Characterising Traffic Streams in Networks of MAP|MAP|1 Queues

Ramin Sadre, Boudewijn Haverkort
Laboratory for Performance Evaluation and Distributed Systems
Department of Computer Science, RWTH-Aachen, D-52056 Aachen
{ramin|haverkort}@lvs.informatik.rwth.aachen.de

Abstract

In this paper we present an approach to characterise, in an approximate but very accurate fashion, the output process of a MAP|MAP|1 queue as a MAP. In our approximations, the interdeparture time distribution is correctly represented; the correlation structure is approximated to any required degree of accuracy. We derive our approximations directly from the queueing process itself. Various case studies show the feasability and accuracy of our approach.

By combining our results with the exact results that are available to merge and split MAPs, we have established all ingredients for a decomposition procedure for open networks of MAP|MAP|1 queues. In a separate paper we will report on the details of such a procedure.

1 Introduction

Efficient and exact algorithms for the evaluation of open queueing network models only exist when the considered models are of a very simple class. Best-known in this context are Jackson queueing networks [7], in which the service and interarrival time distributions are negative exponential, and in which the routing is fixed and Markovian. A decomposition via the solution of the (first-order) traffic equations then naturally leads to the study of single queues at a time. However, to model modern network systems, more realistic packet ("job") arrival streams have to be coped with, as well as non-exponential service times. In such cases, the decomposition using the traffic equations becomes more complex and, in fact, only possible in an approximate way. Early work in this area, that is, for the approximate analysis of open networks of G|G|1 queues, has been reported by Kühn [8] and Whitt [17, 16]. An extension ot this approach has been developed by Haverkort *et al.* [3, 14] in which the service and interarrival time distributions are of phase-type and in which the individual queues are analysed using matrix-gemeotric methods. This is possible due to the quasi-birth-death structure of the resulting per-queue underlying Markov chains. Note that in the latter approach, also finite-buffer queues can be handled.

In this paper, we take the previous analyses a step further. In particular, we attempt to study networks of MAP|MAP|1 queues, that is, networks of queues in which both the service and arrival processes are non-renewal and described as Markovian arrival processes (MAPs; see Section 2). The motivation for this choice is the fact that complex arrival patterns, for instance those appearing in the Internet, can best be described using

MAPs. Furthermore, also complex services that may depend upon one another can also be described using MAPs. Attempts to approximate the non-renewal service and arrival processes with renewal processes often lead to inaccurate analysis results. In order to still study networks of such queues, we have to characterise the output process of MAP|MAP|1 queues, in order to use these as input processes for queues further downstream in the queueing network. Unfortunately, the output process of a MAP|MAP|1 queue is not a finite MAP. Hence, the aim of this paper is to construct approximate representations of the output process of MAP|MAP|1 queues that are still finite MAPs. If we succeed in this, we can develop a decomposition approach as has been done in [3, 14], but now in an more general, non-renewal context.

The problem we are addressing in this paper is not new and has been recognised by authors as well. Recently, Green *et al.* [2] analysed the output process of MAP|PH|1 queues; also in their case, the output process is not a finite MAP. Heindl recently proposed to approximate the output process of an MMPP|G|1[|K] queueing systems using a semi-Markov process [4], to be used for the analysis of tandem queueing networks. In this paper we extend the approach of Green *et al.* in that we consider MAPs as service processes, and in that we allow for more freedom in the process of making the infinite MAP finite. We compare our results with the approach based on renewal processes (as advocated in [3, 14]), as well as with discrete-event simulations. Early results of our implementation are promising.

The paper is further organised as follows. In Section 2 we introduce MAPs and discuss the most important properties of such processes. Section 3 then presents the analysis of the MAP|MAP|1 queue, thereby focussing on the exact departure process, as well as on the families of departure process approximations. In Section 4 we then present evaluation results and comparisons to discrete-event simulations. Section 5 concludes the paper.

2 The Markovian arrival process

We first consider MAPs in Section 2.1. Then, we discuss their interarrival time characteristic in Section 2.2 and their joining and splitting in Section 2.3.

2.1 Definition

MAPs are stochastic processes where the interarrival times are controlled by a continuous-time Markov chain with marked transitions. Such a CTMC has an infinitesimal generator matrix $\mathbf{Q}$ which is splitted into two matrices $\mathbf{Q_0}$ and $\mathbf{Q_1}$ as follows:

$$\mathbf{Q_0} = \begin{pmatrix} -q_1 & q_{12} & \cdots & q_{1m} \\ q_{21} & -q_2 & \cdots & q_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ q_{m1} & q_{m2} & \cdots & -q_m \end{pmatrix}, \quad \mathbf{Q_1} = \begin{pmatrix} 0 & a_{12} & \cdots & a_{1m} \\ a_{21} & 0 & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & 0 \end{pmatrix},$$

with $\mathbf{Q} = \mathbf{Q_0} + \mathbf{Q_1}$ and $q_i = \sum_{j=1, j\neq i}^{m}(q_{ij} + a_{ij})$. The elements of the matrix $\mathbf{Q_1}$ give the transition rates of the marked transitions: passing through such a marked transition triggers an arrival event. Some general results of the Markov-modulated Poisson process (MMPP) [15] can be easily adapted to the MAP. In the following we will state the most important ones.

