Annals of Operations Research (2023) 331:629–647
https://doi.org/10.1007/s10479-022-04807-z

ORIGINAL RESEARCH

The impact of reneging on a fluid on-off queue with strategic customers

Dimitrios Logothetis[1] · **Athanasia Manou**[1] · **Antonis Economou**[1]

Accepted: 23 May 2022 / Published online: 23 June 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

In the recent strategic queueing literature, there is a large number of papers that study the join-or-balk dilemma in queueing systems with server's on-off periods, modeling vacations and failures. These studies consider the customers as discrete units and adopt the assumption that reneging is not permitted. In the present paper, we depart from this framework and study the effect of the reneging option in such systems. We consider the fluid on-off model of the basic queue with vacations/failures and study reneging vs. no-reneging when customers are strategic. We derive the equilibrium customer strategies and the corresponding performance measures of the system, and we use them to study the equilibrium throughput and social welfare. The main finding is that the existence of the reneging option is very beneficial for overloaded systems, i.e., for such systems balking alone is not sufficient to achieve good outcomes. On the contrary, for underloaded systems the reneging option is not particularly valuable.

Keywords Fluid queue · Strategic customers · Equilibrium strategies · Balking and reneging · Throughput · Social welfare

Mathematics Subject Classification 60K25 · 90B22

1 Introduction

The study of customer strategic behavior in queueing systems goes back at least to the pioneering papers of Naor (1969), and Edelson and Hildebrand (1975). A detailed overview

D. Logothetis was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.).

✉ Athanasia Manou
amanou@math.uoa.gr

Dimitrios Logothetis
dlogothetis@math.uoa.gr

Antonis Economou
aeconom@math.uoa.gr

[1] National and Kapodistrian University of Athens, Department of Mathematics, Panepistemiopolis, 15784 Athens, Greece

of this area of queueing can be found in Hassin & Haviv (2003), Stidham (2009), Hassin (2016) and Economou (2021). The consideration of queueing systems with server vacations has recently gained attention within the general literature. Indeed, chapter 10 of Hassin (2016) is solely devoted on strategic issues in queueing systems with vacations and summarizes more than 50 papers. However, the standard assumption in all these studies, is that the customers are prohibited from abandoning the system and they are only concerned with the join-or-balk dilemma. In the present paper, the main goal is to relax this quite artificial assumption of no-reneging. To keep the framework simple, we consider the Markovian single-server queue where the server is alternating between on and off periods according to a two-state Markov chain. This model, with strategic customers who are deciding whether to join or to balk (reneging is forbidden) and when the number of customers in the system is observable, has been initiated in Economou and Kanta (2008).

In the present paper, we consider the fluid version of this model where reneging is possible (FR) and the emphasis is on the comparison with the fluid no-reneging counterpart (FNR), with regard to the strategic customer behavior and the corresponding equilibrium throughput and social welfare. Our results show that, in regards to the equilibrium social welfare, reneging improves significantly an overloaded system while it has negligible effect for underloaded systems. However, the equilibrium throughput is the same in FR and FNR models.
The main contributions can be summarized as follows:

1. We derive double threshold equilibrium strategies with one threshold for each state of the server (on or off).
2. We compute the steady-state distributions of the system under the customer equilibrium strategies in the fluid reneging (FR) and the fluid no-reneging (FNR) cases.
3. We compare the equilibrium throughput and the equilibrium social welfare for the two cases (FR and FNR) of systems with identical parameters.
4. We explore the effect of system parameters on equilibrium throughput and social welfare for the FR and FNR cases.

The paper is organized in the following manner: In Sect. 2 we present a review of the related literature. In Sect. 3 we first describe the fundamental on-off Markovian queueing model and then its fluid counterpart which is the central model of the present paper. Moreover, we define the corresponding reward/cost structure. In Sects. 4 and 5, we derive the equilibrium customer strategies and compute the steady-state measures of the system in the two variants (FR and FNR), when the corresponding equilibrium strategies are employed. Finally, in Sect. 6, we compare the two systems, FR and FNR with regard to the equilibrium throughput and social welfare.

2 Literature review

This study falls within the stream of literature centered on the analysis of the customer strategic behavior regarding the stay-or-renege dilemma in queueing models with vacations or interruptions, while it is also related with the thread of research that contains the study of fluid models as approximations of service systems with discrete units.

The performance evaluation of queueing systems with vacations or interruptions has been studied intensively, usually in a Markovian framework of M/M/1-type queues in a multiple-phase environment, see e.g. Yechiali and Naor (1971) and Yechiali (1973).

The first study on strategic issues for this family of models was reported in Burnetas and Economou (2007) for the M/M/1 queue with multiple vacations. In that paper, the authors characterize customer equilibrium joining strategies for four levels of information according to whether the customers are informed or not about the number of customers in the system and/or the status of the server (serving or on vacation) upon their arrivals. Subsequently, this work was extended by many authors who considered Markovian variations and extensions of the model, most notably in (Huang et al., 2012; Ma et al., 2013; Liu et al., 2015; Zhang et al., 2013; Sun & Li, 2014; Yang et al., 2014a; Sun et al., 2010; Economou & Manou, 2013). A non-Markovian counterpart with generally distributed service and vacation times was studied by Economou et al. (2011).

The other fundamental vacation model is the M/M/1 queue with the N-policy. The strategic joining behavior for that model was initiated by Guo and Hassin (2011) and continued in a number of studies in (Tian et al., 2015; Sun et al., 2016; Chen et al., 2015; Guo & Hassin, 2012; Guo & Zhang, 2013; Guo & Li, 2013).

The strategic viewpoint in models with forced vacations was initially studied for the M/M/1 on-off queue. More specifically, Economou and Kanta (2008) studied the strategic customer behavior in the observable and almost-observable versions of this model, where the customers make their joining decisions knowing the number of customers in the system. Various authors studied variations and extensions of this model, by treating other information cases or by adding queueing features (Li & Han, 2011; Yang et al., 2014a, ?) (discrete-time models); (Do et al., 2012; Li et al., 2014; Jagannathan et al., 2012) (unobservable models); (Do et al., 2012a) (a model with working breakdowns); (Wang & Zhang, 2011) (a model with 2-phase off-periods). However, all these studies focus on joining behavior while reneging is prohibited. Economou et al. (2022) considered the M/M/1 on-off queue with discrete units, when reneging is allowed.

Regarding the second main related stream of the literature, we note that there is a large number of studies that go back to Kosten (1974a, b), Kosten and Vrieze (1975), and Anick et al. (1982). See also the books of Gautam (2012) and Schwartz (1996), and the review paper of Kulkarni (1997), for a smooth introduction in the area and a literature review of the classical references. The study of fluid models under an economic perspective has its roots in the 90s while the consideration of strategic customers, particularly in this framework, is a quite recent attempt.

From this viewpoint, Stidham et al. (1995) is one of the first studies that considered a fluid model under an economic perspective. The authors assumed a cost-reward structure for the customers and derived the socially optimal fluid rate. There are also several studies that examined the customer strategic behavior in fluid models. See the papers of Maglaras (2006), Jain et al. (2011), Juneja and Shimkin (2013), Haviv (2013). However, the service systems that are considered in these studies differ from ours. A key reference is the study of Economou and Manou (2016). In this study, the authors determined the equilibrium strategies as well as the optimal policy in the fluid version of an M/M/1 model with alternating service rate and with two levels of information. The model studied in this paper differs from our model, because the authors assumed that reneging is not allowed and the service rate is always positive. However, the non-reneging version of our model (FNR case) can be seen as limiting case of this paper, when the lower service rate tends to zero.

The approximation of an M/M/1 system where the server alternates between on and off periods with state dependent service rate as a fluid model, was investigated by Boxma et al. (2005). In this study, the authors determined the stationary distribution. Similar systems under the game theoretic perspective were also considered. Wang and Xu (2021) studied a fluid model with working vacation and an exponential reward-cost structure and determined the equilibrium strategies. Also, Liu et al. (2021) analyzed both the equilibrium strategies and the socially optimal policy for a fluid model with Markovian breakdowns and repairs

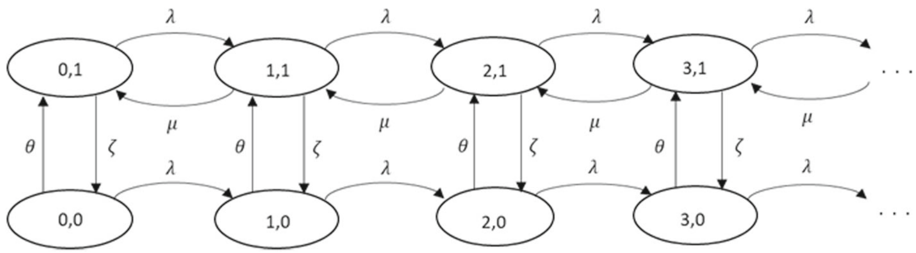

Fig. 1 Transition diagram of the original process

and with two types of customers. Two informational levels were considered in regards to whether the customers can inspect their type or not.

