



**HAL**  
open science

# The Fluid Limit of the Multiclass Processor Sharing Queue

Abdelghani Ben Tahar, Alain Jean-Marie

► **To cite this version:**

Abdelghani Ben Tahar, Alain Jean-Marie. The Fluid Limit of the Multiclass Processor Sharing Queue. [Research Report] RR-6867, 2009, 63 p. inria-00368246v1

**HAL Id: inria-00368246**

**<https://inria.hal.science/inria-00368246v1>**

Submitted on 17 Mar 2009 (v1), last revised 21 Apr 2009 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

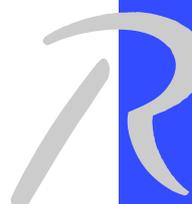
*The Fluid Limit of the Multiclass Processor Sharing  
Queue*

Abdelghani Ben Tahar — Alain Jean-Marie

N° 6867

March 2009

Thème COM



*R*apport  
*de recherche*



## The Fluid Limit of the Multiclass Processor Sharing Queue

Abdelghani Ben Tahar\*, Alain Jean-Marie†

Thème COM — Systèmes communicants  
Équipe-Projet Maestro

Rapport de recherche n° 6867 — March 2009 — 63 pages

**Abstract:** Consider a single server queueing system with several classes of customers, each having its own renewal input process and its own general service times distribution. Upon completing service, customers may leave, or reenter the queue, possibly as customers of a different class. The server is operating under the egalitarian processor sharing discipline. Building on prior work by Gromoll et al. [15] and Puha et al. [19], we provide existence of the fluid model for multiclass processor sharing queue and their asymptotic behavior. In addition, we establish the convergence of a properly normalized state process to a fluid limit.

**Key-words:** Processor Sharing, Fluid Limit

\* Laboratoire de Mathématiques Raphaël Salem UMR 6085 CNRS-Université de Rouen, Avenue de l'Université, BP.12, F76801 Saint-Étienne-du-Rouvray, [abdelghani.bentahar@univ-rouen.fr](mailto:abdelghani.bentahar@univ-rouen.fr). Part of this work was performed while the author was a CNRS Postdoctoral fellow at LIRMM.

† INRIA and LIRMM, CNRS/Université Montpellier 2, 161 Rue Ada, F-34392 Montpellier, [ajm@lirmm.fr](mailto:ajm@lirmm.fr).

## La limite fluide de la file d'attente Processor Sharing multiclasse

**Résumé :** Dans une file d'attente "Processor Sharing", chaque client présent reçoit une part égale de la capacité de service du serveur. Nous étudions le comportement de cette file quand il y a plusieurs classes de clients, susceptibles de ré-entrer dans la file quand leur service se termine. Les différentes classes sont caractérisées par des distributions de temps de service et des probabilités de routage différentes. A partir des résultats de Gromoll *et al.* [15] et Puha *et al.* [19], nous montrons l'existence de cette limite fluide, et calculons les solutions de cette limite ainsi que ses propriétés asymptotiques.

**Mots-clés :** Limite fluide, Processor Sharing

# 1 Introduction

## 1.1 Setting of the problem

We consider a queueing system composed of one station and  $K$  classes of customers. The station is assumed to have a single server and an infinite storage capacity. All customers present in the system are served simultaneously according to the egalitarian processor sharing rule: at any time, each of them is served at a rate that is the inverse of the total number of customers in the system. Customers arrive from the outside with a given class. Their service time distribution depends on the class. Upon service completion, customers may leave the system, or reenter it as customers of a different class, according to a probabilistic, Jackson-like routing mechanism. We call this system a Multiclass Processor Sharing (MPS) queue. This paper is devoted to the study of *fluid limits* for this stochastic system. When properly normalized, the dynamics of this system can be described by a deterministic system of integro-differential equations. The purpose of this paper is to prove limit theorems (law of large number) that justifies this fluid approximation, and to study the basic properties of the fluid system, such as existence, uniqueness and construction of the solution, and its asymptotic behavior.

## 1.2 Literature

When the number of classes is one in the MPS queue, the system reduces to the standard  $GI/GI/1/PS$ . The literature features several fluid or deterministic results for this system. Jean-Marie and Robert [18] studied the asymptotic behavior of the overloaded PS-queue. They established that the queue length grows asymptotically linearly with time, and gave the value of the growth rate. Chen, Kella and Weiss [8] studied the fluid approximation for the PS queue and under the condition “traffic intensity less than one”. Other authors such as Gromoll et al. [15] and Puha et al. [19] studied the fluid approximation of the  $GI/GI/1/PS$  by considering the so-called state descriptor to model the system. This is a measure-valued process  $\mu(\cdot)$  with values in the space of finite, non-negative Borel measures on  $\mathbb{R}_+$  endowed with the topology of weak convergence. At any time  $t$  the measure  $\mu(t)$  puts a unit of mass at the residual service times of each customer in the system. The first authors have studied the heavily loaded queue while the second have extended the results to the overloaded queue.

Bramson [4, 5, 6] and Williams [22] have established a framework for studying the asymptotic behavior of solutions of critical fluid models and obtaining a diffusion approximation, for an open multiclass networks operating under the Head-Of-The-Line processor sharing service discipline. Puha and Williams [20] and Gromoll [14] established the analogue results for the PS queue. Recent developments include the analysis of PS queues with impatience [16, 17].

The present paper generalizes the analysis of [15, 19] to the multiclass case.

## 1.3 Summary of the results

The key quantity attached to a customer in a MPS queue is its *residual service time*. Since customers of different classes have a service characteristics, we model the dynamics of the MPS queue by means of three  $K$ -dimensional processes:  $A(\cdot)$ ,  $D(\cdot)$  and  $\mu(\cdot)$  with values respectively in  $\mathbb{N}_+^K$ ,  $\mathbb{N}_+^K$ ,  $\mathcal{M}^K$ , where  $\mathcal{M}$  is the space of finite, non-negative Borel measures on  $\mathbb{R}_+$  endowed with the topology of weak convergence.

For each class  $k$  and any time  $t$ ,  $A_k(t)$ ,  $D_k(t)$  represent respectively the number of customers arrived at, departed from class  $k$  by time  $t$  and,  $\mu_k(t)$  is a measure valued process that keeps track of all residual service times of class  $k$ . The evolution of these processes is governed by equations (2.1)–(2.5), which have the same structure, albeit multidimensional, than the equations used for the  $GI/GI/1/PS$  in [15] and [19]. In particular, the dynamics for  $\mu_k(\cdot)$ , in Equation (2.5), has the same form as that for  $\mu(\cdot)$  in the single-class queue, except that the exogenous arrival process  $E(t)$  is replaced by the endogenous process  $A_k(t)$  (Equation (2.5)). Starting with a sequence of MPS queues indexed by  $r$ , and under mild

assumptions on initial conditions, we define the normalized processes  $\bar{A}^r(t) = A(rt)/r$  and likewise for  $D$  and  $\mu$ , and we prove that the limit of the process  $(\bar{A}^r(\cdot), \bar{D}^r(\cdot), \bar{\mu}^r(\cdot))$  is a fluid solution.

The fluid model here has three parameters:  $\alpha = (\alpha_1, \dots, \alpha_K)$  is the vector of exogenous arrival rates,  $\nu = (\nu_1, \dots, \nu_K)$  is a vector of probability laws in which each component  $\nu_k$  corresponds to the distribution of i.i.d. service times within class  $k$ , and  $P = (p_{kl})$  corresponds to the routing matrix associated with an open network (but routing consists in re-entry in the queue with possible change of class). These parameters constitute the *data* of our system. A fluid model solution associated to data  $(\alpha, \nu, P)$  is a family of two real-valued, and one measure-valued vectors of continuous functions  $\bar{A}(\cdot)$ ,  $\bar{D}(\cdot)$ , and  $\bar{\mu}(\cdot)$  respectively, that satisfy the flow conservation equations:  $\bar{A}(t) = \alpha t + P' \bar{D}(t)$ ,  $\langle 1, \bar{\mu}_k(t) \rangle = \langle 1, \bar{\mu}_k(0) \rangle + \bar{A}_k(t) - \bar{D}_k(t)$  and,

$$\bar{\mu}_k(t)([x, \infty)) = \bar{\mu}_k(0)([x + S(t), \infty)) + \int_0^t \nu_k([x + S(t) - S(s), \infty)) d\bar{A}_k(s). \quad (1.1)$$

Here, for each  $t \geq s \geq 0$ ,  $S(t) - S(s)$  is the accumulated service quantity devoted to any present customer over  $[s, t]$ , in the fluid model. The analysis strategy requires three steps; each one is based on the analysis used in [15] and [19].

First, since the transition of one customer between the classes does not modify the total number of customers, and consequently the processing speed of the other customers, we argue that the total number of customers in the system evolves according to an equation equivalent to that obtained for  $GI/GI/1/PS$  in which the service of each customer is the sum of all its individual service requirements in the multiclass model. We identify the distribution of the total service times of each customer in the system, and exploit existing results on  $GI/GI/1/PS$  to prove the fluid limit for the total number of customers in the system.

In the second step, we concentrate on the visits a given customer makes to some class  $k$  in its route (the route is the sequence of the successive classes to which a customer belongs during its stay in the system). Conditionally on its route, it is possible to define for each customer and each time  $t$ , the ‘‘remaining service’’ that it should receive until it is able to complete its  $n^{th}$  visit to class  $k$ . This value is given by the sum of all individual service requirements from its arrival to the system up to its  $n^{th}$  visit of class  $k$ , minus the total amount of service he has received up to time  $t$ . We will call this quantity the residual service times of visits. As above, we introduce the measure  $\gamma_{lk}(t)$  which accounts for all the total residual service times at the moments of all visits to class  $k$  (past, present or to come), for all customers present in the system at time  $t$  currently of class  $k$ , who entered the system as customers of class  $l$ . In particular, the total mass of the measure  $\gamma_{lk}(t)$  is the number of visits to class  $k$  made by customers who have arrived to the system as customers of class  $l$ , and are still present in the system at time  $t$ . The family of state descriptors  $\{\gamma_{lk}(\cdot), l, k = 1, \dots, K\}$  encompasses most of the useful information on the system, such as the number of visits of, the departures from and, the arrival process at, each class. It turns out that this state descriptor admits a fluid limit, given for each  $l, k = 1, \dots, K$ , by:

$$\bar{\gamma}_{lk}(t)([x, \infty) = (\mathcal{B}_{kl} * \bar{\mu}_l(0))( [x + S(t), \infty) ) + \int_0^t (\mathcal{B}_{kl} * \nu_l)( [x + S(t) - S(s), \infty) ) \alpha_l ds. \quad (1.2)$$

Here  $\mathcal{B}(\cdot)$  is the matrix-valued function defined as:  $\mathcal{B}(t) = \sum_{n \geq 0} (BP')^{*n}(t)$ , where  $B$  is a diagonal  $K \times K$ -matrix which diagonal entries are the distribution functions associated with  $\nu_k$ . The structure is therefore similar to (1.1), except that the unknown, endogenous processes  $A_k(t)$  have been replaced by the known function  $\alpha_l t$ .

In final step, having proved the fluid limit for the arrival process at each class, the proof of the limit for the state descriptor  $\mu_k(\cdot)$  for each class  $k$ , is a slight modification of the framework used in [15], and [19].

The paper is organized as follows. This introductory part continues with, in Section 2, the different definitions and preliminaries which are necessary for stating the results. This includes the definition of the model and the description of queueing equations (Sections 2.1 and 2.2), then that of the fluid model and equations (Section 2.3) and results on the correspondence with the single-server queue (Section 3.1).

The second part is devoted to the presentation of results, in Section 4. The proofs of these results are presented in Section 5 for results concerning the existence of solutions to the fluid model, and their asymptotic behavior, and in Section 6 for the principal convergence result: Theorem 4.5. Section 7 presents several illustrations and applications of the results. The appendix presents results from the literature and proofs of intermediate results which are not essential to the main presentation.

## 1.4 Notation

The notation used throughout the paper are collected here. For real numbers  $a, b$ ,  $a \vee b = \max(a, b)$  and  $a \wedge b = \min(a, b)$ , in particular,  $a^+ = a \vee 0$  and  $a^- = a \wedge 0$ , and the integer part of  $a$  is denoted as  $[a]$ . Let  $\mathbb{N}$  denote the set of natural numbers,  $\mathbb{R}_+$  denote the non-negative real numbers,  $\mathbb{N}^K$  denote the Cartesian product of  $\mathbb{N}$  and  $\mathbb{R}_+^K$  denote the  $K$ -dimensional Euclidean space. Vectors will be column, when  $e$  denotes a vector, it stands for a row vector of ones. The transpose of a vector or matrix is denoted by prime. For a vector  $x \in \mathbb{R}^K$ , the  $K \times K$  diagonal matrix whose entries are given by the components of  $x$  will be denoted by  $diag\{x\}$ . Let  $\mathcal{K}$  be the set  $\{1, \dots, K\}$  of all classes,  $I, J \subseteq \mathcal{K}$  and the  $K \times K$ -matrix  $R$ , the sub-matrix  $R_{IJ}$  of  $R$  is a matrix whose row indices are in  $I$  and column indices are in  $J$ . For a function  $g : \mathbb{R}_+ \mapsto \mathbb{R}$ , let  $\|g\|_T = \sup_{t \in [0, T]} |g(t)|$  for each  $T \geq 0$  and  $\|g\|_\infty = \sup_{t \in \mathbb{R}_+} |g(t)|$ . The indicator function of a set  $A$  is the function  $1_{\{A\}}$  which equals 1 for  $x \in A$  and 0 otherwise. The following real valued functions will be used repeatedly:  $\chi(x) = x$  for  $x \in \mathbb{R}_+$ , and  $\varphi$ :

$$\varphi(x) = 1/x \text{ for } x \in (0, \infty), \text{ and } \varphi(0) = 0. \quad (1.3)$$

For two matrices of measurable functions  $F(\cdot)$  and  $G(\cdot)$  defined on  $\mathbb{R}_+$ , we denote by the matrix-valued functions  $(F * G)(x)$  for  $x \in \mathbb{R}_+$ , the matrix convolution formed of the elements:  $(F * G)_{ij}(x) = \sum_k (F_{ik} * G_{kj})(x)$ . This operation is associative and distributive over matrix addition. The multiplication by a constant matrix  $C$  can be seen as a convolution, where each element  $C_{ij}$  is interpreted as the function  $C_{ij} 1_{x \geq 0}$ . Associativity therefore holds for mixed scalar products and convolutions. The  $n$ -th convolution power of a matrix  $F(x)$  is denoted with  $F^{*n}(x)$ .

For an interval  $I \subseteq \mathbb{R}$ , let  $\mathbf{C}_b(I)$  and  $\mathbf{C}_b^1(I)$  denote respectively the sets of continuous, bounded, real valued functions defined on  $I$ , and continuously differentiable functions with bounded first derivatives on  $I$ . For  $g \in \mathbf{C}_b^1(I)$  we write  $\dot{g}(x) = \frac{d}{dx}g(x)$  for  $x \in I$ .

The set of finite, nonnegative and Borel measures on  $\mathbb{R}_+$  is denoted by  $\mathcal{M}$  and  $\mathcal{M}^K$  is Cartesian product of  $\mathcal{M}$ . The measures 0 and  $\delta_x^+$  denote respectively the zero measure and the Borel measure on  $\mathbb{R}_+$  with mass one at  $x > 0$ . We write  $\langle g, \mu \rangle = \int g d\mu$  for  $\mu \in \mathcal{M}$  and a Borel measurable function  $g$  which is integrable with respect to  $\mu$ . The space  $\mathcal{M}$  is embedded with the weak topology, for which it is a Polish space.

For a Polish space  $E$  and for  $T \geq 0$ ,  $D([0, T], E)$  respectively  $D([0, \infty), E)$  denotes the space of r.c.l.l. (right continuous with left limits) functions from  $[0, T]$  respectively from  $[0, \infty)$  to  $E$ . We will use  $\mathbb{P}$  and  $\mathbb{E}$  to denote the probability measure and expectation operator with whatever space the relevant random element is defined on, and  $\Rightarrow$  to denote convergence in distribution of a sequence of random elements of a metric space.

## 2 The Queuing Model and the Fluid Model

### 2.1 Primitive data and initial conditions

We construct in this section the evolution equations for the system, which will be the basis for the analysis.

For each  $k \in \mathcal{K}$ , we assume that there are two i.i.d. sequences of random variables,  $u_k = \{u_k(i), i \geq 1\}$  and  $v_k = \{v_k(i), i \geq 1\}$  and an i.i.d. sequence of  $K$ -dimensional random vectors,  $\varphi^k = \{\varphi^k(i), i \geq 1\}$ . Each element of  $u_k$ ,  $v_k$  and  $\varphi^k$  takes values respectively in  $\mathbb{R}_+$ ,  $\mathbb{R}_+^*$  and  $\{e_0, e_1, \dots, e_K\}$ , where  $e_0$  is the

$K$ -dimensional vector of all components 0, and  $e_k$  is the  $K$ -dimensional vector with  $k^{th}$  component 1 and other components 0. Assume that the sequences

$$u_1, \dots, u_K, v_1, \dots, v_K, \varphi^1, \dots, \varphi^K$$

are mutually independent. These sequences have the following interpretation: for each  $i \geq 1$  and for each  $k \in \mathcal{K}$ ,  $u_k(i)$  is the interarrival time between the  $(i-1)^{th}$  and the  $i^{th}$  externally arriving customer at class  $k$ ,  $v_k(i)$  is the service times for the  $i^{th}$  class  $k$  customer and for each  $l = 0, 1, \dots, K$ ,  $\varphi^k(i) = e_l$  means that the  $i$ -th customer of class  $k$  which completes service, becomes a class  $l$  customer. These data constitute the *primitive data* of the network.

From them are derived the following parameters. The real valued vector  $\alpha = (\alpha_1, \dots, \alpha_K)$  is defined as  $\alpha_k = [\mathbb{E}(u_k(1))]^{-1}$  for each  $k \in \mathcal{K}$ . The vector  $\nu = (\nu_1, \dots, \nu_K)$  is formed of  $\nu_k$ , the Borel probability measure of  $v_k$ . Finally, the nonnegative matrix  $P$  is formed from the components  $p_{kl}, k, l \in \mathcal{K}$ , where  $p_{kl} = \mathbb{P}(\varphi^k(1) = e_l)$ . We allow that  $\alpha_k = 0$  for some class  $k$ , and we set  $\mathcal{A} = \{k : \alpha_k \neq 0\}$ . The network is assumed to be open, that is,  $\mathcal{A}$  is nonempty and the matrix

$$Q = I + P' + (P')^2 + \dots$$

is finite, which is equivalent to requiring that  $(I - P')$  be invertible, or that  $P$  has a spectral radius less than 1. In that case,  $Q = (I - P')^{-1}$ . It is assumed that for each  $k \in \mathcal{K}$ , the distribution  $\nu_k$  does not charge the origin,  $\nu_k(\{0\}) = 0$ , and satisfies:  $\langle \chi, \nu_k \rangle < \infty$  (finite expectation).

Associated with class  $k$ , let  $E_k(t) = \sup\{n : \sum_{i=1}^n u_k(i) \leq t\}$  be the number of exogenous arrivals of class  $k$  by time  $t$  and  $\Phi_{lk}(n) = \sum_{i=1}^n \varphi_k^l(i)$  be the number of customers that move from class  $l$  to class  $k$ , among the  $n$  first customers of class  $l$ . Denote by  $E(t) = (E_1(t), \dots, E_K(t))$  and  $\Phi(n) = (\varphi_l^k(n), l, k \in \mathcal{K})$ . These processes are called *primitive processes* in [6]. For each  $k \in \mathcal{K}$ , we assume that there exists an integer random variable with finite mean  $Z_k(0)$  and an i.i.d. sequence of strictly positive random variables  $v_k^0 = \{v_k^0(i), i \geq 1\}$  with a common Borel probability measure  $\nu_k^0$ , such that

$$v_1^0, \dots, v_K^0, v_1, \dots, v_K, \varphi^1, \dots, \varphi^K, Z_1(0), \dots, Z_K(0)$$

are mutually independent. Any customer belonging class  $k$  at time zero in the system is referred to as an “initial customer of class  $k$ ”. Then let  $Z_k(0)$  be the number of initial customer of class  $k$  and  $v_k^0(i)$  be the service times requirement of the  $i^{th}$  initial customer of class  $k$ . Assume that an initial customer of class  $k$  becomes a customer of class  $l$  with service in the sequence  $v_l$  or leaves the station according to the random routing  $\varphi_l^k$ .

## 2.2 Queuing equations

Given the primitive data and the primitives processes  $(E(\cdot), \Phi(\cdot))$  defined in section 2.1, we can consider

$$A(t) = (A_1(t), \dots, A_K(t)), D(t) = (D_1(t), \dots, D_K(t)), Z(t) = (Z_1(t), \dots, Z_K(t))$$

be a random processes such that  $A_k(t), D_k(t)$  and  $Z_k(t)$  are respectively, the total number of arrivals by time  $t$  at, the number of departures by time  $t$  from, and the number of customers present at time  $t$  in, class  $k$ . Jointly, those processes satisfy the following queueing equations:

$$A_k(t) = E_k(t) + \sum_{l=1}^K \Phi_{lk}(D_l(t)) \quad (2.1)$$

$$D_k(t) = \sum_{i=1}^{Z_k(0)} 1_{\{v_k^0(i) \leq S(t)\}} + \sum_{i=1}^{A_k(t)} 1_{\{v_k(i) \leq S(t) - S(\sigma_k(i))\}} \quad (2.2)$$

$$Z(t) = Z(0) + A(t) - D(t) \quad (2.3)$$

$$S(t) = \int_0^t \varphi(eZ(s)) ds. \quad (2.4)$$

Here, for all  $k \in \mathcal{K}$ ,  $A_k(0) = 0$ ,  $\sigma_k(i)$  is the arrival epoch of the  $i^{\text{th}}$  customer to arrive at class  $k$ , and  $\varphi(x)$  is defined in (1.3).

The interpretation of this definition with respect to the model is the following. The functions  $A_k(t)$ ,  $D_k(t)$  represent, respectively, the amount of customers of class  $k$  arrived at, and departed from the queue at time  $t$ . Equation (2.1) is then simply the flow conservation equations, taking external arrivals and internal routing into account.

The function  $S(t)$  is known as the cumulative service. It represents the amount of service received by one particular customer in the interval  $[0, t]$ . Since the Processor Sharing gives the same amount of service to all present customers, this quantity is the same for all customers present in the interval. Hence,  $S_{s,t} := S(t) - S(s)$  is the amount of service received in the interval  $[s, t]$ .

For each  $k \in \mathcal{K}$ , define the measure-valued function of time  $\mu_k : [0, \infty) \rightarrow \mathcal{M}$  by

$$\mu_k(t) = \sum_{j=1}^{Z_k(0)} \delta_{(v_k^0(j) - S(t))^+} + \sum_{i=1}^{A_k(t)} \delta_{(v_k(i) - (S(t) - S(\sigma_k(i))))^+}. \quad (2.5)$$

At each time  $t$ ,  $(v_k^0(j) - S(t))^+$  and  $(v_k(i) - (S(t) - S(\sigma_k(i))))^+$  are the residual service times within class  $k$  of, respectively  $j^{\text{th}}$  initial customer, and  $i^{\text{th}}$  customer. Recall that  $\delta_x^+$  is the Borel measure on  $\mathbb{R}_+$  with mass one at  $x > 0$ , and that the random measure  $\mu_k(t)$  takes values in the space  $\mathcal{M}$  of finite, positive Borel measures on  $\mathbb{R}_+$ .  $\mu_k(\cdot)$  is measure-valued stochastic process with simple path in the polish space  $D([0, +\infty), \mathcal{M})$ . In [15], this process is referred to as the state descriptor. The equation (2.5) is equivalent to

$$\langle g, \mu_k(t) \rangle = \sum_{j=1}^{Z_k(0)} (1_{(0, \infty)} g)((v_k^0(j) - S(t))^+) + \sum_{i=1}^{A_k(t)} (1_{(0, \infty)} g)(v_k(i) - (S(t) - S(\sigma_k(i))))^+ \quad (2.6)$$

for  $t \geq 0$  and all bounded, Borel measurable function  $g : \mathbb{R}_+ \rightarrow \mathbb{R}$ . In particular, for  $g = 1$ ,  $t \geq 0$  and  $k \in \mathcal{K}$ , the number of customers of class  $k$  at time  $t$  is given by

$$Z_k(t) = \langle 1, \mu_k(t) \rangle. \quad (2.7)$$

## 2.3 Fluid Model

The fluid model shares the following parameters with the discrete model: the nonnegative vector  $\alpha = (\alpha_1, \dots, \alpha_K)$ , the vector (or diagonal matrix, depending on the context) of Borel probability measures  $\nu = (\nu_1, \dots, \nu_K)$  and the nonnegative routing matrix  $P$ .

The required assumptions on this parameters are the same as above: for  $k \in \mathcal{K}$ , the measure  $\nu_k$  does not charge the origin,  $\beta_k := \langle \chi, \nu_k \rangle < \infty$ , and the routing matrix  $P$  is substochastic and has spectral radius strictly less than one. Hence the matrix  $Q = (I - P)^{-1}$  is well defined. Define the vector  $\lambda = Q\alpha$ . The global arrival rate to the class  $k$  is then  $\lambda_k$ , and the load factor of the queue is  $\rho = \sum_{k=1}^K \lambda_k \langle \chi, \nu_k \rangle$ . The adjectives *subcritical*, *critical* and *supercritical* will be used to refer to data  $(\alpha, \nu, P)$  that satisfy  $\rho < 1$ ,  $\rho = 1$ ,  $\rho > 1$  respectively. The analysis is made easier using a matrix notation. Denote by  $B(x) = \text{diag}\{B_k(x); k \in \mathcal{K}\}$ ,  $\beta = \text{diag}\{\beta_k; k \in \mathcal{K}\}$  and  $\hat{B}(x) = \text{diag}\{\hat{B}_k(x); k \in \mathcal{K}\}$ , where  $B_k(\cdot)$  the distribution function of  $\nu_k$ ,  $\beta_k = \langle \chi, \nu_k \rangle$  is the average service time for class  $k$ , and  $\hat{B}_k(\cdot)$  the Laplace transform of  $B_k(\cdot)$ .

Let  $\mathcal{M}^{c,K} = \{\xi \in \mathcal{M}^K : \xi_k(\{x\}) = 0 \text{ for all } x \in \mathbb{R}_+ \text{ and } k \in \mathcal{K}\}$  be a set of finite, non-negative Borel measures on  $\mathbb{R}_+$  that have no atoms, and let  $\mathcal{M}^{c,p,K} = \{\xi \in \mathcal{M}^{c,K} : \xi \neq 0\}$  be positive measure of  $\mathcal{M}^{c,K}$

**Definition 2.1** (Fluid Solution Model). *Let  $(\alpha, \nu, P)$  be some data and  $\xi \in \mathcal{M}^{c,K}$  be an initial state. A fluid solution is a triple  $(\bar{A}(t), \bar{D}(t), \bar{\mu}(t))$  of two real-, and one measure-valued vectors of continuous functions:  $\bar{A}, \bar{D} : \mathbb{R}_+ \rightarrow \mathbb{R}_+^K$ , and  $\bar{\mu} = (\bar{\mu}_1, \dots, \bar{\mu}_K) : \mathbb{R}_+ \rightarrow \mathcal{M}^K$  such that  $\bar{\mu}(0) = \xi$ , and*

i)  $\bar{A}$  and  $\bar{D}$  are increasing componentwise,

ii) The triple satisfies the relations

$$\bar{A}(t) = \alpha t + P' \bar{D}(t) \quad (2.8)$$

$$\langle 1, \bar{\mu}_k(t) \rangle = \langle 1, \xi_k \rangle + \bar{A}_k(t) - \bar{D}_k(t) \quad (2.9)$$

$$\langle 1_{[x, \infty[}, \bar{\mu}_k(t) \rangle = \langle 1_{[x, \infty[}(\cdot - \bar{S}(t)), \xi_k \rangle + \int_0^t \langle 1_{[x, \infty[}(\cdot - (\bar{S}(t) - \bar{S}(s))), \nu_k \rangle d\bar{A}_k(s) \quad (2.10)$$

for every  $k \in \mathcal{K}$ ,  $x \in \mathbb{R}_+$  and:

$$\bar{S}(t) = \int_0^t \varphi(\langle 1, e.\bar{\mu}(s) \rangle) ds \quad (2.11)$$

for all  $t < t_\rho(\xi)$  defined below. For  $t \geq t_\rho(\xi)$ ,  $\bar{A}(t) = \bar{D}(t) = \lambda t$ ,  $\bar{\mu}(t) = 0$ . The number  $t_\rho(\xi)$  is the time range of the solution, and is defined as:

$$\begin{cases} t_\rho(\xi) = \inf\{t : e.\bar{\mu}(t) = 0\} & \text{if } \xi \neq 0 \\ t_\rho(0) = 0 & \text{if } \rho \leq 1 \\ t_\rho(0) = \infty & \text{if } \rho > 1. \end{cases} \quad (2.12)$$

The interpretation of this definition parallels that of equations (2.1)–(2.5), with the difference that quantities are now assumed to be continuous, and that routing becomes a linear operation on flows.

Denote the total mass of  $\bar{\mu}_k(t)$  by

$$\bar{Z}_k(t) = \langle 1, \bar{\mu}_k(t) \rangle. \quad (2.13)$$

Since  $\xi_k$  is a finite measure for each  $k \in \mathcal{K}$ , let  $\nu_k^0$  be a probability measure such that  $\xi_k = \bar{Z}_k(0)\nu_k^0$ , where  $\bar{Z}_k(0) = \langle 1, \xi_k \rangle$ . For each  $k \in \mathcal{K}$ , let  $v_k$  and  $v_k^0$  be random variables with distributions  $\nu_k$  and  $\nu_k^0$ , respectively. As a particular case of (2.10), we have the law of evolution for  $Z_k$ :

$$\bar{Z}_k(t) = \bar{Z}_k(0)\mathbb{P}(v_k^0 > \bar{S}(t)) + \int_0^t \mathbb{P}(v_k > \bar{S}(t) - \bar{S}(s)) d\bar{A}_k(s) \quad (2.14)$$

since  $\langle 1_{[x, \infty[}, \mu_k(0) \rangle = Z_k(0)\mathbb{P}(v_k^0 > x)$  for all  $x \geq 0$ . In the following, we shall denote  $Z(t) = (Z_1(t), \dots, Z_K(t))'$ .

## 3 Preliminaries

### 3.1 Mapping to the single-class case

Since the processor sharing queue allocates processing speed to customer in function of their total number, the fact that one customer leaves the queue and re-enters it immediately does not modify the processing speed of the other customers. Accordingly, from the point of view of the global number of customers in the system, the system we have described evolves the same way as a single-class processor-sharing queue in which customers have a global service requirement equivalent to the sum of all individual service requirements in the multi-class model. We proceed with the identification of this service time distribution, and exploit existing results single class queues. For each  $k \in \mathcal{K}$ , we introduce the sequences  $\{V_k(i); i \geq 1\}$ , where  $V_k(i)$  is the total service time that is required by the  $i$ th exogenous customer of class  $k$  until its departure from the station. Likewise, let  $\{V_k^0(i); i \geq 1\}$  be the sequence of total service times for initial customers of class  $k$ . These sequences are i.i.d. Where there is no ambiguity, we shall denote also with

$V_k(x)$  the common distribution function of the random variable  $V_k$ , and similarly for  $V_k^0$ . Let  $\{\gamma(t), t \geq 0\}$  be a family of measures of  $\mathcal{M}$  defined by

$$\gamma(t) = \sum_{k=1}^K \left( \sum_{i=1}^{Z_k(0)} \delta_{(V_k^0(i) - S(t))^+}^+ + \sum_{i=1}^{E_k(t)} \delta_{(V_k(i) - S(t) + S(U_k(i)))^+}^+ \right). \quad (3.1)$$

Here,  $U_k(i)$  is the time at which the  $i^{\text{th}}$  arrival of class  $k$  enters the system. The quantity  $(V_k(i) - S(t) + S(U_k(i)))^+$  is the total residual service time of customer of class  $k$ , that is, the residual service times from its arrival to its departure of the system, and in the same way  $(V_k^0(i) - S(t))^+$  for the initial customers of class  $k$ . The measure  $\gamma(t)$  describes the sum of all total residual service times of customers which are present. Indeed, a customer is present at time  $t$  if and only if its residual service time is positive. The measure  $\delta^+$  eliminates those which have a zero residual service time. In particular, the quantity  $\langle 1_A, \gamma(t) \rangle$  is the number of customers in the system having residual service times comprised in the set  $A$ . Then  $\langle 1, \gamma(t) \rangle = e \cdot \bar{Z}(t)$  is the number of customers present in the system at time  $t$ . Finally, define the matrix:

$$\mathcal{B}(x) = \sum_{n=0}^{\infty} (\mathcal{B}P')^{*n}(x) \quad (3.2)$$

It is easy to see that this matrix is well defined since the series  $\sum_{n=0}^{\infty} (\|B\|_{\infty} P')^n$  converges. We can now state the result:

**Lemma 3.1.** *We have the following properties*

i) *The sequence  $\{V_k(i); i \geq 1\}$  is i.i.d. with common distribution function given by:*

$$V_k(x) = (e(I - P')(\mathcal{B} * B)(x))_k, \quad (3.3)$$

*with Laplace transform given by:*

$$\widehat{V}_k(s) = \left( e(I - P')(I - \widehat{B}(s)P')^{-1} \widehat{B}(s) \right)_k, \quad (3.4)$$

*and with the two first moments:*

$$\mathbb{E}(V_k) = (e\beta Q)_k, \quad \mathbb{E}(V_k^2) = \left( e(\beta^{(2)} + 2\beta P'Q\beta)Q \right)_k, \quad (3.5)$$

*where  $\beta = \text{diag}\{\langle \chi, \nu_k \rangle\}$  and  $\beta^{(2)} = \text{diag}\{\langle \chi^2, \nu_k \rangle\}$ .*

ii) *The sequence  $\{V_k^0(i); i \geq 1\}$  is i.i.d. with common distribution function given by:*

$$V_k^0(x) = (e(I - P')(\mathcal{B} * B^0)(x))_k. \quad (3.6)$$

*Proof.* The proofs of i) and ii) are similar. Let us prove i). Two techniques are possible.

