Queueing Systems 19 (1995) 105–130

Rejection rules in the $M/G/1$ queue

David Perry

Department of Statistics, University of Haifa, Haifa 31999, Israel

Søren Asmussen

Institute of Electronic Systems, Aalborg University, Fr. Bajersv. 7, 9220 Aalborg Ø, Denmark

Received 29 October 1993; revised 14 October 1994

We consider a $M/G/1$ queue modified such that an arriving customer may be totally or partially rejected depending on a r.v. (the barricade) describing his impatience and on the state of the system. Three main variants of this scheme are studied. The steady-state distribution is expressed in terms of Volterra equations and the relation to storage processes, dams and queues with state-dependent Poisson arrival rate is discussed. For exponential service times, we further find the busy period Laplace transform in the case of a deterministic barricade, whereas for exponential barricade it is shown by a coupling argument that the busy period can be identified with a first passage time in an associated birth-death process.

Keywords: Birth-death process; busy period; coupling; impatient customers; martingale stopping theorem; queue; storage process; virtual waiting time; Volterra equation.

1. Introduction

A queueing system has restricted accessibility if not every customer is admitted to the system. For such a queueing system the admittance of a customer will, in general, depend on the state at the moment of his arrival. A system with a limited number of waiting places is a typical example of restricted accessibility. An arriving customer who finds all the waiting places occupied is not admitted (cf. Cohen [10]).

In this paper we will focus on three generic queueing versions of the $M/G/1$-type in which the restricted accessibility is introduced in the sense of the impatience. We consider two rejection rules based on different quantities:

(i) The waiting time in front of a customer should not exceed a specific (random) time (the impatience);
(ii) the sojourn time of a customer should not exceed a specific (random) time.

In case (i), the rejected customer is fully rejected (Model I). In case (ii), a rejected customer may either enter the system but his sojourn time is restricted (Model II) or he is totally rejected (Model III).

In a traditional queueing setting, rejection based upon the workload upon arrival would require that the service times of the customer in line and the residual service time of the customer in service were known either to the server or to the arriving customer. Examples of the first type are a queue of jobs in front of a CPU where the execution time (= service time) would often be known and a queue of data messages waiting to be transmitted (here the service time is roughly proportional to the number of packets in the message). However, sometimes completely interpretations (say in a dam context) are possible. In particular, Model I arises in perishable inventory systems [17, 18].

We denote the Poisson arrival rate by λ, let G be the distribution of the service requirement and $\hat{G}[\theta] = \int_0^\infty e^{-\theta t} dG(t)$ its Laplace–Stieltjes transform (LT). We number the customers $n = 0, 1, 2, \ldots$, let S_n be the service requirement of the nth customer, τ_n his arrival time and W_n the workload upon arrival, $W_n = V(\tau_n -)$ where $\{V(t)\}_{t \geq 0}$ is the workload (virtual waiting time) process. For each of the three models the restricted accessibility are described by barricade random variables $U_0, U_1, \ldots$ with common distribution H. We assume that $\{S_n\}$ and $\{U_n\}$ are two independent i.i.d. sequences and independent also of the arrival process. If the customer is admitted he waits for service which is according to the time of arrival (FCFS). If the customer is fully rejected, it is assumed that he never returns.

In the first model to be studied (Model I), customer n is admitted to the system if and only if $W_n < U_n$, where U_n is the maximal amount of time he is willing to wait in line. That is,

$$V(\tau_n) = W_n + S_n \cdot 1_{\{W_n < U_n\}}. \tag{1.1}$$

In the second model (Model II), customer n will have his service requirement reduced, if necessary, so that his time in system does not exceed U_n. That is, he is admitted if and only if $W_n < U_n$ and if so, his reduced service requirement is $\min(S_n, U_n - W_n)$:

$$V(\tau_n) = W_n + \min(S_n, (U_n - W_n)^+). \tag{1.2}$$

The third model (Model III) differs from Model II in that customer n does not have his service requirement reduced if $W_n < U_n$ but is totally lost,

$$V(\tau_n) = W_n + S_n \cdot 1_{\{W_n + S_n < U_n\}}. \tag{1.3}$$

This admittance policy is of the so-called 'all or nothing' type.

This paper gives various new results on these models combined with some review material. Both steady-state distributions and busy period analysis are considered. We concentrate most of our efforts on exponential service times ($G(x) = 1 - e^{-\mu x}$) and the two most important cases of the distribution H of the barricade r.v. U, the deterministic case ($U \equiv 1$) and the exponential case

($H(x) = 1 - e^{-\eta x}$); for the steady-state theory also some more general results are given. Basically, Model III turns out to be the more difficult one to analyse.

It may be worthwhile to notice that in the case of a deterministic barricade, Model II is just the finite dam [2, Ch. XIII] and quite a lot is known about steady-state distributions; in particular, if the corresponding infinite dam (i.e., the corresponding $M/M/1$ queue) is stable, the stationary distribution can just be obtained by truncation. A further important observation is the following (see section 2 for the proof). Let $M^{(x)}/M/1$ refer to an $M/M/1$ queue with arrival rate $\lambda(x)$ depending on the current state $x = V(t)$.

PROPOSITION 1.1

Assume $G(x) = 1 - e^{-\mu x}$, $H(x) = 1 - e^{-\eta x}$. Then the workload process for the queue with restricted accessibility is the same as for a $M^{(x)}/M/1$ queue with $\lambda(x) = \lambda e^{-\eta x}$ for Models I, II,

$$\lambda(x) = \frac{\lambda\mu}{\mu + \eta} e^{-\eta x}$$

for Model III, and the service rate μ for Model I, $\mu + \eta$ for Models II, III. □

The paper is organised as follows. Section 2 deals with steady-state distributions. In particular, some basic Volterra integral equations are considered and the relation to storage (dam) processes is discussed. In section 3, we compute the LT of the busy period for a deterministic barricade and exponential service times. Two approaches are given, one based upon differential equations and one upon martingales. Finally, section 4 deals with the LT of the busy period for exponential barricades and exponential service times. An associated birth-death process is exhibited, such that the busy period is the same as the passage time from state 1 to state 0 for the birth-death process, and we discuss certain computational aspects, in particular moments. The crucial feature of the analysis is a coupling argument, representing the birth-death process and the queue with restricted accessibility on a common probability space.

Aspects of Models I, II have been studied in Gnedenko and Kovalenko [14] by viewing the models as special case of the $M^{(x)}/G^{(x)}/1$ queue (the $M/G/1$ queue with arrival rate $\lambda(x)$ and service distribution depending on the current state $x = V(t)$). Most of the relevant literature up to 1982 cited in Cohen [11]. Queuing models with restrictions on the waiting times are introduced in Cohen [10], Daley [12] and Gavish and Schweitzer [13]. A recent paper by Knessl et al. [19] gives approximations for the busy period distribution. However, as far as is known to the authors, there are only few exact results on busy period analysis in the literature.

Models of queues with restricted accessibility are extensive, and the reader is referred to Stoyan [28] and Whitt [29] for a general discussion. For the FCFS case, most of the $M/G/1$-type models consider queues with finite waiting rooms (see Wolff [30] and Sonderman [27]). Two recent relevant works by Bhattacharya and Ephremides [7, 8] have considered a similar problem; the question of interest is how the departure process varies in accordance with the service discipline and the further parameters of the system.

2. The workload process – stationary distributions

We first find the necessary and sufficient conditions for positive recurrence. For general results on $M^{(x)}/G^{(x)}/1$ queues, see Afaneseva [1] (reproduced in [14]) and Browne and Sigman [9]. Write $\bar{G}(x) = 1 - G(x)$, $\bar{H}(x) = 1 - H(x)$.

