Probability in the Engineering and Informational Sciences (2024), 1–20
doi:10.1017/S0269964823000244

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Equilibrium analysis of the fluid model with two types of parallel customers and incomplete fault

Yitong Zhang[1], Xiuli Xu[2], Pei Zhao[2] and Mingxin Liu[3]

[1]School of Economics and Management, Yanshan University, Qinhuangdao, China
[2]School of Science, Yanshan University, Qinhuangdao, China
[3]School of Electronics and Information Engineering, Guangdong Ocean University, Zhanjiang, China
Corresponding author: Xiuli Xu; Email: xxl-ysu@163.com

Keywords: entrance fee; equilibrium strategy; fluid queue; incomplete fault; parallel customers

MSC: primary 60K25; 90B22

Abstract
This article considers the individual equilibrium behavior and socially optimal strategy in a fluid queue with two types of parallel customers and incomplete fault. Assume that the working state and the incomplete fault state appear alternately in the buffer. Different from the linear revenue and expenditure structure, an exponential utility function can be constructed to obtain the equilibrium balking thresholds in the fully observable case. Besides, the steady-state probability distribution and the corresponding expected social benefit are derived based on the renewal process and the standard theory of linear ordinary differential equations. Furthermore, a reasonable entrance fee strategy is discussed under the condition that the fluid accepts the globally optimal strategies. Finally, the effects of the diverse system parameters on the entrance fee and the expected social benefit are explicitly illustrated by numerical comparisons.

1. Introduction

The customer behavior analysis of queueing systems has received considerable attention due to the widespread application of random service systems. Most references assume that the server is always available, but perfectly reliable servers are virtually nonexistent. Economou and Kanta [3] considered the equilibrium behavior in the observable queues with breakdowns and repairs. Subsequently, Li *et al.* [7] extended the research results in [3] and derived the joining strategies for customers in unobservable cases. Wang and Zhang [10] studied an unreliable M/M/1 queue with delayed repairs in observable cases. Inspired by Wang and Zhang [10], Yu *et al.* [15] explored the same settings in unobservable cases. Besides, Yang *et al.* [14] analyzed the Geo/Geo/1 queue with fault characteristics and determined the Nash equilibrium and the social optimum. Chen and Zhou [2] incorporated the setup times into a repairable queue and derived the mixed strategies of customers. Moreover, Li *et al.* [5] obtained the equilibrium balking strategies and socially optimal behavior in an M/M/1 queue with partial breakdowns and immediate repairs, where the systems can continue the service at a lower rate during the breakdowns. Yu *et al.* [16] compensated the game analysis in [5] by studying the corresponding partially observable cases. Xu and Xu [12] further discussed the individual Nash equilibrium in an M/M/1 queue with partial failures and delayed repairs. Considering more complex service systems, Zhang and Xu [18] made an economic analysis in a queue with setup times and partial failures. Aghsami and Jolai [1] studied the equilibrium balking strategies in the single-server Markovian queue with partial breakdowns and interruptible setup policy. The economic analysis of queueing systems with breakdowns has been widely applied in the inventory management, communication networks, and supply-chain systems.

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0269964823000244 Published online by Cambridge University Press

In a complex queueing system, the interarrival time of customers is getting smaller, and the service time is getting shorter. Discrete customers can be considered as the continuous fluid. There is an emerging tendency to study the fluid models strategically due to the limitation and the complexity of discrete queueing systems. Economou and Manou [4] considered the fluid model with two service modes and derived the individual equilibrium thresholds and socially optimal strategies. Subsequently, Wang and Xu [11] studied the strategic behavior in a fluid vacation queue. Recently, Xu and Wang [13] explored an on-off fluid queue and characterized the mixed equilibrium strategies in the unobservable cases. Logothetis *et al.* [9] also analyzed a fluid on-off model and studied the equilibrium strategies of customers under conditions of reneging and no-reneging, respectively. However, none of the above literature considers the entrance fee problems, which can eliminate the gap between the individual equilibrium and socially optimal strategies.

In real life, parallel customers who have the same service levels are widespread in public services, such as volunteer clinics for specific populations, psychological counseling, and legal advice. Li *et al.* [6] analyzed the service policy in an M/M/1 queue with two types of customers and different waiting costs. Zhang and Xu [17] incorporated two types of parallel customers into a fault queueing system and obtained the equilibrium balking strategy in the observable cases. Based on Zhang and Xu [17], Liu *et al.* [8] studied a repairable fluid queue with two types of parallel customers from an economic viewpoint. However, in many real-life situations, the server may not stop working completely in the maintenance period due to the durability and complexity of modern operating systems and system failures provide lower service rates. For example, a running computer system may slow down due to aging hardware or virus attacks. Therefore, incorporating the incomplete fault schedule into the fluid queue is reasonable and valuable, and this work compensates the equilibrium game analysis in [8] by studying the incomplete breakdowns.

The proposed model has broad applications and contributions in many fields, one of which is wireless communication technology. With the rapid development of telecommunication networks, wireless resources are shared among various services, and suppliers divide the information flowing into the processor into data, voice, and video categories to provide communication services. The unreliable router can be interrupted when transmitting information due to hardware failures and needs a random amount of time for recovery. During the maintenance periods, the router reduces the efficiency and can still transmit the information fluid at a lower rate. In this scenario, the transmitted information, the router, and the breakdown correspond to the arriving fluids, the buffer, and the incomplete fault, respectively, in the queueing terminology. The single service and pricing standard in the queueing system cannot satisfy the requirements of the growing variety of scenarios. Therefore, considering the equilibrium behavior of two or even more types of customers in an unreliable fluid queue is significant and necessary.

Motivated and inspired by the above practical cases, this paper makes an economic analysis for two types of the fluid in an unreliable server under the fully observable case. Different from [8], this paper assumes that the service can provide relatively low service rates after failure. To the best of the author's knowledge, no research has been found in the literature on conducting equilibrium strategies for fluid queues with incomplete fault and parallel customers. The individual equilibrium strategy can be obtained based on the non-cooperative game theory, and an exponential utility function is constructed to compute the expected social benefit. The reasonable entrance fee strategy can dynamically regulate system parameters and optimize the expected social benefits. Besides, the sensitivity analysis between the globally optimal thresholds and the expected social benefit is illustrated by several numerical examples. The research results provide a theoretical basis and valid suggestions for signal transmissions in network systems, which can obtain more stable response performance in the practical production life.

This article is organized as follows. Section 2 gives a detailed description of the fluid model. Sections 3 and 4 are devoted to the individual balking strategy and social equilibrium analysis in the fully observable case, respectively. Furthermore, the effects of the expected social benefit on globally optimal thresholds are illustrated by numerical examples in Section 4. Section 5 presents an entrance fee policy and analyzes it theoretically and numerically. Finally, we briefly conclude the paper in Section 6.

2. Model description

Assume that two types of parallel customers flow into the buffer according to the exponential distribution with rates λ_1 and λ_2 ($\lambda_1 < \lambda_2$), respectively, and the arrival processes are independent of each other. $I(t)$ is the state of the buffer at time t, and $I(t) = 0, 1$ indicates that the buffer is in maintenance period and normal working period, respectively. The buffer has an exponentially distributed lifetime with parameter θ and the output rate of the buffer in the normal working state is μ_1. When the system breaks down, the buffer immediately enters the maintenance state and provides services for the fluid at a lower rate of μ_0 ($\mu_0 < \mu_1$). The maintenance time follows the exponential distribution with parameter ξ. Let $X(t)$ be the fluid level at time t, and the net input rate of the buffer can be expressed as:

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t} = \begin{cases} \lambda_1 + \lambda_2 - \mu_1, & I(t) = 1, X(t) > 0, \\ \lambda_1 + \lambda_2 - \mu_0, & I(t) = 0, X(t) > 0, \\ \max\{\lambda_1 + \lambda_2 - \mu_1, 0\}, & I(t) = 1, X(t) = 0. \end{cases}$$

The condition $\lambda_j > \mu_0, j = 1, 2$ ensures that the buffer can reach the steady-state.

Suppose that the unit of the type $j (j = 1, 2)$ fluid receives a reward of R_j when it flows out of the buffer and costs C_j per unit time during its sojourn time, and the inequality $R_j > \frac{C_j}{\mu_0}$ is satisfied hereinafter to ensure that the fluid prefers to enter when the buffer is empty.

In this paper, the fluid level $X(t)$ and the buffer's state $I(t)$ can be observed when the fluid arrives the buffer, and the decisions are irreversible, that is, neither retrials of balking customers or reneging of entering customers are permitted. When the fluid reaches the buffer at time t, we assume that the observed buffer state is $(X(t), I(t)) = (x, i)$.

3. Individual equilibrium strategy

In the financial and insurance industry, the exponential utility function is more practical and significant than the linear utility function. The expected net benefit per unit time for the type j fluid after its service when the buffer is in state (x, i) can be denoted as:

$$B_j(x, i) = R_j - C_j E\left[\mathrm{e}^{\alpha S_j(x,i)}\right],$$

where $S_j(x, i), j = 1, 2$ represents the expected sojourn time for the type j fluid. $\mathrm{e}^{\alpha S_j(x,i)}$ is integrable and α is a constant, and $0 < \alpha < \theta, 0 < \alpha < \xi$.

The pre-condition of the fluid flowing into the buffer is that the expected net benefit $B_j(x, i)$ is positive, then we get the following result.

Theorem 3.1. *In the fully observable case, there exists a pair of thresholds* $\left(x^*{}_j(0), x^*{}_j(1)\right)$ *for type* $j\,(j = 1, 2)$ *customers in the fluid model with two types of parallel customers and incomplete fault, and the equilibrium strategies* $x^*{}_j(0)$ *and* $x^*{}_j(1)$*, respectively, are the unique roots of the equations described below.*

$$R_j - C_j \frac{[\alpha(\mu_0 - \mu_1) - \theta\mu_0]\,\mathrm{e}^{\frac{\alpha-\xi}{\mu_0}x} + \xi\mu_1 \mathrm{e}^{\frac{\alpha-\theta}{\mu_1}x}}{\alpha(\mu_0 - \mu_1) - \theta\mu_0 + \xi\mu_1} = 0,$$

$$R_j - C_j \frac{[\alpha(\mu_0 - \mu_1) + \xi\mu_1]\,\mathrm{e}^{\frac{\alpha-\theta}{\mu_1}x} + \theta\mu_0 \mathrm{e}^{\frac{\alpha-\xi}{\mu_0}x}}{\alpha(\mu_0 - \mu_1) - \theta\mu_0 + \xi\mu_1} = 0.$$

The equilibrium strategy of the fluid is defined as "while arriving at time t, observe the system is in state $(X(t), I(t))$*, the type j customers enter if* $X(t) \le x^*{}_j(I(t))$ *and balk otherwise."*

https://doi.org/10.1017/S0269964823000244 Published online by Cambridge University Press

Proof. Assume that T_i is the remaining sojourn time of the buffer in the state $i, i = 0, 1$ when the fluid arrives, then

$$S_j(x,i) = \begin{cases} \frac{x}{\mu_i}, & T_i \geq \frac{x}{\mu_i}, \\ T_i + S_j(x - \mu T_i, 1-i), & T_i < \frac{x}{\mu_i}. \end{cases} \tag{1}$$

Using the conditional expectation formula, we have:

$$E\left(e^{\alpha S_j(x,0)}\right) = e^{(\alpha-\xi)\frac{x}{\mu_0}} + \int_0^{\frac{x}{\mu_0}} E\left(e^{\alpha S_j(x-\mu_0 t,1)}\right) \xi e^{(\alpha-\xi)t} dt,$$

$$E\left(e^{\alpha S_j(x,1)}\right) = e^{(\alpha-\theta)\frac{x}{\mu_1}} + \int_0^{\frac{x}{\mu_1}} E\left(e^{\alpha S_j(x-\mu_1 t,0)}\right) \theta e^{(\alpha-\theta)t} dt.$$

After manipulating, we obtain the differential equations,

$$\frac{\xi-\alpha}{\mu_0} E\left(e^{\alpha S_j(x,0)}\right) + \frac{dE(e^{\alpha S_j(x,0)})}{dx} = \frac{\xi}{\mu_0} E\left(e^{\alpha S_j(x,1)}\right), \tag{2}$$

$$\frac{\theta-\alpha}{\mu_1} E\left(e^{\alpha S_j(x,1)}\right) + \frac{dE(e^{\alpha S_j(x,1)})}{dx} = \frac{\theta}{\mu_1} E\left(e^{\alpha S_j(x,0)}\right), \tag{3}$$

with the boundary condition $E\left(e^{\alpha S_j(0,i)}\right) = 1$.

