*The Annals of Applied Probability*
1994, Vol. 4, No. 2, 414–431

# INSTABILITY OF FIFO QUEUEING NETWORKS

By Maury Bramson[1]

*University of Wisconsin*

Consider a queueing network with customers arriving according to a rate-1 Poisson process. Each customer proceeds along the same prescribed route, waiting at the different queues until exiting from the system. The service times are assumed to be independent and exponentially distributed. Individual queues may be visited more than once by a customer, with the mean service time perhaps depending on the stage along the route. The network is assumed to be first-in, first-out. An obvious necessary condition for such a queueing network to have an equilibrium distribution is that the sum of the mean service times at each queue be less than 1. We show by means of a class of examples that this condition does not suffice, these networks being unstable. Each such network possesses two queues, the first with one slow and one quick stage, and the other with one slow and numerous quick stages.

**1. Introduction.** A topic of considerable recent interest within queueing theory is the existence/nonexistence of equilibria for queueing networks. It seems intuitively clear that equilibria will exist for systems whose customers are served substantially more quickly than the rate at which they enter. The appropriate conditions on the service times are, however, not obvious, and it is easy to succumb to the temptation of overly broad conjectures. We present here a class of examples which illustrates this problem.

As motivation, we first review the situation for the standard M/M/1 or *simple* queue. Customers are assumed to enter a queue according to a rate-1 Poisson process. Equivalently, successive customers arrive after rate-1 exponential holding times. Customers are served one at a time, leaving the queue according to independent rate-$\lambda$ exponential holding times. Let $N_t$ denote the number of customers waiting to be served at time $t$. The process $N_t$ is a birth–death process on $\{0, 1, 2, \dots\}$, and so any equilibrium distribution must be reversible. Using this, one obtains the equilibria distributions

$$\pi_\lambda(k) = (1 - \lambda^{-1})\lambda^{-k}, \qquad k = 0, 1, 2, \dots, \tag{1}$$

for $\lambda > 1$. It is easy to check that for $\lambda \le 1$, there are no equilibria, and in fact for each $\lambda < 1$, the system is transient, with $N_t \to \infty$ as $t \to \infty$. Also, from elementary Markov chain theory, the above equilibria are, for given $\lambda$, unique.

---

Received April 1993; revised September 1993.

[1] Partially supported by NSF Grant DMS-93-00612.

*AMS* 1991 *subject classifications.* 60K25, 68M20.

*Key words and phrases.* Queueing networks, equilibrium distribution, instability, first-in, first-out.

A natural generalization of the above model is to networks consisting of a series of $m$ queues, $m \geq 1$. Customers enter the system, as before, according to a rate-1 Poisson process. Each customer proceeds along the same prescribed *route*, visiting all the queues, and then exiting from the system. The service times are again assumed to be independent and exponentially distributed. Individual queues may be visited more than once, and the rate a customer is served at a given queue may depend on his/her position, or *stage*, along the route. Denote by $\lambda_{i,j}$ the rate for the $j$th visit to the $i$th queue. There is also the issue of a priority for the order of service among customers at a given queue. A natural assumption is that the network is *first-in, first-out* (FIFO), that is, customers at a given queue are served in the order they arrive there, irrespective of the number of visits previously made.

A fundamental question is under what conditions on $\lambda_{i,j}$ do equilibria exist for the above FIFO queueing networks. If no equilibrium exists, the network is said to be *unstable*. For this, we introduce the notation $\mu_{i,j} = \lambda_{i,j}^{-1}$, $J(i) =$ the number of times the $i$th queue is visited along the route, and $\mu_i = \sum_{j=1}^{J(i)} \mu_{i,j}$. Here, $\mu_{i,j}$ is the mean service time for a customer during a visit at a queue, and $\mu_i$ can be interpreted as the total mean service time at $i$. It is not difficult to check that if $\mu_i \geq 1$ for some $i$, then the network is unstable.

A natural conjecture is that an equilibrium will exist under the condition

$$\mu_i < 1 \quad \text{for all } i. \tag{2}$$

It is well known that this is in fact the case if $\mu_{i,j}$ does not depend on $j$, that is, the rate a customer is served depends only on the queue, not on the stage along the route (see [2], [3]). Inspired by the M/M/1 queue, one can, under this restriction, construct equilibria for which the probability of there being $k_i$ customers at the $i$th queue, $i = 1, \dots, m$, is given by

$$\pi(k_1, \dots, k_m) = \pi_{\lambda_1}(k_1) \cdots \pi_{\lambda_m}(k_m), \tag{3}$$

where $\pi_\lambda$ is as in (1) and $\lambda_i = \mu_i^{-1}$. The different possible configurations of customers at different stages along the route with given $k_1, \dots, k_m$ are equally likely here.

It has been generally believed that the condition (2) should suffice even when $\mu_{i,j}$ is dependent on $j$ (e.g., [4] and [9]). In the context of priority queues (without the FIFO assumption), counterexamples are known ([5] and [6]). Here, we present a class of FIFO networks satisfying (2) which are unstable, the number of customers in the system tending to infinity as $t \to \infty$. These models possess two queues, 1 and 2, and customers whose route through the system is of the form $1 \to 2 \to 2 \to \cdots \to 2 \to 1$. The mean service times $\mu_{1,2}$ and $\mu_{2,1}$ are close to 1, and all other service times are small. The number of visits to the second queue by a customer is large.

Such examples raise the question of the nature of the critical value of the mean service times $\mu_i$, below which an equilibrium must always exist. A more complicated class of unstable FIFO networks with arbitrarily small bound, that is, for given $\mu > 0$, $\mu_i \leq \mu$ for all $i$, is given in [1].

The author has been informed of related work on FIFO networks by T. Seidman and A. Yershov. They construct deterministic systems with four types of customers whose service rates correspond to (2) [7], and deterministic systems with a single customer type [8]. The systems possess instantaneous stages.

The paper is structured as follows. In Section 2, we present the models, the main results and a summary of the proof. In Sections 3–6, the proof is broken down primarily according to the evolution of the network over four successive time intervals, after which the network returns to an amplified version of its original state. Repetition of this procedure demonstrates the instability of the network.

