Biased Random Walks, Lyapunov Functions, and Stochastic Analysis of Best Fit Bin Packing
(Preliminary Version)

CLAIRE KENYON† YUVAL RABANI‡ ALISTAIR SINCLAIR§

Abstract

We study the average case performance of the Best Fit algorithm for on-line bin packing under the distribution $U\{j,k\}$, in which the item sizes are uniformly distributed in the discrete range $\{1/k, 2/k, \ldots, j/k\}$. Our main result is that, in the case $j = k-2$, the expected waste for an infinite stream of items remains bounded. This settles an open problem posed recently by Coffman et al [4]. It is also the first result which involves a detailed analysis of the infinite multi-dimensional Markov chain underlying the algorithm.

1 Introduction

In the one-dimensional bin packing problem, one is given a sequence $a_1, \ldots, a_n \in (0,1]$ of items and asked to pack them into bins of unit capacity in such a way as to minimize the number of bins used. This problem is well known to be NP-hard, and a vast literature has developed around the design and analysis of efficient approximation algorithms for it. The most widely studied among these is the Best Fit algorithm, in which the items are packed on-line, with each successive item going into a partially filled bin with the smallest residual capacity large enough to accommodate it; if no such bin exists, a new bin is started.

†CNRS, Ecole Normale Supérieure de Lyon, France. E-mail: `kenyon@lip.ens-lyon.fr`. Work done while this author was visiting UC Berkeley, partially supported by a NATO Fellowship.

‡Faculty of Computer Science, Technion, Haifa 32000, Israel. E-mail: `rabani@cs.technion.ac.il`. Work done while this author was a postdoctoral fellow in the Department of Computer Science, University of Toronto.

§Computer Science Division, University of California, Berkeley CA 94720-1776. Email: `sinclair@cs.berkeley.edu`. Supported by NSF grant CCR-9505448 and a UC Berkeley Faculty Research Grant.

Best Fit was first analyzed in the worst case by Johnson *et al* [9], who proved that the number of bins used is always within a factor 1.7 of that used by an optimal algorithm. When items are drawn from the uniform distribution on $(0,1]$, the expected waste of Best Fit was shown by Shor [12] to be $\Theta(n^{1/2}\log^{3/4} n)$. (The *waste* is the difference between the number of bins used and the sum of the sizes of all the items, and is the standard measure of performance of bin packing algorithms in the average case.) Thus among on-line algorithms Best Fit is currently the best available, in the sense that no algorithm is known which beats it both in the worst case and in the uniform average case. This, together with its intuitive appeal and ease of implementation, makes it the algorithm of choice in most applications.

With the goal of achieving a better understanding of the Best Fit algorithm, researchers have recently considered its behavior under various other input distributions, notably the class of discrete distributions $U\{j,k\}$ for integers $j < k$. Here the item sizes, instead of being chosen from the continuous real interval $(0,1]$, are selected uniformly from the finite set of equally spaced values i/k, for $1 \le i \le j$. Equivalently, we may think of the bins as having capacity k and the item sizes being uniformly distributed on the integers $\{1,\ldots,j\}$. This family of distributions is of interest for two reasons. Firstly, it is an important step towards exploring the robustness of Best Fit under *non-uniform* distributions (because the distribution is biased towards smaller items); and secondly, it applies to the more realistic case of discrete rather than continuous item sizes. (For more extensive background, the reader is referred to [4] and the upcoming survey [3].)

Very little is known in rigorous terms about the performance of Best Fit under this distribution, with the exception of a few extreme cases: when $j = k-1$, the behavior can be related to that for the continuous uniform distribution on $(0,1]$, yielding expected waste

$\Theta(n^{1/2})$ [2]; and if j is very small compared to k (specifically, if $j \le \sqrt{2k+2.25}-1.5$) then the expected waste is known to be bounded by a constant as $n \to \infty$ [2]. The expected waste is also easily seen to be bounded when $j \le 2$ for all $k > 3$.

Nonetheless, there is much experimental evidence to suggest that the behavior of Best Fit for various pairs (j, k) is complex and interesting. For example, it appears that the expected waste remains bounded when j is sufficiently close to k or to 1, but that it grows linearly when the ratio j/k is close to a critical value around 0.80. Moreover, in all cases (except $j = k - 1$) where the expected waste is unbounded it appears to grow linearly with n. Some large scale simulation results, together with some conjectures, are described in [4].

In an attempt to explain this behavior, Coffman *et al* introduced an interesting approach based on a view of the algorithm as a multi-dimensional Markov chain [4]. The states of the chain are non-negative integer vectors $(s_1, \ldots, s_{k-1})$, where s_i represents the current number of open bins of residual capacity i. Note that such a vector contains all relevant information about the state of the algorithm: in Best Fit, the ordering on the open bins is insignificant, and since we are measuring waste we need not concern ourselves with bins that have already been filled. It is a simple matter to write down the new vector s' that results from the arrival of any item $i \in \{1, \ldots, j\}$; since each item arrives with probability $1/j$, this immediately gives the transition probabilities of the chain. (See Section 2.1 below for a more formal definition.) Thus we have a Markov chain on the infinite $(k-1)$-dimensional space $\mathbf{Z}_+^{\mathbf{k-1}}$. The expected waste of Best Fit is intimately related to the asymptotic behavior of this chain.

