*Adv. Appl. Prob.* **38,** 487–504 (2006)
*Printed in Northern Ireland*


# STABILITY OF SPATIAL QUEUEING SYSTEMS

C. BORDENAVE,* *Ecole Normale Supérieure and INRIA*

**Abstract**

In this paper, we analyze a queueing system characterized by a space–time arrival process of customers served by a countable set of servers. Customers arrive at points in space and the server stations have space-dependent processing rates. The workload is seen as a Radon measure and the server stations can adapt their power allocation to the current workload. We derive the stability region of the queueing system in the usual stationary ergodic framework. The analysis of this stability region gives some counter-intuitive results. Some specific subclasses of policy are also studied. Wireless communications networks is a natural field of application for the model.

*Keywords:* Stability region; space–time point process; wireless network; dynamic scheduling; optimal resource allocation

2000 Mathematics Subject Classification: Primary 60K25; 60K30; 90B15
Secondary 60G55

## 1. Introduction and model description

In this paper, we analyze a space–time arrival process of customers served by a countable set of servers. This model is motivated by large-scale wireless communications networks, but could suit other types of infinite queueing systems. The model follows along the lines of other optimal allocation problems studied previously. In seminal work, Tassiulas and Ephremides [14], [15], have examined the stability region of a multihop network with a general topology and set of constraints. Their analysis was performed in a Markovian setting. The stability region is defined as the set of traffic loads such that there exists a scheduling policy under which the system is stable (positive recurrent). A series of papers, [16], [1], [4], [17], and [5], has extended in two directions the framework of Tassiulas and Ephremides; first considering stationary ergodic traffic flows, and then introducing randomness in the network topology. In this paper, we examine a spatial model and take interest in the geometrical properties of the stability region. Applications of this work to wireless networks were presented in [7].

We consider a system in which some jobs arrive exogenously and the jobs are located in the space. Some server stations serve the incoming jobs. Each server station can process the jobs at a rate depending on the position of the job and a random environment variable. We derive some results on the stability of this queueing system. An important aspect of the model is that the workload is an atomic measure with a possibly infinite total mass, and not a vector in $\mathbb{R}_+^n$. This aspect of the model is well suited to large spatial queueing systems and could be used successfully in other spatial models. Another feature is the possibly infinite number of server stations in the system. Our results are all proved in the stationary ergodic framework and

Received 19 January 2005; revision received 6 February 2006.
* Postal address: DI/TREC, Ecole Normale Supérieure, 45 rue d'Ulm, F-75230 Paris Cedex 05, France.
Email address: charles.bordenave@ens.fr
Presented at the ICMS Workshop on Spatial Stochastic Modelling with Applications to Communications Networks (Edinburgh, June 2004).

independency is never required. The proofs of stability are based on generalizations of Loynes' sequences to general metric spaces. Some ideas from optimization theory are also used in the analysis of the stability region, and pave the way to other developments.

The rest of this section describes the model under consideration. Section 2 is dedicated to the stability analysis of our model. In Section 3, we consider a subclass of policy which preserves a kind of monotonicity for the workload measure. In Section 4 we study the stability region when the spatial intensity measure is absolutely continuous with respect to the Lebesgue measure.

### 1.1. Customer arrival point process

All the random variables we are going to introduce in this section are defined on a common probability space $(\Omega, \mathcal{T}, \mathrm{P})$. This space is endowed with a measurable flow $\{\theta_t\}$, $t \in \mathbb{R}$. We suppose that $(\mathrm{P}, \theta_t)$ is *ergodic*.

In our system, the customers (or jobs) are seen as points of a marked point process $A$, a spatial marked point process on $\mathbb{R} \times \mathbb{R}^d$ with marks on $\mathbb{R}_+$. We use the notation $A = \sum_n \delta_{\{T_n, X_n, \sigma_n\}}$ to represent the point process. The $n$th job arrives at time $T_n \in \mathbb{R}$ at $X_n \in \mathbb{R}^d$ and requires a service time of $\sigma_n \in \mathbb{R}_+$. We suppose that $A$ is compatible with the flow $\theta$; that is, if $A(\omega) = \sum_n \delta_{\{T_n, X_n, \sigma_n\}}$ then $A \circ \theta_t(\omega) = \sum_n \delta_{\{T_n - t, X_n, \sigma_n\}}$, where $\delta_{\{t,x,s\}}$ is the usual Dirac measure with a unit mass at $\{t, x, s\}$. For a bounded Borel set $B \subset \mathbb{R}^{d+1}$,

$$\mathrm{E}(A(B)) = \mathrm{E}\bigg(\sum_n \mathbf{1}((T_n, X_n) \in B)\bigg)$$

is supposed to be finite. Thus, the intensity of $A$ is a *Radon measure* of the form $\lambda(\mathrm{d}x)\,\mathrm{d}t$ (see, for example, [10, Lemma A2.7.11, p. 634]). Note that $\lambda$ is not necessarily a finite measure on $\mathbb{R}^d$ and that between times $t$ and $t' > t$ there may arrive an infinite number of jobs.

Define $\mathrm{P}_A^{t,x}(\cdot)$ to be the Palm probability of the point process $A$ at $(t, x) \in \mathbb{R}^{d+1}$ (see [10, Chapter 12]). Since $A$ is compatible with $\theta_t$, we have $\mathrm{P}_A^{t,x}(\cdot) = \mathrm{P}_A^{0,x}(\theta_{-t} \times \cdot)$. Under $\mathrm{P}_A^{t,x}$, let $\sigma_{t,x}$ be the required service time of a customer arriving at time $t$ at position $x$. Denote by $\mathrm{E}^{t,x}$ the expectation with respect to the probability $\mathrm{P}^{t,x}$. We will suppose that $0 < \mathrm{E}^{0,x}(\sigma_{0,x}) < \infty$, $\lambda(\mathrm{d}x)$-almost everywhere ($\lambda(\mathrm{d}x)$-a.e.). Recall that $\mathrm{E}^{0,x}(\sigma_{0,x})$ can be understood as the mean number of service time requirements of a typical customer arriving at $x$.

### 1.2. Server station adaptative policy

We have a *countable* set of server stations, denoted by $\mathcal{J}$. The servers can provide service to all points in space at different processing rates: the server $j$ serves a customer located at $x$ at rate $r_j(x)$. We suppose that $x \mapsto r_j(x)$ is a positive, measurable function and that

$$\lim_{|x| \to \infty} r_j(x) = 0, \tag{1}$$

where $|\cdot|$ denotes the Euclidean norm. The server stations are in a random environment and their processing powers vary over time. At time $t$, the total processing power available to server $j$ is $\varepsilon_j(t) \in \mathbb{R}_+$. We suppose that the driving process $\varepsilon_j(t)$ is compatible with the shift, that is, $\varepsilon_j(t) = \varepsilon_j(0) \circ \theta_t$ and

$$p_j := \mathrm{E}(\varepsilon_j(0)) < \infty.$$

The workload at time $t$ is the set of all the jobs waiting to be processed. It is denoted by $W_t$ and is an atomic measure on $\mathbb{R}^d$, with $\int_B W_t(\mathrm{d}x)$ representing the total remaining service time at time $t$ of all customers located in $B$.

https://doi.org/10.1239/aap/1151337081 Published online by Cambridge University Press

The server stations divide their processing power between the required jobs according to a policy scheme. This power allocation depends on the current workload. We suppose that our queueing system cannot handle an infinite amount of service in finite time at a given location $x$: more precisely,

$$\sum_j p_j \sup_{x\in B} r_j(x) < \infty \quad \text{for all bounded Borel sets } B. \tag{2}$$

**Definition 1.** Let $\mathcal{M}$ be the set of Radon measures on $\mathbb{R}^d$ endowed with the vague topology (see [10, pp. 615–631]). A *policy* $\pi = (\pi_j)_{j\in\mathcal{J}}$ is a measurable mapping from $\mathcal{M}\times\Omega$ to $\mathcal{M}^{\mathcal{J}}$ such that

$$\int_{\mathbb{R}^d} \pi_j(m,\omega)(\mathrm{d}x) \le \varepsilon_j(0)(\omega) \tag{3}$$

and

$$\pi_j(m,\omega) \text{ is } \textit{absolutely continuous} \text{ with respect to } m, \tag{4}$$

for all $\omega\in\Omega$.

The policy enforced at time $t$ is $\pi(W_t, \theta_t\omega)$.

Equation (3) implies that the server stations cannot allocate more than their total processing power. If the total workload on a Borel set is 0, it is useless to dedicate some processing power to this set. At time $t$ the server station $j$ achieves an instantaneous service rate of $r_j(x)\pi_j(W_t)(\{x\})$ for a job located at $x$. (To simplify our notation, $\pi_j(W_t,\theta_t)$ be simply be written as $\pi_j(W_t)$.)

The policies we are considering in our model are stationary, and if $W$ is stationary then $\pi(W_t,\theta_t\omega)$ is also stationary. A study of nonstationary policies was made in [14] in another framework. In that article the authors showed that the nonstationary and stationary policies have the same performances (as far as stability is concerned).