**2. The models.** The models are assumed to possess two queues, 1 and 2. Upon entering the system, customers move along the prescribed route $1 \to 2 \to 2 \to \cdots \to 2 \to 1$, at which point they exit from the system. The stage of a customer will be denoted by $(i, j)$, with $j = 1, 2$ for $i = 1$ and $j = 1, \dots, J$ for $i = 2$: the first coordinate denotes the queue, and the second coordinate denotes the number of times that the queue has been visited up to then. The mean service time for a customer is given by

$$\begin{aligned} &c \text{ at } (1,2) \text{ and } (2,1), \\ &\delta \text{ at } (1,1) \text{ and } (2,j), \qquad j = 2, \dots, J. \end{aligned} \tag{4}$$

We will assume that

$$\frac{399}{400} \le c < 1, \qquad c^J \le \frac{1}{50}, \qquad \delta \le \frac{(1-c)}{50J^2}. \tag{5}$$

($\delta = 0$ is allowed.) Under (5), the condition (2) is clearly satisfied. These restrictions can be weakened by a more parsimonious choice of bounds, but for our argument to work as organized, $c$ needs to be chosen fairly close to 1, and $\delta$ quite small. For the sake of concreteness, the reader can set $c = 399/400$, $J = 1600$ and $\delta = 10^{-11}$. Our main result is the following theorem.

THEOREM 1. *Any FIFO queueing network satisfying* (4) *and* (5) *is unstable, with the number of customers in the system approaching infinity as* $t \to \infty$.

We note that variations in the model can be made without affecting the main structure of the proof. For instance, the long string of 2's given above, $2 \to 2 \to \cdots \to 2$, can be replaced by the route segment $2 \to 3 \to 2 \to 3 \to \cdots \to 2 \to 3$, where the mean service time at $(3, j)$ is the same as at $(2, j)$, which is again chosen as in (4) and (5).

Let $\Xi_t$ denote the state at time $t$ of a queueing network with prescribed route $1 \to 2 \to 2 \to \cdots \to 2 \to 1$ as above, and satisfying (4). The symbol $\xi_t(i, j)$ will denote the number of customers at $(i, j)$ at time $t$, and $\xi_t$ will denote the total number of customers in the system. By $(i, j)^+$ [resp., $(i, j)^-$], we will mean the set of stages in the system strictly beyond (resp., before) $(i,j)$, and by $\xi_t(i,j)^+$ [resp., $\xi_t(i,j)^-$], we will mean the number of customers in $(i,j)^+$ [resp., $(i,j)^-$]. For instance,

$$\xi_t(2,j)^+ = \xi_t(1,2) + \sum_{l=j+1}^{J} \xi_t(2,l).$$

The proof of Theorem 1 rests on the following induction step.

THEOREM 2. *Assume that* (5) *holds for the process* $\Xi_t$, *and that*

$$\xi_0(1,1) = M, \qquad \xi_0(1,1)^+ \le M/50. \tag{6}$$

*Then for some* $\varepsilon_1 > 0$, *large enough* $M$ *and appropriate random* $T$ (*depending on* $M$),

$$P\big(\xi_T(1,1) \ge 100M,\ \xi_T(1,1)^+ \le M\big) \ge 1 - \exp(-\varepsilon, M) \tag{7}$$

*and*

$$P(\xi_t \ge M/4, \forall\, t \in [0,T]) \ge 1 - \exp(-\varepsilon, M). \tag{8}$$

We will later choose $T \approx 2cM/(1-c)$. ($a \approx b$ will mean that $a/b$ and $b/a$ are bounded.) Note that the factor 50 in (6) is not special, although we require the ratio $\xi_0(1,1)^+/\xi_0(1,1)$ to be small. The bound in (8) can be improved.

Suppose that $\Xi_0$ satisfies (6) for some large $M$. Repeated application of Theorem 2 yields

$$P(\xi_t < M/4 \text{ for some } t \ge 0) \le 2 \sum_{k=0}^{\infty} \exp\big[-\varepsilon, M(100)^k\big], \tag{9}$$

which approaches 0 as $M \to \infty$. Since all states in the system are accessible from one another, (9) implies that $\xi_t \to \infty$ as $t \to \infty$ w.p.1 for any $\Xi_0$. Theorem 1 thus follows from Theorem 2.

The conditions given in (4) are related to those for the priority queue example in [5], where the route is of the form $1 \to 2 \to 2 \to 1$. The model is deterministic, with the time between arrivals equal to 1 and service times having duration $2/3$ for the 2nd and 4th stages, and 0 (that is, instantaneous) for the 1st and 3rd stages. Priority is given to customers at the 2nd and 4th stages. One can check that if there are initially $M$ customers at the 1st stage and none elsewhere, then at $t = 4M$, the state of the system will have doubled, with $2M$ customers at the 1st stage [cf. (7)]. This behavior repeats, with the number of customers approaching infinity as $t \to \infty$.

The remainder of the paper is devoted to demonstrating Theorem 2. [We will actually examine $\xi_T$ in somewhat greater detail and under modest generalizations of (5) and (6).] We proceed by introducing some notation, and then outline the basic argument.

Customers at the $i$th queue, $i = 1, 2$, at time $t$ may be ordered according to the times at which they are next served, so one can talk about a "first" or "last" customer in this sense. Note that customers entering the network earlier on may be behind more recent arrivals according to this ordering. Let $S_1$ denote the time at which the last of the original customers (customers at $t = 0$) at the 1st queue is next served. Let $S_2, S_3, \dots, S_l, \dots$ denote the successive times at which the last customer at the 2nd queue is served, where the ordering is made at $t = S_{l-1}$. Clearly $S_1, \dots, S_l, \dots$ are all stopping times for $\Xi_t$. Set $S_\infty = \lim_{l \to \infty} S_l$ and let $S_l = S_{l+1} = \cdots = S_\infty$ when the 2nd queue becomes empty. [$S_\infty < \infty$ w.p.1 because of (2).] We can think of the intervals $(S_l, S_{l+1}]$, $l = 1, 2, \dots$, as "cycles", for which *each customer starting at* $(2, j)$, $j < J$, *ends up at* $(2, j + 1)$. Note that no customer can be served twice at the 2nd queue before every other customer there is served once. We also let $T$ (as in Theorem 2) denote the time at which the last customer at $(1, 2)$ at time $S_{2J}$ leaves the queue. We will write $\beta_{s,t}(i, j)$ [resp., $\gamma_{s,t}(i, j)$] for the number of customers arriving at (resp., departing from) $(i, j)$ over the time $(s, t]$, and $\beta_t(i, j) = \beta_{0,t}(i, j)$ ($\gamma_t(i, j) = \gamma_{0,t}(i, j)$). Here $\beta_{s,t}(1, 1)$ is the number of customers entering the network, $\gamma_{s,t}(1, 2)$ is the number leaving the network and $\gamma_{s,t}(x) = \beta_{s,t}(y)$ if $y$ is the stage immediately following $x$ along the route. Note that

$$\xi_t(x) = \xi_s(x) + \beta_{s,t}(x) - \gamma_{s,t}(x) \tag{10}$$

for all $x$ and $s \le t$.

We close this section by summarizing the argument for Theorem 2. The following six steps illustrate the main ideas, although the reader should keep in mind that they involve some oversimplifications. In particular, we are ignoring the contributions of the exceptional events over which the individual steps fail to hold. (They are all exponentially small in $M$.)

