axioms

MDPI

Article

Optimal Control Problems for Erlang Loss Systems

Mario Lefebvre

Department of Mathematics and Industrial Engineering, Polytechnique Montréal, P.O. Box 6079, Station Centre-Ville, Montréal, QC H3C 3A7, Canada; mlefebvre@polymtl.ca

Abstract: An Erlang loss system, which is an $M/G/s/s$ queue, is a model used in various applications. In this paper, a controlled version of the process is defined. The objective is to maximize the expected time until the system is full when the service time is exponentially distributed. The control variable is the service rate. The dynamic programming equation satisfied by the value function F, from which the optimal control follows at once, is derived, and F is found explicitly when $s = 2$ and $s = 3$. The problem of minimising the probability of the system being saturated is also considered.

Keywords: $M/G/s/s$ model; dynamic programming; first-passage time; homing problem; difference equations

MSC: 93E20; 60K30

1. Introduction

In an $M/G/s/s$ queue, called an *Erlang loss system*, customers arrive according to a Poisson process of rate λ. The service time S can follow any distribution. There are s servers, and no room for potential customers to wait for service because the capacity of the system is also equal to s. As a result, if a customer arrives while the system is in state s, this customer is lost.

This model is used in various applications. In Smith et al. [1], it was applied to traffic distribution in cellular networks. Lee and Dudin [2] analysed the sojourn time and loss probability in a model of traffic control in communication networks.

Restrepo et al. [3] wanted to find out how to allocate a fleet of ambulances in order to keep response times to calls as small as possible. van Dijk and Kortbeek [4] computed bounds for the probability of rejection in operating theatre and intensive care unit systems.

In de Bruin et al. [5], the problem of interest was to determine the size of hospital wards. Hyytiä et al. [6] proposed a model related to an Erlang loss system to describe a single vehicle in a dial-a-ride transport system. They presented closed-form expressions for the blocking probability, the acceptance rate, and the mean sojourn time of the vehicle in the system.

Rastpour et al. [7] worked on the duration of red alerts for ambulance systems, which corresponds to times at which all ambulances are busy. Morozov et al. [8] were interested in the stability of a model for cognitive wireless networks and used a modified Erlang loss system. Finally, in Zimmerman et al. [9], the aim was to effectively manage the distribution of ventilators during the COVID-19 pandemic.

Recently, Lefebvre and Yaghoubi studied optimal control problems for queueing systems. In [10], they assumed that in the $M/G/1$ model, the server can choose between two service time distributions. The aim was to minimize the expected time needed to clear the queue of customers. In [11], there were k servers, and the objective was to determine

Academic Editors: Giovanni Fusco and Valery Y. Glizer

Received: 5 January 2025
Revised: 8 February 2025
Accepted: 10 February 2025
Published: 11 February 2025

Citation: Lefebvre, M. Optimal Control Problems for Erlang Loss Systems. *Axioms* **2025**, *14*, 130. https://doi.org/10.3390/axioms14020130

how many servers should be used at any time to reduce the number of customers from $k+l$ to $k+r$, where $0 \leq r < l$, as rapidly as possible. In both cases, the optimizer had to take quadratic control costs into account.

In this paper, we first suppose that the service time S has an exponential distribution with the parameter μ, so that the Erlang loss system becomes an $M/M/s/s$ queue. However, we assume that the parameter μ is actually a function, $\mu[X(t)]$, of the number of customers in the system at time t and that the server can choose any (positive) service rate. The aim will be to determine the value of $\mu[X(t)]$ that maximizes the expected value of the time $\tau(k)$ until the system is full, where k is the initial number of customers in the system. Then, we will consider the problem of minimising the probability that the system, when in equilibrium, is saturated in the interval $[0, \tau(k))$. Finally, we will suppose that the parameter μ is a random variable R and compute the probability that the system (when in equilibrium) is saturated for various important distributions of R.

2. Optimal Service Rate

Let $X(t)$ denote the number of customers at time t in an $M/M/s/s$ queue. We assume that $X(0) = k \in \{0, 1, 2, \ldots, s-1\}$. We define the random variable as follows:

$$\tau(k) = \inf\{t > 0 : X(t) = s \mid X(0) = k\}. \tag{1}$$

In probability theory, $\tau(k)$ is called the *first-passage time*. It is the first time that the system is full, given that there were k customers at the initial time 0. All the servers are assumed to serve at the same rate $\mu[X(t)]$. The larger $\mu[X(t)]$ is, the longer it will take until $X(t) = s$. However, we take it for granted that serving faster is more costly for the system. More precisely, we consider the cost functional:

$$J(k) := \int_0^{\tau(k)} \left\{ \frac{1}{2} q_0 \mu^2[X(t)] + \theta \right\} \mathrm{d}t + K[\tau(k)], \tag{2}$$

where $q_0 > 0$ and $\theta < 0$ are constants and $K(\cdot)$ is a terminal cost function. Since θ is negative, a reward is received as long as the system is not full. There must therefore be a trade-off between working faster (so that $\mu[X(t)]$ is larger) and the quadratic costs thus incurred. The optimizer must also take into account the final cost $K[\tau(k)]$. In this paper, we assume that $K[\tau(k)]$ is a constant K_0.

