Asymptotics of overflow probabilities in Jackson networks *

Jiyeon Lee

Department of Statistics, Yeungnam University, Kyungsan 712-749, Korea

Abstract

We consider the probability that the total population of a Jackson network exceeds a given large value. By using the relation to the stationary distribution, we derive upper and lower bounds on this probability. These bounds imply a stronger logarithmic limit when multiple nodes have the same maximal load.

Keywords: Jackson network, Overflow probability, Asymptotics

1 Introduction

We analyze a rare-event probability in queueing networks. The probability in question is

\[ p_K := P\{\text{network population reaches } K \text{ before returning to 0, starting from 0}\}, \]

a type of overflow probability if we think of \( K \) as an upper limit on the network population. The network we consider is a Jackson network; a network of \( n \) exponential servers with Bernoulli routing and Poisson exogenous arrivals. It is generally accepted that this overflow probability is analytically intractable. From this point of view, the problem of estimating the overflow

*This work was supported by Korea Research Foundation Grant (KRF-99-015-DP0045). The author wishes to thank anonymous referees and the associate editor for their valuable comments and suggestions that improved the paper significantly.
probability by simulation has been studied by Parekh and Walrand [12],
McDonald [11] and Lee [10]. Simulation approaches to this problem include
an asymptotic limit based on the large deviation theory. Glasserman and
Kou [6] proved the following asymptotic limit:
\[
\lim_{K \to \infty} \frac{1}{K} \log p_K = \log \rho^*,
\]
where \( \rho^* \) is the load of the most highly loaded node in the network. In this
paper we obtain the stronger logarithmic limit for \( p_K \) when the multiple
nodes have the same maximal load \( \rho^* \). To do so, we derive upper and lower
bounds on \( p_K \) in two steps: we first bound the stationary probability of the
overflow set - the set of states with population \( K \); we then use the time
reversal and the fluid limit to convert the stationary bounds to the bounds
on the transient probability \( p_K \).

2 Asymptotics of overflow probabilities

A Jackson network consists of \( n \) nodes or service stations that operate on
a first-come-first-served basis. Customers arrive at a typical node \( i \) from
outside the system according to a Poisson process with rate \( \bar{\lambda}_i \) and, if nec-
essary, wait in a buffer until the station gets free to get served. Service time
is exponentially distributed with mean \( 1/\mu_i \). Once service is completed, the
customer is routed to another node, say \( j \), with probability \( r_{ij} \) or leaves the
system with probability \( r_i := 1 - \sum_{j=1}^n r_{ij} \).

We say that node \( i \) feeds node \( j \) if there is a sequence \( k_1, k_2, \ldots, k_q \) such
that \( r_{i_1} r_{i_2} \cdots r_{i_q} j > 0 \). A network is exogenously supplied if each node
\( i \) has an exogenous arrival rate \( \tilde{\lambda}_i \neq 0 \) or can be fed by another node \( j \) for
which \( \tilde{\lambda}_j \neq 0 \). The network is open if every node \( i \) has an exit probability
\( r_i \neq 0 \) or feeds a node \( j \) for which \( r_j \neq 0 \). We assume that the network is
both exogenously supplied and open.

A Jackson network can be described as a Markov jump process \( \{X(t); t \geq 0\} \) on
\( S \equiv \mathbb{N}^n \), where the state \( \vec{x} = (x_1, x_2, \ldots, x_n) \in S \) depicts the system
when there are \( x_i \) customers waiting or being served at node \( i \).
Jackson [9] gave an expression for the invariant measure of a Jackson network. His results are restated in the following theorem (see Brémaud [4]).

**Theorem 1** For an exogenously supplied and open Jackson network for which the solution \((\lambda_1, \cdots, \lambda_n)\) to the traffic equations

\[
\lambda_i = \bar{\lambda}_i + \sum_{j=1}^{n} \lambda_j r_{ji}, \quad i = 1, 2, \cdots, n
\]
satisfies the light traffic conditions

\[
\rho_i := \frac{\lambda_i}{\mu_i} < 1, \quad i = 1, 2, \cdots, n, \quad (2)
\]

the stationary distribution \(\pi(\vec{x})\) of \(\vec{x} = (x_1, \cdots, x_n) \in S\) is given by the product

\[
\pi(\vec{x}) = \prod_{i=1}^{n} (1 - \rho_i) \rho_i^{x_i}.
\]

The ratio \(\rho_i\) is called the load on node \(i\). We call a Jackson network stable if the light traffic conditions (2) hold. We assume that the light traffic conditions hold. We further assume that \(s\) nodes, say \(\{1, 2, \cdots, s\}\), have the same maximal load \(\rho_\ast\), that is, \(\rho_\ast = \rho_1 = \rho_2 = \cdots = \rho_s > \rho_{s+1} \geq \cdots \geq \rho_n\).

