Probability in the Engineering and Informational Sciences, **33**, 2019, 417–437.
doi:10.1017/S0269964818000232

A NOTE ON MANY-SERVER FLUID MODELS WITH TIME-VARYING ARRIVALS

ZHENGHUA LONG and JIHENG ZHANG
Department of Industrial Engineering and Decision Analytics,
The Hong Kong University of Science and Technology, Clear Water Bay, HK
E-mail: zlong@conncet.ust.hk; jiheng@ust.hk

We extend the measure-valued fluid model, which tracks residuals of patience and service times, to allow for time-varying arrivals. The fluid model can be characterized by a one-dimensional convolution equation involving both the patience and service time distributions. We also make an interesting connection to the measure-valued fluid model tracking the elapsed waiting and service times. Our analysis shows that the two fluid models are actually characterized by the same one-dimensional convolution equation.

Keywords: abandonment, fluid model, many-server queue, time-varying

1. INTRODUCTION

There has been increasing interest in developing and analyzing fluid models of many-server queues with general service and patience time distributions since the pioneering work Whitt [12]. As an example of how powerful the fluid model approach is that it can be used to approximate a system with dependent service and patience times, see Bassamboo and Randhawa [1], Wu, Bassamboo and Perry [13]. The research community has developed measure-valued processes and two-parameter processes to describe the system dynamics due to the generality of the distributions. Existing studies can be divided into two categories. The first tracks the elapsed waiting and service times of all customers in the system, see Whitt [12] and Kang and Ramanan [4]. The second tracks the residual patience and service times, see Zhang [14].

The first line of works is represented by Kang and Ramanan [4], which is based on Kaspi and Ramanan [6] on the model without abandonment. Kang and Ramanan [4] requires rather complicated conditions on the hazard rate of the distributions (see Assumption 3.3 in Kang and Ramanan [4]). Zuniga [15] extends Kang and Ramanan [4] by relaxing their assumptions. However, both in Kang and Ramanan [4] and Zuniga [15], the existence of a solution to the fluid model is proved using a stochastic approximation.

The fluid model tracking elapsed times is also developed in Liu and Whitt [7,8], which adapt the approach in Whitt [12] to allow the number of servers and service/patience time distributions to vary with time. Moreover, they provide a direct analysis on the fluid model tracking elapsed times to obtain existence and uniqueness by assuming two key assumptions: (i) the system alternates between overloaded and underloaded intervals, and

 0269-9648/18 $25.00

(ii) the functions specifying the fluid model are suitably smooth. The direct analysis on the fluid model tracking elapsed times is also studied in Kang [5], which assumes that the service time distribution has a density and the hazard rate function of the patience time distribution is locally bounded.

In the second line of works tracking residual times, Zhang [14] directly proves the existence and uniqueness of the many-server fluid model with a constant arrival rate only requiring continuity of the service time distribution and Lipschitz continuity of the patience time distribution. Moreover, it builds the foundation to prove the convergence to the equilibrium state in Long and Zhang [9]. However, the modeling approach in Zhang [14] seems a bit inflexible as extending the analysis of the fluid model with a constant arrival rate to time-varying arrival rates is not that straightforward. Another downside of Zhang [14] is the condition on the initial state of the queue, which assumes that initial customers are those who arrived in the past following an arrival process with the same arrival rate.

This paper extends the measure-valued fluid model tracking residual times in Zhang [14], where a fluid model is studied for the $G/GI/n + GI$ queue, to allow for time-varying arrivals. In this paper, we focus on the study of the fluid model of many-server queues with time-varying arrival rates, and general service and patience time distributions. The queueing model is denoted by $G_t/GI/n + GI$. The G_t represents a general time-varying arrival process. The first GI indicates that service times are independent and identically distributed (i.i.d.) with a general distribution. The n denotes the number of homogeneous servers. There is an unlimited waiting space, called the buffer, where customers wait to be served according to the first-come-first-served (FCFS) discipline. Customers are only allowed to abandon if their patience times expire before their service starts. Again, the patience times are i.i.d. and with a general distribution (the second GI).

We also provide a unified approach to study the two types of fluid models tracking elapsed times and residual times. We show that both types of fluid models are characterized by the same convolution equation, which is proved to possess a unique solution. Thus, both types of fluid models are alternative to approximate the original stochastic processes. We address the following open issues regarding the fluid model.

1. Can we extend the measure-valued fluid model of Zhang [14], which tracks residuals, to allow for time-varying arrivals?
2. What is the fundamental mathematical law driving the dynamics of both types of fluid models (tracking elapsed and residual times)?

We aim to address these two questions by first extending Zhang [14] to allow for time-varying arrivals in Section 2. We derive a one-dimensional convolution equation (2.17) as the key insight of the fluid model in Section 2.1, where several properties of the fluid model are also developed. Analysis of the key equation (2.17) is presented in Section 3. We identify a connection between the initial conditions required by both fluid models when analyzing the fluid model tracking elapsed times in Section 4. We show that (2.17) also serves as the foundation of the fluid model tracking elapsed times.

2. THE FLUID MODEL TRACKING RESIDUAL TIMES

Let $\mathbb{R}$ denote the set of real numbers and $\mathbb{R}_+ = [0, \infty)$. For $a, b \in \mathbb{R}$, write a^+ for the positive part of a and $a \wedge b$ for the minimum. For convenience of notation, define $C_x = (x, \infty)$. We append a bar sign on processes to indicate that they are fluid model processes and to be consistent with the notations in Zhang [14].

https://doi.org/10.1017/S0269964818000232 Published online by Cambridge University Press

We consider a fluid model of the $G_t/GI/n+GI$ queue with time-varying arrival process

$$\bar{E}(t) = \int_0^t \lambda(s)ds, \quad \lambda(\cdot) \geq 0. \tag{2.1}$$

For $t < 0$, let $\lambda(t)$ be the arrival rate of the fluid arriving before time 0. Following the modeling approach in Zhang [14], we introduce the *virtual buffer* which holds all the fluid that has not yet scheduled to enter service even when their patience is exhausted. When the fluid is admitted to service, the system will check whether the fluid has positive remaining patience time or not. Only the fluid with positive remaining patience time will enter service, otherwise, it will abandon the system. Thus, the fluid in the virtual buffer is allowed to have negative remaining patience time. For any time $t \in [0, \infty)$, let $\bar{\mathcal{R}}(t)(C_x)$ denote the amount of fluid in the virtual buffer with remaining patience time larger than $x \in \mathbb{R}$; and $\bar{\mathcal{Z}}(t)(C_x)$ denote the amount of fluid in service with remaining service time larger than $x \geq 0$. We assume customers' patience times and service times are mutually independent and follow the distributions F and G, respectively. See Bassamboo and Randhawa [1], Wu, Bassamboo and Perry [13] for the study of dependent service and patience time distributions.

Denote by $\bar{R}(t)$, $\bar{Q}(t)$ and $\bar{Z}(t)$ the amount of fluid in the virtual buffer, in the queue and in service at time t, respectively. Then they can be recovered from $\bar{\mathcal{R}}$ and $\bar{\mathcal{Z}}$ as follows

$$\bar{R}(t) = \bar{\mathcal{R}}(\mathbb{R}), \quad \bar{Q}(t) = \bar{\mathcal{R}}(t)(C_0) \quad \text{and} \quad \bar{Z}(t) = \bar{\mathcal{Z}}(t)(C_0), \tag{2.2}$$

where $C_0 = (0, \infty)$ since the fluid in the queue or in service cannot have 0 remaining times. Let $\bar{X}(t) = \bar{Q}(t) + \bar{Z}(t)$ denote the total fluid content in the system. We assume that the initial fluid arrives at some negative time $t \in (-\infty, 0)$. So our arrival process $\bar{E}(t)$ extends to the negative axis and we introduce $\omega(t)$ as the solution to

$$\bar{R}(t) = \int_{t-\omega(t)}^t d\bar{E}(s).$$

Intuitively, $\omega(t)$ can be considered as the waiting time of the earliest arrived fluid in the virtual buffer at time t. And $t - \omega(t)$ can be thought of as the arrival time of the earliest arrived fluid in the virtual buffer at time t. We also introduce

$$\bar{B}(t) = \bar{E}(t) - \bar{R}(t),$$

and $\bar{B}(t) - \bar{B}(s)$ can be regarded as the fluid content leaving the virtual buffer during the time interval $(s, t]$. Note that the processes $\bar{R}$, $\bar{Q}$, $\bar{Z}$ and $\bar{X}$ are all derived directly from the measure-valued process $(\bar{\mathcal{R}}, \bar{\mathcal{Z}})$, and the processes ω and $\bar{B}$ are derived by combining the measure-valued process and the arrival process. The fluid model is defined as follows.