When the service time S has an exponential distribution, we can use dynamic programming to determine the value of $\mu[X(t)]$ in the interval $[0, \tau(k))$ that minimizes the expected value of the cost criterion $J(k)$.

Many papers have been published on the optimal control of queueing systems; see, for instance, Chen and Xia [12], Su and Li [13], and Chen et al. [14]. The main difference between these papers and the present one is the fact that in our case, the final time is a random variable. Such problems are known as *homing problems* (see Whittle [15]). In a homing problem, the optimizer controls a stochastic process until a certain event occurs. In many applications, the final time is indeed neither fixed nor infinite, but rather random. For example, suppose that $X(t)$ denotes the number of people suffering from a certain disease during an epidemic, and that we want to put an end to this epidemic as quickly as possible. In reality, the duration of the epidemic is obviously a random variable. Similarly, in the problem that we are studying, we cannot know how long it will take until the system is saturated.

Homing problems were first considered by Whittle [15] as n-dimensional diffusion processes. In practice, it is very difficult to find explicit and exact solutions to these problems in two or more dimensions, as this usually requires solving non-linear partial differential

equations. Whittle [16] also generalised homing problems by considering a cost functional that takes into account the optimizer's sensitivity to risk.

In Rishel [17], a homing problem for diffusion processes that can model the wear of devices was created. In addition to their papers on the optimal control of queueing systems, the author has solved homing problems for autoregressive processes ([18]) and in discrete times ([19]). In [19], the problem considered was in fact a stochastic dynamic game.

Now, we define the *value function*:

$$F(k) = \inf_{\substack{\mu[X(t)] \\ t \in [0, \tau(k))}} E[J(k)], \tag{3}$$

where "E" denotes the *expected value*. This function gives us the expected cost (or reward) obtained if the optimizer chooses the optimal service rate in the interval $[0, \tau(k))$. Moreover, it satisfies the boundary condition

$$F(s) = K_0. \tag{4}$$

Remark 1. *Contrary to the problem in [10], here the problem that we study is time-invariant. Therefore, the value function depends only on the current state k. In addition, we now assume that the optimizer can choose any positive value of $\mu[X(t)]$. In [10], there were only two possible values for the service rate.*

In an $M/M/s/s$ queue, the random time A needed for a new customer to arrive is exponentially distributed with parameter λ. It follows that the probability that a customer will arrive in the interval $(0, \Delta t]$ is given by

$$P[A \leq \Delta t] = 1 - e^{-\lambda \Delta t} = \lambda \Delta t + o(\Delta t). \tag{5}$$

When there are $k \in \{1, 2, \ldots, s-1\}$ customers in the system, the time $S(k)$ until a customer departs has an exponential distribution with the parameter $k\mu(k)$. Hence,

$$P[S(k) \leq \Delta t] = k\mu(k)\,\Delta t + o(\Delta t). \tag{6}$$

The probability of at least two events (arrivals or departures) in any interval of length Δt is equal to $o(\Delta t)$. Moreover, the random variables A and $S(k)$ are assumed to be independent, which implies that $\min\{A, S(k)\} \sim \text{Exp}(\lambda + k\mu(k))$.

When $X(0) = k \in \{1, 2, \ldots, s-1\}$, the random number of customers in the system at time Δt is

$$X(\Delta t) = \begin{cases} k+1 & \text{with probability } \lambda \Delta t + o(\Delta t), \\ k-1 & \text{with probability } k\mu(k)\,\Delta t + o(\Delta t), \\ k & \text{with probability } 1 - [\lambda + k\mu(k)]\,\Delta t + o(\Delta t). \end{cases} \tag{7}$$

If $X(0) = 0$, we find that

$$X(\Delta t) = \begin{cases} k+1 & \text{with probability } \lambda \Delta t + o(\Delta t), \\ k & \text{with probability } 1 - \lambda \Delta t + o(\Delta t). \end{cases} \tag{8}$$

Next, since $\tau(0)$ has the same distribution as $A + \tau(1)$, we can write that

$$E[\tau(0)] = \frac{1}{\lambda} + E[\tau(1)] \quad \Longrightarrow \quad F(0) = \frac{\theta}{\lambda} + F(1). \tag{9}$$

