

TRANSIENT ANALYSIS OF THE BMAP/PH/1 QUEUE

L. M. LE NY and B. SERICOLA

*IRISA-INRIA**Campus universitaire de Beaulieu**35042 Rennes cedex, France*

Email : {Louis-Marie.LeNy, Bruno.Sericola}@irisa.fr,

Abstract: We derive in this paper the transient queue length and the busy period distributions of a single server queue with phase-type (PH) service times and a batch Markovian arrival process (BMAP). These distributions are obtained by the uniformization technique which gives simple recurrence relations. The main advantage of this technique is that it leads to stable algorithms for which the precision of the result can be given in advance.

Keywords: BMAP, busy period, transient queue length, uniformization.

1 INTRODUCTION

We consider a single server queue with infinite waiting room for which the arrival process is a Batch Markovian Arrival Process (BMAP) and the service times are given by a Phase-type (PH) distribution. The BMAP has received considerable interest during the last few years. It was first introduced by [Neuts, 1979] as the versatile Markovian point process. It generalizes the Markovian Arrival Process (MAP) introduced by [Lucantoni et al, 1990]. A tutorial on the BMAP/G/1 queue is presented in [Lucantoni, 1993] and the transient behavior of this queue is analyzed in [Lucantoni et al, 1994] and extended in [Lucantoni, 1998] using numerical inversions of Laplace transforms and generating functions.

More recently, in [Hofmann, 2001] a generalization of the BMAP/G/1 queue where the arrival process is allowed to depend on the state level of the queue was analyzed. The distribution of the stationary queue length at service completion times and at arbitrary time is given using matrix inversions. In [Dudin, 2001] and [Machihara, 1999], the BMAP/SM/1 queue with negative arrivals and service times depending on the arrival process is proposed for modeling some realistic situations like cancellations of calls.

We consider here the BMAP/PH/1 queue and we focus on the transient queue length and busy period distributions. The main advantage with respect to the BMAP/G/1 queue is that the behavior of the BMAP/PH/1 queue is governed by a homogeneous

Markov process and so it can be analyzed avoiding the use of Laplace transforms and generating functions inversion algorithms which can lead to severe numerical errors and overflow problems. Moreover the set of PH probability distributions is dense in the set of probability distributions on $(0, \infty)$, see [Asmussen, 1987] for instance. The transient behavior of queues, see [Duffield et al, 1998], is of great importance for the design and control of new communications networks, above all when the interarrivals and service times are not exponential. In [Takine, 2000] an alternative recursion formula to compute the stationary BMAP/G/1 queue length distribution at a random point in time is shown to be identical to the stationary distribution of the corresponding BMAP/G/1 queue with multiple vacations and exhaustive services.

For the analysis of the BMAP/PH/1 queue, we consider the uniformization technique that can be found in [Ross, 1983]. Its main advantage is that it leads to the analysis of a discrete time Markov chain for which all the required quantities to evaluate are given by recurrence relations involving only additions and multiplications of non negative numbers bounded by one and thus leading to stable algorithms. Moreover, the precision of the results can be given in advance. The remainder of the paper is organized as follows. In the next section we review some definitions and properties of the BMAP that can also be found with more detail in [Lucantoni, 1993] and we describe the Markov process governing the behavior of the BMAP/PH/1 queue. In Section 3 and 4 we derive

the recurrence relations used for the computation of the transient queue length distribution and the busy period distribution respectively. We illustrate our method through numerical examples in Section 5.

2 THE BMAP/PH/1 QUEUE

2.1 The Batch Markovian Arrival Process

The BMAP is a two dimensional Markov process $\{A(t), J(t)\}$ on the state space $\{(i, j) | i \geq 0, 1 \leq j \leq m\}$ with infinitesimal generator given by

$$\begin{bmatrix} D_0 & D_1 & D_2 & D_3 & \cdot & \cdot & \cdot \\ & D_0 & D_1 & D_2 & \cdot & \cdot & \cdot \\ & & D_0 & D_1 & \cdot & \cdot & \cdot \\ & & & D_0 & \cdot & \cdot & \cdot \\ & & & & \cdot & \cdot & \cdot \\ & & & & & \cdot & \cdot \\ & & & & & & \cdot \end{bmatrix}$$

where $D_k, k \geq 0$, are $m \times m$ matrices such that matrix D_0 has negative diagonal elements and non negative off diagonal elements; matrices $D_k, k \geq 1$, have non negative elements and matrix D , defined by

$$D = \sum_{k=0}^{\infty} D_k,$$

is an irreducible infinitesimal generator. We also assume that $D \neq D_0$, which ensures that arrivals will occur.

The variable $A(t)$ counts the number of arrivals during $[0, t)$ and the variable $J(t)$ represents the phase of the arrival process. Let π be the stationary probability of the Markov process with infinitesimal generator D , i.e. π satisfies

$$\pi D = 0 \text{ and } \pi \mathbf{1} = 1,$$

where $\mathbf{1}$ is a column vector of 1's; its dimension being specified by the context. Then the component π_j is the stationary probability that the arrival process is in state j . The stationary arrival rate of the process is then

$$\lambda = \pi \sum_{k=1}^{\infty} k D_k \mathbf{1}. \tag{1}$$

More detail and results concerning this process can be found in [Lucantoni, 1993] for instance.

2.2 The Queueing Model

We consider a single server queue with a BMAP specified by the sequence $\{D_k, k \geq 0\}$. Let the service times be i.i.d. and independent of the arrival process. The service times have a phase-type (PH) distribution [Neuts, 1981] with $l + 1$ states given by its

initial probability distribution (β, β_{l+1}) and infinitesimal generator

$$\begin{bmatrix} T & T_0 \\ 0 & 0 \end{bmatrix}.$$

For the sake of simplicity, we assume that $\beta_{l+1} = 0$. Note that β is a non negative l dimensional row vector satisfying $\beta \mathbf{1} = 1$. Matrix T is a $l \times l$ non singular matrix and T_0 is a l dimensional column vector satisfying $T_0 = -T \mathbf{1}$. The mean service time μ is then given by

$$\mu = \frac{-1}{\beta T^{-1} \mathbf{1}}. \tag{2}$$

The process describing the behaviour of the BMAP/PH/1 queue is then a Markov process $X(t)$, over the state space

