Asymptotic Results on Infinite Tandem Queueing Networks*

F. Baccelli[†], A. Borovkov[‡], and J. Mairesse[§]

March 19, 1999

Abstract

We consider an infinite tandem queueing network consisting of $\cdot/GI/1/\infty$ stations with i.i.d. service times. We investigate the asymptotic behavior of $t(n,k)$, the inter-arrival times between customers n and $n+1$ at station k, and that of $w(n,k)$, the waiting time of customer n at station k. We establish a duality property by which $w(n,k)$ and the "idle times" $y(n,k)$ play symmetrical roles. This duality structure, interesting by itself, is also instrumental in proving some of the ergodic results of the paper. We consider two versions of the model: the quadrant and the half-plane. In the quadrant version, the sequences of boundary conditions $\{w(0,k),\ k \in \mathbb{N}\}$ and $\{t(n,0),\ n \in \mathbb{N}\}$, are given. In the half-plane version, the sequence $\{t(n,0), n \in \mathbb{Z}\}$ is given. Under appropriate assumptions on the boundary conditions and on the services, we obtain ergodic results for both versions of the model. For the quadrant version, we prove the existence of temporally ergodic evolutions and of spatially ergodic ones. Furthermore, the process $\{t(n,k), n \in \mathbb{N}\}$ converges weakly with k to a limiting distribution, which is invariant for the queueing operator. In the more difficult half plane problem, the aim is to obtain evolutions which are both temporally and spatially ergodic. We prove that $1/n \sum_{k=1}^{n} w(0,k)$ converges almost surely and in L_1 to a finite constant. This constitutes a first step in trying to prove that $\{t(n,k), n \in \mathbb{Z}\}$ converges weakly with k to an invariant limiting distribution.

Keywords: Tandem queueing networks, stationarity, ergodic theory, last-passage percolation, Azuma's inequality.

*This work was partially supported by the European Community Framework IV programme through the research network ALAPEDES ("The ALgebraic Approach to Performance Evaluation of Discrete Event Systems").

[†]INRIA-ENS, Département de Mathématiques et d'Informatique, 45 rue d'Ulm, Paris 75005, France.

[‡]Institute of Mathematics, Novosibirsk, Russia. The work of this author was supported in part by the INTAS grant No.10820.

[§]LIAFA, CNRS-Université Paris VII, case 7014, 2 place Jussieu, 75251 Paris Cedex 05, France.

Contents

1 Introduction

The subject of investigation in this paper is an infinite tandem of FIFO single server queues with infinite buffers. The services are assumed to be i.i.d. (within a queue and between the different queues) with a general distribution. Using Kendall's notation, each queue is of type $./GI/1/\infty$. The queues are connected in such a way that the departure times from station $k \in \mathbb{N}$ are the arrival times at station $k+1$. We denote by $t(n,k)$ the inter-arrival time between customer n and $n+1$ at station k and by $s(n,k)$ and $v(n,k)$ the service time and the sojourn time of customer n at station k, respectively.
We establish a duality property by which $t(n,k)$ and $v(n,k)$ play symmetrical roles (there is a similar duality between the 'waiting times' $w(n,k)$ and the 'idle times' $y(n,k)$). This duality structure holds in a pathwise way, without any assumption on the timing variables of the model. Interesting by itself, the duality is also instrumental in proving some of the ergodic results of the paper.

We consider two versions of the model: the *quadrant* and the *half-plane*. In the quadrant version, the sequences of boundary conditions $\{w(0,k),\ k \in \mathbb{N}\}$ and $\{t(n,0),\ n \in \mathbb{N}\}$, are given. In the half-plane version, the sequence $\{t(n,0), n \in \mathbb{Z}\}$ is given. Under appropriate probabilistic assumptions, ergodic results are obtained for both versions. Working with processes defined on the rich state space $\mathbb{R}^{\mathbb{N}}$ (or $\mathbb{R}^{\mathbb{Z}}$), it is hardly possible to obtain results of convergence in total variation (no matter what the probabilistic assumptions are). Such results would require using renovating events (in the stochastic recursive sequence case, see [6, 7]) or Harris recurrence techniques (in the Markov chain case, see [31]). Therefore, we concentrate on weak convergence type of results.

In the quadrant case, we assume that $\{w(0,k),\ k \in \mathbb{N}\}$ is stationary and ergodic and verifies $E[w(0,0)] > 0$. In Theorem 4.1, we prove that the process $\{t(n,k), n \in \mathbb{N}\}$ converges weakly with k to a limiting distribution, which is invariant for the queueing operator.

In the half plane case, we assume that $\{t(n,0), n \in \mathbb{Z}\}$ is ergodic, integrable, independent of the services and verifies $E[t(0,0)] > E[s(0,0)]$ (stability condition). We conjecture that $\{t(n,k), n \in \mathbb{Z}\}$ converges weakly with k to a limiting distribution, which is invariant for the queueing operator. As a first step, we prove in Theorem 5.3 a Strong Law of Large Numbers for the total time spent by a customer in the system. More precisely we prove, under a mild moment condition on the services, that $1/n \sum_{k=1}^{n} v(0,k)$ converges almost surely to a finite constant. This constant depends on the initial inter-arrival process only through its mean $E[t(0,0)]$. Under an additional assumption involving the initial inter-arrival process, we prove that the convergence also holds in L_1.
Throughout the paper, we use the exponential model, i.e. the model with i.i.d. and exponentially distributed service times, to illustrate the results.

As far as we know, considering the quadrant problem is original. On the other hand, several papers, including recent contributions, have been dealing with the half-plane problem. Most of these papers are concerned with the exponential model. The main references are [2, 10, 33, 34]. The precise results will be recalled in the course of the paper.

The type of models and problems studied in this paper have often been investigated starting from a non-queueing motivation. This is particularly true for the tandem model with exponential services which is connected with several other models that we now detail. The *totally asymmetric exclusion process* and the *totally asymmetric zero range process* are classical interacting particle systems, whose asymptotic behaviors have been widely studied, see for instance [39, 3, 26, 13]. The correspondence between the two processes and exponential tandem networks has been often noticed and used, see for instance [25, 41, 16]. A complete discussion appears in [42]. *Young tableaux* are central in group representation and combinatorics, see [17]. The correspondence between randomly growing Young tableaux and tandem networks is explained for instance in [40]. The asymptotics for several variants of randomly growing Young tableaux have been considered, see [46] and the references therein. The following three problems are closely related, see [1]: (i) longest increasing subsequences in a random permutation; (ii) longest oriented paths in a Poisson random field; and (iii) asymptotics of Hammersley's interacting particle system. These problems have some common features with tandem queues (although the connection is not as precise as the previous ones). A first attempt at applying results developped for these problems in the context of tandem of queues is performed in [20], section 6.

There is another important connection, specifically used in this paper, between tandem networks and last-passage percolation problems. Precisely, sojourn times in a tandem network can be represented as longest paths in a directed graph. As opposed to the previous correspondences, this one is pathwise and valid for any service distribution. It has been used by several authors, see [19, 44]. We detail the correspondence in §5.2.1 and we use it both directly and by borrowing results developped for lattice animals in [11, 18].

The paper is organized as follows: section 2 gives the basic equations capturing the time-space evolution of events in such a network. The probabilistic setting is defined in section 3. Section 4 focuses on the quadrant version of the problem. The half plane question is addressed in Section 5. At last, in section 6, we discuss the relations between the results obtained for the quadrant and the half-plane problem.

We denote by $\mathbb{N}^*$ and $\mathbb{R}^*$ the sets $\mathbb{N}\backslash\{0\}$ and $\mathbb{R}\backslash\{0\}$ respectively. We denote the (lower) integer part of a real a by $[a]$ and we use the notation $a^+ = \max(a, 0)$. The indicator function of an event A is denoted by $\mathbb{1}\{A\}$.

2 The Model

2.1 Duality

In this section, no probabilistic assumption is made. We focus on pathwise properties of the main processes describing the tandem system. The basic idea is that there is a symmetry in the role played by stations and customers. The evolution of the network can be visualized as customers moving from station to station or equivalently as stations moving from customer to customer. The waiting time of a customer and the idle time of a station are dual quantities. The inter-arrival times and the sojourn times are also dual variables. We provide a precise mathematical setting for this intuition. The general idea of duality between stations and customers is present in the work of several authors, see for instance Glynn and Whitt [19] and the references therein. However, the main equations below, (6), (12) and (13), seem to be new.

We consider a model constituted by an infinite sequence of customers going through an infinite number of FIFO queues. Each customer has a number, $n \in \mathbb{Z}$, which it carries when moving from station to station. The different quantities (resp. formulas) below are a priori defined (resp. verified) for all $n \in \mathbb{Z}$, and all $k \in \mathbb{Z}$, where k is the station index. Let

- $A(n, k) \in \mathbb{R}$ be the instant of arrival of customer n at station k, with $A(n, k) \leq A(n + 1, k), \forall n, \forall k$. By convention, $A(0, 0) = 0$;
- $D(n, k) \in \mathbb{R}$ be the instant of departure of customer n from station k, with $A(n, k) \leq D(n, k), \forall n, \forall k$;
- $s(n, k) \in \mathbb{R}_+$ be the service time of the n-th customer at station k.

The stations are single server queues with an infinite buffer and a FIFO (First-In-First-Out) policy. Using Kendall's notation, each station is of type $./ ./1/\infty$ FIFO. The equation relating arrival and service times on the one hand and departure times on the other hand

is: $D(n,k) = \max(A(n,k), D(n-1,k)) + s(n,k)$. The queues are in tandem, meaning that a customer leaving station $k-1$ immediately enters station k: $A(n,k) = D(n,k-1)$.
We introduce some further notations. Let

- $t(n,k) \in \mathbb{R}_+$ be the n-th inter-arrival time at station k, i.e. $t(n,k) = A(n+1,k) - A(n,k)$;
- $v(n,k) \in \mathbb{R}_+$ be the sojourn time of customer n in station k. We have $v(n,k) = A(n,k+1) - A(n,k)$.

The process $\{t(n,k), n \in \mathbb{Z}\}$ is called the *inter-arrival process* (at station k); $\{A(n,k), n \in \mathbb{Z}\}$ being the arrival process (at station k). It would be consistent to call $\{v(n,k), k \in \mathbb{Z}\}$ the inter-arrival process (of customer n) since $\{A(n,k), k \in \mathbb{Z}\}$ is the arrival process (of customer n). To avoid confusions, we prefer the classical denomination of *sojourn time process* for $\{v(n,k), k \in \mathbb{Z}\}$.
Let us now introduce some variables which are less classical. They will become natural in view of the duality structure to be exhibited.

- $\tilde{A}(n,k)$ is the "instant of arrival" of station k at customer n, i.e. the instant at which station k becomes ready to serve customer n (whether customer n has arrived or not). We have $\tilde{A}(n,k) = D(n-1,k)$;
- $\tilde{D}(n,k)$ is the "instant of departure" of station k from customer n, i.e. the instant of departure of customer n from station k, i.e. $\tilde{D}(n,k) = D(n,k)$;
- $\tilde{t}(n,k)$ is the inter-arrival time between stations k and $k+1$ at customer n. We have $\tilde{t}(n,k) = \tilde{A}(n,k+1) - \tilde{A}(n,k)$.

The servers are operating without any vacations: $\tilde{D}(n,k) = \tilde{A}(n+1,k)$. Next equality summarizes the relations between the different quantities introduced.

$$A(n,k+1) = D(n,k) = \tilde{D}(n,k) = \tilde{A}(n+1,k)\,. \tag{1}$$

We now define the *waiting time* (or *workload*) and *idle time* variables. Let

- $w(n,k) \in \mathbb{R}_+$ be the waiting time of the n-th customer, namely the time that elapses between its arrival in station k and the beginning of its service there. Due to the FIFO assumption, it coincides with the workload at station k upon arrival of customer n. We have $w(n,k) = D(n,k) - A(n,k) - s(n,k)$;
- $y(n,k) \in \mathbb{R}_+$ be the idle time of station k before the arrival of customer n. We have $y(n,k) = \tilde{D}(n,k) - \tilde{A}(n,k) - s(n,k)$.

We have the following relations

$$\begin{cases} v(n,k) & = w(n,k) + s(n,k) \;=\; \tilde{t}(n+1,k-1) \\ t(n-1,k+1) & = y(n,k) + s(n,k)\,. \end{cases} \tag{2}$$

We have two conflicting notations, $\tilde{t}(n+1, k-1)$ and $v(n, k)$, to denote the same quantity. Despite the redundancy, it is worth carrying both of them. On one hand, $v(n, k)$ corresponds to the intuitive and classical numbering, on the other hand $\tilde{t}(n+1, k-1)$ enables us to have nice and symmetrical formulas in (3) and (4) below.

The two equations in (2) are illustrated in Fig. 1. The horizontal axis corresponds to time and the vertical one to the total workload in the station.

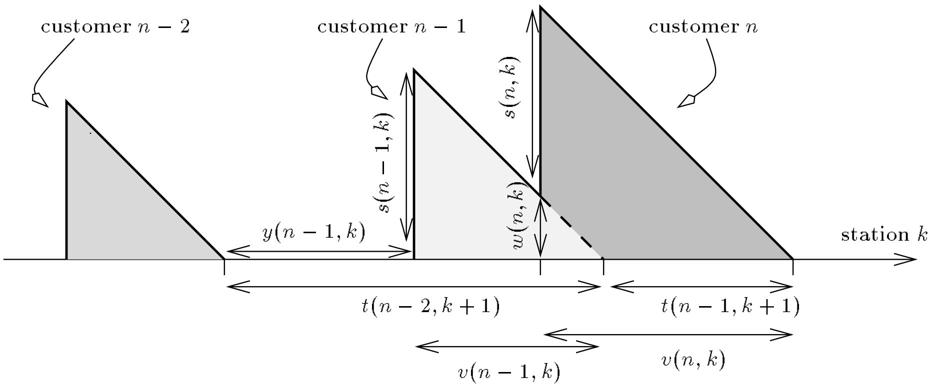


Figure 1: Customers $n-2$, $n-1$ and n at station k.

It follows directly from the definition that there is an equivalent way of expressing the variables $w(n, k)$ and $y(n, k)$.

$$\begin{cases} w(n,k) & = [\tilde{A}(n,k) - A(n,k)]^+ = [\tilde{t}(n,k-1) - t(n-1,k)]^+ \\ y(n,k) & = [A(n,k) - \tilde{A}(n,k)]^+ = [t(n-1,k) - \tilde{t}(n,k-1)]^+ \, . \end{cases} \tag{3}$$

From this, we easily obtain two recurrence equations, where the first one is very classical and known as *Lindley's equation*:

$$\begin{cases} w(n,k) & = [w(n-1,k) + s(n-1,k) - t(n-1,k)]^+ \\ y(n,k) & = [y(n,k-1) + s(n,k-1) - \tilde{t}(n,k-1)]^+ \, . \end{cases} \tag{4}$$

Let us define

$$\xi(n,k) = s(n-1,k) - s(n,k-1) \, . \tag{5}$$

When eliminating $t(n-1, k)$ and $\tilde{t}(n, k-1)$ using eqn (2), we obtain two new and symmetrical equations

$$\begin{cases} w(n,k) & = [w(n-1,k) + \xi(n,k) - y(n,k-1)]^+ \\ y(n,k) & = [y(n,k-1) - \xi(n,k) - w(n-1,k)]^+ \, . \end{cases} \tag{6}$$

Equation (6) expresses the *duality property* of the variables $w(n,k)$ and $y(n,k)$. It follows from (6) that

$$\begin{cases} w(n,k)+y(n,k) & = |w(n-1,k)+\xi(n,k)-y(n,k-1)| \\ w(n,k)-y(n,k) & = w(n-1,k)+\xi(n,k)-y(n,k-1) \ . \end{cases} \tag{7}$$

If $\xi(n,k)$ is a random variable and if its distribution is continuous, then a simple consequence is

$$y(n,k)=0 \Leftrightarrow w(n,k)>0, \quad w(n,k)=0 \Leftrightarrow y(n,k)>0 \text{ a.s..} \tag{8}$$

In words, the probability of an arrival coinciding with a departure is null.
The variables w and v on one hand, y and t on the other hand, are linked by eqn (2). It enables us to obtain couples of equations analogous to (6) for each of the following combinations of variables: (w,t), (v,y) or (v,t).
Let us write the equations for v and t.

$$\begin{cases} v(n,k) & = s(n,k)+[v(n-1,k)-t(n-1,k)]^+ \\ t(n-1,k+1) & = s(n,k)+[t(n-1,k)-v(n-1,k)]^+ \ . \end{cases} \tag{9}$$

Equation (9) shows the duality between the variables $v(n,k)$ and $t(n-1,k+1)$.
The analog of the second equation in (7) for the variables v and t is an interesting equation in which the services have disappeared:

$$v(n,k)-t(n-1,k+1)=v(n-1,k)-t(n-1,k) \ . \tag{10}$$

Now, let us develop the first recurrence in eqn (4). We obtain the classical *developed Lindley's equation.* For all $N,K\in\mathbb{N}^*$,

$$\begin{aligned} w(n,k) \ &= \max \ \Bigg[\max_{i=1}^{N-1} \sum_{j=1}^{i} s(n-j,k) - \sum_{j=1}^{i} t(n-j,k), \\ &\qquad w(n-N,k)+\sum_{j=1}^{N} s(n-j,k)-\sum_{j=1}^{N} t(n-j,k) \Bigg]^+ \ . \end{aligned} \tag{11}$$

An analogous equation could be obtained for $y(n,k)$ by developing the second recurrence in eqn (4). Now, developing the recurrences in (6), we obtain two dual equations.

$$w(n,k) = \max\left[\max_{i=1}^{N-1}\sum_{j=1}^{i}\xi(n-j+1,k) - \sum_{j=1}^{i}y(n-j+1,k-1),\right. \tag{12}$$

$$\left. w(n-N,k) + \sum_{j=1}^{N}\xi(n-j+1,k) - \sum_{j=1}^{N}y(n-j+1,k-1)\right]^+$$

$$y(n,k) = \max\left[\max_{i=1}^{K-1}\sum_{j=1}^{i}-\xi(n,k-j+1) - \sum_{j=1}^{i}w(n-1,k-j+1),\right. \tag{13}$$

$$\left. y(n,k-K) + \sum_{j=1}^{K}-\xi(n,k-j+1) - \sum_{j=1}^{K}w(n-1,k-j+1)\right]^+ .$$

In some cases, equations (11),(12) or (13) can be written for N (resp. K) equal to infinity, see §3.2.

