A Queueing Approximation of MMPP/PH/1

Azam Asanjarani and Yoni Nazarathy

Abstract We consider the well-studied MMPP/PH/1 queue and illustrate a method to find an almost equivalent model, the MTCP/PH/1. MTCP stands for Markovian Transition Counting Process. It is a counting process that has similar characteristics to MMPP (Markov Modulated Poisson Process). We prove that for a class of MMPPs there is an equivalent class of MTCPs. We then use this property to suggest an approximation for MMPP/PH/1 in terms of the first two moments. We numerically show that the steady state characteristics of MMPP/PH/1 are well approximated by the associated MTCP/PH/1 queue. Our numerical analysis leaves some open problems on bounds of the approximations. Of independent interest, this paper also contains a lemma on the workload expression of MAP/PH/1 queues which to the best of our knowledge has not appeared elsewhere.

Keywords Markov modulated poisson process · MAP/PH/1 · Queueing

1 Introduction

Queuing theory finds a variety of applications such as telecommunication networks, healthcare and manufacturing, see for instance [6]. One of the most useful queueing models is the MAP/PH/1 queue, see for example [14]. The Markovian Arrival Process (MAP) is a counting process based on a background finite-state Continuous-Time Markov Chain (CTMC). MAP can be considered as a generalisation of the Poisson process where the inter-arrival times of a MAP are not necessarily independent of each other, nor exponentially distributed. The Phase type (PH) distribution is a generalization of the exponential distribution and is based on the distribution of time until absorption in a finite-state CTMC. These two matrix-analytic objects make up the MAP/PH/1 queue: the arrival process is MAP, and the service times are assumed i.i.d. from a PH distribution.

A. Asanjarani(✉) · Y. Nazarathy
The University of Queensland, Brisbane, Queensland, Australia
e-mail: a.asanjarani@uq.edu.au

T.V. Do et al. (eds.), *Queueing Theory and Network Applications*,
Advances in Intelligent Systems and Computing 383,
DOI: 10.1007/978-3-319-22267-7_4

Comparison of different stochastic processes to find a versatile model for describing observed data in an accurate manner is a fundamental objective in stochastic modelling. In modelling a variety of phenomena such as queueing processes, the Markov Modulated Poisson Process (MMPP), a special case of MAP, can be applied. The MMPP has a variety of applications in modelling bursty traffic. The motivation behind the vast applications of MMPP is that MMPP keeps the tractability of the Poisson process while enabling non-zero correlation between inter-arrival times. See for example [5], [10] and [13].

In this paper we introduce an alternative model to MMPP which we refer to as the Markovian Transition Counting Process (MTCP). MTCP is a MAP which counts every transition of the background CTMC. We believe it is more tractable and more computationally convenient than the MMPP. We find relations between MTCP and MMPP, focusing on the case of a two state background CTMC for the MMPP. We prove that in some cases, the first two moments of MMPP and MTCP can be matched. We refer to these cases as slow MMPPs. This implies that the intensity of arrivals is greater than the total intensity of state changes per state. From a modelling perspective, slow MMPPs are perhaps the most useful MMPPs because non-slow MMPPs have characteristics quite similar to the Poisson process.

In using MTCP for queues, we investigate the behaviour of the MTCP/PH/1 queue as an alternative to the MMPP/PH/1 queue. Here, we address this question empirically through extensive numerical experiments. We show that the basic steady state characteristics (mean and variance of the queue) of a given MMPP/PH/1 queue can be emulated by an MTCP/PH/1 queue almost without relative error in most cases, and with relative errors that are bounded at the worst case by 9%. These preliminary results are significant for the emerging body of research dealing with finding alternative (but similar) queueing models.

As a stochastic modeller chooses a suitable queueing model for a given situation, there is typically more than one choice. Knowing that MTCP/PH/1 is similar to MMPP/PH/1 allows the modeller to have more freedom in model choice. In future research we shall integrate this within a statistical model-selection framework, fitting queueing models to data. Towards that end, a key advantage of using MTCP/PH/1 instead of MMPP/PH/1 is that the MTCP is more informative than the MMPP. In fact we believe that our MTCP is better suited for parameter estimation since for this model, each observed event corresponds to exactly one transition in the background (unobserved) CTMC.

The remainder of this paper is structured as follows: In Section 2 we overview the MMPP/PH/1 queue and treat it as a Quasi-Birth-Death (QBD) process. We also present a lemma on the workload expression of MAP/PH/1 queues which to the best of our knowledge has not appeared elsewhere. In Section 3 we introduce the new model, MTCP, as a special MAP. In Section 4 we show that for a slow MMPP_2, a useful substitute MTCP_4 exists. In fact, we prove that the first and second moments of these two model classes (slow MMPP and MTCP) can be matched. In Section 5 numerical results for approximating a given $\text{MMPP}_2/\text{PH}_2/1$ with an $\text{MTCP}_4/\text{PH}_2/1$ are presented. We conclude in Section 6.