$$S = \bigcup_{n=0}^{\infty} C_n,$$

where

$$C_0 = \{(0, j) | 1 \leq j \leq m\}$$

and, for $n \geq 1$,

$$C_n = \{(n, j, k) | 1 \leq j \leq m, 1 \leq k \leq l\}.$$

The couple $(0, j)$ represents the state for which there are no customer in the queue and the arrival process is in phase j . For $n \geq 1$, the triple (n, j, k) represents the state for which there are n customers in the queue, and the arrival process is in phase j and the customer being served is in phase k . With this description of the state space, the infinitesimal generator of the process is given by

$$A = \begin{bmatrix} A_{0,0} & A_{0,1} & A_{0,2} & A_{0,3} & \cdot & \cdot & \cdot \\ A_{1,0} & A_1 & A_2 & A_3 & \cdot & \cdot & \cdot \\ & A_0 & A_1 & A_2 & \cdot & \cdot & \cdot \\ & & A_0 & A_1 & \cdot & \cdot & \cdot \\ & & & A_0 & \cdot & \cdot & \cdot \\ & & & & \cdot & \cdot & \cdot \end{bmatrix}$$

where

$$A_{0,0} = D_0, A_{0,j} = D_j \otimes \beta \text{ for } j \geq 1,$$

$$A_{1,0} = I_m \otimes T_0, A_1 = (D_0 \otimes I_l) + (I_m \otimes T),$$

$$A_j = D_{j-1} \otimes I_l \text{ for } j \geq 2 \text{ and } A_0 = I_m \otimes (T_0 \beta),$$

where the matrix I_r denotes the $r \times r$ identity matrix and \otimes denotes the classical Kronecker product. Let us define $\nu = \sup\{-A(u, u), u \in S\}$. It is easy to verify that

$$\begin{aligned} \nu &= \max\{-D_0(j, j), j = 1, \dots, m\} \\ &\quad + \max\{-T(k, k), k = 1, \dots, l\}. \end{aligned}$$

Using now the uniformization technique [Ross, 1983], we denote by $Z(n)$ the uniformized Markov chain associated to $X(t)$, with $Z(0) = X(0)$. Its transition probability matrix P is given by $P = I + A/\nu$, where I denotes the infinite identity matrix. Matrix P is a stochastic matrix and has the same structure as matrix A with

$$\begin{aligned} P_{0,0} &= I_m + A_{0,0}/\nu, \quad P_{0,j} = A_{0,j}/\nu \text{ for } j \geq 1, \\ P_{1,0} &= A_{1,0}/\nu, \quad P_1 = I_m + A_1/\nu, \\ P_j &= A_j/\nu \text{ for } j \geq 2 \text{ and } P_0 = A_0/\nu. \end{aligned}$$

For $i \geq 1$, we define the subsets

$$B_i = \bigcup_{n=0}^i C_n.$$

For every $i \geq 0$, B_i represents the states of S corresponding to at most i customers in the queue.

3 THE QUEUE LENGTH

We denote by α the row vector containing the initial probability distribution of the Markov process $X(t)$ and by $\mathbf{1}_{B_i}$, $i \geq 0$, the infinite column vector with the first $|B_i|$ entries equal to 1 and the others equal to 0.

We define $Y(t)$ as the number of customers in the queue at time t . The distribution of $Y(t)$ is then given by

$$\begin{aligned} \Pr(Y(t) \leq i) &= \Pr(X(t) \in B_i) \\ &= \alpha e^{At} \mathbf{1}_{B_i} \\ &= \sum_{n=0}^{\infty} e^{-\nu t} \frac{(\nu t)^n}{n!} \Pr(Z(n) \in B_i), \end{aligned}$$

where $\Pr(Z(n) \in B_i)$ is given by

$$\Pr(Z(n) \in B_i) = \alpha P^n \mathbf{1}_{B_i}.$$

We now derive recurrence relations for the computation of $\Pr(Z(n) \in B_i)$.

For every $n \geq 0$ and $i \geq 0$ we define the infinite column vector $U_{B_i}^{(n)}$ as

$$U_{B_i}^{(n)} = P^n \mathbf{1}_{B_i}.$$

The infinite column vector $U_{B_i}^{(n)}$ can be decomposed into subvectors $U_{h,i}^{(n)}$, $h \geq 0$. The vector $U_{0,i}^{(n)}$ is of dimension m and for $h \geq 1$, $U_{h,i}^{(n)}$ is of dimension ml . With this notation, we have for $h \geq 1$, $1 \leq j \leq m$ and $1 \leq k \leq l$,

$$\begin{aligned} U_{0,i}^{(n)} &= \left(\Pr(Z(n) \in B_i \mid Z(0) = (0, j)) \right)_j \\ U_{h,i}^{(n)} &= \left(\Pr(Z(n) \in B_i \mid Z(0) = (h, j, k)) \right)_{j,k}. \end{aligned}$$

It follows that for $h \geq i + n + 1$, we have $U_{h,i}^{(n)} = 0$. Indeed, if $h \geq i + n + 1$, it is not feasible to have at most i customers in the queue in n transitions, starting with h customers in the queue at time 0.

For $n = 0$, we have $U_{B_i}^{(0)} = \mathbf{1}_{B_i}$ and for $n \geq 1$, writing $U_{B_i}^{(n)} = P U_{B_i}^{(n-1)}$, we get the following recurrence relations, for $2 \leq r \leq i + n$.

$$U_{0,i}^{(n)} = \sum_{h=0}^{i+n-1} P_{0,h} U_{h,i}^{(n-1)} \quad (3)$$

$$U_{1,i}^{(n)} = P_{1,0} U_{0,i}^{(n-1)} + \sum_{h=1}^{i+n-1} P_{0,h} U_{h,i}^{(n-1)} \quad (4)$$

$$U_{r,i}^{(n)} = \sum_{h=r-1}^{i+n-1} P_{h-r+1} U_{h,i}^{(n-1)}. \quad (5)$$

We now decompose the initial probability distribution α as $\alpha = (\alpha_h)_{h \geq 0}$ where the subvector α_h corresponds to the initial probability distribution when there are exactly h customers in queue, that is, for $h \geq 1$, $1 \leq j \leq m$ and $1 \leq k \leq l$,

$$\begin{aligned} \alpha_0 &= \left(\Pr(Z(0) = (0, j)) \right)_j \\ \alpha_h &= \left(\Pr(Z(0) = (h, j, k)) \right)_{j,k}. \end{aligned}$$