Definition 2.1 (The fluid model tracking residual): *The fluid model* $\{\bar{\mathcal{R}}(t), \bar{\mathcal{Z}}(t)\}$ *satisfies the dynamic equations*

$$\bar{\mathcal{R}}(t)(C_x) = \int_{t-\omega(t)}^t F^c(x + t - s)d\bar{E}(s), \quad x \in \mathbb{R}, \tag{2.3}$$

$$\bar{\mathcal{Z}}(t)(C_x) = \bar{\mathcal{Z}}(0)(C_{x+t}) + \int_0^t F^c\left(\omega(s)\right) G^c(x + t - s)d\bar{B}(s), \quad x \in \mathbb{R}_+, \tag{2.4}$$

and the non-idling constraints

$$\bar{Q}(t) = (\bar{X}(t) - 1)^+, \tag{2.5}$$

$$\bar{Z}(t) = \bar{X}(t) \wedge 1. \tag{2.6}$$

https://doi.org/10.1017/S0269964818000232 Published online by Cambridge University Press

Moreover, the initial state $(\bar{\mathcal{R}}(0), \bar{\mathcal{Z}}(0))$ *satisfying* (2.3) *and* (2.4) *at time* $t = 0$ *has no atoms.*

The intuition behind the above definition resembles that of (3.1)–(3.2) in Zhang [14]. The difference is that $\omega(t)$ simply reduces to $\bar{R}(t)/\lambda$ when the arrival rate is constant and equals λ. We want to emphasize here that the dynamic equations (2.3) and (2.4) implicitly assume the FCFS policy. In general, only specifying the remaining patience times in the queue does not give a full picture of the status of the queue. For example, assuming there are two customers with remaining patience times 1 and 10 in the queue, the measure does not tell us who is the first in the queue. To overcome this issue, we incorporate FCFS into the dynamic equations. For any $s \in [t - \omega(t), t]$, among the infinitesimal amount of arriving fluid $d\bar{E}(s)$, the fraction of remaining patience time larger than x at time t is $F^c(x + t - s)d\bar{E}(s)$ as shown in (2.3). By the definition of $\bar{B}$ and ω, it is easy to see that

$$\bar{B}(t) = \bar{E}(t - \omega(t)). \tag{2.7}$$

The infinitesimal amount of fluid $d\bar{B}(t)$ that is about to enter service at time t actually arrived at time $t - \omega(t)$. Only a fraction $F^c(\omega(t))$, with the original patience time larger than the waiting time $\omega(t)$, actually enters service. This is characterized by (2.4).

For direct analysis of the fluid model, we need the following assumption on the service and patience time distributions throughout this paper.

Assumption 2.1: *The service time distribution* G *is continuous with finite mean* $1/\mu$*, and the patience time distribution* F *is Lipschitz continuous.*

2.1. Properties and Analysis

Preliminary analysis. We first perform some preliminary analysis to arrive at the key equation (2.17). It follows from (2.3) that

$$\begin{aligned}\bar{Q}(t) &= \bar{\mathcal{R}}(t)(C_0) \\ &= \int_{t-\omega(t)}^{t} F^c(t-s)\lambda(s)ds = \int_0^{\omega(t)} F^c(s)\lambda(t-s)ds.\end{aligned} \tag{2.8}$$

For any $t \geq 0$, introduce two new functions

$$F_t(x) = \int_0^x f(s)\lambda(t-s)ds, \tag{2.9}$$

$$F_{d,t}(x) = \int_0^x F^c(s)\lambda(t-s)ds, \tag{2.10}$$

where $f(x) = (d/dx)F(x)$ exists since every Lipschitz continuous function is absolutely continuous (Page 112 in Royden [11]). The domain for both functions is $x \in [0, t + \omega(0)]$ since the fluid model at t only depends on the arrival process from time $-\omega(0)$ to time t. For any $t \geq 0$, denote by $N_{F,t}$ the maximum value of $F_{d,t}(\cdot)$;

$$N_{F,t} = F_{d,t}(t + \omega(0)). \tag{2.11}$$

https://doi.org/10.1017/S0269964818000232 Published online by Cambridge University Press

Using (2.10), (2.8) becomes

$$\bar{Q}(t) = F_{d,t}(\omega(t)). \tag{2.12}$$

It follows from (2.4) and (2.7) that

$$\begin{aligned}\bar{Z}(t) &= \bar{\mathcal{Z}}(t)(C_0)\\ &= \bar{\mathcal{Z}}(0)(C_t) + \int_0^t F^c(\omega(s))G^c(t-s)d\bar{E}(s-\omega(s)).\end{aligned}$$

Since a monotone function is of bounded variation, it follows from Lemma A.1 and (2.5) that $\bar{Q}(t)$ is also of bounded variation. Thus, applying the chain rule to (2.8) gives

$$\begin{aligned}d\bar{Q}(t) &= \lambda(t)dt - F^c(\omega(t))\lambda(t-\omega(t))d(t-\omega(t)) - \int_{t-\omega(t)}^t f(t-s)\lambda(s)ds\\ &= \lambda(t)dt - F^c(\omega(t))d\bar{E}(t-\omega(t)) - F_t(\omega(t))dt,\end{aligned} \tag{2.13}$$

so that

$$\bar{Z}(t) = \bar{\mathcal{Z}}(0)(C_t) + \int_0^t G^c(t-s)\left[\lambda(s) - F_s(\omega(s))\right]ds - \int_0^t G^c(t-s)d\bar{Q}(s).$$

Performing change of variable and integration by parts, we have

$$\begin{aligned}\bar{Z}(t) = \bar{\mathcal{Z}}(0)(C_t) &+ \frac{1}{\mu}\int_0^t \lambda(t-s) - F_{t-s}(\omega(t-s))dG_{\mathrm{e}}(s)\\ &- \bar{Q}(t) + \bar{Q}(0)G^c(t) + \int_0^t \bar{Q}(t-s)dG(s),\end{aligned} \tag{2.14}$$

where $G_{\mathrm{e}}(\cdot)$ is the equilibrium distribution associated with G defined as

$$G_{\mathrm{e}}(x) = \mu\int_0^x G^c(y)dy. \tag{2.15}$$

Based on (2.9) and (2.10), we introduce the following function for all $t \geq 0$,

$$H_t(y) = \begin{cases}\lambda(t) - F_t(F_{d,t}^{-1}(y)), & \text{if } 0 \leq y < N_{F,t},\\ \lambda(t) - F_t(F_{d,t}^{-1}(N_{F,t})), & \text{if } y \geq N_{F,t},\end{cases} \tag{2.16}$$

where $N_{F,t}$ is defined in (2.11) and $F_{d,t}^{-1}(y) = \inf\{x \geq 0 : F_{d,t}(x) \geq y\}$ for all $y \in [0, N_{F,t}]$. By (2.5) and (2.12), $\omega(t) = F_{d,t}^{-1}((\bar{X}(t)-1)^+)$. Combining this with (2.5), (2.14) and (2.16), we obtain the following *key equation*

$$\begin{aligned}\bar{X}(t) &= \bar{\mathcal{Z}}(0)(C_t) + \bar{Q}(0)G^c(t)\\ &\quad + \frac{1}{\mu}\int_0^t H_{t-s}((\bar{X}(t-s)-1)^+)dG_{\mathrm{e}}(s) + \int_0^t (\bar{X}(t-s)-1)^+dG(s).\end{aligned} \tag{2.17}$$

https://doi.org/10.1017/S0269964818000232 Published online by Cambridge University Press

Existence and uniqueness of a solution to the fluid model. It follows from the proof of Theorem 3.1 in Zhang [14] that there is a one-to-one correspondence between the measure-valued process $(\bar{\mathcal{R}}, \bar{\mathcal{Z}})$ and the one-dimensional process $\bar{X}$. Thus, the existence and uniqueness of a solution to the fluid model in Definition 2.1 is equivalent to the existence and uniqueness of the solution to the key equation (2.17), which is proved in Proposition 3.1 in Section 3. Below we immediately have the following theorem.

THEOREM 2.1 (Existence and uniqueness): *Under Assumption 2.1, there exists a unique solution to the measure-valued fluid model* $\{\bar{\mathcal{R}}(t), \bar{\mathcal{Z}}(t)\}$ *in* (2.3)–(2.6).

Time shift of the fluid model. For any $\tau \geq 0$, denote $(\bar{\mathcal{R}}_\tau(t), \bar{\mathcal{Z}}_\tau(t)) = (\bar{\mathcal{R}}(\tau + t), \bar{\mathcal{Z}}(\tau + t))$. The time shift for all the derived "status" quantities such as $\omega_\tau(\cdot)$, $\bar{R}_\tau(\cdot)$, $\bar{Q}_\tau(\cdot)$, $\bar{Z}_\tau(\cdot)$ and $\bar{X}_\tau(\cdot)$ is defined in the same way, for example, $\omega_\tau(t) = \omega(\tau + t)$. However, the time shift for the "cumulative" process $\bar{E}_\tau(t)$ is defined as $\bar{E}_\tau(t) = \bar{E}(\tau + t) - \bar{E}(\tau)$ (similarly for $\bar{B}_\tau(\cdot)$). If we think of the arrival rate, then $\bar{E}'_\tau(s) = \lambda_\tau(s) = \lambda(\tau + s)$. The following proposition shows that the fluid model can be "restarted" at time $\tau > 0$ by viewing $(\bar{\mathcal{R}}(\tau), \bar{\mathcal{Z}}(\tau))$ as the initial condition.