2 The MMPP/PH/1 Queue

The MMPP/PH/1 queue is a special case of the general single-server queue MAP/G/1, where the stream of arrivals and service mechanism are modelled by MMPP and PH distribution respectively. Figure 1 illustrates an example of an $MMPP_2/PH_2/1$ queue. Methods of analysing the MMPP/PH/1 queueing models can be found in [7] and [10]. In this paper we use the uniform framework of QBD processes which is an efficient way to analyse more general models using matrix-analytic methods, see [9].

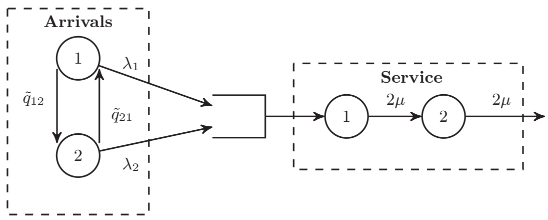


Fig. 1 A schematic illustration of the $MMPP_2/E_2/1$ queue (E_2 is a special case of PH_2 and stands for Erlang, where in this case it has a mean of μ^{-1}). The circles illustrate phases of the arrival and/or service mechanism.

MMPP. An MMPP is simply an arrival process which consists of a finite number of Poisson processes, modulated by a CTMC. In other words, MMPP is a special case of a doubly stochastic Poisson process whose arrival rate is modulated by the states of an irreducible finite-state CTMC, which is referred to as the phase process. The parameters of an MMPP of order p are the vector of Poisson arrival rates associated with each phase, $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_p)'$ as well as the parameters of the p-state background CTMC: the transition rate matrix Q and the initial distribution of the background CTMC, taken as a row vector $\boldsymbol{\alpha}$.

PH Distribution. The time until absorption into state 0 (absorbing state) of a finite-state CTMC with q transient state and one absorbing state is said to have a phase type (PH) distribution of order q. A PH distribution of order q is parametrised by $\boldsymbol{\eta}$ and T, where $\boldsymbol{\eta}$ is the initial distribution over the transient states (taken as a row vector) and the matrix $T = \{t_{ij}\}_{i,j=1,\ldots,q}$ specifies the transition rates between the transient states of the CTMC. PH distributions are very versatile and are dense in the class of distributions defined on the non-negative real numbers [3]. Moreover, PH distributions are used in a wide range of applications, see for instance [1] and [8].

QBD and MMPP/PH/1. A continuous-time homogeneous QBD_r is a Markov process characterised by a two dimensional state space $\{(n, i) : 0 \leq n, 1 \leq i \leq r\}$, which are called the level and the phase of the state, respectively. A transition from (n, i) to (n', j) is possible only when $|n' - n| < 2$ and the transition rate from (n, i) to (n', j) may depend on i, j and $|n' - n|$, but not on the specific values of n and n'. When ordering the states in lexicographic order, the transition rate matrix of a

QBD_r has the following form:

$$A = \begin{pmatrix} B_0 & B_1 & & & 0 \\ B_{-1} & A_0 & A_1 & & \\ & A_{-1} & A_0 & A_1 & \\ & & A_{-1} & A_0 & A_1 \\ 0 & & & \ddots & \ddots & \ddots \end{pmatrix}. \tag{1}$$

In representing the $\text{MMPP}_p/\text{PH}_q/1$ queue as a QBD_r, where $r = p \times q$, the phase records (in lexicographic order) both the background state of the MMPP (arrival) and the current phase of the service (see Figure 1 for illustration of the phases in the special case of $\text{MMPP}_2/\text{E}_2/1$). The level, represents the number of items in the system.

Modelled as a QBD, we have:

$$B_{-1} = I_p \otimes \mathbf{t}, \quad B_0 = C, \quad B_1 = \text{diag}(\boldsymbol{\lambda}) \otimes \boldsymbol{\eta},$$

where $C = Q - \text{diag}(\boldsymbol{\lambda})$ and where $\otimes$ is the Kronecker product. Here $\mathbf{t} = -T\mathbf{1}$, where $\mathbf{1}$ is a column vector of 1's with appropriate dimension.

Further,

$$A_{-1} = I_p \otimes \mathbf{t}\boldsymbol{\eta}, \quad A_0 = I_p \otimes T + C \otimes I_q, \quad A_1 = \text{diag}(\boldsymbol{\lambda}) \otimes I_q.$$

As is well known in the theory of QBDs, the stationary distribution of a positive-recurrent QBD, $\boldsymbol{\pi}$, admits a matrix-geometric form $\pi_n = \pi_{n-1}R$, where R is the solution of a quadratic fixed-point matrix equation $R = A_1 + RA_0 + R^2 A_{-1}$ and π_n are row vectors of dimension r, see [9]. We use the-state-of-the-art SMC solver to find the matrix R and the stationary distribution of a given QBD, see [4]. It is easy to show that A is irreducible due to the properties of the building blocks and irreducibility of Q. Moreover, characterizing the positive-recurrence can be done as follows[1].

