Evaluation of the Maximum Level
Reached by a Queue
Over a Finite Period

Gerardo Rubino
Irisa/Inria/EnstB
Irisa
Campus de Beaulieu
35042 Rennes Cedex, France
rubino@irisa.fr

Abstract

This paper deals with the performance analysis of a system modeled by a queue. If we are interested in occupation problems and if we look at the transient phase, then it makes sense to study the maximum backlog observed in the queue over a finite period. This paper proposes an efficient algorithmic scheme to evaluate the distribution of this maximum backlog level, based on the uniformization technique. The approach is illustrated using the classical $M/M/1$ model, but it can be extended to more complex ones.

1 Introduction

Consider some shared resource, for instance a server, or consider a router in a packet-switching communication network. The kind of problem we are interested in is the evaluation of how much storage must be allocated to (or will be used by) some specific flow or aggregation of flows, at the device and during some fixed and known period of time, which will be denoted here by $[0, T]$. If T is small, the classical asymptotic results of queueing theory can be irrelevant, and a transient analysis can then be necessary to perform such an evaluation. This paper proposes an efficient computational scheme to compute the probability that the backlog at the device does not exceed some given level over a finite period. The idea is to use this distribution for dimensionning purposes.

To be specific, let us consider the problem of deciding how much room H must be assigned to a system modeled as a $M/M/1/H$ queue, for instance in order to reduce (probabilistically) the number of rejected units due to saturation during a fixed and finite time period. The fact that we are interested in the transient phase makes this task more difficult than the classical approach in applied queuing theory, where stationarity is assumed and where we usually limit ourselves to the computation of the loss probability. For instance, in the stationary $M/M/1/H$ model, the so-called PASTA property (Poisson Arrivals See Time Averages, see for instance [1]), tells us that the loss probability is equal to the probability that the system is full, that is, in state H; denoting it by π_H, we have

$$\pi_H = \frac{(1 - \varrho)\varrho^H}{1 - \varrho^{H+1}}, \qquad (1)$$

where $\varrho = \lambda/\mu \neq 1$ is the load (λ is the arrival rate and μ is the service rate), thus leading to

$$\pi_H \leq \delta \iff H \geq \left\lceil \ln\left(\frac{\delta}{\delta\varrho + 1 - \varrho}\right)/\ln \varrho \right\rceil.$$

As usual, $\lceil x \rceil$ denotes the smallest integer greater than or equal to x. If $\lambda = \mu$, then $\pi_H = (H + 1)^{-1}$ and $\pi_H \leq \delta$ is equivalent to $H \geq \left\lceil \delta^{-1} - 1 \right\rceil$.

When the system is to be analyzed over a fixed period in time, the previous development can be irrelevant, or can lead to very poor precision, and, to the best of our knowledge, the only available approach is to analyze process (N_t) where N_t is the number of units in the system at time t. For instance, in [6] a technique is proposed which is able to compute the distribution of the random variable "fraction of the interval $[0, T]$ where the system is saturated".

Here, we propose to work with the random variable

$$M_t = \max\{N_s, s \leq t\},$$

1

which allows a different and detailed analysis of the problem. For instance, given $\delta \ll 1$ and $T > 0$, one can compute

$$L = \min\{K : \Pr(M_T > K) < \delta\}$$

with (M_t) defined on the open model $M/M/1$, and then make some choice H satisfying $H \geq L$. Coming back to the first example, observe that we may have, say, $\Pr(M_T > L) = 1 - 10^{-4}$ and $L \ll \mathrm{E}(N_\infty)$. Using L instead of $\mathrm{E}(N_\infty)$, may lead to a more efficient way to allocate room to the server.

The necessity of analyzing the transient dynamics of ressource sharing systems in the communications world is, of course, not new. For an interesting and recent development of some non standard transient aspects related to these communication systems see [2]. The random variable M_t is studied in [3] where analytical expressions are proposed. Here, we deal with algorithms allowing to efficiently evaluate the distribution of this variable and which can be extended to other models. Next section establishes the preliminary transformations to perform the analysis, which is developed in section 3. Section 4 completes the analysis leading to the computational scheme, and Section 5 illustrates it with some numerical examples. Section 6 concludes the paper.