We note in passing that similar Markov chains have been an object of study in queueing theory for over four decades (see, e.g., [13]); in computer science they have also received attention, for example in the stochastic analysis of packet routing [10]. Despite this extensive body of research few general analytical tools exist, and even the simplest questions, such as whether such a chain is ergodic, seem hard to answer. The chief exception is the method of constructive use of Lyapunov functions, developed in recent years mainly by Malyshev, Menshikov and Fayolle (see [5] for a comprehensive account). The range of situations in which they are able to apply their method appears to be quite limited, however; the highlights are a complete classification of two- and three-dimensional *jump-bounded* space-homogeneous Markov chains (i.e., the transitions are limited to geometrically close states). Obviously, the Markov chains that arise in the analysis of Best Fit are jump-bounded, but of much higher dimension.

This was the starting point for Coffman *et al*, who proceeded to analyze the Best Fit Markov chain for small values of j and k, using a novel approach. In the absence of analytical tools for high dimensional Markov chains, they used a computer program to search in an appropriate class of functions for a Lyapunov function (i.e., a potential function obeying certain properties, notably a systematic drift over some bounded number of steps). The existence of a suitable Lyapunov function for a given pair (j, k) implies bounded or linear expected waste. Coffman *et al* were able to classify the expected waste as bounded or linear for values of k up to 14 and most corresponding values $j < k - 1$.

This approach, while interesting, suffers from several obvious drawbacks, as observed by the authors themselves. Evidently, there is no prospect that it can lead to proofs for infinite sequences of (j, k) pairs; in fact, the time and space resources consumed by the search make it infeasible to extend the study beyond a very small finite range of values for j and k. Perhaps most importantly, the technique seems to yield almost no useful insight into *why* the algorithm performs as it does: for example, the Lyapunov function that proves bounded expected waste for $j = 5$, $k = 7$ is a linear function based on 23 steps of the Markov chain, while that for $j = 7$, $k = 10$ is a 15-step quadratic function, neither of which has any intuitive basis.

In this paper, we aim at *analytical* results on the behavior of Best Fit for an infinite sequence of values (j, k). Specifically, we explore the line $j = k - 2$; this is the "smallest" interesting case beyond $j = k - 1$, which is the discrete analog of the continuous uniform distribution. Coffman *et al* exhibit computer proofs that the expected waste is bounded in this case for $k \le 10$, and also conjecture on the basis of simulations that it remains bounded for larger values of k. Our main result proves this conjecture for all k:

Theorem 1 *The expected waste of the Best Fit algorithm under the discrete uniform distribution $U\{k-2, k\}$ is bounded for all k.*

Note the dramatic contrast with the apparently very similar case $j = k - 1$, in which the expected waste grows unboundedly with n.

Of at least as much interest as this result itself, in our view, are the techniques we use to prove it. Our starting point is again the multi-dimensional Markov chain of Coffman *et al.* However, we develop an alternative view of the chain that seems rather easier to visualize: in this view, the state s of the chain at any time is represented by $k-1$ *tokens* placed on the non-negative integers, with token i at position s_i. The tokens move around as a dynamical system under the influence of item insertions. With the aid of this view, and the intuition that comes with it, we are able to design an explicit Lyapunov function that proves bounded expected waste for all pairs (j, k) with $j = k - 2$. The Lyapunov function is essentially just the waste in the algorithm.

The analysis of the Lyapunov function is somewhat subtle, which perhaps explains why it had not been done before. In order to establish the drift in the Lyapunov function, we have to consider $T(j)$ steps of the Markov chain, where $T(j)$ is an exponential function of j; the drift is proved by a detailed comparison of the time evolution of the Lyapunov function with a random walk on the non-negative integers. More specifically, the Lyapunov function is a linear combination of the coordinate values of the multi-dimensional chain, and we are able to relate the behavior of the individual coordinates to one-dimensional symmetric random walks that are biased by a limited adversary. This adversary model corresponds to a worst case assumption on the effect of other coordinates, and we believe it to be of independent interest. It is similar in flavor to, but differs essentially from, the biased random walk model considered by Azar *et al* [1] in a different context. The model in [1] is allowed to bias the transition probabilities slightly on every step, whereas our adversary may intervene overwhelmingly but only on a limited number of steps.

In addition to settling an open problem posed in [4], our result, more significantly, is the first proof that exploits the detailed structure of the multi-dimensional Markov chain, and thus the first that provides an understanding of its behavior. We believe it is likely that our techniques can be extended to analyze the Best Fit Markov chain for other pairs of values (j, k), and perhaps also to other situations in the analysis of algorithms in which homogeneous multi-dimensional Markov chains of this kind arise.

The remainder of the paper is structured as follows. In Section 2 we introduce the token model as a convenient representation of the Markov chain underlying the algorithm, and establish various fundamental properties of it. In Section 3 we define our Lyapunov function and analyze its behavior using comparisons with biased random walks, concluding with the proof of Theorem 1.