However, in all the aforementioned studies, the strategic behavior was studied when the reneging option was not available. The primary purpose of the present study is to shed light on the effect of allowing reneging on the equilibrium customer behavior and associated throughput and social welfare.

3 The model

The M/M/1 on-off queue is specified in the following manner: Customers arrive according to a Poisson process, at rate λ, at a single-server queue with exponential service times with rate μ and infinite waiting space that operates under the FCFS discipline. The server alternates between on and off periods that are also exponentially distributed with respective rates ζ and θ. Let $\rho = \frac{\lambda}{\mu}$. The state of the system at an arbitrary instant t is described by a vector of random variables $(N(t), I(t))$, where $N(t)$ is the number of customers in the system and $I(t)$ records the state of the server ($I(t) = 1 \rightarrow$ on, $I(t) = 0 \rightarrow$ off). The transition diagram of this process is given in Fig. 1.

The reward/cost structure that is imposed on the system consists of a service reward R and a unit waiting cost C. In other words, a customer receives R for completing service in the system and accumulates waiting costs at rate C.

We consider two models. The no-reneging model (NR case), where the customers decide, upon arrival, whether to join or to balk with the reneging option prohibited, and the reneging model (R case). In the latter, the customers consider the same dilemma but they may abandon the system at any time without being served. Both systems are assumed to be continuously observable. In particular, the information regarding the queue length and the server's state is shared across the customer population. Moreover, as long as they remain in the system, they can watch the evolution of the queue length and the server's state, and can decide to renege at any time (only in the R case).

In the present paper, we focus our study on the fluid limit version of the model. Formally, the fluid limit of the model corresponds to the scaled queue length process as both the arrival rate λ and the service rate μ tend to infinity, while keeping their ratio fixed. That is, we consider the sequence of processes $\{(N^{(m)}(t), I^{(m)}(t))\}$ with arrival rates $\lambda^{(m)} = \lambda m$, service rates $\mu^{(m)} = \mu m$ and constant rates $\zeta^{(m)} = \zeta, \theta^{(m)} = \theta$ (i.e., the processes $\{I^{(m)}(t)\}$ are all identical) and then we are interested in the behavior of the process $\{(X(t), I(t))\}$ where $I(t) = I^{(m)}(t),\ t \geq 0,$ and

$$X(t) = \lim_{m\to\infty} \frac{N^{(m)}(t)}{m}, \quad t \geq 0$$

(for details on the formal construction of fluid limit processes and interpretations see e.g. Gautam (2012)). Intuitively, the fluid model corresponds to a facility that processes a fluid, which alternates between on and off periods, which are independent and exponentially distributed with rates ζ and θ respectively. The input rate of the fluid is λ and its output rate is μ during on periods and 0 otherwise.

Thus, the customers in this model are infinitesimal units of a fluid. Such a model is a good approximation of the standard Markovian model with discrete units when the customer processes (arrivals and services) evolve at a much faster time-scale than the machine-related processes (on and off periods). And this is the rule in a number of applications like high-speed networks, automated manufacturing systems, etc. Moreover, the fluid version enables a more in-depth analysis, because of its deterministic nature regarding the customer processes and in particular due to the continuity of the queue-length state-space that simplifies various calculations. Indeed, we will see that the fluid model is more tractable than its standard Markovian counterpart which was studied in Economou et al. (2022). We will consider separately the R and the NR case for the fluid model which we will refer as the FR and FNR cases.

4 The case with reneging

In this section, we focus on the FR case and wish to determine the equilibrium strategy and the performance measures when all customers follow the equilibrium strategy. We consider a tagged customer who arrives at the system in the FR case. This customer has to decide whether to join or to balk, and, if she joins, she continuously faces the dilemma to stay or to renege. However, because of the Markovian nature of the model and its continuous and perfect observability, a customer may have an incentive to renege only at instants when the server is turned off. Moreover, at these instants, a customer makes her decision identically to a newly arriving customer who observes the server off and makes the decision to join or balk. We also note that the expected utility of a customer who observes the system at a specific state depends on the state of the system and on the reneging (not the joining) behavior of customers that are currently in the system. It depends neither on the history of the actions taken by previous customers nor on the behavior of future customers. Thus, using arguments similar to Yechiali (1971, 1972), one can show that only threshold strategies, i.e., strategies that are specified by a pair of thresholds (x_0^{FR}, x_1^{FR}), can be subgame-perfect equilibrium strategies.

Under (x_0^{FR}, x_1^{FR}) threshold strategy, a customer who finds upon arrival the server at state i joins if the fluid level is below x_i^{FR}, balks if the fluid level is above x_i^{FR}, and joins with any probability if the fluid level is x_i^{FR}. Moreover, a customer reneges only at the instants when the server's state changes and the level of fluid in front of her is above the corresponding threshold.

We now argue that it suffices to find equilibrium strategies within the set of threshold strategies (x_0^{FR}, x_1^{FR}) with $x_0^{FR} \leq x_1^{FR}$.

First, we justify that a threshold strategy (x_0^{FR}, x_1^{FR}) with $x_0^{FR} > x_1^{FR}$ cannot be equilibrium. We assume that all customers follow such a strategy and we consider a customer that finds the system at state $(x, 0)$ with $x_1^{FR} < x \leq x_0^{FR}$ upon arrival. According to this strategy, the customer joins, waits until the activation of the server (the fluid level in front of her remains $x > x_1^{FR}$), and reneges once the server is activated. Thus, the expected utility of this customer is negative, since she waits for a positive amount of time and does not receive service. Hence, this customer wouldn't prefer to join. Consequently, a threshold strategy (x_0^{FR}, x_1^{FR}) with $x_0^{FR} > x_1^{FR}$ cannot be equilibrium.

Therefore, we will assume hereafter that all customers follow a (x_0^{FR}, x_1^{FR}) threshold strategy, with $x_0^{FR} \leq x_1^{FR}$, and we will determine the expected reward of an arriving customer who also follows the same strategy, observes the system at state (x, i) and joins. We denote this expected reward by $S_{(x,i)}^{FR}(x_0^{FR}, x_1^{FR})$. Note that this is also the remaining expected reward of a customer who is already in the system when the server's state is i, the fluid level in front of her is x and decides to stay. The result is presented in the following theorem.

Theorem 1 (Expected reward - FR case) *In the fully observable fluid queue alternating between on and off periods, in the case with reneging, when all customers follow the (x_0^{FR}, x_1^{FR}) threshold strategy, with $x_0^{FR} \leq x_1^{FR}$, the expected reward from joining/staying of an arrival that finds fluid level x when the server's state is i is given by*

$$S_{(x,i)}^{FR}(x_0^{FR}, x_1^{FR}) = \begin{cases} R - C\left((1-i)\frac{1}{\theta} + \frac{x}{\mu}\left(1+\frac{\zeta}{\theta}\right)\right) & \text{if } x \leq x_0^{FR} \text{ and } i = 0, 1, \\ \left(R - \frac{Cx_0^{FR}(\theta+\zeta)}{\mu\theta} + \frac{C}{\zeta}\right)e^{-\frac{\zeta(x-x_0^{FR})}{\mu}} - \frac{C}{\zeta} & \text{if } x_0^{FR} < x \leq x_1^{FR} \text{ and } i = 1. \end{cases} \tag{1}$$

Proof Throughout this proof we will assume that all customers follow the (x_0^{FR}, x_1^{FR}) threshold strategy, with $x_0^{FR} \leq x_1^{FR}$, but we will suppress the notation regarding the strategy, i.e., $S_{(x,i)}^{FR}(x_0^{FR}, x_1^{FR})$ becomes $S_{(x,i)}^{FR}$. Note that when customers follow (x_0^{FR}, x_1^{FR}) threshold strategy, with $x_0^{FR} \leq x_1^{FR}$, an arriving customer may find the system at the states of the set $S^{FR} = \{(x, 0), x \leq x_0^{FR}\} \cup \{(x, 1), x \leq x_1^{FR}\}$. The expected reward of a tagged customer who decides to join given that she finds fluid level x and server's state i is

$$S_{(x,i)}^{FR} = RP_{(x,i)}^{FR} - CT_{(x,i)}^{FR}, \quad (x, i) \in S^{FR}, \tag{2}$$

where $P_{(x,i)}^{FR}$ is the probability of the tagged arrival to be served and $T_{(x,i)}^{FR}$ is the expected sojourn time.