Lemma 1. *The QBD representing a $MAP_p/PH_q/1$ queue is positive-recurrent if and only if,*

$$\rho := \frac{\boldsymbol{\beta} A_1 \mathbf{1}}{\boldsymbol{\beta} A_{-1} \mathbf{1}} = \frac{\Lambda}{\frac{1}{-\boldsymbol{\eta} T^{-1} \mathbf{1}}} < 1,$$

where $-\boldsymbol{\eta}T^{-1}\mathbf{1}$ is the first moment of PH_q with parameters $(\boldsymbol{\eta}, T)$, $\Lambda = \boldsymbol{\pi} D \mathbf{1}$ is the first moment of a time-stationary MAP_p with parameters $(\boldsymbol{\pi}, C, D)$[2], and $\boldsymbol{\beta}$ is the stationary distribution of $A_{-1} + A_0 + A_1$.

[1] To the best of our knowledge, the algebra behind this intuitive lemma has not appeared elsewhere.

[2] The QBD representation of $\text{MAP}_p/\text{PH}_q/1$ generalises the $\text{MMPP}_p/\text{PH}_q/1$ representation, with $\text{diag}(\boldsymbol{\lambda})$ being replaced by D (see next Section for MAPs).

Proof. The fact that the left hand side of ρ is a necessary and sufficient condition for positive recurrence follows from the theory of QBDs (see [9], Theorem 7.2.4). It remains to show that both representations of ρ agree.

First we show that $\boldsymbol{\beta} = \boldsymbol{\pi} \otimes \boldsymbol{\gamma}$, where $\boldsymbol{\gamma}$ is the unique solution of $\boldsymbol{\gamma}(T + \mathbf{t}\boldsymbol{\eta}) = \mathbf{0}'$ and $\boldsymbol{\gamma}\mathbf{1} = 1$[3]. It is immediate that $(\boldsymbol{\pi} \otimes \boldsymbol{\gamma})\mathbf{1} = 1$. Further, we have

$$\begin{aligned}(\boldsymbol{\pi} \otimes \boldsymbol{\gamma})(A_{-1} + A_0 + A_1) &= (\boldsymbol{\pi} \otimes \boldsymbol{\gamma})\big(I_p \otimes \mathbf{t}\boldsymbol{\eta} + (I_p \otimes T + C \otimes I_q) + D \otimes I_q\big) \\ &= (\boldsymbol{\pi} \otimes \boldsymbol{\gamma})\big(I_p \otimes (\mathbf{t}\boldsymbol{\eta} + T) + (C + D) \otimes I_q\big) \\ &= (\boldsymbol{\pi} \otimes \boldsymbol{\gamma})\big((C + D) \otimes (\mathbf{t}\boldsymbol{\eta} + T)\big) \\ &= \mathbf{0}',\end{aligned}$$

where the last two steps follow since $(A \otimes B)(A' \otimes B') = AA' \otimes BB'$ when matrix dimensions agree for the multiplication.

Now we need to show that $\frac{\boldsymbol{\beta} A_1 \mathbf{1}}{\boldsymbol{\beta} A_{-1} \mathbf{1}} = \frac{\Lambda}{\frac{1}{-\boldsymbol{\eta} T^{-1} \mathbf{1}}}$ or equivalently:

$$\boldsymbol{\beta} A_1 \mathbf{1} = (\boldsymbol{\beta} A_{-1} \mathbf{1})\, \Lambda\, (-\boldsymbol{\eta} T^{-1} \mathbf{1})$$

which for the MAP$_p$/PH$_q$/1 queue is written as:

$$(\boldsymbol{\pi} \otimes \boldsymbol{\gamma})(D \otimes I_q)\mathbf{1} = (\boldsymbol{\pi} \otimes \boldsymbol{\gamma})(I_p \otimes \mathbf{t}\boldsymbol{\eta})\mathbf{1}(\boldsymbol{\pi} D \mathbf{1})(-\boldsymbol{\eta} T^{-1} \mathbf{1}). \tag{2}$$

For the left hand side, we have

$$(\boldsymbol{\pi} \otimes \boldsymbol{\gamma})(D \otimes I_q)\mathbf{1} = (\boldsymbol{\pi} D \otimes \boldsymbol{\gamma})\mathbf{1} = \boldsymbol{\pi} D \mathbf{1}.$$

Therefore we need to show that the right hand side of (2) is equal to $\boldsymbol{\pi} D \mathbf{1}$, or equivalently:

$$(\boldsymbol{\pi} \otimes \boldsymbol{\gamma})(I_p \otimes \mathbf{t}\boldsymbol{\eta})\mathbf{1}(-\boldsymbol{\eta} T^{-1} \mathbf{1}) = 1.$$

