Queueing Systems 48, 339–363, 2004

Stochastic Bandwidth Packing Process: Stability Conditions via Lyapunov Function Technique

DAVID GAMARNIK gamarnik@watson.ibm.com
T.J. Watson Research Center, IBM, Yorktown Heights, NY 10598, USA

Received 11 February 2003; Revised 2 April 2004

Abstract. We consider the following stochastic bandwidth packing process: the requests for communication bandwidth of different sizes arrive at times $t = 0, 1, 2, \ldots$ and are allocated to a communication link using "largest first" rule. Each request takes a unit time to complete. The unallocated requests form queues. Coffman and Stolyar [6] introduced this system and posed the following question: under which conditions do the expected queue lengths remain bounded over time (queueing system is stable)? We derive exact constructive conditions for the stability of this system using the Lyapunov function technique. The result holds under fairly general assumptions on the distribution of the arrival processes.

Keywords: bin packing, queueing networks, positive recurrence

AMS subject classification: 60C05, 60G10, 60G50, 60J20, 60K25, 90B18, 90B36

1. Introduction

We consider the following model of processing online bandwidth requests. The requests arrive over time and allocated to a communication link. Requests have different bandwidth sizes and at any time the total allocated bandwidth should not exceed the maximum bandwidth of the link. The allocation process is done according to some scheduling policy. In this paper we consider the *Best Fit* scheduling policy. This is a very simple rule according to which first the largest in size bandwidth requests are allocated, then the next largest are allocated, and so on. This is done until no available bandwidth requests can fit the remaining capacity of the link or the link is full. Unallocated requests form queues. The focus of the present paper is the long term dynamics of these queues.

Before we provide a formal description of the bandwidth allocation model we mention another application which can be addressed by the model considered in the present paper. It is memory allocation in a multiprocessor system. Here the requested bandwidths correspond to memory requests and link capacity becomes the total memory capacity of the multiprocessor. This model was considered by Kipnis and Robert [14] under the First-In-First-Out scheduling policy. This policy is simpler to analyze than the Best Fit policy and the authors obtained a complete probabilistic distributional solution of this system. See also [4], where a very similar model of slotted communication systems is considered.

In the following subsection we introduce a stochastic online bin packing process as a queueing model of our online bandwidth allocation process.

1.1. *Model description and related work*

The bandwidth allocation process was modelled by Coffman and Stolyar [6] as the following stochastic online bin packing process. Item (bandwidth) sizes $1 > a_1 > a_2 > \cdots > a_N > 0$ are fixed. For each $i = 1, 2, \dots, N$ and $t = 0, 1, 2, \dots$, one or several items of size a_i (possibly none) arrive at time t. The number of size a_i items arriving at time t is denoted by $A_i(t)$, which has a discrete probability distribution $\Pr\{A_i(t) = l\} = p_l^i, l = 1, 2, \dots, L, \sum_{l=0}^{L} p_l^i = 1$. Here L is a fixed positive integer. We assume that $A_i(t)$ is independent from $A_{i'}(t)$ and from $A_i(t')$ for $i' \neq i$ and $t' \neq t$. The independence between $A_i(t)$ and $A_{i'}(t)$ for $i \neq i'$ is not something one necessarily expects to see in real systems, but is required here for a certain recursive argument to go through. It is conceivable that this condition may be relaxed. We also assume $L < \infty$ for this paper, although we suspect that this condition could be replaced by an assumption that the distributions of $A_i(t)$ has an infinite support and finite exponential moment generating function. The arrived items are packed into a bin (are allocated to a link). The size of the bin is normalized to be equal to one. The bins also arrive at integer times $t = 0, 1, 2, \dots$, one at a time. The bin that arrived at time t represents the communication link at time t. At each time instance t, the items are packed into the bin according to the Best Fit packing algorithm: largest first. That is, first size a_1 items are placed until the remaining capacity is less than a_1 or no more size a_1 items are available. Then size a_2 items are placed, and so on. This is done until no available item can fit the residual capacity or the bin is fully packed. At this time the bin departs the system before time $t + 1$, and at time $t + 1$ a new empty bin arrives. The items that could not fit the residual capacity of the bin form queues. In particular, N queues are formed by items with sizes $a_1, a_2, \dots, a_N$. We denote by $\lambda_i = \sum_{l=0}^{L} l p_l^i$ the arrival rate of the process i and by $Q_i(t)$ the queue length of the ith process at time $t = 0, 1, 2, \dots$.

The following question was posed in [6]: under what conditions on the item sizes and the arrival processes does the total expected queue length remains bounded as a function of time? In other words when is the queue length process stable? It is proven in [6] under a fairly general assumptions that if $a_i = i/N$, $i = 1, 2, \dots, N - 1$, and if the rates λ_i of the arrival processes satisfy

$$\lambda_i = \lambda_{N-i}, \tag{1}$$

then the system is stable if and only if $\sum_i \lambda_i a_i < 1$. It is fairly easy to prove that, in general, the load condition

$$\sum_i \lambda_i a_i < 1 \tag{2}$$

is a necessary condition for stability for every arrival process distribution (satisfying some mild technical assumptions) and every sequence $a_1, a_2, \dots, a_N$. The main technical part of [6] was to prove sufficiency which is done using the fluid model technique. It is known that the condition (2) is not sufficient for stability (an example is demonstrated in [6]) and exact conditions for stability has been an open question prior to this work.

The model considered in [6] is motivated by earlier works on stochastic packing problems like random interval packing problem and some other variations of the stochastic bin packing problem. There exists by now a vast research literature on both of these topics starting as early as in 1958 by Reiny [18], who considered the first version of a random interval packing problem. A queueing version of an interval packing model was considered by Coffman et al. [5]. Recently Dantzer et al. [7] and Dantzer and Robert [8] considered exactly the model of the present paper under the First Fit packing policy. A stability characterization is obtained for the cases $N = 2$, $N = 3$ in [7,8] respectively. The case $N = 3$ is analyzed only under a specific assumption $a_1 = 3/4$, $a_2 = 2/4$, $a_3 = 1/4$.

A different line of research leading to the model of this paper starts with the classical off-line bin packing problem. It is a problem of packing a given set of items of various sizes into the smallest number of bins of a fixed size, assuming that all of the items and bins are provided initially. This is a combinatorial optimization problem well-known to be NP-complete, although Best Fit and First Fit heuristics are known to provide a constant factor approximation of the optimal value of this problem [11]. Later these algorithms were analyzed in stochastic models using the notion of the *expected waste*. In these models exactly one item of some random size arrives at each time $t = 0, 1, 2, \ldots$ and is packed into a bin, an infinite amount of which is available at time zero. The bins leave the system only when they are fully filled. The waste at time t is the residual capacity, or the sum of available capacities in partially filled bins. The goal is to understand the behavior of the expected waste as function of t. Shor [19] showed that when the arriving items are uniformly distributed over the real interval (0, 1), and the First Fit rule is used, the expected waste grows as $\Theta(t^{2/3})$. Shor [19], and Leighton and Shor [15] showed that, under the same distributional assumption, the Best Fit algorithm leads to waste $\Theta(t^{1/2}\log^{3/4}(t))$. Coffman et al. [3] showed that if the items are uniformly distributed over the set $1/k, 2/k, \ldots, j/k$ (denote this distribution by $U(j, k)$) and $j = k$ or $j = k - 1$, then the expected waste grows as $\Theta(tk^{1/2})$ and $\Theta(t^{1/2}\log k)$ for the First Fit and Best Fit packing schemes respectively. On the other hand, Kenyon et al. [13] showed that when $j = k - 2$, and the Best Fit scheme is used, the expected waste is a bounded function of t. The analysis of the latter work uses the Lyapunov function technique and the stability theory of Markov chains in $\mathcal{Z}_+^d$, see [9,16]. Later Albers and Mitzenmacher [1], using Random Fit packing algorithm as an intermediate analysis step, proved that First Fit also gives bounded expected waste for the distribution $U(k - 2, k)$. It is conjectured that the expected waste grows linearly with time if items are uniformly distributed over $1/k, 2/k, \ldots, j/k$ and $j = \alpha k$, $k \to \infty$, for any $\alpha < 1$. The conjecture was proven by Kenyon and Mitzenmacher [12] for $0.66 \leqslant \alpha < 2/3$.

The fact that the analysis remains to be difficult even if a uniform distribution is assumed demonstrates the complexity of the problem. The necessary and sufficient conditions for bounded expected waste under either Best Fit or First Fit algorithms are not known to the day. The difficulty is in analyzing infinite Markov chains in $\mathcal{Z}_+^d$ in which the transition probabilities depend on which coordinates are strictly positive and which are zero. It was established by the author [10] that, in general, stability of such random walks in $\mathcal{Z}_+^d$ is not a decidable property. That is, no algorithm can exist which checks stability of such Markov chains. In light of this result, it is not even clear whether for the bin packing models we consider in this paper, the stability checking algorithm exists. As the current paper demonstrates, stability is decidable when the Best Fit scheduling policy is used.

1.2. Results

We establish constructive necessary and sufficient conditions for stability of the Best Fit scheduling algorithm for the queueing bin packing model described in the previous subsection. The core of the analysis is a certain computable quantity called *effective drift*. We use the method of Lyapunov function for computing effective drifts. Below we describe at a high level the main ideas of this paper. Our first observation is that the queue length processes corresponding to the first $i \leqslant N-1$ arrival processes are conditionally independent from the remaining $N-i$ queue length processes, when Best Fit algorithm is used. For example, the first queue length process $Q_1(t)$ corresponding to items of size a_1 is a simple one dimensional random walk $Q_1(t+1) = Q_1(t)+A_1(t)-\min\{Q_1(t), f_1\}$. In particular, the processes corresponding to other item types do not affect the first process. This one dimensional random walk is stable if and only $E[A_1(t)] - f_1 \equiv -\gamma_1 < 0$. The behavior of the second queue length process is tightly related to the first one. At each time t we pack as many items of size a_2 as can fit *after* size a_1 items are packed to maximum. The latter depends solely on $Q_1(t)$. Thus the behavior of the second process is also a random walk for which the jump distribution is itself a function of the random state of the first random walk $Q_1(t)$. The following result, proven in this paper, is then intuitively clear: the queue length process $Q_2(t)$ is stable if and only if

$$E\big[A_2(t)\big] - \sum_{k=1}^{\infty} f_2(k)\mathrm{Pr}_{\pi_1}\big\{Q_1(t) = k\big\} \equiv -\gamma_2 < 0, \tag{3}$$

where $f_2(k)$ is the maximum number of type a_2 items that can fit into a bin when the first queue has exactly size k, and $\mathrm{Pr}_{\pi_1}\{\cdot\}$ is the stationary distribution of the first queue length process. In other words, the second queue length is stable if and only if on average (with respect to the stationary distribution of the first process and the arrival process $A_2(t)$) the jumps of the second process are negative. This result is generalized to the remaining queue length processes. The quantity on the left-hand side of (3) and its analogues for $Q_i(t)$, $i \geqslant 3$, is defined to be the *effective drift* of the corresponding process. We prove that the system is stable if and only if the effective drift $-\gamma_i$ corresponding to each of the queue length process is negative ($\gamma_i > 0$).

