POINT PROCESSES IN FAST JACKSON NETWORKS

J. B. MARTIN[1]

INRIA and Ecole Normale Supérieure, Paris

We consider a Jackson-type network, each of whose nodes contains N identical channels with a single server. Upon arriving at a node, a task selects m of the channels at random, and joins the shortest of the m queues observed. We fix a collection of channels in the network, and analyse how the queue-length processes at these channels vary as $N \to \infty$. If the initial conditions converge suitably, the distribution of these processes converges in local variation distance to a limit under which each channel evolves independently. We discuss the limiting processes which arise, and in particular we investigate the point processes of arrivals and departures at a channel when the networks are in equilibrium, for various values of the system parameters.

1 Introduction

The class of *Jackson networks* was introduced in [4] and [5] and has since been one of the most widely studied in queueing network theory. The basic model consists of a network of J nodes; each node j, $1 \le j \le J$, has an infinite buffer and a single server with service rate μ_j; tasks arrive at node j from outside the network as a Poisson process of rate λ_j, and when a task completes its service at node j, it immediately joins the queue at node k with probability p_{jk}, $1 \le k \le J$, and leaves the network with probability $p_j^* = 1 - \sum_k p_{jk}$. All service times, external arrival times and routing decisions are independent.

One of the attractions of this model is the appealing *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. As specified, 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 [7]) 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.

[1]Work supported in part by EPSRC studentship no. 95007341.

AMS 1991 subject classifications: 90B15; 60G55.

Keywords and phrases: Queueing network, Point process, Jackson network.

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.

We consider a modification of the Jackson network model, in which each node j of the network contains N channels, each with an infinite buffer and a single server with service rate μ_j. External arrivals arrive at node j at rate $N\lambda_j$, and routing between nodes is as before, according to the matrix $\mathbf{P} = (p_{jk})$. Upon arrival at a node, (either from outside the network or after being served at the same or another node), each task now inspects m channels chosen uniformly at random from the N available (with replacement, though this is unimportant for large N), and joins the shortest of the m queues observed (breaking ties at random). The behaviour of the network is therefore specified by the parameters J, N, $\underline{\lambda}$, $\underline{\mu}$, $\mathbf{P}$ and m.

Systems corresponding to a single node in this model were studied by Vvedenskaya *et al.* in [9]. The network model just described was introduced in [8]. There it is shown that, as $N \to \infty$, the evolution of the system may be asymptotically represented by the solution of a countably infinite system of ODEs; under a standard non-overload condition on the parameters $\underline{\lambda}$, $\underline{\mu}$ and $\mathbf{P}$, the system has an invariant distribution π_N for each N, and, as $N \to \infty$, π_N converges to a limiting invariant distribution which is concentrated at a single point, corresponding to the fixed point of the system of ODEs. Under this limiting distribution, for $m > 1$, the tail of the distribution of queue lengths decays super-exponentially, rather than exponentially as in the case of standard Jackson networks — hence the term "Fast Jackson Networks".

In this paper we again let $N \to \infty$, but now consider 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 introduce notation and restate results from [8] which we will use. In Section 3 the main theorem is proved, using results of Kabanov and Liptser from [6] which relate convergence in variation distance of multivariate point processes to the convergence of their compensators. In Section 4 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 5 we discuss possible extensions of the results; in particular we compare our approach with that of Brown and Pollett [2], 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.

2 Preliminaries

The state of a network as described above with N channels at each node may be described by a vector $\underline{\mathbf{r}} = \{r_j(n),\ 1 \le j \le J,\ n \in \mathbb{Z}_+\}$, (here and below $\mathbb{Z}_+$ is the set of non-negative integers), where $r_j(n) = N^{-1} \sum_{n' \ge n} M_j(n')$ and $M_j(n')$ is the number of channels at node j with queue length n'. Hence $r_j(n)$ is the proportion of channels at node j whose queue length is at least n. The process $\underline{\mathbf{r}}(t) = \{r_j(n,t),\ n \in \mathbb{Z}_+,\ t \ge 0\}$, describing the state of the network at times $t \ge 0$, is easily seen to be a Markov process for each N, with state space $\mathcal{U}_N^J$ where

$$\mathcal{U}_N = \big\{\mathbf{g} = \big(g(n), n \in \mathbb{Z}_+\big) : g(0) = 1,\ g(n) \ge g(n+1) \ge 0,\ Ng(n) \in \mathbb{N},\ \forall n, \\ \text{and } g(n) = 0 \text{ for sufficiently large } n\big\}. \quad (2.1)$$

