Introduction

Air-cargo operations generate significant revenue for passenger airlines, most of which carry cargo shipments in the belly of their planes. Air-cargo volumes are expected to grow 6 percent annually over the next two decades (Airbus 2007). The growth of air-cargo traffic partly results from global trade liberalization, and the emerging implementation of supply chain management strategies, which emphasize on short lead times.

Many carriers (airlines), especially those in Asia-Pacific, reserve large portions of cargo space on specific flights over a period of time for their key freight forwarders either as part of a binding contract or as part of a goodwill and gentleman's agreement (Billings et al. 2003). The allotted space is referred as an *allotment*. The forwarders re-sell their allotments to their own customers. They collect individual packages from shippers and transport the consolidated shipments to the airline. In addition, they provide value-adding services such as picking up from the shipper location and customs clearance for international cargo. Consequently, shippers typically prefer to use the forwarders unless a shipment is an emergency, or contains perishable or hazardous materials (Thuermer 2005).

An allotment contract is normally valid for the entire *season*, whose duration is specified by the International Air Transport Association (Slager and Kapteijns 2004). The airline awards the medium-term contracts to the forwarders before the season starts. At that time, the forwarders do not know the exact amount of their customer demand. After the forwarder's demand is materialized, the airline often allows the forwarder to return the unused portion of its allotment and charges only the actual usage. The airline's problem is to choose the allotments so that the expected total contribution is maximized. The *contribution* of the shipment is the contract rate minus the associated operational costs, e.g., the incremental fuel cost, and the handling costs.

In this article, we consider a single airline, which wants to allocate cargo capacity to multiple freight forwarders. During the booking horizon, the request from the forwarder arrives one by one. The forwarder's request is accepted, if its space requirement does not exceed the allotment the forwarder currently has. At the end of the booking horizon, the airline receives the contribution, proportional to the forwarder's actual allotment usage. We show that the sequence of the cumulative usages of the accepted shipments for each forwarder forms a discrete Markov chain. The expected actual allotment usage is neither concave nor differentiable. The airline's problem can be solved via dynamic programming, in which the number of computational steps quadratically increases in the capacity.

To develop some heuristic solutions, we assume that the shipment can be partially accepted. Under this assumption, the objective function is concave but still not differentiable. We present two heuristic approaches. In the first, the expected actual usage is approximated by a continuously differentiable function, and the heuristic solution is derived from the Karush-Kuhn-Tucker (KKT) conditions. The latter is based on Lagrangian relaxation, in which the capacity constraint is dualized. Through a series of numerical experiments, we show that the latter outperforms the first.

Since air-cargo capacity can be sold at different prices to heterogeneous customers but cannot be sold after the flight departure, it is a prime candidate for revenue management (RM) strategies. Books on RM theory and practice are, e.g., Ingold et al. (2000), Tallüri and van Ryzin (2004), Yeoman and McMahon-Beattie (2004), and Phillips (2005). Descriptive papers on air-cargo RM can be found in e.g., Kasilingam (1996), Bazaraa et al. (2001), Billings et al. (2003), Slager and Kapteijns (2004) and Becker and Dill (2007). In general, there is a vast literature on capacity management and on supply chain contracts; see Van Mieghem (2003), Cachon (2003), and Lariviere (1999) for reviews. Unfortunately, "revenue managing the contractual terms under which inventory is sold, remains almost untouched in the RM literature in spite of the vast majority of business that is transacted under negotiated contracts (Boyd and Bilegan 2003)." Despite a large number of papers on a nonlinear resource allocation problem (cf. Patriksson 2008), there are few papers that present mathematical models of the air-cargo allocation problem. Below, we review some of them.

Hellermann (2006) proposes a capacity-option pricing contract between a single forwarder and a single airline. The capacity is sold in two stages. In the first stage, the capacity is sold up front through the medium-term contract. In the second stage, the airline sells on the spot market, in which price and demand are random. The interaction between the two parties is specified by a Stackelberg game. Under the terms of the contract, the forwarder pays a reservation fee to acquire the right (but not the obligation) to use capacity, and later pays an execution fee if it eventually uses the capacity. The airline, the Stackelberg leader, first announces the reservation and execution fees. Then, the forwarder decides on the amount of capacity to reserve. After the demand and the spot price materialize, it determines how many reservations to call on. Hellermann compares the performance of the option contract to that of a fixed-commitment contract. Hellermann consider the single-forwarder case, whereas we consider the multiple-forwarder case.

Gupta (2008) studies flexible contracts between a single forwarder and a single airline, which receives two demand streams from the forwarder and the direct shippers. The paper identifies two contract schemes, which allow the carrier to achieve an efficient capacity allocation. In our article, the airline receives multiple demand streams from multiple forwarders. The functional form of the allotment usage in both Hellermann and Gupta resembles that in the newsvendor; specifically, the actual usage equals the minimum of the forwarder's demand and the allotment. We present a more detailed formulation; the sequence of the cumulative usages from one request to the next is modeled as a discrete Markov chain, and the expected actual usage is derived from the property of the Markov chain.

Kasilingam (1996) presents an allocation model in a network setting. A total contribution is maximized subject to a capacity constraint, which ensures that for each flight the total allotment from all associated routings and for all forwarders does not exceed the