Acta Cybernetica 15 (2002) 361–367.

Closed On-Line Bin Packing

E. Asgeirsson[*], U. Ayesta[†], E. Coffman[‡], J. Etra[§],
P. Momčilovič[‡], D. Phillips[*], V. Vokhshoori[‡], Z. Wang[¶]
and J. Wolfe[‡]

Abstract

An optimal algorithm for the classical bin packing problem partitions (packs) a given set of items with sizes at most 1 into a smallest number of unit-capacity bins such that the sum of the sizes of the items in each bin is at most 1. Approximation algorithms for this NP-hard problem are called *on-line* if the items are packed sequentially into bins with the bin receiving a given item being independent of the number and sizes of all items as yet unpacked. *Off-line* algorithms plan packings assuming full (advance) knowledge of all item sizes. The *closed* on-line algorithms are intermediate: item sizes are not known in advance but the number n of items is. The uniform model, where the n item sizes are independent uniform random draws from [0,1], commands special attention in the average-case analysis of bin packing algorithms. In this model, the expected wasted space produced by an optimal off-line algorithm is $\Theta(\sqrt{n})$, while that produced by an optimal on-line algorithm is $\Theta(\sqrt{n \log n})$. Surprisingly, an optimal closed on-line algorithm also wastes only $\Theta(\sqrt{n})$ space on the average. A proof of this last result is the principal contribution of this paper. However, we also identify a class of optimal closed algorithms, extend the main result to other probability models, and give an estimate of the hidden constant factor.

1 Introduction

An instance of the one-dimensional bin packing problem is a list $L_n = \langle a_1, a_2, \ldots, a_n \rangle$ of *items* that must be packed into, i.e., partitioned among, a minimum-cardinality set of *bins* $B_1, B_2, \ldots$ subject to the constraint that the set of items in any bin fits within that bin's capacity. In the usual way, we will take the bin capacity to be 1 for convenience, so a set of items fits into a bin if and only if the item sizes sum to no more than 1. The unused space in B_i is called a *gap*

[*]Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027
[†]INRIA, 2004 Route des Lucioles, Sophia Antipolis Cedex, France
[‡]Department of Electrical Engineering, Columbia University, New York, NY 10027
[§]Columbia Law School, Columbia University, New York, NY 10027
[¶]Graduate School of Business, Columbia University, New York, NY 10027

and is denoted by g_i. The sum of the gaps in the occupied bins of a packing is the *wasted space* of the packing.

The bin packing problem has countless applications in operations research and engineering. To name just a few, we mention storage allocation for computer networks, assigning advertisements to newspaper columns, assigning commercials to station breaks on television, writing a collection of files to several floppy disks, packing trucks with a given weight limit, and the cutting-stock problems of various industries like those producing lumber and cable.

Let A denote an arbitrary approximation algorithm for the NP-hard bin packing problem and let OPT denote an algorithm that produces optimal packings. Let $A(L_n)$ and $\mathrm{OPT}(L_n)$ denote the numbers of bins used by algorithms A and OPT. In the classical analysis of bin packing approximation algorithms, combinatorial methods are used to derive worst-case performance ratios

$$R_A := \sup_L \{A(L)/\mathrm{OPT}(L)\}$$

and their asymptotic variants. Less often, probabilistic studies that are typically quite difficult are conducted in order to obtain average-case performance. The average-case approach is followed in this paper. *The item-size distribution is taken to be the uniform distribution on* $[0,1]$, denoted as usual by $U(0,1)$. This is the distribution of choice in bin packing analysis, along with the assumption that item sizes are independent. For general coverage of the probabilistic analysis of bin packing algorithms, see the monograph by Coffman and Lueker(1991).

A bin packing algorithm is called *on-line* if it packs every item a_i solely on the basis of the sizes of the items a_j, $1 \le j \le i$, i.e., without any information on subsequent items. The decisions of an on-line algorithm are irrevocable; packed items cannot be repacked at later times. Two classical on-line algorithms are First Fit and Best Fit. Each of these algorithms begins by putting a_1 into B_1. Thereafter, First Fit places the next item into the lowest indexed (first) gap no smaller than the item, and Best Fit puts the next item into a smallest gap no smaller than the item with ties resolved in favor of the lowest indexed bins.