Since we wish to let $N \to \infty$, we will also consider the limiting space $\mathcal{U}^J$, where

$$\mathcal{U} = \big\{\mathbf{g} = \big(g(n), n \in \mathbb{Z}_+\big) : g(0) = 1,\ g(n) \ge g(n+1) \ge 0,\ \forall n,\ \text{and } \sum_{n=0}^{\infty} g(n) < \infty\big\}.$$

Then $\mathcal{U}_N^J \subset \mathcal{U}^J$ for all N. Following [9], we define the metric

$$d(\underline{\mathbf{u}}, \underline{\mathbf{u}}') = \sup_{1 \le j \le J} \sup_{n \ge 1} \frac{|u_j(n) - u'_j(n)|}{n} \quad (2.2)$$

on the spaces $\mathcal{U}^J$ and $\mathcal{U}_N^J$.

We will consider the following infinite system of non-linear differential equations for $\underline{\mathbf{u}}(t) = \{u_j(n,t), 1 \le j \le J, n \in \mathbb{Z}_+\}$, $t \ge 0$, with initial condition $\underline{\mathbf{g}} \in \mathcal{U}^J$:

$$\underline{\mathbf{u}}(0) = \underline{\mathbf{g}}, \quad (2.3)$$

$$\dot{\underline{\mathbf{u}}}(t) = \underline{\mathbf{h}}(\underline{\mathbf{u}}(t)), \quad (2.4)$$

where, for all j.

$$h_j(0, \underline{\mathbf{u}}) = 0, \quad (2.5)$$

$$h_j(n, \underline{\mathbf{u}}) = \left[\lambda_j + \sum_{1 \le k \le J} \mu_k p_{kj} u_k(1)\right] [u_j(n-1)^m - u_j(n)^m] - \mu_j[u_j(n) - u_j(n+1)] \quad (2.6)$$

for all $n \ge 1$.

The following result, proved in [8], then describes how the solution of this system asymptotically represents the behaviour of the network:

Theorem 2.1
(i) If $\underline{\mathbf{g}} \in \mathcal{U}^J$, the system (2.3)–(2.6) has a unique solution $\underline{\mathbf{u}}(t, \underline{\mathbf{g}})$, $t \ge 0$ in $\mathcal{U}^J$.
(ii) For any continuous function $f : \mathcal{U}^J \to \mathbb{R}$ and $t \ge 0$,

$$\lim_{N \to \infty} \sup_{\underline{\mathbf{g}} \in \mathcal{U}_N^J} \big|\mathbb{E}_N \left[f\big(\underline{\mathbf{r}}(s)\big) | \underline{\mathbf{r}}(0) = \underline{\mathbf{g}}\right] - f\big(\underline{\mathbf{u}}(s, \underline{\mathbf{g}})\big)\big| = 0,$$

uniformly in $s \in [0, t]$, where $\mathbb{E}_N$ denotes the expectation under the dynamics of the network with N channels at each node.

3 Convergence as $N \to \infty$

We fix the parameters J, $\underline{\lambda}$, $\underline{\mu}$, $\mathbf{P}$ and m, and consider a sequence of networks indexed by N, with the Nth network having N channels at each node. We fix a set of K tagged channels among the various nodes; formally, we fix a function i from $\{1, 2, \ldots, K\}$ to $\{1, 2, \ldots, J\}$ such that, in each network, the kth tagged channel belongs to node $i(k)$, $1 \le k \le K$. We will analyse the behaviour of the process of queue lengths at the tagged channels as $N \to \infty$.

We will describe the evolution of the network by constructing, for each N, a process $(\underline{\mathbf{r}}(t), \mathbf{x}(t))$, $t \ge 0$, with state space $\mathcal{U}^J \times \mathbb{Z}_+^K$. Here $x_k(t)$ will be the length of the queue in the kth tagged channel at time t, and $r_j(n,t)$ will be 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. The topology used on $\mathcal{U}^J \times \mathbb{Z}_+^K$ is the product of the topology induced by the metric (2.2) on $\mathcal{U}^J$ and the discrete topology on $\mathbb{Z}_+^K$.