Since $\boldsymbol{\pi}\mathbf{1} = \boldsymbol{\eta}\mathbf{1} = 1$, we have $(\boldsymbol{\pi} \otimes \boldsymbol{\gamma})(I_p \otimes \mathbf{t}\boldsymbol{\eta})\mathbf{1} = (\boldsymbol{\pi} \otimes \boldsymbol{\gamma}\mathbf{t}\boldsymbol{\eta})\mathbf{1} = \boldsymbol{\gamma}\mathbf{t}$. Moreover, from $\boldsymbol{\gamma}(T + \mathbf{t}\boldsymbol{\eta}) = \mathbf{0}'$ we have $\boldsymbol{\gamma}\mathbf{t}\boldsymbol{\eta} = -\boldsymbol{\gamma} T$ which results in $\boldsymbol{\gamma}\mathbf{t}(-\boldsymbol{\eta} T^{-1} \mathbf{1}) = 1$. □

3 MAPs and the Markovian Transition Counting Process

MAP. A MAP is a pure birth process which can be considered as a special case of the QBD: a MAP is a two-dimensional Markov process with parameters $(\boldsymbol{\alpha}, C, D)$ where $\boldsymbol{\alpha}$ is the initial distribution of the finite-state CTMC and the matrix $C = B_0 = A_0$ records the transitions of the background CTMC with no arrival. The event intensity matrix $D = B_1 = A_1$ has non-negative elements and describes the transitions of the background CTMC with an arrival. The matrices A_{-1} and B_{-1} are zero matrices.

[3] Note that $\boldsymbol{\gamma}$ is the limiting distribution of the phase in a PH$_q$-renewal process.

Moreover, we have $C + D = Q$, where Q is the transition rate matrix of the CTMC. A MAP with parameters $(\boldsymbol{\alpha}, C, D)$ is time-stationary if $\boldsymbol{\alpha} = \boldsymbol{\pi}$, where $\boldsymbol{\pi}$ is the stationary distribution of the phase process, i.e. $\boldsymbol{\pi} Q = \mathbf{0}'$ and $\boldsymbol{\pi}\mathbf{1} = 1$.

For a time-stationary MAP, the mean and variance of the number of counts at any time t are given by the following formulas, see Chapter XI of [3]:

$$\mathbb{E}[N(t)] = \Lambda t = \boldsymbol{\pi} D \mathbf{1}\, t, \tag{3}$$

$$\mathrm{Var}\Big(N(t)\Big) = \{\Lambda - 2\Lambda^2 + 2\,\boldsymbol{\pi} D Q^{-} D\mathbf{1}\}\, t + 2\,\boldsymbol{\pi} D Q^{-}(e^{Qt} - I) Q^{-} D\mathbf{1}, \tag{4}$$

where $Q^{-} = (\mathbf{1}\boldsymbol{\pi} - Q)^{-1}$.

The class of MAPs contains most of the commonly used point processes such as the Poisson process ($D = \lambda$, where λ is the Poisson rate and $C = -\lambda$) and MMPP ($D = \mathrm{diag}(\boldsymbol{\lambda})$ where $\boldsymbol{\lambda}$ is the vector of Poisson rates and $C = Q - D$). In this research, we introduce and investigate a class of MAPs as follows:

Definition 1. *A* ***Markovian Transition Counting Process (MTCP)*** *is a two-dimensional Markov process* $\{(\bar{N}(t), X(t)); t \geq 0\}$ *where* $\bar{N}(t)$ *counts every transition of an irreducible CTMC* $X(\cdot)$ *on* $[0, t]$.

Therefore MTCP is a special type of MAP where we have $\bar{D} = \bar{Q} - \mathrm{diag}(\bar{Q})$ and the parameters of MTCP are just the parameters of the background CTMC. MTCPs and MMPPs are in a sense the extreme cases of MAPs. In an MMPP, the events do not coincide with state transitions (with probability 1). In contrast, in an MTCP the events are precisely all the transitions of the CTMC. This fact motivates the idea of finding relations between MTCP and MMPP. An early reference that analyses both MTCPs and MMPPs (although not using these names) is [12]. We now show some further relations.

4 Relations between MTCP and MMPP

In Proposition 3.2 of [11], the authors showed that every MTCP has an associated MMPP with the same two first moments. For completeness, we present this proposition of [11] in an alternative form here, including the proof.

