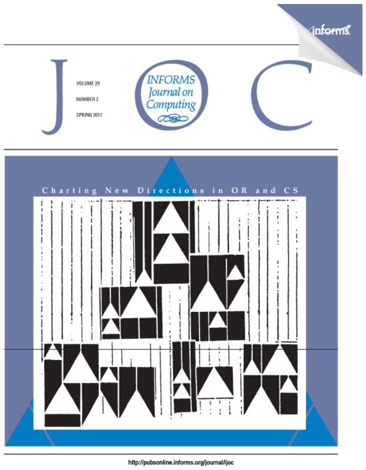


INFORMS Journal on Computing

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Technical Note: The $MAP_t/Ph_t/\#$ Queueing System and Multiclass $[MAP_t/Ph_t/\#]^K$ Queueing Network

Ira Gerhardt, Barry L. Nelson, Michael R. Taaffe

INFORMS JOURNAL ON COMPUTING
Vol. 29, No. 2, Spring 2017, pp. 367–376
ISSN 1091-9856 (print), ISSN 1526-5528 (online)
http://pubsonline.informs.org/journal/ijoc/

Technical Note: The $MAP_t/Ph_t/\infty$ Queueing System and Multiclass $[MAP_t/Ph_t/\infty]^K$ Queueing Network

Ira Gerhardt,[a] **Barry L. Nelson,**[b] **Michael R. Taaffe**[c]

[a] Department of Mathematics, Manhattan College, Riverdale, New York 10471; [b] Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois 60208; [c] Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, Virginia 24061

Contact: ira.gerhardt@manhattan.edu (IG); nelsonb@northwestern.edu (BLN); taaffe@vt.edu (MRT)

Received: April 1, 2016
Revised: September 1, 2016
Accepted: September 13, 2016
Published Online: April 25, 2017

http://dx.doi.org/10.1287/ijoc.2016.0736

Abstract. In this paper we demonstrate how a key adjustment to known numerically exact methods for evaluating time-dependent moments of the number of entities in the $Ph_t/Ph_t/\infty$ queueing system and $[Ph_t/Ph_t/\infty]^K$ queueing network may be implemented to capture the effect of autocorrelation that may be present in arrivals to the more general $MAP_t/Ph_t/\infty$ queueing system and multiclass $[MAP_t/Ph_t/\infty]^K$ queueing network. The MAP_t is more general than the Ph_t arrival process in that it allows for stationary *nonrenewal* point processes, as well as the time-dependent generalization of nonrenewal point processes. Modeling real-world systems with bursty arrival processes such as those in telecommunications and transportation, for example, necessitate the use of nonrenewal processes. Finally, we show that the covariance of the number of entities at different nodes and times may be described by a single closed differential equation.

History: Accepted by Winfried K. Grassmann, Area Editor for Computational Probability and Analysis.
Funding: This research was partially supported by the National Science Foundation [Grants DMI-9622065 and DMI-9821011].
Supplemental Material: The online supplement is available at https://doi.org/10.1287/ijoc.2016.0736.

Keywords: queues • queueing • algorithms • phase-type distribution • nonstationary processes • infinite server • *MAP*s • queueing networks • time-dependent • transient

1. Introduction

A primary difficulty in analyzing steady-state behavior of stationary non-product-form networks is that the nodal departure processes are *not* renewal processes in general. When we consider arrival and service processes at a queueing node that have time-varying parameters, yielding time-dependent queueing networks, the composite arrival processes at a node are also, obviously, nonrenewal. As a result numerical analysis is exceedingly difficult in almost all time-dependent queueing system models. An exception is the $[M_t/M_t/\infty]^K$ network (the superscript indicates the number of nodes in the network), which has *closed* moment differential equations (Whitt and Massey 1993).

In Nelson and Taaffe (2004a, b) we showed that the $Ph_t/Ph_t/\infty$ system possesses characteristics such that evaluation of time-dependent performance measures are easy to compute numerically, and computational effort increases only linearly with the product of the number of arrival-process and service-process phases (independent of the number of states, which is infinite). In a *network* of such nodes the computational requirements increase only linearly with the product of the total number of network arrival phases and service phases—nodal or network capacity does *not* affect the computational demands. The analysis in these papers used moment, partial-moment, and marginal-moment differential equations and showed that the time-dependent moments of the number of entities in the system (at a node or in the network) and the moments and time-dependent distribution of the virtual sojourn times can be evaluated *without* explicit knowledge of any state probabilities. Therefore, the cardinality of the state space (which is infinite) has *no* effect on the computation effort.

Although Ph distributions are dense on the space of all distributions having support on the nonnegative real numbers (Asmussen 1987), they are limited in their ability to represent arrival and composite-arrival processes in general queueing networks where these processes are nonrenewal. Similarly, there is a limit on the variety of time-dependent point processes that a Ph_t process can approximate. Real-world studies of systems in manufacturing, transportation, and telecommunication networks have brought to light that standard assumptions regarding independence of interarrival times may actually be inappropriate (Casale et al. 2010a). More realistic models need to involve processes with nonnegligible dependence structures (i.e., nonzero autocovariance and autocorrelation; Asmussen 2000). Therefore, a mechanism is needed to efficiently represent nonrenewal arrival processes with a variety of marginal-distribution shapes to (eventually) construct highly accurate and efficient algorithms and approximations of composite-arrival processes for use in general stationary queueing network models. Similarly, there is a need for an efficient time-dependent generalization of the Ph_t process for use in time-dependent queueing network models.

In this technical note we develop the time-dependent generalization of the well-known stationary *Markov arrival process* (MAP)—a *nonrenewal* point process—using it as the arrival process to an infinite-server, time-dependent, phase-type-service queue and then in a network of such queues with (perhaps) multiple customer classes.

Several results we present in this paper are analogous to those in Nelson and Taaffe (2004a, b) for the $[Ph_t/Ph_t/\infty]^K$ queueing network. In fact some differential-equation results in this paper are identical to results in the previous papers and are only briefly discussed. However, we demonstrate numerically that behavior of the $[Ph_t/Ph_t/\infty]^K$ queueing network and the $[MAP_t/Ph_t/\infty]^K$ queueing network can be dramatically different. Although the qualitative effect of autocorrelation on queueing network behavior has been known for some time, the contribution of this note is to provide a method to quantify it for a useful class of queueing networks. A second key contribution of this paper is we show that the covariance of the number of entities at (potentially) different nodes and times for the $[MAP_t/Ph_t/\infty]^K$ queueing network is described by a closed, finite system of ordinary differential equations. This result is new.

