Amina Angelika Bouchentouf (Sidi Bel Abbes)

STABILITY CONDITIONS OF A QUEUEING SYSTEM MODEL VIA FLUID LIMITS

Abstract. We study the ergodicity of a multi-class queueing model via fluid limits which have the advantage of describing the model in macroscopic form. We consider a model of processing bandwidth requests. Our system is defined by a network of capacity $C = N$, and a queue which contains an infinite number of items of various sizes 1, a' and b' with $1 < a' < b' < N$. The problem considered is: Under what conditions on the parameters of some large classes of networks, do they reach the stationary regime?

1. Introduction. In this paper we study a queueing model of storage and transmission of bandwidth in computer and communication systems, considering a description of allocation of different streams of messages in a communication network.

A model of processing bandwidth requests of different sizes is considered. Messages in the network are of different nature, and to be transmitted they require different throughput. In this paper the First Fit algorithm policy is considered: the sum of the sizes of messages in the network is less than its capacity; following every event (arrival or departure), the queue is scanned from the beginning in search of an item whose size is smaller than the empty space left in the network. This procedure is repeated until the end of the queue is reached.

The network is a bin of size $C = N$ (N can be large but it is finite), messages are items of different sizes 1, a' and b' such that $1 < a' < b'$ and the bandwidth required by a message is the size of the item. Items have the same distribution as some random variable S_1. The items arrive to the bin according to a Poisson process $\mathcal{N}_\lambda$ with parameter λ; for $t \geq 0$, the quantity

[431]

$\mathcal{N}_\lambda(]0, t])$ denotes the number of arrivals between 0 and t, and each item requires service time of mean 1 (all the service times are i.i.d. exponentially distributed random variables with parameter 1).

The sizes (S_i) of the items form an i.i.d. sequence with a common distribution $F(dx)$ given by

$$F(dx) = p\delta_1 + q\delta_{a'} + r\delta_{b'},$$

where δ_x is the Dirac measure at x and p, q, r are nonnegative numbers such that $p + q + r = 1$.

The set of possible sizes is denoted by $\mathcal{T} = \{1, a', b'\}$ and $\mathcal{T}^{(\mathbb{N})}$ is the set of finite vectors with coordinates in $\mathcal{T}$; if $x \in \mathcal{T}^{(\mathbb{N})}$, $\|x\|$ denotes the number of coordinates of x.

An element X of the state space $\mathcal{S}$ of the Markov process describing the storage process can be written as $X = (B, L)$, where L and B are elements of $\mathcal{T}^{(\mathbb{N})}$, the set of finite vectors with coordinates in $\mathcal{T}$.

The following notation will be employed in the remainder of this paper:

$$\overbrace{(a', \ldots, a')}^{x\,\text{items}\,a'} = [x \text{ items } a'], \qquad \overbrace{(a', \ldots, a', b', \ldots, b')}^{x\,\text{items}\,a',\,y\,\text{items}\,b'} = [x \text{ items } a', y \text{ items } b'].$$

Important point: The choice of a' and b' depends on the choice of the capacity of the bin, N.

Notice that the order of the components in B has no importance but it is important for the vector L since the First Fit discipline checks if the first coordinate fits in the bin, then the second, the third, and so on. The vector L is a string of 1, a' and b'.

If $(X(t)) = (B(t), L(t))$ is the state of the system at time t, then $(X(t))$ is an irreducible Markov process on $\mathcal{S}$; our model is stable when $(X(t))$ is ergodic on $\mathcal{S}$. In Dantzer et al. [1] it has been proved that the condition $\lambda \mathbb{E}(S_1) \leq C$ is necessary for the stability of the system.

Then we define by

$$W(X(t)) = \sum_{i=1}^{\|B(t)\|} b_i \sigma_i^0(t) + \sum_{j=1}^{\|L(t)\|} l_j \sigma_j(t)$$

the load of $(X(t)) = (B(t), L(t)) = ((b_i(t)), (l_j(t)))$. Moreover $\|X\|$ is the norm of the state $X = (B, L) \in \mathcal{S}$, equal to the sum of $\|B\|$ and $\|L\|$, $\sigma_i^0(t)$ is the residual service time of the item $b_i(t)$, and $\sigma_j(t)$ the service time of the item $l_j(t)$.

The probabilistic description of this model involves an infinite-dimensional vector space (a space of strings).

The First Fit algorithm with items having two possible sizes has been analyzed in Dantzer et al. [1]; a stability condition has been established and, more interestingly, a curious transient behaviour was analyzed.

The First Fit algorithm with three possible sizes 1, 2 and 3 and capacity of the bin C equal to 4 has been analyzed by Dantzer and Robert [2].

The present paper proposes some stability results for some large classes of networks with items of different sizes 1, a' *and* b' *such that* $1 < a' < b'$; *the sizes of items depend formally on the capacity of the bin which is* $C = N$. *We prove that under the conditions* $\lambda p > N - b'$ *and* $\lambda \mathbb{E}(S_1) < N$ *our system is ergodic; then in the case when* $\lambda p < N - b'$ *the concept of smooth distribution on the state space is introduced, which is used together with the concept of piecewise linear processes in* $\mathbb{R}_+^2$ *to establish ergodicity and transience for our system.*

In the next section we present two results; the first gives an estimate of the wasted space when there are only two possible sizes: 1 and a', and where it is not important to have much information on the structure of the queue of the initial state. Then we consider the general case with three sizes, and we show that the condition $\lambda \mathbb{E}(S_1) < N$ turns out to be sufficient for ergodicity when $\lambda p > N - b'$.