2 The token model

2.1 Definitions

As advertised in the Introduction, we describe the behavior of the Best Fit algorithm over time in terms of the evolution of a dynamical system. In this system, $k-1$ *tokens* move among the non-negative integer points under the influence of item insertions, as follows. The tokens are labeled $1, 2, \ldots, k-1$. At any time instant t, the position of token i is the number of open bins at time t with residual capacity exactly i. We shall denote the state of the system at time t by $s(t) = (s_1(t), \ldots, s_{k-1}(t))$, a vector random variable taking values in $\mathbf{Z}_+^{\mathbf{k-1}}$. Initially, the state of the system is $s(0) = (0, \ldots, 0)$, reflecting the fact that there are no open bins.

Now suppose the state of the system at time t is $s(t)$ and the next item to be inserted is ℓ, where ℓ has been chosen uniformly from the set $\{1, \ldots, j\}$. Let i be the smallest index such that $i \geq \ell$ and $s_i(t) > 0$, if such exists: in this case, the algorithm inserts item ℓ into a bin with capacity i, so we have $s_i(t+1) = s_i(t)-1$ and, if $i > \ell$, $s_{i-\ell}(t+1) = s_{i-\ell}(t)+1$; all other components of $s(t)$ are unchanged. If no such i exists, then the algorithm inserts item ℓ into an empty bin, so we have $s_{k-\ell}(t+1) = s_{k-\ell}(t)+1$ and all other components of s are unchanged. This completes the description of the dynamical system.

Note that the above system is nothing other than a convenient pictorial representation of a multi-dimensional Markov chain, with state space $\mathbf{Z}_+^{\mathbf{k-1}}$, in which token i executes a random walk in dimension i. The motions of individual tokens are, of course, not independent. However, the transition probabilities of any given token at any time depend only on which of the tokens are at zero at that time, i.e., on the set $\{i : s_i = 0\}$. This is an important property which makes analysis of the chain feasible.

Our goal is to investigate the behavior of the *waste* in the algorithm after packing t items, which is just the quantity $\sum_{i=1}^{k-1} i s_i(t)$. (The weights i appear in this sum because we have scaled the bin sizes from 1 to k.) In particular, we will be concerned with determining, for given pairs (j, k), whether or not the expected waste remains bounded for an infinite stream of items, i.e., as $t \to \infty$. As explained in the Introduction, we will focus on the case where $j = k-2$, and we assume this relationship from now on.

2.2 Classification of tokens

It will be convenient for us to partition the tokens into two classes, which we will call "large" and "small." This idea is motivated by the fact that tokens behave in two distinct ways, as we shall see in a moment. The *small* tokens are tokens i with $1 \leq i \leq \lceil \frac{j}{2} \rceil$. The *large* tokens are tokens $k-i$ with $1 \leq i \leq \lceil \frac{j}{2} \rceil$. Note that the numbers of small and large tokens are equal. In the case that j is even there is actually an additional token, namely $\lceil \frac{j}{2} \rceil + 1$, which is neither small nor large: we call this the *middle* token.

We first establish a fundamental constraint on the states that are reachable from the initial state $s(0)$. This fact is implicit in [4]; the proof, which is a straightforward induction on time, is left as an instructive exercise for the reader.

Proposition 2 *State s is reachable from the initial state $s(0)$ only if*

1. *For distinct indices i and i' with $i + i' \geq k$, either $s_i = 0$ or $s_{i'} = 0$. (I.e., no two tokens whose index sum is k or greater can simultaneously be at non-zero positions.)*

2. $\sum_{i \text{ not small}} s_i \leq 1$. *(I.e., the large and middle tokens cannot move beyond position 1; moreover, at most one of them can be away from 0 at any time.)* □

It is not hard to see that all states satisfying the conditions of Proposition 2 are in fact reachable from the initial state. From now on, we shall therefore assume that the state space of our Markov chain is precisely this set S of reachable states.

The above proposition expresses general constraints on the motions of the tokens. In the following three subsections, we establish further properties of the behavior of tokens under certain assumptions about the distribution of other tokens. These properties will be used in our analysis in the next section.

2.3 Behavior of large and middle tokens

We have already seen that the large and middle tokens behave in an extremely restricted fashion: namely, they can take on values only 0 and 1, and at most one of them can be non-zero at any time. Their behavior becomes even more restricted under the condition that $s_{\lceil j/2 \rceil} > 0$. This condition will arise naturally in our analysis in the next section.

