The Last Buffer First Serve Priority Policy is Stable for Stochastic Re–entrant Lines*

Sunil Kumar and P.R. Kumar†

Abstract

In this paper, the stability of the Last Buffer First Serve (LBFS) buffer priority policy, in the sense of positive Harris recurrence, is shown under fairly general assumptions on the interarrival and service time distributions, whenever the nominal load at each machine is less than one. This is done by first describing the fluid limits, and then showing that the fluid limits empty in finite time from all suitable initial conditions, which Dai [1] has recently shown to be sufficient for positive Harris recurrence.

1 Introduction

The Last Buffer First Serve (LBFS) buffer priority policy is a well known scheduling policy for re–entrant lines. In this policy, the priority order of buffers contending for service at a machine is the reverse of the order in which they are visited by a part. That is, a buffer receives priority over all contending buffers which are upstream of it. The stability of this policy was first studied by Lu and Kumar [2], for the case when the interarrival and service times are deterministic and bursty. They showed that the condition that the nominal load at each machine be strictly less than one, which is obviously necessary for stability, is also sufficient for stability.

Recently, Dai [1] has proposed a method of establishing positive Harris recurrence, which is considered “stability” in this paper, for queueing networks using the fluid limit models. He

---

*Please address all correspondence to the first author.
†The research reported here has been supported in part by the National Science Foundation Grant Nos. ECS–90–25007 and ECS–92–16487, and the U. S. Army Research Office under Contract No. DAAL–03–88–K–0046.
‡Department of Electrical and Computer Engineering, and the Coordinated Science Laboratory, University of Illinois, 1308 West Main Street, Urbana, IL 61801.
has shown that if the fluid limits empty in finite time from every suitable initial condition, and stay empty, then the underlying stochastic queueing network is stable. He has also illustrated the use of this technique by proving the stability of the First Buffer First Serve (FBFS) policy. This paper analyzes the fluid limits by identifying a sequence of bottlenecks, and then invokes Dai’s results, to establish the stability of the LBFS policy. It also identifies techniques for analyzing the fluid limits through the integral equations identified in Dai [1], which may be useful in further work in the area.

The network we consider consists of $S$ machines \{1, 2, \ldots, S\}. Parts arrive at buffer $b_1$, located at machine $\sigma(1) \in \{1, \ldots, S\}$, with the interarrival times being an i.i.d. sequence $\{\xi(n)\}$. Upon completing service, they proceed to buffer $b_2$ located at machine $\sigma(2) \in \{1, \ldots, S\}$. Let $b_L$ at machine $\sigma(L)$ be the last buffer visited. The sequence $\{\sigma(1), \ldots, \sigma(L)\}$ is the route of the part. Since one may have $\sigma(i) = \sigma(j)$ for some pairs $i$ and $j$ with $i \neq j$, we say that the system is a re-entrant line. Let us suppose that parts in $b_i$ require i.i.d. service times $\{\eta_i(n)\}$, with mean $1/\mu_i = m_i < \infty$, from machine $\sigma(i)$. Let us also suppose that $E[\xi(1)] = \lambda = 1/\alpha < \infty$, and that $\{\xi(n)\}$ and $\{\eta_i(n)\}$ for $i = 1, 2, \ldots, L$ are mutually independent. We assume that the interarrival times are unbounded. That is for any $x > 0$, $\operatorname{Prob}(\xi(1) > x) > 0$. Finally, we assume that\footnote{We are grateful to Sean Meyn for pointing out the need for this additional assumption.} for some integer $n$ and some function $p(x) \geq 0$ on $IR_+$ with $\int_{0}^{\infty} p(x)dx > 0$,

$$\operatorname{Prob} \left[ a \leq \sum_{i=1}^{n} \xi(i) \leq b \right] \geq \int_{a}^{b} p(x)dx, \text{ for any } 0 \leq a \leq b.$$ 

For each machine $\sigma \in \{1, 2, \ldots, S\}$, define

$$\rho_{\sigma} = \lambda \left( \sum_{1 \leq k \leq L, \sigma(k) = \sigma} m_k \right).$$

We call $\rho_{\sigma}$ the nominal load for machine $\sigma$.