2. Stability of the system when $\lambda p > N - b'$

LEMMA 2.1. *Under the condition* $\lambda \mathbb{E}(S_1) < N$, *suppose the system processes only items* 1 *and* a', τ *is the first time where the queue is empty, and* D *is the duration of time during which the bin is not full during a busy period* (*between* 0 *and* τ). *Then* $\mathbb{E}_x(D) \leq K$ *with* $K = K_1 \log(1 + \|x\|) + K_2$ *and* $x = (l, b) \in \mathcal{S}$,

Proof. First, we note that if there are some items of size 1 in the queue there is no waste of space in the bin. So, we can assume that l is a string of items a'.

The only possibility to waste space with a nonempty queue and an empty space of size 1 in the bin is when the state of the bin is $[N-1$ items of size 1$]$, $[1, \frac{N-2}{a'}$ items $a']$, $[a', (N - (a' + 1))$ items 1$]$, $[\frac{N-1}{a'}$ items $a']$, $[q_1$ items 1, q_2 items $a']$, $q_1 \neq q_2$, and if $q_1 = q_2$ we will have $b = [\frac{N-1}{a'+1}$ items 1, $\frac{N-1}{a'+1}$ items $a']$. We set $A_0 = l$, $T_0 = 0$ and

$$T_{n+1} = \inf\{t > T_n : C(t-) = N, \ C(t) < N, \ \text{and}$$
$$\text{all the items } a' \text{ present at time } T_n \text{ are served at time } t\}$$

with $A_n = \|L(T_n)\|$ for $n \geq 1$. Then $L(T_n)$ is necessarily a string of items a' which can be empty. The sequence $(B(T_n), A_n)$ is clearly a Markov chain.

• $b = [(N - 1)$ *items of size* 1$]$. If there is at least one item 1 in the queue the items a' are ignored (the First Fit algorithm picks only items of size 1). Since $\lambda p \leq \lambda \mathbb{E}(S_1) < N$, after an integrable amount of time not depending on $\|l\|$, some places will be vacant in the bin and consequently an item a'

will enter the bin. In this situation the number of items 1 is the number of customers of an $M/M/N$ queue.

- $b = \left[\frac{N-1}{a'+1} \text{ items } 1, \frac{N-1}{a'+1} \text{ items } a'\right]$.

  1. $\lambda p < 2$. In this case all items of size a' will be served consecutively. The expected value of this duration of time is bounded with respect to $\|l\|$. When the initial items a' have been served, the queue is an i.i.d. string of items 1 and a'. Then with probability 1 at least $\frac{N+a'}{a'+1}$ items 1 enter the bin. Later, when the number of items 1 in the system is $\frac{N-1}{a'+1}$ the system will waste some space; this is the definition of time T_1, and A_1 is the number of items at that time.

  2. $\lambda p > 2$. If the state of the bin does not change, the arriving items 1 will saturate $\frac{N+a'}{a'+1}$ places in the bin. In this case, the number of items 1 is the number of customers of a transient $M/M/\frac{N+a'}{a'+1}$ queue. A change in the state of the bin may occur only if this transient queue is empty.

     (a) *The $M/M/\frac{N+a'}{a'+1}$ queue never reaches the empty state.* After a short period of time, the bin will be full with $\frac{N-1}{a'+1}$ items a' and $\frac{N+a'}{a'+1}$ items 1. The condition $\lambda\mathbb{E}(S_1) < N$ implies that $\lambda q < 1$ ($\lambda p > 2$), therefore with probability 1 after some period of time the system will not contain any items a'. At that time the state of the bin will be $[N$ items 1$]$. Then, with probability 1, the total number of items 1 will be less than $\frac{N+a'}{a'+1}$. Moreover, $\frac{N-1}{a'+1}$ items a' will be in the bin; this is the starting situation.

     (b) *The queue reaches the empty state.* There are only a few cases where this transient queue reaches the empty space. The items a' occupy the bin with no empty space. All the initial items a' are served. In this situation T_1 is the time when there is some wasted space and after that time all items a' were served.

Notice that case (a) occurs only a geometrically distributed number of times. Hence, the duration of time between time 0 and T_1 when the bin is not full has a bounded expected value (with respect to $\|l\|$).

Consequently, there exist a_0 and $\alpha_1 < 1$ such that for $\|x\| > a_0$,

$$(2.1) \qquad \mathbb{E}_x(A_1) \leq \alpha_1\|x\|, \quad \text{where} \quad \alpha_1 = 1 + c\frac{a'+1}{N+a'}(\lambda\mathbb{E}(S_1) - N),$$

where c is a positive constant such that

$$\lim_{\|x\|\to+\infty} \frac{\mathbb{E}_x(T_1)}{\|x\|} = c \quad \text{in } L_1 \text{ and almost surely,}$$

and

$$(2.2) \qquad \gamma_1 = -\log\left(\frac{1+\alpha_1\|x\|}{1+\|x\|}\right) > 0.$$

