Comments on the hierarchically structured bin packing problem

Thomas Lambert, Loris Marchal, Bora Uçar

▶ To cite this version:

HAL Id: hal-01071414

https://hal.inria.fr/hal-01071414v2

Submitted on 8 Oct 2014

Comments on the hierarchically structured bin packing problem

Thomas Lambert[a], Loris Marchal[b], Bora Uçar[b,*]

[a]*ENS Lyon, France*
[b]*LIP, UMR5668 (CNRS - ENS Lyon - UCBL - Université de Lyon - INRIA), Lyon, France*

Abstract

We study the hierarchically structured bin packing problem. In this problem, the items to be packed into bins are at the leaves of a tree. The objective of the packing is to minimize the total number of bins into which the descendants of an internal node are packed, summed over all internal nodes. We investigate an existing algorithm and make a correction to the analysis of its approximation ratio. Further results regarding the structure of an optimal solution and a strengthened inapproximability result are given.

1. Introduction

We study a variant of the classical bin packing problem, called the hierarchically structured bin packing (HSBP) problem. In this problem, the items to be packed into bins are the leaves of a tree. The objective of the packing is to minimize the total number of bins into which the descendants of an internal node are packed, summed over all internal nodes. Such a packing problem has applications in document organization and retrieval [1], and sparse matrix computations domain [2]. Both of these papers investigate an approximation algorithm, which is claimed to be a $3/2$ approximation on a variant of the problem. Our main contributions are two folds: (i) to show that this approximation does not hold unless a particular condition is met by the given tree; (ii) to

---

*Corresponding author. Tel: +33 472728932, 46 allée d'Italie, 69007, Lyon, France
Email addresses: thomas.lambert@ens-lyon.fr (Thomas Lambert),
loris.marchal@ens-lyon.fr (Loris Marchal), bora.ucar@ens-lyon.fr (Bora Uçar)

strengthen the result that there is no PTAS (polynomial time approximation scheme) or APTAS (asymptotic PTAS) for a variant of the problem. We also investigate the optimality of a class of solutions.

2. Notation and problem definition

Let T be a rooted tree. We use $T(v)$ to denote the (sub)tree rooted at a node v; $T(v)$ contains v and all of its descendants. The set of the leaf nodes and the set of the internal nodes of a tree T are denoted by $\mathcal{L}(T)$ and $\mathcal{N}(T)$. $\mathcal{P}(\mathcal{L}(T))$, or $\mathcal{P}$ for brevity, is used to denote a partition of $\mathcal{L}(T)$.

The *node dispersal number* of a node v under a given partition $\mathcal{P}(\mathcal{L}(T))$ is

$$\rho(v, \mathcal{P}) = |\{A \in \mathcal{P} : A \cap \mathcal{L}(T(v)) \neq \emptyset\}| \ .$$

This number counts the number of bins in which the leaves in $T(v)$ are partitioned. The total node dispersal number of a partition $\mathcal{P}$ of $\mathcal{L}(T)$ is defined as the sum of the dispersal numbers of the internal nodes

$$\rho(\mathcal{P}) = \sum_{v \in \mathcal{N}(T)} \rho(v, \mathcal{P}) \ . \tag{1}$$

Let $w : \mathcal{L}(T) \to N^+$ be a positive weight function associated with the leaves of T. This is extended to a set of leaf nodes $A \subseteq \mathcal{L}(T)$ in such a way that $w(A) = \sum_{v \in A} w(v)$. Let B be a positive integer such that $B \geq \max_{v \in \mathcal{L}(T)} w(v)$. Then, the HSBP problem asks for a partition $\mathcal{P}$ of the leaves of T such that for each part $A \in \mathcal{P}$ we have $w(A) \leq B$ and the total node dispersal number (1) is minimum. In the following, we use $\mathcal{P}^\star$ to denote an optimal partition for HSBP.