The remainder of this paper is organized as follows: We first define the MAP_t process and a representation of it that is particularly useful in our work. We next describe the $MAP_t/Ph_t/\infty$ queueing system and introduce the necessary adjustment to the numerically exact method for evaluating the time-dependent performance measures introduced in Nelson and Taaffe (2004a, b) to this more general system. We utilize a numerical example to quantify the effect of introducing nonzero autocorrelation into the arrival stream as well as the effects of adjusting nonstationarity in the arrival process and variation in service. We close the paper with a brief analysis of the time-dependent covariance of the number of entities at a node (or in the network) at a fixed point of time s and the number of entities at that same node or some other node (or in the network) at point of time t, $t \geq s$. This last result can be useful in constructing approximations for departure point processes.

2. MAPs

Lucantoni (1991), Takine and Hasegawa (1994), Asmussen and Koole (1993), and others develop the *Markovian arrival process* as a generalization of the phase-type, or Ph, renewal process, originally proposed by Neuts (1979). The Ph process is a stochastic point process described by a continuous-time finite-phase Markov process having exactly one absorbing phase, with distribution of initial phase being independent of any history of the Ph-type process. Our definition of MAP generalizes the Ph process to allow for one or more absorbing phases, with the distribution of initial phase dependent upon the absorbing phase most recently visited. Let $m_A < \infty$ and $v_A \geq 1$ represent the number of transient and absorbing phases, respectively, of the embedded discrete-time Markov chain (DTMC). Our representation of the MAP characterizes the DTMC along with a vector of transition rates and a matrix of the initial transient phase probabilities. This representation is analogous to that of the Ph process presented in Nelson and Taaffe (2004b).

We let $\mathbf{A}$ denote the one-step transition probability matrix of the embedded DTMC:

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_0 & \mathbf{A}_2 \\ \boldsymbol{\alpha} & \mathbf{0} \end{pmatrix}.$$

The $m_A \times m_A$ matrix $\mathbf{A}_1$ represents the one-step transition probabilities between the m_A transient phases, whereas the $m_A \times v_A$ matrix $\mathbf{A}_2$ represents the one-step transition probabilities from the m_A transient phases to the v_A absorbing phases. For $i = 1, 2, \ldots, m_A$, if we let a_{ij} represent the one-step transition probability from transient phase i to either transient or absorbing phase j, then for $j = 1, 2, \ldots, m_A$ we set $(\mathbf{A}_1)_{ij} = a_{ij}$ and for $j = v_A + 1, v_A + 2, \ldots, v_A + m_A$ we set $(\mathbf{A}_2)_{i,\, j-v_A} = a_{ij}$. Notice that "absorbing phase" is really a misnomer in this representation because rather than being absorbed, the process is reinitialized for the next interevent time by $v_A \times m_A$ initial probability matrix $\boldsymbol{\alpha}$.

We further define the $(m_A + v_A) \times 1$ vector $\boldsymbol{\lambda}$, whose jth argument is λ_j, the nonnegative rate corresponding to phase j, for $j = 1, 2, \ldots, m_A + v_A$. We use the convention $\lambda_{m_A+h} = \infty$, for $h = 1, 2, \ldots, v_A$, corresponding to an instantaneous sojourn time in any absorbing phase. Thus, our MAP representation is the pair $(\mathbf{A}, \boldsymbol{\lambda})$.

To generalize our parameterization of the MAP to the time-dependent MAP_t process, we make all nonzero and noninfinite rates and one-step transitions probabilities integrable functions of time and add the argument "(t)." Thus, we represent the MAP_t as the pair $(\mathbf{A}(t), \boldsymbol{\lambda}(t))$.

Lucantoni (1991) and other authors define the MAP using notation that is slightly different from ours. Their definition of MAP utilizes a single absorbing phase; however, the distribution of initial phase is now dependent upon the last transient phase visited prior to absorption. The Lucantoni MAP representation is the set of $m_A \times m_A$ matrices $(\mathbf{D}_0, \mathbf{D}_1)$ such that $(\mathbf{D}_l)_{jh}$ is the transition rate from transient phase j to transient phase h upon an arrival of size l, for $l = 0, 1$ and $j, h = 1, 2, \ldots, m_A$. Notice we can construct the Lucantoni representation directly from our representation $(\mathbf{A}, \boldsymbol{\lambda})$ and vice versa. We leave the details of the two-way translation for the online supplement accompanying this journal.

Both definitions correctly describe the same MAP. Similarly, the number of Kolmogorov forward equations (KFEs) describing the queueing process of $MAP_t/Ph_t/\infty$ nodes and $[MAP_t/Ph_t/\infty]^K$ networks is the same regardless of the choice of representation since absorbing phases are instantaneous and therefore do not require KFEs. However, key differences between the two representations are worth mentioning. First, the Lucantoni representation explicitly describes the stochastic process $\{(N(t), J(t)); t \geq 0\}$, where $N(t)$ is the number of arrivals triggered by absorption by time $t \geq 0$ and $J(t)$ is the current phase of the underlying Markov chain. Notice the Lucantoni process has infinite state space, whereas our representation, characterizing transitions in the embedded DTMC, has a (typically finite) space consisting of m_A transient and v_A absorbing phases, respectively. Second, a large body of literature exists that describes techniques for specifying MAPs to capture various properties of stationary point processes. One reason for this is that stationary MAPs are *weakly dense* in the space of all stationary point processes (Asmussen and Koole 1993), so technically MAPs are sufficiently general to serve as approximations for any stationary point process. An additional justification for this prolific body of work is that approximating point processes with MAPs tends to yield queueing models that are analytically tractable. Techniques exist for fitting computationally friendly MAPs including Markov-modulated Poisson processes (MMPPs; see Ferng and Chang 2001, Fischer and Meier-Hellstern 1993, Heffes 1980); general MAPs with exactly two transient phases ($MAP(2)$, e.g., Diamond and Alfa 2000, Horváth and Telek 2006); and nonrenewal analogs of Ph-type renewal processes (e.g., Bitran and Dasu 1993, Johnson 1998)—to name a few—to various properties of a stationary point process or a sample of data. Other fitting techniques utilize Kronecker products for MAP representations (Casale et al. 2010b) or sequentially fit the marginal Ph process and autocorrelation sequence (Horváth et al. 2005). Additional sources of MAP-based methods and applications can be found in Artalejo et al. (2010), Liu and Neuts (1991), Narayana and Neuts (1992), Neuts et al. (1992). Important in this is that these fitting techniques target either properties of the interevent time or properties of the counting process; in the former case a MAP representation that describes transitions in the embedded DTMC, such as ours, would prove more useful, whereas for techniques that target the counting process the Lucantoni representation may be more appropriate.

3. The $MAP_t/Ph_t/\infty$ System

In this section we define the $MAP_t/Ph_t/\infty$ queueing system and present the KFEs (and resulting MDEs) guiding the trajectory of the time-dependent system-state probabilities. The arrival process was defined in Section 2. The service process is parameterized in exactly the same manner as the $Ph_t/Ph_t/\infty$ service process in Nelson and Taaffe (2004a, b).

For completeness we repeat the definition of the $(m_B + 1)$-dimensional Ph_t service process here.