1. $S_1$ is small relative to $M$; at this time, most customers in the network are at $(2, 1)$.
2. Assuming that $\xi_{S_l}(2, 1) \approx c^{l-1}M$, then $S_{l+1} - S_l \approx c^l M$, and so $\xi_{S_{l+1}}(2, 1) \approx c^l M$. By induction, the first two estimates hold for $l \le J$.
3. The $M$ original customers at $(1, 1)$ arrive at $(1, 2)$ at time approximately $S_{J+1} \approx \sum_{l=1}^{J} c^l M \approx cM/(1 - c)$. Until these customers leave, new customers entering the network cannot be served at $(1, 1)$.
4. The customers remaining at the 2nd queue after time $S_{J+1}$ thus advance quickly toward $(1, 2)$, without other customers arriving at this queue. So the 2nd queue is almost empty by time $S_{2J} \approx cM/(1 - c)$, and few customers remain at $(1, 2)^-$.
5. At time $S_{2J}$, there are on the order of $\sum_{l=0}^{J-1} c^l M \approx M/(1 - c)$ customers at $(1, 2)$. On account of step 4, most have arrived within time const. $c^J M$ of one another.
6. The additional time $T - S_{2J}$ required for these customers to leave $(1, 2)$ is on the order of $cM/(1 - c)$; the total time $T \approx 2cM/(1 - c)$. On the order of $cM/(1 - c)$ new customers arrive at $(1, 1)$ during $(S_{2J}, T]$; they remain at $(1, 1)$ until the above customers leave $(1, 2)$.

Step 1 follows from Lemma 1 and Proposition 1, step 2 follows from Proposition 3 and (37), although the actual reasoning is somewhat different, and step

3 follows from step 2. Step 4 is essentially a consequence of Lemma 5 and its second corollary, step 5 is a consequence of Proposition 4, and step 6 is a consequence of Lemma 6 and Proposition 5.

Step 6 states that at time $T$ there are about $cM/(1-c)$ customers at $(1,1)$ and comparatively few customers elsewhere in the network. This yields (7) of Theorem 2, since $c$ is close to 1. It is not difficult to see that (8) follows from the above scenario as well: Using step 1 and the cycles $(S_l, S_{l+1}]$, one can check that not too many of the original customers at $(1,1)$ will leave the network before $S_{2J}$. By then, the number of customers in the system will be about $M/(1-c)$. Reasoning analogous to that in step 6 (Proposition 6) shows that before most of these customers have left $(1,2)$, at least $M$ customers have entered the network. So throughout $[0,T]$, there are always not much fewer than $M$ customers in the network, and (8) holds.

**3. Behavior at $S_1$; elementary bounds for longer times.** In this section, we analyze the behavior of $\Xi_t$ at $t = S_1$. As we will see, $S_1$ will typically be small, and nearly all of the customers will be at $(2,1)$. Simple upper bounds on $\xi_t(1,2) + \gamma_t(1,2)$, $t \le S_J$, and consequently on $\xi_t(1,1)$, $t \in [S_1, S_J]$, will also follow. Here and later on, we will find it convenient to analyze $\Xi_t$ under a modest generalization of the conditions (5) and (6) of Theorem 2, which will be replaced by

$$(5') \qquad \frac{399}{400} \le c < 1, \qquad c^J \vee \frac{\delta J^2}{1-c} \le \eta \le \frac{1}{50}$$

and

$$(6') \qquad \xi_0(1,1) = M, \qquad \xi_0(1,1)^+ \le \eta M.$$

(The bounds for $\Xi_T$ in Proposition 5 are, for instance, more transparent in the context of general $\eta$.)

We will repeatedly be making use of elementary large deviation estimates for the times $S_{l+1} - S_l$, $l = 1,2,\dots$, and for the numbers of customers who have entered and left different stages of the route over these times. These estimates all reduce to applying the strong Markov property in conjunction with the following basic bounds: Let $X_1, X_2, \dots$ be i.i.d. mean-1 exponential random variables, with $Y_n = X_1 + \cdots + X_n$. Then for each $\alpha > 0$, there exists an $\varepsilon > 0$, so that for all $n \ge 1$,

$$(11) \qquad P\left(\frac{1}{n}|Y_n - n| > \alpha\right) \le e^{-\varepsilon n}.$$

(11) can be demonstrated in the usual way by applying Markov's inequality to the Laplace transform of $Y_n$. Note that (11) immediately extends to exponential distributions with other means. We will be applying (11) throughout the paper, with different choices of $\alpha$ and $\varepsilon$. Rather than do precise bookkeeping with the different values of $\varepsilon$, we will typically label them as $\varepsilon_1, \varepsilon_2, \dots$, without worrying about their exact relationship.

We start with Lemma 1, which provides upper bounds for $S_1$, $\beta_{S_1}(1,1)$, and $\gamma_{S_1}(2,1)$.

LEMMA 1. *Suppose that* (5′) *and* (6′) *hold. For appropriate* $\varepsilon_k > 0$ *and large enough* $M$,

$$P(S_1 > 2\eta M) \le \exp(-\varepsilon_1 M), \tag{12}$$

$$P\big(\beta_{S_1}(1,1) > 3\eta M\big) \le \exp(-\varepsilon_2, M) \tag{13}$$

*and*

$$P\big(\gamma_{S_1}(2,1) > 3\eta M\big) \le \exp(-\varepsilon_3 M). \tag{14}$$

PROOF. Customers at $(1,1)$ are served at rate $1/\delta$, those at $(1,2)$ at rate $1/c$. Applying (6′) together with (11) twice (to mean-$\delta$ and mean-$c$ exponentials), it follows that for any $\alpha > 0$,

$$P(S_1 > (c\eta + \delta + \alpha)M) \le \exp(-\varepsilon_1 M)$$

for appropriate $\varepsilon_1 > 0$ and large enough $M$. Since by (5′), $\delta \le \eta/2$, (12) follows by choosing $\alpha = \eta/2$. If one substitutes $2\eta M$ for $S_1$ in (13), then the conclusion follows from (11) and the assumption that customers enter the system at rate 1. Adding in the exceptional probability in (12) gives (13) itself. One obtains (14) in the same way, since customers leave $(2,1)$ at rate $1/c$. □

Using Lemma 1, we can analyze $\xi_{S_1}(1,1)$, $\xi_{S_1}(2,1)$, and $\xi_{S_1}(2,1)^+ + \gamma_{S_1}(1,2)$.