Then, for the cases $\left[1, \frac{N-2}{a'} \text{ items } a'\right]$, $[a', N-(a'+1) \text{ items } 1]$, $\left[\frac{N-1}{a'} \text{ items } a'\right]$ and $[q_1 \text{ items } 1, q_2 \text{ items } a']$ the values of α_i and γ_i are:

$$\begin{cases} \gamma_i = -\log\left(\dfrac{1+\alpha_i\|x\|}{1+\|x\|}\right) > 0, \quad i = 2,\ldots,5, \\[2mm] \alpha_2 = 1 + \dfrac{c}{2}(\lambda\mathbb{E}(S_1) - N), \quad \alpha_3 = 1 + \dfrac{c}{N-a'}(\lambda\mathbb{E}(S_1) - N), \\[2mm] \alpha_4 = 1 + c(\lambda\mathbb{E}(S_1) - N), \quad \alpha_5 = 1 + \dfrac{c}{q_1+1}(\lambda\mathbb{E}(S_1) - N). \end{cases}$$

Now, we define $\nu_i = \inf\{n \geq 1 : A_n \leq a_0\}$; then the sequence $(Z_n) = (\log(1 + A_{n\wedge\nu_i}) + \gamma_i(n \wedge \nu_i))$ is a super-martingale, $i = 1,\ldots,5$. Then the contribution of the ν_i cycles in the integral defining D is bounded by $K\mathbb{E}_x(\nu_i) \leq K\log(1 + \|x\|)/\gamma_i$, for some constant K.

Proposition 6 of Dantzer et al. [1] shows that the system is ergodic. Consequently, starting from the state $(B(T_{\nu_i}), A_{\nu_i})$, the hitting time of the empty state $\emptyset$ is integrable and with an expected value bounded with respect to $\|x\|$. Therefore, the expected value of the contribution of this period in the integral defining D is bounded with respect to $\|x\|$. ∎

PROPOSITION 2.2. *If $\lambda\mathbb{E}(S_1) < N$ and $\lambda p > N - b'$, then $(X(t))$ is an ergodic Markov process.*

Proof. To prove this result we have to suppose that D, the duration of time (between 0 and τ) during which the bin is not full during a busy period, is integrable and

$$\lim_{n\to\infty} \frac{\mathbb{E}_{x_n}(D)}{\|x_n\|} = 0.$$

If this assumption is true we get

$$(2.3) \qquad \limsup_{n\to\infty} \frac{\mathbb{E}_{x_n}(\tau)}{\|x_n\|} \leq \frac{1}{N - \lambda\mathbb{E}(S_1)},$$

$$(2.4) \qquad \limsup_{n\to\infty} \frac{\mathbb{E}_{x_n}(W(X(\tau)))}{W(x_n)} \leq \frac{\lambda\mathbb{E}(S)}{N} < 1.$$

Here τ is the first time when the bin is not full after all the initial items a' have left the system (τ is a stopping time). Notice that for the states (x_n) the bin is not full, thus the queue is a string of items a' and b' (see Proposition 5 of Dantzer et al. [1]).

The two relations (2.3) and (2.4) and Theorem 1 of [2] show that the Markov process $(X(t))$ is ergodic.

Then to prove that really $(\mathbb{E}_{x_n}(D))$ is negligible with respect to $\|x_n\|$ when n is large, we have to discuss several possibilities for b.

(1) b *is* $[N-1 \text{ items } 1]$ *or* $[N \text{ items } 1]$. If there is at least one item 1 in the queue, all the other items are ignored. Since $\lambda p \leq \lambda\mathbb{E}(S_1) < N$, from the

point of view of items 1, the system is a stable $M/M/N$ queue. Hence the first time there will be some empty places (maybe equal to the size of the item a') is a bounded integrable variable. At that time, an item a' will be inserted in the bin. Notice that for this period, the duration of time during which the bin is not full is a bounded integrable variable.

(2) *b has some items a'.*

  (a) *By adding an item of size b' one will exceed the maximum capacity of the bin N.* The items b' are not taken into account as long as the bin contains some items a' or some items 1 such that by adding b' one will exceed the capacity of the bin. In this case the items b' are ignored. Consequently, a string of the initial items b' builds up at the beginning of the queue. Since the condition $\lambda \mathbb{E}(S_1) < N$ implies $\lambda p + a'q < N$, the system with items 1 and a' is stable (see Dantzer et al. [1]). Lemma 1 shows that until an item b' enters the bin, the wasted space is negligible compared to the number of initial items a'.

  (b) *Else.* All items b' are selected by the First Fit algorithm. Since the condition $\lambda \mathbb{E}(S_1) < N$ implies $\lambda q + b'r < N$, the system with items a' and b' is stable; after some time, the condition $\lambda p > N - b'$ implies that the residual space in the bin left by items a' or b' is saturated by items 1. Consequently, the duration of time when the bin is not full is a bounded integrable variable.

(3) *b contains some items b' such that by adding an item of size b' or a' one will exceed the maximum capacity of the bin, N.* Now l_n is a string of items b'. Otherwise, if at some time some items a' enter the bin, items 1 and a' will be cleared from the system until some items b' are in the bin. The condition $\lambda p > N - b'$ implies that the residual space in the bin left by items b' is saturated by items 1. Consequently, the duration of time when the bin is not full is a bounded integrable variable.

This discussion shows that the assertion is proved. ∎

3. Stability of the system when $\lambda p < N - b'$. When $\lambda p < N - b'$, the situation is more delicate, and for this reason the concept of smoothing initial state is introduced.

