|\cdot\sqrt{\sum_{k=0}^{s}\frac{1}{\pi_k}}\cdot\sum_{j=1}^{s}\frac{\exp(-\sigma_j t) - \exp(-\mu_+ t)}{\mu_+ - \sigma_j}, \tag{6.8}$$

where $\overline{\rho} = \max(\rho_+, \rho_-)$. Moreover, we get from the other bounds

$$
\begin{aligned}
|(\mathbf{p}^*(t) - \mathbf{p}(t))\mathbf{K}| &\leq \mu_+|\rho_+ - \rho_-|e^{-\mu_+ t}\int_0^t \beta_s\Big(m_\infty(\tau)\Big)\Big(s - m_\infty(\tau)\Big(1 - \beta_s\Big(m_\infty(\tau)\Big)\Big)\Big)d\tau \\
&\leq \mu_+|\rho_+ - \rho_-|te^{-\mu_+ t}\beta_s(\overline{\rho})(s - \underline{\rho}(1 - \beta_s(\underline{\rho})))
\end{aligned}
$$

where $\underline{\rho} = \min(\rho_+, \rho_-)$.

In addition to these error estimates, we get the following stochastic dominance results by applying Proposition 5.1,

$$\rho_- \leq \rho_+ \implies \mathbf{p}^*(t) \leq_{st} \mathbf{p}(t) \text{ for all } t \geq 0, \tag{6.9}$$

and

$$\rho_- \geq \rho_+ \implies \mathbf{p}^*(t) \geq_{st} \mathbf{p}(t) \text{ for all } t \geq 0. \tag{6.10}$$

7 Proving the Main Theorem

Lemma 7.1 *If* $\mathbf{x} = \sum_{\alpha\in C} x_\alpha \mathbf{e}_\alpha$ *and we define*

$$\pi(\mathbf{k}, \mathbf{x}) \equiv \frac{\mathbf{x}^\mathbf{k}}{\mathbf{k}!} \Big/ \sum_{j=0}^{s} \frac{|\mathbf{x}|^j}{j!}, \tag{7.1}$$

then it follows that

$$\frac{\partial}{\partial x_\alpha}\pi(\mathbf{k}, \mathbf{x}) = \pi(\mathbf{k} - \mathbf{e}_\alpha, \mathbf{x}) \cdot \operatorname{sgn}(k_\alpha) - \pi(\mathbf{k}, \mathbf{x})\left(1 - \sum_{|\mathbf{j}|=s} \pi(\mathbf{j}, \mathbf{x})\right) \tag{7.2}$$

and

$$\pi(\mathbf{k}, \mathbf{x}) \cdot x_\alpha = \pi(\mathbf{k} + \mathbf{e}_\alpha, \mathbf{x}) \cdot (k_\alpha + 1). \tag{7.3}$$

Proof of Theorem 4.1: We first observe that $p^*(\mathbf{k}, t) = \pi(\mathbf{k}, \mathbf{m}_\infty(t))$ and

$$\beta_s\Big(m_\infty(t)\Big) = \sum_{|\mathbf{k}|=\mathbf{s}} \pi(\mathbf{k}, \mathbf{m}_\infty(t)). \tag{7.4}$$