Using this notation we get

$$\Pr(Z(n) \in B_i) = \sum_{h=0}^{\infty} \alpha_h U_{h,i}^{(n)} = \sum_{h=0}^{n+i} \alpha_h U_{h,i}^{(n)}. \quad (6)$$

Let ε be the desired error tolerance for the computation of $\Pr(Y(t) \leq i)$, we define the truncation step N as

$$N = \min \left\{ n \in \mathbb{N} \mid \sum_{j=0}^n e^{-\nu t} \frac{(\nu t)^j}{j!} \geq 1 - \varepsilon \right\}. \quad (7)$$

The distribution of $Y(t)$ can then be written as

$$\Pr(Y(t) \leq i) = \sum_{n=0}^N e^{-\nu t} \frac{(\nu t)^n}{n!} \Pr(Z(n) \in B_i) + e(N),$$

where $e(N)$ verifies

$$\begin{aligned} e(N) &= \sum_{n=N+1}^{\infty} e^{-\nu t} \frac{(\nu t)^n}{n!} \Pr(Z(n) \in B_i) \\ &\leq \sum_{n=N+1}^{\infty} e^{-\nu t} \frac{(\nu t)^n}{n!} \\ &= 1 - \sum_{n=0}^N e^{-\nu t} \frac{(\nu t)^n}{n!} \\ &\leq \varepsilon. \end{aligned}$$

Remark 1: The computation of integer N can be made without any numerical problems even for large values of νt by using the method described in [Bowerman et al, 1990].

Remark 2: The truncation level N is in fact a function of t , say $N(t)$ and it can be easily shown that for a fixed value of ε , $N(t)$ is an increasing function of t . It follows that if we want to compute $\Pr(Y(t) \leq i)$ for M distinct values of t denoted by $t_1 < \dots < t_M$ we only need to compute $\Pr(Z(n) \in B_i)$ for $n = 1, \dots, N(t_M)$ since these values are independent of the parameter t .

We introduce the functions $F_i(t)$ defined by

$$F_i(t) = \sum_{n=0}^{N(t_M)} e^{-\nu t} \frac{(\nu t)^n}{n!} \Pr(Z(n) \in B_i). \quad (8)$$

We then have, for every $t \leq t_M$, by definition of $e(N)$

$$\begin{aligned} \Pr(Y(t) \leq i) - F_i(t) &= e(N(t_M)) \\ &\leq \sum_{n=N(t_M)+1}^{\infty} e^{-\nu t} \frac{(\nu t)^n}{n!} \\ &= 1 - \sum_{n=0}^{N(t_M)} e^{-\nu t} \frac{(\nu t)^n}{n!} \\ &\leq 1 - \sum_{n=0}^{N(t)} e^{-\nu t} \frac{(\nu t)^n}{n!} \\ &\leq \varepsilon. \end{aligned}$$

Below we give the algorithm which we developed to compute the transient queue length distribution.

input : $i, t_1 < \dots < t_M, \varepsilon$
output : $F_i(t_1), \dots, F_i(t_M)$
 Compute $N = N(t_M)$ from relation (7)
for $h = 0$ **to** i **do**
 $U_{h,i}^{(0)} = \mathbf{1}$
endfor
for $n = 1$ **to** N **do**
 Compute $U_{0,i}^{(n)}$ from relation (3)
 Compute $U_{1,i}^{(n)}$ from relation (4)
 for $r = 2$ **to** $i + n$ **do**
 Compute $U_{r,i}^{(n)}$ from relation (5)
 endfor
 Compute $\Pr(Z(n) \in B_i)$ from relation (6)
endfor
for $j = 1$ **to** M **do**
 Compute $F_i(t_j)$ from relation (8)
endfor

4 THE BUSY PERIOD

Let us define the subset B'_0 as the subset of states corresponding to at least one customer in the queue,

that is

$$B'_0 = S - B_0 = \bigcup_{n=1}^{\infty} C_n.$$

The busy period, denoted by BP , is defined by the first passage time from subset B'_0 to subset B_0 .

We denote by β an initial probability distribution concentrated on subset B'_0 , that is $\beta \mathbf{1} = 1$. This distribution can then be written as $\beta = (\beta_h)_{h \geq 1}$ where subvector β_h , which is of dimension ml , corresponds to the initial probability distribution when there are exactly h customers in the queue. The distribution of the busy period BP is then given by

$$\Pr(BP > t) = \beta e^{Gt} \mathbf{1},$$

where matrix G is the submatrix obtained from the infinitesimal generator A by deleting the first m rows and columns which correspond to subset B_0 .

In the same way, we define Q as the submatrix obtained from matrix P by deleting the first m rows and columns. Matrix Q is then a substochastic matrix given by the relation $Q = I + G/\nu$, and so

$$Q = \begin{bmatrix} P_1 & P_2 & P_3 & \cdot & \cdot & \cdot \\ P_0 & P_1 & P_2 & \cdot & \cdot & \cdot \\ & P_0 & P_1 & \cdot & \cdot & \cdot \\ & & P_0 & \cdot & \cdot & \cdot \\ & & & \cdot & \cdot & \cdot \\ & & & & \cdot & \cdot \end{bmatrix}.$$

We thus get

$$\Pr(BP > t) = \sum_{n=0}^{\infty} e^{-\nu t} \frac{(\nu t)^n}{n!} \beta Q^n \mathbf{1}.$$

Defining the infinite column vector $U^{(n)}$ as $U^{(n)} = Q^n \mathbf{1}$, we get the recurrence relation $U^{(n)} = QU^{(n-1)}$ with $U^{(0)} = \mathbf{1}$.

Let $V^{(n)}$ be defined by $V^{(n)} = \mathbf{1} - U^{(n)}$. We then get

$$V^{(n)} = V^{(1)} + QV^{(n-1)}, \quad (9)$$

with $V^{(0)} = 0$ and $V^{(1)} = \mathbf{1} - Q\mathbf{1}$.