For $k \in \{1, 2, \ldots, s-1\}$, we find that

$$\begin{aligned} F(k) &:= \inf_{\substack{\mu[X(t)] \\ t \in [0, \tau(k))}} E\left[\int_0^{\tau(k)} \left\{\frac{1}{2} q_0 \mu^2[X(t)] + \theta\right\} \mathrm{d}t\right] \\ &= \inf_{\substack{\mu[X(t)] \\ t \in [0, \tau(k))}} E\left[\int_0^{\Delta t} \left\{\frac{1}{2} q_0 \mu^2[X(t)] + \theta\right\} \mathrm{d}t + \int_{\Delta t}^{\tau(k)} \left\{\frac{1}{2} q_0 \mu^2[X(t)] + \theta\right\} \mathrm{d}t\right] \\ &= \inf_{\substack{\mu[X(t)] \\ t \in [0, \tau(k))}} E\left[\left(\frac{1}{2} q_0 \mu^2(k) + \theta\right) \Delta t + \int_{\Delta t}^{\tau(k)} \left\{\frac{1}{2} q_0 \mu^2[X(t)] + \theta\right\} \mathrm{d}t \right. \\ &\qquad \left. + o(\Delta t)\right]. \end{aligned} \tag{10}$$

Making use of Bellman's principle of optimality, we can express the expected value of the integral from Δt to $\tau(k)$ in terms of the value function at time Δt:

$$\inf_{\substack{\mu[X(t)] \\ t \in [\Delta t, \tau(k))}} E\left[\int_{\Delta t}^{\tau(k)} \left\{\frac{1}{2} q_0 \mu^2[X(t)] + \theta\right\} \mathrm{d}t\right] = E[F[X(\Delta t)]] + o(\Delta t). \tag{11}$$

We deduce from Equation (7) that

$$E[F[X(\Delta t)]] = F(k+1)\lambda \Delta t + F(k-1) k \mu(k) \Delta t + F(k)\{1 - [\lambda + k\mu(k)]\Delta t\} + o(\Delta t). \tag{12}$$

It then follows from Equation (10) that

$$\begin{aligned} 0 &= \inf_{\mu(k)} \left\{\left(\frac{1}{2} q_0 \mu^2(k) + \theta\right) \Delta t + F(k+1)\lambda \Delta t + F(k-1) k \mu(k) \Delta t - F(k)[\lambda + k\mu(k)]\Delta t \right. \\ &\qquad \left. + o(\Delta t)\right\}. \end{aligned} \tag{13}$$

That is,

$$0 = \inf_{\mu(k)} \left\{\frac{1}{2} q_0 \mu^2(k) + \theta + F(k+1)\lambda + F(k-1) k \mu(k) - F(k)[\lambda + k\mu(k)] + \frac{o(\Delta t)}{\Delta t}\right\}. \tag{14}$$

Letting Δt decrease to zero, we can state the following proposition.

Proposition 1. *The value function $F(k)$ satisfies the dynamic programming equation (DPE)*

$$0 = \inf_{\mu(k)} \left\{\frac{1}{2} q_0 \mu^2(k) + \theta + F(k+1)\lambda + F(k-1) k \mu(k) - F(k)[\lambda + k\mu(k)]\right\} \tag{15}$$

for $k \in \{1, 2, \ldots, s-1\}$. Moreover, we obtain Equation (9) and the boundary condition in Equation (4).

Now, by differentiating the expression within the curly brackets in Equation (15) with respect to $\mu(k)$ and setting it to zero, we find that

$$q_0 \mu(k) + k F(k-1) - k F(k) = 0. \tag{16}$$

Corollary 1. *The optimal control can be expressed in terms of the value function as follows:*

$$\mu^*(k) = \frac{k}{q_0}\left[F(k) - F(k-1)\right] \tag{17}$$

for $k \in \{1, 2, \ldots, s-1\}$.

Substituting the expression for $u^*(k)$ in Equation (17) into the DPE, we find (after simplification) that to obtain the value function $F(k)$ we must solve the non-linear second-order difference equation

$$0 = \theta - \frac{k^2}{2q_0}\left[F(k) - F(k-1)\right]^2 + \lambda\left[F(k+1) - F(k)\right], \tag{18}$$

where $k \in \{1, 2, \ldots, s-1\}$, subject to the boundary condition $F(s) = K_0$. In the remainder of this paper, we will assume that $K_0 = 0$, so that $F(s) = 0$.

Remark 2. *Suppose that we set* $\mu(k) = 0$ *for any k. That is, we consider the uncontrolled process* $\{X_0(t), t \geq 0\}$. *Then, we find that*

$$J_0(k) := \int_0^{\tau_0(k)} \theta \, \mathrm{d}t \quad \Longrightarrow \quad E[J_0(k)] = \theta\, E[\tau_0(k)]. \tag{19}$$

Moreover, we can write that

$$\tau_0(k) = A_k + A_{k+1} + \cdots + A_{s-1}, \tag{20}$$

where $A_i \sim Exp(\lambda)$ *for* $i = 1, 2, \ldots, s-1$. *Hence,*

$$E[\tau_0(k)] = \frac{s-k}{\lambda} \quad \Longrightarrow \quad F_0(k) = \frac{(s-k)\,\theta}{\lambda}. \tag{21}$$