PROPOSITION 2.1

The workload process $\{V(t)\}$ is positive recurrent if and only if

$$\mathbb{E}S < \infty \quad \text{Model I,} \tag{2.1}$$

$$\mathbb{E}\min(S, U) < \infty \quad \text{Model II,} \tag{2.2}$$

$$\mathbb{E}[S; S < U] < \infty \quad \text{Model III.} \tag{2.3}$$

Proof

The distribution of $\{V(t)\}$ just after a jump out of 0 is that of S, resp. $\min(S, U)$, resp. S given $S < U$ for the three models. Hence the stated conditions are necessary for state 0 to have a recurrence time with finite mean, hence for positive recurrence.

For sufficiency, let for $x > 0$

$$m(x) = \lim_{h \downarrow 0} \frac{\mathbb{E}_x V(h) - x}{h} = \begin{cases} \lambda\bar{H}(x)\mathbb{E}S - 1 & \text{Model I,} \\ \lambda\mathbb{E}\min(S, (U - x)^+) - 1 & \text{Model II,} \\ \lambda\mathbb{E}[S; x + S < U] - 1 & \text{Model III.} \end{cases}$$

Clearly, $m(x)$ is a decreasing function of x, so it suffices ([2, pp. 18–19] or [9]) to show that $\limsup_{x\to\infty} m(x) < 0$, which in turn follows if each of the expressions in front of -1 converge to 0. This is obvious for Model I, and follows by dominated convergence for Models II, III since $(U - x)^+ \downarrow 0$, $I(x + S < U) \downarrow 0$. $\square$

PROPOSITION 2.2

$\{V(t)\}$ has a unique stationary distribution, given by an atom of size $\pi = \lim_{t\to\infty} \mathbb{P}(V(t) = 0)$ at 0 and a density f on $(0,\infty)$ satisfying

$$f(x) = \pi K(x,0) + \int_0^x K(x,y)f(y)dy, \tag{2.4}$$

where $K(x,y) = 0$, $y > x$, and for $y \le x$

$$K(x,y) = \begin{cases} \lambda \bar{H}(y)\bar{G}(x-y) & \text{Model I,} \\ \lambda \bar{H}(x)\bar{G}(x-y) & \text{Model II,} \\ \lambda \int_x^\infty (G(z-y) - G(x-y))dH(z) & \text{Model III.} \end{cases} \tag{2.5}$$

Proof

We use a familiar up- and downcrossing argument, see e.g. [2, p. 291] or [21, Th. 1]. Redefining $K(x,y)$ as

$$K(x,y) = \lambda \mathbb{P}(V_{\tau_n} > x - y \mid V_{\tau_n -} > y), \tag{2.6}$$

the rate of downcrossing level x in the steady state in the time interval $[t, t+dt]$ is $f(x)dt$, the rate of upcrossing is dt multiplied by the r.h.s. of (2.4). Since these two rates are equal in stationarity, (2.4) follows (for a different proof, see [14]). Now check that (2.6) has the form (2.5) in each of the three models. □

Except for the easily verified form (2.5) of K, Proposition 2.2 is essentially contained in [14]. The integral equation (2.4) is of Volterra type (see e.g. [5]) and it is well known that the solution can be found in the following general form:

PROPOSITION 2.3

For all three models I, II, III,

$$\frac{1}{\pi} = 1 + \int_0^\infty K^*(x,0)dx, \quad f(x) = \pi K^*(x,0), \tag{2.7}$$

where K^* is finite and defined by $K^{(1)} = K$,

$$K^{(n+1)}(x,y) = \int_y^x K(x,z)K^{(n)}(z,y)dy, \quad K^* = \sum_{n=1}^{\infty} K^{(n)}. \tag{2.8}$$

□

Volterra equations also come up in storage processes (Harrison and Resnick [15] or [2, Ch. XIII]), and we shall see below that in many cases one can in fact link queues with restricted accessibility and storage processes together. The relevant background is the following. Let $\{V_0(t)\}_{t\geq 0}$ be a storage process with arrival intensity λ_0, release rate $r(x)$ and jump size distribution G_0. Assume for convenience that the mean of G_0 is finite, that $r(x) \to \infty$, $x \to \infty$, and that $\omega(x) = \lambda_0 \int_0^x 1/r(y)dy$ is finite for all $x > 0$. Then $\{V_0(t)\}$ is positive recurrent and the stationary distribution is given by an atom π_0 at 0 and a density $f_0(x)$ on $(0, \infty)$, and we have

$$f_0(x) = \pi_0 K_0(x,0) + \int_0^x K_0(x,y)f_0(y)dy, \tag{2.9}$$

$$K_0(x,y) = \frac{1}{r(x)}\bar{G}_0(x-y). \tag{2.10}$$

Given π_0, (2.9) can be solved by iteration as in Proposition 2.3 by letting $K_0^{(1)} = K$,

$$K_0^{(n+1)}(x,y) = \int_y^x K_0(x,z)K_0^{(n)}(z,y)dy, \quad K_0^* = \sum_{n=1}^{\infty} K_0^{(n)}. \tag{2.11}$$

Then K_0^* is well-defined and $f_0(x) = \pi_0 K^*(x,0)$. In particular, the stationary distribution is given by

$$\frac{1}{\pi_0} = 1 + \int_0^{\infty} K_0^*(x,0)dx, \quad f_0(x) = \pi_0 K_0^*(x,0). \tag{2.12}$$

The cases the authors know of where an explicit solution of (2.9) can be obtained are:

(a) exponential jumps, $\bar{G}_0(x) = e^{-\mu x}$, where

$$K_0^*(x,0) = \frac{\lambda_0}{r(x)}e^{\omega(x)-\mu x}; \tag{2.13}$$

(b) constant release rule $r(x) \equiv 1$ and phase-type jump distribution, where $f_0(x)$ has a closed matrix-exponential form (cf. Neuts [22]);

(c) two-step release rule

$$r(x) = \begin{cases} r_1, & x \le x_0, \\ r_2, & x > x_0, \end{cases}$$

and phase-type jump distribution, where $f_0(x)$ is found in Asmussen and Bladt [4] on closed matrix-exponential form (though somewhat more complicated than in (b)). This formula may in fact have some relevance for queues with restricted accessibility, because it gives the stationary distribution for Models I, II for the case where the barricade takes on two values x_0, ∞ (i.e., with a certain probability there is no barricade for the customer, otherwise the barricade is deterministically equal to x_0), cf. Corollary 2.1 and Theorem 2.2.

Before specializing to Models I, II, III, we give:

Proof of Proposition 1.1

Let S^* be a r.v. having the distribution of the service requirement of an accepted customer. For Model I, clearly $S^* \stackrel{\mathscr{D}}{=} S$ and

$$\lambda(x) = \lambda\mathbb{P}(U > x) = \lambda e^{-\eta x}.$$

In Model II, the customer is accepted when $S > x$, so that the same expression for $\lambda(x)$ holds. Further

$$\mathbb{P}(S^* > y) = \mathbb{P}(\min(S, (U-x)^+) > y \,|\, U > x) = \mathbb{P}(\min(S, U) > y) = e^{-(\mu+\eta)x}.$$

For Model III, similarly,

$$\lambda(x) = \lambda\mathbb{P}(x + S < U) = \lambda\mathbb{E}e^{-\eta(x+S)} = \lambda e^{-\eta x}\frac{\mu}{\mu+\eta},$$

$$\mathbb{P}(S^* \in dy) = \mathbb{P}(S \in dy \,|\, x + S < U) = \frac{\mathbb{P}(S \in dy, x + y < U)}{\mathbb{P}(x + S < U)}$$

$$= \frac{\mu e^{-\mu y} e^{-\eta(x+y)}}{\lambda(x)/\lambda} = (\mu+\eta)e^{-(\mu+\eta)y}.$$

□

2.1. MODEL I

We view this model as the particular case $\lambda(x) = \lambda\bar{H}(x)$ of the $M^{(x)}/G/1$

queue. A sufficient condition for positive recurrence is that $\lambda(x) \to 0$, $x \to \infty$, and that G has finite mean.