If buffers $b_j$ and $b_k$ share the same machine, i.e., $\sigma(j) = \sigma(k)$, and $j < k$, then priority is given to $b_k$. The priority discipline is assumed to be preemptive resume. Therefore if $k < L$, $b_k$ can be worked on by $\sigma(k)$ only if $b_j = 0$ for all $j = k + 1, \ldots, L$, with $\sigma(j) = \sigma(k)$. Note
that $b_L$ is never preempted. We also assume that the service discipline is non-idling, that is, a machine cannot be idle when any of the buffers located at that machine is non-empty. This is the LBFS discipline.

The main result of this paper is the following theorem.

**Theorem** Under the assumptions stated above, the re-entrant line operating under the LBFS policy is positive Harris recurrent if

$$\rho_\sigma < 1, \text{ for all } \sigma \in \{1, 2, \ldots, S\}. \quad (1)$$

The rest of the paper is devoted to proving this theorem.

## 2 The Proof of the Theorem

The main steps of the proof are as follows:

- Describe the fluid limits of the stochastic system.

- Show that the fluid limits empty in finite time from every possible initial condition, and stay empty. This is done as follows:

  - Show that the last buffer $b_L$ empties at some finite time $t_L$.

  - At $t_L$, identify an upstream bottleneck which caused $b_L$ to empty, and show that all the buffers downstream of that bottleneck (including $b_L$) remain empty for all $t > t_L$.

  - Iterate the argument, at each time identifying a new upstream bottleneck and a new finite time, and showing that all buffers downstream of it remain empty after that time.

  - Eventually, the iterations terminate, showing that the fluid limit is completely empty at some finite time and remains so thereafter.
First we introduce the notation and some preliminaries from Dai [1]. Let $Q_k(t)$ denote the queue length, and $v_k(t)$ the residual service time at buffer $b_k$ at time $t$. Let $u(t)$ denote the residual interarrival time at time $t$. Let $E^u(t) := \max \{ n : u + \xi(1) + \cdots + \xi(n-1) \leq t \}$. Let $D_k(t)$ denote the number of departures from buffer $b_k$ up to time $t$. Then the total number of arrivals to buffer $b_k$ is given by:

$$A_1(t) = Q_1(0) + E^u(t),$$

$$A_k(t) = Q_k(0) + D_{k-1}(t), \quad k = 2, \ldots, L.$$  

Immediately, we have

$$Q_k(t) = A_k(t) - D_k(t).$$

Let $T_k(t)$ be the amount of time in $[0, t]$ that $\sigma(k)$ spends on $b_k$, and let $B_k(t) = \sum_{\{i : 1 \leq j \leq k, \text{ and } \sigma(i) = \sigma(k)\}} T_j(t)$ be the amount of time spent on buffers $b_k$ and those with higher priority by machine $\sigma(k)$. Let $I_k(t) = t - B_k(t)$ denote the time not spent on $\{b_j : k \leq j \leq L, \text{ and } \sigma(j) = \sigma(k)\}$ by $\sigma(k)$. Because of the non-idling assumption and the LBFS priority, we have

$$\int_0^\infty \left[ \left( \sum_{\{i : 1 \leq j \leq L, \text{ and } \sigma(i) = \sigma(k)\}} Q_i(t) \right)^1 \right] dI_k(t) = 0.$$

The state of the system is given by $X(t) := (Q_1(t), \ldots, Q_L(t), u(t), v_1(t), \ldots, v_L(t))$. For $x = (q_1, \ldots, q_L, u, v_1, \ldots, v_L)$ we define $|x| := \sum_{k=1}^L (q_km_k + u + v_k)$. In the sequel, we shall denote explicit dependence on an initial condition $x$ by adding a superscript $x$ to the variable of interest. For any function $f$, let

$$\tilde{f}^x(t) := \frac{1}{|x|} f^x(|x|t), \text{ for } t \geq 0 \text{ and } x > 0,$$

denote its scaled version.

For simplicity, let us make the additional assumption that all the service and interarrival time distributions are exponential. This allows us to completely ignore the effects of $u$ and $v_k$. The state $x(t)$ then consists only of the queue lengths $(Q_1(t), \ldots, Q_L(t))$. This will be relaxed later.