Proposition 3 *Suppose that $s_{\lceil j/2 \rceil}$ remains strictly positive throughout some time interval. Then during this interval:*

- *all large tokens remain at 0; and*
- *the middle token (if it exists) oscillates between 0 and 1 independently of the positions of all other tokens.*

Proof. The first claim is immediate from condition 1 of Proposition 2. To see the second claim, note that, because all larger tokens are at 0, insertions of item $\lceil \frac{j}{2} \rceil + 1$ are placed alternately in an empty bin (thus creating a bin with capacity $\lceil \frac{j}{2} \rceil + 1$) and in this newly opened bin. No other insertions can affect token $\lceil \frac{j}{2} \rceil + 1$, which therefore oscillates between 0 and 1 as claimed. □

In view of the second claim of Proposition 3, we will assume from now on that j is odd, so that there is no middle token to worry about. This assumption is justified because our analysis will hinge on the behavior of the system when $s_{\lceil j/2 \rceil} > 0$; but Proposition 3 then tells us that the behavior of the middle token under this condition is degenerate. With this observation, the argument we will give for j odd trivially extends to the case when j is even.

2.4 Behavior of small tokens

Most of this paper is concerned with the detailed behavior of the small tokens: since the other tokens remain very severely bounded, it is really only the small tokens that are interesting from the point of view of the asymptotic behavior of the algorithm. In the next proposition, we isolate an essential feature of the motion of the small tokens under a certain condition that will again arise naturally from our analysis in the next section.

Proposition 4 *The motion of a small token i has the following properties:*

Property A *Whenever $s_{i-1} > 0$, the motion of s_i at all positions other than 0 is a (non-time-homogeneous) random walk on $\mathbf{Z}_+$ with non-negative drift and holding probability at most $1 - \frac{2}{j}$.*

Property B *The time spent by s_i on each visit to 0 is stochastically dominated*[†] *by a random variable D with constant expectation (that depends only on j).*

Proof. Consider first the case when $s_i > 0$, and assume also that $s_{i-1} > 0$. Since $s_{i-1} > 0$, the only way in which s_i can decrease is through the insertion of item i. On the other hand, s_i will certainly increase on insertion of item $k-i$; to see this, note from condition 1 of Proposition 2 that $s_{i'} = 0$ for all $i' \geq k-i$, so the algorithm must insert item $k-i$ into an empty bin. Hence s_i decreases with probability $\frac{1}{j}$ and increases with probability at least $\frac{1}{j}$, which is exactly equivalent to Property A.

Now consider what happens when $s_i = 0$. If $s_{i'} = 0$ for all $i' \geq k-i$, then as above we can conclude that s_i moves to 1 with probability at least $\frac{1}{j}$. However, now we cannot exclude the possibility that $s_{i'} = 1$ for some $i' \geq k-i$, in which case item $k-i$ will be inserted into the bin with capacity i' so s_i cannot leave 0. On the other hand, in this situation we see that two consecutive insertions of item $k-i$ will certainly have the effect of moving s_i to 1. This crude argument indicates that the time spent by token i at 0 is stochastically dominated by the random variable D defined as follows over the sequence of item insertions immediately following the arrival of s_i at 0:

$$D = \begin{cases} 1 & \text{if first insertion is } k-i; \\ N & \text{otherwise,} \end{cases}$$

where N is the number of insertions until the first pair of consecutive insertions of $k-i$ has occurred. Notice that the events that item $k-i$ is inserted at time t are mutually independent for all t, and all have probability $\frac{1}{j}$. Hence it is easy to see that the tail of D has the form $\Pr[D > n] \leq \beta^n$ for some constant $\beta < 1$ that depends only on j. This in turn implies that the expectation of D is bounded above by a constant that depends only on j. □

[†]Recall that a random variable X is *stochastically dominated by* a random variable Y if $\Pr[X \geq r] \leq \Pr[Y \geq r]$ for all r.

2.5 Behavior of the waste

Next we investigate what happens to the waste, again under the assumption that $s_{\lceil j/2\rceil} > 0$. Define $f(t) = \sum_{i=1}^{\lceil j/2\rceil} i s_i(t)$, which is just the waste due to the small tokens at time t. By Proposition 2, the total waste is bounded above by $f(t)+k-1$. The following proposition shows that $f(t)$ has negative drift under our assumption about $s_{\lceil j/2\rceil}$.

Proposition 5 *Suppose that* $s_{\lceil j/2\rceil}(t) > 0$. *Then* $\mathrm{E}[f(t+1) - f(t) \mid f(t)] = -1/j$.

Proof. We consider the change in f under the insertion of each possible item i. There are two cases.

Case 1: $1 \le i \le \lceil \frac{j}{2}\rceil$. Suppose first that $s_i > 0$. Then the item is inserted into a bin of capacity i, so s_i decreases by 1 and the change in f is $-i$. Suppose on the other hand that $s_i = 0$. Then the item is inserted into a bin of capacity ℓ for some $i < \ell \le \lceil \frac{j}{2}\rceil$, so s_ℓ decreases by 1 and $s_{\ell-i}$ increases by 1, and the net change in f is again $-i$.

Case 2: $i = k - i'$ with $2 \le i' \le \lceil \frac{j}{2}\rceil$. In this case, since $s_{\lceil \frac{j}{2}\rceil} > 0$, Proposition 2 implies that there can be no open bin large enough to accommodate the item, so it must be inserted into an empty bin. As a result $s_{i'}$ increases by 1, and the change in f is $+i'$.