Proposition 1. *Let* $\bar{N}(t)$ *be the counting processes of a time-stationary* $MTCP_p$*. Then there is an* $MMPP_p$*, with the counting processes* $\tilde{N}(t)$*, such that their first and second moments are matched. That is, for* $\forall t \geq 0$*,*

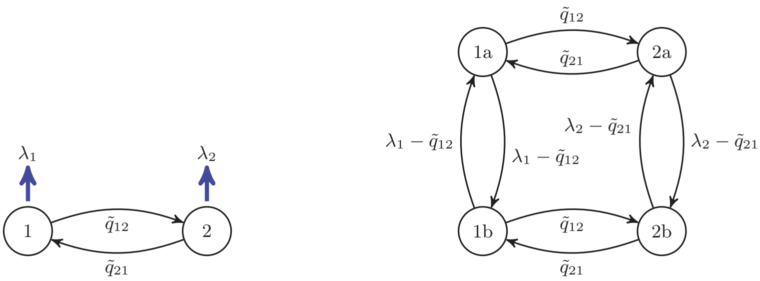
$$\mathbb{E}[\tilde{N}^k(t)] = \mathbb{E}[\bar{N}^k(t)]\,, \qquad \textit{for} \quad k = 1, 2\,.$$

Proof. Assume that the event matrix of the MTCP_p is given by $\bar{D} = \bar{Q} - \mathrm{diag}(\bar{Q})$. Consider an MMPP_p with the same background Markov chain and set $\tilde{D} = -\mathrm{diag}(\bar{Q})$. Now from (3) and (4), we just need to show that $\bar{D}\mathbf{1} = \tilde{D}\mathbf{1}$ and $\boldsymbol{\pi}\bar{D} = \boldsymbol{\pi}\tilde{D}$. Since $\bar{Q}\mathbf{1} = 0$ and $\boldsymbol{\pi}\bar{Q} = \mathbf{0}'$ the result follows. □

The proof shows that in order to construct an MMPP matching an MTCP with parameter $\bar{Q}$: set $\boldsymbol{\lambda} = -\text{diag}(\bar{Q})$ and $\tilde{Q} = \bar{Q}$. The question is now how to construct MTCPs matching MMPPs. Based on the above proposition, the answer is given for the special case of MMPPs where $\boldsymbol{\lambda} = -\text{diag}(\tilde{Q})$, i.e. $\lambda_i = \sum_{j \neq i} \tilde{q}_{ij}$. But this is a very restricted case since it does not leave any freedom with λ_i.

We now show that for each instance in a class of MMPPs (of order 2), where $\lambda_i > \sum_{j \neq i} \tilde{q}_{ij}$ which we call "slow MMPPs", there is an associated MTCP (of order 4) that exhibits the same first and second moments for the counting process. We believe a similar construction holds for arbitrary $p > 2$ (relating MTCP_{2p} to MMPP_p), this remains the subject of future work.

Definition 2. *A* ***slow Markov Modulated Poisson Process (slow MMPP)*** *is an MMPP where for any phase i in the phase process, the arrival rate is greater than the rate of leaving that phase, i.e.* $\lambda_i > \sum_{j \neq i} \tilde{q}_{ij}$.



(a) Transition diagram of the phase process of an MMPP_2.

(b) Transition diagram of the phase process of related MTCP_4.

Fig. 2 An MMPP_2 and its associated MTCP_4

Given MMPP parameters, $\boldsymbol{\lambda}$ and $\tilde{Q}$, we can associate an MTCP_4 to any slow MMPP_2 as illustrated in Figure 2. The transition rate matrix $\bar{Q}$ and the event intensity matrix $\bar{D}$ of the associated MTCP_4 are given as follows:

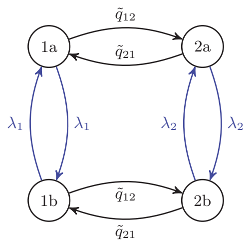
$$\bar{Q} = \left(\begin{array}{cc|cc} -\lambda_1 & \lambda_1 - \tilde{q}_{12} & \tilde{q}_{12} & 0 \\ \lambda_1 - \tilde{q}_{12} & -\lambda_1 & 0 & \tilde{q}_{12} \\ \hline \tilde{q}_{21} & 0 & -\lambda_2 & \lambda_2 - \tilde{q}_{21} \\ 0 & \tilde{q}_{21} & \lambda_2 - \tilde{q}_{21} & -\lambda_2 \end{array}\right), \quad \bar{D} = \bar{Q} - \text{diag}(\bar{Q}). \quad (5)$$

We now have the following:

Proposition 2. *Let $\tilde{N}(t)$ and $\bar{N}(t)$ be the counting processes of a time-stationary slow $MMPP_2$ and its associated $MTCP_4$, respectively. Then, these processes have the same first and second moment. That is, for $\forall t \geq 0$,*

$$\mathbb{E}[\tilde{N}^k(t)] = \mathbb{E}[\bar{N}^k(t)]\,, \qquad \textit{for} \quad k = 1, 2\,.$$