Let us define an overflow set by

\[
C_K = \{\vec{x} \in S : x_1 + x_2 + \cdots + x_n = K\},
\]

the set of states in which the network population is exactly \(K\). We bound \(p_K\) by first bounding the stationary probability of the set \(C_K\), \(\pi(C_K)\).

**Lemma 1**

\[
b_1 \rho_\ast^K (K + 1)^{s-1} \leq \pi(C_K) \leq b_2 \rho_\ast^K (K + 1)^{s-1}
\]

for positive constants \(b_1\) and \(b_2\), which are independent of \(K\).
Proof. From Theorem 1, the stationary probability \( \pi(C_K) \) is given by

\[
\pi(C_K) = \sum_{x_1 + \cdots + x_n = K} \prod_{i=1}^{n} (1 - \rho_i) \rho_i^{x_i}
\]

\[
= \prod_{i=1}^{n} (1 - \rho_i) \rho_i^K \sum_{x_1 + \cdots + x_n = K} \prod_{i=1}^{n} (\rho_i/\rho_s)^{x_i}
\]

\[
= \prod_{i=1}^{n} (1 - \rho_i) \rho_i^K \sum_{k=0}^{K} \sum_{x_{s+1} + \cdots + x_n = k} \sum_{x_1 + \cdots + x_s = K-k} \prod_{i=1}^{n} (\rho_i/\rho_s)^{x_i}
\]

where the last equality follows from \( \rho_i/\rho_s = 1 \) for \( i = 1, 2, \cdots, s \) and \( \sum_{x_1 + \cdots + x_s = K-k} = \binom{K-k+s-1}{s-1} \) in Feller [5], p.38.

Since \( \rho_i/\rho_s < 1 \) for \( i = s+1, \cdots, n \), we can get the upper bound on \( \pi(C_K) \):

\[
\pi(C_K) \leq b_2 \rho_s^K (K+1)^{s-1},
\]

where \( b_2 = (1 - \rho_s)^s \rho_s^{n-s} \prod_{i=s+1}^{n} (1 - \rho_i)/(\rho_s - \rho_i) \), independent of \( K \).

To obtain the lower bound on \( \pi(C_K) \), we define

\[
B_K = \{ \vec{x} \in S : x_1 + \cdots + x_s = K, x_{s+1} = \cdots = x_n = 0 \},
\]

the set of states in which there are \( K \) customers at maximally loaded nodes \( \{1, 2, \cdots, s\} \) and no customers anywhere else. Clearly \( B_K \subseteq C_K \). Thus we have

\[
\pi(C_K) \geq \pi(B_K)
\]

\[
= \sum_{x_1 + \cdots + x_s = K} \prod_{i=1}^{s} (1 - \rho_i) \rho_i^{x_i} \prod_{i=s+1}^{n} (1 - \rho_i)
\]

\[
= (1 - \rho_s)^s \rho_s^K \prod_{i=s+1}^{n} (1 - \rho_i) \left( \frac{K+s-1}{s-1} \right) \text{ (Feller [5] p.38)}
\]

\[
\geq b_1 \rho_s^K (K+1)^{s-1},
\]
where $b_1 = (1 - \rho_*) \prod_{i=s+1}^{n}(1 - \rho_i)/(s-1)!$.

**Remark** Glasserman and Kou [7] obtained other bounds on $\pi(C_K)$ as follows:

$$\rho^K \prod_{i=1}^{n}(1 - \rho_i) \leq \pi(C_K) \leq \rho^K \prod_{i=1}^{n}(1 - \rho_i)(K+1)^{n-1}. \tag{3}$$

In general, the lower bound in Lemma 1 appears to be stronger than that in (3). When $s = 1$, the lower bounds of two cases are the same. For the upper bound, when $s = n$, the bound in (3) meets that in Lemma 1, whereas for a large value $K$ the bound derived in Lemma 1 is sharper than that in (3).

**Theorem 2** Consider an exogenously supplied and open Jackson network which satisfies the stability condition. Then, for some positive constants $c_1, c_2$ that do not depend on the population size $K$ but may depend on the network parameters, we have

$$c_1 \pi(C_K) \leq p_K \leq c_2 \pi(C_K). \tag{4}$$

Further, we have the explicit estimate

$$c_2 = \frac{\sum(\bar{\lambda}_i + \mu_i)}{\pi(0) \sum \lambda_i}. \tag{5}$$

**Proof.** Let $\hat{X}(n)$ be a discrete-time Markov chain obtained by embedding at the virtual jump times of the original process $X(t)$. The virtual jump process is the sum of the exogenous arrival process and the virtual departure processes of the individual nodes. These are independent Poisson processes, with the future independent of the current state; hence the virtual jump process is Poisson with rate $\sum(\bar{\lambda}_i + \mu_i)$ with the future independent of the current state. It is an accepted consequence that the embedded chain $\hat{X}(n)$ has the same stationary distribution $\pi$ as the original process $X(t)$ (Walrand [13], p.279).