PROPOSITION 1. *Suppose that* (5′) *and* (6′) *hold. For* $\varepsilon_2$ *and* $\varepsilon_3$ *as above, and large enough* $M$,

$$P\big(\xi_{S_1}(1,1) > 3\eta M\big) \le \exp(-\varepsilon_2 M), \tag{15}$$

$$P\big(\xi_{S_1}(2,1) < (1 - 3\eta)M\big) \le \exp(-\varepsilon_3 M), \tag{16}$$

$$P\big(\xi_{S_1}(1,1) + \xi_{S_1}(2,1) > (1 + 4\eta)M\big) \le \exp(-\varepsilon_2 M) \tag{17}$$

*and*

$$P\Big(\xi_{S_1}(2,1)^+ + \gamma_{S_1}(1,2) > 4\eta M\Big) \le \exp(-\varepsilon_3 M). \tag{18}$$

PROOF. From the definition of $S_1$,

$$\xi_{S_1}(1,1) = \beta_{S_1}(1,1).$$

So, (15) follows from (13). Note that

$$\xi_{S_1}(2,1) \ge \beta_{S_1}(2,1) - \gamma_{S_1}(2,1).$$

Since $\beta_{S_1}(2,1) = M$, (14) implies that the right-hand side is at least $(1 - 3\eta)M$ except on a set of probability at most $\exp(-\varepsilon_3 M)$. So (16) holds. Also,

$$\begin{aligned}\xi_{S_1}(1,1) + \xi_{S_1}(2,1) &\le \xi_0(1,1) + \xi_0(2,1) + \beta_{S_1}(1,1)\\ &\le (1+\eta)M + \beta_{S_1}(1,1),\end{aligned}$$

so (17) follows from (13). Since

$$\xi_{S_1}(2,1)^+ + \gamma_{S_1}(1,2) = \xi_0(2,1)^+ + \gamma_{S_1}(2,1),$$

(18) follows from (14). □

We next give a convenient upper bound on $S_\infty$ and on the number of customers who have visited the system by that time.

LEMMA 2. *Suppose that* (5′) *and* (6′) *hold. For appropriate* $\varepsilon_4 > 0$ *and large enough* $M$,

$$P(S_\infty > (1+8\eta)cM/(1-c)) \le \exp(-\varepsilon_4 M) \tag{19}$$

*and*

$$P\big(\xi_{S_\infty} + \gamma_{S_\infty}(1,2) > (1+9\eta)M/(1-c)\big) \le \exp(-\varepsilon_4 M). \tag{20}$$

PROOF. There are, by (6′), initially at most $(1+\eta)M$ customers in the system. For $t = (1+8\eta)cM/(1-c)$, it follows from (11) that the number of customers entering the system up to time $t$ satisfies

$$P(\beta_t(1,1) > (1+(1-c)\eta)t) \le \exp(-\varepsilon_5 M)$$

for appropriate $\varepsilon_5 > 0$. (The coefficient $1-c$ of $\eta$ is important for our calculations.) Each customer in these two groups needs to be served at the 2nd queue at most $J-1$ times at stages with rate $1/\delta$ and once at the stage $(2,1)$ with rate $1/c$. So by (11), the total amount of time spent serving all these customers at the 2nd queue is at most

$$((1+\eta)M + (1+(1-c)\eta)t)(c + \delta(J-1) + (1-c)\eta) \tag{21}$$

off of a set of exponentially small probability.

By the second condition in (5′),

$$c + \delta(J-1) + (1-c)\eta \le c(1 + 2(1-c)\eta).$$

After plugging this into (21), one can show after some estimation that (21) is strictly less than $t - 2\eta M$. So the amount of time spent serving all the customers already in the system by time $t$ at the 2nd queue is typically strictly less than $t - 2\eta M$. Consequently,

$$P\left(\sum_{j=1}^{J} \xi_s(2,j) > 0 \text{ for all } s \in (2\eta M, t]\right) \le \exp(-\varepsilon_6 M) \tag{22}$$

for appropriate $\varepsilon_6 > 0$; that is, the 2nd queue will typically be empty at some time in $(2\eta M, t]$. By (12), $S_1 \le 2\eta M$ off of an exceptional set. Since $S_\infty$ is the next time after $S_1$ at which the 2nd queue is empty, (19) follows from (22). One obtains (20) by applying (11) once again. □

Recall that over each cycle $(S_l, S_{l+1}]$, a customer starting at $(2,j)$, $j < J$, ends up at $(2, j+1)$. It therefore follows from (18) that for $j = 1, \dots, J$,

$$P\big(\xi_{S_j}(2,j)^+ + \gamma_{S_j}(1,2) > 4\eta M\big) \le \exp(-\varepsilon_3 M). \tag{23}$$

Consequently,

$$P\big(\xi_{S_J}(1,2) + \gamma_{S_J}(1,2) > 4\eta M\big) \le \exp(-\varepsilon_3 M). \tag{24}$$

That is, the total number of customers ever at $(1,2)$ up to time $S_J$ is small. We use (24) to obtain an upper bound on $\xi_t(1,1)$.

PROPOSITION 2. *Suppose that* (5′) *and* (6′) *hold. For appropriate* $\varepsilon_7 > 0$ *and large enough* $M$,

$$P\big(\xi_t(1,1) > 10\eta M \text{ for some } t \in [S_1, S_J]\big) \le \exp(-\varepsilon_7 M). \tag{25}$$

PROOF. Partition $[0,\infty)$ into the two sets $I_1$ and $I_2$, where $s \in I_1$ if either the 1st customer at the 1st queue is at $(1,1)$ or the queue is empty, and $s \in I_2$ if the 1st customer is at $(1,2)$. Each $I_i$ is the union of a countable number of half-open intervals. We can bound the process $\xi_t(1,1)$ by

$$\xi_t(1,1) \le U_t + V_t,$$

where $V_t$ denotes the number of customers entering the network during $I_2 \cap [0,t]$, and $U_t$ is a birth–death process on $I_1$ with birth rate 1 and death rate $1/\delta$, and is constant on $I_2$. (The inequality is due to the departure of customers arriving during $I_2$.)

It follows from (24) and (11) that for appropriate $\varepsilon_8 > 0$ and large $M$,

$$P(|I_2 \cap [0, S_J]| > 5\eta M) \le \exp(-\varepsilon_8 M) \tag{26}$$

(where $|\cdot|$ denotes Lebesgue measure). Applying (11) to (26) implies that

$$P\big(V_{S_J} > 6\eta M\big) \le \exp(-\varepsilon_9 M), \qquad \varepsilon_9 > 0. \tag{27}$$

One can check that $W_t = \exp(U_t) - t$ is a supermartingale, since $\delta$ is small. Set

$$G = \big\{\xi_{S_1}(1,1) \le 3\eta M;\ S_1 \le e^{\eta M}\big\}$$

and

$$\bar{S} = \min\{t \ge S_1 \colon U_t \ge 4\eta M\}.$$

Application of the Optional Sampling Theorem to $W_t$ at $t = \bar{S} \wedge e^{\eta M}$, together with Markov's inequality, shows that

$$\begin{aligned} P\big(U_t \ge 4\eta M \text{ for some } t \in \big[S_1, e^{\eta M}\big]; G\big) &= P(\bar{S} \le e^{\eta M}; G) \\ &\le 2e^{-\eta M}. \end{aligned} \tag{28}$$