Let Ω be the space of paths $[0, \infty) \to \mathcal{U}^J \times \mathbb{Z}_+^K$ which are right continuous with left limits, (representing the paths of $(\underline{\mathbf{r}}(t), \mathbf{x}(t))$). For each $t \ge 0$ we define the functions $\underline{\mathbf{r}}(t)$ and $\mathbf{x}(t)$ on Ω by setting $(\underline{\mathbf{r}}(t), \mathbf{x}(t))(w) = w(t)$. Define the σ-algebra $\mathcal{G}$ on Ω by $\mathcal{G} = \sigma(\underline{\mathbf{r}}(s), \mathbf{x}(s), s \ge 0)$.

For each $N \ge K$, let ψ_N be a distribution on $\mathcal{U}^J \times \mathbb{Z}_+^K$, 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 the measurable space $(\Omega, \mathcal{G})$ describing the behaviour of the Nth network. We write $\mathbb{E}_N$ for the expectation with respect to $\bar{\mathbb{P}}_N$.

We define the filtration $G = \{\mathcal{G}_t\}_{t \ge 0}$ on $\mathcal{G}$ by $\mathcal{G}_t = \sigma(\underline{\mathbf{r}}(s), \mathbf{x}(s), 0 \le s \le t)$. For each N, $(\underline{\mathbf{r}}(t), \mathbf{x}(t), t \ge 0)$ is a stochastic process defined on $(\Omega, \mathcal{G}, \bar{\mathbb{P}}_N)$ and adapted to the filtration G. Note that $\bar{\mathbb{P}}_N(\underline{\mathbf{r}}(t) \in \mathcal{U}_N^J \; \forall t) = 1$. Ultimately we will be particularly interested in the smaller filtration $F = \{\mathcal{F}_t\}_{t \ge 0}$, where $\mathcal{F}_t = \sigma(\mathbf{x}(s),\ 0 \le s \le t)$, and $\mathcal{F} = \bigvee_t \mathcal{F}_t$. Clearly $\mathcal{F}_t \subset \mathcal{G}_t \; \forall t$ 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 3.1 *Suppose that $\psi_N \to \psi$ weakly, where ψ is a distribution on $\mathcal{U}^J \times \mathbb{Z}_+^K$ under which the marginal distribution of $\underline{\mathbf{r}}(0)$ on $\mathcal{U}^J$ 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 $\boldsymbol{\xi} = \{\xi_{k,l}(t), t \ge 0, k, l \in \{0, 1, \ldots, K\}, (l,k) \ne (0,0)\}$ 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 during 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 $k=l$). So, for example, the total number of departures from the kth tagged channel during the time interval $(0,t]$ is $\sum_{l=0}^{K} \xi_{k,l}(t)$. Note that, $\bar{\mathbb{P}}_N$-a.s. 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 [3].

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 $\mathcal{G}_t = \sigma\left(\{\underline{\mathbf{r}}(s), \mathbf{x}(s)\}, 0 \le s \le t\right)$, 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

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

using a version of the Dominated Convergence Theorem for conditional expectations; see for example [10], 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 $\boldsymbol{\beta}^{(N)}$.

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 [6] then shows that a sufficient condition for $\mathrm{Var}_t(\mathbb{P}_N, \mathbb{P}) \to 0$ as $N \to \infty$ is that

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

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 [10], 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

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

for all k, l.

Consider a channel at node i whose queue length is x, while the "environment" is $\underline{\mathbf{u}} \in \mathcal{U}^J$. 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, if the overall state of the network is described by $\underline{\mathbf{u}}$, 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 $\underline{\mathbf{u}}, \underline{\mathbf{u}}' \in \mathcal{U}^J$,

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

and

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

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 $I\{x_k(t) > 0\}\mu_{i(k)}p_{i(k)i(l)}$, so that we have

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

(Here and below we write $\underline{\mathbf{e}}_i(n)$ for the vector in $\mathcal{U}^J$ whose only non-zero entry is the (i, n) entry, which is 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

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

so that putting

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

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

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

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

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

where $\underline{\mathbf{u}}(t,\underline{\mathbf{g}}) = \{u_j(n,t,\underline{\mathbf{g}}), 1 \le j \le J, n \in \mathbb{Z}_+\}$ is defined by Theorem 2.1(i).

