ANALYSIS OF SOME LARGE MARKOVIAN QUEUEING NETWORKS

JAMES BENEDICT MARTIN
St John's College and Statistical Laboratory
University of Cambridge

A dissertation submitted for the degree of
Doctor of Philosophy

January 1999

Preface

It has been a tremendous pleasure to work under the guidance of Yuri Suhov, and I would like to record here my huge gratitude to him for his never-failing enthusiasm, generosity and patience.

I would also like to thank everyone in the Statistical Laboratory for making it such a congenial place to work, and in particular my contemporaries Giles, Katie and Meena for their friendship and support.

This dissertation is not substantially the same as any that I have submitted for any degree or diploma or other qualification at any university.

Apart from Chapter 2, which is partly the result of joint work done with Yuri Suhov, the work in this dissertation is entirely my own, and contains nothing which is the outcome of work done in collaboration with others.

James Martin
St John's College
24th January, 1999

Contents

Chapter 1

Introduction

Recent years have seen a great increase in the study of queueing network theory. This theory initially emerged as a natural extension of the theory of a single queue. However, as the number of components of a network increases, and as their interaction becomes more complex, the exact analysis of the behaviour of the network can quickly become intractable. Hence it is valuable to obtain approximations which, although not exact for any finite network, become increasingly accurate as the size of the network grows. This approach is closely related to the mean-field methods used in statistical mechanics to obtain approximations for the behaviour of large systems of interacting particles. Much of the work presented here, particularly in Chapters 2 and 3, is in this spirit.

A recent paper of this kind, which provided a motivation for much of the work presented here, is that of Vvedenskaya, Dobrushin and Karpelevich [51]. (In fact this was one of the last papers on which Dobrushin worked before his death). There a join-the-shortest-queue (JSQ) policy is considered for a system of many parallel queues.

JSQ-type policies can be regarded as one of the simplest types of "dynamic routing", and have an obvious appeal due to their simplicity and generality. They have been widely studied, and, from among the many available references, one might mention the early paper of Kingman [33], Whitt's paper [57] in which the efficiency of JSQ is assessed for a variety of models and many references are given, and Walrand's book [53], which contains proofs of the optimality of JSQ for various models with a variety of cost functionals. The models considered tend to consist of a small number of servers in parallel, out of which an arriving customer chooses the one with the shortest queue at the time of his arrival.

For a real system with a large number of servers in parallel, the overheads involved in allowing each arrival to observe and choose between all possible queues may become prohibitive, and thus it is interesting to consider systems in which each customer chooses the shortest from some subset of the queues. The model considered in [51] is as follows. The system consists of N identical channels, each with an infinite buffer and a single server with service rate μ. The input flow is Poisson with rate $N\lambda$; service times and arrival times are all independent. Upon arrival, each task selects m channels at random, and joins the one whose queue is the shortest. We will refer to such a system as a *VDK node* (after the authors of [51]).

The state of the system may be described by the vector $\underline{r}_N = \{r_N(n), n \in \mathbb{Z}_+\}$, (here and below $\mathbb{Z}_+$ is the set of non-negative integers), where $r_N(n) = N^{-1}\sum_{n' \geq n} M(n')$, and $M(n')$ is the number of channels with queue length n'. Hence $r_N(n)$ is the proportion of queues in the system whose length is at least n. The process $\underline{r}_N(t) = \{r_n(n,t), n \in \mathbb{Z}_+, t \geq 0\}$ describing the state of the network at times $t \geq 0$ is easily seen to be a Markov process, with state space given by

$$\overline{\mathcal{U}}_N = \{\underline{g} = (g(n), n \in \mathbb{Z}_+) : g(0) = 1, g(n) \geq g(n+1) \geq 0, Ng(n) \in \mathbb{N} \text{ for all } n\} \tag{1.1}$$

If $\lambda < \mu$, the process is positive recurrent, with a unique stationary distribution π_N, and the main result of [51] is that

$$\lim_{N\to\infty} \mathbb{E}_{\pi_N} r_N(n) = \left(\frac{\lambda}{\mu}\right)^{(m^n-1)/(m-1)}, \ n \geq 0. \tag{1.2}$$

Thus, as $N \to \infty$, a "typical server" in the system will have $\geq n$ tasks in its buffer with a probability that decays super-exponentially as $n \to \infty$. We may compare this to a "linear" system, in which arriving tasks choose a channel at random; this system is equivalent to N isolated $M/M/1$ queues with arrival and service rates λ and μ. A typical server in the linear system has $\geq n$ tasks in the buffer with probability $(\lambda/\mu)^n$, $n \geq 0$, (independently of N), which is larger than the RHS of (1.2) and decays only exponentially as $n \to \infty$.

In fact, as is shown in [51], the whole process $\{\underline{r}_N(t)\}$ is asymptotically deterministic as $N \to \infty$. More precisely, extend (1.1) in the natural way by defining

$$\overline{\mathcal{U}} = \{\underline{g} = (g(n), n \in \mathbb{Z}_+) : g(0) = 1, g(n) \geq g(n+1) \geq 0 \text{ for all } n\}. \tag{1.3}$$

Then if the distribution of the initial state $\underline{r}_N(0)$ approaches the Dirac delta-measure concentrated at a point $\underline{g} \in \overline{\mathcal{U}}$, the distribution of $\{\underline{r}_N(t)\}$ is concentrated in the limit on the trajectory $\underline{u}(t,\underline{g}) = \{u(n,t,\underline{g}), n \in \mathbb{Z}_+\}$, $t \geq 0$, giving the solution to the following system of differential equations:

$$\dot{u}(n,t) = \lambda[u(n-1,t)^m - u(n,t)^m] - \mu[u(n,t) - u(n+1,t)], \ n \geq 1 \tag{1.4}$$

with boundary conditions

$$u(n,0) = g(n), \ n \geq 1, \quad \text{and} \quad u(0,t) = 1 \text{ for all } t.$$

The RHS of (1.2) provides a stationary solution to (1.4).

A similar model is considered by Turner in [50], who obtains a functional law of large numbers for the convergence of the paths of the network process as $N \to \infty$ if the initial conditions converge to a deterministic limit. The methods of Vvedenskaya *et al.* in [51] and of Turner in [50] (and related methods used in earlier papers by Hunt and others) are compared in Section 2.4. Turner also develops a coupling which, for fixed N, compares the behaviour of the system for different values of m. Mitzenmacher, in [41] and [42] investigates similar models and makes some observations concerning systems in which the information about queue-lengths available to arrivals is not necessarily up-to-date. See also [52], where a similar model with a more general class of routing regimes and service distributions is considered, and [17] and the review paper [28] for the discussion of mean-field-type approximations for a wide range of queueing network models.

In Chapters 2 to 4 we will analyse a model of a network whose nodes have the form of the VDK nodes described above. In order to introduce this, we recall the famous model of a *Jackson network*. This model, which was introduced by Jackson in [24], [25], has been one of the most popular and widely studied in queueing network theory. Attractions of a Jackson network lie in the simplicity of its construction, and in the partial exact solvability expressed by the product formula for the stationary distribution of the Markov process describing the evolution of the state of the network. For a detailed treatment of the model and its extensions, see for example [31] or [58]; here we give brief details which will be needed for later reference. In an open Jackson network on a finite set of nodes S, each node j has a single server with an infinite buffer and service rate μ_j. The network is fed by a family of independent Poisson flows of rate $\lambda_j, j \in S$, arriving at the corresponding nodes. After being served at node j, a task joins the queue at node k with probability p_{jk}, and leaves the network with probability $p^*_j = 1 - \sum_k p_{jk}$, where $0 \leq p_{jk} \leq 1$ for all j, k, and $\sum_k p_{jk} \leq 1$ for all j. In general we will assume that the $|S| \times |S|$ matrix $(\mathbf{I} - \mathbf{P})$ is invertible; then $(\mathbf{I} - \mathbf{P})^{-1} = \sum_{i=0}^{\infty} \mathbf{P}^i$. This condition guarantees that the expected total number of nodes visited by a particular task is finite.

The state of a Jackson network at time $t \geq 0$ can be described by the vector $\mathbf{n}(t) = (n_j(t), j \in S)$, where $n_j(t)$ is the length of the queue at node j at time t. It is easy to see that $\mathbf{n}(t)$ is a Markov process, with state space $\mathbb{Z}_+^J$. Its stationary distribution, if one exists, may be determined as follows. Consider the system of equations

$$\rho_j = \lambda_j + \sum_{k \in S} \rho_k p_{kj},$$

for all $j \in S$, which we may rewrite in vector notation as

$$\boldsymbol{\rho} = \boldsymbol{\lambda} + \boldsymbol{\rho}\mathbf{P}. \tag{1.5}$$

Here the row vector $\boldsymbol{\lambda}$ indexed by the set of nodes S is the vector of external arrival rates, and $\mathbf{P} = \{p_{jk}\}$ is the routing matrix. The row vector $\boldsymbol{\rho} = (\rho_j, j \in S)$ is the unknown, and an entry ρ_j will represent the "effective arrival rate" at node j, including arrivals from inside as well as outside the network. (1.5) has solution

$$\boldsymbol{\rho} = \boldsymbol{\lambda}(\mathbf{I} - \mathbf{P})^{-1}. \tag{1.6}$$

Under the *non-overload condition*

$$\boldsymbol{\rho} < \boldsymbol{\mu}, \quad \text{(i.e. } \rho_j < \mu_j \text{ for all } j), \tag{1.7}$$

the network is positive recurrent, and its stationary distribution may be given by

$$\mathbb{P}(N_j \geq r_j \; \forall j \in S) = \prod_{j \in S} \left(\frac{\rho_j}{\mu_j} \right)^{r_j}, \quad \mathbf{r} = (r_j, j \in S) \in \mathbb{Z}_+^S, \tag{1.8}$$

which is the product of the geometric distributions with paramters $1 - (\rho_j/\mu_j)$ at node j, $j \in S$.

The model we consider in Chapters 2 to 4 has the form of a Jackson network whose nodes are VDK nodes. In Chapter 2, we give a formal definition of the model, and we analyse a mean-field limit of the type that was considered by Vvedenskaya *et al.* in [51]. We show that as $N \to \infty$, the process representing the evolution of the system converges to a limiting trajectory which corresponds to the solution of a countably infinite system of non-linear ODEs (corresponding to the system (1.4) for the model of [51]). Under a non-overload condition corresponding to (1.7), we then show that the system has a unique invariant distribution for each N, and that, as $N \to \infty$, these invariant distributions converge to a limit corresponding to the unique fixed point of the system of ODEs. In the limit, the tail of the distribution of queue lengths decays super-exponetially, rather than exponentially as in the case of a standard Jackson network. The resulting decrease in average waiting times which occurs prompts the name "Fast Jackson Network" which we give to the model. The techniques used involve an extension and simplification of the differential equation and Markov generator methods of [51], and the development of coupling methods to provide estimates that are uniform in N on the characteristics of the equilibrium distributions.

In Chapter 3, we again consider the limit as $N \to \infty$, but now focus on the processes of arrivals and departures at individual channels within the various nodes of the network, rather than on the overall distribution of queue-lengths. If the initial conditions converge suitably, the distributions of these processes converge in local variation distance to a limit under which each channel evolves independently. We discuss the limiting processes which arise, and investigate the point processes of arrivals and departures at a channel when the networks are in equilibrium, for various values of the system parameters. For a network in equilibrium, another interpretation of the decoupling that occurs in the limit is that a typical task, given its route through the network, experiences a sequence of independent waiting times.

In Chapter 4, we consider such systems with fixed (finite) N, and use coupling techniques to provide stochastic bounds which compare the behaviour of the system with that of a collection of independent VDK nodes. This is a generalisation of a result of Massey [40], who gave bounds for the behaviour of a standard Jackson network in terms of that of independent $M/M/1$ queues. The bounds apply when the network is transient as well as in the case where an equilibrium exists; in fact they are asymptotically tight exactly when every node is at least fully loaded. They compare the probabilities of certain increasing sets of network states, and cannot be derived by means of standard pathwise, or even pointwise, stochastic orderings of the states of the two systems.

In the final chapter, we return to the situation of a "standard" Jackson network, in which each node of the network contains a single server; but now the set S of nodes is taken to be countably infinite. In this case, the Markov process representing the evolution of the network may not be unique, and even when a unique construction is possible, the equilibrium distribution for the network is not necessarily unique. Some conditions are given for when such uniqueness properties hold, and the structure of the set of invariant distributions and of their basins of attraction is investigated in the general case. The model is an extension of that considered by Kel'bert, Kontsevich and Rybko in [30], and the analysis of invariant distributions is closely related to that carried out by Andjel in [2] for the zero-range process.

I have attempted to keep the notation as consistent as possible between the chapters, although the different frameworks considered have required a few new notations to be introduced in each case. Bold vectors refer throughout to vectors indexed by the set S of nodes of the network; in particular $\boldsymbol{\lambda} = \{\lambda_j, j \in S\}$ is the vector of arrival rates and $\boldsymbol{\mu} = \{\mu_j, j \in S\}$ is the vector of service rates. Such vectors are to be considered as row vectors when matrix notation is used. The routing matrix is represented throughout by $\mathbf{P} = \{p_{jk}, j, k \in S\}$, where p_{jk} is the probability of moving to node k on leaving node j. Then the vector $\mathbf{p}^* = \{p_j^*, j \in S\}$, where $p_j^* = 1 - \sum_{k \in S} p_{jk}$ gives the probability of leaving the network entirely after being served at node j. $\mathbb{Z}_+$ represents the set of non-negative integers; vectors indexed by $\mathbb{Z}_+$ are written as underlined, and vectors indexed by both S (representing the set of nodes) and $\mathbb{Z}_+$ (representing the set of possible queue-lengths) are written as bold and underlined.

One aspect which is not especially consistent is the naming of the entities which receive service in the network; sometimes these will be referred to as "tasks", (reflecting perhaps the manufacturing origins of the Jackson network), at other times as "customers", which is a comfortable anthropomorphism, and at other times as "particles", which feels natural in the case of a rather physical model such as that considered in Chapter 5. In any case, no essential mathematical difference is intended by the use of these various alternatives.

Chapter 2

Fast Jackson Networks

2.1 Introduction and Main Results

We now specify the model of a Fast Jackson Network that we will analyse in the next three chapters. S is a finite set of nodes; each node $j \in S$ contains N channels, each of which has an infinite buffer and a single server with service rate μ_j. The network is fed by a family of independent Poisson flows of rates $N\lambda_j, j \in S$, arriving at the corresponding nodes. As before, after being served at node j, a task jumps to node k with probability p_{jk} and leaves the network with probability p_j^*. Upon arrival at node j, either from outside or from inside the network, a task selects m of the N channels there uniformly at random (with replacement, though this becomes unimportant as N becomes large), and enters the one with the shortest queue (breaking ties at random if necessary). The network is therefore specified by the set S and the parameters $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$, $\mathbf{P}$, N and m.

The state of the network at time $t \geq 0$ may be described by the collection $\underline{\mathbf{r}}(t) = (r_j(n,t), j \in S, n \in \mathbb{Z}_+)$ where $r_j(n,t) = N^{-1}\sum_{n' \geq n} M_j(n',t)$ and $M_j(n',t)$ is the number of channels at node j with queue length n' at time t. Hence $r_j(n,t)$ measures the proportion of queues at j whose length is at least n at time t. The process $\{\underline{\mathbf{r}}(t), t \geq 0\}$ (we will generally write $\{\underline{\mathbf{r}}_N(t)\}$ to stress the dependence on N) forms a Markov process, whose state space is the Cartesian product $\overline{\mathcal{U}}_N^S$, where $\overline{\mathcal{U}}_N$ is defined in (1.1).

The generator of $\{\underline{\mathbf{r}}_N(t)\}$ is given by:

$$\begin{aligned}\mathbf{A}_N f(\underline{\mathbf{g}}) = N \sum_{n\geq 1}\sum_{j\in S} \lambda_j \left[g_j(n-1)^m - g_j(n)^m\right] \left[f\left(\underline{\mathbf{g}} + \frac{\underline{\mathbf{e}}_j(n)}{N}\right) - f(\underline{\mathbf{g}})\right] \\ + N \sum_{n\geq 1}\sum_{j\in S} \mu_j p_j^* \left[g_j(n) - g_j(n+1)\right] \left[f\left(\underline{\mathbf{g}} - \frac{\underline{\mathbf{e}}_j(n)}{N}\right) - f(\underline{\mathbf{g}})\right] \\ + N \sum_{n,n'\geq 1}\sum_{j,k\in S} \mu_k p_{kj} \left[g_k(n') - g_k(n'+1)\right] \\ \times \left[\{g_j(n-1) - \frac{1}{N}\delta_{j,k}\delta_{n-1,n'}\}^m - \{g_j(n) - \frac{1}{N}\delta_{j,k}\delta_{n,n'}\}^m\right] \\ \times \left[f\left(\underline{\mathbf{g}} + \frac{\underline{\mathbf{e}}_j(n)}{N} - \frac{\underline{\mathbf{e}}_k(n')}{N}\right) - f(\underline{\mathbf{g}})\right],\end{aligned} \tag{2.1}$$

where $\underline{\mathbf{g}} = (g_j(n), n \in \mathbb{Z}_+, j \in S) \in \overline{\mathcal{U}}_N^S$ and f is a function from $\overline{\mathcal{U}}_N^S$ to $\mathbb{R}$; here $\underline{\mathbf{e}}_k(n')$ denotes the vector $(e_i(n) : n \in \mathbb{Z}_+, i \in S)$ whose only non-zero entry is $e_i(n') = 1$, and $\delta_{j,k} = 1$ if $j = k$ and 0 otherwise. We denote by $\mathbf{T}_N(t)$ the semigroup of transition operators generated by $\mathbf{A}_N$, defined formally by $\mathbf{T}_N(t) = \exp(t\mathbf{A}), t \geq 0$, and acting on functions $f : \overline{\mathcal{U}}_N^S \to \mathbb{R}$.

As well as the space $\overline{\mathcal{U}}^S$, where $\overline{\mathcal{U}}$ is defined in (1.3), we will also use the space $\mathcal{U}^S$, where

$$\mathcal{U} = \{\underline{g} = (g(n), n \in \mathbb{Z}_+) : g(0) = 1, g(n) \geq g(n+1) \geq 0 \text{ for all } n, \sum_{n=0}^{\infty} g(n) < \infty\}. \tag{2.2}$$

The space $\mathcal{U}^S$ represents limiting states where the average queue length per channel is finite at each node. We will define the norm

$$\|\underline{\mathbf{u}}\| = \sup_{j\in S} \sup_{n\in\mathbb{Z}_+} \frac{|u_j(n)|}{\max(n,1)} \tag{2.3}$$

on the space $(\mathbb{R}^{\mathbb{Z}_+})^S$; this norm is not proper on $(\mathbb{R}^{\mathbb{Z}_+})^S$, in that it sometimes takes the value ∞, but it induces a proper metric on the subsets $\mathcal{U}^S$, $\overline{\mathcal{U}}^S$ and $\overline{\mathcal{U}}_N^S$. Under this metric, $\overline{\mathcal{U}}^S$ becomes a complete compact metric space; the associated topology is to be understood when we refer to continuity and convergence below. Note that $\overline{\mathcal{U}}$ may be interpreted as the space of sub-probability measures on $\mathbb{Z}_+$; and $\mathcal{U}$ as the subspace of probability measures with a finite first moment; the topology induced by (2.3) then corresponds to the topology of weak convergence on the one-point compactification $(\mathbb{Z}_+ \cup \{\infty\})$ of $\mathbb{Z}_+$. The choice of these spaces leads to a convenient description in the limit $N \to \infty$.

The main results of the chapter are stated in Theorems 2.1.1–2.1.4 below. We will frequently refer to the non-overload condition (1.7), whose form is unchanged in the modified setting. An important role will be played by the following infinite system of non-linear differential equations (corresponding to the system (1.4) used in [51]) for $\underline{\mathbf{u}}(t) = (u_j(n,t), j \in S, n \in \mathbb{Z}_+)$, $t \geq 0$, with initial condition $\underline{\mathbf{g}} \in \overline{\mathcal{U}}^S$:

$$\underline{\mathbf{u}}(0) = \underline{\mathbf{g}}, \tag{2.4}$$

$$\dot{\underline{\mathbf{u}}}(t) = \underline{\mathbf{h}}(\underline{\mathbf{u}}(t)), \tag{2.5}$$

where, for all $j \in S$,

$$h_j(0, \underline{\mathbf{u}}) = 0, \tag{2.6}$$

$$h_j(n, \underline{\mathbf{u}}) = \left[\lambda_j + \sum_{k \in S} \mu_k p_{kj} u_k(1)\right] [u_j(n-1)^m - u_j(n)^m] - \mu_j [u_j(n) - u_j(n+1)] \tag{2.7}$$

for all $n \geq 1$.

The first theorem establishes various properties of this system, and the second theorem describes the way in which the behaviour of the processes $\underline{\mathbf{r}}_N(t)$ converges asymptotically to that of the limiting system of ODEs as $N \to \infty$, for finite t.

Theorem 2.1.1
(i) If $\underline{\mathbf{g}} \in \overline{\mathcal{U}}^S$*, the system (2.4)–(2.7) has a unique solution* $\underline{\mathbf{u}}(t, \underline{\mathbf{g}}), t \geq 0$ *in* $\overline{\mathcal{U}}^S$.
(ii) Under (1.7) there exists a unique fixed point $\underline{\mathbf{a}} \in \mathcal{U}^S$ *such that* $\underline{\mathbf{u}}(t, \underline{\mathbf{a}}) = \underline{\mathbf{a}}$ *for all* t*, with*

$$a_j(n) = \left(\frac{\rho_j}{\mu_j}\right)^{(m^n - 1)/(m-1)}. \tag{2.8}$$

(iii) Under (1.7)

$$\lim_{t \to \infty} \underline{\mathbf{u}}(t, \underline{\mathbf{g}}) = \underline{\mathbf{a}} \quad \text{for all } \underline{\mathbf{g}} \in \mathcal{U}^S. \tag{2.9}$$

Thus there exists a unique probability measure π *on* $\mathcal{U}^S$ *which is invariant under the map* $\underline{\mathbf{g}} \mapsto \underline{\mathbf{u}}(t, \underline{\mathbf{g}})$*, so that*

$$\int f(\underline{\mathbf{g}}) d\pi(\underline{\mathbf{g}}) = \int f(\underline{\mathbf{u}}(t, \underline{\mathbf{g}})) d\pi(\underline{\mathbf{g}})$$

for all $t \geq 0$, $f : \overline{\mathcal{U}}^S \to \mathbb{R}$*; and* $\pi = \delta_{\underline{\mathbf{a}}}$*, the probability measure concentrated at the fixed point* $\underline{\mathbf{a}}$.

Theorem 2.1.2 *For any continuous function* $f : \overline{\mathcal{U}}^S \to \mathbb{R}$ *and* $t \geq 0$*,*

$$\lim_{N \to \infty} \sup_{\underline{\mathbf{g}} \in \overline{\mathcal{U}}_N^S} |\mathbf{T}_N(t) f(\underline{\mathbf{g}}) - f(\underline{\mathbf{u}}(t, \underline{\mathbf{g}}))| = 0,$$

and the convergence is uniform in t *within any bounded interval.*

The next result provides a coupling which compares the behaviour of the network described by the process $\underline{\mathbf{r}}_N(t)$ with that of a "linear network" where $m = 1$. See the remark after the proof of Theorem 2.1.3 for suggestions regarding an extension of the theorem.

Theorem 2.1.3 *Let* Q *and* R *be two networks with the same set of nodes* S*, the same parameters* $\boldsymbol{\lambda}$*,* $\boldsymbol{\mu}$*,* $\mathbf{P}$ *and* N*, and the same state at* $t = 0$*, and with* $m(Q) = 1$*,* $m(R) \geq 1$*. Then there is a coupling of the processes* $\{\underline{\mathbf{r}}_N^{(Q)}(t)\}$ *and* $\{\underline{\mathbf{r}}_N^{(R)}(t)\}$ *such that, for all* $t \geq 0$*, the total number of tasks present in* R *at time* t *is no greater than the total number in* Q *at* t.

The final result demonstrates the convergence of the stationary distributions, where they exist, to the probability measure concentrated at the fixed point of the system of ODEs. Thus, from the RHS of (2.8), we have, for $m > 1$, super-exponential decay of the tail of the queue length distribution in the limiting stationary regime, in contrast to the exponential decay occurring in the case $m = 1$.

Theorem 2.1.4 *Under (1.7),*
(i) The Markov process $\underline{\mathbf{r}}_N(t)$ is positive recurrent for all N, and hence has a unique invariant distribution π_N for each N.
(ii) $\pi_N \to \delta_{\underline{\mathbf{a}}}$ weakly as $N \to \infty$, where $\delta_{\underline{\mathbf{a}}}$ is given by Theorem 2.1.1, (ii) and (iii); i.e.

$$\lim_{N\to\infty} \mathbb{E}_{\pi_N} f(\underline{\mathbf{g}}) = f(\underline{\mathbf{a}})$$

for all continuous functions $f : \overline{\mathcal{U}}^S \to \mathbb{R}$.

The next two sections are devoted to the proof of Theorems 2.1.1–2.1.4. The methods used to prove Theorems 2.1.1 and 2.1.2 are essentially those that may be used in the case of a single VDK node, with some modifications; we are also able to simplify in some aspects the approach used in [51]. Later results, especially the proof of Theorem 2.1.3, require more particular attention to the network structure.

2.2 Limiting equations and convergence for finite t

Proof of Theorem 2.1.1(i): Define $\xi(x) = [\min(x,1)]_+$, where $[y]_+ = \max(y,0)$, and consider the following modification of (2.4)–(2.7):

$$\underline{\mathbf{u}}(0) = \underline{\mathbf{g}}, \tag{2.4$'$}$$

$$\dot{\underline{\mathbf{u}}}(t) = \tilde{\underline{\mathbf{h}}}(\underline{\mathbf{u}}(t)), \tag{2.5$'$}$$

where, for all $j \in S$.

$$\tilde{h}_j(0,\underline{\mathbf{u}}) = 0, \tag{2.6$'$}$$

$$\begin{aligned}\tilde{h}_j(n,\underline{\mathbf{u}}) = {} & \left[\lambda_j + \sum_{k\in S} \mu_k p_{kj} \xi(u_k(1))\right] \times [\xi(u_j(n-1))^m - \xi(u_j(n))^m]_+ \\ & - \mu_j[\xi(u_j(n)) - \xi(u_j(n+1))]_+ \quad \text{for all } n \geq 1.\end{aligned} \tag{2.7$'$}$$

Since the RHS of (2.7) and (2.7$'$) are the same if $\underline{\mathbf{u}} \in \overline{\mathcal{U}}^S$, the system (2.4)–(2.7) has the same solutions in $\overline{\mathcal{U}}^S$ as the system (2.4$'$)–(2.7$'$). Also, if $\underline{\mathbf{g}} \in \overline{\mathcal{U}}^S$, then any solution of (2.4$'$)–(2.7$'$) remains within $\overline{\mathcal{U}}^S$, since, under (2.7$'$), if $u_j(n) \leq u_j(n+1)$ for some j, n, then $\tilde{h}_j(n,\underline{\mathbf{u}}) \geq 0$ and $\tilde{h}_j(n+1,\underline{\mathbf{u}}) \leq 0$, and if $u_j(n) \leq 0$ for some j, n, then $\tilde{h}_j(n,\underline{\mathbf{u}}) \geq 0$.