Let

$$\mathbf{B}(t) = \begin{pmatrix} \mathbf{B}_1(t) & \mathbf{B}_2(t) \\ \boldsymbol{\beta}(t)^T & 0 \end{pmatrix},$$

where

$$\mathbf{B}_1(t) = \begin{pmatrix} b_{11}(t) & b_{12}(t) & \cdots & b_{1m_B}(t) \\ b_{21}(t) & b_{22}(t) & \cdots & b_{2m_B}(t) \\ \vdots & \vdots & \vdots & \vdots \\ b_{m_B 1}(t) & b_{m_B 2}(t) & \cdots & b_{m_B m_B}(t) \end{pmatrix}$$

is the underlying Markov chain one-step transition matrix for transient-to-transient phase transitions, and

$$\mathbf{B}_2(t) = \begin{pmatrix} b_{1,m_B+1}(t) \\ b_{2,m_B+1}(t) \\ \vdots \\ b_{m_B,m_B+1}(t) \end{pmatrix}$$

is the vector of transition probabilities from transient phases to the instantaneous absorbing phase, $m_B + 1$, which represents a service completion (a departure from the queue). The vector $\boldsymbol{\beta}(t) = [\beta_1(t), \beta_2(t), \ldots, \beta_{m_B}(t)]^T$ contains the initial service-phase probabilities for an entity completing its arrival process.

Let $\boldsymbol{\mu}(t) = [\mu_1(t), \mu_2(t), \ldots, \mu_{m_B}(t)]^T$ be the vector of real-valued integrable service-rate functions for the transient phases of the service process so that $[\boldsymbol{\mu}(t)^T, \infty]$ is the $(m_B + 1)$-dimensional rate vector for the entire phase service process.

The state of the $MAP_t/Ph_t/\infty$ process at time t is given by

$$[\mathbf{N}(t), A(t)] = [\{N_1(t), N_2(t), \ldots, N_{m_B}(t)\}, A(t)],$$

where $\{A(t); t \geq 0\}$ is the arrival phase of the next arrival to the system at time t, with $A(t) \in \{1, 2, \ldots, m_A\}$ and where $\{N_i(t); t \geq 0\}$ for $i = 1, 2, \ldots, m_B$ is the random process representing the number of entities who, at time t, are in the ith phase of their service. Therefore, the total number of entities in service at time t is $N(t) = \sum_{i=1}^{m_B} N_i(t)$. The instantaneous absorbing phases in the arrival and service processes need not be explicitly represented.

Naturally the state space for $MAP_t/Ph_t/\infty$ system is identical to the state space for the $Ph_t/Ph_t/\infty$ system; however, the KFEs for these two models are *not* the same. As in Nelson and Taaffe (2004b), if we let

$$\mathrm{P}(t;n_1,n_2,\ldots,n_{m_B},k) \equiv \mathrm{P}(N_1(t)=n_1,\ldots,N_{m_B}(t)=n_{m_B},A(t)=k)$$

and

$$\mathrm{P}(t;n_1,n_2,\ldots,n_{m_B},k)' \equiv \frac{d}{dt}\mathrm{P}(N_1(t)=n_1,\ldots,N_{m_B}(t)=n_{m_B},A(t)=k),$$

then we can show

$$\begin{aligned}
\mathrm{P}(t;&n_1,n_2,\ldots,n_{m_B},k)' \\
&= -[1-a_{kk}(t)]\lambda_k(t)\mathrm{P}(t;n_1,n_2,\ldots,n_{m_B},k) \\
&\quad -\sum_{l=1}^{m_B} n_l\mu_l(t)[1-b_{ll}(t)]\mathrm{P}(t;n_1,n_2,\ldots,n_{m_B},k) \\
&\quad +\sum_{l=1}^{m_A}\lambda_l(t)\left(\sum_{j=1}^{v_A}a_{l,m_A+j}(t)\alpha_{j,k}(t)\right) \\
&\quad \cdot\left\{\sum_{h=1}^{m_B}\mathrm{I}_{[n_h>0]}\beta_h(t)\mathrm{P}(t;n_1,\ldots,n_h-1,\ldots,n_{m_B},k)\right\} \\
&\quad +\sum_{\substack{l=1\\l\neq k}}^{m_A}a_{lk}(t)\lambda_l(t)P(t;n_1,n_2,\ldots,n_{m_B},l) \\
&\quad +\sum_{l=1}^{m_B}b_{l,m_B+1}(t)[n_l+1]\mu_l(t) \\
&\quad \cdot\mathrm{P}(t;n_1,\ldots,n_l+1,\ldots,n_{m_B},k) \\
&\quad +\sum_{l=1}^{m_B}\mathrm{I}_{[n_l>0]}\left\{\sum_{\substack{h=1\\h\neq l}}^{m_B}b_{hl}(t)[n_h+1]\mu_h(t)\right. \\
&\quad \left.\cdot\mathrm{P}(t;n_1,\ldots,n_l-1,\ldots,n_h+1,\ldots,n_{m_B},k)\right\}, \qquad (1)
\end{aligned}$$

for $k=1,2,\ldots,m_A$, $n_h=0,1,2,\ldots,\infty$, $h=1,2,\ldots,m_B$, and $t\geq 0$, and where for all t,

$$\sum_{k,n_1,\ldots,n_{m_B}}\mathrm{P}(t;n_1,n_2,\ldots,n_{m_B},k)=1.$$

We let $\mathrm{I}_{[a>0]}$ be the indicator function for the event "$a>0$."

We now describe the closed system of moment differential equations (MDEs) and partial-moment differential equations (PMDEs) that we utilize in calculating $\mathrm{E}[N(t)]$ and $\mathrm{Var}[N(t)]$ for the $MAP_t/Ph_t/\infty$ queueing system at all times $t\geq 0$. As in Nelson and Taaffe (2004b), an MDE or PMDE is *closed* if it contains no state probabilities on the right-hand side; closedness is necessary to calculate the time-dependent moments or partial moments without having to compute the time-dependent state probabilities.