2.2 Boundary conditions

Let us assume that we consider a field of variables $\{Z(n,k) = (w(n,k), y(n,k))\}$ verifying the recurrence of eqn (6) for $n, k \in \mathbb{Z}$. Let us consider a broken line defined by a two-sided sequence of points of $\mathbb{Z}\times\mathbb{Z}$ such that two consecutive points are of the type

$$(n,k) \to (n+1,k),\ (n,k) \to (n,k-1) \text{ or } (n,k) \to (n+1,k-1)\ ,$$

and such that no three consecutive points are of the type

$$(n,k) \to (n+1,k) \to (n+1,k-1) \text{ or } (n,k) \to (n,k-1) \to (n+1,k-1)\ .$$

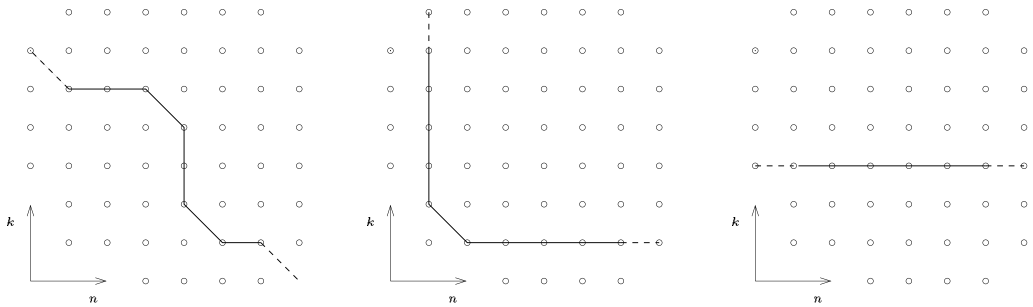


Figure 2: Generic broken line; quadrant case; half-plane case.

Given $Z(n,k)$ on any such line, the variables $Z(n_1,k_1)$ for points (n_1,k_1) lying above and to the right of the line are independent of the variables $Z(n_2,k_2)$ for points (n_2,k_2) lying below and to the left of the line. Hence any such broken line can be used to define the

initial conditions for our model. Three examples have been represented in Fig. 2. As detailed below, the quadrant case corresponds to the line defined by the points $\{(0,k), k \in \mathbb{N}\} \cup \{(n,-1), n \in \mathbb{N}^*\}$, and the half-plane case to an horizontal straight line.
The above considerations also show that one can hardly construct finite-dimensional sequences describing the evolution of the system. Apparently, it is only possible to do so in infinite-dimensional spaces, for instance in the space $\mathbb{R}^{\mathbb{N}}$.

We study the problem under two different specifications for the boundary conditions.

The quadrant. The variables $Z(n,k)$ are defined for $n, k \in \mathbb{N}$. We consider the boundary values:

$$w(0,k), k \in \mathbb{N}, \quad y(n,-1), n \in \mathbb{N}^* .$$

The initial values $\{y(n,-1), n \in \mathbb{N}^*\}$ can be obtained from an initial inter-arrival process $\{t(n,0), n \in \mathbb{N}\}$ using eqn (2). They are interpreted as the idle times in a fictive queue numbered -1.
All the variables $\{Z(n,k), n \in \mathbb{N}^*, k \in \mathbb{N}\}$ can be computed from the boundary values using the recurrence in (6).

The half-plane. The variables $Z(n,k)$ are defined for $n \in \mathbb{Z}, k \in \mathbb{N}$. Let us consider the boundary values:

$$y(n,-1), n \in \mathbb{Z} .$$

These values can be obtained from an initial inter-arrival process $\{t(n,0), n \in \mathbb{Z}\}$. For each $k \geq 0$, knowing the process $\{y(n,k-1), n \in \mathbb{Z}\}$, we can compute the process $\{w(n,k), n \in \mathbb{Z}\}$ according to (eqn (12) with N equal to infinity).

$$w(n,k) = \sup_{i \geq 0} \left[\sum_{j=0}^{i} \xi(n-j,k) - y(n-j,k-1) \right]^+ . \tag{14}$$

Formally, eqn (14) corresponds to eqn (12) with the conventions: "$N = +\infty$" and "$w(-\infty,k) = 0$". One can say that the boundary values defining the problem are $w(-\infty,k) = 0, k \in \mathbb{N}, \quad y(n,-1), n \in \mathbb{Z}$. This way we get a closer parallel with the situation in the quadrant version.
Assume that the variables $w(n,k), n \in \mathbb{Z}$, are finite. Now, knowing the processes $\{y(n,k-1), n \in \mathbb{Z}\}$ and $\{w(n,k), n \in \mathbb{Z}\}$, we can compute the process $\{y(n,k), n \in \mathbb{Z}\}$ using eqn (6). We conclude that all the variables $Z(n,k), n \in \mathbb{Z}, k \in \mathbb{N}$, are recursively computable. For conditions under which the variables $w(n,k)$ are finite, see §3.2.

We consider successively the quadrant problem in §4, and the half plane problem in §5.

3 Probabilistic Preliminaries

In order to get asymptotic results for the models defined above, we need to make some probabilistic assumptions. In this section, we recall the definitions needed and also Loynes' construction for an ergodic queue.

3.1 Definitions

Stationary and ergodic sequences Let us consider a probability space $(\Omega, \mathcal{F}, P)$. Let $(K, \mathcal{K})$ be a measurable space. We denote by $\mathcal{L}(K)$ the set of K-valued random variables; and by $\mathcal{M}(K)$ the set of probability measures on K.
Let $\theta : \Omega \to \Omega$ be a measurable transformation. We say that θ is *stationary* if $P\{\theta^{-1}(\mathcal{A})\} = P\{\mathcal{A}\}, \forall \mathcal{A} \in \mathcal{F}$. A stationary transformation θ is *ergodic* if $\mathcal{A} \in \mathcal{F}$ and $\mathcal{A} = \theta^{-1}(\mathcal{A})$ implies $P\{\mathcal{A}\} = 0$ or 1. We also say that the quadruple $(\Omega, \mathcal{F}, P, \theta)$ is stationary (resp. ergodic).
Let $(K, \mathcal{K})$ be a measurable space, let $I = \mathbb{N}$ or $\mathbb{Z}$ and let $(K^I, \mathcal{K}^I)$ be the corresponding product space (with the product σ-algebra). Given a r.v. $X \in \mathcal{L}(K)$, we define the r.v. $X \circ \theta$ by $X \circ \theta(\omega) = X(\theta\omega), \omega \in \Omega$. A random sequence $\{X(n), n \in I\} \in \mathcal{L}(K^I)$, $I = \mathbb{N}$ or $\mathbb{Z}$, is said to be *compatible* with θ if $X(n) = X(0) \circ \theta^n$. In the case $I = \mathbb{Z}$, we have to assume that θ is a bijection. When θ is stationary (resp. ergodic), a compatible sequence $\{X(n)\}$ is said to be *stationary* (resp. *ergodic*).

It is also convenient to define the stationarity and ergodicity of a random sequence directly on the realization space. Let T be the *shift transformation* defined by $T : \quad K^I \to K^I, \quad [Tx](n) = x(n+1)$. Let $X \in \mathcal{L}(K^I)$ be a random sequence and let $P_X \in \mathcal{M}(K^I)$ be its distribution. We say that X is stationary (resp. ergodic) if $(K^I, \mathcal{K}^I, P_X, T)$ is stationary (resp. ergodic).

It is clear that the first definition implies the second one. Barring some technicalities, the converse is also true. It means that given a stationary (ergodic) process for the second definition, there exists a stationary (ergodic) transformation θ such that the process is compatible with θ. For details, see Doob [14], Chapter X.

In this paper, it will be convenient to work with the two definitions of stationary-ergodic random processes. To avoid confusions, we take the following convention. We consider a single probability quadruple $(\Omega, \mathcal{F}, P, \theta)$ where θ is assumed to be ergodic. When we say that a process is stationary or ergodic, without further precision, it will be according to the second definition. Otherwise, we say (with some redundancy) that the process is ergodic *and* compatible with θ.

For doubly indexed random processes, we also define the notions of temporal and spatial stationarity and ergodicity. Let I and J be equal to $\mathbb{N}$ or $\mathbb{Z}$ and let us consider the product space $(K^I \times K^J, \mathcal{K}^I \times \mathcal{K}^J)$. Let $\{X(n,k), (n,k) \in I \times J\} \in \mathcal{L}(K^I \times K^J)$ be a doubly indexed random sequence and let P_X be its distribution. We say that $\{X(n,k), (n,k) \in I \times J\}$ is

- *temporally stationary* (resp. *ergodic*) if (1) $X(n,k) \circ \theta = X(n+1,k)$, for all k, where θ is a stationary (resp. ergodic) transformation; or if (2) P_X is stationary (resp. ergodic) with respect to the shift $T_1 : x(n,k) \to x(n+1,k)$;

- *spatially stationary* (resp. *ergodic*) if (1) $X(n,k) \circ \theta = X(n,k+1)$, for all n, where θ is a stationary (resp. ergodic) transformation; or if (2) P_X is stationary (resp. ergodic) with respect to the shift $T_2 : x(n,k) \to x(n,k+1)$.

This terminology follows from our use of random sequences which are doubly indexed by customers and stations.

Modes of convergence Let $P_n, n \in \mathbb{N}$, and P be probability measures in $\mathcal{M}(S)$, where $(S, \mathcal{S})$ is a metric space equipped with its Borel σ-algebra. Let $X_n, Y_n, n \in \mathbb{N}$, and X be random variables in $\mathcal{L}(S)$.
There is *weak convergence* of P_n to P, denoted by $P_n \xrightarrow{w} P$, if

$$\forall f \in C_b(\mathcal{S}), \int f dP_n \stackrel{n\to+\infty}{\longrightarrow} \int f dP \, ,$$

where C_b is the set of real-valued continuous bounded functions. By extension, we say that there is weak convergence of X_n to X, and we write $X_n \xrightarrow{w} X$, if there is weak convergence of the corresponding distributions.
There is *coupling in finite time* (or, merely, coupling) of $\{X_n\}$ with $\{Y_n\}$ if

$$P\{X_{n+l} = Y_{n+l}, \ \forall l \geq 0\} \stackrel{n\to+\infty}{\longrightarrow} 1 \, .$$

Coupling of X_n with a *stationary* sequence Y_n implies weak convergence (in fact, even total variation convergence) of the distribution of X_n to the one of Y_0.

The spaces $\mathbb{R}^{\mathbb{N}}$ and $\mathbb{R}^{\mathbb{Z}}$ Again, let the index set I denote either $\mathbb{N}$ or $\mathbb{Z}$. For any finite subset S of I, we denote by π_S the canonical projection of $\mathbb{R}^I$ into $\mathbb{R}^S$ defined by $\pi_S(u) = (u(i), i \in S)$.
We consider the product topology, or topology of coordinatewise convergence, on $\mathbb{R}^I$. If $u_n, n \in \mathbb{N}$ and u belong to $\mathbb{R}^I$, we have $\lim_n u_n = u$ if $\lim_n u_n(i) = u(i), \forall i \in I$. This topology is metrizable using the metric

$$d(u,v) = \sum_{i\in I} \frac{1}{2^i} \frac{|u(i) - v(i)|}{1 + |u(i) - v(i)|} \, . \tag{15}$$

The metric space $(\mathbb{R}^I, d)$ is Polish, i.e. complete and separable. We consider the measurable space $(\mathbb{R}^I, \mathcal{R}^I)$ where $\mathcal{R}^I$ is the Borel σ-algebra generated by the product topology. On $\mathcal{M}(\mathbb{R}^I)$, the topology of weak convergence is metrizable (using for instance the Prohorov metric), and $\mathcal{M}(\mathbb{R}^I)$ is a Polish space. Weak convergence can be characterized as follows. Let us consider $U_n, n \in \mathbb{N}$, and U in $\mathcal{M}(\mathbb{R}^I)$. We have $U_n \xrightarrow{w} U$ if and only if $U_n \pi_S^{-1} \xrightarrow{w} U \pi_S^{-1}$ for all finite subset S of I (on $\mathbb{R}^k, k \in \mathbb{N}^*$, the topology is the usual one). For more details, see Billingsley [5] or Dudley [15].

3.2 Loynes' construction for a single queue

Let us recall some classical results of Loynes [28], a complete presentation may be found in [4]. We consider a single server queue of type $G/G/1/\infty$ FIFO. The inter-arrival process $t = \{t(n), n \in \mathbb{Z}\}$ and the service process $s = \{s(n), n \in \mathbb{Z}\}$ are ergodic, compatible with θ and $t(0)$ and $s(0)$ are integrable (note the shortened notations for inter-arrivals and services). Two types of arrival streams are considered.

First, consider the restricted inter-arrival process $\bar{t} = \{t(n), n \in \mathbb{N}\}$. We denote the departure, inter-departure, workload and idle time variables associated with this process by $\overline{D}(n), \overline{t}(n,1), \overline{w}(n)$ and $\overline{y}(n)$, $n \in \mathbb{N}$. We assume the initial conditions $\overline{w}(0)$ and $\overline{y}(0)$ to be a.s. finite and to verify (8). The above variables are defined recursively using the equations in §2.1.

Second, consider the two-sided inter-arrival process $t = \{t(n), n \in \mathbb{Z}\}$. In this case, possible candidates for the waiting, idle and departure times are (as can be seen by going to the limit in (11)):

$$
\begin{aligned}
w(n) &= [\sup_{i=1}^{\infty}(\sum_{j=1}^{i} s(n-j) - t(n-j))\,]^{+}\,; \qquad (16)\\
y(n) &= [\inf_{i=1}^{\infty}(\sum_{j=1}^{i} t(n-j) - s(n-j))\,]^{+};\quad D(n) = A(n) + w(n) + s(n)\,.
\end{aligned}
$$

It is the limit when k goes to infinity of the systems starting empty and with a restricted inter-arrival process $\{t(n), n \geq -k\}$. The above quantities can be shown to exist (in $\mathbb{R}_{+} \cup \{+\infty\}$). There are two possible cases.

- $E[s(0)] > E[t(0)]$: the instability case.