Next, let $\hat{Y}(n)$ be obtained from $\hat{X}(n)$ by watching it in the set $\{\hat{0}\} \cup C_K$. Then, $\hat{Y}(n)$ is also a discrete-time Markov chain with the stationary
distribution $\hat{\pi}$ given by

$$\hat{\pi}(\vec{x}) = \left( \sum_{\vec{y} \in \{\vec{0}\} \cup C_K} \pi(\vec{y}) \right)^{-1} \pi(\vec{x})$$

(Walrand [13] p.69) and the transition matrix $\hat{P}$ defined by $\hat{P}(\vec{x}, \vec{y}) = P\{\hat{Y}(n+1) = \vec{y}|\hat{Y}(n) = \vec{x}\}$ for all $\vec{x}, \vec{y} \in \{\vec{0}\} \cup C_K$. Specifically, we have

$$\hat{P}(0,0) = P\{\hat{X}(1) \neq \vec{0} | \hat{X}(0) = \vec{0}\} + P\{\hat{X}(1) = \vec{0} | \hat{X}(0) = \vec{0}\}$$

$$\sum_{i} \frac{\bar{\lambda}_i}{\sum (\lambda_i + \mu_i)} P\{\text{return to } \vec{0} \text{ before hitting } C_K | \hat{X}(0) = \vec{0}\}$$

$$= \sum_{i} \frac{\bar{\lambda}_i}{\sum (\lambda_i + \mu_i)} (1 - p_K) + \sum_{i} \frac{\mu_i}{\sum (\lambda_i + \mu_i)}$$

$$= 1 - p_K \sum_{i} \frac{\bar{\lambda}_i}{\sum (\lambda_i + \mu_i)}. \quad \text{(Anantharam and Ganesh [3])} \quad \text{(6)}$$

Therefore,

$$p_K = \sum_{\vec{x} \in C_K} \hat{P}(0, \vec{x}) \frac{\sum (\bar{\lambda}_i + \mu_i)}{\sum \lambda_i}$$

using $1 - \hat{P}(0,0) = \sum_{\vec{x} \in C_K} \hat{P}(0, \vec{x})$.

Now, let $\tilde{Y}(n)$ be the time reversal of $\hat{Y}(n)$, so $\tilde{Y}(n)$ is a Markov chain with the same stationary distribution $\hat{\pi}$ and its transition matrix $\tilde{P}$ given by

$$\tilde{P}(\vec{x}, \vec{y}) = \frac{\hat{\pi}(\vec{y})\hat{P}(\vec{y}, \vec{x})}{\hat{\pi}(\vec{x})}, \quad \vec{x}, \vec{y} \in \{\vec{0}\} \cup C_K.$$

Thus, $p_K$ can be rewritten as

$$p_K = \frac{1}{\pi(0)} \sum_{\vec{x} \in C_K} \pi(\vec{x}) \tilde{P}(\vec{x}, \vec{0}) \frac{\sum (\bar{\lambda}_i + \mu_i)}{\sum \lambda_i}. \quad \text{(7)}$$

Since $\tilde{P}(\vec{x}, \vec{0}) \leq 1$ for all $\vec{x} \in C_K$, the upper bound in (4) and the estimate $c_2$ in (5) are obtained from (7).
Similarly, let \( \hat{Z}(n) \) be a discrete-time Markov chain obtained from \( \hat{X}(n) \) by watching it in the set \( \{\vec{0}\} \cup B_K \), and let \( \hat{Q} \) be its transition matrix. Then, since \( B_K \subseteq C_K \), it is easy to see that

\[
\hat{P}(\vec{0}, \vec{0}) \leq \hat{Q}(\vec{0}, \vec{0}).
\]

Therefore, from the equation (6) and the stochasticity of \( \hat{Q} \) it follows that

\[
p_K \geq \sum_{\vec{x} \in B_K} \hat{Q}(\vec{0}, \vec{x}) \frac{\sum(\tilde{\lambda}_i + \mu_i)}{\sum \lambda_i}.
\]  

(8)

Let \( \tilde{Q} \) denote the transition matrix of the time reversal of \( \hat{Z}(n) \). Then we have

\[
\tilde{Q}(\vec{x}, \vec{0}) = \frac{\pi(\vec{0})\hat{Q}(\vec{0}, \vec{x})}{\pi(\vec{x})}, \quad \vec{x} \in \{\vec{0}\} \cup B_K.
\]

