Approximation Schemes for Deal Splitting and Covering Integer Programs with Multiplicity Constraints

(Extended Abstract)

Hadas Shachnai* Oded Shmueli† Robert Sayegh‡

Abstract

We consider the problem of splitting an order for R goods, $R \geq 1$, among a set of sellers, each having *bounded* amounts of the goods, so as to minimize the total cost of the deal. In *deal splitting with packages (DSP)*, the sellers offer packages containing combinations of the goods; in *deal splitting with price tables (DST)*, the buyer can generate such combinations using price tables. Our problems, which often occur in online reverse auctions, generalize *covering integer programs with multiplicity constraints (CIP)*, where we must fill up an R-dimensional bin by selecting (with bounded number of repetitions) from a set of R-dimensional items, such that the overall cost is minimized. Thus, both DSP and DST are NP-hard, already for a single good, and hard to approximate for arbitrary number of goods.

In this paper we focus on finding efficient approximations, and exact solutions, for DSP and DST instances where the number of goods is some fixed constant. In particular, we show that when R is fixed both DSP and DST can be optimally solved in pseudo-polynomial time, and develop *polynomial time approximation schemes (PTAS)* for several subclasses of instances of practical interest. Our results include a PTAS for CIP in fixed dimension, and a more efficient (combinatorial) scheme for CIP_∞, where the multiplicity constraints are omitted. Our approximation scheme for CIP_∞ is based on a non-trivial application of the fast scheme for the *fractional covering problem*, proposed recently by Fleischer [Fl-04].

1 Introduction

An increasing number of companies are using online *reverse auctions* for their sourcing activities. In reverse auctions, multiple sellers bid for a contract from a buyer for selling goods and/or services. We consider the *deal splitting* problems arising in these reverse auctions. Suppose that a buyer needs to order multiple units from a set of R goods. The number of units required from the j-th good, $1 \leq j \leq R$, is $n_j \geq 1$. The goods can be obtained from m sellers, $S_1, \ldots, S_m$. Each seller offers certain amount from each good (or some combination of the goods); the maximum

*Computer Science Dept., Technion, Haifa 32000, Israel. E-mail: hadas@cs.technion.ac.il. Part of this work was done while the author was on leave in Bell Laboratories, Lucent Technologies, 600 Mountain Ave., Murray Hill, NJ 07974.

†Computer Science Dept., Technion, Haifa 32000, Israel. E-mail: oshmu@cs.technion.ac.il.

‡22 Anilevitch St. Haifa 35025, Israel. E-mail: abufayez@hotmail.com.

number of units of the j-th good available from S_i is T_{ij}, $1 \leq j \leq R$, $1 \leq i \leq m$. In any deal, we may split the order for the goods among a subset of the sellers. We say that a deal is *feasible* if (i) the number of units obtained from the j-th good is at least n_j, $1 \leq j \leq R$, and (ii) the amount of the j-th good obtained from S_i does not exceed T_{ij}, its supply from that good, $1 \leq i \leq m$, $1 \leq j \leq R$. The goal is to find a feasible deal of minimum total cost. Deal splitting naturally models a procurement auction to obtain raw materials with flexible sized lots and many other services. We consider two variants of the problem.

In *deal splitting with packages (DSP)*, each of the sellers, S_i, offers a set of N_i packages. The ℓ-th package, p^i_ℓ, $1 \leq \ell \leq N_i$, has a non-negative cost $c(p^i_\ell)$ and is given by the R-tuple $(n^i_{\ell 1}, \ldots, n^i_{\ell R})$; that is, S_i offers in this package $0 \leq n^i_{\ell j} \leq n_j$ units from the j-th good, $1 \leq j \leq R$. We need to find a feasible deal that minimizes the total cost.

In *deal splitting with price tables (DST)*, each seller S_i, has m_i price ranges. The minimal and maximal numbers of units of the j-th good available from S_i in the ℓ-th price range are $r_{\ell j}$ and $u_{\ell j}$, respectively. The unit cost for the j-th good in the ℓ-th range is $c_{\ell j}$, $1 \leq \ell \leq m_i$, $1 \leq j \leq R$.[1] Thus, the ℓ-th entry in the price table of S_i is given by the vector $\{(r_{\ell 1}, u_{\ell 1}, c_{\ell 1}), \ldots, (r_{\ell R}, u_{\ell R}, c_{\ell R})\}$. We need to find a feasible deal in which the sale of S_i, $1 \leq i \leq m$, corresponds to a valid entry in its price table, and the total cost is minimized.

We note that DSP is NP-hard already for $R = 1$, by reduction from Partition, and hard to approximate within factor $\ln R$ for arbitrary $R > 1$, as it includes as a special case the *multi-set multi-cover* problem.[2] For DST, we note that each price range of a seller "encodes" a possibly large number of packages (each formed by choosing the number of units from each good), as well as a simple rule for computing the price of a particular package (via the unit costs). Thus, in the special case where each price table consists of a *single* price range, which allows to form a *single* combination of the goods, we get an instance of the *constrained multi-set multi-cover*. It follows that DST is also hard to approximate within factor $\ln R$.

Note that DSP generalizes also *covering integer program with multiplicity constraints (CIP)*. In this core problem, we must fill up an R-dimensional bin by selecting (with bounded number of repetitions) from a set of R-dimensional items, such that the overall cost is minimized. Formally, let $A = \{a_{ij}\}$ denote the sizes of the items in the R dimensions, $1 \leq j \leq R$, $1 \leq i \leq n$; the cost of item i is $c_i \geq 0$. Let x_i denote the number of copies selected from item i, $1 \leq i \leq n$. We seek an n-vector $\mathbf{x}$ of non-negative integers, which minimizes $c^T\mathbf{x}$, subject to the R constraints given by $A\mathbf{x} \geq b$, where $b_j \geq 0$ is the size of the bin in dimension j. In addition, we have multiplicity constraints for the vector $\mathbf{x}$, given by $\mathbf{x} \leq d$, where $d \in \{1, 2, \ldots\}^n$. Recall that, in DSP, each seller S_i has T_{ij} units from the j-th good. Consider, for example, the case where $R = 2$, and suppose that S_i has $T_{i1} = 10$ units from the first good and $T_{i2} = 20$ units from the second good. S_i offers two possible packages: $p^i_1 = (5, 7)$ and $p^i_2 = (6, 2)$; then if we obtain two copies of p^i_1, no copies of p^i_2 are available. This dependence among the packages makes DSP a generalization of CIP.[3] Indeed, an instance of CIP can be formulated as a special case of DSP, where each seller offers a *single* package, whose "multiplicity" reflects the precise amount that is available from each of the goods.

[1] See an example in the Appendix.

[2] We elaborate in the Appendix on the relation of our problems to *set cover* and its generalizations.

[3] In the corresponding integer program, we get dependencies among the variables that give the number of copies obtained from each package.

1.1 Our Results

Since our deal splitting problems are harder than set cover, the best approximation ratio that we can expect for arbitrary R is $O(\log R)$ (see, e.g., in [Va-01]); thus, we focus on finding efficient approximations, and exact solutions, for subclasses of instances in which R is a fixed constant. We summarize below our main results.