2.2 Interarrival time characteristics

In order to compute the behaviour of a MAP $(\mathbf{Q_0}, \mathbf{Q_1})$ we first need to choose the initial probability vector $\underline{p}$ of the MAP. In analogy to phase-type distributions we start the MAP at an "arbitrary" arrival epoch by choosing

$$\underline{p} = \frac{1}{\underline{\pi}\mathbf{Q_1}\underline{e}}\underline{\pi}\mathbf{Q_1},$$

where $\underline{\pi}$ is the steady state probability vector of the MAP, i.e., $\underline{\pi}(\mathbf{Q_0} + \mathbf{Q_1}) = \underline{0}$. The thus-obtained process is said to be *interval-stationary* [15]. The interarrival time distribution function of the interval-stationary process is given by

$$F(t) = 1 - \underline{p}\exp(\mathbf{Q_0}t)\underline{e},$$

which leads to the following expression for the kth moment of the interarrival time [15]:

$$E[T^k] = k!\underline{p}(-\mathbf{Q_0})^{-(k+1)}\mathbf{Q_1}\underline{e}.$$

Let T_i be the time between the ith and the $(i+1)$st arrivals in a MAP. Then, the autocovariance function $R(k)$ for T_1 and T_{k+1} with $k \geq 1$ is given by

$$\begin{aligned} R(k) &= E\left[(T_1 - E[T_1])(T_{k+1} - E[T_{k+1}])\right] \\ &= \underline{p}(-\mathbf{Q_0})^{-2}\mathbf{Q_1}\left\{\left[(-\mathbf{Q_0})^{-1}\mathbf{Q_1}\right]^{k-1} - \underline{e}\underline{p}\right\}(-\mathbf{Q_0})^{-2}\mathbf{Q_1}\underline{e}. \end{aligned}$$

The limiting index of dispersion I of a MAP is given by [4]

$$I = \lim_{t\to\infty} \frac{\mathrm{Var}[N(t)]}{E[N(t)]} = 1 + 2\left(\underline{\pi}\mathbf{Q_1}\underline{e} - \frac{1}{\underline{\pi}\mathbf{Q_1}\underline{e}}\underline{\pi}\mathbf{Q_1}(\mathbf{Q} + \underline{e}\underline{\pi})^{-1}\mathbf{Q_1}\underline{e}\right),$$

where $N(t)$ is the counting process of the MAP.

2.3 Joining and splitting

The class of MAPs is closed under joining and probabilistic splitting. The superposition of two MAPs $(\mathbf{A_0}, \mathbf{A_1})$ and $(\mathbf{B_0}, \mathbf{B_1})$ is a new MAP $(\mathbf{C_0}, \mathbf{C_1})$ with

$$\mathbf{C_0} = \mathbf{A_0} \oplus \mathbf{B_0}, \quad \mathbf{C_1} = \mathbf{A_1} \oplus \mathbf{B_1},$$

where $\mathbf{L} \oplus \mathbf{M} = \mathbf{L} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{M}$, and $\otimes$ is the tensor product operator.

The probabilistic splitting of a MAP $(\mathbf{A_0}, \mathbf{A_1})$ with probability r gives two MAPs $(\mathbf{B_0}, \mathbf{B_1})$ and $(\mathbf{C_0}, \mathbf{C_1})$ with

$$\begin{aligned} (\mathbf{B_0}, \mathbf{B_1}) &= (\mathbf{A_0} + (1-r)\mathbf{A_1}, r\mathbf{A_1}), \\ (\mathbf{C_0}, \mathbf{C_1}) &= (\mathbf{A_0} + r\mathbf{A_1}, (1-r)\mathbf{A_1}). \end{aligned}$$

3 The MAP|MAP|1 queue

We first present the CTMC underlying a MAP|MAP|1 queue in Section 3.1, after which we discuss the departure process of such a queue in detail in Section 3.2.

3.1 The underlying Markov chain

In a MAP|MAP|1 queue both the arrival process and the service process are MAPs. The underlying infinite Markov chain of such a queue can be considered as a special case of a quasi-birth-and-death process (QBD) [10] where each *level* of the QBD state space corresponds to a specific number of customers in the queueing system.

Let $(\mathbf{\Lambda_0}, \mathbf{\Lambda_1})$ be the arrival MAP with l states and $(\mathbf{S_0}, \mathbf{S_1})$ the service MAP with m states. Then the underlying QBD has the following generator matrix

$$\mathbf{Q} = \begin{pmatrix} \mathbf{B_0} & \mathbf{A_0} & & \\ \mathbf{B_1} & \mathbf{A_1} & \mathbf{A_0} & \\ \mathbf{0} & \mathbf{A_2} & \mathbf{A_1} & \\ & \ddots & \ddots & \end{pmatrix}, \qquad (1)$$

with

$$\begin{aligned} \mathbf{B_0} &= \mathbf{\Lambda_0} \otimes \mathbf{I}, \\ \mathbf{B_1} &= \mathbf{I} \otimes \mathbf{S_1}, \\ \mathbf{A_0} &= \mathbf{\Lambda_1} \otimes \mathbf{I}, \\ \mathbf{A_1} &= \mathbf{\Lambda_0} \oplus \mathbf{S_0}, \\ \mathbf{A_2} &= \mathbf{I} \otimes \mathbf{S_1}. \end{aligned}$$