THEOREM 2.1

For a $M^{(x)}/G/1$ queue, the stationary distribution of the workload process $\{V(t)\}$ is given by an atom of size π at 0 and a density $f(x)$ on $(0,\infty)$, where

$$\frac{1}{\pi} = 1 + \int_0^\infty \frac{\lambda(0)}{\lambda(x)} K_0^*(x,0)dx, \quad f(x) = \frac{\lambda(0)\pi}{\lambda(x)} K_0^*(x,0), \tag{2.14}$$

and K_0^* is given by (2.11) corresponding to $G_0 = G$, $\lambda_0 = 1$, $r(x) = 1/\lambda(x)$.

Proof

By the same argument as in theorem 2.2, the stationary density satisfies a Volterra equation of the form (2.4) with $K(x,y) = \lambda(y)\bar{G}(x-y)$. Letting $f_0(x) = f(x)\lambda(x)/\lambda(0)$, (2.4) takes the form (2.9) and we get $f_0(x) = \pi K_0^*(x,0)$, from which (2.14) immediately follows. □

Remark 2.1

The probabilistic significance of Theorem 2.1 is that the storage process $\{V_0(t)\}$ and the workload process $\{V(t)\}$ of the $M^{(x)}/G/1$ queue can be coupled as (random) time transformation of each other,

$$V_0(t) = V(\Lambda^{-1}(t)) \quad \text{where} \quad \Lambda(t) = \int_0^t \lambda(V_y)dy. \tag{2.15}$$

To see this, let $\{V(t)\}$ be given and define $V_0(t)$, $\Lambda(t)$ by (2.15). Then the derivative of $\Lambda^{-1}(t)$ is $1/\lambda(V(\Lambda^{-1}(t)))$ and hence the probability that $\{V_0(t)\}$ has an arrival in $[t, t+dt]$ is

$$(\Lambda^{-1}(t+dt) - \Lambda^{-1}(t))\lambda(V(\Lambda^{-1}(t))) = \frac{d}{dt}\Lambda^{-1}(t)dt \cdot \lambda(V(\Lambda^{-1}(t))) = 1.$$

Also, in between jumps we have

$$\frac{d}{dt}V_0(t) = \frac{d}{dt}\Lambda^{-1}(t) \cdot \frac{d}{dt}V(t) = \frac{1}{\lambda(V_{\Lambda^{-1}(t)})} - 1 = -\frac{1}{\lambda(V_0(t))} = -r(V_0(t)),$$

so that indeed $\{V_0(t)\}$ evolves as the proper storage process. □

COROLLARY 2.1

The stationary distribution of Model I is given by

$$\frac{1}{\pi} = 1 + \int_0^\infty \frac{1}{\bar{H}(x)} K_0^*(x,0)dx, \quad f(x) = \frac{\pi}{\bar{H}(x)} K_0^*(x,0), \tag{2.16}$$

and K^* is given by (2.11) corresponding to $G_0 = G$, $\lambda_0 = 1$, $r(x) = 1/\lambda\bar{H}(x)$.

COROLLARY 2.2

For a $M^{(x)}/M/1$ queue with $\bar{G}(x) = e^{-\mu x}$,

$$\frac{1}{\pi} = 1 + \lambda(0) \int_0^\infty e^{\omega(x) - \mu x} dx, \quad f(x) = \pi\lambda(0) e^{\omega(x) - \mu x}, \tag{2.17}$$

where $\omega(x) = \int_0^x \lambda(y)dy$. In particular, if $\bar{G}(x) = e^{-\mu x}$ in Model I, then

$$\frac{1}{\pi} = 1 + \lambda \int_0^\infty e^{\omega(x) - \mu x} dx, \quad f(x) = \pi\lambda e^{\omega(x) - \mu x}, \tag{2.18}$$

where $\omega(x) = \lambda \int_0^x \bar{H}(y)dy$.

Proof

By (2.13). □

Assuming also H to be exponential, we obtain a closed form solution in terms of an incomplete Gamma integral:

COROLLARY 2.3

If $\bar{G}(x) = e^{-\mu x}$, $\bar{H}(x) = e^{-\eta x}$ in Model I, then

$$\frac{1}{\pi} = 1 + \frac{e^{\lambda/\eta}\eta^{\mu/\eta}}{\eta\lambda^{\mu/\eta - 1}} \int_0^{\lambda/\eta} y^{\mu/\eta - 1} e^{-y} dy, \quad f(x) = \pi\lambda e^{\lambda/\eta} e^{-\lambda e^{-\eta x}/\eta - \mu x}. \tag{2.19}$$

Proof

We get $\omega(x) = \lambda(1 - e^{-\eta x})/\eta$,

$$\int_0^\infty e^{\omega(x) - \mu x} dx = e^{\lambda/\eta} \int_0^\infty e^{-\lambda e^{\eta x}/\eta - \mu x} dx$$

$$= \frac{e^{\lambda/\eta}\eta^{\mu/\eta}}{\eta\lambda^{\mu/\eta}} \int_0^{\lambda/\eta} y^{\mu/\eta - 1} e^{-y} dy.$$

□

In the case of a deterministic barricade ($\bar{H}(x) = 1_{\{x<1\}}$), similar straightforward calculus yields:

COROLLARY 2.4

If $\bar{G}(x) = e^{-\mu x}$, $\bar{H}(x) = 1_{\{x<1\}}$ in Model I, then

$$\frac{1}{\pi} = 1 + \lambda\left(\frac{1 - e^{-\gamma}}{\gamma} + \frac{e^{-\gamma}}{\mu}\right),$$

$$f(x) = \begin{cases} \pi\lambda e^{-\gamma x}, & x \le 1, \\ \pi\lambda e^{\lambda} e^{-\mu x}, & x \ge 1. \end{cases}$$

Here and in the following $\gamma = \mu - \lambda$; note that γ is the rate parameter of the exponential stationary workload distribution of the $M/M/1$ queue with full accessibility.

2.2. MODEL II

THEOREM 2.2

The stationary distribution of Model II is given by

$$\frac{1}{\pi} = 1 + \int_0^\infty K_0^*(x, 0) dx, \quad f(x) = \pi K_0^*(x, 0), \tag{2.20}$$

and K_0^* is given by (2.11) corresponding to $G_0 = G$, $\lambda_0 = \lambda$, $r(x) = 1/\bar{H}(x)$.

Proof

This follows by an immediate comparison of (2.5) and (2.10). □

COROLLARY 2.5

If $\bar{G}(x) = e^{-\mu x}$ in Model II, then

$$\frac{1}{\pi} = 1 + \lambda \int_0^\infty \bar{H}(x) e^{\omega(x) - \mu x} dx, \quad f(x) = \pi \lambda \bar{H}(x) e^{\omega(x) - \mu x}, \tag{2.21}$$

where $\omega(x) = \lambda \int_0^x \bar{H}(y) dy$.

COROLLARY 2.6

If $\bar{G}(x) = e^{-\mu x}$, $\bar{H}(x) = e^{-\eta x}$ in Model II, then the stationary distribution is obtained by replacing μ by $\mu + \eta$ in Corollary 2.3.

Proof

By Proposition 1.1. □

COROLLARY 2.7

If $\bar{G}(x) = e^{-\mu x}$, $\bar{H}(x) = 1_{\{x<1\}}$ in Model II, then

$$\frac{1}{\pi} = 1 + \lambda \frac{e^{\lambda - \mu} - 1}{\lambda - \mu},$$

$$f(x) = \begin{cases} \pi \lambda e^{-\gamma x}, & x \leq 1, \\ 0, & x > 1. \end{cases}$$

This follows easily from Corollary 2.5. Note, however, that in the case of a deterministic barricade Model II coincides with the finite dam and thus the solution must be considered to be well known.