Substituting for \( \hat{Q}(\vec{0}, \vec{x}) \) in (8) now gives

\[
p_K \geq \frac{1}{\pi(\vec{0})} \sum_{\vec{x} \in B_K} \pi(\vec{x})\tilde{Q}(\vec{x}, \vec{0}) \frac{\sum(\tilde{\lambda}_i + \mu_i)}{\sum \lambda_i}.
\]  

(9)

If we can show, for each \( \vec{x} \in B_K \), that \( \tilde{Q}(\vec{x}, \vec{0}) \) is bounded below by a positive constant, which is independent of \( K \), the lower bound for \( p_K \) is given by

\[
p_K \geq \text{Const} \cdot \pi(B_K),
\]  

(10)

where \( \text{Const} \) denotes a positive constant, independent of \( K \).

Let \( \hat{X}(t) \) be the time reversal of the original process \( X(t) \). Then, it is known that the time reversal \( \hat{X}(t) \) is a Markov jump process for a different Jackson network, with the same number of nodes but different parameters as follows (Walrand [13] p.61):

\[
\begin{align*}
\tilde{\lambda}_i &= \lambda_i r_i, \quad i = 1, 2, \ldots, n \\
\tilde{\mu}_i &= \mu_i, \quad i = 1, 2, \ldots, n \\
\tilde{r}_{ij} &= \frac{\lambda_j}{\lambda_i} r_{ji}, \quad i, j = 1, 2, \ldots, n \\
\tilde{r}_i &= \frac{\lambda_i}{\lambda_i}, \quad i = 1, 2, \ldots, n,
\end{align*}
\]
where tildes refer to the corresponding quantities in the reversed Jackson network. Also, if the original network is exogenously supplied and open, then so is its time reversal. The solutions to the traffic equations in the time-reversed network are also the same, i.e.,

\[ \tilde{\lambda}_i = \lambda_i, \quad i = 1, 2, \cdots, n. \]

Let \( \tilde{X}_K(t) \) denote the Markov jump process when the time reversal \( \tilde{X}(t) \) is started with \( \vec{x} \in B_K \). It was shown in Anantharam et al. [2] that the process \( \tilde{X}_K(t) \) converges to a fluid limit \( X^f(t) \) in the sense that, for any \( \epsilon_0 > 0 \) and all \( \epsilon > \epsilon_0 \),

\[
\lim_{K \to \infty} P \{ \sup_{0 \leq t \leq T} \left\| \frac{1}{K} \tilde{X}_K(Kt) - X^f(t) \right\| \geq \epsilon \mid \left\| \frac{1}{K} \tilde{X}_K(0) - X^f(0) \right\| < \epsilon_0 \} = 0, \tag{11}
\]

where \( \|X\| = \max|X_i| \) and \( T = \inf \{ t > 0 : X^f_i(t) = 0 \text{ for all } i = 1, \cdots, n \} \). It was also proved in Anantharam and Ganesh [3] that \( \sum_{i=1}^{n} X^f_i(t) \), the total quantity of fluid in the network, is strictly decreasing at a positive rate as long as the amount of fluid is not zero. Furthermore, the fluid limit \( X^f_i(t) \) at node \( i \) stays at zero after it reaches zero until the total amount of fluid becomes empty. In other words, if we let \( T_i := \inf \{ t > 0 : X^f_i(t) = 0 \} \) for \( i = 1, 2, \cdots, n \), then

\[
X^f_i(t) = 0 \quad \text{for all } T_i \leq t \leq T, \tag{12}
\]

where \( T \) is the time at which the total amount of fluid \( \sum_{i=1}^{n} X^f_i(t) \) hits zero. By using the relation to this fluid limit we can obtain bounds on the process \( \sum_{i=1}^{s} \tilde{X}_K^i(t) \) of the total number of customers in the maximally loaded nodes \( \{1, 2, \cdots, s\} \).

Observe that

\[
\lim_{K \to \infty} \frac{1}{K} \tilde{X}_K^i(0) = X^f_i(0)
\]

exists for \( i = 1, 2, \cdots, n \). Then, from (11) and (12) we determine that the following statements are true with probability going to one as \( K \) goes to
infinity;

\[ \sum_{i=1}^{s} \tilde{X}_i^K(t) < \epsilon K \quad \text{for all} \quad K \cdot \max(T_1, T_2, \cdots, T_s) \leq t \leq KT \quad (13) \]

and

\[ \sum_{i=1}^{n} \tilde{X}_i^K(KT) < \epsilon K \quad (14) \]

for all \( \epsilon > 0 \).