Theorem 1. *The $MAP_t/Ph_t/\infty$ first MDE is*

$$\begin{aligned}
\frac{d}{dt}\mathrm{E}[N(t)] &= \sum_{l=1}^{m_A}\lambda_l(t)\left(\sum_{k=1}^{v_A}a_{l,m_A+k}(t)\right)\mathrm{P}(t;\cdot,l) \\
&\quad -\sum_{j=1}^{m_B}\mu_j(t)b_{j,m_B+1}(t)\mathrm{E}[N_j(t)] \qquad (2)
\end{aligned}$$

where $\mathrm{P}(t;\cdot,k)\equiv\sum_{n_1,n_2,\ldots,n_{m_B}}\mathrm{P}(t;n_1,n_2,\ldots,n_{m_B},k)$ *is the marginal probability of the arrival process being in phase k at time t.*

At this point it is worth noticing that the sole difference between Theorem 1 and its counterpart in Nelson and Taaffe (2004b) is that portion of the positive flux term specified by the arrival process being reinitialized in phase k, that is,

$$\lambda_l(t)\left(\sum_{k=1}^{v_A}a_{l,m_A+k}(t)\right)\mathrm{P}(t;\cdot,l),$$

which here accounts for the $v_A\geq 1$ absorbing phases in the MAP_t, and the corresponding term in Nelson and Taaffe (2004b), namely,

$$\lambda_l(t)a_{l,m_A+1}(t)\mathrm{P}(t;\cdot,l)$$

represents the single absorbing phase in the arrival Ph_t. This results directly from the summation

$$\lambda_l(t)\left(\sum_{j=1}^{v_A}a_{l,m_A+j}(t)\alpha_{j,k}(t)\right) \qquad (3)$$

in Equation (1) for the $MAP_t/Ph_t/\infty$ node replacing the single term

$$\lambda_l(t)a_{l,m_A+1}(t)\alpha_k(t) \qquad (4)$$

in the KFEs for the $Ph_t/Ph_t/\infty$ node. Since this is the sole difference between the KFEs in the two systems, we argue that the remainder of the closed system of moment differential equations (MDEs) and the partial-moment differential equations (PMDEs) needed for calculating $\mathrm{E}[N(t)]$ and $\mathrm{Var}[N(t)]$ for the $MAP_t/Ph_t/\infty$ node may be obtained similarly from the corresponding theorems in Nelson and Taaffe (2004b) by replacing the single term in (4) with the summation in (3) that accounts for the (potentially) multiple absorbing phases in the MAP_t. To this end we leave the presentation (with derivation) of Theorems 3–6 for the online supplement accompanying this journal.

As in the $Ph_t/Ph_t/\infty$ system, the number of differential equations required to numerically evaluate $\mathrm{E}[N(t)]$ is m_A+m_B-1 and the number of additional differential equations required to numerically evaluate $\mathrm{Var}[N(t)]$ (by evaluating $\mathrm{E}[N(t)^2]$) is $m_Am_B(m_B+1)$.

Remark 1. We can also consider the K-node network case where we have R independent, time-dependent, MAP_t-type arrival processes for each of several entity classes and time-dependent, class-specific, Markov routing (MR_t) among K nodes. Two key results are that (1) the single-class K-node network of time-dependent, Ph_t-type service nodes having MR_t among nodes is mathematically equivalent to a single-class *single-node system* with a number of service phases equal to the total number of service phases in the network of service nodes and (2) an R-class $[MAP_t/Ph_t/\infty]^K$ queueing network with class-specific, independent, time-dependent, MAP_t arrival processes and MR_t routing is mathematically equivalent to R independent $[MAP_t/Ph_t/\infty]^K$ queueing networks.

Notice that the description of the set of service nodes in this network is identical to the description of the set of service nodes for the $[Ph_t/Ph_t/\infty]^K$. It follows directly then that the single-node-to-network equivalence is true whether the arrival process is Ph_t or is MAP_t. Of course the behavior of the network is very much a function of the arrival process. We refer the reader to the details laid out in Nelson and Taaffe (2004a, b) and the online supplement.

Remark 2. The $[MAP_t/Ph_t/\infty]^K$ network for systems having R arrival sources (or entity classes) can be represented by R single-arrival-source $[MAP_t/Ph_t/\infty]^K$ networks. Since there is an infinite number of servers at every node, the entity classes are stochastically independent and thus the single-node network model (and software) can be used to analyze each of the entity classes separately at the nodal and network levels. The computational effort to analyze the multiple-arrival-process network is a linear function of the number of arrival sources (entity classes) because the independence of the entity-by-class random processes allows us to evaluate the class-performance measures by looking at the each of the R classes separately. Doing this for each arrival type (or entity class) means that we analyze the entire network (including all arrival classes) by decomposing the network into R separate networks—one for each entity class. This result is identical to the multiple arrival process result for the infinite-server network having Ph_t arrival processes (Nelson and Taaffe 2004a).

4. Examples

In this section we present a series of examples specifically constructed to illustrate the effects of autocorrelation, nonstationarity, and variability.

4.1. The Base $[MAP_t/Ph_t/\infty]^2$ Network Model

We utilize the two-node network illustrated in Figure 1 to quantify the effects of changing network attributes such as autocorrelation in the arrival stream, nonstationarity in arrival rates, and variation in service. A single arrival process feeds the network, which includes immediate feedback at node 1 and feedback from node 2 to node 1. Here we provide values and time functions for the initial parameters of the MAP_t with eight transient and two absorbing phases that we use to model arrivals to the network:

$$\mathscr{A}_1(t)=\begin{bmatrix} 0 & 0.3 & 0.4 & 0.3 & 0 & 0 & 0 & 0\\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0.2 & 0 & 0.8 & 0\\ 0 & 0 & 0 & 0 & 0 & \mathscr{A}_1^{55} & 0 & \mathscr{A}_1^{57}\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0.5\\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix},$$

with $\mathscr{A}_1^{55}=0.25-0.25\cos(\frac{1}{3}t\pi)$,
$\mathscr{A}_1^{57}=0.75+0.25\cos(\frac{1}{3}t\pi)$,

$$\mathscr{A}_2(t)=\begin{bmatrix} 0 & 0\\ 0 & 0\\ 0.5+0.2\cos(\pi t/4) & 0.5-0.2\cos(\pi t/4)\\ 0 & 0\\ 0 & 0\\ 0.6 & 0.4\\ 0 & 0\\ 0 & 0 \end{bmatrix},$$

$$\boldsymbol{\alpha}(t)=\begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix},$$

and

$$\boldsymbol{\lambda}(t)=[7.0\ \ 9.0\ \ 7.0\ \ 9.0\ \ 7.0\ \ 9.0+5.0\cos(\tfrac{1}{2}t\pi)\ \ 7.0\ \ 9.0].$$

Notice that $\operatorname{rank}(\boldsymbol{\alpha}(t))>1$ indicating nonzero autocorrelation in the MAP_t.