From formulas (2) and (3), the exponential form of the expected sojourn time in the buffer can be obtained as:

$$E(e^{\alpha S_j(x,0)}) = \frac{[\alpha(\mu_0-\mu_1) - \theta\mu_0]\, e^{\frac{\alpha-\xi}{\mu_0}x} + \xi\mu_1 e^{\frac{\alpha-\theta}{\mu_1}x}}{\alpha(\mu_0-\mu_1) - \theta\mu_0 + \xi\mu_1}, i = 0,$$

$$E(e^{\alpha S_j(x,1)}) = \frac{[\alpha(\mu_0-\mu_1) + \xi\mu_1]\, e^{\frac{\alpha-\theta}{\mu_1}x} + \theta\mu_0 e^{\frac{\alpha-\xi}{\mu_0}x}}{\alpha(\mu_0-\mu_1) - \theta\mu_0 + \xi\mu_1}, i = 1.$$

Obviously, $E(e^{\alpha S_j(x,i)}), i = 0, 1$ is a monotonically increasing function with respect to x, then the unique roots $x^*{}_j(0)$ and $x^*{}_j(1)$ for type j ($j = 1, 2$) customers are given in Theorem 3.1. □

4. Social equilibrium analysis

This section studies the steady-state probability distribution of the fluid level and the socially optimal strategy in a global optimization problem. The fluid considers its own interests first when encountering a joining-balking problem, then the buffer will be over-crowded and cannot achieve the global optimum. Assume that all customers follow the socially optimal threshold strategy $\left(x_{ej}(0), x_{ej}(1)\right), j = 1, 2$, and the fluid flows into the buffer if $X(t) < x_{ej}(i)$ when the system is in state (x, i), otherwise the fluid balks. Besides, the fluid prefers to enter when the buffer is in a normal working period because the expected sojourn time of the fluid in state 1 is always less than the expected sojourn time in state 0 with the same fluid level. Therefore, the inequalities $x_{e1}(0) < x_{e1}(1)$ and $x_{e2}(0) < x_{e2}(1)$ can be obtained. Then, we suppose that the thresholds satisfy $x_{e1}(0) < x_{e2}(0) < x_{e1}(1) < x_{e2}(1)$, and the others can be calculated based on the similar calculation method.

https://doi.org/10.1017/S0269964823000244 Published online by Cambridge University Press

The overall expected social benefit in unit time is:

$$B\left(x_{e1}\left(0\right),x_{e2}\left(0\right),x_{e1}\left(1\right),x_{e2}\left(1\right)\right)=\sum_{j=1}^{2}B_j(x_{ej}(0),x_{ej}(1))=\sum_{j=1}^{2}\left(\lambda_{ej}R_j-C_jE\left(\mathrm{e}^{\alpha X_j}\right)\right),$$

where $B_j(x_{ej}(0),x_{ej}(1))$ and λ_{ej} are the expected social benefit per unit time and the effective arrival rate for type j customers, respectively. The effective arrival rate λ_{ej} can be calculated from the arrival rate and the entrance probability, and X_j is a random variable that denotes the fluid level of type j customers.

4.1. Steady-state probability distribution

The stationary probability distribution of the fluid level in state i can be defined as:

$$F_i(x)=\lim_{t\to\infty}F_i(t,x)=p\{X(t)\le x,I(t)=i\},\quad x\ge 0,i=0,1.$$

According to the alternating renewal process, we can get the steady-state distribution of the buffer is $\pi_0=\frac{\theta}{\theta+\xi},\pi_1=\frac{\xi}{\theta+\xi}$.

Theorem 4.1. *In the fully observable fluid queue with two types of parallel customers and incomplete fault, if all customers follow the threshold strategy* $\left(x_{ej}\left(0\right),x_{ej}\left(1\right)\right)$ *and* $x_{e1}\left(0\right)<x_{e2}\left(0\right)<x_{e1}\left(1\right)<x_{e2}\left(1\right)$*, the steady-state probability distribution of the fluid level is as follows.*

Case I. $\lambda_2>\mu_1$.

$$F_0(x)=\begin{cases}0, & x<x_{e2}(0),\\ \pi_0\dfrac{\xi\left(q\lambda_1+\lambda_2-\mu_1\right)\mathrm{e}^{-\left(\frac{\theta}{q\lambda_1+\lambda_2-\mu_1}-\frac{\xi}{\mu_0}\right)\left(x-x_{e2}(0)\right)}-\theta\mu_0}{\xi\left(q\lambda_1+\lambda_2-\mu_1\right)\mathrm{e}^{-\left(\frac{\theta}{q\lambda_1+\lambda_2-\mu_1}-\frac{\xi}{\mu_0}\right)\left(x_{e2}(1)-x_{e2}(0)\right)}-\theta\mu_0}, & x_{e2}(0)\le x\le x_{e2}(1),\\ \pi_0, & x\ge x_{e2}(1).\end{cases}\tag{4}$$

$$F_1(x)=\begin{cases}0, & x\le x_{e2}(0),\\ \pi_1\dfrac{\theta\mu_0\left[\mathrm{e}^{-\left(\frac{\theta}{q\lambda_1+\lambda_2-\mu_1}-\frac{\xi}{\mu_0}\right)\left(x-x_{e2}(0)\right)}-1\right]}{\xi\left(q\lambda_1+\lambda_2-\mu_1\right)\mathrm{e}^{-\left(\frac{\theta}{q\lambda_1+\lambda_2-\mu_1}-\frac{\xi}{\mu_0}\right)\left(x_{e2}(1)-x_{e2}(0)\right)}-\theta\mu_0}, & x_{e2}(0)\le x\le x_{e2}(1),\\ \pi_1, & x>x_{e2}(1).\end{cases}\tag{5}$$

Case II. $\lambda_1+\lambda_2>\mu_1$ *and* $\lambda_2\le\mu_1$.

$$F_0(x)=\begin{cases}0, & x<x_{e2}(0),\\ \pi_0\dfrac{\xi\left(\lambda_1+\lambda_2-\mu_1\right)\mathrm{e}^{-\left(\frac{\theta}{\lambda_1+\lambda_2-\mu_1}-\frac{\xi}{\mu_0}\right)\left(x-x_{e2}(0)\right)}-\theta\mu_0}{\xi\left(\lambda_1+\lambda_2-\mu_1\right)\mathrm{e}^{-\left(\frac{\theta}{\lambda_1+\lambda_2-\mu_1}-\frac{\xi}{\mu_0}\right)\left(x_{e1}(1)-x_{e2}(0)\right)}-\theta\mu_0}, & x_{e2}(0)\le x\le x_{e1}(1),\\ \pi_0, & x\ge x_{e1}(1).\end{cases}\tag{6}$$

https://doi.org/10.1017/S0269964823000244 Published online by Cambridge University Press

$$F_1(x)=\begin{cases}0, & x\le x_{e2}(0),\\ \pi_1\dfrac{\theta\mu_0\left[\mathrm{e}^{-\left(\frac{\theta}{\lambda_1+\lambda_2-\mu_1}-\frac{\xi}{\mu_0}\right)(x-x_{e2}(0))}-1\right]}{\xi(\lambda_1+\lambda_2-\mu_1)\mathrm{e}^{-\left(\frac{\theta}{\lambda_1+\lambda_2-\mu_1}-\frac{\xi}{\mu_0}\right)(x_{e1}(1)-x_{e2}(0))}-\theta\mu_0}, & x_{e2}(0)\le x\le x_{e1}(1),\\ \pi_1, & x>x_{e1}(1).\end{cases} \tag{7}$$

Case III. *$\lambda_1+\lambda_2=\mu_1$. The fluid level is stabilized at $x_{e2}(0)$.*
Case IV. *$\mu_0<\lambda_1+\lambda_2<\mu_1$.*

$$F_0(x)=\begin{cases}0, & x\le 0,\\ \pi_0\dfrac{\xi(\lambda_1+\lambda_2-\mu_1)\left[1-\mathrm{e}^{-\left(\frac{\xi}{p\lambda_1+\lambda_2-\mu_0}+\frac{\theta}{\lambda_1+\lambda_2-\mu_1}\right)x}\right]}{\theta(p\lambda_1+\lambda_2-\mu_0)\mathrm{e}^{-\left(\frac{\xi}{p\lambda_1+\lambda_2-\mu_0}+\frac{\theta}{\lambda_1+\lambda_2-\mu_1}\right)x_{e2}(0)}+\xi(\lambda_1+\lambda_2-\mu_1)}, & 0\le x\le x_{e2}(0),\\ \pi_0, & x>x_{e2}(0).\end{cases} \tag{8}$$

$$F_1(x)=\begin{cases}0, & x<0,\\ \pi_1\dfrac{\xi(\lambda_1+\lambda_2-\mu_1)+(p\lambda_1+\lambda_2-\mu_0)\theta\mathrm{e}^{-\left(\frac{\xi}{p\lambda_1+\lambda_2-\mu_0}+\frac{\theta}{\lambda_1+\lambda_2-\mu_1}\right)x}}{\theta(p\lambda_1+\lambda_2-\mu_0)\mathrm{e}^{-\left(\frac{\xi}{p\lambda_1+\lambda_2-\mu_0}+\frac{\theta}{\lambda_1+\lambda_2-\mu_1}\right)x_{e2}(0)}+\xi(\lambda_1+\lambda_2-\mu_1)}, & 0\le x\le x_{e2}(0),\\ \pi_1, & x\ge x_{e2}(0).\end{cases} \tag{9}$$

Proof. By comparing the arrival rate and the outflow rate, the steady-state probability distribution of the fluid level can be discussed in the following four cases.

Case I. $\lambda_2>\mu_1$.