The steady-state probability vector $\underline{\pi}$ has the following structure

$$\underline{\pi} = \begin{pmatrix} \underline{v_0} & \underline{v_1} & \cdots \end{pmatrix},$$

where each $\underline{v_i}$ is a vector of size $l \cdot m$ and contains the steady-state probabilities for the states of level i in the QBD. In the following, we assume that matrix-geometric solution methods, like the LR-approach [1], are used to compute $\underline{\pi}$. Matrix-geometric solution methods are based on the fact that the vectors $\underline{v_i}$ have the so-called matrix-geometric form

$$\underline{v_i} = \underline{v_0}\mathbf{R}^i, \quad \mathbf{R} \in R^{lm \times lm}, \quad i = 0, 1, \ldots,$$

where $\mathbf{R}$ is the entry-wise smallest non-negative solution of the matrix-quadratic equation

$$\mathbf{A_0} + \mathbf{R}\mathbf{A_1} + \mathbf{R}^2\mathbf{A_2} = \mathbf{0}.$$

Having obtained $\underline{\pi}$, many performance measures of the queue can be computed easily, e.g., the moments of the queue length distribution:

$$E[N^k] = \sum_{i=0}^{\infty} i^k \underline{v_i}\,\underline{e} = \sum_{i=0}^{\infty} i^k \underline{v_0}\mathbf{R}^i \underline{e},$$

which yields, in case $k = 1$:

$$E[N] = \underline{v_0}\mathbf{R}(\mathbf{I} - \mathbf{R})^{-2}\underline{e}.$$

3.2 The departure process and its approximations

We present the exact departure process in Section 3.2.1 and present a simple and a more complex approximation in Section 3.2.2 and 3.2.3, respectively.

3.2.1 The infinite departure process

When analysing the departure process of a queue one generally is interested in a description that is of the same class as the description of the arrival process, since then one is able to analyse networks of queues by simply feeding the output of a queue as input for the subsequent queue after the splitting and joining of traffic streams has been handeld. However, in general, queues with infinite buffer produce departure processes with an infinite structure.

In the case of MAP|MAP|1 queues, the output process is an infinite MAP $(\mathbf{D_0}, \mathbf{D_1})$ which can easily be obtained by modifying the underlying Markov chain of the queue: since each service completion corresponds to an "arrival" in the departure process we obtain $\mathbf{D_0}$ and $\mathbf{D_1}$ directly from the generator (1) by marking the transitions of the sub-matrices $\mathbf{B_1}$ and $\mathbf{A_2}$ as "arrival" transitions, as follows:

$$\mathbf{D_0} = \begin{pmatrix} \mathbf{B_0} & \mathbf{A_0} & & \\ \mathbf{0} & \mathbf{A_1} & \mathbf{A_0} & \\ & \mathbf{0} & \mathbf{A_1} & \\ & & \ddots & \ddots \end{pmatrix}, \quad \mathbf{D_1} = \begin{pmatrix} \mathbf{0} & & & \\ \mathbf{B_1} & \mathbf{0} & & \\ & \mathbf{A_2} & \mathbf{0} & \\ & & \ddots & \ddots \end{pmatrix}.$$

This representation of the departure process is exact but —due to its infinite size— not practical for further use, e.g., as arrival MAP for a subsequent MAP|MAP|1 queue. Hence, our primary goal is to find a finite approximation of the departure MAP which is as small as possible but well matches the characteristics, i.e., the interarrival time distribution and the correlation structure, of the original MAP.

3.2.2 Simple finite approximation of the departure process

In [13], Green presented an approach to approximate the departure process of MAP|PH|1 queues by a finite MAPs. This approach can be easily extented[1] to MAP|MAP|1 queues. The thus obtained MAP will have two important characteristics:

1. It *exactly* describes the inter-departure time distribution of the original departure process.
2. It is able to describe the correlation structure of the original departure process *at arbitrary precision* albeit at the cost of an arbitrary large state space.

The second characteristic implies that, depending on the required quality of the approximation, the obtained MAP may become so large that other reduction methods must be applied afterwards.

The approximation is based on the structural equality of the infinite departure MAP and the underlying CTMC of the QBD: each level in the QBD has its corresponding entries in the matrices of the departure MAP. However, it often is not required to represent all levels of the QBD in the departure process. Since the probability $\underline{v_i}\underline{e}$ to be in level i decreases with increasing i, there is an integer s where $\underline{v_s}\underline{e}$ becomes so small that one may decide not to represent the levels s and higher in full detail in the departure process.

In the following we assume $s > 1$, i.e., level s is not adjacent to the border level 0. We transform the infinite (exact) departure process into a finite (approximated) departure process by replacing the levels s and higher by only *one* level, the s-called called *clipping level* $\hat{s}$. The transition rates leaving and entering level $\hat{s}$ are computed as follows:

[1]Actually, we developed our approach while not being aware of [13].

- Changing to level $\widehat{s}$ from level $s-1$ corresponds to changing to level s from $s-1$ in the original departure process.