3. Related work and contributions

There are two studies that are of immediate interest [1, 2], dealing with some variants of the HSBP problem. Codenotti et al. [1] investigate two variants in which the leaf nodes have unit and non-unit weights. Amestoy et al. [2] investigate a variant in which the leaf nodes have unit weights and the internal nodes have costs. In this latter variant, the total node dispersal number of

a partition is defined as the *weighted* sum of the node dispersal numbers, i.e., $\rho(\mathcal{P}) = \sum_{v \in \mathcal{N}(T)} c(v) \cdot \rho(v, \mathcal{P})$, where $c(v)$ is the cost of the internal node v. Both studies present NP-completeness results for the problems under investigation.

The two aforementioned studies independently describe a similar algorithm called SIMPLE [1] and POPART [2] for the variant with unit weight leaves. In this algorithm, the leaf nodes of the given tree T are first sorted according to their order in a post-order of T (i.e., in a depth-first traversal of T). The first B leaves are put into the first bin, the following B leaves are put into the second bin and so on so forth. Amestoy et al. show that POPART has an approximation ratio of 2 for their variant (internal nodes have costs, and leaves have unit-weights). Codenotti et al. claim that SIMPLE has an approximation ratio of $3/2$ for unit weights and costs, when each internal node has at least two children. However, we provide examples in Section 4.1 on which SIMPLE obtains an approximation ratio worse than $3/2$ (tends to 2), under the same hypothesis, contradicting to the claimed result. We then show that SIMPLE has an approximation ratio of $3/2$ with a restricted condition on internal nodes.

A curious observation is that in the two existing NP-completeness proofs [1, 2], the optimum value is achieved by a post-order based partition. This implies that the problem of finding the best post-order, e.g., the best initial ordering for the algorithm SIMPLE, is NP-complete as well. This then begs another question. Is there always an optimal, post-order based partition? In Section 4.2, we answer this in the negative: SIMPLE with the best post-order is not necessarily optimal.

Both of the previous studies [1, 2] give inapproximability results. While Amestoy et al. [2] show that there is no $1 + o(1/m)$ approximation for the variant with internal node costs, where m is the number of nodes, Codenotti et al. [1] prove that the structured bin packing problem with non-unit weights on the leaves is not approximable within $3/2 - \varepsilon$ for any $\varepsilon > 0$. We improve this latter inapproximability result in Section 5 by showing that there is no PTAS, nor an APTAS for the structured bin packing problem with non-unit weights on the leaves (while the classical bin packing problem admits an APTAS [3, Chapter 9]). This answers a question left open by Codenotti et al.
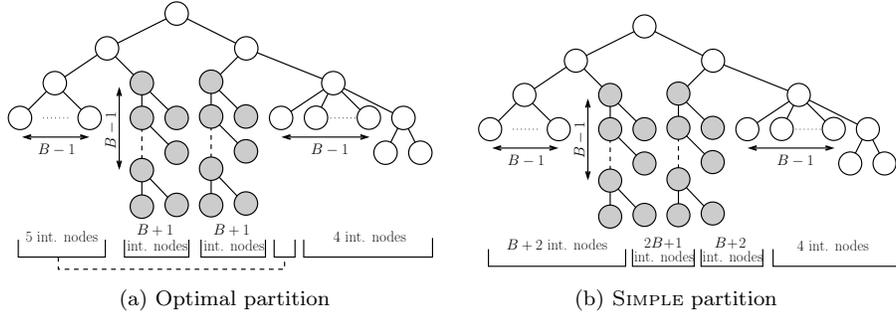
Figure 1: Optimal partition and the SIMPLE partition. The number of internal nodes (int. nodes) per bin is shown to facilitate the computation of the total node dispersal number.