In the next sections, we will define some interesting classes of policy. A simple example of a deterministic policy defined for atomic measures is

$$\pi_j^+(m) = \begin{cases} 0 & \text{if } m \text{ is the zero measure,} \\ \varepsilon_j\delta_{x_j^+} & \text{otherwise,} \end{cases}$$

where $x_j^+ = \operatorname{argmax}\{x : r_j(x)\,\mathbf{1}(m(\{x\})>0)\}$. If multiple choices of $x$ are possible, we choose the first in lexicographical order. With this policy, the server stations serve the user with the greatest processing rate first. Note, in particular, that this policy is *work conserving*: if $W_t$ is not the null measure, the server is active.

## 1.3. Evolution equation

The dynamics of our queueing system is given by the following integral equation, for all Borel sets $B$ and $t' > t$:

$$W_{t'}(B) = W_t(B) + \int_t^{t'}\int_B \sigma_{s,x} A(\mathrm{d}s\times\mathrm{d}x) - \sum_j \int_t^{t'}\int_B r_j(x)\pi_j(W_s)(\mathrm{d}x)\,\mathrm{d}s. \tag{5}$$

In Borovkov's terminology (see [9, Chapter 4]), (5) defines a *stochastic recursive process*: we can write $W_{t+h} = f_h(W_t, A_t^{t+h})$, where $A_t^{t+h}$ denotes the trajectory of the arrival point process $A$ between $t$ and $t+h$, and, for all $h$, $f_h$ is a suitably chosen functional.

Under a given policy $\pi$, for a Radon measure $m$ define $W_t^m$ as the workload at time $t$ when $W_0 = m$.

https://doi.org/10.1239/aap/1151337081 Published online by Cambridge University Press

### 1.4. Some examples

Here are several examples which illustrate the model.

**Example 1.** All the jobs arrive on a countable set of points, $\{x_i,\ i \in \mathbb{N}\}$, with no accumulation points. These points are waiting rooms and the arrival intensity at each point is $\lambda_i = \mathrm{E}(A([0,1] \times \{x_i\})) < \infty$. In this example, the system reduces to multiclass job traffic with processing rates depending on the class and the server station. Stability results on this type of system were presented in [16].

**Example 2.** In a wireless communications scenario with $d = 2$, the server stations are base antennae and the customers are mobile users who want to receive data from the network. Server station $j$ is located at $Y_j \in \mathbb{R}^2$. The processing rate can be written $r_j(x) = L(x, Y_j)/I(x)$, where $L(x, y)$ is an attenuation function on the channel between $x$ and $y$ and $I(x) = \sum_{j \in \mathcal{J}} L(x, Y_j)$ is a shot noise process. A natural assumption is to assume that $L(x, y)$ depends only on $|y - x|$. Motivations for this model were given in [7].

**Example 3.** When $\varepsilon_j \in \{0, 1\}$, the model exhibits random connectivity. The server stations are either switched on or switched off. At time $t$, if $\varepsilon_j(t)$ equals 1 or 0 then the base station $j$ is switched on or, respectively, off (see also [17]).

## 2. Stability analysis

### 2.1. Stability region

In this paper we are interested in the stability of the queueing system described above. A policy $\pi$ is *stable* if there exists a finite stationary workload $\{M_t = M \circ \theta_t\}$, $t \in \mathbb{R}$, for an atomic random measure $M$ satisfying (5). The queueing system is said to be stable if there exists a stable policy. This definition was called *stochastic stability* in [4]. In Subsection 2.4, we will consider a stronger definition of stability.

The parameter of the queueing system is the marked point (arrival) process $A$. Let $\mathcal{N}$ be the set of point processes of finite intensity and with a service time with finite expectation under the Palm measure. We define

$$\mathcal{F} = \Big\{ f = (x \mapsto f_j(x),\ j \in \mathcal{J})\colon \text{ for all } j, \\ f_j \text{ is nonnegative and measurable and } \sum_j f_j(x) = 1\ \lambda(\mathrm{d}x)\text{-a.e.} \Big\},$$

$$\mathcal{N}^s = \bigg\{ A \in \mathcal{N} : \text{there exists an } f \in \mathcal{F} \text{ such that, for all } j, \int_{\mathbb{R}^d} \frac{\mathrm{E}_A^{0,x}(\sigma_{0,x}) f_j(x)}{r_j(x)}\, \lambda(\mathrm{d}x) < p_j \bigg\},$$

$$\bar{\mathcal{N}}^s = \bigg\{ A \in \mathcal{N} : \text{there exists an } f \in \mathcal{F} \text{ such that, for all } j, \int_{\mathbb{R}^d} \frac{\mathrm{E}_A^{0,x}(\sigma_{0,x}) f_j(x)}{r_j(x)}\, \lambda(\mathrm{d}x) \le p_j \bigg\}.$$

These sets are generalized continuous versions of the stability sets derived in [16].

Note that

$$\int_{\mathbb{R}^d} [\mathrm{E}_A^{0,x}(\sigma_{0,x}) f_j(x)/r_j(x)] \lambda(\mathrm{d}x)$$

is a traffic load: $\mathrm{E}_A^{0,x}(\sigma_{0,x})\lambda(\mathrm{d}x)$ is the mean number of service requirements per unit of surface and $r_j(x)$ is the processing rate at $x$ for server station $j$. We can now state the following stability theorem.

**Theorem 1.** *For the queueing model described above,*

- *if $A \in \mathcal{N}^s$ then there exists a stable policy, and*
- *if there is a stable policy then $A \in \bar{\mathcal{N}}^s$.*

Note that, as in Loynes' theorem, the stability region depends on the distribution of the point process $A$ only through its means. To be precise, suppose that there is only one server station and that all jobs arrive at the same place, say 0. The stability region is then characterized by the condition $\mathrm{E}_A^{0,0}(\sigma_{0,0})\lambda(0) < p_1 r_1(0)$. This result is the usual condition, namely $\rho < 1$, for G/G/1 queues.

In the proof of the theorem, we establish that to a given stable policy $\pi$ there corresponds an $(f_j)_{j\in\mathcal{J}}$ in $\mathcal{F}$ such that

$$\int_{\mathbb{R}^d} \frac{\mathrm{E}_A^{0,x}(\sigma_{0,x}) f_j(x)}{r_j(x)} \lambda(\mathrm{d}x) \le p_j.$$

We interpret $f_j$ as the proportion of service carried by the server $j$ for customers at $x$ in the stationary regime.

The converse mapping is also available: for a set of functions $(f_j)_{j\in\mathcal{J}}$ in $\mathcal{F}$ such that the above inequality is satisfied with strict inequality, there exists a stable policy. In fact, as will be seen in the proof of Theorem 1, the last assertion will only be proved for a dense subset of functions in $\mathcal{F}$. The weakness of this theorem is the lack of precision about the policy which achieves the maximum permissible loading. The policy we construct in the proof of the theorem is not practical.

### 2.2. Necessary conditions for stability in Theorem 1

This technical lemma is needed in what follows (see, for example, [2, Lemma 2.2.1, p. 87] for a proof).

**Lemma 1.** *Let $Z$ be a nonnegative, almost surely (a.s.) finite random variable such that, for a given $t$, if* $\mathrm{E}(|Z - Z \circ \theta_t|) < \infty$ *then* $\mathrm{E}(Z - Z \circ \theta_t) = 0$.

Suppose that there exists a stable policy $\pi$ and an a.s. finite stationary workload measure $W_t = W \circ \theta_t$. Let $B$ be a bounded Borel set. From (5), we have

$$W \circ \theta_t(B) = W(B) + \int_0^t \int_B \sigma_{s,x} A(\mathrm{d}s \times \mathrm{d}x) - \sum_j \int_t^{t+h} \int_B r_j(x)\pi_j(W \circ \theta_s)(\mathrm{d}x)\,\mathrm{d}s.$$

The Campbell formula for marked point processes implies that

$$\mathrm{E}\biggl(\int_0^t \int_B \sigma_{s,x} A(\mathrm{d}s \times \mathrm{d}x)\biggr) = t \int_B \mathrm{E}_A^{0,x}(\sigma_{0,x})\lambda(\mathrm{d}x).$$

We define $\bar{\pi}_j(B) = \mathrm{E}(\pi_j(W)(B))$, which is a Radon measure, and obtain $\int_{\mathbb{R}^d} \bar{\pi}_j(\mathrm{d}x) \le p_j$ from (3). Using condition (2), we deduce that

$$\mathrm{E}(|W \circ \theta_t(B) - W(B)|) \le t \int_B \mathrm{E}_A^{0,x}(\sigma_{0,x})\lambda(\mathrm{d}x) + t \sum_j p_j \sup_{x\in B} r_j(x) < \infty.$$

https://doi.org/10.1239/aap/1151337081 Published online by Cambridge University Press

Thus, $W \circ \theta_t(B) - W(B) \in L^1(P)$ and we can apply Lemma 1 to conclude that

$$0 = \int_B \mathrm{E}_A^{0,x}(\sigma_{0,x})\lambda(\mathrm{d}x) - \sum_j \int_B r_j(x)\bar{\pi}_j(\mathrm{d}x). \tag{6}$$

As (6) holds for all bounded Borel sets, the measures $\sum_j r_j(x)\bar{\pi}_j(\mathrm{d}x)$ and $\mathrm{E}_A^{0,x}(\sigma_{0,x})\lambda(\mathrm{d}x)$ are equal. In particular, the measure $r_j(x)\bar{\pi}_j(\mathrm{d}x)$ is absolutely continuous with respect to $\mathrm{E}_A^{0,x}(\sigma_{0,x})\lambda(\mathrm{d}x)$. Let $\tilde{\pi}_j(x)$ be the Radon–Nikodým derivative of $\bar{\pi}_j(\mathrm{d}x)$ with respect to $\lambda(\mathrm{d}x)$, and let $f_j(x) = r_j(x)\tilde{\pi}_j(x)/\,\mathrm{E}_A^{0,x}(\sigma_{0,x})$. We deduce from (6) that, for all Borel sets $B$,

$$\int_B \mathrm{E}_A^{0,x}(\sigma_{0,x})\lambda(\mathrm{d}x) = \sum_j \int_B f_j(x)\,\mathrm{E}_A^{0,x}(\sigma_{0,x})\lambda(\mathrm{d}x).$$

Thus, $\sum_j f_j(x) = 1$ and $f \in \mathcal{F}$ $\lambda(\mathrm{d}x)$-a.e. Finally, $\int_{\mathbb{R}^d} \bar{\pi}_j(\mathrm{d}x) \le p_j$ reads

$$\int_{\mathbb{R}^d} \frac{\mathrm{E}_A^{0,x}(\sigma_{0,x}) f_j(x)}{r_j(x)}\,\lambda(\mathrm{d}x) \le p_j,$$

and the second assertion of Theorem 1 follows.