Proof. We first construct a MAP_4 with the same counting process as the $MMPP_2$ by coupling the events of the phase process of $MMPP_2$. When the process is in phase k, Figure 3 shows the structure of a coupled MAP that results in transition from phase k_a to k_b or vice versa. $\tilde{Q}$ is the phase transition matrix and $\tilde{D}$ is the event intensity matrix of the resulting MAP_4.



$$\tilde{Q} = \left(\begin{array}{cc|cc} -(\lambda_1+\tilde{q}_{12}) & \lambda_1 & \tilde{q}_{12} & 0 \\ \lambda_1 & -(\lambda_1+\tilde{q}_{12}) & 0 & \tilde{q}_{12} \\ \hline \tilde{q}_{21} & 0 & -(\lambda_2+\tilde{q}_{21}) & \lambda_2 \\ 0 & \tilde{q}_{21} & \lambda_2 & -(\lambda_2+\tilde{q}_{21}) \end{array}\right)$$

$$\tilde{D} = \begin{pmatrix} 0 & \lambda_1 & 0 & 0 \\ \lambda_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_2 \\ 0 & 0 & \lambda_2 & 0 \end{pmatrix}$$

Fig. 3 Transition diagram of the phase process of the coupled MAP_4 and its matrices

To find the stationary distribution of the associated MAP_4, $\tilde{\boldsymbol{\pi}}$, we need to solve $\tilde{\boldsymbol{\pi}}\tilde{Q} = \mathbf{0}'$, $\tilde{\boldsymbol{\pi}}\mathbf{1} = 1$. In the same way, we can find the stationary distribution of $MTCP_4$, $\bar{\boldsymbol{\pi}}$, i.e. we have the following systems of equations:

$$\begin{cases} -(\lambda_1+\tilde{q}_{12})\tilde{\pi}_1 + \lambda_1\tilde{\pi}_2 + \tilde{q}_{21}\tilde{\pi}_3 = 0 \\ \lambda_1\tilde{\pi}_1 - (\lambda_1+\tilde{q}_{12})\tilde{\pi}_2 + \tilde{q}_{21}\tilde{\pi}_4 = 0 \\ \tilde{q}_{12}\tilde{\pi}_1 - (\lambda_2+\tilde{q}_{21})\tilde{\pi}_3 + \lambda_2\tilde{\pi}_4 = 0 \\ \tilde{\pi}_1 + \tilde{\pi}_2 + \tilde{\pi}_3 + \tilde{\pi}_4 = 1 \end{cases} \qquad \begin{cases} -\lambda_1\bar{\pi}_1 + (\lambda_1-\tilde{q}_{12})\bar{\pi}_2 + \tilde{q}_{21}\bar{\pi}_3 = 0 \\ (\lambda_1-\tilde{q}_{12})\bar{\pi}_1 - \lambda_1\bar{\pi}_2 + \tilde{q}_{21}\bar{\pi}_4 = 0 \\ \tilde{q}_{12}\bar{\pi}_1 - \lambda_2\bar{\pi}_3 + (\lambda_2-\tilde{q}_{21})\bar{\pi}_4 = 0 \\ \bar{\pi}_1 + \bar{\pi}_2 + \bar{\pi}_3 + \bar{\pi}_4 = 1 \end{cases}$$

Both of the above are uniquely solved by

$$\pi_1 = \pi_2 = \frac{\tilde{q}_{21}}{2(\tilde{q}_{12}+\tilde{q}_{21})}, \qquad \text{and} \qquad \pi_3 = \pi_4 = \frac{\tilde{q}_{12}}{2(\tilde{q}_{12}+\tilde{q}_{21})}.$$

Therefore, these two processes have the same stationary distribution $\boldsymbol{\pi}$.

Now since $\tilde{\boldsymbol{D}}\mathbf{1} = \bar{\boldsymbol{D}}\mathbf{1} = (\lambda_1, \lambda_1, \lambda_2, \lambda_2)'$ one can find from (3):

$$\mathbb{E}[\tilde{N}(t)] = \mathbb{E}[\bar{N}(t)].$$

To compute the variance, first we verify that:

$$\boldsymbol{\pi}\tilde{D} = \boldsymbol{\pi}\bar{D} = \left(\frac{\tilde{q}_{21}\lambda_1}{2(\tilde{q}_{12}+\tilde{q}_{21})} \quad \frac{\tilde{q}_{21}\lambda_1}{2(\tilde{q}_{12}+\tilde{q}_{21})} \quad \frac{\tilde{q}_{12}\lambda_2}{2(\tilde{q}_{12}+\tilde{q}_{21})} \quad \frac{\tilde{q}_{12}\lambda_2}{2(\tilde{q}_{12}+\tilde{q}_{21})}\right).$$