Case 1: $x \leq x_0^{FR}$.

In this case we have that $P_{(x,i)}^{FR} = 1$, since a joining customer will not renege later. Suppose that $i = 0$. Then, the customer will wait for the completion of the remaining off time of the server, which is exponentially distributed with rate θ, and then she will see fluid level x in front of her and server's state 1. Hence,

$$T_{(x,0)}^{FR} = \frac{1}{\theta} + T_{(x,1)}^{FR}. \tag{3}$$

Now, consider a tagged arriving customer who finds fluid level x and server's state $i = 1$. Let T_1 denote her sojourn time and U be the remaining on time for the server. We have that U follows an exponential distribution with rate ζ. Conditioning on $U = u$ and taking into account that the fluid decreases at rate μ, we have that

$$T_{(x,1)}^{FR} = \int_0^{\infty} E[T_1|U = u]dF_U(u), \tag{4}$$

with

$$E[T_1|U = u] = \begin{cases} u + T_{(x-\mu u,0)}^{FR} & \text{if } 0 \leq u < \frac{x}{\mu}, \\ \frac{x}{\mu} & \text{if } u \geq \frac{x}{\mu}, \end{cases} \tag{5}$$

where the upper branch corresponds to the case where the change of the server's state occurs before the service completion of the customer (the customer stays in the system for u units and then the server is turned off and the fluid level in front of her is $x - \mu u$) and the lower branch corresponds to the case where the change occurs after the service completion (the customer stays in the system for $\frac{x}{\mu}$ units until she gets served).

From (3), (4), and (5), we obtain

$$T^{FR}_{(x,1)} = \int_0^{\frac{x}{\mu}} \left(u + \frac{1}{\theta} + T^{FR}_{(x-\mu u,1)} \right) dF_U(u) + \int_{\frac{x}{\mu}}^{\infty} \frac{x}{\mu} dF_U(u),$$

which reduces to

$$T^{FR}_{(x,1)} = \left(\frac{1}{\zeta} + \frac{1}{\theta} \right) \left(1 - e^{-\frac{\zeta x}{\mu}} \right) + \frac{\zeta}{\mu} e^{-\frac{\zeta x}{\mu}} \int_0^x T^{FR}_{(t,1)} e^{\frac{\zeta t}{\mu}} dt$$

after some straightforward algebraic manipulations. Multiplying by $e^{\frac{\zeta x}{\mu}}$ and differentiating with respect to x, we have that $\frac{d}{dx} T^{FR}_{(x,1)} = \frac{1}{\mu} \left(1 + \frac{\zeta}{\theta} \right)$. Taking into account (3) and (2), we have

$$T^{FR}_{(x,i)} = (1-i)\frac{1}{\theta} + \frac{x}{\mu} \left(1 + \frac{\zeta}{\theta} \right), \quad x \le x_0^{FR}, \tag{6}$$

$$S^{FR}_{(x,i)} = R - C \left((1-i)\frac{1}{\theta} + \frac{x}{\mu} \left(1 + \frac{\zeta}{\theta} \right) \right), \quad x \le x_0^{FR}. \tag{7}$$

Case 2: $x > x_0^{FR}$.

We now consider a joining customer seeing fluid level x and server's state $i = 1$. The customer will not abandon the system before she is served if the fluid level in front of her has decreased below x_0^{FR} when the next change of server's state from on to off occurs. Given that the fluid decreases at rate μ when the server is on, it equivalently means that the customer will not abandon if the change of the server's state occurs at least $\frac{x - x_0^{FR}}{\mu}$ time units after her arrival. Denoting by U the remaining on time for the server, we have

$$P^{FR}_{(x,1)} = \Pr[U \ge \frac{x - x_0^{FR}}{\mu}] = e^{-\frac{\zeta(x - x_0^{FR})}{\mu}}, \quad x > x_0^{FR}. \tag{8}$$

Denoting by T_2 the sojourn time of the customer, we have now that

$$T^{FR}_{(x,1)} = \int_0^{\infty} E[T_2 | U = u] dF_U(u), \tag{9}$$

with

$$E[T_2 | U = u] = \begin{cases} u & \text{if } 0 \le u < \frac{x - x_0^{FR}}{\mu}, \\ u + T^{FR}_{(x-\mu u,0)} & \text{if } \frac{x - x_0^{FR}}{\mu} \le u < \frac{x}{\mu}, \\ \frac{x}{\mu} & \text{if } u \ge \frac{x}{\mu}, \end{cases} \tag{10}$$

where the upper branch corresponds to the case where the change of the server's state occurs while the fluid level in front of the customer is still above x_0^{FR} and consequently the customer reneges at the instant of the change (the customer stays in the system for u units and then reneges without receiving service), the middle branch corresponds to the case where the change occurs while the fluid level is at most x_0^{FR} but positive and the customer does not renege (the customer stays in the system for u units and then the server is turned off and the fluid level in front of her is $x - \mu u$), and the lower branch corresponds to the case where the change of the server's state occurs after the service completion of the customer (the customer stays in the system for $\frac{x}{\mu}$ units until she gets served).

Now, using (9), (10), and (3), we obtain

$$T^{FR}_{(x,1)} = \int_0^{\frac{x-x_0^{FR}}{\mu}} u dF_U(u) + \int_{\frac{x-x_0^{FR}}{\mu}}^{\frac{x}{\mu}} \left(u + \frac{1}{\theta} + T^{FR}_{(x-\mu u,1)}\right) dF_U(u) + \int_{\frac{x}{\mu}}^{\infty} \frac{x}{\mu} dF_U(u).$$

Note that $T^{FR}_{(x-\mu u,1)}$ in the right-hand side of the above equation is given by (6), so after substituting it and computing the integrals we obtain

$$T^{FR}_{(x,1)} = \frac{1}{\zeta} + \left(\frac{x_0^{FR}(\theta+\zeta)}{\mu\theta} - \frac{1}{\zeta}\right) e^{-\frac{\zeta(x-x_0^{FR})}{\mu}}, \quad x > x_0^{FR}. \tag{11}$$

Then, using (2), (8), and (11), we obtain

$$S^{FR}_{(x,1)} = \left(R - \frac{Cx_0^{FR}(\theta+\zeta)}{\mu\theta} + \frac{C}{\zeta}\right) e^{-\frac{\zeta(x-x_0^{FR})}{\mu}} - \frac{C}{\zeta}, \quad x > x_0^{FR}.$$

□

The next result gives the equilibrium threshold strategy in the FR case.