Deal Splitting with Packages: We show (in Section 2.1) that when R is fixed DSP can be solved in pseudo-polynomial time. In Section 2.2 we develop a PTAS for instances where the i-th seller offers a set of $N_i \geq 1$ packages, $p_1^i, \ldots, p_{N_i}^i$, and the buyer can obtain at most r_ℓ^i copies from p_ℓ^i, for some $r_\ell^i \geq 1$; the total amount of the j-th good available from S_i is $T_{ij} = \sum_{\ell=1}^{N_i} n_{\ell j}^i r_\ell^i$, $1 \leq j \leq R$, $1 \leq i \leq m$. Indeed, such instances can be formulated as CIP with $\sum_{i=1}^m N_i$ variables. Thus, we get a PTAS for CIP in fixed dimension. In Section 2.3 we consider DSP instances with unbounded supply. Such instances model deals in which the buyer's need is much smaller than the supply from each of the goods. For these instances we develop a faster (combinatorial) scheme. This gives a combinatorial approximation scheme for CIP_∞.

Deal Splitting with Price Tables: We show (in Section 3) that when R is fixed DST is solvable in pseudo-polynomial time. We then develop a PTAS for DST instances in which the price tables satisfy some natural properties such as *volume discount*, that is widely used in reverse auctions (see, e.g., in [KPS-03], [BK$^+$02]).[4]

Techniques: Our PTAS for unbounded DSP (in Section 2.3) is based on a non-trivial application of a *fully polynomial time approximation scheme (FPTAS)* for the *fractional covering problem*, proposed recently by Fleischer [Fl-04]. We use this combinatorial scheme to obtain an *approximate* fractional solution for a linear programming formulation of our problem, building on a technique of Chandra et al. [CHW-76]. We show that by rounding an approximate solution for the LP we increase the cost of the optimal (integral) solution for the DSP instance only by factor of ε. Thus, we get a fast combinatorial implementation for our LP-based scheme. The overall running time of the scheme is $O(N^{\lceil R/\varepsilon \rceil} \cdot \frac{1}{\varepsilon^2} \log C)$, where $N = \sum_{i=1}^m N_i$ is the total number of distinct packages offered by the sellers, and $C = \max_{1 \leq i \leq N} c_i$ is the maximal cost of any package. Since unbounded DSP is equivalent to CIP_∞, this yields a combinatorial approximation scheme for CIP_∞ in fixed dimension. With slight modification, we get the first combinatorial scheme for *multi-dimensional multiple choice knapsack*.

In our PTAS for DST (in Section 3), we combine the guessing technique of Chekuri and Khanna [CK-00] with a novel application of the technique of Frieze and Clarke [FC-84], to the minimum binary multiple choice knapsack problem in fixed dimension. Indeed, due to the constraints imposed on the solution for DST — the amounts chosen from the goods for each seller must correspond to a valid entry in its price table — we cannot apply the rounding technique of [FC-84] to the fractional solution obtained by our scheme; instead, we apply non-standard rounding, which relies heavily on the mathematical properties of the price tables.

1.2 Related Work

Procurement auctions: Our deal splitting problems belong to the class of *winner determination* problems in reverse auctions. Generally, in reverse auction we have a single buyer that needs

[4] We elaborate on these properties in Section 3.

to obtain multiple goods, and a set of sellers offers *bids* for selling the goods. Bidding may follow various mechanisms (a survey of common mechanisms is given in [W-96]). The DST problem with single good (i.e., $R = 1$) and price tables that satisfy the volume discount property[5] was studied in [KPS-03]. The paper shows that DST is NP-hard already in this case and presents an FPTAS for the problem. There has been some previous work on deal splitting with *multiple* goods, however, these papers present either experimental studies or software that implements a given mechanism (see, e.g., [BK$^+$02]). Heuristic methods and preliminary analytic results related to deal splitting are given in [SG$^+$02].

Multiple Choice Knapsack (MCK): As shown in Section 2.2, DSP can be reduced to the *minimum R-dimensional binary MCK (R-MMCK)* problem. The *maximum* variant of this problem was studied since the mid-1970's (see, e.g., [Lu-75], [IH$^+$78], [I-80]). For a single dimension, the best known result is a PTAS by Chandra et al. [CHW-76]. Most of the published work on the maximum multi-dimensional binary MCK presented heuristic solutions (see a survey in [AM$^+$97]). Recently, Shachnai and Tamir developed in [ST-03] a PTAS for the problem in fixed dimension. Our scheme in Section 2.2 includes a PTAS for the minimum R-dimensional binary MCK in fixed dimension.

In Section 2.3, we reduce unbounded DSP to the minimum (non-binary) *R-dimensional MCK*. Chandra et al. [CHW-76] gave a PTAS for the *maximum* version of this problem in fixed dimension; their scheme solves as a procedure a linear program. Our scheme yields the first combinatorial scheme for this problem.

Set Cover/Covering Integer Programs: As mentioned above, our problems include as a special case the multi-set multi-cover problem. Set cover and its generalizations have been extensively studied. (A comprehensive survey is given in [Va-01].) Feige showed that in general set cover is hard to approximate within factor $\ln |E|$, where E is the set of elements to be covered. This hardness result carries over to multi-set multi-cover. The best approximation ratio for set cover is $(1 + \ln |E|)$ [C-79]. For multi-set multi-cover, the best ratio is $O(\log \max_S |S|)$, where $|S|$ is the size of the multi-set S when counting elements with multiplicity [RV-98]. This yields an $O(\log n)$-approximation algorithm for general instances of DSP with unbounded supply, where $n = \sum_{j=1}^{R} n_j$.

Covering integer programs form a large subclass of integer programs encompassing such NP-hard problems as minimum knapsack and set cover. This implies the hardness of CIP in fixed dimension (i.e., where R is a fixed constant). For general instances, the hardness of approximation results for set cover carry over to CIP. Dobson [D-82] gave an algorithm that outputs a solution of cost $O(\max_{1 \le i \le n}\{\log(\sum_{j=1}^{m} A_{ij})\})$ times the integral optimum. It was unknown until recently whether an $O(\log R)$-approximation existed. Kolliopoulos and Young [KY-01] settled this question. Their $O(\log R)$-approximation yields the first constant approximation for CIP in fixed dimension. A comprehensive survey of other results is given in [K-03] (see also in [KY-01]). The best known bounds for the CIP_∞ problem (that include existential improvements on the $O(\log R)$ factor) are due to Srinivasan ([S-99] and [S-96]). In this paper, we give the first pseudo-polynomial time algorithms and approximation schemes for CIP and CIP_∞ in fixed dimension.

[5] See in Section 3.

2 Deal Splitting with Packages

2.1 Exact algorithms

In the Appendix we show that when R is fixed DSP is solvable in pseudo-polynomial time.

Theorem 2.1 *DSP can be solved optimally in $O(m \cdot \max_{1 \le i \le m} N_i \cdot \prod_{j=1}^{R} n_j^3)$ steps.*

This yields a pseudo-polynomial time algorithm for CIP in fixed dimension.