We then apply the identities of Lemma 7.1 and get

$$
\begin{aligned}
\frac{d}{dt}p^*(\mathbf{k}, t) &= \sum_{\alpha\in C}\Big[p^*(\mathbf{k} - \mathbf{e}_\alpha, t) \cdot \operatorname{sgn}(k_\alpha) - p^*(\mathbf{k}, t) \cdot \Big(1 - \beta_s\Big(m_\infty(t)\Big)\Big)\Big] \cdot \frac{d}{dt}m_\infty^\alpha(t) \\
&= \sum_{\alpha\in C} p^*(\mathbf{k} - \mathbf{e}_\alpha, t) \cdot \operatorname{sgn}(k_\alpha) \cdot \frac{d}{dt}m_\infty^\alpha(t) - p^*(\mathbf{k}, t) \cdot \Big(1 - \beta_s\Big(m_\infty(t)\Big)\Big) \cdot \frac{d}{dt}m_\infty(t) \\
&= \sum_{\alpha\in C} p^*(\mathbf{k} - \mathbf{e}_\alpha, t) \cdot \operatorname{sgn}(k_\alpha) \cdot \left(\lambda_\alpha(t) + \sum_{\beta\in C}\mu_\beta(t)p_{\beta\alpha}(t)m_\infty^\beta(t) - \mu_\alpha(t)m_\infty^\alpha(t)\right) \\
&\quad - p^*(\mathbf{k}, t) \cdot \sum_{\alpha\in C}\lambda_\alpha(t) - \mu_\alpha(t)q_\alpha(t)m_\infty^\alpha(t) + p^*(\mathbf{k}, t) \cdot \beta_s\Big(m_\infty(t)\Big) \cdot \frac{d}{dt}m_\infty(t)
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{\alpha\in C}\Big[\lambda_\alpha(t)\mathrm{sgn}(k_\alpha)p^*(\mathbf{k}-\mathbf{e}_\alpha,t)+\mu_\alpha(t)(k_\alpha+1)q_\alpha(t)p^*(\mathbf{k}+\mathbf{e}_\alpha,t)\\
&\quad+\sum_{\beta\in C}\mu_\beta(t)(k_\beta+1)p_{\beta\alpha}(t)\mathrm{sgn}(k_\alpha)p^*(\mathbf{k}-\mathbf{e}_\alpha+\mathbf{e}_\beta,t)\\
&\quad-\mu_\alpha(t)k_\alpha p^*(\mathbf{k},t)-\lambda_\alpha(t)p^*(\mathbf{k},t)\Big]+p^*(\mathbf{k},t)\cdot\beta_s\big(m_\infty(t)\big)\cdot\frac{d}{dt}\mathbf{m}_\infty(\mathbf{t})\\
&= \sum_{\alpha\in C}\Big[\lambda_\alpha(t)\mathrm{sgn}(k_\alpha)p^*(\mathbf{k}-\mathbf{e}_\alpha,t)+\mu_\alpha(t)(k_\alpha+1)q_\alpha(t)p^*(\mathbf{k}+\mathbf{e}_\alpha,t)\\
&\quad+\sum_{\beta\in C}\mu_\beta(t)(k_\beta+1)p_{\beta\alpha}(t)\mathrm{sgn}(k_\alpha)p^*(\mathbf{k}-\mathbf{e}_\alpha+\mathbf{e}_\beta,t)\\
&\quad-(\lambda_\alpha(t)+\mu_\alpha(t)k_\alpha)p^*(\mathbf{k},t)\Big]+p^*(\mathbf{k},t)\cdot\beta_s\big(m_\infty(t)\big)\cdot\frac{d}{dt}m_\infty(t).
\end{aligned}
$$

The above relation holds for all $\mathbf{k}$, but we can also write it as

$$
\begin{aligned}
\frac{d}{dt}p^*(\mathbf{k},t) \;=\; &\sum_{\alpha\in C}\Big[\lambda_\alpha(t)\mathrm{sgn}(k_\alpha)p^*(\mathbf{k}-\mathbf{e}_\alpha,t)\\
&+\sum_{\beta\in C}\mu_\beta(t)(k_\beta+1)p_{\beta\alpha}(t)\mathrm{sgn}(k_\alpha)p^*(\mathbf{k}-\mathbf{e}_\alpha+\mathbf{e}_\beta,t)-\mu_\alpha(t)k_\alpha p^*(\mathbf{k},t)\Big]\\
&-p^*(\mathbf{k},t)\cdot\Big(1-\beta_s\big(m_\infty(t)\big)\Big)\cdot\frac{d}{dt}m_\infty(t).
\end{aligned}
$$