Theorem 2 (Equilibrium strategies - FR case) *In the fully observable fluid queue alternating between on and off periods, in the case with reneging, the equilibrium strategy is given by the pair of thresholds*

$$(x_{e,0}^{FR}, x_{e,1}^{FR}) = \left(\frac{R\mu\theta}{C(\zeta+\theta)} - \frac{\mu}{\zeta+\theta}, \frac{R\mu\theta}{C(\zeta+\theta)} - \frac{\mu}{\zeta+\theta} + \frac{\mu}{\zeta}\log\left(1+\frac{\zeta}{\theta}\right)\right). \tag{12}$$

Proof The threshold strategy $(x_{e,0}^{FR}, x_{e,1}^{FR})$ is equilibrium if it is best response against itself. Thus, considering first the states where the server is off, we should have

$$S^{FR}_{(x,0)}(x_{e,0}^{FR}, x_{e,1}^{FR}) > 0, \quad 0 \le x < x_{e,0}^{FR}, \tag{13}$$

$$S^{FR}_{(x_{e,0}^{FR},0)}(x_{e,0}^{FR}, x_{e,1}^{FR}) = 0. \tag{14}$$

We can see from the upper branch of (1) that $S^{FR}_{(x,0)}(x_{e,0}^{FR}, x_{e,1}^{FR})$ is decreasing in $x \in [0, x_{e,0}^{FR}]$. Thus, condition (14) implies condition (13). Solving (14) for $x_{e,0}^{FR}$, we get the first threshold of the equilibrium strategy

$$x_{e,0}^{FR} = \frac{R\mu\theta}{C(\zeta+\theta)} - \frac{\mu}{\zeta+\theta}.$$

Moreover, we have that $R = \frac{C(\zeta+\theta)x_{e,0}^{FR}}{\mu\theta} + \frac{C}{\theta}$ and substituting it into (1) yields

$$S^{FR}_{(x,i)}(x_{e,0}^{FR}, x_{e,1}^{FR}) = \begin{cases} \left(\frac{C}{\theta} + \frac{C}{\zeta}\right) \frac{\zeta(x_{e,0}^{FR} - x)}{\mu} + i\frac{C}{\theta} & \text{if } x \le x_{e,0}^{FR} \text{ and } i = 0, 1, \\ \left(\frac{C}{\theta} + \frac{C}{\zeta}\right) \exp\left(-\frac{\zeta(x - x_{e,0}^{FR})}{\mu}\right) - \frac{C}{\zeta} & \text{if } x_{e,0}^{FR} \le x \le x_{e,1}^{FR} \text{ and } i = 1. \end{cases}$$

Thus, the expected reward under the equilibrium strategy is also decreasing in the level of fluid x when the server is on.

Now, considering the states where the server is on, in order $(x_{e,0}^{FR}, x_{e,1}^{FR})$ to be equilibrium strategy we should have

$$S_{(x,1)}^{FR}(x_{e,0}^{FR}, x_{e,1}^{FR}) > 0, \quad 0 \le x < x_{e,1}^{FR}, \tag{15}$$

$$S_{(x_{e,1}^{FR},1)}^{FR}(x_{e,0}^{FR}, x_{e,1}^{FR}) = 0. \tag{16}$$

Again, because of the monotonicity of $S_{(x,1)}^{FR}(x_{e,0}^{FR}, x_{e,1}^{FR})$, condition (16) implies condition (15). Solving (16) for $x_{e,1}^{FR}$, we get the second threshold of the equilibrium strategy

$$x_{e,1}^{FR} = \frac{R\mu\theta}{C(\zeta+\theta)} - \frac{\mu}{\zeta+\theta} + \frac{\mu}{\zeta}\log\left(1+\frac{\zeta}{\theta}\right).$$

□

We now proceed to compute the steady-state distribution of the process $\{(X(t), I(t))\}$, when the customers follow the equilibrium strategy. That is, we compute the functions

$$F_i(x) = \lim_{t\to\infty} \Pr[X(t) \le x, I(t) = i], \quad x \ge 0, \quad i \in \{0, 1\}. \tag{17}$$

For simplicity, we will denote in the rest of the subsection $x_{e,0}^{FR}$ and $x_{e,1}^{FR}$ by x_0 and x_1, respectively. Moreover, we set

$$\alpha_0 = \frac{\theta}{\lambda}, \quad \alpha_1 = \frac{\zeta}{\lambda-\mu} \quad \text{and} \quad \alpha = \alpha_0 + \alpha_1. \tag{18}$$

Theorem 3 (Steady-state distribution under equilibrium - FR case) *In the fully observable fluid queue alternating between on and off periods, in the case with reneging, when the equilibrium threshold strategy is employed, the steady-state distribution of the fluid process is given as follows:*

Case 1: $\lambda > \mu$. *The fluid level oscillates in* $[x_0, x_1]$. *In particular,*

$$F_0(x) = \begin{cases} 0 & \text{if } 0 \le x < x_0, \\ \frac{\zeta}{\theta+\zeta} & x \ge x_0 \end{cases} \tag{19}$$

and

$$F_1(x) = \begin{cases} 0 & \text{if } 0 \le x < x_0, \\ \frac{\theta}{\theta+\zeta}\left(1-e^{-\alpha_1(x-x_0)}\right) & \text{if } x_0 \le x < x_1, \\ \frac{\theta}{\theta+\zeta} & \text{if } x \ge x_1. \end{cases} \tag{20}$$

Note that $F_0(x)$ *has a point probability mass at* x_0 *and* $F_1(x)$ *is mixed with a continuous probability density in* $[x_0, x_1]$ *and a point probability mass at* x_1. *The mean fluid level is*

$$E[X] = x_0 + \frac{\theta}{\theta+\zeta} \cdot \frac{1-e^{-\alpha_1(x_1-x_0)}}{\alpha_1}. \tag{21}$$

Case 2: $\lambda = \mu$. *After an initial transient phase, the fluid level stabilizes on* x_0.
Case 3: $\lambda < \mu$. *The fluid level oscillates in* $[0, x_0]$. *In particular,*

$$F_0(x) = \begin{cases} \frac{\zeta}{\theta+\zeta} \cdot \frac{\alpha_0(1-e^{-\alpha x})}{\alpha_0+\alpha_1 e^{-\alpha x_0^e}} & \text{if } 0 \le x < x_0, \\ \frac{\zeta}{\theta+\zeta} & \text{if } x \ge x_0 \end{cases} \tag{22}$$

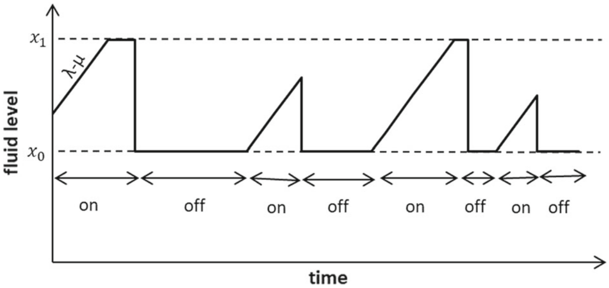

Fig. 2 Typical path of fluid level when $\lambda > \mu$

and

$$F_1(x) = \begin{cases} \frac{\theta}{\theta+\zeta} \cdot \frac{\alpha_0+\alpha_1 e^{-\alpha x}}{\alpha_0+\alpha_1 e^{-\alpha x_0}} & \text{if } 0 \le x < x_0, \\ \frac{\theta}{\theta+\zeta} & \text{if } x \ge x_0. \end{cases} \tag{23}$$

Note that both $F_0(x)$ *and* $F_1(x)$ *are mixed with continuous probability densities in* $[0, x_0]$ *and point probability masses at the points* x_0 *and* 0*, respectively. The mean fluid level is*

$$E[X] = \frac{\alpha_1 e^{-\alpha x_0}}{\alpha_0 + \alpha_1 e^{-\alpha x_0}} \cdot x_0 - \frac{\mu}{\theta+\zeta} \cdot \frac{\alpha_0 \alpha_1 (1 - e^{-\alpha x_0})}{(\alpha_0 + \alpha_1 e^{-\alpha x})\alpha}. \tag{24}$$

Proof We will consider separately the various cases of the theorem.
Case 1: $\lambda > \mu$.

Suppose that the fluid is initially at some level $x < x_0$. As long as it remains below x_0, all customers join and it will increase linearly at rate $\lambda - \mu$ if the server is on or at rate λ if the server is off. Therefore, sooner or later, it will reach level x_0. Then, after reaching x_0, it will remain there if the server is off or will grow at rate $\lambda - \mu$ if the server is on till it reaches level x_1 where it will remain till the next change of the server's state from on to off. At such a time, customers that have fluid level above x_0 in front of them renege and the fluid level will drop immediately to level x_0 and will stay there till the next switch of the server state from off to on etc.

If the level is initially at some level $x \ge x_0$, it will drop to x_0 at the first change of the server's state from on to off and afterwards the dynamics are as described above. Therefore, the fluid level will oscillate in $[x_0, x_1]$. Figure 2 shows how the fluid oscillates in $[x_0, x_1]$ in this case.

