Shipping Consolidation across Two Warehouses
with Delivery Deadline and Expedited Shipment Options

The total costs due to logistics usually accounts 7.9% of the US GDP in 2015, of which shipping cost alone comprises more than 60% (27th State of Logistic Report). For a typical online retailer, Amazon.com, the shipping costs accounts for as high as 11.89% of the net sales (Amazon, 2017). One commonly used strategy to save on shipping costs is shipping consolidation, i.e., combining multiple small shipments into one large one (Cetinkaya, 2005). Although the value of consolidation is well recognized in the supply chain, its saving potential in the Business-to-Customer (B2C) setting, especially for e-commerce and omni-channel retailers has not been fully exploited yet and multiple service providers continue experimentation in this area. Retailers often have significant opportunities to take advantage of various forms of flexibility when satisfying customers' orders, especially choosing from which warehouse to fulfill the orders and when to ship the orders. In this paper, we focus on the latter, taking advantage of a time window between the time the retailer receives an order and the time by which the order must be delivered (Lee et al., 2001; Xu et al., 2009). This window provides an opportunity for the seller to combine existing orders with new incoming orders so that several orders can be shipped together, reducing total shipping costs. However, since orders must be delivered by their guaranteed due date (or deadline), delaying the shipment of some orders may increase total shipping costs due to the need to use expedited shipping. Many logistics firms such as Expedited Logistics and Freight Services, ASAP Expedited Logistics, and major carriers such as UPS, FedEx, and USPS offer shorter lead time deliveries (3 Day Select, Overnight Delivery, etc.) for extra costs.

Our work is motivated by the potential benefit of consolidation in e-commerce and omni-channel settings. For e-commerce, the frequency of orders per customer has continued to increase
while the size of each order has become smaller, partly due to the popularity of free-shipping service offered by retailers such as Neimanmarcus.com and others (Lewis, 2006; Gil, 2014). In this domain, consolidation can be, and is, implemented for individual customers. Indeed, one of the co-founders of one of the largest e-commerce retailers in China recently told us that the proportion of its customers who place two orders within an hour and one day is 5% and 20%, respectively. In addition, the proportion of customers who place three or more orders within one day and one week is 1% and 10%, respectively. Given that many orders are shipped for free, the high customer ordering frequency has intensified pressure to take advantage of such situations. To save shipping cost, online retailers can consolidate multiple orders placed by one customer, being aware that a faster, more expeditious mode of transportation may be used to meet order deadlines. Similarly, in omni-channel retail, especially for those retailers that offer one-day (or several hours) delivery guarantee, orders from different customers in nearby locations can be consolidated and executed through multiple drop-offs.

In this paper, we study the optimal shipping and consolidation policy taking into account both the delivery deadlines and the availability of expedited shipping options, in the setting of two warehouses/stores (typically considered as the primary warehouses, or local stores, for some regions). This reflects the current practice which is quite limited in terms of geographical consolidation (Acimovic and Graves, 2015). To take advantage of consolidation, the retailer may either consolidate orders for the same customer or for multiple customers living in nearby regions. Each shipment incurs both fixed and variable costs, each of which is a function of delivery speed. We consider a finite-horizon problem where, in each period, the retailer needs to make three joint decisions: (1) Which orders should be shipped? (2) From which warehouse/store should the orders be shipped? (3) How should the shipment be split into multiple packages/delivery tasks? The
structure of the optimal policy in three simplified cases is derived: (1) For one warehouse with fixed cost, we show that the optimal policy can be directly characterized by a sequence of slack-time (remaining time until the due date)-dependent thresholds. (2) For two warehouses, with fixed cost and overlapping availability of products, the optimal policy, in the symmetric settings, can be characterized by six non-linear boundaries in three-dimensional space. We show that in asymmetric settings, the optimal policy can be approximated by constant-threshold heuristics, with less than a 2% optimality gap. (3) For one warehouse and both fixed and variable costs, the optimal policy is a function of volume-dependent thresholds, which can be approximated by constant thresholds, with less than a 0.3% optimality gap. Based on these special cases, three easily implementable threshold-based heuristic policies are proposed for the setting of two warehouse and both fixed and variable costs, which significantly outperform the best benchmark policies found in practice/literature and perform near optimal, with less than a 2.5% optimality gap in the cases we tested.

Reference: