Proceedings of Joint Conference of The Korean Data Analysis Society
and Korean Data And Information Science Society,
October 17-18, 2003, pp. 155-162

Asymptotics in Load-Balanced Tandem Networks

Jiyeon Lee[1]

Abstract

A tandem network in which all nodes have the same load is considered. We derive bounds on the probability that the total population of the tandem network exceeds a large value by using its relation to the stationary distribution. These bounds imply a stronger asymptotic limit than that in the large deviation theory.

Keywords : Tandem network, Asymptotic limit, Large deviation theory, Overflow probability.

1. Introduction

We consider n M/M/1 nodes in tandem as shown in Figure 1. A customer arrives at node 1 from outside according to a Poisson process with rate λ and, if necessary, waits in a buffer until the node gets free to get served. Once service is completed at node i, the customer is routed to the next node $i+1$ $(1 \le i \le n-1)$. After getting served at node n the customer finally leaves the network. Service time at node i is exponentially distributed with mean $1/\mu_i$ $(1 \le i \le n-1)$. Each node operates on a first-in-first-out basis.

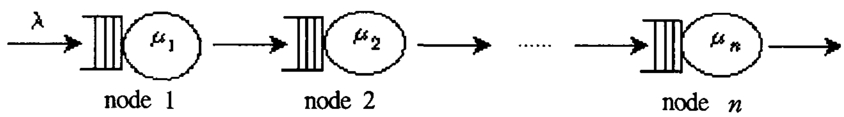

Figure 1. A tandem network of n M/M/1 nodes

We analyze an overflow probability p_K that the network population reaches a large value K before returning to 0, starting from 0. Glasserman and Kou(1995) proved the following asymptotic limit for p_K in the tandem network:

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

[1]Department of Statistics, Yeungnam University, Kyeongbuk 214-1, Korea

where $\rho_{.}$ is the load of the most highly loaded node in the network.

In this paper, we derive upper and lower bounds on p_K by the stationary distribution, and we then obtain the stronger asymptotic limit than (1) for the load-balanced tandem network in which all nodes have the same load.

2. Main Results

A tandem network can be described as a Markov jump process $X(t), t\geq 0$ on $S\equiv N^n$, where the state $\vec{x}=(x_1,x_2,\cdots,x_n)\in S$ depicts the system when there are x_i customers waiting or being served at node i. Under the condition that $\rho_i:=\lambda/\mu_i<1$ for all $i=1,2,\cdots,n$, the stationary distribution of $X(t)$ is given by

$$\pi(\vec{x})=\prod_{i=1}^{n}(1-\rho_i)\rho_i^{x_i},\ \vec{x}=(x_1,x_2,\cdots,x_n)\in S\ \text{(Walrand(1988))}.$$

In particular, if all node have the same service rate μ, the stationary distribution can be simplified as

$$\pi(\vec{x})=(1-\rho)^n\rho^{\sum_{i=1}^{n}x_i},\ \vec{x}=(x_1,x_2,\cdots,x_n)\in S$$

where $\rho:=\lambda/\mu$ denotes the same load of nodes.

Let us define an overflow set by

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

the set of states in which the network population is exactly K. Notice that the stationary distribution of C_K, $\pi(C_K)$ is given by

$$\begin{aligned}\pi(C_K) &= \sum_{\sum_{i=1}^{n}x_i=K}(1-\rho)^n\rho^{\sum_{i=1}^{n}x_i}\\ &= (1-\rho)^n\rho^K\sum_{\sum_{i=1}^{n}x_i=K}1 \\ &= \binom{K+n-1}{n-1}(1-\rho)^n\rho^K.\end{aligned}\tag{2}$$

In the following theorem, we show that the overflow probability p_K can be bounded by the stationary distribution of C_K, $\pi(C_K)$.