Corollary 2.2 *CIP in fixed dimension, R, and n variables can be solved optimally in $O(n \cdot \max_{i,j}(a_{ij}d_i)^{2R})$.*

Consider a restricted version of DSP, in which we require that the total number of packages used in the deal is bounded by some fixed constant, $k \geq 1$. It can be shown that the problem then becomes easy to solve.

Theorem 2.3 *Restricted DSP is solvable in polynomial time.*

2.2 DSP with Bounded Multiplicity

2.2.1 Approximation Scheme

Suppose that the packages offered by each of the sellers have bounded multiplicity. Specifically, there are r_ℓ^i copies available from p_ℓ^i, $1 \le \ell \le N_i$. In this case, if $p_\ell^i = (n_{\ell 1}^i, \ldots, n_{\ell R}^i)$, $1 \le \ell \le N_i$, then the number of units available from the j-th good is $T_{ij} = \sum_{\ell=1}^{N_i} n_{\ell j}^i r_\ell^i$, for $1 \le j \le R$, $1 \le i \le m$. We now develop a PTAS for these instances, assuming that R is fixed.

Reduction to the R-MMCK Problem: Assume that we know the optimal cost, C, for our instance, then we reduce our problem to the minimum R-dimensional binary multiple choice knapsack problem. Recall that for some $R \geq 1$, an instance of binary R-MMCK consists of a single R-dimensional knapsack, of size b_j in the j-th dimension, and m sets of items. Each item has an R-dimensional size and is associated with a cost. The goal is to pack a subset of items, by selecting at most one item from each set, such that the total size of the packed items in dimension j is at least b_j, $1 \le j \le R$, and the overall cost is minimized.

For given values of C and ε, we define an instance for R-MMCK, such that if there is an optimal solution for DSP with cost C, we can find a solution for the DSP instance, whose cost is at most $C(1+\varepsilon)$. Note that C can be 'guessed' in polynomial time within factor $(1+\varepsilon)$, using binary search over the range $(0, \sum_{i=1}^{m} \sum_{\ell=1}^{N_i} r_\ell^i c(p_\ell^i))$.

Formally, given the value of C, the parameter ε and a DSP instance with bounded multiplicity, we construct an R-MMCK instance in which the knapsack capacities in the R dimensions are $b_j = n_j$, $1 \le j \le R$. Also, we have $N = \sum_{i=1}^{m} N_i$ sets of items, denoted by A_ℓ^i, $1 \le i \le m$, $1 \le \ell \le N_i$. Let $\hat{K}_\ell^i$ be the value satisfying $r_\ell^i c(p_\ell^i) \in [\hat{K}_\ell^i \varepsilon C/N, (\hat{K}_\ell^i + 1)\varepsilon C/N)$, then the number of items in A_ℓ^i is $K_\ell^i = \min(\hat{K}_\ell^i, \lfloor N/\varepsilon \rfloor)$. The set A_ℓ^i represents a sale of the package p_ℓ^i which partially fulfills the order. In particular, the k-th item in A_ℓ^i, denoted (i, ℓ, k), represents a sale of at most r_ℓ^i copies of p_ℓ^i such that $c(i, \ell, k)$, the total cost incurred by these copies, is in $[k\varepsilon C/N, (k+1)\varepsilon C/N)$. This total cost is rounded down to the nearest integral multiple of $\varepsilon C/N$; thus, $c(i, \ell, k) = k\varepsilon C/N$. The size of the item (i, ℓ, k) in dimension j, $1 \le j \le R$, denoted by

$s_j(i,\ell,k)$, is the total number of units of the j-th good that we can obtain, such that the total (rounded down) cost is $c(i,\ell,k)$.

Approximating the Optimal Solution for R-MMCK: Given an instance of R-MMCK, we 'guess' the set S of items of maximal costs in the optimal solution, where $|S| = h = \min(m, \lfloor \frac{2R(1-\varepsilon)}{\varepsilon} \rfloor)$. We choose the value of h such that the resulting solution is guaranteed to be within $1+\varepsilon$ from the optimal, as computed below. Let $E(S)$ be the subset of items with costs that are larger than the minimal cost of any item in S, that is, $E(S) = \{(i,\ell,k) \notin S \mid c(i,\ell,k) > c_{min}(S)\}$, where $c_{min}(S) = \min_{(i,\ell,k)\in S} c(i,\ell,k)$. We select all the items $(i,\ell,k) \in S$, and eliminate from the instance all the items $(i,\ell,k) \in E(S)$ and the sets A^i_ℓ from which an item has been selected. In the next step we find an optimal *basic solution* for the following linear program, in which $x_{i,\ell,k}$ is an indicator variable for the selection of the item $(i,\ell,k) \notin S \cup E(S)$.

$$
\begin{aligned}
(LP(S)) \qquad & \text{minimize} \qquad && \sum_{i=1}^{m}\sum_{\ell}^{N_i}\sum_{k=1}^{K^i_\ell} x_{i,\ell,k}\cdot c(i,\ell,k) \\
& \textit{subject to:} && \sum_{k=1}^{K^i_\ell} x_{i,\ell,k} \le 1 \text{ for } i=1,\ldots,m, \quad \ell=1,\ldots,N_i \\
& && \sum_{i=1}^{m}\sum_{\ell=1}^{N_i}\sum_{k=1}^{K^i_\ell} s_j(i,\ell,k) \ge n_j \text{ for } j=1,\ldots,R \\
& && 0 \le x_{i,\ell,k} \le 1 \text{ for } (i,\ell,k) \notin S\cup E(S)
\end{aligned}
$$

Rounding the Fractional Solution: Given an optimal fractional solution for R-MMCK, we get an integral solution as follows. For any i, $1 \le i \le m$ and ℓ, $1 \le \ell \le N_i$ let $k_{max} = k_{max}(\ell, i)$ be the maximal value of $1 \le k \le K^i_\ell$ such that $x_{i,\ell,k} > 0$; then we set $x_{i,\ell,k_{max}} = 1$ and, for any other item in A^i_ℓ, $x_{i,\ell,k} = 0$. Finally, we return to the DSP instance and take the maximum number of copies of the package p^i_ℓ whose total (rounded down) cost is $c(i,\ell,k_{max})$.

2.2.2 Analysis of the Scheme

We use the next three lemmas to show that the scheme yields a $(1+\varepsilon)$-approximation to the optimum cost, and that the resulting integral solution is feasible. We give the proofs in the Appendix.

Lemma 2.4 *If there exists an optimal (integral) solution for DSP with cost C, then the integral solution obtained from the rounding for R-MMCK has the cost $\hat{z} \le (1+\varepsilon)C$.*

Lemma 2.5 *The scheme yields a feasible solution for the DSP instance.*

Lemma 2.6 *The cost of the integral solution for the DSP instance is at most $\hat{z} + \varepsilon C$.*

Combining the above lemmas we get:

Theorem 2.7 *There is a polynomial time approximation scheme for DSP instances with fixed number of goods and bounded multiplicity.*

Consider an instance of CIP in fixed dimension, R. We want to minimize $\sum_{i=1}^{n} c_i x_i$ subject to the constraints $\sum_{i=1}^{n} a_{ij} x_i \geq b_j$ for $j = 1, \ldots, R$, and $x_i \in \{0, 1, \ldots d_i\}$ for $i = 1 \ldots, n$. We can represent such a program as an instance of DSP with $m = n$ sellers, each offering a single package i of multiplicity d_i. The number of units required from the j-th good is $n_j = b_j$.