Our subsequent goal is constructing an algorithm for computing the stationary distribution π_i of the first i queue length processes (provided they are stable), so that the next effective drift $-\gamma_{i+1}$ can be computed. Specifically, we will show that γ_{i+1} can be computed to within an arbitrary accuracy. This will provide us with stability checking algorithm for an arbitrary system, except for the critical case $\gamma_i = 0$ for some $i = 2, \ldots, N$, which corresponds to an unstable (but not detectable by our algorithm) case. Our algorithm is based on constructing a linear Lyapunov function on the queue length vector process and using powerful results of Meyn and Tweedie [17] establishing exponential mixing rate of infinite Markov chains allowing a suitable Lyapunov function. As a corollary of our construction, we establish that the stability of the online bin packing queueing model considered in this paper is always witnessed by some constructive linear Lyapunov function.

Although our algorithm is guaranteed to terminate in finite time, except for the critical case, the computation time grows very rapidly in N. One of the reasons for rapid growth of the computation effort is poor bounds on mixing rates in [17]. Reducing the complexity of the stability checking algorithm remains an interesting open question. We note also that, while our model is similar to the expected waste model described before, our results do not seem to lead to any immediate results for the latter model, although the techniques used could be found applicable.

The remainder of the paper is structured as follows. In the following section we introduce the notion of the fit function, which is central for the steady-state analysis of the system. In section 3 we provide some preliminary technical results. Specifically, we introduce Lyapunov functions and mention its relevance to the analysis of stationary distribution and mixing rates of infinite Markov chains. In section 4 we introduce the notion of the effective drift and we formulate the exact stability conditions in terms of the effective drift. Section 5 contains the most technical part of the paper. We obtain a stability checking algorithm which is based on a linear Lyapunov function. Section 6 contains some concluding remarks. The proofs of some technical auxiliary results of the paper is delayed till appendix.

2. A stochastic bin packing queueing model and a fit function

Observe that the queue length process $\overline{Q}(t) \equiv (Q_1(t), \ldots, Q_N(t))$ is a Markov chain, when the Best Fit scheduling policy is used – the state $\overline{Q}(t+1)$ depends only on the state $\overline{Q}(t)$, and is independent from $\overline{Q}(t')$ for $t' < t$, for all t. Moreover, the truncated process $\overline{Q}_i(t) \equiv (Q_1(t), \ldots, Q_i(t))$ is also a Markov chain since, by the Best Fit rule, it is independent from the processes $Q_j(t)$, $j > i$. We define the process $\overline{Q}(t)$ to be stable if $\sup_t E[\sum_{i=1}^N Q_i(t)] < \infty$. That is, the expected total queue length remains uniformly bounded. The stability of the partial processes $\overline{Q}_i(t)$ is defined similarly. The stability implies the existence of a stationary distribution [16]. We assume throughout the paper that the Markov chain $\overline{Q}(t)$ is irreducible and aperiodic. A simple way to ensure that would be to have a very large L and assume $p_0^i > 0$ for all $1 \leqslant i \leqslant N$, $0 \leqslant l \leqslant L$, as the next lemma states.

Lemma 1. Suppose $p_0^i > 0$ for all $1 \leqslant i \leqslant N$. Then the chain $Q(t)$ is irreducible aperiodic. Moreover, if $p_l^i > 0$ for all $1 \leqslant i \leqslant N$, $0 \leqslant l \leqslant L$ and $L > 1/a_N$, then for every $q_0, q \in \mathcal{Z}_+^N$ there exists t such that $\Pr\{\overline{Q}(t) = q | \overline{Q}(0) = q_0\} > 0$.

Proof. See appendix. □

From now on we assume $L > 1/a_N$ and $p_l^i > 0$ for all $1 \leqslant i \leqslant N, 0 \leqslant l \leqslant L$. Under these assumptions, the stability implies the existence of the unique stationary distribution π such that $\Pr_\pi\{\overline{Q}(t) = q\} > 0$ for every $q \in \mathcal{Z}_+^N$. Here and below $\Pr_\pi\{\cdot\}$ and $E_\pi[\cdot]$ stand for the probability distribution and the expectation operator with respect to the measure π.

We now introduce a *fit function* – a concept critical to our analysis. The first fit function f_1 is a constant $f_1 = \lfloor 1/a_1 \rfloor$. That is f_1 is simply a maximal number of size a_1 items that can fit into a single bin. For each index $i = 2, \ldots, N$ the fit function $f_i : \mathcal{Z}_+^{i-1} \to \mathcal{Z}_+$ is defined as

$$f_i(q_1, \ldots, q_{i-1}) = \max\bigl\{n\colon \min\{q_1, f_1\}a_1 + \min\bigl\{q_2, f_2(q_1)\bigr\}a_2 + \cdots + \min\bigl\{q_{i-1}, f_{i-1}(q_1, \ldots, q_{i-2})\bigr\}a_{i-1} + na_i \leqslant 1\bigr\}. \tag{4}$$

The fit function has the following physical interpretation. If at time t there are q_i items with size a_i then Best Fit algorithm will place $\min\{q_1, f_1\}$ items of size a_1. This leaves $1 - \min\{q_1, f_1\}a_1$ free capacity of which $\max\{n\colon na_2 \leqslant 1 - \min\{q_1, f_1\}a_1\}$ can be used to place size a_2 items. Note that this maximum is a function of q_1 only (the rest are the parameters of the model). This function is defined to be $f_2(q_1)$. The interpretation for $f_i(\cdot), i \geqslant 3$, is similar. In other, words $f_i(q_1, \ldots, q_{i-1})$ is the maximal number of size a_i items that will be placed into a bin by the Best Fit algorithm, when there are q_j items with size a_j waiting in the queue, $j = 1, 2, \ldots, i-1$. Using the fit function notation we obtain $Q_i(t+1) = Q_i(t) + A_i(t) - \min\{Q_i(t), f_i(Q_1(t), \ldots, Q_{i-1}(t))\}$.

3. Preliminary technical results

In this section we present several results from a general theory of infinite Markov chains and Lyapunov functions. These results will be our main instruments in the remainder of the paper.

Consider a discrete time Markov chain $Q(t)$ with a countable state space $\mathcal{X}$. Denote the transition probabilities by $p(x, y) = \Pr\{Q(t+1) = y \mid Q(t) = x\}$. A probability distribution $\pi : \mathcal{X} \to [0, 1]$ is defined to be stationary if $\pi(x) = \sum_y p(y, x)\pi(y)$ for every state $x \in \mathcal{X}$. We say that the Markov chain is positive recurrent or stable if there exist at least one stationary distribution. If the chain is stable and irreducible then there exists the unique stationary distribution.

Definition 1. A nonnegative functions $\Phi : \mathcal{X} \to \Re_+$ is defined to be a Lyapunov function with drift $-\gamma < 0$ and exception set $\mathcal{B} \subset \mathcal{X}$, if $|\mathcal{B}| < \infty$ and

$$E\bigl[\Phi\bigl(Q(t+1)\bigr) \mid Q(t) = x\bigr] - \Phi(x) = \sum_y \Phi(y)p(x, y) - \Phi(x) \leqslant -\gamma \tag{5}$$

for all $x \notin \mathcal{B}$. We put $B = B^{\Phi} = \max\{\Phi(x)\colon x \in \mathcal{B}\}$ and call it the exception parameter. Also a nonnegative function $\Phi^{\mathrm{g}} : \mathcal{X} \to [1, +\infty)$ is defined to be a geometric Lyapunov function with geometric drift $0 < \gamma^{\mathrm{g}} < 1$ and exception set $\mathcal{B} \subset \mathcal{X}$, if $|\mathcal{B}| < \infty$ and

$$E\big[\Phi^{\mathrm{g}}\big(Q(t+1)\big) \mid Q(t) = x\big] \leqslant \gamma^{\mathrm{g}}\Phi^{\mathrm{g}}(x) \tag{6}$$

for all $x \notin \mathcal{B}$.

We say that a Lyapunov function has bounded jumps if

$$\nu \equiv \max\big\{\big|\Phi(y) - \Phi(x)\big|\colon\ p(x, y) > 0,\ x, y \in \mathcal{X}\big\} < \infty. \tag{7}$$

A geometric Lyapunov function is defined to have bounded jumps if

$$\nu_{\mathrm{g}}^{\Phi} \equiv \max\left\{\frac{\Phi(y)}{\Phi(x)}\colon\ p(x, y) > 0,\ x, y \in \mathcal{X}\right\} < \infty. \tag{8}$$

The existence of a Lyapunov function with bounded jumps guarantees the existence of a stationary distribution [9,16]. Several technical results on Lyapunov functions are included below. The proof for the following result can be found in [2].

Theorem 2 [2, theorem 1]. Let Φ be a Lyapunov function with a drift $-\gamma$, exception parameter B and maximal jump ν. Then every stationary distribution π satisfies the following geometric bounds

$$\Pr_{\pi}\big\{\Phi\big(Q(t)\big) \geqslant B + 2\nu m\big\} \equiv \sum_{x:\Phi(x) \geqslant B + 2\nu m} \pi(x) \leqslant \left(\frac{\nu}{\nu + \gamma}\right)^{m+1}. \tag{9}$$

for all $m = 0, 1, 2, \ldots,$ and

$$E_{\pi}\big[\Phi\big(Q(t)\big)\big] \equiv \sum_{x} \Phi(x)\pi(x) \leqslant B + \frac{2\nu^2}{\gamma}. \tag{10}$$

The next technical result required for our analysis is exponential mixing rates for infinite Markov chains, for which a geometric Lyapunov function exists. This is a powerful result established by Meyn and Tweedie [17]. It holds for discrete time Markov chains in discrete or continuous state spaces, where, in the case of the continuous state space so called *small sets* (typically compact sets) play the role of exception sets. We adopt here a simpler version applicable for our Markov chains with countably many states.