When the buffer is in state 1, the fluid in the buffer increases at the rate of $\lambda_1+\lambda_2-\mu_1$ until the fluid level reaches the threshold $x_{e1}(1)$. When the fluid level exceeds the optimal threshold $x_{e1}(1)$, only type 2 customers flow in and the fluid level increases to the threshold $x_{e2}(1)$ at the rate of $\lambda_2-\mu_1$. If the buffer is still in state 1, the fluid will enter the system with probability $\frac{\mu_1}{\lambda_2}$ to ensure that the fluid level is stabilized at $x_{e2}(1)$. When the buffer is in state 0, the fluid decreases at the rate of μ_0. The system repeats in this way, and the fluid level fluctuates within $[x_{e2}(0),x_{e2}(1)]$.

Considering the fluid level in a tiny time interval, we have

$$F_0(t+\Delta t,x)=F_0(t,x+\mu_0\Delta t)\mathrm{e}^{\xi\Delta t}+F_1(t,x-(q\lambda_1+\lambda_2-\mu_1)\Delta t)(1-\mathrm{e}^{\theta\Delta t}),$$
$$F_1(t+\Delta t,x)=F_1(t,x-(q\lambda_1+\lambda_2-\mu_1)\Delta t)\mathrm{e}^{\theta\Delta t}+F_0(t,x+\mu_0\Delta t)(1-\mathrm{e}^{\xi\Delta t}),$$

Substituting the exponential form of the Taylor formulas with Peano terms, we obtain:

$$F_0(t+\Delta t,x)=F_0(t,x+\mu_0\Delta t)(1-\xi\Delta t)+F_1(t,x-(q\lambda_1+\lambda_2-\mu_1)\Delta t)\theta\Delta t+o(\Delta t),$$
$$F_1(t+\Delta t,x)=F_1(t,x-(q\lambda_1+\lambda_2-\mu_1)\Delta t)(1-\theta\Delta t)+F_0(t,x+\mu_0\Delta t)\xi\Delta t+o(\Delta t),$$

Dividing both sides by Δt and making $\Delta t\longrightarrow 0$, then we can get:

$$\begin{cases}\dfrac{\partial F_0(t,x)}{\partial t}-\mu_0\dfrac{\partial F_0(t,x)}{\partial x}=-\xi F_0(t,x)+\theta F_1(t,x),\\ \dfrac{\partial F_1(t,x)}{\partial t}+(q\lambda_1+\lambda_2-\mu_1)\dfrac{\partial F_1(t,x)}{\partial x}=-\theta F_1(t,x)+\xi F_0(t,x),\end{cases}$$

https://doi.org/10.1017/S0269964823000244 Published online by Cambridge University Press

When the system is stable, there exists $\lim_{t\to\infty}\frac{\partial F_i(t,x)}{\partial t}=0, i=0,1.$

The ordinary differential equations can be constructed as:

$$\begin{cases} -\mu_0\dfrac{\mathrm{d}F_0(x)}{\mathrm{d}x}=-\xi F_0(x)+\theta F_1(x), \\ (q\lambda_1+\lambda_2-\mu_1)\dfrac{\mathrm{d}F_1(x)}{\mathrm{d}x}=\xi F_0(x)-\theta F_1(x), \end{cases} \tag{10}$$

with boundary conditions

$$F_0(x_{e2}(1))=\pi_0,\ F_1(x_{e2}(0))=0,$$

where $q=\frac{x_{e1}(1)-x_{e2}(0)}{x_{e2}(1)-x_{e2}(0)}, 0\le q\le 1$ is the probability that the fluid level in the system is less than the threshold $x_{e1}(1)$ in state 1.

The formulas (4) and (5) can be obtained by solving the differential equation (10). The point masses and the probability density functions in states 0 and 1, respectively, are:

$$P_0(x_{e2}(0))=F_0(x_{e2}(0)),\ P_1(x_{e2}(1))=\pi_1-F_1(x_{e2}(1)),$$

$$f_0(x)=\frac{\pi_0\xi(q\lambda_1+\lambda_2-\mu_1)\left(\frac{\xi}{\mu_0}-\frac{\theta}{q\lambda_1+\lambda_2-\mu_1}\right)\mathrm{e}^{-\left(\frac{\theta}{q\lambda_1+\lambda_2-\mu_1}-\frac{\xi}{\mu_0}\right)(x-x_{e2}(0))}}{\xi(q\lambda_1+\lambda_2-\mu_1)\mathrm{e}^{-\left(\frac{\theta}{q\lambda_1+\lambda_2-\mu_1}-\frac{\xi}{\mu_0}\right)(x_{e2}(1)-x_{e2}(0))}-\theta\mu_0},$$

$$f_1(x)=\frac{\pi_1\theta\mu_0\left(\frac{\xi}{\mu_0}-\frac{\theta}{q\lambda_1+\lambda_2-\mu_1}\right)\mathrm{e}^{-\left(\frac{\theta}{q\lambda_1+\lambda_2-\mu_1}-\frac{\xi}{\mu_0}\right)(x-x_{e2}(0))}}{\xi\,(q\lambda_1+\lambda_2-\mu_1)\,\mathrm{e}^{-\left(\frac{\theta}{q\lambda_1+\lambda_2-\mu_1}-\frac{\xi}{\mu_0}\right)(x_{e2}(1)-x_{e2}(0))}-\theta\mu_0}.$$

Case II. $\lambda_1+\lambda_2>\mu_1$ and $\lambda_2\le\mu_1$.

When the buffer is in state 1, the fluid in the buffer increases at the rate of $\lambda_1+\lambda_2-\mu_1$ until the fluid level reaches the threshold $x_{e1}(1)$, then the fluid level remains unaltered. When the buffer is in state 0, the fluid decreases at the rate of μ_0. The system repeats in this way, and the fluid level fluctuates within $[x_{e2}(0), x_{e1}(1)]$.

Considering the fluid level in a tiny time interval, we have:

$$F_0(t+\Delta t,x)=F_0(t,x+\mu_0\Delta t)\mathrm{e}^{\xi\Delta t}+F_1(t,x-(\lambda_1+\lambda_2-\mu_1)\,\Delta t)(1-\mathrm{e}^{\theta\Delta t}),$$
$$F_1(t+\Delta t,x)=F_1(t,x-(\lambda_1+\lambda_2-\mu_1)\,\Delta t)\mathrm{e}^{\theta\Delta t}+F_0(t,x+\mu_0\Delta t)(1-\mathrm{e}^{\xi\Delta t}),$$

Substituting the exponential form of the Taylor formulas with Peano terms, we obtain:

$$F_0(t+\Delta t,x)=F_0(t,x+\mu_0\Delta t)(1-\xi\Delta t)+F_1(t,x-(\lambda_1+\lambda_2-\mu_1)\,\Delta t)\theta\Delta t+o(\Delta t),$$
$$F_1(t+\Delta t,x)=F_1(t,x-(\lambda_1+\lambda_2-\mu_1)\,\Delta t)(1-\theta\Delta t)+F_0(t,x+\mu_0\Delta t)\xi\Delta t+o(\Delta t),$$

Dividing both sides by Δt and making $\Delta t\to 0$, then we can get:

$$\begin{cases} \dfrac{\partial F_0(t,x)}{\partial t}-\mu_0\dfrac{\partial F_0(t,x)}{\partial x}=-\xi F_0(t,x)+\theta F_1(t,x), \\ \dfrac{\partial F_1(t,x)}{\partial t}+(\lambda_1+\lambda_2-\mu_1)\dfrac{\partial F_1(t,x)}{\partial x}=-\theta F_1(t,x)+\xi F_0(t,x), \end{cases}$$

When the system is stable, there exists $\lim_{t\to\infty}\frac{\partial F_i(t,x)}{\partial t}=0, i=0,1.$

The ordinary differential equations can be constructed as:

$$\begin{cases} -\mu_0 \dfrac{\mathrm{d}F_0(x)}{\mathrm{d}x} = -\xi F_0(x) + \theta F_1(x), \\ (\lambda_1 + \lambda_2 - \mu_1)\,\dfrac{\mathrm{d}F_1(x)}{\mathrm{d}x} = \xi F_0(x) - \theta F_1(x), \end{cases} \tag{11}$$

with boundary conditions $F_0(x_{e1}(1)) = \pi_0$ and $F_1(x_{e2}(0)) = 0$.

The formulas (6) and (7) can be obtained by solving the differential equation (11). The point masses and the probability density functions in two states, respectively, are:

$$P_0(x_{e2}(0)) = F_0(x_{e2}(0)),\ P_1(x_{e1}(1)) = \pi_1 - F_1(x_{e1}(1)),$$

$$f_0(x) = \frac{\pi_0\xi(\lambda_1+\lambda_2-\mu_1)\left(\frac{\xi}{\mu_0} - \frac{\theta}{\lambda_1+\lambda_2-\mu_1}\right)\mathrm{e}^{-\left(\frac{\theta}{\lambda_1+\lambda_2-\mu_1}-\frac{\xi}{\mu_0}\right)(x-x_{e2}(0))}}{\xi(\lambda_1+\lambda_2-\mu_1)\mathrm{e}^{-\left(\frac{\theta}{\lambda_1+\lambda_2-\mu_1}-\frac{\xi}{\mu_0}\right)(x_{e1}(1)-x_{e2}(0))} - \theta\mu_0},$$

$$f_1(x) = \frac{\pi_1\theta\mu_0\left(\frac{\xi}{\mu_0} - \frac{\theta}{\lambda_1+\lambda_2-\mu_1}\right)\mathrm{e}^{-\left(\frac{\theta}{\lambda_1+\lambda_2-\mu_1}-\frac{\xi}{\mu_0}\right)(x-x_{e2}(0))}}{\xi\,(\lambda_1+\lambda_2-\mu_1)\,\mathrm{e}^{-\left(\frac{\theta}{\lambda_1+\lambda_2-\mu_1}-\frac{\xi}{\mu_0}\right)(x_{e1}(1)-x_{e2}(0))} - \theta\mu_0}.$$

Case III. $\lambda_1{+}\lambda_2{=}\mu_1$.

When the buffer is in state 0, the net input rate of the fluid is $\lambda_1{+}\lambda_2 - \mu_0$ or $\lambda_2 - \mu_0$ and the fluid level increases to x_{e2} (0), then the fluid level remains unaltered. When the buffer is in state 1, the fluid level remains constant. The fluid level eventually stabilizes at x_{e2} (0) with the alternation of the server status.

Case IV. $\mu_0 < \lambda_1 + \lambda_2 < \mu_1$.

When the buffer is in state 0, the fluid in the system increases at the rate of $\lambda_1 + \lambda_2 - \mu_0$ until the fluid level reaches the threshold $x_{e1}(0)$. When the fluid level exceeds the optimal threshold $x_{e1}(0)$, only type 2 customers flow in and the fluid level increases to $x_{e2}(0)$ at the rate of $\lambda_2 - \mu_0$. If the buffer is still in state 0, the fluid will enter the system with probability $\frac{\mu_0}{\lambda_2}$ to ensure that the fluid level is stabilized at $x_{e2}(0)$. When the buffer is in state 1, the fluid decreases at the rate of $\lambda_1 + \lambda_2 - \mu_1$. The system repeats in this way, and the fluid level fluctuates within $[0, x_{e2}(0)]$.

