Fluid limits for an ALOHA-type model with impatient customers

Maria Frolkova · Sergey Foss · Bert Zwart

Received: 12 March 2012 / Revised: 12 March 2012 / Published online: 21 April 2012
© Springer Science+Business Media, LLC 2012

Abstract Random multiple-access protocols of type ALOHA are used to regulate networks with a star configuration where client nodes talk to the hub node at the same frequency (finding a wide range of applications among telecommunication systems, including mobile telephone networks and WiFi networks). Such protocols control who talks at what time sharing the common idea “try to send your data and, if your message collides with another transmission, try resending later”.

In the present paper we consider a time-slotted ALOHA model where users are allowed to renege before transmission completion. We focus on the scenario that leads to overload in the absence of impatience. Under mild assumptions, we show that the fluid (or law-of-large-numbers) limit of the system workload coincides a.s. with the unique solution to a certain integral equation. We also demonstrate that the

M. Frolkova (✉) · B. Zwart
CWI, P.O. Box 94079, 1098 XG Amsterdam, The Netherlands
e-mail: M.Frolkova@wii.nl

S. Foss
Heriot-Watt University, EH14 4AS Edinburgh, UK
e-mail: s.foss@hw.ac.uk

S. Foss
S.L. Sobolev Institute of Mathematics, Novosibirsk, Russia

B. Zwart
EURANDOM, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
e-mail: Bert.Zwart@wii.nl

B. Zwart
VU University Amsterdam, Amsterdam, The Netherlands

B. Zwart
Georgia Institute of Technology, Atlanta, GA, USA
fluid limits for distinct initial conditions converge to the same value as time tends to infinity.

**Keywords**  ALOHA protocol · Queues with impatience · Queues in overload · Fluid limits

**Mathematics Subject Classification** 60K25 · 60F17 · 90B15 · 90B22

1 Introduction

ALOHA-type algorithms are intended to govern star networks in which multiple client machines send data packets to the hub machine at the same frequency. Thus, collisions of packets being transmitted simultaneously are possible (clients know nothing about each other’s intentions to transmit data and cannot prevent collisions). Such algorithms assume the following acknowledgment mechanism. If data has been received correctly at the hub, which is possible only if no collisions occurred during its transmission, then a short acknowledgment packet is sent to the client. If a client has not received an acknowledgment after a short wait time, then it retransmits the packet after waiting a randomly selected time interval with distribution specified by the ALOHA protocol that governs the network.

The pioneering ALOHA computer network, also known as the ALOHAnet, was developed at the university of Hawaii under the leadership of Norman Abramson (see [1], where Abramson first proposed the ALOHA multi-access algorithm). The goal was to use low-cost commercial radio equipment to connect users on Oahu and the other Hawaiian islands with the central computer on the main Oahu campus. The ALOHAnet became operational in 1971, providing the first demonstration of a wireless data network. Nowadays the ALOHA random access techniques are widely used in WiFi, mobile telephone networks and satellite communications.

To give an example, we describe the conventional centralised time-slotted ALOHA model. Here “slotted time” means that users enter the system and abandon it, initiate and finish transmissions at times \( n = 1, 2, \ldots \). The arrival process forms an i.i.d. sequence \( \{\xi(n); \ n \geq 1\} \); all service times are assumed to equal 1. “Centralised model” means that the total number of users in the system is always known. Let \( W(n) \) denote the total number of users at time \( n \). For any \( n \), at the beginning of the \( n \)th time slot, which is the time interval \([n, n + 1)\), each of the \( W(n) \) customers present in the system starts his transmission with probability \( p(n) \) (and does not with probability \( 1 - p(n) \)) independently of the others. If two or more users attempt transmissions simultaneously, then the transmissions collide and hence fail, causing the users to remain in the system in order to retransmit later. After a successful transmission the user leaves immediately. Note that, for any time slot, given there are \( m \) customers each starting his transmission with probability \( p \), the probability of a successful transmission equals \( mp(1 - p)^{m-1} \) and is maximised by \( p = 1/m \). So we assume that \( p(n) = 1/W(n) \). In such a setting, the population process \( \{W(n); \ n \geq 0\} \) forms a Markov chain that is positive recurrent if \( E\xi(1) < e^{-1} \) (the system is stable) and transient if \( E\xi(1) > e^{-1} \) (the system is unstable). Stability conditions for other ALOHA-type models can be found in [5, 8, 10].

\( \copyright \) Springer
In the present paper we extend the framework described above by allowing impatience of users. Each user knows how long he can stay in the system and abandons the system as soon as he succeeds in transmitting his packet or his patience time expires, depending on which happens earlier. To distinguish between different levels of patience, we assume multiple customer classes. We also assume that the input process is non-trivial, i.e. \( P\{\xi(1) \geq 2\} > 0 \), and that \( E\xi(1) > e^{-1} \). The latter condition would mean the overload regime if not for impatience of customers.

Abandonment before service completion is a very natural behaviour in the case when waiting time exceeds some threshold because dwelling in the system means spending resources, and whatever one can imagine them to be: time itself, money, etc., there is always a limited amount of them at disposal. As for overload causing long waiting times, a system may experience it even if it is not expected to, due to fluctuations of the actual traffic from the designed traffic. These are the reasons of our interest in combination of the overload regime and impatience of customers.

The results of the paper concern fluid limits for the workload process, i.e. weak limits that arise under a law-of-large-numbers scaling. We propose a fluid analogue of the stochastic model and, under mild assumptions, show that the family of the scaled workload processes is relatively compact with weak limit points a.s. satisfying the fluid model equation. Besides the last argument, we provide an independent analytical proof of the existence of a fluid model solution for any initial state, and show that it is unique. We also demonstrate that there exists a unique equilibrium point for the fluid model equation and that all fluid limits stabilise to this point as time tends to infinity.

One of the possible generalisations of the model treated in the present paper is to allow interference of transmissions only if the distance between the corresponding client machines is small, and this is a subject of our future research. Such an extension of the network topology was proposed by Bordenave, Foss and Shneer in [3], where impatience of customers is not taken into account, though. The authors study fluid limits in order to find out whether the stochastic model is stable or not. The description of fluid limits for our model with impatience omitted coincides with the description of fluid limits in [3] adapted to the network topology we consider here.

The paper is organised as follows. In Sect. 2 we present a detailed description of the stochastic model. In Sect. 3 we define the fluid model solutions and formulate their properties. In Sect. 4 we state our main result concerning convergence of sequences of the fluid scaled workload processes to fluid model solutions. Sections 5, 6 and 7 contain the proofs of the results stated in Sects. 3 and 4. Namely, in Sect. 5 we establish the existence and uniqueness of a fluid model solution for any initial state. In Sect. 6 the existence, uniqueness and global asymptotic stability of an equilibrium fluid model solution are shown. In Sect. 7 we prove the result of Sect. 4. In the remainder of this section we list the notation we use throughout the paper.

**Notation** The standard sets are denoted as follows: the real line \( \mathbb{R} = (-\infty, \infty) \), the non-negative half-line \( \mathbb{R}_+ = [0, \infty) \), the non-negative integers \( \mathbb{Z}_+ = \{0, 1, 2, \ldots\} \). All vectors in the paper are \( K \)-dimensional with integer or real coordinates. The coordinates of a vector are denoted by the same symbol as the vector with lower indices \( 1, \ldots, K \) added. If a vector has superscripts, overlining, or tilde sign, they remain in its coordinates. For example, \( \bar{x}^0 \in \mathbb{R}^K_+, \bar{x}^0 = (\bar{x}^0_1, \ldots, \bar{x}^0_K) \). The supremum
and $L_1$-norm in $\mathbb{R}^K$ are $\|x\| = \max_{1 \leq i \leq K} |x_i|$ and $\|x\|_1 = \sum_{i=1}^K |x_i|$, respectively. Vector inequalities hold coordinate-wise. The coordinate-wise product of vectors is $x * y = (x_1y_1, \ldots, x_Ky_K)$.

The signs $\Rightarrow$, $\Leftarrow$ and $\leq_{st}$ stand for weak convergence, equality in distribution and stochastic order, respectively. Recall that, for real-valued r.v.'s $X$ and $Y$, $X \leq_{st} Y$ if $\mathbb{P}[X > t] \leq \mathbb{P}[Y > t]$ for any $t \in \mathbb{R}$. For an r.v. $X$ having a distribution $G$, we write $X \sim G$. The notations $\Pi(a)$, with $a \in (0, \infty)$, and $B(m, x)$, with $m \in \mathbb{Z}_+, x \in [0, 1]$, are used for the Poisson distribution with parameter $a$ and binomial distribution with parameters $m, x$. With $B(0, x)$ we mean the degenerate distribution localised at 0.

The complement of an event $E$ is denoted by $\overline{E}$.

For a function $f : S \to \mathbb{R}^K$, its supremum norm is $\|f\|_S = \max_{1 \leq i \leq K} \sup_{x \in S} |f_i(x)|$. Let $\mathcal{D}$ be the space of right-continuous functions $f : [0, \infty) \to \mathbb{R}^K$ with finite left limits. Endow this space with the Skorokhod $J_1$-topology. Let $\mathcal{G}$ be the class of continuous functions $g : [0, \infty) \to \mathbb{R}^K_+$ such that $\|g(t)\| > 0$ for all $t \neq 0$.

We use the signs $=:=$ and $=:=$ to introduce a definition in the body of a formula.

## 2 Stochastic model description

This section contains a detailed description of the stochastic model under study. In particular, it derives the dynamic equation for the system workload. All stochastic primitives introduced here are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with expectation operator $\mathbb{E}$. The characteristic function of an event is denoted by $\mathbb{I}$.

**Stochastic assumptions and the protocol** We consider an ALOHA-type service system with impatient customers. The system includes a waiting room where customers arrive to, and a server. Time is slotted, i.e. arrivals and abandonments may occur only at time instants $n = 1, 2, \ldots$. Time slot $n$ is the time interval $[n, n + 1)$. We assume that there are $K < \infty$ classes of customers.

The arrival process is denoted by $\{A(n); \ n \geq 1\}$, where $A(n) = (A_1(n), \ldots, A_K(n))$, and $A_i(n)$ is the number of class $i$ customers arriving at time $n$. We suppose that $\{A(n); \ n \geq 1\}$ is an i.i.d. sequence and that $\mathbb{E}A(1) =: \Lambda = (A_1, \ldots, A_K) \in (0, \infty)^K$. The coordinates of the vectors $A(n)$ are allowed to be dependent.

Let $W_i(n)$ be the number of class $i$ customers present in the system at time $n \geq 0$. As can be seen from the further description, $W_i(n)$ coincides with the workload at time $n$ due to class $i$ customers. Hence, $\{W(n) = (W_1(n), \ldots, W_K(n)); \ n \geq 0\}$ denotes the workload process.

Each customer brings a packet that takes exactly a single time slot to be transmitted to the server. He also sets a deadline for transmission: once the deadline expires, the customer leaves the system even if his packet has not been transmitted yet. In this case we say that the customer has abandoned the system due to impatience. Patience times of class $i$ customers have geometrical distribution with parameter $p_i$ and take values greater than or equal to 1. By $p$ we denote the vector having parameters $p_i$ as its coordinates, $p = (p_1, \ldots, p_K)$. Patience times of different customers are mutually independent.
We now describe how a transmission occurs. At the beginning of time slot $n$, each customer, independently of the others, starts transmission with probability $1/\|W(n)\|_1$ (and does not with probability $1 - 1/\|W(n)\|_1$). If there is one customer transmitting, then the transmission is going to be successful. Otherwise a collision happens. At the end of the time slot, customers learn the result. If a customer has succeeded in sending his packet, he immediately leaves the system. If a customer has failed and he is from class $i$, then with probability $p_i$ he leaves due to impatience, and with probability $1 - p_i$ he stays in the system to try to retransmit his packet later.

Throughout the paper we assume the following.

**Assumption 1** The input process is non-trivial, i.e. $\mathbb{P}\{\|A(1)\|_1 \geq 2\} > 0$.

**Assumption 2** The mean amount of work arriving to the system per time slot exceeds the stability threshold for the model with no impatience of customers, i.e. $\|A\|_1 > e^{-1}$.

**Remark 1** The results of the paper can be generalised to the case where patience times are not simply geometrical random variables but finite mixtures of those. It suffices, for all $i = 1, \ldots, K$, to split customer class $i$ into $k_i$ new classes, where $k_i$ is the number of mixture components in the patience time distribution for class $i$ customers.

**Workload dynamics** The sequence $\{W(n); n \geq 0\}$ forms a time-homogeneous Markov chain. Its dynamics can be described as follows: given a history $\{W(m); m = 0, \ldots, n\}$ up to time $n$ with $W(n) = x$, we have, for $i = 1, \ldots, K$,

$$W_i(n + 1) \overset{\text{d}}{=} x_i + A_i - T_i(x) - I_i(x),$$

(1)

where

- $A = (A_1, \ldots, A_K) \overset{\text{d}}{=} A(1)$;
- $T_i(x)$ represents the number of class $i$ customers who are present in the system at time $n$ but will leave at the end of time slot $n$ because of a successful transmission,

$$T_i(x) = \mathbb{I}\{B_i(x) = 1\} \prod_{j \neq i} \mathbb{I}\{B_j(x) = 0\},$$

where $B_i(x) \sim B(x_i, 1/\|x\|_1)$ if $x \neq (0, \ldots, 0)$ and $B_i(0, \ldots, 0) = 0$;
- $I_i(x)$ represents the number of class $i$ customers who are present in the system at time $n$ but will leave at the end of time slot $n$ due to impatience rather than a successful transmission; given $x - T(x) = y$,

$$I_i(x) = \tilde{B}_i(y) \sim B(y_i, p_i);$$

- for any $x, y \in \mathbb{Z}_+^K$, the random vector $A$ and the binomial r.v.’s $B_i(x)$, $\tilde{B}_j(y)$, $i, j = 1, \ldots, K$, are mutually independent.

**Remark 2** The number $T_i(x)$ of successful transmissions by class $i$ customers at a time slot may take only values 0 and 1. Moreover, $\|T(x)\|_1 \leq 1$. 