Next we state the result from Dai [1].
Lemma 1 For every sequence \( \{x_n\} \) with \( |x_n| \to \infty \), there is a subsequence \( \{x_{n_l}\} \) such that as \( l \to \infty \),
\[
(E^{x_{n_l}}, D^{x_{n_l}}, \bar{A}^{x_{n_l}}, \bar{Q}^{x_{n_l}}, \bar{T}^{x_{n_l}}, \bar{I}^{x_{n_l}}) \Rightarrow (\bar{E}, \bar{D}, \bar{A}, \bar{Q}, \bar{T}, \bar{I}),
\]
where "\( \Rightarrow \)" denotes uniform convergence on compact sets. Furthermore, the limit processes, called the fluid limits, satisfy:

\[
\begin{align*}
E(t) &= \lambda t, \\
D_k(t) &= \mu_k \bar{T}(t), \\
\bar{A}_1(t) &= q_1 + \bar{E}(t), \\
\bar{A}_k(t) &= q_k + \bar{D}_{k-1}(t), \quad k \geq 2, \\
\bar{Q}_k(t) &= \bar{A}_k(t) - \bar{D}_k(t), \\
\bar{I}_k(t) &= t - \sum_{\{j: k \leq j \leq L \text{ and } \sigma(j) = \sigma(k)\}} \bar{T}_j(t),
\end{align*}
\]

where \( 0 \leq \bar{T}_k(t) \leq \bar{T}_j(t) \leq t \) for all \( j > k \) with \( \sigma(j) = \sigma(k) \),
\[
0 \leq \bar{I}_k(t_1) \leq \bar{I}_k(t_2) \quad \text{and} \quad 0 \leq \bar{I}_k(t_1) \leq \bar{I}_k(t_2) \quad \text{if} \ t_1 \leq t_2,
\]
\[
\int_0^\infty \left( \sum_{\{j: k \leq j \leq L \text{ and } \sigma(j) = \sigma(k)\}} \bar{Q}_j(t) \wedge 1 \right) d\bar{I}_k(t) = 0,
\]
\[
|\bar{I}_k(t) - \bar{I}_k(s)| \leq |t - s|, \quad |\bar{T}_k(t) - \bar{T}_k(s)| \leq |t - s|, \quad \text{and} \quad \bar{Q}_k(t) \text{ is also Lipschitz},
\]
\[
\text{and} \quad \sum_{k=1}^{L} m_k q_k \leq 1.
\]

Proof. Recall that all distributions are exponential, and \( u, \{v_k\} \) are identically zero under the simplifying assumption. Equation (3) follows from Lemma 4.2 of Dai [1], and (4) follows from the proof of Corollary 4.2. Equations (5–7) are the queue length variants of equations (4.18–4.19) of Theorem 4.1 of [1]. Equations (8–11) are obtained trivially from the fact that these hold for each sample path of the stochastic process, because of the preemptive LBFS policy. Equation (12) follows from (8,9,10). Equation (13) is a consequence of the scaling by \( 1/|x| \).

Having described a set of necessary conditions (3–13) for the fluid limits, we now show through a sequence of lemmas that these limits empty in finite time, and stay empty, from
every initial condition $q$ satisfying (13).

Lemma 2  

(i) $\bar{Q}_{L}(t_{L}) = 0$ for some finite $t_{L} > 0$ which depends only on $\lambda$ and $\{\mu_{k}\}$.

(ii) Let $b_{k}$ be the buffer satisfying

$$k := \max\{i < L \mid \sum_{\{j : i \leq j \leq L \text{ and } \sigma(j) = \sigma(i)\}} m_{j} \geq m_{L}\}. \tag{14}$$

If no such $i$ exists, take $k := 0$. Then, the buffers $b_{k+1}, \ldots, b_{L}$ are empty at time $t_{L}$.