For the two-sided arrival process, we have $P\{w(n) = \infty\} = P\{D(n) = \infty\} = 1$ and $P\{y(n) = 0\} = 1$. The action of the station on the two-sided arrival process is not well defined. For the restricted arrival process, we have $\lim_n \overline{w}(n)/n = E[s(0)] - E[t(0)]$ P-a.s.; $\lim_n \overline{y}(n) = 0$ P-a.s. and, as l goes to infinity, $\{\overline{t}(l+n,1), n \in \mathbb{N}\} \xrightarrow{w} \{s(n,0), n \in \mathbb{N}\}$.
If we assume furthermore that the services are i.i.d., that t and s are independent and that $P\{s(0) \neq E[s(0)]\} > 0$, then the above results also hold in the limit case $E[s(0)] = E[t(0)]$.

- $E[s(0)] < E[t(0)]$: the stability case.

For the two-sided arrival process, we have $P\{w(n) < \infty\} = P\{D(n) < \infty\} = 1$ and $P\{y(n) > 0\} > 0$. The processes $\{w(n), n \in \mathbb{Z}\}, \{y(n), n \in \mathbb{Z}\}$ and $\{t(n,1), n \in \mathbb{Z}\}$ are ergodic and θ-compatible, i.e $w(n) = w(0) \circ \theta^{n}$, $y(n) = y(0) \circ \theta^{n}$ and $t(n,1) = t(0,1) \circ \theta^{n}$. Furthermore, we have $E[t(0,1)] = E[t(0)]$. For the restricted arrival process, we have that $\overline{w}(n), \overline{y}(n)$ and $\overline{t}(n,1), n \in \mathbb{N}$, couple in finite time with $w(0) \circ \theta^{n}, y(0) \circ \theta^{n}$ and $t(0,1) \circ \theta^{n}$ respectively.

In the stability case, we need a strengthening of the above results. Let us consider a restricted arrival process $\hat{t} = \{\hat{t}(n), n \in \mathbb{N}\}$ which couples in finite time with the ergodic process $t = \{t(n), n \in \mathbb{Z}\}$. Then we have that $\hat{w}(n), \hat{y}(n)$ and $\hat{t}(n,1)$ couple in finite time with the ergodic and θ-compatible processes $w(n), y(n)$ and $t(n,1)$. This is proved for instance in [4]. Other results in the same vein, for several modes of convergence of $\hat{t}$ to t, can be found in Szczotka [43] or Szczotka & Kelly [44].

Switching back to the notations of §2.1, let us consider a finite number of queues numbered from 0 to K. If we assume that the processes $\{t(n,0), n \in \mathbb{Z}\}$, $\{s(n,k), n \in \mathbb{Z}\}, k \in \{0, \ldots, K\}$ are ergodic and θ-compatible and that $E[t(0,0)] > E[s(0,k)], \forall k$, then we can apply Loynes' results recursively at each queue. More precisely, for all $k \in \{0, \dots, K\}$, we obtain that the processes $\{w(n,k), n \in \mathbb{Z}\}, \{y(n,k), n \in \mathbb{Z}\}$ and $\{t(n,k), n \in \mathbb{Z}\}$ are ergodic and θ-compatible and we have $E[t(0,k)] = E[t(0,0)]$.

Example 3.1 (Exponential case). Let us consider a $M/M/1/\infty$ queue of parameters (α, β), i.e. a single server queue with exponential services of mean β and a Poisson arrival process of rate $1/\alpha$. More precisely, we assume that $\{s(n), n \in \mathbb{Z}\}$ and $\{t(n), n \in \mathbb{Z}\}$ are i.i.d., mutually independent, and with marginal laws given by, $\forall u \in \mathbb{R}_+$,

$$P\{s(0) > u\} = \exp(-u/\beta), \quad P\{t(0) > u\} = \exp(-u/\alpha) .$$

We assume that $\rho = \beta/\alpha < 1$, which implies that we are in the stability case. Classically, the stationary processes have the following characteristics.

The departure process is a Poisson process of the same rate as the arrival process. In other words, the process $\{t(n,1), n \in \mathbb{Z}\}$ is i.i.d. with a marginal law given by

$$P\{t(0,1) > u\} = \exp(-u/\alpha), \quad \forall u \in \mathbb{R}_+ .$$

The idle time process $\{y(n), n \in \mathbb{Z}\}$ is ergodic (but not i.i.d.) with marginal law given by

$$P\{y(0) = 0\} = \rho \text{ and } P\{y(0) > u\} = (1-\rho)\exp(-u/\alpha), \quad \forall u \in \mathbb{R}_+ .$$

The sojourn time process $\{v(n), n \in \mathbb{Z}\}$ is ergodic (but not i.i.d.) with marginal law given by

$$P\{v(0) > u\} = \exp(-(1/\beta - 1/\alpha)u), \quad \forall u \in \mathbb{R}_+ .$$

The waiting time process $\{w(n), n \in \mathbb{Z}\}$ is ergodic (but not i.i.d.) with a marginal law, for $u \in \mathbb{R}_+$,

$$P\{w(0) = 0\} = 1 - \rho \text{ and } P\{w(0) > u\} = \rho \exp\left(-(1/\beta - 1/\alpha)u\right) .$$

As $t(n-1,1) = y(n) + s(n)$, eqn (2), we see that the processes $\{y(n) + s(n), n \in \mathbb{Z}\}$ and $\{s(n), n \in \mathbb{Z}\}$ are i.i.d. but not the process $\{y(n), n \in \mathbb{Z}\}$. On the other hand, neither the waiting time process $\{w(n), n \in \mathbb{Z}\}$ nor the sojourn times process $\{w(n) + s(n), n \in \mathbb{Z}\}$ are i.i.d.

Now, let us consider a *finite* number of queues (numbered from 0 to K) in tandem. We assume that the services are i.i.d., exponentially distributed of mean β; and that the initial arrival process is Poisson of rate $1/\alpha$ and independent of the services. Then, each queue is equivalent to a $M/M/1/\infty$ queue of parameters (α, β). In particular, for $k \in \{0, \dots, K\}$ and using the notations of §2.1, the processes $\{w(n,k), n \in \mathbb{Z}\}, \{y(n,k), n \in \mathbb{Z}\}$, $\{v(n,k), n \in \mathbb{Z}\}$ and $\{t(n,k), n \in \mathbb{Z}\}$ are distributed as the processes $\{w(n)\}$, $\{y(n)\}$, $\{v(n)\}$ and $\{t(n,1)\}$ described above.
The workloads $w(0,k), k = 0, \dots, K$, are dependent but the sojourn times $v(0,k), k = 0, \dots, K$, are independent random variables.

To summarize, the variables $t(n,0), n = 0,1,\ldots$, and the variables $v(0,k), k = 0,1,\ldots$, are i.i.d., exponentially distributed, and if the mean of $t(n,0)$ or $v(0,k)$ is $x > \beta$, then the mean of the other variable is $\beta x/(x-\beta)$.
This is another illustration of the duality between sojourn and inter-arrival times.

The above results are exposed in most textbooks on queueing theory. The original references are Burke [8, 9], Reich [37, 38] and Jackson [21].

4 The Quadrant Version

We consider the quadrant problem as defined in §2.2. We work with the variables w and y. The duality between these variables, see §2.1, plays a central role in the proof of Theorem 4.1 below.

4.1 Ergodic theorem

Let us consider the one-sided infinite vectors:

$$\begin{aligned} W(k) &= \{w(1,k), w(2,k), \ldots, w(n,k), \ldots\}, k \in \mathbb{N} \\ Y(k) &= \{y(1,k), y(2,k), \ldots, y(n,k), \ldots\}, k \geq -1\,. \\ \hat{W}(n) &= \{w(n,0), w(n,1), \ldots, w(n,k), \ldots\}, n \in \mathbb{N} \\ \hat{Y}(n) &= \{y(n,0), y(n,1), \ldots, y(n,k), \ldots\}, n \in \mathbb{N}^*\,. \end{aligned} \tag{17}$$

We are concerned with the problem of the convergence (in a sense to be precised) of the processes $Y(k), W(k), \hat{Y}(n)$ and $\hat{W}(n)$ to a limit distribution. The convergence of the processes $Y(k)$, $W(k)$, $\hat{Y}(n)$ and $\hat{W}(n)$ to a limit distribution, will imply the same result for the processes $T(k) = \{t(n,k), n \in \mathbb{N}\}$, $V(k) = \{t(n,k), n \in \mathbb{N}\}$, $\hat{T}(n) = \{t(n,k), k \in \mathbb{N}\}$ and $\hat{V}(n) = \{v(n,k), k \in \mathbb{N}\}$ respectively (see eqn (2)).

Theorem 4.1. *Let us assume that $\{s(n,k), n \in \mathbb{Z}, k \in \mathbb{Z}\}$ is spatially ergodic and θ-compatible and that $E[s(n,0)] = E[s(0,0)], \forall n$. We also assume that $\{w(0,k), k \in \mathbb{N}\}$ is ergodic and θ-compatible and verifies $0 < E[w(0,0)] < \infty$. At last, we assume that the variables $y(n,-1), n \in \mathbb{N}^*$, are a.s. finite. Let M_k be the law of $(Y(k), W(k))$. Then there exists a measure M in $\mathcal{M}(\mathbb{R}_+^{\mathbb{N}} \times \mathbb{R}_+^{\mathbb{N}})$ such that $M_k \xrightarrow{w} M$. Furthermore, M is invariant, meaning that if the joint law of $(Y(0), W(0))$ is M then so is the one of $(Y(1), W(1))$.*
Dually, assume that $\{s(n,k), n \in \mathbb{Z}, k \in \mathbb{Z}\}$ is temporally ergodic and θ-compatible and that $E[s(0,k)] = E[s(0,0)], \forall k$; that $\{y(n,-1), n \in \mathbb{N}^*\}$ is ergodic and θ-compatible and verifies $0 < E[y(1,-1)] < \infty$; and that the variables $w(0,k), k \in \mathbb{N}$, are a.s. finite. Let $\hat{M}_k$ be the law of $(\hat{Y}(k), \hat{W}(k))$. There exists an invariant measure $\hat{M}$ such that $\hat{M}_k \xrightarrow{w} \hat{M}$.

Proof. We prove the result for the processes $\{Y(k)\}$ and $\{W(k)\}$. For the processes $\{\hat{Y}(n)\}$ and $\{\hat{W}(n)\}$, the proof is analogous due to the symmetry between w and y in eqn (6).
By assumption, we have $s(n,k) = s(n,0) \circ \theta^k$ and $w(0,k) = w(0,0) \circ \theta^k$, for $n \in \mathbb{N}, k \in \mathbb{N}$.
Let us focus on customer 1. We have, eqn (4), for $k \in \mathbb{N}$,

$$y(1,k) = [y(1,k-1) + s(1,k-1) - (s(0,0) + w(0,0)) \circ \theta^k]^+\,.$$

We interpret this as the equation of a single server queue, see §3.2, y playing the role of w, w the role of y and $s+w$ the role of t. It follows from the assumptions that $s = \{s(0,k), k \in \mathbb{N}\}$ and $\{(s(0,0)+w(0,0)) \circ \theta^k, k \in \mathbb{N}\}$ are ergodic and θ-compatible and that $E[s(0,0)+w(0,0)] > E[s(0,0)]$. Hence, we can apply Loynes' results and we are in the stability case. We deduce that $y(1,k)$ and $w(1,k)$ couple in finite time with two ergodic and θ-compatible processes which we denote $y_1 \circ \theta^k$ and $w_1 \circ \theta^k$. Furthermore, we obtain from (7) that $E[w_1] = E[w(0,0)] > 0$.

Now let us consider customer 2. We have

$$y(2,k) = [y(2,k-1) + s(2,k-1) - (s(1,k) + w(1,k))]^+ ,$$

and $s(1,k) + w(1,k), k \in \mathbb{N}$, couples in finite time with the ergodic and θ-compatible process $(s(1,0)+w_1) \circ \theta^k$. We are still in the stability case as $E[s(1,0)+w_1] > E[s(0,0)]$, and we can apply the strengthening of Loynes' result recalled in §3.2. We obtain that $y(2,k)$ and $w(2,k)$ couple in finite time with two ergodic and θ-compatible processes that we denote by $y_2 \circ \theta^k$ and $w_2 \circ \theta^k$. Furthermore, we have $E[w_2] = E[w(0,0)] > 0$.

Inductively, we obtain that for all $n \in \mathbb{N}$, the processes $\{y(n,k), k \in \mathbb{N}\}$ and $\{w(n,k), k \in \mathbb{N}\}$ couple in finite time with the ergodic and θ-compatible processes $\{y_n \circ \theta^k, k \in \mathbb{N}\}$ and $\{w_n \circ \theta^k, k \in \mathbb{N}\}$. Hence, for any finite $n \in \mathbb{N}^*$, we obtain that $(y(1,k), \ldots, y(n,k))$ and $(w(1,k), \ldots, w(n,k))$ couple in finite time with $(y_1 \circ \theta^k, \ldots, y_n \circ \theta^k)$ and $(w_1 \circ \theta^k, \ldots, w_n \circ \theta^k)$, respectively. It implies that the finite dimensional marginals of $(Y(k), W(k))$ converge weakly to the ones of (Y, W) with $Y = (y_1, \ldots, y_n, \ldots)$ and $W = (w_1, \ldots, w_n, \ldots)$. We conclude, see §3.1, that $(Y(k), W(k))$ converges weakly to (Y, W).

Let M be the distribution of (Y, W). It remains to be shown that M is an invariant distribution. Let us consider the quadrant problem with the boundary conditions $\{w(0,k), k \in \mathbb{N}\}$ and $y(n,-1) = y_n, n \geq 1$. It follows directly from the above construction that $(Y(k), W(k)) = (Y(0), W(0)) \circ \theta^k$, hence M is an invariant distribution. □

The above proof shows in fact a result slightly stronger than just the weak convergence of M_k to M. Indeed, the finite-dimensional marginals of M_k converge in total variation to the corresponding marginals of M. We recall that there is *total variation convergence* of P_n to P if $\sup_{A \in \mathcal{B}} |P_n\{A\} - P\{A\}| \stackrel{n \to +\infty}{\longrightarrow} 0$ and we recall that coupling implies total variation convergence of the distributions.

In the statement of Theorem 4.1, let us assume that we have $w(0,k) = 0$ a.s., the other assumptions remaining the same. We can follow the same steps as in the proof of Theorem 4.1 but we are in the instability case, see §3.2. Hence we obtain that $y(n,k)$ converges a.s. to $+\infty$ with k, and that $w(n,k)$ converges a.s. to 0 with k. If we have that $w(0,k) = 0$ a.s. and $E[y(n,-1)] > 0$ (or vice-versa) and that the field $\{s(n,k), n \in \mathbb{N}, k \in \mathbb{N}\}$ is spatially and temporally ergodic, then we have a degenerate limit in one of the two directions and a non-degenerate one in the other direction.

The last argument in the proof of Theorem 4.1 shows that the processes $\{y_n \circ \theta^k, n \in \mathbb{N}, k \in \mathbb{N}\}$ and $\{w_n \circ \theta^k, n \in \mathbb{N}, k \in \mathbb{N}\}$ are spatially ergodic. On the other hand, they have no reason to be temporally stationary or ergodic as the initial workload process $\{w(0,k)\}$ is an arbitrary ergodic process.

The limit distribution M depends only on the joint distribution of $\{w(0,k), k \in \mathbb{N}\}$ and the service process. It does not depend on $\{y(n,-1), n \in \mathbb{N}^*\}$ on which no assumption is made. A natural (and open) question is whether the process $\{y_n, n \geq l\}$ admits a (weak) limit when l goes to infinity. This question seems to be closely related to the half-plane problem. We will come back to these questions in §6.

The above remarks and questions admit dual counterparts for $\hat{Y}(n)$ and $\hat{W}(n)$.

Example 4.2 (Exponential case). Here, it is more convenient to give the results using as variables the inter-arrival times $t(n,k)$ and the sojourn times $v(n,k) = w(n,k)+s(n,k)$. We assume that the services are i.i.d. and exponentially distributed of mean β. We assume that the boundary processes $\{v(0,k), k \in \mathbb{N}\}$ and $\{t(n,0), n \in \mathbb{N}\}$ are i.i.d., mutually independent, independent of the services, and with marginal laws given by, for $u \in \mathbb{R}_+$,

$$P\{v(0,k) > u\} = \exp(-u/\alpha_1), \text{ and } P\{t(n,0) > u\} = \exp(-u/\alpha_2)\,.$$

The corresponding boundary conditions $\{w(0,k), k \in \mathbb{N}\}$ and $\{y(n,-1), n \in \mathbb{N}^*\}$ are mutually independent, independent of the services, and distributed respectively as the stationary workload in a $M/M/1/\infty$ queue of parameters (α_1, β) and the stationary idle time in a $M/M/1/\infty$ queue of parameters (α_2, β).