Corollary 2.8 *The above is a PTAS for CIP in fixed dimension.*

2.3 Unbounded DSP

Consider now the special case where the sellers have unbounded supply from each of the goods. As before, we formulate our problem as a linear program, however, instead of applying standard techniques to solve this program, we use a fast combinatorial approximation scheme of [Fl-04] to get a fractional solution that is within factor of $(1 + \varepsilon)$ from the optimal; then, we round the solution to obtain an integral solution that is close to the optimal.

2.3.1 Overview of the Scheme

Our scheme, called *multi-dimensional cover with parameter* ε (MDC_ε), proceeds in the following steps.

(i) For a given $\varepsilon \in (0, 1)$, let $\delta = \lceil R \cdot ((1/\varepsilon) - 1) \rceil$.

(ii) Let c_i denote the cost of package i. Recall that $N = \sum_{i=1}^{m} N_i$ is the total number of packages. We number the packages by $1, \ldots, N$, such that $c_1 \geq c_2 \geq \cdots \geq c_N$.

(iii) Denote by Ω the set of integer vectors $\mathbf{x} = (x_1, \ldots, x_N)$ satisfying $x_i \geq 0$ and $\sum_{i=1}^{N} x_i \leq \delta$. For any vector $\mathbf{x} \in \Omega$:

- Let $d \geq 1$ be the maximal integer i for which $x_i \neq 0$. Find a $(1 + \varepsilon)$-approximation to the optimal (fractional) solution of the following linear program.

$$
\begin{array}{lll}
(LP') & \text{minimize} & \displaystyle\sum_{i=d+1}^{N} c_i z_i \\
& \textit{subject to}: & \displaystyle\sum_{i=d+1}^{N} a_{ij} z_i \geq n_j - \sum_{i=1}^{N} a_{ij} x_i \text{ for } j = 1, \ldots, R \qquad (1) \\
& & z_i \geq 0, \text{ for } i = d + 1, \ldots, N
\end{array}
$$

  The constraints (1) reflect the fact that we need to obtain from each of the goods at least $n_j - \sum_{i=1}^{N} a_{ij} x_i$, units, once we obtained the vector $\mathbf{x}$.

- Let $\hat{z}_i$, $d + 1 \leq i \leq N$ be a $(1 + \varepsilon)$ -approximate solution for LP'. We take $\lceil \hat{z}_i \rceil$ as the integral solution. Denote by $C_{MDC}(\mathbf{x}) = \sum_{i=d+1}^{N} \lceil z_i \rceil$ the value obtained from the rounded solution, and let $c(\mathbf{x}) = \sum_{i=1}^{N} c_i x_i$.

(iv) Select the vector $\mathbf{x}$ for which $C_{MDC_\varepsilon}(\mathbf{x}) = \min_{\mathbf{x}} (c(\mathbf{x}) + C_{MDC}(\mathbf{x}))$.

2.3.2 Analysis

We now show that MDC_ε is a PTAS for DSP with unbounded supply. Let C_o be the optimal cost for DSP (in which we take an *integral* number of units from each package).

Theorem 2.9 *(i) If $C_o \neq 0, \infty$ then $C_{MDC_\varepsilon}/C_o < 1 + \varepsilon$. (ii) The running time of algorithm MDC_ε is $O(N^{\lceil R/\varepsilon \rceil} \cdot \frac{1}{\varepsilon^2} \log C)$, where $C = \max_{1 \leq i \leq N} c_i$ is the maximal cost of any package, and its space complexity is $O(N)$.*

We use in the proof the next lemma.

Lemma 2.10 *For any $\varepsilon > 0$, a $(1+\varepsilon)$-approximation to the optimal solution for LP' can be found in $O(1/\varepsilon^2 R \log(C \cdot R))$ steps.*

Proof: We note that for a system of inequalities as given in LP' there is a solution in which at most R variables get non-zero values. This follows from the fact that the number of non-trivial constraints is R. Hence, it suffices to solve LP' for the $\binom{N-d}{R}$ possible subsets of R variables, out of $(z_{d+1}, \ldots, z_N)$. This can be done in polynomial time since R is a fixed constant. Now, for each subset of R variables we have an instance of the *fractional covering problem*, for which we can find a $(1+\varepsilon)$-approximate solution using, e.g., the fast scheme of Fleischer [Fl-04]. This gives the statement of the lemma. ■

Proof of Theorem 2.9: For showing (i), assume that the optimal (integral) solution for the DSP instance is obtained by the vector $\mathbf{y} = (y_1, \ldots, y_N)$. If $\sum_{i=1}^N y_i \leq \delta$ then $C_{MDC_\varepsilon} = C_o$, since in this case $\mathbf{y}$ is a valid solution, and $\mathbf{y} \in \Omega$, therefore, in some iteration MDC_ε will examine $\mathbf{y}$.

Suppose that $\sum_{i=1}^N y_i > \delta$, then we define the vector $\mathbf{x} = (y_1, \ldots, y_{d-1}, x_d, 0, \ldots, 0)$, such that $y_1 + \cdots + y_{d-1} + x_d = \delta$. (Note that $x_d \neq 0$.) Let $\tilde{C}_o(\mathbf{x}) = \sum_{i=d+1}^N \hat{z}_i$ be the approximate fractional solution for LP'. We have that $\mathbf{x} \in \Omega$, therefore

$$C_{MDC}(\mathbf{x}) - \tilde{C}_o(\mathbf{x}) \leq R c_d, \tag{2}$$

Let $C_o(\mathbf{x})$ be the optimal fractional solution for LP' with the vector $\mathbf{x}$. Note that C_o, the optimal (integral) solution for DSP, satisfies $C_o > c(\mathbf{x}) + C_o(\mathbf{x})$, since $C_o(\mathbf{x})$ is a lower bound for the cost incurred by the integral values $y_{d+1}, \ldots, y_N$. In addition, $c(\mathbf{x}) + C_{MDC}(\mathbf{x}) \geq C_{MDC_\varepsilon}$. Hence, we get that

$$\begin{aligned}
\frac{C_o}{C_{MDC_\varepsilon}} &\geq \frac{c(\mathbf{x}) + C_o(\mathbf{x})}{c(\mathbf{x}) + C_{MDC}(\mathbf{x})} \\
&\geq \frac{c(\mathbf{x}) + \tilde{C}_o(\mathbf{x})(1-\varepsilon)}{c(\mathbf{x}) + C_{MDC}(\mathbf{x})} \\
&> (1-\varepsilon)\left(1 - \frac{C_{MDC}(\mathbf{x}) - \tilde{C}_o(\mathbf{x})}{c(\mathbf{x}) + C_{MDC}(\mathbf{x}) - \tilde{C}_o(\mathbf{x})}\right) \\
&\geq (1-\varepsilon)\left(1 - \frac{C_{MDC}(\mathbf{x}) - \tilde{C}_o(\mathbf{x})}{\delta c_d + C_{MDC}(\mathbf{x}) - \tilde{C}_o(\mathbf{x})}\right)
\end{aligned}$$

The second inequality follows from the fact that $C_o(\mathbf{x}) \geq \tilde{C}_o(\mathbf{x})(1-\varepsilon)$, and the last inequality follows from the fact that $c(\mathbf{x}) \geq \delta c_d$.