Let \( T_K \) denote the first time that the actual queue length process \( \tilde{X}^K(t) \) hits the state \( \vec{0} \). Then, by applying Corollary 1 in Anantharam [1] it follows from (14) that \( T_K - KT \) is stochastically dominated by the sum of \( \epsilon K \) independent, identically distributed random variables of finite mean and variance. Since the exogenous arrival process is Poisson of rate \( \sum \bar{\lambda}_i \), the total number of exogenous arrivals in the period \([KT, T_K]\) (taken to be empty if \( KT > T_K \)) is less than a constant times \( \epsilon K \), with probability going to one as \( K \to \infty \). This implies that

\[ \sum_{i=1}^{n} \tilde{X}_i^K(t) < \text{Const} \cdot \epsilon K \quad \text{for all} \quad KT \leq t \leq T_K \]

with asymptotic probability one, where the constant is independent of \( K \) and \( \epsilon > 0 \) is arbitrary. This allows us to extend the validity of (13) through the period \([K \cdot \max(T_1, T_2, \cdots, T_s), T_K]\), that is,

\[ \sum_{i=1}^{s} \tilde{X}_i^K(t) < \text{Const} \cdot \epsilon K \quad \text{for all} \quad K \cdot \max(T_1, T_2, \cdots, T_s) \leq t \leq T_K, \quad (15) \]

with asymptotic probability one.

Now, we investigate the process \( \sum_{i=1}^{s} \tilde{X}_i^K(t) \) during the time period \((0, K \cdot \max(T_1, T_2, \cdots, T_s))\). Let \( \tilde{X}^a(t) \) denote the process started in the same initial condition as \( \tilde{X}^K(t) \), but with the output of the nodes in \( \{1, 2, \cdots, s\} \) replaced by their virtual departure processes of the reversed network \( \tilde{X}(t) \). Then, the queue length process \( \tilde{X}^a(t) \) dominates \( \tilde{X}^K(t) \), that is, for all sample paths \( \omega \),

\[ \tilde{X}_i^a(t; \omega) \geq \tilde{X}_i^K(t; \omega), \quad i = 1, \cdots, n, \quad t \geq 0. \]
To see this, color red the virtual departures from nodes in \{1, 2, \cdots, s\} that are not actual departures, color blue all other departures from all nodes and exogenous arrivals. Note that red customers can arrive only when at least one node in \{1, 2, \cdots, s\} is empty. The idea is that when a service occurs at a node with nonempty queue, we are free to decide which customer in the queue departs without affecting the process of total number of customers at the nodes. Blue customers always have precedence over red customers, i.e. when a service takes place at node \(i\), red customer in queue at node \(i\) does not move unless there are no blue customers in queue at node \(i\).

Then, we can see that \(\tilde{X}^K(t)\) is the process of blue customers, while \(\tilde{X}^a(t)\) is the process of all customers (Anantharam and Ganesh [3])). Observe that \(\tilde{X}^a(t)\) evolves like an unstable Jackson network, the inflow rates into whose queues are given by the solution \((\lambda^a_1, \lambda^a_2, \cdots, \lambda^a_n)\) to the generalized traffic equations (Goodman and Massey [8]):

\[
\lambda^a_i = \tilde{\lambda}_i + \sum_{j=1}^{s} \tilde{\mu}_{ji} \tilde{r}_{ij} + \sum_{j=s+1}^{n} \min(\lambda^a_j, \tilde{\mu}_j) \tilde{r}_{ij}, \quad i = 1, 2, \cdots, n. \tag{16}
\]

We aggregate the customers in the nodes \{1, 2, \cdots, s\} to create a combined node \(\kappa\) for the process \(\tilde{X}^a(t)\). Exogenous customers arrive at node \(\kappa\) at rate \(\bar{\lambda}_\kappa := \sum_{i=1}^{s} \tilde{\lambda}_i\). Since the departure processes out of the nodes \{1, 2, \cdots, s\} are replaced by the corresponding virtual departure processes, the service rate of the new node is to be \(\mu_\kappa := \sum_{i=1}^{s} \tilde{\mu}_i\). After the service completion, the customer returns to itself with rate \(\mu_\kappa r_{\kappa\kappa} := \sum_{i=1}^{s} \tilde{\mu}_i \tilde{r}_{ii}\), leaves the network with rate \(\mu_\kappa r_{\kappa} := \sum_{i=1}^{s} \tilde{\mu}_i \tilde{r}_{i}\), or transfers to the node \(j \in \{s+1, \cdots, n\}\) with rate \(\mu_\kappa r_{\kappa j} := \sum_{i=1}^{s} \tilde{\mu}_i \tilde{r}_{ij}\). Customers are transferred from \(j \in \{s+1, \cdots, n\}\) to the node \(\kappa\) with probability \(r_{j\kappa} := \sum_{i=1}^{s} \tilde{r}_{ij}\).