We decompose the infinite column vector $V^{(n)}$ into subvectors $V_i^{(n)}$, $i \geq 1$ of dimension ml . For $i \geq 1$, vector $V_i^{(n)}$ can be interpreted using the uniformized Markov chain $Z(n)$ as

$$V_i^{(n)} = \left(\Pr(NV \leq n | Z(0) = (i, j, k)) \right)_{1 \leq j \leq m, 1 \leq k \leq l},$$

where NV denotes the number of states visited during the busy period. It follows immediately that for $i \geq n + 1$, we have $V_i^{(n)} = 0$. Indeed, the number of states visited during a busy period cannot be less than or equal to n if the initial number of customers

in the queue is greater than or equal to $n + 1$. The vectors $V^{(n)}$ thus have a finite number of non zero entries; that is why it is preferable to use these vectors instead of the $U^{(n)}$.

Using relation (9), we obtain the following recurrence relation, using the fact that $P_{1,0}\mathbf{1} = P_0\mathbf{1}$,

$$V_1^{(n)} = P_0\mathbf{1} + \sum_{h=1}^{n-1} P_h V_h^{(n-1)} \quad (10)$$

$$V_i^{(n)} = \sum_{h=i-1}^{n-1} P_{h-i+1} V_h^{(n-1)} \text{ for } 2 \leq i \leq n \quad (11)$$

The matrix P being a substochastic matrix the sequence of vectors $U^{(n)}$ decreases to 0 when n tends to infinity, for $\rho < 1$ where $\rho = \lambda/\mu$ is the traffic intensity, given by relations (1) and (2). It follows that the computation can be stopped at step N_1 when the number $\beta U^{(N_1)}$ is sufficiently small or equivalently when $\beta V^{(N_1)}$ is sufficiently close to 1. More precisely, for a given error tolerance ε , let us define N_1 as

$$N_1 = \min \left\{ n \in \mathbb{N} \mid \beta V^{(n)} \geq 1 - \varepsilon \right\}. \quad (12)$$

Using the truncation step N defined by (7), and defining $N' = \min(N, N_1)$, we get

$$\begin{aligned} \Pr(BP > t) &= \sum_{n=0}^{\infty} e^{-\nu t} \frac{(\nu t)^n}{n!} \beta U^{(n)} \\ &= \sum_{n=0}^{N'} e^{-\nu t} \frac{(\nu t)^n}{n!} \beta U^{(n)} + e_1(N') \\ &= \sum_{n=0}^{N'} e^{-\nu t} \frac{(\nu t)^n}{n!} \\ &\quad - \sum_{n=0}^{N'} e^{-\nu t} \frac{(\nu t)^n}{n!} \beta V^{(n)} + e_1(N'), \end{aligned}$$

where

$$e_1(N') = \sum_{n=N'+1}^{\infty} e^{-\nu t} \frac{(\nu t)^n}{n!} \beta U^{(n)}.$$

If $N' = N$ then we have, by definition of N ,

$$\begin{aligned} e_1(N') &= \sum_{n=N+1}^{\infty} e^{-\nu t} \frac{(\nu t)^n}{n!} \beta U^{(n)} \\ &\leq \sum_{n=N+1}^{\infty} e^{-\nu t} \frac{(\nu t)^n}{n!} \\ &= 1 - \sum_{n=0}^N e^{-\nu t} \frac{(\nu t)^n}{n!} \\ &\leq \varepsilon. \end{aligned}$$

If $N' = N_1$ then we have, by definition of N_1 ,

$$\begin{aligned} e_1(N') &= \sum_{n=N_1+1}^{\infty} e^{-\nu t} \frac{(\nu t)^n}{n!} \beta U^{(n)} \\ &\leq \varepsilon \sum_{n=N_1+1}^{\infty} e^{-\nu t} \frac{(\nu t)^n}{n!} \\ &\leq \varepsilon. \end{aligned}$$

We thus have $e_1(N') \leq \varepsilon$. We will see in the next section that the truncation step N' can be significantly less than N . Note that both Remarks 1 and 2 of the previous section also apply here for the computation of $\Pr(BP > t)$. So suppose that we need to compute the busy period distribution for M distinct values of t , denoted by $t_1 < \dots < t_M$, we introduce the functions $G(t)$ defined by

$$G(t) = \sum_{n=0}^{N'(t_M)} e^{-\nu t} \frac{(\nu t)^n}{n!} - \sum_{n=0}^{N'(t_M)} e^{-\nu t} \frac{(\nu t)^n}{n!} \beta V^{(n)},$$

where $N'(t) = \min(N(t), N_1)$. $N(t)$ being increasing with t , so is $N'(t)$. We then have by definition of $e_1(N)$, for $t \leq t_M$,

$$\Pr(BP > t) - G(t) = e_1(N'(t_M)) \leq e_1(N'(t)) \leq \varepsilon.$$

Below we give the algorithm which we developed to compute the busy period distribution.

input : $t_1 < \dots < t_M, \varepsilon$
output : $G(t_1), \dots, G(t_M)$
 Compute $N = N(t_M)$ from relation (7)
 $N' = N, V_1^{(0)} = 0, n = 0$
while [$n < N'$] **do**
 $n = n + 1$
 Compute $V_1^{(n)}$ from relation (10)
 for $i = 2$ **to** n **do**
 Compute $V_i^{(n)}$ from relation (11)
 endfor
 if $\left(\sum_{i=1}^n \beta_i V_i^{(n)} \geq 1 - \varepsilon \right)$ **then** $N' = n$ **endif**
endwhile
for $j = 1$ **to** M **do** Compute $G(t_j)$ **endfor**

5 NUMERICAL RESULTS

We consider a BMAP which is the superposition of h identical Markov Modulated Poisson Processes (MMPPs). Each MMPP alternates between two states: state 0 and state 1. The transition rate from state 0 to state 1 is denoted by μ_0 and the transition

rate from state 1 to state 0 is denoted by μ_1 . The arrival rate from state 0 (resp. 1) is denoted by λ_0 (resp. λ_1). The auxiliary phase in the overall BMAP can be characterized by the number of MMPPs that are in state 1. This number is initially supposed to be equal to 0, that is, all the MMPPs are initially in state 0. The service times distribution is assumed to be Erlang of order l (denoted by $E_l, l \geq 1$) with unit mean time so that the time units are in mean service times.

