2000 Proceedings of the Autumn Conference, Korean Statistical Society

Bounds on the Overflow Probability in Jackson Networks

Jiyeon Lee

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 the stronger logarithmic limit than that in Glasserman and Kou(1995) when several nodes have the same maximal load.
KEY WORDS: Jackson network, Overflow probability, Asymptotics

1 Introduction

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

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

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 probability by simulation has been studied by Parekh and Walrand(1989), McDonald(1999) and Lee(2000), etc. Simulation-base approaches to this problem include an asymptotic approach based on the large deviation theory. Glasserman and Kou(1995) proved the following logarithmic limit;

$$\lim_{K\to\infty}\frac{1}{K}\log p_K = \log\rho_*,$$

where ρ_* 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 several nodes have the same maximal load ρ_*. 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 Bounds on the Overflow Probability

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

*Assistant Professor, Department of Statistics, Yeungnam University, Daedong 214-1, Kyeongsan, Kyeongbuk, Korea

a Poisson process with rate $\bar{\lambda}_i$ and, if necessary, wait in a buffer to get served until the station gets free. 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 assume that the network is both exogenously supplied and open. A network is *exogenously* supplied if each node i has an exogenous arrival rate $\bar{\lambda}_i \neq 0$ or can be fed by another node j for which $\bar{\lambda}_j \neq 0$. The network is also *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 say that node i feeds node j if there is a sequence $k_1, k_2, \cdots, k_q$ such that $r_{i\,k_1} r_{k_1\,k_2} \cdots r_{k_q\,j} > 0$.

A Jackson network can be described as a Markov jump process $\{X(t); t \geq 0\}$ on $\mathcal{S} \equiv \mathbf{N}^n$, where the state $\vec{x} = (x_1, x_2, \cdots, x_n) \in \mathcal{S}$ depicts the system when there are x_i customers waiting or being served at node i.

Jackson (1957) gave an expression for the invariant measure of a Jackson network. His results are restated in the following theorem (see Brémaud (1981)).

Theorem 2.1 *For an exogenously supplied and open Jackson network for which the solution* $(\lambda_1, \cdots, \lambda_n)$ *to the traffic equations*

$$\lambda_i = \bar{\lambda}_i + \sum_{j=1}^{n} \lambda_j r_{ji}, \quad i = 1, 2, \cdots, n \tag{2.1}$$

satisfies the light traffic conditions

$$\rho_i := \frac{\lambda_i}{\mu_i} < 1, \quad i = 1, 2, \cdots, n, \tag{2.2}$$

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

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

The ratio ρ_i is called the *load* on node i. Call a Jackson network *stable* if the light traffic conditions (2.2) hold. We assume the light traffic conditions hold. We further assume that s nodes, say $\{1, 2, \cdots, s\}$, have the same maximal load ρ_*, that is, $\rho_* = \rho_1 = \rho_2 = \cdots = \rho_s > \rho_{s+1} \geq \cdots \geq \rho_n$.

Define the *overflow set*

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

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

Lemma 2.2

$$b_1 \rho_*^K (K+1)^{s-1} \leq \pi(C_K) \leq b_2 \rho_*^K (K+1)^{s-1}$$

for positive constants b_1 *and* b_2.

Proof: The left inequality follows from the fact that the state in which there are K customers at maximal-loaded nodes $\{1, \cdots, s\}$ and no customers anywhere else is an element of C_K.

$$\begin{aligned} \pi(C_K) &\geq \sum_{x_1 + \cdots + x_s = K} \prod_{i=1}^{s} (1 - \rho_*) \rho_*^{x_i} \prod_{i=s+1}^{n} (1 - \rho_i) \\ &\geq b_1 \rho_*^K (K+1)^{s-1}, \end{aligned}$$

where $b_1 = (1-\rho_*)^s \prod_{i=s+1}^n (1-\rho_i)/(s-1)!$.

From Theorem 2.1, on the other hand, we have

$$\begin{aligned}
\pi(C_K) &= \sum_{x_1+\cdots+x_n=K} \prod_{i=1}^n (1-\rho_i)\rho_i^{x_i} \\
&= \prod_{i=1}^n (1-\rho_i)\rho_*^K \sum_{k=0}^K \sum_{x_{s+1}+\cdots+x_n=k} \binom{K-k+s-1}{s-1} \prod_{i=s+1}^n (\rho_i/\rho_*)^{x_i} \\
&\le b_2 \rho_*^K (K+1)^{s-1}
\end{aligned}$$

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

Theorem 2.3 *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*

$$c_1 \pi(C_K) \le p_K \le c_2 \pi(C_K).$$

Further, we have the explicit estimate

$$c_2 = \frac{\sum \bar{\lambda}_i + \mu_i}{\pi(\vec{0}) \sum \bar{\lambda}_i}. \tag{2.3}$$

Proof: Let $\hat{X}(n)$ be a discrete-time Markov chain obtained by embedding at the virtual jump times of the original process $X(t)$. Then the embedded process $\hat{X}(n)$ has the same stationary distribution π as the original process $X(t)$. Next, let $\hat{Y}(n)$ be obtained from $\hat{X}(n)$ by watching it in the set $\{\vec{0}\} \cup C_K$. Then, $\hat{Y}(n)$ is also a discrete-time Markov chain with the stationary distribution $\hat{\pi}$ given by

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

and the transition matrix $\hat{P}$. Specifically, we have

$$\hat{P}(\vec{0}, \vec{0}) = 1 - \frac{p_K \sum \bar{\lambda}_i}{\sum \bar{\lambda}_i + \mu_i}.$$

(see Walrand(1987) and Anantharam and Ganesh(1994)) Therefore,

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

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

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

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

Thus, p_K can be rewritten as

$$p_K = \frac{1}{\pi(\vec{0})} \sum_{\vec{x}\in C_K} \pi(\vec{x})\tilde{P}(\vec{x},\vec{0}) \frac{\sum \bar{\lambda}_i + \mu_i}{\sum \bar{\lambda}_i}. \tag{2.4}$$

Since $\tilde{P}(\vec{x},\vec{0}) \le 1$, we have the upper bounds in (2.3).