*Forward reasoning.* The total service time of a customer of class  $k$  is the sum of one service time  $v_k$  and of the service it requires after the end of this service. The latter duration is distributed according to  $V_j$  with probability  $p_{kj}$  and is zero with probability  $p_{k0}$ . The service time after re-entering the queue is independent from the first service. Accordingly, we have the identity for distribution functions:

$$V_k(x) = \sum_j p_{kj} (B_k * V_j)(x) + p_{k0} B_k(x). \quad (3.7)$$

Expressed in vector-matrix form, with  $V(x) = (V_k(x); k \in \mathcal{K})$  (a row vector), we have:

$$V(x) = (V * (P'B))(x) + e(I - P')B(x).$$

By successive iterations of this identity, and using the fact that  $(P'B)^{*n}(x) \rightarrow 0$  when  $n \rightarrow \infty$ , we obtain:

$$V(x) = e(I - P')(\mathcal{B} * B)(x) . \quad (3.8)$$

*Extensive reasoning.* The total service time of a customer entering the system in the class  $k$  can be written as:

$$V_k = \sum_{m=0}^N v_{X(\ell_m)}^m ,$$

where  $\{X(i); i \in \mathbb{N}\}$  is a Markov chain on the space  $\mathcal{K} \cup \{0\}$  with the probability transition matrix  $P$  completed with the probabilities  $p_{k0}, k \in \mathcal{K}$  and  $p_{00} = 1$ . The random integer number  $N$  is the time at which a transition into state 0 occurs. The equality above is conditioned on the fact that  $k = X(0)$ . The random variables  $\{v_j^m; m \in \mathbb{N}\}$  are i.i.d. with common distribution  $B_j$ . Clearly,

$$\mathbb{P}\{X(1) = \ell_1, \dots, X(N-1) = \ell_{n-1}, N = n \mid X(0) = k\} = p_{k\ell_1} \dots p_{\ell_{n-2}\ell_{n-1}} p_{\ell_{n-1}0} . \quad (3.9)$$

Conditioning on the path of classes followed by the customer during its stay in the system, we obtain:

$$\begin{aligned} V_k(x) &= \sum_{n=1}^{\infty} \sum_{\ell_1, \dots, \ell_{n-1}} p_{k\ell_1} \dots p_{\ell_{n-2}\ell_{n-1}} p_{\ell_{n-1}0} (B_k * B_{\ell_1} * \dots * B_{\ell_{n-2}} * B_{\ell_{n-1}})(x) \\ &= \sum_{n=1}^{\infty} \sum_{\ell_{n-1}, \dots, \ell_1} p'_{0\ell_{n-1}} ((B_{\ell_{n-1}} p'_{\ell_{n-1}\ell_{n-2}}) * \dots * (B_{\ell_2} p'_{\ell_2\ell_1}) * (B_{\ell_1} p'_{\ell_1 k}) * B_k)(x) \\ &= \sum_{n=1}^{\infty} e(I - P') \left( [(BP')^{*(n-1)} * B](x) \right)_k \\ &= (e(I - P')(\mathcal{B} * B)(x))_k . \end{aligned}$$

This gives the result in Equation (3.8).

Next, applying the Laplace-Stieltjes transform to the identity (3.7), one obtains:

$$\widehat{V}_k(s) = \sum_j p_{kj} \widehat{B}_k(s) \widehat{V}_j(s) + p_{k0} \widehat{B}_k(s) .$$

The solution of this linear system of equations is given by (3.4). Consider now the moments of  $V_k$ . Taking expectations in (3.7) leads to the relations:

$$\begin{aligned} \mathbb{E}V_k &= \sum_j p_{kj} (\beta_k + \mathbb{E}V_j) + p_{k0} \beta_k \\ \mathbb{E}V_k^2 &= \sum_j p_{kj} (\beta_k^2 + 2\beta_k \mathbb{E}V_j + \mathbb{E}V_j^2) + p_{k0} \beta_k^2 . \end{aligned}$$

These linear systems of equations can be written as:

$$\begin{aligned} \mathbb{E}V &= eP'\beta + \mathbb{E}VP' + e(I - P')\beta = \mathbb{E}VP' + e\beta \\ \mathbb{E}V^2 &= eP'\beta^{(2)} + 2\mathbb{E}VP'\beta + \mathbb{E}V^2P' + e(I - P')\beta^{(2)} = \mathbb{E}V^2P' + 2\mathbb{E}VP'\beta + e\beta^{(2)} . \end{aligned}$$

The solution of this system leads to (3.5), since  $Q = (I - P')^{-1}$ .  $\square$

The distributions corresponding to  $V_k(\cdot)$  and  $V_k^0(\cdot)$  are respectively

$$\zeta_k = (e(I - P')(\mathcal{B} * \nu))_k \quad (3.10)$$

$$\zeta_k^0 = (e(I - P')(\mathcal{B} * \nu^0))_k . \quad (3.11)$$

It turns out that the state descriptor  $\gamma(\cdot)$  admits as fluid limit the solution of equation (see Proposition 6.2):

$$\langle 1_{[x,\infty)}, \bar{\gamma}(t) \rangle = \langle 1_{[x,\infty)}(\bar{S}(t)), \bar{\gamma}(0) \rangle + \sum_{k=1}^K \int_0^t \langle 1_{[x,\infty)}(\cdot - \bar{S}(t) + \bar{S}(s)), \zeta_k \rangle \alpha_k ds . \quad (3.12)$$

For all  $t < t_\rho$ , and  $\bar{\gamma}(t) = 0$  for all  $t \geq t_\rho$ . We have used the notation  $\bar{S}(t) - \bar{S}(s) = \int_s^t \varphi(\langle 1, \bar{\gamma}(u) \rangle) du$ ,  $\bar{\gamma}(0) = \sum_{k=1}^K \bar{Z}_k(0) \zeta_k^0$ , and  $t_\rho$  is defined by (2.12). Denote by  $C(\cdot)$  the matrix-function  $C(x) := (I - B^0(x)) + (I - B(x))P'Q$ . Observe that

$$\begin{aligned} (\mathcal{B} * \nu^0)([x, \infty)) &= (Q - \mathcal{B} * B^0(x)) \\ &= (\mathcal{B} * C)(x) . \end{aligned} \quad (3.13)$$

Likewise,  $(\mathcal{B} * \nu)([x, \infty)) = (\mathcal{B} * (I - B))Q(x)$ . Where  $\nu = \text{diag}\{\nu_k\}$  and  $\nu^0 = \text{diag}\{\nu_k^0\}$ . This implies by Equation (3.12)

$$\langle 1, \bar{\gamma}(t) \rangle = e(I - P')( \mathcal{B} * C)(\bar{S}(t)) \bar{Z}(0) + \int_0^t e(I - P')( \mathcal{B} * (I - B))(\bar{S}(t) - \bar{S}(s)) \lambda ds . \quad (3.14)$$

For all  $t < t_\rho$ , and  $\langle 1, \bar{\gamma}(t) \rangle = 0$  for all  $t \geq t_\rho$ .

Coming back to our multiclass queue, a customer taken ‘‘at random’’ in the external input flow will be of class  $k$  with probability  $\alpha_k/\alpha_e$ , where  $\alpha_e := \sum_{k \in \mathcal{K}} \alpha_k = e.\alpha$  is the ‘‘equivalent’’ arrival rate of single-class customers. Accordingly, the service time distribution of such a typical customer should be a mixture of the distributions  $V_k$  with these probabilities. The following result summarizes the properties of this distribution, which shall be useful in the remainder of the analysis.

**Lemma 3.2.** *Consider the random variable  $v^s$  which distribution  $B^s$  is formed as a mixture of the  $V_k$ , proportionally to the arrival rates  $\alpha_k$ :  $B^s(x) = \sum_{k \in \mathcal{K}} \alpha_k V_k(x)/\alpha_e$ . The Laplace transform of this distribution is given by:*

$$\widehat{B}^s(\theta) = \frac{1}{\alpha_e} e(I - P')(I - P'\widehat{B}(\theta))^{-1} \widehat{B}(\theta) \alpha .$$

Its two first moments are given by:

$$\begin{aligned} \mathbb{E}v^s &= \frac{e\beta\lambda}{\alpha_e} = \frac{\rho}{\alpha_e} \\ \mathbb{E}(v^s)^2 &= \frac{1}{\alpha_e} e(\beta^{(2)} + 2\beta P'Q\beta) \lambda . \end{aligned}$$

Let the excess lifetime distribution associated to  $B^s$  be denoted by  $B_e^s$ , with first and second moments  $\beta_e^s, \beta_e^{s,2}$ , respectively. Its Laplace transform satisfies the identities:

$$\widehat{B}_e^s(\theta) = \frac{1}{\rho\theta} \left( e.\alpha - e(I - P')(I - \widehat{B}(\theta)P')^{-1} \widehat{B}(\theta) \alpha \right) \quad (3.15)$$

$$= \frac{1}{\rho\theta} e(I - P')(I - \widehat{B}(\theta)P')^{-1} (I - \widehat{B}(\theta)) \lambda \quad (3.16)$$

$$= 1 - \theta \frac{1}{\rho} e \left( \frac{1}{2} \beta^{(2)} + \beta P'Q\beta \right) \lambda + o(\theta) . \quad (3.17)$$

*Proof.* The formulas for the Laplace-Stieltjes transform and the moments are direct consequences of Lemma 3.1. Expression (3.15) is the application of the classical formula  $\widehat{B}_e^s(\theta) = (1 - \widehat{B}^s(\theta))/(\theta\beta^s)$ . The second expression (3.16) is derived from the first one as:

$$\begin{aligned}\widehat{B}_e^s(\theta) &= \frac{1}{\rho\theta} \left( e.\alpha - e(I - P')(I - \widehat{B}(\theta)P')^{-1}\widehat{B}(\theta)\alpha \right) \\ &= \frac{1}{\rho\theta} \left( e(I - P')\lambda - e(I - P')(I - \widehat{B}(\theta)P')^{-1}\widehat{B}(\theta)(I - P')\lambda \right) \\ &= \frac{1}{\rho\theta} e(I - P')(I - \widehat{B}(\theta)P')^{-1}(I - \widehat{B}(\theta)P' + \widehat{B}(\theta)P' - \widehat{B}(\theta))\lambda\end{aligned}$$

and simplifying. Finally, the expansion (3.17) follows from the fact that  $\beta_e^s = \beta^{s,2}/2\beta^s$ .  $\square$

Consider now the Processor Sharing queue with a single class of customers having the service time distribution  $B^s$  with their first and second moment are denoted by  $\beta^s$  and  $\beta^{s,2}$ , and the arrival rate  $\alpha_e$ . We have the following results from the literature.

**Proposition 3.1.** *Assume that  $\alpha_e\beta^s > 1$ . There exists a unique positive solution  $\theta_0$  to the equation:*

$$\theta_0 = \alpha_e \left( 1 - \widehat{B}^s(\theta_0) \right).$$

If  $L(t)$  is the number of customers in the system at time  $t$ , then almost surely:

$$\lim_{t \rightarrow \infty} \frac{L(t)}{t} = \theta_0.$$

If  $D(t)$  is the number of customers that have departed the system at time  $t$ , then almost surely:

$$\lim_{t \rightarrow \infty} \frac{D(t)}{t} = \alpha_e - \theta_0.$$

### 3.2 A descriptor for the multiclass case

Our strategy is to introduce a family of state descriptors  $\gamma_{k_0k}(\cdot)$  where  $k_0, k \in \mathcal{K}$ . Each one describes the evolution of customers, who begin their service in class  $k_0$  and are presently in class  $k$ . For each  $k_0, k \in \mathcal{K}$  and for each  $i \geq 1$ , let  $N_{k_0k}(i)$  be the total number of visits to class  $k$  by the  $i^{\text{th}}$  customer arriving to the system as a customer of class  $k_0$ , and for each  $n = 1, \dots, N_{k_0k}(i)$  let  $V_{k_0k}(i, n)$  be the sum of service times required by this customer from its arrival until its  $n^{\text{th}}$  visit to class  $k$  (included). Likewise for initial customers,  $V_{k_0k}^0(i, n)$  is the total service time required by the  $i^{\text{th}}$  initial customer of class  $k_0$  until its  $n^{\text{th}}$  visit to class  $k$ . The departure process from class  $k$  can be expressed from these quantities as:

$$D_k(t) = \sum_{k_0=1}^K \left( \sum_{i=1}^{Z_{k_0}(0)} \sum_{n=1}^{N_{k_0k}(i)} \mathbf{1}_{\{V_{k_0k}^0(i,n) \leq S(t)\}} + \sum_{i=1}^{E_{k_0}(t)} \sum_{n=1}^{N_{k_0k}(i)} \mathbf{1}_{\{V_{k_0k}(i,n) \leq S(t) - S(U_{k_0}(i))\}} \right). \quad (3.18)$$

We define for each  $t \geq 0$ , the measure

$$\gamma_{k_0k}(t) = \sum_{i=1}^{Z_{k_0}(0)} \sum_{n=1}^{N_{k_0k}(i)} \delta_{(V_{k_0k}^0(i,n) - S(t))^+} + \sum_{i=1}^{E_{k_0}(t)} \sum_{n=1}^{N_{k_0k}(i)} \delta_{(V_{k_0k}(i,n) - S(t) + S(U_{k_0}(i)))^+}. \quad (3.19)$$

The interpretation for these equations is as follows. Consider the  $i^{\text{th}}$  customer to arrive in class  $k_0$ . The arrival epoch of this customer is  $U_{k_0}(i)$ . At time  $t$ , this customer has completed its  $n^{\text{th}}$  visit to class  $k$

(and therefore provoked a departure from class  $k$ ) if and only if  $V_{k_0k}(i, n) \leq S(t) - S(U_{k_0}(i))$ . Still at time  $t$ , the quantity  $(V_{k_0k}(i, n) - S(t) + S(U_{k_0}(i)))^+$  represents the remaining service quantity until this customer completes its  $n^{\text{th}}$  visit to class  $k$ , a sort of residual service time, but measured with respect to this  $n^{\text{th}}$  departure as a customer of class  $k$ . Similar observations apply to initial customers.

The quantity  $\langle 1_A, \gamma_{k_0k}(t) \rangle$  is therefore the the *remaining* number of visits to class  $k$  to be made by customers: a) present in the system at time  $t$ ; b) who have entered as customers of class  $k_0$ , and c) having cumulative residual service times for visits to class  $k$  comprised in the set  $A$ . Then  $\langle 1, \gamma_{k_0k}(t) \rangle$  is simply the total number of visits to class  $k$  remaining to be done by customers present in the system at time  $t$  and who have started as customers of class  $k_0$ . Let

$$\mathcal{Q}_k(t) = \sum_{k_0=1}^K \langle 1, \gamma_{k_0k}(t) \rangle \quad (3.20)$$

be the number of visits to class  $k$  remaining for customers present in the system at time  $t$ . Denote by

$$N_k(t) = \sum_{k_0=1}^K \left( \sum_{i=1}^{Z_{k_0}(0)} N_{k_0k}(i) + \sum_{i=1}^{E_{k_0}(t)} N_{k_0k}(i) \right) \quad (3.21)$$

the total number of visits to class  $k$  made by customers who arrived to the system in the time interval  $[0, t]$ . Then we have the following correspondence between processes  $D(\cdot)$ ,  $\mathcal{Q}(\cdot)$  and  $N$ ,

$$D(t) = N(t) - \mathcal{Q}(t) , \quad (3.22)$$

stating that the number of departures at time  $t$  is just the total potential number of departures for customers that have arrived, minus the number of remaining departures.

**Lemma 3.3.** *For each  $k_0, k \in \mathcal{K}$ , the sequence  $\{N_{k_0k}(i), i \geq 1\}$  is i.i.d. with distribution given by*

$$\mathbb{P}(N_{k_0k}(i) = m) = f_{k_0k}(f_{kk})^{m-1}(1 - f_{kk}) . \quad (3.23)$$

Here,  $f_{k_0k} = Q_{kk_0}/Q_{kk}$  if  $k_0 \neq k$  and  $f_{kk} = (P'Q)_{kk}/Q_{kk}$ .

*Proof.* Since the routing mechanism of customers are independent from one another, for each  $k_0, k \in \mathcal{K}$ , the sequence  $(N_{k_0k}(i), i \geq 1)$  is i.i.d. It is easy to observe that  $N_{k_0k}(i)$  is the number of visits of state  $k$  starting from state  $k_0$  in a Markov chain with state space  $\mathcal{K}$  and transition matrix  $P$ . Since  $\rho(P) < 1$  we have (3.23) (cf. [10, Section 5.3. Chapter 5]).  $\square$

**Lemma 3.4.** *Let  $k_0, k \in \mathcal{K}$  be such that  $Q_{kk_0} \neq 0$ . For each measurable function  $g : \mathbb{R}_+ \rightarrow \mathbb{R}$ , we denote by  $X_{k_0k}(i) := \sum_{n=1}^{N_{k_0k}(i)} g(V_{k_0k}(i, n))$  for all  $i \geq 1$ . The sequence  $(X_{k_0k}(i), i \geq 1)$  is i.i.d. with*

$$\mathbb{E}(X_{k_0k}(1)) = \langle g, (\mathcal{B} * \nu)_{kk_0} \rangle .$$

*Proof.* Since the routing and the service of each customer are independent of that of an other, the sequence  $(X_{k_0k}(i), i \geq 1)$  is i.i.d. Fix  $k_0, k \in \mathcal{K}$ . To compute the mean of  $X_{k_0k}(1)$ , observe that

$$\mathbb{E}(X_{k_0k}(1)) = \sum_{n=1}^{\infty} \mathbb{E}(g(V_{k_0k}(1, n))1_{\{N_{k_0k}(1) \geq n\}}) . \quad (3.24)$$

By iteration on  $n$  we compute  $\mathbb{E}(g(V_{k_0k}(1, n))1_{\{N_{k_0k}(1) \geq n\}})$ . First if  $k_0 \neq k$  then

$$\begin{aligned} & \mathbb{E}(g(V_{k_0k}(1, 1))1_{\{N_{k_0k}(1) \geq 1\}}) \\ &= \mathbb{E}(g(v_{k_0}(1, 1) + v_k(2, 1))p_{k_0k}) \\ &+ \sum_{m=1}^{\infty} \sum_{l_1, \dots, l_m \in \bar{\mathcal{K}}} (p_{k_0l_1}p_{l_1l_2} \dots p_{l_{m-1}l_m} \mathbb{E}(g(v_{k_0}(1, 1) + \sum_{j=1}^m v_{l_j}(j+1, 1) + v_k(m+2, 1)))) \\ &= \langle g, ((BP')_{k_0k} + \sum_{n=0}^{\infty} (BP')_{k\bar{\mathcal{K}}} ((BP')_{\bar{\mathcal{K}}\bar{\mathcal{K}}})^{*n} (BP')_{\bar{\mathcal{K}}k_0}) * \nu_{k_0} \rangle, \end{aligned} \quad (3.25)$$

where  $\bar{\mathcal{K}} = \mathcal{K} \setminus \{k\}$  and  $v_{l_j}(j+1, 1)$  denotes the service time of the first customer in  $(j+1)^{th}$  step at class  $l_j$ . On the other hand, if  $k_0 = k$ , then  $V_{kk}(1, 1) = v_k(1, 1)$  and

$$\mathbb{E}(g(V_{kk}(1, 2))1_{\{N_{kk}(1) \geq 2\}}) = \langle g, \left( \sum_{n=0}^{\infty} (BP')_{k\bar{\mathcal{K}}} ((BP')_{\bar{\mathcal{K}}\bar{\mathcal{K}}})^{*n} (BP')_{\bar{\mathcal{K}}k} \right) * \nu_k \rangle. \quad (3.26)$$

A shorthand notation will be useful in the following: for each  $x \in \mathbb{R}_+$ , we denote by  $B := B(x)$ ,

$$\Psi_{kl} = (BP')_{kl} + \sum_{n=0}^{\infty} (BP')_{k\bar{\mathcal{K}}} ((BP')_{\bar{\mathcal{K}}\bar{\mathcal{K}}})^{*n} (BP')_{\bar{\mathcal{K}}l}$$

and  $\Psi_k$  the vector function with components  $(\Psi_{kl}, l \in \mathcal{K})$ . From Equation (3.25), we have for all  $k_0, k$  such that  $k_0 \neq k$ ,

$$\mathbb{E}(g(V_{k_0k}(1, 1))1_{\{N_{k_0k}(1) \geq 1\}}) = \langle g, \Psi_{kk_0} * \nu_{k_0} \rangle.$$

For each  $n \geq 1$  we introduce  $V_{kk}^*(i, n+1)$  as the service times for  $i^{th}$  customer computed just after the  $(n)^{th}$  visit in class  $k$  until that of the  $(n+1)^{th}$  visit. For each  $k \in \mathcal{K}$ , the sequence  $(V_{kk}^*(1, j), j \geq 2)$  is i.i.d. Proceeding as above, one deduces  $\mathbb{E}(g(V_{kk}^*(1, 2))1_{\{N_{kk}(1) \geq 2\}}) = \langle g, \Psi_{kk} \rangle$ . Since for all  $n \geq 2$  we have  $V_{kk}(1, n) = v_k(1) + \sum_{j=2}^n V_{kk}^*(1, j)$ . Then by independence of  $v_k(1)$  and  $(V_{kk}^*(1, j), j \geq 2)$ , we have

$$\mathbb{E}(g(V_{kk}(1, n))1_{\{N_{kk}(1) \geq n\}}) = \langle g, \Psi_{kk}^{*(n-1)} * \nu_k \rangle.$$

Let  $l, k \in \mathcal{K}$  such that  $l \neq k$ . Since for each  $n \geq 2$ , we have  $V_{lk}(1, n) = V_{lk}(1, 1) + \sum_{j=2}^n V_{kk}^*(1, j)$ . Then by independence of  $V_{lk}(1, 1)$  and  $V_{kk}^*(1, j)$ , we have

$$\mathbb{E}(g(V_{lk}(1, n))1_{\{N_{lk}(1) \geq n\}}) = \langle g, \Psi_{kk}^{*(n-1)} * \Psi_{kl} * \nu_l \rangle.$$

Hence the Equation (3.24) is,

$$\mathbb{E}(g(X_{k_0k}(1))) = \begin{cases} \langle g, \sum_{n=0}^{\infty} \Psi_{kk}^{*(n)} * \Psi_{kk_0} * \nu_{k_0} \rangle & \text{if } k_0 \neq k \\ \langle g, \sum_{n=0}^{\infty} \Psi_{kk}^{*(n)} * \nu_k \rangle & \text{if } k_0 = k. \end{cases}$$

To conclude the proof of lemma 3.4, it suffices to prove that

$$\sum_{n=0}^{\infty} (\Psi_{kk})^{*(n)} = \mathcal{B}_{kk} \text{ and } \sum_{n=0}^{\infty} (\Psi_{kk})^{*(n)} * \Psi_{kk_0} = \mathcal{B}_{kk_0}. \quad (3.27)$$

For each  $k \in \mathcal{K}$ , let  $\hat{B}_k$  be the Laplace Transform of the distribution  $B_k$  and  $\hat{\mathcal{B}}$  be the Laplace transform of the matrix function  $\mathcal{B} := \sum_{n=0}^{\infty} (BP')^{*n}$ , which is  $\hat{\mathcal{B}} = (I - \hat{BP}')^{-1}$ . From Schur's formula we have for the partition  $\{k\}, \bar{\mathcal{K}}$ :

$$\hat{B}_{kk} = \left( 1 - (\hat{BP}')_{k\bar{\mathcal{K}}} (I - (\hat{BP}')_{\bar{\mathcal{K}}\bar{\mathcal{K}}})^{-1} (\hat{BP}')_{\bar{\mathcal{K}}k} \right)^{-1} \quad (3.28)$$

$$\hat{B}_{kk}^{-1} \hat{\mathcal{B}}_{k\bar{\mathcal{K}}} = (\hat{BP}')_{k\bar{\mathcal{K}}} (I - (\hat{BP}')_{\bar{\mathcal{K}}\bar{\mathcal{K}}})^{-1}. \quad (3.29)$$

Since the Laplace transform of  $\Psi_{kk}$  is  $(\hat{B}P')_{k\bar{K}}(I - (\hat{B}P')_{\bar{K}\bar{K}})^{-1}(\hat{B}P')_{\bar{K}k}$  and  $\hat{B}_{kk} > 0$ , we have from (3.28)

$$\hat{B}_{kk} = \frac{1}{1 - \hat{\Psi}_{kk}}.$$

Hence,  $\sum_{n=0}^{\infty} (\hat{\Psi}_{kk})^n = \hat{B}_{kk}$  and by the uniqueness of the Laplace transform, the first identity in (3.27) is satisfied. For the second identity, we have by definition of  $\Psi_k$

$$\begin{aligned} \hat{\Psi}_{k\bar{K}} &= (\hat{B}P')_{k\bar{K}} + (\hat{B}P')_{k\bar{K}}(I - (\hat{B}P')_{\bar{K}\bar{K}})^{-1}(\hat{B}P')_{\bar{K}\bar{K}} \\ &= (\hat{B}P')_{k\bar{K}} \left( I + (\hat{B}P')_{\bar{K}\bar{K}}(I - (\hat{B}P')_{\bar{K}\bar{K}})^{-1} \right) \\ &= (\hat{B}P')_{k\bar{K}}(I - (\hat{B}P')_{\bar{K}\bar{K}})^{-1} \end{aligned}$$

By (3.29), for  $k_0 \in \bar{K}$  we have  $\hat{\Psi}_{kk_0} = \hat{B}_{kk}^{-1} \hat{B}_{kk_0}$ . Then the Laplace transform of  $\sum_{n=0}^{\infty} (\Psi_{kk})^{*(n)} * \Psi_{kk_0}$  is  $\hat{B}_{kk_0}$ .  $\square$

It turns out that the state descriptor  $\gamma_{k_0k}(\cdot)$  admits a fluid limit (see Proposition 6.3), the solution of,

$$\begin{aligned} \langle 1_{[x,\infty)}, \bar{\gamma}_{k_0k}(t) \rangle &= \langle 1_{[x,\infty)}(\cdot - \bar{S}(t)), \bar{\gamma}_{k_0k}(0) \rangle \\ &\quad + \int_0^t \langle 1_{[x,\infty)}(\cdot - \bar{S}(t) + \bar{S}(s)), (\mathcal{B} * \nu)_{kk_0} \rangle \alpha_{k_0} ds. \end{aligned} \quad (3.30)$$

For all  $t < t_\rho$ , and  $\bar{\gamma}_{k_0k}(t) = 0$  for all  $t \geq t_\rho$ . We have used the notation  $\bar{S}(t) - \bar{S}(s) = \int_s^t \varphi(\langle 1, \bar{\gamma}(u) \rangle) du$ ,  $\bar{\gamma}_{k_0k}(0) = (\mathcal{B} * \nu^0)_{kk_0} \bar{Z}_{k_0}(0)$ , and  $t_\rho$  is defined by (2.12). Applying Equation (3.13), the fluid model of the total visits is a solution to the following equation:

$$\bar{Q}(t) = (\mathcal{B} * C)(\bar{S}(t)) \bar{Z}(0) + \int_0^t (\mathcal{B} * (I - B))(\bar{S}(t) - \bar{S}(s)) \lambda ds. \quad (3.31)$$

For all  $t < t_\rho$ , and  $\bar{Q}(t) = 0$  for all  $t \geq t_\rho$ .

## 4 Principal Results

The presentation begins with the properties of the trajectories of Fluid Solutions according to Definition 2.1. We obtain results of two types: first, we have existence and uniqueness results, and the construction of solutions in analytical form; second, we have general asymptotic results. We finish the section with the presentation of the fluid normalization, the assumptions and the convergence result.

### 4.1 Existence and Uniqueness

Let  $\mathcal{M}^{c,K} = \{\xi \in \mathcal{M}_+^K : \xi_k\{x\} = 0 \text{ for all } x \in \mathbb{R}_+ \text{ and } k \in \mathcal{K}\}$  be a set of finite, non-negative Borel measures on  $\mathbb{R}_+$  that have no atoms.

**Theorem 4.1.** *Given data  $(\alpha, \nu, P)$  and  $\xi \in \mathcal{M}^{c,K}$ , there exists a unique fluid solution  $(\bar{A}(t), \bar{D}(t), \bar{\mu}(t))$  of the model such that  $\bar{\mu}(0) = \xi$ . Moreover, this solution is such that:*

$$\bar{W}(t) = (\bar{W}(0) + (\rho - 1)t)^+, \quad (4.1)$$

where  $\bar{W}(\cdot)$  is the total workload defined as

$$\bar{W}(t) := \langle \chi, e \cdot \bar{\mu}(t) \rangle + \beta P' Q \bar{Z}(t). \quad (4.2)$$

**Lemma 4.1.** *Let  $(\alpha, \nu, P)$  be a supercritical data. Then there exists a unique positive real number  $\theta_0$  solution to the equation:*

$$\theta_0 = e (I - \widehat{B}(\theta_0))(I - P'\widehat{B}(\theta_0))^{-1} \alpha . \quad (4.3)$$

*Proof.* Rewrite (4.3) as:

$$\begin{aligned} \theta_0 &= e (I - P'\widehat{B}(\theta_0) + P'\widehat{B}(\theta_0) - \widehat{B}(\theta_0))(I - P'\widehat{B}(\theta_0))^{-1} \alpha \\ &= e.\alpha - e(I - P')\widehat{B}(\theta_0)(I - P'\widehat{B}(\theta_0))^{-1} \alpha \\ &= \alpha_e \left( 1 - \widehat{B}^s(\theta_0) \right) . \end{aligned}$$

Proposition 3.1 applies, and we have the existence and uniqueness of  $\theta_0$ .  $\square$

Define the vector  $m = (m_1, \dots, m_K)'$  as:

$$m = (I - \widehat{B}(\theta_0))(I - P'\widehat{B}(\theta_0))^{-1} \alpha . \quad (4.4)$$

Given a supercritical data  $(\alpha, \nu, P)$ , define  $p_k : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ , for each  $k \in \mathcal{K}$  by

$$p_k(x) = \frac{m_k}{1 - \widehat{B}_k(\theta_0)} \int_x^\infty \theta_0 e^{-\theta_0(y-x)} dB_k(y) , \quad (4.5)$$

and let  $s_k \in \mathcal{M}$  denote the measure that is absolutely continuous with respect to Lebesgue measure and which Radon-Nikodym derivative is  $p_k(\cdot)$ :

$$s_k(x) = p_k(x) dx \quad \text{for all } x \in \mathbb{R}_+ . \quad (4.6)$$

Note that  $\int_0^\infty p_k(x) dx = m_k$  and  $\langle 1, s_k \rangle = m_k$ . Finally, define  $s := (s_1, \dots, s_K)'$ .

**Theorem 4.2.** *Assume that  $(\alpha, \nu, P)$  is a supercritical data, and let  $\theta_0$  and  $s$  be given respectively by (4.3) and (4.4)–(4.6). Then the triple*

$$(A^0(t), D^0(t), \mu^0(t)) = ((I - \widehat{B}(\theta_0))^{-1} mt, (I - \widehat{B}(\theta_0))^{-1} \widehat{B}(\theta_0) mt, ts) \quad (4.7)$$

*is the unique fluid solution of the model starting from the origin, that is, with  $\bar{\mu}(0) \equiv 0$ . As a consequence,  $\bar{Z}(t) = mt$ .*

## 4.2 Asymptotics of fluid solutions for critical and supercritical data

Denote  $\mathcal{M}^{c,p,K} = \{\xi \in \mathcal{M}^{c,K} : \xi \neq 0\}$  the set of non-zero measures. Let  $\beta_k^{(2)}$  be the second moment of the distribution  $B_k$ , and  $\beta^{(2)} = \text{diag}(\beta_k^{(2)}; k \in \mathcal{K})$ . Let  $\nu_k^e$  be the excess life (or forward recurrence time) probability measure associated with  $\nu_k$ .