Let $f(w) = w/(a+w)$, then $f(w)$ is monotone increasing. Define $w = C_{MDC}(\mathbf{x}) - \tilde{C}_o(\mathbf{x})$, and $a = \delta c_d$; then, using (2), we get that $1 - w/(a+w) \geq 1 - Rc_d/(\delta c_d + Rc_d) \geq 1 - \varepsilon$. Thus, we get that $C_o/C_{MDC_\varepsilon} \geq (1-\varepsilon)^2$. By taking in the scheme $\tilde{\varepsilon} = \varepsilon/2$ we get the statement of the theorem.

Next, we show (ii). Note that $|\Omega| = O(N^\delta)$, since the number of possible choices of N non-negative integers, whose sum is at most δ is bounded by $\binom{N+\delta}{\delta}$. Now, given a vector $\mathbf{x} \in \Omega$, we can compute $C_{MDC}(\mathbf{x})$ in $O(N^R)$ steps, since at most R variables out of $z_{d+1}, \ldots, z_N$ can have non-zero values. Multiplying by the complexity of the FPTAS for fractional covering, as given in Lemma 2.10, we get the statement of the theorem. ∎

Recall that DSP with unbounded supply is equivalent to CIP_∞. Hence, we have

Corollary 2.11 *There is a PTAS for CIP_∞ with n variables and fixed dimension, R, whose running time is $O(n^{R/\varepsilon} \cdot \frac{1}{\varepsilon^2} \log C)$.*

3 Deal Splitting with Price Tables

In the following we show that when R is fixed, DST can be solved in pseudo-polynomial time.

Theorem 3.1 *The DST problem can be optimally solved in $O(\sum_i m_i \cdot \prod_{j=1}^R n_j^2)$.*

The proof is similar to the proof of Theorem 2.1. We omit the details.

3.1 A PTAS for DST

We now describe a PTAS for DS with price tables and fixed number of goods. Our scheme applies to any instance of DST satisfying the following properties. **(P1)** *Volume discount.* If we increase the quantity bought from each of the goods, the unit cost can only decrease; that is, let $(a_1^1, \ldots, a_R^1)$, $(a_1^2, \ldots, a_R^2)$ be two vectors representing feasible sales for S_i, for some $1 \leq i \leq m$. If $a_j^2 \geq a_j^1$ for all $1 \leq j \leq R$, then the unit costs corresponding to the two vectors satisfy $c_j^2 \leq c_j^1$ for all j. **(P2)** *Dominance.* If the vectors $(d_1^1, \ldots, d_R^1)$, $(d_1^2, \ldots, d_R^2)$ represent valid sales (vis a vis the price table) for S_i, then the vector $\max((d_1^1, \ldots, d_R^1), (d_1^2, \ldots, d_R^2))$ also represents a valid sale for S_i, where the maximum is taken coordinate-wise. Table 1 (in the Appendix) satisfies the volume discount and the dominance properties.

We note that the properties (P1) and (P2) are quite reasonable in commercial scenarios, reflecting the desire of each seller to increase its part in the deal, by selling more units from each of the goods. (P1) implies that as the quantities increase, the unit prices decrease; (P2) allows for more combinations of the goods for the buyer.[6] It can be shown (by reduction from Partition) that DST is NP-hard even for instances that satisfy properties (P1) and (P2), already for $R = 1$.

Assume that we know the optimal cost, C, for our instance. Then, for a given value of $\varepsilon > 0$, we define an instance of R-MMCK, whose optimal solution induces a solution for DST with cost at most $(1+\varepsilon)C$. We then find an optimal fractional solution for the R-MMCK instance. This gives an almost optimal fractional solution for the DST instance. Finally, we use non-standard rounding to obtain an integral solution whose cost is within factor $(1+\varepsilon)$ from the fractional

[6] It is easy to modify any price table to one that satisfies (P2). We elaborate on that in the full version of the paper.

solution. Note that C can be 'guessed' in polynomial time within factor $(1+\varepsilon)$, using binary search over the range $(0, mR \cdot \max_{i,j} \max_{1 \le \ell \le m_i} u_{\ell j} c_{\ell j})$, i.e., we allow to take the maximum number of units from the j-th good in the ℓth range, for $1 \le \ell \le m_i$ $1 \le i \le m$, $1 \le j \le R$.

Reduction to the R-MMCK Problem: Given the value of C, the parameter ε and a DST instance with m price tables, we construct an R-MMCK instance which consists of a single R-dimensional knapsack with capacities $b_j = n_j$, $1 \le j \le R$, and m sets of items; each set A_i has $m_i \cdot (m/\varepsilon)^R$ items, $1 \le i \le m$. Each of the items in A_i represents a sale of the i-th seller, which (partially) satisfies the order. Specifically, each item in A_i is an integer vector $(i, \ell, k_1, \ldots, k_R)$, where ℓ is the range in the i-th price table from which we choose the goods, and $0 \le k_j \le m/\varepsilon$ is the contribution of the j-th good, bought from the i-th seller, to the total cost. We take this contribution as an integral multiple of $\varepsilon C/m$; for each vector we find the maximal number of units of each good that can be bought with this vector. If for some integer $g \ge 1$, $k_j \varepsilon C/m < g c_{\ell j} \le (k_j+1)\varepsilon C/m$ then we buy g units from the good and round down the cost to $k_j \varepsilon C/m$. The cost of an item $(i, \ell, k_1 \ldots, k_R)$ in A_i is given by $c(i, \ell, k_1 \ldots, k_j) = \varepsilon C/m \sum_{j=1}^{R} k_j$. We denote by $s_j(i, \ell, k_1 \ldots, k_R)$ the maximum total number of units of the j-th good that can be bought from S_i at the cost $k_j \varepsilon C/m$, $1 \le j \le R$.

Approximating the Optimal Solution for R-MMCK: Given an instance of R-MMCK, we 'guess' the set S of items of maximal costs in the optimal solution, where $|S| = h = \min(m, \lfloor \frac{2R(1-\varepsilon)}{\varepsilon} \rfloor)$. Let $E(S)$ be the subset of items with costs that are larger than the minimal cost of any item in S, that is, $E(S) = \{(i, \ell, k_1, \ldots, k_R) \notin S \mid c(i, \ell, k_1, \ldots, k_R) > c_{min}(S)\}$, where $c_{min}(S) = \min_{(i,\ell,k_1,\ldots,k_R) \in S} c(i, \ell, k_1, \ldots, k_R)$.