PROPOSITION 2.1 (Time-shifted fluid model): *The time-shifted fluid solution* $(\bar{\mathcal{R}}_\tau(t), \bar{\mathcal{Z}}_\tau(t))$ *satisfies*

$$\bar{\mathcal{R}}_\tau(t)(C_x) = \int_{t-\omega_\tau(t)}^{t} F^c(x + t - s) d\bar{E}_\tau(s), \quad x \in \mathbb{R}, \tag{2.18}$$

$$\bar{\mathcal{Z}}_\tau(t)(C_x) = \bar{\mathcal{Z}}(\tau)(C_{x+t}) + \int_0^t F^c\left(\omega_\tau(s)\right) G^c(x + t - s) d\bar{B}_\tau(s), \quad x \in \mathbb{R}_+. \tag{2.19}$$

And the shifted key equation becomes

$$\begin{aligned} \bar{X}_\tau(t) = {} & \bar{\mathcal{Z}}(\tau)(C_t) + \bar{Q}(\tau) G^c(t) \\ & + \frac{1}{\mu} \int_0^t H_{\tau+t-s}((\bar{X}_\tau(t-s) - 1)^+) dG_e(s) + \int_0^t (\bar{X}_\tau(t-s) - 1)^+ dG(s). \end{aligned} \tag{2.20}$$

PROOF: Replacing t in (2.3) by $\tau + t$ gives

$$\begin{aligned} \bar{\mathcal{R}}_\tau(t)(C_x) &= \bar{\mathcal{R}}(\tau + t)(C_x) \\ &= \int_{\tau+t-\omega(\tau+t)}^{\tau+t} F^c(x + \tau + t - s) d\bar{E}(s) \\ &= \int_{t-\omega(\tau+t)}^{t} F^c(x + t - s) d\bar{E}(\tau + s), \end{aligned}$$

https://doi.org/10.1017/S0269964818000232 Published online by Cambridge University Press

where the last equation is due to change of variable. This implies (2.18) by using the definition of the time shift. Similarly, replacing t by $\tau + t$ in (2.4) yields

$$\begin{aligned}\bar{\mathcal{Z}}_\tau(t)(C_x) &= \bar{\mathcal{Z}}(\tau + t)(C_x)\\ &= \bar{\mathcal{Z}}(0)(C_{x+\tau+t}) + \int_0^{\tau+t} F^c(\omega(s))G^c(x+\tau+t-s)d\bar{B}(s)\\ &= \bar{\mathcal{Z}}(0)(C_{x+\tau+t}) + \int_0^{\tau} F^c(\omega(s))G^c(x+\tau+t-s)d\bar{B}(s)\\ &\quad + \int_\tau^{\tau+t} F^c(\omega(s))G^c(x+\tau+t-s)d\bar{B}(s)\\ &= \bar{\mathcal{Z}}(\tau)(C_{x+t}) + \int_0^{t} F^c(\omega(\tau+s))G^c(x+t-s)d\bar{B}(\tau+s),\end{aligned}$$

which implies (2.19) by the definition of the time shift.

Replacing (t, x) in (2.4) by (τ, t) yields

$$\bar{\mathcal{Z}}(\tau)(C_t) = \bar{\mathcal{Z}}(0)(C_{\tau+t}) + \int_0^{\tau} F^c\left(\omega(s)\right) G^c(\tau+t-s)d\bar{B}(s).$$

Combining the above with (2.7), (2.13) and (2.16), we can verify that

$$\begin{aligned}\bar{\mathcal{Z}}(\tau)(C_t) = \bar{\mathcal{Z}}(0)(C_{\tau+t}) + \frac{1}{\mu}\int_t^{\tau+t} H_{\tau+t-s}(\bar{Q}(\tau+t-s)dG_{\mathrm{e}}(s)\\ - \bar{Q}(\tau)G^c(t) + \bar{Q}(0)G^c(\tau+t) + \int_t^{\tau+t}\bar{Q}(\tau+t-s)dG(s).\end{aligned}$$

Thus, the right-hand side of (2.20) becomes

$$\begin{aligned}&\bar{\mathcal{Z}}(0)(C_{\tau+t}) + \bar{Q}(0)G^c(\tau+t) + \frac{1}{\mu}\int_0^{\tau+t} H_{\tau+t-s}(\bar{Q}(\tau+t-s))dG_{\mathrm{e}}(s)\\ &\quad + \int_0^{\tau+t}\bar{Q}(\tau+t-s)dG(s),\end{aligned}$$

which equals $\bar{X}(\tau+t)$ by (2.17). Thus (2.20) follows by applying the time-shift definition. ■

Special case with a constant arrival rate. We specialize the time-varying arrival rate to be constant, that is, $\lambda(\cdot) \equiv \lambda$. It can be seen from Lemma A.3 that any solution to (2.17) satisfies

$$(\bar{X}(t) - 1)^+ \le \lambda \int_0^{t+\omega(0)} F^c(s)ds \quad \text{for all } t \ge 0.$$

It follows from (2.16) that for any $t \ge 0$, $H_t(y) = \lambda H(y)$ for all $y \in [0, \lambda\int_0^{t+\omega(0)} F^c(s)ds]$, where

$$H(y) = \begin{cases} F^c\left(F_d^{-1}\left(\frac{y}{\lambda}\right)\right), & \text{if } 0 \le y < \lambda N_F,\\ 0, & \text{if } y \ge \lambda N_F,\end{cases}$$

https://doi.org/10.1017/S0269964818000232 Published online by Cambridge University Press

with $F_d(x) = \int_0^x F^c(s)ds$ and $N_F = \int_0^\infty F^c(s)ds$. Thus we can replace $H_{t-s}(\cdot)$ in (2.17) by $\lambda H(\cdot)$ and obtain the following key equation for this special case:

$$\bar{X}(t) = \bar{\mathcal{Z}}(0)(C_t) + \bar{Q}(0)G^c(t) + \frac{\lambda}{\mu}\int_0^t H((\bar{X}(t-s)-1)^+)dG_e(s) + \int_0^t (\bar{X}(t-s)-1)^+ dG(s),$$

which is consistent with the key equation (4.6) in Zhang [14].

Balance equations. Regarding the last term in (2.4), we introduce an auxiliary process

$$\bar{A}(t) = \int_0^t F^c(\omega(s))d\bar{B}(s),$$

which can be interpreted as the amount of fluid that actually enters service. By (2.7), (2.12), (2.13) and (2.16), the auxiliary process can be written as

$$\bar{A}(t) = \int_0^t H_s(\bar{Q}(s))ds - \bar{Q}(t) + \bar{Q}(0).$$

Denote by $\bar{L}(t)$ the *abandonment* process, which can be derived from the following balance equation of the physical queue,

$$\bar{Q}(t) = \bar{Q}(0) + \bar{E}(t) - \bar{L}(t) - \bar{A}(t).$$

From the above two equations, we get

$$\bar{L}(t) = \bar{E}(t) - \int_0^t H_s(\bar{Q}(s))ds.$$

Using (2.9), (2.12) and (2.16) yields

$$\bar{L}(t) = \int_0^t \int_0^{\omega(s)} f(x)\lambda(s-x)dxds.$$

According to the fluid dynamic equation (2.4),

$$\bar{Z}(t) = \bar{\mathcal{Z}}(0)(C_t) + \int_0^t G^c(t-s)d\bar{A}(s).$$

Then the *service completion* process, denoted by $\bar{S}(t)$, can be derived from the following balance equation of the server pool,

$$\bar{Z}(t) = \bar{Z}(0) + \bar{A}(t) - \bar{S}(t).$$

That is

$$\bar{S}(t) = \bar{\mathcal{Z}}(0)((0,t]) + \int_0^t G(t-s)d\bar{A}(s).$$

It is clear that the balance equation of the fluid content in the system satisfies

$$\bar{X}(t) = \bar{X}(0) + \bar{E}(t) - \bar{L}(t) - \bar{S}(t).$$

Note that the introduced processes $\bar{A}$, $\bar{L}$, $\bar{S}$ and the balance equations are not needed in the definition and analysis of the fluid model. We only provide them here for completeness and potential future use.

https://doi.org/10.1017/S0269964818000232 Published online by Cambridge University Press

3. THE ONE-DIMENSIONAL CONVOLUTION EQUATION

We analyze the key equation (2.17) in this section. Denote by $\mathbf{C}[0, \infty)$ the space of continuous functions on the interval $[0, \infty)$. The following Proposition 3.1 showing the existence and uniqueness of the solution to (2.17) is the main result of this paper. The auxiliary Lemmas A.1–A.5, which also reveal some additional properties of the solution to (2.17), are placed in the appendix.

PROPOSITION 3.1: *There exists a unique solution* $\bar{X} \in \mathbf{C}[0, \infty)$ *to* (2.17).

PROOF: We first prove that there exists a number $b > 0$ and a unique continuous function $\bar{X}(t)$ satisfies (2.17) when $0 \le t \le b$. And then we extend the solution indefinitely. According to the value of $\bar{X}(0)$, we consider the following two cases.

Case 1: $\bar{X}(0) \le 1$. This implies $\omega(0) = 0$ by (2.8). Deduce from Lemma A.3 that for all $t \ge 0$,

$$(\bar{X}(t) - 1)^+ \le \int_0^t F^c(s)\lambda(t-s)ds = N_{F,t}. \tag{3.1}$$

Let M be any strictly positive number and $S_F = \inf\{x \ge 0 : F(x) = 1\}$. From (2.16) the following derivative is bounded for all $t \in [0, S_F \wedge M/2]$:

$$\frac{d}{dy}H_t(y) = -\frac{f(F_{d,t}^{-1}(y))}{F^c(F_{d,t}^{-1}(y))} \ge -\frac{L_F}{F^c(t)} \ge -\frac{L_F}{F^c((S_F \wedge M)/2)}, \quad \text{if } 0 \le y \le N_{F,t},$$

where L_F is denoted to be the Lipschitz constant of F by Assumption 2.1. So we can pick $b_1 = S_F \wedge M/2$ and then for any $t \in [0, b_1]$ the function $H_t(\cdot)$ in (2.17) is Lipschitz continuous. Let $L = L_F/F^c((S_F \wedge M)/2)$ be the Lipschitz constant. By Assumption 2.1, there exists a $b_2 > 0$ such that

$$\kappa := \frac{1}{\mu}L[G_e(b_2) - G_e(0)] + [G(b_2) - G(0)] < 1.$$

Let $b = \min\{b_1, b_2\}$. For any $x \in C[0, b]$, define

$$\Psi(x)(t) = \bar{\mathcal{Z}}(0)(C_t) + \bar{Q}(0)G^c(t) + \frac{1}{\mu}\int_0^t H_{t-s}((x(t-s)-1)^+)dG_e(s) + \int_0^t (x(t-s)-1)^+ dG(s).$$