2   Main Transformations

The solution process starts by performing two transformations. First, an auxiliary Markov process (Y_t) is defined and it is shown that solving a specific problem on (Y_t) gives the solution to the original one. Then, the uniformization technique allows to transform the continuous time problem on (Y_t) into a discrete time problem on a third associated process (Z_k). The rest of the paper exploits the specific structure of (Z_k) to derive an efficient computation scheme, using standard markovian analysis methods.

Denote here by N_s the number of customers in the *open* $M/M/1$ model, assumed to be empty at time 0 to simplify the presentation. Define

$$Y_t = (N_t, M_t).$$

It is clear that (Y_t) is a continuous time homogeneous Markov chain while (M_t) is not Markov. This last fact can be easily formally proved after the transformations performed in this section. The fact that (Y_t) is Markov can be formally proved by checking the Markov property directly from the definition, or by using the classical construction from Poisson processes (see, for instance, the compact presentation in [1, 10.1]). The space state of (Y_t) is the set

$\{(n, m) \in \mathbb{N}^2 : n \leq m\}$. Denoting by $Q(x, y)$ the transition rate from state x to state y, the non-null transition rates are

$$
\begin{aligned}
Q((n, m), (n+1, m)) &= \lambda, \quad n < m, \\
Q((n, m), (n-1, m)) &= \mu, \quad n > 0, \\
Q((m, m), (m+1, m+1)) &= \lambda.
\end{aligned}
$$

The idea is to use the uniformization technique to analyze the (Y_t) process. It consists of the following procedure (see, for instance, [4]): call A the infinitesimal generator of (Y_t) (if $x \neq y$, $A(x, y) = Q(x, y)$, and $A(x, x) = -\sum_{y:y \neq x} Q(x, y)$), and choose any real number $\Lambda \geq -A(x, x)$ for all state x. This real Λ is called *uniformization rate*. Then build matrix P by $P = I - A/\Lambda$. Matrix P is stochastic (its rows are probability distributions) and we have

$$e^{At} = e^{-(I-P)\Lambda t} = e^{-\Lambda t} e^{P \Lambda t}$$

(since P and I commute). Let (Z_k) denote a discrete time homogeneous Markov chain on the same state space than (Y_t), with the same initial distribution, and having transition probabilities P. By a Taylor expansion of $e^{P \Lambda t}$,

$$\Pr(Y_t = y) = \sum_{k=0}^{\infty} e^{-\Lambda t} \frac{(\Lambda t)^k}{k!} \Pr(Z_k = y). \qquad (2)$$

What relation (2) says is that we can obtain the distribution of the continuous time process (Y_t) basically by an exponential transform of the distribution of the discrete time process (Z_k).

Let us choose $\Lambda = \lambda + \mu$. The non-null elements of P are

$$
\begin{aligned}
P((0, m), (0, m)) &= q, \\
P((n, m), (n+1, m)) &= p, \quad n < m, \\
P((n, m), (n-1, m)) &= q, \quad n > 0, \\
P((m, m), (m+1, m+1)) &= p,
\end{aligned}
$$

where

$$p = \frac{\lambda}{\lambda + \mu} \quad \text{and} \quad q = \frac{\mu}{\lambda + \mu}.$$

Figure 1 illustrates the structure of chain (Z_k).

The remaining (and main) task is the analysis of the discrete time process (Z_k). So, the price to pay to be able to work with a discrete time Markov chain, that is, to deal with *recurrences* instead of differential equations, is to work now with a bi-dimensional random walk. Let us denote
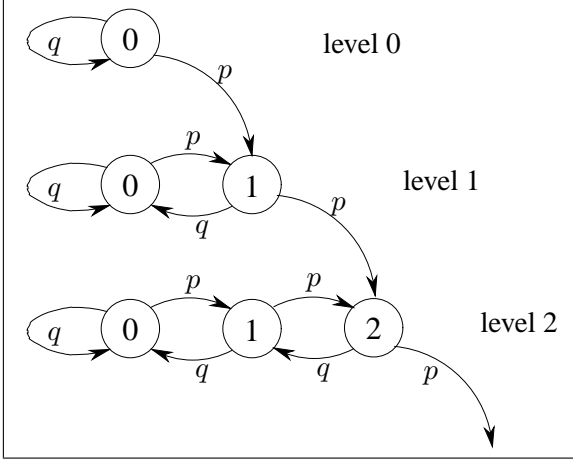
$$p_{n,l}(k) = \Pr(Z_k = (n, l)).$$

Figure 1. The chain Z.