### 2.3. Sufficient conditions for stability in Theorem 1

Suppose that $A \in \mathcal{N}^s$. There exist $(f_j)$, $j \in \mathcal{J}$, in $\mathcal{F}$ such that, for all $j$,

$$\int_{\mathbb{R}^d} [\mathrm{E}_A^{0,x}(\sigma_{0,x}) f_j(x)/r_j(x)]\lambda(\mathrm{d}x) < 1.$$

We can suppose that $f$ has the properties given in Proposition 8 below. For a given policy $\pi$ and an initial atomic workload $m$, we define the set

$$\mathcal{A}_j^m(t) = \left\{x : r_j(x)\int_0^t \pi_j(W_s^m)(\{x\})\,\mathrm{d}s \le f_j(x)\left(m(\{x\}) + \int_{[0,t)\times\{x\}} \sigma_{s,x} A(\mathrm{d}s \times \mathrm{d}x)\right)\right\}.$$

That $x$ is in $\mathcal{A}_j^m(t)$ means that server $j$ has contributed less than $f_j(x)$ in fulfilling the service requirements of customers located in $x$.

Consider the following nonstationary policy. For all $j \in \mathcal{J}$ and for $t \ge 0$,

$$\pi_j(t) = \begin{cases} \varepsilon_j(t)\delta_{x_j^*} & \text{if } \mathcal{A}_j^{W_0}(t) \ne \varnothing, \\ 0 & \text{if } \mathcal{A}_j^{W_0}(t) = \varnothing, \end{cases} \tag{7}$$

where $x_j^* = \operatorname{argmax}\{x : \mathbf{1}(x \in \mathcal{A}_j^{W_0}(t))\,\mathbf{1}(W_t(\{x\}) > 0) r_j(x)\}$. If multiple choices of $x_j^*$ are possible, choose the first in lexicographical order.

The existence of this policy follows from condition (1). The policy $\pi$ divides the workload among all servers in proportion to the $f_j$, and processes the jobs at the fastest available rate. By $\pi_j(t)$ we denote the policy enforced at time $t$ when at time 0 the workload is equal to $m$: in the usual notation, $\pi_j(t) = \pi_j(W_t^m)$. Note that if $\mathcal{A}_j^m(T) = \varnothing$ for some $T$, then after time $T$ the server $j$ will only serve jobs that arrive after time $T$, and that, for $t > 0$,

$$\mathcal{A}_j^m(t+T) = \mathcal{A}_j^0(t) \circ \theta_T. \tag{8}$$

Let $B$ be a bounded Borel set. Then

$$\begin{aligned} W_t(B) &= \int_0^t \int_B \sigma_{s,x} A(\mathrm{d}s \times \mathrm{d}x) - \sum_j \int_0^t \int_B r_j(x)\pi_j(s)(\mathrm{d}x)\,\mathrm{d}s \\ &= \sum_j \int_0^t \int_B f_j(x)\sigma_{s,x} A(\mathrm{d}s \times \mathrm{d}x) - \sum_j \int_0^t \int_B r_j(x)\pi_j(s)(\mathrm{d}x)\,\mathrm{d}s \\ &= \sum_j \int_B r_j(x)\tilde{W}_t^j(\mathrm{d}x), \end{aligned}$$

where

$$\tilde{W}_t^j = \int_0^t \int_B \frac{f_j(x)\sigma_{s,x}}{r_j(x)} A(\mathrm{d}s \times \mathrm{d}x) - \int_0^t \int_B \pi_j(s)(\mathrm{d}x)\,\mathrm{d}s. \tag{9}$$

We now define a total order '$\prec_j$' on $\mathbb{R}^d$, such that $x \succ_j y$ if $r_j(x) > r_j(y)$ or $r_j(x) = r_j(y)$ and $x$ is smaller than $y$ in lexicographical order. A Borel set $B$ is a *$j$-max set* if $B = \{x : x \succeq_j y$ for all $y \in B\}$.

**Lemma 2.** *For a $j$-max set $B$,*

$$\tilde{W}_t^j(B) = \Bigg(\max\Bigg(\int_B \frac{f_j(x)W_0(\mathrm{d}x)}{r_j(x)} + \int_0^t \int_B \frac{f_j(x)\sigma_{s,x}}{r_j(x)} A(\mathrm{d}s \times \mathrm{d}x) - \int_0^t \varepsilon_j(s)\,\mathrm{d}s, \\ \sup_{0\le h\le t} \int_0^h \int_B \frac{f_j(x)\sigma_{s,x}}{r_j(x)} A(\mathrm{d}s \times \mathrm{d}x) - \int_0^h \varepsilon_j(s)\,\mathrm{d}s\Bigg)\Bigg)^+.$$

*Proof.* The policy $\pi$ divides the global queueing system with the server set $\mathcal{J}$ into distinct queues, one for each server in $\mathcal{J}$ with an incoming workload equal to

$$\sum_j \int_0^t \int_B f_j(x)\sigma_{s,x} A(\mathrm{d}s \times \mathrm{d}x).$$

$\tilde{W}_t^j(B)$ is the rescaled workload, such that the server $j$ serves the user at $x$ with a unit processing rate. If $B$ is a $j$-max set then, from the definition of policy $\pi_j$, the server $j$ dedicates all of its processing power to $B$ if the workload in $B$ is not 0. The customers in $\mathbb{R}^d \setminus B$ are served if there is no customer in $B$. Thus, the statement of the lemma follows from the usual formula for the G/G/1 queue.

Let $M_t = W_t^0 \circ \theta_{-t}$ be the Loynes sequence under policy $\pi$, and let $\tilde{M}_t^j = \tilde{W}_t^0 \circ \theta_{-t}$. We have

$$M_t(B) = \sum_j \int_B r_j(x)\tilde{M}_t^j(\mathrm{d}x).$$

Note from Lemma 2 that $\tilde{M}_t^j(B)$ is a nondecreasing sequence for a $j$-max set.

**Lemma 3.** *As $t$ tends to $\infty$, $\tilde{M}_t^j$ couples a.s. with $\tilde{M}_\infty^j$, a finite random measure.*

*Proof.* Let $B$ be a $j$-max set. Then, from Birkhoff's theorem,

$$\lim_{t\to\infty} \frac{1}{t} \int_{-t}^0 \int_B \frac{f_j(x)\sigma_{s,x}}{r_j(x)} A(\mathrm{d}s \times \mathrm{d}x) = \int_B \frac{f_j(x)\,\mathrm{E}_A^{0,x}(\sigma_{0,x})}{r_j(x)} \lambda(\mathrm{d}x) < 1 \quad \text{a.s.}$$

https://doi.org/10.1239/aap/1151337081 Published online by Cambridge University Press

Therefore,

$$\tilde{M}_t^j(B) = \left(\sup_{0\le h\le t} \int_{-h}^{0}\int_B \frac{f_j(x)\sigma_{s,x}}{r_j(x)} A(\mathrm{d}s\times \mathrm{d}x) - h\right)^+$$

is a bounded increasing sequence and it couples with

$$\tilde{M}_\infty^j(B) = \left(\sup_{h\in\mathbb{R}_+} \int_{-h}^{0}\int_B \frac{f_j(x)\sigma_{s,x}}{r_j(x)} A(\mathrm{d}s\times \mathrm{d}x) - h\right)^+.$$

Since $\mathbb{R}^d$ is a $j$-max set and $\tilde{M}_t^j(B) \le \tilde{M}_t^j(\mathbb{R}^d) \le \tilde{M}_\infty^j(\mathbb{R}^d)$ for any set $B$, a subsequence $\tilde{M}_{t_n}^j(B)$ converges to $\tilde{M}_\infty^j(B)$ and $\tilde{M}_\infty^j$ is well defined and a finite random measure.