Theorem 3 [17, theorem 2.3]. Given an irreducible Markov chain $Q(t)$, suppose Φ^{g} is a geometric Lyapunov function with a geometric drift γ^{g} and the exception set $\mathcal{B}$. Suppose also π is the unique stationary distribution. Then, there exist constants $R > 0$,

$0 < \rho < 1$ such that for any state $x \in \mathcal{X}$ and any function $\phi : \mathcal{X} \to \Re$ satisfying $\phi(x) \leqslant \Phi^{\mathrm{g}}(x), \forall\, x \in \mathcal{X}$, the following bound holds

$$\left| \sum_{y \in \mathcal{X}} \phi(y) \big(\Pr\big\{ Q(t) = y \mid Q(0) = x \big\} - \pi(y) \big) \right| \leqslant \Phi^{\mathrm{g}}(x) R \rho^{t}, \tag{11}$$

where the constants R, ρ are computable functions which depend on $\gamma^{\mathrm{g}}, \nu_{\mathrm{g}}^{\Phi}$, $\max_{x \in \mathcal{B}} \Phi^{\mathrm{g}}(x)$ and

$$p_{\min}^{\mathcal{B}} \equiv \min_{x, y \in \mathcal{B}} p(x, y). \tag{12}$$

Remark. Exact formulas for computing R, ρ are provided in [17]. They are quite lengthy and we do not repeat them here. These formulas give meaningful bounds only in case $0 < \gamma^{\mathrm{g}} < 1$; $\nu_{\mathrm{g}}^{\Phi}, \max_{x \in \mathcal{B}} \Phi^{\mathrm{g}}(x) < \infty$; $p_{\min}^{\mathcal{B}} > 0$. We will make sure that this is the case in all the instances where we use this theorem.

Note the dependence of the mixing time on the initial state x. Such dependence is natural in infinite Markov chains since one can make mixing time arbitrarily large by selecting the initial state very far from the equilibrium. Unfortunately, the constants R, ρ depend exponentially on $|\mathcal{B}|$ which makes computation effort grow rapidly with the size of the problem. The importance of this result, however, is the possibility of computing mixing rates solely in terms of basic parameters of the Lyapunov function.

We close this section with additional technical results that we use in the main proof.

Lemma 4. Given a Markov chain $Q(t)$ in $\mathcal{Z}_{+}^{d}$, suppose $\Phi(Q(\cdot)) \equiv \sum_{j=1}^{d} w_j Q_j(\cdot)$ is a linear Lyapunov function with drift $-\gamma$, exception parameter $B < \infty$ and maximum jump $\nu < \infty$. Then, for all $s \leqslant \min\{1/\nu, \gamma/(\nu^2 \mathrm{e})\}$, there exists computable $0 < \gamma^{\mathrm{g}} < 1$, which depends on the parameters of the model and on γ, B, ν, w_j, $j = 1, 2, \ldots, d$, such that $\Phi^{\mathrm{g}}(Q(\cdot)) \equiv \exp(s \sum_{j=1}^{d} w_j Q_j(\cdot))$ is a geometric Lyapunov function with geometric drift γ^{g}. The exception parameter of this Lyapunov function can be taken to be $B^{\mathrm{g}} \equiv \exp(sB)$.

Proof. See appendix. □

Lemma 5. Under the assumptions of lemma 4, for any $\tau \geqslant 0$, the following bound holds

$$E\left[\sum_{j=1}^{d} w_j Q_j(\tau) \mid Q(0) \right] - \sum_{j=1}^{d} w_j Q_j(0) \leqslant B + \frac{\nu \exp(s\nu)}{\gamma^{\mathrm{g}} (1 - \gamma^{\mathrm{g}})^2},$$

for any state $\overline{Q}(0) = (Q_1(0), \ldots, Q_d(0))$, where s and γ^{g} are defined as in lemma 4.

Proof. See appendix. □

Lemma 6. Under the assumption of lemma 4, for any

$$\tau \geqslant \frac{B + \nu \exp(s\nu)/(\gamma^{g}(1-\gamma^{g})^2) + 1 + \gamma}{\gamma},$$

the function $\Phi(Q(\cdot)) \equiv \sum_{j=1}^{d} w_j Q_j(\cdot)$ is also a linear Lyapunov function for the Markov (sub)chain $Q(\tau t), t = 0, 1, \ldots,$ with drift $-\gamma' = -1$, maximum jump $\nu' \leqslant \nu\tau$, and the exception parameter

$$B' = B + \frac{\nu(B + \nu \exp(s\nu)/(\gamma^{g}(1-\gamma^{g})^2) + 1 + \gamma)}{\gamma},$$

where s and γ^{g} are defined as in lemma 5.

Proof. See appendix. □

4. Effective drift and the stability conditions

In this subsection we establish necessary and sufficient conditions for stability of our queueing bin packing system, using the notion of a fit function.

Theorem 7. Consider a stochastic bin packing queueing process with item sizes $1 > a_1 > a_2 > \cdots > a_N > 0$ and arrival probabilities $p_l^i > 0$, $1 \leqslant i \leqslant N$, $0 \leqslant l \leqslant L$. The following holds:

1. The first queue length process $Q_1(t)$ is stable if and only if
$$-\gamma_1 \equiv \lambda_1 - f_1 < 0. \tag{13}$$
2. Suppose $1 \leqslant i \leqslant N-1$ is such that the process $\overline{Q}_i(t) \equiv (Q_1(t), \ldots, Q_i(t))$ is stable with the unique stationary distribution π_i. Then the process $Q_{i+1}(t)$ is stable if and only if
$$-\gamma_{i+1} \equiv \lambda_{i+1} - \sum_{q \in \mathcal{Z}_+^i} f_{i+1}(q)\pi_i(q) < 0. \tag{14}$$

The quantity $-\gamma_i$ is defined to be the *effective drift* of the ith process. The queueing system is then stable if and only if the effective drift of every process $Q_i(t), i = 1, 2, \ldots, N$, is negative.

Remark. This result has a simple intuitive explanation. The first part is simply stability condition for a one-dimensional random walk with bounded jumps – the expected change of the walk (drift) must be negative. For the second part, note that $f_{i+1}(q)$ is the maximal number of size a_{i+1} items that Best Fit algorithm will place into a bin when $\overline{Q}_i(t) = q$. $Q_{i+1}(t)$ can be viewed as a one-dimensional random walk with transition probabilities depending on the state of the Markov chain $\overline{Q}_i(t)$. This random walk is stable if and only if its average (with respect to the chain $\overline{Q}_i(t)$) drift $-\gamma_{i+1}$ is negative.

Proof of theorem 7. The sufficiency of conditions (13), (14) will be established in theorem 8 below which constructs algorithmically a linear Lyapunov function on the process $\overline{Q}_{i+1}(t)$ whenever the process $\overline{Q}_i(t)$ is stable and $\gamma_{i+1} > 0$. We now prove the necessity.

We start with the case $i = 1$. If the condition (13) is violated then we obtain a one-dimensional random walk with zero or positive drift $-\gamma_1$ (that is, $\gamma_1 \leqslant 0$). Since, $p_l^1 > 0, l = 0, 1, \dots, L$, and as a result $\mathrm{Var}(A_1(t)) > 0$, then the variance of the step of this random walk is positive. It follows that $E[Q_1(t)] \to \infty$ as $t \to \infty$.

Suppose now that the chain $\overline{Q}_i(t)$ is stable, but $\gamma_{i+1} \leqslant 0$. For the purposes of contradiction assume that the Markov chain $\overline{Q}_{i+1}(t)$ is stable with the unique stationary distribution π_{i+1} and $E_{\pi_{i+1}}[Q_{i+1}(t)] < \infty$. By stationarity,

$$E_{\pi_{i+1}}\big[Q_{i+1}(t+1)\big] - E_{\pi_{i+1}}\big[Q_{i+1}(t)\big] = 0. \tag{15}$$

On the other hand,

$$E_{\pi_{i+1}}\big[Q_{i+1}(t+1)\big] - E_{\pi_{i+1}}\big[Q_{i+1}(t)\big] = E\big[A_{i+1}(t)\big] - E_{\pi_{i+1}}\big[\min\big\{Q_{i+1}(t), f_{i+1}\big(\overline{Q}_i(t)\big)\big\}\big]. \tag{16}$$

Note $E[A_{i+1}(t)] = \lambda_{i+1}$. Also

$$\begin{aligned}
&E_{\pi_{i+1}}\big[\min\big\{Q_{i+1}(t), f_{i+1}\big(\overline{Q}_i(t)\big)\big\}\big] \\
&\quad = \sum_{k=0}^{\infty} \sum_{q \in \mathcal{Z}_+^i} \min\big\{k, f_{i+1}(q)\big\} \mathrm{Pr}_{\pi_{i+1}}\big\{Q_{i+1}(t) = k \mid \overline{Q}_i(t) = q\big\}\pi_i(q) \\
&\quad \leqslant \sum_{k=1}^{\infty} \sum_{q \in \mathcal{Z}_+^i} f_{i+1}(q) \mathrm{Pr}_{\pi_{i+1}}\big\{Q_{i+1}(t) = k \mid \overline{Q}_i(t) = q\big\}\pi_i(q) \\
&\quad = \sum_{q \in \mathcal{Z}_+^i} f_{i+1}(q)\pi_i(q) \mathrm{Pr}_{\pi_{i+1}}\big\{Q_{i+1}(t) > 0 \mid \overline{Q}_i(t) = q\big\}.
\end{aligned} \tag{17}$$

Note, by irreducibility, that for every state $(q, k) \in \mathcal{Z}_+^{i+1}$ we have $\mathrm{Pr}_{\pi_{i+1}}\{\overline{Q}_i(t) = q, Q_{i+1}(t) = k\} > 0$. Then for every $q \in \mathcal{Z}_+^i$

$$\mathrm{Pr}_{\pi_{i+1}}\big\{Q_{i+1}(t) = 0 \mid \overline{Q}_i(t) = q\big\} = \frac{\mathrm{Pr}_{\pi_{i+1}}\{Q_{i+1}(t) = 0, \overline{Q}_i(t) = q\}}{\pi_i(q)} > 0.$$

As a result

$$\begin{aligned}
&\mathrm{Pr}_{\pi_{i+1}}\big\{Q_{i+1}(t) > 0 \mid \overline{Q}_i(t) = q\big\}\pi_i(q) \\
&\quad = \big(1 - \mathrm{Pr}_{\pi_{i+1}}\big\{Q_{i+1}(t) = 0 \mid \overline{Q}_i(t) = q\big\}\big)\pi_i(q) < \pi_i(q).
\end{aligned}$$

It follows that the expression in (17) is strictly smaller than

$$\sum_{q \in \mathcal{Z}_+^i} f_{i+1}(q)\pi_i(q).$$

Combining with (16), we obtain that

$$E_{\pi_{i+1}}\big[Q_{i+1}(t+1)\big] - E_{\pi_{i+1}}\big[Q_{i+1}(t)\big] > \lambda_{i+1} - \sum_{q\in\mathcal{Z}_+^i} f_{i+1}(q)\pi_i(q) = -\gamma_{i+1} \geqslant 0.$$

But this contradicts (15).