It follows that the value function must satisfy the inequation

$$F(k) \leq \frac{(s-k)\,\theta}{\lambda} < 0. \tag{22}$$

2.1. The Case When $s = 2$

When $s = 2$, we find that $F(2) = 0$. Equation (18) for $k = 1$ becomes

$$0 = \theta - \frac{1}{2q_0}\left[F(1) - F(0)\right]^2 - \lambda F(1). \tag{23}$$

Moreover, Equation (9) implies that

$$-\lambda q_0 F(0) + 2\theta q_0 - \frac{\theta^2}{2\lambda^2} = 0, \tag{24}$$

so that

$$F(0) = \frac{2\theta}{\lambda} - \frac{\theta^2}{2\lambda^3 q_0}, \tag{25}$$

and $F(1) = F(0) - \theta/\lambda$. From Equation (17), we see that the optimal control is given by

$$\mu^*(1) = \frac{1}{q_0}\left[F(1) - F(0)\right] = -\frac{\theta}{q_0\lambda}. \tag{26}$$

2.2. The Case When $s = 3$

We now use $F(3) = 0$ and the following non-linear difference equations:

$$0 = \theta - \frac{1}{2q_0}[F(1) - F(0)]^2 + \lambda[F(2) - F(1)], \tag{27}$$

$$0 = \theta - \frac{2}{q_0}[F(2) - F(1)]^2 - \lambda F(2). \tag{28}$$

Making use of the mathematical software program *Mathematica* (https://www.wolfram.com/mathematica/online/), we find that the solution to systems (27) and (28), together with Equation (9), is

$$F(0) = \frac{\theta(6\lambda^6 q_0^3 - 5\lambda^4 q_0^2\theta + 4\lambda^2 q_0\theta^2 - \theta^3)}{2\lambda^7 q_0^3}, \tag{29}$$

$$F(1) = \frac{\theta(4\lambda^6 q_0^3 - 5\lambda^4 q_0^2\theta + 4\lambda^2 q_0\theta^2 - \theta^3)}{2\lambda^7 q_0^3}, \tag{30}$$

$$F(2) = \frac{\theta(2\lambda^6 q_0^3 - 4\lambda^4 q_0^2\theta + 4\lambda^2 q_0\theta^2 - \theta^3)}{2\lambda^7 q_0^3}. \tag{31}$$

In the particular case where $\theta = -1$ and $\lambda = 1$, we obtain

$$F(0) = -\frac{6q_0^3 + 5q_0^2 + 4q_0 + 1}{2q_0^3}. \tag{32}$$

This function of q_0 is shown in Figure 1. Notice that

$$\lim_{q_0 \to \infty} F(0) = -3, \tag{33}$$

which is the bound deduced from Equation (22). This result follows from the fact that when q_0 tends to infinity, the optimizer should choose $\mu(k) = 0$ for any $k \leq s - 1$. Similarly, with $\theta = -1$ and $\lambda = 1$, we obtain $\lim_{q_0 \to \infty} F(1) = -2$ and $\lim_{q_0 \to \infty} F(2) = -1$.

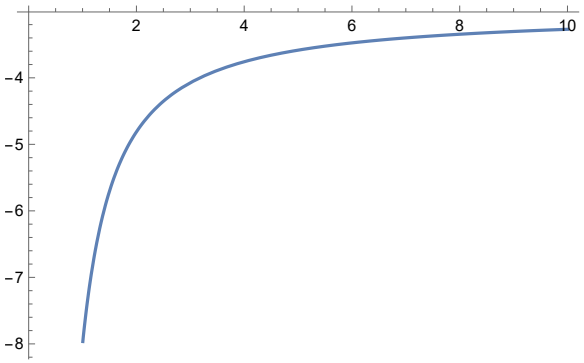

Figure 1. Value function $F(0)$ for $q_0 \in [0, 10]$ when $s = 3$; $\theta = -1$; and $\lambda = 1$.

The optimal controls are

$$\mu^*(1) = -\frac{\theta}{q_0\lambda}, \tag{34}$$

$$\mu^*(2) = -\frac{2\theta}{q_0\lambda} + \frac{\theta^2}{q_0^2\lambda^3}. \tag{35}$$

3. Minimising the Probability of Saturation

When the Erlang loss system is in statistical equilibrium (or a steady state), the probability that it is in state k is given by the following (see [20]):

$$\pi_k := \lim_{t\to\infty} P[X(t) = k] = \frac{(\lambda\, E[S])^k/k!}{\sum_{j=0}^{s}(\lambda\, E[S])^j/j!} \tag{36}$$

for $k \in \{0, 1, \ldots, s\}$. Thus, the probability that the system in equilibrium is saturated (or full), such that all the servers are busy, is

$$\pi_s = \frac{(\lambda\, E[S])^s/s!}{\sum_{j=0}^{s}(\lambda\, E[S])^j/j!}. \tag{37}$$

Suppose now that we start controlling the process when it is in equilibrium and that $X(0) = k \in \{0, 1, \ldots, s-1\}$. Moreover, we assume, as in the preceding section, that the service time is an exponential random variable with parameter μ that depends on the state of the process, such that $\pi_s = \pi_s[\mu(X(t))]$. We want to minimize the probability of saturation in the interval $[0, \tau(k))$. We consider the cost functional

$$J_2(k) := \int_0^{\tau(k)} \left\{ \frac{1}{2}\, q_0\, \mu^2[X(t)] + \pi_s[\mu(X(t))] \right\} \mathrm{d}t. \tag{38}$$

Proceeding as in Section 2, we obtain the following corollary to Proposition 1.