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

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

Thus

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

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(ii) and the definition (2.2) of the metric d,

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

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. □

4 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 3.1.

We will require that no node is *overloaded*. Assume that the matrix $\mathbf{I}-\mathbf{P}$ is invertible. (A physical interpretation of this condition is that, almost surely, every task entering the network eventually leaves the network). Define the vector $\underline{\rho}=(\rho_1,\ldots,\rho_J)$ by

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

Then we have, for all i,

$$\rho_i=\lambda_i+\sum_{1\leq j\leq J}\rho_j p_{ji}, \tag{4.1}$$

and if

$$\underline{\rho}<\underline{\mu}\ (\text{i.e. } \rho_j<\mu_j \text{ for all } j), \tag{4.2}$$

then the networks are said to be non-overloaded, and $N\rho_i$ may be interpreted as the "effective arrival rate" in equilibrium at node j in the Nth network, including arrivals from inside as well as outside the network.

In [8] the following result is established concerning the equilibrium behaviour of the networks considered:

Theorem 4.1 *If (4.2) holds, then:*

(i) There exists a unique fixed point $\underline{\mathbf{a}}$ *in* $\mathcal{U}^J$ *of the system (2.3)-(2.6), i.e. such that* $\underline{\mathbf{u}}(t,\underline{\mathbf{a}})=\underline{\mathbf{a}}$ *for all* t*, and* $\underline{\mathbf{a}}$ *is given by*

$$a_j(n)=\left(\frac{\rho_j}{\mu_j}\right)^{\frac{(m^n-1)}{(m-1)}}. \tag{4.3}$$

(ii) The Markov process $\underline{\mathbf{r}}_N(t)$ *is positive recurrent for all* N*, and so has a unique invariant probability distribution* π_N *for each* N*.*

(iii) $\pi_N\to\delta_{\underline{\mathbf{a}}}$ *weakly as* $N\to\infty$*, where* $\delta_{\underline{\mathbf{a}}}$ *is the probability measure concentrated at the fixed point* $\underline{\mathbf{a}}$*.*

So from here on we assume that (4.2) holds and discuss the situation in which, for each N, the distribution under ψ_N of the initial state $\underline{\mathbf{r}}(0)$ on $\mathcal{U}_N^J$ is the equilibrium distribution π_N given by Theorem 4.1(ii). Then the sequence ψ_N has a weak limit ψ under which $\underline{\mathbf{r}}(0)$ is equal to $\underline{\mathbf{a}}$ with probability 1, and we can apply Theorem 3.1. Since $\underline{\mathbf{u}}(\underline{\mathbf{a}},t)=\underline{\mathbf{a}}$ for all t, (3.12) becomes

$$\beta_{0,l}(t)=\rho_{i(l)}c_{i(l)}\big(\underline{\mathbf{a}},x_l(t)\big), \tag{4.4}$$

using (4.1). 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 $\beta_{k,0}(t) = I\{x_k(t) > 0\}\mu_{i(k)}$ 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 4.1(i), 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 discuss 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 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 — 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

