FLUID LIMIT OF A PS-QUEUE WITH MULTISTAGE SERVICE

Frolkova, Maria; Zwart, Bert

published in
Probability in the Engineering and Informational Sciences
2019

DOI (link to publisher)
10.1017/S0269964817000456

document version
Publisher's PDF, also known as Version of record

Probability in the Engineering and Informational Sciences, **33**, 2019, 1–27.
doi:10.1017/S0269964817000456

FLUID LIMIT OF A PS-QUEUE WITH MULTISTAGE SERVICE

MARIA FROLKOVA
Department of Mathematics at Vrije Universiteit Amsterdam, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands
E-mail: m.frolkova@vu.nl

BERT ZWART
Centrum Wiskunde en Informatica, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands
and
Eindhoven University of Technology, Eindhoven, The Netherlands
E-mail: Bert.Zwart@cwi.nl

We consider a variation of the processor-sharing (PS) queue, inspired by freelance job websites where multiple freelancers compete for a single job. We develop fluid limit approximations for the overloaded PS-model with multiple (possibly infinitely many) service stages. Based on this approximation, we estimate what proportion of freelancers get the job they apply for. In addition, the PS model studied here is an instance of PS with routing and impatience, for which no Lyapunov function is known, and we suggest some partial solutions.

Keywords: fluid limits, freelance job websites, Lyapunov functions, processor-sharing, routing

1. INTRODUCTION

We consider a stylized model of a freelance job website, which has two kinds of visitors: Customers offering jobs and freelancers, or servers, looking for jobs. The key feature of such websites is that multiple servers compete for a single job there. The most common situation is competition at the stage of application, that is, to get the job. Along with that, the applicants might have to do the job, and then the one who has done it best gets paid – this is, for example, the former *modus operandi* of flightfox.com, a website for searching cheap flight connections.

Specifically, the model we consider is the following. We consider a renewal process modeling customer arrivals, and a Poisson process modeling the arrival of servers (a.k.a. freelancers) of rates λ and μ, respectively. Each customer upon arrival posts a job on the website main page and sets a patience clock that is distributed exponentially with parameter ν. Each server upon arrival picks a job from the main page at random and applies for it, in the form of leaving a comment. If there are no jobs, then the freelancer leaves. We assume that at most I applications are allowed for every job, where I can be finite or infinite. Once

 0269-9648/17 $25.00

a customer receives the Ith application or his patience expires, he should remove the job from the main page and continue communication with the applicants via private messaging. The state of this system has I components where component i is the number of jobs on the main page that have collected $i-1$ applications and are waiting for an ith application.

Observe that (using a similar argument as in Borst et al. [5]) is that this model is equivalent to a processor-sharing (PS) queue[1] where

- customers arrive according to the same process as in the freelance model,
- customers re-enter the queue for I times with independent service requirements distributed exponentially with parameter μ,
- patience times of customers are exponentially distributed with parameter ν.

The state of the PS-queue has I components, too, and the ith component is the number of customers who have entered the queue i times, or we also say "*customers at stage i of service*". This interpretation will be followed in the remainder of the paper, and we extend the model by allowing service requirements at different stages of service have different parameters μ_i (the distribution is still exponential). Note that multistage service is essentially tandem routing: From stage/class i to the next stage/class $i+1$. While in the present paper the service rate is stage-varying for every individual customer, the case of PS with the service rate varying in time universally for all customers has been also treated in the literature, see, for example, Mandelbaum, Massey, and Reiman [15], Hampshire, Harchol-Balter, and Massey [12].

The results of this paper concern fluid limit approximations of the suggested model in overload. We show that trajectories of the per-stage population process, when scaled properly, converge to solutions of a system of differential equations, which in turn stabilize to the unique invariant solution over time. Then we use the fluid limit approximation to estimate the chance for a freelancer to get a job. An interesting aspect of our model and an important ingredient of our proofs is that the model's total population is a single-stage/class PS-queue. In particular, this enables us to allow for infinitely many classes at little extra cost compared to the finite-dimensional case. The extra work is to establish a uniform convergence property of the limit deterministic model, which is easier to establish than to analyze the infinite-dimensional stochastic model directly. With the latter approach, for example, the compact containment condition is not straightforward to verify anymore since, in infinite-dimensional spaces, compact sets are not simply closed and bounded. For the necessary results for single-class PS, we refer to Gromoll, Robert, and Zwart [11], so the present paper can be viewed as a generalization of the fluid limit part of [11] to the tandem setting with multiple or infinitely many classes.

The tools used in the proof of convergence of the scaled stochastic model to a fluid limit are: A representation of the system dynamics in terms of time-changed Poisson processes as in Mandelbaum et al. [15], the relative compactness criteria of compact containment and oscillation control by Ethier and Kurtz [6], the continuous mapping theorem and the random-time change theorem. The mentioned relative compactness criteria work if the model is finite-dimensional, otherwise they work for finite-dimensional projections of the model. In the latter case, the fluid limit result by Gromoll et al. [11] for the total population implies the convergence of the entire infinite-dimensional model.

[1] A PS queue is a single server queue where the server works at a unit rate whenever there are customers present and, if there are $n>0$ customers present, each of them is being served at rate $1/n$.

https://doi.org/10.1017/S0269964817000456 Published online by Cambridge University Press

As for the convergence of fluid limits to the invariant point, in addition to generalizing the single-class proof by Gromoll et al. [11], we also discuss the method of Lyapunov functions. The model treated in this paper is an instance of a more general open problem: No Lyapunov function is known for a PS-queue with routing and impatience. We present some partial solutions to this open problem.

As mentioned before, our inspiration for this problem is drawn from a specific application. A natural next step would therefore be to incorporate the service stage in addition to the application stage. We are mostly interested in the scenario when the same job is done multiple times, which mathematically is a particlar kind of dependence of job sizes. A close phenomenon of job redundancy has been studied by Gardner et al. [8–10], who assume there are multiple copies of each job in different queues and the job is served when one of its copies is served. There are also optimization questions that arise in practice. For example, if freelancers are ranked in a way, what strategies should they follow to build and maintain a strong reputation? The majority of freelance websites exist at the cost of transaction fees, then what are the ways to increase website profits while keeping transaction fees affordable to visitors? Recently there has been more interest in related problems, cf. Adlakha, Johari, and Weintraub [1], Arcaute et al. [2], Iyer, Johari, and Moallemi [13], Wu, Bui, and Johari [17], Zhang, Johari, and Rajagopal [18].

The paper is organized as follows. In Section 2, we discuss in detail how the PS-queue with multistage service arises from the basic freelance model. In Section 3, we introduce two equivalent deterministic systems of equations that are analogues of the stochastic model and discuss their properties. We also discuss Lyapunov functions and estimate the probability of a freelancer getting a job. Section 4 specifies the fluid scaling under which the stochastic model converges to its deterministic analogues. Section 5 contains the proofs of some of the results of Sections 2 and 3. Section 6 proves the fluid limit theorem of Section 4. Finally, the appendix contains the derivation of the Lyapunov functions claimed in Section 3. In the remainder of this section, we list the notation we use throughout the paper.

Notation To define x as equal to y, we write $x := y$ or $y =: x$. We abbreviate the left-hand side and right-hand side of an equation as "LHS" and "RHS", respectively.

The standard sets are: The natural numbers $\mathbb{N} := \{1, 2, \ldots\}$, the real line $\mathbb{R} := (-\infty, \infty)$, the non-negative half-line $\mathbb{R}_+ := [0, \infty)$. We also define the index set

$$\mathcal{I} := \begin{cases} \{1, 2, \ldots, I\} & \text{if } I < \infty, \\ \mathbb{N} & \text{if } I = \infty, \end{cases}$$

and the sequence space

$$l_{1,+} := \left\{ (x_i)_{i=1}^{\infty} \colon x_i \in \mathbb{R}_+ \text{ for all } i \text{ and } \sum_{i=1}^{\infty} x_i < \infty \right\}.$$

We use boldface notations for I-dimensional vectors only. The coordinate i of an I-dimensional vector is denoted by the same symbol as the vector itself but in regular font instead of bold, and subscript i is added. Overlining and superscripts of vectors remain in their coordinates as well, for example $\overline{\mathbf{Q}}^r(t) = (\overline{Q}_i^r)_{i=1}^I(t)$. Most of the infinite-dimensional vectors that appear in the paper are elements of the space $l_{1,+}$. Both $\mathbb{R}_+^I$ in case $I < \infty$ and $l_{1,+}$ are endowed with the norm $\|\mathbf{x}\|_1 := \sum_{i=1}^I |x_i|$. We also work with (finite-dimensional) projections of I-dimensional vectors, of the type $(\overline{Q}_i^r)_{i=1}^j(t)$, $j \in \mathcal{I}$.

https://doi.org/10.1017/S0269964817000456 Published online by Cambridge University Press

For the metric space $S = \mathbb{R}$, $\mathbb{R}_+$, $\mathbb{R}_+^j$, $j \in \mathcal{I}$, or $l_{1,+}$ the notation $\mathbf{D}(\mathbb{R}_+, S)$ stands for the space of functions $f\colon \mathbb{R}_+ \to S$ that are right-continuous with left limits. This space is endowed with the Skorokhod J_1-topology.

2. STOCHASTIC MODEL

In this section, we introduce two stochastic models: A basic model of a freelance job website and a PS queue with multistage service. We then discuss in what sense the latter model generalizes the former.

2.1. Basic Model of a Freelance Job Website

There are two types of visitors on a freelance job website: Customers, who publish job descriptions, and freelancers, who apply for those jobs. We assume that new jobs appear on the website main page according to a delayed renewal process of rate λ, and that freelancers intending to find a job visit the website according to a Poisson process of rate μ. As a freelancer looking for a job visits the website, he picks a job from the main page at random and applies for it. If there are no jobs on the main page, then the freelancer immediately leaves. Each job is allowed to collect at most $I \leq \infty$ applications, where $I = \infty$ means that there is no application limit. In addition, each job has a patience time that is distributed exponentially with parameter ν. All the random elements mentioned: The arrival processes of jobs and freelancers, and patience times of different jobs are mutually independent. As soon as a job either gets I applications or its patience time expires, the customer removes the job description from the main page. In this model, our focus is on the process

$$\mathbf{Q}^{\mathrm{FL}}(t) = (Q_i^{\mathrm{FL}})_{i=1}^{I}(t), \quad t \geq 0,$$

where $Q_i^{\mathrm{FL}}(t)$ is the number of jobs on the main page that have $i-1$ applications and are waiting for an ith application at time instant t. In case the application limit I is finite, the process $\mathbf{Q}^{\mathrm{FL}}(\cdot)$ is a random element of the Skorokhod space $\mathbf{D}(\mathbb{R}_+, \mathbb{R}_+^I)$, and in case $I = \infty$, it is a random element of $\mathbf{D}(\mathbb{R}_+, l_{1,+})$. That is, the jobs are divided into infinitely many classes when $I = \infty$, but the total number of jobs is always finite.

2.2. PS-queue with Multistage Service

Now consider a PS-queue with the same arrival process of customers as in the basic freelance model. We assume that each customer of this queue should undergo $I \leq \infty$ stages of service, with stage $i+1$ starting immediately upon completion of stage i and the service requirement at stage i distributed exponentially with parameter μ_i. A customer is supposed to leave the queue upon service completion, but if his patience time expires earlier, he abandons then. As in the previous model, patience times are distributed exponentially with parameter ν. The arrival process, service requirements of all customers at all stages, and patience times of all customers are mutually independent. Here we analyze the process

$$\mathbf{Q}(t) = (Q_i)_{i=1}^{I}(t), \quad t \geq 0,$$

where $Q_i(t)$ stands for the number of customers in stage i of service at time instant t. Similarly to the population process $\mathbf{Q}^{\mathrm{FL}}(\cdot)$ of the basic freelance model, the process $\mathbf{Q}(\cdot)$ is a random element of the Skorokhod space $\mathbf{D}(\mathbb{R}_+, S)$, where $S = \mathbb{R}_+^I$ if $I < \infty$ and $S = l_{1,+}$ if $I = \infty$.

https://doi.org/10.1017/S0269964817000456 Published online by Cambridge University Press

Throughout the paper, we use the following notation. Let a random variable B_i be a generic service requirement of stage i and D – a generic patience time. We assume all the B_i's and D are defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and are mutually independent. Introduce also

$$B_j^i := \begin{cases} \sum_{l=j}^{i} B_l, & 1 \le j \le i \le I, \quad j < \infty, \\ 0, & 1 \le i < j \le I. \end{cases}$$

In particular, B_1^I denotes the total service requirement of a customer. In case the number of service stages $I = \infty$, the total service requirement might be infinite, for example if all μ_i's are the same, and it might be a proper random variable, for example if $\sum_{i=1}^{\infty} 1/\mu_i < \infty$ and $\sum_{i=1}^{\infty} 1/\mu_i^2 < \infty$, by Chebyshev's inequality. The following lemma gives a criterion to distinguish between the two scenarios.