The service processes are represented by three- and two-phase Ph_t processes, respectively. The node 1 service process has parameters

$$\mathscr{B}^{[1]}(t)=\left[\begin{array}{ccc|c} 0 & \frac{6}{10} & \frac{1}{10} & \frac{3}{10}\\ \frac{1}{4}-\frac{1}{4}\sin\frac{2\pi t}{7} & \frac{1}{4} & 0 & \frac{2}{4}+\frac{1}{4}\sin\frac{2\pi t}{7}\\ \frac{4}{7} & \frac{3}{7} & 0 & 0\\ \hline \frac{2}{5}-\frac{1}{5}\cos\frac{5\pi t}{11} & \frac{3}{5}+\frac{1}{5}\cos\frac{5\pi t}{11} & 0 & 0 \end{array}\right],$$

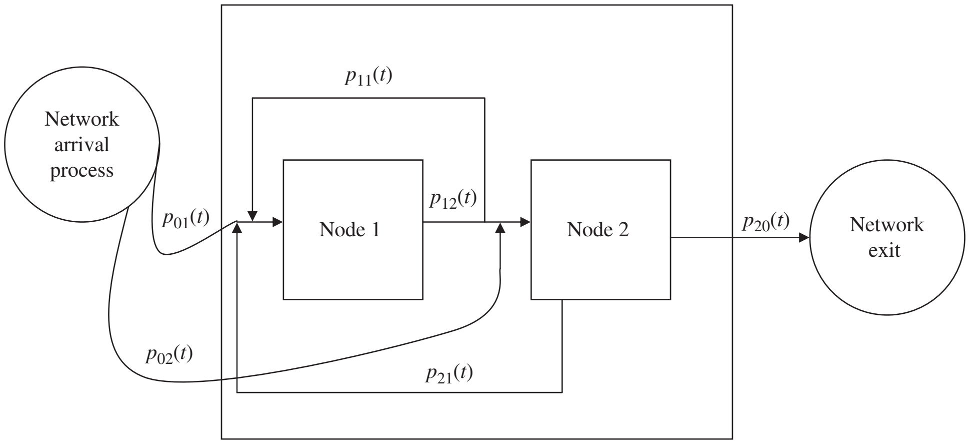
and $\quad \mu^{[1]}(t)=[2\quad 3.5+3\cos(5\pi t/7))\quad 1.5]$

and the node-2 service process parameter values are

$$\mathscr{B}^{[2]}(t)=\left[\begin{array}{cc|c} 0 & 1 & 0\\ \frac{1}{3}+\frac{1}{3}\sin\frac{2\pi t}{5} & \frac{1}{3} & \frac{1}{3}-\frac{1}{3}\sin\frac{2\pi t}{5}\\ \hline \frac{2}{3} & \frac{1}{3} & 0 \end{array}\right],$$

and $\quad \mu^{[2]}(t)=[1\quad 1]$.

Figure 1. Diagram of the Two-Node Network



The Markov routing matrix among the nodes is

$$\mathbf{P}(t) = \left[\begin{array}{cc|c} \frac{1}{2} - \frac{1}{3}\sin\frac{3\pi t}{4} & \frac{1}{2} + \frac{1}{3}\sin\frac{3\pi t}{4} & 0 \\ \frac{1}{6} & 0 & \frac{5}{6} \\ \hline \frac{7}{10} - \frac{2}{10}\cos\frac{2\pi t}{3} & \frac{3}{10} + \frac{2}{10}\cos\frac{2\pi t}{3} & 0 \end{array}\right].$$

Thus, there exists nonstationarity in each of the arrival, service, and routing processes. For the remainder of this section we will refer to this as the *base* network model. Since having R classes is equivalent to R independent K-node networks, we illustrate only a single entity class here.

4.2. Quantifying Network Variations

We first use our example network to quantify the effect of autocorrelation in the arrival process. We compute time-dependent performance measures by numerically integrating the closed system of MDEs for both the base network model and an analogous network where the arrival process has alternative initialization matrix

$$\tilde{\boldsymbol{\alpha}}(t) = \begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 & 0 & 0 & 0 & 0 \\ 0.25 & 0.25 & 0.25 & 0.25 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Since $\operatorname{rank}(\tilde{\boldsymbol{\alpha}}(t)) = 1$, this specifies a Ph_t arrival process with representation $(\tilde{\mathbf{A}}(t), \boldsymbol{\lambda}(t))$. Curves for the mean and the variance of the number of entities (from time $t = 0$ to $t = 10$ when the system begins empty and idle) for each network at each node appear in Figures 2 and 3, respectively, with the base network nodal moments represented by solid thin lines and the alternative network by thick dots. As an aside, even though the $\mathrm{E}[N(t)]$ and $\mathrm{Var}[N(t)]$ curves (respectively at each node) have similar shapes, neither pair of $\mathrm{E}[N(t)]$ and $\mathrm{Var}[N(t)]$ curves is identical, demonstrating that the number of entities at each node (regardless of arrival type) does not follow a time-dependent Poisson distribution.

Notice the top curve in each plot is associated with the network with MAP_t arrivals. Specifically, we see the curve corresponding to the MAP_t representing respective moments nearly 20% higher (at its most extreme) above the curve representing Ph_t arrivals. Thus, we see that failing to capture the autocorrelation present in the MAP_t arrivals causes us to significantly understate both the mean congestion levels at both nodes as well as the variability of this congestion.

Additionally, we can utilize our example to easily quantify effects caused by other simple adjustments to the network parameters. One such adjustment could be to the nonstationarity in the arrival process. To demonstrate, we increase the amplitude in the periodic arrival