Let $B_K = \{\vec{x} \in \mathcal{S} | x_1 + \cdots + x_s = K, x_{s+1} = \cdots = x_n = 0\}$. Clearly $B_K \subseteq C_K$. For $x_1 = i_1, x_2 = i_2, \cdots, x_{s-1} = i_{s-1}$, we have

$$\begin{aligned}
&\sum_{x_s+\cdots+x_n=K-i_1-\cdots-i_{s-1}} \pi(i_1,\cdots,i_{s-1},x_s,\cdots,x_n)\\
&= (1-\rho_*)^{i_1+\cdots+i_{s-1}}\rho_*^{i_1+\cdots+i_{s-1}} \sum_{x_s+\cdots+x_n=K-i_1-\cdots-i_{s-1}} \prod_{i=s}^{n}(1-\rho_i)\rho_i^{x}\\
&\le (1-\rho_*)^{i_1+\cdots+i_{s-1}}\rho_*^{K} \prod_{i=s}^{n}(1-\rho_i) \prod_{i=s+1}^{n} \frac{\rho_*}{\rho_*-\rho_i}\\
&= \pi(i_1,\cdots,i_{s-1},K-i_1-\cdots-i_{s-1},0,\cdots,0) \prod_{i=s+1}^{n} \frac{\rho_*}{\rho_*-\rho_i}.
\end{aligned}$$

Therefore

$$\pi(C_K) \le \sum_{\vec{x}\in B_K} \pi(\vec{x}) \prod_{i=s+1}^{n} \frac{\rho_*}{\rho_*-\rho_i}. \tag{2.5}$$

Notice that p_K is no smaller than the sum in (2.4) restricted to B_K. Thus,

$$p_K \ge \frac{1}{\pi(\vec{0})} \sum_{\vec{x}\in B_K} \pi(\vec{x})\tilde{P}(\vec{x},\vec{0}) \frac{\sum \bar{\lambda}_i + \mu_i}{\sum \bar{\lambda}_i}. \tag{2.6}$$

If we can show, for each $\vec{x} \in B_K$, that $\tilde{P}(\vec{x},\vec{0})$ is bounded below by a positive constant, which is independent of K, then it follows from (2.5) and (2.6) that

$$p_K \ge c_1 \pi(C_K) \tag{2.7}$$

for some constant c_1. Let $\tilde{X}(t)$ denote the time reversal of $X(t)$ and let $\tilde{X}^K(t)$ the Markov jump process when $\tilde{X}(t)$ is started from B_K. It is shown in Anantharam et al.(1990) 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 $\epsilon > \epsilon_0$,

$$\lim_{K\to\infty} P(\sup_{0\le t\le T} \| \frac{1}{K}\tilde{X}^K(Kt) - X^f(t)\| \ge \epsilon \mid \|\frac{1}{K}\tilde{X}^K(0) - X^f(0)\| < \epsilon_0) = 0,$$

where $\|X\| = \max|X_i|$ and $T = \inf\{t > 0 | \tilde{X}_i^K(t) = 0 \text{ for all } i = 1,\cdots,n\}$. They also proved that the fluid limit $X^f(t)$ has a negative drift. More precisely, $\sum_{i=1}^n X_i^f(t)$, the total quantity of fluid in the network, is strictly decreasing at a positive rate as long as the network is not empty. Thus $\tilde{X}(t)$ with initial state $\vec{x} \in B_K$ satisfies

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

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 $\vec{x} \in B_K$, $\tilde{P}(\vec{x},\vec{0}) > 0$, independent of K.

Combining Lemma 2.2 and Theorem 2.3 gives the following corollary.

Corollary 2.4 *For an exogenously supplied and open stable Jackson network in which s nodes have the same maximal load* ρ_*,

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

References

Anantharam, V., Heidelberger, P. and Tsoucas, P. (1990) Analysis of Rare Events in Continuous Time Markov Chains via Time Reversal and Fluid Approximation, IBM Research Report.

Anantharam, V. and Ganesh, A. J. (1994) Correctness Within a Constant of an Optimal Buffer Allocation Rule of Thumb, *IEEE Transactions on Information Theory* **40**, 871-882.

Brémaud, P. (1981). *Point Processes and Queues*, Springer-Verlag. 354 p.

Glasserman, P. and Kou, S. G. (1995) Analysis of an Importance Sampling Estimator for Tandem Queues, *ACM Transactions on Modelling and Computer Simulation* **5**, 22-42.

Jackson, J.R. (1957). Networks of waiting lines, *Operations Research* **5**, 518-521.

J. Lee (2000). Fast Simulation for Excessive Backlogs in Tandem Networks, *The Korean Communications in Statistics* **7**, 499-511.

McDonald, D. (1999). Asymptotics of first passage times for random walk in a orthant. *Annals of Applied Probability* **9**, No. 1, 110-145.

Parekh, S. and Walrand, J. (1989) A quick simulation method for excessive backlogs in networks of queues, *IEEE Transactions on Automatic Control* **34**, no. 1, 54-66.

Walrand, J. (1987) *An Introduction to Queueing Networks*, Englewood Cliffs, NJ, Prentice-Hall.