LEMMA 2.1: *Let $I = \infty$. If $\sum_{i=1}^{\infty} 1/\mu_i = \infty$, then $B_1^{\infty} := \sum_{i=1}^{\infty} B_i = \infty$ a.s. If $\sum_{i=1}^{\infty} 1/\mu_i < \infty$, then $B_1^{\infty} < \infty$ a.s. and the distribution of B_1^{∞} is absolutely continuous.*

The proof of this result follows in Section 5, it relies on the Markov inequality.

2.3. Equivalence of the Two Models in Case All μ_i's are the Same

Suppose that, in the second model, all service stages have the same distribution parameter μ. In this case, the processes $\mathbf{Q}^{\mathrm{FL}}(\cdot)$ and $\mathbf{Q}(\cdot)$ are distributed identically. The idea is that the jobs with $i-1$ applications that are waiting for an ith application can be viewed as customers of the PS-queue who are undergoing stage i of service. The time epochs jobs receive applications indicate the completion of stages of service in the PS-queue. When a freelancer applies for a job, he picks one at random. Correspondingly, if there is a service stage completion in the PS-queue, all of the service stages that have been ongoing are equally likely to be the one that has finished. That is due to the memoryless property of the exponential distribution and because all the μ_i's are the same.

The above insight originally belongs to Borst et al. [5], who discussed the equivalence of PS and random order of service in the context of the $G/M/1$ queue. To formalize the idea they constructed a probabilistic coupling, which can be generalized in a straightforward way to the two models we consider here.

Note that, if the arrival process of customers is Poisson, the processes $\mathbf{Q}^{\mathrm{FL}}(\cdot)$ and $\mathbf{Q}(\cdot)$ are Markov and have the same infinitesimal generators, and hence have the same distribution.

2.4. Dynamic Equations

Most of the results presented in this paper are developed for the more general model of PS with multistage service. This generalized model is defined the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ mentioned earlier in the section. Denote the arrival process of customers, which is delayed renewal of rate λ, by $E(\cdot)$. These are arrivals to stage 1 of service. Let $D_i^{\mathrm{s}}(\cdot)$ stand for the process of service completions at stage i. Note that, for $i < I$, $D_i^{\mathrm{s}}(\cdot)$ is the arrival process to stage $i+1$, and if $I < \infty$, $D_I^{\mathrm{s}}(\cdot)$ is the process of departures due to total service completions. Finally, denote by $D_i^{\mathrm{a}}(\cdot)$ the process of abandonments due to impatience at stage i. Since service requirements at all stages and patience times of all customers are distributed exponentially, and since the exponential distribution is memoryless, the processes $D_i^{\mathrm{s}}(\cdot)$ and $D_i^{\mathrm{a}}(\cdot)$ are doubly stochastic Poisson with instantaneous rates $\mu_i Q_i(\cdot)/\|Q(\cdot)\|$ (zero by convention when the system is empty) and $\nu Q_i(\cdot)$, respectively. That is, the population

https://doi.org/10.1017/S0269964817000456 Published online by Cambridge University Press

process $\mathbf{Q}(\cdot) = (Q_i)_{i=1}^{\infty}(\cdot)$ can be represented as the unique solution to the following system of equations: For $t \geq 0$,

$$\begin{aligned} Q_1(t) &= Q_1(0) + E(t) - D_1^{\mathrm{s}}(t) - D_1^{\mathrm{a}}(t), \\ Q_i(t) &= Q_i(0) + D_{i-1}^{\mathrm{s}}(t) - D_i^{\mathrm{s}}(t) - D_i^{\mathrm{a}}(t), \quad i \in \mathcal{I} \setminus \{1\}, \end{aligned} \tag{1}$$

with

$$\begin{aligned} D_i^{\mathrm{s}}(t) &= \Pi_i^{\mathrm{s}}\left(\mu_i \int_0^t \frac{Q_i(u)}{\|\mathbf{Q}(u)\|_1}\, du\right), \\ D_i^{\mathrm{a}}(t) &= \Pi_i^{\mathrm{a}}\left(\nu \int_0^t Q_i(u) du\right), \end{aligned} \tag{2}$$

where $\Pi_i^{\mathrm{s}}(\cdot), \Pi_i^{\mathrm{a}}(\cdot)$ are Poisson processes of unit rate for all i, and also the initial state $\mathbf{Q}(0) = (Q_i)_{i=1}^{I}(0)$, the arrival process $E(\cdot)$ and the processes $\Pi_i^{\mathrm{s}}(\cdot)$, $\Pi_i^{\mathrm{a}}(\cdot)$ are mutually independent. The above representation in case of Poisson arrivals is proven in Mandelbaum et al. [15]; see Theorem 9.2. However, the proof of this result generalizes to an arbitrary arrival process. Finally, throughout the rest of the paper, we assume the following.

Assumption 2.2: *The system is overloaded, that is,* $\lambda \sum_{i=1}^{I} 1/\mu_i > 1$.

3. FLUID MODEL

In this section, we define and analyze a fluid model – a deterministic analogue of the PS-queue with multistage service introduced above. We use the fluid model to estimate the chance of a freelancer getting a job when the application limit I is large. In the next section, the fluid model is shown to approximate the stochastic PS-model, where the time and space are appropriately normalized.

Definition 3.1: *A function* $\mathbf{z}(\cdot) = (z_i)_{i=1}^{I}(\cdot)\colon \mathbb{R}_+ \to S$, *where* $S = \mathbb{R}_+^I$ *if* $I < \infty$ *and* $S = l_{1,+}$ *if* $I = \infty$, *that is continuous and such that* $\inf_{t \geq \delta} \|\mathbf{z}(t)\|_1 > 0$ *for any* $\delta > 0$ *is called a fluid model solution (FMS) if it solves the following system of differential equations: For* $t > 0$,

$$\begin{aligned} z_1'(t) &= \lambda - \mu_1 \frac{z_1(t)}{\|\mathbf{z}(t)\|_1} - \nu z_1(t), \\ z_i'(t) &= \mu_{i-1} \frac{z_{i-1}(t)}{\|\mathbf{z}(t)\|_1} - \mu_i \frac{z_i(t)}{\|\mathbf{z}(t)\|_1} - \nu z_i(t), \quad i \in \mathcal{I} \setminus \{1\}. \end{aligned} \tag{3}$$

When investigating properties of FMS's, we will also use an alternative description of them. Recall that, in the stochastic PS-model, B_i and D stand for the generic service requirement at stage i and patience time of a customer, and $B_i^j := \sum_{l=i}^{j} B_l$ is the cumulative service requirement from stage i up to stage j. It turns out that the system (3) is equivalent

https://doi.org/10.1017/S0269964817000456 Published online by Cambridge University Press

to the following system of integral equations: For all $i \in \mathcal{I}$ and $t \geq 0$,

$$z_i(t) = \sum_{j=1}^{i} z_j(0)\, \mathbb{P}\left\{ B_j^{i-1} \leq \int_0^t \frac{du}{\|\mathbf{z}(u)\|_1} < B_j^i,\, D > t \right\} + \lambda \int_0^t \mathbb{P}\left\{ B_1^{i-1} \leq \int_s^t \frac{du}{\|\mathbf{z}(u)\|_1} < B_1^i,\, D > t - s \right\} ds. \tag{4}$$

The two systems are equivalent in the sense that they have the same set of continuous, non-negative, non-zero outside $t = 0$ solutions.

The differential Eq. (3) capture the drift of the system, they are direct analogues of the stochastic Eqs. (1) and (2). The integral Eq. (4) mimic the evolution of the stochastic system from an individual customer's prospective. Given a customer arrived at time instant s, he is undergoing stage i of service at time instant $t \geq s$ if his patience time allows it, and if the amount of service he has received up to t covers the service requirements of the first $i-1$ stages completely and the service requirement of stage i only partially. This explains the second term in the RHS of Eq. (4). The first term has the same interpretation but in the context of customers who were present in the system at $t = 0$. Due to the memoryless property of the exponential distribution, the residual service requirements of the service stages that are ongoing at $t = 0$ are still exponentially distributed with the corresponding parameters.

A rigorous proof of the equivalence of the two descriptions of FMS's follows in Section 5. It exploits a specific relation between the distributions of the phase-type random variables B_1^i.

We now proceed with the analysis of FMS's.

THEOREM 3.2: *For any initial state* $\mathbf{z}(0)$*, a FMS exists and is unique.*

PROOF: Both existence and uniqueness follow because the total population of a multistage PS-queue is, in fact, a single-stage PS-queue with the generic service requirement $B_1^I = \sum_{i=1}^I B_i$. The latter model is studied in Gromoll et al. [11]. The summation of the equations of (4) implies that the norm $\|\mathbf{z}(\cdot)\|_1$ solves the equation

$$\|\mathbf{z}(t)\|_1 = \sum_{j=1}^{I} z_j(0)\, \mathbb{P}\left\{ B_j^I > \int_0^t \frac{du}{\|\mathbf{z}(u)\|_1}, D > t \right\} + \lambda \int_0^t \mathbb{P}\left\{ B_1^I > \int_s^t \frac{du}{\|\mathbf{z}(u)\|_1},\, D > t - s \right\} ds, \quad t \geq 0. \tag{5}$$

The last equation has a unique solution by [11, Theorem 2.2] applied to the data λ (the arrival rate), $\theta(F \times G) = \mathbb{P}\{B_1^\infty \in F, D \in G\}$ (the joint service-requirement-and-patience-time distribution) and $\zeta_0(F \times G) = \sum_{i=1}^I z_i(0)\mathbb{P}\{B_i^I \in F, D \in G\}$ (the measure-valued initial condition). The existence and uniqueness of the norm $\|\mathbf{z}(\cdot)\|_1$ in (4) implies that the individual coordinates $z_i(\cdot)$ exist and are unique as well because they are defined by $\|\mathbf{z}(\cdot)\|_1$.

Note also that, if I is finite and the initial condition is non-zero, the uniqueness follows easily from the description (3) by the Gronwall inequality: It applies because the RHS of Eq. (3) is Lipschitz continuous on sets $\{\mathbf{z} \in \mathbb{R}_+^I \colon ||\mathbf{z}||_1 \geq a\}$, $a > 0$. ■

In the next theorem, we characterize the invariant (constant) FMS.

https://doi.org/10.1017/S0269964817000456 Published online by Cambridge University Press

THEOREM 3.3: *There exists a unique invariant FMS, which is given by*

$$\begin{aligned} z_1^* &= \frac{\lambda}{\mu_1 + \nu\|\mathbf{z}^*\|_1}\,\|\mathbf{z}^*\|_1, \\ z_i^* &= \frac{\mu_{i-1}}{\mu_i + \nu\|\mathbf{z}^*\|_1}\,z_{i-1}^*, \quad i \in \mathcal{I}\setminus\{1\}, \end{aligned} \tag{6}$$

where $\|\mathbf{z}^*\|_1$ *solves*

$$f(\|\mathbf{z}^*\|_1) := \lambda \sum_{i=1}^{I} \frac{1}{\mu_i} \prod_{j=1}^{i} \frac{\mu_j}{\mu_i + \nu\|\mathbf{z}^*\|_1} = 1. \tag{7}$$

In particular, if $I = \infty$ *and* $\sum_{i=1}^{\infty} 1/\mu_i = \infty$, *then*

$$\|\mathbf{z}^*\|_1 = \frac{\lambda}{\nu}.$$

Remark 3.4: Note that, in case of infinite total service requirements, the total population of the stochastic model forms an infinite server queue $GI/M/\infty$ with arrival rate λ and service rate ν, and $\|\mathbf{z}^*\|_1$, respectively, is the invariant point of the fluid model $x'(\cdot) = \lambda - \nu x(\cdot)$ for this infinite server queue.