- Staying in level $\widehat{s}$ without an arrival in the departure process corresponds to staying in level s' or changing from level s' to $s'+1$ (for all $s' \geq s$) in the original departure process.

- Finally, an arrival in the original departure process from a state i in the level $s' \in \{s, \dots\}$ may either lead to:

  (i) level $s'-1 = s-1$: this corresponds to an arrival that leads from level $\widehat{s}$ to $s-1$, which happens with probability

  $$\frac{1}{|v_{s,i}| + |v_{+,i}|} v_{s,i},$$

  (ii) level $s'-1 \neq s-1$: this corresponds to an arrival that leads from level $\widehat{s}$ to $\widehat{s}$, which happens with probability

  $$\frac{1}{|v_{s,i}| + |v_{+,i}|} v_{+,i},$$

  where $\underline{v_+}$ is the steady-state probability to be in a level higher than s in the original QBD, that is:

  $$\begin{aligned} \underline{v_+} &= \sum_{i>s} \underline{v_i} = \sum_{i=s+1}^{\infty} \underline{v_0}\mathbf{R}^i \\ &= \underline{v_0}\mathbf{R}^{s+1}(\mathbf{I}-\mathbf{R})^{-1}. \end{aligned}$$

These considerations yield the finite appromixation MAP of the departure process:

$$\mathbf{D_0} = \begin{pmatrix} \mathbf{B_0} & \mathbf{A_0} & & & \\ \mathbf{0} & \mathbf{A_1} & \mathbf{A_0} & & \\ & \mathbf{0} & \mathbf{A_1} & & \\ & \ddots & \ddots & \ddots & \\ & & & \mathbf{A_0} & \\ & & & \mathbf{A_0}+\mathbf{A_1} & \end{pmatrix}, \text{ and}$$

$$\mathbf{D_1} = \begin{pmatrix} \mathbf{0} & & & & \\ \mathbf{B_1} & \mathbf{0} & & & \\ & \mathbf{A_2} & 0 & & \\ & & \ddots & \ddots & \\ & & & \mathbf{0} & \\ & & & \mathbf{A_{2,down}} & \mathbf{A_{2,stay}} \end{pmatrix},$$

with

$$\begin{aligned} (\mathbf{A_{2,down}})_{ij} &= \frac{1}{v_{s,i}+v_{+,i}} v_{s,i} A_{2,ij}, \\ (\mathbf{A_{2,stay}})_{ij} &= \frac{1}{v_{s,i}+v_{+,i}} v_{+,i} A_{2,ij}. \end{aligned}$$

3.2.3 Extended approximation method

In the previous section we explained how the set $\{s, s+1, \ldots\}$ of levels for given s can be merged into one level. Certainly, this merging operation can be applied to any finite set $\{s, s+1, \ldots, t\}$ of levels, too. All we need to know are the probabilities to be in one of the border levels s or t of the set and the probability to be in the set $\{s, s+1, \ldots, t\}$ which are given by subvectors of $\underline{\pi}$. We define $\underline{v_{seq}} = \sum_{i=s}^{t} \underline{v_i}$. Replacing the levels s through t in the original departure process by one level results in the following MAP:

$$
\mathbf{D_0} = \begin{pmatrix}
\mathbf{B_0} & \mathbf{A_0} & & & & \\
\mathbf{0} & \mathbf{A1} & & & & \\
& 0 & \ddots & & & \\
& & & \mathbf{A_0} & & \\
& & & \mathbf{A_1} + \mathbf{A_{0,stay}} & \mathbf{A_{0,up}} & \\
& & & 0 & \mathbf{A_1} & \\
& & & & 0 & \ddots
\end{pmatrix},
$$

$$
\mathbf{D1} = \begin{pmatrix}
0 & & & & & \\
\mathbf{B_1} & 0 & & & & \\
& \mathbf{A_2} & \ddots & & & \\
& & & 0 & & \\
& & & \mathbf{A_{2,down}} & \mathbf{A_{2,stay}} & \\
& & & & \mathbf{A_2} & \ddots
\end{pmatrix},
$$

with

$$
\begin{aligned}
(\mathbf{A_{0,up}})_{ij} &= \frac{1}{v_{seq,i}} v_{t,i} A_{0,ij}, \\
(\mathbf{A_{0,stay}})_{ij} &= \frac{1}{v_{seq,i}} (v_{seq,i} - v_{t,i}) A_{0,ij}, \\
(\mathbf{A_{2,stay}})_{ij} &= \frac{1}{v_{seq,i}} (v_{seq,i} - v_{s,i}) A_{2,ij}, \\
(\mathbf{A_{2,down}})_{ij} &= \frac{1}{v_{seq,i}} v_{s,i} A_{2,ij}.
\end{aligned}
$$

Combined with the method of the previous section we are able to create finite approximations of the departure process whose quality and size can be controlled in many ways by choosing the number and the position of the here described compacting operations. It should be noted that this approach, as well as the approach presented in the previous section, always yields a MAP with exactly the same interarrival distribution as the original departure process.

4 Evaluation of the approximation methods

In Section 4.1 we evaluate our approach when used with renewal external arrivals, whereas we apply non-renewal arrival processes in Section 4.2. We study effects of queue utilisation in Section 4.3.