4. Unitary weights on leaves

4.1. Refinement of the analysis of SIMPLE

Codenotti et al. claim that the algorithm SIMPLE obtains an approximation ratio of 3/2 for trees where each internal node has at least two children [1, p.220]. This is unfortunately not exact. We depict an example from a family of trees, parameterized by B, in Fig. 1. The optimal partition is shown in Fig. 1a. There are five buckets that are shown under the tree, and each bucket contains the leaves just above it. The first and the fourth buckets, linked with a dashed line, constitute the first bin, and the other buckets each constitute a single bin. For this example, the total node dispersal number can be computed easily as follows. For each bin, we count the number of internal nodes that have at least one descendant in that bin, and then we sum up all these numbers. The number of such internal nodes per bin is shown in Fig. 1. For example, there are five internal nodes that have at least one descendant in the first bin (these are all the white internal nodes except the rightmost one); there are $B + 1$ internal nodes that have at least one descendant in the second bin. With this computation, we see that the partition in Fig. 1a has a cost of $\rho(\mathcal{P}^\star) = 2B + 11$. SIMPLE fails to put all the leaves of the two deep subtrees (in gray on the figure) in a single bucket each, and obtains the partition given in Fig. 1b, where each bin now corresponds to a single bucket containing the leaves above it. The total dispersal number is $4B + 9$. As B goes to infinity, the ratio becomes arbitrarily

4

close to 2.

However, the approximation ratio of $3/2$ can be preserved for a restricted set of trees. The following lemma helps us to correct the existing approximation ratio analysis [1], with a more constrained conditions on the tree.

Lemma 1. *Let T be a tree such that no internal node has only one internal node among its children. Let h be its height and $\mu = |\mathcal{N}(T)|$ be the number of internal nodes. Then, $\mu \geq 2h - 1$.*

PROOF — By induction on h. Let r be the root of T. If $h = 1$, then all the children of r are leaves, so r is the only internal node and $\mu = 1 = 2h - 1$. Else, let c be a child of r such that $T(c)$, the corresponding subtree, is of height $h-1$. By the inductive hypothesis, the number μ' of internal nodes of $T(c)$ is greater than or equal to $2(h-1)-1$. In addition, r has another child which is not a leaf by the condition in the lemma. Therefore, $\mu \geq 2 + \mu' \geq 2 + 2h - 2 - 1 \geq 2h - 1$.
□

With this new condition, each internal node has either at least two children which are internal or any number of children which are leaves. This lemma lets us prove the following approximation result for the algorithm SIMPLE.

Theorem 1. *Let T be a tree such that no internal node has only one internal node among its children. Then, $\rho(\mathcal{P}_S) \leq 3/2\ \rho(\mathcal{P}^\star)$, where $\mathcal{P}_S$ is the partition given by the algorithm SIMPLE.*

PROOF — We follow the main steps of Codenotti et al. [1, Theorem 5], and highlight where we need Lemma 1 to prove the result. First, the internal nodes are split into two sets: N_1 is the set of internal nodes whose dispersal number is 1 in $\mathcal{P}^\star$; and N_2 is the set of internal nodes whose dispersal number is at least 2 in $\mathcal{P}^\star$. Codenotti et al. first observe that $\rho(N_2, \mathcal{P}^\star) \geq 2|N_2|$ and that $\rho(N_2, \mathcal{P}_S) \leq \rho(N_2, \mathcal{P}^\star) + |N_2|$, and thus, $\frac{\rho(N_2, \mathcal{P}_S)}{\rho(N_2, \mathcal{P}^\star)} \leq 3/2$. Next, the nodes in N_1 are considered. N_1 can be seen as a set of disjoint subtrees: if an internal node is in N_1, then all of its descendants are in N_1 too. Two cases for subtrees in
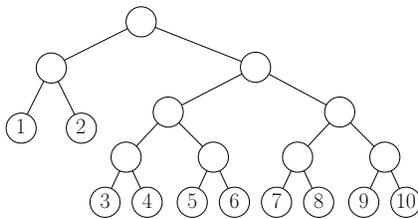
Figure 2: A tree for which no post-order based partition is optimal with $B = 5$.

N_1 are considered. If the number of internal nodes μ' of a subtree T' is at least twice its height h', then $\rho(T', \mathcal{P}_S) \leq \mu' + h'$ [1, p.220]. Since $\rho(T', \mathcal{P}^\star) = \mu'$, one has $\frac{\rho(T', \mathcal{P}_S)}{\rho(T', \mathcal{P}^\star)} \leq 3/2$. Otherwise if $\mu' < 2h'$, then one has $\mu' = 2h' - 1$ thanks to Lemma 1 (without the condition of the lemma this does not hold). In this case again, one obtains $\rho(T', \mathcal{P}^\star) = 2h' - 1$ and $\rho(T', \mathcal{P}_S) \leq 3h' - 1$, guaranteeing $3/2$ approximation for each node in N_1. □

4.2. Partitions based on post-order traversals

As noted before, the two existing NP-completeness proofs [1, 2] use an optimal partition that is based on a post-order. A natural question arises from this observation: does there always exist a post-order based partition with optimal cost? We prove that this is not the case.