Considering the fluid level in a tiny time interval, we have:

$$F_0(t+\Delta t, x) = F_0(t, x - (p\lambda_1 + \lambda_2 - \mu_0)\,\Delta t)\mathrm{e}^{\xi\Delta t} + F_1(t, x - (\lambda_1 + \lambda_2 - \mu_1)\,\Delta t)(1 - \mathrm{e}^{\theta\Delta t}),$$
$$F_1(t+\Delta t, x) = F_1(t, x - (\lambda_1 + \lambda_2 - \mu_1)\,\Delta t)\mathrm{e}^{\theta\Delta t} + F_0(t, x - (p\lambda_1 + \lambda_2 - \mu_0)\,\Delta t)(1 - \mathrm{e}^{\xi\Delta t}),$$

Substituting the exponential form of the Taylor formulas with Peano terms, we obtain:

$$\begin{aligned} F_0(t+\Delta t, x) =& F_0(t, x - (p\lambda_1 + \lambda_2 - \mu_0)\,\Delta t)(1 - \xi\Delta t) \\ &+ F_1(t, x - (\lambda_1 + \lambda_2 - \mu_1)\,\Delta t)\theta\Delta t + o(\Delta t), \\ F_1(t+\Delta t, x) =& F_1(t, x - (\lambda_1 + \lambda_2 - \mu_1)\,\Delta t)(1 - \theta\Delta t) \\ &+ F_0(t, x - (p\lambda_1 + \lambda_2 - \mu_0)\,\Delta t)\xi\Delta t + o(\Delta t), \end{aligned}$$

https://doi.org/10.1017/S0269964823000244 Published online by Cambridge University Press

Dividing both sides by Δt and making $\Delta t \longrightarrow 0$, then we can get:

$$\begin{cases} \dfrac{\partial F_0(t,x)}{\partial t} + (p\lambda_1 + \lambda_2 - \mu_0)\dfrac{\partial F_0(t,x)}{\partial x} = -\xi F_0(t,x) + \theta F_1(t,x), \\ \dfrac{\partial F_1(t,x)}{\partial t} + (\lambda_1 + \lambda_2 - \mu_1)\dfrac{\partial F_1(t,x)}{\partial x} = -\theta F_1(t,x) + \xi F_0(t,x), \end{cases}$$

When the system is stable, there exists $\lim\limits_{t\longrightarrow\infty} \frac{\partial F_i(t,x)}{\partial t} = 0, i = 0, 1.$

The ordinary differential equations can be constructed as:

$$\begin{cases} (p\lambda_1 + \lambda_2 - \mu_0)\dfrac{\mathrm{d}F_0(x)}{\mathrm{d}x} = \theta F_1(x) - \xi F_0(x), \\ (\lambda_1 + \lambda_2 - \mu_1)\dfrac{\mathrm{d}F_1(x)}{\mathrm{d}x} = \xi F_0(x) - \theta F_1(x), \end{cases} \tag{12}$$

with boundary conditions $F_0(0) = 0$ and $F_1(x_{e2}(0)) = \pi_1$, where $p = \frac{x_{e1}(0)}{x_{e2}(0)}, 0 \le p \le 1$ is the probability that the fluid level in the system is less than the threshold $x_{e1}(0)$ in state 0.

The formulas (8) and (9) can be obtained by calculating equation (12). The point masses and the probability density functions in two states, respectively, are:

$$P_0(x_{e2}(0)) = \pi_0 - F_0(x_{e2}(0)),\ P_1(0) = F_1(0),$$

$$f_0(x) = \frac{\pi_0\xi\,(\mu_1 - \lambda_1 - \lambda_2)\left(\frac{\theta}{\mu_1-\lambda_1-\lambda_2} - \frac{\xi}{p\lambda_1+\lambda_2-\mu_0}\right)\mathrm{e}^{-\left(\frac{\xi}{p\lambda_1+\lambda_2-\mu_0}+\frac{\theta}{\lambda_1+\lambda_2-\mu_1}\right)x}}{\theta\,(p\lambda_1+\lambda_2-\mu_0)\,\mathrm{e}^{-\left(\frac{\xi}{p\lambda_1+\lambda_2-\mu_0}+\frac{\theta}{\lambda_1+\lambda_2-\mu_1}\right)x_{e2}(0)} + \xi\,(\lambda_1+\lambda_2-\mu_1)},$$

$$f_1(x) = \frac{\pi_1(p\lambda_1 + \lambda_2 - \mu_0)\left(\frac{\theta}{\mu_1-\lambda_1-\lambda_2} - \frac{\xi}{p\lambda_1+\lambda_2-\mu_0}\right)\theta\mathrm{e}^{-\left(\frac{\xi}{p\lambda_1+\lambda_2-\mu_0}+\frac{\theta}{\lambda_1+\lambda_2-\mu_1}\right)x}}{\theta(p\lambda_1+\lambda_2-\mu_0)\mathrm{e}^{-\left(\frac{\xi}{p\lambda_1+\lambda_2-\mu_0}+\frac{\theta}{\lambda_1+\lambda_2-\mu_1}\right)x_{e2}(0)} + \xi(\lambda_1+\lambda_2-\mu_1)}.$$

□

4.2. Analysis of the expected social benefit

Based on Theorem 4.1, when all customers follow the optimal balking strategies $(x_{e1}(0), x_{e2}(0), x_{e1}(1), x_{e2}(1))$, the utility functions of the expected social benefit per unit time are as follows.

Case I. $\lambda_2 > \mu_1$.

In this case, the fluid level fluctuates within $[x_{e2}(0), x_{e2}(1)]$. The type 1 customers refuse to flow in when the fluid level reaches the socially optimal thresholds $x_{e1}(i), i = 0, 1$. The entrance probabilities of type 2 customers are $\frac{\mu_0}{\lambda_2}$ and $\frac{\mu_1}{\lambda_2}$ when the fluid levels satisfy $x = x_{e2}(0)$ and $x = x_{e2}(1)$, respectively. Based on Theorem 4.1, the stationary probability distribution $F_1(x)$ is continuous when the fluid level reaches the thresholds $x_{e2}(0)$ and $x_{e1}(1)$, then the transient probabilities are $P_1(x_{e2}(0)) = 0$ and $P_1(x_{e1}(1)) = 0$.

The expected social benefit per unit time can be expressed as:

$$B\,(x_{e1}(0), x_{e2}(0), x_{e1}(1), x_{e2}(1)) = \lambda_{e1}R_1 + \lambda_{e2}R_2 - \left[C_1E\left(\mathrm{e}^{\alpha X_1}\right) + C_2E\left(\mathrm{e}^{\alpha X_2}\right)\right],$$

https://doi.org/10.1017/S0269964823000244 Published online by Cambridge University Press

where the effective arrival rate and the exponential expected fluid level, respectively, are:

$$\lambda_{e1} = \lambda_1 \int_{x_{e2}(0)}^{x_{e1}(1)} f_1(x)\mathrm{d}x = \lambda_1 \frac{\pi_1\mu_0\theta \mathrm{e}^{-A_1(x_{e1}(1)-2x_{e2}(0))}}{\xi\left(q\lambda_1+\lambda_2-\mu_1\right)\mathrm{e}^{-A_1(x_{e2}(1)-x_{e2}(0))}-\theta\mu_0},$$

$$\begin{aligned}\lambda_{e2} &= \lambda_2\left[P_0(x_{e2}(0))\frac{\mu_0}{\lambda_2}+P_1(x_{e2}(1))\frac{\mu_1}{\lambda_2}+\int_{x_{e2}(0)}^{x_{e2}(1)} f_1(x)\mathrm{d}x\right]\\ &= \lambda_2\frac{\pi_1\mu_0\theta\mathrm{e}^{-A_1(x_{e2}(1)-2x_{e2}(0))}+\pi_0B_1\frac{\mu_0}{\lambda_2}+\pi_1B_1\mathrm{e}^{-A_1(x_{e2}(1)-x_{e2}(0))}\frac{\mu_1}{\lambda_2}}{\xi\left(q\lambda_1+\lambda_2-\mu_1\right)\mathrm{e}^{-A_1(x_{e2}(1)-x_{e2}(0))}-\theta\mu_0},\end{aligned}$$

$$\begin{aligned}E(\mathrm{e}^{\alpha X_1}) &= \int_{x_{e2}(0)}^{x_{e1}(1)} f_1(x)\mathrm{e}^{\alpha\frac{\lambda_1}{\lambda_1+\lambda_2}x}\mathrm{d}x+\int_{x_{e2}(0)}^{x_{e2}(1)} f_0(x)\mathrm{e}^{\alpha\frac{\lambda_1}{\lambda_1+\lambda_2}x}\mathrm{d}x\\ &= \frac{\pi_1\theta(q\lambda_1+\lambda_2-\mu_1+\mu_0)A_1\left[\mathrm{e}^{-A_1(x_{e1}(1)-2x_{e2}(0))}+\mathrm{e}^{-A_1(x_{e2}(1)-2x_{e2}(0))}\right]}{\left[\xi\left(q\lambda_1+\lambda_2-\mu_1\right)\mathrm{e}^{-A_1(x_{e2}(1)-x_{e2}(0))}-\theta\mu_0\right](A_1-E_1)},\end{aligned}$$

$$\begin{aligned}E\left(\mathrm{e}^{\alpha X_2}\right) &= \int_{x_{e2}(0)}^{x_{e1}(1)} f_1(x)\mathrm{e}^{D_1x}\mathrm{d}x+\int_{x_{e1}(1)}^{x_{e2}(1)} f_1(x)\mathrm{e}^{\alpha x}\mathrm{d}x+P_1(x_{e2}(1))\mathrm{e}^{\alpha\left[x_{e2}(1)-\frac{\lambda_1}{\lambda_1+\lambda_2}x_{e1}(1)\right]}\\ &\quad+\int_{x_{e2}(0)}^{x_{e2}(1)} f_0(x)\mathrm{e}^{\alpha\frac{\lambda_2}{\lambda_1+\lambda_2}x}\mathrm{d}x+P_0(x_{e2}(0))\mathrm{e}^{\alpha x_{e2}(0)}\\ &= \frac{\pi_1\theta\mu_0A_1\mathrm{e}^{A_1x_{e2}(0)}\left(\frac{\mathrm{e}^{(D_1-A_1)(x_{e1}(1)-x_{e2}(0))}}{A_1-D_1}+\frac{\mathrm{e}^{(\alpha-A_1)(x_{e2}(1)-x_{e1}(1))}}{A_1-\alpha}\right)}{\xi\left(q\lambda_1+\lambda_2-\mu_1\right)\mathrm{e}^{-A_1(x_{e2}(1)-x_{e2}(0))}-\theta\mu_0}\\ &\quad+\frac{\pi_0B_1\mathrm{e}^{\alpha x_{e2}(0)}+\pi_0\xi(q\lambda_1+\lambda_2-\mu_1)A_1\frac{\mathrm{e}^{(D_1-A_1)(x_{e2}(1)-x_{e2}(0))+A_1x_{e2}(0)}}{A_1-D_1}}{\xi\left(q\lambda_1+\lambda_2-\mu_1\right)\mathrm{e}^{-A_1(x_{e2}(1)-x_{e2}(0))}-\theta\mu_0}\\ &\quad+\frac{\pi_1B_1\mathrm{e}^{[(\alpha-A_1)x_{e2}(1)-E_1x_{e1}(1)]}\mathrm{e}^{A_1x_{e2}(0)}}{\xi\left(q\lambda_1+\lambda_2-\mu_1\right)\mathrm{e}^{-A_1(x_{e2}(1)-x_{e2}(0))}-\theta\mu_0},\end{aligned}$$

where $A_1=\frac{\theta}{q\lambda_1+\lambda_2-\mu_1}-\frac{\xi}{\mu_0}, B_1=\xi\left(q\lambda_1+\lambda_2-\mu_1\right)-\theta\mu_0, D_1=\alpha\frac{\lambda_2}{\lambda_1+\lambda_2}, E_1=\alpha\frac{\lambda_1}{\lambda_1+\lambda_2}$.