Note that cases 1 and 2 together cover all items since we are assuming that j is odd. Putting together the above cases, we see that the average change in f over all item insertions is

$$\frac{1}{j}\left(\sum_{i=1}^{\lceil \frac{j}{2}\rceil}(-i) + \sum_{i'=2}^{\lceil \frac{j}{2}\rceil} i'\right) = -\frac{1}{j},$$

as claimed. $\square$

3 Analysis of the Markov chain

This section is devoted to proving our main result, Theorem 1 stated in the Introduction. Our proof makes use of the following result of [5], which establishes a general condition, in terms of the existence of a suitable Lyapunov function, for a multi-dimensional Markov chain to be ergodic. For more specialized variations on this theme, see [7, 11, 8, 4].

Lemma 6 [5, Corollary 7.1.3] *Let $\mathcal{M}$ be an irreducible, aperiodic Markov chain with state space $S \subseteq \mathbf{Z}^{\mathbf{k}}$, and b a positive integer. Denote by $p^b_{ss'}$ the transition probability from s to s' in $\mathcal{M}^b$, the b-step version of $\mathcal{M}$. Let $\Phi : S \to R_+$ be a non-negative real-valued function on S which satisfies the following conditions:*

1. *There are constants $C_1, \mu > 0$ such that $\Phi(s) > C_1 \|s\|^\mu$ for all $s \in S$.*
2. *There is a constant $C_2 > 0$ such that $p^b_{ss'} = 0$ whenever $|\Phi(s) - \Phi(s')| > C_2$, for all $s, s' \in S$.*
3. *There is a finite set $B \subset S$ and a constant $\epsilon > 0$ such that $\sum_{s' \in S} p^b_{ss'}(\Phi(s') - \Phi(s)) < -\epsilon$ for all $s \in S \setminus B$.*

Then $\mathcal{M}$ is ergodic with stationary distribution π satisfying $\pi(s) < C\mathrm{e}^{-\delta\Phi(s)}$ for all $s \in S$, where C and δ are positive constants. $\square$

To interpret this lemma, view Φ as a potential function that maps the state space to the non-negative reals, so that the image of the Markov chain under Φ becomes a dynamical system on the real line. Condition 1 requires that Φ grows polynomially with $\|s\| = \sum_i s_i$, while condition 2 requires that the process is well-behaved in the sense that it has bounded variation. The key is condition 3, which says that, except for a finite set of states, Φ has negative drift over an interval of some constant length b. This implies that $\mathcal{M}$ is ergodic with a stationary distribution that decays at least exponentially with Φ.

In our application, $\mathcal{M}$ will be the Markov chain that governs the movements of the tokens, whose state space is the subset S of $\mathbf{Z}_+^{\mathbf{k-1}}$ defined by Proposition 2, and Φ will be the function $\Phi(s) = \sum_{i=1}^{\lceil j/2\rceil} i s_i + k - 1$. Note that Φ is just $f+k-1$, where f is the waste function appearing in Proposition 5. Since $k-1$ is an upper bound on the waste due to the large and middle tokens, $\Phi(s)$ is an upper bound on the total waste in any state s. It is easy to check that conditions 1 and 2 hold for this Φ, with any choice of constant b. All our work will be devoted to proving the negative drift condition 3, for suitably chosen b, B and ϵ. Theorem 1 will then follow immediately, since asymptotically the expected waste is at most $\sum_{s\in S} \pi(s)\Phi(s)$, which by Lemma 6 is bounded.

The following is an informal sketch of our strategy for proving condition 3:

(i) We consider an interval of length b, and show that Φ has negative drift over this interval provided it is large enough at the start of the interval: i.e., we will take B to be the finite set of points on which Φ is "small." Thus for $s \in S \setminus B$, we can be sure that, for some small token i, s_i is large at the start of the interval, and hence positive throughout the interval.

(ii) Since $s_i > 0$ throughout the interval, by Proposition 4 the motion of s_{i+1} is a random walk with non-negative drift: hence the time that s_{i+1} spends at 0 during the interval is dominated by the time spent at 0 by a symmetric random walk, which is small (about const $\times \sqrt{b}$).

(iii) Iterating this argument, appealing to Proposition 4 each time, we can conclude that each of the tokens $s_{i+2}, \ldots, s_{\lceil j/2\rceil}$ spends little time at 0.

(iv) Finally, since we have established that $s_{\lceil j/2 \rceil} > 0$ during most of the interval, Proposition 5 tells us that f (and hence Φ) has negative drift on most steps, and hence an overall negative drift over the entire interval.

The tricky part of the above argument is step (iii): at each stage we need to use the fact that $s_{i'} > 0$ to deduce from Proposition 4 that $s_{i'+1}$ has non-negative drift. However, occasionally $s_{i'}$ will be at 0, and at these times we have no control over the motion of $s_{i'+1}$. We therefore assume that $s_{i'+1}$ has non-negative drift most of the time, but that an adversary is able to bias its motion on a small number of steps. Accordingly, we need to prove a lemma that quantifies the effect that such an adversary can have on the amount of time $s_{i'+1}$ spends at 0. This we now do. Actually, for simplicity we will analyze the effect of such an adversary on a *symmetric* random walk: since $s_{i'+1}$ has non-negative drift, it is clear that the time it spends at 0 will be stochastically dominated by the symmetric walk (see Proposition 9 below for a precise statement).