This last equation resembles the forward equations for $p(\mathbf{k},t)$ when $|\mathbf{k}|=s$. Recasting these results in operator form gives us

$$
\frac{d}{dt}\mathbf{p}^*(t)=\mathbf{p}^*(t)\mathbf{A}(t)+\frac{d}{dt}m_\infty(t)\cdot\sum_{|\mathbf{k}|=s}p^*(\mathbf{k},t)(\mathbf{p}^*(t)-\mathbf{e}_{\mathbf{k}}). \tag{7.5}
$$

We can then write the solution to this inhomogeneous ordinary differential equation as

$$
\mathbf{p}^*(t)=\mathbf{p}^*(0)\mathbf{E}_{\mathbf{A}}(t)+\int_0^t\frac{d}{d\tau}m_\infty(\tau)\cdot\sum_{|\mathbf{k}|=s}p^*(\mathbf{k},\tau)(\mathbf{p}^*(\tau)-\mathbf{e}_{\mathbf{k}})\mathbf{E}_{\mathbf{A}}(\tau,t)d\tau. \tag{7.6}
$$

Combining the above with (2.5) yields the desired (4.2). ∎

References

[1] Brown, M. and Ross, S. M. (1969), Some Results for Infinite Server Poisson Queues, *Journal of Applied Probability*, **6**, 604–611.

[2] Daley, D.J. and Vere-Jones, D. (1988), *An Introduction to the Theory of Point Processes*, Springer, New York.

[3] Davis, J., Massey, W. and Whitt, W. (1994), Sensitivity to the Service-Time Distribution in the Nonstationary Erlang Loss Model. *Management Science*, to appear.

[4] Dollard, J. D. and Friedman, C. N. (1979), *Product Integration with Applications to Differential Equations*, Encyclopedia of Mathematics and its Applications, **10**, Addison-Wesley, Reading MA.

[5] Eick, S., Massey, W. A., and Whitt, W. (1993), The Physics of the $M_t/G/\infty$ Queue, *Operations Research*, 41, 731–742.

[6] Eick, S., Massey, W. A., and Whitt, W. (1993), Infinite Server Queues with Sinusoidal Input, *Management Science*, 39, 241–252.

[7] Erlang, A. K. (1918), Solutions of Some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges, *The Post Office Electrical Engineers' Journal*, **10**, 189–197. (Translated from the 1917 article in Danish in *Elektroteknikeren* vol. 13).

[8] Jagerman, D. L. (1975), Nonstationary Blocking in Telephone Traffic, *Bell System Technical Journal*, **54**, 625–661.

[9] Karlin, S. and McGregor, J.L. (1957), The Differential Equations of Simple Birth-and-Death Process and the Stieljes Moment Problem. *Transactions of the American Mathematical Society*, **85**, 321–369.

[10] Keilson, J. (1979), *Markov Chain Models - Rarity and Exponentiality*, Springer-Verlag, New York.

[11] Khintchine, A. Y. (1955), Mathematical Methods in the Theory of Queueing, *Trudy Mat. Inst. Steklov* **49**, (in Russian). (English translation by Charles Griffin and Co., London, 1960).

[12] Ledermann, W. and Reuter, G.E.H. (1954), Spectral Theory for the Differential Equations of Simple Birth-and-Death Process. *Philosophical Transactions of the Royal Society of London, Series A*, **246**, 321–369.

[13] Massey, W. A. (1984), Open Networks of Queues: Their Algebraic Structure and Estimating Their Transient Behavior, *Advances in Applied Probability*, **16**, 176–201.

[14] Massey, W. A. (1987), Stochastic Orderings for Markov Processes on Partially Ordered Spaces, *Mathematics of Operations Research*, **12**, 350–367.

[15] Massey, W. A. and Whitt, W. (1993). Networks of Infinite-Server Queues with Non-stationary Poisson Input, *Queueing Systems*, 13, 183–250.

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

[17] Prékopa, A. (1958), On Secondary Processes Generated by a Random Point Distribution of Poisson Type, *Annales Univ. Sci. Budapest de Eötvös Nom. Sectio Math.* **1**, 153–170.

[18] Whitt, W. (1981) Comparing Counting Processes and Queues, *Advances in Applied Probability*, **13**, 207–220.