Proof. (i) Consider $\sum_{k=1}^{L} \bar{Q}_{k}(t)$. Suppose $\bar{Q}_{L}(s) > 0$ for all $s \in (0, t]$. Then $\bar{I}_{L}(t) = 0$ because of (11), and the fact that $\bar{I}_{L}(0) = 0$. So by (8), $T_{L}(t) = t$. This means that

$$\sum_{i=1}^{L} \bar{Q}_{k}(t) = \sum_{k=1}^{L} q_{k} - (\mu_{L} - \lambda)t,$$

by (4–7). By (1) this means that $\bar{Q}_{L}(t) = 0$ for some $t \leq \left(\sum_{k=1}^{L} q_{k}\right)/(\mu_{L} - \lambda)$. From (13), $t_{L} \leq K/(\mu_{L} - \lambda)$, where $K$ depends only on $\lambda$ and $\{\mu_{k}\}$.

(ii) Recall that $\bar{Q}, \bar{I}, T$ are Lipschitz. If $k = L - 1$, there is nothing to prove as we already know that $\bar{Q}_{L}(t_{L}) = 0$. So we consider the case when $k \leq L - 2$. The proof is by induction. We note that from (4,6,7), for $\bar{Q}_{L}(t_{L}) = 0$, we must have

$$\bar{D}_{L-1}(t_{L}) - \bar{D}_{L-1}(t_{L} - \delta) \leq \mu_{L}\delta, \text{ for all } \delta \in [0, \delta_{0}], \tag{15}$$

for suitably small $\delta_{0} > 0$. Now if $\bar{Q}_{L-1}(t_{L}) > 0$, then by (4,7), $\bar{Q}_{L-1}(s) > 0$, for all $s \in [t_{L} - \varepsilon, t_{L}]$ for suitable $\varepsilon > 0$. By (11), since $\sigma(k + j) \neq \sigma(L)$ for $j = 1, \ldots, L - k - 1$, we have $\bar{I}_{L-1}(t_{L}) - \bar{I}_{L-1}(s) = 0$ for all $s \in [t_{L} - \varepsilon, t_{L}]$. By (4,8), this implies that $\bar{D}_{L-1}(t_{L}) - \bar{D}_{L-1}(s) > \mu_{L}(t_{L} - s)$, since $b_{L-1}$ satisfies the condition $m_{L-1} < m_{L}$, because $k < L - 1$. This contradicts (15). So we have $\bar{Q}_{L-1}(t_{L}) = 0$, and $\bar{D}_{L-1}(t_{L}) - \bar{D}_{L-1}(t_{L} - \delta) \leq \mu_{L}\delta$, for all $\delta \in [0, \delta_{0}]$.

Now assume that for some $j > k$, and all $i = j + 1, \ldots, L$, we have $\bar{Q}_{i}(t_{L}) = 0$ and $\bar{D}_{i}(t_{L}) - \bar{D}_{i}(t_{L} - \delta) \leq \mu_{L}\delta$, for all $\delta \in [0, \delta_{0}]$ for some $\delta_{0} > 0$ sufficiently small. Then, we claim that these two relations also hold for $b_{j}$. Suppose not. Then either $\bar{Q}_{j}(t_{L}) > 0$, or, for every $\delta_{0} > 0$, there is a $\delta < \delta_{0}$ such that $\bar{D}_{j}(t_{L}) - \bar{D}_{j}(t_{L} - \delta) > \mu_{L}\delta$. Now if $\bar{Q}_{j}(t_{L}) > 0$, then
by (4,7), $\bar{Q}_j(s) > 0$, for all $s \in [t_L - \epsilon, t_L]$ for suitably small $\epsilon > 0$. But by the induction hypothesis, $\bar{D}_l(t_L) - \bar{D}_l(t_L - \delta) \leq \mu_L \delta$ for all $l > j$, $\delta < \delta_0 \wedge \epsilon$. From (4), this implies that $\bar{T}_l(t_L) - \bar{T}_l(t_L - \delta) \leq m_l \mu_L \delta$ for all $l > j$. So

$$\sum_{\{l:j\leq L\, \text{and}\, \sigma(l) = \sigma(j)\}} \left[\bar{T}_l(t_L) - \bar{T}_l(t_L - \delta)\right] \leq \left[\sum_{\{l:j\leq L\, \text{and}\, \sigma(l) = \sigma(j)\}} m_l\right] \mu_L \delta.$$ 

Thus, from (8–11), we have

$$\bar{T}_j(t_L) - \bar{T}_j(t_L - \delta) \geq \left[1 - \sum_{\{l:j\leq L\, \text{and}\, \sigma(l) = \sigma(j)\}} m_l \mu_L\right] \delta.$$ 