© Springer
3 Fluid model

In the present section we define a fluid analogue of the stochastic model described in the previous section. As time and space are appropriately normalised, we expect that the difference equation (1) can be approximated by a differential equation where the rate of increase is due to arrival rates, and the rate of decrease due to service completions and abandonments. In the single-class case, one may expect such a differential equation to look like (we omit the class index)

\[ z'(t) = \Lambda - e^{-1} - pz(t), \]

since the throughput of ALOHA is \( e^{-1} \). In the multiclass case, this naturally extends to

\[ z'_i(t) = \Lambda_i - e^{-1}z_i(t)/(\sum_{j=1}^K z_j(t)) - p_i z_i(t), \quad i = 1, \ldots, K. \]

This will be made rigorous in Sect. 4. We now proceed more formally.

**Definition 1** For \( z^0 \in \mathbb{R}^+_K \), a solution to the integral equation

\[ z(t) = z^0 + t\Lambda - e^{-1} \int_0^t m(z(s)) \, ds - p* \int_0^t z(s) \, ds, \quad (2) \]

that comes from \( \mathcal{G} \), i.e. that is continuous, non-negative and such that, for all \( t \neq 0 \), \( \|z(t)\| > 0 \), is called a fluid model solution (FMS) with initial state \( z^0 \). The function \( m : \mathbb{R}^+_K \to \mathbb{R}^+_K \) is given by

\[ m(x) = \begin{cases} x/\|x\|_1 & \text{if } x \neq (0, \ldots, 0), \\ \Lambda/\|\Lambda\|_1 & \text{if } x = (0, \ldots, 0). \end{cases} \]

**Remark 3** For a function from \( \mathcal{G} \), (2) is equivalent to the Cauchy problem

\[ z'(t) = \Lambda - p*z(t) - e^{-1}m(z(t)), \quad z(0) = z^0, \quad (3) \]

where (3) holds for \( t \geq 0 \) if \( z^0 \neq (0, \ldots, 0) \) and for \( t > 0 \) if \( z^0 = (0, \ldots, 0) \).

**Remark 4** Although \( m(0, \ldots, 0) \) does not appear in (3), it needs to be defined for further use in Sect. 7. We assign to \( m(0, \ldots, 0) \) the value of \( \Lambda/\|\Lambda\|_1 \), which is the limit of \( m(\cdot) \) along FMS’s trajectories as they approach \((0, \ldots, 0)\). Indeed, the only point where a fluid model solution can take the value of \((0, \ldots, 0)\) is \( t = 0 \). Let \( z(\cdot) \) be an FMS starting from \( z(0) = (0, \ldots, 0) \). For the moment, suppose that \( z(\cdot) \) is continuously differentiable at \( t = 0 \). Then (3) and Taylor’s expansion give, for small \( t > 0 \),

\[ z'(t) = \Lambda - p*z(t) - e^{-1}t z'(0) + o_K(t)/\sum_{i=1}^K z'_i(0) t + o(t), \quad (4) \]

where \( o(t) \in \mathbb{R} \), \( o_K(t) \in \mathbb{R}^K \), and \( o(t)/t \to 0 \), \( o_K(t)/t \to (0, \ldots, 0) \) as \( t \to 0 \). The continuity of \( z'(\cdot) \) at \( t = 0 \) and Assumption 2 guarantee that \( \sum_{i=1}^K z'_i(0) > 0 \), so we pass to the limit as \( t \to 0 \) on both sides of (4) and get

\[ z'(0) = \Lambda - e^{-1}z'(0)/\sum_{i=1}^K z'_i(0). \]

The last equation has a unique solution \( z'(0) = (1 - e^{-1}/\|\Lambda\|_1) \Lambda \), which implies the existence of the limit \( \lim_{t \to 0} m(z(t)) = \)
If \( \delta(T) > \Delta_{1m}(z) \) are continuous and non-zero at every point, and hence, for each with other parameters. The whole idea of this proof is adopted from [4].

In the remainder of the section, we discuss properties of FMSs.

Existence and uniqueness of FMS’s follow by the classical results from the theory of ordinary differential equations if the initial state is non-zero. Otherwise the proof is rather involved. The outline follows below; see Sect. 5 for the full proof.

Theorem 1 For any initial state, a fluid model solution exists and is unique.

One-dimensional case Equation (2) turns into (we omit the class index) \( z(t) = z^0 + (A - e^{-1})t - p \int_0^t z(s) \, ds \), and its unique solution is given by \( z(t) = z^0 e^{-pt} + ((A - e^{-1})/p)(1 - e^{-pt}) \).

Multidimensional case, non-zero initial state. Uniqueness follows easily by the Gronwall inequality (see for example [9, Chap. III, Sect. 1]).

Proposition 1 (Gronwall inequality) Suppose that functions \( u(\cdot) \) and \( v(\cdot) \) are non-negative and continuous in \([a, b] \) and that there exists a constant \( C \geq 0 \) such that \( v(t) \leq C + \int_0^t v(s)u(s) \, ds, \ a \leq t \leq b \). Then \( v(t) \leq C \exp \int_a^t u(s) \, ds, \ a \leq t \leq b \). In particular, if \( C = 0 \), then \( v(t) \equiv 0 \).

Since the first-order partial derivatives of the function \( m(\cdot) \) are bounded on all sets \( \mathbb{R}^K_\delta := \{ x \in \mathbb{R}^K : \| x \| \geq \delta \}, \ \\delta > 0 \), this function is Lipschitz continuous on all such sets. Let \( c(\delta) \) be a Lipschitz constant for \( m(\cdot) \) on \( \mathbb{R}^K_\delta \) with respect to the supremum norm, i.e. \( \| m(x) - m(y) \| \leq c(\delta)\| x - y \| \) for all \( x, y \in \mathbb{R}^K_\delta \).

Suppose that \( z(\cdot) \) and \( \tilde{z}(\cdot) \) are two FMSs with the same non-zero initial state. They are continuous and non-zero at every point, and hence, for each \( T > 0 \), there exists a \( \delta(T) > 0 \) such that \( z(t), \tilde{z}(t) \in \mathbb{R}^K_{\delta(T)}, 0 \leq t \leq T \). We put \( \Delta z(\cdot) = z(\cdot) - \tilde{z}(\cdot) \) and \( \Delta m(z) = m(z) - m(\tilde{z}) \). Then \( -\Delta z(t) = e^{-1} \int_0^t \Delta m(z(s)) \, ds + p^* \int_0^t \Delta z(s) \, ds \), and, for \( t \in [0, T] \), we have \( \| \Delta z \|_{[0, t]} \leq \left( e^{-1}c(\delta(T)) + \| p^* \| \right) \int_0^t \| \Delta z \|_{[0, s]} \, ds \). By the last inequality and the Gronwall inequality, we have \( \| \Delta z \|_{[0, t]} \leq 0, 0 \leq t \leq T \). Since \( T \) is arbitrary, \( z(\cdot) \) and \( \tilde{z}(\cdot) \) coincide on \([0, \infty) \).

Existence of an FMS with a non-zero initial state can be shown by the Peano existence theorem (the proof is postponed to Sect. 5).

Multidimensional case, zero initial state. The techniques used in the previous case do not work here because they are based on the continuity properties of the function \( m(\cdot) \) that fail as zero comes into play. So a different approach is required. We introduce a family of integral equations depending on parameters \( (\varepsilon, a) \in \mathbb{R}_+ \times \mathbb{R}^K_\delta \) that includes (for \( (\varepsilon, a) = (0, p) \)) an equation equivalent to the Cauchy problem (3) with \( z^0 = (0, \ldots, 0) \). We show that each equation of this family has a solution. If \( \varepsilon > 0 \), then the uniqueness of an \((\varepsilon, a)\)-solution is straightforward to show. In order to prove the uniqueness of a \((0, a)\)-solution, we derive a proper estimate for it via solutions with other parameters. The whole idea of this proof is adopted from [4].
Equilibrium FMS  We now discuss the existence, uniqueness and asymptotic stability of a constant solution to the fluid model equation (3).

Definition 2 If a differential equation $z'(t) = f(t, z(t))$, $t \geq 0$, $z, f \in \mathbb{R}^K$, admits a constant solution, then this solution is called an equilibrium point.

By means of a Lyapunov function, we prove the following result; see Sect. 6.

Theorem 2 Suppose Assumption 2 holds. Then there exists a unique equilibrium point for the fluid model equation (3), which is given by

$$z^e_i = \frac{A_i}{x + p_i}, \quad i = 1, \ldots, K,$$

where $x$ solves

$$\sum_{i=1}^K \frac{p_i A_i}{x + p_i} = \|A\|_1 - e^{-1},$$

(5)

and any fluid model solution $z(t)$ converges to $z^e$ as $t \to \infty$.

4 Main theorem

In this section we characterise the asymptotic behaviour under a fluid scaling of the workload process of the stochastic model introduced in Sect. 2, justifying the heuristics given in the previous section. First we specify the fluid scaling. Let $R$ be a positive number. Consider the stochastic model from Sect. 2 with the impatience parameters $p_i$ replaced by $p_i/R$, $i \in \{1, \ldots, K\}$. Denote the workload process of the $R$th model by $W^R(\cdot)$, and scale it by $R$ both in space and time,

$$\overline{W}^R(t) := \frac{1}{R} W^R(\lfloor Rt \rfloor), \quad t \geq 0.$$ (6)

We refer to the processes (6) as the fluid scaled workload processes. They take values in the Skorokhod space $\mathcal{D}$.

Remark 5 If the limit is continuous (which is the case whenever we prove convergence in $\mathcal{D}$ in this paper), then convergence in the Skorokhod $J_1$-topology is equivalent to uniform convergence on compact sets.

Now we formulate the main theorem of the paper and highlight the basic steps of the proof; the detailed proof will follow in Sect. 7.

Theorem 3 Let Assumptions 1 and 2 hold. Then any sequence of the fluid scaled workload processes $\overline{W}^R(\cdot)$ such that $\overline{W}^R(0) \Rightarrow z^0 = \text{const}$, $R \to \infty$, converges weakly in $\mathcal{D}$ to the (unique by Theorem 1) FMS with initial state $z^0$.

The first step of the proof of Theorem 3 is to show that the fluid scaled workload satisfies an integral equation that differs from the fluid model equation (2) by negligible terms (for $R$ large enough). This appears to be the most difficult part of the proof.
Then we obtain the relative compactness of the family \(\{\overline{W}^R(\cdot); R > 0\}\) and continuity of its weak limits by applying a general result on relative compactness in \(D\). We also show that the weak limits are bounded away from zero outside \(t = 0\), which, together with their continuity, implies that they a.s. satisfy the fluid model equation (2), and hence all of them coincide a.s. with the unique FMS with initial state \(z^0\).

**Remark 6** In the literature on the fluid approximation approach, a fluid limit is any weak limit of the processes (6) along a subsequence \(R \to \infty\). As it follows from Theorem 3, for a sequence of the processes (6) with \(\overline{W}^R(0) \Rightarrow z^0 = \text{const, } R \to \infty\), there is a unique fluid limit, and it is a deterministic function, namely, the fluid model solution with initial state \(z^0\).

### 5 Proof of Theorem 1

We split the proof into two parts, for a non-zero and zero initial state.

#### 5.1 Non-zero initial state

Here we have to prove the existence result only, see Sect. 3 for the proof of uniqueness. First we derive bounds for an FMS using the following lemma.

**Lemma 1** Let \(S\) be either a finite interval \([0, T]\) or the half-line \([0, \infty)\), and let a real-valued function \(x(t)\) be continuous in \(S\) and differentiable in \(S \setminus \{0\}\). Suppose that a constant \(C\) is such that \(x(t) \geq C\) for \(t \in S \setminus \{0\}\) implies \(x'(t) \leq 0\). Then
\[
\sup_{t \in S} x(t) \leq \max\{x(0), C\}.
\]

**Proof** Let \(\varepsilon > 0\). Suppose that \(x(t) \in (\max\{x(0), C\}, \max\{x(0), C\} + \varepsilon]\). Then, starting from time \(t\), the function \(x(t)\) is decreasing at least until it reaches level \(C\). So \(\sup_{t \in S} x(t) \leq \max\{x(0), C\} + \varepsilon\). Since \(\varepsilon > 0\) is arbitrary, \(\sup_{t \in S} x(t) \leq \max\{x(0), C\}\). \(\square\)

**Bounds for a fluid model solution** Let \(S\) be either a finite interval \([0, T]\) or the half-line \([0, \infty)\). Suppose that a function \(z(\cdot)\) is non-negative with \(\|z(\cdot)\|_1 > 0\) and continuous in \(S\), and that it solves the fluid model equation (3) in \(S\). Then \(\|z(t)\|_1 = \sum_{i=1}^K z_i(t)\), and the derivative \(\|z(t)\|'_1\) exists for all \(t \in S\). Summing up the coordinates of (3), we get
\[
\|A\|_1 e^{-1} - p^* \|z(t)\|_1 \leq \|z(t)\|'_1 \leq \|A\|_1 e^{-1} - p^* \|z(t)\|_1, \quad t \in S,
\]
where \(p^* = \min_{1 \leq i \leq K} p_i\) and \(p^* = \max_{1 \leq i \leq K} p_i\). Then Lemma 1 applied to \(x(\cdot) = \|z(\cdot)\|'_1\) and \(C = (\|A\|_1 e^{-1})/p^*\), and \(x(\cdot) = -\|z(\cdot)\|'_1\) and \(C = (\|A\|_1 e^{-1})/p^*\), implies that
\[
\sup_{t \in S} \|z(t)\|_1 \leq \max\left\{\|z(0)\|_1, \frac{\|A\|_1 e^{-1}}{p^*}\right\} =: u(z(0)), \quad (7)
\]
By Assumption 2, we have \( u(z(0)) > 0 \) for any non-negative \( z(0) \), and \( l(z(0)) > 0 \) for any non-negative and non-zero \( z(0) \).

Inequality (7) implies that \( z_i'(t) \leq A_i - e^{-1}u^{-1}(z(0))z_i(t) - p_i z_i(t) \), \( i = 1, \ldots, K \).