An oscillation cycle comprises two phases with exponential duration with parameters θ and ζ, corresponding to an off and an on period of the server. During the off period the fluid stays at level x_0, whereas during the on period it grows linearly at rate $\lambda - \mu$ till reaching x_1 and stays there till the end of the period. Therefore, $F_0(x)$, for $x \ge x_0$ is just the long-run fraction of time that the server is off which is $\frac{\zeta}{\theta+\zeta}$ and we obtain (19). Similarly, $F_1(x)$ is the long-run fraction of time that the server is on and the fluid level is below x. Since at the beginning of each on period the fluid is at level x_0 and during on periods it grows at rate $\lambda - \mu$, it stays below x, for $x \in [x_0, x_1)$, as long as the duration of on period does not exceed $\frac{x-x_0}{\lambda-\mu}$ time units. Thus, by using the renewal-reward theorem, we can focus in an oscillation

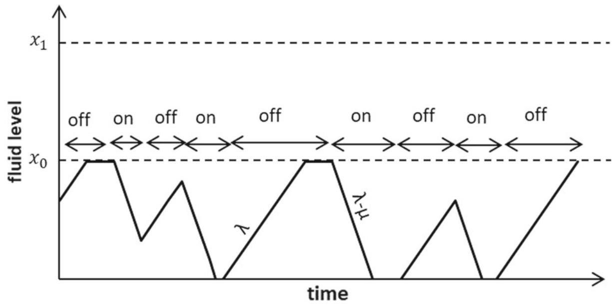

Fig. 3 Typical path of fluid level when $\lambda < \mu$

cycle and we have that

$$F_1(x) = \frac{E\left[\min\left(\frac{x-x_0}{\lambda-\mu}, U\right)\right]}{\frac{1}{\theta} + \frac{1}{\zeta}}, \quad x \in [x_0, x_1),$$

where U represents an on time of the server. After some straightforward computations we obtain (20). The mean fluid level is then directly computed using that $E[X] = \int_0^\infty (1 - F_0(x) - F_1(x))dx$ which yields (21).

Case 2: $\lambda = \mu$.

Suppose that the fluid is initially at some level $x < x_0$. Then, it increases linearly at rate λ as long as the server is off, and stays at the same level when the server is on, till it reaches the state x_0 (where it remains forever thereafter). If the fluid starts from a level $x \geq x_0$, it remains constant (if $x < x_1$) or decreases at rate μ (if $x \geq x_1$) till the first switch of the server from the on to the off mode. At that time, it drops to level x_0 and remains there thereafter.

Case 3: $\lambda < \mu$.

Suppose that the fluid is initially at some level $x < x_0$. Then, it increases linearly at rate λ whenever the server is off, till it reaches level x_0, or decreases linearly at rate $|\lambda - \mu|$ whenever the server is on, till it reaches level 0. Therefore, the fluid oscillates in $[0, x_0]$.

If the fluid starts from a level $x \geq x_0$, it decreases at rate $|\lambda - \mu|$ (if $x < x_1$) or decreases at rate μ (if $x \geq x_1$) till the first switch of the server from the on to the off mode. At that time, it drops to level x_0 and afterwards the dynamics are as described above. Therefore, the fluid will again oscillate in $[0, x_0]$. Figure 3 shows how the fluid oscillates in $[0, x_0]$ in this case.

Using the standard methodology for the analysis of fluid queues (Kulkarni, 1997; Section 9.2 in Gautam (2012); Economou and Manou (2016), we have that the functions $F_0(x)$ and $F_1(x)$ are differentiable in $(0, x_0)$ and satisfy the linear system of ODEs

$$(\lambda - \mu)\frac{dF_1(x)}{dx} = -\zeta F_1(x) + \theta F_0(x),$$
$$\lambda\frac{dF_0(x)}{dx} = \zeta F_1(x) - \theta F_0(x),$$

with boundary conditions

$$F_0(0) = 0, \qquad F_1(x_0) = \frac{\theta}{\zeta + \theta}.$$

Solving this system using the standard theory for first-order linear ODE systems with constant coefficients (see e.g., Braun, 1983), we have that the system has a unique solution given by (22) and (23), where α_0, α_1 and α are given by (18). The mean fluid level is then directly computed using that $E[X] = \int_0^\infty (1 - F_0(x) - F_1(x))dx$ which yields (24). □

Using the above results we can now compute the equilibrium throughput and the equilibrium social welfare functions denoted by TH_e^{FR} and SW_e^{FR}, respectively.

Corollary 1 (Equilibrium throughput and social welfare–FR case) *In the fully observable fluid queue alternating between on and off periods, in the case with reneging, we have the following cases regarding the throughput and the social welfare per time unit under the equilibrium strategy:*

Case 1: $\lambda > \mu$.

$$TH_e^{FR} = \frac{\mu\theta}{\theta + \zeta}, \tag{25}$$

$$\begin{aligned} SW_e^{FR} &= \frac{R\mu\theta}{\theta + \zeta} - C\left(x_0 + \frac{\theta}{\theta + \zeta} \cdot \frac{1 - e^{-\alpha_1(x_1 - x_0)}}{\alpha_1}\right) \\ &= \frac{C\mu}{\theta}\left(1 - \frac{\theta}{\zeta} \cdot \frac{1 - \left(\frac{\theta}{\theta+\zeta}\right)^{\mu/(\lambda-\mu)}}{\mu/(\lambda - \mu)}\right). \end{aligned} \tag{26}$$

Case 2: $\lambda = \mu$.

$$TH_e^{FR} = \frac{\mu\theta}{\theta + \zeta}, \tag{27}$$

$$SW_e^{FR} = \frac{C\mu}{\theta + \zeta}. \tag{28}$$

Case 3: $\lambda < \mu$.

$$TH_e^{FR} = \frac{\lambda\theta}{\theta + \zeta} \cdot \frac{\alpha_0 + \alpha_1}{\alpha_0 + \alpha_1 e^{-\alpha x_0}} - \frac{\mu\theta}{\theta + \zeta} \cdot \frac{\alpha_1(1 - e^{-\alpha x_0})}{\alpha_0 + \alpha_1 e^{-\alpha x_0}}, \tag{29}$$

$$\begin{aligned} SW_e^{FR} &= \frac{R\lambda\theta}{\theta + \zeta} \cdot \frac{\alpha_0 + \alpha_1}{\alpha_0 + \alpha_1 e^{-\alpha x_0}} - \frac{C\alpha_1 x_0}{\alpha_0 + \alpha_1 e^{-\alpha x_0}} \\ &\quad - \frac{C\mu\alpha_1^2(1 - e^{-\alpha x_0})}{(\theta + \zeta)(\alpha_0 + \alpha_1 e^{-\alpha x_0})\alpha}. \end{aligned} \tag{30}$$

Proof We have that

$$SW_e^{FR} = RTH_e^{FR} - CE[X]. \tag{31}$$

We consider three cases according to the ordering of λ and μ.
Case 1: $\lambda > \mu$.

In this case, the system is never empty and the server serves customers at rate μ as long as he is on. Therefore, the throughput is μ times the long-run fraction of time that the server is on, which yields (25). Regarding the social welfare, plugging (25) and (21) into (31), yields (26).
Case 2: $\lambda = \mu$.

As in case 1,the system is never empty and we obtain that the throughput is given by (27). Moreover, by (31), we have that $SW_e^{FR} = RTH_e^{FR} - Cx_0$ and substituting x_0 by (12) yields (28).

Case 3: $\lambda < \mu$.

The server serves customers at rate μ as long as he is on and the fluid level is greater than 0, and at rate λ as long as he is on and the fluid level is 0. Therefore,

$$TH_e^{FR} = (1 - F_1(0))\mu + F_1(0)\lambda,$$

where $F_1(0) = \frac{\theta}{\theta+\zeta} \cdot \frac{\alpha_0+\alpha_1}{\alpha_0+\alpha_1 e^{-\alpha x_0}}$ (using (23)). After a few simplification, we deduce (29). For computing SW_e^{FR}, we use again (31) and substitute TH_e^{FR} by (29) and $E[X]$ by (24). □

5 The case without reneging

In this section, we assume that reneging is not allowed. Thus, a customer who joins the system stays in it until her service completion. As in the case with reneging, we will search for equilibrium strategies among the set of threshold strategies, i.e., strategies that are specified by a pair of thresholds (x_0^{FNR}, x_1^{FNR}). However, in this case, under (x_0^{FNR}, x_1^{FNR}) threshold strategy, a customer who finds the server at state i upon arrival, joins if the fluid level is below x_i^{FNR}, balks if the level of fluid exceeds x_i^{FNR}, and joins with any probability if the level of fluid is exactly x_i^{FNR}. A joining customer cannot renege later.