Proof of Theorem 3.3: By definition, an invariant FMS is non-zero. It follows from the description (3) that an invariant FMS $\mathbf{z}^* = (z_1^*, \ldots, z_I^*)$ is defined by the following system of equations:

$$\begin{aligned} &\lambda - \mu_1 \frac{z_1^*}{\|\mathbf{z}^*\|_1} - \nu z_1^* = 0, \\ &\mu_{i-1}\frac{z_{i-1}^*}{\|\mathbf{z}^*\|_1} - \mu_i \frac{z_i^*}{\|\mathbf{z}^*\|_1} - \nu z_i^* = 0, \quad i \in \mathcal{I}\setminus\{1\}. \end{aligned} \tag{8}$$

As we solve the i-th equation in the system (8) with respect to z_i^*, we obtain Eq. (6). Now, the system (6) is equivalent to

$$\begin{aligned} z_1^* &= \frac{\lambda\|\mathbf{z}^*\|_1}{\mu_1 + \nu\|\mathbf{z}^*\|_1}, \\ z_i^* &= \frac{\lambda\|\mathbf{z}^*\|_1}{\mu_i} \prod_{j=1}^{i} \frac{\mu_j}{\mu_i + \nu\|\mathbf{z}^*\|_1}, \quad i \in \mathcal{I}\setminus\{1\}. \end{aligned}$$

As we sum up over the last set of equations and divide by $\|\mathbf{z}^*\|_1$ on both sides, Eq. (7) follows.

Note that Eqs. (6) and (7) have a unique solution. Indeed, if $\sum_{i=1}^{I} 1/\mu_i < \infty$, the function $f(\cdot)$ is strictly decreasing in $(0, \infty)$ and takes all values between $\lambda\sum_{i=1}^{I} 1/\mu_i$ (which is bigger than 1 by Assumption 2.2) and 0 as it arguments runs from 0 to ∞. If $I = \infty$ and $\sum_{i=1}^{\infty} 1/\mu_i = \infty$, note that

$$f(\|\mathbf{z}^*\|_1) = \frac{\lambda}{\nu\|\mathbf{z}^*\|_1}\mathbb{P}\left\{D < \frac{B_1^{\infty}}{\|\mathbf{z}^*\|_1}\right\} = \frac{\lambda}{\nu\|\mathbf{z}^*\|_1}\mathbb{P}\{D < \infty\} = \frac{\lambda}{\nu\|\mathbf{z}^*\|_1}.$$

Hence, Eq. (7) uniquely defines the norm $\|\mathbf{z}^*\|_1$, and then Eq. (6) uniquely defines the individual coordinates z_i via $\|\mathbf{z}^*\|_1$. ■

https://doi.org/10.1017/S0269964817000456 Published online by Cambridge University Press

Finally, we show that the invariant FMS found above is asymptotically stable.

THEOREM 3.5: *Any FMS* $\mathbf{z}(t)$ *converges to the unique invariant FMS* $\mathbf{z}^*$ *as* $t \to \infty$.

When proving the above result, we again refer to the paper Gromoll et al. [11] on PS with single-stage service. As we put $\mathbf{z}(\cdot) \equiv \mathbf{z}^*$ in Eq. (5) and then take $t \to \infty$, it follows that the norm $\|\mathbf{z}^*\|_1$ of the invariant FMS should solve the equation

$$\|\mathbf{z}^*\|_1 = \lambda \mathbb{E} \min\{\|\mathbf{z}^*\|_1 B_1^I, D\}. \tag{9}$$

Note that Eq. (9) reduces to Eq. (7) as one employs the fact that the B_i's that constitute $B_1^I = \sum_{i=1}^I B_i$ and D are exponentially distributed and mutually independent. Also note that, while Eq. (5) defines FMS's in [11] for certain data specified in the proof of Theorem 3.3, Eq. (9) defines the invariant FMS in [11]. By [11, Theorem 2.4], all solutions $\|\mathbf{z}(t)\|_1$ to Eq. (5) converge to the unique solution $\|\mathbf{z}^*\|_1$ of Eq. (9) (and hence, of Eq. (7)) as $t \to \infty$. Now that we have the convergence $\|\mathbf{z}(t)\|_1 \to \|\mathbf{z}^*\|_1$ for any FMS $\mathbf{z}(\cdot)$, the coordinate-wise convergence can be shown with the use of the same ideas as in [11, Theorem 2.4]. In case of $I = \infty$, the convergence of the norm and the coordinate-wise convergence imply the convergence in $l_{1,+}$. We provide the proof in Section 5 for completeness.

Remark 3.6 Asymptotic stability of the invariant point via Lyapunov functions: For $I < \infty$, an alternative way to establish the asymptotic stability of the invariant solution to the fluid model (3) would be to suggest a Lyapunov function, that is, a function $L\colon (0,\infty)^I \to \mathbb{R}_+$ such that $L(\mathbf{z}) \to \infty$ as $\|\mathbf{z}\|_1 \to \infty$ and whose derivative with respect to the system (3) is non-positive. It is known that a PS-queue with I classes of customers, with Markovian routing and no impatience admits the entropy Lyapunov function (see Bramson [4])

$$L_{\ln}(\mathbf{z}) := \sum_{i=1}^I z_i \ln\left(\frac{z_i/\|\mathbf{z}\|_1}{z_i^*/\|\mathbf{z}^*\|_1}\right). \tag{10}$$

It can also be checked along the lines of Frolkova, Foss, and Zwart [7, Theorem 2] that a PS-queue with impatience (different rates ν_i for different classes are allowed) and no routing admits the quadratic Lyapunov function

$$\widetilde{L}_{\mathrm{qd}}(\mathbf{z}) = \sum_{i=1}^I \frac{(z_i - z_i^*)^2}{\mu_i z_i^*/\|\mathbf{z}^*\|_1}.$$

Whether there is a Lyapunov function for a PS-queue with both routing and impatience is an open problem. In the particular case of a PS-queue with multistage service, where the routing is tandem (from class i to $i+1$) and the impatience parameters are the same for all classes, the mentioned open problem does not seem to become easier. We have, however, come up with some partial solutions (all derivations follow in the appendix). For a general Markovian routing and the same impatience parameters for all classes, the entropy Lyapunov function (10) works if there are $I = 2$ classes of customers, and if there are $I > 2$ classes, it can be shown to work everywhere except for a compact set (see Lemmas A.2 and A.3 in the Appendix). In case of $I = 2$ classes, a general Markovian routing $(P_{i,j})_{i,j=1}^2$ and different impatience parameters ν_1, ν_2, the following quadratic Lyapunov function works (see Lemma A.5 in the Appendix):

$$L_{\mathrm{qd}}(\mathbf{z}) = \frac{(z_1 - z_1^*)^2}{[(1-P_{1,1})\mu_1 + P_{2,1}\mu_2]z_1^*/\|\mathbf{z}^*\|_1} + \frac{(z_2 - z_2^*)^2}{[(1-P_{2,2})\mu_2 + P_{1,2}\mu_1]z_2^*/\|\mathbf{z}^*\|_1}. \tag{11}$$

https://doi.org/10.1017/S0269964817000456 Published online by Cambridge University Press

Probability for a freelancer to get a job In the next section, we discuss in what sense the fluid model approximates the stochastic PS-model. Here we heuristically estimate the chance of a freelancer getting a job based on the fluid model and under the following additional assumptions:

(A1) the application limit I is large,

(A2) all freelancers that applied for the same job have equal chances to get the job.

Recall that, in the context of the basic model of a freelance website presented in the previous section, the ith coordinate z_i^* of the invariant FMS stands for the equilibrium number of jobs on the website that already have $i-1$ applications and are waiting for an ith application. By Eq. (8), out of λ jobs arriving per time unit, the fraction $\nu z_i^*/\lambda$ will leave with $i-1$ applications due to impatience, $i \in \mathcal{I}$. If $I < \infty$, the fraction $\mu z_I^*/(\lambda\|\mathbf{z}^*\|_1)$ will be patient enough to collect I applications. Then, by the assumption (A2), the probability for a freelancer to get a job under the application limit $I \le \infty$ is

$$P_I = \sum_{i=2}^{I} \frac{1}{i-1}\frac{\nu z_i^*}{\lambda} + \mathbb{I}\{I<\infty\}\frac{1}{I}\frac{\mu z_I^*}{\lambda\|\mathbf{z}^*\|_1}. \tag{12}$$

The next lemma suggests a simple expression/approximation for the above probability under the assumption (A1).

LEMMA 3.7: *As $I \to \infty$, then $P_I \to P_\infty$, and $P_\infty = -(\lambda/(\lambda+\mu))\ln(\lambda/(\lambda+\mu))$.*

PROOF: For the basic model of a freelance website, the invariant point Eqs. (6) and (7) can be rewritten as

$$z_i^* = \frac{\lambda}{\nu}u^{i-1}(1-u), \quad i \in \mathcal{I},$$

where u is the unique solution to

$$\frac{\lambda}{\mu}\sum_{i=1}^{I} u^i = 1. \tag{13}$$

(We make the substitution $u := \mu/(\mu+\nu\|\mathbf{z}^*\|_1)$ to obtain the equation for u.)

Note that, if $I=\infty$, then $u = \mu/(\lambda+\mu)$ and, by (12),

$$P_\infty = \sum_{i=1}^{\infty}\frac{1}{i}u^i(1-u) = -(1-u)\ln(1-u) = -\frac{\lambda}{\lambda+\mu}\ln\frac{\lambda}{\lambda+\mu}.$$

Similarly, if $I \to \infty$, then $u \to \mu/(\lambda+\mu)$ and, by (12),

$$P_I = \sum_{i=1}^{I-1}\frac{1}{i}u^i(1-u) + \frac{1}{I}u^I \to P_\infty. \qquad \blacksquare$$

Remark 3.8: As follows from the above proof, neither u nor P_I depend on the impatience parameter ν. This insensitivity is surprising given the intuition that bigger ν's, that is, smaller patience times, should result in jobs collecting less applications meaning less competition among freelancers and a higher chance to get a job.

https://doi.org/10.1017/S0269964817000456 Published online by Cambridge University Press

4. FLUID LIMIT THEOREM

In this section, we discuss under what scaling the PS-queue with multistage service converges to the fluid model introduced in Section 3.

Consider a family of stochastic PS-queues upper-indexed by positive numbers r, all of them defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let the arrival rate λ and the parameters μ_i of service stages be the same in all models (and satisfy Assumption 2.2), and the impatience parameter of model r be ν/r. Define the fluid scaled population processes

$$\overline{\mathbf{Q}}^r(t) := \mathbf{Q}^r(rt)/r, \quad t \geq 0. \tag{14}$$

We refer to weak limits along subsequences of the processes (14) as *fluid limits*. They can be characterized as solutions to the differential/integral Eqs. (3) and (4), as the next theorem asserts.

THEOREM 4.1: *Suppose that* $\overline{\mathbf{Q}}^r(0) \Rightarrow \mathbf{z}(0)$ *as* $r \to \infty$, *where the limit initial condition* $\mathbf{z}(0)$ *is deterministic. Then the processes* $\overline{\mathbf{Q}}^r(\cdot)$ *converge weakly in the Skorokhod space* $\mathbf{D}(\mathbb{R}_+, S)$, *where* $S = \mathbb{R}_+^I$ *if* $I < \infty$ *and* $S = l_{1,+}$ *if* $I = \infty$, *to the unique FMS with initial state* $\mathbf{z}(0)$.

The proof is given in Section 6. First, we show that, for any $j \in \mathcal{I}$, the family of the scaled finite-dimensional projections $(\overline{Q}_i^r)_{i=1}^j(\cdot)$ is relatively compact in $\mathbf{D}(\mathbb{R}_+, \mathbb{R}_+^j)$ by checking the compact containment and oscillation control conditions. Then we check that the weak limits along the sequences $(\overline{Q}_i^r)_{i=1}^j(\cdot)$ coincide with the FMS projection $(z_i)_{i=1}^j(\cdot)$ by deriving the fluid model Eq. (3) from the scaled versions of the stochastic dynamic Eqs. (1) and (2). Finally, if $I = \infty$, we check that the weak convergence of finite-dimensional projections and of the total population (the latter is established by Gromoll et al. [11]) implies the weak convergence in $\mathbf{D}(\mathbb{R}_+, l_{1,+})$. In particular, we establish the following property of FMS's: The partial sums $\sum_{i \leq j} z_i(\cdot)$ converge to the total population $\|\mathbf{z}(\cdot)\|_1$ uniformly on compact sets.

Remark 4.2: The above fluid limit theorem generalizes to phase-type service times. For a representation of the per-stage population process similar to that in Pender and Ko [16] and Ko and Pender [14], the proof provided in Section 6 carries through.

5. PROOFS FOR SECTIONS 2 AND 3

Here we prove Lemma 2.1, show that the descriptions (3) and (4) of the fluid model are equivalent, and provide the missing details of the proof of Theorem 3.5.