We select all the items $(i, \ell, k_1, \ldots, k_R) \in S$ and eliminate from the instance all the items $(i, \ell, k_1, \ldots, k_R) \in E(S)$ and the sets A_i from which an item has been selected. In the next step we find an optimal *basic solution* for the following linear program, in which $x_{i,\ell,k_1,\ldots,k_R}$ is an indicator variable for the selection of an item $(i, \ell, k_1, \ldots, k_R) \notin S \cup E(S)$.

$$
\begin{array}{lll}
(LP(S)) & \text{minimize} & \displaystyle\sum_{i=1}^{m} \sum_{\ell=1}^{m_i} \sum_{k_1,\ldots,k_R} c(i, \ell, k_1, \ldots, k_R) x_{i,\ell,k_1,\ldots,k_R} \\
& \textit{subject to}: & \displaystyle\sum_{\ell=1}^{m_i} \sum_{k_1,\ldots,k_R} x_{i,\ell,k_1,\ldots,k_R} \le 1 \text{ for } i = 1, \ldots, m \\
& & \displaystyle\sum_{i=1}^{m} \sum_{\ell=1}^{m_i} \sum_{k_1,\ldots,k_R} s_j(i, \ell, k_1, \ldots, k_R) x_{i,\ell,k_1,\ldots,k_R} \ge n_j \text{ for } j = 1, \ldots, R \\
& & 0 \le x_{i,\ell,k_1,\ldots,k_R} \le 1 \text{ for } (i, \ell, k_1, \ldots, k_R) \notin S \cup E(S)
\end{array}
$$

Rounding the Fractional Solution: Given an optimal fractional solution for R-MMCK, we now return to the DST instance and get an integral solution as follows. Suppose that we have D fractional variables for some set A_i, $x_{i,\ell_1,k_{11},\ldots,k_{1R}}, \ldots, x_{i,\ell_D,k_{D1},\ldots,k_{DR}}$, then we buy from the i-th seller $max_{1 \le d \le D} s_j(i, \ell_d, k_{d1}, \ldots, k_{dR})$ units of the j-th good, $1 \le j \le R$.

3.2 Analysis

We now show that the above scheme yields a $(1+\varepsilon)$-approximation for the optimum cost for DST, and that the resulting (integral) solution is feasible. We prove three lemmas.

Lemma 3.2 *If there exists an optimal (fractional) solution with cost C for the R-MMCK instance, then there exists a (fractional) solution with cost at most $(1+\varepsilon)C$ for the DST instance.*

Proof: For any $\varepsilon' > 0$, in any fractional solution for R-MMCK with ε', the cost of each of the selected items $(i, \ell, k_1 \ldots, k_R)$ in the DST instance is at most $(c(i, \ell, k_1 \ldots, k_R) + R\varepsilon' C/m) x_{i,\ell,k_1\ldots,k_R}$. Since for all $1 \leq i \leq m$ $\sum_{\ell=1}^{m_i} \sum_{k_1,\ldots,k_R} x_{i,\ell,k_1,\ldots,k_R} \leq 1$, this yields an increase of at most $R\varepsilon' C/m$ for the seller S_i. By taking $\varepsilon' = \varepsilon/R$, we get that the overall increase in the cost is $R\varepsilon' C = \varepsilon C$. ■

Lemma 3.3 *The integral solution obtained from the fractional solution for LP(S) yields a ratio of at most $(1+\varepsilon)$ to the optimal cost for the DST instance.*

Proof: Let $\mathbf{x}^*$ be an optimal (integral) solution for the linear program LP(S), and let $S^* = \{(i, \ell, k_1, \ldots, k_R) | \ x^*_{i,\ell,k_1,\ldots,k_R} = 1\}$ be the corresponding subset of items. As in the proof of Lemma 2.4, we may assume that $|S^*| \geq h$, otherwise we are done. Let

$$S^* = \{(i_1, \ell_1, k_{11}, \ldots, k_{1R}), \ldots, (i_r, \ell_r, k_{r1}, \ldots, k_{rR})\},$$

such that $c(i_1, \ell_1, k_{11}, \ldots, k_{1R}) \geq \cdots \geq c(i_r, \ell_r, k_{r1}, \ldots, k_{rR})$, for some $r > h$, and let

$$S^*_h = \{(i_1, \ell_1, k_{11}, \ldots, k_{1R}), \ldots, (i_h, \ell_h, k_{h1}, \ldots, k_{hR})\}.$$

We define $\sigma = \sum_{t=1}^{h} c(i_t, \ell_t, k_{t1}, \ldots, k_{tR})$ to be the total cost of the items in S^*_h. Then, for any item $(i, \ell, k_1, \ldots, k_R) \notin (S^*_h \cup E(S^*_h))$, we have $c(i, \ell, k_1, \ldots, k_R) \leq \sigma/h$.

We use below the notation $s_j(d)$ when referring to $s_j(i, \ell_d, k_{d1}, \ldots, k_{dR})$. Let $c(max_{1 \leq d \leq D} s_j(d))$ be the total cost of buying the j-th good in the entry of the price table where we obtain $max_{1 \leq d \leq D} s_j(d)$ units form good j, $1 \leq j \leq R$. The heart of the proof is the following claim.

Claim 3.4 *For any $1 \leq i \leq m$, the cost of buying from the i-th seller satisfies*

$$\sum_{j=1}^{R} c(max_{1 \leq d \leq D} s_j(d)) \leq \sum_{d=1}^{D} c(i, \ell_d, k_{d1}, \ldots, k_{dR}).$$

Proof: Note that, by our rounding technique, the vector giving the amounts bought from S_i from each of the goods satisfies

$$(max_{1 \leq d \leq D} s_1(d), \ldots, max_{1 \leq d \leq D} s_R(d)) \geq (s_1(i, \ell_d, k_{d1}, \ldots, k_{dR}), \ldots, s_R(i, \ell_d, k_{d1}, \ldots, k_{dR})),$$

for all $1 \leq d \leq D$. By the volume discount property, the total cost of the rounded solution satisfies $c(max_{1 \leq d \leq D} s_1(d), \ldots, max_{1 \leq d \leq D} s_R(d)) \leq \sum_{d=1}^{D} c(i, \ell_d, k_{d1}, \ldots, k_{dR})$. ■

Let z^* denote the optimal (integral) solution for the R-MMCK instance. Denote by $\mathbf{x}^B(S^*_h)$ a basic solution for LP(S), and let $\mathbf{x}^I(S^*_h)$ be an integral solution obtained by setting $x_{i,\ell_d,k_{d1},\ldots,k_{dR}} = 1$ for all $1 \leq d \leq D$. From Claim 3.4, we can bound the total cost of the solution output by the scheme by comparing z^* to the cost of $\mathbf{x}^I(S^*_h)$. In particular,

$$\begin{aligned} z^* &\geq \sum_{i=1}^{m} \sum_{\ell=1}^{m_i} \sum_{k_1,\ldots,k_R} c(i, \ell, k_1, \ldots, k_R) x^B_{i,\ell,k_1,\ldots,k_R}(S^*_h) \\ &\geq \sum_{i=1}^{m} \sum_{\ell=1}^{m_i} \sum_{k_1,\ldots,k_R} c(i, \ell, k_1, \ldots, k_R) x^I_{i,\ell,k_1,\ldots,k_R}(S^*_h) - \delta \end{aligned}$$

where $\delta = \sum_{(i,\ell,k_1,\ldots,k_R)\in F} c(i,\ell,k_1,\ldots,k_R)$, and F is the set of items for which the basic variable was a fraction, that is, $F = \{(i,\ell,k_1,\ldots,k_R)|\ x^B_{i,\ell,k_1,\ldots,k_R}(S^*_h) < 1\}$

Assume that in the optimal (fractional) solution of $LP(S^*_h)$ there are L tight constraints, where $0 \le L \le m+R$, then in the basic solution $\mathbf{x}^B(S^*_h)$, at most L variables can be strictly positive. Thus, at least $L-2R$ variables get an integral value (i.e. '1'), and $|F| \le 2R$. By an argument similar to the argument in the proof of Lemma 2.4, we get the statement of the lemma.