We are in the range of application of Theorem 4.1. Coupling Theorem 4.1 with the results recalled in Example 3.1, we obtain the following when k goes to infinity. The processes $T(k) = \{t(n,k), n \in \mathbb{N}\}$ and $V(k) = \{v(n,k), n \in \mathbb{N}\}$ converge weakly to the *i.i.d.* process $T = \{t_n, n \in \mathbb{N}\}$ and to the *ergodic* process $V = \{v_n, n \in \mathbb{N}\}$, with respective marginal distributions given by, for $u \in \mathbb{R}_+$,

$$P\{t_n > u\} = \exp(-(1/\beta - 1/\alpha_1)u) \quad \text{and} \quad P\{v_n > u\} = \exp(-u/\alpha_1)\,. \tag{18}$$

The processes T and V are distributed respectively as the inter-arrival (or inter-departure) and sojourn time processes in a $M/M/1/\infty$ queue with parameters $(\beta\alpha_1/(\alpha_1 - \beta), \beta)$.

The dual processes $\hat{T}(n) = \{t(n,k), k \in \mathbb{N}\}$ and $\hat{V}(n) = \{v(n,k), k \in \mathbb{N}\}$ converge weakly to the *ergodic* process $\hat{T} = \{\hat{t}_k, k \in \mathbb{N}\}$ and to the *i.i.d.* process $\hat{V} = \{\hat{v}_k, k \in \mathbb{N}\}$ with respective marginal distributions given by, for $u \in \mathbb{R}_+$,

$$P\{\hat{t}_n > u\} = \exp(-u/\alpha_2) \quad \text{and} \quad P\{\hat{v}_n > u\} = \exp(-(1/\beta - 1/\alpha_2)u)\,. \tag{19}$$

The processes $\hat{T}$ and $\hat{V}$ are respectively distributed as the sojourn time and inter-departure processes in a $M/M/1/\infty$ queue of parameters $(\beta\alpha_2/(\alpha_2 - \beta), \beta)$. We summarize some of the results in Figure 3.
If $1/\beta \neq 1/\alpha_1 + 1/\alpha_2$, then the two limit processes are not consistent as the laws of t_0 and $\hat{t}_0$ (respectively of v_0 and $\hat{v}_0$) are not the same.
If $1/\beta = 1/\alpha_1 + 1/\alpha_2$, for all $n \in \mathbb{N}$ and $k \in \mathbb{N}$, the processes $T(k), V(k), \hat{T}(n)$ and $\hat{V}(n)$ are distributed as the limit processes $T, V, \hat{T}$ and $\hat{V}$ in (18) and (19). In particular,

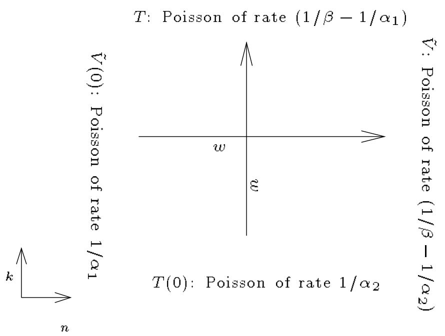


Figure 3: Limit processes for the model with exponential queues.

the random fields $\{t(n,k), n \in \mathbb{N}, k \in \mathbb{N}\}$ and $\{v(n,k), n \in \mathbb{N}, k \in \mathbb{N}\}$ are spatially and temporally ergodic.
When $\alpha_1 = \alpha_2 = 2\beta$, we obtain a completely symmetric situation where the laws of T and $\hat{V}$ on one hand, $\hat{T}$ and V on the other hand, are the same.

Applications of the result

Theorem 4.1 is obtained under the assumption that the boundary sequence $\{w(0,n), n \in \mathbb{N}\}$ is ergodic and verifies $E[w(0,0)] > 0$. Let us mention some models where it would be natural for this assumption to be verified.

Assume that each station requires a *warm-up* time before serving the first customer to visit it (i.e. customer 0). From that customer on, no more warming-up is needed, and the system behaves in the usual way. Let $\delta(k)$ be the warm-up time of station k, we have $w(0,k) = \delta(k)$.

In a slightly different way, we may assume that the stations "arrive" in the system (i.e. get ready to operate) at the epochs $\{\tilde{A}(0,k), k \in \mathbb{N}\}$. In this case, the boundary values $w(0,k)$ can be computed using the equation $w(0,k) = [\tilde{A}(0,k) - A(0,k)]^+$, see (3).

5 The Half-Plane Version

We consider the half-plane problem as defined in §2.2. The half-plane problem can be reduced to the quadrant one, with the boundary conditions $w(0,k), k \in \mathbb{N}$, being computed according to eqn (14). The difficulty is that we have no information on the process $\{w(0,k), k \in \mathbb{N}\}$. In particular it has 'a-priori' no reason to be ergodic (whatever the assumptions on the services and the initial arrival process are) and we cannot apply Theorem 4.1.
In fact, the half plane problem is more difficult than the quadrant one. Below, we make stronger assumptions on the services and we obtain weaker results than in §4. Also, as opposed to the quadrant case, we work with the variables v and t instead of w and y. These variables are in duality as shown in eqn (9).

Stochastic assumptions The service times $s(n,k), n \in \mathbb{Z}, k \in \mathbb{Z}$, are assumed to be i.i.d. random variables, with $E[s(0,0)] < \infty$. The only assumption on the service distribution is that it is not degenerate, i.e.

$$P\{s(0,0) \neq E[s(0,0)]\} > 0 .$$

The case $s(0,0) \equiv E[s(0,0)]$ is not interesting. Indeed, for an inter-arrival process such that $P\{t(0,0) \geq E[s(0,0)]\} = 1$, the only effect of the queue is to delay all arriving customers by a constant. Hence, for any inter-arrival process $\{t(n,0), n \in \mathbb{Z}\}$, we have $t(n,k) = t(n,1), n \in \mathbb{Z}, \forall k \geq 1$.
The initial inter-arrival process $\{t(n,0), n \in \mathbb{Z}\}$ is ergodic and independent of the services. We assume that

$$\alpha = E[t(0,0)] \; > \; E[s(0,0)] = \beta .$$

We say that (α, β) are the parameters of the problem. It follows from Loynes' result, recalled in §3.2, that at each station k, the processes $\{t(n,k), n \in \mathbb{Z}\}$ and $\{v(n,k), n \in \mathbb{Z}\}$ are well defined and ergodic and we have $E[t(0,k)] = \alpha$. Using Kendall's notation, each queue is of type $G/GI/1/\infty$ FIFO.

5.1 Ergodic conjecture

Let us define the two-sided ergodic processes

$$T(k) = \{t(n,k), n \in \mathbb{Z}\}, \;\; V(k) = \{v(n,k), n \in \mathbb{Z}\} .$$

For the reasons detailed in §2.2, $(T(k), V(k))$ is a Markov Chain. However as $\mathbb{R}^{\mathbb{Z}}$ is a complicated space, there is no hope to obtain a Harris-type Markov Chain. Once again, we concentrate on weak convergence results. To get a parallel with the results in §4, we would like to prove the weak convergence of $T(k)$ and $V(k)$ to a limit. Indeed, we expect the next result to be true. A similar conjecture was mentioned in [33].

Conjecture 5.1. *We consider the half-plane problem and we fix $\alpha > \beta$.*
(i) There exists a unique distribution $P_\alpha \in \mathcal{M}(\mathbb{R}_+^{\mathbb{Z}})$ which is ergodic, has one-dimensional marginals of mean α, and such that $\{t(n,1), n \in \mathbb{Z}\}$ is distributed as P_α if $\{t(n,0), n \in \mathbb{Z}\}$ is distributed as P_α. Let Q_α be the distribution of the corresponding sojourn time process.
(ii) Let $\{t(n,0), n \in \mathbb{Z}\}$ be an ergodic inter-arrival process of mean α. Let P_k and Q_k, $k \in \mathbb{N}^*$, be the distributions of $\{t(n,k), n \in \mathbb{Z}\}$ and $\{v(n,k), n \in \mathbb{Z}\}$ respectively. We have $P_k \xrightarrow{w} P_\alpha$ and $Q_k \xrightarrow{w} Q_\alpha$.
(iii) Let $\{t(n,0), n \in \mathbb{Z}\}$ be distributed as P_α and let $N(\alpha) = E[v(0,0)]$. Then the random field $\{t(n,k), v(n,k), n \in \mathbb{Z}, k \in \mathbb{N}\}$ is temporally and *spatially ergodic. Furthermore, for all n, the two-sided extensions $\{t(n,k), k \in \mathbb{Z}\}$ and $\{v(n,k), k \in \mathbb{Z}\}$ are distributed as $Q_{N(\alpha)}$ and $P_{N(\alpha)}$ respectively.*

Let us assume that points (i) and (ii) are verified and let us comment on point (iii). For the random field $\{t(n,k), v(n,k), n \in \mathbb{Z}, k \in \mathbb{N}\}$, temporal ergodicity and spatial stationarity are immediate. However, it is not clear a-priori how to prove the spatial ergodicity.

Assume that spatial ergodicity holds. We obtain readily that the quadrant random field $\{v(n,k), n \in \mathbb{N}, k \in \mathbb{N}\}$ is temporally and spatially ergodic. Using Kolmogorov's theorem, we can extend the random processes to get a temporally and spatially ergodic random field $\{v(n,k), n \in \mathbb{N}, k \in \mathbb{Z}\}$. Now, by duality, see §2.1, we conclude that $\{v(n,k), k \in \mathbb{Z}\}$ is distributed as $P_{N(\alpha)}$ where

$$N(\alpha) = E[v(0,0)].$$

By applying the same argument a second time, we obtain that we must have $N(N(\alpha)) = \alpha$.

When the conjecture is verified, we obtain by considering the r.v. $v(0,0)$ that the one-dimensional marginal of $P_{N(\alpha)}$ is equal to the one of Q_α. On the other hand there is no reason why we should have $P_{N(\alpha)} = Q_\alpha$, see for instance Example 5.2.

Several partial results support Conjecture 5.1. Firstly, the conjecture is proved in the case of exponential service times, see [33] and Examples 3.1 and 5.2.
Secondly, Chang [10] proved that if the services have unbounded support and if there exists a distribution P_α satisfying point (i) of the conjecture, then it is unique.
Thirdly, Mountford and Prabhakar [34] proved that if point (i) is verified and if the services have an *increasing hazard rate* and an unbounded support then the Cesaro averages of the inter-arrival processes weakly converge to P_α, where α is the mean of an inter-arrival.
In Theorem 5.3 below, we prove a result which provides further evidence for the conjecture, although it is still far from completely solving it. In particular it provides a lot of information on the potential function $N(.)$ defined in (iii).

Example 5.2 (Exponential case). Let us assume that the service times are exponentially distributed, i.e. $P\{s(0,0) > u\} = \exp(-u/\beta), u \in \mathbb{R}_+$.
Then Conjecture 5.1 is true. The limit distribution for any arrival process of rate $1/\alpha$ $(\alpha > \beta)$ is a Poisson process of rate $1/\alpha$. The existence in point (i) was proved by Burke, see Example 3.1. The uniqueness was proved first by Anantharam [2]. The proof of point (ii) is due to Mountford and Prabhakar [33]. Preliminary results were obtained by Liggett and Shiga [27].
At last, point (iii) can be deduced from classical results on $M/M/1/\infty$ queues in tandem, see also Example 4.2. We obtain that P_α (resp. Q_α) is equal to the distribution of an i.i.d. (resp. ergodic but not i.i.d.) sequence of exponentials of parameter α (resp. $\alpha\beta/(\alpha-\beta)$). The function $N(.)$ defined in point (iii) of Conjecture 5.1 is the homographic function $N(x) = \beta x/(x-\beta)$. Note here the simple form of the function $\tilde{N}(x) = 1/N(1/x) = (1/\beta) - x$ (it is the function relating the rate of the arrival process at a station and the rate of the arrival process of a customer). The one dimensional marginals of P_α and $Q_{N(\alpha)}$ are identical but we do not have $P_\alpha = Q_{N(\alpha)}$.

The proofs of Anantharam and Mountford & Prabhakar heavily rely on the characteristics of the exponential distribution. It can be adapted to cover geometric service times. But, it is not clear how to generalize the approach to other service times. Indeed, apart from the exponential and geometric cases, there is no service distribution for which either point (i) or point (ii) of Conjecture 5.1 is proved.

5.2 Strong law of large numbers

If true, Conjecture 5.1 would be a powerful stability result. In particular, it would prove the weak convergence of $v(0,k)$ to a limit distribution which depends on the inter-arrival process only through its mean. As a first step, we establish a Strong Law of Large Numbers for $v(0,k), k \in \mathbb{N}$. This can be described as a stability result in an *average sense* or as a *hydrodynamic limit.*

Theorem 5.3. *We consider the half-plane problem with parameters (α, β). Let us assume that for some $a > 0$, we have*

$$E[s(0,0)^{3+a}] < \infty \,. \tag{20}$$

Then there exists a finite constant $N(\alpha)$ such that

$$\frac{1}{n}\sum_{k=0}^{n-1} v(0,k) \xrightarrow{n} N(\alpha) \;\; P-a.s. \tag{21}$$

The above convergence also holds in L_1, if we assume furthermore that the initial inter-arrival process verifies:

$$\exists c, \; \beta < c < \alpha, \;\; E\,[\,\sup_{n\in\mathbb{N}^*}[\sum_{i=-n}^{-1} c - t(i)]^+] \; < \; \infty \,. \tag{22}$$

We have $N(\alpha) = \sup_{x\geq 0}(\gamma(x) - \alpha x)$, where the function $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$ depends only on the service process. The function $N(.)$ is continuous, strictly decreasing, strictly convex, bijective on the interval $(\beta, +\infty)$, and verifies $N = N^{-1}$.

The limit $N(\alpha)$ is the asymptotic average time spent by a customer in a queue. An interesting feature is that $N(\alpha)$ depends on the inter-arrival process only through its mean. This provides a new justification for Conjecture 5.1.

Let us comment on the Assumptions (20) and (22). In order to get the L_1 convergence in (21), it is obviously necessary to have $E[v(0,0)] < \infty$. We now show that conditions (20) and (22) are not too far from this minimal condition. For simplicity, we use the shortened notations of §3.2. In a $G/GI/1/\infty$ FIFO stable queue, we have, see for instance [32] (the result is implicit in [23])

$$E[v(0)] < \infty \;\;\Longrightarrow\;\; E[s(0)^2] < \infty \,. \tag{23}$$

Our requirement for a stronger condition (moments of order 3+ for the services) appears in the technical Lemma 5.10.

Let $v_c(0)$ be the sojourn time in the $G/D/1/\infty$ stable queue obtained by keeping the arrival process unchanged and considering deterministic services equal to c (with $c < E[t(0)]$). Let $\beta = E[s(0)]$, it is proved in [12] that if $E[s(0)^2] < \infty$, then we have

$$\begin{aligned} E[v(0)] < \infty \quad &\Longrightarrow \quad E[v_\beta(0)] < \infty \\ \exists c > \beta, \; E[v_c(0)] < \infty \quad &\Longrightarrow \quad E[v(0)] < \infty \,. \end{aligned}$$

This can be rephrased as (using the waiting times instead of the sojourn times, and developping according to (11))

$$
\begin{aligned}
E\,[\sup_{n\geq 1}[\sum_{i=-n}^{-1} s(i)-t(i)]^{+}] < \infty \quad &\Longrightarrow \quad E\,[\sup_{n\geq 1}[\sum_{i=-n}^{-1} \beta - t(i)]^{+}] < \infty \\
\exists c > \beta,\ E\,[\sup_{n\geq 1}[\sum_{i=-n}^{-1} c - t(i)]^{+}] < \infty \quad &\Longrightarrow \quad E\,[\sup_{n\geq 1}[\sum_{i=-n}^{-1} s(i)-t(i)]^{+}] < \infty\,.
\end{aligned}
$$

We conclude that condition (22) is very close to the necessary condition $E[v(0,0)] < \infty$. Condition (22) appears in the decomposition technique of §5.2.4.