5.1. Proof of Lemma 2.1

In case $\sum_{i=1}^{\infty} 1/\mu_i = \infty$, Markov's inequality implies that, for any $j \in \mathbb{N}$ and $M > 0$,

$$\mathbb{P}\{B_1^j \leq M\} = \mathbb{P}\{e^{-B_1^j} \geq e^{-M}\} \leq e^M \mathbb{E} \prod_{i=1}^{j} e^{-B_i} = e^M \prod_{i=1}^{j} \frac{\mu_i}{1+\mu_i}$$
$$= \frac{e^M}{\prod_{i=1}^{j}(1+1/\mu_i)} \leq \frac{e^M}{1+\sum_{i=1}^{j} 1/\mu_i}.$$

Hence, $\mathbb{P}\{B_1^j \leq M\} \to 0$ as $j \to \infty$ for any $M > 0$, that is, $B_1^j \to \infty$ a.s.

https://doi.org/10.1017/S0269964817000456 Published online by Cambridge University Press

In case $\sum_{i=1}^{\infty} 1/\mu_i < \infty$, Markov's inequality implies that

$$\mathbb{P}\{B_1^\infty > M\} \leq \frac{\mathbb{E}B_1^\infty}{M} = \frac{\sum_{i=1}^{\infty} 1/\mu_i}{M} \to 0 \quad \text{as } M \to \infty,$$

and hence B_1^∞ is a proper random variable. Note that B_2^∞ is a proper random variable, too. Since B_1 is absolutely continuous and independent from B_2^∞, their sum $B_1^\infty = B_1 + B_2^\infty$ is absolutely continuous as well.

5.2. Equivalence of the Two Fluid Model Descriptions

This proof partly relies on the ideas of the proof of a similar result in Frolkova et al. [7], see Lemma 3.7, but it is more involved. In particular, it uses (and establishes) the special property (17) of the phase-type distribution.

Let a function $\mathbf{z}\colon \mathbb{R}_+ \to S$, where $S = \mathbb{R}_+^I$ if $I < \infty$ and $S = l_{1,+}$ if $I = \infty$, be continuous and non-zero outside $t = 0$.

Proof of (3) $\Rightarrow$ (4) Suppose that $\mathbf{z}(\cdot)$ is a solution to the system (3). Consider the following Cauchy problem with respect to $\mathbf{u}(\cdot)$: for $t > 0$,

$$\begin{aligned} u_1'(t) &= \lambda - \mu_1 \frac{u_1(t)}{\|\mathbf{z}(t)\|_1} - \nu u_1(t), \\ u_i'(t) &= \mu_{i-1} \frac{u_{i-1}(t)}{\|\mathbf{z}(t)\|_1} - \mu_i \frac{u_i(t)}{\|\mathbf{z}(t)\|_1} - \nu u_i(t), \quad i \in \mathcal{I} \setminus \{1\}, \\ \mathbf{u}(0) &= \mathbf{z}(0). \end{aligned} \tag{15}$$

This problem has at most one continuous solution. Indeed, let $\mathbf{u}(\cdot)$ and $\widetilde{\mathbf{u}}(\cdot)$ be two continuous solutions to the system (15). Then the difference $\mathbf{w}(\cdot) := (\mathbf{u} - \widetilde{\mathbf{u}})(\cdot)$ satisfies: for $t > 0$,

$$\begin{aligned} w_1'(t) &= -w_1(t) \left(\frac{\mu_1}{\|\mathbf{z}(t)\|_1} + \nu \right), \\ w_i'(t) &= w_{i-1}(t) \frac{\mu_{i-1}}{\|\mathbf{z}(t)\|_1} - w_i(t) \left(\frac{\mu_i}{\|\mathbf{z}(t)\|_1} + \nu \right), \quad i \in \mathcal{I} \setminus \{1\}, \\ \mathbf{w}(0) &= \mathbf{0}. \end{aligned}$$

Note that if $w_1(t) > 0$, then $w_1'(t) < 0$, and the other way around. Then $w_1(\cdot) \equiv 0$ (see, e.g. Frolkova et al. [7, Lemma 1]) and $w_2'(t) = -w_2(t)(\mu_i/\|\mathbf{z}(t)\|_1 + \nu)$, $t > 0$. To each pair $w_i(\cdot)$ and $w_{i+1}(\cdot)$, we apply the same reasoning as to $w_1(\cdot)$ and $w_2(\cdot)$, and thus obtain $\mathbf{w}(\cdot) \equiv \mathbf{0}$.

It is straightforward to check that the LHS and RHS of Eq. (4) both satisfy (15). Since a solution to the system (15) must be unique, the LHS and RHS of Eq. (4) must coincide.

https://doi.org/10.1017/S0269964817000456 Published online by Cambridge University Press

Proof of (4) ⇒ (3) Suppose now that $\mathbf{z}(\cdot)$ solves the system (4). As we differentiate the RHS of Eq. (4), it follows that, for $t > 0$,

$$\begin{aligned}
z_1'(t) &= \lambda - \mu_1 \frac{z_1(t)}{\|\mathbf{z}(t)\|_1} - \nu z_1(t),\\
z_i'(t) &= \sum_{j=1}^{i} \left(f_{B_j^{i-1}} - f_{B_j^i}\right) \left(\int_0^t \frac{du}{\|\mathbf{z}(u)\|_1}\right) \frac{\mathbb{P}\{D > t\}}{\|\mathbf{z}(t)\|_1}\\
&\quad + \lambda \int_0^t \left(f_{B_1^{i-1}} - f_{B_1^i}\right) \left(\int_s^t \frac{du}{\|\mathbf{z}(u)\|_1}\right) \frac{\mathbb{P}\{D > t-s\}}{\|\mathbf{z}(t)\|_1}\, ds\\
&\quad - \nu z_i(t), \quad i \in \mathcal{I} \setminus \{1\},
\end{aligned}$$

where $f_{B_j^i}(\cdot)$ denotes the probability density function of the phase-type random variable B_j^i with finite indices i and j.

At this stage, in order to have Eq. (3), it suffices to show that, for all $i \in \mathcal{I}$ and $t > 0$,

$$\begin{aligned}
\mu_i z_i(t) &= \sum_{j=1}^{i} f_{B_j^i} \left(\int_0^t \frac{du}{\|\mathbf{z}(u)\|_1}\right) \mathbb{P}\{D > t\}\\
&\quad + \lambda \int_0^t f_{B_1^i} \left(\int_s^t \frac{du}{\|\mathbf{z}(u)\|_1}\right) \frac{\mathbb{P}\{D > t-s\}}{\|\mathbf{z}(t)\|_1}\, ds.
\end{aligned} \tag{16}$$

In turn, in order to have Eq. (16) under the assumption (4), it suffices to show that, for all $i \in \mathcal{I}$ and $x \in \mathbb{R}$,

$$\frac{1}{\mu_i} f_{B_1^i}(x) = \mathbb{P}\{B_1^{i-1} \le x < B_1^i\},$$

or equivalently,

$$\mathbb{P}\{B_1^i > x\} = \sum_{j=1}^{i} \frac{1}{\mu_j} f_{B_1^j}(x). \tag{17}$$

We prove the identity (17) by induction: it holds for $i = 1$, assume that it holds for an $i \ge 1$, we have to check that it holds for $i+1$ as well. By the convolution formula,

$$\begin{aligned}
\mathbb{P}\{B_1^{i+1} > x\} &= \int_0^\infty \mathbb{P}\{(y + B_2^{i+1} > x\} f_{B_1}(y) dy\\
&= \int_x^\infty f_{B_1}(y) dy + \int_0^x \mathbb{P}\{B_2^{i+1} > x - y\} f_{B_1}(y) dy.
\end{aligned}$$

Now we incorporate the induction hypothesis and obtain

$$\begin{aligned}
\mathbb{P}\{B_1^{i+1} > x\} &= \mathbb{P}\{B_1 > x\} + \sum_{j=2}^{i+1} \frac{1}{\mu_j} \int_0^x f_{B_2^j}(x-y) f_{B_1}(y) dy\\
&= \frac{1}{\mu_1} f_{B_1}(x) + \sum_{j=2}^{i+1} \frac{1}{\mu_j} \int_{-\infty}^\infty f_{B_2^j}(x-y) f_{B_1}(y) dy\\
&= \frac{1}{\mu_1} f_{B_1}(x) + \sum_{j=2}^{i+1} \frac{1}{\mu_j} f_{B_1^j}(x).
\end{aligned}$$

https://doi.org/10.1017/S0269964817000456 Published online by Cambridge University Press

So the identity (17) indeed holds and implies Eq. (16); and Eq. (16), in turn, imply the system (3).

5.3. Proof of Theorem 3.5

It is shown in Gromoll et al. [11, Theorem 2.4] that, for any FMS $\mathbf{z}(\cdot)$, we have $\|\mathbf{z}(t)\|_1 \to \|\mathbf{z}^*\|_1$ as $t \to \infty$. Here we derive the coordinate-wise convergence and convergence in $l_{1,+}$ when $I = \infty$ from the convergence of the $l_{1,+}$-norm.

It follows from the fluid model description (4), that the coordinates of the invariant FMS $\mathbf{z}^*$ are uniquely defined by its norm via

$$z_i^* = \lambda \mathbb{E} \min\{\|\mathbf{z}^*\|_1 B_1^i, D\} - \lambda \mathbb{E} \min\{\|\mathbf{z}^*\|_1 B_1^{i-1}, D\}, \quad i \in \mathcal{I}. \tag{18}$$

As we compare the system (4) to the system (18), it follows that, in order to have $z_i(t) \to z_i^*$ as $t \to \infty$, it suffices to show that

$$\int_0^t f_i(s,t)\, ds \to \mathbb{E} \min\{\|\mathbf{z}^*\|_1 B_1^i, D\} \quad \text{for any } i, \tag{19}$$

where

$$f_i(s,t) := \mathbb{P}\left\{ B_1^i > \int_s^t \frac{du}{\|\mathbf{z}(u)\|_1}, D > t - s \right\}.$$

Fix an $\varepsilon \in (0, \|\mathbf{z}^*\|_1)$ and let t_ε be such that

$$\|\mathbf{z}^*\|_1 - \varepsilon \le \|\mathbf{z}(t)\|_1 \le \|\mathbf{z}^*\|_1 + \varepsilon \quad \text{for all } t \ge t_\varepsilon.$$

For any fixed s, $f_i(s,t) \to 0$ as $t \to \infty$, and then, by the dominated convergence theorem,

$$\int_0^{t_\varepsilon} f_i(s,t)\, ds \to 0 \quad \text{as } t \to \infty. \tag{20}$$

For all $t \ge t_\varepsilon$, we have

$$\begin{aligned} \int_{t_\varepsilon}^t f_i(s,t)\, ds &\le \int_{t_\varepsilon}^t \mathbb{P}\left\{ B_1^i > \int_s^t \frac{du}{\|\mathbf{z}^*\|_1 + \varepsilon}, D > t - s \right\} ds \\ &\le \int_0^{t-t_\varepsilon} \mathbb{P}\left\{ \min\{(\|\mathbf{z}^*\|_1 + \varepsilon) B_1^i, D\} \ge s \right\} ds, \end{aligned}$$

which, in combination with the convergence (20), implies that

$$\limsup_{t\to\infty} \int_0^t f_i(s,t) ds \le \mathbb{E} \min\{(\|\mathbf{z}^*\|_1 + \varepsilon) B_1^i, D\}.$$

Similarly, we obtain

$$\liminf_{t\to\infty} \int_0^t f_i(s,t) ds \ge \mathbb{E} \min\{(\|\mathbf{z}^*\|_1 - \varepsilon) B_1^i, D\}.$$

As we take $\varepsilon \to 0$ in the last two equations, the convergence (19) follows.

https://doi.org/10.1017/S0269964817000456 Published online by Cambridge University Press

Now we check that, if $I = \infty$, then $\mathbf{z}(t) \to \mathbf{z}^*$ in $l_{1,+}$ as $t \to \infty$. Since the coordinates of $\mathbf{z}(t)$ and $\mathbf{z}^*$ are non-negative, we have, for any $j \in \mathbb{N}$,

$$\|\mathbf{z}(t) - \mathbf{z}^*\|_1 = \sum_{i=1}^{\infty} |z_i(t) - z_i^*| \le \sum_{i \le j} |z_i(t) - z_i^*| + \sum_{i>j} z_i(t) + \sum_{i>j} z_i^*,$$

where

$$\sum_{i>j} z_i(t) = \|\mathbf{z}(t)\|_1 - \sum_{i \le j} z_i(t) = (\|\mathbf{z}(t)\|_1 - \|\mathbf{z}^*\|_1) - \sum_{i \le j} (z_i(t) - z_i^*) + \sum_{i>j} z_i^*,$$

and hence,

$$\|\mathbf{z}(t) - \mathbf{z}^*\|_1 \le 2 \sum_{i \le j} |z_i(t) - z_i^*| + |\|\mathbf{z}(t)\|_1 - \|\mathbf{z}^*\|_1| + 2 \sum_{i>j} z_i^*. \tag{21}$$

By the last inequality, the finiteness of $\sum_{i=1}^{\infty} z_i^*$, the convergence $\|\mathbf{z}(t)\|_1 \to \|\mathbf{z}^*\|_1$ and the coordinate-wise convergence $z_i(t) \to z_i^*$ imply the convergence $\mathbf{z}(t) \to \mathbf{z}^*$ in $l_{1,+}$.