5. Linear Lyapunov function and the stability checking algorithm

In this section we establish our main result – an algorithm for checking stability of a bin packing queueing system whenever $|\gamma_i| > 0$ for all $i = 1, 2, \dots, N$. Simultaneously we prove that whenever the system is stable, the stability is witnessed by some linear Lyapunov function. We first outline the approach and explain the reason for the condition $|\gamma_i| > 0$. We assume inductively that we can check stability of the subprocess $\overline{Q}_i(t)$, and our goal is to check stability of the process $Q_{i+1}(t)$. We assume, moreover, by induction, that we have constructed values $w_j > 0$, $j = 1, 2, \dots, i$ and $\tau_i > 0$ such that the linear function $\sum_{j=1}^{i} w_j Q_j(t)$ is a Lyapunov function for the Markov (sub)chain $\overline{Q}_i(\tau_i t)$, $t = 0, 1, 2, \dots$. We select a large constant C_i and use theorem 2 to obtain an exponentially small upper bound on $\Pr_\pi\{\sum_{j=1}^{i} w_j Q_j(t) > C_i\}$. Suitably modifying our Lyapunov function into a geometric Lyapunov function and applying theorem 3, we obtain estimates $\tilde{\pi}_i(q)$ of the stationary probability $\pi_i(q)$, for any given state $q \in \mathcal{Z}_+^i$ satisfying $\sum_{j=1}^{i} w_j q_j \leqslant C_i$, to within an arbitrary desired accuracy. This can be achieved by selecting large t and computing the transient distribution $\Pr\{\overline{Q}_i(t) = q \mid \overline{Q}_i(0) = 0\}$. The transient distribution can be computed for any fixed t since there exists only a finite number of states that can be reached in t steps from the initial state $q = 0$. We use the bound and the approximations above to obtain an approximating interval $[\tilde{\gamma}_{i+1} - \varepsilon, \tilde{\gamma}_{i+1} + \varepsilon]$ containing γ_{i+1} in (14) with an arbitrary desired accuracy $\varepsilon > 0$. If $\gamma_{i+1} > 0$, then by subsequently decreasing ε we eventually obtain an estimate interval $[\tilde{\gamma}_{i+1} - \varepsilon, \tilde{\gamma}_{i+1} + \varepsilon]$ with $\tilde{\gamma}_{i+1} - \varepsilon > 0$. In this case our algorithm detects that $\gamma_{i+1} > 0$ and the algorithm terminates with the output "stable", by virtue of the second part of theorem 7. Similarly, if $\gamma_{i+1} < 0$ we construct eventually an estimation interval with $\tilde{\gamma}_{i+1} + \varepsilon < 0$. In this case the algorithm detects that $\gamma_{i+1} < 0$ and terminates with the output "unstable". Our algorithm will never terminate if $\gamma_{i+1} = 0$ since we will always have the lower end of the estimation interval negative and the upper end positive.

To complete the proof we need to be able to extend the linear Lyapunov function to the process $\overline{Q}_{i+1}(t)$ so that the similar analysis can be used to determine the stability of the process $Q_{i+2}(t)$, when $i \leqslant N-2$. This is done by selecting a large time gap τ_{i+1} and parameters B_{i+1}, $w_{j+1} > 0$, and showing that the expected change $E[\sum_{j=1}^{i+1} w_j(Q_j(t + \tau_{i+1}) - Q_j(t))]$ is negative whenever $\sum_{j=1}^{i+1} w_j Q_j(t) > B_{i+1}$.

We now state rigorously and prove the main result.

Theorem 8. Consider a stochastic bin packing queueing process with item sizes $1 > a_1 > a_2 > \cdots > a_N > 0$ and arrival probabilities $p_l^i > 0$, $1 \leqslant i \leqslant N$, $0 \leqslant l \leqslant L$.

For $i \leqslant N-1$ suppose we are given $\tau_i, B_i > 0, w_1, w_2, \ldots, w_i > 0$ such that for any $q \in \mathcal{Z}_+^i$ satisfying $\sum_{j \leqslant i} w_j q_j > B_i$, the following holds

$$E\left[\sum_{j=1}^{i} w_j Q_j(t+\tau_i) \mid \overline{Q}_i(t) = q\right] \leqslant -1 + \sum_{j=1}^{i} w_j q_j, \tag{18}$$

and

$$\min_{q',q''} \Pr\{\overline{Q}_i(t+\tau_i) = q'' \mid \overline{Q}_i(t) = q'\} > 0, \tag{19}$$

where minimum runs over q', q'' satisfying $\sum_{j \leqslant i} w_j q'_j, \sum_{j \leqslant i} w_j q''_j \leqslant B_i$. In other words, $\Phi_i(\overline{Q}_i(t)) \equiv \sum_{j=1}^{i} w_j Q_j(t)$ is a Lyapunov function with drift -1 and the exception parameter B_i for the Markov chain $\overline{Q}_i(\tau_i t)$, $t = 0, 1, 2, \ldots$. Let π_i be the unique stationary distribution of $\overline{Q}_i(t)$ and let γ_{i+1} be defined as in (14).

Then

1. For any $\varepsilon > 0$ a value $\tilde{\gamma}_{i+1}$ can be computed such that the interval $[\tilde{\gamma}_{i+1} - \varepsilon, \tilde{\gamma}_{i+1} + \varepsilon]$ contains γ_{i+1}. The value $\tilde{\gamma}_{i+1}$ depends on τ_i, B_i, w_j and the parameters of the model. If $|\gamma_{i+1}| > 0$, then it can be determined in finite time whether $\gamma_{i+1} > 0$ or $\gamma_{i+1} < 0$.

2. If $\gamma_{i+1} > 0$, then the values $\tau_{i+1}, w_{i+1}, B_{i+1} > 0$ can be computed such that for any $q \in \mathcal{Z}_+^{i+1}$ satisfying $\sum_{j \leqslant i+1} w_j q_j > B_{i+1}$, the following holds

$$E\left[\sum_{j=1}^{i+1} w_j Q_j(t+\tau_{i+1}) \mid \overline{Q}_{i+1}(t) = q\right] \leqslant -1 + \sum_{j=1}^{i+1} w_j q_j, \tag{20}$$

   and

$$\min_{q',q''} \Pr\{\overline{Q}_{i+1}(t+\tau_{i+1}) = q'' \mid \overline{Q}_{i+1}(t) = q'\} > 0, \tag{21}$$

   where the minimum runs over q', q'' satisfying $\sum_{j \leqslant i+1} w_j q'_j, \sum_{j \leqslant i+1} w_j q''_j \leqslant B_i$. That is,

$$\Phi_{i+1}(\overline{Q}_{i+1}(t)) \equiv \sum_{j=1}^{i+1} w_j Q_j(t)$$

   is a Lyapunov function for the Markov chain $\overline{Q}_{i+1}(\tau_{i+1} t)$, $t = 0, 1, 2, \ldots$, with drift -1 and the exception parameter B_{i+1}. In particular, Markov chain $\overline{Q}_i(t)$ is stable.

The theorem above establishes a recursive computational procedure for checking stability of the system whenever $|\gamma_i| > 0$ for all $i = 1, 2, \ldots, N$. The drift of the constructed Lyapunov functions $\sum_{j \leqslant i} w_j Q_j(\cdot)$ is normalized to be -1 for convenience. It can take any negative values by rescaling the weights w_j.

Proof. One difficulty in estimating the effective drift $-\gamma_{i+1}$ in (14) is the presence of an infinite summation. We circumvent this difficulty by first observing that the values of $f_{i+1}(q)$ are in the finite range $\{0, 1, 2, \ldots, \lfloor 1/a_{i+1} \rfloor\}$, and by cutting off states $q \in \mathcal{Z}_+^i$ with the high value of $\sum_{j=1}^{i} w_j q_j$. Let $w_{\max}^i = \max_{j \leqslant i} w_j$ and $w_{\min}^i = \min_{j \leqslant i} w_j > 0$. Note that the Lyapunov function $\sum_{j=1}^{i} w_j Q_j(t)$ has the maximal jump at most $w_{\max}^i L i \tau_i$, (recall that by assumption $L > 1/a_N > 1/a_i$, for all i). Applying theorem 2 to any $C > 0$

$$\Pr_{\pi_i}\left\{\sum_{j=1}^{i} w_j Q_j(t) > B_i + 2w_{\max}^i L i \tau_i C\right\} \leqslant \left(\frac{w_{\max}^i L i \tau_i}{w_{\max}^i L i \tau_i + 1}\right)^{C+1}. \tag{22}$$

Fix $\varepsilon > 0$ and select a constant $C > 0$, large enough, so that

$$\Pr_{\pi_i}\left\{\sum_{j=1}^{i} w_j Q_j(t) > C\right\} < \frac{\varepsilon a_{i+1}}{2}. \tag{23}$$

We now estimate the stationary probability $\pi_i(q)$ for any state $q \in \mathcal{Z}_+^i$ such that $\sum_{j=1}^{i} w_j q_j \leqslant C$. By assumption of the theorem, $\sum_{j \leqslant i} w_j Q_j(\cdot)$ is a Lyapunov function for the chain $\overline{Q}_i(t\tau_i)$. Then, combining lemma 4 with theorem 3 and using the function $\phi(x)$ that is equal to 1 for $x = q$ and equal to zero otherwise, we obtain $|\Pr\{\overline{Q}_i(t) = q \mid \overline{Q}_i(0) = 0\} - \pi_i(q)| \leqslant R\rho^t$ for some computable $R > 0$, $0 < \rho < 1$. The assumption (19) ensures that the condition (12) is satisfied.