Case II. $\lambda_1+\lambda_2>\mu_1$ and $\lambda_2\le\mu_1$.

In this case, the fluid level fluctuates within $[x_{e2}(0), x_{e1}(1)]$. The type 1 customers refuse to flow in when the fluid level reaches the socially optimal thresholds $x_{e1}(0)$ in state 0. The entrance probability of type 1 customers is $\frac{\mu_1}{\lambda_1+\lambda_2}$ when the fluid levels satisfy $x=x_{e1}(1)$. The entrance probabilities of type 2 customers are $\frac{\mu_0}{\lambda_2}$ and $\frac{\mu_1}{\lambda_1+\lambda_2}$ when the fluid levels satisfy $x=x_{e2}(0)$ and $x=x_{e1}(1)$, respectively. Based on Theorem 4.1, the stationary probability distribution $F_1(x)$ is continuous when the fluid level reaches the threshold $x_{e2}(0)$, then the transient probability is $P_1(x_{e2}(0))=0$.

The expected social benefit per unit time can be expressed as:

$$B\left(x_{e1}(0), x_{e2}(0), x_{e1}(1), x_{e2}(1)\right)=\lambda_{e1}R_1+\lambda_{e2}R_2-\left[C_1E\left(\mathrm{e}^{\alpha X_1}\right)+C_2E\left(\mathrm{e}^{\alpha X_2}\right)\right],$$

https://doi.org/10.1017/S0269964823000244 Published online by Cambridge University Press

where the effective arrival rates for two types of customers and the exponential expected fluid levels in the system, respectively, are:

$$\begin{aligned}\lambda_{e1} &= \lambda_1\left[\int_{x_{e2}(0)}^{x_{e1}(1)} f_1(x)\mathrm{d}x + \frac{\mu_1}{\lambda_1+\lambda_2}P_1(x_{e1}(1))\right]\\ &= \lambda_1\frac{\pi_1\theta\mu_0 \mathrm{e}^{-A_2(x_{e1}(1)-2x_{e2}(0))} + \pi_1 F_2\frac{\mu_1}{\lambda_1+\lambda_2}\mathrm{e}^{-A_2(x_{e1}(1)-x_{e2}(0))}}{\xi\left(\lambda_1+\lambda_2-\mu_1\right)\mathrm{e}^{-A_2(x_{e1}(1)-x_{e2}(0))}-\theta\mu_0},\end{aligned}$$

$$\begin{aligned}\lambda_{e2} &= \lambda_2\left[P_0(x_{e2}(0))\frac{\mu_0}{\lambda_2} + \int_{x_{e2}(0)}^{x_{e1}(1)} f_1(x)\mathrm{d}x + P_1(x_{e1}(1))\frac{\mu_1}{\lambda_1+\lambda_2}\right]\\ &= \lambda_2\frac{\pi_0 D_2\frac{\mu_0}{\lambda_2} + \pi_1\theta\mu_0\mathrm{e}^{-A_2(x_{e1}(1)-2x_{e2}(0))} + \pi_1 F_2\frac{\mu_1}{\lambda_1+\lambda_2}\mathrm{e}^{-A_2(x_{e1}(1)-x_{e2}(0))}}{\xi\left(\lambda_1+\lambda_2-\mu_1\right)\mathrm{e}^{-A_2(x_{e1}(1)-x_{e2}(0))}-\theta\mu_0},\end{aligned}$$

$$\begin{aligned}E(\mathrm{e}^{\alpha X_1}) &= \int_{x_{e2}(0)}^{x_{e1}(1)} (f_0(x)+f_1(x))\mathrm{e}^{\alpha\frac{\lambda_1}{\lambda_1+\lambda_2}x}\mathrm{d}x + P_1(x_{e1}(1))\mathrm{e}^{\alpha\frac{\lambda_1}{\lambda_1+\lambda_2}x_{e1}(1)}\\ &= \frac{\pi_0\xi(\lambda_1+\lambda_2+\mu_0-\mu_1)A_2\frac{\mathrm{e}^{A_2x_{e2}(0)+(E_1-A_2)(x_{e1}(1)-x_{e2}(0))}}{A_2-E_1}}{\xi\left(\lambda_1+\lambda_2-\mu_1\right)\mathrm{e}^{-A_2(x_{e1}(1)-x_{e2}(0))}-\theta\mu_0}\\ &\quad+\frac{\pi_1F_2\mathrm{e}^{E_1x_{e1}(1)-A_2(x_{e1}(1)-x_{e2}(0))}}{\xi\left(\lambda_1+\lambda_2-\mu_1\right)\mathrm{e}^{-A_2(x_{e1}(1)-x_{e2}(0))}-\theta\mu_0},\end{aligned}$$

$$\begin{aligned}E\left(\mathrm{e}^{\alpha X_2}\right) &= \int_{x_{e2}(0)}^{x_{e1}(1)} (f_0(x)+f_1(x))\,\mathrm{e}^{\alpha\frac{\lambda_2}{\lambda_1+\lambda_2}x}\mathrm{d}x + P_0(x_{e2}(0))\mathrm{e}^{\alpha x_{e2}(0)}\\ &\quad+P_1(x_{e1}(1))\mathrm{e}^{\alpha\frac{\lambda_2}{\lambda_1+\lambda_2}x_{e1}(1)}\\ &= \frac{\pi_0\xi(\lambda_1+\lambda_2+\mu_0-\mu_1)A_2\frac{\mathrm{e}^{A_2x_{e2}(0)}}{A_2-D_1}\mathrm{e}^{(D_1-A_2)(x_{e1}(1)-x_{e2}(0))}}{\xi\left(\lambda_1+\lambda_2-\mu_1\right)\mathrm{e}^{-A_2(x_{e1}(1)-x_{e2}(0))}-\theta\mu_0}\\ &\quad+\frac{\pi_1F_2\mathrm{e}^{-A_2(x_{e1}(1)-x_{e2}(0))}\mathrm{e}^{D_1x_{e1}(1)}+\pi_1D_2\mathrm{e}^{\alpha x_{e2}(0)}}{\xi\left(\lambda_1+\lambda_2-\mu_1\right)\mathrm{e}^{-A_2(x_{e1}(1)-x_{e2}(0))}-\theta\mu_0},\end{aligned}$$

where $A_2=\frac{\theta}{\lambda_1+\lambda_2-\mu_1}-\frac{\xi}{\mu_0}$, $D_2=\xi(\lambda_1+\lambda_2-\mu_1)-\mu_0\theta$, $F_2=\xi\left(\lambda_1+\lambda_2-\mu_1\right)+\theta\mu_0$.

Case III. $\lambda_1+\lambda_2=\mu_1$.

In this case, the fluid level stabilizes at $x_{e2}(0)$. The type 1 customers refuse to flow in and the entrance probability of type 2 customers is $\frac{\mu_0}{\lambda_2}$ when the system is in state 0. Besides, due to $x_{e2}(0)<x_{e1}(1)<x_{e2}(1)$, two types of parallel customers flow into the buffer when the system is in state 1.

The expected social benefit per unit time is:

$$\begin{aligned}&B\left(x_{e1}(0),x_{e2}(0),x_{e1}(1),x_{e2}(1)\right)\\ =&\lambda_1\pi_1R_1+\lambda_2\left(\pi_1+\pi_0\frac{\mu_0}{\lambda_2}\right)R_2-\left[C_1E\left(\mathrm{e}^{\alpha\frac{\lambda_1}{\lambda_1+\lambda_2}x_{e2}(0)}\right)+C_2E\left(\mathrm{e}^{\alpha\frac{\lambda_2}{\lambda_1+\lambda_2}x_{e2}(0)}\right)\right].\end{aligned}$$

Case IV. $\mu_0<\lambda_1+\lambda_2<\mu_1$.

In this case, the fluid level fluctuates within $[0,x_{e2}(0)]$. The fluids choose to flow in when the fluid level satisfies $x\le x_{ej}(0), j=1,2$, and the entrance probability of type 2 customers is $\frac{\mu_0}{\lambda_2}$ when the fluid level satisfies $x=x_{e2}(0)$ in state 0. Besides, due to $x_{e2}(0)<x_{e1}(1)<x_{e2}(1)$, two types of parallel customers flow into the buffer when the system is in state 1. Based on Theorem 4.1, the stationary

https://doi.org/10.1017/S0269964823000244 Published online by Cambridge University Press

probability distribution $F_0(x)$ is continuous when the fluid levels satisfy $x = 0$ and $x = x_{e1}(0)$, then the transient probabilities are $P_0(0) = 0$ and $P_0(x_{e1}(0)) = 0$. The stationary probability distribution $F_1(x)$ is continuous when the fluid level satisfies $x = x_{e2}(0)$, then the transient probability is $P_1(x_{e2}(0)) = 0$.

The expected social benefit per unit time is:

$$B(x_{e1}(0), x_{e2}(0), x_{e1}(1), x_{e2}(1)) = \lambda_{e1}R_1 + \lambda_{e2}R_2 - \left[C_1 E\left(\mathrm{e}^{\alpha X_1}\right) + C_2 E\left(\mathrm{e}^{\alpha X_2}\right)\right],$$

where the effective arrival rates for two types of customers and the exponential expected fluid levels in the system respectively are:

$$\lambda_{e1} = \lambda_1 \left(\int_0^{x_{e1}(0)} f_0(x)\mathrm{d}x + \pi_1\right) = \lambda_1 \left(\pi_1 - \frac{\pi_0 \xi E_3 \mathrm{e}^{-A_3 x_{e1}(0)}}{\theta D_3 \mathrm{e}^{-A_3 x_{e2}(0)} + \xi E_3}\right),$$

$$\begin{aligned}\lambda_{e2} =& \lambda_2 \left(\int_0^{x_{e2}(0)} f_0(x)\mathrm{d}x + \pi_1 + P_0(x_{e2}(0))\frac{\mu_0}{\lambda_2}\right)\\ =& \lambda_2 \frac{\left[\pi_1 \theta D_3 + \pi_0\left((\theta D_3 + \xi E_3)\frac{\mu_0}{\lambda_2} - \xi E_3\right)\right]\mathrm{e}^{-A_3 x_{e2}(0)} + \xi E_3}{\theta D_3 e^{-A_3 x_{e2}(0)} + \xi E_3},\end{aligned}$$