It is clear that $\Psi(x)(t)$ is continuous in t, so Ψ is a mapping from $\mathbf{C}[0, b]$ to $\mathbf{C}[0, b]$. Let $\rho(x, x') = \sup_{t\in[0,b]} |x(t) - x'(t)|$ denote the uniform distance between two functions in $\mathbf{C}[0, b]$. For any $x, x' \in C[0, b]$, we have

$$\begin{aligned}
\rho(\Psi(x), \Psi(x')) &\le \sup_{t\in[0,b]} \frac{1}{\mu}\int_0^t L|(x(t-s)-1)^+ - (x'(t-s)-1)^+|dG_e(s) \\
&\quad + \sup_{t\in[0,b]} \int_0^t |(x(t-s)-1)^+ - (x'(t-s)-1)^+|dG(s) \\
&\le \frac{1}{\mu}L\int_0^b \rho(x, x')dG_e(s) + \int_0^b \rho(x, x')dG(s) \\
&\le \kappa\rho(x, x').
\end{aligned}$$

https://doi.org/10.1017/S0269964818000232 Published online by Cambridge University Press

Since $\kappa < 1$, Ψ is a contraction mapping on $\mathbf{C}[0, b]$ under the uniform topology ρ. Note that $\mathbf{C}[0, b]$ is complete under the uniform topology of ρ (cf. p. 80 in Billingsley [2]). Thus, by the contraction mapping theorem (e.g., Theorem 3.2 in Hunter and Nachtergaele [3]), Ψ has a unique fixed point x, that is, $x = \Psi(x)$. This proves that (2.17) has a unique solution on $[0, b]$ in this case.

Case 2: $\bar{X}(0) > 1$. Due to the continuity of the solution to (2.17) (if there is any) proved in Lemma A.2, there exists $b_3 > 0$ such that

$$\bar{X}(t) \geq 1 \quad \text{for all } t \in [0, b_3]. \tag{3.2}$$

For notational simplicity, denote $q(t) = (\bar{X}(t) - 1)^+$ and

$$a(t) = \int_0^t H_s(q(s))ds - q(t) + q(0). \tag{3.3}$$

For $t \in [0, b_3]$, (A.2) obtained in the proof of Lemma A.1 becomes

$$a(t) = 1 - \bar{\mathcal{Z}}(0)(C_t) + \int_0^t a(t-s)dG(s).$$

Let G^{n*} be the n-fold convolution of G with itself, and denote $U_G(t) = \sum_{i=0}^{\infty} G^{n*}$. The solution to the above renewal equation is

$$a(t) = \int_0^t (1 - \bar{\mathcal{Z}}(0)(C_{t-s}))dU_G(s), \quad t \in [0, b_3].$$

It is clear that $a(t)$ is continuous. Since $H_t(\cdot)$ is continuous, with a known $a(t)$ there exists a continuous solution $q(t)$ to the Eq. (3.3) following from Theorem II.1.1 in Miller [10].

Next we prove the uniqueness. Assume that $q_1(t)$ and $q_2(t)$ satisfy (3.3) on the interval $[0, b_3]$. Let

$$\mathcal{L}(t) := (q_1(t) - q_2(t))^2, \quad t \in [0, b_3].$$

Then, on the interval $[0, b_3]$ we can see from (3.3) that

$$\mathcal{L}'(t) = 2[q_1(t) - q_2(t)][H_t(q_1(t)) - H_t(q_2(t))] \leq 0,$$

where the last inequality is due to the fact that $H_t(\cdot)$ is non-increasing from (2.16). Thus $\mathcal{L}(t)$ is non-increasing on $[0, b_3]$. Since $\mathcal{L}(0) = 0$ and $\mathcal{L}(t) \geq 0$, $\mathcal{L}(t) = 0$ for all $t \in [0, b_3]$. Hence

$$q_1(t) = q_2(t) \quad \text{for all } t \in [0, b_3].$$

Thus (2.17) only has one solution on the interval $[0, b_3]$. By Corollary II.2.6 in Miller [10], we can further extend the solution to a point $\tau > 0$, where $\bar{X}(\tau) = 1$. If there is no such a finite time point, the existence and uniqueness immediately follow. Otherwise, starting from τ, we can apply a similar argument as the above Case 1 to extend the solution to an extra interval with length b. Since the argument involves the time-shifted fluid model equation (2.20), we provide a rigorous proof in Lemma A.5. As a result, we can at least get the unique solution of (2.17) on the interval $[0, b]$ in this case.

Combing the above two cases yields that there exists a unique continuous function $\bar{X}(t)$ satisfying (2.17) when $0 \leq t \leq b$. Here, the definition of b is same as the one in Lemma A.5. Thus, applying Lemma A.5 consecutively at $\tau = b, 2b, \ldots$, we can extend the existence and uniqueness to $[kb, (k+1)b]$, $k = 1, 2, \ldots$ to the whole interval $[0, \infty)$, proving the result. ■

https://doi.org/10.1017/S0269964818000232 Published online by Cambridge University Press

4. THE FLUID MODEL TRACKING ELAPSED TIMES

We now present the fluid model tracking elapsed times following earlier works in this direction, for example, Kang and Ramanan [4] and Liu and Whitt [8]. Let $\bar{\mathcal{R}}_a(t)([0, x])$ denote the amount of fluid in the *potential queue* with elapsed waiting time no larger than x. A potential queue holds all the fluid that has arrived but has not yet abandoned, no matter whether it has entered service or not. Note that the virtual buffer is employed in the fluid model tracking the residual times, and the potential queue is used in the fluid model tracking elapsed times. Let $\bar{\mathcal{Z}}_a(t)([0, x])$ denote the amount of fluid in the server pool with elapsed service time no larger than x. The head count processes of the fluid amount in the potential queue, in the queue and in service can be recovered from $\bar{\mathcal{R}}_a$ and $\bar{\mathcal{Z}}_a$ as follows

$$\bar{R}_a(t) = \bar{\mathcal{R}}_a(t)([0,\infty)), \quad \bar{Q}(t) = \bar{\mathcal{R}}_a(t)([0,\omega(t)]) \quad \text{and} \quad \bar{Z}(t) = \bar{\mathcal{Z}}_a(t)([0,\infty)),$$

where, as in Section 2, $\omega(t)$ represents the waiting time of the earliest arrived fluid content in the physical queue. For the convenience of notations, let $r(t, x)$ and $z(t, x)$ be the densities of the measures $\bar{\mathcal{R}}_a(t)$ and $\bar{\mathcal{Z}}_a(t)$, respectively. In details, $r(t, x) = (d/dx)\bar{\mathcal{R}}_a(t)([0, x])$ and $z(t, x) = (d/dx)\bar{\mathcal{Z}}_a(t)([0, x])$, which exist almost everywhere since $\bar{\mathcal{R}}_a(t)([0, x])$ and $\bar{\mathcal{Z}}_a(t)([0, x])$ are non-decreasing in x (see Royden [11], Page 100). We have the following definition for the fluid model tracking elapsed times.

DEFINITION 4.1 (The fluid model tracking elapsed times): *The fluid model* $\{\bar{\mathcal{R}}_a(t), \bar{\mathcal{Z}}_a(t)\}$ *satisfies the following dynamic equations*

$$\bar{\mathcal{R}}_a(t)([0,x]) = \int_0^{(x-t)^+} \frac{F^c(s+t)}{F^c(s)} r(0,s)ds + \int_{(t-x)^+}^{t} F^c(t-s)d\bar{E}(s), \quad x \in \mathbb{R}_+, \tag{4.1}$$

$$\bar{\mathcal{Z}}_a(t)([0,x]) = \int_0^{(x-t)^+} \frac{G^c(s+t)}{G^c(s)} z(0,s)ds + \int_{(t-x)^+}^{t} G^c(t-s)d\bar{A}(s), \quad x \in \mathbb{R}_+, \tag{4.2}$$

where $\bar{A}(s)$ *is the amount of fluid that enters service by time* s*. Moreover, the abandonment process, denoted by* $\bar{L}(t)$*, satisfies*

$$\bar{L}(t) = \int_0^t \left(\int_0^{\omega(s)} \frac{f(x)}{1-F(x)} r(s,x)dx \right) ds, \tag{4.3}$$

where f *is the density function of* F*. The fluid model needs to satisfy the balance equation*

$$\bar{Q}(t) = \bar{Q}(0) + \bar{E}(t) - \bar{L}(t) - \bar{A}(t), \tag{4.4}$$

and the non-idling constraints (2.5) *and* (2.6)*.*

Note that when $x \geq t$, the fluid content in the potential queue with elapsed waiting time less than or equal to x consists of two parts: initial fluid in the queue with age $s \in [0, (x-t)^+]$ at time 0 and fluid that arrived during $[(t-x)^+, t]$. For the initial fluid content, only a fraction $F^c(s+t)/F^c(s)$ of the infinitesimal amount of fluid $r(0, s)ds$ would still be in the potential queue at time t. For the fluid that arrived at time $s \in [(t-x)^+, t]$, a proportion $F^c(t-s)$ of the infinitesimal amount $d\bar{E}(s)$ will not reach its patience time at time t. When $x < t$, the fluid content in the potential queue with elapsed waiting time less than or equal to x only consists the fluid arriving at $s \in [(t-x)^+, t]$ and the explanation is exactly the same. The explanations for $\tilde{\mathcal{Z}}(t, x)$ and (4.3) are similar. We refer to Kang and

https://doi.org/10.1017/S0269964818000232 Published online by Cambridge University Press

Ramanan [4] and Liu and Whitt [8] for more detailed discussions on the intuition behind this definition.

It is worth pointing out that the waiting time $\omega(t)$, abandonment process $\bar{L}(t)$ and the balance equation (4.4) are needed in Definition 4.1, while they are derived from the model defined in Definition 2.1. The reason is that the same measure $(\bar{\mathcal{R}}_a(t), \bar{\mathcal{Z}}_a(t))$ at time t could represent two different states if we are given two different $\omega(t)$. However, for the fluid model in Section 2, we can uniquely determine $\omega(t)$ once the measure $\bar{\mathcal{R}}(t)$ is given. Of course, this hinges on the validity of the initial condition. In the following, we show a connection between the initial conditions in both types of fluid models.

Correspondence between initial conditions. Given any initial state $(r(0,\cdot), z(0,\cdot))$ in the fluid model tracking elapsed times, we can construct a corresponding initial state in the fluid model tracking residual times. Let

$$\begin{aligned}\bar{\mathcal{R}}(0)(C_x) &:= \int_{-\omega(0)}^{0} F^c(x-s)\lambda(s)ds, \quad x \in \mathbb{R},\\ \bar{\mathcal{Z}}(0)(C_x) &:= \int_0^{\infty} \frac{G^c(s+x)}{G^c(s)} z(0,s)ds, \quad x \in \mathbb{R}_+,\end{aligned} \tag{4.5}$$

where $\lambda(s)$, $s \in (-\omega(0), 0)$, is set to be

$$\lambda(s) = \frac{r(0,-s)}{F^c(-s)} \tag{4.6}$$

and can be regarded as the arrival rate of the fluid arriving before time 0. It is easy to see that the initial state $(\bar{\mathcal{R}}(0), \bar{\mathcal{Z}}(0))$ satisfies (2.3) and (2.4) at time $t = 0$.

PROPOSITION 4.1 (Identical key equation): *The measure-valued fluid model* $\{\bar{\mathcal{R}}_a(t), \bar{\mathcal{Z}}_a(t)\}$ *that tracks elapsed times is characterized via the same key equation* (2.17).

PROOF: It follows from (4.6) that for all $x > t$ the measure-valued process (4.1) becomes

$$\begin{aligned}\bar{\mathcal{R}}_a(t)([0,x]) &= \int_0^{x-t} \frac{F^c(s+t)}{F^c(s)} r(0,s)ds + \int_0^t F^c(t-s)d\bar{E}(s)\\ &= \int_0^{x-t} F^c(s+t)\lambda(-s)ds + \int_0^t F^c(t-s)d\bar{E}(s)\\ &= \int_{t-x}^{t} F^c(t-s)d\bar{E}(s).\end{aligned} \tag{4.7}$$

It can also be seen from (4.1) that (4.7) still holds for all $x \le t$. From the above we have

$$\begin{aligned}\bar{Q}(t) &= \bar{\mathcal{R}}_a(t)([0,\omega(t)])\\ &= \int_{t-\omega(t)}^{t} F^c(t-s)d\bar{E}(s) = \int_0^{\omega(t)} F^c(s)\lambda(t-s)ds\\ &= F_{d,t}(\omega(t)),\end{aligned} \tag{4.8}$$

https://doi.org/10.1017/S0269964818000232 Published online by Cambridge University Press

where the last equation follows from the definition of $F_{d,t}(\cdot)$ in (2.10). From (4.7), we obtain $r(t, x) = F^c(x)\lambda(t-x)$. Thus (4.3) becomes

$$\begin{aligned}\bar{L}(t) &= \int_0^t \int_0^{\omega(s)} f(x)\lambda(s-x)dxds \\ &= \bar{E}(t) - \int_0^t H_s(\bar{Q}(s))ds,\end{aligned}$$

where the last equation follows from $H_s(\cdot)$ in (2.16) and (4.8). Combining this with (4.4) yields

$$\bar{A}(t) = \int_0^t H_s(\bar{Q}(s))ds - \bar{Q}(t) + \bar{Q}(0).$$

Plugging $x = \infty$ and the above into (4.2) then combining with (4.5), we have

$$\bar{Z}(t) = \bar{\mathcal{Z}}(0)(C_t) + \int_0^t G^c(t-s)H_s(\bar{Q}(s))ds - \int_0^t G^c(t-s)d\bar{Q}(s),$$

Performing change of variable and integration by parts, we have

$$\begin{aligned}\bar{X}(t) = \bar{\mathcal{Z}}(0)(C_t) + \bar{Q}(0)G^c(t) + \frac{1}{\mu}\int_0^t H_{t-s}((\bar{X}(t-s)-1)^+)dG_e(s) \\ + \int_0^t (\bar{X}(t-s)-1)^+ dG(s),\end{aligned}$$

which is exactly the same as the key equation in (2.17). ■

Acknowledgements

The authors are grateful to the AE and the anonymous referees for constructive comments and suggestions. The research is supported in part by the Hong Kong Research Grants Council [Grants GRF-16501015 and GRF-16201417].

References

1. Bassamboo, A. & Randhawa, R.S. (2016). Scheduling homogeneous impatient customers. *Management Science* 62(7): 2129–2147.
2. Billingsley, P. (1999). Convergence of probability measures. *Wiley Series in Probability and Statistics: Probability and Statistics*. 2nd ed. New York: John Wiley & Sons Inc.
3. Hunter, J.K. & Nachtergaele, B. (2001). *Applied analysis*. River Edge, NJ: World Scientific Publishing Co. Inc.
4. Kang, W. & Ramanan, K. (2010). Fluid limits of many-server queues with reneging. *The Annals of Applied Probability* 20(6): 2204–2260.
5. Kang, W. (2014). Existence and uniqueness of a fluid model for many-server queues with abandonment. *Operations Research Letters* 42(6–7): 478–483.
6. Kaspi, H. & Ramanan, K. (2011). Law of large numbers limits for many-server queues. *The Annals of Applied Probability* 21(1): 33–114.
7. Liu, Y. & Whitt, W. (2011). Large-time asymptotics for the $G_t/M_t/s_t + GI_t$ many-server fluid queue with abandonment. *Queueing Systems* 67(2): 145–182.
8. Liu, Y. & Whitt, W. (2012). The $G_t/GI/s_t + GI$ many-server fluid queue. *Queueing Systems* 71(4): 405–444.
9. Long, Z. & Zhang, J. (2014). Convergence to equilibrium states for fluid models of many-server queues with abandonment. *Operations Research Letters* 42(6–7): 388–393.

https://doi.org/10.1017/S0269964818000232 Published online by Cambridge University Press

10. Miller, R. (1971). *Nonlinear Volterra integral equations. Mathematics lecture note series*. Menlo Park, CA: W. A. Benjamin.
11. Royden, H.L. (1988). *Real analysis*. 3rd ed. New York: Macmillan Publishing Company.
12. Whitt, W. (2006). Fluid models for multiserver queues with abandonments. *Operations Research* 54(1): 37–54.
13. Wu, C.A., Bassamboo, A. & Perry, O. (2018). Service systems with dependent service and patience times. *Management Science*. https://doi.org/10.1287/mnsc.2017.2983
14. Zhang, J. (2013). Fluid models of many-server queues with abandonment. *Queueing Systems* 73(2): 147–193.
15. Zuñiga, A.W. (2014). Fluid limits of many-server queues with abandonments, general service and continuous patience time distributions. *Stochastic Processes and their Applications* 124(3): 1436–1468.

APPENDIX A. AUXILIARY LEMMAS

LEMMA A.1: *If there is any function* $\bar{X}(t)$ *satisfying* (2.17)*, then*

$$\int_0^t H_s((\bar{X}(s)-1)^+)ds - (\bar{X}(t)-1)^+ + (\bar{X}(0)-1)^+$$

is non-decreasing.

PROOF: To simplify the notation, let $q(t) = (\bar{X}(t)-1)^+$ and

$$a(t) = \int_0^t H_s(q(s))ds - q(t) + q(0). \tag{A.1}$$

So we just need to prove that $a(\cdot)$ is non-decreasing. By (2.15) and (2.17),

$$\begin{aligned}\bar{X}(t) &= \bar{\mathcal{Z}}(0)(C_t) + q(0)G^c(t) + \int_0^t H_{t-s}(q(t-s))[1-G(s)]ds + \int_0^t q(t-s)dG(s)\\ &= \bar{\mathcal{Z}}(0)(C_t) + q(0)G^c(t) + \int_0^t H_s(q(s))ds - \int_0^t H_s(q(s))G(t-s)ds + \int_0^t q(t-s)dG(s).\end{aligned}$$

The second last term on the above equation satisfies

$$\begin{aligned}\int_0^t H_s(q(s))G(t-s)ds &= \int_0^t \int_0^{t-s} H_s(q(s))dG(x)ds\\ &= \int_0^t \int_0^{t-x} H_s(q(s))dsdG(x),\end{aligned}$$

where the last equality follows by changing the order of integration. So we obtain

$$\bar{X}(t) = \bar{\mathcal{Z}}(0)(C_t) + q(0) + \int_0^t H_s(q(s))ds - \int_0^t \left[\int_0^{t-x} H_s(q(s))ds - q(t-x) + q(0)\right] dG(x).$$