2.3. MODEL III

For this model, there appears to be no direct storage process interpretation of the stationary distribution in general. However, we can still find the solution for the case $\bar{G}(x) = e^{-\mu x}$, $\bar{H}(x) = e^{-\eta x}$. In fact, Proposition 1.1 yields immediately

COROLLARY 2.8

If $\bar{G}(x) = e^{-\mu x}$, $\bar{H}(x) = e^{-\eta x}$ in Model III, then the stationary distribution is obtained by replacing λ by $\lambda\mu/(\mu+\eta)$ and μ by $\mu+\eta$ in Corollary 2.3. □

The case $\bar{G}(x) = e^{-\mu x}$, $\bar{H}(x) = 1_{\{x<1\}}$ of exponential service times and a deterministic barricade appears remarkably intricate, and we do not yet have a solution.

3. Busy period analysis – deterministic barricade

In this section we compute the LT $\beta(\theta) = \mathbb{E}e^{-\theta B}$ of the busy period for the particular case $\bar{G}(x) = e^{-\mu x}$ and $\bar{H}(x) = 1_{\{x<1\}}$.

As preparation, note that the standard expression p. 93 in [2] for $\beta(\theta)$ in the $M/M/1$ queue with full accessibility can be rewritten as

$$\beta(\theta) = \frac{\mu}{\mu - \alpha_-}, \tag{3.1}$$

where $\alpha_\pm = \alpha_\pm(\theta)$ are the two roots of

$$\alpha^2 + (\lambda + \theta - \mu)\alpha - \mu\theta = 0, \tag{3.2}$$

i.e.

$$\alpha_\pm = \frac{\mu - \lambda - \theta \pm \sqrt{(\lambda + \theta - \mu)^2 + 4\theta\mu}}{2} = \frac{\mu - \lambda - \theta \pm \sqrt{(\lambda + \mu + \theta)^2 - 4\lambda\mu}}{2} \tag{3.3}$$

(note that for the ease of notation we often omit θ and write α_+ instead of $\alpha_+(\theta)$ etc.). More generally,

$$\beta(x;\theta) = \mathbb{E}[e^{-\theta B} \mid V(0) = x] = e^{\alpha_- x}. \tag{3.4}$$

We shall see here that for Models I, II, III we get a similar solution but with both α_+, α_- involved:

THEOREM 3.1

The LT of busy period starting from $V(0) = x$ is given by

$$\beta(x;0) = C_+(\theta)e^{\alpha_+ x} + C_-(\theta)e^{\alpha_- x}, \quad 0 \le x \le 1, \tag{3.5}$$

where

$$C_+(\theta) = \frac{a_-(\theta)}{a_-(\theta) - a_+(\theta)}, \quad C_-(\theta) = \frac{a_+(\theta)}{a_+(\theta) - a_-(\theta)} \tag{3.6}$$

with

$$a_\pm = \begin{cases} e^{\alpha_\pm}\left(\lambda + \theta + \alpha_\pm - \dfrac{\lambda\mu}{\mu+\theta}\right) & \text{Model I,} \\ e^{\alpha_\pm}(\theta + \alpha_\pm) & \text{Model II,} \\ e^{\alpha_\pm}(\lambda + \theta + \alpha_\pm) & \text{Model III.} \end{cases} \tag{3.7}$$

For Model I, one has in addition that

$$\beta(x;\theta) = e^{-\theta(x-1)}\beta(1;\theta), \quad x > 1. \tag{3.8}$$

Note that (3.6) is equivalent to the set of linear equations

$$C_+(\theta) + C_-(\theta) = 1, \quad a_+(\theta)C_+(\theta) + a_-(\theta)C_-(\theta) = 0; \tag{3.9}$$

the first of these equations amounts to $\beta(0;\theta) = 1$, which is probabilistically obvious, whereas the second will come out by considering the behaviour at the other boundary $x = 1$.

The proof of Theorem 3.1 is given in section 3.1 via differential equations (all three models) and an alternative proof using martingales is given in section 3.2 (Models I, II). Section 3.3 further exploits the martingale approach by direct calculations of the two first moments of B.

Given (3.5), (3.9), we can compute the busy period LT $\beta(\theta)$ itself by integrating x w.r.t. the distribution F_B (say) of $V(0)$ initiating a proper busy period:

COROLLARY 3.1

The LT $\beta(0)$ of the busy period starting with a jump from $V(0-) = 0$ is

$$C_+\left(\frac{\mu(e^{\alpha_+-\mu}-1)}{\alpha_+-\mu} + e^{\alpha_+-\mu}\frac{\mu}{\mu+\theta}\right) + C_-\left(\frac{\mu(e^{\alpha_--\mu}-1)}{\alpha_--\mu} + e^{\alpha_--\mu}\frac{\mu}{\mu+\theta}\right) \quad \text{Model I,}$$

$$C_+\frac{\alpha_+e^{\alpha_+-\mu}-\mu}{\alpha_+-\mu} + C_-\frac{\alpha_-e^{\alpha_--\mu}-\mu}{\alpha_--\mu} \quad \text{Model II,}$$

$$C_+\frac{e^{\alpha_+-\mu}-1}{(\alpha_+-\mu)(1-e^{-\mu})} + C_-\frac{e^{\alpha_--\mu}-1}{(\alpha_--\mu)(1-e^{-\mu})} \quad \text{Model III.}$$

Proof

Let S be exponential with parameter μ and note that:

for Model I, F_B is the distribution of S;
for Model II, F_B is the distribution of $S \wedge 1$;
for Model III, F_B is the conditional distribution of S given $S \leq 1$.

Thus, e.g., for Model II we get

$$\begin{aligned}\beta(\theta) &= \int_0^1 \beta(x;\theta) F_B(dx) \\ &= \int_0^1 \beta(x;\theta)\mu e^{-\mu x} dx + e^{-\mu}\beta(1;\theta).\end{aligned}$$

Inserting (3.5), this is easily seen to reduce to the stated expression. □

3.1. THE BUSY PERIOD LT VIA DIFFERENTIAL EQUATIONS – MODELS I, II, III

Consider first Model III. Conditioning upon arrival or non-arrival in $[0, dt]$, we get

$$\begin{aligned}\beta(x) &= (1-\lambda dt)\beta(x-dt)e^{-\theta dt} + \lambda dt \int_x^1 \beta(y)\cdot e^{-\mu(y-x)} dy \\ &= \beta(x)\left(1 - \lambda dt - \frac{\beta'(x)}{\beta(x)} dt - \theta dt\right) + \lambda dt \int_x^1 \beta(y)\mu e^{-\mu(y-x)} dy, \qquad (3.10)\end{aligned}$$

$$0 = -(\lambda+\theta)\beta(x) - \beta'(x) + \lambda e^{\mu x} \int_x^1 \beta(y)\mu e^{-\mu y} dy, \qquad (3.11)$$

$$0 = -(\lambda+\theta)\beta'(x) - \beta''(x) + \lambda e^{\mu x} \int_x^1 \beta(y)\mu e^{-\mu y} dy - \lambda\mu\beta(x),$$

$$0 = \beta''(x) + (\lambda+\theta-\mu)\beta'(x) - \mu\theta\beta(x). \qquad (3.12)$$

The characteristic equation of the differential equation (3.12) is (3.2), and thus the general solution is of the form (3.5). The second boundary condition in (3.9) is obtained by letting $x = 1$ in (3.11), giving

$$0 = -(\lambda + \theta)(C_+ e^{\alpha_+} + C_- e^{\alpha_-}) - C_+ \alpha_+ e^{\alpha_+} - C_- \alpha_- e^{\alpha_-}.$$

For Model I, the event of a possible upcrossing of level 1 gives an additional term

$$\lambda dt e^{-\mu(1-x)} \mathbb{E} e^{-\theta S} \beta(1) = \lambda dt e^{-\mu(1-x)} \frac{\mu}{\mu + \theta} \beta(1)$$

in (3.10) so that (3.11) takes the form

$$0 = -(\lambda + \theta)\beta(x) - \beta'(x) + \lambda e^{\mu x} \int_x^1 \beta(y) \mu e^{-\mu y} dy + \lambda e^{-\mu(1-x)} \frac{\mu}{\mu + \theta} \beta(1). \tag{3.13}$$

Differentiating, this leads again to (3.12), and letting $x = 1$ in (3.13) yields

$$0 = -(\lambda + \theta)(C_+ e^{\alpha_+} + C_- e^{\alpha_-}) - C_+ \alpha_+ e^{\alpha_+} - C_- \alpha_- e^{\alpha_-} + \lambda \frac{\mu}{\mu + \theta}(C_+ e^{\alpha_+} + C_- e^{\alpha_-}).$$

The calculations for Model II are similar except that $\mathbb{E} e^{-\theta S}$ has to be replaced by 1.