Since $\tilde{W}_t^j(\mathbb{R}^d)$ is a G/G/1 queue, we can define $T$, the first time after the coupling time with $\tilde{M}_\infty^j \circ \theta_t(\mathbb{R}^d)$ such that $\tilde{W}_t^j(\mathbb{R}^d) = 0$. This time $T$ is a.s. finite. In view of (8), $\tilde{W}_t^j = \tilde{M}_\infty^j \circ \theta_t$ for $t \ge T$. Thus, $\tilde{M}_t^j = \tilde{W}_t^j \circ \theta_{-t} = \tilde{M}_\infty^j$ and $\tilde{M}_t^j$ couples with $\tilde{M}_\infty^j$.

Let $B$ be a bounded set. For a choice of $f$ given by Proposition 8,

$$M_t(B) = \sum_{j\in\mathcal{J}_B} \int_B r_j(x)\tilde{M}_t^j(\mathrm{d}x).$$

Since $|\mathcal{J}_B|$ is finite, we deduce from Lemma 3 that $M_t(B)$ couples a.s. with $M_\infty(B) = \sum_{j\in\mathcal{J}_B}\int_B r_j(x)\tilde{M}_\infty^j(\mathrm{d}x)$. We have thus proved the existence of the limit: $\lim_{t\to\infty} M_t = M_\infty$. To conclude the proof, it remains to prove that $M_\infty$ is a stationary solution to (5) for a policy $\pi'$.

Along the lines of the proof of Lemma 3 we can also prove that the process $\{\tilde{M}_{t+s}^j \circ\theta_s\}_{0\le s\le h}$ couples with $\{\tilde{M}_\infty^j \circ \theta_s\}_{0\le s\le h}$ for any positive $h$. From (9), for a bounded set $B$ we have

$$\begin{aligned}\tilde{M}_{t+h}^j \circ\theta_h(B) &= \int_{-t-h}^{0}\int_B \frac{f_j(x)\sigma_{s,x}}{r_j(x)} A(\mathrm{d}s\times\mathrm{d}x) - \int_0^{t+h}\int_B \pi_j(\theta_{-t}, s)(\mathrm{d}x)\,\mathrm{d}s \\ &= \tilde{M}_t^j(B) + \int_0^h\int_B \frac{f_j(x)\sigma_{s,x}}{r_j(x)} A(\mathrm{d}s\times\mathrm{d}x) - \int_0^h\int_B \pi_j(\theta_{-t}, t+s)(\mathrm{d}x)\,\mathrm{d}s.\end{aligned}$$

If $Z$ is the maximum coupling time of $\{\tilde{M}_{t+s}^j \circ\theta_s(B)\}_{0\le s\le h}$, then for $t \ge Z$ we have

$$\int_0^h\int_B \pi_j(\theta_{-t}, t+s)(\mathrm{d}x)\,\mathrm{d}s = \tilde{M}_\infty^j\circ\theta_h(B) - \tilde{M}_\infty^j(B) + \int_0^h\int_B \frac{f_j(x)\sigma_{s,x}}{r_j(x)} A(\mathrm{d}s\times\mathrm{d}x).$$

For all $h$, the right-hand side of this equation does not depend on $t$; thus, $\pi_j(\theta_{-t}, t+s)$ couples with a measure $\pi_j'(s) = \lim_t \pi_j(\theta_{-t}, t+s)$ a.e. Let $t_0$ be such that the coupling occurs; then

$$\pi_j'(s) = \lim_t \pi_j(\theta_{-t+s-t_0}\circ\theta_s, t+t_0-s+s) = \pi_j'(t_0)\circ\theta_{s-t_0}.$$

In consequence, $\pi'$ is a stationary policy and $M_\infty$ is a stationary solution to (5). Theorem 1 is thus proved.

**Remark 1.** In the particular case described in Example 1, a simpler proof is available. We have

$$\sum_i \frac{\lambda_i E_A^{0,x_i}(\sigma_{0,x_i}) f_j(x_i)}{r_j(x_i)} = \rho_j < 1.$$

Consider the following deterministic policy defined for an atomic measure with atoms in $\{x_i,\ i \in \mathbb{N}\}$:

$$\pi_j(m)(\{x_i\}) = \varepsilon_j\, \mathbf{1}(m(\{x_i\}) \neq 0) \frac{f_j(x_i)\lambda_i\, \mathrm{E}_A^{0,x_i}(\sigma_{0,x_i})}{r_j(x_i)\rho_j}.$$

In computing $M_t(\{x_i\})$, it appears that this policy is stable.

### 2.4. Convergence toward a stationary solution

When there exists a stationary regime in a queueing system it is important to know if for any initial condition the workload converges in some sense to the stationary regime. The following proposition gives a positive answer to this for the policy defined in the proof of Theorem 1.

**Proposition 1.** *If the policy scheme defined by (7) is enforced, then for any finite initial workload at time $t = 0$ and for all bounded Borel sets $B$, $\{W_{t+T}(B)\}$, $t \in \mathbb{R}_+$, converges in variation to $\{M \circ \theta_t(B)\}$, $t \in \mathbb{R}_+$, as $T$ tends to $\infty$.*

Note that the workload measure does not converge in variation; rather, convergence happens only on bounded sets. The proposition states that the workload converges in variation in the vague topology.

*Proof of Proposition 1.* The proof relies on the following fact: if a stochastic process $\{X_t\}$ couples with $\{Y \circ \theta_t\}$ then $\{X_{t+T}\}$, $t \in \mathbb{R}_+$, converges in variation to $\{Y \circ \theta_t\}$, $t \in \mathbb{R}_+$, as $T$ tends to $\infty$ (see [11, p. 102] or [2, Property 2.4.1, p. 100]).

From Lemma 2, $\tilde{W}_t^j(B)$ is G/G/1 queue and, therefore, the coupling of $\tilde{W}_t^j(B)$ for any initial condition follows from Property 2.4.1 of [2, p. 100]. For a general Borel set $B$, it suffices to note that $\tilde{W}_t^j(B) \leq \tilde{W}_t^j(\mathbb{R}^d)$. The arguments used in the proof of Theorem 1 work to show that $\tilde{W}_t^j(B)$ couples for any initial condition. If $B$ is a bounded set then $W_t(B) = \sum_{j \in \mathcal{J}_B} \int_B r_j(x) \tilde{M}_t^j(\mathrm{d}x)$. Since $|\mathcal{J}_B|$ is finite, the coupling also occurs in this case.

## 3. Monotone policies

The policy we have defined to prove the sufficiency part of Theorem 1 is not of any special interest. In particular, it requires the knowledge of the mappings $f_j(x)$. Along the lines of the work done in [14], [1], and [4], it would be very appealing to find some stable policy which does not rely on knowledge of the parameters of the system.

In this section, we are going to use the key ideas of Loynes' theorem for general Polish spaces (see [2, paragraph 2.5.2, pp. 107–109]). Let the space $\mathcal{M}$ be the set of Radon measures on $\mathbb{R}^d$, which is a Polish space in the vague topology (see [10, pp. 615–631]). We define a partial order '$\preceq$' on $\mathcal{M}$. Let $m$ and $m'$ be two Radon measures with $m \preceq m'$ if $m(B) \leq m'(B)$ for all bounded Borel sets $B$ in $\mathbb{R}^d$. A policy is said to be *monotone* if $m' \preceq m$ implies that $W_t^{m'} \preceq W_t^m$ for all $t \in \mathbb{R}_+$. The *Loynes sequence* $(M_t,\ t \in \mathbb{R}_+)$ is defined as the workload found at time 0 supposing that there was zero workload at time $-t$; that is, $M_t = W_t^0 \circ \theta_{-t}$.

The following two classical lemmas are straightforward to prove.

**Lemma 4.** *Let $(m_n,\ n \in \mathbb{N})$ be a monotone sequence in $\mathcal{M}$, with respect to the order '$\preceq$', such that $m_n(B)$ is bounded for all bounded Borel sets $B$. Then $(m_n,\ n \in \mathbb{N})$ converges in $\mathcal{M}$ in the vague topology.*

**Lemma 5.** *Suppose that $\pi$ is a monotone policy. Then $(M_t,\ t \in \mathbb{R}_+)$ is a nondecreasing sequence (with respect to the order '$\preceq$').*

https://doi.org/10.1239/aap/1151337081 Published online by Cambridge University Press

In view of Lemma 5, the Loynes sequence is of particular interest for the class of monotone policies. Indeed, if $(M_t,\ t \in \mathbb{R}_+)$ is a nondecreasing sequence and is bounded by a random Radon measure $Z$ (with respect to '$\preceq$'), from Lemma 4 $M_t$ converges a.s. in $\mathcal{M}$ and we can then define the so-called *Loynes variable* as

$$M_\infty = \lim_{t\to\infty} M_t.$$

Monotone policies are quite natural in our queueing setting. Most known processing policies are monotone. In this section we discuss some conditions which guarantee the existence of a stationary workload for this class of policy. Here is an important example.

**Example 4.** *(Cone policies.)* An interesting class of policies has emerged in the literature; see [14], [1], and [4]. Let $\alpha > 0$ and for an atomic policy $m$ define

$$A_j(m) = \operatorname{argmax}\{x : m(\{x\})^\alpha r_j(x)\},$$

$$\pi_j(m) = \begin{cases} C(m)\varepsilon_j \displaystyle\sum_{x\in A_j(m)} r_j(x)^{-(\alpha+1)/\alpha}\delta_x & \text{if } m \neq 0, \\ 0 & \text{if } m = 0. \end{cases}$$

where $C(m)$ is the constant such that $\int_{\mathbb{R}^d} \pi_j(m)(\mathrm{d}x) = \varepsilon_j$ if $m \neq 0$.