The second component of each state (n, l) will be called the *level*, and we will denote

$$\mathcal{N}_l = \{(n, l) : 0 \le n \le l\}.$$

The set of states $\mathcal{N}_l$ will be called *level l*. Remark that if we denote by (U_k) the discrete time Markov chain obtained from (N_t) by uniformization with respect to Λ, then

$$\{Z_k = (n, l)\} \equiv \{U_k = n, \ \max\{U_j, \ j \le k\} = l\}.$$

In other words, observe that if we denote by $\mathcal{O}()$ the operator transforming (N_t) into (Y_t) (by defining the associated process (M_t)), that is, $(Y_t) = \mathcal{O}((N_t))$, and if we write $(Z_k) = \mathrm{Unif}((Y_t), \Lambda)$, then $(Z_k) = \mathrm{Unif}(\mathcal{O}((N_t)), \Lambda) = \mathcal{O}(\mathrm{Unif}((N_t), \Lambda))$, that is, the $\mathcal{O}$ operator and the uniformization one, with respect to some fixed uniformization rate, commute.

We have

$$\Pr(Y_t = (n, l)) = \sum_{k=0}^{\infty} e^{-\Lambda t} \frac{(\Lambda t)^k}{k!} p_{n,l}(k).$$

Observing that if $k < l$, Z_k can not be in level l,

$$\Pr(Y_t = (n, l)) = \sum_{k=l}^{\infty} e^{-\Lambda t} \frac{(\Lambda t)^k}{k!} p_{n,l}(k). \qquad (3)$$

Define

$$q_l(k) = \Pr(Z_k \in \mathcal{N}_l) = \sum_{n=0}^{l} p_{n,l}(k).$$

See that

$$q_l(k) = \Pr(\max\{U_j, j \le k\} = l).$$

From (3), we have

$$\Pr(M_t = l) = \sum_{k=l}^{\infty} e^{-\Lambda t} \frac{(\Lambda t)^k}{k!} q_l(k). \qquad (4)$$

So, the problem reduces to obtain a computational scheme for the $q_l(k)$'s.

Let us show how to exploit the previous development in a numerical procedure. Choose some $\varepsilon > 0$. Then compute (as usual when we use uniformization)

$$K = \min \left\{ J : \sum_{k=0}^{J} e^{-\Lambda t} \frac{(\Lambda t)^k}{k!} > 1 - \varepsilon \right\}. \qquad (5)$$

Define

$$m(t, l, \varepsilon) = \sum_{k=l}^{K} e^{-\Lambda t} \frac{(\Lambda t)^k}{k!} q_l(k). \qquad (6)$$

We have

$$m(t, l, \varepsilon) < \Pr(M_t = l)$$

and

$$\Pr(M_t = l) - m(t, l, \varepsilon) < \varepsilon.$$

Resuming, given λ, μ, L, T and ε, the following procedure returns the probability $\Pr(M_T = L)$ with absolute error less than ε:

- if $L = 0$, return $e^{-\lambda T}$ as the (exact) probability $\Pr(M_T = 0)$;

- if $L > 0$, compute K using (5) (a function of λ, μ, T and ε):
    - if $K < L$, return 0 (since, in that case, relation (6) says us that we have $\Pr(M_T = L) < \varepsilon$);
    - in the remaining case $0 < L \le K$, return $m(T, L, \varepsilon)$ defined in (6).

The rest of the paper will focus on deriving recurrences allowing to work efficiently with the $q_l()$ functions.

3 Analysis of the (Uniformized) Bi-Dimensional Random Walk

The main relation to derive an efficient computational scheme is given in the following result, where it is shown that the probability for process (Z_k) to be at level l by time k, can be given as a function of the probabilities of being in the "border" states (j, j) only.