3.2. THE BUSY PERIOD LT $\mathbb{E} e^{-\theta B}$ VIA MARTINGALES – MODELS I, II

Let $\{\tilde{V}(t)\}$ be the input process. I.e., $\{\tilde{V}(t)\}$ is obtained from $\{V(t)\}$ by removing the reflection at 0 so that $\{\tilde{V}(t)\}$ decreases deterministically at unit rate between the Poisson arrival times, and has an upwards jump governed by G at arrival times. It is then well known (and easily verified) that

$$\mathbb{E} e^{\alpha \tilde{V}(t)} = e^{t\kappa(\alpha)} \quad \text{where} \quad \kappa(\alpha) = \lambda(\mathbb{E} e^{\alpha S} - 1) - \alpha = \lambda \frac{\alpha}{\mu - \alpha} - \alpha,$$

and that

$$\{e^{\alpha \tilde{V}(t) - t\kappa(\alpha)}\}_{t \geq 0} \tag{3.14}$$

is a martingale. Furthermore, for a given θ the equation $\kappa(\alpha) = \theta$ is readily seen to be the same as (3.2), i.e. there are two roots which are precisely α_+, α_-.

In the following, let

$$L = B \wedge \inf\{t > 0; \tilde{V}(t) > 1\} = \inf\{t > 0 : \tilde{V}(t) \notin (0, 1]\},$$

$$\beta_*(x;\theta) = \mathbb{E}_x[e^{-\theta L}; L = B] = \mathbb{E}_x[e^{-\theta L}; \tilde{V}(L) = 0],$$

$$\beta^*(x;\theta) = \mathbb{E}_x[e^{-\theta L}; L < B] = \mathbb{E}_x[e^{-\theta L}; \tilde{V}(L) > 1].$$

To compute $\beta_*(x;\theta)$, $\beta^*(x;\theta)$, we apply optional stopping of (3.14) with $\alpha = \alpha_+$ or $\alpha = \alpha_-$, and get

$$\begin{aligned} e^{\alpha_\pm x} &= \mathbb{E}_x e^{\alpha_\pm \tilde{V}(0) - 0 \cdot \kappa(\alpha)} \\ &= \mathbb{E}_x e^{\alpha_\pm \tilde{V}(L) - L\kappa(\alpha)} = \mathbb{E}_x e^{\alpha_\pm \tilde{V}(L) - \theta L} \\ &= \mathbb{E}_x[e^{-\theta L}; \tilde{V}(L) = 0] + \mathbb{E}_x[e^{\alpha_\pm(1+S) - \theta L}; \tilde{V}(L) > 1] \\ &= \beta_*(x;\theta) + b_\pm \beta^*(x;\theta), \end{aligned}$$

where $b_\pm = \mu e^{\alpha_\pm}/(\mu - \alpha_\pm)$. It follows that

$$\beta_*(x;\theta) = \frac{b_- e^{\alpha_+ x} - b_+ e^{\alpha_- x}}{b_- - b_+}, \quad \beta^*(x;\theta) = \frac{-e^{\alpha_+ x} + e^{\alpha_- x}}{b_- - b_+}. \tag{3.15}$$

For Model I,

$$B = L + (S + B_1) 1_{\{L<B\}}. \tag{3.16}$$

Here L, S, B_1 are independent, $\mathbb{P}(S > x) = e^{-\mu x}$ and B_1 has the P_1-distribution of B. Hence

$$\beta(x;\theta) = \beta_*(x;\theta) + \beta^*(x;\theta)\frac{\mu}{\mu+\theta}\beta(1;\theta). \tag{3.17}$$

Here β_*, β^* are known by (3.15), so that we can let $x = 1$ to solve for the remaining unknown $\beta(1;\theta)$. For Model II,

$$B = L + B_1 1_{\{L<B\}}, \tag{3.18}$$

$$\beta(x;\theta) = \beta_*(x;\theta) + \beta^*(x;\theta)\beta(1;\theta), \tag{3.19}$$

which can be solved in a similar way.

The verification that this procedure leads to the same expressions as Theorem 3.1 is somewhat tedious. We shall carry out the check for Model II, where we get

$$\beta(1;\theta) = \frac{\beta_*(1;\theta)}{1-\beta^*(1;\theta)} = \frac{b_- e^{\alpha_+} - b_+ e^{\alpha_-}}{b_- - b_+ + e^{\alpha_+} - e^{\alpha_-}},$$

$$\beta(x;\theta) = \frac{b_- e^{\alpha_+ x} - b_+ e^{\alpha_- x}}{b_- - b_+} + \frac{-e^{\alpha_+ x} + e^{\alpha_- x}}{b_- - b_+} \frac{b_- e^{\alpha_+ x} - b_+ e^{\alpha_- x}}{b_- - b_+ + e^{\alpha_+} - e^{\alpha_-}}. \tag{3.20}$$

Noting that $\alpha_+\alpha_- = \mu\theta$ and that

$$b_+ - e^{\alpha_+} = \frac{\alpha_+ e^{\alpha_+}}{\mu - \alpha_+}, \quad b_- - e^{\alpha_-} = \frac{\alpha_- e^{\alpha_-}}{\mu - \alpha_-},$$

it follows that the coefficient to $e^{\alpha_+ x}$ in (3.20) is

$$\begin{aligned}
\frac{1}{b_- - b_+}\left(b_- - \frac{b_- e^{\alpha_+} - b_+ e^{\alpha_-}}{b_- - b_+ + e^{\alpha_+} - e^{\alpha_-}}\right) &= \frac{b_- - e^{\alpha_-}}{b_- - b_+ + e^{\alpha_+} - e^{\alpha_-}} \\
&= \frac{(\mu - \alpha_+)\alpha_- e^{\alpha_-}}{(\mu - \alpha_+)\alpha_- e^{\alpha_-} - (\mu - \alpha_-)\alpha_+ e^{\alpha_+}} \\
&= \frac{\mu e^{\alpha_-}(\theta + \alpha_-)}{\mu e^{\alpha_-}(\theta + \alpha_-) - \mu e^{\alpha_+}(\theta + \alpha_+)} \\
&= \frac{a_-}{a_- - a_+} = C_+.
\end{aligned}$$

Similarly, the coefficient to $e^{\alpha_- x}$ can be verified to be C_-, completing the check.

Remark 3.1

For the $M/M/1$ queue with full accessibility, we can use the same argument with L replaced by B. Here, however, optional stopping is only allowed with $\alpha = \alpha_-$ (cf. [2, p. 267]), and we get

$$e^{\alpha_- x} = \mathbb{E}_x e^{\alpha_- \tilde{V}(B) - \theta B} = \mathbb{E}_x e^{-\theta B},$$

proving (3.4). □

For earlier work on martingales methods in a $M/G/1$ setting, see in particular Rosenkranzt [25] and Baccelli and Makowski [6].