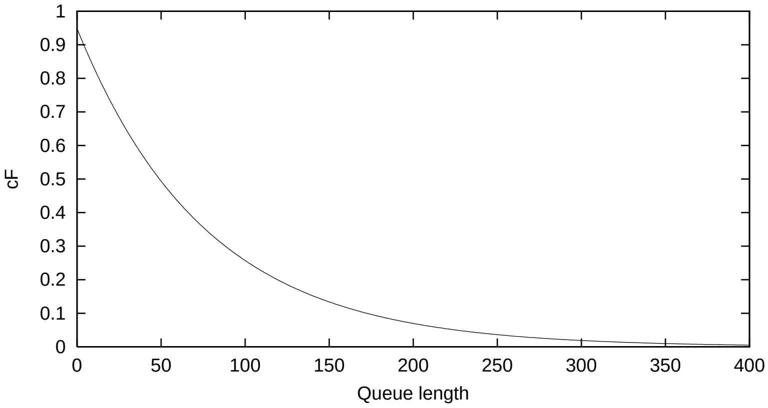

Figure 1: Complementary distribution function (cF) of the queue length (Case 1)

4.1 Case 1: renewal input processes

The queueing system studied

We start our evaluation with the analysis of a simple MAP|MAP|1 queue using an Erlang-2 service process with service rate 1.9. The queue is fed by a hyper-exponential renewal process with MAP representation

$$\left(\begin{pmatrix} -3.39 & 0 \\ 0 & -0.21 \end{pmatrix}, \begin{pmatrix} 3.19 & 0.2 \\ 0.2 & 0.01 \end{pmatrix} \right).$$

The arrival rate and coefficient of variation are 1.8, respectively 8.0. When analysing this queue we can see that the high load of nearly 95% and the high variance of the input process lead to a high mean queue length of 73.11 with the corresponding slowly decaying queue length distribution function (Figure 1).

Characteristics of the approximated departure process

First, we investigate the departure process approximated by the simple approximation method presented in Section 3.2.2. How does the quality of the approximation depend on the choice of the clipping level? In view of the remarks made about the queue length distribution it is clear that a high clipping level is required to preserve the characteristics of the original departure process. To illustrate this, we have computed the approximated departure process for four different clipping levels, namely 17, 33, 49 and 65. For these cases, the autocovariance functions $R(k)$, as defined in Section 2.2, are shown in Figure 2. Choosing 33 as clipping level seems to result in a good approximation to the autocovariance function of the exact departure process (which has not been shown here because it nearly is equal to the 65-approximation).

These results show that even MAPs with low clipping level can give a good approximation of the correlation structure at small time lags k. For the approximation of correlation at large time lags, MAPs must be choosen which respect the structure of the higher levels of the queue QBD. Figure 2 shows that the autocovariance function of MAPs with low clipping level quickly drops to 0 (i.e., to the behaviour of uncorrelated renewal processes) with increasing time lag. However, the figure additionally shows that clipping levels which are considerably smaller than the mean queue length yield good approximations to the exact departure process. This may be surprising since clipping levels like 65 only catch 55% of the queue length distribution probability (Figure 1). But we should note that many important characteristics of queue processes, like the busy period, are mainly represented

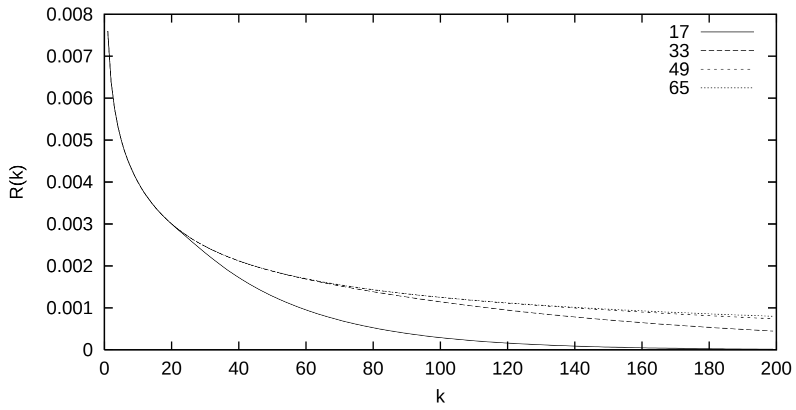

Figure 2: Autocovariance function $R(k)$ of the approximate departure MAP for clipping levels 17, 33, 49 and 65 (Case 1)

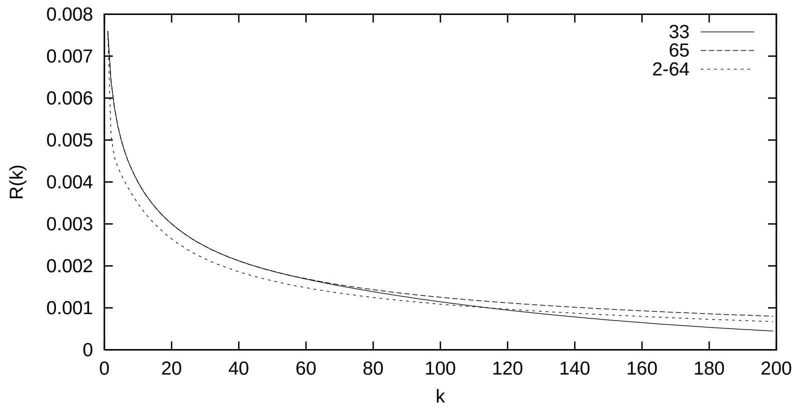

Figure 3: Autocovariance function $R(k)$ of the 2-64-level MAP (Case 1)

by the lower levels. In fact, Green proved for MAP|PH|1 queues in [2] that the k-level approximation captures the first $k-1$ correlation coefficients of the departure process exactly.