It follows that the overall BMAP is given by the matrices D_0 and D_1 ($D_k = 0$ for $k \geq 2$) of dimension $m \times m$ where $m = h + 1$. The non-zero entries of matrices D_0 and D_1 are, for $0 \leq i \leq h$,

$$\begin{aligned} D_1(i, i) &= (h - i)\lambda_0 + i\lambda_1 \\ D_0(i, i) &= -[(h - i)(\lambda_0 + \mu_0) + i(\lambda_1 + \mu_1)] \end{aligned}$$

and,

$$\begin{aligned} D_0(i, i + 1) &= (h - i)\mu_0 \text{ for } 0 \leq i \leq h - 1 \\ D_0(i, i - 1) &= i\mu_1 \text{ for } 1 \leq i \leq h \end{aligned}$$

The arrival rate of the overall BMAP is then

$$\lambda = h \left[\lambda_0 \frac{\mu_1}{\mu_0 + \mu_1} + \lambda_1 \frac{\mu_0}{\mu_0 + \mu_1} \right].$$

The mean service times being equal to 1, the traffic intensity ρ is given by $\rho = \lambda$. In all figures we assume that $\lambda_0 = 0.01, \lambda_1 = 0.04, \mu_0 = 0.04, \mu_1 = 0.01$ and the error tolerance is $\varepsilon = 0.0001$.

Figure 1 shows the complementary cumulative distribution function of the transient queue length for the $\sum_{r=1}^{20} MMPP_r/E_4/1$ model, which gives a traffic intensity $\rho = 0.68$. The initial queue length is $i' = 20$. It can be noted that for values of t greater than 100 the curves coincide with the curve obtained for $t = 100$ and so the steady state seems to be reached.

Figure 2 shows the complementary emptiness function for the $\sum_{r=1}^{20} MMPP_r/E_4/1$ model which gives a traffic intensity $\rho = 0.68$, for different initial queue length. Note that the steady state value of this function ($\rho = 0.68$) is reached for $t = 200$.

Figure 3 shows the complementary emptiness function for the $\sum_{r=1}^{20} MMPP_r/E_l/1$ model which gives a traffic intensity $\rho = 0.68$, for different values of the number l of phases in the Erlang service distribution. The initial queue length is $i' = 20$. It is interesting to note that for t greater than 80 there are no more differences between the curves and so it seems that the number of service phases does not have a great

influence on the emptiness function. Here the steady state value of this function ($\rho = 0.68$) is reached for $t = 100$.

Figures 4 and 5 show the complementary emptiness function for the $\sum_{r=1}^h MMPP_r/E_4/1$ model which gives a traffic intensity $\rho = h \times 0.034$, for different values of the number h of MMPPs. The initial queue length is $i' = 20$ in Figure 4 and $i' = 0$ in Figure 5. Note that for $h = 29$ we get $\rho = 0.986$, for $h = 30$ we get $\rho = 1.02$, and for $h = 40$ we get $\rho = 1.36$. Note also the convergence to the steady state value $\rho < 1$ for $h \leq 29$, and the convergence to 1 for $h \geq 30$, since in that case $\rho > 1$.

Figure 6 shows the complementary cumulative distribution function of the busy period for the $\sum_{r=1}^{20} MMPP_r/E_4/1$ model which gives a traffic intensity $\rho = 0.68$, for different values of the initial queue length. Note that the rate of convergence to 0 depends on the initial queue length. Concerning the truncation steps, for $t = 200$, we have $N = 1125$ and the following values of N' depending on the initial queue length. These results show that the use of the truncation step N' can significantly reduce the computation time of the algorithm.

i'	1	5	10	20	≥ 30
N'	86	269	506	878	1125

Figure 7 shows the complementary cumulative distribution function of the busy period for the $\sum_{r=1}^{20} MMPP_r/E_l/1$ model which gives a traffic intensity $\rho = 0.68$, for different values of the number l of phases in the Erlang service distribution. The initial queue length is $i' = 20$. As for Figure 3, it is interesting to note that there are not many differences between the curves and so it seems that the number of service phases does not have a great influence on the busy period distribution. In this case, for $t = 200$ and $l = 6$, we have $N = 1549$ and $N' = 1172$.

Figure 8 shows the complementary cumulative distribution function of the busy period for the $\sum_{r=1}^h MMPP_r/E_4/1$ model which gives a traffic intensity $\rho = h \times 0.034$, for different values of the number h of MMPPs. The initial queue length is $i' = 20$. As for Figure 5, we have $\rho < 1$ for $h \leq 29$ and $\rho > 1$ for $h \geq 30$.

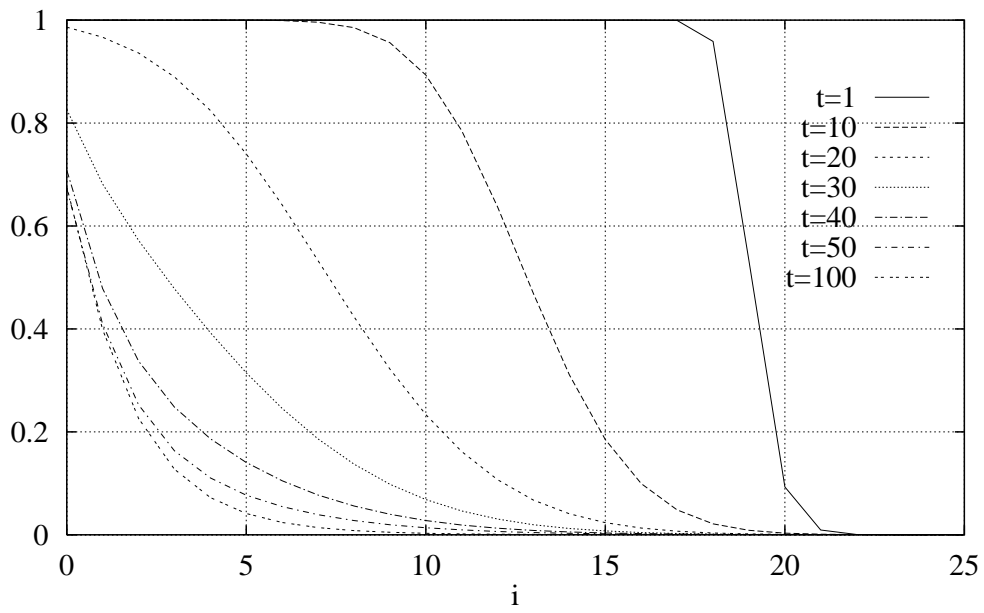


Figure 1: $\Pr(Y(t) > i)$, as a function of i and t in the $\sum_{r=1}^{20} MMPP_r/E_4/1$ queue with initial queue length $i' = 20$.