So

$$\bar{D}_j(t_L) - \bar{D}_j(t_L - \delta) \geq \mu_j \left[1 - \sum_{\{l:j\leq L\, \text{and}\, \sigma(l) = \sigma(j)\}} m_l \mu_L\right] \delta.$$ 

Rewriting $\sum_{\{l:j\leq L\, \text{and}\, \sigma(l) = \sigma(j)\}} m_l$ as $\left[\sum_{\{l:j\leq L\, \text{and}\, \sigma(l) = \sigma(j)\}} m_l\right] - m_j$, we have

$$\bar{D}_j(t_L) - \bar{D}_j(t_L - \delta) \geq \left[\mu_L + (1 - \frac{\mu_L}{\mu_j^*}) \mu_j\right] \delta$$

(16)

where $\mu_j^* := \frac{1}{\sum_{\{l:j\leq L\, \text{and}\, \sigma(l) = \sigma(j)\}} m_l}$. But by (14), $\mu_j^* > \mu_L$. Thus, we conclude that in either case, for every $\delta_0 > 0$, there is a $\delta < \delta_0$ such that $\bar{D}_j(t_L) - \bar{D}_j(t_L - \delta) > \mu_L \delta$. But by the induction assumption, $\bar{D}_{j+1}(t_L) - \bar{D}_{j+1}(t_L - \delta) \leq \mu_L \delta$, for all $\delta \in [0, \delta_0]$. By (6,7), this implies that $Q_{j+1}(t_L) > 0$, which is a contradiction. So $Q_j(t_L) = 0$ and $\bar{D}_j(t_L) - \bar{D}_j(t_L - \delta) \leq \mu_L \delta$, for all $\delta \in [t_L - \delta_0, t_L]$. 

\[\Box\]

**Lemma 3** Let $b_k$ and $t_L$ be as in the previous lemma. Then

$$\bar{Q}_L(\delta + t_L) = \cdots = \bar{Q}_{k+1}(\delta + t_L) = 0, \text{ for all } \delta \geq 0.$$

**Proof.** If $k = 0$, define $\bar{D}_k(t) := \lambda t$. Otherwise $\bar{D}_k(t)$ is as in (4). Suppose $\bar{Q}_L(t) > 0$ for some $t > t_L$. We know $\bar{Q}_L(t_L) = 0$. So, by (6,7), $\bar{D}_{L-1}(t) - \bar{D}_{L-1}(t_L) > \mu_L(t - t_L)$. But $\bar{Q}_{L-1}(t_L) = 0$. So $\bar{D}_{L-2}(t) - \bar{D}_{L-2}(t_L) > \mu_L(t - t_L)$, and so on. Thus, we can conclude that
$D_j(t) - D_j(t_L) > \mu_L(t - t_L)$, for all $j = k, \ldots, L - 1$. If $k = 0$, we immediately obtain a contradiction. If $k > 0$, by (4, 8-11), this implies that
\[
\sum_{\{i: k \leq i \leq L \text{ and } \sigma(i) = \sigma(j)\}} [T_i(t) - T_i(t_L)] > \left[ \sum_{\{i: k \leq i \leq L \text{ and } \sigma(i) = \sigma(j)\}} m_j \right] \mu_L(t - t_L).
\]
But the right hand side above is at least as large as $(t - t_L)$, by (14). This is a contradiction. So $\bar{Q}_L(t_L + \delta) = 0$, for all $\delta > 0$.

If $k = L - 1$, there is nothing else to prove. So assume that $k < L - 1$. Assume for induction that $\bar{Q}_L(t_L + \delta) = \cdots = \bar{Q}_{j+1}(t_L + \delta) = 0$, for all $\delta > 0$, for some $j > k$. For this, by (6,7), we require that
\[
\bar{D}_l(t) - \bar{D}_l(t - \delta) \leq \mu_L \delta,
\] (17)
for all $l = j, \ldots, L$, all $t > t_L$ and $0 < \delta < t - t_L$. Now if for some $t > t_L$, $\bar{Q}_j(t) > 0$. Then arguing exactly as we did while obtaining equation (16), we get $\bar{D}_j(t) - \bar{D}_j(t - \delta) > \mu_L \delta$, which is a contradiction to (17). So the proof is complete by induction. 