**Theorem 4.3.** *Given a critical data  $(\alpha, \nu, P)$  and  $\xi \in \mathcal{M}^{c,p,K}$  and assuming that  $\beta_k^0 = \langle \chi, \xi_k \rangle < \infty$  and  $\beta_k = \langle \chi, \nu_k \rangle < \infty$  for all  $k \in \mathcal{K}$ , then, as  $t \rightarrow \infty$ ,*

$$\mu_k(t)(\cdot) \implies \frac{e(\beta^0 + \beta Q P') \bar{Z}(0)}{e(\frac{1}{2}\beta^{(2)} + \beta P' Q \beta) \lambda} \beta_k \lambda_k \nu_k^e(\cdot) .$$

*If  $\beta_j^{(2)} = +\infty$  for some  $j \in \mathcal{K}$ , the limit is 0.*

**Theorem 4.4.** *Given a supercritical data  $(\alpha, \nu, P)$  and  $\xi \in \mathcal{M}^{c,K}$ , there holds:*

$$\frac{\mu_k(t)}{t}(\cdot) \implies s_k(\cdot) .$$

*As a consequence,*

$$\lim_{t \rightarrow \infty} \frac{\bar{A}(t)}{t} = \lambda - Q P' m \quad \lim_{t \rightarrow \infty} \frac{\bar{D}(t)}{t} = \lambda - Q m .$$

### 4.3 Extension to the Discriminatory Processor Sharing queue

A natural generalization of multiclass processor sharing (egalitarian), commonly encountered in the literature, is the “discriminatory” processor sharing (DPS), where all customers present in the system are served simultaneously with rates controlled by a vector of weights  $\{g_k > 0, k \in \mathcal{K}\}$ . Under the DPS discipline, any individual customer of class  $k$  is served at a speed which is proportional to  $g_k$ . Accordingly, it is quickly seen that the cumulative of service per customer of class  $k$  should be,

$$S_k(t) = \int_0^t g_k \varphi(\langle 1, g \cdot \bar{\mu}(s) \rangle) ds, \quad (4.8)$$

and that the dynamics of the measure  $\bar{\mu}_k$  is:

$$\langle 1_{[x, \infty)}, \mu_k(t) \rangle = \langle 1_{[x, \infty)}(\cdot - S_k(t)), \mu_k(0) \rangle + \int_0^t \langle 1_{[x, \infty)}(\cdot - (S_k(t) - S_k(s))), \nu_k \rangle dA_k(s). \quad (4.9)$$

When all the weights are multiplied by some scalar  $\gamma$ , the dynamics are not changed since  $\langle 1, \gamma g \cdot \bar{\mu}(s) \rangle = \gamma \langle 1, g \cdot \bar{\mu}(s) \rangle$  and  $\gamma g_k \varphi(\langle 1, \gamma g \cdot \bar{\mu}(s) \rangle) = g_k \varphi(\langle 1, g \cdot \bar{\mu}(s) \rangle)$ . In particular, when all weights are equal, this coincides with the equations of the multiclass egalitarian processor sharing system.

This fluid queueing model is described by the data  $(\alpha, P, \nu, g)$ . Definition 2.1 naturally extends to this system: we shall call a DPS Fluid Solution a triple of vector functions and measures  $(\bar{A}(t), \bar{D}(t), \bar{\mu}(t))$  that satisfy Equations (2.8)–(2.9) and (4.8)–(4.9). We proceed to show that such a DPS Fluid solution can be constructed from an equivalent (egalitarian) PS Fluid solution. To that end, consider the following transformations.

Let  $G$  be the diagonal matrix obtained from  $g$ . Define  $(\alpha^g, P^g, \nu^g)$  by

$$\begin{cases} \alpha^g &= G^{-1} \alpha \\ P^g &= G^{-1} P G \\ \nu_k^g(\cdot) &= \nu_k(g_k \times \cdot). \end{cases} \quad (4.10)$$

Now consider a triple  $(A^g(t), D^g(t), \mu^g(t))$  of vector functions and measures as usual. Define then the transformed functions  $A, D$  and measures  $\mu$  by:

$$\begin{cases} \bar{A}(t) &= G^{-1} A^g(t) \\ \bar{D}(t) &= G^{-1} D^g(t) \\ \bar{\mu}_k(\cdot)(t) &= \frac{1}{g_k} \bar{\mu}_k^g(\frac{1}{g_k} \times \cdot)(t). \end{cases} \quad (4.11)$$

Observe that the triple  $(\alpha^g, P^g, \nu^g)$  may not be a valid data, since it may be that some entries in  $P$  are larger than 1. However, it is always true that  $P^g$  is a positive matrix with the same spectral radius as  $P$ . We conjecture that the results for the fluid process do hold even if  $P$  is not substochastic, under the condition  $\rho(P) < 1$ , and although the interpretation of the entries  $p_{ij}$  as routing *probabilities* does not necessarily hold.

**Proposition 4.1.** *Assume that the triple  $(\alpha^g, P^g, \nu^g)$  is a valid data. The triple  $(A^g(t), D^g(t), \mu^g(t))$  is a solution of the egalitarian Fluid model with data  $(\alpha^g, P^g, \nu^g)$  defined in (4.10), and initial state described by the measures  $\bar{\mu}_k^g(\cdot)(0) = (g_k)^{-1} \bar{\mu}_k((g_k)^{-1} \times \cdot)(0)$ , if and only if the triple  $(\bar{A}(t), \bar{D}(t), \bar{\mu}(t))$  defined by (4.11) is a DPS Fluid solution for the DPS model, and initial measure  $\bar{\mu}(0)$ .*

Let  $T_k(s) = \bar{S}_k^{-1}(s)$ . We shall see later that this function represents the response time of customers, as a function of their service time. We have the following ordering property.

**Proposition 4.2.** *Let  $(\alpha, \nu, P)$  be a given data, let  $g = \{g_k > 0, k \in \mathcal{K}\}$  be a vector of weights and  $\xi$  be a non-zero initial state. If  $g_k \geq g_l$ , then for all  $t \geq 0$ :*

$$T_k(t) \leq T_l(t).$$

#### 4.4 Convergence to the fluid model solution

Consider a sequence of multiclass processor sharing queues indexed by integer numbers  $r$ . Assume that this model is defined on probability space  $(\Omega, \mathbb{P}^r)$ , and to have the same basic structure as described in Section 2.1. Furthermore the number of classes  $K$  and the set  $\mathcal{A} = \{k : \alpha_k^r \neq 0\}$  remain fixed for all  $r$ . The primitive increments are denoted by  $u_k^r = \{u_k^r(i), i \geq 1\}$  and  $v_k^r = \{v_k^r(i), i \geq 1\}$  and  $\varphi^{r,k} = \{\varphi^{r,k}(i), i \geq 1\}$ , for all  $k \in \mathcal{K}$ , and  $\alpha^r, \nu^r$  and  $P^r$  their parameters respectively.

**Assumptions for primitive data.** It is assumed that

$$u_1^r, \dots, u_K^r, v_1^r, \dots, v_K^r, \varphi^{1,r}, \dots, \varphi^{K,r} \quad (4.12)$$

are mutually independent and  $\langle \chi, \nu_k^r \rangle < \infty$ , for all  $k \in \mathcal{K}$ .

$$\alpha_k^r \rightarrow \alpha_k, \text{ for all } k \in \mathcal{K} \quad (4.13)$$

$$p_{lk}^r \rightarrow p_{lk} \text{ for all } k, l \in \mathcal{K} . \quad (4.14)$$

The matrix  $P = (p_{lk})_{l,k \in \mathcal{K}}$  is nonnegative such that  $\rho(P) < 1$ .

$$\nu_k^r \Rightarrow \nu_k \quad (4.15)$$

$$\langle \chi, \nu_k^r \rangle \rightarrow \langle \chi, \nu_k \rangle \quad (4.16)$$

$$\mathbb{E}(u_k^r(1); u_k^r(1) > r) \rightarrow 0 . \quad (4.17)$$

Assumptions (4.14) and (4.17) guarantee that the traffic intensity converges:

$$\rho^r = \sum_{k=1}^K \alpha_k^r \langle \chi, \nu_k^r \rangle \rightarrow \rho = \sum_{k=1}^K \alpha_k \langle \chi, \nu_k \rangle .$$

**Assumption for initial conditions.** We assume that the sequence of initial service times for each  $k \in \mathcal{K}$  is denoted by  $v_k^{0,r} = \{v_k^{0,r}(i), i \geq 1\}$  with  $\nu_k^{0,r}$  its distribution, and the initial number of customers is  $\bar{Z}_k^r(0)$ .

$$v_1^{0,r}, \dots, v_K^{0,r}, v_1^r, \dots, v_K^r, \varphi^{1,r}, \dots, \varphi^{K,r}, \bar{Z}_1^r(0), \dots, \bar{Z}_K^r(0) \quad (4.18)$$

are mutually independent and  $\langle \chi, \nu_k^{0,r} \rangle < \infty$ , for all  $k \in \mathcal{K}$ . Moreover, assume that there exist a vector  $\bar{Z}(0) = (\bar{Z}_1(0), \dots, \bar{Z}_K(0)) \in \mathbb{R}_+^K$  and measure of valued vector  $\nu^0 = (\nu_1^0, \dots, \nu_K^0) \in \mathcal{M}^K$  such that for each  $k \in \mathcal{K}$  we have

$$\bar{Z}^r(0) \Rightarrow \bar{Z}(0) \quad (4.19)$$

$$\nu_k^{0,r} \Rightarrow \nu_k^0 \quad (4.20)$$

$$\langle \chi, \nu_k^{0,r} \rangle \rightarrow \langle \chi, \nu_k^0 \rangle \quad (4.21)$$

$$\langle 1_{\{x\}}, \nu_k^0 \rangle = 0 \quad \text{for all } x \in \mathbb{R}_+ . \quad (4.22)$$

**The normalized process.** The scaled processes that will give rise to a fluid limit is defined as:

$$\begin{aligned} \bar{A}_k^r(t) &= \frac{A_k^r(rt)}{r}, \bar{D}_k^r(t) = \frac{D_k^r(rt)}{r}, \bar{E}_k^r(t) = \frac{E_k^r(rt)}{r}, \bar{\Phi}_{lk}^r(t) = \frac{\Phi_{lk}^r([rt])}{r} \\ \bar{Z}_k^r(t) &= \frac{Z_k^r(rt)}{r}, \bar{W}^r(t) = \frac{W^r(rt)}{r} \\ \bar{S}_{t,t+h}^r &= S_{rt,r(t+h)}^r = \int_{rt}^{r(t+h)} \varphi(\langle 1, e \cdot \mu^r(s) \rangle) ds = \int_t^{t+h} \varphi(e \cdot \bar{Z}^r(s)) ds . \end{aligned}$$

**Theorem 4.5.** *Consider a sequence of multiclass processor sharing queues as defined above, satisfying assumptions (4.12)-(4.22). Then the sequence  $(\bar{A}^r, \bar{D}^r, \bar{\mu}^r)$  converges in distribution to  $(\bar{A}, \bar{D}, \bar{\mu})$  solution to the fluid model such that  $\bar{\mu}_k(0) = \bar{Z}_k(0)\nu_k^0$  for all  $k \in \mathcal{K}$ .*

## 5 Proofs for the fluid model

The proof of the results follows the lines of [19]. We begin with a reduction of the system to a single-class processor-sharing queue, which provides an useful insight.

### 5.1 Results for a nonzero initial state

The first step is to establish that the function  $\bar{S}(t)$  does not converge, which means, in practice, that every customer in the system eventually completes its service and leaves the system. Since this property is a global one, it should be possible to use the equivalent single-class system of Section 3.1. However, the equations (2.9) or (2.14) concerning the number of customers do not lead immediately to the analogous ones in the single class queue. A separate proof is therefore necessary.

This section is devoted to the proof of Theorem 4.1 in the case of positive measures, a result that we state as a separate proposition for easier reference.

**Proposition 5.1.** *Given data  $(\alpha, \nu, P)$  and  $\xi \in \mathcal{M}^{c,p,K}$ , there exists a unique fluid solution  $(\bar{A}(t), \bar{D}(t), \bar{\mu}(t))$  of the model such that  $\bar{\mu}(0) = \xi$ . Moreover, this solution satisfies (4.1).*

#### 5.1.1 Time range of the solution

The time range of a fluid solution is defined by (2.12). Since  $\xi \neq 0$  in this section, we have  $t_\rho(\xi) = \inf\{t : e.\bar{\mu}(t) = 0\}$ .

**Lemma 5.1.** *Consider a fluid solution for data  $(\alpha, \nu, P)$  and  $\xi \in \mathcal{M}^{c,p,K}$  with  $\xi \neq 0$ . Denote  $t^* := t_\rho(\xi) = \inf\{t : e.\bar{\mu}(t) = 0\}$  the first time at which the fluid queue empties. Then the function  $\bar{S}(t)$  defined by (2.11) is continuous, strictly increasing and differentiable on  $[0, t^*)$ . Moreover  $\lim_{t \rightarrow t^*} \bar{S}(t) = +\infty$ .*

*Proof.* From (2.13) we have  $e.\bar{Z}(t) = \langle 1, e.\mu(t) \rangle$  and therefore  $e.\bar{Z}(t) > 0$  for all  $t \in [0, t^*)$ . Consequently,

$$\bar{S}(t) = \int_0^t \frac{1}{e.\bar{Z}(u)} du \quad \text{for all } t < t^* .$$

Since  $t \rightarrow \bar{\mu}(t)$  is assumed to be continuous then  $t \rightarrow \bar{Z}(t)$  is continuous and  $t \rightarrow \bar{S}(t)$  is increasing and differentiable with  $d\bar{S}(t)/dt = 1/e.\bar{Z}(t)$  for all  $t < t^*$ . For the limit  $t \rightarrow t^*$ , we have two cases.

**Case 1:**  $t^* = +\infty$ . Using Equations (2.8), (2.9) and (2.14), we have

$$\bar{Z}(t) = \bar{Z}(0) + \alpha t - (I - P')\bar{D}(t)$$

which implies that

$$\begin{aligned} e\bar{Z}(t) &= e\bar{Z}(0) + e.\alpha t - e(I - P')\bar{D}(t) \\ &\leq e\bar{Z}(0) + e.\alpha t . \end{aligned}$$

Indeed, since  $P$  is substochastic,  $Pe' \leq e'$  which implies that  $e(I - P') \geq 0$ . It follows that

$$\begin{aligned} \bar{S}(t) &\geq \int_0^t \frac{1}{e.\bar{Z}(0) + e.\alpha u} du \\ &\geq \frac{1}{e.\alpha} \log\left(1 + \frac{e.\alpha t}{e.\bar{Z}(0)}\right) \end{aligned}$$

hence  $\lim_{t \rightarrow t^*} \bar{S}(t) = +\infty$ .

**Case 2:**  $t^* < +\infty$ . The proof is by contradiction. Assume that there exists a finite limit  $\bar{S}(t^*)$ . Let  $k$  be one index such that  $\alpha_k \neq 0$ . Fix  $\varepsilon > 0$ . Since the distribution of  $v_k$  has no mass at 0 (one of the general assumptions of this paper), for this  $\varepsilon$  there exists a  $\delta > 0$  such that  $\mathbb{P}(v_k > \delta) > \varepsilon$ . Next by continuity of  $\bar{S}(t)$ , there exists a  $t_\delta < t^*$  such that  $\bar{S}(t^*) - \bar{S}(u) < \delta$  for all  $u \in [t_\delta, t^*]$ . Finally, observe that, as a consequence of Equation (2.8),  $dA_k(t)/dt \geq \alpha_k$ . It follows that, with Equation (2.14), for all  $t \in [t_\delta, t^*]$ ,

$$\begin{aligned} Z_k(t) &\geq \alpha_k \int_0^t \mathbb{P}(v_k > \bar{S}(t) - \bar{S}(u)) du \\ &\geq \alpha_k \int_0^t \mathbb{P}(v_k > \bar{S}(t^*) - \bar{S}(u)) du \\ &\geq \alpha_k \int_0^t \mathbb{P}(v_k > \delta) du \\ &\geq \alpha_k \int_0^{t_\delta} \mathbb{P}(v_k > \delta) du + \alpha_k \int_{t_\delta}^t \mathbb{P}(v_k > \delta) du \\ &\geq \alpha_k \varepsilon (t - t_\delta) . \end{aligned}$$

Therefore, by continuity of  $\bar{Z}(\cdot)$ , we have from the above  $Z_k(t^*) \geq \alpha_k \varepsilon (t^* - t_\delta) > 0$ . This is a contradiction since  $Z_k(t^*)$  must be 0.  $\square$

### 5.1.2 Existence and uniqueness for $\xi \neq 0$

According to Lemma 5.1, the function  $\bar{S}(t)$  maps  $[0, t^*)$  to  $[0, +\infty)$  and can be inverted. Let  $T : [0, +\infty) \rightarrow [0, t^*)$  be defined as  $T(s) = \bar{S}^{-1}(s)$ . This function is also continuous and differentiable since  $\dot{\bar{S}}(t) = 1/e \cdot \bar{Z}(t) < +\infty$ , strictly increasing because  $e \cdot \bar{Z}(t) > 0$ , and with  $T(0) = 0$ . Let:

$$\tilde{A}(t) = \bar{A}(T(t)), \quad \tilde{D}(t) = \bar{D}(T(t)), \quad \tilde{Z}(t) = \bar{Z}(T(t)) .$$

The vector functions  $\tilde{A}(\cdot)$ ,  $\tilde{D}(\cdot)$ , and  $\tilde{Z}(\cdot)$  are defined for  $t \in [0, +\infty)$  and are continuous and increasing. Performing the change of variables and functions in equations (2.8), (2.9), (2.11) and (2.14) gives the new functional equations:

$$\tilde{A}(t) = \alpha T(t) + P' \tilde{D}(t) \tag{5.1}$$

$$\tilde{Z}(t) = \bar{Z}(0) + \tilde{A}(t) - \tilde{D}(t) \tag{5.2}$$

$$\tilde{Z}_k(t) = Z_k(0) \mathbb{P}(v_k^0 > t) + \int_0^t \mathbb{P}(v_k > t - s) d\tilde{A}_k(s) \tag{5.3}$$

for every  $k \in \mathcal{K}$  and  $t \geq 0$ , and:

$$T(t) = \int_0^t e \cdot \tilde{Z}(s) ds . \tag{5.4}$$

Observe that the input/output equations (5.1)–(5.2) are equivalent, eliminating  $\bar{D}(\cdot)$  or  $\bar{A}(\cdot)$ , to:

$$\bar{A}(t) = \lambda t + QP'(\bar{Z}(0) - \bar{Z}(t)) \tag{5.5}$$

$$\bar{D}(t) = \lambda t + Q(\bar{Z}(0) - Q\bar{Z}(t)) \tag{5.6}$$

which gives, after the change of variable:

$$\tilde{A}(t) = \lambda T(t) + QP'(\bar{Z}(0) - \tilde{Z}(t)) \tag{5.7}$$

$$\tilde{D}(t) = \lambda T(t) + Q(\bar{Z}(0) - \tilde{Z}(t)) . \tag{5.8}$$

**Lemma 5.2.** *Given a data  $(\alpha, \nu, P)$  and  $\xi \in \mathcal{M}^{c,p,K}$ . There exists a unique solution  $(\tilde{A}, \tilde{D}, \tilde{Z})$  to equations (5.1)–(5.4) such that:*

(i) *the functions  $\tilde{A}$ ,  $\tilde{D}$  and  $\tilde{Z}$  are continuous;*

(ii) *the functions  $\tilde{A}$  and  $\tilde{D}$  are increasing with  $\tilde{A}(0) = \tilde{D}(0) = 0$ .*

*Proof.* The proof consists in reducing the problem to a matrix-functional equation, which is shown to have a unique solution. The solution is constructed in the process. Using the matrix-convolution notation introduced in Section 1.4, Equation (5.3) writes as:

$$\tilde{Z}(t) = (I - B^0(t))\tilde{Z}(0) + ((I - B(\cdot)) * \tilde{A})(t) .$$

Next, using (5.7), the short-hand notations  $C(t) = (I - B^0(t)) + (I - B(t))QP'$  and  $T^m = T(t)I$ , then rearranging, we obtain successively:

$$\begin{aligned} \tilde{Z}(t) &= C(t)\tilde{Z}(0) + ((I - B) * T^m)(t)\lambda - (I - B(\cdot)) * (QP' \tilde{Z}(\cdot)) \\ (I + (I - B(\cdot))(QP')) * \tilde{Z}(t) &= C(t)\tilde{Z}(0) + ((I - B) * T^m)(t)\lambda . \end{aligned} \quad (5.9)$$

We have introduced in Section 3.1 the function  $\mathcal{B}(t) = \sum_{n \geq 0} (BP')^{*n}(t)$ , which satisfies the identity  $\mathcal{B} = I + \mathcal{B} * (BP')$ . Using this identity and the fact that  $QP' = P'Q$ , it is easily seen that:

$$\mathcal{B} * (I + (I - B)(QP')) = Q . \quad (5.10)$$

Convolving Equation (5.9) by  $\mathcal{B}$ , we therefore have:

$$Q\tilde{Z}(t) = (\mathcal{B} * C)(t)\tilde{Z}(0) + (\mathcal{B} * (I - B) * T^m)(t)\lambda \quad (5.11)$$

$$\tilde{Z}(t) = (I - P')(\mathcal{B} * C)(t)\tilde{Z}(0) + (I - P')(\mathcal{B} * (I - B) * T^m)(t)\lambda . \quad (5.12)$$

Finally, multiplying by the vector  $e$  and using the fact that  $\dot{T}(t) = e.\tilde{Z}(t)$ , we obtain an integro-differential equation, of the renewal type, of which the function  $T$  is a solution:

$$\dot{T}(t) = e(I - P')(\mathcal{B} * C)(t)\tilde{Z}(0) + (K * T)(t) , \quad (5.13)$$

with  $K(t) = e(I - P')(\mathcal{B} * (I - B))(t)\lambda$ . We now show that this equation is of the same type, and has the same convolution kernel as the corresponding single-class equation in [19]. Indeed, these authors show (Lemma 4.2) that  $T$  is solution of the equation:

$$T(u) = H(u) + \rho(F_e * T)(u) \quad (5.14)$$

for some function  $H$  and where  $\rho$  is the load factor of the queue and  $F_e$  the excess life distribution associated to the service time. The unique locally bounded solution of this equation is given by  $T(u) = (H * U_e)(u)$ , where  $U_e(u) = \sum_{n=0}^{\infty} \rho^n (F_e)^{*n}(u)$ . Integrating Equation (5.13) with respect to  $t$  and setting

$$H(u) = \int_0^t e(I - P')(\mathcal{B} * C)(u)\tilde{Z}(0)du \quad (5.15)$$

and  $F_e(t) = \rho^{-1} \int_0^t K(u)du$ , we reduce it to Equation (5.14). In order to see that the convolution kernels indeed coincide, we compute the Laplace-Stieltjes transforms of the functions  $F_e$ , and that of the  $B_e^s$ , the excess lifetime distribution associated to  $B^s$ . This gives:

$$\hat{F}_e(\theta) = \frac{1}{\rho\theta} \hat{K}(\theta) = \frac{1}{\rho\theta} e(I - P')(I - \hat{B}(\theta)P')^{-1}(I - \hat{B}(\theta))\lambda ,$$

and this expression coincides with (3.16). The result of [19] therefore applies, and the function  $T$  exists and is unique. As a consequence, we obtain the existence and uniqueness of  $\tilde{Z}$  given by Equation (5.12) and therefore  $\tilde{A}$  and  $\tilde{D}$  by Equations (5.7)–(5.8).

In order to conclude, it remains to prove the continuity of the functions, and the increasingness of  $\tilde{A}$  and  $\tilde{D}$ . For this purpose, we proceed with rewriting some expressions. First, using the identity (5.10), we have:

$$\mathcal{B} * C = \mathcal{B} * (I + (I - B)(QP') - B^0) = Q - \mathcal{B} * B^0 .$$

Consequently, replacing this value in (5.11), and then in (5.8) we obtain:

$$\begin{aligned} Q(\tilde{Z} - \bar{Z}(0)) &= -(\mathcal{B} * B^0)(t)\bar{Z}(0) + (\mathcal{B} * (I - B) * T^m)(t)\lambda \\ \tilde{D}(t) &= \lambda T(t) - (\mathcal{B} * (I - B) * T^m)(t)\lambda + (\mathcal{B} * B^0)(t)\bar{Z}(0) \\ &= ((I - \mathcal{B} + \mathcal{B} * B) * T^m)(t)(I - P)^{-1}\alpha + (\mathcal{B} * B^0)(t)\bar{Z}(0) \\ &= (\mathcal{B} * B * T^m)(t)\alpha + (\mathcal{B} * B^0)(t)\bar{Z}(0) . \end{aligned} \quad (5.16)$$

Next, we have from (5.7)

$$\tilde{A}(t) = \lambda T(t) - P'(\mathcal{B} * (I - B) * T^m)(t)\lambda + P'(\mathcal{B} * B^0)(t)\bar{Z}(0) .$$

Introducing the function  $\mathcal{B}^\dagger(t) = \sum_{n=0}^{\infty} (P'B)^{*n}(t)$ , we observe that the following identities are satisfied:  $\mathcal{B}^\dagger = I + (P'B) * \mathcal{B}^\dagger = I + \mathcal{B}^\dagger * (P'B)$  and  $P'\mathcal{B} = \mathcal{B}^\dagger P'$ ,  $\mathcal{B} * B = B * \mathcal{B}^\dagger$ . Using them, we reduce  $\tilde{A}$  to:

$$\tilde{A}(t) = (\mathcal{B}^\dagger * T^m)(t)\alpha + (\mathcal{B}^\dagger * (P'B^0))(t)\bar{Z}(0) . \quad (5.17)$$

The function  $T(t) = (H * \mathcal{B})(t)$  is continuous and increasing because  $H(t)$  is positive, continuous and increasing, and  $\mathcal{B}(t)$  is increasing. The continuity of  $\tilde{A}(t)$  is a consequence of (5.17), given that the function  $B^0$  is continuous, by assumption. The continuity of  $\tilde{Z}$  results from Equation (5.3), still due to the continuity of  $B^0$ . The continuity of  $\tilde{D}$  results from Equation (5.8) or (5.16). This proves (i). Finally, since the functions  $\mathcal{B}$ ,  $\mathcal{B}^\dagger$ ,  $B^0$ ,  $B$  and  $T$  are increasing, so are the functions  $\tilde{A}$  and  $\tilde{D}$ . This proves (ii).  $\square$

### 5.1.3 The workload of the queue

In this paragraph, we are interested in the workload of the system, per class and globally. In particular, we solve the problem of the “missing mass”.

**Lemma 5.3.** *Given a data  $(\alpha, \nu, P)$  and  $\xi \in \mathcal{M}^K$ . Assume that  $(\bar{A}, \bar{D}, \bar{\mu})$  is a fluid solution such that  $\bar{\mu}(0) = \xi$ . Then the workload  $\bar{W}(t)$  defined by (4.2) satisfies equation (4.1)*

*Proof.* Denote by  $\bar{W}_k(t) = \int_0^{+\infty} \langle 1_{[x, +\infty)}, \bar{\mu}_k(t) \rangle dx$  the workload per class  $k$ . It suffices to prove that

$$\bar{W}_k(t) = \bar{W}_k(0) - \int_0^t \langle 1, \bar{\mu}_k(s) \rangle d\bar{S}(s) + \beta_k \bar{A}_k(t) \quad \text{for all } t < t_\rho(\xi) . \quad (5.18)$$

By this and (5.5), we have  $\bar{W}(t) = \sum_{k \in \mathcal{K}} \bar{W}_k(t) + e.\beta P'Q\bar{Z}(t)$ , and so (4.1) is easily verified. Now let us prove (5.18). Integrating equation (2.10) with respect to variable  $x$  on  $\mathbb{R}_+$  gives

$$\bar{W}_k(t) = \int_0^{+\infty} \langle 1_{[x + \bar{S}(t), \infty)}, \xi \rangle dx + \int_0^{+\infty} \int_0^t \langle 1_{[x + \bar{S}(t) - \bar{S}(s), \infty)}, \nu_k \rangle d\bar{A}_k(s) .$$

Next, adding and subtracting  $\bar{W}_k(0) + \beta_k \bar{A}_k(t)$ , interchanging the order of integration in the integrals, and performing a change of variable  $x = \bar{S}(u)$  in the first integral and  $y = \bar{S}(u)$  in the last integral, successively

give:

$$\begin{aligned}
\bar{W}_k(t) &= \bar{W}_k(0) + \beta_k \bar{A}_k(t) - \int_0^{+\infty} \langle 1_{[x, x+\bar{S}(t)], \xi_k} \rangle dx \\
&\quad - \int_0^{+\infty} \int_0^t \langle 1_{[x, x+\bar{S}(t)-\bar{S}(s)], \nu_k} \rangle d\bar{A}_k(s) dx \\
&= \bar{W}_k(0) + \beta_k \bar{A}_k(t) - \int_0^{\bar{S}(t)} \langle 1_{[x, +\infty), \xi_k} \rangle dx \\
&\quad - \int_0^t \int_{\bar{S}(s)}^{\bar{S}(t)} \langle 1_{[y-\bar{S}(s), +\infty), \nu_k} \rangle d\bar{A}_k(s) dy \\
&= \bar{W}_k(0) + \beta_k \bar{A}_k(t) - \int_0^t \langle 1_{[\bar{S}(u), +\infty), \xi_k} \rangle d\bar{S}(u) \\
&\quad - \int_0^t \int_s^t \langle 1_{[\bar{S}(u)-\bar{S}(s), +\infty), \nu_k} \rangle d\bar{A}_k(s) d\bar{S}(u) .
\end{aligned}$$

Interchanging the order of integration in last integral finally gives

$$\begin{aligned}
\bar{W}_k(t) &= \bar{W}_k(0) + \beta_k \bar{A}_k(t) - \\
&\quad \int_0^t \left( \langle 1_{[\bar{S}(u), +\infty), \xi_k} \rangle + \int_0^u \langle 1_{[\bar{S}(u)-\bar{S}(s), +\infty), \nu_k} \rangle d\bar{A}_k(s) \right) d\bar{S}(u) \\
&= \bar{W}_k(0) + \beta_k \bar{A}_k(t) - \int_0^t \langle 1, \bar{\mu}_k(u) \rangle d\bar{S}(u) .
\end{aligned}$$

□

As a corollary, we obtain a refinement on the time range of the solutions with non-zero initial state:

**Lemma 5.4.** *Assume that  $\xi \neq 0$ . The the value of  $t_\rho(\xi)$  is*

$$\begin{cases} t_\rho(\xi) = +\infty & \text{if } \rho \geq 1 \\ t_\rho(\xi) = \frac{e^{(\beta^0 + \beta Q P')} \bar{Z}(0)}{1 - \rho} & \text{if } \rho < 1 . \end{cases}$$

*Proof.* The value of  $W(0)$  in Equation (4.1) is readily seen to be  $e^{(\beta^0 + \beta P' Q)} \bar{Z}(0)$ . If  $[0, u]$  is an interval over which  $e.\bar{Z}(t)$  is not zero, then  $\bar{W}(t) > 0$  over this interval, and  $\bar{W}(t) = \bar{W}(0) + (\rho - 1)t$ . Conversely, if  $\bar{W}(t) = 0$ , then  $\bar{Z}(t) = 0$  and  $e.\bar{\mu}(t) = 0$ . If  $\rho < 1$ , then it follows that  $\bar{W}(t)$  is positive over the interval  $[0, \bar{W}(0)/(1 - \rho)]$ . If  $\rho > 1$ , this implies that  $\bar{W}(t) \geq \bar{W}(0)$  for all  $t$ , so that  $\bar{W}$  is never 0, and neither is  $e.\bar{\mu}(t)$ . □

#### 5.1.4 Continuity properties of the measure $\mu(\cdot)$

**Lemma 5.5.** *Let  $(\bar{A}, \bar{D}, \bar{\mu})$  be a fluid solution and let  $\xi \in \mathcal{M}^{c,p,K}$  be an initial state. We have the following properties:*

- i) *For each  $k \in \mathcal{K}$  and  $t \geq 0$ ,  $\bar{\mu}_k(t)$  has no atoms.*
- ii) *For each  $k \in \mathcal{K}$  and  $x \geq 0$ , the function  $t \rightarrow \langle 1_{[x, \infty), \bar{\mu}_k(t) \rangle$  belongs to  $\mathbf{C}([0, \infty))$ . In particular,  $\bar{\mu}_k(\cdot)$  is continuous on  $[0, \infty)$ .*

*Proof.* Let  $(\bar{A}, \bar{D}, \bar{\mu})$  be a fluid solution that satisfies the conditions of Definition 2.1. For  $i$ ), by (2.10) we have for each  $k \in \mathcal{K}$ ,  $t \geq 0$  and  $0 \leq x < y \leq +\infty$

$$\langle 1_{[x,y]}, \bar{\mu}_k(t) \rangle = \langle 1_{[x,y]}(\cdot - \bar{S}(t)), \xi_k \rangle + \int_0^t \langle 1_{[x,y]}(\cdot - (\bar{S}(t) - \bar{S}(s)), \nu_k) \rangle d\bar{A}_k(s).$$

By letting  $y \searrow x$  and using the bounded convergence theorem,

$$\langle 1_{\{x\}}, \bar{\mu}_k(t) \rangle = \langle 1_{\{x\}}(\cdot - \bar{S}(t)), \xi_k \rangle + \int_0^t \langle 1_{\{x\}}(\cdot - (\bar{S}(t) - \bar{S}(s)), \nu_k) \rangle d\bar{A}_k(s).$$

The first term of the second member is clearly 0 since  $\xi \in \mathcal{M}^{c,p,K}$ . For the second term we have

$$\int_0^t \langle 1_{\{x\}}(\cdot - (\bar{S}(t) - \bar{S}(s)), \nu_k) \rangle d\bar{A}_k(s) = \int_0^t \langle 1_{\{x\}}(\cdot - (\bar{S}(t) - \bar{S}(s)), \nu_k) \rangle \dot{\bar{A}}_k(s) ds.$$

Since  $\nu_k\{x + \bar{S}(t) - \bar{S}(s)\} = 0$  for all  $s \in [0, t]$  except for a countable set, the second term of the second member of the above equality is also 0, which proves  $i$ ).