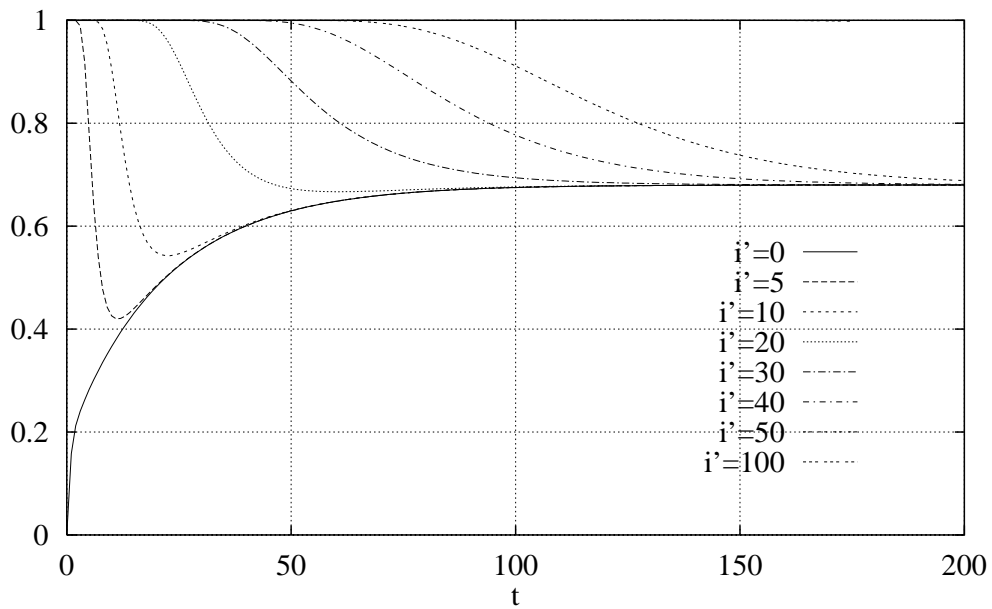


Figure 2: $\Pr(Y(t) > 0)$, as a function of t and the initial queue length i' in the $\sum_{r=1}^{20} MMPP_r/E_4/1$ queue.

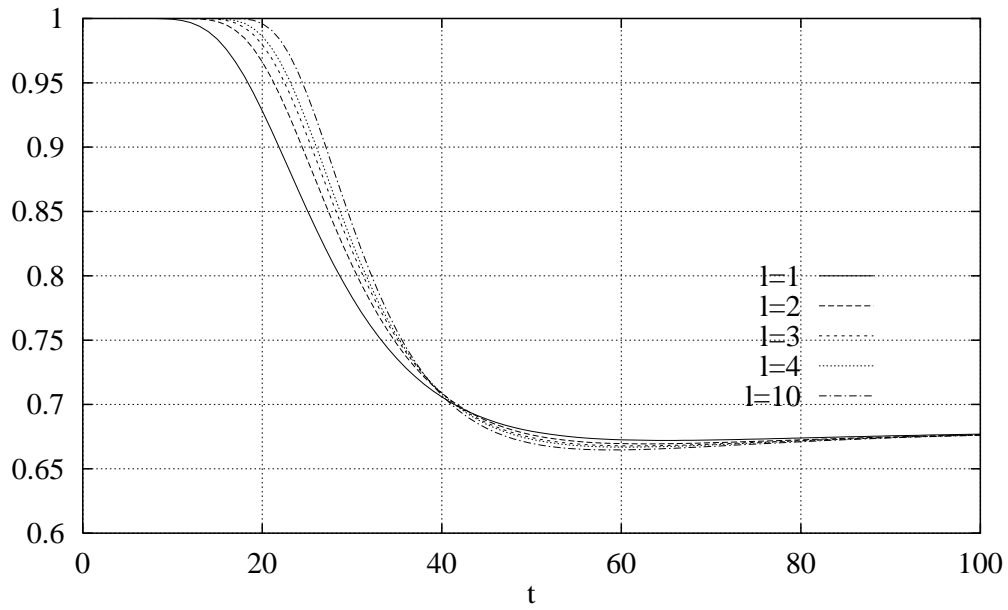


Figure 3: $\Pr(Y(t) > 0)$, as a function of t for different values of the number l of phases in the $\sum_{r=1}^{20} MMPP_r/E_l/1$ queue with initial queue length $i' = 20$.

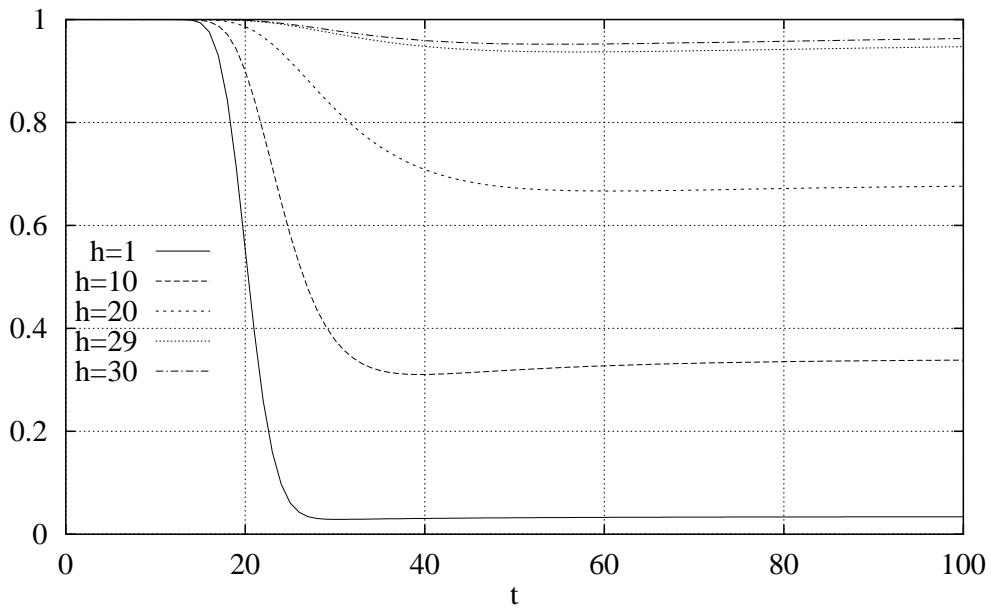


Figure 4: $\Pr(Y(t) > 0)$, as a function of t for different values of the number h of MMPPs in the $\sum_{r=1}^h MMPP_r/E_4/1$ queue with initial queue length $i' = 20$.