Then, by Lemma 1, we get

\[
\sup_{t \in S} z_i(t) \leq \max \left\{ z_i(0), \frac{A_i}{e^{-1}u^{-1}(z(0)) + p_i} \right\} =: u_i(z(0)).
\]  

(9)

Similarly, if \( z(0) \neq (0, \ldots, 0) \), then inequality (8) and Lemma 1 yield, for \( i = 1, \ldots, K \),

\[
\inf_{t \in S} z_i(t) \geq \min \left\{ z_i(0), \frac{A_i}{e^{-1}l^{-1}(z(0)) + p_i} \right\} =: l_i(z(0)).
\]  

(10)

**Remark 7** If \( z(0) \neq (0, \ldots, 0) \), then, by the bound (10) and since \( z_i'(0) = A_i > 0 \) if \( z_i(0) = 0 \), we have

\[
\text{for all } \delta > 0, \quad \inf_{t \in S, \delta > 0} \min_{1 \leq i \leq K} z_i(t) > 0.
\]

Existence of a fluid model solution with a non-zero initial state The key tool used in this proof is the Peano existence theorem (see for example [9, Chap. II, Sect. 2]).

**Proposition 2** (Peano) Suppose that a function \( f : \mathbb{R} \times \mathbb{R}^K \to \mathbb{R}^K \) is continuous in the rectangle \( B = \{(t, y) : t_0 \leq t \leq t_0 + a, \max_{1 \leq i \leq K} |y_i - y_i^0| \leq b\} \). Let \( M \geq \|f\|_B \) and \( \alpha = \min\{a, b/M\} \). Then the Cauchy problem

\[
y'(t) = f(t, y(t)),
\]

\[
y(t_0) = y^0,
\]

has a solution in the interval \([t_0, t_0 + \alpha]\) such that \( y(t) \in B \).

First note that it suffices to show the existence of a non-negative solution to (3). By Remark 7, it will not hit zero at \( t > 0 \) and hence will be an FMS.

Further note that it suffices to consider initial states with all coordinates positive. Indeed, if \( z(0) \) is non-zero, there exists a rectangle \( B = \{ (t, y) : t_0 \leq t \leq t_0 + a, \max_{1 \leq i \leq K} |y_i - y_i^0| \leq b \} \) that does not contain zero, and, consequently, the mapping \( f(\cdot) = e^{-1}m(\cdot) + p \) is continuous in \( B \). Let \( M = \|f\|_B \) and \( \alpha = b/M \). By the Peano theorem, there exists a solution to (3) in the interval \([0, \alpha]\). If \( z_i(0) > 0 \), then, by the continuity of \( z(\cdot) \), there exists a \( t_i \leq \alpha \) such that \( z_i(t) \geq z_i(0)/2, t \leq t_i \). If \( z_i(0) = 0 \), then \( z_i'(0) = A_i > 0 \), and there exists a \( t_i \leq \alpha \) such that \( z_i(t) = A_i t (1 + o(1)) \geq A_i t/2, t \leq t_i \). Therefore, with \( \beta = \min_{1 \leq i \leq K} t_i \), we have \( \inf_{t \leq \beta} \|z(t)\|_1 > 0 \) and \( z_i(\beta) > 0, 1 \leq i \leq K \).

Suppose now that \( z_i(0) > 0, 1 \leq i \leq K \). We show that there exists a constant \( \alpha^* > 0 \) such that any non-negative solution \( z^{(T)}(\cdot) \) to (3) that is defined in an interval
[0, T] can be continued onto [0, T + α*] remaining non-negative (α* does not depend on T and z(0)). This will complete the proof. Define the rectangle
\[ B^* = \left\{ 0 < l_i(z(0))/2 \leq z_i \leq u_i(z(0)) + l_i(z(0))/2, \ 1 \leq i \leq K \right\} \]
and the constants
\[ M^* = \|f\|_{B^*}, \quad b^* = \min_{1 \leq i \leq k} l_i(z(0))/2, \quad \alpha^* = b^*/M^*. \]

Consider T = 0. Let \( B_0 = \{ \max_{1 \leq i \leq K} |z_i - z_i(0)| \leq b^* \} \); then \( M^* \geq \|f\|_{B_0} \) because \( B_0 \subseteq B^* \). By the Peano theorem, there exists a solution to (3) in the interval \([0, \alpha^*]\), and it is non-negative because \( B_0 \subseteq \mathbb{R}^K_+ \). Consider \( T > 0 \) and a non-negative solution \( z(T) \) to (3) defined in \([0, T]\). By the bounds (9) and (10), we have \( l_i(z(0)) \leq z_i(T)(T) \leq u_i(z(0)), \ i = 1, \ldots, K \). Let \( B_T = \{ \max_{1 \leq i \leq K} |z_i - z_i(T)(T)| \leq b^* \} \); then \( M^* \geq \|f\|_{B_T} \) because \( B_T \subseteq B^* \). By the Peano theorem, in the interval \([0, \alpha^*]\), there exists a solution \( y(T) \) to the Cauchy problem
\[
y'(t) = \Lambda - p^* y(t) - e^{-1} m(y(t)), \\
y(0) = z(T)(T),
\]
and it is non-negative because \( B_T \subseteq \mathbb{R}^K_+ \). Then
\[ z(T+\alpha^*)(t) := \begin{cases} z(T)(t), & t \in [0, T], \\
y(T)(t - T), & t \in [T, T + \alpha^*], \end{cases} \]
is a non-negative solution to (3) in \([0, T + \alpha^*]\).

5.2 Zero initial state

This proof is based on the ideas from [4]. We introduce a family of auxiliary integral equations parameterised by \((\varepsilon, a) \in \mathbb{R}_+ \times \mathbb{R}^K_+ \). By further Lemma 2, the equation with parameters \((0, p)\) is equivalent to the fluid model equation with zero initial condition. By Property 1, each auxiliary equation has a solution. By Property 3, for any \( a \in \mathbb{R}^K_+ \), a solution to the equation with parameters \((0, a)\) is unique.

**Lemma 2** (Equivalent description of the fluid model) *For any initial state \( z^0 \), the set of fluid model solutions coincides with the set of functions from \( G \) that solve the following system of integral equations: for \( i = 1, \ldots, K, \ t \geq 0, \)
\[
z_i(t) = z_i^0 \exp \left( -p_i t - \int_0^t \frac{e^{-1}}{\|z(s)\|_1} ds \right) \\
+ \Lambda_i \int_0^t \exp \left( -p_i (t - s) - \int_s^t \frac{e^{-1}}{\|z(x)\|_1} dx \right) ds. \tag{11}
\]
Proof As we differentiate (11), the fluid model equation (3) follows. We now show that (3) implies (11). Let \( z(\cdot) \) be an FMS with initial state \( z^0 \) and consider the following Cauchy problem with respect to \( u(\cdot) \): for \( i = 1, \ldots, K \),
\[
\begin{align*}
    u_i'(t) &= \Lambda_i - \left( p_i + \frac{e^{-1}}{\|z(t)\|_1} \right) u_i(t), \quad t > 0, \\
    u_i(0) &= z^0_i.
\end{align*}
\] (12)

If (12) has a continuous solution, it must be unique. Indeed, suppose that \( u(\cdot), \tilde{u}(\cdot) \) are two continuous solutions to (12). Let \( w(\cdot) = u(\cdot) - \tilde{u}(\cdot) \). Then, for \( i = 1, \ldots, K \),
\[
\begin{align*}
    w_i'(t) &= -\left( p_i + \frac{e^{-1}}{\|z(t)\|_1} \right) w_i(t), \quad t > 0, \\
    w_i(0) &= 0.
\end{align*}
\]
Lemma 1 applied to \( x(\cdot) = w_i(\cdot) \) and \( C = 0 \), \( x(\cdot) = -w_i(\cdot) \) and \( C = 0 \), \( i = 1, \ldots, K \), implies that \( w(\cdot) \equiv 0 \).

Finally, \( z(\cdot) \) is a solution to (12) by (3). For \( i = 1, \ldots, K, t \geq 0 \), put \( f_i(t) \) to be the RHS of (11). Differentiating of \( f(\cdot) \) implies that it is a solution to (12), too. Since (12) has a unique continuous solution, \( z(\cdot) \) and \( f(\cdot) \) coincide, and hence \( z(\cdot) \) is a solution to (11). \( \square \)

Auxiliary fluid model solutions and their properties For each \( (\varepsilon, a) \in \mathbb{R}^+ \times \mathbb{R}^K_+ \), introduce the operator \( F^{(\varepsilon, a)} : \mathcal{G} \to \mathcal{G} \) defined by
\[
F^{(\varepsilon, a)}_i(u)(t) = \varepsilon + \Lambda_i \int_0^t \exp\left(-a_i(t-s) - \int_s^t \frac{e^{-1}}{\|u(x)\|_1} \, dx \right) \, ds, 
\] 
\[ t \geq 0, \quad i = 1, \ldots, K. \]

Definition 3 Let \( (\varepsilon, a) \in \mathbb{R}^+ \times \mathbb{R}^K_+ \). A function \( z(\cdot) \in \mathcal{G} \) solving the integral equation
\[
z(t) = F^{(\varepsilon, a)}(z)(t), \quad t \geq 0, \] (13)

is called an \( (\varepsilon, a) \)-fluid model solution (for short, we write “\( (\varepsilon, a) \)-FMS”).

Further we establish a number of properties of the auxiliary fluid model solutions defined above, including the existence and uniqueness of an \( (\varepsilon, a) \)-FMS for any \( (\varepsilon, a) \in \mathbb{R}^+ \times \mathbb{R}^K_+ \).

For \( a \in \mathbb{R}^K_+ \), put
\[
a^* = \max_{1 \leq i \leq K} a_i, \quad a_* = \min_{1 \leq i \leq K} a_i, \quad a^u = (a^*, \ldots, a^*), \quad a^l = (a_*, \ldots, a_*).
\]

Property 1 In what follows, \( z^{(\varepsilon, a)}(\cdot) \) denotes an \( (\varepsilon, a) \)-FMS.

(i) For any \( (\varepsilon, a) \in \mathbb{R}^+ \times \mathbb{R}^K_+ \), there exists an \( (\varepsilon, a) \)-FMS.
(ii) If \( \varepsilon > 0 \), then an \( (\varepsilon, a) \)-FMS is unique.
If $a_1 = \cdots = a_K$, then a $(0, a)$-FMS is unique and given by
\[
(z^{(0,a)}(t) = \begin{cases}
\frac{A_i}{\|A\|_1}(\|A\|_1 - e^{-1})t & \text{if } a_1 = 0, \\
\frac{A_i}{\|A\|_1}e^{-1}a_1(1 - e^{-a_1t}) & \text{if } a_1 > 0,
\end{cases}
\]

(iv) If $\varepsilon > \delta \geq 0$, then $z^{(\varepsilon,a)}(t) \geq z^{(\delta,a)}(t)$, $t \geq 0$.

(v) A $(0, a)$-FMS admits the bounds $z^{(0,a)}(t) \leq z^{(\varepsilon,a)}(t) \leq z^{(0,a)}(t)$, $t \geq 0$.

Proof (i) for $\varepsilon > 0$. Put $z^0(\cdot) = (\varepsilon, \ldots, \varepsilon)$ and $z^{n+1}(\cdot) = F(\varepsilon,a)(z^n)(\cdot)$, $n \geq 0$. Then $z^1(\cdot) \geq (\varepsilon, \ldots, \varepsilon) = z^0(t)$, $t \geq 0$. The operator $F(\varepsilon,a)$ is monotone, that is, $u(t) \geq v(t)$ for all $t \geq 0$ implies $F(\varepsilon,a)(u)(t) \geq F(\varepsilon,a)(v)(t)$ for all $t \geq 0$. Also $F(\varepsilon,a)(u)(t) \leq (\varepsilon, \ldots, \varepsilon) + tA$ for all $u(\cdot) \in \mathcal{G}$ and all $t \geq 0$. Hence, for each $t \geq 0$, the sequence $\{z^n(t); n \geq 0\}$ is non-decreasing and bounded from above, and there exists the pointwise limit of $z^n(\cdot)$ as $n \to \infty$; denote it by $z(\cdot)$. Now we show that $z(\cdot)$ is an $(\varepsilon,a)$-FMS. Take an arbitrary $t \geq 0$. For all $n \geq 0$, we have $z^n(\cdot) \geq (\varepsilon, \ldots, \varepsilon)$, and hence $1/\|z^n(\cdot)\|_1 \leq 1/(K\varepsilon)$. Then, by the dominated convergence theorem, for all $s \in [0, t]$, $\int_s^t e^{-1}/\|z^n(x)\|_1 dx \to \int_s^t e^{-1}/\|z(x)\|_1 dx$ as $n \to \infty$. The last argument, together with $\exp(-a_1(t - s) - \int_s^t e^{-1}/\|z^n(x)\|_1 dx) \leq 1$, $s \in [0, t]$, and, again, the dominated convergence theorem, implies that $F(\varepsilon,a)(z^n)(t) \to F(\varepsilon,a)(z)(t)$ as $n \to \infty$. So, indeed, $z(\cdot)$ satisfies (13) for all $t \geq 0$.

(ii) Let $z(\cdot)$ and $\tilde{z}(\cdot)$ be two $(\varepsilon,a)$-FMS’s. Take an arbitrary $T > 0$. Since $a_i(t - s) + \int_s^t e^{-1}/\|z(x)\|_1 dx \leq (\|a\| + (eK\varepsilon)^{-1})T$, $0 \leq s \leq t \leq T$, and the same is true for $\tilde{z}(\cdot)$, by the Lipschitz continuity of $\exp(\cdot)$ on compact sets, there exists a constant $\alpha(T)$ such that, for $t \leq T$,
\[
\|z(t) - \tilde{z}(t)\|_1 \leq \alpha(T)e^{-1}\int_0^t \int_s^t |1/\|z(x)\|_1 - 1/\|\tilde{z}(x)\|_1| dx ds
\leq \alpha(T)e^{-1}T \int_0^t |1/\|z(x)\|_1 - 1/\|\tilde{z}(x)\|_1| dx.
\]

Then, by the Lipschitz continuity of $1/\|\cdot\|_1$ on the set $\mathbb{R}^K$ and by the Gronwall inequality, $z(\cdot)$ and $\tilde{z}(\cdot)$ must coincide in all finite intervals $[0, T]$, and hence they coincide on $[0, \infty)$.