6. PROOF OF THEOREM 4

This proof makes use of the fact that the total population of the PS-queue with multistage service forms a PS-queue with single-stage service. In particular, the fluid limit result by Gromoll et al. [11] for the latter type of PS-queues implies that $\|\overline{\mathbf{Q}}^r(\cdot)\|_1$ converges weakly in $\mathbf{D}(\mathbb{R}_+, \mathbb{R}_+)$ to the unique solution $\|\mathbf{z}(\cdot)\|_1$ of Eq. (5). (To be precise, the convergence of the total population process follows by [11, Theorem 2.3] applied to the following sequence of arrival processes, joint service-requirement-and-patience-time distributions and measure-valued initial conditions:

$$E^r(\cdot) := E(\cdot),$$
$$\theta^r(F \times G) := \mathbb{P}\{B_1^I \in F, D \in G\},$$
$$\mathcal{Z}^r(0) := \sum_{i=1}^{I} \sum_{l=1}^{Q_i^r(0)} \delta^+_{(B_{i,l}^I, D_{i,l})},$$

where the collections $(B_{i,l}^I, D_{i,l})_{i=1}^I$ are i.i.d. copies of $(B_i^I, D_i)_{i=1}^I$ and the D_i's are i.i.d. copies of D that are also independent from $(B_i^I)_{i=1}^I$.)

The remainder of the proof consists of three parts. First, we show that for any $j \in \mathcal{I}$, the family of the fluid scaled projections $(\overline{Q}_i^r)_{i=1}^j(\cdot)$ is $\mathbf{C}$-tight in $\mathbf{D}(\mathbb{R}_+, \mathbb{R}_+^j)$, that is, that any sequence of the projections $(\overline{Q}_i^r)_{i=1}^j(\cdot)$ such that $r \to \infty$ has a subsequence that converges weakly in $\mathbf{D}(\mathbb{R}_+, \mathbb{R}_+^j)$ and all such weak limits are continuous. Then we check that any weak limit of the scaled projections solves the first j equations in (3), that is, any such weak limit is the projection of the FMS. The third part of the proof is only necessary in case $I = \infty$: we show that the weak convergence of the finite-dimensional projections and of the total population of the scaled queue to those of the FMS imply the weak convergence to the FMS in $\mathbf{D}(\mathbb{R}_+, l_{1,+})$.

https://doi.org/10.1017/S0269964817000456 Published online by Cambridge University Press

Throughout the proof, we use the following representation of the processes $\mathbf{Q}^r(\cdot)$ (they all are defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$): for $t \geq 0$,

$$\begin{aligned} Q_1^r(t) &= Q_1^r(0) + E(t) - D_1^{r,\mathrm{s}}(t) - D_1^{r,\mathrm{a}}(t), \\ Q_i^r(t) &= Q_i^r(0) + D_{i-1}^{r,\mathrm{s}}(t) - D_i^{r,\mathrm{s}}(t) - D_i^{r,\mathrm{a}}(t), \quad 2 \leq i \leq I, \end{aligned} \tag{22}$$

with

$$\begin{aligned} D_i^{r,\mathrm{s}}(t) &= \Pi_i^{\mathrm{s}}\left(\mu_i \int_0^t \frac{Q_i^r(u)}{\|\mathbf{Q}^r(u)\|_1}\, du\right), \\ D_i^{r,\mathrm{a}}(t) &= \Pi_i^{\mathrm{a}}\left(\frac{\nu}{r}\int_0^t Q_i^r(u) du\right), \end{aligned} \tag{23}$$

where the processes $E(\cdot)$ and $\Pi_i^{\mathrm{s}}(\cdot), \Pi_i^{\mathrm{a}}(\cdot)$ are the same as in the dynamic Eqs. (1) and (2), except that this time we assume them to be independent from the family of the initial states $\mathbf{Q}^r(0)$.

6.1. C-tightness of Finite-Dimensional Projections

Fix a $j \in \mathcal{I}$. In order to prove that the family of the processes $(\overline{Q}_i^r)_{i=1}^j(\cdot)$ is $\mathbf{C}$-tight in $\mathbf{D}(\mathbb{R}_+, \mathbb{R}_+^j)$, it suffices to show that the following two properties hold (see Ethier and Kurtz [6]): For any $T > 0$ and $\varepsilon > 0$, there exists an $M < \infty$ and a $\delta > 0$ such that

$$\liminf_{r\to\infty} \mathbb{P}\left\{\sum_{i=1}^j \overline{Q}_i^{\,r}(T) \leq M\right\} \geq 1 - \varepsilon, \tag{24}$$

and

$$\liminf_{r\to\infty} \mathbb{P}\left\{\sup_{\substack{s,t\in[0,T],\\ |s-t|<\delta}} \sum_{i=1}^j |\overline{Q}_i^{\,r}(s) - \overline{Q}_i^{\,r}(t)| \leq \varepsilon\right\} \geq 1 - \varepsilon. \tag{25}$$

The compact containment condition (24) follows easily by the upper bound

$$\sum_{i=1}^j \overline{Q}_i^r(T) \leq \|\overline{\mathbf{Q}}^r(T)\|_1 \leq \|\overline{\mathbf{Q}}^r(0)\|_1 + E(rT)/r \Rightarrow \|\mathbf{z}(0)\|_1 + \lambda T \quad \text{as } r \to \infty.$$

To establish the oscillation control condition (25), it is enough to have oscillations of the scaled departure processes $D_i^{r,\mathrm{s}}(r\cdot)/r$ and $D_i^{r,\mathrm{a}}(r\cdot)/r$ bounded.

Define the modulus of continuity for functions $x\colon \mathbb{R}_+ \to \mathbb{R}$,

$$\omega(x, T, \delta) := \sup\{|x(s) - x(t)|\colon s, t \in [0, T], |s - t| < \delta\}.$$

First we estimate oscillations of $D_i^{r,\mathrm{s}}(r\cdot)/r$. We have, for all $t \geq s \geq 0$,

$$\left|\frac{D_i^{r,\mathrm{s}}(rs)}{r} - \frac{D_i^{r,\mathrm{s}}(rt)}{r}\right| \leq |G_i^{r,\mathrm{s}}(s)| + |G_i^{r,\mathrm{s}}(t)| + \mu_i \int_s^t \frac{\overline{Q}_i^{\,r}(u)}{\|\overline{\mathbf{Q}}^{\,r}(u)\|_1}\, du,$$

where, for all $t \geq 0$,

$$G_i^{r,\mathrm{s}}(t) := \frac{1}{r}\Pi_i^{\mathrm{s}}\left(r\mu_i \int_0^t \frac{\overline{Q}_i^{\,r}(u)}{\|\overline{\mathbf{Q}}^{\,r}(u)\|_1}\, du\right) - \mu_i \int_0^t \frac{\overline{Q}_i^{\,r}(u)}{\|\overline{\mathbf{Q}}^{\,r}(u)\|_1}\, du. \tag{26}$$

https://doi.org/10.1017/S0269964817000456 Published online by Cambridge University Press

Then

$$\omega\left(\frac{D_i^{r,\mathrm{s}}(r\cdot)}{r}, T, \delta\right) \le 2 \sup_{t\in[0,\mu_i T]} \left|\frac{\Pi_i^{\mathrm{s}}(rt)}{r} - t\right| + \delta. \tag{27}$$

Now we switch to $D_i^{r,\mathrm{a}}(r\cdot)/r$. Consider a family of $GI/M/\infty$-queues with a common arrival process $E(\cdot)$, queue r starting with $\|\mathbf{Q}^r(0)\|_1$ customers, and service times in queue r being patience times of the corresponding customers in the r-th PS-queue with multistage service. Denote the departure process of the r-th $GI/M/\infty$-queue by $\widetilde{D}^r(\cdot)$. We have, for all i and $s, t \ge 0$,

$$\left|\frac{D_i^{r,\mathrm{a}}(rs)}{r} - \frac{D_i^{r,\mathrm{a}}(rt)}{r}\right| \le \left|\frac{\widetilde{D}^r(rs)}{r} - \frac{\widetilde{D}^r(rt)}{r}\right|,$$

and hence,

$$\omega(D_i^{r,\mathrm{a}}(r\cdot)/r, T, \delta) \le \omega(\widetilde{D}^r(r\cdot)/r, T, \delta). \tag{28}$$

By, for example, Gromoll et al. [11], where the $GI/M/\infty$ queue is included as the special case of the PS-queue with infinite service times, the scaled processes $\widetilde{D}^r(r\cdot)/r$ converge weakly in the Skorokhod space $\mathbf{D}(\mathbb{R}_+, \mathbb{R}_+)$ to a continuous limit, which we denote by $\widetilde{D}(\cdot)$. Since the modulus of continuity $\omega(\cdot, T, \delta)$ as a function on $\mathbf{D}(\mathbb{R}_+, \mathbb{R})$ is continuous at any continuous $x(\cdot)$, we have, by the continuous mapping theorem,

$$\omega(\widetilde{D}^r(r\cdot)/r, T, \delta) \Rightarrow \omega(\widetilde{D}(\cdot), T, \delta) \quad \text{as } r \to \infty. \tag{29}$$

Since continuity implies uniform continuity on compact sets, we also conclude that

$$\omega(\widetilde{D}(\cdot), T, \delta) \Rightarrow 0 \quad \text{as } \delta \to 0. \tag{30}$$

Finally, as we put together the FLLN for the process $E(\cdot)$, the bound (27) and the FLLN's for the processes $\Pi_i^{\mathrm{s}}(\cdot)$, and also the results (28)–(30), it follows that one can pick a δ such that the bound (25) holds.

6.2. Convergence to the FMS Projections

Consider a weak limit $(\widetilde{Q}_i)_{i=1}^j(\cdot)$ along a subsequence $(\overline{Q}_i^q)_{i=1}^j(\cdot)$, $q \to \infty$. By the first part of the proof, this weak limit is a.s. continuous, and now we show that it a.s. coincides with the FMS projection $(z_i)_{i=1}^j(\cdot)$. Note that we have the joint weak convergence

$$((\overline{Q}_i^q)_{i=1}^j(\cdot), \|\overline{\mathbf{Q}}^q(\cdot)\|_1) \Rightarrow ((\widetilde{Q}_i)_{i=1}^j(\cdot), \|\mathbf{z}(\cdot)\|_1) \quad \text{as } q \to \infty$$

in $\mathbf{D}(\mathbb{R}_+, \mathbb{R}_+^j) \times \mathbf{D}(\mathbb{R}_+, \mathbb{R}_+)$ since the limit of the total population is deterministic.