These rates define a new Jackson network on the nodes \{\kappa\} \cup \{s+1, \cdots, n\}. Notice that \(\sum_{i=1}^{s} \tilde{X}^a_i(t)\) is the process of the number of customers at node \(\kappa\). We first solve the traffic equations for this new Jackson network with \(\lambda_\kappa := \sum_{i=1}^{s} \lambda^a_i\) and the \(\lambda^a_i\) for \(i = s+1, \cdots, n\) in (16) as follows:

\[
\tilde{\lambda}_\kappa + \mu_\kappa r_{\kappa\kappa} + \sum_{j=s+1}^{n} \min(\lambda^a_j, \tilde{\mu}_j) r_{j\kappa} = \sum_{i=1}^{s} \tilde{\lambda}_i + \sum_{i=1}^{s} \sum_{j=1}^{s} \tilde{\mu}_{ji} \tilde{r}_{ji} + \sum_{i=1}^{s} \sum_{j=s+1}^{n} \min(\lambda^a_j, \tilde{\mu}_j) \tilde{r}_{ji}\]
\[
\sum_{i=1}^{s} \lambda_i^a = \lambda_\kappa, \\
\text{and for } i = s+1, \cdots, n, \\
\tilde{\lambda}_i + \mu_\kappa r_{ni} + \sum_{j=s+1}^{n} \min(\lambda_j^a, \bar{\mu}_j) \tilde{r}_{ji} = \tilde{\lambda}_i + \sum_{j=1}^{s} \tilde{\mu}_j \tilde{r}_{ji} + \sum_{j=s+1}^{n} \min(\lambda_j^a, \bar{\mu}_j) \tilde{r}_{ji} = \lambda_i^a.
\]

Secondly, we check the light traffic condition of this new Jackson network. Let us define
\[
\phi_i(\bar{\eta}) = \tilde{\lambda}_i + \sum_{j=1}^{s} \tilde{\mu}_j \tilde{r}_{ji} + \sum_{j=s+1}^{n} \min(\eta_j, \bar{\mu}_j) \tilde{r}_{ji}, \quad i = 1, 2, \cdots, n
\]
for a vector \( \bar{\eta} := (\eta_1, \eta_2, \cdots, \eta_n) \). Then, the solution \( (\lambda_1^a, \lambda_2^a, \cdots, \lambda_n^a) \) to the generalized traffic equation in (16) is the unique fixed point of \( \phi \). For \( \tilde{\lambda} := (\lambda_1, \lambda_2, \cdots, \lambda_n) \), the solution to the traffic equation of the original Jackson network, we have
\[
\phi_i\left(\frac{1}{\rho^*} \tilde{\lambda}\right) = \tilde{\lambda}_i + \sum_{j=1}^{s} \tilde{\mu}_j \tilde{r}_{ji} + \sum_{j=s+1}^{n} \min\left(\frac{1}{\rho^*} \lambda_j, \bar{\mu}_j\right) \tilde{r}_{ji} \leq \frac{1}{\rho^*} \tilde{\lambda}_i + \sum_{j=1}^{s} \tilde{\mu}_j \tilde{r}_{ji} + \sum_{j=s+1}^{n} \frac{1}{\rho^*} \lambda_j \tilde{r}_{ji} = \frac{1}{\rho^*} \lambda_i,
\]
since \( \rho^* = \rho_1 = \cdots = \rho_n < 1 \). From the fact that \( \phi_i \) is increasing it follows that \( \lambda_i^a \leq \lambda_i/\rho^* \) for all \( i = 1, 2, \cdots, n \). Therefore for \( i = 1, 2, \cdots, s \), we have \( \lambda_i^a \leq \mu_i \) and for \( i = s+1, \cdots, n \), we obtain \( \lambda_i^a < \mu_i = \bar{\mu}_i \) from \( \rho_i < \rho^* \). Hence the light traffic conditions on nodes \( \{s+1, \cdots, n\} \) hold. Next, we check the stability of node \( \kappa \). Assume that \( \lambda_i^a = \mu_i \) for all \( i = 1, \cdots, s \). Since \( \mu_i = \tilde{\mu}_i \) for \( i = 1, \cdots, n \) and \( \lambda_i^a < \bar{\mu}_i \) for \( i = s+1, \cdots, n \), \( (\lambda_1^a, \lambda_2^a, \cdots, \lambda_n^a) \) satisfies the traffic equation of the original time-reversed Jackson network as follows:
\[
\lambda_i^a = \tilde{\lambda}_i + \sum_{j=1}^{s} \tilde{\mu}_j \tilde{r}_{ji} + \sum_{j=s+1}^{n} \min(\lambda_j^a, \bar{\mu}_j) \tilde{r}_{ji}.
\]
\[ \tilde{\lambda}_i + \sum_{j=1}^{s} \lambda_j \tilde{r}_{ji} + \sum_{j=s+1}^{n} \lambda_j \tilde{r}_{ji}, \quad i = 1, 2, \ldots, n. \]