3.1. Main result for the smoothing case. In this section we shall assume that $\lambda p < N - b'$ and that the initial states are strings of items a' and b'. The next definition formalizes the notion of "smooth random" state, in fact the notion of a smooth distribution on $\mathcal{S}$.

DEFINITION 3.1. For integers l, m, n we define a distribution $R_{l,m,n}(dx)$ on $\mathcal{T}^{(\mathbb{N})}$ by

(3.1) $$R_{l,m,n}(dx) = \delta_{b'}(du)^{(l)} \otimes F_{a',b'}(du)^{(m)} \otimes F(du)^{(n)},$$

where $F(dx)^{(n)}$ is the nth power of the distribution $F(dx)$, and $F_{a',b'}(dx)$ is the conditional distribution $F_{a',b'}(dx) = (q\delta_{a'} + r\delta_{b'})/(q+r)$.

A distribution μ on $\mathcal{S}$ is smooth if its L-component is in the convex hull of the $R_{l,m,n}^{l \times m \times n}$, $l, m, n \in \mathbb{N}$, i.e. there exists a probability distribution (q_i) on $\mathbb{N}^3$ such that

$$\mu(L \in dx) = \sum_{i \in \mathbb{N}^3} q_i R_i(dx).$$

The distribution $R_{l,m,n}$ is the distribution of the concatenation of several i.i.d. strings. The L-component of a distribution of type $R_{0,0,n}(dx)$ is just an i.i.d. string of length n with distribution F.

PROPOSITION 3.2. *If $\lambda\mathbb{E}(S_1) < N$, then for any stopping time U greater than the first time when all the initial items have left the queue, the distribution of $X(U)$ is smooth.*

In [2] it was proved that if $\lambda\mathbb{E}(S_1) < 4$, $C = 4$ and $\mathcal{T} = \{1, 2, 3\}$ the distribution of $X(U)$ is smooth. For a large class of systems, neither the capacity of the bin nor the sizes of the items affect the result, and under the hypotheses of our proposition we will get a smooth distribution.

PROPOSITION 3.3. *If $\lambda\mathbb{E}(S_1) < N$, $\lambda p < N - b'$ and U_0 is the first time t after all the initial items have left the queue that $B(t) = [b', N - (b' + 1)$ items 1] then*

(3.2) $$\sup_{x \in \mathcal{S}_1} \mathbb{E}_x\left(\left(\frac{U_0}{\|x\|}\right)^2\right) < \infty \quad and \quad \sup_{x \in \mathcal{S}_1} \mathbb{E}_x\left(\left(\frac{\|X(U_0)\|}{\|x\|}\right)^2\right) < \infty$$

where $\mathcal{S}_1$ is the subset of the states of S for which the bin is not full.

Proof. The initial state is given by $x = (B, L)$ with $L = (l_1, \ldots, l_p)$ for some $p \geq 1$. Let T be the first time when all the initial items have left the queue. Then T is a stopping time bounded by $T_{a'} + T_{b'}$, where $T_{a'}$ [resp. $T_{b'}$] is the time when all the initial items a' [resp. b'] have left the queue.

For a fixed $k \in \{1, \ldots, p-1\}$, we define $\check{x} = (B, \check{L})$ where $\check{L}$ is the same string as L except that the components k and $k+1$ are permuted. For $1 \leq i < p$, the quantities τ_i, $\check{\tau}_i$ denote respectively the waiting time necessary for the ith item l_i to enter the bin when the initial state is respectively x, $\check{x}$. We assume that for these two initial states, the arrival stream and the services associated with the items are the same. There are two cases: $\tau_{k+1} < \tau_k$ and $\tau_{k+1} \geq \tau_k$ and the initial state is $\check{x}$. In any case $\tau_k \leq \check{\tau}_k$. By induction, the quantity $\mathbb{E}_x(T_{a'})$ is thus bounded by $\mathbb{E}_{x'}(T_{a'})$ where $x' = (B, L')$ with $L' = (b', \ldots, b', a', \ldots, a')$. Similarly, $\mathbb{E}_x(T_{a'}) \leq \mathbb{E}_{x''}(T_{a'})$ if $T_{a'}$ is the time to get rid of the initial items a' and $x'' = (B, L'')$, where L'' is a permutation of the L-component of x when all items b' are at the head of the queue.

To bound $\mathbb{E}_x(T^2)$ it is sufficient to give an upper bound for $\mathbb{E}_{x''}(T_{a'}^2)$ and $\mathbb{E}_{x'}(T_{b'}^2)$.

- *Items a' are at the beginning.* We get, for some constant A_1, $\mathbb{E}_{x'}(T_{b'}^2) \leq A_1 \|x\|^2$ (notice that $\|x'\| = \|x\|$).

- *Items b' are at the beginning.*
  - If the initial state of the bin is $[k$ items $b']$, for a convenient constant A_2, one easily gets $\mathbb{E}_{x''}(T_{a'}^2) \leq A_2 \|x\|^2$.
  - If there is a b' in the bin and at least $N - (b' + 1)$ items 1, we get $\mathbb{E}_{x''}(T_{a'}) \leq A_3 \|x\|$ for some constant A_3.