Now, we note that, from Lemma 3.2, we have $(1+\varepsilon)^2$-approximation for DST, and since C is guessed within factor $1+\varepsilon$, we get a $(1+\varepsilon)^3$-approximation. By taking $\varepsilon' = \varepsilon/4$ we get the statement of the theorem. ∎

Lemma 3.5 *The integral solution obtained by the rounding is feasible for DST.*

We give the proof in the Appendix. Combining the above lemmas we get:

Theorem 3.6 *There is a polynomial time approximation scheme for any DST instance satisfying properties (P1) and (P2), with fixed number of goods.*

References

[AM+97] M.M. Akbar, E.G., Manning, G.C. Shoja and S. Khan, "Heuristic Solutions for the Multiple-Choice Multi-Dimension Knapsack Problem". *Int. Conference on Computational Science*, (2), 2001: 659–668.

[BK+02] M. Bichler, J. Kalagnanam, H.S. Lee, J. Lee, "Winner Determination Algorithms for Electronic Auctions: A Framework Design". *EC-Web* 2002: 37–46.

[CHW-76] A.K. Chandra, D.S. Hirschberg and C.K. Wong, "Approximate Algorithms for Some Generalized Knapsack Problems". *Theoretical Computer Science* 3, pp. 293–304, 1976.

[CK-00] C. Chekuri and S. Khanna, "A PTAS for the Multiple Knapsack Problem". In *Proc. of SODA*, pp. 213–222, 2000.

[C-79] V. Chvátal, V. "A Greedy Heuristic for the Set Covering Problem", *Math. Oper. Res.* 4, 1979, 233–235.

[D-82] G. Dobson, "Worst-case Analysis of Greedy for Integer Programming with Nonnegative Data". *Math. of Operations Research*, 7, pp. 515–531, 1982.

[FC-84] A. M. Frieze and M.R.B. Clarke, Approximation Algorithms for the m-dimensional 0-1 knapsack problem: worst-case and probabilistic analyses. In *European J. of Operational Research*, 15(1):100–109, 1984.

[Fe-96] U. Feige. "A threshold of ln n for approximating set cover". In *Proc. of 28th Symposium on Theory of Computing*, pp. 314–318, 1996.

[Fl-04] L. Fleischer, "A Fast Approximation Scheme for Fractional Covering Problems with Variable Upper Bounds". In *Proc. of the 15th ACM-SIAM Symposium on Discrete Algorithm*, 2004, pp. 994–1003.

[GJ-79] M.R. Garey and D.S. Johnson, Computers and intractability: *A Guide to the Theory of NP-Completeness.* W.H. Freeman, 1979.

[I-80] T. Ibaraki, "Approximate algorithms for the multiple-choice continuous knapsack problems", *J. of Operations Research Society of Japan*, 23, 28–62, 1980.

[IH+78] T. Ibaraki, T. Hasegawa, K. Teranaka and J. Iwase. " The Multiple Choice Knapsack Problem". *J. Oper. Res. Soc. Japan* 21, 59–94, 1978.

[K-03] S. G. Kolliopoulos, "Approximating covering integer programs with multiplicity constraints", *Discrete Applied Math.*, 129:2–3, 461 –473, 2003.

[KY-01] S. G. Kolliopoulos and N. E. Young, "Tight Approximation Results for General Covering Integer Programs". In *Proc. of FOCS*, 522–528, 2001.

[KPS-03] A. Kothari, D. Parkes and S. Suri, "Approximately-Strategyproof and Tractable Multi-Unit Auctions". *Proc of Fourth ACM Conf. on Electronic Commerce*, 2003.

[La-76] E. L. Lawler, *Combinatorial Optimization: Networks and Metroids.* Holt, Reinhart and Winston, 1976.

[Lu-75] G. S. Lueker, "Two NP-complete problems in nonnegative integer programming". Report # 178, Computer science Lab., Princeton Univ., 1975.

[RV-98] S. Rajagopalan and V. V. Vazirani, "Primal-Dual RNC Approximation Algorithms for Set Cover and Covering Integer Programs". *SIAM J. Comput.*, 28(2) pp. 525–540, 1998.

[ST-03] H. Shachnai and T. Tamir, "Approximation Schemes for Generalized 2-dimensional Vector Packing with Application to Data Placement". In *proc of Random-APPROX*, 2003.

[S-96] A. Srinivasan, "An Extension of the Lovász Local Lemma, and its Applications to Integer Programming". In *Proc. of SODA*, 6–15, 1996.

[S-99] A. Srinivasan, "Improved Approximation Guarantees for Packing and Covering Integer Programs". *SIAM J. Comput.*, 29(2): 648–670, 1999.

[SG+02] O. Shmueli, B. Golany, R. Sayegh, H. Shachnai, M. Perry, N. Gradovitch and B. Yehezkel, "Negotiation Platform". *International Patent Application* WO 02077759, 2001-2.

[Va-01] V. V. Vazirani, Approximation Algorithms. Springer-Verlag, 2001.

[W-96] E. Wolfstetter, "Auctions: An Introduction". *J. of Economic Surveys*, 10, 1996: 367–420.

A Deal Splitting with Price Tables - An Example

Suppose that $R = 3$ and the goods are printers, cartridges and paper boxes. Table 1 gives the possible combinations of goods for the seller S_1, specified by amounts and unit costs, in 3 price ranges (i.e., $m_1 = 3$)..

Price range	Printers	Cartridges	Paper
1	$(0, 2, 300)$	$(0, 5, 30)$	$(0, 9, 15)$
2	$(3, 5, 280)$	$(7, 9, 25)$	$(10, 100, 10)$
3	$(6, 20, 250)$	$(10, 50, 23)$	$(10, 100, 10)$

Table 1: A price table for multiple (3) goods

Thus, if we buy 2 printers or less, the unit cost is 300, whereas the unit cost for buying $3 \leq p \leq 5$ printers is 280. A valid sale for S_1 is the combination $(1, 0, 7)$, in which we obtain a printer and 7 paper boxes. The cost of this sale, which corresponds to the first price range, is 405.