Theorem 2. *There exists trees for which no post-order based partition has the optimal cost for the HSBP problem with unit weighted leaves.*

PROOF — Consider the tree depicted in Figure 2, and $B = 5$. Any post-order based partition will put leaves 1 and 2 in the same bin (yielding, for example, the first bin containing the leaves 1-to-5, and the second bin containing the leaves 6-to-10). In such a setting, the cost will be 13 (for the given example, the number of internal nodes that have at least one descendant in the first bin and the second bin are 6 and 7, respectively). However, it is possible to obtain a partition with cost 12 if the leaves 1 and 2 are put in different bins (the first bin contains the leaves 1,3,4,5,6 and the second bin contains the leaves 2,7,8,9,10). □

5. Arbitrary weights on leaves

We now return to the general case of the HSBP problem in which the leaves have arbitrary positive weights. Codenotti et al. [1] prove that there is no $3/2 - \varepsilon$ approximation for any $\varepsilon > 0$ and ask whether this can be improved asymptotically. We answer this question in Theorem 3 below by showing that there is no asymptotic approximation algorithm with a ratio strictly better than $3/2$. This shows that even for arbitrarily large instances, the $3/2$ ratio cannot be improved. Since we answer a question asked by Codenotti et al., we work under their assumption (in which every internal node has at least two children).

The proof of Theorem 3 uses a reduction from the NP-complete PARTITION problem [4, p.47]. An instance of PARTITION is defined as follows: given a set S of positive integers $a_1, \ldots, a_n$, determine if there exists a partition of S into two sets S_1 and $S_2 = S \setminus S_1$ such that $\sum_{s \in S_1} s = \sum_{s \in S_2} s$.

Theorem 3. *Let H be a polynomial time heuristic for the hierarchically structured bin packing problem and $\mathcal{P}_H$ be the partition given by H. Consider two constants α and β such that for every instance of the HSBP problem $\rho(\mathcal{P}_H) \leq \alpha \rho(\mathcal{P}^\star) + \beta$. Then, $\alpha \geq 3/2$, unless $P = NP$.*

PROOF — The proof is based on an inapproximability proof for a scheduling problem [5]. For the sake of contradiction, suppose that there exists a polynomial time heuristic such that $\rho(\mathcal{P}_H) \leq \alpha \rho(\mathcal{P}^\star) + \beta$ for any problem instance. Assume that $\alpha = 3/2 - \varepsilon$, where $\varepsilon < 1/2$, and let q be an integer such that $q > \frac{\beta}{2\varepsilon}$.

We show that with these assumptions, one can solve the PARTITION problem in polynomial time, which is a contradiction unless P = NP. For an instance I of PARTITION, we create the following instance J of HSBP, shown in Fig. 3. The instance J consists of a chain of q nodes, where each node has two leaves of weight 1, and the last node has n additional children (which are leaves) with weights $b_1, \ldots, b_n$, where $b_i = (q + 1) \times a_i$ for $i = 1, \ldots, n$. Let $W = \sum_{i=1}^{n} a_i$ be the sum of the weights of the integers in I, and the size B of the bins be

equal to $(q + 1) \times W/2 + q$. It is easy to see that a lower bound on the optimal solution for J is $2q$ (the leaves b_1 to b_n cannot be put into a single bin), and that the weight of any subset of b_is is a multiple of $q + 1$.