We now turn to  $ii$ ). The proof focuses on the case  $t \leq t^*$ , where  $t^*$  is as before the time range of the solution. Indeed, in the subcritical case,  $\bar{\mu}_k(t) = 0$  for all  $k \in \mathcal{K}$  and  $t \geq t^*$  by definition, so that  $ii$ ) holds for  $t \geq t^*$ .

Let  $k \in \mathcal{K}$  and  $x \geq 0$ , to show that the function  $t \mapsto \langle 1_{[x,\infty)}, \bar{\mu}_k(t) \rangle$  is continuous on  $[0, t^*)$ , it suffices to show that each term on the right side of (2.10) is continuous. Since  $t \mapsto \bar{S}(t)$  is continuous on  $[0, t^*)$  and  $\xi_k$  has no atoms, then  $t \mapsto \langle 1_{[x,\infty)}(\cdot - \bar{S}(t)), \xi_k \rangle$  is continuous on  $[0, t^*)$ . For the second term, since  $t \mapsto \bar{A}_k(t)$  is continuous and nondecreasing and  $u \mapsto \langle 1_{[x,\infty)}(\cdot - (\bar{S}(u) - \bar{S}(s))), \nu_k \rangle$  is bounded, we have

$$\int_0^u \langle 1_{[x,\infty)}(\cdot - (\bar{S}(u) - \bar{S}(s)), \nu_k) \rangle d\bar{A}_k(s) = \int_0^u \langle 1_{[x,\infty)}(\cdot - (\bar{S}(u) - \bar{S}(s)), \nu_k) \rangle \dot{\bar{A}}_k(s) ds.$$

By the bounded convergence theorem, the function of  $u$  in the right-hand side of the above equation is continuous. In order to show that  $t \mapsto \bar{\mu}_k(t)$  is continuous on  $[0, t^*)$ , we must show that  $t \mapsto \langle g, \bar{\mu}_k(t) \rangle$  is continuous on  $[0, t^*)$  for each  $g \in C_b(\mathbb{R}_+)$ . Fix  $x \in \mathbb{R}_+$  and  $g \in C_b(\mathbb{R}_+)$ , denote  $G_k(t, x)$  the function of  $t$  given by

$$G_k(t, x) = \begin{cases} \frac{\langle g, \bar{\mu}_k(t) \rangle}{\bar{Z}_k(t)} & \text{if } \bar{Z}_k(t) \neq 0 \\ 0 & \text{if } \bar{Z}_k(t) = 0 \end{cases}$$

and prove that this function is well defined and continuous on  $[0, t^*)$ . Since the functions  $\bar{Z}_k(\cdot)$  and  $\langle 1_{[x,\infty)}, \bar{\mu}_k(\cdot) \rangle$  belong to  $C([0, t^*))$ ,  $t \mapsto \langle 1_{[x,\infty)}, \bar{\mu}_k(t) \rangle / \bar{Z}_k(t)$  is well defined and continuous on the set  $\{t < t^* : \bar{Z}_k(t) \neq 0\}$ . Then for each  $t < t^*$  such that  $\bar{Z}_k(t) \neq 0$ , we have

$$\lim_{u \rightarrow t} \langle 1_{[x,\infty)}, \frac{\bar{\mu}_k(u)}{\bar{Z}_k(u)} \rangle = \langle 1_{[x,\infty)}, \frac{\bar{\mu}_k(t)}{\bar{Z}_k(t)} \rangle.$$

Since for each  $t \in [0, t^*)$  such that  $\bar{Z}_k(t) \neq 0$ ,  $\bar{\mu}_k(t)(\cdot) / \bar{Z}_k(t)$  is a probability measure that have no atoms, by [3, Theorem 2.2], the above limit is satisfied for all  $g \in C_b(\mathbb{R}_+)$ . Hence, the function  $t \mapsto G_k(t, x)$  is continuous on  $\{t < t^* : \bar{Z}_k(t) \neq 0\}$ . Finally let  $t \in [0, t^*)$  such that  $\bar{Z}_k(t) = 0$ , since  $\langle g, \bar{\mu}_k(t) \rangle \leq \|g\|_\infty \bar{Z}_k(t)$  then  $\langle g, \bar{\mu}_k(t) \rangle = 0$  this proves the continuity of  $G_k(t, x)$  on  $[0, t^*)$ . This completes the proof.  $\square$

### 5.1.5 Relationship with the single-class case

We have now gathered enough information to state the correspondence results with the single-class case. We refer to Section 3.1 for the definitions of symbols, and [19] for the terminology associated to the single-class queue.

**Lemma 5.6.** Consider a processor sharing queue with arrival rate  $\alpha_e$  and service time distribution  $B^s$ , associated to the Borel measure  $\nu^s$ . Assume that the data  $(\alpha_e, \nu^s)$  is supercritical, and let  $T^{(s)}$  be the function corresponding to the initial measure  $\xi$  such that

$$\langle 1_{(x, +\infty)}, \xi \rangle = e(I - P')(B * C)(x)\bar{Z}(0) ,$$

where  $\mathcal{B}(t) = \sum_{n \geq 0} (BP')^{*n}(t)$  and  $C(t) = (I - B^0(t)) + (I - B(t))QP'$ . Then the function  $T^{(s)}$  is equal to  $T$ . It is given by:

$$T(u) = (H * U_e)(u)$$

where  $U_e(u) = \sum_{n=0}^{\infty} \rho^n (B_e^s)^{*n}(u)$ ,  $H(x) = \int_0^x \langle 1_{(y, +\infty)}, \xi \rangle dy$ , and  $B_e^s$  is the excess lifetime distribution associated to  $B^s$ . Its Laplace-Stieltjes transform is given by

$$\hat{T}(\theta) = \frac{\hat{H}(\theta)}{1 - \psi(\theta)} , \quad (5.19)$$

where:

$$\hat{H}(\theta) = \theta^{-1} e \left( I - (I - P')(I - \hat{B}(\theta)P')^{-1} \hat{B}^0(\theta) \right) \bar{Z}(0) \quad (5.20)$$

$$\psi(\theta) = \theta^{-1} e(I - P')(I - \hat{B}(\theta)P')^{-1}(I - \hat{B}(\theta))\lambda = \rho \hat{B}_e^s(\theta) . \quad (5.21)$$

*Proof.* From Equation (5.14), it is easy to see that the Laplace-Stieltjes transform of  $T$  must satisfy:  $\hat{T}(\theta) = \hat{H}(\theta) + \rho \hat{F}_e(\theta) \hat{T}(\theta)$ , which readily gives (5.19). The expression for  $\hat{H}(\theta)$  uses the simplification obtained above for  $\mathcal{B} * C$ , and the fact that if  $H(t) = \int_0^t h(u)du$ , then  $\hat{H}(\theta) = \theta^{-1} \hat{h}(\theta)$ .  $\square$

### 5.1.6 Proof of Proposition 5.1

Fix a data  $(\alpha, \nu, P)$ , an initial measure  $\xi \in \mathcal{M}^{c,p,K}$ . From condition (2.10) of Definition 2.1, for each  $k \in \mathcal{K}$ ,  $x \in \mathbb{R}_+$  and  $t \geq 0$ , the function  $\langle 1_{[x, +\infty)}, \mu_k(t) \rangle$  is uniquely determined by  $\bar{S}(t)$ ,  $\bar{A}(t)$ ,  $\bar{D}(t)$ ,  $\bar{Z}(t)$ ,  $(\alpha, \nu, P)$  and  $\xi$ . Because the sets  $\{[x, +\infty) \mid x \in \mathbb{R}_+\}$  generate the Borel  $\sigma$ -algebra on  $\mathbb{R}_+$ , this uniquely determines  $\mu_k(t)$  for each  $t \geq 0$ . It is therefore sufficient to prove the existence and uniqueness of the functions  $\bar{S}(t)$ ,  $\bar{A}(t)$ ,  $\bar{D}(t)$ ,  $\bar{Z}(t)$ .

The uniqueness follows from Lemma 5.2: to each Fluid Solution corresponds a system of functions  $\bar{A}$ ,  $\bar{D}$ ,  $\bar{Z}$ . Since these functions are uniquely determined, there cannot be more than one Fluid Solution.

On the other hand, Lemma 5.2, guarantees the existence of the functions  $\bar{A}$ ,  $\bar{D}$ ,  $\bar{Z}$  for every data. In the proof of this lemma, it is shown that  $T$  is also uniquely defined, and is continuous, strictly increasing, and maps  $[0, +\infty)$  to  $[0, \bar{t})$ , where  $\bar{t} = \lim_{t \rightarrow +\infty} T(t)$ . Using Lemmas 5.4 and 5.8 (the proof of which is given later in the paper, but depends only on the already proved Lemma 5.6), we have  $\bar{t} = t^*$ . The inverse function of  $T(\cdot)$  is well defined for  $t < \bar{t}$ . Let  $\bar{S}(t) = T^{-1}(t)$ . Let then define  $\bar{A}(t) = \bar{A}(\bar{S}(t))$ ,  $\bar{D}(t) = \bar{D}(\bar{S}(t))$  and  $\bar{Z}(t) = \bar{Z}(\bar{S}(t))$  for all  $t < \bar{t}$ . Since  $\bar{A}$ ,  $\bar{D}$  and  $\bar{S}(\cdot)$  are (componentwise) continuous and increasing function, so are  $\bar{A}$  and  $\bar{D}$ . These functions solve the system (2.8), (2.9), (2.11), (2.14). Therefore, the existence of a solution is proved over the interval  $[0, \bar{t})$ . If  $\rho \geq 1$ , this interval is  $[0, +\infty)$  and the proof of existence is complete.

In the case  $\rho < 1$ , it is sufficient to prolong the solution on  $[\bar{t}, +\infty)$  with  $\bar{A}(t) = \bar{D}(t) = \lambda t$  and  $\bar{Z}(t) = 0$ . Equation (4.1) is proved from Lemma 5.3.

## 5.2 Existence of a solution for $\xi = 0$

In this paragraph, we prove the existence part of Theorem 4.2, since we need this result in our analysis. We state it as a separate result for easier reference.

**Proposition 5.2.** *The functions  $A^0$ ,  $D^0$  and the measure  $\mu^0$  defined in (4.7) are a Fluid Solution when  $\bar{\mu}(0) = 0$ .*

*Proof.* The proof is by inspection. Let us check that  $A^0$ ,  $D^0$  and  $\mu^0$  indeed solve the equations (2.8)–(2.11). We have, using the notation  $\widehat{B} = \widehat{B}(\theta_0)$ ,

$$\begin{aligned} \alpha t + P'D^0(t) &= (I - P'\widehat{B}(\theta_0))(I - \widehat{B}(\theta_0))^{-1}mt + P'(I - \widehat{B}(\theta_0))^{-1}\widehat{B}(\theta_0)mt \\ &= (I - P'\widehat{B}(\theta_0) + P'\widehat{B}(\theta_0))(I - \widehat{B}(\theta_0))^{-1}mt \\ &= (I - \widehat{B}(\theta_0))^{-1}mt \\ &= A^0(t), \end{aligned}$$

which is equation (2.8). Also,

$$\begin{aligned} A^0(t) - D^0(t) &= (I - \widehat{B}(\theta_0))^{-1}mt - (I - \widehat{B}(\theta_0))^{-1}Bmt \\ &= (I - \widehat{B}(\theta_0))^{-1}(I - \widehat{B}(\theta_0))mt \\ &= mt. \end{aligned}$$

Since from (4.6), we have  $\langle 1, \bar{s}_k(t) \rangle = m_k t$ , this implies equation (2.9). Finally let us check (2.10). First, we have

$$\begin{aligned} \langle 1_{[x, +\infty)}, s_k \rangle &= \int_x^{+\infty} p_k(y) dy \\ &= \frac{\theta_0}{1 - \widehat{B}_k} m_k \int_x^{+\infty} \int_y^{+\infty} \exp(\theta_0(y - z)) d\nu_k(z) dy \\ &= \frac{\theta_0}{1 - \widehat{B}_k} m_k \int_x^{+\infty} \int_x^z \exp(\theta_0(y - z)) dy d\nu_k(z) \\ &= \frac{m_k}{1 - \widehat{B}_k} \int_x^{+\infty} (1 - \exp(\theta_0(x - z))) d\nu_k(z). \end{aligned}$$

On the other hand, Since  $\bar{S}(t) - \bar{S}(s) = \log(t/s)/\theta_0$  for all  $0 < s < t$ ,

$$\begin{aligned} &\int_0^t \langle 1_{[x, \infty[}(\cdot - (\bar{S}(t) - \bar{S}(s))), \nu_k \rangle dA_k(s) \\ &= \frac{m_k}{1 - \widehat{B}_k} \int_0^t \langle 1_{[x + \log(t/s)/\theta_0, \infty[}, \nu_k \rangle ds \\ &= t\theta_0 \frac{m_k}{1 - \widehat{B}_k} \int_0^{+\infty} \langle 1_{[x+u, \infty[}, \nu_k \rangle \exp(-\theta_0 u) du \\ &= t\theta_0 \frac{m_k}{1 - \widehat{B}_k} \int_x^{+\infty} \int_v^{+\infty} \exp(-\theta_0(v - x)) d\nu_k(y) dv \\ &= t \frac{m_k}{1 - \widehat{B}_k} \int_x^{+\infty} (1 - \exp(-\theta_0(y - x))) d\nu_k(y). \end{aligned}$$

By identification, we have (2.10). □

### 5.3 Convergence to initial Fluid solution for critical and supercritical data

We first prove a continuity property of the fluid solutions at the zero initial condition. For this we have adopted the convention that,

$$D^0(t) = A^0(t) = \lambda t, \quad \theta_0 = 0 \quad \text{and} \quad \mu^0(t) = 0 \quad \text{if} \quad \rho \leq 1.$$

**Theorem 5.1.** Let  $(\alpha, \nu, P)$  be a data of MPS. Let  $\{\xi^n, n \geq 0\} \subseteq \mathcal{M}^{c,p,K}$  such that  $\xi^n \implies 0$ . For each  $n \geq 0$  let  $(A^n, D^n, \mu^n)$  be a fluid solution with initial state  $\xi^n$ . Then for all  $T > 0$ ,

$$\lim_{n \rightarrow +\infty} \sup_{0 \leq t \leq T} \|D^n(t) - D^0(t)\| = 0 \quad (5.22)$$

$$\lim_{n \rightarrow +\infty} \sup_{0 \leq t \leq T} \|A^n(t) - A^0(t)\| = 0 \quad (5.23)$$

and for all  $k \in \mathcal{K}$ ,  $g \in C_b(\mathbb{R}_+)$  and  $T > 0$ ,

$$\lim_{n \rightarrow +\infty} \sup_{0 \leq t \leq T} |\langle g, \mu_k^n(t) \rangle - \langle g, \mu_k^0(t) \rangle| = 0. \quad (5.24)$$

The proof uses the next lemma.

**Lemma 5.7.** Let  $\{\xi^n, n \geq 0\} \subset \mathcal{M}^{c,p,K}$  such that  $\xi^n \implies 0$ . Then

$$\lim_{n \rightarrow +\infty} \sup_{0 \leq t \leq T} |e \cdot Z^n(t) - \theta_0 t| = 0 \quad \text{for all } T > 0. \quad (5.25)$$

If moreover  $\rho > 1$ , then for all  $0 < T_1 < T_2 < +\infty$ ,

$$\lim_{n \rightarrow +\infty} \sup_{T_1 \leq s \leq t \leq T_2} \left| S^n(t) - S^n(s) - \frac{1}{\theta_0} \log\left(\frac{t}{s}\right) \right| = 0. \quad (5.26)$$

*Proof.* Since the global number of customers in the system evolves the same as a single class processor sharing queue, the proof of this lemma is similar of that of Theorem 4.3 in [19]  $\square$

*Proof of Theorem 5.1.* We begin by proving the theorem for  $\rho \leq 1$ . Since  $g$  is a bounded function,

$$\sup_{0 \leq t \leq T} \langle g, \mu_k^n(t) \rangle \leq \|g\|_\infty \sup_{0 \leq t \leq T} \langle 1, \mu_k^n(t) \rangle.$$

By (5.25) we have  $\lim_{n \rightarrow +\infty} \sup_{0 \leq t \leq T} \langle 1, \mu_k^n(t) \rangle = 0$  because in this case  $\theta_0 = 0$ . By (2.8) and (2.9)  $\lim_{n \rightarrow +\infty} A^n(t) = \lim_{n \rightarrow +\infty} D^n(t) = \lambda t$ . This implies (5.22)–(5.24). Let us proceed with the case  $\rho > 1$ . Consider Equation (5.16) with  $\bar{D}^n(u)$  and performing a change of variable  $u = S^n(t)$  one obtains

$$D^n(t) = \int_0^t (\mathcal{B} * B)(S^n(t) - S^n(s)) ds + (\mathcal{B} * \Gamma^n)(S^n(t))\alpha, \quad (5.27)$$

where  $\Gamma^n(x) = \text{diag}\{\langle 1_{[0,x)}, \xi_k^n \rangle\} = B^{0,n}(x)Z^n(0)$ . The analogue equation of (5.27) for  $D^0(\cdot)$  is

$$D^0(t) = \int_0^t (\mathcal{B} * B)\left(\frac{1}{\theta_0} \log\left(\frac{t}{s}\right)\right)\alpha ds.$$

Since  $t \rightarrow \mathcal{B} * B(t)$  is r.c.l.l. and by Lemma 5.7  $S^n(t) - S^n(s) \rightarrow \frac{1}{\theta_0} \log\left(\frac{t}{s}\right)$  for all  $t \geq s > 0$ , we have

$$\lim_{n \rightarrow \infty} (\mathcal{B} * B)(S^n(t) - S^n(s)) = (\mathcal{B} * B)\left(\frac{1}{\theta_0} \log\left(\frac{t}{s}\right)\right)$$

for all  $0 < s \leq t$  except  $\{(s, t) : 0 < s \leq t \text{ and } \frac{1}{\theta_0} \log\left(\frac{t}{s}\right) \text{ is an atom of } \mathcal{B} * \nu\}$ . This set has the Lebesgue measure zero. Then

$$\lim_{n \rightarrow \infty} \int_0^t (\mathcal{B} * B)(S^n(t) - S^n(s))\alpha ds = D^0(t).$$

Since,  $\|(\mathcal{B} * \Gamma^n(\cdot))(S^n(t))\| \leq \|Z^n(0)\| \|Q\|$  and by assumption  $\|Z^n(0)\| \rightarrow 0$ , the second term in the right-hand side of (5.27) tends to 0. Because the limit  $D^0(t)$  is continuous, the limit (5.22) holds. The limit (5.23) follows by equation (2.8).

We turn now to the proof of (5.24). We have defined the dynamics for measures (2.10) for indicator functions  $1_{[x, +\infty)}$ . However, by the monotone class theorem for functions [10, Chapter 5, Theorem 1.5 p. 280], and using an argument similar as that of [19] (*cf.* the proof of Lemma 4.14), we have for all  $g \in C_b(\mathbb{R}_+)$ ,  $k \in \mathcal{K}$  and  $t \geq 0$ :

$$\langle g, \mu_k^n(t) \rangle = \langle g(\cdot - S^n(t)), \xi_k^n \rangle + \int_0^t \langle g(\cdot - (S^n(t) - S^n(s)), \nu_k \rangle dA_k^n(s) . \quad (5.28)$$

On the other hand,

$$\langle g, \mu_k^0(t) \rangle = \int_0^t \langle g(\cdot - \frac{1}{\theta_0} \log(\frac{t}{s})), \nu_k \rangle dA_k^0(s) . \quad (5.29)$$

So from (5.28) and (5.29),

$$\begin{aligned} & \sup_{0 \leq t \leq T} |\langle g, \mu_k^n(t) \rangle - \langle g, \mu_k^0(t) \rangle| \\ & \leq \sup_{0 \leq t \leq T} \langle g(\cdot - S^n(t)), \xi_k^n \rangle \\ & \quad + \sup_{0 \leq t \leq T} \left| \int_0^t \langle g(\cdot - (S^n(t) - S^n(s))), \nu_k \rangle dA_k^n(s) - \int_0^t \langle g(\cdot - \frac{1}{\theta_0} \log(\frac{t}{s})), \nu_k \rangle dA_k^0(s) \right| . \end{aligned} \quad (5.30)$$

The first term on the right-hand side of (5.30) tends to zero because,

$$\sup_{0 \leq t \leq T} |\langle g(\cdot - S^n(t)), \xi_k^n \rangle| \leq \|g\| \langle 1, \xi_k^n \rangle$$

and  $\lim_{n \rightarrow +\infty} \langle 1, \xi_k^n \rangle = 0$ . For the second term we use Lemma B.1, *iv*). The conditions of the lemma are satisfied with:

- the sequence of increasing functions  $A_k^n$ , which are such that  $A_k^n(0) = 0$ , and which converge uniformly to the Lipschitz-continuous function  $A_k^0(t) = m_k t$  on every compact, according to Theorem 5.1, (5.23);
- the sequence of functions  $f^n(t, s) = S^n(t) - S^n(s)$ , which converges uniformly on every compact not containing 0, to the function  $f(t, s) = \log(t/s)/\theta_0$ , according to Lemma 5.7, (5.26), and such that  $s \mapsto f(t, s)$  is continuously decreasing to  $f(t, t) = 0$ ;
- the probability measure  $\nu_k$ .

The second term of (5.30) goes therefore also to 0, and the convergence (5.24) follows.  $\square$

**Proposition 5.3** (Translation property of Fluid solution). *Let  $\tau > 0$  and let  $(\bar{A}, \bar{D}, \bar{\mu})$  be a fluid solution with initial state  $\xi \in \mathcal{M}^K$ . Define  $(A^\tau, D^\tau, \mu^\tau)$  by*

$$A^\tau = \bar{A}(\tau + t) - \bar{A}(\tau), \quad D^\tau(t) = \bar{D}(\tau + t) - \bar{D}(\tau), \quad \mu^\tau(t) = \bar{\mu}(\tau + t) \quad \text{for all } t \geq 0 .$$

*Then  $(A^\tau, D^\tau, \mu^\tau)$  is a fluid solution with initial state  $\bar{\mu}(\tau)$ . The process  $(S^\tau(t), t \geq 0)$  associated to a translated fluid solution is given by*

$$S^\tau(t) = \bar{S}(t + \tau) - \bar{S}(\tau) = \int_0^t \varphi(\langle 1, e \cdot \mu^\tau(s) \rangle) ds .$$

*Proof.* It is not difficult to check the equations (2.8) and (2.9). For equation (2.10), we have

$$\begin{aligned} \langle 1_{[x,\infty)}, \mu_k^\tau(t) \rangle - \langle 1_{[x,\infty)}(\cdot - (S^\tau(t))), \mu_k(\tau) \rangle &= \int_\tau^{t+\tau} \langle 1_{[x,\infty)}(\cdot - (\bar{S}(t+\tau) - \bar{S}(s))), \nu_k \rangle dA_k(s) \\ &= \int_0^t \langle 1_{[x,\infty)}(\cdot - (S^\tau(t) - S^\tau(s))), \nu_k \rangle dA_k^\tau(s). \end{aligned}$$

This concludes the proof.  $\square$

### 5.3.1 Proof of Theorems 4.1 and 4.2

The proof of Theorem 4.2 has two parts. Proposition 5.2 provides the existence part. For the uniqueness, let  $(\alpha, \nu, P)$  be a supercritical data ( $\rho > 1$ ) and let  $(A^0, D^0, \mu^0)$  be the fluid solution for zero initial state given by (4.7).

Let  $(\bar{A}, \bar{D}, \bar{\mu})$  be a fluid solution such that  $\bar{\mu}(0) = 0$ . We must show that  $(\bar{A}, \bar{D}, \bar{\mu}) = (A^0, D^0, \mu^0)$ . Let  $\{\tau_n, n \in \mathbb{N}\} \subset (0, +\infty)$  be a sequence such that  $\tau_n \rightarrow 0$  as  $n \rightarrow +\infty$ . From Proposition 5.3, for each  $n \in \mathbb{N}$ ,  $(A^{\tau_n}, D^{\tau_n}, \mu^{\tau_n})$  is a Fluid Solution with initial state  $\bar{\mu}(\tau_n)$ . By Lemma 5.3, we have  $\bar{W}(\tau_n) > 0$ , so  $e \cdot \bar{\mu}(\tau_n) > 0$  for all  $n \in \mathbb{N}$ . Using the continuity of  $t \rightarrow \bar{\mu}(t)$ ,  $e \cdot \bar{\mu}(\tau_n) \Rightarrow 0$  as  $n \rightarrow +\infty$ , which implies that  $\bar{\mu}(\tau_n) \Rightarrow 0$ . Consequently by Theorem 5.1 we have

$$(A^{\tau_n}, D^{\tau_n}, \mu^{\tau_n}) \longmapsto (A^0, D^0, \mu^0) \tag{5.31}$$

uniformly on compacts as  $n \mapsto +\infty$ , in the sense of (5.22)–(5.24). On the other hand, by the continuity of the fluid solution, we have also:  $(A^{\tau_n}(t), D^{\tau_n}(t), \mu^{\tau_n}(t)) \rightarrow (\bar{A}(t), \bar{D}(t), \bar{\mu}(t))$  as  $n \rightarrow +\infty$  for all  $t \geq 0$ . By uniqueness of limits we have  $(\bar{A}(t), \bar{D}(t), \bar{\mu}(t)) = (A^0(t), D^0(t), \mu^0(t))$  for all  $t \geq 0$ . This concludes the proof of Theorem 4.2.

For Theorem 4.1, we have two cases. When  $\xi \neq 0$ , Proposition 5.1 provides the existence and uniqueness.

When  $\xi = 0$ , we have three sub-cases. For a supercritical data, the existence and uniqueness is provided by Theorem 4.2. For a subcritical or critical data, the functions  $\bar{A}(t) = \bar{D}(t) = \lambda t$  and the measure  $\bar{\mu}(t) = 0$  provide a Fluid solution. According to Lemma 5.3,  $\bar{W}(t) = \langle \chi, e \cdot \bar{\mu}(t) \rangle + e \cdot \beta P' Q \bar{Z}(t)$  is identically 0. As a consequence,  $\bar{\mu}(t) = 0$  is the unique possible measure for a Fluid Solution. This uniquely determines  $\bar{A}(t)$  and  $\bar{D}(t)$  with Equations (2.8) and (2.9). The solution is therefore unique. This concludes the proof of Theorem 4.1.

## 5.4 A characterization of the measure $\mu_k$

The following proposition gives a characterization of fluid solutions according to Definition 2.1. It features a characterization of the process  $\bar{\mu}(t)$  (Equation (5.32)) of central importance for the forthcoming proof of convergence. Let  $\mathcal{C} = \{g \in C_b^1(\mathbb{R}_+) : g(0) = \dot{g}(0) = 0\}$ .

**Proposition 5.4.** *Consider the data  $(\alpha, \nu, P)$ . Let  $\bar{A}, \bar{D} : \mathbb{R}_+ \rightarrow \mathbb{R}_+^K$  be two vectors of continuous and increasing functions and,  $\bar{\mu} = (\bar{\mu}_1, \dots, \bar{\mu}_K) : \mathbb{R}_+ \rightarrow \mathcal{M}^K$  be a measure-valued vector of functions, such that equations (2.8), (2.9) and (2.11) are satisfied. The triple  $(\bar{A}(t), \bar{D}(t), \bar{\mu}(t))$  is a fluid solution if and only if the vector of measure  $\bar{\mu}$  satisfies the following conditions*

- (i)  $\bar{\mu}(\cdot)$  is continuous
- (ii)  $\bar{\mu}_k(t)\{0\} = 0$ , for all  $k \in \mathcal{K}$  and  $t \geq 0$

(iii) For all  $g \in \mathcal{C}$ ,  $k \in \mathcal{K}$ ,  $\bar{\mu}_k(\cdot)$  satisfies

$$\langle g, \bar{\mu}_k(t) \rangle = \langle g, \bar{\mu}_k(0) \rangle - \int_0^t \langle \dot{g}, \bar{\mu}_k(s) \rangle d\bar{S}(s) + \bar{A}_k(t) \langle g, \nu_k \rangle \quad (5.32)$$

for all  $t < t_\rho(\xi)$ , and  $\bar{A}(t) = \bar{D}(t) = \lambda t$ ,  $\bar{\mu}(t) = 0$  for all  $t \geq t_\rho(\xi)$ ,  $t_\rho(\xi)$  being defined by (2.12).

*Proof.* Let  $\bar{A}$ ,  $\bar{D}$  be nonnegative real-valued vectors of continuous and increasing functions and,  $\bar{\mu}$  a nonnegative measure-valued vector of functions, such that equations (2.8), (2.9) and (2.11) are satisfied and let  $\xi \in \mathcal{M}^K$ .

**Sufficient condition.** Suppose that conditions (i), (ii) and (iii) of proposition 5.4 such that  $\bar{\mu}(0) = \xi$  are verified, and prove that equation (2.10) of Definition 2.1 is satisfied. Let  $\bar{W}(\cdot)$  be the function defined as in (4.2) with  $\bar{\mu}_k$  is solution of Equation (5.32) and  $\bar{Z}_k(t) = \langle 1, \bar{\mu}_k(t) \rangle$  for each  $k \in \mathcal{K}$ . Substituting a sequence of nonnegative functions  $(g_n, n \geq 0) \subset \mathcal{C}$  into (5.32), where  $g_n \nearrow \chi$  and  $g'_n \nearrow 1_{(0,\infty)}$  as  $n \rightarrow \infty$ , one obtains by monotone convergence,

$$\begin{aligned} \langle \chi, e.\bar{\mu}(t) \rangle &= \langle \chi, e.\bar{\mu}(0) \rangle - \int_0^t \langle 1, e.\bar{\mu}(s) \rangle d\bar{S}(s) + e.\beta \bar{A}(t) \\ &= \langle \chi, e.\bar{\mu}(0) \rangle + (1 - \rho)t + e\beta QP' \bar{Z}(0) - e\beta QP' \bar{Z}(t) \end{aligned}$$

for all  $t < t_\rho(\xi)$ . Using this and Equation (5.5), the function  $\bar{W}(\cdot)$  satisfies Equation (4.1). We terminate the proof of (2.10) in three separate cases:

i) Case  $\langle 1, e.\xi \rangle > 0$ . For each  $k \in \mathcal{K}$ , In similar manner of proof of Lemma 4.1 in [15], an alternative equation for time dependent function can be derived from equation (5.32),

$$\begin{aligned} \langle f(t, \cdot), \bar{\mu}_k(t) \rangle &= \langle f(0, \cdot), \bar{\mu}_k(0) \rangle + \int_0^t \left\langle \frac{\partial f}{\partial s}(s, \cdot), \bar{\mu}_k(s) \right\rangle \\ &\quad - \int_0^t \frac{\left\langle \frac{\partial f}{\partial x}(s, \cdot), \bar{\mu}_k(s) \right\rangle}{\langle 1, e.\bar{\mu}(s) \rangle} ds + \int_0^t \langle f(s, \cdot), \nu_k \rangle d\bar{A}_k(s) \end{aligned}$$

for all  $f \in C_b^1([0, \infty) \times \mathbb{R}_+)$  such that  $f(\cdot, 0) \equiv 0$  and  $\frac{\partial f}{\partial x}(\cdot, 0) \equiv 0$ , and for all  $t < t^*$ . Next, by Lemma 4.3 of [15], the above equation implies (2.10).

ii) Case  $\langle 1, e.\xi \rangle = 0$  and  $\rho \leq 1$ . Since  $\bar{W}(t) = 0$ , we have  $\bar{\mu}(t) = 0$  and  $\bar{A}(t) = \bar{D}(t) = \lambda t$ . This is the unique fluid solution.

iii) Case  $\langle 1, e.\xi \rangle = 0$  and  $\rho > 1$ . By the translation property, for each  $\eta > 0$ , the triple  $(\bar{A}^\eta(t), \bar{D}^\eta(t), \bar{\mu}^\eta(t))$  defined by

$$\bar{A}^\eta(t) := \bar{A}(t + \eta) - \bar{A}(\eta), \quad \bar{D}^\eta(t) := \bar{D}(t + \eta) - \bar{D}(\eta) \quad \text{and} \quad \bar{\mu}^\eta(t) := \bar{\mu}(t + \eta)$$

for all  $t \geq 0$ , satisfies equations (2.8), (2.9) and (5.32) with  $\bar{S}^\eta(t) = \bar{S}(t + \eta) - \bar{S}(\eta)$ . We have  $\langle 1, e.\bar{\mu}^\eta(0) \rangle > 0$ . Then by the first case the equation (2.10) is satisfied for all  $\eta > 0$ . Letting  $\eta \rightarrow 0$ , by Theorem 5.1,  $(\bar{A}^\eta, \bar{D}^\eta, \bar{\mu}^\eta)$  converges uniformly on compacts to the unique solution of fluid model  $(A^0, D^0, \mu^0)$  starting from initial state zero, which satisfies Equation (2.10).