Therefore \( \mu_i = \lambda_i = \lambda_i \) for all \( i = 1, \ldots, s \), which contradicts to the stability of the original network. Hence for at least one \( i \in \{1, \ldots, s\} \), \( \lambda_i < \mu_i \), from which we can get \( \sum_{i=1}^{s} \lambda_i < \sum_{i=1}^{s} \mu_i \), that is, \( \lambda_{\kappa} < \mu_{\kappa} \). Consequently, the Jackson network on the nodes \( \{\kappa\} \cup \{s + 1, \ldots, n\} \) is stable.

If we consider an initial condition where queues outside \( \{1, 2, \ldots, s\} \) are in their stationary distributions, the external arrival process into node \( \kappa \) will be Poisson of rate \( \lambda_{\kappa} - \mu_{\kappa} \). This dominates the external arrival process into node \( \kappa \) in the process \( \tilde{X}^a(t) \), wherein the queues outside \( \{1, 2, \ldots, s\} \) were initially empty. This can also be shown using the coloring technique employed above. Color red all customers who are in \( \{s + 1, \ldots, n\} \) initially and color blue all other customers who arrive from outside the network. If we let blue customers have preemptive service priority at each node in \( \{s + 1, \ldots, n\} \), then blue customers who get into the node \( \kappa \) constitute the external arrival process while all customers including red customers who arrive at node \( \kappa \) form a Poisson process. We thus see that \( \sum_{i=1}^{s} \tilde{X}^a_i(t) \) is dominated by a Markov jump process \( M(t) \), of arrival rate \( \lambda_{\kappa} - \mu_{\kappa} r_{\kappa\kappa} \), service rate \( \mu_{\kappa} \), the transition probability to itself after the service \( r_{\kappa\kappa} \), and started at \( M(0) = K \). It can be seen that for the process \( M(t) \),

\[
P\{M(t) = 0 \text{ before } M(t) = K\} = \frac{\mu_{\kappa}(1 - r_{\kappa\kappa})}{\lambda_{\kappa} + \mu_{\kappa}(1 - r_{\kappa\kappa})} P_{K-1},
\]

where the first term on the right-hand side is the probability that \( M(t) \) is decreased by one before it is increased by one or jumps to itself and \( P_{K-1} \) denotes the probability that \( M(t) \) hits 0 before \( K \), started at \( K - 1 \). Here we can find \( P_{K-1} \) by the first step method (Walrand [13] p.56). Let, for \( 0 \leq i \leq K \), we define

\[
P_i := P\{M(t) = 0 \text{ before } M(t) = K\mid M(0) = i\}.
\]

Clearly, \( P_0 = 1 \) and \( P_K = 0 \). The first step equations give

\[
P_i = \frac{\mu_{\kappa}(1 - r_{\kappa\kappa})}{\lambda_{\kappa} + \mu_{\kappa}(1 - r_{\kappa\kappa})} P_{i-1} + \frac{\mu_{\kappa} r_{\kappa\kappa}}{\lambda_{\kappa} + \mu_{\kappa}(1 - r_{\kappa\kappa})} P_i + \frac{\lambda_{\kappa} - \mu_{\kappa} r_{\kappa\kappa}}{\lambda_{\kappa} + \mu_{\kappa}(1 - r_{\kappa\kappa})} P_{i+1},
\]

\[1 \leq i \leq K - 1\]
which can be rewritten as

\[ P_i = qP_{i-1} + pP_{i+1}, \quad 1 \leq i \leq K - 1 \]

with \( p := \frac{\lambda_\kappa - \mu_\kappa r_{\kappa\kappa}}{\lambda_\kappa + \mu_\kappa (1 - 2r_{\kappa\kappa})} \) and \( q := 1 - p \). Note that \( 0 < p < q < 1 \) because \( \mu_\kappa r_{\kappa\kappa} < \lambda_\kappa < \mu_\kappa \) and \( r_{\kappa\kappa} < 1 \). The solution of these linear equations is well known as the gambler’s ruin probability (Feller [5], p.345) given by

\[ P_i = \left( \frac{q}{p} \right)^{K - 1} - \left( \frac{q}{p} \right)^{i - 1}, \quad 0 \leq i \leq K. \]

