European Journal of Operational Research 246 (2015) 462–470

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Production, Manufacturing and Logistics

Decentral allocation planning in multi-stage customer hierarchies

Sebastian Vogel[a,*], Herbert Meyr[b]

[a] Fachgebiet Produktion & Supply Chain Management, Technische Universität Darmstadt, Hochschulstr. 1, D-64 289 Darmstadt, Germany
[b] Lehrstuhl für Supply Chain Management (580C), Universität Hohenheim, D-70 593 Stuttgart, Germany

ARTICLE INFO

Article history:
Received 6 October 2013
Accepted 4 May 2015
Available online 9 May 2015

Keywords:
Supply chain management
Demand fulfillment
Allocation planning
Customer hierarchies
Customer heterogeneity

ABSTRACT

This paper presents a novel allocation scheme to improve profits when splitting a scarce product among customer segments. These segments differ by demand and margin and they form a multi-level tree, e.g. according to a geography-based organizational structure. In practice, allocation has to follow an iterative process in which higher level quotas are disaggregated one level at a time, only based on local, aggregate information. We apply well-known econometric concepts such as the Lorenz curve and Theil's index of inequality to find a non-linear approximation of the profit function in the customer tree. Our resulting Approximate Profit Decentral Allocation (ADA) scheme ensures that a group of truthfully reporting decentral planners makes quasi-coordinated decisions in support of overall profit-maximization in the hierarchy. The new scheme outperforms existing simple rules by a large margin and comes close to the first-best theoretical solution under a central planner and central information.

1. Problem statement

Many manufacturing firms are faced with the problem of ensuring that their planned product supply matches customer demand. This demand fulfillment problem is often challenging since purchasing and production need to be planned with sufficient lead time. Customer demand, by contrast, is often volatile and changes over time, and not all customers are of equal importance or profitability. Hence, an up-front allocation of scarce supply to the most important customer segments is beneficial.

Consider a manufacturer who makes a single, quasi-continuous product to stock (MTS) and sells it to a large number of customers via a multi-level sales and distribution system across continents, countries, and local sales areas (*customer hierarchy*). The constituent nodes of this hierarchy form the set $\mathcal{N} \subseteq \mathbb{Z}_{\geq 0}$, with each node $k \in \mathcal{N}$ representing a particular customer segment. The uppermost level with the root node $k = 0$ (representing the whole world) will be referred to as level 0. The leaf nodes $l \in \mathcal{L} \subset \mathcal{N}$ of the hierarchy will be referred to as *base* customer segments or simply *customers*. Each customer $l \in \mathcal{L}$ has demand $d_l \in \mathbb{R}_{\geq 0}$ and entails a profit per unit $p_l \in \mathbb{R}_{\geq 0}$. Intermediate nodes $i \in \mathcal{I} := \mathcal{N} \setminus \mathcal{L}$ represent *aggregate* customer segments like countries or continents which are characterized by aggregate demand $d_i \in \mathbb{R}_{\geq 0}$ and aggregate profitability $p_i \in \mathbb{R}_{\geq 0}$. This overall hierarchy setting is depicted in Fig. 1 where $\mathcal{D}_i$ denotes the set of (immediate) successor nodes of a node $i \in \mathcal{I}$.

Besides d_l and p_l at each leaf node $l \in \mathcal{L}$ of the hierarchy also the total available supply $0 < S \leq \sum_{l \in \mathcal{L}} d_l$ is given. The focus in this paper is on the following problem: Determine allocations $x_l \in \mathbb{R}_{\geq 0}$ to the leaf nodes l to maximize total profits $\sum_{l \in \mathcal{L}} (p_l \cdot x_l)$ in the hierarchy, subject to the constraints $0 \leq x_l \leq d_l$ and $\sum_{l \in \mathcal{L}} x_l \leq S$. In addition, also specify the allocations to all intermediate nodes x_i for $i \in \mathcal{I}$.