According to the above definition of $a(t)$, we have

$$a(t) = (\bar{X}(t)\wedge 1) - \bar{\mathcal{Z}}(0)(C_t) + \int_0^t a(t-s)dG(s). \tag{A.2}$$

We now use (A.1) and (A.2) to show that $a(\cdot)$ is non-decreasing. Choose $b>0$ such that $G(b)<1$. We first show that $a(\cdot)$ is non-decreasing on the interval $[0, b]$. Let

$$a^* = \inf_{0\le t\le t'\le b} a(t') - a(t).$$

We will prove by contradiction that $a^* \ge 0$, which implies that $a(\cdot)$ is non-decreasing on $[0, b]$. Assume to the contrary that $a^* < 0$. Choose any $t_1, t_2 \in [0, b]$ with $t_1 \le t_2$ and consider the following two cases.

https://doi.org/10.1017/S0269964818000232 Published online by Cambridge University Press

Case 1: If $\bar{X}(t_2) \geq 1$, then $\bar{X}(t_2) \wedge 1 = 1$. Applying (A.2), we have

$$a(t_2) - a(t_1) = (\bar{X}(t_2) \wedge 1) - (\bar{X}(t_1) \wedge 1) - \bar{\mathcal{Z}}(0)(C_{t_2}) + \bar{\mathcal{Z}}(0)(C_{t_1}) + \int_{t_1}^{t_2} a(t_2 - s) - a(0) dG(s) + \int_0^{t_1} a(t_2 - s) - a(t_1 - s) dG(s),$$

where $a(0) = 0$ from (A.1). So due to the fact $\bar{\mathcal{Z}}(0)(C_t)$ is non-increasing that

$$a(t_2) - a(t_1) \geq \int_0^{t_2} a^* dG(s) = a^* G(t_2) \geq a^* G(b),$$

where the last inequality follows from the assumption that a^* is negative.

Case 2: If $\bar{X}(t_2) < 1$. Let $\tau = \sup\{s < t_2 : \bar{X}(s) \geq 1\} \vee 0$ be the last time that $\bar{X}$ is larger than or equal to 1. Thus $\bar{X}(t) < 1$ for all $t \in (\tau, t_2]$. Then by (A.1) and (2.16),

$$a(t_2) - a(t) = \int_t^{t_2} H_s(0) ds + q(t) \geq \int_t^{t_2} \lambda(s) ds \quad \text{for all } t \in [\tau, t_2]. \tag{A.3}$$

If $t_1 \in [\tau, t_2]$, then from the above we have $a(t_2) - a(t_1) \geq 0 \geq a^* G(b)$. If $t_1 \in [0, \tau)$, then it is only possible when $\tau > 0$. If $\bar{X}(\tau) \geq 1$, we can apply the same analysis in the above case (where $\bar{X}(t_2) \geq 1$) at time τ to obtain $a(\tau) - a(t_1) \geq a^* G(b)$. This together with (A.3) shows that $a(t_2) - a(t_1) \geq a^* G(b)$. Otherwise, if $\bar{X}(\tau) < 1$, from the definition of τ we can find a sequence $\tau_n \in (t_1, \tau)$ satisfying $\tau_n \to \tau$ as n goes to infinity and $\bar{X}(\tau_n) \geq 1$ for all $n \in \mathbb{N}$. Applying Case 1 at each time epoch τ_n obtains $a(\tau_n) - a(t_1) \geq a^* G(b)$. Combining this with (A.1) and (A.3) yields

$$\begin{aligned} a(t_2) - a(t_1) &= a(t_2) - a(\tau) + a(\tau) - a(\tau_n) + a(\tau_n) - a(t_1) \\ &\geq 0 + \int_{\tau_n}^{\tau} H_s(q(s)) ds - q(\tau) + q(\tau_n) + a^* G(b). \end{aligned}$$

Note that $q(\tau) = 0$ since we have $\bar{X}(\tau) < 1$. Thus, the above inequality also yields $a(t_2) - a(t_1) \geq a^* G(b)$ since $q(\tau_n) \geq 0$ and $\lim_{n \to \infty} \tau_n = \tau$. Summarizing both cases of $\bar{X}(t_2)$, we have

$$a(t_2) - a(t_1) \geq a^* G(b).$$

Taking infimum over $0 \leq t_1 \leq t_2 \leq b$ gives $a^* \geq a^* G(b)$. Since $G(b) < 1$, it contradicts the assumption $a^* < 0$. So we must have $a^* \geq 0$, which implies that $a(t)$ is non-decreasing on $[0, b]$.

We next extend the monotonicity to $\mathbb{R}_+$ proving by induction. Suppose we can show that $a(\cdot)$ is non-decreasing on the interval $[0, nb]$ for some $n \in \mathbb{N}$. Let

$$\bar{\mathcal{Z}}(nb)(C_t) = \bar{\mathcal{Z}}(0)(C_{nb+t}) + \int_0^{nb} G^c(nb + t - s) da(s). \tag{A.4}$$

It is clear that the shifted fluid versions of (A.1) and (A.2) satisfy

$$a_{nb}(t) = \int_0^t H_{nb+s}(q_{nb}(s)) ds - q_{nb}(t) + q_{nb}(0),$$

$$a_{nb}(t) = (\bar{X}_{nb}(t) \wedge 1) - \bar{\mathcal{Z}}(nb)(C_t) + \int_0^t a_{nb}(t - s) dG(s).$$

To show that $a(\cdot)$ is non-decreasing on $[nb, (n+1)b]$ is the same as showing that $a_{nb}(\cdot)$ is non-decreasing on $[0, b]$. For this purpose, it is enough to verify that $\bar{\mathcal{Z}}(nb)(C_t)$ is non-increasing. This is obviously true due to the fact that $a(\cdot)$ is non-decreasing on $[0, nb]$ and by the definition of

https://doi.org/10.1017/S0269964818000232 Published online by Cambridge University Press

$\bar{\mathcal{Z}}_{nb}(C_t)$ in (A.4). Thus we extend the non-decreasing interval to $[0, (n+1)b]$. By induction, the function $a(\cdot)$ is non-decreasing on the whole interval $[0, \infty)$. ■

LEMMA A.2: *If there is any function $\bar{X}(t)$ satisfying (2.17), then $\bar{X}(t)$ is a continuous function, that is, $\bar{X}(t) \in \mathbf{C}[0, \infty)$.*

PROOF: Let us denote the non-decreasing formula in Lemma A.1 by

$$a(t) = \int_0^t H_s((\bar{X}(s)-1)^+)ds - (\bar{X}(t)-1)^+ + (\bar{X}(0)-1)^+.$$

Then we can transform (2.17) to be

$$\bar{X}(t) = \bar{\mathcal{Z}}(0)(C_t) + (\bar{X}(0)-1)^+ + \int_0^t H_s((\bar{X}(s)-1)^+)ds - \int_0^t G(t-s)da(s).$$

It suffices to prove the continuity of $\int_0^t G(t-s)da(s)$. For any $0 \le t_1 < t_2$, we can see from the monotonicity of $a(t)$ that

$$\begin{aligned} 0 &\le \int_0^{t_2} G(t_2-s)da(s) - \int_0^{t_1} G(t_1-s)da(s) \\ &= \int_{t_1}^{t_2} G(t_2-s)da(s) + \int_0^{t_1} [G(t_2-s) - G(t_1-s)]da(s). \end{aligned}$$

Obviously, the right-hand side of the above equality could be arbitrarily small as long as t_1 and t_2 are close enough. Thus, the result holds. ■

LEMMA A.3: *If there is any function $\bar{X}(t)$ satisfying (2.17), then*

$$(\bar{X}(t)-1)^+ \le N_{F,t} = \int_0^{t+\omega(0)} F^c(s)\lambda(t-s)ds \quad \textit{for all } t \ge 0, \tag{A.5}$$

where $N_{F,t}$ is denoted in (2.11).

PROOF: For notational simplicity, let $q(t) = (\bar{X}(t)-1)^+$. By (2.2) and (2.3), the initial state satisfies $q(0) = \int_0^{\omega(0)} F^c(s)\lambda(-s)ds$. This implies that (A.5) holds at $t=0$. Suppose there exists $t_1 > 0$ such that $q(t_1) > N_{F,t_1}$. Let $t_0 = \sup\{s < t_1 : q(s) \le N_{F,s}\}$. Then due to the continuity proven in Lemma A.2, we have $q(t) - N_{F,t} \ge 0$ for all $t \in [t_0, t_1]$. By Lemma A.1 and (2.16),

$$\begin{aligned} q(t_1) - q(t_0) &\le \int_{t_0}^{t_1} H_s(q(s))ds \\ &= \int_{t_0}^{t_1} \lambda(s)ds - \int_{t_0}^{t_1}\int_0^{s+\omega(0)} f(x)\lambda(s-x)dxds. \end{aligned} \tag{A.6}$$

Apply change of variable to the last term

$$\begin{aligned} &\int_{t_0}^{t_1}\int_{-\omega(0)}^{s} f(s-x)\lambda(x)dxds \\ &= \int_{-\omega(0)}^{t_0} dx \int_{t_0}^{t_1} f(s-x)\lambda(x)ds + \int_{t_0}^{t_1} dx \int_x^{t_1} f(s-x)\lambda(x)ds \\ &= \int_{t_0}^{t_1} \lambda(x)dx - \int_0^{t_1+\omega(0)} F^c(x)\lambda(t_1-x)dx + \int_0^{t_0+\omega(0)} F^c(x)\lambda(t_0-x)dx, \end{aligned}$$

https://doi.org/10.1017/S0269964818000232 Published online by Cambridge University Press

where the first equality follows by changing the order of integration. Plugging the above into (A.6) yields

$$q(t_1) \le q(t_0) + \int_0^{t_1+\omega(0)} F^c(x)\lambda(t_1 - x)dx - \int_0^{t_0+\omega(0)} F^c(x)\lambda(t_0 - x)dx.$$

Then by the definition of t_0, the above implies $q(t_1) \le \int_0^{t_1+\omega(0)} F^c(x)\lambda(t_1 - x)dx = N_{F,t_1}$. This is a contradiction. So (A.5) follows. ■

LEMMA A.4: *If there is any function $\bar{X}(t)$ satisfying (2.17), then*

$$t - F_{d,t}^{-1}((\bar{X}(t) - 1)^+) \tag{A.7}$$

is non-decreasing.