Thus it follows that

\[ P_{K-1} = \left( \frac{q}{p} \right)^{K - 1} - \left( \frac{q}{p} \right)^{K - 1} > (1 - \frac{q}{p})^{-1} \]

for all \( K \geq 1 \). Hence, for all \( K \geq 1 \), we have

\[ P\{ M(t) = 0 \text{ before } M(t) = K \} > \frac{\mu_\kappa - \lambda_\kappa}{\lambda_\kappa + \mu_\kappa (1 - r_{\kappa\kappa})}. \quad (17) \]

Let \( T_0 := \inf\{ t > 0 : M(t) = 0 \} \). Since the stable Markov jump process \( M(t) \) does not grow by \( K \) in time linear in \( K \), with probability one, we can have

\[
\lim_{K \to \infty} P\{ M(t) < K \text{ for all } T_0 \leq t \leq K \cdot \max(T_1, \ldots, T_s) \} = 1.
\]

Hence \( \sum_{i=1}^{s} \tilde{X}_i^K(t) \), which is dominated by \( M(t) \), does not hit \( B_K \) before the time \( K \cdot \max(T_1, \ldots, T_s) \) with probability bounded away from zero since \( \lambda_\kappa < \mu_\kappa \) in (17). Combining this with (15) gives \( \sum_{i=1}^{s} \tilde{X}_i^K(t) < K \) for all \( 0 \leq t \leq T_K \) with a positive probability, independent of \( K \). Thus \( \tilde{X}(t) \) with initial state \( \vec{x} \in B_K \) satisfies

\[
\liminf_{K \to \infty} P\{ \tilde{X}(t) = \vec{0} \text{ before } \tilde{X}(t) \text{ hits } B_K \} > 0.
\]

Then, since the time reversal of the watching of the embedding is the same as the watching of the embedding the time reversal, we have that for all \( \vec{x} \in B_K \), \( \tilde{Q}(\vec{x}, \vec{0}) > 0 \), independent of \( K \).
Now we need to bound $\pi(B_K)$ below by $\pi(C_K)$ in order to complete the proof. Given $\vec{x} \in C_K$ with $x_1 = i_1, x_2 = i_2, \ldots, x_{s-1} = i_{s-1}$ we have

$$\sum_{x_s + \cdots + x_n = K - i_1 - \cdots - i_{s-1}} \pi(i_1, \ldots, i_{s-1}, x_s, \ldots, x_n) = (1 - \rho_*)^{s-1} \rho_*^{i_1 + \cdots + i_{s-1}} \prod_{x_s + \cdots + x_n = K - i_1 - \cdots - i_{s-1}, i = s} (1 - \rho_i) \rho_i^x_i \leq (1 - \rho_*)^{s-1} \rho_*^K \prod_{i=s+1}^n (1 - \rho_i) \prod_{i=s+1}^n \frac{\rho_s}{\rho_s - \rho_i} = \pi(i_1, \ldots, i_{s-1}, K - i_1 - \cdots - i_{s-1}, 0, \ldots, 0) \prod_{i=s+1}^n \frac{\rho_s}{\rho_s - \rho_i}.$$

Therefore

$$\pi(C_K) \leq \pi(B_K) \prod_{i=s+1}^n \frac{\rho_s}{\rho_s - \rho_i}. \quad (18)$$

From (10) and (18) we finally obtain the lower bound on $p_K$ given by

$$p_K \geq c_1 \pi(C_K),$$

where $c_1$ is a positive constant, independent of $K$. For large enough $K$, from (9), (17) and (18) the explicit estimate for $c_1$ is given by

$$c_1 = \frac{(\mu_* - \lambda_n) \sum (\tilde{\lambda}_i + \mu_i)}{\pi(0)[\lambda_n + \mu_n(1 - r_{\infty n})] \sum \lambda_i} \prod_{i=s+1}^n \frac{\rho_s}{\rho_s - \rho_i}.$$

Combining Lemma 1 and Theorem 2 results in the following corollary which is a stronger result than (1).

**Corollary 1** For an exogenously supplied and open stable Jackson network in which $s$ nodes have the same maximal load $\rho_*$,

$$\lim_{K \to \infty} \frac{\log p_K - \log \rho_*^K}{\log K} = s - 1.$$
3 Concluding remarks

We obtained new bounds on the overflow probability $p_K$ in Jackson networks. The improved lower bound of $p_K$ is useful for analyzing the performance of an importance sampling estimator for $p_K$. If there is only one maximally loaded node in the Jackson tandem network, without the additional assumption in Glasserman and Kou [7] we can conclude that the importance sampling estimator proposed in Parekh and Walrand [12] has the bounded relative error in certain parameter regions.

References