As we did in the FR case, we will first assume that all customers follow the threshold strategy (x_0^{FNR}, x_1^{FNR}) and determine the expected reward of a tagged customer that finds the system at state (x, i) and joins. This will allow as to determine the equilibrium strategy. The results are presented in Theorems 4 and 5.

Theorem 4 (Expected reward - FNR case) *In the fully observable fluid queue alternating between on and off periods, in the case without reneging, assuming that all customers follow the threshold strategy (x_0^{FNR}, x_1^{FNR}), the expected reward from joining of an arrival that finds fluid level x when the server's state is i is given by*

$$S_{(x,i)}^{FNR}(x_0^{FNR}, x_1^{FNR}) = R - C\left((1-i)\frac{1}{\theta} + \frac{x}{\mu}\left(1 + \frac{\zeta}{\theta}\right)\right),$$
$$x \le \max(x_0^{FNR}, x_1^{FNR}), i = 0, 1. \quad (32)$$

Proof We assume that all customers follow (x_0^{FNR}, x_1^{FNR}) threshold strategy. Then, an arriving customer may find the system at the states of the set $S^{FNR} = \{(x, i), \ x \in [0, \max(x_0^{FNR}, x_1^{FNR})], i = 0, 1\}$. We want to compute the expected reward of a tagged customer that finds the system at state $(x, i) \in S^{FNR}$ and joins, $S_{(x,i)}^{FNR}(x_0^{FNR}, x_1^{FNR})$. Since customers do not renege, the derivation of $S_{(x,i)}^{FNR}(x_0^{FNR}, x_1^{FNR})$ is basically a repetition of the derivation of (7). Thus, we obtain (32). □

We will now proceed to the computation of equilibrium strategies.

Theorem 5 (Equilibrium strategies - FNR case) *In the fully observable fluid queue alternating between on and off periods, in the case without reneging, the equilibrium strategy is given by the pair of thresholds*

$$(x_{e,0}^{FNR}, x_{e,1}^{FNR}) = \left(\frac{R\mu\theta}{C(\zeta+\theta)} - \frac{\mu}{\zeta+\theta}, \frac{R\mu\theta}{C(\zeta+\theta)}\right). \quad (33)$$

Proof The threshold strategy $(x_{e,0}^{FNR}, x_{e,1}^{FNR})$ is equilibrium if it is best response against itself. Since the expected reward $S_{(x,i)}^{FNR}(x_{e,0}^{FNR}, x_{e,1}^{FNR})$ is decreasing in x, for $i = 0, 1$, it suffices to satisfy

$$S_{(x_{e,i}^{FNR},i)}^{FNR}(x_{e,0}^{FNR}, x_{e,1}^{FNR}) = 0, \ i = 0, 1.$$

Solving the above equations with respect to $x_{e,0}^{FNR}$ and $x_{e,1}^{FNR}$, we get the equilibrium strategy given by (33). □

Regarding the steady-state distribution of the fluid model, we consider again the functions $F_i(x)$, $i = 0, 1$, defined in (17). In the rest of this subsection, we will denote the equilibrium strategy by (x_0, x_1).

Theorem 6 (Steady-state distribution - FNR case) *In the fluid queue alternating between on and off periods, in the case without reneging, when customers join according to the equilibrium threshold strategy, the steady-state distribution of the fluid process is given as follows:*

Case 1: $\lambda > \mu$. *After an initial transient phase, the fluid level stabilizes on* x_1.
Case 2: $\lambda = \mu$. *Let* x_s *the initial state of the fluid. After an initial transient phase, the fluid level stabilizes on* x_* *which depends on* x_s *as follows:*

$$x_* = \begin{cases} x_0 \text{ if } x_s \in [0, x_0], \\ x_s \text{ if } x_s \in (x_0, x_1), \\ x_1 \text{ if } x_s \in [x_1, \infty). \end{cases} \tag{34}$$

Case 3: $\lambda < \mu$. *The fluid level oscillates in* $[0, x_0]$. *The distribution functions* $F_0(x)$ *and* $F_1(x)$ *are given by (22) and (23). Moreover, the mean fluid level is given by (24).*

Proof We again consider the three cases regarding the relative values of λ and μ.
Case 1: $\lambda > \mu$.

If the fluid starts at a level $x \geq x_1$, then it decreases at rate μ whenever the server is on till it reaches x_1. If the fluid starts at a level $x < x_1$, then it increases at rate λ whenever the server is off and its level is below x_0, it stays at the same level whenever the server is off and its level is greater than x_0, and it increases at rate $\lambda - \mu$ whenever the server is on, till it reaches x_1. After reaching x_1, the fluid remains forever at this level.
Case 2: $\lambda = \mu$.

If the initial level of the fluid is $x_s \in [0, x_0]$, then it increases with rate λ when the server is off, and remains constant when the server is on, till it reaches x_0. It remains at level x_0 thereafter. If the initial level is $x_s \in (x_0, x_1]$, then it remains constant forever. Finally, if it starts from a state greater than x_1, then it decreases with rate μ when the server is on, and remains constant when the server is off, till it reaches x_1. Then, it stabilizes at x_1.
Case 3: $\lambda < \mu$.

This case is identical to case 3 of Theorem 3 and the proof is the same. □

We can apply these results and obtain the equilibrium throughput and the equilibrium social welfare functions, denoted by TH_e^{FNR} and SW_e^{FNR}, respectively.

Corollary 2 (Throughput and social welfare - FNR case) *In the fluid queue alternating between on and off periods, in the case without reneging, when customers join according to the equilibrium threshold strategy, we have the following cases regarding the throughput and the social welfare per time unit:*

Case 1: $\lambda > \mu$.

$$TH_e^{FNR} = \frac{\mu\theta}{\theta + \zeta}, \quad SW_e^{FNR} = 0. \tag{35}$$

Case 2: $\lambda = \mu$.

$$TH_e^{FNR} = \frac{\mu\theta}{\theta + \zeta}, \quad SW_e^{FNR} = \begin{cases} \frac{C\mu}{\theta+\zeta} & \text{if } x_s \in [0, x_0], \\ \frac{C\mu}{\theta+\zeta} - C(x_s - x_0) & \text{if } x_s \in (x_0, x_1), \\ 0 & \text{if } x_s \in [x_1, \infty), \end{cases} \tag{36}$$

where x_s is the initial state of the fluid.

Case 3: $\lambda < \mu$.

$$TH_e^{FNR} = TH_e^{FR}, \quad SW_e^{FNR} = TH_e^{FR}. \tag{37}$$

where TH_e^{FR} and SW_e^{FR} are given by (29) and (30), respectively.

Proof In cases 1 and 2, the system is never empty. Thus, the throughput is the long-run fraction of time that the server is on times the service rate μ. For case 3, the proof is identical to case 3 in Corollary 1. Regarding the social welfare, we use that $SW_e^{FNR} = RTH_e^{FNR} - CE[X]$ and we substitute the appropriate formula for $E[X]$ from Theorem 6. □

6 Numecical experiments–conclusions

In this section, we study the impact of reneging in more detail by using both the theoretical results of the previous sections, as well as several numerical experiments. The general insights of the present paper are summarized bellow.

1. $x_{e,0}^{FR} = x_{e,0}^{FNR} < x_{e,1}^{FNR} < x_{e,1}^{FR}$. The first inequality is obvious from (12) and (33) while for the second we notice that $x_{1,e}^{FNR} < x_{1,e}^{FR} \Leftrightarrow x \log x > x - 1$ which is valid for $x > 1$. That is, customers are more likely to join the system when reneging is allowed, which does not strike as a surprise.
2. When $\lambda < \mu$, i.e., when the arrival rate is lower than the service rate, the evolution of the system is the same in both FNR and FR cases. This is due to the fact that the fluid level never gets above the equilibrium threshold $x_{0,e}^{FR} = x_{0,e}^{FNR}$ and thus, even in the FR case, the customers never renege. Thus, in this case $TH_e^{FR} = TH_e^{FNR}$ and $SW_e^{FR} = SW_e^{FNR}$.
3. When $\lambda \geq \mu$, we have that $TH_e^{FR} = TH_e^{FNR}$. Thus, although some customers may renege in the FR model, the fraction of the customers who receive service, stays the same in both models.
4. Additionally, when $\lambda \geq \mu$, we have that $SW_e^{FR} \geq SW_e^{FNR} = 0$. That is, although the same fraction of customers receive service, they always seem to be able to achieve higher gain when they are not restricted to stay in the system. Thus, in terms of the social welfare, allowing reneging should always be preferred in this case.
5. We also note that, the throughput, TH_e, is always independent of both economic parameters whereas the social welfare, SW_e, when $\lambda \geq \mu$, is linear with respect to C and does not depend on R.