At time T all the initial items have left the queue. Since $X(T) \leq \|x\| + \mathcal{N}_\lambda(]0, T])$, Wald's formula and the above estimate show that for some constant A, $\|X(T)\| \leq A\|x\|$.

Then to estimate the first time $\bar{T}$ when the state of the bin is $[b', N - (b' + 1)$ items 1$]$ it is sufficient to prove that if the initial state is x, then $\bar{T}$ has a second moment of order $\|x\|^2$. The first step is to get rid of items a'. There are many possibilities for the bin:

- The bin has some items 1 (at least $N - (b' + 1)$). Since $\lambda p < N - b'$, after some time the items 1 which were in the queue will enter the bin and the state will reach $[N - b'$ items 1, $\cdot]$.

  * If, at that time, there are sufficiently many items b' in the queue, the state of the bin will reach the state $[N - b'$ items 1, $b']$. Then, with probability 1, the state of the bin will be $[b', N - (b' + 1)$ items 1$]$.
  * Else. All the items a' in the queue at that time are served ($[N - b'$ items 1, $a']$). When this is finished, the condition $\lambda p < N - b'$ implies that the number of items 1 is tight. The items b' accumulated during that time are served; consequently, with probability 1, the state of the bin will be $[b', N - (b' + 1)$ items 1$]$.

- $b = [N/b'$ items $b']$ (*the bin is full*). The initial items b' are served consecutively at rate N/b'. Then with probability 1, an item a' will enter the bin and then all the other items a' will be processed consecutively. When this is finished, since $\lambda p < N - b'$, $N - b'$ items 1 enter the bin. This is the situation of the previous case.

- $b = [N/a'$ items $a']$ (*the bin is full*). As in the previous case, the initial items a' are served consecutively at rate N/a'. When this is finished, always with probability 1, $N - a'$ items 1 enter the bin; then we will have $N - b'$ items 1 in the bin, the situation of the previous case.

For the case where the bin is not full (there is an empty space of size 1), we follow the same discussion with a very small change which depends on the sizes of the items.

- $b = [k_1$ items a', k_2 items $b']$ (*the bin is full*)

  * If b' is served before some items a', the First Fit algorithm selects all the items b', then all the initial items a' are served at that time; we see the arriving of items 1 wastes the empty space, until we have $N - b'$ items 1 in the bin; since $\lambda p < N - b'$ the bin reaches the state $[N - b'$ items 1, $a']$, the situation of the previous case.
  * If a' is served before item b', the First Fit algorithm selects all the items a'. When this is finished, some items b' will be served, then with probability 1, $N - (b' + 1)$ items 1 enter the bin ($[b', N - (b' + 1)$ items 1]).

- $b = [k_1$ items a', k_2 items 1] (*the bin is full*).

  * $b = [a', N - a'$ items 1]
  ** If a' is served before some items 1, the First Fit algorithm selects all the initial items a', then with probability 1, we will have $N - (b' + 1)$ items 1 in the bin; at that time an item b' will enter the bin.
  ** If 1 is served before a', the items 1 which are in the queue will be served consecutively, then all the items a' will be served; after some time we will have, with probability 1, $N - (b' + 1)$ items 1 in the bin; at that time an item b' enters the bin.

When $b = [1, \frac{N-1}{a'}$ items $a']$, $b = [\frac{N}{a'+1}$ items a', $\frac{N}{a'+1}$ items 1] we will follow the same discussion. And there is not much difference in the cases where the bin is not full: $b = [1, \frac{N-2}{a'}$ items $a']$, $b = [a', N - (a' + 1)$ items 1], $b = [\frac{N-1}{a'+1}$ items 1, $\frac{N-1}{a'+1}$ items $a']$.

- $b = [k_1$ items b', k_2 items 1] (*the bin is full or not*). This case is easy (we follow the same discussion).

Notice that each of the steps we have described has a duration with second moment of order $\|x\|^2$. And the last inequality is a consequence of Wald's formula applied to the stopping time U_0. ∎

3.2. A random dynamical system in $\mathbb{R}^2_+$ for a large class of networks. In this section we assume that (μ_n) is a sequence of smooth distributions on $\mathcal{S}$ such that

(3.3) $\qquad \mu_n(B = [b', N - (b' + 1)$ items 1], $L \in dx) = \mathbb{E}(R_{y_n, w_n, 0}(dx))$,

where y_n, w_n are random variables such that

$$\lim_{n \to \infty} \frac{y_n}{n} = y, \quad \text{and} \quad \lim_{n \to \infty} \frac{w_n}{n} = w \quad \text{in } L_1,$$

y and w are nonnegative integrable random variables and $\mathbb{P}(y + w > 0) = 1$. The B-component of μ_n is $[b', N - (b' + 1)$ items 1] and the L-component

of the distribution μ_n does not have an item 1 in the queue; it is the concatenation of y_n items b' followed by an i.i.d. string of length w_n of a' and b' with respective probability $q/(q+r)$ and $r/(q+r)$.

In all that follows, $a' = 2$.

If the distribution of $X(0)$ is given by μ_n, then $\|X(0)\|$ is equivalent to $((y+w)n)$. (Note that a sequence (X_n) of random variables is equivalent to (α_n) if the sequence (X_n/α_n) converges to 1 in $L_1(\mathbb{P})$.)

The next proposition proves that the distribution of X at a stopping time has a property similar to Identity (3.3).