Together with (15) and (27), (28) implies that

$$P\big(\xi_t(1,1) > 10\eta M \quad \text{for some } t \in \big[S_1, S_J \wedge e^{\eta M}\big]\big) \le \exp(-\varepsilon_{10} M)$$

for appropriate $\varepsilon_{10} > 0$ and large enough $M$. Since $S_J \le S_\infty$, (25) follows from (19) of Lemma 2. □

**4. Behavior of the system up through $S_{J+1}$.** On account of Proposition 1, we know the behavior of $\Xi_{S_1}$: Most of the customers are located at

(2, 1). From (23) and Proposition 2, we have upper bounds on $\xi_t(i,j)$ at other stages of the route. We now examine the behavior of $\Xi_t$ in detail at the times $S_l$, $l = 2,\dots,J+1$. As we will show, for $l \le J$, the greatest number of customers will be at $(2,l)$, with $(2,l-1),\dots,(2,1)$ possessing geometrically decreasing numbers of customers; at $S_{J+1}$ a corresponding statement holds if one starts the sequence with (1, 2). To show such behavior, we will make use of the following induction hypothesis.

INDUCTION HYPOTHESIS (IH). *For given $l$,*

$$P\big(\xi_{S_l}(1,1) + \xi_{S_l}(2,1) < (c^{l-1} - 3\eta)M\big) \le \exp(-\varepsilon_{l,1}M), \tag{29}$$

$$P\big(\xi_{S_l}(1,1) + \xi_{S_l}(2,1) > (c^{l-1} + 13\eta)M\big) \le \exp(-\varepsilon_{l,2}M) \tag{30}$$

*and*

$$P\big(|\xi_{S_t}(2,j) - c^{l-j}M| > 13\eta M\big) \le \exp(-\varepsilon_{l,3}M), \qquad j = 2,\dots,l\wedge J, \tag{31}$$

*all hold for appropriate $\varepsilon_{l,1}, \varepsilon_{l,2}, \varepsilon_{l,3} > 0$ and large enough $M$.*

Our main result in this section is the following proposition. We will be primarily interested in the conclusions for $l = J$ and $l = J+1$.

PROPOSITION 3. *Suppose that (5′) and (6′) hold. Then (29)–(31) are valid for all $l = 1,\dots,J+1$.*

We first observe that for $l = 1$, (29) and (30) follow from (16) and (17); (31) is, in this case, vacuous. It also follows immediately from Proposition 2 that given (29),

$$P\big(\xi_{S_l}(2,1) < (c^{l-1} - 13\eta)M\big) \le \exp(-\varepsilon_{l,4}M), \qquad \varepsilon_{l,4} > 0 \tag{32}$$

holds for $l = 1,\dots,J$. [(32) is not true for $l = J+1$.] It is easy to check that (31), for $l = 2,\dots,J+1$, follows from (30) and (32) if $j = 2$, and from itself at the previous time (at $j-1$) for $j > 2$. To demonstrate Proposition 3, it therefore suffices to show that (29) and (30) hold for $l+1$, given that they hold for $l$. We note that by checking the evolution of $\Xi_t$ over the cycles $(S_l, S_{l+1}]$, $l = 1,\dots,J$, one obtains from (16)–(18) the extension of (31),

$$P\big(|\xi_{S_{J+1}}(1,2) + \gamma_{S_{J+1}}(1,2) - M| > 8\eta M\big) \le \exp(-\varepsilon_{11}M). \tag{33}$$

[One can also apply (31) itself at $j = J$ and $t = S_J$, with 13 replacing 8.]

To examine IH, we first obtain bounds on the number of customers entering the system over $(S_l, S_{l+1}]$.

LEMMA 3. *Assume (5′) and that IH holds for given $l, 1 \le l \le J$. Then*

$$P\left(\frac{1}{M}|\beta_{S_l,S_{l+1}}(1,1) - c\xi_{S_l}(2,1)| > (1-c)\eta\right) \le \exp(-\varepsilon_{l,5}M) \tag{34}$$

*for appropriate $\varepsilon_{l,5} > 0$ and large $M$.*

PROOF. On account of (30) and (32), $|\xi_{S_l}(2,1) - c^{l-1}M| > 13\eta M$ holds only with exponentially small probability. One can employ (11) to show that off this and another exceptional set, the time it takes to serve these customers is within $(1-c)\eta M/4$ of $c\xi_{S_l}(2,1)$. By (31) and (18), the number of customers at other stages of the 2nd queue is

$$\leq \left( \sum_{j=1}^{l} c^{l-j} + 13\eta(J-1) \right) M \leq 2JM,$$

again, off an exceptional set. By (5′) and (11), the time required to serve all of these customers is usually

$$\leq 2\delta JM + (1-c)\eta M/4 \leq (1-c)\eta M/2.$$

So, off of the exceptional sets,

$$|S_{l+1} - S_l - c\xi_{S_l}(2,1)| \leq 3(1-c)\eta M/4. \tag{35}$$

Now (34) follows from a final application of (11). □

We note that direct estimation of $S_{l+1} - S_l$ in terms of $c^l M$ instead of as in (35) would not allow us to retain the factor $1-c$ in (34). This factor is used in the following proof.

PROOF OF PROPOSITION 3. We need to verify (29) and (30) for $l+1$, given that they hold for $l$. Abbreviating $\xi_t(\cdot,1)$ for $\xi_t(1,1) + \xi_t(2,1)$, we first note that

$$\xi_{S_{l+1}}(\cdot,1) = \xi_{S_l}(1,1) + \beta_{S_l, S_{l+1}}(1,1), \tag{36}$$

since those customers at $(2,1)$ at $S_l$ have already been served by time $S_{l+1}$. Off of the exceptional set given in Lemma 3, this is

$$\begin{aligned}
&\geq \xi_{S_l}(1,1) + c\xi_{S_l}(2,1) - (1-c)\eta M \\
&= c\xi_{S_l}(\cdot,1) + (1-c)\big(\xi_{S_l}(1,1) - \eta M\big) \\
&\geq c\xi_{S_l}(\cdot,1) - (1-c)\eta M.
\end{aligned}$$

Off of the exceptional set in (29), this is at least $(c^l - 3\eta)M$. So

$$P\big(\xi_{S_{l+1}}(\cdot,1) < (c^l - 3\eta)M\big) \leq \exp(-\varepsilon_{l,1}M) + \exp(-\varepsilon_{l,5}M).$$

This implies (29) for $l+1$.