A bin packing algorithm that can use full knowledge of all items in packing L_n is called *off-line*. One of the first results in the average-case theory was a proof by Lueker(1982) that an optimal off-line algorithm has the following asymptotic bound.

$$E\mathrm{OPT}_{offline}(L_n) = \frac{n}{2} + \Theta(\sqrt{n}).$$

where $\Theta(\sqrt{n})$ bounds the expected wasted space, since the expected total item size gives the $n/2$ term. More recently, Shor(1991) proved that on-line packings must produce greater expected wasted space by at least a log factor. In particular, he showed that

$$E\mathrm{OPT}_{online}(L_n) = \frac{n}{2} + \Theta(\sqrt{n \log n}).$$

Although there is no known simple algorithm for achieving this bound, the Best

Fit (BF) algorithm comes close in that (see Shor(1986))

$$E\mathrm{BF}(L_n) = \frac{n}{2} + \Theta(\sqrt{n}\log^{3/4} n)$$

The *closed* on-line algorithms are intermediate between the classes of on-line and off-line algorithms: item sizes are not known in advance, but the number n of items is. As noted by Shor(1986), it is surprising that one can produce an algorithm that achieves $O(\sqrt{n})$ expected wasted space without knowing item sizes in advance. However, the algorithm must know n, i.e., it must be a closed on-line algorithm. According to one such algorithm, which we call Closed Best Fit (CBF), the first $\lceil n/2 \rceil$ items are packed one to a bin and the remaining $\lfloor n/2 \rfloor$ items are packed by Best Fit. The claim is that CBF wastes at most $O(\sqrt{n})$ space on average, and so the following bound on closed on-line packing holds.

Theorem 1

$$E\mathrm{OPT}_{closed}(L_n) = \frac{n}{2} + \Theta(\sqrt{n})$$

We have seen no proof of this result, and while it is true that standard techniques may be applied in such a proof, the way in which they are applied has novel features. For this reason, and because the improvement possible in closed on-line bin packing is indeed unexpected, the next section sets down for the record a proof of Theorem 1. Still more reasons are provided by the additional results to which the analysis leads. For example, we derive a compact upper bound on the hidden constant factor from the analysis of a random walk. Further, as discussed in Section 3, Theorem 1 will be seen to apply to a number of practical matching algorithms, and to be extendible to distributions other than the uniform.

2 Proof of Theorem 1

For convenience, we assume hereafter that n is even; this will not affect our asymptotic results. Let $L_n^{(1)}$ and $L_n^{(2)}$ be the sublists of the first $n/2$ and last $n/2$ items of L_n, respectively. We begin by proving $O(\sqrt{n})$ wasted space for the modification of CBF which closes any bin B_j, $j \le n/2$, after it receives a second item, and closes any bin B_j, $j > n/2$, after it receives its first item. Denote the modified algorithm by CBF$_*$. An example is shown in Figure 1(a). After proving that Theorem 1 holds for CBF$_*$, we will show that $\mathrm{CBF}_*(L_n) \ge \mathrm{CBF}(L_n)$ for all L_n, thus completing the proof of Theorem 1.

We begin with a key property of CBF$_*$ packings.

Lemma 1 *Let L_n and L'_n differ only in the permutations of their last $n/2$ items. Then $CBF_*(L_n) = CBF_*(L'_n)$.*

Proof. Consider the *ordered* CBF$_*$ packing of L_n in which the bins are arranged so that the first $n/2$ items are in decreasing size order, as illustrated in Figure 1(b). We say that this packing is a *canonical* packing if in addition the last $n/2$ items

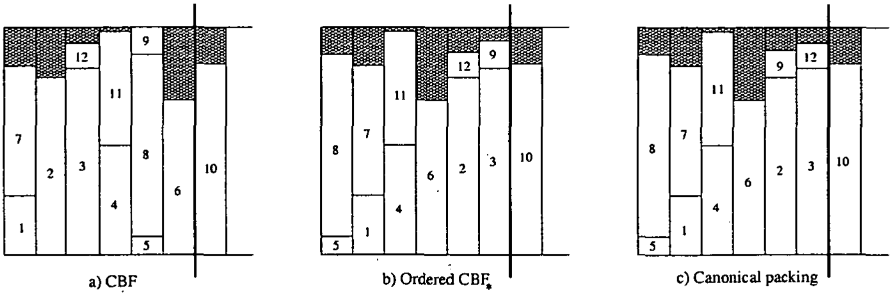

Figure 1: An example with items 1 through $n = 12$ having sizes

$$.26, .78, .82, .48, .08, .68, .57, .8, .12, .84, .5, .11.$$

are in *increasing* size order. Roughly speaking, CBF$_*$ attempts to pack bins with matched items, one large and one small. However, ordered matchings (packings) are not necessarily canonical unless $L_n^{(2)}$ is in increasing size order. For example, in Figure 1(b) a canonical matching requires that items a_9 and a_{12} be interchanged.