Consider the mappings $\varphi_l \colon \mathbf{D}(\mathbb{R}_+, \mathbb{R}_+^j) \times \mathbf{D}(\mathbb{R}_+, \mathbb{R}_+) \to \mathbf{D}(\mathbb{R}_+, \mathbb{R})$ given by

$$\begin{aligned}
\varphi_1((y_i)_{i=1}^j, x)(t) =: {} & y_1(t) - y_1(0) - \lambda t + \mu_1 \int_0^t \frac{y_1(u)}{x(u)}\, du + \nu \int_0^t y_1(u)\, du, \\
\varphi_l((y_i)_{i=1}^j, x)(t) =: {} & y_l(t) - y_l(0) - \mu_{l-1} \int_0^t \frac{y_{l-1}(u)}{x(u)}\, du + \mu_l \int_0^t \frac{y_l(u)}{x(u)}\, du \\
& + \nu \int_0^t y_l(u)\, du, \quad 2 \le l \le j.
\end{aligned}$$

https://doi.org/10.1017/S0269964817000456 Published online by Cambridge University Press

These mappings are continuous at any continuous $(y_i)_{i=1}^j(\cdot)$ and $x(\cdot)$ such that $x(\cdot)$ is non-zero outside $t=0$. Then, by the continuous mapping theorem, for all $l=1,\dots,j$,

$$\varphi_l((\overline{Q}^q)_{i=1}^j, \|\overline{\mathbf{Q}}^q\|_1) \Rightarrow \varphi_l((\widetilde{Q})_{i=1}^j, \|\mathbf{z}\|_1) \quad \text{as } q\to\infty. \tag{31}$$

On the other hand, it follows from the stochastic dynamics (22) and (23) that, for all q and all $t\geq 0$,

$$\begin{aligned}
\varphi_1((\overline{Q}_i^q)_{i=1}^j, \|\overline{\mathbf{Q}}^q\|_1)(t) &= (E(qt)/q-\lambda t) - G_1^{q,\mathrm{s}}(t) - G_1^{q,\mathrm{a}}(t),\\
\varphi_l((\overline{Q}_i^q)_{i=1}^j, \|\overline{\mathbf{Q}}^q\|_1)(t) &= G_{l-1}^{q,\mathrm{s}}(t) - G_l^{q,\mathrm{s}}(t) - G_l^{q,\mathrm{a}}(t), \quad 2\leq l\leq j,
\end{aligned} \tag{32}$$

where, for all i and all $t\geq 0$,

$$G_i^{q,\mathrm{a}}(t) := \frac{1}{q}\Pi_i^{\mathrm{a}}\left(q\nu\int_0^t \overline{Q}_i^q(u)du\right) - \nu\int_0^t \overline{Q}_i^q(u)du,$$

and the processes $G_i^{q,\mathrm{s}}(\cdot)$ were defined earlier by Eq. (26).

Next we use the following result (see, e.g. Billingsley [3]).

PROPOSITION 6.1 (Random time change theorem): *Consider stochastic processes $X^q(\cdot)\in\mathbf{D}(\mathbb{R}_+,S)$, where S is a complete and separable metric space, and non-decreasing stochastic processes $\Phi^q(\cdot)\in\mathbf{D}(\mathbb{R}_+,\mathbb{R}_+)$. Assume that the joint convergence $(X^q,\Phi^q)(\cdot)\Rightarrow(X,\Phi)(\cdot)$ holds as $q\to\infty$, and that the limits $X(\cdot)$ and $\Phi(\cdot)$ are a.s. continuous. Then $X^q(\Phi^q(\cdot))\Rightarrow X(\Phi(\cdot))$ in $\mathbf{D}(\mathbb{R}_+,S)$ as $q\to\infty$.*

Put $X^q(t)=\Pi_i^{\mathrm{s}}(qt)/q-t$ and $\Phi^q(t)=\nu\int_0^t\overline{Q}_i^q(u)du$ for all $t\geq 0$. The marginal weak limits of these processes are $X(\cdot)\equiv 0$ and $\Phi(\cdot)=\nu\int_0^{\cdot}\widetilde{Q}_i(u)du$, respectively. Since one of the marginal limits is deterministic, we actually have the joint weak convergence, and then Proposition 6.1 implies that, as $q\to\infty$,

$$G_i^{q,\mathrm{a}}(\cdot)\Rightarrow 0 \quad \text{in } \mathbf{D}(\mathbb{R}_+,\mathbb{R}). \tag{33}$$

Similarly,

$$G_i^{q,\mathrm{s}}(\cdot)\Rightarrow 0 \quad \text{in } \mathbf{D}(\mathbb{R}_+,\mathbb{R}). \tag{34}$$

As we put the relations (32) and the convergence results (33), (34), (31) together, it follows that

$$\varphi_l((\widetilde{Q}_i)_{i=1}^j, \|\mathbf{z}\|_1)\equiv 0 \quad \text{a.s.}, \quad 1\leq l\leq j.$$

Along the lines of the proof of the implication (3) $\Rightarrow$ (4) in Section 5, we can show that the last display yields $\widetilde{Q}(\cdot)\equiv z_i(\cdot)$ a.s., $1\leq i\leq j$.

6.3. Convergence to the FMS if $I=\infty$

By Gromoll et al. [11] and the first two parts of the proof, we have $\|\overline{\mathbf{Q}}^r(\cdot)\|_1\Rightarrow\|\mathbf{z}(\cdot)\|_1$ in $\mathbf{D}(\mathbb{R}_+,\mathbb{R}_+)$ and $(\overline{Q}_i^r)_{i=1}^j(\cdot)\Rightarrow(z_i)_{i=1}^j(\cdot)$ in $\mathbf{D}(\mathbb{R}_+,\mathbb{R}_+^j)$ for any $j\in\mathcal{I}$. If $I=\infty$, we still need to prove that $\overline{\mathbf{Q}}^r(\cdot)\Rightarrow\mathbf{z}(\cdot)$ in $\mathbf{D}(\mathbb{R}_+,l_1)$. Note that convergence in the Skorokhod topology

https://doi.org/10.1017/S0269964817000456 Published online by Cambridge University Press

is equivalent to uniform convergence on compact sets if the limit is continuous. Hence, for any $T > 0$ and $j \in \mathbb{N}$, we have

$$\sup_{t\in[0,T]} |\|\overline{\mathbf{Q}}^r(t)\|_1 - \|\mathbf{z}(t)\|_1| \Rightarrow 0, \quad \sup_{t\in[0,T]} \sum_{i=1}^{j} |\overline{Q}_i^r(t) - z_i(t)| \Rightarrow 0,$$

and we need to check that

$$\sup_{t\in[0,T]} \|\overline{\mathbf{Q}}^r(t) - \mathbf{z}(t)\|_1 \Rightarrow 0.$$

In analogy with the inequality (21), for any $j \in \mathbb{N}$,

$$\sup_{t\in[0,T]} \|\overline{\mathbf{Q}}^r(t) - \mathbf{z}(t)\|_1 \le 2 \sup_{t\in[0,T]} \sum_{i\le j} |\overline{Q}_i^r(t) - z_i(t)| + \sup_{t\in[0,T]} |\|\overline{\mathbf{Q}}^r(t)\|_1 - \|\mathbf{z}(t)\|_1| + 2 \sup_{t\in[0,T]} \sum_{i>j} z_i(t).$$

Hence, it suffices to check that, for any $\varepsilon > 0$, there exists a j such that

$$\sup_{t\in[0,T]} \sum_{i>j} z_i(t) \le \varepsilon. \tag{35}$$

By the fluid model description (4), for all indices $j \ge j_0$,

$$\sup_{t\in[0,T]} \sum_{i>j} z_i(t) \le \overbrace{\sup_{t\in[0,T]} \sum_{i\le j_0} z_i(0)\mathbb{P}\left\{B_i^j \le \int_0^t \frac{du}{\|\mathbf{z}(u)\|_1} < B_i^\infty\right\}}^{=: F_j} + \sum_{i>j_0} z_i(0) + \underbrace{\lambda \sup_{t\in[0,T]} \int_0^t \mathbb{P}\left\{B_1^j \le \int_s^t \frac{du}{\|\mathbf{z}(u)\|_1} < B_1^\infty\right\} ds}_{=: G_j}.$$

Fix a j_0 such that

$$\sum_{i>j_0} z_i(0) \le \frac{\varepsilon}{3}.$$

Next consider G_j. If $B_1^\infty = \infty$ a.s., then, for any $\delta \in (0, T)$,

$$G_j = \sup_{t\in[0,T]} \int_0^t \mathbb{P}\left\{B_1^j \le \int_s^t \frac{du}{\|\mathbf{z}(u)\|_1}\right\} ds$$

$$\le \delta + \sup_{t\in[\delta,T]} \int_\delta^T \mathbb{P}\left\{B_1^j \le \underbrace{\int_\delta^T \frac{du}{\|\mathbf{z}(u)\|_1}}_{=: M_\delta}\right\} ds \le \delta + T\mathbb{P}\{B_1^j \le M_\delta\},$$

and since $M_\delta < \infty$ and $B_1^j \to \infty$ a.s., we have

$$\limsup_{j\to\infty} G_j \le \delta, \quad \delta \in (0, T).$$

https://doi.org/10.1017/S0269964817000456 Published online by Cambridge University Press

If $B_1^\infty < \infty$ a.s., then $\int_s^t du/\|\mathbf{z}(u)\|_1 < \infty$for all $t \geq s > 0$ and

$$G_j \leq T \sup_{x \in \mathbb{R}_+} |\mathbb{P}\{B_1^j \leq x\} - \mathbb{P}\{B_1^\infty \leq x\}| \to 0 \quad \text{as } j \to \infty$$

because $B_1^j \to B_1^\infty$ a.s. as $j \to \infty$ and the distribution function of B_1^∞ is continuous. Hence, in both cases, there exists a j_1 such that

$$\lambda G_j \leq \frac{\varepsilon}{3}, \quad j \geq j_1.$$

Finally, consider F_j. If $\mathbf{z}(0) = 0$, then $F_j = 0$. Otherwise $\int_0^t du/\|\mathbf{z}(u)\|_1 < \infty$ for all $t \geq 0$ and F_j can be treated in the same way as G_j. By considering the two cases $B_1^\infty = \infty$ a.s. and $B_1^\infty < \infty$ a.s. separately, one can prove that there exists a j_2 such that

$$F_j \leq \frac{\varepsilon}{3}, \quad j \geq j_2.$$

With any $j \geq \max(j_0, j_1, j_2)$, we have (35), and the proof of the theorem is now complete.

References

1. Adlakha, S., Johari, R., & Weintraub, G. Y. (2015). Equilibria of dynamic games with many players: Existence, approximation, and market structure. *Journal of Economic Theory* 156: 269–316.
2. Arcaute, E., Dyagilev, K., Johari, R., & Mannor, S. (2013). Dynamics in tree formation games. *Games and Economic Behavior* 79: 1–29.
3. Billingsley, P. (1999). *Convergence of probability measures*, 2nd ed. Series in Probability and Statistics. New York: Wiley.
4. Bramson, M. (1996). Convergence to equilibria for fluid models of head-of-the-line proportional processor sharing Queueing Networks. *Queueing Systems* 23(1–4): 1–26.
5. Borst, S.C., Boxma, O.J., Morrison, J.A., & Nunez Queija, R. (2003). The equivalence between processor sharing and service in random order. *Operations Research Letters* 31(4): 254–262.
6. Ethier, S.N. & Kurtz, T.G. (1986). *Markov processes: characterization and convergence.* New York: Wiley.
7. Frolkova, M., Foss, S., & Zwart, B. (2012). Fluid limits for an ALOHA-type model with impatient customers. *Queueing Systems* 72: 69–101.
8. Gardner, K., Zbarsky, S., Doroudi, S., Harchol-Balter, M., Hyytia, E., & Scheller-Wolf, A. (2016). Queueing with redundant requests: exact analysis. *Queueing Systems* 83(3): 227–259.
9. Gardner, K., Zbarsky, S., Velednitsky, M., Harchol-Balter, M., & Scheller-Wolf, A. (2016). Understanding Response Time in the Redundancy-d System. Workshop on Mathematical Performance Modeling and Analysis (MAMA 2016), Antibes Juan-les-Pins, France.
10. Gardner, K., Harchol-Balter, M., & Scheller-Wolf, A. (2016). A Better Model for Job Redundancy: Decoupling Server Slowdown and Job Size. IEEE Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS 2016), London, UK.
11. Gromoll, H.C., Robert, Ph., & Zwart, B. (2008). Fluid limits for processor sharing queues with impatience. *Mathematics of Operations Research* 33: 375–402.
12. Hampshire, R.C., Harchol-Balter, M., & Massey, W.A. (2006). Fluid and diffusion limits for transient sojourn times of processor sharing queues with time varying rates. *Queueing Systems* 53: 19–30.
13. Iyer, K., Johari, R., & Moallemi, C.C. (2014). Information aggregation and allocative efficiency in smooth markets. *Management Science* 60(10): 2509–2524.
14. Ko, Y.M. & Pender, J. (2016). Strong approximations for time-varying infinite-server queues with non-renewal arrival and service processes. Submitted for publication, preprint available at https://people.orie.cornell.edu/jpender/MAP_MAP_INF.pdf.
15. Mandelbaum, A., Massey, W.A., & Reiman, M.I. (1998). Strong approximations for Markovian service networks. *Queueing Systems* 30: 149–201.
16. Pender, J. & Ko, Y.M. (2016). Approximations for the queue length distributions of time-varying many-server queues. Submitted for publication, preprint available at https://people.orie.cornell.edu/jpender/Phase_paper.pdf.

https://doi.org/10.1017/S0269964817000456 Published online by Cambridge University Press

17. Wu, Y., Bui, L., & Johari, R. (2012). Heavy traffic approximation of equilibria in resource sharing games. *IEEE Journal on Selected Areas in Communications* 30(11): 2200–2209.
18. Zhang, B., Johari, R., & Rajagopal, R. (2015). Competition and coalition formation of renewable power producers. *IEEE Transactions on Power Systems* 30(3): 1624–1632.