PROOF: As in the proof of Lemmas A.1–A.3, we also denote $q(t) = (\bar{X}(t) - 1)^+$ and

$$a(t) = \int_0^t H_s(q(s))ds - q(t) + q(0). \tag{A.8}$$

Meanwhile, let $\varpi(t) = F_{d,t}^{-1}(q(t))$ to simplify the notation. Then by (2.10) and (A.5) we obtain

$$q(t) = \int_0^{\varpi(t)} F^c(s)\lambda(t - s)ds = \int_{t-\varpi(t)}^t F^c(t - s)\lambda(s)ds. \tag{A.9}$$

Applying the chain rule to the above equation yields

$$\begin{aligned} dq(t) &= \lambda(t)dt - F^c(\varpi(t))\lambda(t - \varpi(t))d(t - \varpi(t)) - \int_{t-\varpi(t)}^t f(t - s)\lambda(s)dsdt \\ &= \lambda(t)dt - F^c(\varpi(t))d\bar{E}(t - \varpi(t)) - F_t(\varpi(t))dt, \end{aligned}$$

where F_t is given in (2.9). Combining the above with (A.8) and (2.16), it is easy to verify

$$da(t) = F^c(\varpi(t))d\bar{E}(t - \varpi(t)). \tag{A.10}$$

To arrive at the result of this lemma, our first step is to show that

$$\bar{E}(t - \varpi(t)) \text{ is non-decreasing.} \tag{A.11}$$

Let $S_F = \inf\{x \ge 0 : F(x) = 1\}$. According to the value of S_F we consider the following two cases.

Case 1: $S_F = \infty$. Since $\varpi(t) < \infty$ on any finite time interval by (A.5) and (A.9), one can see from (A.10) that

$$\bar{E}(t - \varpi(t)) - \bar{E}(0 - \varpi(0)) = \int_0^t \frac{1}{F^c(\varpi(s))} da(s). \tag{A.12}$$

Due to the fact that $a(\cdot)$ is non-decreasing from Lemma A.1, the above immediately yields that $\bar{E}(t - \varpi(t))$ is non-decreasing.

Case 2: $S_F < \infty$. In this case, it is possible that $\varpi(\cdot) = S_F$ within a finite time. So (A.12) may not hold. Therefore, we choose any $0 \le t_1 < t_2$ and consider the following two situations.

https://doi.org/10.1017/S0269964818000232 Published online by Cambridge University Press

If $\varpi(t_1) = S_F$, then

$$t_2 - \varpi(t_2) - (t_1 - \varpi(t_1)) = t_2 - t_1 + S_F - \varpi(t_2) \geq 0, \tag{A.13}$$

where the inequality holds due to the fact that $\varpi(t) = F_{d,t}^{-1}(q(t)) \leq S_F$ for all $t \geq 0$ following from (2.10). Thus, the above inequality (A.13) and (2.1) yield $\bar{E}(t_2 - \varpi(t_2)) \geq \bar{E}(t_1 - \varpi(t_1))$.

If $\varpi(t_1) < S_F$. Let $\tau = \inf\{s \geq t_1 : \varpi(s) \geq S_F\}$ be the first time that $\varpi(t)$ is larger than or equal to S_F. Once $\tau = \infty$, it becomes the same issue as Case 1. So we just need to consider $\tau < \infty$. Similar to (A.12) we have

$$\bar{E}(t - \varpi(t)) - \bar{E}(t_1 - \varpi(t_1)) = \int_{t_1}^{t} \frac{1}{F^c(\varpi(s))} da(s) \geq 0 \quad \text{for all } t \in [t_1, \tau],$$

where the last inequality holds due to the fact that $F^c(\varpi(s)) > 0$ on the interval (t_1, t) and $a(t)$ is non-decreasing proved in Lemma A.1. If $t_2 \in (t_1, \tau]$, the above yields that $\bar{E}(t_2 - \varpi(t_2)) \geq \bar{E}(t_1 - \varpi(t_1))$. If $t_2 \in (\tau, \infty)$, then similar to (A.13) we can apply the situation $\varpi(\tau) = S_F$ to obtain $t_2 - \varpi(t_2) \geq \tau - \varpi(\tau)$. This implies that $\bar{E}(t_2 - \varpi(t_2)) \geq \bar{E}(\tau - \varpi(\tau))$ by (2.1). This together with the above inequality yields $\bar{E}(t_2 - \varpi(t_2)) \geq \bar{E}(t_1 - \varpi(t_1))$. From the above analysis we can conclude that (A.11) holds in any case.

With the help of (A.11), we prove (A.7) by contradiction and assume to the contrary that there exist $0 \leq \tau < t$ such that $t - \varpi(t) < \tau - \varpi(\tau)$. This implies

$$\bar{E}(\tau - \varpi(\tau)) - \bar{E}(t - \varpi(t)) = \int_{t-\varpi(t)}^{\tau-\varpi(\tau)} \lambda(s) ds \geq 0,$$

where the equation comes from (2.1) and the inequality follows since $\lambda(\cdot) \geq 0$. On the other hand, one can see from (A.11) that $\bar{E}(t - \varpi(t)) \geq \bar{E}(\tau - \varpi(\tau))$ since $t > \tau$. Therefore there must be $\bar{E}(t - \varpi(t)) = \bar{E}(\tau - \varpi(\tau))$. This together with (A.9) yields

$$\begin{aligned} q(t) &= \int_{t-\varpi(t)}^{t} F^c(t-s)\lambda(s) ds = \int_{\tau-\varpi(\tau)}^{t} F^c(t-s)\lambda(s) ds \\ &= \int_{0}^{t-\tau+\varpi(\tau)} F^c(s)\lambda(t-s) ds, \end{aligned}$$

where the last equation follows by applying change of variable. By (2.10) and (A.9) we have

$$q(t) = \int_{0}^{\varpi(t)} F^c(s)\lambda(t-s) ds = F_{d,t}(\varpi(t)).$$

Recall the definition of $F_{d,t}^{-1}$ below (2.16). We can see from the above two equations that

$$\varpi(t) \leq t - \tau + \varpi(\tau).$$

The above just means $t - \varpi(t) \geq \tau - \varpi(\tau)$. This contradicts the assumption. Thus we must have (A.7) to be non-decreasing. ■

LEMMA A.5: *If the existence and uniqueness of the solution to* (2.17) *hold on* $[0, \tau]$ *for some* $\tau > 0$, *then there exists a number* $b > 0$ *such that the unique solution can be extended to* $[0, \tau + b]$.

PROOF: To prove this lemma, we analyze the following two cases.

https://doi.org/10.1017/S0269964818000232 Published online by Cambridge University Press

Case 1: $\bar{X}(\tau) \le 1$. It follows from Proposition 2.1 that we can obtain the same shifted key equation as (2.20). Thus, it is enough to prove the existence and uniqueness of the solution to (2.20) on $[0, b]$. Deduce from Lemma A.4 that for all $t \ge 0$,

$$\tau + t - F^{-1}_{d,\tau+t}((\bar{X}_\tau(t) - 1)^+) \ge \tau - F^{-1}_{d,\tau}((\bar{X}_\tau(0) - 1)^+) = \tau.$$

Combining the above with (2.10) yields

$$(\bar{X}_\tau(t) - 1)^+ \le \int_0^t F^c(s)\lambda(\tau + t - s)ds = F_{d,\tau+t}(t). \tag{A.14}$$

Let

$$\tilde{H}_{\tau+t}(y) = \begin{cases} \lambda(\tau + t) - F_{\tau+t}(F^{-1}_{d,\tau+t}(y)), & \text{if } 0 \le y < F_{d,\tau+t}(t), \\ \lambda(\tau + t) - F_{\tau+t}(t), & \text{if } y \ge F_{d,\tau+t}(t), \end{cases} \tag{A.15}$$

which actually is a truncation of $H_{\tau+t}(y)$. This is because that one can see from (2.16),

$$\tilde{H}_{\tau+t}(y) = H_{\tau+t}(y) \quad \text{for all } 0 \le y \le F_{d,\tau+t}(t). \tag{A.16}$$

Thus, deduce from (A.14) that any function $\bar{X}_\tau(t)$ satisfying (2.20) also satisfies the following convolution equation,

$$\begin{aligned} \bar{X}_\tau(t) = {} & \bar{\mathcal{Z}}(\tau)(C_t) + \bar{Q}(\tau)G^c(t) \\ & + \frac{1}{\mu}\int_0^t \tilde{H}_{\tau+t-s}((\bar{X}_\tau(t-s) - 1)^+)dG_e(s) + \int_0^t (\bar{X}_\tau(t-s) - 1)^+ dG(s). \end{aligned} \tag{A.17}$$