The main conclusion is that the reneging option is particularly valuable for the case where $\lambda > \mu$. The basic insight that reneging is valuable for overloaded systems is also valid for the corresponding queueing model with discrete units (for details see Economou et al. (2022)).

6.1 Effect of the server's alternating speed

In this subsection, we investigate the effect of server's alternating speed when the fraction of time that the server stays on-line (consequently off-line) is kept fixed. To this end, we set $\delta = \frac{1}{\theta}$ and $\gamma = \frac{\theta}{\theta+\zeta}$, we keep γ fixed and let δ vary. We explore the effect of δ on the equilibrium strategy and on the equilibrium throughput and social welfare.

Writing the equilibrium threshold with respect to γ and δ, we have

$$x_{e,0}^{FR}(\delta) = x_{e,1}^{FR}(\delta) = \mu\gamma(\frac{R}{C} - \delta),$$

$$x_{e,1}^{FR}(\delta) = \mu\gamma\left[\frac{R}{C} + \delta\left(\frac{\log(\frac{1}{\gamma})}{1-\gamma} - 1\right)\right],$$

$$x_{e,1}^{FNR}(\delta) = \mu\gamma\frac{R}{C}.$$

Thus, $x_{e,0}^{FR}$ is linearly decreasing in δ and becomes 0 when $\delta \geq \frac{R}{C}$, $x_{e,1}^{FR}$ is increasing in δ and $x_{e,1}^{FNR}$ is independent of δ. Therefore, as the alternating speed decreases ($\delta = \frac{1}{\theta}$ increases), in the FR case, customers tend to join more when find the server on, but they join less when they find the server off. Also, the reneging threshold, $x_{e,0}^{FR}$, decreases. In the case without reneging, the threshold $x_{e,1}^{FNR}$ is independent of δ, so a decrease in the alternating speed does affect the joining behavior when the server is on. However, customers join less when the server is off as they expect to stay at the off state for a longer period of time.

Regarding the effect of server's alternating speed on equilibrium social welfare and throughput, we have that, for $\lambda > \mu$,

$$TH_e^{FNR}(\delta) = TH_e^{FR}(\delta) = \mu\gamma,$$
$$SW_e^{FNR}(\delta) = 0,$$
$$SW_e^{FR}(\delta) = \delta\left[C\mu\left(1 - \frac{\gamma}{1-\gamma}\frac{1-\gamma^b}{b}\right)\right],$$

where $b = \frac{\mu}{\lambda-\mu}$. Thus, when $\lambda > \mu$, the server's alternating speed does not affect the equilibrium throughput (in both cases) and the equilibrium welfare in the case without reneging. However, as the alternating speed decreases, the expected social welfare linearly increases in the case with reneging. This is due to the fact that the customers who find the server on and join, will receive service in the first on-time with higher probability.

When $\lambda = \mu$, we have that

$$TH_e^{FNR}(\delta) = TH_e^{FR}(\delta) = \mu\gamma,$$

$$SW_e^{FNR}(\delta) = \begin{cases} \mu\gamma C\delta & \text{if } x_s \in [0, x_0], \\ \mu\gamma R - Cx_s & \text{if } x_s \in (x_0, x_1), \\ 0 & \text{if } x_s \in [x_1, \infty), \end{cases}$$

$$SW_e^{FR}(\delta) = \mu\gamma C\delta.$$

Thus, in this case, the equilibrium throughput is independent of the alternating speed whereas the equilibrium social welfare increases linearly as the alternating speed decreases in the FR case and also in the FNR when $x_s \in [0, x_0]$. However, in the FNR case with $x_s > x_0$, the social welfare is independent of the alternating speed.

The effect of the server's alternating speed on throughput and social welfare, when $\lambda < \mu$, is shown in Fig. 4, where we consider a numerical example with $(\mu, \zeta, \theta, R, C) =$

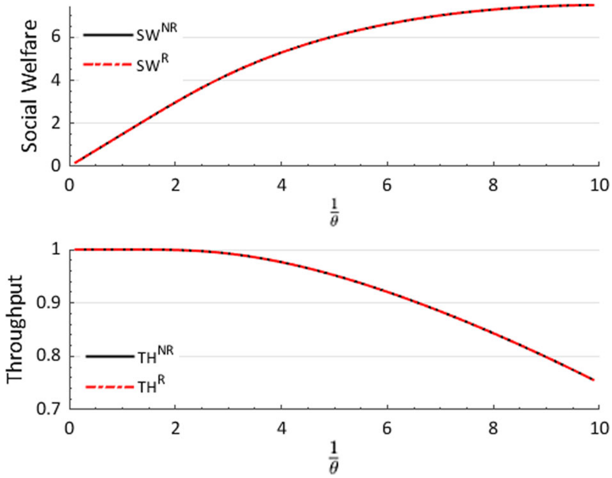

Fig. 4 Equilibrium social welfare and throughput with respect to $\delta = \frac{1}{\theta}$ for $(\lambda, \mu, R, C) = (1.5, 2, 10, 1)$

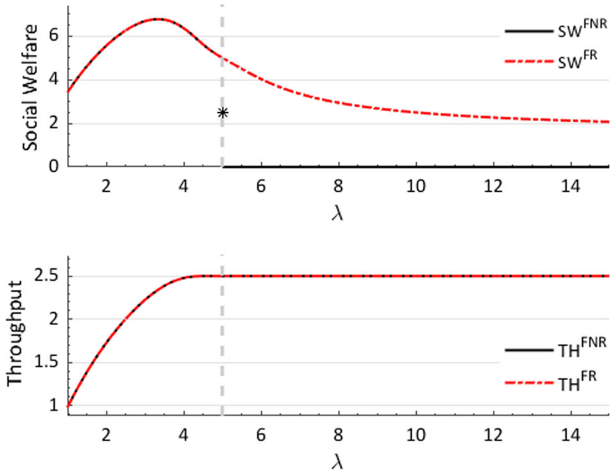

Fig. 5 Equilibrium social welfare and throughput with respect to λ for $(\mu, \zeta, \theta, R, C) = (5, 0.5, 0.5, 5, 1)$

$(5, 0.5, 0.5, 5, 1)$. We notice that the equilibrium social welfare is increasing and concave and the equilibrium throughput is decreasing and concave. Thus, as the alternating speed decreases (δ increases), the fraction of customers that receive service decreases but the social welfare increases.

6.2 Effect of the arrival rate

In this subsection, we explore the effect of the arrival rate λ, on the equilibrium throughput and social welfare for the FR and FNR cases through a numerical example. Specifically, we set the parameters $(\mu, \zeta, \theta, R, C) = (5, 0.5, 0.5, 5, 1)$. The results are depicted in Fig. 5. Consistently to our theoretical analysis, the equilibrium throughput is equal in the FR and FNR cases. Moreover, the equilibrium social welfare is equal in the FR and FNR cases for

$\lambda < \mu$, equals 0 in the FNR case for $\lambda > \mu$, and is not continuous in λ in the FNR case. As λ increases, the equilibrium social welfare first increases and then decreases.

Acknowledgements D. Logothetis was supported by the Hellenic Foundation for Research and Innovation (HFRI) under the HFRI PhD Fellowship grant (Fellowship Number: 1158).