$$\begin{aligned}E\left(\mathrm{e}^{\alpha X_1}\right) &= \int_0^{x_{e1}(0)} f_0(x)\mathrm{e}^{\alpha\frac{\lambda_1}{\lambda_1+\lambda_2}x}\mathrm{d}x + \int_0^{x_{e2}(0)} f_1(x)\mathrm{e}^{\alpha\frac{\lambda_1}{\lambda_1+\lambda_2}x}\mathrm{d}x\\ &= \frac{\pi_1 \theta A_3\left[E_3 \mathrm{e}^{(E_1 - A_3)x_{e1}(0)} - D_3 \mathrm{e}^{(E_1 - A_3)x_{e2}(0)}\right]}{\left[\theta D_3 \mathrm{e}^{-A_3 x_{e2}(0)} + \xi E_3\right](E_1 - A_3)},\end{aligned}$$

$$\begin{aligned}E\left(\mathrm{e}^{\alpha X_2}\right) =& \int_0^{x_{e1}(0)} f_0(x)\mathrm{e}^{D_1 x}\mathrm{d}x + \int_{x_{e1}(0)}^{x_{e2}(0)} f_0(x)\mathrm{e}^{\alpha x}\mathrm{d}x + \int_0^{x_{e2}(0)} f_1(x)\mathrm{e}^{D_1 x}\mathrm{d}x\\ &+ P_0(x_{e2}(0))\mathrm{e}^{\alpha\left[x_{e2}(0) - \frac{\lambda_1}{\lambda_1+\lambda_2}x_{e1}(0)\right]}\\ =& \frac{\pi_0 \xi A_3\left[\frac{E_3 \mathrm{e}^{(D_1 - A_3)x_{e1}(0)} + D_3 \mathrm{e}^{(D_1 - A_3)x_{e2}(0)}}{D_1 - A_3} + \frac{E_3 \mathrm{e}^{(\alpha - A_3)(x_{e2}(0) - x_{e1}(0))}}{\alpha - A_3}\right]}{\theta D_3 \mathrm{e}^{-A_3 x_{e2}(0)} + \xi E_3}\\ &+ \frac{\pi_0\left[\xi E_3 + \theta D_3\right]\mathrm{e}^{\alpha\left[x_{e2}(0) - \frac{\lambda_1}{\lambda_1+\lambda_2}x_{e1}(0)\right] - A_3 x_{e2}(0)}}{\theta D_3 \mathrm{e}^{-A_3 x_{e2}(0)} + \xi E_3},\end{aligned}$$

where $A_3 = \frac{\xi}{p\lambda_1 + \lambda_2 - \mu_0} + \frac{\theta}{\lambda_1 + \lambda_2 - \mu_1}$, $D_3 = p\lambda_1 + \lambda_2 - \mu_0$, $E_3 = \lambda_1 + \lambda_2 - \mu_1$.

4.3. Numerical comparisons of the expected social benefit

The optimal balking strategy $(x_{e1}(0), x_{e2}(0), x_{e1}(1), x_{e2}(1))$ is assumed to obtain the maximal expected social benefit per unit time. However, the optimal balking strategy cannot be accomplished by conventional mathematical methods due to the complexity of the utility function $B(x_{e1}(0), x_{e2}(0), x_{e1}(1), x_{e2}(1))$. This subsection continues to investigate the effects of the optimal balking strategy on the expected social benefit by some numerical examples. Then, we assume that $\lambda_1 = 2.8$, $\lambda_2 = 3$, $\theta = 0.4$, $\xi = 0.3$, $\mu_0 = 1$, $R_1 = 10$, $R_2 = 15$, $C_1 = 3$, $C_2 = 4$, $\alpha = 0.2$ in the following cases.

Figure 1 shows the variation of the expected social benefit per unit time with the thresholds $x_{e2}(0)$ and $x_{e1}(1)$ when $\mu_1 = 3$. In this case, the type 2 customers' arrival rate is equal to the normal service rate, that is $\lambda_2 = \mu_1 < \lambda_1 + \lambda_2$. The expected social benefit decreases with the increase of the type 2

https://doi.org/10.1017/S0269964823000244 Published online by Cambridge University Press

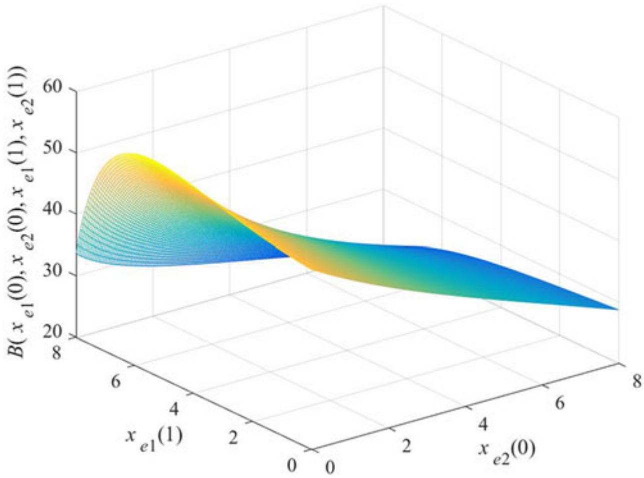

Figure 1. *Expected social benefit versus thresholds $x_{ej}(i)$ when $\mu_1 = 3$.*

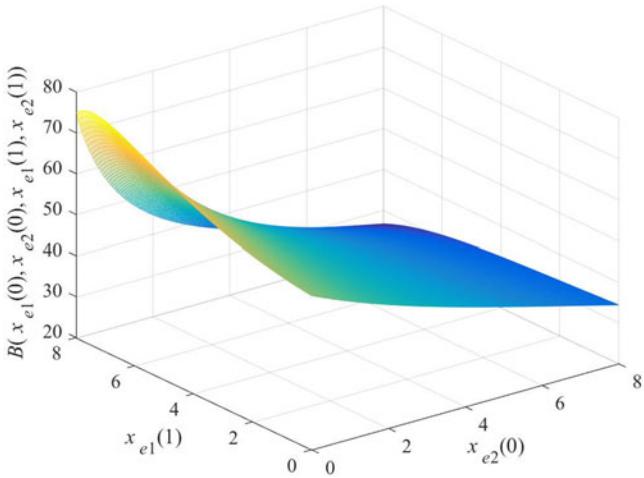

Figure 2. *Expected social benefit versus thresholds $x_{ej}(i)$ when $\mu_1 = 3.3$.*

customers' threshold $x_{e2}(0)$ when the type 1 customers' threshold $x_{e1}(1)$ is fixed. On the contrary, the expected social benefit per unit time increases first and then decreases with the increase of the threshold $x_{e1}(1)$ when the threshold $x_{e2}(0)$ is fixed. This indicates that if each fluid flows into the buffer selfishly, the system will be over-congested and cannot achieve the global maximum. Therefore, the system designers could appropriately regulate the joining thresholds for better deployment management. The expected social benefit is concave, and the maximal benefit $B_{\max} = 53.7$ can be obtained at the thresholds $x_{e2}(0) = 0$ and $x_{e1}(1) = 6.3$.

Figure 2 describes the sensitivity between the expected social benefit per unit time and the thresholds $x_{e2}(0)$ and $x_{e1}(1)$ when the outflow rate is $\mu_1 = 3.3$. When parameters satisfy $\lambda_2 < \mu_1 < \lambda_1 + \lambda_2$, the variable tendency of the expected social benefit is positively correlated with the type 1 customers' threshold $x_{e1}(1)$ while negatively correlated with the type 2 customers' threshold $x_{e2}(0)$. As the optimal threshold $x_{e2}(0)$ continues to increase, substantial customers are emerging in the system, and the buffer may be heavily loaded, which inevitably has a negative impact on society under the fully observable case. Evidently, when the threshold $x_{e1}(1)$ is sufficiently large, the expected social benefit is more likely to

https://doi.org/10.1017/S0269964823000244 Published online by Cambridge University Press

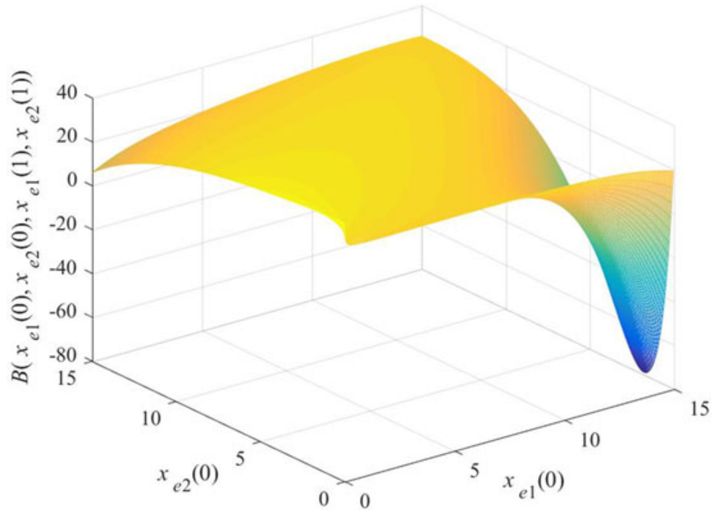

Figure 3. *Expected social benefit versus thresholds $x_{ej}(0)$ when $\mu_1 = 6$.*

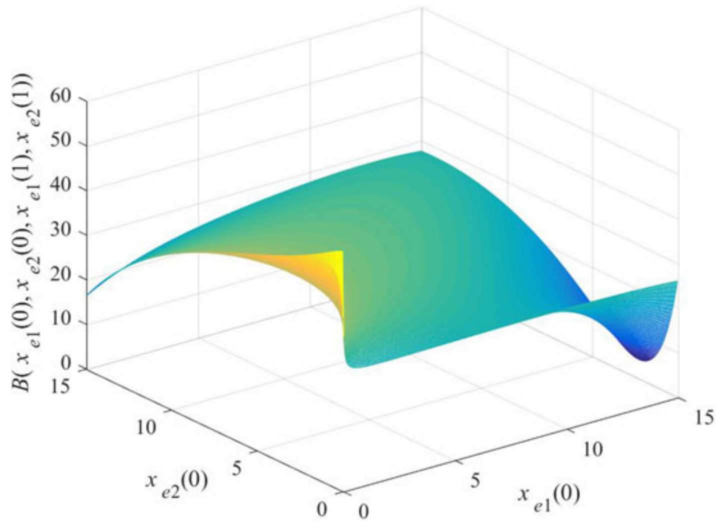

Figure 4. *Expected social benefit versus thresholds $x_{ej}(0)$ when $\mu_1 = 6.5$.*

reach its maximum. The expected social benefit achieves a maximum $B_{\max} = 76$ at thresholds $x_{e2}(0) = 0$ and $x_{e1}(1) = 7.8$.

Figures 3 and 4 present the sensitivity between the socially optimal thresholds and the expected social benefit per unit time when the sum of the arrival rates is less than the outflow rate under the normal working state, that is $\lambda_1 + \lambda_2 < \mu_1$. The expected social benefit accelerates its decline significantly when the type 1 customers' threshold $x_{e1}(0)$ increases. Moreover, when the thresholds for two types of customers are almost equal, the expected social benefit per unit time grows slowly and gradually stabilizes. Then, the expected social benefit reaches its maximum when the fluid flows into the buffer at this time.