Corollary 2. *The value function $F(k)$ satisfies the DPE*

$$0 = \inf_{\mu(k)} \left\{ \frac{1}{2}\, q_0\, \mu^2(k) + \frac{[\lambda/\mu(k)]^s/s!}{\sum_{j=0}^{s}[\lambda/\mu(k)]^j/j!} + F(k+1)\,\lambda + F(k-1)\,k\,\mu(k) - F(k)\,[\lambda + k\,\mu(k)] \right\} \tag{39}$$

for $k \in \{1, 2, \ldots, s-1\}$. The boundary condition is $F(s) = 0$.

Even in the case where $s = 2$, trying to express the optimal control $\mu^*(k)$ in terms of the value function does not seem possible. Therefore, we will simplify the problem. Assume that the optimizer can only choose between $\mu = \mu_1$ or μ_2 and cannot switch from μ_1 to μ_2 (or vice versa) in the interval $[0, \tau(k))$. Then,

$$E[J_2(k)] = E\left[\int_0^{\tau(k)} \left\{ \frac{1}{2}\, q_0\, \mu_i^2 + \pi_s(\mu_i) \right\} \mathrm{d}t \right] = \left\{ \frac{1}{2}\, q_0\, \mu_i^2 + \pi_s(\mu_i) \right\} E[\tau(k)], \tag{40}$$

where $\mu_i = \mu_1$ or μ_2.

3.1. The Case When $s = 2$

Let $F_{i,j}$ be the event: the first transition of the process from state i is to state j. Using the fact that

$$P[A < S] = \frac{\lambda}{\mu_i + \lambda}, \tag{41}$$

we can write that

$$\begin{aligned} E[\tau(1)] &= E[\tau(1) \mid F_{1,0}]\, P[F_{1,0}] + E[\tau(1) \mid F_{1,2}]\, P[F_{1,2}] \\ &= \left\{ \frac{1}{\mu_i + \lambda} + \frac{1}{\lambda} + E[\tau(1)] \right\} \frac{\mu_i}{\mu_i + \lambda} + \frac{1}{\mu_i + \lambda}\, \frac{\lambda}{\mu_i + \lambda} \end{aligned} \tag{42}$$

for $i = 1, 2$. It follows that

$$E[\tau(1)] = \frac{\mu_i + \lambda}{\lambda^2}. \tag{43}$$

Therefore, in order to determine whether $E[J_2(1)]$ is smaller when $\mu = \mu_1$ or μ_2, we must calculate

$$E[J_2(1)] = \left[\frac{1}{2}q_0\mu_i^2 + \frac{(\lambda/\mu_i)^2/2}{1+(\lambda/\mu_i)+\frac{1}{2}(\lambda/\mu_i)^2}\right]\frac{\mu_i+\lambda}{\lambda^2} \tag{44}$$

for $i = 1, 2$ and compare the two values of $E[J_2(1)]$.

Suppose that $\lambda = 1$, $\mu_1 = 1$ and $\mu_2 = 2$. We must compare

$$E[J_2(1)] \overset{\mu=1}{=} q_0 + \frac{2}{5} \quad \text{with} \quad E[J_2(1)] \overset{\mu=2}{=} 6q_0 + \frac{3}{13}. \tag{45}$$

The two straight lines cross at $q_0 = q^* := 11/325$. As can be seen in Figure 2, the optimal solution is to choose $\mu = 2$ for $q_0 \in [0, q^*]$ and $\mu = 1$ for $q_0 \geq q^*$.

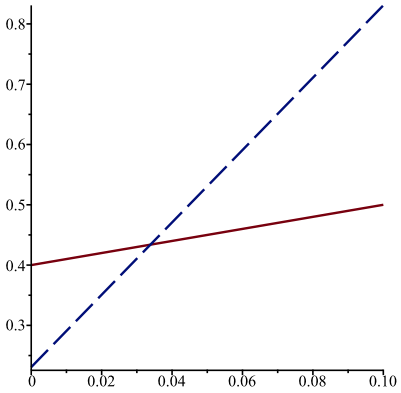

Figure 2. $E[J_2(1)]$ for $q_0 \in [0, 0.1]$ when $s = 2$, $\lambda = 1$, $\mu = 1$ (solid line), and $\mu = 2$ (dotted line).