3.3. $\mathbb{E}B$ AND $\mathbb{E}B^2$ VIA MARTINGALES – MODELS I, II

In addition to (3.14), we shall use the further related martingales

$$\{\tilde{V}(t) - (\rho - 1)t\}_{t \geq 0}, \tag{3.21}$$

$$\{[\tilde{V}(t) - t\kappa'(\alpha)]e^{\alpha \tilde{V}(t) - t\kappa(\alpha)}\}_{t \geq 0}, \tag{3.22}$$

$$\{[\tilde{V}(t) - (\rho - 1)t]^2 - t\sigma^2\}_{t \geq 0}, \tag{3.23}$$

where $\rho = \lambda/\mu$ and $\sigma^2 = 2\lambda/\mu^2$. Note that (3.22) is the martingale obtained by differentiating (3.14) w.r.t. α; (3.21) is then obtained by letting $\alpha = 0$. Similarly, (3.23) is the derivative of (3.22) at $\alpha = 0$. See Neveu [23] for analogous discussion in discrete time and for verification of the conditions for the optional stopping schemes used below.

LEMMA 3.1

Let $\gamma = \mu - \lambda$, $p(x) = \mathbb{P}_x(L < B) = \mathbb{P}_x(\tilde{V}(L) > 1)$. Then

$$p(x) = \frac{\lambda(e^{\gamma x} - 1)}{\mu e^{\gamma} - \lambda}, \tag{3.24}$$

$$\mathbb{E}[L; L < B] = \frac{\lambda^2 x e^{\gamma x} - p(x)e^{\gamma}\mu(\lambda + 1) + \lambda\gamma\mathbb{E}_x L}{\lambda(\lambda - \mu e^{\gamma})}. \tag{3.25}$$

Proof

(3.24) is shown in Asmussen and Perry [3], but since the proof is short, we reproduce it here. Noting that $\kappa(\gamma) = 0$, optional stopping of (3.14) (at time L) with $\alpha = \gamma$ yields

$$e^{\gamma x} = 1 - p(x) + p(x)\mathbb{E}e^{\gamma(1+S)} = 1 - p(x) + p(x)\frac{\mu e^{\gamma}}{\lambda},$$

and (3.24) follows easily. From (3.22) and $\kappa'(\gamma) = \mu/\lambda - 1 = \gamma/\lambda$, we get similarly

$$x e^{\gamma x} = \mathbb{E}_x\Big[\tilde{V}(L) - L\frac{\gamma}{\lambda}\Big]e^{\gamma \tilde{V}(L)}.$$

Here

$$\begin{aligned}\mathbb{E}_x\tilde{V}(L)e^{\gamma\tilde{V}(L)} &= p(x)\mathbb{E}(1+S)e^{\gamma(1+S)} = p(x)e^{\gamma}\left(\frac{\mu}{\lambda}+\frac{\mu}{\lambda^2}\right),\\ \mathbb{E}_xLe^{\gamma\tilde{V}(L)} &= \mathbb{E}_x[L; L=B] + \mathbb{E}_x[Le^{\gamma(1+S)}; L<B]\\ &= \mathbb{E}_xL + \mathbb{E}_x[L; L<B]\{\mathbb{E}e^{\gamma(1+S)}-1\}\\ &= \mathbb{E}_xL + \mathbb{E}_x[L; L<B]\left\{\frac{\mu e^{\gamma}}{\lambda}-1\right\}.\end{aligned}$$

Performing some algebra yields (3.25). □

PROPOSITION 3.1

$$\mathbb{E}_xL = \frac{\mu x - p(x)(\mu+1)}{\gamma} \tag{3.26}$$

$$\mathbb{E}B_1 = \mathbb{E}_1B = \frac{\mathbb{E}_1L + p(1)/\mu}{1-p(1)}, \tag{3.27}$$

$$\mathbb{E}_xB = \mathbb{E}_xL + p(x)\left(\frac{1}{\mu}+\mathbb{E}B_1\right). \tag{3.28}$$

Proof

Optimal stopping of (3.21) yields

$$x = \mathbb{E}_x\tilde{V}(L) - (\rho-1)\mathbb{E}_xL = p(x)\left(1+\frac{1}{\mu}\right) - (\rho-1)\mathbb{E}_xL.$$

Solving for $\mathbb{E}_xL$ yields (3.26). The expression (3.28) follows by taking expectations in (3.16), and (3.27) then follows by letting $x=1$ in (3.28) and solving for $\mathbb{E}_1B$. □

It remains to calculate $\mathbb{E}_xB^2$. By (3.16),

$$\mathbb{E}_xB^2 = \mathbb{E}_xL^2 + p(x)\left\{\frac{2}{\mu^2}+\mathbb{E}_1B^2+2\mathbb{E}_1B/\mu\right\} + 2\mathbb{E}[L; L<B]\left\{\frac{1}{\mu}+E_1B\right\}. \tag{3.29}$$

If we can compute $\mathbb{E}_x L^2$, (3.29) with $x = 1$ then yields an expression for $E_1 B^2$, and everything is known for a general x. Thus, consider $\mathbb{E}_x L^2$. Optimal stopping of (3.23) yields

$$\sigma^2 \mathbb{E}_x L = \mathbb{E}_x(\tilde{V}(L) - (\rho - 1)L)^2 \tag{3.30}$$

$$= \mathbb{E}_x \tilde{V}(L)^2 + (\rho - 1)^2 \mathbb{E}_x L^2 + 2(1 - \rho)\mathbb{E}_x L \tilde{V}(L). \tag{3.31}$$

Here

$$\mathbb{E}_x \tilde{V}(L)^2 = p(x)\mathbb{E}_x(1 + S)^2 = p(x)\left(1 + \frac{2}{\mu} + \frac{2}{\mu^2}\right), \tag{3.32}$$

$$\mathbb{E}_x L \tilde{V}(L) = \mathbb{E}[L(1 + S); L < B] = \left(1 + \frac{1}{\mu}\right)\mathbb{E}_x[L; L < B]. \tag{3.33}$$

Thus, everything in (3.30) except for $\mathbb{E}_x L^2$ has been computed, and solving for $\mathbb{E}_x L^2$ completes the analysis.

The analysis of Model II follows the same pattern, with only minor changes. Formulas (3.24), (3.25), (3.26) hold without changes. Replacing (3.16) by (3.18), formulas (3.27), (3.28) become

$$\mathbb{E}_1 B = \frac{\mathbb{E}_1 L}{1 - p(1)}, \quad \mathbb{E}_x B = \mathbb{E}_x L + p(x)\mathbb{E}_1 B. \tag{3.34}$$

Also (3.32), (3.33) are unchanged, whereas (3.29) has to be replaced by

$$\mathbb{E}_x B^2 = \mathbb{E}_x L^2 + p(x)\mathbb{E}_1 B^2 + 2\mathbb{E}[L; L < B]\mathbb{E}_1 B. \tag{3.35}$$

4. Busy period analysis – exponential barricade

In this section we compute the LT $\beta(\theta) = \mathbb{E}e^{-\theta B}$ of the busy period for the special case $\bar{G}(x) = e^{-\mu x}$, $\bar{H}(x) = e^{-\eta x}$. In view of Proposition 1.1, it suffices to develop the theory for Model I; the formulas for Models II, III then follow by trivial parameter substitutions.

The form of the solution is in terms of an associated birth-death process $\{N(t)\}_{t \geq 0}$ with birth and death rates

$$\lambda_n = \lambda, \quad \mu_n = \mu + (n - 1)\eta \quad (n \geq 1). \tag{4.1}$$

If $\omega(n)$ is the first passage time of $\{N(t)\}$ from level $n \geq 1$ to level $n - 1$ and $\Gamma_n(\theta) = \mathbb{E}e^{-\theta\omega(n)}$, we show that $\beta(\theta) = \Gamma_1(\theta)$ by coupling $\{V(t)\}$ and $\{N(t)\}$ on a common probability space in such a way that $B = \omega(1)$.

In this section, $Y_1(t), Y_2(t), \ldots$ denote random variables which are independent for fixed t, with $Y_i(t)$ being exponential with rate μ_i, and with paths of $Y_i(t)$ which are right-continuous with left limits $Y_i(t-)$. The idea behind the coupling is to observe that when $N(t) = n$, the residual work $V_N(t)$ in $\{N(t)\}$ can obviously be represented as

$$Y_1(t) + \cdots + Y_n(t) \tag{4.2}$$

and next to prove (see (4.4)–(4.10) below) that $V(t)$ has a similar representation in distribution. A careful step-by-step construction then yields the coupling of sample paths (which it is needed to conclude that B and $\omega(1)$ have the same distribution).