Notice that $\pi(cm) = \pi(m)$ for $c > 0$. For finite workload measures, it can be shown that this policy is monotone. However, it is not clear whether or not this policy is stable when $A \in \mathcal{N}^s$.

A way to ensure that the Loynes variable is a stationary solution is to impose some continuity conditions on $\pi_j$.

**Definition 2.** Let $f$ be a measurable mapping from $\mathcal{M}$ to $\mathcal{M}$. We say that $f$ is left continuous or right continuous if for all nondecreasing or, respectively, nonincreasing converging sequences $(m_n,\ n \in \mathbb{N})$ in $\mathcal{M}$, we have $\lim_n f(m_n) = f(\lim_n m_n)$.

It is consistent to define some continuity properties in terms of converging sequences since $\mathcal{M}$ is a complete metric space. Right-continuous policies are not of practical interest. Indeed, a work-conserving policy cannot be right continuous. The cone policies of Example 4 are left continuous.

We define the *discontinuity set* of a mapping $h$ as follows: $\operatorname{disc}(h) = \{x : h(x)$ is not continuous at $x\}$.

**Proposition 2.** *Suppose that $\pi_j$ is left continuous and that $\lambda(\operatorname{disc}(r_j)) = 0$ for all $j \in \mathcal{J}$. Then when $M_\infty$ is a Radon measure, it is a stationary solution to (5).*

*Proof.* By definition, $W_s^{M_t} = W_{t+s}^0 \circ \theta_{-t} = W_{t+s}^0 \circ \theta_{-s-t} \circ \theta_s = M_{t+s} \circ \theta_s$. Therefore, from (5), for a Borel set $B$ and $t \in \mathbb{R}^+$ we have

$$\begin{aligned} M_{t+h} \circ \theta_h(B) &= W_h^{M_t}(B) \\ &= M_t(B) + \int_0^h \int_B \sigma_{s,x} A(\mathrm{d}s, \mathrm{d}x) - \sum_j \int_0^h \int_B r_j(x)\pi_j(M_{t+s} \circ \theta_s)(\mathrm{d}x)\,\mathrm{d}s. \end{aligned}$$

If $(t_k,\ k \in \mathbb{N})$ is an increasing sequence converging to $\infty$ then $(M_{t_k+s} \circ \theta_s,\ k \in \mathbb{N})$ is a nondecreasing sequence converging to $M_\infty \circ \theta_s$. Since $\pi_j$ is left continuous, we have $\lim_{k\to\infty} \pi_j(M_{t_k+s} \circ \theta_s) = \pi_j(M_\infty \circ \theta_s)$ for this vague convergence.

https://doi.org/10.1239/aap/1151337081 Published online by Cambridge University Press

For $A = \sum_n \delta_{\{T_n, X_n, \sigma_n\}}$ we define $C = \{X_n,\ n \in \mathbb{N}\}$ and let $B$ be a bounded Borel set such that $C \cap \partial B = \varnothing$, where $\partial B$ (the boundary of $B$) avoids a countable set of points in $\mathbb{R}^d$. From (5) and condition (4), $M_\infty \circ \theta_s$ and $\pi_j(M_\infty) \circ \theta_s$ are atomic measures with support $\sum_k \delta_{X_k} \mathbf{1}(T_k < s)$. Thus, for a set $B$ as above and for all $s$ in $\mathbb{R}$, we have $\pi_j(M_\infty \circ \theta_s)(\partial B) = 0$. Moreover, since $\lambda(\mathrm{disc}(r_j)) = 0$, $\pi(M_\infty \circ \theta_s)(\mathrm{disc}(r_j)) = 0$ a.s. From Lemma 7 below, we deduce that

$$\lim_{k\to\infty} \int_B r_j(x)\pi_j(M_{t_k+s} \circ \theta_s)(\mathrm{d}x) = \int_B r_j(x)\pi_j(M_\infty \circ \theta_s)(\mathrm{d}x) \quad \text{a.s.}$$

Now, from (2),

$$\sum_j \int_0^h \int_B r_j(x)\pi_j(M_{t+s} \circ \theta_s)(\mathrm{d}x) \le \sum_j h \sup_{x\in B} r_j(x) < \infty$$

by the dominated convergence theorem, whence

$$M_\infty \circ \theta_h(B) = M_\infty(B) + \int_0^h \int_B \sigma_{s,x} A(\mathrm{d}s, \mathrm{d}x) - \sum_j \int_0^h \int_B r_j(x)\pi_j(M_\infty \circ \theta_s)(\mathrm{d}x) \quad \text{a.s.}$$

From Lemma 8, below, this equation is indeed satisfied for all Borel sets, and $M_\infty$ is a stationary solution.

The assumptions of Proposition 2 can be changed as follows. Let

$$E = \{\omega \in \Omega\colon \text{ there exists a } T \text{ such that, for all } t > T,\ M_t(\omega) = M_\infty(\omega)\}.$$

On $E$, $M_t$ converges in variation (or couples). We can easily check that $E$ is a $\theta_t$-invariant event and, by ergodicity, that $\mathrm{P}(E) \in \{0, 1\}$. If $\mathrm{P}(E) = 1$ then $M_t$ couples with $M_\infty$ a.s. and the assumptions on the continuity of $\pi_j$ and $r_j$ are no longer needed to ensure the stationarity of $M_\infty$.

**Corollary 1.** *For a given policy $\pi$, if $M_t = W_t \circ \theta_{-t}$ couples with $M_\infty$ then $M_\infty$ is a stationary workload solution to (5).*

## 4. Spatial allocation

In this section, we suppose that the spatial arrival intensity $\lambda(\mathrm{d}x)$ is absolutely continuous with respect to the Lebesgue measure, i.e. it can be written as $\lambda(x)\,\mathrm{d}x$. We have seen in Theorem 1 that stability relies on the value of

$$\rho = \inf_{f\in\mathcal{F}} \sup_{j\in\mathcal{J}} \frac{1}{p_j} \int_{\mathbb{R}^d} \frac{\mathrm{E}_A^{0,x}(\sigma_{0,x}) f_j(x)}{r_j(x)} \lambda(x)\,\mathrm{d}x. \tag{10}$$

If $\rho < 1$ then the system is stable, and if $\rho > 1$ then the system is unstable. In this section, we analyze this optimization problem.

This analysis of the stability region is of particular interest in wireless communications networks (see Example 2) for which the geometry of the network is contained in the processing rates $r_j$. The server $j$ is a base station and we look for the optimal way to divide the traffic load between servers. We know that there is a mapping from a function $f$ in $\mathcal{F}$ to a stable policy $\pi$. The optimal choice is thus given by an $f$ that maximizes the permissible traffic load.

### 4.1. Optimal spatial allocation

We define

$$\rho_j(f) = \frac{1}{p_j} \int_{\mathbb{R}^d} \frac{\mathrm{E}_A^{0,x}(\sigma_{0,x}) f_j(x)}{r_j(x)} \lambda(x)\,\mathrm{d}x \quad \text{and} \quad \rho(f) = \sup_j \rho_j(f).$$

The set $\mathcal{F}$ is convex and closed and $f \mapsto \rho(f)$ is a convex function; thus, the minimum of (10) is attained. Here we consider the optimal subset of $\mathcal{F}$, defined as

$$\mathcal{F}^* = \{f \in \mathcal{F} : \rho(f) = \rho\}.$$

The extremal points of the convex set $\mathcal{F}$ are the measurable functions such that $f_j(x) = \mathbf{1}(x \in V_j)$, for a Borel set $V_j$. This function of this class is called a *tessellation*. A tessellation is a partition of the space: there exists a set $D$ with a null measure such that each point $x \in \mathbb{R}^d \setminus D$ belongs to a unique $V_j$. The policy scheme which corresponds to a tessellation is a cellular-type policy, i.e. one in which a customer is served by only one base station.

**Proposition 3.** *If $\rho$ is finite then there exists a function $f$ such that*

$$\rho_j(f) = \rho \quad \textit{for all } j.$$

*If there is a finite number of server stations, then all $f \in \mathcal{F}^*$ satisfy the above equation.*

*Proof.* Let $f \in \mathcal{F}$ and suppose, for example, that $\rho_1(f) < \rho_2(f)$. Since $\rho_2(f) > 0$, $f_2$ is not a.e. equal to 0. Thus, there exists a measurable, nonnegative function $x \mapsto \epsilon(x)$ such that $f_2^\epsilon(x) := f_2(x) - \epsilon(x) \geq 0$, $f_1^\epsilon(x) := f_1(x) + \epsilon(x) \leq 1$, and $\rho_2(\epsilon) > 0$. Let $f_j^\epsilon(x) = f_j(x)$ for $j \notin \{1, 2\}$. Then $f^\epsilon \in \mathcal{F}$ and we have $\rho_1(f^\epsilon) = \rho_1(f) + \rho_1(\epsilon)$ and $\rho_2(f^\epsilon) = \rho_2(f) - \rho_2(\epsilon)$. Thus, if $\rho_1(\epsilon)$ is small enough, $\sup_{j\in\{1,2\}} \rho_j(f^\epsilon) < \sup_{j\in\{1,2\}} \rho_j(f)$ and $\rho(f^\epsilon) \leq \rho(f)$.