Take t large enough so that

$$R\rho^t < \frac{\varepsilon a_{i+1}}{2(C/w_{\min}^i + 1)^i} \equiv \delta \tag{24}$$

and compute $\tilde{\pi}(q) \equiv \Pr\{\overline{Q}_i(t) = q \mid \overline{Q}_i(0) = 0\}$ for every q with $\sum_{j=1}^{i} w_j q_j \leqslant C$. The obtained values satisfy

$$\left|\tilde{\pi}_i(q) - \pi_i(q)\right| < \delta. \tag{25}$$

For each q with $\sum_{j=1}^{i} w_j q_j \leqslant C$ we compute $f_{i+1}(q)$. We then compute the difference

$$\lambda_{i+1} - \sum_{q\colon \sum_{j=1}^{i} w_j q_j \leqslant C} f_{i+1}(q)\tilde{\pi}_i(q) \equiv -\tilde{\gamma}_{i+1}. \tag{26}$$

We have

$$\begin{aligned} -\gamma_{i+1} &= \lambda_{i+1} - \sum_{q} f_{i+1}(q)\pi_i(q) \\ &= \lambda_{i+1} - \sum_{q\colon \sum_{j=1}^{i} w_j q_j \leqslant C} f_{i+1}(q)\pi_i(q) - \sum_{q\colon \sum_{j=1}^{i} w_j q_j > C} f_{i+1}(q)\pi_i(q). \end{aligned}$$

Since $|\{q\colon \sum_{j=1}^{i} w_j q_j \leqslant C\}| \leqslant (C/w_{\min}^i + 1)^i$ and $f_{i+1}(q) \leqslant 1/a_{i+1}$, then, using (25)

$$\left| \sum_{q\colon \sum_{j=1}^{i} w_j q_j \leqslant C} f_{i+1}(q)\pi_i(q) - \sum_{q\colon \sum_{j=1}^{i} w_j q_j \leqslant C} f_{i+1}(q)\tilde{\pi}_i(q) \right| \leqslant \frac{\varepsilon}{2}.$$

Also, applying (23) we obtain

$$\sum_{q\colon \sum_{j=1}^{i} w_j q_j > C} f_{i+1}(q)\pi_i(q) \leqslant \frac{\varepsilon}{2}.$$

Combining with (26) we obtain

$$|\tilde{\gamma}_{i+1} - \gamma_{i+1}| \leqslant \varepsilon.$$

Now if $|\gamma_{i+1}| > 0$ then by computing $[\tilde{\gamma}_{i+1} - \varepsilon, \tilde{\gamma}_{i+1} + \varepsilon]$ for $\varepsilon = 1/2, 1/4, \ldots, 1/2^m, \ldots$ we eventually obtain an interval which lies completely in $(0, \infty)$ or $(-\infty, 0)$. The first case implies $\gamma_{i+1} > 0$, which as will be shown below implies stability, the second case implies $\gamma_{i+1} < 0$, which by theorem 7 implies instability. This completes the proof of the first part of the theorem.

We now turn to the second, more difficult part of the theorem. We first assume that the parameters $\tau_{i+1}, w_{i+1}, B_{i+1} \geqslant B_i$ are already fixed and do the analysis. Later we show that these parameters can be selected to satisfy all the requirements that result from the analysis. We will show that (20) and (21) hold when a set of to be defined constraints is satisfied. The main bulk of the proof is showing that (20) is satisfied, under a to be specified collection of constraint. The proof of (21) is delayed to the end.

Since the process $\overline{Q}_{i+1}(t)$ is Markovian, we assume without the loss of generality that $t = 0$. Select a value

$$C \in [B_i, B_{i+1}], \tag{27}$$

which is specified later. Suppose

$$\sum_{j \leqslant i+1} w_j Q_j(0) > B_{i+1}. \tag{28}$$

We consider two cases.

Case 1.

$$\sum_{j \leqslant i} w_j Q_j(0) \leqslant C. \tag{29}$$

This implies $w_{i+1} Q_{i+1}(0) > B_{i+1} - C$. We specify the following constraint on B_{i+1}, τ_{i+1}:

$$B_{i+1} - C \geqslant \frac{w_{i+1}(\tau_{i+1} + 1)}{a_{i+1}}. \tag{30}$$

Since $Q_{i+1}(t+1) \geqslant Q_{i+1}(t) - (1/a_{i+1})$, then this constraint insures that $Q_{i+1}(t) \geqslant 1/a_{i+1}$ for all $t \leqslant \tau_{i+1}$. In particular, for every such t, the number of type $i+1$ items that leave the system is exactly $f_{i+1}(\overline{Q}_i(t))$. We now estimate $Q_{i+1}(\tau_{i+1})$. We have

$$\begin{aligned}
&E\big[Q_{i+1}(\tau_{i+1}) \mid \overline{Q}_{i+1}(0)\big]\\
&\quad = Q_{i+1}(0) + \sum_{t=1}^{\tau_{i+1}} E\big[A_{i+1}(t)\big] - \sum_{t=0}^{\tau_{i+1}-1} \sum_{q\in\mathcal{Z}_+^i} f_{i+1}(q)\Pr\big\{\overline{Q}_i(t) = q \mid \overline{Q}_{i+1}(0)\big\}\\
&\quad = Q_{i+1}(0) + \lambda_{i+1}\tau_{i+1} - \sum_{t=0}^{\tau_{i+1}-1} \sum_{q\in\mathcal{Z}_+^i} f_{i+1}(q)\pi_i(q)\\
&\qquad - \sum_{t=0}^{\tau_{i+1}-1} \sum_{q\in\mathcal{Z}_+^i} f_{i+1}(q)\big(\Pr\big\{\overline{Q}_i(t) = q \mid \overline{Q}_{i+1}(0)\big\} - \pi_i(q)\big)\\
&\quad = Q_{i+1}(0) - \gamma_{i+1}\tau_{i+1} - \sum_{t=0}^{\tau_{i+1}-1} \sum_{q\in\mathcal{Z}_+^i} f_{i+1}(q)\big(\Pr\big\{\overline{Q}_i(t) = q \mid \overline{Q}_{i+1}(0)\big\} - \pi_i(q)\big).
\end{aligned} \tag{31}$$

We now estimate the sum in the expression above. For each time $t \leqslant \tau_{i+1}$, let $\tau(t)$ be the unique time $< \tau_i$ such that $t - \tau(t)$ divides τ_i. Then

$$\sum_{j\leqslant i} w_j Q_j\big(\tau(t)\big) \leqslant \sum_{j\leqslant i} w_j Q_j(0) + w_{\max}^i L i \tau_i.$$

Note $\Pr\{\overline{Q}_i(t) = q \mid \overline{Q}_{i+1}(0)\} = \Pr\{\overline{Q}_i(t) = q \mid \overline{Q}_i(0)\}$. From lemma 4 the function $\Phi_i^{\mathrm{g}}(\overline{Q}_i(\cdot)) = \exp(s_i \sum_{j\leqslant i} w_j Q_j(\cdot))$ is a geometric Lyapunov function for some computable s_i with some computable geometric drift γ_i^{g}. Note $\Phi_i^{\mathrm{g}}(\overline{Q}_i(\cdot)) \geqslant 1$ since $\sum_{j\leqslant i} w_j Q_j(\cdot) \geqslant 0$. Recall that $f_{i+1}(q) \leqslant 1/a_{i+1}$ for any q. Thus $a_{i+1} f_{i+1}(q) \leqslant \Phi_i^{\mathrm{g}}(q)$ for any q. Applying theorem 3 for the function $f = a_{i+1} f_{i+1}(\cdot)$, and noting (19) is assumed to hold, we obtain that there exist computable values s_i, R_i, ρ_i which depend on parameters of the system and on B_j, w_j, τ_j, $j \leqslant i$, such that

$$\begin{aligned}
&\left|\sum_{q\in\mathcal{Z}_+^i} f_{i+1}(q)\big(\Pr\big\{\overline{Q}_i(t) = q \mid \overline{Q}_{i+1}\big(\tau(t)\big)\big\} - \pi_i(q)\big)\right|\\
&\quad \leqslant \frac{1}{a_{i+1}} \exp\bigg(s_i \sum_{j\leqslant i} w_j Q_j\big(\tau(t)\big)\bigg) R_i \rho_i^{(t-\tau(t))/\tau_i}\\
&\quad \leqslant \frac{1}{a_{i+1}} \exp\bigg(s_i \sum_{j\leqslant i} w_j Q_j(0) + s_i w_{\max}^i L i \tau_i\bigg) R_i \rho_i^{t/\tau_i - 1}.
\end{aligned}$$

For simplicity, we incorporate $(1/a_{i+1})\exp(s_i w^i_{\max} L i\tau_i)\rho_i^{-1}$ into the constant R_i. Combining with the assumption (29), we obtain from (31) that

$$
\begin{aligned}
E\big[Q_{i+1}(\tau_{i+1}) \mid \overline{Q}_{i+1}(0)\big] &\leqslant Q_{i+1}(0) - \gamma_i\tau_{i+1} + R_i \mathrm{e}^{s_i C}\sum_{t=0}^{\tau_{i+1}-1}\rho_i^{t/\tau_i} \\
&\leqslant Q_{i+1}(0) - \gamma_{i+1}\tau_{i+1} + \frac{R_i \mathrm{e}^{s_i C}}{1-\rho_i^{1/\tau_i}}. \qquad (32)
\end{aligned}
$$

Our next goal is to estimate

$$
E\Bigg[\sum_{j\leqslant i} w_j Q_j(\tau_{i+1}) \mid \overline{Q}_{i+1}(0)\Bigg] = E\Bigg[\sum_{j\leqslant i} w_j Q_j(\tau_{i+1}) \mid \overline{Q}_{i}(0)\Bigg]
$$

under the assumption (29). Since the Markov chain $\overline{Q}_i(\tau_i t)$ has a maximal jump $\nu \leqslant w^i_{\max} L i\tau_i$, then applying lemma 5 to $\tau = \tau_{i+1}$, we obtain

$$
E\Bigg[\sum_{j\leqslant i} w_j Q_j(\tau_{i+1}) \mid \overline{Q}_{i}(0)\Bigg] \leqslant \sum_{j\leqslant i} w_j Q_j(0) + B_i + \frac{w^i_{\max} L i\tau_i \exp(s_i w^i_{\max} L i\tau_i)}{\gamma_i^{\mathrm{g}}(1-\gamma_i^{\mathrm{g}})^2}.
$$

The important implication of this bound is that the expected difference of $\sum_{j\leqslant i} w_j Q_j(\cdot)$ during a time gap of length τ_{i+1} is upper bounded by a value, which does not depend on τ_{i+1}, C, B_{i+1}.