(iii) Due to Lemma 2, $(0, a)$-FMSs are defined by the Cauchy problem
\[
\begin{align*}
z'(t) & = A - a_1z(t) - e^{-1}z(t)/\|z(t)\|_1, \quad t > 0, \\
z(0) & = 0.
\end{align*}
\]

Summing up its coordinates, we get the Cauchy problem
\[
\begin{align*}
\|z(t)\|_1 &= (\|A\|_1 - e^{-1}) - a_1\|z(t)\|_1, \quad t > 0, \\
\|z(0)\|_1 &= 0,
\end{align*}
\]

which admits a unique solution
\[
\|z(t)\|_1 = \begin{cases}
(\|A\|_1 - e^{-1})t, & \text{if } a_1 = 0, \\
(\|A\|_1 - e^{-1})(1 - e^{-a_1t})/a_1, & \text{if } a_1 > 0.
\end{cases}
\]
In the case of \( \varepsilon = 0 \) and \( a_1 = \cdots = a_K \), (13) implies that \( z(\cdot)/\|z(\cdot)\|_1 \equiv A/\|A\|_1 \). Then the unique \((0,a)\)-FMS is given by (14).

(i) for \( \varepsilon = 0 \), and (v) In order to prove the existence, consider the sequence \( z(\cdot) := z(0,a^v)(\cdot), \ z_n(\cdot) := F(0,a^v)(z_n(\cdot)), n \geq 0 \). By reasoning analogous to that in the case of \( \varepsilon > 0 \), the point-wise limit of this sequence is a \((0,a)\)-FMS. To prove the lower bound, consider the sequence \( z(\cdot) := z(0,a^v)(\cdot), \ z_{n+1}(\cdot) := F(0,a^v)(z_n(\cdot)), n \geq 0 \). It is non-increasing in \( n \), and its point-wise limit is the \((0,a^v)\)-FMS. Then \( z(0,a^v)(t) \leq z(0)(t) = z(0,a)(t) \) for all \( t \geq 0 \). Similarly, the upper bound holds.

(iv) Consider the sequence \( z(t) := z(0,a^v)(\cdot), \ z_{n+1}(\cdot) := F(\varepsilon,a)(z_n(\cdot)), n \geq 0 \). It is non-decreasing in \( n \), and its point-wise limit is the \((\varepsilon,a)\)-FMS. Then \( z(\varepsilon,a)(t) \geq z(0)(t) = z(\delta,a)(t) \) for all \( t \geq 0 \).

We proceed with properties of \((0,a)\)-FMS’s at \( t = 0 \) (cf. Remark 4).

**Property 2** For any \((0,a)\)-FMS \( z(0,a)(\cdot) \), its right derivative at \( t = 0 \) is well defined, and \( z(0,a)(t) \) is non-decreasing in \( \varepsilon \) for \( 0, a^* > a_\ast \) and \( a^* > a_\ast = 0 \). By Property 1, (iii) and (v), for \( i = 1, \ldots, K, t \geq 0 \),

\[
\frac{A_i}{\|A\|_1} \left( \left\|A\right\|_1 - e^{-1} \right) \leq \frac{A_i}{\|A\|_1} \frac{\|A\|_1 - e^{-1} \right) a_\ast (1 - e^{-a_\ast t}).
\]

Then, for any sequence \( t_n \to 0, n \to \infty \),

\[
\limsup_{n \to \infty} \frac{z(0,a)(t_n)}{t_n} \leq \frac{A_i}{\|A\|_1} \frac{\|A\|_1 - e^{-1}}{a_\ast t_n} = \frac{A_i}{\|A\|_1} \left( \frac{\|A\|_1 - e^{-1}}{a_\ast t_n} \right).
\]

Similarly, \( \liminf_{n \to \infty} z(0,a)(t_n)/t_n \geq (A_i/\|A\|_1)(\|A\|_1 - e^{-1}) \). Hence, the derivative exists. The second result follows from Taylor’s expansion, as was discussed in Remark 4.

Finally, we show uniqueness of a \((0,a)\)-FMS by estimating it via the auxiliary FMSs with other parameters.

**Property 3** Fix \( a \in \mathbb{R}_+^K \). For short, \( z^\varepsilon \) denotes an \((\varepsilon,a)\)-FMS.

(i) For \( \varepsilon > 0 \) and the function \( \varphi_\varepsilon(t) \) := \( \int_0^t K e^{-1}/\|z(\varepsilon)(s)\|_1 \) \( ds \),

\[
\|z(\varepsilon)(t) - z(0)(t)\|_1 \leq \varepsilon (K + \|a\|_1 + \varphi_\varepsilon(t)), \quad t \geq 0.
\]

(ii) If \( \varepsilon > 0, \varepsilon \to 0 \), then \( \varepsilon \varphi_\varepsilon(t) \to 0 \) for any \( t > 0 \).

(iii) \((0,a)\)-FMS is unique.
Proof (i) Let $\varepsilon \geq 0$. By differentiating (13), we get, for $i = 1, \ldots, K$,

\[
(z_\varepsilon^i)'(t) = \Lambda_i - \left(a_i + \frac{e^{-1}}{\|z_\varepsilon(t)\|_1}\right)(z_\varepsilon^i(t) - \varepsilon), \quad t > 0, \\
z_\varepsilon^i(0) = \varepsilon.
\]

Then integrating over $[0, t]$ yields

\[
z_\varepsilon^i(t) - \varepsilon = \Lambda_i t - \int_0^t \left(a_i + \frac{e^{-1}}{\|z_\varepsilon(s)\|_1}\right)(z_\varepsilon^i(s) - \varepsilon) ds, \quad t \geq 0,
\]

which, after taking the sum in all coordinates, can be rewritten as

\[
\sum_{i=1}^K z_\varepsilon^i(t) + \sum_{i=1}^K \int_0^t a_i z_\varepsilon^i(s) ds = (\|\Lambda\|_1 - e^{-1})t + \varepsilon K + \varepsilon a \|1\| t + \varepsilon \varphi_\varepsilon(t), \quad t \geq 0.
\]

The last equation implies that, for $\varepsilon > 0$,

\[
\sum_{i=1}^K (z_\varepsilon^i(t) - z_\varepsilon^0(t)) + \sum_{i=1}^K \int_0^t a_i (z_\varepsilon^i(s) - z_\varepsilon^0(s)) ds \\
= \varepsilon K + \varepsilon a \|1\| t + \varepsilon \varphi_\varepsilon(t), \quad t \geq 0.
\]

Due to Property 1, (v), for $\varepsilon > 0$, both sums in the LHS of (16) have non-negative summands. By omitting the second sum, we obtain the bound (15).

(ii) Suppose that $a^* > 0$ (the other case can be treated similarly). Property 1, (iv) and (v), together with $\|z_\varepsilon(\cdot)\|_1 \geq K\varepsilon$, implies that, for $\varepsilon > 0$,

\[
\varphi_\varepsilon(t) = \int_0^t Ke^{-1} \|z_\varepsilon(s)\|_1 ds \leq \int_0^t \frac{Ke^{-1}}{\max\{K\varepsilon, \|z_{(\varepsilon,a^*)}(s)\|_1\}} ds.
\]

By (14), $\|z_{(\varepsilon,a^*)}(s)\|_1 \leq K\varepsilon$ if and only if $s \leq f(\varepsilon) := \frac{1}{a^*} \ln \frac{\|\Lambda\|_1 - e^{-1}}{\|\Lambda\|_1 - e^{-1} - Ke^*}$. We have $f(\varepsilon) > 0$ for $\varepsilon$ small enough and $f(\varepsilon) \to 0$, $\varepsilon \to 0$. Put $\beta = Ke^{-1}/(\|\Lambda\|_1 - e^{-1})$; then

\[
\varepsilon \varphi_\varepsilon(t) = e^{-1} f(\varepsilon) + \varepsilon \beta a^* \int_{f(\varepsilon)}^t \frac{1}{1 - e^{-a^*s}} ds \\
= e^{-1} f(\varepsilon) + \varepsilon \beta a^* \left(t - f(\varepsilon) + \frac{1}{a^*} \ln(1 - e^{-a^*t})\right) \\
- \varepsilon \frac{\beta}{f(\varepsilon)} \left(f(\varepsilon) \ln(1 - e^{-a^*f(\varepsilon)})\right).
\]

In the very RHS of the last equation, convergence of the first two summands to 0 as $\varepsilon \to 0$ is clear. The first multiplier of the last summand tends to a finite constant, and the second multiplier tends to 0.
(iii) Suppose that \( z^0(\cdot) \) and \( \tilde{z}^0(\cdot) \) are two \((0, \pi)\)-FMS’s. For any \( t > 0 \), by (i) and (ii), \( \| z^0(t) - \tilde{z}^0(t) \|_1 \leq \| z^0(t) - z^\varepsilon(t) \|_1 + \| z^\varepsilon(t) - \tilde{z}^0(t) \|_1 \to 0 \) as \( \varepsilon \to 0 \). Hence, \( z^0(\cdot) \) and \( \tilde{z}^0(\cdot) \) must coincide in \([0, \infty)\).

\[ \square \]

6 Proof of Theorem 2

Existence and uniqueness. The function \( f(s) = \sum_{i=1}^{K} p_i \Lambda_i / (s + p_i) \) is continuous and strictly decreasing in \((0, \infty)\), and takes all values between \( \|\Lambda\|_1 \) and 0 as \( s \) goes along \((0, \infty)\). Then, by Assumption 2, there exists a unique \( z^e \) satisfying (5), and all its coordinates are positive. Equilibrium FMSs are defined by the equation

\[ \Lambda - p * z^e - e^{-1} z^e / \|z^e\|_1 = 0. \] (17)

In order to prove the first part of the theorem, we have to check that (5) is a solution to (17) and that, if there is a solution to (17), then it is necessarily (5).

By plugging (5) into (17) multiplied by \( \|z^e\|_1 \), we obtain, for \( i = 1, \ldots, K \),

\[
\left( \Lambda_i - \frac{p_i \Lambda_i}{x + p_i} \right) \sum_{j=1}^{K} \frac{\Lambda_j}{x + p_j} - \frac{e^{-1} \Lambda_i}{x + p_i} = \frac{\Lambda_i x}{x + p_i} \sum_{j=1}^{K} \frac{\Lambda_j}{x + p_j} - \frac{e^{-1} \Lambda_i}{x + p_i} = \frac{\Lambda_i}{x + p_i} \left( \sum_{j=1}^{K} \frac{\Lambda_j x}{x + p_j} - e^{-1} \right) = \frac{\Lambda_i}{x + p_i} \left( \sum_{j=1}^{K} \Lambda_j - \sum_{j=1}^{K} \frac{\Lambda_j p_j}{x + p_j} - e^{-1} \right) = 0.
\]

So, indeed, (5) is an equilibrium FMS.

Suppose now that \( z^e \) is a solution to (17). By solving coordinate \( i \) of (17) with respect to \( z^e_i \), we get

\[ z^e_i = \frac{\Lambda_i}{p_i + e^{-1}/\|z^e\|_1} \quad i = 1, \ldots, K. \]

Plug the last relation into the sum of the coordinates of (17); then

\[ \sum_{i=1}^{K} \frac{p_i \Lambda_i}{p_i + e^{-1}/\|z^e\|_1} = \|\Lambda\|_1 - e^{-1}, \]

Hence, \( z^e \) satisfies (5) with \( x = e^{-1}/\|z^e\|_1 \).
Stability. When proving the second part of the theorem, we refer to the following result (see [9, Chap. XIV, Paragraph 11]).

Consider an autonomous system of differential equations

\[ z'(t) = f(z(t)), \quad t \geq 0, \quad z, f \in \mathbb{R}^K. \]  \tag{18}

**Proposition 3**  
Suppose that the function \( f(z) \) is continuous in an open set \( E \subseteq \mathbb{R}^K \), and that, for any \( z(0) \in E \), there exists a unique solution \( z(\cdot) \) to system (18) and \( z(t) \in E \) for all \( t \geq 0 \). Suppose also that there exists a non-negative continuously differentiable in \( E \) function \( V(z) \) such that \( V(z) \to \infty \) as \( \|z\| \to \infty \), and the trajectory derivative of \( V(z) \) with respect to system (18), defined by

\[ V'(z) = \sum_{i=1}^{K} f_i(z) \frac{\partial V}{\partial z_i}(z), \]

is non-positive. If there exists a unique point \( \tilde{z} \) where \( V'(\tilde{z}) = 0 \), then the solution \( \tilde{z}(t) \equiv \tilde{z} \) of system (18) is asymptotically stable in \( E \); i.e. any solution to (18) with initial condition in \( E \) converges to \( \tilde{z} \) as \( t \to \infty \).

By Remark 7 and Property 1, (iii) and (v), any FMS at any time \( t > 0 \) has all coordinates positive. Then it suffices to show convergence to the equilibrium point for FMSs that start in the interior of \( \mathbb{R}^K_+ \). This, in turn, follows from Proposition 3 with the open set \( E = (0, \infty)^K \), (3) and the Lyapunov function

\[ V(z) = \sum_{i=1}^{K} \frac{y_i^2}{z^e_i}, \quad \text{where } y_i := z_i - z^e_i. \]

By plugging \( y_i \)'s into (3), we get

\[ z'_i = -p_i y_i + e^{-1} \frac{z^e_i}{\|z^e\|_1} - e^{-1} \frac{y_i + z^e_i}{\|y + z^e\|_1}, \quad i = 1, \ldots, K. \]

Then,

\[ V'(z) = \sum_{i=1}^{K} \frac{2y_i z'_i}{z^e_i} = -\sum_{i=1}^{K} \frac{2p_i y_i^2}{z^e_i} - \frac{2e^{-1}}{\|y + z^e\|_1} \left( \sum_{i=1}^{K} \frac{y_i^2}{z^e_i} - \frac{(\sum_{i=1}^{K} y_i)^2}{\sum_{i=1}^{K} z^e_i} \right). \]

We have to check that \( V'(z) \) is non-positive on \((0, \infty)^K \). For \( k = 1, \ldots, K \), let

\[ u_k(y) = \sum_{i=1}^{k} \frac{y_i^2}{z^e_i} - \frac{(\sum_{i=1}^{k} y_i)^2}{\sum_{i=1}^{k} z^e_i}. \]

In particular,

\[ V'(z) = -\sum_{i=1}^{K} \frac{2p_i y_i^2}{z^e_i} - \frac{2e^{-1}}{\|y + z^e\|_1} u_K(y). \]  \tag{19}

The quadratic form \( u_K(y) \) is non-negative on the set \( \{ y : y + z^e \in (0, \infty)^K \} \), which we can show by applying the Lagrange reduction to canonical form. We start from
separating out the terms with \( y_K \) and obtain

\[
u_K(y) = \frac{1}{\sum_{i=1}^{K} z_i^e} \left( \sqrt{ \frac{z_K^e}{\sum_{i=1}^{K-1} z_i^e} \sum_{i=1}^{K-1} z_i^e y_i - \sum_{i=1}^{K-1} z_i^e y_K } \right)^2 + \nu_{K-1}(y).
\]

Iterating the procedure for \( y_{K-1}, \ldots, y_2 \), we transform \( u_K(y) \) into a sum of squares. Then, by (19), \( V'(z) \) is non-positive on \((0, \infty)^K\).