Thus, in order to show that there exists a unique solution to (2.4)–(2.7) in $\overline{\mathcal{U}}^S$, it suffices to show that there exists a unique solution to (2.4$'$)–(2.7$'$) in $(\mathbb{R}^{\mathbb{Z}+})^S$. We use the Picard successive approximation method. From (2.7′) and the definition (2.3) of the norm, we have that, for all $\underline{\mathbf{u}}, \underline{\mathbf{u}}' \in (\mathbb{R}^{\mathbb{Z}_+})^S$,

$$\begin{aligned} \|\tilde{\underline{\mathbf{h}}}(\underline{\mathbf{u}})\| &\le \max_j \lambda_j + (|S|+1)\max_j \mu_j, \\ &:= C_1, \end{aligned} \tag{2.10}$$

and

$$\begin{aligned} \|\tilde{\underline{\mathbf{h}}}(\underline{\mathbf{u}}) - \tilde{\underline{\mathbf{h}}}(\underline{\mathbf{u}}')\| &\le (2m\max_j \lambda_j + (6+2m)\max_j \mu_j)\|\underline{\mathbf{u}} - \underline{\mathbf{u}}'\| \\ &:= C_2\|\underline{\mathbf{u}} - \underline{\mathbf{u}}'\|. \end{aligned} \tag{2.11}$$

Here we have used the facts that $|\xi(u) - \xi(u')| \le |u - u'|$, and, for $a_1, a_2, b_1, b_2 \in [0,1]$,

$$|a_1 b_1^m - a_2 b_2^m| \le |a_1 - a_2| + m|b_1 - b_2|.$$

For $t \ge 0$, let $\underline{\mathbf{u}}^{(0)}(t) = \underline{\mathbf{g}}$, and, recursively, let

$$\underline{\mathbf{u}}^{(r)}(t) = \underline{\mathbf{g}} + \int_0^t \tilde{\underline{\mathbf{h}}}(\underline{\mathbf{u}}^{(r-1)}(s))ds, \quad r \in \mathbb{N}. \tag{2.12}$$

It follows by induction that $\underline{\mathbf{u}}^{(r)}(t)$ is continuous in t on $[0,\infty)$ for all r, and that

$$\|\underline{\mathbf{u}}^{(r+1)}(t) - \underline{\mathbf{u}}^{(r)}(t)\| \le \frac{C_1 C_2^r t^{r+1}}{(r+1)!} \quad \text{for all } r, t.$$

Hence, for all $t \ge 0$, $\underline{\mathbf{u}}(s) = \lim_{r\to\infty} \underline{\mathbf{u}}^{(r)}(s)$ exists uniformly for $s \in [0,t]$. Since $\tilde{\underline{\mathbf{h}}}$ is uniformly continuous (see (2.11)), (2.12) then implies that

$$\underline{\mathbf{u}}(t) = \underline{\mathbf{g}} + \int_0^t \tilde{\underline{\mathbf{h}}}(\underline{\mathbf{u}}(s))ds, \quad t \ge 0,$$

and hence, by differentiating, that $\underline{\mathbf{u}}(t)$ is a solution to (2.4′)–(2.7′).

To show uniqueness, suppose that $\underline{\mathbf{w}}(t)$ is any solution of (2.4′)–(2.7′). Then

$$\underline{\mathbf{w}}(t) = \underline{\mathbf{g}} + \int_0^t \tilde{\underline{\mathbf{h}}}(\underline{\mathbf{w}}(s))ds, \tag{2.13}$$

$$\text{and so} \quad \underline{\mathbf{w}}(t) - \underline{\mathbf{u}}^{(r)}(t) = \int_0^t [\tilde{\underline{\mathbf{h}}}(\underline{\mathbf{w}}(s)) - \tilde{\underline{\mathbf{h}}}(\underline{\mathbf{u}}^{(r-1)}(s))]ds. \tag{2.14}$$

It follows by induction, as before, that

$$\|\underline{\mathbf{w}}(t) - \underline{\mathbf{u}}^{(r)}(t)\| \le \frac{C_1 C_2^r t^{r+1}}{(r+1)!},$$

so that $\underline{\mathbf{w}}(t) = \lim_{r\to\infty} \underline{\mathbf{u}}^{(r)}(t) = \underline{\mathbf{u}}(t)$. This completes the proof of Theorem 2.1.1(i). □

Before proving the rest of Theorem 2.1.1, we turn to the proof of Theorem 2.1.2, which provides a result which we will use in showing the convergence in Theorem 2.1.1(iii). The next two lemmas show the existence of the first and second derivatives of $\underline{\mathbf{u}}(t, \underline{\mathbf{g}})$ with respect to the entries in $\underline{\mathbf{g}}$, and give bounds on their size.

Lemma 2.2.1 *Consider the infinite system of differential equations*

$$\begin{aligned} z_x(0) &= c_x \\ \frac{dz_x(t)}{dt} &= \sum_{y\in\mathcal{C}} a_{x,y}(t)z_y(t) + b_x(t), \end{aligned}$$

for $x \in \mathcal{C}$, $t \geq 0$, *where* $\mathcal{C}$ *is a countable index set and where* $a_{x,y}$ *and* b_x *are continuous functions of* t *for all* $x, y \in \mathcal{C}$, *and suppose that* $\sum_{y\in\mathcal{C}} |a_{x,y}(t)| \leq a$, $|b_x(t)| \leq b_0 e^{bt}$, $|c_x| \leq c$ *for all* x, t. *Then there exists a solution* $\{z_x(t),\ x \in \mathcal{C},\ t \geq 0\}$ *which satisfies*

$$|z_x(t)| \leq ce^{at} + \frac{b_0}{b-a}(e^{bt} - e^{at}) \text{ for all } x, t. \tag{2.15}$$

Proof: We use a successive approximation argument similar to that in the proof of Theorem 2.1.1(i) above. Let

$$z_x^{(0)}(t) = c_x + \int_0^t b_x(s)ds,$$

and

$$z_x^{(r+1)}(t) = c_x + \int_0^t \left[\sum_{y\in\mathcal{C}} a_{x,y}(s)z_y^{(r)}(d) + b_x(s)\right] ds$$

for $x \in \mathcal{C}$ and for $r \in \mathbb{Z}_+$. Then one can show by induction that

$$|z_x^{(r+1)}(t) - z_x^{(r)}(t)| \leq \frac{a^{r+1}t^{r+1}(ac + b_0e^{bt})}{(r+1)!}$$

for $x \in \mathcal{C}$, $r \in \mathbb{Z}_+$. It follows that the limit $z_x(t) = \lim_{r\to\infty} z_x^{(r)}(t)$ exists uniformly in x and in t from bounded intervals, and so $\{z_x(t),\ x \in \mathcal{C},\ t \geq 0\}$ gives a solution to the system above.

Also

$$|z_x^{(0)}(t)| \leq c + \frac{b_0}{b}(e^{bt} - 1),$$

and another induction gives

$$|z_x^{(r+1)}(t) - z_x^{(r)}(t)| \leq a^{r+1}\left(\frac{ct^{r+1}}{(r+1)!} + \frac{b_0}{b^{r+2}}\sum_{l=r+2}^{\infty}\frac{(bt)^l}{l!}\right),$$

for $x \in \mathcal{C}$, $r \in \mathbb{Z}_+$. Summing over $r \geq 0$ gives the bound (2.15) as desired. □

Lemma 2.2.2 *The derivatives* $\dfrac{\partial \underline{\mathbf{u}}(t,\underline{\mathbf{g}})}{\partial g_j(n)}$, $\dfrac{\partial^2 \underline{\mathbf{u}}(t,\underline{\mathbf{g}})}{\partial g_j(n)^2}$ *and* $\dfrac{\partial^2 \underline{\mathbf{u}}(t,\underline{\mathbf{g}})}{\partial g_j(n)\partial g_{j'}(n')}$ *exist, for* $\underline{\mathbf{g}} \in \overline{\mathcal{U}}^S$ *and* $t \geq 0$, *and satisfy*

$$\left|\frac{\partial u_k(r,t,\underline{\mathbf{g}})}{\partial g_j(n)}\right| \leq e^{A_1 t} \tag{2.16}$$

$$\left|\frac{\partial^2 u_k(r,t,\underline{\mathbf{g}})}{\partial g_j(n)^2}\right|, \quad \left|\frac{\partial^2 u_k(r,t,\underline{\mathbf{g}})}{\partial g_j(n)\partial g_{j'}(n')}\right| \leq \frac{A_2}{A_1}(e^{2A_1 t} - e^{A_1 t}), \tag{2.17}$$

where

$$A_1 = 2(|S|m + |S| + 1)\max_j \mu_j + 2m \max_j \lambda_j,$$
$$A_2 = 2|S|(m(m-1)+2)\max_j \mu_j + 2m(m-1)\max_j \lambda_j.$$

Proof: Fix $j, n, \underline{\mathbf{g}}$, and write $\underline{\mathbf{u}}(t) = \underline{\mathbf{u}}(t, \underline{\mathbf{g}})$, the unique solution to (2.4)-(2.7) in $\overline{\mathcal{U}}^S$ given by Theorem 2.1.1. Suppose that $\underline{\mathbf{u}}'(t)$ satisfies $u'_k(0) = 0$ for all $k \in S$, and

$$\begin{aligned}\frac{du'_k(r)}{dt} = &\left[\lambda_k + \sum_{i\in S}\mu_i p_{ik} u_i(1,t)\right][m u_k(r-1,t)^{m-1}u'_k(r-1,t) + m u_k(r,t)^{m-1}u'_k(r,t)] \\ &+ \left[\sum_{i\in S}\mu_i p_{ik} u'_i(1,t)\right][u_k(r-1,t)^m - u_k(r,t)^m] - \mu_k[u'_k(r,t) - u'_k(r+1,t)]\end{aligned}$$

for all $k \in S$ and $r \geq 1$. Then, by differentiating (2.7), $\underline{\mathbf{u}}'(t)$ is the derivative $\partial\underline{\mathbf{u}}(t, \underline{\mathbf{g}})/\partial g_j(n)$, and so the system has at most one such solution $\underline{\mathbf{u}}'(t)$. Using the fact that $|u_k(r,t)| \leq 1$ for all k, r, t, and that $u_k(r,t)$ is continuous in t for all k and r, we may apply Lemma 2.2.1 with $a = A_1$, $b_0 = 0$, $c = 1$ and $\mathcal{C} = \mathbb{Z}_+ \times S$ to show that the solution $\underline{\mathbf{u}}'(t)$=$\partial\underline{\mathbf{u}}(t, \underline{\mathbf{g}})/\partial g_j(n)$ does exist, and satisfies the bound (2.16).

By differentiating the above system again with respect to $g_j(n)$ or $g_{j'}(n')$, one arrives at systems of equations for the required second partial derivatives, with coefficients now involving $\underline{\mathbf{u}}$ and $\underline{\mathbf{u}}'$, to which Lemma 2.2.1 can be applied with $a = A_1$, $b = 2A_1$, $b_0 = A_2$, $c = 0$ to give the bound (2.17). □

Proof of Theorem 2.1.2: Let L be the set of continuous functions $f : \overline{\mathcal{U}}^S \to \mathbb{R}$, and let D be the set of those $f \in L$ for which the derivatives

$$\frac{\partial f(\underline{\mathbf{g}})}{\partial g_j(n)}, \frac{\partial^2 f(\underline{\mathbf{g}})}{\partial g_j(n)^2} \quad \text{and} \quad \frac{\partial^2 f(\underline{\mathbf{g}})}{\partial g_j(n)\partial g_{j'}(n')}$$

exist for all $\underline{\mathbf{g}}, j, j', n, n'$ and are uniformly bounded in modulus by some constant $C = C(f) < \infty$. Observe that D is dense in L, using the topology induced by the norm (2.3) on $\overline{\mathcal{U}}^S$ and the sup norm on L. For $f \in D$,

$$N\left(f\left(\underline{\mathbf{g}} + \frac{\underline{\mathbf{e}}_j(n)}{N}\right) - f(\underline{\mathbf{g}})\right) \to \frac{\partial f(\underline{\mathbf{g}})}{\partial g_j(n)}$$
$$N\left(f\left(\underline{\mathbf{g}} - \frac{\underline{\mathbf{e}}_j(n)}{N}\right) - f(\underline{\mathbf{g}})\right) \to -\frac{\partial f(\underline{\mathbf{g}})}{\partial g_j(n)}$$

and

$$N\left(f\left(\underline{\mathbf{g}} + \frac{\underline{\mathbf{e}}_j(n)}{N} - \frac{\underline{\mathbf{e}}_k(n')}{N}\right) - f(\underline{\mathbf{g}})\right) \to \frac{\partial f(\underline{\mathbf{g}})}{\partial g_j(n)} - \frac{\partial f(\underline{\mathbf{g}})}{\partial g_k(n')}$$

uniformly in $\underline{\mathbf{g}}$ from $\overline{\mathcal{U}}^S$. Thus, using (2.1), we have

$$\begin{aligned} A_N f(\underline{\mathbf{g}}) \to & \sum_{n\geq 1}\sum_{j\in S}\lambda_j[g_j(n-1)^m - g_j(n)^m]\frac{\partial f(\underline{\mathbf{g}})}{\partial g_j(n)} \\ & - \sum_{n\geq 1}\sum_{j\in S}\mu_j p_j^*[g_j(n) - g_j(n+1)]\frac{\partial f(\underline{\mathbf{g}})}{\partial g_j(n)} \\ & + \sum_{n,n'\geq 1}\sum_{j,k\in S}\mu_k p_{kj}[g_k(n') - g_k(n'+1)][g_j(n-1)^m - g_j(n)^m] \\ & \quad \times \left[\frac{\partial f(\underline{\mathbf{g}})}{\partial g_j(n)} - \frac{\partial f(\underline{\mathbf{g}})}{\partial g_k(n')}\right], \end{aligned} \tag{2.18}$$

uniformly in $\underline{\mathbf{g}}$. The RHS of (2.18) may be rewritten as

$$\sum_{n\geq 1}\sum_{j\in S}\frac{\partial f(\underline{\mathbf{g}})}{\partial g_j(n)}\left\{\left[\lambda_j + \sum_{k\in S}\mu_k p_{kj} g_k(1)\right][g_j(n-1)^m - g_j(n)^m] - \mu_j[g_j(n) - g_j(n+1)]\right\}$$

which coincides with

$$\frac{d}{dt}f(\underline{\mathbf{u}}(t,\underline{\mathbf{g}}))\Big|_{t=0} \quad , \tag{2.19}$$

where $\underline{\mathbf{u}}(t,\underline{\mathbf{g}})$ is the solution to (2.4)–(2.5).

Setting $\mathbf{T}(t)f(\underline{\mathbf{g}}) = f(\underline{\mathbf{u}}(t,\underline{\mathbf{g}}))$ defines a semigroup of operators $\mathbf{T}(t), t \geq 0$ in L, corresponding to shifts along the solutions of (2.4)–(2.5). The generator $\mathbf{A}$ of this semigroup is given by (2.19). Thus we have

$$\lim_{N\to\infty}\mathbf{A}_N f = \mathbf{A}f \tag{2.20}$$

in the sup norm, for all $f \in D$.

Define $D_0 \subset D$ as the set of those functions in D which depend on only finitely many variables $g_j(n)$. By the definition of the norm (2.3), D_0 is dense in D, and hence in L. Also, it follows from Lemma 2.2.2 that $\mathbf{T}(f_0) \in D$ for all $f_0 \in D_0$. Finally, observe that the semigroups $\mathbf{T}_N(t)$ and $\mathbf{T}(t)$ are continuous and contracting in L. These facts, together with (2.20) and with Proposition 3.3 and Theorem 6.1 from Chapter 1 of [13], give the result of Theorem 2.1.2. □

We write $\underline{\mathbf{g}} \leq \underline{\mathbf{g}}'$ to mean that all the inequalities $g_j(n) \leq g_j'(n)$ hold; similarly $\min[\underline{\mathbf{g}}, \underline{\mathbf{g}}']$ and $\max[\underline{\mathbf{g}}, \underline{\mathbf{g}}']$ are defined componentwise.

Lemma 2.2.3 *If $\underline{\mathbf{g}} \leq \underline{\mathbf{g}}'$, where $\underline{\mathbf{g}}, \underline{\mathbf{g}}' \in \overline{\mathcal{U}}^S$, then $\underline{\mathbf{u}}(t,\underline{\mathbf{g}}) \leq \underline{\mathbf{u}}(t,\underline{\mathbf{g}}')$ for all t.*

Proof: Define the coordinate functions $f_{j,n}(\underline{\mathbf{g}}) = g_j(n), n \in \mathbb{Z}_+, j \in S$. Fix N, and consider two networks with initial states $\underline{\mathbf{g}}_N, \underline{\mathbf{g}}_N' \in \mathcal{U}_N^S$, $\underline{\mathbf{g}}_N \leq \underline{\mathbf{g}}_N'$. A simple coupling shows that

$$\mathbf{T}_N(t)f_{j,n}(\underline{\mathbf{g}}_N) \leq \mathbf{T}_N(t)f_{j,n}(\underline{\mathbf{g}}_N'), \quad \text{for all} \quad t \geq 0, j, n. \tag{2.21}$$

Now for given $\underline{\mathbf{g}}, \underline{\mathbf{g}}' \in \overline{\mathcal{U}}^S$ with $\underline{\mathbf{g}} \leq \underline{\mathbf{g}}'$, choose sequences $\{\underline{\mathbf{g}}_N\}, \{\underline{\mathbf{g}}'_N\}, N \in \mathbb{N}$ with $\underline{\mathbf{g}}_N, \underline{\mathbf{g}}'_N \in \mathcal{U}_N^S$, such that $\underline{\mathbf{g}}_N \leq \underline{\mathbf{g}}'_N$ for all N and $\underline{\mathbf{g}}_N \to \underline{\mathbf{g}}$, $\underline{\mathbf{g}}'_N \to \underline{\mathbf{g}}'$ as $N \to \infty$. Then it follows from Theorem 2.1.2 and (2.21) that $f_{j,n}(\underline{\mathbf{u}}(t, \underline{\mathbf{g}})) \leq f_{j,n}(\underline{\mathbf{u}}(t, \underline{\mathbf{g}}'))$ for all t, j, n, which is exactly what we need. □

For $\underline{\mathbf{g}} \in \mathcal{U}^S$, define the vector $\underline{\mathbf{v}}(\underline{\mathbf{g}}) = (v_j(n, \underline{\mathbf{g}}), n \in \mathbb{Z}_+, j \in S)$ by $v_j(n, \underline{\mathbf{g}}) = \sum_{n' \geq n} g_j(n')$. Observe that $v_j(1, \underline{\mathbf{g}})$ represents the average queue length of the channels at node j in the state $\underline{\mathbf{g}}$, which is guaranteed to be finite by the definition of $\mathcal{U}$ in (2.2); also note that $v_j(n, \underline{\mathbf{g}}) \geq v_j(n+1, \underline{\mathbf{g}})$ for all j, n. In addition, if $\underline{\mathbf{u}}(t, \underline{\mathbf{g}})$ is the solution of the system (2.4)–(2.7), write $\underline{\mathbf{v}}(t, g) = \underline{\mathbf{v}}(\underline{\mathbf{u}}(t, \underline{\mathbf{g}}))$. Define $\mathbf{x} \cdot \mathbf{y}$ by $(\mathbf{x} \cdot \mathbf{y})_j = x_j y_j$ for $j \in S$.

Lemma 2.2.4 *If $\underline{\mathbf{g}} \in \mathcal{U}^S$ then $\underline{\mathbf{u}}(t, \underline{\mathbf{g}}) \in \mathcal{U}^S$ for all $t \geq 0$, and*

$$\dot{v}_j(n, t, \underline{\mathbf{g}}) = \left[\lambda_j + \sum_{k \in S} \mu_k p_{kj} u_k(1, t, \underline{\mathbf{g}})\right] u_j(n-1, t, \underline{\mathbf{g}})^m - \mu_j u_j(n, t, \underline{\mathbf{g}}).$$

In particular,

$$\dot{v}_j(1, t, \underline{\mathbf{g}}) = \lambda_j + \sum_{k \in S} \mu_k p_{kj} u_k(1, t, \underline{\mathbf{g}}) - \mu_j u_j(1, t, \underline{\mathbf{g}})$$

or, in vector notation,

$$\dot{\mathbf{v}}(1, t, \underline{\mathbf{g}})(\mathbf{I} - \mathbf{P})^{-1} = \boldsymbol{\rho} - \boldsymbol{\mu} \cdot \mathbf{u}(1, t, \underline{\mathbf{g}}). \tag{2.22}$$

Proof: Summing the RHS of (2.7) over $n' \geq n$ gives the result. □

Proof of Theorem 2.1.1, (ii) and (iii): For a fixed point $\underline{\mathbf{a}}$ of the map $\underline{\mathbf{u}}$, we need $\underline{\mathbf{h}}(\underline{\mathbf{a}}) = 0$ in (2.6) and (2.7). Restricting ourselves to the space $\mathcal{U}^S$, we have $a_j(0) = 1$ for all j, and can use (2.22), setting the LHS to 0, to give $a_j(1) = \rho_j/\mu_j$ for all $j \in S$. Putting these values back into the RHS of (2.7) gives

$$a_j(n+1) = a_j(n) - \frac{\rho_j}{\mu_j}[a_j(n-1)^m - a_j(n)^m],$$

which can be solved recursively to confirm the unique solution $\underline{\mathbf{a}}$ given in (2.8).

To prove the convergence in (2.9), it will be sufficient to show that the conclusion $\underline{\mathbf{u}}(t, \underline{\mathbf{g}}) \to \underline{\mathbf{a}}$ holds for all those $\underline{\mathbf{g}} \in \mathcal{U}^S$ for which either $\underline{\mathbf{g}} \leq \underline{\mathbf{a}}$ or $\underline{\mathbf{g}} \geq \underline{\mathbf{a}}$, since Lemma 2.2.3 implies that

$$\underline{\mathbf{u}}\big(t, \min[\underline{\mathbf{g}}, \underline{\mathbf{a}}]\big) \leq \underline{\mathbf{u}}(t, \underline{\mathbf{g}}) \leq \underline{\mathbf{u}}\big(t, \max[\underline{\mathbf{g}}, \underline{\mathbf{a}}]\big) \quad \text{for all } \underline{\mathbf{g}} \in \overline{\mathcal{U}}^S, t \geq 0.$$

First, we need to check that for such a $\underline{\mathbf{g}}$, the quantities $v_j(1, t, \underline{\mathbf{g}})$ (and hence also $v_j(n, t, \underline{\mathbf{g}})$, $n > 1$) remain bounded uniformly in t. If $\underline{\mathbf{g}} \leq \underline{\mathbf{a}}$ then, by Lemma 2.2.3, $\underline{\mathbf{u}}(t, \underline{\mathbf{g}}) \leq \underline{\mathbf{a}}$, and so $\underline{\mathbf{v}}(1, t, \underline{\mathbf{g}}) \leq \underline{\mathbf{v}}(1, \underline{\mathbf{a}})$, for all t.

On the other hand, if $\underline{\mathbf{g}} \geq \underline{\mathbf{a}}$, then by the same lemma, $\underline{\mathbf{u}}(t, \underline{\mathbf{g}}) \geq \underline{\mathbf{a}}$. In particular, $u_j(1, t, \underline{\mathbf{g}}) \geq (\rho_j/\mu_j)$ for all j, or, in vector notation, $\boldsymbol{\mu} \cdot \mathbf{u}(1, t, \underline{\mathbf{g}}) \geq \boldsymbol{\rho}$. Returning to (2.22) yields

$$\dot{\underline{v}}(1, t, \underline{\mathbf{g}})(\mathbf{I} - \mathbf{P})^{-1} \leq \underline{0}. \tag{2.23}$$

Since $(\mathbf{I}-\mathbf{P})^{-1}$ has diagonal entries ≥ 1 and all other entries ≥ 0, (2.23) implies that the entries of vectors $\mathbf{v}(1,t,\underline{\mathbf{g}})(\mathbf{I}-\mathbf{P})^{-1}$ and $\mathbf{v}(1,t,\underline{\mathbf{g}})$ remain bounded uniformly in t.

The convergence $\underline{\mathbf{u}}(t,\underline{\mathbf{g}}) \to \underline{\mathbf{a}}$ will follow from

$$\int_0^\infty \left[u_j(n,s,\underline{\mathbf{g}}) - a_j(n)\right] ds < \infty, \quad j \in S, \quad n \in \mathbb{Z}_+ \tag{2.24}$$

in the case $\underline{\mathbf{g}} \geq \underline{\mathbf{a}}$, and from

$$\int_0^\infty \left[a_j(n) - u_j(n,s,\underline{\mathbf{g}})\right] ds < \infty, \quad j \in S, \quad n \in \mathbb{Z}_+$$

in the case $\underline{\mathbf{g}} \leq \underline{\mathbf{a}}$; the integrands are non-negative for all s in each case. The two bounds may be proved similarly, and we discuss, say, (2.24). We use induction in n, starting with $n = 1$. Using (2.22),

$$\begin{aligned}\int_0^t \left[u_j(1,s,\underline{\mathbf{g}}) - a_j(1)\right] ds &= \frac{1}{\mu_j}\int_0^t ds[\boldsymbol{\mu}\cdot\mathbf{u}(1,s,\underline{\mathbf{g}}) - \boldsymbol{\rho}]_j \\ &= \frac{1}{\mu_j}\int_0^t ds\left[-\dot{\mathbf{v}}(1,s,\underline{\mathbf{g}})(\mathbf{I}-\mathbf{P})^{-1}\right]_j \\ &= \frac{1}{\mu_j}\left[(\mathbf{v}(1,0,\underline{\mathbf{g}}) - \mathbf{v}(1,t,\underline{\mathbf{g}}))(\mathbf{I}-\mathbf{P})^{-1}\right]_j\end{aligned}$$

The RHS remains bounded as $t \to \infty$, so the integral on the LHS converges.

Now assume that the integral (2.24) converges for all $n \leq L-1$. Using Lemma 2.2.4 and the relation

$$\mu_j a_j(L) = \left[\lambda_j + \sum_{k\in S}\mu_k p_{kj} a_k(1)\right] a_j(L-1)^m,$$

we have

$$\begin{aligned} v_j(L,0,\underline{\mathbf{g}}) - v_j(L,t,\underline{\mathbf{g}}) &= -\int_0^t \dot{v}_j(L,s,\underline{\mathbf{g}})ds \\ &= \int_0^t \left[\mu_j u_j(L,s,\underline{\mathbf{g}}) - \left(\lambda_j + \sum_{k\in S}\mu_k p_{kj} u_k(1,s,\underline{\mathbf{g}})\right) u_j(L-1,s,\underline{\mathbf{g}})^m\right] ds \\ &= \mu_j \int_0^t [u_j(L,s,\underline{\mathbf{g}}) - a_j(L)]ds - \lambda_j \int_0^t [u_j(L-1,s,\underline{\mathbf{g}})^m - a_j(L-1)^m]ds \\ &\quad - \sum_{k\in S}\mu_k p_{kj}\int_0^t [u_k(1,s,\underline{\mathbf{g}})u_j(L-1,s,\underline{\mathbf{g}})^m - a_k(1)a_j(L-1)^m]ds \end{aligned}$$

By the induction hypothesis, the last two integrals converge as $t \to \infty$. The LHS also remains bounded, so we have $\int_0^\infty [u_j(L,s,\underline{\mathbf{g}}) - a_j(L)]ds < \infty$ as required.

The second statement in part (iii) of the theorem follows immediately from the convergence properties in the space $\mathcal{U}^S$ just established. □

Remark: There are many other invariant distributions which are not concentrated on $\mathcal{U}^S$. In fact, for all $\boldsymbol{\epsilon}$ in a neighbourhood of $\mathbf{0}$ in $[0,1]^J$, there exists a fixed point $\underline{\mathbf{a}}^{(\boldsymbol{\epsilon})}$ of the maps

$\underline{\mathbf{u}} : \underline{\mathbf{g}} \to \underline{\mathbf{u}}(t, \underline{\mathbf{g}})$ with $\lim_{n\to\infty} a_j^{(\epsilon)}(n) = \epsilon_j$ for all j, and if $\underline{\mathbf{g}} \in \overline{\mathcal{U}}^S$ satisfies $\lim_{n\to\infty} g_j(n) = \epsilon_j$ for all j, then $\underline{\mathbf{u}}(t, \underline{\mathbf{g}}) \to \underline{\mathbf{a}}^{(\epsilon)}$ as $t \to \infty$. This corresponds to a situation where a proportion ϵ_j of the channels at node j are considered "saturated", i.e. their queues are infinitely long and remain so always, irrespective of arrivals or departures. (Of course $\underline{\mathbf{a}}^{(\mathbf{1})}$, given by $a_j^{(\mathbf{1})}(n) = 1$ for all j, n, is also a fixed point; here every channel in the network is saturated). In fact one may show that $\underline{\mathbf{a}}^{(\epsilon)} \to \underline{\mathbf{a}}$ as $\boldsymbol{\epsilon} \to \mathbf{0}$; then, by using the fact that any state with no saturated channels may be majorised by a state with an arbitrarily small proportion of saturated channels at each node and with finite expected average queue length at the other channels, one can extend the results of Theorem 2.1.1 (ii) and (iii), on the existence and uniqueness of an attractive fixed point, so as to apply not only to to the space $\mathcal{U}^S$ (representing states with finite average queue length at each node) but also to the larger space representing all network states in which no channel is saturated.

2.3 Convergence of Stationary Distributions

We first prove Theorem 2.1.3, from which we can deduce the existence and various properties of the stationary distributions of the Markov processes $\underline{\mathbf{r}}_N(t)$ under the non-overload condition (1.7).

Proof of Theorem 2.1.3: The argument is an extension of that used by Turner to prove Theorem 4 in [50]. For a network B, and $r \geq 1$, $1 \leq k \leq J$, let $A_{k,r}(B)$ be the time of the rth arrival after time 0 (from outside or from within the network) at node k, and $D_{k,r}(B)$ the time of the rth departure from node k.