As next step, we use the fact that high clipping levels lead to increased correlation for large time lags by introducing the approximation method presented in Section 3.2.3. To evaluate its impact, we choose the 33-level MAP as reference. It is of moderate size (134 states; which can be easily handled computationally) and is a good approximation to the exact solution. We now compare the 33-level MAP with a MAP *of the same size* but constructed as follows: (i) levels 0 and 1 are modelled without approximation, (ii) levels 2 through 63 are merged two by two, and (iii) the levels beyond 63 are merged into one level. This MAP (called 2-64-level MAP) is compared with the 33-level and the 65-level MAP in Figure 3. For $k \geq 110$ this MAP better approximates the exact solution as the 33-level MAP. However for $k \leq 110$ its autocovariance is apparently lower than the 33-level solution: since the 2-64-level MAP approximates the levels from 2 through 63 two by two it gives a worse description of correlation between subsequent queue departure.

Even slightly better approximations can be obtained by using more complex level schemes, e.g., a 4-40-84-level MAP with the following structure: (i) levels 0 through 3 are modelled exactly, (ii) levels 4 through 39 are merged two by two and (iii) levels 40 through 83 are merged four by four (clipping level is 84). Figure 4 shows this MAP in comparision with the other approximations.

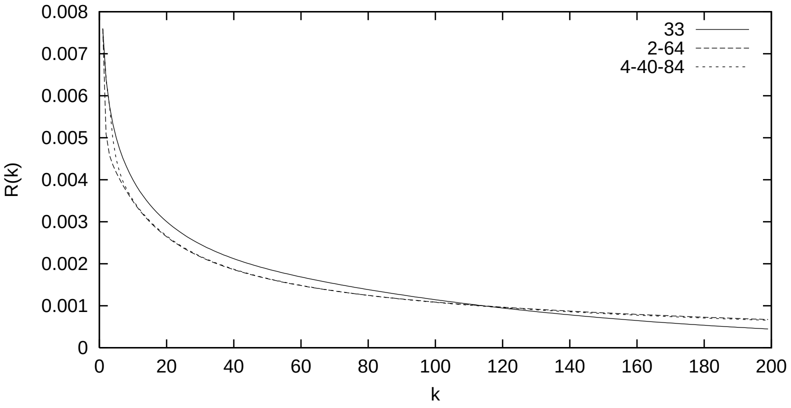

Figure 4: Autocovariance function $R(k)$ of the 4-40-84-level MAP (case 1)

	$E[N]$	RE
Simulation	22.2 ± 1.4	
FiFiQueues	16.3	26.6%
33	19.8	10.8%
2-64	20.8	6.4%
4-40-84	21.0	5.4%
46	21.0	5.4%

Table 1: Mean queue length for different approximations (Case 1)

Analysis results of MAP|MAP|1 tandem queues

In the previous section we have presented and compared different approximations to the output process of an MAP|MAP|1 queue. These approximations have the same interarrival time distribution but differ in higher order statistics as the autocovariance. It is well known that the presence of autocorrelation in arrival processes heavily impacts the performance measures of queueing systems [12]. Hence, in this section, we will examine how the different approximations affect the performance of a subsequent second queue which takes the (approximated) departure process of the first queue as arrival process.

As for the first queue, we choose for this queue as service process an Erlang-2 process with service rate 1.9. Table 1 shows the mean value of its queue length as a function of the chosen approximated departure process. The first row also presents the 95% confidence interval produced by simulation. The second row states the results computed by the queueing analysis tool FiFiQueues [14] which approximates queue departure processes by renewal processes. The remaining rows show the results obtained with the new method and compares them with the simulation. The relative error (RE) stated in the table is defined as

$$\frac{\text{simulation} - \text{approximation}}{\text{simulation}} \cdot 100\%.$$

As reported in other publications, these cases shows that increased correlation leads to higher mean queue lengths. This can be explained in a quite intuitive manner by looking at the power spectrum (see also Appendix A) of the discrete function $f(k) = T_k$ where T_k is the interarrival time as defined in Section 2.2. In Figure 5, we show the power spectrum for clipping levels 2, 6 and 11; we see that low-frequency energy, i.e., around $\omega = 0$, increases with the clipping level. This is an important fact, since San-qi Li has shown in [9] that input power in the low-frequency band has a dominent impact on queueing performance.