Explicit calculation of the fundamental matrices $\tilde{Q}^-$ and $\bar{Q}^-$ shows that even though these matrices are not the same, it holds that $\boldsymbol{\pi}\tilde{D}\tilde{Q}^-\tilde{D}\mathbf{1} = \boldsymbol{\pi}\bar{D}\bar{Q}^-\bar{D}\mathbf{1}$. In addition, by explicitly calculating the matrix exponential, we have:

$$\boldsymbol{\pi}\tilde{D}\tilde{Q}^-(e^{\tilde{Q}t} - I)\tilde{Q}^-\tilde{D}\mathbf{1} = \frac{(e^{-(\tilde{q}_{12}+\tilde{q}_{21})t} - 1)\tilde{q}_{12}\tilde{q}_{21}(\lambda_1 - \lambda_2)^2}{(\tilde{q}_{12} + \tilde{q}_{21})^4} = \boldsymbol{\pi}\bar{D}\bar{Q}^-(e^{t\bar{Q}} - I)\bar{Q}^-\tilde{D}\mathbf{1}.$$

Therefore, from (4), $\mathrm{Var}\Big(\tilde{N}(t)\Big) = \mathrm{Var}\Big(\bar{N}(t)\Big)$ and the proof is complete. □

Remark 1. Note that Proposition 2 only holds for slow MMPPs. Otherwise the construction of a MAP_4 from a given MMPP_2 does not hold due to some non-positive off-diagonal elements $\lambda_i - \tilde{q}_{ij}$ in the matrices $\bar{Q}$ and $\bar{D}$.

5 The Steady-State Queue Approximation

In this section we use the results of the previous section to approximate a given (slow) $\mathrm{MMPP}_2/\mathrm{PH}_2/1$ with an $\mathrm{MTCP}_4/\mathrm{PH}_2/1$. In general, our computations are for MAP/PH/1 queues where the service time distributions are parametrized by their workloads and their Squared Coefficient of Variations (SCVs) which we denote by c^2. We have $c^2 = \frac{1}{2}$ in the case of Erlang-2 (E_2) distribution: the sum of two i.i.d. exponential random variables with rate $\frac{2\Lambda}{\rho}$, where Λ is the arrival rate as in (3) and ρ is the workload. In the case of $c^2 = 1$, we use exponentially distributed random variables with rate $\mu = \frac{\Lambda}{\rho}$. For the case of $c^2 > 1$, we use the Hyperexponential-2 (H_2) distribution which is a mixture of two independent exponential random variables. With probability $p = \frac{1}{2c^2-1}$ we take an exponential distribution with rate $\frac{\Lambda}{c^2\rho}$ and with probability $1 - p$ we take an exponential distribution with rate $\frac{2\Lambda}{\rho}$. It is easy to verify that this H_2 random variable has mean 1 and the desired c^2.

We compute the matrix R and the stationary distribution of $\mathrm{MMPP}_2/\mathrm{PH}_2/1$ and $\mathrm{MTCP}_4/\mathrm{PH}_2/1$ as QBDs by using the SMC solver. The numerical computation for finding the relative errors, $\frac{\text{true value-approximate value}}{\text{true value}}$, shows the same properties for the curves of the relative error of mean and SCV of steady state queue for all of the above mentioned processes.

Figure 4 (left) shows different relative errors of the steady state mean for various service time SCVs. The bigger the SCV of service time, the less relative error of the mean. Figure 4 (right) shows different relative errors of the steady state SCV for various service time SCVs. The minimum absolute value of the relative error is again for the case that the service distribution is hyperexponential, i.e. the bigger the SCV of service time, the less absolute value of the relative error of SCV of steady state queue.

Both of these families of curves are bell-shaped. The only difference is that in contrast to the relative error of means which has positive values, the relative error of SCV of the steady state queue has negative values. This shows that the true value for mean is always greater than the approximate one and the opposite holds for SCV.

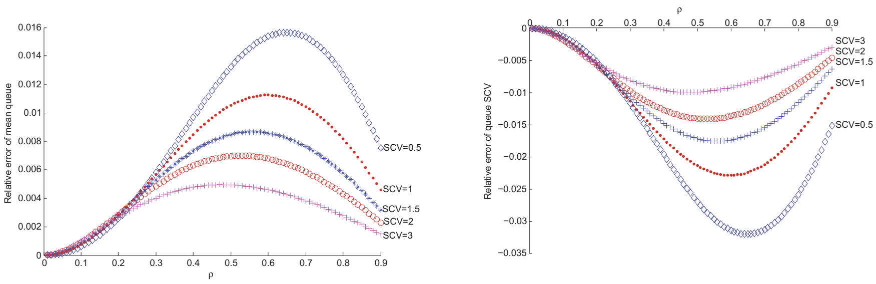


Fig. 4 The relative mean error (left) and relative SCV error (right) of a steady state queue. The $MMPP_2$ model used has $\tilde{q}_{12} = \tilde{q}_{21} = 5$, $\lambda_1 = 10$, $\lambda_2 = 20$. Then the mean service time is varied to accommodate for the desired ρ.