So, consider a symmetric random walk of a given length on $\mathbf{Z}_+$, started at some specified position, and an adversary whose goal is to maximize the number of times the walk hits 0.[†] The adversary is allowed to intervene at some specified number of steps, selected according to any strategy: on these steps, the adversary may specify any desired probability distribution on the legal moves of the process from the current state; on all other steps, the process behaves as a symmetric random walk with a perfectly reflecting barrier at 0. (Note that, since the only legal move from 0 is to 1, the adversary is not able to intervene at 0.)

It is perhaps not surprising that the optimal strategy for the adversary is always to intervene by driving the process deterministically towards the origin, and to use up all these interventions as early as possible. However, this claim requires some justification, which we now provide.

Lemma 7 *Let $p(i,n,y,m)$ be the probability that a symmetric random walk of n steps, starting at i and with y adversary steps, hits the origin at least m times. Let $q(i,n,y,m)$ be the same quantity for the particular adversary strategy in which downward steps are used as early as possible. Then $p(i,n,y,m) \le q(i,n,y,m)$ for all i,n,y,m.*

To prove Lemma 7, we need a simple technical observation about symmetric random walk.

Proposition 8 *Let $p_0(i,n,m) = p(i,n,0,m)$ denote the probability that a symmetric random walk of length n started at i hits 0 at least m times. Then $p_0(2,n,m-1) \ge p_0(1,n+1,m)$.*

[†]Throughout, for convenience, we shall take "hits 0" to mean "makes the transition $0 \to 1$."

Proof. Let W_2 be the random walk started from 2 and W_1 the random walk started from 1. Consider the first time T when W_2 reaches 1, and the first time T' when W_1 reaches 1 by a $0 \to 1$ transition. Then it is easy to see that T' is equal to $T+1$ in distribution. The remainder of W_2 after T is a random walk started at 1 which must have at least $m-1$ hits. The remainder of W_1 after T' is also a random walk started at 1 which must have at least $m-1$ hits. Thus if $m > 1$, the probabilities are the same for both walks (conditioning on the event $T = T' - 1 \le n$), and if $m = 1$, we trivially have $p_0(2,n,0) = 1 \ge p_0(1,n+1,1)$. □

We now prove Lemma 7. We will consider only deterministic strategies: the randomized case follows by averaging.

Proof of Lemma 7. We use induction on n. Let $D^y R$ denote the strategy which uses the y forced down steps as soon as possible, and then follows the symmetric random walk. Let RD^y denote the strategy which starts with a truly random step as soon as possible, and then uses the y forced down steps as soon as possible. Notice that a transition from 0 to 1 is neither a forced down step (obviously) nor a truly random step, since it has probability 1. Let $q(i,n,y,m)$ denote the probability that the walk of length n defined by the strategy $D^y R$, started at i and with y forced down steps has at least m hits; let $r(i,n,y,m)$ be defined similarly for the strategy RD^y. We claim that

$$q(i,n,y,m) \ge r(i,n,y,m) \qquad (1)$$

Note that the lemma will then follow by induction on n: consider the first time the adversary may intervene. Either way, after this step we are left to deal with fewer than n steps. If the adversary does force a down step, by induction the best strategy to continue is $D^{y-1}R$, so the strategy for the entire walk is $D^y R$. If the adversary does not intervene, using induction again, the best strategy to continue is $D^y R$, so the strategy for the entire walk is $RD^y R = RD^y$. Inequality (1) shows that the strategy $D^y R$ is better than RD^y.

To prove (1), we also use induction on n. If $i = 0$, both strategies start the same way and we are done by induction. If $i \ge y+1$, both strategies give the same distribution of positions after $y+1$ steps, and neither has hit 0 yet, so the two quantities are equal. The interesting case is for $1 \le i \le y$. Then, let $n' = n - (2y - i + 2)$ and $m' = m - (y - i + 1)$. It is easy to see that

$$q(i,n,y,m) = \tfrac{1}{2} p_0(0,n',m') + \tfrac{1}{2} p_0(2,n',m')$$

and

$$r(i,n,y,m) = \tfrac{1}{2} p_0(0,n',m') + \tfrac{1}{4} p_0(0,n',m'+1) + \tfrac{1}{4} p_0(2,n',m'+1).$$

Since $p_0(1,n'+1,m'+1) = \frac{1}{2}\big(p_0(0,n',m'+1) + p_0(2,n',m'+1)\big)$, the difference is

$$q - r = \tfrac{1}{2}\Big(p_0(2,n',m') - p_0(1,n'+1,m'+1)\Big),$$

which is non-negative by Proposition 8. □

Recall that we simplified our adversary argument by considering symmetric random walk rather than the more general random walk with non-negative drift and holding probabilities that is actually executed by the tokens according to Property A of Proposition 4. However, it should be intuitively clear that this simplification can only increase the number of hits on 0. To make this precise, consider an arbitrary stochastic process over the non-negative integers. Assume it has arbitrary holding probabilities except at 0, where the holding probability is zero, and non-negative drift everywhere. Let Z denote the number of hits on 0 during the first T steps of this process. Let U be the similar quantity for the symmetric random walk with zero holding probabilities and perfectly reflecting barrier at 0, starting at the same point. The following fact can be proved by a simple coupling argument, which we omit.