By comparison, allocation would be easy if a *central planner* was available. For her, the hierarchy would be 'flat', i.e. she would have full information and access to all leaf node customer segments. Then, she would only consider a simple continuous linear knapsack problem comprising all leaf nodes. This problem could be solved with the greedy approach first described by Dantzig (1957), i.e. by serving the leaf nodes in decreasing order of p_l and by setting $x_l := d_l$ until running out of supply. The allocations to the intermediate nodes i would merely be a consequence which results from summing the allocations to all successor nodes in the sub-tree $\mathcal{D}_i$. We will consider this approach as a first-best benchmark to our original problem and refer to it as *Optimal Central Allocation* (OCA).

In practice, there are typically multiple *decentral planners* present at the intermediate nodes. They only have decentralized information regarding demand and profit per unit of their immediately following nodes. Then, the allocation has to occur in a top-down manner by first setting the allocations to the intermediate nodes at the first level below the root node and then working further downwards until reaching the leaf nodes. This approach corresponds to solving a set of

* Corresponding author. Tel.: +491753789082.
E-mail addresses: sebastian.vogel@gmx.net (S. Vogel), h.meyr@uni-hohenheim.de (H. Meyr).

http://dx.doi.org/10.1016/j.ejor.2015.05.009
0377-2217/

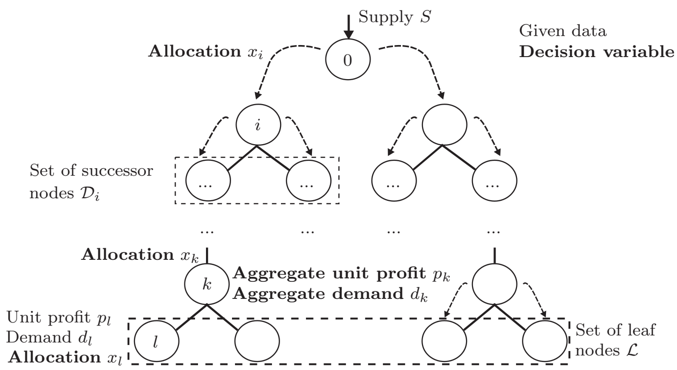

Fig. 1. Naming conventions in customer hierarchy.

continuous linear knapsack problems. Since each of these resulting decentral problems is of the same type as the single problem under OCA, the same greedy solution approach can be used per individual knapsack problem. We will refer to this decentral solution approach as *decentral average margin allocation* (DAMA) since subsequent nodes are served in decreasing order of average profits. However, many firms even fail to exploit differences in customer profitability in such a decentral setting. They perform a mere *quantity-based* allocation at each intermediate node, e.g. available supply is split proportionally to demand (proportional allocation; PA).

Surprisingly, this problem has not received much attention despite its preeminence in business. Most major manufacturing firms are faced with the challenge to serve a geographically dispersed customer base where margins differ and where some products are short at times. Apple Inc., for example, has repeatedly shifted its international roll-out of its tablet computer due to heavy demand and better margins in its domestic market (Paul & Madway, 2010). Roitsch and Meyr (2015) describe how a leading European refinery operator set up a hierarchical product allocation system based on DAMA to optimize margins by splitting scarce volumes between business units, sales districts and different types of contracted customers.

This paper will show that these existing decentral schemes DAMA and PA leave money on the table. We present the ADA scheme under which overall profit comes very close to the first best OCA benchmark with a central planner. To the knowledge of the authors, this allocation problem in multi-stage customer hierarchies was first characterized by Vogel (2013); our paper carves out the key insights developed there. We make the following contributions:

We extend the simple resource allocation problem for 'flat' customer segments to multi-level hierarchies and limited information transparency. To solve the resulting set of nested continuous linear knapsack problems, we propose a non-linear approximation of the total profit function in the customer hierarchy, which is otherwise piece-wise linear. The resulting approximation allows for a less data intensive, decentral representation of the allocation problem: rather than employing a central planner who requires comprehensive information from the leaves of the hierarchy, a very good allocation can be performed by a group of decentral planners who only need local data if they use our approximation function. Our new *Approximated Profit Decentral Allocation* (ADA) scheme helps to coordinate their decentral allocation decisions.