Now suppose that $f \in \mathcal{F}^*$; then $f^\epsilon$ is also in $\mathcal{F}^*$. By iterating the construction above for all $j$ and $j'$ such that $\rho_{j'}(f) < \rho_j(f)$, the proposition follows.

Proposition 3 has an intuitive meaning: for an optimal spatial allocation, the traffic load is the same at each server station. We can similarly prove a more surprising result. We say that the processing rates are *singular* if there exist servers $j$ and $k$ in $\mathcal{J}$, a constant $C > 0$, and a Borel set $A$ of positive Lebesgue measure such that

$$r_j(x) = C r_k(x) \quad \text{for all } x \in A.$$

**Proposition 4.** *Suppose that $\rho$ is finite. If the processing rates are not singular, then there is an $f \in \mathcal{F}^*$ which is a tessellation. If there are finite number of server stations, then every $f \in \mathcal{F}^*$ is a tessellation.*

This is a counter-intuitive result: the server stations do not need to share the jobs to reach the stability region. The difficulty is to find an optimal tessellation solving (10). This result is not very surprising from the point of view of convex optimization, as it only asserts that the extremum is reached at an extremal point.

The definition of singular processing rates is purely technical and does not rely on any natural assumption on the processing rates. In the wireless scenario (see Example 2), if $Y_j \neq Y_k$ for all $j, k \in \mathcal{J}$ and $l$ is a strictly convex mapping, then the processing rates are nonsingular.

https://doi.org/10.1239/aap/1151337081 Published online by Cambridge University Press

*Proof of Proposition 4.* We consider the $f \in \mathcal{F}^*$ introduced in Proposition 3. Let $E = f_1((0,1))^{-1} \cap f_2((0,1))^{-1}$. In this proof, $\mu$ will denote the Lebesgue measure. We want to show that $\mu(E) = 0$. Suppose instead that $\mu(E) > 0$ and, without loss of generality, that $\mu(E) < \infty$. Let $A$ and $B$ be disjoint compact sets of positive Lebesgue measure contained in $E$. Such sets exist in view of Theorem 2.14 of [13, p. 42] (the Riesz representation theorem). We consider the mapping $\phi(x) = \mathbf{1}(x \in A) - \nu\,\mathbf{1}(x \in B)$, $\nu > 0$.

Let $f_1^\epsilon(x) = f_1(x) + \epsilon\phi(x)$, $f_2^\epsilon(x) = f_2(x) - \epsilon\phi(x)$, and $f_i^\epsilon(x) = f_i(x)$ for $i \notin \{1,2\}$. If $\epsilon > 0$ is small enough then $f^\epsilon$ and $f^{-\epsilon}$ are in $\mathcal{F}$ and, for $i \in \{1,2\}$,

$$\rho_i(f^{\pm\epsilon}) = \rho_i(f) \pm \epsilon\rho_i(\phi) = \rho \pm \epsilon\rho_i(\mathbf{1}_A) \mp \nu\epsilon\rho_i(\mathbf{1}_B).$$

Since $f \in \mathcal{F}^*$, we have $\max(\rho_1(f^{\pm\epsilon}), \rho_2(f^{\pm\epsilon})) \geq \rho$ and we deduce that $\operatorname{sgn}(\rho_1(\phi)) = \operatorname{sgn}(\rho_2(\phi))$. It follows that, for all real $\nu$, $\rho_1(\mathbf{1}_A) - \nu\rho_1(\mathbf{1}_B)$ and $\rho_2(\mathbf{1}_A) - \nu\rho_2(\mathbf{1}_B)$ have the same sign. Therefore, the vectors $(\rho_1(\mathbf{1}_A), \rho_1(\mathbf{1}_B))$ and $(\rho_2(\mathbf{1}_A), \rho_2(\mathbf{1}_B))$ are collinear, i.e. there exists a $C_{A,B}$ such that $\rho_1(\mathbf{1}_A) = C_{A,B}\rho_2(\mathbf{1}_A)$, and $C_{A,B}$ cannot depend on $B$ and, by symmetry, does not depend on $A$ either. Thus, there exists a $C > 0$ such that

$$\rho_1(\mathbf{1}_A) = C\rho_2(\mathbf{1}_A).$$

This equality has been proved for any compact set contained in $E$. From Theorem 2.14 of [13, p. 42], it can be extended to any Borel set contained in $E$. Thus, for all sets $A$ contained in $E$ such that $\mu(A) > 0$, we have

$$\frac{1}{\mu(A)} \int_A \left( \frac{\mathrm{E}_A^{0,x}(\sigma_{0,x})}{p_1 r_1(x)} - C \frac{\mathrm{E}_A^{0,x}(\sigma_{0,x})}{p_2 r_2(x)} \right) \mathrm{d}x = 0.$$

We can now apply Theorem 1.40 of [13, p. 31] to conclude that

$$C' r_1(x) = r_2(x) \quad \text{a.e. in } E.$$

This contradicts our hypothesis on the processing rates. Therefore, $\mu(E) = 0$ and we have proved that there exists an $f$ in $\mathcal{F}^*$ such that $f_j(x) = \mathbf{1}(x \in V_j)$ a.e. and $\mu(V_j \cap V_k) = 0$ for $j \neq k$. We deduce that $(\mathbf{1}(V_j))_{j \in \mathcal{J}}$ is a tessellation in $\mathcal{F}^*$.

### 4.2. Cellular policies

**Definition 3.** Let $\{V_j\}$, $j \in \mathcal{J}$, be a tessellation. A *cellular policy* with cells $\{V_j\}_j$ is a policy scheme satisfying

$$\pi_j(m)(\mathbb{R}^d \setminus V_j) = 0 \quad \text{for all } m \in \mathcal{M} \text{ and all } j.$$

We have seen in Proposition 4 that a cellular policy reaches the stability region under some conditions. We say that a cellular policy is *work conserving* if $m(V_j) > 0$ implies that $\pi_j(m)(V_j) = 1$.

**Proposition 5.** *Let $\{V_j\}$, $j \in \mathcal{J}$, be a tessellation with bounded cells. Any work-conserving cellular policy with cells $\{V_j\}_j$ is stable if*

$$\int_{V_j} \frac{\mathrm{E}_A^{0,x}(\sigma_{0,x})}{r_j(x)} \lambda(\mathrm{d}x) < p_j \quad \textit{for all } j.$$

*If there is a $j$ such that $\int_{V_j} [\mathrm{E}_A^{0,x}(\sigma_{0,x})/r_j(x)]\lambda(\mathrm{d}x) > p_j$ then* any *cellular policy with cells $\{V_j\}_j$ is unstable.*

https://doi.org/10.1239/aap/1151337081 Published online by Cambridge University Press

This proposition is similar to the result on single-server queues which asserts that the stability does not depend on the discipline, provided that it is work conserving. Since there are no interactions between server stations when a cellular policy is enforced, to prove the proposition it is sufficient to prove the following result. Suppose that the intensity measure $\lambda$ is finite and that there is a single server. If $A \in \mathcal{N}^s$ then any work-conserving policy is stable, and if $A \notin \bar{\mathcal{N}}^s$ then any policy is unstable. This result on multiclass queues is quite well known. A proof is given in the appendix.

### 4.3. Homogeneous networks in wireless communications

In this subsection, we give some results on the stability region in the wireless communications scenario (see Example 2).

4.3.1. *Spatially ergodic network.* We suppose the arrival point process $A$ to be stationary in time and space. The intensity of $A$ is denoted by $\lambda$, and $\mathrm{E}_A^{0,x}(\sigma_{0,x}) =: \sigma$. We assume that the attenuation function is radial and positive, i.e. $L(x, Y_y) = l(|Y_j - x|)$ for some positive measurable function $l$, and suppose that $\int_{\mathbb{R}_+} rl(r)\,\mathrm{d}r < \infty$. We further suppose that the point pattern $\{Y_j\}_{j\in\mathbb{N}}$ is a realization of an ergodic point process of intensity $\nu > 0$ on the plane $\mathbb{R}^2$. From the Campbell formula, we have $\mathrm{E}(I(x)) = \nu \int_{\mathbb{R}_+} rl(r)\,\mathrm{d}r < \infty$. The stability of the system depends on the value of

$$\rho_{\mathrm{c}}^{-1} = \inf_{f\in\mathcal{F}} \sup_{j\in\mathbb{N}} \int_{\mathbb{R}^2} \frac{f_j(x)I(x)}{l(|x - Y_j|)}\,\mathrm{d}x.$$

If $\lambda\sigma < \rho_{\mathrm{c}}$ then the system is stable, and if $\lambda\sigma > \rho_{\mathrm{c}}$ then the system is unstable.

**Lemma 6.** *$\rho_{\mathrm{c}}$ is a.s. constant.*

*Proof.* The mapping $(f_j)_{j\in\mathbb{N}} \mapsto (f_j(\cdot - y))_{j\in\mathbb{N}}$ is a bijection on $\mathcal{F}$. It follows that, for all $y$, $\rho$ is invariant under translations by $y$. Thus, for all $a \geq 0$, from ergodicity, $\mathrm{P}(\rho > a) \in \{0, 1\}$.