To obtain (30) for $l+1$, we again apply Lemma 3 to (36) to get, off of the exceptional set,

$$\begin{aligned}
\xi_{S_{l+1}}(\cdot,1) &\leq \xi_{S_l}(1,1) + c\xi_{S_l}(2,1) + (1-c)\eta M \\
&= c\xi_{S_l}(\cdot,1) + (1-c)\big(\xi_{S_l}(1,1) + \eta M\big).
\end{aligned}$$

Off of the exceptional sets in (30) and Proposition 2, this is

$$\leq c\xi_{S_l}(\cdot,1) + 11(1-c)\eta M \leq (c^l + 13\eta)M.$$

So,

$$P\Big(\xi_{S_{l+1}}(\cdot,1) > (c^l + 13\eta)M\Big) \le \exp(-\varepsilon_7 M) + \exp(-\varepsilon_{l,2} M) + \exp(-\varepsilon_{l,5} M).$$

This implies (30) for $l+1$. □

We also wish to keep track of $S_l$. Along the lines of Lemma 3, one can demonstrate for $l = 1,\dots,J$, that

$$P\left(\frac{1}{M}|S_{l+1} - S_l - c^l M| > 14\eta\right) \le \exp(-\varepsilon_{l,6} M) \tag{37}$$

for appropriate $\varepsilon_{l,6} > 0$ and large $M$. In particular, setting $l = J$, by (5′),

$$P(S_{J+1} - S_J > 15\eta M) \le \exp(-\varepsilon_{J,6} M). \tag{38}$$

**5. Behavior of the system over $(S_J, S_{2J}]$.** The nature of $\Xi_t$ changes over $(S_J, S_{J+1}]$. By (24), few customers have visited $(1,2)$ by time $S_J$; by (33), this has changed completely by time $S_{J+1}$. Lemma 4 gives us a lower bound on the number of customers still at $(1,2)$ at this time.

Lemma 4. *Assume that* (5′) *and* (6′) *hold. For appropriate* $\varepsilon_{12} > 0$ *and large* $M$,

$$P\Big(\xi_{S_{J+1}}(1,2) < 22\eta M\Big) \le \exp(-\varepsilon_{12} M). \tag{39}$$

Proof. On account of (24),

$$P\Big(\gamma_{S_J}(1,2) > 4\eta M\Big) \le \exp(-\varepsilon_3 M).$$

By (38) and (11),

$$P\Big(\gamma_{S_J,S_{J+1}}(1,2) > 16\eta M\Big) \le \exp(-\varepsilon_{13} M), \qquad \varepsilon_{13} > 0.$$

These two bounds show that

$$P\Big(\gamma_{S_{J+1}}(1,2) > 20\eta M\Big) \le \exp(-\varepsilon_3 M) + \exp(-\varepsilon_{13} M).$$

Since by (6′), $\eta \le 1/50$, (39) follows from (33). □

Let $\sigma$ denote the further time after $S_{J+1}$ until the last customer at $(1,2)$ at time $S_{J+1}$ is served. The following bound is a direct consequence of (39) and (11).

Corollary 1. *Assume that* (5′) *and* (6′) *hold. For appropriate* $\varepsilon_{14} > 0$ *and large* $M$,

$$P(\sigma < 21\eta M) \le \exp(-\varepsilon_{14} M). \tag{40}$$

No customer entering the system after time $S_{J+1}$ will be served at the 1st queue until the customers already at $(1,2)$ have been served. On the other hand, customers already waiting at the 2nd queue continue to be served. Lemma 5 gives an upper bound on the further time $S_{2J} - S_{J+1}$ it takes for the customers at $(2,j)$, $j = 2,\dots,J$, to enter $(1,2)$.

LEMMA 5. *Assume that* (5′) *and* (6′) *hold. For appropriate* $\varepsilon_{15} > 0$ *and large* $M$,

$$P(S_{2J} - S_{J+1} > 16\eta M) \le \exp(-\varepsilon_{15} M). \tag{41}$$

PROOF. By (31) with $l = J + 1$, the number of customers at all of the sites $(2, j)$, $j = 2, \dots, J$, at time $S_{J+1}$ is, except for exponentially small probability,

$$\le \left( \sum_{j=2}^{J} c^{J+1-j} + 13\eta(J-1) \right) M. \tag{42}$$

By (30) and (5′), the number of customers at $(1,1)$ and $(2,1)$ is, off an exceptional set,

$$\le (c^J + 13\eta) M \le 14\eta M. \tag{43}$$

Note that (42) and (43) sum to at most $JM/(1-c)$.

Each customer starting at some $(2, j)$, $j \ge 2$, needs to be served at most $J - 1$ times before reaching $(1,2)$, each of which occurs at rate $1/\delta$. Those customers starting at $(1,1)$ and $(2,1)$ in addition need to be served at $(2,1)$, at rate $1/c$. Let $\tau$ denote the amount of time spent serving all of these customers at the 2nd queue. We may apply (11) to the total service times at $(2,1)$ and at $(2, j)$, $j \ge 2$. Since by (5′),

$$14\eta + \delta J^2/(1-c) \le 15\eta,$$

it follows from (42) and (43) that

$$P(\tau > 16\eta M) \le \exp(-\varepsilon_{16} M), \qquad \varepsilon_{16} > 0. \tag{44}$$

One obtains from Corollary 1 and (44) that

$$P(\sigma \le \tau) \le \exp(-\varepsilon_{14} M) + \exp(-\varepsilon_{16} M).$$

Since no customers entering the system after $S_{J+1}$ arrive at $(2,1)$ until time $\sigma$, this means that off an exceptional set, all customers at $(1,1)^+$ at $S_{J+1}$ arrive at $(1,2)$ by time $S_{J+1} + \tau$. [This need not be true for customers at $(1,1)$, since their service at $(1,1)$ might be delayed by customers at $(1,2)$, during which the 2nd queue may be empty.] The inequality (41) therefore follows from (44). □

The following bound on the number of customers exiting from the system by time $S_{2J}$ is a consequence of Lemma 5.

COROLLARY 2. *Assume that* (5′) *and* (6′) *hold. For appropriate* $\varepsilon_{17} > 0$ *and large* $M$,

$$P\big(\gamma_{S_{2J}}(1,2) > 36\eta M\big) \le \exp(-\varepsilon_{17} M). \tag{45}$$

PROOF. Together with (38), (41) implies that

$$P(S_{2J} - S_J > 31\eta M) \le \exp(-\varepsilon_{J,6} M) + \exp(-\varepsilon_{15} M).$$

So by (11),

$$P\big(\gamma_{S_J, S_{2J}}(1,2) > 32\eta M\big) \le \exp(-\varepsilon_{18} M), \qquad \varepsilon_{18} > 0.$$

Then (45) follows from (24). □

Lemma 5 also provides the following bound for the number of customers at $(1,2)^-$ at time $S_{2J}$.