Proposition 9 *Z is stochastically dominated by U.* □

We are now in a position to proceed with the proof of our main result, following the sketch given after Lemma 6. As observed there, the main difficulty lies in step (iii): assuming that $s_i > 0$ throughout some interval, we want to conclude that $s_{\lceil j/2\rceil} > 0$ during most of that interval. This is the subject of the next lemma, which makes essential use of our adversary result, Lemma 7.

Lemma 10 *Let T and a be positive constants, and suppose that $s_i(0) > T$ for some $i \le \lceil \frac{j}{2}\rceil$. With probability at least $1-C/a$, where C is a constant that depends only on j, $s_{\lceil j/2\rceil}(t)$ is strictly positive at all but at most $a\sqrt{T}$ time instants t within the interval $[0,T]$.*

Proof. We will prove the claim for $i=1$; as will become apparent, the proof for general i is exactly the same. So, assume that $s_1(0) > T$. For each i, let the random variable T_i denote the time spent at 0 by token i during the interval $[0,T]$. Clearly $T_1 = 0$ with probability 1.

Next let us consider the behavior of the sequence s_2. Consider a modified process s_2' which is defined as follows. First, run s_2 for T steps. Then, have the token s_2' follow a symmetric random walk with a holding time at 0 distributed according to D. Finally, delete from s_2' all stationary steps at 0. Let T_2' be the number of hits on 0 of s_2' during the time interval $[0,T]$. Then we have

$$T_2 \le \sum_{r=1}^{T_2'+1} D_r, \tag{2}$$

where the D_r are i.i.d. with the same distribution as D, and T_2' is independent of all of the D_r. (Here $\le$ denotes stochastic domination between the random variables.)

To analyze T_2', we compare it with U, the number of hits on 0 of symmetric random walk with no holding probabilities and perfectly reflecting barrier at 0. By Proposition 9, T_2' is stochastically dominated by U. Taking expectations in (2) and using this observation, we get

$$\begin{aligned} \mathrm{E}T_2 &\le \mathrm{E}\Bigg[\sum_{r=1}^{T_2'+1} D_r\Bigg] \\ &= \sum_{t=0}^{\infty}\sum_{r=1}^{t+1} \mathrm{E}[D_r \mid T_2' = t]\,\Pr[T_2' = t] \\ &\le (\mathrm{E}U + 1)\mathrm{E}D \\ &= d(\mathrm{E}U+1)\,, \end{aligned} \tag{3}$$

where the constant d is the expectation of D from Proposition 4.

Now consider token s_ℓ, where $3 \le \ell \le \lceil \frac{j}{2}\rceil$. Define a modified process s_ℓ' and a random variable T_ℓ' in exactly similar fashion to s_2' and T_2'. By analogy with (2) we may write

$$T_\ell \le \sum_{r=1}^{T_\ell'+1} D_r. \tag{4}$$

Now our adversary argument, Lemma 7, implies that T_ℓ' is stochastically dominated by $U + T_{\ell-1}$. To see this, note that

$$\begin{aligned} \Pr[T_\ell' \ge m \mid T_{\ell-1} = y] &\le p(s_\ell(0), T, y, m) \\ &\le q(s_\ell(0), T, y, m) \\ &\le \Pr[U \ge m - y]. \end{aligned}$$

Taking expectations in equation (4), and using this fact, we get

$$\mathrm{E}T_\ell \le (\mathrm{E}U + \mathrm{E}T_{\ell-1} + 1)\mathrm{E}D = d(\mathrm{E}U + \mathrm{E}T_{\ell-1} + 1).$$

Iterating this bound, and using the base case (3), gives

$$\mathrm{E}T_\ell \le \left(\sum_{r=1}^{\ell-2} d^r\right)(\mathrm{E}U+1) + d^{\ell-2}\mathrm{E}T_2 \le \ell d^\ell(\mathrm{E}U+1).$$

But $\mathrm{E}U \le c\sqrt{T}$ for some universal constant c. Hence by Markov's inequality

$$\Pr\big[T_{\lceil j/2\rceil} > a\sqrt{T}\big] < \frac{\lceil j/2\rceil d^{\lceil \frac{j}{2}\rceil}(c + 1/\sqrt{T})}{a},$$

which is bounded above by C/a for some constant C as required. □

Remarks: (a) The above proof actually demonstrates the stronger conclusion that $s_{i'}(t) > 0$ for all i' in the range $i \le i' \le \lceil \frac{j}{2}\rceil$, for a similar majority of the interval.

(b) It is interesting to note that the only properties of the sequence $s(t)$ we have used in the above proof are properties A and B of Proposition 4. Thus Lemma 10 actually applies to any sequence of vector random variables satisfying these rather natural properties. We believe that this fact may be of independent interest.