On the other hand, the CBF$_*$ packing can be put into canonical form without changing CBF$_*(L_n)$. To see this, suppose items i_1, i_2 are in $L_n^{(1)}$ and matched with j_1, j_2, respectively, in bins of the ordered CBF$_*$ packing. If $a_{j_1} > a_{j_2}$ and $a_{i_1} > a_{i_2}$ then $a_{j_1} + a_{i_2} \le a_{j_1} + a_{i_1} \le 1$ and $a_{j_2} + a_{i_1} \le a_{j_1} + a_{i_1} \le 1$, so we can interchange items j_1 and j_2 without exceeding bin capacity. Iterating these interchanges at most $O(n^2)$ times brings the CBF$_*$ packing into canonical form up through $B_{n/2}$. Trivially, the items in the singleton bins beyond $B_{n/2}$ can then be sorted into increasing order, at which point the entire packing is a canonical packing. In addition, the set of items in singleton bins can not have changed, since the Best Fit rule depends only on gap sizes and not on the bin (gap) indexing. We conclude that the cardinality of the CFB$_*$ packing is left unchanged at the end of the ordering process. It remains only to observe that CBF$_*$ packings for lists L_n and L'_n that differ only in the permutation of their last $n/2$ items will converge under the ordering process to the same equal-cardinality canonical packing. □

We now prove that Theorem 1 holds for CBF$_*$, and in the process, find the hidden constant factor.

Lemma 2

$$E\text{CBF}_*(L_n) - \frac{n}{2} \sim \sqrt{\frac{\pi n}{8}}$$

as $n \to \infty$.

Proof. By Lemma 1 we need consider only canonical CBF$_*$ packings. Let $N_1(y)$ be the number of items in $L_n^{(1)}$ with sizes less than y and let $N_2(y)$ be the number of items in $L_n^{(2)}$ with sizes greater than $1-y$. Define

$$\delta(y) := N_2(y) - N_1(y)$$

and note that $\max_{0\le y\le 1}\delta(y) \ge \delta(0) = 0$. It is easy to see that

$$\mathrm{CBF}_*(L_n) = \frac{n}{2} + \max_{0\le y\le 1}\delta(y). \tag{1}$$

As an example, note that $\max_{0\le y\le 1}\delta(y) = 1$ in Figure 1(b), and that $\delta(y)$ achieves its maximum for any $y \in (1-a_{11}, a_6)$. To verify (1), one can argue in terms of the number of singleton bins beyond $B_{n/2}$ in the CBF_* packing, which is just $\mathrm{CBF}_*(L_n) - n/2$. To the right of the rightmost singleton bin B_j with $j \le n/2$, the number of items from $L_n^{(2)}$ less the number of items from $L_n^{(1)}$ gives the maximum of $\delta(y)$ over $[0,1]$ and is equal to the number of singleton bins beyond $B_{n/2}$.

We now interpret $\delta(y)$ as a random walk that evolves as y increases from 0 to 1. For each size a in $L_n^{(2)}$ a plus is plotted at point a, and for each size a in $L_n^{(1)}$ a minus is plotted at point $1-a$. For each minus encountered as y increases from 0 to 1, $\delta(y)$ steps down by 1, and for each plus encountered, $\delta(y)$ steps up by 1. Let δ_i, $0 \le i \le n$, be the position of this random walk after the ith jump. As can be seen, $\{\delta_i\}$ is a classical n-step symmetric random walk with the constraint that its paths start and end at the origin, i.e., $\delta_0 = \delta_n = 0$. Letting $n = 2\nu$, the number of such paths is $\binom{2\nu}{\nu}$. By the reflection principle (see e.g, Feller(1968), p. 72), the number of such paths that hit or exceed k is $\binom{2\nu}{\nu+k}$, $0 \le k \le \nu$, and so

$$E \max_{0\le i\le 2\nu}\delta_i = \frac{1}{\binom{2\nu}{\nu}} \sum_{1\le k\le \nu} \binom{2\nu}{\nu+k}.$$