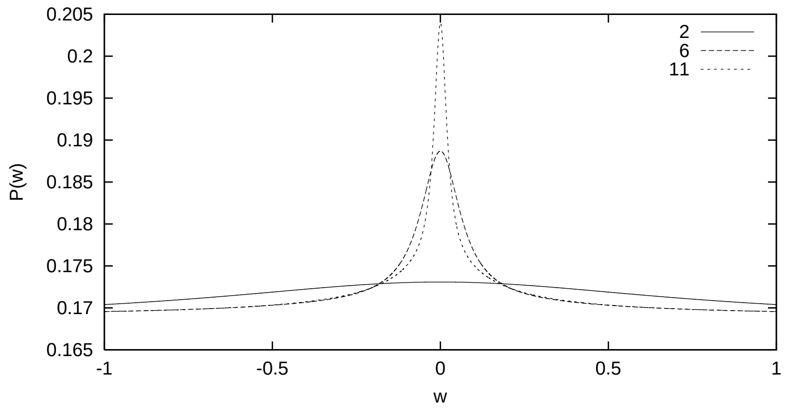

Figure 5: The power spectrum of MAPs for three different clipping levels

	$E[N]$	RE
Simulation	27.7 $\pm$ 2.3	
FiFiQueues	24.1	13.0%
33	26.4	4.7%
4-40-84	27.4	1.1%

Table 2: Mean queue length for different approximations (Case 2; Poisson arrivals)

Low frequencies produce long periods of increased workload entering the queue [12].

This example also shows that better approximations do not necessarily require more states, as becomes clear from the last row: a 4-40-84-level MAP with *134* states produces the same mean queue length as a 46-level MAP with *186* states. This is an significant improvement since the complexity of MAP|MAP|1 queue analysis cubically depends on the number of states in the MAP describing the arrival process. It seems that the good modelling of correlation for large time lags, as provided by the 4-40-84-level MAP, is more important than for small time lags. This observation confirms the results of studies made in the area of self-similar traffic which report that correlation over many time scales may noticeably affect the performance of queueing systems [11].

4.2 Case 2: true Markovian arrival processes

In this second case, we show that comparably good results (as before) are obtained with true Markovian arrival processes. The queueing system under study remains unchanged (two queues with Erlang-2 service and service rate 1.9). For the sake of compactness we present only the results of the tandem system analysis; the results for the departure process per se are as good as before.

Table 2 shows the mean queue length at the second queue, obtained by feeding a Poisson process with arrival rate 1.86 into the first queue. The mean queue length of the first queue is 35.1. Again, our results are rather good.

Our next arrival process is the superposition of an hyper-exponential renewal process with arrival rate 1.4 and squared coefficient of variation 8.0 with an Erlang-2 process with arrival rate 0.4. The thus-obtained process has the following characteristics:

$$\begin{aligned} \text{arrival rate} &= 1.8, \\ c^2 &= 2.2. \end{aligned}$$

Since this process is not a renewal process, its limiting index of dispersion for counts

	$E[N]$	RE
Simulation	20.2 ± 2.0	
33	18.9	6.4%
2-64	19.2	4.9%
4-40-84	19.4	4.0%

Table 3: Mean queue length for different approximations (Case 2; non-renewal arrivals)

	$[\mu = 1.9]$ $E[N]$	$[\mu = 1.7]$ $E[N]$
Simulation	7.4	45.0
22	7.16	38.6
2-42	7.09	39.2

Table 4: Mean queue length for different arrival processes (Case 3)

considerably differs from its squared coefficient of variation: $I = 6.3$. The mean queue length of the first queue is 58.2. The results for the mean queue length of the second queue are shown in Table 3. The confidence interval for the simulation result is ± 2 with 95% confidence level. Although slightly worse as before, the results are still acceptable.

4.3 Case 3: different load scenarios

Finally, we present a scenario for which the more complex departure process approximation does not improve the simple clipping-level approximation. We take the same tandem queueing system as in Case 1, but lower the arrival intensity to 1.6. This results in a mean queue length of 19.3 for the first queue. The mean queue length of the second queue is shown in Table 4, in case the service rate $\mu = 1.9$ (middle column) or 1.7 (right column).

Here, we can find an interesting fact: the last two rows in the middle column (Case $\mu = 1.9$) show that the 2-42-level approximation provides worse results than the 22-level approximation. Apperantly, this effect is not present for $\mu = 1.7$ (right column). This may be explained by the fact that for $\mu = 1.9$ the queue runs with a substantially lower load than in all other experiments presented in this paper. This seems to lead to a lower sensitivity to long-time correlation and the 2-42-level MAP loses its advantage against the 22-level approximation. However, more investigations are needed here, to make this more firm.

5 Conclusion

In this paper we have presented an approach to characterise, in an approximate but very accurate fashion, the output process of a MAP|MAP|1 queue as a MAP. In any case, the interarrival time distribution is correctly represented; the correlation structure is approximated to any required degree of accuracy. By combining this result with the exact results that are available to merge and split MAPs, we have established all ingredients for a decomposition procedure for open networks of MAP|MAP|1 queues. In a separate paper we will report on that; most probably we will incorporate the new technique in the FiFiQueues tool [14]. It should be noted at this point, that the MAP respresentations of the traffic streams may become rather large, especially in case we allow for feedback loops in the queueing network. We will then have to use algorithms to reduce the size of the MAPs. We are currently investigating the use of Markovian bisimulation algorithms [5, 6] to perform such reductions; early results are promising.