Theorem 1 *If we denote $y_l(k) = \Pr(Z_k = (l,l)) = p_{l,l}(k)$, then, for all $k > l > 0$, we have*

$$q_l(k) = p[y_{l-1}(k-1) - y_l(k-1)] + q_l(k-1). \quad (7)$$

Proof. If we range the states by levels, and inside level l, from $(0,l)$ to (l,l) (see Figure 1), then the transition probability matrix P has the block structure

$$P = \begin{pmatrix} P_0 & Q_0 & 0 & 0 & \cdots \\ 0 & P_1 & Q_1 & 0 & \cdots \\ 0 & 0 & P_2 & Q_2 & \cdots \\ & & & \cdots & \end{pmatrix},$$

where matrix P_l has $l+1$ rows and $l+1$ columns and is tridiagonal, and matrix Q_l has $l+1$ rows and $l+2$ columns. Indexing rows and columns from 0, the former is given, for $l \geq 1$, by

$$\begin{aligned} (P_l)_{0,0} &= q, \\ (P_l)_{n,n} &= 0, \ 1 \leq n \leq l, \\ (P_l)_{n,n-1} &= q, \ 1 \leq n \leq l, \\ (P_l)_{n,n+1} &= p, \ 0 \leq n \leq l-1 \end{aligned}$$

Matrix P_0 is reduced to the scalar q. Matrix Q_l has all its elements null except $(Q_l)_{l,l+1} = p$. For instance,

$$P_2 = \begin{pmatrix} q & p & 0 \\ q & 0 & p \\ 0 & q & 0 \end{pmatrix} \quad \text{and} \quad Q_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & p \end{pmatrix}.$$

Let $\vec{z}(k)$ be the distribution of the uniformized chain Z_k, seen as a row vector, with the previously given order in the state space. If we write

$$\vec{z}^{(l)}(k) = (\, p_{0,l}(k) \ \ p_{1,l}(k) \ \ \cdots \ \ p_{n,l}(k) \ \ \cdots \ \ p_{l,l}(k) \,),$$

then

$$\vec{z}(k) = (\, \vec{z}^{(0)}(k) \ \ \vec{z}^{(1)}(k) \ \ \cdots \ \ \vec{z}^{(l)}(k) \ \ \cdots \,).$$

From $\vec{z}(k) = \vec{z}(k-1)P$ (the balance equations for (Z_k)), we have

$$\begin{aligned} \vec{z}^{(0)}(k) &= \vec{z}^{(0)}(k-1)P_0, \\ \vec{z}^{(1)}(k) &= \vec{z}^{(0)}(k-1)Q_0 + \vec{z}^{(1)}(k-1)P_1, \\ \vec{z}^{(2)}(k) &= \vec{z}^{(1)}(k-1)Q_1 + \vec{z}^{(2)}(k-1)P_2, \\ &\vdots \\ \vec{z}^{(l)}(k) &= \vec{z}^{(l-1)}(k-1)Q_{l-1} + \vec{z}^{(l)}(k-1)P_l. \end{aligned}$$

Now, $q_l(k) = \Pr(Z_k \in \mathcal{N}_l) = \vec{z}^{(l)}(k)1^{\mathrm{T}}$, where 1 denotes here a row vector having all its components equal to

1, its dimension being defined by the context, and 1^{T} is its transpose. We have

$$Q_l 1^{\mathrm{T}} = p\vec{e}_l^{\mathrm{T}}, \quad \vec{e}_l = (\, 0 \ \ 0 \ \ \cdots \ \ 0 \ \ 1 \,)$$

$$\text{and } P_l 1^{\mathrm{T}} = \begin{pmatrix} 1 \\ 1 \\ \cdots \\ 1 \\ 1-p \end{pmatrix} = 1^{\mathrm{T}} - p\vec{e}_l^{\mathrm{T}}.$$

From the balance equations, we obtain

$$\begin{aligned} q_l(k) &= \vec{z}^{(l-1)}(k-1)Q_{l-1}1^{\mathrm{T}} + \vec{z}^{(l)}(k-1)P_l 1^{\mathrm{T}} \\ &= \vec{z}^{(l-1)}(k-1)p\vec{e}_{l-1}^{\mathrm{T}} + \vec{z}^{(l)}(k-1)\left(1^{\mathrm{T}} - p\vec{e}_l^{\mathrm{T}}\right) \\ &= pz_{l-1}^{(l-1)}(k-1) + q_l(k-1) - pz_l^{(l)}(k-1). \end{aligned}$$