**Necessary condition:** Conditions (i) and (ii) of Lemma 5.5 imply conditions (i) and (ii) of Proposition 5.4. It suffices to prove that Equation (5.32) is satisfied. Using Equation (2.10), the monotone class theorem for functions ([10, Chapter 4, Theorem 1.5, p. 280]) and by a similar argument as that of ([15], Equation 4.37) and ([19], the proof of lemma 4.14) we have for each  $k \in \mathcal{K}$ ,  $g \in C_b(\mathbb{R}_+)$  and  $0 \leq t < t^*$

$$\langle g, \bar{\mu}_k(t) \rangle = \langle g(\cdot - \bar{S}(t)), \xi_k \rangle + \int_0^t \langle g(\cdot - (\bar{S}(t) - \bar{S}(s))), \nu_k(t) \rangle d\bar{A}_k(s) \quad (5.33)$$

$$\langle g, \mu_k^0(t) \rangle = \int_0^t \langle g(\cdot - (-\frac{1}{\theta_0} \log(\frac{t}{s}))), \nu_k(t) \rangle dA_k^0(s). \quad (5.34)$$

Recall the convention that any function  $g$  that is defined on  $\mathbb{R}_+$  is extended to be identically equal to zero on  $(-\infty, 0)$  so that for all  $a > 0$ ,  $g(\cdot - a)$  is well defined on  $\mathbb{R}_+$ . Fix  $g \in \mathcal{C}$ , then  $g \in C_b^1(\mathbb{R})$  with  $g(x) = 0$  and  $\dot{g}(x) = 0$  for  $x \leq 0$ . In particular Equation (5.33) holds for  $g$  and  $\dot{g}$ . We will proceed as in the proof of Lemma 4.8 in [15], by computing the derivative of each term of (5.33), which yields

$$\frac{d}{dt} \langle g, \bar{\mu}_k(t) \rangle = - \frac{\langle \dot{g}, \bar{\mu}_k(t) \rangle}{e \cdot \bar{Z}(t)} + \langle g, \nu_k \rangle \dot{A}_k(t) \quad (5.35)$$

for all  $k \in \mathcal{K}$ ,  $g \in \mathcal{C}$  and  $0 \leq t < t^*$  such that  $\dot{A}_k(t)$  exists. Similarly, from Equation (5.34) we have

$$\frac{d}{dt} \langle g, \mu_k^0(t) \rangle = - \frac{\langle \dot{g}, \mu_k^0(t) \rangle}{\theta_0 t} + \langle g, \nu_k \rangle \dot{A}_k^0(t) \quad (5.36)$$

for all  $k \in \mathcal{K}$ ,  $g \in \mathcal{C}$  and  $t \geq 0$ . Then Equation (5.32) follows by integrating (5.35) and (5.36).  $\square$

## 5.5 Asymptotic results

### 5.5.1 Asymptotics for the response time

The following asymptotic results concern the scalar function  $T(\cdot)$ . Since we know that this function is the same as in the single-class model, the results of [19] or [20] apply. We state and prove the results again in our context for the sake of completeness.

**Lemma 5.8.** *Given a data  $(\alpha, \nu, P)$  and  $\xi \in \mathcal{M}^{c,p,K}$ , we have:*

(i) *If  $\rho < 1$  then  $\lim_{t \rightarrow +\infty} T(t) = \frac{e(\beta^0 + \beta Q P') \bar{Z}(0)}{1 - \rho}$ ;*

(ii) *If  $\rho = 1$  and  $\beta_k^{(2)} < +\infty$  for all  $k \in \mathcal{K}$ , then  $\dot{T}(t) \sim c_1$  and  $T(t) \sim c_1 t$  as  $t \rightarrow +\infty$ , where:*

$$c_1 = \frac{e(\beta^0 + \beta Q P') \bar{Z}(0)}{e(\frac{1}{2}\beta^{(2)} + \beta P' Q \beta) \lambda}. \quad (5.37)$$

(iii) *If  $\rho > 1$  then  $T(t) \sim c_2 \exp(\theta_0 t)$  as  $t \rightarrow +\infty$ , where*

$$c_2 = - \frac{\hat{H}(\theta_0)}{\rho \theta_0 \hat{B}_e^s(\theta_0)}. \quad (5.38)$$

*Proof.* According to Lemma 5.6 the Laplace-Stieltjes transform of  $T$  is given by:

$$\hat{T}(\theta) = \frac{\hat{H}(\theta)}{1 - \rho \hat{B}_e^s(\theta)},$$

the functions being defined in (5.20) and (3.15). On the other hand, the function  $\hat{H}(\theta)$  is regular for  $\Re(\theta) > 0$  and according to the Tauberian theorem [13, p. 446],

$$\lim_{\theta \rightarrow 0^+} \hat{H}(\theta) = \lim_{x \rightarrow +\infty} H(x) = e(\beta^0 + \beta Q P') \bar{Z}(0). \quad (5.39)$$

**Case  $\rho < 1$ .** In this case, the denominator of  $\hat{T}(\theta)$  is defined for all  $\theta$  with positive real part, and  $\lim_{\theta \rightarrow 0^+} (1 - \rho \hat{B}_e^s(\theta)) = 1 - \rho$ . Using (5.39), it follows that:

$$\lim_{\theta \rightarrow 0^+} \hat{T}(\theta) = \frac{e(\beta^0 + \beta Q P') \bar{Z}(0)}{1 - \rho},$$

and according to the Tauberian theorem, the limit of (i) holds.

**Case  $\rho = 1$ .** In the present case,  $T(t)$  solves the renewal equation (5.14), and  $\dot{T}(t)$  solves:

$$\dot{T}(u) = \dot{H}(u) + (B_e^s * \dot{T})(u) .$$

Thanks to the Renewal Theorem [13, Chap XI.I, page 363], it follows that

$$\lim_{t \rightarrow +\infty} \dot{T}(t) = \frac{1}{\beta_e^s} \int_0^{+\infty} \dot{H}(u) du = \frac{1}{\beta_e^s} \lim_{x \rightarrow \infty} H(x) = \frac{\widehat{H}(0)}{\beta_e^s} .$$

The formulas (3.17) and (5.39) allow to conclude that this limit is  $c_1$  as defined in (5.37). The Renewal Theorem applies since the function  $H(\cdot)$  is directly integrable (it is monotone and integrable), and the convolution kernel  $B_e^s(\cdot)$  is not lattice since excess lifetime distributions never are. We therefore have  $\dot{T}(t) \sim c_1$  which implies  $T(t) \sim c_1 t$ . The latter result is also obtained through the Tauberian theorem since in when  $\rho = 1$ , the denominator of  $\widehat{T}(\theta)$  vanishes when  $\theta = 0$ , and using (3.17),

$$\lim_{\theta \rightarrow 0^+} \frac{1}{\theta} \left( 1 - \rho \widehat{B}_e^s(\theta) \right) = e \left( \frac{1}{2} \beta^{(2)} + \beta P' Q \beta \right) \lambda .$$

**Case  $\rho > 1$ .** In that case, the denominator of  $\widehat{T}(\theta)$  vanishes when  $\theta = \theta_0 > 0$ . It follows that, as  $\theta \rightarrow \theta_0^+$ ,

$$\widehat{T}(\theta) \sim \frac{\widehat{H}(\theta_0)}{\rho \widehat{B}_e^s(\theta_0)} \frac{1}{\theta - \theta_0} .$$

Where  $\widehat{B}_e^s(t) = d\widehat{B}_e^s(t)/dt$ . Applying the Tauberian theorem to the shifted Laplace-Stieltjes transform  $\widehat{T}(\theta - \theta_0)$ , one obtains:  $T(t) \sim e^{\theta_0 t} \widehat{H}(\theta_0) / \rho \theta_0 \widehat{B}_e^s(\theta_0)$ . Alternately,  $T(t)e^{-\theta_0 t}$  solves a standard renewal equation, see [13, p. 468]. This proves (iii).  $\square$

### 5.5.2 Asymptotics for the measure $\bar{\mu}(t)$

In this section, we are interested in the behavior of the measure  $\bar{\mu}(t)$  when  $t \rightarrow \infty$ . Accordingly, we suppose that  $\rho \geq 1$  and  $\xi \in \mathcal{M}^{c,p,K}$ , which implies (Lemma 5.4) that  $t^* = \inf\{t \geq 0 : e.\mu(t) = 0\} = +\infty$ . The functions  $\bar{S}(t)$  and  $T(t)$  are therefore defined on  $\mathbb{R}_+$  (see Paragraph 5.1.2). Actually, the results obtained in that paragraph coupled with the asymptotic results for  $T(t)$ , allow to derive the asymptotic behavior of the functions  $\bar{A}(t)$ ,  $\bar{D}(t)$ ,  $\bar{Z}(t)$  and the measure  $\bar{\mu}(t)$ . We choose to use an alternative derivation which emphasizes on an underlying multi-dimensional renewal equations and uses general asymptotic results existing for this class of problems. We begin with some preparatory computations, then state the principal result.

Let us define:

$$\widetilde{M}_k^x(t) = \langle 1_{[x, +\infty)}, \mu_k(T(t)) \rangle .$$

Observe that  $\widetilde{M}_k^0(s) = \widetilde{Z}_k(s)$ . Using Equation (2.10) and performing a change of variable  $t = T(s)$ , we obtain:

$$\widetilde{M}_k^x(t) = \mathbb{P}(v_k^0 > x + t) Z_k(0) + \int_0^s \mathbb{P}(v_k > x + t - s) d\widetilde{A}_k(s) .$$

Equation (5.3) is a special case of this identity for  $x = 0$ . With the vector notation  $\widetilde{M}^x(t) = (\widetilde{M}_1^x(t), \dots, \widetilde{M}_K^x(t))$ , this rewrites in matrix form as:

$$\widetilde{M}^x(t) = (I - B^0(x + t)) \widetilde{Z}(0) + \int_0^t (I - B(x + t - s)) d\widetilde{A}(s) .$$

Making use of Equation (5.7) for  $\tilde{A}(s)$ , the fact that  $\dot{T}(t) = e \cdot \tilde{Z}(t)$  and integrating by parts, one deduces:

$$\tilde{M}^x(t) = C(x+t)\tilde{Z}(0) - (I - B(x))P'Q\tilde{Z}(t) + (G^x * (Q\tilde{Z}))(t), \quad (5.40)$$

where we recall that  $C(t) = I - B^0(t) + (I - B(t))P'Q$ , and where  $G^x(t)$  is the matrix defined as:

$$G^x(t) = \int_x^{x+t} (I - B(u))du \lambda e(I - P') + (B(x+t) - B(x))P'. \quad (5.41)$$

As a special case of (5.41) when  $x = 0$ , the vector  $\tilde{Z}(t)$  is solution of the equation:

$$\tilde{Z}(t) = C(t)\tilde{Z}(0) - P'Q\tilde{Z}(t) + (G^0 * (Q\tilde{Z}))(t),$$

which is equivalent, since  $I + P'Q = Q$ , to:

$$Q\tilde{Z}(t) = C(t)\tilde{Z}(0) + (G^0 * (Q\tilde{Z}))(t). \quad (5.42)$$

The function  $Q\tilde{Z}$  is therefore solution of a multidimensional renewal equation with kernel  $G^0$ . We shall use the results of [1] devoted to this family of equations. Define the renewal matrix function  $U_G(t) = \sum_{n=0}^{\infty} (G^0)^{*n}(t)$ , where  $(G^0)^{*n}(t)$  is the  $n$ -th fold convolution of  $G^0$  with itself. This function is defined for all  $t \geq 0$  since  $G^0(0) = 0$  (see Lemma A.1). According to the same result, the unique locally bounded solution to (5.42) is, for all  $t \geq 0$ ,

$$Q\tilde{Z}(t) = (U_G * C)(t) \tilde{Z}(0).$$

Finally, equation (5.40) can then be written as

$$\tilde{M}^x(t) = C(t+x)\tilde{Z}(0) - (I - B(x))P'Q\tilde{Z}(t) + (G^x * U_G * C)(t)\tilde{Z}(0). \quad (5.43)$$

We are now in position to prove the properties of these various objects.

**Lemma 5.9.** *The matrix of real valued functions  $G^0(t)$  is nonnegative, irreducible for all  $t \in (0, +\infty]$ , non-lattice and has the following properties:*

- (i) if  $\rho = 1$ , then  $\rho(G(\infty)) = 1$  and left- and right- eigenvectors of  $G(\infty)$  are, respectively,  $e$  and  $v = Q\beta\lambda$ ;
- (ii) if  $\rho > 1$ , then the matrix  $\hat{G}(\theta_0)$  has a spectral radius of one, and satisfies  $\hat{G}(\theta_0)Qm = Qm$  and the vector  $\tilde{u} = e(I - P')(I - \hat{B}(\theta_0)P')^{-1}$  is such that  $\tilde{u}\hat{G}(\theta_0) = \tilde{u}$ .

*Proof.* According to the definition (5.41),  $G^0(t)$  is:

$$G^0(t) = \int_0^t (I - B(u))du \lambda e(I - P') + B(t)P'.$$

For any  $t \in (0, +\infty]$ , we have therefore:  $G^0(t) = X\lambda e(I - P') + YP'$  where  $X$  and  $Y$  are strictly positive diagonal matrices.

The matrix is non-lattice thanks to the integral in its definition. It is nonnegative because the vector  $e(I - P')$  is nonzero and nonnegative, which implies that both matrices in the sum are nonnegative. For the irreducibility, assume by contradiction that there exists a partition  $\mathcal{I} \cup \mathcal{J} = \mathcal{K}$  such that  $G^0(t)_{ij} = 0$  for all  $i \in \mathcal{I}$  and  $j \in \mathcal{J}$ . Decomposing the matrices according to the blocs  $\mathcal{I}$  and  $\mathcal{J}$ , we obtain, with obvious notation:

$$G^0(t)_{\mathcal{I}\mathcal{J}} = 0 = (X\lambda e(I - P'))_{\mathcal{I}\mathcal{J}} + (YP')_{\mathcal{I}\mathcal{J}}.$$

Since both matrices are nonnegative, this implies that both are null, with in particular  $P'_{\mathcal{I}\mathcal{J}} = 0$  and therefore that  $(I - P')_{\mathcal{I}\mathcal{J}} = 0$ . Developing with a bloc-matrix product, we obtain that it is necessary that:

$$0 = X_{\mathcal{I}\mathcal{I}} \lambda_{\mathcal{I}} e_{\mathcal{J}}(I_{\mathcal{J}\mathcal{J}} - P'_{\mathcal{J}\mathcal{J}}) .$$

But since, by assumption,  $\lambda_{\mathcal{I}} > 0$  and  $I_{\mathcal{J}\mathcal{J}} - P'_{\mathcal{J}\mathcal{J}}$  is invertible, this equation cannot hold.

The proof of the (i) and (ii) is by inspection, given that:

$$G(\infty) = \beta \lambda e(I - P') + P'$$

and

$$\widehat{G}(\theta) = \frac{1}{\theta}(I - \widehat{B}(\theta))\lambda e(I - P') + \widehat{B}(\theta)P' .$$

The fact that  $G(\infty)e = e$  implies that  $\rho(G(\infty)) = 1$ , according to [2, Corollary 1.12, p. 28].  $\square$

**Lemma 5.10.** *Assume that  $\xi \in \mathcal{M}^{c,p,K}$ , then*

(i) *if  $\rho = 1$ , then:*

$$\lim_{t \rightarrow +\infty} \widetilde{Z}(t) = c_1 \beta \lambda \quad (5.44)$$

$$\lim_{t \rightarrow +\infty} \widetilde{M}^x(t) = c_1 (1 - B_e(x)) \beta \lambda , \quad (5.45)$$

where  $B_e(\cdot)$  is the excess lifetime distribution associated to  $B(\cdot)$ , and where  $c_1$  is defined by (5.37) in Lemma 5.8.

(ii) *if  $\rho > 1$ , then*

$$\lim_{t \rightarrow +\infty} \widetilde{Z}(t)e^{-\theta_0 t} = c_2 m \quad (5.46)$$

$$\lim_{t \rightarrow +\infty} \widetilde{M}^x(t)e^{-\theta_0 t} = c_2 \left( \int_x^\infty (1 - e^{\theta_0(x-y)}) dB(y) \right) (I - \widehat{B}(\theta_0))^{-1} m , \quad (5.47)$$

where  $c_2$  is defined by (5.38) in Lemma 5.8.

*Proof.* Assume first that  $\rho = 1$ . The function  $Q\widetilde{Z}$  is solution of the multidimensional renewal equation (5.42). From Lemma 5.9 (i), the kernel  $G^0(t)$  of this equation is irreducible, non-lattice, and has a spectral radius of one. Left- and right- eigenvectors are  $u = e$  and  $v = Q\beta\lambda$ . Proposition A.1 (i) therefore applies. Straightforward computations lead to:

$$\int_0^\infty C(s)ds = \beta^0 + \beta QP' , \quad (5.48)$$

and then to (5.44).

Next, consider  $x \geq 0$  and Equation (5.43). The convolution matrix in the right-hand side has components that can be written as:

$$(G^x * U_G * C)(t)_{ij} = \sum_{k,\ell} (G_{ik}^x * (U_G)_{k\ell} * C_{\ell j})(t) = \sum_{k,\ell} ((U_G)_{k\ell} * G_{ik}^x * C_{\ell j})(t) ,$$

using the commutativity of scalar convolutions. Using again Proposition A.1 (i), we see that:

$$(G^x * U_G * C)(t)_{ij} \sim \sum_{k,\ell} \frac{v_k u_\ell}{u \Gamma v} \int_0^\infty (G_{ik}^x * C_{\ell j})(s) ds = \sum_{k,\ell} G_{ik}^x(\infty) \frac{v_k u_\ell}{u \Gamma v} \int_0^\infty C_{\ell j}(s) ds .$$

Using (5.48), we have, in matrix form:

$$\lim_{t \rightarrow +\infty} (G^x * U_G * C)(t) = G^x(\infty) \frac{vu}{u\Gamma v} (\beta^0 + \beta QP') .$$

Finally, letting  $t \rightarrow \infty$  in (5.43) and using (5.44), we obtain (5.45).

Now assume that  $\rho > 1$ . According to Lemma 5.9 (ii), the matrix  $\widehat{G}(\theta_0)$  is of spectral radius 1. Left- and right- eigenvectors are  $\tilde{u} = e(I - P')(I - \widehat{B}(\theta_0)P')^{-1}$  and  $\tilde{v} = Qm$ . This matrix is irreducible, as a consequence of the irreducibility of  $G(t)$  for all  $t$ . Proposition A.1 (ii) therefore applies. For the function  $\tilde{Z}(t)$ , this gives:

$$\lim_{t \rightarrow +\infty} Q\tilde{Z}(t)e^{-\theta_0 t} = \frac{\tilde{v}\tilde{u}}{\tilde{u}\tilde{\Gamma}\tilde{v}} \int_0^\infty e^{-\theta_0 s} C(s) ds \tilde{Z}(0) = \frac{\tilde{u}\widehat{C}(\theta_0)\tilde{Z}(0)}{\theta_0 \tilde{u}\tilde{\Gamma}\tilde{v}} Qm .$$

That the constant in this expression coincides with  $c_2$  defined in (5.38) can be checked by inspection. Indeed, we have, using the definition (5.15),

$$\widehat{H}(\theta_0) = \frac{1}{\theta_0} e(I - P')(I - \widehat{B}(\theta_0)P')^{-1} \widehat{C}(\theta_0) \tilde{Z}(0) = \frac{1}{\theta_0} \tilde{u}\widehat{C}(\theta_0)\tilde{Z}(0) .$$

It remains to prove that  $\tilde{u}\widehat{G}(\theta_0)\tilde{v} = -\rho\theta_0\dot{\widehat{B}}^s(\theta_0)$ . Before proceeding, observe that the following relation holds:

$$\begin{aligned} e(I - P')(I - \widehat{B}P')^{-1}(I - \widehat{B})\lambda &= e(I - \widehat{B})(I - \widehat{B}P')^{-1}(I - P')\lambda \\ &= e(I - \widehat{B})(I - \widehat{B}P')^{-1}\alpha = e.m = \theta_0 , \end{aligned} \quad (5.49)$$

as a consequence of the definition of  $m$  in (4.4) and the general matrix identity:

$$(I - A)(I - BA)^{-1}(I - B) = (I - B)(I - BA)^{-1}(I - A) . \quad (5.50)$$

Then, with the argument  $\theta_0$  omitted for more readability,

$$\widehat{G}(\theta_0)Qm = -\frac{1}{\theta_0}(I - \widehat{B})\lambda + \dot{\widehat{B}}P'Qm - \frac{1}{\theta_0}\dot{\widehat{B}}\lambda$$

because  $e(I - P')Qm = \theta_0$ . Next,

$$\tilde{u}\widehat{G}(\theta_0)Qm = -\frac{1}{\theta_0} e(I - P')(I - \widehat{B}P')^{-1}(I - \widehat{B})\lambda + e(I - P')(I - \widehat{B}P')^{-1}\dot{\widehat{B}}(P'Qm - \lambda) ,$$

and the first term in the right-hand side is equal to -1 according to (5.49). Now let us turn to  $\widehat{B}_e^s$ . Starting from Equation (3.16), we get:

$$\begin{aligned} \rho\dot{\widehat{B}}_e^s &= -\frac{1}{\theta_0^2} e(I - P')(I - \widehat{B}P')^{-1}(I - \widehat{B})\lambda + \frac{1}{\theta_0} e(I - P')(I - \widehat{B}P')^{-1}\dot{\widehat{B}}P'(I - \widehat{B}P')^{-1}(I - \widehat{B})\lambda \\ &\quad - \frac{1}{\theta_0} e(I - P')\dot{\widehat{B}}\lambda \end{aligned}$$

$$\theta_0\rho\dot{\widehat{B}}_e^s = -1 + e(I - P')(I - \widehat{B}P')^{-1}\dot{\widehat{B}}\left(P'(I - \widehat{B}P')^{-1}(I - \widehat{B}) - I\right)\lambda ,$$

where we have used again (5.49) in the first term. Finally, the two formulas are identical if

$$P'Qm - \lambda = \left(P'(I - \widehat{B}P')^{-1}(I - \widehat{B}) - I\right)\lambda ,$$

which turns out to be true, using (5.49) a last time.

Turning to the function  $\widetilde{M}^x(t)$ , following the same reasoning as above, we have:

$$(G^x * U_G * C)(t)_{ij} e^{-\theta_0 t} \sim \frac{\tilde{v}_i \tilde{u}_j}{\tilde{u} \tilde{\Gamma} \tilde{v}} \int_0^\infty e^{-\theta_0 s} (G_{ik}^x * C_{lj})(s) ds = \frac{\tilde{v}_i \tilde{u}_j}{\tilde{u} \tilde{\Gamma} \tilde{v}} \frac{1}{\theta_0} \widehat{G}_{ik}^x(\theta_0) \widehat{C}_{lj}(\theta_0).$$

The calculations towards the final formula use in particular the fact that:

$$\begin{aligned} \widehat{G}^x(\theta_0) Qm &= \int_0^\infty e^{-\theta_0 t} [(I - B(x+t)) \lambda e.m dt + B(dt) P' Qm] \\ &= \int_x^\infty e^{\theta_0(x-y)} (I - B(y)) \lambda \theta_0 dt + \int_x^\infty e^{\theta_0(x-y)} r dB(y) P' Qm \\ &= I - B(x) - \int_x^\infty e^{\theta_0(x-y)} dB(y) \lambda + \int_x^\infty e^{\theta_0(x-y)} dB(y) P' Qm \\ &= I - B(x) - \int_x^\infty e^{\theta_0(x-y)} dB(y) (\lambda - P' Qm). \end{aligned}$$

It turns out that  $\lambda - P' Qm = (I - \widehat{B}(\theta_0))^{-1} m$ . The result (5.47) follows by replacing in (5.43).  $\square$

### 5.5.3 Proof of theorems 4.3 and 4.4

The proof of theorems 4.3 and 4.4 follows from Lemmas 5.8 and 5.10. When  $\rho \geq 1$ , we have  $T(t) \rightarrow \infty$ .

Define  $M^x(t) = \langle 1_{[x,+\infty)}, \mu_k(t) \rangle = \widetilde{M}^x(\bar{S}(t))$ . In the case  $\rho = 1$ , we have:

$$\lim_{t \rightarrow \infty} M^x(t) = \lim_{s \rightarrow \infty} \widetilde{M}^x(s) = \frac{e^{(\beta^0 + \beta Q P')} \bar{Z}(0)}{e^{(\frac{1}{2} \beta^2) + \beta P' Q \beta} \lambda} (1 - B_e(x)) \beta \lambda,$$

from Lemma 5.10 (i).

For  $\rho > 1$ , we write:

$$\lim_{t \rightarrow \infty} M^x(t) = \lim_{s \rightarrow \infty} \frac{M^x(T(s))}{T(s)} = \lim_{s \rightarrow \infty} \frac{\widetilde{M}(s)}{T(s)}.$$

According to Lemmae 5.8,  $T(s) \sim c_2 e^{\theta_0 s}$  and by Lemma 5.10 (ii),

$$\lim_{t \rightarrow \infty} M^x(t) = \left( \int_x^\infty (1 - e^{\theta_0(x-y)}) dB(y) \right) (I - \widehat{B}(\theta_0))^{-1} m.$$

In order to see that this coincides with the measure  $s_k(\cdot)$ , it suffices to compute:

$$\begin{aligned} \int_x^\infty p_k(y) dy &= \int_x^\infty \int_y^\infty \theta_0 e^{\theta_0(y-z)} \nu_k(z) dy \\ &= \int_x^\infty \int_x^z \theta_0 e^{\theta_0(y-z)} dy \nu_k(z) \\ &= \int_x^\infty (1 - e^{\theta_0(y-x)}) \nu_k(z). \end{aligned}$$

Since  $\bar{Z}(t) = \langle 1, \bar{\mu}(t) \rangle$ , it follows that  $\bar{Z}(t)/t \rightarrow m$ . Using (2.8) and (2.9), the limits  $\bar{A}(t)/t$  and  $\bar{D}(t)/t$  follow.

## 5.6 Extensions

We present here the proofs of the results of Section 4.3.

### 5.6.1 Proof of Proposition 4.1

We prove the result in the direction “discriminatory to egalitarian”. The converse statement follows easily since only invertible linear transformations and change of variables are involved.

Assume therefore that the triple  $(\bar{A}(t), \bar{D}(t), \bar{\mu}(t))$  is a solution of the DPS Fluid model. Clearly, Equations (2.8) and (2.9) for the egalitarian model are satisfied by the transformations (4.10) and (4.11).

For Equation (2.10) we have: the cumulative of service  $S^g(t)$  of the egalitarian model is such that  $S_k(t) = g_k S^g(t)$ . It follows that:

$$\begin{aligned}
& \langle \mathbf{1}_{[x, \infty)}, \bar{\mu}_k(t) \rangle \\
&= \langle \mathbf{1}_{[x, \infty)}, g_k \mu_k^g(g_k \cdot)(t) \rangle = g_k \langle \mathbf{1}_{[g_k x, \infty)}, \mu_k^g(\cdot)(t) \rangle \\
&= g_k \langle \mathbf{1}_{[g_k x, \infty)}(\cdot - S_k(t)), \mu_k^g(\cdot)(0) \rangle + g_k \int_0^t \langle \mathbf{1}_{[g_k x, \infty)}(\cdot - S_k(t) + S_k(s)), \nu_k \rangle dA_k^g(s) \\
&= g_k \langle \mathbf{1}_{[g_k(x+S^g(t)), \infty)}(\cdot), \mu_k^g(\cdot)(0) \rangle + g_k \int_0^t \langle \mathbf{1}_{[g_k(x+S^g(t)-S^g(s)), \infty)}(\cdot), \nu_k \rangle dA_k^g(s) \\
&= \langle \mathbf{1}_{[x+S^g(t), \infty)}(\cdot), \bar{\mu}_k(\cdot)(0) \rangle + \int_0^t \langle \mathbf{1}_{[x+S^g(t)-S^g(s), \infty)}(\cdot), \nu_k^g(\cdot) \rangle dA_k(s) .
\end{aligned}$$

In the first line of the above equality, one replaces  $\bar{\mu}_k(\cdot)$  by  $g_k \mu^g(g_k \times \cdot)$ , according to (4.11), then one performs a change of variable in integral. The second line follows from the evolution equation of  $\bar{\mu}_k^g$ , (4.9). The third and last lines use respectively the fact that  $S_k(t) = g_k S^g(t)$ ,  $A_k(t) = g_k A_k^g(t)$  and a change of variable.

### 5.6.2 Proof of Proposition 4.2

Let  $S_k(t)$  be the cumulative service of the DPS fluid model and let  $S^g(t)$  be the cumulative service of PS fluid model with data  $(\alpha^g, P^g, \nu^g)$  defined by (4.10). Since the initial state is not zero,  $S_k(t)$  and  $S^g(t)$  are well defined on  $[0, t^*)$ , where  $t^*$  is the time range associated with data  $(\alpha^g, P^g, \nu^g)$ . Let  $T^g(t)$  be the inverse time of  $S^g(t)$  and  $T_k(t)$  be the inverse time of  $S_k(t)$ . Since  $S^g(t) = S_k(t)/g_k$ , we have  $T_k(t) = T^g(t/g_k)$ . Because  $T^g$  is increasing, this implies the claim of Proposition 4.2.

## 6 Proof of the convergence

The objective of this section is to prove the convergence result Theorem 4.5. To that end, we shall state the results which are essential to the proof. The remainder of the section is devoted to the auxiliary results and their proofs. Here we consider all variables and parameters of our system are indexed by  $r$ .

### 6.1 Auxiliary results

Denote by  $\mathcal{B}^r(\cdot)$ ,  $\zeta^r$ , and  $\zeta^{0,r}$  the renewal matrix function defined by (3.2) and the measures defined by (3.10)-(3.11) respectively. Recall that the average values of  $\zeta^r$ , and  $\zeta^{0,r}$  are  $\langle \chi, \zeta_k^r \rangle = (e\beta^r Q^r)_k$ ,  $\langle \chi, \zeta_k^{0,r} \rangle = e.(\beta^{0,r} + \beta^r P^{r'} Q^r)_k$ . Consequently, by assumptions (4.14)-(4.16) and (3.10) we have

$$\begin{aligned}
\zeta_k^r &\Rightarrow \zeta_k \\
\langle \chi, \zeta_k^r \rangle &\rightarrow \langle \chi, \zeta_k \rangle .
\end{aligned}$$

We have used the notation  $\zeta = e(I - P')(\mathcal{B} * \nu)$  and  $\mathcal{B} = \sum_{n \geq 0} (BP')^{*n}$  as in Section 4.4. And from assumptions (4.14)-(4.16), (4.20)-(4.21) and (3.11) we have also

$$\begin{aligned} \zeta_k^{0,r} &\Rightarrow \zeta_k^0 \\ \langle \chi, \zeta_k^{0,r} \rangle &\rightarrow \langle \chi, \zeta_k^0 \rangle \end{aligned}$$

where  $\zeta_k^0 = (e(I - P')(\mathcal{B} * \nu^0))_k$  and  $\beta_k^0 = \langle \chi, \nu_k^0 \rangle$ . Assumption (4.22) implies

$$\langle 1_{\{x\}}, \zeta_k^0 \rangle = 0 \text{ for all } x \in \mathbb{R}_+ \text{ for all } k \in \mathcal{K} .$$

Since the measure valued process  $\{\gamma^r(\cdot)\}$  defined by (3.1) evolves as of that of a single class processor sharing in Gromoll *et al.* [15] and in Puha *et al.* [19], and the above conditions correspond to the assumptions in that case, we have the two following proposition.

**Proposition 6.1.** *Consider a sequence of multiclass processor sharing queues as defined in Section 4.4, satisfying assumptions (4.12)-(4.22). Denote by  $\bar{\gamma}^r(t) = \gamma^r(rt)/r$ . The sequence  $\langle \chi, \bar{\gamma}^r(\cdot) \rangle$  converges in distribution as  $r \rightarrow \infty$  to  $\bar{W}(\cdot)$  solution to the Equation (4.1).*

**Proposition 6.2.** *Consider a sequence of multiclass processor sharing queues as defined in Section 4.4, satisfying assumptions (4.12)-(4.22). Then the sequence  $\bar{\gamma}^r(\cdot)$  converges in distribution as  $r \rightarrow \infty$  to a measure valued process  $\bar{\gamma}(\cdot)$ , the fluid solution of Equation (3.12).*

We state the following Proposition which justifies the fluid limit of Equation (3.30). An outline of its proof will be given in Appendix C.

**Proposition 6.3.** *Consider a sequence of multiclass processor sharing queues as defined in Section 4.4, satisfying assumptions (4.12)-(4.22). For each  $k_0, k \in \mathcal{K}$ , denote by  $\bar{\gamma}_{k_0 k}^r(t) = \gamma_{k_0 k}^r(rt)/r$  the fluid scaled of the process  $\gamma_{k_0 k}^r(t)$ . Then the sequence  $\bar{\gamma}_{k_0 k}^r(\cdot)$  converges in distribution as  $r \rightarrow \infty$  to a measure valued process  $\bar{\gamma}_{k_0 k}(\cdot)$ , the unique fluid solution of the Equation (3.30).*

The following theorem is a corollary of Proposition 6.3.