By the binomial theorem, the sum evaluates to $\frac{1}{2}(2^{2\nu} - \binom{2\nu}{\nu})$, so routine applications of Stirling's formula yield

$$E \max_{0\le y\le 1}\delta(y) = E \max_{0\le i\le n}\delta_i \sim \sqrt{\frac{\pi n}{8}}$$

as $n \to \infty$. □

We will be done once we have proved

Lemma 3

$$\mathrm{CBF}_*(L_n) \ge \mathrm{CBF}(L_n).$$

Proof. Let $a_{\ell_1}, \ldots, a_{\ell_k}$ be the subsequence of items in $L_n^{(2)}$ that are packed by CBF into bins B_j, $j \le n/2$, that already have at least two items, or bins B_j, $j > n/2$, that already have at least one item. Remove these items from the CBF packing and repack them best-fit into the singleton bins B_j, $j \le n/2$. That is, item a_{ℓ_i} is put into a singleton bin B_j, $j \le n/2$, with the smallest gap no smaller than $1 - a_{\ell_i}$, if such a bin exists; if no such bin exists, a_{ℓ_i} is put into an empty bin (necessarily beyond $B_{n/2}$). In either case, the bin receiving a_{ℓ_i} is then closed. The final packing is a CBF$_*$ packing of a list L'_n that can differ from L_n only in the permutation of the last $n/2$ items. Moreover, the final packing has a cardinality at least that of the original CBF packing. In particular, $\mathrm{CBF}_*(L'_n) - \mathrm{CBF}(L_n) \geq 0$ is the number of new singleton bins produced in the new packing. By Lemma 1 we can then conclude that

$$\mathrm{CBF}_*(L'_n) = \mathrm{CBF}_*(L_n) \geq \mathrm{CBF}(L_n)$$

which proves the lemma. □

3 Final Remarks

Consider the closed on-line algorithm that (i) packs the first $n/2$ items one to a bin, (ii) sorts the bins so that the items are in decreasing order, and (iii) packs the remaining items First Fit. This is the algorithm actually proposed by Shor(1986). Let us call this algorithm Closed First Fit (CFF) and define CFF$_*$ just as we defined the variant CBF$_*$ of CBF (limiting bins to at most two items). In comparing CFF$_*$ and CBF$_*$, we observe that packing best fit is like packing first fit into a decreasing sequence, so the two algorithms give, for all L_n, exactly the same packing.

Theorem 1 is easily generalized to any distribution symmetric around $1/2$ that is not concentrated entirely at $1/2$. Further, we can apply the same ideas to distributions $U(0, 1/p)$, with p an integer. For example, suppose $p = 3$. Then we take $n/6$ bins and divide each into thirds. The top thirds of these bins are packed as before as if they were bins themselves; only the scaling by a factor of 3 has any effect. Similarly, the middle thirds are packed after top thirds and then the bottom thirds are packed last. Bins beyond $B_{n/6}$ are introduced as needed and packed as if they consisted of 3 bins with capacity $1/3$. The extension of Theorem 1 follows easily.

References

[1] Coffman, E. G., Garey, M. R., and Johnson, D. S.: (1995), Approximation algorithms for bin packing: A survey, in *Approximation algorithms for NP-hard problems*, D. Hochbaum (ed.), PWS Publishing Co., Boston.

[2] Coffman, E. G. and Lueker, G. S.: (1991), *Probabilistic Analysis of Packing and Partitioning Algorithms*, Wiley, New York.

[3] Csirik, J., Frenk, J. B. G., Galambos, G., and Rinnooy Kan, A. H. G.: (1991), Probabilistic analysis of algorithms for dual bin packing problems, *J. Algorithms*, **12**, 189–203.

[4] Feller, W.: (1968), *Probability Theory and Its Applications, Vol. 1*, Wiley, New York.

[5] Lueker, G. S.: (1982), An average-case analysis of bin packing with uniformly distributed item sizes, *Technical Report 181*, University of California at Irvine, Department of Information and Computer Science.

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

[7] Shor, P. W.: (1991), How to pack better than Best Fit: Tight bounds for average-case on-line bin packing, *Proc. 32nd Ann. Symp. Found. Comp. Sci.*, 752–759, IEEE Comp. Soc. Press, New York.

Received January, 2002