By comparing Figures 3 and 4, we observe that when the service rate μ_1 increases, an arriving customer who is served in a shorter waiting time tends to have a stronger joining willingness. Concretely, the customers have more chances to be served at a higher service rate μ_1, which can reduce their overall sojourn time and increase the expected social benefit. Besides, the impact of the type 1 customers'

https://doi.org/10.1017/S0269964823000244 Published online by Cambridge University Press

threshold $x_{e1}(0)$ on the expected social benefit significantly reduces and the overall benefit tends to be stable when the outflow rate increases.

5. Analysis of the entrance fee strategy

Arrivals are self-interested and consider their own benefits first when encountering a joining/balking problem based on the non-cooperative game theory. The fluid inclines to flow out of the system as quickly as possible, then the buffer is saturated ultimately, which results in the over-utilization of resources and overcrowding of the system. The expected social benefit per unit time may not be globally optimal. Consequently, the decision-makers can impose an entrance fee $p_j (j = 1, 2)$ on type j customers, which can make the fluid comply with the globally optimal threshold $x^*_{ej}(i), x^*_{ej}(i) > x_{ej}(i)$. The entrance fees can relieve the queue congestion for systems and provide management insights for the optimal design of the fluid queues, which has theoretical and practical significance. An example is the entrance tickets for "fast-pass" in the amusement parks, and the separate fees and reservations for popular attractions can limit the queue length. Besides, some patients with life-threatening situations must choose to go to the emergency room and pay an expensive fee in order to receive services immediately and reduce the pressure on normal clinics.

The expected net benefit of the type j customers who observe the buffer at state i and decide to enter the system is yielded as:

$$B(x_j(i)) = R_j - p_j - C_j \frac{x_j(i) + 1}{\mu_i}, \tag{13}$$

where $x_j(i)$ is the fluid level of the type j customers in state i.

The expected net benefit should be non-negative to ensure that customers are willing to join. When the buffer is in a maintenance period, the entrance fee of the type j customers satisfies:

$$\begin{cases} B(x^*_{ej}(0) - 1) = R_j - C_j \frac{x^*_{ej}(0)}{\mu_0} - p_j \geq 0, \\ B(x^*_{ej}(0)) = R_j - C_j \frac{x^*_{ej}(0)+1}{\mu_0} - p_j < 0, \end{cases}$$

that is, the entrance fee p_j satisfies:

$$R_j - C_j \frac{x^*_{ej}(0) + 1}{\mu_0} < p_j \leq R_j - C_j \frac{x^*_{ej}(0)}{\mu_0}.$$

Then the maximal entrance fee p^*_j in state 0 satisfies:

$$p^*_j = R_j - C_j \frac{x^*_{ej}(0)}{\mu_0} < R_j - C_j \frac{x_{ej}(0)}{\mu_0}. \tag{14}$$

On the other hand, when the buffer is in a normal working period, we have:

$$\begin{cases} B(x^*_{ej}(1) - 1) = R_j - C_j \frac{x^*_{ej}(1)}{\mu_1} - p_j \geq 0, \\ B(x^*_{ej}(1)) = R_j - C_j \frac{x^*_{ej}(1)+1}{\mu_1} - p_j < 0, \end{cases}$$

that is, the entrance fee p_j satisfies:

$$R_j - C_j \frac{x^*_{ej}(1) + 1}{\mu_1} < p_j \leq R_j - C_j \frac{x^*_{ej}(1)}{\mu_1}.$$

https://doi.org/10.1017/S0269964823000244 Published online by Cambridge University Press

Thus, the maximal entrance fee p_j^* in state 1 satisfies:

$$p_j^* = R_j - C_j \frac{x_{ej}^*(1)}{\mu_1} < R_j - C_j \frac{x_{ej}(1)}{\mu_0}. \tag{15}$$

The benefit for a service provider can be expressed as:

$$Z\left(x_1(i), x_2(i)\right) = \lambda_1[1 - P_i(x_1(i))]p_1 + \lambda_2[1 - P_i(x_2(i))]p_2,$$

where $P_i(x_j(i))$ is the probability that the type j fluid reaches the threshold $x_j(i)$ in state i.

Theorem 5.1. *In the fully observable fluid queue with two types of parallel customers and incomplete fault, if the fluid observes the state $\left(x_j(i), i\right)$ and decides to join, the maximal entrance fee per unit time can be expressed as*

Case I. $\lambda_2 > \mu_1$.

$$Z\left(x_{e1}^*(1), x_{e2}^*(1)\right) = \lambda_1 \left(R_1 - \frac{C_1 x_{e1}^*(1)}{\mu_1}\right) + \lambda_2 \left[1 - P_1(x_{e2}^*(1))\right] \left(R_2 - \frac{C_2 x_{e2}^*(1)}{\mu_1}\right),$$

where

$$P_1(x_{e2}^*(1)) = \frac{\pi_1 \left[\xi\left(q\lambda_1 + \lambda_2 - \mu_1\right) - \theta\mu_0\right] \mathrm{e}^{-\left(\frac{\theta}{q\lambda_1+\lambda_2-\mu_1} - \frac{\xi}{\mu_0}\right)\left(x_{e2}^*(1) - x_{e2}(0)\right)}}{\xi\left(q\lambda_1 + \lambda_2 - \mu_1\right) \mathrm{e}^{-\left(\frac{\theta}{q\lambda_1+\lambda_2-\mu_1} - \frac{\xi}{\mu_0}\right)\left(x_{e2}^*(1) - x_{e2}(0)\right)} - \theta\mu_0}.$$

Case II. $\lambda_1 + \lambda_2 > \mu_1$ *and* $\lambda_2 \leq \mu_1$.

$$Z\left(x_{e1}^*(1), x_{e2}^*(1)\right) = \lambda_1[1 - P_1(x_{e1}^*(1))] \left(R_1 - \frac{C_1 x_{e1}^*(1)}{\mu_1}\right) + \lambda_2 \left(R_2 - \frac{C_2 x_{e2}^*(1)}{\mu_1}\right),$$

where

$$P_1(x_{e1}^*(1)) = \frac{\pi_1 \left[\xi\left(\lambda_1 + \lambda_2 - \mu_1\right) - \theta\mu_0\right] \mathrm{e}^{-\left(\frac{\theta}{\lambda_1+\lambda_2-\mu_1} - \frac{\xi}{\mu_0}\right)\left(x_{e1}^*(1) - x_{e2}(0)\right)}}{\xi\left(\lambda_1 + \lambda_2 - \mu_1\right) \mathrm{e}^{-\left(\frac{\theta}{\lambda_1+\lambda_2-\mu_1} - \frac{\xi}{\mu_0}\right)\left(x_{e1}^*(1) - x_{e2}(0)\right)} - \theta\mu_0}.$$

Case III. $\lambda_1+\lambda_2=\mu_1$.

$$Z\left(x_{e1}^*(1), x_{e2}^*(1)\right) = \lambda_1 \left(R_1 - \frac{C_1 x_{e1}^*(1)}{\mu_1}\right) + \lambda_2 \left(R_2 - \frac{C_2 x_{e2}^*(1)}{\mu_1}\right).$$

https://doi.org/10.1017/S0269964823000244 Published online by Cambridge University Press

Case IV. $\mu_0 < \lambda_1 + \lambda_2 < \mu_1$.

$$Z\left(x_{e1}^*(0), x_{e2}^*(0)\right) = \lambda_1 \left(R_1 - \frac{C_1 x_{e1}^*(0)}{\mu_0}\right) + \lambda_2 [1 - P_0(x_{e2}^*(0))] \left(R_2 - \frac{C_2 x_{e2}^*(0)}{\mu_0}\right),$$

$$Z\left(x_{e1}^*(1), x_{e2}^*(1)\right) = \lambda_1 \left(R_1 - \frac{C_1 x_{e1}^*(1)}{\mu_1}\right) + \lambda_2 \left(R_2 - \frac{C_2 x_{e2}^*(1)}{\mu_1}\right),$$

where

$$P_0(x_{e2}^*(0)) = \frac{\pi_0 \left[\theta\left(p\lambda_1 + \lambda_2 - \mu_0\right) + \xi\left(\lambda_1 + \lambda_2 - \mu_1\right)\right] \mathrm{e}^{-\left(\frac{\xi}{p\lambda_1+\lambda_2-\mu_0} + \frac{\theta}{\lambda_1+\lambda_2-\mu_1}\right)x_{e2}^*(0)}}{\theta\left(p\lambda_1 + \lambda_2 - \mu_0\right) \mathrm{e}^{-\left(\frac{\xi}{p\lambda_1+\lambda_2-\mu_0} + \frac{\theta}{\lambda_1+\lambda_2-\mu_1}\right)x_{e2}^*(0)} + \xi\left(\lambda_1 + \lambda_2 - \mu_1\right)}.$$

Proof. By comparing the arrival rate and the outflow rate, the entrance fee per unit time can be obtained in the following four cases.

Case I. $\lambda_2 > \mu_1$.

When the buffer is in state 0, the fluid level is always greater than the balking thresholds of the customers, then the fluid is reluctant to flow into the system. When the buffer is in state 1, the entrance fee benefit shows a strictly growing trend with the increase of the fluid level. Therefore, the total entrance fees of the service provider can maximize when the fluid level reaches its maximum. According to (14) and (15), the maximal unit price of the entrance fee is $R_j - C_j \frac{x_{ej}^*(1)}{\mu_1}$. The fluid flows into the buffer when the fluid level is less than the threshold strategy $x_{ej}^*(1), j = 1, 2$, then the inflow probability per unit time is $\lambda_j P\left(x < x_{ej}^*(1)\right) = \lambda_j[1 - P_1(x_{ej}^*(1))], j = 1, 2$. According to the steady-state probability distribution in Theorem 4.1, the function $F_1(x)$ is continuous when the fluid level reaches the threshold $x_{e1}^*(1)$, then the transient probability is $P_1(x_{e1}^*(1)) = 0$. Therefore, the maximal income of service providers per unit time is $Z\left(x_{e1}^*(1), x_{e2}^*(1)\right) = \lambda_1 p_1 + \lambda_2[1 - P_1(x_{e2}^*(1))]p_2$.

Case II. $\lambda_1 + \lambda_2 > \mu_1$ and $\lambda_2 \le \mu_1$.

When the buffer is in state 0, the fluid level is greater than the thresholds of two types of customers, and the fluid chooses to balk. When the buffer is in state 1, the maximal unit price of the entrance fee is $R_j - C_j \frac{x_{ej}^*(1)}{\mu_1}$, and the inflow probability of the type 1 customers is $1 - P_1(x_{e1}^*(1))$. The type 2 customers flow into the buffer when the fluid level is less than the threshold $x_{e2}^*(1)$, then the maximal income of service providers per unit time is $Z\left(x_{e1}^*(1), x_{e2}^*(1)\right) = \lambda_1[1 - P_1(x_{e1}^*(1))]p_1 + \lambda_2 p_2$.

Case III. $\lambda_1 + \lambda_2 = \mu_1$.

When the buffer is in state 0, the fluid chooses to balk and the inflow probability is 0. When the buffer is in state 1, the probability that the fluid level reaches the threshold $x_{ej}^*(1)$ is $P_1(x_{ej}^*(1)) = 0$, then we get the maximal income $Z\left(x_{e1}^*(1), x_{e2}^*(1)\right) = \lambda_1 p_1 + \lambda_2 p_2$.