The remainder of the section is devoted to the proof of Theorem 5.3. It is carried out in several steps. First, we recall in §5.2.1 a formula, interesting in itself, expressing $\sum_k v(0,k)$ as a longest path in a graph. It enables us to use results developed in this context. The whole methodology is close to the one of Glynn and Whitt [19]. In §5.2.2, we prove a Strong Law of Large Numbers (SLNN) for the longest paths (associated with $\sum_k v(0,k)$) when restricted to a fixed direction. In §5.2.3, we control the speed of convergence in the SLLN, using Azuma's inequality. At last, in §5.2.4, the different pieces are brought together to provide a proof of Theorem 5.3.

5.2.1 Cumulative sojourn times as longest paths

It is convenient to start the numbering of the stations at a value which might be different from 0. For $K \in \mathbb{N}^*$, let us consider a tandem network with stations numbered $(-K+k, k \in \mathbb{N})$. In this network, the initial condition is $\{t(n,-K), n \in \mathbb{Z}\}$, the ergodic inter-arrival process in station $-K$.

We define the following quantities, for $k, l \in \mathbb{Z}, k < l$,

$$
V(n,[k,l]) = \sum_{i=k}^{l} v(n,i), \quad V(-k) = V(0,[-k,-1]), \quad V(k) = V(0,[0,k-1])\,. \tag{24}
$$

In words, $V(n,[k,l])$ is the total time spent in the system by customer n between its arrival at station k and its departure from station l. It is called a *cumulative sojourn time.*

If the processes $\{t(n,0), n \in \mathbb{Z}\}$ and $\{t(n,-k), n \in \mathbb{Z}\}$ are equal in distribution and are both independent of the services, then obviously the r.v.'s $V(k)$ and $V(-k)$ are equal in distribution. It is more convenient to work with the variables $\{V(-k), k \in \mathbb{N}^*\}$ (we use a Loynes' backward scheme with respect to the stations). More precisely, let $\{t(i), i \in \mathbb{Z}\}$ be our initial inter-arrival process. It is assumed to be independent of the services $\{s(n,k), n \in \mathbb{Z}, k \in \mathbb{Z}\}$. Then, for each $k \in \mathbb{N}^*$, we define $V(-k)$ with respect to the inter-arrival process $\{t(n,-k), n \in \mathbb{Z}\} = \{t(n), n \in \mathbb{Z}\}$.

Below, we prove our results (and in particular Theorem 5.3) on the "backward" variables $\{V(-k), k \in \mathbb{N}^*\}$. Then, we explain, at the end of §5.2.4, how to extend the results to the "forward" variables $\{V(k), k \in \mathbb{N}^*\}$.

Proposition 5.4. *Let $V(-K)$ be the cumulative sojourn time defined in (24). We have*

$$V(-K) = \sup_{n\geq 0} \left\{ \max_{0=i_0\leq i_1\leq\cdots\leq i_K=n} \left(\sum_{i=i_0}^{i_1} s(-i,-1) + \sum_{i=i_1}^{i_2} s(-i,-2) + \cdots + \sum_{i=i_{K-1}}^{i_K} s(-i,-K) \right) - \sum_{i=1}^{n} t(-i,-K) \right\}, \tag{25}$$

where we use the convention $\sum_{i=1}^{0} = 0$.

Proposition 5.4 was proved by several authors in the case $t(n,-K) \equiv 0$, see Tembe & Wolff [45], Muth [35] and Glynn & Whitt [19]. In the general case, it appears in Szczotka & Kelly [44], Theorem 2. For the sake of completeness, we recall the proof.

Proof. Let us prove that for $k > -K$, we have

$$\begin{aligned} V(n,[-K,k]) &= s(n,k) + \max \; [V(n,[-K,k-1]), \\ &\qquad V(n-1,[-K,k]) - t(n-1,-K)] \\ V(n,[-K,-K]) &= s(n,-K) + \max \; [V(n-1,[-K,-K]) - t(n-1,-K), 0] . \end{aligned} \tag{26}$$

The second equation is a simple transformation of Lindley's equation, see (4). We prove the first one. We distinguish between the cases $w(n,k) > 0$ and $w(n,k) = 0$. Let us assume that $w(n,k) > 0$. The inter-departure time from station k between customers $(n-1)$ and n is equal to $s(n,k)$, whereas the inter-arrival time between the same two customers in station $-K$ is $t(n-1,-K)$. We deduce that $V(n,[-K,k]) - V(n-1,[-K,k]) = s(n,k) - t(n-1,-K)$. Let us now assume that $w(n,k) = 0$. It is immediate from (24) that $V(n,[-K,k]) = V(n,[-K,k-1]) + s(n,k)$. Putting the two cases together, we deduce (26). Developing the recurrence (26) and using the equation for $V(n,[-K,-K])$, we obtain the formula of Proposition 5.4.

□

Equation (25) can be compared with eqn (24) when developed using eqn (11):

$$V(-K) = \sum_{k=1}^{K} s(0,-k) + \sum_{k=1}^{K} \left[\sup_{i\geq 1} \{ \sum_{j=1}^{i} s(-j,-k) - t(-j,-k) \} \right]^{+} . \tag{27}$$

The crucial point is that only the initial inter-arrivals $t(n,-K) = t(n)$ (and not $t(n,k), k > -K$) do appear in eqn (25).

Let $G(-K)$ be the weighted directed graph represented in Fig. 4. More precisely, the set of nodes is $\mathbb{Z} \times \{-K, -K+1, \ldots\}$, the set of arcs is $\{(n,k) \to (n-1,k), n \in \mathbb{Z}, k \geq -K\} \cup \{(n,k) \to (n,k-1), n \in \mathbb{Z}, k \geq -K+1\}$. The weights are

- $s(n,k)$ on node (n,k);

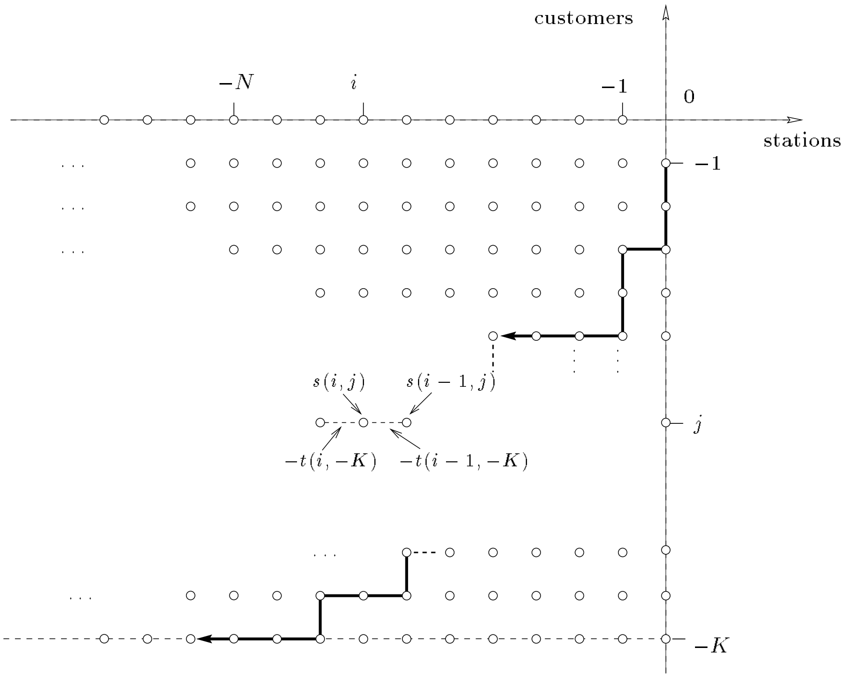


Figure 4: Interpretation of the cumulative sojourn time as a maximum-weight-path.

- $-t(n,-K)$ on arc $(n,k) \to (n-1,k), \forall k \geq -K$.

It follows from eqn (25) that $V(-K)$ can be interpreted as the weight of the maximum-weight-path originating from node $(0,-1)$ and ending in any node of the form $(n,-K), n \leq 0$. We deduce that $V(-K)$ can be interpreted as a last passage time in an oriented percolation model.

Remark 5.5. Let us define another directed graph, $\tilde{G}(-K)$, having the same sets of nodes and arcs as $G(-K)$ but with weights:

- $s(n,k)$ on node (n,k); $-t(n,k)$ on arc $(n,k) \to (n-1,k)$.

Using eqn (27), we see that $V(-K)$ also has a simple interpretation in graph $\tilde{G}(-K)$. It is the sum for k ranging from $-K$ to -1 of the weights of the maximum-weight-segment originating in $(0,k)$ and ending at one of the points $(n,k), n \leq 0$.

Going back to the graph $G(-K)$, let us consider a maximum-weight-path originating in (n,k). It follows from the proof of Prop. 5.4 that this path switches from line k to line $(k-1)$ at the first point $(n-i,k), i \geq 0$, such that $w(n-i,k) = 0$. The integer i can also be defined as the minimal integer l attaining the sup in the following expression (see eqn (16)):

$$w(n,k) = \sup_{l=1}^{+\infty} \left[\sum_{j=1}^{l} s(n-j,k) - t(n-j,k) \right]^+ ,$$

with the convention $i = 0$ if $w(n,k) = 0$. We deduce that a maximal-weight-path in $G(-K)$ is also a maximal-weight-path in $\tilde{G}(-K)$.
Here is a summary of the results proved.

1. In the graph $\tilde{G}(-K)$, $V(-K)$ is the sum (for k ranging from $-K$ to -1) of the weights of the maximum-weight-segments originating in $(0,k)$ and ending at one of the points $(n,k), n \leq 0$.

2. In the graph $G(-K)$, $V(-K)$ is the weight of the maximum-weight-path originating in $(0,-1)$ and ending at one of the points $(n,-K), n \leq 0$.

3. In the graphs $\tilde{G}(-K)$ and $G(-K)$, maximum-weight-paths originating in $(0,-1)$ and ending at one of the points $(n,-K), n \leq 0$, coincide (but, of course, not their respective weights).

5.2.2 Longest paths in a fixed direction

We work with the graph $G(-k)$, for $k \in \mathbb{N}^*$. Let $\Pi(-k)$ denote the set of all oriented paths originating at $(0,-1)$ and ending at one of the points $(-n,-k), n \in \mathbb{N}$. We define $\Pi_x(-k)$, $x \in \mathbb{R}_+$, to be the set of all paths of $\Pi(-k)$ having direction x, i.e. originating in $(0,-1)$ and ending at point $(-[xk],-k)$. Let π be a path in the graph $G(-k)$. We denote S_π the sum of the service times along this path, i.e.

$$S_\pi = \sum_{(u,v)\in\pi} s(u,v) .$$

We also introduce the notations

$$M(-k) = \sup_{\pi\in\Pi(-k)} S_\pi, \;\; M_x(-k) = \sup_{\pi\in\Pi_x(-k)} S_\pi .$$

By rewriting (25), we obtain

$$V(-K) = \sup_{x\in\mathbb{R}_+} \{M_x(-K) - \sum_{i=1}^{[xK]} t(-i)\} . \tag{28}$$

The aim of this section is to prove that $M_x(-K)/K$ converges a.s. and in L_1 to a finite constant.

Lemma 5.6. *If for some $a > 0$, $E[s(0,0)^2(\log^+ s(0,0))^{2+a}] < \infty$, then there exists a finite constant M_x such that $\limsup_K M_x(-K)/K \leq M_x$, $P - a.s.$*

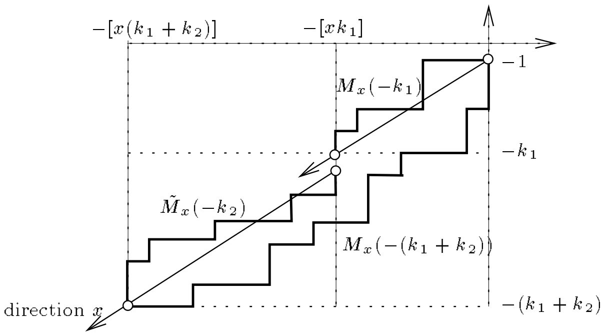
Lemma 5.6 is proved under a slightly different (and more general) form in Cox & al [11], Theorem 1. The reader is referred to this paper for the subtle proof of the result.

The moment condition in Lemma 5.6 is quite tight. Indeed, the result (23) shows that if $E[s(0,0)^2] = \infty$ then eqn (21) and a fortiori Lemma 5.6 fail to be true.

Theorem 5.7. *Let the service times verify the moment condition of Lemma 5.6. Then for all $x \in \mathbb{R}_+$, there exists a finite constant $\gamma(x)$ such that*

$$\lim_{K\to\infty} \frac{M_x(-K)}{K} = \gamma(x) \ \ P - a.s. \text{ and in } L_1 \, . \tag{29}$$

Proof. Under the assumption of exponential moments for the service time, this theorem is proved in Glynn and Whitt [19]. However, the exponential moment assumption was only used to get the existence of an upper bound for $M_x(-K)/K$. With the addition of Lemma 5.6, their proof carries over under the assumption $E[s(0,0)^2(\log^+ s(0,0))^{2+a}] < \infty$. It is based on Kingman's sub-additive ergodic Theorem, [24].



Indeed, the maximal weight $M_x(-k)$ is super-additive as illustrated in the figure above. For details, see Theorem 6.3 in [19]. Using additional arguments, we prove at the end of §5.2.4 that the convergence in (29) also holds for the variables $M_x(K)$. $\square$

In the next proposition, we list only the properties of $\gamma(x)$ that are needed for the proof of Theorem 5.3. Other results, as well as simulations, can be found in [19] and [20].

Proposition 5.8. *The function $\gamma(x)$, defined in (29), satisfies the following properties. We recall that $\beta = E[s(0,0)]$.*

(i) *γ is a continuous, strictly increasing and strictly concave function ;*

(ii) $\gamma(x) = x\gamma(1/x)$ *;*

(iii) $\gamma(x+y) - \gamma(x) \geq \beta y, \ \forall x, y > 0$ *;*

(iv) $\gamma(0) = \beta, \ \lim_{x\to\infty} \gamma(x)/x = \beta, \ \lim_{x\to\infty} \gamma'(x) = \beta$ *;*

(v) *if* $P\{s(0,0) \neq \beta\} > 0$, *then* $\gamma(x) > \beta + \beta x$, *for* $x > 0$.

Proof. Points (i), (ii), (iii) and (v) are proved in [19], Theorems 5.1 and 6.3. We prove the remaining point. The strong law of large numbers for the service times gives

$$\gamma(0) = \lim_{k\to\infty} \frac{1}{k} \sum_{i=-k}^{0} s(0,i) = E[s(0,0)] = \beta \, .$$

By applying (ii), we obtain $\lim_{x\to\infty}\gamma(x)/x = \gamma(0) = \beta$.
As the function $\gamma(.)$ is concave, it has left and right derivatives at each point. It follows from (iii) (for the left inequality) and from the concavity of γ (for the right inequality) that:

$$\forall x, y > 0,\ \beta \leq \frac{\gamma(x+y)-\gamma(x)}{y} \leq \frac{\gamma(x)-\gamma(0)}{x}\,.$$

As $\lim_{x\to\infty}\gamma(x)/x = \beta$, we obtain that $\lim_{x\to\infty}\gamma'(x) = \beta$ (for both the left and right derivatives).

□

Example 5.9 (Exponential case). Let us assume that the services are exponential of mean β. Then, we have

$$\gamma(x) = \beta + \beta x + 2\beta\sqrt{x}\,. \tag{30}$$

This was proved in [19], Theorem 6.1, see also Srinivasan [42]. The proof uses a hydrodynamic limit for the totally asymmetric exclusion process proved by Rost [39].
Let us now assume that the services are geometric of parameter $0 < p < 1$, i.e. $P\{s(0,0) = k\} = (1-p)p^k, k \in \mathbb{N}$. Let $\beta = p/(1-p)$ be the mean of the services. Then we have

$$\gamma(x) = \beta + \beta x + 2\sqrt{\beta(\beta+1)x}\,.$$

This was proved by Seppäläinen [40]. To the best of our knowledge, the exponential and geometric distributions are the only two cases where an explicit formula is known for $\gamma(x)$.