Theorem 1. For an open stable load-balanced tandem network,

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

Proof. Let $\hat{X}(n)$ be the uniformized(Walrand(1988)) Markov chain of the original process $X(t)$, where we assume without loss of generality that we have re-scaled time such that $\lambda+n\mu=1$. Then, the Markov chain $\hat{X}(n)$ 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

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

and the transition matrix $\hat{P}(\vec{x},\vec{y})$ for $\vec{x}$, $\vec{y}\in\vec{0}\cup C_K$. Specifically, we have

$$\begin{aligned}\hat{P}(\vec{0},\vec{0}) &= P\ \hat{Y}(1)=\vec{0}\,|\,\hat{Y}(0)=\vec{0}\\ &= P\ \hat{X}(n)=\vec{0}\ \mathit{before}\ \hat{X}(n)\in C_K|\,\hat{X}(1)\neq\vec{0},\ \hat{X}(0)=\vec{0}\\ &\quad\times P\ \hat{X}(1)\neq\vec{0}\,|\,\hat{X}(0)=\vec{0}+P\ \hat{X}(1)=\vec{0}\,|\,\hat{X}(0)=\vec{0}\\ &= \lambda P\ \hat{X}(n)=\vec{0}\ \mathit{before}\ \hat{X}(n)\in C_K|\,\hat{X}(1)\neq\vec{0},\ \hat{X}(0)=\vec{0}+n\mu\\ &= 1-p_K\lambda\end{aligned}$$

because $p_K=P\ \hat{X}(n)\in C_K\ \mathit{before}\ \hat{X}(n)=\vec{0}\,|\,\hat{X}(1)\neq\vec{0},\ \hat{X}(0)=\vec{0}$. Therefore,

$$p_K=\frac{1}{\lambda}\sum_{x\in C_K}\hat{P}(\vec{0},\vec{x}),$$

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

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

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

Thus, p_K can be rewritten as

$$p_K=\frac{1}{\lambda\,\pi(\vec{0})}\sum_{x\in C_K}\pi(\vec{x})\,\check{P}(\vec{x},\vec{0}).$$

Since $\check{P}(\vec{x},\vec{0})\le 1$ for all $\vec{x}\in C_K$, the estimate c_2 in the upper bound is given by

$$c_1 = [\lambda(1-\rho)^n]^{-1}.$$

Notice that for $\vec{x} \in C_K$ with $x_1 = 0$, $\tilde{P}(\vec{0}, \vec{x}) = 0$ because the process $\tilde{X}(n)$ needs at least one external arrival at node 1 to hit C_K for the first time. Thus $\tilde{P}(\vec{x}, \vec{0}) = 0$ for $\vec{x} \in C_K$ with $x_1 = 0$. If we let

$$C_K' := C_K - \vec{x} \in C_K : x_1 = 0, x_2 + \cdots + x_n = K,$$

clearly from (2) we have

$$\begin{aligned} \pi(C_K') &= \frac{K}{K+n-1}\pi(C_K) \\ &\geq \frac{1}{n}\pi(C_K) \end{aligned} \tag{3}$$

for all $K \geq 1$. Therefore, if we can show, for each $\vec{x} \in C_K'$, that $\tilde{P}(\vec{x}, \vec{0})$ is bounded below by a positive constant, which is independent of K, the lower bound for p_K is given by

$$p_K \geq c_1\, \pi(C_K), \tag{4}$$

where c_1 denotes a positive constant, independent of K.

Let $\tilde{X}(t)$ be the time reversal of the original process $X(t)$. Then, it is well known that the time reversal $\tilde{X}(t)$ is a Markov jump process for another tandem network with the same number of nodes and the same service rates but different arrival pattern (Walrand(1988)). In the reversed tandem network there is an external arrival process at node n and after service completion at node i the customer moves to node $i-1$, $2 \leq i \leq n$ and finally leaves the system at node 1. Figure 2 depicts the reversed tandem network evolved by $\tilde{X}(t)$.