THEOREM 4.1

There exist a coupling of $\{V(t)\}$ and $\{N(t)\}$ such that $V(t) = V_N(t)$, where $V_N(t)$ has the sample path representation (4.2) with the $Y_i(t)$ independent of $\sigma(N(v) : 0 \le v \le t)$.

Proof

We assume $\tau_1 = 0$, $N_0 = 1$, and shall construct $\{N(t)\}$ and the $Y_i(t)$ recursively on the intervals $[\tau_n, \tau_{n+1})$. To this end, we start from a representation

$$V(\tau_n) = \sum_{i=1}^{N(\tau_n)} Y_i(\tau_n). \tag{4.3}$$

For $n = 1, N(\tau_n) = 1$ and we can just take $Y_1(\tau_1)$ as the service time of the first customer.

Given (4.3) with $N(\tau_n) = m$ (say), let

$$\begin{aligned} I_0 &= [0, Y_m(\tau_n)), \\ I_1 &= [Y_m(\tau_n), Y_m(\tau_n) + Y_{m-1}(\tau_n)), \\ &\vdots \\ I_m &= [Y_m(\tau_n) + \cdots + Y_1(\tau_n), \infty). \end{aligned}$$

For $t \in I_k$, define $\tilde{N}(\tau_n + t) = m - k$.

$$\tilde{Y}_i(\tau_n + t) = \begin{cases} Y_i(\tau_n), & i < m - k, \\ Y_{m-k}(\tau_n) - (t - Y_m(\tau_n) - \cdots - Y_{m-k-1}(\tau_n)), & i = m - k. \end{cases}$$

Then the $\tilde{Y}_i$ have similar properties as the Y_i and are independent of $\sigma(N(v) : v \le \tau_n + t)$ (this is obvious for $i < m - k$ and for $i = m - k$, it follows by interpreting $Y_{m-k}(\tau_n + t)$ as an excess value of $Y_{m-k}(\tau_n)$, using the memoryless property of the exponential distribution). Now $\{\tilde{N}(\tau_n + t)\}_{t \ge 0}$ and the $\tilde{Y}_i$ describe the way in which $\{N(\tau_n + t)\}$ would have evolved if all arrivals after τ_n were cancelled, and thus we can just let

$$N(\tau_n + t) = \tilde{N}(\tau_n + t), \quad Y_i(\tau_n + t) = \tilde{Y}_i(\tau_n + t), \quad t < \tau_{n-1} - \tau_n.$$

To finish the construction, we need to show that we can obtain the representation (4.3) for $n + 1$.

If $N(\tau_{n+1}) = N(\tau_{n+1}-) + 1 = l$ (say), then clearly

$$V(\tau_{n+1}-) = \sum_{i=1}^{l-1} Y_i(\tau_{n+1}-), \tag{4.4}$$

with LT

$$\prod_{i=1}^{l-1} \frac{\mu_i}{\mu_i + \theta}. \tag{4.5}$$

Now

$$\mathbb{E}e^{-\theta S 1_{\{U > x\}}} = 1 - e^{-\eta x} + e^{-\eta x} \frac{\mu}{\mu + \theta} = 1 - \frac{\theta}{\mu + \theta} e^{-\eta x}, \tag{4.6}$$

$$\mathbb{E}e^{-\theta(x + S 1_{\{U > x\}})} = e^{-\theta x} - \frac{\theta}{\mu + \theta} e^{-(\eta + \theta)x}, \tag{4.7}$$

and thus on $\{N(\tau_{n+1}) = l\}$

$$\begin{aligned} \mathbb{E}[e^{-\theta V(\tau_{n+1})} \mid \mathscr{G}_n] &= \prod_{i=1}^{l-1} \frac{\mu_i}{\mu_i + \theta} - \frac{\theta}{\mu + \theta} \prod_{i=1}^{l-1} \frac{\mu_i}{\mu_i + \eta + \theta} & (4.8) \\ &= \prod_{i=1}^{l-1} \frac{\mu_i}{\mu_i + \theta} - \frac{\theta}{\mu_l + \theta} \prod_{i=1}^{l-1} \frac{\mu_i}{\mu_i + \theta} & (4.9) \\ &= \prod_{i=1}^{l} \frac{\mu_i}{\mu_i + \theta}, & (4.10) \end{aligned}$$

where

$$\mathscr{G}_n = \sigma(N(v) : v \le \tau_{n+1}; Y_1(t), Y_2(t), \ldots, 0 \le t < \tau_{n+1})$$

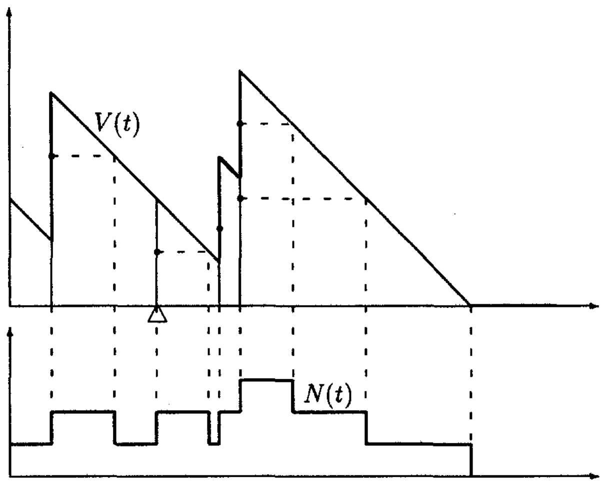

Fig. 1.

Thus, given $\mathscr{G}_n$, (4.3) holds in distribution, and we can choose r.v.'s $Y_i(\tau_{n+1})$ such that (4.3) holds a.s. □

The construction is illustrated in fig. 1. Here the ●'s on the vertical lines at an arrival instant τ_n are separators for the $Y_i(\tau_n)$, and the △ marks a rejected arrival.

COROLLARY 4.1

In the setting of Theorem 4.1, $B = \omega(1)$, $\beta(\theta) = \Gamma_1(\theta)$.

Proof

Obviously,

$$B = \inf\{t > 0 : V(t) = 0\} = \inf\{t > 0 : V_N(t) = 0\}. \qquad \square$$

For the computation of $\Gamma_1(\theta)$ we note the following equation for a general birth-death process, which follows immediately by conditioning upon the first transition out of state n:

$$\Gamma_n(\theta) = \frac{\mu_n}{\lambda_n + \mu_n + \theta} + \frac{\lambda_n}{\lambda_n + \mu_n + \theta}\Gamma_n(\theta)\Gamma_{n+1}(\theta). \tag{4.11}$$

EXAMPLE 4.1

For the standard $M/M/1$ queue, we have a coupling similar to Theorem 4.1, with $N(t)$ simply being the number of customers. Here $\lambda = \lambda_n$, $\mu = \mu_n$, $\Gamma(\theta) = \Gamma_n(\theta)$ do not depend on n, so that (4.11) takes the form

$$\beta(\theta) = \frac{\lambda}{\lambda + \mu + \theta} + \frac{\mu}{\lambda + \mu + \theta}\beta(\theta)^2$$

which immediately yields the second expression for $\beta(\theta)$ in (3.1). □

In the general case, expressing $\Gamma_2(\theta)$ in terms of $\Gamma_2(\theta)$, $\Gamma_2(\theta)$ in terms of $\Gamma_3(\theta)$ and so on via (4.11) leads to an expression of $\beta(\theta) = \Gamma_1(\theta)$ as a continued fraction, which does not appear too useful. Neither have any other approach we know of provided tractable expressions for $\Gamma_n(\theta)$. However, by truncation we can get an approximate solution via a finite system of recursive equations:

PROPOSITION 4.1

Consider a positive recurrent birth-death process. Let $\Gamma_n^{(M)}, n = 1, \cdots, M$, be the solution of (4.11) obtained by letting $\Gamma_M^{(M)}(\theta) = \mu_n/(\lambda_n + \mu_n + \theta)$ and solving for $\Gamma_{M-1}^{(M)}(\theta), \cdots, \Gamma_1^{(M)}(\theta)$. Then $\Gamma_n^{(M)}(\theta) \to \Gamma_n(\theta)$ as $M \to \infty$.