5.2.3 Speed of convergence

In order to carry on with the proof of Theorem 5.3, it is necessary to have a control on the speed of convergence of $M_x(-K)/K$ to $\gamma(x)$.
Due to the a-priori complete lack of independence or mixing properties of the sequence $\{M_x(-K), K \in \mathbb{N}\}$, we cannot expect to control this speed using large deviations or central limit theorems.
In fact, we apply a martingale method based on Azuma's inequality. For a survey on the method, called there "method of bounded differences", see McDiarmid [30]. For an utilization in a closely related model (percolation), and the proof of a stronger inequality than the one used here, see Kesten [22].

First of all, we need to truncate the service times. We define, for $a > 0$,

$$\hat{s}(i,j) = \min\left[s(i,j), (|i|+|j|)^{\frac{1}{2+a}}\right]\,. \tag{31}$$

We denote by $\hat{S}_\pi, \hat{M}_x, \ldots$, the quantities corresponding to $S_\pi, M_x, \ldots$, when $\{s(i,j)\}$ is replaced by $\{\hat{s}(i,j)\}$.

Lemma 5.10. *Let us assume that there exists $a > 0$ such that $E[s(0,0)^{3+a}] < \infty$. For the same a, we define the truncated variables $\hat{s}(i,j)$ as in (31). Then, we have*

$$\lim_{K\to+\infty} \frac{1}{K} \sup_{\pi\in\Pi(-K)} (S_\pi - \hat{S}_\pi) = 0, \quad P-a.s. \text{ and in } L_1 .$$

It implies that $\lim_{k\to\infty} \hat{M}_x(-k)/k = \hat{\gamma}(x)$ exists P-a.s. and in L_1 for all $x \in \mathbb{R}_+$ and that $\hat{\gamma}(x) = \gamma(x)$.

Proof. In the graph $G(-K)$, we define the set of paths $\mathcal{L}(-K) = \{\pi \in \Pi(-K) \mid (0,-1) \in \pi \text{ and } |\pi| = K\}$, where $|\pi|$ is the length (number of nodes) of π. The following result holds:

$$\lim_K \frac{1}{K} \max_{\pi\in\mathcal{L}(-K)} (S_\pi - \hat{S}_\pi) = 0 \quad P-a.s. \text{ and in } L_1 . \tag{32}$$

The a.s. convergence is proved in Gandolfi and Kesten [18], Lemma 1. In fact, it is even proved under a weaker moment assumption (the one of Lemma 5.6) and for non-oriented paths.
We obtain the L_1 convergence as an application of the Dominated-Convergence Theorem. Indeed, we have by Lemma 5.6,

$$\frac{1}{K} \max_{\pi\in\mathcal{L}(-K)} (S_\pi - \hat{S}_\pi) \leq \frac{1}{K} M_1(-K) \text{ and } \limsup_K \frac{1}{K} M_1(-K) \leq M_1 .$$

The result in (32) is not sufficient to prove the lemma as the length of the paths in $\Pi(-K)$ is unbounded. This is why the moment condition on the services is stronger here than in [18].
Let $\Delta = 1/K \sup_{\pi\in\Pi(-K)}(S_\pi - \hat{S}_\pi)$. We have, for $X > 0$,

$$\Delta \leq \underbrace{\frac{1}{K} \max_{\pi\in\Pi_X(-K)} (S_\pi - \hat{S}_\pi)}_{A} + \underbrace{\frac{1}{K} \sum_{k=0}^{K} \sum_{i=[XK]+1}^{\infty} s(-i,-k) - \hat{s}(-i,-k)}_{B(X,K)} .$$

To obtain the above inequality, we use the pathwise upper bound illustrated in Figure 5. Let us control the term A. Any path in $\Pi_X(-K)$ is of length $[(1+X)K]$. For a fixed $X > 0$, we obtain, using (32) and using that $S_\pi \geq \hat{S}_\pi$,

$$A = (1+X) \frac{1}{(1+X)K} \max_{\pi\in\mathcal{L}(-[(1+X)K])} (S_\pi - \hat{S}_\pi) \xrightarrow{K} 0 , \tag{33}$$

the convergence holding P-a.s. and in L_1.
Now, we have to control the term $B(X,K)$. Using that $s \geq \hat{s}, P-a.s.$, we get, for all $k \in \mathbb{N}^*$,

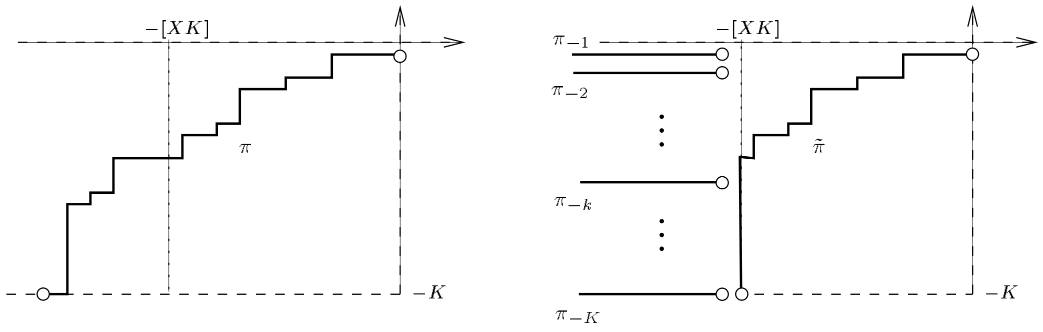


Figure 5: **Graphical proof of:** $S_\pi - \hat{S}_\pi \leq S_{\tilde{\pi}} - \hat{S}_{\tilde{\pi}} + \sum_{i=-K}^{-1} S_{\pi_i} - \hat{S}_{\pi_i}$.

$$\begin{aligned}
\sup_{n\geq 0} \sum_{i=0}^{n} s(-i,-k) - \hat{s}(-i,-k) &= \sum_{i=0}^{\infty} s(-i,-k) - \hat{s}(-i,-k) \\
&= \sum_{i=0}^{\infty} [s(-i,-k) - (i+k)^{\frac{1}{2+a}}]^+ .
\end{aligned}$$

Taking expectations and using the stationarity of the services and Fubini's Theorem, we get

$$\begin{aligned}
E[\sum_{i=0}^{\infty}[s(-i,-k) - (i+k)^{\frac{1}{2+a}}]^+] &= \sum_{i=k}^{\infty} \int_{i^{\frac{1}{2+a}}}^{\infty} P\{s(0,0) > t\} dt \\
&\leq \sum_{i=1}^{\infty} \int_{i^{\frac{1}{2+a}}}^{\infty} P\{s(0,0) > t\} dt \\
&\leq \int_0^{\infty} di \int_{i^{\frac{1}{2+a}}}^{\infty} P\{s(0,0) > t\} dt \\
&\leq \int_0^{\infty} t^{2+a} P\{s(0,0) > t\} dt \\
&= \frac{E[s(0,0)^{3+a}]}{3+a} < \infty . \qquad (34)
\end{aligned}$$

For a given $X > 0$, we have by Birkhoff's Ergodic Theorem,

$$\begin{aligned}
B(X,K) &= \frac{1}{K} \sum_{k=1}^{K} \sum_{i=[XK]+1}^{\infty} s(-i,-k) - \hat{s}(-i,-k) \\
&\leq \frac{1}{K} \sum_{k=1}^{K} \sum_{i=X+1}^{\infty} s(-i,-k) - \hat{s}(-i,-k) \stackrel{K\to\infty}{\longrightarrow} E[B(X,1)] ,
\end{aligned}$$

the convergence holding $P-a.s.$ and in L_1. Using (34), we obtain

$$E[B(X,1)] = E[\sum_{i=[X]+1}^{\infty} s(-i,-1) - \hat{s}(-i,-1)\ ] \overset{X\to\infty}{\longrightarrow}\ 0\,.$$

We conclude that for any $\varepsilon > 0$, we can choose $X > 0$ such that

$$\lim_K E[B(X,K)] \leq \varepsilon \quad \text{and} \quad \lim_K B(X,K) \leq \varepsilon \quad P-a.s.$$

Together with the convergence in (33), this concludes the proof of the first assertion. Using the inequality $M_x(-k) \geq \hat{M}_x(-k)$, we obtain P-a.s.,

$$\begin{aligned} 0 \ \leq\ \limsup_{k\to+\infty}\frac{1}{k}\sup_x M_x(-k) \ &- \ \limsup_{k\to+\infty}\frac{1}{k}\sup_x \hat{M}_x(-k) \\ &\leq\ \limsup_{k\to+\infty}\frac{1}{k}\sup_x (M_x(-k) - \hat{M}_x(-k)) \\ &\leq\ \limsup_{k\to+\infty}\frac{1}{k}\sup_x \max_{\pi\in\Pi_x(-k)} (S_\pi - \hat{S}_\pi) \ = \ 0\,. \end{aligned} \tag{35}$$

We deduce immediately the other assertions of the lemma. □

From now on, we work with the truncated variables $\hat{s}(i,j)$.

Lemma 5.11. *Under the moment condition (20), and setting $c = a/(2+a)$ and $d = (4+a)/(2+a)$, we obtain*

$$P\left\{\left|\frac{\hat{M}_x(-K)}{K} - E[\frac{\hat{M}_x(-K)}{K}]\right| > \varepsilon\right\} \leq 2\exp\left(-\frac{\varepsilon^2 K^c}{2(1+x)^d}\right)\,.$$

The remainder of the section is devoted to the proof of Lemma 5.11. We first recall Azuma's inequality in a general setting.

Azuma's inequality Consider a finite sequence of increasing σ-algebras

$$(\mathcal{F}_0 = \{\emptyset,\Omega\}) \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_N \subset \mathcal{F},\ N<\infty\,.$$

Let X be a $\mathcal{F}_N$-measurable random variable verifying $E[|X|] < \infty$. Define the *martingale difference sequence* $\{Y_n\}$ in the following way:

$$Y_n = E[X \mid \mathcal{F}_n] - E[X \mid \mathcal{F}_{n-1}],\ n = 1,\ldots,N\,.$$

Note that $E[X \mid \mathcal{F}_n]$ is a martingale with respect to the filtration $\{\mathcal{F}_n\}$, hence the name given to the sequence Y_n. Also note that we have $\sum_{n=1}^N Y_n = X - E[X]$.

Lemma 5.12 (Azuma's inequality). *Let X be a $\mathcal{F}_N$-measurable random variable and let $\{Y_n, n = 1, \dots, N\}$ be the associated martingale difference sequence. Then for all $u > 0$,*

$$P\{|X - E[X]| > u\} = P\{|\sum_{n=1}^{N} Y_n| > u\} \leq 2\exp\left(-\frac{u^2}{2\sum_{n=1}^{N} \|Y_n\|^2}\right),$$

where $\|Y\| = \inf\{c \, : \, |Y| < c, P - a.s.\}$.

For a proof of Lemma 5.12, see for instance [30], Lemma 4.1.

Let $\mathcal{F}_k, k \in \mathbb{N}^*$, be the sub-$\sigma$-algebra of $\mathcal{F}$ generated by $\{s(i,j), -k \leq i + j \leq -1\}$ and let $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Any directed path π intersects the set $\{s(i,j), i + j = -k\}$ at one point at most. For a given $K \in \mathbb{N}^*$, the r.v. $\hat{M}_x(-K)/K$ is $\mathcal{F}_{[(1+x)K]}$-measurable. We consider the martingale difference sequence:

$$Y_k = E[\hat{M}_x(-K)/K \mid \mathcal{F}_k] - E[\hat{M}_x(-K)/K \mid \mathcal{F}_{k-1}], \; k = 1, \dots, [(1+x)K].$$

Let us estimate $\|Y_k\|$. We define $\hat{M}_x^{(k)}(-K)$ to be the value of the maximum-weight-path in direction x when the service times $\{\hat{s}(i,j), i + j = -k\}$ are all replaced by 0, the other service times being unchanged. We have

$$\hat{M}_x^{(k)}(-K) \leq \hat{M}_x(-K) \leq \hat{M}_x^{(k)}(-K) + \max_{i+j=-k} \hat{s}(i,j).$$

Using the bound $\hat{s}(u,v) \leq [(1+x)K]^{1/(2+a)}$, for $(u,v) \in \Pi_x(-K)$, and taking conditional expectations, we obtain

$$E\left[\hat{M}_x^{(k)}(-K) \mid \mathcal{F}_k\right] \leq E\left[\hat{M}_x(-K) \mid \mathcal{F}_k\right] \leq E\left[\hat{M}_x^{(k)}(-K) \mid \mathcal{F}_k\right] + [(1+x)K]^{\frac{1}{2+a}}. \quad (36)$$

As $\hat{M}_x^{(k)}(-K)$ is independent of the service times $\{\hat{s}(i,j), i + j = -k\}$, we have $E[\hat{M}_x^{(k)}(-K) \mid \mathcal{F}_k] = E[\hat{M}_x^{(k)}(-K) \mid \mathcal{F}_{k-1}]$. Writing the inequalities in (36) for k and $k-1$, we get

$$\left|E\left[\hat{M}_x(-K) \mid \mathcal{F}_k\right] - E\left[\hat{M}_x(-K) \mid \mathcal{F}_{k-1}\right]\right| \leq [(1+x)K]^{\frac{1}{2+a}}.$$

This implies that $\|Y_k\| \leq (1+x)^{\frac{1}{2+a}} K^{\frac{1}{2+a}-1}$. Now applying Lemma 5.12, we obtain Lemma 5.11.

5.2.4 Proof of Theorem 5.3

We now have all the ingredients which are necessary to prove Theorem 5.3. We prove it first for the backward variables $V(-k), k \in \mathbb{N}^*$. Then we extend the results to the forward variables $V(k), k \in \mathbb{N}^*$.

Almost sure convergence for the backward variables

As detailed at the beginning of §5.2.1, there is a unique initial inter-arrival process, $\{t(i), i \in \mathbb{Z}\}$, and when the initial station is the one numbered $-k$, we set $\{t(i,-k), i \in \mathbb{Z}\} = \{t(i), i \in \mathbb{Z}\}$. In a consistent way with previous notations, we define

$$V_x(-k) = M_x(-k) - \sum_{j=1}^{[xk]} t(-j) . \tag{37}$$

We have $\sup_{x \in \mathbb{R}_+} V_x(-k) = V(-k)$, where $V(-k)$ is the cumulative sojourn time defined in eqn (25). There is no immediate way to prove that $\lim_{k \to +\infty} V(-k)/k$ exists and is finite, as $V(-k)$ is neither sub-additive nor super-additive. It follows from Theorem 5.7 and Birkhoff's Ergodic Theorem that we have, $P - a.s.$ and in L_1,

$$\lim_{k \to +\infty} \frac{V_x(-k)}{k} = \gamma(x) - \alpha x . \tag{38}$$

We set $\delta(x) = \gamma(x) - \alpha x$. As the function γ is concave, verifies $\lim_{x \to \infty} \gamma'(x) = \beta$ and $\gamma(x) > \beta + \beta x$ for $x > 0$ (Proposition 5.8), we deduce that there exists $x^* > 0$ such that

$$\delta(x^*) = \sup_{x \in \mathbb{R}_+} \delta(x) . \tag{39}$$

In particular, we have $\delta(x^*) > \delta(0) = \beta$. We want to prove the following interchange between a supremum and a limit.

$$\begin{aligned} \lim_{k \to +\infty} V(-k)/k &= \lim_{k \to +\infty} \sup_{x \in \mathbb{R}_+} V_x(-k)/k \\ &\stackrel{?}{=} \sup_{x \in \mathbb{R}_+} \lim_{k \to +\infty} V_x(-k)/k = \delta(x^*) . \end{aligned} \tag{40}$$

If we manage to do so, we will have proved the a.s. convergence part of Theorem 5.3 (for the variables $V(-k), k \in \mathbb{N}^*$), the function γ introduced in the statement of the Theorem being equal to the one defined in Theorem 5.7.