Recall that if $k < l$, then $q_l(k) = 0$. So, resuming,

- for $k \geq 0$, $q_0(k) = q^k$,

- for all l, $q_l(l) = p^l$,

- and for $k > l > 0$,

$$q_l(k) = p\left[y_{l-1}(k-1) - y_l(k-1)\right] + q_l(k-1),$$

which is the announced relation. ───────────●

Since (M_t) is the aggregation of (Y_t) over the levels, it is easy to verify that it is not Markov (for instance, by checking that the sojourns in the levels are not exponential –this can be done using Laplace transforms). See [5] for a general work on this topic.

4   Computational Scheme

Let us define a new discrete time homogeneous Markov chain $W^{(l)} = (W_k^{(l)})_{k \geq 0}$, for any $l \geq 0$, having state space $\{0, 1, \cdots, l, l+1\}$, transition probability matrix given in block form by

$$\begin{pmatrix} P_l & p\vec{e}_l^{\mathrm{T}} \\ 0 & 1 \end{pmatrix}$$

and initial state l; state $l+1$ is then absorbing. Let us denote the absorbing time of $W^{(l)}$ by $T^{(l)}$ (that is, $T^{(l)} = \min\{k : W_k^{(l)} = l+1\}$). The idea is of course that the space state of $W^{(l)}$ is naturally associated with level l (subset $\mathcal{N}_l$) of the state space of (Z_k), and that the evolution of $W^{(l)}$ *before*
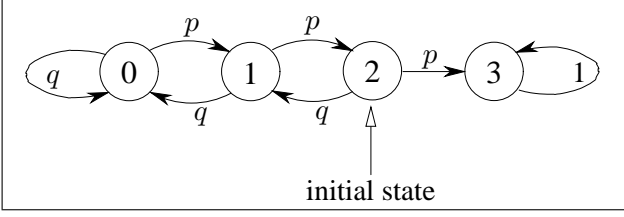
Figure 2. Chain $W^{(2)}$.

absorption is stochastically equivalent to that of (Z_k) in $\mathcal{N}_l$. In Figure 2 the graph associated with chain $W^{(2)}$ is given.

The distribution of the absorption time $T^{(l)}$

Denote $t_l(k) = \Pr(T^{(l)} = k)$. For level 0 (a two-state model), we have

$$t_0(0) = 0, \quad t_0(k) = q^{k-1}p \quad \forall k \geq 1.$$

For any level $l \geq 1$ we trivially have by "direct inspection" of chain $W^{(l)}$ that $t_l(1) = p$ and $t_l(2) = 0$. We obviously have $t_l(0) = 0$, and it is not hard to obtain other values as $t_l(3) = qp^2$. A general relation allowing us to compute the distribution of $T^{(l)}$ is given in the following result.

Relation 1 *For any $l \geq 1$ and $k \geq 3$, we have*

$$t_l(k) = q \sum_{j=2}^{k-1} t_{l-1}(j-1) t_l(k-j). \tag{8}$$

Proof. To prove this relation, we first write

$$t_l(k) = \Pr(T^{(l)} = k \mid W_1^{(l)} = l-1) \Pr(W_1^{(l)} = l-1)$$

(recall that $W_0^{(l)} = l$). Now, let us denote by R the time of the first return of $W^{(l)}$ to state l *after* time 1 (thus, $R > 1$). We have

$$\Pr(T^{(l)} = k \mid W_1^{(l)} = l-1) =$$
$$\sum_{j=2}^{k-1} \Pr(T^{(l)} = k \mid R = j, W_1^{(l)} = l-1) \Pr(R = j \mid$$
$$W_1^{(l)} = l-1).$$

From the strong Markov property,

$$\Pr(T^{(l)} = k \mid R = j, W_1^{(l)} = l-1) = \Pr(T^{(l)} = k-j)$$

and

$$\Pr(R = j \mid W_1^{(l)} = l-1) = \Pr(T^{(l-1)} = j-1).$$

Since we have $\Pr(W_1^{(l)} = l-1) = q$, the result follows. •

There is no room here to give more details on the properties of the $t_l()$ distribution. For instance, we can easily check that $t_l(2h) = 0$ if $h = 0, 1, \ldots, l$. Also, $t_l(k) = t_{l-1}(k)$ for $k = 0, 1, \ldots, 2l-1$.