Figure 2. Node 1 Moments: Autocorrelation Effect

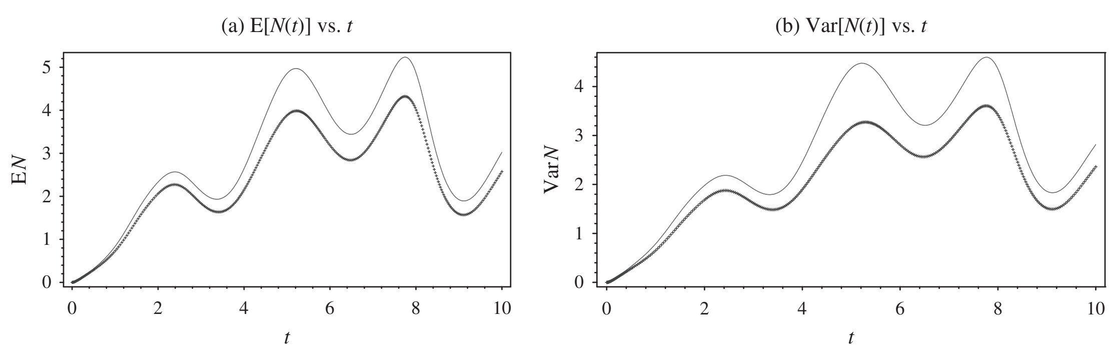


Figure 3. Node 2 Moments: Autocorrelation Effect

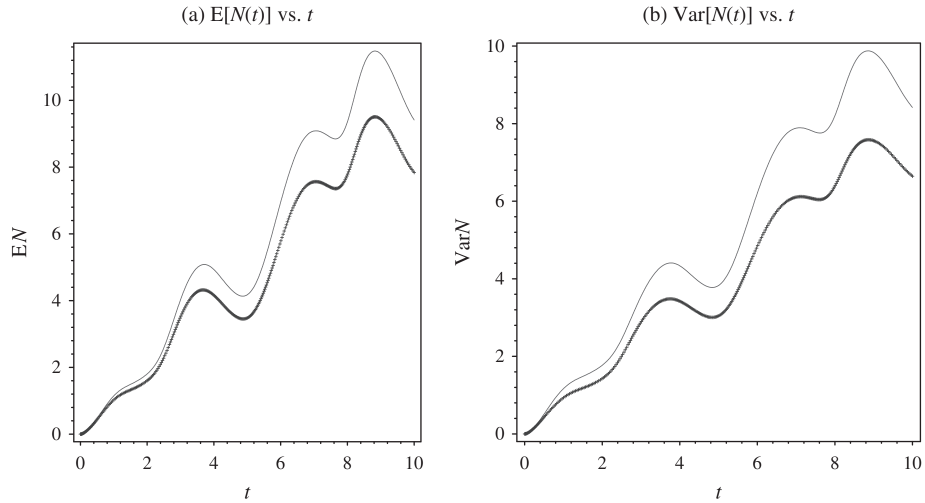
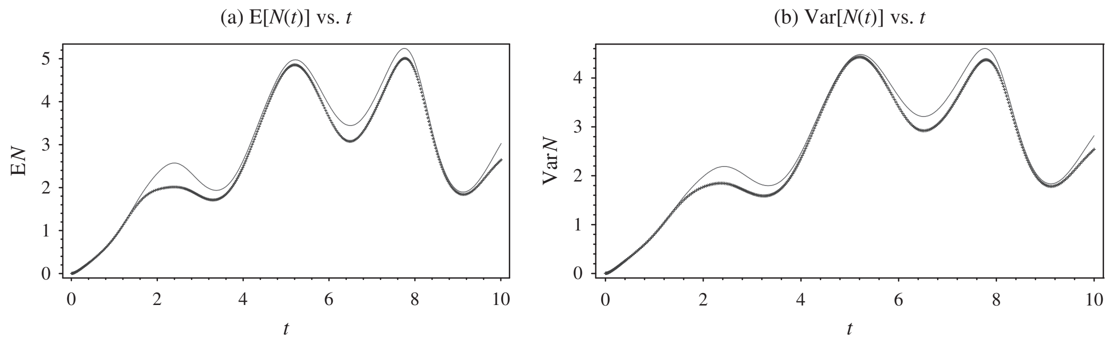


Figure 4. Node 1 Moments: Arrival Nonstationarity Effect



rate in phase 6 of the base network model's initial rate vector, defining

$$\tilde{\lambda}(t) = [7.0\ \ 9.0\ \ 7.0\ \ 9.0\ \ 7.0\ \ 9.0 + 8.0\cos(\tfrac{1}{2}t\pi)\ \ 7.0\ \ 9.0].$$

We then compare the respective nodal moments in the base network model to those in an identical service network fed by an alternative MAP_t with representation $(\mathbf{A}(t), \tilde{\lambda}(t))$, presenting the results in Figures 4 and 5; as above, the base network model is represented by solid thin lines. This adjustment appears to have less effect than was seen in the autocorrelation comparisons; however, this is not unexpected because the adjustment here solely affects the mean time spent in a single phase of the MAP_t.

One other effect we calculate here results from adjusting the variation in service. We define

$$\widetilde{\mathscr{B}^{[1]}}(t) = \left[\begin{array}{ccc|c} 0 & \frac{6}{10} & \frac{1}{10} & \frac{3}{10} \\ \frac{1}{4} - \frac{1}{4}\sin(2\pi t) & \frac{1}{4} & 0 & \frac{2}{4} + \frac{1}{4}\sin(2\pi t) \\ \frac{4}{7} & \frac{3}{7} & 0 & 0 \\ \hline \frac{2}{5} - \frac{1}{5}\cos\frac{5\pi t}{11} & \frac{3}{5} + \frac{1}{5}\cos\frac{5\pi t}{11} & 0 & 0 \end{array}\right],$$

created simply by replacing "$\sin\frac{2\pi t}{7}$" with "$\sin(2\pi t)$" in those relevant terms in the second row of the $\mathscr{B}^{[1]}(t)$ matrix from the base network model. We again compare the respective nodal moments in the base network model to those in a network with the same arrival and node-2 service processes but with node 1 service Ph_t process having representation $(\widetilde{\mathscr{B}^{[1]}}(t), \mu^{[1]}(t))$. The respective plots appear in Figures 6 and 7.

In this section we have utilized our example network to quantify the specific effects on nodal moments from changes to autocorrelation and nonstationarity in the arrival process as well as variation in service; these represent only a few potential network variations that can be analyzed. It is worth noting that although the network presented here includes a single external MAP_t, our technique does allow for more than one independent MAP_t arrival process.

5. Covariance of the Number of Entities at Different Nodes in the $[MAP_t/Ph_t/\infty]^K$ Network

We now show that for the $[MAP_t/Ph_t/\infty]^K$ queueing network the covariance of the number of entities present at time t at node i and the number of entities at time s at node j has a describing differential equation that is *closed*.