**Theorem 6.1.** *Consider a sequence of multiclass processor sharing queues as defined in Section 4.4, satisfying assumptions (4.12)-(4.22). Then the sequence  $(\bar{A}^r, \bar{D}^r, \bar{Z}^r)$  converges to  $(\bar{A}, \bar{D}, \bar{Z})$  solution to the fluid model.*

*Proof of Theorem 6.1.* Fix  $k_0, k \in \mathcal{K}$ . By assumption (4.14), we have  $f_{k_0 k}^r \rightarrow f_{k_0 k}$ , and by Lemma 3.3 we have  $N_{k_0 k}^r(1) \Rightarrow N_{k_0 k}$  with  $\mathbb{P}(N_{k_0 k} = m) = f_{k_0 k} (f_{kk})^{m-1} (1 - f_{kk})$ . In addition,  $\mathbb{E}(N_{k_0 k}^r(1)) = f_{k_0 k}^r (1 - f_{kk}^r)^{-1} \rightarrow \mathbb{E}(N_{k_0 k})$ , which is finite by assumption on  $P$ . This implies,

$$\lim_{r \rightarrow \infty} \mathbb{E}(N_{k_0 k}^r(1) 1_{\{N_{k_0 k}^r(1) > r\}}) = 0 .$$

Since  $\mathbb{E}(N_{k_0 k}) = Q_{kk_0}$ , by the weak law of large number for triangular arrays [10], we have

$$\frac{1}{r} \sum_{i=1}^{\lfloor rt \rfloor} N_{k_0 k}^r(i) \Rightarrow Q_{kk_0} t .$$

Assumptions (4.13) and (4.17) imply, by the weak law of large number for renewal processes, that for each  $k \in \mathcal{K}$ ,  $\bar{E}_k^r(t) \Rightarrow \alpha_k t$  as  $r \rightarrow \infty$ . So from the random time change formula (*cf.* [3, Section 17]),

$$\frac{1}{r} \sum_{i=1}^{r \bar{E}_{k_0}^r(t)} N_{k_0 k}^r(i) \Rightarrow Q_{kk_0} \alpha_{k_0} t . \quad (6.1)$$

From assumption (4.19),

$$\frac{1}{r} \sum_{i=1}^{r\bar{Z}_{k_0}^r(0)} N_{k_0k}^r(i) \Rightarrow Q_{kk_0} \bar{Z}_{k_0}(0) . \quad (6.2)$$

Now consider  $\bar{N}_k^r(t)$  the fluid scaled of process  $N_k^r(t)$  defined by (3.21). Since the limits in (6.2) and (6.1) are deterministic,

$$\bar{N}_k^r(t) \Rightarrow (Q\bar{Z}(0))_k + \lambda_k t .$$

Rewriting Equation (3.22) in fluid scaled version, by Proposition 6.3, we have  $\bar{Q}^r(\cdot) \Rightarrow \bar{Q}(\cdot)$ , where  $\bar{Q}(\cdot)$  is the unique fluid solution to (3.31). So  $\bar{Q}(t) = Q\bar{Z}(t)$ , and

$$\bar{D}^r(\cdot) \Rightarrow \bar{D}(\cdot) . \quad (6.3)$$

where  $\bar{D}(t)$  satisfies Equation (5.6). The fluid scaled of Equation (2.1) is

$$\bar{A}_k^r(t) = \bar{E}_k^r(t) + \sum_{l=1}^K \bar{\Phi}_{lk}^r(\bar{D}_l^r(t))$$

By the weak law of large numbers and assumption (4.14), we have  $\bar{\Phi}_{lk}^r([s]) \Rightarrow p_{lk}s$  as  $s \rightarrow \infty$ . Since  $\bar{D}(\cdot)$  is deterministic, we combine the last convergence with that of (6.3) using again the random time change formula to obtain

$$\bar{\Phi}_{lk}^r(\bar{D}_l^r(\cdot)) \Rightarrow p_{lk} \bar{D}_l(\cdot) .$$

Since for all  $l, k \in \mathcal{K}$  the limits are deterministic,  $\sum_{l=1}^K \bar{\Phi}_{lk}^r(\bar{D}_l^r(\cdot)) \Rightarrow \sum_{l=1}^K p_{lk} \bar{D}_l(\cdot)$ , this jointly with  $\bar{E}_k^r(t) \Rightarrow \alpha_k t$ , implies

$$\bar{A}^r(\cdot) \Rightarrow \bar{A}(\cdot) , \quad (6.4)$$

where  $\bar{A}(t)$  jointly with  $\bar{D}(\cdot)$  satisfy Equation (2.8). The limits in (6.3) and in (6.4) together with Equation (2.3) applied to the fluid scaled process give finally  $\bar{Z}^r(\cdot) \Rightarrow \bar{Z}(\cdot)$ .  $\square$

## 6.2 Proof of Theorem 4.5

In this section we give the background for the proof of Theorem 4.5. As we will see, having proved Proposition 6.3 which involves the law of large numbers for the input processes then the proof is an extension of that of Gromoll *et al.* [15] and Puha *et al.* [20]. To complete the proof it will suffice to verify that the sequence  $\{\bar{\mu}_k^r(\cdot)\}_{r>0}$  is tight in  $D([0, \infty), \mathcal{M})$ . By Jakubowski's criterion [9], we must show that:

**C1** For each  $T > 0$  and  $0 < \eta < 1$ , there is a compact subset  $C_{T,\eta}$  of  $\mathcal{M}$  such that

$$\liminf_{r \rightarrow \infty} \mathbb{P}^r(\bar{\mu}_k^r(t) \in C_{T,\eta} \text{ for all } t \in [0, T]) \geq 1 - \eta .$$

**C2** For each  $g \in C_+^1(\mathbb{R}_+)$ , the sequence of real-valued processes  $\{\langle g, \bar{\mu}_k^r \rangle, r \geq 0\}$  is tight.

The strategy outlined to prove C2 is similar to that in [15] and [19]. The essential difference is that the result on controlled oscillations of  $\{\langle g, \bar{\mu}_k^r(\cdot) \rangle\}$  for each  $g \in C_b^1(\mathbb{R}_+)$  is proved here by considering the fact that the global number of customers in the system admits a limit fluid  $\langle 1, \bar{\gamma}(\cdot) \rangle$ , instead of considering the global workload alone. More precisely, we consider two cases: (a)  $\langle 1, \bar{\gamma}(0) \rangle > 0$  and (b)  $\langle 1, \bar{\gamma}(0) \rangle = 0$  and  $\rho > 1$ . Note that the remaining case  $\langle 1, \bar{\gamma}(0) \rangle = 0$  and  $\rho \leq 1$ , convergence is a consequence of Lemma 6.1 below and Lemma C.1 (in Appendix C): in that case  $\bar{\mu}_k^r(\cdot) \Rightarrow 0$ . In case (a) the fluid scaled  $\langle 1, \bar{\gamma}^r(\cdot) \rangle$  is not small on each compact  $[0, T]$  for  $T < t_\rho$ , where  $t_\rho$  is given by (2.12), in that case, the proof of tightness of  $\{\langle g, \bar{\mu}_k^r(\cdot) \rangle\}$  is in the same way as that of Gromoll *et al.* [15]. In case (b), let

$$s = (1 - \hat{B}_k(\theta_0))/4m_k . \quad (6.5)$$

Observe that, under that case, by time  $s\varepsilon/2$  the fluid scaled process  $\langle 1, \bar{\mu}_k^r(\cdot) \rangle$  is of order of  $\varepsilon$  as  $r \rightarrow \infty$  and therefore does not admit large oscillations on the compact time interval  $[0, s\varepsilon/2]$ . Thereafter, it suffices to prove the controlled oscillation on each compact time interval  $[s\varepsilon/2, T]$ , which can be derived as in case (a). To handle the established results in proving Proposition 6.3 and give a simple proof, we need the Lemmas 6.1 and 6.2 below. The first lemma proves that the measure valued process  $\{\bar{\mu}_k^r(\cdot)\}$  is bounded by  $\{\sum_{k_0=1}^K \bar{\gamma}_{k_0 k}^r(\cdot)\}_{r>0}$  and the second summarizes some conditions that are necessary for condition C2. The proofs that any limit of  $\{\bar{\mu}_k^r(\cdot)\}$  satisfies conditions (i), (ii) and (iii) of proposition 5.4, are much like in Gromoll and al.[15] and Puha and al. [20]. We begin by stating the dynamic Equation. This equation is derived from (2.6). For each  $g \in \mathcal{C}_b^1(\mathbb{R}_+)$  and for each  $t \geq 0, h \geq 0$  we have,

$$\begin{aligned} \langle g(\cdot), \bar{\mu}_k^r(t+h) \rangle &= \langle 1_{(0,\infty)} g(\cdot - \bar{S}^r(t+h) + \bar{S}^r(t)), \bar{\mu}_k^r(t) \rangle \\ &+ \frac{1}{r} \sum_{i=r\bar{A}_k^r(t)+1}^{r\bar{A}_k^r(t+h)} (1_{(0,\infty)} g)(v_k^r(i) - \bar{S}^r(t+h) + \bar{S}^r(U_k^r(i)/r)) . \end{aligned} \quad (6.6)$$

We use here the convention that, any function defined on  $\mathbb{R}_+$ , is extended to be identically equal to zero on  $(-\infty, 0)$  so that for all  $a > 0$ , the function  $g(\cdot - a)$  is well defined on  $\mathbb{R}$ .

**Lemma 6.1.** *For each  $k \in \mathcal{K}$  and for each constant  $\kappa \geq 0$ , we have for each  $r > 0$*

$$\langle 1_{[0,\kappa]}, \bar{\mu}_k^r(t) \rangle \leq \langle 1_{[0,\kappa]}, \sum_{k_0 \in \mathcal{K}} \bar{\gamma}_{k_0 k}^r(t) \rangle .$$

*Proof.* Let  $i$  be some customer arrived at class  $k$  by time  $t$ , (i.e  $i \in \{1, \dots, A_k^r(t)\}$ );  $\sigma_k^r(i)$  is the time at which he arrived. If  $i$  is an initial customer in the system, then there exist  $k_0 \in \mathcal{K}$  and  $j \in \{1, \dots, \bar{Z}_{k_0}^r(0)\}$  and  $n \in \{1, \dots, N_{k_0 k}^r(j)\}$  (Note that  $n \in \{2, \dots, N_{k_0 k}^r(j)\}$  if  $k_0 = k$ ) such that

$$\bar{S}^r(\sigma_k^r(i)/r) + v_k^r(i) = V_{k_0 k}^{0,r}(j, n) .$$

Otherwise, there exist  $k_0 \in \mathcal{K}$  and  $j \in \{1, \dots, \bar{E}_{k_0}^r(t)\}$  and  $n \in \{1, \dots, N_{k_0 k}^r(j)\}$  such that

$$\bar{S}^r(\sigma_k^r(i)/r) - \bar{S}^r(U_k^r(i)/r) + v_k^r(i) = V_{k_0 k}^r(j, n) .$$

This implies,

$$\begin{aligned} \sum_{i=1}^{A_k^r(t)} 1_{\{0 \leq v_k^r(i) - \bar{S}^r(t) + \bar{S}^r(\sigma_k^r(i)/r) \leq \kappa\}} &\leq \sum_{k_0 \in \mathcal{K} \setminus k} \sum_{j=1}^{Z_{k_0}^r(0)} \sum_{n=1}^{N_{k_0 k}^r(j)} 1_{\{0 \leq V_{k_0 k}^{0,r}(j, n) - \bar{S}^r(t) \leq \kappa\}} \\ &+ \sum_{j=1}^{Z_k^r(0)} \sum_{n=2}^{N_{kk}^r(j)} 1_{\{0 \leq V_{kk}^{0,r}(j, n) - \bar{S}^r(t) \leq \kappa\}} \\ &+ \sum_{k_0 \in \mathcal{K}} \sum_{j=1}^{E_{k_0}^r(t)} \sum_{n=1}^{N_{k_0 k}^r(j)} 1_{\{0 \leq V_{k_0 k}^r(j, n) - \bar{S}^r(t) + \bar{S}^r(U_{k_0}^r(j)/r) \leq \kappa\}} . \end{aligned}$$

By adding to both terms, the number of customers that are initially in class  $k$ ,

$$\sum_{j=1}^{Z_k^r(0)} 1_{\{0 \leq v_k^r(j) - \bar{S}^r(t) \leq \kappa\}}$$

and using the fact that  $V_{kk}^r(j, 1) = v_k^r(j)$ , we obtain the desired inequality.  $\square$

In the following we consider the convention that the constant  $s$  defined by (6.5) is extended in case (a) to  $s = 0$ . The proofs of Lemma 6.2 and Lemma 6.3 below are omitted since they are similar to the proofs of Gromoll [15]. The key ingredient for the first lemma is the functional law of large number for input process  $A^r(\cdot)$ , service times, and Lemma 6.1 that gives a fine estimate for the amount of mass that  $\bar{\mu}_k^r(t)$  can have concentrated near zero. Lemma 6.3 summarizes the useful results that implies tightness.

**Lemma 6.2.** *For each  $k \in \mathcal{K}$ , For each  $T \in (s, t^*)$  and for each  $\epsilon, \eta < 1$ , there exist constants  $M_T, M_0, \Gamma, l, \kappa, r_0$ , and a sequence of events  $\{B^r\}$  such that  $\mathbb{P}^r(B^r) \geq 1 - \eta$  for  $r > r_0$ , and on  $B^r$  we have:*

$$l \leq s\epsilon/2 \quad (6.7)$$

$$\bar{A}_k^r(t+l) - \bar{A}_k^r(t) \leq \epsilon \quad (6.8)$$

$$\bar{A}_k^r(t) \leq 2\bar{A}_k(t) \quad (6.9)$$

$$\inf_{t \in [\epsilon s/2, T]} \langle 1, \bar{\gamma}^r(t) \rangle > \frac{1}{\Gamma} \quad (6.10)$$

$$\langle 1, \bar{\mu}_k^r(0) \rangle \leq M_s \quad (6.11)$$

$$\sup_{t \in [0, T]} \langle 1, \bar{\mu}_k^r(t) \rangle < M_T \quad (6.12)$$

$$\sup_{t \in [0, T]} \langle 1_{[0, \kappa]}, \bar{\mu}_k^r(t) \rangle < \frac{\epsilon}{2}, \quad (6.13)$$

The constant  $M_s$  in condition (6.11) is defined by : in case  $\langle 1, \bar{\mu}_k(0) \rangle > 0$ :  $M_s = M_0$ , and in case  $\langle 1, \bar{\mu}_k(0) \rangle = 0$ :  $M_s = \epsilon/2$ .

As a consequence of the above lemma we have the following which implies condition C2.

**Lemma 6.3.** *Let  $g \in C_b^1(\mathbb{R}_+)$ ,  $s \leq T < t^*$ , where  $s$  is defined by (6.5), and  $0 < \epsilon, \eta < 1$ . Set  $\epsilon = \frac{\epsilon}{2(\|g\|_\infty \vee 1)}$ . Let  $M_T, M_0, \Gamma, l, \kappa, r_0$  be the constants, and  $(B^r, r > 0)$  be the events, given by Lemma 6.2. Set  $M = (\|g\|_\infty \vee 1)M_T$  and*

$$\delta = \min\{T/2, l, \frac{\epsilon}{4\Gamma M_T(\|g'\|_\infty \vee 1)}, \kappa/\Gamma, 1\}.$$

Then

$$\mathbb{P}^r \left( \sup_{t \in [0, T]} |\langle g, \bar{\mu}_k^r(t) \rangle| \leq M \right) \geq 1 - \eta \quad (6.14)$$

$$\mathbb{P}^r \left( \sup_{t \in [\epsilon s/2, T-\delta]} \sup_{h \in [0, \delta]} |\langle g, \bar{\mu}_k^r(t+h) \rangle - \langle g, \bar{\mu}_k^r(t) \rangle| \leq \epsilon \right) \geq 1 - \eta. \quad (6.15)$$

Furthermore, in case (b) we have

$$\mathbb{P}^r \left( \sup_{t \in [0, \epsilon s/2]} \sup_{h \in [0, \delta]} |\langle g, \bar{\mu}_k^r(t+h) \rangle - \langle g, \bar{\mu}_k^r(t) \rangle| \leq \epsilon \right) \geq 1 - \eta. \quad (6.16)$$

*Proof of tightness of  $\{\bar{\mu}_k^r(\cdot)\}$ .* It suffices to verify both properties C1 and C2. The property C2 follows directly from Lemma 6.3. For Property C1, we fix  $T > 0$  and  $0 < \epsilon, \eta < 1$ . From Equation (2.6) we have

$$\langle \chi, \bar{\mu}_k^r(t) \rangle \leq \frac{1}{r} \sum_{i=1}^{r\bar{Z}_k^r(0)} v_k^{0,r}(i) + \frac{1}{r} \sum_{i=1}^{r\bar{A}_k^r(t)} v_k^r(i)$$

By Theorem 6.1,  $\bar{A}_k^r(\cdot) \Rightarrow \bar{A}_k(\cdot)$ . This together with condition (4.19) implies, by using the weak law of large numbers and the random time change formula, that the second member of the above inequality converges to  $\beta_k^0 \bar{Z}_k(0) + \beta_k \bar{A}_k(t)$ . Since the limit is deterministic and continuous, the above inequality yields

$$\lim_{r \rightarrow \infty} \mathbb{P}^r \left( \sup_{t \in [0, T]} \langle \chi, \bar{\mu}_k^r(t) \rangle < 2(\beta_k^0 \bar{Z}_k(0) + \beta_k \bar{A}_k(T)) \right) = 1 . \quad (6.17)$$

Let  $N_T = \max(M_T, 2(\beta_k^0 \bar{Z}_k(0) + \beta_k \bar{A}_k(T)))$ . Using (6.10) and (6.17) we obtain

$$\lim_{r \rightarrow \infty} \mathbb{P}^r \left( \sup_{t \in [0, T]} \langle 1, \bar{\mu}_k^r(t) \rangle \vee \langle \chi, \bar{\mu}_k^r(t) \rangle < N_T \right) = 1 . \quad (6.18)$$

Let  $C_{T, \eta}$  be the set  $C_{T, \eta} = \{\xi \in \mathcal{M} : \langle 1, \xi \rangle \vee \langle \chi, \xi \rangle \leq N_T\}$ . By the same argument of Gromoll [15],  $C_{T, \eta}$  is relatively compact and (6.18) implies

$$\liminf_{r \rightarrow \infty} \mathbb{P}^r (\bar{\mu}_k^r(t) \in C_{T, \eta} \text{ for all } t \in [0, T]) \geq 1 - \eta$$

which proves C1. □

## 7 Illustrations

### 7.1 Trajectories

We illustrate in this section the effective construction of trajectories, in a case where computations can be performed in closed form. The first step is to compute the function  $\bar{T}(\cdot)$ . Lemma 5.6 provides its Laplace Transform.

Consider a queue with two classes. Customers of class 1 have no external arrivals ( $\alpha_1 = 0$ ), their service is distributed as  $\text{Exp}(\mu_1)$  and when they complete service, they turn into customers of class 2:  $P_{12} = 1$ . Customers of class 2 arrive from the exterior with rate  $\alpha$ . Their service is distributed as  $\text{Exp}(\mu_2)$  and when the complete service, they exit the system:  $P_{21} = P_{22} = 0$ . The initial situation is that there is one unit of fluid of class 1, and no fluid of class 2:  $\bar{Z}(0) = (1, 0)'$ . Service times of customers present initially have the same distribution as that of regular customers. In that case, we have:

$$\hat{T}(\theta) = \frac{\theta + \mu_1 + \mu_2}{(\theta + \mu_1)(\theta + \mu_1 - \alpha)} ,$$

and inversion of the Laplace transform gives:

$$\bar{T}(t) = \frac{\mu_1 + \mu_2}{\mu_1(\mu_1 - \alpha)} + \frac{\mu_2 e^{-\mu_1 t}}{\mu_1(\mu_1 - \mu_2 + \alpha)} + \frac{(\alpha + \mu_1) e^{(\alpha - \mu_2)t}}{(\alpha - \mu_2)(\mu_1 - \mu_2 + \alpha)} .$$

The next step is to compute the function  $\bar{S}(s)$  by solving  $s = \bar{T}(t)$ , which cannot always be done in closed form, even in this simple case. We proceed with specific values.

If  $\mu_1 = 4$  and  $\mu_2 = 1$ , and the arrival rate is  $\alpha = 1/2$ , the load factor is  $\rho = 1/2$ . The solution is given by:

$$\begin{aligned} \bar{T}(t) &= 10 - 16e^{-t/4} + 6e^{-t/2} \\ \bar{S}(s) &= -4 \log \left( \frac{4}{3} - \frac{\sqrt{4+6s}}{6} \right) \\ Z_1(s) &= \frac{4}{3} - \frac{\sqrt{4+6s}}{6} \\ Z_2(s) &= 3 Z_0(s) (1 - Z_0(s)) . \end{aligned}$$

The corresponding trajectories are displayed in Figure 1. The figure also displays one normalized trajectory obtained by simulation with an initial population of 1000 customers. Simulations with a smaller initial population exhibit significant differences with the fluid trajectory, certainly due to the fact that the approximation of the initial random workload by its fluid counterpart has a bad precision then.

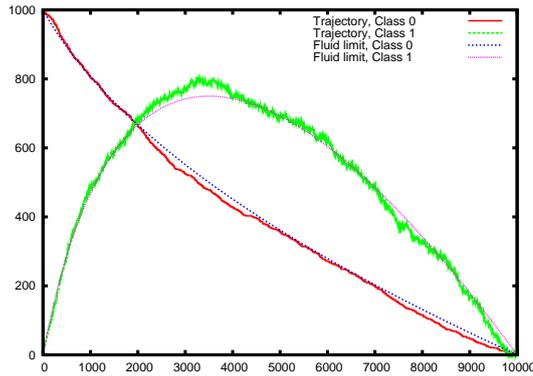


Figure 1: Trajectories in a stable case

For the same values of  $\mu_1$  and  $\mu_2$ , when the arrival rate is  $\alpha = 5/4$ , the load factor is  $\rho = 5/4$ . The solution is given by:

$$Z_1(s) = \frac{5}{4} + \frac{s}{16} - \frac{\sqrt{16 + 40s + s^2}}{16}$$

$$Z_2(s) = 3 \left( \frac{1}{Z_0(s)} - Z_0(s) \right) .$$

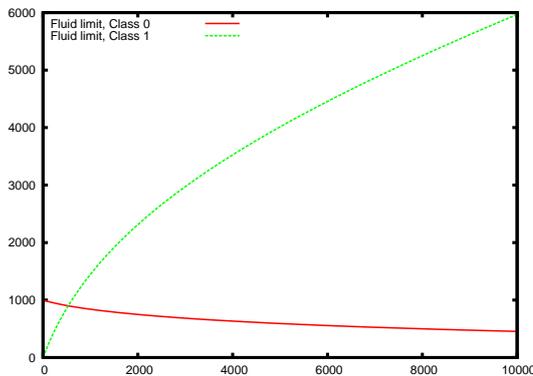


Figure 2: Trajectories in an unstable case

## 7.2 Slowness as a function of the service time

Consider a single-class, processor sharing with service time distribution  $v$ . Assume that this distribution is discrete, with  $\mathbb{P}(v = k\delta) = \pi_k$ ,  $\sum_{k=1}^{\infty} \pi_k = 1$ , for some parameter  $\delta > 0$ . Denote also  $f_k = \sum_{j=k}^{\infty} \pi_k = \mathbb{P}(v \geq k\delta)$ .

According to the discussion in Section 3.1, this single-class queue can be seen as a multiclass queue. Let  $k$  denote the class of customers having their service time equal to  $k\delta$  in the single-class queue. In the multiclass queue, such customers are considered to have a service time equal to  $\delta$ , and a routing probability  $P_{k,k+1} = \mathbb{P}(v \geq (k+1)\delta | v \geq k\delta) = f_{k+1}/f_k$ . If the support of the distribution is not bounded, there is an infinite number of classes. For the discussion to follow, we shall informally consider that the results of Section 4 apply with infinitely many classes. For a rigorous discussion, we might as well truncate the distribution, then let the truncation threshold go to infinity.

In the multi-class view of the system, the matrix  $Q = (I - P')^{-1}$  is given by:

$$Q = \begin{pmatrix} 1 & 0 & 0 & \dots \\ f_2/f_1 & 1 & 0 & \dots \\ f_3/f_1 & f_3/f_2 & 1 & \ddots \\ \vdots & & & \ddots \end{pmatrix},$$

in other words,  $Q_{ij} = f_i/f_j$  for  $i \geq j$ , 0 otherwise. Observe that  $f_1 = 1$ . The vector of external arrival rates is  $(\alpha, 0, 0, \dots)'$ , and the vector of (theoretical) global arrival rates is, as expected by construction,  $\lambda = Q\alpha = \alpha(1, f_2, f_3, \dots)$ .

Assuming that the system is supercritical, we define  $\theta_0$  as the solution to the equation:

$$\theta_0 = \alpha \left( 1 - \sum_{k \geq 0} \pi_k e^{-k\delta\theta_0} \right),$$

and apply (4.4) to obtain the value of the vector  $m$ . We have simply  $\widehat{B}(\theta_0) = e^{-\delta\theta_0}I$ , and it is easily seen that:

$$(I - P'\widehat{B}(\theta_0))^{-1}\alpha = \alpha(1, f_2\widehat{B}(\theta_0), f_3\widehat{B}(\theta_0)^2, \dots) = \alpha(1, f_2e^{-\delta\theta_0}, f_3e^{-2\delta\theta_0}, \dots).$$

Accordingly, applying Theorem 4.4, we obtain for the asymptotic growth rate, arrival rate and departure rate of customers of class  $k$ , respectively:

$$m_k = \alpha (1 - e^{-\delta\theta_0}) e^{-(k-1)\delta\theta_0}, \quad a_k = \alpha e^{-(k-1)\delta\theta_0}, \quad d_k = \alpha e^{-k\delta\theta_0}.$$

This shows that the departure rate of customers is decreased *exponentially* as a function of their service length. The factor of this exponential decay is the factor  $\theta_0$ . This is in accordance with prior findings that the response time of customers grows exponentially with their service time.

## 7.3 Competition between classes

We illustrate here how the processor sharing discipline “distorts” the throughputs of classes, in the case of overload. Consider a multiclass queue in which customers of class  $k$  arrive with a rate  $\alpha_k$ , receive service, then leave the system. The routing matrix is  $P' = 0$ .

The reference situation is that the available service capacity is “fairly” shared among classes, proportionally to their load factor  $\rho_k = \alpha_k\beta_k$ . This situation is that of a stable server, and that of an overloaded FIFO queue.

In the supercritical case, the “fair” situation is therefore that the throughput of class  $k$  is  $\alpha_k/\rho$ . We introduce the index

$$\phi_k = \frac{d_k}{\alpha_k} \rho,$$

where  $d_k$  is the effective throughput of customers of class  $k$ . If this index is smaller than 1, the class is unfairly treated with respect to the ideal situation. If it is larger than 1, the class obtains more throughput than it would in a FIFO queue. According to Theorem 4.4, and the definition of  $m$  (4.4), the throughput  $d_k$  is  $\alpha_k - m_k = \alpha_k \widehat{B}_k(\theta_0)$ . The index for class  $k$  is therefore  $\phi_k = \rho \widehat{B}_k(\theta_0)$ .

### 7.3.1 The effect of service length.

Here an example with two classes, in which the distributions have the same “shape”, the same load, but not the same average service time. Figure 3 represents the proportion of customers of class 1 in queue, as a function of  $\rho$  and  $\xi = \mu_2/\mu_1$ .

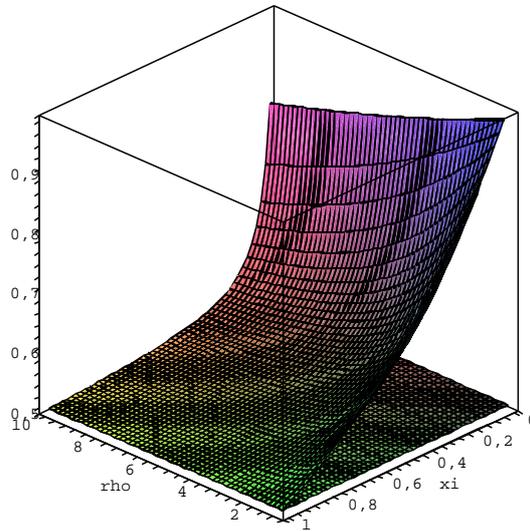


Figure 3: Proportion of customers of class 1 & 2 in queue, exponential distributions

### 7.3.2 The effect of the distribution

Here an example with two classes, in which the distributions have the same load and average service time, but not the same “shape”. The service for the first class has an exponential distribution, whereas the service for the second class has a Pareto distribution. Figure 4 represents the proportion of customers of class 1 & 2 in queue, as a function of  $\rho$ , for different Pareto shape parameters.

Finally, when customers of class 2 have a deterministic service time, the proportions are as displayed in Figure 5.

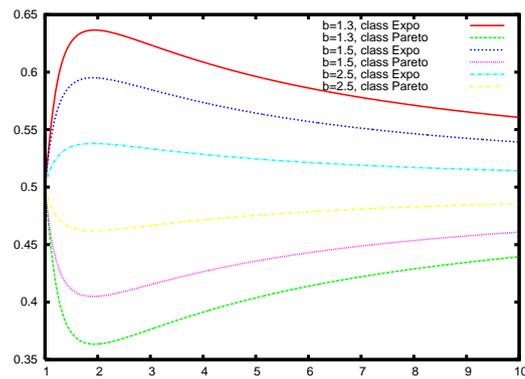


Figure 4: Proportion of customers of class 1 & 2 in queue, Exponential *vs.* Pareto

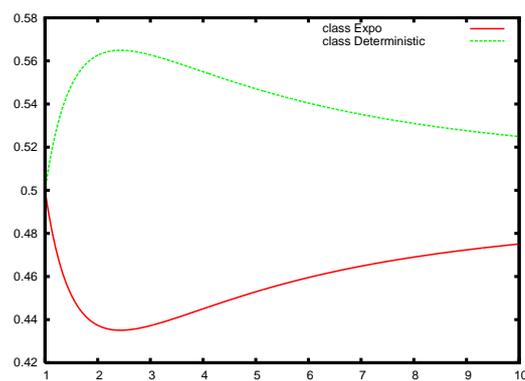


Figure 5: Proportion of customers of class 1 & 2 in queue, Exponential *vs.* Deterministic

## References

- [1] Athreya, K.B. and Rama Murthy, K. “Feller’s Renewal Theorem for Systems of Renewal Equations”, IISC, 1976.
- [2] Berman, A. and Plemmons, A.J. *Nonnegative Matrices in the Mathematical Sciences*. SIAM Classics in Applied Mathematics, Vol. 9, 1994.
- [3] Billingsley, P. *Convergence of probability Measures*. John Wiley and Sons, INC, New York, 1968.
- [4] Bramson, M. Convergence to equilibria for fluid models of FIFO queueing networks. *Queueing Systems Theory Appl.*, 22, no. 1-2, 5–45, 1996.
- [5] Bramson, M. Convergence to equilibria for fluid models of head-of-the-line proportional processor sharing queueing networks. *Queueing Systems Theory Appl.*, 23, no. 1-4, 1–26, 1997.
- [6] Bramson, M. State space collapse with application to heavy traffic limits for multiclass queueing networks. *Queueing Systems Theory Appl.* 30, no. 1-2, 89–148, 1998.
- [7] Bramson, M. Dai, J. G. Heavy traffic limits for some queueing networks. *Ann. Appl. Probab.*, 11, no. 1, 49–90, 2001.
- [8] Chen, H., Kella, O. and Weiss, G., Fluid approximations for a processor sharing queue. *Queueing Systems Theory Appl.*, 27:99-125, 1997.
- [9] Dawson, D.A. Measure-valued Markov processes, école d’été de probabilités de Saint Flour, volume XXI, lecture notes in Mathematics NO 1541, Springer, 1993.
- [10] Durrett, R. T., *Probability: Theory and Examples*, 2nd ed. Duxbury Press, Belmont, CA., 1996.
- [11] Ethier, S. N. and Kurtz, T. G. *Markov Processes: Characterization and Convergence*. Wiley, New York, 1986.
- [12] W. Feller, *An Introduction to Probability and Applications*, vol. 1, 3rd edition, J. Wiley & Sons, 1968.
- [13] W. Feller, *An Introduction to Probability and Applications*, vol. 2, 2nd edition, J. Wiley & Sons, 1971.
- [14] Gromoll, H, C. Diffusion approximation for a processor sharing queue in heavy traffic. *Ann. Appl. Probab.*, 14 555-611, 2004.
- [15] Gromoll, H. C., A. L. Puhá, R. J. Williams, “The fluid limit of a heavily loaded processor sharing queue”, *Ann. Appl. Probab.*, 12, 797–859, 2002.
- [16] Gromoll, H. C. and Kruk, L. Heavy traffic limit for a processor sharing queue with soft deadlines. *Ann. Appl. Probab.*, 17, no. 3, 1049–1101, 2007.
- [17] Gromoll, H. C., Robert, Ph. and Zwart, B. Fluid Limits for Processor-Sharing Queues with Impatience, *Math. Ops. Res.*, 33 (2), pp. 375–402, 2008.
- [18] Jean-Marie, A. Robert, P. 1994. On the transient behavior of the processor sharing queue. *Queueing Systems: Theory Appl.* 17 129-136.
- [19] Puhá, A.L., Stolyar, A.L. and Williams R.J., “The Fluid Limit of an Overloaded Processor Sharing Queue”, *Math. Ops. Res.*, Vol. 31, No. 2, May 2006, pp. 316–350.