PROPOSITION 3.4. *Let U_1 be the first time when all the initial items a' have left the queue, the initial items b' in the bin have been served and the state of the bin is $[b', N - (b'+1)$ items $1]$. Then there exist $\mathcal{F}_{U_1}$-measurable random variables Y_n and W_n such that*

$$\mathbb{P}_{\mu_n}(B(U_1)) = ([b', N - (b'+1) \, items \, 1], L(U_1) \in dx) = \mathbb{E}_{\mu_n}(R_{Y_n, W_n, 0}(dx)),$$

and a random matrix M such that

$$(3.4) \qquad \lim_{n \to \infty} \frac{1}{n}(Y_n, W_n) = M \cdot (y, w) \qquad almost \ surely \ and \ in \ L_1.$$

The random matrix M has two possible values with equal probability,

$$(3.5) \qquad m_1 = \begin{pmatrix} 1 & \frac{1-p-q}{(N-b')(1-p)} \\ 0 & \lambda(q + \frac{\lambda p(1-p-q)}{(N-b')(N-\lambda p)}) \end{pmatrix}$$

and

$$m_2 = \begin{pmatrix} \frac{\lambda}{N-b'}(1-p-q) & \frac{1-p-q}{(N-b')(1-p)} \\ \lambda^2(1-p)(q + \frac{\lambda p(1-p-q)}{(N-b')(N-\lambda p)}) & \lambda(q + \frac{\lambda p(1-p-q)}{(N-b')(N-\lambda p)}) \end{pmatrix}.$$

M is independent of (y, w) if $\mathbb{P}(y > 0, w > 0) = 1$.

Proof. Let $b = [b', N - (b'+1)$ items $1]$ at time 0. If a new item 1 arrives, then the bin is full and, during that time, the b' in the bin is replaced by the initial items b'. So, the state of the bin will return to $[b', N - (b'+1)$ items $1]$. In this manner, a finite number of initial items b' in the queue are served in the bin before a significant change occurs. And if there is a departure of an item 1, an item of size $a' = 2$ replaces it, and since (w_n) converges almost surely to infinity, there will be an item a' in the queue with probability 1 at the occasion of such a departure. So the fact that an item b' or an item a' enters the bin is independent of the limit of $(y_n, w_n)/n$, as long as y and w are positive with probability 1. (y_n/n) and (w_n/n) converge almost surely (see Skorokhod's representation theorem with a change of probability space).

Throughout this discussion, we shall ignore small strings in our statements. Then if the initial distribution is given by μ_n, the initial state of the

bin is $[b', N - (b' + 1)$ items 1]. If there is a departure before a new arrival, there are two possibilities:

1) With probability $\frac{1}{1+[N-(b'+1)]}$ the item b' leaves first. In this case the first item b' enters the bin and all the other $y_n - 1$ items b' will follow it. During that time, since $\lambda p < N - b'$, the items 1 are processed by the empty space in the bin. So the time τ_1 to get rid of the initial items b' is equivalent to $y_n \sim yn$.

   At time τ_1 the length of the queue is thus equivalent to $(w+\lambda(q+r)y)n$ [2]. As soon as an item of size a' is in the bin, it is easy to check that after an integrable period of time the bin will be filled up with items a' and some items 1. The time τ_2 to serve all the items a' that arrive before the bin is filled up either with only items of size a' or items of size a' and one item 1 is equivalent to the quantity $(w + \lambda(q+r)y)\frac{q}{q+r}n$. At time $\tau_1 + \tau_2$ there is a string of b''s at the head of the queue of length equivalent to

$$(3.6) \qquad (w + \lambda(q + r)y)\frac{r}{(N - b')(q + r)}n$$

followed by an i.i.d. string with distribution $F(du)$ whose length is equivalent to the quantity $\lambda(w+\lambda(q+r)y)\frac{q}{q+r}n$. At that time we have to discuss two cases, but at the end we will have the same result.

♣ *The bin is filled up with items of size a'.* If there is a departure of an item a' it is replaced immediately by an item a' ($a' = 2$) or two items 1. The items b' cannot be served at that time. Due to the i.i.d. structure of the queue at that time, after an integrable period of time, the bin reaches $[N$ items 1]; from that time on, all items 1 are served at rate N. The time τ_3 it takes to empty the queue of items 1 and to have exactly one b' and $N - (b' + 1)$ items of size 1 in the bin is equivalent to

$$\frac{\lambda p(w + \lambda(q + r)y)r}{(N - b')(q + r)(N - \lambda p)}n.$$

At time $\tau_1 + \tau_2 + \tau_3$ there is a string of b''s whose length is equivalent to (3.6), followed by a string of a''s and b''s of length equivalent to

$$(3.7) \qquad \lambda(w + \lambda(q + r)y)\left(q + \frac{\lambda rp}{(N - b')(N - \lambda p)}\right).$$

♣ *The bin is filled up with some items a' and one item 1.*

   1) If there is a departure of an item a' we will have the same discussion as in the previous case.
   2) If there is a departure of an item 1 it is replaced by an item 1, then all items 1 which are in queue will be served (the essential point in this phase is that the items b' cannot be served); after that each

item a' which departs will be replaced immediately by an item a'
or items 1, then we will follow the same discussion.

Finally the distribution of $L(U)$ is given by $\mathbb{E}_{\mu_n}(R_{Y_n,W_n,0}(dx))$ and (Y_n, W_n)
satisfies the relation (3.4) with the matrix $M = m_2$.