References

Anick, D., Mitra, D., & Sondhi, M. M. (1982). Stochastic theory of a data-handling system with multiple sources. *The Bell System Technical Journal, 61*, 1871–1894.
Braun, M. (1983). *Differential Equations and their Applications* (3rd ed.). New York: Springer-Verlag.
Boxma, O., Kaspi, H., Kella, O., & Perry, D. (2005). On/off storage systems with state-dependent input, output, and switching rates. *Probability in the Engineering and Informational Sciences, 19*(1), 1–14.
Burnetas, A., & Economou, A. (2007). Equilibrium customer strategies in a single server Markovian queue with setup times. *Queueing Systems, 56*, 213–228.
Chen, P., Zhou, W., & Zhou, Y. (2015). *Equilibrium customer strategies in the queue with threshold policy and setup times* (p. 361901). Article ID: Mathematical Problems in Engineering.
Do, C. T., Hong, C. S., Hong, J. (2012a). Pricing control for hybrid overlay/underlay spectrum access in cognitive radio networks. *APNOMS (2012).*
Do, C. T., Tran, N. H., Nguyen, M. V., Hong, C. S., & Lee, S. (2012). Social optimization strategy in unobserved queueing systems in cognitive radio networks. *IEEE Communications letters, 16*, 1944–1947.
Economou, A. (2021). The impact of information structure on strategic behavior in queueing systems. Mathematics and StatisticsIn V. Anisimov & N. Limnios (Eds.), *Queueing theory 2, advanced trends*. London: Sciences, ISTE & Wiley.
Economou, A., & Kanta, S. (2008). Equilibrium balking strategies in the observable single-server queue with breakdowns and repairs. *Operations Research Letters, 36*, 696–699.
Economou, A., Gómez-Corral, A., & Kanta, S. (2011). Optimal balking strategies in single-server queues with general service and vacation times. *Performance Evaluation, 68*, 967–982.
Economou, A., Logothetis, D., & Manou, A. (2022). The value of reneging for strategic customers in queueing systems with server vacations/failures. *European Journal of Operational Research, 299*, 960–976.
Economou, A., & Manou, A. (2013). Equilibrium balking strategies for a clearing queueing system in alternating environment. *Annals of Operations Research, 208*, 489–514.
Economou, A., & Manou, A. (2016). Strategic behavior in an observable fluid queue with alternating service process. *European Journal of Operational Research, 254*, 148–160.
Edelson, N. M., & Hildebrand, K. (1975). Congestion tolls for Poisson queueing processes. *Econometrica, 43*, 81–92.
Gautam, N. (2012). *Analysis of queues: methods and applications*. Boca Raton: CRC Press.
Guo, P., & Hassin, R. (2011). Strategic behavior and social optimization in Markovian vacation queues. *Operations Research, 59*, 986–997.
Guo, P., & Hassin, R. (2012). Strategic behavior and social optimization in Markovian vacation queues: the case of heterogeneous customers. *European Journal of Operational Research, 222*, 278–286.
Guo, P., & Li, Q. (2013). Strategic behavior and social optimization in partially-observable Markovian vacation queues. *Operations Research Letters, 41*, 277–284.
Guo, P., & Zhang, Z. G. (2013). Strategic queueing behavior and its impact on system performance in service systems with the congestionbased staffing policy. *Manufacturing & Service Operations Management, 15*, 118–131.
Hassin, R. (2016). *Rational queueing*. Boca Raton: CRC Press, Taylor and Francis Group.
Hassin, R., & Haviv, M. (2003). *To queue or not to queue: equilibrium behavior in queueing systems*. Boston: Kluwer Academic Publishers.
Haviv, M. (2013). When to arrive at a queue with tardiness costs? *Performance Evaluation, 70*(6), 387–399.
Huang, P., Wang, J., & Fu, L. (2012). Equilibrium balking strategies in the single-server queues with Erlangian service and setup times. *LISS*.
Jagannathan, K., Menache, I., Modiano, E., & Zussman, G. (2012). Non-cooperative spectrum access: the dedicated vs. free spectrum choice. *IEEE Journal on Selected Areas in Communications, 30*, 2251–2261.
Jain, R., Juneja, S., & Shimkin, N. (2011). The concert queueing game: to wait or to be late? *Discrete Event Dynamic Systems, 21*, 103–138.
Juneja, S., & Shimkin, N. (2013). The concert queueing game: strategic arrivals with waiting and tardiness costs. *Queueing Systems, 74*(4), 369–402.

Kosten, L. (1974). Stochastic theory of multi-entry buffer, part 1. *Delft Progress Report, Series F, 1*, 10–18.
Kosten, L. (1974). Stochastic theory of multi-entry buffer, part 2. *Delft Progress Report, Series F, 1*, 44–55.
Kosten, L., & Vrieze, O. J. (1975). Stochastic theory of multi-entry buffer, part 3. *Delft Progress Report, Series F, 1*, 103–115.
Kulkarni, V. G. (1997). Fluid models for single buffer systems. In J. H. Dshalalow (Ed.), *Frontiers in queueing: models and applications in science and engineering* (pp. 321–338). CRC Press.
Li, H., & Han, Z. (2011). Socially optimal queuing control in cognitive radio networks subject to service interruptions: to queue or not to queue? *IEEE Transactions on Wireless Communications, 10*, 1656–1666.
Li, X., Wang, J., & Zhang, F. (2014). New results on equilibrium balking strategies in the single-server queue with breakdowns and repairs. *Applied Mathematics and Computation, 241*, 380–388.
Liu, J., Xu, X., Wang, S., & Yue, D. (2021). Equilibrium analysis of the fluid model with two types of parallel customers and breakdowns. *Communications in Statistics-Theory and Methods, 50*(24), 5792–5805.
Liu, Z., Ma, Y., & Zhang, Z. G. (2015). Equilibrium mixed strategies in a discrete-time Markovian queue under multiple and single vacation policies. *Quality Technology & Quantitative Management, 12*, 367–380.
Ma, Y., Liu, W., & Li, J. (2013). Equilibrium balking behavior in the Geo/Geo/1 queueing system with multiple vacations. *Applied Mathematical Modelling, 37*, 3861–3878.
Maglaras, C. (2006). Revenue management for a multiclass single-server queue via a fluid model analysis. *Operations Research, 54*(5), 914–932.
Naor, P. (1969). The regulation of queue size by levying tolls. *Econometrica, 37*, 15–24.
Schwartz, M. (1996). *Broadband integrated networks*. Prentice-Hall, Inc.
Stidham, S., Jr. (2009). *Optimal design of queueing systems*. Boca Raton: CRC Press, Taylor and Francis Group.
Stidham, S., Rajagopal, S., & Kulkarni, V. G. (1995). Optimal flow control of a stochastic fluid-flow system. *IEEE Journal on Selected Areas in Communications, 13*(7), 1219–1228.
Sun, W., Guo, P., & Tian, N. (2010). Equilibrium threshold strategies in observable queueing systems with setup/closedown times. *Central European Journal of Operations Research, 18*, 241–268.
Sun, W., & Li, S. (2014). Equilibrium and optimal behavior of customers in Markovian queues with multiple working vacations. *TOP, 22*, 694–715.
Sun, W., Li, S., & Cheng-Guo, E. (2016). Equilibrium and optimal balking strategies of customers in Markovian queues with multiple vacations and N-policy. *Applied Mathematical Modelling, 40*, 284–301.
Tian, R., Yue, D., & Yue, W. (2015). Optimal balking strategies in an M/G/1 queueing system with a removal server under N-policy. *Journal of Industrial and Management Optimization, 11*, 715–731.
Wang, S., & Xu, X. (2021). Equilibrium strategies of the fluid queue with working vacation. *Operational Research, 21*(2), 1211–1228.
Wang, J., & Zhang, F. (2011). Equilibrium analysis of the observable queues with balking and delayed repairs. *Applied Mathematics and Computation, 218*, 2716–2729.
Yang, B., Hou, Z., Wu, J. Liu, Z. (2014a). Analysis the equilibrium strategies in the Geo/Geo/1 queue with multiple working vacations.*Preprint*.
Yang, T., Wang, J., & Zhang, F. (2014). Equilibrium balking strategies in the Geo/Geo/1 queues with server breakdowns and repairs. *Quality Technology & Quantitative Management, 11*, 231–243.
Yechiali, U. (1971). On optimal balking rules and toll charges in the GI/M/1 queuing process. *Operations Research, 19*(2), 349–370.
Yechiali, U. (1972). Customers' optimal joining rules for the GI/M/s queue. *Management Science, 18*(7), 434–443.
Yechiali, U. (1973). A queuing-type birth-and-death process defined on a continuous-time Markov chain. *Operations Research, 21*(2), 604–609.
Yechiali, U., & Naor, P. (1971). Queuing problems with heterogeneous arrivals and service. *Operations Research, 19*(3), 722–734.
Zhang, F., Wang, J., & Liu, B. (2013). Equilibrium balking strategies in Markovian queues with working vacations. *Applied Mathematical Modelling, 37*, 8264–8282.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.