

capacity, and a chance constraint, which states that for each forwarder the probability of the allotment exceeding the demand is within a specified threshold. The model assumes that the contribution for each route and for each forwarder is equal to the product of the amount of the allotment and the per-unit contribution. Putting it differently, Kasilingam (1996) assumes that the forwarder is charged for the whole allotment. Under this assumption, the total contribution is deterministic, and the objective function is linear in the allotment. However, we assume that the contribution is linear in the actual allotment usage, not the whole amount that the forwarder originally receives. Our assumption is consistent with the industry practice; see Bazaraa et al. (2001) and Hellermann (2006) Section 2.2.4.

The rest of the paper is organized as follows. In Section 1, we formulate the carrier's problem. The probabilistic analysis and the optimal solution of the carrier's problem are given in Section 2. We develop heuristic procedures in Section 3 and test their performance in Section 4. Finally, concluding remarks are provided in Section 5.

1 Problem Formulation

Throughout this article, we use \mathbb{N} to denote the set of natural numbers, \mathbb{Z}_+ the set of nonnegative integers, and \mathbb{R}_+ the set of nonnegative real numbers.

The carrier has κ units of cargo capacity per flight, which would be allocated as allotments for m forwarders before the season starts. Let $x_i \geq 0$ denote the per-flight allotment for forwarder $i \in \{1, 2, \dots, m\} := \mathcal{M}$. We assume that cargo capacity is one dimensional, because in practice most airlines manage their allotments only in terms of weight. (Nonetheless, modern short-haul planes are constrained by volume not weight due to recent advances in aircraft engine.) Alternatively, cargo capacity may be thought of as a number of standard containers in a cargo hold.

The carrier charges forwarder i based on its actual allotment usage $U_i(x_i)$ and receives

a contribution of p_i per unit usage. Assume that the carrier is risk neutral, since it operates multiple repeat flights in one season. The carrier wants to maximize the expected total contribution subject to a capacity constraint:

$$\max \left\{ E \left[\sum_{i=1}^m p_i U_i(x_i) \right] : \sum_{i=1}^m x_i \leq \kappa, \quad x_i \geq 0 \text{ for each } i \in \mathcal{M} \right\} \quad (1)$$

The first constraint states that the carrier does not allocate more than its cargo capacity. In other words, it does not overbook their allotments. Theoretically, we could include an overbooking policy in (1): the objective function would include an expected oversale cost, and the right-hand-side of the capacity constraint would include an overbooking pad. In practice, the penalty cost from delaying forwarders' shipments is much higher than that from bumping passengers. Unlike passenger airlines which typically overbook their seats, air-cargo carriers rarely overbook their allotments.

In practice, the forwarders' usages are usually dependent, because they are affected by same factors, e.g., a gross domestic product (GDP), a weather condition, and a fuel price, etc. Even if the forwarders' usages $U_1(x_1), \dots, U_m(x_m)$ are dependent, (1) remains valid. Recall that the expectation of a finite sum of random variables is the finite sum of their expected values, and that the expectation of a constant times a random variable is the constant times the expectation of the random variable. The objective function becomes

$$E \left[\sum_{i=1}^m p_i U_i(x_i) \right] = \sum_{i=1}^m p_i E[U_i(x_i)]$$

The expectation $E[U_i(x)]$ can be calculated with respect to the marginal distribution of random variable $U_i(x)$.

Assume that the request to book a shipment arrives to the carrier one by one throughout the booking horizon. When there is a medium-term contract between the forwarder and the carrier, the booking request from the forwarder is accepted, if its space requirement is within the contracted space (Bazaraa et al. 2001). In practice, the requests from

the forwarder with the allotment are accepted on the first-come, first-served basis, but those from the forwarder without the allotment need not be. The latter is sometimes referred to as *ad-hoc/free sale* (Slager and Kapteijns 2004). The carrier might apply some rudimentary RM techniques in controlling the free sale space; e.g., it might reject the current request, if it anticipates the larger contribution from the future requests. We do not model the free sale, because many Asia Pacific airlines reserve large portions of their capacity as allotments (Hendricks and Elliott 2005). An omnibus model, which includes both allotment and free sale, is an interesting future research direction.

Consider the booking requests from forwarder i , which receives allotment x . Let N_i be its total number of requests, and let $W_{i,k}$ be the space requirement of its k -th request, where $k \in \{1, 2, \dots, N_i\}$. Let $X_{i,k}(x)$ denote the cumulative usage after the carrier makes the k -th accept/reject decision [just before the arrival of the $(k+1)$ -st booking request] for each $k \in \{1, 2, \dots, N_i\}$. At the end of the booking horizon, the actual usage of forwarder i that receives allotment x is $U_i(x) = X_{i,N_i}(x)$. The finite sequence of the cumulative usages $\{X_{i,k}(x) : k = 1, 2, \dots, N_i\}$ is determined by the following recurrence equation:

$$X_{i,k}(x) = \begin{cases} X_{i,k-1}(x) + W_{i,k} & \text{if } W_{i,k} \leq x - X_{i,k-1}(x) \\ X_{i,k-1}(x) & \text{otherwise} \end{cases} \quad (2)$$

for each $k = 1, 2, \dots, N_i$, and $X_{i,0}(x) = 0$. Equation (2) can be explained as follows. Prior to the arrival of the k -th request, the cumulative usage of the accepted shipments is $X_{i,k-1}(x)$, and the unused portion of the allotment is $x - X_{i,k-1}(x)$. If the unused portion is at least the space requirement of the k -th booking request, then the request is accepted, and the cumulative usage increases to $X_{i,k-1}(x) + W_{i,k}$. Otherwise, the request is rejected, and the cumulative usage stays the same.

Given that its total number of booking requests is $N_i = n \in \mathbb{N}$, we assume that the