We will construct a coupling such that for all k, r

$$\begin{aligned} &\text{(i)} \quad D_{k,r}(R) \leq D_{k,r}(Q), \\ &\text{(ii)} \quad A_{k,r}(R) \leq A_{k,r}(Q). \end{aligned} \tag{2.25}$$

As well as the real tasks in Q, we invent "shadow" tasks, whose behaviour links the networks Q and R. For each k and r, a shadow task arrives at node k in Q at time $A_{k,r}(R)$, and remains until time $A_{k,r}(Q)$, when it is replaced by the arriving real task. This definition makes sense if (2.25)(ii) holds. The shadow task inhabits the same channel as the real task which replaces it. We will write $\overline{Q}$ to mean the network Q including shadows as well as real tasks; thus a node in the network $\overline{Q}$ experiences arrivals at the same time as the corresponding node in R, and departures at the same time as that in Q.

For a network B, and $x \in \mathbb{Z}_+$, define

$$\psi_{k,x}(B, t) = \sum_{n=1}^{N} [l_{k,n}(B, t) - x]_+,$$

where $l_{k,n}(B, t)$ is the queue length of the nth channel at node k in T at time t.

In order to show (2.25), we will also establish that under the coupling

$$\psi_{k,x}(R,t) \leq \psi_{k,x}(\overline{Q},t) \quad \text{for all } k \in S, x \in \mathbb{Z}_+, t \geq 0. \tag{2.26}$$

The following observation from [50] will be used here. For a given network B, the difference between $\psi_{k,x}(B,t)$ and $\psi_{k,x+1}(B,t)$ is equal to the number of queues at node k in B with length greater than x at time t. Thus if $\psi_{k,y}(R,t) \leq \psi_{k,y}(\overline{Q},t)$ for all y, and $\psi_{k,x}(R,t) = \psi_{k,x}(\overline{Q},t)$, then it follows from the inequality (2.26) at $x+1$ and $x-1$ that

$$\#\{n : l_{k,n}(R,t) \leq x\} \leq \#\{n : l_{k,n}(\overline{Q},t) \leq x\} \tag{2.27}$$

and

$$\#\{n : l_{k,n}(R,t) \geq x\} \leq \#\{n : l_{k,n}(\overline{Q},t) \geq x\}. \tag{2.28}$$

The coupling works as follows. For each $k \in S$, generate a Poisson process of rate $N\lambda_k$ as the external arrival stream at node k, and one of rate $N\mu_k$ to represent potential departures from node k. These are the same in both systems. Also, the rth task to depart from node k in Q will have the same destination as the rth task to depart from node k in R (generated i.i.d. according to the probabilities $p_{k,j}$ and p^*_k). This ensures that if (2.25)(i) holds for the departure process up to a given time, then so does (2.25)(ii) for the arrival process up to that time. We now need to describe the way in which channels within nodes are chosen for arrivals and departures, verifying that departures preserve (2.25) and that all network events preserve (2.26). (Certainly (2.26) holds at $t = 0$ since the initial states of $\overline{Q}$ and R are the same).

At a potential departure time at node k, arrange the channels at k in $\overline{Q}$ and in R in order of queue lengths (including shadow tasks), and let a departure occur from the corresponding queue in each system; for example, if it occurs from the longest queue in $\overline{Q}$, let it occur from the longest queue in R also. Departures from an empty queue, or from one containing only shadow tasks, are lost. Now let D be a departure point, and suppose that properties (2.25) and (2.26) hold for $t < D$; we wish to show that they are preserved. Firstly, the only way in which (2.26) could be violated by a departure point at node k would be if $\psi_{k,x}(R,D-) = \psi_{k,x}(\overline{Q},D-)$ for some x, and if queues of length a in $\overline{Q}$ and length b in R were chosen for departure at time D, with $0 \leq b \leq x < a$. But this is impossible, by (2.27) and the coupling method used. Secondly, for (2.25)(i) to be violated, the number of previous departures from node k must be the same in R as in Q (and hence also $\overline{Q}$). Then, since arrivals in R and $\overline{Q}$ occur at the same times, there must be the same total number of tasks at node k in R and in $\overline{Q}$; i. e. $\psi_{k,0}(R) = \psi_{k,0}(\overline{Q})$. But $\psi_{k,1}(R) \leq \psi_{k,1}(\overline{Q})$ by assumption, so, by (2.27) again, there are at least as many empty queues in $\overline{Q}$ as in R. Thus if a departure occurs from $\overline{Q}$, the coupling ensures that one must also occur from R, preserving (2.25)(i) and hence also (2.25)(ii).

At an arrival time $A_{k,r}(R)$, generated either by an external arrival or by a departure within the network, order the channels in $\overline{Q}$ and in R by queue length as before. Choose $m(R)$ channels at random and send the arrival in R to the one with the shortest queue. Choose one of the corresponding $m(R)$ channels in $\overline{Q}$ at random and send the shadow task to this channel. The effect is that if an arrival occurs to the nth longest queue in R, the corresponding arrival in $\overline{Q}$ must occur to a queue at least as the long as the nth longest there. Again, suppose (2.26) has held up to this point - how might it now fail? If the chosen queue in $\overline{Q}$ has length a, and that in R length b, then it would require $\psi_{k,x}(R) = \psi_{k,x}(\overline{Q})$ to hold for some x before the arrivals, with $a < x \leq b$. Then, following (2.28), we would have

$$\#\{n : l_{k,n}(R, A-) \geq x\} \leq \#\{n : l_{k,n}(\overline{Q}, A-) \geq x\};$$

but this is impossible, since $a < x \leq b$ and, according to our coupling scheme, we must have

$$\#\{n : l_{k,n}(R, A-) \geq b\} \geq \#\{n : l_{k,n}(\overline{Q}, A-) \geq a\},$$

(since a queue with length a in $\overline{Q}$ is at least as long as one whose ranking corresponds to a queue with length b in R), and $\#\{n : l_{k,n}(\overline{Q}, A-) = a\} \geq 1$ (a queue of length a must exist in $\overline{Q}$ since a queue of length a has been chosen there for the arrival).

Finally, at an arrival time $A_{k,r}(Q)$, the shadow task which arrived at time $A_{k,r}(R) \leq A_{k,r}(Q)$ is replaced by a real task in the same channel. This does not affect the queue lengths in $\overline{Q}$ so (2.26) is preserved.

Note that in fact no time elapses between a departure and an arrival which it generates at the same or another node; we have separated the two events in the above description, but if the two separately do not violate the properties (2.25) and (2.26) then nor will the two combined. Similarly, it may be that $A_{k,r}(R) = A_{k,r}(Q)$ for some k and r; in this case the shadow task arriving at this time is never seen since it is instantaneously erased by the arriving real task.

Under the coupling described, arrivals in the network from outside occur at the same time in R as in Q, and departures from the network to the outside occur no later in R than in Q; hence the total number of tasks in R at a given time is no greater than that in Q. □

Remark: We have not resolved the question of whether the conclusion of Theorem 2.1.3, or a similar one, holds also in the case $1 < m(Q) < m(R)$. The complication introduced by the network structure is that if one part of the network works more "slowly", this can lead to delayed arrivals in another part; and these delayed arrivals may, in the light of more recent information, be able to make a better choice about which queue to join than if they had arrived earlier. Counterexamples to the result can certainly be constructed if the parameter m is allowed to vary between nodes, even though $m_j(Q) \leq m_j(R)$ for all $j \in S$. Related questions concerning the optimal assignment of customers to servers in the case of multi-server stations in tandem are discussed by Hordijk and Koole in [20].

Proof of Theorem 2.1.4:

(i) If $m = 1$, then the network described by $\underline{\mathbf{r}}_N(t)$ is equivalent to a standard Jackson network, with $N|S|$ nodes (j,n), $j \in S$, $1 \leq n \leq N$, arrival and service rates $\lambda_{j,n} = \lambda_j$ and $\mu_{j,n} = \mu_j$, and routing matrix $p_{(j,n)(j',n')} = p_{jj'}/N$. Then $\rho_{(j,n)} = \rho_j$ also, and under the condition (1.7) the network is positive recurrent, with an invariant distribution analogous to (1.8).

For $m > 1$, Theorem 2.1.3 allows us to couple the network with a "slower" one for which $m = 1$ in such a way that the faster network is empty whenever the slower one is empty. Then, since the slower network is positive recurrent, the faster one must be also.

(ii) Since $\overline{\mathcal{U}}^S$ is compact, so is the set $\mathcal{P}(\overline{\mathcal{U}}^S)$ of probability measures on $\overline{\mathcal{U}}^S$. Hence the sequence $\{\pi_N, N \in \mathbb{N}\}$ has limit points. We wish to check that any limit point is the Dirac delta-measure concentrated at $\underline{\mathbf{a}}$.

From Lemma 2.2.2, the maps $\underline{\mathbf{g}} \mapsto \underline{\mathbf{u}}(t, \underline{\mathbf{g}}), t \geq 0, \underline{\mathbf{g}} \in \overline{\mathcal{U}}^S$ are continuous in $\underline{\mathbf{g}}$, and it follows from Theorem 2.1.2 that any limit point π of $\{\pi_N\}$ must be an invariant distribution for these maps. Hence, by Theorem 2.1.1(iii), it is sufficient to show that π is concentrated on $\mathcal{U}^S$, i.e. $\pi(\{\underline{\mathbf{g}} : v_j(1, \underline{\mathbf{g}}) < \infty \; \forall j\}) = 1$. We show in fact that $\sum_{j \in S} \mathbb{E}_\pi v_j(1) < \infty$. If $m = 1$, $\mathbb{E}_{\pi_N} r_j(n) = (\rho_j/\mu_j)^n$ for all N, by the argument in the proof of part (i) and the expression (1.8), so that $\sum_{j \in S} \mathbb{E}_{\pi_N} v_j(1) = \sum_{j \in S} \rho_j/(\mu_j - \rho_j)$ for all N. Hence, by Theorem 2.1.3 and standard results on the convergence of Markov processes to equilibrium, $\sum_{j \in S} \mathbb{E}_{\pi_N} v_j(1) \leq \sum_{j \in S} \rho_j/(\mu_j - \rho_j)$ for all N and any m, giving $\sum_{j \in S} \mathbb{E}_\pi v_j(1) \leq \sum_{j \in S} \rho_j/(\mu_j - \rho_j)$ also, as required. $\square$

2.4 Related Work

Vvedenskaya, Dobrushin and Karpelevich, in [51], consider a model which corresponds to a single node of our network. The methods involving the convergence of generators used here to prove Theorem 2.1.2 are also used there to prove their Theorem 2, which is similar in form. One may compare the approach of [51] to that of Turner in [50], who considers a very similar model. There a functional law of large numbers (FLLN) is proved for the convergence of the processes $\underline{\mathbf{r}}_N$ to the limiting trajectory in $D_{\overline{\mathcal{U}}}[0, \infty)$, the space of paths in $\overline{\mathcal{U}}$ which are right-continuous with left limits, endowed with the Skorohod topology. (See also [48], where some flaws in the argument of [50], arising from the fact that the state space on which the network processes live is not naturally compact, are rectified). The FLLN result is not formally equivalent to a result of the form of our Theorem 2.1.2 or of Theorem 2 in [51], but the two may loosely be said to be similar in strength; the convergence under the FLLN is in the stronger sense of pathwise convergence rather than that of uniform convergence of single time-point distributions, but is not obtained uniformly over all initial conditions as in Theorem 2.1.2. It seems very likely that the FLLN result could also be extended to the network model we consider here.

Similar FLLNs are obtained for the asymptotic analysis of large loss networks and queueing networks with a variety of structures and routing regimes by Hunt and others, in [10], [15], [22], [23] and [56], for example. The techniques used involve the application of a martingale convergence theorem of the sort given in [13], p.339, once relative compactness of the sequence of processes in the path space has been established. Our methods involving convergence of generators could also be used in most of these situations, although not perhaps so easily for models such as those considered in [22] or [47], where the coefficients of the limiting systems of differential equations obtained are not continuous at the boundary of the state space. In [22], a limit for a loss network with dynamic routing is considered under which the set of links remains fixed while the arrival rates and link capacities become large with their ratio held constant. The routing rules, expressed in terms of the number of free circuits at each link, remain fixed. In the limit one observes a *separation of time-scales* (see e.g. [34]) between the free-circuit process and the process describing the average occupancy levels of the circuits at each link.

The FLLNs of the sort mentioned above do not however yield convergence of equilibrium distributions of the type described by Theorem 2.1.4 here. So far this inversion of limits has seemed to require approaches more tailored to each particular model. Recent work by Graham [18] strengthens Theorem 5 of [51] (which has a similar form to Theorem 2.1.4 here) by proving a FLLN for the convergence of the paths of processes in equilibrium. In [21] (see also [17]), a fully symmetric star-shaped loss network is considered, and a mean-field limit is taken under which the number of nodes increases to infinity while the capacity of the links remains constant; there a functional central limit theorem is obtained for the convergence of the paths of the processes in equilibrium.

Chapter 3

Point Processes in Fast Jackson Networks

3.1 Introduction

One of the attractions of the model of a standard Jackson network is its partial exact solvability expressed by the product form of the stationary distribution for the network; in equilibrium (if the network is not overloaded), the lengths of the queues at the various nodes in the network at a given point in time are independent, and are each distributed as for an equilibrium $M/M/1$ queue with appropriately chosen arrival and service rates. However, this simplicity of description for single points in time does not extend to the pathwise behaviour of the network. The arrivals from outside the network to the various nodes form a family of independent Poisson processes, and so by a reversibility argument (see for example [31]) one can show that in equilibrium the same is true of the processes of departures leaving the network from the various nodes. But the process of all arrivals, or of all departures, at a given node is not in general easy to describe, and of course the queue-length processes at different nodes of the network are not independent. Similarly, the joint distribution of the waiting times experienced by a particular task at the various nodes it visits in the network is not easy to obtain.

In this chapter we again consider the "Fast Jackson Network" model introduced in the previous chapter, and once more take the limit $N \to \infty$; but instead of concentrating on the overall distribution of queue-lengths at single time-points, in the context of which any individual channel is unimportant, we now investigate how the paths of the queue-length processes at individual channels behave as the size of the network grows. We show that, if the initial state of the network converges suitably, the distribution of the queue length processes at a fixed collection of channels at the same or different nodes converges in "local variation distance" as $N \to \infty$, and that under the limit the component processes are independent. We describe the limiting processes which arise, and analyse them in particular in the case where the networks are positive recurrent and are started in equilibrium. Then for $m = 1$ (in which case the networks are standard Jackson networks for finite N), the limiting point processes of arrivals and departures are Poisson processes, and we examine how they and the relationship between them change as m increases. For networks in equilibrium, another interpretation of the decoupling which occurs in the limit is that a typical task, given its route through the network, experiences a sequence of independent waiting times.

In the next section we state and prove the main theorem, using results of Kabanov and Liptser from [26] which relate convergence in variation distance of multivariate point processes to the convergence of their compensators. In Section 3.3 we interpret this result for the case of networks in equilibrium, and analyse and illustrate the particular point processes that arise for various different values of the network parameters. Finally in Section 3.4 we discuss some possible extensions of the results; in particular we compare our approach with that of Brown and Pollett [8], who investigate how the distance of arrival processes in a standard Jackson network from appropriate Poisson processes varies as the number of nodes in the network is increased.

3.2 Convergence as $N \to \infty$

We fix S and the parameters $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$, $\mathbf{P}$ and m, and consider a sequence of networks indexed by $N = 1, 2, \ldots$, with the Nth network having N channels at each node. Fix a set of K tagged channels among the various nodes, and let $i(k)$ be the node to which the kth tagged channel belongs, $1 \leq k \leq K$. We will analyse the behaviour of the process of queue lengths at the tagged channels as $N \to \infty$.

We now describe the evolution of the network by a process $(\underline{\mathbf{r}}(t), \mathbf{x}(t))$, $t \geq 0$, with state space $\overline{\mathcal{U}}^S \times \mathbb{Z}_+^K$. Here $x_k(t)$ is the length of the queue in the kth tagged channel at time t, and $r_j(n,t)$ is the proportion of channels (including tagged channels) at node j whose queue length is at least n at time t. Thus we do not distinguish between different untagged channels at the same node. Where relevant, the topology we will use on $\overline{\mathcal{U}}^S \times \mathbb{Z}_+^K$ is the product of the topology induced by the norm (2.3) on $\overline{\mathcal{U}}^S$ and the discrete topology on $\mathbb{Z}_+^K$.

Let Ω be the space of paths $[0,\infty) \to \overline{\mathcal{U}}^S \times \mathbb{Z}_+^K$ which are right continuous with left limits, representing the paths of $(\underline{\mathbf{r}}(t), \mathbf{x}(t))$. Define the filtration $G = \{\mathcal{G}_t\}_{t\geq 0}$ by $\mathcal{G}_t = \sigma(\{\underline{\mathbf{r}}(s), \mathbf{x}(s)\},\ 0 \leq s \leq t)$, and put $\mathcal{G} = \bigvee_t \mathcal{G}_t$. For each N, let ψ_N be a distribution on $\overline{\mathcal{U}}^S \times \mathbb{Z}_+^K$, representing the distribution of $(\underline{\mathbf{r}}(0), \mathbf{x}(0))$ for the Nth network. This, together with the dynamics of the Nth network described earlier, yields a probability measure $\bar{\mathbb{P}}_N$ on $(\Omega, \mathcal{G})$ describing the behaviour of the Nth network; the process $(\underline{\mathbf{r}}(t), \mathbf{x}(t))$ is adapted to the filtration G. We will write $\mathbb{E}_N$ for the expectation with respect to $\bar{\mathbb{P}}_N$. Note that $\bar{\mathbb{P}}_N(\underline{\mathbf{r}}(t) \in \overline{\mathcal{U}}_N^S\ \forall t) = 1$.

Ultimately we will be particularly interested in the smaller filtration $F = \{\mathcal{F}_t\}_{t\geq 0}$, where

$\mathcal{F}_t = \sigma(\mathbf{x}(s),\ 0 \le s \le t)$, and $\mathcal{F} = \bigvee_t \mathcal{F}_t$. Clearly $\mathcal{F}_t \subset \mathcal{G}_t$ for all t and $\mathcal{F} \subset \mathcal{G}$, and the process $\mathbf{x}(t)$ is adapted to the filtration F. Let $\mathbb{P}_N$ be the restriction of $\bar{\mathbb{P}}_N$ to $(\Omega, \mathcal{F})$.

For two measures $\mathbb{P}$ and $\mathbb{P}'$ on $(\Omega, \mathcal{F})$, we will write $\mathrm{Var}_t(\mathbb{P}, \mathbb{P}')$ for the variation distance between $\mathbb{P}$ and $\mathbb{P}'$ restricted to $\mathcal{F}_t$:

$$\mathrm{Var}_t(\mathbb{P}, \mathbb{P}') = \sup_{A \in \mathcal{F}_t} |\mathbb{P}(A) - \mathbb{P}'(A)|.$$

The following theorem states that, if the initial conditions converge suitably, then the processes $\mathbf{x}$ governed by $\mathbb{P}_N$ converge in this local variation distance for each t.

Theorem 3.2.1 *Suppose that $\psi_N \to \psi$ weakly as $N \to \infty$, where ψ is a distribution on $\overline{\mathcal{U}}^S \times \mathbb{Z}_+^K$ under which the marginal distribution of $\underline{\mathbf{r}}(0)$ on $\overline{\mathcal{U}}^S$ is concentrated at a single point. Then:*

(i) There exists a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ such that, for all $t \ge 0$, $\mathrm{Var}_t(\mathbb{P}_N, \mathbb{P}) \to 0$ as $N \to \infty$.

(ii) Under the limiting measure $\mathbb{P}$, $\{\mathbf{x}(t), t \ge 0\}$ is a Markov process (not in general time-homogeneous) and if $x_1(0), \ldots, x_K(0)$ are independent under ψ, then the component processes $\{x_k(t), t \ge 0\}$, $1 \le k \le K$, are independent under $\mathbb{P}$.

Proof: Given the initial state $\mathbf{x}(0)$, we may represent the process $\mathbf{x}(t)$ by a multivariate point process $\boldsymbol{\xi}$ with $K^2 + 2K$ components. Each component is an increasing integer-valued process, which has value 0 at time 0, and whose value at time t is the number of *points* which occur in that component in the time interval $(0, t]$. For $1 \le k \le K$, let the points of the component process $\xi_{0,k}(t)$ record times of arrivals at the kth tagged channel which do not originate from a tagged channel (so are external arrivals or tasks transferring from an untagged channel at the same or another node), and let those of $\xi_{k,0}(t)$ record times of departures from the kth tagged channel which do not proceed to another tagged channel. Finally, for $1 \le k, l \le K$, let the points of $\xi_{k,l}(t)$ record times of transfers from the kth tagged channel to the lth (which is the same if $k = l$). Note that with probability 1, for all N, no two points occur simultaneously, in the same or different components.

For each N, we associate with the point process $\boldsymbol{\xi}$ and the filtered probability space $(\Omega, \mathbb{P}_N, \mathcal{F}, F)$ to which it is adapted the multivariate *compensator* $\mathbf{B}^{(N)}$ (indexed in the same way as $\boldsymbol{\xi}$) which is the (unique) F-previsible process such that $\mathbf{B}^{(N)}(0) = \mathbf{0}$ and each component of $\boldsymbol{\xi} - \mathbf{B}^{(N)}$ is a $(\mathbb{P}_N, F)$-martingale. If the limit

$$\boldsymbol{\beta}^{(N)}(t) = \lim_{h \downarrow 0} h^{-1} \mathbb{E}_N \left(\boldsymbol{\xi}(t+h) - \boldsymbol{\xi}(t) \middle| \mathcal{F}_t \right). \tag{3.1}$$

exists a.s. for all $t \ge 0$, then $\boldsymbol{\beta}^{(N)}$ is called the *conditional intensity process* for $\boldsymbol{\xi}$ (with respect to F), and we have

$$\mathbf{B}^{(N)}(t) = \int_0^t \boldsymbol{\beta}^{(N)}(s) ds \text{ a.s.,}$$

for all t. (All such relations are to be understood componentwise). For further details, see for example Section 13.2 of [11].

To show the existence of $\boldsymbol{\beta}^{(N)}$ for all N, and to demonstrate the convergence as $N \to \infty$, we will additionally consider the conditional intensity of the point process $\boldsymbol{\xi}$ with respect to $(\Omega, \bar{\mathbb{P}}_N, \mathcal{G}, G)$, denoting this by $\boldsymbol{\alpha}^{(N)}$. Analogously to (3.1), we have

$$\boldsymbol{\alpha}^{(N)}(t) = \lim_{h \downarrow 0} h^{-1} \mathbb{E}_N (\boldsymbol{\xi}(t+h) - \boldsymbol{\xi}(t) | \mathcal{G}_t). \tag{3.2}$$

Since, for each N, $(\underline{\mathbf{r}}(s), \mathbf{x}(s))$ is a countable state-space Markov process under $\bar{\mathbb{P}}_N$, with $\mathcal{G}_t = \sigma(\{\underline{\mathbf{r}}(s), \mathbf{x}(s)\}, 0 \le s \le t)$, the limit (3.2) exists a.s. for all t, and corresponds to a vector of certain instantaneous transition probabilities from the state $(\underline{\mathbf{r}}(t), \mathbf{x}(t))$. The transitions concerned are those representing a departure or an arrival at a tagged channel, giving rise to a point in the process $\boldsymbol{\xi}$. We later give the exact forms of the components of $\boldsymbol{\alpha}^{(N)}$. Note for the moment that

$$\alpha_{k,l}^{(N)}(t) \le \max_i \mu_i + m \max_i (\lambda_i + \sum_j \mu_j p_{ji})$$

a.s., for all k, l and t.

Then

$$\begin{aligned}
\boldsymbol{\beta}^{(N)}(t) &= \lim_{h \downarrow 0} h^{-1} \mathbb{E}_N \left(\boldsymbol{\xi}(t+h) - \boldsymbol{\xi}(t) | \mathcal{F}_t \right) \\
&= \lim_{h \downarrow 0} h^{-1} \mathbb{E}_N \left(\mathbb{E}_N (\boldsymbol{\xi}(t+h) - \boldsymbol{\xi}(t) | \mathcal{G}_t) | \mathcal{F}_t \right) \\
&= \mathbb{E}_N \left(\lim_{h \downarrow 0} h^{-1} \mathbb{E}_N \left(\boldsymbol{\xi}(t+h) - \boldsymbol{\xi}(t) | \mathcal{G}_t \right) \big| \mathcal{F}_t \right) \\
&= \mathbb{E}_N \left(\boldsymbol{\alpha}^{(N)}(t) \big| \mathcal{F}_t \right),
\end{aligned} \tag{3.3}$$

using a version of the Dominated Convergence Theorem for conditional expectations; see for example [59], section 9.7.

Conversely to the above, a given conditional intensity process yields uniquely the law of a corresponding point process, subject to the condition, which we shall require, that no two points occur simultaneously. See for example [1]. Here, the limiting measure $\mathbb{P}$ will be specified by the initial distribution ψ and by a conditional intensity process $\boldsymbol{\beta}(t)$ which we will construct; $\boldsymbol{\beta}$ will be adapted to F and represents the limit of the processes $\boldsymbol{\beta}^{(N)}$.

Kabanov and Liptser [26] show that convergence in local variation distance of measures defining multivariate point processes is implied by a condition on the convergence of their compensators. (In fact they regard both point processes and their compensators as random measures, so that they compare directly the convergence of (random) variation distance of point processes with that of their compensators).

We assume that $\psi_N \to \psi$ weakly; hence the marginal distribution of $\mathbf{x}(0)$ on $\mathbb{Z}_+^K$ under ψ_N converges weakly to that under ψ. Thus, since we use the discrete topology on $\mathbb{Z}_+^K$,

$\mathrm{Var}_0(\mathbb{P}_N, \mathbb{P}) \to 0$. (Here Var_0 is the variation distance between the two measures restricted to $\mathcal{F}_0$, the σ-algebra containing information only about the initial state $\mathbf{x}(0)$ of the K tagged channels). Theorem 1 of [26] then shows that a sufficient condition for $\mathrm{Var}_t(\mathbb{P}_N, \mathbb{P}) \to 0$ as $N \to \infty$ is that

$$\int_0^t \mathbb{E}_N \left|\beta_{k,l}^{(N)}(s) - \beta_{k,l}(s)\right| ds \to 0 \text{ as } N \to \infty, \tag{3.4}$$

for all k, l, where $\boldsymbol{\beta}$ is the conditional intensity of the point process $\boldsymbol{\xi}$ with respect to $(\Omega, \mathbb{P}, \mathcal{F}, F)$.

Since $\boldsymbol{\beta}$ is F-adapted, we have, by (3.3) and the $\mathcal{L}^1$-contraction property of conditional expectation, (see again [59], section 9.7), that

$$\begin{aligned}\int_0^t \mathbb{E}_N \left|\beta_{k,l}^{(N)}(s) - \beta_{k,l}(s)\right| ds &= \int_0^t \mathbb{E}_N \left|\mathbb{E}_N [\alpha_{k,l}^{(N)}(s) - \beta_{k,l}(s) | \mathcal{F}_s]\right| ds \\ &\le \int_0^t \mathbb{E}_N \left|\alpha_{k,l}^{(N)}(s) - \beta_{k,l}(s)\right| ds.\end{aligned}$$

Hence it suffices to construct a non-negative F-adapted process $\boldsymbol{\beta}$ such that

$$\int_0^t \mathbb{E}_N \left|\mathbb{E}_N [\alpha_{k,l}^{(N)}(s) - \beta_{k,l}(s) | \mathcal{F}_s]\right| ds \to 0 \text{ as } N \to \infty, \tag{3.5}$$

for all k, l.

Consider a channel at node i whose queue length is x, while the "environment" is $\underline{\mathbf{u}} \in \overline{\mathcal{U}}^S$. We define the quantity $c_i(\underline{\mathbf{u}}, x)$ by

$$c_i(\underline{\mathbf{u}}, x) = u_i(x)^{m-1} + u_i(x)^{m-2} u_i(x+1) + \ldots + u_i(x+1)^{m-1}$$

which, provided $u_i(x) > u_i(x+1)$, is equal to

$$\frac{u_i(x)^m - u_i(x+1)^m}{u_i(x) - u_i(x+1)}.$$

Here the numerator is the probability, in the environment $\underline{\mathbf{u}}$, that out of m randomly chosen queues at node i, the shortest has length x, and the denominator is the proportion of queues at node i with length x. Hence $N^{-1} c_i(\underline{\mathbf{u}}, x)$ may be interpreted as the probability that a new customer arriving at the node chooses this particular channel to join.

Note that for all $x \in \mathbb{Z}_+^K$ and $\underline{\mathbf{u}}, \underline{\mathbf{u}}' \in \overline{\mathcal{U}}^S$,

$$|c_i(\underline{\mathbf{u}}, x)| \le m \tag{3.6}$$

and

$$|c_i(\underline{\mathbf{u}}, x) - c_i(\underline{\mathbf{u}}', x)| \le (x+1) m^2 \|\underline{\mathbf{u}} - \underline{\mathbf{u}}'\|. \tag{3.7}$$

We now consider the conditional intensities of the three types of component process $\boldsymbol{\xi}_{k,l}$, depending on which of the subscribts k and l are zero.