Proof

Obviously, $\Gamma_n^{(M)}(\theta) = \mathbb{E}e^{-\theta\omega_M(n)}$, where $\omega_M(n)$ refers to the truncated birth-death process on $\{0, 1, \ldots, M\}$ obtained by replacing λ_n by 0 for $n \geq M$. As $M \to \infty$,

$$P\left(\max_{0 \leq t \leq \omega(n)} N(t) > M\right) \to 0$$

and hence $\omega_M(n) \to \omega(n)$, $\Gamma_n^{(M)}(\theta) \to \Gamma_n(\theta)$. □

Rewriting (4.11) as

$$(\lambda_n + \mu_n + \theta)\Gamma_n = \mu_n + \lambda_n\Gamma_n\Gamma_{n+1}$$

and differentiating yields

$$\Gamma_n + (\lambda_n + \mu_n + \theta)\Gamma_n' = \lambda_n(\Gamma_n'\Gamma_{n+1} + \Gamma_n\Gamma_{n+1}'),$$
$$2\Gamma_n' + (\lambda_n + \mu_n + \theta)\Gamma_n'' = \lambda_n(\Gamma_n''\Gamma_{n+1} + 2\Gamma_n'\Gamma_{n+1}' + \Gamma_n\Gamma_{n+1}'').$$

Letting $m_n = \mathbb{E}\omega(n) = -\Gamma'_n(0)$, $m_n^{(2)} = \mathbb{E}\omega(n)^2 = \Gamma''_n(0)$, it follows for $\theta = 0$ that

$$\begin{aligned} 1-(\lambda_n+\mu_n)m_n &= -\lambda_n m_n - \lambda_n m_{n+1}, \\ m_n &= \frac{1}{\mu_n}\{\lambda_n m_{n+1}+1\}, \qquad (4.12) \\ -2m_n+(\lambda_n+\mu_n)m_n^{(2)} &= \lambda_n m_n^{(2)} + 2\lambda_n m_n m_{n+1} + \lambda_n m_{n+1}^{(2)}, \\ m_n^{(2)} &= \frac{1}{\mu_n}\{\lambda_n m_{n+1}^{(2)} + 2\lambda_n m_n m_{n+1} + 2m_n\}. \qquad (4.13) \end{aligned}$$

In the same way as in Proposition 4.1, an approximate solution is obtained by solving backwards starting from $m_M = (\lambda_n + \mu_n)^{-1}, m_M^{(2)} = 2(\lambda_n + \mu_n)^{-2}$.

In fact, (4.12) leads to an explicit solution:

COROLLARY 4.2

$$\mathbb{E}B = m_1 = 1 + \frac{\lambda}{\mu} + 2\sum_{n=1}^{\infty} \lambda^n \prod_{i=1}^{n} \frac{1}{\mu + (i-1)\eta}.$$

Proof

Noting that the general solution of the recursion $m_k = c_k m_{k+1} + d_k$ is $m_k = \Sigma_{n=k}^{\infty} d_n \Pi_{i=k}^{n-1} c_i$, we obtain

$$m_1 = \sum_{n=1}^{\infty}\left(1 + \frac{\lambda}{\mu + (n-1)\eta}\right) \prod_{i=1}^{n-1} \frac{\lambda}{\mu + (i-1)\eta},$$

which is the same as the asserted expression. □

References

[1] L.G. Afaneseva, Existence of a limit distribution in queueing systems with bounded sojourn time, Th. Prob. Appl. 10 (1965) 515–522.

[2] S. Asmussen, *Applied Probability and Queues* (Wiley, Chichester New York, 1987).

[3] S. Asmussen and D. Perry, On cycle maxima, first passage problems and extreme value theory for queues, Stoch. Models 8 (1992) 421–458.

[4] S. Asmussen and M. Bladt, Phase-type distributions and risk processes with premiums dependent on the current reserve, to appear in Scand. Act. J.

[5] C.T.H. Baker, *The Numerical Solution of Integral Equations* (Clarendon Press, Oxford, 1977).

[6] F. Bacelli and A.M. Makowski, Dynamic, transient and stationary behavior of the M/G/1 queue via martingales, Ann. Prob. 17 (1989) 1691–1699.
[7] P.P. Bhattacharya and A. Eprhremides, Stochastic monotonicity of multiserver queues with impatient customers, J. Appl. Prob. 28 (1991) 673–682.
[8] P.P. Bhattacharya and A. Eprhemides, Optimal scheduling under strict deadlines, IEEE Trans. Autom. Control 34 (1989) 721–728.
[9] S. Browne and K. Sigman, Work-modulated queues with applications to storage processes, J. Appl. Prob. 29 (1992) 699–712.
[10] J.W. Cohen, Single server queue with restricted accessibility, J. Eng. Math. 3 (1969) 265–285.
[11] J.W. Cohen, *The Single Server Queue*, revised ed. (North-Holland, Amsterdam, 1982).
[12] D.J. Daley, Single server queueing system with uniformly limited queuing time, J. Aust. Math. Soc. 4 (1964) 489–505.
[13] Gavish and P.J. Schweitzer, The Markovian queue with bounded waiting time, Man. Sci. 23 (1977) 1349–1357.
[14] B.V. Gnedenko and I.N. Kovalenko, *Introduction to Queueing Theory*, 2nd Ed. (Birkhäuser, Basel, 1989).
[15] J.M. Harrison and S.I. Resnick, The stationary distribution and first exit probabilities of a storage process with general release rule, Math. Oper. Res. 1 (1976) 347–358.
[16] J. Keilson, A gambler's ruin type problem in queueing theory, Oper. Res. 11 (1963) 570–576.
[17] H. Kaspi and D. Perry, Inventory systems of perishable commodities, Adv. Appl. Prob. 15 (1983) 674–685.
[18] H. Kaspi and D. Perry, Inventory systems of perishable commodities with renewal input and Poisson output, Adv. Appl. Prob. 16 (1984) 402–421.
[19] C. Knessl, B.J. Matowsky, Z. Schuss and C. Tier, Busy period distribution on state-dependent queues, Queing Systems 2 (1987) 285–305.
[20] M. Loeve, *Probability Theory*, 3rd ed. (Van Nostrand, Princeton, 1963).
[21] M. Miyazawa, The intensity conservation law for queues with randomly changed service rates, J. Appl. Prob. 22 (1985) 408–418.
[22] M.F. Neuts, *Matrix-Geometric Solutions in Stochastic Models* (Johns Hopkins University Press, Baltimore. London, 1981).
[23] J. Neveu, *Martingales á Temps Discret* (Masson, Paris, 1972).
[24] D. Perry and M.J.M. Posner, Analysis of production/inventory systems with several production rates, Stoch. Models 6 (1990) 99–116.
[25] W.A. Rosenkrantz, Calculation of the Laplace transform of the length of the busy period for the M/GI/1 queue via martingales, Ann. Prob. 11 (1989) 817–818.
[26] S. Schock Petersen, Calculation of ruin probabilities when the premium depends on the current reserve, Scand. Act. J. (1988).
[27] D. Sonderman, Comparing multi-server queues with finite waiting rooms, I, II, Appl. Prob. 11 (1979) 439–455.
[28] D. Stoyan, *Comparison Method for Queues and Other Stochastics Models* (Wiley, 1983).
[29] W. Whitt, Comparing counting processes and queues, Adv. Appl. Prob. 13 (1981) 207–220.
[30] R.W. Wolff, An upper bound for multi-channel queues, J. Appl. Prob. 14 (1977) 884–888.