From further investigation of the variance (not appearing in the figures) it also holds that the true variance is less than or equal to the approximated variance.

As is evident from the figures, in any case, the relative error is negligible. Note though, that for more bursty arrival processes we may have bigger relative errors than those in the figure, yet we carried out an extensive computational study to find an empirical boundary for relative error. Assuming that λ_1 is constant (=10) and varying the values of λ_2, $\tilde{q}_{12}$ and $\tilde{q}_{21}$ gives the results in Table 1 for the maximum relative error. These empirical results indeed suggest that the MTCP/PH/1 is a very sensible alternative model to MMPP/PH/1.

Table 1 Maximum relative error of mean queue in approximation of $MMPP_2/PH_2/1$ queue by $MTCP_4/PH_2/1$ queue where $\lambda_1 = 10$. Note that the H_2 case corresponds to $c^2 = 1.1$.

Model	λ_2	$\tilde{q}_{12}$	$\tilde{q}_{21}$	Max Relative Error of Mean Queue
$MMPP_2/E_2/1$	500	8	70	0.0893
$MMPP_2/M/1$	300	9	70	0.0725
$MMPP_2/H_2/1$	400	5	70	0.0715

6 Conclusions and Future Work

As illustrated in this paper, MMPPs can perhaps be replaced by MTCPs for modelling purposes. We have shown a theoretical relationship between the two processes and an empirical relationship between their associated queueing models. Our focus in this conference paper is on being expository, hence we focused on the case of $p = 2$. A question that arises is: "Can we construct an MTCP to match a non-slow MMPP with the same mean and variance?".

In further work we plan to handle the general case, for $p > 2$, where we believe similar results may hold. Proving the empirical bounds that we found for the queueing approximations remains a challenge.

Of further interest is the issue of parameter estimation of MTCPs. Our belief is that since data traces generated by MTCPs are more informative than those generated by MMPPs, there is a promise in devising a good parameter estimation method for MTCPs.

Acknowledgments We thank Sophie Hautphenne for useful discussions and advice. Yoni Nazarathy is supported by Australian Research Council (ARC) grants DP130100156 and DE130100291.

Azam Asanjarani is supported by DE130100291.

References

1. Asmussen, S.: Phase-type representations in random walk and queueing problems. Ann. of Prob., 772–789 (1992)
2. Asmussen, S.: Matrix-analytic models and their analysis. Scand. Jour. of Stat. **27**(2), 193–226 (2000)
3. Asmussen, S.: Applied probability and queues. Stochastic Modelling and Applied Probability, vol. 51. Springer (2003)
4. Bini, D.A., Meini, B., Steffé, S., Van Houdt, B.: Structured Markov chains solver: software tools. In: Proc. 2006 Workshop on Tools for Solving Structured Markov Chains, Article No. 14. ACM (2006)
5. Fischer, W., Meier-Hellstern, K.: The Markov-modulated Poisson process (MMPP) cookbook. Perf. Eval. **18**(2), 149–171 (1993)
6. Fomundam, S., Herrmann, J.W.: A survey of queuing theory applications in healthcare (2007)
7. Gun, L.: An Algorithmic Analysis of the MMPP/G/1 Queue (No. ISR-TR-88-40). Maryland Univ. College Park Inst. For Systems Research (1988)
8. Horváth, A., Telek, M.: Markovian modeling of real data traffic: heuristic phase type and MAP fitting of heavy tailed and fractal like samples. In: Calzarossa, M.C., Tucci, S. (eds.) Performance 2002. LNCS, vol. 2459, pp. 405–434. Springer, Heidelberg (2002)
9. Latouche G., Ramaswami, V.: Introduction to matrix analytic methods in stochastic modeling, vol. 5. SIAM (1999)
10. Lucantoni, D.M.: The BMAP/G/1 queue: a tutorial. In: Donatiello, L., Nelson, R. (eds.) SIGMETRICS 1993 and Performance 1993. LNCS, vol. 729, pp. 330–358. Springer, Heidelberg (1993)
11. Nazarathy, Y., Weiss, G.: The asymptotic variance rate of the output process of finite capacity birth-death queues. Queueing Sys. **59**(2), 135–156 (2008)
12. Rudemo, M.: Point processes generated by transitions of Markov chains. Adv. Appl. Prob., 262–286 (1973)
13. Ramesh, N.I.: Statistical analysis on Markov-modulated Poisson processes. Environmetrics **6**(2), 165–179 (1995)
14. Riska, A., Squillante, M., Yu, S.Z., Liu, Z., Zhang, L.: Matrix-analytic analysis of a MAP/PH/1 queue fitted to web server data. Matrix-Analytic Methods; Theory and Applications, 333–356 (2002)