The probabilities $\Pr(W_k^{(l)} = l)$

The second useful tool in deriving the distribution we are looking for is a relation to compute efficiently the numbers

$$u_l(k) = \Pr(W_k^{(l)} = l).$$

This is because the goal is to compute the $y_l()$ functions which lead to the $q_l()$ ones through Theorem 1, and because, as we will see, the $y_l()$ can be computed from the $t_l()$ functions and the $u_l()$ ones (see (12)).

Observe first that, for level 0,

$$u_0(k) = q^k \quad k \geq 0.$$

As for the $t_l()$ functions, we can prove the following result (which is done as for Relation 1):

Relation 2 *For any $l \geq 1$ and $k \geq 2$,*

$$u_l(k) = q \sum_{j=1}^{k-1} t_{l-1}(j) u_l(k-1-j). \tag{9}$$

But instead of using (9), see that

$$t_l(k) = p u_l(k-1), \tag{10}$$

which follows simply by observing that to be absorbed at time k, process $W^{(l)}$ must be back at l at time $k-1$ and that it must then go to the absorbing state $l+1$.

Hitting time of (Z_k) at level l

Let us now denote by F_l the hitting time of chain (Z_k) at level l. Denote

$$f_l(k) = \Pr(F_l = k).$$

First, $F_0 = 0$, so, $f_0(0) = 1$ and for all $k \geq 1$, $f_0(k) = 0$. For level 1,

$$f_1(0) = 0, \quad f_1(k) = q^{k-1}p \quad \forall k \geq 1.$$

For each $l \geq 1$, $F_l \in \{l, l+1, \cdots\}$, so, if $k < l$ then $f_l(k) = 0$.

Relation 3 *For $l \geq 2$ and $k \geq l$,*

$$f_l(k) = \sum_{j=l-1}^{k-1} t_{l-1}(k-j) f_{l-1}(j) \qquad (11)$$

Proof. By conditioning and standard Markov techniques,

$$
\begin{aligned}
f_l(k) &= \sum_{j=l-1}^{k-1} \Pr(F_l = k \mid F_{l-1} = j) f_{l-1}(j) \\
&= \sum_{j=l-1}^{k-1} \Pr(T^{(l-1)} = k-j) f_{l-1}(j) \\
&= \sum_{j=l-1}^{k-1} t_{l-1}(k-j) f_{l-1}(j).
\end{aligned}
$$

$\bullet$

An expression for $y_l(k)$

Recall that $y_l(k) = \Pr(Z_k = (l,l))$. This means that $y_l(k) = 0$ if $k < l$. We also immediately have $y_l(l) = p^l$.

For level 0, $y_0(k) = q^k$, $k \geq 0$. The general relation closing the analysis process is given in the following result.

Relation 4 *For $k \geq l \geq 1$,*

$$y_l(k) = \sum_{j=l}^{k} u_l(k-j) f_l(j) = \frac{1}{p} \sum_{j=l}^{k} t_l(k-j+1) f_l(j) \quad (12)$$

Proof. We have, for $k \geq l$,

$$
\begin{aligned}
y_l(k) &= \sum_{j=l}^{k} \Pr(Z_k = (l,l) \mid F_l = j) f_l(j) \\
&= \sum_{j=l}^{k} \Pr(W_{k-j}^{(l)} = l \mid F_l = j) f_l(j) \\
&= \sum_{j=l}^{k} u_l(k-j) f_l(j).
\end{aligned}
$$

Using (10) to replace $u_l(k-j)$ by $t_l(k-j+1)/p$ concludes the proof. $\bullet$

Resuming, we must compute $m(t, L, \varepsilon)$ given by (6) (in the case of $0 < L \leq K$, where K is defined in (5)). To do this, we need numbers $q_L(L), \cdots, q_L(K)$, which in turn need, to be computed, numbers $y_L(L), \cdots, y_L(K-1)$ and $y_{L-1}(L), \cdots, y_{L-1}(K-1)$. So, in fact, we only need these last $2(K-L)$ values, which are computed using (12). This means that we must only run the recurrences given in Relation 1 for the $t_l()$ functions and in Relation 3 for the $f_l()$ ones, that is, we must only compute $t_l(k)$ and $f_l(k)$, for $l = 0, 1, \ldots$, in order to reach levels $L-1$ and L, and obtain the final values $t_{L-1}(i)$ for $i = 1, \cdots, K-L+1$ and $f_{L-1}(j)$ for $j = L-1, \cdots, K-1$, and values $t_L(i)$ for $i = 1, \cdots, K-L$ and $f_L(j)$ for $j = L, \cdots, K-1$.