APPENDIX

Here we present our attempts to find a Lyapunov funtion for the (finite) system of differential equations (recall that $\mathcal{I} = \{1, 2, \dots, I\}$)

$$z_i'(t) = \lambda_i + \sum_{j=1}^{I} P_{j,i}\mu_j \frac{z_j(t)}{\|\mathbf{z}(t)\|} - \mu_i \frac{z_i(t)}{\|\mathbf{z}(t)\|} - \nu_i z_i(t), \quad i \in \mathcal{I}, \quad t > 0. \tag{A.1}$$

The system (A.1) is a generalisation of the fluid model (3) and a deterministic analogue of a PS-queue with $I < \infty$ classes of customers, where λ_i is the arrival rate to class i, $1/\mu_i$ and ν_i are the mean service time and abandonment rate of a class i customer, respectively, and $P_{i,j}$ is the probability that a class i customer, upon finishing service, is rerouted to class j. Naturally, we assume that the $P_{i,j}$'s form a sub-stochastic matrix:

$$\text{for any } i, \quad \sum_{j=1}^{I} P_{i,j} \le 1.$$

Additionally, we assume that the system (A.1) is overloaded and has a unique invariant solution $\mathbf{z}^*$ (as in the particular case (3)), that is, there exists a unique solution to

$$\lambda_i + \sum_{j=1}^{I} P_{j,i}\mu_j \frac{z_j^*}{\|\mathbf{z}^*\|} - \mu_i \frac{z_i^*}{\|\mathbf{z}^*\|} - \nu_i z_i^* = 0, \quad i \in \mathcal{I}. \tag{A.2}$$

We have tested two kinds of candidates for a Lyapunov function for (A.1). The first one is the entropy function (10). It turned out to work for two classes with the same abandonment rates; and if there are more than two classes, it works everywhere except for a compact set. The second candidate is a quadratic function. It works for two classes with different abandonment rates. The details follow below.

Entropy Lyapunov function Note that the entropy function $L_{\ln}(\cdot)$ given by (10) is non-negative on $(0, \infty)^I$ since $L_{\ln}(\mathbf{z})/\|\mathbf{z}\|$ is the Kullback–Leibler distance between the distributions $\{z_i/\|\mathbf{z}\|\}_{i=1}^I$ and $\{z_i^*/\|\mathbf{z}^*\|\}_{i=1}^I$. The next two lemmas assume that the abandonment rates are the same in all classes and check the sign of the derivative of $L_{\ln}(\cdot)$ with respect to (A.1), which is given by

$$L'_{\ln}(\mathbf{z}) := \sum_{i=1}^{I} R_i(\mathbf{z}) \ln\left(\frac{z_i/\|\mathbf{z}\|}{z_i^*/\|\mathbf{z}^*\|}\right), \tag{A.3}$$

where

$$\begin{aligned} R_i(\mathbf{z}) &:= \lambda_i + \sum_{j=1}^{I} P_{j,i}\mu_j \frac{z_j}{\|\mathbf{z}\|} - \mu_i \frac{z_i}{\|\mathbf{z}\|} - \nu z_i \\ &= \sum_{j=1}^{I} P_{j,i}\mu_j \left(\frac{z_j(t)}{\|\mathbf{z}(t)\|} - \frac{z_j^*}{\|\mathbf{z}^*\|}\right) - \mu_i \left(\frac{z_i(t)}{\|\mathbf{z}(t)\|} - \frac{z_i^*}{\|\mathbf{z}^*\|}\right) - \nu(z_i(t) - z_i^*). \end{aligned}$$

First, we consider the case of two classes.

https://doi.org/10.1017/S0269964817000456 Published online by Cambridge University Press

LEMMA A.2: *If $I = 2$ and $\nu_1 = \nu_2 = \nu$, then $L'_{\ln}(\cdot) \le 0$ on $(0,\infty)^2$.*

PROOF: Fix a $\mathbf{z} \in (0,\infty)^2$ and, to shorten notation, put

$$p_i := z_i/\|\mathbf{z}\|, \quad q_i := z_i^*/\|\mathbf{z}^*\|, \quad i,j = 1,2.$$

We make the following rearrangements in (A.3):

$$\begin{aligned} L'_{\ln}(z) &= \sum_{i=1}^{2} \left(\sum_{j=1}^{2} P_{j,i}\mu_j(p_j - q_j) - \mu_i(p_i - q_i) - \nu(z_i - z_i^*) \right) \ln(p_i/q_i) \\ &= \Sigma_1 + \Sigma_2 - \nu\Sigma_3, \end{aligned}$$

where

$$\begin{aligned} \Sigma_1 &= \sum_{i=1}^{2} (P_{i,i} - 1)\mu_i(p_i - q_i)\ln(p_i/q_i), \\ \Sigma_2 &= P_{2,1}\mu_2(p_2 - q_2)\ln(p_1/q_1) + P_{1,2}\mu_1(p_1 - q_1)\ln(p_2/q_2) \\ \Sigma_3 &= \|\mathbf{z}\| \sum_{i=1}^{2} p_i \ln(p_i/q_i) + \|\mathbf{z}^*\| \sum_{i=1}^{2} q_i \ln(q_i/p_i). \end{aligned}$$

Since $(p_i - q_i)$ and $\ln(p_i/q_i) = \ln(p_i) - \ln(q_i)$ are of the same sign, and $P_{i,i} \le 1$, we have $\Sigma_1 \le 0$. Now note that $p_i \le q_i$ implies that $p_{3-i} \ge q_{3-i}$. Then $(p_i - q_i)$ and $\ln(p_{3-i}/q_{3-i})$ are of different signs, and hence $\Sigma_2 \le 0$. Finally, $\Sigma_3 \ge 0$ because it is a sum of two Kullback–Leibler distances with non-negative weights. ■

In case there are more than two classes, we managed to check the sign of $L'_{\ln}(\cdot)$ everywhere except for a compact set, and the proof becomes much trickier. We used the ideas of Bramson [4] here, who proved $L_{\ln}(\cdot)$ to be a Lyapunov function for (A.1) without impatience.

LEMMA A.3: *If $I < \infty$ and $\nu_i = \nu$ for all i, then, for $\mathbf{z} \in (0,\infty)^I$, $\|\mathbf{z}\| \ge \|\mathbf{z}^*\|$, we have $L'_{\ln}(\mathbf{z}) \le 0$.*

Remark A.4: We have run numerical tests on the fluid model of a freelance website, that is (3) with all μ_i's being the same. According to those tests, $L'_{\ln}(\mathbf{z})$ should be non-positive for $\|\mathbf{z}\| \le \|\mathbf{z}^*\|$ as well. Moreover, as we omit the non-positive impatience term

$$-\nu \sum_{i=1}^{I} z_i \ln\left(\frac{z_i/\|\mathbf{z}\|}{z_i^*/\|\mathbf{z}^*\|} \right),$$

what is left should still be non-positive, that is,

$$\sum_{i=1}^{I} \left(\lambda_i + \sum_{j=1}^{I} P_{j,i}\mu_j \frac{z_j}{\|\mathbf{z}\|} - \mu_i \frac{z_i}{\|\mathbf{z}\|} \right) \ln\left(\frac{z_i/\|\mathbf{z}\|}{z_i^*/\|\mathbf{z}^*\|} \right) \le 0 \quad \text{for all } \|\mathbf{z}\| \le \|\mathbf{z}^*\|.$$

On the other hand, the proof of Lemma A.3 relies on the impatience term. So the two sets $\|\mathbf{z}\| \le \|\mathbf{z}^*\|$ and $\|\mathbf{z}\| \ge \|\mathbf{z}^*\|$ seem to need different approaches.

https://doi.org/10.1017/S0269964817000456 Published online by Cambridge University Press

Proof of Lemma A.3: Fix a $\mathbf{z} \in (0,\infty)^I$ such that $\|\mathbf{z}\| \geq \|\mathbf{z}^*\|$ and, to shorten notation, put

$$a_i := \frac{z_i/\|\mathbf{z}\|}{z_i^*/\|\mathbf{z}^*\|}, \quad q_i := z_i^*/\|\mathbf{z}^*\|, \quad i \in \mathcal{I}.$$

In the new notation,

$$L'_{\ln}(\mathbf{z}) = \Sigma - \nu \sum_{i=1}^{I} z_i \ln(a_i), \tag{A.4}$$

where

$$\Sigma := \sum_{i=1}^{I} \left(\lambda_i + \sum_{j=1}^{I} P_{j,i}\mu_j q_j a_j - \mu_i q_i a_i \right) \ln(a_i).$$

Also introduce

$$a_0 := 1, \quad q_0 := 1, \quad \mu_0 := \sum_{i=1}^{I} \lambda_i, \quad P_{0,0} := 0,$$

$$P_{0,i} := \lambda_i / \sum_{j=1}^{I} \lambda_j, \quad P_{i,0} := 1 - \sum_{j=1}^{I} P_{i,j}, \quad i \in \mathcal{I}.$$

Note that, by the fixed point Eq. (A.2),

$$\sum_{j=0}^{I} P_{j,i}\mu_j q_j - \mu_i q_i = \gamma_i, \quad i \in \mathcal{I}, \tag{A.5}$$

where

$$\gamma_0 := -\nu\|\mathbf{z}^*\| \quad \text{and,} \quad \text{for } i \in \mathcal{I}, \quad \gamma_i := \nu z_i^*.$$

Now Σ can be rewritten as

$$\Sigma = \sum_{i=0}^{I} \left(\sum_{j=0}^{I} P_{j,i}\mu_j q_j a_j - \mu_i q_i a_i \right) \ln(a_i). \tag{A.6}$$

Let $\sigma\colon \mathcal{I} \to \mathcal{I}$ be a permutation such that $a_{\sigma(i)}$ is non-decreasing in i. After reordering the classes according to σ, we apply to (A.6) the Abel partial summation rule, which reads as

$$\sum_{n=0}^{N} \alpha_n \beta_n = \alpha_N \sum_{m=0}^{N} \beta_m - \sum_{n=0}^{N-1} (\alpha_{n+1} - \alpha_n) \sum_{m=0}^{n} \beta_m.$$

Then we obtain

$$\Sigma = \ln(a_{\sigma(I)})(B_I - C_I) - \sum_{i=0}^{I-1} (\ln(a_{\sigma(i+1)}) - \ln(a_{\sigma(i)}))(B_i - C_i). \tag{A.7}$$

where

$$B_i := \sum_{l=0}^{i} b_l, \quad b_l := \sum_{j=0}^{I} P_{\sigma(j),\sigma(l)}\mu_{\sigma(j)} q_{\sigma(j)} a_{\sigma(j)},$$

$$C_i := \sum_{l=0}^{i} c_l, \quad c_l := \mu_{\sigma(l)} q_{\sigma(l)} a_{\sigma(l)}.$$

https://doi.org/10.1017/S0269964817000456 Published online by Cambridge University Press

Note that

$$B_I - C_I = \sum_{i,j=0}^{I} P_{j,i}\mu_j q_j a_j - \sum_{i=1}^{I} \mu_i q_i a_i = \sum_{j=0}^{I} \mu_j q_j a_j \left(\sum_{i=1}^{I} P_{j,i} - 1 \right) = 0. \tag{A.8}$$

We will prove that

$$\begin{aligned} &(\ln(a_{\sigma(i+1)}) - \ln(a_{\sigma(i)}))(B_i - C_i) \\ &\qquad \geq \Gamma_i (a_{\sigma(i+1)} \ln(a_{\sigma(i+1)}) - a_{\sigma(i)} \ln(a_{\sigma(i)})), \quad i = 0, \dots, I-1, \end{aligned} \tag{A.9}$$

where

$$\Gamma_i := \sum_{l=0}^{i} \gamma_{\sigma(l)},$$

but first we demonstrate how this implies the lemma.

