Int. J. Production Economics 148 (2014) 110–121

Contents lists available at ScienceDirect

Int. J. Production Economics

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

The dynamic fleet management problem with uncertain demand and customer chosen service level

Ning Shi [a], Haiqing Song [b,*], Warren B. Powell [c]

[a] Business School, Sun Yat-sen University, No. 135 XinGangXi Road, Guangzhou 510275, China
[b] Lingnan College, Sun Yat-sen University, No. 135 XinGangXi Road, Guangzhou 510275, China
[c] Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544, United States

ARTICLE INFO

Article history:
Received 19 January 2012
Accepted 2 August 2013
Available online 9 October 2013

Keywords:
Stochastic programming
Approximate dynamic programming
Dynamic fleet management
Decomposition method
Multistage networks

ABSTRACT

In this paper, we study a dynamic fleet management problem with uncertain demands and customer chosen service levels. We first show that the problem can be transformed into a dynamic network with partially dependent random arc capacities, and then develop a structural decomposition approach which decomposes the network recourse problem into a series of tree recourse problems (TRPs). As each TRP can be solved by an efficient algorithm, the decomposition approach can solve the problem very efficiently. We conduct numerical experiments to compare its performance with two alternative methods. Numerical experiments show that the performance of our method is quite encouraging.

1. Introduction

We consider the problem faced by a carrier who needs to manage a fleet of homogeneous vehicles over space and time to serve a number of transportation requests. Each transportation request is defined by an origin–destination pair. The service for the request must start at a specific instant in time or be lost. The carrier may provide customers a portfolio of service levels that vary in travel time (24 h, 48 h or 3 days). For example, when customers place orders from E-commercial companies (such as amazon.com, 360buy.com, Taobao.com), they can choose delivery service level ranged from 3 days, 2 days and 1 day with different rates. Carriers, given orders with various service level and demand, need to allocate their fleets to fulfill the tasks. If customer demands a service level with shorter travel time, the carrier may need to pay additional transportation cost and ask for a higher rate. Customers select one service level when they place the order. This problem belongs to a class of dynamic fleet management problems (DFMPs), which can be viewed as a type of spatial, dynamic inventory management problem with reusable vehicles. At any point in space and time, a vehicle may be assigned to satisfy a revenue generating activity. It may be repositioned empty to another point in space and time or held in inventory.

The DFMP has received a lot of attention from the research community. Comprehensive reviews can be found in Dejax and Crainic (1987), Crainic and Laporte (1998), Powell and Topaloglu (2005) and Flatberg et al. (2007). We focus on the most relevant literature. Early fleet management models are deterministic and appear as the first applications of linear programming and min-cost network flow algorithms (see Dantzig and Fulkerson, 1954; White, 1972). These models formulate the problem over a time-space network (also called a *dynamic network*, or a *time-staged network*). A number of authors have studied the problem of random demands. The first to explicitly incorporate uncertainty in demands is Jordan and Turnquist (1983) in the context of the allocation of rail freight cars. Powell (1986) formulates the problem with random demands as a dynamic network with random arc capacities. Powell (1988) provides an overview of alternative modeling and algorithmic strategies for the stochastic fleet management problem. This model has been applied to dynamic truck allocation (Powell, 1987; Powell et al., 1988), empty container repositioning (Crainic et al., 1993; Chu, 1995) and transportation planning (Barbarosoglu and Arda, 2004). Godfrey and Powell (2002) and Topaloglu and Powell (2006) introduce the idea of adaptively estimating piecewise-linear approximations of the value of vehicles in the future.

All of this literature assumes that the time for completing the transportation request is fixed. In our motivating application, the time is not fixed but determined by the service level chosen by the customer. Therefore, the service time for a particular demand can be regarded as a random variable, and it becomes known as soon as the demand is known. But for orders moving in the future, we do not yet know the customer's choice, and for this reason the travel time is stochastic. We note that this type of randomness in the time

* Corresponding author. Tel.: +86 20 84113598; fax: +86 20 84114823.
E-mail address: songhq@mail.sysu.edu.cn (H. Song).

0925-5273/$ - see front matter
http://dx.doi.org/10.1016/j.ijpe.2013.09.010

to complete a service contrasts with randomness resulting from delays due to weather, congestion and equipment failures. In these settings, the service completion time only becomes known after the demand has been served.

A few authors have considered the problem of random travel times. Cheung et al. (2005) proposes a labeling method to handle uncertain service times. This method is able to handle a number of operational constraints, but does not scale to larger problems. Glockner and Nemhauser (2000) use scenario trees in a stochastic programming model to handle uncertainty. Simao et al. (2009) use approximate dynamic programming to model truckload operations which includes random travel times which become known only after a trip is completed.

There is a separate literature that includes customer choice. For example, Zhang and Adelman (2009) incorporate customer choice in airline revenue management. However, customer choice has not been considered in the fleet management literature.