4.3.2. *Periodic network* We suppose that the set of base stations is located on a regular hexagonal grid of radius $R$. We index our base station by $\mathbb{Z}^2$ and, with a complex representation of $\mathbb{R}^2$, the base station $(p, q)$ is located at $Y_{p,q} = R(p + q\mathrm{e}^{\mathrm{i}\pi/3})$. Let $\{V_j\}$, $j \in \mathbb{Z}^2$, be the Voronoi tessellation of the hexagonal network (that is, $x \in V_j$ if $|x - Y_j| < |x - Y_{j'}|$ for all $j' \neq j$). A simple argument based on the symmetry of the hexagonal grid leads to the following proposition, which implies that the Voronoi cellular network is optimal for the hexagonal grid.

**Proposition 6.** *For the hexagonal network,*

$$\rho_{\mathrm{c}}^{-1} = \int_{V_{0,0}} \frac{I(x)}{l(|x|)}\,\mathrm{d}x.$$

4.3.3. *Poisson network.* The hexagonal grid is a regular point pattern. It is interesting to analyze the stability region when the base station point pattern is more irregular. To this end, we suppose now that the base stations are located according to a realization of a Poisson process of finite intensity $\nu > 0$.

**Proposition 7.** *Suppose that* $\limsup_{r\to\infty} l(4r)/l(r) > 0$*, $l$ is nonincreasing, and $r^2 l(r) \in L^1(\mathbb{R}_+)$. If $\{Y_j\}_{j\in\mathbb{N}}$ is a homogeneous Poisson point process of finite intensity $\nu > 0$, then $\rho_{\mathrm{c}} = 0$ a.s.*

Thus, in the homogeneous Poisson case, the network cannot be stable. Note that if $l(r) \sim r^{-\alpha}$, $\alpha > 3$, then the assumptions of the proposition hold. Whatever the intensity of the base stations is, the local behavior of the Poisson point pattern will lead to a global instability. This negative result is similar to the results in the static case given in [3] and [8]. The assumptions of this proposition are not optimal; in particular, the factor of 4 in the limit supremum is arbitrary.

*Proof of Proposition 7.* Suppose that $\rho_c > 0$ and let $B(0, R)$ denote the open ball of radius $R$ centered at the origin. We define $\mathcal{C}_n = B(0, (n+1)R) \setminus B(0, nR)$. The area of $\mathcal{C}_n$ is $|\mathcal{C}_n| = (2n+1)\pi R^2$. Let $\theta > \mathrm{e}^2 \nu$ and $S_R = (\theta/\rho_c) \sum_{n \geq 4} |\mathcal{C}_n| l((n-1)R)/l(4R) < \infty$.

For a Poisson point process $\Phi$, the event

$$A_R = \{\Phi(B(0, 2R)) = 0\} \cap \{\Phi(\mathcal{C}_2) > 2S_R\} \cap \bigcap_{n>2} \{\Phi(\mathcal{C}_n) \leq \theta |\mathcal{C}_n|\}$$

has positive probability. Indeed, the sets $\mathcal{C}_n$ are disjoint and Lemma 1.2 of [12, p. 17] yields

$$\mathrm{P}\bigg(\bigcap_n \Phi(\mathcal{C}_n) \leq \theta |\mathcal{C}_n|\bigg) = \prod_n \mathrm{P}(\Phi(\mathcal{C}_n) \leq \theta |\mathcal{C}_n|) > 0.$$

Let $N = \Phi(\mathcal{C}_2)$. On $A_R$, for $x \in B(0, R)$ we have $I(x) = \sum_j l(|x - Y_j|) \geq N l(4R)$. Moreover, if $Y_j \in \mathcal{C}_n$ then, since

$$\int_{B(0,R)} f_j(x) \frac{I(x)}{l(|x - Y_j|)} \,\mathrm{d}x \leq \rho_c^{-1},$$

we have

$$\int_{B(0,R)} f_j(x) \,\mathrm{d}x \leq \frac{1}{N\rho_c} \frac{l((n-1)R)}{l(4R)}. \tag{11}$$

It follows that

$$\begin{aligned}
\sum_j \mathbf{1}(|Y_j| \geq 3R) \int_{B(0,R)} f_j(x) \,\mathrm{d}x &= \sum_{n>2} \sum_{Y_j \in \mathcal{C}_n} \int_{B(0,R)} f_j(x) \,\mathrm{d}x \\
&\leq \frac{\theta}{N\rho_c} \sum_n |\mathcal{C}_n| \frac{l((n-1)R)}{l(4R)} \\
&\leq \tfrac{1}{2}.
\end{aligned}$$

Since $\sum_j \int_{B(0,R)} f_j(x) \,\mathrm{d}x = \pi R^2$ and $\Phi(B(0, 2R)) = 0$, we deduce that

$$\sum_j \mathbf{1}(Y_j \in \mathcal{C}_2) \int_{B(0,R)} f_j(x) \,\mathrm{d}x \geq \frac{\pi R^2}{2}.$$

By using this equation with (11) for $n = 2$, we obtain

$$\rho_c \leq \frac{l(R)}{l(4R)} \frac{2}{\pi R^2}.$$

From the hypotheses on $l$ and $l(4R)/l(R)$, for large enough $R$ we find a contradiction. We have argued in terms of the event $A_R$; since $\mathrm{P}(A_R) > 0$, by ergodicity of the Poisson point process, the result extends to the whole $\sigma$-algebra.

## Appendix A.

### A.1. Property of the stability set

In this subsection, we find some properties of the set $\mathcal{N}^s$ as it is defined in Section 2.1. To simplify the notation, $\nu_j(\mathrm{d}x)$ will denote the measure $[\mathrm{E}_A^{0,x}(\sigma_{0,x})/r_j(x)]\lambda(\mathrm{d}x)$ and $\mathcal{J}$ is chosen to be $\mathbb{N}$. For $x \in \mathbb{R}^d$ let $\mathcal{J}_x = \{j\colon f_j(x) > 0\}$, and for a set $B$ let $\mathcal{J}_B = \bigcup_{x\in B} \mathcal{J}_x$.

**Proposition 8.** *If $\mathcal{N}^s$ is not empty then there exists an $f \in \mathcal{F}$ such that $x \mapsto f_j(x)$ is continuous for all $j$ and $|\mathcal{J}_B|$ is finite for all bounded sets $B$.*

*Proof.* Let $f \in \mathcal{F}$ be such that $\rho_j = \int_{\mathbb{R}^d} f_j(x)\nu_j(\mathrm{d}x) < 1$ for all $j$. Let $G$ be a bounded open set. For all $j$, from Lusin's theorem (see, for example, [13, p. 56]) there exists a sequence of continuous functions on $G$, $g_j^n(\cdot)$, such that $0 \le g_j^n(x) \le f_j(x)$ $\nu_j(\mathrm{d}x)$-a.e. and

$$\lim_n \int_G g_j^n(x)\nu_j(\mathrm{d}x) = \int_G f_j(x)\nu_j(\mathrm{d}x).$$

We have $\sum_j g_j^n(x) \le 1 = \sum_j f_j(x)$. Let $\epsilon^n(x) = 1 - \sum_{j\ne 1} g_j^n(x)$. We define

$$f_j^n(x) = \begin{cases} g_j^n(x) + \mathbf{1}(j = 1)\epsilon^n(x) & \text{if } x \in G,\\ f_j(x) & \text{if } x \notin G. \end{cases}$$

We check that $f^n$ is in $\mathcal{F}$ and that, for $n$ large enough, $\int_{\mathbb{R}^d} f_j^n(x)\nu_j(\mathrm{d}x) < 1$ for all $j$. By iterating this construction for a set of open sets covering $\mathbb{R}^d$, we deduce that there exists an $f \in \mathcal{F}$ such that $\rho_j = \int_{\mathbb{R}^d} f_j(x)\nu_j(\mathrm{d}x) < 1$ and $f_j$ is continuous, for all $j$.

Now we turn to the second part of the proposition. Let $K$ be a compact subset, define $f_j$ as above, and let $\rho'_j$ satisfy $\max(\rho_j, \frac{1}{2}) < \rho'_j < 1$. We then define $g_j(x) = f_j(x)/\rho'_j$, such that $\int_{\mathbb{R}^d} g_j(x)\nu_j(\mathrm{d}x) = \rho_j/\rho'_j < 1$ and $\sum_j g_j(x) > 1$. For all $x$, there exists a $j_x$ such that $\sum_{j=1}^{j_x} g_j(x) > 1$. By continuity, since $K$ is compact, $j_K = \sup_{x\in K} j_x$ is finite. It follows immediately that $\tilde{f}_j(x) = g_j(x)\,\mathbf{1}(j \le j_x)/\sum_{j\le j_x} g_j(x)$ has all the properties required in the proposition.

### A.2. Vague convergence in $\mathcal{M}$

The following lemma is an adaptation of Theorem 5.2 of Billingsley [6, p. 31] to the vague topology.