We are finally in a position to complete the proof of Theorem 1, following our earlier sketch.

Proof of Theorem 1. Recall from the discussion immediately following the statement of Lemma 6 that it suffices to establish condition 3 of the lemma for the function $\Phi(s) = \sum_{i=1}^{\lceil j/2 \rceil} is_i + k - 1$, with suitable choices of b, B and ϵ. The set $B \subseteq S$ will be defined as

$$B = \{s \in S : \forall i, s_i \leq T\},$$

where T is some constant to be specified shortly.

Assume that $s = s(0) \in S \setminus B$, i.e., that for some coordinate i we have $s_i(0) > T$; necessarily $i \leq \lceil \frac{j}{2} \rceil$. Consider the time interval $[0, T]$. We will show that the expected drift of Φ over this interval is less than $-\epsilon$ for some $\epsilon > 0$, thus establishing condition 3. Since we are analyzing drift, we may equivalently work with the function f of Proposition 5, which differs from Φ only by a constant.

By Lemma 10, with probability at least $1 - C/a$ we have $s_{\lceil j/2 \rceil} > 0$ at all but at most $a\sqrt{T}$ time instants in the time interval $[0, T]$.

Now consider the change Δf in f after one step. By Proposition 5, if $s_{\lceil j/2 \rceil} > 0$ then $\mathrm{E}\Delta f \leq -1/j$. In all other situations, there is the trivial bound $\mathrm{E}\Delta f \leq j$. Putting these facts together, and conditioning on the event $\mathcal{A}$ that $s_{\lceil j/2 \rceil} > 0$ at all but at most $a\sqrt{T}$ time instants in the interval, we see that the drift of f over the entire interval $[0, T]$ is

$$\begin{aligned} &\mathrm{E}[f(b) - f(0) \mid f(0)] \\ &\quad \leq \mathrm{E}[f(b) - f(0) \mid f(0) \wedge \mathcal{A}] \\ &\qquad + (1 - \Pr[\mathcal{A}])\, \mathrm{E}[f(b) - f(0) \mid f(0) \wedge \neg\mathcal{A}] \\ &\quad \leq \left\{ -\frac{1}{j}\left(T - a\sqrt{T}\right) + ja\sqrt{T} \right\} + \frac{Cj}{a} T. \end{aligned}$$

By taking $a = 2Cj^2$ large enough, and then T large enough, we can make this expression less than some negative constant $-\epsilon$.

This completes the verification of condition 3, and hence the proof of the theorem. $\square$

Acknowledgments

We thank David Aldous and Ashwin Nayak for helpful comments on an earlier draft of this paper.

References

[1] Y. Azar, A.Z. Broder, A.R. Karlin, N. Linial and S. Phillips. Biased random walks. In *Proceedings of the 24th Annual ACM Symposium on Theory of Computing*, 1992, pp. 1–9.

[2] 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 *Proceedings of the 23rd Annual ACM Symposium on Theory of Computing*, 1991, pp. 230–240.

[3] E.G. Coffman Jr., M.R. Garey and D.S. Johnson. Approximation algorithms for bin packing: A survey. To appear in *Approximation Algorithms for NP-Hard Problems*, D.S. Hochbaum, ed., PWS Publishing Company, Boston, 1996.

[4] E.G. Coffman Jr., D.S. Johnson, P.W. Shor and R.R. Weber. Markov chains, computer proofs and average-case analysis of Best Fit bin packing. In *Proceedings of the 25th Annual ACM Symposium on Theory of Computing*, 1993, pp. 412–421.

[5] G. Fayolle, V.A. Malyshev and M.V. Menshikov. *Constructive Theory of Countable Markov Chains, Part* I. Manuscript, May 1992.

[6] W. Feller. *An Introduction to Probability Theory and its Applications, Volume* I (3rd ed.) John Wiley, New York, 1968.

[7] F.G. Foster. On stochastic matrices associated with certain queueing processes. *Annals of Mathematical Statistics* **24** (1953), pp. 355–360.

[8] B. Hajek. Hitting-time and occupation-time bounds implied by drift analysis with applications. *Advances in Applied Probability* **14** (1982), pp. 502–525.

[9] D.S. Johnson, A. Demers, J.D. Ullman, M.R. Garey and R.L. Graham. Worst case performance bounds for simple one-dimensional packing algorithms. *SIAM Journal on Computing* **3** (1974), pp. 299–325.

[10] F.T. Leighton. Average case analysis of greedy routing algorithms on arrays. In *Proceedings of the 2nd Annual ACM Symposium on Parallel Algorithms and Architectures*, 1990, pp. 2–10.

[11] V.A. Malyshev and M.V. Menshikov. Ergodicity, continuity and analyticity of countable Markov chains. *Transactions of the Moscow Mathematical Society* **39** (1979), pp. 3–48.

[12] P.W. Shor. The average case analysis of some on-line algorithms for binpacking. *Combinatorica* **6** (1986), pp. 179–200.

[13] J. Walrand. *An introduction to queuing networks.* Prentice Hall, Englewood Cliffs, New Jersey, 1988.