2) With probability $\frac{N-(b'+1)}{1+(N-(b'+1))}$ an item 1 is served first. This situation is
simpler since the initial items b' are not served at time U_1.

REMARK 3.5. The dynamic of the system is influenced by the fact that
either the b' leaves first or not. This is also true at the fluid level. The case
where the random bifurcation may depend on the current state was shown
in Robert [4].

The discussion is similar on the set $\{y = 0, w \neq 0\} \cup \{y \neq 0, w = 0\}$;
the difference is that the duration of some transitions described above is
negligible in this case. ∎

The following proposition presents the most important result on the
ultimate behavior of the fluid limits.

PROPOSITION 3.6. *If (M_n) is an i.i.d. sequence of random matrices with
the same distribution as M in Proposition* 3.4 *and $P_n = M_n \cdot M_{n-1} \cdots M_1$,
then there exist $\alpha, \beta > 0$ and a function η on $\mathbb{R}_+$ such that for any $n \in \mathbb{N}$
and $x \in \mathbb{R}_+^2$,*

$$(3.8) \qquad \mathbb{E}(\langle(\alpha, \beta), P_n \cdot x\rangle) = \eta(\lambda)^n \langle(\alpha, \beta), x\rangle$$

where $\langle\cdot, \cdot\rangle$ is the usual scalar product in $\mathbb{R}^2$. If

$$(3.9) \quad \lambda^* = \frac{1}{2(-qNp + qb'p)}[(-qN^2 + Nq + qb'N - N + b'p)$$
$$+ (N^2 - 2qN^2 + q^2N^2 + 2pqb'N - 2q^2N^3b' + 2N^2q^2b' + q^2b'^2N^2 - 2qb'N^2$$
$$- 2Nb'p + q^2N^4 - 2q^2N^3 + 2qN^3 + b'^2p^2 + 6qN^2b'p - 2qb'^2Np - 4qN^3p)^{1/2}]$$

then $\eta(\lambda) < 1$ if $\lambda < \lambda^*$ and $\eta(\lambda) > 1$ if $\lambda^* < \lambda < N/p$.

The proof of this proposition is based on the fact that if $\lambda p < N$, the
same property is true for the largest eigenvalue $\eta(\lambda)$ of the matrix $\mathbb{E}(M_1)$,
where $\mathbb{E}(P_n) = \mathbb{E}(M_1)^n$. The smallest root of the equation $\eta(\lambda) = 1$ is given
by $\lambda = \lambda^*$. Then it can be shown that the term under the square root in
(3.9) is non-negative if $p + q \leq 1$ and that $\lambda^* p < N$.

3.3. Ergodicity and transience theorem

THEOREM 3.7. *When the arrival rate of the items is λ, the distribution
of their sizes is given by $F(dx) = p\delta_1 + q\delta_{a'} + (1 - p - q)\delta_{b'}$, and the size of*

the bin is N, if

$$(3.10) \qquad \lambda_{\mathrm{FF}} := \min\left\{\lambda^*, \frac{N}{b' + (1 - b')p + (a' - b')q}\right\}$$

then the Markov process $(X(t))$ describing the First Fit algorithm is ergodic when $\lambda < \lambda_{\mathrm{FF}}$, and transient when $\lambda > \lambda_{\mathrm{FF}}$.

Proof. Ergodicity. If $\lambda p > N - b'$, Proposition 2.2 shows that the condition $\lambda \mathbb{E}(S_1) < N$, i.e. $\lambda(b' + (1 - b')p + (a' - b')q) < N$, is sufficient for the ergodicity of $(X(t))$. One can check that in that case

$$\frac{N}{b' + (1 - b')p + (a' - b')q} < \lambda^*.$$

Now, we assume that conditions (3.10) and $\lambda p < N - b'$ are satisfied. To prove the ergodicity under these hypotheses it is sufficient to prove that there exists a stopping time V such that for any sequence $(x_n) = (b_n, l_n)$ in $\mathcal{S}_\infty$ with $\|x_n\| = n$, the following inequalities hold (Theorem 1 of [2]):

$$(3.11) \qquad \limsup_{n\to\infty} \frac{\mathbb{E}_{x_n}(\|X(V)\|)}{n} \le 1 - \varepsilon, \qquad \limsup_{n\to\infty} \frac{\mathbb{E}_{x_n}(V)}{n} \le K,$$

where $K > 1$ and $\varepsilon > 0$ are constants independent of the sequence (x_n).

Transience. In this setting we suppose that the initial distribution of $(L(t))$ is given by $R_{y,w,0}$. Using the notation of Proposition 3.4, let U_1 be the first time when all the initial items a' have left the queue, the items b' in the bin have been served and the state of the bin is $[b', N - (b' + 1)$ items $1]$. We denote by $U_{p+1} = U_p + U_1 \circ \theta_{U_p}$ the first time when all the items a' present at time U_p have left the queue and the state of the bin is $[b', N - (b' + 1)$ items $1]$. It is clear that the sequence $(L(U_p))$ is a homogeneous irreducible Markov chain on $\mathcal{T}^{(\mathbb{N})}$. We suppose that the Markov process $(X(t))$ is recurrent and it visits the state $\xi_0 = (\emptyset, [b', N - (b' + 1)$ items $1])$ infinitely often, in other words, with probability 1 the queue will be empty and the state of the bin will be $[b', N - (b' + 1)$ items $1]$. The first time the process $(X(t))$ visits the state ξ_0 is necessarily at one of the times U_p, $p \ge 1$. Consequently, the Markov chain $(L(U_p))$ visits the state ξ_0 with probability 1.