If the answer to a given instance I of PARTITION is YES, then we can split the leaves with weights $b_1, \ldots, b_n$ in two bins of weight $(q + 1) \times W/2$. We can then add in each bin q leaves of weight 1 and obtain a solution to instance J of HSBP with two bins. Then, we have $\rho(\mathcal{P}^\star) = 2q$ because each internal node has a dispersal number of two. If the answer to the instance I of PARTITION is NO, then for any two way partition of S into S_1 and $S_2 = S \setminus S_1$, one of the partitions has a weight strictly larger than $W/2$. The corresponding leaves in J will therefore weigh more than $(q + 1) \times (W/2 + 1)$, which is larger than B. Therefore, at least three bins are necessary to partition $b_1, \ldots, b_n$ in J. Since each bin increases the dispersal number of each internal node by one, we have $\rho(\mathcal{P}^\star) \geq 3q$.

We now describe how to obtain a YES/NO answer to the instance I of the PARTITION problem by applying the heuristic H to the instance J of the HSBP problem. If $\rho(\mathcal{P}_H) < 3q$, then we know that it is possible to partition $b_1, \ldots b_n$ into two bins. Each bin should then contain a weight of $(q + 1) \times W/2$ from these items (no more, because it should be less than B, no less because the other bin should then have a larger weight than B). In the corresponding partition of S into two, each part should have a weight of $W/2$, and hence the answer to PARTITION is YES. If $\rho(\mathcal{P}_H) \geq 3q$, then $\rho(\mathcal{P}^\star) \geq \frac{\rho(\mathcal{P}_H) - \beta}{\alpha} > \frac{3q - 2q\varepsilon}{3/2 - \varepsilon} = 2q$. Therefore, it is not possible to partition S into two sets of equal weights. Hence, in this case the answer to PARTITION is NO. Thus, we are able to solve the PARTITION problem in polynomial time, which contradicts its being NP-complete. Hence, $\alpha \geq 3/2$, unless P=NP. □

The previous theorem proves that there is no APTAS, and thus no PTAS, for the hierarchically structured bin packing problem.
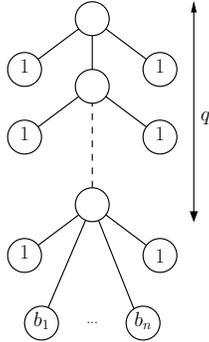
Figure 3: An instance J of HSBP corresponding to an instance I of PARTITION where $b_i = (q+1) \times a_i$ for $a_i \in S$ of I.

6. Conclusion

We have investigated two variants of the hierarchically structured bin packing problem in which the leaves have unit or arbitrary weights. We have shown that a known algorithm guarantees an approximation ratio of $3/2$ for the unit weighted case only when no internal node has only one internal node among its children. This condition was not specified before and fixes a glitch in an existing proof [1]. We note that this condition is likely to be satisfied by a large class of trees (e.g., trees that are not very unbalanced). We have also shown that there is no PTAS nor APTAS for the variant with arbitrary weighted leaves, and that the best post-order based partition (which is NP-complete to find) cannot always guarantee optimality.

The problem studied in the present paper has a generalization [2]. This generalization involves two trees T_1 and T_2, whose leaves are mapped by a bijection a: for any leaf ℓ of T_1, $a(\ell)$ is the corresponding leaf of T_2. The problem is to partition the set of pairs $\{(\ell, a(\ell))$ for ℓ a leaf of $T_1\}$ into bins with at most B pairs per bin in such a way that the total dispersal number of the induced partitions on the two trees is minimized. Currently, there is a practical heuristic for this problem [2], but not much is known about approximability.

References

[1] B. Codenotti, G. De Marco, L. Leoncini, M. Montangero, M. Santini, Approximation algorithms for a hierarchically structured bin packing problem, Information Processing Letters 89 (5) (2004) 215–221.

[2] P. R. Amestoy, I. S. Duff, J.-Y. L'Excellent, Y. Robert, F.-H. Rouet, B. Uçar, On computing inverse entries of a sparse matrix in an out-of-core environment, SIAM Journal on Scientific Computing 34 (4) (2012) 1975–1999.

[3] V. V. Vazirani, Approximation Algorithms, Springer-Verlag New York, Inc., New York, NY, USA, 2001.

[4] M. R. Garey, D. S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, W. H. Freeman & Co., New York, NY, USA, 1979.

[5] M. Queyranne, Bounds for assembly line balancing heuristics, Operations Research 33 (6) (1985) 1353–1359.