First consider the case $1 \leq k, l \leq K$, so that we are interested in transfers from the kth tagged channel to the lth tagged channel. At time t, the instantaneous rate at which departures destined for node $i(l)$ occur at the kth tagged channel is $\mathbf{I}\{x_k(t) > 0\}\mu_{i(k)} p_{i(k)i(l)}$, so that we have

$$\alpha_{k,l}^{(N)}(t) = \mathbf{I}\{x_k(t) > 0\}\mu_{i(k)} p_{i(k)i(l)} N^{-1} c_{i(l)} \big(\underline{\mathbf{r}}(t) - N^{-1}\underline{\mathbf{e}}_{i(k)}(x_k(t)), x_l(t) - \mathbf{I}\{k = l\}\big).$$

(Here and below we write $\underline{\mathbf{e}}_i(n')$ for the vector $(e_j(n), j \in S, n \in \mathbb{Z}_+)$ whose only non-zero entry is $e_i(n') = 1$). Then

$$|\alpha_{k,l}^{(N)}(t)| \leq N^{-1} m \max_i \mu_i \max_{i,j} p_{ij} \tag{3.8}$$

$$\to 0 \text{ as } N \to \infty, \tag{3.9}$$

so putting

$$\beta_{k,l}(t) = 0 \tag{3.10}$$

gives (3.5) for this case.

Next take $1 \leq k \leq K$, $l = 0$. We have

$$\alpha_{k,0}^{(N)}(t) = \mathbf{I}(x_k(t) > 0)\mu_{i(k)} - \sum_{l=1}^{K} \alpha_{k,l}^{(N)}(t),$$

so that putting

$$\beta_{k,0}(t) = \mathbf{I}(x_k(t) > 0)\mu_{i(k)} \tag{3.11}$$

gives (3.5) again.

The case $k = 0$, $1 \leq l \leq K$ is the most difficult – it is when considering arrivals at tagged channels which come from outside the set of tagged channels that the effect of the environment is most greatly felt. We have

$$\begin{aligned}\alpha_{0,l}^{(N)}(t) = \lambda_{i(l)} c_{i(l)} \big(\underline{\mathbf{r}}(t), x_l(t)\big) \\ + \sum_j \Bigg\{ \left(1 - \frac{\#\{k : i(k) = j\}}{N}\right) \sum_{n=1}^{\infty} \Big\{ [r_j(n,t) - r_j(n+1,t)] \\ \times \mu_j p_{ji(l)} c_{i(l)} \big(\underline{\mathbf{r}}(t) - N^{-1}\underline{\mathbf{e}}_j(n), x_l(t)\big)\Big\}\Bigg\}.\end{aligned}$$

Let $\underline{\mathbf{g}}$ be the point at which the marginal distribution of ψ on $\overline{\mathcal{U}}^S$ is concentrated. We will set

$$\beta_{0,l}(t) = \left[\lambda_{i(l)} + \sum_j u_j(1,t,\underline{\mathbf{g}})\mu_j p_{ji(l)}\right] c_{i(l)} \big(\underline{\mathbf{u}}(t,\underline{\mathbf{g}}), x_l(t)\big), \tag{3.12}$$

where $\underline{\mathbf{u}}(t,\underline{\mathbf{g}})$ is defined by Theorem 2.1.1(i).

Using (3.6) and (3.7) one can find constants D_1, D_2 and D_3, depending on m, $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$ and $\mathbf{P}$ but not on N, such that

$$|\alpha_{0,l}^{(N)}(t) - \beta_{0,l}(t)| \le D_1 \min\left(D_2, x_l(t)\|\underline{\mathbf{r}}(t) - \underline{\mathbf{u}}(t, \underline{\mathbf{g}})\|\right) + \frac{D_3}{N}.$$

Thus

$$\begin{aligned}\limsup_{N\to\infty} \int_0^t \mathbb{E}_N\, |\alpha_{0,l}^{(N)}(s) - \beta_{0,l}(s)| ds \\ &\le D_1 \limsup_{N\to\infty} \int_0^t \mathbb{E}_N \min\left(D_2, x_l(s)\|\underline{\mathbf{r}}(s) - \underline{\mathbf{u}}(s, \underline{\mathbf{g}})\|\right) ds \\ &\le D_1 \limsup_{N\to\infty} \int_0^t \left[y\, \mathbb{E}_N\left(\|\underline{\mathbf{r}}(s) - \underline{\mathbf{u}}(s, \underline{\mathbf{g}})\|\right) + D_2 \mathbb{P}_N(x_l(s) > y)\right] ds\end{aligned} \tag{3.13}$$

for any y. The arrival process at channel l may be dominated for all N by a Poisson process of rate $m(\lambda_{i(l)} + \sum_j \mu_j p_{ji(l)})$, and the distribution of $x_l(0)$ under ψ_N converges to that under ψ, so we have

$$\mathbb{P}_N(x_l(s) > y) \to 0 \text{ as } y \to 0,$$

uniformly in N and in $s \in [0, t]$. Also, from Theorem 2.1.2,

$$\mathbb{E}_N\left(\|\underline{\mathbf{r}}(s) - \underline{\mathbf{u}}(s, \underline{\mathbf{g}})\|\right) \to 0 \text{ as } N \to \infty,$$

uniformly in $s \in [0, t]$. This shows that the RHS of (3.13) is 0, so that (3.5) holds in this final case also.

Thus, since $\boldsymbol{\beta}$ defined by (3.10), (3.11) and (3.12) is non-negative and F-adapted and satisfies (3.5) for all k and l, the first part of the theorem is proved.

For the second part, notice that $\beta_{k,l}(t)$ is identically zero if $k > 0$ and $l > 0$, and that $\beta_{k,0}(t)$ and $\beta_{0,k}(t)$ depend on $\mathbf{x}(s), 0 \le s \le t$ only through $x_k(t)$. Conversely, $x_k(t)$ is a function only of $\xi_{k,l}(t)$ and $\xi_{l,k}(t)$, $0 \le l \le K$. Hence $\mathbf{x}$ is Markov, and, if the components of $\mathbf{x}$ are independent at time 0, their paths thereafter are also independent. □

3.3 Systems in Equilibrium

If the networks considered in the previous section are each started in equilibrium, we can describe more precisely the limiting point process arising from Theorem 3.2.1.

Define $\boldsymbol{\rho}$ by (1.6) as before, so that, for all $j \in S$,

$$\rho_j = \lambda_j + \sum_{k\in S} \rho_k p_{kj}. \tag{3.14}$$

From here on we assume that the non-overload condition $\boldsymbol{\rho} < \boldsymbol{\mu}$ for the network holds, and that for each N the distribution under ψ_N of the initial state $\underline{\mathbf{r}}(0)$ on $\overline{\mathcal{U}}_N^S$ is the equilibrium distribution π_N given by Theorem 2.1.4(i). If ψ is the weak limit of ψ_N then, from the same theorem, $\underline{\mathbf{r}}(0)$ is equal to $\underline{\mathbf{a}}$ (as defined by (2.8)) with probability 1 under ψ, and we can apply Theorem 3.2.1. Since $\underline{\mathbf{u}}(\underline{\mathbf{a}}, t) = \underline{\mathbf{a}}$ for all t, (3.12) becomes

$$\beta_{0,l}(t) = \rho_{i(l)} c_{i(l)}\big(\underline{\mathbf{a}}, x_l(t)\big), \tag{3.15}$$

using (3.14). For $m = 1$, $c_{i(l)}(.) = 1$ identically, and for $m > 1$ we have

$$c_i(\underline{\mathbf{a}}, x) = \frac{(\rho_i/\mu_i)^{m(m^x-1)/m-1}\left(1-(\rho_i/\mu_i)^{m^{x+1}}\right)}{(\rho_i/\mu_i)^{m^x-1/m-1}\left(1-(\rho_i/\mu_i)^{m^x}\right)}.$$

As before, $\beta_{k,l}(t) = 0$ for $k, l \geq 1$, and $\beta_{k,0}(t) = \mathbf{I}\{x_k(t) > 0\}\mu_{i(k)}$ for $k \geq 1$. Thus under the limiting measure $\mathbb{P}$, the queue length processes at the tagged channels are independent time-homogeneous Markov birth-and-death processes. From the fixed point result in Theorem 2.1.1(ii), it follows that the process at a channel belonging to node i has an equilibrium distribution under which the probability of the queue length being n or longer is $\{a_i(n), n \in \mathbb{Z}_+\}$.

In the case $m = 1$, the queue behaves simply as an $M/M/1$ queue, whose arrival rate does not depend on the queue length. For $m > 1$, the arrival rate decreases as the length increases.

The arrival rate when the queue is empty is

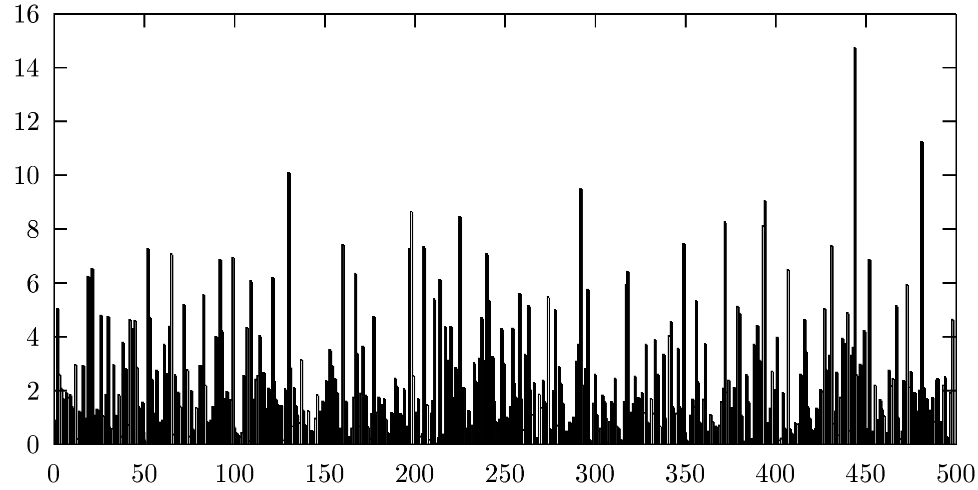
$$\frac{1-(\rho_i/\mu_i)^m}{1-(\rho_i/\mu_i)},$$

which increases as m increases. Thus the average length of an idle period of the queue decreases as m increases; the overall intensity of arrivals and the average service time stay constant, so the average length of a busy period decreases also.

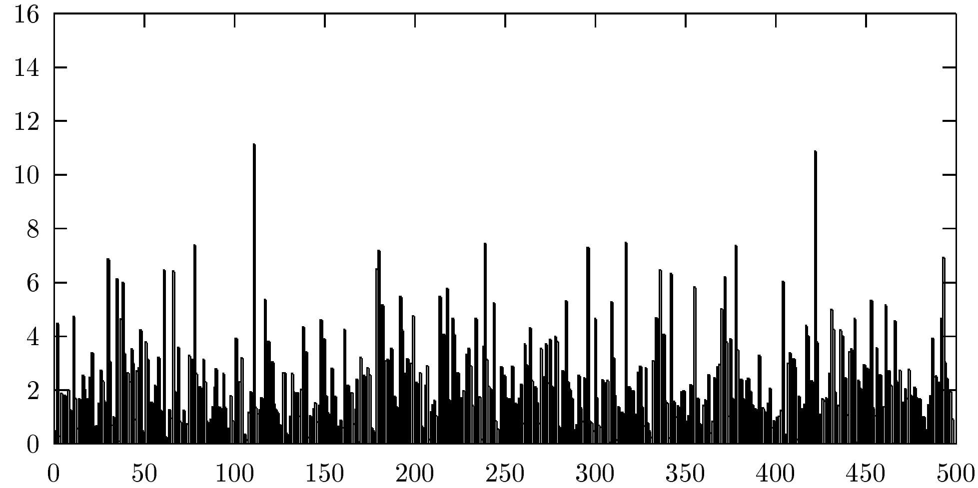
We now consider specifically the point process of arrivals at such a queue. (If the queue is started in equilibrium, then by a reversibility argument one can show that this has the same distribution as the point process of departures). If $m = 1$, this is simply a Poisson process of rate ρ_i. As m increases, the average intensity remains the same, but the points tend to become more "evenly spread"; the probability of a given time interval containing no points at all decreases, for example. In the limit $m = \infty$, the points of the process are those of a renewal process, whose renewal intervals are the independent sum of a service time which is exponentially distributed with rate μ_i and an arrival time which is exponentially distributed with rate $(1-\rho_i/\mu_i)^{-1}$. This corresponds to a situation where arrivals at a node choose freely between all channels, and so only ever join empty queues.

This change is illustrated by Figure 3.1, which shows the results of simulations of sequences of 500 interarrival times in the cases $m = 1$, $m = 2$ and $m = \infty$ for $\rho_i = 0.5$ and μ_i=1. As m increases, very long interarrival distances become less frequent and tend to bunch together less; the same applies to very short interarrival distances.

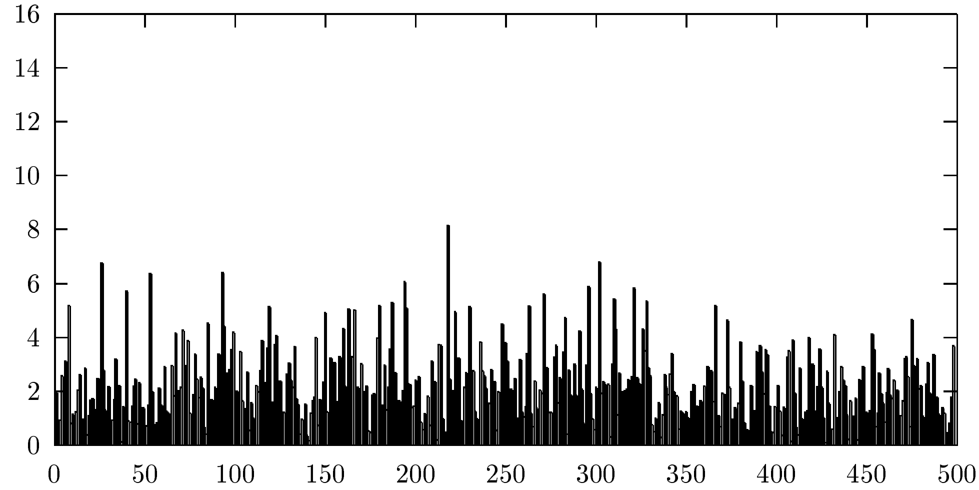
A different phenomenon occurs if ρ_i is nearly 1. Then the difference between the Poisson process of arrivals for $m = 1$ and the renewal process of arrivals for $m = \infty$ is very slight –



$m = 1$: total time 978.96



$m = 2$: total time 1013.60



$m = \infty$: total time 975.94

Figure 3.1: Simulations of 500 successive interarrival times for $\rho = 0.5$, $\mu = 1$, various m.

and, as noted above, in each case the departure process is distributed exactly as the arrival process so again is very similar for $m = 1$ and $m = \infty$. However, the process of queue lengths at the channel, which is determined by the *joint* distribution of arrivals and departures, is extremely different: for example, for $m = 1$ its stationary distribution is geometric with mean $\rho_i/(1 - \rho_i)$, while for large m the queue length hardly ever exceeds 1.

We can alternatively consider the network from the viewpoint of a particular task progressing through it. Since the routing is Markovian, its route may as well be considered fixed as soon as it enters the network – the route may be taken to include both the order in which the task visits nodes and the particular channels it inspects on each arrival at a node.

The above observation that the queue length processes at the tagged channels are independent in the limit can then be interpreted as follows: as $N \to \infty$, we approach a situation in which the waiting times of the task at each stage of the route are independent, and where the queue lengths of the m channels that the task inspects at node i are independently drawn from the distribution represented by $a_i(n), n \in \mathbb{Z}_+$. Thus the probability that the task joins a queue of length n at node i is $a_i(n)^m - a_i(n+1)^m$, which, from (2.8), can be written as

$$\left(\frac{\rho_i}{\mu_i}\right)^{\frac{m(m^n-1)}{m-1}} \left(1 - \left(\frac{\rho_i}{\mu_i}\right)^{m^{n+1}}\right),$$

decaying superexponentially in n for $m > 1$.

3.4 Extensions

An analagous result to Theorem 3.2.1 holds if we remove the condition that the marginal distribution of $\underline{\mathbf{r}}(0)$ on $\overline{\mathcal{U}}^S$ under the limiting initial condition ψ is concentrated at a single point. Then, however, we must replace 3.12 by

$$\alpha_{0,l}(t) = \left[\lambda_{i(l)} + \sum_j u_j\big(1, t, \underline{\mathbf{r}}(0)\big)\mu_j p_{ji(l)}\right] c_{i(l)}\Big(\underline{\mathbf{u}}\big(t, \underline{\mathbf{r}}(0)\big), x_l(t)\Big)$$

and

$$\beta_{0,l}(t) = \mathbb{E}[\alpha_{0,l}(t)|\mathcal{F}_t].$$

Now part (ii) of Theorem 3.2.1 fails, since the observed paths of $\{\mathbf{x}(s), 0 \le s \le t\}$ provide information about the "hidden" initial environment $\underline{\mathbf{r}}(0)$ and hence about $\underline{\mathbf{r}}(t)$; so in general $\mathbf{x}$ is no longer Markov and the components of $\mathbf{x}(t)$ are not independent even if those of $\mathbf{x}(0)$ are. One can formulate an interesting non-parametric inference problem concerning the estimation of the initial environment $\underline{\mathbf{r}}(0)$ given the observed paths $\{\mathbf{x}(s), 0 \le s \le t\}$ of the queue length processes at the tagged channels.

Following more closely the approach of Brown and Pollett in [8], we can consider letting $|S|$, the number of nodes, rather than N, the number of channels at each node, tend to infinity. In

[8], this limit is considered for single-class Markovian queueing networks with state-dependent service rates, and, under various conditions, bounds are derived for the variation distance between the equilibrium arrival process at a node and a Poisson process, which tend to 0 as $|S| \to \infty$. For example, for a standard Jackson network, one appropriate condition under which the distance between the arrival process at node j and a Poisson process tends to 0 in equilibrium is that

$$\sum_{k \in S} (\mu_k p_{kj})^2 \to 0$$

as the set S of nodes becomes large.

It seems likely that similar results hold also in the situations we have considered above. However, the networks considered in [8] all have the property that the equilibrium distribution of the state of the network has a product form, and this is an important element of the methods used there. In the networks we have considered here, the lengths of the queues at channels at the same or at different nodes are not in general independent for finite N, even in equilibrium, and it seems that different methods will be needed to establish such results in this case.

Chapter 4

Stochastic Bounds for Fast Jackson Networks

4.1 Introduction

In this chapter, we again consider the "Fast Jackson Network" model introduced in Section 2.1. In the previous two chapters we have considered a limit as N, the number of servers at each node of the network, becomes large; now we will consider properties of networks with fixed finite N. Another difference is that we will no longer focus particularly on the case where the non-overload condition $\boldsymbol{\rho} < \boldsymbol{\mu}$ holds and the network has an equilibrium; in fact the results we prove will be in some sense "best possible" exactly when every node of the network is at least fully loaded.

Recall from the introduction the idea of a *VDK node*, which corresponds to a single node of the network models which we have been considering. Here we provide stochastic bounds which compare the behaviour of our network with the behaviour of a system of $|S|$ independent VDK nodes, whose arrival and service rates depend on the parameters $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$ and $\mathbf{P}$ of the network.

The bounds compare the probabilities of certain increasing sets of network states, and cannot be derived by means of standard pathwise, or even pointwise, stochastic orderings of the states of the two systems. They are an extension of a result of Massey [40], who provides bounds on the evolution of a standard Jackson network in terms of the behaviour of independent $M/M/1$ queues.

The original proof by Massey involved the use of some rather intricate operator methods. A much simpler proof based on a coupling was proposed by Tsoucas and Walrand in [49]. However, there appears to be a gap in the proof which they present – we discuss this after the proof of Theorem 4.2.1 below. Nevertheless their methods are certainly valuable and we draw on them for the basis of the proof presented in this chapter.

4.2 Stochastic Bounds

We will now describe the state of the network at time t by a vector $\underline{\mathbf{X}}(t) = \{X_{j,r}(t), j \in S, 1 \le r \le N\} \in (\mathbb{Z}_+^N)^S$; here $X_{j,r}(t)$ represents the length of the queue in the rth channel of the jth node at time t. For convenience, we will identify S with the set $\{1, 2, \ldots, J\}$, where $J = |S|$.

We call $\Gamma \subset \mathbb{Z}_+^N$ an *increasing set* if, whenever $\underline{x} \in \Gamma$ and $\underline{x}' \ge \underline{x}$, we have $\underline{x}' \in \Gamma$ also. For random variables $\underline{\mathbf{X}}, \underline{\mathbf{Y}}$ taking values in $(\mathbb{Z}_+^N)^S$, write $\underline{\mathbf{X}} \ge_d \underline{\mathbf{Y}}$ if

$$\mathbb{P}(\underline{X}_1 \in \Gamma_1, \ldots, \underline{X}_J \in \Gamma_J) \ge \mathbb{P}(\underline{Y}_1 \in \Gamma_1, \ldots, \underline{Y}_J \in \Gamma_J)$$

for all increasing sets $\Gamma_1, \ldots, \Gamma_J \subset \mathbb{Z}_+^N$.

In the case of a standard Jackson network, one may freely assume that $p_{ii} = 0$ for all i, since a departure which immediately reenters the same node does not change the state of the system. Here, this is not the case, since the task may leave one channel and join another at the same node – so we extend slightly the model of a VDK node considered in [51] to include the possibility that some departures from the system are immediately reintroduced (and join the shortest of m randomly selected queues in the normal way, independently of previous choices).

Theorem 4.2.1 *Let $\underline{\mathbf{Y}}(t)$ be a fast Jackson network with parameters $J = |S|$, N, m, $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$ and $\mathbf{P}$.*

(i) Let $\underline{\mathbf{X}}(t)$ be a system of J independently evolving VDK nodes, with parameters N and m, arrivals at node i at rate $N(\lambda_i + \sum_{j \ne i} \mu_j p_{ji})$ and service at node i at rate μ_i, and with a proportion p_{ii} of departures at node i immediately reintroduced to the same node. Suppose that the time 0 distributions of $\underline{\mathbf{X}}$ and $\underline{\mathbf{Y}}$ satisfy $\underline{\mathbf{X}}(0) \ge_d \underline{\mathbf{Y}}(0)$. Then $\underline{\mathbf{X}}(t) \ge_d \underline{\mathbf{Y}}(t)$ for all $t \ge 0$.

(ii) Let $Z(t)$ be a single $M/M/1$ queue, with arrival rate $N \sum_i \lambda_i$ and service rate $N \sum_i \mu_i p_i^*$, and with

$$\mathbb{P}(Z(0) \ge z) \le \mathbb{P}\left(\sum_{i=1}^{J} \sum_{r=1}^{N} Y_{i,r}(0) \ge z\right)$$

for all r. Then, for all t, z,

$$\mathbb{P}(Z(t) \ge z) \le \mathbb{P}\left(\sum_{i=1}^{J} \sum_{r=1}^{N} Y_{i,r}(t) \ge z\right).$$

Proof:

(i) We will couple the evolution of $\underline{\mathbf{X}}$ and $\underline{\mathbf{Y}}$. Consider a collection of Poisson processes A and $\tilde{A}$, all independent of each other, indexed variously by $i, j \in S \cup \{0\} = \{0, \ldots, J\}$ and $a, b_1, b_2, \ldots, b_m \in \{1, \ldots, N\}$. The indices have the following meanings, where relevant: i and a represent the node and channel respectively from which a customer departs; j represents the node to which a customer arrives; and $b_1, \ldots, b_m$ represent the channels selected on arrival at the node – the customer joins the one with the shortest queue (choosing the first one listed in the case of a tie). If one of i or j is 0, this indicates an arrival from, or a departure to, outside the network. The word "potential" below means that the transfer or departure referred to only occurs if the channel from which a customer is to depart is non-empty. The collection of processes may be divided into five types:

(i) $A_{i,a,j,b_1,\ldots,b_m}, i, j \geq 1, i \neq j$, of rates $\mu_i p_{ij} N^{-m}$, whose points represent, simultaneously, potential transfers from node i to node j in $\underline{\mathbf{Y}}$ and potential departures from node i in $\underline{\mathbf{X}}$.

(ii) $\tilde{A}_{i,a,j,b_1,\ldots,b_m}, i, j \geq 1, i \neq j$, of rates $\mu_i p_{ij} N^{-m}$, whose points represent arrivals at node j in $\underline{\mathbf{X}}$. (Here the nature of the event is not affected by i and a, but the indices are retained for comparison with type (i)).

(iii) $A_{i,a,i,b_1,\ldots,b_m}, i \geq 1$, of rates $\mu_i p_{ii} N^{-m}$ representing, in both $\underline{\mathbf{X}}$ and $\underline{\mathbf{Y}}$, potential transfers within node i.

(iv) $A_{0,0,j,b_1,\ldots,b_m}, j \geq 1$, of rates $\lambda_j N^{-m+1}$ representing, in both $\underline{\mathbf{X}}$ and $\underline{\mathbf{Y}}$, arrivals at node j.

(v) $A_{i,a,0,0,\ldots,0}, i \geq 1$, of rates $\mu_i p_i^*$ representing, in both $\underline{\mathbf{X}}$ and $\underline{\mathbf{Y}}$, potential departures from node i.

Under the coupling thereby described, $\underline{\mathbf{X}}(t)$ and $\underline{\mathbf{Y}}(t)$ are processes of the required types, and the pair $(\underline{\mathbf{X}}(t), \underline{\mathbf{Y}}(t), t \geq 0)$ forms a Markov process, whose jump times are a subset of the points of the various processes A and $\tilde{A}$.

Let $T_0 = 0$ and, for $n \geq 1$, let T_n be the time of the nth point after time 0 of all the processes $A, \tilde{A}$. Write $\underline{\mathbf{X}}^{(n)} = \underline{\mathbf{X}}(T_n+), \underline{\mathbf{Y}}^{(n)} = \underline{\mathbf{Y}}(T_n+)$. For appropriate i, a, j, $b_1, \ldots, b_m$, say that the event $G_n(i, a, j, b_1, \ldots, b_m)$ occurs if the point at T_n is a point in a process $A_{i,a,j,b_1,\ldots,b_m}$ or $\tilde{A}_{i,a,j,b_1,\ldots,b_m}$.

Since all the $A, \tilde{A}$ are independent and Poisson, the points $\{T_n, n \geq 1\}$ form a Poisson process; the process $\{T_n, n \geq 1\}$ is independent of the events $\{G_n, n \geq 1\}$ and so also of $\{\underline{\mathbf{X}}^{(n)}, \underline{\mathbf{Y}}^{(n)}, n \geq 0\}$. Then

$$\begin{aligned}\mathbb{P}(\underline{\mathbf{X}}(t) \in \Gamma_1 \times \ldots \times \Gamma_J) &= \sum_{n=0}^{\infty} \mathbb{P}(T_n \leq t < T_{n+1}, \underline{\mathbf{X}}^{(n)} \in \Gamma_1 \times \ldots \times \Gamma_J) \\ &= \sum_{n=0}^{\infty} \mathbb{P}(T_n \leq t < T_{n+1}) \mathbb{P}(\underline{\mathbf{X}}^{(n)} \in \Gamma_1 \times \ldots \times \Gamma_J).\end{aligned}$$

Also

$$\begin{aligned}\mathbb{P}(\underline{\mathbf{X}}^{(n)} \in \Gamma_1 \times \ldots \times \Gamma_J) = &\sum_{i,j,a,b_1,\ldots,b_m} \mathbb{P}(G_n(i, a, j, b_1, \ldots, b_m)) \\ &\times \mathbb{P}(\underline{\mathbf{X}}^{(n)} \in \Gamma_1 \times \ldots \times \Gamma_J \,\big|\, G_n(i, a, j, b_1, \ldots, b_m)),\end{aligned}$$

and similarly for $\underline{\mathbf{Y}}(t)$, $\underline{\mathbf{Y}}^{(n)}$.