Now we define a Lyapunov function on the state space of $(L(U_p))$ by

$$f(l) = \log(1 + \alpha p + \beta(\|l\| - p)) \quad \text{if } l = (l_i), \ p = \inf\{k - 1 : l_k \ne a'\}.$$

With the notations above, we have

$$f(L(U_1)) = \log(1 + \alpha Y_{y,w} + \beta W_{y,w}).$$

Now by using Proposition 3.4 and Lebesgue's Theorem, we get

$$\mathbb{E}_{R_{y,w,0}}(|f(L(U_1)) - f(L(U_0))|^2) \leq n_1 \log(\alpha \vee \beta)^2$$
$$+ n_2 \mathbb{E}_{R_{y,w,0}}\left(\log^2\left(\frac{1 + Y_{y,w} + W_{y,w}}{1 + \alpha y + \beta w}\right)\right), \quad n_1, n_2 \in \mathbb{N}.$$

Hence

$$\sup_{y,w;y+w>k} \mathbb{E}_{R_{y,w,0}}(|f(L(U_1)) - f(L(U_0))|^2) < \infty.$$

These two inequalities imply that our Markov chain $(L(U_p))$ is transient (see Lamperti [3]). This contradicts our assumption on the recurrence of $(X(t))$. ∎

EXAMPLE (The case of symmetrical distributions). The distribution F is symmetrical if $F(dx) = p\delta_1 + (1 - 2p)\delta_{a'} + p\delta_{b'}$ for $p \in [0, 1/2]$. According to Theorem 3.7 the critical value of λ for the First Fit algorithm is given by

$$\lambda_{\mathrm{FF}} = \frac{1}{2(-1+2p)p(N-b')}[(-N^2 + 2N^2p - 2Np + b'N - 2Nb'p + b'p)$$
$$+ (b'^2p^2 - 4Nb'p^2 - 2n^3b' - 8N^3b'p^2 + 8N^3b'p - 4N^2b'p^2 + 2N^2b'p + b'^2N^2$$
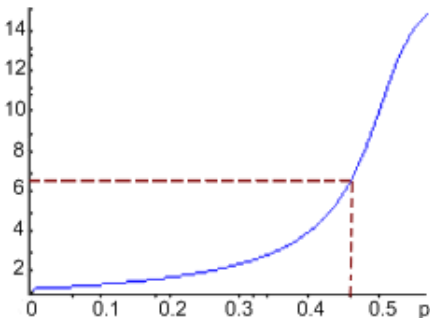$$+ 4b'^2N^2p^2 - 4b'^2N^2p + 4N^2p^2 + N^4 - 4N^4p + 4N^4p^2 - 2b'^2Np + 4b'^2Np^2)^{1/2}].$$

Now we give some examples of the effective bandwidth of First Fit policies for symmetrical distributions on $\{1, 2, b'\}$.
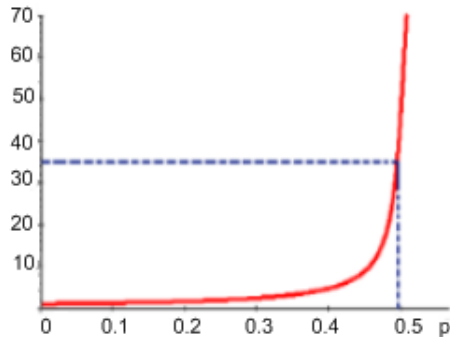
- For $C = 13$, $b' = 5$,
$$\lambda_{\mathrm{FF}} = \frac{1}{16}\frac{-104 + 187p + \sqrt{10816 - 42224p + 41625p^2}}{(-1+2p)p}.$$

- For $C = 68$, $b' = 34$,
$$\lambda_{\mathrm{FF}} = \frac{1}{2}\frac{-68 + 133p + \sqrt{4624 - 18360p + 18233p^2}}{(-1+2p)p}.$$

The effective bandwidth of the First Fit policy for symmetrical distribution on $\{1, 2, 5\}$ and $C = 13$.

The effective bandwidth of the First Fit policy for symmetrical distribution on $\{1, 2, 34\}$ and $C = 68$.

References

[1]   J.-F. Dantzer, M. Haddani and P. Robert, *On the stability of a bandwidth packing algorithm*, Probab. Engrg. Inform. Sci. 41 (2000), 57–79.
[2]   J.-F. Dantzer and P. Robert, *Fluid limits of string valued Markov processes*, Ann. Appl. Probab. 12 (2002), 860–889.
[3]   J. Lamperti, *Criteria for the recurrence or transience of stochastic process. I*, J. Math. Anal. Appl. 1 (1960), 314–330.
[4]   P. Robert, *Smooth initial distributions and fluid limits for multi-class queueing systems*, in preparation.

Amina Angelika Bouchentouf
Department of Mathematics
University Djillali Liabes of Sidi Bel Abbes
B.P. 89, Sidi Bel Abbes 22000, Algeria
E-mail: bouchentouf_ amina@yahoo.fr