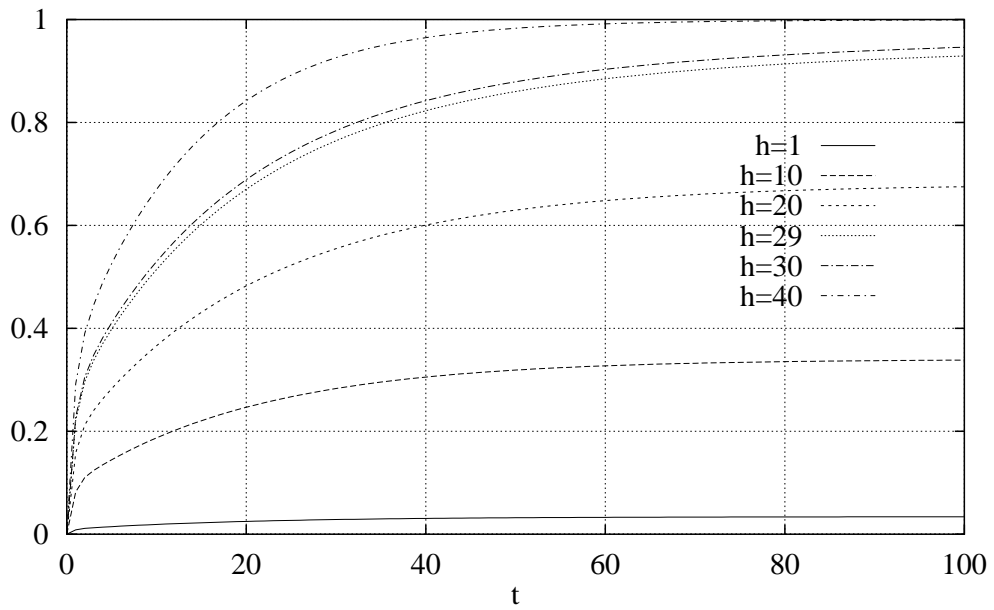


Figure 5: $\Pr(Y(t) > 0)$, as a function of t for different values of the number h of MMPPs in the $\sum_{r=1}^h MMPP_r/E_4/1$ queue with initial queue length $i' = 0$.

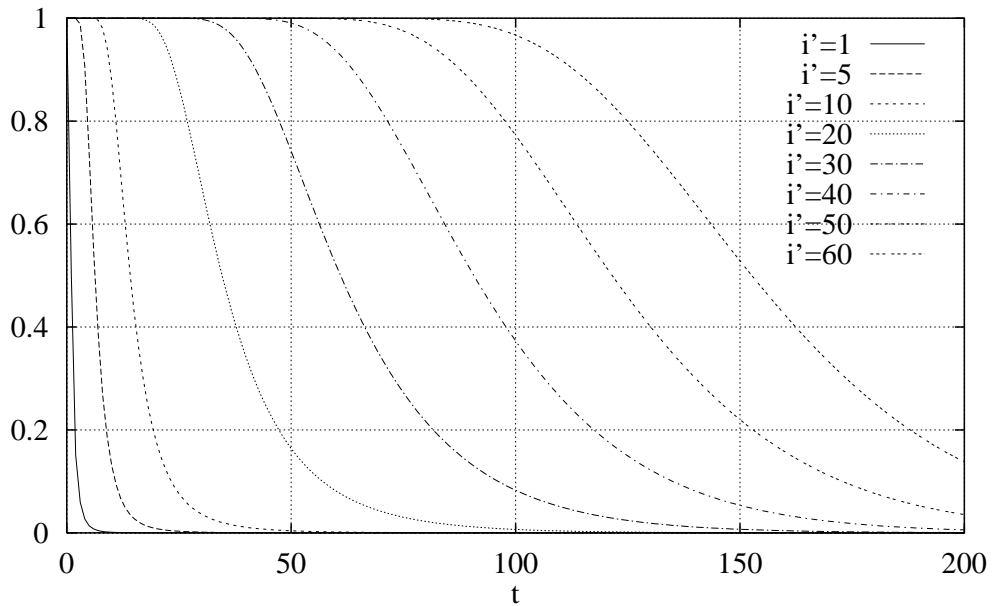


Figure 6: $\Pr(BP > t)$, as a function of t and the initial queue length i' in the $\sum_{r=1}^{20} MMPP_r/E_4/1$ queue.