3.2. The Case When $s = 3$

We know that

$$\begin{aligned} E[\tau(1)] &= E[\tau(1) \mid F_{1,0}]\,P[F_{1,0}] + E[\tau(1) \mid F_{1,2}]\,P[F_{1,2}] \\ &= \left\{\frac{1}{\mu_i+\lambda} + \frac{1}{\lambda} + E[\tau(1)]\right\}\frac{\mu_i}{\mu_i+\lambda} + \left\{\frac{1}{\mu_i+\lambda} + E[\tau(2)]\right\}\frac{\lambda}{\mu_i+\lambda} \end{aligned} \tag{46}$$

and

$$\begin{aligned} E[\tau(2)] &= E[\tau(2) \mid F_{2,1}]\,P[F_{2,1}] + E[\tau(2) \mid F_{2,3}]\,P[F_{2,3}] \\ &= \left\{\frac{1}{\mu_i+\lambda} + E[\tau(1)]\right\}\frac{\mu_i}{\mu_i+\lambda} + \frac{1}{\mu_i+\lambda}\frac{\lambda}{\mu_i+\lambda}. \end{aligned} \tag{47}$$

We find that the solution to systems (46) and (47) is

$$E[\tau(1)] = \frac{1}{\lambda^3}(2\lambda^2 + 2\lambda\mu + \mu^2) \quad \text{and} \quad E[\tau(2)] = \frac{1}{\lambda^3}(\lambda^2 + \lambda\mu + \mu^2). \tag{48}$$

With these values, we can proceed as in the case when $s = 2$ to determine whether $\mu = \mu_1$ or $\mu = \mu_2$ is the optimal choice for $k = 1, 2$. First, when $k = 1$, we compute

$$E[J_2(1)] = \left[\frac{1}{2}q_0\mu_i^2 + \frac{(\lambda/\mu_i)^3/6}{1+(\lambda/\mu_i)+\frac{1}{2}(\lambda/\mu_i)^2+\frac{1}{6}(\lambda/\mu_i)^3}\right]\frac{1}{\lambda^3}(2\lambda^2 + 2\lambda\mu_i + \mu_i^2) \tag{49}$$

for $i = 1, 2$. Let us again take $\lambda = 1$, $\mu_1 = 1$, and $\mu_2 = 2$. We obtain

$$E[J_2(1)] \overset{\mu=1}{=} \frac{5}{2} q_0 + \frac{5}{16} \quad \text{and} \quad E[J_2(1)] \overset{\mu=2}{=} 20 q_0 + \frac{10}{79}. \tag{50}$$

The optimal service rate is $\mu = 2$ for $q_0 \in [0, 47/4424]$ and $\mu = 1$ for $q_0 \geq 47/4424$; see Figure 3.

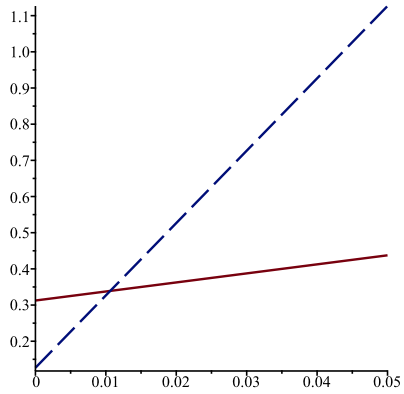

Figure 3. $E[J_2(1)]$ for $q_0 \in [0, 0.05]$ when $s = 3$, $\lambda = 1$, $\mu = 1$ (solid line), and $\mu = 2$ (dotted line).

Similarly, when $k = 2$, we find that

$$E[J_2(2)] \overset{\mu=1}{=} \frac{3}{2} q_0 + \frac{3}{16} \quad \text{and} \quad E[J_2(2)] \overset{\mu=2}{=} 14 q_0 + \frac{7}{79}. \tag{51}$$

This time, the optimal service rate is $\mu = 2$ for $q_0 \in [0, 5/632]$ and $\mu = 1$ for $q_0 \geq 5/632$; see Figure 4.

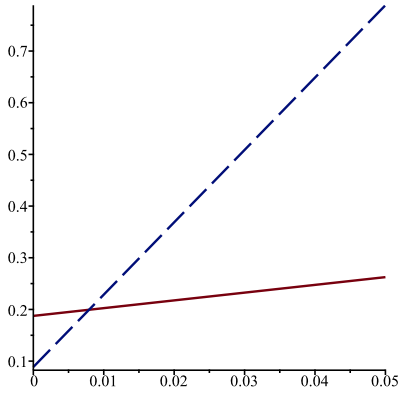

Figure 4. $E[J_2(2)]$ for $q_0 \in [0, 0.05]$ when $s = 3$, $\lambda = 1$, $\mu = 1$ (solid line), and $\mu = 2$ (dotted line).