**Proof of the Theorem** In the previous lemmas we have identified a time $t_L$ at which the buffers $b_{k+1}, \ldots, b_L$ are all empty, and remain empty thereafter. If we now show that there exists a time $t_k > t_L$, which depends only on $\lambda$ and $\{\mu_i\}$, such that $b_k$ empties and remains empty thereafter, then we are done by iteration, since there are only $L$ buffers. Now the existence of such a $t_k$ follows analogously to Lemma 2 (i). Suppose $\bar{Q}_k(s) > 0$, for all $s \in (t_L, t_L + \delta]$. Then we have, from (8-11),
\[
\bar{T}_k(t_L + \delta) - \bar{T}_k(t_L) = \delta - \sum_{\{i: k < i \leq L \text{ and } \sigma(i) = \sigma(k)\}} [\bar{T}_i(t_L + \delta) - \bar{T}_i(t_L)].
\] (18)
But from Lemma 3, we can conclude that $\bar{D}_k(t_L + \delta) - \bar{D}_k(t_L) = \bar{D}_k(t_L + \delta) - \bar{D}_k(t_L)$, for all $i = k + 1, \ldots, L$. Using (4) and (18), we get
\[
\bar{D}_k(t_L + \delta) - \bar{D}_k(t_L) = \frac{\delta}{\sum_{\{i: k \leq i \leq L \text{ and } \sigma(i) = \sigma(k)\}} m_i}.
\]
Thus we conclude that
\[
0 \leq \sum_{i=1}^{L} \bar{Q}_i(t_L + \delta) = \sum_{i=1}^{L} \bar{Q}_i(t_L) + \left( \lambda - \frac{1}{\sum_{\{i: k \leq i \leq L \text{ and } \sigma(i) = \sigma(k)\}} m_i} \right) \delta.
\]
Since the second term on the right hand side above decreases linearly as long as $Q_k$ stays positive, we conclude that there is a time $t_k$ at which $\tilde{Q}_k(t_k) = 0$. The fact that $t_k$ depends only on $\lambda$ and $\{\mu_i\}$ follows from

$$\sum_{i=1}^{L} \tilde{Q}_i(t_L) < \lambda t_L + \sum_{i=1}^{L} q_i,$$

since the right hand side above is no larger than $K \left(1 + \frac{1}{(\lambda L - \lambda)}\right)$ by (13) and Lemma 2 (i).

At time $t_k$ we next identify a $k'$ such that

$$k' = \max\{j < k \mid \sum_{\{i : j \leq i \leq L \text{ and } \sigma(i) = \sigma(j)\}} m_i \geq \sum_{\{n : k \leq n \leq L \text{ and } \sigma(n) = \sigma(k)\}} m_n\}.$$

This is the bottleneck which caused $b_k$ to empty. Then we repeat the arguments of Lemmas 3 and 4 to show that $b_{k+1}, \ldots, b_k$ are empty at $t_k$ and remain empty thereafter. Iterating the argument proves the fact the fluid limits remain empty after a finite time.

We note that we can relax the assumption that residual service times are zero, i.e., the exponential assumption, since it is easy to see that when all downstream buffers are empty, a buffer emptying at time $t_k$ in the exponential case, would empty at time $t_k + v_k$ in general, and $\sum v_k < 1$. Also see Theorem 5.2 of Chen [3].

Now we apply Theorem 4.3 of Dai [1] to conclude the queueing network is positive Harris recurrent.

\[\Box\]

## 3 Conclusions

We have shown that the LBFS policy is stable whenever the arrival rate is within capacity, i.e., (1) is satisfied. This provides an useful extension of the stability of bursty deterministic models, established in [2], to the stochastic setting, and an interesting application of the results of [1], which proves their power. It also exhibits techniques of deducing the behavior of the fluid limits from the integral equations which they satisfy, and hence may be useful in future work. Finally, we note that Dai and Weiss [4] have simultaneously and independently obtained a proof of the stability of LBFS. We also refer the reader to [4] for other stability/instability results for some examples of systems.
References