### 7 Proof of Theorem 3

The proof is organised as follows. Section 7.1 contains a representation of the workload process before and after the fluid scaling. In Sect. 7.2, we formulate two auxiliary results (Lemmas 3 and 4), and then proceed with the proof of relative compactness of the family \( \{WR(\cdot); \; R > 0\} \) and weak convergence (as \( R \to \infty \)) along subsequences from this family. Proofs of Lemmas 3 and 4 are given in Sects. 7.3 and 7.4, respectively.

#### 7.1 Workload representation

Throughout the proof, unless otherwise stated, we use the following representation of the workload dynamics:

\[
W^R(n+1) = W^R(n) + A(n+1) - T(n, W^R(n)) - I^R(n, W^R(n)),
\]

where, for \( x \in \mathbb{Z}_+^K \) and \( i = 1, \ldots, K \),

\[
p_i(x) = \begin{cases} B(x_i, 1/\|x\|_1)(\{1\}) \prod_{j \neq i} B(x_j, 1/\|x\|_1)(\{0\}) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0, \end{cases}
\]

\[
T_i(n, x) = \mathbbm{1}\left\{ \sum_{j=1}^{i-1} p_j(x) \leq U(n) < \sum_{j=1}^{i} p_j(x) \right\},
\]

\[
I_i^R(n, x) = \sum_{j=1}^{x_i - T_i(n, x)} \xi_i^R(n, j),
\]

and

- \( \{U(n); \; n \geq 0\} \) is an i.i.d. sequence, and \( U(0) \) is distributed uniformly over \([0, 1]\),
- \( \{\xi_i^R(n, j); \; j \geq 1, \; n \geq 0, \; i = 1, \ldots, K\} \) are independent i.i.d. sequences of Bernoulli r.v.’s, and \( P[\xi_i^R(n, 1) = 1] = p_i/R = 1 - P[\xi_i^R(n, 1) = 0] \),
- the sequences \( \{A(n); \; n \geq 1\} \), \( \{U(n); \; n \geq 0\} \) and \( \{\xi_i^R(n, j); \; j \geq 1, \; n \geq 0, \; i = 1, \ldots, K\} \) are mutually independent and also do not depend on the initial condition \( W^R(0) \).
For short, we put

\[ h(x) = \begin{cases} 
0, & x = 0, \\
(1 - 1/x)^{x-1}, & x \geq 1. 
\end{cases} \]

Then, in particular,

\[ p_i(x) = h(\|x\|_1)m_i(x), \]

\[ E[T(i, W^R(i))|W^R(i)] = h(\|W^R(i)\|_1)m(W^R(i)). \]

\[ E[I^R(i, W^R(i))|W^R(i)] = \frac{p}{R} * (W^R(i) - h(\|W^R(i)\|_1)m(W^R(i))). \]

We now transform the workload dynamics into an integral equation that, as we show later, differs from the fluid model equation (2) by the terms that vanish as \( R \to \infty \). For any \( n \in \mathbb{Z}_+ \), we have

\[
W^R(n) = W^R(0) + \sum_{i=1}^{n} A(i) - \sum_{i=0}^{n-1} T(i, W^R(i)) - \sum_{i=0}^{n-1} I^R(i, W^R(i)) \\
= W^R(0) + n \Lambda - \sum_{i=0}^{n-1} h(\|W^R(i)\|_1)m(W^R(i)) \\
- \frac{p}{R} \sum_{i=0}^{n-1} (W^R(i) - h(\|W^R(i)\|_1)m(W^R(i))) + M^R(n), \tag{21}
\]

where the sequence \( \{M^R(n); n \geq 0\} \) forms a zero-mean martingale since

\[
M^R(n) = \sum_{i=1}^{n} (A(i) - E(A(i)) - \sum_{i=0}^{n-1} (T(i, W^R(i)) - E[T(i, W^R(i))|W^R(i)]) \\
- \sum_{i=0}^{n-1} (I^R(i, W^R(i)) - E[I^R(i, W^R(i))|W^R(i)])].
\]

Introduce the fluid scaled version of the martingale \( \{M^R(n); n \geq 0\} \) analogous to that of the workload process:

\[
\overline{W}^R(t) = \frac{1}{R} W^R(\lfloor Rt \rfloor), \quad \overline{M}^R(t) = \frac{1}{R} M^R(\lfloor Rt \rfloor).
\]

Then (21) turns into the integral equation

\[
\overline{W}^R(t) = \overline{W}^R(0) + \frac{\lfloor Rt \rfloor}{R} \Lambda - \left( (1, \ldots, 1) - \frac{p}{R} \right) \\
\times \int_0^{\lfloor Rt \rfloor/R} h(R\|\overline{W}^R(s)\|_1)m(\overline{W}^R(s)) \, ds \\
- p \int_0^{\lfloor Rt \rfloor/R} \overline{W}^R(s) \, ds + \overline{M}^R(t). \tag{22}
\]
Finally, we rewrite (22) as
\[
W^R(t) = W^R(0) + t\Lambda - e^{-1}\int_0^t m(W^R(s)) \, ds - p\int_0^t W^R(s) \, ds + G^R(t),
\]
where
\[
G^R(t) = M^R(t) + G^{1,R}(t) + G^{2,R}(t) + G^{3,R}(t),
\]
\[
G^{1,R}(t) = \left(\frac{\lfloor R_t \rfloor}{R} - t\right)\Lambda,
\]
\[
G^{2,R}(t) = e^{-1}\int_0^t m(W^R(s)) \, ds - (1, \ldots, 1) - \frac{p}{R} \int_0^{\lfloor R_t \rfloor/R} h(R\|W^R(s)\|_1)m(W^R(s)) \, ds,
\]
\[
G^{3,R}(t) = p\int_0^{\lfloor R_t \rfloor/R} W^R(s) \, ds.
\]

7.2 Relative compactness and limiting equations

We first discuss convergence \(G^R(\cdot) \Rightarrow (0, \ldots, 0)\) as \(R \to \infty\) in \(\mathcal{D}\). By Remark 5, \(G^{1,R}(\cdot) \Rightarrow (0, \ldots, 0)\) as \(R \to \infty\). Weak convergence to zero of the three other summands in \(G^R(\cdot)\) follows from Lemmas 3 and 4.

**Lemma 3** Let Assumptions 1 and 2 hold. Then
(i) for any \(\delta > 0\) and \(\varepsilon > 0\), there exists a \(\gamma = \gamma(\delta, \varepsilon) > 0\) such that
\[
\liminf_{R \to \infty} P\{\varphi^R(\gamma R) \leq \delta R\} \geq 1 - \varepsilon,
\]
where \(\varphi^R(\gamma R) := \inf\{n \geq 0 : W^R(n) \geq \gamma R\}\),
(ii) for any \(\varepsilon > 0\) and \(\Delta > \delta > 0\), there exists a \(C = C(\varepsilon, \delta, \Delta) > 0\) such that
\[
\liminf_{R \to \infty} P\left\{\inf_{\delta \leq t \leq \Delta} \|W^R(t)\|_1 \geq C\right\} \geq 1 - \varepsilon,
\]
(iii) \(G^{2,R}(\cdot) \Rightarrow (0, \ldots, 0)\) in \(\mathcal{D}\) as \(R \to \infty\).

**Lemma 4** Suppose \(W^R(0) \Rightarrow z^0\) as \(R \to \infty\). Then \(G^{3,R}(\cdot) \Rightarrow (0, \ldots, 0)\) and \(M^R(\cdot) \Rightarrow (0, \ldots, 0)\) in \(\mathcal{D}\) as \(R \to \infty\).

Now we are in a position to prove the theorem. The proof consists of two steps. First, we establish the relative compactness of a sequence \(\{W^R(\cdot) ; R \to \infty\}\) such that \(W^R(0) \Rightarrow z^0\) (note that the assumption of \(z^0\) being deterministic is not essential here; it is used in the next step). Second, we show that all weak limits of such a sequence a.s. coincide with the FMS starting from \(z^0\), which means the convergence of the sequence to the FMS starting from \(z^0\).
Relative compactness. We apply the following known result (see [6, Chap. 3, Corollary 7.4 and Theorem 10.2(a)]).

Proposition 4 Let \( \{X^n(\cdot); n \geq 1\} \) be a sequence of processes with sample paths in \( \mathcal{D} \). Suppose that, for any \( \eta > 0 \) and \( t \geq 0 \), there exists a compact set \( \Gamma_{\eta,t} \subset \mathbb{R}^K \) such that

\[
\liminf_{n \to \infty} P\{X^n(t) \in \Gamma_{\eta,t}\} \geq 1 - \eta.
\]

Suppose further that, for any \( \eta > 0 \) and \( T > 0 \), there exists a \( \delta > 0 \) such that

\[
\limsup_{n \to \infty} P\{\omega(X^n, \delta, T) \geq \eta\} \leq \eta,
\]

where \( \omega(x, \delta, T) := \max_{1 \leq i \leq K} \sup_{s \leq t \leq T} \{|x_i(s) - x_i(t)| : s, t \in [0, T], |t - s| < \delta\} \). Then the sequence \( \{X^n(\cdot); n \geq 1\} \) is relatively compact and all its weak limit points are a.s. continuous.

Consider a sequence \( \{\overline{W}^R(\cdot); R \to \infty\} \) such that \( \overline{W}^R(0) \Rightarrow z^0 \). By (20), we have \( 0 \leq \overline{W}^R(t) \leq \overline{W}^R(0) + \sum_{n=1}^{[Rt]} A(n)/R \Rightarrow z^0 + t\Lambda \), and the first condition of the proposition holds. Now we check the second condition. Equation (23) implies, for \( t > s, i = 1, \ldots, K \),

\[
\overline{W}^R_i(t) - \overline{W}^R_i(s) = \Lambda_i(t - s) - e^{-1} \int_s^t m_i(\overline{W}^R(x)) \, dx - p_i \int_s^t \overline{W}^R_i(x) \, dx + G^R_i(t) - G^R_i(s).
\]

Since \( m_i(\cdot) \leq 1 \), we have, for \( t, s \in [0, T], |t - s| < \delta \),

\[
|\overline{W}^R_i(t) - \overline{W}^R_i(s)| \leq \Lambda_i \delta + e^{-1} \delta + p_i \delta \|\overline{W}^R_i\|_{[0,T]} + 2 \|G^R_i\|_{[0,T]}.
\]

Equation (23) also implies the following upper bound for the workload process:

\[
\overline{W}^R_i(t) \leq \overline{W}^R_i(0) + \Lambda_i t + G^R_i(t) \leq \|\overline{W}^R(0)\| + \|\Lambda\| T + \|G^R\|_{[0,T]}.
\]

By (25) and (26),

\[
\omega(\overline{W}^R(\cdot), \delta, T) \leq (\|\Lambda\| + e^{-1}) \delta + \|p\| \delta (\|\overline{W}^R(0)\| + \|\Lambda\| T + \|G^R\|_{[0,T]}) + 2 \|G^R\|_{[0,T]},
\]

which implies that the second condition of Proposition 4 holds, since \( \overline{W}^R(0) \Rightarrow z^0 \) and, by Remark 5, \( \|G^R\|_{[0,T]} \Rightarrow 0 \).

Limiting equations. Now we show that, if \( \{\overline{W}^R(\cdot); R \to \infty\} \) converges weakly in \( \mathcal{D} \), then its limit point \( \overline{W}(\cdot) \) a.s.

(i) is continuous,
(ii) does not vanish at \( t > 0 \),
(iii) satisfies the fluid model equation (2).

Then, by the uniqueness of FMSs, \( \tilde{W}(\cdot) \) a.s. coincides with the FMS starting from \( z^0 \).