4. The Case in Which the Service Rate Is a Random Variable

To conclude this paper, we will consider the following problem: suppose that the service time S has an exponential distribution with parameter R, where R is itself a random variable. We would like to determine for which distribution of R, as it has a given expected value, the probability of saturation is minimized when the system is in equilibrium. To do so, assuming that R possesses a probability density function $f_R(r)$, we would have to minimize

$$\pi_s = \int_0^\infty \frac{(\lambda/r)^s/s!}{\sum_{j=0}^{s} (\lambda/r)^j/j!} f_R(r)\, \mathrm{d}r, \tag{52}$$

under the constraints that

$$\int_0^\infty f_R(r)\,\mathrm{d}r = 1 \quad \text{and} \quad \int_0^\infty r f_R(r)\,\mathrm{d}r = r_0, \tag{53}$$

where r_0 is a given constant. The random variable R could be discrete, or even of mixed type, so that the density function $f_R(r)$ may contain Dirac delta functions.

Instead of trying to find the function $f_R(r)$ that minimizes π_s, we will consider the most important probability distributions for R. We will treat this as a case in which $s = 2$ and, for the sake of simplicity, set λ equal to 1. Moreover, we assume that $E[R] = 1$. Hence, we must compute

$$\pi_2 = \int_0^\infty \frac{(1/r)^2/2}{1 + (1/r) + \frac{1}{2}(1/r)^2} f_R(r)\,\mathrm{d}r \tag{54}$$

for random variables such that

$$\int_0^\infty r f_R(r)\,\mathrm{d}r = 1. \tag{55}$$

Suppose first that R is actually the constant 1. Then, we find that

$$f_R(r) = \delta(r-1) \quad \Longrightarrow \quad \pi_2 = \frac{1/2}{1+1+\frac{1}{2}} = \frac{1}{5}. \tag{56}$$

Next, if R has a uniform distribution at the interval $(0,2)$, we find that

$$\pi_2 = \frac{1}{2}\int_0^2 \frac{(1/r)^2/2}{1 + (1/r) + \frac{1}{2}(1/r)^2}\,\mathrm{d}r = \frac{\arctan(5)}{2} - \frac{\pi}{8} \approx 0.2940. \tag{57}$$

Assume now that S has a gamma distribution with both parameters equal to α, so that

$$f_R(r;\alpha) = \frac{1}{\Gamma(\alpha)}(\alpha r)^{\alpha-1}\alpha e^{-\alpha r} \quad \text{for } r > 0. \tag{58}$$

We obtain $E[R] = 1$, as required. If $\alpha = 1$, $R \sim \mathrm{Exp}(1)$. We compute

$$\int_0^\infty \frac{(1/r)^2/2}{1 + (1/r) + \frac{1}{2}(1/r)^2} f_R(r;\alpha = 0.1)\mathrm{d}r \approx 0.7474. \tag{59}$$

The value of π_2 when $\alpha = 1, 2, \ldots, 20$ is shown in Figure 5. We see that π_2 decreases with an increasing α. For large α, the limiting probability π_2 is approximately equal to 0.21.

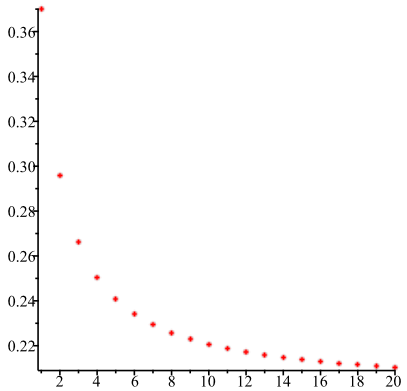

Figure 5. Value of the probability π_2 of saturation when R has the gamma distribution defined in Equation (58), for $\alpha = 1, 2, \ldots, 20$.

Let

$$f_R(r;\sigma) := \frac{r}{\sigma^2} e^{-r^2/(2\sigma^2)} \quad \text{for } r > 0. \tag{60}$$

That is, the random variable R has a Rayleigh distribution with parameter σ. When $\sigma = \sqrt{2/\pi}$, we find that $E[R] = 1$. We calculate

$$\int_0^\infty \frac{(1/r)^2/2}{1 + (1/r) + \frac{1}{2}(1/r)^2} f_R(r;\sigma = \sqrt{2/\pi})\mathrm{d}r \approx 0.2619. \tag{61}$$

Notice that for all the distributions considered above, the value of π_2 is larger than when $R = 1$.

If R is a discrete random variable taking the values 0.5 and 1.5 with probability of $1/2$, we find that $\pi_2 \approx 0.2588$. When $R = r_1$ or $2 - r_1$ with a probability of $1/2$, where $r_1 \in (0,1)$, $\pi_2 \in (1/5, 7/13)$ and decreases as r_1 increases. More generally, if $R = r_1 < 1$ or $r_2 > 1$, such that $E[R] = 1$, the value of π_2 is greater than $1/5$.

5. Conclusions

In this paper, a homing problem for an $M/G/s/s$ queueing model was considered. As mentioned in Section 1, this model is used in many applications. We were interested in maximising the expected time until the system becomes saturated. We were able to determine the optimal value of the service rate when the service time S has an exponential distribution, such that we are dealing with an $M/M/s/s$ queue. In this case, we can appeal to dynamic programming to obtain the equation satisfied by the value function. This equation is a non-linear second-order difference equation. The optimal control can be deduced from the value function.