COROLLARY 3. *Assume that* (5′) *and* (6′) *hold. For appropriate* $\varepsilon_{19} > 0$ *and large* $M$,

$$P\Big(\xi_{S_{2J}}(1,2)^- > 31\eta M\Big) \le \exp(-\varepsilon_{19} M). \tag{46}$$

PROOF. Contributions to $\xi_{S_{2J}}(1,2)^-$ consist of customers either (a) at $(2,j)$, $j = 2,\dots,J$, at time $S_{J+1}$, (b) at $(1,1)$ or $(2,1)$ at that time, or (c) arriving at $(1,1)$ later. Those customers originally at $(2,j)$ will automatically have arrived at $(1,2)$ by time $S_{2J}$. On account of (41) and (11),

$$P(\beta_{J+1,2J}(1,1) > 17\eta M) \le \exp(-\varepsilon_{20} M), \qquad \varepsilon_{20} > 0,$$

so not many customers enter the system over $(S_{J+1}, S_{2J}]$. Together with (30), for $l = J+1$, which bounds the number of customers at $(1,1)$ and $(2,1)$, this implies that

$$P\Big(\xi_{S_{2J}}(1,2)^- > 31\eta M\Big) \le \exp(-\varepsilon_{J+1,2} M) + \exp(-\varepsilon_{20} M)$$

(recall that $c^J \le \eta$). Inequality (46) follows. □

We see from (46) that at time $S_{2J}$, not many customers are left at $(1,2)^-$. On the other hand, by (45), most of the customers ever in the system are still at $(1,2)$ at time $S_{2J}$, not having had enough time to be served there. More detailed information on $S_{2J}$ and $\xi_{S_{2J}}(1,2)$ is given in Proposition 4.

PROPOSITION 4. *Assume that* (5′) *and* (6′) *hold. For appropriate* $\varepsilon_{21} > 0$ *and large* $M$,

$$P\left(\frac{c\rho_1 M}{1-c} \le S_{2J} \le \frac{c\theta_1 M}{1-c}\right) \ge 1 - \exp(-\varepsilon_{21} M) \tag{47}$$

*and*

$$P\left(\frac{\rho_1 M}{1-c} \le \xi_{S_{2J}}(1,2) \le \frac{\theta_1 M}{1-c}\right) \ge 1 - \exp(-\varepsilon_{21} M), \tag{48}$$

*where* $\rho_1 = \rho_1(\eta) > 1/3$, $\theta_1 = \theta_1(\eta)$ *and* $\rho_1, \theta_1 \to 1$ *as* $\eta \to 0$.

PROOF. Since $S_{2J} \le S_\infty$, it follows from (19) that

$$P(S_{2J} > (1 + 8\eta)cM/(1-c)) \le \exp(-\varepsilon_4 M).$$

This gives the upper bound in (47). For the lower bound, we use (37), with $l = 1,\dots,J$, to get

$$P\left(S_{2J} < \sum_{j=1}^{J} (c^j - 14\eta)^+ M\right) \le \exp(-\varepsilon_{22} M), \qquad \varepsilon_{22} > 0. \tag{49}$$

We can bound the preceding sum as follows:

$$\sum_{j=1}^{J} (c^j - 14\eta)^+ \ge \sum_{j=1}^{j_0} c^j - 14 j_0 \eta, \tag{50}$$

where $c^{j_0 - 1} \ge b$, $c^{j_0} < b$, and $b$ satisfies

$$\eta = bc/[14\log(1/b)]. \tag{51}$$

One can check that, since $c^J \le \eta$ [by (5′)], $j_0 \le J$ holds. Since $\eta \le 1/50$, $b \le 1/3$ is always satisfied. With a little estimation, (51) implies that $14 j_0 \eta \le bc/(1-c)$. On the other hand,

$$\sum_{j=1}^{j_0} c^j \ge (1-b)c/(1-c), \tag{52}$$

so the right side of (50) is at least $(1-2b)c/(1-c)$. For $b = 1/3$, this is $c/[3(1-c)]$; as $\eta \to 0$, $b \to 0$. Plugging this bound into (49), we see that the lower bound in (47) holds.

Since $\xi_{S_{2J}}(1,2) \le \xi_{S_\infty} + \gamma_{S_\infty}(1,2)$, it follows from (20) that

$$P\big(\xi_{S_{2J}}(1,2) > (1+9\eta)M/(1-c)\big) \le \exp(-\varepsilon_4 M).$$

This gives the upper bound in (48). For the lower bound, note that by time $S_{2J}$ all the customers at the 2nd queue at time $S_J$ have entered (1, 2). So by (29) and (31), with $l = J$,

$$P\left(\beta_{S_{2J}}(1,2) < \sum_{j=1}^{J} (c^{J-j} - 13\eta)^+ M\right) \le \exp(-\varepsilon_{23} M), \qquad \varepsilon_{23} > 0. \tag{53}$$

Together with (45), this shows that

$$P\left(\xi_{S_{2J}}(1,2) < \sum_{j=1}^{J} \left((c^{J-j} - 13\eta)^+ - 36\eta\right) M\right) \le \exp(-\varepsilon_{24} M), \qquad \varepsilon_{24} > 0. \tag{54}$$

The estimates for the sum in (54) are the same as in (49) after absorbing the term $36\eta$, and they give the same bounds except for the factor of $c$. So the lower bound of (48) also holds. [One is, of course, free to use the same coefficients $\rho_1$, $\theta_1$ and $\varepsilon_{21}$ in (47) and (48). $\square$

**6. Behavior over $(S_{2J}, T]$; conclusion.** Let $T$ denote the time at which the last of the customers at (1, 2) at time $S_{2J}$ leaves the system. On account of Proposition 4 and Corollary 3, we know that nearly all of the customers at time $S_{2J}$ are at $(1,2)$. This enables us to analyze $\Xi_T$ and $T$ without much trouble. We first estimate $T - S_{2J}$.

LEMMA 6. *Assume that* (5′) *and* (6′) *hold. For appropriate* $\varepsilon_{25} > 0$ *and large* $M$,

$$P\left(\frac{c\rho_2 M}{1-c} \le T - S_{2J} \le \frac{c\theta_2 M}{1-c}\right) \ge 1 - \exp(-\varepsilon_{25} M), \tag{55}$$

*where* $\rho_2 = \rho_2(\eta) > 1/3$, $\theta_2 = \theta_2(\eta)$ *and* $\rho_2, \theta_2 \to 1$ *as* $\eta \to 0$.

PROOF. The lower bound follows from (48) and (11). For the upper bound, note that $T$ will occur by the time all of the customers in the 1st queue at time $S_{2J}$ have been served. By Corollary 3, $\xi_{S_{2J}}(1,1) \le 31\eta M$ off of an exceptional set, and by (48), the same is true for $\xi_{S_{2J}}(1,2) \le \theta_1 M/(1-c)$. The bound therefore follows by applying (11) separately to the total service times for the customers at $(1,1)$ and $(1,2)$. □

Proposition 5 gives us the desired behavior of $\Xi_T$.