(i) The continuity of \( \tilde{W}(\cdot) \) follows from Proposition 4.
(ii) By Lemma 3, for any \( \varepsilon > 0 \) and \( n \geq 1 \), there exists a constant \( \tilde{C}(\varepsilon, 1/n) > 0 \) such that
\[
\liminf_{R \to \infty} P\{\inf_{1/n \leq t \leq n} \|\tilde{W}(t)\|_1 \geq \tilde{C}(\varepsilon, 1/n)\} \geq 1 - \varepsilon.
\]
In the interval \( (0, \tilde{C}(\varepsilon, 1/n)] \), choose a continuity point \( \tilde{C}(\varepsilon, 1/n) \) for the distribution of \( \inf_{1/n \leq t \leq n} \|\tilde{W}(t)\|_1 \). The mapping \( x(\cdot) \to \inf_{1/n \leq t \leq n} \|x(t)\| \) is continuous in \( D \); hence,
\[
1 - \varepsilon \leq \lim_{R \to \infty} P\{\inf_{1/n \leq t \leq n} \|\tilde{W}(t)\|_1 \geq \tilde{C}(\varepsilon, 1/n)\} \leq P\{\inf_{1/n \leq t \leq n} \|\tilde{W}(t)\|_1 > 0\}.
\]

By taking the limits as \( \varepsilon \to 0 \) in the last inequality, we obtain
\[
P\{\inf_{1/n \leq t \leq n} \|\tilde{W}(t)\|_1 > 0\} = 1 \quad \text{for any } n \geq 1,
\]
which, in turn, implies that
\[
P\{\|\tilde{W}(t)\|_1 > 0 \text{ for all } t > 0\} = 1. \quad (27)
\]

(iii) Fix \( t \geq 0 \). We introduce the mappings \( \varphi^1_t, \varphi^2_t : \mathbb{R}^K \to \mathbb{R}^K \) defined by
\[
\varphi^1_t(x) = x(t) + p^* \int_0^t x(s) ds, \quad \varphi^2_t(x) = e^{-1} \int_0^t m(x(s)) ds.
\]
By Remark 5, the mapping \( \varphi^1_t \) is continuous at any continuous \( x(\cdot) \). By \( m_i(\cdot) \leq 1 \), \( i = 1, \ldots, K \), and the dominated convergence theorem, the mapping \( \varphi^2_t \) is continuous at any continuous \( x(\cdot) \) that differs from zero everywhere except the points forming a set of zero Lebesgue measure. Then, by the continuity of \( \tilde{W}(\cdot) \), (27) and the continuous mapping theorem, we have
\[
\varphi^1_t(\tilde{W}^R) + \varphi^2_t(\tilde{W}^R) \Rightarrow \varphi^1_t(\tilde{W}) + \varphi^2_t(\tilde{W}).
\]

On the other hand, by (23),
\[
\varphi^1_t(\tilde{W}^R) + \varphi^2_t(\tilde{W}^R) = \tilde{W}^R(0) + tA + G^R(t) \Rightarrow z^0 + tA.
\]
Since \( z^0 + tA \) is deterministic, we have, for any fixed \( t \geq 0 \),
\[
P\{\tilde{W}(t) = z^0 + tA - e^{-1} \int_0^t m(\tilde{W}(s)) ds - p \int_s^t \tilde{W}(s) ds\} = 1.
\]
Let \( \Omega_t \) denote the event in the last equality. Then, again due to the continuity of \( \tilde{W}(\cdot) \) and (27),
\[
P\{\tilde{W}(\cdot) \text{ satisfies (2) in } [0, \infty)\} = P\{\bigcap \Omega_t \text{ over all rational } t \in [0, \infty)\} = 1.
\]
7.3 Proof of Lemma 3

We split the proof into four parts. In the first two parts we show that Assumptions 1 and 2 imply (i), and that (i) implies (ii), both for the single-class case. In the third part we show that the total workload $\|W(\cdot)\|_1$ of a multiclass model is bounded from below by that of a single-class model with suitable parameters. Then (i) and (ii) hold for the multiclass case, too. In the last part we show that (ii) implies (iii).

**Assumptions 1 and 2 imply (i), single class case**

For every $\gamma$ from an interval $(0, \gamma^*)$, we construct a Markov chain (see $\{V(\gamma)(n); n \geq 0\}$ below) that, for all $R$ large enough, is a lower bound for the workload process until the latter first hits the set $[\gamma R, \infty)$. Then we choose a $\gamma$ so as to have (i) with $\{V(\gamma)(n); n \geq 0\}$ in place of $\{W^{R}(n); n \geq 0\}$, and this completes the proof.

Throughout the proof, $\delta$ and $\epsilon$ are fixed.

Without loss of generality, we can assume that, for all $R$, a.s. $W^{R}(0) = 0$. Indeed, for all $n$ and $x \geq y$, $x - T(n, x) - I^{R}(n, x) \geq y - T(n, y) - I^{R}(n, y)$ a.s. (28)

Property (28) says that the process $\{W^{R}(n); n \geq 0\}$ admits path-wise monotonicity: the bigger is the initial value $W^{R}(0)$, the bigger are all the other values $W^{R}(n), n \geq 1$.

Further we make preparations needed to construct the lower-bound Markov chains. Let

$$h^{*} = e^{-1} + (\Lambda - e^{-1})/2 \quad \text{and} \quad B(k, 1/k)(1) \leq h^{*}, \quad k \geq N.$$

Let $\{B(n); n \geq 1\}$ be an i.i.d. sequence with $B(1) \sim B(N, p)$.

We apply the following proposition (see the Appendix for the proof) with $a = (\Lambda - e^{-1})/4$.

**Proposition 5**  For any $a > 0$, there exists a $\gamma^{*} = \gamma^{*}(a)$ and a family of r.v.'s $\{\theta_{\gamma}; 0 \leq \gamma \leq \gamma^{*}\}$ with the following properties:

1. the family $\{\theta_{\gamma}; 0 \leq \gamma \leq \gamma^{*}\}$ is uniformly integrable;
2. for any $\gamma \in [0, \gamma^{*}]$, $E\theta_{\gamma} \leq a$;
3. $\theta_{\gamma} \Rightarrow \theta_{0}$ as $\gamma \to 0$;
4. for any $\gamma \in (0, \gamma^{*}]$, there exists an $R_{\gamma}$ such that $\theta_{\gamma} \geq_{st} B^{R}_{\gamma} \sim B([\gamma R], p/R)$, $R \geq R_{\gamma}$.

For $\gamma \in (0, \gamma^{*}]$, let $\{\theta_{\gamma}(n); n \geq 1\}$ be an i.i.d. sequence with $\theta_{\gamma}(1) \overset{d}{=} \theta_{\gamma}$, and assume that this sequence does not depend on $\{B(n); n \geq 1\}$.

Now we construct the lower-bound Markov chains. For $R$ large enough, given $W^{R}(n) < N$,

$$W^{R}(n + 1) - W^{R}(n) \overset{a.s.}{\geq} A(n + 1) - 1 - \sum_{i=1}^{N} \xi^{R}(n, i) \geq_{st} A(1) - 1 - B(1).$$
and, given $N \leq W^R(n) < \gamma R$,

$$W^R(n + 1) - W^R(n) \geq \text{a.s.} A(n + 1) - I\{U(n) \leq h^*\} - \sum_{i=1}^{\lfloor \gamma R \rfloor} \xi^R(n,i)$$

$$\geq_{st} A(1) - I\{U(0) \leq h^*\} - \theta^R. \quad (31)$$

Introduce the two i.i.d. sequences:

$$x(n) = A(n) - 1 - B(n), \quad n \geq 1,$$

$$y^R(n) = A(n) - I\{U(n - 1) \leq h^*\} - \theta^R(n), \quad n \geq 1,$$

and the two auxiliary Markov chains:

$$V^R_{\gamma}(0) = 0,$$

$$V^R_{\gamma}(n + 1) = \begin{cases} \text{max}\{0, V^R_{\gamma}(n) + x(n + 1)\} & \text{if } V^R_{\gamma}(n) < N, \\
\text{max}\{0, V^R_{\gamma}(n) + y^R(n + 1)\} & \text{if } N \leq V^R_{\gamma}(n) < \gamma R, \\
V^R_{\gamma}(n) + A(n + 1) - T(n, V^R_{\gamma}(n)) - I^R(n, V^R_{\gamma}(n)) & \text{if } V^R_{\gamma}(n) \geq \gamma R, \\
\end{cases}$$

$$V^\gamma(0) = 0,$$

$$V^\gamma(n + 1) = \begin{cases} \text{max}\{0, V^\gamma(n) + x(n + 1)\} & \text{if } V^\gamma(n) < N, \\
\text{max}\{0, V^\gamma(n) + y^\gamma(n + 1)\} & \text{if } V^\gamma(n) \geq N. \\
\end{cases}$$

Put $\psi^R_{\gamma}(\gamma R)$ and $\psi_{\gamma}(\gamma R)$ to be the first hitting times of the set $[\gamma R, \infty)$ for the processes $\{V^R_{\gamma}(n); \ n \geq 0\}$ and $\{V^\gamma(n); \ n \geq 0\}$, respectively. Then $\psi^R_{\gamma}(\gamma R) = \psi_{\gamma}(\gamma R)$ for all $R$.

The processes $\{V^R_{\gamma}(n); \ n \geq 0\}$ and $\{W^R(n); \ n \geq 0\}$ are related in the following way: $V^R_{\gamma}(n) = x$, $W^R(n) = y$, where $x \leq y$, implies $W^R(n + 1) \geq_{st} V^R_{\gamma}(n + 1)$. Indeed, due to inequalities (28), (30) and (31),

$$W^R(n + 1) = y + A(n + 1) - T(n, y) - I^R(n, y) \geq_{st} x + A(n + 1) - T(n, x) - I^R(n, x).$$

Then we get $\varphi^R(\gamma R) \geq_{st} \psi^R_{\gamma}(\gamma R) = \psi_{\gamma}(\gamma R)$ as a consequence of the following result (see the Appendix for the proof).

**Proposition 6** Suppose $\{X(n); \ n \geq 0\}$ and $\{Y(n); \ n \geq 0\}$ are Markov chains with a common state space $S$, where $S$ is a closed subset of $\mathbb{R}$, and deterministic initial states $X(0) \leq Y(0)$. Suppose also that, for any $x \leq y$ and any $z$,

$$P\{X(n + 1) \geq z \mid X(n) = x\} \leq P\{Y(n + 1) \geq z \mid Y(n) = y\}.$$
Then there exist Markov chains \{\tilde{X}(n); \ n \geq 0\} and \{\tilde{Y}(n); \ n \geq 0\} defined on a common probability space, distributed as \{X(n); \ n \geq 0\} and \{Y(n); \ n \geq 0\}, respectively, and such that \(\tilde{X}(n) \leq \tilde{Y}(n)\) a.s. for all \(n\).

Now our goal is to choose a \(\gamma \in (0, \gamma^*)\) so as to have

\[
\liminf_{R \to \infty} \mathbb{P}\{\psi_\gamma(\gamma R) \leq \delta R\} \geq 1 - \varepsilon. \tag{33}
\]

To track the moments when the process \(\{V_\gamma(n); \ n \geq 0\}\) reaches level \(N\) from below and above, we recursively define the hitting times

\[
\tau^{(0)}_\gamma = 0, \quad \tau^{(i)}_\gamma = \inf\{n \geq \nu^{(i-1)}_\gamma: V_\gamma(n) \geq N\}, \quad i \geq 1,
\]

\[
\nu^{(0)}_\gamma = 0, \quad \nu^{(i)}_\gamma = \inf\{n \geq \tau^{(i)}_\gamma: V_\gamma(n) < N\}, \quad i \geq 1.
\]

By convention, infimum over the empty set is \(\infty\). So, if either \(\tau^{(i)}_\gamma = \infty\), or \(\nu^{(i)}_\gamma = \infty\), then \(\tau^{(j)}_\gamma = \nu^{(j)}_\gamma = \infty\) for all \(j > i\).

Note that the r.v. \(\tau^{(1)} := \tau^{(1)}_\gamma\) is a.s. finite and does not depend on \(\gamma\) because, for \(n \leq \tau^{(1)}\), the process \(\{V_\gamma(n); \ n \geq 0\}\) is a reflected homogeneous random walk with i.i.d. increments \(\{x(n); \ n \geq 1\}\), which do not depend on \(\gamma\). By Assumption 1, \(\mathbb{P}[x(1) > 0] > 0\), and then \(\mathbb{E}\tau^{(1)} < \infty\). Further, for any \(i\), if \(\nu^{(i-1)}_\gamma\) is finite, then \(\tau^{(i)}_\gamma\) is finite, too, and the difference \(\tau^{(i)}_\gamma := \tau^{(i)}_\gamma - \nu^{(i-1)}_\gamma\) is stochastically bounded from above by \(\tau^{(1)}\).

Let \(q^{(i)}_\gamma = \mathbb{P}\{\nu^{(i)}_\gamma < \infty | \nu^{(i-1)}_\gamma < \infty\}\). Then there exists a constant \(\tilde{q} < 1\) such that,

\[
\text{for all } i \text{ and } \gamma \text{ small enough, } \quad q^{(i)}_\gamma \leq \tilde{q}. \tag{34}
\]

Indeed, for all \(\gamma \in [0, \gamma^*]\), consider the random walks \(Y_\gamma(n) := \sum_{i=1}^n y_\gamma(i)\) (here \(y_0(n), \ n \geq 1, \) are defined by (32) with \(\gamma = 0\)). By Proposition 5, the family \(\{y_\gamma(1); \ 0 \leq \gamma \leq \gamma^*\}\) is uniformly integrable, and \(y_\gamma(1) \Rightarrow y_0(1)\) as \(\gamma \to 0\), which, together with \(\mathbb{E}y_0(1) \geq (\Lambda - e^{-1})/4 > 0\), implies that \(\inf_{n \geq 0} Y_\gamma(n) \Rightarrow \inf_{n \geq 0} Y_0(n)\) as \(\gamma \to 0\) (see [2, Chap. X, Theorem 6.1]). Also \(\mathbb{E}y_0(1) > 0\) implies \(\mathbb{P}[\inf_{n \geq 0} Y_0(n)] =: p_0 > 0\) (see [2, Chap. VIII, Theorem 2.4]). Then, for all \(i\), we have \(\mathbb{P}[\nu^{(i)}_\gamma = \infty | \nu^{(i-1)}_\gamma < \infty] \leq \mathbb{P}[\inf_{n \geq 0} Y_\gamma(n) \geq 0] \to p_0 > 0\), and, for all \(i\) and \(\gamma\) small enough, \(q^{(i)}_\gamma \leq 1 - p_0/2 =: \tilde{q} < 1\).