Case IV. $\mu_0 < \lambda_1 + \lambda_2 < \mu_1$.

When the buffer is in state 0, the maximal unit price of the entrance fee is $R_j - C_j \frac{x_{ej}^*(0)}{\mu_0}$. The fluid flows into the buffer when the fluid level is less than the threshold strategy $x_{ej}^*(0), j = 1, 2$, then the inflow probability per unit time is $\lambda_j P\left(x < x_{ej}^*(0)\right) = \lambda_j[1 - P_0(x_{ej}^*(0))], j = 1, 2$. According to the steady-state probability distribution in Theorem 4.1, the function $F_0(x)$ is continuous when the fluid level reaches the threshold $x_{e1}^*(0)$, then the transient probability is $P_0(x_{e1}^*(0)) = 0$. Therefore, the maximal income of service providers per unit time is $Z\left(x_{e1}^*(0), x_{e2}^*(0)\right) = \lambda_1 p_1 + \lambda_2[1 - P_0(x_{e2}^*(0))]p_2$. When the buffer is in state 1, the probability that the fluid level reaches the threshold $x_{ej}^*(1)$ is $P_1(x_{ej}^*(1)) = 0$, and the maximal income per unit time is $Z\left(x_{e1}^*(1), x_{e2}^*(1)\right) = \lambda_1 p_1 + \lambda_2 p_2$. □

https://doi.org/10.1017/S0269964823000244 Published online by Cambridge University Press

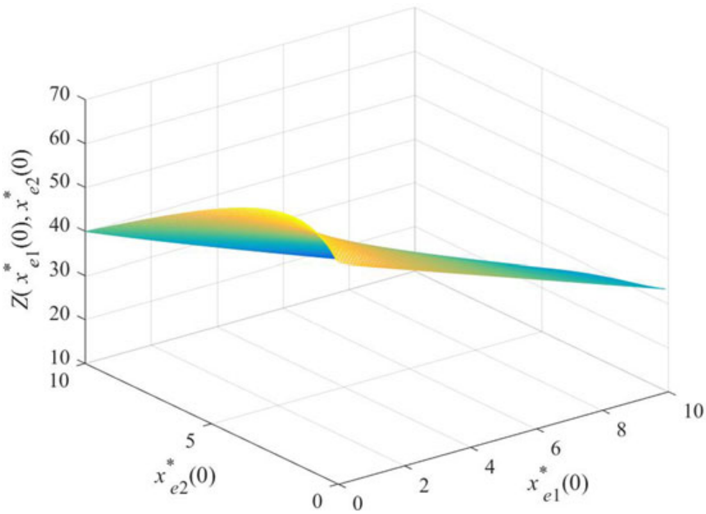

Figure 5. *Entrance fee income* $Z\left(x^*_{e1}(0), x^*_{e2}(0)\right)$ *versus thresholds* $x^*_{ej}(0)$ *when* $\mu_0 = 1.5$.

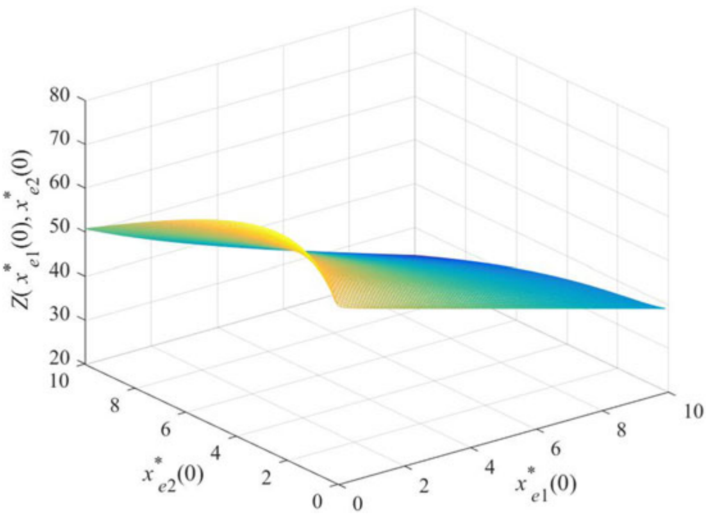

Figure 6. *Entrance fee income* $Z\left(x^*_{e1}(0), x^*_{e2}(0)\right)$ *versus thresholds* $x^*_{ej}(0)$ *when* $\mu_0 = 1.9$.

Next, we present some numerical examples intuitively with case IV due to the complexity of the maximal income per unit time in Theorem 5.1. We assume that $\lambda_1 = 2$, $\lambda_2 = 3$, $\theta = 0.6$, $\xi = 0.4$, $\mu_1 = 6$, $R_1 = 15$, $R_2 = 25$, $C_1 = 2$, $C_2 = 3, \alpha = 0.2$ in the following cases.

When the buffer is in state 1, $Z\left(x^*_{e1}(1), x^*_{e2}(1)\right)$ decreases monotonically with the thresholds $x^*_{ej}(1), j = 1, 2$ according to the expression of the maximal income in Theorem 5.1. Figure 5 displays the sensitivity between the maximal income and the thresholds $x^*_{ej}(0), j = 1, 2$ in state 0 with parameter $\mu_0 = 1.5$. $Z\left(x^*_{e1}(0), x^*_{e2}(0)\right)$ first increases and then decreases with the threshold $x^*_{e2}(0)$, and always decreases with the threshold $x^*_{e1}(0)$. The income of the entrance fee is concave and the optimal thresholds $x^*_{e1}(0)$ and $x^*_{e2}(0)$ can be set to 0 and 2.6 to maximize the entrance fee $Z\left(x^*_{e1}(0), x^*_{e2}(0)\right) = 62$. Figure 6 shows the variation between the maximal income and the optimal thresholds $x^*_{ej}(0)$ in state 0

https://doi.org/10.1017/S0269964823000244 Published online by Cambridge University Press

with parameter $\mu_0 = 1.9$. The overall trend of $Z\left(x^*_{e1}(0), x^*_{e2}(0)\right)$ is similar to Figure 5, and the entrance fee varies significantly when the threshold $x^*_{e1}(0)$ is relatively small.

From Figures 5 and 6, we can observe that the maximal income per unit time increases with faster service rates and more frequent deliveries, which has a positive impact on society under the fully observable case. This suggests that the social planners can appropriately accelerate the service rate and reasonably control the optimal thresholds for the fluid to gain more entrance fees. However, if the fee-collecting organization imposes excessively expensive fees on customers, the arrivals are reluctant to join the system due to the greater waiting costs, which affects society negatively and cannot achieve the global optimum.

6. Conclusion

Based on the queueing theory and non-cooperative game theory, this paper explores the equilibrium strategies in a fluid model with incomplete fault and parallel arrivals. The existence and uniqueness of the strategic behavior are derived, and an exponential utility function is constructed in the fully observable case. The entrance fee strategy for the arrivals and social designers could be imposed on the fluid and regulated the system parameters dynamically, which can maximize the social benefits without compromising individual interests. Finally, the effects of the inflow and outflow rates on the expected social benefit and the entrance fee income are illustrated by several numerical examples. The research results provide a feasible method and valuable insights for computer technologies, digital communication networks, and flexible manufacturing systems. Further extension of this work may explore the equilibrium behavior of the fluid in the unobservable cases. The economic analysis of the fluid queue with two types of priority customers and incomplete fault is also an intriguing and challenging direction.

Funding statement. This research was supported by the National Natural Science Foundation of China under grant no. 62171143.

Competing interest. The authors declare no conflict of interest.

References

[1] Aghsami, A. & Jolai, F. (2020). Equilibrium threshold strategies and social benefits in the fully observable Markovian queues with partial breakdowns and interruptible setup/closedown policy. *Quality Technology and Quantitative Management* 17(6): 685–722.

[2] Chen, P. & Zhou, Y. (2015). Equilibrium balking strategies in the single server queue with setup times and breakdowns. *Operational Research* 15(2): 213–231.

[3] Economou, A. & Kanta, S. (2008). Equilibrium balking strategies in the observable single-server queue with breakdowns and repairs. *Operations Research Letters* 36(6): 696–699.

[4] Economou, A. & Manou, A. (2016). Strategic behavior in an observable fluid queue with an alternating service process. *European Journal of Operational Research* 254(1): 148–160.

[5] Li, L., Wang, J. & Zhang, F. (2014). Equilibrium customer strategies in Markovian queues with partial breakdowns. *Computers and Industrial Engineering* 66(4): 751–757.

[6] Li, W., Xu, X., & Ni, G. (2018). Policy of customer-intensive public service based on queueing system. *Chinese Journal of Management* 15(3): 450–458.

[7] 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.

[8] Liu, J., Xu, X., Wang, S., & Yue, D. (2021). Equilibrium analysis of the fluid model with two types of parallel customers and breakdowns. *Communication in Statistics-Theory and Methods* 50(24): 5792–5805.

[9] Logothetis, D., Manou, A., & Economou, A. (2023). The impact of reneging on a fluid on-off queue with strategic customers. *Annals of Operations Research* 331: 629–647.

[10] Wang, J. & Zhang, F. (2011). Equilibrium analysis of the observable queues with balking and delayed repairs. *Applied Mathematics and Computation* 218(6): 2716–2729.

[11] Wang, S. & Xu, X. (2021). Equilibrium strategies of the fluid queue with working vacation. *Operational Research* 21(2): 1211–1228.

[12] Xu, B. & Xu, X. (2016). Equilibrium strategic behavior of customers in the M/M/1 queue with partial failures and repairs. *Operational Research* 18(2): 273–292.

https://doi.org/10.1017/S0269964823000244 Published online by Cambridge University Press

[13] Xu, X. & Wang, S. (2022). Equilibrium strategies in an unobservable on-off fluid queue. *Acta Mathematicae Applicatae Sinica-English Series* 38(2): 324–336.

[14] Yang, T., Wang, J., & Zhang, F. (2014). Equilibrium balking strategies in the Geo/Geo/1 queues with server breakdowns and repairs. *Quality Technology and Quantitative Management* 11(3): 231–243.

[15] Yu, S., Liu, Z., & Wu, J. (2016). Equilibrium strategies of the unobservable M/M/1 queue with balking and delayed repairs. *Applied Mathematics and Computation* 290: 56–65.

[16] Yu, S., Liu, Z., & Wu, J. (2017). Strategic behavior in the partially observable Markovian queues with partial breakdowns. *Operations Research Letters* 45(5): 471–474.

[17] Zhang, S. & Xu, X. (2019). Equilibrium analysis of queueing system with two types of parallel customers and partial failures. *Journal of Systems Science and Mathematical Sciences* 39(4): 637–647.

[18] Zhang, S. & Xu, X. (2020). Equilibrium customer strategies in Markovian queues with setup times and partial failures. *Journal of Systems Science and Complexity* 33(4): 1163–1178.

Cite this article: Zhang Y., Xu X., Zhao P. and Liu M. (2024). Equilibrium analysis of the fluid model with two types of parallel customers and incomplete fault. *Probability in the Engineering and Informational Sciences*, 1–20. https://doi.org/10.1017/S0269964823000244

https://doi.org/10.1017/S0269964823000244 Published online by Cambridge University Press