- [20] Puha, A. L., R. J. Williams. 2004. “Invariant states and rates of convergence for the fluid limit of a heavily loaded processor sharing queue”, *Ann. Appl. Probab.*, Vol. 14, pp. 517–554.
- [21] B. de Saporta, *Étude de la solution stationnaire de l'équation  $Y(n+1) = a(n)Y(n)+b(n)$  à coefficients aléatoires*, PhD thesis, University of Rennes 1, 2004.
- [22] Williams, R. J. 1998. Diffusion approximation for open multiclass queueing networks: Sufficient conditions involving state space collapse. *Queueing Systems Theory Appl.* 30 27-88.
- [23] Zhang, J. Dai, J.G. Zwart, B. Law of Large Number Limits of Limited Processor Sharing Queues. Processor sharing. Tech. Rep., Georgia Institute of Technology [http://www2.isye.gatech.edu/people/faculty/dai/publications/draft\\_zhangDaiZwart08.pdf](http://www2.isye.gatech.edu/people/faculty/dai/publications/draft_zhangDaiZwart08.pdf), 2008.

## A Multidimensional renewal equations

Consider  $F(t) = (F_{ij}(t))_{i,j}$  a matrix of increasing and r.c.c.l. functions, such that  $F_{ij}(t) = 0$  when  $t < 0$ . Let  $\widehat{F}(\theta)$  denote the Laplace-Stieltjes transform of  $F$ . With the definition of matrix-matrix and matrix-vector convolutions assumed in Section 3.1, define the renewal matrix as:

$$U(t) = \sum_{k=0}^{\infty} F^{(*k)}(t).$$

Let  $H$  be a vector of measurable and bounded functions. The matrix-renewal equation is the system of equations:

$$V_i(t) = H_i(t) + \sum_j \int_0^t V_j(t-s) dF_{ij}(s),$$

for all  $i$ . In vector notation, this equation can be written as:

$$V(t) = H(t) + (F * V)(t).$$

The following results is quoted from [21] (see also [1, Lemma 2.1]).

**Lemma A.1** ([21], Lemma 3, p. 23). *The function  $U(t)$  is finite for all  $t$  if, and only if,  $\rho(F(0)) < 1$ .*

*If  $\rho(F(0)) < 1$ , then  $V = U * H$  is the unique measurable and bounded solution to the matrix-renewal equation.*

The matrix  $F$  will be called *lattice* if there exists a real value  $\lambda$  such that:  $F_{ij}$  is concentrated on  $b_{ij} + n_{ij}\lambda\mathbb{Z}$  (with  $n_{ij} \in \mathbb{N}$ ,  $b_{ii} = 0$ ) and if, assuming that  $a_{ij}$  is an increasing point for  $F_{ij}$ , all values  $(a_{ij} + a_{jk} - a_{ki})/\lambda$  are integer.

The following proposition summarizes the asymptotic results we need. They are a direct consequence of the results of [1] or [21]. The irreducibility condition, which is implicit in [1], is explicitly added here.

**Proposition A.1.** *Let  $g$  be a directly integrable function. Then:*

- (i) *If  $\rho(F(\infty)) = 1$ ,  $F(\infty)$  is irreducible and not lattice, then, for  $u$  and  $v$  left- and right- eigenvectors of  $F(\infty)$  for the eigenvalue 1, such that  $u.v = 1$ ,*

$$\lim_{t \rightarrow \infty} (U_{ij} * g)(t) = \frac{v_i u_j}{u \Gamma v} \int_0^{\infty} g(s) ds,$$

with

$$\Gamma = \int_0^\infty s \, dF(s) = -\frac{d\widehat{F}}{d\theta}(0).$$

If one component of  $\Gamma$  is infinite, the limit above is 0.

(ii) If  $\rho(F(\infty)) \neq 0$ , and if there exists a positive  $\theta_0$  such that  $\rho(\widehat{F}(\theta_0)) = 1$  and  $\widehat{F}(\theta_0)$  is irreducible, then, for  $\tilde{u}$  and  $\tilde{v}$  left- and right- eigenvectors of  $\widehat{F}(\theta_0)$  for the eigenvalue 1, such that  $\tilde{u} \cdot \tilde{v} = 1$ , then

$$\lim_{t \rightarrow \infty} (U_{ij} * g)(t) e^{-\theta_0 t} = \frac{\tilde{v}_i \tilde{u}_j}{\tilde{u} \tilde{\Gamma} \tilde{v}} \int_0^\infty e^{-\theta_0 s} g(s) \, ds,$$

where

$$\tilde{\Gamma} = \int_0^\infty s e^{-\theta_0 s} dF(s) = -\frac{d\widehat{F}}{d\theta}(\theta_0).$$

If one component of  $\tilde{\Gamma}$  is infinite, the limit above is 0.

## B Convergence of shifted convolutions

This section is devoted to the following technical result:

**Lemma B.1.** *Consider:*

- A sequence of increasing functions  $A^n : \mathbb{R}_+ \rightarrow \mathbb{R}$ , which are such that  $A^n(0) = 0$  and which converges uniformly on compacts to a function  $A$  which is Lipschitz-continuous, and such that  $A(t) > 0$  if  $t > 0$ .
- A sequence of functions  $f^n : \mathbb{R}_+^2 \rightarrow \mathbb{R}$  which converges uniformly on every compact of  $(0, \infty) \times (0, \infty)$ , to a continuous function  $f$  which satisfies:
- for every  $t > 0$  the function  $f_t(\cdot) : (0, t] \rightarrow \mathbb{R}_+$  defined by  $f_t(s) = f(t, s)$  is continuous and strictly decreasing, and  $f(t, t) = 0$ .

Then for every probability measure  $\nu$  on  $\mathbb{R}_+$ , the following statements hold, for every function  $g \in C_b(\mathbb{R}_+)$ , extended with  $g(x) = 0$  for  $x < 0$ :

- i) for each  $t > 0$  the function  $s \in (0, t] \mapsto \langle g(\cdot - f(t, s)), \nu \rangle$  is continuous, except for the at most countable set of values  $s$  where  $\nu(\{f(t, s)\}) \neq 0$ ;
- ii) for every  $\varepsilon > 0$  and for every  $T > T_0 > 0$  there exists  $\eta_0 > 0$  and  $N_0 > 0$  such that

$$\int_{T_0}^t \langle 1_{[(f(t, s) - \eta_0)^+, (f(t, s) + \eta_0)]}, \nu \rangle dA^n(s) < \varepsilon \quad (\text{B.1})$$

for all  $n \geq N_0$  and  $t \in [T_0, T]$ ;

- iii) the sequence of functions  $h_n(t) = \int_0^t \langle g(\cdot - f(t, s)), \nu \rangle dA^n(s)$  is equicontinuous on finite intervals;
- iv) for every  $T > 0$ , the following convergence holds:

$$\lim_{n \rightarrow +\infty} \sup_{0 \leq t \leq T} \left| \int_0^t \langle g(\cdot - f^n(t, s)), \nu \rangle dA^n(s) - \int_0^t \langle g(\cdot - f(t, s)), \nu \rangle dA(s) \right| = 0. \quad (\text{B.2})$$

*Proof of i).* Denote with  $h_t(s)$  the function:

$$h_t(s) = \int_{f(t,s)}^{\infty} g(x - f(t, s)) \nu(dx) .$$

Let  $s \leq t$  be such that  $\nu(\{f(t, s)\}) = 0$ . Let  $s_n$  be a decreasing sequence converging to  $s$ . We have:

$$\begin{aligned} & |h_t(s_n) - h_t(s)| \\ &= \left| \int_{f(t,s)}^{\infty} g(x - f(t, s)) \nu(dx) - \int_{f(t,s_n)}^{\infty} g(x - f(t, s_n)) \nu(dx) \right| \\ &\leq \left| \int_{f(t,s)}^{f(t,s_n)} g(x - f(t, s)) \nu(dx) \right| + \int_{f(t,s)}^{\infty} |g(x - f(t, s_n)) - g(x - f(t, s))| \nu(dx) \\ &\leq \|g\|_{\infty} \int_{f(t,s)}^{f(t,s_n)} \nu(dx) + \int_{f(t,s)}^{\infty} |g(x - f(t, s_n)) - g(x - f(t, s))| \nu(dx) . \end{aligned} \quad (\text{B.3})$$

When rewriting the first term in (B.3), we have used the fact that  $s < s_n$  and that  $f_t(\cdot)$  is decreasing. This first term tends to 0 as  $n \rightarrow +\infty$  because  $\nu(\{f(t, s)\}) = 0$ . By the continuity of both  $f$  and  $g$ ,  $g(x - f(t, s_n)) - g(x - f(t, s)) \rightarrow 0$  as  $n \rightarrow \infty$ . The dominated convergence theorem implies that the second term in (B.3) tends to 0 as well. This proves that  $h$  is right-continuous. A similar argument proves that  $h_t$  is left-continuous, hence continuous at  $s$ .  $\square$

*Proof of ii).* Note that, for all  $t_1 < t_2$ ,

$$A^n(t_2) - A^n(t_1) = A^n(t_2) - A(t_2) - (A^n(t_1) - A(t_1)) + A(t_2) - A(t_1) .$$

Let  $L$  be the Lipschitz constant of the function  $A(\cdot)$ . Using the uniform convergence of  $A^n(\cdot)$  on the interval  $[0, T]$ , there exists  $N_0 \in \mathbb{N}$  such that

$$A^n(t_2) - A^n(t_1) \leq \frac{\varepsilon}{4} + L(t_2 - t_1) \quad (\text{B.4})$$

for all  $0 \leq t_1 \leq t_2 \leq T$  and  $n \geq N_0$ . Let  $\Delta = \frac{\varepsilon}{4} + L(T - T_0)$ . Note that  $A^n(T) - A^n(T_0) \leq \Delta$ . The difficulty in proving (B.1) lies in the fact that  $\nu$  can have atoms. Let  $\mathcal{A} \subset \mathbb{R}_+$  denotes the set of all the atoms of  $\nu$ , which is countable. Let  $\nu^d = \sum_{a \in \mathcal{A}} \nu(\{a\}) \delta_a$  be the Borel measure formed with the atoms of  $\nu$ , and  $\nu^c = \nu - \nu^d$  be the measure that has no atoms. By Lemma A.1 of [15], there exists  $\eta_1 > 0$  such that for all  $\eta \leq \eta_1$  we have

$$\sup_{y \in \mathbb{R}_+} \langle 1_{(y-\eta, y+\eta)}, \nu^c \rangle < \frac{\varepsilon}{4\Delta} . \quad (\text{B.5})$$

Since  $\sum_{a \in \mathcal{A}} \nu(a) \leq 1$ , there exists a finite set  $\mathcal{A}_\varepsilon \subset \mathcal{A}$  such that

$$\sum_{a \in \mathcal{A} \setminus \mathcal{A}_\varepsilon} \nu(\{a\}) \leq \frac{\varepsilon}{4\Delta} . \quad (\text{B.6})$$

Let  $T_0 < T$ ,  $t \in [T_0, T]$ ,  $n \geq N_0$  and  $\eta \leq \eta_1$ . We have:

$$\begin{aligned}
& \int_{T_0}^t \langle \mathbf{1}_{[(f(t,s)-\eta)^+, f(t,s)+\eta]}, \nu \rangle dA^n(s) \\
&= \int_{T_0}^t \langle \mathbf{1}_{[(f(t,s)-\eta)^+, f(t,s)+\eta]}, \nu^c \rangle dA^n(s) + \int_{T_0}^t \langle \mathbf{1}_{[(f(t,s)-\eta)^+, f(t,s)+\eta]}, \nu^d \rangle dA^n(s) \\
&\leq \frac{\varepsilon}{4\Delta} (A^n(t) - A^n(T_0)) + \sum_{a \in \mathcal{A} \setminus \mathcal{A}_\varepsilon} \int_{T_0}^t \langle \mathbf{1}_{[(f(t,s)-\eta)^+, f(t,s)+\eta]}, \delta_a \rangle dA^n(s) \\
&\quad + \sum_{a \in \mathcal{A}_\varepsilon} \int_{T_0}^t \langle \mathbf{1}_{[(f(t,s)-\eta)^+, f(t,s)+\eta]}, \delta_a \rangle dA^n(s) \\
&\leq \left( \frac{\varepsilon}{4\Delta} + \frac{\varepsilon}{4\Delta} \right) \Delta + \sum_{a \in \mathcal{A}_\varepsilon} \nu(\{a\}) \int_{T_0}^t \langle \mathbf{1}_{[(f(t,s)-\eta)^+, f(t,s)+\eta]}, \delta_a \rangle dA^n(s). \tag{B.7}
\end{aligned}$$

In the third line, we have used (B.5) and in the last line we have used the fact that  $A^n(t) - A^n(T_0) \leq \Delta$  and (B.6). If  $\mathcal{A}_\varepsilon = \emptyset$ , let  $\eta_0 = \eta_1$ . The right-hand side in (B.7) is  $\varepsilon/2$ . Otherwise, for any  $0 < \eta \leq \eta_1$ ,  $T_0 \leq s \leq t$  and  $a \in \mathcal{A}_\varepsilon$ , we have

$$a \in [(f(t,s) - \eta)^+, f(t,s) + \eta] \implies s \in [f_t^{-1}(a + \eta), f_t^{-1}((a - \eta)^+)].$$

This implies for  $0 < \eta \leq \eta_1$ ,  $t \in [T_0, T]$ ,  $a \in \mathcal{A}$  and  $n \geq N_0$ , that:

$$\int_{T_0}^t \langle \mathbf{1}_{[(f(t,s)-\eta)^+, f(t,s)+\eta]}, \delta_a \rangle dA^n(s) \leq A^n(f_t^{-1}((a - \eta)^+)) - A^n(f_t^{-1}(a + \eta)). \tag{B.8}$$

Since  $n > N_0$  and  $f_t^{-1}(a + \eta) \leq f_t^{-1}((a - \eta)^+) \leq f_t^{-1}(0) = t \leq T$ , by (B.4) we have

$$A^n(f_t^{-1}((a - \eta)^+)) - A^n(f_t^{-1}(a + \eta)) \leq \frac{\varepsilon}{4} + L(f_t^{-1}((a - \eta)^+) - f_t^{-1}(a + \eta)). \tag{B.9}$$

On the other hand, since  $\nu(\{a\})(f_t^{-1}((a - \eta)^+) - f_t^{-1}(a + \eta)) \leq t\nu(\{a\})$  and  $\sum_{a \in \mathcal{A}_\varepsilon} t\nu(\{a\}) \leq t$ , the function of  $\eta$ :

$$\eta \mapsto \sum_{a \in \mathcal{A}_\varepsilon} \nu(\{a\}) (f_t^{-1}((a - \eta)^+) - f_t^{-1}(a + \eta))$$

is continuous, and there exists  $0 < \eta_0 < \eta_1$  such that

$$\sum_{a \in \mathcal{A}_\varepsilon} \nu(\{a\}) (f_t^{-1}((a - \eta_0)^+) - f_t^{-1}(a + \eta_0)) \leq \frac{\varepsilon}{4L}. \tag{B.10}$$

Finally, using (B.8), (B.9) and (B.10) in (B.7), we have, for  $\eta_0$ ,

$$\int_{T_0}^t \langle \mathbf{1}_{[(f(t,s)-\eta_0)^+, f(t,s)+\eta_0]}, \nu \rangle dA^n(s) \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{4} + \frac{L\varepsilon}{4L} \leq \varepsilon.$$

The inequality (B.1) is therefore proved. This concludes the proof of *ii*).  $\square$

*Proof of iii*). For each  $n \geq 0$ , the function  $h_n(\cdot)$  is continuous on each interval  $[0, T]$ . Indeed,  $s \mapsto A^n(s)$  is a continuous and increasing function, so that it is differentiable almost everywhere, and  $h_n$  can be written as

$$h_n(t) = \int_0^T \langle g(\cdot - f(t,s)), \nu \rangle \mathbf{1}_{[0,t]}(s) \dot{A}^n(s) ds$$

for all  $t \leq T$ . On the other hand, the function

$$\begin{aligned} \lambda : [0, T] \times [0, T] &\rightarrow \mathbb{R}^+ \\ (t, s) &\mapsto \langle g(\cdot - f(t, s)), \nu \rangle 1_{[0, t]} \dot{A}^n(s) \end{aligned}$$

has the following properties. First, for each  $T \geq t \geq 0$  the function  $s \mapsto \lambda(t, s)$  is bounded by  $\|g\|_\infty \dot{A}^n(s)$  which is integrable independently of  $t$ . Second, for each  $T \geq s \geq 0$  the function  $t \mapsto \lambda(t, s)$  is continuous except for a denumerable set of  $t \leq T$ .

Hence by Lebesgue's theorem the function  $h_n$  is continuous on  $[0, T]$ . To prove the equicontinuity of  $(h_n(\cdot), n \geq 0)$  on finite intervals, it suffices to show that, for every  $T > 0$  and  $\varepsilon > 0$ , there exists  $\eta > 0$  and  $N_0 \in \mathbb{N}$  such that if  $0 \leq t_1, t_2 \leq T$  and  $|t_2 - t_1| < \eta$  then

$$|h_n(t_2) - h_n(t_1)| < \varepsilon$$

for all  $n \geq N_0$ . Note that for each  $0 \leq t_1, t_2 \leq T$ ,

$$\begin{aligned} &|h_n(t_2) - h_n(t_1)| \\ &= \left| \int_{t_1}^{t_2} \langle g(\cdot - f(t_2, s)), \nu \rangle dA^n(s) + \int_0^{t_1} \langle g(\cdot - f(t_2, s)) - g(\cdot - f(t_1, s)), \nu \rangle dA^n(s) \right| \\ &\leq \|g\|_\infty |A^n(t_2) - A^n(t_1)| + \int_0^{t_1} \langle |g(\cdot - f(t_2, s)) - g(\cdot - f(t_1, s))|, \nu \rangle dA^n(s). \end{aligned} \tag{B.11}$$

Let  $\varepsilon > 0$ . By the uniform convergence of  $A^n(\cdot)$  on  $[0, T]$ , there exists  $N_0 \in \mathbb{N}$  such that

$$|A^n(t_2) - A^n(t_1)| \leq \frac{\varepsilon}{16(\|g\|_\infty \vee 1)} + L|t_2 - t_1|. \tag{B.12}$$

Let  $T_1 = \frac{\varepsilon}{16L(\|g\|_\infty \vee 1)}$ . Inequality (B.12) implies, for  $t_2 = T_1$  and  $t_1 = 0$

$$A^n(T_1) \leq \frac{\varepsilon}{16(\|g\|_\infty \vee 1)} + LT_1 \leq \frac{\varepsilon}{8(\|g\|_\infty \vee 1)}, \tag{B.13}$$

For all  $n \geq N_0$ . Thus for  $t_1 \in [0, T_1]$  and  $n \geq N_0$ , using the fact that  $A^n(0) = 0$ ,

$$\begin{aligned} \int_0^{t_1} \langle |g(\cdot - f(t_2, s)) - g(\cdot - f(t_1, s))|, \nu \rangle dA^n(s) &\leq 2\|g\|_\infty A^n(t_1) \\ &\leq 2\|g\|_\infty A^n(T_1) \\ &\leq \frac{\varepsilon}{4}. \end{aligned} \tag{B.14}$$

The second inequality follows from the increasingness of  $A^n(\cdot)$ , and the third one uses (B.13). Then by (B.11), (B.12) and (B.14) we have

$$|t_2 - t_1| \leq \frac{11\varepsilon}{16L(\|g\|_\infty \vee 1)} \implies |h_n(t_2) - h_n(t_1)| \leq \varepsilon.$$

Since in equation (B.11) the variables  $t_1, t_2$  play a symmetrical role, the same estimate as above is derived when  $t_2 \in [0, T_1]$  and  $t_1 \in [0, T]$ .

There remains to consider the case where  $t_1, t_2 \in [T_1, T]$ . Let therefore be two such real numbers with  $|t_2 - t_1| \leq \frac{3\varepsilon}{16L(\|g\|_\infty \vee 1)}$ . We have

$$\begin{aligned} & |h_n(t_2) - h_n(t_1)| \\ & \leq \|g\|_\infty |A^n(t_2) - A^n(t_1)| + \int_0^{T_1} \langle |g(\cdot - f(t_2, s)) - g(\cdot - f(t_1, s))|, \nu \rangle dA^n(s) \\ & \quad + \int_{T_1}^{t_1} \langle |g(\cdot - f(t_2, s)) - g(\cdot - f(t_1, s))|, \nu \rangle dA^n(s) \\ & \leq \frac{\varepsilon}{2} + \int_{T_1}^{t_1} \langle |g(\cdot - f(t_2, s)) - g(\cdot - f(t_1, s))|, \nu \rangle dA^n(s) \end{aligned} \quad (\text{B.15})$$

for all  $n \geq N_0$ . We have used (B.11) in the first inequality, and equations (B.12) and (B.14) in the last inequality.

In order to conclude on the equicontinuity of  $h_n$ , it suffices to show that there exists  $\eta_0 > 0$  such that if  $t_1, t_2 \in [T_1, T]$  are such that  $|t_2 - t_1| < \eta_0$ , then

$$\int_{T_1}^{t_1} \langle |g(\cdot - f(t_2, s)) - g(\cdot - f(t_1, s))|, \nu \rangle dA^n(s) \leq \frac{\varepsilon}{2} \quad (\text{B.16})$$

for all  $n \geq N_0$ . An estimate for the integrand in (B.16) can be derived as follows. We first introduce the shorthand notation  $m_1(s) := f(t_1, s)$  and  $m_2(s) := f(t_2, s)$ , where we shall omit the argument  $s$  when no ambiguity occurs. Let  $\Delta = \frac{\varepsilon}{16(\|g\|_\infty \vee 1)} + L(T - T_1)$ . From (B.12),  $A^n(T) - A^n(T_1) \leq \Delta$  for all  $n \geq N_0$ . Next, let  $M > 0$  be such that

$$\langle 1_{[M, \infty)}, \nu \rangle \leq \frac{\varepsilon}{8\Delta(\|g\|_\infty \vee 1)}. \quad (\text{B.17})$$

Such an  $M$  exists, since  $\nu$  is a proper probability measure. Therefore, for every fixed  $s$ , the integral:

$$\langle |g(\cdot - m_2) - g(\cdot - m_1)|, \nu \rangle = \int_0^\infty |g(x - m_2) - g(x - m_1)| \nu(dx)$$

can be decomposed according to the intervals  $[0, m_1 \wedge m_2)$ ,  $[m_1 \wedge m_2, m_1 \vee m_2)$ ,  $[m_1 \vee m_2, M]$  and  $(M, +\infty)$ . Using the fact that  $g(\cdot)$  is zero on negative half line, we obtain the bound:

$$\begin{aligned} \langle |g(\cdot - m_2) - g(\cdot - m_1)|, \nu \rangle & \leq \|g\|_\infty \langle 1_{[m_1 \wedge m_2, m_1 \vee m_2)}, \nu \rangle \\ & \quad + \int_{m_1 \vee m_2}^M |g(x - m_2) - g(x - m_1)| \nu(dx) \\ & \quad + 2\|g\|_\infty \langle 1_{[M, \infty)}, \nu \rangle. \end{aligned} \quad (\text{B.18})$$

Since  $g(\cdot)$  is continuous, it is uniformly continuous on  $[0, M]$  and there exists  $\delta > 0$  such that

$$|g(x) - g(y)| \leq \frac{\varepsilon}{4\Delta} \quad (\text{B.19})$$

for all  $x, y \in [0, M]$  verifying  $|x - y| \leq \delta$ . On the other hand the function  $f(\cdot, \cdot)$  is continuous on  $(0, \infty) \times (0, \infty)$ , it is uniformly continuous on  $[T_1, T] \times [T_1, T]$  then for each  $\delta' \in (0, \delta]$  there exists  $\eta(\delta') > 0$  such that if  $(t_1, s_1), (t_2, s_2) \in [T_1, T] \times [T_1, T]$  and  $|t_2 - t_1| + |s_2 - s_1| \leq \eta(\delta')$  then  $|f(t_1, s_1) - f(t_2, s_2)| < \delta'$ . In particular for all  $t_1, t_2 \in [T_1, T]$  such that  $|t_2 - t_1| \leq \eta(\delta')$ ,

$$\sup_{T_1 \leq s \leq T} |m_1(s) - m_2(s)| = \sup_{T_1 \leq s \leq T} |f(t_1, s) - f(t_2, s)| < \delta'. \quad (\text{B.20})$$

Thus for all  $s \in [T_1, T]$ ,  $x \in [f(t_1, s) \wedge f(t_2, s), M]$  and  $t_1, t_2 \in [T_1, T]$  such that  $|t_2 - t_1| < \eta(\delta')$ . So, by (B.20) we have  $|(x - f(t_2, s)) - (x - f(t_1, s))| = |f(t_2, s) - f(t_1, s)| \leq \delta'$  and by (B.19), we have

$$|g(x - m_2) - g(x - m_1)| \leq \frac{\varepsilon}{4\Delta}. \quad (\text{B.21})$$

Note that (B.20) implies  $(f(t_1, s) - \delta')^+ \leq f(t_2, s) \leq f(t_1, s) + \delta'$ . Consequently we have

$$(f(t_1, s) - \delta')^+ \leq f(t_1, s) \wedge f(t_2, s) \leq f(t_1, s) \vee f(t_2, s) \leq f(t_1, s) + \delta'. \quad (\text{B.22})$$

We combine the estimates (B.17), (B.21) and (B.22) in (B.18) to obtain the following bound:

$$\langle |g(\cdot - f(t_2, s)) - g(\cdot - f(t_1, s))|, \nu \rangle \leq \|g\|_\infty \langle \mathbf{1}_{[(f(t_1, s) - \delta')^+, f(t_1, s) + \delta')}, \nu \rangle + \frac{\varepsilon}{2\Delta}. \quad (\text{B.23})$$

Coming back to the left-hand side of (B.16), we conclude that for all  $\delta' \in (0, \delta]$ ,  $t_1, t_2 \in [T_1, T]$  such that  $|t_2 - t_1| \leq \eta(\delta')$  and  $n \geq N_0$

$$\begin{aligned} & \int_{T_1}^{t_1} \langle |g(\cdot - f(t_2, s)) - g(\cdot - f(t_1, s))|, \nu \rangle \, dA^n(s) \\ & \leq \frac{\varepsilon}{2\Delta} (A^n(t_1) - A^n(T_1)) + \|g\|_\infty \int_{T_1}^{t_1} \langle \mathbf{1}_{[(f(t_1, s) - \delta')^+, f(t_1, s) + \delta')}, \nu \rangle \, dA^n(s) \\ & \leq \frac{\varepsilon}{2} + \|g\|_\infty \int_{T_1}^{t_1} \langle \mathbf{1}_{[(f(t_1, s) - \delta')^+, f(t_1, s) + \delta')}, \nu \rangle \, dA^n(s), \end{aligned} \quad (\text{B.24})$$

where the second inequality uses the fact that  $A^n(T) - A^n(T_1) \leq \Delta$  for all  $n \geq N_0$  and  $t_1 \leq T$ . Replacing  $\varepsilon$  by  $\varepsilon/2 \|g\|_\infty$  in *ii*), there exists  $\delta_0 > 0$  and  $N_0 \in \mathbb{N}$  such that

$$\int_{T_1}^{t_1} \langle \mathbf{1}_{[(f(t_1, s) - \delta_0)^+, f(t_1, s) + \delta_0)}, \nu \rangle \, dA^n(s) \leq \frac{\varepsilon}{2 \|g\|_\infty}$$

for all  $n \geq N_0$ . Therefore, if  $\delta_0 \geq \delta$  then the second term of right-hand side of (B.24) is less than  $\varepsilon/2$  in this case it suffices to take  $\eta_0 = \eta(\delta)$  and (B.16) holds. Otherwise there exists  $\eta(\delta_0)$  such that (B.24) is verified, consequently (B.16) holds for  $\eta_0 = \eta(\delta_0)$ .  $\square$

*Proof of iv).* Fix  $T > 0$ . Using a double difference, we have:

$$\sup_{0 \leq t \leq T} \left| \int_0^t \langle g(\cdot - f^n(t, s)), \nu \rangle \, dA^n(s) - \int_0^t \langle g(\cdot - f(t, s)), \nu \rangle \, dA(s) \right| \quad (\text{B.25})$$

$$\leq \sup_{0 \leq t \leq T} \left| \int_0^t \langle g(\cdot - f(t, s)), \nu \rangle \, dA^n(s) - \int_0^t \langle g(\cdot - f(t, s)), \nu \rangle \, dA(s) \right| \quad (\text{B.26})$$

$$+ \sup_{0 \leq t \leq T} \int_0^t \langle |g(\cdot - f^n(t, s)) - g(\cdot - f(t, s))|, \nu \rangle \, dA^n(s). \quad (\text{B.27})$$

For the term (B.26), we use the following reasoning. Let  $0 < t \leq T$ . Choose  $N > 0$  such that  $A^n(t) > 0$  for all  $n \geq N$ . This  $N$  exists because  $A^n(t) \rightarrow A(t)$ . Define for all  $n \geq N$ ,  $F_n(s) = A_n(s)/A_n(t)$  if  $s < t$  and  $F_n(s) = 1$  if  $t \leq s$ . It is clear that  $\{F_n(\cdot); n \geq N\}$  is a sequence of distribution functions that satisfies  $F_n(\cdot) \Rightarrow F(\cdot)$ , where  $F(s) = A(s)/A(t)$  if  $s < t$  and  $F(s) = 1$  if  $s \geq t$ . From *i*), the function  $s \mapsto \langle g(\cdot - f(s)), \nu \rangle$  is bounded and continuous, except for countably many values of  $s \in (0, t]$ . Then, by the continuous mapping theorem (*cf.* [10, Theorem 2.3, Chapter 2]):

$$\lim_{n \rightarrow \infty} \int_0^t \langle g(\cdot - f(t, s)), \nu \rangle \, dF^n(s) = \int_0^t \langle g(\cdot - f(t, s)), \nu \rangle \, dF(s).$$

Then, replacing  $F_n$  and  $F$  by their expressions and using the fact that  $A^n(t) \rightarrow A(t)$  one deduces

$$\int_0^t \langle g(\cdot - f(t, s)), \nu \rangle dA^n(s) \rightarrow \int_0^t \langle g(\cdot - f(t, s)), \nu \rangle dA(s).$$

Then by *iii*) the uniform convergence of the above limit holds on every finite interval, thus (B.26) tends to 0. For the term (B.27), we adopt the following strategy which is similar to that of the proof *ii*). First, we isolate and bound the integral when  $s$  is close to 0. Indeed, we have not specified the behavior of the function  $f(t, s)$  when  $s \rightarrow 0$ , and the difference  $|f_n(t, s) - f(t, s)|$  is not necessarily uniformly bounded for  $(t, s) \in [0, T] \times [0, T]$ . Next, we eliminate the unbounded part of the integral with respect to the measure  $\nu$ , in order to reduce the integral to a compact. Finally, we bound the difference on this compact. The objective is therefore to prove that, for each given  $T > 0$ ,  $g \in C_b(\mathbb{R}_+)$  and  $\varepsilon > 0$ , there exists  $N > 0$  such that for all  $n \geq N$ ,

$$\sup_{0 \leq t \leq T} \int_0^t \langle |g(\cdot - f^n(t, s)) - g(\cdot - f(t, s))|, \nu \rangle dA^n(s) < \varepsilon. \quad (\text{B.28})$$

Observe first that the steps (B.12)–(B.14) of the proof of *iii*) do not depend on the nature of the shift inside  $g$ , and are still valid here. For each  $\varepsilon > 0$ , let therefore  $T_1 = \frac{\varepsilon}{16L(\|g\|_\infty \vee 1)}$ : for all  $t \in [0, T_1]$  and  $n \geq N_0$  we have:

$$\int_0^t \langle |g(\cdot - f^n(t, s)) - g(\cdot - f(t, s))|, \nu \rangle dA^n(s) \leq \frac{\varepsilon}{4}.$$

The next step is to bound the part of the integral in (B.28) for the range  $t \in [T_1, T]$ . In order to obtain (B.28) it suffices to find some  $N_1 > 0$  such that for  $n \geq N_1$ :

$$\int_{T_1}^t \langle |g(\cdot - f^n(t, s)) - g(\cdot - f(t, s))|, \nu \rangle dA^n(s) \leq \frac{3\varepsilon}{4}. \quad (\text{B.29})$$

Set  $m_1(s) := f^n(t, s)$  and  $m_2(s) := f(t, s)$ . The arguments of the proof of *iii*), between (B.17) and (B.18) still apply. Set therefore  $\Delta = \frac{\varepsilon}{16(\|g\|_\infty \vee 1)} + L(T - T_1)$ , and fix  $M > 0$  be such that (B.17) holds. We then (B.18) holds as well. Since  $f^n(\cdot, \cdot) \rightarrow f(\cdot, \cdot)$  uniformly on  $[T_1, T] \times [T_1, T]$ , for each  $\delta' \leq \delta$  there exists  $N_1(\delta') > 0$  such that

$$\sup_{T_1 \leq s \leq T} |m_1(s) - m_2(s)| \leq \sup_{T_1 \leq t, s \leq T} |f^n(t, s) - f(t, s)| \leq \delta'. \quad (\text{B.30})$$