References

[1] G. Latouche and V. Ramaswami. A logarithmic reduction algorithm for quasi birth and death processes. *Journal of Applied Probability*, 30:650–674, 1993.

[2] D. Green. Lag correlations of approximating departure processes for MAP|PH|1 queues. In G. Latouche and P. Taylor, editors, *Advances in Algorithmic Methods for Stochastic Models — Proceedings of the 3rd International Conference on Matrix Analytic Methods*, pages 135–151. Notable Publications, Inc., 2000.

[3] B. R. Haverkort. Approximate analysis of networks of PH|PH|1|K queues with customer losses: Test results. *Annals of Operations Research*, 79:271–291, 1998.

[4] A. Heindl. Decomposition of general tandem queueing networks with MMPP input. *Performance Evaluation*, 44:5–23, 2001.

[5] H. Hermanns. *Interactive Markov Chains.* PhD thesis, University of Erlangen-Nürnberg, 1999.

[6] H. Hermanns and J.-P. Katoen. Automated compositional Markov chain generation for a plain-old telephone system. *Science of Computer Programming*, 36(1):97–127, 2000.

[7] J.R. Jackson. Networks of waiting lines. *Operations Research*, 5:518–521, 1957.

[8] P.J. Kühn. Approximate analysis of general queueing networks by decomposition. *IEEE Transactions on Communications*, 27(1):113–126, 1979.

[9] San-qi Li. Queue response to input correlation functions: Continuous spectral analysis. *IEEE/ACM Transactions on Networking*, 1(6):678–692, December 1993.

[10] M. F. Neuts. *Matrix-Geometric Solutions in Stochastic Models — An Algorithmic Approach.* Dover Publications, Inc., 1981.

[11] M. Grossglauser and J.-C. Bolot. On the relevance of long-range dependence in network traffic. *IEEE/ACM Transactions on Networking*, 7(5):629–640, October 1999.

[12] M. Livny, B. Melamed, and A. K. Tsiolis. The impact of autocorrelation on queueing systems. *Management Science*, 39(3):322–339, March 1993.

[13] N. G. Bean, D. A. Green, and P. G. Taylor. Approximations to the output process of MAP|PH|1 queues. In *Advances in Matrix Analytic Methods for Stochastic Models — Proceedings of the 2nd International Conference on Matrix Analytic Methods*, pages 151–159. Notable Publications, Inc., 1998.

[14] R. Sadre, B.R. Haverkort, and A. Ost. An efficient and accurate decomposition method for open finite- and infinite-buffer queueing networks. In W. Stewart and B. Plateau, editors, *Proc. 3rd Int. Workshop on Numerical Solution of Markov Chains*, pages 1–20. Zaragosa University Press, 1999.

[15] W. Fischer and K. Meier-Hellstern. The Markov-modulated Poisson process (MMPP) cookbook. *Performance Evaluation*, 18:149–171, 1992.

[16] W. Whitt. Performance of the queueing network analyzer. *The Bell System Technical Journal*, 62(9):2817–2843, 1983.

[17] W. Whitt. The Queueing Network Analyzer. *The Bell System Technical Journal*, 62(9):2779–2815, 1983.

A The power spectrum of a MAP

The power spectrum of the discrete process $T_1, T_2, \ldots$ is the Fourier transform of the non-centralized autocovariance function $V(k) = E[X_1 X_{k+1}]$. In the spcial case of a MAP $(\mathbf{Q_0}, \mathbf{Q_1})$, we have:

$$\begin{aligned} V(k) &= E\{X_1 X_{k+1}\} \\ &= \underline{p}(-\mathbf{Q_0})^{-2}\mathbf{Q_1}\left[(-\mathbf{Q_0})^{-1}\mathbf{Q_1}\right]^{k-1}(-\mathbf{Q_0})^{-2}\mathbf{Q_1}\underline{e} \end{aligned}$$

For $V(0)$ we find:

$$V(0) = E[X_1^2] = 2\underline{p}(-\mathbf{Q_0})^{-3}\mathbf{Q_1}\underline{e}.$$

The Fourier transform $\Phi(\omega)$ is given by:

$$\Phi(\omega) = \sum_{k=-\infty}^{\infty} V(k)e^{-i\omega k}.$$

Since the process is weakly stationary, we have $V(-k) = V(k)$, which leads to

$$\begin{aligned} \Phi(\omega) &= V(0) + \sum_{k=1}^{\infty} V(k)(e^{i\omega k} + e^{-i\omega k}) \\ &= V(0) + 2\underline{C_1} \cdot \sum_{k=1}^{\infty}\left[(-\mathbf{Q_0})^{-1}\mathbf{Q_1}\right]^{k-1}\cos(\omega k) \cdot \underline{C_2}, \end{aligned} \tag{2}$$

where $\underline{C_1} = \underline{p}(-\mathbf{Q_0})^{-2}\mathbf{Q_1}$ and $\underline{C_2} = (-\mathbf{Q_0})^{-2}\mathbf{Q_1}\underline{e}$. Using the eigenvectors and eigenvalues of the matrix $(-\mathbf{Q_0}^{-1}\mathbf{Q_1}$, a closed-form solution for $\Phi(\omega)$ can be obtained.