By assumption, $\underline{\mathbf{X}}^{(0)} \geq_d \underline{\mathbf{Y}}^{(0)}$, so to prove the theorem, it suffices, by induction, to show that if $\underline{\mathbf{X}}^{(n)} \geq_d \underline{\mathbf{Y}}^{(n)}$ then, for all $\Gamma_1, \dots, \Gamma_J$ increasing, all appropriate $i, j, a, b_1, \dots, b_m$,

$$\begin{aligned} \mathbb{P}(\underline{\mathbf{X}}^{(n+1)} &\in \Gamma_1 \times \ldots \times \Gamma_J \,|\, G_{n+1}(i, a, j, b_1, \dots, b_m)) \\ &\geq \mathbb{P}(\underline{\mathbf{Y}}^{(n+1)} \in \Gamma_1 \times \ldots \times \Gamma_J \,\big|\, G_{n+1}(i, a, j, b_1, \dots, b_m)). \end{aligned} \tag{4.1}$$

We make the following observation. Fix $a, b_1, \dots, b_m$, and take an increasing set $\Gamma_k \subset \mathbb{Z}_+^N$ describing a collection of states at node k. Define the set Γ_k^+ of states at node k which, after a departure from channel a, yield a state in Γ_k. Then Γ_k^+ is also an increasing set, and $\Gamma_k^+ \subset \Gamma_k$. Similarly, define the set Γ_k^- of states which, after an arrival at the channel out of $b_1, \dots, b_m$ which has the shortest queue, yield a state in Γ_k; then Γ_k^- is increasing and $\Gamma_k \subset \Gamma_k^-$. Finally, consider a transfer from channel a back to the same node, joining the shortest queue from $b_1, \dots, b_m$; the set $\Gamma_k^=$ of states from which this transfer yields a state in Γ_k is increasing.

We return to (4.1). First suppose that $i = j$, so that we have a point in a process of type (iii). Without loss of generality, take $i = 1$. Then

$$\begin{aligned} \mathbb{P}(\underline{\mathbf{X}}^{(n+1)} \in \Gamma_1 \times \ldots \times \Gamma_J \,|\, &G_{n+1}(1, a, 1, b_1, \dots, b_m)) \\ &= \mathbb{P}(\underline{\mathbf{X}}^{(n)} \in \Gamma_1^= \times \Gamma_2 \times \ldots \times \Gamma_J) \\ &\geq \mathbb{P}(\underline{\mathbf{Y}}^{(n)} \in \Gamma_1^= \times \Gamma_2 \times \ldots \times \Gamma_J) \\ &= \mathbb{P}(\underline{\mathbf{Y}}^{(n+1)} \in \Gamma_1 \times \ldots \times \Gamma_J \,\big|\, G_{n+1}(1, a, 1, b_1, \dots, b_m)) \end{aligned}$$

as required. The cases $i = 0$ (type (iv)) and $j = 0$ (type (v)) are similar.

If $i, j \geq 1$ and $i \neq j$, then given $G_{n+1}(i, a, j, b_1, \dots, b_m)$ we have one out of type (i) and type (ii), with equal probability. Without loss of generality, let $i = 1$ and $j = 2$. Then

$$\begin{aligned} &\mathbb{P}(\underline{\mathbf{X}}^{(n+1)} \in \Gamma_1 \times \ldots \times \Gamma_J \,\big|\, G_{n+1}(1, a, 2, b_1, \dots, b_m)) \\ &\quad = \tfrac{1}{2}\mathbb{P}(\underline{\mathbf{X}}^{(n)} \in \Gamma_1^+ \times \Gamma_2 \times \Gamma_3 \times \ldots \times \Gamma_J) + \tfrac{1}{2}\mathbb{P}(\underline{\mathbf{X}}^{(n)} \in \Gamma_1 \times \Gamma_2^- \times \Gamma_3 \times \ldots \times \Gamma_J) \qquad (4.2) \\ &\quad \geq \tfrac{1}{2}\mathbb{P}(\underline{\mathbf{X}}^{(n)} \in \Gamma_1^+ \times \Gamma_2^- \times \Gamma_3 \times \ldots \times \Gamma_J) + \tfrac{1}{2}\mathbb{P}(\underline{\mathbf{X}}^{(n)} \in \Gamma_1 \times \Gamma_2 \times \Gamma_3 \times \ldots \times \Gamma_J) \qquad (4.3) \\ &\quad \geq \tfrac{1}{2}\mathbb{P}(\underline{\mathbf{Y}}^{(n)} \in \Gamma_1^+ \times \Gamma_2^- \times \Gamma_3 \times \ldots \times \Gamma_J) + \tfrac{1}{2}\mathbb{P}(\underline{\mathbf{Y}}^{(n)} \in \Gamma_1 \times \Gamma_2 \times \Gamma_3 \times \ldots \times \Gamma_J) \\ &\quad = \mathbb{P}(\underline{\mathbf{Y}}^{(n+1)} \in \Gamma_1 \times \ldots \times \Gamma_J \,\big|\, G_{n+1}(1, a, 2, b_1, \dots, b_m)). \end{aligned}$$

Here, to get from (4.2) to (4.3), we use the fact that if $C \subseteq C'$ and $D \subseteq D'$, then

$$\mathbb{P}(C \cap D') + \mathbb{P}(C' \cap D) \leq \mathbb{P}(C \cap D) + \mathbb{P}(C' \cap D').$$

Thus (4.1) holds for all i, j, which completes the proof of part (i).

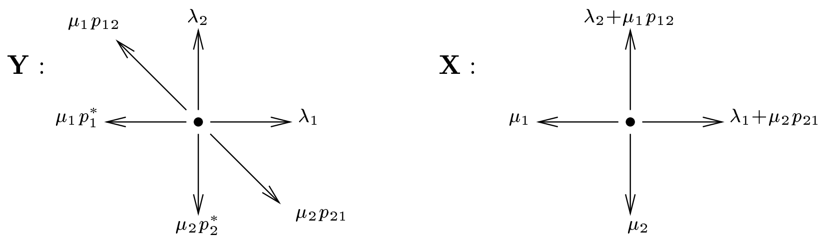
(ii) One can construct a simple coupling under which $Z(t) \leq \sum_{i=1}^J \sum_{r=1}^N Y_{i,r}(t)$ for all t. Let arrivals to Z occur at the same time as external arrivals to any node in $\underline{\mathbf{Y}}$, and let potential departures from Z occur at the same time as potential departures to outside the network from any of the channels in the nodes of $\underline{\mathbf{Y}}$. Departures from Z are only lost if the queue is already empty, while departures in $\underline{\mathbf{Y}}$ are lost whenever the selected channel is empty, even if other channels at the same or another node are not. □

Remark: Putting $N = 1$ reduces to the case of a standard Jackson network and $M/M/1$ queues. Then any increasing set Γ of the type considered has the simple form $\{x : x \geq x_0\}$ for some $x_0 \in \mathbb{Z}_+$, so that $\mathbf{X} \geq_d \mathbf{Y}$ if and only if $\mathbb{P}(\mathbf{X} \geq \mathbf{r}) \geq \mathbb{P}(\mathbf{Y} \geq \mathbf{r})$ for all $\mathbf{r} \in \mathbb{Z}_+^S$.

The result in this case was proved by Massey in [40] using operator methods. A simpler probabilistic proof was proposed by Tsoucas and Walrand in [49]; however, their presentation is not entirely clear and it appears that there is a gap in their proof. They seem to require that if $\mathbf{X}^{(n)} \geq_d \mathbf{Y}^{(n)}$ and $(X_1^{(n+1)}, X_2^{(n+1)}) \geq_d (Y_1^{(n+1)}, Y_2^{(n+1)})$, with $(X_i^{(n)}, Y_i^{(n)}) = (X_i^{(n+1)}, Y_i^{(n+1)})$ for $3 \leq i \leq J$, and with $(X_1^{(n+1)}, X_2^{(n+1)}, Y_1^{(n+1)}, Y_2^{(n+1)})$ independent of $(X_3^{(n)}, \ldots, X_J^{(n)}, Y_3^{(n)}, \ldots, Y_J^{(n)})$ given $(X_1^{(n)}, X_2^{(n)}, Y_1^{(n)}, Y_2^{(n)})$, then $\mathbf{X}^{(n+1)} \geq_d \mathbf{Y}^{(n+1)}$. This is not true in general.

Alternatively, they may regard $\mathbf{X}^{(n)}, \mathbf{Y}^{(n)}$ as fixed (by the conditioning on their collection Ψ) when considering the distribution of $(\mathbf{X}^{(n+1)}, \mathbf{Y}^{(n+1)})$; but then one may not assume that $\mathbf{X}^{(n)} \geq \mathbf{Y}^{(n)}$ always. For the construction of Massey in [40] shows not merely that one may not be able to make a pathwise coupling of the processes $\mathbf{X}(t)$ and $\mathbf{Y}(t)$, $t \geq 0$ in such a way that $\mathbb{P}(\mathbf{X}(t) \geq \mathbf{Y}(t) \text{ for all } t \geq 0) = 1$ (as noted by Tsoucas and Walrand in their Remark (3)), but, more strongly, that for at least some fixed t, one may not even be able to couple the two random variables $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ in such a way that $\mathbb{P}(\mathbf{X}(t) \geq \mathbf{Y}(t)) = 1$.

For example, let $J = 2$, let $p_{11} = p_{22} = 0$, let $p_{12} + p_{21} > 0$, and let $\mathbf{X}(0) = \mathbf{Y}(0) = (1, 1)$ with probability 1. Define $B = \{\mathbf{x} : x_1 + x_2 \geq 2\}$; then B is an increasing set. The transition rates for the processes $\mathbf{X}$ and $\mathbf{Y}$ at time 0 are shown in the diagram (where movement to the right indicates an increase in the first co-ordinate and movement upwards an increase in the second):



Since the only possible transitions that lead to the process leaving the set B from the state $(1, 1)$ are those directly down or directly to the left, one sees that

$$\begin{aligned}
\lim_{t \downarrow 0} \frac{1}{t} \mathbb{P}(\mathbf{Y}(t) \notin B) &= \mu_1 p_1^* + \mu_2 p_2^* \\
&< \mu_1 + \mu_2 \\
&= \lim_{t \downarrow 0} \frac{1}{t} \mathbb{P}(\mathbf{X}(t) \notin B).
\end{aligned}$$

Similarly, one cannot necessarily couple $\mathbf{X}^{(n)}$ and $\mathbf{Y}^{(n)}$ to give $\mathbb{P}(\mathbf{X}^{(n)} \geq \mathbf{Y}^{(n)}) = 1$; the stochastic inequalities here hold only in the more restricted sense given by $\geq_d$.

The proof which we present above (which may of course be written more straightforwardly in the case $N = 1$) avoids these problems by considering more explicitly the influence of $(\underline{\mathbf{X}}^{(n)}, \underline{\mathbf{Y}}^{(n)})$ on $(\underline{\mathbf{X}}^{(n+1)}, \underline{\mathbf{Y}}^{(n+1)})$, by means of the sets Γ^+, Γ^-, $\Gamma^=$.

4.3 Tightness of Bounds

Define $\tilde{\rho}_i = \lambda_i + \sum_{j=1}^J \mu_j p_{ji}$. The bounds in Theorem 4.2.1 are asymptotically tight (in a sense to be made precise in Theorem 4.3.2) if $\tilde{\rho}_i \geq \mu_i$ for all i. Equivalently, $\hat{\rho}_i \geq \mu_i$ for all i, where $\hat{\boldsymbol{\rho}} = \{\hat{\rho}_i, i \in S\}$ is the unique solution to the "generalised throughput equations" $\hat{\rho}_i = \lambda_i + \sum_{j=1}^J \min(\mu_j, \hat{\rho}_j) p_{ji}$ (see [16] for details); thus every node in the network is working at full capacity. If $\tilde{\rho}_j < \mu_j$ for some j, then (with probability 1) the proportion of time for which a typical channel at node j is busy does not converge to 1 in the long run, so that the terms $\mu_j p_{ji}$, representing the long-run average rate of transfer from node j to node i, are overestimates, and the bound is not optimal (unless $p_j^* = 1$).

The following lemma is a restatement of Theorem 2.1.3 in the special case where S consists of only a single node. (Alternatively, Theorem 4 of [50] provides an easier route to more sophisticated comparisons between VDK nodes with different values of m, though a small change to the coupling method used there is needed if some departures are to be reintroduced to the node).

Lemma 4.3.1 *Let $\mathbf{V}(t)$ and $\tilde{\mathbf{V}}(t)$ be VDK nodes with the same number of channels N, the same arrival rate, the same service rate, and the same proportion of departures immediately reintroduced to the node. Let arrivals in $\mathbf{V}$ inspect m channels, and arrivals in $\tilde{\mathbf{V}}$ inspect only one channel. If $\mathbf{V}(0)$ and $\tilde{\mathbf{V}}(0)$ have the same distribution, then there exists a coupling between the two systems such that, for all t, $\sum_{r=1}^N V_r(t) \leq \sum_{r=1}^N \tilde{V}_r(t)$.*

We can now prove the theorem on tightness of bounds, which is analogous to Theorem 3.6 in [40].

Theorem 4.3.2 *Let $\underline{\mathbf{X}}$, $\underline{\mathbf{Y}}$ and Z be as in Theorem 4.2.1. Assume that $\mathbb{E}[X_{i,r}(0)] < \infty$ for all i, r. Then if $\tilde{\rho}_i \geq \mu_i$ for all i,*

$$\lim_{t\to\infty} t^{-1}\mathbb{E}[Z(t)] = N \sum_{i \in S} (\tilde{\rho}_i - \mu_i),$$

and, for all i and r,

$$\lim_{t\to\infty} t^{-1}\mathbb{E}[Y_{i,r}(t)] = \lim_{t\to\infty} t^{-1}\mathbb{E}[X_{i,r}(t)] = \tilde{\rho}_i - \mu_i.$$

Proof: For an $M/M/1$ queue $Q(t)$ with arrival rate α, service rate $\beta \leq \alpha$ and $\mathbb{E}[Q(0)] < \infty$,

$$\lim_{t\to\infty} t^{-1}\mathbb{E}[Q(t)] = \alpha - \beta. \tag{4.4}$$

Hence

$$\lim_{t\to\infty} t^{-1}\mathbb{E}[Z(t)] = N\sum_{i\in S}\lambda_i - N\sum_{i\in S}\mu_i p_i^*, \\ = N\sum_{i\in S}(\tilde{\rho}_i - \mu_i), \tag{4.5}$$

as required. From Theorem 4.2.1, we also have, for all i, r, t,

$$\mathbb{E}[Y_{i,r}(t)] \le \mathbb{E}[X_{i,r}(t)] \tag{4.6}$$

and

$$\mathbb{E}[Z(t)] \le \sum_{i\in S}\sum_{r=1}^{N}\mathbb{E}[Y_{i,r}(t)]. \tag{4.7}$$

Consider a system $\underline{\mathbf{X}}'(t)$ evolving in the same way as $\underline{\mathbf{X}}(t)$, but with initial distribution given by

$$\mathbb{P}(X'_{i,1}(0) = X'_{i,2}(0) = \ldots = X'_{i,N}(0)) = 1$$

for all i, and, for all i, r, z,

$$\mathbb{P}(X'_{i,r}(0) = z) = \mathbb{P}(\max_{1\le r'\le N} X_{i,r'}(0) = z).$$

Then one may easily construct a coupling between $\underline{\mathbf{X}}$ and $\underline{\mathbf{X}}'$ under which $\underline{\mathbf{X}}(t) \le \underline{\mathbf{X}}'(t)$ for all t, so that, for all i, r, t,

$$\mathbb{E}[X_{i,r}(t)] \le \mathbb{E}[X'_{i,r}(t)]. \tag{4.8}$$

Next, let $\underline{\mathbf{X}}''$ be a system identical to $\underline{\mathbf{X}}'$, except that $m = 1$ in the system $\underline{\mathbf{X}}''$. By Lemma 4.3.1, there is a coupling between the two systems $\underline{\mathbf{X}}'$ and $\underline{\mathbf{X}}''$ such that, for all i and t, $\sum_{r=1}^{N} X'_{i,r}(t) \le \sum_{r=1}^{N} X''_{i,r}(t)$. Using the symmetry among the channels at a node of $\underline{\mathbf{X}}'$ or of $\underline{\mathbf{X}}''$, we then have, for all i, r, t,

$$\mathbb{E}[X'_{i,r}(t)] \le \mathbb{E}[X''_{i,r}(t)]. \tag{4.9}$$

Now note that $\underline{\mathbf{X}}''$ may be regarded as a collection of J independent standard Jackson networks, each with N nodes. By Theorem 4.2.1(i), for each i, r, x, t,

$$\mathbb{P}(X''_{i,r}(t) \ge x) \le \mathbb{P}(W_i(t) \ge x),$$

where W_i is an $M/M/1$ queue with arrival rate $\lambda_i + \sum_{j\ne i}\mu_j p_{ji} + \mu_i p_{ii}$, service rate μ_i, and with $W_i(0)$ distributed as $\max_{r'} X_{i,r'}(0)$. Then $\mathbb{E}[W_i(0)] \le \sum_{r'}\mathbb{E}[X_{i,r'}(0)] < \infty$, so, by (4.4),

$$\lim_{t\to\infty} t^{-1}\mathbb{E}[X''_{i,r}(t)] \le \lim_{t\to\infty} t^{-1}\mathbb{E}[W_i(t)] \\ = \lambda_i + \sum_{j\ne i}\mu_j p_{ji} + \mu_i p_{ii} - \mu_i \\ = \tilde{\rho}_i - \mu_i. \tag{4.10}$$

Combining (4.5)-(4.10) gives the required result. □

Chapter 5

Jackson Networks on Countably Infinite Graphs

5.1 Introduction and Preliminaries

In this chapter we return to the original Jackson network setting in which each node of the network contains a single server, but now consider situations in which the set S of nodes is countably infinite.

This setting provides several interesting questions which are straighforward in the case of standard Jackson networks on a finite set S. Firstly, the Markov process representing the evolution of the network may not always be uniquely defined, due to phenomena such as "influence from infinity". Secondly, even when the process can be uniquely constructed, the equilibrium distibution for the network is not necessarily unique; a solution to the network throughput equation provides an invariant product distribution, as in the case where S is finite, but now the throughput equation may have more than one solution, and it is not always guaranteed that every invariant distribution will be a mixture of such product distributions.

We look in particular for conditions determining when these uniqueness properties hold, and, in the case where the network equilibrium is not unique, we investigate the structure of the set of invariant distributions, and their basins of attraction.

The network is described by the parameters $\boldsymbol{\lambda}, \boldsymbol{\mu}$ and $\mathbf{P}$. As before, particles arrive at node $j \in S$ from outside the network as a Poisson process of rate λ_j. The single server at node $j \in S$ serves at rate μ_j when the node is non-empty, and after being served at node j a particle moves to node $k \in S$ with probability p_{jk}, and leaves the network entirely with probability $p^* = 1 - \sum_k p_{jk}$. We will assume throughout that $\sum_k \mu_k p_{kj} < \infty$ for all j, in order to exclude the possibility that the total arrival rate at a node ever becomes infinite.

The model was studied by Kel'bert, Kontsevich and Rybko [30], who treated the case where μ_j is constant in j and $\mathbf{P}$ is doubly substochastic (i.e. $\sum_{k\in S} p_{kj} \leq 1$ for all j as well as $\sum_{k \in S} p_{jk} \leq 1$), finding conditions for the system to have a unique invariant product distribution, and proving convergence to this distribution from any initial conditions. See also [29], where a Central Limit Theorem for the long-term occupancy levels of the nodes is proved for networks of this type started in a (not necessarily unique) invariant product distribution.

A closely related model is the *zero-range process* introduced by Spitzer in [45]. The model is of a closed network (i.e. in our notation $\boldsymbol{\lambda} = \mathbf{0}$ and $\mathbf{p}^* = \mathbf{0}$), and the service rate at node j does not depend on j, but may depend on the number of particles currently occupying node j. Product form equilibria are available as in the Jackson network case, and many of the same questions concerning the classification of the invariant distributions arise. Early work on the construction of the process was done by Holley [19] and Liggett [37], and the structure of the set of invariant distributions was described in a range of cases by Andjel [2], whose methods we draw on during this chapter. (See also [55] for the case where $\mathbf{P}$ is a positive recurrent matrix). The model was extended by Cocozza-Thivent [9] and by Andjel, Cocozza-Thivent and Roussignol [3] who introduce the so-called *misanthrope process*; here certain closed translation-invariant networks are considered in which the rate of transfers from one node to another is a function both of the occupancy number of the departure node (in which it is an increasing function) and of the occupancy number of the destination node (in which it is decreasing).

Recently a good deal of work has been done on convergence to equilibrium (or to saturation in overloaded networks) for the zero-range process and related models. In particular, the *hydrodynamical limit* for translation-invariant systems is considered, involving LLN-type or CLT-type results in which both space and time are rescaled. For example, Ferrari, Presutti and Vares [14] study a zero-range process on $\mathbb{Z}$ where $\mathbf{P}$ is symmetric and nearest-neighbour, for which the initial distribution is drawn from one equilibrium on the positive half-line and another on the negative half-line; they identify convergence to a limiting distribution of the configuration in the vicinity of site xn at time tn^2, as $n \to \infty$, in terms of a non-linear diffusion equation in x and t. In [6], similar results are obtained for asymmetric $\mathbf{P}$ (see also [4]), while Srinivasan in [46] considers a series of queues in tandem, which are all empty at time 0. Landim and others, in [7], [35] and [36], derive large deviations principles for hydrodynamical limits for a range of such processes in $\mathbb{Z}$ and $\mathbb{Z}^d$.

As a general reference on the theory of interacting particle systems, to which all these models belong, Liggett's book [38] is invaluable; for a recent overview of the field, see also [39] by the same author, although this concentrates mainly on models such as the contact process, the simple exclusion process and the voter model, in which the state space is naturally compact.

We now make some preliminary definitions and comments converning the Jackson network model on the infinite set S which will be the subject of the results of this chapter. Consider the random walk $(X_n, n \geq 0)$ on S with transition matrix $\mathbf{P}$, which moves from node j to node k with probability p_{jk}, and which is "cut off", i.e. leaves the network, never to return, after leaving node j, which probability p_j^*. For $j, k \in S$, write $j \Rightarrow k$ if there exists $r > 0$ such that

$$\mathbb{P}(X_r = k | X_0 = j) > 0.$$

Some kind of irreducibility condition on $\mathbf{P}$ is certainly desirable, and it is often assumed that $j \Rightarrow k$ for all $j, k \in S$. However, this is more restrictive than we require, and excludes some natural models that we would like to include, (for example, a collection of queues in series). Instead we will assume:

(i) for all $j, k \in S$ with $j \neq k$, either $j \Rightarrow k$ or $k \Rightarrow j$, and

(ii) for all $j \in S$: if $\lambda_j > 0$, or if $\exists i$ such that $i \Rightarrow j$ and $j \not\Rightarrow i$, then either:

  (a) $p_j^* > 0$, or

  (b) $\exists k$ such that $j \Rightarrow k$ and either $k \not\Rightarrow j$ or $p_k^* > 0$.

These conditions ensure that we may properly describe the matrix $\mathbf{P}$ as positive recurrent, or null recurrent, or transient, since if one node (considered as a state of the random walk according to $\mathbf{P}$) has such a property, then it is shared by every other node. If $\mathbf{P}$ is recurrent then we must have $p_j^* = 0$ for all $j \in S$, and from our irreducibility conditions it also follows that $\lambda_j = 0$ for all $j \in S$; (in any case if we had $\lambda_j > 0$ for any $j \in S$ then the network would inevitably become saturated). In the case where $\mathbf{P}$ is transient, both open and closed networks may be interesting.

If $\mathbf{P}$ is recurrent then all the entries of $\sum_{n=0}^{\infty} \mathbf{P}^n$ are infinite, and there exists an essentially unique (up to multiplication by a constant) non-zero vector $\boldsymbol{\delta}$ such that $\delta_j = \sum_{k \in S} \delta_k p_{kj}$ for all $j \in S$. (See for example [32], Chapter 5). $\mathbf{P}$ is then positive recurrent if $\sum_{j \in S} \delta_j < \infty$, and null recurrent otherwise.

We will make particular use of what will be referred to below as the $(\mathbf{P}, \boldsymbol{\mu})-$*random walk*; this is the continuous time Markov chain on S whose jump chain is a random walk according to $\mathbf{P}$ and whose holding times at node j are exponential with rate μ_j. Of course, the $(\mathbf{P}, \boldsymbol{\mu})$-random walk is recurrent exactly when $\mathbf{P}$ is recurrent; it is then positive recurrent if the $\boldsymbol{\delta}$ above has $\sum_{j \in S} \delta_j / \mu_j < \infty$ and null recurrent if $\sum_{j \in S} \delta_j / \mu_j = \infty$

In Section 5.2 we focus on the construction of the process, identifying conditions under which the dynamics of the system specify a unique process, and giving examples of how non-uniqueness can arise when these conditions do not hold. In Section 5.3 we study the invariant product distributions, identifying, in certain situations, conditions for the network throughput equation to have a unique solution. In Section 5.4 we consider the entire set of invariant distributions for the network; when $\mathbf{P}$ is recurrent this may be described completely, but in the transient case only partial results are produced.

5.2 Construction of the Process

When S is a finite set, the set $\mathbb{Z}_+^S$ of network configurations is countable, and the countable state-space Markov process describing the evolution of the network may be uniquely constructed in a standard way, for example using the jump-chain/holding-time representation. When S is infinite, however, $\mathbb{Z}_+^S$ is uncountable, and the states of the process will normally be instantaneous. We can no longer rely in general on the specified dynamics of the system yielding a unique process: if certain network parameters grow unboundedly, then "influence from infinity" can give rise to non-uniqueness of the realisation of the process.

Any time-homogeneous Markov process representing the evolution of the system can be characterised by giving, for each initial configuration $\boldsymbol{\eta}(0) \in \mathbb{Z}_+^S$ and time t, and each collection $j_1, j_2 \ldots, j_r \in S$ and $x_1, x_2, \ldots, x_r \in \mathbb{Z}_+$, the probability of the increasing cylinder set

$$\{\eta_{j_1}(t) \geq x_1, \ldots, \eta_{j_r}(t) \geq x_r\}. \tag{5.1}$$

More generally, if π and π' are measures on $\mathbb{Z}_+^S$, we will say that π' *majorises* π on a finite set $\Delta \subset S$, and write $\pi \preceq_\Delta \pi'$, if $\pi(\Gamma) \leq \pi'(\Gamma)$ for any increasing set Γ which depends only on coordinates in Δ. (Any such Γ may be written as a finite union of sets of the form (5.1) with $j_1, \ldots, j_r \in \Delta$). If $\pi \preceq_\Delta \pi'$ for all finite $\Delta \subset S$, we say simply that π' majorises π, and write $\pi \preceq \pi'$. This is equivalent to the existence of a joint distribution $\overline{\pi}$ on $\mathbb{Z}_+^S \times \mathbb{Z}_+^S$ which has marginals π and π' and for which $\overline{\pi}(\{(\boldsymbol{\eta}, \boldsymbol{\xi}) : \eta_i \leq \xi_i \ \forall i \in S\}) = 1$ – see [27] for details. Note that this is stronger than the condition that all sets of the form (5.1) have at least as great probability under π' as under π, which corresponds to the relation $\leq_d$ as discussed in Chapter 4.

We will say that one process majorises another on a set $\Delta \subset S$ if, for all $t \geq 0$, the distribution of the first process at time t majorises that of the second process at time t, on Δ.

For a finite subset Λ of S, we will make frequent use of the finite volume processes on Λ with *zero boundary conditions* and with *saturated boundary conditions*, in order to provide lower and upper bounds for the paths in Λ of any realisation of the infinite volume Jackson network process. The process on Λ with saturated boundary conditions is obtained by letting every node outside Λ have an infinitely long queue for all time; such a node $k \in S \setminus \Lambda$ produces an uninterrupted output stream of rate μ_k. This yields a Jackson network process on the nodes of Λ with new parameters $\tilde{\mu}_j = \mu_j$, $\tilde{\lambda}_j = \lambda_j + \sum_{i \in S\setminus\Lambda} \mu_i p_{ij}$, $\tilde{p}_{jk} = p_{jk}$ and $\tilde{p}_j^* = p_j^* + \sum_{i \in S\setminus\Lambda} p_{ji}$ for $j, k \in \Lambda$. The finite volume process on Λ with zero boundary conditions is obtained by fixing $\eta_i(t) = 0$ for all $j \in S \setminus \Lambda$, $t \geq 0$. Thus no transfer ever occurs from a node in $S \setminus \Lambda$ to a node in Λ, and any particle transferring from a node in Λ to one outside Λ is deemed to have left the network. This yields a Jackson network process inside Λ with parameters $\tilde{\mu}_j = \mu_j$, $\tilde{\lambda}_j = \lambda_j$, $\tilde{p}_{jk} = p_{jk}$ and $\tilde{p}_j^* = p_j^* + \sum_{i \in S\setminus\Lambda} p_{ji}$ for $j, k \in \Lambda$.

One may then see, by a simple coupling, that any process realising the infinite volume Jackson network majorises, on any $\Delta \subset \Lambda$, the finite volume process on Λ with zero boundary conditions. Similarly, any such process is majorised on any $\Delta \subset \Lambda$ by the finite volume process on Λ with saturated boundary conditions.