Let \(K_\gamma = \inf\{i \geq 1: v^{(i)}_\gamma = \infty\}\). By (34), the \(K_\gamma\)'s are stochastically bounded from above by a geometric r.v. uniformly in \(\gamma\) small enough,

\[
\mathbb{P}[K_\gamma > i] \leq \tilde{q}^i, \quad i \geq 0. \tag{35}
\]

Further, for \(i = 1, \ldots, K_\gamma\), define the hitting times

\[
\tilde{v}^{(i)}_\gamma = \inf\left\{n \geq 0: \sum_{j=1}^n y_\gamma(\tau^{(i)}_\gamma + j) \geq \gamma R\right\}.
\]
Since $E\gamma(1) > 0$, these r.v.'s are finite. We have $\psi\gamma(\gamma R) = \sum_{i=1}^{K\gamma}(\tilde{\tau}_\gamma^{(i)} + \tilde{\nu}_\gamma^{(i)})$. Indeed, if $\min\{i \geq 1: \nu^{(i)} \gamma - \tau^{(i)} \gamma \geq \tilde{\nu}^{(i)} \gamma\} = k$, then $k \leq K\gamma$ because $\nu^{(i)} \gamma = \infty$, and

$$\psi\gamma(\gamma R) \leq \tau^{(k)}_\gamma + \tilde{\nu}^{(k)}_\gamma = (\tau^{(1)}_\gamma - \nu^{(0)}_\gamma) + (\nu^{(1)}_\gamma - \tau^{(1)}_\gamma) + \cdots + (\tau^{(k)}_\gamma - \nu^{(k)}_\gamma) + \tilde{\nu}^{(k)}_\gamma$$

$$\leq \tilde{\tau}^{(1)}_\gamma + \tilde{\nu}^{(1)}_\gamma + \cdots + \tilde{\tau}^{(k)}_\gamma + \tilde{\nu}^{(k)}_\gamma \leq \tilde{\tau}^{(1)}_\gamma + \tilde{\nu}^{(1)}_\gamma + \cdots + \tilde{\tau}^{(K\gamma)}_\gamma + \tilde{\nu}^{(K\gamma)}_\gamma.$$

Now we are ready to complete the proof. By (35), there are $k_0$ and $\tilde{\gamma}$ such that $P\{K\gamma > k_0\} \leq \varepsilon$ for all $\gamma \leq \tilde{\gamma}$. Put $d_0 = \delta/(2k_0)$ and $\gamma = \min\{\gamma^*, \tilde{\gamma}, d_0(\Lambda - e^{-1})/8\}$. Then

$$P\{\psi\gamma(\gamma R) > \delta R\} \leq P\left\{\sum_{i=1}^{K\gamma}(\tilde{\tau}_\gamma^{(i)} + \tilde{\nu}_\gamma^{(i)})I\{K\gamma \geq i\} > \delta R\right\}$$

$$\leq P\left\{\sum_{i=1}^{k_0}(\tilde{\tau}_\gamma^{(i)} + \tilde{\nu}_\gamma^{(i)})I\{K\gamma \geq i\} > \delta R\right\} + \varepsilon$$

$$\leq \sum_{i=1}^{k_0}P\left\{\tilde{\tau}_\gamma^{(i)}I\{K\gamma \geq i\} > \delta_0 R\right\}$$

$$+ \sum_{i=1}^{k_0}P\left\{\tilde{\nu}_\gamma^{(i)}I\{K\gamma \geq i\} > \delta_0 R\right\} + \varepsilon.$$

(36)

Since $\{K\gamma \geq i\} \subseteq \{\nu^{(i-1)} < \infty\}$ and $\tau^{(1)}$ is a.s. finite,

$$t_i^R \leq P\left\{\tilde{\tau}_\gamma^{(i)}I\{\nu^{(i-1)} < \infty\} > \delta_0 R\right\}$$

$$= P\left\{\tilde{\tau}_\gamma^{(i)} > \delta_0 R \mid \nu^{(i-1)} < \infty\right\}P\{\nu^{(i-1)} < \infty\}$$

$$\leq P\{\tau^{(1)} > \delta_0 R\} \rightarrow 0 \text{ as } R \rightarrow \infty.$$

(37)

For $i = 1, \ldots, K\gamma$, we have $\{\sum_{j=1}^{\delta_0 R}\gamma(y_{\gamma}(j) + j) \geq \gamma R\} \subseteq \{\tilde{\nu}_\gamma^{(i)} \leq \delta_0 R\}$. Then

$$s_i^R = P\{\tilde{\nu}_\gamma^{(i)} > \delta_0 R \mid K\gamma \geq i\}P\{K\gamma \geq i\}$$

$$\leq P\left\{\sum_{j=1}^{\delta_0 R}\gamma(y_{\gamma}(j) + j) < \gamma R \mid K\gamma \geq i\right\}$$

$$= P\{Y_{\gamma}(\delta_0 R) < \gamma R\} \rightarrow 0 \text{ as } R \rightarrow \infty$$

(38)

because a.s. $Y_{\gamma}([\delta_0 R]/R) \rightarrow \delta_0 E\gamma(1) \geq \delta_0(\Lambda - e^{-1})/4 > \delta_0\gamma$.

Finally, (36), (37) and (38) imply (33).
(i) implies (ii), single class case  By (i), we can choose a \( \gamma > 0 \) such that, for large \( R \), the process \( W^R(\cdot) \) reaches level \( \gamma R \) in time \( \varphi^R(\gamma R) \leq \delta R \) with high probability. Now we prove that, within the time horizon \([\varphi^R(\gamma R), \Delta R]\), there exists a minorant for \( W^R(\cdot) \) that, for large \( R \), stays close to level \( \gamma R \) with high probability. Then \( W^R(\cdot) \) stays higher than, for example, level \( \gamma R/2 \).

We now proceed more formally. Fix \( \delta, \Delta \) and \( \varepsilon \). Take \( h^* \) and \( N \) the same as in (29). Take \( \gamma \) and r.v. \( \theta_i \) that satisfy (i) of Lemma 3, (ii) (with \( a = (\Delta - \varepsilon^{-1})/4 \)) of Proposition 5 and (iv) of Proposition 5. Let \( \{\theta(n); \ n \geq 1\} \) be an i.i.d. sequence with \( \theta(1) \equiv \theta_\gamma \), and assume that this sequence does not depend on \( \{A(n); \ n \geq 1\} \) and \( \{U(n); \ n \geq 0\} \). Let \( \{v^R(n); \ n \geq 1\} \) and \( \{y(n); \ n \geq 1\} \) be i.i.d. sequences with \( v^R(n) = A(n) - I\{U(n - 1) \leq h^*\} - \sum_{i=1}^{\lfloor \gamma R \rfloor} \xi^R(n - 1, i) \) and \( y(n) = A(n) - I\{U(n - 1) \leq h^*\} - \delta(n) \). Define the auxiliary processes

\[
V^R(n) = \begin{cases} 
W^R(n), & n < \varphi^R(\gamma R), \\
\lfloor \gamma R \rfloor, & n = \varphi^R(\gamma R), \\
\min\{\lfloor \gamma R \rfloor, V^R(n - 1) + v^R(n)\}, & n > \varphi^R(\gamma R),
\end{cases}
\]

and

\[
\tilde{V}^R(0) = 0, \quad \tilde{V}^R(n) = \min\{0, \tilde{V}^R(n - 1) + v^R(n)\}, \quad n \geq 1, \\
Y(0) = 0, \quad Y(n) = \min\{0, Y(n - 1) + y(n)\}, \quad n \geq 1.
\]

The processes \( W^R(\cdot) \) and \( V^R(\cdot) \) coincide within the time interval \([0, \varphi^R(\gamma R) - 1]\). Starting from time \( \varphi^R(\gamma R) \), as long as \( V^R(\cdot) \) stays above level \( N \), it stays a minorant for \( W^R(\cdot) \). Indeed, for \( R \) large enough, given \( N \leq V^R(i) \leq W^R(i), i = \varphi^R(\gamma R), \ldots, n \), if \( W^R(n) \geq \lfloor \gamma R \rfloor \), then, by (28),

\[
W^R(n + 1) \overset{a.s.}{=} \lfloor \gamma R \rfloor + A(n + 1) - T(n, \lfloor \gamma R \rfloor) - I^R(n, \lfloor \gamma R \rfloor)
\]

\[
= \lfloor \gamma R \rfloor + v^R(n + 1) \geq V^R(n + 1),
\]

and, if \( W^R(n) < \lfloor \gamma R \rfloor \), then

\[
W^R(n + 1) \geq W^R(n + 1) + A(n + 1) - I\{U(n) \leq h^*\} - \sum_{i=1}^{\lfloor \gamma R \rfloor} \xi^R(n, i)
\]

\[
\geq V^R(n) + v^R(n + 1) \geq V^R(n + 1).
\]

Further, by independence arguments, for \( R \) large enough, \( y(1) \leq \delta_1 \), \( v^R(1) \) and

\[
\min_{0 \leq i \leq n} Y(i) \leq \min_{0 \leq i \leq n} \tilde{V}^R(i). \quad \text{Since } E_y(1) > 0, \min_{0 \leq i \leq n} Y(i)/n \to 0 \text{ a.s. Hence}
\]

\[
\min_{0 \leq i \leq [\Delta R]} \tilde{V}^R(i)/R \to 0 \quad \text{as } R \to \infty.
\]

(39)

Now we are ready to complete the proof. Put \( C = \gamma/2 \) and define the events

\[
E^R = \left\{ \min_{[\delta R] \leq n \leq [\Delta R]} W^R(n) < RC \right\}.
\]
\[ A^R = \{ \varphi^R(\gamma R) \leq \delta R \}, \]
\[ B^R = \left\{ \min_{0 \leq i \leq |\Delta R|} V^R(\varphi^R(\gamma R) + i) \geq 3\gamma R/4 \right\}. \]

Then \( P\{E^R \} \leq P\{E^R \cap A^R \cap B^R \} + P\{AR\} + P\{BR\}, \)
- \( E^R \cap A^R \cap B^R \subseteq \{3\gamma R/4 \leq \min_{|\Delta R| \leq n \leq |\Delta R|} W^R(n) < \gamma R/2 \} = \emptyset, \)
- \( \lim_{R \to \infty} P\{AR\} \leq \varepsilon, \)
- by \( \{V^R(\varphi^R(\gamma R) + n); \ n \geq 0\} = \{\bar{V}^R(n) + |\gamma R|; \ n \geq 0\} \) and (39),
\[ P\{BR\} = P\left\{ \min_{0 \leq n \leq |\Delta R|} \bar{V}(n) < 3\gamma R/4 - |\gamma R| \right\} \to 0 \quad \text{as } R \to \infty. \]

Hence, (ii) of Lemma (3) holds.

**Single-class bound for a multiclass model**  Now we show that a model with multiple classes of customers can be coupled with a suitable single-class model in such a way that the workload process of the single-class model is majorised by the total workload of the multiclass model within the whole time horizon \([0, \infty)\). This, in particular, implies that statements (i) and (ii) of the lemma, proven in the single-class case, are valid in the multiclass case, too.

For the multiclass model, we slightly modify the representation of the workload process suggested in Sect. 7.1. We only change the terms that represent departures due to impatience. For \( x \in \mathbb{Z}^K_+ \), let
\[ I^R_i(n, x) = \sum_{j=x_1-T_i(n, x)+\cdots+x_i-T_i(n, x)} \mathbb{I}\left\{ U(n, j) \leq \frac{p_i R}{R} \right\}, \]
where the r.v.'s \( U(n, i), \ i \geq 1, \ n \geq 0, \) are mutually independent and distributed uniformly over the interval \([0, 1]\). We also assume that the \( U(n, i) \)'s do not depend on the random elements \( W_0^R, \ \{A(n); \ n \geq 1\} \) and \( \{U(n); \ n \geq 0\} \).

Consider a single-class model with
- initial condition \( \tilde{W}^R(0) = \|W^R(0)\|_1, \)
- arrival process \( \tilde{A}(n) = \|A(n)\|_1, \)
- reneging probability \( \tilde{p} = \max_{1 \leq i \leq K} p_i, \)
and define its dynamics as follows:
\[ \tilde{W}^R(n + 1) = \tilde{W}^R(n) + \tilde{A}(n + 1) - \tilde{\tau}(n, \tilde{W}^R(n)) - \tilde{I}^R(n, \tilde{W}^R(n)), \quad (40) \]
where, for \( k \in \mathbb{Z}_+, \)
\[ \tilde{\tau}(n, k) = \mathbb{I}\left\{ U(n) \leq h(k) \right\}, \]
\[ \tilde{I}^R(n, k) = \sum_{i=1}^{k-\tilde{\tau}(n, k)} \mathbb{I}\left\{ U(n, i) \leq \frac{\tilde{p}}{R} \right\}, \]
and the r.v.’s $U(n)$, $U(n,i)$, $n \geq 0$, $i \geq 1$, are those defining the multiclass model. Then, in particular,

$$
\| T(n,x) \|_1 \overset{a.s.}{=} \mathbf{1} \left\{ U(n) \leq \sum_{j=1}^{K} p_j(x) \right\} = \tilde{T}(n, \|x\|_1),
$$

(41)

We show by induction that $\tilde{W}^{R}()$ bounds $\| W^{R}(\cdot) \|_1$ from below. Let $N^{R}(n) = \| W^{R}(n) \|_1 - T(n, \| W^{R}(n) \|_1)$, and suppose that $\tilde{W}^{R}(n) \leq \| W^{R}(n) \|_1$ a.s., then

$$
\| W^{R}(n+1) \|_1 \overset{a.s.}{=} \| W^{R}(n) \|_1 + \tilde{A}(n+1) - \tilde{T}(n, \| W^{R}(n) \|_1)
$$

$$
- \sum_{i=1}^{N^{R}(n)} \mathbf{1}\{U(n,i) \leq \tilde{\rho}/R\}
$$

$$
= \| W^{R}(n) \|_1 + \tilde{A}(n+1) - \tilde{T}(n, \| W^{R}(n) \|_1)
$$

$$
- \tilde{I}^{R}(n, \| W^{R}(n) \|_1)
$$

$$
\overset{a.s.}{=} \tilde{W}^{R}(n) + \tilde{A}(n+1) - \tilde{T}(n, \tilde{W}^{R}(n)) - \tilde{I}^{R}(n, \tilde{W}^{R}(n))
$$

$$
= \tilde{W}^{R}(n+1),
$$

where the first and last inequalities hold by (41) and (28) respectively, and the identity is due to representation (40).