In Section 3, we examined a related optimal control problem. However, we had to simplify the problem because it is not possible to explicitly express the optimal control in terms of the value function.

In Section 4, we wanted to minimize the probability of saturation when the service rate μ is a random variable R. Various types of distribution of R were considered.

In general, when the service time is not exponentially distributed, it is not possible to make use of dynamic programming. Indeed, if the probability in Equation (6) is not proportional to Δt, theIS technique will fail. The same holds true if the random variable A does not have an exponential distribution. In those cases, another technique must be used for the $M/G/s/s$ queue.

Finally, we could modify the assumptions leading us to the $M/M/s/s$ model. For instance, the servers could serve at different rates and/or perhaps more than one customer at a time, or there could be batch arrivals, and so on.

Funding: This research was supported by the Natural Sciences and Engineering Research Council of Canada.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: No new data were created or analysed in this study.

Acknowledgments: The author wishes to thank the anonymous reviewers of this paper for their constructive comments.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Smith, P.J.; Sathyendran, A.; Murch, A.R. Analysis of traffic distribution in cellular networks. In Proceedings of the 1999 IEEE 49th Vehicular Technology Conference, Houston, TX, USA, 16–20 May 1999; Volume 3, pp. 2075–2079. [CrossRef]
2. Lee, M.H.; Dudin, S.A. An Erlang loss queue with time-phased batch arrivals as a model for traffic control in communication networks. *Math. Probl. Eng.* **2008**, *2008*, 814740. [CrossRef]
3. Restrepo, M.; Henderson, S.G.; Topaloglu, H. Erlang loss models for the static deployment of ambulances. *Health Care Manag. Sci.* **2009**, *12*, 67–79. [CrossRef] [PubMed]
4. van Dijk, N.M.; Kortbeek, N. Erlang loss bounds for OT–ICU systems. *Queueing Syst.* **2009**, *63*, 253–280. [CrossRef]
5. de Bruin, A.M.; Bekker, R.; van Zanten, L.; Koole, G.M. Dimensioning hospital wards using the Erlang loss model. *Ann. Oper. Res.* **2010**, *178*, 23–43. [CrossRef]
6. Hyytiä, E.; Aalto, S.; Penttinen, A.; Sulonen, R. A stochastic model for a vehicle in a dial-a-ride system. *Oper. Res. Lett* **2010**, *38*, 432–435. [CrossRef]
7. Rastpour, A.; Ingolfsson, A.; Kolfal, B. Modeling yellow and red alert durations for ambulance systems. *Prod. Oper. Manag.* **2020**, *29*, 1972–1991. [CrossRef]
8. Morozov, E.; Rogozin, S.; Nguyen, H.Q.; Phung-Duc, T. Modified Erlang loss system for cognitive wireless networks. *Mathematics* **2022**, *10*, 2101. [CrossRef]
9. Zimmerman, S.L.; Rutherford, A.R.; van der Waall, A.; Norena, M.; Dodek, P. A queuing model for ventilator capacity management during the COVID-19 pandemic. *Health Care Manag. Sci.* **2023**, *26*, 200–216. [CrossRef] [PubMed]
10. Lefebvre, M.; Yaghoubi, R. Optimal service time distribution for an $M/G/1$ queue. *Axioms* **2024**, *13*, 594. [CrossRef]
11. Lefebvre, M.; Yaghoubi, R. Optimal control of a queueing system. *Optimization* **2024**, 1–14. [CrossRef]
12. Chen, S.; Xia, L. Optimal control of admission prices and service rates in open queueing networks. *IFAC-PapersOnLine* **2017**, *50*, 928–933. [CrossRef]
13. Su, Y.; Li, J. Optimality of admission control in an $M/M/1/N$ queue with varying services. *Stoch. Model.* **2021**, *37*, 317–334. [CrossRef]
14. Chen, G.; Liu, Z.; Xia, L. Event-based optimization of service rate control in retrial queues. *J. Oper. Res. Soc.* **2023**, *74*, 979–991. [CrossRef]
15. Whittle, P. *Optimization over Time*; Wiley: Chichester, UK, 1982; Volume I.
16. Whittle, P. *Risk-Sensitive Optimal Control*; Wiley: Chichester, UK, 1990.
17. Rishel, R. Controlled wear process: Modeling optimal control. *IEEE Trans. Autom. Control* **1991**, *36*, 1100–1102. [CrossRef]
18. Lefebvre, M. The homing problem for autoregressive processes. *IMA J. Math. Control Inform.* **2022**, *39*, 322–344. [CrossRef]
19. Lefebvre, M. A discrete-time homing problem with two optimizers. *Games* **2023**, *14*, 68. [CrossRef]
20. Ross, S.M. *Introduction to Probability Models*, 13th ed.; Elsevier/Academic Press: Amsterdam, The Netherlands, 2023.