We now combine this bound with (32) to conclude that if (28)–(30) hold, then

$$
\begin{aligned}
&E\Bigg[\sum_{j\leqslant i+1} w_j Q_j(\tau_{i+1}) \mid \overline{Q}_{i+1}(0)\Bigg] - \sum_{j\leqslant i+1} w_j Q_j(0) \\
&\quad\leqslant -w_{i+1}\gamma_{i+1}\tau_{i+1} + w_{i+1}\mathrm{e}^{s_i C} V_1 + V_2,
\end{aligned}
$$

where

$$
V_1 \equiv \frac{R_i}{1-\rho_i^{1/\tau_i}}, \qquad V_2 = B_i + \frac{w^i_{\max} L i\tau_i \exp(s_i w^i_{\max} L i\tau_i)}{\gamma_i^{\mathrm{g}}(1-\gamma_i^{\mathrm{g}})^2}. \qquad (33)
$$

Observe that V_1, V_2 only depend on w_j, $j \leqslant i$, τ_i, B_i and the parameters of the model. Recall from part 1 of theorem 8 that for any $\varepsilon > 0$ we can compute an interval $[\tilde{\gamma}_{i+1}, \tilde{\gamma}'_{i+1}]$ with length ε which contains γ_{i+1}. Specifically, since by assumption of the theorem, $\gamma_{i+1} > 0$, then by making ε sufficiently small we will obtain $\tilde{\gamma}'_{i+1} > \tilde{\gamma}_{i+1} > 0$. Observe that the value γ_{i+1} of the effective drift, is a physical notion which depends only on the parameters of the system and is independent from the values of w_j, τ_i, B_i, etc.

After computing $\tilde{\gamma}_{i+1} \leqslant \gamma_{i+1}$ we specify the following constraint

$$
\tau_{i+1} \geqslant \frac{\mathrm{e}^{s_i C} V_1}{\tilde{\gamma}_{i+1}} + \frac{V_2 + 1}{w_{i+1}\tilde{\gamma}_{i+1}}. \qquad (34)
$$

Constraint (34) and the fact $\tilde{\gamma}_{i+1} < \gamma_{i+1}$ guarantee that (20) holds, provided that constraints (30), (34) can be satisfied.

We now analyze the second case, derive some additional constraints and, finally, show that all of the constraints can be satisfied simultaneously.

Case 2. We fix $\overline{Q}_{i+1}(0) = q \in \mathcal{Z}_+^{i+1}$ such that

$$\sum_{j\leqslant i} w_i q_i > C. \tag{35}$$

First, note

$$Q_{i+1}(\tau_{i+1}) \leqslant Q_{i+1}(0) + L\tau_{i+1}. \tag{36}$$

We now analyze the expected difference

$$E\left[\sum_{j\leqslant i} w_j Q_j(\tau_{i+1}) \mid \overline{Q}_{i+1}(0) = q\right] - \sum_{j\leqslant i} w_j q_j.$$

Let

$$\tau = \frac{a_i(C - B_i)}{w^i_{\max}}. \tag{37}$$

Since at each time step $t = 0, 1, 2, \ldots$, the value of $\sum_{j\leqslant i} w_j Q_j(t)$ can decrease by at most $w^i_{\max}(1/a_i)$ (at most $1/a_i$ items of types $j = 1, 2, \ldots, i$ can be placed into a bin at a time), then, from assumption (35), for any $t \leqslant \tau$,

$$\sum_{j\leqslant i} w_j Q_j(t) > C - w^i_{\max}\frac{1}{a_i}t \geqslant B_i. \tag{38}$$

Observe that

$$\sum_{j\leqslant i} w_j Q_j(\tau) - \sum_{j\leqslant i} w_j Q_j\left(\left\lfloor\frac{\tau}{\tau_i}\right\rfloor\tau_i\right) \leqslant w^i_{\max} i L\tau_i, \tag{39}$$

We now apply assumption (18) of the theorem for times $0, \tau_i, 2\tau_i, \ldots, (\lfloor\tau/\tau_i\rfloor - 1)\tau_i$. For each of these time instances we have by (38) that the state belongs to the region $\sum_{j\leqslant i} w_j Q_j(\cdot) > B_i$. Therefor using (18) and (39) we obtain

$$E\left[\sum_{j\leqslant i} w_j Q_j(\tau) \mid \overline{Q}_{i+1}(0) = q\right] - \sum_{j\leqslant i} w_j q_j \leqslant -\frac{\tau}{\tau_i} + 1 + w^i_{\max} i L\tau_i. \tag{40}$$

Applying lemma 5, for any $q' \in \mathcal{Z}^i_+$,

$$E\left[\sum_{j\leqslant i} w_j Q_j(\tau_{i+1}) \mid \overline{Q}_i(\tau) = q'\right] \leqslant V_2 + \sum_{j\leqslant i} w_j q'_j + w^i_{\max} i L\tau_i \tag{41}$$

where V_2 is defined by (33) and the extra $w^i_{\max} i L\tau_i$ term comes from the fact that $\tau_{i+1} - \tau$ does not necessarily divide τ_i, and the approximation similar to (39) is needed.

Combining this bound with (40), and summing over q', we obtain

$$E\left[\sum_{j\leqslant i} w_j Q_j(\tau_{i+1}) \mid \overline{Q}_{i+1}(0) = q\right] - \sum_{j\leqslant i} w_j q_j \leqslant -\frac{\tau}{\tau_i} + 1 + 2w^i_{\max} i L\tau_i + V_2.$$

Finally, we combine this inequality with (36) and (37) to obtain

$$E\left[\sum_{j\leqslant i+1} w_j Q_j(\tau_{i+1}) \mid \overline{Q}_{i+1}(0) = q\right] - \sum_{j\leqslant i+1} w_j q_j \leqslant -\frac{a_i(C-B_i)}{w^i_{\max}\tau_i} + 1 + 2w^i_{\max} i L\tau_i + V_2 + w_{i+1}L\tau_{i+1}. \tag{42}$$

We introduce our last constraint

$$\frac{a_i(C-B_i)}{w^i_{\max}\tau_i} \geqslant 2w^i_{\max} i L\tau_i + V_2 + w_{i+1}L\tau_{i+1} \tag{43}$$

which guarantees that the left-hand side of (42) is at most -1 and, as a result, inequality (20) holds.

Our next task is to show that $w_{i+1}, \tau_{i+1}, C, B_{i+1}$ can be selected in such a way that all the constraints (27), (30), (34), (43) are satisfied. Note, that (43) implies $C > B_i$ and B_{i+1} is present only in the constraints (27) and (30), which can be satisfied by taking

$$B_{i+1} = C + \frac{w_{i+1}(\tau_{i+1}+1)}{a_{i+1}}. \tag{44}$$

We rewrite the remaining constraints below, simplifying them by introducing new notations in an obvious way

$$\tau_{i+1} \geqslant \mathrm{e}^{s_i C} V_3 + \frac{V_4}{w_{i+1}}, \tag{45}$$

$$V_5 C \geqslant w_{i+1}L\tau_{i+1} + V_6. \tag{46}$$

Note that V_i, $i = 1, 2, \ldots, 7$, depend only on τ_i, B_i, w_j, $j \leqslant i$, and the parameters of the model. We take $C = (2V_4L + V_6)/V_5$, $\tau_{i+1} = 2\mathrm{e}^{s_i C} V_3$, $w_{i+1} = 2V_4/\tau_{i+1}$. An elementary calculations show that the constraints are satisfied with equality. We constructed the values $\tau_{i+1}, w_{i+1}, B_{i+1}$ which satisfy (20). In particular, $\sum_{j\leqslant i+1} w_j Q_j(\cdot)$ is a linear Lyapunov function for the Markov chain $\overline{Q}_{i+1}(\tau_{i+1}t)$.

To complete the proof of the theorem we need to arrange for (21) to hold. Let s_{i+1} and $\gamma^{\mathrm{g}}_{i+1}$ be defined as in lemma 4. We first increase B_{i+1} to a higher value. Specifically, let

$$B'_{i+1} = B_{i+1} + w^{i+1}_{\max}(i+1) \times L\tau_{i+1}\left(B_{i+1} + \frac{w^{i+1}_{\max}(i+1)L\tau_{i+1}\exp(s_{i+1}w^{i+1}_{\max}(i+1)L\tau_{i+1})}{\gamma^{\mathrm{g}}_{i+1}(1-\gamma^{\mathrm{g}}_{i+1})^2} + 2\right).$$

Note that the constraints (27) and (30) are still satisfied. Let $\hat{\tau} > 0$ be large enough so that the following conditions are satisfied: $\hat{\tau}/\tau_{i+1}$ is an integer,

$$\min_{q',q'':\, \sum_{j\leqslant i} w_j q'_j, \sum_{j\leqslant i} w_j q''_j \leqslant B'_{i+1}} \Pr\{\overline{Q}_{i+1}(\hat{\tau}) = q'' \mid \overline{Q}_{i+1}(0) = q'\} > 0$$

and

$$\hat{\tau} \geqslant B_{i+1} + \frac{w_{\max}^{i+1}(i+1)L\tau_{i+1}\exp(s_{i+1}w_{\max}^{i+1}(i+1)L\tau_{i+1})}{\gamma_{i+1}^{g}(1-\gamma_{i+1}^{g})^2} + 2.$$

The inequality condition is satisfied for sufficiently large τ since the chain $\overline{Q}(t)$ and therefore the chain $\overline{Q}_{i+1}(\hat{\tau}t)$ is irreducible and aperiodic. Applying lemma 6, by the choice of B'_{i+1}, the function $\sum_{j\leqslant i+1} w_j Q_j(\cdot)$ is also a linear Lyapunov function for the Markov chain $\overline{Q}_{i+1}(\hat{\tau}t)$, $t = 0, 1, 2, \ldots$. The exception parameter for this Markov chain is B'_{i+1} and the drift is again -1. Now we have that both (20) and (21) hold when B'_{i+1} and $\hat{\tau}$ take the role of B_{i+1} and τ_{i+1} respectively. This completes the proof of the theorem. □

We conclude this section by providing an example of application of theorem 7.