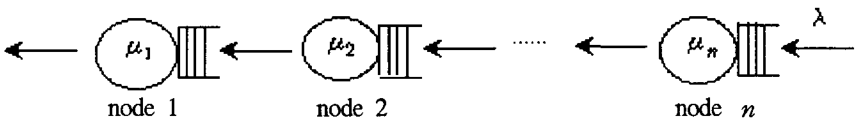

Figure 2. The reversed tandem network

Let $\tilde{X}^K(t)$ denote the Markov jump process when the time reversal $\tilde{X}(t)$ is started with $\vec{x} \in C_K'$. Anantharam et al.(1990) showed that the process $\tilde{X}^K(t)$ converges to a fluid limit $F(t)$ in the sense that, for any $\varepsilon_0 > 0$ and all $\varepsilon > \varepsilon_0$,

$$\lim_{K\to\infty} P\{\sup_{0\leq t\leq T} \mid \frac{1}{K}\tilde{X}^K(Kt) - F(t) \mid \geq \varepsilon \mid \mid \frac{1}{K}\tilde{X}^K(0) - F(0) \mid < \varepsilon_0\} = 0, \tag{5}$$

where $|X| = \max|X_i|$ and $T = \infty\ t > 0 : F_i(t) = 0$ for *all* $i = 1, \cdots, n$.

Anantharam and Ganesh(1994) also proved that $\sum_{i=1}^{n} F_i(t)$, the total amount of fluid in the network, is strictly decreasing at a positive rate as long as the amount of fluid is not zero, and the fluid limit $F_i(t)$ at node i stays at zero after it reaches zero until the total amount of fluid $\sum_{i=1}^{n} F_i(t)$ becomes empty. That is, if we let $T_i = \infty\ t > 0 : F_i(t) = 0$ for $i = 1, 2, \cdots, n$, then

$$F_i(t) = 0 \text{ for all } T_i \le t \le T, \tag{6}$$

where T is the time at which $\sum_{i=1}^{n} F_i(t)$ hits zero.

Notice that

$$\lim_{K \to \infty} \frac{1}{K} \tilde{X}_i^K(0) = F_i(0)$$

exists for $i = 1, 2, \cdots, n$. Then, from (5) and (6) we can determine that the following statements are true with probability going to one as K goes to infinity;

$$\sum_{i=1}^{n} \tilde{X}_i^K(t) < \varepsilon K$$

for all $K \cdot \max(T_1, T_2, \cdots, T_n) \le t \le KT$ and in particular,

$$\sum_{i=1}^{n} \tilde{X}_i^K(KT) < \varepsilon K. \tag{7}$$

Let T_K denote the first time when the queue length process $\tilde{X}^K(t)$ hits the state $\vec{0}$. Then, by applying Corollary 1 in Anantharam(1989) it follows from (7) that $T_K - KT$ is stochastically dominated by the sum of εK independent, identically distributed random variables of finite mean and variance. Since the external arrival process is Poisson of rate λ, the total number of external arrivals in the period $[KT, T_K]$ is less than a constant times εK, with probability going to one as $K \to \infty$. This implies that with asymptotic probability one,

$$\sum_{i=1}^{n} \tilde{X}_i^K(t) < const \cdot \varepsilon K$$

where the constant *const* is independent of K and is $\varepsilon > 0$ arbitrary. So it enables us to extend the validity of (7) through the period $[K \cdot \max(T_1, T_2, \cdots, T_n), T_K]$, that is,

$$\sum_{i=1}^{n} \tilde{X}_i^{K}(t) < const \cdot \varepsilon K \text{ for all } K \cdot \max(T_1, T_2, \cdots, T_n) \leq t \leq T_K \tag{8}$$

with asymptotic probability one.

Now, we investigate the process $\sum_{i=1}^{n} \tilde{X}_i^{K}(t)$ during the time period $(0, K \cdot \max(T_1, T_2, \cdots, T_n))$. Let $Z(t)$ denote the process started in the same initial condition as $\tilde{X}^{K}(t)$, but with the output of the nodes replaced by their virtual departure. Then, the queue length process $Z(t)$ dominates $\tilde{X}^{K}(t)$, that is, for any sample paths ω,

$$Z_i(t, \omega) \geq \tilde{X}_i^{K}(t, \omega), \ i = 1, 2, \cdots, n, \ t \geq 0.$$

To see this, we use the coloring method employed in Anantharam and Ganesh(1994). Color red the virtual departures from each node that are not actual departures and color blue all other departures from all nodes and external arrivals. Note that red customers can arrive only when at least one node is empty. The idea is that when a service occurs at a node with *non-empty 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 is no blue customer in queue at node i. Then, we can see that $\tilde{X}^{K}(t)$ is the process of blue customers, while $Z(t)$ is the process of all customers.