Figure 5. Node 2 Moments: Arrival Nonstationarity Effect

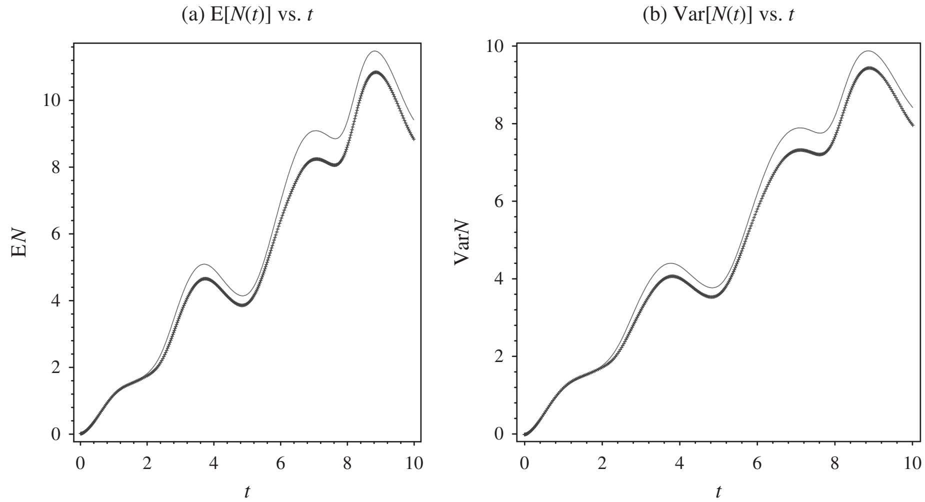


Figure 6. Node 1 Moments: Node 1 Service Variation Effect

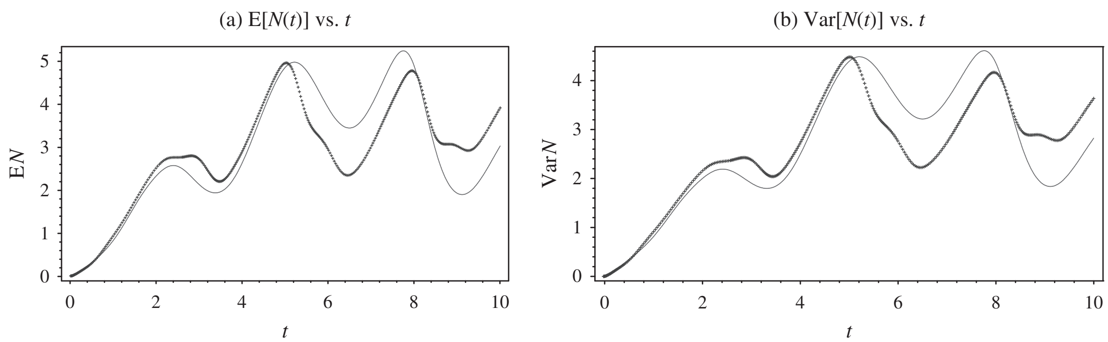
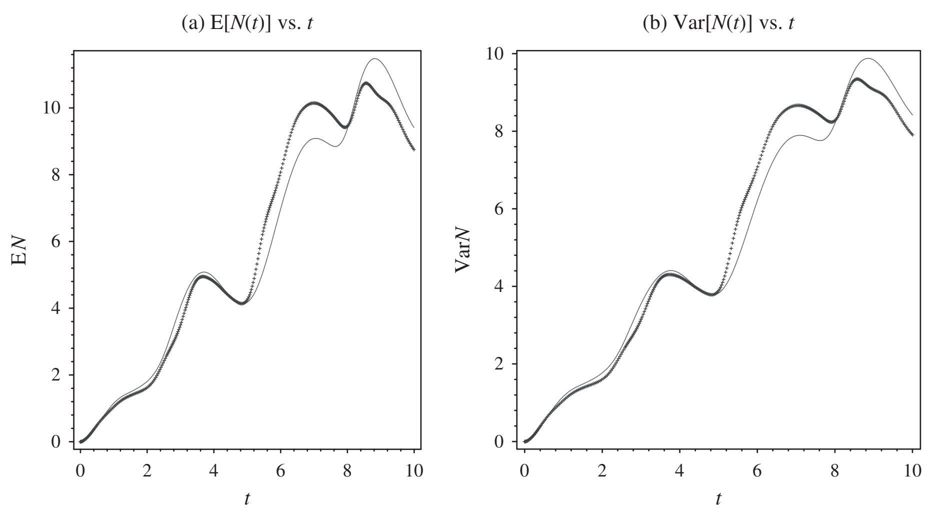


Figure 7. Node 2 Moments: Node 1 Service Variation Effect



We begin with useful notation, presented in Table 1. Notice that for any time-dependent notation M defined in Table 1, we let $M' \equiv (\mathrm{d}/\mathrm{d}t)M$; the sole exception is $C_{i,j}(s,t)' \equiv (\partial/\partial t)C_{i,j}(s,t)$.

For the $MAP_t/Ph_t/\infty$ the partial marginal KFEs for phases i and j are

$$\frac{\partial}{\partial t}\mathrm{P}(N^{[i]}(s)=n_i, N^{[j]}(t)=n_j)$$

$$\begin{aligned}
&= -\beta_j(t)\sum_{h=1}^{m_A}\lambda_h(t)\sum_{r=1}^{v_A} a_{h,m_A+r}(t)\\
&\quad \cdot \mathrm{P}(N^{[i]}(s)=n_i, N^{[j]}(t)=n_j, A(t)=h)\\
&\quad - [1-b_{jj}(t)]\mu_j(t)n_j\mathrm{P}(N^{[i]}(s)=n_i, N^{[j]}(t)=n_j)\\
&\quad - \sum_{\substack{r=1\\ r\neq j}}^{m_B}\mu_r(t)b_{rj}(t)\sum_{n_r=0}^{\infty} n_r
\end{aligned}$$

$$
\begin{aligned}
&\cdot \mathrm{P}(N^{[i]}(s)=n_i, N^{[j]}(t)=n_j, N_r(t)=n_r)\\
&+\beta_j(t)\sum_{h=1}^{m_A}\lambda_h(t)\sum_{r=1}^{v_A}a_{h,m_A+r}(t)\\
&\cdot \mathrm{P}(N^{[i]}(s)=n_i, N^{[j]}(t)=n_j-1, A(t)=h)\\
&+\mu_j(t)[1-b_{jj}(t)](n_j+1)\\
&\cdot \mathrm{P}(N^{[i]}(s)=n_i, N^{[j]}(t)=n_j+1)\\
&+\sum_{\substack{r=1\\ r\neq j}}^{m_B}\mu_r(t)b_{rj}(t)\sum_{n_r=0}^{\infty}n_r\\
&\cdot \mathrm{P}(N^{[i]}(s)=n_i, N^{[j]}(t)=n_j-1, N_r(t)=n_r).
\end{aligned}
$$

From these equations we can compute the required partial derivatives of the form $(\partial/\partial t)\mathrm{E}_{ij}(s,t)$ for all $t \geq s$ and then use these results to compute the $C_{i,j}(s,t)'$ terms. The proof of the following theorem, found in the online supplement, makes use of the expression for $\mathrm{dE}_j(t)/\mathrm{dt}$ developed in Nelson and Taaffe (2004b).

Theorem 2. *For the $MAP_t/Ph_t/\infty$ with $t \geq s$, $s \geq 0$, and $i, j = 1,2,\ldots m_B$,*

$$C_{i,j}(s,t)' \equiv -\mu_j(t)C_{i,j}(s,t)+\sum_{r=1}^{m_B}\mu_r(t)b_{r,j}(t)C_{i,j}(s,t). \tag{5}$$

Notice that Theorem 2 is written in the notation used for the single-node system, which we can apply to finding covariances between all pairs of phases within a node as well as the covariances among all pairs of nodes in the network case since the $[MAP_t$ (or $Ph_t)/Ph_t/\infty]^K$ network system is mathematically equivalent to the single-node system. To implement the notationally cumbersome network case, we need to use the fact that a *node* is made up of (possibly) many *phases* and make use of the standard covariance-decomposition:

$$
\begin{aligned}
\mathrm{Cov}[Q+R, Y+Z] = {}& \mathrm{Cov}[Q,Y]+\mathrm{Cov}[Q,Z]\\
&+\mathrm{Cov}[R,Y]+\mathrm{Cov}[R,Z].
\end{aligned}
$$