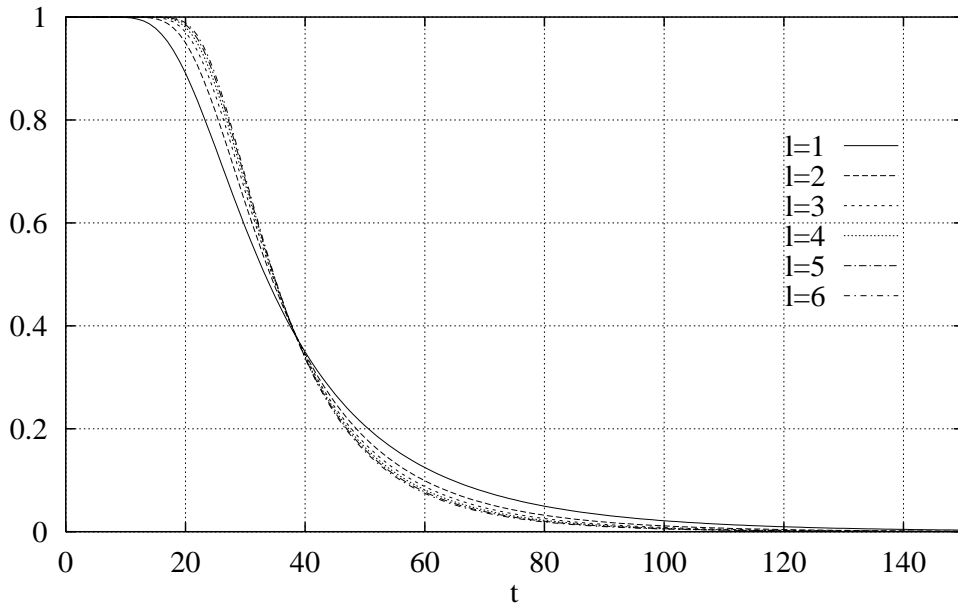


Figure 7: $\Pr(BP > t)$, as a function of t for different values of the number l of phases in the $\sum_{r=1}^{20} MMPP_r/E_l/1$ queue with initial queue length $i' = 20$.

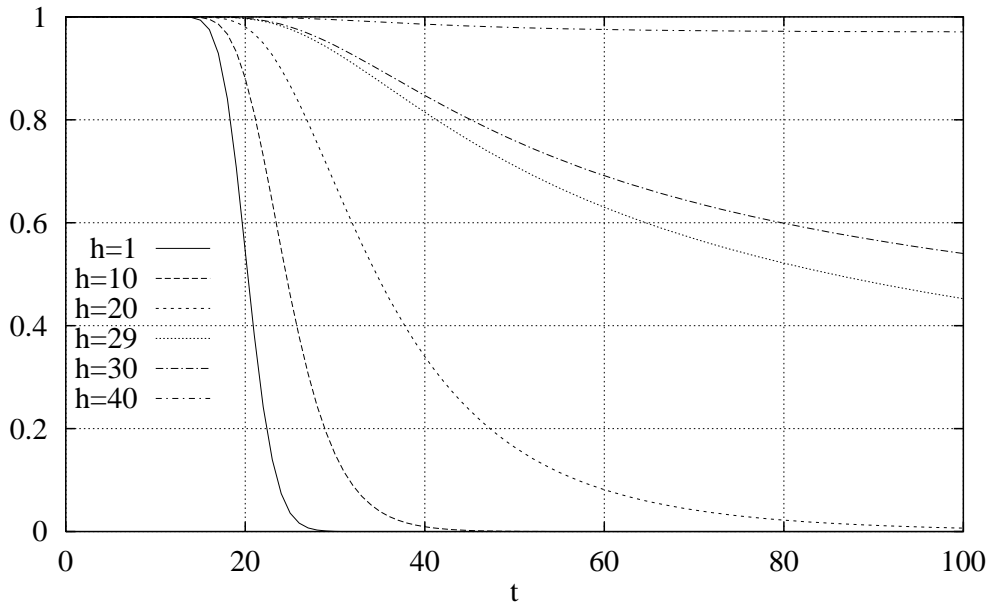


Figure 8: $\Pr(BP > t)$, as a function of t for different values of the number h of MMPPs in the $\sum_{r=1}^h MMPP_r/E_4/1$ queue with initial queue length $i' = 20$.

REFERENCES

Asmussen S. 1987, *Applied Probability and Queues*. John Wiley and Sons.

Bowerman P. N., Nolty R. G. and Scheuer E. M. 1990, "Calculation of the poisson cumulative distribution function". *IEEE Transactions on Reliability*, Vol. 39. Pp:158–161.

Dudin A. N., Karolik A. V. 2001, "BMAP/SM/1 queue with Markovian input of disasters and non-instantaneous recovery". *Performance Evaluation*, Vol. 45. Pp:19–32.

Duffield N. G., Whitt W. 1998, "A source traffic model and its transient analysis for network control". *Stochastic Models*, Vol. 14. Pp:51–78.

Hofmann J. 2001, "The BMAP/G/1 queue with Level-Dependent Arrivals - An Overview". *Telecommunication Systems*, Vol. 16. Pp:347–360.

Lucantoni D. M., Meier-Hellstern K. S. and Neuts M. F. 1990, "A single server queue with server vacations and a class of non-renewal arrival processes". *Adv. Appl. Prob.*, Vol. 22. Pp:676–705.

Lucantoni D. M. 1993, "The BMAP/G/1 queue: A tutorial". In L. Donatiello and R. Nelson, editors, *Performance Evaluation of Computer and Communications Systems*. Lectures Notes in Computer Science 729, Springer Verlag, Pp:330–358.

Lucantoni D. M., Choudhury G. L. and Whitt W. 1994, "The transient BMAP/G/1 queue". *Stochastic Models*, Vol. 10. Pp:145–182.

Lucantoni D. 1998, "Further Transient Analysis of the BMAP/G/1 queue". *Stochastic Models*, Vol. 14. Pp:461–478.

Machihara F. 1999, "A BMAP/SM/1 queue with service times depending on the arrival process".

Queueing Systems - Theory and Application, Vol. 33. Pp:277–292.

Neuts M. F. 1979, "A versatile Markovian point process". *J. Appl. Prob.*, Vol. 16. Pp:764–779.

Neuts M. F. 1981, *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*. Baltimore: The Johns Hopkins University Press.

Ross S. M. 1983, *Stochastic Processes*. John Wiley and Sons.

Takine T. 2000, "A new recursion for the queue length distribution in the stationary BMAP/G/1 queue". *Stochastic Models*, Vol. 16. Pp:335–341.

BIOGRAPHY



LOUIS-MARIE LE NY is an Associate Professor in Applied Mathematics at Rennes 1 university (France). He is also part of the ARMOR (Networks architectures and Modeling) group at IRISA laboratory. His research interests include queueing networks analysis and performance evaluation of Internet Protocols.



BRUNO SERICOLA received the Ph.D. degree in computer science from the University of Rennes I in 1988. He has been with INRIA (Institut National de Recherche en Informatique et Automatique, a public research French laboratory) since 1989. His main research activity is in computer and communication systems performance evaluation, dependability and performability analysis of fault-tolerant architectures and applied stochastic processes.