Observe that the process $\sum_{i=1}^{n} Z_i(t)$ is a Markov jump process with the arrival rate λ, the service rate $n\mu$, and the transition probability to itself after the service $(n-1)/n$, started from $\sum_{i=1}^{n} Z_i(0) = K$. Since $\lambda < \mu$, the process $\sum_{i=1}^{n} Z_i(t)$ is stable. Then, it can be seen that for the process $\sum_{i=1}^{n} Z_i(t)$,

$$P \sum_{i=1}^{n} Z_i(t) = 0 \ \ before \ \sum_{i=1}^{n} Z_i(t) = K$$
$$> \mu \times P \sum_{i=1}^{n} Z_i(t) = 0 \ \ before \ \sum_{i=1}^{n} Z_i(t) = K \,|\, \sum_{i=1}^{n} Z_i(0) = K-1,$$

where μ is the probability that $\sum_{i=1}^{n} Z_i(t)$ is decreased by one before it is increased by one or jumps to itself. From the gambler's ruin probability(Feller(1968)) we obtain

$$P\sum_{i=1}^{n} Z_i(t)=0 \textit{ before } \sum_{i=1}^{n} Z_i(t)=K| \sum_{i=1}^{n} Z_i(0)=K-1 = \frac{(\mu/\lambda)^K-(\mu/\lambda)^{K-1}}{(\mu/\lambda)^K-1} > 1-(\frac{\mu}{\lambda})^{-1} \tag{9}$$

for all $K\geq 1$.

Let $T_0 := \infty\ t>0: \sum_{i=1}^{n} Z_i(t)=0$. Since the stable Markov jump process $\sum_{i=1}^{n} Z_i(t)$ does not grow by K in time linear in K, with probability one, we can have

$$\lim_{K\to\infty} P\sum_{i=1}^{n} Z_i(t)<K \quad \text{for } \textit{all } T_0\leq t\leq K\cdot \max(T_1,\cdots,T_n)=1.$$

Hence, $\sum_{i=1}^{n} \tilde{X}_i{}^K(t)$ which is dominated by $\sum_{i=1}^{n} Z_i(t)$, does not hit C_K before the time $K\cdot\max(T_1,\cdots,T_n)$ with probability bounded away from zero because $\lambda<\mu$ in (9).

Combining this with (8) gives $\sum_{i=1}^{n} \tilde{X}_i{}^K(t)<K$ for all $0\leq t\leq T_K$ with a positive probability, independent of K. Thus $\tilde{X}(t)$ with initial state $\vec{x}\in C_K'$ satisfies

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

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 $\vec{x}\in C_K'$,

$$\tilde{P}(\vec{x},\vec{0})>0,$$

independent of K. So 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 (3) and (9) the explicit estimate for c_1 is given by

$$c_1=[n\,\rho(1-\rho)^{n-1}]^{-1}. \qquad \blacksquare$$

Corollary. For an open stable load-balanced tandem network of which load is ρ,

$$\lim_{K\to\infty}\frac{\log p_K-\log\rho^K}{\log K}=n-1. \tag{10}$$

Proof. Substituting (2) into the bounds on p_K in Theorem 1 and noting that

$$\lim_{K\to\infty} \frac{\log\binom{K+n-1}{n-1}}{\log K} = n-1$$

give the asymptotic limit in (10), which is stronger than (1). ■

References

1. Anantharam, V. (1989). The optimal buffer allocation problem, *IEEE transactions in Information Theory*, 35, 721-725.
2. Anantharam V., Heidelberger, P., and Tsoucas, P. Analysis of rare events in continuous time Markov chains via time reversal and fluid approximation, *IBM Research Report*, 1990.
3. 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.
4. 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.
5. Feller, W. *An Introduction to Probability Theory and Its Applications*, Volume I, Third Edition, John Wiley & Sons, Inc. 1968.
6. Walrand, J. *An Introduction to Queueing Networks*, Englewood Cliffs, NJ, Prentice-Hall, 1988.