(ii) implies (iii) By Remark 5, it is enough to show that, for any $\Delta > 0$ and $i = 1, \ldots, K$,

$$
\| G_{i}^{2,R} \|_{[0,\Delta]} \Rightarrow 0 \text{ as } R \to \infty.
$$

(42)

Fix $\Delta$ and $i$. Recall that

$$
G_{i}^{2,R}(t) = e^{-1} \int_{0}^{t} m_{i}^{R}(s) ds - \left(1 - \frac{p_i}{R}\right) \int_{0}^{[Rt]/R} h^{R}(s)m_{i}^{R}(s) ds,
$$

(43)

where $m_{i}^{R}(s) = m_{i}(R\tilde{W}^{R}(s))$ and $h^{R}(s) = h(R\| \tilde{W}^{R}(s) \|_1)$. First, we estimate the subtractor in (43). Since $R\| \tilde{W}^{R}(\cdot) \|_1$ is integer-valued and non-negative, $h^{R}(\cdot) \leq 1$. Also $m_{i}^{R}(\cdot) \leq 1$. Then, for $t \leq \Delta$, we have

$$
\left| \int_{0}^{t} h^{R}(s)m_{i}^{R}(s) ds - \left(1 - \frac{p_i}{R}\right) \int_{0}^{[Rt]/R} h^{R}(s)m_{i}^{R}(s) ds \right|
$$

$$
= \int_{[Rt]/R}^{t} h^{R}(s)m_{i}^{R}(s) ds + \frac{p_i}{R} \int_{0}^{[Rt]/R} h^{R}(s)m_{i}^{R}(s) ds \leq \frac{1}{R} + \frac{p_i \Delta}{R}.
$$
Take $\delta < \Delta$; then
\[
\|G^2_i, R\|_{[0, \Delta]} \leq \frac{1 + p_i \Delta}{R} + \sup_{0 \leq t \leq \Delta} \left| \int_0^t m_i^R(s)(e^{-1} - h^R(s)) \, ds \right|
\leq \frac{1 + p_i \Delta}{R} + (e^{-1} + 1)\delta + \Delta \sup_{\delta \leq s \leq \Delta} |e^{-1} - h^R(s)|. \tag{44}
\]

Now we show that
\[
x^R(\delta, \Delta) := \|e^{-1} - h^R(\cdot)\|_{[\delta, \Delta]} \Rightarrow 0 \text{ as } R \to \infty. \tag{45}
\]

For any $\sigma > 0$ and $\varepsilon > 0$ and $C(\delta, \Delta, \varepsilon)$ satisfying (24),
\[
\{x^R(\delta, \Delta) \geq \sigma \} \subseteq \{x^R(\delta, \Delta) \geq \sigma, \inf_{\delta \leq s \leq \Delta} \|\overline{W}^R(s)\|_1 \geq C(\delta, \Delta, \varepsilon)\}
\cup \left\{ \inf_{s \geq \sup \{s \geq RC(\delta, \Delta, \varepsilon) \}} |e^{-1} - h(s)| \geq \sigma \right\}
\cup \left\{ \inf_{\delta \leq s \leq \Delta} \|\overline{W}^R(s)\|_1 < C(\delta, \Delta, \varepsilon) \right\}.
\]

Here the first event in the very RHS is empty for $R$ large enough, and hence
\[
\limsup_{R \to \infty} P\{x^R(\delta, \Delta) \geq \sigma \} \leq \limsup_{R \to \infty} P\left\{ \inf_{\delta \leq s \leq \Delta} \|\overline{W}^R(s)\|_1 < C(\delta, \Delta, \varepsilon) \right\} \leq \varepsilon.
\]

Since in the last inequality $\varepsilon > 0$ is arbitrary, for any $\sigma > 0$, as $R \to \infty$, we have $P\{x^R(\delta, \Delta) \geq \sigma \} \to 0$, which is (45).

Finally, (44) and (45) imply (42).

7.4 Proof of Lemma 4

We prove the result in the single-class case. The same proof is valid for each coordinate in the multiclass case.

Convergence of $G^3, R(\cdot)$ By Remark 5, it suffices to show that, for any $T > 0$,
\[
\mu^R(T) := \sup_{0 \leq t \leq T} \int_{[\lfloor Rt \rfloor/R, R]} \overline{W}^R(s) \, ds \Rightarrow 0 \text{ as } R \to \infty. \tag{46}
\]

Since $\overline{W}^R(\cdot)$ is a constant function within the interval $[[\lfloor Rt \rfloor/R, R]]$, we have
\[
\mu^R(T) = \sup_{t \leq T} \overline{W}^R(t)(Rt - \lfloor Rt \rfloor) / R \leq \sup_{0 \leq t \leq T} \overline{W}^R(t) / R
= W^R(0) / R^2 + \sum_{i=1}^{\lfloor TR \rfloor} A(i) / R^2,
\]
which implies (46).

\[\square\] Springer
Convergence of $\overline{M}^R(\cdot)$ We represent the martingale $\{M^R(n); n \geq 0\}$ as a sum of three other zero-mean martingales,

$$M^R(n) = S^R(n) - I^R(n) - T^R(n),$$

$$S^R(n) = \sum_{i=1}^{n} A(i) - n \Lambda,$$

$$I^R(n) = \sum_{i=0}^{n-1} \sum_{m=1}^{W^R(i)} \left( \xi^R(i, m) - \frac{P}{R} \right),$$

$$T^R(n) = \sum_{i=0}^{n-1} \left( T(i, W^R(i)) - h(W^R(i)) \right) + \sum_{i=0}^{n-1} \sum_{m=W^R(i)-T(i,W^R(i))+1}^{W^R(i)} \left( \xi^R(i, m) - \frac{P}{R} h(W^R(i)) \right).$$

It suffices to show that, for $N \in \mathbb{Z}_+, N \to \infty$,

$$\max_{1 \leq n \leq N} \frac{|X^N(n)|}{N} \Rightarrow 0, \quad X = S, T, I. \quad (47)$$

For $X = S$, (47) follows from the functional law of large numbers.

For all $n$ and $N$, we have $|T^N(n+1) - T^N(n)| \leq 4$. Then we get (47) with $X = T$ by applying Doob’s inequality for non-negative submartingales and the following known result (see for example [7, Chap. VII, Sect. 9, Theorem 3]).

**Proposition 7** Let $\{X(n); n \geq 1\}$ be a sequence of r.v.’s such that, for all $n$,

$$\mathbb{E}\{X(n) | X(1), \ldots, X(n-1)\} = 0.$$

Suppose that $b(1) < b(2) < \cdots \to \infty$ and that $\sum_{k=1}^{\infty} b(k)^{-2} \mathbb{E}X(k)^2 < \infty$. Then a.s. $b(n)^{-1} \sum_{k=1}^{n} X(k) \to 0$.

Now we prove (47) for $X = I$. The key tool of this proof is Markov’s inequality. We have to show that, for any $\varepsilon > 0$, as $N \to \infty$,

$$\mathbb{P}\left\{ \max_{1 \leq n \leq N} I^N(n)/N > \varepsilon \right\} \to 0, \quad (48a)$$

$$\mathbb{P}\left\{ \min_{1 \leq n \leq N} I^N(n)/N < -\varepsilon \right\} \to 0. \quad (48b)$$

By Taylor’s expansion, there exists an $\alpha^* > 0$ such that, for any $\alpha \in [0, \alpha^*]$, $\rho \in [0, 1]$ and r.v. $\xi(\rho)$ with $\mathbb{P}\{\xi(\rho) = 1\} = \rho = 1 - \mathbb{P}\{\xi(\rho) = 0\}$,

$$\mathbb{E}e^{\alpha(\xi(\rho) - \rho)} \leq e^{\alpha^2 \rho} \quad \text{and} \quad \mathbb{E}e^{\alpha(\rho - \xi(\rho))} \leq e^{\alpha^2 \rho}. \quad (49)$$
Since \( W_N(n) \leq W_N(0) + \sum_{i=1}^{N} A(i), \ n \leq N, \) and \( W_N(0)/N \Rightarrow z^0, \) and \( \sum_{i=1}^{N} A(i)/N \to \Lambda \) a.s., for any \( \delta > 0, \) there exists an \( M(\delta) \in \mathbb{Z}_+ \) such that
\[
\limsup_{N \to \infty} P\left\{ \max_{0 \leq n \leq N} W_N(n) > M(\delta)N \right\} \leq \delta. \tag{50}
\]
Denote the event in (50) by \( E_N^{\delta}(\delta) \) and put
\[
\alpha(\delta) = \min\{\alpha^*, \varepsilon/(2M(\delta))\}. \tag{51}
\]
We introduce the auxiliary martingale
\[
I_{N,\delta}(n) = \sum_{i=0}^{n-1} \sum_{m=1}^{W_{N,\delta}(i)} \left( N(i, m) - \frac{p}{N} \right), \quad n \geq 1,
\]
where \( W_{N,\delta}(i) = \max\{W_N(i), M(\delta)N\}. \) Note that, on \( E_{N}(\delta) \), we have \( I_{N,\delta}(n) = I^N(n), \ n = 0, \ldots, N. \) Hence,
\[
P\left\{ \max_{1 \leq n \leq N} I^N(n)/N > \varepsilon \right\} \leq P\left\{ \max_{1 \leq n \leq N} I_{N,\delta}(n) > \varepsilon N \right\} + P\left\{ E^N(\delta) \right\}
\]
\[
\leq \sum_{n=1}^{N} P\left\{ I_{N,\delta}(n) > \varepsilon N \right\} + P\left\{ E^N(\delta) \right\}. \tag{52}
\]
By Markov’s inequality, (51) and (49), for \( n \leq N, \)
\[
\exp(\alpha(\delta)\varepsilon N) P\left\{ I_{N,\delta}(n) > \varepsilon N \right\} \leq E \exp(\alpha(\delta)I_{N,\delta}(n))
\]
\[
= E\left[ E\left[ \prod_{i=0}^{n-1} \prod_{m=1}^{W_{N,\delta}(i)} \exp(\alpha(\delta)(N(i, m) - p/N)) \mid W_N(0), \ldots, W_N(n-1) \right]\right]
\]
\[
\leq E\left[ \exp\left( \alpha^2(\delta)\frac{p}{N}M(\delta)Nn \right) \right] \leq \exp(\alpha^2(\delta)M(\delta)N). \tag{53}
\]
By (50) and (51), bounds (52) and (53) imply that
\[
\limsup_{N \to \infty} P\left\{ \max_{1 \leq n \leq N} I^N(n)/N > \varepsilon \right\} \leq \delta,
\]
where \( \delta > 0 \) is arbitrary. Hence, convergence (48a) holds. Convergence (48b) can be treated similarly.

**Acknowledgements** The research of M. Frolkova and B. Zwart is supported by an NWO VIDI grant.

**Appendix**

**Proof of Proposition 5** There exists an \( R^* > 0 \) such that, for all \( n \in \mathbb{Z}_+, \ \gamma \in (0, 1] \) and \( R \geq R^*, \)
\[
e^n P\left\{ B^R_{\gamma} \geq n \right\} \leq E \exp(B^R_{\gamma}) = (ep/R + 1 - p/R)^{\lfloor\gamma R\rfloor} \leq \exp((e - 1)p) =: \mu.
\]
Take $N^*$ such that $\sum_{n>N^*} \mu ne^{-n} \leq a/2$, and $\gamma^* = \min\{1, a/(4p)\}$. Fix $\gamma \in (0, \gamma^*)$.

Since, as $R \to \infty$, $B([\gamma R], p/R) \Rightarrow \Pi(\gamma p)$, there exists an $R_\gamma \geq R^*$ such that

$$\mathbb{P}\{B^R_\gamma = n\} \leq 2\Pi(\gamma R)(\{n\}) \quad \text{for } R \geq R_\gamma \text{ and } n = 0, \ldots, N^*.$$

For $\gamma \in (0, \gamma^*)$, put

$$\mathbb{P}\{\theta_\gamma \geq n\} = \mu e^{-n}, \quad n > N^*, \quad \mathbb{P}\{\theta_\gamma = n\} = \min\{2\Pi(\gamma R)(\{n\}), 1 - \mathbb{P}\{\theta_\gamma \geq n + 1\}\} \quad n = N^*, \ldots, 0. \quad (54)$$

For $\gamma = 0$, put (54) and

$$\mathbb{P}\{\theta_0 \geq 0\} = 1 - \mu e^{-N^*-1}, \quad \mathbb{P}\{\theta_0 = n\} = 0, \quad n = 1, \ldots, N. \quad \square$$

**Proof of Proposition 6** Define

$$P(x, \geq z) = \mathbb{P}\{X(n+1) \geq z \mid X(n) = x\}, \quad Q(y, \geq z) = \mathbb{P}\{Y(n+1) \geq z \mid Y(n) = y\}.$$

Let $\{U(n); n \geq 0\}$ be an i.i.d. sequence with $U(0)$ distributed uniformly over $[0, 1]$. Then put

$$\tilde{X}(0) = X(0), \quad \tilde{X}(n+1) = \sup\{z \in \mathcal{S} : U(0) \geq 1 - P(\tilde{X}(n), \geq z)\},$$

$$\tilde{Y}(0) = Y(0), \quad \tilde{Y}(n+1) = \sup\{z \in \mathcal{S} : U(n) \geq 1 - Q(\tilde{Y}(n), \geq z)\}. \quad \square$$

**References**

1. Abramson, N.: The ALOHA system—another alternative for computer communications. AFIPS Conf. Proc. 36, 295–298 (1970)
2. Asmussen, S.: Applied Probability and Queues. Springer, New York (2003)
3. Bordenave, Ch., Foss, S., Shneer, S.: A random multiple access protocol with spatial interactions. J. Appl. Probab. 46, 844–865 (2009)
4. Borst, S., Egorova, R., Zwart, B.: Fluid limits for bandwidth-sharing networks in overload. Preprint (2009)
5. Ephremides, A., Hajek, B.: Information theory and communication networks: an unconsummated union. IEEE Trans. Inf. Theory 44, 2416–2434 (1998)
6. Ethier, N.S., Kurtz, T.G.: Markov Processes: Characterization and Convergence. Wiley, New York (1986)
7. Feller, W.: An Introduction to Probability Theory and Its Applications, vol. 2. Wiley, New York (1966)
8. Hajek, B.: Hitting-time and occupation-time bounds implied by drift analysis with applications. Adv. Appl. Probab. 14, 502–525 (1982)
9. Hartman, P.: Ordinary Differential Equations. Birkhäuser, Boston (1982)
10. Mikhailov, V.: Geometric analysis of stability of Markov chains in $\mathbb{R}^n$ and its applications to throughput evaluation of the adaptive random multiple access algorithm. Probl. Inf. Transm. 24, 47–56 (1988)