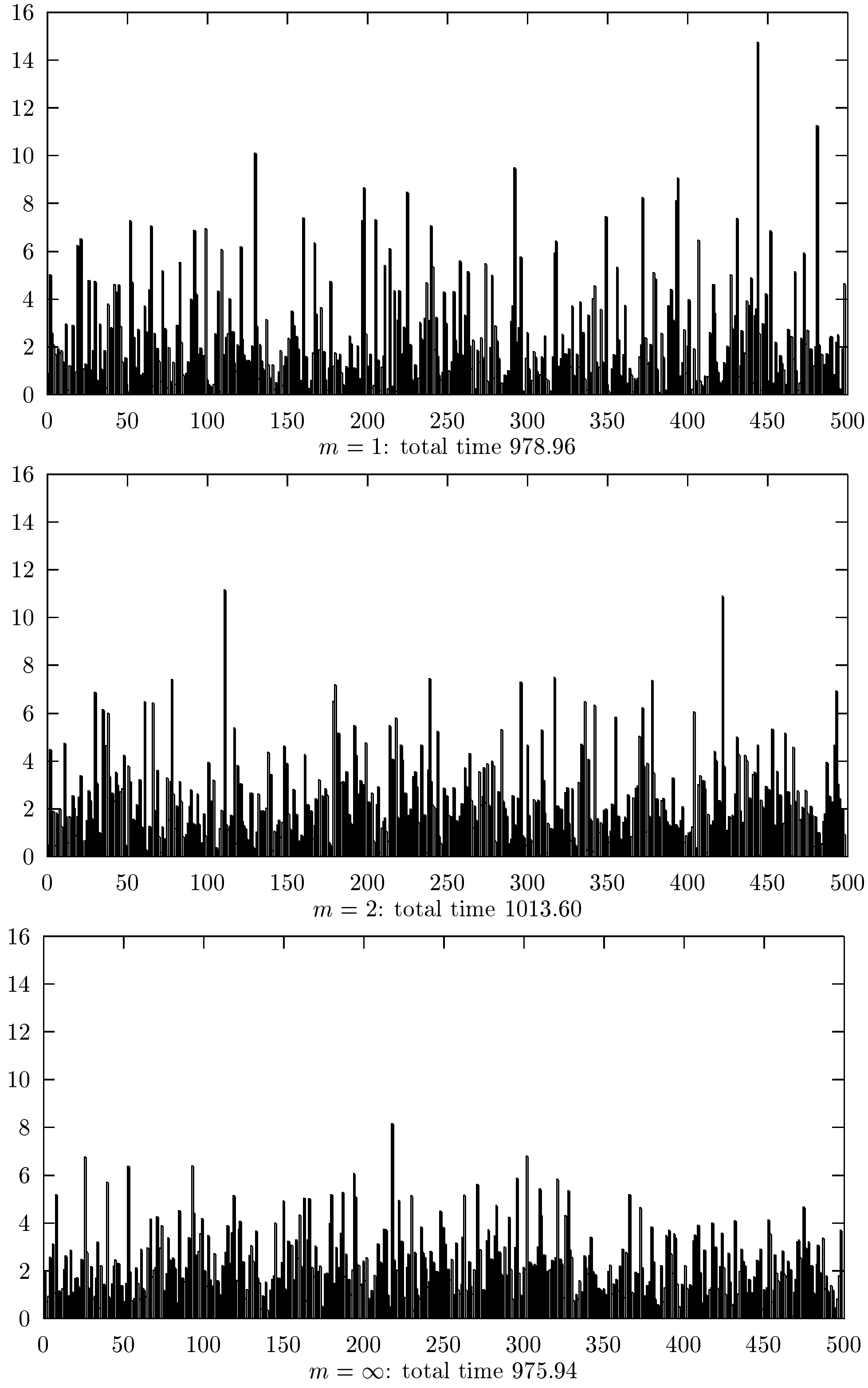

Figure 1: Simulations of 500 successive interarrival times for $\rho = 0.5$, $\mu = 1$, various m.

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 (4.3), 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 $m > 1$.

5 Extensions

An analagous result to Theorem 3.1 holds if we remove the condition that the marginal distribution of $\underline{\mathbf{r}}(0)$ on $\mathcal{U}^J$ 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 3.1 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 [2], we can consider letting J, the number of nodes, rather than N, the number of channels at each node, tend to infinity. In [2], 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 $J \to \infty$. For example, an appropriate condition for standard Jackson networks is that

$$\sum_{j=1}^{J} (\mu_j p_{ji})^2 \to 0 \text{ as } J \to \infty.$$

It seems likely that similar results hold also in the situations we have considered above. However, the networks considered in [2] 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.

Acknowledgments

I would like to thank Dr. Yuri Suhov for many valuable conversations during the course of this work, and a referee for several helpful suggestions and, in particular, comments leading to a more rigorous presentation of Theorem 3.1.

References

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

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

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

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

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

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

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

[8] Martin, J. B. and Suhov, Y. M. Fast Jackson Networks. To appear in *Ann. Appl. Prob.*

[9] 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.

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

Département de Mathématiques et d'Informatique,
Ecole Normale Supérieure,
45 rue d'Ulm,
75230 PARIS Cédex 05,
FRANCE
James.Martin@ens.fr