Using the inequality $V(-k) \geq V_{x^*}(-k)$ and the a.s. convergence of $V_{x^*}(-k)/k$ to $\delta(x^*)$, we obtain that

$$\liminf_{k \to +\infty} V(-k)/k \geq \delta(x^*), \quad P - a.s. \tag{41}$$

For the converse, we proceed as follows. We consider the truncated variables $\{\hat{s}(i,j)\}$ defined as in Lemma 5.10. We define the corresponding variables $\hat{V}(-k)$ and $\hat{V}_x(-k)$, see (28) and (37). We prove that we have

$$\limsup_{k} \hat{V}(-k)/k \leq \delta(x^*), \quad P - a.s. \tag{42}$$

Using that $V_x(-k) - \hat{V}_x(-k) = M_x(-k) - \hat{M}_x(-k)$, we deduce that the inequalities in (35) remain valid for the variables $V_x(-k)$ and $\hat{V}_x(-k)$. We deduce that (42) implies that

$$\limsup_k V(-k)/k \leq \delta(x^*), \quad P-a.s. \tag{43}$$

Together with (41), this completes the proof of (40).
It remains to be shown that (42) is verified. For $k \in \mathbb{N}^*$, we define the countable set $k^{-1}\mathbb{N} = \{0, 1/k, 2/k, \ldots\}$. We have

$$\hat{V}(-k) = \sup_{x \in \mathbb{R}_+} \hat{V}_x(-k) = \sup_{x \in k^{-1}\mathbb{N}} \hat{V}_x(-k) \, .$$

We now apply a decomposition technique. A similar trick was used to study the finiteness of the moments of the waiting-time in a single server queue in [12], Proposition 4. Let $\eta > 0$ be such that $\beta < \alpha - \eta < \alpha$. We have

$$\begin{aligned}
\frac{\hat{V}(-k)}{k} &= \sup_{x \in k^{-1}\mathbb{N}} \{\frac{\hat{M}_x(-k)}{k} - \frac{1}{k}\sum_{i=-xk}^{-1} t(i)\} \\
&= \sup_{x \in k^{-1}\mathbb{N}} \{\frac{\hat{M}_x(-k)}{k} - (\alpha-\eta)x + \frac{1}{k}\sum_{i=-xk}^{-1} \alpha - \eta - t(i)\} \\
&\leq \sup_{x \in k^{-1}\mathbb{N}} \{\frac{\hat{M}_x(-k)}{k} - (\alpha-\eta)x\} + \frac{1}{k}\sup_{n\in\mathbb{N}}\{\sum_{i=-n}^{-1} \alpha - \eta - t(i)\}. \qquad (44)
\end{aligned}$$

Roughly speaking, we replace the original inter-arrival process of mean α by a deterministic inter-arrival process of marginals $\alpha - \eta$.
By Birkhoff's Theorem, we have

$$\lim_n \frac{1}{n}\sum_{i=-n}^{-1} \alpha - \eta - t(i) = -\eta < 0, \quad P-a.s.$$

We deduce that the rv $\sup_{n\in\mathbb{N}} \sum_{i=-n}^{-1} \alpha - \eta - t(i)$ is almost surely finite. We conclude that for all $\eta > 0$, we have

$$\frac{1}{k}\sup_{n\in\mathbb{N}}\{\sum_{i=-n}^{-1} \alpha - \eta - t(i)\} \stackrel{k\to\infty}{\longrightarrow} 0, \quad P-a.s.$$

Let

$$\frac{\hat{V}^{(\eta)}(-k)}{k} = \sup_{x \in k^{-1}\mathbb{N}} \{\frac{\hat{M}_x(-k)}{k} - (\alpha-\eta)x\}, \quad \frac{\hat{V}_x^{(\eta)}(-k)}{k} = \frac{\hat{M}_x(-k)}{k} - (\alpha-\eta)x \, .$$

By going to the limit in (44), we get that for all η,

$$\limsup_k \hat{V}(-k)/k \leq \limsup_k \hat{V}^{(\eta)}(-k)/k \, ,$$

which implies that

$$\limsup_k \hat{V}(-k)/k \le \lim_{\eta \searrow 0} \limsup_k \hat{V}^{(\eta)}(-k)/k \,.$$

Hence, if we show that

$$\lim_\eta \limsup_k \hat{V}^{(\eta)}(-k)/k \le \delta(x^*), \quad P-a.s.\,, \tag{45}$$

then the proof of (42) will follow. We have

$$P\{\hat{V}^{(\eta)}(-k)/k \ge \delta(x^*) + u\} \le \sum_{x \in k^{-1}\mathbb{N}} P\{\hat{V}_x^{(\eta)}(-k)/k \ge \delta(x^*) + u\} \,.$$

We are going to bound the terms on the right-hand side in two different ways for x small and x large. We need some preliminary results.
We have, from Theorem 5.7, $\lim_k \hat{V}_x^{(\eta)}(-k)/k = \gamma(x) - (\alpha - \eta)x = \delta(x) + \eta x$, P-a.s. It easily follows from the properties of $\gamma(.)$ that

$$\sup_x \{\gamma(x) - (\alpha - \eta)x\} \ge \delta(x^*) \quad \text{and} \quad \lim_{\eta \searrow 0} \sup_x \{\gamma(x) - (\alpha - \eta)x\} = \delta(x^*) \,.$$

Let us consider $\varepsilon > 0$. From now on, we assume η to be such that for all $0 < \eta' \le \eta$, $\delta(x^*) + \varepsilon \ge \sup_x\{\gamma(x) - (\alpha - \eta')x\}$.
A real-valued super-additive sequence $\{u_n, n \in \mathbb{N}\}$ satisfies $\lim_n u_n/n = \sup_n u_n/n$. As $E[\hat{V}_x^{(\eta)}(-k)/k]$ is super-additive, we obtain for all x,

$$\delta(x^*) + \varepsilon \ge \sup_x\{\gamma(x) - (\alpha - \eta)x\} \ge \delta(x) + \eta x \ge E[\hat{V}_x^{(\eta)}(-k)/k] \,. \tag{46}$$

Let us consider the function $f(x) = \delta(x^*) - \delta(x)$. It is positive, convex, decreasing on $[0, x^*]$ and increasing on $[x^*, \infty)$. Furthermore, we have $\lim_{+\infty} f(x)/x = \alpha - \beta$ (Prop. 5.8, (iv)). Let us fix $\nu < \alpha - \beta$. It follows from the above analysis that there exists $\xi^* > x^*$ such that

$$\delta(x^*) - \delta(x) \ge \nu x \mathbb{1}\{x \ge \xi^*\} \,. \tag{47}$$

This lower bound is illustrated on Figure 6. Let us set (the dependence over k and u being implicit)

$$A_x(\eta) = P\{\hat{V}_x^{(\eta)}(-k)/k \ge \delta(x^*) + u\}. \tag{48}$$

We have

$$\begin{aligned} A_x(\eta) &= P\{\, \frac{\hat{V}_x^{(\eta)}(-k)}{k} - E[\frac{\hat{V}_x^{(\eta)}(-k)}{k}] \ge \delta(x^*) - E[\frac{\hat{V}_x^{(\eta)}(-k)}{k}] + u \,\} \\ &= P\{\frac{\hat{M}_x(-k)}{k} - E[\frac{\hat{M}_x(-k)}{k}] \ge \delta(x^*) - E[\frac{\hat{V}_x^{(\eta)}(-k)}{k}] + u\} \,. \end{aligned}$$

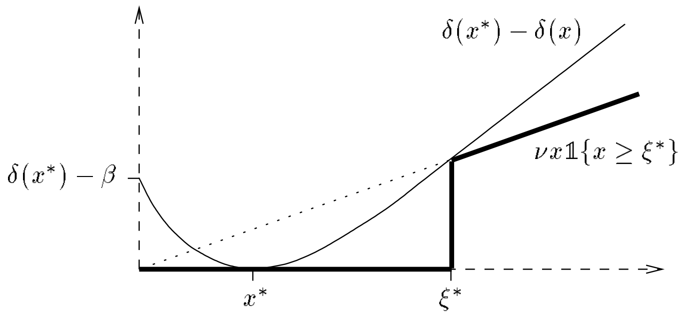


Figure 6: Lower bound.

For any direction x, we have, using (46), $\delta(x^*) - E[\hat{V}_x^{(\eta)}(-k)/k] \geq -\varepsilon$. We deduce that

$$A_x(\eta) \quad \leq \quad P\{\frac{\hat{M}_x(-k)}{k} - E[\frac{\hat{M}_x(-k)}{k}] \geq u - \varepsilon\} \, .$$

We assume from now on that we have $u > \varepsilon > 0$. By application of Azuma's inequality, Lemma 5.11, we obtain

$$A_x(\eta) \leq 2\exp\left(-\frac{(u-\varepsilon)^2 k^c}{2(1+x)^d}\right) \, .$$

We deduce that we have, for a given $X > 0$,

$$\sum_{x \in k^{-1}\mathbb{N}, x \leq X} A_x(\eta) \leq 2kX \exp\left(-\frac{(u-\varepsilon)^2 k^c}{2(1+X)^d}\right) \, . \tag{49}$$

Now, let us consider the directions $x > X$. We assume furthermore that we have $X \geq \xi^*$, where ξ^* is the constant defined in (47). Using (46) and (47), we obtain that

$$\delta(x^*) - E[\frac{\hat{V}_x^{(\eta)}(-k)}{k}] \geq \delta(x^*) - \delta(x) - \eta x \geq \nu x \mathbb{1}\{x \geq \xi^*\} - \eta x = (\nu - \eta)x \, .$$

We deduce that, for $x > X$,

$$\begin{aligned} A_x(\eta) \quad &\leq \quad P\{\frac{\hat{M}_x(-k)}{k} - E[\frac{\hat{M}_x(-k)}{k}] \geq (\nu-\eta)x + u\} \\ &\leq \quad P\{\frac{\hat{M}_x(-k)}{k} - E[\frac{\hat{M}_x(-k)}{k}] \geq (\nu-\eta)x\} \, . \end{aligned}$$

From now on, we assume that we have $\nu > \eta > 0$. By applying Lemma 5.11, we get

$$A_x(\eta) \leq 2\exp\left(-\frac{(\nu-\eta)^2 x^2 k^c}{2(1+x)^d}\right) \, .$$

Using that $1 + x = x(1+1/x) \leq x(1+1/\xi^*)$, and using that $d = 2 - c$, we obtain

$$A_x(\eta) \leq 2\exp\left(-\frac{(\nu-\eta)^2 x^c k^c}{2(1+1/\xi^*)^d}\right) \, .$$

Let $c_1(\eta) = (\nu - \eta)^2/2(1+1/\xi^*)^d$. We obtain

$$\sum_{x \in k^{-1}\mathbb{N}, x>X} A_x(\eta) \leq 2 \sum_{i=[Xk]+1}^{\infty} \exp(\, -c_1(\eta)(i/k)^c k^c \,) = 2 \sum_{i=[Xk]+1}^{\infty} \exp(\, -c_1(\eta) i^c \,)$$

As the function $f(x) = x^c$ is concave, we have $(m+n)^c/2^c \geq m^c/2 + n^c/2$, for $m, n \geq 0$. We deduce that

$$\begin{aligned} \sum_{x \in k^{-1}\mathbb{N}, x>X} A_x(\eta) &\leq 2 \exp\left(-c_1(\eta)\frac{(2Xk)^c}{2}\right) \sum_{i=0}^{\infty} \exp\left(-c_1(\eta)(\frac{(2XK+2i)^c}{2^c} - \frac{(2Xk)^c}{2})\right) \\ &\leq 2 \exp\left(-c_1(\eta)\frac{(2Xk)^c}{2}\right) \sum_{i=0}^{\infty} \exp\left(-c_1(\eta)\frac{(2i)^c}{2}\right) . \end{aligned}$$

Let us set $c_1(\eta) := c_1(\eta)2^{c-1}$ and $c_2(\eta) = 2\sum_{i=0}^{\infty} \exp\left(-c_1(\eta)2^{c-1}i^c\right)$. We have proved that

$$\sum_{x \in k^{-1}\mathbb{N}, x>X} A_x(\eta) \leq c_2(\eta) \exp\left(-c_1(\eta) X^c k^c\right) . \tag{50}$$

Using (49) and (50), we conclude that we have $\sum_{k \in \mathbb{N}^*} P\{\hat{V}^{(\eta)}(-k)/k \geq \delta(x^*) + u\} < \infty$. By the Borel-Cantelli lemma, we obtain that

$$P\left\{ \bigcap_{k \geq K} \{\hat{V}^{(\eta)}(-k)/k < \delta(x^*) + u \right\} \stackrel{K\to\infty}{\longrightarrow} 1 .$$

We have proved (45), hence (42), hence (43). It completes the proof of the a.s. convergence in Theorem 5.3 (for the backward variables).

L_1 convergence

Let us prove that we also have convergence in L_1 in Theorem 5.3. As $V(-k)/k$ converges P-a.s. to $\delta(x^*)$, there is L_1 convergence if and only if the family $\{V(-k)/k, k \in \mathbb{N}^*\}$ is UI (Uniformly Integrable).

Using the same type of inequalities as in (35), we see that $\{V(-k)/k, k \in \mathbb{N}^*\}$ is UI if and only if $\{\hat{V}(-k)/k, k \in \mathbb{N}^*\}$ is UI. According to Assumption (22) in Theorem (5.3), there exists η such that $\beta < \alpha - \eta < \alpha$ and verifying

$$E\,[\,\sup_{n \geq 1}[\,\sum_{i=-n}^{-1} (\alpha - \eta) - t(i)]^+ \,] \,<\, \infty .$$

Using the same decomposition as in (44) for this η, we have

$$\frac{\hat{V}(-k)}{k} \leq \frac{\hat{V}^{(\eta)}(-k)}{k} + \frac{1}{k} \sup_{n \in \mathbb{N}} \sum_{i=-n}^{-1} (\, \alpha - \eta - t(i)\,) . \tag{51}$$

It is immediate that the family $\{\sup_n[\ \sum_{i=-n}^{-1} \alpha - \eta - t(i)\ ]/k,\ k \in \mathbb{N}^*\}$ is UI. Hence, if the family $\{\hat{V}^{(\eta)}(-k)/k, k \in \mathbb{N}^*\}$ is UI, it implies that the family $\{\hat{V}(-k)/k, k \in \mathbb{N}^*\}$ is also UI.

Let us set $\chi = \sup_x\{\gamma(x) - \alpha x + \eta x\}$. Using the first part of the proof, we have $\lim_k \hat{V}^{(\eta)}(-k)/k = \chi$, P-a.s.
The family $\{\hat{V}^{(\eta)}(-k)/k,\ k \in \mathbb{N}^*\}$ is UI if we have

$$\sup_{k\in\mathbb{N}^*} E\ [\ \frac{\hat{V}^{(\eta)}(-k)}{k}\mathbb{1}\{\frac{\hat{V}^{(\eta)}(-k)}{k} > C\}\ ] \stackrel{C\to\infty}{\longrightarrow} 0\,. \tag{52}$$

We have

$$\begin{aligned} E[\frac{\hat{V}^{(\eta)}(-k)}{k}\mathbb{1}\{\frac{\hat{V}^{(\eta)}(-k)}{k} > C\}] &= \int_C^\infty P\{\frac{\hat{V}^{(\eta)}(-k)}{k} > u\}du \\ &= \int_{C-\chi}^\infty P\{\ \frac{\hat{V}^{(\eta)}(-k)}{k} > \chi + u\ \}du\,. \end{aligned}$$

Let us set $\Delta_x^{(\eta)} = \hat{V}^{(\eta)}(-k)/k - E[\hat{V}^{(\eta)}(-k)/k] = M_x(-k)/k - E[M_x(-k)/k]$. Using (46), we obtain for $x \in \mathbb{R}_+$,

$$\{\frac{\hat{V}_x^{(\eta)}(-k)}{k} > \chi + u\} = \{\Delta_x^{(\eta)} > \chi - E[\frac{\hat{V}^{(\eta)}(-k)}{k}] + u\} \subset \{\Delta_x^{(\eta)} > u\}\,.$$

Let us consider ν such that $\eta < \nu < \alpha - \beta$ and let us define ξ^* accordingly, see (47). We set $\tilde{C} = C - \chi$ and we assume that $\tilde{C} > \xi^*$. We obtain, using (49) and (50) (the trick is to decompose the infinite sum in a way which depends on u),

$$\begin{aligned} E[\frac{\hat{V}^{(\eta)}(-k)}{k}\mathbb{1}\{\frac{\hat{V}^{(\eta)}(-k)}{k} > C\}] &\le \int_{\tilde{C}}^\infty \sum_{x\in k^{-1}\mathbb{N}} P\{\Delta_x^{(\eta)} > u\}\ du \\ \le&\ \int_{\tilde{C}}^\infty \sum_{x\le u} P\{\Delta_x^{(\eta)} > u\} + \sum_{x>u} P\{\Delta_x^{(\eta)} > u\}\ du \\ \le&\ \int_{\tilde{C}}^\infty 2ku\exp\left(-\frac{u^2k^c}{2(1+u)^d}\right) + c_2(\eta)\exp(-c_1(\eta)u^ck^c)\ du \\ \le&\ \int_{\tilde{C}}^\infty 2ku\exp\left(-\frac{u^ck^c}{2(1+1/\tilde{C})^d}\right) + c_2(\eta)\exp(-c_1(\eta)u^ck^c)\ du\,. \end{aligned}$$

We deduce from the above expression that (52) holds. It follows that the families $\{\hat{V}^{(\eta)}(-k)/k, k \in \mathbb{N}^*\}$, $\{\hat{V}(-k)/k, k \in \mathbb{N}^*\}$ and $\{V(-k)/k, k \in \mathbb{N}^*\}$ are UI. It completes the proof of the L_1 convergence in Theorem 5.3.