The approach to proving uniqueness is to fix a finite set Δ and to analyse the difference between the behaviour on Δ of the finite volume processes on Λ with saturated and with zero boundary conditions as Λ becomes large. In fact we aim to couple these two processes such that for any fixed t they are identical on Δ during $[0,t]$ with a probability which tends to 1 as $\Lambda \uparrow S$. If this works for all Δ, this is enough to give uniqueness of the process, since the probabilities of all sets of the form (5.1) are specified.

Theorem 5.2.1 provides two alternative sufficient conditions for the system to have a unique realisation. It will be seen that boundedness of λ_j or of μ_j are neither necessary nor sufficient for uniqueness: rather one needs some control over the quantities $\sum_{k\in S}\mu_k p_{kj}$ representing the total possible amount of flow to a node j from within the network. After the proof of the theorem some examples are given to illustrate these conditions and ways in which non-uniqueness can occur if they do not hold; we also see that the given conditions take advantage of particular features of the Jackson network model and are not strong enough to guarantee uniqueness for other related systems, such as a simple-exclusion-type process, sharing the same system parameters. We also describe some approaches to defining the process in a natural way in situations where non-uniqueness occurs, by confining the process to a restricted part of the subspace.

Note: A more technical approach to existence and uniqueness of the process is via the *martingale problem*. We can specify a generator Ω for the process, which acts on functions defined on the state space of the process. In order to provide good boundedness properties, it will be convenient to compactify the space $\mathbb{Z}_+$ by adding the point at infinity, corresponding to the situation where a node has an infinitely long queue and is saturated. The state space of the process may then be taken to be $(\mathbb{Z}_+\cup\{\infty\})^S$, which is also compact under the usual product topology which we will assume from here on whenever referring to the space. The domain $\mathcal{C}$ of the generator can then be defined as the set of continuous functions from $(\mathbb{Z}_+\cup\{\infty\})^S$ to $\mathbb{R}$, and the generator Ω may be given by:

$$(\Omega f)(\boldsymbol{\eta}) = \sum_{j\in S}\mu_j p_j^*[f(\boldsymbol{\eta}^{(j\infty)}) - f(\boldsymbol{\eta})] + \sum_{j\in S}\lambda_j[f(\boldsymbol{\eta}^{(\infty j)}) - f(\boldsymbol{\eta})] + \sum_{j,k\in S}\mu_j p_{jk}[f(\boldsymbol{\eta}^{(jk)}) - f(\boldsymbol{\eta})], \quad (5.2)$$

for all $f \in \mathcal{C}, \boldsymbol{\eta} \in \mathbb{Z}_+^S$.

Here $\boldsymbol{\eta}^{(j\infty)}, \boldsymbol{\eta}^{(\infty j)}$ and $\boldsymbol{\eta}^{(jk)}$ are defined by

$$\eta_i^{(j\infty)} = \eta_i - \mathbf{I}(\eta_j \geq 1)\mathbf{I}(i=j),$$
$$\eta_i^{(\infty j)} = \eta_i + \mathbf{I}(i=j),$$

and

$$\eta_i^{(jk)} = \eta_i + \mathbf{I}(\eta_j \geq 1)[\mathbf{I}(i = k) - \mathbf{I}(i = j)], \tag{5.3}$$

for all i, j and $k \in S$.

A process is then said to solve the martingale problem for Ω if

$$f(\boldsymbol{\eta}(t)) - \int_0^t (\Omega f)(\boldsymbol{\eta}(s))ds$$

is a martingale with respect to the natural filtration, for all $f \in \mathcal{C}$. See for example [38] for details. In our situation we can show the existence of maximal and minimal solutions (in the sense of majorisation described above) to the martingale problem for Ω by taking the limit as $\Lambda \uparrow S$ of the saturated and zero boundary condition processes on Λ: if these two coincide, as for example under the conditions of Theorem 5.2.1, then a unique solution exists as desired (and the martingale problem is said to be *well-posed*).

Theorem 5.2.1 *If*

$$\sum_{k \in S} \mu_k p_{kj} \leq c \textit{ for all } j \in S, \textit{ for some } c \in \mathbb{R}, \tag{5.4}$$

or if

$$\frac{\sum_{k \in S} \mu_k p_{kj}}{\mu_j} \leq 1 \textit{ for all } j \in S, \tag{5.5}$$

then the process is uniquely defined.

Proof: For any finite set of nodes $\Lambda \subset S$, we can couple the processes on Λ with saturated and with zero boundary conditions as follows. Distribute red particles within Λ according to the initial state $\boldsymbol{\eta}(0)$. Let red particles arrive from outside the system to node $j \in \Lambda$ at rate λ_j, and let black particles arrive from outside the system to node $j \in \Lambda$ at rate $\sum_{k \in S \setminus \Lambda} \mu_k p_{kj}$. Serve within Λ as normal: a particle leaving node $j \in \Lambda$ goes to node $k \in \Lambda$ with probability p_{jk} and leaves the system with probability $p_j^* + \sum_{k \in S \setminus \Lambda} p_{jk}$. But give red particles priority at all times, so that a black particle never leaves a node in Λ until all the red particles there have already left.

Under this scheme, the red particles behave as a system in Λ with zero boundary conditions, and the red and black particles together constitute a system in Λ with saturated boundary conditions. Thus if, for all fixed $t > 0$ and $j_0 \in S$, we can show that the expected number of black particles at node j_0 at time t tends to 0 as $\Lambda \uparrow S$, then this implies the uniqueness of the process as required.

First suppose that (5.4) holds. With any black particle which arrives at j_0 before time t, we can associate some $n \geq 1$ and collections $j_1, \ldots, j_{n-1} \in \Lambda$ and $0 < t_n < \ldots < t_1 < t$

such that at time t_n a black particle arrives from outside Λ at node j_{n-1}, and at time t_i for $1 \le i \le n-1$, a transfer occurs from node j_i to node j_{i-1}.

Now black particles arrive from outside Λ at node j_{n-1} at rate $\sum_{k\in S\setminus\Lambda} \mu_k p_{kj_{n-1}}$, and potential transfers from j_i to j_{i-1} occur at rate $\mu_{j_i} p_{j_i j_{i-1}}$. So the expected number of black particles to have arrived at j_0 before time t cannot be greater than

$$\sum_{k\in S\setminus\Lambda} \sum_{n=1}^{\infty} \sum_{j_1,\dots,j_{n-1}\in S} \int_{0<s_n<\dots<s_1<t} \mu_k p_{kj_{n-1}} \mu_{j_{n-1}} p_{j_{n-1}j_{n-2}} \dots \mu_{j_1} p_{j_1 j_0} ds_1 \dots ds_n,$$

which is

$$\sum_{k\in S\setminus\Lambda} \sum_{n=1}^{\infty} \sum_{j_1,\dots,j_{n-1}\in S} \frac{t^n}{n!} \mu_k p_{kj_{n-1}} \mu_{j_{n-1}} p_{j_{n-1}j_{n-2}} \dots \mu_{j_1} p_{j_1 j_o}. \tag{5.6}$$

But, from (5.4), the expression in (5.6), summed over all $k \in S$ rather than $k \in S \setminus \Lambda$, can be seen to be no greater than $\sum_{n=1}^{\infty} \frac{t^n}{n!} c^n$, which is $e^{ct} - 1$; hence the sum (5.6) over $k \in S \setminus \Lambda$ must tend to 0 as $\Lambda \uparrow S$, as required.

Alternatively, suppose that (5.5) holds. If in fact $\sup_{j\in S} \sum_{k\in S} \mu_k p_{kj}/\mu_j < 1$ then the proof of uniqueness is straightforward, but more care is needed to include the case when the supremum is equal to 1.

As soon as a particle enters the system (or at time 0 if it is already there), let us fix the future route that it will take through the network. Then we can associate with every particle in the system a "level", namely the number of jumps which that particle will need in order to reach j_0. (If in fact it will never reach j_0, this number can be taken to be ∞). The level of a particle decreases by 1 every time it makes a jump.

We will show firstly that, for any n, the expected number of black particles entering Λ with level n during $[0, t]$ tends to 0 as $\Lambda \uparrow S$; and secondly that the expected number of particles which have level n or above at or after time 0, and which visit j_0 during $[0, t]$, tends to 0 as $n \to \infty$. We can then deduce that the expected number of black particles to reach j_0 during $[0, t]$ tends to 0 as $\Lambda \uparrow S$.

The times at which black particles enter Λ with level n form a Poisson process with rate

$$\sum_{k\in S\setminus\Lambda} \sum_{j_1,\dots,j_n\in\Lambda} \mu_k p_{kj_n} p_{j_n j_{n-1}} \dots p_{j_1 j_0}. \tag{5.7}$$

We have

$$\begin{aligned} &\sum_{k\in S} \sum_{j_1,\dots,j_n\in S} \mu_k p_{kj_n} p_{j_n j_{n-1}} \dots p_{j_1 j_0} \\ &\qquad = \sum_{k\in S} \sum_{j_1,\dots,j_n\in S} \frac{\mu_k p_{kj_n}}{\mu_{j_n}} \frac{\mu_{j_n} p_{j_n j_{n-1}}}{\mu_{j_{n-1}}} \dots \frac{\mu_{j_1} p_{j_1 j_0}}{\mu_{j_0}} \mu_{j_0} \\ &\qquad \le \mu_{j_0}, \end{aligned} \tag{5.8}$$

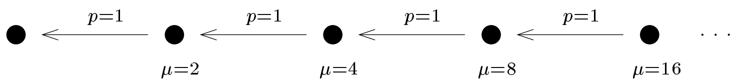
using (5.5). Thus the expression (5.7), which is (5.8) with the summation over S replaced by a summation over $S \setminus \Lambda$ and with $j_1, \dots, j_n$ restricted to Λ rather than being taken from the whole of S, must converge to 0 as $\Lambda \uparrow S$. Thus the expected number of black particles entering Λ from outside Λ with level n during $[0, t]$ tends to 0 as $\Lambda \uparrow S$ for fixed t, as required.

Now we consider the set of times at which a particle of level n is served somewhere in Λ, and hence moves down a level. This set of times is a subset of a Poisson process of rate

$$\sum_{j_1, \ldots, j_n \in S} \mu_{j_n} p_{j_n j_{n-1}} \cdots p_{j_1 j_0},$$

which is no greater than μ_{j_0} as before. Hence, for a single particle whose level is n or above during $[0, t]$ to reach j_0 before time t, we require a sequence of n events each of rate no greater than μ_{j_0} to occur one after the other during the interval $[0, t]$. The probability that this happens tends to 0 as $n \to \infty$, which completes the proof. $\square$

Example 5.2.2 Let $S = \mathbb{Z}_+$, let $\boldsymbol{\lambda} = \mathbf{0}$, let $\mu_0 = 0$, and $\mu_j = 2^j$ for all $j \geq 1$, and let $p_{j(j-1)} = 1$ for all $j \geq 1$.

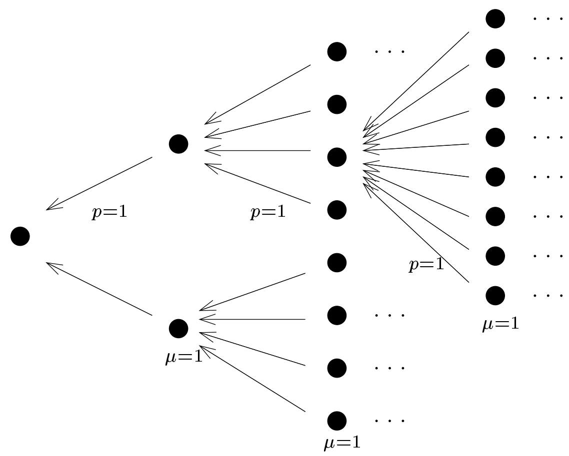


This is one of the simplest examples of non-uniqueness arising from "influence from infinity". Suppose that the system is empty at time 0: $\boldsymbol{\eta}(0) = \mathbf{0}$. Then, under the process obtained by taking the limit of the zero boundary condition processes, the system will certainly remain empty for all time. However, if a single particle is introduced at node n, it will certainly reach node 0, and after a time whose expectation is bounded above by 1 for all n. So, under the process obtained as the limit of the saturated boundary condition processes, we have for example $\mathbb{P}(\eta_0(1) > 0) > 0$. Notice here that $\sum_{k \in S} \mu_k p_{kj} / \mu_j$ is bounded above, but not by 1 as required for (5.5).

Example 5.2.3 Let S consist of countably many levels, $0, 1, 2, \ldots$, and let the nth level have $2^{n(n+1)/2}$ nodes. Every node at the nth level has 2^{n+1} "ancestors" at the $(n+1)$st level, and receives all the particles that its ancestors serve. Let the service rate be 1 at every node, and let the external arrival rate be 0 everywhere.

Here again we have non-uniqueness, although μ_j and λ_j are both bounded. For instance, let the network be empty at time 0. As before, under the process obtained by taking the limit of the zero boundary condition processes, the network stays empty for all time. However, if we introduce a particle to every node of level n, the expected time until every node of level $n-1$ has had an arrival is less than $(n+1)(\ln 2)/(2^n)$. We have $\sum_{n=1}^{\infty} (n+1)/(2^n) < \infty$, so, under the process obtained by taking the limit of the saturated boundary condition processes, particles will reach level 0 "from infinity" in a finite time with probability one.

This example is a special case of the zero-range process described in the introduction to the chapter. Recall that for these processes, the service rate does not depend on the node but may depend on the number of particles currently present in the node. Let $g(r), r \geq 1$ be the



service rate at a node containing r particles. If g is unbounded, questions about the existence as well as the uniqueness of the process may not be trivial, since one needs to exclude the possibility of infinitely many particles arriving at a single node in a finite time. Andjel, in [2], considers zero range processes for which the function g satisfies

$$\sup_{r \geq 1} |g(r+1) - g(r)| < \infty.$$

Let α be a strictly positive real function on S such that

$$\sum_{k \in S} p_{jk} \alpha(k) \leq M \alpha(j)$$

for all $j \in S$, for some M. For example in this case we can take

$$\alpha(j) = M^{-l(j)},$$

where $l(j)$ is the level of node j. Define the set

$$H = \{\boldsymbol{\eta} : \sum_{j \in S} g(\eta_j) \alpha(j) < \infty\}.$$

(In our example we have $g(r) = 1$ for all $r \geq 1$). Then Andjel's construction shows that if the initial configuration $\boldsymbol{\eta}(0)$ is restricted to the set H, there exists a unique process with the specified dynamics under which $\boldsymbol{\eta}(t)$ remains within H for all $t > 0$ with probability 1.

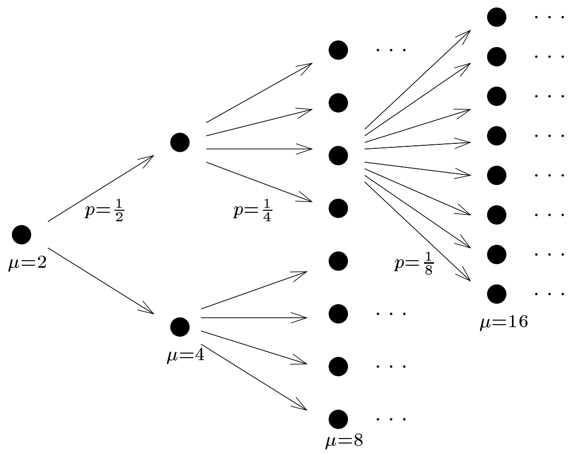
This approach can be extended to our setting in cases where μ_j is bounded and $\boldsymbol{\lambda}$ satisfies certain conditions; but it does not extend well in general to cases where μ_j is unbounded, since then influence from infinity may result from the movement of only finitely many particles, as in Example 5.2.2, rather than requiring the involvement of unboundedly many as in Example 5.2.3.

Contrastingly, it is sometimes possible to guarantee uniqueness by insisting that the network is sufficiently busy rather than that it is sufficiently empty. For example, suppose that in Example 5.2.2, $\eta_j(0) = 4^j$ for all $j \in S$. Then under any possible realisation of the process,

$$\mathbb{P}(\eta_j(t) \to \infty \text{ as } j \to \infty \text{ for all } t) = 1,$$

and the ensuing traffic jam excludes any possible influence from infinity.

Example 5.2.4 We reverse the directions of the arrows in Example 5.2.3, while keeping the maximum rate of flow along each link the same as before. Now each node at level n serves at rate 2^{n+1}, and sends its output with equal probability to one of its 2^{n+1} "descendant nodes" at level $n+1$.



Here both (5.4) and (5.5) are fulfilled, and the process has a unique realisation; this realisation is easy to construct recursively since the behaviour of a node at level n only depends on that of the nodes in levels $0, 1, \ldots, n-1$.

We can associate with the same parameters a *simple exclusion*-type system: the dynamics are the same except that each node may contain at most one particle at any time, and any particle which tries to move to an already occupied node is unable to do so and remains where it is. In this case there may not be a unique process associated with the system. For example, if every node is occupied at time 0, it is possible either for the system to remain gridlocked everywhere for all time (under the limit of the *occupied boundary condition* processes) or for gradual rightwards movement to occur in finite time (under the limit of the zero boundary condition processes).

The difference between the two cases is that in the models without "blocking" that we consider, the rate of flow from one node to another is determined only by the occupancy number of the departure node and is not affected by the occupancy of the destination node. Loosely speaking, we can say that influence in the network is only felt in the direction of the arrows, so that the conditions (5.4) and (5.5), which control the maximum amount of flow into a node from the other nodes inside the network, but not the maximum amount of flow out of a node into other nodes inside the network, are sufficient to give uniqueness.

5.3 Invariant Product Distributions

From now on we will assume that the martingale problem for the generator Ω in (5.2) is indeed well-posed, so that a unique Markov process representing the evolution of the state of the network may be identified. We now turn our attention to the invariant distributions for the process.

It follows from the results of Echeverría in [12] that a distribution π is invariant if and only if $\int \Omega f(\boldsymbol{\eta}) d\pi(\boldsymbol{\eta}) = 0$ for all $f \in \mathcal{C}$. (See also the discussion in [38], pp. 51-53).

Although we will be particularly interested in what we will call *non-degenerate* distributions, concentrated on $\mathbb{Z}_+^S$, we will also consider more general distributions on $(\mathbb{Z}_+ \cup \{\infty\})^S$ under which one or more nodes of the network have a non-zero probability of being saturated.

We will say that a distribution π on $(\mathbb{Z}_+ \cup \{\infty\})^S$ is a *limiting distribution* if there exists a non-degenerate initial distribution for $\boldsymbol{\eta}(0)$ such that the distribution of $\boldsymbol{\eta}(t)$ converges weakly to π as $t \to \infty$. (Note that weak convergence in the space of measures on $(\mathbb{Z}_+ \cup \{\infty\})^S$ is equivalent to the convergence of all the probabilities of increasing cylinder sets of the form (5.1)). Certainly every limiting distribution is invariant, (see for example [38], p.10), and clearly every non-degenerate invariant distribution is a limiting distribution (merely by taking $\boldsymbol{\eta}(0)$ to have that distribution). The aim of the definition is to exclude physically unnatural cases in which a node is deemed to be saturated but is not in fact overloaded; as an extreme example, the trivial distribution under which every node is saturated with probability 1 is invariant for every network, but it is not a limiting distribution unless every node of the network is indeed inevitably overloaded.

In this section we investigate invariant product distributions for the network; that is, invariant distributions under which the queue lengths at different nodes are independent.

Let $\boldsymbol{\rho}$ be a non-negative vector indexed by the nodes of S, satisfying the non-linear *generalised throughput equation* for the network:

$$\rho_j = \lambda_j + \sum_{k \in S} \min(\mu_k, \rho_k) p_{kj} \text{ for all } j \in S. \tag{5.9}$$

Let $\pi_{\boldsymbol{\rho}}$ be the probability distribution on $(\mathbb{Z}_+ \cup \{\infty\})^S$ specified by the following probabilities of increasing cylinder sets:

$$\pi_{\boldsymbol{\rho}}(\eta_j \geq r_j \forall j \in J) = \prod_{j \in J} \left[\min\left(\frac{\rho_j}{\mu_j}, 1 \right) \right]^{r_j} \tag{5.10}$$

for all finite subsets J of S, and all $\mathbf{r} \in \mathbb{Z}_+^J$.

Then $\pi_{\boldsymbol{\rho}}$ is a product distribution, under which the marginal distribution at node j is geometric with parameter (ρ_j/μ_j) whenever $\rho_j < \mu_j$, and under which any node j for which $\rho_j \geq \mu_j$ is saturated with probability 1.

From the form (5.2) of the generator of the process, one can show that $\pi_{\boldsymbol{\rho}}$ is an invariant distribution, as in the case where S is finite. When S is finite and $\boldsymbol{\lambda} \neq \mathbf{0}$, the throughput equation (5.9) has a unique solution $\boldsymbol{\rho}$, and the corresponding product distribution $\pi_{\boldsymbol{\rho}}$ is the unique limiting invariant distribution for the network, as shown by Goodman and Massey in [16]. If S is infinite, there may be many solutions to (5.9), whether the network is open or closed. The next result identifies minimal and maximal solutions, which provide bounds on possible limiting invariant distributions for the network:

Proposition 5.3.1 *There exist minimal and maximal non-negative solutions $\boldsymbol{\rho}^0$ and $\boldsymbol{\rho}^\infty$ to the generalised throughput equation (5.9). The corresponding product distributions $\pi_{\boldsymbol{\rho}^0}$ and $\pi_{\boldsymbol{\rho}^\infty}$ are invariant, and if π is any limiting invariant distribution for the network, then $\pi_{\boldsymbol{\rho}^0} \preceq \pi \preceq \pi_{\boldsymbol{\rho}^\infty}$.*

Proof: For each finite subset J of S, let $\boldsymbol{\rho}^{0,J}$ and $\boldsymbol{\rho}^{\infty,J}$ be the solutions to

$$\rho_j^{0,J} = \lambda_j + \sum_{k \in J} \min(\rho_k^{0,J}, \mu_k) p_{kj} \tag{5.11}$$

and

$$\rho_j^{\infty,J} = \lambda_j + \sum_{k \notin J} \mu_k p_{kj} + \sum_{k \in J} \min(\rho_k^{\infty,J}, \mu_k) p_{kj} \tag{5.12}$$

for all $j \in J$.

These correspond to the throughput equations for the Jackson networks on J with zero and with saturated boundary conditions respectively, and, since J is finite, have unique solutions.

Let $J \uparrow S$. As J increases, $\rho_j^{0,J}$ increases and $\rho_j^{\infty,J}$ decreases; both $\rho_j^{0,J}$ and $\rho_j^{\infty,J}$ are bounded below by 0 and above by $\lambda_j + \sum_{k \in S} \mu_k p_{kj}$; so the limits $\boldsymbol{\rho}^0$ and $\boldsymbol{\rho}^\infty$ exist and, since the limits on both sides of (5.11) and of (5.12) are monotone, both $\boldsymbol{\rho}^0$ and $\boldsymbol{\rho}^\infty$ satisfy (5.9).

Thus $\pi_{\boldsymbol{\rho}^0}$ and $\pi_{\boldsymbol{\rho}^\infty}$ are invariant distributions for the network.

Now for any cylinder set A,

$$\pi_{\boldsymbol{\rho}^0}(A) = \lim_{J \uparrow S} \pi_{\boldsymbol{\rho}^{0,J}}(A),$$

by the continuity of the probabilities (5.10) in $\boldsymbol{\rho}$.

But $\pi_{\boldsymbol{\rho}^{0,J}}$ is the unique limiting invariant distribution for the Jackson network on the finite subset J with zero boundary conditions, so if π is any limiting invariant distribution for the whole network, then, by a simple coupling and standard results on convergence to equilibrium of countable state space Markov chains, $\pi_{\boldsymbol{\rho}^{0,J}}$ must be majorised by the restriction of π to the space $(\mathbb{Z}_+ \cup \{\infty\})^J$ of configurations on J. Then, since any distribution is characterised by its values on cylinder sets, π majorises the limiting distribution $\pi_{\boldsymbol{\rho}^0}$. Similarly $\pi \preceq \pi_{\boldsymbol{\rho}^\infty}$, as required. □

In particular we are interested in non-negative solutions to (5.9) for which $\boldsymbol{\rho} < \boldsymbol{\mu}$, (i.e. $\rho_j < \mu_j$ for all $j \in S$), or, equivalently, non-negative solutions to $\boldsymbol{\rho} = \boldsymbol{\lambda} + \boldsymbol{\rho}\mathbf{P}$, i.e.

$$\rho_j = \lambda_j + \sum_{k \in S} \rho_k p_{kj} \text{ for all } j \in S \tag{5.13}$$

for which $\boldsymbol{\rho} < \boldsymbol{\mu}$; these lead to non-degenerate invariant product measures.

If $\mathbf{P}$ is recurrent, then $\boldsymbol{\lambda} = \mathbf{0}$, and $\boldsymbol{\rho}^0 = \mathbf{0}$ is a solution. If $\mathbf{P}$ is transient, then we can form the matrix $\sum_{n=0}^{\infty} \mathbf{P}^n$; then the minimal non-negative solution to (5.13) is given by

$$\boldsymbol{\rho}^0 = \boldsymbol{\lambda} \left(\sum_{n=0}^{\infty} \mathbf{P}^n \right). \tag{5.14}$$

In either case, any non-negative solution to (5.13) has the form

$$\boldsymbol{\rho} = \boldsymbol{\rho}^0 + \boldsymbol{\delta}, \tag{5.15}$$

where $\boldsymbol{\delta}$ is non-negative and satisfies $\boldsymbol{\delta} = \boldsymbol{\delta}\mathbf{P}$. (See for example [32] for details).

For the rest of this section, we will assume that $\boldsymbol{\rho}^0 < \boldsymbol{\mu}$, so that the two definitions of $\boldsymbol{\rho}^0$ in Proposition 5.3.1 and in (5.14) agree, and so that at least one non-degenerate invariant product distribution exists. (Note also that if this is not the case, then in fact no non-degenerate invariant distribution exists at all, by Proposition 5.3.1).

Extending the methods of [30], we can now find some situations in which convenient conditions can be given for the uniqueness or otherwise of this non-degenerate invariant product distribution. From (5.15), the distribution is unique if and only if there exists no non-negative solution to $\boldsymbol{\delta} = \boldsymbol{\delta}\mathbf{P}$ which is $\boldsymbol{\beta}$*-tempered*, i.e. for which δ_j/β_j is bounded, where $\boldsymbol{\beta} = \boldsymbol{\mu} - \boldsymbol{\rho}^0$.

Putting $\sigma_j = \delta_j/\beta_j$, we then require bounded solutions $\boldsymbol{\sigma}$ to $\boldsymbol{\sigma} = \boldsymbol{\sigma}\tilde{\mathbf{P}}$, or

$$\sigma_j = \sum_{k \in S} \sigma_k \tilde{p}_{kj} \tag{5.16}$$

for all $j \in S$, where $(\tilde{\mathbf{P}})_{kj} = \tilde{p}_{kj} = p_{kj}\beta_k/\beta_j$.

Suppose that

$$\sum_{k \in S} p_{kj} \frac{\beta_k}{\beta_j} \leq 1 \tag{5.17}$$

for all $j \in S$. Then the transposed matrix $\tilde{\mathbf{P}}^T$ is sub-stochastic, and we can consider a random walk according to $\tilde{\mathbf{P}}^T$, which moves from node j to node k with probability $p_{kj}\beta_k/\beta_j$, and leaves the network entirely from node j with probability $1 - \sum_{k \in S} p_{kj}\beta_k/\beta_j$. Then, from basic results on absorption probabilities, (see for example [43], p. 112), the probability that the walk leaves the network eventually, starting from node j, is γ_j, where $\boldsymbol{\gamma}$ is the minimal non-negative solution to

$$\gamma_j = \sum_{k \in S} p_{kj} \frac{\beta_k}{\beta_j} \gamma_k + \left(1 - \sum_{k \in S} p_{kj} \frac{\beta_k}{\beta_j} \right) \tag{5.18}$$

for all $j \in S$.

But if $\boldsymbol{\gamma}$ satisfies (5.18) then, putting $\sigma_j = 1 - \gamma_j$, we have that $\boldsymbol{\sigma}$ satisfies (5.16). Thus $\boldsymbol{\sigma} = \mathbf{0}$ is the only non-negative bounded solution to (5.16) if and only if the random walk with transition matrix $\tilde{\mathbf{P}}^T$ leaves the network with probability 1, whatever its starting position.