Table 1. Useful Notation for the Covariance MDEs

Notation	Description
m_B	The number of phases (nodes) in the service process (network)
m_A, v_A	Number of nonabsorbing and absorbing (respectively) phases in the arrival process
$N^{[i]}(s)$	Number of entities in the system at node i at time s for $s \geq 0$ and $i = 1,2,\ldots,m_B$
$\mathrm{E}_i(s)$	$\mathrm{E}[N^{[i]}(s)]$
$\mathrm{E}_{ij}(s,t)$	$\mathrm{E}[N^{[i]}(s)N^{[j]}(t)]$
$\mathrm{E}_{ijj}(s,t,t)$	$\mathrm{E}[N^{[i]}(s)(N^{[j]}(t))^2]$
$\mathrm{E}_{ij,h}(s,t)$	$\mathrm{E}[N^{[i]}(s)N^{[j]}(t), A(t)=h]$
$b_{i,j}(t)$	Markov routing probability for entities proceeding to phase j after having finished their service at phase i at time t for $i, j = 1,2,\ldots,m_B$, $t \geq 0$
$C_{i,j}(s,t)$	$\mathrm{Cov}[N^{[i]}(s), N^{[j]}(t)]$, for $s \geq 0$, $t \geq 0$, $i, j = 1,2,\ldots,m_B$

Clearly at time $t = s$, $C_{i,i}(s,s) = \mathrm{Var}[N^{[i]}(s)]$. If the "initial" condition (i.e., the state of the system at time s) is empty-and-idle then $C_{i,i}(s,s) = 0$ and $C_{i,i}(s,t) = 0$, for all $t \geq s$. Likewise if at time $t = s$ we have an empty-and-idle system, then $C_{i,j}(s,s) = \mathrm{Cov}[N^{[i]}(s), N^{[j]}(s)] = 0$ and $C_{i,j}(s,t) = \mathrm{Cov}[N^{[i]}(s), N^{[j]}(t)] = 0$ for all $t \geq s$.

The interesting case is if at some arbitrarily selected time s we need to compute the effects of the *current* node-i-time-s-to-node-j-time-s covariance or the node-i-time-s-to-node-j-time-t covariance. Thus we need to be able to compute the covariance starting from arbitrary (or random) initial (time s) conditions. For instance, we may need to evaluate the queueing systems of interest for some interval ending at time s_1^- and then compute the covariance for the system from s_1^+ until time s_2^-, and so on. The indicated discontinuity at times s_l is allowed so that at those epochs the state of the system can be deterministically altered (for instance by adding some entities to the system). The ability to compute results for a system in which a deterministic number of entities can be added or deleted at fixed points in time allows for time-based control protocols to be examined. Of course, time-based control policies could by computed and perhaps optimized, but that optimization is beyond the scope of this paper.

Finally, the covariance partial differential equations for the $Ph_t/Ph_t/\infty$ and $MAP_t/Ph_t/\infty$ are exactly the same as those for the covariance in a $[M_t/M_t/\infty]^K$ system, or equivalently the $M_t/Ph_t/\infty$ system. Whitt and Massey (1993) showed this result for the $[M_t/M_t/\infty]^K$ and it is well known that the $[M_t/M_t/\infty]^K$ is equivalent to the $M_t/Ph_t/\infty$. We provide a detailed proof using our notation in the online supplement and on our Nonstationary Queueing Network website (http://users.iems.nwu.edu/~nelsonb/PhPhInf/).

References

Artalejo J, Gómez-Corral A, He QM (2010) Markovian arrivals in stochastic modeling: A survey and some new results. *Statist. Oper. Res. Trans.* 34(2):101–144.

Asmussen S (1987) *Applied Probability and Queues* (John Wiley & Sons, New York).

Asmussen S (2000) Matrix-analytic models and their analysis. *Scandinavian J. Statist.* 27(2):193–226.

Asmussen S, Koole G (1993) Marked point processes as limits of Markovian arrival streams. *J. Appl. Probab.* 30(2):365–372.

Bitran GR, Dasu S (1993) Approximating nonrenewal processes by Markov chains: Use of Super-Erlang (SE) chains. *Oper. Res.* 41(5):903–923.

Casale G, Zhang EZ, Smirni E (2010a) Trace data characterization and fitting for Markov modeling. *Performance Eval.* 67(2):61–79.

Casale G, Zhang EZ, Smirni E (2010b) KPC-toolbox: Best recipes for automatic trace fitting using Markovian arrival processes. *Perform. Eval.* 67(9):873–896.

Diamond JE, Alfa AS (2000) On approximating higher order MAPs with MAPs of order two. *Queueing Systems* 34(1): 269–288.

Ferng HW, Chang JF (2001) Connection-wise end-to-end performance analysis of queuing networks with MMPP inputs. *Performance Evaluation* 43(1):39–62.

Fischer W, Meier-Hellstern K (1993) The Markov-modulated Poisson process (MMPP) cookbook. *Performance Evaluation* 18(2): 149–171.

Heffes H (1980) A class of data traffic processes: Covariance function characterization and related queueing results. *Bell System Tech. J.* 59(6):897–929.

Horváth A, Telek M (2006) Formal methods and stochastic models for performance evaluation. *Third Euro. Performance Engrg. Workshop (EPEW 2006), Budapest, Hungary*.

Horváth G, Telek M, Buchholz P (2005) A MAP fitting approach with independent approximation of the inter-arrival time distribution and the lag correlation. *QEST '05 Proc. Second Internat. Conf. Quantitative Eval. Systems* (IEEE Computer Society, Los Alamitos, CA), 124–133.

Johnson MA (1998) Markov MECO: A simple Markovian model for approximating nonrenewal arrival processes. *Comm. Statist.–Stochastic Models* 14(1&2):419–442.

Liu DM, Neuts MF (1991) Counter-examples involving Markovian arrival processes. *Commun. Statist.-Stochastic Models* 7(3): 499–509.

Lucantoni DM (1991) New results on the single server queue with a batch Markovian arrival process. *Comm. Statist.–Stochastic Models* 7(1):1–46.

Narayana S, Neuts MF (1992) The first two moments matrices of the counts for the Markovian arrival processes. *Commun. Statist.-Stochastic Models* 8(3):459–477.

Nelson BL, Taaffe MR (2004a) The $[Ph(t)/Ph(t)/\infty]^K$ queueing system: Part II—The multiclass network. *INFORMS J. Comput.* 16(3):275–283.

Nelson BL, Taaffe MR (2004b) The $Ph(t)/Ph(t)/\infty$ queueing system: Part I—The single node. *INFORMS J. Comput.* 16(3):266–274.

Neuts MF (1979) A versatile Markovian point process. *J. Appl. Probab.* 16(4):764–779.

Neuts MF, Liu D, Narayana S (1992) Local Poissonification of the Markovian arrival process. *Stochastic Models* 8(1):87–129.

Takine T, Hasegawa TT (1994) The workload in the $MAP/G/1$ queue with state-dependent services: Its application to a queue with preemptive resume priority. *Commun. Statist.-Stochastic Models* 10(1):183–204.

Whitt W, Massey WA (1993) Networks of infinite-server queues with nonstationary Poisson input. *Queueing Systems* 13(1):183–250.