Back to the forward variables

We have just proved the a.s. and L_1 convergence of the backward variables $V(-k)/k$ to the finite constant $\delta(x^*)$. It implies that the forward variables $V(k)/k$ converge in probability and in L_1 to $\delta(x^*)$. The a.s. convergence remains to be proved.
Let $M_x(k), V(k), V_x(k), \hat{V}(k), \hat{V}_x(k), \hat{V}^{(\eta)}(k)$ and $\hat{V}_x^{(\eta)}(k)$ be the forward analogues of the backward variables $M_x(-k), V(-k), V_x(-k), \hat{V}(-k), \hat{V}_x(-k), \hat{V}^{(\eta)}(-k)$ and $\hat{V}_x^{(\eta)}(-k)$. A backward variable and its forward counterpart have the same distribution.

The same decomposition as in (44) can be applied to the forward variables

$$\frac{\hat{V}(k)}{k} \leq \frac{\hat{V}^{(\eta)}(k)}{k} + \frac{1}{k} \sum_{i=-xk}^{-1} (\ \alpha - \eta - t(i)\ )\ .$$

Lemma 5.11 remains valid for the forward variables as it states a result in distribution. We deduce that the bounds in (49) and (50), which were obtained using Lemma 5.11, are valid for the forward variables. We conclude that we have $\lim_k \hat{V}(k)/k = \delta(x^*)$, P-a.s.

Lemma 5.10 is also true for the forward variables. The only point to check is that eqn (32) is true when replacing $\mathcal{L}(-K)$ by $\mathcal{L}(K)$ the set of all oriented paths originating in $(0, K-1)$ and of length K. Any path $\pi \in \mathcal{L}(K)$ can be extended into a non-oriented path originating in $(0,0)$ and of length $2K-1$. Now, Lemma 1 in [18] (the result used to prove eqn (32)) was proved for non-oriented paths, hence can still be applied.
Using the forward version of Lemma 5.10, we obtain (cf (35))

$$\begin{aligned} 0 \ \leq\ \limsup_{k\to+\infty} \frac{1}{k} V(k) \ &-\ \limsup_{k\to+\infty} \frac{1}{k} \hat{V}(k) \\ &\leq\ \limsup_{k\to+\infty} \frac{1}{k} \sup_x (M_x(k) - \hat{M}_x(k)) \ =\ 0\ . \end{aligned}$$

We deduce that we have a.s. convergence of the forward variables $V(k)/k$ to $\delta(x^*)$.

As a side remark, using Lemma 5.10 and 5.11, we also obtain that Theorem 5.7 is valid for the forward variables $M_x(K)$. There is no other obvious way to get the a.s. convergence as the variables $M_x(K)$ are not super-additive.

Example 5.13 (Exponential case). For exponential service times of mean β, we have $\delta(x) = \beta + (\beta - \alpha)x + 2\beta\sqrt{x}$, see eqn (38) and (30). The real number x^* is determined by $\delta'(x^*) = 0$. We obtain $x^* = \beta^2/(\alpha - \beta)^2$, hence

$$\lim_K \frac{1}{K} \sum_{k=0}^{K-1} v(0,k) = \delta(x^*) = \frac{\alpha\beta}{\alpha - \beta} \quad P - a.s. \text{ and in } L_1\ .$$

This is perfectly consistent with the following two results (also given in the case $\alpha = E[t(0,0)] > E[s(0,0)] = \beta$), see Examples 3.1 and 5.2:

1. In the half-plane problem with exponential service times, the departure process $\{D(n,k), n \in \mathbb{Z}\}$ converges weakly to a Poisson process of rate $1/\alpha$.
2. In a $M/M/1/\infty$ queue, the stationary sojourn time $v(0,0)$ verifies $E[v(0,0)] = \alpha\beta/(\alpha - \beta)$.

6 Conclusion

Let us mention some interconnections between the quadrant and the half-plane versions of our model.

We consider i.i.d., non-degenerate, service times of mean $\beta > 0$. We consider the half-plane problem first, and we assume that Conjecture 5.1 is verified for these service times. For $\alpha > \beta$, we denote by P_α the unique invariant ergodic distribution with one-dimensional marginals of mean α (whose existence and uniqueness is precisely ensured by Conjecture 5.1). We set the initial inter-arrival process to be $\{t_\alpha(n,0), n \in \mathbb{Z}\}$, a process distributed as P_α and independent of the services. Then, according to point (iii) in Conjecture 5.1, the random field $\{t_\alpha(n,k), v_\alpha(n,k), n \in \mathbb{Z}, k \in \mathbb{N}\}$ is spatially and temporally ergodic (with obvious notations). It implies in particular that the sojourn time process $\{v_\alpha(0,k), k \in \mathbb{N}\}$ is ergodic. Applying Theorem 5.3, we get that $E[v_\alpha(0,0)] = N(\alpha) = \sup_{x\in\mathbb{R}_+} \gamma(x) - \alpha x > 0$, where $\gamma(x)$ is the function defined in Theorem 5.7-Proposition 5.8.

Let us concentrate on the quadrant version. Let the boundary conditions be given by $\{v_\alpha(0,k), k \in \mathbb{N}\}$ and $\{t_\alpha(n,0), n \in \mathbb{N}\}$, the processes that we just defined. The random field $\{v_\alpha(n,k), t_\alpha(n,k), n \in \mathbb{N}, k \in \mathbb{N}\}$ is spatially and temporally ergodic. This is perfectly coherent with Theorem 4.1 (but much stronger of course), which can be applied here as the boundary processes are ergodic and verify $E[v_\alpha(0,0)] = N(\alpha) > \beta$ and $E[t_\alpha(0,0)] = \alpha > \beta$.

Now, let us consider the following boundary conditions: $\{v(0,k), k \in \mathbb{N}\}$ is ergodic and verifies $E[v(0,0)] = N(\alpha)$, and $\{t(n,0), n \in \mathbb{N}\}$ is arbitrary (a.s. finite). Then, applying Theorem 4.1, we obtain the existence of a process $T = \{t_n, n \in \mathbb{N}\}$ such that $T(k) = \{t(n,k), n \in \mathbb{N}\}$ converges weakly to T (with k). Following the proof of Theorem 4.1, we also have that the processes $\{t_{n,k}, k \geq l\}$ couple in finite time with an ergodic process $\mathcal{T}(n)$ with t_n as its one-dimensional marginal. Applying point (ii) of Conjecture 5.1 to the two-sided extensions of $\mathcal{T}(n)$, we get the following result (as l goes to infinity):

$$\{t_n, n \geq l\} \xrightarrow{w} \{t_\alpha(n,0), n \in \mathbb{N}\} .$$

We prove similarly that $\{v_n, n \geq l\} \xrightarrow{w} \{v_\alpha(n,0), n \in \mathbb{N}\}$ (with obvious notations). Dually, we can start from the following boundary conditions: $\{v(0,k), k \in \mathbb{N}\}$ is arbitrary (a.s. finite) , and $\{t(n,0), n \in \mathbb{N}\}$ is ergodic and verifies $E[t(0,0)] = \alpha$. In this case, we obtain the existence of a limit process $\hat{V} = \{\hat{v}_k, k \in \mathbb{N}\}$ such that $\hat{V}(n) = \{v(n,k), k \in \mathbb{N}\} \xrightarrow{w} \hat{V}$ and such that $\{\hat{v}_k, k \geq l\} \xrightarrow{w} \{v_\alpha(0,k), k \in \mathbb{N}\}$ (as l goes to infinity).

Example 6.1 (Exponential case). Let the services be i.i.d., exponentially distributed with mean β. We consider the half-plane problem with the following boundary conditions. The processes $W(0)$ and $Y(-1)$ are ergodic, independent of the services and verify $E[w(0,0)] = \alpha_1 - \beta$ and $E[y(1,-1)] = \alpha_2 - \beta$ (with $\alpha_1 > \beta$ and $\alpha_2 > \beta$). We apply Theorem 4.1 and we proceed as above to define the limit processes $\{t_n, n \in \mathbb{N}\}$, $\{v_n, n \in \mathbb{N}\}$, $\{\hat{t}_k, k \in \mathbb{N}\}$ and $\{\hat{v}_k, k \in \mathbb{N}\}$. This time, the limit processes are not stationary. Using the results recalled in Example 5.2, we obtain that $\{t_n, n \geq l\}$ and $\{v_n, n \geq l\}$ converge weakly (with l) to the processes defined in (18). Similarly, $\{\hat{y}_k, k \geq l\}$ and $\{\hat{w}_k, k \geq l\}$ converge weakly (with l) to the processes defined in (19).

Further research

While the present article was being prepared, Conjecture 5.1 has been partially solved. The new results use in a crucial way Theorem 5.3. They will be reported in [36, 29].

Acknowledgment

The authors would like to thank Geoffrey Grimmett for pointing out several relevant references and Balaji Prabhakar and James Martin for helpful suggestions. The authors are also very grateful to Ephie Deriche for carrying out a large portion of the typing work.

References

[1] D. Aldous and P. Diaconis. Hammersley's interacting particle process and longest increasing subsequences. *Prob. Theory and Related Fields*, 103:199–213, 1995.

[2] V. Anantharam. Uniqueness of stationary ergodic fixed point for a ./M/K node. *Annals Appl. Prob.*, 3(1):154–172, 1993. Correction published in 4(2):607, 1994.

[3] E. Andjel. Invariant measures for the zero-range process. *Annals of Probability*, 10:525–547, 1982.

[4] F. Baccelli and P. Brémaud. *Elements of Queueing Theory*. Number 26 in Applications of Mathematics. Springer Verlag, Berlin, 1994.

[5] P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, New York, 1968.

[6] A. Borovkov. *Asymptotic Methods in Queueing Theory*. John Wiley & Sons, New York, 1984. Russian edition: Nauka, 1980.

[7] A. Borovkov and S. Foss. Stochastically recursive sequences and their generalizations. *Siberian Adv. in Math.*, 2:16–81, 1992.

[8] P. Burke. The output of a queueing system. *Operations Research*, 4:699–704, 1956.

[9] P. Burke. The dependence of delays in tandem queues. *Ann. Math. Stat.*, 35:874–875, 1964.

[10] C.S. Chang. On the input-output map of a G/G/1 queue. *J. Appl. Prob.*, 31(4):1128–1133, 1994.

[11] J. Cox, A. Gandolfi, P Griffin, and H. Kesten. Greedy lattice animals I: upper bounds. *Annals Appl. Prob.*, 3(4):1151–1169, 1993.

[12] D. Daley and T. Rolski. Finiteness of waiting-time moments in general stationary single-server queues. *Annals Appl. Prob.*, 2(4):987–1008, 1992.

[13] B. Derrida, M. Evans, V. Hakim, and V. Pasquier. Exact solution of a 1d asymmetric exclusion model using a matrix formulation. *J. Phys. A: Math. Gen.*, 26:1493–, 1993.

[14] J.L. Doob. *Stochastic Processes*. Wiley, 1953.

[15] R. Dudley. *Real Analysis and Probability*. Brooks/Cole Pub. Co., New York, 1989.

[16] P. Ferrari and L. Fontes. The net output process of a system with infinitely many queues. *Ann. Applied Prob.*, 4(4), 1994.

[17] W. Fulton. *Young tableaux: With applications to representation theory and geometry.* Cambridge Univ. Press, 1997.

[18] A. Gandolfi and H. Kesten. Greedy lattice animals II: linear growth. *Annals Appl. Prob.*, 4(1):76–107, 1994.

[19] P. Glynn and W. Whitt. Departures from many queues in series. *Annals Appl. Prob.*, 1(4):546–572, 1991.

[20] A. Greenberg, O. Schlunk, and W. Whitt. Using distributed-event parallel simulation to study departures from many queues in series. *Proba. Engineering and Informational Sc.*, 7:159–186, 1993.

[21] J.R. Jackson. Networks of waiting lines. *Oper. Res.*, 5:518–521, 1957.

[22] H. Kesten. On the speed of convergence in first-passage percolation. *Annals Appl. Proba.*, 3(2):296–338, 1993.

[23] J. Kiefer and J. Wolfowitz. On the characteristics of the general queueing process with application to random walks. *Ann. Math. Statist.*, 27:147–161, 1956.

[24] J. Kingman. Subadditive ergodic theory. *Annals of Probability*, 1:883–909, 1973.

[25] C. Kipnis. Central limit theorems for infinite series of queues and applications to simple exclusion. *Annals of Probability*, 14:397–408, 1986.

[26] T. Liggett. *Interacting Particle Systems.* Springer Verlag, Berlin, 1985.

[27] T. Liggett and T. Shiga. Note on departure processes in a G/M queue. Working paper, 1994.

[28] R. Loynes. The stability of a queue with non-independent interarrival and service times. *Proc. Camb. Philos. Soc.*, 58:497–520, 1962.

[29] J. Mairesse and B. Prabhakar. On the existence of fixed points for the ./GI/1 queue. Technical report, 1999. In preparation.

[30] C. McDiarmid. On the method of bounded differences. In J. Siemons, editor, *Surveys in Combinatorics*, number 141 in London Math. Soc. Lect. Note Series, pages 148–188. Camb. Univ. Press, 1989.

[31] S. Meyn and R. Tweedie. *Markov Chains and Stochastic Stability.* Springer-Verlag, Berlin, 1993.

[32] M. Miyazawa. A formal approach to queueing processes in the steady state and their applications. *J. Appl. Prob.*, 16:332–346, 1979.

[33] T. Mountford and B. Prabhakar. On the weak convergence of departures from an infinite sequence of ./M/1 queues. *Annals Appl. Prob.*, 5(1):121–127, 1995.

[34] T. Mountford and B. Prabhakar. The cesaro limit of departures from certain ./GI/1 queueing tandems. In F. Kelly, S. Zacharie, and I. Ziedins, editors, *Stochastic Networks: Theory and Applications*, pages 309–322. Oxford Univ. Press, 1996.

[35] E. Muth. The reversibility property of production lines. *Management science*, 25(2):152–158, 1979.

[36] B. Prabhakar. On the attractiveness of the fixed points of a ./GI/1 queueing operator. Technical report, 1999. In preparation.

[37] E. Reich. Waiting times when queues are in tandem. *Ann. Math. Stat.*, 28:527–530, 1957.

[38] E. Reich. Note on queues in tandem. *Ann. Math. Stat.*, 34:338–341, 1963.

[39] H. Rost. Non-equilibrium behavior of a many particle process: density profile and local equilibria. *Z. Warsch. Verw. Gebiete*, 58:41–53, 1981.

[40] T. Seppäläinen. Hydrodynamic scaling, convex duality, and asymptotic shapes of growth models. *Markov Process. Related Fields*, 4:1–26, 1998.

[41] R. Srinivasan. Stochastic comparisons of density profiles for the road-hog process. *J. Appl. Prob.*, 28:852–863, 1991.

[42] R. Srinivasan. Queues in series via interacting particle systems. *Math. Oper. Res.*, 18(1):39–50, 1993.

[43] W. Szczotka. Stationary representation of queues. I. *Adv. Appl. Probability*, 18:815–848, 1986.

[44] W. Szczotka and F. Kelly. Asymptotic stationarity of queues in series and the heavy traffic approximation. *Annals of Probability*, 18:1232–1248, 1990.

[45] S. Tembe and R. Wolff. The optimal order of service in tandem queues. *Oper. Res.*, 24:824–832, 1974.

[46] A. Vershik. Asymptotic combinatorics and algebraic analysis. In Birkhäuser, editor, *Proceedings of the Int. Congress of Mathematicians, Zürich*, 1995.