**Lemma 7.** *Let $(m_n)_n$ be a sequence in $\mathcal{M}$ converging to $m$ in the vague topology. Let $h$ be a bounded, measurable function for which $m(\mathrm{disc}(h)) = 0$, and let $B$ be a bounded Borel set of $\mathbb{R}^d$ with $m(\partial B) = 0$. Then $\lim_n \int_B h(x)m_n(\mathrm{d}x) = \int_B h(x)m(\mathrm{d}x)$.*

**Lemma 8.** *Let $C$ be a countable set of points in $\mathbb{R}^d$ and let $\mathcal{C}$ be the set of bounded Borel sets, $B$, of $\mathbb{R}^d$ with $C \cap \partial B = \varnothing$. Then*

- *$\mathcal{C}$ is an algebra and the $\sigma$-algebra generated by $\mathcal{C}$, $\sigma(\mathcal{C})$, is a Borel $\sigma$-algebra $\mathcal{B}$; and*
- *for a measure $m$ defined on $\mathcal{C}$, there is a unique extension of $m$ on $\mathcal{B}$.*

*Proof.* From the relations $\partial(A \cap B) \subset \partial A \cap \partial B$, $\partial(A \cup B) \subset \partial A \cup \partial B$, and $\partial(A^{\mathrm{c}}) = \partial A$, it follows that $\mathcal{C}$ is an algebra on the set of subsets of $\mathbb{R}^d$.

$\mathcal{B}$ is the $\sigma$-algebra of the algebra generated by the open rectangles of $\mathbb{R}^d$. To prove the first assertion of the lemma, it suffices to prove that the rectangle $(0, 1)^d$ can be written as $\bigcup_{n\in\mathbb{N}} B_n$,

https://doi.org/10.1239/aap/1151337081 Published online by Cambridge University Press

where $B_1 \subset B_2 \subset \cdots \subset B_n \in \mathcal{C}$. To this end, consider the rectangle $R_\varepsilon = (\varepsilon, 1-\varepsilon)^d$ with $0 < \varepsilon < \frac{1}{2}$. If $\varepsilon \neq \varepsilon'$ then $\partial R_\varepsilon \cap \partial R_{\varepsilon'} = \varnothing$. Since $C$ is countable there can only be a countable set of $\varepsilon$s such that $C \cap \partial R_\varepsilon \neq \varnothing$. In particular, there exists an increasing sequence $(\varepsilon_n)_n$ such that $C \cap \partial R_{\varepsilon_n} = \varnothing$. This proves the first statement of the lemma. The second assertion follows from the Caratheodory extension theorem.

### A.3. Spatial queueing system with one server

When there is only one server in the system, Theorem 1 can be made more precise. The system reduces to a multiclass queue. The condition $\lambda \in \mathcal{N}^s$ can be restated as

$$\int_{\mathbb{R}^d} \frac{\mathrm{E}_A^{0,x}(\sigma_{0,x})}{r(x)} \lambda(\mathrm{d}x) < p,$$

where $r$ is the processing rate for the server providing service for a user located at $x$ and $p$ is the expectation of the available processing power.

We make the following proposition.

**Proposition 9.** *Suppose that the intensity measure $\lambda(\mathrm{d}x)$ is finite and that there is a single server. If $A \in \mathcal{N}^s$ then any work-conserving policy is stable.*

*Proof.* Since $\lambda$ is a finite Radon measure, $N = \sum_n \delta_{\{T_n, \sigma_n / r(X_n)\}}$ is a simple marked point process on $\mathbb{R}$ with finite intensity $\mathrm{E}(N[0,1]) = \int_{\mathbb{R}^d} \lambda(\mathrm{d}x)$. For a given work-conserving policy $\pi$, define $Y_t = \int_{\mathbb{R}^d} W_t(\mathrm{d}x)/r(x)$. From (5), we deduce that

$$Y_t = \left( Y(T_n-) + \frac{\sigma_n}{r(X_n)} - \int_{T_n}^t \varepsilon(t)\,\mathrm{d}t \right)^+ \quad \text{for } t \in [T_n, T_{n+1}),$$

where $\varepsilon(t)$ is the total processing power available to the server station. $Y_t$ does not depend on the policy and is the usual workload for the G/G/1 queue. The workload for our queue is

$$\int_{\mathbb{R}^d} \mathrm{E}_N^0 \left( \frac{\sigma_0}{r(X_0)} \right) \lambda(\mathrm{d}x) = \int_{\mathbb{R}^d} \frac{\mathrm{E}_A^{0,x}(\sigma_{0,x})}{r(x)} \lambda(\mathrm{d}x) < p.$$

By writing $W_n = W_{T_n-}$, it appears that $(W_n, n \geq 0)$ is generated by a stochastic recurrence; see [9, Chapter 3] and [2, p. 104].

If $W_0$ is an atomic measure with a finite set of atoms on $\mathbb{R}^d$ then a.s. so is $W_t$ for $t \geq 0$. We define the following policy on atomic measures with a finite set of atoms, where $x^- = \operatorname{argmin}\{x \colon \mathbf{1}(m(\{x\}) > 0) r(x)\}$:

$$\pi^-(m) = \begin{cases} 0 & \text{if } m \text{ is the zero measure,} \\ \varepsilon \delta_{x^-} & \text{otherwise.} \end{cases}$$

If multiple choices of $x^-$ are possible, we choose the first in lexicographical order.

This policy is the work-conserving policy which dedicates all the processing power to the slowest customer. It is monotone, and $M_t^-(B) \leq M_t^-(\mathbb{R}^d)$. As has already been pointed out, $M_t^-(\mathbb{R}^d)$ is the Loynes sequence for the usual stable G/G/1 queue. From Lemmas 4 and 5, we deduce that $M_t^-$ converges a.s. to the Loynes variable $M_\infty^-$.

Now consider any work-conserving policy $\pi$. We similarly define the Loynes variable $M_t$ for this policy, such that $M_t(\mathbb{R}^d) \leq M_t^-(\mathbb{R}^d) \leq M_\infty^-(\mathbb{R}^d)$. The event $A = \{M_{T_n}^- = 0\}$ is a

https://doi.org/10.1239/aap/1151337081 Published online by Cambridge University Press

renovating event for $M_{T_n}$ and, since the workload of the G/G/1 queue is strictly less than 1, $\mathrm{P}_N^0(A) \geq \mathrm{P}_N^0(M_\infty^-(\mathbb{R}^d) = 0) > 0$. From Theorem 2.5.3 and Property 2.5.5 of [2, p. 115, p. 117], we deduce that $M_t$ converges to a stationary solution $M_\infty$, and that $M_t$ couples with $M_\infty$ (in the strong backward sense).

Restating Property 2.4.1 of [2, p. 100], we can also prove that from any finite initial condition $m$, $W_t^m$ couples with $M_\infty \circ \theta_t$ (as $t$ tends to $\infty$).

## References

[1] Armony, M. and Bambos, N. (2003). Queueing dynamics and maximal throughput scheduling in switched processing systems. *Queueing Systems* **44,** 209–252.

[2] Baccelli, F. and Brémaud, P. (2003). *Elements of Queueing Theory. Palm Martingale Calculus and Stochastic Recurrences* (Appl. Math. **26**), 2nd edn. Springer, Berlin.

[3] Baccelli, F., Blaszczyszyn, B. and Tournois, F. (2003). Downlink admission/congestion control and maximal load in CDMA networks. In *Proc. INFOCOM 2003*, Vol. 1, ed. F. Bauer, IEEE, Piscataway, NJ, pp. 723–733.

[4] Bambos, N. and Michailidis, G. (2004). Queueing and scheduling in random environments. *Adv. Appl. Prob.* **36,** 293–317.

[5] Bambos, N. and Michailidis, G. (2005). Queueing networks of random link topology: stationary dynamics of maximal throughput schedules. *Queueing Systems* **50,** 5–52.

[6] Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley, New York.

[7] Bordenave, C. (2004). Stability properties of data flows on a CDMA network in macrodiversity. INRIA Tech. Rep. RR-5257. Available at http://www.inria.fr/rrrt/rr-5257.html.

[8] Bordenave, C. (2006). Spatial capacity of multiple access networks. Submitted. Draft version available at http://www.inria.fr/rrrt/rr-5102.html.

[9] Borovkov, A. (ed.) (1998). *Ergodicity and Stability of Stochastic Processes* (Wiley Ser. Prob. Statist.). John Wiley, Chichester.

[10] Daley, D. J. and Vere-Jones, D. (1988). *An Introduction to the Theory of Point Processes* (Springer Ser. Statist.). Springer, New York.

[11] Lindvall, T. (1992). *Lectures on the Coupling Method* (Wiley Ser. Prob. Math. Statist. Prob. Math. Statist.). John Wiley, New York.

[12] Penrose, M. (2003). *Random Geometric Graphs* (Oxford Stud. Prob. **5**). Oxford University Press.

[13] Rudin, W. (1974). *Real and Complex Analysis* (McGraw-Hill Ser. Higher Math.), 2nd edn. McGraw-Hill, New York.

[14] Tassiulas, L. and Ephremides, A. (1992). Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. *IEEE Trans. Automatic Control* **37,** 1936–1948.

[15] Tassiulas, L. and Ephremides, A. (1996). Throughput properties of a queueing network with distributed dynamic routing and flow control. *Adv. Appl. Prob.* **28,** 285–307.

[16] Wasserman, K., Michailidis, G. and Bambos, N. (2001). Differentiated processors scheduling on heterogeneous task flows. Tech. Rep. SU-Netlab-2001-12/1, Engineering Library, Stanford University.

[17] Wasserman, K. M., Michailidis, G. and Bambos, N. (2005). Optimal processor allocation to differentiated job flows. *Performance Eval.* **63,** 1–14.