PROPOSITION 5. *Assume that* (5′) *and* (6′) *hold. For appropriate* $\varepsilon_{26} > 0$ *and large* $M$,

$$P\left(\frac{2c\rho_3 M}{1-c} \le T \le \frac{2c\theta_3 M}{1-c}\right) \ge 1 - \exp(-\varepsilon_{26} M) \tag{56}$$

*and*

$$P\left(\frac{c\rho_3 M}{1-c} \le \xi_T(1,1) \le \frac{c\theta_3 M}{1-c}\right) \ge 1 - \exp(-\varepsilon_{26} M), \tag{57}$$

*where* $\rho_3 = \rho_3(\eta) > 1/3$, $\theta_3 = \theta_3(\eta)$ *and* $\rho_3$, $\theta_3 \to 1$ *as* $\eta \to 0$. *Moreover*,

$$P\left(\xi_T(1,1)^+ \le M\right) \ge 1 - \exp(-\varepsilon_{26} M). \tag{58}$$

PROOF. Inequality (56) follows immediately from (47) and (55). Application of (11) and (55) implies that

$$P\left(\frac{c\rho_4 M}{1-c} \le \beta_{S_{2J},T}(1,1) \le \frac{c\theta_4 M}{1-c}\right) \ge 1 - \exp(-\varepsilon_{27} M), \qquad \varepsilon_{27} > 0, \tag{59}$$

where $\rho_4$ and $\theta_4$ have the same properties as $\rho_2$ and $\theta_2$. None of the customers entering the system after time $S_{2J}$ can be served by time $T$, since $(1,2)$ is still occupied by earlier customers. Therefore,

$$\xi_T(1,1) \ge \beta_{S_{2J},T}(1,1),$$

from which the lower bound in (57) follows. On the other hand,

$$\xi_T(1,1) \le \xi_{S_{2J}}(1,1) + \beta_{S_{2J},T}(1,1),$$

together with (46) and (59), implies the upper bound in (57).

No customers entering the system after time $S_{2J}$ can be served at (1, 1) before $T$. Also, all customers at (1, 2) at time $S_{2J}$ will be gone from the system by time $T$. So

$$\xi_T(1,1)^+ \le \xi_{S_{2J}}(1,2)^-,$$

and (58) follows from (46), since $\eta < 1/50$. □

We need an elementary lower bound on $\xi_t$, $t \in [0,T]$, for (8) of Theorem 2.

LEMMA 7. *Assume that* (5′) *and* (6′) *hold. For appropriate* $\varepsilon_{28} > 0$ *and large* $M$,

$$(60) \qquad P(\xi_t \ge (1-36\eta)M \quad \forall\, t \in [0,T]) \ge 1 - \exp(-\varepsilon_{28}M).$$

PROOF. Corollary 2 states that the number of customers leaving the system by time $S_{2J}$ is comparatively small:

$$P\big(\gamma_{S_{2J}}(1,2) > 36\eta M\big) \le \exp(-\varepsilon_{17}M), \qquad \varepsilon_{17} > 0.$$

So, on account of (6′),

$$(61) \qquad P(\xi_t < (1-36\eta)M \text{ for some } t \in [0, S_{2J}]) \le \exp(-\varepsilon_{17}M).$$

[This bound can be improved to $(1-4\eta)M$ by considering $\xi_t$ separately over $[0, S_J]$ and $(S_J, S_{2J}]$, rather than over $[0, S_{2J}]$.] To examine the behavior of $\xi_t$ over $(S_{2J}, T]$, we introduce the time $T' = S_{2J} + 2M$. It follows from (11) that

$$(62) \qquad P\big(\beta_{S_{2J},T'}(1,1) < M\big) \le \exp(-\varepsilon_{29}M), \qquad \varepsilon_{29} > 0$$

and

$$(63) \qquad P\big(\gamma_{S_{2J},T'}(1,2) > 4M\big) \le \exp(-\varepsilon_{30}M), \qquad \varepsilon_{30} > 0.$$

Together with (48) of Proposition 4, (63) shows that

$$(64) \quad P(\xi_t < M \text{ for some } t \in (S_{2J}, T']) \le \exp(-\varepsilon_{21}M) + \exp(-\varepsilon_{30}M).$$

Since customers entering the system after $S_{2J}$ remain at (1, 1) until time $T$, (62) implies that

$$(65) \qquad P(\xi_t < M \text{ for some } t \in (T' \wedge T, T]) \le \exp(-\varepsilon_{29}M).$$

Hence (60) follows from (61), (64) and (65). □

Proposition 5 provides us with a detailed description of the behavior of $\Xi_T$. Since $c \ge 399/400$, the lower bound in (57) states that $\xi_T(1,1) \ge 100M$. Together with (58), this implies (7) of Theorem 2. Display (8) of Theorem 2 is an immediate consequence of Lemma 7, with $\eta = 1/50$. So Theorem 2, and, consequently, Theorem 1 hold. The process $\Xi_t$ is therefore unstable, with $\xi_t \to \infty$ as $t \to \infty$. More detail on the asymptotics of $\Xi_t$ under (6′) is supplied by Propositions 1–5.

**Acknowledgment.** The author thanks Jim Dai for various helpful conversations.

## REFERENCES

[1] BRAMSON, M. (1993). Instability of FIFO queueing networks with quick service times. Unpublished manuscript.

[2] KELLY, F. P. (1975). Networks of queues with customers of different types. *J. Appl. Probab.* **12** 542–554.

[3] KELLY, F. P. (1979). *Reversibility and Stochastic Networks.* Wiley, New York.

[4] KUMAR, P. R. (1993). Re-entrant lines. *Queueing Systems Theory Appl.* **13** 87–110.

[5] LU, S. H. and KUMAR, P. R. (1991). Distributed scheduling based on due dates and buffer priorities. *IEEE Trans. Automat. Control* **36** 1406–1416.

[6] RYBKO, S. and STOLYAR, A. L. (1992). Ergodicity of stochastic processes that describe the functioning of open queueing networks. *Problems Inform. Trans.* **28** 3–26 (in Russian).

[7] SEIDMAN, T. I. (1993). 'First come, first served' is unstable! Unpublished manuscript.

[8] SEIDMAN, T. I. and YERSHOV, A. (1993). A single product instability example for FCFS. Unpublished manuscript.

[9] WHITT, W. (1993). Large fluctuations in a deterministic multiclass network of queues. *Management Sci.* **39** 1020–1028.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WISCONSIN, MADISON
VAN VLECK HALL
480 LINCOLN DRIVE
MADISON, WISCONSIN 53706