The remainder of this paper is organized as follows: First, similarities and differences to related literature areas will be pointed out (Section. 2). Then, key model assumptions will be stated (Section. 3) and disadvantages of the existing allocation schemes will be discussed (Section. 4). In Section. 5, ADA will be derived while in Section. 6 results of numerical experiments will be reported. Last, Section. 7 contains conclusions and further research needs.

2. Related literature

Our allocation problem in multi-stage hierarchies can be related to the following research areas:

Demand fulfillment and available-to-promise (ATP). Likewise, the Demand Fulfillment literature also addresses the problem which orders to serve if supplies are short. While most papers also allow customer classes to differ in terms of profitability, these classes are typically 'flat'. Hence, hierarchical settings with distributed decision making are not modeled. For MTS environments, it is often proposed to reserve or allocate quotas (so-called *available-to-promise* quantities, ATP) upfront based on expected demand per customer class to be able to serve more profitable orders from higher-order customer classesupon order arrival by gradually consuming these quotas ("push-based" ATP, see Ball, Chen & Zhao, 2004). Summaries of this stream of literature can be found in Fleischmann and Meyr (2003) or Pibernik (2005). Deterministic linear programs for this ATP allocation planning and consumption logic were introduced by Meyr (2009); a first stochastic variant for ATP allocation was described in Quante (2009).

Our paper adds predominantly to this literature area, not only by formalizing the ATP reservation problem in a customer hierarchy, but also by proposing a novel solution approach which is suited for distributed decision making.

Revenue management. There is a close relationship between ATP and some aspects of Revenue Management (RM), hence also to our problem setting. In general, the RM literature primarily discusses pricing, capacity reservation, and overbooking to resolve shortage situations in service industries (for an overview, see Talluri & van Ryzin, 2004). Harris and Pinder (1995) were the first to apply such tactics to manufacturing environments, but especially pricing decisions are often not available in practice for the allocation planning problem in MTS settings. Ball et al. (2004) and Quante (2009) highlight that ATP allocation problems, and thus also our problem, are structurally similar to the basic RM problem of allocating a given capacity to multiple booking classes which represent different customer segments. But with most conventional RM approaches being based on a central planner, they neither consider hierarchical settings nor distributed decision making with decentral information.

Capacity allocation. When implementing rule-based allocation, especially in decentral settings with human decision makers, it may turn out to be difficult to ensue that the allocation rule is being followed strictly if incentives are not aligned. This aspect may be investigated in a follow-up paper and a good starting point is the capacity allocation literature (for an introduction, see Lariviere, 2011).

In contrast to our assumptions in this paper, the actual allocation scheme is taken as given in capacity allocation. This literature rather focuses on the induced behavior of the planners and proposes ways to align incentives of decentral planners. As an example consider PA, the predominant allocation scheme used in industry, which is prone to manipulation: if shortage situations can be anticipated and if higher-level sales managers in the customer hierarchy cannot verify the information they receive from planners at lower levels, the latter have an incentive to bias their demand signals upwards in order to obtain a larger allocation. This *shortage gaming* behavior has been shown to be a major source of the bull-whip effect in supply chains (see Lee, Padmanabhan, & Whang, 1997).

Most capacity allocation papers only study quantity-based allocation without exploiting differences in customer profitability. Their setup is also different from ours by not addressing hierarchical settings.

Multi-stage customer hierarchies. In fact, only few aspects of multi-stage customer hierarchies are discussed in the literature. Formal hierarchy models based on a standard mathematical tree have a long tradition in organizational economics, e.g. Radner (1992) or van Zandt (2003). Some aspects of decentral coordination and allocation problems in hierarchies are covered in the budgeting literature (e.g.,