Example. Let $a_1 = 3/5$, $a_2 = 1/7$, $p_0^1 = 2/3$, $p_1^1 = 0$, $p_2^1 = 1/3$, $p_0^2 = p_8^2 = 1/2$, $p_l^2 = 0, l = 1, 2, \ldots, 7$. From this we obtain $\lambda_1 = 2/3$, $\lambda_2 = 4$, $f_1 = 1$, $f_2(0) = 7$, $f_2(k) = 2$ for $k \geqslant 1$. Note that $Q_1(t)$ is a simple random walk with $\Pr\{Q_1(t+1) = Q_1(t)+1\} = 1/3$, $\Pr\{Q_1(t+1) = Q_1(t)-1\} = 2/3$ when $Q_1(t) > 0$ and $\Pr\{Q_1(t+1) = 1\} = 1/3$, $\Pr\{Q_1(t+1) = 0\} = 2/3$ when $Q_1(t) = 0$. For this random walk it is easy to obtain the stationary distribution: $\pi_1(k) = 1/2^{k+1}$, $k = 0, 1, 2, \ldots$. Applying theorem 7 to the second process $Q_2(t)$ we obtain

$$-\gamma_2 = \lambda_2 - f_2(0)\pi_1(0) - \sum_{k=1}^{\infty} f_2(k)\pi_1(k) = 4 - 7 \times \frac{1}{2} - 2 \times \frac{1}{2} = -\frac{1}{2} < 0$$

and the system is stable. Note that the arrival rates λ_1, λ_2 do not satisfy the symmetry condition (1) and the results by Coffman and Stolyar [6] are not applicable. In the example above we had some of the $p_j^i = 0$, violating the condition of theorem 7. It is immediate to see that making these values just slightly higher than zero does not change the argument.

Now suppose the parameters for the first process are the same, but for the second process $a_2 = 1/2$, $p_0^2 = 0$, $p_1^2 = 1 - \varepsilon$, $p_2^2 = \varepsilon$, for some small $\varepsilon > 0$. Then $\lambda_2 = 1 + \varepsilon$ and $f_2(0) = 2$, $f_2(k) = 0$, $k > 0$. We obtain

$$-\gamma_2 = 1 + \varepsilon - 2 \times \frac{1}{2} = \varepsilon > 0$$

and the system is unstable. Note, though, that the load condition (2) $\lambda_1 a_1 + \lambda_2 a_2 = 2/3 \times 3/5 + (1+\varepsilon) \times 1/2 = 9/10 + \varepsilon/2 < 1$ is satisfied when $\varepsilon < 1/5$.

6. Conclusion

We considered a stochastic bin packing model as a model for online bandwidth allocation process. In our model the items to be packed arrive over time according to some stochastic arrival process, and unpacked items form queues. We established constructive necessary and sufficient condition for stability of the underlying queueing process when the Best Fit packing algorithm is used. Our analysis is built on using a linear Lyapunov function and exponential mixing property of infinite Markov chains allowing a Lyapunov function. The result is established by introducing a computable notion of effective drifts of the individual queueing processes. Unfortunately, our algorithm is very non-efficient as the computation time grows badly as the size of the model increases. Reducing the complexity of the algorithm is an interesting open problem.

Acknowledgments

Many thanks to Ed. Coffman, Sasha Stolyar, Philippe Robert and anonymous referees for insightful discussions and corrections.

Appendix

Proof of lemma 1. For every starting state $Q(0) = q$ and every finite t, with positive probability there are no new arriving items during the next $\tau = 1, 2, \ldots, t$ time units. Then for t sufficiently large all the N queue length become empty. Thus, state 0 is reachable with positive probability from any starting state q. The origin is also aperiodic since $\Pr\{Q(t+1) = 0 \mid Q(t) = 0\} = \prod_{1 \leqslant i \leqslant N} p_0^i > 0$.

We now prove the second part of the lemma and thus suppose $p_l^i > 0$ for all $1 \leqslant i \leqslant N$, $0 \leqslant l \leqslant L$. Applying the first part we may assume $q_0 = 0$. The proof is by induction in $|q| = \sum_i q_i$. We already know that origin is reachable from itself in one step with positive probability. Suppose the assertion holds for all q with $|q| \leqslant l$ and consider any q with $|q| = l + 1$. Let q_{i_0} be any positive component of q. Then $\hat{q} = (q_1, \ldots, q_{i_0-1}, q_{i_0} - 1, q_{i_0+1}, \ldots, q_N)$ satisfies $|\hat{q}| = |q| - 1 \leqslant l$ and by assumption $\Pr(Q(t) = \hat{q} \mid Q(0) = 0) > 0$ for some $t > 0$. Let $b_i = b_i(\hat{q}) \geqslant 0$ denote the number of size a_i items that will be packed into a bin when the state is $\hat{q}$. Note $b_i \leqslant 1/a_N < L$. During any time step, with probability $\hat{p} \equiv p_{b_1}^1 p_{b_2}^2 \cdots p_{b_{i_0}+1}^{i_0} \cdots p_{b_N}^N > 0$ there will be $b_{i_0} + 1$ arrivals of size a_{i_0} items and b_i arrivals of size a_i, $i \neq i_0$, items. Then $\Pr(Q(t+1) = q \mid Q(t) = \hat{q}) = \hat{p} > 0$. We conclude

$$
\begin{aligned}
&\Pr\bigl(Q(t+1) = q \mid Q(0) = 0\bigr) \\
&\quad \geqslant \Pr\bigl(Q(t) = \hat{q} \mid Q(0) = 0\bigr) \Pr\bigl(Q(t+1) = q \mid Q(t) = \hat{q}\bigr) > 0.
\end{aligned}
$$

This completes the proof of the lemma. □

Proof of lemma 4. Let b be a (positive or negative) fixed constant. Using second order Taylor expansion of $g(s) = \mathrm{e}^{sb}$ around $s = 0$, the inequality $\mathrm{e}^{sb} \leqslant 1 + sb + s^2b^2\mathrm{e}^{s|b|}/2$ holds. Assume $\sum_{j \leqslant i} w_j Q_j(t) > B$. Then

$$\begin{aligned}
&E\left[\exp\left(s\sum_j w_j\big(Q_j(t+1) - Q_j(t)\big)\right) \mid Q(t)\right] \\
&\quad\leqslant 1 + sE\left[\sum_j w_j\big(Q_j(t+1) - Q_j(t)\big) \mid Q(t)\right] \\
&\qquad + \frac{1}{2}s^2 E\left[\left(\sum_j w_j\big(Q_j(t+1) - Q_j(t)\big)\right)^2\right. \\
&\qquad\qquad \left.\times \exp\left(s\left|\sum_j w_j\big(Q_j(t+1) - Q_j(t)\big)\right|\right) \mid Q(t)\right] \\
&\quad\leqslant 1 - s\gamma + \frac{1}{2}s^2\nu^2 \exp(s\nu),
\end{aligned}$$

where the Lyapunov condition $E[\sum_j w_j(Q_j(t+1) - Q_j(t)) \mid Q(t)] \leqslant -\gamma$ and the definition of ν is used. For the specified values of s, the resulting expression is at most $\gamma^{\mathrm{g}} \equiv 1 - s\gamma/2 < 1$. We conclude that $\exp(s\sum_j Q_j(t))$ is a geometric Lyapunov function. The exception parameter of this function can be taken $B^{\gamma} = \exp(sB)$. □

Proof of lemma 5. Consider the random trajectory $Q(t), t = 0, 1, \ldots, \tau$, with the deterministic initial state $Q(0)$. It suffices to prove that

$$E\left[\sum_j w_j Q_j(\tau) \mid Q(0)\right] \leqslant \max\left\{\sum_j w_j Q_j(0), B\right\} + \frac{\nu\exp(s\nu)}{\gamma^{\mathrm{g}}(1-\gamma^{\mathrm{g}})^2}. \tag{A.1}$$

Let

$$T = \max\left\{t \leqslant \tau \colon \sum_j w_j Q_j(t) \leqslant \max\left\{\sum_j w_j Q_j(0), B\right\}\right\}.$$

The set of such t is non-empty since $t = 0$ clearly belongs to it. If $T = \tau$, then the statement of the lemma holds. Otherwise, $\sum_j w_j Q_j(t) > \max\{\sum_j w_j Q_j(0), B\}$ for all $T < t \leqslant \tau$. For any fixed $0 \leqslant t_0 < \tau$ we now prove that

$$\Pr\big\{T = t_0 \mid Q(0)\big\} \leqslant \big(\gamma^{\mathrm{g}}\big)^{\tau - t_0 - 1}\exp(s\nu). \tag{A.2}$$

Before we prove this bound, we show that it actually implies (A.1). By definition we have

$$\sum_j w_j Q_j(T) \leqslant \max\left\{\sum_j w_j Q_j(0), B\right\}. \tag{A.3}$$

Note also, that

$$\sum_j w_j Q_j(\tau) \leqslant (\tau - T)\nu + \sum_j w_j Q_j(T), \tag{A.4}$$

since, at each time step, $\sum w_j Q_j(t)$ can increase by at most ν. Using the bounds (A.2)–(A.4), we obtain

$$\begin{aligned}
&E\biggl[\sum_j w_j Q_j(\tau) \mid Q(0)\biggr] \\
&\quad\leqslant E\biggl[\sum_j w_j Q_j(T) \mid Q(0)\biggr] + \sum_{t_0=0}^{\tau-1}(\tau - t_0)\nu \Pr\bigl\{T = t_0 \mid Q(0)\bigr\} \\
&\quad\leqslant \max\biggl\{\sum_j w_j Q_j(0), B\biggr\} + \sum_{0\leqslant t_0\leqslant\tau}(\tau - t_0)\nu\bigl(\gamma^{\mathrm{g}}\bigr)^{\tau-t_0-1}\exp(s\nu).
\end{aligned}$$

Using the fact $\sum_{x=0}^{\infty} x\alpha^x = 1/(1-\alpha)^2$ for any $0 < \alpha < 1$, and applying it to $\alpha = \gamma^{\mathrm{g}}$, we obtain the bound (A.1).

Thus, we are left with proving (A.2). We fix any vector $q \in \mathcal{Z}_+^d$ satisfying $\sum_j w_j q_j \leqslant \max\{\sum_j w_j Q_j(0), B\}$. We show that the bound (A.2) holds, when $\Pr\{T = t_0 \mid Q(0)\}$ is replaced by $\Pr\{T = t_0 \mid Q(t_0) = q\}$, for any q satisfying $\sum_j w_j q_j \leqslant \max\{\sum_j w_j Q_j(0), B\}$. Since $Q(t)$ is a Markov chain, this would imply (A.2). The probability in (A.2) is upper bounded by

$$\begin{aligned}
\Pr\biggl\{&\sum_j w_j Q_j(\tau) > \sum_j w_j q_j \text{ and } \forall k = 1, \dots, \tau - t_0 - 1, \\
&\sum_j w_j Q_j(\tau - k) > B \mid Q(t_0) = q\biggr\}.
\end{aligned}$$

For every $r = 1, \dots, \tau - t_0 - 1$, we denote by E_r the event "$\forall r \leqslant k \leqslant \tau - t_0 - 1$, $\sum_j w_j Q_j(\tau - k) > B$". Now we use lemma 4 and the corresponding parameters s, γ^{g}, $B^{\mathrm{g}} = \exp(sB)$. We rewrite the probability above as

$$\begin{aligned}
&\Pr\biggl\{\exp\biggl(s\sum_j w_j Q_j(\tau)\biggr) > \exp\biggl(s\sum_j w_j q_j\biggr) \mid E_1\biggr\}\Pr\bigl\{E_1 \mid Q(t_0) = q\bigr\} \\
&\quad\leqslant \frac{E[\exp(s\sum_j w_j Q_j(\tau))|E_1]\Pr\{E_1|Q(t_0) = q\}}{\exp(s\sum_j w_j q_j)},
\end{aligned} \tag{A.5}$$

where we use Markov's inequality. Note that the event E_1 implies the event $\sum_j w_j \times Q_j(\tau - 1) > B$ which, in turn, implies $\exp(s\sum_j w_j Q_j(tau - 1)) > \exp(sB) = B^{\mathrm{g}}$.

Since by the conclusion of lemma 4, $\exp(s\sum w_j Q_j(\cdot))$ is a geometric Lyapunov function, then

$$E\left[\exp\left(s\sum_j w_j Q_j(\tau)\right) \mid E_1\right] \leqslant \gamma^{\mathrm{g}} E\left[\exp\left(s\sum_j w_j Q_j(\tau-1)\right) \mid E_1\right].$$

Note

$$\begin{aligned}
&E\left[\exp\left(s\sum_j w_j Q_j(\tau-1)\right) \mid E_1\right]\Pr\left\{E_1 \mid Q(t_0)=q\right\}\\
&\quad= E\left[\exp\left(s\sum_j w_j Q_j(\tau-1)\right)\mathbb{1}\{E_1\} \mid Q(t_0)=q\right]\\
&\quad\leqslant E\left[\exp\left(s\sum_j w_j Q_j(\tau-1)\right)\mathbb{1}\{E_2\} \mid Q(t_0)=q\right]\\
&\quad= E\left[\exp\left(s\sum_j w_j Q_j(\tau-1)\right) \mid E_2\right]\Pr\left\{E_2 \mid Q(t_0)=q\right\}.
\end{aligned}$$

Again, the event E_2 implies the event $\sum_j w_j Q_j(\tau-2) > B$ or $\exp(s\sum_j w_j Q_j(\tau-2)) > B^{\mathrm{g}}$. Therefore,

$$E\left[\exp\left(s\sum_j w_j Q_j(\tau-1)\right) \mid E_2\right] \leqslant \gamma^{\mathrm{g}} E\left[\exp\left(s\sum_j w_j Q_j(\tau-2)\right) \mid E_2\right].$$

Continuing, we obtain the following upper bound on the right-hand side of (A.5)

$$\left(\gamma_i^{\mathrm{g}}\right)^{\tau-t_0-1}\frac{E[\exp(s\sum_{j\leqslant i} w_j Q_j(t_0+1)) \mid Q(t_0)=q]}{\exp(s\sum_j w_j q_j)}.$$

Note

$$\sum_j w_j Q_j(t_0+1) \leqslant \sum_j w_j q_j + \nu.$$

Then, we obtain a bound

$$\left(\gamma^{\mathrm{g}}\right)^{\tau-t_0-1}\frac{\exp(s(\sum_j w_j q_j+\nu))}{\exp(s\sum_j w_j q_j)} = \left(\gamma^{\mathrm{g}}\right)^{\tau-t_0-1}\exp(s\nu),$$

which is the required bound (A.2). □

Proof of lemma 6. Clearly, for the Markov subchain $Q(\tau)$ the maximal jump satisfies $\nu' \leqslant \nu\tau$. Now we need to show that

$$E\left[\sum_j w_j Q_j(\tau) \mid Q(0)\right] \leqslant -1 + \sum_j w_j Q_j(0),$$

VTEX(JK) PIPS No:5384235 artty:res (Kluwer BO v.2002/10/03)
q5384235.tex; 2/09/2004; 12:34; p. 23

whenever $\sum_j w_j Q_j(0) > B'$. Let

$$t_0 = \left\lceil \left(B + \frac{\nu \exp(s\nu)}{\gamma^{g}(1-\gamma^{g})^2} + 1 \right) \Big/ \gamma \right\rceil < \left(B + \frac{\nu \exp(s\nu)}{\gamma^{g}(1-\gamma^{g})^2} + 1 + \gamma \right) \Big/ \gamma. \tag{A.6}$$

We have

$$E\left[\sum_j w_j Q_j(\tau) \mid Q(0) \right] = \sum_q E\left[\sum_j w_j Q_j(\tau) \mid Q(t_0) = q \right] \Pr\{Q(t_0) = q \mid Q(0)\}.$$

Applying lemma 5 to $\tau - t_0$ instead of τ, we have

$$\begin{aligned} E\left[\sum_j w_j Q_j(\tau) \mid Q(0) \right] &\leqslant \sum_q \left(B + \frac{\nu \exp(s\nu)}{\gamma^{g}(1-\gamma^{g})^2} + \sum_j w_j q_j \right) \Pr\{Q(t_0) = q \mid Q(0)\} \\ &= B + \frac{\nu \exp(s\nu)}{\gamma^{g}(1-\gamma^{g})^2} + E\left[\sum_j w_j Q_j(t_0) \mid Q(0) \right]. \end{aligned}$$

Note from (A.6) that $B' - \nu t \geqslant B$ for all $t \leqslant t_0$. Then, since $\sum_j w_j Q_j(\cdot)$ is a Lyapunov function with drift $-\gamma$, and the exception parameter B, we have

$$\begin{aligned} E\left[\sum_j w_j Q_j(t_0) \mid Q(0) \right] &\leqslant -\gamma t_0 + \sum_j w_j Q_j(0) \\ &\leqslant -\left(B + \frac{\nu \exp(s\nu)}{\gamma^{g}(1-\gamma^{g})^2} + 1 \right) + \sum_j w_j Q_j(0), \end{aligned}$$

where we use (A.6) in the second inequality. Combining, we obtain

$$E\left[\sum_j w_j Q_j(\tau) \mid Q(0) \right] \leqslant -1 + \sum_j w_j Q_j(0).$$

This completes the proof of the lemma. □

References

[1] S. Albers and M. Mitzenmacher, Average-case analysis of first fit and random fit bin packing, Random Struct. Algorithms 16 (2000) 240–259.

[2] D. Bertsimas, D. Gamarnik and J. Tsitsiklis, Performance of multiclass Markovian queueing networks via piecewise linear Lyapunov functions, Ann. Appl. Probab. 11(4) (2001) 1384–1428.

[3] E.G. Coffman, Jr., C. Courcoubetis, M.R. Garey, D.S. Johnson, L.A. McGeoch, P.W. Shor, R.R. Weber and M. Yannakakis, Fundamental discrepancies between average-case analyses under discrete and continuous distributions: A bin packing case study, in: *Proc. of the 23d Ann. ACM Symposium on the Theory of Computing (STOC)* (1991).

[4] E.G. Coffman, Jr., S. Halfin, A. Jean-Marie and Ph. Robert, Stochastic analysis of a slotted, FIFO communication channel, IEEE Trans. Inform. Theory 39 (1993) 1555–1566.

[5] E.G. Coffman, Jr., P. Robert and A.L. Stolyar, The interval packing process of linear networks, Performance Evaluation Rev. 27(3) (1999).

[6] E.G. Coffman, Jr. and A.L. Stolyar, Bandwith packing, Algorithmica 29 (1999) 70–88.
[7] J.F. Dantzer, M. Haddani and P. Robert, On the stability of a bandwidth packing algorithm, Probab. Engrg. Inform. Sci. 14(1) (2000) 57–79.
[8] J.F. Dantzer and P. Robert, Analysis of a multi-class queueing system, Ann. Appl. Probab. 12(3) (2002) 860–889.
[9] G. Fayolle, V.A. Malyshev and M.V. Menshikov, *Topics in the Constructive Theory of Countable Markov Chains* (Cambridge Univ. Press, Cambridge, 1995).
[10] D. Gamarnik, On deciding stability of constrained homogeneous random walks and queueing systems, Math. Oper. Res. 27(2) (2002) 272–293.
[11] D.S. Johnson, A. Demers, J.D. Ullman, M.R. Garey and R.L. Graham, Worst case bounds for simple one-dimensional packing algorithms, SIAM J. Comput. 3 (1974) 299–325.
[12] C. Kenyon and M. Mitzenmacher, Linear waste of best fit bin packing on skewed distributions, in: *Proc. of the 41st Symposium on the Foundations of Computer Science* (2000).
[13] C. Kenyon, A. Sinclair and Y. Rabani, Biased random walks, Lyapunov functions, and stochastic analysis of best fit bin packing, in: *Proc. of the 7th ACM–SIAM Symposium on Discrete Algorithms* (1995).
[14] C. Kipnis and Ph. Robert, A dynamic storage process, Stochastic Process. Appl. 34 (1990) 155–169.
[15] T. Leighton and P.W. Shor, Tight bounds for minimax grid matching with applications to average case analysis of algorithms, Combinatorica 9 (1989) 161–187.
[16] S.P. Meyn and R.L. Tweedie, *Markov Chains and Stochastic Stability* (Springer, Berlin, 1993).
[17] S.P. Meyn and R.L. Tweedie, Computable bounds for geometric convergence rates of Markov chains, Ann. Appl. Probab. 4 (1994) 981–1011.
[18] A. Rényi, On a one-dimensional random space filling problem, MTA Mat. Kut. Internat. Kzl. 3 (1958) 109–127.
[19] P.W. Shor, The average case analysis of some on-line algorithms for binpacking, Combinatorica 6 (1986) 179–200.