This bound implies:  $|(x - m_1(s)) - (x - m_2(s))| = |m_1(s) - m_2(s)| \leq \delta' \leq \delta$ , for all  $T_1 \leq t, s \leq T$ ,  $f^n(t, s) \vee f(t, s) \leq x \leq M$  and  $n \geq N_1(\delta')$ . Then by (B.19), we have

$$|g(x - f^n(t, s)) - g(x - f(t, s))| \leq \frac{\varepsilon}{4\Delta} \quad (\text{B.31})$$

for all  $T_1 \leq t, s \leq T$ ,  $f^n(t, s) \vee f(t, s) \leq x \leq M$  and  $n \geq N_1(\delta')$ . From (B.30) we have  $(f(t, s) - \delta')^+ \leq f^n(t, s) \leq f(t, s) + \delta'$ . So,

$$(f(t, s) - \delta')^+ \leq f^n(t, s) \wedge f(t, s) \leq f^n(t, s) \vee f(t, s) \leq f(t, s) + \delta'. \quad (\text{B.32})$$

Coming back to (B.18), we have

$$\begin{aligned} \langle |g(\cdot - f^n(t, s)) - g(\cdot - f(t, s))|, \nu \rangle &\leq \|g\|_\infty \langle 1_{[(f(t, s) - \delta')^+, f(t, s) + \delta']}, \nu \rangle \\ &\quad + \frac{\varepsilon}{4\Delta} + 2 \|g\|_\infty \frac{\varepsilon}{8\Delta(\|g\|_\infty \vee 1)} \\ &\leq \|g\|_\infty \langle 1_{[(f(t, s) - \delta')^+, f(t, s) + \delta']}, \nu \rangle + \frac{\varepsilon}{2\Delta} \end{aligned} \quad (\text{B.33})$$

for all  $T_1 \leq t, s \leq T$  and  $n \geq N_1(\delta')$ . The first inequality is by (B.32), (B.31) and (B.17). Then the estimate of the integral in (B.29) is obtained by integrating (B.33) on  $[T_1, t]$  with respect to  $dA^n(s)$ ,

$$\begin{aligned} & \int_{T_1}^t \langle |g(\cdot - f^n(t, s)) - g(\cdot - f(t, s))|, \nu \rangle dA^n(s) \\ & \leq \frac{\varepsilon}{2\Delta} (A^n(t) - A^n(T_1)) + \|g\|_\infty \int_{T_1}^t \langle \mathbf{1}_{[(f(t,s)-\delta')^+, f(t,s)+\delta')]} \nu \rangle dA^n(s) \\ & \leq \frac{\varepsilon}{2} + \|g\|_\infty \int_{T_1}^t \langle \mathbf{1}_{[(f(t,s)-\delta')^+, f(t,s)+\delta')]} \nu \rangle dA^n(s) \end{aligned} \quad (\text{B.34})$$

for all  $T_1 \leq t \leq T$  and  $n \geq N_1(\delta') \vee N_0$ . The estimate of the last term in (B.34) is obtained as that of the last term in (B.24) by using *ii*).  $\square$

## C Proof of Proposition 6.3

To prove Proposition 6.3 it suffices to prove the following result:

**Proposition C.1.** *Let  $k_0, k \in \mathcal{K}$  be such that  $Q_{kk_0} \alpha_{k_0} \neq 0$ . Consider a sequence of multiclass processor sharing queues as defined in Section 4.4, satisfying assumptions (4.12)–(4.22). The sequence of fluid scaled processes  $\bar{\gamma}_{k_0 k}^r(\cdot)$  is tight. Moreover, any limit  $\bar{\gamma}_{k_0 k}^*(\cdot)$  is a fluid solution of Equation (3.30).*

This section is organized as follows. In section C.1, we prove the convergence in particular cases where no elaborate technique is necessary. Tightness of  $\bar{\gamma}_{k_0 k}^r(\cdot)$  is proved in Section C.2. In Section C.3, we state the characterization of the fluid limit which allows to complete the proof. Recall the convention that any function defined on  $\mathbb{R}_+$ , is extended to be identically equal to zero on  $(-\infty, 0)$  so that for all  $a > 0$ , the function  $g(\cdot - a)$  is well defined on  $\mathbb{R}$ . The basic equation used in the proof of tightness and the characterization of the fluid limit is known as the dynamic equation. Let  $g : \mathbb{R}_+ \rightarrow \mathbb{R}$  be a Borel-measurable function and  $t, h \geq 0$ . Integrating the functions  $g$  and  $(\mathbf{1}_{(0, \infty)} g)(\cdot - \bar{S}_{t, t+h}^r)$  with respect to measures  $\bar{\gamma}_{k_0 k}^r(t+h)$  and  $\bar{\gamma}_{k_0 k}^r(t)$  respectively, then subtracting the integral  $\langle (\mathbf{1}_{(0, \infty)} g)(\cdot - \bar{S}_{t, t+h}^r), \bar{\gamma}_{k_0 k}^r(t) \rangle$  from  $\langle g, \bar{\gamma}_{k_0 k}^r(t+h) \rangle$  gives,

$$\begin{aligned} \langle g, \bar{\gamma}_{k_0 k}^r(t+h) \rangle &= \langle (\mathbf{1}_{(0, \infty)} g)(\cdot - \bar{S}_{t, t+h}^r), \bar{\gamma}_{k_0 k}^r(t) \rangle \\ &+ \frac{1}{r} \sum_{i=r\bar{E}_{k_0}^r(t)+1}^{r\bar{E}_{k_0}^r(t+h)} \sum_{n=1}^{N_{k_0 k}^r(i)} (\mathbf{1}_{(0, \infty)} g) \left( V_{k_0 k}^r(i, n) - \bar{S}_{U_{k_0}^r(i)/r, t+h}^r \right). \end{aligned}$$

### C.1 Cases without input

**Lemma C.1.** *Assume that the assumptions (4.12)–(4.22) hold,  $\rho \leq 1$  and the initial state is zero i.e  $\bar{\gamma}(0) = 0$ . Then,*

$$\bar{\gamma}_{k_0 k}^r(\cdot) \Rightarrow 0.$$

*Proof.* Recall that for each  $k, l \in \mathcal{K}$ ,  $\Phi_k^{r,l}(N)$  is the number of customers among  $N$  which moves from the class  $l$  towards the class  $k$ . Then the corresponding between  $Z(t)$  and  $Q(t)$  is,

$$\bar{Z}_k^r(t) = \bar{Q}_k^r(t) - \sum_{l=1}^K \bar{\Phi}_k^{r,l}(\bar{Q}_l^r(t)). \quad (\text{C.1})$$

By using Equations (3.20)-(3.22), we have

$$\bar{Q}_k^r(t) \leq \sum_{k_0=1}^K \frac{1}{r} \left( \sum_{i=1}^{r\bar{Z}_{k_0}^r(0)} N_{k_0k}^r(i) + \sum_{i=1}^{r\bar{E}_{k_0}^r(t)} N_{k_0k}^r(i) \right).$$

Then  $\limsup_{r \rightarrow \infty} \bar{Q}_k^r(t)$  exists and is finite in probability. Thanks to (C.1)

$$\limsup_{r \rightarrow \infty} \bar{Q}_k^r(t) \leq \limsup_{r \rightarrow \infty} \bar{Z}_k^r(t) + \sum_{l=1}^K p_{lk} \limsup_{r \rightarrow \infty} \bar{Q}_l^r(t).$$

This is verified for all  $t \geq 0$  and  $k \in \mathcal{K}$ , since  $\rho(P) < 1$ , we have

$$\limsup_{r \rightarrow \infty} \bar{Q}^r(t) \leq Q \limsup_{r \rightarrow \infty} \bar{Z}^r(t).$$

Proposition 6.2 implies that  $\limsup_{r \rightarrow \infty} \bar{Z}_k^r(t) = 0$  in probability for all  $k \in \mathcal{K}$ , and  $t \geq 0$ , then  $\lim_{r \rightarrow \infty} \bar{Q}^r(t) = 0$  in probability. Because,

$$\langle g, \bar{\gamma}_{k_0k}^r(t) \rangle \leq \|g\|_\infty \bar{Q}_k^r(t)$$

for all bounded measurable function  $g : \mathbb{R}_+ \rightarrow \mathbb{R}$ , this proves Lemma C.1.  $\square$

**Lemma C.2.** *Suppose that there are  $k_0, k \in \mathcal{K}$  such that  $Q_{kk_0} \alpha_{k_0} = 0$  and assumptions (4.12)-(4.22). Then, for each Borel-measurable function  $g : \mathbb{R}_+ \rightarrow \mathbb{R}$*

$$\langle g, \bar{\gamma}^r(t) \rangle \Rightarrow \langle 1_{(0,\infty)} g(\cdot - \bar{S}(t)), \bar{\gamma}(0) \rangle.$$

*Proof.* From the dynamic equation we have,

$$\begin{aligned} \langle g, \bar{\gamma}_{k_0k}^r(t) \rangle &= \langle (1_{(0,\infty)} g)(\cdot - \bar{S}_{0,t}^r), \bar{\gamma}_{k_0k}^r(0) \rangle \\ &\quad + \frac{1}{r} \sum_{i=1}^{r\bar{E}_{k_0}^r(t)} \sum_{n=1}^{N_{k_0k}^r(i)} (1_{(0,\infty)} g) \left( V_{k_0k}^r(i, n) - \bar{S}_{U_{k_0}^r(i)/r, t}^r \right). \end{aligned} \quad (\text{C.2})$$

Here,

$$\bar{\gamma}_{k_0k}^r(0) = \frac{1}{r} \sum_{i=1}^{r\bar{Z}_{k_0}^r(0)} \sum_{n=1}^{N_{k_0k}^r(i)} \delta_{V_{k_0k}^r(i, n)}.$$

As in Lemma C.3, we have

$$\langle g, \bar{\gamma}_{k_0k}^r(0) \rangle \Rightarrow \langle g, \bar{\gamma}_{k_0k}(0) \rangle, \quad (\text{C.3})$$

with  $\bar{\gamma}_{k_0k}(0) = \mathcal{B} * \nu^0 \bar{Z}_{k_0}(0)$ . The second term on the right-hand side of (C.2) is less than,

$$w^r(t) := \|g\|_\infty \frac{1}{r} \sum_{i=1}^{r\bar{E}_{k_0}^r(t)} N_{k_0k}^r(i).$$

By the law of large number and random time change formula,  $w^r(t) \Rightarrow \|g\|_\infty Q_{kk_0} t \alpha_{k_0}$  which is 0 by assumption. Then the second term on the right-hand side of (C.2) tends to 0 in distribution. On the other hand, observe that, if the initial state is zero in fluid limit that is  $\bar{\gamma}(0) = 0$  then the limit in (C.3) is zero. So  $\langle g, \bar{\gamma}^r(t) \rangle \Rightarrow 0$ . Otherwise, in that case, we suppose  $\langle 1, \bar{\gamma}(0) \rangle > 0$  and  $t < t^*$ . By Proposition 6.2, we have  $\bar{S}_{0,t}^r \Rightarrow \bar{S}_{0,t}$ , this jointly with (C.3) proves that the first term of the right-hand side of (C.2) converges to,

$$\langle 1_{(0,\infty)} g(\cdot - \bar{S}(t)), \bar{\gamma}(0) \rangle.$$

This proves the lemma.  $\square$

## C.2 Proof of tightness

In this section we prove the first half of Proposition C.1 that is for each  $T > 0$  the sequence of measure valued process  $\{\bar{\gamma}_{k_0k}^r(\cdot)\}_{r>0}$  is tight in  $D([0, T], \mathcal{M})$ . By Jakubowski's criterion [9], it suffices to show that this process satisfies conditions C1 and C2 instead of  $\{\bar{\mu}_k^r(\cdot)\}_{r>0}$ .

As a consequence of Lemmas C.2 and C.1, we assume that  $Q_{kk_0}\alpha_{k_0} \neq 0$  for  $k_0, k \in \mathcal{K}$  and we consider the additional case: either (a)  $\langle 1, \bar{\gamma}(0) \rangle > 0$ , or (b)  $\langle 1, \bar{\gamma}(0) \rangle = 0$  and  $\rho > 1$ . The same strategy discussed in Section 6.2 is adapted here, the difference is that the constant  $s$  is defined by

$$s = (4Q_{kk_0}\alpha_{k_0})^{-1}. \quad (\text{C.4})$$

### C.2.1 Law of large numbers

For each  $k_0, k \in \mathcal{K}$  and for each  $r > 0$ , recall that  $N_{k_0k}^r(i)$  is the number of visits to class  $k$  made by  $i^{\text{th}}$  customer that starts service in class  $k_0$ , and  $V_{k_0k}^r(i, n)$  its required service times for the  $n^{\text{th}}$  visit.

**Lemma C.3.** *Assume that assumptions (4.13)-(4.16) hold. For each  $g: \mathbb{R} \rightarrow \mathbb{R}$  bounded, measurable and  $\nu_k$ -continuous for all  $k \in \mathcal{K}$ ,*

$$\frac{1}{r} \sum_{i=1}^{r\bar{E}_{k_0}^r(t)} \sum_{n=1}^{N_{k_0k}^r(i)} g(V_{k_0k}^r(i, n)) \Rightarrow \langle g, (\mathcal{B} * \nu)_{kk_0} \rangle \alpha_{k_0} t.$$

*Proof.* Denote by  $X_{k_0k}^r(i) = \sum_{n=1}^{N_{k_0k}^r(i)} g(V_{k_0k}^r(i, n))$ . By Lemma 3.4, the sequence  $(X_{k_0k}^r(i), i \geq 1)$  is i.i.d. such that

$$\mathbb{E}_{k_0k}^r(X_{k_0k}^r(1)) = \langle g, \mathcal{B}^r * \nu^r \rangle_{kk_0}.$$

Conditions (4.13)-(4.16) implies  $(\mathcal{B}^r * \nu^r)_{kk_0} \Rightarrow (\mathcal{B} * \nu)_{kk_0}$ . Since  $g$  is  $\nu_k$ -continuous for all  $k \in \mathcal{K}$ , then it is  $(\mathcal{B} * \nu)_{kk_0}$ -continuous. So by the Portementeau theorem,

$$\mathbb{E}_{k_0k}^r(X_{k_0k}^r(1)) \rightarrow \langle g, (\mathcal{B} * \nu)_{kk_0} \rangle. \quad (\text{C.5})$$

Since  $g$  is bounded we have  $X_{k_0k}^r(1) \leq \|g\|_\infty N_{k_0k}^r(1)$ , and

$$\mathbb{E}_{k_0k}^r(X_{k_0k}^r(1)1_{\{X_{k_0k}^r(1) > r\}}) \leq \|g\|_\infty \mathbb{E}_{k_0k}^r(N_{k_0k}^r(1)1_{\{N_{k_0k}^r(1) > \frac{r}{\|g\|_\infty}\}}).$$

Lemma 3.3 implies that  $\lim_{r \rightarrow \infty} \mathbb{E}_{k_0k}^r(N_{k_0k}^r(1)1_{\{N_{k_0k}^r(1) > \frac{r}{\|g\|_\infty}\}}) = 0$ . Thanks to the last inequality, we have

$$\lim_{r \rightarrow \infty} \mathbb{E}_{k_0k}^r(X_{k_0k}^r(1)1_{\{X_{k_0k}^r(1) > r\}}) = 0.$$

This involves by Equation (C.5)

$$\lim_{r \rightarrow \infty} \mathbb{P}^r(X_{k_0k}^r(1)1_{\{X_{k_0k}^r(1) \leq r\}}) = \langle g, (\mathcal{B} * \nu)_{kk_0} \rangle.$$

Then the necessary conditions for weak Law of Large Number for Triangular Arrays ([10] Theorem 5.5 and Theorem 5.6) are verified. So,

$$\frac{1}{r} \sum_{i=1}^r X_{k_0k}^r(i) \Rightarrow \langle g, (\mathcal{B} * \nu)_{kk_0} \rangle.$$

The random time change formula [3] implies the convergence in Lemma C.3. □

### C.2.2 Weak estimates

In the following the constant  $s$  is extended to 0 in case (a). Estimates are prepared in the following lemma, the proofs use the law of large number of Lemma C.3 and the assumptions on primitive and initial data.

**Lemma C.4.** *Let  $k_0, k \in \mathcal{K}$  be such that  $Q_{kk_0}\alpha_{k_0} \neq 0$ . Fix  $s < T < t_\rho$ , where  $s$  is given by (C.4). For each  $0 < \varepsilon, \eta < 1$ , there exist strictly positive constants  $l, M_0, M_T, \Gamma, \kappa, r_0$ , and events  $(B^r, r > 0)$  such that  $\mathbb{P}(B^r) \geq 1 - \eta$  for all  $r > r_0$ , and on  $B^r$  the following hold,*

$$l \leq \frac{\varepsilon}{16Q_{kk_0}\alpha_{k_0}} \quad (\text{C.6})$$

$$\sup_{[0, T-l]} \left( \frac{1}{r} \sum_{i=r\bar{E}_{k_0}^r(t)+1}^{r\bar{E}_{k_0}^r(t+l)} N_{k_0k}^r(i) \right) \leq \frac{\varepsilon}{4} \quad (\text{C.7})$$

$$\frac{1}{r} \sum_{i=1}^{r\bar{E}_{k_0}^r(t)} N_{k_0k}^r(i) \leq 2Q_{kk_0}\alpha_{k_0}t \quad (\text{C.8})$$

$$\sup_{t \in [0, T]} \langle 1, \bar{\gamma}^r(t) \rangle \leq M_T \quad (\text{C.9})$$

$$\inf_{t \in [s\varepsilon/2, T]} \langle 1, \bar{\gamma}^r(t) \rangle > \frac{1}{\Gamma} \quad (\text{C.10})$$

$$\langle 1, \bar{\gamma}_{k_0k}^r(0) \rangle < M_s \quad (\text{C.11})$$

$$\kappa < \frac{l}{2M_T} \quad (\text{C.12})$$

$$\sup_{x \in \mathbb{R}_+} \langle 1_{[x, x+\kappa]}, \bar{\gamma}_{k_0k}^r(0) \rangle < \frac{\varepsilon}{4}. \quad (\text{C.13})$$

For  $N = \lceil T\Gamma/\kappa \rceil$ , and  $m = 0, \dots, N$

$$\begin{aligned} & \sup_{[0, T-l]} \frac{1}{r} \sum_{i=r\bar{E}_{k_0}^r(t)+1}^{r\bar{E}_{k_0}^r(t+l)} \sum_{n=1}^{N_{k_0k}^r(i)} \mathbf{1}_{[m\kappa, (m+1)\kappa)}(V_{k_0k}^r(i, n)) \\ & < \frac{\varepsilon}{8Q_{kk_0}} \langle \mathbf{1}_{[(m-\frac{1}{2})\kappa, (m+\frac{2}{3})\kappa)}, (\mathcal{B} * \nu)_{kk_0} \rangle \end{aligned} \quad (\text{C.14})$$

where the initial measure of  $\bar{\gamma}_{k_0k}^r(\cdot)$  in the fluid limit is  $\bar{\gamma}_{k_0k}(0) = \mathcal{B} * \nu^0 \bar{Z}_{k_0}(0)$ , and the constant  $M_s$  in (C.11) is defined as:  $M_s = M_0$  if  $\langle 1, \bar{\gamma}_{k_0k}(0) \rangle > 0$ , and  $M_s = \varepsilon/2$  if  $\langle 1, \bar{\gamma}_{k_0k}(0) \rangle = 0$ .

The next lemma gives, on  $B^r$ , an upper bound for the amount of mass that  $\bar{\gamma}_{k_0k}^r(t)$  can have concentrated near zero for  $t \in [0, T]$ . The proof is adapted of that of Lemma 5.5 in [15].

**Lemma C.5.** *Let  $k_0, k \in \mathcal{K}$  be such that  $Q_{kk_0}\alpha_0 \neq 0$ . Let  $s < T < t^*$  and  $\eta, \varepsilon < 1$  be given. Let  $l, M_0, M_T, \Gamma, \kappa, r_0$  be the constants and  $(B^r, r > 0)$  be the events given by Lemma C.4. Then on  $B^r$  for  $r > r_0$*

$$\sup_{t \in [s\varepsilon/2, T]} \langle \mathbf{1}_{[0, \kappa]}, \bar{\gamma}_{k_0k}^r(t) \rangle \leq \frac{\varepsilon}{2}$$

where  $s$  is given by (C.4).

*Proof.* The proof in the case (b) is the same one in the case (a). To prove the lemma in the last case, we proceed in similar manner to the proof of (Lemma 5.5, [15]) to obtain following property: For all  $m \geq 0$ ,  $t', t \in [0, T]$  such that  $l \leq t - t'$ , and for all  $i, n$  satisfying  $U_{k_0}^r(i)/r \in [t', t]$ ,  $n \leq N_{k_0k}^r(i)$ , and

$$\mathbf{1}_{[m\kappa, (m+1)\kappa)}(V_{k_0k}^r(i, n)) \mathbf{1}_{(0, \kappa]}(V_{k_0k}^r(i, n) - \bar{S}_{U_{k_0}^r(i)/r, t}^r) = 1.$$

There exists  $v \in [0, t-l]$  such that  $U_{k_0}^r(i)/r \in (v, v+l]$ . This property jointly with conditions (C.6)-(C.14) concludes the proof.  $\square$

### C.2.3 Controlled oscillations

The following lemma summarizes the established estimates involving sufficient conditions for the proof of tightness.

**Lemma C.6.** *Let  $g \in C_b^1(\mathbb{R}_+)$ ,  $s < T < t^*$  and  $0 < \epsilon, \eta < 1$ . Set  $\epsilon = \frac{\epsilon}{2(\|g\|_\infty \vee 1)}$ . Let  $l, M_0, M_T, \Gamma, \kappa, r_0$  be the constant, and  $(B^r, r > 0)$  be the events, given by lemma C.4. Set  $\tilde{M}_T = M_0 \vee (\epsilon/4) + 2Q_{kk_0}\alpha_{k_0}T$ ,  $M = (\|g\|_\infty \vee 1)\tilde{M}_T$  and*

$$\delta = \min \left\{ T/2, l, \frac{\epsilon}{4\Gamma\tilde{M}_T(\|g\|_\infty \vee 1)}, \kappa/\Gamma, 1 \right\}. \quad (\text{C.15})$$

Then we have

$$\mathbb{P}^r \left( \sup_{t \in [0, T]} |\langle g, \tilde{\gamma}_{k_0 k}^r(t) \rangle| \leq M \right) \geq 1 - \eta \quad (\text{C.16})$$

$$\mathbb{P}^r \left( \sup_{t \in [s\epsilon/2, T-\delta]} \sup_{h \in [0, \delta]} |\langle g, \tilde{\gamma}_{k_0 k}^r(t+h) \rangle - \langle g, \tilde{\gamma}_{k_0 k}^r(t) \rangle| \leq \epsilon \right) \geq 1 - \eta \quad (\text{C.17})$$

$$\mathbb{P}^r \left( \sup_{t \in [0, s\epsilon/2]} \sup_{h \in [0, \delta]} |\langle g, \tilde{\gamma}_{k_0 k}^r(t+h) \rangle - \langle g, \tilde{\gamma}_{k_0 k}^r(t) \rangle| \leq \epsilon \right) \geq 1 - \eta. \quad (\text{C.18})$$

### C.2.4 Proof of Tightness for Proposition C.1

The property C2 follows by Lemma C.6. Let us verify C1. Recall that  $V_{k_0}^r(i)$  is the service time of the  $i^{\text{th}}$  customer in the system from beginning in class  $k_0$  to end its service, and  $V_{k_0 k}^r(j, n)$  is the service times of  $j^{\text{th}}$  in  $n^{\text{th}}$  visit of class  $k$  from beginning in class  $k_0$ . For each  $i \geq 1$  and  $k \in \mathcal{K}$  we have,  $\sum_{n=1}^{N_{k_0 k}^r(i)} V_{k_0 k}^{0,r}(i, n) \leq V_{k_0}^r(i)$ . By definition of  $\{\tilde{\gamma}_{k_0 k}^r(\cdot)\}$  we have

$$\begin{aligned} \langle \chi, \tilde{\gamma}_{k_0 k}^r(t) \rangle &= \frac{1}{r} \sum_{i=1}^{r\bar{Z}_{k_0}^r(0)} \sum_{n=1}^{N_{k_0 k}^r(i)} (V_{k_0 k}^{0,r}(i, n) - \bar{S}_t^r)^+ + \frac{1}{r} \sum_{i=1}^{r\bar{E}_{k_0}^r(t)} \sum_{n=1}^{N_{k_0 k}^r(i)} (V_{k_0 k}^r(i, n) - \bar{S}_{t, U_{k_0}^r(i)}^r)^+ \\ &\leq \frac{1}{r} \sum_{i=1}^{r\bar{Z}_{k_0}^r(0)} V_{k_0}^{0,r}(i) + \frac{1}{r} \sum_{i=1}^{r\bar{E}_{k_0}^r(t)} V_{k_0}^r(i). \end{aligned} \quad (\text{C.19})$$

According to the conditions of Section 6.1 (the four first equations), the Weak Law of Large Number for triangular arrays (Theorem 5.5 of [10]) shows that the right-hand side of (C.19) converges in law to  $\langle \chi, \zeta_{k_0}^0 \rangle \bar{Z}_{k_0}(0) + \langle \chi, \zeta_{k_0} \rangle \alpha_{k_0} t$ . Let  $\bar{M}_T = 2(\langle \chi, \zeta_{k_0}^0 \rangle \bar{Z}_{k_0}(0) + \langle \chi, \zeta_{k_0} \rangle \alpha_{k_0} T)$ . Since the limit is deterministic and continuous, we have,

$$\lim_{r \rightarrow \infty} \mathbb{P}^r \left( \sup_{[0, T]} \frac{1}{r} \sum_{i=1}^{r\bar{Z}_{k_0}^r(0)} V_{k_0}^{0,r}(i) + \frac{1}{r} \sum_{i=1}^{r\bar{E}_{k_0}^r(t)} V_{k_0}^r(i) \leq \bar{M}_T \right) = 1.$$

Then, from Inequality (C.19) we have

$$\lim_{r \rightarrow \infty} \mathbb{P}^r \left( \sup_{[0, T]} \langle \chi, \tilde{\gamma}_{k_0 k}^r(t) \rangle \leq \bar{M}_T \right) = 1. \quad (\text{C.20})$$

Let  $N_T = \bar{M}_T \vee \tilde{M}_T$ . By the dynamic equation we have  $\sup_{[0,T]} \langle 1, \bar{\gamma}_{k_0 k}^r(t) \rangle \leq \tilde{M}_T$ . So by (C.20) we have

$$\lim_{r \rightarrow \infty} \mathbb{P}^r \left\{ \sup_{[0,T]} (\langle \chi, \bar{\gamma}_{k_0 k}^r(t) \rangle \vee \langle 1, \bar{\gamma}_{k_0 k}^r(t) \rangle) \leq N_T \right\} = 1. \quad (\text{C.21})$$

Define  $C_{T,\eta} = \{\xi \in \mathcal{M} : \langle 1, \xi \rangle \vee \langle \chi, \xi \rangle \leq N_T\}$ , since  $\sup_{\xi \in C_T} \langle 1_{[c,\infty]}, \xi \rangle \rightarrow 0$  as  $c \rightarrow \infty$ . As in [15],  $C_{T,\eta}$  is relatively compact. By (C.21) we have

$$\lim_{r \rightarrow \infty} \mathbb{P}^r \left\{ \bar{\gamma}_{k_0 k}^r(t) \in C_{T,\eta} \text{ for all } t \in [0, T] \right\} = 1.$$

This is also verified with the closure of  $C_{T,\eta}$  which proves C1.

### C.3 Limit point properties

Having established tightness of  $\bar{\gamma}_{k_0 k}^r(\cdot)$ , we return to the second part of Proposition C.1. The following equivalence of fluid solution  $\bar{\gamma}_{k_0 k}(\cdot)$  proceeds in a manner nearly identical to the equivalence for single class processor sharing queue in Gromoll and al.

**Lemma C.7.** *The measure  $\bar{\gamma}_{k_0 k}(\cdot)$  is a fluid solution of Equation (3.30) if and only if it satisfies:*

S1  $\bar{\gamma}_{k_0 k}(\cdot)$  is continuous,

S2  $\langle 1_{\{0\}}, \bar{\gamma}_{k_0 k}(t) \rangle = 0$  for all  $t \geq 0$ ,

S3 For each  $g \in \mathcal{C}$ , we have

$$\langle g, \bar{\gamma}_{k_0 k}(t) \rangle = \langle g, \bar{\gamma}_{k_0 k}(0) \rangle - \int_0^t \frac{\langle g', \bar{\gamma}_{k_0 k}(u) \rangle}{\langle 1, \bar{\gamma}(u) \rangle} du + \langle g, (\mathcal{B} * \nu)_{k k_0} \rangle \alpha_{k_0} t \quad (\text{C.22})$$

for all  $t < t_\rho$ , and  $\langle g, \bar{\gamma}_{k_0 k}(t) \rangle = 0$  for all  $t \geq t_\rho$ ,  $t_\rho$  being defined by (2.12).

On each route,  $\bar{\gamma}_{k_0 k}(\cdot)$  describes the fluid dynamics of customers who start their service in class  $k_0$  to reach class  $k$ . Recall that the service rate per unit of mass in the system equals the reciprocal of the total mass in the system. Because, when a customer finishes service and reenters immediately the queue, there is no effect on the other customers and on the rate of service. Thus Equation (C.22) evolves as a fluid model equation in a monoclase processor sharing queue. We complete the proof of Proposition C.1 by verifying that any limit of the sequence  $\{\bar{\gamma}_{k_0 k}^r(\cdot)\}$  satisfies the conditions S1, S2 and S3. The proofs are similar of that of Gromoll et al. [15] in the case (a), and of that of Puha et al. [20] in case (b).

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Setting of the problem . . . . .	3
1.2	Literature . . . . .	3
1.3	Summary of the results . . . . .	3
1.4	Notation . . . . .	5
<b>2</b>	<b>The Queuing Model and the Fluid Model</b>	<b>5</b>
2.1	Primitive data and initial conditions . . . . .	5
2.2	Queuing equations . . . . .	6
2.3	Fluid Model . . . . .	7
<b>3</b>	<b>Preliminaries</b>	<b>8</b>
3.1	Mapping to the single-class case . . . . .	8
3.2	A descriptor for the multiclass case . . . . .	12
<b>4</b>	<b>Principal Results</b>	<b>15</b>
4.1	Existence and Uniqueness . . . . .	15
4.2	Asymptotics of fluid solutions for critical and supercritical data . . . . .	16
4.3	Extension to the Discriminatory Processor Sharing queue . . . . .	17
4.4	Convergence to the fluid model solution . . . . .	18
<b>5</b>	<b>Proofs for the fluid model</b>	<b>19</b>
5.1	Results for a nonzero initial state . . . . .	19
5.1.1	Time range of the solution . . . . .	19
5.1.2	Existence and uniqueness for $\xi \neq 0$ . . . . .	20
5.1.3	The workload of the queue . . . . .	22
5.1.4	Continuity properties of the measure $\mu(\cdot)$ . . . . .	23
5.1.5	Relationship with the single-class case . . . . .	24
5.1.6	Proof of Proposition 5.1 . . . . .	25
5.2	Existence of a solution for $\xi = 0$ . . . . .	25
5.3	Convergence to initial Fluid solution for critical and supercritical data . . . . .	26
5.3.1	Proof of Theorems 4.1 and 4.2 . . . . .	29
5.4	A characterization of the measure $\mu_k$ . . . . .	29
5.5	Asymptotic results . . . . .	31
5.5.1	Asymptotics for the response time . . . . .	31
5.5.2	Asymptotics for the measure $\bar{\mu}(t)$ . . . . .	32
5.5.3	Proof of theorems 4.3 and 4.4 . . . . .	36
5.6	Extensions . . . . .	37
5.6.1	Proof of Proposition 4.1 . . . . .	37
5.6.2	Proof of Proposition 4.2 . . . . .	37
<b>6</b>	<b>Proof of the convergence</b>	<b>37</b>
6.1	Auxiliary results . . . . .	37
6.2	Proof of Theorem 4.5 . . . . .	39

---

<b>7 Illustrations</b>	<b>42</b>
7.1 Trajectories . . . . .	42
7.2 Slowness as a function of the service time . . . . .	44
7.3 Competition between classes . . . . .	44
7.3.1 The effect of service length. . . . .	45
7.3.2 The effect of the distribution . . . . .	45
 <b>Bibliography</b>	 <b>47</b>
 <b>Appendices</b>	 <b>48</b>
<b>A Multidimensional renewal equations</b>	<b>48</b>
<b>B Convergence of shifted convolutions</b>	<b>49</b>
<b>C Proof of Proposition 6.3</b>	<b>56</b>
C.1 Cases without input . . . . .	56
C.2 Proof of tightness . . . . .	58
C.2.1 Law of large numbers . . . . .	58
C.2.2 Weak estimates . . . . .	59
C.2.3 Controlled oscillations . . . . .	60
C.2.4 Proof of Tightness for Proposition C.1 . . . . .	60
C.3 Limit point properties . . . . .	61



---

Centre de recherche INRIA Sophia Antipolis – Méditerranée  
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex  
Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier  
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq  
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique  
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex  
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex  
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex  
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex

---

Éditeur  
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)  
<http://www.inria.fr>  
ISSN 0249-6399