B Relation of Deal Splitting to Set Cover

Recall that in *multi-set multi-cover* we have a set E of elements and a collection of multi-subsets of the elements; a multi-set S may contain multiple copies of some elements. Each element j, $1 \le j \le |E|$, has a coverage requirement, $r_j \ge 1$; each multi-set S is associated with some positive cost c_S. We need to find a collection of multi-sets that satisfy the coverage requirements, such that the total cost is minimized. Indeed, any instance of multi-set multi-cover can be formulated as the following special case of DSP. There are m sellers: S_i offers a single package p^i, such that the number of units of good j in any package p^i is at most r_j; T_{ij} is unbounded for all i, j. Recall that, in a general DSP instance, the supply of the j-th good is bounded for $1 \le i \le m$ and $1 \le j \le R$, and each seller S_i may offer a set of packages. We need to find a minimum cost deal for $R = |E|$ goods, by ordering at least $n_j = r_j$ units from the j-th good. As multi-set multi-cover is hard to approximate within factor $\ln |E|$ [Fe-96], we get that, for arbitrary R, DSP is hard to approximate within factor $\ln R$. When each seller offers a single package, in which we can buy T_{ij} units from the j-th good (i.e., the total supply from that good), we get the *constrained multi-set multi-cover* problem which generalizes *set cover* [Va-01]; thus, for arbitrary R, such instances of DSP are also hard to approximate within factor $\ln R$.

C Some Proofs

Proof of Theorem 2.1: Given an instance of DSP, we construct a directed graph $G = (V, E)$, in which the vertices are *configurations* and the edges give the transition costs between configurations. Each configuration $(i, x_1, \ldots, x_R, \ell, y_1, \ldots, y_R)$ represents a (partial) deal in which we obtain a total of x_j units from the j-th good, from the sellers $S_1, \ldots, S_i$, and additional y_j units of the j-th good, by obtaining copies of the first ℓ packages of S_{i+1}, where $0 \le x_j + y_j \le n_j$, $1 \le j \le R$.[7] We add to the graph the initial configuration $(0, 0, \ldots, 0)$.

The edges in G give the costs of the following transitions.

- From the configuration $(i, x_1, \ldots, x_R, \ell, y_1, \ldots, y_R)$ to $(i, x_1, \ldots, x_R, \ell{+}1, y_1{+}z_1, \ldots, y_R{+}z_R)$, i.e., we obtain copies of the $(\ell + 1)$-th package of S_{i+1}.
- From the configuration $(i{-}1, x_1, \ldots, x_R, N_i, y_1, \ldots, y_R)$ to $(i, x_1{+}y_1, \ldots, x_R{+}y_R, 1, z_1, \ldots, z_R)$; that is, once we obtain the desired number of packages from S_i, we start obtaining copies of the first package of S_{i+1}.

The *length* of each edge is the cost of the corresponding transition. We compute the optimal solution by finding the shortest path from the configuration $(0, 0, \ldots, 0)$ to the configuration $(m - 1, x_1, \ldots, x_R, N_R, y_1, \ldots, y_R)$, where $x_j + y_j = n_j$, $1 \le j \le R$. Since G is a directed acyclic graph, this can be done in $O(|V| + |E|)$ steps. This gives the statement of the theorem. ∎

Proof of Lemma 2.4: Let $\mathbf{x}^*$ be an optimal solution for the linear program LP(S), and let S^* be the corresponding subset of items, that is, $S^* = \{(i, \ell, k) | \, x^*_{i,\ell,k} = 1\}$. If $|S^*| < h$ then we are done

[7] We truncate the amounts of the goods obtained from the sellers, so that the total amount from the j-th good does not exceed n_j; the costs of the packages remain unchanged.

(the scheme outputs a $(1+\varepsilon)$-approximation to the optimal cost: this is due to the initial guess of C); otherwise, let $S^* = \{(i_1, \ell_1, k_1), \ldots, (i_g, \ell_g, k_g)\}$, such that $c(i_1, \ell_1, k_1) \geq \cdots \geq c(i_g, \ell_g, k_g)$, for some $g > h$. Let $S_h^* = \{(i_1, \ell_1, k_1), \ldots, (i_h, \ell_h, k_h)\}$, and $\sigma = \sum_{t=1}^{h} c(i_t, \ell_t, k_t)$. Then, for any item $(i, \ell, k) \notin (S_h^* \cup E(S_h^*))$, we have $c(i, \ell, k) \leq \sigma/h$. Let $z^*, \hat{z}$ denote the optimal (integral) solution and the solution output by the scheme for the R-MMCK instance, respectively. We denote by $\mathbf{x}^B(S_h^*), \mathbf{x}^I(S_h^*)$ the basic and integral solutions of LP(S) as computed by the scheme, for the initial guess S_h^*.

We note that, by our rounding technique, for any $1 \leq i \leq m$, $1 \leq \ell \leq N_i$, the cost of the item selected from A_ℓ^i is $c(i, \ell, k_{max})$. Hence, we get that

$$
\begin{aligned}
z^* \quad & \geq \sum_{i=1}^{m} \sum_{\ell=1}^{N_i} \sum_{k=1}^{K_\ell^i} c(i, \ell, k) x_{i,\ell,k}^B(S_h^*) \\
& \geq \sum_{i=1}^{m} \sum_{\ell=1}^{N_i} \sum_{k=1}^{K_\ell^i} c(i, \ell, k) x_{i,\ell,k}^I(S_h^*) - \delta
\end{aligned}
$$

where $\delta = \sum_{(i,\ell,k) \in F} c(i, \ell, k)$, and F is the set of items for which the basic variable was a fraction, that is, $F = \{(i, \ell, k) |\ x_{i,\ell,k}^B(S_h^*) < 1\}$. This follows from the fact that we round up to 1 only $x_{i,\ell,k_{max}}$, for all $1 \leq i \leq m$ and $1 \leq \ell \leq N_i$. Recall that in any *basic* solution for a linear program, the number of non-zero variables is bounded by the number of tight constraints in some optimal solution (since non-tight constraints can be omitted). Assume that in the optimal (fractional) solution of $LP(S_h^*)$ there are L tight constraints, where $0 \leq L \leq N + R$. Then in the basic solution $\mathbf{x}^B(S_h^*)$, at most L variables can be strictly positive. Thus, at least $L - 2R$ variables get an integral value (i.e. '1'), and $|F| \leq 2R$. Note that, for any $(i, \ell, k) \in F$, $c(i, \ell, k) \leq \sigma/h$, since $F \cap (S_h^* \cup E(S_h^*)) = \emptyset$. Hence, we get that $z^* \geq \hat{z} + \frac{2R\sigma}{h} \geq \hat{z} + \frac{2R\hat{z}}{h} \geq \frac{\hat{z}}{1-\varepsilon}$. ■

Proof of Lemma 2.5: The feasibility of the solution follows from the definition of K_ℓ^i and the fact that $k_{max}(i, \ell) \leq K_\ell^i$. Thus, the number of copies bought from p_ℓ^i is at most r_ℓ^i. ■

Proof of Lemma 2.6: We note that for any package p_ℓ^i, the cost incurred by the integral solution for the DSP instance is at most $c(i, \ell, k_{max}(i, \ell)) + \varepsilon C/N$. It follows that the total cost value for the R-MMCK instance is increased at most by εC. ■

Proof of Lemma 3.5: Given the set of items with fractional variables in A_i, recall that we buy from S_i

$\max_{1 \leq d \leq D} s_j(d)$ units from the j-th good, $1 \leq j \leq R$. By property (P2), there exists an entry in the price table of seller i, in which the number of units available from the jth good is $\max_{1 \leq d \leq D} s_j(d)$. This entry guarantees that the covering constraint is satisfied for all goods, that is, we buy at least n_j units of the jth good, and the stock constraints of the seller are not violated.

■