The cost of running (8) to obtain those $t_l(k)$ values is $O(L(K-L)^2/2)$ and the same for computing the $f_l(k)$ ones using (11). So, the algorithm runs with computational cost $O(L(K-L)^2)$. When $\Lambda T \gg 1$ (so that $K \approx \lambda T$), the cost can be written $O(L(\Lambda T - L)^2)$.

It is particularly interesting to compute $\Pr(M_T > L)$, instead of $\Pr(M_T = L)$. Slight modifications of the preceeding developments allow to do this directly. Define $V_k = \max\{U_j, j \leq k\}$ (recall that (U_j) is the occupation process in the discrete time chain obtained by uniformization of (N_t)). Writing then (4) as

$$\Pr(M_t = l) = \sum_{k=l}^{\infty} e^{-\Lambda t} \frac{(\Lambda t)^k}{k!} \Pr(V_k = l),$$

we derive that

$$\Pr(M_t > l) = \sum_{k=l}^{\infty} e^{-\Lambda t} \frac{(\Lambda t)^k}{k!} \Pr(V_k > l).$$

Now, $\Pr(V_k > l) = \Pr(F_{l+1} \leq k) = f_{l+1}(l+1) + \cdots + f_{l+1}(k)$. Thus, defining for $k \geq l$,

$$g_l(k) = \sum_{j=l}^{k} f_l(j),$$

we obtain the approximation

$$\Pr(M_t > l) \approx \sum_{k=l}^{K} e^{-\Lambda t} \frac{(\Lambda t)^k}{k!} g_l(k),$$

with K as before, thus with absolute error less than ε.

5 Some numerical examples

To illustrate the method, consider the case of $\lambda = 0.95$ and $\mu = 1$ (we scale, using as unit of time the mean service time), leading to a load of 0.95 and an average of 19
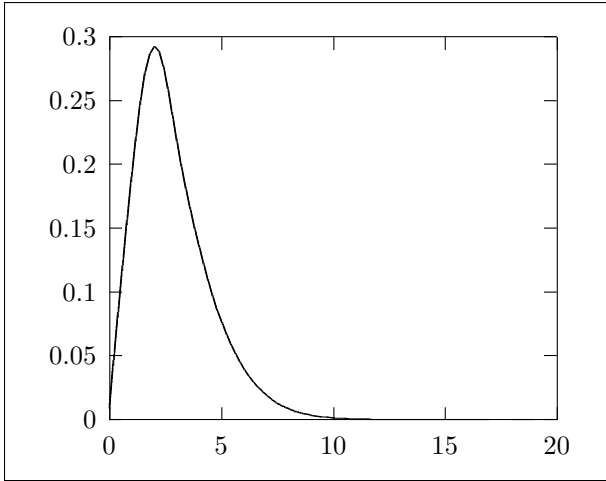
Figure 3. For $\lambda = 0.95$ **and** $\mu = 1$**, this is the plot of** $\Pr(M_5 = l)$**, for level** l **from 0 to 20 (the curve appears as "continuous" just for a better presentation.**

Figure 4. For $\lambda = 0.95$ **and** $\mu = 1$**, this is the plot of** $\Pr(M_{15} = l)$**, for level** l **from 0 to 20 (the curve appears as "continuous" just for a better presentation).**

customers in equilibrium. If the system is going to be considered for a very short period of time, for instance, for a few units of time, the probability of reaching values close to the mean backlog in equilibrium are, of course, very low. For instance, Figure 3 plots the probability of reaching levels 0 to 20 on the interval $[0, 5]$.

In Figure 4 we plot the same quantities but on the period $[0, 15]$.