Regarding τ as a starting point, the above equation thus becomes a key equation of a fluid model with initial state $(\bar{\mathcal{R}}_\tau(0), \bar{\mathcal{Z}}_\tau(0))$ satisfying (2.18) and (2.19) at time $t = 0$ and external arrival rate being $\lambda_\tau(t) := \lambda(\tau + t)$. Thus for any function $\bar{X}_\tau(t)$ satisfying (A.17) we can obtain similar results as Lemmas A.1–A.4 using the same argument (with different initial states and the external arrival processes). Especially, replacing $\omega(0)$ and $\lambda(t)$ in (A.5), respectively with $\omega_\tau(0)$ and $\lambda_\tau(t)$, we can obtain the following inequality for any solution $\bar{X}_\tau(t)$ satisfying (A.17),

$$(\bar{X}_\tau(t) - 1)^+ \le \int_0^{t+\omega_\tau(0)} F^c(s)\lambda_\tau(t - s)ds.$$

The proof is essentially the same as Lemma A.3, so we omit it for brevity. Since $\omega_\tau(0) = 0$ due to the fact $\bar{X}_\tau(0) \le 1$ and it satisfies (2.18) at $t = 0$ that the right-hand side of the above inequality equals $F_{d,\tau+t}(t)$ by (2.10). This together with (A.14) and (A.16) immediately yields that the convolution equations (2.20) and (A.17) have same solution $\bar{X}_\tau(t)$ (if any) for all $t \ge 0$. Thus instead of analyzing (2.20) we just need to prove the existence and uniqueness of the solution to (A.17) on $[0, b]$. Let M be any strictly positive number and $S_F = \inf\{x \ge 0 : F(x) = 1\}$. From (A.15) the following derivative is bounded for all $t \in [0, S_F \wedge M/2]$:

$$\frac{d}{dy}\tilde{H}_{\tau+t}(y) = -\frac{f(F^{-1}_{d,\tau+t}(y))}{F^c(F^{-1}_{d,\tau+t}(y))} \ge -\frac{L_F}{F^c(t)} \ge -\frac{L_F}{F^c(\frac{S_F \wedge M}{2})}, \quad \text{if } 0 \le y \le F_{d,\tau+t}(t),$$

where L_F is denoted to be the Lipschitz constant of F by Assumption 2.1. We can pick $b_1 = S_F \wedge M/2$ and then for any $t \in [0, b_1]$ the function $\tilde{H}_{\tau+t}(\cdot)$ in (A.17) is Lipschitz continuous.

https://doi.org/10.1017/S0269964818000232 Published online by Cambridge University Press

Let $L = L_F/F^c((S_F \wedge M)/2)$ be the Lipschitz constant. By Assumption 2.1, there exists a $b_2 > 0$ such that

$$\kappa := \frac{1}{\mu} L[G_e(b_2) - G_e(0)] + [G(b_2) - G(0)] < 1.$$

Let $b = \min\{b_1, b_2\}$. For any $x \in C[0, b]$, define

$$\Psi(x)(t) = \bar{\mathcal{Z}}(\tau)(C_t) + \bar{Q}(\tau)G^c(t) + \frac{1}{\mu}\int_0^t \tilde{H}_{\tau+t-s}((x(t-s)-1)^+)dG_e(s) + \int_0^t (x(t-s)-1)^+ dG(s).$$

It is clear that $\Psi(x)(t)$ is continuous in t, so Ψ is a mapping from $\mathbf{C}[0, b]$ to $\mathbf{C}[0, b]$. Let $\rho(x, x') = \sup_{t\in[0,b]} |x(t) - x'(t)|$ denote the uniform distance between two functions in $\mathbf{C}[0, b]$. For any $x, x' \in C[0, b]$, we have

$$\begin{aligned}
\rho(\Psi(x), \Psi(x')) &\le \sup_{t\in[0,b]} \frac{1}{\mu}\int_0^t L|(x(t-s)-1)^+ - (x'(t-s)-1)^+|dG_e(s) \\
&\quad + \sup_{t\in[0,b]} \int_0^t |(x(t-s)-1)^+ - (x'(t-s)-1)^+|dG(s) \\
&\le \frac{1}{\mu} L\int_0^b \rho(x, x')dG_e(s) + \int_0^b \rho(x, x')dG(s) \\
&\le \kappa\rho(x, x').
\end{aligned}$$

Since $\kappa < 1$, Ψ is a contraction mapping on $\mathbf{C}[0, b]$ under the uniform topology ρ. Note that $\mathbf{C}[0, b]$ is complete under the uniform topology of ρ (cf. p. 80 in Billingsley [2]). Thus, by the contraction mapping theorem (e.g., Theorem 3.2 in Hunter and Nachtergaele [3]), Ψ has a unique fixed point x, i.e., $x = \Psi(x)$. This proves that (A.17) has a unique solution on $[0, b]$. It is also the unique solution to (2.20) on $[0, b]$ as argued in the above.

Case 2: $\bar{X}(\tau) > 1$. As in Case 1, we also have the shifted key equation (2.20). Due to the continuity there exists $b_3 > 0$ such that

$$\bar{X}_\tau(t) \ge 1 \quad \text{for all } t \in [0, b_3]. \tag{A.18}$$

For notational simplicity, denote $q_\tau(t) = (\bar{X}_\tau(t) - 1)^+$ and

$$a_\tau(t) = \int_0^t H_{\tau+s}(q_\tau(s))ds - q_\tau(t) + q_\tau(0). \tag{A.19}$$

By (2.15) and (2.20),

$$\begin{aligned}
\bar{X}_\tau(t) &= \bar{\mathcal{Z}}_\tau(0)(C_t) + q_\tau(0)G^c(t) + \int_0^t H_{\tau+t-s}(q_\tau(t-s))[1-G(s)]ds + \int_0^t q_\tau(t-s)dG(s) \\
&= \bar{\mathcal{Z}}_\tau(0)(C_t) + q_\tau(0)G^c(t) + \int_0^t H_{\tau+s}(q_\tau(s))ds - \int_0^t H_{\tau+s}(q_\tau(s))G(t-s)ds \\
&\quad + \int_0^t q_\tau(t-s)dG(s).
\end{aligned}$$

The second last term on the above equation satisfies

$$\begin{aligned}
\int_0^t H_{\tau+s}(q_\tau(s))G(t-s)ds &= \int_0^t \int_0^{t-s} H_{\tau+s}(q_\tau(s))dG(x)ds \\
&= \int_0^t \int_0^{t-x} H_{\tau+s}(q_\tau(s))dsdG(x),
\end{aligned}$$

https://doi.org/10.1017/S0269964818000232 Published online by Cambridge University Press

where the last equality follows by changing the order of integration. So we obtain

$$\bar{X}_\tau(t) = \bar{\mathcal{Z}}_\tau(0)(C_t) + q_\tau(0) + \int_0^t H_{\tau+s}(q_\tau(s))ds - \int_0^t \left[\int_0^{t-x} H_{\tau+s}(q_\tau(s))ds - q_\tau(t-x) + q_\tau(0)\right] dG(x).$$

According to the above definition of $a_\tau(t)$, we have

$$a_\tau(t) = \bar{X}_\tau(t) - q_\tau(t) - \bar{\mathcal{Z}}_\tau(0)(C_t) + \int_0^t a_\tau(t-s)dG(s).$$

By (A.18), the above becomes

$$a_\tau(t) = 1 - \bar{\mathcal{Z}}_\tau(0)(C_t) + \int_0^t a_\tau(t-s)dG(s), \quad t \in [0, b_3].$$

Let G^{n*} be the n-fold convolution of G with itself, and denote $U_G(t) = \sum_{i=0}^{\infty} G^{n*}$. The solution to the above renewal equation is

$$a_\tau(t) = \int_0^t (1 - \bar{\mathcal{Z}}_\tau(0)(C_{t-s}))dU_G(s), \quad t \in [0, b_3].$$

It is clear that $a_\tau(t)$ is continuous. Since $H_{\tau+t}(\cdot)$ is continuous, with a known $a_\tau(t)$ there exists a continuous solution $q_\tau(t)$ to the Eq. (A.19) following from Theorem II.1.1 in Miller [10].

Next, we prove the uniqueness. Assume that $q_1(t)$ and $q_2(t)$ satisfy (A.19) on the interval $[0, b_3]$. Let

$$\mathcal{L}(t) := (q_1(t) - q_2(t))^2, \quad t \in [0, b_3].$$

Then, on the interval $[0, b_3]$ we can see from (A.19) that

$$\mathcal{L}'(t) = 2[q_1(t) - q_2(t)][H_{\tau+t}(q_1(t)) - H_{\tau+t}(q_2(t))] \le 0,$$

where the last inequality is due to the fact that $H_{\tau+t}(\cdot)$ is non-increasing; see (2.16). Thus $\mathcal{L}(t)$ is non-increasing on $[0, b_3]$. Since $\mathcal{L}(0) = 0$ and $\mathcal{L}(t) \ge 0$, $\mathcal{L}(t) = 0$ for all $t \in [0, b_3]$. Hence

$$q_1(t) = q_2(t) \quad \text{for all } t \in [0, b_3].$$

Thus (2.20) only has one solution on the interval $[0, b_3]$. So we have the existence and uniqueness of the solution to (2.17) on the interval $[0, \tau + b_3]$. In fact, our analysis shows that we can further extend the solution to a point where $\bar{X}(\cdot)$ reaches 1. Starting from there, we can apply Case 1 to extend the solution to an extra interval with length b. Again, we can at least extend the unique solution of (2.17) to the interval $[0, \tau + b]$ in this case. ■

https://doi.org/10.1017/S0269964818000232 Published online by Cambridge University Press