Equation (5.17) can be expanded to read

$$\sum_{k\in S}\frac{\mu_k-\rho_k^0}{\mu_j-\rho_j^0}p_{kj}\le 1 \text{ for all } j.$$

Since, from (5.14), $\sum_{k\in S}\rho_k^0 p_{kj}=\rho_j^0-\lambda_j$, (5.17) is then equivalent to

$$\sum_{k\in S}\mu_k p_{kj}+\lambda_j\le\mu_j \text{ for all } j. \tag{5.19}$$

So in summary we have the following result:

Proposition 5.3.2 *Suppose $\boldsymbol{\rho}^0<\boldsymbol{\mu}$ and (5.19) holds. Then the product distribution $\pi_{\boldsymbol{\rho}^0}$ is the unique non-degenerate invariant product distribution for the network if and only if the random walk on S with transition matrix $\tilde{\mathbf{P}}^T$ leaves the network with probability 1 from any starting state.*

Corollary 5.4.7 in the next section will show further that if there is a unique non-degenerate invariant product distribution, then it is the unique non-degenerate invariant distribution for the network.

Similar methods to the above can be used to establish conditions like that of Proposition 5.3.2 in other cases in which $\boldsymbol{\rho}^0<\boldsymbol{\mu}$ but (5.19) does not hold. For example, if β_j is bounded above and bounded away from 0, then any bounded solution $\boldsymbol{\sigma}$ to (5.16) corresponds to a bounded solution $\boldsymbol{\delta}$ to $\boldsymbol{\delta}=\boldsymbol{\delta}\mathbf{P}$, and vice versa. In that case we can replace (5.17) by the condition that $\sum_{k\in S}p_{kj}\le 1$ for all $j\in S$, i.e. that $\mathbf{P}$ is doubly substochastic; then, by the same reasoning as above, $\pi_{\boldsymbol{\rho}^0}$ is the unique non-degenerate invariant product distribution for the network if and only if the random walk on S with transition matrix $\mathbf{P}^T$ leaves the network with probability 1 from any starting state.

We now mention some situations in which the set of solutions to the throughput equation (5.13) has a very simple form. Recall from the introduction that if $\mathbf{P}$ is a recurrent matrix then the set of $\boldsymbol{\delta}$ such that $\boldsymbol{\delta}=\boldsymbol{\delta}\mathbf{P}$ is one-dimensional. Let $\boldsymbol{\delta}^0$ be an arbitrary (non-zero) solution. We assume that $\boldsymbol{\lambda}=\mathbf{0}$ (as required for non-overload and for our irreducibility condition when $\mathbf{P}$ is recurrent), and define the set R by

$$R=\{\rho\ge 0:\rho\delta_j^0<\mu_j \text{ for all } j\in S\}. \tag{5.20}$$

Then the set of solutions $\boldsymbol{\rho}$ to (5.13) with $\boldsymbol{\rho}<\boldsymbol{\mu}$ is $\{\rho\boldsymbol{\delta}^0;\rho\in R\}$. R is an interval with 0 as its left endpoint and $\inf_{j\in S}\mu_j/\delta_j^0$ as its right endpoint; it is open on the right if and only if this infimum is obtained for some j. The following result shows that the set of solutions is also one-dimensional if S is $\mathbb{Z}^d$, $\mathbf{P}$ is a translation invariant matrix, and $\boldsymbol{\beta}=\boldsymbol{\mu}-\boldsymbol{\rho}^0$ is bounded:

Proposition 5.3.3 *If $S=\mathbb{Z}^d$, $d\ge 1$, and $\mathbf{P}$ is translation invariant, and $\boldsymbol{\delta}=\{\delta_j,j\in S\}$ is bounded in j and satisfies*

$$\delta_j=\sum_{k\in S}p_{kj}\delta_k \tag{5.21}$$

for all $j\in S$, then $\boldsymbol{\delta}$ is constant in j.

Proof: Since $\mathbf{P}$ is translation invariant and substochastic, we have

$$\begin{aligned}\sum_{k\in\mathbb{Z}^d} p_{kj} &= \sum_{l\in\mathbb{Z}^d} p_{(j-l)l} \\ &= \sum_{l\in\mathbb{Z}^d} p_{0l} \\ &= \sum_{k\in\mathbb{Z}^d} p_{jk} \leq 1\end{aligned}$$

so that $\mathbf{P}^T$ is also translation invariant and substochastic.

Let $X_n, n \geq 0$ on $\mathbb{Z}^d$ be a random walk on $S = \mathbb{Z}^d$ with transition matrix $\mathbf{P}^T$. If $p^* = 1 - \sum_k p_{jk} > 0$ then the random walk leaves the network with probability p^* at each step. Suppose that $\boldsymbol{\delta}$ is bounded in j and satisfies (5.21), and consider the process δ_{X_n}, $n \geq 1$, (where δ_{X_n} is taken to be 0 if the random walk has left the network by time n). Then δ_{X_n} is a bounded martingale, since

$$\begin{aligned}\mathbb{E}(\delta_{X_{n+1}}|X_n = j) &= \sum_k (\mathbf{P}^T)_{jk}\delta_k \\ &= \sum_k p_{kj}\delta_k \\ &= \delta_j.\end{aligned}$$

By the Martingale Convergence Theorem (see for example [59], Chapter 11), $D = \lim_{n\to\infty}\delta_{X_n}$ exists almost surely, and

$$\delta_j = \mathbb{E}(D|X_0 = j).$$

If $p^* > 0$, then $D = 0$ almost surely, since $\mathbb{P}(\delta_{X_n} = 0) \geq 1 - (1-p^*)^n$, so $\delta_j = 0$ for all j.

So assume $p^* = 0$, so that the matrix $\mathbf{P}^T$ is stochastic and the random walk never leaves the network. For any j and $j' \in \mathbb{Z}^d$, let X and X' be two such random walks, with $X_0 = j$, $X'_0 = j'$. By the translation invariance of $\mathbf{P}$, we may couple X and X' so that at each step they make the "same jump": that is,

$$X_n - X'_n = j - j' \in \mathbb{Z}^d \tag{5.22}$$

for all n.

By the irreducibility of $\mathbf{P}$, either $j \Rightarrow j'$ or $j' \Rightarrow j$ under $\mathbf{P}$; without loss of generality assume the second, so that under the random walk according to $\mathbf{P}^T$, there exist $r \in \mathbb{N}$ and $\alpha > 0$ such that

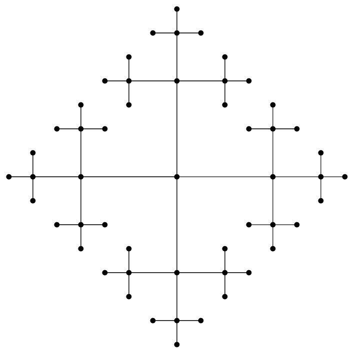
$$\mathbb{P}(X_r = j'|X_0 = j) = \alpha > 0.$$

Then, by translation invariance again, and from (5.22), the event $\{X_{n+r} = X'_n\}$ has probability α for every n, and, using also the Markov property, the events $\{X_r = X'_0\}$, $\{X_{2r} = X'_r\}$, $\{X_{3r} = X'_{2r}\}$, ... are mutually independent. Thus

$$\mathbb{P}(X_{n+r} = X'_n \text{ for infinitely many values of } n) = 1. \tag{5.23}$$

Since $D = \lim_{n\to\infty} \delta_{X_n}$ and $D' = \lim_{n\to\infty} \delta_{X'_n}$ exist almost surely, (5.23) gives $\mathbb{P}(D = D') = 1$. Thus $\mathbb{E}(D) = \mathbb{E}(D')$; but $\delta_j = \mathbb{E}(D)$ and $\delta_{j'} = \mathbb{E}(D')$, so $\delta_j = \delta_{j'}$. Thus $\boldsymbol{\delta}$ is constant in j, as required. □

Remark: The result holds similarly for other regular lattices, but more general, non-lattice, translation-invariant graphs need not have the same property. For example, let S represent the homogeneous tree of degree 4:



etc.

and let $p_{jk} = 1/4$ if j and k are nearest neighbours on the tree, $p_{jk} = 0$ otherwise.

Now identify one particular node of the tree as a root, and define $\boldsymbol{\delta}$ as follows:

- $\delta_j = 1$ if j is the root;
- $\delta_j = \left(\frac{1}{3}\right)^{n(j)}$ where $n(j)$ is the length of the path from the root to j, if this path begins by moving up or down; and
- $\delta_j = 2 - \left(\frac{1}{3}\right)^{n(j)}$ if the path from the root to j begins by moving left or right.

Then δ_j is bounded in j, and one can easily check that $\delta_j = \sum_{k\in S} \delta_k p_{kj}$ for all $j \in S$, but δ_j is not constant in j. Thus for this S and $\mathbf{P}$, there can exist non-translation-invariant non-degenerate invariant product distributions for the network, even when S, $\mathbf{P}$, $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ are translation-invariant. The proof given for Proposition 5.3.3 fails here since the group representing paths between nodes of the network is not commutative.

5.4 The Set of Invariant Distributions

We will denote by $\mathcal{I}$ the set of non-degenerate invariant distributions for the network. In this section we explore the structure of $\mathcal{I}$ in various cases. When $\mathbf{P}$ is recurrent, we are able to describe $\mathcal{I}$ completely; for transient $\mathbf{P}$ the results are much sketchier.

π is an *extreme* member of $\mathcal{I}$ if there do not exist α with $0 < \alpha < 1$, and $\pi_1, \pi_2 \in \mathcal{I}$ with $\pi_1 \neq \pi_2$, such that $\pi = \alpha\pi_1 + (1-\alpha)\pi_2$. We will write $\mathcal{I}_e$ for the set of extreme members of $\mathcal{I}$. $\mathcal{I}$ is convex, and it follows from the Choquet Representation Theorem, (for example [44], pp. 19-20; see also [5]), that $\pi \in \mathcal{I}$ if and only if one may write $\pi = \int \tilde{\pi} d\gamma(\tilde{\pi})$ for some probability distribution γ on $\mathcal{I}_e$. Thus we may concentrate on describing the set $\mathcal{I}_e$, since this essentially provides a description of the whole set $\mathcal{I}$.

$\pi \in \mathcal{I}$ is said to be *ergodic* if, for all $A \subset \mathbb{Z}_+^S$,

$$\lim_{t\to\infty} \int_0^t \frac{\mathbf{I}(\boldsymbol{\eta}(s) \in A)}{t} ds = \pi(A)$$

with probability 1, when $\boldsymbol{\eta}(0)$ is distributed according to π. By arguments similar to those in Theorem 6.10(iii) in [54], for example, one can show that $\pi \in \mathcal{I}$ is ergodic if and only if it is extreme.

The following lemma shows that even if $\pi \in \mathcal{I}$ is not a product distribution, the occupancy probabilities of the various nodes under π (or equivalently the average rates of their arrival or departure processes) can be related to one another via the network throughput equation.

Lemma 5.4.1 *Let $\pi \in \mathcal{I}$, and let $w_j = \pi(\eta_j > 0)$ for $j \in S$. Then, for all $j \in S$, $w_j < 1$ and*

$$w_j \mu_j = \lambda_j + \sum_{k\in S} w_k \mu_k p_{kj}. \tag{5.24}$$

Proof: For any $j \in S$ and $N \in \mathbb{Z}_+$, let $f_j^{(N)} = \min(\eta_j, N)$. Since $\pi \in \mathcal{I}$ and $f_j^{(N)} \in \mathcal{C}$, we have $\int \Omega f_j^{(N)}(\boldsymbol{\eta}) d\pi(\boldsymbol{\eta}) = 0$. From (5.2) this gives

$$\begin{aligned}\big[w_j - \pi(\eta_j > N+1)\big]\mu_j(1-p_{jj}) = \lambda_j \big[1 - \pi(\eta_j > N-1)\big] \\ + \sum_{k\neq j} \big[w_k - \pi(\eta_k > 0, \eta_j > N-1)\big]\mu_k p_{kj}\end{aligned} \tag{5.25}$$

Since π is non-degenerate, $\pi(\eta_j > N) \to 0$ as $N \to \infty$. Then letting $N \to \infty$ and adding $w_j \mu_j p_{jj}$ to each side of (5.25) gives (5.24).

Since π is non-degenerate, we also have $\pi(\eta_j = x) > 0$ for some $x \in \mathbb{Z}_+$. If $x > 0$, then it follows that $\pi(\eta_j = x-1) > 0$, since the rate of departures from j in a configuration where j is non-empty is non-zero, and the total rate at which events involving node j occur is finite. Hence $\pi(\eta_j = 0) > 0$ as required. □

5.4.1 Positive Recurrent Case

We first consider the case where $\mathbf{P}$ is a recurrent matrix. Then $\mathbf{p}^* = \mathbf{0}$, and we also assume that $\boldsymbol{\lambda} = \mathbf{0}$.

Define $\boldsymbol{\delta}^0$ and R as at (5.20), and, for $\rho \in R$, write $\pi_\rho = \pi_{\rho\boldsymbol{\delta}^0}$ as defined by (5.10), so that $\{\pi_\rho, \rho \in R\}$ is the set of invariant product distributions for the network.

The following theorem describes the set of ergodic invariant distributions in the case where $(\mathbf{P}, \boldsymbol{\mu})$ is positive recurrent. This is in fact the only case in which we exhibit ergodic invariant distributions which do not have a product form. The result was proved by Waymire [55] in the case where μ_j is constant in j, (so that $\mathbf{P}$ is a positive recurrent matrix). Note that if $(\mathbf{P}, \boldsymbol{\mu})$ is positive recurrent then $\sum_{j\in S} \delta_j^0/\mu_j < \infty$; this shows that $\inf_{j\in S} \mu_j/\delta_j^0 > 0$, so that R certainly contains points other than 0.

Theorem 5.4.2 *Suppose* $(\mathbf{P}, \boldsymbol{\mu})$ *is positive recurrent. Choose an arbitrary non-zero* $\rho \in R$. *Then* $\mathcal{I}_e = \{\pi_\rho^{(N)}, N \in \mathbb{Z}_+\}$, *where* $\pi_\rho^{(N)}(\,.\,) = \pi_\rho(\,.\,|\sum_{j\in S}\eta_j = N)$. *(This definition is independent of the particular choice of* ρ*).*

Proof: We first show that any $\pi \in \mathcal{I}$ must be concentrated on configurations with a finite number of particles. From Lemma 5.4.1, (with $w_j = \pi(\eta_j > 0)$ as before),

$$w_j \mu_j = \sum_{k\in S} w_k \mu_k p_{kj}$$

for all $j \in S$, so that $\mathbf{w}.\boldsymbol{\mu}$ is a multiple of $\boldsymbol{\delta}^0$, and $\sum_{j\in S} w_j < \infty$. Then

$$\begin{aligned} \mathbb{E}_\pi\left(\#\{j : \eta_j > 0\}\right) &= \textstyle\sum_{j\in S} w_j \\ &< \infty \end{aligned}$$

so that $\pi(\eta_j > 0 \text{ for infinitely many } j) = 0$. Since π is non-degenerate, this implies $\pi\big(\sum_{j\in S}\eta_j < \infty\big) = 1$ as required.

Next note that the dynamics of the system preserve the total number of particles present. Hence, for each N, if π is any invariant distribution with $\pi\big(\sum_{j\in S}\eta_j = N\big) > 0$, then $\pi^{(N)}$ defined by $\pi^{(N)}(\,.\,) = \pi\big(\,.\,|\sum_{j\in S}\eta_j = N\big)$ is also invariant. Thus the $\pi_\rho^{(N)}$, $N \in \mathbb{Z}_+$ are indeed invariant.

Lastly note that, for given N, the network process considered on the state space of configurations with exactly N particles present constitutes an irreducible Markov chain with countable state space, so that at most one invariant distribution can be concentrated on this state space. Hence the $\pi_\rho^{(N)}$ are extreme in $\mathcal{I}$ and are the only non-degenerate ergodic invariant distributions, and are not affected by the particular choice of ρ. $\square$

5.4.2 A Coupling

In this section we introduce a coupling between two copies of the network process, whose initial distributions need not be the same. We will consider the coupling for general $\mathbf{P}, \boldsymbol{\mu}$ and $\boldsymbol{\lambda}$, although we will use it in particular in the next section to investigate the structure of the set $\mathcal{I}$ in the case where $(\mathbf{P}, \boldsymbol{\mu})$ is null recurrent. The technique is an adaptation of that used by Andjel in [2] to treat the case of a zero-range process with null recurrent $\mathbf{P}$.

The coupling is made by specifying that the external arrival points and potential departure points at each node are the same in both copies, and that a departure occurring at the same time at the same node in the two copies has the same destination in each copy.

More formally we may define the coupled process by specifying the following generator (compare with (5.2)):

$$
\begin{aligned}
(\overline{\Omega} f)(\boldsymbol{\eta},\boldsymbol{\xi}) = \sum_{j\in S} \mu_j p_j^* \big[f(\boldsymbol{\eta}^{(j\infty)},\boldsymbol{\xi}^{(j\infty)}) - f(\boldsymbol{\eta},\boldsymbol{\xi})\big] \\
+ \sum_{j\in S} \lambda_j \big[f(\boldsymbol{\eta}^{(\infty j)},\boldsymbol{\xi}^{(\infty j)}) - f(\boldsymbol{\eta},\boldsymbol{\xi})\big] \\
+ \sum_{j,k\in S} \mu_j p_{jk} \big[f(\boldsymbol{\eta}^{(jk)},\boldsymbol{\xi}^{(jk)}) - f(\boldsymbol{\eta},\boldsymbol{\xi})\big].
\end{aligned} \tag{5.26}
$$

Here $\boldsymbol{\eta}^{(jk)}$ and $\boldsymbol{\xi}^{(jk)}$ are defined as in (5.3), and f is a member of $\mathcal{C}^{(2)}$, the set of continuous functions from $\big((\mathbb{Z}_+ \cup \{\infty\})^S\big)^2$ to $\mathbb{R}$. If $\mathbf{P}$ is in fact recurrent then $\mathbf{p}^* = \mathbf{0}$ and $\boldsymbol{\lambda} = \mathbf{0}$, so that the last term is the only non-zero term on the RHS of (5.26).

The coupled process has a unique realisation iff the original network process does. We will denote by $\overline{\mathcal{I}}$ the set of non-degenerate invariant distributions for the coupled process (here a distribution is said to be non-degenerate if both its marginals are non-degenerate), and by $\overline{\mathcal{I}}_e$ the set of extreme members of $\overline{\mathcal{I}}$, (equivalently the ergodic members of $\overline{\mathcal{I}}$, as in the case of $\mathcal{I}$ and $\mathcal{I}_e$).

We write $\boldsymbol{\eta} \le \boldsymbol{\xi}$ if $\eta_j \le \xi_j$ for all $j \in S$.

Lemmas (5.4.3) to (5.4.5) establish some basic properties of the coupling, and are analogous to some of the results in Section 4 of [2].

Lemma 5.4.3
(i) If $\overline{\pi} \in \overline{\mathcal{I}}$ and $0 < \overline{\pi}(\boldsymbol{\eta} \le \boldsymbol{\xi}) < 1$, then also $\overline{\pi}(\,.\,|\boldsymbol{\eta} \le \boldsymbol{\xi}) \in \overline{\mathcal{I}}$ and $\overline{\pi}(\,.\,|\boldsymbol{\eta} \not\le \boldsymbol{\xi}) \in \overline{\mathcal{I}}$.
(ii) If $\overline{\pi} \in \overline{\mathcal{I}}_e$ then either $\overline{\pi}(\boldsymbol{\eta} \le \boldsymbol{\xi}) = 0$ or $\overline{\pi}(\boldsymbol{\eta} \le \boldsymbol{\xi}) = 1$ (and similarly for $\overline{\pi}(\boldsymbol{\eta} \ge \boldsymbol{\xi})$).

Proof: Part (i) follows from the fact that, under the coupling as described, if $\boldsymbol{\eta}(0) \le \boldsymbol{\xi}(0)$ then, with probability 1, $\boldsymbol{\eta}(t) \le \boldsymbol{\xi}(t)$ for all $t \ge 0$.

Then for part (ii) note that if $0 < \overline{\pi}(\boldsymbol{\eta} \le \boldsymbol{\xi}) < 1$, one can write

$$\overline{\pi}(.) = \overline{\pi}(\boldsymbol{\eta} \le \boldsymbol{\xi})\overline{\pi}(\,.\,|\boldsymbol{\eta} \le \boldsymbol{\xi}) + \overline{\pi}(\boldsymbol{\eta} \not\le \boldsymbol{\xi})\overline{\pi}(\,.\,|\boldsymbol{\eta} \not\le \boldsymbol{\xi}),$$

so that $\overline{\pi}$ is not extremal in $\overline{\mathcal{I}}$. □

Lemma 5.4.4
(i) If $\pi_1, \pi_2 \in \mathcal{I}$ then there exists $\overline{\pi} \in \overline{\mathcal{I}}$ with marginals π_1 and π_2.
(ii) If $\pi_1, \pi_2 \in \mathcal{I}_e$ then there exists $\overline{\pi} \in \overline{\mathcal{I}}_e$ with marginals π_1 and π_2.
(iii) If $\pi_1, \pi_2 \in \mathcal{I}$ (respectively $\mathcal{I}_e$) and $\pi_1 \preceq \pi_2$ then there exists $\overline{\pi} \in \overline{\mathcal{I}}$ (respectively $\overline{\mathcal{I}}_e$) with marginals π_1 and π_2 and such that $\overline{\pi}(\boldsymbol{\eta} \le \boldsymbol{\xi}) = 1$.

Proof: Let $\pi_1, \pi_2 \in \mathcal{I}$. Let $(\boldsymbol{\eta}(0), \boldsymbol{\xi}(0))$ be distributed according to $\pi_1 \times \pi_2$ and, for all $t > 0$, define the distributions $\overline{\pi}^{(t)}$ by

$$\overline{\pi}^{(t)}(A) = \frac{1}{t}\int_0^t \mathbb{P}\big((\boldsymbol{\eta}(s), \boldsymbol{\xi}(s)) \in A\big)\, ds$$

for $A \subset ((\mathbb{Z}_+)^S)^2$. Since π_1 and π_2 are invariant, $\overline{\pi}^{(t)}$ has marginals π_1 and π_2 for all t. Take a sequence $t_n \to \infty$. Since the set of measures with marginals π_1 and π_2 is compact, the sequence $\overline{\pi}^{(t_n)}$ has limit points; let $\overline{\pi}$ be such a limit point. Then $\overline{\pi} \in \overline{\mathcal{I}}$ as required for part (i).

Now suppose π_1 and π_2 are invariant and extreme, and let $\mathcal{A}$ be the subset of $\overline{\mathcal{I}}$ formed by measures with marginals π_1 and π_2. By part (i), $\mathcal{A}$ is non-empty, and $\mathcal{A}$ is compact so has an extreme point. But any extreme point of $\mathcal{A}$ must be an extreme point of $\overline{\mathcal{I}}$, since the marginals π_1 and π_2 are extreme in $\mathcal{I}$, giving part (ii).

If $\pi_1 \preceq \pi_2$ then let $(\boldsymbol{\eta}(0), \boldsymbol{\xi}(0))$ be distributed with marginals π_1 and π_2 and such that $\boldsymbol{\eta}(0) \leq \boldsymbol{\xi}(0)$ with probability 1. Then $\overline{\pi}^{(t)}(\boldsymbol{\eta} \leq \boldsymbol{\xi}) = 1$ for all $t \geq 0$, and part (iii) follows by the same arguments as parts (i) and (ii). □

Lemma 5.4.5 *Let $\overline{\pi} \in \overline{\mathcal{I}}$. If*

$$\overline{\pi}(\eta_i > \xi_i = 0, \eta_j < \xi_j) = 0 \tag{5.27}$$

and

$$\overline{\pi}(0 = \eta_i < \xi_i, \eta_j > \xi_j) = 0 \tag{5.28}$$

whenever $p_{ij} > 0$, then

$$\overline{\pi}(\boldsymbol{\eta} \leq \boldsymbol{\xi} \text{ or } \boldsymbol{\xi} \leq \boldsymbol{\eta}) = 1. \tag{5.29}$$

Proof: To show (5.29), we need to show that there do not exist $i, j \in S$ such that $\overline{\pi}(\eta_i > \xi_i, \eta_j < \xi_j) > 0$. From the irreducibility condition on $\mathbf{P}$, it suffices to show that if $n \geq 1$ and $k_0, k_1, \ldots, k_n$ are distinct members of S with $p_{k_r k_{r-1}} > 0$ for $1 \leq r \leq n$, then

$$\overline{\pi}(\eta_{k_0} < \xi_{k_0}, \eta_{k_r} = \xi_{k_r} \text{ for } 1 \leq r \leq n-1, \eta_{k_n} > \xi_{k_n}) = 0 \tag{5.30}$$

and

$$\overline{\pi}(\eta_{k_0} > \xi_{k_0}, \eta_{k_r} = \xi_{k_r} \text{ for } 1 \leq r \leq n-1, \eta_{k_n} < \xi_{k_n}) = 0. \tag{5.31}$$

We will deduce (5.30) from (5.27); (5.31) follows from (5.28) in the same way.

Suppose (5.30) does not hold. Let N be the smallest value of n for which it fails. Then, for some $k_0, k_1, \ldots, k_N$ as above and some $x \in \mathbb{Z}_+$,

$$\overline{\pi}(\eta_{k_0} < \xi_{k_0}, \eta_{k_r} = \xi_{k_r} \text{ for } 1 \leq r \leq N-1, \eta_{k_N} > \xi_{k_N} = x) > 0. \tag{5.32}$$

Imagine that the event described by (5.32) occurs at time 0. The total rate at which network events occur involving nodes $k_0, \ldots, k_N$ is no greater than $\sum_{r=0}^{N} (\lambda_{k_r} + \mu_{k_r})$. Hence the probability that the first x network events to occur at these nodes after time 0 are all transfers from node k_N to node k_{N-1} is at least

$$\left(\frac{\mu_{k_N} p_{k_N k_{N-1}}}{\sum_{r=0}^{N} (\lambda_{k_r} + \mu_{k_r})} \right)^x.$$

If this happens, the resulting configuration has

$$\{\eta_{k_0} < \xi_{k_0}, \eta_{k_r} = \xi_{k_r} \text{ for } 1 \le r \le N-1, \eta_{k_N} > \xi_{k_N} = 0\}. \tag{5.33}$$

Thus, since $\overline{\pi}$ is invariant, the event described by (5.33) must have positive probability under $\overline{\pi}$. This means that $N > 1$, since otherwise (5.27) is contradicted.

But then consider the situation if the $(x+1)$st event is also a transfer from k_N to k_{N-1}. This yields a configuration for which

$$\{\eta_{k_0} < \xi_{k_0}, \eta_{k_r} = \xi_{k_r} \text{ for } 1 \le r \le N-2, \eta_{k_{N-1}} > \xi_{k_{N-1}}\}. \tag{5.34}$$

Thus (5.34) also has positive probability under $\overline{\pi}$. This is impossible, since N was chosen to be minimal in (5.32). So (5.30) must hold as required. □

The next lemma concerns the situation where a maximal non-degenerate product distribution exists. It allows us to deduce Corollary 5.4.7, which shows that conditions such as those in Proposition 5.3.2, which were shown there to guarantee uniqueness of an non-degenerate invariant product distribution, in fact guarantee also that this distribution is the unique non-degenerate invariant distribution for the network.

Lemma 5.4.6 *Suppose that $\boldsymbol{\rho}$ is a maximal solution to the throughput equation (5.13) among those for which $\rho_j < \mu_j$ for all j, so that $\pi_{\boldsymbol{\rho}}$ is a maximal non-degenerate invariant product distribution. If $\pi \in \mathcal{I}$ and $\pi_{\boldsymbol{\rho}} \preceq \pi$, then $\pi = \pi_{\boldsymbol{\rho}}$.*

Proof: Let $w_j = \pi(\eta_j > 0)$. From Lemma 5.4.1, $\mathbf{w} \cdot \boldsymbol{\mu}$ satisfies the throughput equation (5.13); but $\boldsymbol{\rho}$ is a maximal solution to (5.13), so that

$$\pi(\eta_j > 0) = w_j \le \frac{\rho_j}{\mu_j} = \pi_{\boldsymbol{\rho}}(\eta_j > 0) \tag{5.35}$$

for all $j \in S$.

Let $\overline{\pi} \in \overline{\mathcal{I}}$ with marginals $\pi_{\boldsymbol{\rho}}$ and π, and such that $\overline{\pi}(\boldsymbol{\eta} \le \boldsymbol{\xi}) = 1$ (as given by Lemma 5.4.4(iii)).

Then, because of (5.35), we must have

$$\overline{\pi}(0 = \eta_j < \xi_j) = 0 \text{ for all } j \in S.$$

It then follows by induction that

$$\overline{\pi}(x = \eta_j < \xi_j) = 0 \text{ for all } x \in \mathbb{Z}_+, \text{ all } j \in S,$$

since if $x > 0$ and $\{x = \eta_j < \xi_j\}$ occurs with positive probability under the invariant distribution $\overline{\pi}$, then so does $\{x - 1 = \eta_j < \xi_j\}$, because of the possibility of a departure from node j occurring before any arrival to node j.