Following the ideas given in the Introduction, we plot in Figure 5 the values of $\Pr(M_t > l)$ for a fixed value of t (15 in the plot). We see that in a so short period of time, the probability of observing a backlog greater than l decreases very quickly with l. For instance, we have $\Pr(M_{15} > 10) \approx 0.00025$. Recalling that the mean number of customers in equilibrium is 19 for this system, this shows how to use the algorithmic tools given here for dimensionning purposes.

6   Conclusions

The derivation of an efficient computational scheme for the distribution of the maximum level reached by a simple random walk in continuous time and over a finite period $[0, T]$ could be done by elementary combinatorial techniques mainly due to (i) the uniformization tool which allows to work in discrete time, so, to work with recurrences, and (ii) the particular structure of the model, allowing to
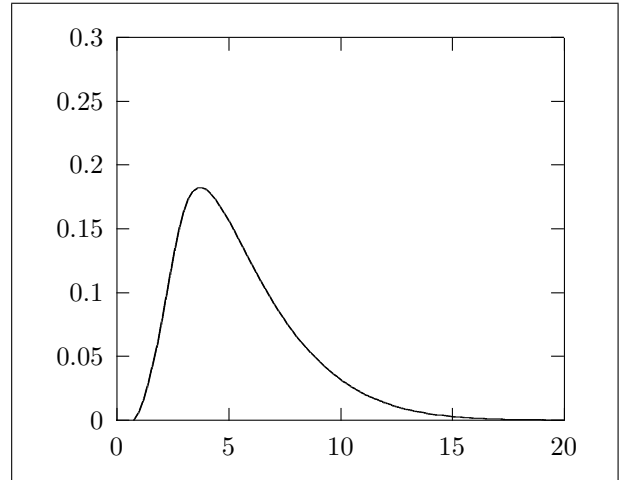

Figure 5. For $\lambda = 0.95$ **and** $\mu = 1$**, this is the plot of** $\Pr(M_{15} > l)$**, for level** l **from 0 to 20 (the curve appears as "continuous" just for a better presentation.**
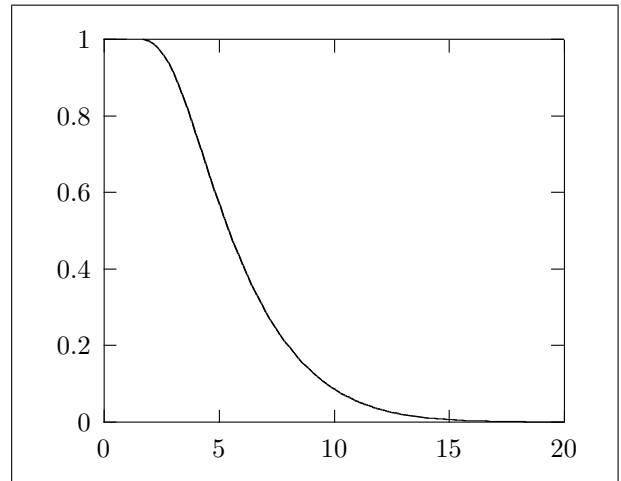
define an iterative process.

This last point allows easily to deal with more complex structures, for instance, to deal with random walks with phase-type distributions. This topic, and the detailed implementation of the resulting algorithms, are now under investigation. The corresponding recurrences are now matrix ones but the general idea is the same.

References

[1] F. Baccelli and P. Brémaud. *Elements of queuing theory*. Springer-Verlag, 1994.

[2] H.-P. H.-P. Schwefel, L. Lipsky, and M. Jobmann. On the necessity of transient performance analysis in telecommunication networks. In *Proceedings of the International Teletraffic Congress (ITC) 17*, Salvador da Bahia, Brazil, December 2-7 2001.

[3] M. Neuts. The distribution of the maximum length of a poisson queue during a busy period. *Oper. Res.*, 18:388–395, 1996.

[4] S. Ross. *Stochastic Processes*. Addison-Wesley, 1995.

[5] G. Rubino and B. Sericola. A finite characterization of weak lumpable markov processes. part ii: the continuous time case. *Stoc. Proc. and their App.*, 45, 1993.

[6] G. Rubino and B. Sericola. Interval availability analysis using denumerable markov processes. application to multiprocessor subject to breakdowns and repairs. *IEEE Trans. on Computers*, 44(2), 1995.