Combining (A.7)–(A.9) and the Abel partial summation rule, we get

$$\begin{aligned} \Sigma &\leq - \sum_{i=1}^{I-1} (a_{\sigma(i+1)} \ln(a_{\sigma(i+1)}) - a_{\sigma(i)} \ln(a_{\sigma(i)})) \Gamma_i \\ &= \sum_{i=0}^{I} a_{\sigma(i)} \ln(a_{\sigma(i)}) \gamma_{\sigma(i)} - a_{\sigma(I)} \ln(a_{\sigma(I)}) \underbrace{\Gamma_I}_{=0} \\ &= \sum_{i=0}^{I} \gamma_i a_i \ln(a_i) = \sum_{i=1}^{I} \gamma_i a_i \ln(a_i) \\ &= \sum_{i=1}^{I} \nu z_i^{\mathrm{e}} \frac{z_i/\|\mathbf{z}\|}{z_i^*/\|\mathbf{z}^*\|} \ln(a_i) = \nu \|\mathbf{z}^*\| \sum_{i=1}^{I} \frac{z_i}{\|\mathbf{z}\|} \ln(a_i). \end{aligned}$$

Now we plug the last bound for Σ into (A.4) and get:

$$\begin{aligned} L'_{\ln}(\mathbf{z}) &\leq \nu \|\mathbf{z}^*\| \sum_{i=1}^{I} \frac{z_i}{\|\mathbf{z}\|} \ln(a_i) - \nu \sum_{i=1}^{I} z_i \ln(a_i) \\ &= \nu (\|\mathbf{z}^*\| - \|\mathbf{z}\|) \sum_{i=1}^{I} \frac{z_i}{\|\mathbf{z}\|} \ln\left(\frac{z_i/\|\mathbf{z}\|}{z_i^*/\|\mathbf{z}^*\|} \right) \leq 0, \end{aligned}$$

where in the second line, $\|\mathbf{z}^*\| - \|\mathbf{z}\| \leq 0$ by the Lemma's assumption, and the summation term is non-negative as a Kullback-Leibler distance.

So it is left to show (A.9) in order to finish the proof. For $i = 0, \dots, I-1$, the following holds with $f_i = a_{\sigma(i)}$ and $f_i = a_{\sigma(i+1)}$:

$$\begin{aligned} B_i &= \sum_{l=0}^{i} \sum_{j=0}^{I} P_{\sigma(j),\sigma(l)} \mu_{\sigma(j)} q_{\sigma(j)} a_{\sigma(j)} \\ &= \sum_{l=0}^{i} \sum_{j=0}^{i} P_{\sigma(j),\sigma(l)} \mu_{\sigma(j)} q_{\sigma(j)} a_{\sigma(j)} + \sum_{l=0}^{i} \sum_{j=i+1}^{I} P_{\sigma(j),\sigma(l)} \mu_{\sigma(j)} q_{\sigma(j)} a_{\sigma(j)} \\ &\geq \sum_{l=0}^{i} \sum_{j=0}^{i} P_{\sigma(j),\sigma(l)} \mu_{\sigma(j)} q_{\sigma(j)} a_{\sigma(j)} + \underbrace{f_i \sum_{l=0}^{i} \sum_{j=i+1}^{I} P_{\sigma(j),\sigma(l)} \mu_{\sigma(j)} q_{\sigma(j)}}_{=: \widetilde{B}_i}, \end{aligned} \tag{A.10}$$

https://doi.org/10.1017/S0269964817000456 Published online by Cambridge University Press

where by (A.5),

$$\begin{aligned}\widetilde{B}_i &= f_i \sum_{l=0}^{i} \left(\sum_{j=0}^{I} P_{\sigma(j),\sigma(l)} \mu_{\sigma(j)} q_{\sigma(j)} - \sum_{j=0}^{i} P_{\sigma(j),\sigma(l)} \mu_{\sigma(j)} q_{\sigma(j)} \right) \\ &= f_i \sum_{l=0}^{i} \left(\gamma_{\sigma(l)} + \mu_{\sigma(l)} q_{\sigma(l)} - \sum_{j=0}^{i} P_{\sigma(j),\sigma(l)} \mu_{\sigma(j)} q_{\sigma(j)} \right) \\ &= f_i \Gamma_i + f_i \sum_{j=0}^{i} \mu_{\sigma(j)} q_{\sigma(j)} - f_i \sum_{j=0}^{i} \sum_{l=0}^{i} P_{\sigma(j),\sigma(l)} \mu_{\sigma(j)} q_{\sigma(j)} \\ &= f_i \Gamma_i + f_i \sum_{j=0}^{i} \mu_{\sigma(j)} q_{\sigma(j)} \sum_{l=i+1}^{I} P_{\sigma(j),\sigma(l)} \\ &\geq f_i \Gamma_i + \sum_{j=0}^{i} \mu_{\sigma(j)} q_{\sigma(j)} a_{\sigma(j)} \sum_{l=i+1}^{I} P_{\sigma(j),\sigma(l)}.\end{aligned}$$

As we plug the last inequality back into (A.10), it follows that

$$B_i \geq C_i + f_i \Gamma_i,$$

and since $(\ln(a_{\sigma(i+1)}) - \ln(a_{\sigma(i)})) \geq 0$, we have

$$(\ln(a_{\sigma(i+1)}) - \ln(a_{\sigma(i)}))(B_i - C_i) \geq (\ln(a_{\sigma(i+1)}) - \ln(a_{\sigma(i)})) f_i \Gamma_i. \tag{A.11}$$

Let i_0 be the index for which $\sigma(i_0) = 0$. For $i < i_0$, we have $\ln(a_{\sigma(i)}) \leq 0$ and $\Gamma_i \geq 0$, yielding

$$\begin{gathered} a_{\sigma(i+1)} \ln(a_{\sigma(i)}) \leq a_{\sigma(i+1)} \ln(a_{\sigma(i)}), \\ \Gamma_i (a_{\sigma(i+1)} \ln(a_{\sigma(i+1)}) - a_{\sigma(i+1)} \ln(a_{\sigma(i)})) \geq \Gamma_i (a_{\sigma(i+1)} \ln(a_{\sigma(i+1)}) - a_{\sigma(i)} \ln(a_{\sigma(i)})), \end{gathered}$$

which, when compared to (A.11) with $f_i = a_{\sigma(i+1)}$, implies (A.9).

Similarly, we prove (A.9) for $i \geq i_0$. For such an i, $\ln(a_{\sigma(i+1)}) \geq 0$ and $\Gamma_i \leq 0$. Hence,

$$a_{\sigma(i)} \ln(a_{\sigma(i+1)}) \leq a_{\sigma(i+1)} \ln(a_{\sigma(i+1)})$$

and

$$\Gamma_i (a_{\sigma(i)} \ln(a_{\sigma(i+1)}) - a_{\sigma(i)} \ln(a_{\sigma(i)})) \geq \Gamma_i (a_{\sigma(i+1)} \ln(a_{\sigma(i+1)}) - a_{\sigma(i)} \ln(a_{\sigma(i)})).$$

As we compare the last inequality to (A.11) with $f_i = a_{\sigma(i)}$, (A.9) follows. ■

Quadratic Lyapunov function In the next lemma, we verify that the quadratic form (11) is a Lyapunov function for (A.1) in case there are two classes of customers with different impatience parameters.

LEMMA A.5: *If $I = 2$, the derivative of* (11) *with respect to* (A.1) *is non-positive on* $(0, \infty)^2$.

https://doi.org/10.1017/S0269964817000456 Published online by Cambridge University Press

PROOF: Fix a $\mathbf{z} \in (0,\infty)^2$. Introduce the notations

$$\mathbf{y} = (y_1, y_2) := \mathbf{z} - \mathbf{z}^*, \quad p_i := z_i/\|\mathbf{z}\|, \quad q_i := z_i^*/\|\mathbf{z}^*\|, \quad i = 1,2,$$
$$\alpha_1 := \frac{2}{[(1-P_{1,1})\mu_1 + P_{2,1}\mu_2]q_1}, \quad \alpha_2 := \frac{2}{[(1-P_{2,2})\mu_2 + P_{1,2}\mu_1]q_2},$$

and note that

$$L_{\mathrm{qd}}(\mathbf{z}) = \sum_{i=1}^{2} (\alpha_i/2) y_i^2, \quad p_i - q_i = \frac{1}{\|\mathbf{z}\|}\left(y_i - q_i \sum_{j=1}^{2} y_j\right), \quad i = 1,2. \tag{A.12}$$

Then

$$\begin{aligned} L'_{\mathrm{qd}}(\mathbf{z}) &= \sum_{i=1}^{2} \alpha_i y_i \left(\lambda_i + \sum_{j=1}^{2} P_{j,i}\mu_j p_j - \mu_i p_i - \nu_i z_i\right) \\ &= \sum_{i=1}^{2} \alpha_i y_i \left(\sum_{j=1}^{2} P_{j,i}\mu_j (p_j - q_j) - \mu_i (p_i - q_i) - \nu_i y_i\right) \\ &\overset{(A.12)}{=} -\alpha_1\nu_1 y_1^2 - \alpha_2\nu_2 y_2^2 - \frac{1}{\|\mathbf{z}\|}\Sigma(\mathbf{y}), \end{aligned}$$

where

$$\begin{aligned} \Sigma(\mathbf{y}) = \alpha_1 y_1 &\left[(1-P_{1,1})\mu_1\left(y_1 - q_1\sum_{j=1}^{2} y_j\right) - P_{2,1}\mu_2\left(y_2 - q_2\sum_{j=1}^{2} y_j\right)\right] \\ + \alpha_2 y_2 &\left[(1-P_{2,2})\mu_2\left(y_2 - q_2\sum_{j=1}^{2} y_j\right) - P_{1,2}\mu_1\left(y_1 - q_1\sum_{j=1}^{2} y_j\right)\right]. \end{aligned}$$

Now, in order to prove the lemma, it suffices to show that the quadratic form $\Sigma(\cdot)$ is non-negative in $\mathbb{R}^2$. Denote the coefficients of $\Sigma(\mathbf{y})$ in front of y_1^2, y_2^2 and $y_1 y_2$ by $a_{1,1}$, $a_{2,2}$, and $2a_{1,2}$, respectively. Then

$$\begin{aligned} a_{1,1} &= \alpha_1[(1-P_{1,1})\mu_1(1-q_1) + P_{2,1}\mu_2 q_2] = \alpha_1[(1-P_{1,1})\mu_1 + P_{2,1}\mu_2]q_2 = 2q_2/q_1, \\ a_{2,2} &= \alpha_2[(1-P_{2,2})\mu_2(1-q_2) + P_{1,2}\mu_1 q_1] = \alpha_2[(1-P_{2,2})\mu_2 + P_{1,2}\mu_1]q_1 = 2q_1/q_2, \end{aligned}$$

and

$$\begin{aligned} 2a_{1,2} &= -\alpha_1[(1-P_{1,1})\mu_1 q_1 + P_{2,1}\mu_2(1-q_2)] - \alpha_2[(1-P_{2,2})\mu_2 q_2 + P_{1,2}\mu_1(1-q_1)] \\ &= -\alpha_1[(1-P_{1,1})\mu_1 + P_{2,1}\mu_2]q_1 - \alpha_2[(1-P_{2,2})\mu_2 + P_{1,2}\mu_1]q_2 = -4. \end{aligned}$$

Hence,

$$\Sigma(\mathbf{y}) = 2\left(\sqrt{q_2/q_1}\,y_1 - \sqrt{q_1/q_2}\,y_2\right)^2 \geq 0,$$

which completes the proof. ■

Remark A.6: As mentioned before, the quadratic function

$$\widetilde{L}_{\mathrm{qd}}(\mathbf{z}) = \sum_{i=1}^{I} \frac{(z_i - z_i^*)^2}{\mu_i z_i^*/\|\mathbf{z}^*\|}$$

is a Lyapunov function for the system (A.1) in case there is no routing (the number I of classes can be arbitrary). We have done numerical tests which indicate that this function should also work

https://doi.org/10.1017/S0269964817000456 Published online by Cambridge University Press

as a Lyapunov function for the freelance fluid model (that is (3) with all μ_i's the same). Unlike for the entropy function (10) (*cf.* Remark A.4), the impatience term

$$-\sum_{i=1}^{I} \frac{\nu_i}{\mu_i z_i^* / \|\mathbf{z}^*\|} (z_i - z_i^*)^2$$

is crucial in this case: Without it the derivative can take positive values.

https://doi.org/10.1017/S0269964817000456 Published online by Cambridge University Press