Hence $\overline{\pi}(\boldsymbol{\eta} = \boldsymbol{\xi}) = 1$, and $\pi = \pi_{\boldsymbol{\rho}}$ as required. □

Corollary 5.4.7 *If there exists a unique non-degenerate invariant product distribution $\pi_{\boldsymbol{\rho}}$ for the network, then $\pi_{\boldsymbol{\rho}}$ is the only member of $\mathcal{I}$.*

Proof: $\pi_{\boldsymbol{\rho}}$ is minimal in $\mathcal{I}$ by Proposition 5.3.1. Thus $\pi_{\boldsymbol{\rho}} \preceq \pi$, and the result follows from Lemma 5.4.6. □

5.4.3 Null Recurrent Case

We now consider in particular the case where $(\mathbf{P}, \boldsymbol{\mu})$ is null recurrent. The coupling methods established in the last section will allow us to show that the invariant product distributions are the only extreme members of $\mathcal{I}$. This was done in similar fashion by Andjel for the zero-range process in [2].

Lemma 5.4.8 *If $(\mathbf{P}, \boldsymbol{\mu})$ is recurrent and $\overline{\pi} \in \overline{\mathcal{I}}$, then $\overline{\pi}(\boldsymbol{\eta} \leq \boldsymbol{\xi}$ or $\boldsymbol{\eta} \geq \boldsymbol{\xi}) = 1$.*

Proof: For given j, let $f_N(\boldsymbol{\eta}, \boldsymbol{\xi}) = \left[\min(\xi_j, N) - \eta_j\right]_+$. Since $f_N \in \mathcal{C}^{(2)}$ and $\overline{\pi} \in \overline{\mathcal{I}}$, we have

$$\int \overline{\Omega} f_N(\boldsymbol{\eta}, \boldsymbol{\xi}) d\overline{\pi}(\boldsymbol{\eta}, \boldsymbol{\xi}) = 0.$$

Expanding this, we have

$$\begin{aligned}\sum_{k \neq j} \overline{\pi}(\eta_j \leq \xi_j \leq N-1, 0 = \eta_k < \xi_k)\mu_k p_{kj} + \overline{\pi}(\eta_j \leq N < \xi_j)\mu_j(1-p_{jj}) \\ -\overline{\pi}(0 = \eta_j < \xi_j)\mu_j(1-p_{jj}) - \sum_{k \neq j} \overline{\pi}(\eta_j \leq N-1, \eta_j < \xi_j, 0 = \xi_k < \eta_k)\mu_k p_{kj} \\ - \sum_{k \neq j} \overline{\pi}(\eta_j \leq N-1 < \xi_j, 0 < \xi_k, 0 < \eta_k)\mu_k p_{kj} = 0.\end{aligned}$$

As $N \to \infty$, $\overline{\pi}(\eta_j \geq N$ or $\xi_j \geq N) \to 0$, since $\overline{\pi}$ is non-degenerate. Hence, letting $N \to \infty$ above, we obtain

$$\begin{aligned}\sum_{k \neq j} \overline{\pi}(\eta_j \leq \xi_j, 0 = \eta_k < \xi_k)\mu_k p_{kj} \\ = \sum_{k \neq j} \overline{\pi}(\eta_j < \xi_j, 0 = \xi_k < \eta_k)\mu_k p_{kj} + \overline{\pi}(0 = \eta_j < \xi_j)\mu_j(1-p_{jj}).\end{aligned}$$

Letting $g_j = \overline{\pi}(0 = \eta_j < \xi_j)$ and adding $g_j \mu_j p_{jj}$ to each side then gives

$$\sum_{k \in S} g_k \mu_k p_{kj} \geq g_j \mu_j + \sum_{k \neq j} \overline{\pi}(\eta_j < \xi_j, 0 = \xi_k < \eta_k)\mu_k p_k j. \tag{5.36}$$

Let $w_j = \overline{\pi}(\xi_j > 0)$. From (5.36) and Lemma 5.4.1,

$$\sum_{k \in S} (w_k - g_k)\mu_k p_{kj} \leq (w_j - g_j)\mu_j. \tag{5.37}$$

But $w_j - g_j \geq 0$ for all $j \in S$, so, since $\mathbf{P}$ is recurrent, Proposition 6.4 in [32] shows that the inequality in (5.37) is an equality. Then there must also be equality in (5.36), so that

$$\sum_{k \neq j} \overline{\pi}(\eta_j < \xi_j, 0 = \xi_k < \eta_k)\mu_k p_{kj} = 0.$$

In a similar way we can show that

$$\sum_{k \neq j} \overline{\pi}(\eta_j > \xi_j, 0 = \eta_k < \xi_k)\mu_k p_{kj} = 0.$$

Thus the hypotheses of Lemma 5.4.5 hold, giving the conclusion (5.29) as desired. □

Theorem 5.4.9 *If $(\mathbf{P}, \boldsymbol{\mu})$ is null recurrent then $\mathcal{I}_e = \{\pi_\rho : \rho \in R\}$.*

Proof: Let $\rho \in R$, and write $\mathbb{E}_\rho$ and Var_ρ for the expectation and variance under π_ρ.

Since $(\mathbf{P}, \boldsymbol{\mu})$ is null recurrent, we have $\sum_{j \in S} \delta_j^0/\mu_j = \infty$, and we can partition S into a sequence of disjoint finite sets J_n, $n = 1, 2, \ldots$, such that $\sum_{j \in J_n} \delta_j^0/\mu_j \geq 1$ for all n.

Define

$$f_n(\boldsymbol{\eta}) = \frac{\sum_{j \in J_n} \mathbf{I}(\eta_j > 0)}{\sum_{j \in J_n} \delta_j^0/\mu_j}.$$

Then, for each $n \geq 1$,

$$\mathbb{E}_\rho f_n(\boldsymbol{\eta}) = \frac{\sum_{j \in J_n} \rho\delta_j^0/\mu_j}{\sum_{j \in J_n} \delta_j^0/\mu_j} = \rho,$$

and

$$\mathrm{Var}_\rho\, f_n(\boldsymbol{\eta}) = \frac{\sum_{j \in J_n} \left[\rho\delta_j^0/\mu_j - \left(\rho\delta_j^0/\mu_j\right)^2\right]}{\left(\sum_{j \in J_n} \delta_j^0/\mu_j\right)^2} \leq \rho,$$

so, by the Strong Law of Large Numbers and the independence under π_ρ of $f_n(\boldsymbol{\eta})$ for different values of n,

$$\pi_\rho\left(\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} f_n(\boldsymbol{\eta}) = \rho\right) = 1. \tag{5.38}$$

Now let $\pi \in \mathcal{I}_e$, and let $\overline{\pi} \in \overline{\mathcal{I}}$ with marginals π and π_ρ, (by Lemma 5.4.4(i)).

By Lemma 5.4.8, $\overline{\pi}_\rho(\boldsymbol{\eta} \leq \boldsymbol{\xi}$ or $\boldsymbol{\eta} \geq \boldsymbol{\xi}) = 1$. Let $\alpha_\rho = \overline{\pi}(\boldsymbol{\eta} \leq \boldsymbol{\xi})$, and suppose that $0 < \alpha_\rho < 1$. Then, by Lemma 5.4.3, we can write

$$\overline{\pi}_\rho = \alpha_\rho \overline{\pi}_\rho^{(1)} + (1 - \alpha_\rho)\overline{\pi}_\rho^{(2)} \tag{5.39}$$

where $\overline{\pi}_\rho^{(1)}(\boldsymbol{\eta} \leq \boldsymbol{\xi}) = 1$ and $\overline{\pi}_\rho^{(2)}(\boldsymbol{\eta} \leq \boldsymbol{\xi}) = 0$.

Since π is extremal in $\mathcal{I}$, the first marginal of both $\overline{\pi}_\rho^{(1)}$ and $\overline{\pi}_\rho^{(2)}$ must be π.

The second marginal of $\overline{\pi}_\rho^{(1)}$ is a distribution $\pi_\rho^{(1)}$ such that $\pi \preceq \pi_\rho^{(1)}$, and the second marginal of $\overline{\pi}_\rho^{(2)}$ is a distribution $\pi_\rho^{(2)}$ such that $\pi_\rho^{(2)} \preceq \pi$. But $\pi_\rho^{(1)}$ and $\pi_\rho^{(2)}$ are also both absolutely continuous with respect to π_ρ, because of (5.39), so in particular, from (5.38),

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f_n(\boldsymbol{\eta}) = \rho \ \pi_\rho^{(1)}\text{-a.s. and } \pi_\rho^{(2)}\text{-a.s.}$$

Hence, since $\pi_\rho^{(2)} \preceq \pi \preceq \pi_\rho^{(1)}$ and the functions $f_r(\boldsymbol{\eta})$ are increasing in $\boldsymbol{\eta}$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f_n(\boldsymbol{\eta}) = \rho \ \pi\text{-a.s.}$$

also.

Clearly this can occur for at most a single value of ρ. For other $\rho \in R$, the assumption that $0 < \alpha_\rho < 1$ must be false, and either $\pi_\rho \preceq \pi$ or $\pi \preceq \pi_\rho$.

But the probabilitites under π_ρ of increasing cylinder sets of the form (5.10) are continuous in $\rho \in R$, so this implies that either $\pi = \pi_\rho$ for some $\rho \in R$, or $\pi_\rho \preceq \pi$ for all $\rho \in R$.

If $\sup R \in R$, then $\pi_\rho \preceq \pi$ for all $\rho \in R$ implies that $\pi = \pi_{\sup R}$, by Lemma 5.4.6.

On the other hand, if $\sup R \notin R$, then there exists $j \in S$ such that $\mu_j = (\sup R)\delta_j^0$. Then $\pi_\rho(\eta_j \geq k) \to 1$ as $\rho \uparrow \sup R$ for all $k \in \mathbb{Z}_+$, and since π is non-degenerate it is impossible that $\pi_\rho \preceq \pi$ for all $\rho \in R$.

Thus we have shown that every $\pi \in \mathcal{I}_e$ has the form π_ρ for some $\rho \in R$. To complete the proof, we need to show that every such π_ρ is indeed in $\mathcal{I}_e$. Let $\rho_0 \in R$. From (5.38), we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f_n(\boldsymbol{\eta}) = \rho_0 \ \pi_{\rho_0}\text{-a.s.} \tag{5.40}$$

Since π_{ρ_0} can be written as a mixture of members of $\mathcal{I}_e$, there must be at least one member π of $\mathcal{I}_e$ such that the event in (5.40) holds π-a.s. But this π can only be π_{ρ_0}, since the event in (5.40) does not hold π_ρ-a.s. for any $\rho \neq \rho_0$. Thus $\pi_{\rho_0} \in \mathcal{I}_e$ as required. □

5.4.4 Transient Case

In this section we turn to the case where $\mathbf{P}$ is a transient matrix. We are not able to provide a complete characterisation of $\mathcal{I}$ as was done for recurrent $\mathbf{P}$. However, we also cannot provide any example in which $\mathcal{I}_e$ is shown to contain distributions which do not have product form; and it certainly seems to be a plausible conjecture that, as in the case where $(\mathbf{P}, \boldsymbol{\mu})$ is null recurrent, $\mathcal{I}_e$ is exactly the set of invariant product distributions.

Approaches to showing this have been made in certain situations. Suppose that $S = \mathbb{Z}^d$, $d \geq 1$, and that λ, $\mathbf{P}$ and $\boldsymbol{\mu}$ are translation invariant. If $\lambda > \mathbf{0}$, then also $\mathbf{p}^* > \mathbf{0}$, (otherwise the network is clearly overloaded), and it follows from Proposition 5.3.2 and Corollary 5.4.7 that $\mathcal{I}$ has at most a single member. If $\boldsymbol{\lambda} = \mathbf{0}$ and $\mathbf{p}^* = \mathbf{0}$, then the results of Andjel in [2] show that all the translation invariant distributions in $\mathcal{I}_e$ have product form. This is proved by an argument similar to the one which we used in the case where $(\mathbf{P}, \boldsymbol{\mu})$ is null recurrent, taking advantage again of the fact that the set of invariant product distributions is one-dimensional (see Proposition 5.3.3) and can be ordered by $\preceq$. The proof is also simplified by the fact that one can show directly that the invariant product distributions are ergodic by using the fact that they are ergodic with respect to the space translations in $\mathbb{Z}^d$.

In general, however, it is not the case that the solutions to the throughput equation $\boldsymbol{\rho} = \boldsymbol{\lambda} + \boldsymbol{\rho}\mathbf{P}$ form a one-dimensional set, and in particular it is not true that if $\pi, \pi' \in \mathcal{I}_e$ then $\pi \preceq \pi'$ or $\pi' \preceq \pi$. Thus the methods used in the previous section do not apply.

In this section we give a result which identifies part of the basin of attraction of any invariant product measure for the network.

Suppose that $\boldsymbol{\rho} < \boldsymbol{\mu}$ and that $\boldsymbol{\rho}$ satisfies the throughput equation (5.13), so that $\pi_{\boldsymbol{\rho}} \in \mathcal{I}$. Let $\boldsymbol{\rho}'$ be another vector indexed by S, such that $\boldsymbol{\rho}' < \boldsymbol{\mu}$, and define $\mathbf{d}$ by

$$d_j = \frac{\mu_j |\rho'_j - \rho_j|}{(\mu_j - \rho'_j)(\mu_j - \rho_j)}$$

for $j \in S$. Then d_j is the absolute difference between the expectation of two geometric random variables with parameters $1 - \rho_j/\mu_j$ and $1 - \rho'_j/\mu_j$, and so between the expectation of the number of particles at node j under $\pi_{\boldsymbol{\rho}}$ and under $\pi_{\boldsymbol{\rho}'}$.

We may regard the product distribution $\pi_{\boldsymbol{\rho}'}$ as a perturbation of $\pi_{\boldsymbol{\rho}}$. The following theorem gives a condition under which the distribution of the state of the network reverts over time to $\pi_{\boldsymbol{\rho}}$, if started with initial distribution $\pi_{\boldsymbol{\rho}'}$.

Theorem 5.4.10 *Let $\boldsymbol{\rho}, \boldsymbol{\rho}'$ and $\mathbf{d}$ be as above, let $\mathbf{Q} = \sum_{n=0}^{\infty} \mathbf{P}^n$, and denote by $\pi^{(t)}$ the distribution of the state of the network at time t if its time-0 distribution is $\pi_{\boldsymbol{\rho}'}$. If $(\mathbf{dQ})_j < \infty$ for all $j \in S$, then $\pi^{(t)} \to \pi_{\boldsymbol{\rho}}$ weakly as $t \to \infty$.*

Proof: Define $\boldsymbol{\rho}^-$ and $\boldsymbol{\rho}^+$ by letting $\rho_j^- = \min(\rho_j, \rho'_j)$ and $\rho_j^+ = \max(\rho_j, \rho'_j)$ for $j \in S$. We will couple a copy $\boldsymbol{\eta}^-(t)$ of the network process whose initial distribution is $\pi_{\boldsymbol{\rho}^-}$ with a copy $\boldsymbol{\eta}^+(t)$ whose initial distribution is $\pi_{\boldsymbol{\rho}^+}$. Since $\pi_{\boldsymbol{\rho}^-} \preceq \pi_{\boldsymbol{\rho}} \preceq \pi_{\boldsymbol{\rho}^+}$ and $\pi_{\boldsymbol{\rho}^-} \preceq \pi_{\boldsymbol{\rho}'} \preceq \pi_{\boldsymbol{\rho}^+}$, the processes starting with distributions $\pi_{\boldsymbol{\rho}}$ and $\pi_{\boldsymbol{\rho}'}$ both majorise the process $\boldsymbol{\eta}^-(t)$ and are both majorised by $\boldsymbol{\eta}^+(t)$. Thus to show the weak convergence of $\pi^{(t)}$ to $\pi_{\boldsymbol{\rho}}$, it will be sufficient to show that for any $j \in S$, the probability that there exists a time T_j such that $\eta_j^-(t) = \eta_j^+(t)$ for all $t \geq T_j$ is 1.

We will consider a network whose nodes may contain marked and unmarked particles. At time 0, let unmarked particles be distributed among the nodes according to $\pi_{\boldsymbol{\rho}^-}$, and let the distribution of the unmarked and marked particles considered together be $\pi_{\boldsymbol{\rho}^+}$. This is possible since $\pi_{\boldsymbol{\rho}^-} \preceq \pi_{\boldsymbol{\rho}^+}$. Then the expected number of marked particles at node j at time 0 is d_j.

Let the normal network dynamics apply, with the provision that unmarked particles have priority at all times; thus, a marked particle only leaves a node if there are no unmarked particles there. Particles entering the network from outside (with rate λ_j at node j) are unmarked. Letting $\eta_j^-(t)$ be the number of unmarked particles at node j at time t and $\eta_j^+(t)$ be the total number of particles at node j at time t then provides a coupling of the two copies of the network process with the desired initial distributions.

We wish to show that, for each $j \in S$, there exists with probability 1 a time T_j such that from T_j onwards there are no marked particles at node j. Since the expected number of marked particles to start at node i is d_i, the expected number of arrivals of marked particles at node j after time 0 is no greater than $\sum_{i\in S} d_i \left(\sum_{n=0}^{\infty} \mathbf{P}^n\right)_{ij} = (\mathbf{dQ})_j$. By hypothesis this is finite, so the number of such arrivals at j is a.s. finite. Thus it suffices to show that no marked particle ever gets "stuck" at a node, arriving and never leaving.

Since $\pi_{\boldsymbol{\rho}}$ is invariant, the distribution of $\boldsymbol{\eta}^-(t)$ is dominated by $\pi_{\boldsymbol{\rho}}$ because of the majorisation referred to earlier. Hence

$$\begin{aligned}\mathbb{E}(\liminf_{t\to\infty} \eta_j^-(t)) &\le \liminf_{t\to\infty} \mathbb{E}(\eta_j^-(t)) \\ &\le \rho_j/(\mu_j - \rho_j)\end{aligned}$$

so that

$$\mathbb{P}(\eta_j^-(t) \to \infty \text{ as } t \to \infty) = 0.$$

Thus, with probability 1, there exists $x_j \in \mathbb{Z}_+$ such that η_j^- enters state x_j infinitely often, or eventually stays in state x_j for all time. Since potential departure points occur at node j with constant rate μ_j, independently of previous network events, it follows that (if $x_j > 0$), η_j^- enters state $x_j - 1$ infinitely often, and so on by induction to show that η_j^- either enters state 0 infinitely often or is in fact eventually 0 for all time.

It follows that, from any point in time onwards, there are infinitely many potential departure points at node j at times when no unmarked particles are present. Thus, since the number of times that marked particles arrive at j is almost surely finite, all marked particles which arrive at j will eventually leave j, as required. $\square$

Remark: Note that $\sum_{i\in S} d_i < \infty$ is certainly a sufficient condition for $(\mathbf{dQ})_j < \infty$, since $(\mathbf{Q})_{ij} \le (\mathbf{Q})_{jj}$ for all $i, j \in S$. For simplicity we assumed that the perturbed distribution, here $\pi_{\boldsymbol{\rho}'}$, was a product distribution. The method of proof makes clear that the same result holds for a general initial distribution π which can be coupled with $\pi_{\boldsymbol{\rho}}$ in such a way that $(\mathbf{dQ})_j < \infty$ holds, where d_j is the expected absolute difference between the number of particles at j in the component with distribution $\pi_{\boldsymbol{\rho}}$ and the number of particles at j in the component with distribution π.

Bibliography

[1] Aalen, O., (1978) Nonparametric Inference for a Family of Counting Processes. *Ann. Statist.* **6**, 701–726.

[2] Andjel, E. D., (1982) Invariant Measures for the Zero Range Process. *Ann. Prob.* **10**, 525–547.

[3] Andjel, E. D., Coccozza-Thivent, C. and Roussignol, M., (1985) Quelques Compléments sur le Processus des Misanthropes et le Processus "Zéro-Range". *Ann. Inst. Henri Poincaré Prob. Stat.* **21**, 363–382.

[4] Andjel, E. D. and Vares, M. E., (1987) Hydrodynamic equations for attractive particle systems on $\mathbb{Z}$. *J. Stat. Phys.* **47**, 265–288.

[5] Apted, M., dir., (1996) Extreme Measures. *Simian Films.* With Hugh Grant, Gene Hackman and Sarah Jessica Parker.

[6] Benassi, A. and Fouque, J.-P., (1988) Hydrodynamical Limit for the Asymmetric Zero-Range Process. *Ann. Inst. Henri Poincaré Prob. Stat.* **24**, 189–200.

[7] Benois, O., Kipnis, C. and Landim, C., (1995) Large Deviations from the Hydrodynamical Limit of Mean Zero Asymmetric Zero Range Processes. *Stoch. Proc. Appl.* **55**, 65–89.

[8] Brown, T. C. and Pollett, P. K., (1982) Some Distributional Approximations in Markovian Queueing Networks. *Adv. Appl. Prob.* **14**, 654–671.

[9] Cocozza-Thivent, C., (1985) Processus des misanthropes. *Z. Wahrsch. Verw. Gebiete* **70**, 509–523.

[10] Crametz, J.-P. and Hunt, P. J., (1991) A Limit Result Respecting Graph Structure For A Fully Connected Loss Network With Alternative Routing. *Ann. Appl. Prob.* **1**, 436–444.

[11] Daley, D. J. and Vere-Jones, D., (1988) *An Introduction to the Theory of Point Processes.* Springer Verlag.

[12] Echeverría, P., (1982) A Criterion for Invariant Measures of Markov Processes. *Z. Wahrsch. Verw. Gebiete* **61**, 1–16.

[13] Ethier, S. N. and Kurtz, T. G., (1985) *Markov Processes: Characterization and Convergence*. John Wiley & Sons.

[14] Ferrari, P. A., Presutti, E. and Vares, M. E., (1987) Local Equilibrium for a One-Dimensional Zero Range Process. *Stoch. Proc. Appl.* **26**, 31–45.

[15] Gibbens, R. J., Hunt, P. J. and Kelly, F. P., (1990) Bistability in Communication Networks. In G. R. Grimmett and D. J. A. Welsh, eds., *Disorder in Physical Systems*, pages 113–128. OUP.

[16] Goodman, J. B. and Massey, W. A., (1984) The Non-Ergodic Jackson Network. *J. Appl. Prob.* **21**, 860–869.

[17] Graham, C., (1996) A Statistical Physics Approach to Large Networks. In *Probabilistic Models for Nonlinear Partial Differential Equations*, vol. 1627 of *Lecture Notes in Math.*, pages 127–147. Springer-Verlag.

[18] Graham, C., (1999) Chaoticity on Path Space for a Queueing Network with Selection of the Shortest Queue among Several. CMAP, Ecole Polytechnique. Preprint.

[19] Holley, R., (1970) A Class of Interaction in an Infinite Particle System. *Adv. in Math* **5**, 291–309.

[20] Hordijk, A. and Koole, G., (1992) On the Shortest Queue Policy for the Tandem Parallel Queue. *Prob. in the Eng. and Inf. Sci.* **6**, 63–79.

[21] Hunt, P. J., (1995) Loss Networks Under Diverse Routing: The Symmetric Star Network. *Adv. Appl. Prob.* **25**, 255–272.

[22] Hunt, P. J. and Kurtz, T. G., (1994) Large Loss Networks. *Stoch. Proc. Appl.* **53**, 363–378.

[23] Hunt, P. J. and Laws, C. N., (1992) Least Busy Alternative Routing in Queueing and Loss Networks. *Prob. in Eng. and Inf. Sci.* **6**, 439–456.

[24] Jackson, J. R., (1957) Networks of Waiting Times. *Operational Research* **5**, 518–527.

[25] Jackson, J. R., (1965) Jobshop-like Queueing Systems. *Management Sci.* **10**, 131–142.

[26] Kabanov, Y. M. and Liptser, R. S., (1983) Convergence of the Distributions of Multivariate Point Processes. *Z. Wahrsch. verw. Gebiete* **63**, 475–485.

[27] Kamae, T., Krengel, U. and O'Brien, G. L., (1977) Stochastic Inequalities on Partially Ordered Spaces. *Ann. Prob.* **5**, 899–912.

[28] Karpelevich, F. I., Pechersky, E. A. and Suhov, Y. M., (1996) Dobrushin's Approach to Queueing Network Theory. *J. Appl. Math. Stoch. Anal.* **9**, 373–397.

[29] Kel'bert, M. Y., (1989) On the Central Limit Theorem for Jackson Networks on Countable Graphs. *Theory Prob. Appl.* **34**, 713–714.

[30] Kel'bert, M. Y., Kontsevich, M. L. and Rybko, A. N., (1989) On Jackson Networks on Denumerable Graphs. *Theory of Probability and its Applications* **33**, 358–361.

[31] Kelly, F. P., (1979) *Reversibility and Stochastic Networks.* John Wiley & Sons.

[32] Kemeny, J. G., Snell, J. L. and Knapp, A. W., (1976) *Denumerable Markov Chains.* Springer-Verlag, New York, 2nd ed.

[33] Kingman, J. F. C., (1961) Two Similar Queues in Parallel. *Ann. Math. Statist.* **32**, 1314–1323.

[34] Kurtz, T. G., (1992) Averaging for Martingale Problems and Stochastic Approximation. In *Applied Stochastic Analysis, Proc. US-French Workshop*, no. 177 in Lecture Notes in Control and Information Sciences, pages 186–209. Springer-Verlag.

[35] Landim, C., (1996) Hydrodynamical Limit for Space Inhomogeneous One-Dimensional Totally Asymmetric Zero-Range Processes. *Ann. Prob.* **24**, 599–638.

[36] Landim, C. and Mourragui, M., (1997) Hydrodynamic Limit of Mean Zero Asymmetric Zero Range Processes in Infinite Volume. *Ann. Inst. Henri Poincaré Prob. Stat.* **33**, 65–82.

[37] Liggett, T. M., (1973) An Infinite Particle System with Zero Range Interactions. *Ann. Prob.* **1**, 240–253.

[38] Liggett, T. M., (1985) *Interacting Particle Systems.* Springer-Verlag, New York.

[39] Liggett, T. M., (1997) Stochastic Models of Interacting Systems. *Ann. Prob.* **25**, 1–29.

[40] Massey, W. A., (1984) An Operator-Analytic Approach to the Jackson Network. *J. Appl. Prob.* **21**, 379–393.

[41] Mitzenmacher, M., (1996) The Power of Two Choices in Randomized Load Balancing. Ph.D. thesis, University of California, Berkeley.

[42] Mitzenmacher, M., (1997) How Useful is Old Information? In *Proceedings of the 16th Annual ACM Symposium on Principles of Distributed Computing.*

[43] Norris, J. R., (1997) *Markov Chains.* Cambridge University Press.

[44] Phelps, R. R., (1965) *Lectures on Choquet's Theorem.* Van Nostrand, Princeton.

[45] Spitzer, F., (1970) Interaction of Markov Processes. *Adv. in Math.* **5**, 246–290.

[46] Srinivasan, R., (1993) Queues in Series via Interacting Particle Systems. *Math. Oper. Res.* **18**, 39–50.

[47] Teh, Y. C., (1998) Alternative Routing in Fully Connected Queueing Networks. Department of Statistics, University of Oxford. Preprint.

[48] Thomas, E., (1998) A Loss Service System With Another Chance. Statistical Laboratory, University of Cambridge.

[49] Tsoucas, P. and Walrand, J., (1984) A Note on Stochastic Bounds for Queueing Networks. *Adv. Appl. Prob.* **16**, 926–928.

[50] Turner, S. R. E., (1998) The Effect of Increasing Routing Choice on Resource Pooling. *Probability in the Engineering and Informational Sciences* **12**, 109–124.

[51] Vvedenskaya, N. D., Dobrushin, R. L. and Karpelevich, F. I., (1996) Queueing System with Selection of the Shortest of Two Queues: An Asymptotic Approach. *Problems of Information Transmission* **32**, 15–27.

[52] Vvedenskaya, N. D. and Suhov, Y. M., (1997) Dobrushin's Mean-Field Approximation for a Queue with Dynamic Routing. *Markov Processes and Related Fields* **3**, 493–526.

[53] Walrand, J., (1986) *An Introduction to Queueing Networks.* Prentice Hall.

[54] Walters, P., (1982) *An Introduction to Ergodic Theory.* Springer-Verlag.

[55] Waymire, E., (1980) Zero Range Interaction at Bose-Einstein Speeds Under a Positive Recurrent Single Particle Law. *Ann. Prob.* **8**, 441–450.

[56] Whitt, W., (1985) Blocking When Service is Required from Several Facilities Simultaneously. *AT & T Tech. J.* **64**, 1807–1856.

[57] Whitt, W., (1986) Deciding Which Queue to Join: Some Counterexamples. *Op. Res.* **34**, 55–62.

[58] Whittle, P., (1986) *Systems in Stochastic Equilibrium.* John Wiley & Sons.

[59] Williams, D., (1991) *Probability with Martingales.* Cambridge University Press.