Our solution strategy extends a line of research in stochastic vehicle allocation using separable approximations. This problem class has been most widely studied using the framework of two-stage stochastic programs with network recourse (Wallace, 1986, 1987; Birge and Wallace, 1988). Wallace (1986) introduces a piecewise linear upper bound for networks and provides a result that is generalized in Birge and Wallace (1988) for stochastic programs. Independently, separable, piecewise linear approximations have been proposed for discrete vehicle allocation problems that arise in the context of fleet management (Powell, 1986). Frantzeskakis and Powell (1990) presents a method called the *Successive Linear Approximation Procedure* (SLAP) for approximating the expected recourse function. It is generalized by the *Successive Convex Approximation Method* (SCAM) in Cheung and Powell (1996). Similar to SLAP, SCAM decomposes the network into a series of trees and expresses the expected recourse function in terms of trees in the network. Furthermore, SCAM generalizes SLAP by using convex, instead of linear, approximations of the expected recourse function. The advantage of the SCAM is that it works quite well even with problems with huge sample space of random outcomes. SCAM and SLAP both estimate piecewise linear functions in a preprocessor. Godfrey and Powell (2002) and Topaloglu and Powell (2006) propose adaptive approximation procedures to estimate piecewise linear, separable approximations.

In this paper, we show that the problem of managing a fleet of vehicles where customers can choose the service level for orders can be formulated as a multistage dynamic network model with partially dependent random variables by an arc transformation. The contributions of this paper are:

1. We formulate for the first time the fleet management problem with customer-chosen service levels. We formulate it as a dynamic network model with partially dependent random arc capacities. As this model retains the network structure, it enables us to apply structural decomposition techniques. Unlike previous dynamic fleet management models where random variables are usually assumed independent, our model allow random variables representing the customer selection of service level be partially dependent. It enables us to consider customer behaviors, which usually introduce a dependency in random demands, in fleet management field.
2. We show that with slight modifications, SCAM works for DFMP with customer chosen service levels. We present a new structural decomposition approach: the *Successive REsource directive Decomposition Method* (SREDM). This approach provides a search mechanism which takes the advantage of the efficiency of solving the sub-problems. Both decomposition methods explicitly take the advantage of the network structure.
3. We evaluate the efficiency of the decomposition method numerically. The results demonstrate that our approach is superior to the modified SCAM, and the gaps between the results of our approach and the lower bound are very tight.

The rest of this paper is organized as follows. Section 2 introduces how to transform the problem to a dynamic network model with partially dependent random arc capacities. Then it provides a multistage stochastic programming formulation. Next, Section 3 presents the structural decomposition approach that formulates the problem as a *Discrete Resource Allocation Model* (DRAM) and decomposes the problem into a number of stochastic tree recourse problems. Section 4 compares this approach with the alternative methods on a set of test problems. The results in Section 4 show the superiority of the new method over the alternative methods. Finally, Section 5 gives some concluding remarks.

2. Problem formulation

In this section, we first introduce an arc transformation. By this transformation, any arc with a random travel time can be transformed into a set of arcs with deterministic travel times and dependent random arc capacities. Then, the problem is formulated as a dynamic network model with dependent arc capacities.

2.1. Arc transformation

The left part of Fig. 1 is a typical time-space network. The vertical dimension represents the geographical locations and the horizonal dimension represents the discrete times. In the example on the left of Fig. 1, there are two locations i and j, and there are three time periods. Let us define,

$\mathcal{L}$	the set of locations
Γ	the set of discrete time periods
u_{ijt}	the demand, in terms of the number of movements, from location i to location j starting at time t

We use i_t to denote a node representing a location i at time t. Accordingly, $(i_t, j_{t'})$ denotes the arc from i_t to $j_{t'}$.

Now we are ready to illustrate the transformation by Fig. 1. Assume that the arc $(i_t, j_{t'})$, where $t' = t + \tau_{ijt}$, represents the movement from location i to location j starting at time t, where

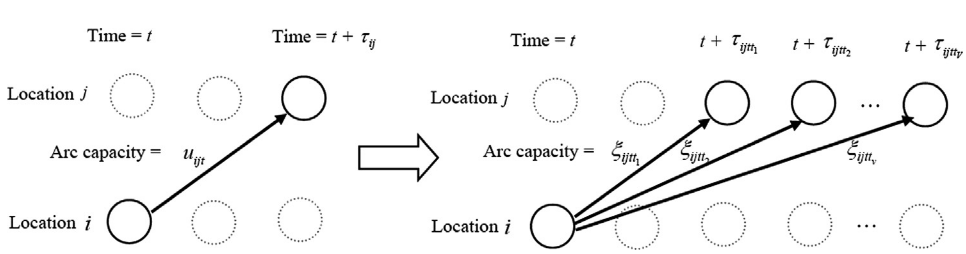

Fig. 1. Transformation of an arc with random travel time and random capacity to a set of arcs with deterministic travel times and random arc capacities.